CECS 551 Artificial Intelligence

Coffee Beans Sales Analysis and Recommendation

Supervisors
Dr. Mahshid Fardadi
Dr. Allen Bolourchi

SPRING 2022

Contents

Abstract

Coffee Shop is a hypothetical store which has multiple products sold across various geographical locations. There are 132 stores of an international coffee shop chain with more than 1000 stores worldwide and $400M revenue.

Over a period of time there is a slight fall in the revenue across few of the stores. As a precautionary and preventive measure the organization decided to consult data scientists to understand the reason behind the slight decline in sales and expecting a few recommendation which will help improve the overall sales across the stores. The purpose of the analysis is to use the sales prediction and inventory alignment to improve sales and reduce waste, resulting in a more efficient operation.

We analyze the data for coffee shop stores and provide few initiative which help maximize the overall sales and profit. This document summarizes our finding.

1 Introduction

In a team of four, in section one, two members mainly focus on analyzing the data of three specified stores (StoreID=18, 117, 332) and two persons mainly focuses on the data of the entire 132 stores. Single store results and corporate-wide results should be communicated between team members to help them with their individual analysis and ensure consistency. But you will be graded as one team. In section two and three all team members work together.

1.1 Problem Statement

The final project is designed to implement two-week sprints of the scrum process, mimicking a real tech company machine learning or software development team environment. Please see the timetable below for more details. It is suggested that one person plays the role of the scrum master to coordinate the communications between team members and ensure on-time delivery at the end of each 2-week sprint.

1.2 Dataset Description

The table below describes the features of the dataset for coffee shop. We will be considering only 13 products for the initial analysis.

Features	Description
StoreID	ID of Store
PLU	ID of inventory
Description	Name of product
ItemType	Type of product
CategoryLvl1Desc	Main level of product's category
CategoryLvl2Desc	The 2nd level of product's category
CategoryLvl3Desc	The 3rd level of product's category
ReceivedQuantity	The amount stores received from the distributor.
SoldQuantity	Quantity sold on a particular day.
EndQuantity	Quantity at the end of the day – inventory condition.
LatestOrder	The number of items they requested.
StockedOut	They record when a customer asks for an out-of-stock item.
GroupID	Not Applicable
MissedSales	Not sure what these data signify.
BusinessDate	Date of record

Table 1: Coffee Store data description.

1.3 Proposed Workflow

We perform the analysis in three phases.

1. Data exploratory analysis.
2. Create a machine leaning model for sales prediction.
3. Inference and recommendation to maximize the profit.

1.3.1 Phase 1

Each phase is covered in individual chapters. The first phase tires to understand the data distribution, relationship among them using correlation matrix and SHAP feature interactions.

- Rank the products based on their sales (best seller and worst seller products), identify where/when the store gets rid of the unpurchased products.
- Investigate whether drive thru feature causes certain products to sell better or worse.
- Understand if the seasonal change impacts the sales of certain product.
- Draw conclusions and suggest a recommendation to optimize the stocking

1.3.2 Phase 2

The purpose of second phase is to come up with a predictor for sales of each product based on which we can optimize the restocking and inventory management of stores 18, 117, 332. We evaluate the impact of GAN on the existing dataset and overall performance of prediction mode, and concluded that adding GAN dataset doesn't help much. We also design a regression based predictive model for understanding the sales (SoldQuantity) based on various below machine learning models.

- Random Forest.
- Gradient Boosting Machine
- Light GBM
- XGBoost
- LSTM

1.3.3 Phase 3

One of the goal is to provide insight and recommend few initiatives based on the sales predictions model from the previous phases and provide minimum of three inventory optimization initiatives and examine them on the data.

We also deploy the model on a dashboard for visualization.

2 Exploratory Data Analysis

We analyze the best seller and worst seller products across the various stores and provide our analysis.

2.1 Product analysis based on inventory

For all stores except 18, 117, and 332, we can visually determine which stores have the highest average received quantities over the span of one year. For products which appear to have some of the highest received quantities such as the everything bagel, butter croissant, and danish classic cheese, this may indicate a high demand for these products as the stores need to stock up or a precursor to how much waste is generated by not selling enough of said product. Likewise, we can apply the same logic towards the lower end of products such as the caprese sandwich and pretzel egg sandwich, where perhaps there is not enough demand for these items, resulting in lower received quantities.

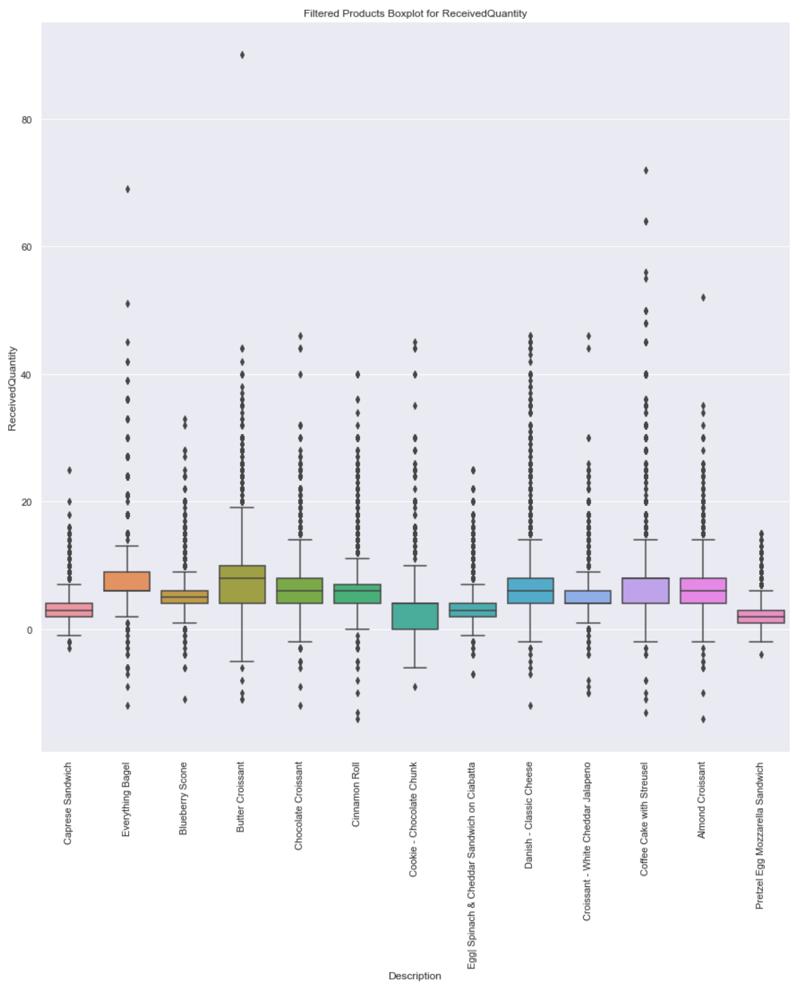


Figure 1: Analysis for highest average received quantities

The last detail worth mentioning about this graph is that every product present in these stores accordingly reaches at least one negative value in their received quantity. What this indicates to us is perhaps products that may never made it to stores during transit or lost in some way. However, it is worth mentioning that by manually searching through said data, these data points are few and far in between, so this may not be an issue to be concerned with overall, but still possibly worth investigating.

Figure 2 graph may be the most impactful in understanding the rank order of how consistently products sell overall compared to one another, and may indicate preferences among customers. For example, the butter croissant, everything bagel, and coffee cake appear to be the most consistent products in this list, with decently well situated averages as shown in the box plot. This may be a suggestion to look at these products more closely to try and understand how to further increase sales due to their sell out quantities. Our last note on this is to see how the received quantities

