

# eda-titanic

April 30, 2025

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

sns.set(style="whitegrid")
df=pd.read_csv("C:\\Users\\shaik\\Downloads\\tested.csv")
df.head()
```

```
[2]:
```

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

	Name	Sex	Age	SibSp	Parch	\
0	Kelly, Mr. James	male	34.5	0	0	
1	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	
2	Myles, Mr. Thomas Francis	male	62.0	0	0	
3	Wirz, Mr. Albert	male	27.0	0	0	
4	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	

	Ticket	Fare	Cabin	Embarked
0	330911	7.8292	NaN	Q
1	363272	7.0000	NaN	S
2	240276	9.6875	NaN	Q
3	315154	8.6625	NaN	S
4	3101298	12.2875	NaN	S

```
[4]: #Understanding the Data
df.info()
df.describe()
df.isnull().sum()
```

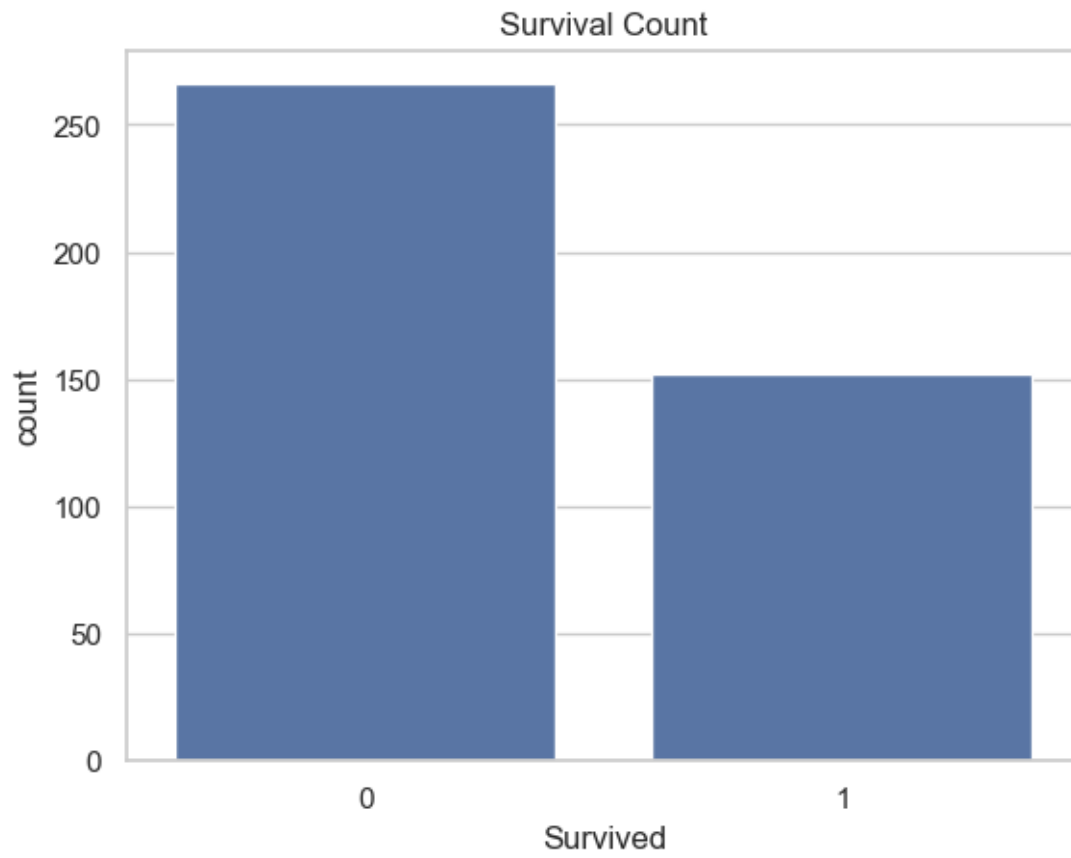
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 12 columns):
```

#	Column	Non-Null Count	Dtype
0	PassengerId	418 non-null	int64
1	Survived	418 non-null	int64
2	Pclass	418 non-null	int64
3	Name	418 non-null	object
4	Sex	418 non-null	object
5	Age	332 non-null	float64
6	SibSp	418 non-null	int64
7	Parch	418 non-null	int64
8	Ticket	418 non-null	object
9	Fare	417 non-null	float64
10	Cabin	91 non-null	object
11	Embarked	418 non-null	object

