

```
In [3]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [4]: df=pd.read_csv(r"C:\Users\USER\OneDrive\Desktop\Titanic dataset.csv")
```

```
In [11]: df.head()
```

```
Out[11]:
```

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

```
In [24]: df.describe()
```

```
Out[24]:
```

	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

In [25]: `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
```

In [27]: `df.drop('Cabin',axis=1,inplace=True)`In [28]: `df.head()`

Out[28]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	S

In [113... `df['Survived'].value_counts()`

```
Out[113]: 0    549
          1    342
          Name: Survived, dtype: int64
```

```
In [26]: numSurvived=df['Survived'].sum()
          print(numSurvived)
          totalPassenger=len(df['PassengerId'])
          print(totalPassenger)
          PerSurvived=[numSurvived/totalPassenger*100]
          print(PerSurvived,'% survived the Titanic disaster')

          342
          891
          [38.38383838383838] % survived the Titanic disaster
```

```
In [111... df['Age'].mean()
```

```
Out[111]: 29.69911764705882
```

```
In [6]: df['Pclass'].value_counts()
```

```
Out[6]: 3    491
          1    216
          2    184
          Name: Pclass, dtype: int64
```

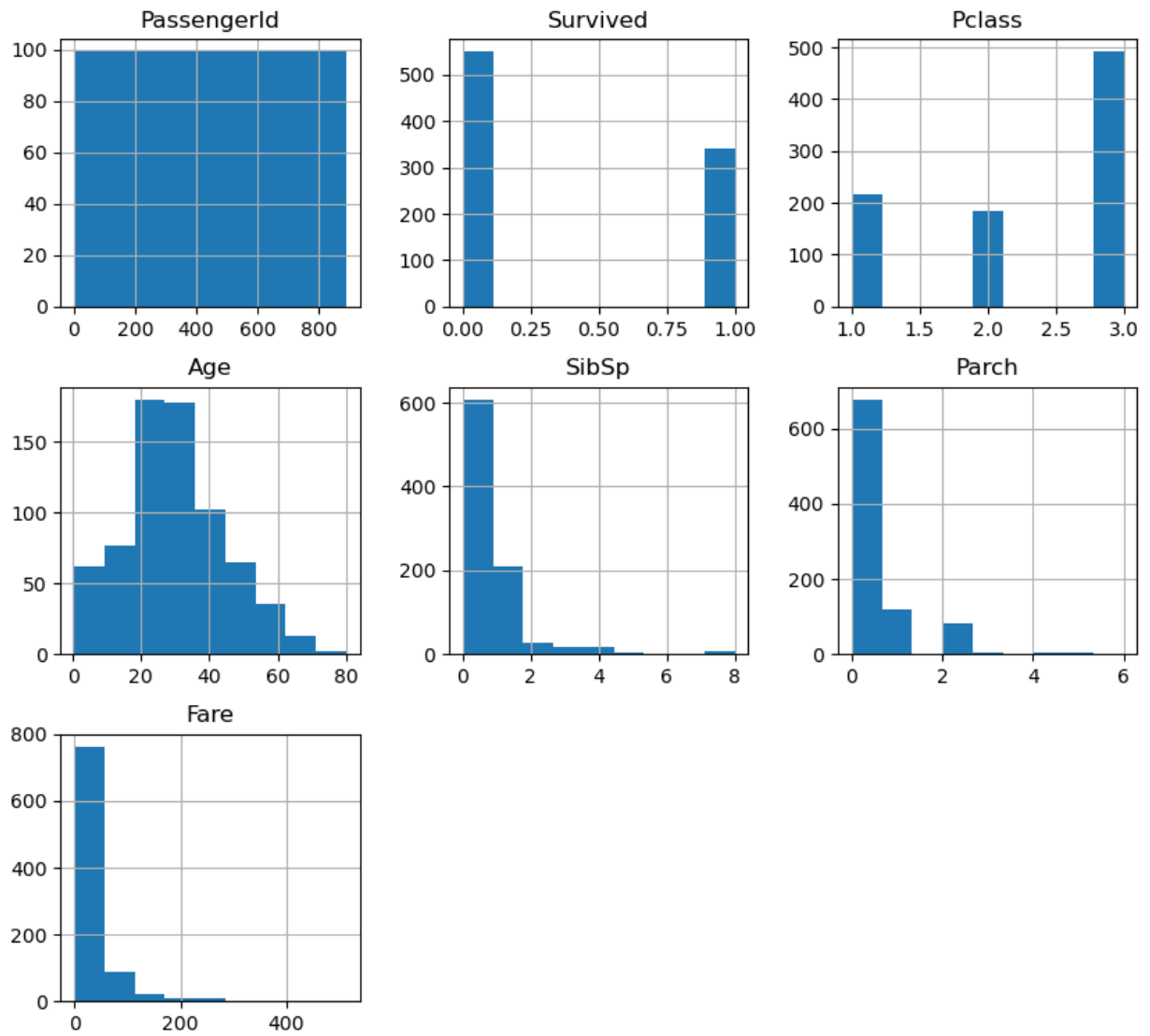
```
In [7]: df.groupby('Sex')['Survived'].mean()
```

