!pip3 install seaborn openpyxl

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

import pandas as pd

df= pd.read\_excel('train.xls.xlsx', engine='openpyxl')



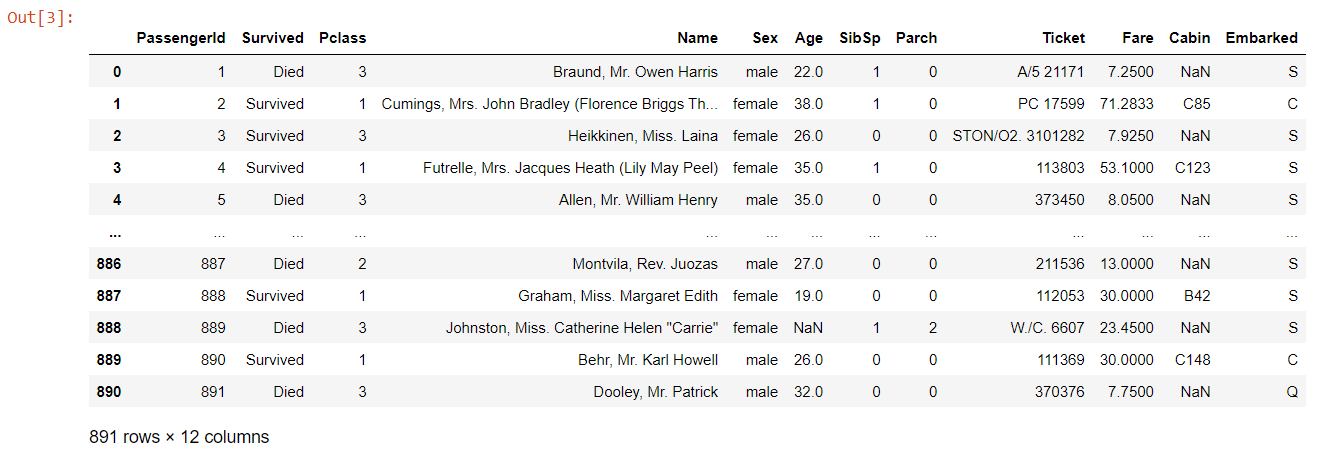
df['Survived'] = df['Survived'].map({

0: 'Died',

1: 'Survived'

})

df

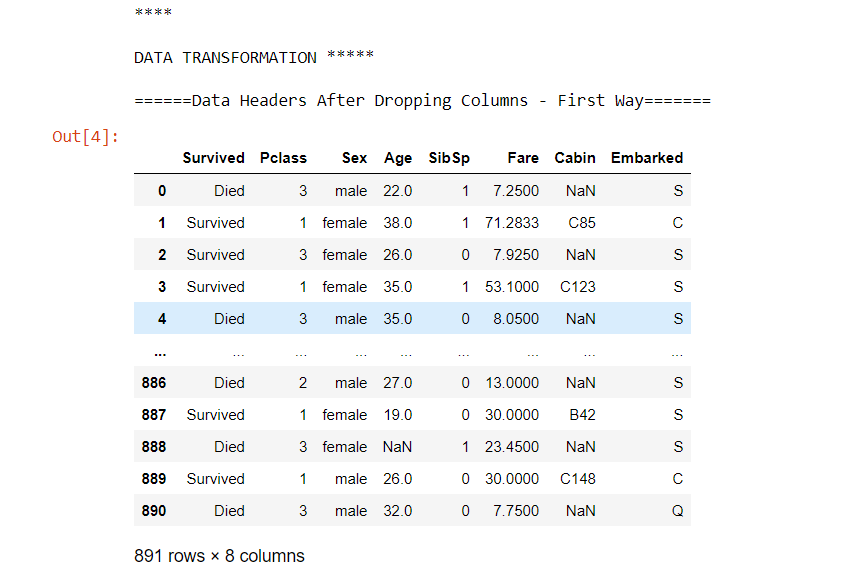


print("\*\*\*\* \n\nDATA TRANSFORMATION \*\*\*\*\*\n")

print("======Data Headers After Dropping Columns - First Way=======")

df.drop(['Parch','PassengerId','Name','Ticket'], axis=1, inplace=True)

df



df['Pclass'] = df['Pclass'].map({

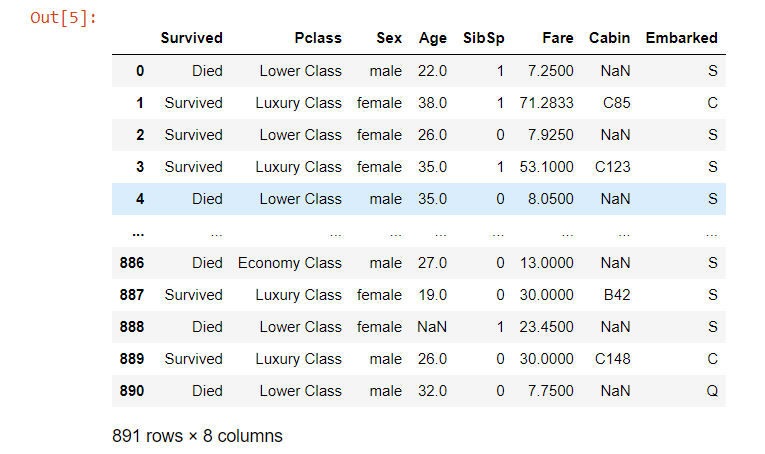
1: 'Luxury Class',

2: 'Economy Class',

3: 'Lower Class'

