Promotional Forecasting Model for Profit Optimization

Yichen Isabel Zhou, Xinmeng Li, Tianshu Chu
Center for Data Science
New York University
New York, NY 10011
{yz6126, xl1575, tc2992}@nyu.edu

Abstract

In this project, we aimed to investigate different promotions' influence on monthly total sales and margins for the Shell Select convenient stores operated by a Brazil company, Grupo Nós. We explored the models, such as Ordinary Least Squares, Random Forest, SARIMAX, and LSTM, based on the company's one year of sales data on Red Bull products to predict total monthly sales and margins under different business strategies. The best model in our case is random forest, achieving the average monthly percentage error of about 3%. Finally, we recommended best promotional strategies based on the predicted results from random forest and built a pipeline to generalize our approach on other products.

1 Introduction

Promotion has become a common and powerful marketing tactic used by retailers to drive sales and margin. Different promotion strategies can have substantial influence on the company's total profits. The Brazil company named Grupo Nós (Group We in English), which was formed by Raizen and Femsa last year, took over the operation of Shell Select convenience stores in gas stations. They want us to create a promotional forecasting model to optimize promotion investments and to attract more customers while understanding price elasticity, cannibalization and optimal price point.

Most of the research in promotion forecasting are focused on causal inference and time series models. However, in this project, we are lack of insightful data for confounding variable analysis. Thus, we considered this problem more as a prediction task instead of causal inference. To be more specific, we investigated through different approaches, including time series, machine learning and deep learning algorithms, to predict total monthly sales and margin under different promotion strategies. Business insights on best promotion strategies are obtained according to models' predicted results. To simplify our task, we started with the sales and promotion data for only Red Bull products, and generalized the whole pipeline of data pre-processing, modeling and inferring strategies to other products.

The best model we trained is random forest, which achieved around 3% of percentage error in the prediction of both monthly sales and monthly margin for Red Bull products. This random forest model also provided us with a more flexible framework compared with other models we built. It could generate reasonable predictions when we alter the price level from original promotions, and thus can be utilized to set promotion strategies that maximize monthly sales and margin. This approach is relatively rare and naive in the area of promotional analysis, but it can be a reasonable method for limited and sparse data.

Submitted to 34th Conference on Neural Information Processing Systems (NeurIPS 2020). Do not distribute.

2 Related Work

Promotional Analysis and Forecasting for Demand Planning: A Practical Time Series Approach In paper [Leo01], the author uses traditional time series models to evaluate promotions by analyzing historical data. Our tasks are very similar so we can also use ARIMA models to forecast demand but one major difference is that they are able to perform intervention analysis. An intervention event is an input series that indicates the presence or absence of an event, which is a promotion in this case. Intervention effect is calculated by how the dependent variable differs in the treatment and control group when hold all other independent variables the same. Our data is real-world data (like an observational study), not an experiment, and we lack of store information that prevents us to pair up stores to study interventions.

Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry This paper [AAF16] also uses a time series model named Seasonal Autoregressive Integrated Moving Average with external variables (SARIMAX), which further accounts all the effects due to the demand influencing factors, to forecast the daily sales of perishable foods in a retail store. This model with external variables is able to outperform the traditional SARIMA model.

Forecasting at Scale Facebook introduces a modular regression model named The Prophet Forecasting Model with interpretable parameters for time series data in paper [TL18]. This model gives us flexibility to accommodate seasonality with multiple periods and interpolating missing values is not required. However, Prophet seems to have less power than SARIMAX because it cannot take account for exogenous variables. We later also see both weekly and monthly seasonality in the data and Prophet fails to capture multi-seasonality, either. As a result, we no longer consider Prophet.

Propensity Score–Matching Methods for Nonexperimental Causal We understand that promotions do not happen randomly and it is important to analyze the causal effect of these promotions. Paper [DW02] considers causal inference in a setting where few units in the non-experimental comparison group are comparable to the treatment units. However, we still lack of information to calculate store similarity so Propensity Score–Matching method is not applicable.

3 Problem Definition and Algorithm

3.1 Task

The whole framework for our task is shown in Figure 1. We have data of hundreds of products, but we have to look at each product individual. Let's start with an example - the energy drink red bull. We first combine and clean store, product, and promotion information. Next, we perform exploratory data analysis and detect seasonality, price elasticity, and correlation from the visualizations. Then we put everything into the models and Random Forest is the best performed model. From the best model, we obtain the optimal price and promotion strategy. Next, we can put all previous steps (in the yellow box) into a pipeline that can be applied for any product.

