**Project Report on**

**Inventory Optimization with Marketing Insights of a US based online retail store**

(A project on Marketing and Supply Chain Analytics with the development of an auto - inventory optimization tool)

**Submitted By:**

**TANVEER AHMAD HURRA** (**2K19/BMBA/19)**

**Under the Guidance of:**

**Mr. V. K Sharma**



**UNIVERSITY SCHOOL OF MANAGEMENT**

**& ENTREPRENEURSHIP**

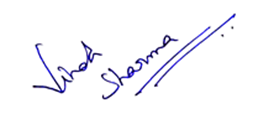
**Delhi Technological University**

**Bawana Road Delhi 110042**

**Jan -May 2021**

**CERTIFICATE**

This is to certify that Tanveer Ahmad Hurra (2K19/BMBA/19) is bona fide students of University School of Management and Entrepreneurship, Delhi and have successfully completed the project work as prescribed by the Delhi Technological University in the partial fulfillment of the requirement of Master Of Business Administration (MBA) , Business Analytics Program for the academic year 2018-2019

****

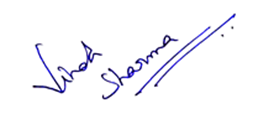
Project Guide

Mr. VK Sharma

**Evaluation:**

Tanveer has express the essential understanding of contents, meet all the expectations and deadlines well in time shown perseverance & commitment, I

Congratulate Tanveer for great work done by him & recommend “O” Outstanding grade (92 out of 100 marks)

****

**Declaration**

I the undersigned solemnly declare that the report of the project work on **Inventory Optimization with Marketing Insights of a US based online retail store** is based on my own work carried out during my study of MBA (Business Analytics) under the supervision of Mr. V. K. Sharma. I assert that the statements made, and conclusions drawn are an outcome of the project work. I further declare that to the best of my knowledge and belief that the project report does not contain any part of any work which has been submitted for the award of any other degree/diploma/certificate in this University or any other University.

**Tanveer Ahmad Hurra**

**2K19/BMBA/19**

**Acknowledgement**

I would like to earnestly acknowledge the sincere efforts and valuable time given by my Supervisor **Mr V. K. Sharma** and our H.O.D. **Dr. Amit Mookerjee**. Their valuable guidance and feedback have helped me in completing this project titled **Inventory Optimization with Marketing Insights of a US based online retail store.** I am further highly thankful to each and every faculty of USME who put their every effort to make me learn the fundamentals of Business Analytics, the concepts of which were highly utilized in completing this project. I further sincerely thank all my friends and batch mates in USME for their timely support and help

**Table of Content**

|  |  |  |  |
| --- | --- | --- | --- |
| **S. No.** | **Topic** | **Sub Topic** | **Page Number** |
| **1** | **Abstract** |  | **6** |
| **2** | **Introduction** |  | **6** |
| **3** | **CLV Calculation** |  | **7** |
| **4** |  | **Aggregate Method** | **7** |
| **5** |  | **Cohort Analysis** | **8** |
| **6** |  | **Predictive Method** | **8** |
| **7** |  | **Probabilistic Method** | **8** |
| **8** | **Inventory Optimization** |  | **9** |
| **9** |  | **Cycle Inventory** | **9** |
| **10** |  | **Cycle inventory calculation** | **10** |
| **11** | **Costs involved with inventory** |  | **10** |
| **12** |  | **Material Cost** | **10** |
| **13** |  | **Fixed Ordering Cost** | **10** |
| **14** |  | **Handling Cost** | **11** |
| **15** | **Bull Whip Effect** |  | **12** |
| **16** | **Economic Order Quantity** |  | **13** |
| **17** | **Solving the EOQ equation** |  | **15** |
| **18** | **Tools used in the project** |  | **17** |
| **19** |  | **MS Excel** | **17** |
| **20** |  | **R-Programming** | **17** |
| **21** |  | **Python Programming** | **18** |
| **22** | **CLV - Calculation** |  | **20** |
| **23** |  | **Data** | **20** |
| **24** | **CLV with different period of analysis** |  | **23** |
| **25** | **R-code** |  | **27** |
| **26** | **Other Marketing Insights** |  | **28** |
| **27** |  | **Pattern of churn rate over time** | **30** |
| **28** |  | **Effect of average discount on churn rate** | **32** |
| **29** |  | **Effect of discount on CLV** | **33** |
|  |  | **Effect of discount on customer growth** | **34** |
| **30** |  | **Effect of discount on customer acquisition** | **39** |
| **31** |  | **Effect of discount on customer retention** | **42** |
| **32** | **Inventory Optimization with Transactional data** |  | **43** |
| **33** |  | **Python Code** | **43** |
| **34** |  | **Backend Code** | **45** |
| **35** | **Conclusion** |  | **45** |
| **36** | **References** |  | **47** |

1. **Abstract:**

The current project attempts to gain insights about the customer lifetime value using the transactional data of a US based online store. Different period of analysis are used to calculate the churn and retention rates for CLV calculation. A separate approach of Cohort Analysis has also been adopted for CLV calculation and the effect of customer acquisition based cohorts on the same. Further, the effect of discounts given by the store on the customer acquisition and retention has also been included.

Besides, getting the marketing insights through the transactional data, the same data has been used to study the inventory of the firm and calculations are done to bring optimization to the same based on various inventory related costs like handling and ordering cost etc. A generalized tool for inventory optimization and calculation of Economic order quantity has also been developed using Python and Tkinter Package for GUI development.

1. **Introduction:**

The customer lifetime value represents the total profit or revenue that a firm expects to earn from a single customer throughout his/her lifetime. The customer lifetime value is an important figure because it helps a firm to decide what portion of total profit must be spent on the customer acquisition or retention policies. One important application is to decide the total advertisement expenditure or total discounts that a firm should consider to acquire new customers or retain the existing ones respectively.

For example, the CLV of a Netflix user might be a 20000 INR on average, depending on his/her frequency of renewal of subscription. Similarly for an average home owner in a metropolitan city like Delhi or Mumbai, the average lifetime value of his/her tenant may depend on the average number of tenants the owner is getting and also the renewal of rent agreement too.

In short, CLV is an indicator of the profit that you expect from a particular customer relationship. The value is used to decide what expenditures should be made to maintain that relationship. As an example if a customer is worth 2000 INR for a firm, the firm might not spent more than that to maintain the relationship with its customers.

1. **CLV Calculation**

There are different methods available to estimate the CLV for a firm. The most notable among these are:

* Aggregate Method.
* Cohort Analysis
* Predictive Method (Machine Learning)
* Probabilistic Method

**3.1 Aggregate Method:** One of the most simple and old method of calculating the customer lifetime value is the aggregate method. The method assumes that there is a constant churn rate and average order value and customer frequency throughout the time. Under this method, the customer lifetime value can be calculated using the following formula:

CLV = Average Order value x Customer Frequency x (1/ Churn rate) \* Profit Margin

The Average order value can be calculated by dividing the total sales amount by the total number of the orders in the period

Average Order Value = Total Sales/Total Orders

Similarly customer frequency can be obtained by dividing the total number of orders by the total number of unique customers in the period. Customer frequency means that on an average how many orders a single customer puts.

Customer Frequency = Total Orders/Total unique customers.

Churn rate refers to the metric indicating how many customers do not come back after their first purchase or simply:

Churn rate = (C0- C1)/C0

C0 = Customers at the start of the period

C1 =Customers at the end of the period

This method does not differentiate between customers and produces a single value for CLV at an overall level. This leads to unrealistic estimates if some of the customers transacted in high value and high volume, which ultimately skews the average CLV value.

If you recall the formula discussed above, except Profit Margin all the other variables can be estimated/calculated.

**3.2 Cohort Analysis:** Aggregate Method of CLV calculation assumes all the cusomers across all the periods form a single group. This has a major drawback because not all custmors share similar relationship with the business in relation to the number of orders and total amount spent.

This major drawback of aggregate method is overcome by an another method called cohort analysis, wherein the total customer base is divided into various cohorts or sub-groups with each sub group sharing a particular characteristic. Normally, customers are grouped on the basis of the month of acquisition but the other characteristics like scale of purchase etc. can also be considered. The best choice will depend on the customer acquisition rate, seasonality of the business, and whether additional customer information can be used.

**3.3 Predictive Model (Machine Learning):** This method leverages the machine learning models like Regression, Neural Networks etc. to predict the customer behaviour using the past purchasing history. Once the future trend of the purchase is known, the total customer lifetime value can be easily calculated.

**3.4 Probabilistic Method:** This method uses probability distributions to calculate the terms in the CLV equation. As an example a normal distribution can be assumed to represent the customer churn rate of average customer order frequency. The distributions are used to get the expected value of the different quantities in the CLV calculation equation.

There are various probability distributions that are used for the Customer Lifetime value calculation and there is not a single universal distribution that could serve the purpose. Normally for the transaction variables i.e. average customer frequency and churn rate, a different probability distribution is used than the monetary variables i.e. the average order value or profit margin. The type of probability distributions that are used are discussed below:

Beta Geometric and Negative Binomial Distribution: BG and NBD models are among the commonly used models for the customer lifetime value estimation. There is an another model called Pareto/NBD model which the current model serves as an alternative, Pareto model is the most used model for the CLV calculation.

Both BG/NBD and Pareto/NBD model tries to accurately predict the customer life time value. These models are then added with the Gamma distribution model to add monetary variables (average order value etc.) and what we finally get is the Customer Lifetime Value through probability distributions.

1. **Inventory Optimization**

So now, we have discussed the customer lifetime value, it’s time to present the theory of inventory optimization. Inventory optimization refers to the techniques used to minimize the total cost involved in inventory upkeep. The various costs involved here are handling cost and ordering cost.

**4.1 Cycle Inventory:** When we speak of the lot in a supply chain we refer to the batch of material that the firm either purchases or produces at a given time. As an instance, a company that sells 5 products per day, the inventory manager orders 100 products each time the inventory needs to be refilled or replenished. The lot size here is 100 and keeping in mind that the firm sells 5 products per day, it means that takes on an average 100/5 = 20 days before the inventory manager takes action and places the order for another lot. The store of the firm holds inventory because the mangers order the lot size which is 20 times greater than the daily demand of the firm. Cycle inventory is the average inventory that the firm holds throughout its inventory cycle.

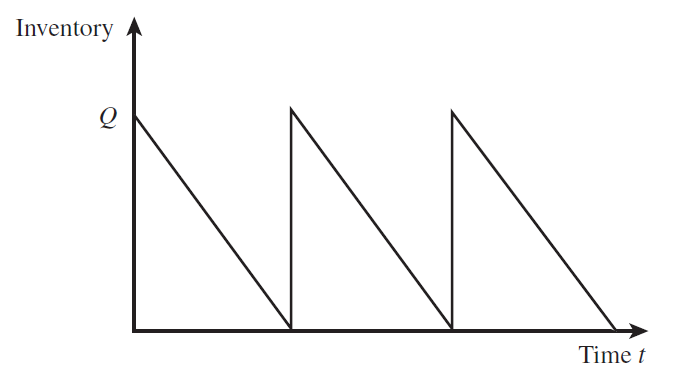


Figure : Inventory over time

The above image shows the inventory cycles and the replenishment periods that are considered for placing a new order.

**4.2 Cycle Inventory Calculation & relation with demand:** Suppose at the start of an inventory cycle, the quantity in the inventory is “Q” units. At the end of the cycle before the next order is placed, the quantity reduces to zero.

Hence average inventory held = (Q + 0)/2 = Q/2

Hence if the lot size is Q, the cycle inventory is Q/2.

Now, if we assume that the demand faced by the firm is steady at “D” units each day. The average time of storage of the inventory items would be:

Average time of storage = Average Inventory/Demand

= Q/2D

1. **Costs Involved with inventory**

**5.1 Material Cost:** The average cost involved with a particular lot is the key decision that an inventory manager has to make. If increasing the lot size reduces the average cost, the inventory manager may increase the lot size and vice-versa. As an instance if an electronics manufacturer charges 30 USD per gadget for ordering up to 100 gadgets and charges 25 USD if more than 100 gadgets are ordered at a time, the inventory manager here will take the decision of ordering more than 100 items as it reduces the cost by 5USD/item. The price that is paid for a particular item is called material cost and is one of the factors involved in the total cost of the inventory.

**5.2 Fixed Ordering Cost:** The fixed ordering cost does not depend on the size of the lot but is fixed for each order. The fixed ordering cost may involve fixed administrative costs, loading and shipping costs etc. As an instance, a company ordering a lot of 200 items with the fixed ordering cost of 1000 USD incur the transportations cost of 1000/200 = 5 USD per item, whereas a lot size of 1000 items results in a transportation cost of 1 USD/item. Given the fixed transportation cost per batch, the store manager can reduce transportation cost per unit by increasing the lot size. The fixed ordering cost per lot or batch is denoted by S (commonly

thought of as a setup cost) and is measured in dollars per lot. The ordering cost also displays

economies of scale—increasing the lot size decreases the fixed ordering cost per unit purchased.

