Pricing Optimization in Consumer Retail

Indhujha Natarajan

Student ID #187342

Harrisburg University

## 

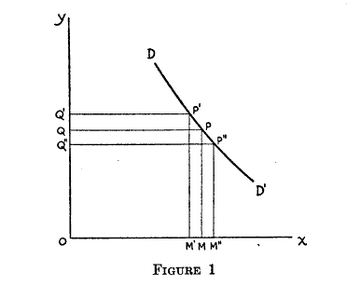
# Abstract

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 I have always been interested in the Pricing field and this paper builds on my previous unpublished work.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 01/12/2010 and 09/12/2011 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

# Introduction

As mentioned in Natarajan (2017), the effective quantitative theory of demand was given by [Antoine Augustin Cournot](https://en.wikipedia.org/wiki/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.* Demand curve. Adapted 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

#### 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. It is very common to have such pricing tests in consumer retail. These types of tests are popularly called A/B tests. Nowadays each retailer has a loyalty program where the consumer can earn points and use these points later to buy items in lieu of cash. These points offered are now being preferred to coupons to incentivize the retailer’s regular shoppers. One example of a test would be to test two alternatives: 1) points offered with slightly increased promotion prices on select apparel products and 2) low promotion prices with no points offered on the same products. In the two weeks tested, the low promotions gave a 42% increase in sales versus the 37% 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):

1. Linear demand model
2. Log Linear demand model
3. Logit model
4. Translog demand system

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

1. **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.

1. **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.

1. **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.

1. **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 obtained from economic demand theory. These model functions are used in my daily work at a retail company.



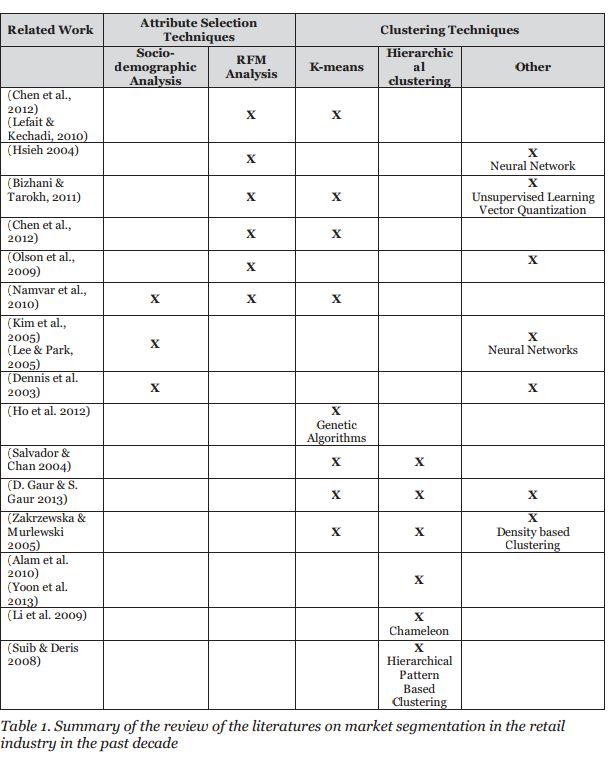
It can be inferred from the equations above that log/log model is easy to implement and interpret as the elasticity is the coefficient as opposed to the other two where the elasticity is a function of the coefficient and price. The other two models are similar. The difference is the second one gives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

### Machine Learning methods and models

1. **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 below:

Table 1



*Note.* 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.

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 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).

1. **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 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.

1. **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).

As explained above, we see 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.

# 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) the specific questions or areas I am interested in are

* 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 (∆Q/∆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

image0.jpg,

where Q =quantity demanded, alpha=shifting parameter, P=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

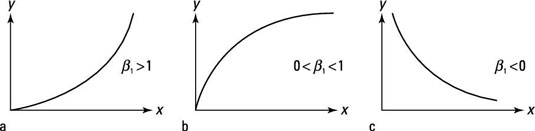
ln(Q) = ln(α) + βln(P)

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

δQ/Q= β\*δP/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 image9.jpgis the elasticity, x is the quantity and y is the price, we get the following curves for various values ofimage9.jpg.



*Figure 2.* 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 complexity. 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.

