
Analytics- Project 1

Sales Prediction at BigMart

By

Jazeem Abdul Jaleel and Aparajita Kar

<https://datahack.analyticsvidhya.com/contest/practice-problem-big-mart-sales-iii/>

Problem Description

- Given: Incomplete sales data for year 2013 in 10 stores (Train Data: 8523)

Item ID	Item Weight	Item Fat Content	Item Visibility
Item Type	Item MRP	Outlet ID	Outlet Est Year
Outlet Size	Outlet Location Type	Outlet Type	Item Outlet Sales

- Objective: Predict the sales amount for certain items in the different stores (Test Data: 5681)
- Problem Significance/ Importance

Project is divided into four phases :

1. Data Cleaning
2. Data Visualization
3. Prediction Modeling
4. Results and Conclusion

1. Data Cleaning

We encountered data cleaning tasks for 4 variables

- ✓ Item Fat Content

```
Levels of tem_Fat_Content  
[1] "LF"    "low fat" "Low Fat" "reg"    "Regular"
```

- ✓ Item Weights - NA 1463 values

We notice that there exist 1-1 mapping of Item_Id to Item_weight
4 items no weight in file. Calculated Avg in category

- Item Visibility

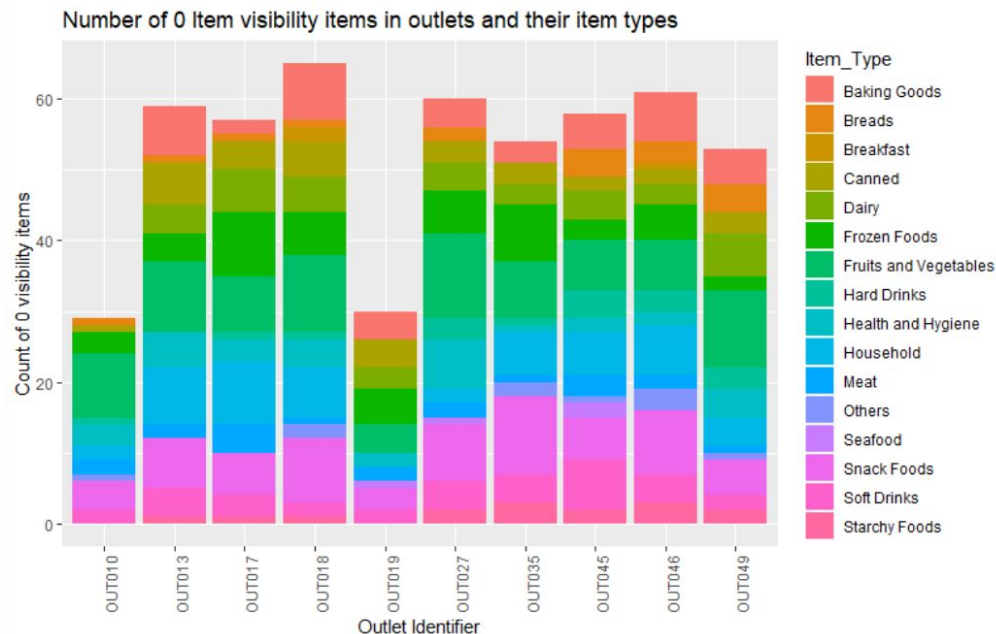
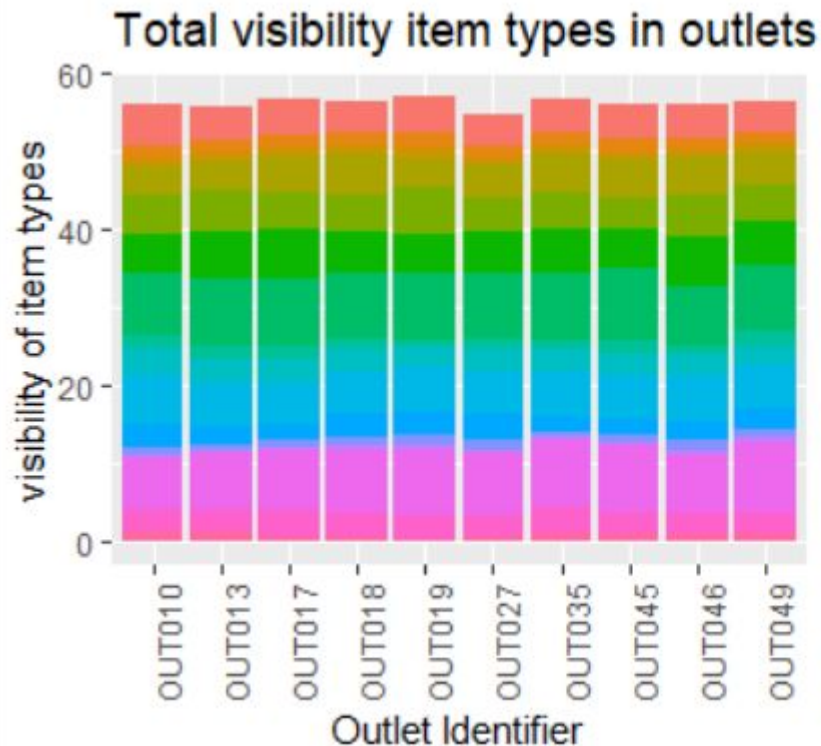
```
sum(Data$Item_Visibility==0)  
[1] 526
```

