```
import pandas as pd
df = pd.read_csv('train.csv')
```

```
df.head()
```

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare | Cabin | Embarked |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 22.0 | 1 | 0 | A/5 21171 | 7.2500 | NaN | S |
| **1** | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | PC 17599 | 71.2833 | C85 | C |
| **2** | 3 | 1 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | 0 | STON/O2. 3101282 | 7.9250 | NaN | S |
| **3** | 4 | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35.0 | 1 | 0 | 113803 | 53.1000 | C123 | S |

Next steps: Generate code with df View recommended plots New interactive sheet

```
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.describe()
```

| | PassengerId | Survived | Pclass | Age | SibSp | Parch | Fare |
|---|---|---|---|---|---|---|---|
| **count** | 891.000000 | 891.000000 | 891.000000 | 714.000000 | 891.000000 | 891.000000 | 891.000000 |
| **mean** | 446.000000 | 0.383838 | 2.308642 | 29.699118 | 0.523008 | 0.381594 | 32.204208 |
| **std** | 257.353842 | 0.486592 | 0.836071 | 14.526497 | 1.102743 | 0.806057 | 49.693429 |
| **min** | 1.000000 | 0.000000 | 1.000000 | 0.420000 | 0.000000 | 0.000000 | 0.000000 |
| **25%** | 223.500000 | 0.000000 | 2.000000 | 20.125000 | 0.000000 | 0.000000 | 7.910400 |
| **50%** | 446.000000 | 0.000000 | 3.000000 | 28.000000 | 0.000000 | 0.000000 | 14.454200 |
| **75%** | 668.500000 | 1.000000 | 3.000000 | 38.000000 | 1.000000 | 0.000000 | 31.000000 |
| **max** | 891.000000 | 1.000000 | 3.000000 | 80.000000 | 8.000000 | 6.000000 | 512.329200 |

```
df.value_counts()
```

```
                                                                                                                                  count
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
```

```
df.dtypes
```

```
                   0
PassengerId    int64
Survived       int64
Pclass         int64
Name          object
Sex           object
Age          float64
SibSp          int64
Parch          int64
Ticket        object
Fare         float64
Cabin         object
Embarked      object

dtype: object
```

```
df.isnull()
```

```
     PassengerId Survived Pclass  Name   Sex   Age SibSp Parch Ticket  Fare Cabin Embarked
0          False    False  False False False False False False  False False  True    False
1          False    False  False False False False False False  False False False    False
2          False    False  False False False False False False  False False  True    False
3          False    False  False False False False False False  False False False    False
4          False    False  False False False False False False  False False  True    False
...          ...      ...    ...   ...   ...   ...   ...   ...    ...   ...   ...      ...
886        False    False  False False False False False False  False False  True    False
887        False    False  False False False False False False  False False False    False
888        False    False  False False False  True False False  False False  True    False
889        False    False  False False False False False False  False False False    False
890        False    False  False False False False False False  False False  True    False
891 rows × 12 columns
```

```
df.isnull().sum()
```

```
                0
PassengerId     0
   Survived     0
     Pclass     0
       Name     0
        Sex     0
        Age   177
      SibSp     0
      Parch     0
     Ticket     0
       Fare     0
      Cabin   687
   Embarked     2

dtype: int64
```

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

```
           count
   Sex
  male       577
female       314

dtype: int64
```

```
df['Embarked'].value_counts()
```

```
             count
Embarked
   S           644
   C           168
   Q            77

dtype: int64
```

```
df.dropna()
```

```
      PassengerId  Survived  Pclass                          Name     Sex   Age  SibSp  Parch   Ticket     Fare    Cabin  Embarked
  1             2         1       1     Cumings, Mrs. John Bradley  female  38.0      1      0       PC  71.2833      C85         C
                                         (Florence Briggs Th...                                   17599
  3             4         1       1   Futrelle, Mrs. Jacques Heath  female  35.0      1      0   113803  53.1000     C123         S
                                                (Lily May Peel)
  6             7         0       1          McCarthy, Mr. Timothy J    male  54.0      0      0    17463  51.8625      E46         S
 10            11         1       3    Sandstrom, Miss. Marguerite  female   4.0      1      1  PP 9549  16.7000       G6         S
                                                             Rut
 11            12         1       1        Bonnell, Miss. Elizabeth  female  58.0      0      0   113783  26.5500     C103         S
...           ...       ...     ...                             ...     ...   ...    ...    ...      ...      ...      ...       ...
871           872         1       1      Beckwith, Mrs. Richard  female  47.0      1      1    11751  52.5542      D35         S
                                      Leonard (Sallie Monypeny)
872           873         0       1        Carlsson, Mr. Frans Olof    male  33.0      0      0      695   5.0000  B51 B53         S
                                                                                                                     B55
879           880         1       1     Potter, Mrs. Thomas Jr (Lily  female  56.0      0      1    11767  83.1583      C50         C
```

