# titanic-analysis

February 2, 2024

[207]:
```
import pandas as pd
```

[208]:
```
data = pd.read_csv(r"E:\PYTHON\Exploratory Data Analysis Projects\Titanic␣
↪Analysis Project\Dataset\Titanic.csv")
```

### Display Top5 Rows

[209]:
```
data.head(5)
```

[209]:

| | 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 |
| 4 | 5 | 0 | 3 | Allen, Mr. William Henry | male | 35.0 | 0 | 0 | 373450 | 8.0500 | NaN | S |

### Display Last5 Rows

[210]:
```
data.tail(5)
```

[210]:

| | PassengerId | Survived | Pclass | Name |
|---|---|---|---|---|
| 886 | 887 | 0 | 2 | Montvila, Rev. Juozas |
| 887 | 888 | 1 | 1 | Graham, Miss. Margaret Edith |
| 888 | 889 | 0 | 3 | Johnston, Miss. Catherine Helen "Carrie" |

```
889          890         1       1                     Behr, Mr. Karl Howell
890          891         0       3                       Dooley, Mr. Patrick

        Sex   Age  SibSp  Parch      Ticket   Fare Cabin Embarked
886    male  27.0      0      0      211536  13.00   NaN        S
887  female  19.0      0      0      112053  30.00   B42        S
888  female   NaN      1      2  W./C. 6607  23.45   NaN        S
889    male  26.0      0      0      111369  30.00  C148        C
890    male  32.0      0      0      370376   7.75   NaN        Q
```

### Shape Of Our Dataset

[211]:
```
data.shape
```

[211]:
```
(891, 12)
```

[212]:
```
print("No. of Rows:", data.shape[0])
print("No. of Columns:" , data.shape[1])
```

```
No. of Rows: 891
No. of Columns: 12
```

### Get Information About Dataset

[213]:
```
data.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
```

### Get Overall Statistics

[214]:
```
data.describe()
```

[214]:
```
       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
```

**Data Filtering**

[215]:
```
data.columns
```

[215]:
```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

[216]:
```
data[['Name','Age']]
```

[216]:
```
                                                 Name   Age
0                             Braund, Mr. Owen Harris  22.0
1    Cumings, Mrs. John Bradley (Florence Briggs Th…  38.0
2                              Heikkinen, Miss. Laina  26.0
3        Futrelle, Mrs. Jacques Heath (Lily May Peel)  35.0
4                            Allen, Mr. William Henry  35.0
..                                                ...   …
886                            Montvila, Rev. Juozas  27.0
887                     Graham, Miss. Margaret Edith  19.0
888         Johnston, Miss. Catherine Helen "Carrie"   NaN
889                            Behr, Mr. Karl Howell  26.0
890                              Dooley, Mr. Patrick  32.0

[891 rows x 2 columns]
```

[217]:
```
sum(data['Sex']=='male')
```

[217]: 577

[218]:
```
data[data['Sex']=='male'].head(5)
```

[218]:
```
   PassengerId  Survived  Pclass                            Name   Sex   Age  \
0            1         0       3         Braund, Mr. Owen Harris  male  22.0
4            5         0       3        Allen, Mr. William Henry  male  35.0
5            6         0       3                Moran, Mr. James  male   NaN
6            7         0       1         McCarthy, Mr. Timothy J  male  54.0
7            8         0       3  Palsson, Master. Gosta Leonard  male   2.0

   SibSp  Parch     Ticket     Fare Cabin Embarked
0      1      0  A/5 21171   7.2500   NaN        S
4      0      0     373450   8.0500   NaN        S
5      0      0     330877   8.4583   NaN        Q
6      0      0      17463  51.8625   E46        S
7      3      1     349909  21.0750   NaN        S
```

[219]:
```
sum(data['Survived']==1)
```

[219]: 342

[220]:
```
data[data['Survived']==1].head()
```

[220]:
```
   PassengerId  Survived  Pclass  \
1            2         1       1
2            3         1       3
3            4         1       1
8            9         1       3
9           10         1       2

