LAB 3

DATA PREPARATION AND WRANGLING WITH PANDAS

Dataset: titanic.csv

Lab3_starting.ipynb

1. Data Loading and Inspection (10 points)

You are provided with the Titanic dataset in a CSV file named "titanic.csv." Load the dataset into a Pandas DataFrame and answer the following questions:

a) Display the first 5 rows of the dataset. (3 points)

b) Display the data types of each column. (3 points)

c) Calculate the summary statistics (mean, median, standard deviation) for the "Age" column. (4 points)

2. Data Cleaning (15 points)

Perform data cleaning on the Titanic dataset:

a) Check for missing values in the dataset and count the number of missing values in each column. (5 points)

b) Check the number of rows; Remove rows with missing values in the "Age" column. Check the number of rows after removing missing values. (5 points)

c) Replace missing values in the "Embarked" column with the mode of that column. Check for the missing value in the "Embarked" column. (5 points)

LAB 3

3. Data Transformation (20 points)

Perform data transformation tasks on the Titanic dataset:

a) Create a new column “FamilySize” that represents the total number of family members on board for each passenger (sum of “SibSp” and “Parch” columns). Display the DataFrame with the new “FamilySize” column.(8 points)

b) Convert the “Sex” column to binary values, where “Male” becomes 1 and “Female” becomes 0. Display the first five records of the “Sex” Column. (8 points)

c) Calculate the average “Fare” for each “Pclass” (passenger class). (4 points)

4. Data Filtering and Export (15 points)

Filter the dataset and export it to a new CSV file:

a) Filter the dataset to include only passengers with “Fare” greater than 50. Display the last 5 records of passengers with a 'Fare' greater than 50 from the filtered DataFrame. (8 points)

b) Export the filtered dataset to a new CSV file named “filtered_titanic_FirstLastName.csv.” (7 points)