Task 1: Exploratory Data Analysis (EDA) And Visualization – Titanic Dataset

Objective:

The goal of this task is to perform exploratory data analysis on a real-world dataset (Titanic) to understand the patterns and relationships in the data, and to identify key insights using visualizations.

Step 1: Dataset Overview

We used a custom version of the Titanic dataset. It contains information about passengers, such as:

Passenger ID

Age

Fare

Sex

Siblings/Spouses Aboard (sibsp)

Parents/Children Aboard (Parch)

Passenger Class (Pclass)

Port of Embarkation (Embarked)

Survival status (Survived)

Step 2: Data Cleaning

Removed unnecessary columns (with names like zero, zero.1, etc.)

Renamed incorrect column name 2urvived to Survived

Verified that no columns had missing values

Checked for duplicates and found none

Step 3: Data Visualization and Insights

1. Survival Count

Most passengers did not survive

Fewer people survived than those who died

2. Survival by Gender

More females survived compared to males

This suggests that women were prioritized during evacuation

3. Survival by Passenger Class

1st class passengers had higher survival rates

3rd class passengers had the lowest survival rate

Socio-economic status affected chances of survival

Conclusion: From the visualizations, we observed that gender and passenger class were two major factors influencing survival on the Titanic. Females and passengers in 1st class had higher survival rates. This analysis helps understand how different features affect the outcome in a real-world tragedy.

Screenshots:











