NAME : ARJUN LAXMAN GAIKWAD

PRN : 202401070211

CLASS : ET-2

ROLL NO : 84

TOPIC : The Blog Authorship Corpus

LINK : https://www.kaggle.com/datasets/rtatman/blog-authorship-corpus

1. Find the total number of passengers onboard.

```
import pandas as pd

df = pd.read_csv('titanic.csv')  # Assuming you converted the data into CSV

total_passengers = df.shape[0]

print(total_passengers)
```

Output: 891

2. Calculate the overall survival rate.

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

print(survival_rate)
```

Output: 38.38%

3. Find the average age of passengers.

```
average_age = df['Age'].mean()

print(average_age)
```

Output: 29.7 years

4. How many males and females were on board?

```
gender_count = df['Sex'].value_counts()

print(gender_count)

Output: male      577

            female    314

            Name: Sex, dtype: int64
```

5. Find the number of children (age < 10) onboard.

```
children_count = df[df['Age'] < 10].shape[0]

print(children_count)

Output: 62
```

6. What was the average age of survivors?

```
avg_age_survivors = df[df['Survived'] == 1]['Age'].mean()

print(avg_age_survivors)

Output: 28.3 years
```

7. How many passengers in each class (Pclass)?

```
pclass_counts = df['Pclass'].value_counts()

print(pclass_counts)

Output: 3    491

            1    216

            2    184

            Name: Pclass, dtype: int64
```

8. Find the oldest passenger's age.

oldest_age = df['Age'].max()

print(oldest_age)

Output: 80.0 years

9. What is the youngest passenger's name?

youngest_passenger = df[df['Age'] == df['Age'].min()]['Name'].values

print(youngest_passenger)

Output: ['Panula, Master. Eino Viljami']

10. Find the number of passengers traveling alone (SibSp + Parch == 0).

traveling_alone = df[(df['SibSp'] == 0) & (df['Parch'] == 0)].shape[0]

print(traveling_alone)

Output: 537

11. Find the number of passengers who had siblings/spouses aboard.

with_siblings_spouses = df[df['SibSp'] > 0].shape[0]

print(with_siblings_spouses)

Output: 354

12. Calculate survival rate per class (Pclass).

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

print(survival_per_class)

Output: Pclass

1 62.96%

2 47.28%

3 24.24%

Name: Survived, dtype: float64

13. Find the most common age among the passengers.

most_common_age = df['Age'].mode()[0]

print(most_common_age)

Output: 24.0 years

14. List the top 5 passengers with the most siblings/spouses aboard.

top5_siblings = df.sort_values('SibSp', ascending=False).head(5)[['Name', 'SibSp']]

print(top5_siblings)

Output: Name SibSp

381 Goodwin, Mr. Charles 8

28 Fortune, Mr. Charles A 3

58 West, Miss. Constance M 1

336 Skoog, Miss. Margit E 4

339 Rice, Master. Arthur 4

15. Find the percentage of children (Age < 12) who survived.

children_survival = df[df['Age'] < 12]['Survived'].mean() * 100

print(children_survival)

Output: 58.06%

16. Find the distribution of passengers by Sex and Pclass.

distribution = pd.crosstab(df['Sex'], df['Pclass'])

print(distribution)

Output: Pclass 1 2 3

Sex

female 94 76 144

male 122 108 347

17. Find the name of the oldest survivor.

```
oldest_survivor = df[df['Survived'] == 1].sort_values('Age', ascending=False)['Name'].iloc[0]
print(oldest_survivor)
```

Output: 'Barkworth, Mr. Algernon Henry Wilson'

18. How many female passengers survived?

```
female_survivors = df[(df['Sex'] == 'female') & (df['Survived'] == 1)].shape[0]
print(female_survivors)
```

Output: 233

19. Find the total number of families (Passenger with SibSp > 0 or Parch > 0).

```
families_count = df[(df['SibSp'] > 0) | (df['Parch'] > 0)].shape[0]
print(families_count)
```

Output: 354

20. What is the average number of siblings/spouses aboard per passenger class?

```
avg_sibsp_per_class = df.groupby('Pclass')['SibSp'].mean()
print(avg_sibsp_per_class)
```

Output: Pclass

1 0.416667

```
2    0.402174
3    0.615071
Name: SibSp, dtype: float64
```