Predicting Item Sales

Author: Robert Ramos

Task to Solve

Using the attributes of products and their outlets, we are tasked with analyzing the data to determine which ones contribute to higher sales as well as creating a machine learning model to predict how to create more sales.

The Data

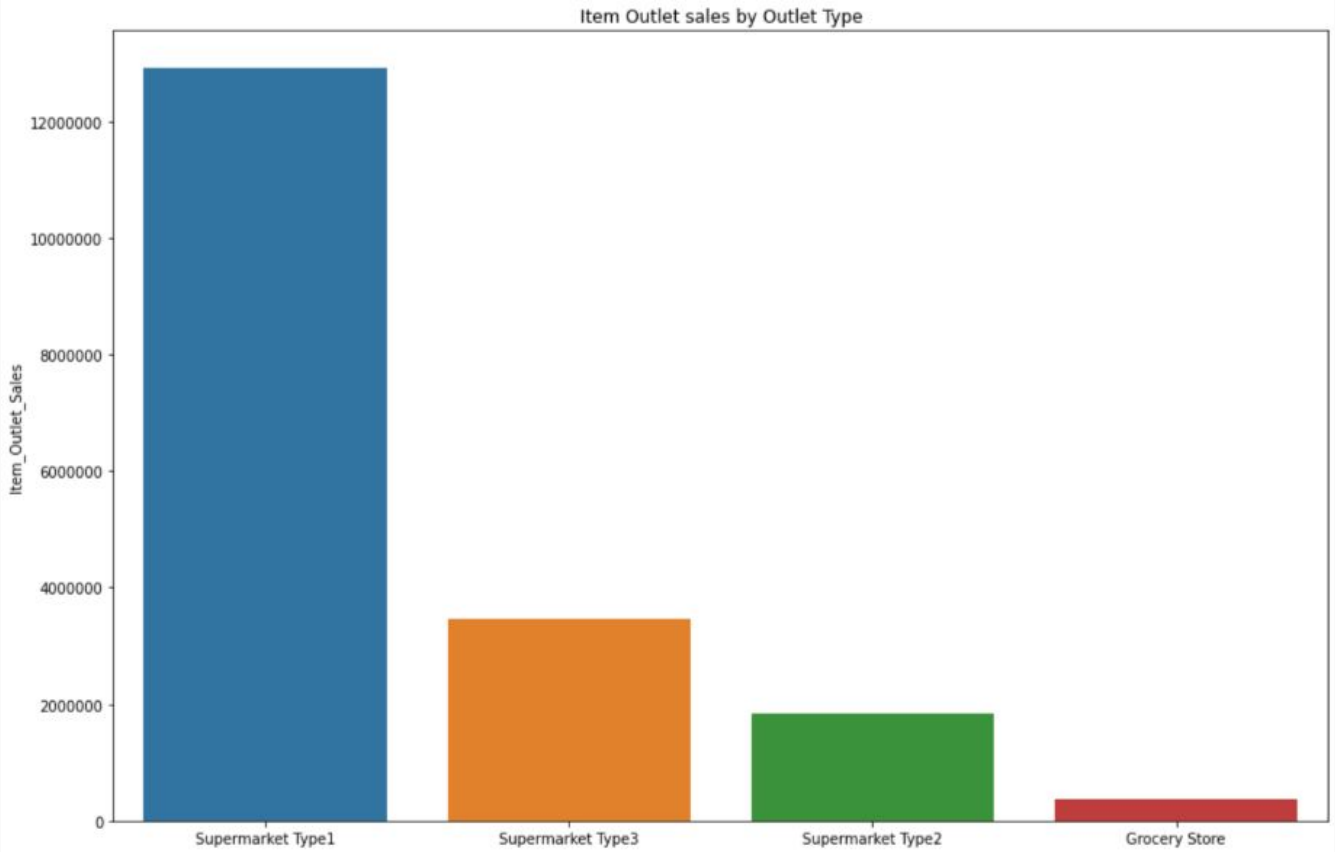
Item_Identifier	1555
Item_Weight	415
Item_Fat_Content	5
Item_Visibility	6519
Item_Type	16
Item_MRP	5225
Outlet_Identifier	8
Outlet_Establishment_Year	8
Outlet_Size	4
Outlet_Location_Type	3
Outlet_Type	3

Variable Name	Description
Item_Identifier	Unique product ID
Item_Weight	Weight of product
Item_Fat_Content	Whether the product is low fat or regular
Item_Visibility	The percentage of total display area of all products in a - store allocated to the particular product
Item_Type	The category to which the product belongs
Item_MRP	Maximum Retail Price (list price) of the product
Outlet_identifier	Unique store ID
Outlet_Establishment_Year	The year in which store was established
Outlet_Size	The size of the store in terms of ground area covered
Outlet_Location_Type	The type of area in which the store is located
Outlet_Type	Whether the outlet is a grocery store or some sort of supermarket
Item_Outlet_Sales	Sales of the product in the particular store. This is the target variable to be predicted

The Problem

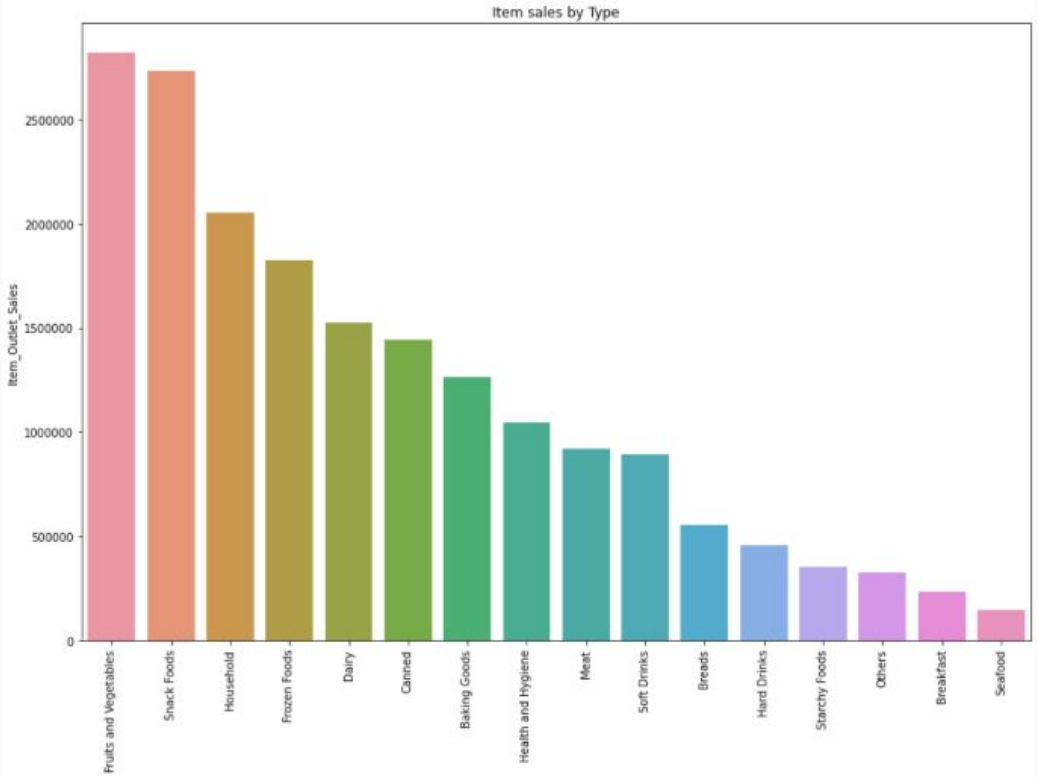
The retailer has collected both product and outlet attribute data for items that they sold but need insight as to what can led to items selling.

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Size	Outlet_Location_Type	Outlet_Type	Item_Outlet_Sales
0	FDA15	9.300	Low Fat	0.016047	Dairy	249.8092	OUT049	1999	Medium	Tier 1	Supermarket Type1	3735.1380
1	DRC01	5.920	Regular	0.019278	Soft Drinks	48.2692	OUT018	2009	Medium	Tier 3	Supermarket Type2	443.4228
2	FDN15	17.500	Low Fat	0.016760	Meat	141.6180	OUT049	1999	Medium	Tier 1	Supermarket Type1	2097.2700
3	FDX07	19.200	Regular	0.000000	Fruits and Vegetables	182.0950	OUT010	1998	Not Provided	Tier 3	Grocery Store	732.3800
4	NCD19	8.930	Low Fat	0.000000	Household	53.8614	OUT013	1987	High	Tier 3	Supermarket Type1	994.7052



- Supermarket Type 1's has the highest volume of sales at over **12.9M** units
- Supermarket Type 3's recorded **3.4M** units sold
- Supermarket Type 2's recorded **1.8M** units sold
- Grocery stores had the least amount of units sold at **368K**

- Top 3 Selling Product Categories
  - Fruits and Vegetables
  - Snack Foods
  - Household
- Bottom 3 Selling Product Categories
  - Others
  - Breakfast
  - Seafood



Model Recommendation

Decision Tree

- Higher Correlation
- More Accurate

```
Linear Regression Model Metrics

Train Evaluation
R^2 0.56105939927833,
RMSE: 1139.747895261957

Test Evaluation
R^2 0.5666580448401798,
RMSE: 1093.4260890154733
```

```
Decision Tree Model Metrics

Train Evaluation
R^2 0.603931621085344,
RMSE: 1082.6572972836627

Test Evaluation
R^2 0.5947201327889511,
RMSE: 1057.4298789236952
```

Thanks!

Contact me:

Robert Ramos

ramos.robertanthony@gmail.com

https://github.com/robertramos89/Predicting-Item-Sales