IMPORT LIBRARIES

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

IMPORT RAW DATA

```
df=pd.read_csv("C:/Users/lekhs/OneDrive/Documents/titanic
ds/train.csv")

df.head(10)
```

```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3
5            6         0       3
6            7         0       1
7            8         0       3
8            9         1       3
9           10         1       2

                                                 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
5                                    Moran, Mr. James    male   NaN
0
6                             McCarthy, Mr. Timothy J    male  54.0
0
7                      Palsson, Master. Gosta Leonard    male   2.0
3
8  Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)  female  27.0
0
9                 Nasser, Mrs. Nicholas (Adele Achem)  female  14.0
1

   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
5      0            330877   8.4583   NaN        Q
6      0             17463  51.8625   E46        S
7      1            349909  21.0750   NaN        S
8      2            347742  11.1333   NaN        S
9      0            237736  30.0708   NaN        C
```

MISSING VALUES

```
df.isnull()

     PassengerId  Survived  Pclass   Name    Sex    Age  SibSp  Parch
Ticket  \
0          False     False   False  False  False  False  False  False
False
1          False     False   False  False  False  False  False  False
False
2          False     False   False  False  False  False  False  False
False
3          False     False   False  False  False  False  False  False
False
4          False     False   False  False  False  False  False  False
False
..           ...       ...     ...    ...    ...    ...    ...    ...
...
886        False     False   False  False  False  False  False  False
False
887        False     False   False  False  False  False  False  False
False
888        False     False   False  False  False   True  False  False
False
889        False     False   False  False  False  False  False  False
False
890        False     False   False  False  False  False  False  False
False

      Fare  Cabin  Embarked
0    False   True     False
1    False  False     False
2    False   True     False
3    False  False     False
4    False   True     False
..     ...    ...       ...
886  False   True     False
887  False  False     False
888  False   True     False
```

```
889  False  False     False
890  False   True     False

[891 rows x 12 columns]
```

REMOVE DUPLICATES

```
df.drop_duplicates()
```

```
     PassengerId  Survived  Pclass  \
0              1         0       3
1              2         1       1
2              3         1       3
3              4         1       1
4              5         0       3
..           ...       ...     ...
886          887         0       2
887          888         1       1
888          889         0       3
889          890         1       1
890          891         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
..                                                 ...     ...   ...
...
886                            Montvila, Rev. Juozas    male  27.0
0
887                     Graham, Miss. Margaret Edith  female  19.0
0
888           Johnston, Miss. Catherine Helen "Carrie"  female   NaN
1
889                            Behr, Mr. Karl Howell    male  26.0
0
890                              Dooley, Mr. Patrick    male  32.0
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
..     ...               ...      ...   ...      ...
886      0            211536  13.0000   NaN        S
887      0            112053  30.0000   B42        S
888      2        W./C. 6607  23.4500   NaN        S
889      0            111369  30.0000  C148        C
890      0            370376   7.7500   NaN        Q

[891 rows x 12 columns]
```

```
print(df.columns)
```

```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age',
'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

DESCRIBE VALUES

```
df.describe()
```

```
       PassengerId    Survived      Pclass         Age       SibSp  \
count   891.000000  891.000000  891.000000  714.000000  891.000000
mean    446.000000    0.383838    2.308642   29.699118    0.523008
std     257.353842    0.486592    0.836071   14.526497    1.102743
min       1.000000    0.000000    1.000000    0.420000    0.000000
25%     223.500000    0.000000    2.000000   20.125000    0.000000
50%     446.000000    0.000000    3.000000   28.000000    0.000000
75%     668.500000    1.000000    3.000000   38.000000    1.000000
max     891.000000    1.000000    3.000000   80.000000    8.000000