In the Modeling section from the framework in Figure 1, for all listed approaches, the main goal is to predict monthly sales and margins with separate models. Detailed implementation for each approach varies a lot and is explained in Section 3.2. In general, LSTM didn't perform well with our data, and thus we didn't further explore this algorithm. OLS is a naive baseline model, and failed to compete with random forest in terms of mean absolute percentage error (MAPE). SARIMAX is more suitable for this seasonal time series data forecasting, while Random Forest provides a more flexible architecture to evaluate the influence from various promotions.

3.2 Algorithm

3.2.1 Ordinary Least Squares

The main idea is to use the given store code, product code, weekday, month, price, and whether adhere to various promotions to predict daily sales and margin separately. The total monthly sales and margins are obtained through the summation of daily predictions group by month over all stores and products. Then, we just estimate the price level's influence and find best promotion strategy based on the fitted model's monthly predictions.

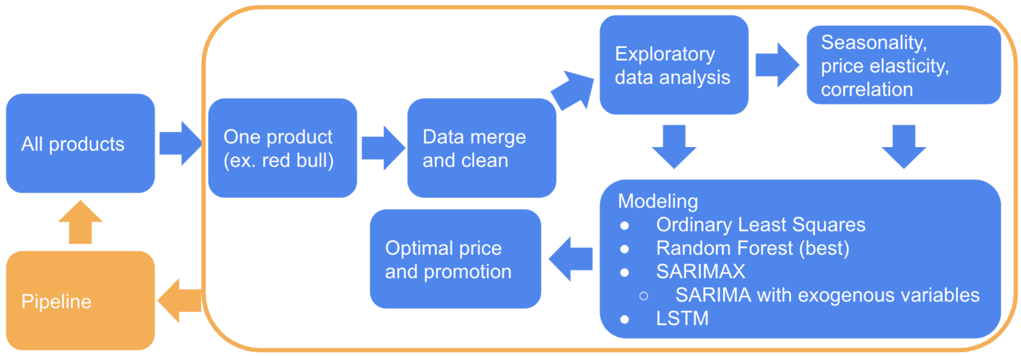

Figure 1: Framework

The Ordinary Least Squares regression model with k explanatory variables writes:

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon$$

where $X_1, \cdots, X_k$ - observations of k number of explanatory variables corresponding to the dependent variable Y
$\beta_1, \cdots, \beta_k$ - regression coefficients of explanatory variables
ϵ - the random error with expectation 0 and variance σ^2 (Normality Assumption)

3.2.2 Random Forest

The input and output are the same as in OLS. The monthly predictions are also obtained by adding up daily predictions over all stores and products. Two versions of random forest models Ver.Price and Ver.Percentage are designed to predict the impact of different promotion strategies. The Price version evaluate the influence when every store applies the same promotion price. Since the regular unit price for the same product varies across stores, the Percentage Version estimate the effect of applying the same discount, i.e. price might be different, across all of the stores.

The Following example illustrates how this algorithm works:

- During training:
  1. For b = 1 to B:
     (a) Draw a bootstrap sample Z^* of size N from the training data.
     (b) Grow a random-forest tree T_b to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the maximum depth of the tree is reached.
        i. Select m variables at random from the p variables.
        ii. Pick the best variable/split-point among the m.
        iii. Split the node into two daughter nodes.
  2. Output the ensemble of tress $\{T_b\}_1^B$
- To make prediction at a new data point x:

$$\hat{f}_{rf}^B(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$$

3.2.3 SARIMA and SARIMAX

The SARIMA $(p, d, q)(P, D, Q)_s$ can be represented as:

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D Z_t = \theta_q(B)\Theta_Q(B^S)\epsilon_t$$

where B - lag operator
$\phi_p(B)$ - autoregressive operator of p-order

$\Phi_P(B)$ - seasonal autoregressive operator of P-order
$\theta_q(B)$ - moving average operator of q-order
$\Theta_q(B)$ - seasonal moving average operator of Q-order
$(1-B)^d$ - differencing operator of d-order
$(1-B)^D$ - seasonal differencing operator of D-order
S - seasonal length
Z_t - Sales (or margin) of a product at time t
ϵ_t - residual error in the model

The SARIMAX$(p, d, q)(P, D, Q)_s(X)$ model adds external variables to SARIMA model, where X is the vector of external variables. The external variables can be modeled by multi linear regression:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_k X_{k,t} + w_t$$

where $X_{1,t}, \cdots, X_{k,t}$ - observations of k number of external variables corresponding to the dependent variable Y_t
$\beta_1, \cdots, \beta_k$ - regression coefficients of external variables
w_t - stochastic residual and can be represented in the form of SARIMA model:

$$w_t = \frac{\theta_q(B)\Theta_Q(B^S)}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D}\epsilon_t$$

Thus, the general SARIMAX model equation can be expressed as:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_k X_{k,t} + (\frac{\theta_q(B)\Theta_Q(B^S)}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D}\epsilon_t)$$

The time series data is the daily sales / margin sum (of all stores) and the predicting time frame is the last month. The external variables - "Tuesday", "Saturday", "Promo Month", "Nonpromo Month" - are the results from Exploratory Data Analysis (explained later).

