Springboard – DSC
Capstone Project 2
Corporacion Favorita Grocery Sales Forecasting

Final Report

George Tang

December, 2019

Table of Contents

List of Figures

List of Tables

1 Introduction

The importance of sales forecasting cannot be overstated for any brick-and-mortar retail business. Managers leveraging accurate forecasts can ensure their stores have the right products in stock at the right time. This in turn provides a positive shopping experience which is critical in maintaining customer satisfaction and retention. Also, it helps businesses to control item inventory and optimize shelf space usage, reduce perishable goods wastage, labor and transportation costs, and ultimately increase revenue and profit.

These benefits motivated Corporacion Favorita to host Kaggle's Corporacion Favorita Grocery Sales Forecasting competition in 2017. Corporacion Favorita is an Ecuador-based grocery retailer that operates hundreds of supermarkets and carries over 200,000 items. At the time, the business relied on subjective forecasting methods backed by little data analytics. They used the competition to challenge data scientists to develop innovative, data-driven solutions that can ultimately improves their sales forecasts.

1.1 Objective

The objectives of this project are:

- To Identify features that affect sales
- To develop sales forecasting models and evaluate their performance
- To provide actionable recommendations for the management

1.2 Significance

Through this project, we have identified important features that determine sales. The company can use the sales forecasted by machine learning models to plan its logistics and inventories, which allows them to better utilize their resources and ultimately maximize profit.

1.3 Programming Code

The programming codes are posted in this github repository.

2 Dataset

The data used in this project were obtained from the Kaggle competition's website, which contains 7 data tables. We will provide a description of each data table in this section.

Sales data

The train dataset (train.csv) consists of the sales data for each item sold in each store. It consists of 125,497,040 rows and 6 columns, namely Id number, Date, Store Number, Item Number, Promotion Status, and Unit Sale, with date ranges from 2013/1/1 to 2017/8/15.

The test dataset (test.csv) contains sales information for date ranges from 2017/8/16 to 8/31, but without the Unit Sales. We will not use this dataset in this project because model evaluation for sales forecasting in that period is impossible without the true target values (Unit Sales).

Transaction data

This dataset (transactions.csv) consists of the transaction information for each store. It contains 83,488 rows, one row for the number of transactions of the date at that store. The three columns are Date, Store Number and Number of Transactions. We did not use this dataset in this project.

Store information

This dataset (stores.csv) consists of 54 rows, one for each of the stores owned and operated by the grocery store chain. The 5 columns are: Store Number, City, State, Type, and Cluster.

Item Information

This dataset (items.csv) consists of 4,400 rows, with each row consisting of information of one grocery item. Note that it is a subset of the 200,000+ products carried by this company. The 4 columns are: Item Number, Family, Class and Perishable Status.

Holidays and special events

This dataset (holiday_events.csv) consists of holidays and special events information of the country.

Oil price

Since Ecuador is an oil-dependent country, Ecuadorians' spending may be affected by changes in oil prices. This dataset (oil.csv) contains two columns, namely Date and Oil Price.

3 Data Cleaning and Wrangling

Since the data is provided for a Kaggle competition, all the datasets are relatively clean and ready for data visualization and analysis. Only the Sales dataset requires some data cleaning and wrangling steps which are summarized as the following:

- Change Boolean data type from string to Boolean
- Change datetime data type from string to datetime
- Create new columns of Year, Month, Day of Month and Day of Week based on the Date column
- Impute missing values for the Promotion Status column

We also merged datasets with common columns (e.g. Date, Item Number, Store Number, etc.) for data visualization and statistical analysis as needed.

On the other hand, we conducted extensive feature engineering for machine learning which we will cover in Section 5.3.

4 Exploratory Data Analysis

4.1 Store Information

The number of stores in each city is shown in Figure 4-1(a). The 54 grocery stores are a subset of the 100+ stores operated by the company. They spread across 22 cities in 16 states. Quito, the capital city of Ecuador, consists of the most stores with 18. Guayaquil, the largest city in the country, contains 8 stores.

Figure 4-1(b) shows the number of stores for each type. The five store types are Megamaxi, Supermaxi, Gran Akí, Super Akí and Akí, although their corresponding labels (A to E) are not provided. The most common store type is D (18), followed by C (15) and A (8), while the least common type is E (3).

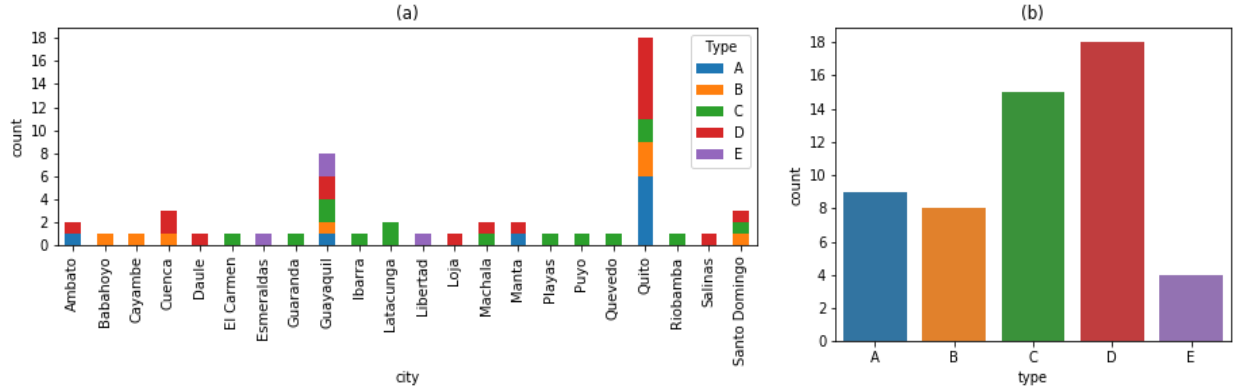


Figure 4-1 Store count (a) by city, (b) by type

The company divides the store into clusters is based on store size, category sales volume and shopper purchase behavior. Figure 4-2 shows that there are 17 clusters. Except for Cluster 10, all other clusters consist of only one store type. Cluster 3 consists of the most stores (7), followed by Clusters 6 and 10 (with 6 each). Each of Clusters 5, 12, 16 and 17 consists of only 1 store.

