NOTES ON BUSINESS ANALYTICS

3/-/-/3

BUSINESSANALYTICSBASICS

COURSEOBJECTIVES

Tohelpstudentsinunderstandinghowthemanagersusebusinessanalyticsformanagerial decision making.

LearningOutcome/s:

Thestudentswillbe familiarwiththepracticesofanalyzingandreportingthebusinessdata useful for the insights of business growth and development.

Unit-I:UnderstandingBusinessAnalytics

Introduction:MeaningofAnalytics -EvolutionofAnalytics-NeedofAnalytics-BusinessAnalysis vs. Business Analytics - Categorization of Analytical Models - Data Scientist vs. DataEngineervs. BusinessAnalyst-BusinessAnalytics inPractice-Types ofData-RoleofBusinessAnalyst.

Unit-II:DealingwithDataandDataScience

Data: Data Collection - Data Management - Big Data Management - Organization/Sources ofData-ImportanceofDataQuality-DealingwithMissingorIncompleteData-DataVisualization-Data Classification.

DataScienceProjectLifeCycle:BusinessRequirement-DataAcquisition-DataPreparation -HypothesisandModeling-EvaluationandInterpretation-Deployment -Operations-Optimization-Applications forDataScience

Unit-III:DataMiningandMachineLearning

Data Mining: The Origins of Data Mining - Data Mining Tasks - OLAP and MultidimensionalDataAnalysis-BasicConceptofAssociationAnalysisandClusterAnalysis.

MachineLearning:HistoryandEvolution-AIEvolution-Statisticsvs.DataMiningvs. DataAnalytics vs. Data Science - Supervised Learning - Unsupervised Learning - ReinforcementLearning-FrameworksforBuildingMachine Learning Systems.

Unit-IV:ApplicationsofBusinessAnalytics

Overview of Business Analytics Applications: Financial Analytics - Marketing Analytics - HRAnalytics - Supply Chain Analytics - Retail Industry - Sales Analytics - Web & Social MediaAnalytics-HealthcareAnalytics-EnergyAnalytics-TransportationAnalytics-LendingAnalytics
-SportsAnalytics-FutureofBusinessAnalytics.

Unit-V:Ethical,LegalandOrganizationalIssues

Issues&Challenges:BusinessAnalyticsImplementationChallenges-PrivacyandAnonymizaiton-HackingandInsider Threats - MakingCustomer Comfortable.

REFERENCES:

- JamesREvans,BusinessAnalytics,GlobalEdition,PearsonEducation
- UDineshKumar,BusinessAnalytics,WileyIndiaPvt.Ltd.,NewDelhi
- GerKoole,AnIntroductiontoBusinessAnalytics,Lulu.com,2019
- J.D.Camm,J.J.Cochran,M.J.Fry,J.W.Ohlmann,D.R.Anderson,D.J.Sweeney,T.A.Williams- Essentials ofBusiness Analytics,2e;Cengage Learning.
- VipinKumar,IntroductiontoDataMining,Pang-NingTan,MichaelSteinbach,PearsonEducationIndia
- BhimasankaramPochiraju,SridharSeshadri,EssentialsofBusinessAnalytics:AnIntroductionto the Methodology anditsApplication, Springer

UnderstandingBusinessAnalytics

Introduction – Meaning of Analytics-Evolution of Analytics-Need of Analytics- Business Analytics vs. Business Analytics – Categorization of Analytical Models – Data Scientist vs. Data Engineer vs. Business Analyst – Business Analytics in practice- Types ofData- Role of Business Analyst.

Introduction

The word analytics has come into the foreground in last decade or so. The increase of the internet and information technology has made analytics very relevant in the current age. Analytics is a field which combines data, information technology, statistical analysis, quantitative methods and computer-based models into one.
This all are combined to provide decision makers all the possible scenarios to make a well thought andresearcheddecision. Thecomputer-based modelensuresthat decisionmakersare able to see performance of decision under various scenarios.

Meaning

Businessanalytics(BA) isa set ofdisciplinesand technologiesfor solving businessproblems using data analysis, statistical modelsand other quantitative methods. It involvesan iterative, methodical exploration of an organization's data, with an emphasis on statistical analysis, to drive decision-making.
Atitscore,businessanalyticsinvolvesacombinationofthefollowing:

- identifying newpatternsandrelationshipswithdatamining;
- usingquantitativeandstatisticalanalysistodesignbusiness models;
- conductingA/Band multi-variabletesting basedonfindings;
- forecasting future business needs, performance, and industry trends with predictive modelling; and
- Communicating your findings in easy-to-digest reports to colleagues, management, and customers.

Definition

- **Business analytics**(**BA**) refers to the skills, technologies, and practices forcontinuous iterative exploration andinvestigation of past businessperformance to gain insight and drive business planning. Business analytics focuses on developing new insights and understanding of business performance based on dataandstatisticalmethods.
- **Business Analytics** is the process of transforming data into insights to improve business decisions.Datamanagement,data visualization,predictivemodelling,data

mining, forecasting simulation, and optimization are some of the tools used to create insights from data.

EvolutionofBusinessAnalytics

- Business analytics has been existence since very long time and has evolved with availability of newer and better technologies. It has its roots in operations research, which was extensively used during World War II.
- Operations research was an analytical way to look at data to conduct military operations. Over a period of time, this technique started getting utilized for business. Here operation's research evolved into management science. Again, basis for management science remained same as operation research in data, decision making models, etc.
- Analytics have been used in business since the management exercises were put into place by Frederick Winslow Taylor in the late 19th century.
- Henry Ford measured the time of each component in his newly established assembly line. But analytics began to command more attention in the late 1960s whencomputers were used in decision support systems.
- Since then, analytics have changed and formed with the development ofenterpriseresource planning (ERP) systems, data warehouses, and a large number of other software tools and processes.

In later years the business analytics have exploded with the introduction of computers. This change has brought analytics to a whole new level and has brought about endlesspossibilities. As far as analytics has come in history, and what the current field ofanalytics is today, manypeople would never think that analytics started in the early1900s with Mr. Ford himself. As the economies started developing and companies became more and more competitive, managementscienceevolvedintobusinessintelligence,decisionsupportsystemsandintoPC software.

- **ScopeofBusinessAnalytics**

Business analytics has a wide range of application and usages. It can be used for descriptive analysis in which data is utilized to understand past and present situation. This kind of descriptive analysis is used to asses' current market position of the company and effectiveness of previous business decision.

It is used for predictive analysis, which is typical used to asses' previous business performance.

Business analytics is also used for prescriptive analysis, which is utilized to formulate optimization techniques for stronger business performance.

For example, business analytics is used to determine pricing of various products in a departmental store based past and present set of information.

❖ Howbusinessanalyticsworks

Beforeanydataanalysistakesplace,BAstartswithseveral foundationalprocesses:

- Determine the businessgoaloftheanalysis.
- Select ananalysismethodology.
- Get businessdatatosupporttheanalysis, oftenfromvarioussystemsandsources.
- Cleanse andintegrate data intoa single repository, such as a data warehouseor datamart.

❖ Need/ImportanceofBusinessAnalytics

- **Businessanalyticsisamethodologyortooltomakeasoundcommercialdecision**. Hence it impacts functioning ofthe whole organization. Therefore, business analytics can help improve profitability ofthe business, increase market share and revenue and provide better return to a shareholder.
- Facilitates better understanding of available primaryand secondarydata, which again affect operational efficiency of several departments.
- Provides a competitive advantage to companies. In this digital age flow ofinformation is almost equal to all the players. It is how this information is utilized makes the company competitive. Business analytics combines available data with various well thought models to improve business decisions.
- Converts available data into valuable information. This information can be presented in any required format, comfortable to the decision maker.

For starters, business analytics is the tool your company needs to make accurate decisions. These decisions are likely to impact your entire organization as they help you to improve profitability, increase market share, and provide a greater return to potential shareholders.

While some companies are unsure what to do with large amounts of data, business analytics works to combine this data with actionable insights to improve the decisions you make as a company

Essentially,the fourmainwaysbusinessanalyticsisimportant,no mattertheindustry,are:

- Improvesperformancebygivingyourbusinessaclearpictureofwhatisandisn't working
- Providesfasterandmoreaccuratedecisions
- Minimizesrisksasithelpsabusinessmaketherightchoicesregardingconsumer behaviour, trends, and performance
- Inspireschangeandinnovationbyansweringquestionsaboutthe consumer.

❖ Essentialsofbusinessanalytics

Businessanalyticshas manyusecases, but whenit comestocommercialorganizations, BAis typically used to:

- Analyzedatafromavarietyofsources.Thiscouldbeanythingfromcloud applications to marketing automation tools and CRM software.

- Use advanced analytics and statistics to find patterns within datasets. These patterns can help you predict trends in the future and access new insights about the consumer and their behaviour.
- Monitor KPIs and trends as they change in real-time. This makes it easy forbusinesses to not only have their data in one place but to also come to conclusions quickly and accurately.
- Support decisions based on the most current information. With BA providing such a vast amount of data that you can use to back up your decisions, you can be sure that you are fully informed for not one, but several different scenarios.

❖ **DataforAnalytics**

- Business analytics uses data from three sources for construction of the businessmodel. It uses business data such as annual reports, financial ratios, marketing research, etc. It uses the database which contains various computer files and information coming from data analysis.

BenefitsofimplementingBAin yourorganization

Apartfromhavingapplicationsinvariousarenas,followingarethebenefitsofBusiness Analytics and its impact on business –

- Accuratelytransferringinformation
- Consequentimprovementinefficiency
- HelpportrayFutureChallenges
- MakeStrategicdecisions
- Asaperfectblendofdatascienceandanalytics
- ReductioninCosts
- ImprovedDecisions
- Shareinformationwithalargeraudience
- EaseinSharinginformationwithstakeholders

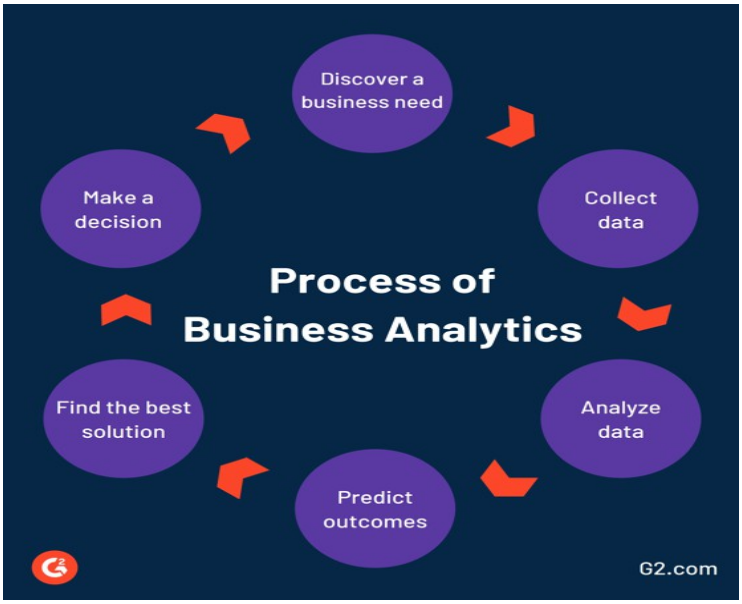
❖ **Challenges**

Moreover, anytechnologyissubject to itsownsetofproblemsand challenges. Following are the challenges in implementing business analytics in an organization.

- Lackoftechnicalskillsinemployees
- FussoveracceptanceofBAbystaff
- DataSecurityandMaintenance
- IntegrityofData
- Deliveringrelevantinformationinthegiventime
- Inabilityto addresscomplex issues
- Costsinvolvedinimplementing BA
- InvestmentofstafftimeinimplementationofBA
- Lackofaproperstrategyto implementBA

➢ Business analytics can be possible only on large volume of data. It is sometime difficult obtain large volume of data and not question its integrity.
➢ Businessanalyticsdependsonsufficientvolumesofhigh-qualitydata.

- The difficulty in ensuring data quality is integrating and reconciling data across different systems, and then deciding what subsets of data to make available.
- Previously,analyticswasconsideredatypeofafter-the-factmethod offorecastingconsumer behaviour by examining the number of units sold in the last quarter or the last year. This type of data warehousing required a lot more storage space than it did speed.
- Nowbusiness analytics is becomingatoolthat can influencetheoutcomeofcustomer interactions. When a specific customer type is considering a purchase, an analytics-enabled enterprise can modify the sales pitch to appeal to that consumer. This means the storage space for all that data must react extremely fast to provide the necessary data in real-time.



❖ Application

Businessanalyticshasawiderangeofapplicationfrom customerrelationshipmanagement,financialmanagement,and marketing,supply-chainmanagement, human-resource management, pricing and even in sports through team game strategies.
In healthcare, business analysis can be used to operate and manage clinical information systems. It can transform medical data from a bewildering array of analytical methods into useful information. Data analysis can also be used to generate contemporary reporting systems which include the patient's latest key indicators, historical trends and reference values.

- **Decisionanalytics**:supportshumandecisionswithvisualanalyticsthattheuser models to reflect reasoning.
- **Descriptive analytics:** gains insight from historical data withreporting, scorecards, clustering etc.
- **Predictiveanalytics:**employspredictivemodelling usingstatisticaland machinelearning techniques

- **Prescriptiveanalytics:**recommendsdecisionsusingoptimization,simulation,etc.

- Behaviouralanalytics
- Cohort analysis
- Competitor analysis
- Cyberanalytics
- Enterpriseoptimization
- Financialservicesanalytics
- Fraud analytics
- Healthcareanalytics
- KeyPerformanceIndicators (KPI's)
- Marketinganalytics
- Pricinganalytics
- Retailsalesanalytics
- Risk&Creditanalytics
- Supplychainanalytics
- Talentanalytics
- Telecommunications
- Transportationanalytics
- CustomerJourneyAnalytics
- MarketBasketAnalysis

❖ **BusinessAnalysisvs.BusinessAnalytics**

The aim of business analytics is data and reporting—examining past business performance and forecasting future business performance. On the other hand, the business analysisfocuseson functions and processes—determining business requirements and suggesting solutions.

- ***BusinessAnalysis:Definitionand Activities***

Business analysis is the practice of assisting firms in resolving their technical difficulties by understanding, defining, and solving those issues.

Theactivitiesthat arecarriedoutwhileperforming Business Analysis:

- **Company analysis**: Business analysis aims at figuring out the requirements ofa firm in general and its strategic direction and determining the initiatives that will enablethe business to address those strategic goals.
- **Requirements planning and management**: It focuses on planning the requirements of the development process, identifying what the top priority is for execution, and managing the changes.
- **Requirements elicitation**: It outlines techniques for collecting needs from relevant members of the project team.
- **Requirements analysis and documentation**: It explains how to establish and define the needs in detail to allow them to be effectively carried out by the team.

- **Requirements communication**: Business analysis explains methods to help stakeholders have a shared understanding of the needs and how they will be carried out.
- **Solution assessment and validation**: It also explains how a business analyst can execute a suggested solution, how to support the execution of a solution, and how to evaluate possible flaws in the implementation.

Businessanalysis isperformedbyFunctionalAnalysts,SystemsAnalysts,BusinessAnalysts, and Business Requirements Analysts.

- ***BusinessAnalytics:Definitionand Its Applications***

Business analyticsis also known as data analytics. It is a process of collecting, evaluating, and drawing valuable outcomes from the enormous amount of data available. Business analytics is widely used in the following applications:

- Finance
- Marketing
- HR
- CRM
- Manufacturing
- BankingandCredit Cards

BusinessanalyticsisperformedbyDataScientistsandData Analysts.

- ***BusinessAnalysisvs.BusinessAnalytics***

Most people believe that business analysis and analytics are the same, but theyare not!The primary differences between business analysis and business analytics:

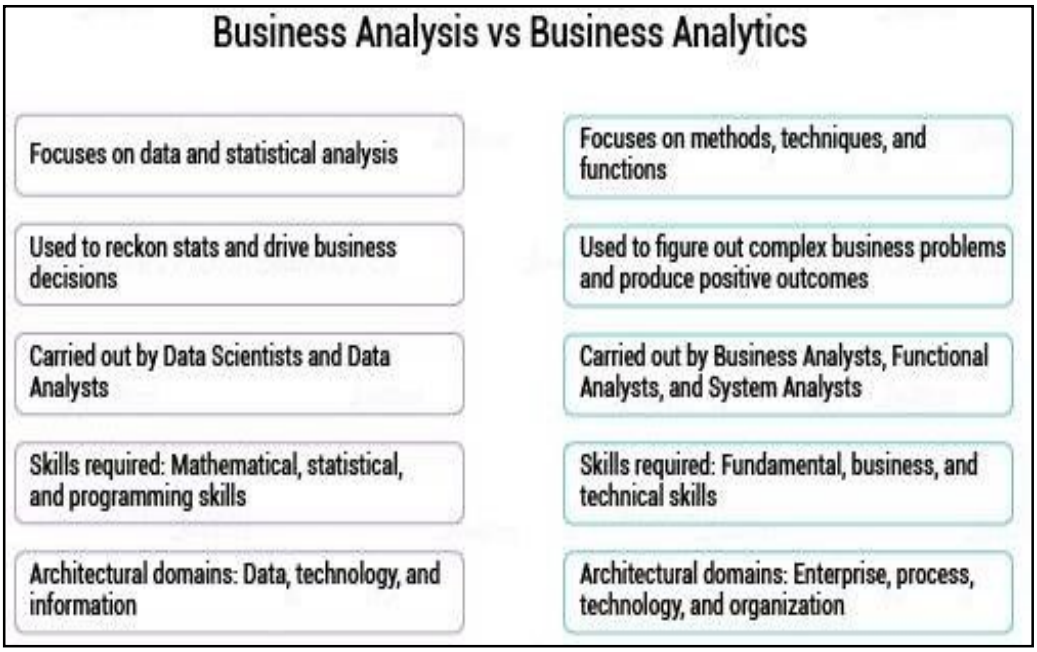
BusinessAnalysis

- Itmainlyaimsatthemethodsanddeterminingthebusinessneeds.
- It is employed tofigure out the organizational needs and possible problems to haveproductive outcomes.
- Here,thetasksarecarriedoutbyFunctionalAnalysts,SystemsAnalysts,and Business Analysts.
- Business,functional,anddomainskillsareneededtoperformbusinessanalysis.
- Thearchitecturaldomainsforbusinessanalysisincludeenterprisearchitecture, process architecture, technology architecture, and organization architecture.

BusinessAnalytics

- Itaimsatdataandreporting.
- Itiswidelypracticedtoreckonfurtherstatsandmakedecisionstobring improvements in the business.
- Here,thetasksarecarried out byDataScientistsand Data Analysts.

- Mathematical,statistical,and programming skillsare neededfor executingbusiness analytics.
- The architecturaldomains for business analytics include data architecture,technology architecture, and information architecture.



➢ ***BusinessAnalysisvs. Analytics:SimilaritiesExplained***

Businessanalysisandbusinessanalyticshavesomecommonalities.Theyboth:

- Examineandenhancebusinesses
- Determinesolutionstoissues
- Establishthingsbasedontherequirements

Businessanalysis isa practice ofidentifying businessrequirementsand figuring out solutions to specific businessproblems. This hasa heavyoverlap withthe analysisofbusinessneedsto function normally and to enhance how they function. Sometimes, the solutions include a system's development feature. It can also incorporate business change, process enhancement or strategic planning, and policy improvement.

On the contrary, business analytics is all about the group of tools, techniques, and skills that help the investigation of previous business performance. It also aids to gain insights into future performance. Ingeneral, businessanalytics aims mostlyat data and statisticalanalysis.

Categorization of Analytical Models

4TypesofBusinessAnalytics

There are mainly four types of Business Analytics, each of these types are increasingly complex. They allow us to be closer to achieving real-time and future situation insight application.Eachofthesetypesofbusinessanalyticshavebeendiscussedbelow.

1. **DescriptiveAnalytics**
2. **DiagnosticAnalytics**
3. **PredictiveAnalytics**
4. **PrescriptiveAnalytics**

1. DescriptiveAnalytics

Itsummarizesanorganisation'sexistingdata tounderstandwhathashappenedinthepast or is happening currently. Descriptive Analytics is the simplest form of analytics as it employsdataaggregationandminingtechniques.Itmakesdatamoreaccessibleto membersofanorganisationsuchastheinvestors,shareholders,marketingexecutives, and sales managers.

It can help identify strengths and weaknesses and provides an insight into customerbehaviourtoo.Thishelpsinformingstrategiesthatcanbedevelopedintheareaof targeted marketing.

2. DiagnosticAnalytics

This type of Analytics helps shift focus from past performance to the current events and determine which factors are influencing trends. To uncover the root cause of events, techniques such as data discovery, data mining and drill-down are employed. Diagnostic analytics makes use of probabilities, and likelihoods to understand why events may occur. Techniques such as sensitivity analysis and training algorithms are employed for classification and regression.

3. PredictiveAnalytics

This type of Analytics is used to forecast the possibility of a future event with the help of statisticalmodelsandMLtechniques.Itbuildson theresultof descriptiveanalyticsto devise models to extrapolate the likelihood of items. To run predictive analysis, Machine Learningexperts are employed.They can achieveahigherlevel of accuracy than by business intelligence alone.

One of the most common applications is sentiment analysis. Here, existing data collected fromsocialmediaandisusedtoprovideacomprehensivepictureofanusersopinion.

Thisdataisanalysedtopredicttheirsentiment(positive,neutral ornegative).

4. PrescriptiveAnalytics

Going a step beyond predictive analytics, it provides recommendations for the next best action to be taken. It suggests all favourable outcomes according to a specific course ofactionandalsorecommendsthespecificactionsneededtodeliverthemostdesiredresult.

Itmainlyreliesontwothings,astrongfeedbacksystemandaconstantiterativeanalysis. It learns the relation between actions and their outcomes. One common use of this type of analytics is to create recommendation systems.

Business Analytics	Questions	Tools	Outcomes	Focus
Prescriptive (Automation)	• What should I do? • Why should I do it?	• Decision modeling • Optimization • Simulation • Expert systems	• Optimization-Best possible business decisions	• Focus on decision making and efficiency
Predictive (Foresight)	• What is likely to happen? • What will happen? • Why will it happen	• Data mining • Text/media mining • Predictive modeling • Artificial Neural Networks (ANN)	• Accurate projections of the future conditions and states	• Identify past patterns to predict the future
Diagnostic (Insight)	• Why did it happen?	• Enterprise data warehouse • Data discovery • Data mining and correlations • Drill-down/roll-up	• Accurate projections of the future conditions and states	• Identify past patterns to predict the future
Descriptive (Hindsight)	• What happened? • What is happening?	• Data modeling • Business reporting • Visualization • Dashboard • Regression	• Well defined business problems or opportunities	• Uncovering patterns that offer insight

❖ BusinessAnalyticsTools

Business Analytics tools help analysts to perform the tasks at hand and generate reportswhich may be easy fora layman to understand. These tools can be obtainedfrom open sourceplatforms,andenablebusinessanalyststomanagetheirinsightsina comprehensivemanner. They tend tobe flexible and user-friendly. Various business analytics tools and techniques like.

- **Python**isveryflexibleandcanalsobeusedinwebscripting.Itismainly applied when there is a need for integrating the data analyzed with a web application or the statistics is to be used in a database production. The I Python Notebook facilitates and makes it easy to work with Python and data. One can share notebooks with other people without necessarily telling them to install anything which reduces code organizingoverhead
- **SAS** The tool has a user-friendly GUI and can churn through terabytes of datawithease.Itcomeswithanextensivedocumentationandtutorialbasewhich can help early learners get started seamlessly.
- **R** is open source software and is completely free to use making it easier for individualprofessionalsorstudentsstartingouttolearn.Graphicalcapabilities ordatavisualizationisthestrongestforteofRwithRhavingaccessto packages like GGPlot, RGIS, Lattice, and GGVIS among others which provide superior graphical competency.
- **Tableau**isthemostpopularandadvanceddatavisualizationtoolinthe market. Story-telling and presenting data insights in a comprehensive way has becomeoneofthetrademarksofacompetentbusinessanalystTableauisa greatplatformtodevelopcustomizedvisualizationsinnotime,thankstothe drop and drag features.

Python,R,SAS,Excel,andTableauhaveallgottheiruniqueplaceswhenitcomesto usage.

❖ DataScientistvs.DataEngineervs.Data Analyst

1. Data scientists use their advanced statistical skills to help improve the models the data engineers implement andtoputproper statisticalrigouronthedatadiscoveryandanalysisthe customer is asking for.

- Companies extract data to analyze and gain insights about various trends and practices. In order to do so, they employ specialized data scientists who possess knowledge of statistical tools and programming skills. Moreover, a data scientist possesses ***knowledge of machine learning algorithms***.
- However, Data Science is not a singular field. It is a quantitative field that shares its background with math, statistics and computer programming. With the help of data science, industries are qualified to make careful data-driven decisions.
- These algorithms are responsible for predicting future events. Therefore, data science can be thoughtofas anoceanthat includes allthe dataoperations like data extraction, data processing, data analysis and data prediction to gain necessary insights.

ADataScientistisrequiredtoperform**responsibilities**–

- Performingdatapre-processingthatinvolvesdatatransformationaswellasdata cleaning.
- Usingvariousmachinelearningtoolstoforecastandclassifypatternsinthedata.
- Increasingtheperformanceandaccuracyofmachinelearningalgorithmsthrough fine-tuning and further performance optimization.
- Understanding the requirements ofthe company and formulating questions that needs to be addressed.
- Usingrobuststorytellingtoolstocommunicateresultswiththeteammembers.

ForbecomingaDataScientist, youmusthavethefollowingkey**skills**–

- ShouldbeproficientwithMathand Statistics.
- Shouldbeabletohandlestructured&unstructuredinformation.
- In-depthknowledgeoftoolslikeR,Pythonand SAS.
- Wellversedinvariousmachinelearningalgorithms.
- Haveknowledgeof SQL(StructuredQueryLanguage)andNoSQL(NonStructured Query Language or not only SQL)
- MustbefamiliarwithBigDatatools.

Someofthe **tools**thatareusedbyDataScientist**are**

- **WebScraping**
- **DataAnalytics**
- **MachineLearning**
- **Reporting**

2. A Data Engineer *is a person who specializes in preparing data for analytical usage*. Data Engineeringalso involvesthedevelopmentofplatformsandarchitecturesfordataprocessing.

- Inother words,adata engineer developsthe foundation for variousdataoperations. A Data Engineer is responsible for designing the format for data scientists and analysts to work on.
- Data Engineers have to work with both structured and unstructured data. Therefore, they***need expertise in SQL***and NoSQL databases both. Data Engineers allow data scientists to carry out their data operations.
- DataEngineers have to dealwithBig Data wheretheyengage innumerousoperations like data cleaning, management, transformation, data deduplication etc.
- A Data Engineer is more experienced with core programming concepts and algorithms. The **role of a data engineer**also follows closely to that of a software engineer. This is because a data engineer is assigned to develop platforms and architecture that utilize guidelines of software development.

For example, developing a cloud infrastructure to facilitate real-time analysis of datarequires various development principles. Therefore, building an interface API is one of thejob responsibilities of a data engineer.

ToolsusedbyDataEngineers

Someofthe**tools**thatareused byDataEngineers**are**–

- ***Hadoop***
- ***Apache Spark***
- ***Kubernetes***
- ***Java***
- ***Yarn***

ADataEngineerissupposedtohavethefollowing**responsibilities–**

- Development,construction, andmaintenanceofdataarchitectures.
- Conductingtestingonlargescaledata platforms.
- Handlingerrorlogsandbuildingrobustdata pipelines.
- Abilitytohandlerawand unstructureddata.
- Providerecommendationsfordataimprovement,quality,andefficiencyofdata.
- Ensureandsupportthedataarchitectureutilizedbydatascientistsand analysts.
- Development ofdataprocesses fordatamodelling,mining,anddata production.

Followingarethekey**skills**required to becomeadataengineer–

- Knowledgeofprogrammingtools likePythonandJava.
- SolidUnderstandingofOperating Systems.
- AbilitytodevelopscalableETLpackages.
- ShouldbewellversedinSQLaswellasNoSQLtechnologieslikeCassandraand MongoDB.
- HeshouldpossessknowledgeofdatawarehouseandbigdatatechnologieslikeHadoop, Hive, Pig, and Spark.
- Shouldpossesscreativeand outoftheboxthinking.

3. A**Data Analyst** is responsible for taking actionable that affect the current scope of the company. A**data engineer**is responsible for developing a platform those data analysts and data scientists work on. And, a**data scientist**is responsible for unearthing future insights from existing data and helping companies to make data-driven decisions.

- A**dataanalyst** doesnot directlyparticipate inthe decision-making process;rather, he helps indirectlythroughproviding static insights about companyperformance. A**data engineer**is not responsible for decision making. And, a **data scientist**participates in the active decision-making process that affects the course of the company.

- A**data analyst** usesstatic modelling techniquesthat summarize the data through descriptiveanalysis.Ontheotherhand,a **dataengineer** isresponsiblefor thedevelopment and maintenance of data pipelines. A**data scientist** usesdynamic techniques like **Machine learning to gain insights about the future**.

- Knowledge of machine learning is not important for **data analysts**. However, this is mandatory for **data scientists**. A**data engineer**need not require the knowledge of machine learning but he is required to have the knowledge of core computingconcepts like programming and algorithms to build robust data systems.

- A**data analyst** only has to deal withstructured data. However, both**data scientists and data engineers** deal with unstructured data as well.

- Dataanalyst and datascientistsare bothrequired to be proficient indatavisualization. However, this is not required in the case of a **data engineer**.

- Both**data scientists and analysts** need not have knowledge of application development and working of the APIs. However, this is the most essentialrequirement for a **data engineer**.

ADataAnalysthasfollowing**responsibilities-**

- Analyzingthedatathroughdescriptive statistics.
- Usingdatabasequerylanguagestoretrieveandmanipulate information.
- Performdata filtering,cleaningand earlystage transformation.
- Communicatingresultswiththeteamusingdatavisualization.
- Workwiththemanagementteamtounderstandbusinessrequirements.

InordertobecomeaDataAnalyst,youmustpossessthefollowing**skills**–

- Shouldpossessthestrong mathematicalaptitude
- ShouldbewellversedwithExcel, Oracle, andSQL.
- Possessionofproblem-solving attitude.
- Proficientinthecommunicationofresultstotheteam.
- Shouldhaveastrong suiteofanalyticalskills.

Someofthe**tools**thatareused byDataAnalyst**are**

- Talend:Talend isoneofthe most powerfuldataanalyticstoolsavailable inthe market and is developed in the eclipse graphical development environment. ...
- QlikSense....
- ApacheSpark....
- PowerBI. ...
- ThoughtSpot....
- RapidMiner....
- Tableau

BusinessAnalyst
Businessanalystsusedatatoformbusinessinsightsandrecommendchangesinbusinessesand other organizations. Business analysts can identify issues in virtually any part of an organization,includingITprocesses,organizationalstructures,orstaffdevelopment.

As businesses seek to increase efficiency and reduce costs, business analytics has become an important componentoftheiroperations. Let'stake acloserlookat what business analystsdo and what it takesto get a job in business analysis.

