

REPORT OF EDA ON BIG MART SALES DATA

The Big Mart sales dataset has a variety of information. The data is of 2013 and was collected from 10 different outlets. The dataset has 8253 rows and after feature engineering, we have 16 columns. To break down the analysis, I have divided it into 3 major sections:

1. Tiers
2. Outlets
3. Items

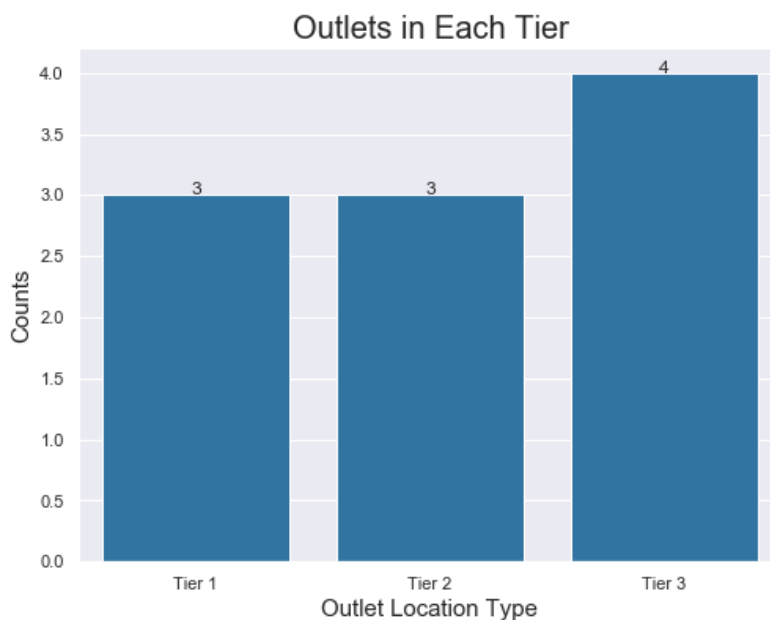
Tiers

There are 3 tiers:

- **Tier 1**
- **Tier 2**
- **Tier 3**

Outlets in each Tier

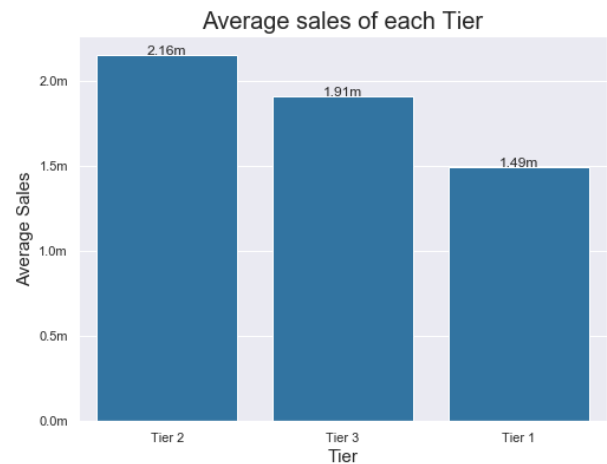
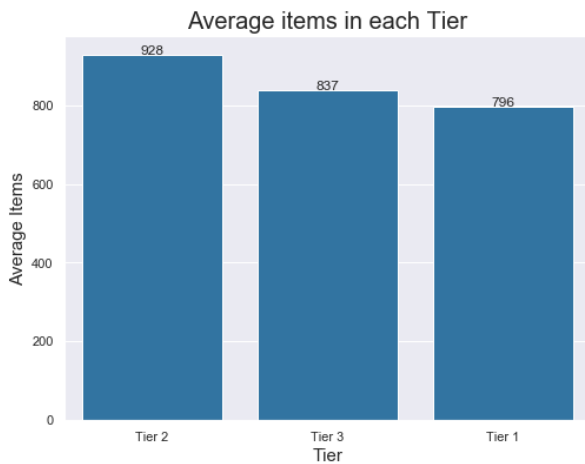
The 1st question which arises is whether the outlets are evenly divided or any tier has dominance in outlets counts. The chart shows that each tier has the same number of outlets except the 3rd Tier which has one more outlet than the others.



