**Introduction**

This report presents the findings from an exploratory data analysis (EDA) conducted on the Titanic dataset. The objective of this analysis is to uncover insights and patterns in the data that can be understood by a non-technical business audience. The dataset includes information about passengers aboard the Titanic, such as their survival status, class, gender, age, and embarkation point.

**Common Values for Each Variable**

To begin with, we examined the most common values for each variable in the dataset.

**Survival Status**

The majority of the passengers did not survive, with over 500 passengers deceased and around 300 passengers surviving.

**Interpretation:** This indicates that the majority of passengers aboard the Titanic did not survive the disaster. For a business audience, this highlights the severe impact of the event on passenger survival rates.

**Passenger Class**

Most passengers were in the third class, followed by the second class, and the fewest were in the first class.

**Interpretation:** The third class had the highest number of passengers, which could imply a lower cost of tickets in this class. This distribution shows the socio-economic diversity of the passengers.

**Gender**

There were more male passengers compared to female passengers.

**Interpretation:** The gender distribution shows a higher number of male passengers, which could impact survival rates and rescue operations during the disaster.

**Embarked Location**

Most passengers embarked from Southampton, followed by Cherbourg and Queenstown.

**Interpretation:** The majority of passengers boarding from Southampton indicates it was a major port of departure, reflecting the travel patterns of that era.

**Covariation Between Variables**

Next, we explored the relationships between different variables to understand how they interact and influence each other.

**Survival by Class**

Survival rates were higher for first-class passengers compared to second and third-class passengers.

Interpretation: Higher survival rates in the first class suggest that passengers in this class had better access to lifeboats and safety measures. This demonstrates the influence of socio-economic status on survival chances.

**Survival by Gender**

Females had a higher survival rate compared to males.

Interpretation: The higher survival rate among female passengers suggests that women and children were given priority during rescue operations. This aligns with historical accounts of the disaster.

**Survival by Age**

Younger passengers, particularly children, had higher survival rates.

**Interpretation**: The survival rate for younger passengers, especially those below 10 years old, was higher. This indicates that children were prioritized during rescue efforts, reflecting the "women and children first" protocol.

**Conclusion**

Based on the analysis, we can draw several key conclusions:

Survival Rates: Survival rates varied significantly based on class, gender, and age, with first-class passengers, females, and younger passengers having higher survival rates.

Embarkation Point: The majority of passengers boarded from Southampton, which influenced the demographic composition of the passengers.

These findings are not surprising given historical accounts of the Titanic disaster, which document the prioritization of women and children for lifeboat seats and the better conditions afforded to first-class passengers.