Business analysts identify business areas that can be improved to increase efficiency and strengthen business processes. They often work closely with others throughout the business hierarchyto communicatetheir findings andhelp implement changes.
Tasksanddutiescaninclude:
- Identifyingandprioritizingtheorganization'sfunctionalandtechnicalneedsand requirements
- UsingSQLandExceltoanalyzelargedatasets
- Compilingcharts,tables, andotherelementsofdatavisualization
- Creatingfinancialmodelstosupportbusinessdecisions
- Understandingbusinessstrategies,goals,andrequirements
- Planningenterprisearchitecture(thestructureofabusiness)
- Forecasting,budgeting,andperformingbothvarianceanalysisandfinancialanalysis

Businessanalystskills

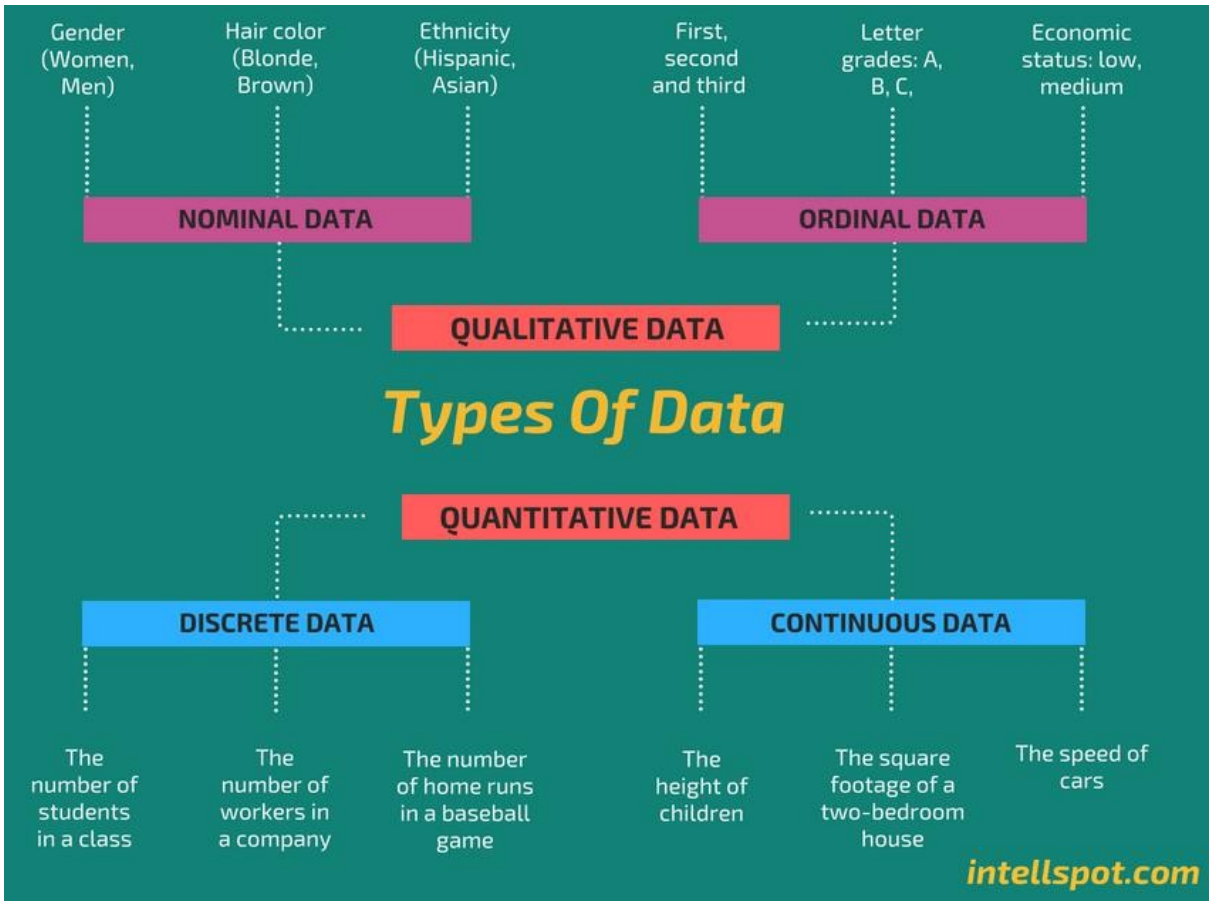
Thekeyskillsbusinessanalystsneed are:

- **Technicalskills:** Theseskillsincludestakeholdermanagement,datamodelingand knowledge of IT.
- **Analytical skills:** Business analysts have to analyze large amounts of data and other business processes to form ideas and fix problems.
- **Communication:**These professionals must communicate their ideas inanexpressive way that is easy for the receiver to understand.
- **Problem-solving:** Itisabusinessanalyst'sprimary responsibility tocomeupwith solutions to an organization's problems.
- **Researchskills:** Thoroughresearchmustbeconductedaboutnewprocessesand software to present results that are effective.

Businessanalystresponsibilities

- Analyzingandevaluatingthecurrentbusinessprocessesacompanyhasand identifying areas of improvement
- Researching and reviewing up-to-date business processes and new IT advancementsto make systems more modern
- Presentingideasandfindingsinmeetings
- Trainingandcoachingstaffmembers
- Creatinginitiativesdependingonthebusiness'srequirementsand needs
- Developingprojectsandmonitoringprojectperformance
- Collaboratingwithusersand stakeholders
- Workingcloselywithseniormanagement,partners,clientsandtechnicians

Typesof Data
Qualitativevs.Quantitative Data



1. Quantitativedata

- Quantitative data seems to be the easiest to explain. It answers key questions such as "how many, "how much" and "how often".
- Quantitative data can be expressed as a number or can be quantified. Simply put, it can be measured by numerical variables.
- Quantitative data are easily amenable to statistical manipulation and can be represented by a wide varietyof statisticaltypes ofgraphsand charts such as line, bar graph, scatter plot, and etc.

Examplesofquantitativedata:

- Scoresontests andexams e.g.85,67,90andetc.

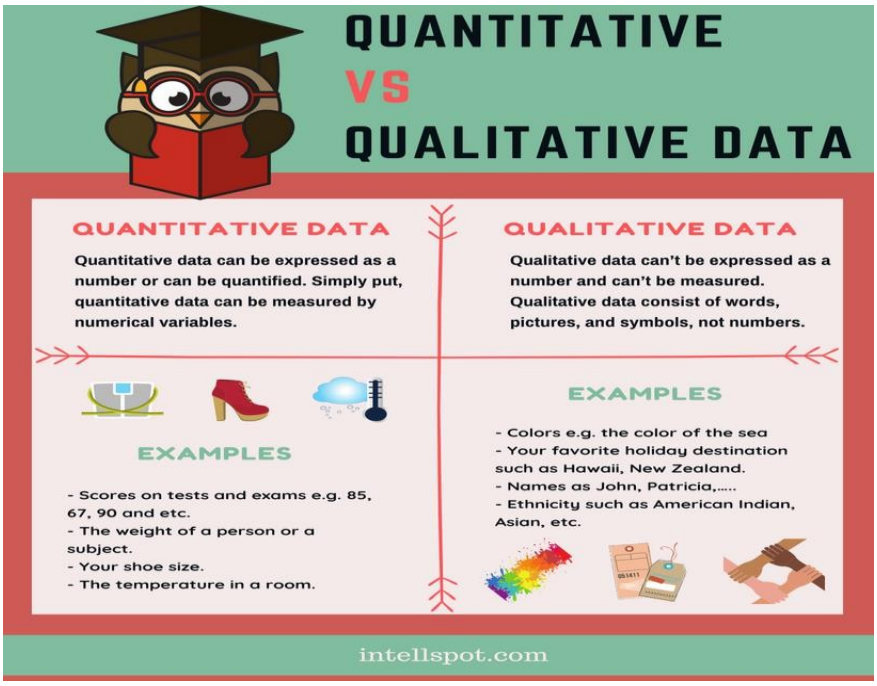
- The weightofapersonorasubject.
- Yourshoesize.
- Thetemperatureinaroom.

2. Qualitativedata

- Qualitative data can’tbe expressed as a number and can’tbemeasured.Qualitative data consist of words, pictures, and symbols, not numbers.
- Qualitative data is also called categorical databecause the information can be sorted by category, not by number.
- Qualitative data can answer questions such as “how this has happened” or and “why this has happened”.

Examplesofqualitative data:

- Colorse.g.the coloroftheh sea
- YourfavoriteholidaydestinationsuchasHawaii,NewZealandand etc.
- NamesasJohn,Patricia..
- EthnicitysuchasAmericanIndian,Asian,etc.



Nominalvs.OrdinalData

3. Nominaldata

Nominal data is used just for labelling variables, without any type of quantitative value. The name ‘nominal’ comes from the Latin word “nomen” which means ‘name’.

The nominal data justname a thing without applying it to order. Actually, the nominal data could just be called “labels.”

ExamplesofNominalData:

- Gender(Women,Men)
- Haircolor(Blonde,Brown,Brunette,Red,etc.)
- Maritalstatus(Married, Single,Widowed)
- Ethnicity(Hispanic,Asian)

Eye color isa nominal variable having a few categories(Blue, Green, Brown) and there is no way to order these categories from highest to lowest.

4. Ordinaldata

Ordinal data shows where a numberis in order. This is the crucial differencefrom nominaltypes of data.

Ordinaldataisdatawhichisplacedintosomekindoforderbytheirpositiononascale. Ordinal data may indicate superiority.

However, **youcannotdoarithmeticwithordinalnumbers**becausetheyonlyshow sequence.

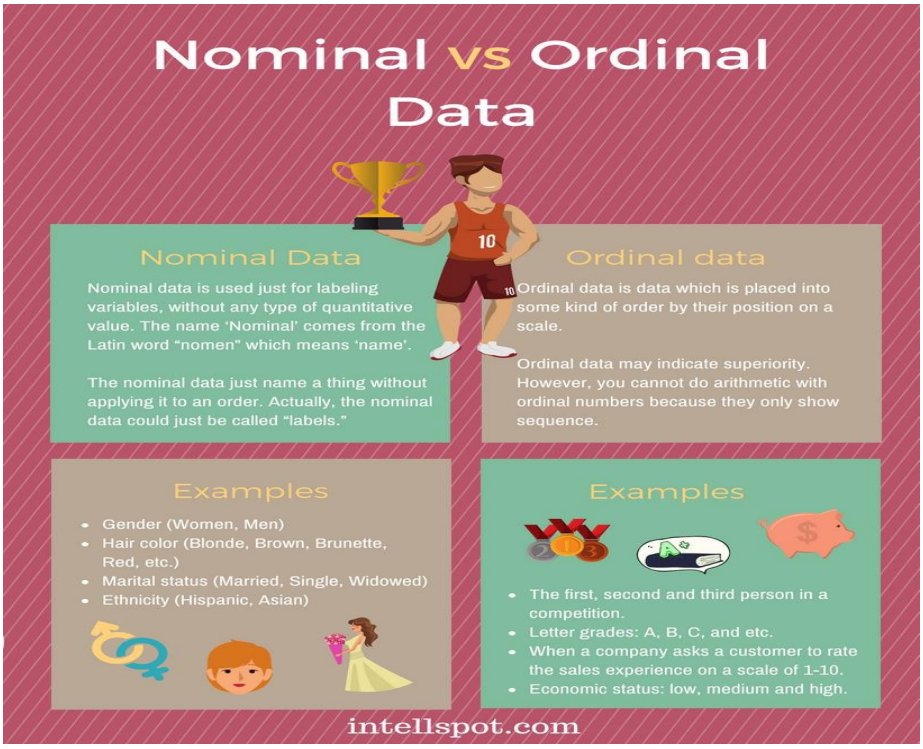
Ordinalvariablesareconsideredas“inbetween”qualitativeandquantitativevariables. In other words, the ordinal data is qualitative data for which the values are ordered.

Incomparisonwithnominaldata,thesecondoneisqualitativedataforwhichthevalues cannot be placed in an ordered.

We can also assign numbers to ordinaldata to show their relative position. But we cannot do math with those numbers. For example: “first, second, third…etc.”

ExamplesofOrdinalData:

- Thefirst,secondandthirdpersonina competition.
- Lettergrades:A, B, C,andetc.
- Whena companyasksa customertoratethesalesexperience onascale of1-10.
- Economicstatus:low,mediumandhigh.



Discretevs.ContinuousData

Instatistics,marketingresearch,anddatascience,manydecisionsdependonwhetherthe basic data is discrete or continuous.

5. Discretedata

Discrete data is a count that involves only integers. The discrete values cannot be subdivided into parts.

Forexample,thenumberofchildreninaclassisdiscretedata.Youcancountwhole individuals. You can’t count 1.5 kids.

Toput inother words, discretedatacantakeonlycertainvalues. Thedatavariablescannot be divided into smaller parts.
Ithasalimitednumber ofpossiblevaluese.g. daysofthemonth.

Examplesofdiscrete data:

- Thenumberofstudentsina class.
- Thenumberofworkersina company.
- Thenumber ofhomerunsinabaseballgame.
- Thenumberoftest questions youanswered correctly

6. Continuousdata

Continuous datais informationthat could be meaningfully divided into finer levels. It canbe measured on a scale or continuum and can have almost any numeric value.
For example, you can measure your height at very precise scales — meters, centimeters, millimeters and etc.
You can record continuous data at so many different measurements – width, temperature, time, and etc. This is where the key difference from discrete types of data lies.
The continuous variables cantake anyvalue betweentwo numbers. For example, between50 and 72 inches, there are literally millions of possible heights: 52.04762 inches, 69.948376 inches and etc.
A good great rule for defining if a data is continuous or discrete is that if the point of measurement can be reduced in half and still make sense, the data is continuous.

Examplesofcontinuousdata:

- Theamountoftimerequiredtocompletea project.
- Theheightofchildren.
- Thesquarefootageofatwo-bedroomhouse.
- Thespeed ofcars.

Conclusion

Allof the different types of data have a criticalplace in statistics, research, and data science. Data types work great together to help organizations and businesses fromall industries build successful data-driven decision-making process.

Working inthedata management areaand having agoodrangeofdatascienceskills involves a deep understanding of various types of data and when to apply them.

❖ ROLES OFA BUSINESSANALYST

1.BALEVELS

Therearefourlevelsthatabusinessanalyst in anorganizationcomprisesof:

- **Strategic management:** This is the analysis level, where a businessanalyst evaluates and calculates the strategic where about if a company. This is one of the most critical levels because unless the evaluation is doneonthe point, noneofthe further steps can work appropriately.
- **Analysis of business model:** This level has to do with evaluating policies that are currently being employed by the company. This not only enables us to implement what's new but also helps in checking the previous ones.
- **Designing the process:** Like an artist creates his imagination, business analysts do that with their skills. The step includes modelling the business processes, whichcomes out to be designing and modelling.
- **Analysis of technology:** Technical systems need a thorough analysis too. This is something that, if not taken care of, leads to severe consequences.

Thekey**businessanalystrolesandresponsibilities**:

- ✓ **What does a business needs:** As a business analyst, it is his key responsibility to understand what stakeholdersneed and passthese requirementstothe developers, and also give on the developer's expectations to the stakeholders. A business analyst's skill for this responsibility is the communication skills that can impress everyone across. While he transfers the information, he is the one who needs to put these insuch words that make a difference. This responsibility is no doubt tome takingbecause he needs to listen and execute, which might seem easy, but only a skilled professional can handle all this.
- ✓ **Conducting meetings with developing team and stakeholders:** Business analysts are supposed to coordinate with both stakeholders and the development team whenever a new featureor updateis added to aproject. This may varyfromproject to project. This facilitates the collection of client feedback and the resolution of issues encounteredbythedevelopmentteamwhenimplementingnewfeatures.The**business analyst role** is to understand and explain the new feature updates to clientsand takefeedback forfurther development.Based onclient feedback, Business Analyst instructs the development team to make amendments or continue as is. At times, the client requests anadditional feature be added to a project, and the BA must determine whether or not it is feasible, and then assign resources if necessary to implement it.
- ✓ **System possibilities:** A business analyst might be considered one among those working in the software team, but their key responsibility Is not what the team does. He has to ensure that he figures out what a project needs. He is the one wholeads the

path to the goals. He might be the one who dreams of targets, but he is also the one who knows how to make those dreams a reality. Looking for the opportunities and grabbing them before they go is what a business analyst is good at.

- ✓ **Present the company:** He can be called the face of a business. A business analyst is responsible for putting a business's thoughts and goals in front ofthe stakeholders. In short, he is the one who needs to impress the stakeholders with his presentation skills and the skill to present what the person on the other side is looking for and not what the company has in store for them.
- ✓ **Present thedetails:**Aproject bringswith itself hundredsofminute detailsthat might be left unseen. A business analyst is the one who is responsible for elaborating the project with the tiniest ofthe loopholes or hidden secrets. This is considered the most crucial role of a business analyst because unless the details are put across the stakeholders, theywon't take an interest, and unless theyshow the part, the project is likely to take a pause.
- ✓ **Implementation of the project:** After going through all the steps mentioned above, the next and the most important role of a business analyst in agile is to implement whatever has beenplanned. Execution is not easyunless the previous steps have been taken care of in a systemized fashion.
- ✓ **Functionalandnon-functionalrequirements ofa business:** Asanorganization, the main goal is to receive an end product that is productive and gives a company a long time.The role of business analyst in it company is to take care of the business's functional aspect, which includes the steps and ways to ensure the working of the project. Sideways he is also supposed to take careofthe non-functionalthat comprise how a project or a business is supposed to work.
- ✓ **Testing:** The role ofa business analyst is waylongerthanexpected. Oncethe product is prepared, the next step is to test it among the users to know it's working capacity and quality. The Business Analyst tests the prototype/interface by involving some clients and recording their experiences with the model that has been developed, according to the role description. Based on their feedback, Business Analyst intendsto make some changes to the modelthat will make it even better. Theyconduct UAT (user acceptance test) to determine whether or not the prototype meets the requirements of the project under consideration.
- ✓ **Decisionmakingandproblem-solving:** The **responsibilitiesofbusinessanalyst** range from developing the required documents to making decisions in the most stringent circumstances, job role of business analyst is**to**do it all. Moreover, a business analyst is expected to be the one who tackles things most easily and calmly because he should also be good at problem-solving, even if that's related to the stakeholders, employees, or the clients.
- ✓ **Maintenance:** Like they say that care is as essential as building something new. No matter how much human resources, energy, or finds you spend on a project, if the maintenance part is not taken care of properly or is neglected, it tends to spoil the entire hard work put across. What is the role of a business analyst here? Is it just limited to the maintenance of the clients or sales; it also has to ensure that the quality and the promised products are maintained throughout.
- ✓ **Building a team:** Everyone is born with varied skills. As a business analyst, the business analyst's responsibility is to make the team with people possessing different skills required for the project. Not onlythe hiring but retaining them is as essential. A wellunitedandskilledteam candowonders.Thethingsthatarerequiredinagreat

section inside co combination, structuring, and skills. A good team tends to take the company to the heights of success.

- ✓ **Presentation and Documentation of the Final Project**:After the business project is completed, the Business Analyst must document the details of the project and share theproject'sfindingswiththeclient.Inmostcases, **BArolesandresponsibilities** include preparing reports and presenting the results of a project tokeystakeholders and clients. During building the project, they must also record all of the lessons learned and challenges they encountered in a concise form. This step aids the business analyst in making better decisions in the future.

CONCLUSION

A business analyst might be another position in an organization but its roles and responsibilities play a vital role in an organization's success. While he needs to be a good orator, he should possess the quality of bringing people closers to his team and across. His rolesare not limited to a specific step inproject management. He isrequired one overstep till the end. Fromthe initial stages ofevaluation to the maintenance, a company needs a business analyst's skill.

UNIT-II
DealingwithDataandDataScience

Data: Data Collection-Data Management-Big Data Management-Organization/sources of Data- Importance of Data Quality- Dealing with missing or incomplete data – Data Visualization- Data Classification.
Data Science project Life Cycle- Business Requirement – Data Acquisition- data Preparation- Hypothesis and Modelling- Evaluation and interpretation- Deployment-Operations-Optimization-Applications for Data Science.

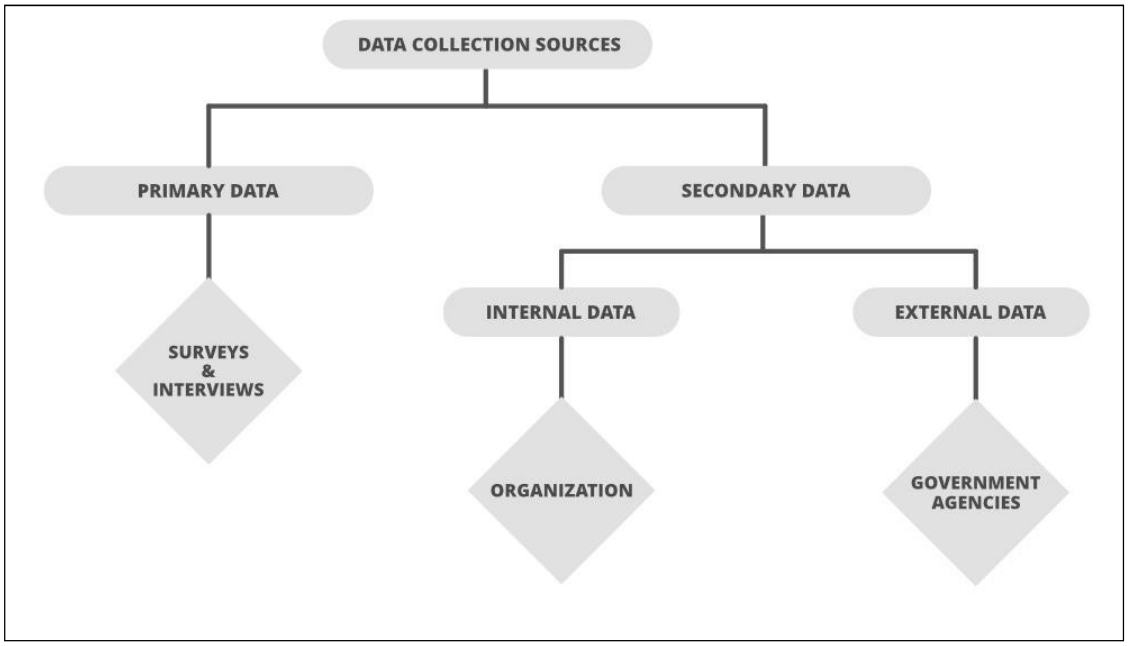
Data

- Knowledge is power, information is knowledge, and data is information in digitized form, at least as defined in IT. Hence, data is power.
- Dataare individual facts, statistics, or items of information, often numeric. In a more technical sense, dataare a set of values of qualitative orquantitativevariablesabout one or more persons or objects
- Data is various kinds of information formatted in a particular way. Therefore, data collection is the process of gathering, measuring, and analyzing accurate data from a variety of relevant sources to find answers to research problems, answer questions, evaluate outcomes, and forecast trends and probabilities.
- Accurate data collection is necessary to make informed business decisions, ensure quality assurance, and keep research integrity.
- The concept of data collection isn't a new one, as we'll see later, but the world has changed. There is far more data available today, and it exists in forms that were unheard of a century ago. The data collection process has had to change and growwith the times, keeping pace with technology.
- Datacollectionbreaksdownintotwomethods: 1.Primary&2.Secondary

❖ **Data Collection**

Datacollectionistheprocessofacquiring,collecting,extracting,andstoringthe voluminous amount of data which may be in the structured or unstructured form like text, video,audio,XMLfiles,records,orotherimagefilesusedinlaterstages ofdataanalysis. In the process of big data analysis, "Data collection" is the initial step before starting to analyzethepatternsorusefulinformationindata.Thedatawhichistobeanalyzedmust be collected from different valid sources.
Theactualdataisthenfurtherdividedmainlyintotwotypesknownas:

1. **Primarydata**
2. **Secondarydata**



1. Primarydata:

ThedatawhichisRaw,original,andextracteddirectlyfrom theofficialsourcesisknown as primary data. This type of data is collected directly by performing techniques such as questionnaires,interviews,andsurveys.Thedatacollectedmustbeaccordingtothe demandandrequirementsofthetargetaudienceonwhichanalysisisperformedotherwise it would be a burden in the data processing.

Fewmethodsofcollectingprimarydata:

- **Interviewmethod:**
  The data collectedduringthisprocess is throughinterviewingthe targetaudienceby a person called interviewer and the person who answers the interview is known as the interviewee. Some basic business or product related questions are asked and noted down in the form of notes, audio, or video and this data is stored for processing. These can be both structuredandunstructuredlikepersonalinterviewsorformalinterviewsthrough telephone, face to face, email, etc.
- **Surveymethod:**
  Thesurveymethodistheprocessof researchwherealistof relevantquestionsareasked and answers are noted down in the form of text, audio, or video. The survey method can be obtained in both online and offline mode like through website forms and email. Then that survey answers are stored for analyzing data. Examples are online surveys or surveysthrough social media polls.
- **Observationmethod:**
  The observation method is a method of data collection in which the researcher keenly observes the behaviour and practices of the target audience using some data collecting tool and stores the observed data in the form of text, audio, video, or any raw formats. In this method, the data is collected directly by posting a few questions on the participants. For example,observingagroupofcustomersandtheirbehaviourtowardstheproducts.The data obtained will be sent for processing.
- **ProjectiveTechnique**
  Projective data gathering isan indirect interview, used whenpotentialrespondentsknow why they're being asked questions and hesitate to answer. For instance, someone may be reluctant

to answer questions about their phone service if a cell phone carrier representative poses the questions. With projective data gathering, the interviewees get an incomplete question, and they must fill in the rest, using their opinions, feelings, and attitudes.

- **DelphiTechnique.**

The Oracle at Delphi, according to Greek mythology, was the high priestess of Apollo's temple, who gaveadvice, prophecies, andcounsel. Intherealmofdatacollection, researchers use the Delphi technique by gathering information from a panel of experts. Each expert answers questions in their field of specialty, and the replies are consolidated into a single opinion.

- **FocusGroups.**

Focus groups, like interviews, are a commonly used technique. The group consists of anywhere from a half-dozen to a dozen people, led by a moderator, brought together to discuss the issue.

- **Questionnaires.**

Questionnairesareasimple, straightforwarddatacollectionmethod. Respondentsget aseries of questions, either open or close-ended, related to the matter at hand.

- **Experimentalmethod:**

Theexperimentalmethodistheprocessofcollectingdatathroughperforming experiments,research,andinvestigation.Themostfrequentlyusedexperimentmethods are CRD, RBD, LSD, FD.

- **CRD-CompletelyRandomizeddesign** isasimpleexperimentaldesignusedin data analytics which is based on randomization and replication. It is mostly used for comparing the experiments.
- **RBD-RandomizedBlockDesign** isanexperimentaldesigninwhichthe experimentisdividedintosmallunitscalledblocks.Randomexperimentsare performed on each of the blocks and results are drawn using a technique known as analysisof variance(ANOVA).RBD was originated from theagriculture sector.
- **LSD – Latin Square Design** is an experimental design that is similar to CRD and RBDblocks butcontainsrows andcolumns.Itisan arrangementof NxNsquares with an equal amount of rows and columns which contain letters that occurs only once in a row. Hence the differences can be easily found with fewer errors in the experiment. Sudoku puzzle is an example of a Latin square design.
- **FD- Factorial design** is an experimental design where each experiment has two factorseachwithpossiblevaluesandonperformingtrailothercombinationalfactors are derived.

2. Secondarydata:

Secondarydataisthedatawhichhasalreadybeencollectedandreusedagainforsome validpurpose. Thistype of data is previously recorded from primary data and ithas two types of sources named internal source and external source.

i. **Internalsource:**

These types of data can easily befoundwithin theorganization such asmarket record,a sales record, transactions, customer data, accounting resources, etc. The cost and time consumption is less in obtaining internal sources.

- FinancialStatements
- SalesReports
- Retailer/Distributor/DealFeedback
- CustomerPersonalInformation(e.g.,name,address,age,contactinfo)
- BusinessJournals
- GovernmentRecords(e.g., census,taxrecords,SocialSecurityinfo)
- Trade/BusinessMagazines
- Theinternet

ii. **Externalsource:**

The data which can't be found at internal organizations and can be gained through external thirdpartyresourcesisexternalsourcedata.Thecostandtimeconsumptionismore becausethiscontainsahugeamountofdata.Examplesofexternalsourcesare Government publications, news publications, Registrar General of India, planning commission, international labour bureau, syndicate services, and other non-governmental publications.

iii. **Othersources:**

- **Sensors data:** With the advancement of IoT devices, the sensors of these devices collectdatawhichcanbeusedforsensordataanalyticstotracktheperformance and usage of products.
- **Satellitesdata:** Satellitescollectalotofimagesanddatainterabytesondaily basisthrough surveillancecameras which can be used to collectusefulinformation.
- **Webtraffic:**Duetofastandcheapinternetfacilitiesmanyformatsofdata Whichis uploaded by users on differentplatforms can be predicted and collected with their permission for data analysis. The search engines also provide their data through keywords and queries searched mostly.

❖ **DataCollectionTools**

1. WordAssociation.

The researcher gives the respondent a set of words and asks them what comes to mind when they hear each word.

2. SentenceCompletion.

Researchers use sentence completion to understand what kind of ideas the respondent has. This tool involves giving an incomplete sentence and seeing how the interviewee finishes it.

3. Role-Playing.

Respondentsarepresented withan imaginarysituationand asked howtheywould act or react if it was real.

4. In-PersonSurveys.

Theresearcherasksquestionsinperson.

5. Online/WebSurveys.

These surveys are easyto accomplish, but some users may be unwilling to answer truthfully, if at all.

6. MobileSurveys.

These surveys take advantage of the increasing proliferation of mobile technology. Mobile collection surveys relyon mobile devices like tablets or smart phones to conduct surveys via SMS or mobile apps.

7. PhoneSurveys.
No researcher can call thousands of people at once, so they need a third party to handle the chore. However, many people have call screening and won't answer.

8. Observation.
Sometimes, the simplest method is the best. Researchers who make direct observationscollect data quickly and easily, with little intrusion or third-party bias. Naturally, it's only effective in small-scale situations.

❖ **DataManagement**

Data management refers to the professional practice of constructing and maintaining a framework for ingesting, storing, mining, and archiving the data integral to a modern business. Data management is the spine that connects all segments of the information lifecycle.

Data management works symbiotically with process management, ensuring that the actions teams take are informed by the cleanest, most current data available — which in today's world means tracking changes and trends in real-time. Below is a deeper look at the practice, its benefits and challenges, and best practices for helping your organization get the most out of its business intelligence.

❖ **7types ofdata management**
Data management experts generally focus on specialties within the field. These specialtiescan fall under one or more of the following areas:
1. **Master data management:** Master data management (MDM) is the process of ensuring the organization is always working with — and making business decisions based on — a single version of current, reliable information. Ingesting data from all of your data sourcesand presenting it as one constant, reliable source, as well as repropagating data into different systems, requires the right tools.

2. Data stewardship: Adata steward doesnot develop information management policies but rather deploys and enforces them across the enterprise. As the name implies, a data steward stands watch over enterprise data collection and movement policies, ensuring practices are implemented and rules are enforced.

3. Data quality management: If a data steward is a kind of digital sheriff, a data quality manager might be thought of as his court clerk. Quality management is responsible for combing through collected data for underlying problems like duplicate records, inconsistent versions, and more. Data quality managers support the defined data management system.

4. Data security: One of the most important aspects of data management today is security. Thoughemergent practices like DevSecOps incorporatesecurityconsiderationsat everylevel ofapplicationdevelopmentanddataexchange,securityspecialistsarestilltaskedwith

encryption management, preventing unauthorized access, guarding against accidental movement or deletion, and other frontline concerns.

5. Datagovernance: Data governance sets the law for anenterprise's stateofinformation. A data governance framework is like a constitution that clearly outlines policies for the intake, flow, and protection of institutional information. Data governors oversee their network of stewards, quality management professionals, security teams, and other people and data management processes in pursuit of a governance policy that serves a master data management approach.

6. Big data management: Big data is the catch-all term used to describe gathering, analyzing, and using massive amounts of digital information to improve operations. In broad terms, this area of data management specializes in intake, integrity, and storage of the tide of raw data that other management teams use to improve operations and security or inform business intelligence.

7. Data warehousing: Information is the building block of modern business. The sheer volume of information presents an obvious challenge: What do we do with all these blocks? Data warehouse management provides and oversees the physical and/or cloud-based infrastructure usedto aggregate rawdataand analyze it in-depthto produce business insights. The unique needs of any organization practicing data management may require a blend of some or all of these approaches. Familiarity with management areas provides data managers with the background they need to build solutions customized for their environments.

❖ **Benefitsofdatamanagement systems**

Data management processes help organizations identify and resolve internal pain points to deliver a better customer experience.

First, data management provides businesses with a way of measuring the amount of data in play. A myriad of interactions occur in the background of any business — between network infrastructure, software applications, APIs, security protocols, and much more — and each presents a potential glitch (or time bomb) to operations if something goes wrong. Data management gives managers a big-picture look at business processes, which helps with both perspective and planning.

Once data is under management, it can be mined for informational gold: businessintelligence. This helps business users across the organization in a varietyof ways, including the following:

- Smartadvertisingthattargetscustomersaccordingtotheirinterestsandinteractions
- Holisticsecuritythatsafeguardscritical information
- Alignmentwithrelevant compliancestandards,savingtimeand money
- Machine learning that grows more environmentally aware over time, powering automatic and continuous improvement
- Reduced operating expenses byrestricting use to onlythe necessarystorage and compute power required for optimal performance

❖ **Datamanagement challenges**

- All these benefits don't come without climbing some hills. The ever-growing, rolling landscape of information technology is constantly changing and data managers will encounter plenty of challenges along the way.
- Therearefour keydatamanagementchallengestoanticipate:

- **The amount of data can be (at least temporarily) overwhelming.** It's hard to overstate the volume of data that must come under management in a modern business, so, when developing systems and processes, be ready to think big. Really big. Specialized third-party services and apps for integrating big data or providing it as a platform are crucial allies.
- **Many organizations silo data.** The development team may work from one data set, the sales team fromanother, operations fromanother, and so on. A modern data management systemreliesonaccesstoallthisinformationtodevelop modernbusinessintelligence. Re
- Real-time data platform services help stream and share clean information between teams from a single, trusted source.
- **The journey from unstructured data to structured data can be steep.**Data oftenpours into organizations in an unstructured way. Before it can be used to generatebusiness intelligence, data preparation has to happen: Data must be organized, de- duplicated, and otherwisecleaned. Datamanagersoftenrelyonthird-partypartnershipsto assist with these processes, using tools designed for on-premises, cloud, or hybrid environments.
- **Managing the culture is essential to managing data.**All of the processes and systems in the world won't do you much good if people don't know how — and perhaps just as importantly, why — to use them. By making team members aware ofthe benefits of data management (andthepotentialpitfalls ofignoring it) and fosteringtheskills ofusingdata correctly, managers engage team members as essentialpieces of the information process.

These and other challenges stand between the old way of doing business and initiatives that harness the power of data for business intelligence. But with proper planning, practices, and partners, technologies likeaccelerated machine learning can turn pinch points into gateways for deeper business insights and better customer experience.

❖ **Datamanagementbestpractices**

Thoughspecific data needs are unique to everyorganization's data strategyand datasystems, preparing a framework will smooth the path to easier, more effective data management solutions. Best practices like the three below are key to a successful strategy.