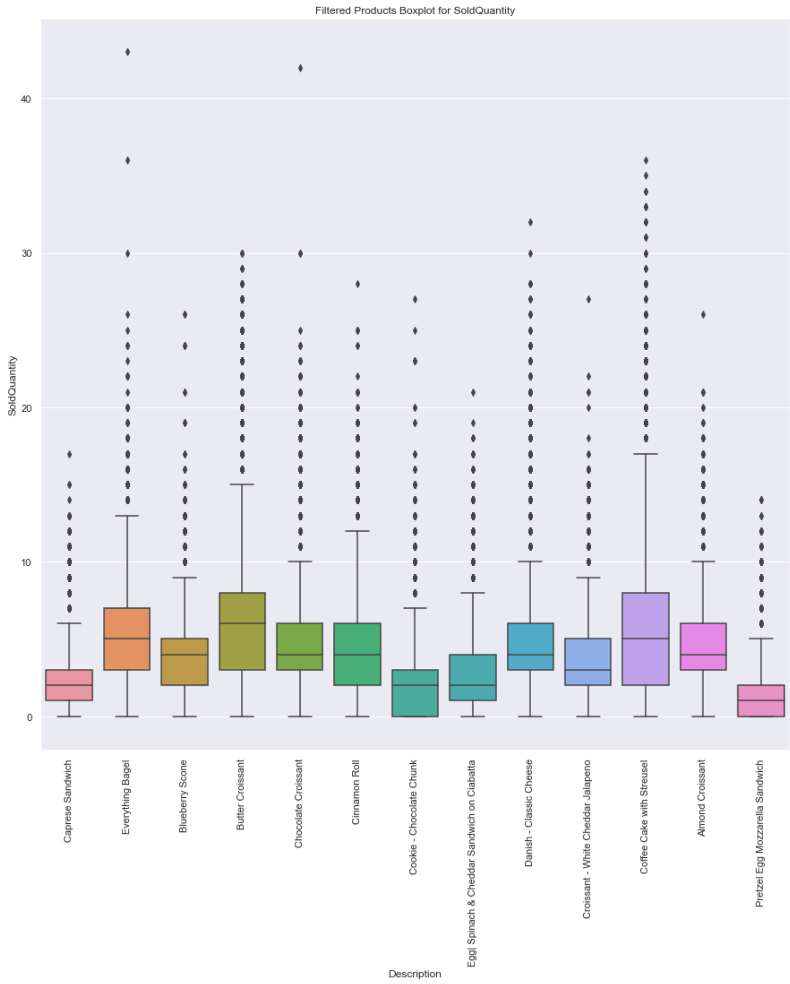


Figure 2: Analysis for highest average received quantities

of all products compare to their sold quantities, as we would preferably want to see a close relation between the two. Previously we have already mentioned that butter croissants and everything bagel have high received quantities, possibly indicating either a high demand or a precursor to waste. Here we can more confidently say that since the sold quantities of said products appear to be fairly consistent, we may believe that those products are not being wasted or tossed.

Figure 3 shows the End quantity where things get much more theoretical, as this graph is radically different compared to the previous two. For one, the majority of the products have their boxes appear as just a line located at the zero mark. This tells us that on average, these products wind up with zero end quantity, which may tell us that these products give off zero waste on average. For a business oriented around selling products, this is a very good indicator that those particular products are not being needlessly tossed away.

The second point of interest are the products which do have boxes present, as each box indicates an average above the zero mark. Considering that this is data over a time span of one year, having an average higher than zero may indicate incredibly high waste of those products, as on average each day at least some of those products are being tossed away. This is a sign to review those products much more closely, and see why each product is being tossed to the bin.

Lastly are the negative values present underneath each product, similar to the received quantity graph. To have negative values indicates to us that perhaps there was a certain amount of demand for those products that were not met, as in a customer wanting said product but it was out of stock. This portion is fairly theoretical however, as we are unsure if said information is being tracked by the stores.

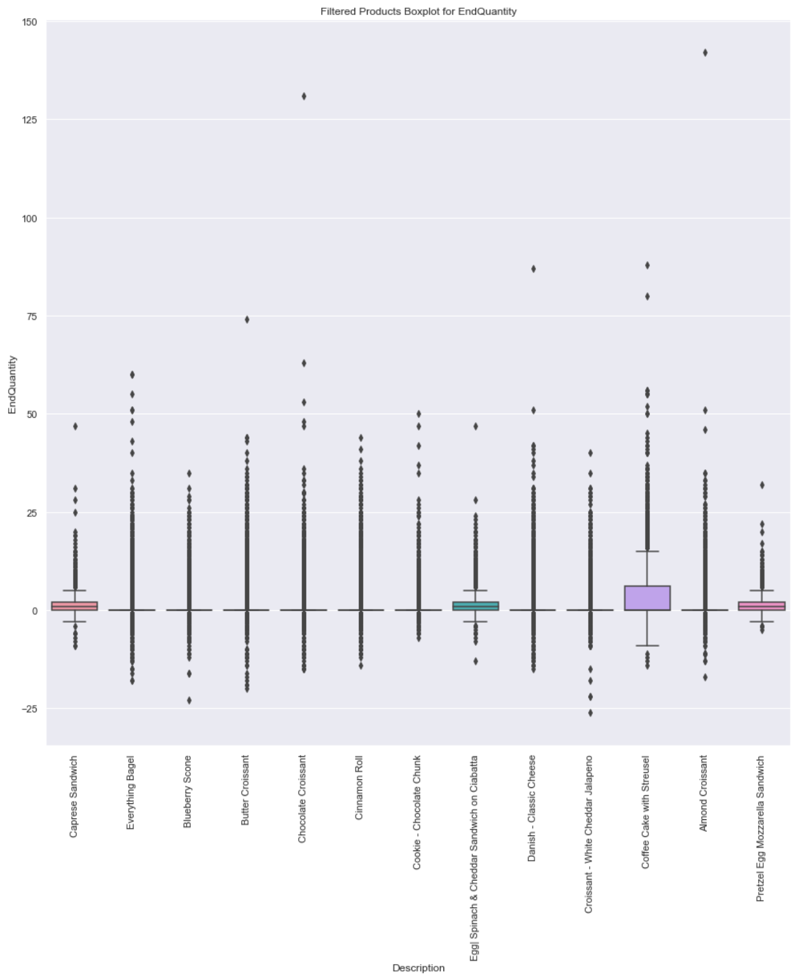


Figure 3: Analysis for highest average received quantities

For missed sales, we can infer which products may prove to have a slightly higher demand than expected. For products with little or no visible average, we may infer that stores are meeting the demand of customers, perhaps noting a more or less stable inventory of said products. For other products however, having high or often missed sales may mean a re-evaluation on the stocking patterns of these products, as perhaps these products in their current ordering quantities may not be meeting certain demand.

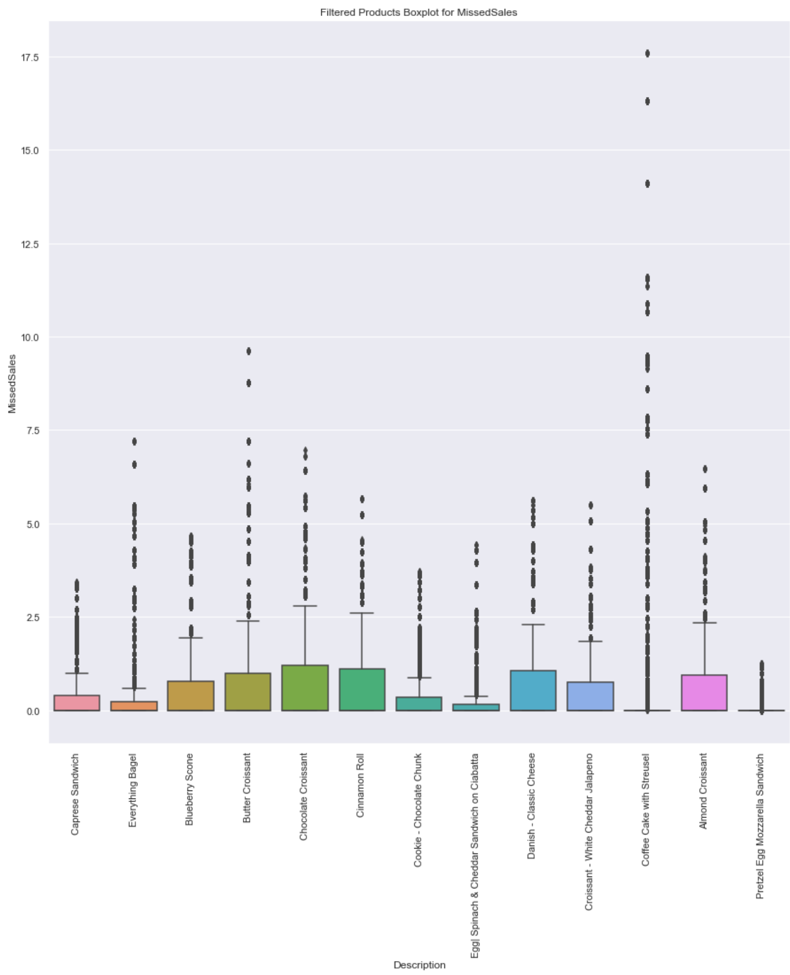


Figure 4: Analysis for highest average received quantities

2.2 Top 25% of the product

We try to further understand the sales trends of the top 25% of products by visualizing a breakdown of how often these products are received in stores and determine if any pattern or trend can be understood. In this example, we see that for the top 25% of products there are two patterns, either the amount of products received is very consistent as we see with the coffee cake and everything bagel, or there is a small variance in the amount of product received. It is also worth noting that these product's received count lie in the range around 5 to 13, true for the top 3 products.