             Parch        Fare
count  891.000000  891.000000
mean     0.381594   32.204208
std      0.806057   49.693429
min      0.000000    0.000000
25%      0.000000    7.910400
50%      0.000000   14.454200
75%      0.000000   31.000000
max      6.000000  512.329200
```

INFO

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype
```

```
---  ------       --------------  -----
 0   PassengerId  891 non-null    int64
 1   Survived     891 non-null    int64
 2   Pclass       891 non-null    int64
 3   Name         891 non-null    object
 4   Sex          891 non-null    object
 5   Age          714 non-null    float64
 6   SibSp        891 non-null    int64
 7   Parch        891 non-null    int64
 8   Ticket       891 non-null    object
 9   Fare         891 non-null    float64
 10  Cabin        204 non-null    object
 11  Embarked     889 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

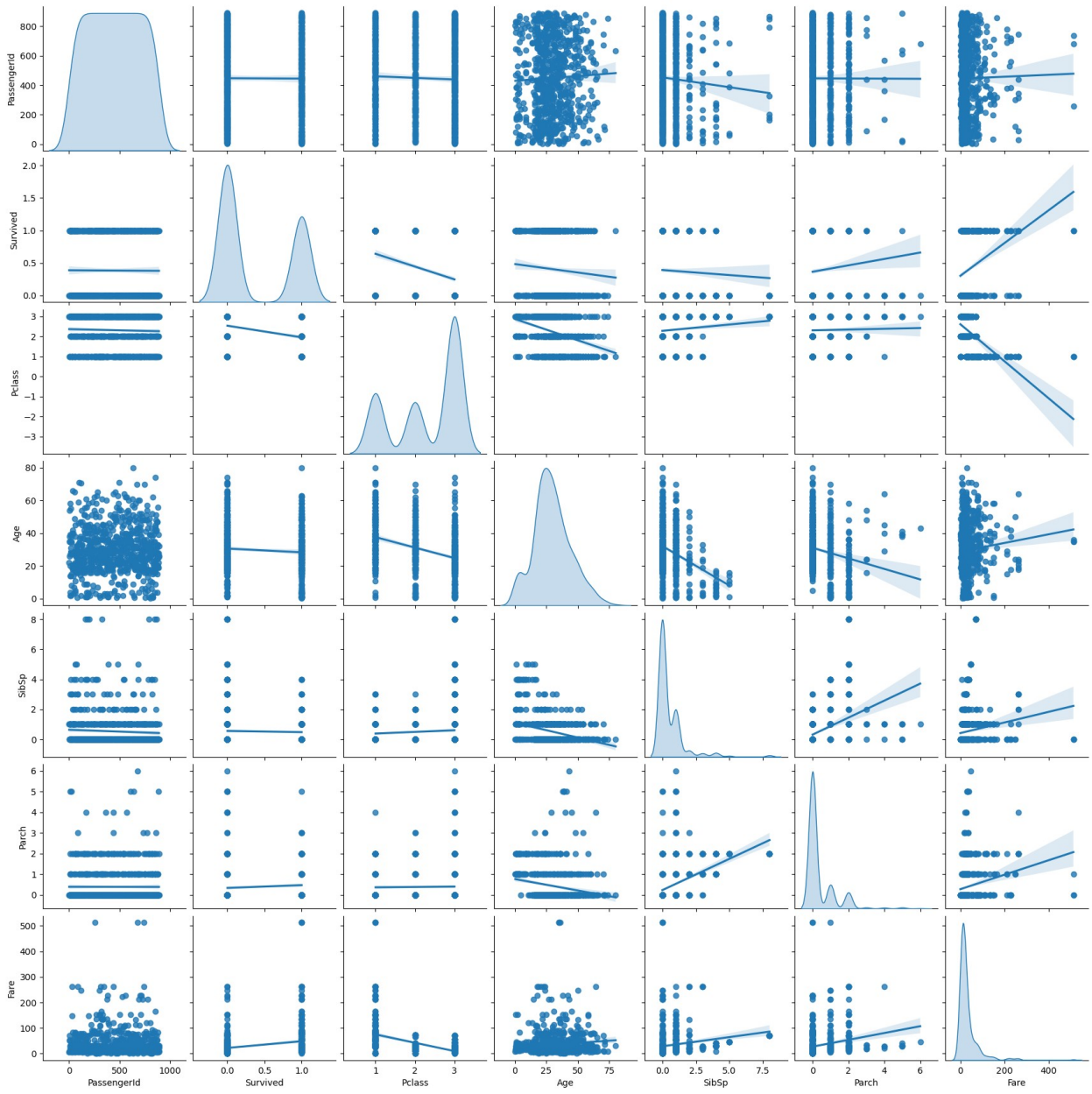
```
df.value_counts()
```

```
PassengerId  Survived  Pclass  Name
Sex     Age   SibSp  Parch  Ticket    Fare     Cabin        Embarked
2            1         1       Cumings, Mrs. John Bradley (Florence
Briggs Thayer)  female  38.0  1      0      PC 17599  71.2833  C85
C           1
4            1         1       Futrelle, Mrs. Jacques Heath (Lily May
Peel)          female  35.0  1      0      113803    53.1000  C123
S           1
7            0         1       McCarthy, Mr. Timothy J
male    54.0  0      0      17463     51.8625  E46          S
1
11           1         3       Sandstrom, Miss. Marguerite Rut
female  4.0   1      1      PP 9549   16.7000  G6           S
1
12           1         1       Bonnell, Miss. Elizabeth
female  58.0  0      0      113783    26.5500  C103         S
1

