

Predictive Sales Data

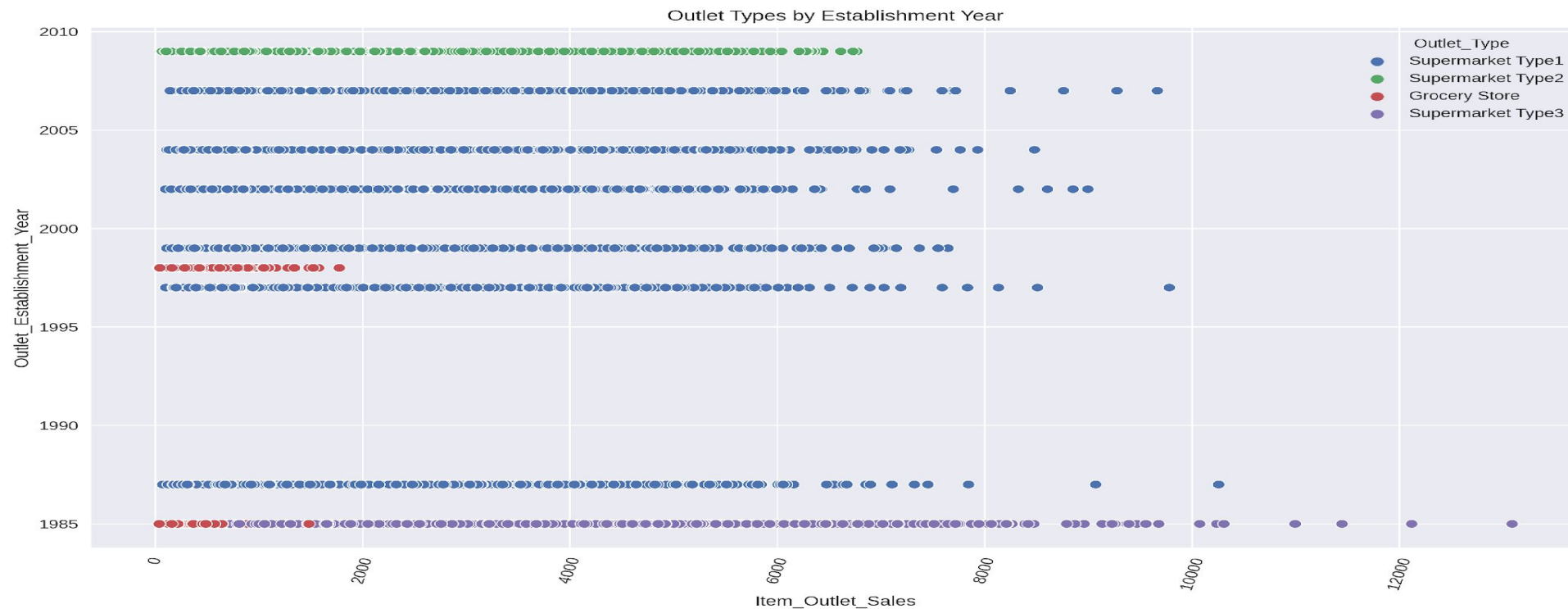


Goals for this presentation

- The task was to review this data to find trends in sales data based on various attributes(There were 12 individual columns we will be using as our features)
- Explore multiple visualization to help shareholders make informative decisions (barplot and boxplot)
- Relay these findings and give insights for areas of improvement



