Exploratory Data Analysis on Titanic Dataset

Introduction

This report presents an exploratory data analysis (EDA) of the Titanic dataset. The primary goal is to clean the data, explore relationships between variables, identify patterns and trends, and visualize key findings in a human-readable manner.

Data Cleaning

- Missing values in 'Age' were filled with median age.

- Missing values in 'Embarked' were filled with the most common port.

- 'Cabin' column was dropped due to excessive missing data.

Python Code Used

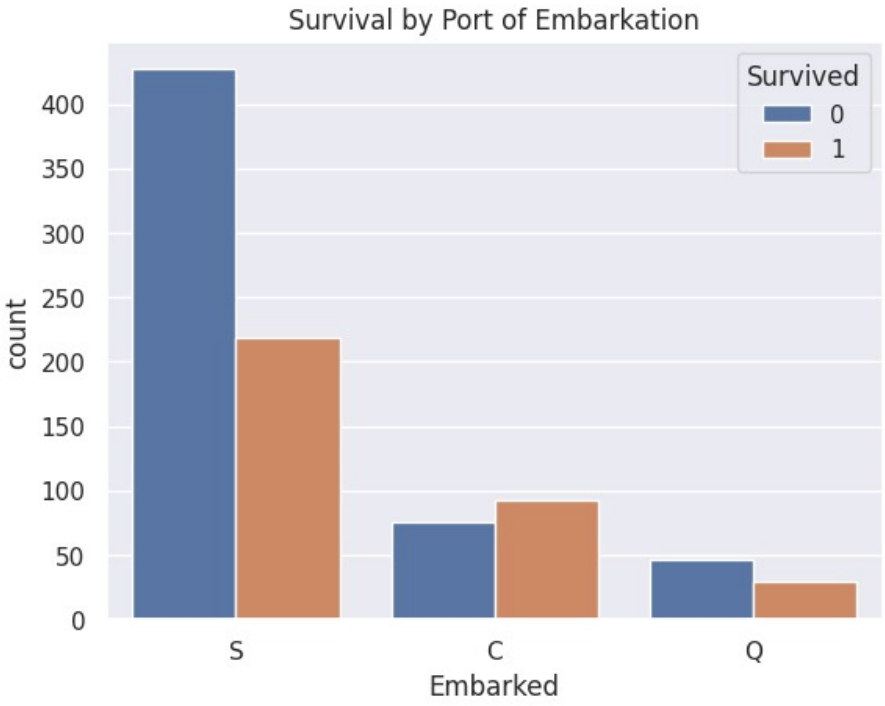
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv('Titanic-Dataset.csv')
df['Age'].fillna(df['Age'].median(), inplace=True)
df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)
if 'Cabin' in df.columns:
    df.drop(columns=['Cabin'], inplace=True)
```

Visualizations and Observations

1. Survival by Gender:

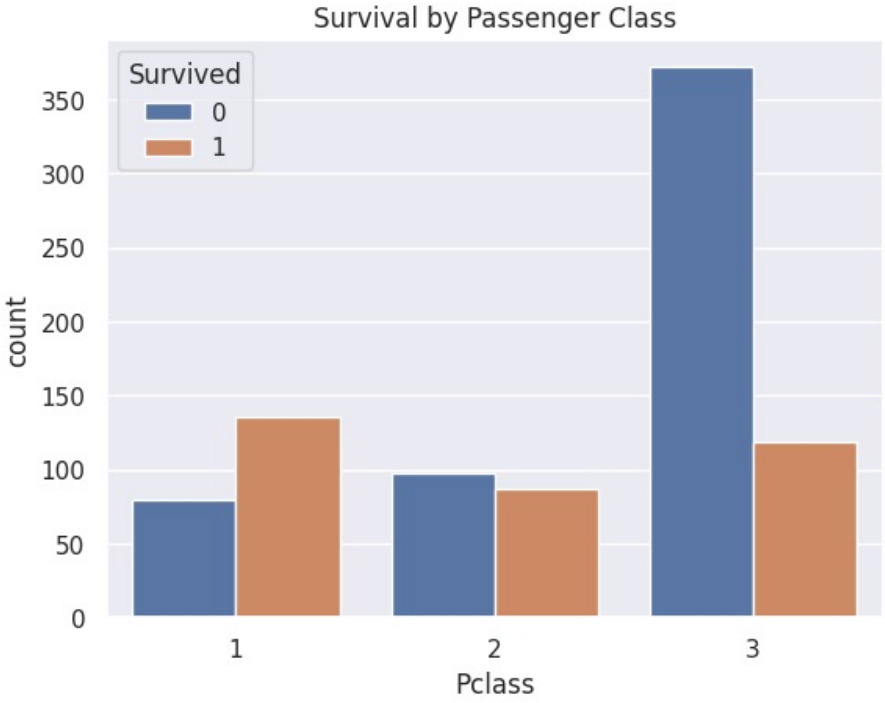