**5.3 Holding Cost:** Holding cost is the cost incurred by a company by holding one unit of inventory for a particular period of time, usually one year. It is the combination of various costs put together like maintenance charges, costs related to the items getting obsolete, opportunity costs etc.

The holding cost is usually denoted by “H” and is measured as USD/year. Holding cost is normally calculated by considering a fraction of material cost contributed to the holding cost.

If the fraction of material cost contribution to holding cost is “h”

And the total material cost per unit is “C”

Then the holding cost per unit will be, H = hC

The total holding cost increases with an increase in lot size and cycle inventory.

The main objective of defining a cycle inventory is to make different levels of supply chain management to order the quantities and lot sizes in the order and magnitude that minimises the total cost of inventory. The total cost here refers to the sum of the material cost, ordering cost which is fixed as stated earlier and the holding cost which may be expressed as the fraction of the material cost only. As an instance, if an inventory manager considers only the material cost then the main tendency of the manager would be to order a lot size as larger as fits the budget to take the benefit of the economies of scale or in other words get benefited from the discounts that come with the larger lot sizes.

Now, on the other hand if a manager only considers the fixed ordering cost, his main tendency again would be to order a lot size as larger as fits the budget to take the benefit of the economies of scale involved here too. Since ordering cost is fixed, a larger lot size would result in the lower ordering cost per item.

Now, if the manager only takes into account the holding cost, his tendency would be to decrease the lot size to reduce the total inventory cost. As we know that the inventory handling cost can be expressed as the fraction of total material cost, it means smaller the lot size or smaller the total items in the inventory, lower would be the inventory cost.

As is evident from the above discussion that there are some factors in the total inventory cost which tends to increase with the increasing lot size and there are other factors which tend to decrease with the increasing lot size, the inventory manager must find the trade-off to calculate the optimum lot size that minimises the total cost.

1. **Bull Whip Effect**

Ideally the decisions of cycle inventory should be made keeping in view the whole supply chain but normally this is rarely being done in the organizations. Usually each level of supply chain take their cycle inventory decisions independently due to which the synergy within the supply chain gets missing an as a result the overall effect on the supply is the inflated demand figures which gets magnified when we go from downstream to upstream.

As an example a local warehouse calculating its total inventory cost may get a figure of lot size that is independent of any other stage up the supply chain. The ware house may add a little margin or buffer to its calculations to compensate any uncertainty that lies with the cost and cycle inventory calculations. The same figures when conveyed up the supply chain will prompt the inventory manager at manufacturing level to calculate their cycle inventory and because they have to add their own margin for uncertainty, the overall cycle inventory will get inflated by all the added margins included at each stage. This effect is known as bull whip effect and must be taken care of by incorporating synergy within the supply chain.

One of the hurdles in calculating the cycle inventory is to estimate the fixed ordering cost and the holding cost of the inventory. It becomes obligatory on an inventory manager not to spend too much resources and time on such estimations but to guesstimate them initially from experience until the calculations are refined further.

Often the main goal of an inventory manager is to calculate the incremental cost change with the incremental change in the lot size. All those parameters whose incremental change does not bring the change in the total inventory cost must be ignored in the calculation. As an instance, the labour cost can be independent on the lot size and thus does not change or contribute to the total inventory cost or cycle inventory. In this case, the labour cost can be eliminated from the cycle inventory calculations. Same goes for the other costs too, that doesn’t change with the lot size.

1. **Economic Order Quantity**

Economic Order Quantity refers to the lot size that a company orders to minimise its total cost. In order to calculate the same , we need to express all the three costs in terms of lot size and then plug these three different costs into the total cost equation and then optimise the equation for the lot size.

The three costs involved are:

1. Material Cost
2. Fixed Ordering Cost
3. Handling Cost

If we assume that the material cost is “C” USD per unit and fixed ordering cost is “S” USD per unit and the handling cost “H” USD per unit. Assuming further that the demand is “D” units per year, then the total cost will be:

Total Inventory Cost = D\*C +(D/Q) S +(Q/2)\* hC -1

Here “h” is the fraction of material cost contributed to handling cost of the inventory.

The above equation upon further refinement can be easily optimised to get the economic order quantity.

The assumptions that are considered for the above calculations are:

1. The annual demand “D” is consistent and do not change over time
2. There are no shortages allowed or in other words, each order is supplied from the stock
3. The lead time for replenishment is fixed and is assumed zero at the start.

The equation 1 can be either solved through the techniques of calculas or can be solved graphically or after building a data table to check costs corresponding to different lot sizes. A close inspection of the equation reveals that the only variable present in it is “Q”, that is the lot size. All the other parameters are obtained as discussed below:

Demand, D: The demand can be forecasted by utilizing the past sales data. A trend in the past data can give us an estimate about the future demand and same can be incorporated in equation-1 to estimate the Economic Order Quantity.

Holding Cost: There is no way to directly calculate the holding cost as it involves a lot sub-costs like monthly rent, labour cost, administrative and maintenance costs etc. Initially this cost is estimated by the experienced professionals and managers and then plugged into equation -1 to estimate the economic order quantity. With due course of time, a further refinement is made into this estimation and the Economic Order Quantity is further calculated with greater accuracy.

Material Cost: This cost is usually known. It refers to the cost the firm purchases each item and is very straight forward in nature. This cost, hence do not involve any complex calculations or estimations.

Fixed Ordering Cost: This cost too is relatively easy to calculate. Although it involves various sub-costs too like shipping & transportation costs, loading costs, etc. but they are comparatively easy to incorporate & ordering cost/item is usually known.

The equation-1 can be graphically represented as below:

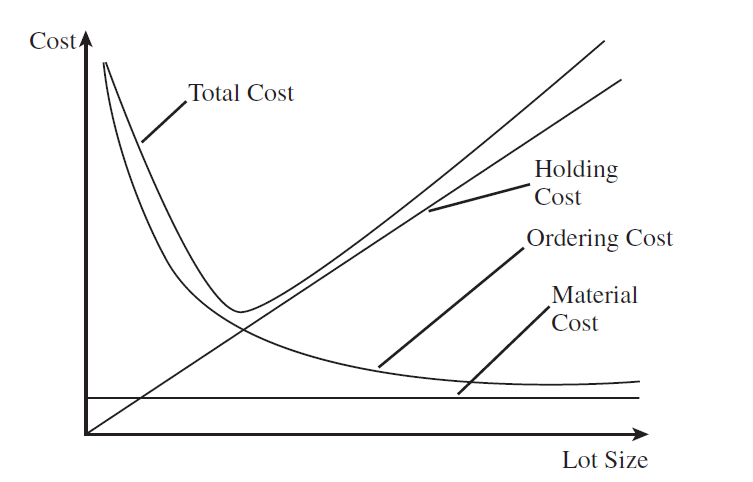


Figure : Relation of various costs with lot size

As you can clearly see, the Holding Cost increases with the increasing lot size, while as ordering cost/item decreases with the increasing lot size. The material cost is practically independent and the total inventory cost see a trough or a minima with the increasing lot size. This is that minima that we are interested in and is usually calculated for finding the Economic Order Quantity.

1. **Solving the EOQ equation**

Data Table: One of the easiest way to find the solution or optimum value of the equation of Economic Order Quantity is generate a data table as shown below:

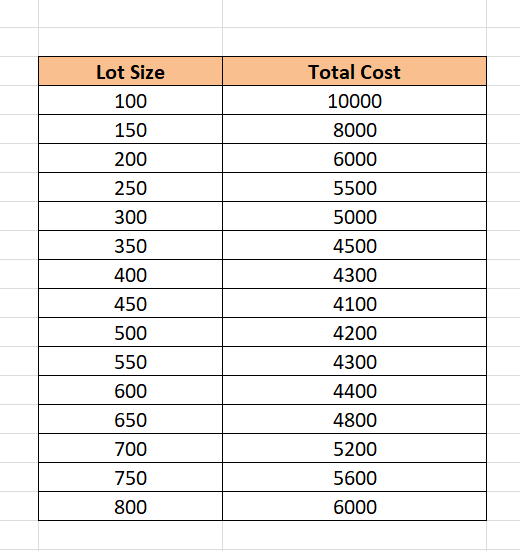


Figure : Data Table of Lot size VS Cost

As is evident from the above data table, the minimum cost is 4100 and corresponds to the lot size of 450. Hence, the economic order quantity in this case would be 450.

Graphical Method: Another very easy method of estimating the economic order quantity is through graphs. We plot total cost against the lot size & the point where we get the trough is our solution.

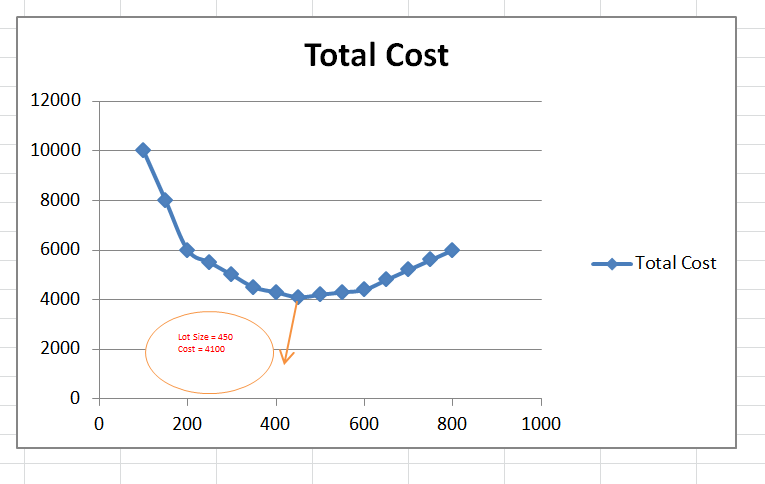
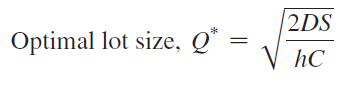


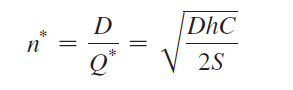
Figure : Total Inventory Cost VS lot size

The above graph clearly shows a trough or a minima at point (450, 4100). Hence the total minimum cost is 4100 corresponding to the lot size of 450.

Method of Calculus: This method involves differentiating the equation-1 with respect to Q and then equating the resulting equation to zero. After we differentiate the equation-1 with respect to the Q (lot size), we get the economic order quantity to be:



Since, the total annual demand is D and now we know the lot size that we are ordering to minimize our cost, we can easily find the total number of orders that we are going to put annually:



As an example, if the demand of a certain computer is 1000 units per month and suppose the fixed ordering cost that the company incur is 4000 USD per order. If the cost of each computer is 500 USD and the holding cost is 20% of the material cost, the economic order quantity for this case can be calculate as:

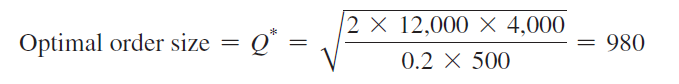
Annual Demand = 1000 \* 12 = 12000 units

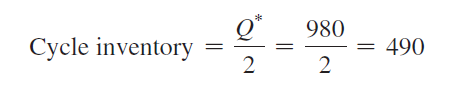
Ordering Cost per lot = 4000 USD

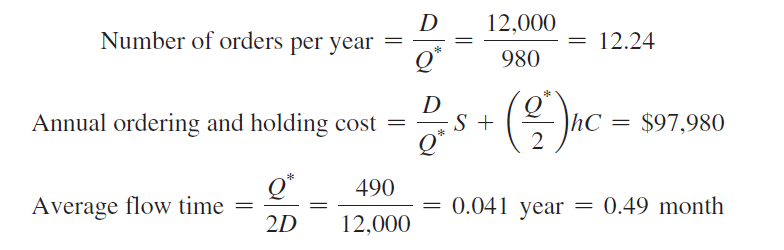
Unit cost of computer = 500 USD

Holding cost fraction = 0.2

Using equation for economic order quantity, we have:







1. **Tools used in the Project**

The main software tools that were used in the project are:

1. MS Excel
2. R – Programming
3. Python Programming

**9.1 MS-Excel:** The customer lifetime value and the plots to get the insights into the data were done in MS Excel. Also, the visual data inspection is done using this tool only. The final calculations of churn rate and average order value etc. are done in MS Excel. The Excel formulae are incorporated for final calculations and representations.