1. Makeaplan
2. Storeyourdata
3. Shareyourdata

1. Makeaplan

- Develop and write a data management plan (DMP). This document charts estimated data usage, accessibility guidelines, archiving approaches, ownership, and more. A DMP serves as both a reference and a living record and will be revised as circumstances change.

- Additionally, DMPs present the organization's overarching strategy for data management to investors, auditors, and other involved parties — which is an important insight into a company's preparedness for the rigors of the modern market.

ThebestDMPsdefinegranulardetails,including:

- Preferredfile formats
- Namingconventions
- Accessparametersforvariousstakeholders
- Backupandarchiving processes
- Defined partnersandthetermsandservicestheyprovide
- Thoroughdocumentation
- TherareonlineservicesthatcanhelpcreateDMPsbyprovidingstep-by-step guidance to creating plans from templates.

2. Storeyourdata

- Among the granular details mentioned above, a solid data storage approach is central to good data management. It begins by determining if your storage needs best suit a data warehouse or a data lake(or both), and whether the company's data belongs on-premises or in the cloud.
- Then outline a consistent, and consistently enforced, agreement for naming files, folders, directories, users, and more.This isa foundationalpiece ofdatamanagement, as these parameters will determine how to store all future data, and inconsistencies will result in errors and incomplete intelligence.

1. **Security and backups.** Insecure data is dangerous, so security must be considered at every layer. Some organizations come under special regulatory burdens like HIPAA, CIPA, GDPR, and others, which add additional security requirements like periodic audits. When security fails, the backup plan can be the difference between business life and death. Traditional models called for three copies of all important data: the original, the locally stored copy, and a remote copy. But emerging cloud models include decentralized data duplication, with even more backup options available at an increasingly affordable cost for storage and transfer.
2. **Documentation is key.** If it's important, document it. If the entire team splits the lottery and runs off to Jamaica, thorough, readable documentation outlining security and backup procedures will give the next team a fighting chance to pick up wheretheyleft off. Without it, knowledge resides exclusivelywith holders who mayor may not be part of a long-term data management approach.

Data storage needs to be able to change as fast as the technology demands, so any approach should be flexible and have a reasonable archiving approach to keep costs manageable.

3. Shareyourdata

Afterallthe plansare laid forstoring,securing,anddocumenting yourdata,youshould begin the process of sharing it with the appropriate people.

Herearesomecriticalquestionstoanswerbeforeotherpeopleaccesspotentiallycritical information:

- Whoownsthedata?
- Canitbecopied?
- Haseveryonecontributingtothedataconsentedtoshareitwithothers?
- Who canaccessitandatwhat times?

- Are there copyrights, corporate secrets, proprietary intellectualproperty, or other off-limits information in the data set?
- Whatelsedoestheorganization'sdatarevealabout itself?

With those and other questions answered, it's time to find a place and means of sharing the data.Once called a repository, thisrole is increasinglyfilled bysoftware and infrastructure as service models that are fine-tuned for big data management.

❖ Big DataManagement

Big data consists of huge amounts of information that cannot be stored or processed using traditional data storage mechanisms or processing techniques. It generally consists of three different variations.

i. **Structured data** (as its name suggests) has a well-defined structure and follows a consistent order. This kind ofinformation is designed so that it can be easilyaccessed and used by a person or computer. Structured data is usually stored in the well-defined rows and columns of a table (such as a spreadsheet) and databases — particularly relational database management systems, or RDBMS.

ii. **Semi-structureddata** exhibitsa fewofthe same propertiesas structureddata, but for the most part,thiskind ofinformationhas no definite structureand cannot conformto the formal rules of data models such as an RDBMS.

iii. **Unstructured data** possesses no consistent structure across its various forms anddoes not obey conventional data models' formal structural rules. In very few instances, it may have information related to date and time.

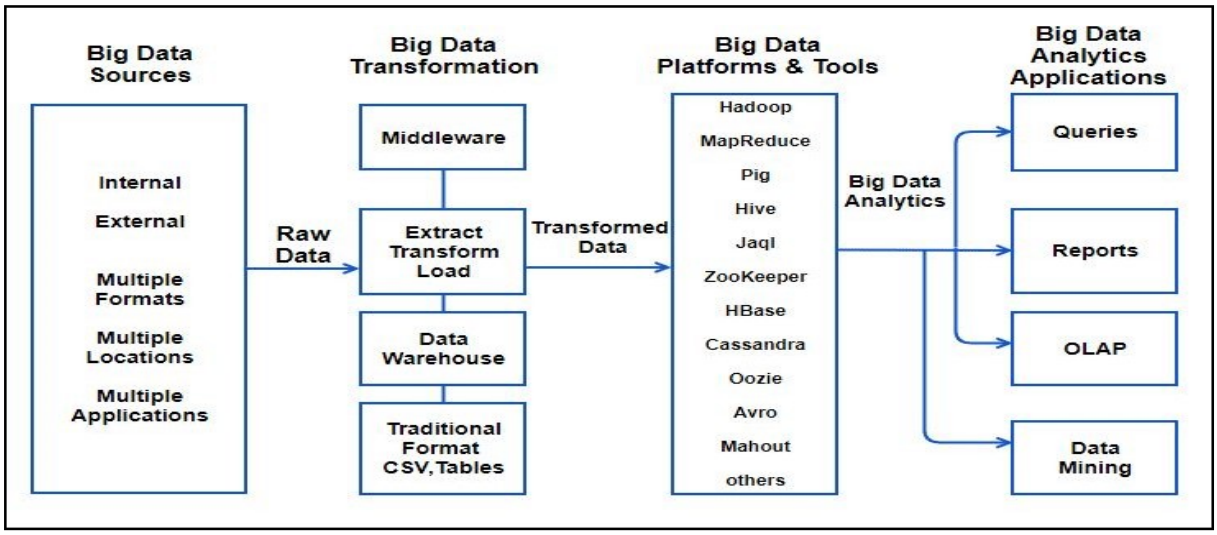
CharacteristicsofBigDataManagement

Inlinewithclassicaldefinitionsoftheconcept,bigdataisgenerallyassociatedwith three core characteristics:

1. **Volume:** This trait refers to the immense amounts of information generated every second via social media, cell phones, cars, transactions, connected sensors, images, video, and text. In petabytes, terabytes, or even zettabytes, these volumes can only be managed by big data technologies.
2. **Variety:** To the existing landscape of transactional and demographic data such as phone numbers and addresses, information inthe formofphotographs, audio streams, video, and a host of other formats now contributes to a multiplicity of data types — about 80% of which are completely unstructured.
3. **Velocity:**Information is streaming into data repositories at a prodigious rate, and this characteristic alludes to the speed of data accumulation.It also refers to the speedwithwhich big datacanbe processed and analyzedto extractthe insights and patterns it contains. These days, that speed is often real-time.

Beyond "the Three Vs," current descriptions of big data management also include two other characteristics, namely:

- **Veracity:** This is the degree of reliability and truth that big data has to offer in terms of its relevance, cleanliness, and accuracy.
- **Value:** Sincetheprimaryaimofbigdatagatheringandanalysis istodiscover insights that can inform decision-making and other processes, this characteristic explores the benefit or otherwise that information and analytics can ultimately produce.



BigDataManagementServices

When it comes to technology, organizations have many different types of big data management solutions to choose from. Vendors offer a variety of standalone or multi-featured big data management tools, and manyorganizations use multiple tools. Some of the most common types of big data management capabilities include the following:

- **Datacleansing:**findingandfixingerrorsindatasets
- **Dataintegration:**combining datafromtwo ormore sources
- **Data migration:** moving data from one environment to another, such as movingdata from in-house data centres to the cloud
- **Datapreparation:**readyingdatatobeusinginanalyticsorother applications
- **Dataenrichment:**improvingthequalityofdatabyaddingnewdatasets, correcting small errors or extrapolating new information from raw data
- **Dataanalytics:**analysingdatawithavarietyofalgorithmsinordertogaininsights
- **Dataquality:**makingsuredataisaccurateand reliable
- **Master data management (MDM) :**linking critical enterprise data to one master set that serves as the single source of truth for the organization

- **Datagovernance:**ensuringtheavailability,usability,integrity andaccuracy of data
- **Extracttransform load(ETL):** movingdata from an existingrepository intoa database or data warehouse.

❖ **Organization/Sourcesof Data**

Data organization is the practice of categorizing and classifying data to make it more usable. Similar to a file folder, where wekeep important documents, you'llneed to arrange yourdata in the most logical and orderly fashion, so you — and anyone else who accesses it — can easily find what they're looking for.

DATAISBEINGCOLLECTED

- Thebigdataincludesinformationproducedbyhumansanddevices.
- Device-drivendataislargelycleanandorganized,
- Butof fargreater interestis human-driven data thatexistin variousformats and need more exquisite tools for proper processing and management.

Thebigdatacollectionisfocusedonthefollowingtypesofdata:

- **Networkdata.** Thistype ofdata isgathered onall kindsofnetworks, including social media, informationandtechnologicalnetworks,theInternet andmobile networks, etc.
- **Real-time data.** They are produced on online streaming media, such as YouTube, Twitch, Skype, or Netflix.
- **Transactional data.** They are gathered when a user makes an online purchase (information on the product, time of purchase, payment methods, etc.)
- **Geographic data.** Location data of everything, humans, vehicles, building, natural reserves, and other objects are continuously supplied with satellites.
- **Natural language data.** These data are gathered mostly from voice searches that can be made on different devices accessing the Internet.
- **Time series data.**This type of data is related to the observation of trends and phenomena taking place at this very moment and over a period of time, for instance, global temperatures, mortality rates, pollution levels, etc.
- **Linked data.**They are based on HTTP, RDF, SPARQL, and URIs web technologies and meant to enable semantic connections between various databases so that computers could read and perform semantic queries correctly.

HOW ISBIG DATA COLLECTED?

Therearedifferentwaysofhowtocollectbigdatafrom users.Thesearethemost popular ones.

- **1.Askingforit**
  the majorityof firms prefer asking users directly to share their personal information. Theygivethesedatawhencreatingwebsiteaccountsorbuyingonline.The minimum information to be collected includes a username and an email address, but some profiles require more details.

- **2.CookiesandWebBeacons**
  Cookies and web beacons are two widely used methods to gather the data on users, namely,what webpagestheyvisit andwhen.Theyprovide basicstatisticsabouthow a website is used. Cookies and webbeacons in no waycompromise your privacy but just serve to personalize your experience with one or another web source.
- **3. Emailtracking**
  Emailtrackersaremeant togive more informationontheuseractionsinthe mailbox. In particular, an email tracker allows detecting when an email was opened. Both Google and Yahoo use this method to learn their users' behavioural patterns and provide personalized advertising.

❖ ImportanceofDataQuality

Dataqualityisdefined as:
"The degree to which data meets a company's expectations of accuracy, validity, completeness, and consistency"

Bytrackingdata quality, a business can pinpoint potential issues harming quality, and ensure that shared data is fit to be used for a given purpose.
When collected data fails to meet the company expectations of accuracy, validity, completeness, and consistency, it can have massive negative impacts on customer service, employee productivity, and key strategies.

Qualitydata iskeyto makingaccurate, informeddecisions. Andwhilealldatahassome level of "quality," a variety of characteristics and factors determines the degree of data quality (high-quality versus low-quality). Furthermore, different data quality characteristics will likely be more important to various stakeholders across the organization.
Alistofpopular dataqualitycharacteristicsanddimensionsinclude:

1. **Completeness:**Completenessisdefinedasameasureofthepercentageofdatathatis missing within a dataset.
2. **Timeliness:**Timelinessmeasureshowup-to-dateorantiquatedthedataisatany given moment.
3. **Validity:**Validityreferstoinformationthatfailstofollowspecificcompanyformats, rules, or processes.
4. **Integrity:**Integrityofdatareferstothelevelatwhichtheinformationisreliableand trustworthy.
5. **Uniqueness:**Uniquenessisadataqualitycharacteristicmostoftenassociatedwith customer profiles.
6. **Consistency**:Itensuresthatthesourceoftheinformationcollectioniscapturingthe correct data based on the unique objectives of the department or company.

❖ DealingwithMissing orincompleteData

The concept ofmissing data is implied inthe name: its datathat is not captured for avariable for the observation in question. Missing data reduces the statistical power of the analysis, which can distort the validity of the results.
Fortunately,thereareproventechniquestodealwithmissing data.

Imputationvs. RemovingData

Whendealing with missing data, data scientistscanuse two primarymethods to solve the error: imputation or the removal of data.
The imputation method develops reasonable guesses for missing data. It's most useful when the percentage of missing data is low. If the portion of missing data is too high, the results lack natural variation that could result in an effective model.
The otheroption is to remove data. Whendealing withdatathat is missing at random, related data can be deleted to reduce bias. Removing data may not be the best option if there are not enoughobservationstoresult inareliableanalysis. Insome situations, observationofspecific events or factors may be required.

Before deciding which approach to employ, data scientists must understand why the data is missing.
MissingatRandom(MAR)
Missing at Randommeansthe data is missing relative tothe observed data.It isnot relatedto thespecific missing values. The data is not missing across all observations but only withinsub-samplesofthedata.Itisnotknownifthedatashouldbethere;instead,itis missinggiven the observed data. The missing data can be predicted based on the complete observed data.

MissingCompletelyat Random(MCAR)
In the MCAR situation, the data is missing across all observations regardless of the expected value or other variables. Data scientists can compare two sets of data, one with missing observations and one without. Using a t-test, if there is no difference between the two data sets, the data is characterized as MCAR.

Data may be missing due to test design, failure in the observations or failure in recording observations. This type of data is seen as MCAR because the reasons for its absence are external and not related to the value of the observation.
It is typically safe to remove MCAR data because the results will be unbiased. The test may not be as powerful, but the results will be reliable.
Missing NotatRandom (MNAR)
The MNAR categoryapplies whenthe missing data has a structureto it. Inother words,there appearto be reasonsthe datais missing. Ina survey, perhaps a specific groupofpeople – say women ages 45 to 55 – did not answer a question. Like MAR, the data cannot be determined bytheobserveddata,becausethemissinginformationisunknown.Datascientists must modelthe missing datato develop an unbiased estimate. Simplyremoving observations with missing data could result in a model with bias.

Deletion
There are two primary methods for deleting data when dealing with missing data: list wise and dropping variables.

Listwise
In this method, all data for an observation that has one or more missing values are deleted. The analysis is run only on observations that have a complete set of data. If the data set is small, it may be the most efficient method to eliminate those cases from the analysis. However, in most cases, the data are not missing completely at random (MCAR). Deletingthe instances with missing observations can result in biased parameters and estimates and reduce the statistical power of the analysis.
Pair wise
Pair wise deletion assumes data are missing completelyat random(MCAR), but allthe cases with data, even those with missing data, are used in the analysis. Pairwise deletion allowsdata scientists to use more of the data. However, the resulting statistics may vary becausethey are based on different data sets. The results may be impossible to duplicate with acomplete set of data.

DroppingVariables
If data is missing for more than 60% of the observations, it may be wise to discard it if the variable is insignificant.

- **Imputation**

When data is missing, it may make sense to delete data, as mentioned above. However, that may not be the most effective option. For example, if too much information is discarded, it may not be possible to complete a reliable analysis. Or there may be insufficient data to generate a reliable prediction for observations that have missing data.

Instead of deletion, data scientists have multiple solutions to impute the value of missingdata. Depending why the data are missing, imputation methods can deliver reasonablyreliable results. These are examples ofsingle imputation methods for replacing missing data.

Mean,MedianandMode
This is one ofthe most common methodsofimputing values whendealing with missing data. Incases where there are a small number of missing observations, data scientists can calculatethe mean or median of the existing observations. However, when there are many missing variables, mean or median results can result in a loss of variation in the data. This method does not use time-series characteristics or depend on the relationship between the variables.

Time-SeriesSpecificMethods
Another option is to usetime-series specific methods whenappropriate to impute data. There are four types of time-series data:

- Notrend orseasonality.
- Trend,butno seasonality.
- Seasonality,butnotrend.
- Bothtrendand seasonality.

The time series methods of imputation assume the adjacent observations will be like the missingdata.Thesemethodsworkwellwhenthatassumptionisvalid.However,these

methods won't always produce reasonable results, particularly in the case of strong seasonality.

Last Observation Carried Forward (LOCF) & Next Observation Carried Backward (NOCB)

These options are used to analyze longitudinal repeated measures data,in which follow-up observations may be missing. In this method, every missing value is replaced with the last observed value. Longitudinaldatatrackthe same instance at different points along atimeline. This method is easyto understand and implement. However, this method may introduce bias when data has a visible trend. It assumes the value is unchanged bythe missing data.

LinearInterpolation

Linear interpolation is often used to approximate a value of some function by using two known values of that function at other points. This formula can also be understood as a weighted average. The weights are inverselyrelatedtothe distance fromthe end pointsto the unknown point. The closer point has more influence than the farther point.

When dealing with missing data, you should use this method in a time series that exhibits a trend line, but it's not appropriate for seasonal data.

SeasonalAdjustment withLinearInterpolation

When dealing with data that exhibits both trend and seasonality characteristics, use seasonal adjustment with linear interpolation. First you would perform the seasonal adjustment by computing a centered moving average or taking the average of multiple averages – say, two one-year averages – that are offset by one period relative to another. You can then complete data smoothing with linear interpolation as discussed above.

MultipleImputations

Multiple imputations is considered a good approach for data sets with a large amount of missing data. Instead of substituting a single value for each missing data point, the missing values are exchanged for values that encompass the natural variability and uncertainty of theright values. Using the imputed data, the process is repeated to make multiple imputed data sets. Each set is then analyzed using the standard analytical procedures, and the multiple analysis results are combined to produce an overall result.

The various imputations incorporate naturalvariabilityinto the missing values, whichcreates a valid statistical inference. Multiple imputations can produce statistically valid results even when there is a small sample size or a large amount of missing data.

KNearestNeighbours

Inthis method, data scientists choose a distance measure for k neighbours, and the average is used to impute an estimate. The data scientist must select the number of nearest neighbours and the distance metric. KNN can identifythe most frequent value among the neighbours and the mean among the nearest neighbours.

- **DataVisualization**
  - Data visualization is the practice oftranslating information into a visualcontext, such as a map or graph, to make data easier for the human brain to understand and pull insights from.

- The main goalofdata visualization is to make it easier to identifypatterns, trends and outliers in large data sets. The term is often used interchangeably with others, including information graphics, information visualization and statistical graphics.
- Datavisualizationisone of thesteps of the data science process,which states that after data has been collected, processed and modelled, it must be visualized for conclusions to be made.
- Data visualization is also an element of the broader data presentation architecture (DPA) discipline, which aims to identify, locate, manipulate, format and deliver data in the most efficient way possible.
- Data visualization is important for almost every career. It can be used by teachers to displaystudenttestresults,bycomputerscientistsexploringadvancementsinartificial intelligence (AI) or by executives looking to share information with stakeholders.
- It also plays an important role inbig dataprojects. As businesses accumulated massive collections of data during the early years of the big data trend, they needed a way to quickly and easily get an overview of their data. Visualization tools were a natural fit.
- Visualizationiscentralto advancedanalytics forsimilarreasons.Whenadata scientist is writing advanced predictive analytics or machine learning (ML) algorithms, it becomesimportanttovisualizetheoutputstomonitorresultsandensure that models are performing as intended. This is because visualizations of complex algorithms are generally easier to interpret than numerical outputs.

Whyisdatavisualizationimportant?

Data visualization provides a quick and effective way to communicate information in a universal manner using visual information. The practice can also help businesses identify which factors affect customer behaviour; pinpoint areas that need to be improved or need more attention; make data more memorable for stakeholders; understand when and where to place specific products; and predict sales volumes.

Otherbenefitsofdatavisualizationincludethefollowing:

- theabilitytoabsorbinformationquickly,improveinsightsandmakefasterdecisions;
- anincreasedunderstandingof thenextstepsthatmustbetakentoimprovethe organization;
- an improved ability to maintain the audience's interest with information they canunderstand;
- aneasydistributionof informationthat increases the opportunityto share insights with everyone involved;
- eliminatetheneedfordatascientistssincedataismoreaccessibleandunderstandable; and
- Anincreasedabilitytoactonfindingsquicklyand,therefore,achievesuccess with greater speed and less mistakes.

Datavisualization andbigdata

- The increased popularityofbig data and data analysis projects has made visualization more important than ever.
- Companiesare increasinglyusing machine learningtogather massiveamountsofdata that can be difficult and slow to sort through, comprehend and explain.
- Visualizationoffersameanstospeedthisupandpresentinformationtobusiness owners and stakeholders in ways they can understand.

- Big data visualization often goes beyond the typical techniques used in normal visualization,such as pie charts, histogramsandcorporate graphs.Itinstead uses more complex representations, such as heat maps and fever charts.
- Bigdatavisualizationrequirespowerfulcomputersystemstocollect rawdata,process it and turn it into graphical representations that humans can use to quickly draw insights.

❖ **Examplesofdata visualization**

Intheearlydaysofvisualization,themostcommonvisualizationtechniquewasusing aMicrosoftExcelspreadsheettotransform theinformationintoatable,bargraphor piechart. While these visualization methods are still commonly used, more intricate techniques are now available, including the following:

- infographics
- bubbleclouds
- bulletgraphs
- heatmaps
- fevercharts
- timeseriescharts

Someotherpopulartechniquesareas follows.

Line charts. This is one of the most basic and common techniques used. Line charts display how variables can change over time.

Area charts. This visualization method is a variation of a line chart; it displays multiple values ina time series -- or asequence ofdata collected at consecutive, equallyspaced points in time.

Scatterplots. Thistechniquedisplaystherelationshipbetweentwovariables.A scatterplot takes the form of an x- and y-axis with dots to represent data points.

Tree maps. This method shows hierarchical data in a nested format. The size of the rectangles used for eachcategoryis proportionalto its percentage ofthe whole. Treemapsare best used when multiple categories are present, andthe goalis to compare different partsofa whole.

Population pyramids. This technique uses a stacked bar graphto displaythe complexsocial narrative of a population. It is best used when trying to display the distribution of a population.

DataVisualizationApplications

Commonusecasesfordatavisualizationincludethefollowing:

SalesandMarketing: Researchfromthe mediaagencyMagnapredictsthat halfofallglobal advertising dollars will be spent online by2020. As a result, marketing teams must payclose attention to their sources of web traffic and how their web properties generate revenue. Data visualization makes it easyto see traffic trends over time as a result of marketing efforts.

Politics: Acommonuse ofdata visualization inpolitics isa geographic map that displaysthe party each state or district voted for.

Healthcare: Healthcare professionals frequently use choropleth maps to visualize important healthdata.Achoroplethmapdisplaysdividedgeographicalareasorregionsthatare

assigned a certain color in relation to a numeric variable. Choropleth maps allow professionals to seehowavariable, suchasthe mortalityrateofheart disease, changesacross specific territories.

Scientists: Scientific visualization, sometimes referred to in shorthand as SciVis, allows scientists and researchers to gain greater insight from their experimental data than everbefore.

Finance: Finance professionals must track the performance of their investment decisions when choosing to buy or sell an asset. Candlestick charts are used as trading tools and help finance professionals analyze price movements over time, displaying important information, such as securities, derivatives, currencies, stocks, bonds and commodities. Byanalyzing how the price has changed over time, data analysts and finance professionals can detect trends.

Logistics: Shipping companies can use visualization tools to determine the best global shipping routes.

❖ Datavisualizationtoolsand vendors

Data visualization tools can be used in a variety of ways. The most common use today is as business intelligence (BI) reporting tool. Users can set up visualization tools to generate automatic dashboards that track company performance across key performance indicators (KPIs) and visually interpret the results.

The generated images may also include interactive capabilities, enabling users to manipulate them or look more closely into the data for questioning and analysis. Indicators designed to alert users when data has been updated or when predefined conditions occur can also be integrated.

Many business departments implement data visualization software to track their own initiatives. For example, a marketing team might implement the software to monitor the performance of an email campaign, tracking metrics like open rate, click-through rate and conversion rate.

As data visualization vendors extend the functionality of these tools, they are increasingly being used as front ends for more sophisticated big data environments. In this setting, data visualization software helps data engineers and scientists keep track of data sources and do basic exploratory analysis of data sets prior to or after more detailed advanced analyses.

The biggest names in the big data tools marketplace include Microsoft, IBM, SAP and SAS.

Some other vendors offer specialized big data visualization software; popular names in thismarket include **Tableau, Qlik and Tibco.**

While **Microsoft Excel continues to be a popular tool for data visualization,** others have been created that provide more sophisticated abilities:

- IBM Cognos Analytics
- QlikSenseand QlikView
- MicrosoftPowerBI
- OracleVisual Analyzer
- SAPLumira
- SASVisualAnalytics
- TibcoSpotfire
- ZohoAnalytics
- D3.js
- Jupyter

- MicroStrategy
- GoogleCharts

❖ **DataClassification**

- ✓ Data classification is broadly defined as the process of organizing data by relevant categoriessothatitmaybeusedandprotectedmoreefficiently.Onabasiclevel, the classification process makes data easier to locate and retrieve.
- ✓ Data classificationis of particular importance when it comes to risk management, compliance, and data security.
- ✓ Dataclassificationinvolvestaggingdatatomakeiteasilysearchableandtraceable.
- ✓ It also eliminates multiple duplications ofdata, whichcanreduce storage and backup costs while speeding up the search process. Though the classification process may sound highly technical, it is a topic that should be understood by your organization'sleadership.

ImportanceofDataClassification:
Dataclassificationisaregulatoryrequirement,asdatamustbesearchableandretrievable within specified timeframes.
For thepurposes of datasecurity,dataclassification is auseful tacticthatfacilitates proper security responses based on the type of data being retrieved, transmitted, or copied.

❖ **TypesofDataClassification**
Data classification involves the use of tags and labels to define the data type, its confidentiality, and its integrity. There are three main types of data classification that are considered the industry standard:

- **Content-based classification** – inspects and interprets files, looking for sensitive information
- **Context-based classification** – lookstothe application, location, metadata,or creator (among other variables) as indirect indicators of sensitive information
- **User-based classification** – requires a manual, end-user selection for each document. User-based classification takes advantage ofthe user knowledge ofthe sensitivity ofthe document, and can be applied or updated upon creation, edit, review, or dissemination.

DETERMININGDATARISK
Inadditionto thetypesofclassification, it'swise foranorganizationto determinetherelative risk associated with the types of data, how that data is handled and where it is stored/sent (endpoints). A common practice is to separate data and systems into three levels of risk

Low risk: If data is public and it's not easy to permanently lose (e.g. recovery is easy), this data collection and the systems surrounding it are likely a lower risk than others.
Moderate risk: Essentially, this is data that isn't public or is used internally (by your organization and/or partners). However, it's also not likely too critical to operations or sensitiveto be"highrisk."Proprietaryoperating procedurescostofgoodsandsomecompany documentation may fall into the moderate category.

High risk: Anything remotely sensitive or crucial to operational security goes into the high risk category. Also, pieces of data those are extremely hard to recover (if lost). All confidential, sensitive and necessary data falls into a high risk category.

- **DataSensitivityLevels**

While we've looked at mapping data out by type, you should also look to segment your organization's data in terms of the level of sensitivity – high, moderate, or low.

- **High sensitivity data (Confidential)** – data that if compromised or destroyed would be expected to have a severe or catastrophic effect on organizational operations, assets, or individuals. Examples can include financial data, medical records, and intellectual property.
- **Moderate sensitivity data (Restricted)** – data that if compromised or destroyed would be expected to have a serious effect on organizational operations, assets, or individuals. Examples can include unpublished research results, information strictly for internal use, and operational documents.
- **Low sensitivity data (Public)** – data that if compromised or destroyed would be expected to have a limited effect on organizational operations, assets, or individuals. Examples can include press releases, job advertisements, and published research.

The following shows common examples of organizational data which may be classified into each sensitivity level:

High:

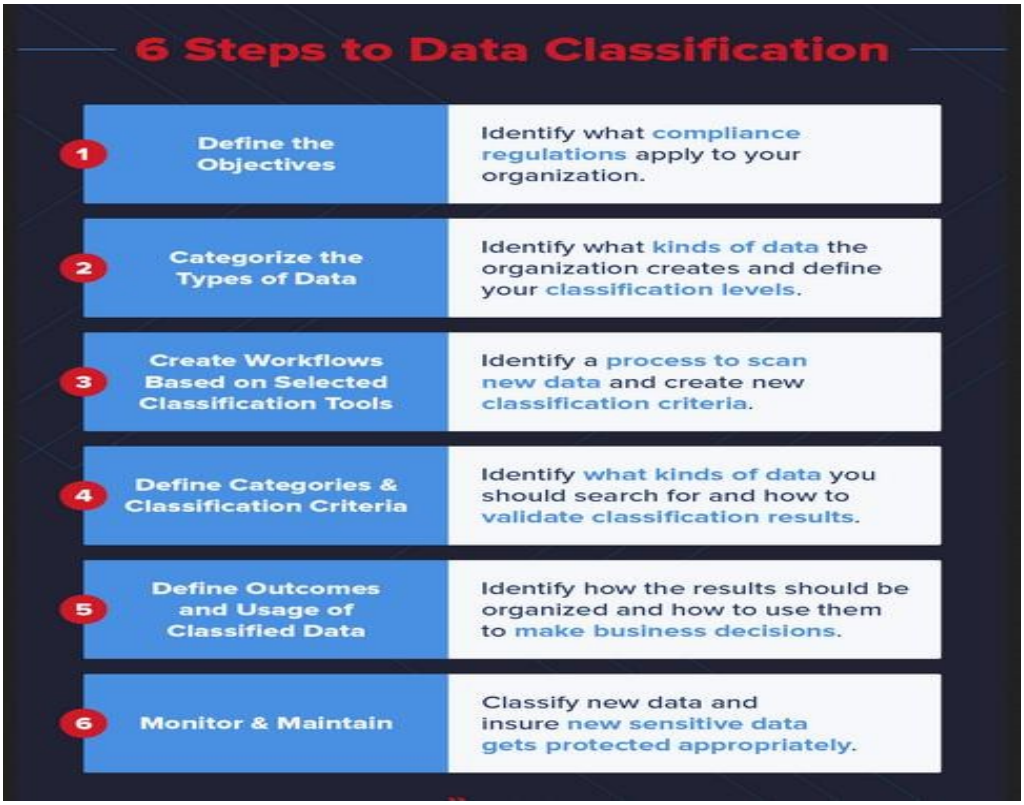
- Personallyidentifiableinformation(PII)
- Creditcarddetails(PCI)
- Intellectualproperty(IP)
- Protectedhealthcareinformation(includingHIPAAregulateddata)
- Financialinformation
- Employeerecords
- ITARmaterials
- Internalcorrespondence includingconfidentialdata

Moderate:

- Studenteducationrecords
- Unpublishedresearchdata
- Operationaldata
- Informationsecurityinformation
- Suppliercontactinformation
- Internalcorrespondencenotcontainingconfidentialdata

Low:

- Publicwebsites
- Publicdirectorydata
- Publiclyavailable research
- Pressreleases
- Job advertisements
- Marketingmaterials



- **DataScienceProject LifeCycle:**

Data Science is a multidisciplinary field that uses scientific methods to extract insights from structured and unstructured data. Data science is such a huge field and concept that's often intermingled with other disciplines, but generally, DS unifies statistics, data analysis,machine learning, and related fields.

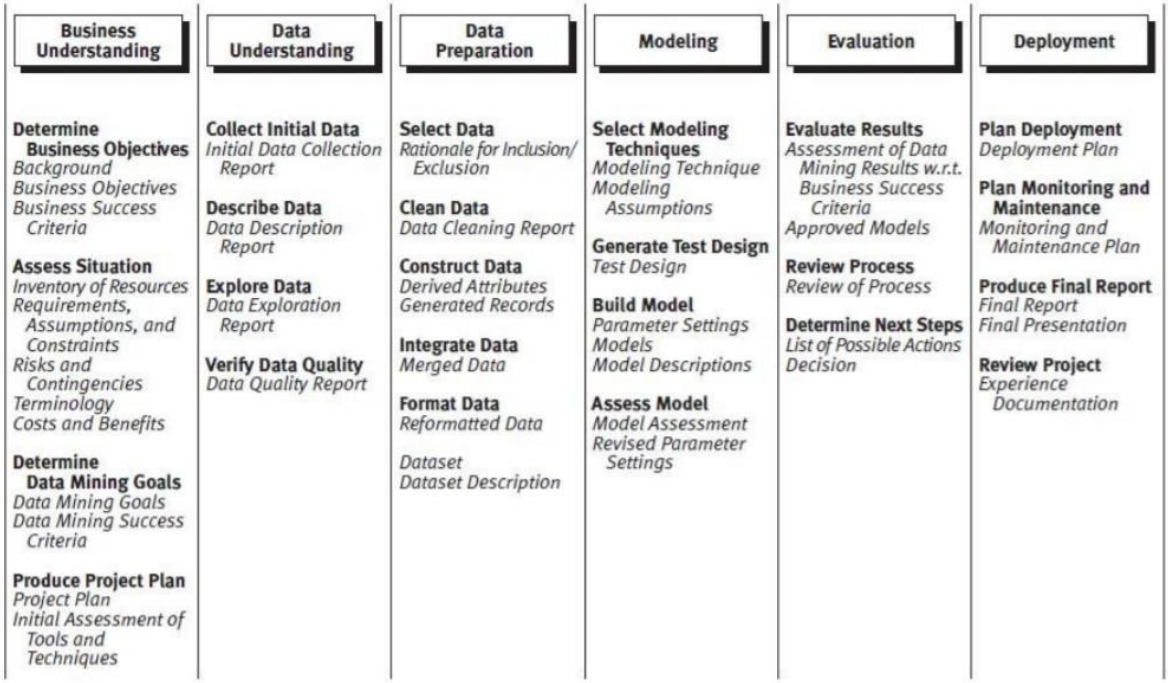
Data Science life cycle provides the structure to the development of a data science project. The lifecycle outlines the major steps, fromstart to finish, that projects usually follow. Now, there are various approaches to managing DS projects, amongst which are **Cross-industry standard process for data mining (aka CRISP-DM),** process of knowledge discovery in databases (aka KDD), any proprietary-based custom procedures conjured up by a company, and a few other simplified processes.

