Pricing Optimization in Consumer Retail

Indhujha Natarajan
Harrisburg University

Abstract

Pricing is a vast, inter-disciplinary field which borrows from marketing, finance, economics and analytics. I have always been interested in the Pricing field and this paper builds on my previous unpublished work. The Four Ps of marketing – Product, Price, Promotion and Place - are the foundation of any transaction. The Price is a crucial driver especially in a B2C (Business to Consumer) relationship as the Price is what nails the buying decision as there is no room for a consumer to negotiate. Pricing being inter-disciplinary uses the principles of marketing, the principal ratios of finance, the quantitative rigor of analytics and the data structures of computer science. I looked at applying this multi-disciplinary approach for pricing in the consumer retail industry. Through an analysis of the transactional data set (created by Chen, Sain & Guo in 2012) which contains all the transactions occurring between 12/01/2010 and 12/09/2011 for a UK-based and registered non-store online retail, I computed the elasticity of the products sold using the log-log model and implemented the model using regression. I also logically grouped products using the K-means clustering algorithm to improve estimations. This will help us understand the price sensitivity of the items/logical item groups and thereby maximize revenue. I also built a Holt Winters time series model to know the high-level demand variation using aggregate units of all items during each day. This helps in forecasting – understanding the trend of sales units. Maximizing revenue or sales dollars is one of the primary goals of business as it will directly impact the bottom line.

Keywords: Marketing, retail, price, elasticity, regression, log-log, clustering, revenue, Holt-Winters

Pricing Optimization in Consumer Retail

Introduction

As mentioned in Natarajan (2017), the effective quantitative theory of demand was given by Antoine Augustin Cournot around 1838. He was the economist who introduced the concept of demand as a mathematical function of price. If D=F(p) is the symbolic expression of the relation between the amount of commodity demanded, D, and the price per unit of commodity, p ,then our aim is to find for what value of p the product p F(p) (Revenue) is the maximum (Moore,1922). In the Figure 1. we have the familiar graph of the law of demand, where x=the amount of commodity demanded and y=price per unit of commodity; and DD' is the demand curve (Natarajan, 2017).

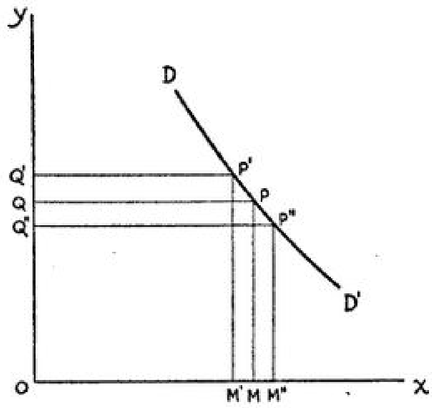

FIGURE 1

Figure 1. Demand curve. Reprinted from "Elasticity of Demand and Flexibility of Prices", by H. Moore,1922, Journal of the American Statistical Association, 18(137), p. 10. doi:10.2307/2277462

We need to come up with an approximate determination of the value of p which renders the product p*F(p) a maximum – an optimization problem. This has been given by Professor Marshall's form of statement where he describes demand as elastic or inelastic according as percentage in quantity demanded/percentage change in price is less or greater than unity. Empirically estimating this demand function has been challenging (Natarajan, 2017).

Literature Review

Various methods have been suggested to assist in estimating the demand function. Some of the methods from the fields of marketing, economics, analytics and machine learning are as below.

Marketing methods and models

These methods can be of two types as explained below

Consumer Interviews and Market Experiments. Two methods widely used for this purpose are Consumer Interviews and Market experiments (Natarajan, 2017). In Consumer interviews firms frequently interview consumers to understand their buying habits, intentions, willingness to pay and value created. However, these surveys have some well-known limitations. These surveys entail answering hypothetical questions regarding pricing which don't result in accurate answers. However, more subtle approaches can be useful. Interviews indicated that most buyers of a particular baby food selected it on their doctor's recommendation and that most of them knew very little about prices of substitutes (Dean, 1980). This information, together with other data, suggested that the price elasticity of demand was quite low in absolute value (Dean, 1980).

Another method of estimating demand functions is to run direct experiments in the real world. The idea is to vary the price of the product keeping other market conditions stable (Natarajan, 2017). The disadvantages of this approach are that direct experimentation can entail risk and loss of revenue. Also, it is really difficult to conduct really controlled experiments and often the tests are run for a relatively short duration and hence don't provide all the information that is needed (DeBruicker, Quelch, & Ward 1980). Nonetheless, these experiments can be of great value as indicated in the L'eggs products example in DeBruicker, Quelch, & Ward (1980) where a coupon and three different promotion price/package combinations were tested to zero in on the one that generated the maximum net cumulative short-term sales increase.

In the two weeks tested, the low promotions gave a 42 percent increase in sales versus the 37 percent increase in sales given by points with slightly increased promotion prices. Another test could be testing promotion price endings using two alternatives: 1) all prices with a .99 cent ending (e.g. 2.99,3.99) and 2) prices with odd number endings (4.23,5.97). The result of this test was that the .99 cent ending performed better. It is also important to align on the goal of the test. In most cases revenue is a good metric for retail. But in some cases, one may need to prioritize margin. These tests dealt with short term price reductions in terms of promotions but may include other tests such as bundling of products, regular price changes etc.

RFM Model. In addition to the above socio-demographic experiments, the RFM model proposed by Hughes (2000) can also be used to understand elasticity of demand. Though this model is widely used for understanding customer lifetime value, it can give an idea regarding how price sensitive your customer segments are (Natarajan, 2017). According to You, Si, Zhang, Zeng, Leung, & Li (2015) the RFM segmentation model is a model that differentiates important customers according to three variables: customers' consumption interval, frequency and amount of money as explained below:

(1) R represents "recency", which is defined as the interval between the time of the latest consuming behavior and the present; the shorter the interval, the greater the value of R.

(2) F represents "frequency", which is defined as the frequency of consuming behavior over a period of time.

(3) M represents "monetary", which is defined as money value of consumption over a period of time. This model uses a more aggregative approach unlike the experiments which may be conducted at a more granular level (Natarajan, 2017). RFM models are easy to understand and don't require knowledge of any statistical software (Webber, 2013). However, the RFM approach only looks at past data and does not involve a predictive approach (Webber, 2013).

Economics Models

We have looked at the basic demand function and the various techniques that are used to collect the data needed to estimate this function (drawn from marketing fields). Now, let us walk through the functional forms that can be used to estimate demand models (drawn from economic theory). These can be classified into 4 categories (Oum, 1989):

a) Linear demand model

b) Log Linear demand model

c) Logit model

d) Translog demand system

Let us look at the intuitive properties of each model as explained by Oum (1989):

(a) Linear demand model The linear function has been extensively used in sales forecasting because it is simple to estimate and easy to interpret. However, as discussed previously the assumption of a linear effect may not be realistic.

(b) Log-Linear demand model The log-linear (double logarithmic or Cobb-Douglas) model specifies the logarithmic of quantity sold as a linear function of the logarithms of potential determinants, such as price and quality variables. This is the most widely used model because (a) the coefficients themselves are the respective elasticities of demand; (b) the log-linear function is capable of modeling non-linear effects and (c) it resembles the demand function obtainable from Cobb-Douglas utility function. The main drawback of this model is that elasticity is invariant across all data points.

(c) Aggregate Logit model The logit model is used because (a) the discrete choice version of the logit model provides an intuitive and theoretical rationale (b) it can be estimated using any regression program (c) the S- shaped curve realistically describes the behavior of decision makers. Oum has used this model for modeling market shares of alternative modes of transport. This model has two forms (a) ratio form and (b) difference form. The difference form is preferred as it is independent of base taken.

(d) Translog demand system Since the mid-1970s, economists have begun to use a demand system derived from a flexible utility or production function. This flexible function provides a quadratic approximation to the unknown true function. As per Oum, this includes the translog (Christensen, Jorgenson and Lau, 1973), generalized Leontief (Diewert, 1971) and generalized Cobb-Douglas functions (Diewert, 1973). The translog function is the most widely used of all.

The demand system derived from the translog utility has is consistent with the neoclassical theory and allows free variation of the elasticities and cross-elasticities. These advantages, of course, come at the cost of substantially increased computation when compared with the above 3 models. Oum had explained the above demand models for the transportation industry. However, he emphasizes that the theoretical and methodological discussions are directly applicable to demand studies for any goods and services. Extending the above discussion to consumer retail, we see the below three model functions to represent demand. These equations are obtained from economic demand theory. The model functions are used in my daily work at a retail company are shown in Figure 2.

It can be inferred from the equations above that log/log model is easy to implement and interpret as the elasticity is the coefficient as opposed to the other two where the elasticity is a function of the coefficient and price. The other two models are similar. The difference is the second one gives the price whereas the third one gives a discount value.

Machine Learning methods and models

Attribute Selection and Clustering methods. Attribute selection techniques and machine learning methods such as clustering are great tools to perform market segmentation. Clustering in particular is very useful as it helps in identifying similarity between items. This has been summarized by Singh & Rumantir (2015) in a concise table which is reproduced in Table 1.

As noted in Singh & Rumantir, all of these techniques segment markets or customers based on the attributed selected. However, I am more interested in the key

Model 1 **Log/Log**

$\log(Q) = \alpha + \beta \log(P)$

$\varepsilon = \beta$

Optimal Price $= \varepsilon * \text{Cost} / (1 + \varepsilon)$

$Q_1 = Q_0 * (P_1/P_0)^\wedge \beta$

Model 2 **Log/Linear**

$\log(Q) = \alpha + \beta * P$

$\varepsilon = \beta * P$

Optimal Price $= -1/\beta + \text{Cost}$

$Q_1 = Q_0 * \exp(\beta * (P_1 - P_0))$

Model 3 **Markdown**

$\log(Q) = \alpha + \gamma * (\text{pct discount})$

$\varepsilon = \gamma * (1 - \text{pct discount})$

Optimal Price $= -1/\beta + \text{Cost}$

$Q_1 = Q_0 * \exp(\beta * (P_1 - P_0))$

(equivalent to Log/Linear)

Figure 2. Equations used in my daily work.

market driver which is Price and how demand is affected by the price of the product.

Like the RFM Model, clustering also uses an aggregative approach (Natarajan, 2017). In You et al. (2015), we see a framework that uses the K-means clustering algorithm on RFM data to predict the amount of inventory that needs to be carried for each customer category. This is done from a customer category perspective and not from a product demand perspective and it mainly looks at identifying the commercial activities of key customers for a particular product (Natarajan, 2017).

Regression. Another way of estimating the demand function is to use a common statistical, machine learning technique called regression. The purpose of a regression analysis is to obtain the mathematical equation for a line that describes the average relationship between the dependent and independent variables. Regression analysis assumes that the mean value of Y, given the value of X, is a linear function of

Table 1

Market segmentation using clustering/attribute selection.

Related work	Attribute Selection Techniques			Clustering Techniques	
	Socio-demographic Analysis	RFM Analysis	K-means	Hierarchical clustering	Other
(Chen et al., 2012) (Lefait & Keehadi,2010)		X	X		
(Hsieh 2004)		X			X - Neural Network
(Bizani & Tarokh, 2011)		X	X		X-Unsupervised Learning Vector Quantization
(Chen et al., 2012)		X	X		
(Olson et al., 2009)		X			
(Namvar et al., 2009)	X	X	X		X
(Kim et al., 2005) & (Lee & Park,2005)	X				X-Neural Networks
(Dennis et al., 2003)	X				X
(Ho et al., 2012)			X-Genetic Algorithms		
(Salvador & Chan, 2004)			X	X	
(D.Gaur & S.Gaur, 2013)			X	X	X
(Zakrzewska & Murlewski, 2005)			X	X	X-Density based clustering
(Alam et al.,2010) (Yoon et al.,2013)				X	
(Li et al.,2009)				X-Chameleon	
(Suib & Deris,2008)				X-Hierarchical Pattern based clustering	

Note. Reprinted from Two-Tiered Clustering Classification Experiments for Market Segmentation of EFTPOS Retailers, by A. Singh and R. Grace, 2015, Australasian Journal of Information Systems,19, p. 119.

