



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

```
df = pd.read_csv("/content/titanic_submission (1).csv")
```

```
df.head()
```




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

Next steps:

[Generate code with df](#)
[View recommended plots](#)

```
df1=pd.read_csv("/content/train.csv")
```

```
df1.head()
```




	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	71.2834

Next steps:


[Generate code with df1](#)
[View recommended plots](#)

```
df1.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
```

```
df1.columns
```




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

```
df1.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	49.693428
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


```
df1.shape
```



```
(891, 12)
```

Handling missing values

```
df1.isnull().sum()
```




```
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
```

Drop Unnecessary Columns

```
df=df1.drop(['PassengerId', 'Name', 'Ticket', 'Cabin'], axis=1)
```

```
df.head()
```



	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
0	0	3	male	22.0	1	0	7.2500	S
1	1	1	female	38.0	1	0	71.2833	C
2	1	3	female	26.0	0	0	7.9250	S
3	1	1	female	35.0	1	0	53.1000	S
4	0	3	male	35.0	0	0	8.0500	S

Next steps:

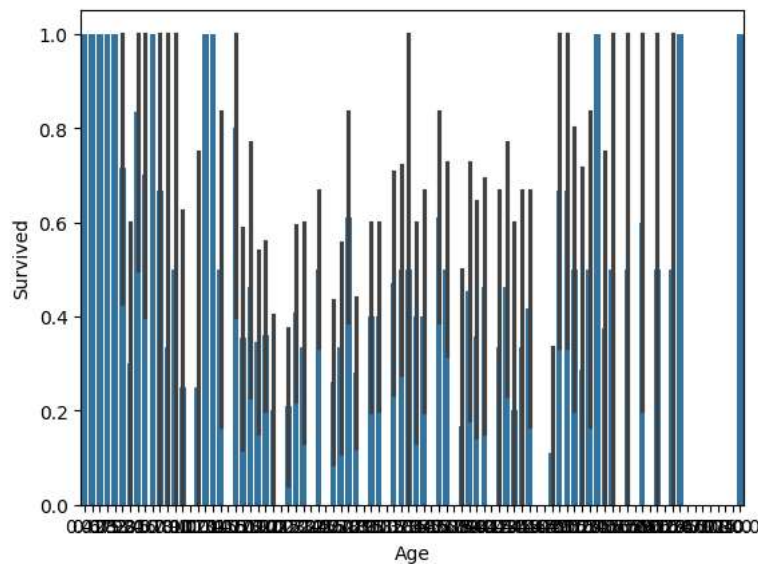
Generate code with df

 View recommended plots

Data Visualization

```
#Survival Rate vs. Age
import seaborn as sns
sns.barplot(x='Age', y='Survived', data=df)
```

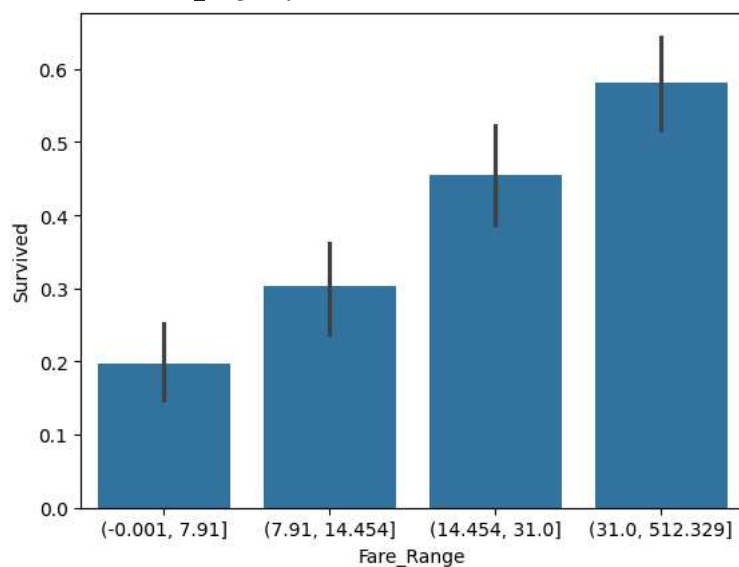
<Axes: xlabel='Age', ylabel='Survived'>



#Survival Rate vs. Fare:

