
Task 5



Titanic Dataset Analysis Summary



Dataset Overview

The dataset `finaltask4.csv` contains information on 1310 passengers from the Titanic.

Initial inspection showed several missing values in columns like `age`, `cabin`, `embarked`, `body`, `boat`, and `home.dest`.



Missing Value Treatment

Dropped columns with excessive missing data: `body`, `boat`, `home.dest`, and `cabin`.

Filled missing `age` values using the median of the column.

Filled missing `embarked` values using the mode (most frequent port).

◆ Exploratory Data Analysis (EDA)

✓ Survival Count

count plot shows that more passengers did not survive (value 0) than those who survived (1).

✓ Survival by Gender

Females had higher survival rates.

Most males did not survive, indicating women were given priority during rescue.

✓ Survival by Passenger Class

First-class passengers had the highest survival rate, followed by second-class.

Third-class had the lowest survival, showing class disparity during evacuation.

◆ Correlation Heatmap Insights

Strong negative correlation between fare and pclass (-0.56) — higher fare = higher class.

survived is weakly positively correlated with fare and negatively with pclass.

age, sibsp, and parch have minimal correlation with survival.

◆ Visualizations Used

`sns.countplot()` for:

- Overall survival count
- Survival by sex
- Survival by pclass

`sns.heatmap()`:

- for correlation visualization

Pair Plot() :

- Visualizes relationships between features like fare, age, pclass, etc.
- Survival distribution shown with blue (0 = not survived) and orange (1 = survived).
- Higher fares seem correlated with higher survival.

Fare vs Pclass Boxplot :

- Passengers in 1st class paid the highest fares, followed by 2nd and 3rd class.
- Many outliers in 1st class fare values.

Violin Plot: Age vs Sex & Survival :

- Females had a higher survival rate compared to males.
- Age range of survivors is mostly between 20–40 years, especially among women

Key Takeaways

- Gender and Class played a significant role in survival probability.
 - Females and higher-class passengers had better survival chances.
 - Handling missing data properly is crucial for accurate analysis.
 - Visualization is key to uncovering trends and patterns in data.
-