

# Data Analysis Report

This report provides an overview of the dataset 'test.csv' based on the automated EDA process. It includes summary statistics, missing value counts, and visualizations to understand the distribution, relationships, and possible outliers in the data.

## Key Findings:

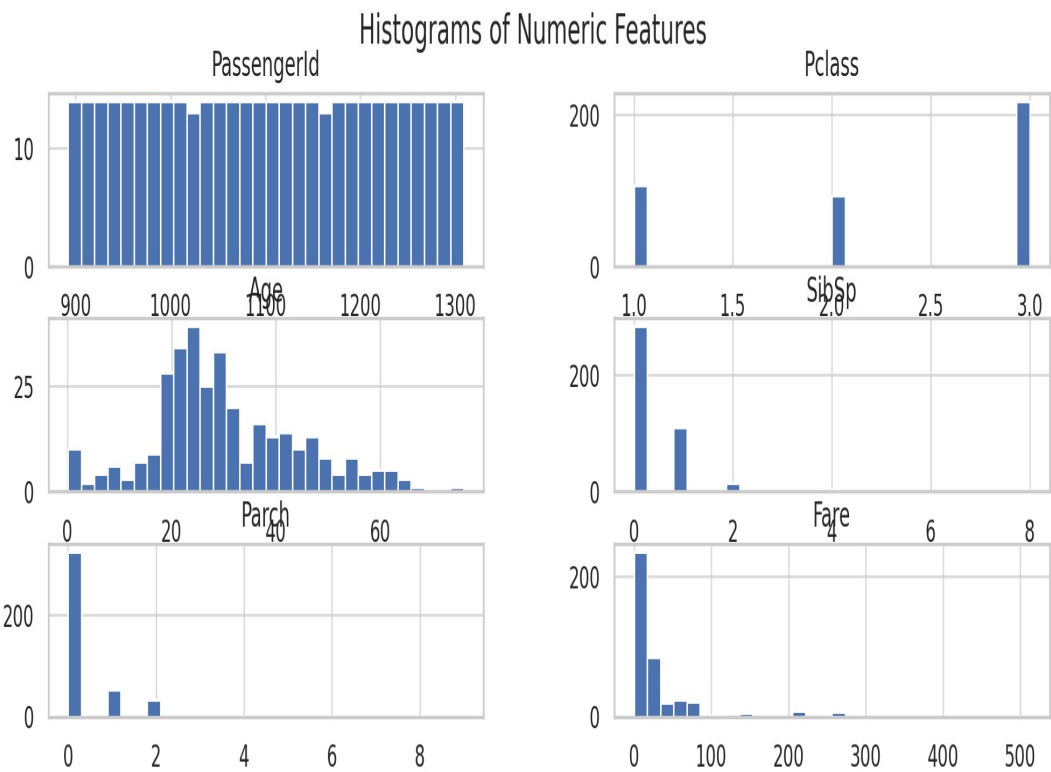
### 1. Data Overview

- **Rows & Columns:** 418 rows, 11 columns.
- **Data Types:** Mix of numeric (e.g., Age, Fare, SibSp, Parch, Pclass) and categorical (Sex, Embarked, Cabin, Ticket, Name).
- **Missing Values:** Significant missing values in Cabin (327), moderate in Age (86), and 1 in Fare.

### 2. Univariate Analysis (Histograms)

- **Age:** Slight right skew, indicating a higher concentration of younger passengers.
- **Fare:** Highly right-skewed, with a few extreme high values (outliers).
- **SibSp & Parch:** Majority of passengers had 0 siblings/spouses or 0 parents/children aboard.