```
Out[7]: Sex
female    0.742038
male      0.188908
          Name: Survived, dtype: float64
```

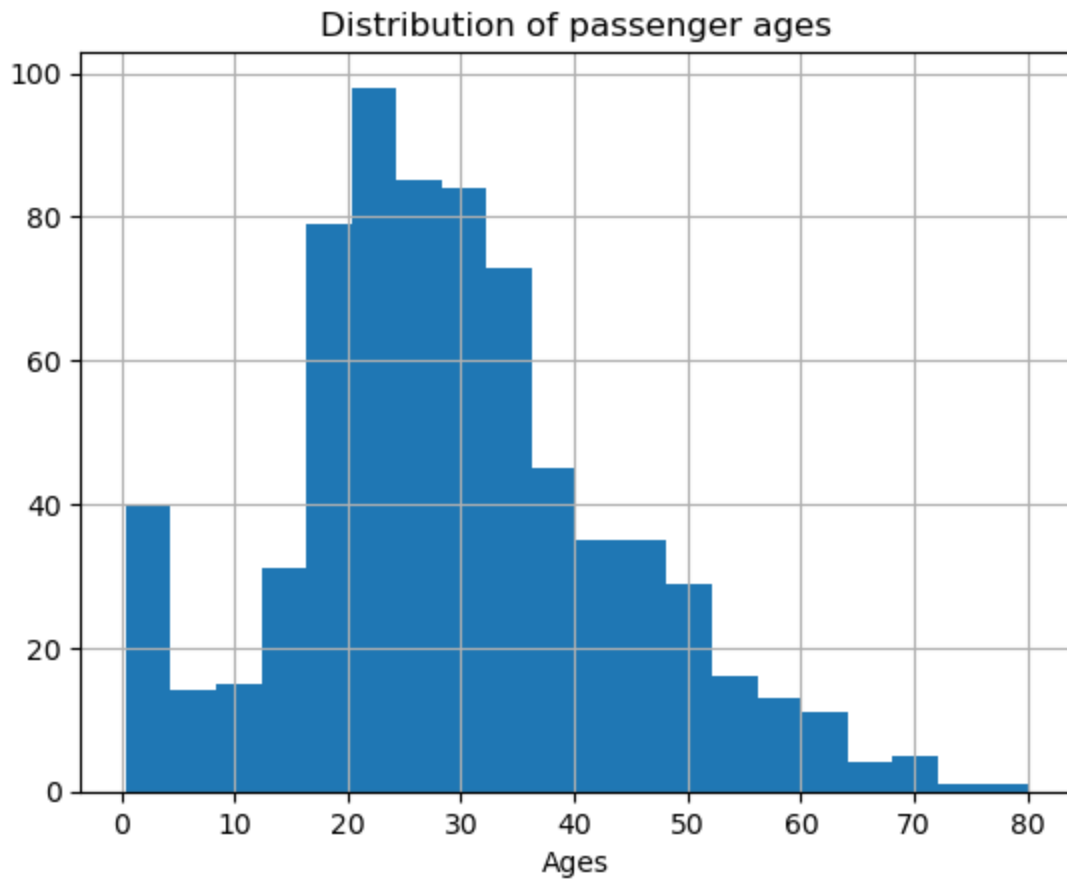
```
In [122... df.groupby('Sex')['Survived'].value_counts()
```