X ("Estimating demand functions", n.d.). The assumptions of regression analysis should at least be approximately satisfied to carry out this analysis (Natarajan, 2017). As shown in Figure 1, the demand curve is a curve-linear function. Hence the linearity assumption is not valid when using the data directly (Natarajan, 2017). Hence suitable transformations need to be applied to use this technique to estimate demand.

Time Series forecasting. Gahirwal and Vijayalakshmi (2009) define time series as a sequence of data points, measured typically at successive times spaced at uniform time intervals. In the same paper Gahirwal and Vijayalakshmi (2009) define time series forecasting as the use of a model to forecast future events based on past events to predict data points before they are measured. There is a need to have a view of sales over time as it helps to estimate the inter-temporal nature of demand. A very common method used for forecasting is the moving average method. In this method, the most recent data points are used to forecast the future. This is very applicable in retail as the recent trend has more weightage than other long-term projections.

Exponentially weighted moving averages (EWMA) were widely used for smoothing data around 1957. However, this forecast was limiting. It was at this time that Holt realized that the EWMA concept can be used not only as a way to smooth the level of a variable but also to smooth trends, seasonals and other components that can be forecasted. Winters worked on assessing the performance of the method (Holt,2004). As indicated in Charles Holt's paper, the Holt-Winters method greatly improved forecast accuracy, especially for retail, as this method was made by Holt when he had to make forecasts for over 200,000 products of a sand-paper company.

It has also been suggested to combine various time series models to improve forecast accuracy. Once the series is decomposed into the seasonal, trend and irregular components, statistical methods like ARIMA and Holt Winters can be used to forecast each of these components. This can be done by an approach called Association mining (Gahirwal and Vijayalakshmi, 2009).

Neural Network. As per Penpece and Elma (2014), Artificial neural networks (ANN) can be defined as a highly connected array of elementary processors called neurons. Forecasting sales quantity and sale revenue is very important in retail as it is key in making pricing and inventory decisions. Neurons help in information processing and pattern identification which aids in forecasting (Penpece and Elma, 2014).

The most used neural network is a feed forward neural network which has three layers - the input layer, the hidden layer and the output layer. At the input layer, different independent variables are used to forecast the output layer, which consists of the dependent variable (Penpece and Elma, 2014). A challenging problem in Neural is identify the number of neurons in the hidden layer.Neural Networks operate in two phases - training phase and the forecasting phase. If optimum number of neurons are chosen in the hidden layer then good results can be obtained within less training time. Karsoliya (as cited in Penpece and Elma, 2014) gives 3 rules for determining the number of neurons in the hidden layer.

(1) The number of hidden layer neurons is 2/3 (or 70 percent to 90 percent) of the size of the input layer.

(2) The number of hidden layer neurons should be less than twice of the number of neurons in input layer.

(3) The size of the hidden layer neurons is between the input layer size and the output layer size.

In their paper, Penpece and Elma (2014) explain how Neural Networks can be used for forecasting sales revenue at grocery retailing industry companies. Their neural network has 3 independent variables as neurons in the input layer and 2 neurons in the hidden layer.

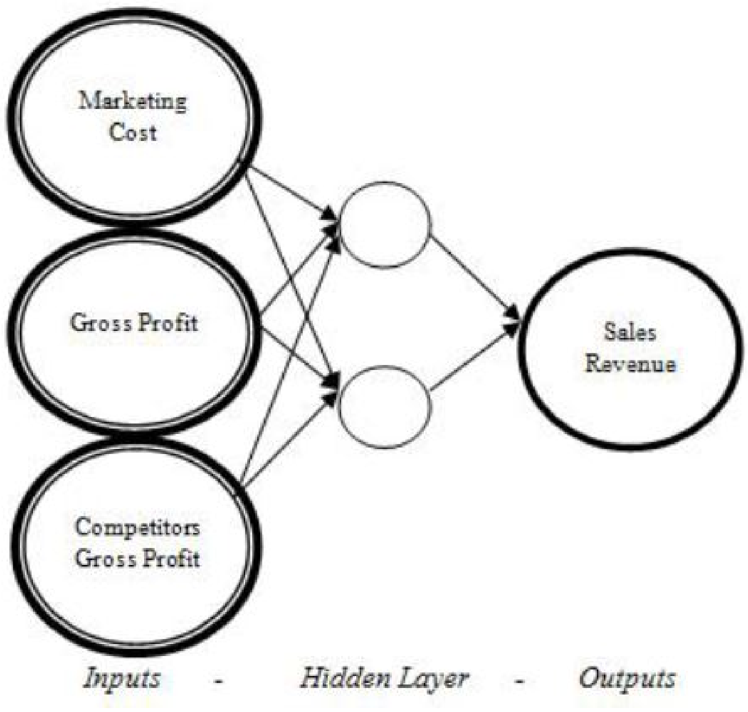

Figure 3. Neural Network to forecast Sales Revenue. Reprinted from "Predicting Sales Revenue by Using Artificial Neural Network in Grocery Retailing Industry: A Case Study in Turkey", by D. Penpece, & O.E. Elma, 2014, International Journal of Trade, Economics and Finance, 5(5), p. 438. https://doi.org/10.7763/IJTEF.2014.V5.411

In Figure 3, it is seen that the independent variables considered are Marketing cost, Gross Profit and Competitors Gross Profit. Though these are important variables in the marketing mix, I feel it is not very easy to get to know a competitor's gross profit in most other retail sectors. Hence I believe that using the company's own and competitor's prices as independent variables is a better approach. Penpece and Elma (2014) say that the forecasted results of their study indicate that the outputs are bigger or smaller than the actual data by only 10 percent which means that the results are highly accurate. One of the reasons for the high forecast accuracy can be attributed to ANN being better able to handle non-linearities in the data (Penpece and Elma, 2014).

Kavyashri, Jayaram, and Jeevitesh (2016) explain how neural networks can be used for dynamic pricing in online retail. They use the back propagation algorithm to minimize error and the hyperbolic tangent as the activation function. Kavyashri, Jayaram, & Jeevitesh (2016) use the neuralnet R package to implement the neural network whereas Penpece and Elma (2014) use the excel add-in called NeuroXL to implement the neural network.

Although Neural networks have the advantages explained above, it also has some disadvantages. The first being that it is not very intuitive to understand how it derives the relationship between variables. The other disadvantage is that neural networks require a huge amount of data for training.

We see from the literature review the advantages and drawbacks of various approaches drawn from the fields of marketing, machine learning and economic theory. Some papers look at the problem from a customer or segmentation perspective. The objective of my paper is to create a simple model that looks at demand from an item/logical item group perspective to find the optimal price thereby leading to the ultimate goal of revenue/sales dollars maximization.

Before delving into the methodology where various analytical and machine learning techniques are used an important area of consideration irrespective of the analytical technique chosen is explained below.

Behavioral Economics – The Psychology of Pricing

The intent behind all the various techniques explained is to come up with the optimum price that's improves sales or increases demand for the retail companies' products. For this to happen we must consider the fact that human beings are NOT completely rational. This is why see that prices ending in .99 or .95 work better for retailers than prices ending in .00. Kenneth J. Wisniewski (as cited in Lindstrom, 2012) in his study gave the example of margarine - Margarine dropped from 89 cents to 71 cents and led to a sales increase of 65%. However, when the price fell by two more cents to 69 cents, the sales jumped by an astounding 222% (Lindstrom, 2012). This shows that two pennies are worth more than seem to be.

The concept of 99 cent ending has been around for decades. Lindstrom (2012) gives examples such as the 99 Cent Only stores started by David Gold and his wife in 1982 and of Steve Jobs who managed to liberate the music industry with his 99-cents-per-song iTunes strategy. In the retail company I work, we conducted an A/B test where one group of stores received prices directly from the model without any properly made endings like .99 or .95 (e.g. 25.71, 19.83) and the for the other group the model prices were made to end in 99 cents (e.g. 24.99,19.99). The second group performed much better than the first one.

Another important criterion is the messaging of the price. As Lindstrom (2012) points out, having a dollar sign or even the word dollar on the store sign makes the customer feels sad about parting with his/her money and hence discourages spending. Therefore, leaving out the dollars sign on the store signs works better to improve sales. A good example of this price messaging is illustrated by Professor Ariely's study (as cited in Greifenberger, 2018) which was about the three subscription pricing tiers for a periodical. As explain in Greifenberger (2018) the three subscription tiers were as below.

(1) On-line only option $59

(2) Print option $125

(3) Print and on-line option $125

The study showed the expected buying behavior as shown in Table 2.

Table 2

Expected buying behavior with three offer options.

Offer	Percent Selected
On-line only option	16%
Print option	0%
Print and on-line option	84%

Note. Reprinted from Behavioral Economics Powered by Machine Learning and AI Techniques: A True Path to Optimized Pricing, by P. Greifenberger, 2018. Retrieved June 10, 2018, from https://www.linkedin.com/pulse/behavioral-economics-powered-machine-learning-ai-true-greifenberger

It is seen that the second option clearly has zero economic utility. But when the second option was eliminated the buying behavior changed as shown in Table 3.

Table 3

Expected buying behavior with three offer options.

Offer	Percent Selected Test One	Percent Selected Test Two
On-line only option	16%	68%
Print option	0%	eliminated
Print and on-line option	84%	32%

Note. Reprinted from Behavioral Economics Powered by Machine Learning and AI Techniques: A True Path to Optimized Pricing, by P. Greifenberger, 2018. Retrieved June 10, 2018, from https://www.linkedin.com/pulse/behavioral-economics-powered-machine-learning-ai-true-greifenberger

This study shows that the third offer looked more desirable when compared to the weak offer 2 (Greifenberger, 2018). This study proves that decision making can be largely influenced by a cleverly constructed offer (Greifenberger, 2018). Advanced algorithms use pricing history to predict the optimal price but they are limited to the parameters available in the training data. Hence behavior economics needs to be incorporated into the model by a process that consists of consulting domain experts, running market surveys and conducting price experiments (A/B testing) in a representative market (Greifenberger, 2018). The learnings from these processes can be used in the training data of machine learning algorithms which bring about a positive cycle of learning (Greifenberger, 2018).

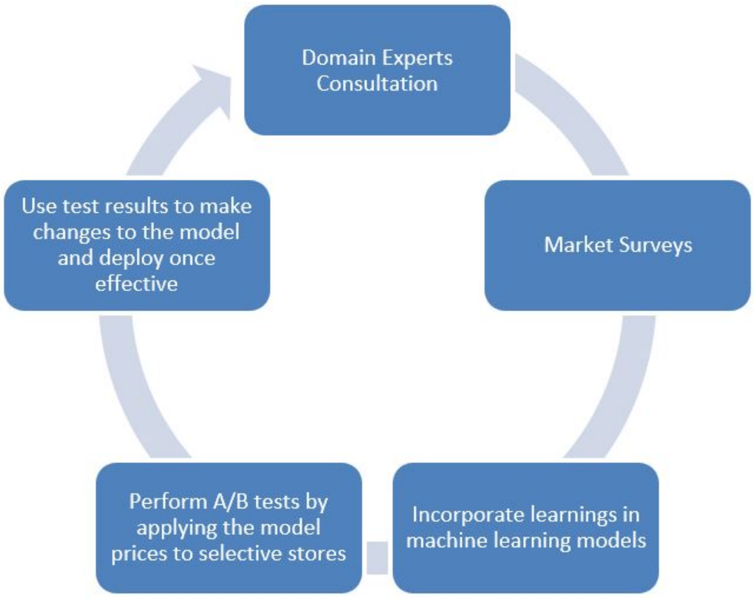

Figure 4. Balanced cycle of Analytics and Behavioral Economics. Data for figure from from Behavioral Economics Powered by Machine Learning and AI Techniques: A True Path to Optimized Pricing, by P. Greifenberger, 2018. Retrieved June 10, 2018, from https://www.linkedin.com/pulse/behavioral-economics-powered-machine-learning-ai-true-greifenberger

As shown in Figure 4, one must maintain a balance between the analytical techniques used and behavioral economic factors. The data crunching should not make us miss the human element from the equation.

