Price Optimization for Revenue Maximization at Scale

Nikhil Gupta, Massimiliano Moro, Kay Ayala, Dr. Bivin Sadler

Master of Science in Data Science, Southern Methodist University,  
Dallas, TX 75275 USA

{guptan, mmoro, kayala, bsadler}@mail.smu.edu

**Abstract.** This study presents a novel approach to price optimization in order to maximize revenue for the distribution market of “non-perishable” products. Data analysis techniques such as association mining, statistical modeling, machine learning, and an automated machine learning platform are used to forecast the demand for products considering the impact of pricing. These techniques used allow for accurate modelling of the customer’s buying patterns including cross effects such as cannibalization and the halo effect. This study uses data from 2013 to 2019 for Super Premium Whiskey and Economy Vodka from a large distributor of alcohol beverages. For each product and customer, the expected demand and the ideal pricing strategy to maximize revenue for the business are shown. This study suggests optimal revenue for the distributor is TBD. This is a N fold increase over their historical revenue. While the techniques presented in this paper have been validated for the distribution market of alcoholic beverages, they don’t rely on any domain specific knowledge from this industry, and thus can be applied to other distribution markets for “non-perishable” products.

# Introduction

In an ever-competitive world, companies need to find optimal pricing for their products in order to maximize business objectives such as revenue and profits. Revenue for a product depends on a concept called “price elasticity” [1], [2]. When price is high demand drops, but more revenue is made per sale. When price is low the opposite effect is observed. Hence, a business must find the ideal price point for their products in order to maximize revenue. This concept becomes much more complex when cross effects such as cannibalization and halo effects are taken into consideration. Cannibalization indicates that the price of a product depends not just on the price of that product but also similar products. A drop in the price of a product can cause switching behavior in customers leading to reduction in the revenue of other products. The halo effect refers to the customer preference for a brand due to a positive experience with another product from the same brand [3]. This implies that a company could charge a premium for certain products due to positive image from other products of the same brand.

There are several factors that further complicate the process of determining the optimal pricing strategy. For example, cross effects are not only non-linear but asymmetric as well. How product “A” influences product “B” is not the same as how product “B” influence product “A” [4]. Moreover, the change in the demand of a product may not only be influenced by switching behavior, but instead caused by other factors such as stockpiling (customer accelerates purchases at a discounted pricing) or a true increase or decrease in demand [3]. This can confound the results of evaluating the long-term impact of a pricing strategy. Furthermore, evaluating cross effects across hundreds of products becomes a herculean task.

Recent advances in data mining and forecasting techniques provide a glimmer of hope. Association mining techniques (such as those used by Netflix and Amazon to make recommendations) can be used to evaluate the most influential set of products that can have an impact on the revenue of other products. Multivariate time series analysis, and deep neural networks (such as Recurrent Neural Networks - RNN and Long Short-Term Memory - LSTM) can consider a complex set of interdependent and sequential variables to unveil patterns as well as provide a more accurate prediction of demand. The advent of Automated Machine Learning (AutoML) can help model the demand for products at scale. Additionally, advances in optimization methods such as Bayesian optimization plus the increase in computing power can help navigate different pricing scenarios more efficiently to find the ideal price point for various products jointly.

Using these techniques, this study provides an answer to one of the most important question for a company – what is the ideal pricing strategy for all products to drive up demand and maximize revenue? While this research focuses on the distribution market for alcoholic beverages, the techniques applied in this study are not industry specific and are generic enough to be applied to other distribution industries.

# Literature Review

## Elasticity and Revenue Management

Revenue management is defined as the science of maximizing revenue by controlling the price of the product [5]. This concept is based on the theory of price elasticity, [1] which states that as the price of a produce is reduced, its demand will increase. Initially the increase in demand will be enough to offset the decrease in price resulting in an increase in revenue. However, after a certain price point, the increase in demand will not be able to make up for the decrease in price and the revenue will eventually start to fall. This points to the existence of an optimal price point for every product to maximize revenue [2].

In terms of increasing revenue, price cuts are the most effective medium, even more so than advertisement [3]. This is an especially important topic for managers since they want to know whether a drop in price is bringing in new customers or simply providing discounts to customers who would have paid the non-discounted price anyway [3]. On the other end of the spectrum, increasing price to drive revenue can still work since core customers are likely to continue purchasing products when price is increased. This can offset the loss of revenue from fringe customers who might stop buying if price increases too much [4].

## Cannibalization and Cross-Elasticity

Any sales bump derived from price reduction can be temporary (customer stockpiling at the reduced price leading to a reduction in future sales), lasting (true increase in demand due to the reduced pricing) or the result of shifting revenue from one product to another (switching behavior, e.g. cannibalization) [3]. It is important to decipher between the various sources of increase in demand. This is because some of them are temporary and only lead to an acceleration of revenue and not a true increase [3].

Cannibalization is of key interest in this study. This has been well studied in various industries such as the airline and hotel industry [5], [3]. Cannibalization is defined as the reduction of sales in one product due to another product [6]. This can be traced to the “cross-elasticity of demand theory” which essentially states that the change in demand in one product is influenced by the change in price of another product [7]. However, the concept of asymmetric price elasticities can lead to more complexity in determining cross-elasticity [4].

The methods proposed in this research are data driven. Hence, they capture the customer’s buying pattern and the underlying reason for any potential increase in demand due to discounts. Any inherent stockpiling or purchase acceleration behavior will be captured in the model as a drop in demand following the stockpiling event. Switching behavior is studied by adding the pricing of the most influential products in the models. This study leverages association mining literature to determine the most influential products that can impact the demand for another product. True increase in demand will be visible if none of the above effects are found.

## Modeling Cannibalization

Not much literature has been devoted to modeling revenue management taking price elasticity into consideration [5], [3]. Most studies focus on assuming a linear dependence of demand on price [8]. In the best case, saturation effects at extremely low and high price points are taken into consideration using a “probit” approximation [5]. However, the nature of cross-effects may be non-linear in nature and not much research seems to have been done in this regard.

Previous literature has focused predominantly on perishable products (a product whose utility expires after a certain time) such as the airline industry and hotel industry. In these products, there is a strict deadline by which the product needs to be sold and after this deadline is crossed, the utility of the product diminishes to zero. This means that as the deadline approaches, companies may need to discount their products heavily to make sure that they are sold. This work is focused on the non-perishable product space where the market dynamics may be very different. At best, heavy discounting may not be needed in order to sell the product and at worst, the product may consume unnecessary warehouse space.

Moreover, some of the literature is based on an expert system approach wherein the opinion of experts is used to construct appropriate pricing curves for elasticity [5]. This approach lacks the scalability. Additionally, not every company has access to the experts, and even if they do, the intuition may be limited to capturing linear elasticity patterns as explained in economic theory [5]. This may in turn miss out on the subtle non-linearities involved in the pricing curves.

Additionally, most of the literature is concentrated on the direct consumer market where the solution in one industry such as the airline industry may not be applicable to the hotel industry [5], [9]. This study is focused on the distributor market. When switching from a direct consumer market to a distribution market, this effect can be exacerbated. For example, for a company selling directly to a consumer, brand switching is a problem since it is a loss of revenue. However, for a distributor who is selling products from multiple brands, this is not an issue since they still get the revenue. In addition, switching stores in the consumer market is much easier than in the distribution market due to the volume of choices available.

While private research may have been done in the area of the impact of price on demand in the distribution market, scholarly literature on this topic is scant. The goal of this paper is to shed light into the nuances of price elasticity and cross effects as it pertains to the distribution market.

## Product and Customer Segmentation

This study focusses on creating an optimal pricing strategy for the distribution market of alcoholic products. Previous studies in this area concluded that demand needs to be modeled on a per product basis due to different buying patterns, product preferences and tolerance to price changes [10], [11]. This has been backed up in previous studies that noted the change in sales due to change in pricing can vary across brands and products [12], [13], [14]. In a consumer market that is focused on thousands of customers with sporadic purchasing patterns, modeling on a per customer basis is generally not possible. Instead aggregate effects are considered by combining the revenue of products across all customers. This can lead to a loss of information such as the buying patterns and price sensitivities of individual customers. However, a distribution market is inherently different. The number of customers is limited, and the sales are more frequent and in larger volumes. Hence modeling on a per customer basis is feasible and gives a nuanced model which captures the buying patterns of individual customers. This is important to consider since some customers may be more price sensitive than others.

In this paper, we propose a method to compute the price point at which a product should be offered to individual customer of the distribution company. By offering different price points for the same products to different customers based on their purchase propensity, a distribution company will be able to maximize their revenue.