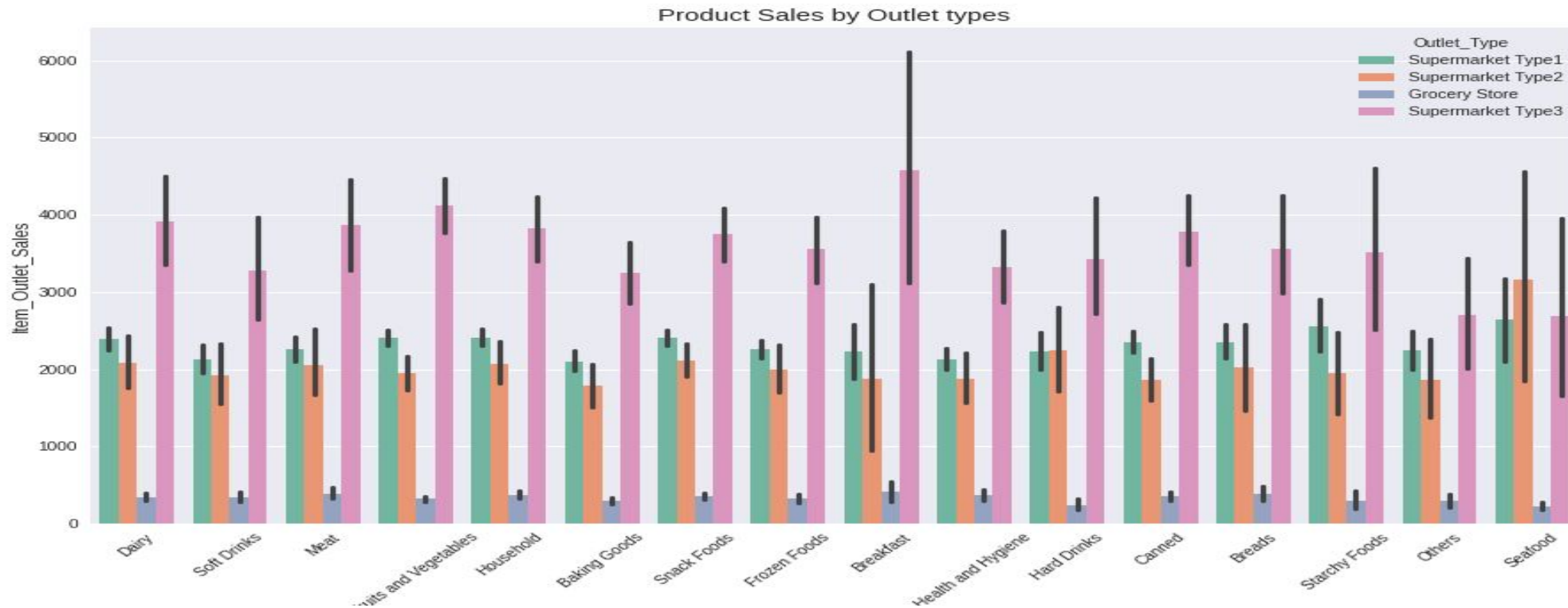
Performance by Year Established



Here we can see a sort visualization of Outlet Sales by Establishment Year. With with Supermarket Type3 having the highest sales performance (largest and oldest) and Grocery stores having the least in regards to sales performance