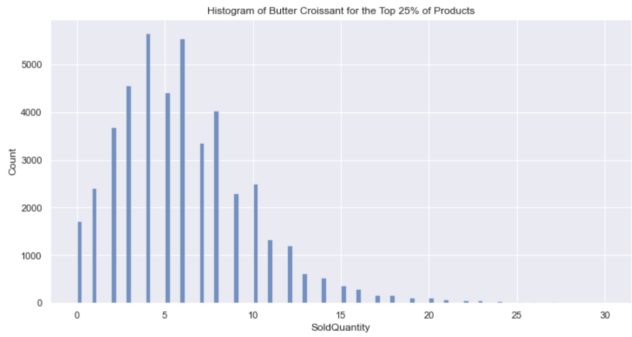


Figure 5: Histogram of Butter Croissant

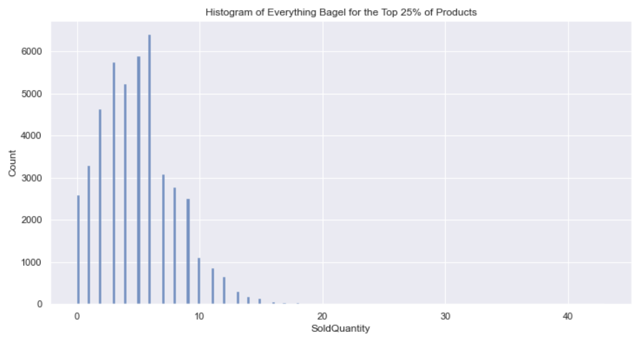


Figure 6: Histogram of Bagel

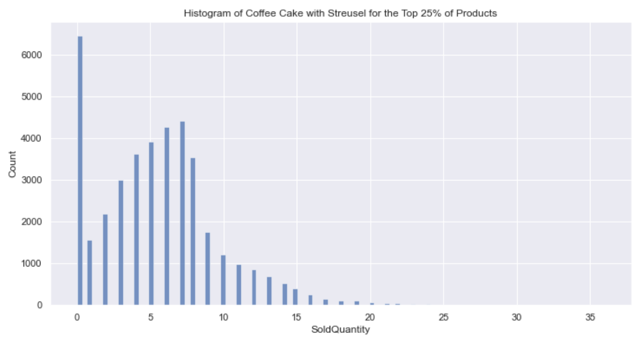


Figure 7: Histogram of Coffee Cake

Continuing on we look at the sold quantities of these products, and immediately notice a consistent bell curve among all three products peaking between 5 and 8.

Here we also want to establish a factor which determines if the products are doing well, and in our opinion we believe that a graph which surpasses half of the count at zero is a good indicator of product performance. When comparing these graphs, what sticks out greatly is the X axis at zero for the coffee cake with streusel graph, as it peaks much higher than the curve's peak. This may bear investigation later.

The last chart worth looking at is the end quantity of these products, and in our investigation we notice a very great trend towards having zero end quantity per product. Of course, there will be minor outliers which tell us that some products were left over, and out of the top 25% of products it appears that the coffee cake with streusel does tend to have the most leftover products. However, we do investigate this further into the report.

2.3 Bottom 25% of the product

Continuing on with our investigation, we inspected the bottom 25% of products to view their patterns, and we immediately noticed how much lower the received quantities for these products are compared to the top 25%. Here we see a trend skewing slightly closer to zero, indicating perhaps a lower demand or need of these products across all stores

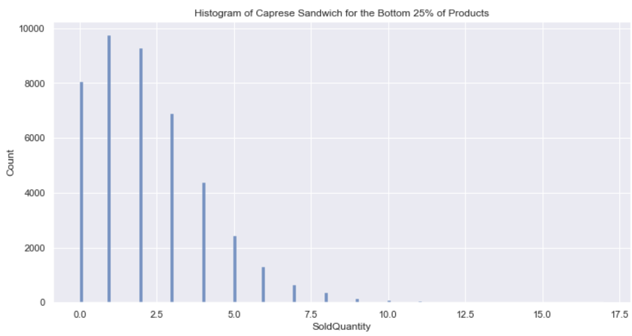


Figure 8: Histogram of Caprese

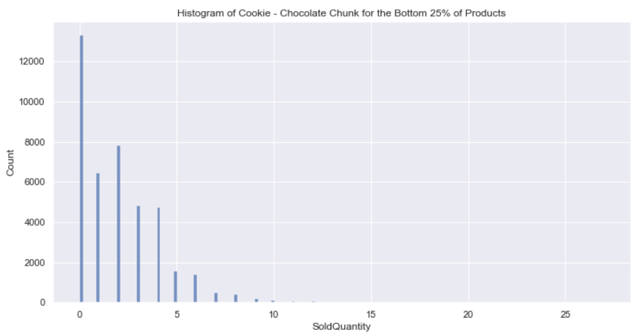


Figure 9: Histogram of Cookie

Of course, considering these to be the bottom 25% of products, we notice just how much lower they tend to sell. Compared to the top 25% of products, these products tend to linearly sell much less and present very little or no curve to follow.

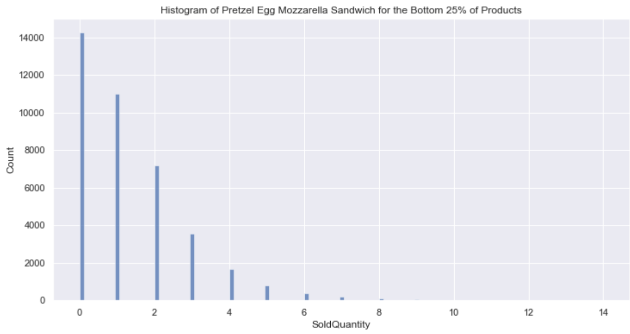


Figure 10: Histogram of Mozzarella

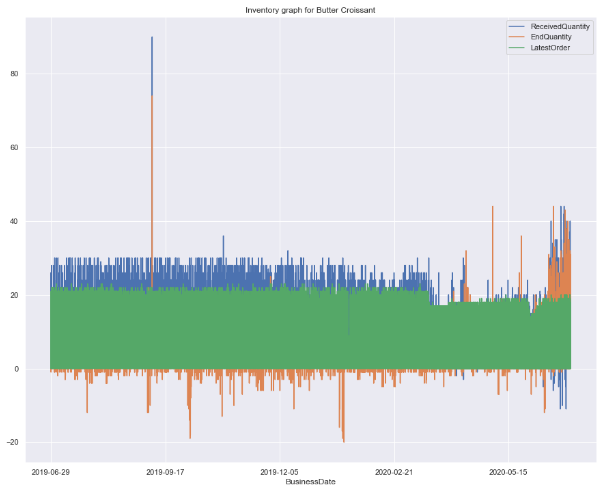
For example the pretzel egg sandwich and the chocolate cookie commonly does not sell as a product, which should give off red flags and demand evaluation in stocking quantities.

Lastly, we want to carefully examine the end quantities of these products and see if these products produce more waste. For context, we want the highest peak to be at zero, with all other values to be very very low. Unfortunately, both the pretzel egg and caprese sandwich tend to produce a lot of waste, resulting in tossing of these items and loss of sales.

2.4 Analysis

For all stores except 17, 118, and 332 we can compare differences between the top 25% and bottom 25% of products based on sales and understand the differences in stocking, selling, receiving, and how often the products are sold out in general. For each product in top 25%, each store will sell, in average, $16.50 worth of product and retain approximately $4.80 worth of product by the end of each day. In comparison, the bottom 25% of products only sell in average $8.10 worth of product and retain $3 by the end of each day. This comes into play when we note that the stores order around $35.40 worth of product for the top 25% of products while stores order $28.44 worth of product for the bottom 25% of product. With this we understand that there exist an unnecessary expense of $17.44 when purchasing products from the bottom 25% category compared to an expense cost of $14.10. We can further break this down by creating a line graph of each product to make inventory trends much more apparent. We first apply this to the top 25% of products.

- Here we see that in June 2020 the end quantity for our products are very high compared to sold quantity.
- For the product Everything Bagel in the month of Jan 2020 The quantity received was very high and thus the end quantity was high where as the sales was less.
- Not all products in these categories were chosen, however these two figures best represent the trends found in both products.



-1

Figure 11: Inventory of Butter Croissant (Top 25%)

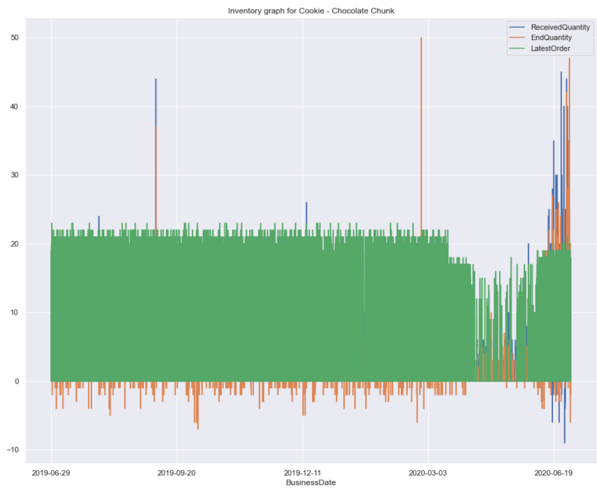


Figure 12: Inventory of Cookie - Chocolate Chunk (Bottom 25%)

- Lastly, there is a noticeable gap at around June 2020. While we have no solid evidence, this may have been due to real world impacts as at around that time the Covid-19 pandemic begun impacting stores and distribution around the world. Only a few products reflect this trend.