3.2.4 LSTM

To capture the time pattern with LSTM on the sales history of one product. We use a rolling window to make predictions.

- Prediction(m,$data$):
  1. initialize $window[1, ..., w] = data[data.length - w + 1, ..., data.length]$
  2. initialize $pred$ to be an empty array with size $m.length$
  3. for d = 1 to $m.length$:
     (a) $pred[d]$ = LSTM($window$)
     (b) $window[1, ..., w - d] = data[data.length - w + 1 + d, ..., data.length]$
     (c) $window[w - d + 1, ..., w] = pred[1, ..., d]$

  return $pred$

where m - month to predict
$data$ - training data
$pred$ - daily prediction for m

4 Experimental Evaluation

4.1 Data

We are given four xlsb files containing information on daily sales, products, promotions, and adherence in one state, which are shown in Figure 8, Figure 9, Figure 10, and Figure 11 in Appendix respectively. In Figure 8, we have sales, volume, and margin information for each store and product in daily bases. In Figure 9, we are given detailed data for each product with categories, supply, brand, and name. For promotion table in Figure 10, we get detailed information on each promotion's time span, name, included products, and price level. Finally in Figure 11, we have access to whether each store have adhered to different promotions at given time spans.

Since the daily sales table they gave us only contains one year of data from 2019-10-01 to 2020-09-30, we first filtered the promotions and adherence tables to make sure all information is within the same period. To start with the "Red Bull" instance, we extracted 11 distinct promotions with name containing "Red Bull" or "Redbull" (ignoring the cases) from the promotion table (in Figure 10). Note that we use promotion name instead of promotion code as primary key to represent each promotion, because we found that the promotion code has not been unified across tables. Then, we extracted all the distinct product codes that get involved in the Red Bull promotions. Some selected products are other snacks such as hot dog or chips, but we only want red bull products in out table. Thus, for the selected product code, we checked their information in product tables (Figure 9), and only remain 14 products that contain "RED BULL" in Brand. Then, we selected all the data points regarding the remaining 14 products from daily sales table (Figure 8), and generated a new column "price" by "Sales"/"Volume". Note that we didn't directly use the price from the original promotions table (Figure 10), because those are the product price after promotions, and we still have more than half of the data points that didn't adhere to any promotions. Finally, we want to add columns for the 11 promotions to indicate whether each data point adhere to them. Basically, for each data point, we checked whether the store code, product code, and date satisfied the requirements for each promotions from adherence table (Figure 11) and promotions table (Figure 10). The resulted pre-processed data is shown in Figure 12, and the last "AD" column is just a binary indicator of whether each data point adhere to any of the 11 promotions.

All the algorithms we trained are based on subsets of the columns as shown in Figure 12, but we also further processed the data so that it could best fit in different approaches. For OLS and Random Forest - Price, we extracted "weekday" and "month" as features from the column "Date", and dropped "Volume" and "AD" when fitting the model. We also encoded the "EAN Prod Code" with integer ranged from 0 to 13. For time series model SARIMA and SARIMAX, we calculated the summation of sales/margin groupby date over all stores and products and sorted the data chronologically.

For Random forest - Percentage, we use Store Code, product, weekday, month, date, cost, price percentage, regular price, promo 250ML, promo 473ML, promo 355ML, promo single, promo 2pack, promo combo as feature attributes. As we discovered in the time series analysis, the sales and margin have weekly and monthly trends, we also assume that they have annual trends. In the store dimension, the prices and costs of the same product within the same store fluctuate over time and also vary across different stores at the same time, so we need to let the model know the average of regular prices in the last non-promoted month, and the cost for this month. The promotion lasts for a month, and not all of the store owners choose to adhere to the promotion. The key definition for a promotion is the percentage of the price change. In this case of the red bull, we categorize the promotion from 2 perspectives. One is to look at the specification including 250ML, 355ML, and 473ML, the other is to look at the promotion types including applying a discount with a minimum number of one, combo pack requiring buy two together, and cross-selling such as the combo of red bull and Dorito for 9.99 reals. The current model generates insight based on the training data for each brand. If we train the model on different categories or brands together in the future, we will need to tell the model which brand and category the product belongs to.

