Task 5: Exploratory Data Analysis (Titanic Dataset)

Objective

The objective of this analysis is to explore the Titanic dataset using statistical and visual methods to identify patterns, trends, and anomalies related to passenger survival.

Dataset Overview

• Number of rows and columns: 891 rows × 12 columns
• Columns include: PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked
• Missing values are mainly in Age, Cabin, and Embarked.

Univariate Analysis

1. Most passengers were between 20–40 years old, with fewer children and elderly passengers.
2. The dataset had more male passengers than female passengers.
3. Most passengers were in 3rd class, followed by 1st and 2nd.

Bivariate Analysis

1. A much higher proportion of females survived compared to males.
2. First-class passengers had the highest survival rate, while third-class had the lowest.
3. Younger passengers had a slightly higher survival probability compared to older ones.

Multivariate Analysis

1. The pairplot showed clear separation in survival rates based on Pclass, Fare, and Sex.
2. The correlation heatmap showed Fare positively correlated with survival, while Pclass was negatively correlated.
3. Family size variables (SibSp, Parch) had weaker relationships with survival.

Summary of Findings

1. Female passengers had higher survival rates than males.
2. First-class passengers were more likely to survive compared to second and third-class passengers.
3. Younger passengers, especially children, had better survival chances.
4. Fare showed a positive relationship with survival, indicating wealthier passengers had higher chances of survival.
5. Some features such as SibSp and Parch had less impact compared to class, sex, and fare.