Titanic Dataset - Exploratory Data Analysis (EDA)

Task: Exploratory Data Analysis (EDA)

Objective: Extract insights using visual and statistical exploration

Tools Used: Python (Pandas, Matplotlib, Seaborn)

Dataset Overview

- ➢ The dataset contains information about passengers aboard the Titanic.
- ➢ Total Records: 891
- ➢ Missing Data: Age, Cabin, Embarked
- ➢ Used .info(), .describe(), .value_counts() to summarize data.
- ➢ Categorical features like Sex, Pclass, and Embarked were analyzed for frequency.
- ➢ Histograms Showed distributions for Age, Fare, SibSp, Parch.
- ➢ Age is right-skewed; Fare has high outliers.
- ➢ Boxplots Compared Age and Fare across Survived values.
- ➢ Scatterplot (Age vs Fare) Colored by survival status. Higher fares and certain age groups had better survival chances.
- ➢ Correlation Heatmap shows Positive correlation between Fare and Survived and Negative correlation between Pclass and Survived.

Insights

- Survival Rate: Only a smaller number of passengers survived.

- Gender Impact: Females had significantly higher survival chances.

- Class Impact: 1st class passengers had the highest survival rate.

- Age Factor: Children and younger adults had better outcomes.

- Embarked Port: Most passengers embarked from Southampton.

Conclusion

The EDA provided a strong foundation for understanding survival patterns on the Titanic. Variables like class, fare, age, and gender had significant influence on survival. Visual tools proved crucial in uncovering hidden relationships.