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 and machine learning models, and an automated machine learning platform are used to forecast the demand for products considering the impact of pricing. The techniques used allow for accurate modelling of the customer’s buying patterns including cross effects such as cannibalization, and the halo effect. The paper uses data from 2013 to 2019 for Super Premium Whiskeys and Economy Vodkas 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. 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 strategy 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 and 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 other 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. 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 non-symmetric as well. How product “A” influences product “B” is not the same as how product “B” influence products “A” [4]. Moreover, the change in the demand of a product may not only be influenced by switching behavior but could be 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 more recently deep neural networks (such as Recurrent Neural Networks - RNN and Long Short Term Memories - LSTM) can consider a complex set of interdependent and sequential variables to unveil patterns and 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 and the increase in computing power can help navigate different pricing scenarios more efficiently and find the ideal price point for various products jointly.

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

The remainder of the paper begins with a review of the literature studying price elasticity, cross effects, demand modeling, automated machine learning and optimization techniques. Next, this study outlines the methods used to accomplish these objectives followed by the results, and concludes with the discussion on the relevance to pricing managers and the ethical impact of this work.

# 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 pricing to drive revenue can still work since core customers are still likely to purchase products when price is increased and this is likely to offset the loss of revenue from fringe customers who are likely to stop buying if price increases too much [4].

## Cannibalization and Cross-Elasticity

Any sales bump derived from price reduction can be either 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 thus, important to decipher between the various sources of increase in demand since some of them are temporary and only lead to an acceleration of revenue and not a true increase [3].

Switching behavior (e.g. cannibalization) is of specific 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, i.e. product “A” may be able to influence the demand for product “B” more so than product “B” can influence the price of product “A”, can lead to more complexity in determining cross-elasticity [4].

## Modeling Cannibalization

Not much literature has been devoted to modeling revenue management taking price elasticity into consideration [5], [3]. What literature is available focuses 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 more non-linear in nature and not much research seems to have been done in this regard.

Additionally, most of the literature is concentrated on perishable products (a product whose utility expires after a certain time) and at the consumer market where approaches to price optimization in one industry such as airlines may not be suitable for another industries such as hotels [5], [9]. 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]. However, this approach lacks the scalability as not every company may have 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.

## Model 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 at a per customer and 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].

## 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 had to be 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 Ensembling 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 based approaches tend to overfit. In addition, while both these papers focused less than 5 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 pricing optimization. Much of the literature though has focused on discrete optimization based approach 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 optimization, have been discussed where heuristics are developed in order to simplify the optimization problem [22], [23].

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.

## Addressing the Gaps

While the literature review identified possible shortcomings, this study aims to address some of these, specifically:

1. Previous literature has focused predominantly on the perishable product space such as the airline industry, hotel industry etc. where there is a strict deadline by which the product needs to be sold. 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 sit in the warehouse for a little longer.
2. This paper is focused on the distributor market. The solution to a direct consumer industry focusing on perishable products such as the airline industry may not be applicable to other direct consumer markets such as the hotel industry [5]. 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, publicly available literature 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.
3. Being a data driven approach, the modeling methods in this paper 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 will be 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. This also helps reduce the dimensionality of the problem space which assists the optimizer in finding the global minima more easily. True increase in demand will be visible if none of the above effects are found.
4. A consumer market is focused on thousands of customers with sporadic purchasing patterns. Hence, 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. 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.
5. Despite modeling demand for fewer customers than a direct consumer market, models need to be built for all products within certain subcategories. This is still quite substantial and warrants the need to develop an automated time series forecasting tool. This study thus, aims to develop this as a by-product of this research.

# Methods

In order to address the research objectives, this paper focuses on data from a distributor of alcoholic beverages. This research uses data from 2013 to 2019, sampled on a weekly basis (maximum of 364 observations per product). This paper focuses on two distinct markets – “Super Premium Whiskeys” and “Economy Vodkas”. These markets represent opposite ends of the spectrum. “Super Premium Whiskeys” are a relatively lower volume higher margin business, whereas “Economy Vodkas” are a higher volume, lower margin business. Decreasing the price of “Super Premium Whiskeys” may spur demand and increase sales and revenue, whereas the ability to charge even a few dollars more for “Economy Vodkas” (without impacting the demand negatively) could add substantially to the bottom-line.