Research Objective

The method used in this paper is the machine learning technique called regression. This starts the epistemological debate if regression can represent demand accurately. The demand is curvilinear so transformations have been done to represent the data in a regression (Natarajan, 2017). The regression equation is derived from the well-known Cournot Theory and this model is widely known as the Log-Log model (Natarajan, 2017). The over-arching research aim is to maximize the revenue or sales dollars at an item/logical item group level. Revenue is taken as a driving factor as that is one that captures demand more accurately (Natarajan, 2017). Revenue is the product of unit price and number of units sold. The number of units sold is a function of the demand. Though Margin is important for the business to survive in the long run the more pertinent question especially in consumer retail is whether you are creating value and what people are willing to pay for the value created (Natarajan, 2017). Also, consumer retail is highly driven by seasonal products and hence the focus should be more on revenue than on margin as the cost is already sunk. Various approaches have been explained in this literature review. The common component among these approaches is that they all use transactional data (Natarajan, 2017). This data is aggregated at appropriate levels to achieve the goals set out by each paper. Granularity is an important decision to make when recommending any form of optimization. As indicated in my previous work (Natarajan, 2017) there are four specific questions that I am interested in which are as follows.

- Based on historical data, is it possible to plot the demand curve at a product/ product group level?
- Is it possible to find the optimal price point on the curve where the revenue is the maximum?
- Is it possible to estimate the elasticity $(\Delta Q)\Delta P$) or price sensitivity at the product or product group level in order to set prices in the future?
- Is it possible to group items based on price/demand to have meaningful item categories?

Once these questions are answered we can achieve the over-arching goal of revenue/sales dollars maximization. This is important in the field of pricing whose goal is the optimization of prices based on the observed demand for the product. In line with Pricing's inter-disciplinary nature, this paper combines marketing concepts, machine learning and economic models to answer the above questions and achieve the goal of pricing.

Methodology

The methodology involves using three different techniques – Simple Linear Regression, K-means Clustering and Time Series using Holt Winters. The granularity of data used for each technique is different – item level data versus aggregated data. The dataset used for this paper is from the below link:

`http://archive.ics.uci.edu/ml/datasets/online+retail`

This dataset is from the UCI Machine Learning repository which contains transactional data of an online retailer based in London - the important columns for this study being Quantity, UnitPrice and StockCode. This company mainly sells unique all occasion gifts.

Reasoning and Approach

The core model of this paper is built using Simple Linear Regression. Although the demand function is curvilinear, the Cournot Theory provides a basis to perform the necessary transformations to linearize the data. I chose the natural log transformation to perform this linearization as the interpretation of the regression coefficients is straightforward. Roberto Pedace (2013) explains the demand function as given below. Consider the demand function

$$Q = \alpha P^{\beta},$$

where Q is the quantity demanded, alpha is the shifting parameter, P is the price of the good, and the parameter beta is less than zero for a downward-sloping demand curve. This equation form needs to be transformed to apply the regression technique. If you take the natural log of both sides, you end up with

$$ln(Q) = ln(\alpha) + \beta ln(P)$$

Using calculus with a simple log-log model aids in the interpretation. Differentiating we get

$$\delta Q/Q = \beta \delta P/P$$

The term on the right-hand side is the percent change in P, and the term on the left-hand side is the percent change in Q, so β measures the elasticity.

Elasticity is the most important metric needed as that is what is used to calculate to optimal price at which the revenue is the maximum. The elasticity drives the shape of the curve. If β_1 is the elasticity, x is the quantity and y is the price, we get the following curves for various values of β_1 (Figure 5).

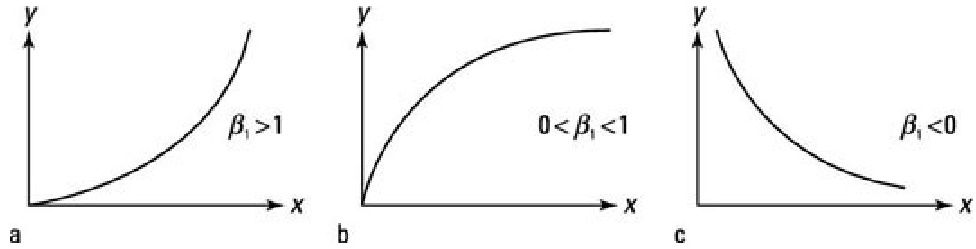

Figure 5. Demand curves for various elasticities. Reprinted from Econometrics and the Log-Log Model In dummies, by R. Pedace, 2013, Retrieved May 11, 2017, from http://www.dummies.com/education/economics/econometrics/econometrics-and-the-log-log-model/

Except for some extreme cases or unique retail segments, the elasticity is generally less than one in consumer retail – companies that sells fast moving consumer goods.

The trade-off for choosing a technique is between ease of implementation and accuracy. There are other more powerful machine learning techniques such as Artificial Neural Networks which have the ability to implicitly detect complex nonlinear relationships between dependent and independent variables. However, its disadvantages include the "black box" nature, greater computational burden and proneness to over-fitting (You et al., 2015). For day-to-day work in the retail environment, a simple model that delivers good results for the time invested is preferred to a computationally complex one.

The regression runs at items level. However, for some items there are not many data points. Hence it becomes necessary to logically group items based on their price and units sold. This is done by using the K-means clustering technique which uses the

Hartigan-Wong algorithm. This algorithm is often the fastest. I also used the Time series technique using the Holt Winters model on aggregated data. The Holt Winters model decomposes the time series into seasonal, trend and irregular components to model the historical data and use the parameters learned for predicting units sold during future time periods. This technique is very suitable for consumer retail as this industry has cyclic and seasonal trends and hence Holt Winters is suitable for this transactional dataset as well.

Implementation

Before the implementation of the modeling process, data cleaning needs to be performed first. Data cleaning involved the following steps (Natarajan, 2017).

(1) Removing missing values. Almost all columns are important and hence all missing values need to be removed.

(2) The transactions where the units sold were less than or equal to zero were removed. These were mostly canceled invoices and accounting adjustments that are not in the scope of this study.

(3) The same alpha-numeric stock code has different casing (e.g. 85123a and 85123A). This makes the model consider it as two different items. To ensure the items are properly identified, I converted all stock codes to uppercase.

After the clean item level data is obtained after the above steps, I do some Exploratory Data Analysis to understand the data better. These will be explained as the first part of results section.

After EDA, the clean item level data is subjected to the process flow of modeling using the three techniques which represented using the diagram in Figure 6.
The process was implemented as explained below.

Simple Regression Model: The regression was implemented using the lm function in R. The equation of the regression model is given below:
$lm(log(Quantity) \sim log(UnitPrice))$

The regression was run at item level using a 'for' loop where the model is applied

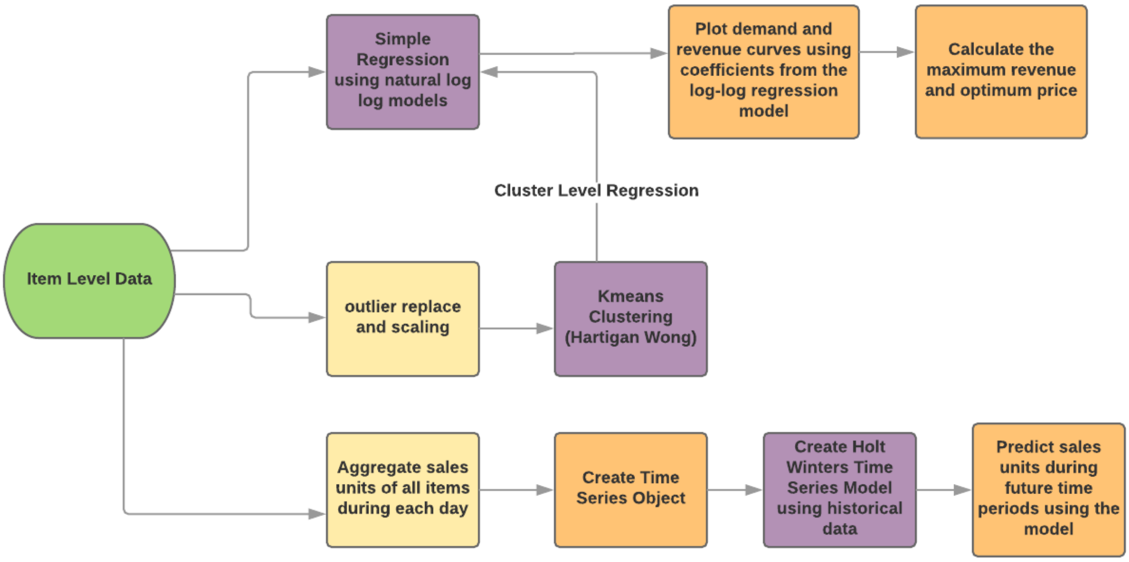

Figure 6. Process Flow of Pricing Optimization.

on a particular item during each iteration of the ‘for’ loop. There were about 3800 items and the for loop runs for each of the items.

The regression output contains the coefficients which was used to build demand and revenue curves. The data from the curve was then used to calculate the optimal price at which revenue was the maximum.

Clustering: The items with Adjusted R-squared as NaN are removed from the regression output. This regression output is then merged with the clean item level data. Before applying clustering, this merged dataset was aggregated to item,quantity and price level using the dplyr package as currently the data has items sold by day. To get unit price at this aggregate level, I first multiply quantity and unit price at the item-day level to get sales. Then while aggregating to item level, I divide the sum of sales by sum of quantity to get the unit price of the item. This aggregated item level data for the whole time period considered was further processed by replacing outliers with the median (using ’outliers’ package). The data was then scaled using the scale function. Scaling the data does not remove outliers and hence the outliers need to be removed before scaling. Outliers need to be removed before clustering as the clustering needs to generalize well without getting biased by outliers. The number of clusters need to be decided. For this, I used the scree plot which uses within groups sum of squares.

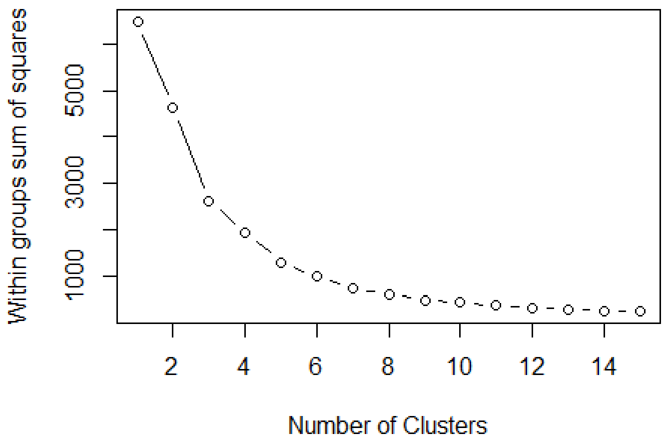

Figure 7. Scree-plot showing the graph of within groups sum of squares.

From Figure 7., we can see that the within groups sum of squares does not decrease much after eight clusters. The clustering was done using K-means after choosing the number of clusters. The approach for the actual K-means cluster modeling was derived from Johnson's (2017) article. After clustering, the clusters assignments were merged with the item level dataset. Then the regression model was applied at cluster level.

Time Series: There are three parts to the Time Series implementation as below:

Data Preparation: Preparing the data involves three steps as below:

(1)The item level data was aggregated to sales units of all items sold on each day for the time frame considered. To do time series forecasting, one needs to ensure that information was available for each day in the time period considered.

(2) On aggregating the data, it was noticed that days that were Saturdays or public holidays were missing in the data. The missing days in the aggregated data were included with units sold as zero as zero units were sold on these days.

(3) This aggregated and modified dataset also needs to be sorted in ascending order of Date in order to provide the sequence for time series.

Creating Time Series Object and Model: A time series object was then created using this sorted dataset. Further, the Holt Winters model was applied on this time series object.

Forecasting: This Holt Winters model was then used to predict sales units during future time periods using predict and forecast functions.

Results

The results obtained at each stage of the process is explained as follows.

Exploratory Data Analysis (EDA) Results

Firstly, I drew the histograms for the three main quantitative variables in the dataset - revenue, quantity and Unit Price. Revenue is derived by multiplying the quantity with the Unit Price. The histogram for revenue is shown in Figure 8. It is seen that the histogram is left skewed and mostly has values between 0 and around 14,500. The histogram for quantity sold is shown in Figure 9. It is seen that the histogram is left skewed and mostly has values between 0 and around 5,000.The histogram for Unit Price is shown in Figure 10. It is seen that the histogram is left skewed and mostly has values between 0 and around 500.

