**Big-Mart Sales Dataset Analysis**

**EXECUTIVE**:

It is the analysis of the data about all the different outlet sales in a big mart mall to determine the factors contributing towards the sales of various items. This analysis shows that the sales prediction made depending on the item weight, fat content, item visibility, type, MRP, outlet size, outlet location type, outlet type and year has the accuracy of only 56%.

**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 was 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 Aahan Gupta.

**LIMITATIONS**:

This analysis has got certain limitations:

* Accuracy percent 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**:

The personal who contributed to the analysis are as follows:

* **Niraj Sharma**: Instructor at SCTPL, taught how to use R language for data analysis.
* **Aahan Gupta**: Student at SCTPL, taught by Niraj Sharma, Made the full analysis on the Titanic Dataset along with this report with the help of Ashish Gupta and Niraj Sharma.
* **Ashish Gupta**: Father of Aahan Gupta, who helped Aahan Gupta in making this report.

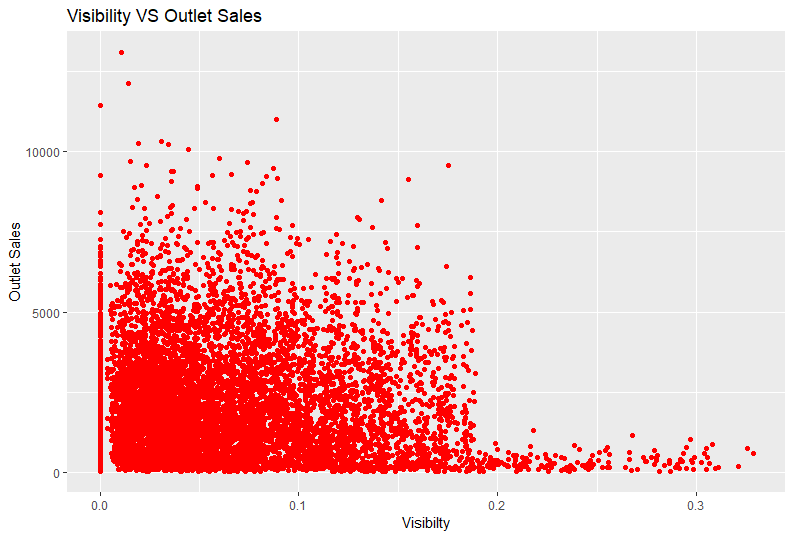
**INSTRUMENTATION**:

The tools used by us are as follows:

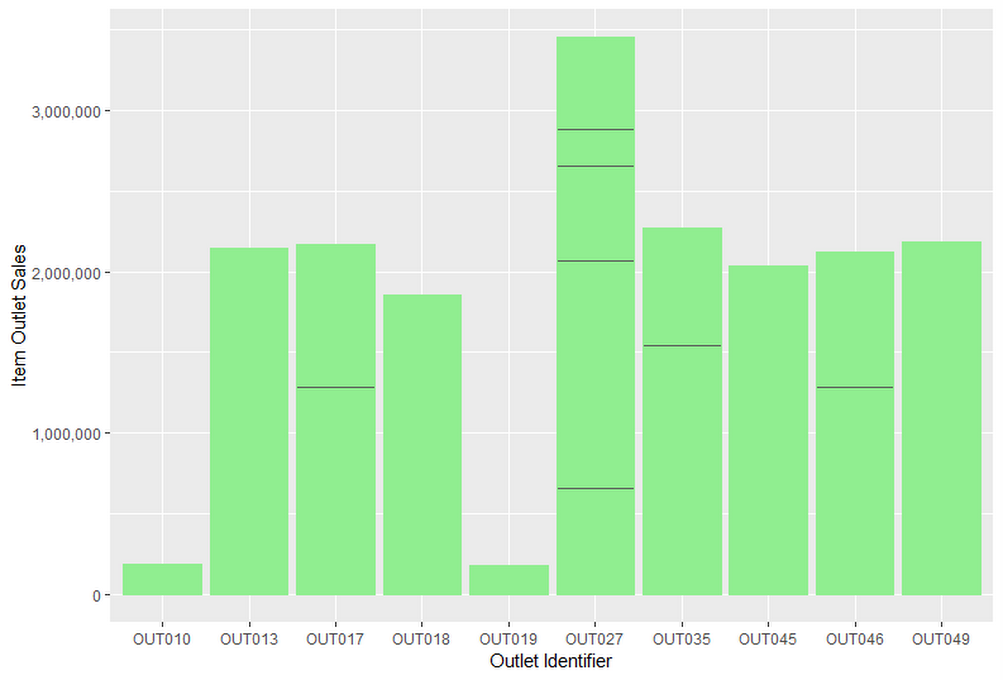
* Rstudio IDE
* R V3.5.1
* Sublime Text Editor

