Marketing Analytics II

Final Project Report

Customer Lifetime Value

Created by:

Arnav Deshwal, Palakh Gupta, Tushar Gupta

McCOMBS SCHOOL OF BUSINESS

1. Background and Business Problem

The most crucial asset for a retail organization is its customers. In the early times, a more frequent customer was provided better recognition by the staff, but how do we achieve this in the new age of e-commerce. The metrics for a superior customer have changed and so has the context of frequent visits.

Study shows that repeat customers spend 3 times more than new customers, therefore, customer retention becomes an important task. In addition to this, the cost of retaining existing customers has historically always been less than new customer acquisition costs.

Customer Retention is the process of engaging existing customers to continue buying products or services from the business. The best customer retention tactics enable the seller to create lasting relationships with consumers who will become loyal to their brand. They might even spread the word within their own circles of influence, which could turn them into brand ambassadors. Hence, we can easily observe that customer retention is necessary. Let's suppose a small business has 30,000 customers for those who have bought an item once or twice. Is it profitable to try and retain all of them? All customers retained would be an ideal scenario but, forming realistic strategies for customer retention involve losing profit margin, providing special offers, etc. Therefore, the first step is to identify which customers are profitable to retain in the long run. The million-dollar question is how do we determine customer lifetime profitability. This can be done using calculating Customer Lifetime Value.

Customer Lifetime Value is an estimated projection of the total revenue a business can obtain from a single customer account. In simpler terms - CLTV presents the amount of revenue or profit that can be generated through a customer over the period of the relationship. Over the years, it has been involved with direct response marketing. For every business, to start investing in a unique and different strategy is time-consuming and demands a high amount of resources. So, understanding why CLTV is crucial for the survival and growth of the business is really important.

Primarily, every business has the same foundation. It begins with investing in customers, it could be acquisition costs, offline ads, promotions, discounts, etc. The motivation behind this is to generate revenue and be profitable. These methods make some customers more valuable in terms of lifetime value. On the other hand, there will be a select few who will pull the profitability curve. Identifying these customers through CLTV value will reduce the slope of the drop significantly.

Let's talk in more detail. Customer Lifetime Value as mentioned has multifold advantages. A few of them are :

- It provides the companies about the approximate duration to recoup the Customer Acquisition Cost (CAC) and also make a decision if the investment will follow through or not
- According to the Pareto principle, 80% of a company's future revenue comes from 20% of their existing customers (Gartner), therefore, it's important to identify those customers. This also helps in achieving the long term goal for a company, to achieve loyalty
- Form customer segments and different potential customer segments by long term profitability to drive decisions to modify market strategy and strategize customized campaigns, group-ons, etc
- Recognize customer defection warnings at an early stage and discover those customer segments to correct the cause
- Customer Satisfaction Analysis can be performed by studying an increase or decrease in their CLV Scores. This will further drive investment in automating internal processes, offering coupons, etc

In our project, we have tried to model the Customer Lifetime Value to find customers who are more probable of providing high revenue to the business in the future. The time window over which we are trying to predict this is 90 days (3 months). We have The first step was data preparation. We have used RFM (Recency, Frequency, and Monetary) and Age on Network features in the model. These features have helped us capture the CLTV value.

For the model implementation, we have followed a two-step approach. We have started with determining the rate at which customers will make future transactions. Then, we have predicted the rate at which customers will drop out off of the system in the future. To identify these, we have implemented Pareto/NBD or BG/NBD. We have used these results to calculate customers' monetary value.

Let's discuss the business outcomes and the areas where CLTV will prove beneficial

1. There will be a reduction in spending on customer retention. Additionally, we know which customers are going to profitable in the long run, we can automate the couponing and special offer system reducing manual labor hours.
2. The % of customers, we focus on retaining are less, therefore, we can perform analysis and further create additional schemes focusing on improving the customer retention rate.
3. We can identify precise segments and strategize accordingly. This helps to create targeted campaigns than using the traditional blanket campaign approach.

2. Data Summary and Exploratory Analysis

For the purpose of our project, we have used the Online Retail Data Set from the UCI Machine Learning Repository.

This is a transactional data set that contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. The dataset contains transaction-level data of customers with 541,909 rows and 8 columns.

The raw data has the following 8 attributes:

- InvoiceNo: Invoice number of the transaction. If this code starts with letter 'c', it indicates a cancellation
- StockCode: Product (item) code
- Description: Product (item) name
- Quantity: The quantities of each product (item) per transaction
- InvoiceDate: Invoice Date and time
- UnitPrice: Product price per unit in sterling
- CustomerID: Customer number
- Country: The name of the country where each customer resides

2.1 Data cleaning

This data had a lot of records that did not contain Customer IDs or had negative order quantities and hence data cleaning is required.