CRISP-DM

CRISP-DM is an open standard process model that describes common approaches used by data mining scientists. In 2015, it was refined and extended by IBM, which released a new methodology called Analytics Solutions Unified Method for Data Mining/Predictive Analytics (aka ASUM-DM).

TheCRISP-DMmodelstepsare:

1. BusinessUnderstanding
2. DataUnderstanding
3. Data Preparation
4. Modelling
5. Evaluationand
6. Deployment



Knowledgediscoveryindatabases**(KDD)**

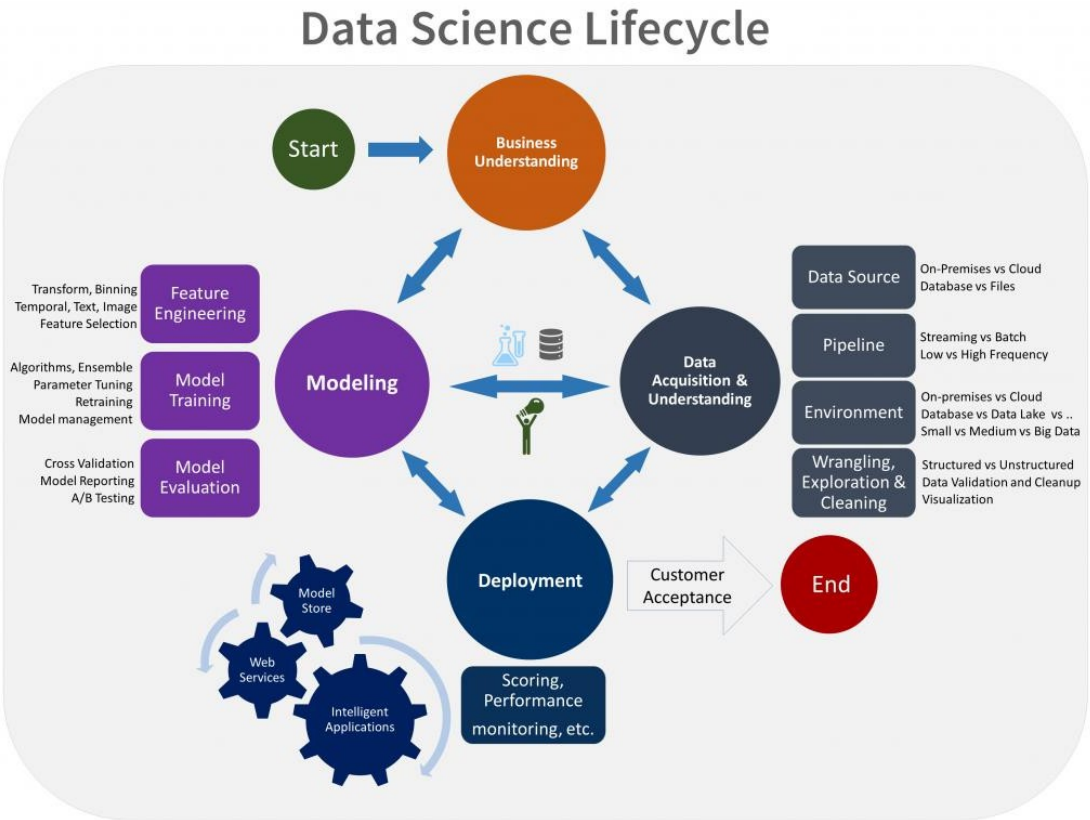
KDDiscommonlydefinedwiththefollowing stages:

- ✓ Selection
- ✓ Pre-processing
- ✓ Transformation
- ✓ Datamining
- ✓ Interpretation/evaluation

The**simplifiedprocess** looksasfollows:(1)Pre-processing,(2)DataMining,and(3) Results Validation.

Suppose, wehaveastandardDSproject (withoutanyindustry-specificpeculiarities), thenthe lifecycle would typically include:

- ✓ **Businessunderstanding**
- ✓ **Dataacquisitionand understanding**
- ✓ **Modelling**
- ✓ **Deployment**
- ✓ **Customeracceptance**



The DS project life cycle is an iterative process of research and discovery that provides guidance on the tasks needed to use predictive models. The goal of this process is to move a DS project to an engagement end-point by providing means for easier and clearer communication between teams and customers with a well-defined set of artifacts and standardized templates to homogenize procedures and avoid misunderstandings.

Eachstagehasthefollowinginformation:

- Goalsand specificobjectivesofthe stage
- Aclearoutlineofspecifictasksandinstructionsonhowto completethem
- Theexpecteddeliverables(artifact)

Businessunderstanding

Before you even embark on a DS project, you need to understand the problem you're tryingto solve and define the central objectives of your project by identifying the variables to predict.

Goals:

- ✓ Identifykeyvariablesthatwillserveasmodeltargetsandserveasthemetricsfor defining the success of the project
- ✓ Identifydatasourcesthatthebusinesshasalreadyaccesstoorneedtoobtainsuch access

Guidelines:

Work with customers and stakeholders to define business problems and formulate questions that data science needs to answer.

The goal here is to identify the key business variables (aka model targets) that your analysis needstopredictandtheproject'ssuccesswouldbeassessedagainst.Forexample,thesales

forecasts. This is what needs to be predicted, and at the end of your project, you'll compare your predictions to the actual volume of sales.

Defineprojectgoalsbyaskingspecificquestionsrelatedtodatascience, suchas:

- Howmuch/many?(regression)
- Whichcategory?(classification)
- Whichgroup?(clustering)
- Doesthismakesense?(anomalydetection)
- Whichoptionshouldbetaken?(recommendation)

❖ **BusinessRequirements**

- ✓ The purposeofbusinessrequirementsisto define a project'sbusinessneed, aswellas the criteria of its success.
- ✓ Business requirements describe why a project is needed, whom it will benefit, when and where it will take place, and what standards will be used to evaluate it.
- ✓ Businessrequirement_generally do not define*how*a project is to be implemented; the requirements of the business need do not encompass a project's implementation details.
- ✓ "Business requirements are higher-level statements of the goals, objectives, or needs of the enterprise."
- ✓ "They describe the reasons why a project has been initiated, the objectives that the project will achieve, and the metrics that will be used to measure its success."
- ✓ In short, business requirements chart where a project is going, not how it's going to get there.

The business requirements the analyst creates forthis project would include (but not be limited to):

- **Identification of the business problem** (key objectives of the project), i.e., "Declining ticket sales require a strategy to increase the number of customers at our theatres."
- **Why the solution has been proposed** (its benefits; why it will produce the desired outcome of returning ticket sales to higher levels), i.e., "Customers have overwhelminglycited the inconvenienceofstanding in lineastheprimaryreasonthey no longer attend our theatre. We will remove this impediment by enabling customers to buyand print their theatre tickets at home with just a few clicks."
- **The scope of the project**. A few examples might be: "1. while the plan is to bringthis project to all 400 theatres eventually, we will start with 50 theatres in the most populated metropolitan areas."
- **Rules, policies, and regulations**. For example, "We will design our web site and commerce so that all other relevant governmental regulations are properly adheredto."
- **Key features of the service** (without details as to how they will be implemented). A fewexamplesmight include: "1. wewillprovidea securesite fortheuserto select the number of tickets and showing they wish, and to enter their payment information. 2. We will give the user the option to store his or her card information in our system so that theydo not have to re-enter it in a later session. 3. The systemwillaccommodate credit, debit, or PayPal payment methods only."

- **Keyperformance features**(without detailsasto howtheywillbe implemented), i.e., "1. The system will be designed so that it can recover within 30 seconds of any downtime. 2. Because our peak audience has been 25,000 customers in all of our theatres on one night, the system will accommodate at least10 times that many users at any given time without any impact on system performance."
- **Keysecurityfeatures**(againwithout details), i.e.,"Wewilldeviseaunique identifier for each ticket that will prohibit photocopies or counterfeits."
- Criteria to measure the project's success, such as: "This project will be deemed successful if ticket sales returnto 2008 levels within 12 months of its launch."

Thisproject'sresultingbusinessrequirementswouldnotinclude:

- Adescriptionofhowtoadheretogovernmentalorregulatoryrequirements.
- A description of how performance requirements will be implemented, such as: "The XYZ server on which customer information is stored will be backed up every five minutes using XYZ program."
- Anydescriptionofhowtheuniqueticketidentifierwouldbeimplemented.
- Any details or specifics related to the service's features, such as: "1. The credit card number text box will be 20 characters long and accommodate simple text. 2. If the user selects Yes (01), the information will be loaded to our XYZ storage server called."

While the above examples accompanying selected bullet points are textual, business requirements may include graphs, models, or any combination of these that best serves the project. Effective business requirements require strong strategic thinking, significant input from a project's business owners, and the ability to clearly state the needs of a project at a high level.

Aswithall<u>requirements</u>,businessrequirementsshouldbe:

- **Verifiable.** Just because business requirements state business needs rather than technical specifications doesn't mean they mustn't be demonstrable.
- Verifiable requirements are specific and objective. A quality control expert must be able to check, for example, that the system accommodates the debit, credit, andPayPal methods specified in the business requirements. (S)he could not do so if the requirements were more vague, i.e., "The system will accommodate appropriate payment methods." (*Appropriate* is subject to interpretation.)
- **Unambiguous**, stating precisely what problem is being solved. For example, "This project will be deemed successful ifticket sales increase sufficiently," is probablytoo vague for all stakeholders to agree on its meaning at the project's end.
- **Comprehensive,** covering every aspect of the business need. Business requirements are indeed big picture, but they are very thorough big picture. In the aforementioned example, if the analyst assumed that the developers would know to design a system that could accommodate many times the number of customers the theatre chain had seen at one time in the past, but did not explicitly state so in the requirements, the developers might designa systemthat could accommodateonly10,000 patronsat any one time without performance issues.

Rememberthatbusinessrequirementsanswerthewhat's,notthehow's,buttheyare meticulouslythoroughindescribingthose's.Nobusinesspointisoverlooked.Ataproject's

end, the business requirements should serve as a methodical record of the initial business problem and the scope of its solution.

Understanding the project objectives and requirements from a domain perspective and then converting this knowledge into a data science problem definition with a preliminary plan designed to achieve the objectives. Data science projects are often structured around the specific needs of an industry sector (as shown below) or even tailored and built for a single organization. A successful data science project starts from a well defined question or need.

❖ **DataAcquisition**

- ✓ **Data acquisition**(commonly abbreviated as **DAQor DAS**)is the process of sampling signals that measure real-world physical phenomena and converting them into a digital form that can be manipulated by a computer and software.

- ✓ Data Acquisitionis generally accepted to be distinctfrom earlierforms of recording to tape recorders or paper charts. Unlike those methods, the signals are convertedfrom the analog domain to the digital domain and then recorded to a digital medium such as ROM, flash media, or hard disk drives.

ThePurposesofData Acquisition

The primarypurpose ofa data acquisitionsystemis to acquire and storethe data.Buttheyare also intended to provide real-time and post-recording visualization and analysis of the data. Furthermore, most data acquisition systems have some analytical and report generation capability built-in.

Engineers in different applications have various requirements, of course, but these key capabilities are present in varying proportion:

- Datarecording
- Datastoring
- Real-timedatavisualization
- Post-recordingdata review
- Dataanalysisusingvariousmathematicalandstatisticalcalculations
- Reportgeneration

❖ **DataPreparation**

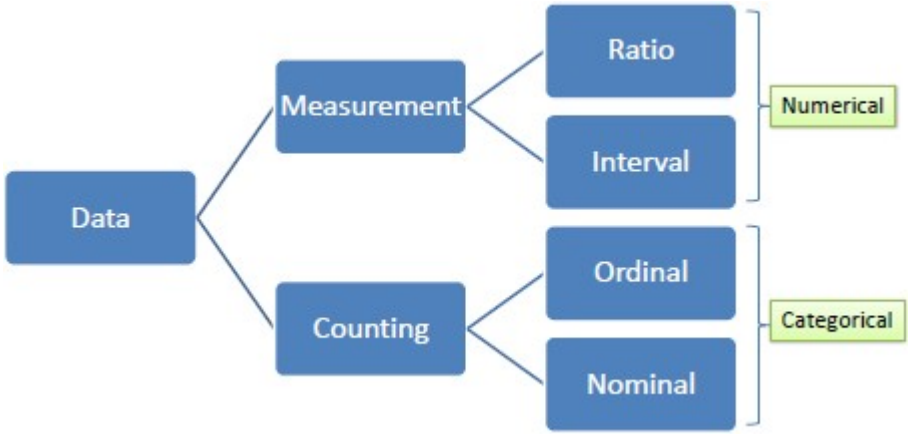
Data preparation is about constructing a dataset from one or more data sources to be used for explorationand modeling. It isasolidpracticeto startwithaninitialdataset to getfamiliar withthe data, to discover first insights into the data and have a good understanding of any possible data quality issues. Data preparation is often a time consuming process and heavily prone to errors. The old saying "garbage-in-garbage-out" is particularly applicable to those data science projects where datagatheredwithmanyinvalid, out-of-rangeand missing values. Analyzingdatathat hasnot been carefully screened for such problems can produce highly misleading results. Then, the success of data science projects heavily depends on the quality of the prepared data.

Data

Data isinformationtypicallytheresultsofmeasurement(numerical)orcounting (categorical). **Variables**serveasplaceholdersfordata.Therearetwotypesof variables, *numerical* and *categorical*.

A**numerical**or **continuousvariable** is one that can accept any value within a finite or infinite interval(e.g.,height,weight,temperature,bloodglucose,).Therearetwotypesofnumerical data,*interval* and *ratio*. Data on an interval scale can be added and subtracted but cannot be meaningfully multiplied or divided because there is no true zero. For example, we cannot say that onedayistwiceas hotasanother day. Ontheother hand, dataonaratio scale has truezeroandcan be added, subtracted, multiplied or divided (e.g., weight).

A**categorical**or**discretevariable** is one thatcan accept two or more values (categories).There are two types of categorical data, *nominal* and *ordinal*. Nominal data does not have an intrinsic ordering inthecategories.Forexample,"gender" withtwocategories, maleand female.Incontrast, ordinal data does have an intrinsic ordering in the categories. For example, "level of energy" with three orderly categories (low, medium and high).



Dataset

Datasetisacollectionofdata,usuallypresentedinatabularform.Eachcolumnrepresentsa particular variable, and each row corresponds to a given member of the data.

Columns

ID	Outlook	Temp	Humidity	Windy	Play Golf
1	Rainy	85	92	False	No
2	Rainy	80	88	True	No
3	Overcast	83	86	False	Yes
4	Sunny	70	80	False	Yes
5	Sunny	68	?	False	Yes
6	Sunny	65	58	True	No
7	Overcast	64	62	True	Yes
8	Rainy	72	95	?	No
9	Rainy	?	70	False	Yes
10	Sunny	75	72	False	Yes
11	Rainy	75	74	True	Yes
12	?	72	78	True	Yes
13	Overcast	81	66	False	Yes
14	Sunny	71	79	True	No

Rows

Values

Therearesomealternatives forcolumns,rowsand values.

- Columns,Fields,Attributes,Variables
- Rows,Records,Objects,Cases,Instances,Examples, Vectors

- Values,Data

Inpredictivemodeling, **predictors**or **attributes**aretheinputvariablesand **target** or**class attribute** is the output variable whose value is determined by the values of the predictors and function of the predictive model.

Database

Database collects, stores and manages information so users can retrieve, add, update or removesuch information. It presents information in tables with rows and columns.A table is referred to asa relation in the sense thatitis a collection of objects of the same type (rows). Data in a table canbe related according to common keys or concepts, and the ability to retrieve related data from related tables is the basis for the term relational database. A Database Management System (**DBMS**) handles the way data is stored, maintained, and retrieved. Most data science toolboxes connect to databases through**ODBC**(Open Database Connectivity) or **JDBC**(Java Database Connectivity) interfaces.

SQL(StructuredQueryLanguage)isadatabasecomputerlanguageformanagingand manipulating data in relational database management systems (RDBMS).

SQLDataDefinitionLanguage (**DDL**) permitsdatabase tablesto be created,alteredordeleted. We can also define indexes (keys), specify links between tables, and impose constraints between database tables.

- CREATE TABLE :createsanewtable
- ALTERTABLE:altersatable
- DROPTABLE:deletesatable
- CREATEINDEX:createsan index
- DROP INDEX :deletesanindex

SQLDataManipulationLanguage(**DML**)isalanguagewhichenablesuserstoaccessand manipulate data.

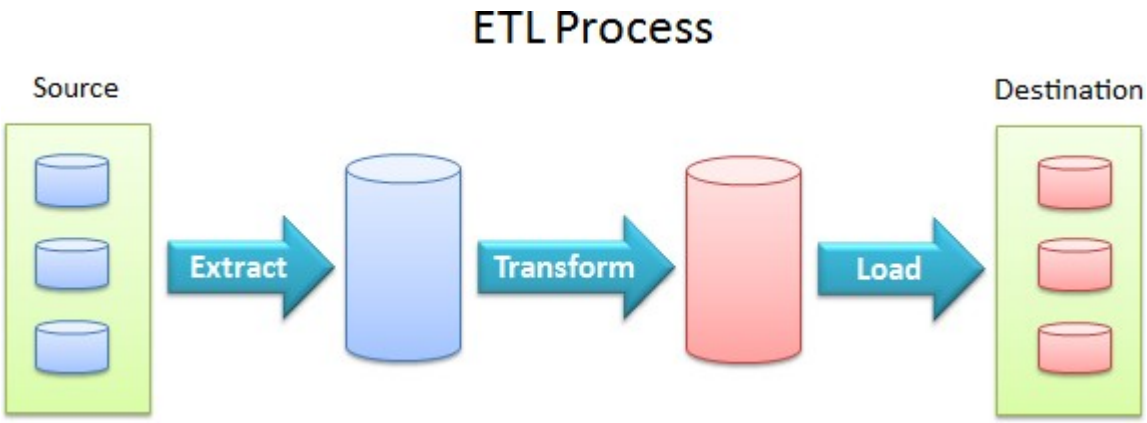
- SELECT:retrievalofdatafromthedatabase
- INSERTINTO:insertion ofnew dataintothedatabase
- UPDATE: modificationofdatainthe database

- DELETE:deletionofdata inthedatabase

ETL(Extraction,TransformationandLoading)

ETL extracts datafromdata sources and loads it into data destinations using a setoftransformation functions.

- **Data extraction** provides the ability to extract data from a variety of data sources, such as flat files, relational databases, streaming data, XML files, and ODBC/JDBC data sources.
- **Data transformation**provides the ability tocleanse,convert,aggregate,merge,and split data.
- **Data loading**provides the ability to load data into destination databases via update, insertor delete statements, or in bulk.

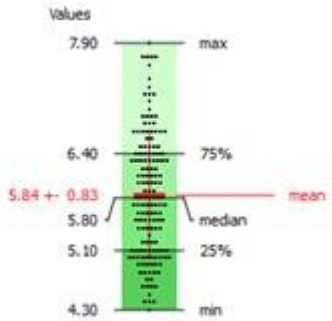


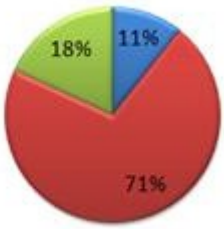
CreditDefaultDatasets

❖ **DataExploration**

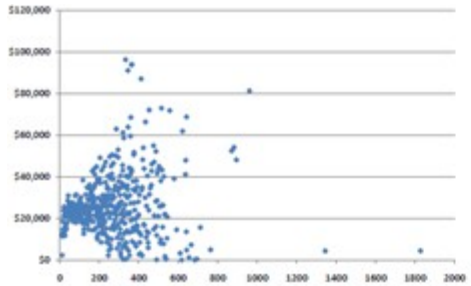
Data Exploration is about describing the data by means of statistical and visualization techniques. We explore data in order to bring important aspects of that data into focus for further analysis.

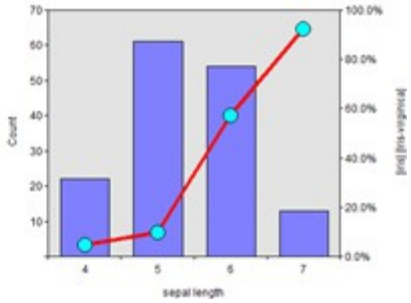
1. UnivariateAnalysis





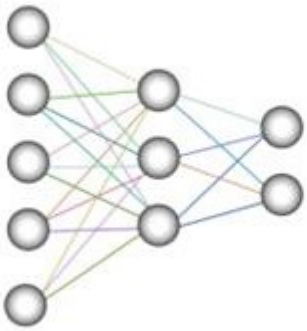
2. BivariateAnalysis

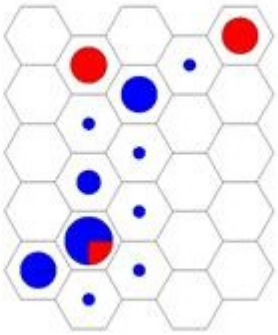




Modeling

Predictive modeling is the process bywhicha modelis createdto predict anoutcome. Ifthe outcome is categorical it is called classificationand if the outcome is numerical it is called regression. Descriptive modelingor clusteringistheassignmentofobservationsinto clusterssothatobservations in the same cluster are similar. Finally, association rulescan find interesting associations amongst observations.





ModelEvaluation

Model Evaluation is an integral part of the model development process.It helps to find the best model that represents our data and how well the chosen model will work in the future. Evaluating model performance with the data used for training is not acceptable in data science because it can easilygenerateoveroptimisticandoverfittedmodels.Therearetwomethodsofevaluatingmodelsin datascience,Hold-OutandCross-Validation.Toavoidoverfitting,bothmethodsuseatestset(not

seenbythemodel)toevaluatemodelperformance.

Hold-Out

Inthismethod,themostlylargedatasetis *randomly*dividedtothreesubsets:

1. **Trainingset**isasubsetofthedatasetusedtobuildpredictivemodels.
2. **Validation set** is a subset ofthe dataset used to assess the performance of model built in the training phase. It provides a test platformfor fine tuning model's parameters and selecting the best-performing model. Not all modeling algorithms need a validation set.
3. **Test set** orunseenexamplesare a subset ofthe dataset to assessthe likelyfuture performance of a model. If a model fit to the training set much better than it fits the test set, overfitting is probably the cause.

Cross-Validation

When only a limited amount of data is available, to achieve an unbiased estimate of the model performanceweuse *k*-foldcross-validation. In*k*-foldcross-validation,wedividethedata into *k*subsets of equal size. We build models *k*times, each time leaving out one of the subsets from training and use it as the test set. If *k* equals the sample size, this is called "leave-one-out".

Modelevaluationcanbe dividedtotwosections:

- ClassificationEvaluation
- RegressionEvaluation

ModelDeployment

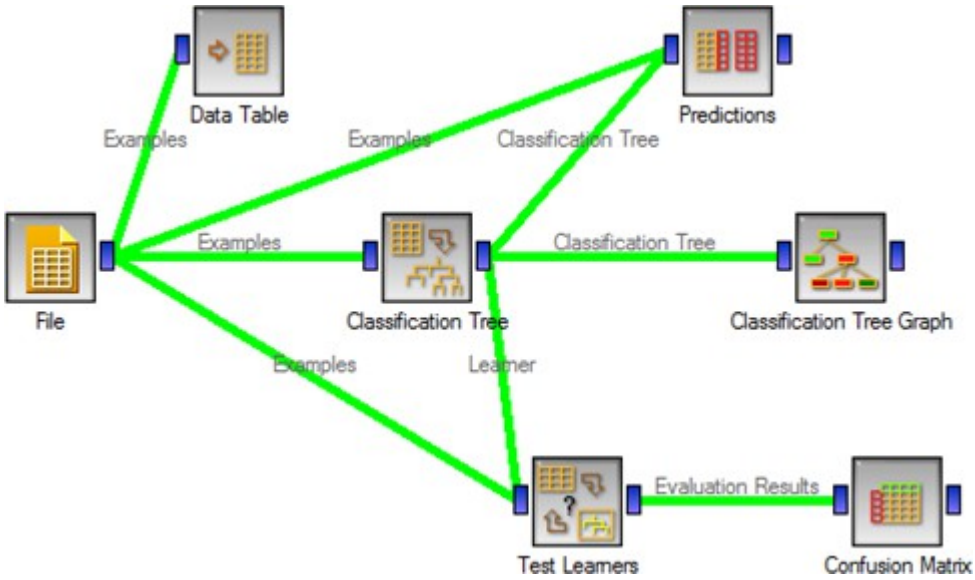
The concept of deployment in data science refers to the application of a model for prediction using a newdata. Building a modelisgenerally not theend oftheproject. Even ifthepurposeofthe modelis to increase knowledge ofthe data,the knowledge gained willneed to beorganized and presented ina way that the customer can use it. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data science process. In manycases, it willbethecustomer, notthedataanalyst, who willcarryoutthedeployment steps. For example, a credit card company may want to deploy a trained model or set of models (e.g., neural networks, meta-learner) to quickly identify transactions, which have a high probability of being fraudulent. However, even if the analyst will not carry out the deployment effort itis important for the customer to understand up front what actionswillneed to be carried out inorderto actually make use of the created models.

Modeldeploymentmethods:

Ingeneral,thereisfour wayofdeployingthemodelsindatascience.

1. Datasciencetools(orcloud)
2. Programminglanguage(Java,C,VB,…)
3. DatabaseandSQLscript(TSQL,PL-SQL,…)
4. PMML(PredictiveModelMarkupLanguage) 5.

Anexampleofusinga dataminingtool(Orange)todeploya decisiontreemodel.



- **OperationsResearch**

Generally, OR is concerned with obtaining extreme values of some real-world objective functions; maximum (profit, performance, utility, or yield), minimum (loss, risk, distance, or cost). It incorporates techniques from mathematical modelling, optimization, and statistical analysis while emphasising the human-technology interface. However, one of the difficulties in answering this question is that there is a lot of overlap in scientific terminology — and sometimes terms become extremely popular, affecting the landscape of the terminology. E.g. the popularity of vague, broad terms such as AI and Big Data that works good for marketing but does nothing for the discussion on theresearch.Therefore,IhavetriedillustratingitintermsofORsrelatedfields,subfields,andthe addressedproblems

Process optimization is an exercise that aims to streamline operations within a projectprocess, maximizing resource use and improving overalloutput. It is a significant element of business decision-making and is used in manydifferent project management areas.

Processoptimizationmethodsandtechniques

Therearemanyprocessoptimizationtechniquesyoucanusetogetyoustarted.Hereare three examples:
Processmining:Thisisagroup of techniqueswith adatascienceapproach.Dataistaken fromeventlogstoanalyzewhatteammembersaredoinginacompanyandwhatstepsthey

take to complete a task. This data can then be turned into insights, helping projectmanagers to spot any roadblocks and optimize their processes.

DMAIC: DMAIC is a data-focused method used in Six Sigma to improve processes. Itstands for Define, Measure, Analyze, Improve, and Control. These five stages combine to form a cycle. First, customers are defined. Then, performance is measured, and the data is analyzed. Finally, improvements are implemented and controlled to ensure the process remains in optimal condition.

PDSA: PDSA is an acronym for Plan, Do, Study, Act. It uses a four-stage cyclical model to improve quality and optimize business processes. Project managers will start by mapping what achievementstheywant to accomplish. Next, theywilltest proposed changesona small scale. After this, they will study the results and determine if these changes were effective. If so, they will implement the changes across the entire business process.

It's good practice for a project manager to take some time to research various process optimization methods before deciding which one is most suited to their business

MajorApplications ofDataScience

Data Science is the deep study of a large quantity of data, which involves extracting some meaningful from the raw, structured, and unstructured data. The extracting out meaningful data from large amounts use processing of data and this processing can be done using statisticaltechniquesandalgorithm,scientifictechniques,differenttechnologies,etc.It usesvarioustoolsandtechniquestoextractmeaningfuldatafromrawdata.DataScience is also known as the **Future of Artificial Intelligence**.

1. InSearchEngines

ThemostusefulapplicationofDataScienceisSearchEngines.Asweknowwhenwe want to search for something on the internet, we mostly used Search engines like Google, Yahoo, Safari,Firefox,etc. So Data Scienceis used to get Searches faster.

ForExample,Whenwesearchsomethingsuppose“DataStructureandalgorithmcourses ” then at that time on the Internet Explorer we get the first link of GeeksforGeeks Courses. ThishappensbecausetheGeeksforGeekswebsiteisvisitedmostinordertoget informationregardingDataStructurecoursesandComputerrelatedsubjects.Sothis analysisis Done usingData Science, and we get theTopmostvisitedWeb Links.

2. InTransport

Data Science also entered into the Transport field like Driverless Cars. With the help of Driverless Cars, it is easy to reduce the number of Accidents.

ForExample, InDriverlessCarsthetrainingdataisfedintothealgorithmandwiththe help of Data Science techniques, the Data is analyzed like what is the speed limitinHighway, Busy Streets, Narrow Roads, etc. And how to handle different situations while driving etc.

3. InFinance

Data Science plays a key role in Financial Industries. Financial Industries always have an issue of fraud and risk of losses. Thus, Financial Industries needs to automate risk of loss analysisinordertocarryoutstrategicdecisionsforthecompany.Also,Financial Industries uses Data Science Analytics tools in order to predict the future. It allows the companies to predictcustomer lifetime value and theirstock marketmoves.

For Example, In Stock Market, Data Science is the main part. In the Stock Market, Data Science is used to examine past behavior with past data and theirgoal is to examine thefuture outcome. Data is analyzed in such a way that it makes it possibleto predict future stock prices over a set timetable.

4. InE-Commerce

E-Commerce Websites like Amazon, Flipkart, etc. uses data Science to make a better user experience with personalized recommendations.

For Example, When we search for something on the E-commerce websites we get suggestions similar to choices according to our past data and also we get recommendations according to most buy the product, most rated, most searched, etc. This is all done with the help of Data Science.

5. InHealthCare

IntheHealthcareIndustrydatascienceactasaboon.DataScienceisusedfor:

- DetectingTumor.
- Drugdiscoveries.
- MedicalImageAnalysis.
- VirtualMedicalBots.
- GeneticsandGenomics.
- PredictiveModelingforDiagnosisetc.

6. ImageRecognition

currently, Data Science is also used in Image Recognition. **For Example,** When we upload our image with our friend on Facebook, Facebook gives suggestions Tagging who is in the picture.Thisisdonewiththehelpofmachinelearningand DataScience.WhenanImage isRecognized,thedataanalysisisdoneonone'sFacebookfriendsandafteranalysis,if thefaceswhicharepresentinthepicturematchedwithsomeoneelseprofilethen Facebook suggests us auto-tagging.

7. TargetingRecommendation

TargetingRecommendationisthemostimportantapplicationofDataScience.Whatever the user searches on the Internet, he/she will see numerous posts everywhere. This can be explainedproperlywithanexample:SupposeIwantamobilephone,soIjustGoogle search it and after that, I changed my mind to buy offline. Data Science helps those companies who are paying for Advertisements for their mobile. So everywhere on theinternet in the social media, in the websites, in the apps everywhere I will see the recommendation of that mobile phone which I searched for. So this will force me to buy online.

8. AirlineRoutingPlanning

With the help of Data Science, Airline Sector is also growing like with the help of it, it becomeseasy topredictflightdelays.Italsohelpstodecidewhethertodirectlylandinto thedestination or takeahaltinbetween likeaflightcanhaveadirectroutefrom Delhi to the U.S.A or itcan haltin between after that reach at the destination.

9. DataScienceinGaming

In most of the gameswhere a user will playwith an opponenti.e.a Computer Opponent, data science concepts are used with machine learning where with the help of past data the Computer will improve its performance. There are many games like Chess, EA Sports, etc. will use Data Science concepts.

10. MedicineandDrugDevelopment

The process of creating medicine is very difficult and time-consuming and has to be done withfulldisciplinedbecauseitisamatterofSomeone'slife.WithoutDataScience,it takes lots of time, resources, and finance or developing new Medicine or drug but with the help of Data Science, it becomes easy because the prediction of success rate can be easily determined based on biological data or factors. The algorithms based on data science will forecasthow this will react tothe human body withoutlab experiments.

11. InDeliveryLogistics

VariousLogisticscompanieslikeDHL,FedEx,etc.makeuseofDataScience.Data Science helps these companies to find the best route for the Shipment of their Products, the besttimesuitedfordelivery,thebestmode of transporttoreach the destination,etc.