                                                 Name     Sex   Age  SibSp  \
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
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
1      0          PC 17599  71.2833   C85        C
2      0  STON/O2. 3101282   7.9250   NaN        S
3      0            113803  53.1000  C123        S
8      2            347742  11.1333   NaN        S
9      0            237736  30.0708   NaN        C
```

### Checking Null Values

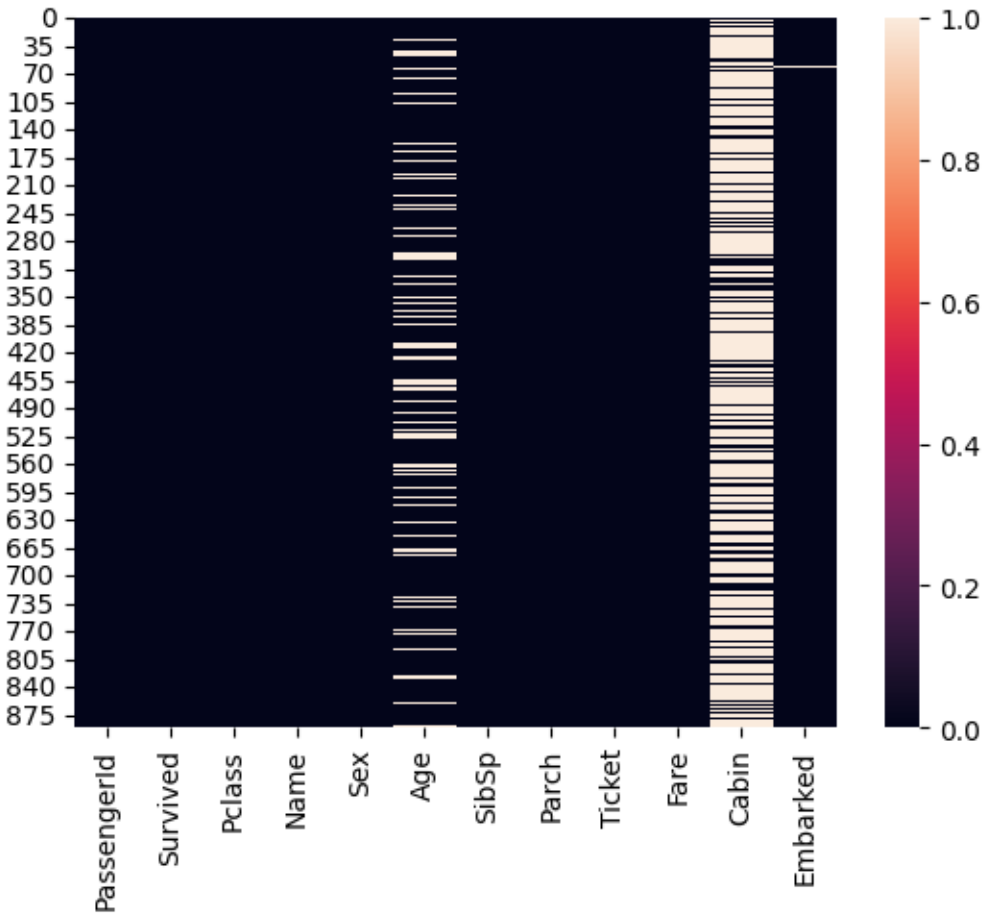
[221]:
```
data.isnull().sum()
```

[221]:
```
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
```

[222]:
```
import seaborn as sns
import matplotlib.pyplot as plt
```

[223]:
```
sns.heatmap(data.isnull())
```

[223]:
```
<Axes: >
```




[224]:
```
per_missing  = data.isnull().sum()*100/len(data)
per_missing
```

[224]:
```
PassengerId     0.000000
Survived        0.000000
Pclass          0.000000
Name            0.000000
Sex             0.000000
Age            19.865320
SibSp           0.000000
Parch           0.000000
Ticket          0.000000
Fare            0.000000
Cabin          77.104377
Embarked        0.224467
dtype: float64
```

### Drop The Column

[225]:
```
data.drop('Cabin', axis=1, inplace=True)
```

[226]:
```
data.isnull().sum()
```

[226]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Embarked         2
dtype: int64
```

### Handle Missing Values

[227]:
```
data['Embarked'].mode()
```

[227]:
```
0    S
Name: Embarked, dtype: object
```

[228]:
```
data['Embarked'].fillna('S', inplace=True)
```

[229]:
```
data.isnull().sum()
```

[229]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Embarked         0
dtype: int64
```

[230]:
```
data['Age'].fillna(data['Age'].mean(), inplace=True)
```

[231]:
```
data.isnull().sum()
```

[231]:
```
PassengerId    0
Survived       0
```

```
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       0
dtype: int64
```

### Categorical Data Encoding

[232]:
```
data['Sex'].unique()
```

[232]:
```
array(['male', 'female'], dtype=object)
```

[233]:
```
data['Gender'] = data['Sex'].map({'male':1, 'female':0})
```

[234]:
```
data.head(1)
```

[234]:
```
   PassengerId  Survived  Pclass                     Name   Sex   Age  SibSp  \
0            1         0       3  Braund, Mr. Owen Harris  male  22.0      1