The following steps were performed to clean the data:

- Records that had negative order quantities and monetary values were filtered out
- Only records with a Customer ID were kept

For the probabilistic approach followed, some additional data processing steps have been taken:

- The orders were grouped by day instead of InvoiceNo because the minimum time unit used by the probabilistic model is a day
- Only customers who bought something in the past 90 days are considered
- Only the fields that were useful for the probabilistic model are retained

2.2 Train-test split

A threshold date needed to be chosen in order to prepare the data for training the model. That date separates the orders into two partitions:

- Orders before the threshold date are used to train the model
- Orders after the threshold date are used to compute the target value

The threshold date chosen for our analysis is *09/09/2011*.

2.3 Data aggregation

After the split of the data into training and target intervals, aggregated data is used to create features and targets for each customer. For the probabilistic model, the aggregation is limited to recency, frequency, and monetary (RFM) fields.

The new features are defined as follows:

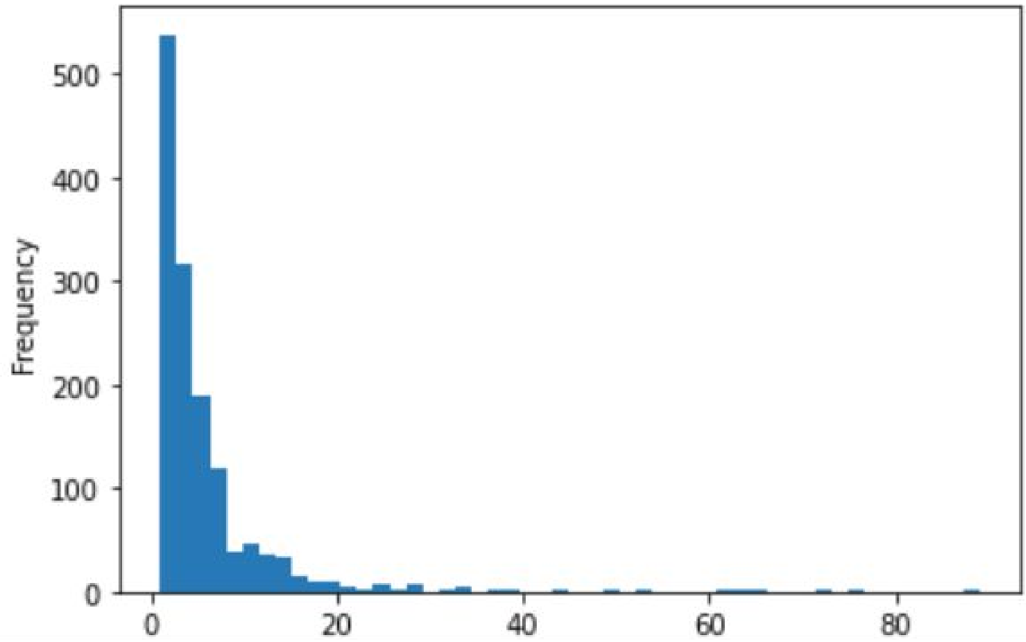
- monetary_btyd: The average of all orders' monetary values for each customer during the features period. The probabilistic model assumes that the value of the first order is 0. This has been manually enforced
- Recency: The time between the first and last orders that were placed by a customer during the features period

- frequency_btyd: The number of orders placed by a customer during the features period minus the first one
- frequency_btyd_clipped: Same as frequency_btyd, but clipped by cap outliers
- monetary_btyd_clipped: Same as monetary_btyd, but clipped by cap outliers
- target_monetary_clipped: Same as target_monetary, but clipped by cap outliers
- Target_monetary: The total amount spent by a customer

2.4 EDA

Among all customers in our data, more than 18% of them only made purchases more than once.

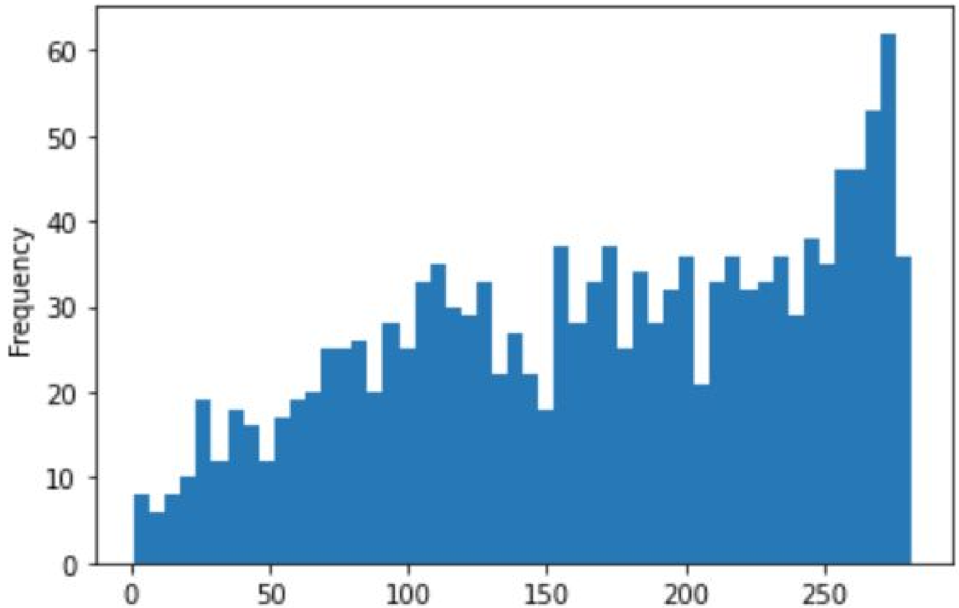
```
count    1389.000000
mean        5.455724
std         7.072419
min         1.000000
25%         2.000000
50%         3.000000
75%         6.000000
max        89.000000
Name: frequency_btyd, dtype: float64
0.18862491000719941
```