One thing noticed for the three histograms was that increasing the number of breaks did not increase the number of bars in the histogram. This means there is wide variation in the metric values between the various items.

A table is constructed to summarize the key statistical aspects of the three variables (Table 4). As explained before the variance is pretty high. These high values could be present because of accounting adjustments and fees applied which is not applicable for this study and hence can be ignored.

Next, I explore the other variables in the dataset against revenue. Firstly, the revenue obtained from each country is seen using a bar graph as shown in Figure 11. It was expected that the highest revenue would be from United Kingdom as the company is based in London. Figure 11 corroborates this expectation. United Kingdom is distantly followed by Netherlands, EIRE (Southern Ireland), Germany and France. This

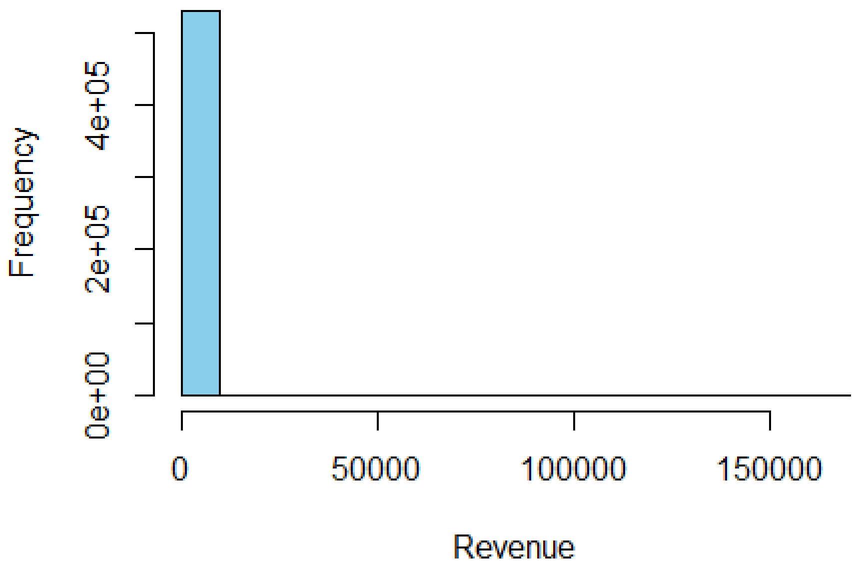

Figure 8. Histogram of Revenue.

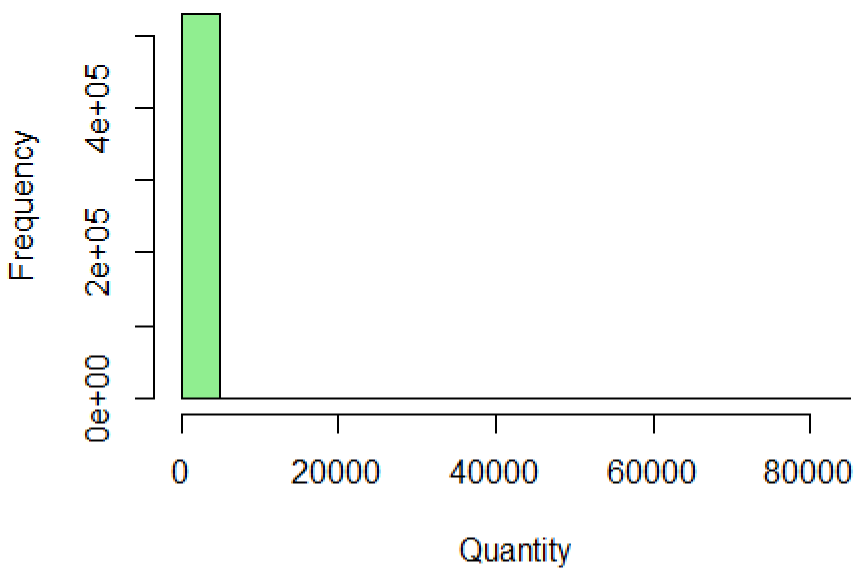

Figure 9. Histogram of Quantity.

indicates even though this is an online only store it has not expanded its customers beyond United Kingdom.

The next graph is a bar graph that plots the top 25 customer IDs based on revenue (Figure 12).It is generally observed in retail that it is the members of a store that are assigned a customer ID. The highest amount of revenue comes NA (Not

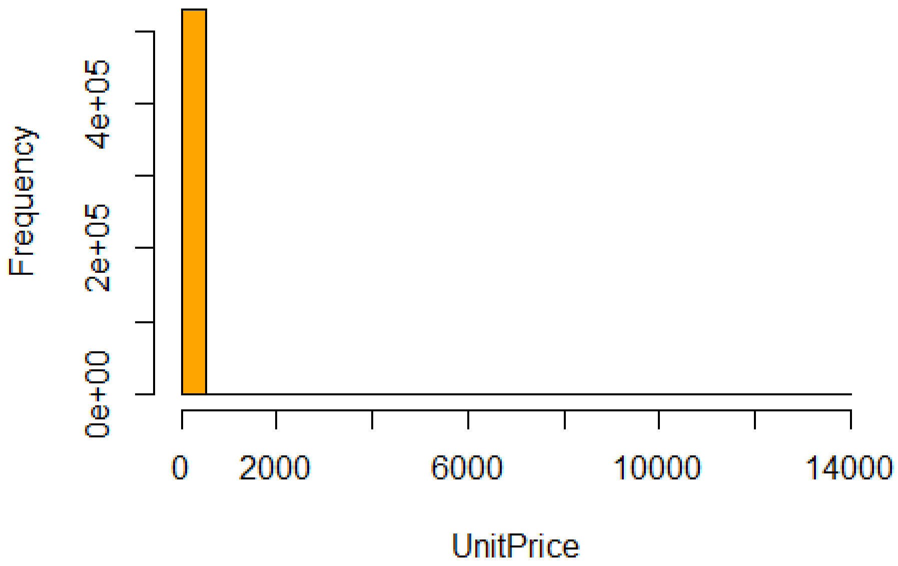

Figure 10. Histogram of Unit Price.

assigned Customer IDs) which means the highest revenue comes from non-members of the store. The NA is distantly followed by Customer ID 14646 and Customer ID 18102. The top 25 Customer IDs contribute 3.79 percent of the total revenue computed from the dataset.

The next graph (Figure 13) is a bar graph of the 15 topmost selling items based on revenue.The item which gives the maximum revenue is Dotcom postage which is closely followed by 3-tier regency cake stand and little birdie paper craft. The company sells to many wholesalers which could be the reason why Dotcom Postage is the topmost selling item.

This is followed by a similar graph (Figure 14). The difference between Figure 13 and Figure 14 is that Figure 14 shows 15 of the least selling items based on revenue. The item that generates the least revenue is 'Hen House with Chick in nest'. This item is followed by set '12 colouring pencils Doley' and 'Vintage blue tinsel reel'. It is important to know the least selling items as it helps retailers evaluate their assortment thereby aiding future stocking decisions. All the graphs are sorted in descending order except the last one which is in ascending order for easy viewing.

A point to note is that the returns or cancellations were removed from the data as

Table 4

Range,Mean and Variance of Revenue, Quantity and Unit Price.

Variable	Min. of Range	Max. of Range	Mean	Variance
Revenue	0.06	168469.60	20.12	73093.32
Quantity	1	80995	10.54	24187.93
Unit Price	0.04	13541.33	3.91	1289.95

they are negative values which means the transactions did not occur. The Invoice number starting with the alphabet 'C' indicates a canceled transaction. When I sum over the negative revenue for the canceled invoices it is -896,812.49 pounds sterling. This indicates about 900,000 pounds sterling of lost sales or expenditure related to the lost sales. These need to be investigated by the company in order to save money. It is known that returns cause the highest margin drain for online retail.

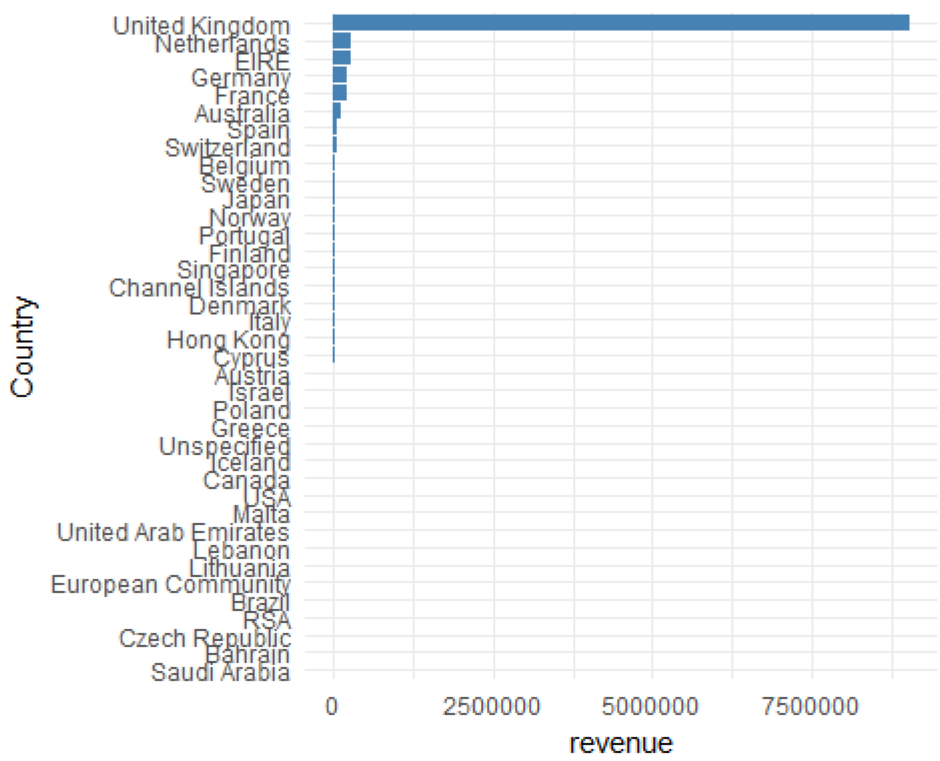

Figure 11. Bar plot of Country versus Revenue.

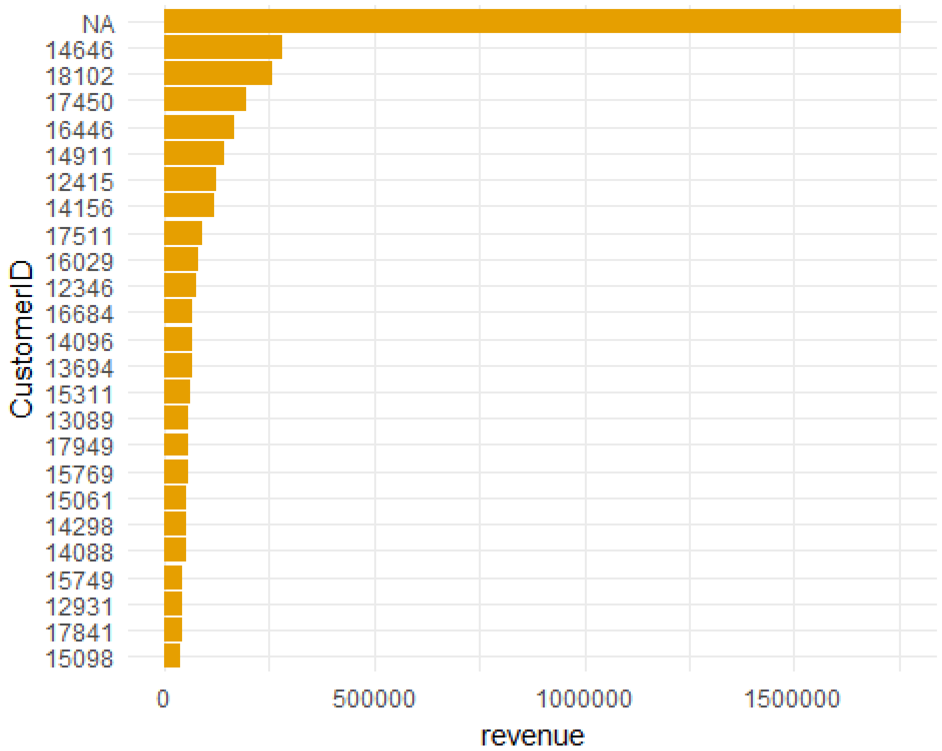

Figure 12. Bar plot of Top 25 Customer IDs based on Revenue.

