UNIT-I
SUBJECT: BA4206/BUSINESS ANALYTICS

2 MARKS:

1. What is meant by business analytics?

Business analytics (BA) is a set of disciplines and technologies for solving business problems using data analysis, statistical models and other quantitative methods. It involves an iterative, methodical exploration of an organization's data, with an emphasis on statistical analysis, to drive decision-making.

2. Define data warehouse.

A **Data Warehousing** (DW) is process for collecting and managing data from varied sources to provide meaningful business insights. A Data warehouse is typically used to connect and analyze business data from heterogeneous sources. The data warehouse is the core of the BI system which is built for data analysis and reporting.

3. What is business intelligence?

BI (Business Intelligence) is a set of processes, architectures, and technologies that convert raw data into meaningful information that drives profitable business actions. It is a suite of software and services to transform data into actionable intelligence and knowledge.BI has a direct impact on organization's strategic, tactical and operational business decisions.

4. What do you mean by dashboards and scorecard?

Dashboards are used daily in organizations as they offer a more operational view of success than scorecards' focus on strategic goals. Data available in dashboards is used to provide a foundation for better decision making and more efficient day-to-day management of teams, resources, and expenses.

Scorecards offer organizations a snapshot of their current performance when compared to their goals. They are useful tools for organizations which need to manage performance and make strategic decisions better based on the distance between current performance and the goal.

5. What is meant by Hadoop?

Hadoop is an open-source framework from Apache and is used to store process and analyze data which are very huge in volume. Hadoop is written in Java and is not OLAP (online analytical processing). It is used for batch/offline processing. It is being used by Facebook, Yahoo, Google, Twitter, LinkedIn and many more. Moreover, it can be scaled up just by adding nodes in the cluster.

6. Define data mining.

Data mining is the process of sorting through large data sets to identify patterns and relationships that can help solve business problems through data analysis. Data mining techniques and tools enable enterprises to predict future trends and make more-informed business decisions.

7. What is text mining?

Text mining, also known as text data mining, is the process of transforming unstructured text into a structured format to identify meaningful patterns and new insights. By applying advanced analytical techniques, such as Naïve Bayes, Support Vector Machines (SVM), and other deep learning algorithms, companies are able to explore and discover hidden relationships within their unstructured data.

8. What is meant by data visualization?

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. Additionally, it provides an excellent way for employees or business owners to present data to non-technical audiences without confusion.

9. Define information repository.

An information repository is a collection of interrelated information maintained across a network on multiple servers. It creates a unified resource for anyone connected with the system to access when they need information. Numerous organizations use information repositories to handle their data and may network with others to share material as necessary.

10. What is data cleansing?

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled. If data is incorrect, outcomes and algorithms are unreliable, even though they may look correct.

11. What are the steps involved in business analytics?

```
1. Define the Business Need
2. Explore the Data
3. Analyse the Data
4. Predict what's likely to happen
5. Optimise – Find the best solution
6. Make a Decision & Measure the Outcome
7. Update the System with the Results of the 'Decision'
```

12. What is big data?

Big Data is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Big data is also a data but with huge size.

13. What is artificial intelligence?

Artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems. Specific applications of AI include expert systems, natural language processing, speech recognition and machine vision. Artificial intelligence is a constellation of many different technologies working together to enable machines to sense, comprehend, act, and learn with human-like levels of intelligence.

14. What is contextual data?

Contextual data is the background information that provides a broader understanding of an event, person, or item. This data is used for framing what you know in a larger picture. These relevant facts can be utilized to analyze your customers' behavior patterns, thereby improving their experience.

15. What are geospatial analytics?

Geospatial analytics gathers, manipulates and displays geographic information system (GIS) data and imagery including GPS and satellite photographs. Geospatial data analytics rely on geographic coordinates and specific identifiers such as street address and zip code. They are used to create geographic models and data visualizations for more accurate modeling and predictions of trends.

16. What do you mean by key performance indicator?

Key performance indicators (KPIs) refer to a set of quantifiable measurements used to gauge a company's overall long-term performance. KPIs specifically help determine a company's strategic, financial, and operational achievements, especially compared to those of other businesses within the same sector.

17. What is variety, velocity, volume?

Volume

It refers to the size of Big Data. Data can be considered Big Data or not is based on the volume.

The rapidly increasing volume data is due to cloud-computing traffic, IoT, mobile traffic etc.

Velocity

It refers to the speed at which the data is getting accumulated. This is mainly due to IoTs, mobile data, social media etc.

Variety

It refers to **Structured**, **Semi-structured** and **Unstructured** data due to different sources of data generated either by humans or by machines.

Structured data: It's the traditional data which is organized and conforms to the formal structure of data. This data can be stored in a relational database. Example: Bank statement containing date, time, amount etc.

18. What is meant by pivot table?

A **pivot table** is a table of grouped values that aggregates the individual items of a more extensive table (such as from a database, spreadsheet, or business intelligence program) within one or more discrete categories. This summary might include sums, averages, or other statistics, which the pivot table groups together using a chosen aggregation function applied to the grouped values.

19. What is meta data?

Metadata is data about data. In other words, it's information that's used to describe the data that's contained in something like a web page, document, or file. Another way to think of

metadata is as a short explanation or summary of what the data is. A simple example of metadata for a document might include a collection of information like the author, file size, the date the document was created, and keywords to describe the document.

20. Write the importance of business analytics.

- Business analytics is a methodology or tool to make a sound commercial decision. Hence it impacts functioning of the whole organization. Therefore, business analytics can help improve profitability of the business, increase market share and revenue and provide better return to a shareholder.
- Facilitates better understanding of available primary and secondary data, which again affect operational efficiency of several departments.
- Provides a competitive advantage to companies. In this digital age flow of information 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.

21. Define prescriptive analytics.

Prescriptive analytics is a process that analyzes data and provides instant recommendations on how to optimize business practices to suit multiple predicted outcomes. In essence, prescriptive analytics takes the "what we know" (data), comprehensively understands that data to predict what could happen, and suggests the best steps forward based on informed simulations.

22. What is mean by SAS?

SAS (Statistical Analysis System) is a programming language for statistical analysis that is useful in various fields and industries for data mining and related data handling. It provides results related to multivariate analysis, predictive analytics and more.

23. Write a note on SPSS.

SPSS Statistics is a statistical software suite developed by IBM for data management, advanced analytics, multivariate analysis, business intelligence, and criminal investigation. It is a suite of software programs that analyzes scientific data related to the social sciences. SPSS offers a fast-visual modeling environment that ranges from the smallest to the most complex models. The data obtained from SPSS is used for surveys, data mining, market research, etc.

24. What is MATLAB?

MATLAB (an abbreviation of "MATrix LABoratory") is a proprietary multi-paradigm programming language and numeric computing environment developed by MathWorks. MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages.

25. What is meant by gap analysis?

Gap analysis is the method organizations use to monitor their actual performance in comparison to their anticipated and projected performance. This research is used to assess if a company is meeting its standards and is making good use of its resources. Gap analysis is the process by which a company can identify its current state through evaluating time, money, and labor, and make a comparison to its target state.

UNIT-II

1. Define business analytics personnel.

A business analytical personnel focuses on statistics, data, reporting, and recording data. They also aim to improve various functions in a business but the difference is they analyze numbers to do so. Business analysis has less to do with data and instead focuses on analyzing and optimizing the processes and functions that make up a business.

2. What are the skills required for business analytics personnel?

- Understanding the Business Objective
- Analytical and Critical Thinking
- Communication and Interpersonal Skills
- Negotiation and Cost-Benefit Analysis
- Decision-Making Skills
- Programming Languages
- Creation of Reports and Dashboards
- Database and SQL
- Microsoft Excel
- Documentation and Presentation

3. Mention the key responsibilities of business analytics personnel.

- Assessing and identifying areas for improvement in the company's present business processes
- Modernizing systems by investigating and evaluating new IT developments and contemporary business practices
- Presenting concepts and results at meetings
- Coaching and training employees
- Developing ideas in accordance with the demands and needs of the business
- Creating projects and monitoring their development
- Cooperating with stakeholders and users
- Maintaining close communication with senior management, partners, clients, and technicians

4. Explain data business analyst.

Business data analysts assess historical data that helps them to optimize business operations by identifying the existing loopholes, leading to optimum utilization of resources and efficiency in the system, saving up a considerable chunk of expenditure for the organization.

5. What is primary source of data?

- Primary data means first-hand information collected by an investigator.
- It is collected for the first time.
- It is original and more reliable.

- For example, the population census conducted by the government of India after every ten years is primary data.

6. What is business intelligence analyst?

A business intelligence analyst, also known as a BI analyst, uses data and other information to help organizations make sound business decisions. Business intelligence (BI) analysts transform data into insights that drive business value. Through use of data analytics, data visualization and data modeling techniques and technologies, BI analysts can identify trends that can help other departments, managers and executives make business decisions to modernize and improve processes in the organization.

7. Define web mining.

Web mining is the process of using data mining techniques and algorithms to extract information directly from the Web by extracting it from Web documents and services, Web content, hyperlinks and server logs. The goal of Web mining is to look for patterns in Web data by collecting and analyzing information in order to gain insight into trends, the industry and users in general.

8. What do you mean by DBMS?

A database management system (or DBMS) is essentially nothing more than a computerized data-keeping system. Users of the system are given facilities to perform several kinds of operations on such a system for either manipulation of the data in the database or the management of the database structure itself.

9. What is secondary data?

Secondary data refers to data that is collected by someone other than the primary user.[1] Common sources of secondary data for social science include censuses, information collected by government departments, organizational records and data that was originally collected for other research purposes. Primary data, by contrast, are collected by the investigator conducting the research.

10. Define data quality.

Data quality measures how well a dataset meets criteria for accuracy, completeness, validity, consistency, uniqueness, timeliness, and fitness for purpose, and it is critical to all data governance initiatives within an organization. Examples of data quality issues include duplicated data, incomplete data, inconsistent data, incorrect data, poorly defined data, poorly organized data, and poor data security.

11. What is data privacy?

Data privacy is a discipline intended to keep data safe against improper access, theft or loss. It's vital to keep data confidential and secure by exercising sound data management and preventing unauthorized access that might result in data loss, alteration or theft.

12. What are the tools used by business analyst?

- Microsoft Office
- Wrike
- Oracle NetSuite
- Pencil
- Trello
- Xplenty
- Other tools

- Inbound marketing can involve dozens of different sub-tasks, so it's no surprise there's a panoply of CRM tools here, including HubSpot, Salesforce, and Zoho, among others.

13. What do you understand by IT business analyst?

An IT Business Analyst, or Business Systems Analyst, is in charge of assessing information technology systems and making suggestions for how they can better meet business needs. Business analysis is the business process of assessing an organization's structure, processes, technology, and capabilities to identify and define solutions to roadblocks that impede the achievement of organizational goals.

14. Difference between data mining and text mining.

S.No.	Data Mining	Text Mining
1.	Data mining is the statistical technique of processing raw data in a structured form.	Text mining is the part of data mining which involves processing of text from documents.
2.	Pre-existing databases and spreadsheets are used to gather information.	The text is used to gather high quality information.
3.	Processing of data is done directly.	Processing of data is done linguistically.
4.	Statistical techniques are used to evaluate data.	Computational linguistic principles are used to evaluate text.
5.	In data mining data is stored in structured format.	In text mining data is stored in unstructured format.
6.	Data is homogeneous and is easy to retrieve.	Data is heterogeneous and is not so easy to retrieve.

15. List the roles of business analytics personnel.

- Conducting meetings with developing team and stakeholders

- System possibilities
- Present the company
- Present the details
- Implementation of the project

- Functional and non-functional requirements of a business
- Testing
- Decision making and problem-solving
- Maintenance
- Building a team
- Presentation and Documentation of the Final Project

16. Define data modeling.

Data modeling is the process of diagramming data flows. When creating a new or alternate database structure, the designer starts with a diagram of how data will flow into and out of the database. This flow diagram is used to define the characteristics of the data formats, structures, and database handling functions to efficiently support the data flow requirements.

17. Distinguish between BI & AI.

S. No.	Factors	Artificial Intelligence	Business Intelligence
1.	Concept	Artificial intelligence involves humans like computer intelligence.	Business intelligence involves intelligent decision-making.
2.	Focus	It deals with the principles of statistical analysis.	It deals with machine learning and deep learning algorithms.
3.	Application	It is mainly used in robotics, image recognition, virtual gaming, fuzzy logic, etc.	It is used in data extraction and data warehousing techniques.
4.	Starts with	It begins with instructing systems to think and act like people, and it concludes with foresight into the future.	The process begins with collecting and analyzing data points from multiple data sources and concludes with visual dashboards and reports.
5.	Scope	Its scope is associated with events of the future.	Its scope is associated with what has happened in the past.

18. Mention the challenges of business analytics.

- Increase in number of Sources
- Shortage of Talent for Data Analytics
- Data Secrecy
- Handling Large Volumes of Data
- Changing technological Landscape
- Quality of storage and retrieving data

19. Define clustering.

Clustering is the task of dividing the unlabeled data or data points into different clusters such that similar data points fall in the same cluster than those which differ from the others. In simple words, the aim of the clustering process is to segregate groups with similar traits and assign them into clusters.

20. What is simulation?

A simulation is a model that mimics the operation of an existing or proposed system, providing evidence for decision-making by being able to test different scenarios or process changes. This can be coupled with virtual reality technologies for a more immersive experience.

21. Define data translation.

Data translation can be defined as the process of converting volumes of data from one syntax to another and performing value lookups or substitutions from the data during the process. Translation can include data validation as well. One example of data translation is to convert EDI purchase order document data into purchase order database files or even flat files while performing data validation on the source data.

22. Difference between intranet and extranet.

| S.NO	Intranet	Extranet
1.	Intranet is a tool for sharing information throughout the organization.	Whereas Extranet is a tool for sharing information between the internal members and external members.
2.	Intranet is owned by a single organization.	While Extranet is owned by either a single or a many organization.
3.	In intranet, security is implemented through a firewall.	Whereas in this, security is implemented through a firewall in order to separate the extranet and the internet.
4.	Intranet is managed by an organization.	Whereas Extranet is managed by many organizations.

S.NO	Intranet	Extranet
5.	Intranet has a limited number of connected devices.	Whereas in the extranet, connected devices are comparable with the intranet.

23. Define systematic sampling.

Systematic sampling is a statistical method used to select a sample from a larger population systematically and randomly. It is a widespread technique for researchers and analysts who want to gather data from a large population without surveying every individual. This method is beneficial when the population is large, diverse, or hard to reach.

24. Explain Bowley's coefficient of skewness.

The coefficient of skewness is a measure of asymmetry in the distribution. A positive skew indicates a longer tail to the right, while a negative skew indicates a longer tail to the left. A perfectly symmetric distribution, like the normal distribution, has a skew equal to zero.
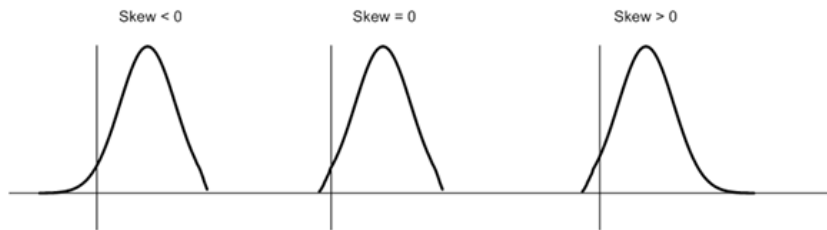
25. What is mean by normal distribution?

A normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme. The middle of the range is also known as the mean of the distribution.

UNIT-III

1. What are descriptive analytics?

Descriptive analytics is a statistical interpretation used to analyze historical data to identify patterns and relationships. Descriptive analytics seeks to describe an event, phenomenon, or outcome. It helps understand what has happened in the past and provides businesses the perfect base to track trends.

2. State the meaning and properties of mode.

It is defined as the value occurring in the given set. This means the value or number in a data set that has a high frequency is called modal value or mode. It is the three measures of central tendency, apart from mean and median. Mode, like median, can be calculated graphically. Mode is not affected by extreme values. The mode can be used for open-ended distribution and qualitative data.

3. What do you mean by measures of central tendency?

A measure of central tendency (also referred to as measures of centre or central location) is a summary measure that attempts to describe a whole set of data with a single value that represents the middle or centre of its distribution. Measures of central tendency are summary statistics that represent the center point or typical value of a dataset. Examples of these measures include the mean, median, and mode.

4. What is statistical estimation?

In statistics, **estimation** refers to the process by which one makes inferences about a population, based on information obtained from a sample. The term "estimation" is often used to describe the process of finding an estimate for an unknown value.

5. What is probability?

Probability means possibility. It is a branch of mathematics that deals with the occurrence of a random event. The value is expressed from zero to one. Probability has been introduced in Maths to predict how likely events are to happen. Probability denotes the possibility of the outcome of any random event. The meaning of this term is to check the extent to which any event is likely to happen.

6. Define mutually exclusive events.

In probability theory, two events are said to be mutually exclusive if they cannot occur at the same time or simultaneously. In other words, **mutually exclusive events** are called disjoint events. If two events are considered disjoint events, then the probability of both events occurring at the same time will be zero.

7. State the types of estimation.

Types of estimates

- Preliminary estimate.
- Detailed estimate.
- Quantity estimate.
- Bid estimate.
- Cost estimates.

8. What is sampling and non- sampling error?

Sampling error is one which occurs due to unrepresentativeness of the sample selected for observation. Conversely, non-sampling error is an error arise from human error, such as error in problem identification, method or procedure used, etc. Sampling error can only be estimated if probability sampling is used. The sampling error is the error caused by observing a sample instead of the whole population. while non-sampling error is error that arises from other sources, such as errors in data collection or data entry.

9. What is probability distribution?

A **probability distribution** is a mathematical function that describes the probability of different possible values of a variable. Probability distributions are often depicted using graphs or probability tables. The probability distribution function of a random variable always lies between 0 and 1.

10. Discuss assumptions of probability distribution.

The requirements for a probability distribution are:
1) The value of the random variable X must be described.
2) All the values of random variable X must have their corresponding probabilities.
3) The sum of all probabilities is equal to one.

11. Difference between probability and non-probability sampling.

Basis for comparison	Probability Sampling	Non-Probability Sampling
Definition	Probability sampling is a technique in which each member in a population has an equal chance of being selected as a sample.	Non-probability sampling is a method in which it is unknown who will be selected as a sample from the given population.
Alternatively known as	Random sampling	Non-random sampling
Selection basis	Random	Arbitrarily
Research	Conclusive	Exploratory
Inferences	Statistical	Analytical
Results	Results generated by probability sampling are unbiased	Results of non-probability sampling are more or less biased

12. What is mean, median, mode?

The Mean, Median and Mode are the three measures of central tendency. Mean is the arithmetic average of a data set. This is found by adding the numbers in a data set and dividing by the number of observations in the data set. The median is the middle number in a data set when the numbers are listed in either ascending or descending order. The mode is the value that occurs the most often in a data set and the range is the difference between the highest and lowest values in a data set.

13. Define standard deviation.

A **standard deviation** (or σ) is a measure of how dispersed the data is in relation to the mean. Low standard deviation means data are clustered around the mean, and high standard deviation indicates data are more spread out. A standard deviation close to zero indicates that data points are close to the mean, whereas a high or low standard deviation indicates data points are respectively above or below the mean.

14. What is co-efficient of quartile deviation?

The coefficient of quartile deviation is a measure of the spread or variability of a set of data. It is calculated by taking the difference between the upper and lower quartiles, squaring it, and then taking the square root. It is expressed as a percentage, making it easy to compare data sets. It can be used to identify unusual values in a set of data and to measure the degree of variation within a data set.

15. Discuss the demerits of range.

The demerits of range are-

1: It is highly affected by extreme values.

2: In case of open-ended distributions, it is difficult to calculate.

3: It doesn't tell us much about distributions of items of the series.

4: It considered only upper and lower limit. It doesn't take rest of the distribution.

16. Explain the coefficient of skewness of Karl Pearson.

Pearson's coefficient of skewness (second method) is calculated by multiplying the difference between the mean and median, multiplied by three. The result is divided by the standard deviation. The coefficient of skewness is a measure of asymmetry in the distribution. A positive skew indicates a longer tail to the right, while a negative skew indicates a longer tail to the left.