## Demand Models

On the one hand, it has been demonstrated that a Vector Autoregressive (VAR) based demand forecasting framework produced good results [3]. On the other hand, “no one sized fit all” model exists for the distribution of alcoholic beverage market [10], [11]. The type of model that worked for one customer-product combination did not work well for another. Hence, a variety of modeling techniques were tried such as Naïve models, statistical models (e.g. ARMA, ARIMA, Seasonal ARIMA, VAR, Signal Plus Noise, Multiple Linear Regression with Correlated Errors), Deep Learning models (based on variants of Long Short Term Memory) and “Ensemble” techniques [10], [11]. These authors also noted that due to limited amount of available demand data (84 observations in their cases – 7 years sampled on a monthly basis) the deep learning approaches tend to overfit. In addition, while both these studies focused less than five customers and products, they indicated that the ability to scale would need an automated machine learning system (AutoML) [10], [11].

## Automated Machine Learning (AutoML) for Time Series

Due to recent advances in computing, and data science, as well as the democratization of machine learning, several AutoML systems have been introduced in the marketplace. One such framework is Auto-sklearn [15] which is built on the popular scikit-learn framework in python. Demand models are inherently time series based as the demand is a function of time, in addition to other exogenous variables. Although the scikit-learn framework has support for time series based sliding window resampling methods as well as imputing and feature engineering methods, they do not support the traditional autocorrelation-based model (such as ARIMA and Vector Auto Regressive models) [16].

Similarly, a recent entry into the world of AutoML is PyCaret, but it also does not support time series models [17]. Other systems such as the H2o.ai Driverless AI support time series forecasting, but these are paid products costing thousands of dollars [18].

Hence, in order to truly scale this across several customers and products, this study develops an AutoML framework that can support a variety of machine learning models and methods such as Seasonal ARIMA, Vector Auto Regressive models, XGBoost, and Stacking. Fortunately, libraries such as Statsmodels have extensive support for time series analysis [19]. In addition, Scikit-learn provides the option to add custom models in the framework [20]. Based on research, using available libraries and frameworks, a free and scalable time series AutoML framework can be developed to perform demand forecasting and revenue optimization at scale.

## Optimization

Optimization techniques have been used in the past for finding optimal pricing strategy. Much of the literature though has focused on discrete optimization-based approaches where the price point could only be set to discrete levels [5]. This may be limiting, especially for larger volume companies such as distributors, since a change of even a few cents may have a substantial impact on the revenue.

The other challenge with optimization techniques is that as the design space becomes non-linear and higher dimensional (i.e. as more products are considered that can cause cross elasticity), optimizers struggle to reach the global minimum [21]. Techniques to aid optimization, specifically for pricing, have been discussed where heuristics are developed in order to simplify the optimization problem [22], [23]. Alternately, this study proposes the use of association mining rule to pick the cross products that are “most influential” in predicting the revenue of another product. Since the number of cross products that are considered can be controlled using this technique, we do not need to take all products into consideration simultaneously when building the demand models for a single product. This helps reduce the dimensionality of the problem space which assists the optimizer in finding the global minima more easily.

More recently, Bayesian optimization has emerged as a promising technique that allows for a more efficient search of the sample space leading to optimal results [24]. By providing a practical approach to exploitation of the best-known search space and the exploration of unknown search spaces, this approach has been shown to yield faster and better optimization compared to other techniques.

# Methods

In order to address the research objectives, this study focuses on data from a large distributor of alcoholic beverages. It particularly focuses on two distinct markets – “Super Premium Whiskey” and “Economy Vodka”. These markets represent opposite ends of the spectrum. Super Premium Whiskey are a relatively lower volume higher margin business, whereas Economy Vodka are a higher volume, lower margin business. Decreasing the price of Super Premium Whiskey may spur demand and increase sales and revenue, whereas the ability to charge even a few dollars more for Economy Vodka (without impacting the demand negatively) could add substantially to the bottom-line.

## Exploratory Data Analysis (EDA)

The data consisted of 201,730 observations of weekly transactions from January 2013 to December 2019. It was broken down by individual customers and products. A total of 278 distinct products from 50 different brands were identified in the data. Majority (215) of these products belonged to the Super Premium Whiskey category, while 64 were in the Economy Vodka category.

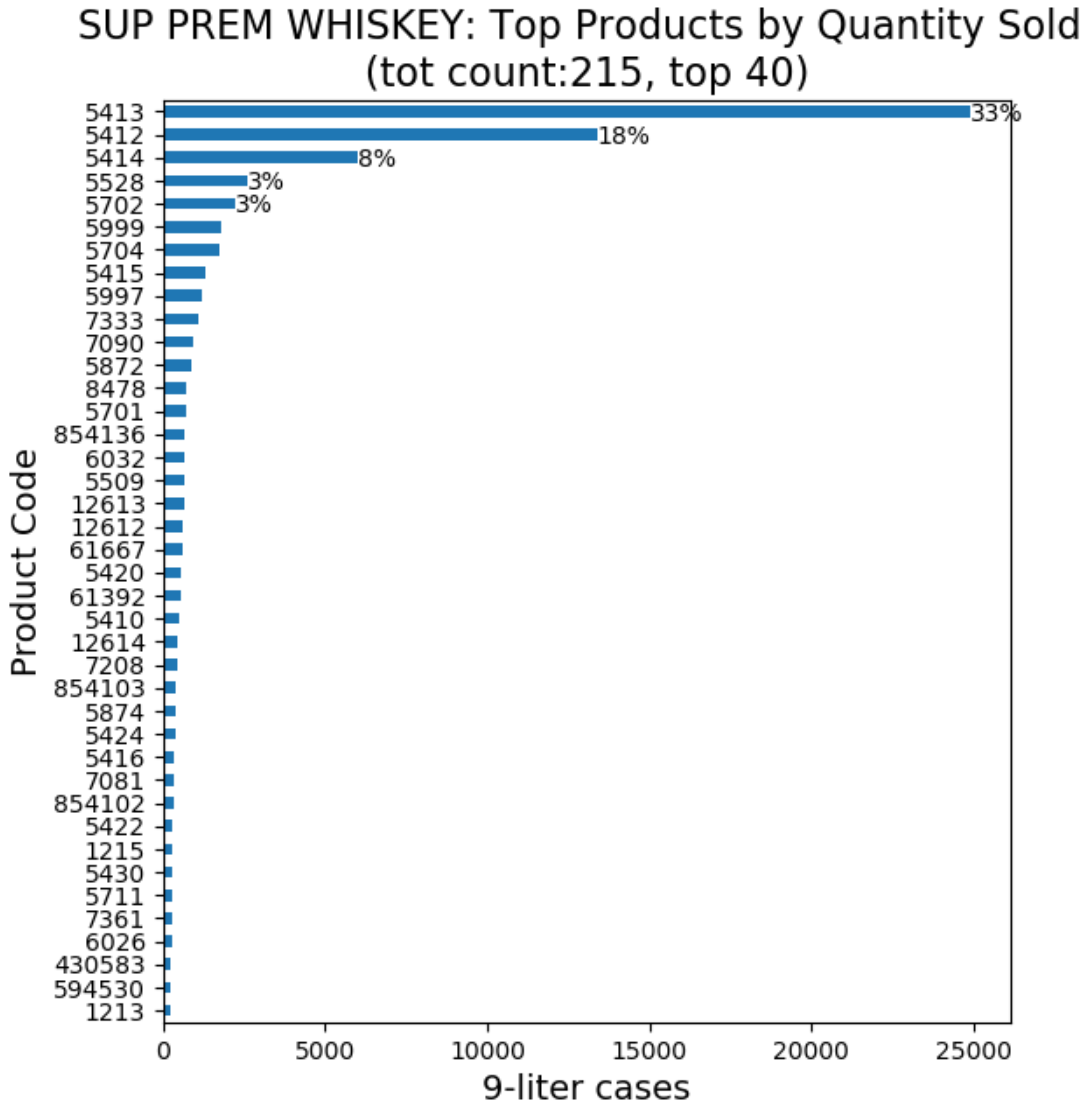
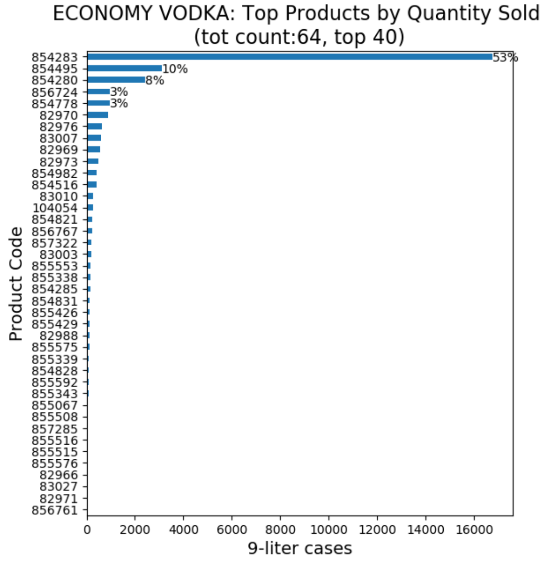


Fig. 1. Distribution of number of transactions vs. 9-liter equivalent cases sold by product in the Super Premium Whiskey. The percentage near the horizontal bar is the proportion vs. the total quantity sold.