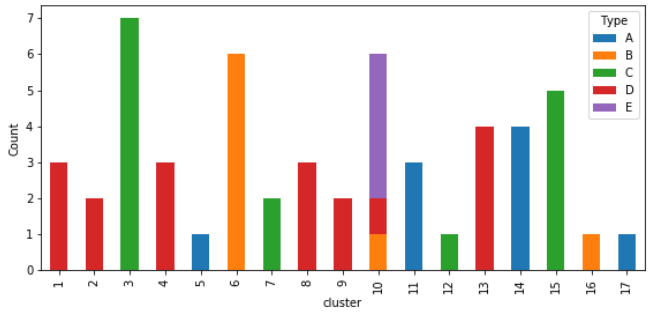


Figure 4-2 Store count by cluster

4.2 Item Information

Figure 4-3 shows the number of items for each product category. In total, there are 4,400 items in the dataset, which is a subset of the 200,000+ items carried by the company. GROCERY I contains the highest number of items (1,314), followed by BEVERAGES (613) and CLEANING (446). Baby care, Books and Home Appliances contain the fewest items with only one each.

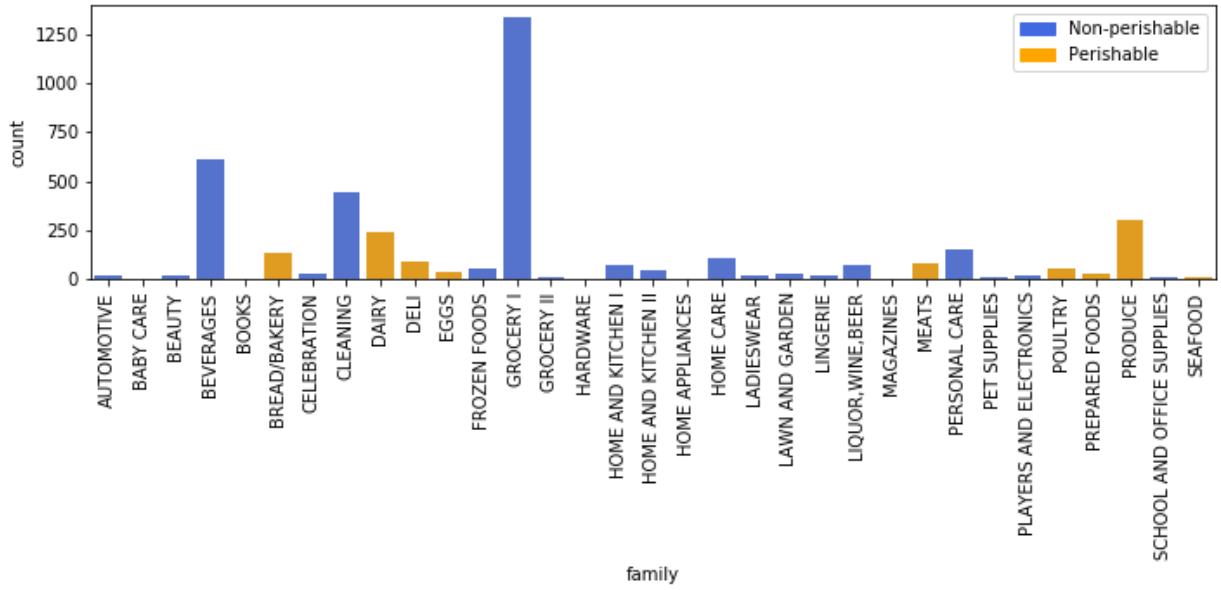


Figure 4-3 Item count by family

Items that belong to one family are either all perishable or non-perishable. Families of Bread/bakery, Deli, Poultry, Eggs, Dairy, Seafood, Prepared Foods, and Produce are perishable, which comprises of 986 items (24% of total items). This information is important because perishable products must be disposed of after their sell-by dates, which will lead to an important loss in revenue.

Additionally, each item is assigned a class number, likely based on their similarities and functions (e.g. one class number assigned to different brands of cereals, another for various soft drinks, etc.)

4.3 Oil Price

Since Ecuador is an oil-producing country, the nation's consumer behavior may be affected by fluctuations in oil price. Daily oil price from 1/2013 to 8/2017 is shown in Figure 4-4. The figure shows that there is a substantial decline in oil price beginning in the end of 2014. Prior to that, oil price hovered around $100/barrel with maximum value of $110.62/barrel. After that, oil price hovered around $50/barrel with minimum value of $26.19/barrel on 2016-02-11.

Figure 4-4 Daily oil price from January 2013 to August 2017

4.4 Holidays and Special Events

This dataset provides information on the holidays and special events in Ecuador. Some holidays are national while others are local. The most notable special events were the World Cup in the Summer of 2014, and the earthquake event that happened in 2016-04-16.

4.5 Sales Data

The sales data (train.csv) consists of item sales information at each store from 2013-01-01 to 2017-08-15.

4.5.1 Effect of Store Types on Unit Sales

In this section, we will look at how total sales are related to the store types. Given the vast number of items, we will present the 'big picture' by the summing the unit sales for each item in a family. Also, we will only consider sales in 2017 since it is more relevant to our prediction of future sales.

The total unit sales in each store for the item families "GROCERY" I and "PRODUCE" in 2017 are shown in Figure 4-5. Notice that Store #52 was not in business until 2017-04-20. The total unit sales can be significantly different even for stores of the same type. For example, the total sales of Store #26 are considerably higher than all other Type A stores (red) across all three families. Overall, for GROCERY I, total sales are the highest for Type A stores followed by Type D stores. For PRODUCE, Type A and Type D stores are comparable where Type C stores are lowest.

