Importing Libraries

```
import pandas as pd
```

+ Code + Text

Upload dataset in the local directory

```
from google.colab import files
uploaded=files.upload()
```

Choose Files Titanic Data.csv

- **Titanic Data.csv**(text/csv) - 28679 bytes, last modified: 7/28/2023 - 100% done

```
Saving Titanic Data.csv to Titanic Data.csv
```

Load Dataset

```
#Reading the csv file
dataset=pd.read_csv("Titanic Data.csv")

#Print top 5 rows
print(dataset.head())
```

```
   PassengerId  Survived  Pclass  \
0          892         0       3
1          893         1       3
2          894         0       2
3          895         0       3
4          896         1       3

                                          Name     Sex   Age  SibSp  Parch  \
0                              Kelly, Mr. James    male  34.5      0      0
1              Wilkes, Mrs. James (Ellen Needs)  female  47.0      1      0
2                     Myles, Mr. Thomas Francis    male  62.0      0      0
3                              Wirz, Mr. Albert    male  27.0      0      0
4  Hirvonen, Mrs. Alexander (Helga E Lindqvist)  female  22.0      1      1

    Ticket     Fare Embarked
0   330911   7.8292        Q
1   363272   7.0000        S
2   240276   9.6875        Q
3   315154   8.6625        S
4  3101298  12.2875        S
```

Getting information about the dataset

```
#No.of rows and columns
print(dataset.shape)
```

```
(418, 11)
```

```
#To display columns and datatype
print(dataset.info)
```

```
<bound method DataFrame.info of      PassengerId  Survived  Pclass  \
0            892         0       3
1            893         1       3
2            894         0       2
3            895         0       3
4            896         1       3
..           ...       ...     ...
413         1305         0       3
414         1306         1       1
415         1307         0       3
416         1308         0       3
417         1309         0       3

                                          Name     Sex   Age  SibSp  Parch  \
```

```
0                              Kelly, Mr. James    male  34.5      0      0
1              Wilkes, Mrs. James (Ellen Needs)  female  47.0      1      0
2                     Myles, Mr. Thomas Francis    male  62.0      0      0
3                              Wirz, Mr. Albert    male  27.0      0      0
4    Hirvonen, Mrs. Alexander (Helga E Lindqvist)  female  22.0      1      1
..                                          ...     ...   ...    ...    ...
413                            Spector, Mr. Woolf    male   NaN      0      0
414                Oliva y Ocana, Dona. Fermina  female  39.0      0      0
415                Saether, Mr. Simon Sivertsen    male  38.5      0      0
416                        Ware, Mr. Frederick    male   NaN      0      0
417                  Peter, Master. Michael J    male   NaN      1      1

                  Ticket      Fare Embarked
0                 330911    7.8292        Q
1                 363272    7.0000        S
2                 240276    9.6875        Q
3                 315154    8.6625        S
4                3101298   12.2875        S
..                   ...       ...      ...
413           A.5. 3236    8.0500        S
414            PC 17758  108.9000        C
415  SOTON/O.Q. 3101262    7.2500        S
416              359309    8.0500        S
417                2668   22.3583        C

[418 rows x 11 columns]>
```

```
#To get the distribution of the data
print(dataset.describe)
```

```
<bound method NDFrame.describe of      PassengerId  Survived  Pclass  \
0            892         0       3
1            893         1       3
2            894         0       2
3            895         0       3
4            896         1       3
..           ...       ...     ...
413         1305         0       3
414         1306         1       1
415         1307         0       3
416         1308         0       3
417         1309         0       3

                                          Name     Sex   Age  SibSp  Parch  \
0                              Kelly, Mr. James    male  34.5      0      0
1              Wilkes, Mrs. James (Ellen Needs)  female  47.0      1      0
2                     Myles, Mr. Thomas Francis    male  62.0      0      0
3                              Wirz, Mr. Albert    male  27.0      0      0
4    Hirvonen, Mrs. Alexander (Helga E Lindqvist)  female  22.0      1      1
..                                          ...     ...   ...    ...    ...
413                            Spector, Mr. Woolf    male   NaN      0      0
414                Oliva y Ocana, Dona. Fermina  female  39.0      0      0
415                Saether, Mr. Simon Sivertsen    male  38.5      0      0
416                        Ware, Mr. Frederick    male   NaN      0      0
417                  Peter, Master. Michael J    male   NaN      1      1

                  Ticket      Fare Embarked
0                 330911    7.8292        Q
1                 363272    7.0000        S
2                 240276    9.6875        Q
3                 315154    8.6625        S
4                3101298   12.2875        S
..                   ...       ...      ...
413           A.5. 3236    8.0500        S
414            PC 17758  108.9000        C
415  SOTON/O.Q. 3101262    7.2500        S
416              359309    8.0500        S
417                2668   22.3583        C

[418 rows x 11 columns]>
```