17. Explain the concept Random Variable.

A random variable is a variable whose value is unknown or a function that assigns values to each of an experiment's outcomes. A random variable can be either discrete (having specific values) or continuous (any value in a continuous range).

18. What is Independent Event?

Independent events are those events whose occurrence is not dependent on any other event. For example, if we flip a coin in the air and get the outcome as Head, then again if we flip the coin but this time, we get the outcome as Tail. In both cases, the occurrence of both events is independent of each other.

19. Explain the term Measures of Dispersion

A measure of dispersion indicates the scattering of data. It explains the disparity of data from one another, delivering a precise view of their distribution. The measure of dispersion displays and gives us an idea about the variation and the central value of an individual item.

In other words, dispersion is the extent to which values in a distribution differ from the average of the distribution.

20. What is Regression analysis?

Regression analysis is a statistical method that shows the relationship between two or more variables. Usually expressed in a graph, the method tests the relationship between a dependent variable against independent variables.

21. Define systematic sampling.

Systematic sampling is a probability sampling method where researchers select members of the population at a regular interval – for example, by selecting every 15th person on a list of the population. If the population is in a random order, this can imitate the benefits of simple random sampling.

22. What is statistical inference?

Statistical inference is the process of analysing the result and making conclusions from data subject to random variation. It is also called inferential statistics. Hypothesis testing and confidence intervals are the applications of the statistical inference. Statistical inference is the process of drawing conclusions about an underlying population based on a sample or subset of the data.

23. Define snowball sampling.

Snowball sampling is a non-probability sampling method where new units are recruited by other units to form part of the sample. Snowball sampling can be a useful way to conduct research about people with specific traits who might otherwise be difficult to identify (e.g., people with a rare disease).

24. What is convenience sampling?

Convenience sampling is a non-probability sampling method where units are selected for inclusion in the sample because they are the easiest for the researcher to access. This can be due to geographical proximity, availability at a given time, or willingness to participate in the research.

25. Define cluster sampling.

Cluster sampling, when used, gives every unit/person in the population an equal and known chance of being selected in the sample group.For this method of sampling, researchers divide the population into internally heterogeneous and externally homogeneous subpopulations known as clusters.

UNIT-IV

1. What are predictive analytics?

Predictive analytics is the use of data, statistical algorithms and machine learning techniques to identify the likelihood of future outcomes based on historical data. The goal is to go beyond knowing what has happened to providing a best assessment of what will happen in the future.

2. What are neural networks?

Neural networks, also known as artificial neural networks (ANNs) or simulated neural networks (SNNs), are a subset of machine learning and are at the heart of deep learning algorithms. Their name and structure are inspired by the human brain, mimicking the way that biological neurons signal to one another.

3. List out the methodologies of data mining.

- Association
- Classification
- Clustering Analysis
- Prediction
- Sequential Patterns or Pattern Tracking
- Decision Trees
- Outlier Analysis or Anomaly Analysis
- Neural Network

4. Define data integration.

Data integration refers to the process of bringing together data from multiple sources across an organization to provide a complete, accurate, and up-to-date dataset for BI, data analysis and other applications and business processes.

5. State the procedure of predictive analysis.

- Define your project's objectives.
- Collect your data.
- Clean and prepare your data.
- Build and test your model.
- Deploy your model.
- Monitor and refine your model.

6. Write the merits of predictive analytics.

- Predictive Analytics Increase production efficiency.

- It minimizes business risks.

- It helps in decision-making purposes in any business organization.

- It drives a competitive environment.

7. Mention the applications of predictive analytics.

Marketing

Consumers are attracted with pool of advertising and marketing,

Retail

Either online or brick and mortar, each retailer looks for managing inventory and logistics, and thereby predictive analytics is extremely important.

Manufacturing

With the modernized technology and fully automated factory machines, predictive analytics tools are very significant in operating and optimizing the manufacturing process at each stage of designing, purchasing, developing, quality and inventory control, delivery, etc.

Healthcare

Healthcare industry is among dominant adapters to consider predictive analytics techniques aimed at facilitating technology to save money and improve health practices efficiencies.

Finance

Applying to a broad spectrum of banking and financial services & activities, predictive analytics is the most valuable process helping from accessing risks to maximizing customer satisfaction.

8. Write the principles of predictive models.

- A predictive model is not fixed; it is validated or revised regularly to incorporate changes in the underlying data.
- Predictive models make assumptions based on what has happened in the past and what is happening now.
- If incoming, new data shows changes in what is happening now, the impact on the likely future outcome must be recalculated, too.

9. What are the applications of data mining?

Banks

Data mining helps banks work with credit ratings and anti-fraud systems, analyzing customer financial data, purchasing transactions, and card transactions.

Healthcare

Data mining helps doctors create more accurate diagnoses by bringing together every patient's medical history, physical examination results, medications, and treatment patterns.

Marketing

If there was ever an application that benefitted from data mining, it's marketing! After all, marketing's heart and soul is all about targeting customers effectively for maximum results.

Retail

The world of retail and marketing go hand-in-hand, but the former still warrants its separate listing.

10. List the challenges of data mining.

- **Data Quality**
  The quality of data used in data mining is one of the most significant challenges.
- **Data Complexity**
  Data complexity refers to the vast amounts of data generated by various sources, such as sensors, social media, and the internet of things (IoT).

- **Data Privacy and Security**
  Data privacy and security is another significant challenge in data mining. As more data is collected, stored, and analyzed, the risk of data breaches and cyber-attacks increases.
- **Scalability**
  Data mining algorithms must be scalable to handle large datasets efficiently. As the size of the dataset increases, the time and computational resources required to perform data mining operations also increase.
- **Interpretability**
  Data mining algorithms can produce complex models that are difficult to interpret. This is because the algorithms use a combination of statistical and mathematical techniques to identify patterns and relationships in the data.
- **Ethics**
  Data mining raises ethical concerns related to the collection, use, and dissemination of data.

11. What are the four pillars of data analytics?

- The four pillars of data analytics —
- Descriptive Analytics
- Diagnostic Analytics
- Predictive Analytics
- Prescriptive Analytics

12. State techniques of predictive analytics.

- **Data mining**
- **Data warehousing**
- **Clustering**
- **Classification**
- **Predictive modeling**
- **Logistic regression**
- **Decision trees**
- **Time series analysis**
- **Neural networks**
- **Artificial intelligence and machine learning**

13. Mention the process of predictive analytics.
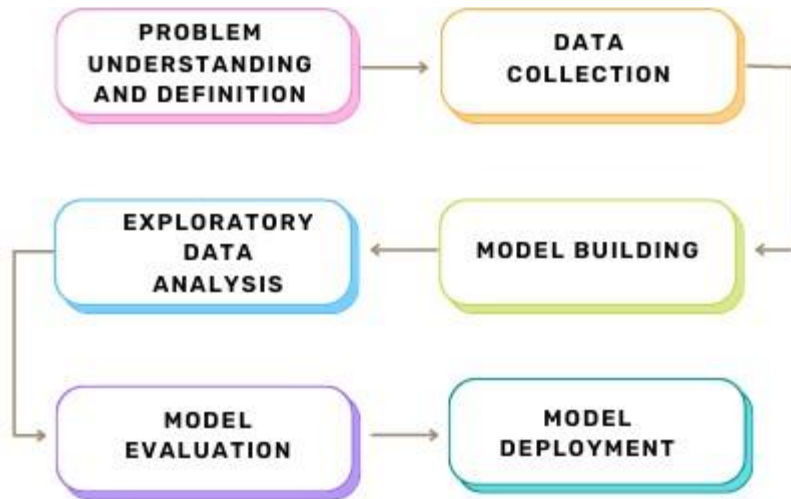
14. Write a note on uses of data mining.

- It helps companies gather reliable information
- It's an efficient, cost-effective solution compared to other data applications
- It helps businesses make profitable production and operational adjustments
- Data mining uses both new and legacy systems
- It helps businesses make informed decisions
- It helps detect credit risks and fraud
- It helps data scientists easily analyze enormous amounts of data quickly

15. List the components of data mining.

- Sources of Data
- Database or Data Warehouse Server
- Data Mining Engine
- Modules for Pattern Evaluation
- GUI or Graphical User Interface
- Knowledge Base

16. What are predictive analytics models?

- **Classification models**

Classification models fall under the branch of supervised machine learning models.
Clustering models

Clustering models fall under unsupervised learning. They group data based on similar attributes.

- **Time series models**

Time series models use various data inputs at a specific time frequency, such as daily, weekly, monthly, et cetera

17. Write the limitations of predictive modeling.

- **Limitation 1: Data Quality**

One of the most significant limitations of predictive analytics is data quality. Predictive models rely on large, accurate, and relevant datasets to produce accurate predictions.

- **Limitation 2: Overfitting**

Another limitation of predictive analytics is overfitting. Overfitting occurs when a model is trained on a specific dataset and becomes too complex, making it difficult to generalize to new data.

- **Limitation 3: Changing Conditions**

Predictive analytics models are designed to predict future outcomes based on historical data.

- **Limitation 4: Ethical Concerns**

Predictive analytics can also raise ethical concerns, particularly around issues of bias and privacy.

18. What are the types of data mining?

- Clustering
- Classification
- Association
- Anomaly detection
- Regression
- Text mining
- Web mining
- Summarization

## 19. Difference between predictive analytics and data mining.

Predictive Analytics	Data Mining
Predictive analytics refers to the use of both new and historical data, statistical algorithms, and machine learning techniques to forecast future activity, patterns, and trends	data mining refers to the computational technique of discovering patterns in huge data sets involving methods at the intersection of AI.
It helps to make predictions based on future events.	It helps to understand the gathered information better.

Business analysts and other SMEs perform it.	Statisticians and engineers perform it.
It applies business knowledge to find patterns to get valid business predictions.	It applies algorithms such as classification and regression on gathered information to find hidden patterns.

20. Write the merits of data mining.

- It helps companies gather reliable information
- It's an efficient, cost-effective solution compared to other data applications
- It helps businesses make profitable production and operational adjustments
- Data mining uses both new and legacy systems
- It helps businesses make informed decisions
- It helps detect credit risks and fraud
- It helps data scientists easily analyze enormous amounts of data quickly

21. State the applications of data mining.
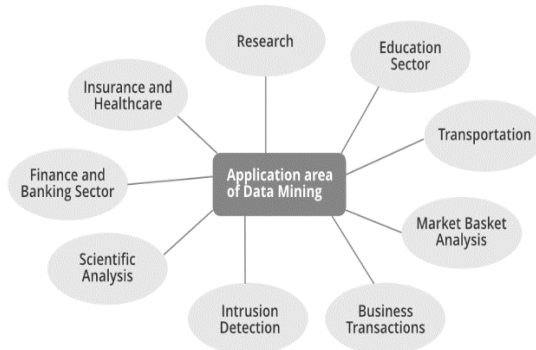
22. How does predictive analytics work?

Statistical models analyze past data to find patterns. In this way, they calculate probabilities for future events. Usually, experts use machine learning methods (a subfield of artificial intelligence and computer science) for this purpose.

It is also essential to know that predictive analytics is an entire field of study that encompasses various mathematical and statistical methods for making predictions. So there is no "one" predictive analytics method.

23. State the components of Data mining.

The major components of data mining are as follows −

- **Databases** − This is one or a set of databases, data warehouses, spreadsheets, and another type of data repository where data cleaning and integration techniques can be implemented.
- **Data warehouse server** − This component fetches the relevant records based on users request from a data warehouse.
- **Knowledge base** − It is a knowledge domain that is employed for discovering interesting patterns.
- **Data mining engine** − It uses a functional module that is used to perform tasks including classification, association, cluster analysis, etc.
- **Pattern evaluation module** − This component uses interestingness measures that communicate with data mining structure to target the search towards interesting patterns.
- **User interface** − This interface enables users to interact with the system by describing a data mining function or a query through the graphical user interface.

24. **What is randaom sampling?**

Random sampling is a part of the sampling technique in which each sample has an equal probability of being chosen. A sample chosen randomly is meant to be an unbiased representation of the total population. If for some reasons, the sample does not represent the population, the variation is called a sampling error.

25. **What is probability distribution?**

Probability distribution yields the possible outcomes for any random event. It is also defined based on the underlying sample space as a set of possible outcomes of any random experiment. These settings could be a set of real numbers or a set of vectors or a set of any entities. It is a part of probability and statistics. Random experiments are defined as the result of an experiment, whose outcome cannot be predicted.

26. Write note on Gradient boosted model.

Gradient Boosting is a system of machine learning boosting, representing a decision tree for large and complex data. It relies on the presumption that the next possible model will minimize the gross prediction error if combined with the previous set of models. The decision trees are used for the best possible predictions. The gradient boosting is also known as the statistical prediction model. It works quite similarly to other boosting methods even though it allows the generalization and optimization of the differential loss functions.

27. What is social analytics?

Social analytics is monitoring, analyzing, measuring and interpreting digital interactions and relationships of people, topics, ideas and content. Interactions occur in workplace and external-

facing communities. Social analytics include sentiment analysis, natural-language processing and social networking analysis (influencer identification, profiling and scoring), and advanced techniques such as text analysis, predictive modeling and recommendations, and automated identification and classification of subject/topic, people or content.

UNIT-V

1. **What are prescriptive analytics?**
   Prescriptive analytics is the use of advanced processes and tools to analyze data and content to recommend the optimal course of action or strategy moving forward. Simply put, it seeks to answer the question, "What should we do?" It analyzes raw data about past trends and performance through machine learning (so very little human input, if any at all) to determine possible courses of action or new strategies generally for the near term.

2. **What are the techniques of prescriptive analytics?**
   Specific techniques used in prescriptive analytics include **optimization, simulation, game theory and decision-analysis methods**. Data science and machine learning tools form the foundation of a prescriptive analytics practice.

3. **Write the applications of prescriptive analytics.**
   - Prescriptive Analytics for Hospitals and Clinics
   - Prescriptive Analytics for Airlines
   - Prescriptive Analytics in Banking
   - Prescriptive Analytics in Marketing

4. **State the merits of prescriptive analytics.**
   - Prescriptive analytics can cut through the clutter of immediate uncertainty and changing conditions.
   - It can help prevent fraud, limit risk, increase efficiency, meet business goals, and create more loyal customers.
   - When used effectively, it can help organizations make decisions based on highly analyzed facts rather than jump to under-informed conclusions based on instinct.

5. **What are the challenges of prescriptive analytics?**
   Data Quality
   Since prescriptive analytics relies heavily on the accuracy and reliability of data, low-quality data can thwart its ability to generate insights and recommendations. With supply chain integrating with multiple systems like enterprise resource planning (ERP), warehouse management systems (WMS), and transportation management systems (TMS), ensuring that they share accurate data is difficult.
   Resistance to Change
   The implementation of prescriptive analytics calls for modifications in processes and workflows. This can get challenging, as employees could be resistant to the training needed to adapt to a technology-driven way of working.

Cost and ROI
Deploying prescriptive analytics requires investments in technology, data infrastructure and training. Businesses may also need to show a positive return on investment (ROI) to justify investments in prescriptive analysis.

6. Explain the prescriptive modeling.
- Models the entire business
- Is 100 percent data driven
- Recommends specific business decisions
- Considers interdependencies
- Is not bound by static rules
- Provides tangible, measurable benefits
- Supports what-if scenarios
- Is free of "gut feel" and personal bias
- Accounts for all inputs, variables and outputs

7. List the types of prescriptive modeling.
There are three types of prescriptive process models. They are:

1. The Waterfall Model
2. Incremental Process model
3. RAD model

8. Why Is Prescriptive Analytics So Important for Businesses?
Prescriptive analytics help businesses identify the best course of action, so they achieve organizational goals like cost reduction, customer satisfaction, profitability etc. While figuring out what you should do is a crucial aspect of any business, the value of prescriptive analytics is often missed.

9. Write the pros and cons of prescriptive analytics.
Pros of Prescriptive Analytics
- Improved Decision-Making
- Increased Efficiency
- Enhanced Customer Experience
- Competitive Advantage

Cons of Prescriptive Analytics
- **Data Quality**: In order to provide accurate insights and recommendations, prescriptive analytics relies on high quality data, which can be difficult to obtain and clean.
- **Complexity:** Implementing prescriptive analytics is a complex process which requires advanced technical skills and expertise.
- **Cost:** Prescriptive analytics requires huge investment in technology and skilled personnel.
- **Resistance to Change:** Prescriptive analytics can require significant changes in existing business processes and practices.

10. What is time series analysis?

Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly.

11. Define doughnut chart.

A doughnut chart is a chart in Excel whose visualization function is similar to pie charts. The categories represented in this chart are parts, and together they express the whole data in the chart.

12. What is stratified random sampling?

Stratified random sampling is a method of sampling that involves the division of a population into smaller subgroups known as strata. In stratified random sampling, or stratification, the strata are formed based on members' shared attributes or characteristics, such as income or educational attainment.

13. Write the types of probability distribution.

The probability distribution is divided into two parts:

1. Discrete Probability Distributions
2. Continuous Probability Distributions

Discrete Probability Distribution
A discrete distribution describes the probability of occurrence of each value of a discrete random variable.

Continuous Probability Distributions
A continuous distribution describes the probabilities of a continuous random variable's possible values. A continuous random variable has an infinite and uncountable set of possible values (known as the range).

14. What is point estimation?

Point estimation, in statistics, the process of finding an approximate value of some parameter—such as the mean (average)—of a population from random samples of the population. It can be used for many data mining tasks such as summarization and time-series prediction. Summarization is the process of extracting or deriving representative information about the data.

15. What is interval estimation?

In statistics, **interval estimation** is the use of sample data to estimate an interval of possible values of a parameter of interest. This is in contrast to point estimation, which gives a single value.

The most prevalent forms of interval estimation are confidence intervals (a frequentist method) and credible intervals (a Bayesian method); less common forms include likelihood intervals and fiducial intervals.

16. List the types of events in probability.

Types of Events in Probability:

- Impossible and Sure Events.
- Simple Events.
- Compound Events.
- Independent and Dependent Events.
- Mutually Exclusive Events.
- Exhaustive Events.
- Complementary Events.
- Events Associated with "OR"

17.Point out the three types of business analytics.

The different types of business analytics are mentioned below:

- Descriptive Analytics: Summarizing and describing past events
- Diagnostic Analytics: Examining past performance to find causes
- Predictive Analytics: Forecasting future events using historical data and models/ML
- Prescriptive Analytics: Recommending specific actions based on data analysis

18.Brief on compensation of business analyst team.
This means their role is to develop technical solutions to problems in a business or to further a company's sales revenue by defining, documenting and analyzing requirements. When they use their role to manage requirements during a project, they begin to help fulfill business needs.

19.What type of analytics is considered most crucial for businesses in general?

Prescriptive analytics is one of the key types of business analytics. It predicts future events in a business and outlines the steps to be taken to achieve the desired outcome. It produces guidelines and recommends actions to be taken, making it highly sought after in the industry.

20.What are the purposes of descriptive analytics?
- It doesn't require a deep understanding of analytical or statistical methods, and it can be performed with readily available tools.
- It can answer many of the most common questions about business performance, such as whether the last quarter's sales were in line with goals.
- This helps the business identify areas in need of improvement.

21.State any two tools of descriptive analytics?
- **Measures of Central Tendency**
  A measure of central tendency is one value that can represent the total population and act like the central gravity towards which all the other values move.
- **Variance and Standard Deviation**
  Standard deviation - as the name suggest is a measure of the deviation.
  Deviation means change or distance.

22.What is predictive analytics with examples?

Predictive analytics is the practice of studying historic data to help predict future events and trends. Predictive analytics typically uses tools like big data, machine learning, and regression analysis–which examines the relationships between historical data points–to improve the accuracy of the predictions.

23. List four steps in predictive analytics.

Five key phases in the predictive analytics process cycle require various types of expertise:

- Define the requirements,
- Explore the data,
- Develop the model,
- Deploy the model and
- Validate the results.

