

Titanic Dataset – Exploratory Data Analysis (EDA) Summary Report

This report presents an exploratory data analysis (EDA) of the Titanic dataset. The analysis was performed in Google Colab using Python libraries such as Pandas, Matplotlib, and Seaborn. The goal was to explore passenger demographics, survival rates, and relationships among key variables.

Dataset Overview

Dataset Name/Location: /content/drive/MyDrive/Intern task/titanic_dataset_sample.csv

Number of Columns: 12

Key Columns: PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked

Shape of Dataset: (Rows × Columns) — as observed from the analysis output.

Data Cleaning Summary

No significant missing values were found in this dataset as per the `df.isnull().sum()` analysis. All columns were kept in their original form for exploratory analysis. The dataset appeared to be pre-cleaned and structured for immediate visualization.

Exploratory Data Analysis (EDA)

The exploratory analysis focused on understanding passenger demographics and survival distribution using Seaborn visualizations.

- **Age Distribution:** The histogram showed that most passengers were young, concentrated in the 20–40 age range.
- **Passenger Class Count:** The majority of travelers belonged to Class 3, followed by Classes 1 and 2.
- **Survival Count by Gender:** Females had a significantly higher survival rate compared to males.
- **Survival Rate by Passenger Class:** Higher-class passengers (Class 1) had a better survival rate than lower-class passengers.
- **Correlation Heatmap:** A positive correlation was found between *Fare* and *Pclass*, and a slight negative correlation between *Age* and *Survived*.

Key Insights

1. Females were more likely to survive than males.
2. Passengers from higher classes (1st class) had better chances of survival.
3. Younger passengers tended to survive more often than older ones.
4. Fare prices varied significantly across passenger classes.
5. The dataset reflects clear socio-economic differences affecting survival.

Conclusion

The analysis of the Titanic dataset highlights the impact of gender, passenger class, and age on survival outcomes. This EDA provides a foundation for building predictive models, such as logistic

regression, to estimate survival probability.