Total items and sales of each tier

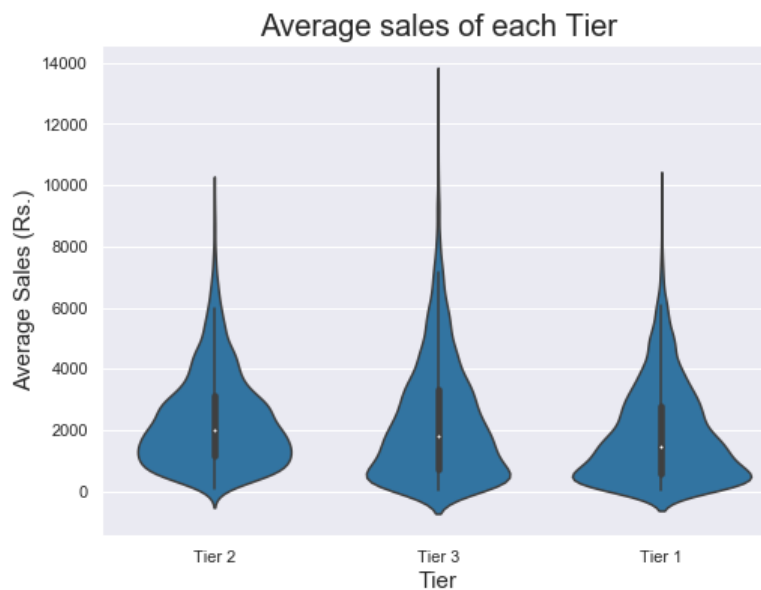
The most important aspect is how well any tier has made in sales. As Tier 3 has the advantage of an extra outlet, we will view the average sales of each tier.

1. Tier 2 had the highest sales of 2.16 million
2. Tier 3 had 1.91 million sales, it had 250K fewer sales than Tier 2.
3. Tier 1 had the least sales of 1.49 million



Although Tier 3 didn't have the highest sales, it still shows the potential to have better sales. Through this violin plot we can see that:

1. Tier 1 and Tier 2 had a threshold at around 11K.
2. Tier 3 almost touched the 14K mark.



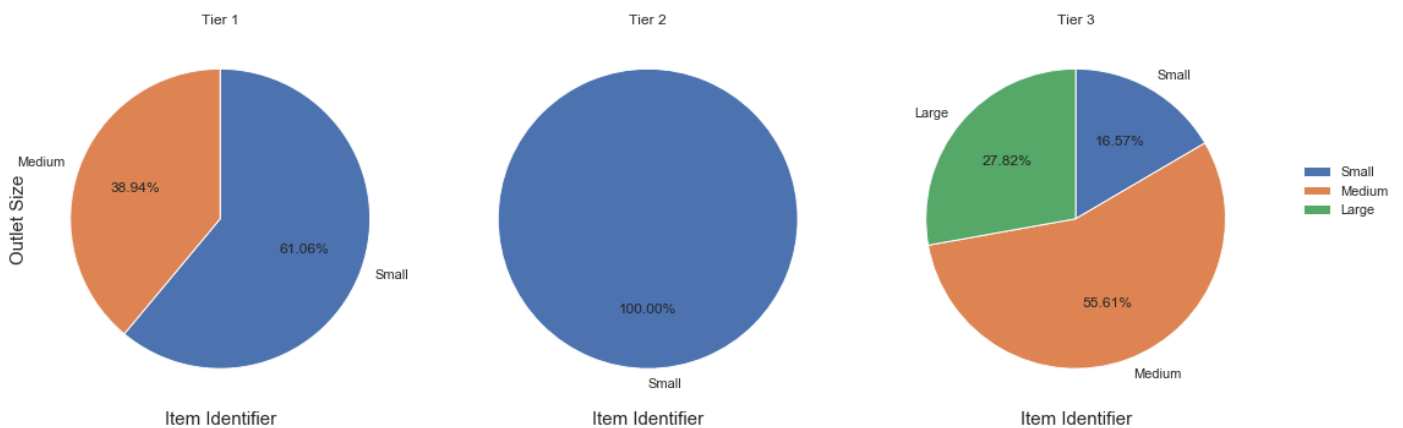
Size of outlets in each Tier

To further investigate the success of Tier 2, we will try to see if Tier 2 has Large outlets which made them achieve such big sales numbers or they tried to manage even with Small and Medium-sized outlets.

We have the following sized outlets:

- **Small**
- **Medium**
- **Large**

Distribution of outlet sizes in each Tier



From the pie chart above, we can that:

- Tier 2 only had Small-sized outlets which seems amazing.
- Tier 3 is the only tier with all sized outlets.
- Tier 1 didn't have Large outlets.