..
872          1         1       Beckwith, Mrs. Richard Leonard (Sallie
Monypeny)     female  47.0  1      1      11751     52.5542  D35
S           1
873          0         1       Carlsson, Mr. Frans Olof
male    33.0  0      0      695       5.0000   B51 B53 B55  S
1
880          1         1       Potter, Mrs. Thomas Jr (Lily Alexenia
Wilson)        female  56.0  0      1      11767     83.1583  C50
C           1
888          1         1       Graham, Miss. Margaret Edith
female  19.0  0      0      112053    30.0000  B42          S
1
890          1         1       Behr, Mr. Karl Howell
```

```
male    26.0  0      0      111369    30.0000  C148          C
1
Name: count, Length: 183, dtype: int64
```

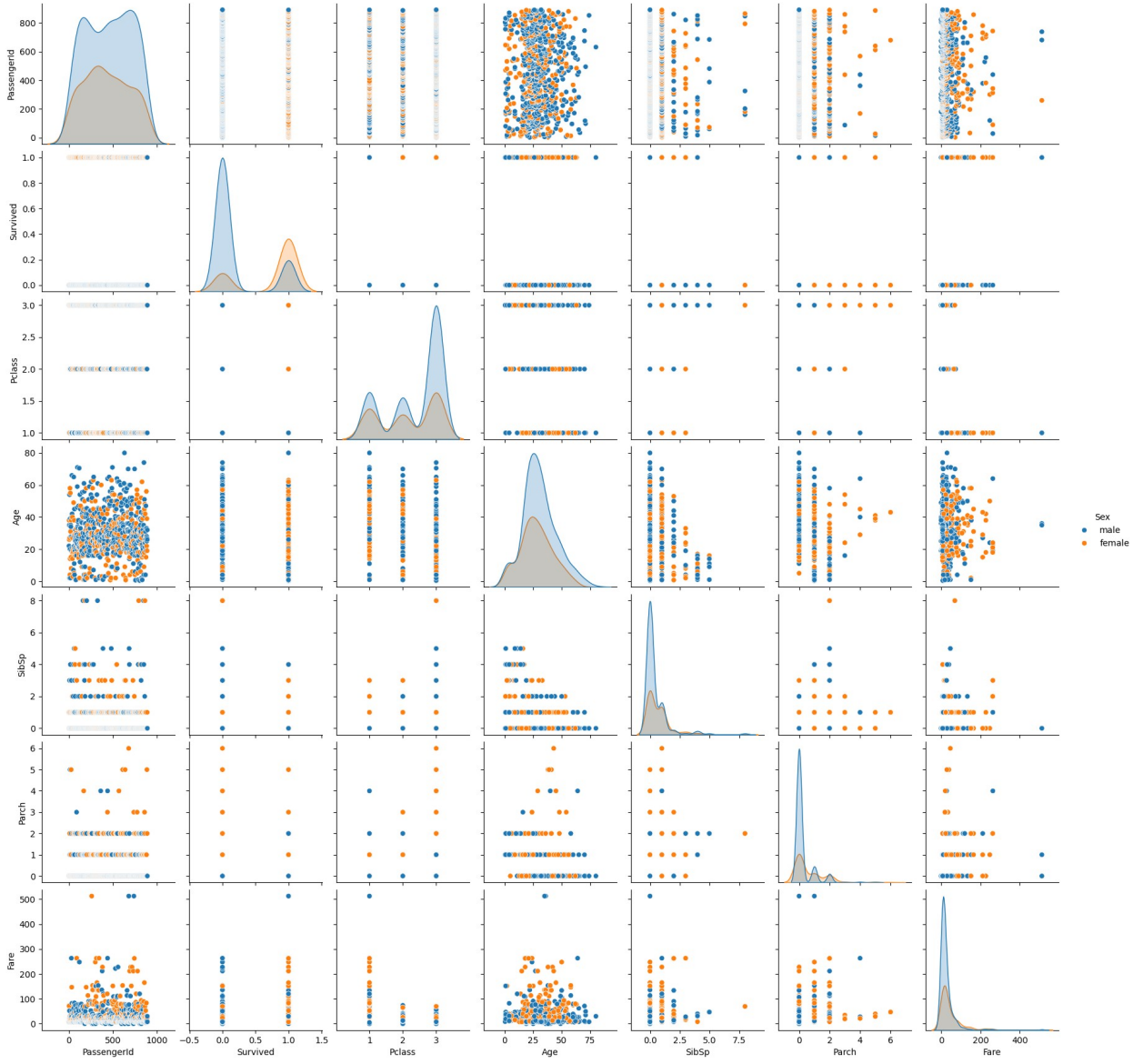
PAIRPLOT THROUGH SEABORN

```
sns.pairplot(df, diag_kind='kde',
             kind='reg')