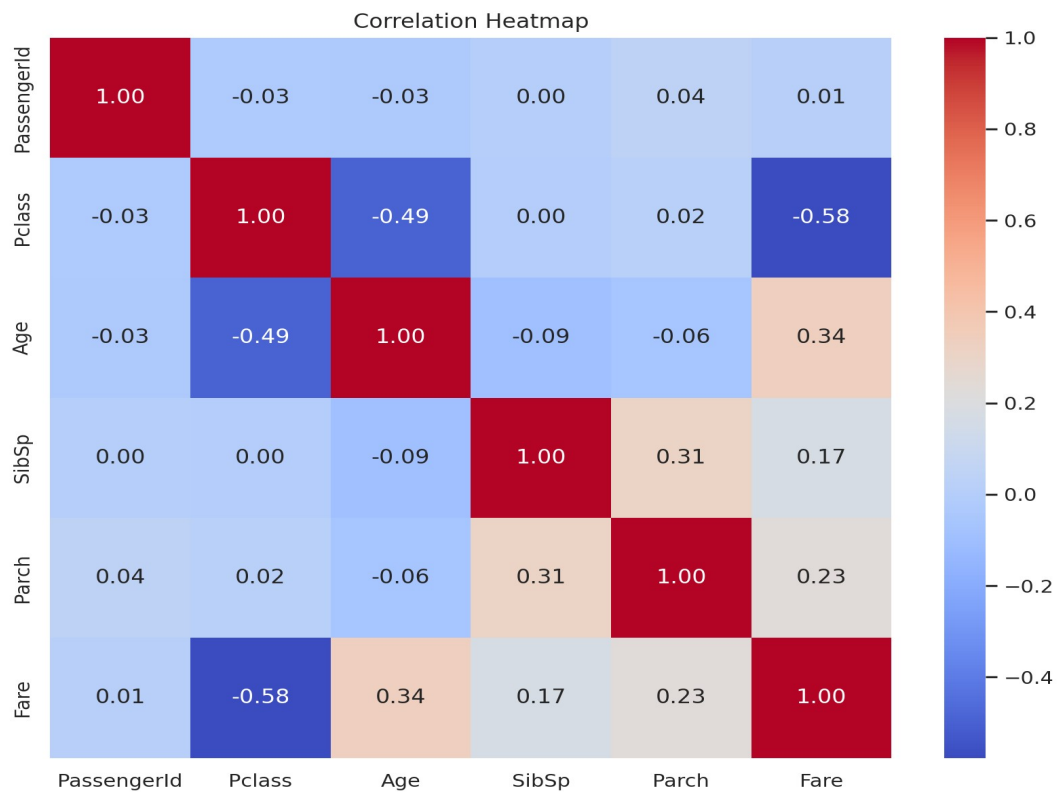
Hamdaan Sayed Task 5



# Hamdaan Sayed Task 5

## 3. Correlation Analysis (Heatmap)

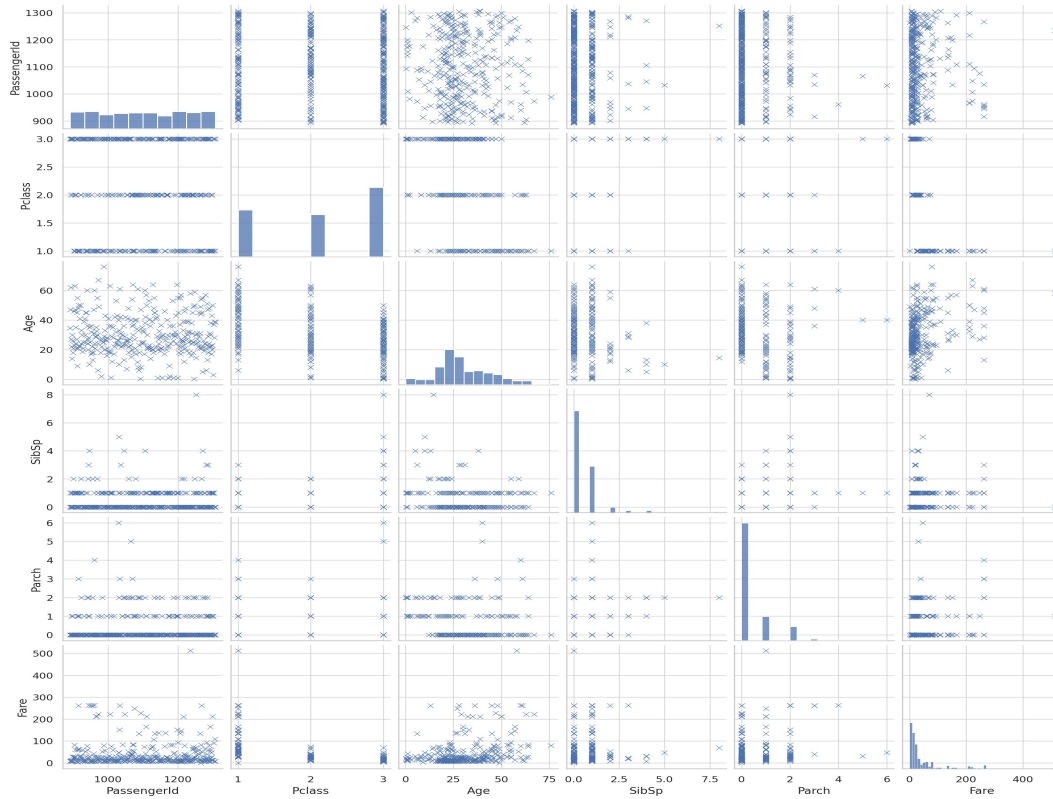
- **Positive Correlations:**
- `Pclass` and `Fare`: Negative correlation (higher class = higher fare). `SibSp` and `Parch`: Moderate positive correlation (traveling with more family often means more parents/children as well).
- **Low correlation** between most numeric variables, suggesting no severe multicollinearity.