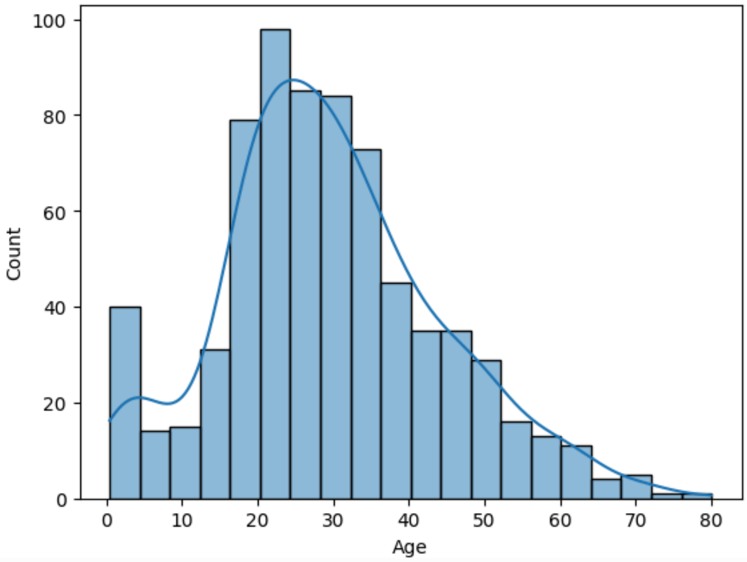
```
import seaborn as sns
import matplotlib.pyplot as plt


# imported Seaborn and matplotlib libreries for data visualisation
```

Histogram

```
sns.histplot(df['Age'], kde = True)
```

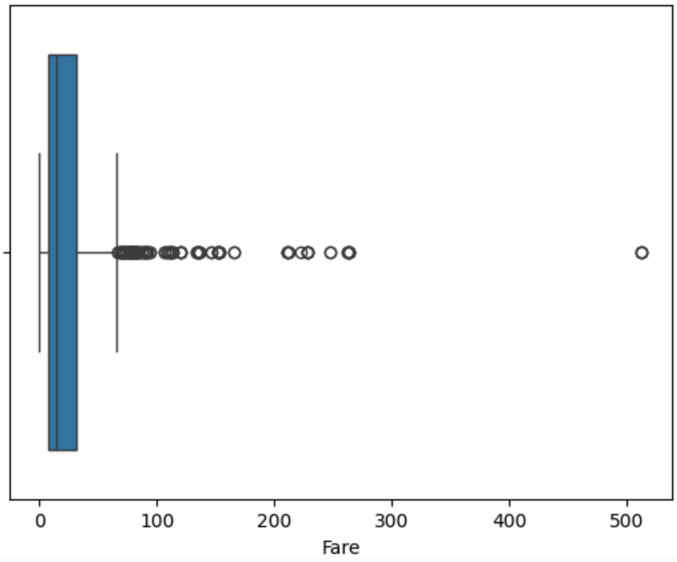
```
<Axes: xlabel='Age', ylabel='Count'>
```



```
#The most number of passangers are from 20-40 Age group.
```

Boxplot

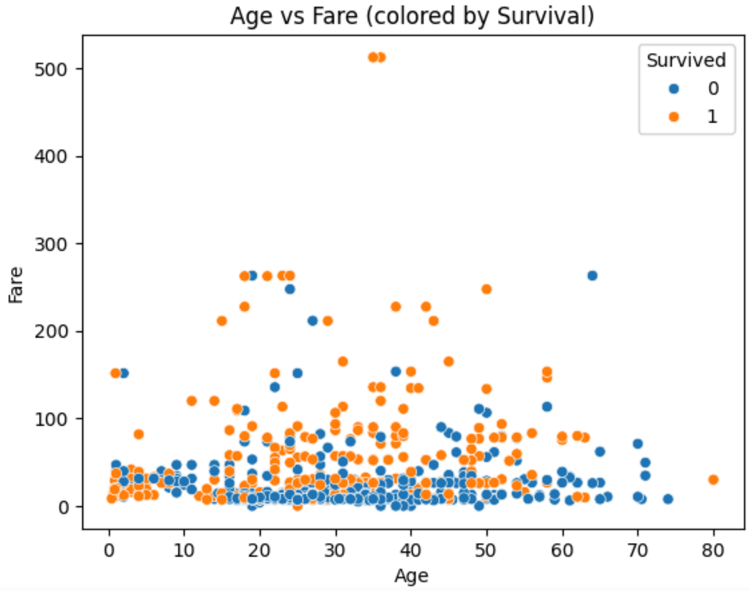
```
sns.boxplot(x=df['Fare'])
plt.show()
```



```
#Most passengers paid fares under 50 units
```