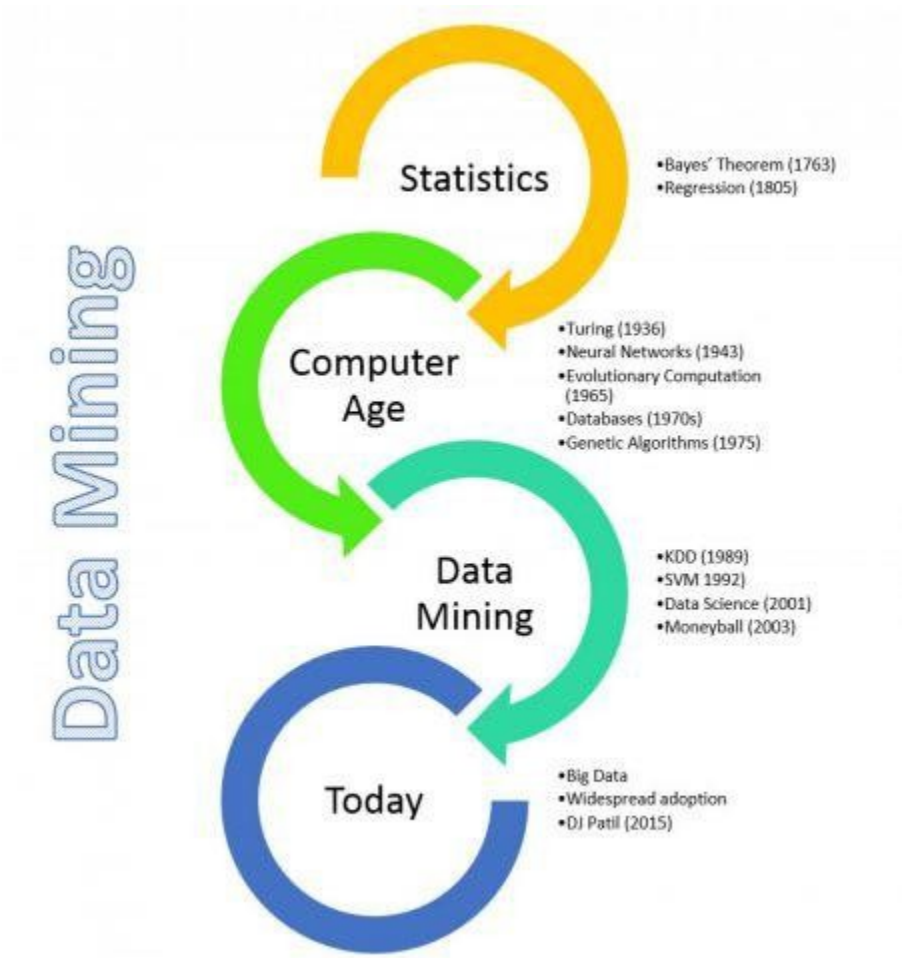
12. Autocomplete

AutoComplete feature isan importantpart of Data Sciencewhere the userwill get the facilitytojusttypeafewlettersorwords,andhewillgetthefeatureofauto-completing the line. In Google Mail, when we are writing formal mail to someone so at that time data scienceconceptofAutocompletefeatureisusedwherehe/sheisanefficientchoiceto auto-complete the whole line.Also in Search Engines in social media, in various apps, AutoComplete feature is widely used.

UNIT-III
DataMiningandMachineLearning

Data Mining: The origins of Data Mining- Data Mining Tasks- OLAP and Multidimensional-Data Analysis-Basic Concept of Association Analysis and ClusterAnalysis.
Machine Learning: History and Evolution – AI Evolution- Statistics vs. Data Mining vs. Data Analytics vs. Data Science – Supervised Learning- Unsupervised Learning- Reinforcement Learning- Frameworks for Building Machine Learning Systems.

DataMining:
TheoriginsofData Mining



- ✓ Data mining is a discipline with a long history. It starts with the early Data Mining methods Bayes' Theorem (1700`s) and Regression analysis (1800`s) which were mostly identifying patterns in data.
- ✓ Data mining is the process of analyzing large data sets (Big Data) from different perspectives and uncovering correlations and patterns to summarize them into useful information.
- ✓ Nowadays it is blended with manytechniques such as artificial intelligence, statistics, data science, database theory and machine learning.
- ✓ Increasing power of technology and complexity of data sets has lead Data Mining to evolvefromstaticdatadeliverytomoredynamicandproactiveinformationdeliveries; from tapes and disks to advanced algorithms and massive databases.

- ✓ In the late 80`s Data Mining term began to be known and used within the research community by statisticians, data analysts, and the management information systems (MIS) communities.
- ✓ By the early 1990`s, data mining was recognized as a sub-process or a step within a largerprocesscalledKnowledgeDiscoveryinDatabases(KDD) .Themostcommonly used definition of KDD is "The nontrivial process of identifying valid, novel, potentiallyuseful, and ultimatelyunderstandable patterns in data" (Fayyad, 1996).

Thesub-processesthatformpartoftheKDDprocessare;

1. Understandingoftheapplicationandidentifyingthegoaloftheprocess
2. Creatingatargetdata set
3. Datacleaningandpre-processing
4. MatchingthegoalsoftheKDDprocess(step1)toaparticulardata-miningmethod.
5. Researchanalysisandhypothesisselection
6. Datamining:Searchingforpatternsofinterestinaparticularform,including classification rules, regression, and clustering
7. Interpretingminedpatterns
8. Actingonthediscoveredanalysis

Evolutionary Step	Business Question	Enabling Technologies	Product Providers	Characteristics
Data Collection (1960s)	"What was my total revenue in the last five years?"	Computers, tapes, disks	IBM, CDC	Retrospective, static data delivery
Data Access (1980s)	"What were unit sales in New England last March?"	Relational databases (RDBMS), Structured Query Language (SQL), ODBC	Oracle, Sybase, Informix, IBM, Microsoft	Retrospective, dynamic data delivery at record level
Data Warehousing & Decision Support (1990s)	"What were unit sales in New England last March? Drill down to Boston."	On-line analytic processing (OLAP), multidimensional databases, data warehouses	Pilot, Comshare, Arbor, Cognos, Microstrategy	Retrospective, dynamic data delivery at multiple levels
Data Mining (Emerging Today)	"What's likely to happen to Boston unit sales next month? Why?"	Advanced algorithms, multiprocessor computers, massive databases	Pilot, Lockheed, IBM, SGI, numerous startups (nascent industry)	Prospective, proactive information delivery

DataMining Tasks:
Data mining, also known as knowledge discovery in data (KDD), is the process ofuncovering patterns and other valuable information from large data sets. Given the evolution ofdata warehousing technology and the growth of big data, adoption of data mining techniques has rapidly accelerated over the last couple of decades, assisting companies by transforming their raw data into useful knowledge.

Data mining functionalities are to perceive the various forms of patterns to be identified in data mining activities. To define the type of patterns to be discovered in data mining activities, data mining features are used. Data mining has a wide application for forecasting and characterizing data in big data.

Dataminingtasksaremajorlycategorizedintotwocategories:descriptiveand predictive.

1. **Descriptivedata mining**:
   Descriptivedata miningoffersadetaileddescriptionofthedata,forexample- it gives insight into what's going on inside the data without anyprior idea. This demonstrates the common characteristics in the results. It includes any information to grasp what's going on in the data without a prior idea.
2. **PredictiveDataMining**:
   Thisallowsuserstoconsider featuresthat arenotspecificallyavailable.Forexample, the projection of the market analysis in the next quarters with the output of the previousquarters,Ingeneral, thepredictiveanalysis forecastsorinfersthe featuresof the data previously available. For an instance: judging by the outcomes of medical records of a patient who suffers from some real illness.

KeyDataMiningTasks
1) CharacterizationandDiscrimination

- **Data Characterization**: The characterization of data is a description of the general characteristics of objects in a target class which creates what are called characteristic rules.
  A database query usually computes the data applicable to a user-specified class and runs through a description component to retrieve the meaning of the data at various abstraction levels.
  Eg;-Bar maps,curves,andpiecharts.
- **Data Discrimination**: Data discriminationcreatesa seriesofrulescalled discriminate rules that is simply a distinction between the two classes aligned with the goal class and the opposite class of the general characteristics of objects.

2) Prediction
To detect the inaccessible data, it uses regression analysis and detects the missing numeric values inthedata.Iftheclass mark isabsent,soclassificationisusedtorender theprediction. Due to its relevance in business intelligence, the prediction is common. If the class mark is absent, so the prediction is performed using classification. There are two methods of predicting data. Due to its relevance in business intelligence, a prediction is common. The prediction ofthe class mark using the previously developed class modeland the prediction of incomplete or incomplete data using prediction analysis are two ways of predicting data.

3) Classification

Classification is used to create data structures of predefined classes, as the model is used to classify new instances whose classification is not understood. The instances used to produce the model are known as data from preparation. Adecision tree or set of classification rules is based on such a form of classification process that can be collected to identify future details, for example by classifying the possible compensation of the employee based on the classification of salaries of related employees in the company.

4) AssociationAnalysis

The link between the data and the rules that bind them is discovered. And two or more data attributes are associated. It associates qualities that are transacted together regularly. They work out what are called the rules of partnerships that are commonly used in the study of stock baskets. To link the attributes, there are two elements. One is the trust that suggeststhe possibilityof both associated together, and another helps, which informs of associations' past occurrence.

5) OutlierAnalysis

Data components that cannot be clustered into a given class or cluster are outliers. They are often referred to as anomalies or surprises and are also very important to remember.
Although in some contexts, outliers can be called noise and discarded, they can disclose useful information in other areas, and hence can be very important and beneficial for their study.

6) ClusterAnalysis

Clusteringis the arrangement of data in groups. Unlike classification, however, class labels are undefined in clustering and it is up to the clustering algorithm to find suitable classes. Clustering is often called unsupervised classification since provided class labels do not execute the classification. Many clustering methods are all based on the concept of maximizing the similarity (intra-class similarity) between objects of the same class and decreasing the similarity between objects in different classes (inter-class similarity).

7) Evolution &DeviationAnalysis

We may uncover patterns and shifts in actions over time, with such distinct analysis, we can find features such as time-series results, periodicity, and similarities in patterns. Many technologies from space science to retail marketing can be found holistically in data processing and features.

- **OLAP(onlineanalyticalprocessing)**

  - OLAP (online analytical processing) is a computing method that enables users to easily andselectivelyextract andquery data inordertoanalyzeit from different points of view. OLAP business intelligence queries often aid in trends analysis, financial reporting, sales forecasting, budgeting and other planning purposes.

- For example, a user can request that data be analyzed to display a spreadsheetshowing all of a company's beach ball products sold in Florida in the month of July, comparerevenue figureswiththoseforthesameproductsinSeptemberandthenseea comparison of other product sales in Florida in the same time period.

HowOLAPsystemswork

- ✓ To facilitate this kind of analysis, data is collected from multiple data sources and stored in data warehouses then cleansed and organized into data cubes.

- ✓ EachOLAP cube contains data categorized by dimensions (such as customers, geographic sales region and time period) derived bydimensional tables in the data warehouses.

- ✓ Dimensions are then populated by members (such as customer names, countries and months) that are organized hierarchically. OLAP cubes are often pre-summarized across dimensions to drastically improve query time over relational databases.

❖ **TypesofOLAPServers**

WehavefourtypesofOLAPservers−

- RelationalOLAP (ROLAP)
- MultidimensionalOLAP (MOLAP)
- HybridOLAP(HOLAP)
- SpecializedSQLServers

1. RelationalOLAP

ROLAP servers are placed between relational back-end server and client front-end tools. To store and manage warehouse data, ROLAP uses relational or extended-relational DBMS.
ROLAPincludesthefollowing−

- Implementationofaggregationnavigation logic.
- OptimizationforeachDBMSbackend.
- Additionaltoolsand services.

2. MultidimensionalOLAP

MOLAP uses array-based multidimensional storage engines for multidimensional views of data. With multidimensional data stores, the storage utilization may be low if the data set is sparse. Therefore, many MOLAP server use two levels of data storage representation to handle dense and sparse data sets.

3. HybridOLAP

Hybrid OLAP is a combination of both ROLAP and MOLAP. It offers higher scalability of ROLAP and faster computation of MOLAP. HOLAP servers allows to store the large data volumes of detailed information. The aggregations are stored separately in MOLAP store.

4. SpecializedSQLServers

Specialized SQL servers provide advanced querylanguage and queryprocessing support for SQL queries over star and snowflake schemas in a read-only environment.

❖ **OLAPOperations**

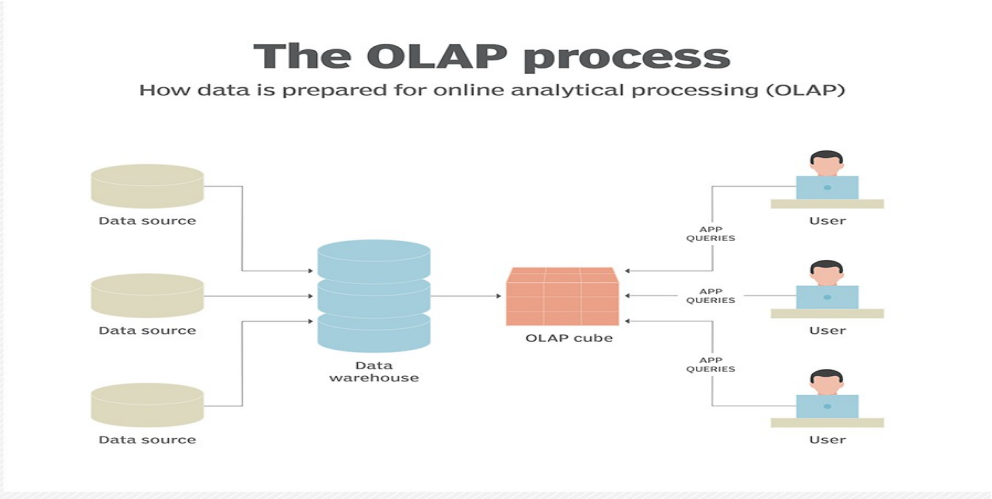
SinceOLAPserversarebasedonmultidimensionalviewofdata,OLAPoperationsin **multidimensionaldata.**

HereisthelistofOLAPoperations−

- Roll-up
- Drill-down
- Sliceanddice
- Pivot(rotate)
  - **Roll-up.** Also known as*consolidation*, or*drill-up*, this operation summarizes the data along the dimension.
  - **Drill-down.** This allows analysts to navigate deeper among the dimensions of data, for example drilling down from "time period" to "years" and "months" to chart sales growth for a product.
  - **Slice.** This enables ananalyst to takeone levelof information for display, suchas "sales in 2017."
  - **Dice.** This allows an analyst to select data from multiple dimensions to analyze, such as "sales of blue beach balls in Iowa in 2017."
  - **Pivot.**Analystscangainanew view ofdata byrotatingthe data axesofthe cube.

OLAP software then locates the intersection of dimensions, such as all products sold in the Eastern region above a certain price during a certain time period, and displays them. The result is the "measure"; each OLAP cube has at least one to perhaps hundreds of measures, which are derived from information stored in fact tables in the data warehouse.

OLAP begins with data accumulated from multiple sources and stored in a data warehouse.The data is then cleansed and stored in OLAP cubes, which users run queries



against.

❖ AssociationRuleMining

Association analysis is useful for discovering interesting relationships hidden in large datasets. The uncovered relationships can be represented in the form of **association rules**or setsof frequent items.

Given a set oftransactions, find rules that will predict the occurrence ofan item based on the occurrences of other items in the transaction

Market-Baskettransactions

TID	Items
1	Bread,Milk
2	Bread,Diaper, Beer, Eggs
3	Milk,Diaper,Beer, Coke
4	Bread,Milk,Diaper,Beer
5	Bread,Milk,Diaper,Coke

Implicationmeansco-occurrence,notcausality! Example of Association Rules

{Beer}□{Diaper}
{Milk,Bread}□{Eggs,Coke}
{Beer,Bread}□{Milk}

SupportCount( )–Frequencyofoccurrenceofaitemset.
Here({Milk,Bread,Diaper})=2

FrequentItemset– Anitemsetwhosesupportisgreaterthanorequaltominsup threshold.

AssociationRule–Animplicationexpressionof theformX->Y,whereXandYareany 2 item sets.

Example:{Milk,Diaper}->{Beer}

RuleEvaluationMetrics–

Support(s) –
The number of transactions that include items in the {X} and {Y} parts of the rule as a percentage of the total number of transaction. It is a measure of how frequently the collection of itemsoccurs togetheras a percentage of all transactions.

Support = (X+Y)÷ total –
Itisinterpretedasfractionoftransactionsthatcontainbothx andY.

Confidence(c) –
It is the ratio of the no of transactions that includes all items in {B} as well as the no of transactionsthatincludesallitemsin{A}tothenooftransactionsthatincludesall items in {A}.

Conf(X=>Y) = Supp(XY) ÷ Supp(X) –
It measures how often each item in Y appears in transactions that containsitemsin X also.

- **Lift (l)** –
  The lift of the rule X=>Y is the confidence of the rule divided by the expected confidence, assuming that the item sets X and Y are independent of each other. The expected confidenceis the confidencedividedby the frequency of {Y}.
- **Lift(X=>Y) = Conf(X=>Y) ÷ Supp(Y)** –
  Liftvaluenear1indicatesXandYalmostoftenappeartogetherasexpected, greaterthan1meanstheyappeartogethermorethanexpectedandlessthan1

meanstheyappearlessthanexpected.Greaterliftvaluesindicatestronger association.

- **Example –** From the above table, {Milk, Diaper}=>{Beer}
  s= ({Milk, Diaper, Beer}) ÷|T|
  =2/5
  =0.4

- c= (Milk,Diaper,Beer)÷(Milk,Diaper)
  =2/3
  =0.67
- l=Supp({Milk,Diaper,Beer})÷Supp({Milk,Diaper})*Supp({Beer})
  =0.4/(0.6*0.6)
  =1.11

TheAssociationruleisveryusefulinanalyzingdatasets.Thedataiscollectedusing bar-code scanners in supermarkets. Such databases consist of a large number of transaction records whichlistallitemsboughtbyacustomeronasinglepurchase.So the manager could know if certain groups of items are consistently purchased togetherandusethisdataforadjustingstorelayouts,cross-selling,promotionsbasedon statistics.

❖ **DataMining–ClusterAnalysis**

ClusterAnalysisisthe process tofindsimilargroups of objects in ordertoformclusters. It isanunsupervisedmachinelearning-basedalgorithm thatacts on unlabelleddata.Agroup of data points would comprise together to form a cluster in which all the objects wouldbelong to the same group.

Cluster:

The given data is divided into different groups by combining similar objects into a group. Thisgroupisnothingbutacluster.Aclusterisnothingbutacollection of similardata which is grouped together.

For example, consider a dataset of vehicles is given in which it contains information about different vehicles like cars, buses, bicycles, etc. As it is unsupervised learning there are no class labels like Cars, Bikes, etc for all the vehicles, all the data is combined and is not in a structured manner.

Now our task is to convert the unlabelled data to labelled data and it can be done using clusters.

The main idea of cluster analysis is that it would arrange all the data points by forming clusters like cars cluster which contains all the cars, bikes clusters which contains all the bikes, etc.

Simplyitispartitioningofsimilarobjectswhichareappliedonunlabelleddata.

PropertiesofClustering:

1. **Clustering Scalability:** Nowadays there is a vast amount of data and should be dealing with huge databases. In order to handle extensive databases, the clustering algorithm should be scalable. Data should be scalable if it is not scalable, then we can't get the appropriate result and would lead to wrong results.
2. **High Dimensionality:** The algorithm should be able to handle high dimensional space along with the data of small size.
3. **Algorithm Usability with multiple data kinds:** Different kinds of data canbe used with algorithms of clustering. It should be capable of dealing with differenttypes of datalikediscrete,categorical andinterval-baseddata, binary data etc.
4. **Dealing with unstructured data:** These would be some databases thatcontainmissingvalues,noisy orerroneous data. If thealgorithmsare sensitivetosuchdatathenitmayleadtopoorqualityclusters.Soitshould be able to handle

- **ClusteringMethods:**

Theclusteringmethodscanbeclassifiedintothefollowingcategories:

- PartitioningMethod
- HierarchicalMethod
- Density-basedMethod
- Grid-BasedMethod
- Model-BasedMethod
- Constraint-basedMethod

1. **Partitioning Method:** It is used to make partitions on the data in order to form clusters.If "n"partitionsaredone on"p" objects of the database theneachpartition is represented by a cluster and n < p.The two conditions which need to be satisfied with this Partitioning Clustering Method are:
   - Oneobjectiveshouldonlybelongtoonlyonegroup.
   - Thereshouldbenogroupwithoutevenasinglepurpose.

In thepartitioningmethod,there isone techniquecallediterativerelocation,whichmeans the objectwill bemovedfrom onegroup toanother toimprovethe partitioning.

2. **HierarchicalMethod:** Inthismethod,ahierarchicaldecompositionofthegiven setof data objects iscreated.Wecan classifyhierarchicalmethods andwill beable to know the purpose of classification on the basis of how the hierarchical decomposition is formed. There are two types of approaches for the creation of hierarchical decomposition, they are:
   - **Agglomerative Approach:** The agglomerative approach is also known as the bottom-upapproach.Initially,thegivendataisdividedinwhichtheobjects form separate groups. Thereafter it keeps on merging the objects or the groupsthat are close to one another which means that they exhibitsimilar properties. Thismergingprocess continues until the termination condition holds.

- **Divisive Approach:** The divisive approach is also known as the top-down approach. In this approach, we would start with the data objects that are in the same cluster. The group of individual clusters is divided into small clusters by continuousiteration.Theiterationcontinuesuntiltheconditionoftermination is met or until each cluster contains one object.

Once the group is split or merged then it can never be undone as it is a rigid method and isnot so flexible. The two approaches which can be used to improve the HierarchicalClustering Quality in Data Mining are: –

- One should carefully analyze the linkages of the object at every partitioning of hierarchical clustering.
- One can use a hierarchical agglomerative algorithm for the integration of hierarchical agglomeration. In this approach, first, the objects are grouped into micro-clusters.Aftergroupingdataobjectsintomicroclusters,macro clustering is performed on the micro cluster.

3. **Density-Based Method:** The density-basedmethod mainly focuses on density.In this method, the given cluster will keep on growing continuously as long as thedensityintheneighbourhoodexceedssomethreshold,i.e,foreachdatapoint withinagivencluster.The radiusof agivenclusterhastocontainatleasta minimum number of points.
4. **Grid-Based Method:** In the Grid-Based method a grid is formed using the object together,i.e,the objectspaceisquantizedintoafinitenumberof cellsthatform a grid structure. One of the major advantages of the grid-based method is fastprocessingtimeanditisdependentonlyonthenumberofcellsineachdimension in the quantized space.The processing time for this method is much faster so it can save time.
5. **Model-BasedMethod:** Inthemodel-basedmethod,alltheclustersare hypothesized in order to find the data which is best suited for the model. The clustering of the density function is used to locate the clusters for a given model. It reflects thespatial distribution of datapointsandalsoprovidesa way to automatically determine the number of clusters based on standard statistics, taking outlier or noise intoaccount. Therefore ityields robust clusteringmethods.

6. **Constraint-BasedMethod:** Theconstraint-basedclusteringmethodisperformed bytheincorporationofapplicationoruser-orientedconstraints.Aconstraintrefers totheuserexpectationortheprepertiesofthedesiredclusteringresults. Constraintsprovideuswithaninteractivewayofcommunicationwiththe clustering process. Constraints can be specified by the user or the application requirement.

ApplicationsofClusterAnalysis:

- Itiswidelyusedinimageprocessing,dataanalysis,andpatternrecognition.
- Ithelpsmarketerstofindthedistinctgroupsintheircustomerbaseandtheycan characterize their customer groups by using purchasing patterns.
- Itcanbeusedinthefieldofbiology,byderivinganimalandplanttaxonomies, identifying genes with the same capabilities.
- Italsohelpsininformationdiscoverybyclassifyingdocumentsontheweb.

❖ MachineLearning

- ✓ **Machine learning** (**ML**) is the study of computer algorithmsthat can improve automaticallythroughexperienceandbytheuseofdata.Itisseenasapart ofartificial intelligence. Machine learning algorithms build a model based on sample data, known astraining data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide varietyofapplications,suchasinmedicine, emailfiltering,speechrecognition,andcomputer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.
- ✓ A subset of machine learning is closely related to computational statistics, which focuses on making predictions using computers; but not all machine learning is statistical learning. The study of mathematical optimizationdelivers methods, theory and application domains to the field of machine learning. Data miningis a relatedfield of study, focusing onexploratory data analysisthroughunsupervised learning. Some implementations of machine learning use data and neural networksin a waythat mimics the working of a biological brain. In its application across business problems, machine learning is also referred to as predictive analytics.

❖ Howmachinelearningworks

1. **A Decision Process**: In general, machine learning algorithms are used to make a prediction or classification. Based on some input data, which can be labelled or unlabeled,youralgorithm will produce an estimate abouta pattern in the data.
2. **AnErrorFunction**:Anerrorfunctionservestoevaluate the predictionof the model. If there are known examples, an error function can make a comparison to assess the accuracy of the model.
3. **A Model Optimization Process**: If the model can fit better to the data points in the trainingset,thenweightsareadjustedtoreducethediscrepancybetweenthe known example and the model estimate. The algorithm will repeat this evaluate and optimize process, updating weights autonomously until a threshold of accuracy has been met.

❖ Machinelearning methods

Machinelearningclassifiersfallintothreeprimary categories.

1. **Supervisedmachinelearning**

Supervised learning, also known as supervised machine learning, is defined by its use of labelled datasets to train algorithms that to classify data or predict outcomes accurately. As input data is fed into the model, it adjusts its weights until the model has been fitted appropriately. This occurs as part of the cross validation process to ensure that the model avoidsoverfittingorunderfitting.Supervised learninghelpsorganizationssolve foravariety of real-world problems at scale, such as classifying spam in a separate folder from yourinbox.Somemethodsusedinsupervisedlearningincludeneuralnetworks,naïvebayes,

linear regression, logistic regression, random forest, support vector machine (SVM), and more.

2. **Unsupervisedmachinelearning**

   Unsupervised learning, also knownas unsupervised machine learning, uses machine learning algorithms to analyze and cluster unlabeled datasets. These algorithms discover hidden patterns or data groupings without the need for human intervention. Its ability to discover similarities and differences in information make it the ideal solution for exploratory data analysis, cross-selling strategies, customer segmentation, image and pattern recognition. It's also used to reduce the number of features in a model through the process of dimensionality reduction; principal component analysis (PCA) and singular value decomposition (SVD) are two common approaches for this. Other algorithms used in unsupervised learning include neural networks, k-means clustering, probabilistic clustering methods, and more.

3. **Semi-supervisedlearning**

   Semi-supervised learning offers a happy medium between supervised and unsupervised learning. During training, it usesa smaller labelled datasetto guide classificationand feature extraction from a larger, unlabeled data set. Semi-supervised learning can solve the problem ofhaving not enough labelled data(or not being able to affordto labelenoughdata)totraina supervised learning algorithm.

- **Reinforcementmachinelearning**

  Reinforcement machine learning is a behavioural machine learning model that is similar to supervised learning, but the algorithm isn't trained using sample data. This model learnsas it goes by using trial and error. A sequence of successful outcomes will be reinforced todevelop the best recommendation or policy for a given problem.

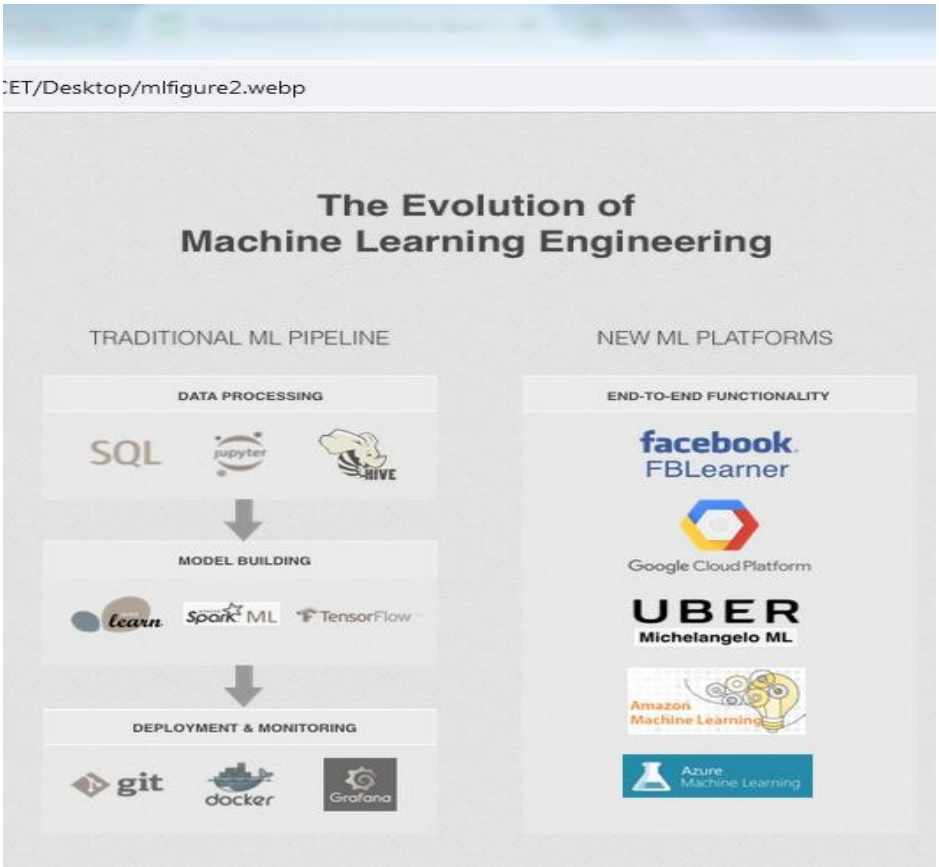
  MachineLearning isa sub-set ofartificialintelligencewherecomputer algorithms areusedtoautonomously learn fromdata and information. In machine learning computers don't have tobe explicitly programmed but can change and improve their algorithms by themselves. Machine learning algorithms enable computers to communicate with humans, autonomouslydrive cars, write and publish sport match reports, and find terrorist suspects.

- **Theoriginsofmachinelearning**
  - Theterm*machinelearning* wascoinedin1959byArthurSamuel,an AmericanIBMer andpioneerinthefieldofcomputergaming and artificialintelligence.Alsothesynonym*self-teachingcomputers*wasusedinthistime period. A representative book of the machine learning research during the 1960s was the Nilsson's book on Learning Machines, dealing mostly with machine learning for pattern classification. Interest related to pattern recognition continued into the 1970s, as described byDuda and Hart in 1973. In 1981 a report was given on using teaching strategies so that a neural network learns to recognize 40 characters (26 letters, 10 digits, and 4 special symbols) from a computer terminal.
  - Tom M. Mitchellprovided a widely quoted, more formal definition of the algorithms studied in the machine learning field: "A computer program is said to learn from experience*E* with respect to some class of tasks *T*and performance measure *P* if its performance at tasks in*T*, as measured by*P*, improves with experience *E*."This definitionofthetasksinwhichmachinelearningisconcernedoffersa fundamentallyoperational definitionrather than defining the field in cognitive terms.

This follows Alan Turing's proposal in his paper "Computing Machinery andIntelligence", in which the question "Can machines think?" is replaced with the question "Can machines do what we (as thinking entities) can do?".

- Modern day machine learning has two objectives, one is to classify data based on models which have been developed, and the other purpose is to make predictions for future outcomes based on these models. A hypothetical algorithm specific to classifying data may use computer vision of moles coupled with supervised learningin order to train it to classify the cancerous moles. A machine learning algorithm for stock trading may inform the trader of future potential predictions.

EVOLUTION OFMACHINELEARNING:



MachineLearning Frameworks

A Machine Learning Framework is an interface, library or tool which allows developers to build machine learning models easily, without getting into the depth of the underlying algorithms. Let's discuss the Top 10 Machine Learning Frameworks in detail:

TensorFlow

Google'sTensorflow is one of the most popular frameworks today. It is an open-source software library for numerical computation using data flow graphs. TensorFlow implements data flow graphs, where batches of data or tensors can be processed by a series of algorithms described by a graph.

Theano

Theano is wonderfully folded over Keras, an abnormal state neuralsystems library, that runs nearlyinparallelwiththe Theano library. Keras' fundamentalfavorable position isthat it isa moderatePython libraryfor profound discovering that can keep running over Theano orTensorFlow.
It was created to make actualizing profound learning models as quick and simple as feasible forinnovativework.DischargedunderthetolerantMITpermit,itkeepsrunningonPython 2.7or3.5andcanconsistentlyexecuteonGPUsandCPUsgiventhe basicstructures.

Sci-KitLearn
Scikit-learnis one ofthe most well-knownML libraries. It is preferable for administered and unsupervised learning calculations. Precedents implement direct and calculated relapses, choice trees, bunching, k-implies, etc.
This framework involves a lot of calculations for regular AI and data mining assignments, including bunching, relapse, and order.
Caffe
Caffe is another popular learning structure made with articulation, speed, and measured quality as the utmost priority. It is created by the Berkeley Vision and Learning Center (BVLC) and by network donors.
Google's DeepDream depends on Caffe Framework. This structure is a BSD-authorized C++ library with Python Interface.

