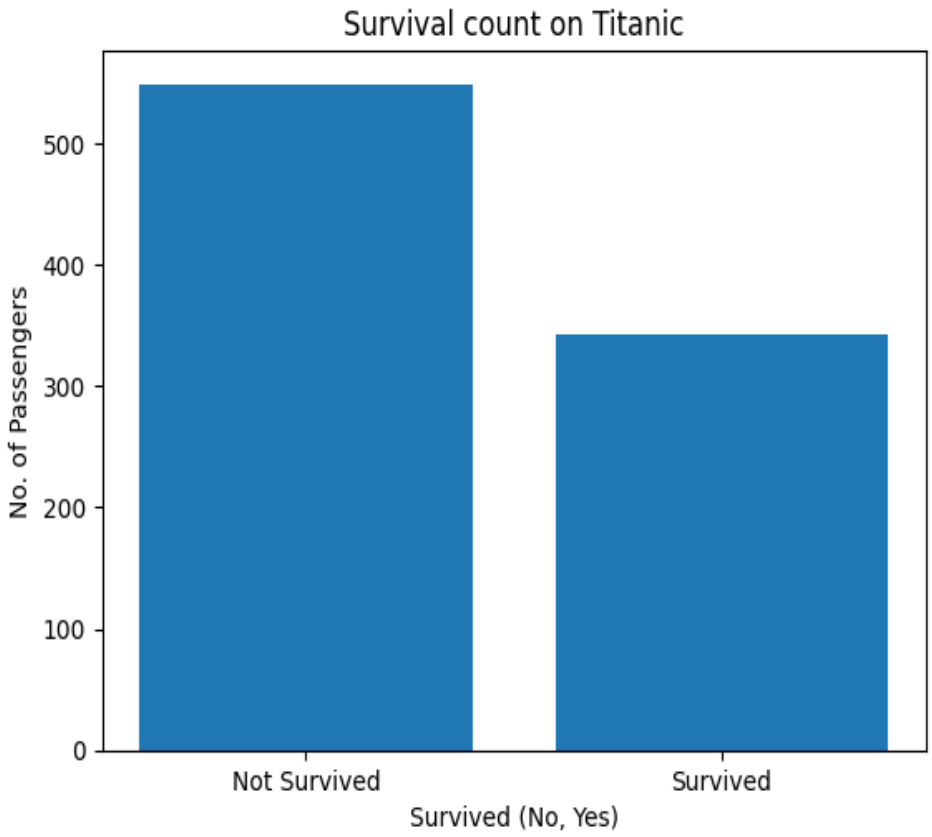
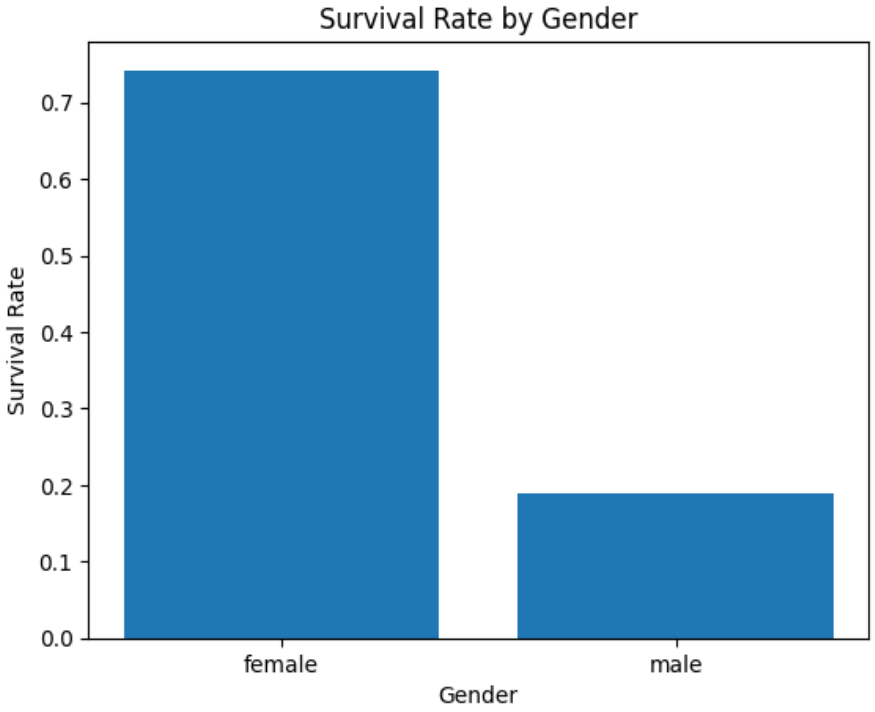
Prodigy InfoTech
Name: Shruti Sirsat.
TASK 2 : Data Cleaning & EDA