**9.2 R- Programming:** R programming was used to find monthly churn and to get unique customer visits along with the first time visitors from the data. The following R-libraries are used to get the job done:

Dplyr: As per CRAN website [1], dplyr helps with the following tasks:

* The package makes the data manipulation easy and convenient.
* The package provides simple and important verbs also called functions to carry out the task of data manipulation and wrangling.
* The package uses the computational backend in most efficient form and hence avoids wastage.

The important dplyr verbs that were used in this project are:

Group\_by: The function groups the data based upon some particular given column. As an instance if we want to group the total sales of a company based upon their stores, we need to group the data for each store using group\_by() function and then proceed with the calculations.

Filter(): The function is used to filter a particular data table based upon a given condition. As an instance to filter the marks data of students to only get the information about those who are from a particular section, the filter() function can be used.

Select(): The select() function is used to select few columns from the whole dataset.

**9.3 Python Programming:** The Python programming [2] is the general purpose programming language which can be used for varied number of purposes ranging from data analytics to GUI building. This project uses python programming for inventory optimization and the calculation of Economic Order Quantity. The plots for inventory optimization and total cost are plotted using Python. The main python packages that were used are as follows:

Numpy [3]: Numpy is a python package providing support for n-dimensional arrays and matrices along with the large set of mathematical functions that can be operated on these matrices. The python programming language is not originally designed for numerical computation; it is the Numpy package that facilitates it. Numpy offers same features to Python users as offered by MATLAB in terms of matrix operation and numerical computation along with the speed. The current project uses Numpy for the creation of Covariance Matrix of the stocks in the portfolio mainly.

Pandas [4]: Pandas is a python package written for the data analysis and manipulation. The main object in Pandas is a data frame which is analogous to an excel sheet and hence make data analysis much easy. Pandas can be used to import the external data (e.g. CSV) into the python environment as a data frame. It offers data structures and operations for manipulating numerical tables and time series in an effective way. The current project uses Pandas to pull data whether local or online into the Python environment. It is also used for return calculation from the share price data.

Tkinter [2]: Tkinter is a standard python package for GUI creation. It offers various toolkits and GUI components to create beautiful looking Graphical User Interfaces. Tkinter Entry, Label, Button, Check Box, and Radio Box etc. are some of the GUI components offered by the package. The package has a easy to implement GUI geometry to arrange the components in a window. The current project uses Tkinter for GUI creation to make the program more user-friendly

Scipy [5]: The Scipy package is used for scientific computation in Python. Scipy contains modules for optimization, linear Algebra, Calculus, ODE solving etc. The basic data structure used by Scipy is the n-dimensional Numpy array. Since it is written over Numpy, the implementation becomes easy for the projects already using Numpy. The current project is using Numpy mainly for optimization operation of the portfolio selected by the user.

MatPlotLib [6]: MatPlotLib is a plotting library for Python. This package is also built upon Numpy and hence is easy to implement in projects where Numpy is already deployed. The MatPlotLib syntax is much similar to that of MATLAB and offers various plotting components like scatter plot, line plot, histograms, and bar plots etc. that can prove very helpful. The current project uses MatPlotLib package to plot return time series and probability distribution of the expected return.

1. **Calculation – Customer Lifetime Value**

**10.1 Data:** The data pertains to a US based online retail store. The data set as shown below covers the transaction data recording the important information like Month of order, total sales etc. This is a 4 year data ranging from April-2017 to April-2021. The first few rows of the data are shown below:

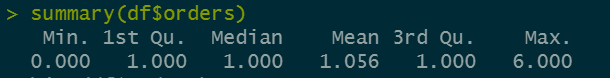


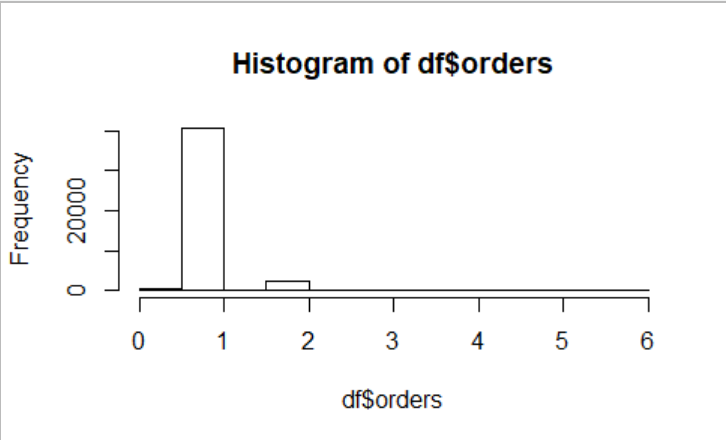
The main variables recorded in the data set are explained below:

Month: It is the month of sale and the value ranges from 2017-04 to 2021-04.

Customer\_id: This is the unique id which recognises a particular customer.

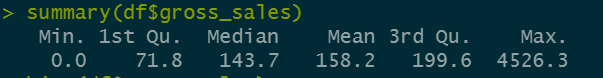
Orders: It shows the number of order that a particular customer places on a given day. The exploratory analysis of this column shows following results:

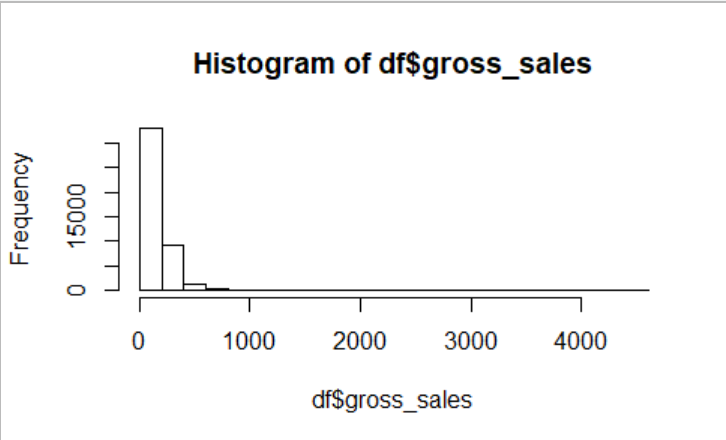




The above analysis shows that the single order is the most frequent one.

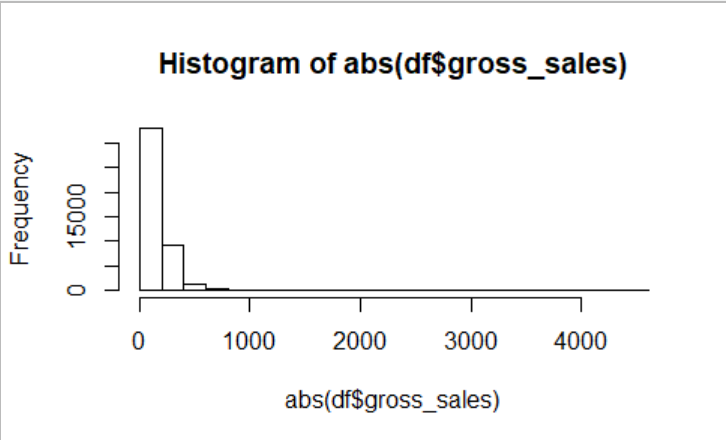
Gross\_sales: The variable captures the sales amount excluding discounts and taxes. The exploratory data analysis of the variable shows following results:



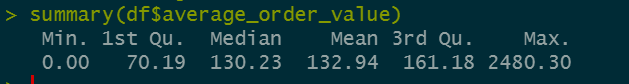


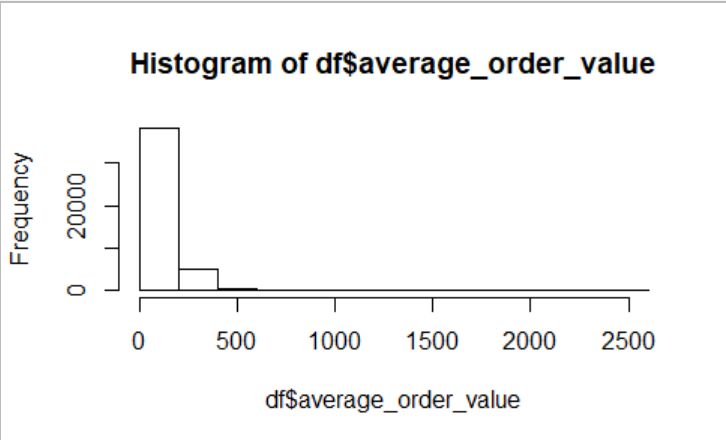
Discounts: This is the amount of discount given on a particular transaction to a customer on a particular day. The exploratory data analysis shows following results:





Average\_order\_value: This is the total sales amount by total orders by a particular customer on a given day. The exploratory data analysis shows following results:





1. **Customer Lifetime Value with different period of analysis**

The CLV can be calculated by taking into consideration different periods. Normally a month is taken for the calculation of churn rate, average order value etc. but that means that the customers who shop with the interval of more than a month are not counted and hence their contribution to the total customer lifetime value calculation is ignored. As such the current project has considered the following period of analysis for CLV calculation:

1. One Month
2. Three Months
3. Six Months
4. Nine Months
5. Twelve Moths

One Month: Here the average order value & customer frequency is calculated for each month and the churn rate is calculated as follows:

Churn Rate =

The CLV & other values for this period of analysis is as:

Table : CLV Calculation with period of analysis = 1 month

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Period = 1 month** | | | | | |  |
| **S.no** | **Month** | **No. of Unique Customers** | **Return Customers from prev month** | **Churn Rate** | **Average Customer Frequency** | **Average Order Value** |
| 1 | 2017-04 | 29 | NA |  | 0.97 | 130.80 |
| 2 | 2017-05 | 1191 | 2 | 93.10% | 1.09 | 115.57 |
| 3 | 2017-06 | 859 | 155 | 86.99% | 1.05 | 95.97 |
| 4 | 2017-07 | 1026 | 141 | 83.59% | 1.07 | 98.00 |
| 5 | 2017-08 | 917 | 181 | 82.36% | 1.06 | 122.88 |
| 6 | 2017-09 | 827 | 154 | 83.21% | 1.04 | 121.00 |
| 7 | 2017-10 | 961 | 165 | 80.05% | 1.06 | 130.80 |
| 8 | 2017-11 | 905 | 181 | 81.17% | 1.06 | 124.02 |
| 9 | 2017-12 | 799 | 145 | 83.98% | 1.04 | 122.08 |
| 10 | 2018-01 | 803 | 85 | 89.36% | 1.03 | 123.84 |
| 11 | 2018-02 | 798 | 105 | 86.92% | 1.04 | 121.74 |
| 12 | 2018-03 | 928 | 139 | 82.58% | 1.05 | 134.28 |
| 13 | 2018-04 | 1116 | 203 | 78.13% | 1.08 | 130.42 |
| 14 | 2018-05 | 1029 | 207 | 81.45% | 1.05 | 113.19 |
| 15 | 2018-06 | 1038 | 161 | 84.35% | 1.07 | 112.61 |
| 16 | 2018-07 | 1122 | 172 | 83.43% | 1.04 | 104.40 |
| 17 | 2018-08 | 1044 | 176 | 84.31% | 1.04 | 107.36 |
| 18 | 2018-09 | 988 | 161 | 84.58% | 1.05 | 116.94 |
| 19 | 2018-10 | 882 | 148 | 85.02% | 1.04 | 116.72 |
| 20 | 2018-11 | 1291 | 236 | 73.24% | 1.08 | 135.53 |
| 21 | 2018-12 | 729 | 183 | 85.82% | 1.03 | 128.14 |
| 22 | 2019-01 | 644 | 93 | 87.24% | 1.03 | 119.48 |
| 23 | 2019-02 | 713 | 99 | 84.63% | 1.05 | 131.83 |
| 24 | 2019-03 | 973 | 151 | 78.82% | 1.03 | 121.02 |
| 25 | 2019-04 | 1052 | 163 | 83.25% | 1.03 | 114.68 |
| 26 | 2019-05 | 878 | 152 | 85.55% | 1.05 | 131.50 |
| 27 | 2019-06 | 1071 | 187 | 78.70% | 1.06 | 140.92 |
| 28 | 2019-07 | 850 | 187 | 82.54% | 1.03 | 118.44 |
| 29 | 2019-08 | 896 | 176 | 79.29% | 1.04 | 127.96 |
| 30 | 2019-09 | 834 | 158 | 82.37% | 1.05 | 133.08 |
| 31 | 2019-10 | 725 | 145 | 82.61% | 1.03 | 135.43 |
| 32 | 2019-11 | 1112 | 217 | 70.07% | 1.10 | 150.37 |
| 33 | 2019-12 | 722 | 125 | 88.76% | 1.02 | 133.01 |
| 34 | 2020-01 | 862 | 137 | 81.02% | 1.09 | 131.92 |
| 35 | 2020-02 | 800 | 170 | 80.28% | 1.06 | 136.69 |
| 36 | 2020-03 | 958 | 199 | 75.13% | 1.09 | 138.92 |
| 37 | 2020-04 | 950 | 204 | 78.71% | 1.05 | 135.38 |
| 38 | 2020-05 | 994 | 189 | 80.11% | 1.06 | 124.24 |
| 39 | 2020-06 | 867 | 208 | 79.07% | 1.05 | 130.35 |
| 40 | 2020-07 | 806 | 157 | 81.89% | 1.05 | 133.80 |
| 41 | 2020-08 | 999 | 189 | 76.55% | 1.05 | 140.55 |
| 42 | 2020-09 | 893 | 197 | 80.28% | 1.04 | 127.57 |
| 43 | 2020-10 | 834 | 165 | 81.52% | 1.06 | 135.01 |
| 44 | 2020-11 | 1248 | 263 | 68.47% | 1.10 | 162.84 |
| 45 | 2020-12 | 623 | 140 | 88.78% | 1.04 | 134.73 |
| 46 | 2021-01 | 791 | 129 | 79.29% | 1.04 | 132.74 |
| 47 | 2021-02 | 715 | 146 | 81.54% | 1.05 | 146.10 |
| 48 | 2021-03 | 1008 | 187 | 73.85% | 1.11 | 134.78 |
| 49 | 2021-04 | 867 | 167 | 83.43% | 1.06 | 142.05 |
| **Average** | | | | **81.82%** | **1.05** | **127.59** |
| **Customer Lifetime Value** | | | | **164.13** | | |