<seaborn.axisgrid.PairGrid at 0x1c9463a9d00>
```



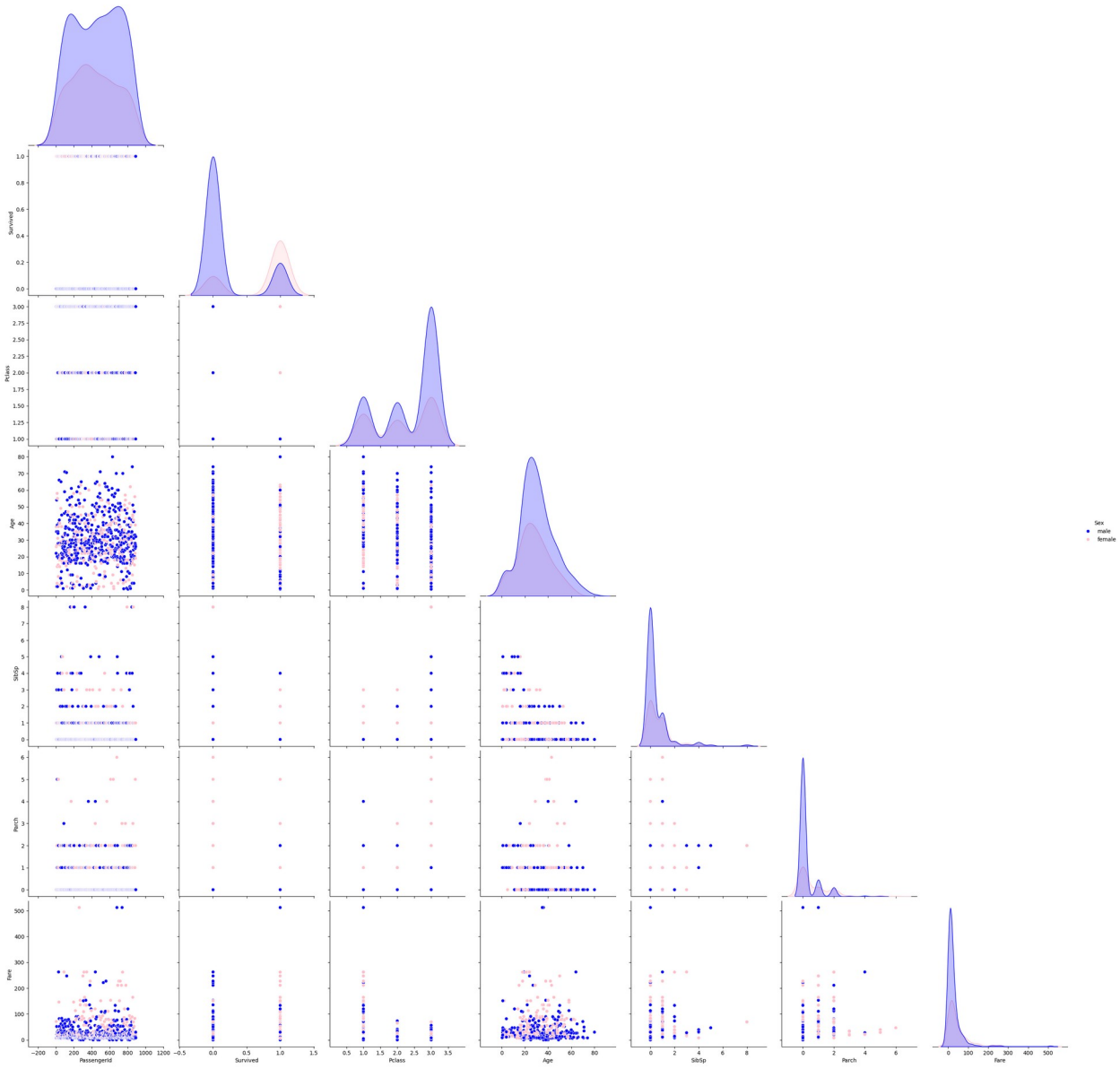