## 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 would be considered in a category while modeling the demand for any product in the category. This will give the most accurate measure of cross effects. However, due to limited amount of data, this approach does not work since lagged variables must also considered in the analysis of the demand. Hence, this study narrows the number of influential products down to a handful. 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.

The ideal measure of similarity between products should consider how much 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 have been 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. 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. 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 mean 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. This final score represents how much two products are similar and interchangeable. The similarity score is calculated for pairs of products purchased together in the same period. This can be visualized using a heatmap as shown in Fig. 1. This figure shows the similarity score matrix for the 15 most popular products in the “Economy Vodkas” product category for a customer chain and clearly depicts how some products with similar name and similar purchase frequency have a higher score.

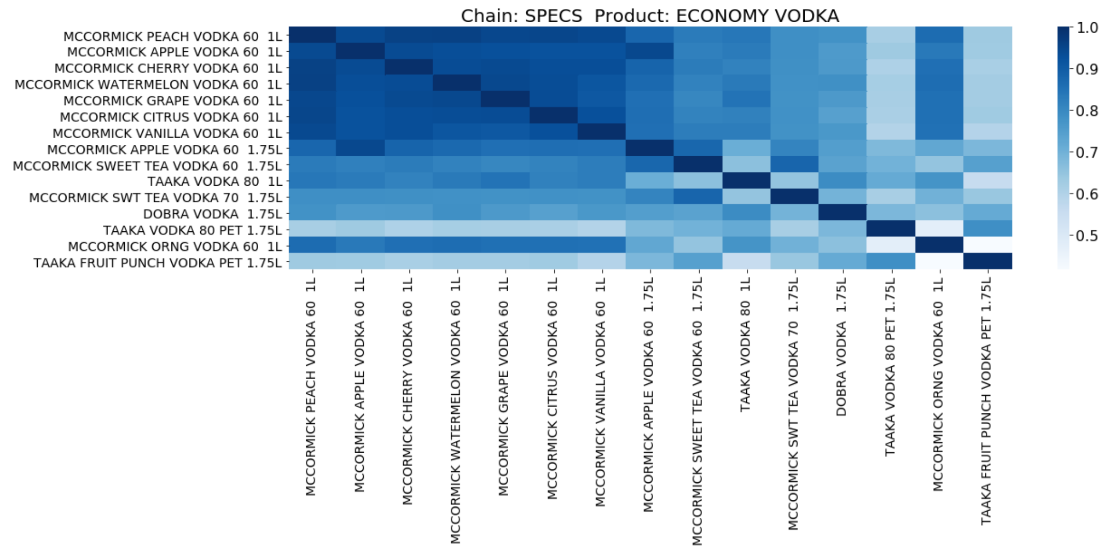


Fig. 1. Similarity matrix for “Economy Vodka” for a customer chain.

## Demand Modeling

<placeholder>

* Seasonal ARIMA with exogenous variables,
* Vector Autoregressive Models,
* Neural Networks such as LSTMs,
* Prophet Library from Facebook,
* PyFlux,
* others

</placeholder>

## AutoML

<placeholder>

* scikit-learn,
* statsmodels,
* rolling window analysis

</placeholder>

## Optimization

<placeholder>

* Bayesian Optimization
* Constraints on optimizer to a “practical” price point (cannot be less than say 80% of previous minimum or more than 120% of previous max). No constraints on supply.

</placeholder>

# Results

<placeholder>

1. Cross-cannibalization and Halo Effects
   1. These items [\_\_\_\_\_] had the highest influence on each other
   2. Demand Forecasting
      1. ASE
      2. Model Fit Metrics
      3. Confidence
   3. Price Elasticity
      1. Metrics, Charts to show the Price Elasticity
      2. Variable Importance?
   4. Optimal Revenue Estimation
      1. Supply Quantity
      2. Associated Demand
      3. Price points
      4. Anonymized projected revenue per product and overall

</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. Lui, D.-R., & Shih, T.-T. (2005). Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences. *Journal of Systems and Software, 77*(2), 181-191.

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