Optimizing Operational Spend by Predicting Product Sales

FCAS Final Project_B3

Aditya Verma - 61910756
Amit Kumar Gupta - 61910279
Neeraj Nathany - 61910791
Prateek Singhvi - 61910308
Varsha Shridhar - 61910351
Vishal Abraham - 61910217

Executive Summary

In the retail business, it is crucial for the firms to accurately forecast the sales in the future to prepare themselves and optimize costs. Over-estimating sales can lead to a significant cost of inventory holding and even losses due to expiry in case of perishable items. There could also be a scenario of underutilized resources in manufacturing be it machinery or labour. Underestimation of sales in forecasts can lead to loss of business opportunity.

Problem description

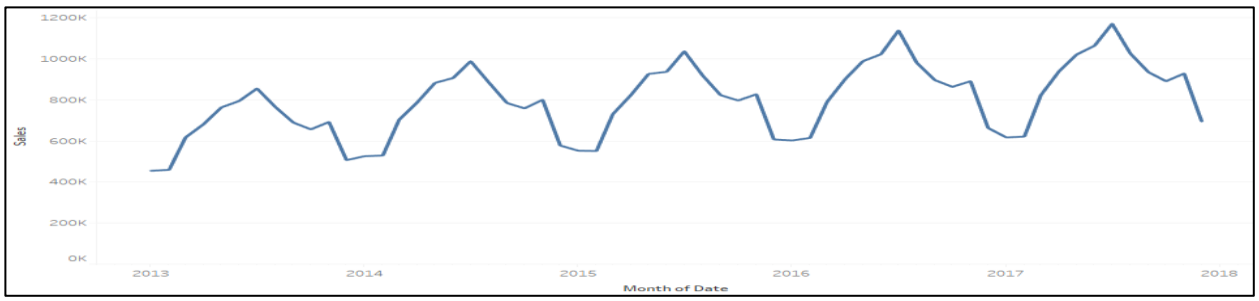
There are three main problems of the retailers that we are looking to solve

- Allocating budget towards increasing manufacturing or production capacity
- Creating scheduling plans for inventory holding of 2 particular stores to account for peak periods and lean periods
- To calculate the required number of counters that need to be operational on a given week day and the manpower required to monitor them

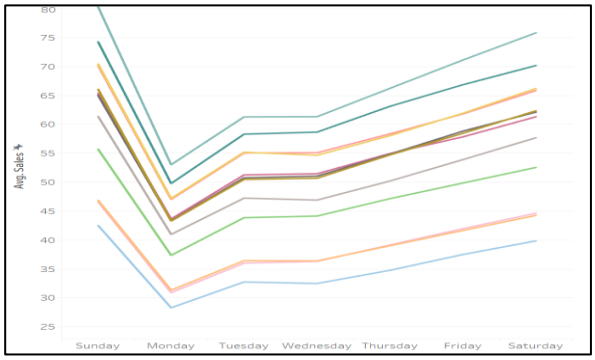
Data description

- We have 5 years (2013 onwards) of time series data at daily level for sales of 50 products at 10 store locations
- The dataset has 9,13,000 records and consists of categorical variables such as store id and product id and continuous variables such as sales
- The data has been obtained from Kaggle.com where it was used as a part of a competition
- There is no indication of the store names, locations, product name or product category. They have been hidden by assigning numbers to the names.
- Hyperlink of the dataset: https://www.kaggle.com/c/demand-forecasting-kernels-only/data

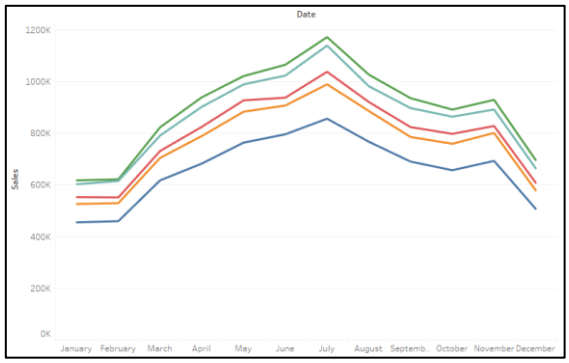
Overall Yearly Sales Trend



Weekly Sales Pattern Across Months



Monthly Sales Trend Across Years



Forecasting methods and performance evaluation

Following forecasting methods were used:

- Linear regression
- Holt-winters Smoothing
- Naïve forecast

We looked at RMSE and MAPE values to evaluate performance of forecasting models. The benchmarking was done against Naïve model which was assumed to be status quo.