Histogram for Distribution of Frequency Feature

Most customers have been inactive initially and made their last purchase later in their lifetime.

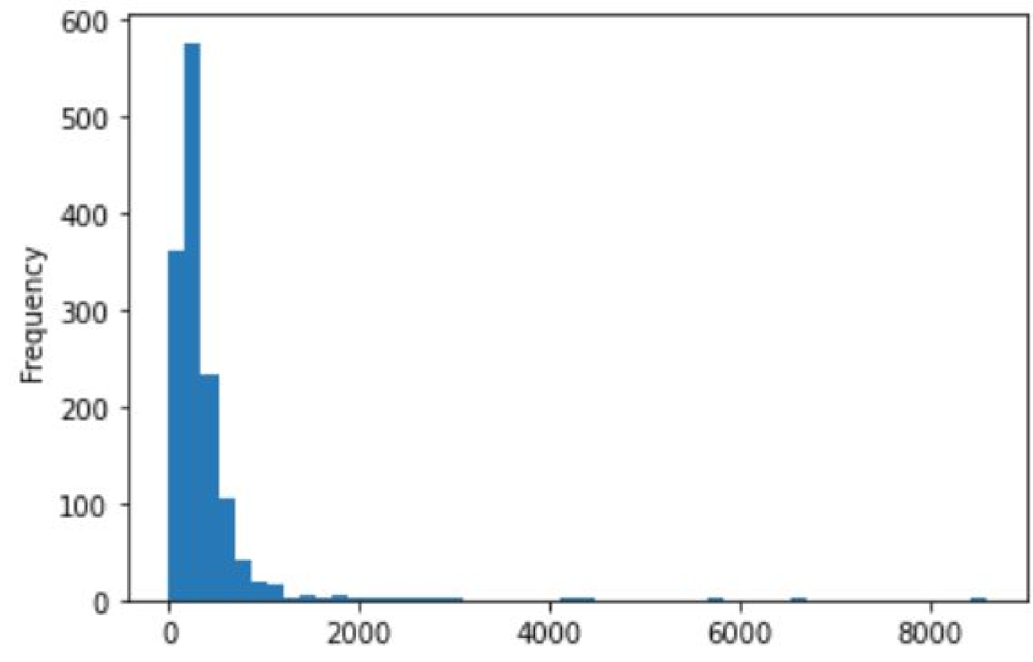
```
count    1389.000000
mean      168.851692
std        75.932794
min         1.000000
25%       108.000000
50%       175.000000
75%       237.000000
max       282.000000
Name: recency, dtype: float64
```



Histogram for Distribution of Recency Feature

Only about 30% of the customers have an average monetary value greater than the mean average monetary value of $365.

```
count     1389.000000
mean       364.974082
std        482.283032
min          0.000000
25%        168.000000
50%        264.000000
75%        397.000000
max       8574.000000
Name: monetary_btyd, dtype: float64
```



Histogram for Distribution of Monetary Feature

3. Data Analyses, Key Findings, and Conclusions

The data that we are dealing with is of non-store online retail. In such a non-contractual business setting, customers are not bound and can end their relationship with the retailer at will without any liability. This makes it difficult to assess customer behavior and understand whether the customer is “alive” i.e. will make future transactions or “dead” i.e. will never make a purchase in the future. However, in this project, we derive some signals from their transactions with the retailer.

“Buy Till You Die” statistical models help to quantify the behavioral characteristics of the customers and calculate their lifetime value by predicting the number of future transactions that the customer will do and assigning a probability to the customer being “alive”.

3.1 CLV Calculation and models

In our project, we calculate the Customer Lifetime Value in two steps:

1. Determine the rate at with customers will make future transactions and the rate at which customers will drop out of the system in the future using Pareto/NBD or BG/NBD
2. Calculate the monetary value of each customer

We have made the following assumptions:

- Number of transactions made by an active customer follows a Poisson Process given transaction rate of λ, which is E[# transactions in a given period of time]
- Heterogeneity in λ among customers follow a Gamma Distribution
- Probability of customer becoming inactive after every transaction is p
- Heterogeneity in p among customers follow a Beta Distribution
- λ and p is independent among different customers

BTYD models (Pareto/NBD or BG/NBD) give us the following three outputs:

