Name : Komal Katare

Roll No. 772

PRN : 202201040142

link to dataset: https://www.kaggle.com/c/titanic/data?select=train.csv

Question 1: What is the survival rate of passengers based on their age group (child, adult, elderly)?

```
import pandas as pd
import numpy as np
df = pd.read_csv('/content/drive/MyDrive/train.csv')
age_bins = [0, 18, 60, np.inf]
age_labels = ['Child', 'Adult', 'Elderly']

df['AgeGroup'] = pd.cut(df['Age'], bins=age_bins, labels=age_labels)

survival_rate_per_age_group = df.groupby('AgeGroup')['Survived'].mean() * 100

print("Survival rate of passengers based on age group:")
print(survival_rate_per_age_group)
```

```
Survival rate of passengers based on age group:
AgeGroup
Child      50.359712
Adult      38.878843
Elderly    22.727273
Name: Survived, dtype: float64
```

Question 2: How many passengers had siblings or spouses on board, and how many of them survived?

```
passengers_with_sibsp = df[df['SibSp'] > 0].shape[0]
survivors_with_sibsp = df[(df['SibSp'] > 0) & (df['Survived'] == 1)].shape[0]

print("Number of passengers with siblings or spouses:", passengers_with_sibsp)
print("Number of survivors with siblings or spouses:", survivors_with_sibsp)
```

```
Number of passengers with siblings or spouses: 283
Number of survivors with siblings or spouses: 132
```

Question 3: What is the average fare paid by passengers in each age group?

```
age_bins = [0, 18, 60, np.inf]
age_labels = ['Child', 'Adult', 'Elderly']

df['AgeGroup'] = pd.cut(df['Age'], bins=age_bins, labels=age_labels)

average_fare_per_age_group = df.groupby('AgeGroup')['Fare'].mean()

print("Average fare paid by passengers in each age group:")
print(average_fare_per_age_group)
```

```
Average fare paid by passengers in each age group:
AgeGroup
Child      32.500721
Adult      34.980318
Elderly    41.371214
Name: Fare, dtype: float64
```

Question 4: How many passengers traveled in each cabin class (A, B, C, D, E, F, G) and what percentage of total passengers does each class represent?

```
passengers_per_cabin_class = df['Cabin'].str[0].value_counts()
total_passengers = df.shape[0]

percentage_per_cabin_class = passengers_per_cabin_class / total_passengers * 100

print("Passengers in each cabin class and their percentage:")
print(passengers_per_cabin_class)
print(percentage_per_cabin_class)
```

```
Passengers in each cabin class and their percentage:
C    59
B    47
```

```
D    33
E    32
A    15
F    13
G     4
T     1
Name: Cabin, dtype: int64
C    6.621773
B    5.274972
D    3.703704
E    3.591470
A    1.683502
F    1.459035
G    0.448934
T    0.112233
Name: Cabin, dtype: float64
```

Question 5: What is the survival rate of passengers who traveled alone (without any siblings, spouses, parents, or children)?

```
alone_passengers = df[(df['SibSp'] == 0) & (df['Parch'] == 0)]
survival_rate_alone_passengers = alone_passengers['Survived'].mean() * 100

print("Survival rate of passengers who traveled alone: {:.2f}%".format(survival_rate_alone_passengers))
```

```
Survival rate of passengers who traveled alone: 30.35%
```

Question 6: How many passengers had a known cabin number assigned?

```
passengers_with_cabin = df['Cabin'].notnull().sum()

print("Number of passengers with a known cabin number assigned:", passengers_with_cabin)
```

```
Number of passengers with a known cabin number assigned: 204
```

Question 7: What is the average fare paid by passengers of each gender?

```
average_fare_per_gender = df.groupby('Sex')['Fare'].mean()

print("Average fare paid by passengers of each gender:")
print(average_fare_per_gender)
```

```
Average fare paid by passengers of each gender:
Sex
female    44.479818
male      25.523893
Name: Fare, dtype: float64
```

Question 8: What is the survival rate of passengers based on their ticket fare category (low, medium, high)?

```
fare_bins = [0, 50, 100, np.inf]
fare_labels = ['Low', 'Medium', 'High']

df['FareCategory'] = pd.cut(df['Fare'], bins=fare_bins, labels=fare_labels)

survival_rate_per_fare_category = df.groupby('FareCategory')['Survived'].mean() * 100

print("Survival rate of passengers based on fare category:")
print(survival_rate_per_fare_category)
```

```
Survival rate of passengers based on fare category:
FareCategory
Low       32.402235
Medium    65.420561
High      73.584906
Name: Survived, dtype: float64
```

Question 10: What is the percentage of passengers who survived based on their cabin class?

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

print("Percentage of passengers who survived based on cabin class:")
print(survival_percentage_per_class)
```

```
Percentage of passengers who survived based on cabin class:
Pclass
1    62.962963
2    47.282609
```

```
3    24.236253
Name: Survived, dtype: float64
```