SNS COLLEGE OF TECHNOLOGY

Coimbatore-35
An Autonomous Institution
Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A++' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF MECHATRONICS ENGINEERING

UNIT 2 –DATA ANALYTICS

BASICS & INTERMEDIATE ANALYSIS

BASICS & INTERMEDIATE ANALYSIS

Basics and intermediate analysis in the context of data analytics encompass a range of techniques and methods to explore, understand, and derive insights from datasets. Here's an overview of the basics and intermediate levels of data analysis:

Basics of Data Analysis:

1.Descriptive Statistics:

1. **Definition:** Descriptive statistics summarize and describe the main features of a dataset.
2. **Key Techniques:** Mean, median, mode, range, standard deviation, and percentiles.

2.Data Cleaning:

1. **Definition:** Cleaning involves handling missing values, removing duplicates, and addressing outliers to ensure data quality.
2. **Key Techniques:** Imputation, filtering, and transformation.

3.Exploratory Data Analysis (EDA):

1. **Definition:** EDA involves visually and statistically exploring datasets to understand their characteristics.
2. **Key Techniques:** Histograms, box plots, scatter plots, and summary statistics.

4.Data Visualization:

1. **Definition:** Visualizing data using charts and graphs to facilitate understanding.
2. **Key Techniques:** Bar charts, pie charts, line charts, scatter plots, and heatmaps.

5.Data Transformation:

1. **Definition:** Transforming data to meet analysis requirements.
2. **Key Techniques:** Normalization, standardization, and log transformation.

BASICS & INTERMEDIATE ANALYSIS

Intermediate Data Analysis:

1.Inferential Statistics:

1. **Definition:** Inferential statistics make inferences about a population based on a sample of data.
2. **Key Techniques:** Hypothesis testing, confidence intervals, and regression analysis.

2.Correlation and Causation:

1. **Definition:** Analyzing the relationship between variables and distinguishing between correlation and causation.
2. **Key Techniques:** Pearson correlation coefficient, Spearman rank correlation, and causal inference methods.

3.Advanced Data Visualization:

1. **Definition:** Creating more sophisticated visualizations to uncover complex patterns and trends.
2. **Key Techniques:** Treemaps, bubble charts, radar charts, and interactive dashboards.

4.Time Series Analysis:

1. **Definition:** Analyzing data collected over time to identify patterns and trends.
2. **Key Techniques:** Time series decomposition, forecasting, and autocorrelation analysis.

5.Clustering and Segmentation:

1. **Definition:** Grouping similar data points together to discover underlying patterns.
2. **Key Techniques:** K-means clustering, hierarchical clustering, and DBSCAN.

6.Classification and Regression:

1. **Definition:** Building models to predict categorical (classification) or continuous (regression) outcomes.
2. **Key Techniques:** Logistic regression, decision trees, random forests, and support vector machines.

BASICS & INTERMEDIATE ANALYSIS

1.Dimensionality Reduction:

1. **Definition:** Reducing the number of features in a dataset while preserving important information.
2. **Key Techniques:** Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE).

2.Statistical Modeling:

1. **Definition:** Using statistical models to understand relationships within data.
2. **Key Techniques:** Generalized Linear Models (GLM), Poisson regression, and Bayesian statistics.

3.A/B Testing:

1. **Definition:** Conducting experiments to compare two versions of a variable to determine which performs better.
2. **Key Techniques:** Hypothesis testing and statistical significance.

4.Machine Learning Evaluation:

1. **Definition:** Assessing the performance of machine learning models.
2. **Key Techniques:** Precision, recall, F1-score, area under the Receiver Operating Characteristic (ROC) curve.

BASICS & INTERMEDIATE ANALYSIS

Integration of Basics and Intermediate Analysis:

1.Iterative Process:

1. Basics and intermediate analysis often occur iteratively, with insights from basic analyses guiding the development of more complex analyses.

2.Holistic Approach:

1. A holistic approach involves combining various techniques to gain a comprehensive understanding of the data and derive actionable insights.

3.Communication:

1. Effectively communicating findings to stakeholders is crucial, involving the use of clear visualizations and concise summaries.

4.Domain Knowledge:

1. Incorporating domain knowledge enhances the analysis process, allowing for more meaningful interpretations and insights.