   Parch     Ticket  Fare Embarked  Gender
0      0  A/5 21171  7.25        S       1
```

[235]:
```
x = data['Sex'].map({'male':1, 'female':0})
```

[236]:
```
data.insert(5,'Gender_New',x)
```

[237]:
```
data.head(1)
```

[237]:
```
   PassengerId  Survived  Pclass                     Name   Sex  Gender_New  \
0            1         0       3  Braund, Mr. Owen Harris  male           1

    Age  SibSp  Parch     Ticket  Fare Embarked  Gender
0  22.0      1      0  A/5 21171  7.25        S       1
```

[238]:
```
data['Embarked'].unique()
```

[238]:
```
array(['S', 'C', 'Q'], dtype=object)
```

[239]:
```
pd.get_dummies(data, columns=['Embarked'])
```

[239]:
```
     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  Gender_New  \
0                                Braund, Mr. Owen Harris    male           1
1    Cumings, Mrs. John Bradley (Florence Briggs Th…  female           0
2                                 Heikkinen, Miss. Laina  female           0
3         Futrelle, Mrs. Jacques Heath (Lily May Peel)  female           0
4                               Allen, Mr. William Henry    male           1
..                                                   …     …         …
886                               Montvila, Rev. Juozas    male           1
887                        Graham, Miss. Margaret Edith  female           0
888            Johnston, Miss. Catherine Helen "Carrie"  female           0
889                               Behr, Mr. Karl Howell    male           1
890                                 Dooley, Mr. Patrick    male           1

           Age  SibSp  Parch            Ticket     Fare  Gender  Embarked_C  \
0    22.000000      1      0         A/5 21171   7.2500       1       False
1    38.000000      1      0          PC 17599  71.2833       0        True
2    26.000000      0      0  STON/O2. 3101282   7.9250       0       False
3    35.000000      1      0            113803  53.1000       0       False
4    35.000000      0      0            373450   8.0500       1       False
..         …    …    …                …      …     …         …
886  27.000000      0      0            211536  13.0000       1       False
887  19.000000      0      0            112053  30.0000       0       False
888  29.699118      1      2        W./C. 6607  23.4500       0       False
889  26.000000      0      0            111369  30.0000       1        True
890  32.000000      0      0            370376   7.7500       1       False

     Embarked_Q  Embarked_S
0         False        True
1         False       False
2         False        True
3         False        True
4         False        True
..          …         …
886       False        True
887       False        True
888       False        True
```

```
889       False       False
890        True       False

[891 rows x 15 columns]
```

[240]:
```
data1 = pd.get_dummies(data, columns=['Embarked'], drop_first=True)
```

[241]:
```
data1.head(1)
```

[241]:
```
   PassengerId  Survived  Pclass                     Name   Sex  Gender_New  \
0            1         0       3  Braund, Mr. Owen Harris  male           1

