Big-Mart Sales Dataset Analysis Report

**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 uses R language and manipulates the dataset to make it consistent before using the data for analysis, making scatter plots of the variables, doing regression test etc.

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 Neha Mirani.

**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:**

Bar plotting of input variables

* **Conversion of data:**

Imputing missing values before doing the data analysis

* **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 us how to use R language for data analysis
* **Neha Mirani**: Student at SCTPL, taught by Niraj Sharma, who did the full analysis on the Bigmart Dataset along with this report preparation

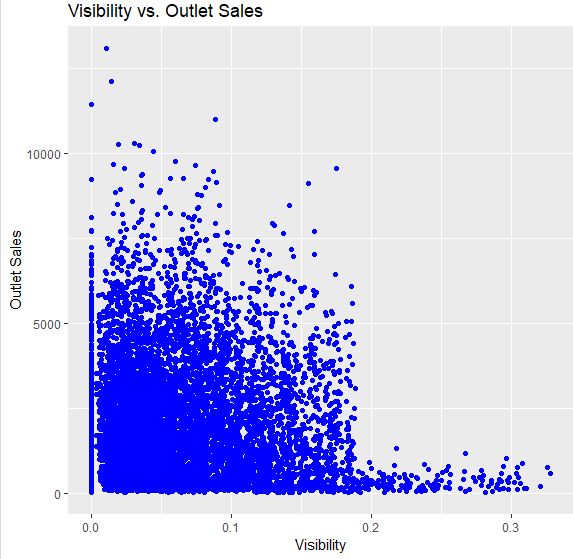
**INSTRUMENTATION**:

The tools used by us are as follows:

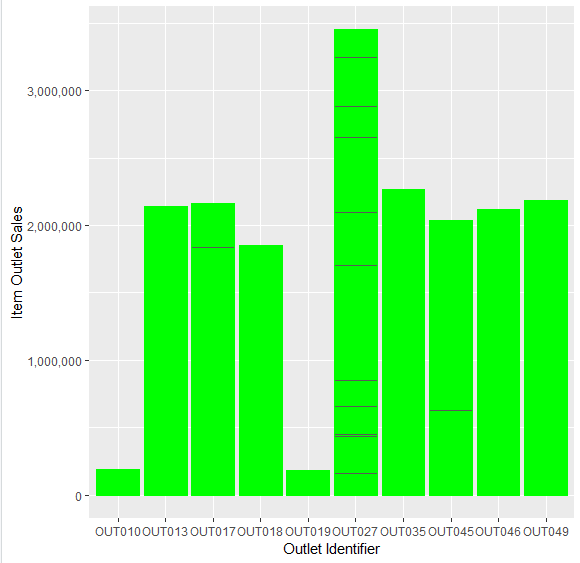
* Rstudio IDE
* R V3.5.1

**RESULTS**:

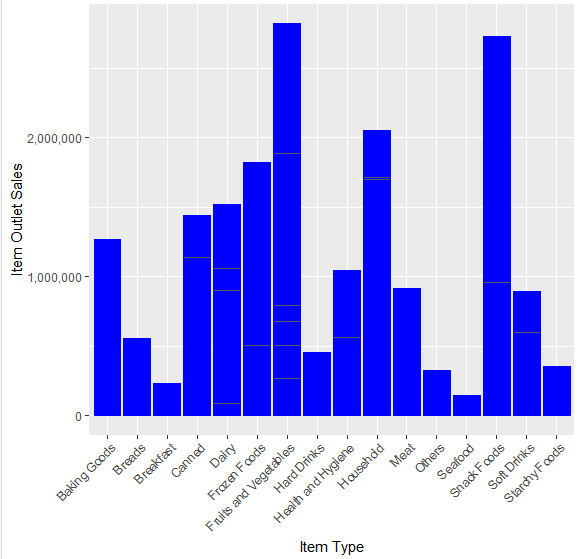
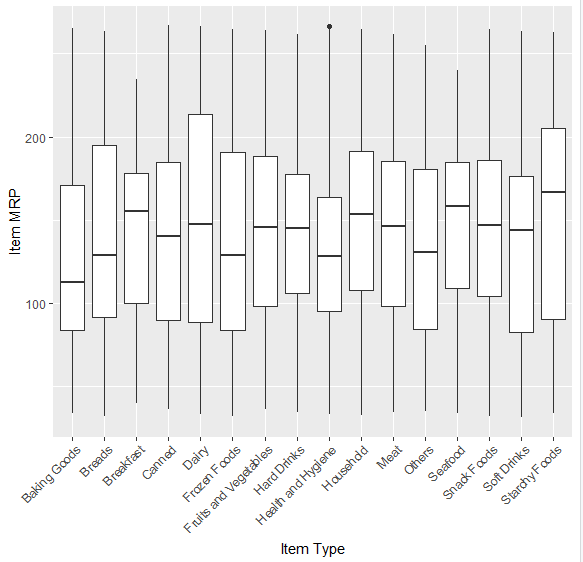
We first made several plots on different features of the dataset for exploratory purposes.



**Finding:** In this scatterplot, we clearly see that the items that have low visibility have much higher sales in the outlets.



**Finding:** Outlet Identifier "OUT027" has contributed to the majority of Sales.

**Finding:** In the above barplot, we see that "Fruits and Vegetables" Item Type are sold more. 

**Finding:** In the above boxplot, we find that "Dairy" Item Type is costlier than the rest other Item Types.

**RECOMMENDATIONS**:

This is just a practice analysis and no real-life recommendations may 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