```
Out[122]: Sex      Survived
female  1           233
          0            81
male     0           468
          1           109
          Name: Survived, dtype: int64
```

```
In [8]: df.hist(bins=9,figsize=(10,9));
```



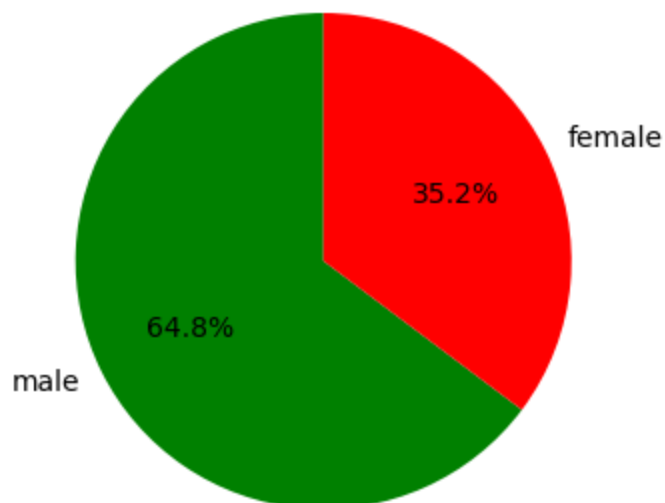
```
In [10]: df['Age'].hist(bins=20);  
plt.title('Distribution of passenger ages')  
plt.xlabel('Ages')  
plt.show()
```



```
In [60]: import matplotlib.pyplot as plt
```

```
In [20]: gender_counts = df['Sex'].value_counts()
plt.figure(figsize=(4, 5))
plt.pie(gender_counts, labels=gender_counts.index, autopct='%1.1f%%', startangle=90, c
plt.title('Percentage of Male and Female Passengers')
plt.show()
```

Percentage of Male and Female Passengers



```
In [85]: # Group the data by 'Pclass' and calculate the survival count for each class
survival_counts = df.groupby(['Pclass', 'Survived']).size().unstack(fill_value=0)

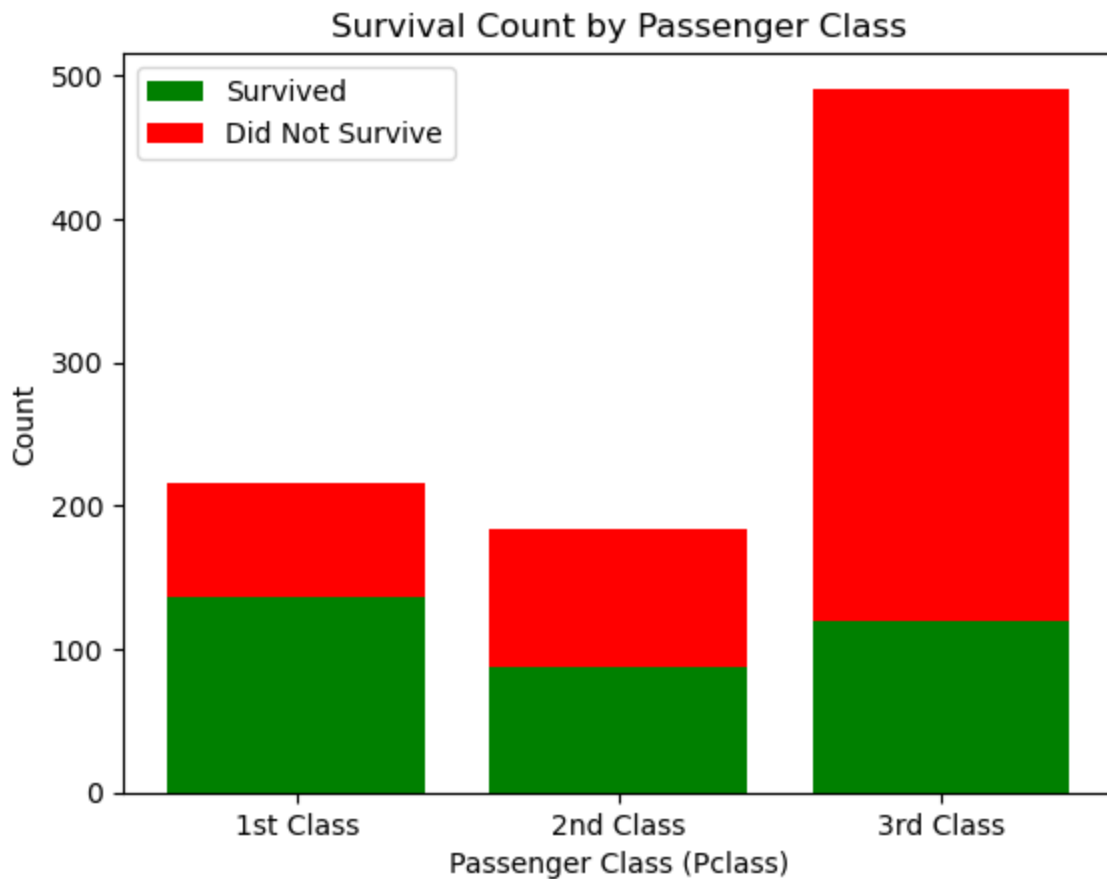
# Create a bar chart for passengers who survived (Survived=1)
plt.bar(survival_counts.index, survival_counts[1], color='green', label='Survived')