4.2 Exploratory Data Analysis

For illustrative purposes, only visualizations for margins are shown and sales trend is very similar to margins. Figure 2 has three subplots: total margin for all stores on each day (top), average margin for all stores on each day (middle), and the number of stores on each day (bottom). The blue lines indicate stores that applied promotions where orange indicates stores that did not apply promotions.

We immediately realize weekly seasonality: Saturday (see •) is the peak of the week and Tuesday (see ▼) is the trough. Therefore, it is important to add features that indicate day of the week.

Also, we see some monthly patterns, which implies the change of price elasticity. No promotions were used in March or April. Very few stores applied promotion in October, November, January, and June and their margin is much higher than non-promoted stores - promotions seem effective. However, we only have one-year data so it is impossible to conclude anything causal. What we can do is to also add features that indicate month that has very few or many promotions.

Figure 2: Red Bull Margin

4.3 Methodology

4.3.1 Ordinary Least Squares

We randomly split the data points into training, validation, and testing sets with size ratio roughly equal to 0.7, 0.15 and 0.15 respectively. In order to let the model learn influences from each promotion, we also make sure that 70% of the data points that adhere to each promotion are in the training set.

We use daily sales and margins in test data, and summed up the daily predictions for each month over all stores and products. The average monthly percentage error in test data will be reported as the final metric that determines the models' performance across approaches.

The mean absolute percentage error (MAPE) is defined as:

$$100 \times \frac{\sum_{i=1,\cdots,N}(|\frac{true\ value_i - predicted\ value_i}{true\ value_i}|)}{N}$$

where $true\ value_i$ and $predicted\ value_i$ are aggregate values for month i.

4.3.2 Random Forest - Price

We continue use the same training, validation, and testing sets as in OLS. Here, we mainly conducted hyper-parameter tuning on trees' max depth, and received the lowest validation mean squared error at max_depth=13 for both sales' and margin's daily predictions. We applied default settings for other hyper-parameters in the random forest. MAPE was also calculated to compare with other approaches.

Model	Sales	Margin
OLS	8.27%	8.27%
Random Forest-Price	**3.32%**	3.68%
Random Forest-Percentage	3.55%	**3.23%**
SARIMA(7,1,2)(1,0,2,7)	8.57%	11.89%
SARIMAX(7,1,2)(1,0,2,7)	8.67%	9.08%

Table 1: MAPE in Test Data

To further investigate how our model would reflect the promotion strategy's influence on sales and margin, we modify the original promotion, named "RED BULL 473ML", in test data with higher and lower price. The predictions of those modified data can provide us with some insights on how to set promotion strategies in the future.

4.3.3 Random Forest - Percentage

Since the data is randomly split into 0.7 training data and 0.3 test data, we cannot guarantee that each promotion is split with the ratio of 70/30. To alleviate this ratio inconsistency, the data is reshuffled 15 times with different random seeds before splitting, and we use the average of 15 random forest predictions errors as the final result. The best max_depth is 12 and n_estimators is 100 after the hyper-parameter tuning.

4.3.4 SARIMA and SARIMAX

SARIMA/SARIMAX models have 6/7 hyper-parameters and their optimization are tuned by grid search and we use Akaike information criterion (AIC) as the criteria. After parameters are estimated, we diagnosis the fitness of model using ACF, PACF, and Diagnostics Diagram. If the residuals are normally distributed, we proceed to forecasting and validation. The train-test split is done chronologically.

4.4 Results

4.4.1 Ordinary Least Squares

For OLS, the MAPE is 8.27% for both sales margin, which is slightly higher than our expectation. The next model we try is Random Forest, which improves on OLS.

4.4.2 Random Forest - Price

In Figure 3, the two plots show the random forest's predictions of monthly sales and monthly margin over a year for the test data. Let's first focus on the ground true curve and predicted curve in the plots, which are denoted with blue and orange respectively. We can see that for both sales and margins, the two lines are highly overlapped. The MAPE for sales is 3.32% and for margin is 3.68%, which are all lower than the 5% requirement from the company.

