

Big Mart Sales Analysis

Samrudha Medhekar | Data Analysis in R | 31/7/2018

# Executive

The Big Mart sales dataset was analysed to predict the sales and find the correlation between different factors influencing the sales, like Item Types, Outlet Types, etc. It was found that sales are dependant on particular outlets, their sizes and locations. The outlets that were situated in Tier 3,belonged to the SuperMarket Type 1, and medium size had the highest sales than any other.

## Introduction

The data analysis on the Big Mart dataset was performed by Samrudha Medhekar. We explore the data to:

Understand the data

Summarize the data

Clean and Prune the data

Understand relationships between atrributes

Think about and source other data which maybe useful in answering the question

Get a preliminary feel for the types of models we think would best fit the data

## background

BigMart is an internatioal brand. BigMart retail corporation is one of the fastest growing express home delivery food and grocery retail store company covering worldwide. It has over 22,000 products and over 1200 brands,you will almost get everything you are looking for. It started it’s journey in 2007 with free home delivery services of food and grocery. In the year of 2009 big mart started its operation through its retail store on small size 100 sft. to 300 sft. area. BigMart has grocery products, dairy products, organic products, bakery goods and frozen foods.

## Purposes

Our field of study is concerned with the sales of BigMart. We have the sales data of BigMart of the year 2013. Also, certain attributes of each product and store have been defined. Our aim is to determine what are the factors that are affecting the sales of bigmart and build a predictive model for sales.

Main objective of this study is to analyze the properties of the products and stores. Various factors that is avalibale to us for the analysis are .

**Store variables:-**

Store location

Store Type

Store Size

Store establishment year

**Item variables :-**

Item identifier

Item weight

Item fat content

Item visibility

Item type

Item MRP

Item outlet sales

## Methods

The following methods and techniques were used to perform the analysis :-

1. Dim() – For finding the dimensions of the dataset.
2. Is.na() – For finding any missing values.
3. Summary and str () – For performing exploratory data analysis.
4. Ggplot + geombar() – For plotting the bar plots of various features.
5. Rbind – To bind the training and test datasets together.
6. Lm and predict() – For performing multiple regression to predict the sales.

## Sample:-

The analysis was carried out by Samrudha Medhekar under the guidance of Neeraj Sir for Suven Consultants.

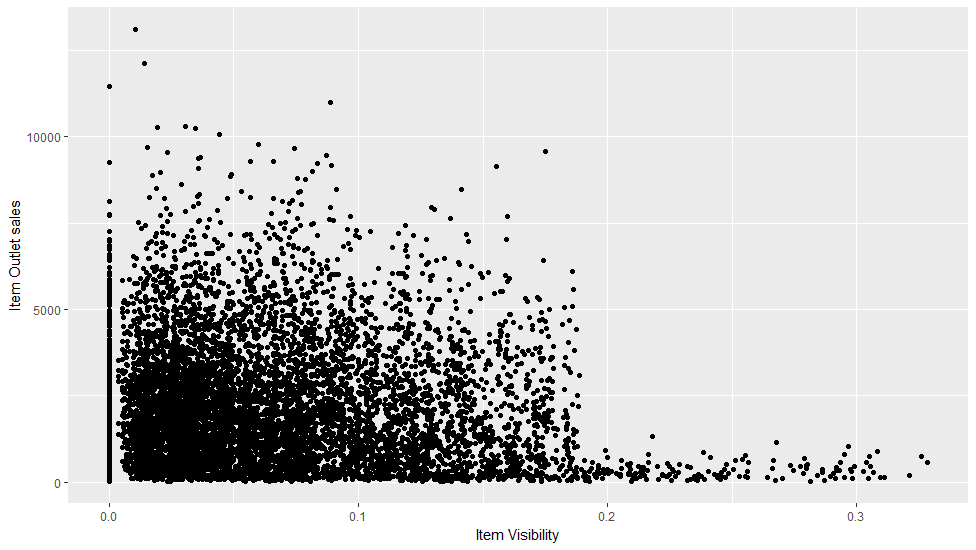
## Instrumentation:-

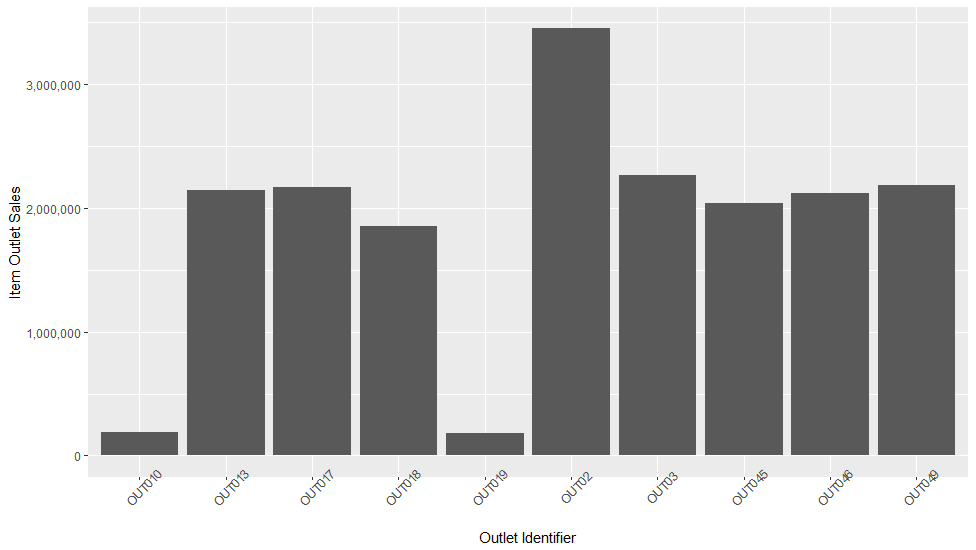
The following tools and libraries were used to perform the analysis:-