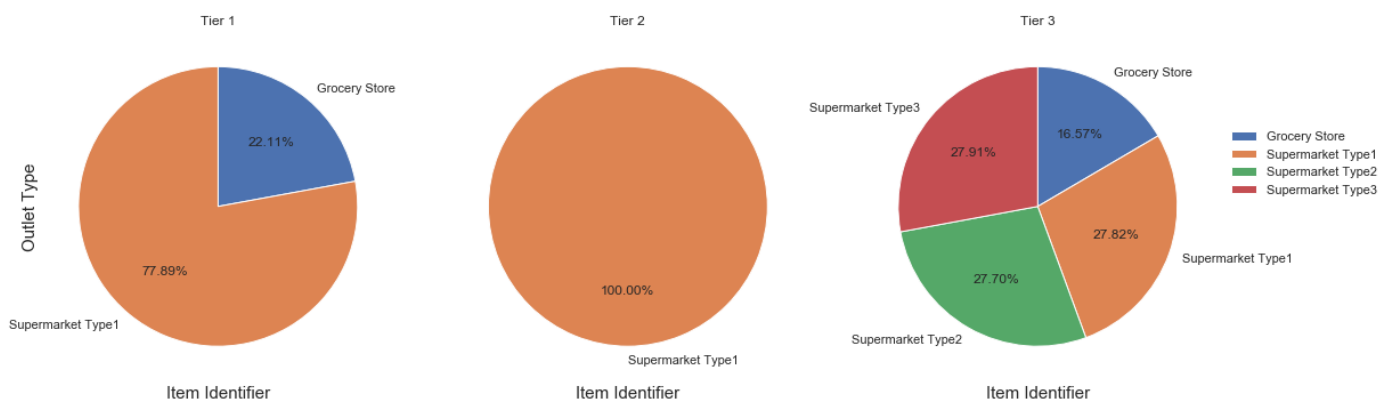
Types of outlets in each Tier

As we have seen that Tier 2 only have small-sized outlets, it is interesting to know what types of outlet they had.

We have the following types of outlets:

1. **Grocery Store**
2. **Supermarket Type 1**
3. **Supermarket Type 2**
4. **Supermarket Type 3**

Distribution of outlet types in each Tier



From the chart above we can see that:

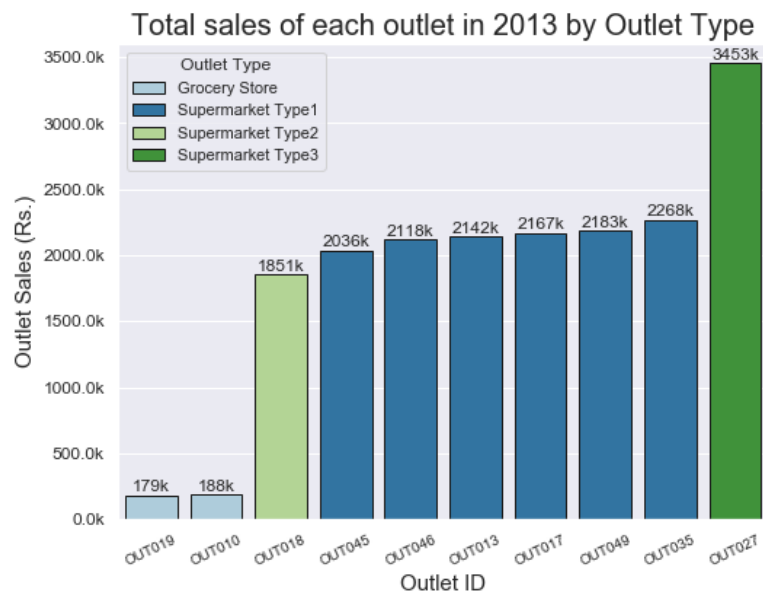
1. Tier 2 only had Supermarket Type 1 outlets.
2. Again Tier 3 had all types of outlets from Grocery stores to Supermarket type 3 and the sales of Supermarket is evenly distributed (around 27 %).
3. Tier 1 only had Grocery Stores and Supermarket Type 1 outlets

Outlets

Highlighting the important finding of each outlet

Sales of each outlet

Now that we can see which tier has shown the best results, its time to see which outlet has shown the best sales result.



