Assignment No. 1
Titanic Dataset

Code:

```
import pandas as pd
import numpy as np
from statistics import mean
import matplotlib.pyplot as plt
import seaborn as sns
data = pd.read_csv("titanic.csv")
```

1) Which people most likely survived?
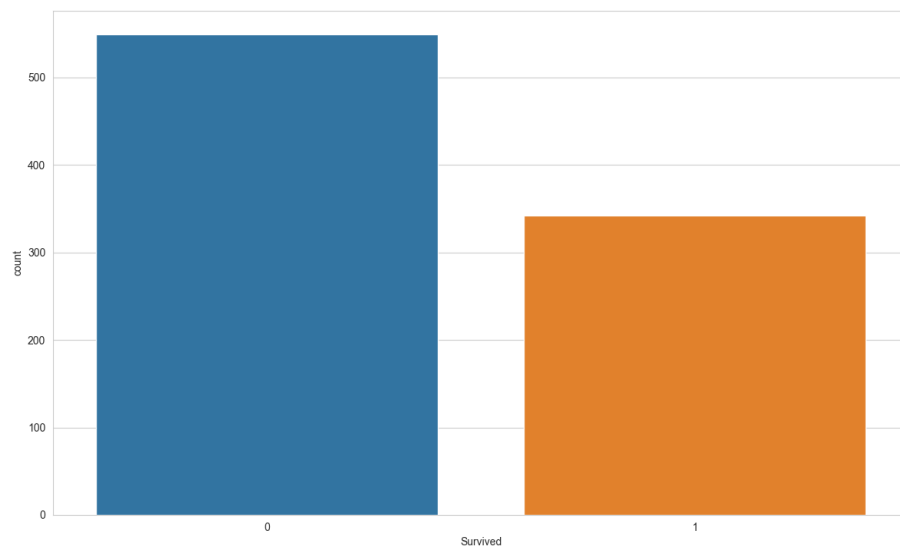   **Code:**
   ```
   print(data["Survived"].value_counts())
   sns.set_style("whitegrid")
   sns.countplot(x = "Survived", data = data)
   plt.show()
   ```

   Output:
   ```
   0    549
   1    342
   Name: Survived, dtype: int64
   ```


In total 342 passengers survived in Titanic crash. Including all females, males, young, middle, old age and rich and poor passengers.

2) Young/Old/Middle?
Code:
```
print((((data["Survived"] == 1) & (data["Age"] <= 17)).value_counts().tolist()[1])

print((((data["Survived"] == 1) & ((data["Age"] >= 18) & (data["Age"] <= 59))).value_counts().tolist()[1])

print((((data["Survived"] == 1) & (data["Age"] >= 60)).value_counts().tolist()[1])

sns.distplot(data["Age"].dropna(), kde = False, color = "darkgreen", bins = 40)
plt.show()
```
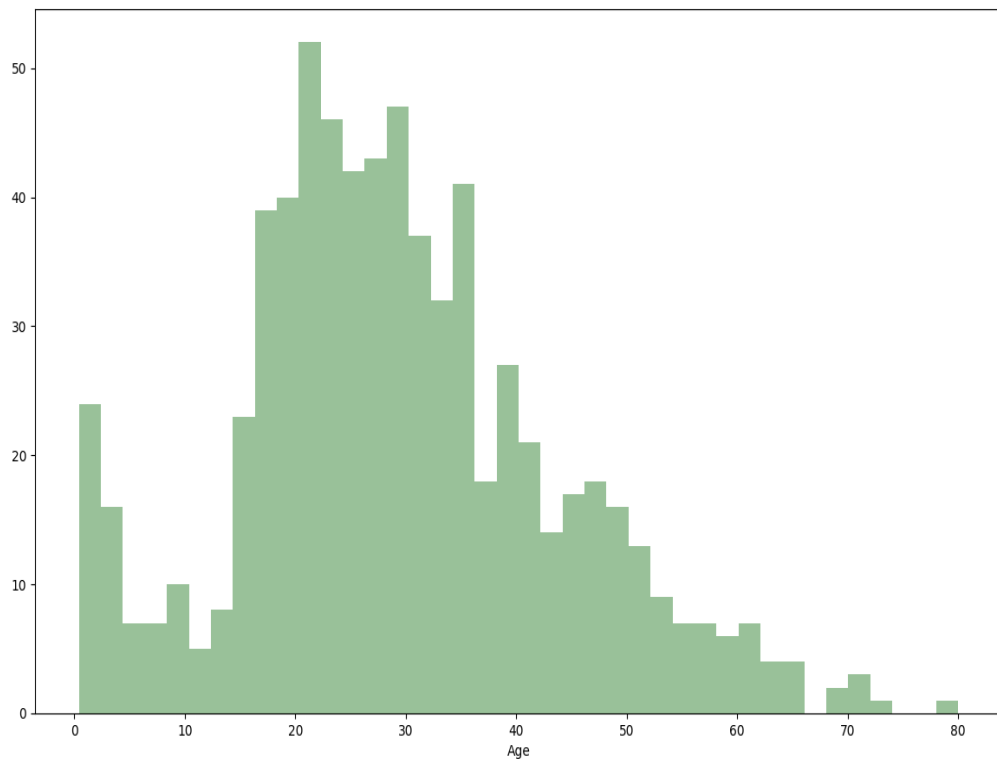
Output:
Young - 61
Middle - 222
Old - 7