H20
H20 is an open-source machine learning platform. It is an artificial intelligence toolwhich is business-oriented and helps in making a decision based on data and enables the user to draw insights. It is mostly used for predictive modeling, risk and fraud analysis, insuranceanalytics, advertising technology, healthcare, and customer intelligence.

Amazon Machine Learning
Amazon Machine Learning provides visualization tools that help you go through the process ofcreatingmachinelearning(ML)modelswithouthavingtolearncomplex MLalgorithms and technology.
It is a service that makes it easy for developers of all skill levels to use machine learning technology. It connects to data stored in Amazon S3, Redshift, or RDS, and can run binary classification, multiclass categorization, or regressionon the datato build a model.

Torch
This framework provides wide support for machine learning algorithms to GPUs first. It is easy to use and efficient because of the easy and fast scripting language, **LuaJIT**, and an underlying **C/CUDA** implementation.
The goal of Torch is to have maximum flexibility and speed in building your scientific algorithms along with an extremely simple process.

GoogleCloudMLEngine

Cloud Machine Learning Engine is a managed service that helps developers and data scientists in building and running superior machine learning models in production.
It offers training and prediction services that can be used together or individually. It is usedby enterprises to solve problems like ensuring food safety, clouds in satellite images, responding four times faster to customer emails, etc.

AzureML Studio
This Framework allows Microsoft Azureuserstocreate andtrain models, thenturntheminto APIs that can be consumed byother services. Also, you canconnect your ownAzure storage to the service for larger models.
To use the Azure ML Studio, you don't even need an account to tryout the service. You can log in anonymously and use Azure ML Studio for up to eight hours.
SparkMLLib
This is Apache Spark's machine learning library. The goal of this framework is to make practical machine learning scalable and easy.

- **ABriefHistory ofArtificialIntelligence**

The beginnings of modern AI can be traced to classical philosophers' attempts to describe humanthinkingasa symbolic system. But the fieldofAI wasn't formallyfoundeduntil1956, at aconferenceat DartmouthCollege, inHanover, NewHampshire, wheretheterm"artificialintelligence" was coined.

The beginnings of modern AI can be traced to classical philosophers' attempts to describe humanthinkingasa symbolic system. But the fieldofAI wasn't formallyfoundeduntil1956, at aconferenceat DartmouthCollege, inHanover, NewHampshire, wheretheterm"artificialintelligence" was coined.

Despite artificialintelligence hasbeenpresent formillennia, it wasnot untilthe 1950sthat its realpotentialwas investigated. Agenerationofscientists, physicists, and intellectualshad the idea of AI, but it wasn't until Alan Turing, a British polymath, proposed that people solve problems and make decisions using available information and also a reason.
The difficulty of computers was the major stumbling block to expansion. They needed to adapt fundamentally before they could expand any further. Machines could execute ordersbut not store them. Until 1974, financing was also a problem.
By1974, computers had become extremely popular. Theywere now quicker, less expensive, and capable of storing more data.

- **AIResearchToday**

AIresearchisongoingandexpandingintoday'sworld.AIresearchhasgrownatapaceof 12.9percentannuallyoverthelastfiveyears,asperAliceBonasio, atechnologyjournalist.

China is expected to overtake the United States as the world's leading source of AI technology in the next 4 years, having overtaken the United States' second position in 2004 and is rapidly closing in on Europe's top rank.
In the area of artificial intelligence development, Europe is the largest and most diverse continent, with significant levels of international collaboration. India is the 3rd largestcountry in AI research output, behind China and the USA.

- **AI inthePresent**

Artificial intelligence is being utilized for so many things and has so much promise that it's difficult to imagine our future without it, related to business.
Artificial intelligence technologies are boosting productivity like never seen before, from workflow management solutions to trend forecasts and even the way companies buy advertisements.
Artificial Intelligence can gather and organize vast volumes of data in order to draw inferences and estimates that are outside of the human ability to comprehend manually. Italso improves organizationalefficiencywhile lowering the risk ofa mistake, and it identifies unusual patterns, such as spam and frauds, instantaneously to alert organizations about suspicious behaviour, among other things. AI has grown in importance and sophistication to the point that a Japanese investment firm became the first to propose an AI Board Memberfor its ability to forecast market trends faster than humans.
Artificialintelligencewillindeed beand isalreadybeingused inmanyaspectsoflife, suchas self-driving cars in the coming years, more precise weather forecasting, and earlier health diagnoses, to mention a few.

- **AIinThe Future**

It has been suggested that we are onthe verge ofthe 4th Industrial Revolution, which will be unlike any of the previous three. From steam and water power through electricity and manufacturing process, computerization, and now, the question of what it is to be human is being challenged.
Smarter technology in our factories and workplaces, as well as linked equipment that will communicate, view the entire production process, and make autonomous choices, are just a few of the methods the Industrial Revolution will lead to business improvements. One of the most significant benefits of the 4thIndustrialRevolution is the abilityto improve the world's populace's quality of life and increase income levels. As robots, humans, and smart devices work on improving supply chains and warehousing, our businesses and organizations are becoming "smarter" and more productive.

- **AIinDifferent Industries**

Artificial intelligence (AI) may help you enhance the value of your company in a variety of ways. It may help you optimize your operations, increase total revenue, and focus your staff on more essential duties if applied correctly. As a result, AI is being utilized in a variety of industries throughout the world, including health care, finance, manufacturing, and others.

HealthCare

AI is proven to be uplift in the healthcare business. It's enhancing nearly every area of the industry, from data security to robot-assisted operations. AI is finally providing this sector, whichhas beenharmedbyinefficient proceduresandgrowingprices, a much-needed facelift.

Automotive

Self-driving vehicles are certainly something you've heard of, and they're a hint that the future is almost here. It's no longer science fiction; the autonomous car is already a reality.As per recent projections, by 2040, roughly 33 million automobiles with self-driving capability are projected to be on the road.

Finance

According to experts, the banking industry and AI are a perfect combination. Real-time data transmission, accuracy, and large-scale data processing are the most important elements driving the financial sector. Because AI is ideal for these tasks, the banking industry is recognizing its effectiveness and precision and incorporating machine learning, statistical arbitrage, adaptive cognition, chatbots, and automation into its business operations.

TransportationandTravel

From recommending the best route for drivers to arranging travel reservations remotely, AI has now become a gigantic trend in this business. End consumers are finding it easier to navigate thanks to artificial intelligence. Furthermore, travel businesses that integrate AI into their systems profit from Smartphone usage.

E-Commerce

Have you ever come upon a picture ofclothing that you were hunting foron one website but couldn't find on another? Well, that is done by AI. It's due to the machine learningtechniques that businesses employ to develop strong client connections. These technologies not only personalize customers' experiences but also assist businesses in increasing sales.

Conclusion

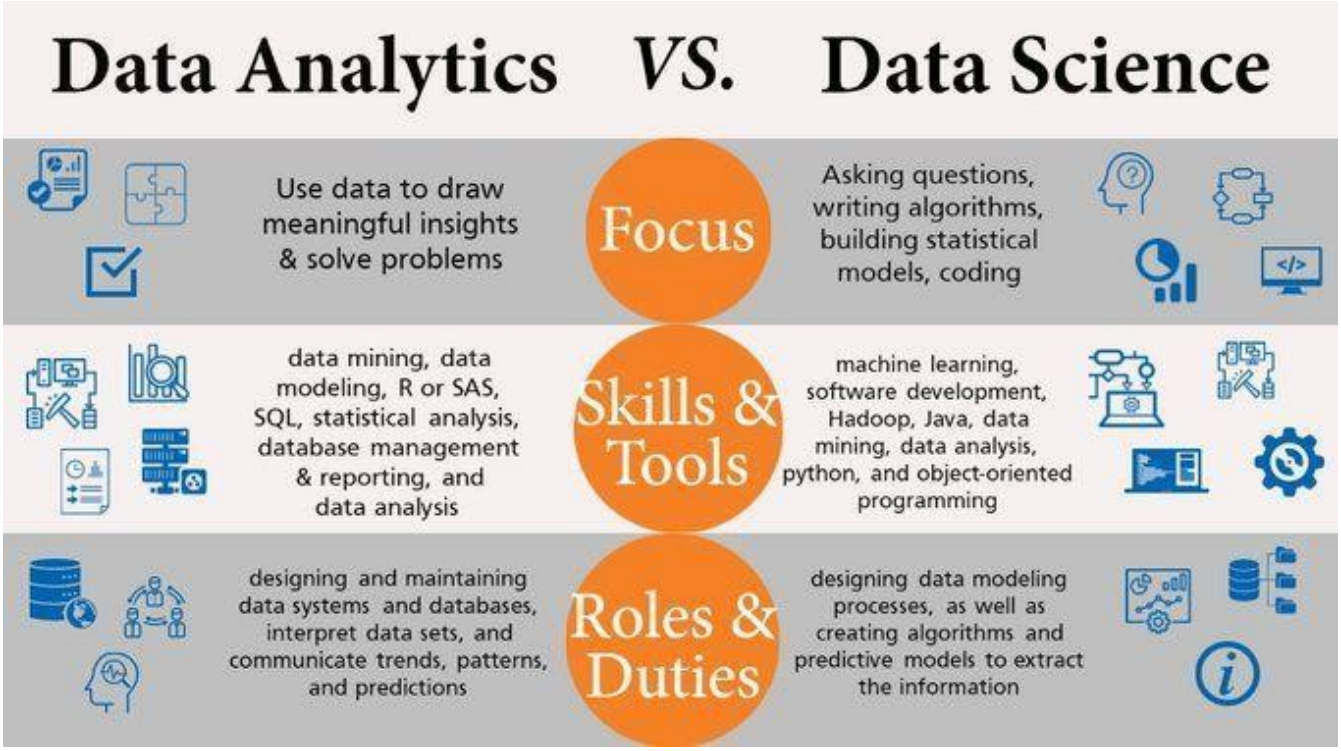
In the early twenty-first century, no place has had a larger influence on AI than theworkplace. Machine-learning techniques are resulting in productivity gains that have never beenobservedbefore.AIistransformingthewaywedobusiness,fromworkflow

management solutions to trend forecasts and even the way businesses buy advertising. AI research has so much promise that it's becoming difficult to envisage a world without it. Be its self-driving vehicles, more precise weather predictions, or space travel, AI will be prevalent in everyday life by 2030.

- **`Statisticsvs.DataMining**

DataMining	Statistics
Data mining is a process of extracting useful information, pattern, and trends from huge data sets and utilizes them to make a data-driven decision.	Statistics refers to the analysis and presentation of numeric data, and it is the major part of all data mining algorithm.
Thedatausedindataminingisnumericornon-numeric.	Thedatausedinthestatisticisnumericonly.
Indatamining,datacollectionisnotmore important.	In statistics, data collection is more important.
Thetypesofdataminingareclustering, classification, association, neural network, sequence-based analysis, visualization, etc.	The types of statistics are descriptive statistical and Inferential statistical.
Itissuitablefor hugedatasets.	Itissuitableforsmallerdataset.
Dataminingisaninductiveprocess.Itmeansthe generation of new theory from data.	Statistics is the deductive process. It does not indulge in making any predictions.
Datacleaningisa partofdata mining.	In statistics,clean data is used toimplement the statistical method.
It requires less user interaction to validate the model, so it is easy to automate.	Itrequiresuserinteractiontovalidatethemodel, so it is complex automate.
Data mining applications include financial Data Analysis, Retail Industry, Telecommunication Industry, Biological Data Analysis, Certain Scientific Applications, etc.	The application of statistics includes biostatistics, quality control, demography, operational research, etc.

- **Data Science vs.DataAnalytics**


Data Analytics VS. Data Science
Use data to draw meaningful insights & solve problems
Focus
Asking questions, writing algorithms, building statistical models, coding
data mining, data modeling, R or SAS, SQL, statistical analysis, database management & reporting, and data analysis
Skills & Tools
machine learning, software development, Hadoop, Java, data mining, data analysis, python, and object-oriented programming
designing and maintaining data systems and databases, interpret data sets, and communicate trends, patterns, and predictions
Roles & Duties
designing data modeling processes, as well as creating algorithms and predictive models to extract the information

UNITIV
ApplicationsofBusiness Analytics

Overview of Business Analytics Application : Financial Analytics- Marketing Analytics- HR Analytics – Supply Chain Analytics- Retail Industry- Sales Analytics- Web & Social Media Analytics- Healthcare Analytics- Energy Analytics- Transportation Analytics-Lending Analytics- Sports Analytics- Future of Business Analytics.

FinancialAnalytics

Financial analytics is a concept that provides different views on the business' financial data.It helps give in-depth knowledge and take strategic actions against them to improve your business'overallperformance. Financialanalyticsisasubset of BI &EPM and hasan impact oneveryaspect of your business. It plays a crucialrole incalculating your business' profit. It helps you answer every business question related to your business while letting your forecast the future of your business.

Sowhyisfinancial analyticsimportant?

- Today'sbusinessesrequiretimelyinformationfordecision-making purposes
- Everycompanyneedsprudentfinancialplanningandforecasting
- Thediverseneedsofthetraditionalfinancialdepartment,andadvancementsin technology, all point to the need for financial analytics.
- Financialanalyticscanhelpshapeupthebusiness'futuregoals.Itcanhelpyou improve the decision-making strategies for your business.
- Financial analyticscanhelpyoufocusonmeasuringandmanagingyourbusiness' tangible assets such as cash and equipment.
- It provides an in-depth insightinto the organization'sfinancial status andimproves the cash flow, profitability, and business value.

Importantfinancialanalyticsyouneedtoknow

In today's data-driven world, analytics is critical for any business that wants to remain competitive. Financial analytics can help you understand your business' past and present performance and make strategic decisions. Here are some of the critical financial analytics that any company, size notwithstanding, should be implementing.

1. Predictivesalesanalytics

Sales revenue is critical for every business. As such, accurate sales projection has essential strategic and technical implications forthe organization. Apredictive salesanalytics involves coming up with an informed sales forecast. There are many approaches to predicting sales, such as the use of correlation analysis or use of past trends to forecast your sales. Predictive sales analytics can help you plan and manage your business' peaks and troughs.

2. Clientprofitabilityanalytics

Every business needs to differentiate between clients that make them money and clients that lose them money. Customer profitability typically falls within the80/20 rule, where 20 percent of the clients account for 80 percent of the profits, and 20 percent of the clients account for 80 percent of customer-related expenses. Understanding of which is vital.

Byunderstanding yourcustomers'profitability, youwillbeableto analyzeeveryclient group and gains useful insight. However, the greatest challenge to customer profitability analytics comes in when you fail to analyze the client's contribution to the organization.

3. Productprofitabilityanalytics

For organizations to remain competitive within an industry, organizations need to know where they are making, and losing money. Product profitability analytics can help you establish the profitability of every product rather than analyzing the business as a whole. To do this, you need to assess each product individually. Product profitability analytics can also help you establish profitability insights across the product range so you can make better decisions and protect your profit and growth over time.

4. Cashflowanalytics

You need a certain amount of cash to run the organization on a day-to-day basis.Cash flowisthe lifebloodofyourbusiness.Understandingcashflow iscrucialforgaugingthehealthof the business. Cash flow analytics involves the use of real-time indicators like the Working Capital Ratio and Cash Conversion Cycle. You can also predict cash flow using tools like regression analysis. Besides helping with cash flow management and ensuring that you have enough money for day-to-day operations, cash flow analytics can also help you support a range of business functions.

5. Value-drivenanalytics

Most organizations have a sense of where they are going to and what they are hoping to achieve. These goals can be formal and listed on a strategy map that pinpoints the business' value drivers. These value drivers are the vital drivers that the organization needs to pull to realize its strategic goals. Value driver analytics assesses these levers to ensure that they can deliver the expected outcome.

6. Shareholdervalueanalytics

The profits and losses, and their interpretation by analysts, investors, and the media can influenceyourbusiness'performanceonthestockmarket.Shareholdervalueanalytics

calculates the value of the company by looking at the returns it is providing to shareholders. In other words, it measures the financial repercussions of a strategy and reports how much value the strategy in question is delivering to the shareholders. Shareholder value analytics is used concurrently with profit and revenue analytics. You can use tools like Economic Value Added (EVA) to measure the shareholder value analytics.

Conclusion

Financial analytics is a valuable tool that every organization, small and large, should use to manage and measure its progress. Done right, it can help the organization adapt to the trends that affect its operations.

❖ **MarketingAnalytics**

Modern marketing is a data-driven process fueled by analytics. Without analyzing relevant key performance indicators (KPI), businesses can't tell whether their marketing efforts are providing the expected return on investment (ROI). Marketing analytics is the key to evaluating past performance and determining how to improve it going forward.

Marketing analytics is a set of technologies and methods used to transform raw data into marketing insights. The goal of marketing analytics is to maximize ROI from an enterprise's marketing initiatives. Marketing analytics encompasses tools for planning, managing, and evaluating these efforts across every channel.

Marketers use established business metrics, and sometimes create new KPIs, to measure the success of their organizations' marketing initiatives. These metrics include:

- Profitabilitysegmentedbydemographic
- Churnrate
- Customerlifetimevalue
- Customersatisfaction
- Publicperception

Businesses can analyze performance indicators alongside other data, such as customerprofiles or demographic trends, to reveal the causal links between marketing decisions and actual sales.

❖ **Importanceofmarketing analytics**

Marketinganalytics makesadvertisingmoreeffectiveandautomates manyrotetasks:

- Marketing analytics helps stakeholders achieve a comprehensive view across all marketing channels, such as pay-per-click (PPC) advertising, email marketing, and social media. Analytics can clarify the big picture, as well as dig down into more granular marketing trends.

- Marketing analyticstools improve lead generationbyproviding the insightsneeded to optimize advertising efforts and target the most profitable consumers. Better leads generate more sales and improved ROI.

- Marketing analytics provides insights into customer behavior and preferences. Businesses can then tailor their marketing initiatives to meet the needs of individual consumers.

- Marketing analytics enables real-time decision support as well as proactive management. Modern analytics tools make it easy for stakeholders to analyze data asit comes in, so marketing can be adjusted as required by changing trends, and theyalso allow businesses to use predictive analytics to anticipate those trends rather than react to them.

❖ BenefitsofMarketingAnalytics

Marketing analytics can benefit organizations' marketing initiatives across all channels. Enterprises should consider the many applications of marketing analytics and determine which may be valuable to them.

Understandsearchmarketing

Many organizations access huge markets through search engines like Google, where consumers often begin their purchasing journeys. Search engine marketing (SEM) promotes businesses and raises online visibility through advertising on search engine results pages (SERP). Revenue from digital advertising in the U.S. breaks new records every year, and search advertising accounts for **almost half of this revenue**. Businesses can use services like Google Ads and Bing Ads to expand their reach.

Organizations need marketing analytics to track and optimize the performance of their SEM efforts. One application of SEM analytics might involve serving different versions of the same ad to a randomized set ofbrowsers and then comparing the performance ofthese ads in real time.

Search engine optimization (SEO) involves adjusting web content and structure to improve organic search engine rankings. An enterprise can use SEO to reach more consumers and enhance its brand. Tools like **Google Analytics**allow businesses to track relevant KPIs and analyze how their SEO initiatives are progressing and how to improve them.

Analyzesocialmediaengagement

More than**a third of the world's population**— including **98% of digital consumers**— spend time on social media, averaging almost two and a half hours per day on theseplatforms. While SEM drives sales from customers who are searching for specific products, social media marketing can generate interest and demand from new groups of consumers. Social media is now the primary or sole marketing channel for many businesses and organizations, such as crafts sellers on Pinterest, fashion brands on Instagram, and nonprofits on Facebook. Manysocial media platforms offer their own analytics tools, such as Facebook Insights or Twitter Analytics, and third-party options are available as well.

Analyzing data obtained through social media platforms can provide valuable insights for building business or customer relationships. For example, marketers can set up an account to automatically post information about new products or features as they come out, use an analytics tool to evaluate consumer sentiment through comments or reactions without manuallysifting throughthe data, andthenreworktheir social media marketing as necessary.

Optimizeemailmarketing

Though businesses can use email promotions to reach new audiences, email marketing is more often concerned with existing customers who have opted in to mailing lists or have already purchased products and services. Email provides a more direct gauge of consumer sentimentthanotherchannels,becauseexistingcustomersaremorelikelytorespondto

surveysorinteractwithadvertised material. Popular emailmarketingtools includeSalesforce Marketing Cloud, Mailjet, and Autopilot.

Enterprises can use analytics to optimize and personalize email marketing efforts. Analyzing how customers interact withdifferent emailpromotionscanhelp businessestarget their email marketingandtailor their messagestomeet customer expectationsandneeds. Enterprisescan use marketing analyticsto determine, for instance, whether customersrespond wellto certain keywords, emails sent at particular times ofday, or links to content on specific topics.

Takeadvantageofpredictive scoring

Predictive lead scoring models leverage marketing data fromall channels, as wellas internal data, to create a full picture of customer behavior, advertising potential, and marketing opportunities. These models use machine learning to build consumer profiles, which organizations can use to predict how consumers may react to different types of advertising and outreach. Campaigns can then target individual customers to maximize efficiency. For example, a predictive scoring system could rank individuals by likelihood of retention and risk of churning, which could help prioritize outreach to an organization's customer base.

❖ HR Analytics

HR Analytics brings “analysis” and “statistics” together to find the application of the data pool created byHR. Plainly enough, it is a data-driven approach to manage the employees.
ThepurposeofHRanalyticsexamplescanbedividedintotwosubgroups:

(i) ReachingBusinessGoals

This dimensionofHR analytics aims to provide the organization insight into the current state of operations. Such insights expedite the fulfilment of business goals according to the set timeframes.

(ii) Data-DrivenStrategyBuilding

It helps build prediction models to identifythe strategythat could lead to the optimum return on investment (ROI) for its human resources.
HRanalyticsmakesiteasyfortheHRprofessionaltocreatejob offersthatcan procure them thebest talentinthemarket,manageandretainthemtogiveaboost to ROI.

DataPointsforHRAnalytics?

The data is collected across multiple points in the organization to create a pool of employee data. HR analytics tools collect data like:

(i) Employeeperformance
(ii) Employeeattendance
(iii) Multi-rateror360-degreereviews
(iv) Salarydata
(v) Promotiondata

(vi) Employeeworkhistory
(vii) Demographic details
(viii) Employeetemperamentdata
HR analytics tools help in a close alignment of employee data and HR initiatives to direct them towards the achievement of the organization's goals.
Once theemployee data is gathered, the analysts feed the same into sophisticated data modelling programs, run them through the algorithms, and predictive tools to gain insights that can be acted upon.
The insights could be represented in the reports, dashboards or visualizations. TheHR analytics examples involve the below steps:
(i) Benchmarkanalysis
(ii) RelevantData gathering
(iii) DataCleansing
(iv) DataAnalysis
(v) Evaluationofgoals
(vi) Strategybuildingbased onanalysis
(vii) Planexecution

PracticalApplicationofHR Analytics
The HRanalyticsexamples areslowlyedgingtowardspracticalityanditfindsitsusein below-mentioned areas:
1. EmployeeRetention
AccordingtothestatsfromtheUSemployers,theaverageemployeereplacementcostis 200% of the annual salary they draw.
When an employeeleaves theorganization,the costof on boardingcostof the recruitment process and lost productivity adds to the total loss to the organization.
Itbecomescritical for theHRdepartment to containtheattritionrateand it isachievableonly when the HR department adapts to a data-driven approach.
Someoftheaspectsthatdriveimpactiterationanalysisare:
(i) ChurnRates
(ii) PerDepartmentAttritionRates
(iii) OnboardingExperience
(iv) EmployeeInterviewData

(v) EmployeePerformanceData

HR analytics helps identify the reasons behind attrition, and develop policies and training programs to dampen the impact of attrition.

2. EmployeePerformance

Around 45% of HR professionals believe that performance reviews are not accurate which makes it difficult to retain the talented employees.

Thus accurate performance evaluation is critical for retention. HR analytics tools are intelligent and leverage the employee data to identify the key players based on multiple performance parameters.

As the performance measurement and career progressionare interdependent, anaccuratedatamining of the two could help HR professionals identify the employee expectations.

For larger organizations, HR leaders can analyze the promotion expectation and average promotion wait time to determine howthe employees can be motivated to perform better and stay.

3. Employeerecruitment

The talent shortage is the biggest nightmare for enterprises. Almost 42% of the employers recruiting today are worried about the selection of not so appropriate candidate.

The HR recruitment team is primarily responsible for finding out the right CV's from thepool but before that, they need to develop a powerful ideal candidate portfolio.

This is wherewhat is HR analytics is answered and its role is identified. The data collected fromhiring managers and the performance data of previous hires for the same role are fed to the HR analytics tools to create an optimized skill set which is desired.

Thedatawhichisconsideredis:

(i) Identifytheaveragenumberoftheapplicantafter analyzingtheapplicantpool

(ii) Number ofinterviewrounds

(iii) Offeracceptancestatistics

4. EmployeeDevelopment

HRplaysacriticalinemployeedevelopmentasaskillgapalwaysexistswithnew recruitment.

According to American Employers, 40% of the recruited resource is not ideal for the job but with HR's employee development programs, the skill gap can be covered substantially.

TheHR Analytics tools help human resource management assess the skill needs, train the employeesaccordinglyandallocate theright resourcestotheteams. This increasesthe agility of the organization as well as enhances employee satisfaction.
Employeedevelopmentprogramsarerunningcurrentlyalsobutdiggingouttheright requirement isstilla challenge. In light ofHRanalytics, more refined employee development programs can be led.

5. EmployeeEngagement

Attracting the besttalent totheorganization is anart and HRstrive hardto achieve it. Having appropriate employee engagement is critical for an organization to attract and retain the employees.
As critical as it is to identify the factors that drive employee engagement, it is equallydifficult to find the right metrics.
TheHRneedstodoStatisticalanalysisofemployeeengagementdatasurveystoidentifythe datawhichleadstobetteremployee engagement.

6. DevelopingCompensation Programs

Employee compensation is knownto contribute 33%to employee retentionand performance. As it still remains the biggest investment by the business expense, its appropriation and accuracy needs go without saying.
Both internal and external factors impact the compensation plans that is why this arearequires more precise automation.
The HR needs to analyze what the competitors are offering their resources and what kind of compensation is inducing higher retention.

OverallBenefits ofHR Analytics

After thoroughlyanalyzing the role ofHR analytics it canbe deduced that it candirectly lead to the operational enhancement and more strategic hiring.
AstheorganizationsareadaptingHRanalyticstools, theROIis increasing. Let ustakealook at what are the changes when HR analytics examples are adopted.

(i) Decreasedattrition
(ii) Manualtaskautomation
(iii) HRProcessimprovement
(iv) Refinedhiringpractices
(v) Enhancementinemployeeproductivity

- **HRanalytics challenges**
  HRanalyticsrequiresalotofdatahandlingandthiscouldprovetopotentialchallenge enterprises need to overcome.
  ThepracticalimplementationofHRanalyticsisnoteasilyachievableandsomeofthe challenges that retard the adoption rates are:
  (i) AccesstotherightskillsetwhocanworkwiththeHRanalyticstools
  (ii) AggressiveData Cleansing
  (iii) Maintenanceofhigh-qualityDataquality
  (iv) Dataprivacy
  (v) Datacompliance
  (vi) Provingthevalueofreportsto the leadershipfor implementation
  (vii) Identifyingtheimpact ofstrategiesonROI

- **SupplyChainManagement**

  - Supply chain analytics refers to the processes organizations use to gain insight and extract value from the large amounts of data associated with the procurement, processinganddistribution of goods.Supply chain analyticsis an essential element of supply chain management (SCM).
  - Supply chains typically generate massive amounts of data. Supply chain analytics helps to make sense of all this data — uncovering patterns and generating insights.

  Differenttypesofsupplychain analyticsinclude:
  1. Descriptiveanalytics
  Provides visibility and a single source of truth across the supply chain, for both internal and external systems and data.

  2. Predictiveanalytics
  Helps an organization understand the most likely outcome or future scenario and its business implications. For example, by using predictive analytics, you can project and mitigate disruptions and risks.

  3. Prescriptiveanalytics
  Helps organizations solve problems and collaborate for maximum business value. Helps businesses collaborate with logistic partners to reduce time and effort in mitigating disruptions.

4. **Cognitiveanalytics**
   Helps an organization answer complex questions in natural language — in the way a person or team of people might respond to a question. It assists companies to think through a complex problem or issue, such as “How might we improve or optimize X?”

❖ Applying cognitive technologies

Supply chain analytics is also the foundation for applying cognitive technologies, such as artificial intelligence (AI), to the supply chain process. Cognitive technologies understand, reason, learn and interact like a human, but at enormous capacityand speed.

This advanced form of supply chain analytics is ushering in a new era of supply chain optimization. It can automatically sift through large amounts of data to help an organization improve forecasting, identify inefficiencies, respond better to customer needs, drive innovation and pursue breakthrough ideas.

Important supply chain analytics: Supply chain analytics can help an organization make smarter, quicker and more efficient decisions. Benefits include the ability to:

Reduce costs and improve margins: Access comprehensive data to gain a continuous integrated planning approach and real-time visibility into the disparate data that drives operational efficiency and actionable insights.

Betterunderstandrisks: Supplychainanalyticscanidentifyknownrisksandhelpto predict future risks by spotting patterns and trends throughout the supply chain.

Increaseaccuracyin planning: Byanalyzingcustomer data, supplychainanalyticscanhelp a business better predict future demand. It helps an organization decide what products can be minimized whentheybecome lessprofitable or understand what customer needswill be after the initial order.

Achieve the lean supply chain: Companies can use supply chain analytics to monitor warehouse, partner responses and customer needs for better-informed decisions.

Prepare for the future: Companies are now offering advanced analytics for supply chain management. Advanced analytics can process both structured and unstructured data, to give organizations an edge by making sure alerts arrive on time, so they can make optimal decisions. Advancedanalyticscanalso buildcorrelationandpatternsamongdifferent sources to provide alerts that minimize risks at little costs and less sustainability impact.

Astechnologiessuchas AI become more commonplace in supply chainanalytics, companies maysee anexplosionoffurther benefits. Informationnot previouslyprocessed because ofthe limitations of analyzing natural language data can now be analyzed in real time. AI can rapidlyand comprehensivelyread, understand and correlate data fromdisparate sources,silos and systems.

It can then provide real-time analysis based on interpretation of the data. Companies willhave far broader supply chain intelligence. They can become more efficient and avoid disruptions — while supporting new business models.

Keyfeaturesofeffectivesupplychain analytics

The supplychain is the most obvious face of the business for customers and consumers. The better a company can perform supply chain analytics, the better it protects its business reputation and long-term sustainability.

Keyfeaturesofeffectivesupplychain optimization include:

Connected: Being able to access unstructured data from social media, structured data from the Internet of Things (IoT) and more traditional data sets available through traditional ERP and B2B integration tools.

Collaborative: Improving collaboration with suppliers increasingly means the use of cloud-based commerce networks to enable multi-enterprise collaboration and engagement.

Cyber-aware: The supply chain must harden its systems from cyber-intrusions and hacks, which should be an enterprise-wide concern.

Cognitively enabled: The AI platformbecomes the modern supplychain's controltower by collating, coordinating and conducting decisions and actions across the chain. Most of the supply chain is automated and self-learning.

Comprehensive: Analyticscapabilities must be scaled withdata inrealtime. Insightswillbe comprehensive and fast. Latency is unacceptable in the supply chain of the future.

❖ **RetailAnalytics**

Retail analytics is the process of tracking business data, such as inventory levels, consumer behavior, sales numbers, and more, to make more informed, strategic decisions. Thisincludes providing insights to understand and optimize the retail business's supply chain, consumer behavior, sales trends, operational processes, and overall performance. With today's high customer expectations for retail, companies must meet those rising needs with personalized omni channel offers, efficient processes, and quick adjustments to upcoming trends—all of which require retail analytics.

Retailanalyticstranslatereal-worldbusinessactivityintoquantifiabledatatodrive betterbusinessdecision-making.Thisdatacancoverconsumerbehaviorpatterns, supplychaininformation,inventoryupdates,andmore.Retailerscanoperationalize theseinsightsinnumerousways,including:

- Optimizingstorelayoutanddesign
- Iteratingonproductdisplays
- Improvingpricingstrategy
- Enhancingpromotionalcampaigns
- Buildingcomprehensivecustomerpersonas
- Drivingpersonalizedproductrecommendations

Whilethesourcesofthisdataareendless,notallanalyticsarecreatedequal.Retail analytics derived from real-time consumer data collected via mobile app usage are particularly high-yield. Retailanalyticsrevolutionizesbusinessdecision-making infourmainsteps:

Datagathering

A team of experts collect data from all retailers, in whichever form they hold it. This is entered into an analytics platform in an intuitive dashboard, saving time and resources compared with internal sales teams doing it manually.

Datacleansingandvalidation

Retailanalytics tools – and a teamofexperts – comb through the data, removing false entries and making sure all remaining data is accurate and consistent.

Dataanalyzing

The analytics platform puts clean data into dashboards, tables and graphs – making it effortless for businesses to analyze performance and informcampaigns. This data canalso be entered into data warehouse tools to inform stock and inventory decisions.