Sales by outlet type in descending order:

Supermarket Type 3 > Supermarket Type 1 > Supermarket Type 2 > Grocery Store

- Outlet 19 and 10
 - Both contributed around 1 percent each
 - Both are Grocery Stores
- Outlet 18
 - The only Supermarket Type 2 type outlet
 - Contributed 10% in overall sales in 2013
- Outlet 45, 46, 13, 17, 49, 35
 - Had 11-12% contribution
 - All outlets are Supermarket Type 1
- Outlet 27
 - Had the highest contribution of 18%
 - Is of Supermarket Type 3

Item Prices in each outlet

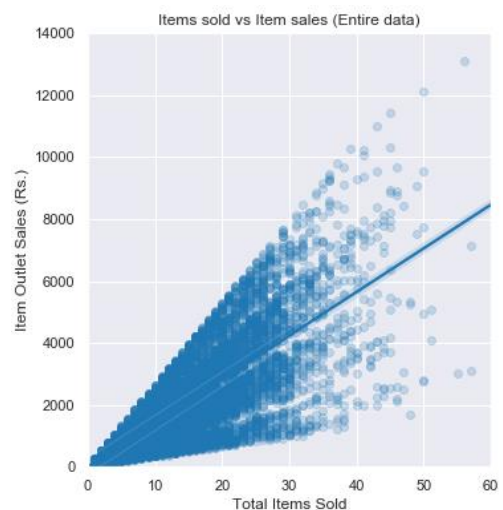
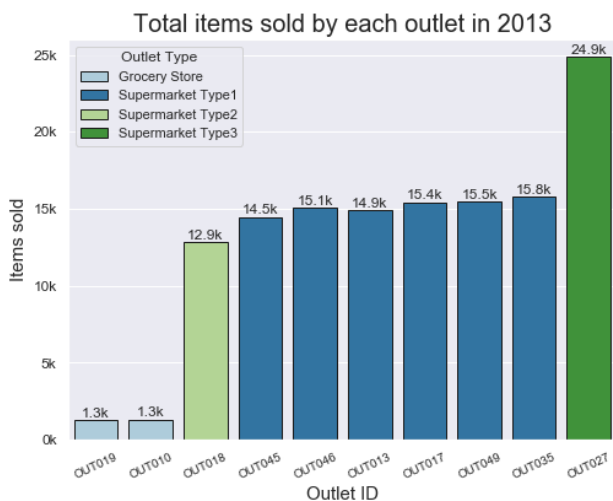
As Outlet 27 had the best sales, it may be true that they have higher-priced items than the rest of the outlets. From the below box plots we can see that outlets didn't have a huge price difference. Each outlet has shown strong competition with its pricing range.



Total items sold by each outlet

As the items price range didn't have a huge difference. We can see that the sales are proportional to the total items sold.

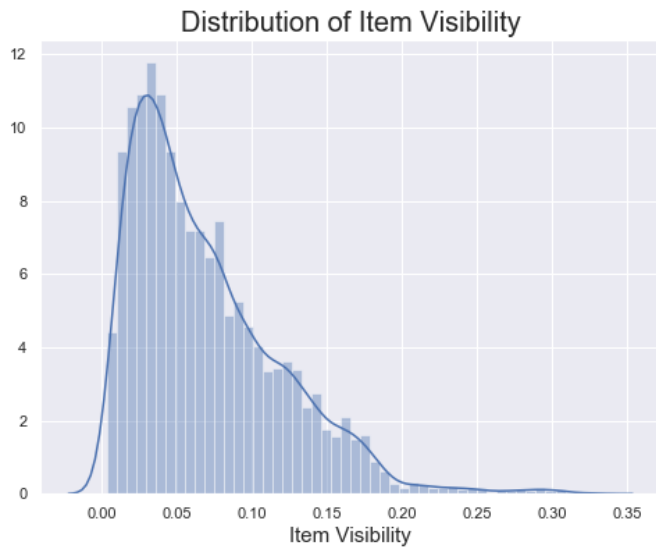
The scatter plot shows the visible relationship between the two. Total items sold and total sales showed a correlation of 0.76.



Items

Insights on items sold by outlets.