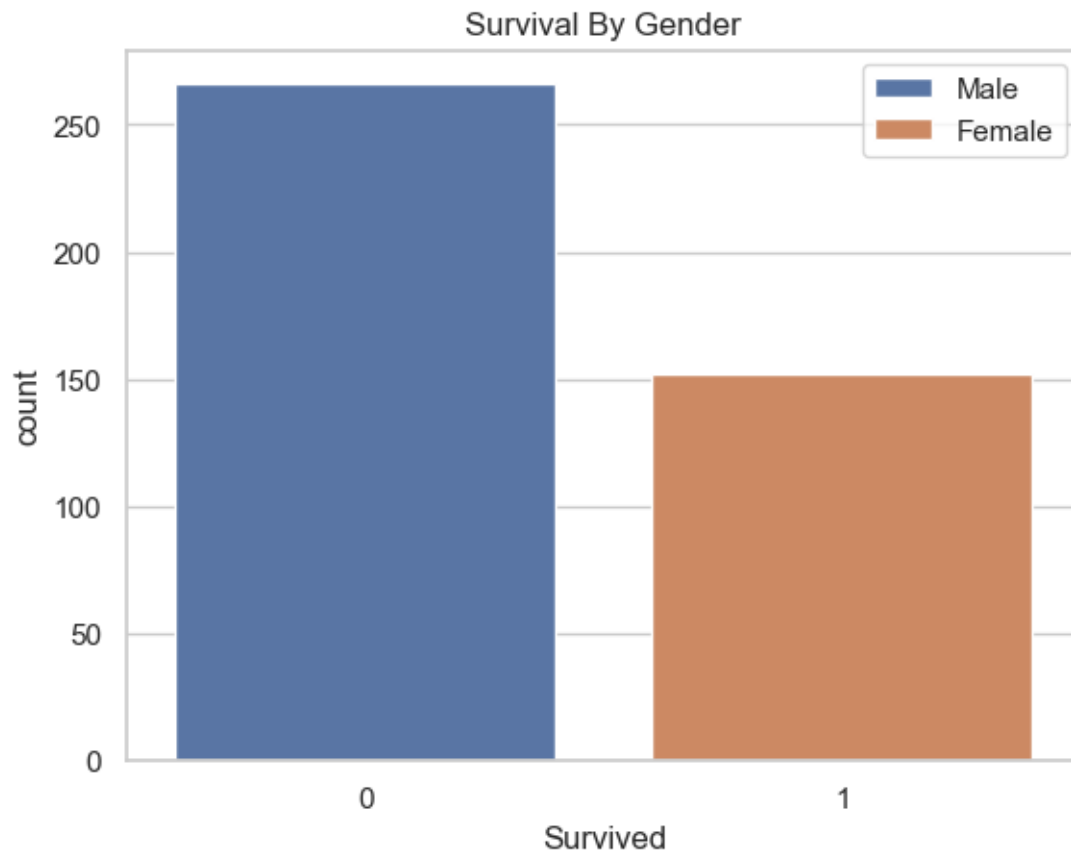
dtypes: float64(2), int64(5), object(5)  
memory usage: 39.3+ KB

```
[4]: PassengerId      0
      Survived        0
      Pclass          0
      Name            0
      Sex             0
      Age            86
      SibSp           0
      Parch           0
      Ticket          0
      Fare            1
      Cabin          327
      Embarked        0
      dtype: int64
```