Driveadoption

Retailanalyticstools – including automated reports, intuitive dashboardsand convenient data files – make it simple for everyone across the supply chain to access and learn from the business' data.

5high-valueretailanalyticsuse

Behavioural analytics

Properly optimizing the in-store experience requires an in-depth understanding of how consumersmovethrough theretailspace,andcustomerflowanalyticsprovides exactly theseinsights.Providedthrough techplatformslikeVera,retailerscandiscoverhow traffic density varies across floor space, identify points of interest, and spot shoppers' navigationpatterns.Inthecaseof Vera,thisdatacanbecollectedandanalyzedinreal-time,ensuringretailershavethemostcurrentinformationpossible.

Behaviouralanalyticslikethesenableretailerstomakerelevant,data-baseddecisionsonin-storeexperiencedesign.Insteadofrelyingonbestpractices,retailerscantailor floorlayouts tomatchtheparticularbehaviouralpatternsoftheirconsumers.

Essentially,thistechnologybringsalevelofexperientialoptimizationonceonly possible on the web to real-world commercial environments, significantly enhancing the efficacy of commercial spaces.

Recommendationengines

For years, online retailers have leveraged the power of analytics to offer their customers personalizedproductrecommendations.Thisinnovationhasbeenintegraltothesuccess ofecommercegiantslikeAmazon.But,emergingtechnologieslikeaugmentedreality (AR) are helping level the playing field for brick and mortar businesses by providing anaturalplatformfordeployingthisanalysistodrivepurchasingbehaviour.

By analyzing purchasing behaviour and predicting future demand, retailers can deliver personalizedproductrecommendationsthroughAR-enhancedin-storeshopping experiences.Likeonlineshopping,thesecustomnotificationscanbepushedto consumersaccordingtotheirproximityandbehaviourduringthenaturalpointsin browsing.Personalizedrecommendationscandramaticallyincreasecustomer

engagement,asresearchbyAccenturehasrevealedthat 65%ofconsumerspreferretailers that know their shopping history. Global retailer Auchan deploys a simplifiedversionofthisapproachtoboostfoottrafficattheirlocations.

Inventoryoptimization

The goal of inventory management is to optimize the relationship between supply and demand.Whilethisprocesswasoncethedomainofeducatedguesses,managerscan nowleverage a deep setof data and analytical tools tomake stocking decisions.For example,America'slargestgrocer—Kroger—hasbeen usinganin-houseanalyticsteamtoanalyzeeverythingfromeconomictrendstoshopperbehaviortoaccurately forecast demand for years.

One of the most exciting innovations in this area is the use of real-time customer flow statistics to anticipateinventory needs.Insteadof relying on relevanthistorical data to projectdemand,retailerscanalsopulldatafromacrosstheentireretailenterpriseto makestockingdecisions.Whilethisinventorymanagementmodelispreferablein almostanycircumstance,it'sevenmoreessentialduringperiodsofunprecedented consumerbehaviorlikehasbeenseenthroughoutthepandemic.Havingcomprehensive andcurrentdataempowersinventorymanagerstoidentifytrendsandrespond appropriately.

Predictivepricing

Priceisoneofthemostpowerfulleversincommerce,andtheincreasingavailabilityof dataandanalyticaltoolsgivesretailersanevenbettergraspofit.Now,retailerscanuse a multitude of inputs to drive their pricing strategies. These include everything from fundamentalslikethecostofgoodssoldandcompetitorpricingtoadvancedanalytics likeweatherforecastsandreal-timecustomerbehavioraldata.Withthisdata,retailers canuseanalyticstopredicttheidealsaleduration,identifycustomerpricetolerance, anddetermineothercriticalelementsoftheirpricingstrategy.

Smartmerchandising

Frompromotionstodisplayoptimization,merchandisingistheartandscienceof sellingretailgoods,andfewretailanalyticsusecaseshaveamoredirectimpacton business performance. Access to store-level data and customer behavior analytics — combined with machine learning-driven analysis — are transforming how retailers are running their in-store campaigns.

ThegrowinguseofARinshoppingexperiencesaddsanotherlayerofactionabledatato this trove, allowing retailers to getfeedback on their display strategies faster than everbefore. ARexperiencesserveasatestingground forproductplacements,displays, signage,andotherpromotionalcollateral.Theseapplicationscanthenrelayreal-time data on the effectiveness of these campaigns, allowing retailers to quickly and cost-effectivelyiterateoncampaignconceptsbeforecommittingsignificantresources.

Howtoimplement retailanalytics

Informationispower—alwayshasbeenandalwayswill be.While21st-century innovations have largely favoured online retailers, emerging technologies like AR have changedthegameforbrickandmortarbusinesses.Now,theytoocanhaveaccessto real-timedataabouttheirconsumerstooptimizetheirstorelayouts,testcampaigns, improveinventory decisions,andmore.

Forforward-thinkingretailers,thisisall excellentnews.If you'reinterestedinlearning more about some of the most exciting retail analytics use cases, then reach out. Our AR platformispoweringthenextgeneration ofin-storedata,andwe'dlovetohelpyou leverage it to grow your business.

❖ **Salesanalytics**

- Sales analytics refers to the technology and processes used to gather sales data and gaugesalesperformance.Salesleadersusethesemetricsto set goals, improve internal processes, and forecast future sales and revenue more accurately.
- The goal of sales analytics is always to simplify the information available to you. It should help you clearly understand your team's performance, sales trends, and opportunities.

Generally,salesanalyticsisdivided intofourcategories:

Descriptive: What happened? Descriptive analytics entails tracking historical sales data—revenue, number of users, etc.—so you can make comparisons and better understand what's currently happening.

Diagnostic: Why did it happen?Diagnostic analytics is examining and drilling down into the data to determine exactly why something occurred.

Predictive: What's going to happen?Predictive analytics is taking what you've learned about past sales and using it to gauge patterns and trends. This allows you to make educated predictions.

Prescriptive: What's the best solution or action? Prescriptive analytics involves assessing all the data and recommending the best plan of action.

Sales analysis should be a priority if you want your business to stand out in a highly competitive world, especially during decision-making scenarios.

Hereareseveral**benefitsofsalesanalyticsforbusinesses:**

1. PromotesBetterDecision-making

Access to data from sales analytics provides a company with the capability to make accurate decisions that can be beneficial in the long run. Companies can utilize sales analytics when they share the discussion with the workforce.

Working collaborativelyallowsbetteranalysis anddecisionsforthebenefitofall, especially when deciding on a marketing or sales strategy to implement.

2. HelpsAchievetheMissionStatement

- With quantified values, it'llpromote the growthof the business, especially the analytical process, since it defines a common objective that the workforce follows. Once these values are quantified, they'llundergo evaluation by employees to better understand their expectations of them. With well-informed employees, they're likely to be more productive.

3. KeepsYourBusiness Updated

Today, consumers readily change their minds as fads come and go. Mostget easily swayed byseeminglygoodoffers. Withsales analytics, it canprovide a companyinsight onthe latest flow in the target market.

Always remember that the fluctuations in the industry occur rapidly. In recent years, you might see big-scale companies succumb to promising startups. Make it a priority to protect your business from the unpredictable nature of the industry with sales analytics. With the data, you can make the right move by innovating according to the current needs and preferences of the consumers.

4. Boosts Efficiency

The availabilityofbusiness analytics has made it possible for businesses to improve in terms of efficiency. Since analytics can rapidly gather large amounts of data and present it appealingly, companies can decide on suitable plans to reach their objectives.

Remember that analytics can encourage a company culture that values efficiency and teamwork. It creates an environment where the workforce can readily express insights and share in the decision-making process.

Additionally, analytics allows businesses to develop better choices, such as the direction to take or figuring out the necessary steps to reach new objectives.

5. ProvidesBetterInsightsviaData Visualization

In recent years, the versions of sales analytics are easy to decipher and even presentable. Businesses can check out highly comprehensive graphs and charts to aid with the decision-making process.

With the visual representations of the data, any business will gain beneficial insights more straightforwardly. The presentation of the data visualization is organized and visually appealing.

6. Better Accessibility

With the help of sales analytics software, businesses can readily access data and produce accurate reports. Depending on the sales analytics software your company uses, it generally allows quick access on any device with Internet connectivity.

The sales team can easily access the progress at any time. The convenient access provides better efficiency and flexibility, which are both crucial to a fast-paced business environment. When the sales analytics software you'reusing has a quick and easy reportinginterface,it can simplify the decision-making process, eventually resulting in better sales.

7. TransparencyofSalesData

The sales data can provide businesses with complete transparency and serves as a tool for mentoringthesalesteam.Withthedata,companiescanutilizethenecessarytactics, such as a cross-sell campaign.

The business has a clear view of the sales team'sprogress, including strengths, priorities, challenges, andpossibleweakpoints. Inthelongrun, it directlyaffectssalessince it provides you with control over variables to boost efficiency and productivity levels.

8. HelpsPinpointProfitableandSlow-MovingProducts

The data from sales analysis will allow a business to pinpoint both profitable and slow-moving products and services.

Depending on the business you're managing, you can make the appropriate modifications to the sales and marketing tactics basing onthe data and distributing resources for products that show the highest chances of future growth.

Aside from identifying the profitable assets, the data can also help you identify slow-moving productsorservices. It'llallowyoutoallocateresourcesefficiently,cutholding expenses, and prevent over-stocking. The data from the analytics will serve as your basis for implementing changes in the prices or providing discounts.

Conclusion

Sales analytics is an indispensable tool for businesses all over the globe. Without this must-have element, your businesswon't last long in a highly competitive industry. Depending on the company you're managing, finding the right sales analytics software is crucial. With the benefits that sales analytics provides, making the most outofthetoolwillkeep your business running efficiently and maintain superior productivity for years to come.

❖ Web&SocialMediaAnalytics

Web Analytics, by definition "Web analytics is the measurement of data, the collection of information, analysis, and reporting of Internet data for the purposes of optimizing and understanding Web usage.

Web analytics uses the data collected directly from a particular business website and Social media analytics uses the data collected from social media networks.
Web analytics Gathers information from the Business website only. In general, Web Analytics tells you about your traffic levels, referralsources, bounce rate, and user behaviour on your website.
Web Analytics mainly used to improve the user experience and conversion rate. Below are the few important analysis can be done through web analytics,

- Wherearеyourwebsite visitorscoming from?
- Onwhichpageofyourwebsite theyare spendingmore time?
- Howaretheyconnectingtoyourwebsite?
- Whichpartofthedayisyour websitegettingmoretraffic?
- Howwelldoyouretainusers?

Inother words,thefour importantkeymetricscanbeanalyzedfromwebanalytics,

- 1.TotalTraffic
- 2.TrafficSource
- 3.BounceRate
- 4.Conversionrate

TotalTraffic
Total Traffic to your website gives insights about where are you getting more traffic, which helps you to understand your target market. In addition, you can analyze which hours of the day and days of the week, you are getting more visits to your websites. Based on this information you can run a campaign to optimize more conversions.

TrafficSource
Traffic source is about how you are getting most visitors to your website. Is it through social media, search engines or Referral Sites? Based on that information you can strategize your marketing campaign or write a blog or focusing on a particular social media network. For example, if most of your visitors are coming from social media networks, use thatinformation to brand your business more on Facebook, Twitter or any other social media platforms to boost your website traffic.

BounceRate

Bounce rate is the percentage ofvisitors to a particular website who are leaving the site after viewing only one page without navigating other pages on the website.

Thiscouldbehigherforanynumberofreasons maybe,

- Irrelevantcontent

- Inappropriatedesigns
- Confusingnavigation
- Frequentpop-ups
- Unnecessaryads
- Or,Annoyingsounds

However,this metrichelpsyoutoimproveyourwebdesignoverall.

ConversionRate

A conversion rate is the percentage of visitors who have taken some action on your websiteor complete the desired goal; it could be purchasing a product, Sign up for newsletters, etc.

❖ **SocialMediaAnalytics**

Social media analytics is the practice of gathering data from social media websites or networks such as Facebook, Twitter, Google plus, etc., and analyzing those metrics to understand insights to make business decisions.

Social media analytics gathers information from social media networking sites and helps businesses better understand customer sentiment, users' attitudes, build rich consumer profiles, and, most importantly, build effective business strategies.

Therearemanytoolsavailable inthe markettotrackandanalyzesocialmediadata. The most common use of social media analytics is to discover

- Sentiments
- Opinions
- Emotions
- Topics

Puttogether knownasSentimentAnalysis.

❖ **SentimentAnalysis**

- Itistheprocessofcomputationallyidentifyingand categorizingopinionsexpressed in a piece oftext, especially in order to determine whether the writer's attitudetowards a particular topic, product, etc. is positive, negative, or neutral. Opinion mining or sentiment analysis refers to the use of natural language processing.
- Importantly, the first step is to define which business goal we are trying to address through social media analytics.
- In general, the business objectives include increasing revenues, reducing customer service costs, getting feedback on products and services, and improving publicopinion of a particular product or business division.
- Aquick recap ofinformationgathering

❖ **Importanceofsocialmediaanalytics**

There is a tremendous amount of information in social media data. In decades past,enterprises paid market research companies to poll consumers and conduct focus groups toget the kind of information that consumers now willingly post to public social media platforms.

In the past few years, businesses have rushed to use Web and Social Media widely. Almost 94% of allbusinesses now use some formof social media to promotetheir brand and engage with customers. While companies sprint to master web and social media marketing, the analysis of data available in web and social media more of a struggle.
The problem is this information is in the form of free text and natural language, the kind of unstructured data that analytics algorithms have traditionally. However, as machine learning and artificial intelligence have advanced, it has become easier for businesses to quantify in a scalable way the information in social media posts.

This allows enterprises to extract information about how the public perceives their brand, what kind of products consumers like and dislike and generally, where markets are going. Social media analytics makes it possible for businesses to quantify allthis without using less reliable polling and focus groups.

❖ **HealthCareAnalytics**

Healthcare analytics refers to the use of data that offers comprehensive insight into patients andtheconditionsthat affectorhavethepotentialtoaffect them.Analyticsin healthcarecan also provide insight into the effectiveness of healthcare providers themselves in terms of productivity, cost and general performance. Essentially, it is all about gathering and leveraging information to improve quality and efficiency of healthcare services.

Thebenefitsofdataanalyticsinhealthcare

The advantages that advanced analytics in healthcare can bring are vast and well-recognised bythose working within the industry. Let's dive into some of the specific ways that data and analytics can support healthcare.

- **Predict risks** - As the saying goes, prevention is betterthan cure. Bycollecting mass data, hospitals can identify common symptoms and causes ofconditions and diseases. This helps doctors spot when a patient may be at risk of developing a certain health problem and treat them as early as possible.
- **Make data-driven decisions** – With more data at their disposal in regard to patient medical history and the health of the wider population, healthcare professionals can make informed decisions about individual treatment and how likely it is to be successful.
- **Increase patient satisfaction** – Insights from data can help doctors to personalise treatment and improve how they care for patients. Software can even assess the performance of healthcare professionals and provide feedback.
- **Improve service delivery** – Hospitals can use software to predict busier times and appropriately plan to meet demand, for example by having more staff on rota. This can help reduce long waiting times and shortages of beds.
- **Electronic record-keeping** - Storing medicalrecords electronically, as opposedtoon paper, also improves productivity by mitigating the problem of having multiple records. It also enables different healthcare professionals to access the same records without transporting paperwork between facilities.
    - **Reducecosts**–Withdataimprovingpatientcareandallowinghealthcarefacilitiesto runmoreproductively,treatmentcostsandotherhospitalexpensescanbeminimised.

Healthcareanalyticsimportance

- Despite the pandemic advancing how technology is used in healthcare, the industry is slowto adaptoverall. For example, it hasbeenreportedthat just over halfofhospitals

have no strategies for data governance or analytics in their day to day practices, and 97% of data produced by hospitals is wasted.

- The Health Foundation similarly discovered that, although the NHS generates masses of data, they lack staff with the right analytical experience to interpret this data. Consequently, opportunities to improve services – such as by improving diagnoses and day-to-day care – are being missed. Similarly, leaders in the industry noted that the pandemic exposed many flaws within many health care systems with poor quality of data, time-consuming analytic processes, and staff lacking the training to use data properly being a common occurrence.
- With much of the industry failing to innovate when it comes to technology, it is important to both understand the benefits that analytics can bring and consider how it can be incorporated into your organisation. After all, technology is only going to develop further, and the industry will need to innovate in order to be ready for future challenges. As industry professionals state, "digital health solutions and technology will play a crucial role in the difficult work of optimizing processes and systems for greater efficiency, financial viability, and enhanced outcomes."

❖ **EnergyAnalytics**

Energy analytics generally describes the process of collecting electrical data and applying sophisticated analytical software and algorithms to deliver insights around consumption and time of use reductions.

Businessbenefitfromenergyanalytics

1. Valuableinsightsintoyourenergydata

Energy analytics can provide you with unique insights into your business energy data that would've been impossible to find manually or with Excel.
For example, energy analytics software can show you the periods where you spend the most on energy. It can also help you understand which areas of your organisation are inefficient and how energy consumption is affected by external factors such as the weather.

2. Improvedenergyefficiencyandreducedenergycosts

Ifyou payattentiontothe insights delivered byenergyanalytics and take appropriate actions, you canquicklyreduce your energycosts. Before you know it,the software you're using will be paying for itself.

As an example, energy data analytics could highlight that one of your buildings uses a surprising amount of energy in non-operational hours. Further investigation could uncover that the heating controls for this building are faulty. Taking action and fixing this problem will likely save thousands over the course of a year.
Researchsuggest that most companiescansaveat least 10%ontheirenergybills withenergy data analytics.

3. Itcanstreamlineyourjobasanenergy manager

Energy managers are usually responsible for more than just electricity use. Chances are you'realso heldaccountable forgasandwateruse,maybe evengenerationifyou've installed renewable energy solutions onsite.
Energy analytics software can connect with all of this data and import it automatically. This means you no longer have to log into four or five different systems to build a completepicture of your organisation's utility use.

4. Automationoftime-consumingactivities

- Energymanagement is a role that sometimes involves doing the same things over and over again. This can include creating and sending out reports, conducting degree day analysis or analysing the success of a project.
- Energy analytics software can help with this by automating these tasks, allowing you to spend more time planning and running energy-saving projects and improving the energy efficiency of your business.

5. Itsimplifiesdatasharingand collaboration

- Asbusinesses have begun to realise the financial and environmental benefitsassociated with energy efficiency, there's been an increase in the number of energy managers employed by companies.
- Someorganisations areevenhiringteamsofenergymanagersto improvetheir energy efficiency.
- When working in a team, it's vitally important that data can be shared and communicated quickly and clearly.

- In other words, energy analytics software helps ensure that essential energy conversationsdon't get lost ina seaofemails. It also meansthat there'sonededicated place for people in your organisation to see, analyse and discuss energy data.

❖ **TransportationAnalytics**

Transportation data analytics increasingly power mobility information and insights– transforming transportation planning by making it easier, faster, cheaper, and safer to collect and understand critical information.
While the transportation industry may not be in crisis, it is certainly being heavily disrupted by multiple forces, including the COVID-19 pandemic. As these changes unfold, transportation experts must:

- Prioritizeprojectsaccuratelytoguideeffectiveresourceinvestmentandmakethe biggest impact.
- Make informed decisions based onrecent, accurate data, not onguesses or input from a few vocal stakeholders.
- Maintainsocialequity andenvironmentaljustice,providingaccessandsupportfor outlying areas and the underserved.
- Fosterpublicengagement,sothatresidents,constituents,andpublicofficials understand, can respond to questions about, and support planned mobility efforts.
- Accuratelyandquicklymeasureresultsoftransportationinitiatives,enablingadjustment and optimization in real time.

Increasing numbers of cities, transit organizations, departments of transportation, and other localities areusingtransportationdataanalyticstosolveproblems, prioritize investments, and win stakeholder support.

❖ **ImplementationofMultimodalTransportSegment-wise Analysis**

With the growth of multi-modal transport, the need for segment-wise analysis is essential. The primary modes of transportation include roadways, railways, waterways, and airways.

Letusexploretheimplementationofanalysis ineachsegment:

Roadways

Using analytics for one of the most used modes of transportation, roadways, has several benefits:

RoadSafetyManagement

Advanced data can be used to analyze where, why, and when accidents happen. With this data,theycancreate**Prognostication Crash Maps**(shown inthe image) that analyze data to shortlist high-risk areas. These maps can help issue warnings to be extra careful at these locations and help authorities take precautionary measurements.

RoadTrafficManagement

Keepingarecordofautomobiles'movingpatterns, velocity, and lanechanging behaviourcan help us understand how different road designs can influence driving. The insights are useful for smartertraffic controland identifying congestion in the road layout when planning future infrastructural developments. The graph depicts the same.

RailTrafficManagement

There is a whole range of possibilities that railways can explore in big data analytics. Applications in the railway industry include booking, improving security, automatic scheduling and planning, network enhancement and ticket management. The existing data fromthepassengeroperatingcontrol, reservationsystem, CCTV,and maintenancedepotscan be used to our advantage to yield business benefits in the above areas. Real-time train information system (RTIS), the Nation Train Enquiry System (NTES), and the control office application (COA) are some examples where data analytics is used.

AirTraffic Management

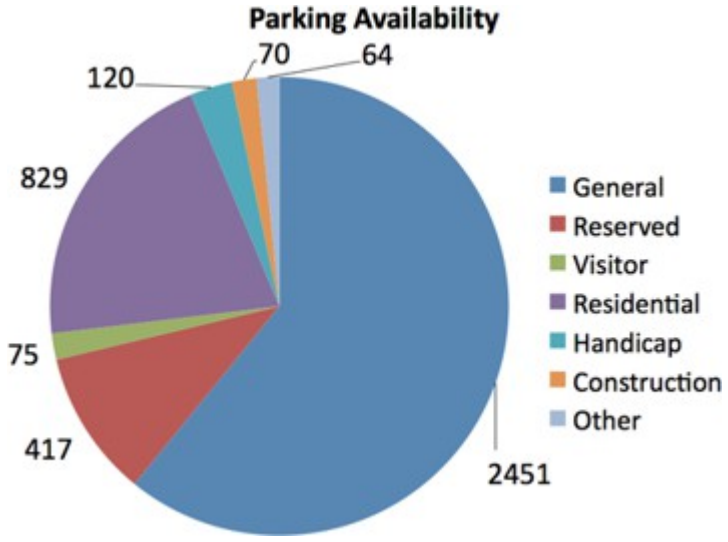
Long queues are a top annoyance of air travellers. However, by accessing data of those travellers coming through the facility, advanced analytics can help airport workers easily visualizethebusiestperiodsfortheirsecuritycheckpoints.Overtime,machinelearning-

powered by AI can generate predictive models that can allow the airport to strategize betterand allocate resources.

Waterways
ShipMonitoring andRouteOptimization
Ship monitoring is one of the most critical factors for seamless planning and execution. Various tools such as vessel's sensors, weather station reports, and satellite reports will increaseships'efficiency.Theentire dataarraycanbeprocessedthroughmachine learning, and the following questions can be answered using the same:
• Whendoesthehullneedcleaningtosavefuel?
• Whenshouldtheshipequipmentbechanged?
• Whichisthebestrouteintermsofweather,safety,andwhichrouteisfuelsustainable?

❖ **AnalyticsinRegularDay-to-dayTransportation**
Ease Traffic Congestion:
Agencies can help ease traffic congestion by using a high occupancy toll (HOT) with real-timeanalytics. Basedonthetraffic, theycandynamicallyadjust pricesandopenHOT lanes at a much higher cost, reducing traffic.

IncreasedefficiencyisfindingParking Slots:
Inmajorcities, peoplewho aretryingto findparkingslotscause15%-30%oftraffic. New technologies like cameras, sensors, and geo-tracking and analytics can help drivers find parking spots.



Analytics can help the transportation industry, especially the multi-modal transportation system, to be sustainable and efficient. However, withrapid advancements intechnologyand data flowing inand out, various factors suchas privacy, regulations, and confidentialitymust be taken care of to use data 100% effectively to provide fruitful analysis.

❖ **Lending Analytics**

Financialinstitutions like banks have beenusingpredictivecustomeranalysis fora longtime. But as the complexity of loans increased, so did the need for more complex and accurate analysis.

Most of the past loan frauds are correlated to the prediction models that were being used for decades, which were not effective to detect bad loan potentials. But, as the days passed by, the financial risk factors started to plummet and it became almost impossible to approve bad loans. Today, thanks to data analytics, the default and fraud probability of loans have decreased significantly.

1. **CustomerSelection**

Customer selection is the fundamental part of any loan proceeding. While the dataverification methods are still present, it's now possible to analytically predict the quality of applications withthe help ofdataanalytics. This kind ofanalysis is deemed betteras it leaves no-to-minimal room for errors.

Data analytics take account ofthe credit card purchases, subscriptions, and loyaltycards then categorize them into financial profiles. These are used to approve loans that have a better probability of being repaid.

Acustomer selection model usually concludes if a customer will be inclined to paythe EMIs regularly and is safe to grant loans to. But, typically, the financial verifications are done by business experts with the help of data analytics to understand their financial behavior, spending pattern, and repayment credibility's for safer loan disbursement.

2. **DesigningCustom Offers**

Most of the loan offers are customized to the needs of individuals. If the applicant'screditscoreis poor, they might have to paya higher interest rate with collateral. Ifthey have faulty loan histories, they might not be able to secure a higher principal.

In contrast, customers, who were efficient at paying previous loans are entitled to higher principal and lower interest rates.

Although these elements were always present in loan processing, with data analytics and the availability of past data, the applicants can get to know the best offers within seconds. Immediate feedback helps customers make an informed decision quickly, in addition to ensuring the experts that no bad loans are being distributed.

Data analytics are also being used by financial institutions to customize the promotionaloffers that a specific demographic is likely to avail. Small interest rates, longer repayment terms, and no-cost EMIs targeted to reach certain customers are such examples.

3. DelinquencyDetection

It was almost impossible to predict the financial behavior of a lender with older prediction models. Acustomer, appearing to bethe perfect candidate could pose asa bad loanwiththeir erratic payments after the loan was approved.

As the problem grew to the extent of jeopardizing the business of the lenders, delinquency prediction models came into play. With extensive data of past loans, records of transactions, late payments, partialpayments, and failed payments, the models are now able to predict the risky loans before they are approved.

Delinquency detection not only helps the banks but the borrowers too. It's possible that an individual somehow missed payments and has shown unusual financial behavior in the past, but is now trying to rework the errors. With ample data available on them, they can check their financial scores – which are usually reversible – to understand what went wrong and decide how it can be improved.

4. StrategizingCollection

Even after probability models are employed, some bad loans pass through the check. When that happens, the only way to recover the principal is by collection methods. In the past, customers were categorized by risk factors. And different contact strategies were used to extract the amount. Which resulted in failures more often than not.

But after data analytic models were introduced to the banking sector, even though bad loans still went through, the applications could be segmented into micro categories depending on demographics, financial activity, and risk ratings to employ more effective intervention techniques.

According to ScoreData, the customers are divided into three categories in collection analytics.

1. Theoneswhohavedefaultedforthefirsttime
2. Thelazyones
3. Self-curecustomers
4. Baddebt
5. Thosewhoarebeyondanyredemption

The first-time defaulters are often the ones who are the safest of the customers but can be flagged as frauds if the account is new. Hence the collection efforts are marked as low.

The lazy customers forget to pay the bills but are the safest bet among the lenders. They usually pay after the period is over with late fines and interest. The collection efforts are marked as low.

Self-cure customers are occasional defaulters and usually are safe bets. Collection efforts for these customers are medium

Bad debt and point of no return customers are the ones who have a comparatively higher spending rate than their earnings but usually pay the minimum fees. Collection Efforts are considered High or non-viable for the borrowers of this category.

FraudDetection

Credit card frauds, loan frauds, and deliberate payment delays have always been an issue for lenders. With data analytics and better customer selection, the problem had been handled quite effectively. Let's look at some of the ways data analytics help to detect fraudulent activities of a profile.

DataAnalysis

Data analysis models provide an all-around view of alarming actions and suspicious transactions of the customers. It isolates attributes and identifies hidden threats to notify the banks about them.

Analytical Frameworks

By calculating statistical parameters like standard deviation and moving averages, analytical frameworks identify fraud patterns in customers. Excessive high or low numbers are also taken into consideration while screening fraudulent activities.

CrossChannelMonitoring

As the financial sectors became more centralized, all the transaction data of customers are available to the models. Which data analytics programs can monitor and analyze to prevent loan fraud.

EasyLoan Processing

The data analytical models were not designed to keep good loans at bay. Moreover, the data analytics capabilities now make the loan proceedings easier than ever if you have a good credit score.

As the financial institutions run their business on loans and interests, the faster and safer proceeding time helps them get the most benefit out of the customers as well.

TheBottomLine

Bad loans will always be there. What would the lenders do if a business goes bankrupt even after showing promise? But, it's now becoming possible to detect patterns of customers who are likely to fail loans and categorize them with the help of data analytics.

Delinquency detection, offer creation, fraud detection, and loan collection are also thebenefits of data analytics in loan processing.

- **SportsAnalytics**

  Sports analytics is the process of plugging statistics into a mathematical model to predictthe outcome of a given play or game. Coaches rely on analytics to scout opponents andoptimize play calls in game, while front offices use it to prioritize player development.Analytics also play a major role off the field, providing fans with both sports betting andfantasy sports insights.

sports analytics is the practice of applying mathematical and statistical principles to sportsand related peripheral activities. While there are many factors and priorities specific to the industry, sports analysts use the same basic methods and approach as any other kind of data analyst. Establishing parameters for measurement, like hit or fumble rate, and consistently collecting data from a broad sample is the basis of the analytics process. This data is then curated and optimized to improve the accuracy and usability of the results.

Sports analytics goes beyond traditional statistics to add accurate analysis to improve many factors in team performance.

On-FieldApplications

Analytics has many on-field applications in a sports environment, including managing both individual and group performance. Coaches can use data to optimize exercise programs for their players and develop nutrition plans to maximize fitness. Analytics is also commonly used in developing tactics and team strategies. With thousands of games worth of data to study, analysts can look for patterns across a broad sample size regarding formation, counter strategies and other key variables.

UsesinTeamManagement

Practical data analysis has plentyofapplications forthe business side ofsports as well. Since most professional sports teams function as businesses, they are always seeking ways to improve sales and reduce expenses across their organization. Some sports analystsspecifically focus on issues regarding the marketing and sale of sports tickets and team merchandise. Modern marketing and fan outreach efforts also rely heavily on analytics to predict their consumer base and identify opportunities to increase brand engagement.

Sport	Statistic	Definition
Baseball	Batting average	ratio of hits and number of at bats
	On-base percentage	times a player reaches base by hitting, walking, or by being hit by a pitch
	Slugging average	the ratio of number of bases earned and the number of at bats
	WHIP	Walks plus Hits allowed per Inning Pitched measure the number of baserunners the pitcher allows on hits and walks
Football	GAP ratings	Generalized Attacking Performance, introduced by Wheatcroft (2020), is a rating system for assessing the attacking and defensive capability of a team with respect to number of corners or shots in football.
	BA ratings	Bivariate attacking ratings is based on the parameter that minimises the mean absolute error (MAE) between estimated and observed match statistic.
Cricket: (Criclytics)	WASP	Winning and Scoring Prediction is a machine learning technique to predict final score of the first innings and probability of winning while chasing in second innings
	Win Probability Statistic	is a model that calculates the team's win probability in real-time using the team's historical data

FutureofBusinessAnalytics

The future of business analytics is bright. That's partly because people generate a massive amount of digital data every day. In 2020, people consumed a whopping 64.2 zettabytes of data, according to Statista. To put that number in perspective, just 1 zettabyte provides enoughstorage for30billion4K movies, 60billionvideo games, or7.5trillionsongs inMP3 format. This number is poised for growth, too:It's projected that the data consumed globally will reach approximately 181 zettabytes in 2025.

It's therefore not surprising that students earning a degree in business analytics will enter a marketwithampleopportunity.AccordingtotheU.S.BureauofLaborStatistics,the

employment of market research analysts is growing much faster than average — by a projected 22% between 2020 and 2030. The job outlook is also above average for related positions, such as management analyst (14%) and operations research analyst (25%). These numbers demonstrate the need for qualified professionals to manage all that digital information and convert it to meet business goals.

BusinessAnalyticsTrends

Technologyisanever-evolving process. Naturally,thisevolutionproducesbusinessanalytics trends that professionals must know about to help companies leverage raw data to reachtheir goals. These pros must understand the following trends in data and business intelligence to optimize opportunities for their companies.

Search-BasedDiscoveryTools

Using raw data to answer specific questions or track trends isn't a foreign concept to many. After all, people do it routinely when they turn to tools such as Google and other search engines to find something. Search queries are second nature to many individuals, but not every company has user-friendly data discovery technology for this purpose.

These tools make it possible to sift through and find actionable data from disparate sources more efficiently. This can boost the capacity to find key insights for effective business strategies that may otherwise be lost in a sea of information. As the number of potential sources increases, the ability to cut through superfluous information and get to what matters becomes more essential.

Bringing the power of search-based discovery tools to the internal operations of enterprises isn't the only area where knowledge of business analytics trends shines. Online marketing isa key initiative for all types of businesses, and an understanding of the data mechanics of Google and other search engines is critical to its success. The future of business analytics in marketing is now. Time spent on-page, social shares, where page visits originate from, and other data can inform marketing decisions and drive relevant traffic.