In conclusion, the middle age passengers whose age was from 18 to 59 survived the most. In bar graph the maximum bar plot is between age 18 to 59 and through age condition also it's same.

3) Female/Male?

Code:
```
survival_gender = data[data["Survived"] == 1]["Sex"].value_counts()
survival_gender = survival_gender.tolist()
print(survival_gender)

total_passengers = data["Survived"].value_counts()
total_passengers = total_passengers.tolist()
total_passengers = total_passengers[0] + total_passengers[1]

female_survival_percentage = float((survival_gender[0] / total_passengers) * 100)
print(female_survival_percentage)

male_survival_percentage = float((survival_gender[1] / total_passengers) * 100)
print(male_survival_percentage)

sns.set_style("whitegrid")
sns.countplot(x = "Sex", data  = data[data["Survived"] == 1])
plt.show()
```
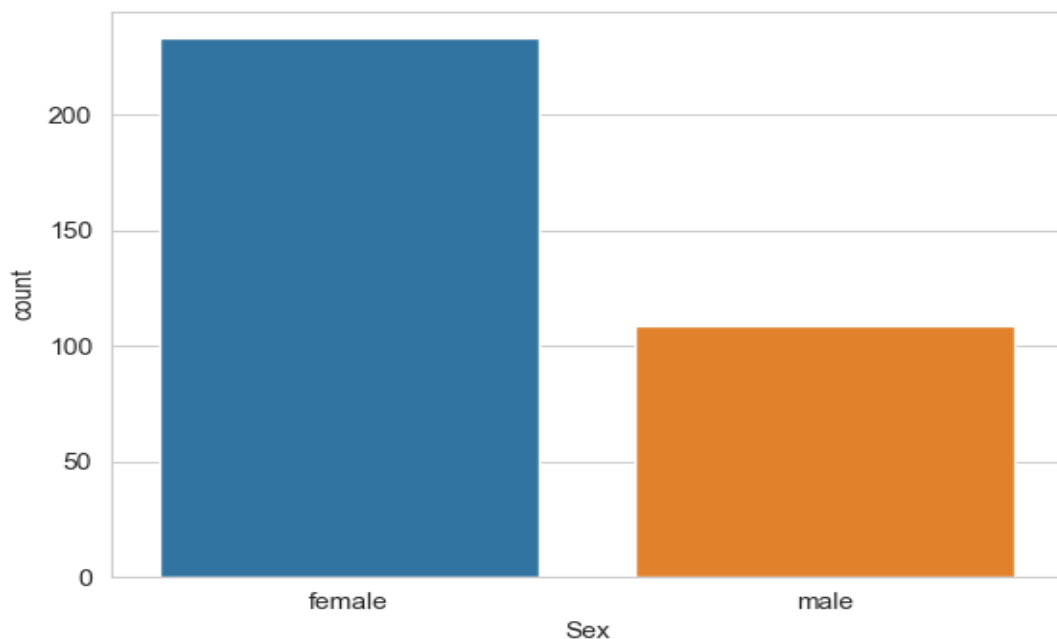
Output:
[233, 109]
Female Survivor Percentage - 26.15039281705948
Male Survivor Percentage - 12.2334455667789

Therefore, female survivors were more than male survivors. The priority is given to female to escape from ship. Maximum female commuter was safely escaped from crash.