```
sns.pairplot(df, hue='Sex')
```

```
<seaborn.axisgrid.PairGrid at 0x1c9471e7860>
```




```
sns.pairplot(df, hue='Sex',
             kind="scatter",
             diag_kind="kde",
             palette={"male":"blue","female":"pink"},
             height=4,
             corner=True)
```

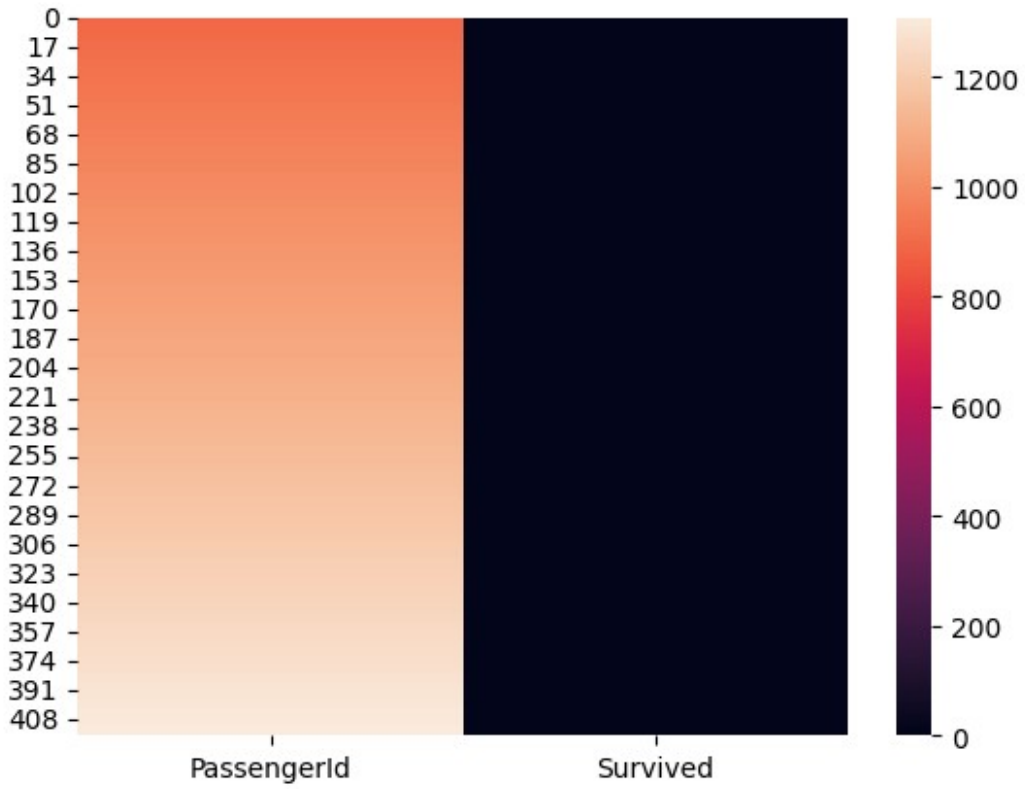
```
<seaborn.axisgrid.PairGrid at 0x1c94e972630>
```



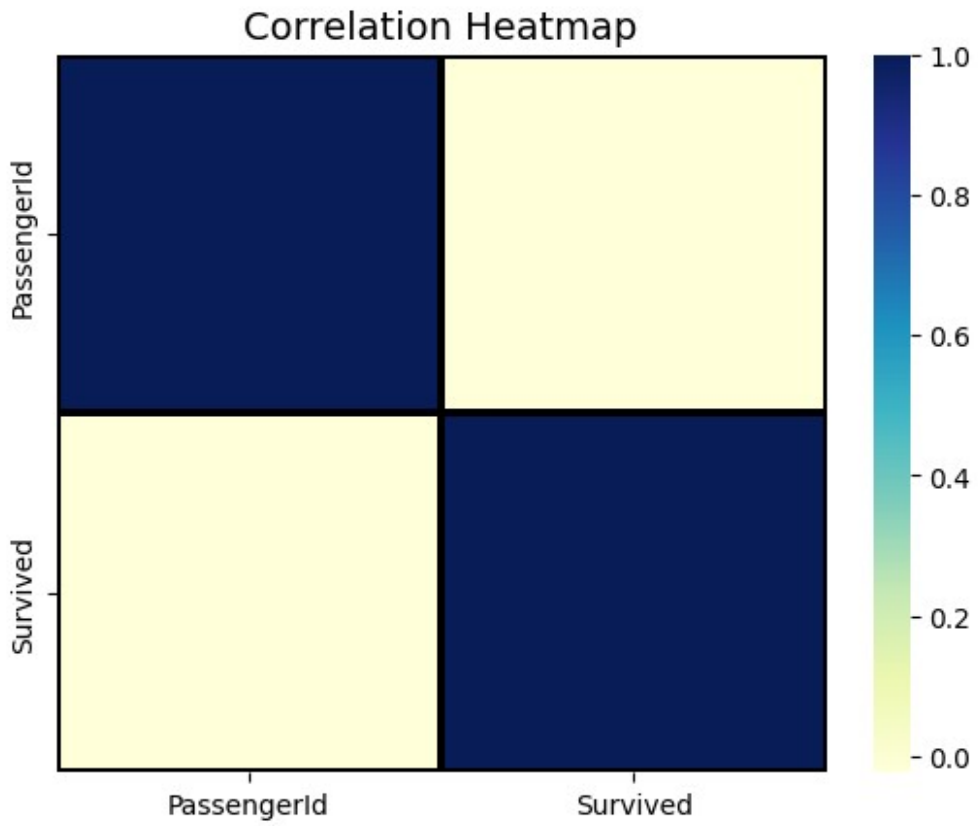