    Age  SibSp  Parch     Ticket  Fare  Gender  Embarked_Q  Embarked_S
0  22.0      1      0  A/5 21171  7.25       1       False        True
```

**Univariate Analysis**

[ ]:
```

```

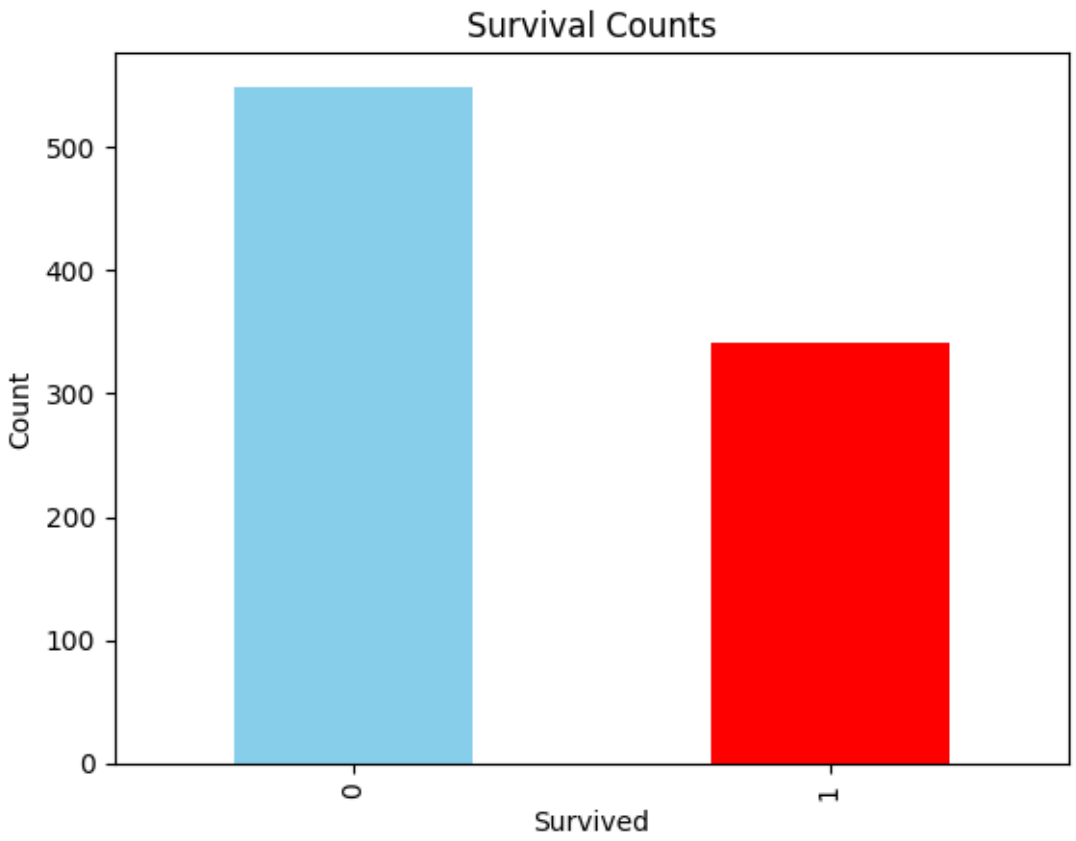
### How Many People Survived And How Many People Died

[242]:
```
data['Survived'].value_counts()
```

[242]:
```
Survived
0    549
1    342
Name: count, dtype: int64
```

[251]:
```
colors = ['skyblue', 'red']
data['Survived'].value_counts().plot(kind='bar', color=colors)
plt.xlabel('Survived')
plt.ylabel('Count')
plt.title('Survival Counts')
```

[251]:
```
Text(0.5, 1.0, 'Survival Counts')
```



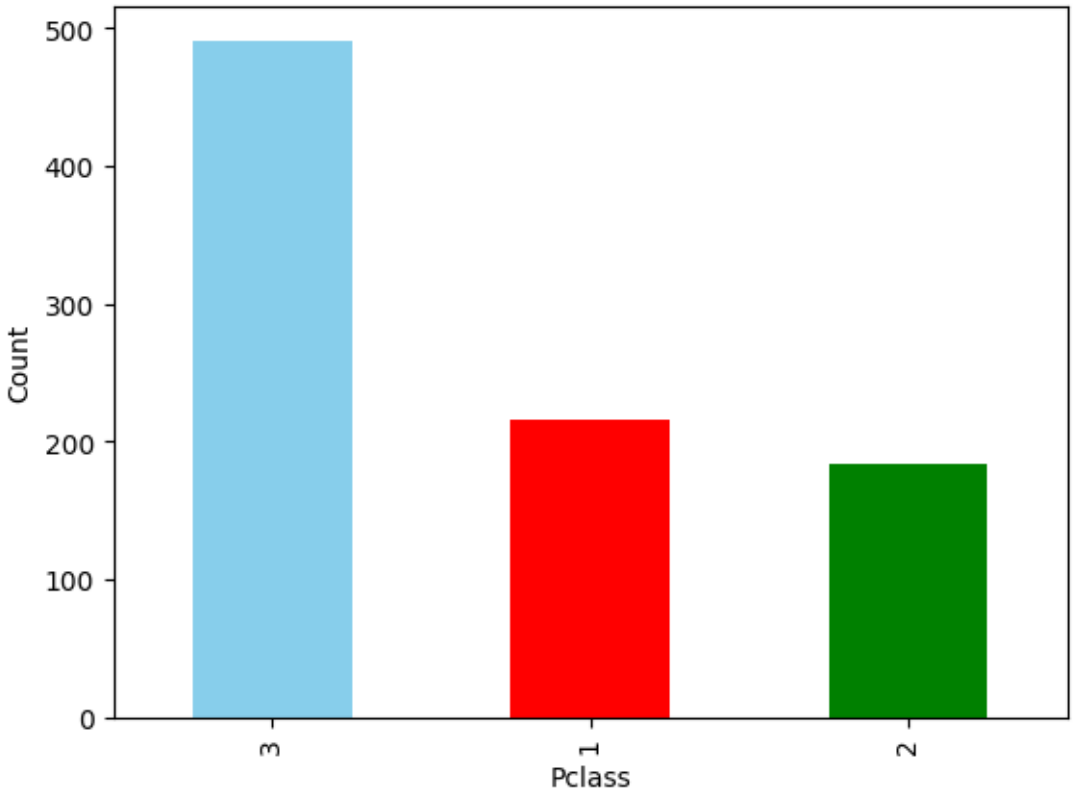