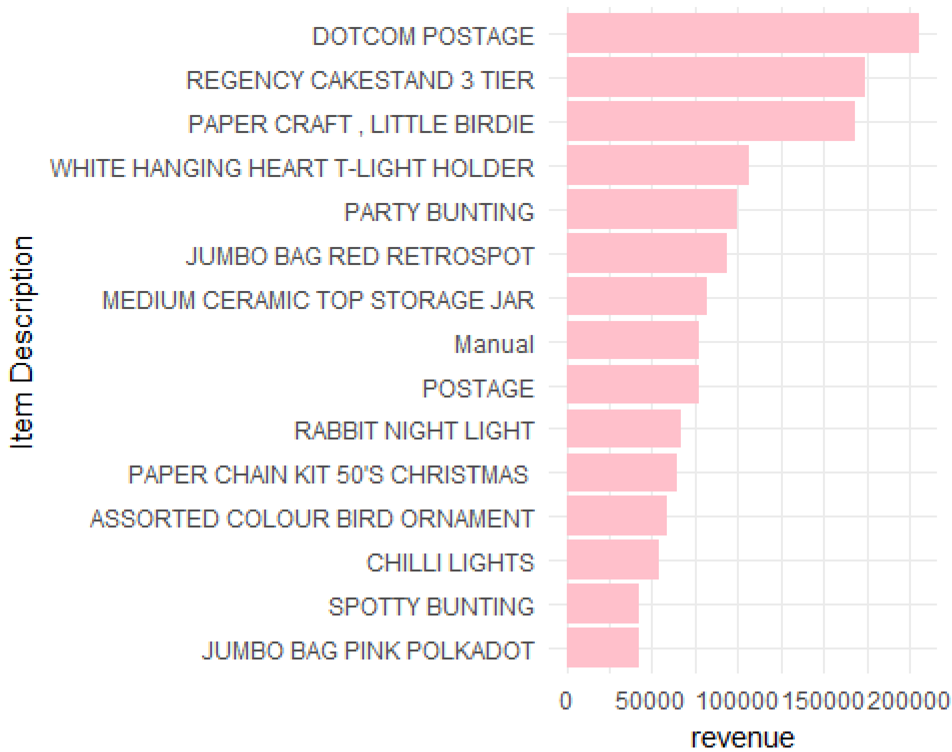

Figure 13. Bar plot of 15 Top Selling items based on Revenue.

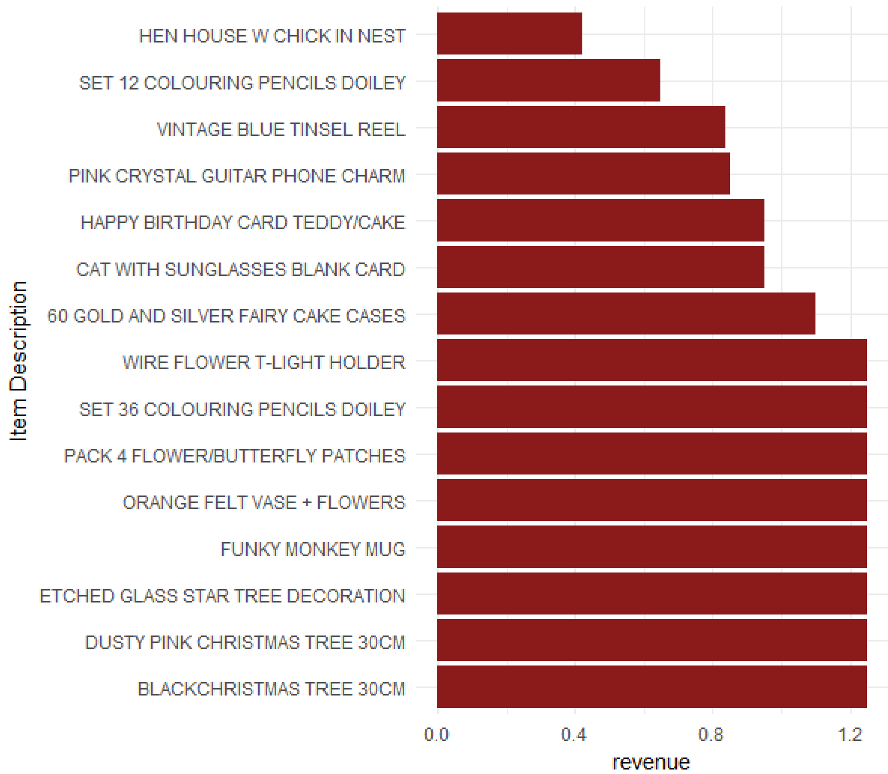

Figure 14. Bar plot of 15 of the Lowest Selling items based on Revenue.

Regression Results

The regression output contains the coefficients and the adjusted r-squared. The regression output for one of the items is as given in Table 5.

Table 5

Regression Analysis Summary for Item Stockcode 22716 - Predicting units sold using unit price

Variable	Coefficient	Std. Error	t-value	$Pr(>\|t\|)$
(Intercept)	-0.42	0.06	-7.44	1.19e-12
log(UnitPrice)	-3.42	0.07	-45.92	< 2e-16

Note. Adjusted R^2=.88 (N=289, $p < .001$). Residual standard error: .3807 on 287 degrees of freedom, F-statistic=2108 on 1 and 287 DF.

The regression output for Stock code 22716 as shown in Table 5 was pretty strong with an Adjusted R-squared of 0.88 and significant coefficients. When the regression model of Stock code 22716 is plotted, the four graphs that check the assumptions of regression are obtained. These plots are as shown in Figure 15. From the Figure 15, it is seen that Stock code 22716 does satisfy the constraints of regression - enough to use the model.

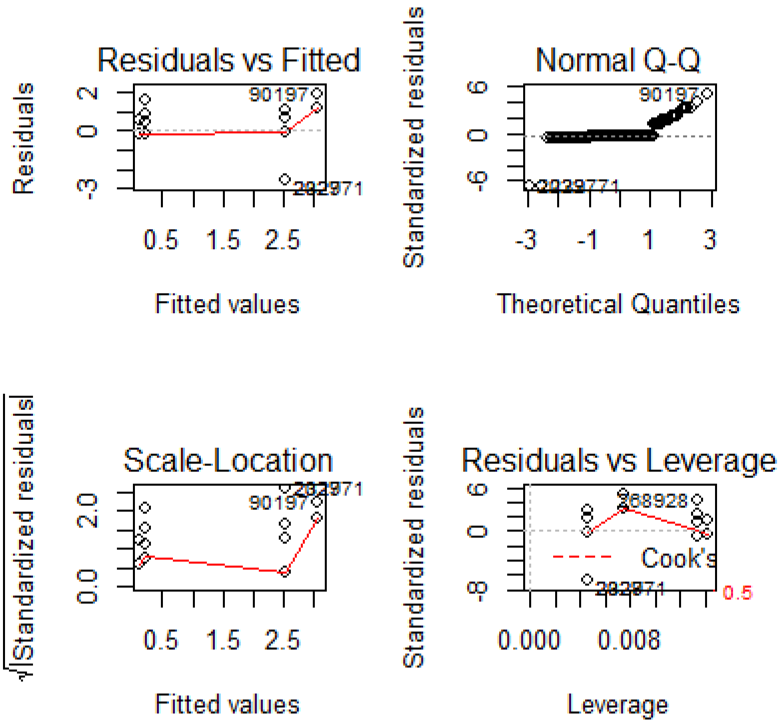

Figure 15. Plotting the Regression model of Stock code 22716.

Since there were about 3800 items, the output of each item was fed into a data frame and which was then written into a comma separated values (CSV) file. A sample of the CSV file is shown in Table 6.

The Adjusted R-squared was not high enough for some items. The reasons could be that there were not enough data points or data quality. I filtered out the items with Adjusted R-squared less than 70 percent before building the demand and revenue curves. The demand and revenue curve for one of the items is shown in Figure 16.

The optimal price at which the revenue was maximum was calculated from the curve. As explained previously, the outputs of all items were fed into a data frame which was then written into a CSV file. This CSV file has three columns - Stock code,

Table 6

A sample of the Regression output which contains all items.

intercept	Coefficient.of.price	Stock Code	Adj.r.squared
-8.24	-13.25	23691	0.81
-9.84	-15.10	21497	0.81
9.02	-2.20	23064	0.81
2.03	-0.85	20835	0.81
19.09	-16.26	47351B	0.80

Note. Output shown for selected items only. Size of dataset = 150 rows

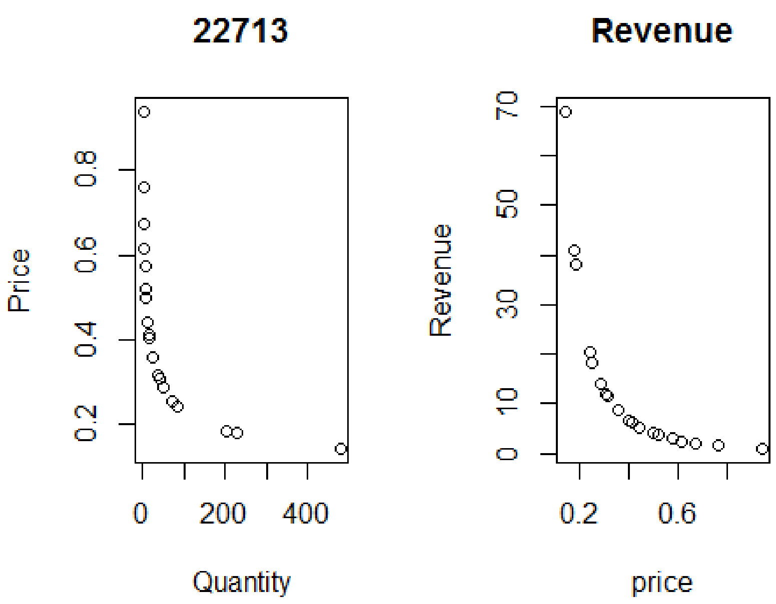

Figure 16. Demand and Revenue Curve for Item Stock code 22714 - using the data from the curve, the maximum revenue = 76 pounds and optimal price = 0.15 pounds.

Revenue at the Optimal price and the Optimal Price. The histogram of the Revenue at Optimal price is shown in Figure 17.

The histogram in Figure 17 gives us an idea of the distribution of the revenue of the items priced. It is seen that the histogram is skewed to the left. Most of the items priced have revenues between 0 and 250 pounds. The frequency of items with revenue between 250 pounds and 450 pounds is less. There are few outliers with revenue greater

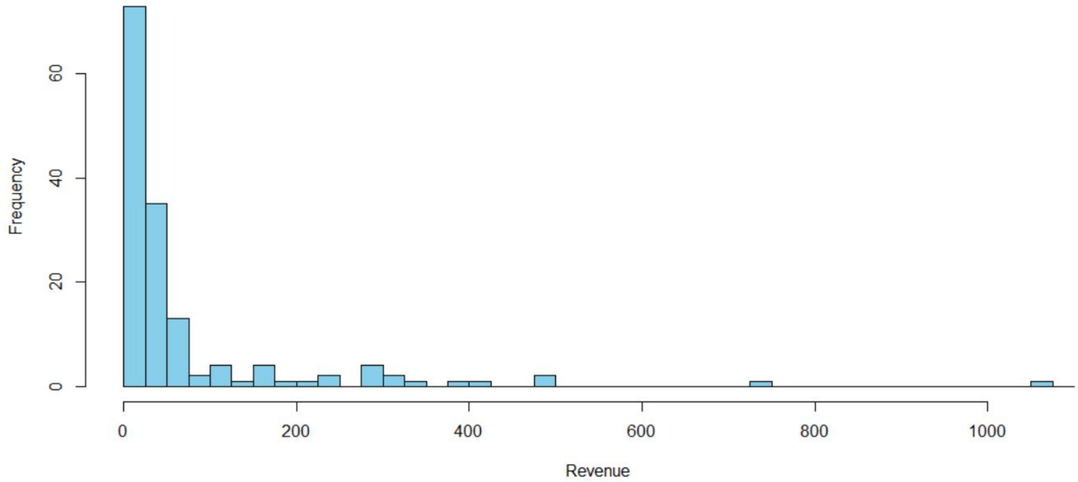

Figure 17. Histogram of Revenue at Optimal Price.

than 450 pounds with the maximum revenue being 1063 pounds.

