**TITANIC ANALYSIS**

**Executive:** The aim is to perform a statistical analysis of the fatalities on the ship using the Titanic dataset.

**Background:** The RMS Titanic was a British passenger liner that sank in the North Atlantic Ocean in the early morning of April 15, 1912 after colliding with an iceberg during her maiden voyage from Southampton to New York City. The ship contained 2,224 passengers and crew, out of which 1,500 died in the unfortunate incident.

**Purpose:** The main purpose is to check that whether there is a statistically significant relation between the death of the person and their passenger class, age, sex and/or the port where they embarked their journey.

**Questions:**

1. Import train.csv file from Titanic\_dataset.

2. Factors and Levels

A. Find number of Passengers according to their Group Class: 1st , 2nd , 3rd

B. Find number of Passengers according to their Group Sex: Male, Female

c. Find stats of Passengers Age –Rectify if Age is less than one, is value fractional ?

D. Find number of Passengers according to their Group Embarked: Place where the passenger embarked their journey. One of Cherbourg, Queenstown or Southampton.

3. Response Variables

A. Validate number of passengers who survived / Not Survived

4. Exploratory Data Analysis: A. Explore / Print first n Records from Dataset.

B. Find mean, median, quartile, max, min data for every feature.

C. For the purposes of this study, we work with only four input variables and one response variable.

Input variables : Passenger Class, Sex, Age, and Port of Embarkment.

Response variable : Survived.

D. Perform data cleaning steps

E. Encode Data

Make Age as a categorical variable as follows:

# If age <= 18, then age = child

# If 18 < age <= 60, then age = adult

# If age > 60, then age = senior - use if else condition

F. Validate above 2 steps.

5. Data Analysis to perform

Computing main effects for all four factors

Validate computed effects if True

Draw Conclusion over above Analysis

A. Plot the barplot of all four input variables:

B. Convert the categorical dataframe into numeric dataframe.

6. Statistical Analysis:

A. Number of survivors on an average from Class & Plot a scatter plot

B. Number of survivors on an average from Gender & Plot a scatter plot

C. Number of survivors on an average from Every Port of Embarkment & Plot a scatter plot.

D. Validate above scatterplots using ANOVA ( 1 way Interaction )

**Methods:**

Read.csv() was used to read and import the dataset from the system to do the analysis.

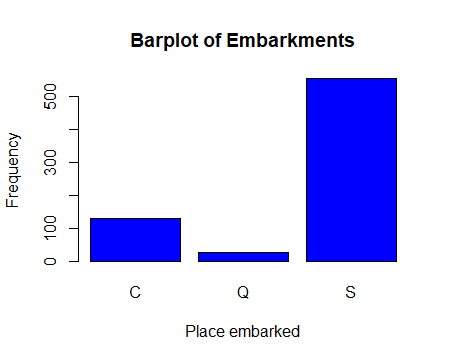
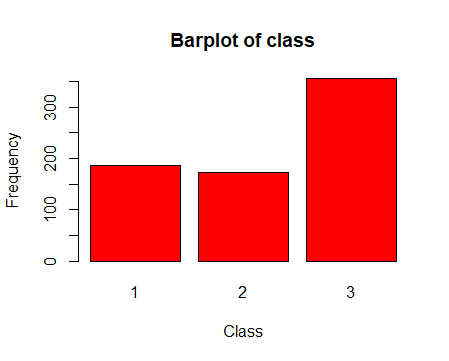
Exploratory analysis was done using as.factor(), head(), dim(), names(), class(), str(), summary(), boxplot(), hist(), subset(), na.omit(), rownames()

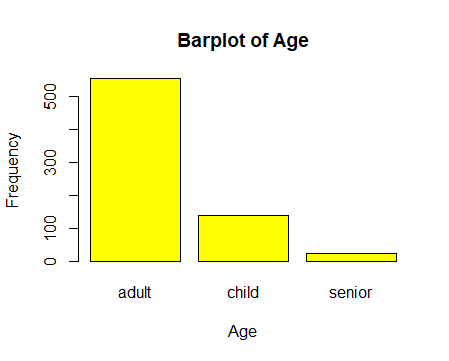
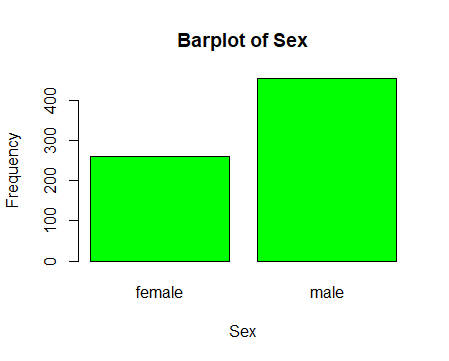
A statistical analysis of the fatalities on the ship using the Titanic dataset which was validated or confirmed with scatterplots and using ANOVA (1 way interaction).

**Instrumentation:** The analysis was done on RStudio with R-programming

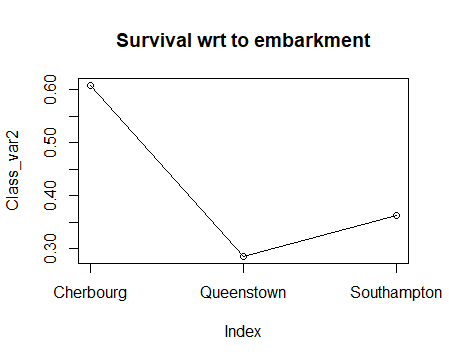
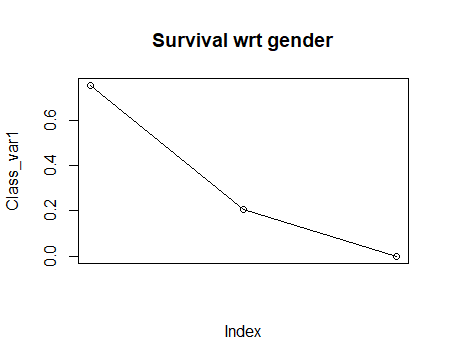
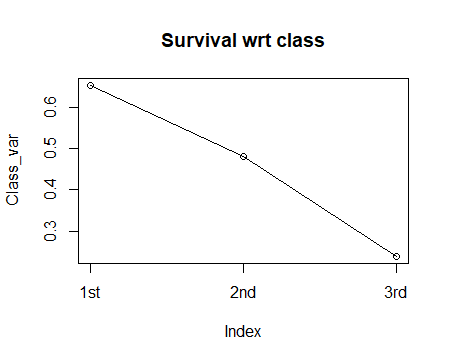
**Results:** The survivors were successfully analyzed by their features. Following are the findings and the plots.

Barplot of all four input variables

Scatterplots of survivors wrt to four features

  …………………………………...

**Summary:**

The Titanic dataset was analyzed. The data contained details of the passengers that included passenger id, class, name, sex, age. Ticket, ticket fare, cabin and their embarkment. The 1st class contained 216 passengers, the 2nd class, 184 passengers and the 3rd class 491 passengers. There were 314 females in all and the rest were males. There were 891 details about the passengers. 168 had their embarkment from Cherbourg, 77 from Queenstown and 644 from Southampton. The survival to non-survival ratio was 549:344. The mean, median, quantile, max and min for every feature of the data was summarized. The ages of 177 passengers were not present in the given dataset, Also 2 had unknown embarkments. Hence that data was omitted to do the evaluation. The data evealuation was finally done with 712 passengers. The scatterplots and bargraphs were plotted. Anova(1 way interaction) was performed