```
df['Fare_Range'] = pd.qcut(df['Fare'], 4)
sns.barplot(x='Fare_Range', y='Survived', data=df)
```

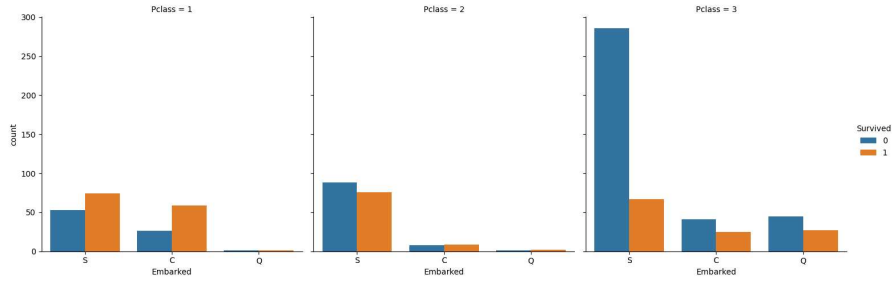
<Axes: xlabel='Fare\_Range', ylabel='Survived'>



#Survival Rate vs. Embarked:

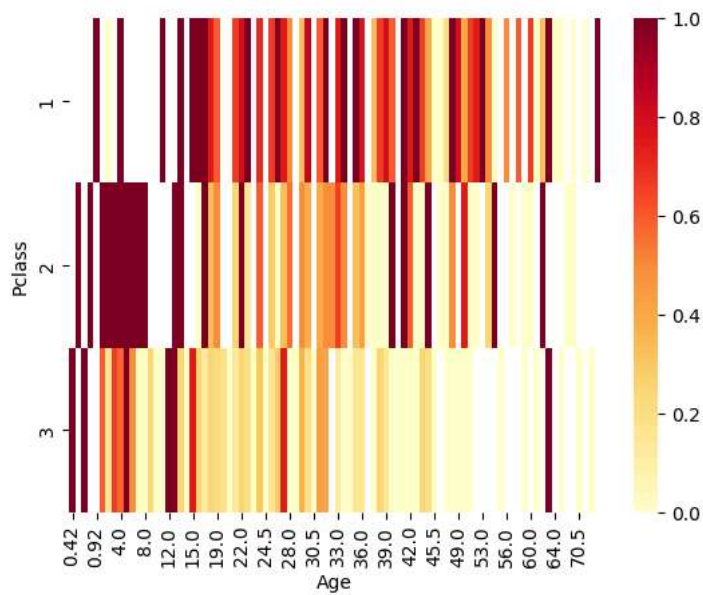
```
sns.catplot(x='Embarked', hue='Survived', kind='count', col='Pclass', data=df)
```

```
<seaborn.axisgrid.FacetGrid at 0x78a82dcf5090>
```

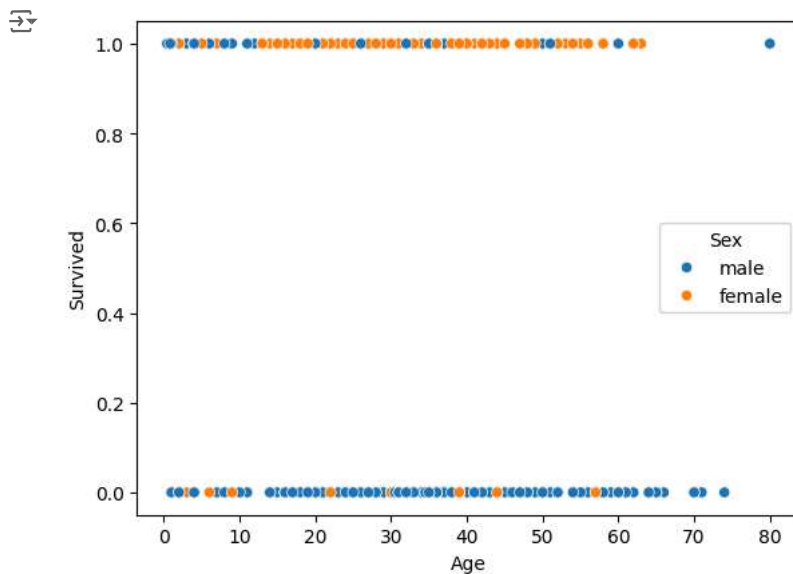