Conclusions and recommendations

- Daily forecast for the 2 stores will be helpful in (a) staffing number of employees on feet, and (b) planning number of counters to be open on which day of the week

As can be seen from the data, there has been a consistent trend in the sales during weekend and sales during weekdays. It is notable that the sales start to peak during weekend i.e., Saturday and achieve at its peak on Sunday followed by a sudden drop on the next day.

Firm can hire contract employees for weekends – Saturday and Sunday. Regular employees can be continued throughout the week. Also, decision about the number of counters to be open can be taken to reduce operational costs. There should be a smaller number of counters open on Monday and other weekdays as compare to Saturday and Sunday.

- Monthly forecasts for the 2 products in a particular store will help determine (a) the inventory levels, and (b) the order levels.

As can be seen from the data, sales peak in the month of July and the minimum sales is observed in the months of December and January. A mini peak is also observed in the month of November. Firm should ensure maximum inventory during the month of July and can use its labor force in the previous months for product packaging and systematic updating of SKUs. Firm should be aware of ordering extra products during July and November. To manage extra capacity handling during July, pre planning about warehouse must be done.

- Quarterly or annual forecasts will help in capacity building at the aggregate production level.

There has been an increase of 11.06% in the sales 2017 to 2018. Based on the available forecast, the firm should manage its capacity. Firm can also decide whether it can improve its efficiency and processing to increase the productivity, or it has to add additional capacity.

Technical Summary

Data Preparation and Data Charts

- Data Cleaning – Few data points were missing in the data for which we assumed means of the population as placeholder values for them
- Exploration – We viewed many cross sections and segments of the data to look for levels, trend and seasonality.

Forecasting

Monthly Sales Forecast for item 5 and item 15 at Store 2:

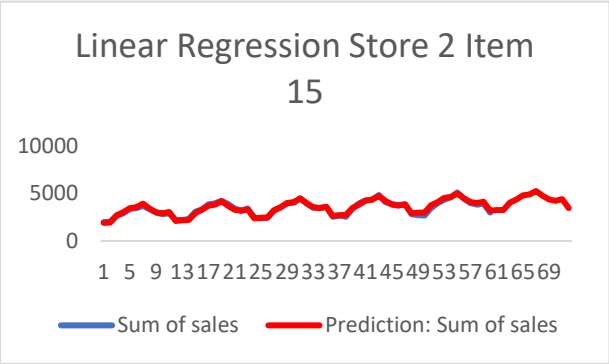
We chose item 5 and item 15 since these are the products with highest and lowest sales at store 2 which had the highest sales among 10 stores. For these 2 items, we started with the Naïve forecast as benchmark followed by linear regression with dummy variables to capture monthly seasonality and Holt Winters smoothing to compare model. Below are the results of various models:

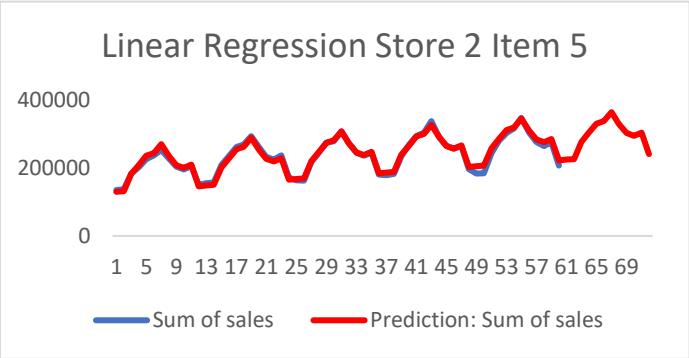
Time Series: Training: 48 months – Jan'13 to Dec'16, Validation: 12 months – Jan'17 to Dec'17, Forecast: 12 months – Jan'18 to Dec'18