### How Many Passengers Were In First Class And Second Class And Third Class

[252]:
```
data['Pclass'].value_counts()
```

[252]:
```
Pclass
3    491
1    216
2    184
Name: count, dtype: int64
```

[255]:
```
colors = ['skyblue', 'red', 'green']
data['Pclass'].value_counts().plot(kind='bar', color=colors)
plt.xlabel('Pclass')
plt.ylabel('Count')
```

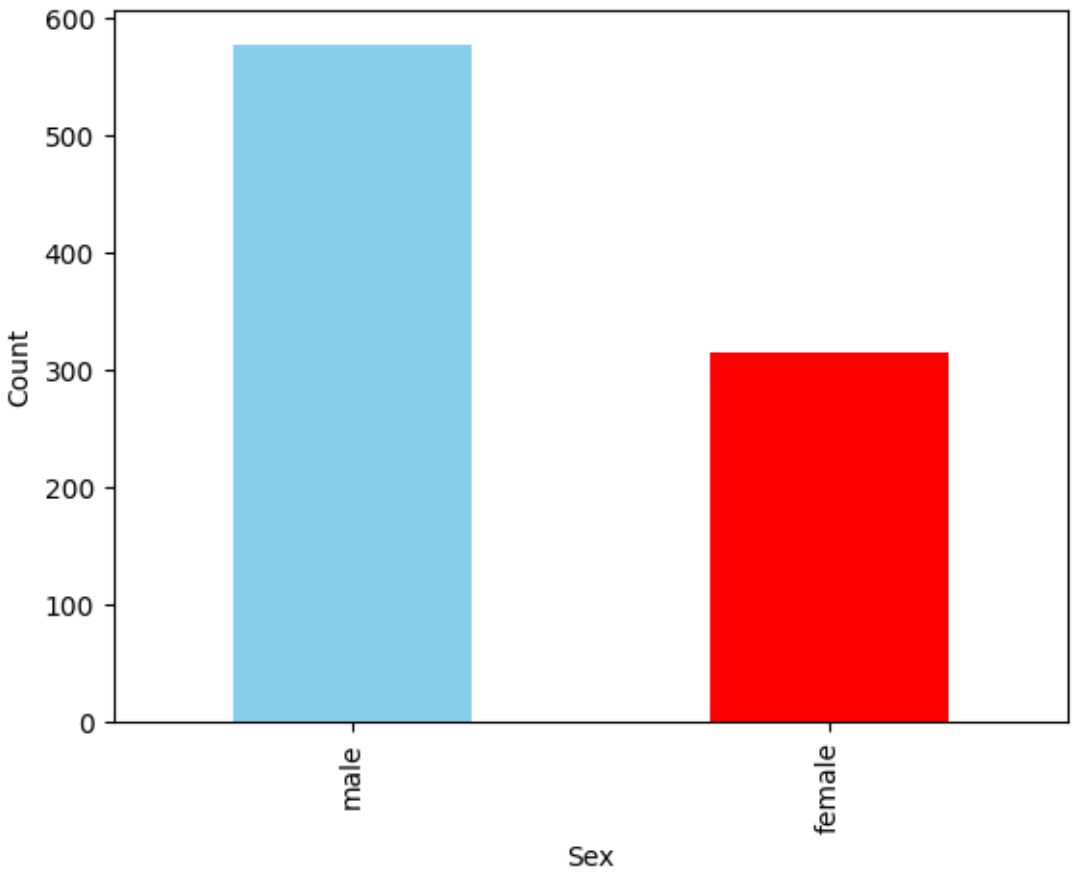
[255]:
```
Text(0, 0.5, 'Count')
```




**No. of Male And Female Passengers**

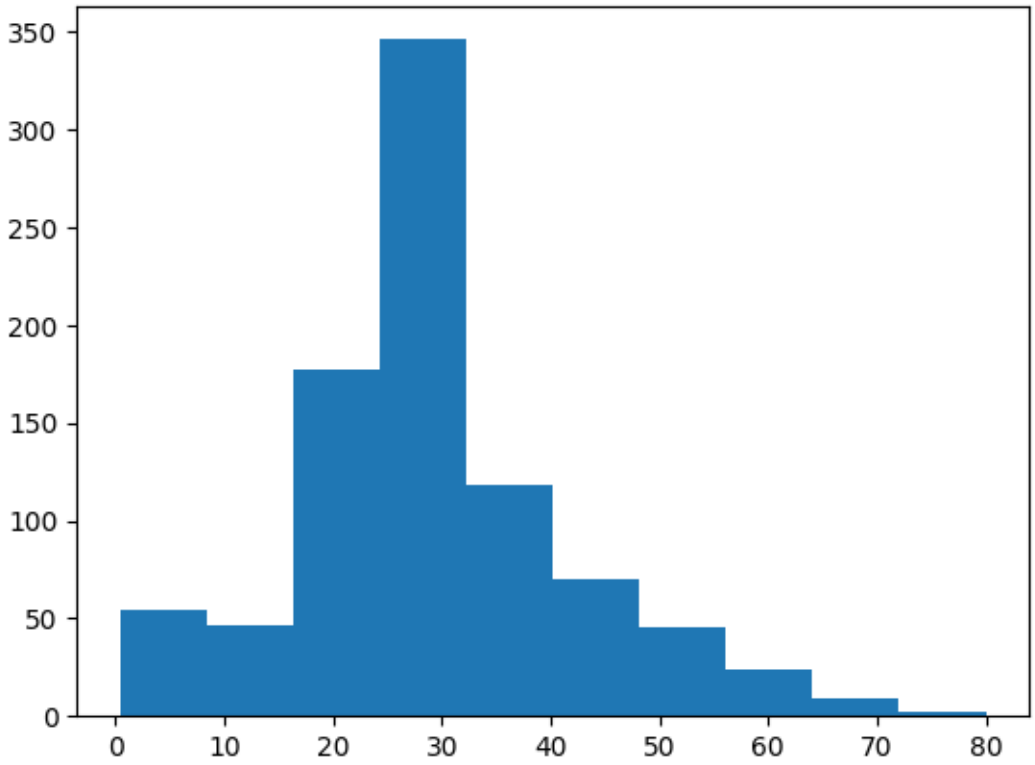
[256]:
```
colors = ['skyblue', 'red']
data['Sex'].value_counts().plot(kind='bar', color=colors)
plt.xlabel('Sex')
plt.ylabel('Count')
```

[256]:
```