Scatterplot

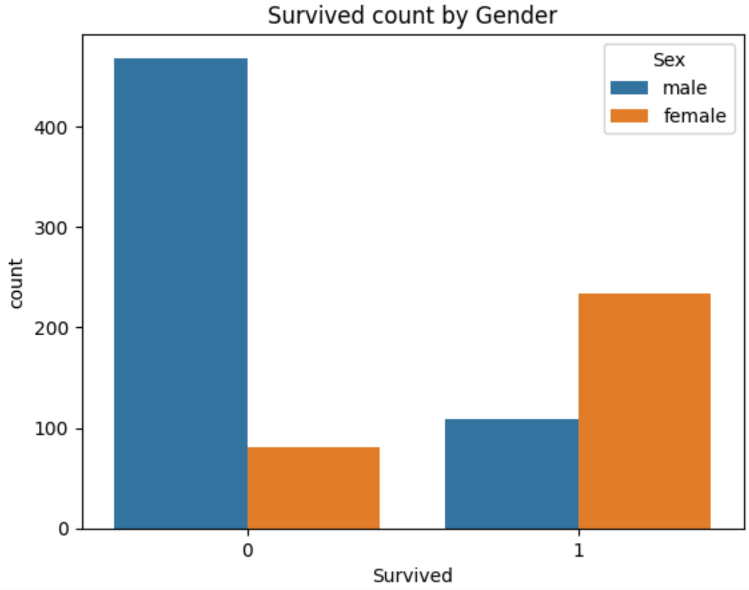
```
sns.scatterplot(x='Age', y='Fare', hue='Survived', data=df)
plt.title('Age vs Fare (colored by Survival)')
plt.show()
```



```
#The Maximum sales of the ticket are from 0-150 price range
```

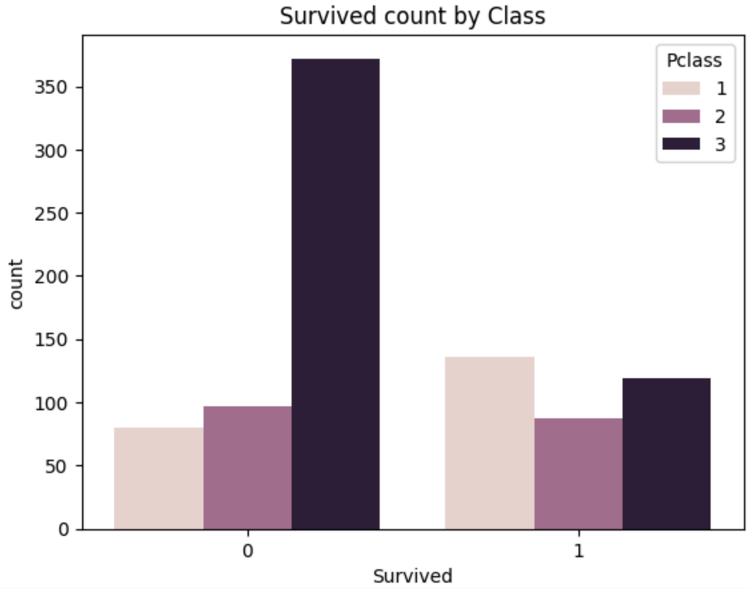
Countplots

```
sns.countplot(x='Survived', hue='Sex', data = df)
plt.title("Survived count by Gender")
plt.show()
```



```
# Females had a significantly higher survival rate compared to males.<br>
```