An initial analysis of the transactions by product shows that the top three products represent a large proportion of the total quantity sold in each category. This is visible in Fig. 1 which shows the quantity sold for each product. This quantity is represented in terms of “9-liter Cases” which is an industry standard measurement term that standardizes the quantity regardless of the size of the bottles in the case (375ml, 750ml, 1L, 1.75L, etc.). This helps in aggregating the quantity across the various transactions. It can be observed that the top three products for Economy Vodka represent 71% of total quantity sold and the top three products in the Super Premium Whiskey category represent 59% of the total quantity sold. In order to build a more effective revenue optimization model, this work focuses on the top three products in each category and the correlation of their price with the demand for the other top products.

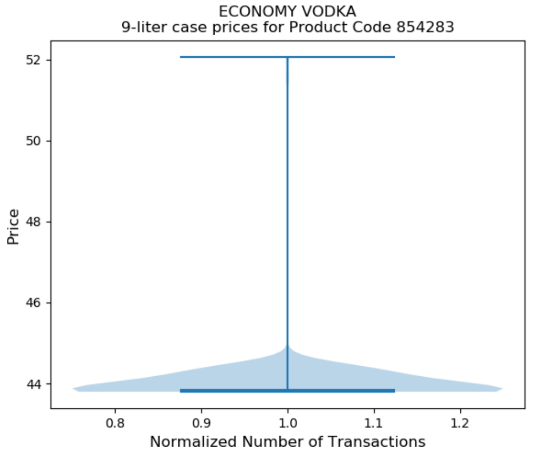
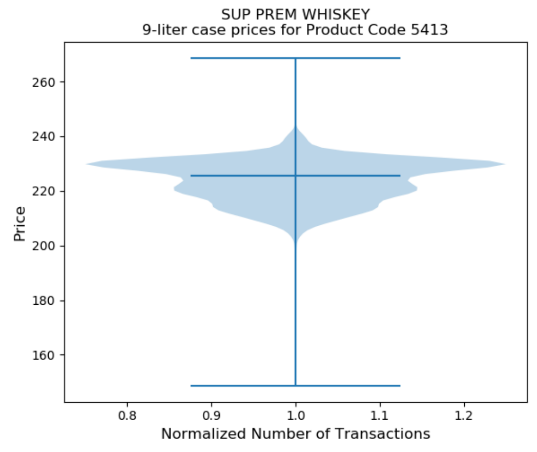


Fig. 2 Distribution of price per 9-liter case for the most sold product in Super Premium Whiskey and Economy Vodka category

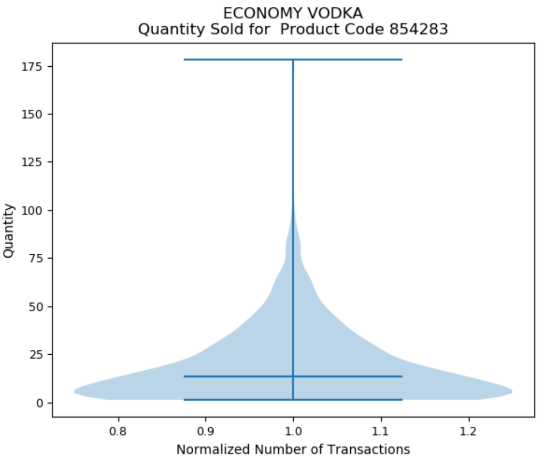
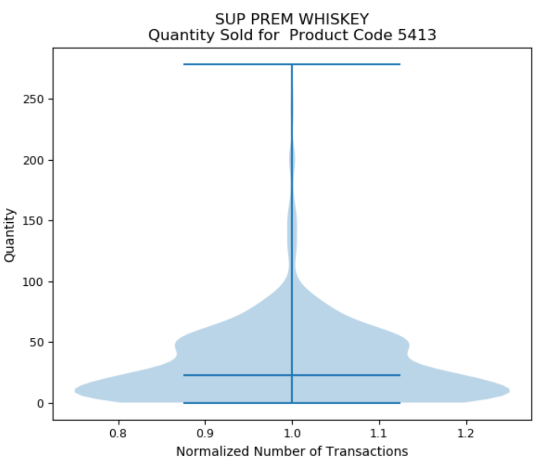


Fig. 3. Distribution of quantity of products sold for the most sold product in the Super Premium Whiskey and Economy Vodka category.

A model that takes price elasticity into account needs to leverage the variance in the price at which the product is sold over time and how the sales vary as the price changes. Fig. 2 shows that the variance in the price of a product is remarkably different for the top product in the Super Premium Whiskey category vs. the top product in the Economy Vodka category. The Super Premium Whiskey product has a very high variance (pricing varying from $200-$240 in most cases) compared to the Economy Vodka product (price varying from only $44-$45). In addition, although the quantity sold is right skewed in both cases, the top product in the Super Premium Whiskey category shows a bimodal behavior, whereas the top product in the Economy Vodka category shows a unimodal behavior (Fig. 3). Similar disparate behavior was seen for the other products in these categories (see Appendix 8.1and 8.2).

The difference in the price distribution is an indication that the two product categories have a very different price and demand behavior and that this needs to be analyzed with separate models. This observation is consistent with the literature review as well.

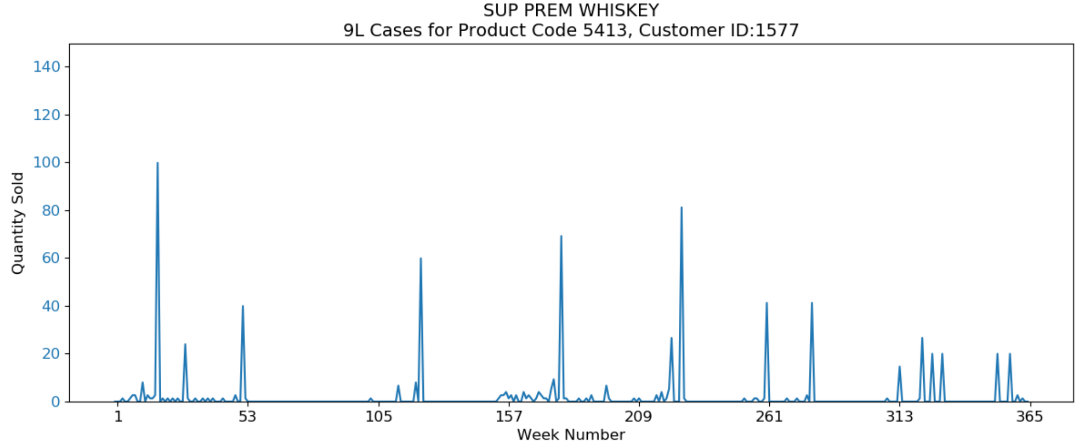


Fig. 4 Quantity sold per week, for the top product in the Super Premium Whiskey category for a Customer.

In the beverage market there is an expectation of seasonal behavior, where some periods have higher sales due to holidays and weather. Fig. 4 shows the quantity sold for the top product in the Super Premium Whiskey category for a customer. It is visible that there are periodic peaks in the quantity sold, and some hints of seasonality. Hence, seasonality is considered a key factor and needs to be incorporated in the models used by this paper.

There is also evidence of outliers in the data that can be caused by human error or other external factors like special sales events. Some of the weeks also show zero quantity sold, which can be caused by external factors like data entry errors, closure, or slow-down in the business. This work considers that these data are based on real-word situations where human error, systems update, and record adjustments may add large noise in the data. In order to reduce the noise, we resample at the monthly level which reduces the outliers and periods of low or no purchase. This can be seen clearly from Fig. 5 and Table 1.

Lastly, the exploratory analysis looks at the variation of price over time and how it impacts demand. Fig. 6 shows this behavior for the top product in the Super Premium Whiskey category. There is evidence of fluctuation in the price over time, with some impact on the quantity sold. When the price increases the total quantity sold tends to reduce. Hence, the models in this research uses price as exogenous variable when predicting demand.

