

Cycle-4

1. Create a Pandas DataFrame with some missing values (NaN) in different columns.

Print the original DataFrame.

Demonstrate different techniques for handling missing data:

a. Drop rows with missing values.

b. Fill missing values with a specified constant.

c. Fill missing values using forward fill and backward fill methods.

Print the DataFrame after each operation.

2. Create a Pandas DataFrame with numerical columns representing different features. Use Scikit-learn to scale the data using:

StandardScaler (mean=0, variance=1).

MinMaxScaler (scaling values between 0 and 1).

Apply normalization to the data and print the transformed DataFrame.

3. Create a Pandas DataFrame with categorical data columns. Convert categorical data into numerical data using: Label Encoding and One-Hot Encoding.

4. Write a Python program using Scikit-learn to perform the following tasks: Create a dataset with multiple features (at least 5) using Pandas. Apply Principal Component Analysis (PCA) to reduce the dataset to 2 principal components. Print the transformed dataset with the reduced dimensions.

5. Create two separate Pandas DataFrames with a common key column and Merge the DataFrames using: Inner join, Left join, Right join

6. Create a Pandas DataFrame with multiple columns, including a categorical column for grouping. Group the data by the categorical column and perform aggregation operations:

- Calculate the mean of numerical columns for each group.
- Count the number of occurrences in each group.
- Calculate the sum of numerical columns for each group