Titanic Dataset Analysis

**EXECUTIVE**:

It is the analysis of the data about all the passengers that were travelling the Titanic.We needed to determine the factors/attributes which contributed in saving the lives of the survived passengers. What we learnt from this analysis is that the age, gender, passenger class and port of embarkment greatly contributed to the passenger’s survival.

**INTRODUCTION AND PURPOSES**:

This data analysis report using R language highlights the verification of the data consistency, bar plots of the variables, Conversion of categorical data into numeric data and validation of the statistical analysis.

The analysis is conducted by Suven Consultants and Technology Pvt. Ltd. for an online internship corresponding to the course of “Data Analytics Using R”.

The analyser of this dataset is Neha Mirani.

**LIMITATIONS**:

This analysis has got certain limitations:

* Accuracy percentage of the prediction model for this dataset is less
* Analysis cannot be automated using R language

**METHODS**:

* **Verification of data consistency:**

Checking and omitting the NA and blank values

* **Bar Plotting:**

Generic bar plotting of input variables

* **Conversion of data:**

Converting the categorical data into numeric values

* **Validation of the statistical analysis:**

Validating the statistical data

**SAMPLE**:

People who contributed to the analysis are as follows:

* **Niraj Sharma**: Instructor at SCTPL, taught us how to use R language for data analysis
* **Neha Mirani**: Student at SCTPL, trained by Niraj Sharma, made the full analysis on the Titanic Dataset along with this analysis report

**INSTRUMENTATION**:

The tools used by us are as follows:

* R Studio IDE
* R v3.5.1

**RESULTS**:

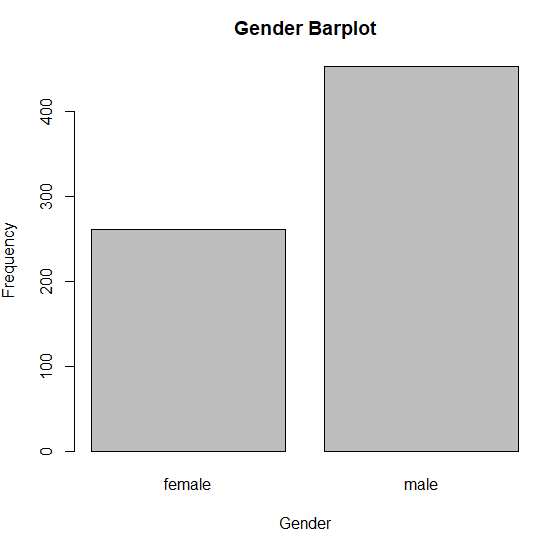
Following are the findings after the analysis has been done:

* The number of passengers travelling that belong to Class 1 are 216, Class 2 are 184 and Class 3 are 491
* There were 314 females and 577 males travelling in the ship in total
* There are records present where age is less than 1 and the age value for such records is fractional
* 168 passengers embarked their journey from Cherbourg, 77 embarked their journey from Queenstown, 644 embarked their journey from Southampton
* 342 passengers survived and 549 did not

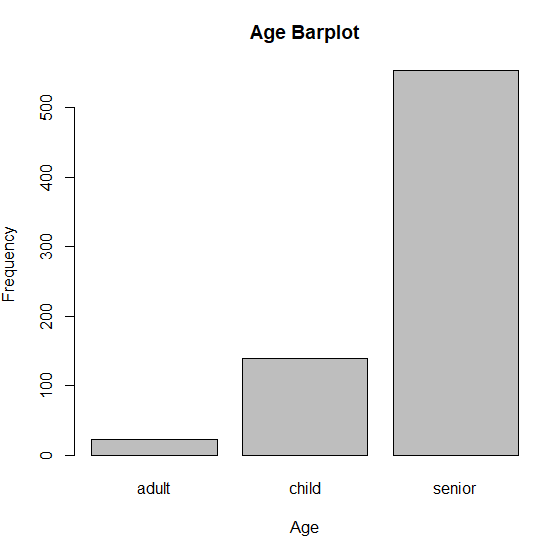
Below are the findings using statistical analysis:



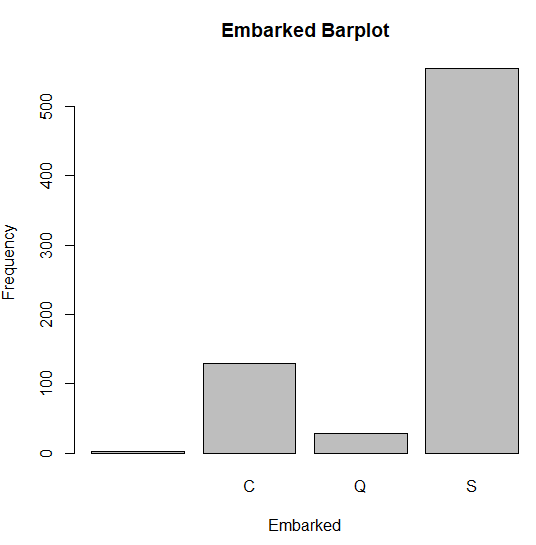
**Finding:** According to the above bar plot we find that majority of the passengers are from Class 3.