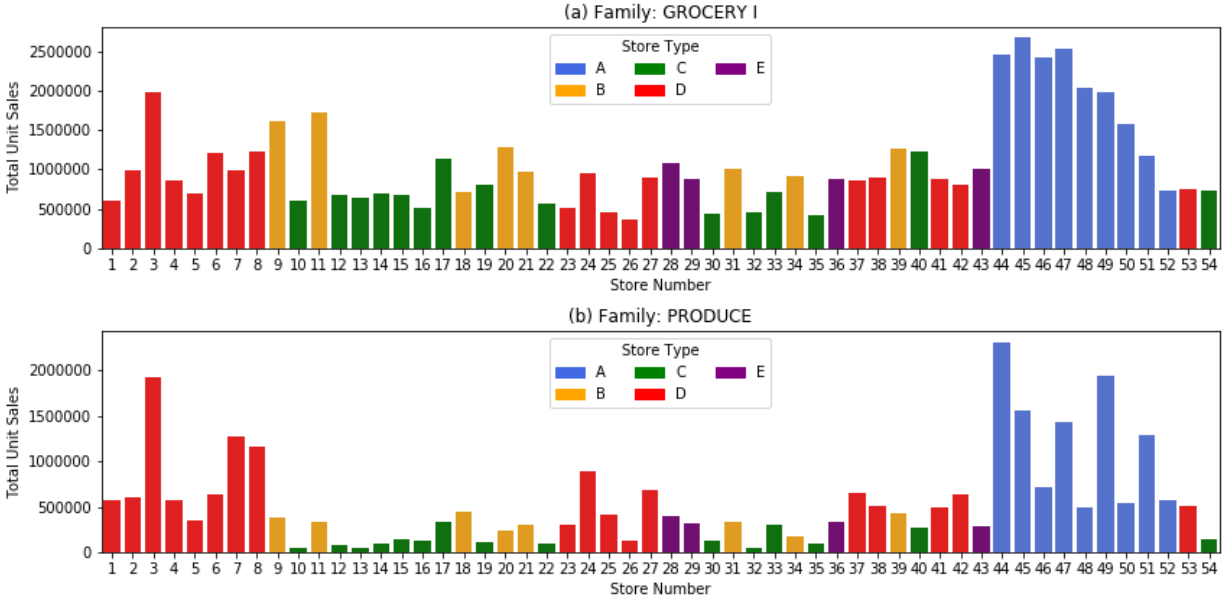


Figure 4-5 Sum of unit sales of (a) GROCERY I, (b) PRODUCE for all stores in 2017

4.5.2 Effect of Locations on Unit Sales

The total unit sales of GROCERY I in 2017 for Quito and Guayaquil, the two cities with the most stores, are shown in Figure 4-6(a) and (b), respectively. The figure shows that, overall, the total sales of stores in Quito are higher than those in Guayaquil.

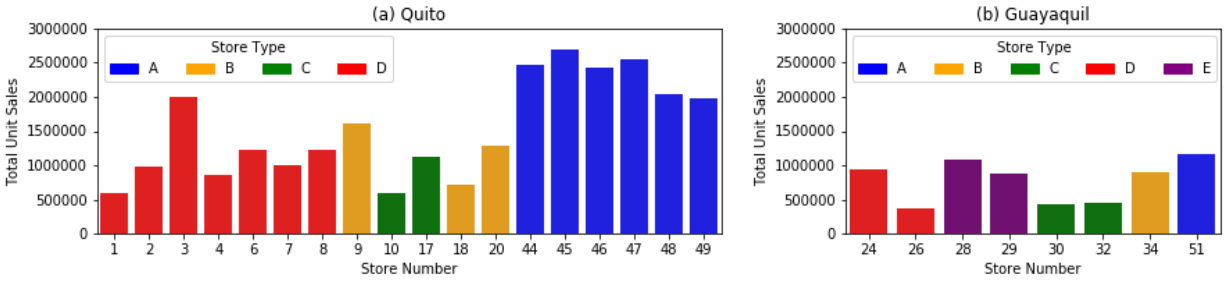


Figure 4-6 Sum of unit sales of GROCERY I in (a) Quito, (b) Guayaquil in 2017

4.5.3 Effect of Day of the Week on Unit Sales

For this analysis, we will focus on Stores #1, #9, #17 and #45. They are all located in Quito, the capital city which consists of the highest number of stores. The four stores encompass the four store types (A, B, C and D) present in the city.

The effect of day of the week on the sum of unit sales for GROCERY I is shown in Figure 4-7. All stores, except for Store #1, show increased sales during the weekends (Saturday and Sunday). For Store #1, sales are similar from Monday to Saturday, but show a significant dip on Sundays. Overall, the effect of day of the week on unit sales is specific to the store.

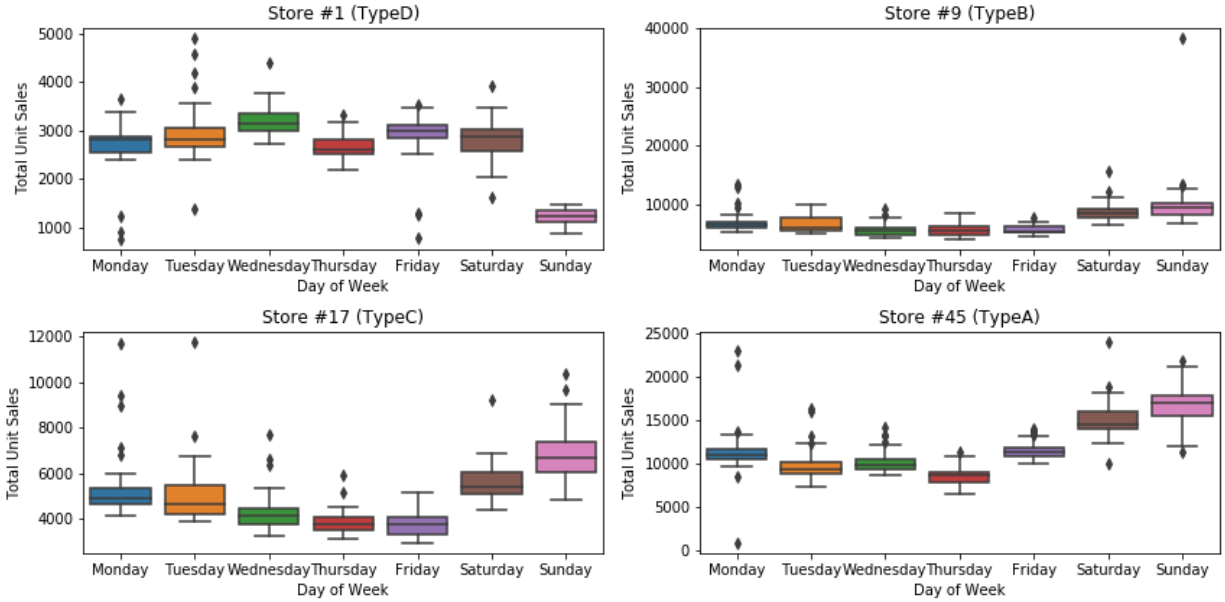


Figure 4-7 Sum of unit sales of GROCERY I items in 2017 for Stores #1, #9, #17 and #45

4.5.4 Effect of Holidays and Special Events on Unit Sales

Christmas and New Year in 2016/2017

Intuitively, the sales of products in categories LIQUOR,WINE, and BEER, among others, will increase during holiday seasons. Figure 4-8 shows the total daily unit sales for this family in Store #45 between December 2016 and January 2017. The figure shows an increase in sales prior to the two holidays. Similar trends are also present in earlier years, as well as in some other stores.