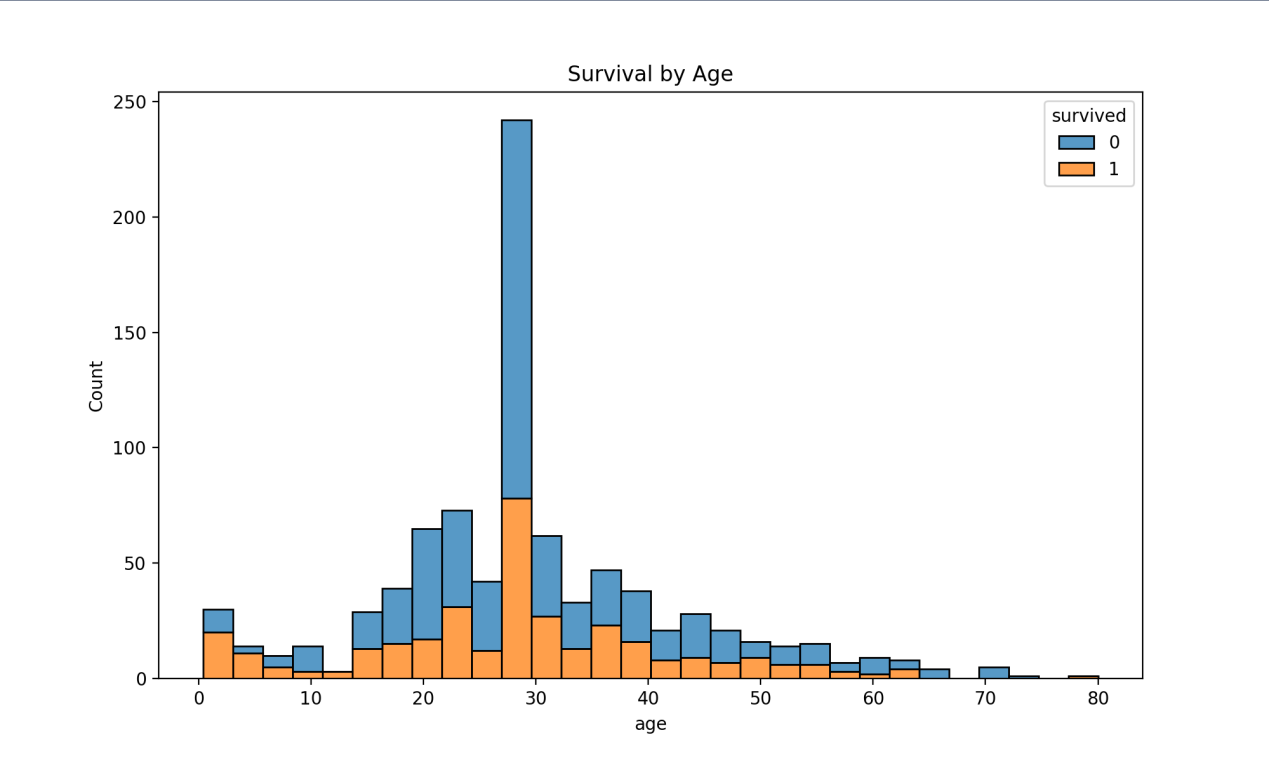
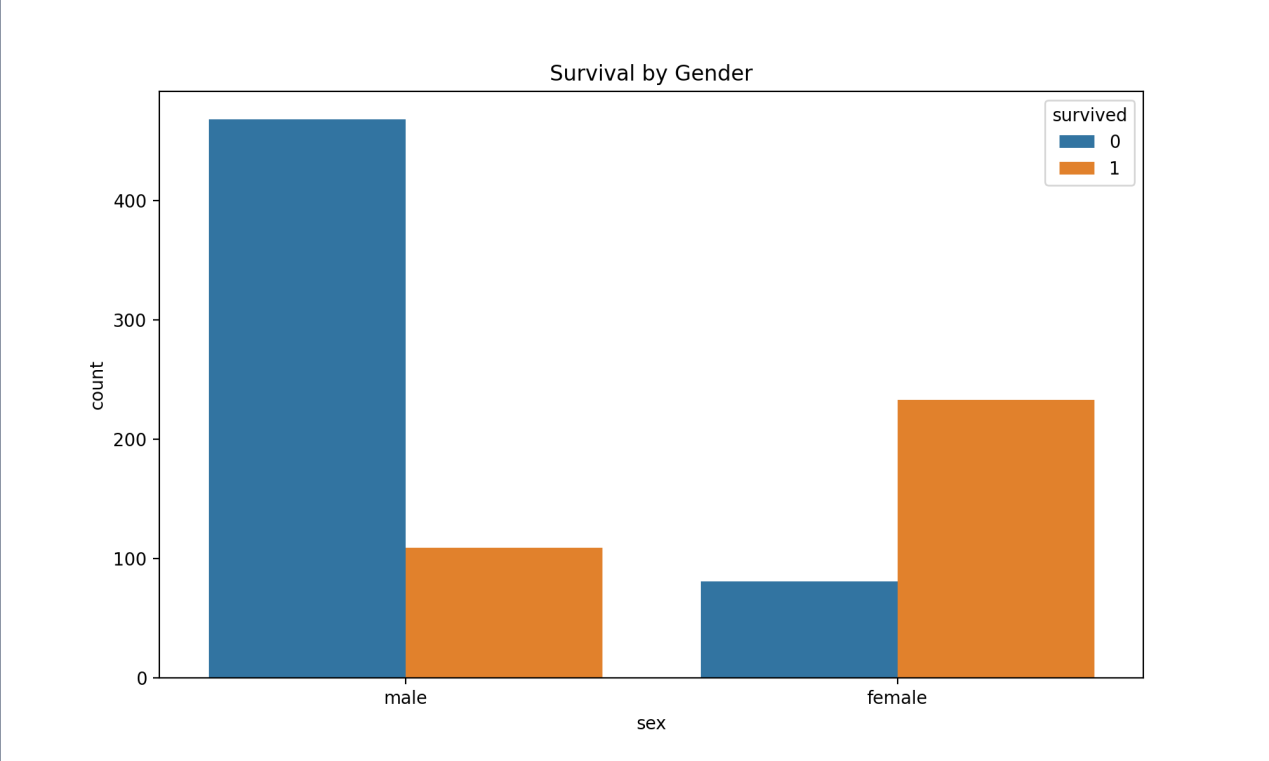