We also made up an example to study the influence for one specific promotion. In the original promotion 2 ("RED BULL 473ML"), it sets the price of the red bull product ("ENERGY RED BULL LATA 473ML") to 10.99. We want to see what would happen if we modify this promotion with higher or lower price. From Figure 3, the green curves in the two graphs represent the predictions of sales and margins if all stores apply promotion 2 with price equal to 9, and the red curves are those predictions if all stores apply promotion 2 with price equal to 12. One example is illustrated with the orange dots in the graphs. In June, if the price is set to 9, total sales would increase by 18.07% and margins would roughly increase by 11.89%. If the price is set to 12, total sales would decrease by 31.52%, and margin would decrease by 19.82%.

4.4.3 Random Forest - Percentage

The model has around 3% monthly prediction error on both sales and margin. First, we look at how will the promotion Buy one RED BULL 473ML with discount affect the product ENERGY RED BULL LATA 473ML in Figure 4. We observe that In June, if the price $\uparrow$30%, sales would $\downarrow$18.39%;

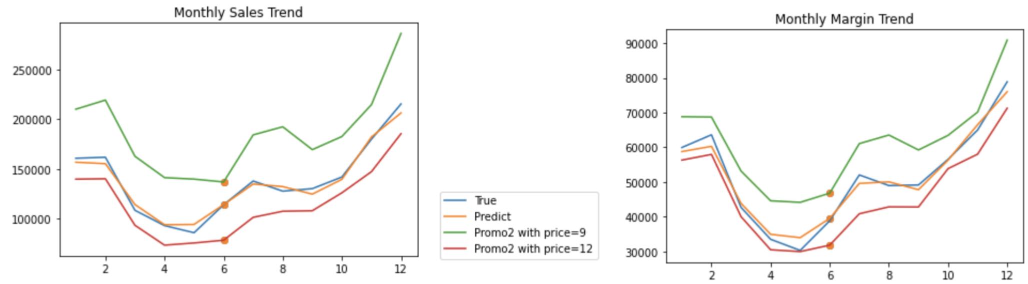

Figure 3: Random forest - price result analysis

if we have 30% off, sales $\uparrow$36.08%. In August, if the price $\uparrow$30%, the margin $\uparrow$13.41%; if we have 30% off, the margin $\uparrow$44.15%. An interesting thing is that In January, if the price $\uparrow$30%, the margin would $\uparrow$21.70%; if we have 30% off, the margin $\uparrow$15.02%. Although the sales in January for 30% off is relatively high, the margin is lower.

The influence of the promotion on all of the red bull products involved in Figure 5 also follows a similar pattern. In June, if the price $\uparrow$30%, sales $\downarrow$ 9.58%; if we have 30% off , sales $\uparrow$18.81%. In August, if the price $\uparrow$30%, margin would $\uparrow$4.95%; if we have 30% off , margin $\uparrow$16.30%. In January, if the price $\uparrow$30%, margin$\uparrow$10.56%; if we have 30% off , margin $\uparrow$7.31%.

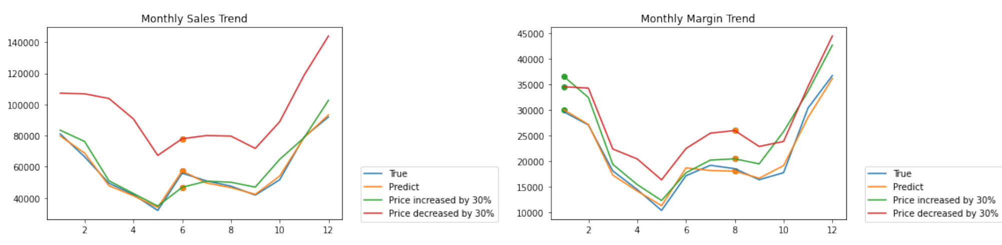

Figure 4: Random Forest - Percentage result analysis for one product

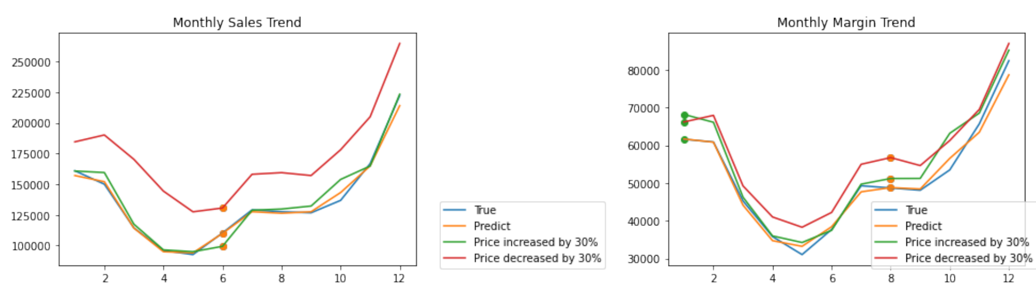

Figure 5: Random Forest - Percentage result analysis for all red bull products involved

