Data Analytics – 315326

UNIT 1 Introduction to Data Analytics

Q.1 Explain Data Analytics

- Data Analytics is a kind of branch of computer science in which procedure can be implement to analyze unstructured and structured data to obtain insightful, relevant information that can be used to support smart decisions
- Analyzing data sets to gain information that can be applied to solve problems across various sectors is a component of the discipline of data analytics.
- Importance of Data Analytics :-
  - **Informed Decision-Making:-**
    - Data Analytics empowers businesses to make data-driven decision rather than relying on guesswork or intuition. By analyzing data, business can identify patterns and trends to make predictions, customer needs, and identify new opportunities
    - This leads to more effective strategies, whether in marketing, product development, or operations.
  - **Improved Efficiency and Cost Savings:-**
    - Data analysis helps identify inefficiencies and bottlenecks in business processes, allowing for streamlining and cost reduction.
    - By analyzing resource allocation and process data, organizations can identify areas where they can cut expenses, boost productivity and save time.
  - **Enhanced Customer Experience**:-
    - Data analytics provides insights into customer behaviour, preferences, and needs, allowing businesses to personalize their offerings and interactions.

- By understanding customers better, businesses can tailor their products, services, and marketing efforts to meet their specific needs.

- **Competitive Advantage**
  - Data analytics helps businesses stay ahead of the competition by identifying emerging trends, predicting future demand, and making informed decisions.
  - Organizations that leverage data analytics can gain a competitive edge by optimizing their operations, improving customer experiences, and developing new products and services.
- **Innovation and Growth**
  - Data analytics can uncover new opportunities for growth and innovation by revealing emerging trends, market gaps, and unmet customer needs.

Q.2 Explain types of Data Analytics

Types of Data Analytics are as follows:-

1. **Descriptive Analytics:-** The form of analysis known as descriptive analytics is straightforward and focuses on the surface level of previous events
2. **Diagnostic Analytics:**- Diagnostics analytics focuses on why of certain occurrences or patterns rather than just the what of descriptive analytics. Further in-depth data analysis is required to identify the variables and reasons that resulted in certain results
3. **Prescriptive Analytics:**- it makes recommendations for the courses of action and choices to be made
4. **Predictive Analytics :**- It seeks to forecast the expected course of events, as the name indicates.

Q.3 Explain Descriptive Analytics

a. It is a statistical interpretation used to analyze historical data to identify the patterns and relationships.

b. It seeks to describe an event, phenomenon, or outcome.

c. Descriptive Analytics is the process of using current and historical data to identify trends and relationships.

d. Its also called as the simples form of data analytics because it describes trends and relationships but doesn't dig deeper.

e. It is especially useful for communicating change over time and uses trends as a springboard for further analysis to drive decision-making.

f. Examples of Descriptive Analytics:-

1. Traffic and Engagement:-
    a. If the organization tracks engagement in the form of social media analytics or web traffic, you are already using descriptive analytics.
    b. These reports are created by taking raw data generated when users interact with your website, advertisements, or social media content and using it to compare current metrics to historical metrics and visualize trends.
    c. Using descriptive analytics, you can analyze the pages traffic data to determine the number of users for each source
2. Financial Statement Analysis:-
    a. Descriptive analytics serve as essential tools for assessing companies financial health, providing insights into its performance over time.
    b. By examining various types of financial statements and the methods of analysis, we can understand how businesses utilize these reports to inform decision-making and strategy
    c. Types of financial statements
        i. Balance Sheet

ii. Income Statement.
iii. Cash Flow statement
iv. Statement of shareholders Equity

Q.4 Explain Diagnostics Analytics

1. It is a branch of data analysis that focuses on identifying the reasons behind trends, patterns, and anomalies in data.
2. Key features of Diagnostic Analytics :-
   a. Root Cause Analysis(RCA):- Determines the underlying causes of business performance issues or successes
   b. Drill-Down and Data Mining:- Explores data at multiple levels to pinpoint specific factors influencing outcomes
   c. Correlation and Causation :- Uses statistical methods to determine relationships between variables
   d. Comparative Analysis:- Examines performances over time, across different segments, or against benchmarks
3. Common Techniques in Diagnostics Analytics:-
   a. Regression Analysis:- identifies relationships between dependent and independent variables
   b. Hypothesis Testing:- Validates assumptions using statistical methods
   c. Machine Learing Models :- Predicts outcomes based on historical patterns.