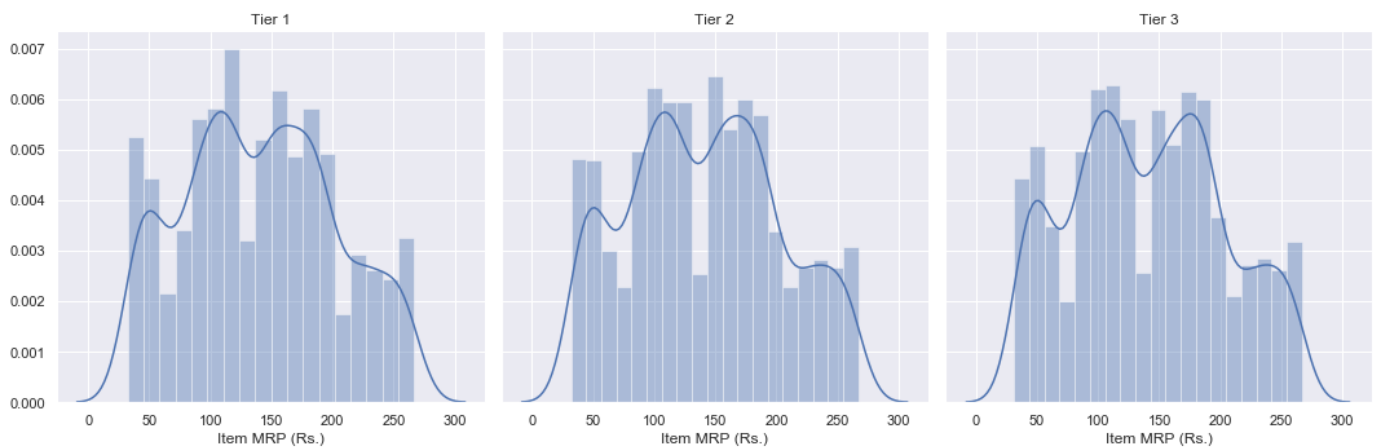
Item Visibility



- 75% of items are below 0.09 % visibility
- Only 1.5% of items have above 0.2% visibility.

Item MRP

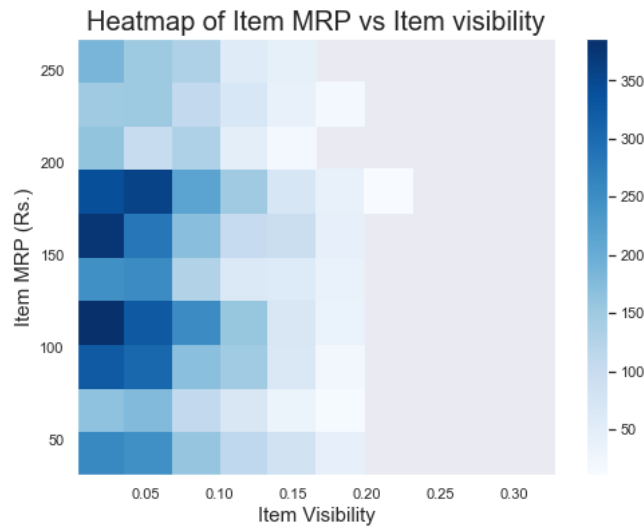
The distribution of items MRP is almost similar in each Tier. They don't have a significant difference.



The relation between Item MRP and Item Visibility

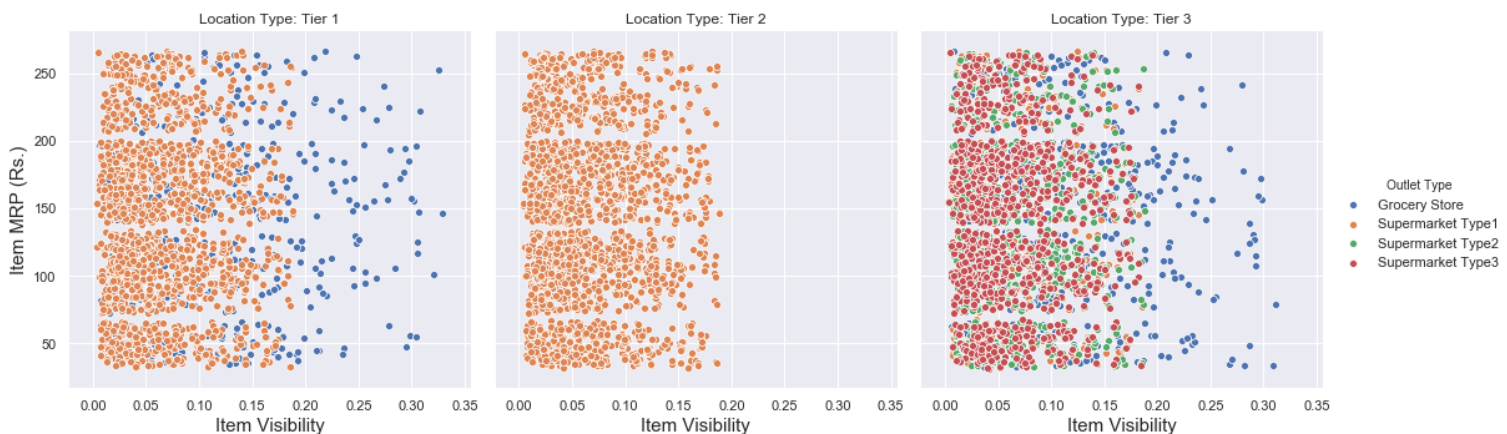
As MRP is almost similar, we can try to find a relation between item MRP and item visibility and the below heatmap shows that:

1. Items within the range of 100 to 200 are having visibility below 0.10.
2. Even the items with the highest price range are mostly below 0.20 visibility.



Deeper insights on the relation between item MRP and item sales

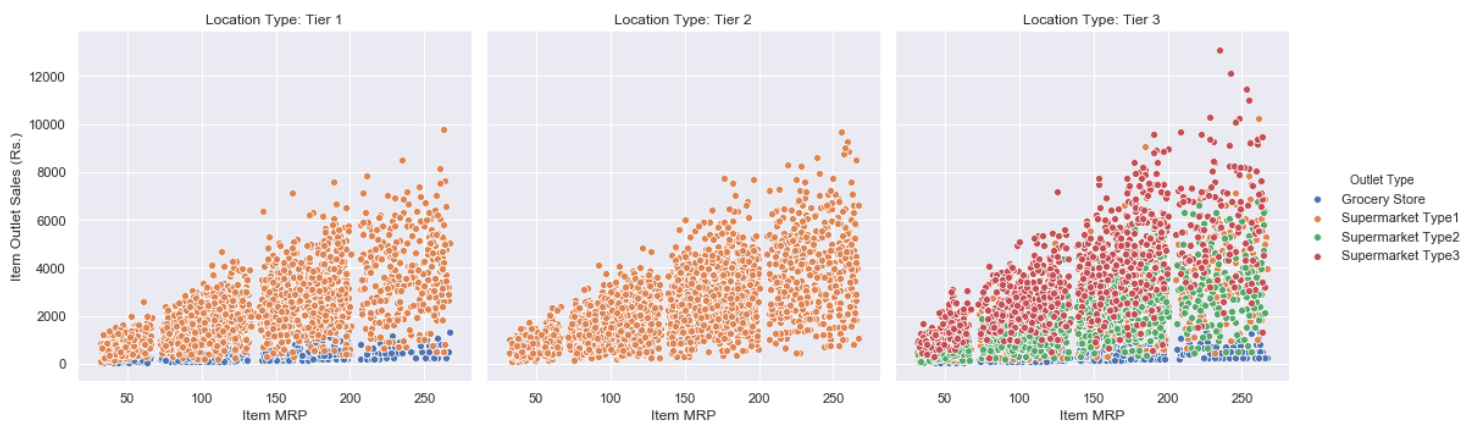
From the below scatter plots, we can see that they do have a relation with Outlet Types.



1. Supermarkets have item visibility range from 0.00 to 0.20. None of the products crosses a visibility mark of 0.20.
2. Only Grocery Stores have item visibility range from 0.00 to 0.33. They are the only outlets that have item visibility above 0.20.

Finding more relations of item MRP and outlet sales with outlet location type

As we have seen earlier that the type of outlets does have a relation with the total number of items they have. We can try to find if item sales and MRP have a relation with outlet type.