4.4.4 SARIMA and SARIMAX

The process for forecasting margin and sales is the same. For illustrative purposes, only results from SARIMAX on margin are presented below but the MAPE for all forecasts can be found in Table 1. After applying SARIMA and SARIMAX, the predicted values align with true values closely in Figure 6. The residuals have very few autocorrelation and partial autocorrelation in Figure 7. Residuals also seem to follow the normal distribution very well in Figure 13. As result, we can conclude the models are effective but not as good as Random Forest.

4.4.5 LSTM

The LSTM performs worse than the expectation, as there are many important factors such as store code, but it only took care of the time factor.

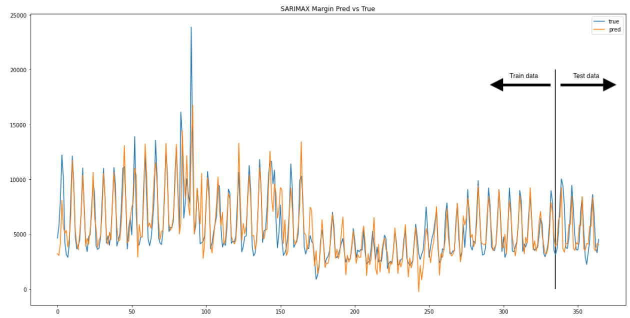

Figure 6: SARIMAX on Margin: Pred vs True

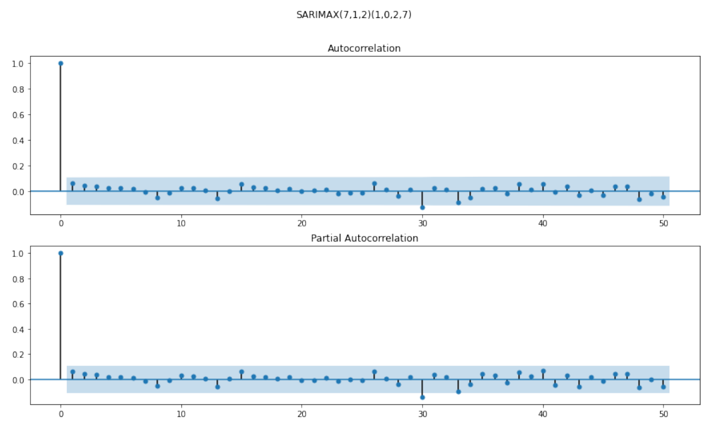

Figure 7: ACF and PACF of residuals from SARIMAX on Margin

4.5 Discussion

The MAPE for all models we built are shown in Table 1. Our goal is to achieve MAPE $\leq 5\%$ and both Random Forest models are able to keep MAPE around 3%. This metric basically means that if the true total sales/margin in next month is 100, the predicted value will most likely fall in the range of [97, 103]. We also summarize the advantages and disadvantages of all deliverable models we built in Table 2. All models are ready to be delivered but the data cleaning part is still challenging. We spent more than half of the time to get the model-ready data frame. Therefore, we also recommend to format data better and keep the vocabulary constant.

5 Conclusions

Our models reached the goal that given a product and its associated promotion, estimating the influence on the monthly total sales and margins of both this product and the brand that the product belongs to. With the sales and margin prediction, we are capable of recommending the optimal price point or price change percentage during a specific period, and thus provide assistance for the company to make data-drive business decisions.

Model	Pros	Cons
RF - Price	High accuracy, directly reflect price effect	Assume universal price for same promotion
RF - Percentage	Can estimate influence of new promos	Unstable, big memory for large-scale data
SARIMA(X)	Chronological, easy to train new data	Not best with few data

Table 2: Model Comparison

As show in Table 2, for the first approach in random forest model, where we use price as input, the major shortcoming is it assumes that all stores will follow the same price when adhered to the same promotion. But in reality, this might not be the case. RF - Percentage actually deals with this issue, and reach a similar MAPE score.

The major shortcomings for RF - Percentage include the instability and the potential issues of memory and efficiency during the deployment. The error fluctuates between 2.5% and 5%, we need more years of data to alleviate this instability. Currently, we are building one model for each brand, if we want to train all of the data on one model, it is very likely that we will not have enough memory on our computers. The random forest model has to be retrained once the data is updated, and this might takes a long time as the data size grows. Using subsampling, spark or AWS might solve these memory and efficiency issues.