As we can see, products in the top 25% category show healthier inventory trends compared to the bottom 25% category. Products in the top 25% category tend to place lower orders than the products received, retaining less product stock and resulting in the items being out of stock more often as noted by the end quantity. However for products in the bottom 25% category we see order amounts outweigh the amount receive, resulting in higher retention of products and leading to higher end quantities in these products. This may demonstrate an unnecessary amount of overstocking in the bottom 25% of products, requiring certain re-evaluation when choosing to re-order these items.

Lastly, we must estimate the amount of sales loss per event when product is unavailable. To achieve this, we scan through the data for entries which state the product is out of stock. If so, we note down the date and the store ID which ran out of stock of the product. Using the date, we calculate the dates of the previous four weeks and store those dates in a list. Lastly, we can filter through the data and scan for the previous four dates alongside the store ID and obtain the SoldQuantity entry for each week. We will use these values to calculate the average amount of product the given store ID has sold, and make the assumption that for the current date the store has missed out on 75% of those sales. We go through the entire data frame one product at a time, summing each of those values and get a rough estimate of how much missed sales each of those products encountered over the course of one year.

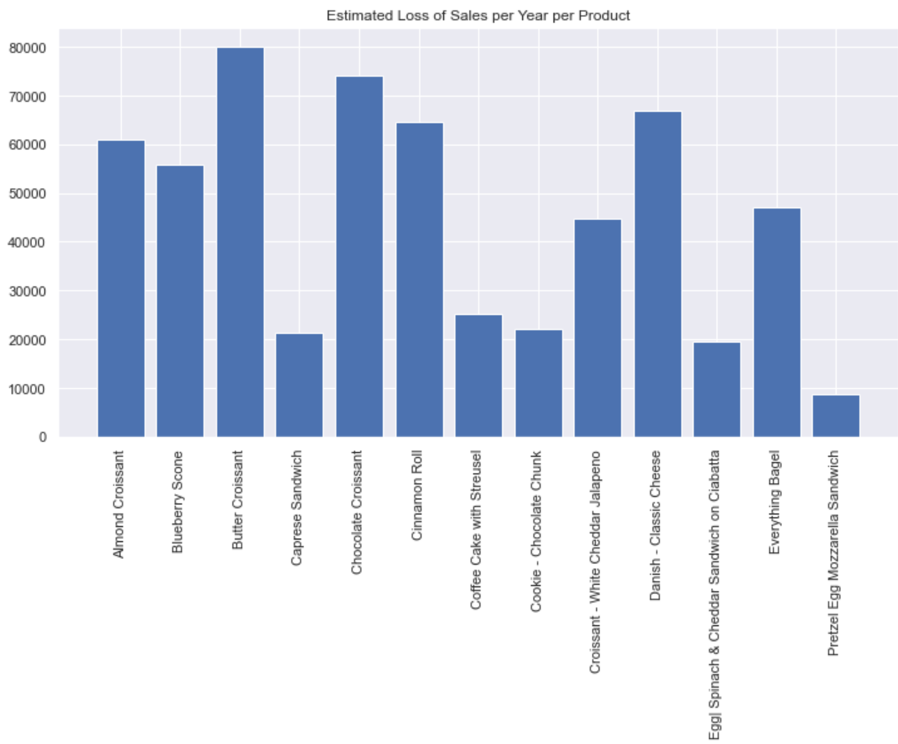


Figure 13: Estimated Loss of Sales in Volume

3 Sales Prediction

3.1 Exploratory data analysis

We analyze the features of the coffee store dataset using SHAP analysis.

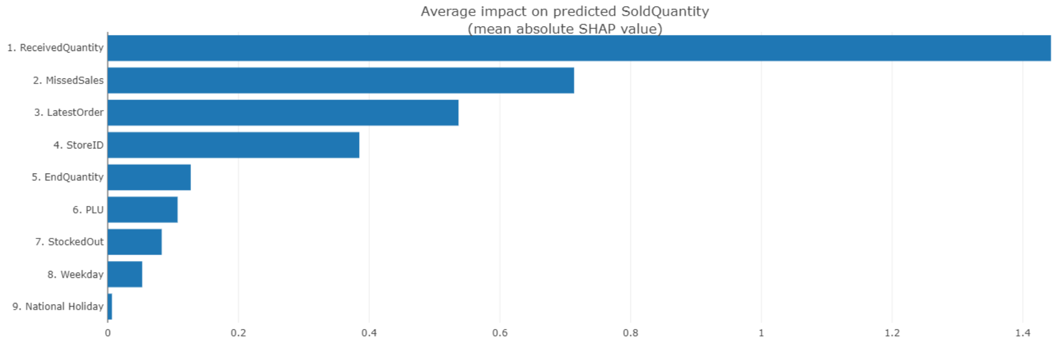


Figure 14: Shows the features sorted from most important to least important. Can be either sorted by absolute SHAP value (average absolute impact of the feature on final prediction) or by permutation importance (how much does the model get worse when you shuffle this feature, rendering it useless)

Figure 15: Plot shows the observed SoldQuantity and the predicted SoldQuantity in the same plot. A perfect model would have all the points on the diagonal (predicted matches observed). The further away point are from the diagonal the worse the model is in predicting SoldQuantity.

3.2 Partial dependence plot

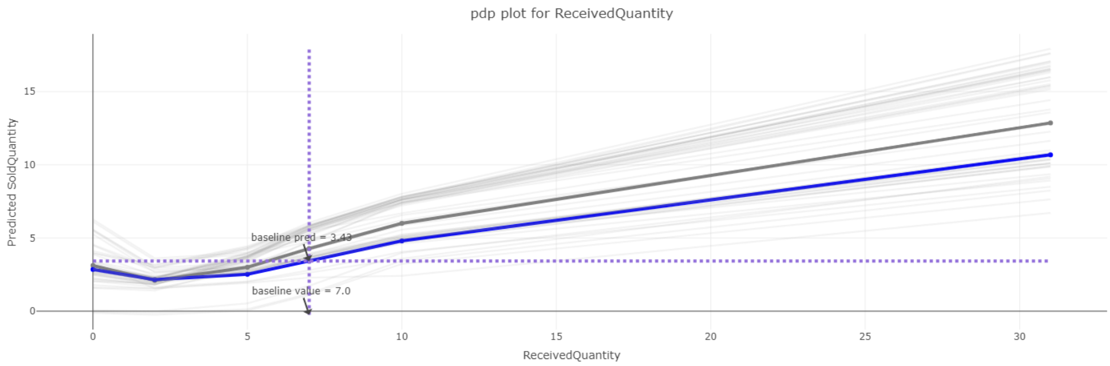


Figure 16: The partial dependence plot (pdp) show how the model prediction would change if you change one particular feature. The plot shows you a sample of observations and how these observations would change with this feature (gridlines). The average effect is shown in grey. The effect of changing the feature for a single Index is shown in blue. You can adjust how many observations to sample for the average, how many gridlines to show, and how many points along the x-axis to calculate model predictions for (gridpoints).

3.3 Predictive Model

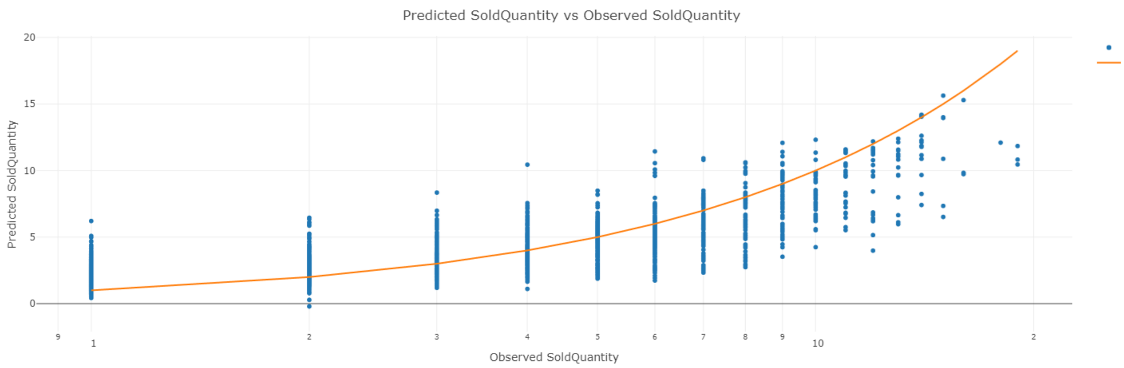


Figure 17: Plot shows the observed SoldQuantity and the predicted SoldQuantity in the same plot. A perfect model would have all the points on the diagonal (predicted matches observed). The further away point are from the diagonal the worse the model is in predicting SoldQuantity.

3.4 Algorithms

We create a predictive model for the sold quantity data using the below machine learning algorithms.

- Random Forest.
- Gradient Boosting Machine
- Light GBM
- XGBoost
- LSTM

After evaluating the mentioned machine learning models, the accuracy achieved among all models was between 75-90% accuracy. Light GBM XGBoost reported the best accuracy of 89%. At the same time, Random Forest and GBM reported 87.93% 81.68%, respectively.

