AI Lab 1
22k-4458
Muhammad Taha
BCS-6H

Q1)

[1]:
```
import pandas as pd

df = pd.read_csv('Titanic-Dataset.csv')

print(df.head())
```

```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3

                                                Name     Sex   Age  SibSp  \
0                            Braund, Mr. Owen Harris    male  22.0      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
2                             Heikkinen, Miss. Laina  female  26.0      0
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                           Allen, Mr. William Henry    male  35.0      0

   Parch            Ticket     Fare Cabin Embarked
0      0         A/5 21171   7.2500   NaN        S
1      0          PC 17599  71.2833   C85        C
2      0  STON/O2. 3101282   7.9250   NaN        S
3      0            113803  53.1000  C123        S
4      0            373450   8.0500   NaN        S
```

1. **Total Passengers:**

```
[2]: passengers = df['PassengerId'].count()
     print(passengers)
```

```
891
```

2. **Survival Rate:**

```
survival_rate = df['Survived'].mean() * 100
print(survival_rate)
```

```
38.38383838383838
```

3. **Average Age:**

```
avg_age = df['Age'].mean()
print(avg_age)
```

```
29.69911764705882
```

4. **Total No of Male and Female Passengers:**

```
male = df[df['Sex'] == "male"]
female = df[df['Sex'] == "female"]

print ("Male Passengers = ",male['Sex'].count())
print ("Female Passengers = ", female['Sex'].count())
```

```
Male Passengers =  577
Female Passengers =  314
```

5. **Survival Rate by Gender:**

```
male_survival_rate = df[df['Sex'] == "male"]
female_survival_rate = df[df['Sex'] == "female"]

print(male_survival_rate['Survived'].mean() * 100)
print(female_survival_rate['Survived'].mean() * 100)
```

```
18.890814558058924
74.20382165605095
```

6. **Passengers in Each Class:**

```
: classes = df.groupby('Pclass')['PassengerId'].count()
  print(classes)
```

```
Pclass
1    216
2    184
3    491
```

7. **Survival Rate by Class:**

```
class_survival_rate = df.groupby('Pclass')['Survived'].mean() * 100
print(class_survival_rate)
```

```
Pclass
1    62.962963
2    47.282609
3    24.236253
```

8. **Average Fare for Each Class:**

```
avg_fare = df.groupby('Pclass')['Fare'].mean()
print(avg_fare)
```

```
Pclass
1    84.154687
2    20.662183
3    13.675550
```

9. **Passengers Travelling with Family Members:**

```
Family_Members = df['SibSp'] + df['Parch']
passengers_wth_family = df[df["Family_Members"] > 0]
print(passengers_wth_family['PassengerId'].count())
```

```
354
```

10. **Handling Missing Values:**

```
#for age column we can use mean to fill up the missing spaces
avg_age = df['Age'].mean()
df['Age'].fillna(avg_age, inplace=True)

#For cabin we can replace null values with a placeholder
df['Cabin'].fillna('Unknown', inplace=True)
```