The main limitation for time series model such as SARIMA(X) is that we only have one year of data, and it is difficult for the model to capture monthly trend with the limited information. For furture improvement, we plan to require roughly three years of data and thus will be able to test the model's performance on longer time spans.

Our potential next steps include evaluating the influence of the product's promotion on the whole category, cannibalization effect, i.e. sales or margin reduction of other products in the same category, and synergy effect, i.e. sales or margin increase of by-products. In the current stage, we assume that all of the stores adhere the promotion to predict the sales or margin change brought by promotion. The recommender system, which is used in shopping websites like amazon, basically recommends products based on users' purchase histories. With enough data, recommender systems might be helpful to suggest the products that a store owner tends to apply a promotion based on the promotion adherence history.

6 Lessons learned

From the project, we learned that the real world data is messy and has lots of unrelated information. We first discussed the representative and necessary attributes and asked our mentor to extract the data from the database based on our schema. After getting the data, the promotion code in two tables did not match, so we used the promotion name instead. There were cases that the names for a same promotion across different tables are slightly differ, and we solve this by matching the time and price.

We collaborated with the pricing coordinator without any technical background, so we have to figure out how to define their business goal in a technical way, and how to dig deeper to get more useful information in our conversation. Our initial presentation uses lots of texts and technical terms, which makes it hard for the mentor to fully understand. Based the mentor's feedback, in our final presentation to the company, we add visualization to each section, translate technical term into the concise language that any person without data science knowledge can follow, and give example of a specific product to explain the model performance. Thus, employees who have attended the presentation focused on our materials and expressed strong interest.

Both of the data processing and business expression experience are valuable for our future data science projects.

Contributions

Data preprocessing: Tianshu Chu, Xinmeng Li
OLS: Yichen Isabel Zhou
Random Forest - Price: Tianshu Chu
Random Forest - Percentage: Xinmeng Li
SARIMA(X): Yichen Isabel Zhou
Report write-up: All

References

[AAF16] Nari Sivanandam Arunraj, Diane Ahrens, and Michael Fernandes. Application of sarimax model to forecast daily sales in food retail industry. *International Journal of Operations Research and Information Systems (IJORIS)*, 7(2):1–21, 2016.

[DW02] Rajeev H Dehejia and Sadek Wahba. Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics*, 84(1):151–161, 2002.

[Leo01] Michael Leonard. Promotional analysis and forecasting for demand planning: a practical time series approach. *with exhibits*, 1, 2001.

[TL18] Sean J Taylor and Benjamin Letham. Forecasting at scale. *The American Statistician*, 72(1):37–45, 2018.

Appendix

2	Store Code	EAN Prod Code	Date	Sales	Volume	Margem
330	1	1220000250000	9/17/2020	15.98	2	15.98
331	1	1220000250000	9/18/2020	31.96	4	31.96
332	1	1220000250000	9/19/2020	23.97	3	23.97
333	1	1220000250000	9/20/2020	7.99	1	7.99
334	1	1220000250000	9/21/2020	7.99	1	7.99

Figure 8: Data snapshot for daily sales/margin

	EAN Cubo	Nível 1	Nível 2	Nível 3	Supply	Brand	Product
1	0	BEBIDAS ALCOOLICAS	VINHOS E ESPUMANTES	REGIONAIS 13	ARGENTO	ARGENTO	VIN ARGENTO VARIETAL MALBEC 750ML
2	0000	BEBIDAS ALCOOLICAS	DESTILADOS	REGIONAIS 12	SHELL SELECT	SHELL SELECT	FF DEST SANGRIA
3	0000078909182	BEBIDAS ALCOOLICAS	CERVEJAS	CERVEJA PILSEN	AMBEV	ANTARCTICA	SNACK BATATA ELMA CHIPS LAYS PICANHA 30G
4	0000078909212	BEBIDAS ALCOOLICAS	CERVEJAS	CERVEJA PILSEN	AMBEV	ANTARCTICA	SORV KIBON SORVETERIA TENTACAOO 1,5LT
5	0000078909229	BEBIDAS ALCOOLICAS	CERVEJAS	CERVEJA PILSEN	AMBEV	ANTARCTICA	CHOC NEUGEBAUER DELIRIO CHOCO 14 GR