24. What techniques do business analysts use?

- **SWOT Analysis:** Assesses Strength, Weakness, Opportunities, and Threat in a business through internal and external forces
- **MOST Analysis:** Mission, Objective, Strategy, Tactics or MOST Analysis is a simple and useful tool used to analyze and plan the details of what an organization does or should do
- **Mindmapping:** A graphic way to represent concepts and ideas to strategize, analyze, comprehend and generate new ideas
- **PESTLE analysis:** Used to identify what drives change in a business. Stands for Political, Economic, Social, Technological, Legal and Environmental
- **Brainstorming:** A technique that is used to problem-solve, find facts and generate ideas

25. Distinguish predictive and prescriptive analytics.

Criteria	Predictive Analysis	Prescriptive Analysis
Summary	What's going to happen?	What should happen?
Function	It looks at historical data and analyzes past data trends to predict what could happen.	It takes the conclusions gleaned from descriptive and predictive analysis and recommends the best future course of action.
Pros	It's a valuable forecasting tool.	It offers critical insights into making the best, most informed decisions.

Criteria	Predictive Analysis	Prescriptive Analysis
Cons	It needs lots of historical data to work. It will never be 100% accurate.	It requires a lot of past data and often cannot account for all possible variables.

26. What do you mean by nonlinear optimization.

Nonlinear programming is minimizing or maximizing a nonlinear objective function subject to bound constraints, linear constraints, or nonlinear constraints, where the constraints can be inequalities or equalities. Example problems in engineering include analyzing design tradeoffs, selecting optimal designs, computing optimal trajectories, and portfolio optimization and model calibration in computational finance.

UNIT-I
1. Discuss the scope and importance of business analytics.
The Future Scope of Business Analytics

The use of business analytics is becoming increasingly important as organisations strive to gain a competitive edge and maximise profits. The future scope of business analytics is vast and will continue to expand as technology advances and the need for data-driven decisions increases.
- Automation

Business analytics is becoming increasingly automated as organisations look to simplify processes and reduce costs. Automated solutions can collect, organise and analyse data quickly and accurately, giving businesses near real-time insights into their operations. This can help organisations make faster decisions with greater accuracy.
- Big Data Analytics

As the amount of available data grows exponentially, businesses are turning to big data solutions for improved performance. By leveraging the power of **big data** tools such as Hadoop or Spark, companies can gain deeper insight into customer behavior, market trends, and more to optimise their strategies for maximum success.

- Artificial Intelligence (AI)

AI technologies such as **machine learning** algorithms are being used more frequently in business analytics to detect patterns, make predictions and optimise decisions. AI-powered solutions are becoming increasingly sophisticated, making them invaluable for organisations seeking a competitive edge in the market.

- Cloud Computing

Cloud computing transforms how businesses handle data and analytics. By moving their analytical processes to the cloud, companies can reduce costs and improve scalability while accessing powerful tools for analysis and leveraging near real-time insights into their operations.

- Internet of Things (IoT)

The IoT revolution is creating vast amounts of corresponding data that can be analysed for improved performance. Companies can use IoT data to gain insights into customer behavior, optimise operations or develop new products and services based on customer needs and requirements.

Importance of Business Analytics

- **Business analytics is a methodology or tool to make a sound commercial decision**. Hence it impacts functioning of the whole organization. Therefore, business analytics can help improve profitability of the business, increase market share and revenue and provide better return to a shareholder.
- Facilitates better understanding of available primary and secondary data, which again affect operational efficiency of several departments.
- Provides a competitive advantage to companies. In this digital age flow of information 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.

2. Explain the components of business analytics.

Components of Business Analytics

The following are the primary components of a typical business analytics dashboard:

Components of Business Analytics
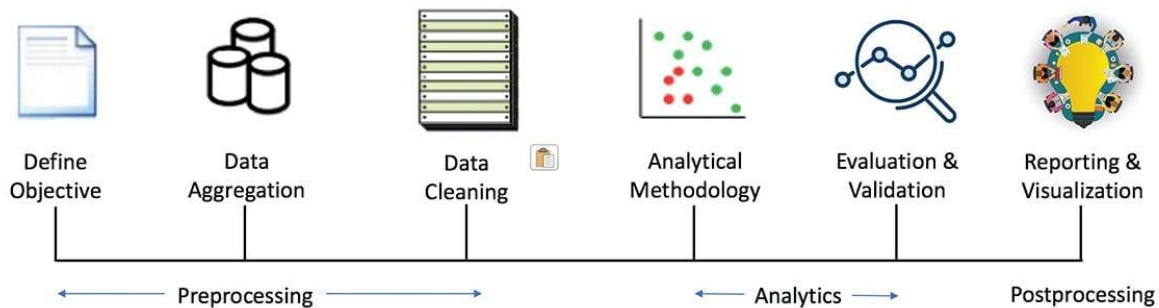
- **Define Objective**
  This is the foremost step. Without having a clear understanding of business goals, questions we need to answer, and problems we ought to solve, none of the following steps will deliver. This also helps us to translate business objectives into analytics objective and map data requirements.

- **Data Aggregation**
  Data must first be obtained, sorted, and filtered, either through volunteered data or transactional records before it can be analysed.

- **Data Mining**
  To detect trends and establish links, data mining for business analytics filters through enormous datasets using databases, statistics, and machine learning.

- **Association and Sequence Identification**
  The identification of predictable activities that are carried out in conjunction with other acts or in a sequential order

- **Text Mining**
  For qualitative and quantitative analysis, examines and organises big, unstructured text databases.

- **Forecasting**
  Analyses historical data from a given time period in order to create educated predictions about future occurrences or behaviours.

- **Predictive Analytics**
  Predictive business analytics employs a number of statistical techniques to build models

that extract data from datasets, discover patterns, and provide a score for a variety of organisational outcomes.

- **Optimization**
  Businesses can use simulation tools to test out best-case scenarios once patterns have been discovered and predictions have been made.

- **Data Visualization**
  It provides visual representations of data, such as charts and graphs, to make data analysis simple and rapid.

- **Data Cleaning**
  Data Cleaning is an extremely important component of business analytics because the data in its raw form sometimes is not directly usable. As the other components of Business Analytics use mathematics, statistics, and computer programming, the data must be compatible with these streams of study.

- **Analytical Methodology**
  Having a detailed understanding of the different type of analytics out there dominate this component as this is where the analysts have to identify the method with which they will go to achieve their end goal. If the end goal is to understand what is the present situation of the business, then that requires a different set of methods while if there is a need to identify which has happened in the past or what can happen in the future, then a different technique is required.

- **Evaluation and Validation**
  Once the results come out, the next task is to understand if the result stands true given a different situation or not. This is where is predictive models are used and their evaluation and validation are conducted whereas, for other methods, various simulation techniques are put to use to identify the most plausible outcome, thus providing provide a very reliable result.

3. Describe the types of business analytics

Types of Business Analytics

The four most popular types of business analytics are descriptive, diagnostic, predictive, and prescriptive. The fifth– cognitive analytics is a new type that employs AI, ML, and deep learning. Whilst each of these business analytics types is effective when used individually, they become extremely powerful when employed together.

1. Descriptive Analytics

It analyses historical data to determine the response of a unit over a set of given variables. It tracks key performance indicators (KPIs) for a better understanding of the present state of a business.

It involves the following five steps:

- Deciding which business metrics will effectively evaluate performance against objectives
- Identifying required data as per the current business state
- Collecting and preparing data using various processes like depublication, transformation, and cleansing.
- Analyzing data for patterns to measure performance
- Presenting data in charts and graphs to make it understandable for non-analytics experts

Examples of Descriptive Analytics

- Summarizing past events, exchange of data, and social media usage
- Reporting general trends

2. Diagnostic Analytics

Diagnostic Analytics is one of those business analytics types that help understand why things happened in the past. Using drill-downs, data mining, data discovery, and correlations, you can comprehend the driving factors.

This advanced analytics method is usually employed as a preceding step of Descriptive Analytics to find the reasoning behind certain results in finance, marketing, cybersecurity, and more.

Examples of Diagnostic Analytics

- Examining market demand
- Identifying technical issues
- Explaining customer behavior
- Improving organization culture

3. Predictive Analytics

It considers historical data trends for determining the probability of particular future outcomes. It uses several techniques like data mining, machine learning algorithms, and statistical modeling to forecast the likelihood of events.

Predictive analytics helps improve business areas, including customer service, efficiency, fraud detection and prevention, and risk management. It allows you to grow the most profitable customers, improve the operations of businesses, and determine customer responses and cross-sell opportunities.

Examples of Predictive Analytics

- Predicting customer preferences
- Detection of employee intentions
- Recommending products
- Predicting staff and resources

4. Prescriptive Analytics

Prescriptive analytics generates recommendations to handle similar future situations relying on past performances. It employs several tools, statistics, and ML algorithms for the available internal data and external data.

It gives you insights into what may happen, when, and why.

Examples of Prescriptive Analytics

- Tracking fluctuating manufacturing prices
- Improving equipment management
- Suggest the best course of action
- Price modeling
- Evaluating rates of readmission
- Identifying testing

5. Cognitive Analytics

Combining Artificial Intelligence and Data Analytics, Cognitive Analytics is one of the newest types of business analytics. It looks at the available data in the knowledge base and discovers the best solutions for the questions posed.

Cognitive analytics covers multiple analytical techniques to analyze large data sets and monitor customer behavior patterns and emerging trends.

Examples of Cognitive Analytics

- Tapping unstructured data sources such as images, text documents, emails, and social posts.

4. Discuss the terminologies of business analytics.

High-level practices and methodologies used with data.

- **Artificial Intelligence (AI)** is a broad term for using vast data sets to provide a high level of understanding and sometimes a higher level of consciousness. Within the realm of analytics, artificial intelligence can apply to machine learning.
- **Big Data** is a widely-used term referring to the derivation of insights from large data sets. Big data is often associated with large unstructured data sets such as social media feeds and IoT data streams. "Big Data" is often overused and misused for the purposes of marketing.

- **Business Analytics (BA)** involves analyzing business data to effect change at a company. It is process-oriented and focused on using data as a functional, decision making tool to improve business efficiencies, business operations, and business profits.
- **Business Intelligence (BI)** describes the act of informing an organization using data. Business intelligence is typically associated with information dashboarding, specifically high-level reporting that summarizes previous traditional tabular reporting into charts and data visualizations. In recent years, business intelligence has been overcome by business analytics because users are asking deeper questions of their data.
- **Business Management** encompasses all operations necessary to run an organization. Business management requires standardized processes to effectively operate all departments. Technologies for data pipelines and data analytics can greatly enhance the ease and efficacy of an organization's business management efforts.
- **Data Analytics** is the reporting and visualization of business information. A more technical definition would include: the process of cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions and supporting decision-making.
- **Data Science** is the act of applying industry domain knowledge and advanced statistical analysis to one or more data sets. It includes bringing together data from various sources, applying statistical analysis, and presenting the results to key decision makers. Data science tools include statistics models, machine learning, artificial intelligence, and data visualization.
- **Database Management** is the act of maintaining complex databases. Databases are computer application systems that require software patches, backups, storage management, performance tuning, and software patches.
- **Descriptive Analytics** explains the current state of the data. It is the first of three stages in the progression of increased sophistication in data analytics. Descriptive Analytics is followed by predictive analytics and prescriptive analytics. Descriptive analytics explains the current state of the data. Predictive analytics applies statistical analysis to predict future behavior. Prescriptive analytics provides a specific action plan to improve a business function.
- **Information Repository** refers to a data set that is organized for decision making. "Decisions are made on information, not data." An information repository can be comprised of one or more databases, documents, or any other electronic data system.
- **Predictive Analytics** applies statistical analysis to predict future behavior. explains the current state of the data. It is the second of three stages in the progression of increased sophistication in data analysis. Predictive Analytics is preceded by descriptive analytics and followed by prescriptive analytics.
- **Prescriptive Analytics** is finding the best course of action for a given situation. It is the third of three stages in the progression of increased sophistication in data analysis. Prescriptive analytics is preceded by descriptive analytics and prescriptive analytics.
- **Statistical Analysis** is the general application of math and statistics to data. It is used in data science, time-series forecasting, predictive analytics.

- **Variety, Velocity, Volume** are often used to describe large data sets used in data analytics. Volume describes the vast amount of data that often feeds into analysis. Velocity describes how quickly the data sets change. Variety describes how many types of data are used for analysis.
- **Business Analysts** analyze an organization's data, processes, and systems to provide guidance for improving business processes, products, services and software through data analysis. Often a member of a data team.
- **Business Expert** is an employee with in-depth domain knowledge of how a business operates, the systems it uses, internal and external processes, company structure and strategies, and technologies. At ReconInsight, we recommend a Business Expert be a part of a company's data team.
- **Data Analyst** is a technical role focused on implementing, supporting, and optimizing data analytics technology. They are skilled administrators, architects, and users of data tools.

-

5.   Explain the Evolution and process of business analytics.

Evolution of Business Analytics

In recent times, business analytics has evolved into a much more advanced set of tools and techniques assisted by automation and big data. Initially, business analytics was limited to a few corporate applications used by only the major MNCs. The first adoption of computing for business was noticed in the use of report-building, presentations and data entry using applications such as Microsoft Excel. Later on, more advanced applications that involved multi-dimensional data processing and data analytics were seen using add-ons such as PowerPivot in Excel. This was still a long way from BI or business intelligence, which the highly accredited firm, Gartner.

Business Analytics Evolution Chart Example Of PPT
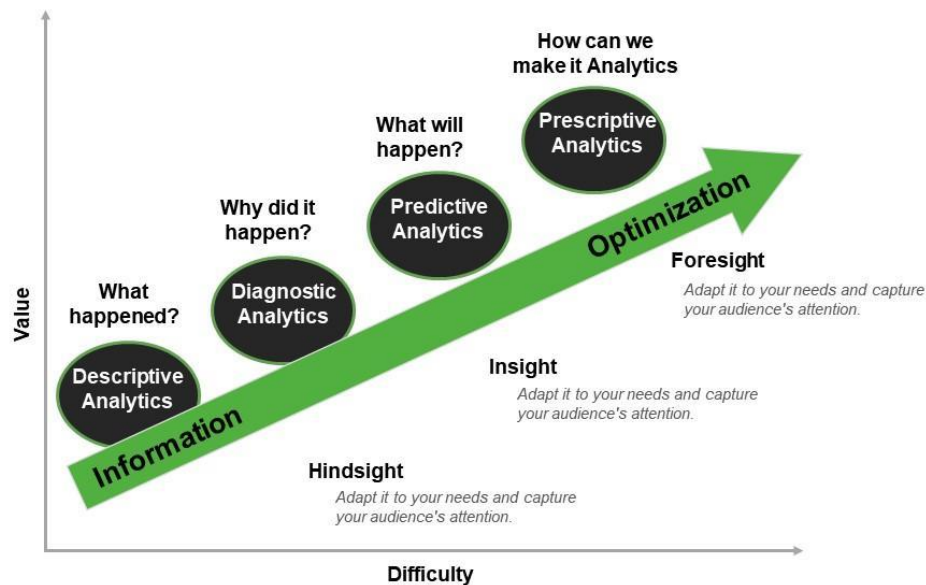
Here are some more historical facts about business analytics:

- During the early 1900s, Henry Ford, inspired by Frederick Taylor's scientific management system, hired him in order to measure the performance of the assembly line of his famous Ford Model T. This led to a series of events that had transformed the manufacturing industry and production lines across the world. This also helped Henry Ford make his assembly line as efficient as possible.

- In 1956, IBM introduced the first hard disk drive that allowed users to store data that can be used for business or corporate purposes.

- During the 1970s, Bill Inmon started discussing the concept of a data warehouse to solve the problem of storing vast amounts of data for business intelligence.

- During the 1980s, the first business data warehouse was developed by IBM researchers Barry Devlin and Paul Murphy.

- In this period between the 1990s and early 2000s, various solutions and software were introduced, such as business intelligence tools by companies like SAP, Microsoft, SAS and IBM alongside relational databases.

- After the early 2000s, common people started using data more proactively for personal purposes. This also led to more corporate use of data through employees extensively using organisational data. More tools were also introduced during this time with which individuals can use business intelligence tools without extensive

training. Eventually, Google Analytics was introduced that allowed website owners to analyse statistics about their website, such as trends in website visits.

- After 2010, business intelligence and analytics truly took off, being adopted worldwide by companies and businesses around the world. This also pushed us to an era of cloud computing and extensive use of Artificial Intelligence or automation.

3. Recent Evolution in Business Analytics

The recent evolution that business analytics has experienced can be fundamentally traced back to the introduction of automation in analytics and the concept of big data. The advent of big data meant that analytics along with various data sources should become more scalable and more powerful. This helped in introducing more advanced tools and systems that are compatible with large volumes of data. The emergence of cloud technologies also meant that data did not need to be on-site. There was also a huge demand for automating analytical tools by this time due to the massive amount of data that needed to be worked upon.

There have been four main spheres where business analytics has evolved greatly, these are:
- Artificial Intelligence and Automated Analytics

- Predictive Analytics

- Real-time Analytics

- Big Data

ANALYTIX LABS

Recent Evolution In Business Analytics

There Have Been Four Main Spheres Where Business Analytics Has Evolved Greatly.

Artificial Intelligence & Automated Analytics   Predictive Analytics   Real-time Analytics   Big Data

Process

Let's look at the business analytics process:

Define Business Requirements

Determine the problem and how to address it. Break the problem down into smaller goals to tackle

them individually. Gather relevant data required to solve issues. Answer key questions like:
- What data is available?
- Is it sufficient?
- How can we leverage it?

Explore Information

Clean datasets by imputing missing values, eliminating outliers, removing duplicates and

combining variables. Plot time series graphs to determine patterns and inconsistencies. Use scatter

plots to discover correlations between variables. Summarize information with the help of

appropriate visuals for clear insights.

Conduct Analysis

Leverage statistical analysis methods like regression, correlation and hypothesis testing to perform advanced analysis. Regression analysis estimates the relationship between dependent and independent variables. Hypothesis testing is a statistical inference technique that draws conclusions from sample data about the entire population.

Make Accurate Predictions

Leverage predictive analytics techniques like decision trees, neural networks and logistic regression to unearth insights and highlight relationships between variables. Compute predictive errors by comparing actual values with the predicted ones. Select the highest-performing model based on accuracy and outcomes.

Select Optimal Solutions

Run what-if scenarios to determine plausible solutions with the given constraints. Select the optimal solution based on minimum error, model coefficients and business objectives.

6. Why Business Analytics Is Essential to Build a Competitive Business? Explain the competitive advantage of Business analytics.

Here are some reasons why Business Intelligence and analytics are essential for a successful business:

- Business analytics is very helpful in reducing risk in business operations.

- It increases revenue and helps companies churn out more profit.

- Analytics allow companies to make better and more informed business decisions that are data-driven.

- It helps increase the operational efficiency of projects or processes.

- It also helps in effectively using resources such as human assets.

- It helps in reducing wastage and in cutting operational costs.

- Business analytics directly allows companies to stay ahead of their competition and outperform them.

- Analytics optimises processes and helps in business process automation.

- It also allows companies to replicate successful results and understand how the success was achieved.

- It helps in forecasting and prediction of outcomes.

- Business analysts help monitor performance and evaluate operations through identifiable metrics or Key Performance Indicators(KPI).

- It helps in the identification of anomalies and factors that affect market and customer behaviour.

- Real-time business analytics helps in taking rapid data-centric decisions.

- It helps in improving the sustainability of businesses and projects in the long run.

Competitive advantage is what distinguishes a company's goods or services from all other options available to a customer. It refers to the elements that enable a company to manufacture goods or services more efficiently or at a lower cost than its competitors. This may result in the company gaining a large market share. These elements enable the producing unit to earn higher sales or higher margins than its competitors. A range of elements contributes to competitive advantages, including cost structure, branding, product quality, distribution network, intellectual property, and customer service improving the company's customer base.

- Business analytics for organizations is becoming a competitive advantage and is now necessary to apply business analytics, particularly its subset of predictive business analytics. When business analytics initiatives are adopted correctly businesses are guaranteed success. The use of business analytics is a skill that is gaining mainstream value due to the increasingly thinner margin for decision error.
- Business analytics enables differentiation through the various analytics models. It is primarily about driving change through analytics priorities. Business analytics drives competitive advantage by generating economies of scale, economies of scope, and quality improvement. Taking advantage of economies of scale is the first way organizations achieve comparative cost efficiencies and drive competitive advantage against their peers.

- Business analytics improves the efficiency of business operations providing business owners with valuable information about the performance of the business. The efficiencies that accumulate when a firm embraces big data technology eventually contribute to a ripple effect of increased production and reduced business costs.