Similarly the calculations for the period of analysis of 3 months is shown below:

Table : CLV Calculation with period of analysis = 3 month

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Period = 3 Months** | | | | | |
| **Period** | **No. of Unique Customers** | **Return Customers from previous period** | **Churn Rate** | **Average Customer Frequency** | **Average Order Value** |
| 2017-04 to 2017-06 | 1917 | NA |  | 1.164840897 | 107.84 |
| 2017-07 to 2017-09 | 2310 | 667 | 65.21% | 1.269264069 | 112.98 |
| 2017-10 to 2017-12 | 2194 | 791 | 65.76% | 1.279398359 | 125.91 |
| 2018-01 to 2018-03 | 2145 | 733 | 66.59% | 1.227972028 | 127.06 |
| 2018-04 to 2018-06 | 2645 | 829 | 61.35% | 1.281663516 | 119.13 |
| 2018-07 to 2018-09 | 2658 | 994 | 62.42% | 1.239277652 | 109.33 |
| 2018-10 to 2018-12 | 2388 | 970 | 63.51% | 1.281407035 | 128.06 |
| 2019-01 to 2019-03 | 1975 | 741 | 68.97% | 1.221772152 | 123.96 |
| 2019-04 to 2019-06 | 2475 | 790 | 60.00% | 1.269090909 | 129.10 |
| 2019-07 to 2019-09 | 2123 | 872 | 64.77% | 1.260009421 | 126.52 |
| 2019-10 to 2019-12 | 2117 | 921 | 56.62% | 1.279168635 | 141.50 |
| 2020-01 to 2020-03 | 2068 | 922 | 56.45% | 1.369922631 | 135.93 |
| 2020-04 to 2020-06 | 2244 | 913 | 55.85% | 1.321746881 | 129.87 |
| 2020-07 to 2020-09 | 2192 | 943 | 57.98% | 1.291970803 | 134.26 |
| 2020-10 to 2020-12 | 2203 | 983 | 55.16% | 1.321833863 | 148.13 |
| 2021-01 to 2021-03 | 2009 | 878 | 60.15% | 1.340965655 | 137.31 |
| **Average** | | | **61.38%** | **1.276269032** | **127.31** |
| Customer Lifetime Value | | | 264.69 | | |
| Increment from last CLV | | | 100.56 | | |

For a 6 month period, the calculations are as follows:

Table : CLV Calculation with period of analysis = 6 month

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Period = 6 Months** | | | | | |
| **Period** | **No. of Unique Customers** | **Return Customers from previous period** | **Churn Rate** | **Average Customer Frequency** | **Average Order Quantity** |
| 2017-04 to 2017-09 | 3560 | NA | NA | 1.45 | 110.76 |
| 2017-10 to 2018-03 | 3606 | 1296 | 63.60% | 1.51 | 126.47 |
| 2018-04 to 2018-09 | 4309 | 1469 | 59.26% | 1.55 | 114.30 |
| 2018-10 to 2019-03 | 3622 | 1568 | 63.61% | 1.51 | 126.25 |
| 2019-04 to 2019-09 | 3726 | 1476 | 59.25% | 1.56 | 127.91 |
| 2019-10 to 2020-03 | 3263 | 1533 | 58.86% | 1.70 | 138.65 |
| 2020-04 to 2020-09 | 3493 | 1532 | 53.05% | 1.66 | 132.02 |
| 2020-10 to 2021-03 | 3334 | 1536 | 56.03% | 1.68 | 142.93 |
| **Average** | | | **59.09%** | **1.58** | 127.41 |
| **Customer Lifetime Value** | | | **340.1900165** | | |
| Increment from last CLV | | | **75.50** | | |

For a 9 month period, the calculations are as follows:

Table : CLV Calculation with period of analysis = 9 months

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Period** | **No. of Unique Customers** | **Return Customers from previous period** | **Churn Rate** | **Average Customer Frequency** | **Average Order Quantity** |
| 2017-04 to 2017-12 | 4738 | NA | NA | 1.68 | 116.09 |
| 2018-01 to 2018-09 | 5406 | 1726 | 63.57% | 1.72 | 117.91 |
| 2018-10 to 2019-06 | 4946 | 1931 | 64.28% | 1.74 | 127.29 |
| 2019-07 to 2020-03 | 4264 | 1891 | 61.77% | 1.93 | 134.70 |
| 2020-04 to 2020-12 | 4444 | 1903 | 55.37% | 1.96 | 137.40 |
| Average | | | 61.25% | 1.84 | 129.33 |
| Customer Lifetime Value | | | 388.10 | | |
| Increment from last CLV | | | 47.91 | | |

For a 12 month period, the calculations are as follows:

Table : CLV Calculation with period of analysis = 12 months

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Period = 12 Months** | | | | | | |
| **Period** | | **No. of Unique Customers** | **Return Customers from previous period** | **Churn Rate** | **Average Customer Frequency** | **Average Order Quantity** |
|  | 2017-04 to 2018-03 | 5870 | NA | NA | 1.81 | 118.82 |
|  | 2018-04 to 2019-03 | 6363 | 2096 | 64.29% | 1.91 | 119.68 |
|  | 2019-04 to 2020-03 | 5456 | 2175 | 65.82% | 2.08 | 133.15 |
|  | 2020-04 to 2021-03 | 5291 | 2174 | 60.15% | 2.16 | 137.38 |
| **Average** | | | | **63.42%** | **1.99** | **127.26** |
| **Customer Lifetime Value** | | | | **399.01517** | | |
| **Increment from last CLV** | | | | **10.91** | | |

The results from the above tables can be consolidated into a single table as shown below:

Table : Period of Analysis VS CLV

|  |  |
| --- | --- |
| **Period Of Analysis (Months)** | **CLV** |
| 1 | 164.13 |
| 3 | 100.56 |
| 6 | 75.50 |
| 9 | 47.91 |
| 12 | 10.91 |
| **Customer Lifetime Value** | **399.02** |

This, upon plotting, shows a graph like this:

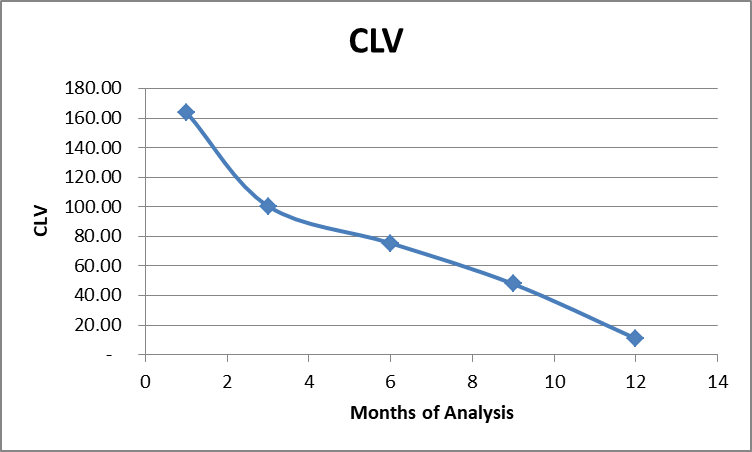


Figure : CLV VS Period of Analysis

The above plot shows how incremental CLV tends to touch the X-axis as the period of analysis is increased. The total customer lifetime value for the present case comes to be 399 USD/customer.

1. **R – Code**

The main task of the table generation as showed above done in R Studio. R did the most heavy-lifting with the packages like TidyVerse (Dplyr being part of it). The function to calculate the number for churn calculation is shown below:



1. **Other Marketing Insights**

The other marketing insights that can be drawn from the transactional data can be:

* Pattern of Churn rate over time.
* Effect of average discount per order on the churn rate
* Effect of discount on Customer Lifetime Value
* Effect of percentage discount on customer growth rate
* Effect of discount on customer acquisition
* Effect of discount of customer retention

To see these effects and draw conclusions based on it, we may need the data of average rate per month along with the average discounts given. The information table is generated is generated using the R-code and shown below:

Table : Churn rate with average discount information (monthly)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.no** | **Month** | **No. of Unique Customers** | **Return Customers from prev month** | **Churn Rate** | **Average Discount /Order** |
| 1 | 2017-04 | 29 | NA | NA | 14.5021429 |
| 2 | 2017-05 | 1191 | 2 | 0.931034483 | 20.0952919 |
| 3 | 2017-06 | 859 | 155 | 0.869857263 | 7.00231451 |
| 4 | 2017-07 | 1026 | 141 | 0.835855646 | 5.47148047 |
| 5 | 2017-08 | 917 | 181 | 0.823586745 | 7.35804954 |
| 6 | 2017-09 | 827 | 154 | 0.832061069 | 7.25715777 |
| 7 | 2017-10 | 961 | 165 | 0.800483676 | 11.6991953 |
| 8 | 2017-11 | 905 | 181 | 0.811654527 | 18.9679709 |
| 9 | 2017-12 | 799 | 145 | 0.839779006 | 26.8271463 |
| 10 | 2018-01 | 803 | 85 | 0.893617021 | 17.74157 |
| 11 | 2018-02 | 798 | 105 | 0.869240349 | 16.9693591 |
| 12 | 2018-03 | 928 | 139 | 0.825814536 | 19.000572 |
| 13 | 2018-04 | 1116 | 203 | 0.78125 | 22.4181411 |
| 14 | 2018-05 | 1029 | 207 | 0.814516129 | 17.8523911 |
| 15 | 2018-06 | 1038 | 161 | 0.843537415 | 14.3004792 |
| 16 | 2018-07 | 1122 | 172 | 0.834296724 | 13.1140839 |
| 17 | 2018-08 | 1044 | 176 | 0.843137255 | 14.5368324 |
| 18 | 2018-09 | 988 | 161 | 0.845785441 | 19.0289135 |
| 19 | 2018-10 | 882 | 148 | 0.850202429 | 15.984163 |
| 20 | 2018-11 | 1291 | 236 | 0.732426304 | 30.9881628 |
| 21 | 2018-12 | 729 | 183 | 0.858249419 | 29.236875 |
| 22 | 2019-01 | 644 | 93 | 0.872427984 | 20.0584163 |
| 23 | 2019-02 | 713 | 99 | 0.846273292 | 19.0818775 |
| 24 | 2019-03 | 973 | 151 | 0.788218794 | 16.7113714 |
| 25 | 2019-04 | 1052 | 163 | 0.832476876 | 19.002373 |
| 26 | 2019-05 | 878 | 152 | 0.855513308 | 18.59027 |
| 27 | 2019-06 | 1071 | 187 | 0.787015945 | 25.4638428 |
| 28 | 2019-07 | 850 | 187 | 0.825396825 | 21.272254 |
| 29 | 2019-08 | 896 | 176 | 0.792941176 | 19.187535 |
| 30 | 2019-09 | 834 | 158 | 0.823660714 | 21.2060894 |
| 31 | 2019-10 | 725 | 145 | 0.826139089 | 21.8619118 |
| 32 | 2019-11 | 1112 | 217 | 0.700689655 | 29.3516544 |
| 33 | 2019-12 | 722 | 125 | 0.887589928 | 27.201069 |
| 34 | 2020-01 | 862 | 137 | 0.810249307 | 21.4081403 |
| 35 | 2020-02 | 800 | 170 | 0.802784223 | 20.9958462 |
| 36 | 2020-03 | 958 | 199 | 0.75125 | 22.6845177 |
| 37 | 2020-04 | 950 | 204 | 0.787056367 | 20.58346 |
| 38 | 2020-05 | 994 | 189 | 0.801052632 | 23.93625 |
| 39 | 2020-06 | 867 | 208 | 0.790744467 | 21.576967 |
| 40 | 2020-07 | 806 | 157 | 0.818915802 | 18.7360235 |
| 41 | 2020-08 | 999 | 189 | 0.765508685 | 24.2660932 |
| 42 | 2020-09 | 893 | 197 | 0.802802803 | 23.5762151 |
| 43 | 2020-10 | 834 | 165 | 0.815229563 | 18.60322 |
| 44 | 2020-11 | 1248 | 263 | 0.684652278 | 36.5536621 |
| 45 | 2020-12 | 623 | 140 | 0.887820513 | 26.4538402 |
| 46 | 2021-01 | 791 | 129 | 0.7929374 | 19.501326 |
| 47 | 2021-02 | 715 | 146 | 0.815423515 | 19.6889614 |
| 48 | 2021-03 | 1008 | 187 | 0.738461538 | 19.3143622 |
| 49 | 2021-04 | 867 | 167 | 0.834325397 | 22.3757734 |