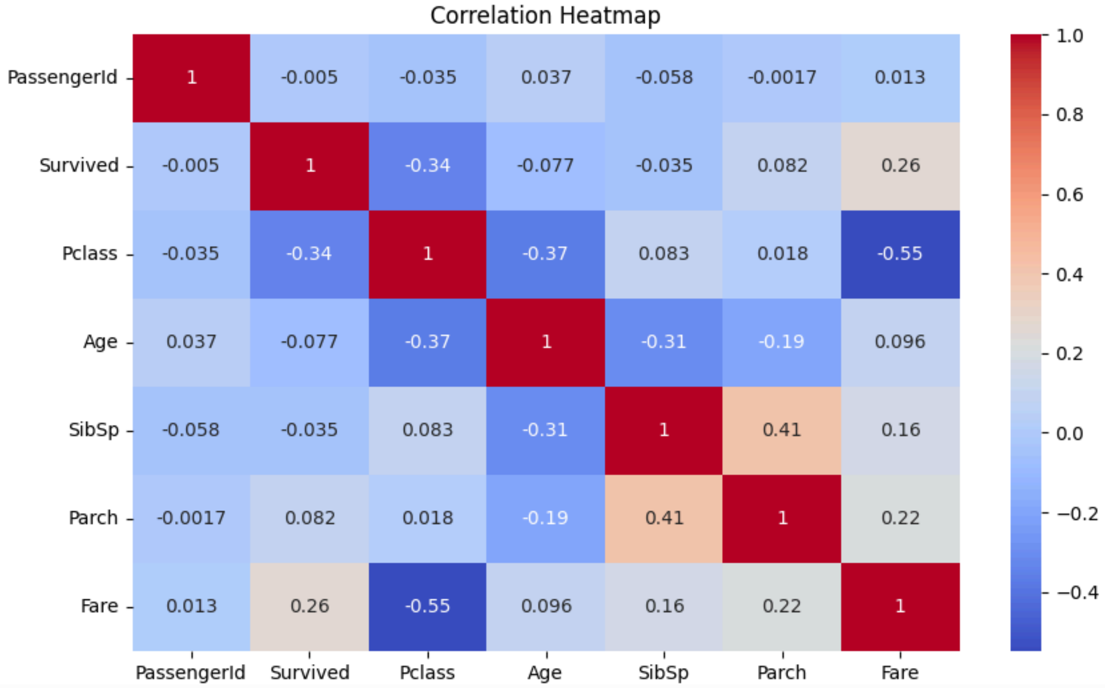
```
sns.countplot(x='Survived', hue='Pclass',data = df)
plt.title("Survived count by Class")
plt.show()
```



```
# Passengers in 1st class had a higher chance of survival compared to 2nd and 3rd class.
# 3rd class passengers were the most affected in terms of survival.
```

Heatmap

```
plt.figure(figsize=(10,6))
sns.heatmap(df.corr(numeric_only=True),annot=True,cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

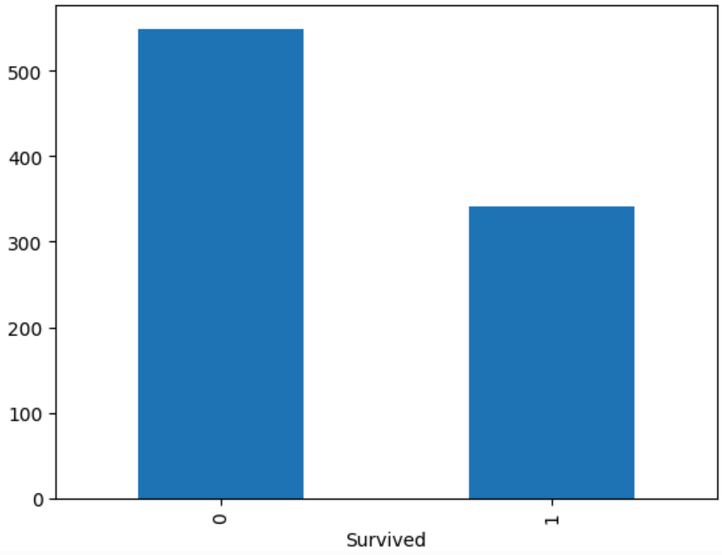


|  | PassengerId | Survived | Pclass | Age | SibSp | Parch | Fare |
|---|---|---|---|---|---|---|---|
| PassengerId | 1 | -0.005 | -0.035 | 0.037 | -0.058 | -0.0017 | 0.013 |
| Survived | -0.005 | 1 | -0.34 | -0.077 | -0.035 | 0.082 | 0.26 |
| Pclass | -0.035 | -0.34 | 1 | -0.37 | 0.083 | 0.018 | -0.55 |
| Age | 0.037 | -0.077 | -0.37 | 1 | -0.31 | -0.19 | 0.096 |
| SibSp | -0.058 | -0.035 | 0.083 | -0.31 | 1 | 0.41 | 0.16 |
| Parch | -0.0017 | 0.082 | 0.018 | -0.19 | 0.41 | 1 | 0.22 |
| Fare | 0.013 | 0.26 | -0.55 | 0.096 | 0.16 | 0.22 | 1 |

```
# Pclass and Fare are negatively correlated.
#Survived has positive correlation with Fare and negative with Pclass, meaning higher-class and higher-paying passengers were more likel
```

```
df['Survived'].value_counts().plot(kind='bar')
```

```
<Axes: xlabel='Survived'>
```



```
#The number of passengers who did not survive (label 0) is significantly higher than those who did. This shows that the majority of pass
```