- Analytics gives companies insight into their customers' behaviour and needs. It also makes it possible for a company to understand its brand's public opinion, follow the results of various marketing campaigns, and strategize how to create a better marketing strategy to nurture long and fruitful relationships with its customers.

UNIT-II

1. What are the skills required for business analytics personnel?

Top Business Analyst Skills

A business analyst must perform multiple duties; hence they need to have a strong skill set that is an amalgam of technical and non-technical skills.

Here, we will take you through the top ten vital business analyst skills that are a dire necessity.

1. Understanding the Business Objective

- A business analyst should be able to comprehend an organization's goals and problems.
- It requires them to recognize business problems and come up with the most suited solution.
- It is good if business analysts have domain knowledge in the organization they are working in. This will help them with the required deliverables.
- In most cases, business analysts work towards enabling a change with the motive of increasing sales, scale-up production, improving revenue streams, etc.

Understanding the business motive is the first step in a business analyst's life; it is a special skill by itself. Now, let's proceed to the next set of business analyst skills.

2. Analytical and Critical Thinking

As the famous quote by Thomas Alva. Edison goes:

"Five percent of the people think;

ten percent of the people think they think;

and the other eighty-five percent would rather die than think."

This shows that, although thinking sounds basic, it is an underrated ability. Analytical and critical thinking is one of the core business analyst skills.

- A business analyst must analyze and translate the client's requirements distinctly.
- Critical thinking helps a business analyst in assessing multiple options before arriving at the aspired solution.
- Business analysts focus on gathering and understanding the client's needs. Critical thinking enables them to prioritize business requirements.
- A good analytical bend of mind will help a business analyst reach the stated goals even when there is a limitation in the resources and the conditions are nonideal.

3. Communication and Interpersonal Skills

The next skill in our list of business analyst skills are commonly heard of skills - communication and interpersonal skills.

- Being understood is as important as understanding. You should have the ability to communicate concisely with the stakeholders and clients with regard to the requirements.
- A business analyst uses communication and interpersonal skills at different phases, for example: when a project is being launched, while collecting requirements, when collaborating with stakeholders, while validating the final solution, and so on.
- Business Analysts use verbal and written communication to convey ideas, facts, and opinions to stakeholders.
- Good communication and interpersonal skills will give confidence to a business analyst while facilitating meetings.

4. Negotiation and Cost-Benefit Analysis

Needless to say, negotiation is a crucial skill every business analyst must-have. The next skill on our list of business analyst skills is the negotiation and cost-benefit analysis.

- Business analysts negotiate at every project phase. At the initial stage of a project, negotiation skills are used to decide what they must include in the project's vision.
- Business analysts then use their negotiation skills to determine which requests turn into requirements and their priority levels.
- As the project progresses, negotiation skills play a significant role in deciding the functional design that fulfills the requirements. Negotiation skills are also used to make technical decisions.
- Business analysts carry out a cost-benefit analysis to assess the costs and benefits expected in a project. When organizations undertake new projects, business analysts make use of cost-benefit analysis to establish if they should embark on those particular projects.

5. Decision-Making Skills

At number five, we have another non-technical skill, and that is decision-making skills.

- The decisions made by a business analyst has a direct and indirect impact on the company's business. Hence, they should think of all the aspects before presenting their decision.
- Before making a decision, a business analyst interprets the problem and finds alternative business approaches.
- They then test all the alternative approaches and make a decision based on their thoughts regarding these approaches. They finally test and implement the solution.

- Business analysts also take the last call in ensuring that a particular technical design conforms to the discussed business requirements or not.

We will now take you through a few technical skills in our list of business analyst skills.

6. Programming Languages

- Business analysts should have hands-on programming knowledge in order to perform quicker and better data analysis.
- Knowledge of R and Python is extremely beneficial. Complex problems can be solved by writing efficient codes.
- R and Python comprise several libraries and packages for data wrangling, data manipulation, data visualization, and analytics. Additionally, a sound understanding of statistical software like SAS and SPSS is recommended.
- With the help of the above programming languages, massive data can be analyzed and visualized finely. Also, business models can be created for making business predictions.

7. Creation of Reports and Dashboards

The next vital skill we have is the creation of reports and dashboards.

- Business analysts must be proficient in using various business intelligence tools for creating reports and dashboards.
- Business analysts develop general reports and dashboard reports to solve decision-making problems.

8. Database and SQL

The next skill every business analyst should have is the knowledge of database and SQL.

- Business analysts most often work with structured data. In order to store and process this heavy data, they should have a sound understanding of relational databases like Microsoft SQL Server, MySQL database, Oracle DB, as well as NoSQL databases.
- Every business analyst must carry hands-on experience with SQL. This will help them access, retrieve, manipulate, and analyze data.
- They must write data definition and data manipulation commands like create, delete, select, update, insert, etc.

9. Microsoft Excel

Next, in our list of business analyst skills is knowledge of Microsoft Excel. This is a fundamental skill that every business analyst must-have.

- Excel is one of the oldest and strongest analytics and reporting tools; business analysts use it to perform several calculations, data, and budget analysis to unravel business patterns.

- They summarize data by creating pivot tables. They make different charts using Excel to generate dynamic reports related to a business problem.
- Excel is used to create revenue growth models for new products based on recent customer forecasts, plan an editorial calendar, list expenses for products, and create charts to show how close the product is to budget across each category.
- Business analysts use Excel to calculate customer discounts based on monthly purchase volume by product. They even summarize customer revenue by product to find areas where there is a need to build stronger customer relationships.

10. Documentation and Presentation

Last, in the list of business analyst skills, we have documentation and presentation.

- A business analyst should document their project teachings and results very well, clearly, and concisely.
- They should confidently present their project findings and outcomes in front of the stakeholders and clients. With the help of organized documentation, business analysts can communicate technical concepts easily to non-technical employees.
- Jotting down project lessons is vital, as this will help them make better decisions in the future.

2. Explain the role and responsibilities of business analytics personnel.

Roles of Business Analyst

The business analysis goes through phases, and there are four levels that the analysts follow to get the best results. These levels are:

- Strategy Management: At this level, the business analyst plans and determines the strategies to analyze the business elements. It is the most significant factor as these strategies decide whether the final results are apt.
- Business Model Analysis: In this phase, the professionals analyze the existing business model, checking where the system is lacking.
- Process Designing: At this level, the analysts design the processes they need to introduce into the system to make necessary improvements.
- Technology Analysis: At this level, the business analyst responsibilities go through the technical analysis to ensure the feasibility of the change. It is another crucial aspect that you should thoroughly check.

Business Analyst Responsibilities

- Domain Knowledge
- Getting signoff from the stake holder
- Knowledge transfer to the developers
- Functional testing or verify the solution developed by the developement
- Support after implementation
- Track the requirements
- Documentation
- Provide solutions to the requirement
- GAP Analysis
- Requirements Elicitation
- Requirements Gathering

The BA roles and responsibilities are all about doing the tasks required to complete these levels. Some of the roles of these experts are listed below. Refer to the business analyst role description to get better clarity and plan your career accordingly.

- **Sorting Business Needs**

The first BA role in agile practices is to sort the requirements. It would need a detailed analysis of the existing business model and the operations. Moreover, the expert should compare it with the market advancements to identify areas with scope for improvement.

- **Meeting stakeholders & development team**

Another role of the business analyst working on a particular project is to meet the stakeholders and the development teams. First, they need to convey the ideas to the stakeholders, and once they get the nod, the next step is to transfer the changes to the development teams. All these tasks should be completed in a seamless manner to yield positive results.

- **Determining Feasibilities**

Once the business analyst discusses the plans with the concerned teams, the task is determining the feasibilities. The developers and other concerned teams will discuss the feasibility, and if the need persists, the experts may have to change the plan.

- **Present the Plan**

After the plan is ready to be implemented, the next thing is to present it to the staff and other people involved in the change. The team will gain familiarity with the changes and may need

the training to adapt to them. It is again the role of a business analyst to present that plan and conduct essential pieces of training.

- **Project Implementation**

Successful implementation is another role that business analysts have to fulfill. They assign tasks to teams concerned about infrastructural changes or technical advancements. The entire implementation process happens under these experts, whose strategies are under question.

- **Testing**

The work does not end with the implementation phase. The analysts also have to stay put during the testing phase, ensuring that everything happens according to the expectations. If there is a technical advancement involved, the software testing team will generate reports to confirm that the result of these changes is working in favor.

- **Problem Solving**

It is another role or responsibility that business analysts have to fulfill. They must identify the problem areas in a business process and find apt solutions. So, their analytics skill will help find the root cause behind the issue and the right way to solve them.

- **Maintenance**

When the organizational processes and operations are on the right track and yielding the best results, the role left is to maintain that pace. The analysts will keep track of the tasks and operations to ensure no further changes are required. If they find bugs or defects in the system, they report it to the concerned team and get it fixed soon.

Key Responsibilities of a Business Analyst

Now you know all the roles of a business analyst. Let us understand the responsibilities that come on their shoulders. It will tell you what a business analyst does, and if you resonate with them, you can plan your career in this domain.

- Get all the information about the current processes that the business follows. The idea behind the process analysis is to identify areas that need improvement.
- Dig in depth and research the market thoroughly to understand what other people in a similar domain are doing. It will help you see where you are behind.
- Get up-to-date information and data related to the organization to create research-based plans for company growth. The idea is to ensure the strategies that you prepare are well-researched and fact-based.

- Creating a presentation, providing details about the required improvements and how to execute them. Moreover, the business analyst would present it in front of the board, staff, and every person directly involved with the change.
- Organize staff training to ensure everyone is ready to adapt to the changes and incorporate them into the system smoothly.
- They convince the stakeholders to nod yes to the changes and work with the senior management and other professionals.
- Test the changes to ensure that strategies work fine and give the expected results.

3. Discuss the concept of business analytics data in detail.
Key concepts in business analytics

Business analytics is a field that aims to use data and computational techniques to help businesses make better decisions. The field encompasses various topics, from analyzing large datasets and developing predictive models to providing insights into consumer behavior and market trends. In this section, we'll discuss some key business analytics concepts.

- **Business Statistics**

Business statistics is the study of numbers businesses use to make decisions. These numbers are typically collected to find patterns, trends, and relationships among them. Statistics is one of the most important tools a business can use to improve its operations and make better decisions.

There are several types of business statistics:
1. Financial statements (including cash flow statements and balance sheets)
2. Demographic information (age distribution and gender-wise breakdown of customers)
3. Consumer purchase behaviors (how much time people spend on different websites when they are searching for something)

There are three broad areas of business statistics:
1. Descriptive statistics include summarizing data in various ways.
2. Inferential statistics use statistical tests and business analytics techniques to infer conclusions about unknown variables.
3. Experimental statistical methods use randomization procedures or experiments designed to answer questions about likelihoods or probabilities.

Want to know the different career prospects in business analytics? Check out our detailed guide on business analytics careers for more information.

- **Business Econometrics**

Business econometrics is a field of economics that uses statistical techniques to study how companies make decisions. It's often used to develop models that can be used to predict how a company will perform in the future or to help companies make better decisions about their operations.

Business econometrics can be applied to any business, from large multinational corporations to small start-ups with just one employee. It has applications in areas like marketing, pricing, and financial management.

- **Time Series Analysis**

Time series analysis is a technique that uses past data to predict future events. It's used in business analytics to study trends, patterns, and changes over time.

Time series analysis is used as a way to understand how markets work and make predictions about what will happen next. For example, it can be used by financial analysts to forecast the value of stocks based on their current price and other factors such as demand and supply.

Want to know the latest trends in business analytics? Check out our detailed blog on business analytics trends for more information.

- **Business Intelligence**

Business intelligence means collecting, analyzing, and interpreting data for better business decisions. It involves using analytics as a business intelligence tool to solve business problems. Business intelligence aims to provide an overview of the data and its relationships so that information can be used in marketing, operations, sales, or customer service. It involves gathering, storing, and analyzing data to create reports that describe key performance indicators (KPIs) for managers.

- **Predictive Analytics**

Predictive analytics is a field of data science that uses statistical models to make predictions. It can predict the performance of a company's products or services or customer behavior.

Predictive analysis techniques aim to identify patterns in large sets of data, which can then be used to make predictions about future events or outcomes. Predictive analytics has many applications in business and finance, such as predicting sales trends or identifying fraudulent banking transactions.

- **Data Visualization**

Data visualization means using data to visualize information so people can understand it. This can be done through graphs, charts, maps, and other visual data displays. Data visualization in business helps you see patterns in your data and make connections between different pieces of information.

Some common techniques used when creating visual displays include the following:

1. Bar charts
2. Line graphs
3. Pie charts
4. Scatter plots
5. Quantile regression
6. Frequency distributions
7. Pictogram

- **Business Analytics Dashboard**

Business analytics dashboards are one of the business intelligence tools used by businesses to visualize the data being collected. A business analytics dashboard aims to show the business owner how their company is doing and what they need to do to improve.

The features of a business analytics dashboard include the following:

1. Data visualizations show how many customers have been acquired and lost in different periods or how much revenue has increased or decreased over time.
2. Interactive charts, graphs, and tables allow users to drill down into specific areas of interest.
3. A chronological timeline shows when certain events occurred to determine correlation or subsequent outcomes (such as increases or decreases in sales).

Want to know the types of business analyst roles? Check out our blog on <u>business analyst types</u> for more information.

- **Marketing Analytics**

<u>Marketing Analytics</u> uses statistical analysis to understand and improve marketing campaigns. It can be used for market segmentation, branding, pricing and promotion strategies, and customer acquisition.

Marketing analytics consists of four main areas:

1. Market Segmentation
2. Market Forecasting
3. Demand Generation
4. Customer Relationship Management (CRM)

- **Digital and Web Analytics**

Digital and Web Analytics is a method of tracking user behavior on a website and using that information to improve the user experience.

Web analytics tools are used by businesses to measure the performance of their websites, including how many people visit them, what they do there, and what they click on. These tools allow you to analyze how users interact with your site, which can help you to improve your site's layout and functionality.

For these tools to work properly, you must first decide on an appropriate set of variables that will be tracked by your tool. These variables can include the following things:

1. User actions (whether or not someone clicked on something)
2. Pages viewed (this can be broken down into individual pages)
3. Links clicked (this can be broken down into individual pages)

Want to know the essential tools used in business analytics? Check out our detailed guide on <u>business analytics tools</u> for more information.

- **Financial Analytics**

Financial analytics is applying data and other information to evaluate and make decisions in a financial context. It can help companies improve their decision-making, understand market trends, manage risk, collect revenue, manage costs, and optimize operations.

4. Explain the business analytics technology.

Here are some key technologies that enable Big Data for Businesses:

1) Predictive Analytics

One of the prime tools for businesses to avoid risks in decision making, <u>predictive analytics can help businesses</u>. Predictive analytics hardware and software solutions can be utilised for discovery, evaluation and deployment of predictive scenarios by processing big data. Such data can help companies to be prepared for what is to come and help solve problems by analyzing and understanding them.

2) NoSQL Databases

These databases are utilised for reliable and efficient data management across a scalable number of storage nodes. NoSQL databases store data as relational database tables, JSON docs or key-value pairings.

3) Knowledge Discovery Tools

These are tools that allow businesses to mine big data (structured and unstructured) which is stored on multiple sources. These sources can be different file systems, APIs, DBMS or similar platforms. With search and knowledge discovery tools, businesses can isolate and utilise the information to their benefit.

4) Stream Analytics

Sometimes the data an organisation needs to process can be stored on multiple platforms and in multiple formats. Stream analytics software is highly useful for filtering, aggregation, and analysis of such big data. Stream analytics also allows connection to external data sources and their integration into the application flow.

5) In-memory Data Fabric

This technology helps in distribution of large quantities of data across system resources such as Dynamic RAM, Flash Storage or Solid State Storage Drives. Which in turn enables low latency access and processing of big data on the connected nodes.

6) Distributed Storage

A way to counter independent node failures and loss or corruption of big data sources, distributed file stores contain replicated data. Sometimes the data is also replicated for low latency quick access on large computer networks. These are generally non-relational databases.

7) Data Virtualization

It enables applications to retrieve data without implementing technical restrictions such as data formats, the physical location of data, etc. Used by Apache Hadoop and other distributed

data stores for real-time or near real-time access to data stored on various platforms, data virtualization is one of the most used big data technologies.

8) Data Integration

A key operational challenge for most organizations handling big data is to process terabytes (or petabytes) of data in a way that can be useful for customer deliverables. Data integration tools allow businesses to streamline data across a number of big data solutions such as Amazon EMR, Apache Hive, Apache Pig, Apache Spark, Hadoop, MapReduce, MongoDB and Couchbase.

9) Data Preprocessing

These software solutions are used for manipulation of data into a format that is consistent and can be used for further analysis. The data preparation tools accelerate the data sharing process by formatting and cleansing unstructured data sets. A limitation of data preprocessing is that all its tasks cannot be automated and require human oversight, which can be tedious and time-consuming.

10) Data Quality

An important parameter for big data processing is the data quality. The data quality software can conduct cleansing and enrichment of large data sets by utilising parallel processing. These softwares are widely used for getting consistent and reliable outputs from big data processing.

5. Write the note on i) Managing information policy ii) Data quality iii) Managing change in Business Analytics

Information management policy

An information management policy provides the authoritative statement of the principles for the management of the public inquiry records. It need not be a lengthy document, but it is a statement of commitment by the inquiry to manage information appropriately so as to comply with information legislation, meet the ongoing needs of the inquiry team and deliver a well-ordered and comprehensive record at the end of the inquiry process.

It is important to establish a successful information management policy from the outset of an inquiry, as it will act as a mandate for supporting processes and procedures, while demonstrating the value of these to inquiry staff.

An information management policy is particularly vital as it will help inquiry staff to manage digital information in accordance with best practice. This is because in a digital environment it is more difficult to ensure that records remain complete, authentic and accessible.

The policy should highlight to staff how good information management can support day-to-day operations, ensure that the inquiry meets its statutory obligations, and deliver time and money efficiencies at the end of the inquiry process.

The National Archives recommends that the policy includes the following points:

- roles and responsibilities for the management of the inquiry's records should be assigned to key officials within the public inquiry

- business processes, decisions, activities and transactions need to be recorded to form a comprehensive record of the inquiry

- information needs to be managed in such a way that it is accessible and meaningful, in the right format, to those who need to use it

- requirements for the management of all digital information created by the inquiry should be recorded (including scanned evidence, the inquiry website and electronic mail)

- any record that the public inquiry creates (including public records tabled as evidence) is Crown property and therefore subject to the Public Records Act 1958

- principles for the selection of records for permanent preservation needs to be defined including the type of information that can be disposed of

- sensitive information needs to be managed in accordance with best practice and in a manner that facilitates an efficient review of selected records prior to transfer

Data Quality

Data quality analysts are responsible for conducting data quality assessments, which involve assessing and interpreting every quality data metric. Then, the analyst creates an aggregate score reflecting the data's overall quality and gives the organization a percentage rating that shows how accurate the data is.

To put the definition in more direct terms, data quality indicates how good the data is and how useful it is for the task at hand. But the term also refers to planning, implementing, and controlling the activities that apply the needed quality management practices and techniques required to ensure the data is actionable and valuable to the data consumers.

Now, let us look at data quality dimensions after you better understand what is data quality.
Data Quality Dimensions

There are six primary, or core, dimensions to data quality. These are the metrics analysts use to determine the data's viability and its usefulness to the people who need it.

- Accuracy

The data must conform to actual, real-world scenarios and reflect real-world objects and events. Analysts should use verifiable sources to confirm the measure of accuracy, determined by how close the values jibe with the verified correct information sources.

- Completeness

Completeness measures the data's ability to deliver all the mandatory values that are available successfully.

- Consistency

Data consistency describes the data's uniformity as it moves across applications and networks and when it comes from multiple sources. Consistency also means that the same datasets stored in different locations should be the same and not conflict. Note that consistent data can still be wrong.