**13.1 Pattern of Churn rate over time:** One of the simplest insights from the data is how churn rate varies over time. This can provide the information about the effect of marketing campaigns. Normally a downward trending churn rate what is desired. If the churn rate has a positive trend, then it may be the matter of concern and may need an immediate attention from the managers of the firm. The pattern of churn rate in the current case is shown below:

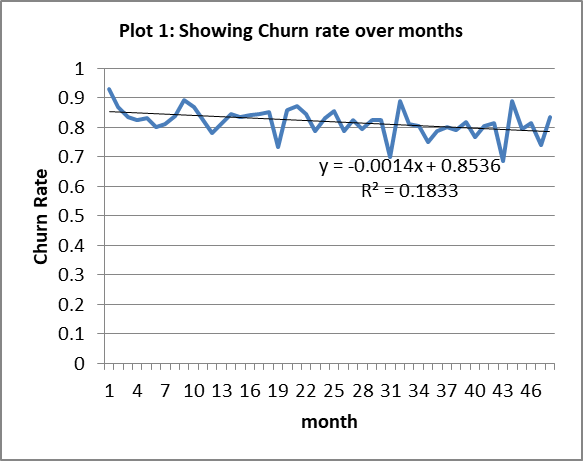


Figure : Churn Rate VS Month

The above chart shows that the coefficient of the time in churn rate VS time graph is slightly negative with the R-sq value of 18.3%. Although the coefficient is negative which shows the company is already working to reduce the churn rate but it is too small, which means there is a further scope for the company to reduce churn rate further. The efficient customer retention programs like loyalty programs etc. can be incorporated to bring a significant change in the churn rate. To understand the quantum of churn rate further a separate table showing the frequency of store visits by customers is shown below:

Table : Store Visit Frequency of customers

|  |  |  |
| --- | --- | --- |
| **Table 3: Showing Store visit frequency of customers** | | |
| **Number of store visits** | **Number of customers** | **Cumulative Percentage** |
| 1 | 9203 | 57.6% |
| 2 | 2455 | 73.0% |
| 3 | 1185 | 80.4% |
| 4 | 718 | 84.9% |
| 5 | 483 | 87.9% |
| 6 | 353 | 90.2% |
| 7 | 276 | 91.9% |
| 8 | 226 | 93.3% |
| 9 | 168 | 94.4% |
| 10 | 149 | 95.3% |
| 11 | 132 | 96.1% |
| 12 | 99 | 96.7% |
| 13 | 73 | 97.2% |
| 14 | 79 | 97.7% |
| 15 | 63 | 98.1% |
| 16 | 46 | 98.4% |
| 17 | 45 | 98.6% |
| 18 | 35 | 98.9% |
| 19 | 24 | 99.0% |
| 20 | 25 | 99.2% |
| 21 | 18 | 99.3% |
| 22 | 16 | 99.4% |
| 23 | 16 | 99.5% |
| 24 | 20 | 99.6% |
| 25 | 9 | 99.7% |
| 26 | 11 | 99.7% |
| 27 | 3 | 99.8% |
| 28 | 11 | 99.8% |
| 29 | 3 | 99.8% |
| 30 | 6 | 99.9% |
| 31 | 3 | 99.9% |
| 32 | 6 | 99.9% |
| 33 | 1 | 99.9% |
| 34 | 1 | 99.9% |
| 35 | 1 | 100.0% |
| 36 | 1 | 100.0% |
| 38 | 1 | 100.0% |
| 39 | 1 | 100.0% |
| 43 | 2 | 100.0% |
| 47 | 2 | 100.0% |
| **Total Unique Customers** | **15969** |  |

**13.2 Effect of average discount per order on the churn rate:** The above generated churn table can be used in this case too. We can use the columns of churn rate and average discount a scatter plot to inspect the same. The plot (generated in Excel) is shown below:

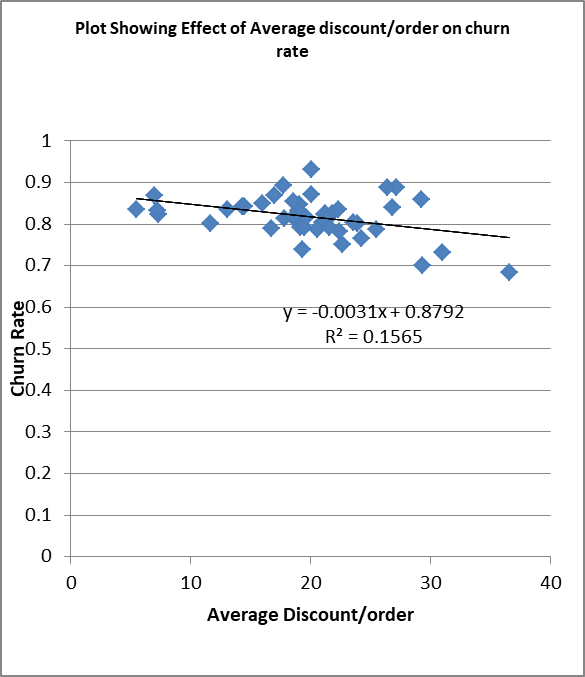


Figure : Churn Rate VS Average Discount

The above plot shows that the churn rate varies negatively with the average discounts given per month. The coefficient of discount being negative with the R-sq value of 15.65%.

**13.3 Effect of discount on Customer Lifetime Value:** As we have seen from the above plot that the churn rate varies inversely with the offered discounts and since churn rate is inversely proportional to the customer lifetime value, the offered discounts tend to have a positive effect on the CLV. On the other hand when a firm offers discounts it directly affects the revenue generation and hence will reduce the customer lifetime value. Since there are two opposing effects of discounts on the customer lifetime value, it means there must be an optimum value of average discount which can maximise the CLV. To find the optimum discount value, we can plug the model fitted in churn rate VS discount in the CLV calculation & generate a data table as shown below:

Table : Average discount & CLV info

|  |  |
| --- | --- |
| **Table showing Effect of average discount on CLV** | |
| **Average Discount** | **CLV** |
|  | 461.6481 |
| 10 | 479.1197 |
| 11 | 477.436 |
| 12 | 475.7399 |
| 13 | 474.0313 |
| 14 | 472.31 |
| 15 | 470.5759 |
| 16 | 468.8288 |
| 17 | 467.0686 |
| 18 | 465.2952 |
| 19 | 463.5083 |
| 20 | 461.7079 |
| 21 | 459.8938 |
| 22 | 458.0659 |
| 23 | 456.2239 |
| 24 | 454.3677 |
| 25 | 452.4971 |
| 26 | 450.612 |
| 27 | 448.7123 |
| 28 | 446.7976 |
| 29 | 444.868 |
| 30 | 442.9231 |
| 31 | 440.9628 |
| 32 | 438.987 |
| 33 | 436.9953 |
| 34 | 434.9877 |
| 35 | 432.964 |
| 36 | 430.9239 |
| 37 | 428.8673 |
| 38 | 426.7939 |
| 39 | 424.7036 |
| 40 | 422.5961 |

The above information when plotted gives a graph like this:

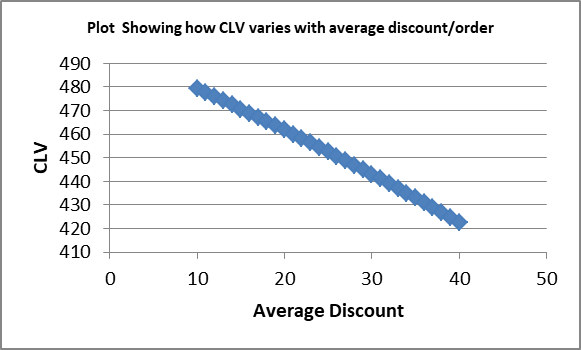


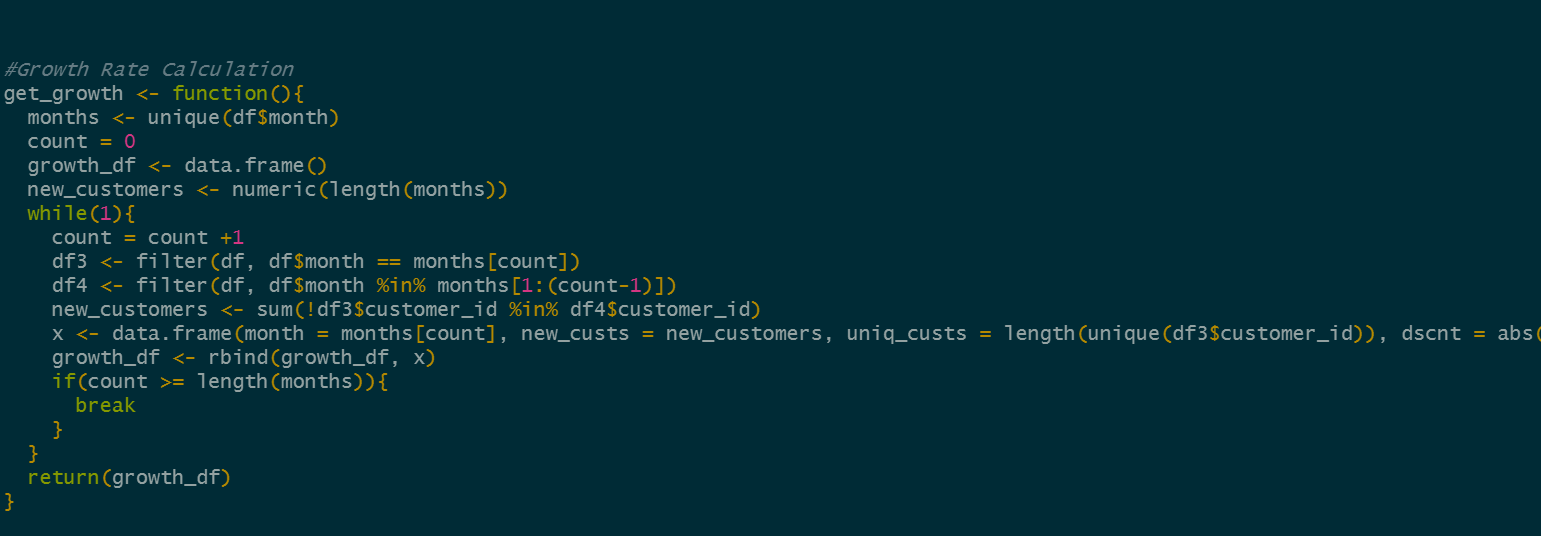
Figure : CLV VS Average Discount

The above plot clearly says that the negative effect of discounts offered is more than its positive effect through reduction of churn. Hence it can be concluded that the discounts offered by the firm are doing no good to it.

**13.4 Effect of percentage discount on customer growth rate**: Growth rate refers to the rate at which the firm acquires new customers. More formally it can be defined as follows:

Growth Rate = New Customers acquired this month/ Total customers in last month.