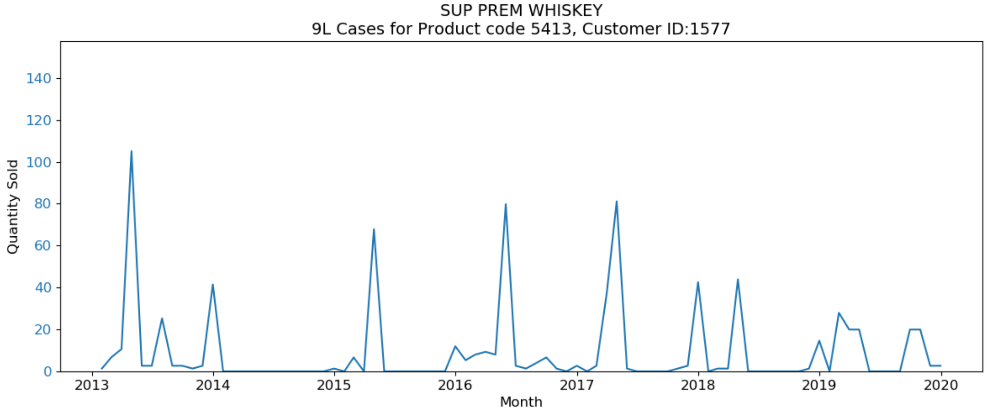


Fig. 5 Quantity sold per month, for the top product in the Super Premium Whiskey category for a Customer

Table 1. Percentage of observations with zero sales (weekly vs. monthly) for the top three Super Premium Whiskey products for a Customer

|  |  |  |
| --- | --- | --- |
| Top Product ID | % of periods with 0 sales (Weekly) | % of periods with 0 sales (Monthly) |
| 0 | 6.3 | 0 |
| 1 | 23.42 | 8.33 |
| 2 | 25.75 | 2.38 |

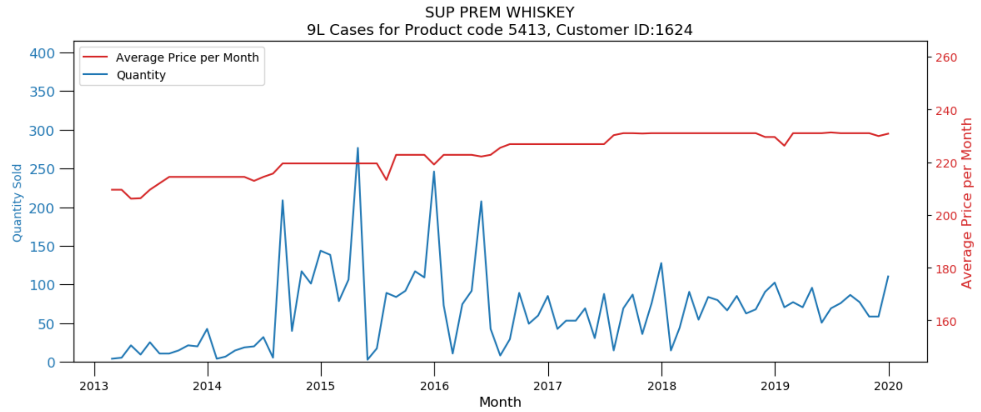


Fig. 6 Quantity sold per month and Price per 9-liter case for the top product in the Super Premium Whiskey category for a Customer.

## Overall Algorithm

The study builds revenue optimization models on a per customer and per product category basis. Hence, the overall algorithm flows as shown in Fig. 7. For each customer and (line 1) product category (line 2 – Super Premium Whiskey and Economy Vodka in this case), we select the top ‘N’ products to analyze based on revenue (line 3). These ‘N’ are the products that will have the most likely impact on the overall revenue moving forward. Then we build ‘K’ demand models (using various modeling techniques) for each of these products (lines 4-6). For building the demand model and in order to consider the price elasticity, we first select the ‘M’ most similar products (line 5) and include their price in the model as exogenous variables along with the price of the product for which the revenue is being modeled (line 6). Finally, we pick the best models from these ‘K’ models using a suitable metric (line 7). In addition, a rolling window is used for model comparison since it offers a more robust estimate of the models forecasting capability by computing “validation scores” across multiple time windows. This ensures that a single “bad” forecast does not eliminate an otherwise good model.

Once a “best” model is chosen for each product, we compute a “revenue” cost function for this customer and product category considering all ‘N’ products at the same time (line 8). Details on how this cost function is computed is discussed in the subsequent sections. Next, we use this cost function in an optimization algorithm and maximize the overall revenue of each customer and product category. The variables used to optimize the revenue are the price for the ‘N’ products. In other words, optimization computes the maximum revenue achievable by considering the joint price of the top ‘N’ products for each customer and product category combination.

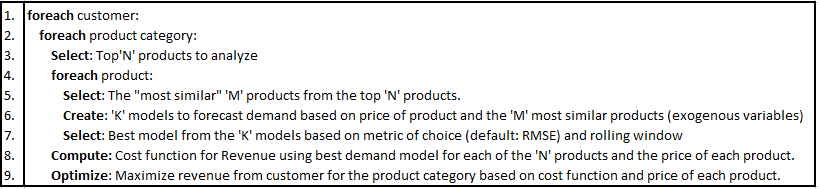


Fig. 7. Overall Algorithm

For this research, ‘N’ was chosen to be 3 since it accounted for majority of the overall revenue for each category. ‘M’ was chosen to be 1 (i.e. consider 1 similar product) for the baseline revenue optimization model. ‘K’ was chosen to be 2 for reasons that will be described in the subsequent sections.

## Identifying Products Similarity through Association Analysis to address Cross Effects

To study the impact of cross-effects on price elasticity, the similarity between products contained in the data is investigated. Ideally, all products in a category would be included when modeling the demand (with cross-effects) for any product. This will give the most accurate measure of cross effects. However, with limited amount of data, this approach does not work since lagged variables must also be considered in the demand analysis. Hence, this study narrows the number of influential products used in modeling. This is accomplished with the help of association analysis (co-purchase patterns) and product characteristic similarity metrics considering brand, product name, flavor, and bottle size.

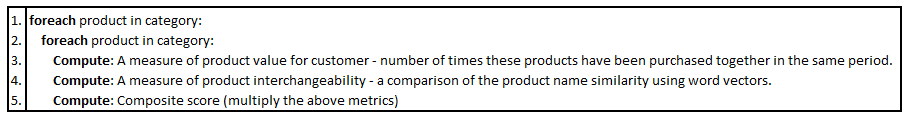


Fig. 8. Algorithm to compute the product similarity

The algorithm to compute the product similarity is shown in Fig. 8. The ideal measure of similarity between products should consider the degree to which two products have similar value for the customers, as they will be more likely to switch between these products. The value of a product for a customer can be defined in terms of the purchase frequency of the product within a defined timeframe [25]. A product with higher purchase frequency shows a higher value for the customer. The measure of similarity in value of two products for a customer can be calculated by counting how many times two products were purchased together in the same time period relative to the total number of purchases during that period, i.e. normalized score between zero and one (line 3). A value of zero represents an absence of similarity in the value of two products for the customer, while a value of one represents maximum similarity.

While this co-purchase measure is an effective way to assess the value of the products for the customers, it does not measure product interchangeability. Instead, similar brand and product name increase the customer’s perceived similarity between product variants in the same family (the halo effect) [26]. Hence, they are inclined to evaluate these products jointly when making their purchasing decisions. Due to this a second measure of product similarity is included by comparing word-vectors of the concatenation of products name, brand, flavor, and bottle size (line 4). Word-vectors are a multi-dimensional representation of word meaning wherein two similar words tends to be close to each other in their vector space. This methodology goes beyond the simple syntactic regularities by also considering the semantic representation of the word [27]. With this technique two products with different names but similar characteristics would have a high similarity score. The calculation of this measure has been performed by using the word2vec algorithm, provided by the Python library spaCy [28]. This algorithm returns a similarity score for two words which is between zero and one. Values close to one imply the products are semantically similar and vice versa for values close to zero.

The two measures of similarity, i.e. “product value” and “characteristic similarity”, are then multiplied together resulting in a single score ranging for zero to one (line 5). This final score represents how much two products are similar and interchangeable.

## Demand Modeling

The models that can be used in this study need to have the ability to pass exogenous variables as inputs to the model. This was needed since in the optimization step, the price was used as the variable for optimization of the revenue.

For the demand modeling, 2 models were considered for the initial baseline flow. Because our data showed seasonal behavior, Seasonal ARIMA models with Exogenous variables (SARIMAX) were included in this analysis. A Machine Learning (ML) based approach was also included where the time series problem was converted to a supervised learning problem by considering the previous observations as inputs [29]. This transformed data could then be fed into a plethora of supervised learning ML models and a best model chosen from amongst these ML models.

From literature it was found that VAR model works well in many cases [3]. Although the VAR models provide an option to include endogenous variables in the data, we don’t have any control over these variables in the model since this type of model will automatically and jointly forecast these variables along with the variable of interest. Since we cannot control these variables directly in the model, we could not include this model in our evaluation.

It was also found [10], [11] that more complex deep learning models were not suitable since they tended to overfit the data. Considering that we only had a maximum of 84 observations per product (when resampled monthly), this conclusion would apply in our case as well. Hence, we decided to exclude these models from our analysis.

Since the study needed to build over 35 models (three customers, two product categories, three products in each category, two models per product), an automated flow was needed. Hence, this research also developed such a system that allows for the user to create all models for a single product using just two lines of code [30]. The complexity of the model building process was encapsulated within the two lines and made transparent from an end user’s perspective.