- $P(X(t) = x \mid \lambda, p)$ - probability of observing x transactions in given time t
- $E(X(t) \mid \lambda, p)$ - expected number of transactions in given time t
- $P(\tau > t)$ - probability of the customer being inactive at time t

These fitted distributions parameters are then used to find the expected number of transactions in a future time period t for a customer having past observed behavior defined by x, t_x, T, where x = number of historical transactions, t_x = time of last purchase and T = Age of the customer. The formula mentioned below is taken from P. Fader's Paper.

$$E(Y(t) \mid X = x, t_x, T, r, \alpha, a, b) =$$
$$\frac{\frac{a+b+x-1}{a-1}\left[1-\left(\frac{\alpha+T}{\alpha+T+t}\right)^{r+x} {}_2F_1\left(r+x, b+x; a+b+x-1; \frac{t}{\alpha+T+t}\right)\right]}{1+\delta_{x>0}\frac{a}{b+x-1}\left(\frac{\alpha+T}{\alpha+t_x}\right)^{r+x}}$$

The expected number of transactions in a future period of length t for an individual with past observed behavior (X = x, t_x, T; where x = n. historical transactions, t_x = time of last purchase and T = Age of a customer) given the fitted model parameters r, α, a, b

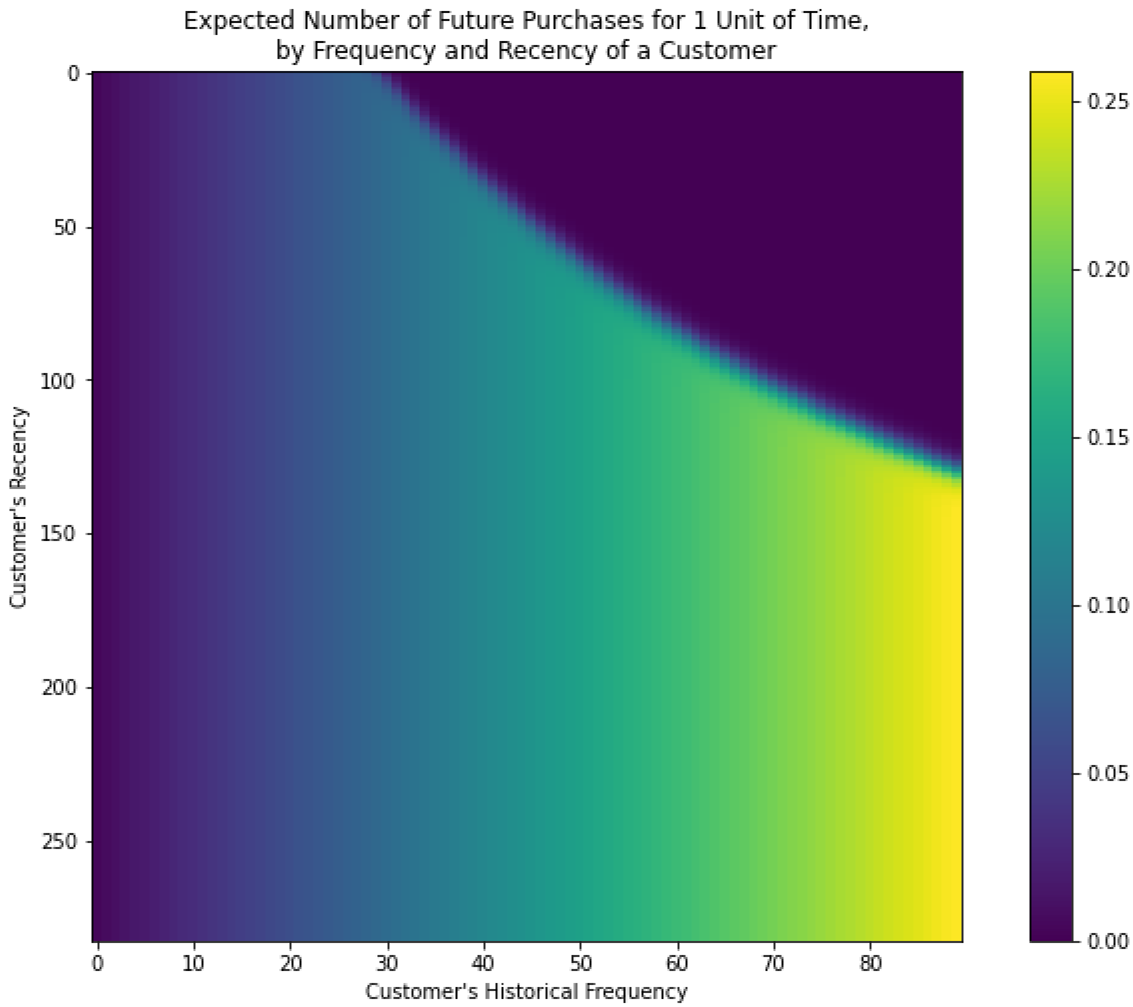
The outputs of the above-mentioned probabilistic model are used to model the future monetary value of the customers. The probabilistic method assumes that monetary value follows a gamma-gamma distribution.