AI andMachineLearning

Artificial intelligence (AI) and machine learning facilitate high efficiency at substantially reduced costs.Expertspredict machine learning willcomplete agrowing number ofcustomer service tasks in the future, but the effectiveness of these technologies depends heavilyon the people behind them. Even machines with the capacity to learn must be fed the right information, and business analysts are often the keepers of these solutions.

AI and machine learningoffer widespreadapplications for businesses. Theonlinepublication Business 2 Community points out that the concepts provide the backbonefor elements that are crucial for business effectiveness and efficiencies, such as personalized marketing, cybersecurity, talent recruitment, and customer relationship management. While these technology-drivenelements are poised to impact the business environment,theystillneed the human touch of business analytics professionals to ensure that their functions translate to success.

CloudComputing

Cloud computing, the process of using remote servers on the internet to store and manage data, provides many of the benefits businesses demand. These benefits include reduced reliance on physical resources that are often outdated as soon as they're installed; increased efficiencies, especiallywhenworking withteams indisparate locations;and plentyofoptions

for redundancymanagement anddisaster planning. Whenutilizedcorrectly, cloudcomputing can increase operational efficiency and reduce costs.

However, taking advantage of these benefits isn't always easy. Companies need to develop viable workflows to pool resources, share information efficiently but securely, oversee data access to protect consumers, and maintain the speed and ease with which data can be accessed. Business analysts manage data flow within cloud-based structures, help design and develop data processes, and analyze both the data and the performance of systems overall to ensure that business goals are supported.

PredictiveAnalyticsTools

Predictive analytics will likelyplaya large role in the future of business analytics. Predictive analytics in businessoftencomesdownto the need to anticipate the customer'sorthe client's next move. By analyzing historical data patterns in consumer behavior, market fluctuations, and even societal trends, businesses can prepare for certain outcomes and performances with increased confidence. This could not only keep them consistently relevant in their industrybut also transform them into industry leaders.

Because the concept's success depends on the future, the information from the past must be properly interpreted for upcoming plans. If not, negative ramifications might follow. The work of a highly skilled business analytics professional can be critical for keeping thepurpose behind predictive analytics beneficialand not something that can inadvertentlycause business strategies to hit unexpected snags.

DataAutomation

With the amount of data reaching the zettabytes, sorting, storing, and managing it can be an increasingly challenging and time-consuming process. This is what makes data automation such an important component for the future of business. Data automation can automatically take care of the mundane yet critical parts of data management, leaving those in business analyticsroles more time to analyze and interpretgathered findings. It canalso be a vitalpart of helping businesses overcome any scalability issues they may have.

The key to effectively using data automation is building an effective strategy that properly curates the automated system. This means the data is stored in a way that makes it easy to accessanduseattheappropriatetime. Businessanalyticsprofessionals canbe instrumentalin this key process, as they can use their knowledge and skills to build strategies that integrate sensibility and efficiency into the data automation process.

UNITV
Ethical, LegalandOrganizationalIssues

Issues&Challenges-BusinessAnalyticsImplementationChallenges– PrivacyandAnonymization- Hacking and Insider Threats- Making Customer Comfortable

The era of technology has given rise to a sea wave of change of how data is used. Business Analytics is an emerging field in data science that aims at using data to develop business insights that can be beneficial to the organization. You'll see implementation of BA invarious areas. Let's take an example: Credit rating companies analyze the credit card transactions of its consumers and can predict the spending patternof the consumer as well as his financialhealth. This informationcanbeusefulto companiesto findtheirtarget audience. Another example can be a mobile company which extracts data about the customer's frequencyofcalls, rechargeamounts, etc.This informationcanbeclassifiedand later usedby the companyas per itsobjective. Infact, BhartiAirtel, has successfullyimplemented BAas a part of their Analytics department, and it has led to positive outcomes.

BenefitsofimplementingBA inyourorganization

Apart from having applications in various arenas, following are the benefits of Business Analytics and its impact on business –

- Accuratelytransferringinformation
- Consequentimprovementinefficiency
- HelpportrayFutureChallenges
- MakeStrategicdecisions
- Asaperfectblendofdatascienceandanalytics
- ReductioninCosts
- ImprovedDecisions
- Shareinformationwithalargeraudience
- EaseinSharinginformationwithstakeholders

Moreover, anytechnologyissubject to itsownsetofproblemsand challenges. Following are the common challenges in implementing business analytics in an organization.

- Lack oftechnicalskillsinemployees
- Fussoveracceptance ofBAbystaff
- DataSecurityandMaintenance
- IntegrityofData
- Deliveringrelevantinformationinthegiventime
- Inabilityto addresscomplex issues
- Costsinvolvedinimplementing BA
- Investmentofstafftimein implementationof BA
- Lackofa properstrategyto implementBA

ThemajorchallengesinBusinessAnalyticsareas follows:

Increasein numberofSources

Whendatasetsbecome biggerand morecomplex, bringingtheminto ananalytics framework poses a huge challenge in business analytics. If this is ignored, it creates gaps that result in incorrect communications and observations.

ShortageofTalentforData Analytics
Data processing is essential in order to render usable this voluminous volume of data that is generated every minute. The immense demand for big data scientists and big data analystshas been generated on the market with the rapid growth in data. It is essential for business organizations, because the role of a data scientist is multidisciplinary, to employ a data scientist with qualifications that are versatile.

DataSecrecy
Gaining valuable insights fromBig Data analytics is crucial for business enterprises and it is therefore criticalthat this intelligence is obtained onlythroughthe appropriate team. A major challenge inbusinessanalyticsposed bytheBigDataanalytics firms isto successfullybridge this large void.

HandlingLargeVolumesof Data
It's not shocking that for everydaythat passes data is through. It clearly means that business entities need to manage a great regular amount of info. The volume and quality of data available these days will overpower every computer engineer and this is why it is deemed important for brand owners and managers to make data accessibilityquick and convenient.

ChangingtechnologicalLandscape
Each day new technology and businesses are being built with the growth of Big Data. Nonetheless, a major challenge in business analytics posed by big data analytics firms is to figure out which technologies can better match them without new challenges and future threats being added.

Qualityofstorageandretrievingdata
Enterprise companies are increasing rapidly. With the exponential development of the companies and major business organizations, the volume ofdata generated increases. Storing a large volume of data is now a huge challenge in business analytics for us. Popular data storage solutions such as data lake/warehouses are widely used to capture and preserve vast volumes of unstructured, organized data in their format.

Considering the above challenges, there is a dearth of professionals who're well-equipped withtheknowledgeofBusinessAnalytics. ABAprofessionalcantakeup acorporateroles in various sectors: marketing, insurance, management, finance, health care & lifestyle, etc. In fact, there is still hesitation to use trends and statistics for making business decisions, and most of themstill are comfortable trusting their gut feeling for making strategic decisions.

- **DataPrivacyandAnonymization**
  Dataprivacy,sometimesalsoreferredtoasinformationprivacy,isanareaof dataprotectionthat concerns the proper handling of sensitive data including, notably, *personal data* , but also other confidential data, such as certain financial data and intellectual property data, to meet regulatory requirements as well as protecting the confidentiality and immutability of the data.

  Roughly speaking, data protection spans three broad categories, namely, traditional data protection (such as backup and restore copies), data security, and data privacy as shown inthe Figure below. Ensuring the privacy of sensitive and personal data can be considered an outcome of best practice in data protection and securitywith theoverall goalofachieving the continual availability and immutability of critical business data.

Please note that the term data privacy contains what the European Union (EU) refers to as "data protection."

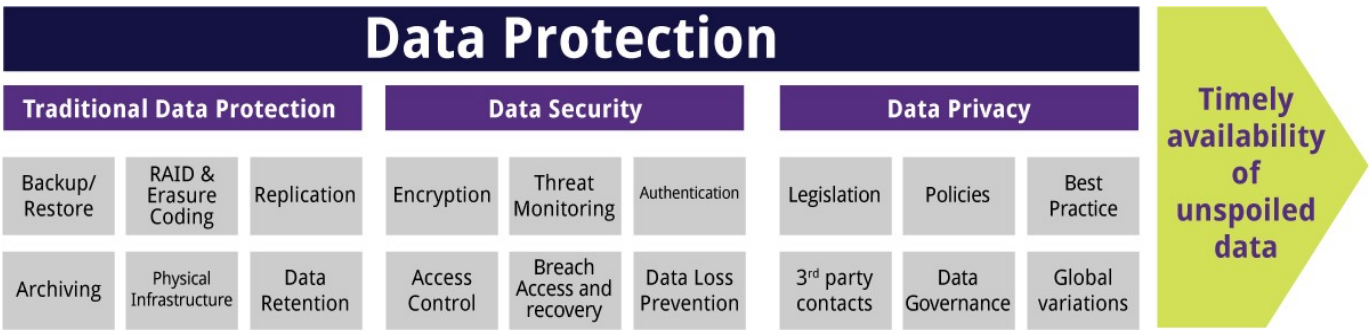


Figure:TheThreeCategoriesofData Protection

Security becomes an important element in protecting the data from external and internal threatsbut also whendetermining what digitallystoreddatacanbe shared and withwhom. In a practical sense, data privacy deals with aspects of the control process around sharing data withthird parties, howand where that datais stored, andthe specific regulations that applyto those processes.

Almostallcountriesintheworldhaveintroducedsomeformoflegislationconcerning data privacy in response to the needs of a particular industryor section of the population.

DataSovereignty

Data sovereignty refers to digital data that is subject to the laws of the country in which it is located.

The increasing adoption ofcloud data services and a perceived lack ofsecurityhas led many countries to introduce new legislation that requires data to be kept within the country inwhich the customer resides.

Current concerns surrounding data sovereignty are related to governments trying to prevent datafrombeingstoredoutsidethegeographic boundariesoftheoriginatingcountry. Ensuring that data exists only in the host country can becomplex and often relies on the detailprovided in the Service Level Agreementwith the Cloud Service Provider.

DataPrivacy -Geographicalvariationsinterms

In the European Union, privacy is recognised as an absolute fundamental right and in some parts of the world privacy has often been regarded as an elementof liberty, the right tobe free from intrusions by the state. In most geographies, privacy is a legal concept and not a technology, and so it is the term data protection that deals with the technical framework of keeping the data secure and available.

WhyisDataPrivacyimportant?

Theanswer to thisquestioncomesdowntobusinessimperatives:

1. **Business Asset Management:** Data is perhaps the most important asset a business owns. We live ina dataeconomywhere companiesfind enormousvalue incollecting, sharing and using data about customers or users, especially from social media. Transparency in how businesses request consent to keep*personal data,*abide bytheir privacy policies, and manage the data that they've collected, is vital to building trust with customers who naturally expect privacy as a human right.
2. **Regulatory Compliance:** Managing datato ensure regulatorycompliance isarguably even more important. A business may have to meet legal responsibilities about how they collect, store, and process personal data, and non-compliance could lead to ahuge fine. If the business becomes the victim to a hack or ransom ware, the consequences in terms of lost revenue and lost customer trust could be even worse.

Data PrivacyisnotDataSecurity

Businesses are sometimes confused by the terms and mistakenly believe that keepingpersonal and sensitive data secure from hackers means that they are automatically compliant with data privacy regulations. This is not the case. Data security protects data from compromise by external attackers and malicious insiders whereas data privacy governs how the data is collected, shared and used.

DifferinglegaldefinitionsofDataPrivacy

Ifthere is agreement onthe importance ofdataprivacyto abusiness, thenthe legaldefinition can be extremely complex.

What is meant by data privacy - it is left to businesses to determine what they consider best practice in their own industry. The legislation often refers to what is considered 'reasonable' which may differ between laws, along with the respective fines.

In practice, this means that companies who work with sensitive and personal data should consider exceeding the legal parameters to ensure that their data practices are well above those outlined in the legislation.

- **DataAnonymization**

Data anonymization is the process ofprotecting private or sensitive information byerasing or encryptingidentifiersthatconnectanindividualtostoreddata.Forexample,youcanrunPersonally Identifiable Information (PII) such as names, social security numbers, and addresses through a data anonymization process that retains the data but keeps the source anonymous.

However, even when you clear data of identifiers, attackers can use de-anonymization methods to retrace the data anonymization process. Since data usually passes throughmultiple sources—some available to the public—de-anonymization techniques can cross- reference the sources and reveal personal information.

TheGeneral Data Protection Regulation (GDPR) outlines a specific set of rules that protect userdata andcreate transparency.WhiletheGDPRis strict,itpermits companies tocollect

anonymized data without consent, use it for anypurpose, and store it for an indefinite time—as long as companies remove all identifiers from the data.

DataAnonymizationTechniques

- **Data masking**—hiding data with altered values. You can create a mirror version of a database and applymodificationtechniquessuchascharacter shuffling, encryption, and word or character substitution. For example, you can replace a value character with a symbol such as "*" or "x". Data masking makes reverse engineering or detection impossible.
- **Pseudonymization**—a data management and de-identification method that replaces private identifiers with fake identifiers or pseudonyms, for example replacing the identifier "JohnSmith"with"MarkSpencer". Pseudonymizationpreserves statisticalaccuracyanddata integrity, allowing the modified data to be used for training, development, testing, and analytics while protecting data privacy.
- **Generalization**—deliberately removes some of the data to make it less identifiable. Data can be modified into a set of ranges or a broad area with appropriate boundaries. You can remove the house number in an address, but make sure you don't remove the road name. The purpose is to eliminate some of the identifiers while retaining a measure of dataaccuracy.
- **Data swapping**—also known as shuffling and permutation, a technique used to rearrange the dataset attribute values so they don't correspond with the original records. Swapping attributes (columns) that contain identifiers values such as date of birth, for example, may have more impact on anonymization than membership type values.
- **Data perturbation**—modifies the original dataset slightly by applying techniquesthat round numbers and add random noise. The range of values needs to be in proportion to theperturbation. Asmall base mayleadto weakanonymizationwhilea large basecanreduce the utilityofthe dataset. Forexample, you canuse a base of5 for rounding values like age or house number because it's proportional to the original value. You can multiply a house number by15 and the value mayretain its credence. However, using higher bases like 15 can make the age values seem fake.
- **Synthetic data**—algorithmically manufactured information that has no connection to real events. Synthetic data is used to create artificial datasets instead of altering the original dataset or using it as is and risking privacy and security. The process involves creating statistical models based on patterns found in the original dataset. You can use standard deviations, medians, linear regression or other statistical techniques to generate the synthetic data.

❖ **DisadvantagesofDataAnonymization**

The GDPR stipulates that websites must obtain consent from users to collect personal information such as IP addresses, device ID, and cookies. Collecting anonymous data and deleting identifiers fromthe database limit your abilityto derive value and insight from your data. For example, anonymised data cannot be used for marketing efforts, or to personalize the user experience.

❖ **HackingandInsiderThreats**

Acommonlyused hacking definition is the actofcompromising digitaldevices and networks through unauthorized access to an account or computer system. Hacking is not always a malicious act, but it is most commonlyassociatedwith illegalactivityand datatheft bycyber criminals.

❖ **Hacking**

Hackingreferstothe misuseofdevices likecomputers, smartphones, tablets, and networksto cause damage to or corrupt systems, gather information on users, steal data and documents,or disrupt data-related activity.

Atraditional view ofhackers is a lone rogueprogrammer who is highlyskilled incoding and modifying computersoftware and hardware systems. Butthisnarrow view doesnot coverthe true technical nature of hacking. Hackers are increasingly growing in sophistication, using stealthy attack methods designed to go completely unnoticed by cybersecurity software and IT teams. They are also highly skilled in creating attack vectorsthat trick users into opening malicious attachments or links and freely giving up their sensitive personal data.

As a result, modern-dayhacking involves far more than just anangrykid intheir bedroom. It isamultibillion-dollarindustrywithextremelysophisticatedandsuccessfultechniques.

HistoryofHacking/Hackers

Hacking first appeared as a term in the 1970s but became more popular through the next decade. An article in a 1980 edition of Psychology Today ran the headline “The Hacker Papers” in an exploration of computer usage's addictive nature. Two years later, two movies, Tron and WarGames, were released, in which the lead characters set about hacking into computer systems, which introduced the concept of hacking to a wide audience and as a potential national security risk.

Sure enough, later that year, a group of teenagers cracked the computer systems of major organizations like Los Alamos National Laboratory, Security Pacific Bank, and Sloan-Kettering Cancer Center. A Newsweek article covering the event became the first to use the word “hacker” in the negative light it now holds.

This event also led Congress to pass several bills around computer crimes, but that did not stop the number of high-profile attacks on corporate and government systems. Of course, the concept of hacking has spiraled with the release of the public internet, which has led to far more opportunities and more lucrative rewards for hacking activity. This saw techniques evolve and increase in sophistication and gave birth to a wide range of types of hacking and hackers.

TypesofHacking/Hackers

Therearetypicallyfourkeydriversthatlead to badactorshacking websitesorsystems:
(1)financialgainthroughthetheftofcreditcarddetailsorbydefraudingfinancialservices,
(2) corporate espionage, (3) to gain notoriety or respect for their hacking talents, and (4)state-sponsored hacking that aims to steal business information and national intelligence. On top of that, there are politically motivated hackers—or hacktivists—who aim to raise public attention by leaking sensitive information, such as Anonymous, LulzSec, and WikiLeaks.

Afewofthemostcommontypesofhackersthatcarryouttheseactivitiesinvolve:

BlackHatHackers

Black hat hackersare the "bad guys" of the hacking scene. They go out of their way to discover vulnerabilities in computer systems and software to exploit them for financial gainor for more malicious purposes, such as to gain reputation, carry out corporate espionage, or as part of a nation-state hacking campaign.

These individuals' actions can inflict serious damage on both computer users and the organizations they work for. They can steal sensitive personal information, compromise computer and financial systems, and alter or take down the functionality of websites and critical networks.

WhiteHat Hackers

White hat hackers can be seen as the "good guys" who attempt to prevent the success ofblack hat hackers throughproactive hacking. They use their technical skills to break into systems to assess and test the level of network security, also known as ethical hacking. This helps expose vulnerabilities in systems before black hat hackers can detect and exploit them.

The techniques white hat hackers use are similar to or even identical to those of black hat hackers, but these individuals are hired by organizations to test and discover potential holesin their security defenses.

GreyHatHackers

Grey hat hackers sit somewhere between the good and the bad guys. Unlike black hathackers, theyattempt to violate standards and principles but without intending to do harmor gain financially. Their actions are typically carried out for the common good. For example, theymayexploit avulnerabilitytoraiseawarenessthat it exists, but unlikewhitehat hackers, they do so publicly. This alerts malicious actors to the existence of the vulnerability.

Other common hacker types include blue hat hackers, which are amateur hackers who carry out malicious acts like revenge attacks, red hat hackers, who search for black hat hackers to prevent their attacks, and green hat hackers, who want to learn about and observe hacking techniques on hacking forums.

Other common hacker types are cyber terrorists, hacktivists, state- or nation-sponsored hackers, script kiddies, malicious insiders, and elite hackers.

DevicesMostVulnerableTo Hacking

SmartDevices

Smart devices, such as smartphones, are lucrative targets for hackers. Android devices, in particular, have a more open-source and inconsistent software development process than Apple devices, which puts them at risk of data theft or corruption. However, hackers are increasinglytargetingthemillionsofdevicesconnectedtotheInternetofThings(IoT).

Webcams

Webcamsbuilt into computersare acommon hacking target,mainlybecause hacking themis a simple process. Hackers typically gain access to a computer using a Remote Access Trojan (RAT) in rootkit malware, which allows them to not only spy on users but also read their messages,seetheirbrowsingactivity,takescreenshots,andhijacktheirwebcam.

Routers

Hacking routers enables an attacker to gain access to data sent and received across them and networks that are accessed on them. Hackers can also hijack a router to carry out wider malicious acts such as distributed denial-of-service (DDoS) attacks, Domain Name System (DNS) spoofing, or cryptomining.

Email

Email is one of the most common targets ofcyber-attacks. It is used to spread malware and ransomware and as a tactic for phishing attacks, whichenable attackersto target victims with malicious attachments or links.

JailbrokenPhones

Jail breaking a phone means removing restrictions imposed on its operating systemto enable the user to install applications or other software not available through its official app store. Aside from being a violation of the end-user's license agreement with the phone developer, jailbreaking exposes many vulnerabilities. Hackers can target jailbroken phones, whichallowsthemto stealanydataonthe device but also extendtheir attackto connected networks and systems.

PreventionfromGettingHacked

There are several key steps and best practices that organizations and users can follow to ensure they limit their chances of getting hacked.

SoftwareUpdate

Hackers are constantly on the lookout for vulnerabilities or holes in security that have not been seen or patched. Therefore, updating software and operating systems are both crucialto preventing users and organizations fromgetting hacked. Theymust enable automatic updates and ensure the latestsoftware version is alwaysinstalled on all of theirdevices and programs.

UseUniquePasswordsforDifferent Accounts

Weak passwords or account credentials and poor password practices are the most common cause of data breaches and cyberattacks. It is vital to not only use strong passwords that are difficult for hackers to crack but also to never use the same password for different accounts. Usinguniquepasswordsiscrucialtolimitinghackers'effectiveness.

HTTPSEncryption

Spoofed websites are another common vehicle for data theft, when hackers create a scam website that looks legitimate but will actually steal the credentials that users enter. It is important to look for the Hypertext Transfer ProtocolSecure (HTTPS) prefix at the start of a web address. For example:https://www.fortinet.com.

AvoidClicking onAdsorStrangeLinks

Advertisements like pop-up ads are also widelyused by hackers. Whenclicked, theylead the usertoinadvertentlydownload malwareorspywareontotheirdevice.Linksshould betreated carefully, and strange links within email messages or on social media, in particular, should never beclicked. Thesecan beused by hackersto install malwareonadeviceor lead usersto spoofed websites.

ChangetheDefaultUsernameandPasswordonYourRouterandSmart Devices

Routers and smart devices come with default usernames and passwords. However, as providers ship millions of devices, there is a risk that the credentials are not unique, which heightens the chances of hackers breaking into them. It is best practice to set a unique username and password combination for these types of devices.

❖ MeasuresagainstHacking

Therearefurtherstepsthat usersandorganizations cantaketo protectthemselvesagainstthe threat of hacking.

DownloadfromFirst-partySources

Only download applications or software from trusted organizations and first-party sources. Downloading content from unknown sources means users do not fully know what they are accessing,andthesoftwarecanbeinfectedwithmalware,viruses,orTrojans.

InstallAntivirusSoftware

Havingantivirus softwareinstalled on devices is crucial to spotting potential malicious files, activity, and bad actors. A trusted antivirus tool protects users and organizations from the latest malware, spyware, and viruses and uses advanced detection engines to block and prevent new and evolving threats.

UseaVPN

Using a virtual private network(VPN) allows users to browse the internet securely. It hides theirlocationandpreventshackersfrominterceptingtheirdataorbrowsingactivity.

Do NotLoginas anAdminbyDefault

"Admin" is one of the most commonly used usernames by IT departments, and hackers use this information to target organizations. Signing in with this name makes you a hackingtarget, so donot loginwithit bydefault.

UseaPassword Manager

Creating strong, unique passwords is a security best practice, but remembering them is difficult. Password managers are useful tools for helping people use strong, hard-to-crack passwordswithouthavingtoworryaboutrememberingthem.

UseTwo-factorAuthentication

Two-factor authentication (2FA) removes people's reliance on passwords and provides more certainty that the person accessing an accountis who they say they are. When a userlogsin to their account,theyare thenprompted to provide another piece ofidentityevidence, suchas their fingerprint or a code sent to their device.

BrushUponAnti-phishingTechniques

Users must understand the techniques that hackers deploy to target them. This is especially the case withant phishing and ransom ware, which help users know the tell-tale signs of a phishing email or a ransomware attack or ransomware settlements.

What isEthicalHacking?HowLegalisEthicalHacking?

Ethical hacking refers to the actions carried out by white hat security hackers. It involves gainingaccesstocomputersystemsand networkstotestforpotentialvulnerabilities,andthen fixing any identified weaknesses. Using these technical skills for ethical hacking purposes is legal, provided the individual has written permission from the system or network owner, protects the organization's privacy, and reports all weaknesses they find to the organization and its vendors.

Examples

The biggest hack in history is thought to be the data breach against Yahoo! The 2013 attackcompromised around 3 billionpeople, and the companyrevealed that everyYahoo! customerwas affected by it.

China is believed to be the country with the highest number of dangerous hackers. Most ofthe major cyber attacks that occurred around the world can be traced back to China.

InsiderThreat

An insider threat is a security risk that originates from within the targeted organization. It typically involves a current or former employee or business associate who has access to sensitive information or privileged accounts within the network of an organization, and who misuses this access.

Traditional security measures tend to focus on external threats and are not always capable of identifying an internal threat emanating from inside the organization.

Typesofinsiderthreatsinclude:

- **Malicious insider**—also known as a Turncloak, someone who maliciously and intentionally abuses legitimate credentials, typically to steal information for financial or personal incentives. For example, an individual who holds a grudge against a former employer, or an opportunistic employee who sells secret information to a competitor. Turncloaks have anadvantageoverother attackers because theyare familiar withthe security policies and procedures of an organization, as well as its vulnerabilities.
- **Carelessinsider**—an innocent pawnwho unknowinglyexposesthe systemto outside threats. This is the most common type of insider threat, resulting from mistakes, such as leaving a device exposed or falling victim to a scam. For example, an employee who intends no harm may click on an insecure link, infecting the system with malware.
- **A mole**—an imposter who is technically an outsider but has managed to gain insider access to a privileged network. This is someone from outside the organization who poses as an employee or partner.

Threetypesofriskybehavior explained

MaliciousInsiderThreat Indicators

Anomalous activity at the network level could indicate an inside threat. Likewise, if an employeeappearstobedissatisfiedorholdsagrudge,orifanemployeestartstotakeon

more tasks with excessive enthusiasm, this could be an indication of foul play. Trackable insider threat indicators include:

- **Activity atunusualtimes**—signingintothenetworkat3am
- **Thevolumeoftraffic**—transferring toomuchdataviathenetwork
- **Thetypeofactivity**—accessingunusualresources

HowtoProtect Against anInsiderAttack: BestPractices

Youcantakethefollowing stepstohelp reducetherisk ofinsiderthreats:

- **Protect critical assets**—these can be physical or logical, including systems, technology, facilities, and people. Intellectual property, including customer data for vendors, proprietary software, schematics, and internal manufacturing processes, are also critical assets. Form a comprehensive understanding of your critical assets. Ask questions such as: What critical assets do we possess? Can we prioritize our assets? And, what do we understand about the current state of each asset?
- **Enforce policies**—clearly document organizational policies so you can enforce them and prevent misunderstandings. Everyone in the organization should be familiar with security procedures and should understand their rights in relation to intellectual property (IP) so they don't share privileged content that they have created.
- **Increase visibility**—deploy solutions to keep track ofemployee actions and correlate information from multiple data sources. For example, you can use deception technologytolurea maliciousinsiderorimposterandgainvisibilityintotheiractions.
- **Promote culture changes**—ensuring security is not only about know-how but also about attitudesand beliefs.To combat negligence andaddressthedriversofmalicious behavior, you should educate your employees regarding security issues and work to improve employee satisfaction.

- **MakingCustomerComfortable**

Delivering excellent customer care and proactive customer supportmake clients feel valued. It is all about knowing what your customers' expectations are and offering the best of your ability.

1. Providereal timesupport

Businesses can align your customer's expectations with reality bydeploying new technology to provide real time support to your customers. Great customer experience can be achievedbyusing live chat software and live engagement tools that boost customer satisfaction rates.

Livechat

Live chatis the most preferred channel over other communication channels such as phoneand email. The real time support it delivers to customers makes it popular. **"79% ofcustomerssay they prefer to live chat because of the immediacy it offers compared to other channels."**

- Livechatinstantlyconnectswithyour customersandassiststheminrealtime.
- You can trigger proactive chat messages to guide customers in their buying journey and improve their experience.

Engagementtools

In real timeby usinglive customer engagementsuch as co-browsing,video& voice chat. These tools allow customers to communicate within minimal wait time and delight your customers.

- Video chat allows us to **identify the issue faster and deliver effective solutions**, which reduces the number of touch points and increases customer satisfaction.
- Co-browsing solutionallows you to **collaborate with customers and guide them to complete a complex form fill up or application process.**
- Having**direct personalized conversationsdevelopstrust in customers** and delivers a delightful customer support experience. Streamline all the customer conversations under one platform and provide a cohesive experience.
- Identify the most preferred channels and be 24×7 active across those channels to reduce average response time.
- Makeuseoftools like livechat, chatbots,visualtoolstogainfasterdetailsofthe issue and deliver first contact resolution.

2. Deliverconsistent omni-channelcustomerservice

- Streamlineall thecustomer conversations under one platform and provide a cohesive experience.
- Identifythemostpreferredchannelsandbe24×7activeacrossthosechannelsto reduce average response time.
- Makeuseoftools like livechat, chatbots,visualtoolstogainfasterdetailsofthe issue and deliver first contact resolution.

Customerdelightexample –BankofAmerica
Bank of Americais one of the biggest known brands following consistent omni-channel servicetoitscustomers. Thebankallows foreverything fromdepositingchecksto scheduling an appointment to be handled by the company's mobile and desktop apps.

Empoweryourteam todelightyourcustomers

- **Freedomofdecision making** – The employees hold the complete authorityto handle customer's queries independently. It is their responsibility to amaze them by meeting and exceeding customer expectations.
- **Perform as a team**– Empowering your teamallows themto performtogether totake a moveto deliver superior customer servicethat surpassesthecustomer delight index.
- **Employees feedback**–The feedback from employees are directly aligned with the company's objective. The mission of the company and the opinion of your team are linked that makes them valued.

Analyzecustomerfeedback

"Asatisfiedcustomeristhebest businessstrategyofall."

Customer feedback is crucial for the growth of all businesses. It provides valuable insights into what is working well about your products or services and what should be done to make the experience better.
Analyzing feedback involves**identifying customer needs and frustrations of customers so that businesses can improve customer satisfaction and reduce churn**. Feedback analysis needs to be done wisely by following the below steps.

- **Categorize all feedback into categories** – The feedback may include product delivery speed, after-sales services, customer service approach, etc. Once categorized you can divide further that deserves immediate attention.
- **Identify the nature of the feedback –**Customer feedback comes with negative and positive comments. The positive ones bring in concrete ideas on what can be extremelyeffective inbuilding customer loyalty. Onthe other hand,the negative ones provide insights on improvement areas.
- **Consolidate results and plan your next move** –Finally, amalgamate all the results to make a plan of action as to how you intend to respond to each of the issues raised. Make a feasible and effective plan that would address all the problems your clients think your business is having while keeping the good services still functioning.

Note: You need to train your customer support teamto ask feedback at the right time via the right channels and by tailoring feedback questionnaires.

Personalizeyourcommunication

- Use a tone that matches your customer personality. Some prefer short and direct communications and some like longer conversations and sharing opinions.
- Listen to your customers and empower them. Asking your customers about their preferences both personalizes the experience and builds their confidence in your brand.
- Understanding your customers' backgrounds by having authentic conversations helps to personalize every relationship.
- Make kind gestures by sending follow-up messages to customers after each purchase or service interaction to thank them and offer further assistance if required.
- Recommend products and services to your customers based ontheir purchase history. Personalized cart recommendations influence 92% of shoppers online.

Empoweryourcustomers with communities

Customers love to be part of a community or group. Building communities that benefit customers create a positive feeling and improve your brand image. Communities can be used as a resource for sharing useful information related to products and services.
Branded communities are 13% more likely to have an impact on customer experiencethan social media communities.
When you foster a special place for customers to interact with one another, your business is adding extra value to the customer experience both before and after the purchase. Generally, customers trust other customers, when they intend to purchase.
So,thesponsoredcommunityforumcanbeusedtohelptoguideintheirbuyingdecisions.

7. Donot underpromiseand over deliver

Don't makeapromiseyoucan'tkeepandkeeptheones youmake.

Over delivering on customer expectations would raise customer satisfaction and be good for business. It developstrust and loyaltyincustomers and staysassociated with your brand for a lifetime.
To keep customers highly satisfied, you must continue to deliver more value because their expectations will keep increasing.
Note: If you set the tone from the start of under-promising and over-delivering, then your customer is going to expect that same experience of getting more than promised with every interaction.You are setting yourself up to fail and for your customer to be disappointed. A better method might be to deliver on your promises.

8. Listenactively todelightyourcustomers

- Understand your customer needs, expectations, and painpointsand align your service to match accordingly to impress them.
- Welcome your customer complaintsand feedback and adopt the right tools/process to deliver a better experience.

Actively listening to customers allows you to use the right empathy statements for customerserviceand deliver a delightful experience.

Stop tryingtodelightyourcustomers

"Towin customerloyalty,forget thebellsandwhistlesandjust fixtheir problems".

The role ofmotivational customer servicecan never be discounted as businesses create loyal customers mainly by resolving their issues faster. Customers resent having to contact customer support repeatedly, to get anissue fixed, having to repeatthe complete information, and switching from one channel to another.

Finalthoughtsoncustomerdelight

If you want to be on the right side of the customer service road then get together with your team today and plan different ways on how to impress and delight customers. Following the right strategies will help to deliver a positive customer experience.