```
sns.pairplot(df)
```

HEATMAP THROUGH SEABORN

```
df=pd.read_csv("C:/Users/lekhs/OneDrive/Documents/titanic ds/gender_submission.csv")

sns.heatmap(df)
plt.show()
```

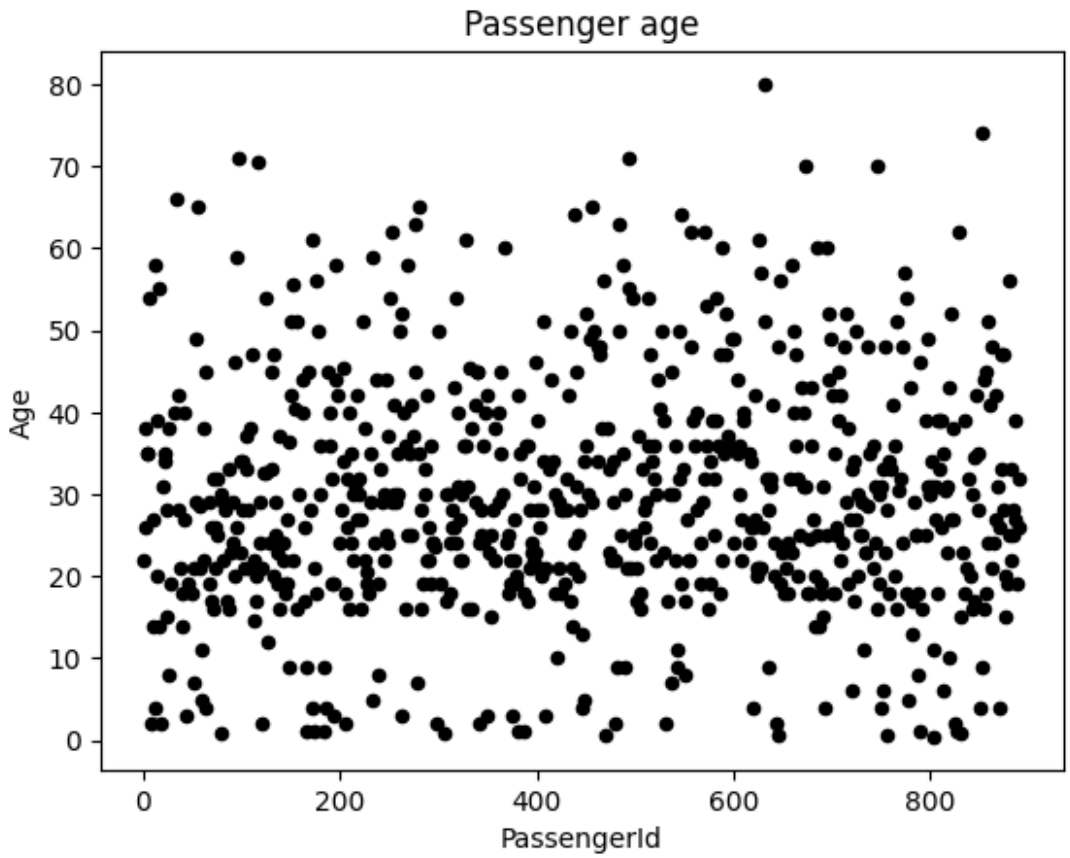



```
sns.heatmap(corr,
            annot=False,
            fmt=".1f",
            cmap='YlGnBu',
            linewidths=1.5,
            linecolor='black',
            cbar=True)
