Titanic Dataset - Exploratory Data Analysis (EDA)

1. Introduction

This report provides a detailed Exploratory Data Analysis (EDA) of the Titanic dataset. The goal is to uncover patterns, relationships, and survival factors among passengers.

2. Dataset Overview

Key features include demographics (age, sex), ticket information (fare, class, deck), travel details (embark_town), and survival outcome. The dataset has missing values in 'age', 'deck', and 'embark_town'.

3. Missing Value Handling

- Missing 'age' values were filled with the median age.
- 'Embark_town' missing values were filled with the mode (most common town).
- 'Deck' missing values were filled with 'Unknown' to retain all records.

4. Univariate Analysis

- Most passengers were aged between 20-40 years.
- The majority of passengers did not survive (more deaths than survivors).
- 3rd class had the largest number of passengers.
- There were more males than females onboard.

5. Bivariate Analysis

- Females had a significantly higher survival rate than males.
- 1st class passengers had better survival rates compared to 2nd and 3rd classes.
- Children under age 10 had higher survival rates compared to other age groups.

6. Multivariate Analysis

- Multivariate analysis combining 'pclass', 'sex', and 'survived' shows that 1st class females had the highest survival rate.
- Higher ticket fares are generally associated with higher survival chances.

Titanic Dataset - Exploratory Data Analysis (EDA)

7. Correlation Analysis

- Positive correlation observed between 'fare' and 'survived'.
- Slight correlation between 'parch' and 'sibsp' (family traveling together).
- Age shows a weak but interesting relation with survival.
- Heatmap reveals Fare and Pclass as important variables.

8. Final Summary of Insights

Survival chances on the Titanic were strongly influenced by gender, age, ticket class, and fare. Females, younger passengers, and those in 1st class had a significantly higher probability of survival. Proper missing value treatment and multivariate visualizations were key to uncovering these patterns.