})

df



df["Embarked"] = df["Embarked"].fillna("S")

df



df['Embarked'] = df['Embarked'].map({

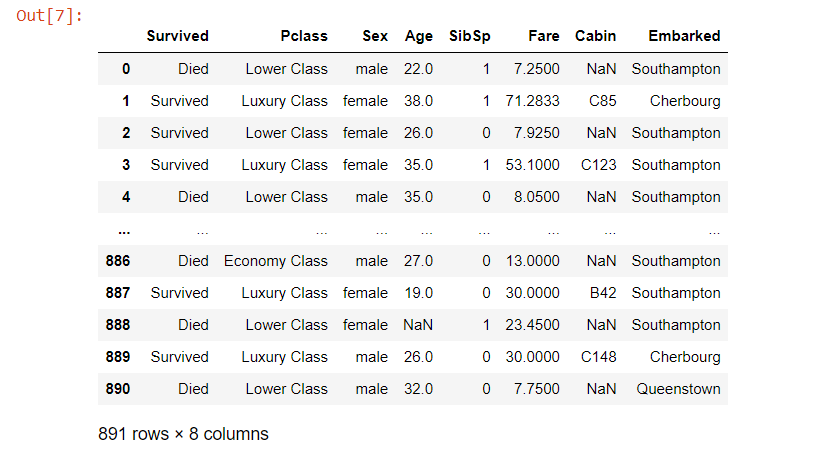
'C':'Cherbourg',

'Q':'Queenstown',

'S':'Southampton'

})

df

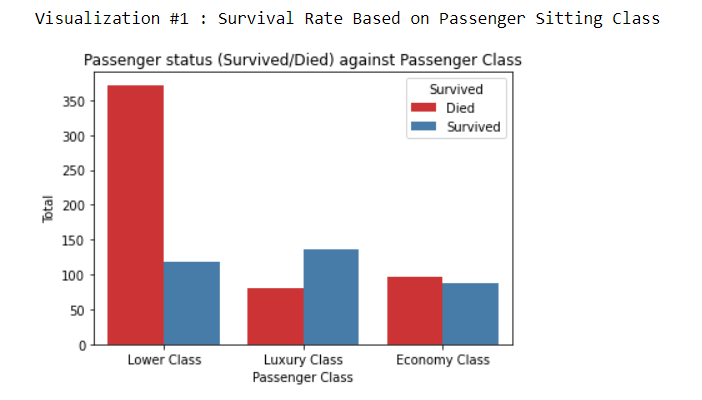


print("Visualization #1 : Survival Rate Based on Passenger Sitting Class")

ax = sns.countplot(x = 'Pclass', hue = 'Survived', palette = 'Set1',data = df)

ax.set(title = 'Passenger status (Survived/Died) against Passenger Class', xlabel = 'Passenger Class', ylabel = 'Total')

plt.show()



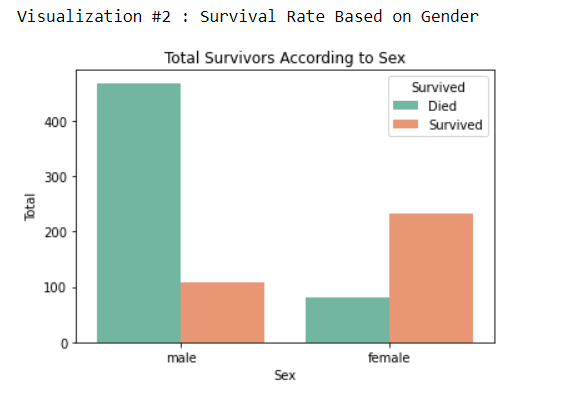
print("Visualization #2 : Survival Rate Based on Gender")

# print(pd.crosstab(df["Sex"],df.Survived))

ax = sns.countplot(x = 'Sex', hue = 'Survived', palette = 'Set2', data = df)

ax.set(title = 'Total Survivors According to Sex', xlabel = 'Sex', ylabel='Total')

plt.show()



print("Visualization #3 : Survival Rate Based on Passenger Age Group")

interval = (0,18,35,60,120)

categories = ['Children','Teens','Adult', 'Old']

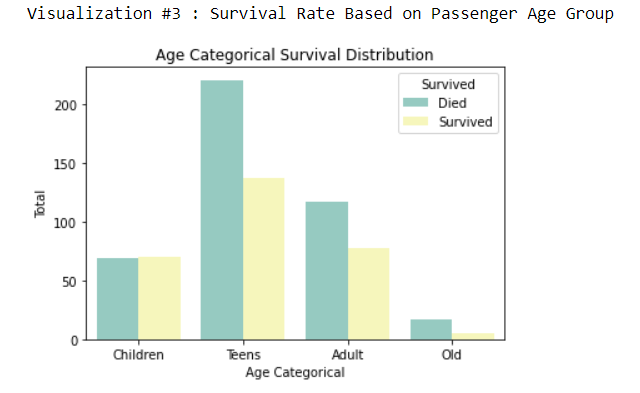
# Pandas cut() function is used to separate the array elements into different bins . The cut function is mainly used to perform statistical analysis on scalar data.

df['Age\_cats'] = pd.cut(df.Age, interval, labels = categories)

ax = sns.countplot(x = 'Age\_cats', data = df, hue = 'Survived', palette = 'Set3')

ax.set(xlabel='Age Categorical', ylabel='Total',title="Age Categorical Survival Distribution")

plt.show()



print("Visualization #4 : Survival Rate Based on Passenger Embarked Port")

print(pd.crosstab(df['Embarked'], df.Survived))

ax = sns.countplot(x = 'Embarked', hue = 'Survived', palette = 'Set1', data = df)

ax.set(title = 'Survival distribution according to Embarking place')

plt.show()