plt.title("Correlation Heatmap", fontsize=14)
plt.show()
```

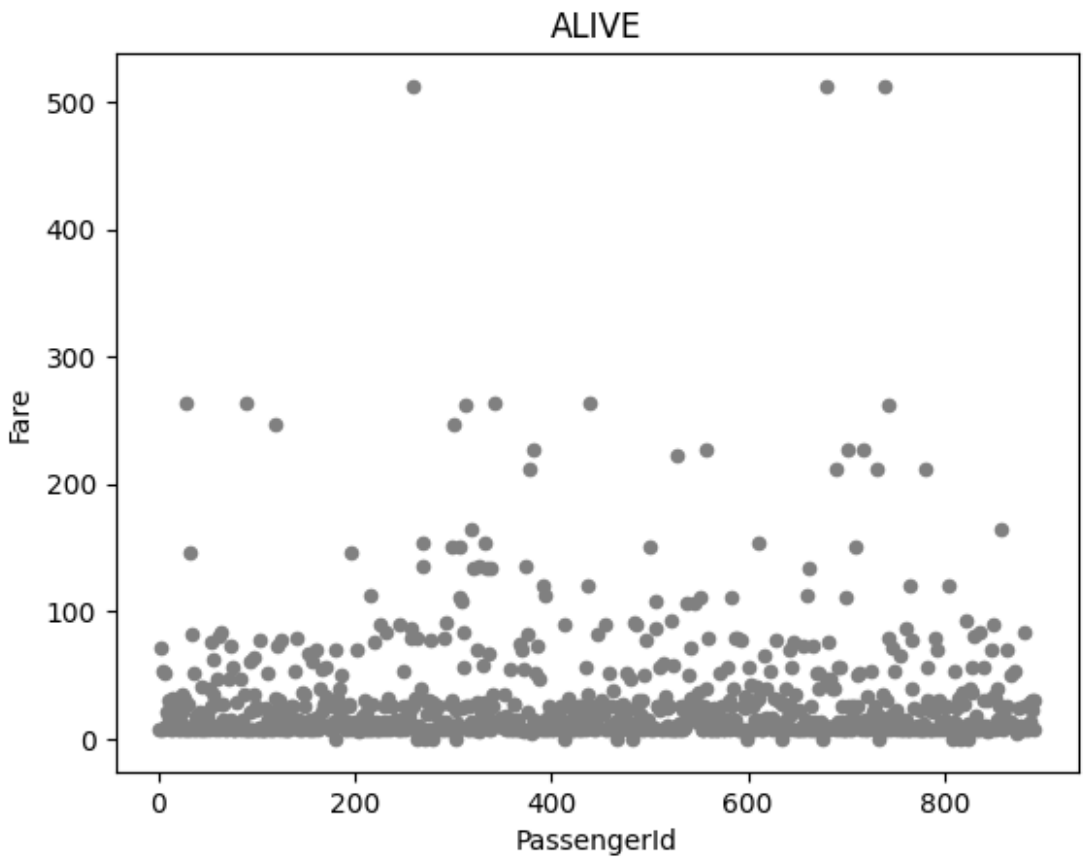


VISUALIZATION BY SCATTER PLOT

```
df=pd.read_csv("C:/Users/lekhs/OneDrive/Documents/titanic
ds/train.csv")

df.plot(kind='scatter', x="PassengerId", y="Age", color="black")
plt.title("Passenger age")
plt.ylabel("Age")
plt.xlabel("PassengerId")
plt.show()
```

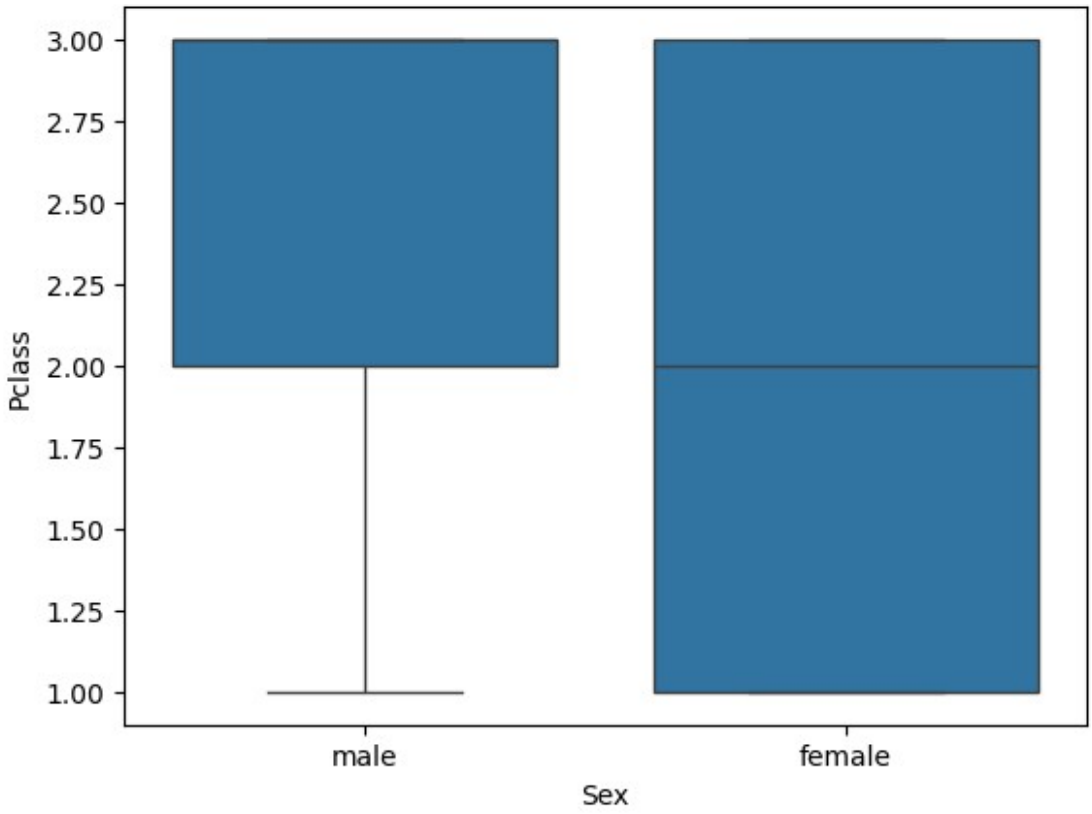