- Timeliness

Timely data is information that is readily available whenever it's needed. This dimension also covers keeping the data current; data should undergo real-time updates to ensure that it is always available and accessible.
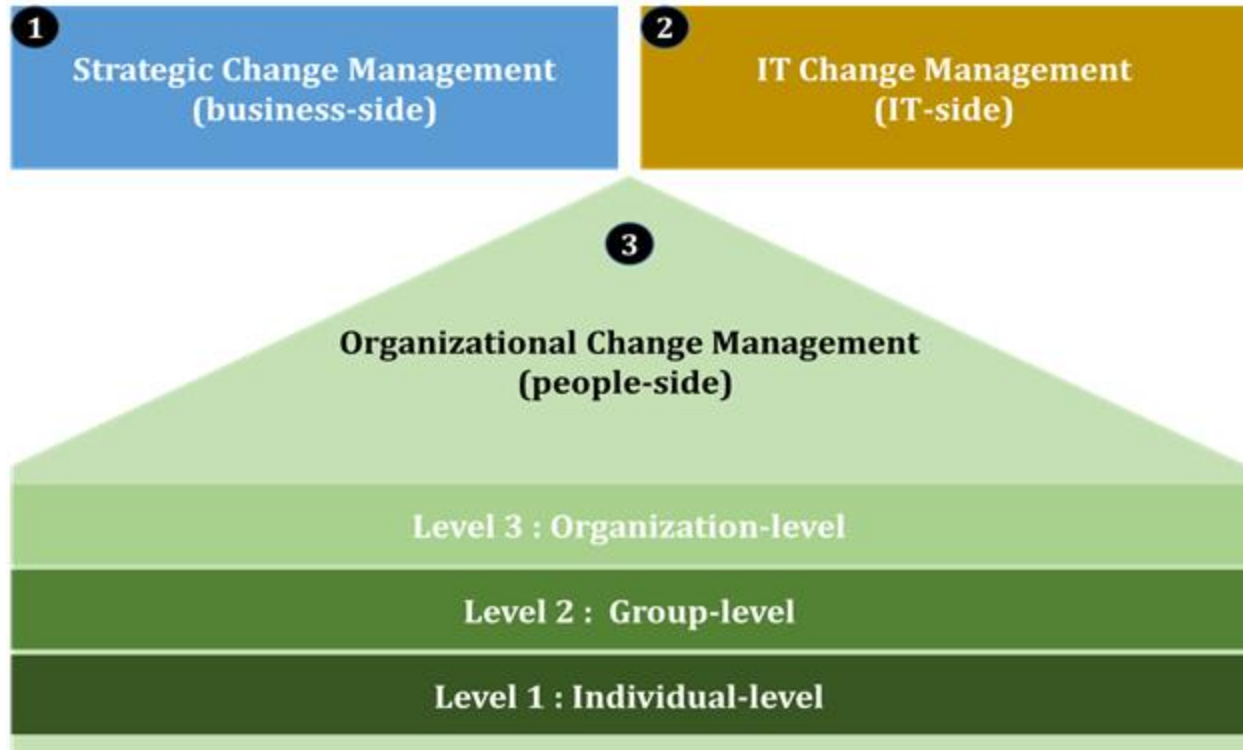
- Uniqueness

Uniqueness means that no duplications or redundant information are overlapping across all the datasets. No record in the dataset exists multiple times. Analysts use data cleansing and deduplication to help address a low uniqueness score.

- Validity

Data must be collected according to the organization's defined business rules and parameters. The information should also conform to the correct, accepted formats, and all dataset values should fall within the proper range.

Managing change in Business Analytics

1 Strategic Change Management (business-side)

2 IT Change Management (IT-side)

3 Organizational Change Management (people-side)

Level 3 : Organization-level

Level 2 : Group-level

Level 1 : Individual-level

Change is hard for most people. There are a variety of reasons why change is hard, from our inherent need for a sense of security to having to deal with too much change at once to not following a process to increase the change's likelihood of success. I know I am personally not looking forward to having to adapt my hyper-productive processes when Windows 8 is released and I may have to re-learn or find new ways to do things efficiently. As Business Analysts, we are often involved in projects or initiatives that cause a great deal of change within an organization. In some cases we are put on the front lines of the change, whether it is gathering requirements from skeptical stakeholders to supporting the review of a solution that was put in place too quickly and is now meeting strong resistance.

6. Explain the different types of organizational structures for business analytics.

The organizational structure determines how information flows within an organization and how decisions are made. It defines roles, responsibilities, reporting lines, and communication pathways. A well-defined and aligned structure is crucial for creating an environment that fosters effective business analysis.

When it comes to business analysis success, the following aspects of organizational structure are particularly relevant:

- **Reporting Structure:** The reporting structure determines the line of communication between business analysts and other stakeholders. A clear reporting structure helps business analysts gather requirements efficiently and collaborate with key decision-makers.
- **Departmentalization:** The way departments are organized impacts business analysts' access to necessary information and subject matter experts. An optimal departmentalization structure ensures business analysts can easily connect with the right people to gather insights and analyze data effectively.
- **Authority and Decision-Making:** The level of authority and decision-making power given to business analysts within the organizational structure directly influences their ability to drive change and implement recommendations.

Now, let's explore how different organizational structures can affect business analysis success.

Functional Organizational Structure

A functional organizational structure groups employees by their specific functions, such as finance, operations, or marketing. While this structure provides a clear reporting line, it can pose challenges to business analysts:

- **Limited Collaboration:** Business analysts in functional organizational structures may find it challenging to collaborate with teams in other functions, hindering their ability to understand the holistic business context.
- **Departmental Silos:** Silos can develop among functional departments, making it difficult for business analysts to collect accurate and comprehensive data.

However, leveraging their reporting line, business analysts can gather insights within their function, ensuring accurate and relevant analysis.

Matrix Organizational Structure

A matrix organizational structure combines functional and project-based reporting lines. This structure offers advantages for business analysis:

- **Enhanced Collaboration:** Matrix structures encourage collaboration across functions, enabling business analysts to access diverse perspectives and data that support comprehensive analysis.
- **Multiple Reporting Lines:** Business analysts have multiple reporting lines, allowing them to seek guidance and input from various stakeholders, ensuring a more complete understanding of the business.

However, business analysts in matrix structures may face challenges in managing conflicting priorities and expectations arising from dual reporting lines.

Flat Organizational Structure

In a flat organizational structure, there are few hierarchical levels, and decision-making authority is decentralized. This structure can greatly benefit business analysis efforts:

- **Direct Communication:** Business analysts can interact directly with key decision-makers, eliminating delays and increasing the speed of analysis and decision-making.
- **A Culture of Collaboration:** Flat structures foster a collaborative culture, encouraging cross-functional collaboration and knowledge sharing among business analysts.

However, the lack of clear reporting lines and project management frameworks can lead to confusion and conflicts if not well-managed.

Understanding the relationship between organizational structure and business analysis success is crucial for organizations aiming to optimize their analysis capabilities. Here are some key takeaways:

- A well-defined reporting structure enables efficient communication and collaboration between business analysts and stakeholders.
- Optimal departmentalization facilitates effective data gathering and analysis by providing easy access to subject matter experts.
- The level of authority and decision-making power given to business analysts empowers them to drive change and implement recommendations.

- Functional structures can hinder collaboration and result in departmental silos, while matrix structures enhance knowledge sharing and collaboration.
- Flat structures facilitate direct communication and foster a collaborative culture, allowing for quicker decision-making and improved analysis outcomes.

By recognizing the impact of organizational structure on business analysis success, organizations can adapt their structures to optimize analysis capabilities, improve decision-making, and drive overall business growth.

Exploring How Organizational Structure Affects Business Analysis Outcomes

Understanding Organizational Structures

Before delving into the impact of organizational structures on business analysis outcomes, let's briefly explore various common structures:

- **Functional Structure:** Typically found in smaller organizations, this structure groups employees based on their specific functions or areas of expertise.
- **Divisional Structure:** Common in larger companies, this structure divides the organization into smaller divisions based on products, services, or geographic locations.
- **Matrix Structure:** Combining elements of both functional and divisional structures, this organizational form fosters collaboration by creating cross-functional teams.
- **Flat Structure:** Encouraging minimal hierarchy, this structure promotes open communication and empowers employees to make decisions.
- **Network Structure:** Common in modern organizations, this structure emphasizes collaboration and partnerships with external entities.

The Impact of Organizational Structure on Business Analysis

Now that we have a foundation in organizational structures, we can explore how they affect business analysis outcomes:

1. Communication Channels

The structure of an organization plays a critical role in establishing communication channels. Effective communication is vital for successful business analysis, as it enables seamless information flow between different stakeholders.

- The choice of organizational structure can either facilitate or hinder communication, impacting the accuracy and timeliness of information provided during the analysis process.

2. Decision-Making Authority

Organizational structures determine how decisions are made within an organization. This authority allocation affects the decision-making process during business analysis, as it determines who has the power to implement recommended changes.

- Organizations with decentralized decision-making authority may experience smoother implementation of analysis outcomes, as decision-makers are closer to the problem.

3. Cross-Functional Collaboration

Business analysis often requires collaboration across different departments and functions. The structure of an organization influences the ease of this collaboration and the level of cooperation between teams.

- Organizations with a matrix or network structure might have an advantage in terms of fostering collaboration, enhancing the accuracy and comprehensiveness of business analysis outcomes.

4. Data Accessibility

Easy accessibility to relevant data is crucial for accurate business analysis. Organizational structures can impact data flow and accessibility, depending on how information is stored and shared within the organization.

- Organizations with a well-established data sharing infrastructure, such as a centralized database, are likely to yield more accurate and comprehensive business analysis outcomes.

The Path to Optimized Business Analysis Outcomes

To ensure the best possible business analysis outcomes, organizations should consider the following:

- 1. Evaluate the existing organizational structure and identify any potential barriers to effective business analysis.
- 2. Foster a culture of open communication and knowledge sharing to improve the accuracy and timeliness of information during the analysis process.
- 3. Incorporate cross-functional collaboration mechanisms, such as forming temporary project teams or utilizing digital collaboration tools.
- 4. Invest in robust data management systems that streamline data accessibility and ensure its integrity.
- 5. Continually reassess and adapt the organizational structure as the business evolves to optimize business analysis outcomes.

UNIT-III

1.Define sampling. Explain the methods of sampling.

Sampling Methods:

Sampling is a technique of selecting individual members or a subset of the population to make statistical inferences from them and estimate the characteristics of the whole population. Different sampling methods are widely used by researchers in market research so that they do not need to research the entire population to collect actionable insights.

It is also a time-convenient and cost-effective method and hence forms the basis of any research design. Sampling techniques can be used in research survey software for optimum derivation.

For example, suppose a drug manufacturer would like to research the adverse side effects of a drug on the country's population. In that case, it is almost impossible to conduct a research study that involves everyone. In this case, the researcher decides on a sample of people from each demographic and then researches them, giving him/her indicative feedback on the drug's behavior.

Learn more about Audience by QuestionPro

Types of sampling: sampling methods
Sampling in market action research is of two types – probability sampling and non-probability sampling. Let's take a closer look at these two methods of sampling.

1. **Probability sampling:** Probability sampling is a sampling technique where a researcher selects a few criteria and chooses members of a population randomly. All the members have an equal opportunity to participate in the sample with this selection parameter.
2. **Non-probability sampling:** In non-probability sampling, the researcher randomly chooses members for research. This sampling method is not a fixed or predefined selection process. This makes it difficult for all population elements to have equal opportunities to be included in a sample.

This blog discusses the various probability and non-probability sampling methods you can implement in any market research study.

Types of probability sampling with examples:
Probability sampling is a technique in which researchers choose samples from a larger population based on the theory of probability. This sampling method considers every member of the population and forms samples based on a fixed process.
For example, in a population of 1000 members, every member will have a 1/1000 chance of being selected to be a part of a sample. Probability sampling eliminates sampling bias in the population and allows all members to be included in the sample.
There are four types of probability sampling techniques:

- **Simple random sampling:** One of the best probability sampling techniques that helps in saving time and resources is the Simple Random Sampling method. It is a reliable method of obtaining information where every single member of a population is chosen randomly, merely by chance. Each individual has the same probability of being chosen to be a part of a sample.
  For example, in an organization of 500 employees, if the HR team decides on conducting team-building activities, they would likely prefer picking chits out of a bowl. In this case, each of the 500 employees has an equal opportunity of being selected.
- **Cluster sampling:** Cluster sampling is a method where the researchers divide the entire population into sections or clusters representing a population. Clusters are identified and included in a sample based on demographic parameters like age, sex, location, etc. This makes it very simple for a survey creator to derive effective inferences from the feedback.
  For example, suppose the United States government wishes to evaluate the number of immigrants living in the Mainland US. In that case, they can divide it into clusters based on states such as California, Texas, Florida, Massachusetts, Colorado, Hawaii, etc. This way of conducting a survey will be more effective as the results will be organized into states and provide insightful immigration data.
- **Systematic sampling:** Researchers use the systematic sampling method to choose the sample members of a population at regular intervals. It requires selecting a starting point for the sample and sample size determination that can be repeated at regular intervals. This type of sampling method has a predefined range; hence, this sampling technique is

the   least   time-consuming.
For example, a researcher intends to collect a systematic sample of 500 people in a population of 5000. He/she numbers each element of the population from 1-5000 and will choose every 10th individual to be a part of the sample (Total population/ Sample Size = 5000/500 = 10).

- **Stratified random sampling:** Stratified random sampling is a method in which the researcher divides the population into smaller groups that don't overlap but represent the entire population. While sampling, these groups can be organized, and then draw a sample   from   each   group   separately. For example, a researcher looking to analyze the characteristics of people belonging to different annual income divisions will create strata (groups) according to the annual family income. Eg – less than $20,000, $21,000 – $30,000, $31,000 to $40,000, $41,000 to $50,000, etc. By doing this, the researcher concludes the characteristics of people belonging to different income groups. Marketers can analyze which income groups to target and which ones to eliminate to create a roadmap that would bear fruitful results.

Uses of probability sampling
There are multiple uses of probability sampling:

- **Reduce Sample Bias:** Using the probability sampling method, the research bias in the sample derived from a population is negligible to non-existent. The sample selection mainly depicts the researcher's understanding and inference. Probability sampling leads to higher-quality data collection as the sample appropriately represents the population.
- **Diverse Population:** When the population is vast and diverse, it is essential to have adequate representation so that the data is not skewed toward one demographic. For example, suppose Square would like to understand the people that could make their point-of-sale devices. In that case, a survey conducted from a sample of people across the US from different industries and socio-economic backgrounds helps.
- **Create an Accurate Sample:** Probability sampling helps the researchers plan and create an accurate sample. This helps to obtain well-defined data.

Types of non-probability sampling with examples
The non-probability method is a sampling method that involves a collection of feedback based on a researcher or statistician's sample selection capabilities and not on a fixed selection process. In most situations, the output of a survey conducted with a non-probable sample leads to skewed results, which may not represent the desired target population. But, there are situations, such as the preliminary stages of research or cost constraints for conducting research, where non-probability sampling will be much more useful than the other type.
Four types of non-probability sampling explain the purpose of this sampling method in a better manner:

- **Convenience sampling:** This method depends on the ease of access to subjects such as surveying customers at a mall or passers-by on a busy street. It is usually termed as convenience sampling because of the researcher's ease of carrying it out and getting in touch with the subjects. Researchers have nearly no authority to select the sample elements, and it's purely done based on proximity and not representativeness. This non-probability sampling method is used when there are time and cost limitations in collecting feedback. In situations with resource limitations, such as the initial stages of

research, convenience sampling is used. For example, startups and NGOs usually conduct convenience sampling at a mall to distribute leaflets of upcoming events or promotion of a cause – they do that by standing at the mall entrance and giving out pamphlets randomly.

- **Judgmental or purposive sampling:** Judgmental or purposive samples are formed at the researcher's discretion. Researchers purely consider the purpose of the study, along with the understanding of the target audience. For instance, when researchers want to understand the thought process of people interested in studying for their master's degree. The selection criteria will be: "Are you interested in doing your masters in …?" and those who respond with a "No" are excluded from the sample.
- **Snowball sampling:** Snowball sampling is a sampling method that researchers apply when the subjects are difficult to trace. For example, surveying shelterless people or illegal immigrants will be extremely challenging. In such cases, using the snowball theory, researchers can track a few categories to interview and derive results. Researchers also implement this sampling method when the topic is highly sensitive and not openly discussed—for example, surveys to gather information about HIV Aids. Not many victims will readily respond to the questions. Still, researchers can contact people they might know or volunteers associated with the cause to get in touch with the victims and collect information.
- **Quota sampling:** In Quota sampling, members in this sampling technique selection happens based on a pre-set standard. In this case, as a sample is formed based on specific attributes, the created sample will have the same qualities found in the total population. It is a rapid method of collecting samples.

Uses of non-probability sampling
Non-probability sampling is used for the following:

- **Create a hypothesis:** Researchers use the non-probability sampling method to create an assumption when limited to no prior information is available. This method helps with the immediate return of data and builds a base for further research.
- **Exploratory research:** Researchers use this sampling technique widely when conducting qualitative research, pilot studies, or exploratory research.
- **Budget and time constraints:** The non-probability method when there are budget and time constraints, and some preliminary data must be collected. Since the survey design is not rigid, it is easier to pick respondents randomly and have them take the survey or questionnaire.

2.Difference between Probability and Non-Probability sampling Methods.

To encapsulate the whole discussion, though, the significant differences between probability sampling methods and non-probability sampling methods are as below:

	Probability Sampling Methods	**Non-Probability Sampling Methods**

Definition	Probability Sampling is a sampling technique in which samples from a larger population are chosen using a method based on the theory of probability.	Non-probability sampling is a sampling technique in which the researcher selects samples based on the researcher's subjective judgment rather than random selection.
Alternatively Known as	Random sampling method.	Non-random sampling method
Population selection	The population is selected randomly.	The population is selected arbitrarily.
Nature	The research is conclusive.	The research is exploratory.
Sample	Since there is a method for deciding the sample, the population demographics are conclusively represented.	Since the sampling method is arbitrary, the population demographics representation is almost always skewed.
Time Taken	Takes longer to conduct since the research design defines the selection parameters before the market research study begins.	This type of sampling method is quick since neither the sample nor the selection criteria of the sample are undefined.
Results	This type of sampling is entirely unbiased; hence, the results are also conclusive.	This type of sampling is entirely biased, and hence the results are biased, too, rendering the research speculative.
Hypothesis	In probability sampling, there is an underlying hypothesis before the study begins, and this method aims to prove the hypothesis.	In non-probability sampling, the hypothesis is derived after conducting the research study.

3. Explain the concept of descriptive analytics.

Descriptive Analytics is a field of business intelligence with expertise in statistical analysis, waiting for history, and other data. Descriptive Analytics professionals find the data to question and study; they pose the questions that need answers; they translate these queries into mathematical models and apply them to their chosen data. Using descriptive analytics in data visualization is a practice that can greatly aid and improve people's decision-making. Their decisions are taken from over-reliance, wishful thinking, and in isolation. Descriptive analytics is a rapidly growing field with a promising future. It offers the ability to make better business decisions and understand how customers interact with companies and products.

There are four main steps in descriptive analytics:

1. **Data collection:** Collecting data from various sources such as sales reports, customer surveys, social media, etc.
2. **Data preparation:** Cleaning and organizing the data so it can be analyzed.
3. **Exploratory data analysis:** Analyzing the data to find trends, patterns, and relationships.
4. **Data visualization:** Creating graphs and charts to visualize the data and make it easy to understand.

The most common techniques used in descriptive analytics are statistical analysis, data visualization, and predictive modeling.

Descriptive Analysis Techniques

The techniques for descriptive analysis are the most common descriptive methods of data analysis for qualitative data. Descriptive data analysis techniques are used to describe the subjects of a study in detail, identifying patterns and trends, and providing insights into how subjects behave.

Some of the most common descriptive analysis methods for descriptive analysis statistics are:

- The frequency distribution is a method that provides an overview of all the responses to a question.
- The bar chart is a visual representation that displays how responses vary on different dimensions.
- The pie chart displays how responses vary on different dimensions.
- A scatterplot displays how two variables relate to each other.
- A histogram provides an overview of all the responses to a question, with each response grouped into bins according to some criterion such as age or income level.

Types of Descriptive Analysis

There are four different types of descriptive analysis: measures of frequency, central tendency, dispersion or variation, position, . These techniques work best when only one variable is present.

1. **Measures of Frequency**

Understanding how frequently a specific event or response is likely to occur is crucial for descriptive analysis. The main goal of frequency measurements is to create something akin to a count or a percentage. For example: Think about a survey where 500 people are questioned about their favorite football team. A list of 500 responses would be challenging to read and organize, but by counting the number of times a specific football team was chosen, the data can be made much more understandable.