A similar type of histogram is plotted for the Optimal Price in Figure 18.From Figure 18, it is seen that the majority of the optimal prices are between 0 and 5 pounds. There are few optimal prices between 5 and 10 pounds. There are some outliers with extremely low frequency values with the maximum price being 145 pounds.

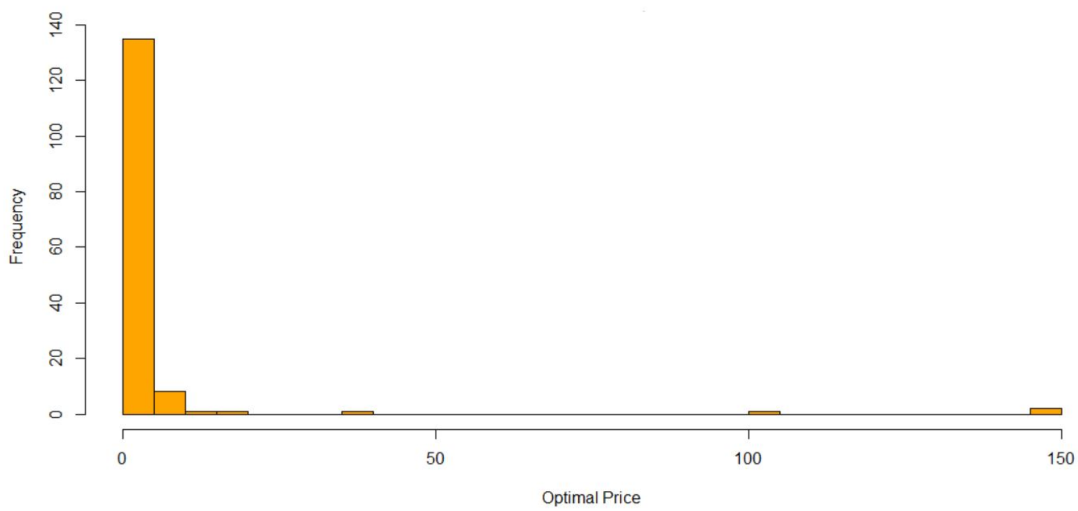

Figure 18. Histogram of Optimal Price.

Clustering Results

From Figure 7, we saw that the within group sum of squares does not decrease much after 8 clusters. Therefore, the K-means function was used to fit 8 clusters to the data. These 8 clusters can be visualized using the plot function as shown in Figure 19. On visualizing the clusters, it is seen that having eight clusters can be too many as data points are mainly present in three areas of the figure - top left, bottom left and bottom right.

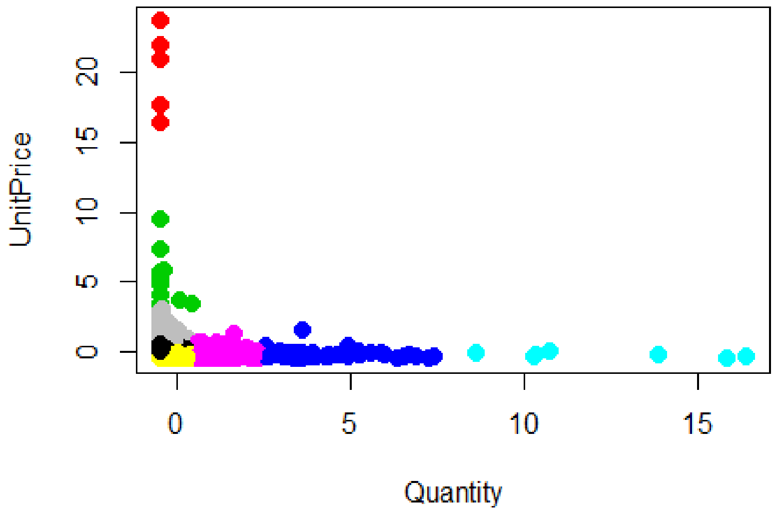

Figure 19. K-means Result with 8 clusters.

I then tried various values for the number of clusters such as 3,4,5 and 6. I decided to take 3 clusters due to the following reasons:

(1) The sharpest drop in the within group sum of squares shown in the scree plot (Figure 7) occurs when going to 3 clusters.

(2) The data should not be spread thin between too many clusters.

(3) Eyeballing the plot also lends itself to 3 clusters.

The three clusters will also be better for running cluster level regression which will be seen subsequently.

The same plot is then plotted after running K-means algorithm for 3 clusters as shown in Figure 20. The Figure 20 appears to be more intuitive than Figure 19 as explained previously.

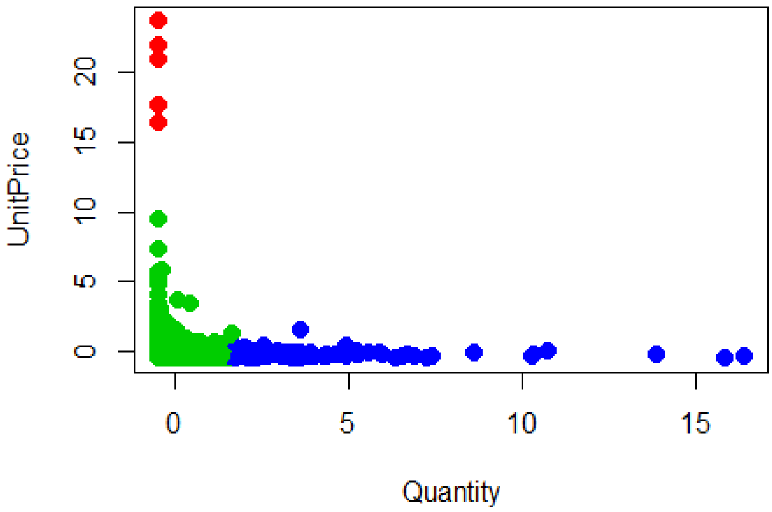

Figure 20. K-means Result with 3 clusters.

The important properties of a cluster are the size and center. These values are summarized in Table 7. Please note that the center values are scaled values as the data is scaled before clustering.

As explained before, the input dataset for K-means was prepared by (1) removing outliers followed by (2) scaling the data. In order to know the true mean values of Quantity and Unit Price of each cluster, the cluster assignments were aggregated with the original K-means input dataset which has the outliers removed but is not scaled. The Revenue is calculated by multiplying the Quantity with the Unit Price. The output of the mean function is summarized in Table 8.

The cluster assignments are combined with the original K-means input dataset which has the outliers removed but is not scaled. This dataset is used as input to Regression. The Regression is run as a 'for' loop. The number of iterations is equal to the number of clusters. This is achieved by sub-setting the dataset based on the cluster number during each iteration. The regression output at cluster level for the 3 clusters is given in Table 9.

The elasticity values that are so important to know to understand price sensitivity is given by the coefficient of price. It is -2.94,-0.65 and -0.03 for clusters 1,2 and 3 respectively. It is clear from Table 9 that the Adjusted R Squared for each of the 3

Table 7

Properties of the 3 K-means clusters - Size, Centers.

Cluster No.	Size	Center of Quantity	Center of Price
Cluster1	5	-0.48	20.11
Cluster2	3,085	-0.16	-0.02
Cluster3	150	3.41	-0.26

Table 8

Mean of Quantity, Unit Price and Revenue for each cluster.

Cluster No.	Quantity	Unit Price	Revenue
Cluster1	35.6	130.32	4,639.33
Cluster2	1,106.23	3.13	3,464.93
Cluster3	13,143.69	1.59	20,976.78

Table 9

Regression output of the 3 clusters.

Cluster No.	Intercept	Coefficient of Price	Adjusted R Squared
Cluster1	17.66	-2.94	0.14
Cluster2	6.48	-0.65	0.15
Cluster3	9.37	-0.03	-0.004

clusters is very low. The clustering was done to logically group items as some items did not have enough data points. The grouping was achieved by K-means but the regression is not strong enough to explain the variation in units.

The quantity and revenue curves for the 3 clusters is drawn using the curve function as shown in Figures 21, 22 and 23. The curves are drawn by giving values for the price in the range of 0 to 100. Revenue is calculated by multiplying revenue with unit price. The optimize function is used to find the optimal price (range given as 1 to 100) and the maximum revenue. The procedure of using curves and optimize function was derived from Vries and Meys (2012).

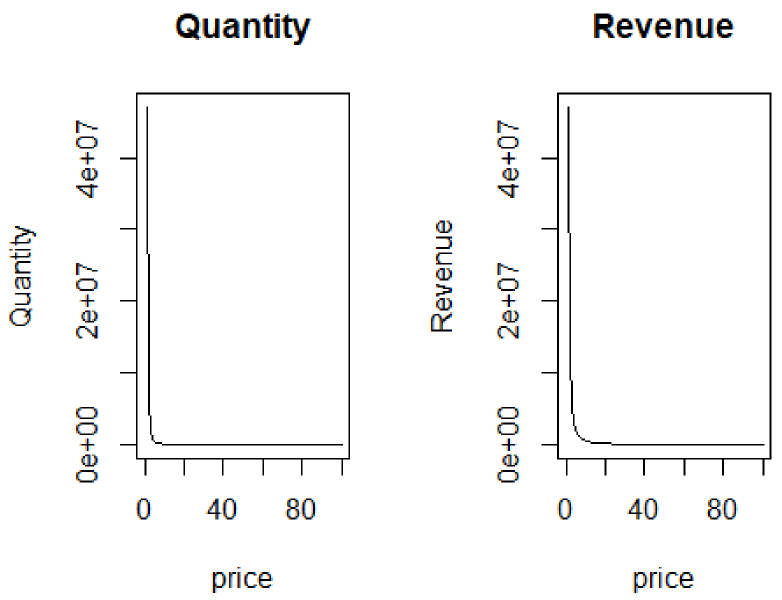

Figure 21. Quantity and Revenue Curve for Cluster 1. Optimal Price = 4.63e-05, Maximum Revenue = 1.2e+16.

Figure 22. Quantity and Revenue Curve for Cluster 2. Optimal Price = 99.99, Maximum Revenue = 3315.53.

Figure 23. Quantity and Revenue Curve for Cluster 3. Optimal Price = 99.99, Maximum Revenue = 1,050,376.

Since the Adjusted R squared is less, the regression equations may not be reliable. Hence the optimum price and revenue obtained from the optimize function may not be reliable too. However, the plots obtained do give us an idea about the shape of the curve for each cluster.

Time Series Results

Building the Time Series Object and Checking Assumptions. As explained in the implementation, I processed the data to get an aggregated and sorted dataset. This dataset is converted to a time series object with frequency equal to 7. The frequency of 7 was taken as I chose the cycle of the time series as a week. Since the data was at day level, the number of observations in a cycle will be 7. There was only about a year of data and hence a cycle value of a year cannot be taken. A lower cycle level of day or hour may become too low for the data. In retail, the seasonality and trends generally vary week to week and hence a frequency of 7 is suitable.

A time series model cannot be built unless the time series is stationary. A time series is said to be stationary if (1) The mean of the series should not be a function of time rather should be a constant; (2) The variance of the series should not a be a function of time (homoscedasticity); (3) The covariance is not constant with time (Srivastava, 2015). To check whether our time series is stationary I use the Augmented Dickey-Fuller test. The output of the test is shown in Table 10.

Table 10

Augmented Dickey-Fuller Test.

Dickey-Fuller	Lag order	p-value
-14.24	0	0.01

Note. Alternative hypothesis: stationary. The p- value is < 0.05 and hence we can reject null hypothesis - not stationary.