Checking NaN Values

```
dataset.isna().any()
```

```
PassengerId    False
Survived       False
Pclass         False
Name           False
```

```
Sex           False
Age            True
SibSp         False
Parch         False
Ticket        False
Fare           True
Embarked      False
dtype: bool
```

```
dataset.isna().sum()
```

```
PassengerId     0
Survived        0
Pclass          0
Name            0
Sex             0
Age            86
SibSp           0
Parch           0
Ticket          0
Fare            1
Embarked        0
dtype: int64
```

Checking Duplicates

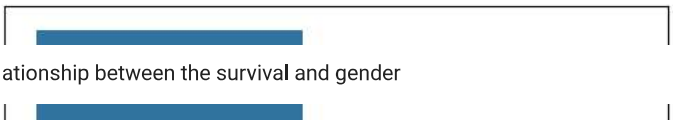
```
dataset.duplicated()
```

```
0      False
1      False
2      False
3      False
4      False
       ...
413    False
414    False
415    False
416    False
417    False
Length: 418, dtype: bool
```

Data Visualization

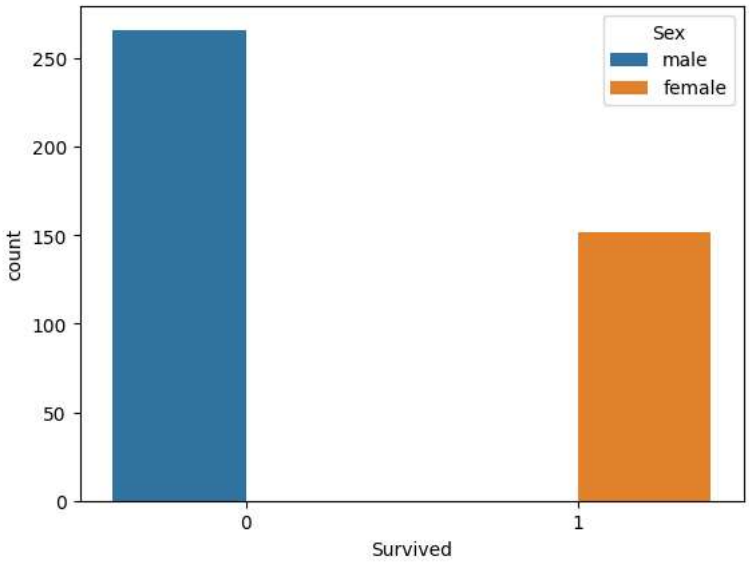
Visualizing the target column

```
#Importing Libraries
import seaborn as sns
import matplotlib.pyplot as plt
sns.countplot(x="Survived",data=dataset)
plt.show()
```