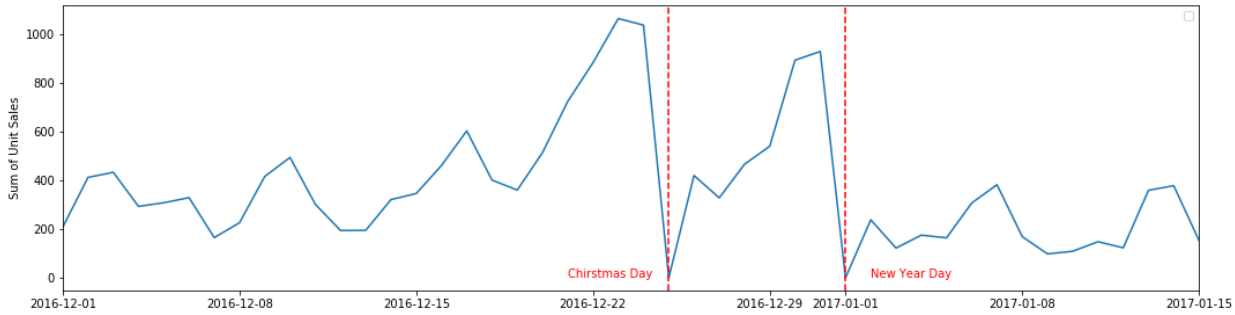


Figure 4-8 Sum of unit sales of LIQUOR,WINE, BEER items in Store #45 between December 2016 and January 2017

National Day in 2016

The national day of Ecuador in 2016 is on 2016-08-10 (Wednesday), and was transferred to 2016-08-12 (Friday). Figure 4-9 shows the daily sum of LIQUOR, WINE, BEER item unit sales for Store #45. Contrarily to Christmas and New Year, the increase in sales during this holiday is minimal.

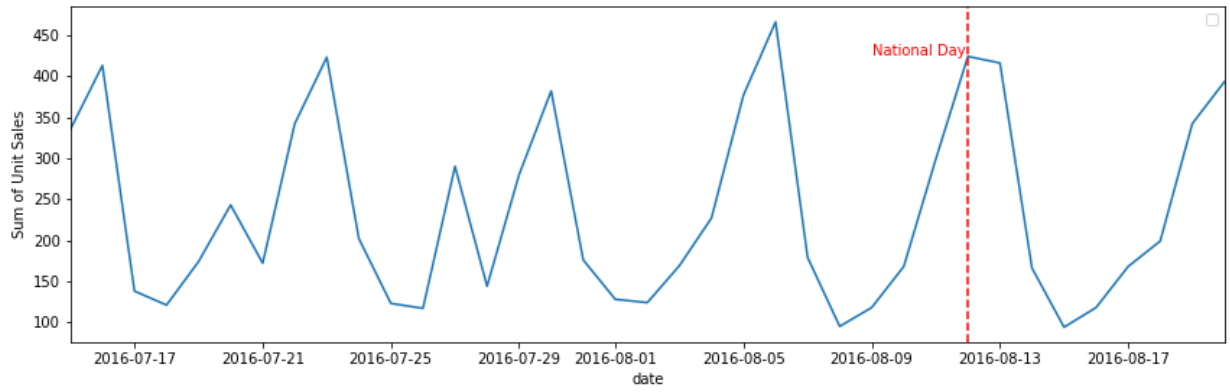


Figure 4-9 Sum of unit sales of LIQUOR,WINE, BEER items in Store #45 in August 2016

World Cup in 2014

The Ecuador national team was part of the World Cup football tournament 2014, which was held in the summer of that year. Ecuadorians are famous for their passion for the sport. Figure 4-9 shows the daily sum of LIQUOR, WINE, BEER item unit sales for Store #45 in June 2014. We see the seasonal pattern of alcoholic beverage sales which peaked on Saturdays during the period. We can also see an increase in sales that corresponded to the dates when the team played. It seems that during this period, alcohol sale was prohibited on Sundays. As such, since Game 1 was held on a Sunday, the sales increase occurred the day before. Similar trends are also present in some other stores. The following World Cup tournament was held in 2018, so there is no similar event before 2017/08/15.

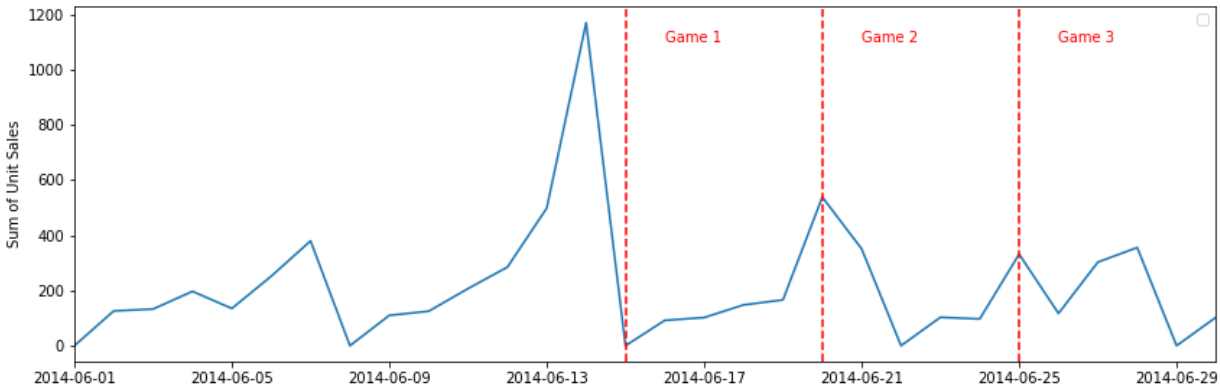


Figure 4-10 Sum of unit sales of LIQUOR,WINE, BEER items in Store #45 in June 2014

Earthquake in 2016

An earthquake occurred on 4/16/2016. Figure 4-11 shows the daily sum of GROCERY I item unit sales for Store #45 in that month. Sales peaked on 4/18 and 4/21, which corresponded to a nationwide charity drive. Similar trends were also present in some other stores. There is no similar event before 2017/08/15.