## Model Evaluation

The metric used to evaluate the models was chosen to be Root Mean Squared Error (RMSE). Since the output variable (demand) was numeric, this was an appropriate metric. This metric essentially computes the error of each observation after disregarding the sign of the error – i.e. underprediction and overprediction by the same amount are treated as the same amount of error. Also, since the resulting error is on the same scale as the output variable, it has a human readable meaning to it. Particularly, in this case, an error of ‘X’ means that the demand predicted was either ‘X’ above or below the actual demand for that period. This is a useful interpretation compared to metrics such as Mean Squared Error (MSE) where the metric is not on the same scale as the output variable. For example, if demand is in the 100’s and RMSE is in the single digits, it could imply that the model is good. It is possible to make these conclusions since the error is on the same scale as the demand. However, with other metrics which are not on the same scale, making this direct comparison becomes difficult.

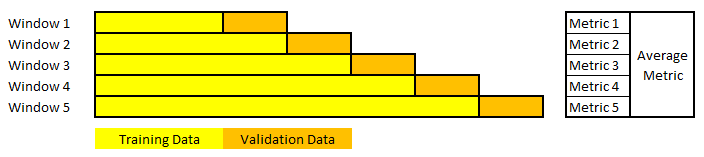


Fig. . Rolling Window Forecasts

Another important consideration is the window to use to “validate” the models. Traditional approaches use a single “window” of comparison. However, this may not lead to choosing the best model for the data. For example, if we choose the window to be the period where an otherwise good model predicts poorly, we can possibly eliminate that model. This will eliminate a model that may otherwise have generalized well to the prediction of unseen future data. Conversely, if we choose a window where an otherwise bad model predicts well, we may end up choosing a model that will not generalize well to future predictions. Hence a better approach is to use a “rolling window” approach as described in Fig. 9. In this approach, the model is evaluated multiple times using different time periods. The evaluation metric is computed for each of these windows and is averaged out to obtain a more robust estimate of the models forecasting capability. It is important to maintain the sequence of the data when creating these windows since the models will still be “autoregressive” in nature and use previous time points to make future forecasts. Mixing future data observations in the training data would lead to data leakage and the model’s performance in production will not be as good as what we would get during the training.

## Optimization

Once the models have been developed for a customer and all the products in a certain product category, a maximum revenue can be computed for this customer and product category combination using the joint pricing information for these products. In order to do so, the pricing can be fed into an optimization algorithm along with a cost function that computes the revenue for the customer and product category combination. This study uses Bayesian Optimization since it offers an efficient search mechanism that minimizes the optimization time.

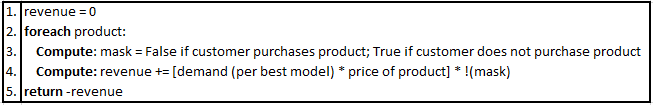


Fig. 10. Algorithm to compute the Cost Function for revenue maximization

The algorithm to compute this cost function is shown in Fig. 10. The algorithm begins by initializing the revenue to 0 (line 1). For each product in the category (line 2), the algorithm decides whether to add the revenue for that product into the cost function or whether to mask it (line 3). This decision is based on whether the customer is going to purchase a product in that time period. This is an important step since this study is not about influencing a customer to purchase a product when they don’t have a need for it. Rather it only looks at the optimization of pricing to maximize the revenue after the customer has indicated the willingness to buy the product (i.e. the distributor has received a request for quote or RFQ). Hence, this revenue maximization flow is applicable only when a customer is already purchasing a product.

Once the mask is computed, the algorithm computes the revenue for the product by multiplying the output of the demand model for that product by the price of the product. The resulting revenue is then added to the overall revenue taking mask into consideration (line 4). Once all the individual revenue calculations are added to the overall revenue, this overall revenue is returned as an output of the cost function which can then be used by the optimizer. It is important to note a small subtlety here that the optimizer only minimizes the value returned by the cost function. Since the goal is to maximize the revenue, the cost function returns the negative revenue value and the optimizer minimizes this which is equivalent to maximizing the revenue.

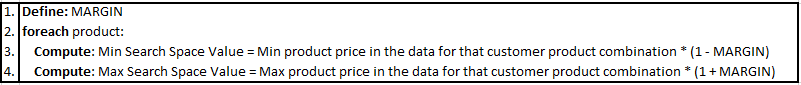


Fig. 11. Algorithm to compute the Search Space for the Optimizer

The other important consideration for the optimizer is the definition of the search space for the pricing of each product. This confines the minimum and maximum value of the pricing for each product that the optimizer will consider. It is important to note that since the price range of each product will be very different, the search space must be computed on a per product basis. The algorithm to compute this search space is shown in Fig. 11. The algorithm starts with the pricing analyst (or equivalent) defining a “MARGIN” (line 1). This is a parameter that is set based on how high or low we want the price to go with respect to historical pricing. Next, the algorithm computes the minimum value of the search space for each product by looking at the minimum value of the price that this product was offered to the customer in the past. This is then multiplied by (1-MARGIN) to get the minimum value of the search space for this customer-product combination. The maximum value of the search space is computed in a similar manner, the exception being that the algorithm looks for the maximum historical pricing and instead of multiplying by (1-MARGIN), it multiplies by (1+MARGIN).

# Results

The first step in the algorithm is the computation of product similarity for the first category (Super Premium Whiskey). The similarity score is calculated for pairs of products purchased together in the same period. This can be visualized using a heatmap (in Fig. 12). This figure shows the similarity matrix for the three most popular products in the Economy Vodka and Super Premium Whiskey product categories for a customer chain and clearly depicts how some products with similar name and purchase frequency have a higher score.

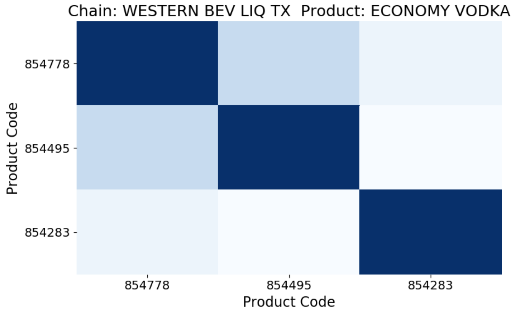
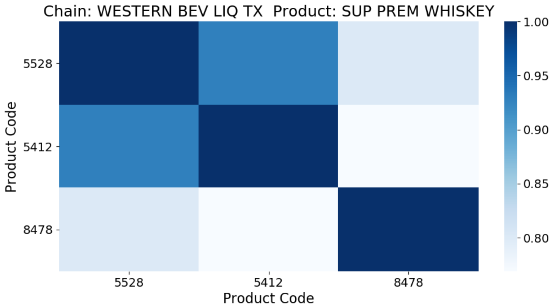
 

Fig. 12. Similarity matrix for Economy Vodka and Super Premium Whiskey for a customer chain.

Table 2. Most Similar Products for the Top three products in the Super Premium Whiskey Category for a customer chain

|  |  |  |
| --- | --- | --- |
| Product Code | Top Product ID | Most Similar Product ID |
| 5528 | 0 | 1 |
| 5412 | 1 | 0 |
| 8478 | 2 | 1 |

From Fig. 12 and Table 2, we can see that the most similar product to Top Product ID 0 (corresponding to the top Product Code 5528) in the Super Premium Whiskey category is Top Product ID 1 and vice versa. For Top Product ID 2, the most similar product is Top Product ID 1. We use this information to build the models for each product (i.e. for Top Product ID 0, we take the price of Top Product ID 0 and the price of Top Product ID 1 into consideration, etc.).

Next, we constructed naïve baseline models for each of these products in this category. This will serve as a reference for comparing the statistical and machine learning models. The naïve model is computed by taking the average of the last two seasonal periods. For example, if seasonality is 12 (for monthly data), then the naïve model will pick the average of the actual value from 12 months and 24 months back as the forecast for the next period (month).

Further, we constructed the ‘K’ models as described in the algorithm. Fig. 13 shows the impact of cross effects as captured by one of the models. It shows how the revenue for Top Product ID 1 is impacted by the pricing of Top Product ID 2. If there was no impact, we would expect the revenue to rise linearly as the price of product 1 changes. However, we see that product 2 has a minor impact on the revenue of product 1, which introduces a slight non-linearity.