In order calculate the monthly growth rate of the firm; we first need to know number of first time customers each month. This calculation is done in R and is shown below:



The above R-function is used to generate the growth rate table first for all the customers in the dataset and then only those customers who availed the discount. The second table was generated to check the effect of discount on the growth rate. The R-function generates a table (for all customers) as shown below:

Table : Growth Rate

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table - Growth Rate: All Customers** | | | | | | |
| **Month** | **First Time Customers** | **Total Customers** | **Growth Rate** | **Discount Amount** | **Total Gross Sales** | **%age Discount Given** |
| 2017-04 | 29 | 29 |  | 406.06 | 4215.17 | 9.63% |
| 2017-05 | 1189 | 1191 | 4100.0% | 26164.07 | 178331.7 | 14.67% |
| 2017-06 | 699 | 859 | 58.7% | 6323.09 | 95009.59 | 6.66% |
| 2017-07 | 703 | 1026 | 81.8% | 6024.1 | 115806.86 | 5.20% |
| 2017-08 | 524 | 917 | 51.1% | 7129.95 | 127584.63 | 5.59% |
| 2017-09 | 416 | 827 | 45.4% | 6255.67 | 112343.13 | 5.57% |
| 2017-10 | 426 | 961 | 51.5% | 11921.48 | 146461.57 | 8.14% |
| 2017-11 | 395 | 905 | 41.1% | 18228.22 | 138161.24 | 13.19% |
| 2017-12 | 357 | 799 | 39.4% | 22186.05 | 124819.4 | 17.77% |
| 2018-01 | 387 | 803 | 48.4% | 14690.02 | 119229.73 | 12.32% |
| 2018-02 | 364 | 798 | 45.3% | 14033.66 | 115989.71 | 12.10% |
| 2018-03 | 381 | 928 | 47.7% | 18601.56 | 151016.47 | 12.32% |
| 2018-04 | 434 | 1116 | 46.8% | 27013.86 | 184999.41 | 14.60% |
| 2018-05 | 441 | 1029 | 39.5% | 19262.73 | 144246.75 | 13.35% |
| 2018-06 | 441 | 1038 | 42.9% | 15816.33 | 141997.51 | 11.14% |
| 2018-07 | 464 | 1122 | 44.7% | 15317.25 | 140937.44 | 10.87% |
| 2018-08 | 411 | 1044 | 36.6% | 15787 | 136218.09 | 11.59% |
| 2018-09 | 357 | 988 | 34.2% | 19790.07 | 142809.7 | 13.86% |
| 2018-10 | 306 | 882 | 31.0% | 14705.43 | 122670.3 | 11.99% |
| 2018-11 | 334 | 1291 | 37.9% | 43011.57 | 234231.55 | 18.36% |
| 2018-12 | 247 | 729 | 19.1% | 21986.13 | 120312.76 | 18.27% |
| 2019-01 | 230 | 644 | 31.6% | 13298.73 | 95392.69 | 13.94% |
| 2019-02 | 232 | 713 | 36.0% | 14330.49 | 114985.1 | 12.46% |
| 2019-03 | 370 | 973 | 51.9% | 16694.66 | 139050.55 | 12.01% |
| 2019-04 | 432 | 1052 | 44.4% | 20579.57 | 146155.66 | 14.08% |
| 2019-05 | 316 | 878 | 30.0% | 17214.59 | 141761.4 | 12.14% |
| 2019-06 | 285 | 1071 | 32.5% | 28825.07 | 190827.89 | 15.11% |
| 2019-07 | 252 | 850 | 23.5% | 18591.95 | 125869.57 | 14.77% |
| 2019-08 | 273 | 896 | 32.1% | 17825.22 | 137761.47 | 12.94% |
| 2019-09 | 230 | 834 | 25.7% | 18491.71 | 135502.45 | 13.65% |
| 2019-10 | 185 | 725 | 22.2% | 16352.71 | 120275.71 | 13.60% |
| 2019-11 | 221 | 1112 | 30.5% | 35838.37 | 221833.17 | 16.16% |
| 2019-12 | 189 | 722 | 17.0% | 20101.59 | 119519.54 | 16.82% |
| 2020-01 | 220 | 862 | 30.5% | 20145.06 | 146174.97 | 13.78% |
| 2020-02 | 197 | 800 | 22.9% | 17741.49 | 134800.54 | 13.16% |
| 2020-03 | 204 | 958 | 25.5% | 23750.69 | 173155.79 | 13.72% |
| 2020-04 | 230 | 950 | 24.0% | 20583.46 | 159298.63 | 12.92% |
| 2020-05 | 246 | 994 | 25.9% | 25276.68 | 158586.21 | 15.94% |
| 2020-06 | 221 | 867 | 22.2% | 19635.04 | 141405.59 | 13.89% |
| 2020-07 | 209 | 806 | 24.1% | 15925.62 | 130620.51 | 12.19% |
| 2020-08 | 212 | 999 | 26.3% | 25527.93 | 174916.36 | 14.59% |
| 2020-09 | 204 | 893 | 20.4% | 21925.88 | 142776 | 15.36% |
| 2020-10 | 190 | 834 | 21.3% | 16408.04 | 137766.32 | 11.91% |
| 2020-11 | 236 | 1248 | 28.3% | 50407.5 | 277686.28 | 18.15% |
| 2020-12 | 170 | 623 | 13.6% | 17221.45 | 106570.86 | 16.16% |
| 2021-01 | 201 | 791 | 32.3% | 16030.09 | 128124.12 | 12.51% |
| 2021-02 | 190 | 715 | 24.0% | 14786.41 | 125460.21 | 11.79% |
| 2021-03 | 257 | 1008 | 35.9% | 21651.4 | 177181.03 | 12.22% |
| 2021-04 | 262 | 867 | 26.0% | 20540.96 | 153523.17 | 13.38% |

The growth rate table for only the customers who availed discount is shown below:

Table : Growth for customers who availed the discount

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 5: Growth Rate - Only those customers who availed discount** | | | | | | |
| **Month** | **First Time Customers** | **Total Customers** | **Growth Rate** | **Discount Amount** | **Total Gross Sales** | **%age Discount Given** |
| 2017-04 | 15 | 15 |  | 406.06 | 2565.7 | 15.83% |
| 2017-05 | 741 | 742 | 4940.00% | 26164.1 | 127816.38 | 20.47% |
| 2017-06 | 319 | 376 | 42.99% | 6323.09 | 43843.87 | 14.42% |
| 2017-07 | 208 | 262 | 55.32% | 6024.1 | 35638.05 | 16.90% |
| 2017-08 | 251 | 347 | 95.80% | 7129.95 | 58252.51 | 12.24% |
| 2017-09 | 124 | 192 | 35.73% | 6255.67 | 35704.59 | 17.52% |
| 2017-10 | 209 | 341 | 108.85% | 11921.5 | 70992.74 | 16.79% |
| 2017-11 | 396 | 628 | 116.13% | 18228.2 | 105591.05 | 17.26% |
| 2017-12 | 456 | 721 | 72.61% | 22186.1 | 118144.35 | 18.78% |
| 2018-01 | 361 | 608 | 50.07% | 14690 | 108083.55 | 13.59% |
| 2018-02 | 299 | 580 | 49.18% | 14033.7 | 103378.72 | 13.57% |
| 2018-03 | 338 | 726 | 58.28% | 18601.6 | 139907.68 | 13.30% |
| 2018-04 | 373 | 890 | 51.38% | 27013.9 | 172697.86 | 15.64% |
| 2018-05 | 300 | 700 | 33.71% | 19262.7 | 127257.85 | 15.14% |
| 2018-06 | 303 | 705 | 43.29% | 15816.3 | 124779.93 | 12.68% |
| 2018-07 | 273 | 712 | 38.72% | 15317.3 | 121806.4 | 12.58% |
| 2018-08 | 254 | 693 | 35.67% | 15787 | 120390.2 | 13.11% |
| 2018-09 | 269 | 720 | 38.82% | 19790.1 | 129708.41 | 15.26% |
| 2018-10 | 217 | 622 | 30.14% | 14705.4 | 109079.55 | 13.48% |
| 2018-11 | 305 | 1078 | 49.04% | 43011.6 | 220036.08 | 19.55% |
| 2018-12 | 211 | 623 | 19.57% | 21986.1 | 114650.06 | 19.18% |
| 2019-01 | 168 | 450 | 26.97% | 13298.7 | 83020.7 | 16.02% |
| 2019-02 | 172 | 529 | 38.22% | 14330.5 | 104374.62 | 13.73% |
| 2019-03 | 266 | 735 | 50.28% | 16694.7 | 127143.81 | 13.13% |
| 2019-04 | 374 | 843 | 50.88% | 20579.6 | 136284.82 | 15.10% |
| 2019-05 | 235 | 678 | 27.88% | 17214.6 | 132230.05 | 13.02% |
| 2019-06 | 213 | 859 | 31.42% | 28825.1 | 179794.59 | 16.03% |
| 2019-07 | 195 | 645 | 22.70% | 18592 | 116754.04 | 15.92% |
| 2019-08 | 201 | 681 | 31.16% | 17825.2 | 126444.79 | 14.10% |
| 2019-09 | 161 | 665 | 23.64% | 18491.7 | 127506.44 | 14.50% |
| 2019-10 | 137 | 558 | 20.60% | 16352.7 | 111917.09 | 14.61% |
| 2019-11 | 191 | 958 | 34.23% | 35838.4 | 210932.73 | 16.99% |
| 2019-12 | 149 | 556 | 15.55% | 20101.6 | 107997.62 | 18.61% |
| 2020-01 | 174 | 687 | 31.29% | 20145.1 | 136245.72 | 14.79% |
| 2020-02 | 153 | 633 | 22.27% | 17741.5 | 126295.76 | 14.05% |
| 2020-03 | 164 | 785 | 25.91% | 23750.7 | 163428.26 | 14.53% |
| 2020-04 | 202 | 817 | 25.73% | 20583.5 | 152009.62 | 13.54% |
| 2020-05 | 233 | 849 | 28.52% | 25276.7 | 151451.17 | 16.69% |
| 2020-06 | 189 | 711 | 22.26% | 19635 | 133359.72 | 14.72% |
| 2020-07 | 163 | 654 | 22.93% | 15925.6 | 122248.88 | 13.03% |
| 2020-08 | 187 | 854 | 28.59% | 25527.9 | 166835.36 | 15.30% |
| 2020-09 | 179 | 736 | 20.96% | 21925.9 | 133847.56 | 16.38% |
| 2020-10 | 158 | 692 | 21.47% | 16408 | 130774.65 | 12.55% |
| 2020-11 | 236 | 1154 | 34.10% | 50407.5 | 272242.42 | 18.52% |
| 2020-12 | 119 | 451 | 10.31% | 17221.5 | 94175.57 | 18.29% |
| 2021-01 | 144 | 614 | 31.93% | 16030.1 | 118441.09 | 13.53% |
| 2021-02 | 146 | 554 | 23.78% | 14786.4 | 112598.19 | 13.13% |
| 2021-03 | 206 | 797 | 37.18% | 21651.4 | 164347.29 | 13.17% |
| 2021-04 | 178 | 687 | 22.33% | 20541 | 144285.77 | 14.24% |

When we plot the numbers from first table we get a graph like this:

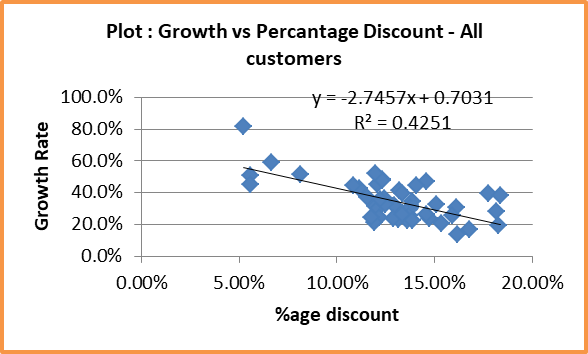


Figure : Growth Rate VS %age discount

The above plot shows a downward trend of growth rate against the offered discounts but in the context of each & every customer, whether they availed the discount or not. To check whether the discount has any positive effect on growth rate, we need to plot the same data but for only those customers who availed the discount. The said graph is shown below:

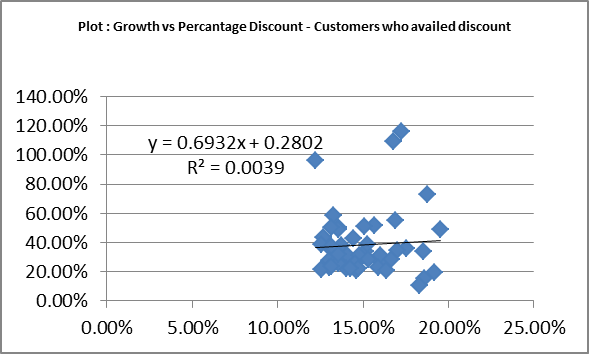
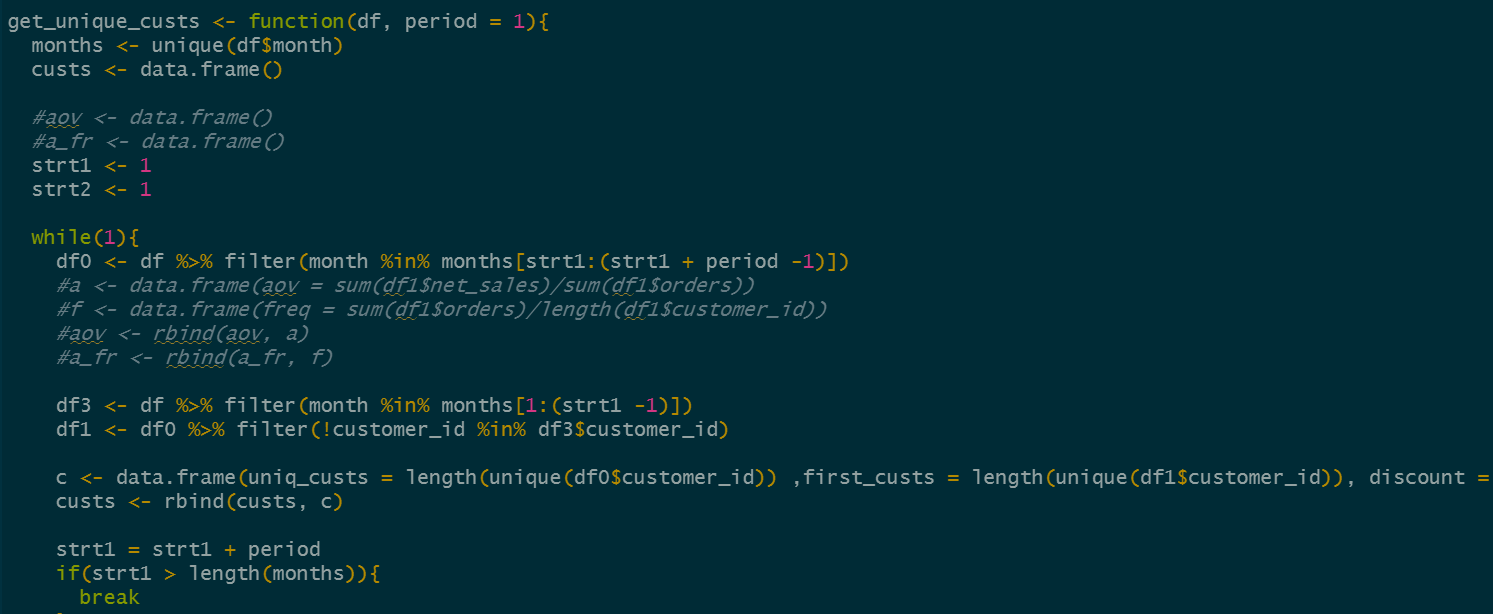
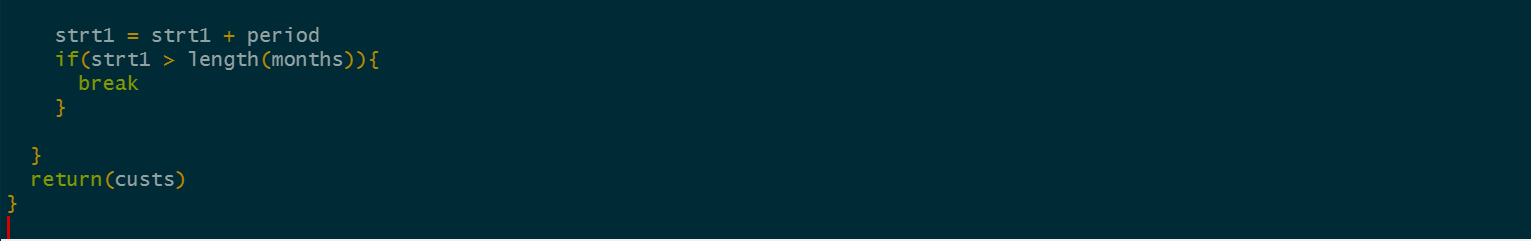


Figure : Growth Rate VS %age discount - only customers who availed discount

The above plot shows that when the data is filtered only for those customers who have actually availed discount, the curve between growth rate & discount deflects in the positive direction.

**13.5 Effect of discount on customer acquisition**: Customer acquisition means to acquire new customers. New customer acquisition is very important for a firm to survive. Companies often run advertisement campaigns, offer discounts etc. to attract new customers. In the present context, an interesting insight that we might draw from the data would be the effect of discounts on customer acquisition. To check the same, the number of first time customers are generated from the R-code as shown below:





The above code generates a table as shown below:

Table : First time and return customers (monthly)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Month** | **Total Customers** | **First Time Customers** | **Return Customers** | **Discount** |
| 2017-04 | 29 | 29 | 0 | 0 |
| 2017-05 | 1191 | 1189 | 2 | 26073 |
| 2017-06 | 859 | 699 | 160 | 4804.38 |
| 2017-07 | 1026 | 703 | 323 | 4533.36 |
| 2017-08 | 917 | 524 | 393 | 4136.08 |
| 2017-09 | 826 | 415 | 411 | 2980.58 |
| 2017-10 | 961 | 426 | 535 | 4498.88 |
| 2017-11 | 905 | 395 | 510 | 5574.7 |
| 2017-12 | 798 | 356 | 442 | 9030.03 |
| 2018-01 | 801 | 385 | 416 | 6160.84 |
| 2018-02 | 796 | 362 | 434 | 5081.27 |
| 2018-03 | 926 | 378 | 548 | 5400.24 |
| 2018-04 | 1110 | 427 | 683 | 7658.61 |
| 2018-05 | 1024 | 436 | 588 | 7999.17 |
| 2018-06 | 1032 | 435 | 597 | 4596.47 |
| 2018-07 | 1109 | 450 | 659 | 4639.7 |
| 2018-08 | 1041 | 407 | 634 | 4556.91 |
| 2018-09 | 988 | 357 | 631 | 5373.17 |
| 2018-10 | 879 | 303 | 576 | 4585.62 |
| 2018-11 | 1289 | 332 | 957 | 8094.26 |
| 2018-12 | 728 | 246 | 482 | 5843.85 |
| 2019-01 | 643 | 229 | 414 | 3220.94 |
| 2019-02 | 711 | 231 | 480 | 3978.62 |
| 2019-03 | 960 | 357 | 603 | 4661.19 |
| 2019-04 | 1027 | 407 | 620 | 6434.3 |
| 2019-05 | 867 | 303 | 564 | 4966.45 |
| 2019-06 | 1066 | 281 | 785 | 4120.9 |
| 2019-07 | 844 | 245 | 599 | 4161.01 |
| 2019-08 | 890 | 266 | 624 | 4250.99 |
| 2019-09 | 823 | 219 | 604 | 3283.39 |
| 2019-10 | 720 | 181 | 539 | 3370.01 |
| 2019-11 | 1106 | 214 | 892 | 4941.28 |
| 2019-12 | 717 | 184 | 533 | 3031.21 |
| 2020-01 | 850 | 208 | 642 | 3339.71 |
| 2020-02 | 794 | 192 | 602 | 2994.26 |
| 2020-03 | 955 | 201 | 754 | 3343.77 |
| 2020-04 | 944 | 224 | 720 | 3414.6 |
| 2020-05 | 987 | 239 | 748 | 5183.59 |
| 2020-06 | 864 | 218 | 646 | 3424.83 |
| 2020-07 | 800 | 204 | 596 | 3069.21 |
| 2020-08 | 994 | 207 | 787 | 3637.3 |
| 2020-09 | 889 | 201 | 688 | 4205.39 |
| 2020-10 | 831 | 187 | 644 | 2801.32 |
| 2020-11 | 1235 | 225 | 1010 | 7444.39 |
| 2020-12 | 622 | 169 | 453 | 2997.54 |
| 2021-01 | 787 | 197 | 590 | 3115.44 |
| 2021-02 | 712 | 185 | 527 | 3111 |
| 2021-03 | 997 | 248 | 749 | 4464.46 |
| 2021-04 | 857 | 251 | 606 | 4373.18 |

To check the effect of discount on customer acquisition, we need to plot the First Tme Customers column against the discount column, the graph is shown below:

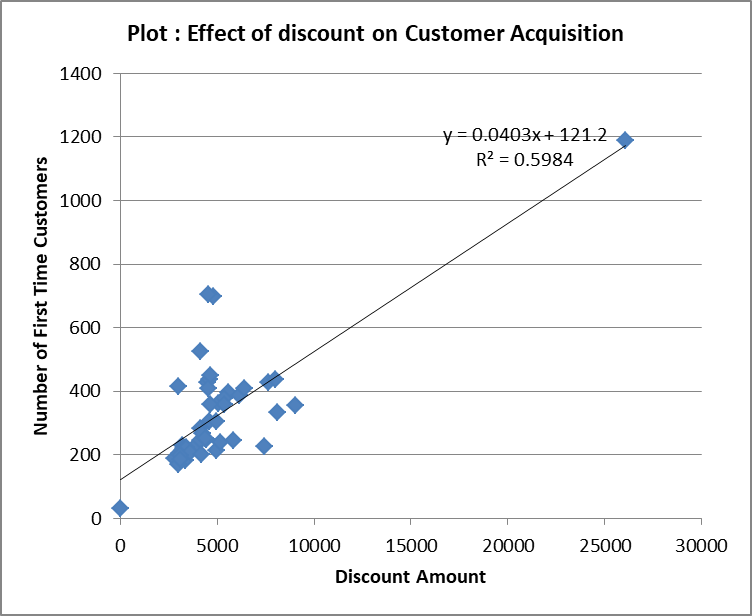


Figure : Effect of discount on Customer Acquisition

The above plot clearly shows a positive trend of increased number of first time customers against the increased trend of discount amount offered. The regression line shows the coefficient of 0.0403 which means for every 100 USD of discount offered, the company acquired an average of 4 new customers. The R-sq. value of 59.8% shows that the regression coefficient is significant.

**13.6 Effect of discount on customer retention**:

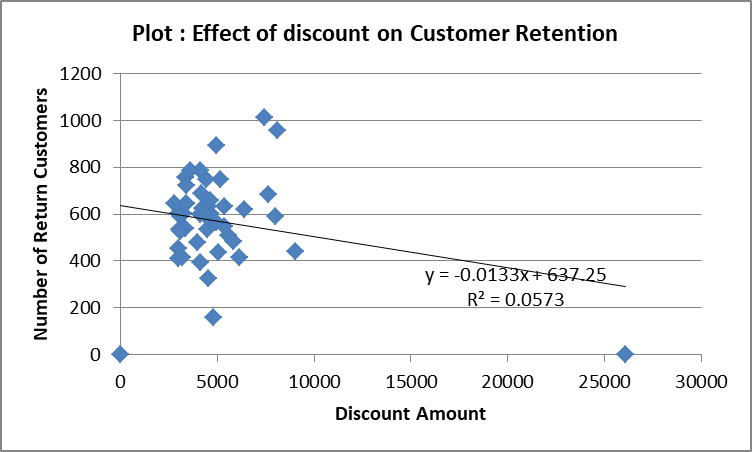


Figure : Effect of discount on customer retention

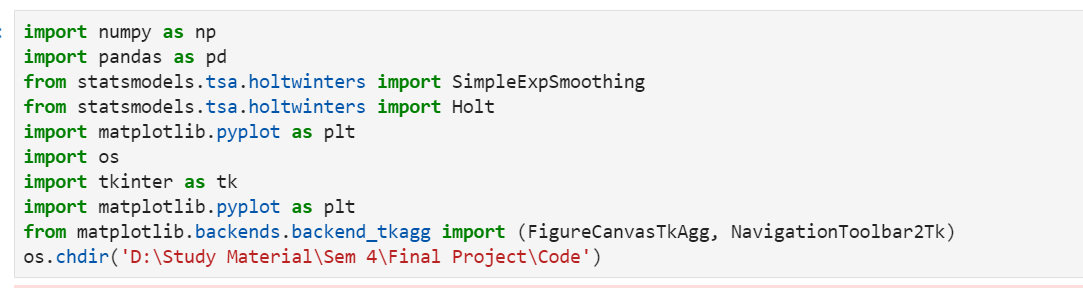
The effect of discount seems to be negative on a first glance on the customer retention but R-sq. value of the regression coefficient implies that it is not significant and hence can be considered as zero. Hence the data suggests that the discount does not help the customer retention.

