**INTRODUCTION**

1. Description of the Titanic Dataset

Overview

The Titanic dataset contains information about the passengers aboard the RMS Titanic, which sank on its maiden voyage in 1912. The dataset includes demographic and travel details for each passenger, allowing for analysis of survival rates and factors influencing survival.

Key Variables

1. PassengerId: Unique identifier for each passenger.

2. Survived: Binary indicator of survival (0 = No, 1 = Yes).

3. Pclass: Ticket class (1st, 2nd, 3rd).

4. Name: Passenger's full name.

5. Sex: Gender of the passenger.

6. Age: Age in years.

7. SibSp: Number of siblings/spouses aboard.

8. Parch: Number of parents/children aboard.

9. Ticket: Ticket number.

10. Fare: Fare paid by the passenger.

11. Cabin: Cabin number.

12. Embarked: Port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton).

Initial Insights

1. Survival Rate: Approximately 38% of passengers survived.

2. Class Distribution: Most passengers were in the 3rd class.

3. Gender Distribution: More males than females were aboard.

4. Age Range: Includes a wide age range from infants to elderly passengers.

5. Family Connections: Many passengers traveled with family members, indicated by SibSp and Parch variables.

6. Embarkation Points: Majority of passengers boarded at Southampton.

Data Quality

- Missing Values: Notable in Age, Cabin, and Embarked columns.

- Categorical Variables: Need to be encoded for analysis.

- Continuous Variables: Need to be scaled for some machine learning algorithms.

Purpose of the Review

The purpose of this review is to conduct a preliminary analysis of the Titanic dataset to identify key variables, data types, and any obvious patterns, trends, or anomalies. This initial review helps in understanding the structure and content of the dataset, which is crucial for developing further in-depth analysis and building predictive models. Identifying missing values, outliers, and significant factors related to survival can guide the next steps in the data analysis process.

**2. OBSERVATION**

Key Observations from Titanic Dataset:

1. Survival Rate:

- Approximately 38% of passengers survived.

- Higher survival rates observed among females compared to males.

2. Passenger Class:

- First-class passengers had a higher survival rate compared to second and third-class passengers.

- Survival rates: 1st Class (63%), 2nd Class (47%), 3rd Class (24%).

3. Gender:

- Female survival rate: 74%

- Male survival rate: 19%

- Gender appears to be a significant factor in survival.

4. Age:

- Children (age < 18) had higher survival rates compared to adults.

- Infants had the highest survival rate among age groups.

5. Family Relations:

- Passengers with fewer family members (SibSp + Parch) on board tended to have higher survival rates.

- Solo travelers had moderate survival rates.

6. Fare:

- Higher ticket fares are associated with higher survival rates, potentially linked to passenger class.

7. Embarkation Port:

- Passengers embarking from Cherbourg (C) had higher survival rates than those from Queenstown (Q) or Southampton (S).

**3. Conclusion**

Initial analysis suggests that survival on the Titanic was influenced by passenger class, gender, age, family size, ticket fare, and port of embarkation. Further in-depth analysis is recommended to uncover deeper patterns and interactions between variables.

Recommendations for Future Analysis

- Perform multivariate analysis to understand the interplay between different features.

- Investigate missing data and consider imputation methods for accurate analysis.

- Apply machine learning models to predict survival based on the identified features.

This preliminary report provides a foundation for a more comprehensive analysis of the Titanic dataset, potentially guiding improvements in safety protocols for modern maritime travel.

This is a stage zero Task For Data Analysis Interns at HNG 11. For more information reference the links below:

<https://hng.tech/internship>,  <https://hng.tech/hire>