Findings & Observations Report

Introduction

This report presents the findings and observations derived from the analysis of the dataset. The focus is on identifying missing values, understanding summary statistics, and highlighting meaningful patterns in the data. The insights are kept minimal and direct, yet sufficiently detailed to provide clarity for interpretation.

Data Overview

The dataset consists of 891 rows and 12 columns. It represents passenger information including demographic details, travel class, and survival status. During preprocessing, it was observed that several columns contain missing values which are important to note before conducting deeper analysis.

Missing Values Summary: Age (177), Cabin (687), Embarked (2).

Summary Statistics

Key numerical insights from the dataset include the following: - The average passenger age is approximately 29.7 years, with the youngest recorded as 0.42 years and the oldest at 80 years. - The average fare paid is 32.2 units, with a wide range extending up to over 500 units. - The survival rate is around 38%, indicating that less than half of the passengers survived. - Passenger distribution across travel classes shows more passengers in the lower classes.

Unique Values in Categorical Columns

- Sex: 2 unique values (Male, Female) - Ticket: 681 unique ticket numbers - Cabin: 147 distinct cabin identifiers - Embarked: 3 unique embarkation points (C = Cherbourg, Q = Queenstown, S = Southampton) This highlights that while categorical variables such as sex and embarkation are concise, others like ticket and cabin are highly diverse and sparse.

Key Observations

1. Passenger Class Distribution: The majority of passengers traveled in the 3rd class, which reflects affordability for the larger population. 2. Missing Data Challenge: A significant portion of the cabin information is missing (over 75%). This makes cabin data less reliable for predictive modeling. 3. Gender and Survival: Survival chances were higher among females compared to males, suggesting priority or protection during evacuation. 4. Age Factor: Younger passengers, particularly children, showed higher survival rates. 5. Fare and Class Relation: Higher fares are strongly associated with higher passenger classes, which in turn correlate with better survival chances.

Conclusion

The analysis provides clear evidence of survival patterns based on passenger class, gender, and age. The presence of missing data, especially in the cabin column, is a limitation but does not diminish the overall trends. These findings offer a foundation for predictive modeling or further exploratory analysis.