```
df.plot(kind='scatter', x="PassengerId", y="Fare", color="grey")
plt.title("ALIVE")
plt.ylabel("Fare")
plt.xlabel("PassengerId")
plt.show()
```



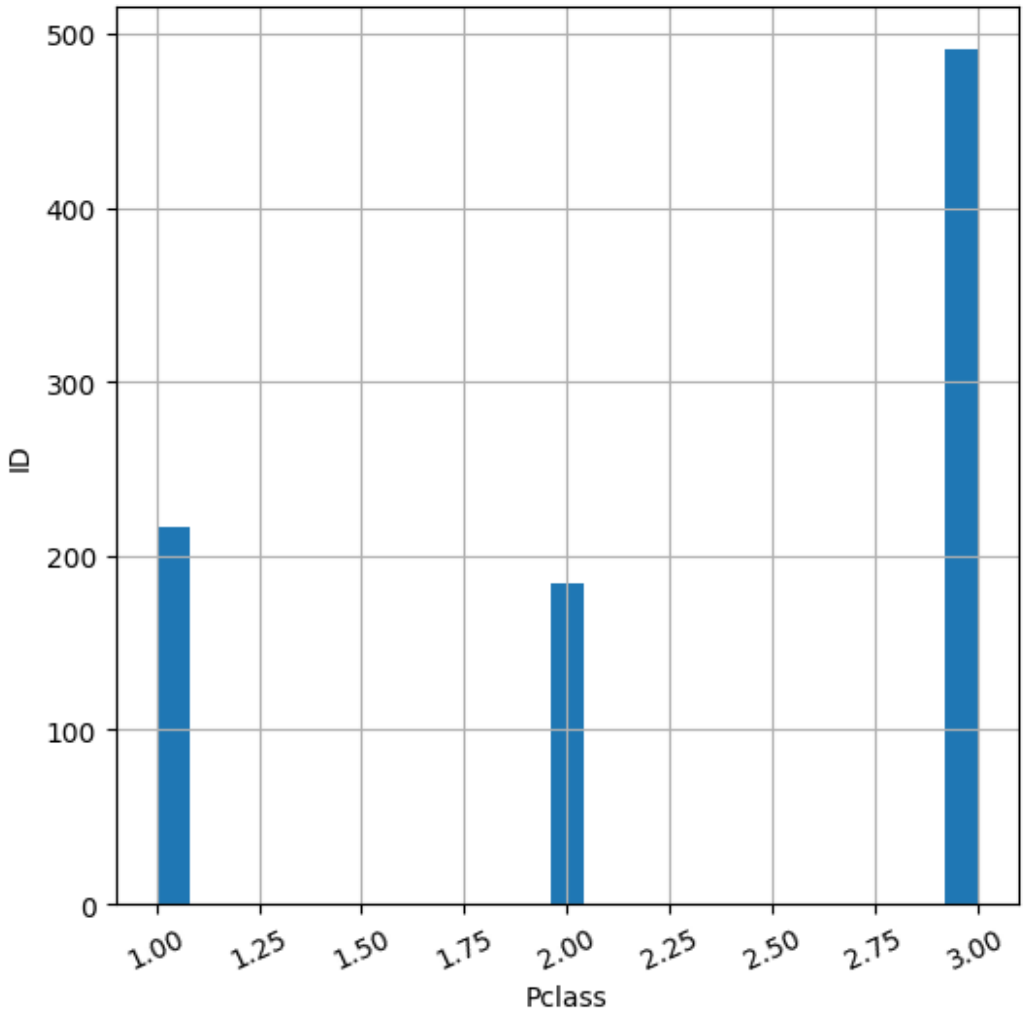
VISUALIZATION BY BOX PLOT

```
sns.boxplot(x="Sex", y="Pclass", data=df)
plt.show()
```



VISUALIZATION BY HISTOGRAMS

```
df["Pclass"].hist(bins=25,figsize=(6,6), xrot=25, grid=True)
plt.xlabel("Pclass"),
plt.ylabel("ID"),
plt.show()
```


500
400
300
200
100
0
ID
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
Pclass