4) Rich/Poor

Code:

```
total_passengers = data["Survived"].value_counts()
total_passengers = total_passengers.tolist()
total_passengers = total_passengers[0] + total_passengers[1]

avg = int(mean(data["Fare"]))
fare = (((data["Survived"] == 1) & (data["Fare"] >= avg)).value_counts()).tolist()
print(fare[1])
rich_passengers = fare[1]

fare1 = (((data["Survived"] == 1) & (data["Fare"] <= avg)).value_counts()).tolist()
print(fare1[1])
poor_passengers = fare1[1]

rich_passengers_survival_percentage = float((rich_passengers / total_passengers) * 100)
print(rich_passengers_survival_percentage)

poor_passengers_survival_percentage = float((poor_passengers / total_passengers) * 100)
print(poor_passengers_survival_percentage)

sns.set_style("whitegrid")
sns.countplot(x = "Survived", hue = "Pclass", data = data)
plt.show()
```
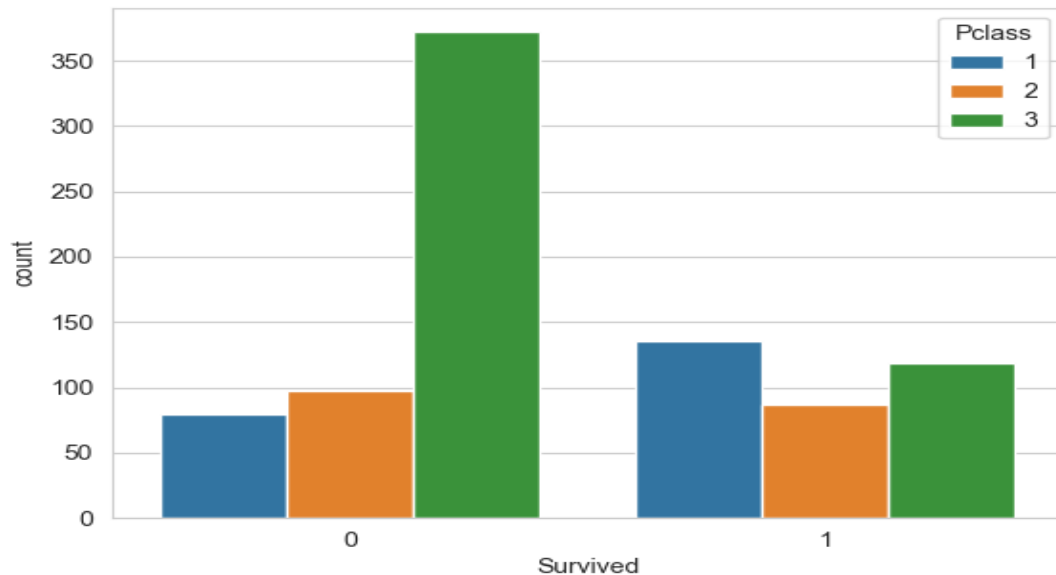
Output:

Rich Passengers Survival - 126
Poor Passengers Survival - 216
Rich Passengers Survival Percentage - 14.14141414141414
Poor Passengers Survival Percentage - 24.242424242424242

Hence, poor passengers survived more than rich passengers. The no. of poor passengers were more than rich passengers. The bar graph is of Pclass. The blue plot is of rich people who survived. The green is of poor people who survived and orange plot is of middle-class people who survived. As I have taken average from fare and calculated rich and poor the orange bar plot got divided into some rich passengers and some poor passengers. Resulting, in poor passengers survival was more than rich passengers.

References:
https://github.com/datasciencedojo/datasets/blob/master/titanic.csv