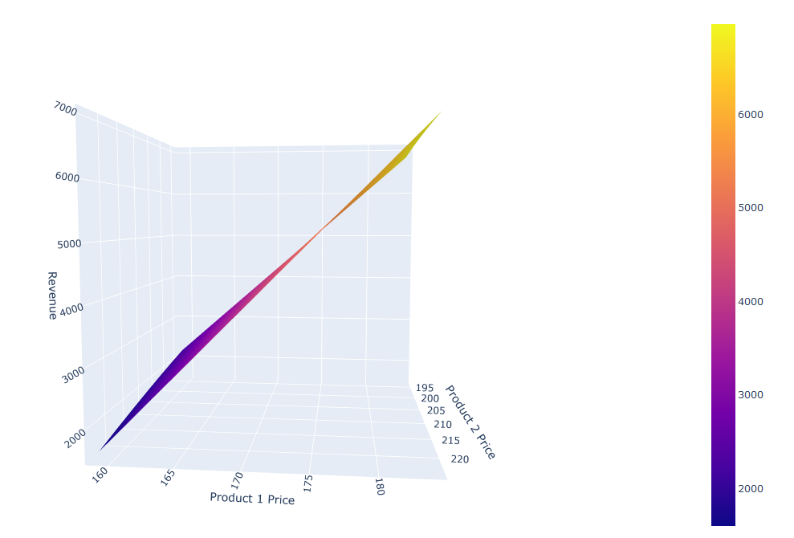


Fig. 13. Impact of Product 1 Pricing on the Revenue for Product 2.

The performance comparison of the various models documented in Table 3. For Product ID 0 and 1, the ML models perform considerably better than the SARIMAX models, whereas for Product ID 2, the SARIMAX model performs marginally better than the ML model. Our AutoML framework automatically picks the best model from the list for each product.

Table 3 Model RMSE for Top three products in the Super Premium Whiskey Category for a Customer

|  |  |  |  |
| --- | --- | --- | --- |
| Top Product ID | Naïve Baseline Model | SARIMAX | Machine Learning (ML) |
| 0 | TBD | 90.84 | 53.36 |
| 1 | TBD | 64.38 | 15.54 |
| 2 | TBD | 18.19 | 19.9 |

Once the best models have been selected for each product, the complex cost function and the search space are computed as described earlier. A MARGIN of 0 was used in the baseline flow (but another value can be used as well). These are then used to find the optimum pricing for maximizing the revenue. An example of the optimizer is shown in Fig. 14. It takes roughly 80 iterations to reach the global maxima of the revenue (note from above section that the optimizer minimizes the negative revenue value, hence the negative values on the y-axis).

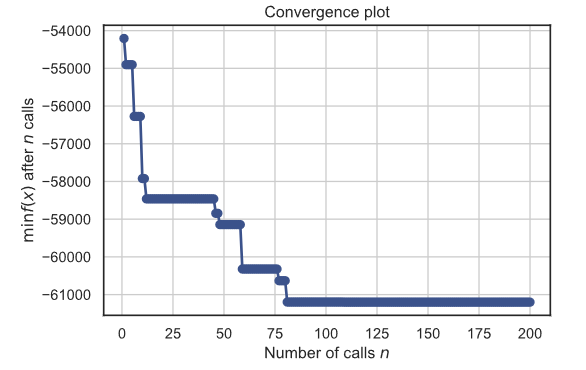


Fig. 14. Optimizer Convergence Progress for the Top three products in the Super Premium Whiskey Category for a Customer

Table 4 shows the output from the optimizer. The price for which products 0, 1, and 2 were sold during that period were $TBD, $TBD and $TBD. The recommendation from the optimizer is to sell them at $TBD, $TBD and $TBD instead. The optimizer predicts that this will likely increase the revenue by (TBD)% from $157421 to $TBD.

Table 4 Optimized Revenue Calculation vs. Actual for Top three products in the Super Premium Whiskey Category for a Customer

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Primary Product ID | Price | Revenue | Optimized Price | Optimized Revenue |
| 0 | $229.53 | $109,290 | $236.35 | $44,286 |
| 1 | $185.59 | $32,788 | $186.4 | $9,439 |
| 2 | $222.36 | $15,343 | $222.6 | $7,476 |
| Total |  | **$157,421** |  | **$61,201** |

<placeholder>

The baseline revenue optimization flow has been developed. However, the model performance does not seem to be very good. In the above case, they are underestimating the demand (so even though the optimized price is close to what was actually offered, they are estimating the optimized revenue to be much lower).

Next steps include.

1. Current AutoML flow only incorporates 1 evaluation window. Incorporate rolling window to get a better estimate of model performance.
2. Incorporating other models such as Prophet library from Facebook
3. More ML techniques
4. Include more “similar” products to see the impact

</placeholder>

# Discussion

## Impact on Pricing Managers

<placeholder>

* + Now that this study has projected demand, estimated optimal price, and calculated projected revenue.
  + This projection is best during [\_\_\_\_] time period due to this study’s forecast horizon
  + This methodology can be used to calculate price points [\_\_\_] time in advance (days, weeks, or months). Based on the forecast horizon and the variability of global politics and other phenomena
  + While [\_\_\_\_] interaction patterns in this study’s cross-cannibalization and halo effects analysis can be seen, ultimately the true interaction pattern is best captured with experimentation.
  + Since this analysis is not an experiment but rather an exploration of data this study cannot make any population or causal claims. However, this study was able to elucidate patterns in the data.
  + A pilot study would be recommended to see how this strategy performs in practice.

</placeholder>

## Ethical Considerations

<placeholder>

* + This paper will add the ethical implication of providing a different price point for the same product to different customers. “A University of Pennsylvania [study](http://edition.cnn.com/2005/LAW/06/24/ramasastry.website.prices/) (Open to Exploitation: American Shoppers Online and Offline, Joseph Turow, Lauren Feldman, and Kimberly Meltzer) that addressed online price discrimination revealed that the majority of those surveyed believed that price customization was illegal, or strongly believed it ought to be. The truth is, it’s usually legal.”.
  + “Price discrimination is illegal if it’s done on the basis of race, religion, nationality, or gender, or if it is in violation of antitrust or price-fixing laws.” This research should not be implemented based on discriminatory practices and based purely on buying propensity of the customers (data driven based on historical behavior). “Retailers can distinguish between customers and price differences (as long as they do not discriminate based on impermissible attributes) and dynamic pricing is perfectly legal.” <https://econsultancy.com/what-is-price-discrimination-and-is-it-ethical/>.

</placeholder>

# Conclusion

<placeholder>

* This methodology can increase profit for products other than just alcohol
  + The economic size of the retail sector is large making this kind of analysis have broad implications
* Future work
  + New products
  + Navigating the pandemic ‘cold start’ problem
  + Including spatial analysis
  + Including a broader feature set outside of just the company’s data to include
    - Economic indicators
    - Social network data
    - Marketing effects
    - Weather and other global effects
  + Building in a control system style experiment for capturing cannabalization and halo effects.
  + Consulting literature on price point psychology

</placeholder>

# References

1. OpenStax. (2016, May 18). *Principles of Economics.* Retrieved from BC Open Textbooks: https://opentextbc.ca/principlesofeconomics

2. Bromley, R. (2006). *Elasticity*. Retrieved from pcecon.com class notes: https://www.pcecon.com/notes/elasticity.html

3. McColl, R., Macgilchrist, R., & Rafiq, S. (2020). Estimating cannibalizing effects of sales promotions: The impact of price cuts and store type. *Journal of Retailing and Consumer Services, 53.*

4. Meredith, L., & Maki, D. (2001). Product cannibalization and the role of prices. *Applied Economics*, 1785-1793.

5. Bayoumi, A.-M., Saleh, M., Atiya, A., Aziz, H., & Atiya, A. (2013). Dynamic Pricing for Hotel Revenue Management Using Price Multipliers

6. Kotler, P., & Keller, K. (2012). *Marketing Management.* Boston: Pearson.

7. Kerin, R. A., Harvey, M. G., & Rothe, J. T. (1978). Cannibalism and new product development. *Business Horizons, 21.5*, 25-31.

8. Pickett, J. C., & Reilly, D. P. (2020, May). *Testing Marketing Hypothesis.* Retrieved from Automatic Forecasting Systems: https://www.autobox.com/pdfs/cannibalize.pdf

9. Akçay, Y., Natarajan, H., & Xu, S. H. (2010). Joint Dynamic Pricing of Multiple Perishable Products Under Consumer Choice. *Management Science, 56*(8), 1345-1361.

10. Jiang, L., Rollins, K. M., Ludlow, M., & Sadler, B. (2020). Demand Forecasting for Alcoholic Beverage Distribution. *SMJ Data Science Review*.

11. Arora, T., Chandna, R., Conant, S., Sadler, B., & Slater, R. (2020). Demand Forecasting In Wholesale Alcohol Distribution: An Ensemble Approach. *SMU Data Science Review*.

12. Chevalier, M. (1975). Increase in Sales Due to In-Store Display. *Journal of Marketing research, 12*(4), 426-431

13. Blattberg , R. C., & Wisniewski, K. J. (1989). Price-Induced Patterns of Competition. *Marketing Science, 8*(4), 291-309.

