TITANIC DATASET REPORT

The Titanic dataset is a well-known dataset that provides information about passengers on the Titanic, including details such as class, sex, age, survival status and more. This report aims to explore and analyze various aspects of the dataset, including summary statistics, distribution of passengers, survival rates and correlation between variables.

The summary statistics often a comprehensive overview of the dataset. Key metrics such as mean, standard deviation, and quartiles are provided for numerical variables, while count and unique values are presented for categorical variables. This section serves as a quick reference to the overall characteristics of the dataset.

Distribution of Passengers

Distribution of Class

The distribution of passengers by class reveals the number of individuals in each class (First, Second and Third). This information provides insights into the socio-economic composition of the Titanic passengers.

Name : class, dtype: int64

Distribution by Sex

Examining the distribution of passengers by sex sheds light on the gender composition of the dataset.

Name : sex, dtype : int64

Distribution by Age

The distribution of passengers by age is visualized through a histogram. This graphical representation offers a glimpse into the age demographics of the Titanic passengers.

Survival Rate by Class

Understanding the survival rate by class helps identify potential patterns related to socio-economic factors. The survival rates for each class are as follows:

Name : survived, dtype: float64

Correlation Matrix

A correlation matrix provides insights into the relationships between numerical variables. Positive or negative correlations close to 1 or -1 suggest strong relationships, while values close to 0 indicate weaker correlations. The correlation matrix is visualized as a heatmap.

In conclusion, this report offers a comprehensive exploration of the Titanic dataset, providing valuable insights into the distribution of passengers, survival rates, and correlations between variables. Further analysis and visualizations can be conducted based on specific research questions or areas of interest.

Name: G. Erla

Member No.: 23040971

GitHub repository: https://github.com/GopiErla/statistics