For both the models, we use the Python package called Lifetimes.

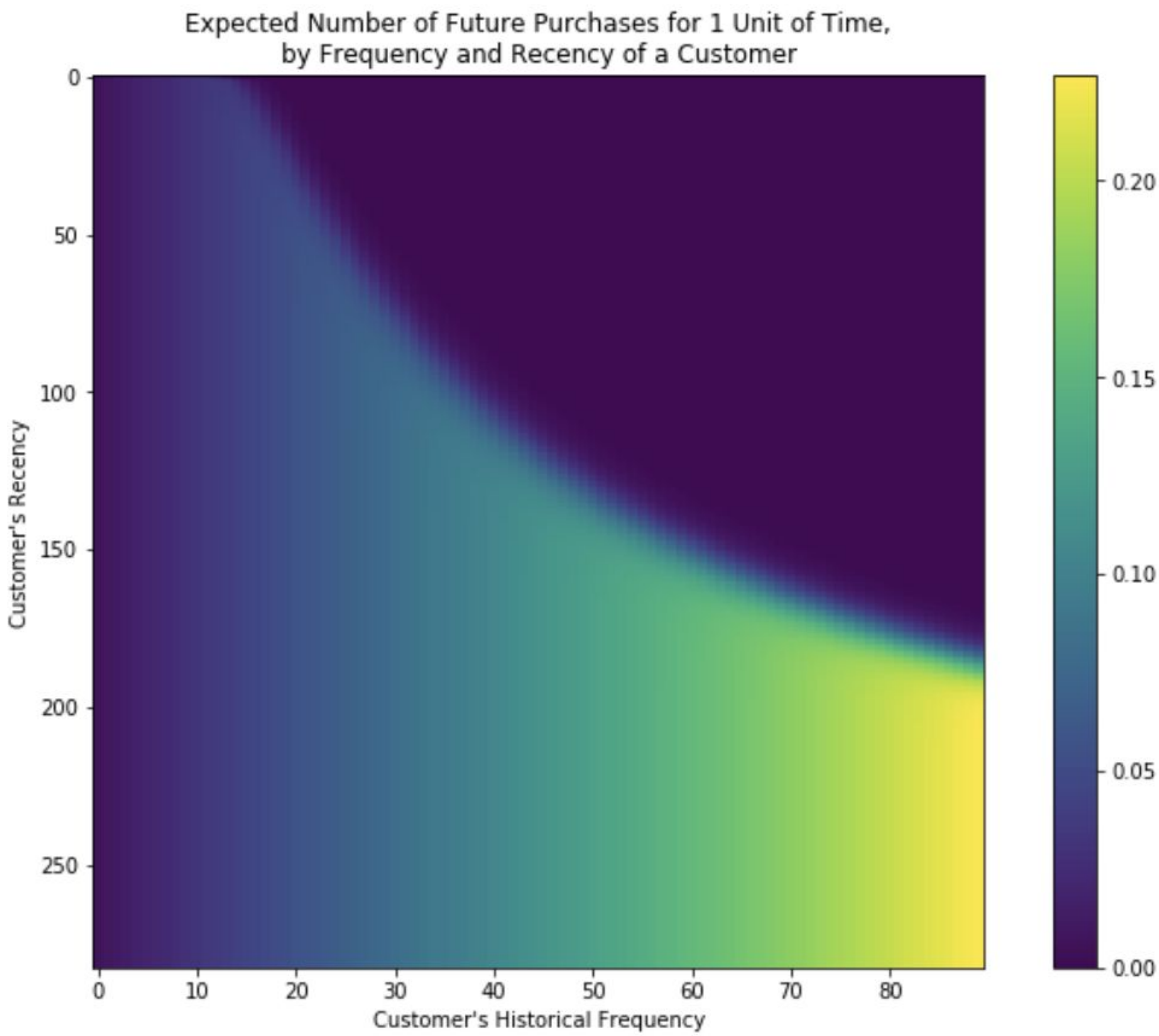
3.2 Predicted RFM Analysis

We have used RMSE to determine the accuracy of our CLTV model prediction. The Pareto/NBD model gives an RMSE of ~ $6435.87 while BG/NBD model gives an RMSE of ~ $6549.15.

Consider for instance a customer that has made a purchase every day for four weeks straight, and then is inactive for months. What are the chances he/she is still “alive”? The chances are pretty slim. On the other hand, a customer that historically made a purchase once a quarter, and again last quarter, is likely still alive. This can be visualized using the frequency/recency matrix, which computes the expected number of transactions an artificial customer is to make in the next time period, given his recency (age at last purchase) and frequency (the number of repeat transactions he has made).



