



Exploratory Data Analysis (EDA) Report

Dataset: Titanic - `train.csv`



Key Findings

1. Correlation Analysis:

- Sex (-0.54) and Pclass (-0.34) are negatively correlated with Survived, showing that **females** and those in **higher classes** had higher survival chances.
- Fare shows a **positive correlation** (+0.26) with Survived.

2. Distribution Insights:

- Age** follows a right-skewed distribution.
- Fare** is highly skewed with few passengers paying significantly more.

3. Boxplots:

- Survivors** tend to be **younger** on average.
- Higher **fare** passengers had better survival outcomes.

4. Scatter Plot (Age vs Fare by Survival):

- Clear clustering shows that **high-fare, low-age** passengers were more likely to survive.
-



Model Accuracy from EDA Features

- A **Logistic Regression** model trained using:
 - Sex, Pclass, Age, Fare, SibSp, Parch, Embarked
 - Achieved an **accuracy of approximately: 0.82 (82%)**
-



Conclusion

- Sex, Pclass, Fare, and Age** are the most **influential factors** in survival.
- EDA helped us uncover meaningful patterns to guide feature selection for modeling.
- The baseline model based purely on EDA-informed features shows promising performance.