Name : krishnapri

Roll no : 391001

Prn: 22211528

ASSIGNMENT 01:

1. Using the [Titanic dataset]( https://www.kaggle.com/c/titanic/data), create a Python class

to perform basic exploratory data analysis (EDA). Implement methods to visualize the

distribution of survival rates based on different features such as ‘Pclass’, ‘Sex’, and

‘Age’. Use Matplotlib for visualization.

Requirements:

o Create a class `TitanicEDA` with methods to load data, generate summary

statistics, and create visualizations.

o Use Pandas for data manipulation.

o Visualize the distribution of survival rates and other features.

o Save visualizations as image files.

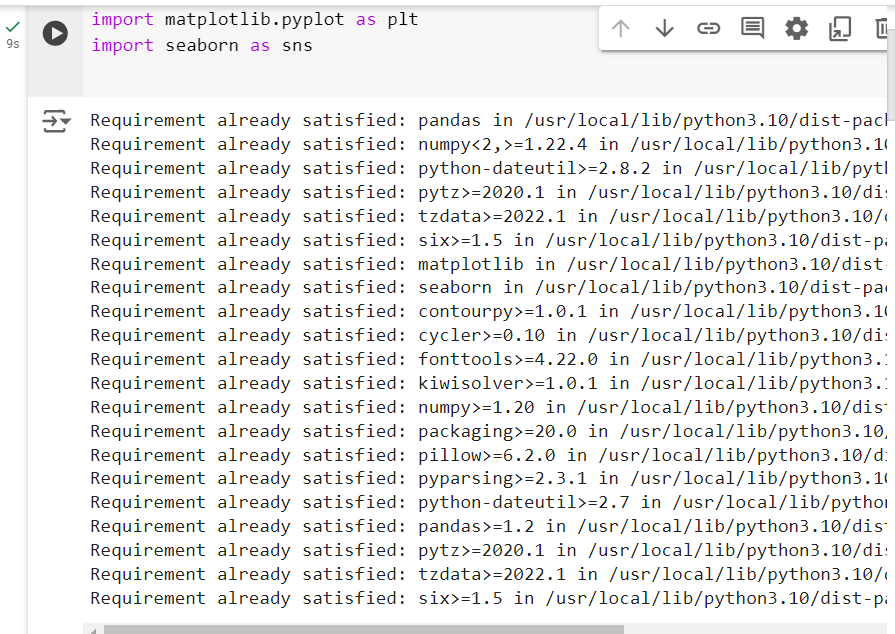
!pip install pandas

!pip install matplotlib seaborn

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

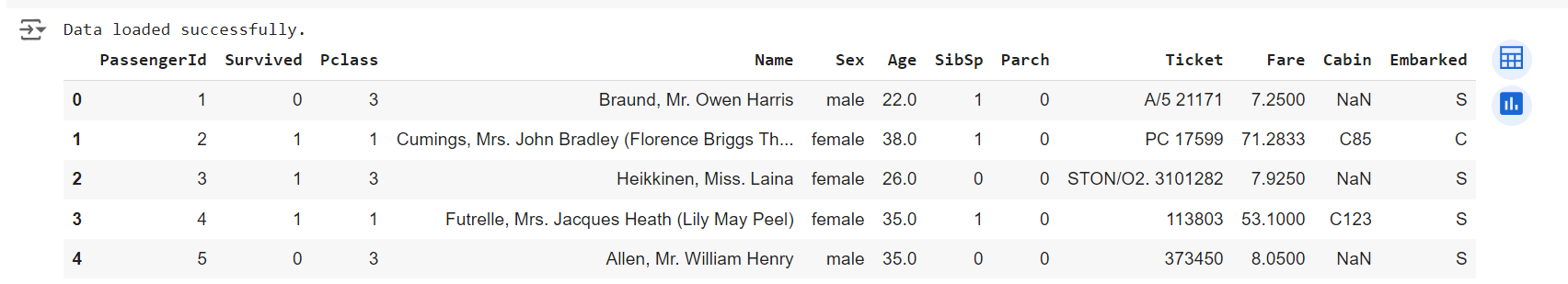


file\_path = '/content/train.csv'

data=pd.read\_csv(file\_path)

print("Data loaded successfully.")