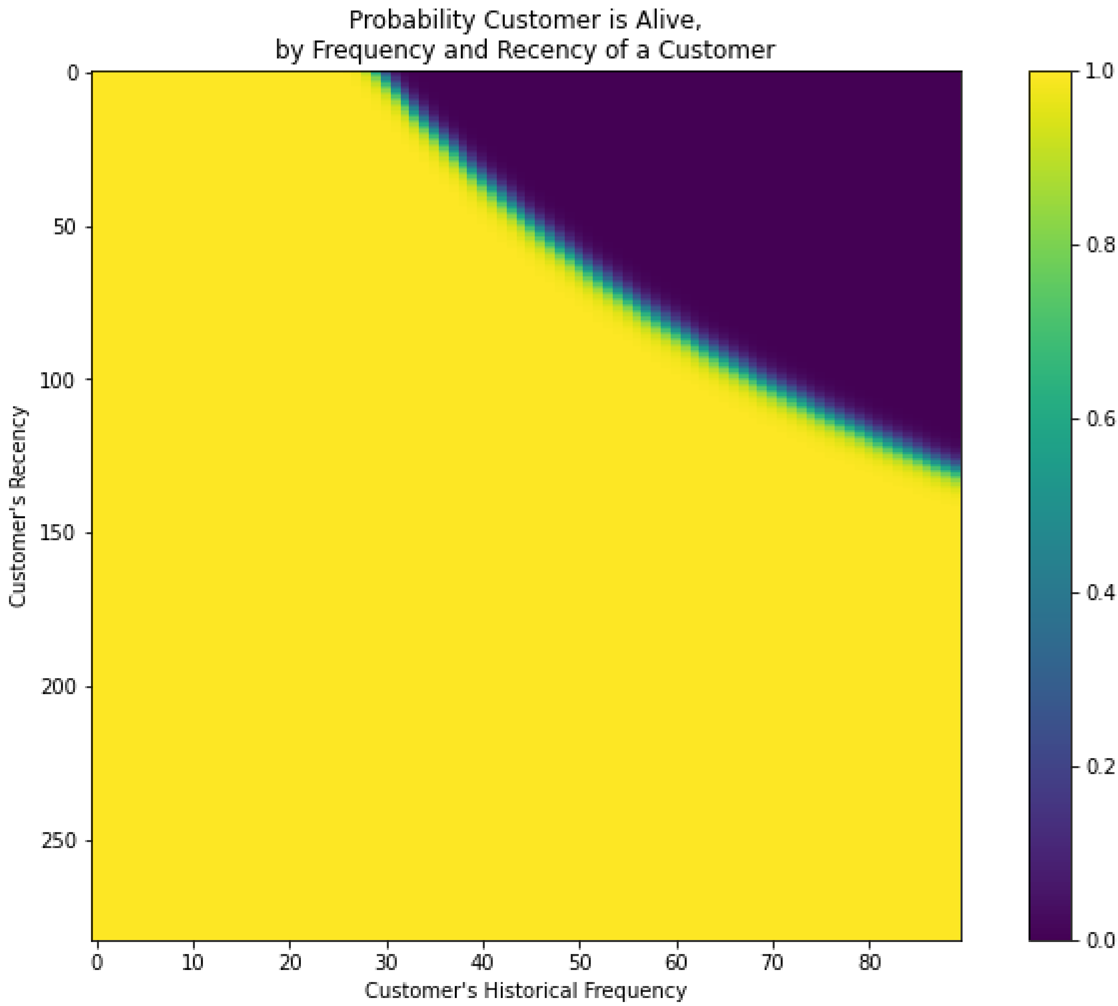
Using the BG/NBD Model



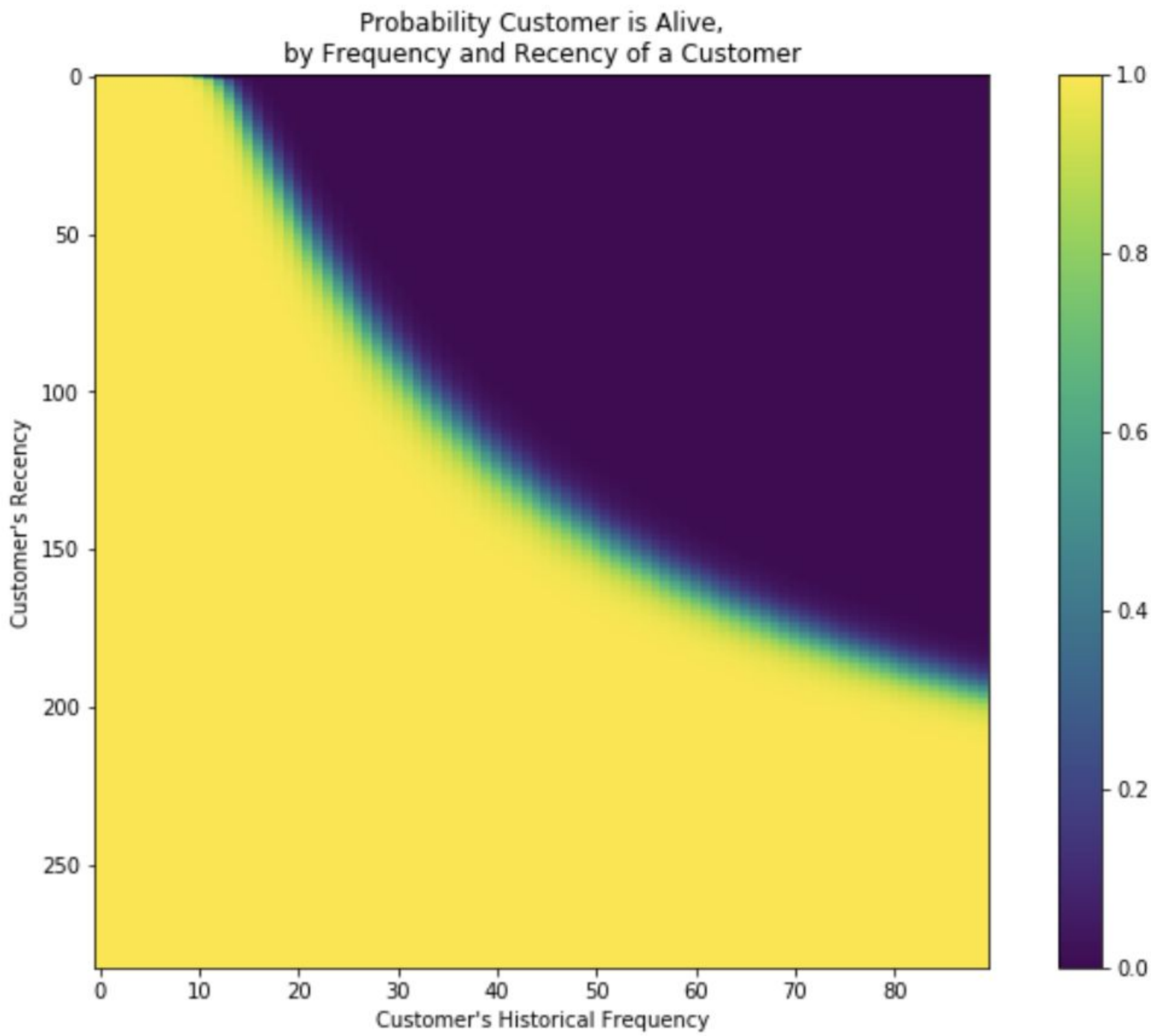
Using the Pareto/NBD Model

In the plot for BG/NBD model, it can be seen that if a customer has made 80 purchases, and his/her latest purchase was when he/she was approximately 125 days old (i.e. Recency: the duration between his first transaction and his latest transaction is 125 days), then he/she is our best customer (bottom-right). Customers who have purchased a lot and purchased recently will likely be the best customers in the future. Customers who have purchased a lot but not recently (top-right corner), have probably been lost. Similarly, we can do this analysis for the Pareto/NBD model.

We can predict which customers are most probable to be alive:



Using the BG/NBD Model



Using the Pareto/NBD Model

In the plot for BG/NBD model, it can be seen that customers who have purchased recently are almost surely “alive”. Customers who have purchased a lot but not recently, are likely to have dropped out. And the more they bought in the past, the more likely they have dropped out. They are represented in the upper-right. Similarly, we can do this analysis for the Pareto/NBD model.