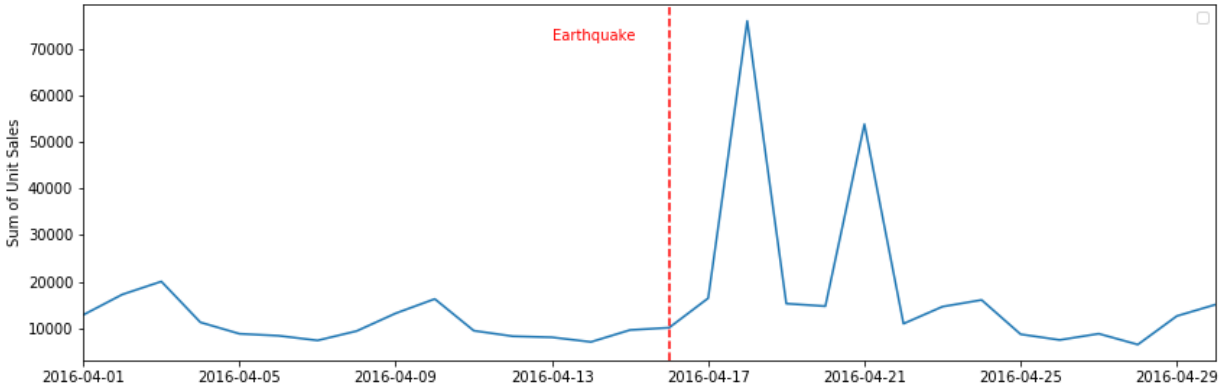


Figure 4-11 Sum of unit sales of GROCERY I items in Store #45 in April 2016

4.6 Statistical Inference

In this section, we will present the results of statistical hypothesis tests to examine whether the unit sales (target variable) at different values of a categorical variable are from the same population or not. This way, we can identify features that may be helpful in predicting sales.

Here we consider the population to be a specific item in one store. Given the numerous store-item combination, we choose Store #45 because its overall unit sales are among the highest in all stores. We choose one of the top sellers in the three families examined above, namely, GROCERY I, PRODUCE and LIQUOR,WINE,BEER.

4.6.1 Effect of Day of Week on Unit Sales

Figure 4-12 shows the distribution of unit sales of three selected items in 2017. The figure shows that the effect of days of the week are different among the products. For GROCERY I Item #314384, unit sales are higher during the weekends than the weekdays. For PRODUCE Item #1503844, unit sales are higher during the weekends and on Wednesdays. The high sale on Wednesdays is due to promotion of some of the

items in this family. For LIQUOR,WINE,BEER Item #1004550, the highest unit sales occurred on Saturdays, followed by Fridays.

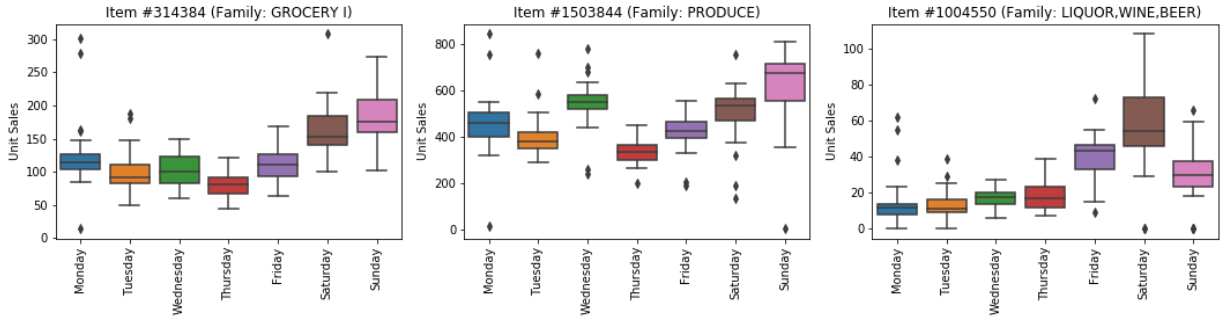


Figure 4-12 Unit sales of selected items with respect to the days of the week in Store #45 in 2017

4.6.2 Effect of Promotion on Unit Sales

Figure 4-13 shows the effect of promotion on the three selected items. Among them, promotion status has the highest effect on PRODUCE Item #1503844.

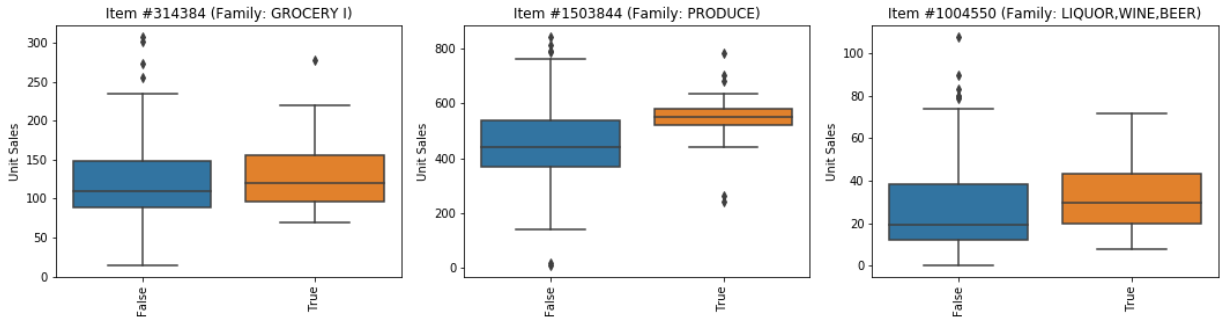


Figure 4-13 Unit sales of selected items with respect to promotion status in Store #45 in 2017

4.6.3 Statistical Analysis

In this section, we will conduct hypothesis tests on the two categorial features, namely days of the week, and promotion status. The null hypothesis is that the distributions of unit sales are the same among all categories within a feature, where the alternative hypothesis is that the distributions are not the same. To account for the potential non-normality of unit sales distribution, the Mann Whitney U (MWU) test will be used for features with binary response (1/0, or t/f, etc), where the Kruskal-Wallis (KW) test will be used for multi-category responses. Both tests are non-parametric and do not assume any distribution of the samples. The hypothesis tested by both MWU test and KW test is whether the samples from the different categories are taken from the same population.

The results show that, for all features, the null hypothesis is rejected, i.e., unit sales do not follow the same distribution among all categories in each feature. It means that the two features can be useful in predicting unit sales.

5 Sales Forecasting with Machine Learning

5.1 Test Dataset

We define the target dataset as the sales of each item for each store for the 15-day period from 8/1/2017 (Tuesday) to 8/15/2017 (Tuesday). It is the last 15-day period in the sales data (train.csv) with sales record available.

