**Titanic Dataset Documentation**

**Overview:**

The Titanic dataset is a historical dataset containing information about passengers aboard the RMS Titanic, which sank on its maiden voyage in 1912. This dataset provides valuable insights into the demographics and survival outcomes of the passengers.

**Dataset Information:**

The dataset consists of the following columns:

Passenger Id: A unique identifier for each passenger.

Survived: Indicates whether the passenger survived (1) or not (0).

Pclass: The passenger class (1 = 1st class, 2 = 2nd class, 3 = 3rd class).

Name: The name of the passenger.

Sex: The gender of the passenger (male or female).

Age: The age of the passenger in years.

SibSp: The number of siblings/spouses aboard the Titanic.

Parch: The number of parents/children aboard the Titanic.

Ticket: The ticket number.

Fare: The fare paid for the ticket.

Cabin: The cabin number.

Embarked: The port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton).

**Data Cleaning and Preprocessing:**

The following steps were performed to clean and preprocess the dataset:

Missing values in the 'Age' column were imputed with the median age.

Missing values in the 'Embarked' column were imputed with the mode.

The 'Cabin' column, which had a large number of missing values, was dropped.

Duplicate rows, if any, were removed.

Categorical variables were mapped to numeric values for modelling purposes.

Additional features were engineered, such as 'Title' (extracted from 'Name'), 'FamilySize' (combining 'SibSp' and 'Parch'), and 'Alone' (indicating if the passenger was traveling alone).

**Exploratory Data Analysis:**

The dataset was analysed to gain insights and answer relevant questions. Some of the key visualizations and findings include:

Distribution of passenger ages: A histogram showing the age distribution of passengers.

Gender distribution among passengers: A bar chart displaying the count of male and female passengers.

Survival rate based on passenger class: A bar chart showing the survival rate for each passenger class.

Fare variation based on passenger class: A box plot comparing the fare distribution for different passenger classes.

Distribution of fares paid by passengers: A histogram illustrating the distribution of fares.

Survival rate based on the port of embarkation: A bar chart displaying the survival rate for each embarkation port.

Survival rate based on the number of siblings/spouses aboard: A bar chart showing the survival rate for different numbers of siblings/spouses.

Age variation based on passenger class: A box plot comparing the age distribution for different passenger classes.

Survival rate based on passenger title: A bar chart showing the survival rate for different passenger titles.

Survival rate based on the number of parents/children aboard: A bar chart illustrating the survival rate for different numbers of parents/children.

**Statistical Analysis**

Statistical analysis was performed to investigate relationships and derive insights from the data. Some of the statistical techniques applied include:

Summary statistics: Descriptive statistics, such as mean, median, minimum, maximum, and quartiles, were calculated for numerical variables.

Correlation analysis: The correlation matrix was computed to measure the relationships between variables.

Hypothesis tests: One-way ANOVA and independent two-sample t-tests were conducted to test for significant differences in survival rates and mean age between different groups.

**Findings and Insights:**

The analysis of the Titanic dataset revealed the following findings and insights:

The survival rate varied based on passenger class, gender, and the number of family members aboard.

Children had a higher chance of survival compared to adults.

Passengers who paid higher fares had a higher survival rate.

The port of embarkation showed variations in survival rates.

The dataset provided insights into the impact of passenger titles on survival.

These findings provide valuable insights into the factors that influenced the survival of passengers on the Titanic and can aid in further analysis or model development.