Q.5 Explain Prescriptive Analytics

1. It is the most advanced form of data analytics, providing actionable recommendations based on historical data, machine learning, and optimization algorithms.

2. It goes beyond descriptive and diagnostic analytics by suggesting the best course of action to achieve desired outcomes

e3. Key features of Prescriptive Analytics:-

a. Decision Optimization :- Suggests the best actions based on available data

b. Predictive Modelling:- Uses machine learning to forecast future outcomes

c. Scenario Analysis:- Evaluates different strategies to determine optimal results.

d. Real-Time Recommendations:- Continuously adapts based on live data.

4. Examples of prescriptive analytics:-

a. Retail:- Suggesting optimal pricing and promotions to maximize sales

b. Healthcare:- Recommending personalized treatment plans for patients.

c. Supply chain: Identifying the best logistics route to minimize delays and costs

d. Finance:- Advising investment strategies to maximize returns

Q.6 Explain Predictive analytics

1. It is a data-driven approach that uses statistical techniques, machine learning, and historical data to forecast future trends, behaviors, and outcomes.

2. Key features of predictive Analytics :-

a. Forecasting Trends:- Uses historical data to predict future patterns

b. Risk Assessment:- Identifies potential risks and opportunities

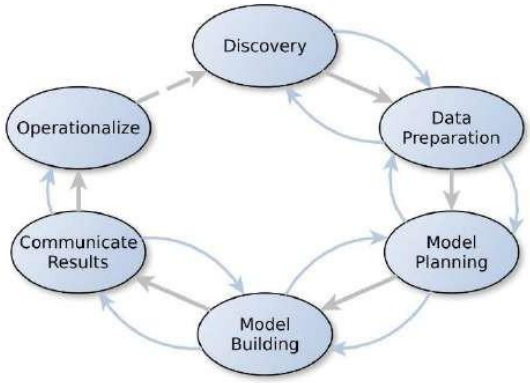
c. Customer behavior prediction:- Anticipates customer preferences and actions

d. Fraud Detection:- Detects anomalies to prevent fraud

e. Market Demand Prediction:- Helps businesses prepare for changes in demand

Q.7 Explain Life Cycle of Data Analytics

1. Data Analytics Life Cycle is designed for Big Data Problems and data science projects.



2. Phase 1:- Discovery

a. Data science team learns and investigates the problem.

b. Develop context and understanding.

c. The team learns business domain, including relevant history, such as business unit has attempted similar project in term of people technology, time, and data. Important activities in this phase including framing the businesss problem as an analytics challenge that can be addressed in subsequent phase and formulating initial hypothesis to test and begin learning the data.

3. Phase 2:- Data Preparation

a. In this team work with data and perform analytics for the duration of the project.

b. Steps to explore preprocess and condition data before modelling and analysis.

c. It requires the presence of an analytic sandbox, the team executes, loads and transforms to get data into the sandbox.

d. Data preparation tasks are likely to be performed multiple times and not in predefined order.

4. Phase 3:- Model Planning

a. The team explores data to learn about relationships between variables and subsequently, selects key variables and the most suitable methods.

b. In this phase, the data science team develops data sets for training, testing and production purposes. Team builds and executes models based on the work done.

5. Phase 4:- Model building

a. Team develops datasets for testing, training and production purposes.

b. Team also considers whether its existing tools will suffice for running the models or if they need more robust environment for executing models .

6. Phase 5:- Communication Results

a. After executing model team need to compare outcomes of modeling to criteria established for success and failure.

b. Team considers how best to articulate findings and outcomes to various team members and stakeholders, taking into account warning, assumptions.

7. Phase 6:- Operationalize

a. Team communicates benefits of project more broadly and sets up pilot project to deploy work in controlled way before broadening the work to full enterprise of users.

b. This approach enables team to learn about performance and related constraints of the model in production environment on small scale which make adjustments before full deployment.

c. The team delivers final reports, briefings and codes.

