**ASSIGNMENT -2**

**BSAN 710: PYTHON FOR BUSINESS**

**Thursday 7.00 PM**

**Dr. Justin Keeler**

**April 27, 2023**

**GROUP 2:**

**NARAYANAM NAGA VAMSI SANDEEP**

**DECHARAJU PHANINDRA**

**SAI KOUSHIK REDDY**

**AARTHI AWASTHI**

**RAGHU VAMSI**

Firstly, we have downloaded the file of Titanic dataset and changed the path of Jupyter notebook

os.chdir("C:\\Users\\user\\Downloads\\Python Assignment 2")

copied the path to the jupyter notebook.

**1)Load the dataset into Pandas Data Frame.**

Using the define function df = pd.read\_csv("Titanic-Dataset.csv")

**2) Explore the dataset using Pandas to answer the following questions:**

1. **How many passengers survived?**

print ("Total number of passengers survived:", (df.Survived == 1).sum())

Here, we can know how many passengers survived by using survived column and counted the number of people survived using sum function.

So, by using print function we will get to know how many passengers survived i.e, 342

1. **What was the average age of the passengers?**

df.describe()

By using this function, we get a table describing mean, standard deviation, count. etc.

print ("Average age of passengers:", round(df.describe().loc['mean', 'Age'],2), "years")

with mean and age by using loc function we can get the average age of the passenger = 29.7 years

1. **What was the survival rate of male and female passengers?**

Here, we calculated Number of Male and Female people survived and dead and calculated the percentage of Male and Female people survived

Male Survival rate = (Male survived/ (Male survived + Male died)) \*100

Female Survival rate = (Female survived/ (Female survived + Female died)) \*100

1. **Which passenger class had the highest survival rate?**

Here we calculated the Total number of people survived and died of each class i.e. (P1, P2, P3) and calculated the percentage of people survived on each class

P1 class survived rate = (p1 survived/ total p1 passengers) \*100

P2 class survived rate = (p2 survived/ total p2 passengers) \*100

P3 class survived rate = (p3 survived/ total p3 passengers) \*100

We calculated the Maximum percentage rate using max function and founded P1 class had highest survival rate

1. **How many passengers had siblings/spouses aboard?**

(df['SibSp'] > 0).value\_counts().loc[True]

Using the data of sibling’s/spouse column we counted the number of passengers having siblings if the column value is greater than 0 and we got the value as 283

1. **How many passengers had parents/children aboard?**

(df['Parch'] > 0).value\_counts().loc[True]

Using the data present in parent’s/children column we counted the number of passengers having parents and children if the column value is greater than 0 and we got the value as 213

**3)Create a dictionary to store the data for each passenger class (1st, 2nd, and 3rd) and their respective survival rates.**

Here we created a dictionary named survival\_rates\_pclass and considered keys as P1, P2, P3 survival rates and values are considered as percentage of P1, P2, P3 survival rates and printed the values present in the dictionary.

**4) Use Matplotlib to create a bar chart showing the survival rates of each passenger class.**

Here we calculated the Total number of people survived and died of each class i.e. (P1, P2, P3) and calculated the percentage of people survived on each class

P1 class survived rate = (p1 survived/ total p1 passengers) \*100

P2 class survived rate = (p2 survived/ total p2 passengers) \*100

P3 class survived rate = (p3 survived/ total p3 passengers) \*100

Survival rates are stored in a list and named as an and class is taken in other list b

X-axis is labeled as Pclass and Y-axis is labeled as survival rate and title is given as Passenger Vs Survival rate.

Bar chart is plotted between an and b by using plt.bar (b, a)

**5)Create a scatter plot showing the relationship between passenger age and fare paid. Color code the plot based on survival status (survived or did not survive)**

Here we plotted scatter plot between age of the passenger and fair paid based on their survival rate. we took the age and fare column from the data set and mapped with survival data column and assigned color as cool warm and created scatter plot function Using scatter plot function (plt. scatter)

X-axis is labeled as age and y-axis is labeled as fare and title is given as Age Vs Fare by Survival

**6) Create a histogram of the passenger age distribution.**

we created a Histogram based on the age of the passengers. We took bin size as 8 and X-axis is labeled as Age and Y-axis as frequency of number of passengers and titled as Distribution of Passenger Age. We created Histogram function (plt. hits)

**7) Create a pie chart showing the proportion of passengers who survived and did not survive.**

**We counted the number of people survived based on the survived column data. It is titled as** Proportion of Passengers Not survived', 'Proportion of passengers who survived and did not survive'. We created a pie chart using(kind=pie) function and labeled as 'Proportion of Passengers Not survived', 'Proportion of Passengers survived’. we found 61.62% of people not survived and 38.38% of people survived.

**8) Create a box plot showing the distribution of passenger fares by passenger class.**

class1\_fares = df[df['Pclass'] == 1] ['Fare']. This line is used to filter the fare price of class P. similarly we filtered for Class 2&3 and stored in a list named fares. X-axis is labeled as Passenger class and sub plots is taken as Pclass=1, Pclass=2, Pclass=3 and y-axis as fare. And box plot is drawn between fare and passenger class using [fig, ax = plt.subplots() ax.boxplot()] function

  