PYTHONE & DATA ANALYTICS BOOTCAMP

PROJECT

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sb

 To read data frame

Data1=pd.DataFrame(pd.read\_csv('/content/train (1).csv'))

To see data set

Data1

# To get fist 5 data information

Data1.head()

# To get last 5 data information

Data1.tail()

# To get number of columns and rows

Data1.shape

# To get column names

Data1.columns

# To get rows information

Data1.index

# To get all null values

x=Data1.isnull().sum()

x

drop\_col=x[x>(35/100\*Data1.shape[0])]

drop\_col

Data1.drop(drop\_col.index, axis=1, inplace=True)

Data1.fillna(Data1.mean(),inplace=True)

Data1.isnull().sum()

# To get information about Embaked columns

Data1['Embarked'].describe()

Data1['Embarked'].fillna('S',inplace=True)

Data1.isnull().sum()

# To get corrlation between other columns

Data1.corr()

Data1['FamilySize']=Data1['SibSp']+Data1['Parch']

Data1.drop(['SibSp','Parch'],axis=1,inplace=True)

Data1.corr()

Data1['Alone']=[0 if Data1['FamilySize'][i]>0 else 1 for i in Data1.index]

Data1.head()

Data1.tail()

Data1.groupby(['Alone'])['Survived'].mean()

Data1.groupby(['Sex'])['Survived'].mean()

Data1.groupby(['Embarked'])['Survived'].mean()

Data1.groupby(['Pclass'])['Survived'].mean()

# CONCLUSION

# There is 891 rows and 12 columns

# When we onsider the corrilation between other colums then nwe can see there is a good corrilation between Pclass and Fare value, that mean if Pclass is increase then Fare value is decrease.

# We can, if person is alone then they have less chance of surviving.

# And also Femeals passengers have more chance of surviving than males.

# We see that the precentage of survive was more for Embarked is "C" and survive was minimum for Embarked is "S".

# finaly we can say that the Pclass 1 has hige chance to survived and less chance to survive Pclass is 3