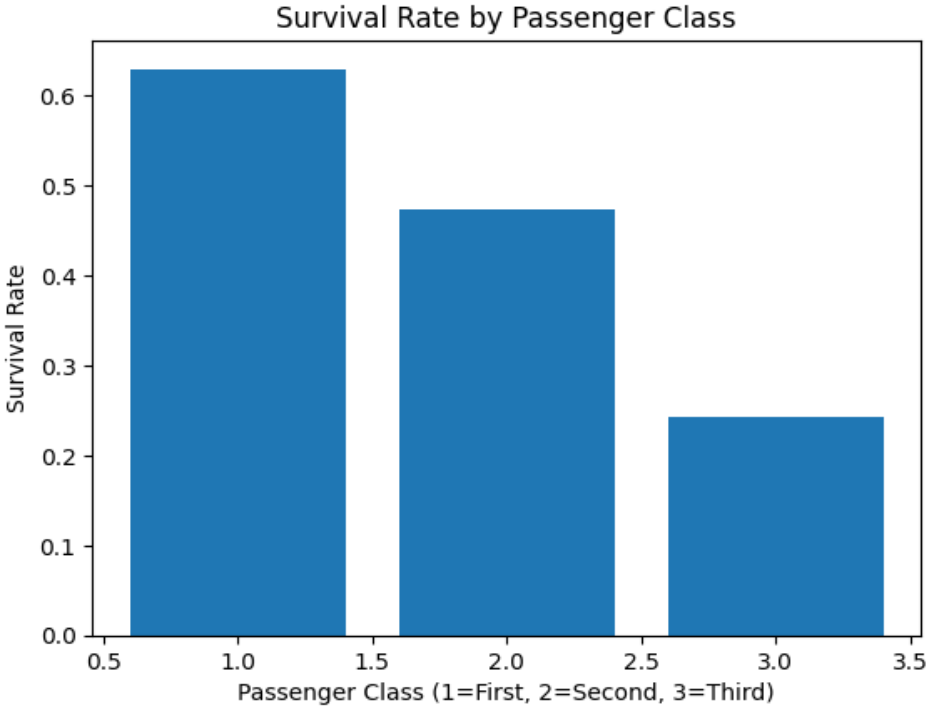
❖ Survival Counts on Titanic



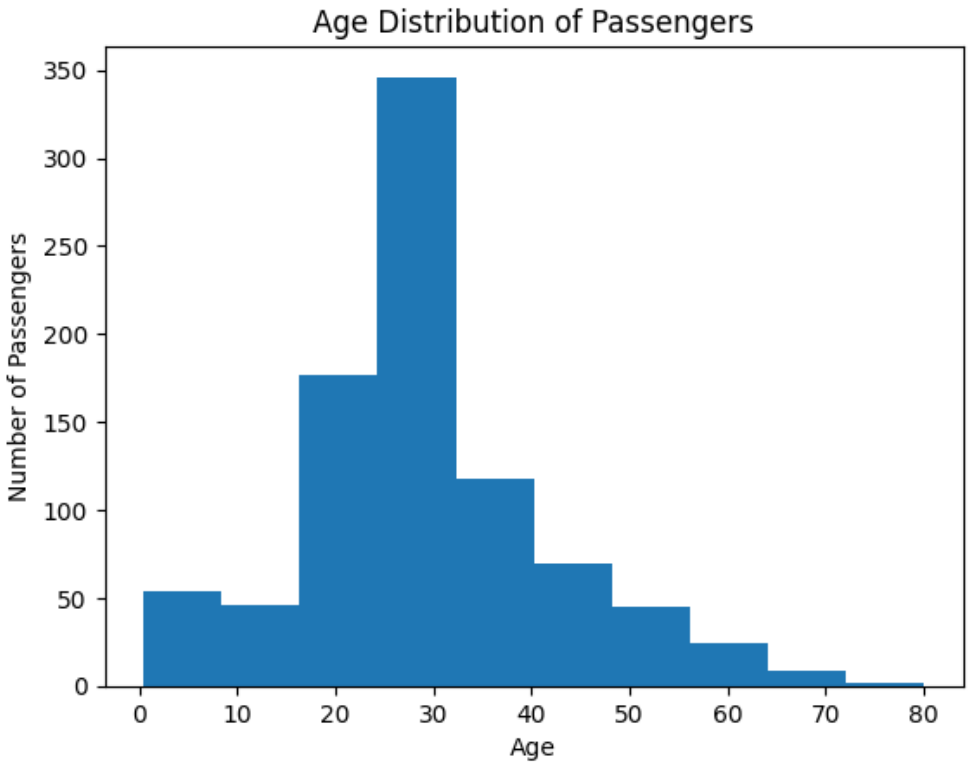
❖ Survival rate by Gender



❖ Survival rate by Passenger Class



❖ Age Distribution of Passengers



1. Dataset Description

The Titanic dataset was used for this task. It contains information about passengers such as age, gender, passenger class, fare, and survival status. The dataset helps analyze patterns and factors that affected survival during the Titanic disaster.

◇ 2. Data Cleaning

The following data cleaning steps were performed:

- Missing values in the **Age** column were filled using the **median age**.

- Missing values in the **Embarked** column were filled using the **mode**.
- The **Cabin** column was removed due to a large number of missing values.
- Irrelevant columns such as **PassengerId, Name, and Ticket** were removed to reduce noise.

These steps ensured the dataset was clean and suitable for analysis.

◇ 3. Exploratory Data Analysis (EDA)

📊 Survival Count

The survival count bar chart showed that **more passengers did not survive than survived** the Titanic disaster.

📊 Survival by Gender

The analysis showed that **female passengers had a significantly higher survival rate** compared to male passengers.

📊 Survival by Passenger Class

Passengers traveling in **first class had the highest survival rate**, followed by second class, while third-class passengers had the lowest survival rate.

📊 Age Distribution

The age distribution histogram indicated that **most passengers were between 20 and 40 years old**, showing that the majority of passengers were young adults.

◇ 4. Key Insights

- Gender played a major role in survival, with females having higher chances.
- Passenger class strongly influenced survival probability.

- Younger adults formed the largest group of passengers.
- Socio-economic status impacted survival outcomes.

◇ 5. Conclusion

Exploratory Data Analysis of the Titanic dataset revealed clear patterns related to survival. Factors such as gender, passenger class, and age significantly influenced survival chances. Data cleaning helped improve the quality of the dataset, making the analysis reliable and meaningful.

🔗 Code & Notebook Reference

The data cleaning and exploratory data analysis were performed using Python in Google Colab. The complete notebook containing the code and outputs can be accessed here:

Google Colab Notebook:
https://colab.research.google.com/drive/1V8yKNjpkA6u8UKSNHDk2bFEl_GTksEKY?usp=sharing