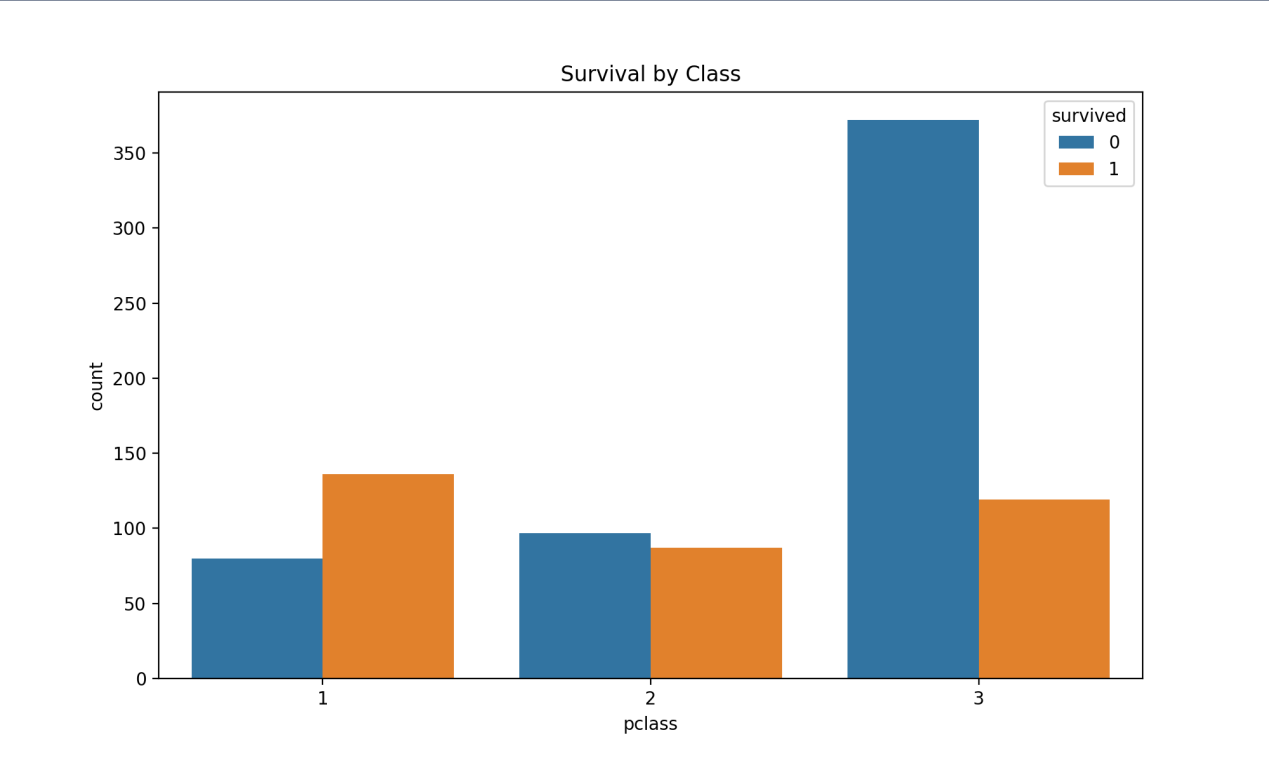