# Create a bar chart for passengers who did not survive (Survived=0)
plt.bar(survival_counts.index, survival_counts[0], color='red', bottom=survival_counts[1])

plt.xlabel('Passenger Class (Pclass)')
plt.ylabel('Count')
plt.title('Survival Count by Passenger Class')
plt.xticks(survival_counts.index, labels=['1st Class', '2nd Class', '3rd Class'])

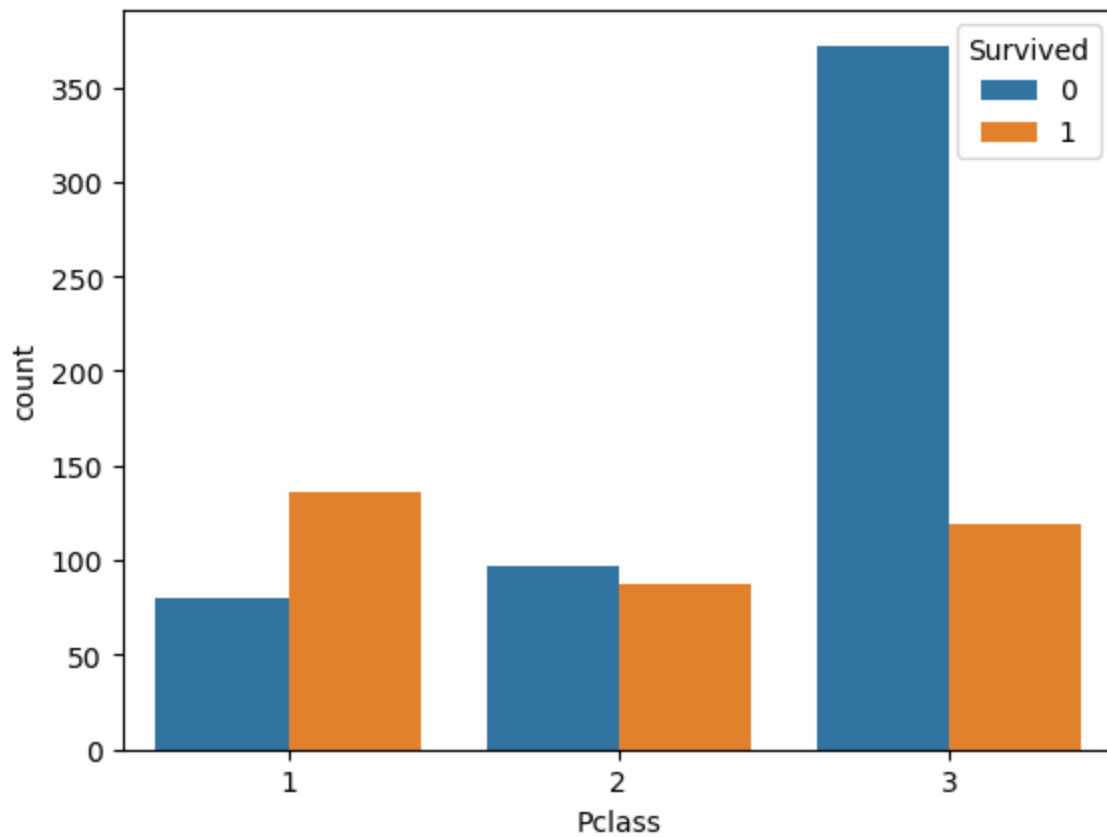
# Add a Legend
plt.legend()