Text(0, 0.5, 'Count')
```



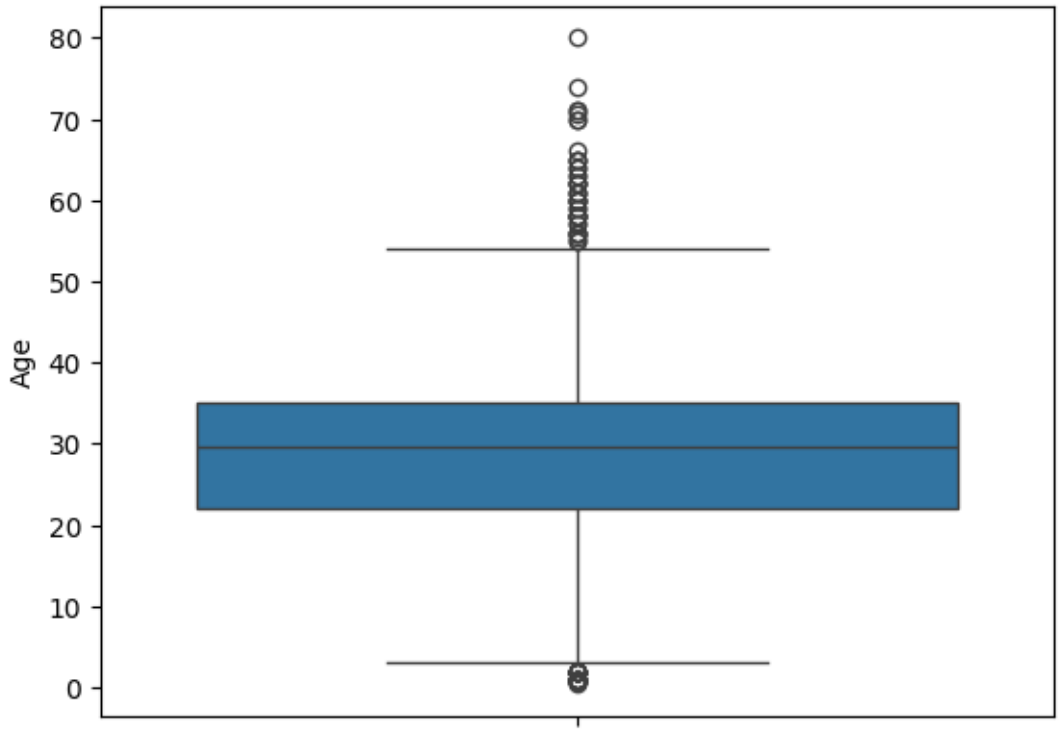

[257]:
```
plt.hist(data['Age'])
```

[257]:
```
(array([ 54.,  46., 177., 346., 118.,  70.,  45.,  24.,   9.,   2.]),
 array([ 0.42 ,  8.378, 16.336, 24.294, 32.252, 40.21 , 48.168, 56.126,
        64.084, 72.042, 80.   ]),
 <BarContainer object of 10 artists>)
```




[258]:
```
sns.boxplot(data['Age'])
```

[258]:
```
<Axes: ylabel='Age'>
```



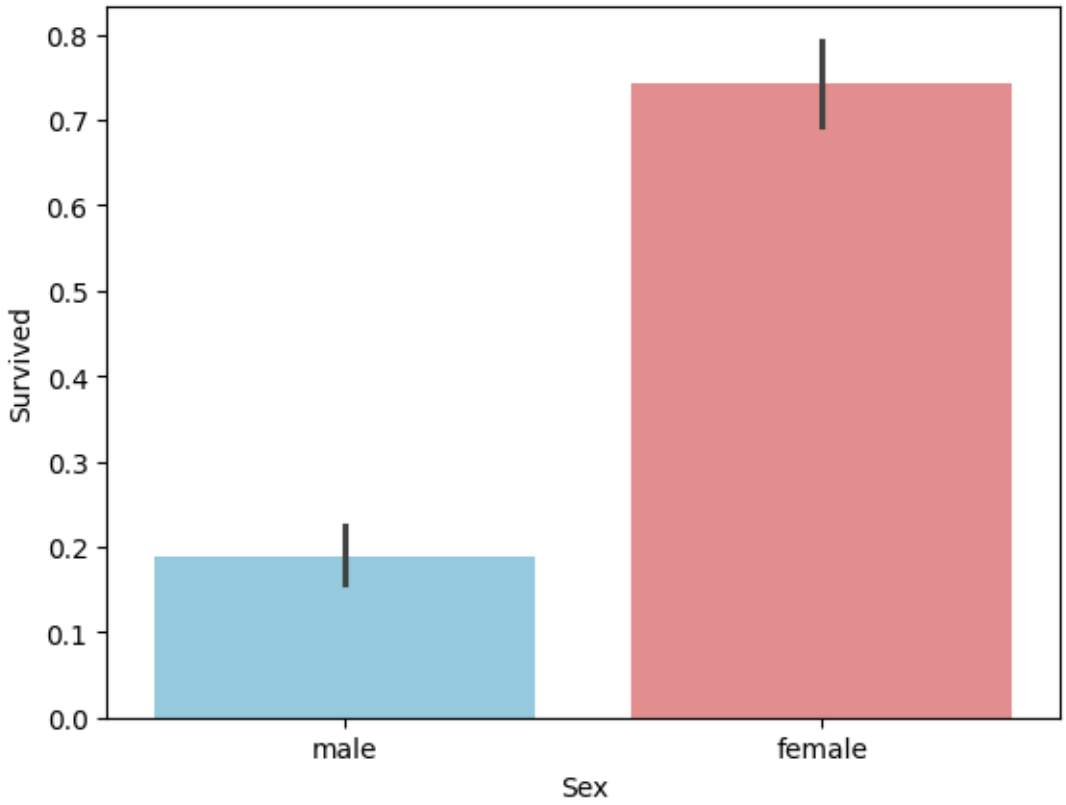

### 0.0.1 Bivariate Analysis

**How Has Better Chance Of Survival Male OR Female?**

[265]:
```
colors = ['skyblue', 'lightcoral', 'lightgreen']
sns.barplot(x='Sex', y='Survived', data=data, palette=colors)
```