Product Sales by Item type



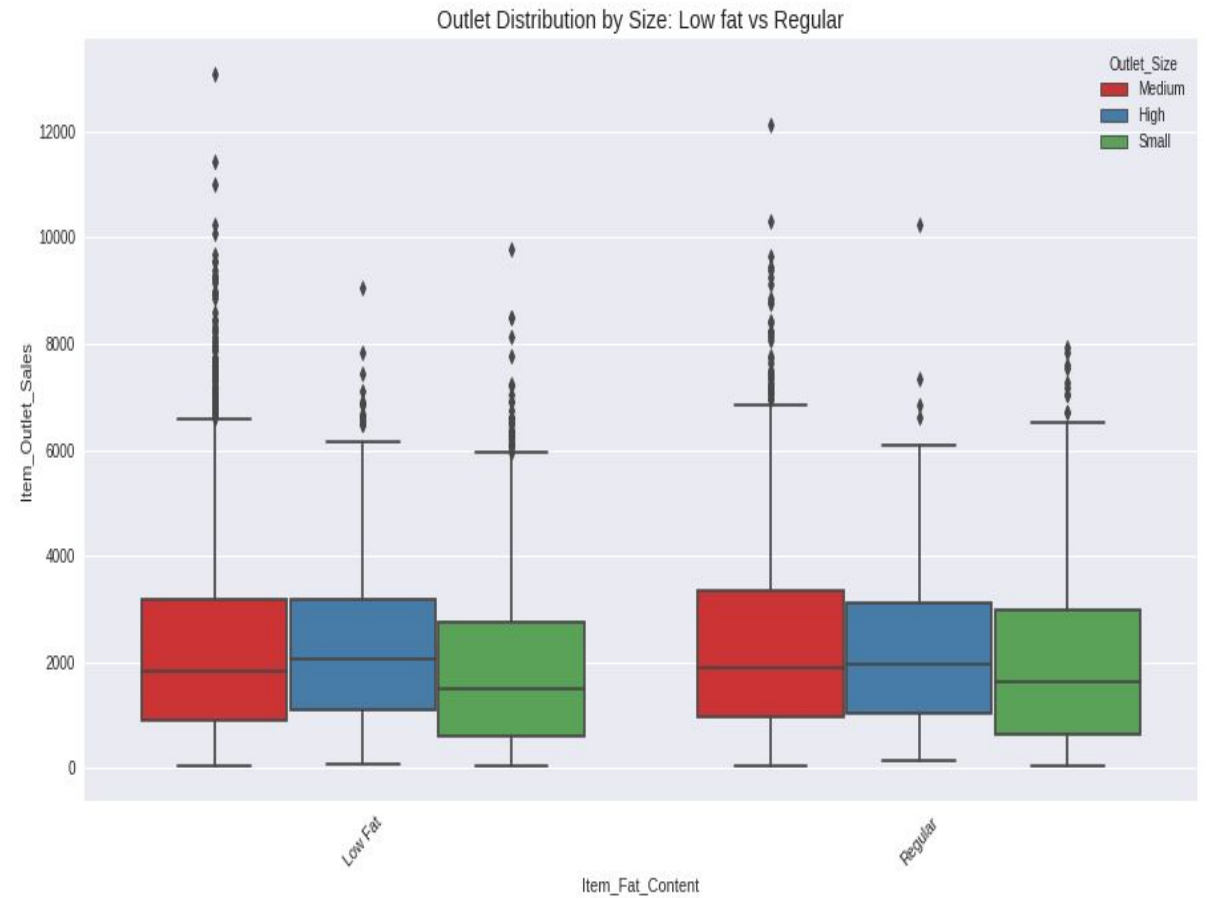
Relatively the outlet sales of supermarket Type 3 out performed the other stores in each product category except seafood. Both graphs indicate Grocery stores have constant minimum sales performance. Neither the year establish or type of product seem to decrease or increase sales performance.



Outlet Distribution by Size “LF” vs “Reg”

The distributions for both are nearly symmetrical for the 25th, median and 75th quartiles. However small size outlets have higher maximums for small sized Regular outlets in regards to sales for fat content compared to small sized Low Fat outlets. Overall the majority of sales come from medium and high size outlets

Suggestions: Based on the data For each product grocery stores have the lowest outlet sales among any Outlet type out of any product. From viewing the first graph we see they have the lowest amount of establishments with a minimum performance in sales across the board. Considering operation cost this organization should reevaluate operating cost of Grocery stores in regards to sales performance. Maybe a consolidation plan of low performing grocery stores in near proximities. Opening Medium or



Decision Tree Model

- ▶ Based on the features used Item Outlet Sales I was able to implement a functioning Machine learning model. We can predict options of the data at a 91% accuracy at handling current and new data. Further modification of features can be adjusted based on business needs.



```
evaluate_model(y_train, pipe4.predict(X_train))
```

```
MAE 218.19225397721672
```

```
MSE 245947.14518762362
```

```
R^2 0.9168944665022553
```

```
RMSE 495.930585049585
```