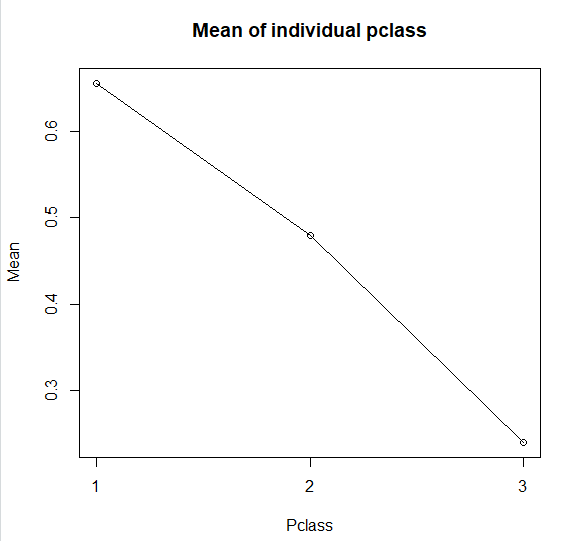
**Finding:** As we can see in the above plot, number of males is higher than females.

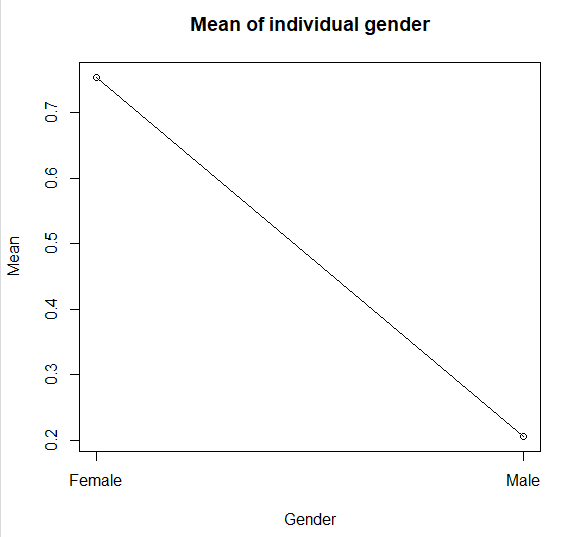


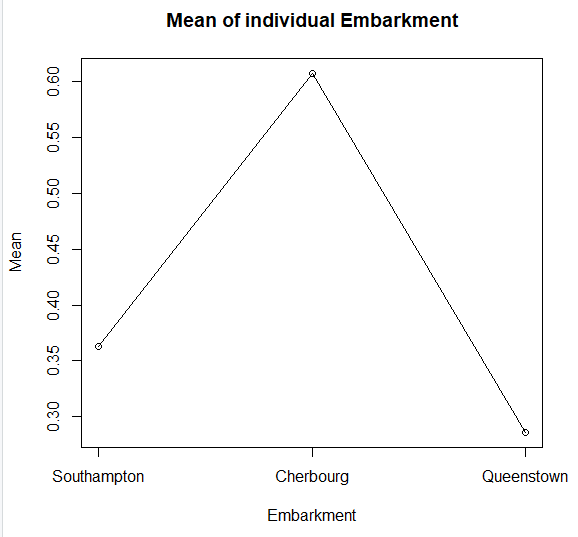
**Finding:** Here we find that majority of the passengers are senior people.



**Finding:** Here we can see that majority of the passengers embarked their journey from Southampton.







**Finding:** The highest average number of survivors belong to Class 1, female gender and port of embarkment is Cherbourg.

We also validated our scatter plots using anova function (1-way interaction) and our finding is that there is a statistically significance of 95% or above for the relation between the death of the person and their passenger class, gender, age and port of embarkment.

**RECOMMENDATIONS**:

This is just a practice analysis and real-life recommendations may not be possible.

**SUMMARY**:

We performed analysis of survivors in the ship which sank in the North Atlantic ocean. We performed a statistical analysis of the fatalities on the ship using the Titanic dataset. Our findings from this analysis is that there is a statistically significance of 95% or above for the relation between the death of the person and their passenger class, gender, age and port of embarkment.

**REFERENCES**:

<https://www.analyticsvidhya.com/>

<https://www.r-bloggers.com/>

<https://www.tutorialspoint.com/index.htm>

Notes provided by SCTPL