

Department : CSE,AI

Year & Semester : 3rd year 5th sem

Sub Code & Sub Name : 23CSD351B & Exploratory Data Analysis

Unit-I

S.No	Part-A Questions
1.	What is Exploratory Data Analysis (EDA)?
2.	Mention two main goals of EDA.
3.	What is the importance of Data Science in modern industries?
4.	What is a NumPy array?
5.	Define a Pandas DataFrame.
6.	How do you check the shape of a dataset in Pandas?
7.	What is the use of the .describe() function in Pandas?
8.	What is Anaconda, and why is it used?
9.	Define Series in Pandas.
10.	Write the difference between Jupyter Notebook and VS Code.

S.No	Part-B Questions
1.	Explain the role and importance of EDA in the Data Science life cycle with examples.
2.	Discuss the steps involved in setting up the Python environment using Anaconda, Jupyter Notebook, and VS Co
3.	Explain the structure and operations of NumPy arrays with examples.
4.	Describe Pandas data structures – Series and DataFrame – with examples
5.	Demonstrate data loading and basic operations (info, describe, shape) on a CSV file using Pandas.

Unit-II

S.No	Part-A Questions
1	What is data wrangling?
2	Name any two methods to handle missing data
3	How do you remove duplicate records in Pandas?
4	Define outlier and give an example.
5	What is Label Encoding?
6	What is the purpose of One-Hot Encoding
7	Define Normalization.
8	What is data type casting in Python?
9	What is binning in data preprocessing?

10	Mention any two scaling techniques.
----	-------------------------------------

S.No	Part-B Questions
1	Explain in detail the techniques to handle missing data with suitable examples.
2	Discuss methods to detect and handle outliers in datasets.
3	Explain categorical data encoding techniques (Label and One-hot Encoding) with Python examples
4	Describe the importance of data transformation – scaling, normalization, and binning – in preprocessing
5	Write a Python program to demonstrate data cleaning and type conversion on a real-world dataset.

Unit-III

S.No	Part-A Questions
1.	Define mean, median, and mode.
2.	What is a histogram used for?
3.	Define variance and standard deviation
4.	What is the use of a boxplot?
5.	What is a scatter plot?
6.	What does a heatmap represent?
7.	Define correlation and covariance
8.	What is a pair plot?
9.	Which Seaborn function is used to create a count plot?
10.	What is the purpose of using bivariate analysis?

S.No	Part-B Questions
1.	Explain measures of central tendency and dispersion with examples.
2.	Compare and contrast univariate and bivariate analysis.
3.	Illustrate different types of plots used in univariate analysis with code examples.
4.	Explain correlation and covariance analysis. How are they used in EDA?
5.	Using Python, demonstrate how to visualize relationships between two variables using scatter plots, pair plots, and heatmaps.

Unit-IV

S.No	Part-A Questions
1	What is Matplotlib?
2	What is Seaborn used for?
3	Define subplot.
4	What is a Violin plot?
5	Differentiate between strip plot and swarm plot.
6	Mention any two customization options in Matplotlib.
7	What is the purpose of legends and labels in visualization?
8	Define multivariate visualization
9	What is Plotly used for?
10	What is the difference between static and interactive plots?

S.No	Part-B Questions
1	Explain the features and advantages of Matplotlib and Seaborn libraries.
2	Write a Python program to customize a Matplotlib plot with title, legends, and axis labels.
3	Discuss advanced visualization techniques: Violin Plots, Strip Plots, and Swarm Plots with examples.
4	Explain multivariate visualization and the use of subplots in EDA.
5	Describe interactive visualization using Plotly and give examples of its basic functionalities.

Unit-V

S.No	Part-A Questions
1	Name two popular datasets used for EDA practice.
2	What is feature engineering?
3	Define outlier detection
4	What is the purpose of an EDA report?
5	Mention any two Python libraries used for EDA.
6	What is the target variable in the Titanic dataset?
7	Why is EDA important before model building?
8	What does data cleaning involve?
9	What are anomalies in datasets?
10	What is the importance of preparing data for ML models?

S.No	Part-B Questions
1	Perform step-by-step EDA on the Titanic dataset and explain your findings.
2	Discuss various outlier detection techniques with examples.
3	Explain different feature engineering techniques used in EDA.
4	Describe the process of generating an EDA report using Python notebooks.
5	Explain how EDA helps in preparing data for machine learning models using a real-world example.