```
C:\Users\Admin\AppData\Local\Temp\ipykernel_6324\2348904149.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.barplot(x='Sex', y='Survived', data=data, palette=colors)
C:\Users\Admin\AppData\Local\Temp\ipykernel_6324\2348904149.py:2: UserWarning:
The palette list has more values (3) than needed (2), which may not be intended.
  sns.barplot(x='Sex', y='Survived', data=data, palette=colors)
```

[265]: `<Axes: xlabel='Sex', ylabel='Survived'>`




**Which Passenger Class Have Better Survival Chance(First or Second or Third)?**

[261]:
```
data.columns
```

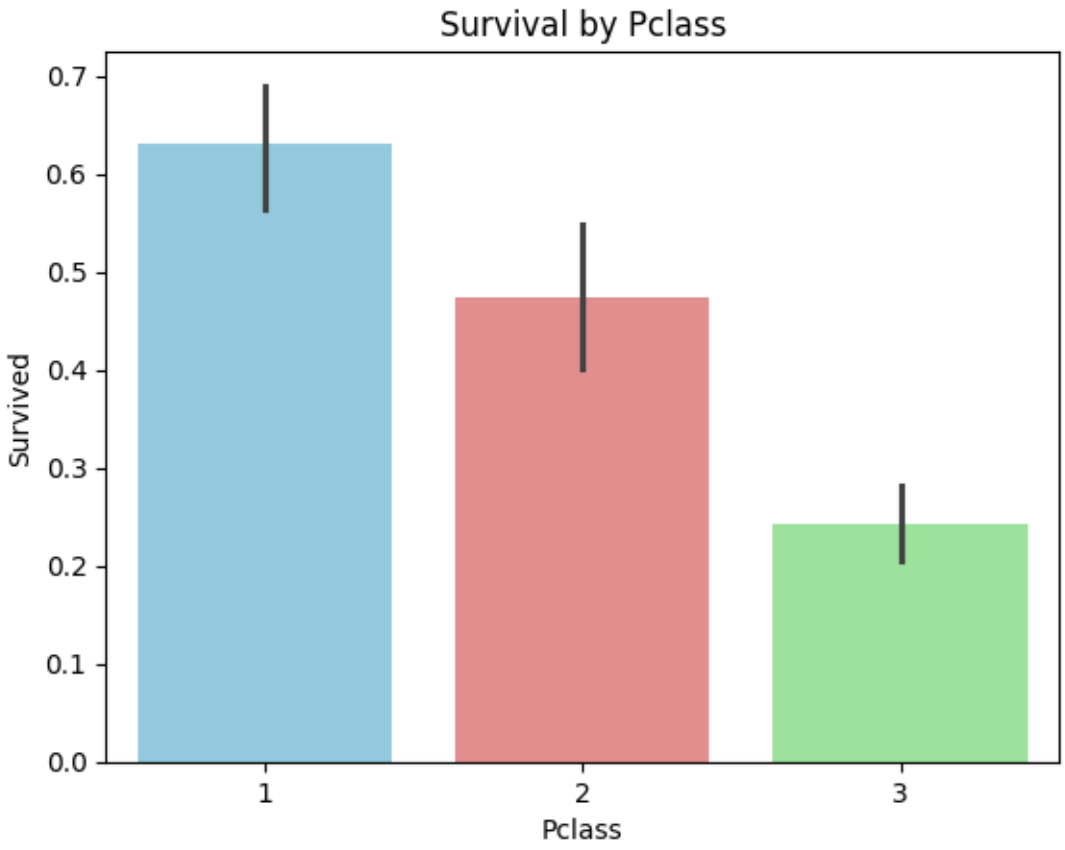
[261]:
```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Gender_New', 'Age',
       'SibSp', 'Parch', 'Ticket', 'Fare', 'Embarked', 'Gender'],
      dtype='object')
```

[264]:
```
colors = ['skyblue', 'lightcoral', 'lightgreen']