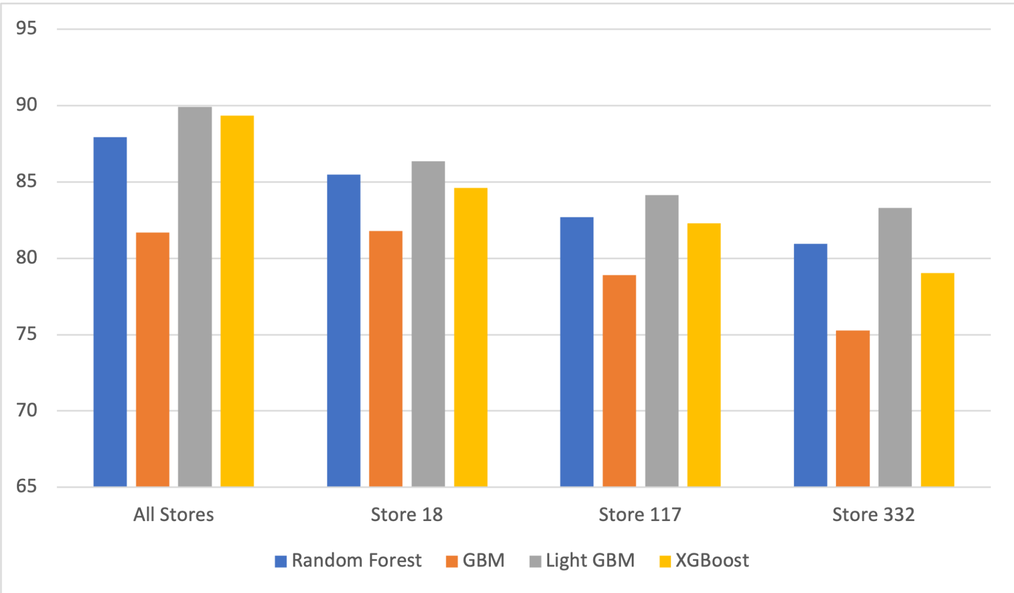


Figure 18: Accuracy comparison for three datastore using machine learning algorithms

	R Square
Random Forest	87.93
GBM	81.68
Light GBM	89.90
XGBoost	89.34

While the Light GBM and XGboost reported the best accuracies among all the best models, our study suggestes that Light GBM performed best for the dataset, as the execution time for Light GBM is seven times faster than XGboost.

	Accuracy	**Execution Time**
Light GBM	89.90	0.22 sec
XGBoost	89.34	12.74 sec

The graph below shows the actual and predicted values of XGBoost and Light GBM

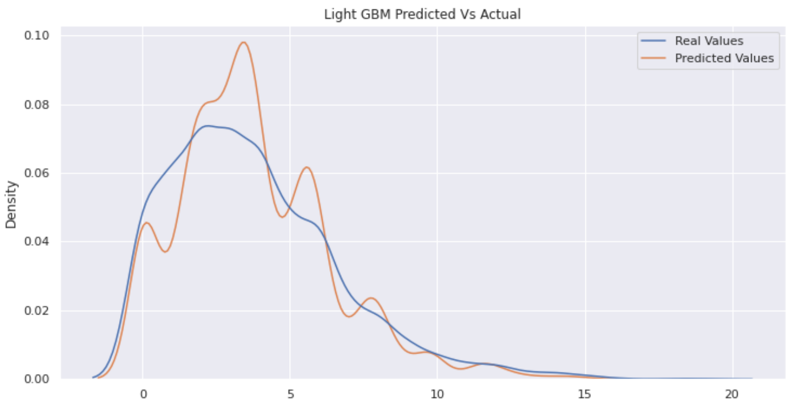


Figure 19: Light Gradient Boosting Machine

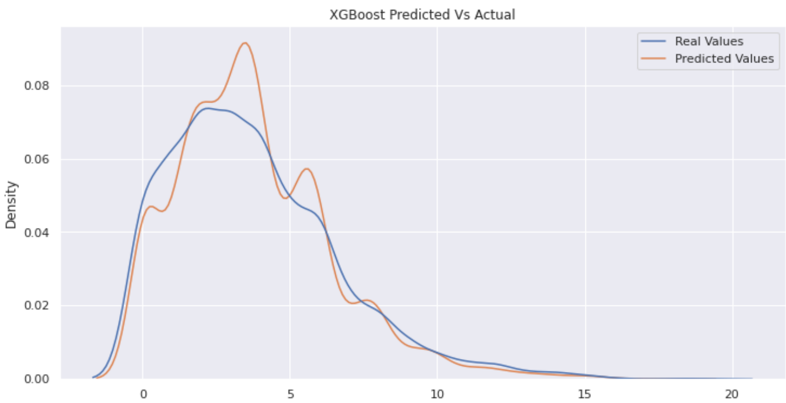


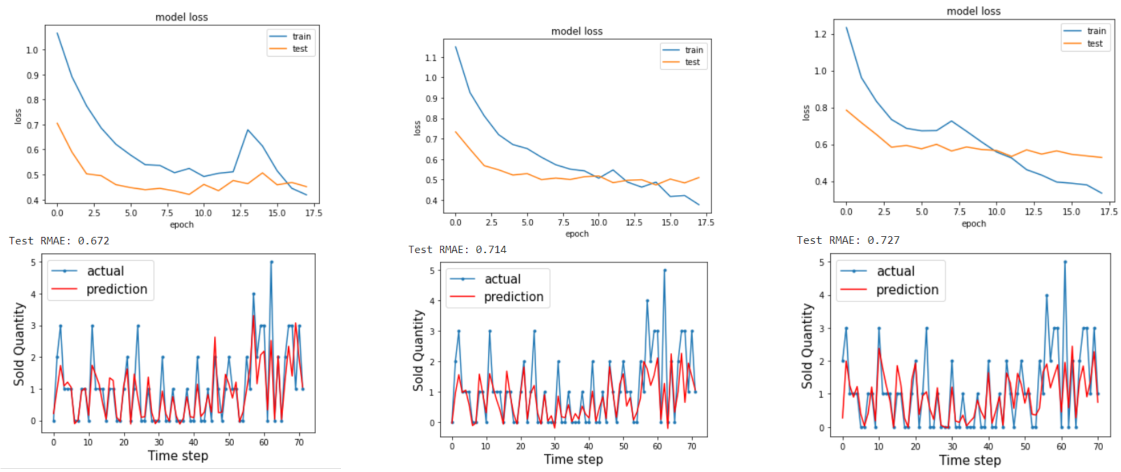
Figure 20: XGBoost

We trained LSTM to predict 1-day, 3-day and 10-day predictions based on the original coffee store data. We split train test by 80/20 and obtained the mean absolute error for each model as shown in Table 2.

Store	1 day prediction	3 day prediction	10 day prediction
18	0.9	0.98	1.04
117	0.94	0.98	1.02
332	1.44	1.5	1.44

Table 2: LSTM model MSE on Test dataset

When using "day of the week" to do prediction, 1-day ahead prediction is more accurate (less error); When added weather data, long term prediction (10-day ahead) is more accurate/less error. Figure 21 shows the prediction on test data set for store 18 for the product of Caprese Sandwich. These findings are consistent across three stores.



(a) 1-day prediction (b) 3-day prediction (c) 10-day prediction

Figure 21: Benchmark of Prediction: 1-day vs. 3-day, vs. 10-day

3.5 Algorithms with synthetic data

The composition of our dataset consists of over 132 different stores, each selling and managing the inventory of their respective 27 list of products. Our models in the future should attempt to correlate all kinds of information to the sales performance of their products, which include their inventory patterns, the date, and potentially even the weather to predict which trend may improve sales performance. However, out of all 132 stores we only had access to the weather information of 3 stores, severely stunting the amount of data points we can use to train a model with. To fix this we tried applying a Generative Adversarial Network (GAN) to the select 3 stores and create more weather information for use in our predictive models.

However, while the execution of our GAN model did generate additional 20 thousand data points, it resulted in mostly random information that serve to confuse our predictive models.

	Random Forest	**GBM**	**Light GBM**	**XGBoost**
All Stores	68.35	53.23	72.4	71.02
Store 18	63.45	53.63	67.43	63.61
Store 117	48.46	32.70	52.87	52.48
Store 332	62.32	47.07	67.22	63.79

As you can see, the model's accuracy ratings are greatly impacted using GAN data, with some models like the GBM experiencing reductions of over 25%. For this reason, we opted to not include GAN when training our models. This is likely the result of an implementation error, as the library used at the time appears to have numerous errors due to updates over time.

4 Insight and Recommendation

4.1 Inventory optimization

We developed three inventory initiatives following below process to show how much extra sales that each initiative can facilitate. Followed with a cost-benefit analysis for each initiative.