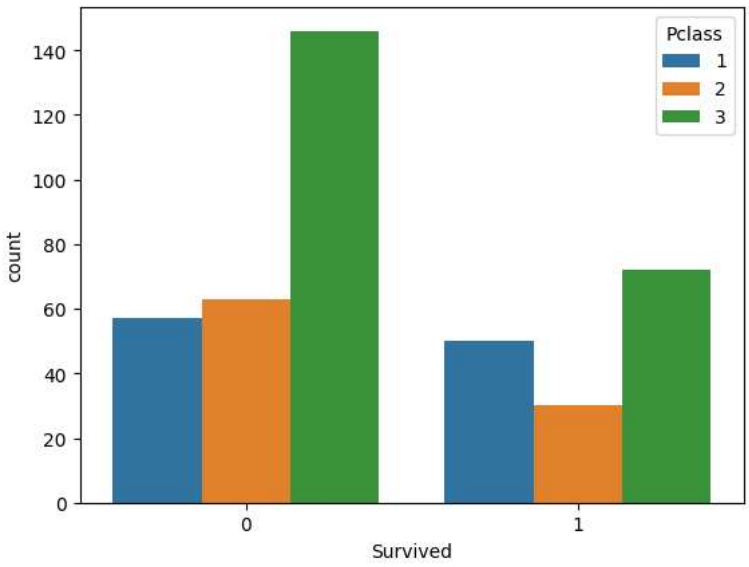
Explore the relationship between the survival and gender

```
sns.countplot(x="Survived",hue="Sex",data=dataset)
plt.show()
```



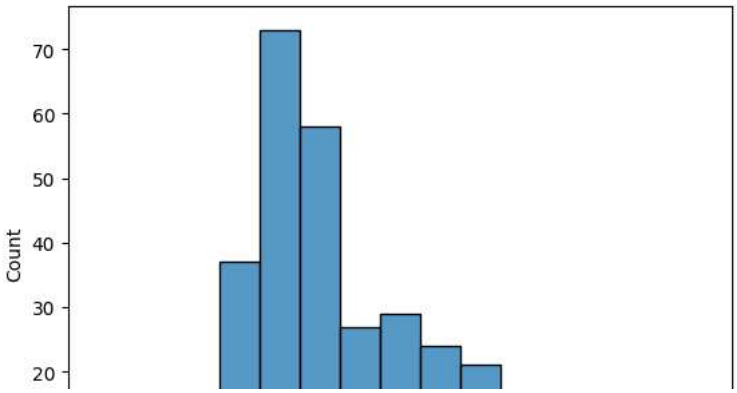
Explore relationship between the survival and passengerclass

```
sns.countplot(x="Survived",hue="Pclass",data=dataset)
plt.show()
```



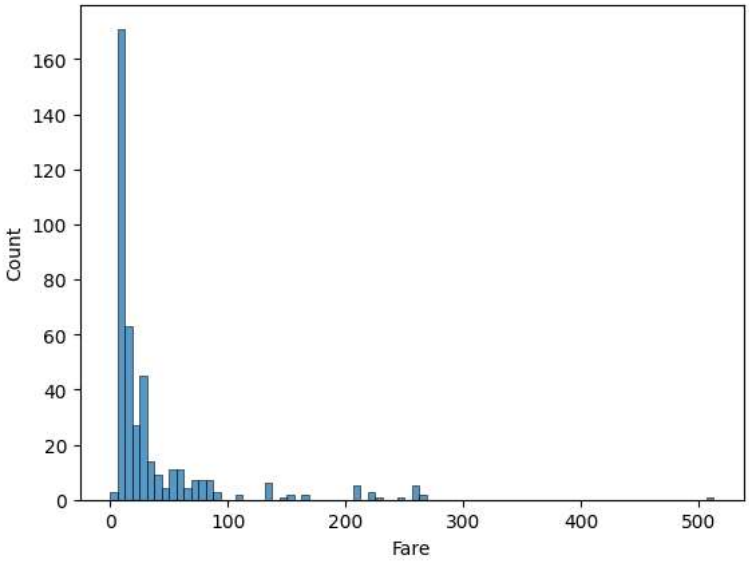
Explore the distribution of age

```
sns.histplot(x="Age",data=dataset)
plt.show()
```



Explore the distribution of fare

```
sns.histplot(x="Fare",data=dataset)
plt.show()
```



Save Dataset

```
dataset.to_csv("Titanicnew.csv",index=False)
```