2. **Measures of Central Tendency**

A single value that seeks to characterize a set of data by pinpointing the central position within that set of data is referred to as a measure of central tendency. As a result, measures of central location are occasionally used to refer to measures of central tendency. They also fit into the category of summary statistics. You are probably most familiar with the mean (also known as the average), but there are other central tendency measures, including the median and the mode.

The mean, median, and mode are all reliable indicators of central tendency, but depending on the situation, some indicators are more useful than others.

Think about a survey where 1,000 people's body weight are recorded as an example. The mean average would be a great descriptive metric to use in this situation to measure mid-values.

3. Measures of Dispersion

At times, understanding how data is distributed across a range is crucial. Consider the average weight of a sample of two people to further explain this. The average weight will be 60 kilograms if both people weigh 60 kilograms. The average weight is still 60 kg even if one person weighs 40 kg and the other 80 kg. This type of distribution can be measured using dispersion metrics like range or standard deviation.

4. Measures of Position

Identifying the position of a single value or its response in relation to others is another aspect of descriptive analysis. In this field of expertise, metrics like percentiles and quartiles are extremely helpful.

Measures: Percentile Ranks, Quartile Ranks

Similar to bivariate analysis, the multivariate analysis examines more than two variables. The two methods for bivariate analysis are listed below:

5. Contingency table

In statistics, a contingency table, also known as a two-way frequency table—is a tabular representation with at least two rows and two columns that are used to present categorical data as frequency counts.

For instance, the contingency table below, which has two rows and five columns, displays the findings of a random sample of 2200 adults categorized by gender and preferred method of eating Icy dessert.

Gender	cup	cone	sundae	sandwich	other
Male	592	300	204	24	80
Female	410	335	180	20	55

2. **Scatter plots**

You can visualize the relationship between two or three different variables using a scatter plot. It represents a relationship's strength in a visual way. One variable should be plotted along the x-axis, and another along the y-axis in a scatter plot. A point in the chart represents each data point.

The Advantages and Disadvantages of Descriptive Analytics in Data Science

Even though it is one of the more straightforward analytical strategies, descriptive analysis in data science has many benefits:

- gives access to information that would otherwise be difficult to understand.
- gives a precise estimation of how frequently important data points occur.
- is cheap, and it only calls for rudimentary mathematical knowledge.
- is easier to complete, particularly with the aid of programmes like Python or Microsoft Excel.
- relies on information that businesses already have, so getting new information is not necessary.
- compared to inferential statistics, it considers the entire population (rather than a data sampling).

We've examined the benefits of descriptive analytics, but what are its drawbacks? The following are some drawbacks of descriptive analytics:

- Although you can summarize the data sets you have access to, they might not provide the full picture.
- Descriptive analytics can't be used to test a theory or figure out why data is presented in a certain way.
- Descriptive analytics cannot be used to make future predictions.
- Your results cannot be applied to a larger population as a whole.
- Descriptive analytics provide no information regarding the method of data collection, so the data set may contain errors.

4.How is Descriptive Analytics Used in businesses?

Businesses utilize descriptive analytics in various areas of their operations to assess how well they are performing and if they are on track to meet their objectives. Common financial measurements generated by descriptive analytics, such as quarterly increases in sales and expenses, are monitored by business executives and financial experts. By tracking stats like conversion rates and the number of social media followers, marketing teams may assess the effectiveness of their campaigns. Production line throughput and downtime are among the variables that manufacturing companies keep an eye on.

There are several uses for the metrics generated by descriptive analytics, including:

1. **Reports:** Descriptive analytics is used to provide the primary financial indicators found in a company's financial statements. Descriptive analytics are often used in other typical reports to emphasize specific areas of business performance.
2. **Visualizations**: Metrics can be more effectively communicated to a larger audience by being displayed in charts and other graphic forms.
3. **Dashboards**: Dashboards are a tool that executives, managers, and other staff members can use to monitor progress and organize their daily workload. Dashboards offer a selection of KPIs and other crucial data that are catered to the needs of each

individual. To help people quickly digest the information, it may be presented as charts or other visualizations.

The Five Steps Descriptive Data Science Involves

Determining the metrics you want to output is typically the first step in applying descriptive analytics, and presenting them in the proper format is the final step. The procedures to generate your own descriptive analytics are listed below.

Step 1: Define business metrics

Defining the metrics you wish to measure is the first step. These should represent the main organization's objectives of each segment or the organization as a whole. For instance, a company that prioritizes expansion may track quarterly revenue growth, and its accounts receivable department may monitor metrics like days sales outstanding and other measures of how long does it take to get payment from a customer?

Step 2: Identify data required

Find the data you require to generate the desired stats. The data may be dispersed over numerous programmes and files at some businesses. Businesses that use ERP systems can already have the majority or all of the data they require in the databases of their systems. Some indicators might also need information from outside sources, like social media platforms, e-commerce websites, and databases used for industry benchmarking.

Step 3: Extract and preprocess data

When data is gathered from several sources, extracting, integrating, and preprocessing it before analysis is a time-consuming but necessary step to ensure accuracy. This procedure could include data cleansing to eliminate conflicts and inaccuracies in data from diverse sources and convert the data into a format compatible with descriptive-analytical tools. Advanced data analytics employ the method of data modeling to help prepare, shape, and organize corporate data. Data modeling is a framework for describing and formatting data inside information systems.

Step 4: Data Analysis

Businesses can utilize a variety of technologies, such as spreadsheets and business intelligence (BI) tools, to do descriptive analytics. In descriptive analytics, applying simple mathematical operations to one or more variables is a common step. For example, sales managers could monitor the average profit per transaction or the monthly revenue from new clients. Executives and financial professionals may keep an eye on financial indicators like the gross profit margin, or the ratio of gross profit to sales.

Step 5: Present data

Data that is presented in visually appealing forms, such as pie charts, bar charts, and line graphs, is typically easier for stakeholders to understand. But some people, like financial professionals, could like information that is provided in the form of figures and tables.

Build the skills to design Data Science and Machine Learning models and get noticed by organizations for your ability to help them harness the power of big data. From Machine Learning and Data Mining to Data Analysis, Knowledgehut have got best <u>Machine Learning with Python training</u> course to help you progress on your career path.

5.Explain the Descriptive Analytics Use Cases of business analytics.

Even though descriptive analytics only considers what occurred rather than why it is still an important first step in the larger data analytics process. Let's look at it.

1. **Monitoring social media activity**

A significant touchpoint in the sales process is social media. Therefore, it is essential to have the ability to measure and present engagement metrics across a complex constellation of campaigns and social networks in order to identify the most effective strategies for digital marketing. Fortunately, descriptive analytics will be included by default in marketing reports on social media engagement. Measures of social media engagement include clicks, likes, shares, detail expands, bounce rates, and more. These metrics are all easily summed up using descriptive techniques.

For instance, a business might be curious to know which social media account is bringing in the most visitors to its website. They can quickly compare data about various channels using dashboards, visualizations, and descriptive statistics. Similarly to this, marketing teams can examine particularly shareable content and compare, for example, blog posts and videos, to determine which generates the most clicks.

Although none of this data draws clear conclusions (since it doesn't assess cause and effect), it is still useful. Teams can use it to create hypotheses or make educated assumptions about where to spend their time and money.

2. Streaming and Online Shopping

Descriptive analytics are used to spot trends by subscription streaming services like Spotify and Netflix as well as e-commerce websites like Amazon and eBay. Descriptive metrics are useful for identifying what users and consumers are currently most interested in. For instance, Spotify uses descriptive analytics to find out which albums or artists its subscribers are enjoying. Amazon compares customer purchases using descriptive analytics. These insights influence the recommendation engines in both cases to work with are influenced by these insights.

While Netflix goes even further in its application of descriptive analytics. Netflix, a business that places a high value on data, uses descriptive analytics to determine which genres and TV shows are most popular with its audience. Decisions about new content creation, marketing strategies, and even which production companies they work with.

3. Learning management systems

Many institutions now use online/offline hybrid learning, from traditional education to corporate training. A common component of this is learning management systems or LMSs for those in the know. LMS platforms keep track of everything, including user participation, attendance, test results, and in the case of e-learning programs even the length of time it takes students to finish a course. Descriptive-analytical reports provide a high-level overview of what is working and what is not by summarizing this data.

Teachers and training specialists can monitor goals at the individual and organizational levels using these data. They can examine grade distributions or discover the most well-liked teaching aids. And even though they won't always understand why, it might be possible to deduce from the data that, for instance, videos are more popular than written documents. The first step in improving course design and improving learner outcomes is to present this information

6.Explain the role of warehousing in business analytics.

Data warehousing involves collecting, storing, and managing enormous data from various sources. It is a centralized repository that offers a single source of truth for consistent and reliable data analysis to support business decision-making. This data is structured, cleansed, transformed, and organized to provide a unified view of your organization's operations and performance.

Data warehouses form the backbone of Business Intelligence (BI), offering a centralized, reliable, and integrated platform for data analysis. Their unique features empower BI in several ways:

Single Source of Truth: Data warehouses eliminate the need to navigate scattered data across different systems. Instead, they act as a single, unified source of truth, ensuring data consistency and accuracy. This reduces the risk of errors or misleading insights, enabling BI to deliver reliable and actionable insights.

Enhanced Data Quality: The data transformation process allows users to clean, standardize, and transform information from various sources. This results in high-quality data suitable for reliable analysis. Additionally, the non-volatile nature of data warehouses preserves historical data integrity for long-term analysis.

Faster Analysis: Data warehouses are designed to utilize parallel processing and columnar data storage for faster retrieval and analysis of large datasets. This optimized approach allows you to spend less time searching for relevant data and more time extracting valuable insights from your information.

Complex Queries and Exploration: The structured architecture of the data warehouse facilitates the efficient execution of complex queries and in-depth data exploration. This empowers you to delve deeper into the data, uncover hidden patterns, and explore various dimensions. This flexibility fuels strategic decision-making by providing comprehensive insights that answer complex business questions.

Modern Data Warehouses for Business Intelligence
While various data warehouse solutions exist, two popular options well-suited for BI needs include:

- **Amazon Redshift**: A cloud-based data warehouse service from Amazon Web Services (AWS) known for its scalability, cost-effectiveness, and ease of use. It is designed for high-performance analytics and can handle petabyte-scale operations. Redshift offers features like columnar storage, parallel query execution, and integration with other AWS services like S3 and AWS Glue.
- **Snowflake:** Another cloud-based data warehouse solution valued for its scalability, performance, and flexibility. Snowflake offers cross-cloud deployment options and supports various data formats, making it adaptable to diverse data scenarios.

Challenges with Data Warehouse and Business Intelligence

While data warehouse and BI offer immense benefits, implementing and utilizing them comes with its own set of challenges:

Data Quality and Consistency: Maintaining high-quality and consistent data is crucial in reporting and analysis. However, addressing issues like missing values, inconsistencies, inaccurate data, and duplicate entries can be challenging. However, implementing CDC technology can minimize copying exact data by capturing only new information from source systems to maintain consistency.

Integration with Existing Systems: Integrating data from multiple sources into a data warehouse with existing IT infrastructure can be time-consuming. Planning with consideration of existing systems and data formats is essential to ensure smooth integration and accessibility. Additionally, testing and rollback plans become crucial to validate functionality as well as address unforeseen issues during integration.

Data Security and Privacy: Robust security measures are essential to protect against unauthorized access, data breaches, and compliance violations. Implementing access controls, encryption, and regular backups is crucial to safeguard data security and privacy.

Cost Management: Building and maintaining a data warehouse and BI infrastructure requires strategic resource allocation and investment in technology and personnel. However, achieving efficient operations and sustained ROI relies on ongoing cost management and technology optimization.

7. Explain the visual analytics and its benefits in details.

Visual analytics is the use of sophisticated tools and processes to analyze datasets using visual representations of the data. Visualizing the data in graphs, charts, and maps helps users identify patterns and thereby develop actionable insights. These insights help organizations make better, data-driven decisions.

Benefits of Visual Analytics

- Visual analytics is a valuable tool for medium and large enterprises across various sectors. In the business and finance sector, visual analytics can assist in analyzing financial data, identifying market trends, and making strategic decisions based on real-time data.

- The global business intelligence and analytics software market is expected to reach $26.78 billion by 2023. In healthcare, visual analytics can help providers analyze patient data, identify patterns in medical records, and improve patient outcomes. The global healthcare analytics market is expected to reach $84.2 billion by 2026.

- Additionally, visual analytics can help marketers track customer behavior, analyze customer data, and create targeted marketing campaigns. It can also help manufacturers optimize production processes, reduce costs, and improve product quality.

- Faster Insights: By using interactive visualizations, you can quickly identify patterns, trends, and outliers in data, leading to faster insights.
- Improved Decision Making: Visual Analytics can help users make better decisions by providing more accurate and relevant information.
- Increased Efficiency: By streamlining the data analysis process, Visual Analytics can save time and resources while improving the quality of insights.
- Enhanced Collaboration: Visual Analytics enables teams to work together and share insights more effectively, leading to better outcomes.

Feature	Data Visualization	Visual Analytics

Purpose	To make data more understandable and accessible to humans	To explore, analyze, and understand large and complex datasets
Process	Creating visual representations of data	Using interactive visual interfaces to explore, analyze, and understand data
Tools	Charts, graphs, maps, etc.	Data mining algorithms, statistical analysis, machine learning, etc.
Output	Visual representations of data	Insights into data

UNIT-IV
1.Explain the procedure and applications involved in predictive analytics.

Predictive analytics is a branch under advanced analytics primarily used to make predictions about the uncertain future events. Predictive analytics makes use of statistics, modelling, data mining, artificial intelligence, machine language to work on the current set of data provided as instructions and predict the future events.

Through the usage of the historical data and transactional data patterns, risks are identified in addition to exploring opportunities for future needs. Through the usage of predictive analysis businesses can effectively interpret the date available to them for their benefits.

Data mining and text analytics along with statistics enables business users to create a predictive intelligence by unravelling relationships among the data patterns both in the structured and unstructured data. Certain data can be used readily for commencing the analysis like the age, gender, income, sales etc. These data are referred to as structured data. In addition to the structured data we also information available around us like the social media, website information and any other information from which data could be picked and then structured for further analysis. This is called unstructured data. The date collected from such activities are then extracted and used in model building process.

Through predictive analytics organisations tend to get more pro-active and forward looking in terms of planning and apt decision making rather than going by emotions or a gutsy feeling!

Process involved in Predictive Analytics

1. Project Definition: Identify what shall be the outcome of the project, the deliverables, business objectives and based on that go towards gathering those data sets that are to be used.
2. Data Collection: This Is more of the big basket where all data from various sources are binned for usage. This gives a picture about the various customer interactions as a single view

item.
3. Analysis: Here the data is inspected, cleansed, transformed and modelled to discover if it really provides useful information and arriving at conclusion ultimately.
4. Statistics: This enables to validate if the findings, assumptions and hypothesis are fine to go ahead with and test them using statistical model.
5. Modelling: Through this accurate predictive models about the future can be provided. From the options available the best option could be chosen as the required solution with multi model evaluation.
6. Deployment: Through the predicative model deployment an option is created to deploy the analytics results into everyday effective decision. This way the results, reports and other metrics can be taken based on modelling.
7. Monitoring: Models are monitored to control and check for performance conformance to ensure that the desired results are obtained as expected.

Applications of Predictive Analytics

4. **CRM:** Through predictive analytics marketing campaigns, sales, and customer services are objectively achieved. This can be used in analytical customer relationship management throughout the customer life cycle right from the acquisition, relationship growth, retention and customer win back can be better planned and strategically addressed for retaining customers and addressing them more clearly.
2. Health care: Usage of predictive analytics in the health care domain can aid to determine and prevent cases and risks of those developing certain health related complications like diabetics, asthma and other life threatening ailments. Through the administering of predictive analytics in health care better clinical decisions can be made.
3. Collection Analytics: These applications optimise the allocation of collection resources by identifying collection agencies, contact strategies to reach out to them, legal actions to increase recovery and cost reduction of collection.
4. Cross Sell: Through predictive analytics applications attached to various touch points connected to the customers a detailed analysis on the customer spends, usage pattern of certain purchases they make regularly, customer behaviour can obtained with which ultimately to efficient cross sales or selling additional products to customers. This way organisations dealing with multiple products can effectively increase its sales volume and profits ultimately.
5. Fraud Detection: Predictive Analytics can aid to spot inaccurate credit application, deviant transactions leading to frauds both online and offline, identity thefts and false insurance claims saving financial and insurance institutions of lots of security issues and damages to their operations.
6. Risk Management: The best portfolio prediction to maximise returns on the capital invested, probabilistic risk assessment to yield accurate forecasts are some of the important benefits of using predictive analytics.
7. Direct Marketing: Predictive Analytics also aids in identifying the most effective combination of product versions, marketing material, communication channels and the timing to be used to target a given consumer in the current environment where the dynamics are constantly changing and gets challenging for a business to compete and run successfully.
8. Underwriting: Perhaps one of the biggest benefits that can infiltrate into underwriting is providing information about the likelihood of illness, default of loan/insurance and

bankruptcy. Predictive Analytics streamline the process of customer acquisition by closely predicting the future risk behaviour of a customer through the application data.

9.Marketing

Individuals working in the marketing domain need to look how consumers will react on a particular market campaign, or what will be their impact on the overall economy while conducting such marketing event, etc. in this case,

- Predictive analytics tools could be helpful in segmenting the marketing leads by displaying ads over websites and social media platforms relating to consumer behavior and interest.
- Predictive analytics tools can explore "expect to purchase" by analyzing consumer's behaviour on past and current available data to find people whose data matches with ideal consumers.
- Marketers could also use predictive analytics for leads scoring by analyzing data to identify which prospects are potentially most valuable for the company, or to identify how likely the prospective consumers will buy products or services and to plan how they should be contacted and with what information.

10.Retail

Either online or brick and mortar, each retailer looks for managing inventory and logistics, and thereby predictive analytics is extremely important. The method allows retailers to correlate huge data information such as historical sales data, purchasing products and behavior, geographical references to optimize operations and efficiencies in the following ways;

- Customer sales data provides personalized recommendations and promotions for individual customers, through predictive analytics, better targeting built over real-time data assists retailers for planning campaigns, making ads and promotions that buyers will respond the most.
- Sales and logistics data analysis using predictive analytics helps retailers to ensure the availability of sufficient inventory/products in warehouses, and good merchandise in stores at the right time.

11.Manufacturing

With the modernized technology and fully automated factory machines, predictive analytics tools are very significant in operating and optimizing the manufacturing process at each stage of designing, purchasing, developing, quality and inventory control, delivery, etc. Moreover,

- Predictive analytics is helpful when combined with machine data in order to help in tracking and comparing machines' performance and equipment maintenance status and predicting which particular machine will fail.

- Predictive analytics insights can lead to decrease in shipping and transportation expenses by accepting all the factors included in transferring manufacturing products at different places under the proper system.
- Considering predictions over supply chain and sales data helps in making more considerable decisions for purchasing and ensuring that no expensive raw materials get purchased unless not required. This data can also be used in aligning manufacturing processes with consumer demands.

12.Healthcare

Healthcare industry is among dominant adapters to consider predictive analytics techniques aimed at facilitating technology to save money and improve health practices efficiencies.
- Predictive analytics can help medical practitioners by analyzing data concerning global diseases statistics, drug interactions, patient diagnostic history individually to provide advanced care and conduct more effective medical practices.
  - Applying predictive analytics on clinics' past appointment data helps in identifying probable no-shows or delays in cancellations more accurately and thus save time and resources.
  - To detect claims frauds, the health insurance industry is using predictive analytics to discover patients at most risk of incurable or chronic disease, it helps companies in finding suitable interventions.

2.Discuss the Predictive Analytics Tools &Techniques of BA.

In todays' industries involving healthcare, life sciences, oil and gas, insurance, etc., predictive analytics is widely employed in these areas and provides most valued anticipations when business strategies and applications are clearly defined.

Predictive analytics incorporates a combination of scientific methods and techniques as discussed below;

Data Mining: In order to manage large amounts of data sets either structured or unstructured to recognize hidden patterns and relationships among variables provided, data mining is aimed to. Once identified, these relationships can be used to understand the behaviour of the event from which data is compiled.