- You can rewrite the ReceivedQuantity feature based on each initiative, keep the SoldQuantity, re-calculate the EndQuantity, and apply your forecasting models.
- In the next step, rewrite the SoldQuantity based on the Sales forecasting models and mention if the ML models result in a conservative estimate or not. Discuss the impact of your three initiatives separately.
- Show how much extra sales you can have based on each inventory optimization initiative. Meantime, we obtained the cost per each initiative based on the following algorithm.
- Do a cost benefit analysis. We compare the profits in each initiative with the original coffee store profit, to suggest the effectiveness of each initiatives.

4.2 Proposed Initiatives to increase the Sales

- Initiative 1: Increase product Received Quantity during weekends holidays, adjust End Quantity accordingly
- Initiative 2: Correlate products based on temperature (Example: Increase Received Quantity for cold products on cooler weather days)
- Initiative 3: Increase all products received quantity by 20%, adjust end quantity accordingly.

4.3 Proposed Algorithm to Obtain Cost

- We experimented with various algorithm to obtain cost, considering the on-shelf life cycle of each product.
- We found that the waste can be obtained by getting the difference between Received Quantity and Sold Quantity during a fixed period of time disregarding each product's on-shelf life cycle.
- In order to calculate the waste, however, it has to be obtained product by product.

We propose the following algorithm for obtaining waste.

```python
def get_waste_sales(df):
  waste_by_products = {}
  total_waste, total_sales = 0, 0
  PLUs = df['PLU'].unique()
  products = df['Description'].unique()
  for i, product in enumerate(products):
    df_name = "df" + str(PLUs[i])
    df_name = df[df['Description'] == product]

    ReceivedQ = df_name['ReceivedQuantity'].values
    EndQ = df_name['EndQuantity'].values
    SoldQ = df_name['PredictedSoldQuantity'].values

    #track the last record in this whole period for EndQuantity
    #those are still in the stock (not thrown away)
    lastEndQ = EndQ[-1]
    sum_ReceivedQ = sum(ReceivedQ)
    sum_EndQ = sum(EndQ)
    sum_SoldQ = sum(SoldQ)
    waste = sum_ReceivedQ - sum_SoldQ - lastEndQ
    total_waste += waste

    #accumulate total sales for all product for this store
    total_sales += sum_SoldQ

    #save the waste and sales per product to a dictionary
    waste_by_products[product] = (waste, sum_SoldQ)

  return total_waste, total_sales, waste_by_products
```

4.4 Summary - Cost Benefit Analysis among original data and three initiatives for Store 18, 117, 332

We applied three inventory improvement initiatives. As below report shows, all of the three initiatives boost up sales more than the increment of wastes to finally increase profit. Table 3 summarizes the overall growth of profits for the three initiatives comparing with base line from the original coffee store dataset.

Store ID	Original	Initiative 01	Initiative 02	Initiative 03
18	baseline	3%	2.9%	10%
117	baseline	2%	0.14%	6%
332	baseline	4%	0.6%	15%
ten stores	baseline	4%	N/A	14%
all stores	baseline	9%	N/A	31%

Table 3: Profit Growth Benchmark between Initiatives vs. Original coffee store data. Profit is calculated as the total sales minus the waste. We applied $3 as the average selling price to calculate the revenue, $0.5 as the waste unit cost to calculate Wastes.

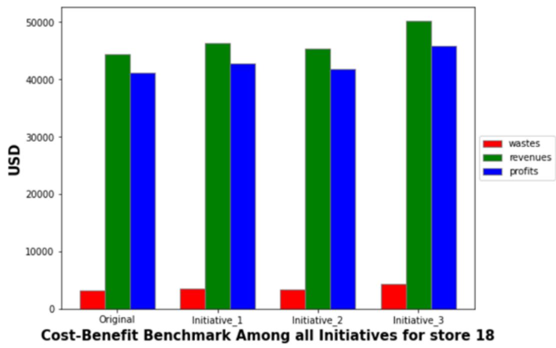
4.4.1 Store 18

Table 4 shows the wastes, predicted sales, and predicted profits after applying three initiatives. We can see initiative 3 facilitates 10% increase in profit.

Metrics	**Original**	**Initiative 1**	**Initiative 2**	**Initiative 3**
Wastes ($)	3221	3595	3451	4352
Sales ($)	44382	46339	45276	50157
Profits ($)	41161	42744	41825	45805

Table 4: Wastes, Sales and Profits for Store 18

Profit growth:
Original coffee store profit: $40,644
Initiative 1 profit: 42,001, profit increased 3%
Initiative 2 profit: 41,825, profit increased 2.9%
Initiative 3 profit: 44,833, profit increased 10%



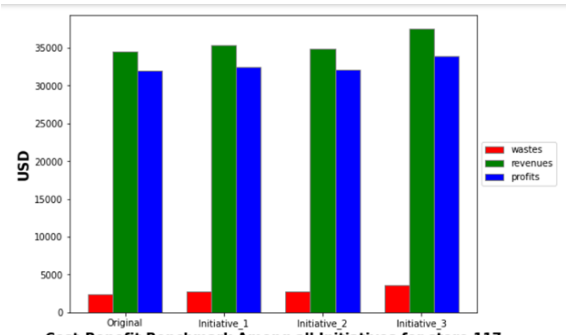
4.4.2 Store 117

Table 5 shows the wastes, predicted sales, and predicted profits after applying three initiatives. We can see initiative 3 facilitates 6% increase in profit.

Metrics	Original	Initiative 1	Initiative 2	Initiative 3
Wastes ($)	2454	2812	2794	3591
Sales ($)	34451	35289	34836	37476
Profits ($)	31997	32477	32042	33885

Table 5: Wastes, Sales and Profits for Store 117

Profit growth:
original coffee store profit: $ 31997
Initiative 1 profit: $ 32477, profit increased 2%
Initiative 2 profit: $ 32042, profit increased 0.14%
Initiative 3 profit: $ 33885, profit increased 6%



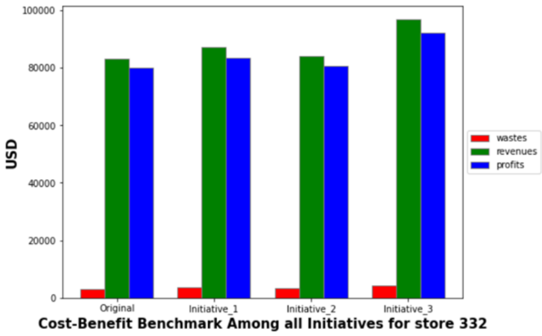
Cost-Benefit Benchmark Among all Initiatives for store 117

4.4.3 Store 332

Table 6 shows the wastes, predicted sales, and predicted profits after applying three initiatives. We can see initiative 3 facilitates 15% increase in profit.
Profit growth:
original coffee store profit: $ 79, 992
Initiative 1 profit: $ 83,445, profits increase 4%
Initiative 2 profit: $ 80.498, profits increase 0.6%
Initiative 3 profit: $ 92,272, profits increase 15%

Metrics	Original	Initiative 1	Initiative 2	Initiative 3
Wastes ($)	3203	3597	3502	4453
Sales ($)	83195	87042	84000	96725
Profits ($)	79992	83445	80498	92272

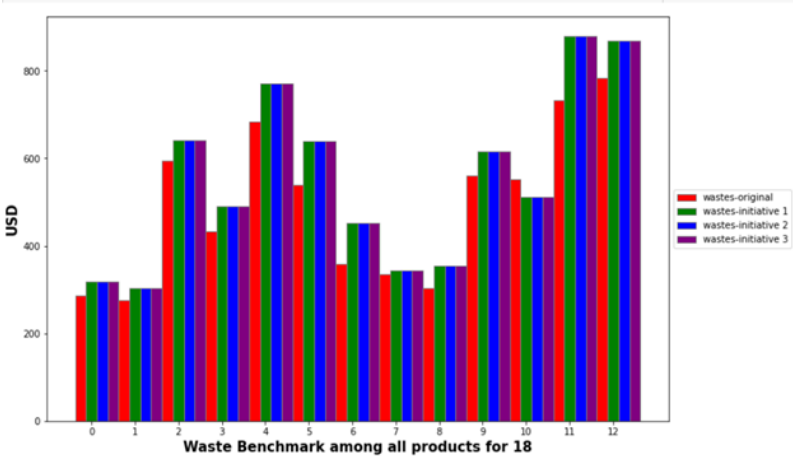
Table 6: Wastes, Sales and Profits for Store 332