- Outlet Size

```
Levels of Outlet_Size  
[1] ""      "High"  "Medium" "Small"
```

Data Cleaning contd.

- Item Visibility - non zero values existed in other Outlet records



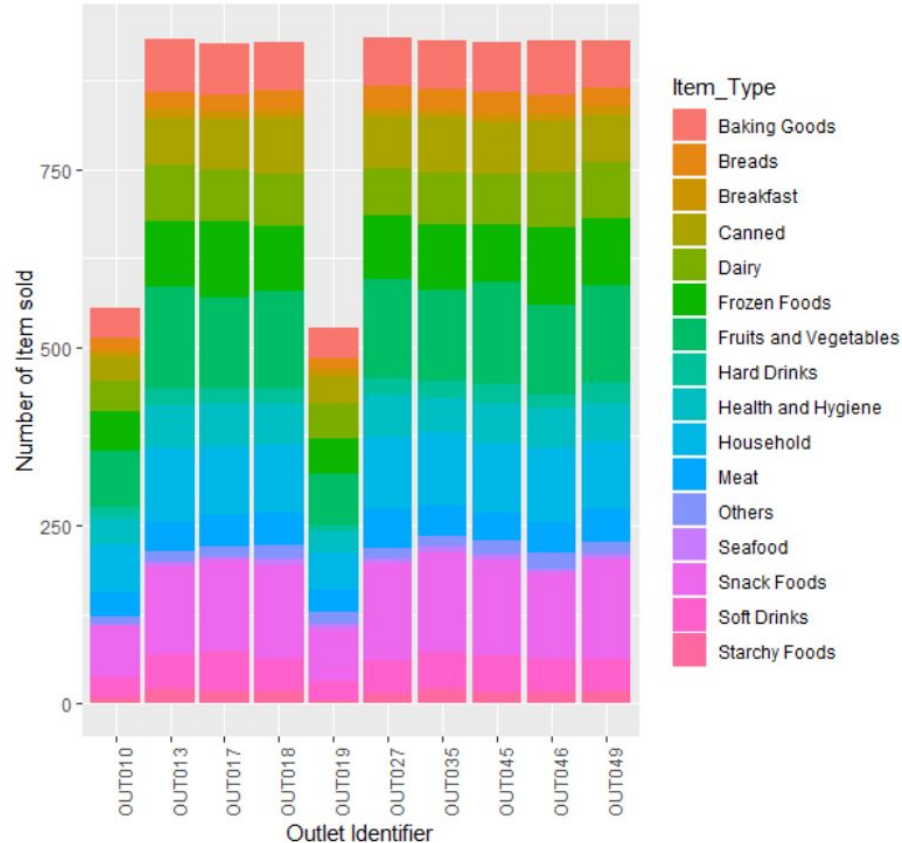
Data Cleaning contd.

- Outlet_Size

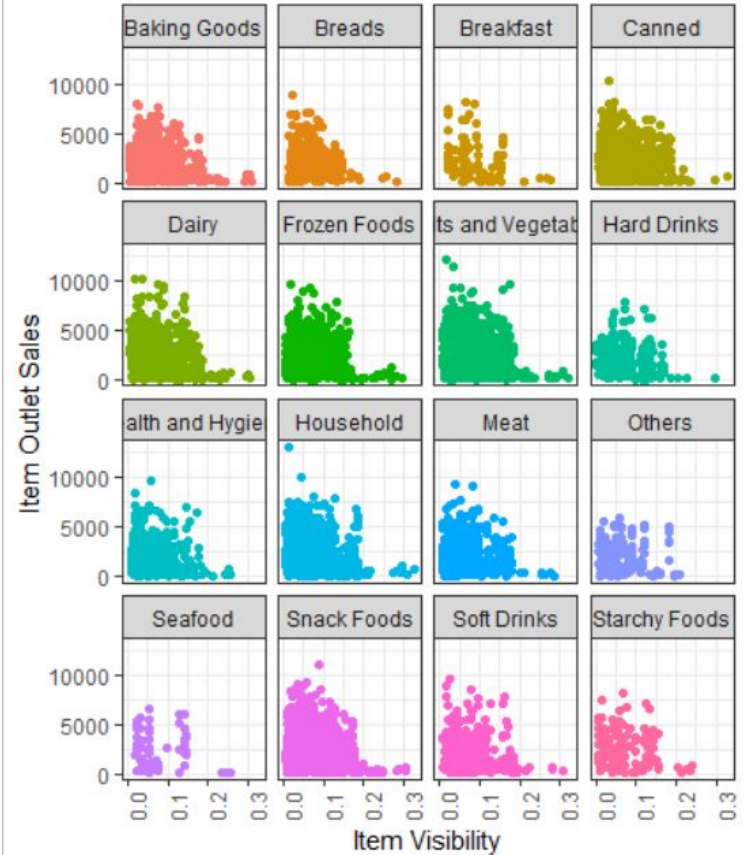
	Outlet_Identifier	Outlet_Type	Outlet_Location_Type	Outlet_Size	total_sales	total_revenue	total_weight	TotalItemVisibility
1	OUT010	Grocery Store	Tier 3	Small	188340.2	32098491	2428398	57.62661
2	OUT013	Supermarket Type1	Tier 3	High	2142663.6	366497258	28103929	59.28537
3	OUT017	Supermarket Type1	Tier 2	Small	2167465.3	364132175	28054721	59.90698
4	OUT018	Supermarket Type2	Tier 3	Medium	1851822.8	316173229	23868355	60.45533
5	OUT019	Grocery Store	Tier 1	Small	179694.1	29734534	2338112	58.54377
6	OUT027	Supermarket Type3	Tier 3	Medium	3453926.1	573386176	44897735	58.11823
7	OUT035	Supermarket Type1	Tier 2	Small	2268122.9	385657664	29266088	59.95604
8	OUT045	Supermarket Type1	Tier 2	Small	2036725.5	343240628	25913962	59.40270
9	OUT046	Supermarket Type1	Tier 1	Small	2118395.2	357320464	27163541	60.12938
10	OUT049	Supermarket Type1	Tier 1	Medium	2183969.8	366999485	28229162	60.37248

Data Visualization

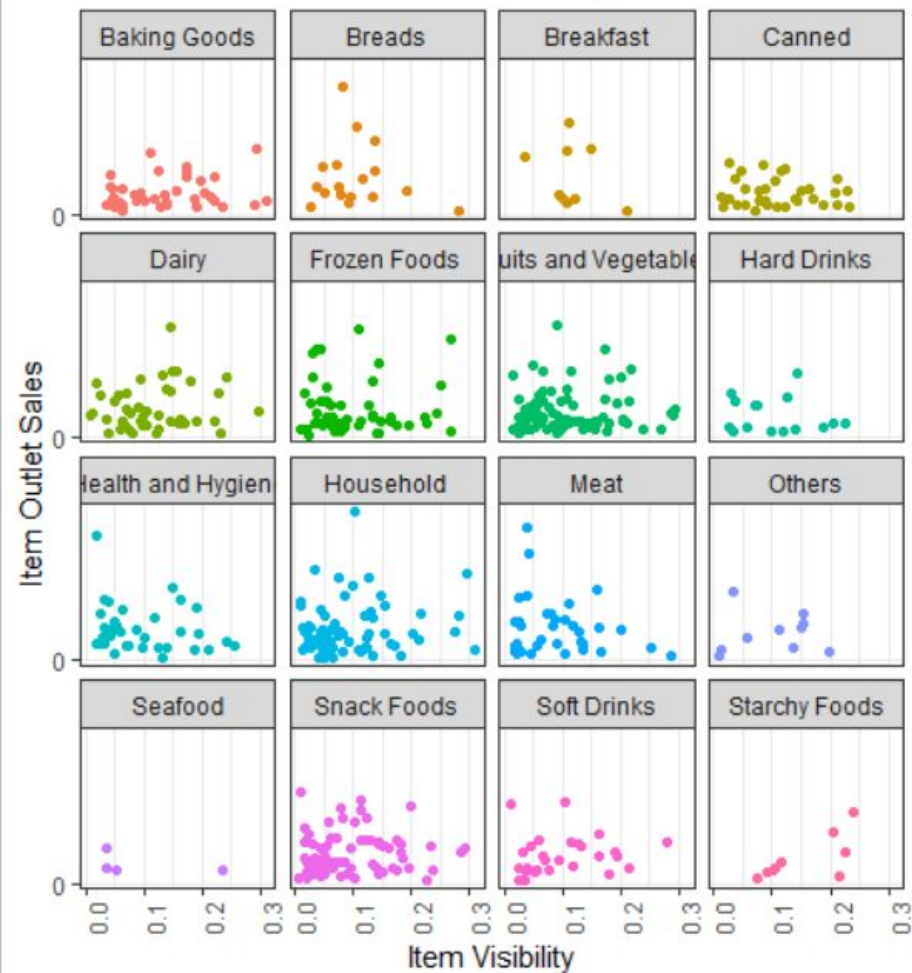
Type of items sold in different Outlet



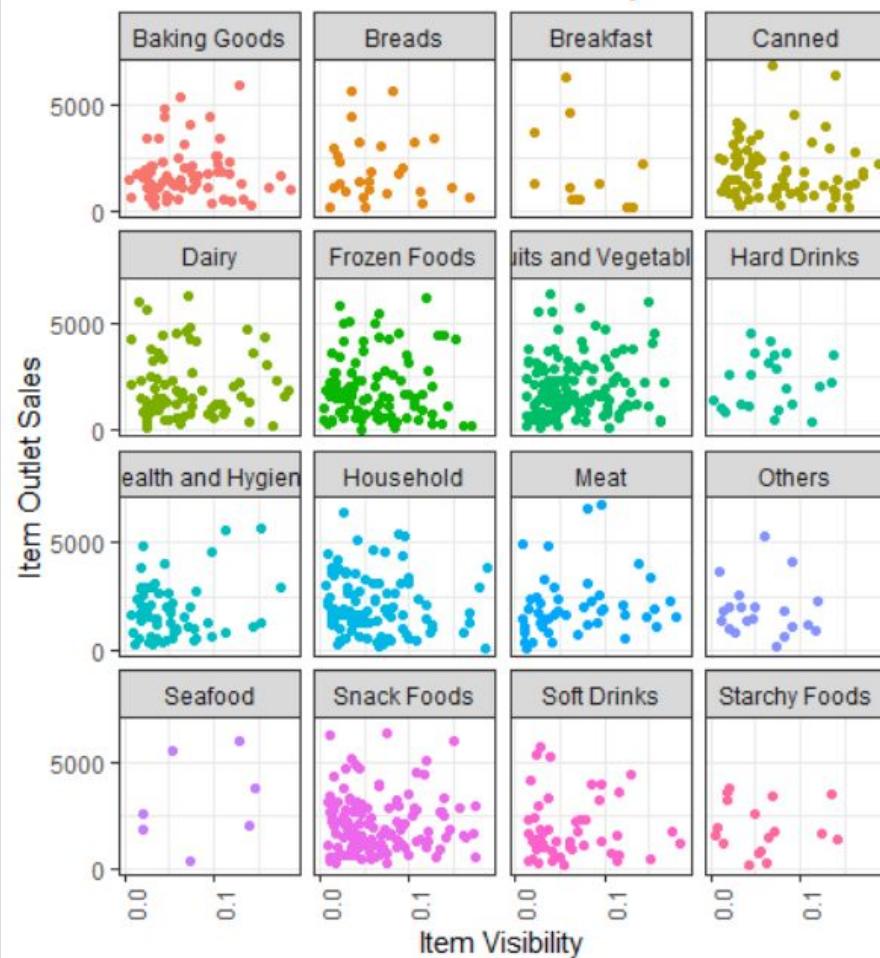
Item Outlet sales v/s Item Visibility



Item Outlet sales v/s Item Visibility - OUT010



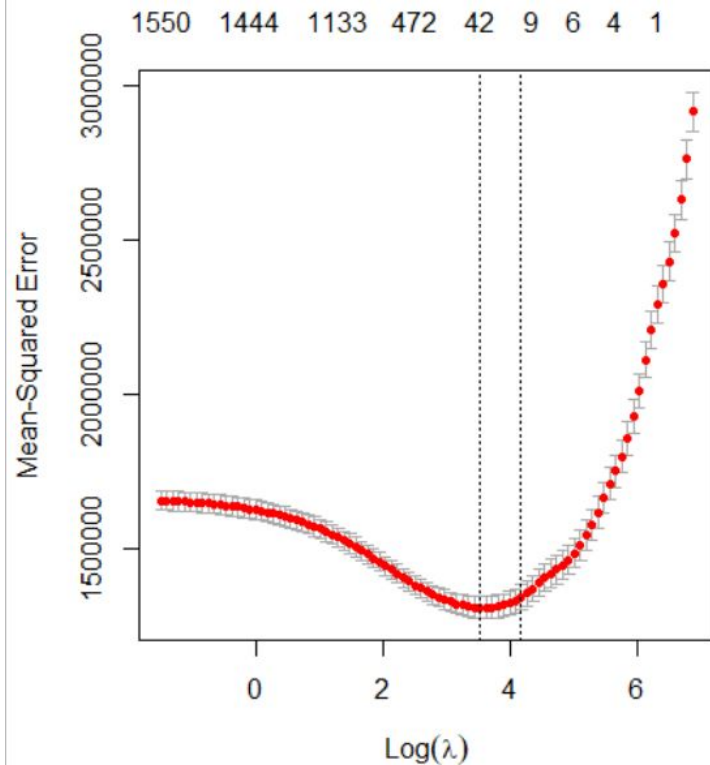
Item Outlet sales v/s Item Visibility - OUT018



Prediction Modeling

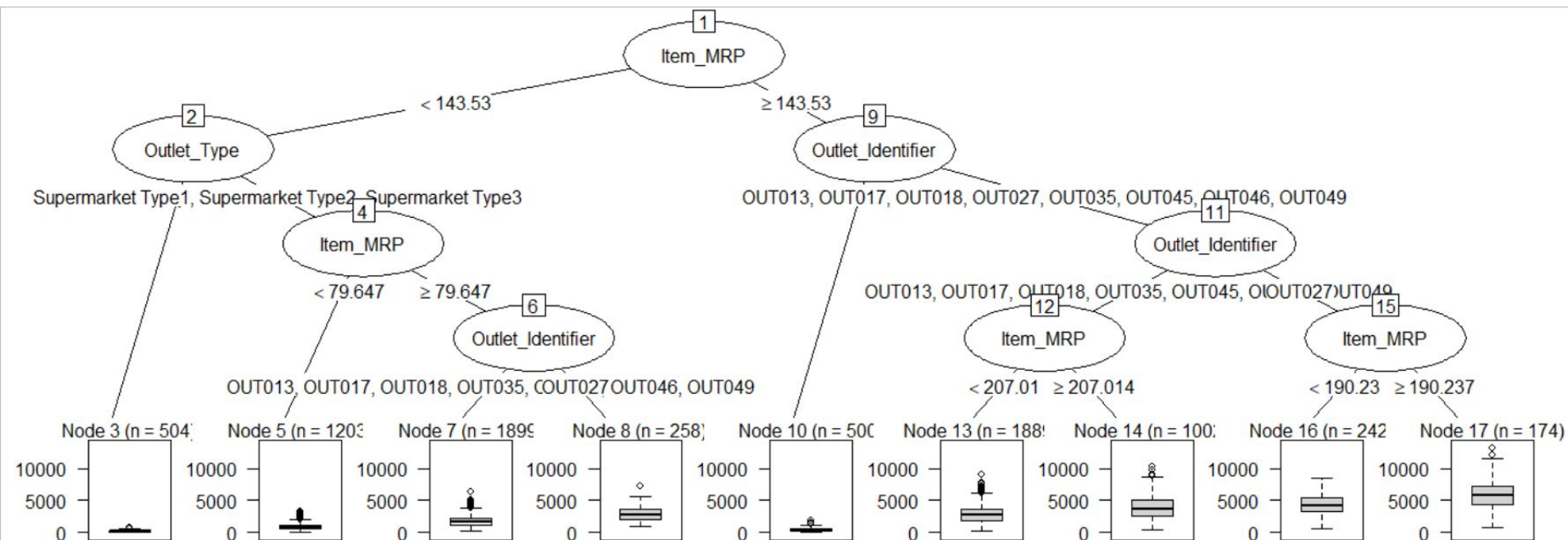
<u>Method</u>	<u>RMSE</u>
Linear Model (Log)	1246 (1245)
LASSO (Log)	1144 (1142)
Ridge (Log)	1229 (1227)
Decision Tree	1148

Lasso



	Lambda	Measure	SE	Nonzero
min	33.86	1305231	29193	42
1se	59.17	1330539	30560	10

Decision Tree



Interesting...

	Outlet_Identifier	Outlet_Type	Outlet_Location_Type	Outlet_Size	total_sales	total_revenue	total_weight
1	OUT010	Grocery Store	Tier 3	Small	188340.2	32098491	2428398
2	OUT013	Supermarket Type1	Tier 3	High	2142663.6	366497258	28103929
3	OUT017	Supermarket Type1	Tier 2	Small	2167465.3	364132175	28054721
4	OUT018	Supermarket Type2	Tier 3	Medium	1851822.8	316173229	23868355
5	OUT019	Grocery Store	Tier 1	Small	179694.1	29734534	2338112
6	OUT027	Supermarket Type3	Tier 3	Medium	3453926.1	573386176	44897735
7	OUT035	Supermarket Type1	Tier 2	Small	2268122.9	385657664	29266088
8	OUT045	Supermarket Type1	Tier 2	Small	2036725.5	343240628	25913962
9	OUT046	Supermarket Type1	Tier 1	Small	2118395.2	357320464	27163541
10	OUT049	Supermarket Type1	Tier 1	Medium	2183969.8	366999485	28229162

New LASSO

Interactions:

- Item_ID and Outlet_Type,
- Item_MRP and Outlet Type

<u>Method</u>	<u>RMSE</u>
Lasso	1144
New Predictor Lasso	

Results and Conclusion

- The decision tree result follows intuition
- Covered data cleaning, data visualization, and prediction modelling
- Results are comparable to online solutions. Challenges of small dataset
- Further study in interaction effects and feature modeling will be beneficial

References

<https://datahack.analyticsvidhya.com/contest/practice-problem-big-mart-sales-iii/>

Behera, Gopal & Nain, Neeta. (2019). A Comparative Study of Big Mart Sales Prediction.

https://www.researchgate.net/publication/336530068_A_Comparative_Study_of_Big_Mart_Sales_Prediction