

Titanic Dataset – EDA Summary

The exploratory data analysis (EDA) of the Titanic dataset revealed important insights about passenger demographics, survival patterns, and key features influencing survival:

- **Passenger Profile:** Most passengers were male (~64%), aged 20–30, and traveled in 3rd class. Children (0–5 years) had higher survival, while seniors survived the least.
- **Survival Factors:** Survival was strongly influenced by Pclass, Sex, Age, Family Size, and Embarkation Port. Females, 1st class travelers, and passengers from smaller families had higher chances of survival.
- **Fare Insights:** Fare data was highly skewed; an Individual_Fare feature was created to better capture per-person fare. Outliers reflect genuine wealthy passengers.
- **Family & Cabin Features:** Large families had lower survival rates; Cabin information was transformed into FloorType for additional predictive insight.
- **Feature Engineering:** Missing values in Age and Cabin were handled, and new features like Age_type and Family_Size were created to improve model performance.
- **Correlation Analysis:** Numerical features showed relationships that guided feature selection for predictive modeling.

Key Takeaway: Younger passengers, females, first-class travelers, and small families had higher survival chances. Feature engineering and EDA insights help in building robust predictive models.