- **Statistical Modelling:** In parallel to the data mining process, statistical data models can be developed depending on the context of what needs to be anticipated using the same collected data as for data mining. Once the model

builts, the new data is fed to models to predict future outcomes. For example, a business expert can build a cross-selling model using current customer data and predict what other items they will likely to purchase from the same company.

- **Machine learning:** ML can deploy iterative methods and techniques to identify patterns from large data sets and build models. For example, recommendation engines are widely used for online shopping recommendations as predictions are made from using customers' prior purchasing and browsing behavior.

Predictive Analytics Tools

Predictive analytics tools use data to help you predict the future. Instead, it informs you of the probability of various scenarios. Knowing these possibilities might assist you in planning various parts of your business.

Predictive analytics is a subset of data analysis. Descriptive analytics, which helps you determine what your data represents, is another part of data analytics. Diagnostic analytics identify the root reasons for what has occurred. Prescriptive analytics is more similar to predictive analytics. This provides you with actionable advice for making better selections.

- **SAS Advanced Analytics**

SAS is the global leader in analytics, with a plethora of various predictive analytics products offered. The list is so broad that it may be difficult to determine which tool(s) you will require for your specific needs. In addition, the firm does not give upfront pricing, making it difficult to compare prices. Nonetheless, with so many different tools available, chances are SAS offers just what you want.

- **IBM SPSS**

IBM SPSS (Statistical Package for the Social Sciences) is a data modeling and statistics-based analytics program. The software can handle both organized and unstructured data. To meet any security and mobility requirements, this software is offered in the cloud, on-premises, or via hybrid deployment.

- **RapidMiner Studio**

RapidMiner Studio blends data preparation and analysis with unique business implementation. You may use this code optimal application to automate reporting based on time intervals or to have events trigger changes in your visualizations.

Using the platform's 60+ native integrations, you may import your own data sets and export them to other programs. Extensions provide you more functionality, for example, anomaly detection, text processing, and web mining, but they may cost more than the basic membership fee.

TIBCO Spotfire

TIBCO Spotfire includes a variety of tools for working with large data sets. Spotfire is simple enough for anybody to utilize when it comes to predictive analytics. Spotfire includes a feature known as one-click predictions. These are pre-programmed methods for classifying and clustering data.

It also displays relationships and forecasts. Spotfire features an attractive data display. It is always constantly reading data and updating in real-time. It is simple to create your apps for use with the platform. Spotfire's machine learning algorithms get a greater in-depth understanding.

- **H2O**

If you're looking for an open-source predictive analytics solution, H2O should be at the top of your list. It provides quick performance, low cost, superior features, and great flexibility. The H2O dashboard provides excellent visualization of data insights.

Predictive analytics are useful in;

- Prohibition of credit card fraud via indicating unusual transactions,
- Credit card scoring to determine whether to approve or deny loan applications,
- Most importantly, analyzing customers' churn data and facilitating banks to approach potential customers before they are likely to switch respective institutions.
- Measuring credit risk, maximizing cross-sell/up-sell opportunities and retaining valuable customers.
- Commonwealth bank implements predictive analytics to anticipate fraud activities for a given transaction before it is accomplished- within 40 milliseconds of the occurrence of transaction.

3.Explain the types of predictive analytics.

Types of Predictive Analytical Models

There are three common techniques used in predictive analytics: Decision trees, neural networks, and regression. Read more about each of these below.

Decision Trees
If you want to understand what leads to someone's decisions, then you may find decision trees useful. This type of model places data into different sections based on certain variables, such as price or market capitalization. Just as the name implies, it looks like a tree with individual branches and leaves. Branches indicate the choices available while individual leaves represent a particular decision.

Decision trees are the simplest models because they're easy to understand and dissect. They're also very useful when you need to make a decision in a short period of time.

Regression
This is the model that is used the most in statistical analysis. Use it when you want to determine patterns in large sets of data and when there's a linear relationship between the inputs. This method works by figuring out a formula, which represents the relationship between all the inputs found in the dataset. For example, you can use regression to figure out how price and other key factors can shape the performance of a security.

Neural Networks
Neural networks were developed as a form of predictive analytics by imitating the way the human brain works. This model can deal with complex data relationships using artificial intelligence and pattern recognition. Use it if you have several hurdles that you need to overcome like when you have too much data on hand, when you don't have the formula, you need to help you find a relationship between the inputs and outputs in your dataset, or when you need to make predictions rather than come up with explanations.

 Cluster Models
Clustering describes the method of aggregating data that share similar attributes. Consider a large online retailer like Amazon. Amazon can cluster sales based on the quantity purchased or it can cluster sales based on the average account age of its consumer. By separating data into similar groups based on shared features, analysts may be able to identify other characteristics that define future activity.

Time Series Modeling
Sometimes, data relates to time, and specific predictive analytics rely on the relationship between what happens when. These types of models assess inputs at specific frequencies such as daily, weekly, or monthly iterations. Then, analytical models seek seasonality, trends, or behavioral patterns based on timing. This type of predictive model can be useful to predict when peak customer service periods are needed or when specific sales will be made.

4.Explain the techniques of Data Mining with an example.

Data mining is the process of searching and analyzing a large batch of raw data in order to identify patterns and extract useful information. Companies use data mining software to learn more about their customers. It can help them to develop more effective marketing strategies, increase sales, and decrease costs. Data mining relies on effective data collection, warehousing, and computer processing.

Data Mining Techniques

The majority of data mining tasks can be classified as either prediction or summary at a high level. Predictive tasks allow you to forecast the value of a variable based on previously collected data. Predicting when a customer will leave a business, predicting whether a transaction is fraudulent, and recognizing the best customers to receive direct marketing offers are all examples of predictive tasks.

- Cleansing and preparing data
- Pattern Recognition
- Classification
- Association
- Detection of Outliers
- Clustering
- Regression
- Prediction
- Sequential trends
- Decision Trees
- Statistical techniques
- Visualization
- Neural Networks
- Data warehousing
- Machine Learning and Artificial intelligence
- Cleansing and preparing data

Cleaning and preparing data is a vital part of the data mining process. Raw data must be cleansed and formatted in order to be useful in various analytic approaches. Different elements of data modeling, transformation, data migration, ETL, ELT, data integration, and aggregation are used in data cleaning and planning. It's a necessary step in determining the best use of data by understanding its basic features and attributes. Cleaning and preparing data has obvious business value. Data is either useless to an entity or inaccurate due to its accuracy if this first phase is not completed. Companies must be able to trust their data, analytics results, and the actions taken as a result of those findings. These measures are also needed for good data quality and data governance.

Pattern Recognition

A basic data mining technique is pattern recognition. It entails spotting and tracking trends or patterns in data in order to draw informed conclusions about business outcomes. When a company notices a pattern in sales data, for example, it has a reason to act. If it's determined that a certain product sells better than others for a specific demographic, a company may use this information to develop similar goods or services, or simply better stock the original product for this demographic.

Classification

The various attributes associated with different types of data are analyzed using classification data mining techniques. Organizations may categorize or classify similar data after identifying the key

characteristics of these data types. This is essential for recognizing personally identifiable information that organizations may wish to shield or redact from records, for example.

Association

The statistical technique of association is a data mining technique. It denotes that some data (or data-driven events) are linked to other data. It's similar to the machine learning concept of co-occurrence, where the existence of one data-driven event indicates the probability of another. Correlation and association are two statistical concepts that are very similar. This means that data analysis reveals a connection between two data occurrences, such as the fact that hamburger purchases are often followed by French fries purchases.

Detecting of Outliers

Outlier detection is used to identify the deviations in datasets. When companies discover anomalies in their records, it becomes easier to understand why they occur and plan for potential events in order to achieve business goals. For example, if there is an increase in the use of transactional systems for credit cards at a certain time of day, businesses can use this information to maximize their income for the day by finding out the cause of it.

Clustering

Clustering is an analytics methodology that employs visual approaches to data interpretation. Graphics are used by clustering mechanisms to demonstrate where data distribution is in relation to various metrics. Different colors are used in clustering techniques to represent data distribution. When it comes to cluster analytics, graph-based methods are perfect. Users can visually see how data is distributed and recognize patterns related to their business goals using graphs and clustering in particular.

Regression

The essence of the relationship between variables in a dataset can be determined using regression techniques. In some cases, such connections may be causal, and in others, they may only be correlations. Regression is a simple white box technique for revealing the relationships between variables. In areas of forecasting and data modeling, regression methods are used .

Prediction
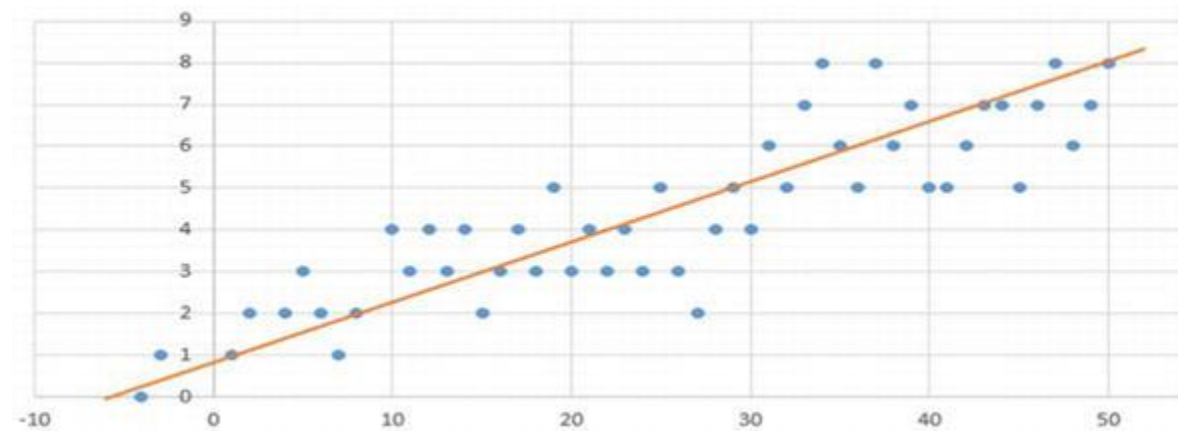
One of the four branches of analytics is prediction, which is a very important feature of data mining. Patterns observed in current or historical data are extended into the future using predictive analytics. As a result, it allows businesses to predict what data patterns will emerge next. Using predictive analytics can take a variety of forms. Machine learning and artificial intelligence are used in some of the more advanced examples. Predictive analytics, on the other hand, does not have to rely on these methods; simpler algorithms can also be used.

Sequential Trends

This data mining technique focuses on identifying a sequence of events. It's particularly useful for transactional data mining. For example, when a customer buys a pair of shoes, this technique will show which pieces of clothing they are more likely to buy. Understanding sequential trends may assist businesses in recommending additional products to consumers in order to increase sales.

Decision trees

Decision trees are a form of predictive model that enables businesses to mine data more effectively. A decision tree is technically a machine learning technique, but because of its simplicity, it is more often referred to as a white box machine learning technique. Users can see how the data inputs influence the outputs using a decision tree. A random forest is a predictive analytics model that is created by combining different decision tree models. Complicated random forest models are referred to as "black box" machine learning techniques because their outputs are not always easy to comprehend based on their inputs. However, in most cases, this simple form of ensemble modeling is more effective than relying solely on decision trees

Statistical techniques
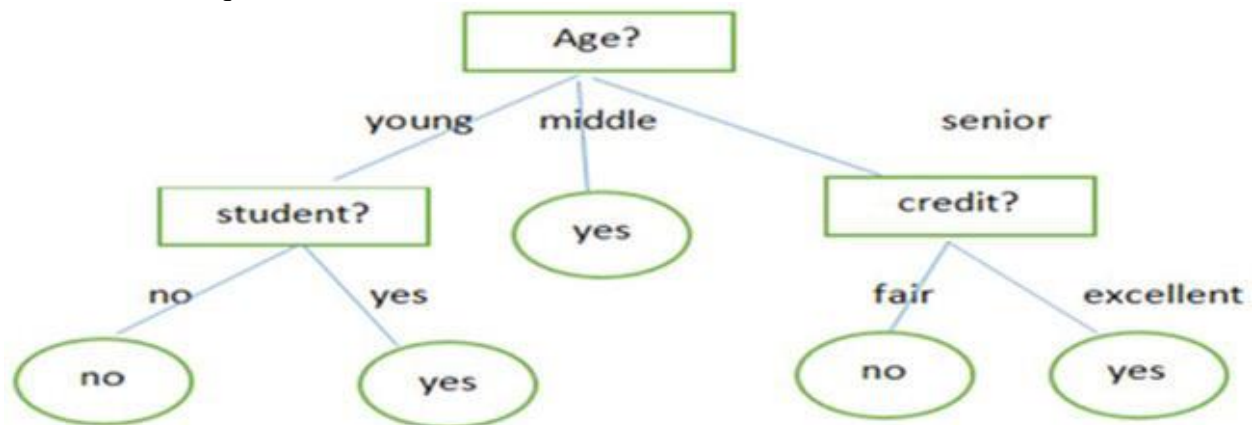
Example of a decision tree

Statistical approaches are at the heart of the majority of data mining analytics. The various analytics models are focused on mathematical principles that produce numerical values that can be used to achieve clear business goals. In image recognition systems, neural networks, for example, use complex statistics based on various weights and measures to decide if a picture is a dog or a cat. Statistical models are one of artificial intelligence's two primary branches. Some mathematical methods have static models, while others that use machine learning improve over time.

Visualization

Another essential aspect of data mining is data visualization which uses sensory impressions that can be seen to provide users with access to data. Today's data visualizations are interactive, useful for streaming data in real-time, and distinguished by a variety of colors that show various data trends and patterns. Dashboards are a valuable tool for uncovering data mining insights using data visualizations. Instead of relying solely on the numerical results of mathematical models, organizations may create dashboards based on a variety of metrics and use visualizations to visually illustrate trends in data.

Neural Networks

A neural network is a type of machine learning model that is frequently used in AI and deep learning applications. Among the most accurate machine learning models used today is neural network. They are named for the fact that they have multiple layers that resemble how neurons function in the human brain. While a neural network can be a powerful tool in data mining,

companies should exercise caution when using it because some of these neural network models are extremely complex, making it difficult to understand how a neural network calculated an output

Data warehousing
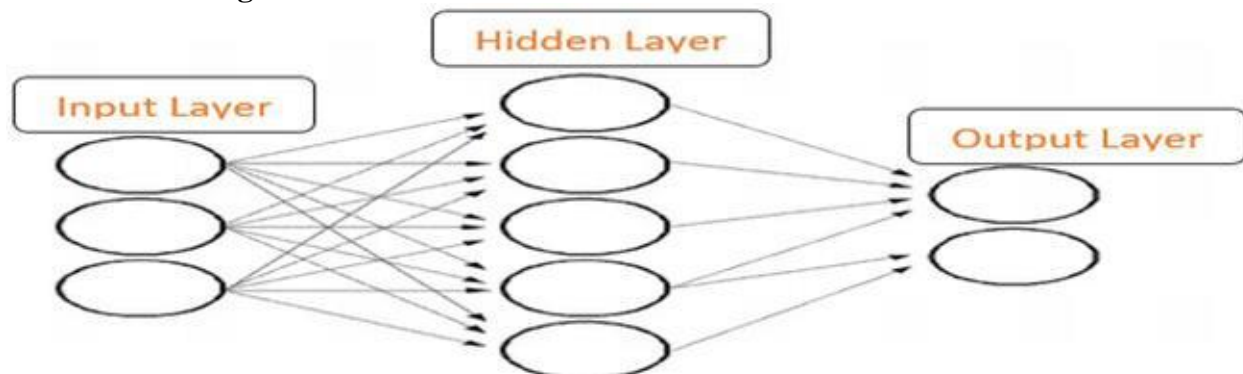
Data warehousing used to imply storing organized data in relational database management systems so that it could be analyzed for business intelligence, reporting, and simple dashboarding. Cloud data centers and data warehouses in semi-structured and unstructured data stores, such as Hadoop, are available today. Although data warehouses have historically been used to store and analyze historical data, many new approaches can now provide in-depth, real-time data analysis.

Machine Learning and Artificial Intelligence

Some of the most advanced advances in data mining are machine learning and artificial intelligence (AI). When operating with large amounts of data, advanced machine learning techniques such as deep learning provide extremely accurate predictions. As a result, they can be used to process data in AI applications such as computer vision, speech recognition, and advanced text analytics using Natural Language Processing. These data mining techniques work well with semi-structured and unstructured data to determine meaning.

5.Discuss the process & applications of data mining

The Data Mining Process

To be most effective, data analysts generally follow a certain flow of tasks along the data mining process. Without this structure, an analyst may encounter an issue in the middle of their analysis that could have easily been prevented had they prepared for it earlier. The data mining process is usually broken into the following steps.

Step 1: Understand the Business

Before any data is touched, extracted, cleaned, or analyzed, it is important to understand the underlying entity and the project at hand. What are the goals the company is trying to achieve by mining data? What is their current business situation? What are the findings of a SWOT analysis? Before looking at any data, the mining process starts by understanding what will define success at the end of the process.

Step 2: Understand the Data

Once the business problem has been clearly defined, it's time to start thinking about data. This includes what sources are available, how they will be secured and stored, how the information will be gathered, and what the final outcome or analysis may look like. This step also includes determining the limits of the data, storage, security, and collection and assesses how these constraints will affect the data mining process.

Step 3: Prepare the Data
Data is gathered, uploaded, extracted, or calculated. It is then cleaned, standardized, scrubbed for outliers, assessed for mistakes, and checked for reasonableness. During this stage of data mining, the data may also be checked for size as an oversized collection of information may unnecessarily slow computations and analysis.

Step 4: Build the Model
With our clean data set in hand, it's time to crunch the numbers. Data scientists use the types of data mining above to search for relationships, trends, associations, or sequential patterns. The data may also be fed into predictive models to assess how previous bits of information may translate into future outcomes.

Step 5: Evaluate the Results
The data-centered aspect of data mining concludes by assessing the findings of the data model or models. The outcomes from the analysis may be aggregated, interpreted, and presented to decision-makers that have largely been excluded from the data mining process to this point. In this step, organizations can choose to make decisions based on the findings.

Step 6: Implement Change and Monitor
The data mining process concludes with management taking steps in response to the findings of the analysis. The company may decide the information was not strong enough or the findings were not relevant, or the company may strategically pivot based on findings. In either case, management reviews the ultimate impacts of the business and recreates future data mining loops by identifying new business problems or opportunities.

Different data mining processing models will have different steps, though the general process is usually pretty similar. For example, the Knowledge Discovery Databases model has nine steps, the CRISP-DM model has six steps, and the SEMMA process model has five steps.1

Applications of Data Mining
In today's age of information, almost any department, industry, sector, or company can make use of data mining.

Sales
Data mining encourages smarter, more efficient use of capital to drive revenue growth. Consider the point-of-sale register at your favorite local coffee shop. For every sale, that coffeehouse collects the time a purchase was made and what products were sold. Using this information, the shop can strategically craft its product line.

Marketing
Once the coffeehouse above knows its ideal line-up, it's time to implement the changes. However, to make its marketing efforts more effective, the store can use data mining to understand where its clients see ads, what demographics to target, where to place digital ads, and what marketing strategies most resonate with customers. This includes aligning marketing campaigns, promotional offers, cross-sell offers, and programs to the findings of data mining.

Manufacturing
For companies that produce their own goods, data mining plays an integral part in analyzing how much each raw material costs, what materials are being used most efficiently, how time is spent along the manufacturing process, and what bottlenecks negatively impact the process. Data mining helps ensure the flow of goods is uninterrupted.

Fraud Detection

The heart of data mining is finding patterns, trends, and correlations that link data points together. Therefore, a company can use data mining to identify outliers or correlations that should not exist. For example, a company may analyze its cash flow and find a reoccurring transaction to an unknown account. If this is unexpected, the company may wish to investigate whether funds are being mismanaged.

Human Resources

Human resources departments often have a wide range of data available for processing including data on retention, promotions, salary ranges, company benefits, use of those benefits, and employee satisfaction surveys. Data mining can correlate this data to get a better understanding of why employees leave and what entices new hires.