sns.barplot(x='Pclass', y='Survived', data=data, palette=colors)

plt.xlabel('Pclass')
plt.ylabel('Survived')
plt.title('Survival by Pclass')
```

```
C:\Users\Admin\AppData\Local\Temp\ipykernel_6324\3925696990.py:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.
```

```
sns.barplot(x='Pclass', y='Survived', data=data, palette=colors)
```

[264]:
```
Text(0.5, 1.0, 'Survival by Pclass')
```




**Feature Engineering**

[266]:
```
data['Family_Size'] = data['SibSp']+data['Parch']
```

[267]:
```
data.head(1)
```

[267]:

| | PassengerId | Survived | Pclass | Name | Sex | Gender_New | Age | SibSp | Parch | Ticket | Fare | Embarked | Gender | Family_Size |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 1 | 22.0 | 1 | 0 | A/5 21171 | 7.25 | S | 1 | 1 |

[268]:
```
data['Fare_Per_Person'] = data['Fare']/ (data['Family_Size']+1)
```

[269]:
```
data.head(1)
```

[269]:

| | PassengerId | Survived | Pclass | Name | Sex | Gender_New |
|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 1 |

| | Age | SibSp | Parch | Ticket | Fare | Embarked | Gender | Family_Size | Fare_Per_Person |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 22.0 | 1 | 0 | A/5 21171 | 7.25 | S | 1 | 1 | 3.625 |