- Females had a significantly higher survival rate than males.

Exploratory Data Analysis on Titanic Dataset



2. Survival by Passenger Class:

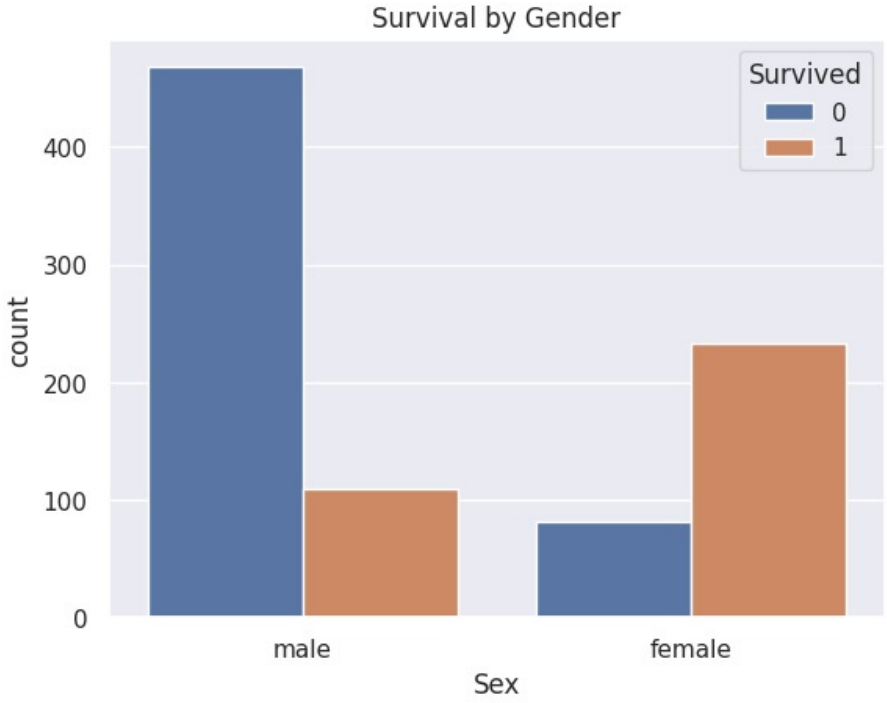
- 1st class passengers had the highest survival rate, followed by 2nd and 3rd class.



3. Survival by Port of Embarkation:

Exploratory Data Analysis on Titanic Dataset

- Passengers who embarked from Cherbourg (C) had higher survival rates.



4. Passenger Class Distribution:

- Most passengers were in 3rd class, followed by 1st and 2nd class.

Exploratory Data Analysis on Titanic Dataset

Conclusion

From the analysis, it is evident that:

- Gender played a major role in survival chances.

- Higher class passengers had better chances of survival.

- Embarkation port was also a subtle indicator of survival patterns.

This EDA provides a foundational understanding of how different factors influenced survival on the Titanic.