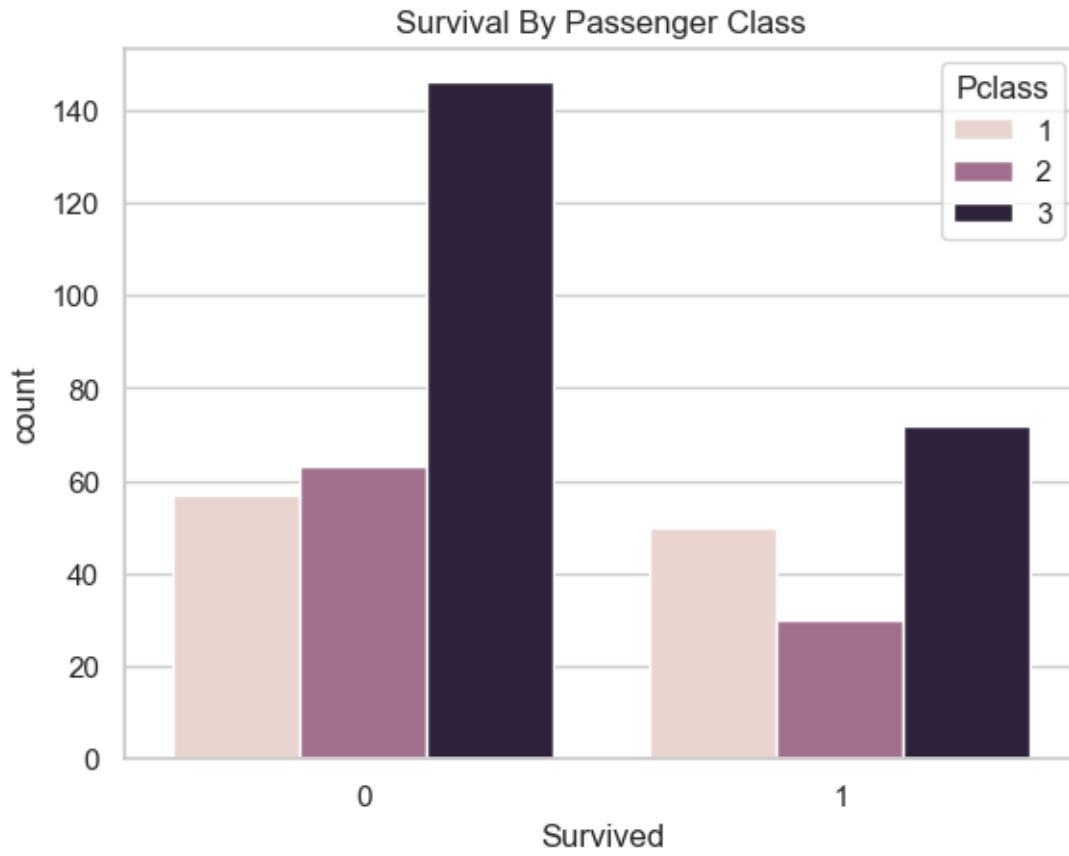
```
[8]: #Exploratory Data Analysis(EDA)
      #Survival Count
      sns.countplot(x='Survived',data=df)
      plt.title('Survival Count')
      plt.show()
```



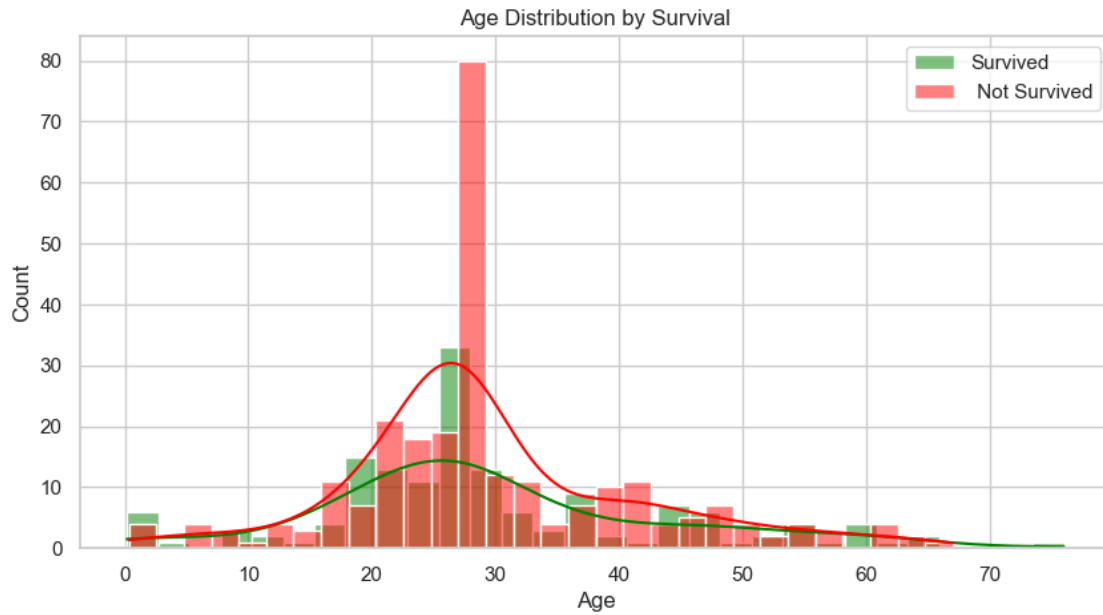
```
[9]: #Survival by gender  
sns.countplot(x='Survived', hue='Sex', data=df)  
plt.title('Survival By Gender')  
plt.legend(labels=['Male', 'Female'])  
plt.show()
```