We have successfully analyzed our models and validated them through their RMSE, which informs us of their CLTV value. We can form strategies according to our business objectives setting a threshold value for customer retention or not.

3.3 Customer Segmentation

One more analysis we can gain from our dataset is implementing the RFM value and segmenting our customers accordingly.

To achieve segmentation, we perform clustering on all 3 metrics individually - Recency, Frequency, and Monetary value. Our model is k-means and the optimum number of clusters obtained through elbow plot is 4. On these different clusters, we perform weighted sum and achieve an overall score

t6_cust['OverallScore'] = t6_cust['RecencyCluster'] + t6_cust['FrequencyCluster'] + t6_cust['RevenueCluster']
t6_cust.groupby('OverallScore')['recency','frequency_btyd','target_monetary'].mean()

Through this process, a total of 8 overall score clusters are obtained.

OverallScore	recency	frequency_btyd	target_monetary
0	49.709821	1.674107	1526.519241
1	182.867508	3.321767	2608.764385
2	116.100313	2.504702	1986.714796
3	232.633745	5.259259	3216.717984
4	215.850000	19.650000	8756.751500
5	256.089686	11.269058	7403.718206
6	263.500000	21.428571	54173.265000
7	269.000000	52.714286	89300.385714
8	275.500000	52.500000	194459.340000