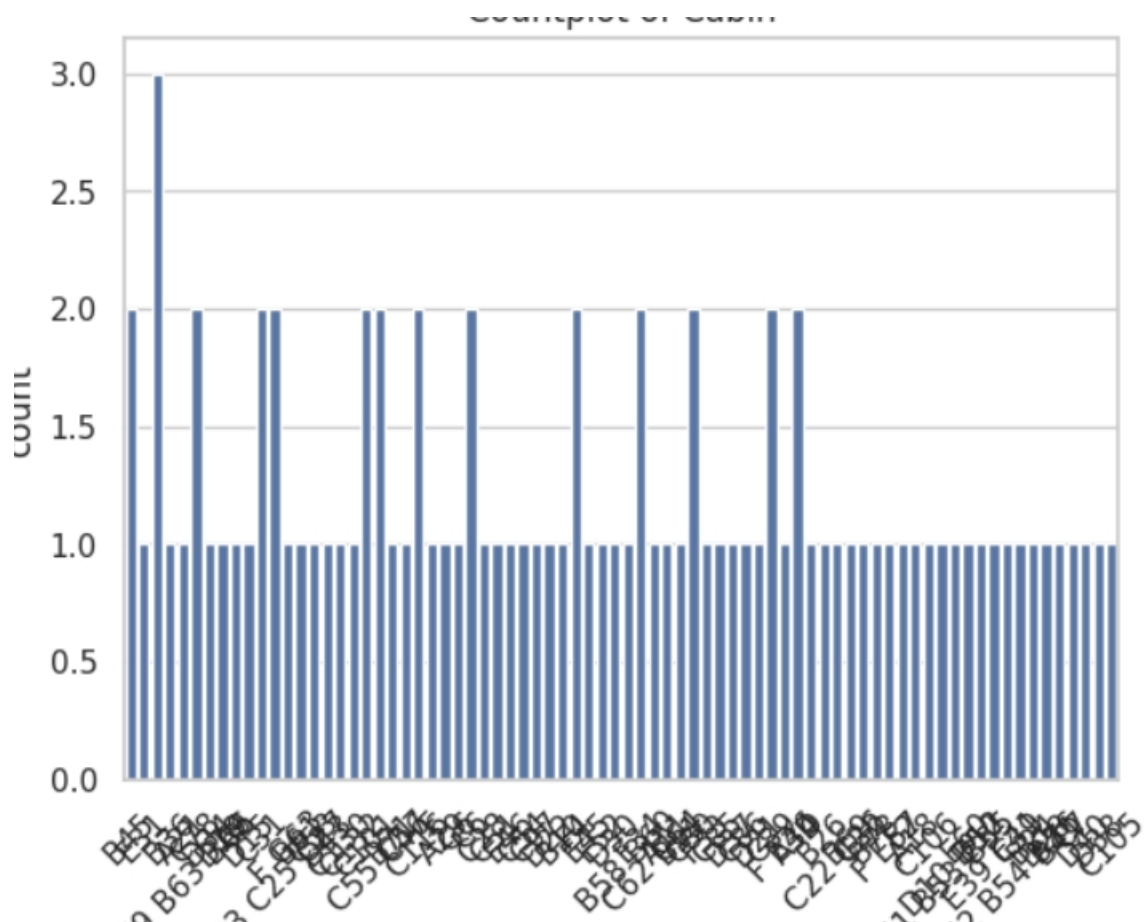
# Hamdaan Sayed Task 5

## 4. Multivariate Analysis (Pairplot)

- Clear visual separation in fare distribution by passenger class.
- High-fare outliers clustered among higher-class passengers.



## Hamdaan Sayed Task 5



## 5.Conclusion :

The dataset exhibits strong class-based differences in fares, notable skewness in certain variables, and missing values that will require handling before predictive modeling. Insights from these patterns can inform feature engineering and guide model selection in further analysis.