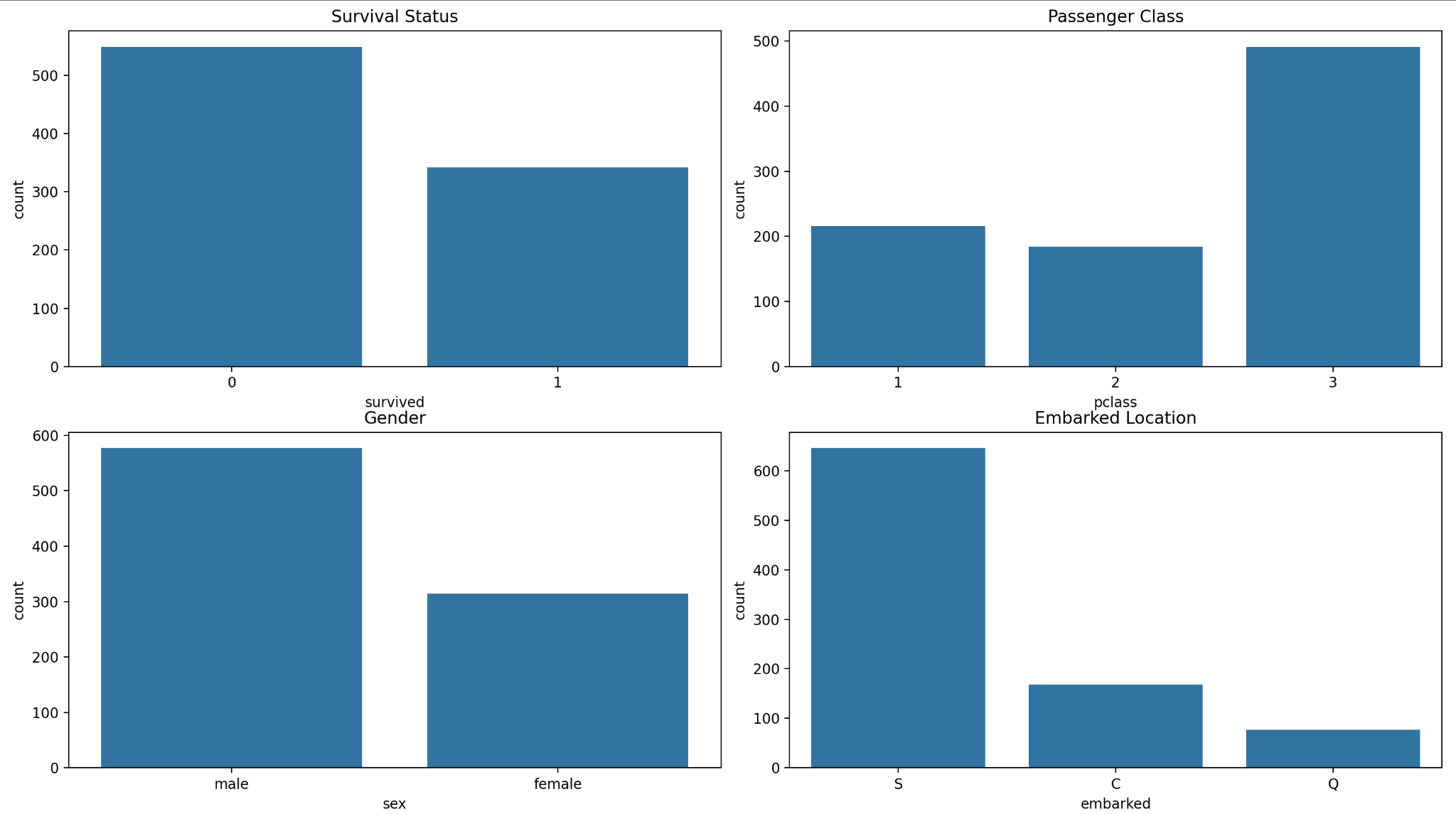
**Predictive Analysis Application**