Naïve Performance: RMSE: Item 15 - 288.17 and Item 5 - 73.73

Linear Regression: RMSE - Item 15 - 191.2 and Item 5 – 57.76

Holt Winter: RMSE - Item 15 – 243.54 and Item 5 – 64.38





Daily Sales Forecast for store 2 and store 7:

We chose store 2 and store 7 since stores have the highest and lowest sales. For these 2 stores, we started with the Naïve forecast as benchmark followed by linear regression with dummy variables to capture weekly and monthly seasonality and Holt Winters smoothing to compare model. Below are the results of various models:

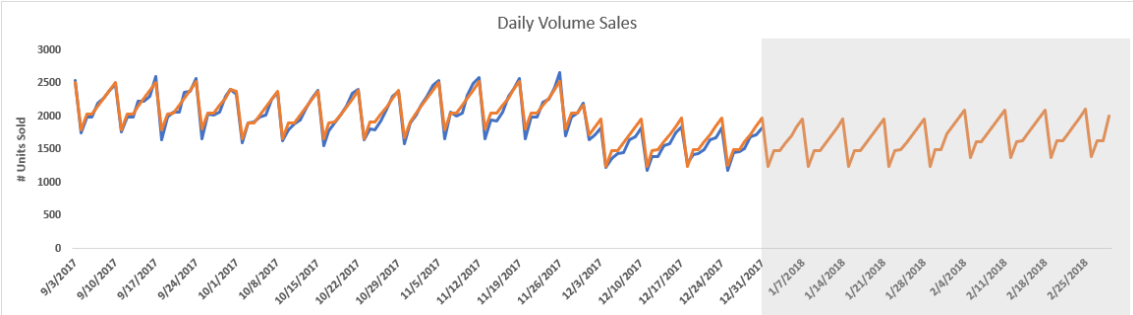
Time Series: Training: 1766 days – Jan'13 to Oct'17, Validation: 60 days – Nov'17 to Dec'17, Forecast: 60 days – Jan'18 to Feb'18

Naïve Performance: RMSE - Store 2 – 928.67 and Store 7 – 650.24

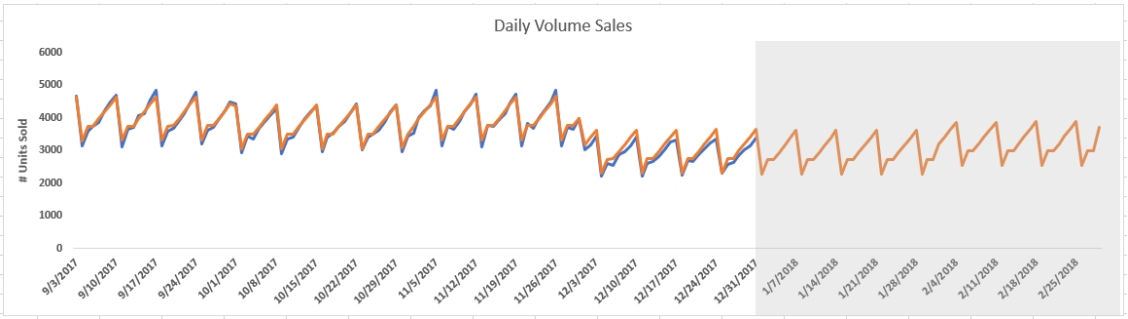
Linear Regression: RMSE - Store 2 - 175.14 and Store 7 – 108.40

Holt Winter: RMSE - Store 2 - 818.74 and Store 7 – 515.69

Store 7 Performance:



Store 2 Performance:



Quarterly Sales Forecast for item 5 and item 15:

We chose item 5 and item 15 since these are the products with highest and lowest overall sales. For these 2 items, we again started with the Naïve forecast as benchmark followed by linear regression with dummy variables to capture monthly seasonality and Holt Winters smoothing to compare model. Below are the results of various models:

Time Series: Training: 16 quarters – Q1'13 to Q4'16, Validation: 4 quarters – Q1'17 to Q4'17, Forecast: 4 quarters – Q1'17 to Q4'17

Naïve Performance: RMSE: Item 15 - 3318 and Item 5 – 695.93

Linear Regression: RMSE - Item 15 – 1794.9 and Item 5 – 363.44