**RESULTS**:

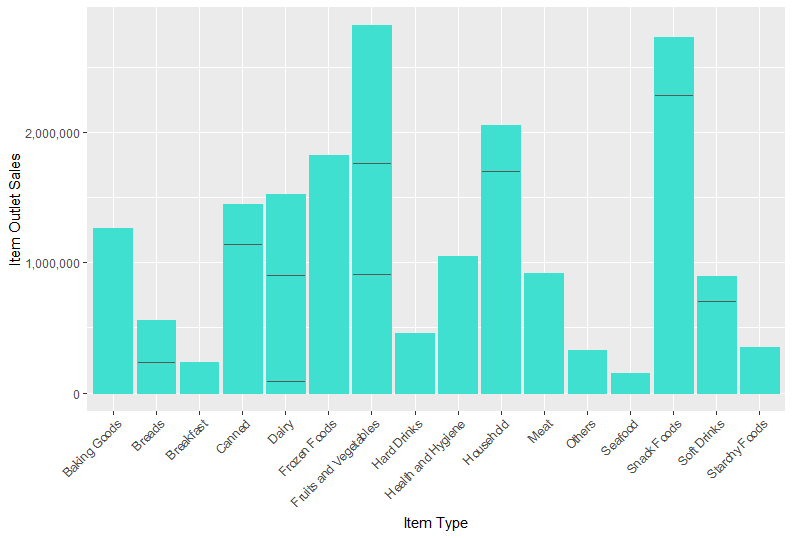
We first made several plots on different features of the dataset for exploratory purposes.



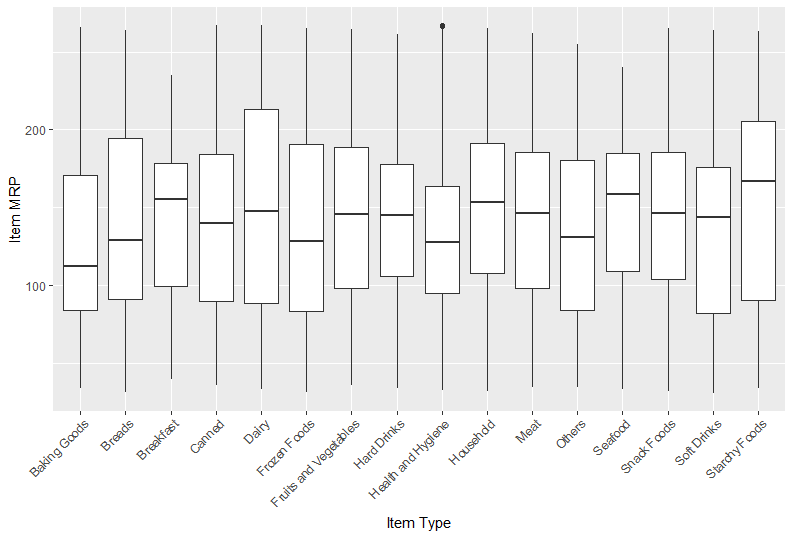
In this scatterplot, we clearly see that the items that have low visibility have much higher sales at the outlets.



Here we see that the outlet with label OUT027 has the highest item sales.



In the above barplot, we find out that fruits and vegetables are sold the most out of all the items.



In the above boxplot, we find that dairy items are the costliest of all the products.

**RECOMMENDATIONS**:

This is just a practice analysis and no real-life recommendations may not be possible.

**SUMMARY**:

This analysis shows that the sales prediction made depending on the item weight, fat content, item visibility, type, MRP, outlet size, outlet location type, outlet type and year has the accuracy of only 56%.

**REFERENCES**:

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

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

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

Notes provided by SCTPL.