The insights from this EDA can be used to create predictive models to estimate survival probabilities based on passenger attributes. For example:

Passenger Class: The class of the passenger can significantly impact the survival probability.

Gender and Age: Female and younger passengers have higher chances of survival.

These predictive models can help in understanding and preparing for similar events in the future, enhancing safety protocols and prioritizing rescue operations effectively.



I finished it by Python, the code is following:

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Load Titanic dataset from seaborn library

titanic = sns.load\_dataset('titanic')

# Display the first few rows of the dataset

titanic.head()

# Check for missing values

titanic.isnull().sum()

# Fill missing age values with the median age

titanic['age'].fillna(titanic['age'].median(), inplace=True)

# Fill missing embarked values with the mode

titanic['embarked'].fillna(titanic['embarked'].mode()[0], inplace=True)

# Drop rows with missing values in 'deck' and 'embark\_town'

titanic.drop(columns=['deck', 'embark\_town'], inplace=True)

# Fill missing 'embarked' values with mode

titanic['embarked'].fillna(titanic['embarked'].mode()[0], inplace=True)

# Display the cleaned dataset

titanic.isnull().sum()

# Plot common values for each variable

fig, axs = plt.subplots(2, 2, figsize=(15, 10))

# Plot survival

sns.countplot(data=titanic, x='survived', ax=axs[0, 0])

axs[0, 0].set\_title('Survival Status')

# Plot class

sns.countplot(data=titanic, x='pclass', ax=axs[0, 1])

axs[0, 1].set\_title('Passenger Class')

# Plot gender

sns.countplot(data=titanic, x='sex', ax=axs[1, 0])

axs[1, 0].set\_title('Gender')

# Plot embarked

sns.countplot(data=titanic, x='embarked', ax=axs[1, 1])

axs[1, 1].set\_title('Embarked Location')

plt.tight\_layout()

plt.show()

# Plot survival by class

plt.figure(figsize=(10, 6))

sns.countplot(data=titanic, x='pclass', hue='survived')

plt.title('Survival by Class')

plt.show()

# Plot survival by gender

plt.figure(figsize=(10, 6))

sns.countplot(data=titanic, x='sex', hue='survived')

plt.title('Survival by Gender')

plt.show()

# Plot survival by age

plt.figure(figsize=(10, 6))

sns.histplot(data=titanic, x='age', hue='survived', multiple='stack', kde=False)

plt.title('Survival by Age')

plt.show()

# Create age groups

titanic['age\_group'] = pd.cut(titanic['age'], bins=[0, 10, 20, 30, 40, 50, 60, 70, 80], labels=['0-10', '10-20', '20-30', '30-40', '40-50', '50-60', '60-70', '70-80'])

# Plot survival by age group and class

plt.figure(figsize=(12, 8))

sns.countplot(data=titanic, x='age\_group', hue='survived', palette='pastel', col='pclass')

plt.title('Survival by Age Group and Class')

plt.show()

# Plot survival by embarkation point and class

plt.figure(figsize=(12, 8))

sns.countplot(data=titanic, x='embarked', hue='survived', palette='pastel', col='pclass')

plt.title('Survival by Embarkation Point and Class')

plt.show()