Lab-02

Aim : Data Preprocessing and Visualization

This lab focuses on performing data cleaning, preprocessing, and visualization to prepare datasets for machine learning tasks. Students will explore real-world datasets to identify and handle missing values, remove inconsistencies, and transform data into a suitable format for analysis. The lab also covers basic preprocessing techniques such as encoding categorical variables and scaling numerical features. In addition, students will use data visualization methods to examine data distributions, detect outliers, and understand relationships between features. These steps help ensure data quality and provide meaningful insights, forming a strong foundation for building accurate and reliable machine learning models.

Experiment 1: Write a Python program to Load a dataset using Pandas Identify missing values using isnull() and notnull()**.** Display the total count of missing values in each column.

Experiment 2: Write a program to Read a CSV dataset Remove rows containing missing values. Compare the dataset size before and after cleaning.

Experiment 3: Write a Python program to load a dataset with missing values Replace missing numerical values using fillna() with mean or median**.** Replace missing categorical values using appropriate constants.

Experiment 4: Write a program to load a dataset, Display column data types, Convert selected columns to appropriate data types using astype(). Verify the changes.

Experiment 5: Write a Python program to load a dataset, rename column names using rename() for better readability. Display the updated DataFrame.

Experiment-6 : Write a program to identify inconsistent or incorrect values in a dataset Replace them using the replace() function Verify the corrected data.

Experiment 7: Write a Python program to load a dataset. Create a line plot to show trends over time. Create a bar plot to compare categorical data.

Experiment 8: Write a program to load a dataset. Create a scatter plot between two numerical features. Interpret the relationship between the features.

Experiment-9 : Write a Python program to load a dataset. Plot a histogram for a selected numerical feature. Analyse data distribution and skewness.

Experiment 10: Write a program to load a dataset. Create a boxplot for numerical attributes. Identify and interpret potential outliers.