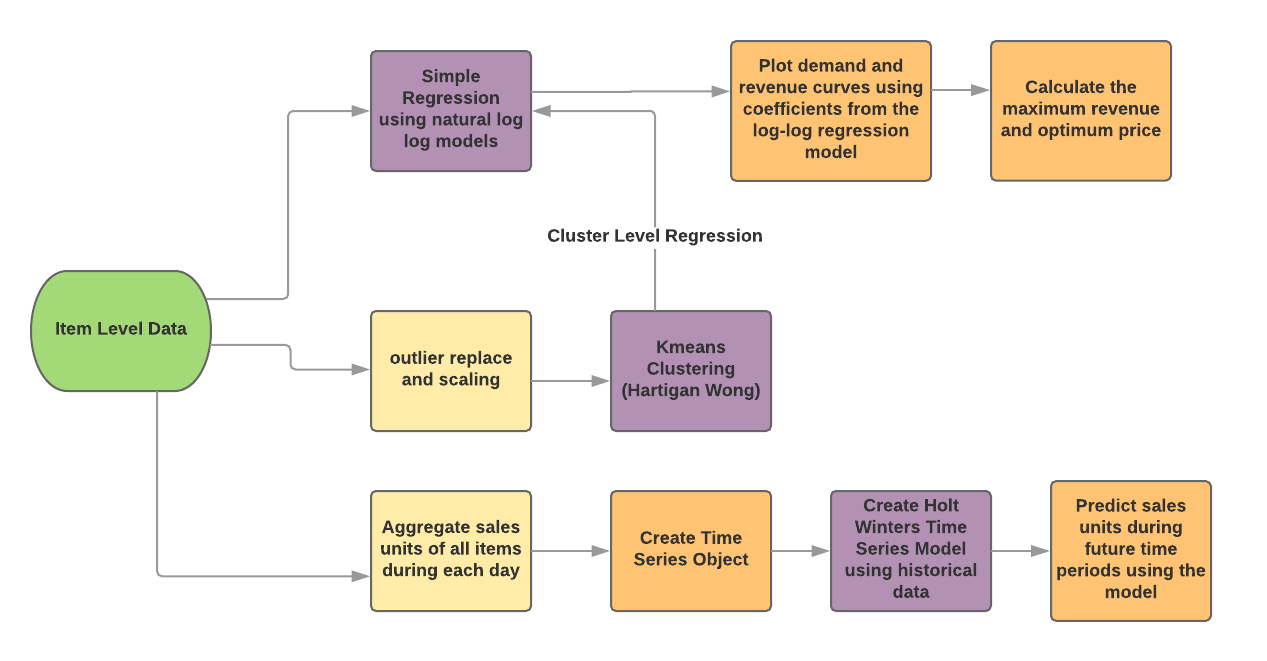
## Implementation

Before the implementation of the modeling process, data cleaning needs to be performed first.

Data cleaning involved the following steps.

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 cancelled 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, the process flow of modeling using the three techniques is represented using the diagram below.



*Figure 3.* Process Flow of Pricing Optimization.

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) ~log(UnitPrice))

The regression is run at item level using a ‘for’ loop where the model is applied on a particular item during each iteration of the ‘for’ loop. There are 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:** Before applying clustering, the cleaned item level data is further processed by replacing outliers with the median (using ‘outliers’ package). The data is 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. After clustering the clusters assignments are merged with the item level dataset. Then the regression model is applied at cluster level.

**Time Series:** The item level data was aggregated to sales units of all items sold on each day for the timeframe considered. A time series object was then created using this aggregated data. The Holt Winters model was built using this aggregated historical data. This model was then used to predict sales units during future time periods.

## Results

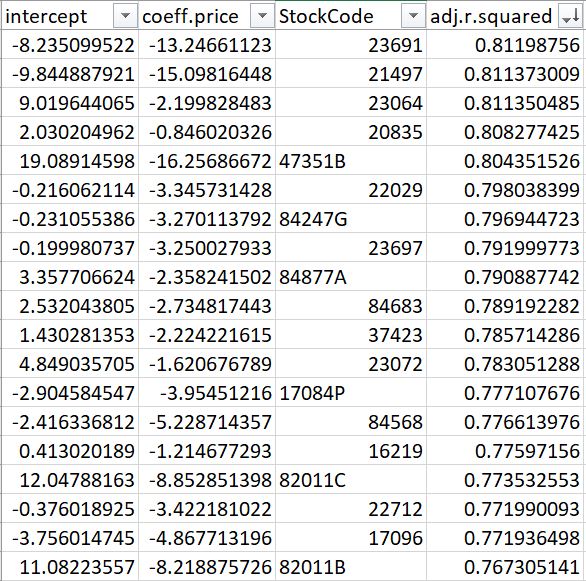
The regression output contains the coefficients and the adjusted r-squared. The regression output for one of the items is as given below.

Table made in overleaf.

The regression output for the above item (Stockcode: 22716) was pretty strong with an Adjusted R-squared of 0.88 and significant coefficients. Since there were about 3800 items, the output of each item was fed into a data frame and which was then written into a .csv file. A snapshot of the .csv file is shown in Table 3.

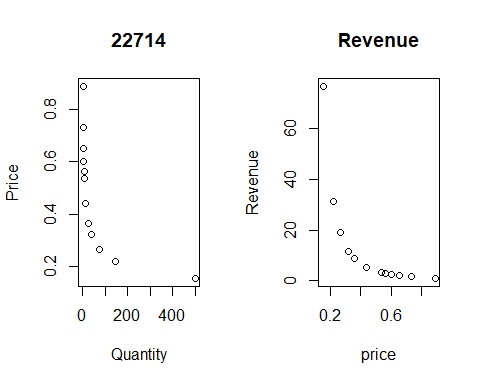
Table 3

Snapshot of the Regression output which contains all items.



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 5.

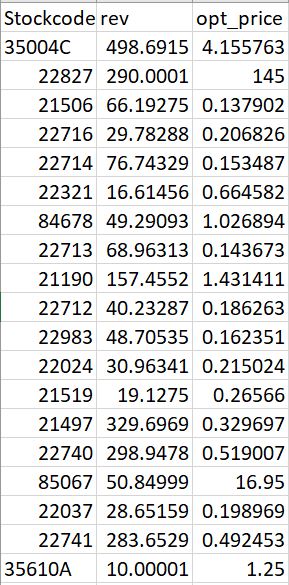


*Figure 4.* Demand and Revenue Curve for Item Stockcode 22714 - using the data from the curve, the maximum revenue = 76 pounds and optimal price = 0.15 pounds.

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. A snapshot of the .csv file is shown in Figure 6.

Table 4

Snapshot of the Maximum Revenue and Optimal Price output which contains all items.



# 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