14. Neslin, S. A. (2002 ). *Sales Promotion, Relevant Knowledge Series.* Marketing Science Institute .

15. Machine Learning Professorship Freiburg. (2019). auto-sklearn. Retrieved from auto-sklearn: https://automl.github.io/auto-sklearn/master/index.html#

16. Rosenthal, E. (2018, MArch 22). *Time Series for scikit-learn People (Part II): Autoregressive Forecasting Pipelines.* Retrieved from Ethan Rosenthal : https://www.ethanrosenthal.com/2018/03/22/time-series-for-scikit-learn-people-part2/

17. Moez, A. (2020, 05). *PyCaret*. Retrieved from PyCaret: https://pycaret.org/

18. H2O.ai. (2018, June 12). *H2O.ai.* Retrieved from Time is Money! Automate Your Time-Series Forecasts with Driverless AI: https://www.h2o.ai/blog/time-money-automate-time-series-forecasts-driverless-ai/

19. Perktold, J., Seabold, S., & Taylor, J. (2020, Feb 21). *Time Series Analysis.* Retrieved from statsmodels v.0.11.1: https://www.statsmodels.org/stable/tsa.html

20. scikit-learn developers. (2019). *Developing scikit-learn estimators*. Retrieved from Scikit Learn: https://scikit-learn.org/stable/developers/develop.html

21. Dauphin, Y. N., Pascanu, R., Gulcehre, C., Cho, K., Ganguli, S., & Bengio, Y. (2014). Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. *Cornell University Library*.

22. Zhang, D., & Cooper, W. L. (2009). Pricing substitutable flights in airline revenue management. *European Journal of Operational Research, 197*(3), 848-861.

23. Gallego, G., & van Ryzin, G. (1994). Optimal Dynamic Pricing of Inventories with Stochastic Demand over Finite Horizons. *Management Science, 40*(8), 999-1020.

24. Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. (2016). Taking the Human Out of the Loop: A Review of Bayesian Optimization. *IEEE*, 148-175.

25. D.-R. Lui and T.-T. S. Shih, "Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences," Journal of Systems and Software, vol. 77, no. 2, pp. 181-191, 2005.

26. Hui, K.-L. (2004). Product Variety Under Brand Influence: An Empirical Investigation of Personal Computer Dema. Management Science, 50(5), 686-700.

27. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space.

28. Explosion.ai. (2020). *Word Vectors and Semantic Similarity.* Retrieved June 10, 2020, from https://spacy.io/usage/vectors-similarity

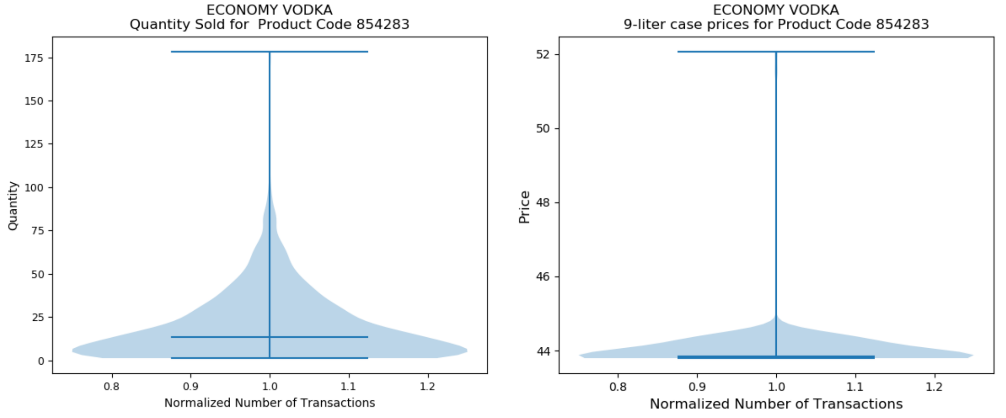
29. Brownlee , J. (2017, May 8). How to Convert a Time Series to a Supervised Learning Problem in Python. (Machine Learning Mastery) Retrieved July 10, 2020, from https://machinelearningmastery.com/convert-time-series-supervised-learning-problem-python/

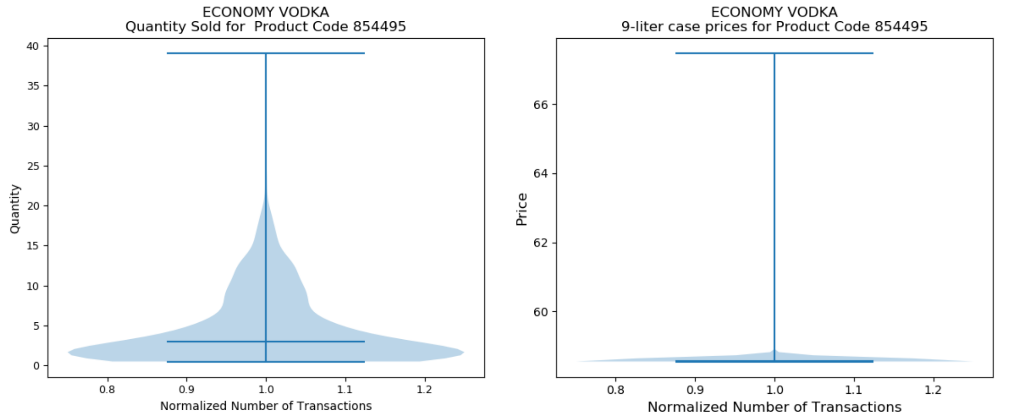
30. Seshadri, R., Chen, K., & Gupta, N. (2020, July). Auto-TS. (Auto-TS) Retrieved July 2020, from https://github.com/AutoViML/Auto\_TS

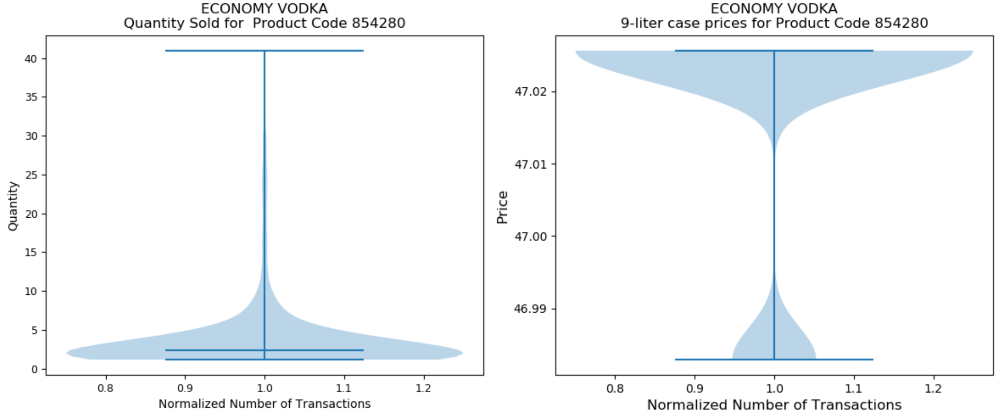
# Appendix A

## Distribution of quantity sold and price for top products and top customers.

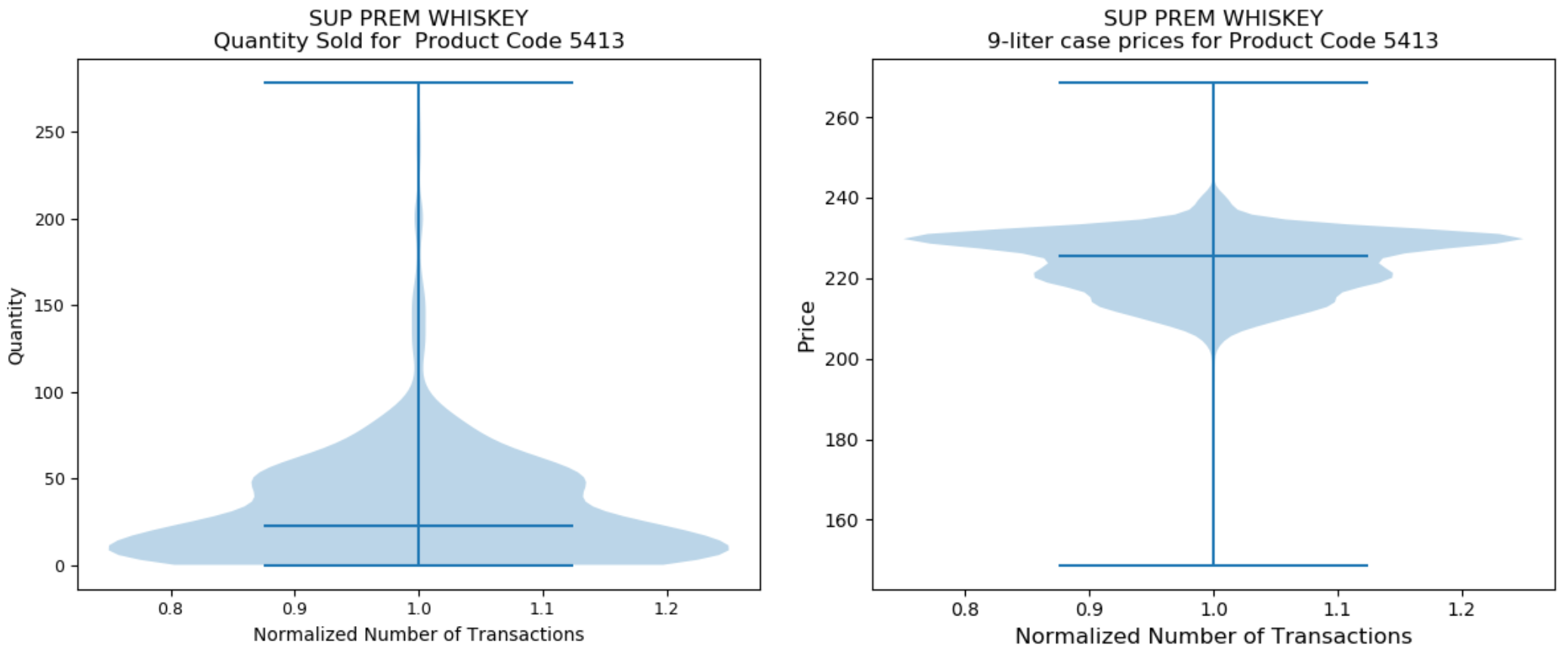
### Economy Vodka category.