We evaluate model performance with the metric provided by the Kaggle competition, known as the Normalized Weighted Root Mean Squared Logarithmic Error (NWRMSLE):

$$NWRMSLE = \sqrt{\frac{\sum_{i=1}^{n} w_i[\ln(\hat{y}_i + 1) - \ln(y_i + 1)]^2}{\sum_{i=1}^{n} w_i}}$$

Where:
n is the total number of item sales across all stores in the 15-day period;
w_i is the item weight, which is 1.25 and 1 for perishable and non-perishable items, respectively;
$\hat{y}_i$ is the predicted unit sales; and
y_i is the actual unit sale.

Note that the metric is calculated is based on the logarithmic transformation of the unit sales. The purpose is to account for the difference in magnitude of unit sale values across all items, which range from 0 to > 10,000. With this metric, predicting 20 unit sales for a true value of 10 (error of 10) would be penalized more heavily than predicting 110 unit sales for a true value of 100 (error of 10 units).

This metric also penalize error more heavily in sales predictions for perishable items. As perishable items must be disposed of after the sell-by date, errors in sale forecasts will lead to higher financial loss.

5.2 Modeling Approach

We adopted the modeling approach from the winner of this Competition with modification:

1. Create a training, validation and test dataset which consists of features as discussed in the next section
2. Use the training dataset to fit machine learning models. In total, we create 15 separate models, one for each day of the 15-day period, with the sales data of each day for all item unit sales across all stores as target variables
3. Use the validation dataset to prevent over-fitting of models
4. Make predictions with the models and test dataset for the 15-day sales forecasting period

5.3 Feature Engineering

We created the following three type of features with the input variables: past sales data, promotion status, and store and item information.

Historical sales data

Sales forecasting is a type of time series problem where the unit sales of items in a store is related to their sales in the past. We created the following features for each item sales for each store:

- Unit sales 1 to 15 days before the reference date (first day of the 15-day forecasting period)
- Mean daily difference for the last 3, 7, 14, and 30 days before the reference date

- Mean, median, minimum and maximum sales and standard deviation of sales 3, 7, 14 and 30 days before the reference date
- Total number of days with sales (i.e. unit sales > 0) 7, 14 and 30 days before the reference date
- Mean sales on days with promotion 3, 7, 14 and 30 days before the reference date
- Mean sales on days without promotion 3, 7, 14 and 30 days before the reference date
- Mean daily sales for each date of week (i.e. Sunday, Monday etc.) 4 weeks and 20 weeks before the reference date

For the test dataset, the reference day is 8/1/2017, which is a Tuesday. As such, we set the reference dates for both the training and validation datasets to be the same day of the week. Also, since it is time series prediction, the time period for the training dataset must precede the validation dataset, which in turn must precede the test dataset to avoid using future data to predict past data.

The chosen reference dates for the training, validation and test datasets along with the corresponding sales forecasting periods are shown in Table 5-1. The sales data during the 15-day sales forecasting periods are the target variables for the training and validation datasets.

Table 5-1 Data time periods for training, validation and test datasets

Dataset	Reference Date	Sales Forecasting Periods (Target variables)
Training	5/30/2017 (Tues) 6/6/2017 (Tues) 6/13/2017 (Tues) 6/20/2017 (Tues)	5/30/2017 (Tues) – 6/13/2017 (Tues) 6/6/2017 (Tues) – 6/20/2017 (Tues) 6/13/2017 (Tues) – 6/27/2017 (Tues) 6/20/2017 (Tues) – 7/4/2017 (Tues)
Validation	7/11/2017 (Tues)	7/11/2017 (Tues) – 7/25/2017 (Tues)
Test	8/1/2017 (Tues)	8/1/2017 (Tues) – 8/15/2017 (Tues)

For the training data, we have chosen four reference dates to increase the size of the training dataset. We concatenate vertically the dataset from the four reference dates and the corresponding target variables to result in one training dataset.

Promotion Status

As discussed in Section 4.6.2, item sales generally increase with promotion status. Intuitively, item sales may decrease after, or before a pending promotion event. We create the following features base on items' promotion status:

- Promotion status from 14 days before to 14 days after the reference date
- Sum of promotion days with 7, 14 and 30 days before the reference date
- Sum of promotion days with 3, 7 and 14 days after the reference date
- First and last promotion days 14 days before the reference date
- First and last promotion days 14 days after the reference date

Store and Item Information

We create dummy variables for the following store and item information:

- Province of the store
- City of the store
- Store type
- Store cluster
- Item family

- Item class

5.4 Machine Learning Models

In this study, we will first use gradient boosting to establish a 'baseline' performance. Then we will explore the use of neural network for additional gain in accuracy.

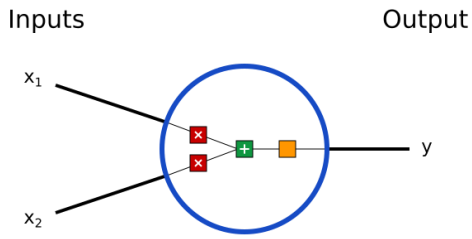
5.4.1 Gradient boosting with LightGBM

Gradient boosting belongs to the family of non-parametric ensemble models. It involves building binary decision trees sequentially with each tree fitted to the residuals of the preceding tree. We will use Microsoft's LightGBM (LGBM), a state-of-the-art gradient boosting package for this project. The model incorporates two improvements to the conventional model: 1. Gradient-based One-Side Sampling which inspect the most informative samples while skipping the less informative samples, and 2. Exclusive Feature Bundling which takes advantage of sparse dataset by grouping features in a near lossless way.

We will use grid search to obtain the best hyperparameters. For each combination of hyperparameters, we will use the validation dataset will determine the early stopping, i.e., if model building will cease to continue if there is no further improvement in cross validation score.

5.4.2 Neural Network

Neural network is another state-of-the-art machine learning algorithm. It consists of multiple layers of neurons which is the basic unit of a neural network. Figure 5-1 shows a simple neuron with two inputs.



(Source)

Figure 5-1 Mathematical operations in a neuron with 2 inputs

The mathematical operation of a neuron consists of the following steps:

1. Each input is multiplied by a weight
   - $x_1 \rightarrow x_1 * w_1$
   - $x_2 \rightarrow x_2 * w_2$
2. All the weighted inputs are added together with a bias
   - $(x_1 * w_1) + (x_2 * w_2) + b$
3. The sum is passed through an activation function to produce an output
   - $y = f(x_1 * w_1 + x_2 * w_2 + b)$

There are multiple neurons in each layer, each with different values of weights. The output is then used as the input of the following layer, and the output of the last layer is the predicted value, $\hat{y}$. Based on the prediction error, the weights and biases of each layer are updated with back-propagation.