As explained in Table 10, it is proved that the time series is stationary.

Plotting graphs of the Time Series Object. I plotted the time series object as shown in Figure 24 where we see cyclic spikes in data around a mean along with some random variations. This time series object is explored further in the subsequent graphs.

In Figure 25, the box plot gives us an idea of the seasonal effect that plays out over the week. The median in Figure 25 rises, then remains flat and then falls towards the end of the cycle. The median becomes zero at the 7th value as those values are all Saturdays when the store was closed. The box plot shows that there were few outliers for values 2,3,4 and 6 but otherwise the series has quite a consistent pattern.

In Figure 26, the values over a cycle were aggregated using the mean function to show a week over week trend for the entire time duration of data available. This helps us understand the cyclic trend of the series. A rising trend towards the end is seen in Figure 26. This is consistent with what is seen in retail - the sales increase as December nears due to the festive season.

The Figure 27 shows the basis on which time series forecasting works. The time series has 3 defined components - the trend component, the season component and the random component. The Figure 27 clearly shows the rising trend towards the end, the repeating seasonal pattern and the random spikes that occur over time.

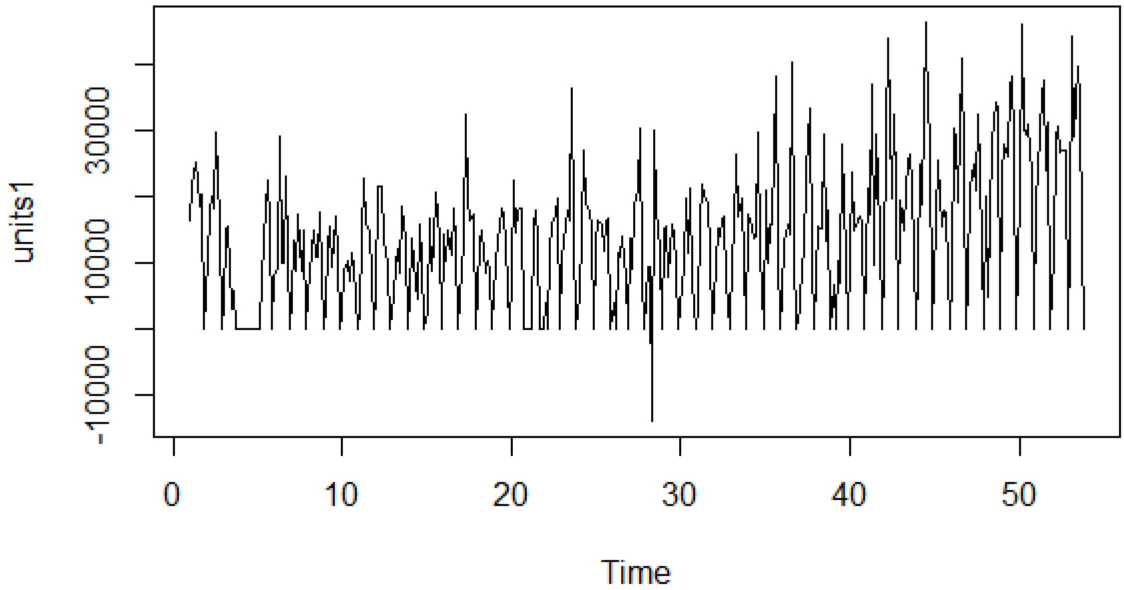

Figure 24. The Plot of the Time Series Object.

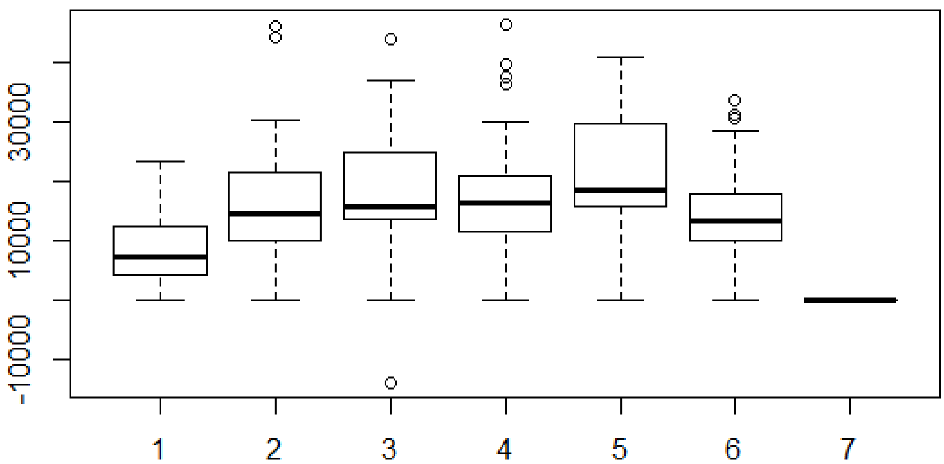

Figure 25. Box plot across days gives us an idea of the seasonal effect.

Holt Winters Model. The Holt Winters model was fit to the time series object. A sample of the Holt Winters fitted values is shown in Table 11.

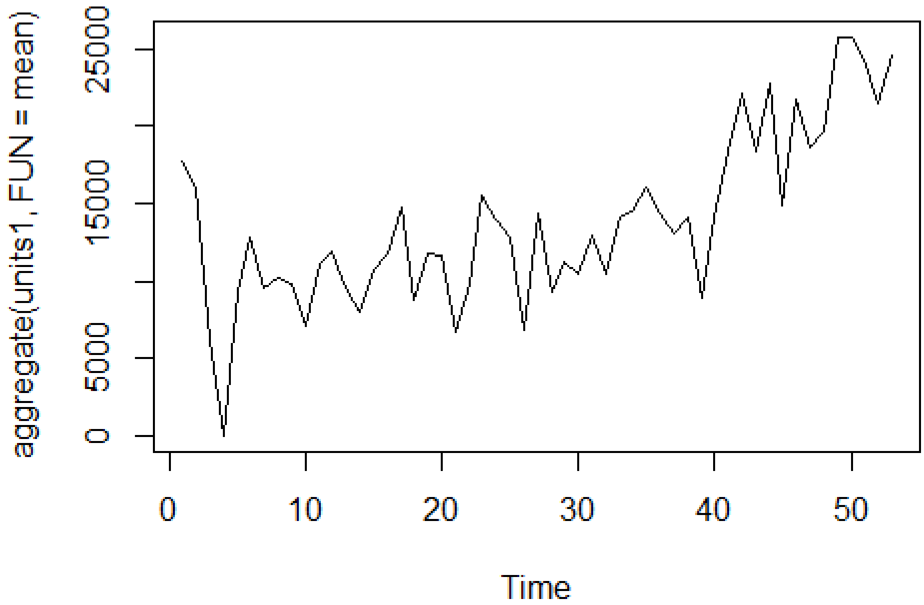

Figure 26. Week over Week trend - values over a cycle were aggregated using the mean function.

Table 11

A sample of Holt Winters Fitted Data.

xhat	level	trend	season
13,388.87	17,186.11	-195.12	-3,602.12
18,254.63	16,598.06	-199.17	1,855.74
21,045.66	16,311.00	-200.08	4,934.74
20,237.79	15,979.06	-201.44	4,460.16
17,671.24	15,498.11	-204.32	2,377.45
21,556.64	16,958.06	-187.15	4,785.74

The alpha parameter specifies how to smooth the level component, the beta parameter specifies how to smooth the trend component and the gamma parameter specifies how to smooth the seasonal component. The Holt Winters model fitted for this data uses an alpha of 0.14, beta of 0.01 and gamma of 0.13. I plotted the fitted model

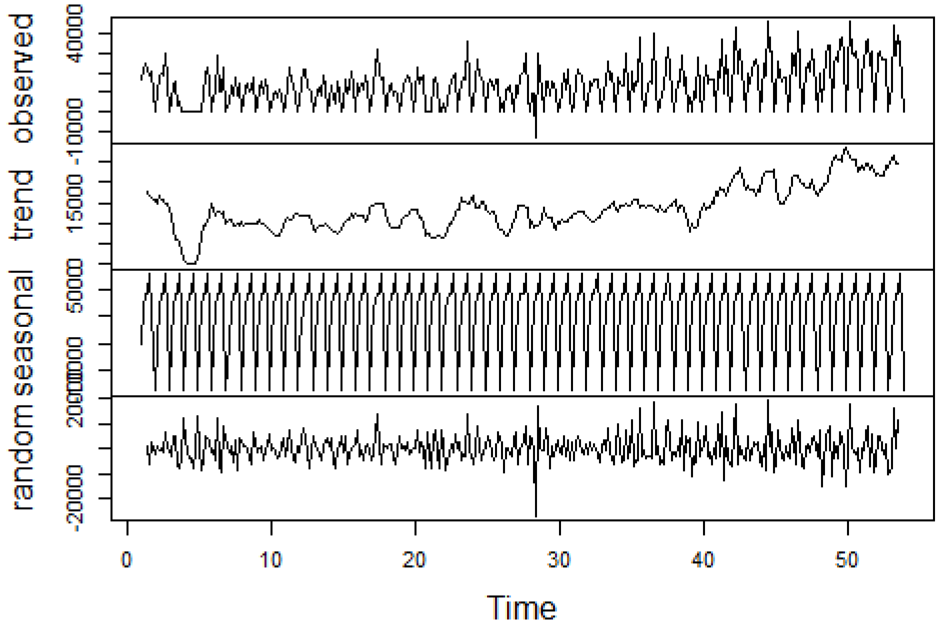

Figure 27. Decomposing the observed into the three defining components of a Time Series.

values and the actual time series data in Figure 28.

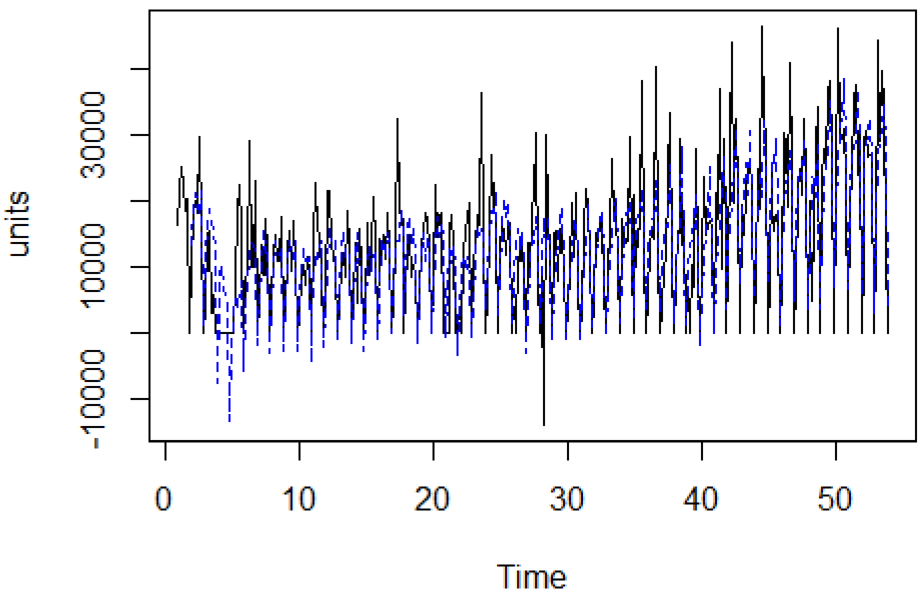

Figure 28. Plotting the Fitted versus the Actual Values.

The root mean square error (RMSE) is 6822.73. The maximum value in the time series data is 46,161. The percentage of RMSE to the maximum value is about 14%. So, relatively the RMSE is not a really large value.

The model built was then used to predict future time periods. The time series data length is 371. The time length for the prediction should be roughly about one third of the input time series data length. Hence I took 126 as the prediction time length which is approximately one third of 371 and is also divisible by 7. Figure 29 shows the predictions of the Holt Winters model created. The actual data is shown in sea green color, the predicted data is shown in red and the upper and lower confidence intervals are shown in blue.