4.5 Waster across product categories

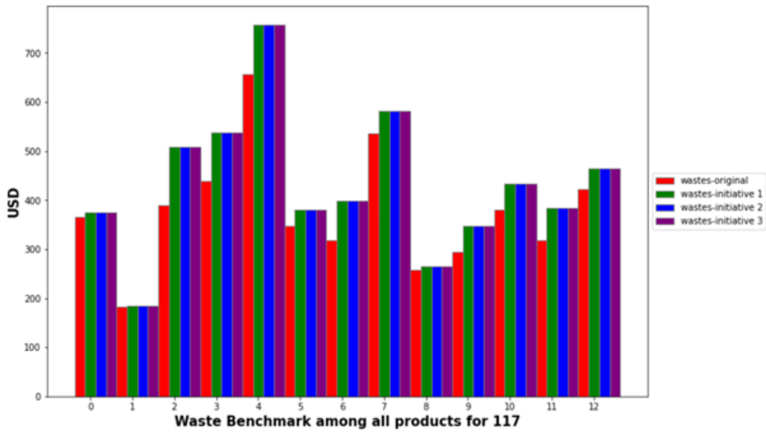
Store 18 waste per product

- Butter Croissant
- Coffee Cake with Streusel
- Almond Croissant



Store 117 waste per product

- Butter Croissant

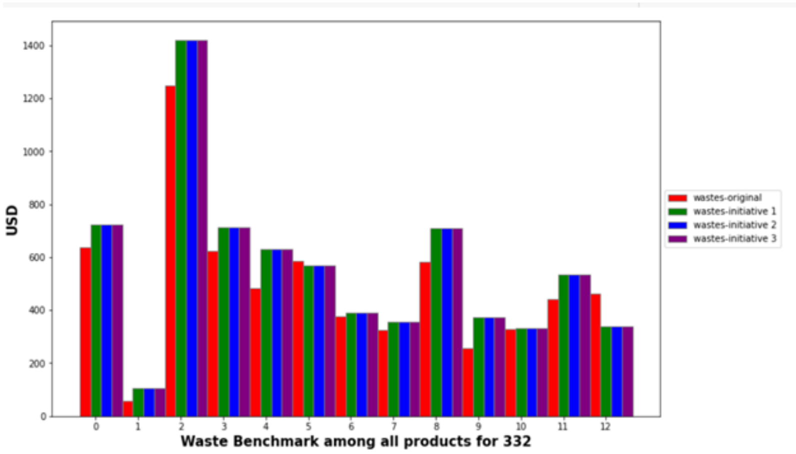


Store 332 waste per product

- Everything Bagel

4.6 Summary - Cost Benefit Analysis among original data and three initiatives for ten other stores and all stores

We extended the cost-benefit analysis for initiative 1 and initiative 3 to ten other stores and across all stores. We found that inventory initiatives consistently boost up sales and profit despite increase in wastes. Profit can grow up to 14% in initiative 3.



4.6.1 10 Other Stores

Key findings: - original coffee store profit: \$ 399,865.
- initiative 1 profit: \$ 415,563, profit increases 4%.
- initiative 3 profit: \$ 457,127, profit increases 14%.

Table 7 reports the Wastes, Sales and Profits we experimented on 10 other stores beyond 18, 117, and 332. The findings is consistent that the initiatives for inventory improvement can grow profit significantly.

Metrics	Original	Initiative 1	Initiative 3
Wastes (\$)	20626	23097	28609
Sales (\$)	420491	438660	485736
Profit (\$)	399865	415563	457127

Table 7: Wastes, Sales and Profits for 10 other stores.

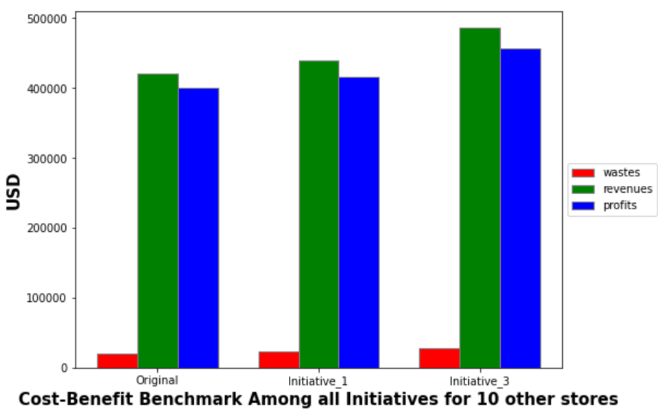


Figure 22: Cost-Benefit for Ten Other Stores: benchmark initiative 1 and 3 with original coffee store data. It shows initiative 3 increases 14% profit overall.

Our waste analysis across other ten stores show that Everything bagel and Blueberry Scone are on the top 2 waste in these stores.

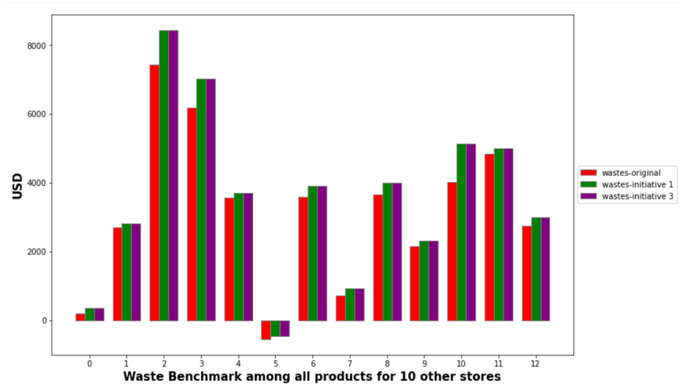


Figure 23: Waste Analysis for Ten Other Stores.

4.6.2 All Stores

Key findings:
- original coffee store profit: $ 6,638,263.
- initiative 1 profit: $ 7,252,146, profit increases 9%.
- initiative 3 profit: $ 8,680,860, profit increases 31%.

Table 8 reports the Wastes, Sales and Profits we experimented on all stores including 18, 117, and 332. The findings is consistent that the initiatives for inventory improvement can grow profit significantly.

Metrics	Original	Initiative 1	Initiative 3
Wastes ($)	349,419	341,087	320,937
Sales ($)	6,987,682	7,593,233	9,001,797
Profit ($)	6,638,263	7,252,146	8,680,860

Table 8: Wastes, Sales and Profits for all stores.

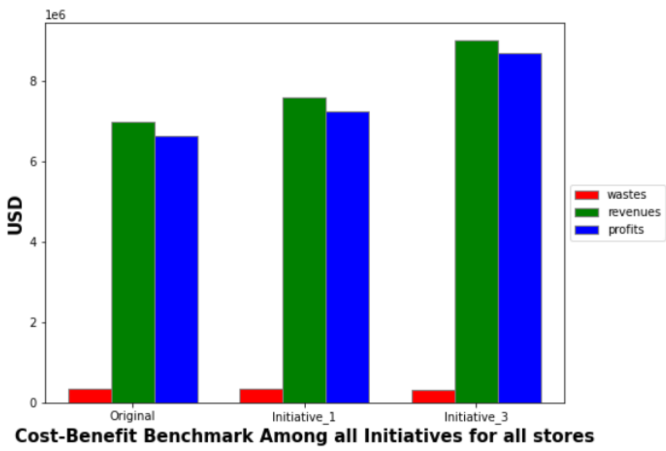


Figure 24: Cost-Benefit for ALL Stores: benchmark initiative 1 and 3 with original coffee store data. It shows initiative 3 increases 31% profit overall.

Our waste analysis across other ten stores show that Everything bagel remains the number one waste in all stores.

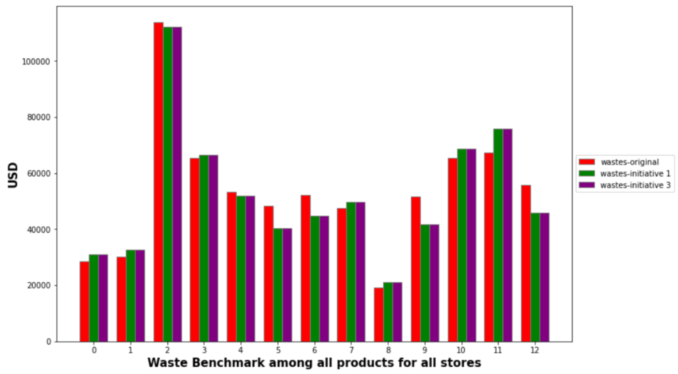


Figure 25: Waste Analysis for All Stores.

4.7 Are the models conservative?

Summary

- LSTM model is conservative to the different dataset. In general, the models are more conservative with the original dataset. With applying with initiatives (notice initiative 3 is more aggressive than initiative 1), we observe the model can predict to meet the higher demand of the market demands.

- Notice that there are outliers in the Sold Quantity (blue sharp up stream) that the model still misses. We believe those are triggers (similar to the initiatives we have applied, such as holidays, weather changes, weekday/weekends) that we can in further observe to apply more combinations to initiatives to catch the market needs.

- We notice that a more risk-taking (risks regarding to potential wastes) results to more sales and higher profit. As a summary, the LSTM model is strong to adapt to different datasets to predict.

Example: Store 18 of 820602 Everything Bagel The blue lines are the Received Quantity (either the original data or the adjusted data for each initiative); The read line are the predicted Sold Quantity for each dataset.