Customer Service

Customer satisfaction may be caused (or destroyed) for a variety of reasons. Imagine a company that ships goods. A customer may be dissatisfied with shipping times, shipping quality, or communications. The same customer may be frustrated with long telephone wait times or slow e-mail responses. Data mining gathers operational information about customer interactions and summarizes the findings to pinpoint weak points and highlight what the company is doing right.

Benefits of Data Mining

- Data mining ensures a company is collecting and analyzing reliable data. It is often a more rigid, structured process that formally identifies a problem, gathers data related to the problem, and strives to formulate a solution. Therefore, data mining helps a business become more profitable, more efficient, or operationally stronger.
- Data mining can look very different across applications, but the overall process can be used with almost any new or legacy application. Essentially any type of data can be gathered and analyzed, and almost every business problem that relies on qualifiable evidence can be tackled using data mining.
- The end goal of data mining is to take raw bits of information and determine if there is cohesion or correlation among the data. This benefit of data mining allows a company to create value with the information they have on hand that would otherwise not be overly apparent. Though data models can be complex, they can also yield fascinating results, unearth hidden trends, and suggest unique strategies.

Limitations of Data Mining

- This complexity of data mining is one of its greatest disadvantages. Data analytics often requires technical skill sets and certain software tools. Smaller companies may find this to be a barrier of entry too difficult to overcome.
- Data mining doesn't always guarantee results. A company may perform statistical analysis, make conclusions based on strong data, implement changes, and not reap any benefits. Through inaccurate findings, market changes, model errors, or inappropriate data populations, data mining can only guide decisions and not ensure outcomes.

6.Explain the web analytics with some examples in detail.

Web analytics is the gathering, synthesizing, and analysis of website data with the goal of improving the website user experience. It's a practice that's useful for

managing and optimizing websites, web applications, or other web products. It's highly data-driven and assists in making high-quality website decisions.

- Web analytics is important to help you:
  - Refine your marketing campaigns
  - Understand your website visitors
  - Analyze website conversions
  - Improve the website user experience
  - Boost your search engine ranking
  - Understand and optimize referral sources
  - Boost online sales

The process of web analytics involves:

- **Setting business goals:** Defining the key metrics that will determine the success of your business and website
- **Collecting data:** Gathering information, statistics, and data on website visitors using analytics tools
- **Processing data:** Converting the raw data you've gathered into meaningful ratios, KPIs, and other information that tell a story
- **Reporting data**: Displaying the processed data in an easy-to-read format
- **Developing an online strategy:** Creating a plan to optimize the website experience to meet business goals
- **Experimenting**: Doing **A/B tests** to determine the best way to optimize website performance

You can use this information to optimize underperforming pages and further promote higher-performing ones across your website.

UNIT-V
1.Explain the concept of prescriptive analytics in BA.

Definition of Prescriptive Analytics

Prescriptive analytics is a type of data analytics that uses statistical algorithms, machine learning techniques, and artificial intelligence to analyze data and provide recommendations on the actions to take to optimize business outcomes. Unlike descriptive and predictive analytics, which focus on describing past events and predicting future outcomes, prescriptive analytics provides actionable

insights on achieving business objectives by suggesting the most optimal course of action. It involves analyzing data from various sources, identifying patterns, and using this information to optimize decision-making processes and improve overall business performance.

Real-world Implementation of Prescriptive Analytics

1. **UPS**: The logistics giant uses prescriptive analytics to optimize its delivery routes and reduce fuel consumption. Using sophisticated algorithms, UPS determines the most efficient ways for its delivery trucks, considering factors such as traffic patterns, weather conditions, and package weight.

2. **Hilton Worldwide**: Hilton uses prescriptive analytics to optimize its hotel room pricing strategy. By analyzing historical data on room occupancy rates, customer preferences, and competitive pricing, Hilton can adjust room rates in real-time to maximize revenue and occupancy.

3. **Royal Bank of Scotland:** The bank uses prescriptive analytics to detect and prevent fraudulent transactions. By analyzing customer transaction data in real time, RBS can identify unusual patterns or suspicious activity and take immediate action to prevent fraud.

4. **The Weather Company**: The Weather Company uses prescriptive analytics to predict the weather and help businesses prepare for severe weather events. Using data from sensors, satellites, and other sources, the company can provide accurate weather forecasts and recommend actions to minimize severe weather's impact.

5. **Procter & Gamble:** P&G uses prescriptive analytics to optimize its supply chain operations. By analyzing demand, inventory levels, and production capacity data, P&G can make real-time decisions on allocating inventory and production resources to maximize efficiency and reduce costs.

Tools Used for Prescriptive Analytics

Tool Type	Examples
Optimization tools	IBM ILOG CPLEX, Gurobi, FICO Xpress
Simulation tools	AnyLogic, Simul8, Arena
Machine learning algorithms	Neural networks, decision trees, regression analysis
Business intelligence tools	Tableau, QlikView, Power BI
Natural language processing (NLP) tools	Google Cloud Natural Language API, Microsoft Azure Cognitive Services, IBM Watson NLU
Prescriptive analytics software	FICO Decision Management Suite, IBM Decision Optimization, SAP Analytics Cloud

Example of Prescriptive Analytics

Recently, a deadly cyclone hit Odisha, India, but thankfully most people had already been evacuated. The Odisha meteorological department had already predicted the arrival of the monstrous cyclone and made the life-saving decision to evacuate the potentially prone regions.

2.Explain the prescriptive analytics modeling.

There are three types of prescriptive process models. They are:

1. The Waterfall Model
2. Incremental Process model
3. RAD model

1. **The Waterfall Model**

- The waterfall model is also called as **'Linear sequential model'** or **'Classic life cycle model'**.

- In this model, each phase is fully completed before the beginning of the next phase.
- This model is used for the small projects.
- In this model, feedback is taken after each phase to ensure that the project is on the right path.
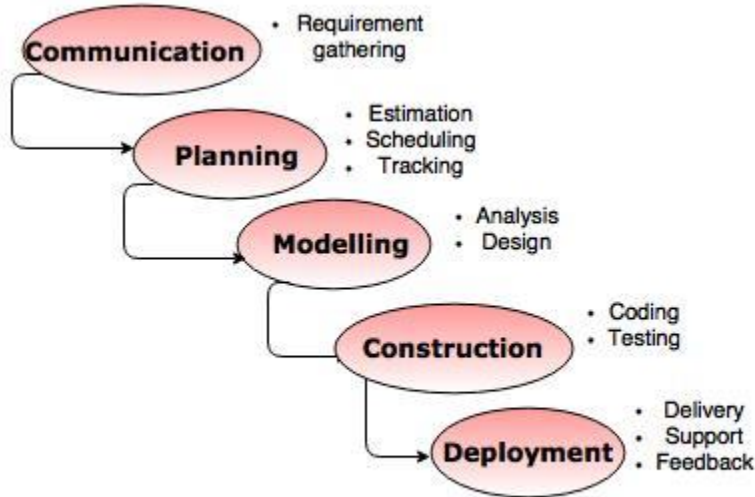- Testing part starts only after the development is complete.

Fig. - The Waterfall model

NOTE: The description of the phases of the waterfall model is same as that of the process model.

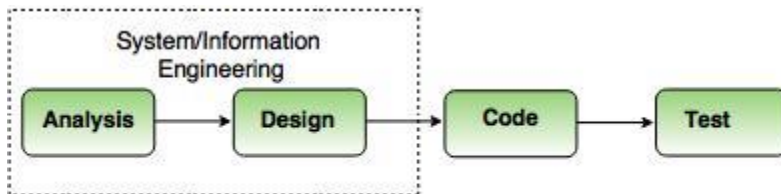**An alternative design for 'linear sequential model' is as follows:**

Fig. - The linear sequential model

Advantages of waterfall model
- The waterfall model is simple and easy to understand, implement, and use.
- All the requirements are known at the beginning of the project, hence it is easy to manage.
- It avoids overlapping of phases because each phase is completed at once.
- This model works for small projects because the requirements are understood very well.
- This model is preferred for those projects where the quality is more important as compared to the cost of the project.

Disadvantages of the waterfall model
- This model is not good for complex and object oriented projects.
- It is a poor model for long projects.
- The problems with this model are uncovered, until the software testing.
- The amount of risk is high.

2. Incremental Process model

- The incremental model combines the elements of waterfall model and they are applied in an iterative fashion.
- The first increment in this model is generally a core product.
- Each increment builds the product and submits it to the customer for any suggested modifications.
- The next increment implements on the customer's suggestions and add additional requirements in the previous increment.
- This process is repeated until the product is finished. **For example,** the word-processing software is developed using the incremental model.
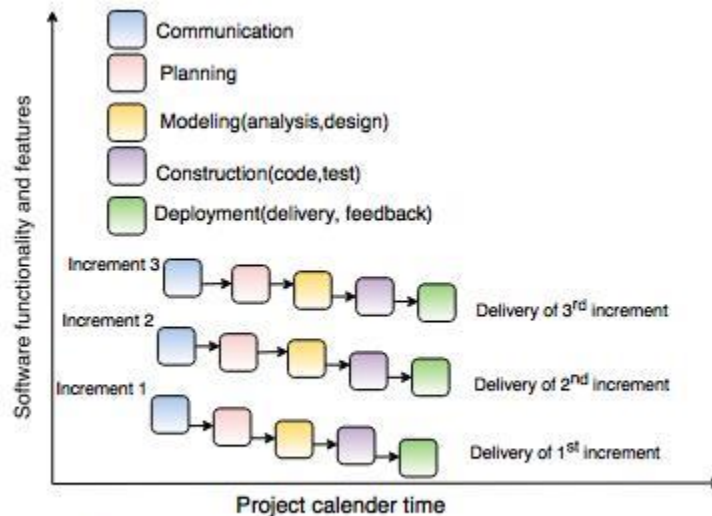

Fig. - Incremental Process Model

Advantages of incremental model
- This model is flexible because the cost of development is low and initial product delivery is faster.
- It is easier to test and debug during the smaller iteration.
- The working software generates quickly and early during the software life cycle.
- The customers can respond to its functionalities after every increment.

Disadvantages of the incremental model
- The cost of the final product may cross the cost estimated initially.
- This model requires a very clear and complete planning.
- The planning of design is required before the whole system is broken into small increments.
- The demands of customer for the additional functionalities after every increment causes problem during the system architecture.

3. RAD model

- RAD is a Rapid Application Development model.
- Using the RAD model, software product is developed in a short period of time.
- The initial activity starts with the communication between customer and developer.

- Planning depends upon the initial requirements and then the requirements are divided into groups.
- Planning is more important to work together on different modules.

The RAD model consist of following phases:

1**. Business Modeling**
- Business modeling consist of the flow of information between various functions in the project.
- For example what type of information is produced by every function and which are the functions to handle that information.
- A complete business analysis should be performed to get the essential business information.

2. Data modeling
- The information in the business modeling phase is refined into the set of objects and it is essential for the business.
- The attributes of each object are identified and define the relationship between objects.

3. Process modeling
- The data objects defined in the data modeling phase are changed to fulfil the information flow to implement the business model.
- The process description is created for adding, modifying, deleting or retrieving a data object.

4. Application generation
- In the application generation phase, the actual system is built.
- To construct the software the automated tools are used.

5. Testing and turnover
- The prototypes are independently tested after each iteration so that the overall testing time is reduced.
- The data flow and the interfaces between all the components are are fully tested. Hence, most of the programming components are already tested.
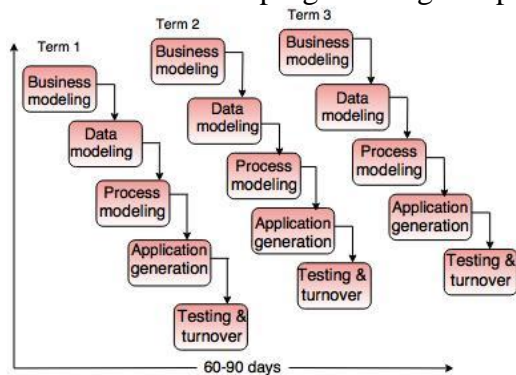

Fig. - RAD Model

3.Explain the components of data mining.
Data Mining Architecture Components

1. **Sources of Data**

The place where we get our data to work upon is known as the data source or the source of the data. There are many documentations presented, and one might also argue that the whole **World Wide Web (WWW)** is a big data warehouse. The data can be anywhere, and some might reside in text files, a standard spreadsheet document, or any other viable source like the internet.

2. **Database or Data Warehouse Server**

The server is the place that holds all the data which is ready to be processed. The fetching of data works upon the user's request, and, thus, the actual datasets can be very personal.

3. **Data Mining Engine**

The field of data mining is incomplete without what is arguably the most crucial component of it, known as a data mining engine. It usually contains a lot of modules that can be used to perform a variety of tasks. The tasks which can be performed can be association, characterization, prediction, clustering, classification, etc.

4. **Modules for Pattern Evaluation**

This module of the architecture is mainly employed to measure how interesting the pattern that has been devised is actually. For the evaluation purpose, usually, a threshold value is used. Another critical thing to note here is that this module has a direct link of interaction with the data mining engine, whose main aim is to find interesting patterns.

5. **GUI or Graphical User Interface**

As the name suggests, this module of the architecture is what interacts with the user. GUI serves as the much-needed link between the user and the system of data mining. GUI's main job is to hide the complexities involving the entire process of data mining and provide the user with an easy to use and understand module which would allow them to get an answer to their queries in an easy-to-understand fashion.

6. Knowledge Base

The base of all the knowledge is vital for any data mining architecture. The knowledge base is usually used as the guiding beacon for the pattern of the results. It might also contain the data from what the users have experienced. The data mining engine interacts with the knowledge base often to both increase the reliability and accuracy of the final result. Even the pattern evaluation module has a link to the knowledge base. It interacts with the knowledge base on a regular interval to get various inputs and updates from it.

4. Explain the uses of prescriptive analytics with an example.

1.Use of Prescriptive Analytics in Hospitals and Clinics:

One of the best **prescriptive analytics examples** is its use in the healthcare sector. Hospitals and clinics use prescriptive analytics to enhance the outcomes for patients. It uses healthcare data to assess the profitability of different processes and treatments. Moreover, it can assess the official clinical methods.

Explore our US Business Analytics Programs from World's Top Universities

Prescriptive Analytics can investigate which hospital patients have the maximum risk of re-admission. Based on this analysis, it instructs healthcare providers to keep the re-admission to the hospital or emergency room at bay.

2. Use of Prescriptive Analytics for Airlines:
Prescriptive Analytics helps airlines' CEOs maximize their company's profits. It automatically modifies ticket prices and accessibility depending on weather, customer demand, and gasoline prices.

For example, the Prescriptive Analytics algorithm can analyze whether the current year's Christmas ticket sales from New York to Los Angeles are lagging or leading compared to last year. Based on this analysis, it automatically lowers prices while also considering the higher fuel prices.

3. Use of Prescriptive Analytics in Banking, Financial Services, and Insurance (BFSI):
You can find various **prescriptive analytics examples** when it comes to financial institutions. These institutions can propose Prescriptive Analytics algorithms for managing risk and cost-effectiveness by scrutinizing historical trading data. Certain insurance companies also use risk assessment models to offer better premium information regarding insurance policies for clients.

4. Use of Prescriptive Analytics in manufacturing:
Big production machines can have a broad range of minor changes. Price prediction is inevitable to tackle these changes. Prescriptive Analytics can accurately predict current production, material handiness, power consumption, and more. It can also help optimize productive capacity, conforming to the delivery schedule and consolidating final assembly lines.

Manufacturers can use Prescriptive Analytics to model prices on different factors like storage, production, and discoveries. It helps determine the optimum settings to boost yield without compromising efficiency.

5. **Use of Prescriptive Analytics for sales and marketing:**

Prescriptive modeling is a mathematical process that benefits brands aiming to reinforce their marketing techniques. It can help run promotional campaigns and predict the customer interests and segments' consumption.

6. Use of Prescriptive Analytics in Supply Chain and Logistics:

Prescriptive Analytics is essential for route optimization in the Supply Chain industry. Generally, logistics companies use it to avoid logistical issues like improper shipping locations. They use Predictive Analytics for improved route planning while saving time, money, and resources.

7. Use of Prescriptive Analytics to improve business efficiency:

Prescriptive analytics ensure businesses can save time and use data to develop a process that will make them stand out from their competitors. Business efficiency significantly increases with the use of Cloud-based prescriptive analytics tools.

8. Use of Prescriptive Analytics in creating Data governance strategy:

Prescriptive analytics also permits a degree of caution from the viewpoint of ethics. For example, generating automated recommendations or decisions depending on a computer's student data analysis can raise questions about privacy and impartiality, such as– Do learners provide consent? Who can access the data and results?

The learner's predictions can be inaccurate if the collected data is not entirely precise. This can lead to wrong decisions or recommendations about the learner. A data governance strategy can be implemented, and the prescriptive analytics models can be used to emphasize validation.

9. Examples of prescriptive analytics in online learning:

Prescriptive Analytics is extensively used in specific learning management systems (LMS) and learning technologies. The following points clarify how it enhances online learning:

Certain online learning tools use prescriptive analytics to recognize the content learned. These tools present content yet to be mastered. So, it is one of the best **prescriptive analytics examples** of exploring adaptive learning.

5.Describe the best practices implementing prescriptive analytics and explain the merits and demerits of its.

Best Practices for Implementing Prescriptive Analytics

In order to successfully implement and overcome the challenges/disadvantages of prescriptive analytics, there are best practices that businesses should follow. These includes:

- **Develop a Clear Understanding of the Problem:** Companies should clearly define what they are trying to fix with standardized assessments. This may include identifying key metrics and outcomes they are trying to to achieve.
- **Identify Data Sources:** The success of prescriptive analytics depends on the data it uses. Therefore, it is important to identify the data sources that are most relevant to the business problem being solved.
- **Build a Cross-Functional Team:** Business should build a team of people with the skills required for analytics like data analysis, machine learning applications, statistics etc. Therefore, businesses should build cross functional team with these necessary skills to ensure successful implementation of prescriptive analytics.
- **Start Small and Scale up:** Prescriptive analytics can be complex and expensive to implement. Therefore, it is important to start with a small project and gradually move or scale up once you gain experience and confidence. This allows businesses to minimize risks and test the effectiveness of prescriptive analytics.
- **Monitoring and Adjusting Algorithm:** Prescriptive analytics is not something that you can just set up and then forget about. It needs to be regularly monitored and adjusted as needed to ensure that it always provides accurate recommendations and insights.

Advantages of Prescriptive Analytics

- **Improved Decision-Making:** Prescriptive analytics provide recommendations or actionable insights to businesses that can be used to make informed decisions. It provides recommendations that are based on data, rather than intuition or guesswork which helps businesses to take decisions that are more likely to achieve their desired outcomes.
- **Increased Efficiency:** Prescriptive analytics can help organizations to optimize their allocation of resources which results in increases efficiency, reduced costs and improved productivity. By identifying inefficiencies and recommending changes, prescriptive analytics can help businesses streamline their operations and save time and money.
- **Enhanced Customer Experience:** Prescriptive analytics helps businesses to enhance customer services by providing personalized recommendations, predicting and preventing issues and engaging proactively with customers. This leads to improved customer satisfaction and loyalty.
- **Competitive Advantage:** By using prescriptive analytics, businesses can optimize their operations and make better decisions. Also, it helps them in improving their products, services and processes which eventually helps them to gain a competitive edge in the market and stay ahead of their competitors.

Disadvantages of Prescriptive Analytics

Although prescriptive analytics provide valuable insights, but it also have some disadvantages that businesses should be aware of. Here are some of the disadvantages of prescriptive analytics:

- **Data Quality**: In order to provide accurate insights and recommendations, prescriptive analytics relies on high quality data, which can be difficult to obtain and clean. If the data collected for analysis is incomplete, outdated or inaccurate, the recommendations generated may also be unreliable.
- **Complexity:** Implementing prescriptive analytics is a complex process which requires advanced technical skills and expertise. Therefore, it becomes difficult for some businesses to execute it effectively.
- **Cost:** Prescriptive analytics requires huge investment in technology and skilled personnel. Therefore, implementing and executing prescriptive analytics can be costly, particularly for small and medium sized business or those with limited resources.
- **Resistance to Change:** Prescriptive analytics can require significant changes in existing business processes and practices. And, it may face resistance from employees or stakeholders who are resistant to change, which can make it difficult for businesses to implement and adopt new processes and tools.