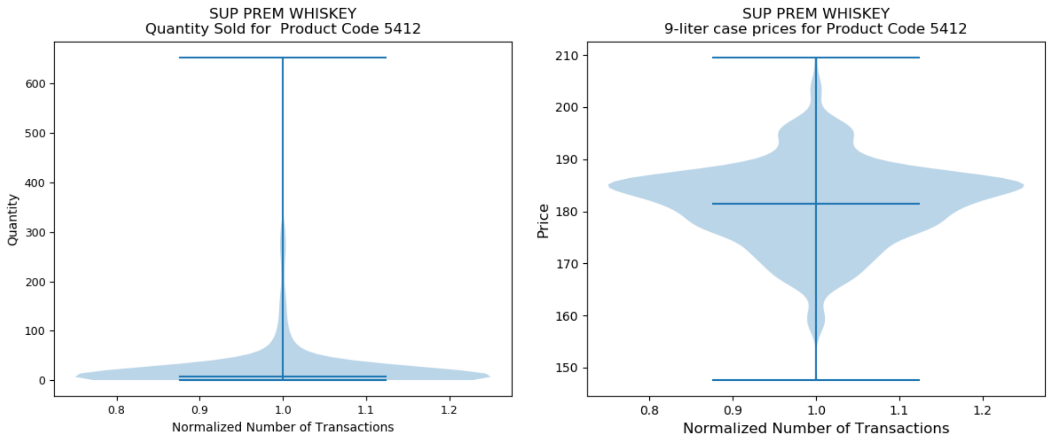


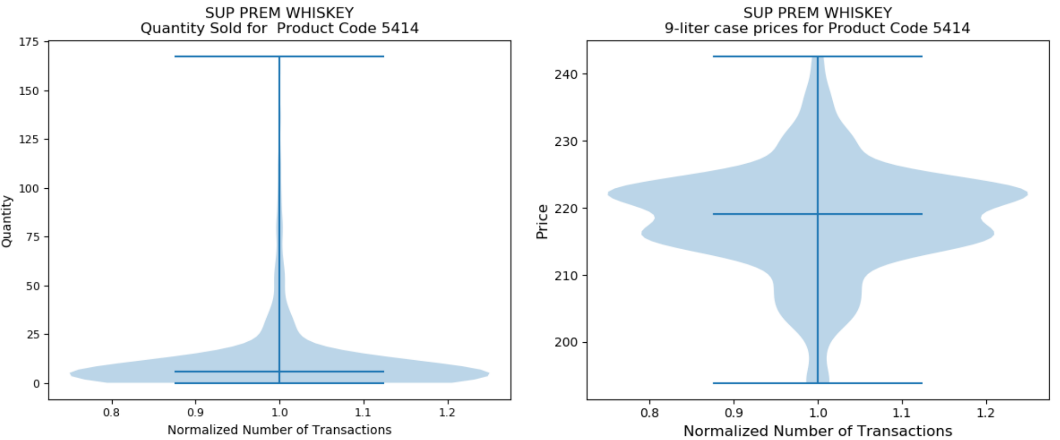




### Super Premium Whiskey category.

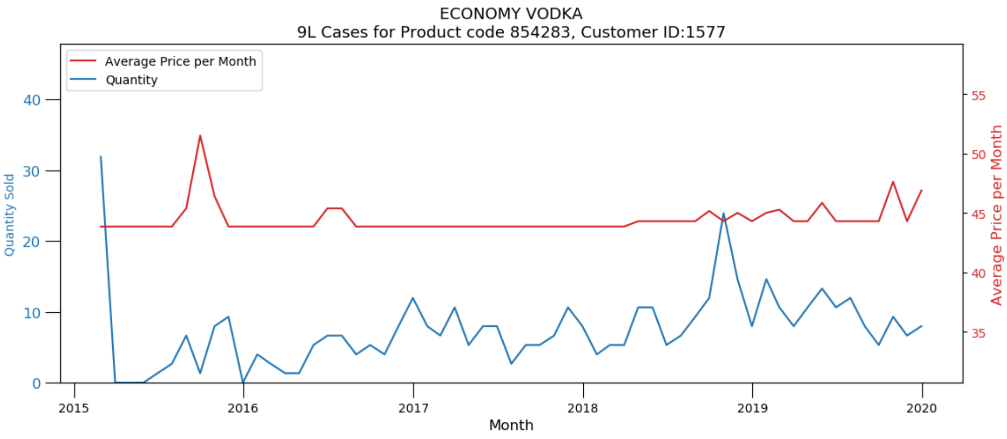


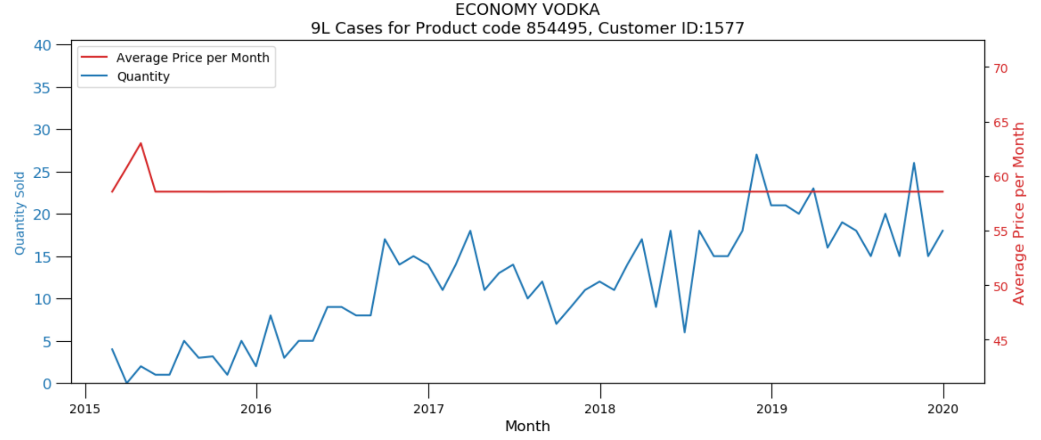


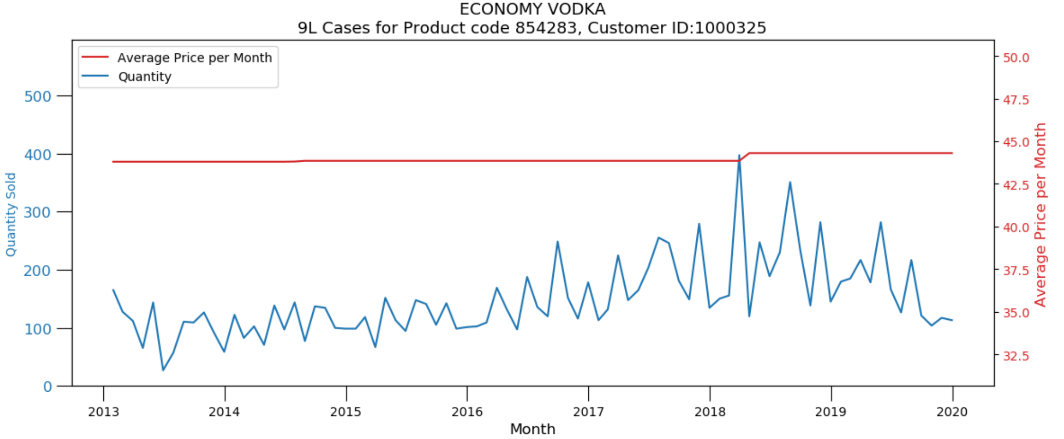


## Distribution of quantity sold and price over time for the top products and top customers

### Economy Vodka category







### Super Premium Whiskey category