Q.8 Explain Quality and Quantity of the data in analytics

a. Quality of Data:-

1. High quality data ensures accurate, reliable and actionable insights.

2. Key dimensions include:-

- **Accuracy:-** Data should be correct and free from errors.
- **Completeness:** No missing or incomplete values
- **Consistency :** Uniform data across multiple sources
- **Timeliness:** Data should be up to date
- **Relevance:** Data should be meaningful to the analysis

b. Quantity of Data:-

1. Amount of data impacts the depth of analysis

2. Key dimensions include:-

- **Small Data**: suitable for simple statistical analysis, small-scale decision-making

- **Big Data**: Large volumes of structured and unstructured data, requiring specialized tools like Hadoop.
- **Balanced Data**: A sufficient amount of high-quality data leads to better model performance and insights.
- **Optimal Data**: It requires both high-quality and sufficient quantity of data to ensure reliable insights and decision-making

Q.9 Explain Data Types in Statistical Analysis

1. Data types in analytics determine how data is categorized and analyzed.

2. They are classified into as follows:

a. Qualitative (Categorical) Data

- Nominal Data – Categories without any order
- Ordinal Data – Categories with a meaningful order but no fixed interval

b. Quantitative (Numerical) Data

- Discrete Data – Whole numbers / countable values
- Continuous Data – Measured values with infinite precision

Q.10 Explain Measures of Central tendency

1. Measures of central tendency describes the center or typical value of a dataset

2. Mean:- Average of the dataset

Formula :- Mean = sum of datapoints / count of datapoints

3. Median :- The middle value when data is arranged in order

4. Mode:- The most frequent occurring value in a dataset

Q.11 Explain Measures of dispersion

1. Measures of dispersion describe the spread or variability in a dataset

2. Range:- Difference between maximum value and minimum value

3. Variance:- Measures how far data points are from the mean .

4. Standard Deviation :- Square root of variance, measuring the spread in original units

5. Interquartile Range (IQR) :- Measure the 50% of the data

$$IQR = Q3 - Q1$$

Q.12 Explain Sampling Funnel

1. Sampling funnel refers to a structured approach for narrowing down a large dataset or population into a smaller, more representative subset for analysis.

2. Need of a Sampling Funnel:-

a. Efficiency :- Reduces the time and cost of analyzing large datasets

b. Accuracy

c. Scalability :- Allows businesses to test hypothesis before full-scale implementation

3. Stages of Sampling Funnel

a. Population Identification:- Define the entire group from which you want a sample

b. Target population selection :- Identify a specific segment relevant to the analysis

c. Sampling Frame creation :- Develop a list of elements that meet selection criteria

d. Sampling Method selection :- Choose an appropriate sampling technique

e. Sample Extraction:- Apply the chosen method to select a subset of data

f. Data Cleaning and Preparation:- Remove inconsistencies, missing values, or irrelevant entries

g. Analysis and Interpretation:- Conduct statistical or machine learning analysis on the sample. User insights to infer conclusions about the broader population

Q.13 Explain Central Limit Theorem (CLT)

1. CLT States that regardless of the original distribution, the sampling distribution of the sample means will approach a normal distribution as the sample size increases, provided the samples are independent

2. Need of CLT :-

- Even if the population is skewed, we can apply normal distribution assumptions for large samples
- Enables techniques like t-tests and confidence intervals in inferential statistics
- Helps in making data-driven decision based on sample data. If we take random samples of size n from a population with any distribution that has a mean and standard deviation then as the sample size increases, the distribution of the sample mean will approach a normal distribution.

3. Key Principles of CLT:-

a. Normality of sampling distribution

b. Mean of sampling distribution

c. Standard Error

Q.14 Explain applications of data analytics:-

1. Education:- Officials may utilize data analytics to improve management and educational decisions. These programs would improve administrative control and learning. You may gather preference information from each student and utilize it to develop a syllabus to enhance the existing syllabus

2. Delivery and logistics:- Analysis are used in logistics to optimize delivery processes and streamline operations. As a result, the sector has fared better, which has increases the number of customers. It boosts productivity by facilitating real-time data exchange of business insights across partners

3. Digital Marketing and advertising:- Marketers utilize data analytics to understand their audience and achieve high conversion rates. Digital advertising professionals employ analytics to learn target markets demographics, including their age, race and gender. Additionally, they filter their audience using this technology in accordance with their tastes and behaviors

4. Transportation:- The transport sector might be completely transformed thanks to data analytics. It is especially helpful when moving a sizable group to a site demanding smooth movement. By enhancing transportation systems and intelligence, data analytics may be used to reduct traffic congestion and enhance travel

5. Security: -Data analytics, particularly predictive analytics, is used by security staff to forecast future instances of crime or security breaches. Additionally, they can investigate recent or current assaults.