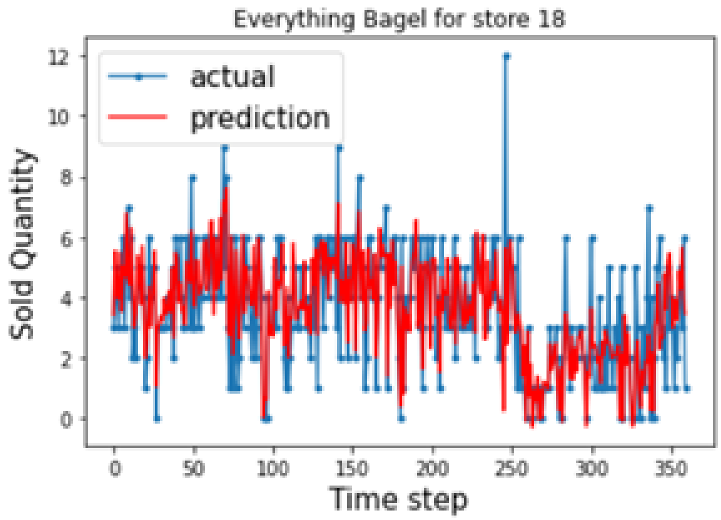


Figure 26: Original

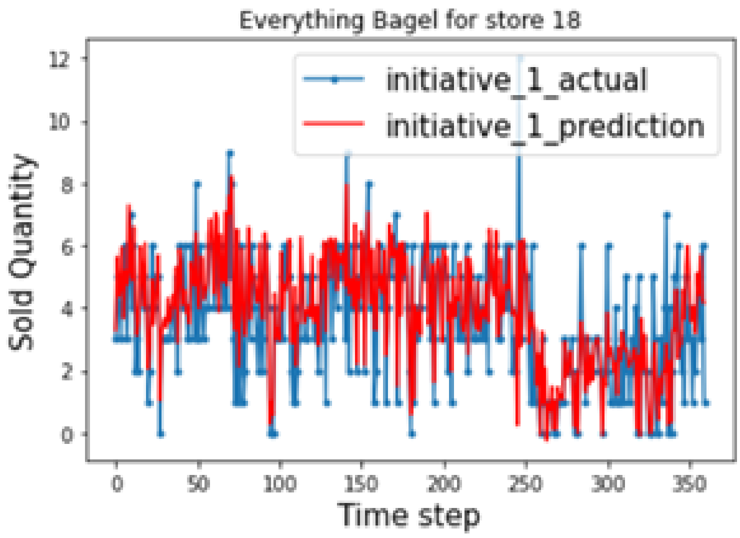


Figure 27: Initiative 1 (predicted sales based on the current initiative's estimated Received Quantity)

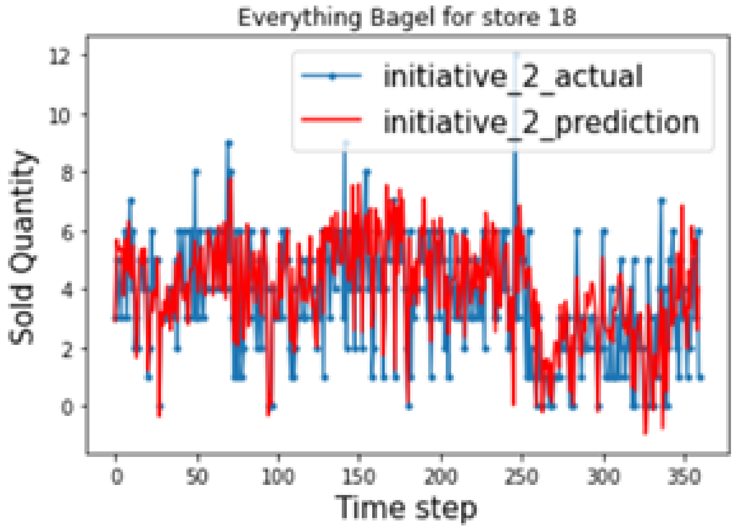


Figure 28: Initiative 2 (predicted sales based on the current initiative's estimated Received Quantity)

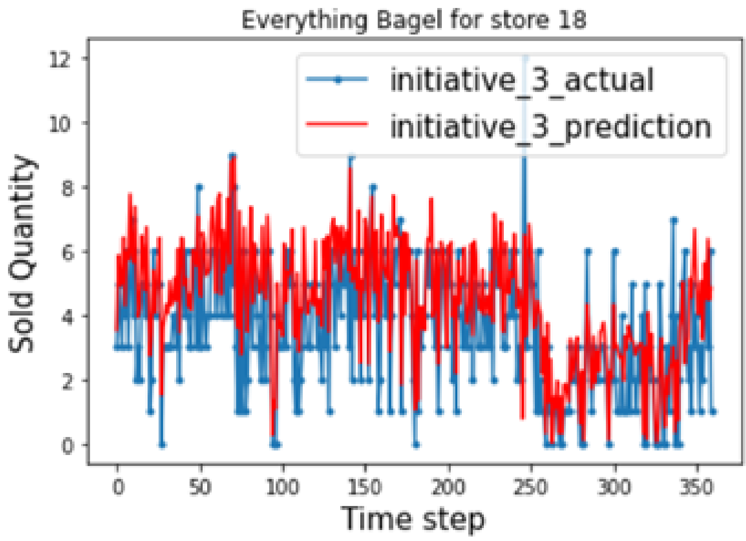


Figure 29: Initiative 3 (predicted sales based on the current initiative's estimated Received Quantity)

5 Conclusion

We used various regression algorithms to predict the sold quanity based on the given parameters. The results show the interest of this method on different stores along with the proposed initiatives.

We observed our models reduce in accuracy when using GAN on the coffee store dataset as with each iteration of generator over-optimizes for a particular discriminator, and the discriminator never manages to learn its way out of the trap.

Furthermore, we performed a comparative study of regression algorithms over the dataset and observed that Light GBM performed better as compared to other discussed algorithms in terms of accuracy and execution time. Moreover, hyperparameter tuning helped improve the accuracy of the model.

We found that inventory initiatives consistently boost up sales and profit despite increase in wastes for the third initiative, and consequently the profit grew up to 14%.

Finally, we deployed a model using explainerdashboard on localhost, but because of resource constraints we could not deploy on heroku app.

Key Findings

- The model(s) is conservative in predicting sales particularly the peak and low changes. We found defining some initiatives/stimulation patterns is a very helpful tactic to stimulate the model to take/recognize the pattern of peak and low (by realizing such feature patterns) of market demand. Due to the conservative nature of the model, increasing inventory by a flat rate (e.g. 20% in our initiative 3) drives the model to be closer to the market reality. Weekend/holiday can be combined with a flat raise to form a new initiative.

- Temperature is a complicated pattern, but we do think it is a helpful pattern to define a new initiative. It can be combined with all other initiatives together to reflect the real market demand.

6 Source Code

- LSTM-Store18 Business Initiatives to Optimize Inventory: [1]
- LSTM-Store117 Business Initiatives to Optimize Inventory: [2]
- LSTM-Store332 Business Initiatives to Optimize Inventory: [3]
- LSTM-10-Stores Business Initiatives to Optimize Inventory: [4]
- LSTM-All-Stores Business Initiatives to Optimize Inventory: [5]
- Phase-1 Data Distribution and Correlation: [6]
- Phase-2 Models to Predict Sales: [7]

7 Dashboard for Model

The model was deployed locally (localhost) and we recorded the screen for the visualization.
https://www.youtube.com/watch?v=g33QfydnPIk

8 Acknowledgement

Throughout this project's problem-solving and reporting process, we received a great deal of invaluable insight and guidance from Dr. Allen Bolourchi. We would like to thank him for sharing his expertise with us. He utilized his private time throughout the semester to discuss with us how to approach the problems, providing his professional insights in the ML/DL/Time Series field, and giving us feedback to improve our findings and reporting skills.

The project, to all of us, is huge and new. This is a tremendous learning experience to tackle an industry's real problems and we appreciate very much the opportunity to work with Dr. Bolourchi throughout this semester.

Meantime, we thank Dr. Mahshid Fardadi for giving us this opportunity in her class to tackle this real-world industry problem, and for connecting us with Dr. Bolourchi. We thank her support and assistance during the project and reporting process in this semester.

References

[1] Lstm store 18, . URL https://drive.google.com/file/d/1OdTsWemrl8erLxwRLZFcdwQTZyEdUyM4/view?usp=sharing.

[2] Lstm store 117, . URL https://drive.google.com/file/d/1ty_kFnYD7gb0RpLli3db68pro6aa6BG1/view?usp=sharing.

[3] Lstm store 332, . URL https://drive.google.com/file/d/1zG33UNRddTZvXDJy0QflEjUzxNI-S673/view?usp=sharing.

[4] Lstm ten other stores, . URL https://drive.google.com/file/d/1I9WfE8jWOjtSHtGr6y0cf82njHsn6H_v/view?usp=sharing.

[5] Lstm all stores, . URL https://drive.google.com/file/d/1vxZbSh2ITq45twKyYzI4QdAmnza5ntfx/view?usp=sharing.

[6] Phase 1, . URL https://colab.research.google.com/drive/1XUMfcz8bqwx1t9_YNGOtq7iJffKzdlOx?authuser=3.

[7] Phase 2, . URL https://colab.research.google.com/drive/1abUlGhniwMMGUmKZh6Mhx8ieRGXl5FCH?usp=sharing.