1. **Inventory Optimization using Transaction Data**

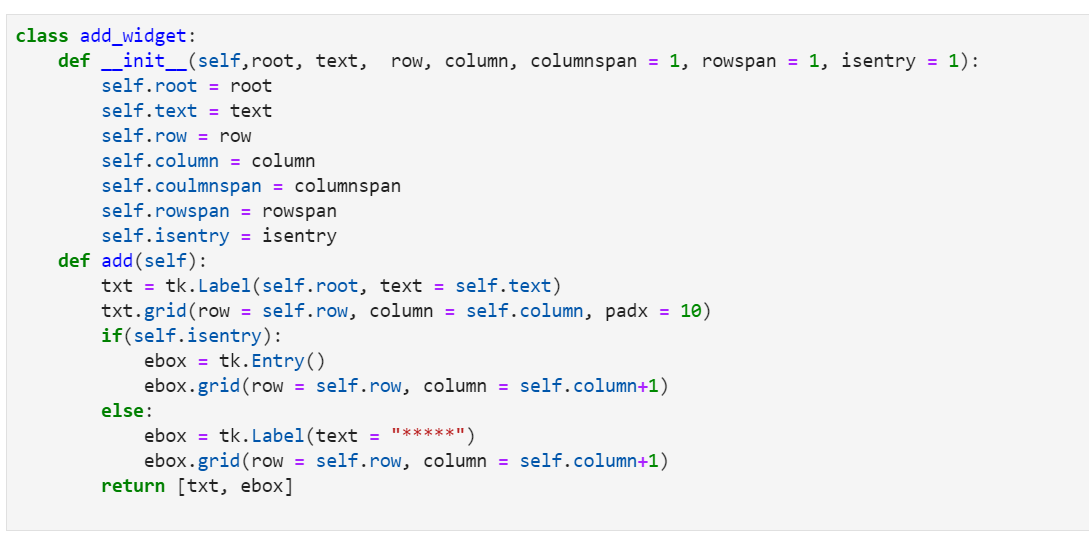
The current project not only attempts to find the optimum inventory or Economic Order Quantity for the firm under consideration but also tries to develop a generalized tool for the calculation and visualization of the same. A graphical user interface (GUI) is developed to make the process more intuitive and interactive using the Tkinter module of python.

**14.1 Python Code:** The python code is mainly divided into two parts as discussed below:

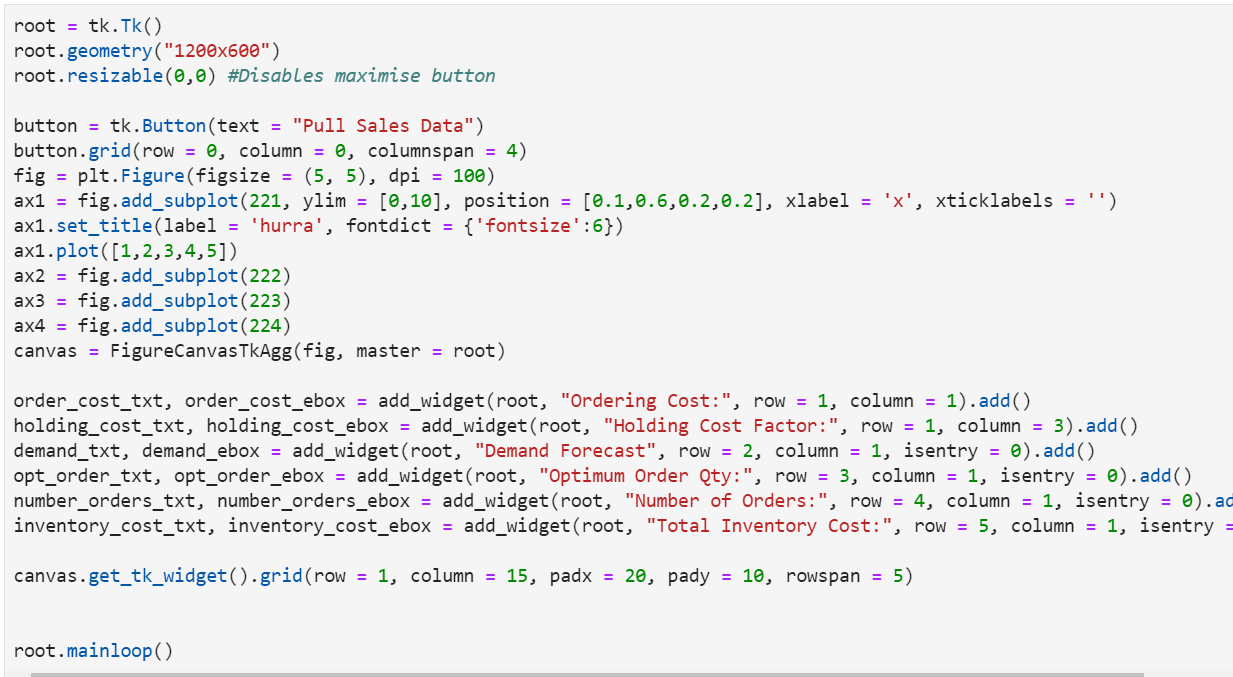
1. Code for GUI development: This is mainly the Tkinter code which creates a generic window add then adds necessary buttons and controls to it. The complete python code is shown below:



The above code snippet loads the necessary modules and packages into the environment e.g. pandas for data-frame operations, matplotlib for plotting, Tkinter for GUI development and controls etc.



This code snippet defines a add\_widget class which then used to create objects to add ontrols like button, entry fields to the GUI window. The main code which actually creates the window, adds controls to it, and starts the main loop is shown below:



The above code when executed creates a window like this:

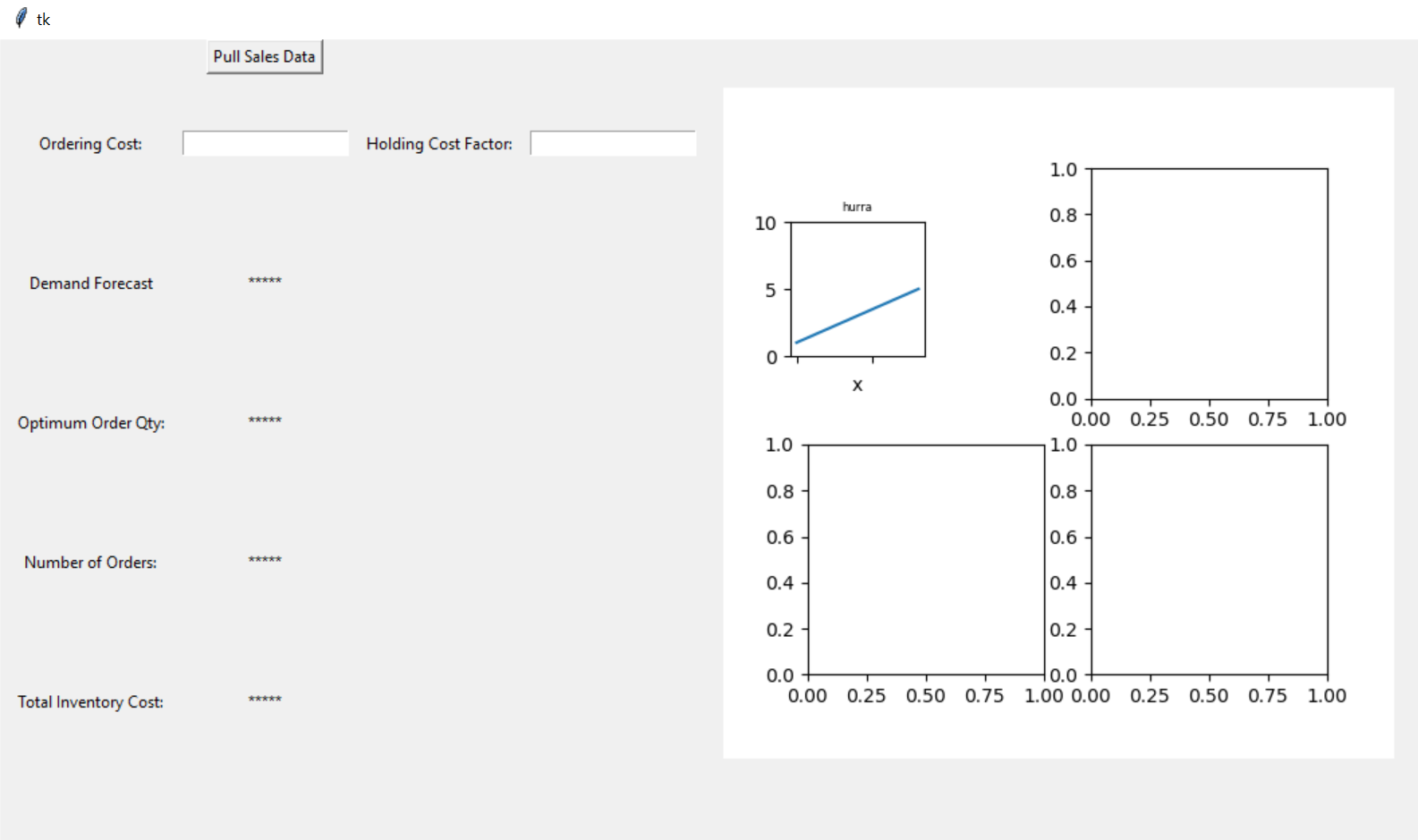
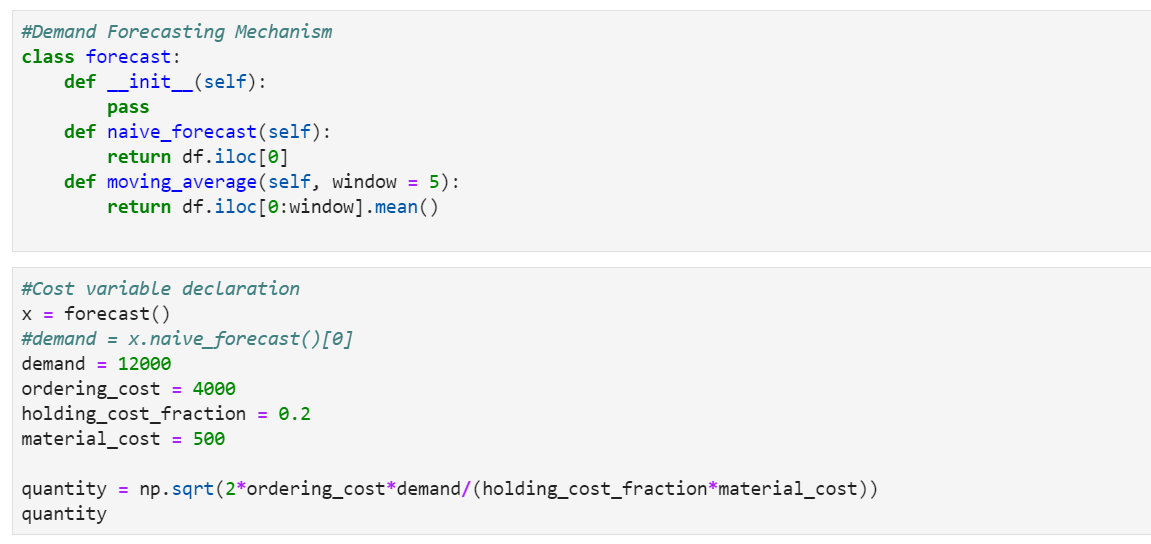


Figure : GUI Showing the main controls in inventory optimization tool

* 1. **Backend Code:** This part of code is the one which does actual heavy lifting of importing the data, making a forecast and making the calculations for inventory optimization. A part of the code is shown below:



1. **Conclusion**

The customer lifetime value is almost 400 USD, which means that the firm can expect to earn a revenue of 400 USD from each unique customer.The rate of Churn (Customers not showing up again after the first order) is very high (82%), which means the firm is not doing enough to retain the customers. To understand it further, refer to Table 8, almost 57.6% customers did not return after their first order and 73% never returned after their 2nd. Although there is a negative trend of churn rate with time (Figure 6) which means the firm has made efforts over time to retain customers but the negative slope is still low meaning the results from the efforts are not appreciable enough. The churn rate is seen to have negative relationship with the discount but its effect on CLV shows that offering discounts is not helping the firm much (refer Figure 8). Since offering discounts is not helping the firm in retaining customers, an alternative approach like better sales experience, better after sales service etc. might be considered. The discount offered appears to have a positive impact on the customer acquisition. The coefficient of discount in Customer Acquisition model is 0.04, which means for every 100 bucks spent as discount, the firm gains 4 new customers on average. The R-sq of 59.8% shows that the customer acquisition model is quite significant. For customer retention, the discount appears to impact the retention negatively at first but the R-sq of 5% implies that the coefficient is not statistically significant

16. References:

[1]: https://cran.r-project.org/web/packages/dplyr/index.html

[2]: <https://www.python.org/about/>

[3]: https://numpy.org/doc/stable/

[4]: <https://docs.python.org/3/library/tkinter.html>

**Plagiarism Report**