```
sns.heatmap(df.pivot_table(index='Pclass', columns='Age', values='Survived', aggfunc='mean'), cmap='YlOrRd')
```

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

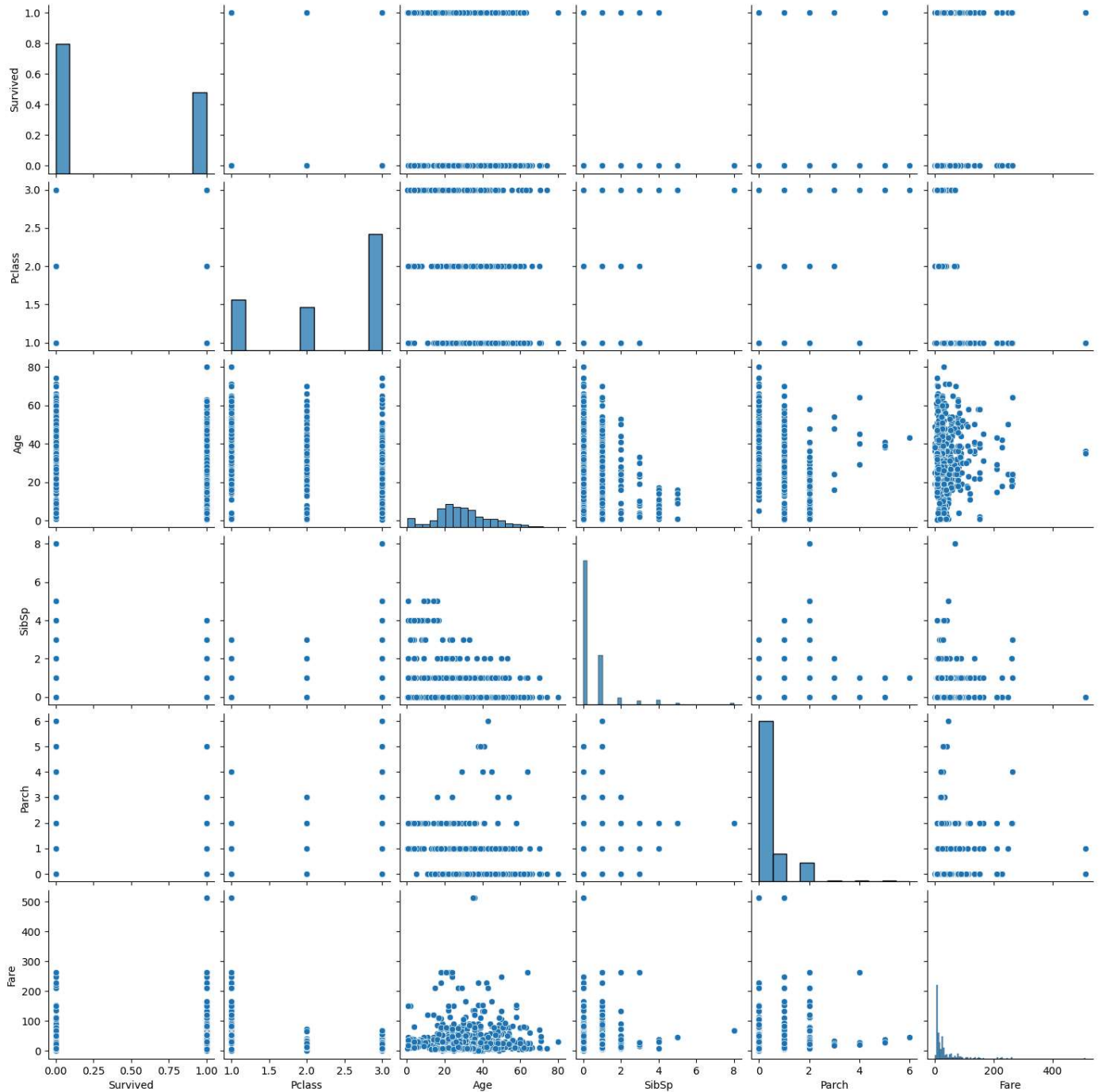


```
sns.scatterplot(x='Age', y='Survived', hue='Sex', data=df)
plt.show()
```



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

```
<seaborn.axisgrid.PairGrid at 0x78a82d4c50c0>
```



### Correlation Analysis

```
numerical_df = df.select_dtypes(include=['number'])
correlation_matrix = numerical_df.corr()
print(correlation_matrix)
```



	Survived	Pclass	Age	SibSp	Parch	Fare
Survived	1.000000	-0.338481	-0.077221	-0.035322	0.081629	0.257307
Pclass	-0.338481	1.000000	-0.369226	0.083081	0.018443	-0.549500
Age	-0.077221	-0.369226	1.000000	-0.308247	-0.189119	0.096067
SibSp	-0.035322	0.083081	-0.308247	1.000000	0.414838	0.159651
Parch	0.081629	0.018443	-0.189119	0.414838	1.000000	0.216225
Fare	0.257307	-0.549500	0.096067	0.159651	0.216225	1.000000

```
fig, ax = plt.subplots(figsize=(12, 9))  
sns.heatmap(correlation_matrix, annot=True, square=True)  
plt.title("Correlation Heatmap", size=16)  
plt.show()
```

