PRODIGY INFOTECH - TASK 02

Exploratory Data Analysis (EDA) on Titanic Dataset

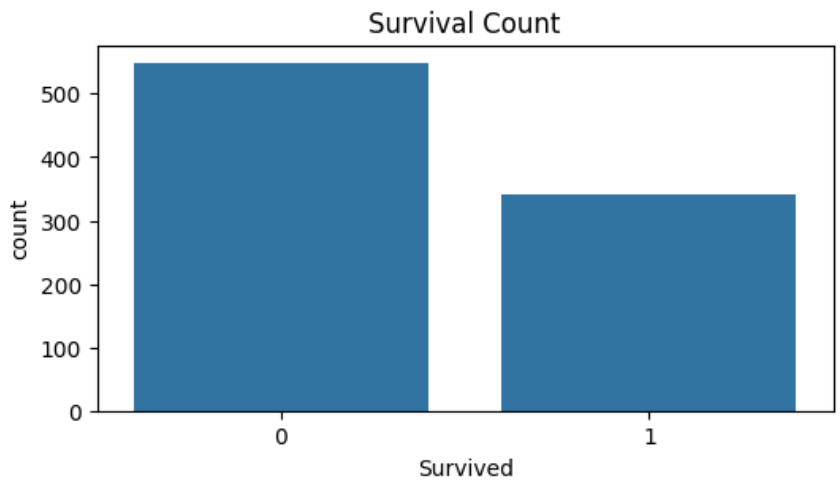
Intern Name: Akash Wasokar

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

df = pd.read_csv("https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv")

df['Age'] = df['Age'].fillna(df['Age'].median())
df['Embarked'] = df['Embarked'].fillna(df['Embarked'].mode()[0])

print(df.head())


print("\nMissing Values Count:\n")
print(df.isnull().sum())
```

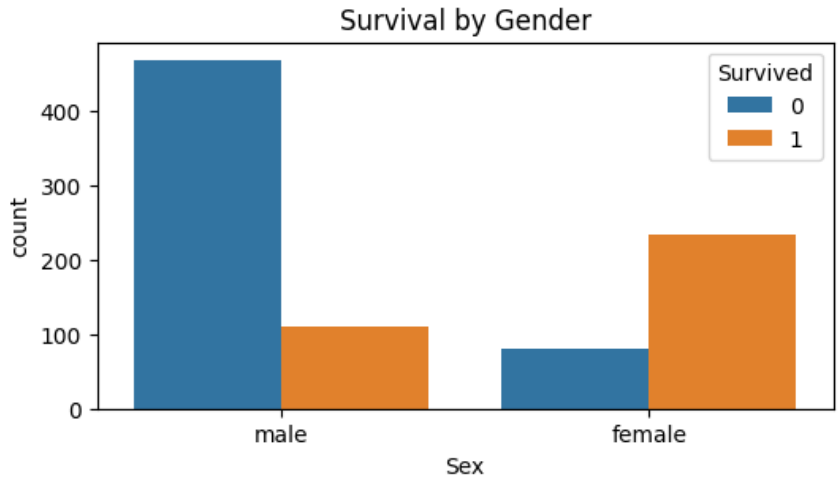
```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3

                                                Name     Sex   Age  SibSp  \
0                            Braund, Mr. Owen Harris    male  22.0      1
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
4                           Allen, Mr. William Henry    male  35.0      0

   Parch            Ticket     Fare Cabin Embarked
0      0         A/5 21171   7.2500   NaN        S
1      0          PC 17599  71.2833   C85        C
2      0  STON/O2. 3101282   7.9250   NaN        S
3      0            113803  53.1000  C123        S
4      0            373450   8.0500   NaN        S

Missing Values Count:

PassengerId      0
Survived         0
```

```
Pclass          0
Name            0
Sex             0
Age             0
SibSp           0
Parch           0
Ticket          0
Fare            0
Cabin         687
Embarked        0
dtype: int64
```

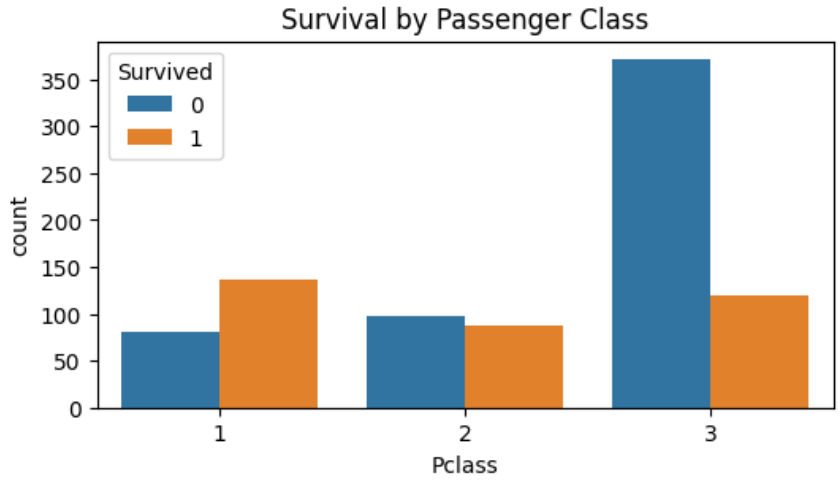
```
plt.figure(figsize=(6, 3))
sns.countplot(x='Survived', data=df)
plt.title("Survival Count")
plt.show()
```




```
plt.figure(figsize=(6, 3))
sns.countplot(x='Sex', hue='Survived', data=df)
plt.title("Survival by Gender")
plt.show()
```




```
plt.figure(figsize=(6, 3))
sns.countplot(x='Pclass', hue='Survived', data=df)
plt.title("Survival by Passenger Class")
plt.show()
```



Conclusion

1. Female passengers had a significantly higher survival rate than male passengers.

2. Passengers traveling in 1st class had better chances of survival.

3. Younger passengers showed higher survival probability.

4. Gender, age and class were major factors influencing survival.

Note: This task helped me understand how data cleaning and EDA is performed on real-world datasets. It improved my understanding of handling missing values and visualization.