Figure 9: Data snapshot for products

numero	data_inicio	data_fim	mecanica	nome	codigo	grupo_produto	produto_nome	produto_marca	produto_EAN	quantidade_minima	preco_sell_out	alcance
1	01/10/18	31/10/18	2	3 UNIDADES DE FINI 15GR 17GR	18928	1	BALA FINI TUBES CITRICO MORANGO 20GR	FINI	7898279799823	3	0.99	Nacional
1	01/10/18	31/10/18	2	3 UNIDADES DE FINI 15GR 17GR	18928	1	BALA FINI TUBES MORANGO 17GR	FINI	7898519450262	3	0.99	Nacional

Figure 10: Data snapshot for promotions

1	Store Code	Promo Code	Promo Desc	Coverage	State	Month	Start	End	Adherence\n(1 - Yes)
2	1	320	2 UNIDADES DE TRIDENT 30,6GR POR	Nacional	NaN	Jun-19	30/6/19	01/6/19	1
3	1	237	2 UNIDADES DE TRIDENT 8GR	Nacional	NaN	Mar-19	31/3/19	01/3/19	1
4	1	589	AGUA CRYSTAL 1,5L	Regional	DF,ES,GO,MG,RJ,RS,SP	May-20	05/6/20	06/5/20	0
5	1	590	AGUA CRYSTAL 500ML	Regional	AL,BA,CE,DF,ES,GO,MA,MG,PB,PE,PI,RJ,RN,RS,SE,SP	May-20	05/6/20	06/5/20	0
6	1	592	AGUA CRYSTAL 500ML	Regional	AL,BA,CE,DF,ES,GO,MA,MG,PB,PE,PI,RJ,RN,RS,SE,SP	Apr-20	05/5/20	06/4/20	0

Figure 11: Data snapshot for adherence

Store Code	EAN Prod Code	Date	Sales	Volume	Margem	price	promo0	promo1	promo2	promo3	promo4	promo5	promo6	promo7	promo8	promo9	promo10	AD
2	9002490214166	2020-01-10	62.7	3.0	29.72	20.9	0	0	0	0	0	0	0	0	0	0	0	0
2	9002490214166	2020-01-11	104.5	5.0	49.54	20.9	0	0	0	0	0	0	0	0	0	0	0	0
2	9002490214166	2020-01-12	20.9	1.0	9.75	20.9	0	0	0	0	0	0	0	0	0	0	0	0
2	9002490214166	2020-01-14	41.8	2.0	19.50	20.9	0	0	0	0	0	0	0	0	0	0	0	0
2	9002490214166	2020-01-15	20.9	1.0	9.75	20.9	0	0	0	0	0	0	0	0	0	0	0	0

Figure 12: Data snapshot after pre-processing

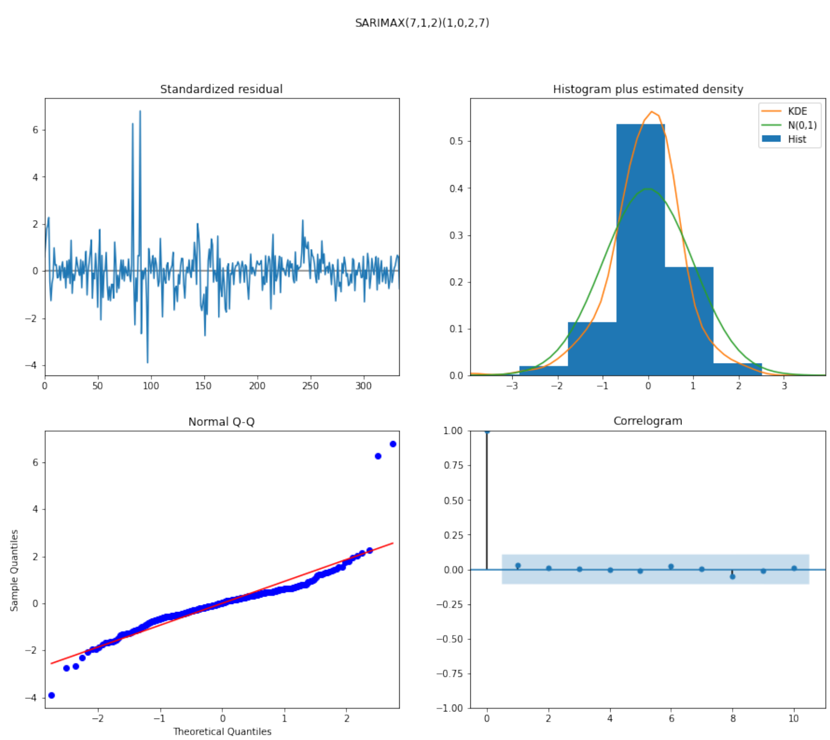

Figure 13: Diagnostics of SARIMAX on Margin