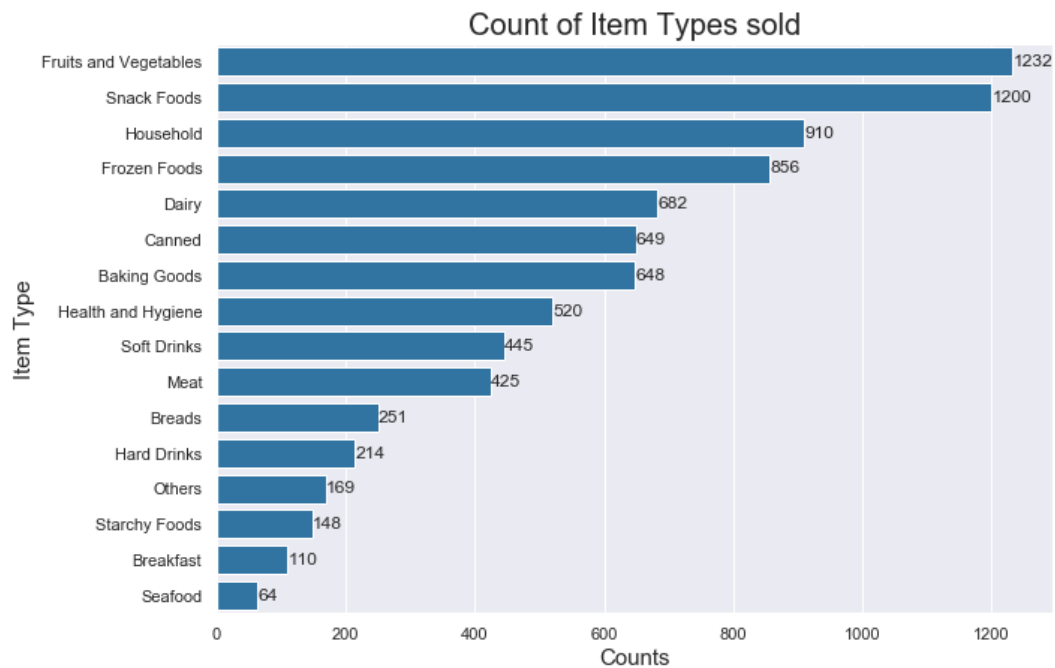


From the above chart, we can see that the sales have shown a certain threshold in each outlet type. The table below shows the highest item sales achieved by each outlet type.

Outlet Type	Sales Threshold (Rs)
Grocery Store	1700 (Only 1 item above it, 1775)
Supermarket Type 1	10000 (Only 1 item above it, 10256)
Supermarket Type 2	6800
Supermarket Type 3	13000 (Only 1 item above it, 13086)

Item types sold in outlets

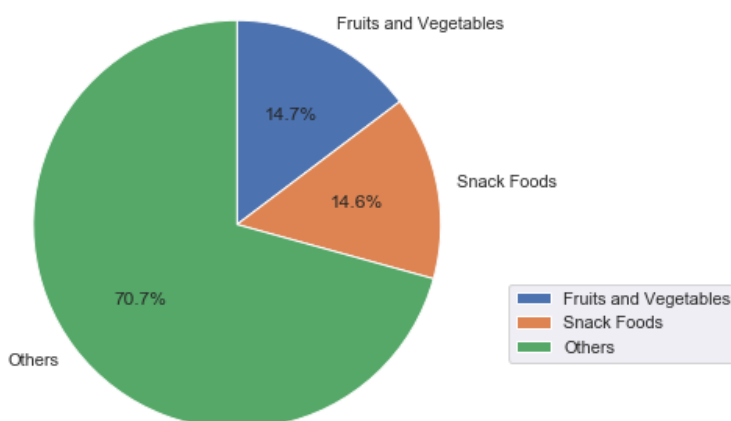
Big Mart had 15 items types and an additional category named Others.



1. Fruits and Vegetables are the most sold items (1232) followed by Snack Foods (1200).
2. Dairy items sold were half to Fruits and Vegetables (682)
3. Seafood is the least sold item.
4. People were not interested in Breakfast items too.

Contribution of Fruits and Vegetables and Snack Food

Sales comparison of item types



Fruits and Vegetables and Snack Foods generated 28% sales.

Each of them contributed 14.6%.

Investing more in them will surely be beneficial in future.

Conclusion

1. The item sales were highly dependent on total items sold
2. Tier 2 was able to generate the most sales. It had very selected types of outlets such as all of the outlets were of small-sized and were located in Supermarket Type 1.
3. Outlet 27 had the highest sales in comparison to other outlets. It is the only outlet which is located in the Supermarket Type 3 location and is of Medium size.
4. Item in the Supermarket don't have visibility of more than 0.20%
5. Item sales do show a threshold depending upon the type of outlet.
6. Food and Vegetables and Food Snacks are the most promising items for generating more sales.