EXPLORATORY DATA ANALYSIS(EDA) REPORT

Objective

The goal of this project is to perform Exploratory Data Analysis (EDA) on the Titanic dataset to identify patterns, relationships, and trends that may have influenced passenger survival.

Dataset Overview:

The Titanic dataset contains passenger information such as name, age, gender, class, ticket, fare, cabin, and survival status.

Using:

df.info()

df.describe()

df.isnull().sum()

df.head()

Dataset contains 200 rows and several columns.Columns like Age, Cabin, and Embarked have missing values.

Survived is the target variable (0 = No, 1 = Yes).

Univariate Analysis:

Survival Count: More passengers didn't survive than survived.

Gender: Males were more, but females had higher survival rate.

Class: 3rd Class had most passengers but lowest survival.

Age: Majority between 20–40 years; few children and elders.

Bivariate Analysis:

Gender vs Survival: Around 75% females survived; only ~20% males.

Class vs Survival: 1st Class had best survival rate.

Age vs Survival: Younger passengers fared slightly better.

Fare vs Survival: Higher fares → higher survival chance (wealthier passengers).

Correlation Analysis:

Fare positively correlated with Survived.

Pclass negatively correlated with Survived.

Encoded Sex variable strongly linked to survival outcome.

Visual Summary:

Visualizations used:

sns.countplot(), sns.histplot(), sns.boxplot(), sns.heatmap(), sns.pairplot()

They reveal clear differences between survived and non-survived groups, especially by class and gender.

Key Insights:

Females had significantly higher survival chances.

1st Class passengers had better survival outcomes.

Younger and wealthier passengers survived more.

Fare and class were strong indicators of survival.

Conclusion:

The EDA shows that gender, class, and fare were major factors influencing survival on the Titanic.

This analysis forms the foundation for future predictive modeling or machine-learning tasks.

Prepared By:

Ritik Srivastava

(Data Analyst Intern – Task 5: Exploratory Data Analysis)