Executive Summary

by - Akash Anandani

Insights from Titanic Dataset Analysis

This report summarizes the exploratory data analysis performed on the Titanic dataset as part of my internship at Prodigy Infotech. Using Jupyter Notebook, I analyzed key passenger data to uncover trends and insights related to survival. The findings highlight important factors that influenced survival rates during the Titanic disaster.

Introduction

The Titanic dataset is one of the most well-known datasets in data science, containing information about passengers aboard the ill-fated RMS Titanic. This analysis explores key features, including survival rates, passenger demographics, and other relevant variables.

- **Gender**, played a significant role in determining survival on the Titanic , Women and children, particularly those with titles like "Mrs." and "Miss.", had a higher chance of survival.

- **First-class** passengers had the **highest survival rate** , Second-class passengers had a moderate survival rate ,third-class passengers had the lowest survival rate.

- Traveling with **family** increased the **chances of survival** on the Titanic. This could be due to various factors, such as families being prioritized during evacuation or having better access to lifeboats.

- Cherbourg had the highest survival rate, followed by Southampton, with Queenstown having the lowest.

- We can observe a weak positive correlation between age and fare, suggesting older passengers tended to pay higher fares.

- **Sex** shows the **strongest correlation** with survival, with females having a significantly higher chance of survival compared to males.

- **Fare** also shows a **moderate** positive correlation, which likely reflects socio-economic factors, such as passengers in higher classes being more likely to survive.

- The Travel with Family and Parch columns suggest that passengers with family or those traveling with children/parents were somewhat more likely to survive.