Since the algorithm of neural network (NN) is considerably more complex, we are interested in whether it has a higher accuracy than LGBM.

The architectures of the three NN models are shown in Table 5-2.

Table 5-2 Architectures of the NN models

Model	Number of layers	Number of neurons in each layer
#1	7	512 x 256 x 128 x 64 x 32 x 16 x 1
#2	8	512 x 512 x 256 x 128 x 64 x 32 x 16 x 1
#3	6	256 x 128 x 64 x 32 x 16 x 1

5.5 Model Evaluation

5.5.1 Model Performance

We chose the best LGBM model and NN model, with their validation and test errors summarized in

Table 5-3.

Table 5-3 Validation and test errors (NWRMSLE) of the best LBGM and NN Models

Model	Validation	Test
LGBM	0.5866	0.5966
NN (Model #3)	0.6099	0.6222

The best LGBM model has hyperparameters of *maximum number of leaves = 200*; *feature fraction = 0.8*; *bagging fraction = 0.8*.

Interestingly, the simplest NN model (Model #3) has the best performance. It is possible that the more complex models overfit to the training data.

Overall, the best LGBM model has better validation and test scores than the best NN model. The validation scores for models with other hyperparameters are shown in **Appendix I**.

5.5.2 Performance Evaluation of Selected Items

In this section, we will specifically look at sales forecasting for a few choice items, as shown in Figure 5-2. Even though the sales forecasting period is the first 15 days of August, we have included sales data from May of 2017, because historical sales data are used as input to the models.

Figure 5-2(a) and (b) shows that both LGBM and NN models perform reasonably well if the sales data follow a seasonal pattern.

On the other hand, Figure 5-2(c) and (d) shows that, for the two items, the unit sales during the sales forecasting period were considerably different from those of previous months. In this case, both models perform poorly, since they are trained with historical data that does not project well to the future.

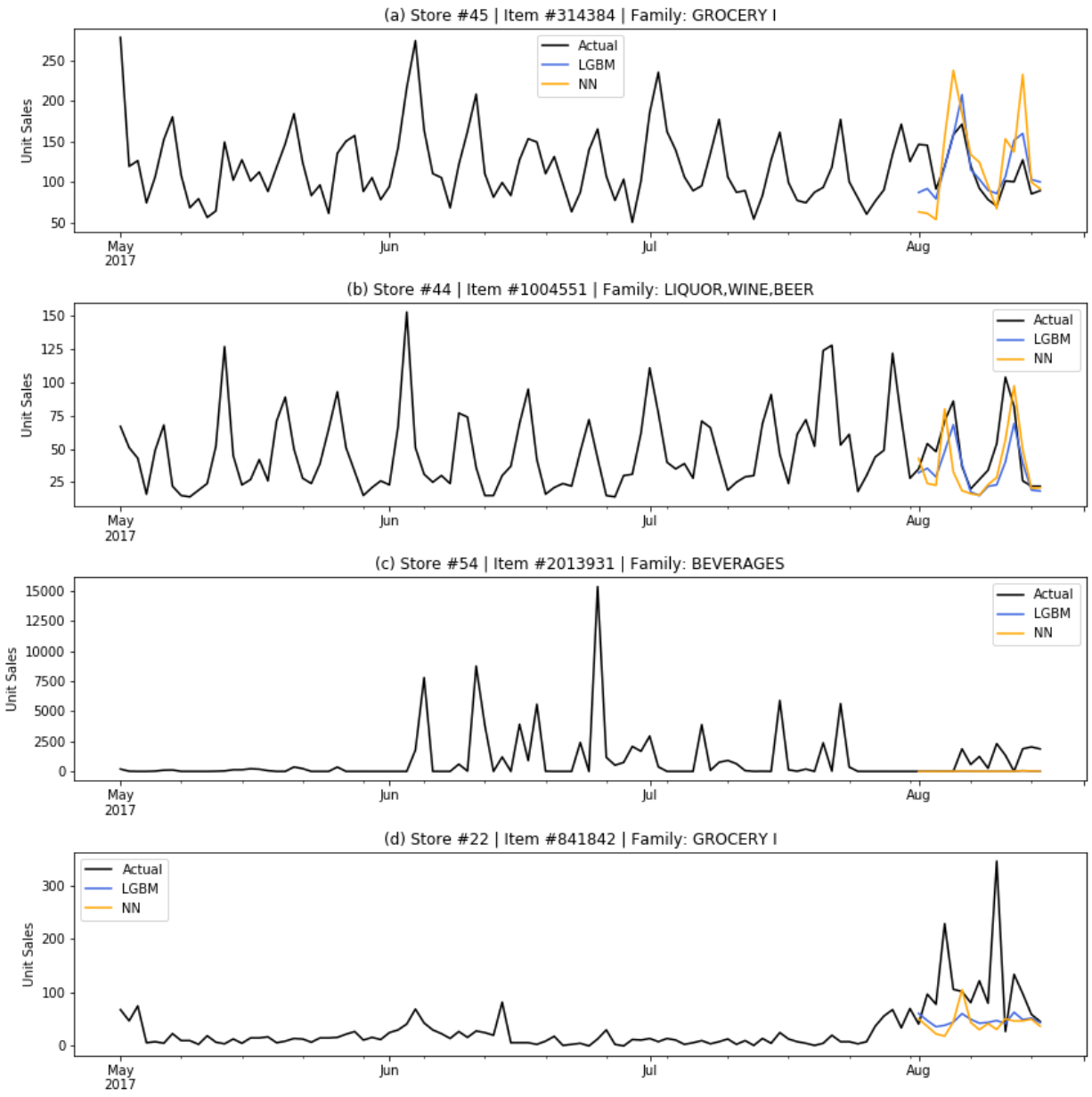


Figure 5-2 Sales forecasting with LGBM and NN models for four selected items

A statistical summary of the prediction errors is summarized in Table 5-4. Note that it is a very small subset out of the 4,400 items sold in 52 stores, which amount to over 160,000 data series.