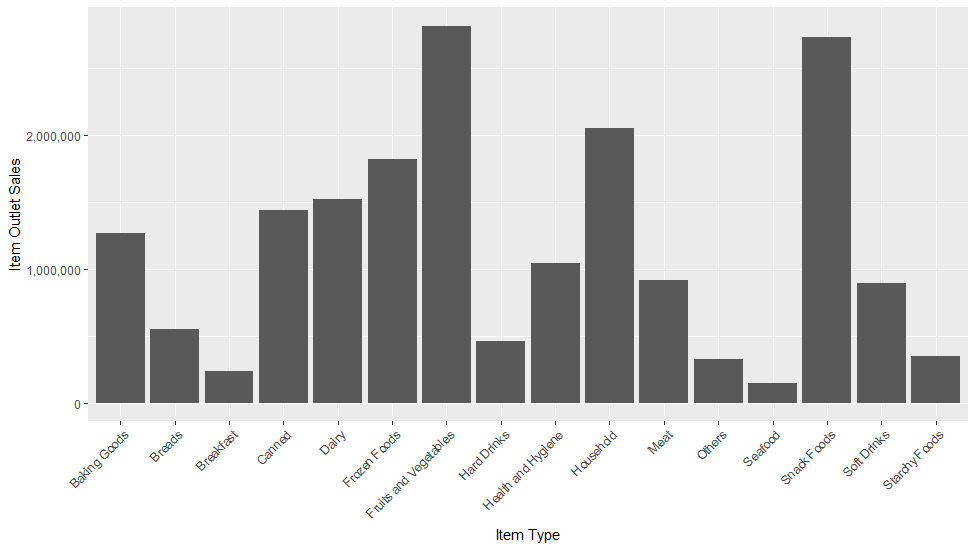
1. R Studio :- For performing the analysis in R language.
2. MS Excel :- For fetching the dataset.
3. Ggplot library :- For plotting the graphs with sophistication and clarity.
4. Libraries plyr and dplyr :- For data manipulation and splitting.

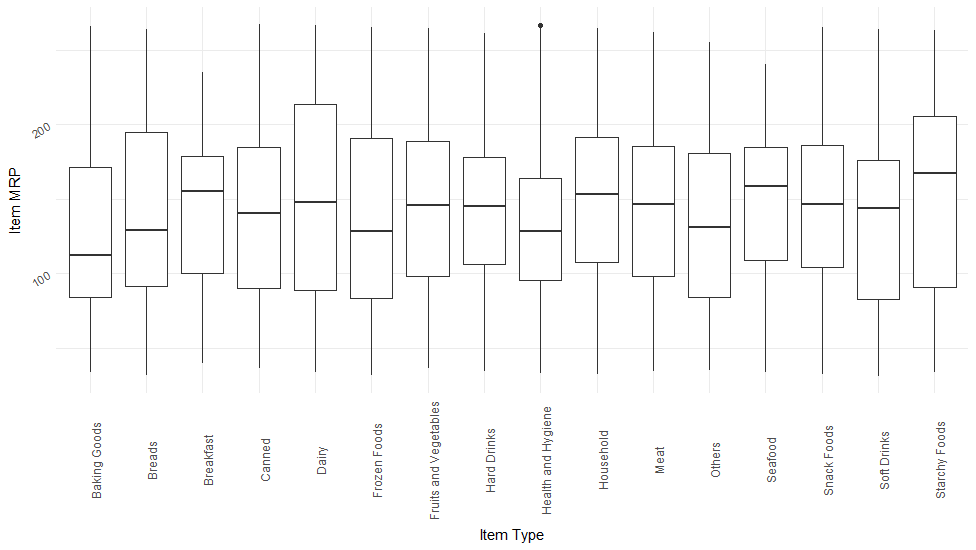
## results

There is a moderate correlation between an Item’s MRP at a Big Mart location and that item’s sales at that location. Also the smallest locations produced the lowest sales. However, the largest location did not produce the highest sales. The location that produced the highest sales was the OUT027 location. This location was Supermarket Type3 and its size was medium. This outlet performed much better than any other location. Its median Item\_Outlet\_Sales were 3364.95. The location that was second was the OUT035 location, which had a median Item\_Outlet\_Sales of 2109.25.









## Recommendations

If Big Mart were to try to increase sales at all locations, it may consider switching more locations to Supermarket Type3. Other things Big Mart could do to increase sales is to see which Items had the highest sales. They may also consider how product visibility affected outlet sales. However, the model built in this report should be good for helping Big Mart predict future sales at its locations.

## summary

The sales analysis suggests that sales are heavily influenced by outlet types, location and sizes. The item types also have a contributing factor in the sales, but item fat content, item visibility, and item weight does not matter much.

## References

<http://www.milanor.net/blog/preparing-the-data-for-modelling-with-r/>

<http://dhrubajitdas44.blogspot.com/2017/10/machine-learning-project-1-big-mart.html>

<https://towardsdatascience.com/data-science-case-study-optimizing-product-placement-in-retail-part-1-2e8b27e16e8d>

<https://www.kaggle.com/aakash2016/big-mart-sales-prediction>

## 