# Show the plot
plt.show()
```



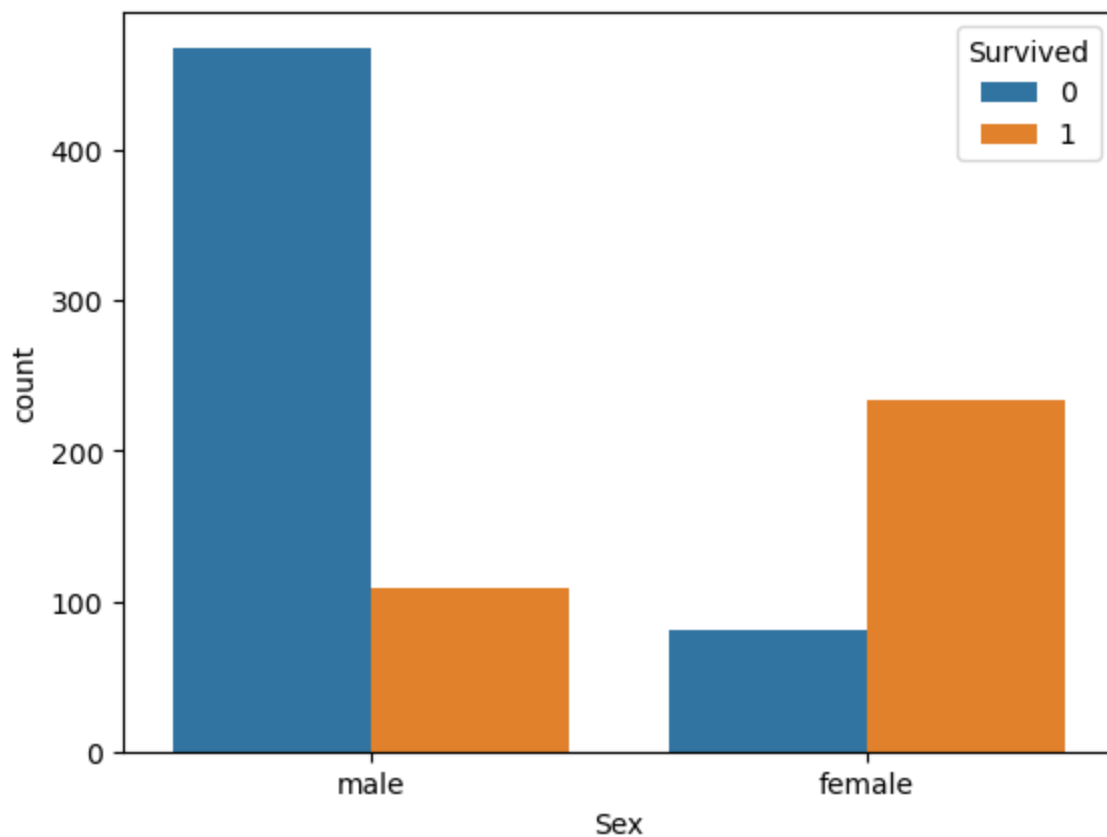
```
In [88]: sns.countplot(data=df, x='Pclass', hue='Survived')
```

```
Out[88]: <AxesSubplot:xlabel='Pclass', ylabel='count'>
```



```
In [89]: sns.countplot(data=df, x='Sex', hue='Survived')
```

```
Out[89]: <AxesSubplot:xlabel='Sex', ylabel='count'>
```



```
In [90]: sns.countplot(data=df,x='Survived')
```

```
Out[90]: <AxesSubplot:xlabel='Survived', ylabel='count'>
```