Table 5-4 Statistical summary of model prediction error for the selected items during the 15-day forecasting period

Store #	Item #	MAE		Max & Min Errors*	
		LGBM	NN	LGBM	NN
45	314384	21.54	40.65	(51.21, -59.25)	(105.25, -83.9)
44	1004551	15.78	18.98	(13.51, -63.79)	(23.5, -53.20)
54	2013931	883.4	883.7	(7.08, -2308)	(7.83, -2307)
22	841842	66.2	70.7	(19.91, -298.3)	(23.85, -315.1)

* Positive number denotes sales overprediction; negative number denotes sales underprediction

5.5.3 Feature Importance of LGBM Models

One advantage of the LGBM models is that they provide information on feature importance. Figure 5-3(a) and (b) shows the top-10 features for the models used for 2017-08-01 (Tuesday, Day 1) and 2017-08-13 (Sunday, Day 13), respectively. Most features are variation of the sales data creating by feature engineering, along with promotion data.

For Model (a), the most important feature is the mean sales in the last 14 days, followed by mean sales in the last 7 days. The next important feature is the mean sales in the last 20 Tuesdays (*mean_20_dow0*, where 0 refers to the number of days from Tuesday). The promotion status on that day (*promo_0*, where 0 refers to the number of days from 2017-08-01) is also a top-10 feature.

For Model (b), the most important feature is the mean sales in the last 30 days (based on the reference day which is 2017-08-01), followed by the mean sales on the last four Sundays (*mean_4_dow5*, where 5 refers to the number of days from Tuesday (Sunday is 5 days from Tuesday)). The promotion status on that day (*promo_12,* where 12 refers to the number of days from 2017-08-01) is also a top-10 feature.

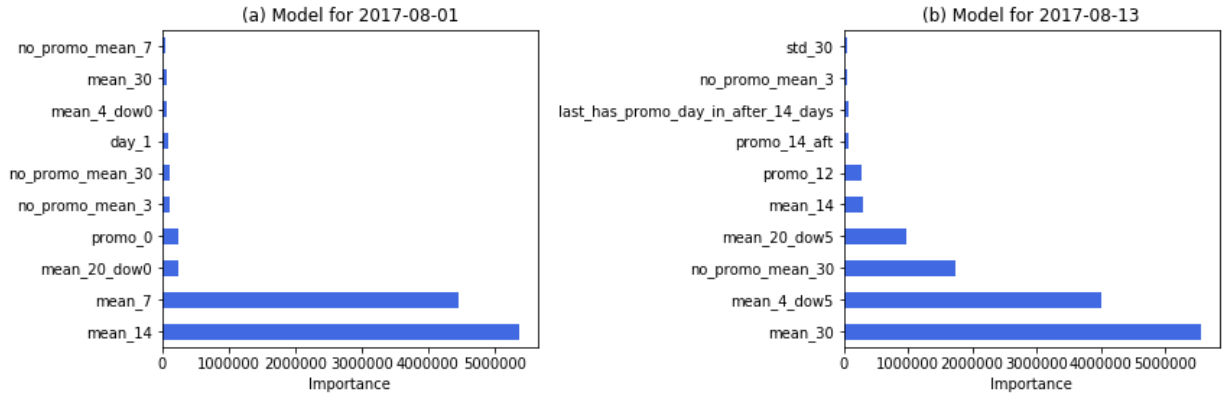


Figure 5-3 Feature Importance for LGBM models used to predict sales on (a) 2017-08-01 (Day 1) and (b) 2017-08-13 (Day 13)

6 Recommendations and Future Work

In this study, we have examined Coporacion Favorita's sales data in detail through visualization and statistical analysis. We have also applied state-of-the-art machine learning models that allows the company to forecast sales.

We have the following recommendations to the management of the company:

1. Our machine learning model shows that past sales data (e.g. mean sales for the last 30 days, and mean sales on the same day of week for the last 4 weeks) are important features in predicting future sales. We suggest the company to keep track of those information and use them to predict sales in the following weeks.
2. Our data analysis shows that holidays and special events (e.g. World Cup) affect sales which are store- and item-specific. To ensure the uninterrupted availability of the popular products, we recommend the company to identify the types of products that are likely to have high demands during those occasions, both from sales data during in previous occasions (e.g. Christmas of yesteryears), and the more recent sales data, and to prepare accordingly.
3. Extreme events such as an earthquake can lead to drastic change in sales pattern, not only for stores in the affected areas, but also for those in other areas due to national charity drives. We suggest the company to develop emergency response plans to ensure adequate products are in-stock in all areas. In particular, the plans should include risk management of potential logistics interruption caused by the events.

We propose the following future works to improve sales forecasting:

Create models with subsets of data

In this study, we adopted a somewhat ambitious approach in creating models that encompass all items across all stores. One way to improve accuracy is to create individual models for a store, a city, or even a family of items. The models may better fit to the trends specific to the subsets of data and produce better accuracy.

Understand population demographic

Customers from different population demographics may have vastly different spending habits. For instance, parents with babies need to spend on baby products, while households with teenagers may regularly buy packaged foods and beverages. Understand the customer demographic, which is specific to each store, along with its dynamics can help improving sales forecasting.

Explore the use of commercially available forecasting services

Many IT service providers, such as Amazon Web Services (AWS) and Databrick, provide sales forecasting services with their proprietary, state-of-the-art machine learning algorithms. We encourage the company to explore the use of those services for better forecasting.

Appendix I

Validation Score for LGBM Models

Model	Max no. of leaves	Feature Fraction	Bagging Fraction	NWRMSLE
0	31	0.8	0.6	0.5877418750255925
1	31	0.8	0.8	0.5876746516156044
2	31	0.8	1.0	0.5881172916574132
3	31	0.9	0.6	0.5878097976764276
4	31	0.9	0.8	0.5878256686637173
5	31	0.9	1.0	0.5882729044227182
6	31	1.0	0.6	0.5878852507920574
7	31	1.0	0.8	0.5879840469762064
8	31	1.0	1.0	0.5884032028033929
9	200	0.8	0.6	0.5867102587602677
10	200	0.8	0.8	0.5866896155206477
11	200	0.8	1.0	0.5868599336087861
12	200	0.9	0.6	0.5868187773976103
13	200	0.9	0.8	0.586774476729484
14	200	0.9	1.0	0.5869875794626516
15	200	1.0	0.6	0.5869545868239956
16	200	1.0	0.8	0.5869062287139597
17	200	1.0	1.0	0.5871591986955973

Validation Score for NN Models

Model	Number of layers	Number of neurons in each layer	NWRMSLE
1	7	512 x 256 x 128 x 64 x 32 x 16 x 1	0.6143938249014856
2	8	512 x 512 x 256 x 128 x 64 x 32 x 16 x 1	0.6158113158278086
3	6	256 x 128 x 64 x 32 x 16 x 1	0.6098121112836306