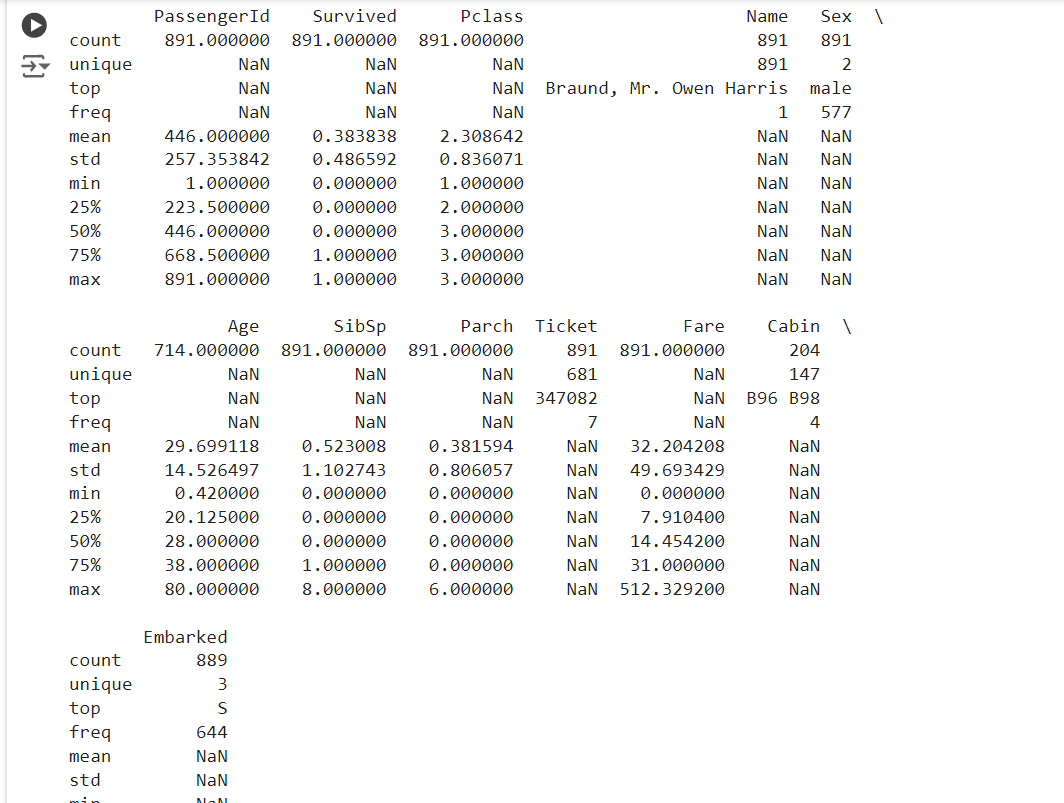
data.head()



def summary\_statistics(data):

    print(data.describe(include='all'))

summary\_statistics(data)



o Visualize the distribution of survival rates and other features.

def visualize\_survival\_rate(data, feature):

    plt.figure(figsize=(18, 4))

    sns.barplot(x=feature, y='Survived', data=data, ci=None)

    plt.title(f'Survival Rate by {feature}')

    plt.ylabel('Survival Rate')

    plt.savefig(f'survival\_rate\_by\_{feature}.png')

    plt.show()

visualize\_survival\_rate(data, 'Pclass')

visualize\_survival\_rate(data, 'Sex')

def visualize\_age\_distribution(data):

    plt.figure(figsize=(8, 4))

    sns.histplot(data, x='Age', hue='Survived', multiple='stack', kde=True)

    plt.title('Age Distribution with Survival Overlay')

    plt.savefig('age\_distribution\_with\_survival.png')

    plt.show()

visualize\_age\_distribution(data)