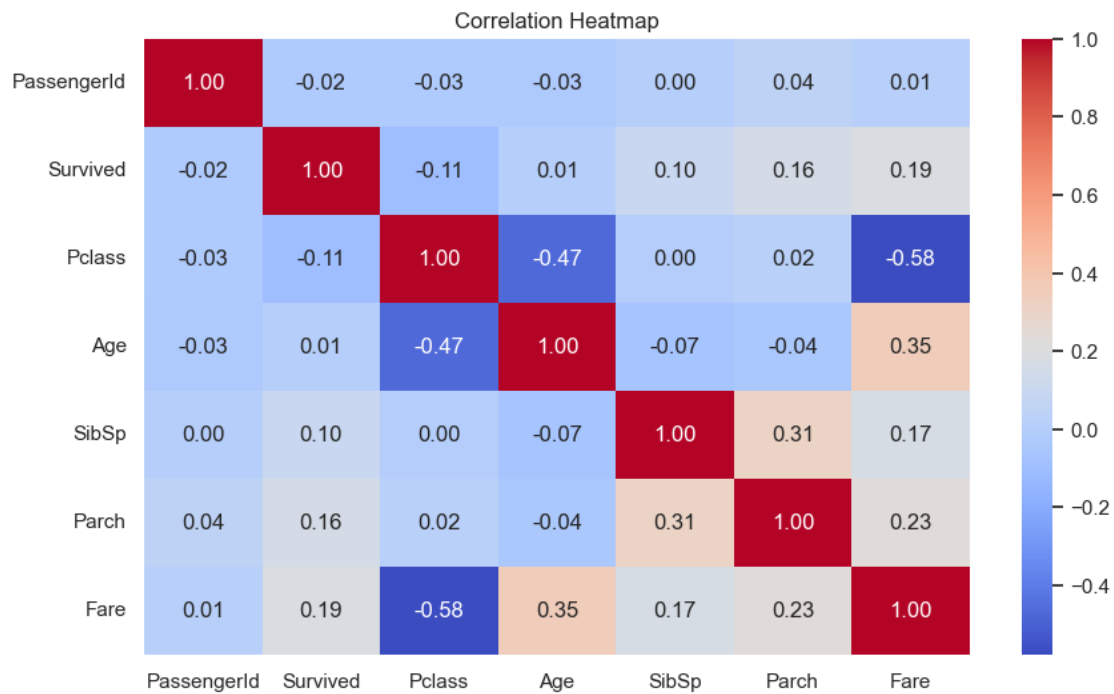
```
[11]: #Survival by passenger class
sns.countplot(x='Survived',hue='Pclass', data=df)
plt.title('Survival By Passenger Class')
plt.show()
```



```
[12]: #AGE Distribution and Survival
plt.figure(figsize=(10,5))
sns.
    ↳histplot(df[df['Survived']==1]['Age'],bins=30,color='green',label='Survived',
    ↳kde=True)
sns.histplot(df[df['Survived']==0]['Age'],bins=30,color='red',label=' Not
    ↳Survived', kde=True)
plt.legend()
plt.title('Age Distribution by Survival')
plt.show()
```



```
[14]: numeric_df=df.select_dtypes(include=[np.number])
plt.figure(figsize=(10,6))
sns.heatmap(numeric_df.corr(),annot=True,cmap='coolwarm',fmt=".2f")
plt.title('Correlation Heatmap')
plt.show()
```



[ ]: