

# Sales Prediction Project



## PROJECT 1 - FINAL

The goal of this is to help the retailer understand the properties of products and outlets that play crucial roles in predicting sales.





## The 'PROBLEM'

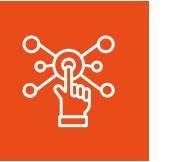
### Big Supermarkets or Small Tiered Stores are more profitable

Sales in the bloodline of any store. With more umbrella cooperations buying out various sized supermarkets, the question rises as to which is the best store sizes and what are the factors that best influence the maximal amounts of sales.

# Method



Import & explore data to determine the information



Exploratory Data Analysis



Data Visualization



Machine Learning modeling



Recommendations

# The Data

The data Source

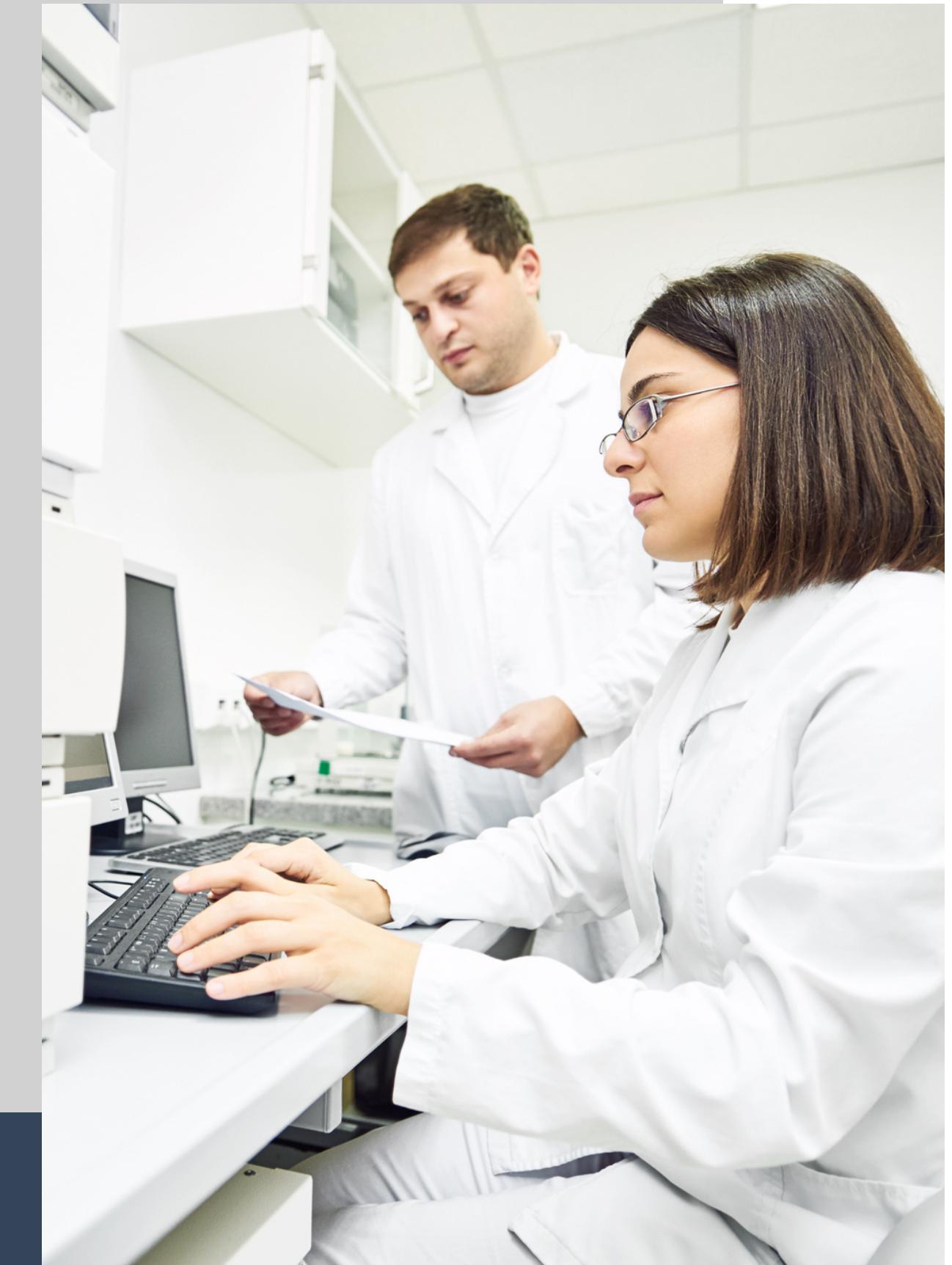


**Big Mart Sales Prediction**

Big Mart Sales Prediction | Practice Problem

 analyticsvidhya.com

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

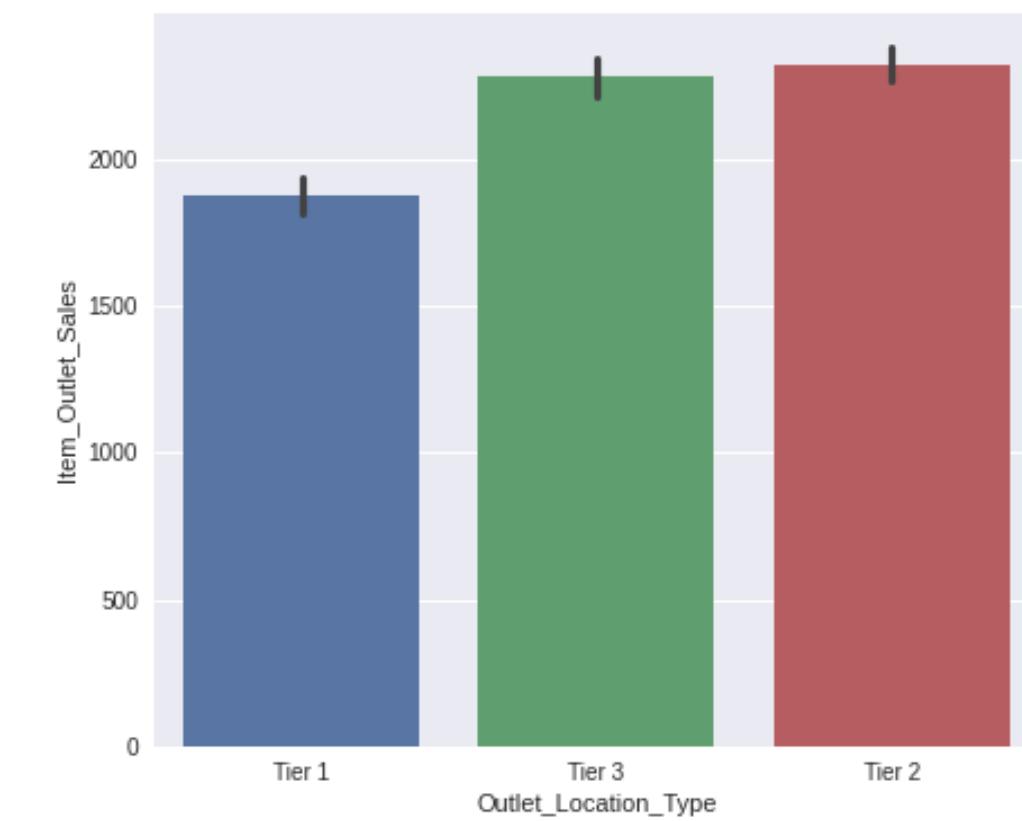
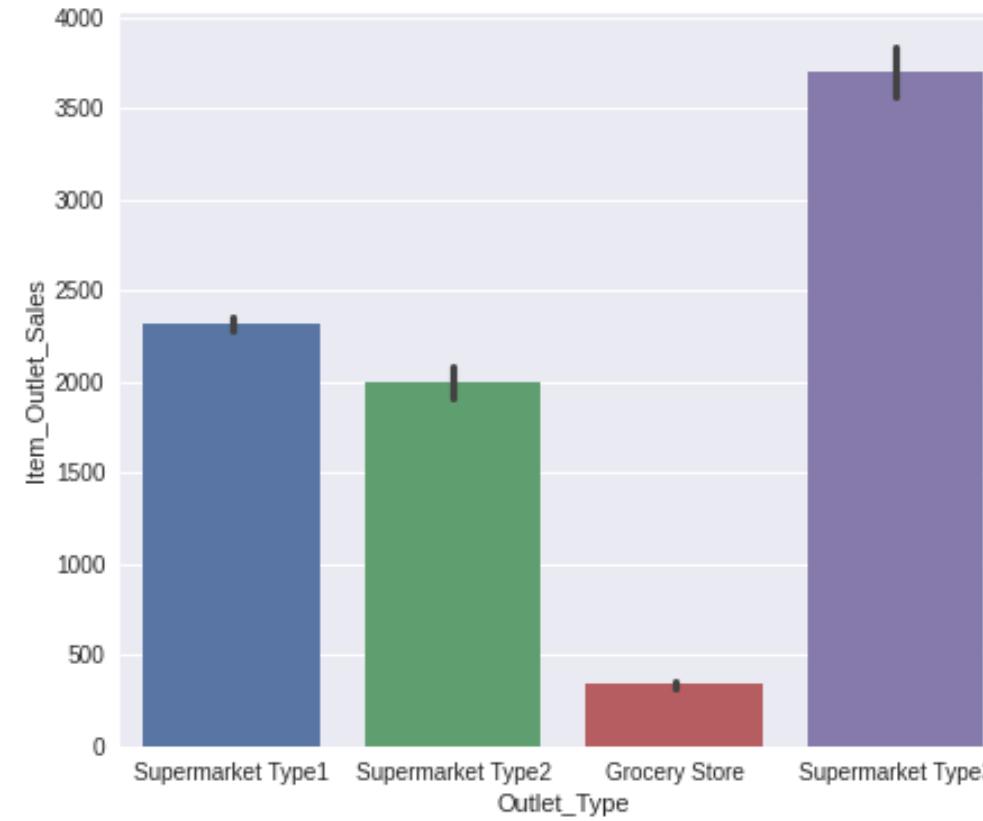
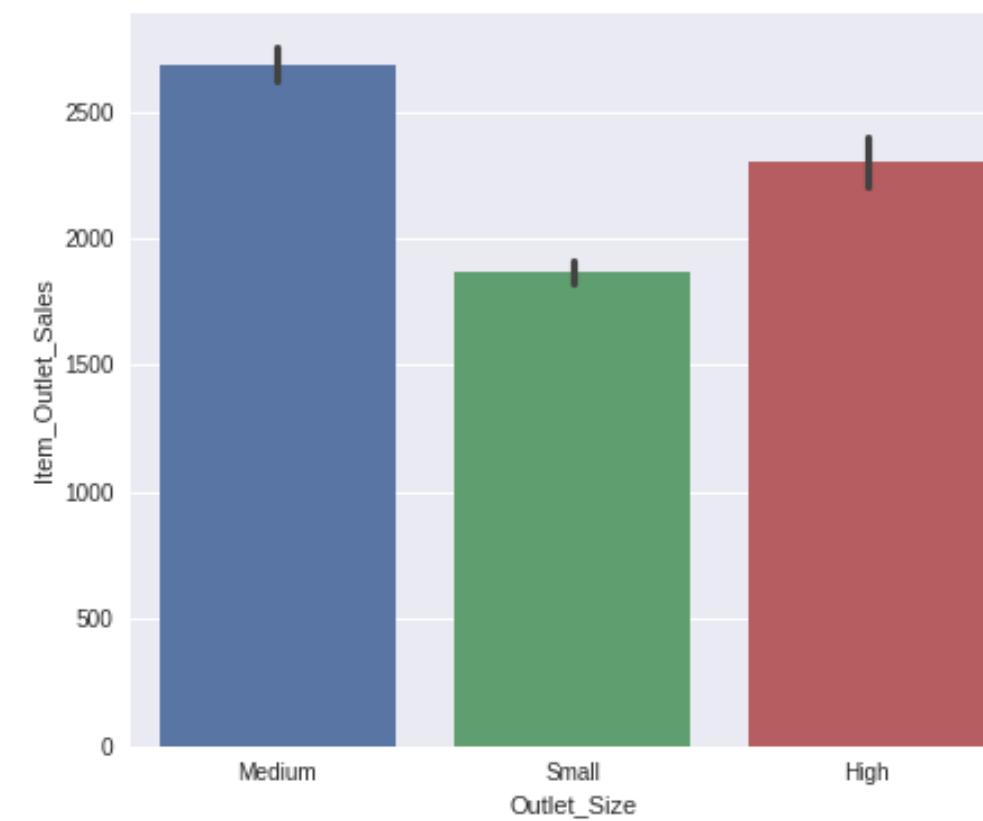


# THE DATA DICTIONARY

Variable	Description
Item_Identifier	Unique product ID
Item_Weight	Weight of product
Item_Fat_Content	Whether the product is low fat or not
Item_Visibility	The % 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 city in which the store is located
Outlet_Type	Whether the outlet is just a grocery store or some sort of supermarket
Item_Outlet_Sales	Sales of the product in the particular store. This is the outcome variable to be predicted.

# Insights #1 - How does overall Sales compare with Supermarket Size, Type and Location ?

Comparing Supermarket Size, Type and location to overall sales



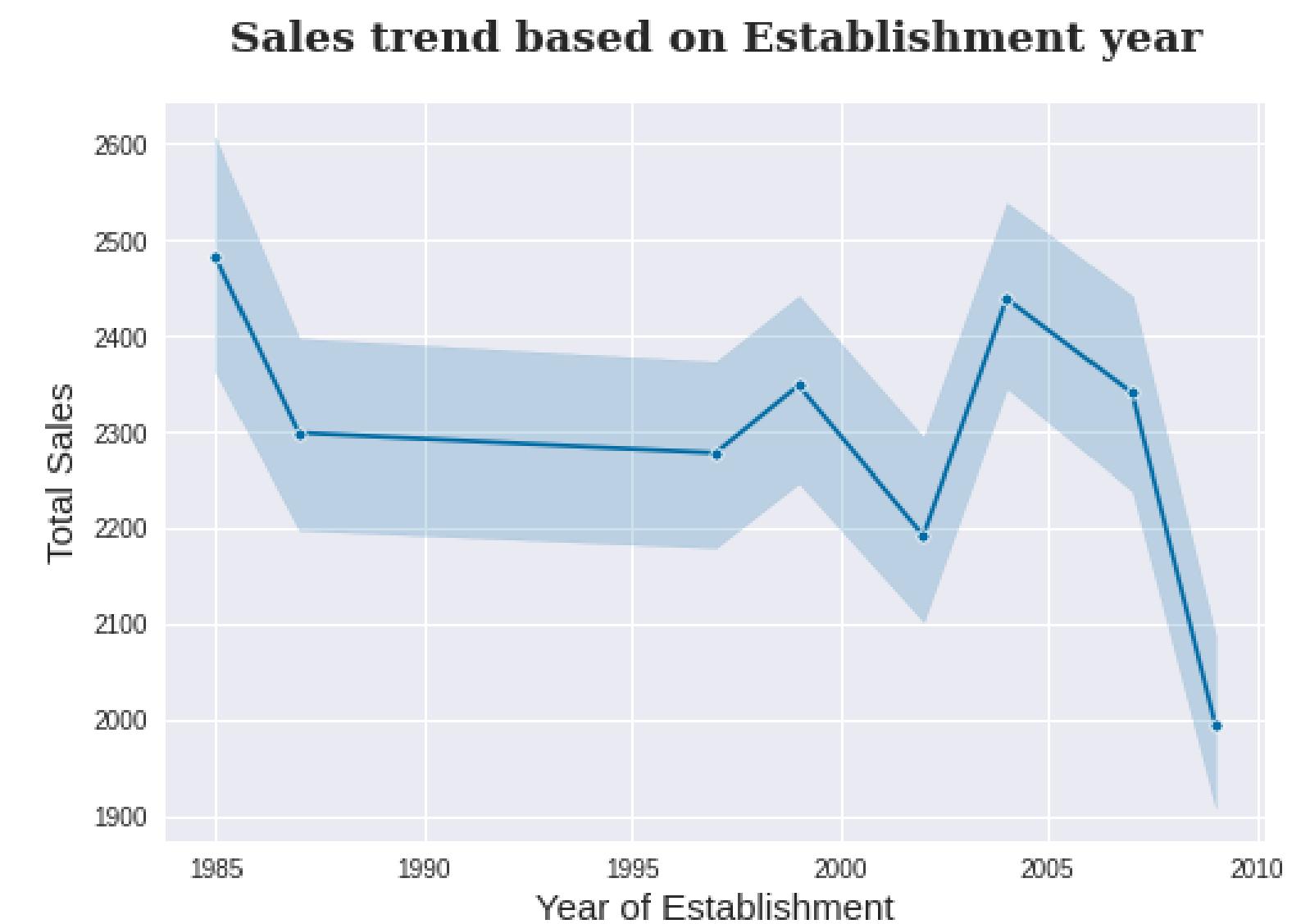
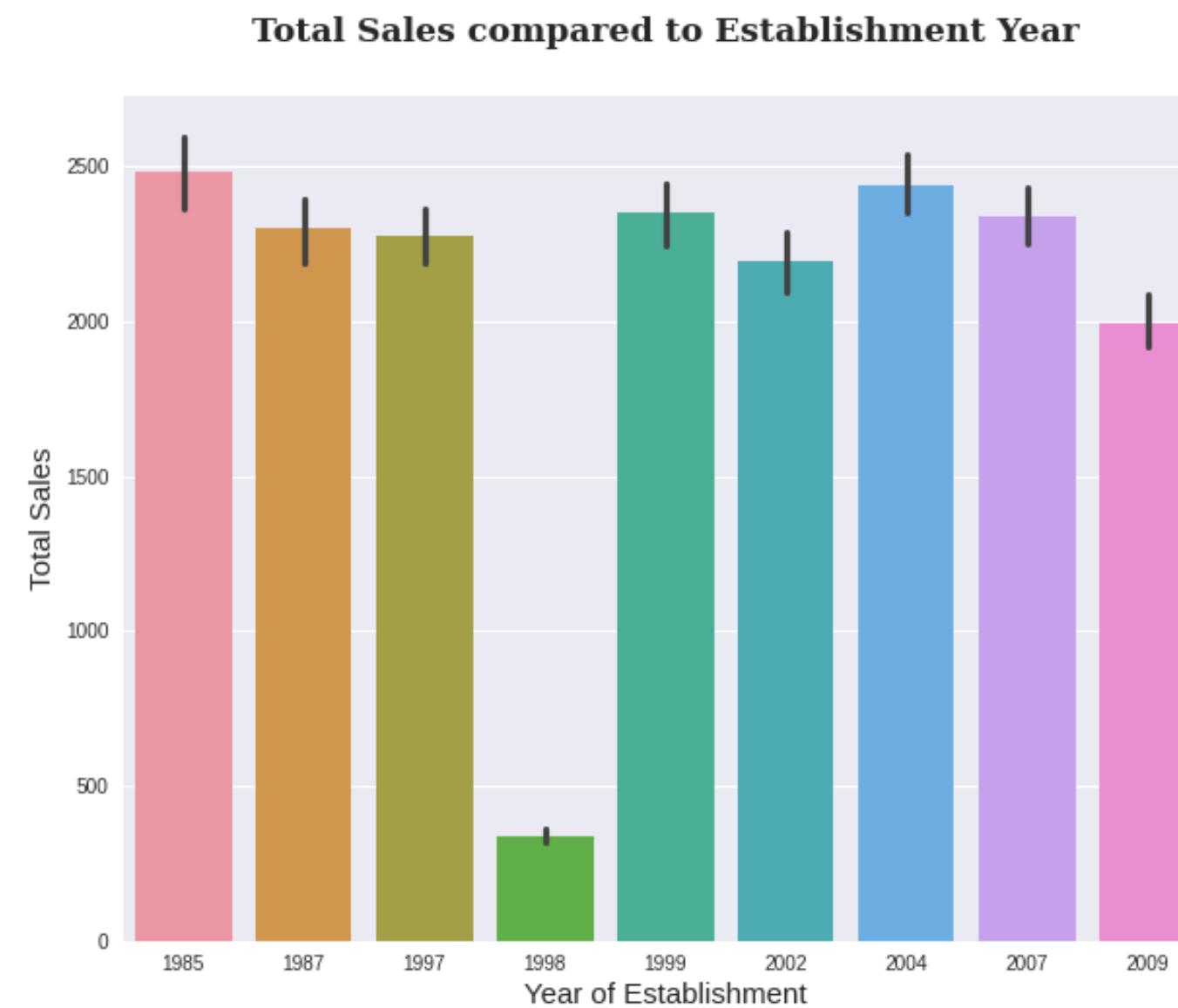
Over all the most favored were the  
**Medium sized Supermarkets**

Outlet Type  
**Supermarket**  
Type 3

Location Type  
**Tier 3 & 2**



# Insights #2 - How does establishment year have an impact on sales?



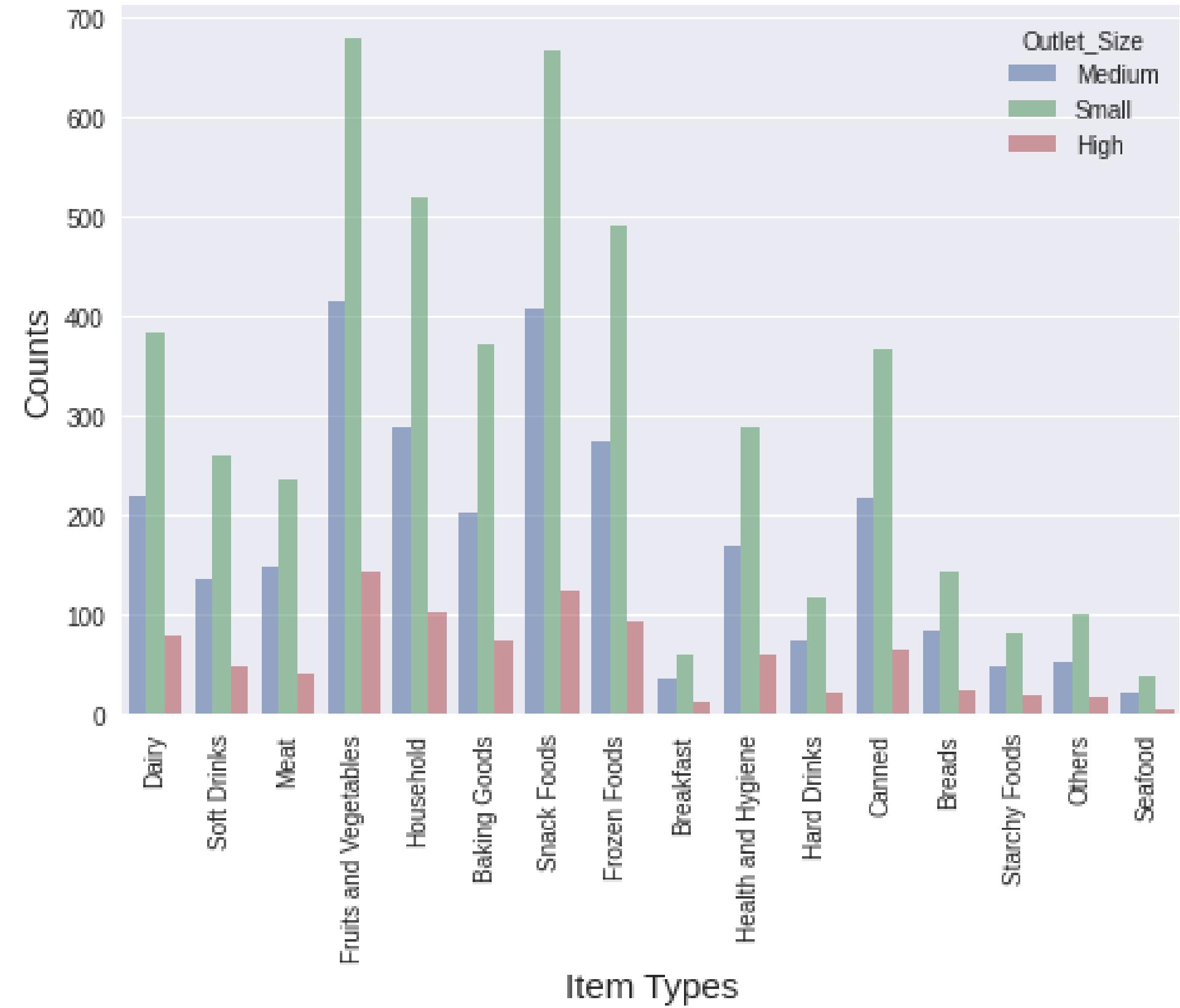
Over all a slight downward trend on sales seen with Establishment years

# Insights #3

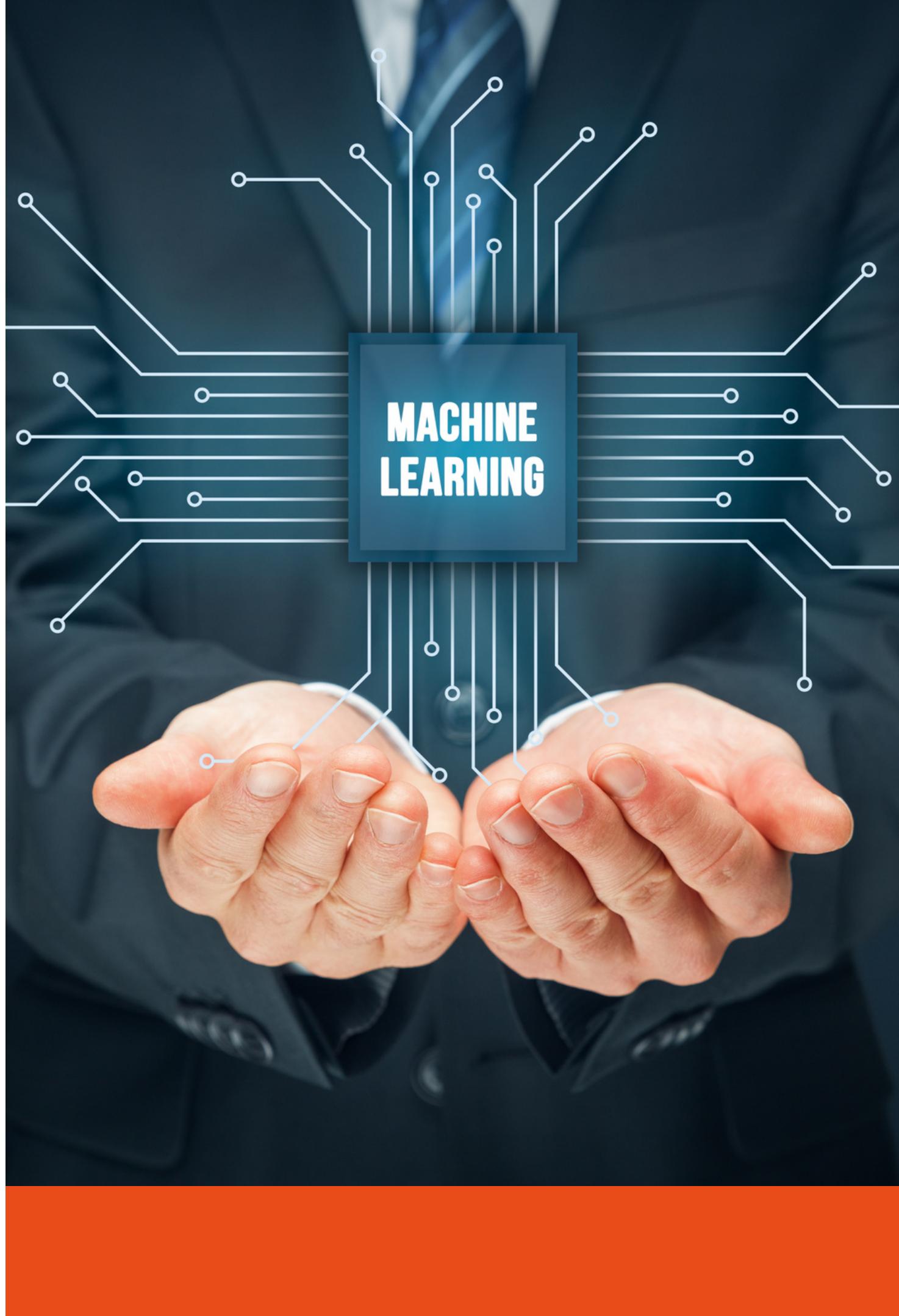
## What are the most selling Items and is there a difference based on Outlet Size ?



Most selling items based on Outlet Size.



Fruits & Vegetables and Snack foods seem to be the top selling items and fairly consistant across the Outlet sizes.



# USING MACHINE LEARNING in PREDICTION

As forecasting sales based on various categories helps companies stay ahead of the game.

Can machine learning tools used to better predict Sales outcomes based on the various features.

Let's explore that next.



# Comparing Machine Learning Models

## Model

### RMSE Value

### R2 Score

### Simplified Explanation

**Training**    **Testing**

**Training**    **Testing**

**RMSE Value -**  
Lesser the better

Baseline

1720.30    1664.97

0.0    -0.004

**R2 Score -**  
Closer to 1 the better & and lesser  
the difference between training  
and testing the better

Linear Regression

1208.78    1168.36

0.506    0.505

Regression Decision  
Tree

1082.58    1057.44

0.604    0.594

Random Forest

1073.58    1047.14

0.610    0.602



**Best Tested  
Model**

# Recommendations

## Insights recommendations

The most preferred type of supermarkets seems to be medium sized markets with a preference towards type 3 Supermarkets. Investing in most of these can be beneficial towards profitability

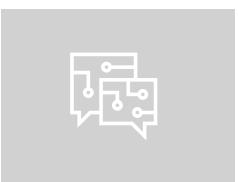
The main selling products remain to be fruits / Vegetables and Snack food.

## Further Recommendations!

With supermarket sales starting to trend downwards with establishment years, retailers can use all the advantage to better forecast/ predict dependent features and associations.

Continued enhancement and use of further machine learning models could assist retailers.

Use further datasets with median income and housing costs to evaluate correlations and preferences of grocery types.





# Thank You

**Looking forward to your feedback !**

The files and analysis for this project can be found in the below GitHub repository.

**Website**

<https://github.com/vinosraj/Project-01-Supermarket-Analysis>

**Contact - Vino Raj**

<https://www.linkedin.com/in/vinosraj/>