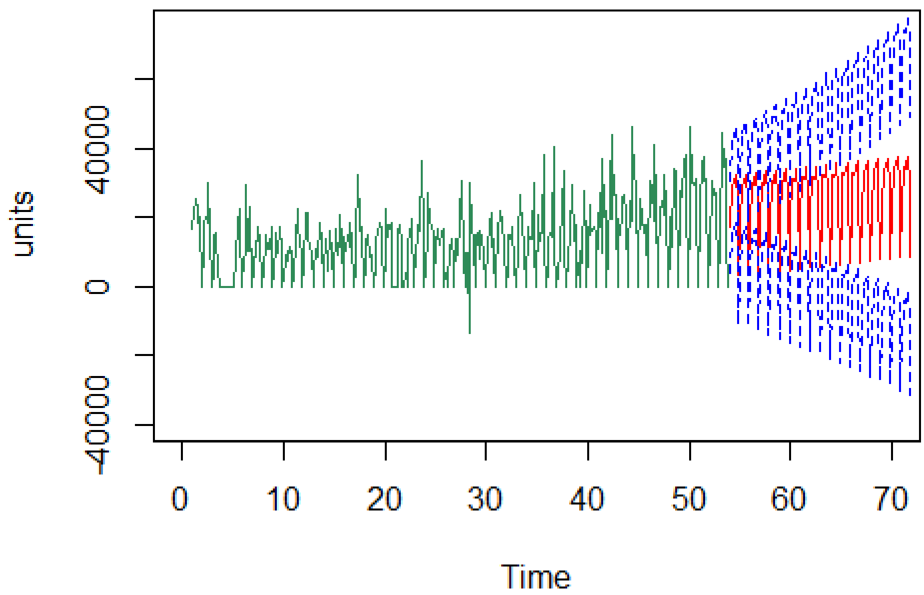

Figure 29. Plot the predictions of the Holt Winters model.

For time series, the assumption is that the data patterns will repeat in the future. This means an observation in the time series cannot be dependent on another observation in the time series. For this, I plotted the Auto Correlation Function (ACF). The correlation of the time series is the correlation for the time series observations with previous values in the same series. These are called lags. Figure 30 shows the ACF for the data. It is seen some of the lags below zero do touch the significance line slightly. It needs to be checked if this is significant enough to call the time series as an auto-correlated time series.

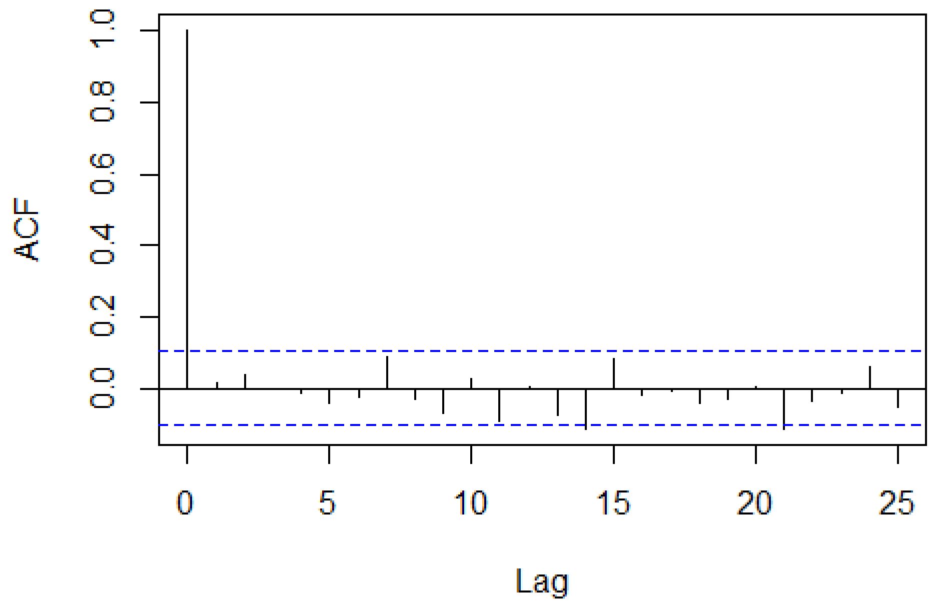

Figure 30. The Auto Correlation Function.

To test the significance, I performed the Box test on the residuals of the time series model. The Box–Pierce or Ljung–Box test statistic is used for examining the null hypothesis of independence in a given time series. The results of the Box test are shown in Table 12.

Table 12

Box-Ljung Test.

X-squared	df	p-value
0.092	1	0.76

Note. Null hypothesis: the time series is independent. The p- value is > 0.05 and hence we cannot reject the null hypothesis.

I also plotted the residuals of the time series model to check if they follow a normal distribution. This is shown in Figure 31. As seen in Figure 31, the data values are spread out about a mean. Hence it can be said that the residuals of the model do follow the normal distribution.

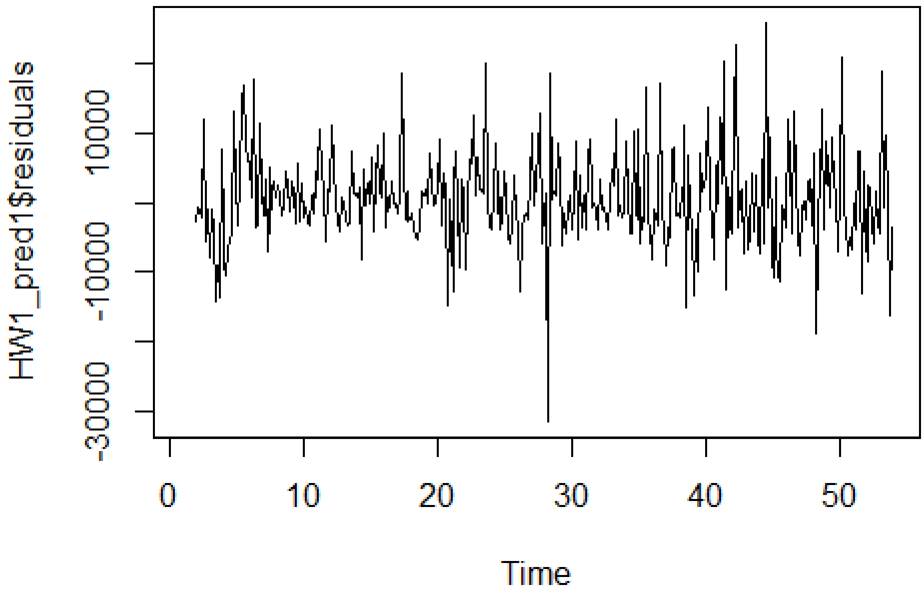

Figure 31. Residuals of the Time Series Model.

The time series data satisfies all assumptions needed to build the model. The model residuals satisfy the independence assumptions of a time series. The model also manages to predict the data pattern well except for the spikes observed which can be thought of as coming from the random component.

Conclusion

The research conducted goes through techniques and models that are a must-have in a retailer's toolkit for price optimization and demand forecasting. It is the price that I am looking to optimize but it is always taken as the independent variable. This is because changing the price is under a retailer's control. The goal is finding a price that maximizes demand. In line with that, the demand curves are built to understand the price sensitivity or elasticity. This is achieved by a simple log-log linear regression model that helps to predict the optimum price at which the revenue is the maximum.

An important issue that needs to be addressed with this model is the data sparsity issue. It is not possible to build demand curves for items with few points. I tried to address this problem by clustering the items using K-means and then applying the regression. However, the results of the regression were not satisfactory. Hence, future work should aim at addressing this issue. Other types of regression such as hierarchical regression can be tried in order to address this issue.

An important goal of this research is to forecast demand or the units sold. To achieve this the time series model was built using Holt Winters. The model effectively predicted the future demand for all of the products of the retail store in aggregate. This can help the store in pricing and product planning. In order to further tune the model and reduce the RMSE, a time series that uses neural networks can be tried. This is an interesting combination of techniques but it would definitely need more training data to achieve good results.

An important point to note is that these models need to be constantly updated with latest data to get suitable results. Elasticity is actually a random variable that keeps changing at various points in time. Hence, it needs to be constantly updated in order to be useful. This drives the need to build simple, non-cumbersome models that are easy to use. The models I have built have these virtues. But one is always walking the tightrope of balancing between performance and simplicity.

As explained in the Behavioral economics section of the literature review, it is important to test out all the three models in the real world either in the form of A/B experiments or by comparing the model with the actual results. It is also important to consult Domain experts to augment the data given to the model. It is the combination of the three - Analytics, Real world tests and Domain Expert Consultation; that leads to successful results. Hence, the future work would entail lot of experiments and interactions with various retail industry veterans in order to fine tune the model and improve its results.

References

Dean, J. (1980). Estimating the Price Elasticity of Demand. E. Mansfield (4th Ed.),Managerial Economics and Operations Research. New York: Norton.

DeBruicker, F., Quelch J., & Ward S. (1986). Cases in Consumer Behavior (2d ed.). Englewood Cliffs, NJ: Prentice-Hall.

Webber, R. J (2013). The evolution of direct, data and digital marketing. Journal of Direct, Data and Digital Marketing Practice, 14, 291-309

Moore, H. (1922). Elasticity of Demand and Flexibility of Prices. Journal of the American Statistical Association, 18(137), 8-19. doi:10.2307/2277462

Singh, A. and Grace, R. (2015). Two-Tiered Clustering Classification Experiments For Market Segmentation of EFTPOS Retailers. Australasian Journal of Information Systems,19, 117-132

Gahirwal, M., Vijayalakshmi M. (2009). Inter Time Series Sales Forecasting. Retrieved from http://arxiv.org/abs/1303.0117

Holt, C. C. (2004). Forecasting trends and seasonals by exponentially weighted moving averages. ONR Memorandum, 52, 2004.

You, Zhen et al. A Decision-Making Framework for Precision Marketing. Expert Systems with Applications 42.7 (2015): 3357-3367.

Webber, Richard. "The Evolution of Direct, Data and Digital Marketing". Journal of Direct, Data and Digital Marketing Practice 14.4 (2013): 291-309.

Oum, T. H. (1989). Alternative Demand Models and their Estimates. Journal of Transport Economics and Policy, 23(2), 163–187.

Estimating demand functions. (2017). Retrieved 11 May 2017, from http://www.econ.jku.at/members/WinterEbmer/files/Teaching/managerial/ws14-15/estimating.pdf

Johnson, W. (2017). K Means Clustering in R Example – Learn by Marketing. Learnbymarketing.com. Retrieved 11 May 2017, from http://www.learnbymarketing.com/tutorials/k-means-clustering-in-r-example/

Pedace, R. (2013). Econometrics and the Log-Log Model - dummies. dummies. Retrieved 11 May 2017, from http://www.dummies.com/education/economics/econometrics/econometrics-and-the-log-log-model/

Vries, A., & Meys, J. (2012). How to Find Minimum or Maximum Values for Functions in R - dummies. dummies. Retrieved 11 May 2017, from http://www.dummies.com/programming/r/how-to-find-minimum-or-maximum-values-for-functions-in-r/

Natarajan, I. (2017). Pricing – A Key Driver of Demand. Harrisburg: Unpublished Manuscript.

Chen, D., Sain, S. L., & Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. Journal of Database Marketing & Customer Strategy Management, 19(3), 197–208. https://doi.org/10.1057/dbm.2012.17

Kavyashri, Jayaram, M. N., & Jeevitesh. (2016). Dynamic Pricing in Ecommerce using Neural Network approach. International Journal of Research, 03(10), 1272–1278.

Penpece, D., & Elma, O. E. (2014). Predicting Sales Revenue by Using Artificial Neural Network in Grocery Retailing Industry : A Case Study in Turkey. International Journal of Trade, Economics and Finance, 5(5). https://doi.org/10.7763/IJTEF.2014.V5.411

Greifenberger, P. (2018). Behavioral Economics Powered by Machine Learning and AI Techniques: A True Path to Optimized Pricing. Retrieved June 10, 2018, from https://www.linkedin.com/pulse/behavioral-economics-powered-machine-learning-ai-true-greifenberger

Lindstrom, M. (2012). The Psychology Behind The Sweet Spots Of Pricing. Retrieved June 10, 2018, from https://www.fastcompany.com/1826172/psychology-behind-sweet-spots-pricing

Srivastava, T. (2015). A Complete Tutorial on Time Series Modeling in R. Retrieved June 14, 2018, from https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/