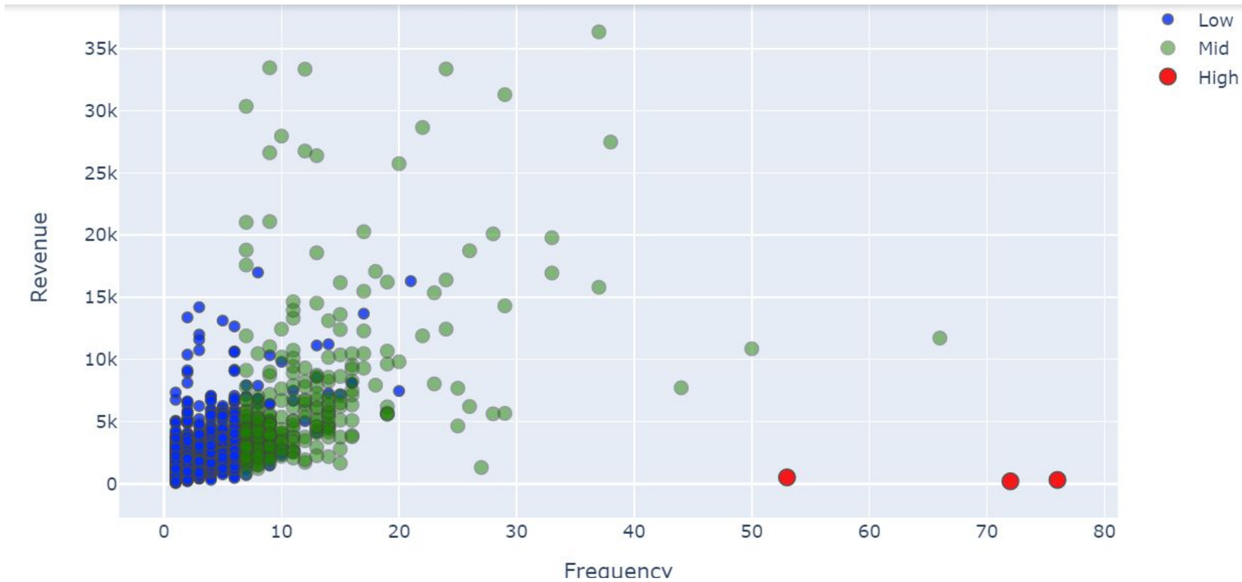
After analyzing the mean Recency, Frequency, and monetary values of these clusters, we can observe three major groups being formed. We will label then Low Value, Mid Value, and High-Value customers.

Customer	Overall Score
Low Value	0 - 3
Mid Value	4 - 6
High Value	7+

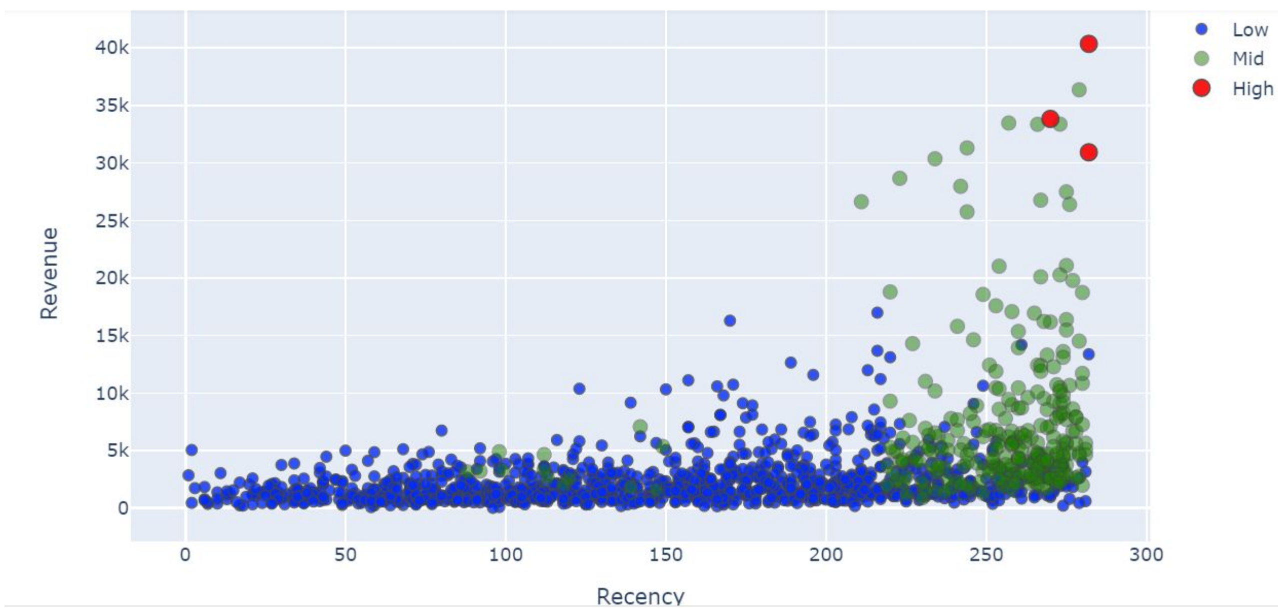
Since we have three groups, we want to strategize customer retention on their individual attributes as that would provide a higher retention value. Therefore, we want to understand where these customers lag i.e. whether they buy less frequently, or they buy low-value or less number of items, or they buy sporadically or have they not bought recently, and then plan to mitigate these issues.

We plot our customers against

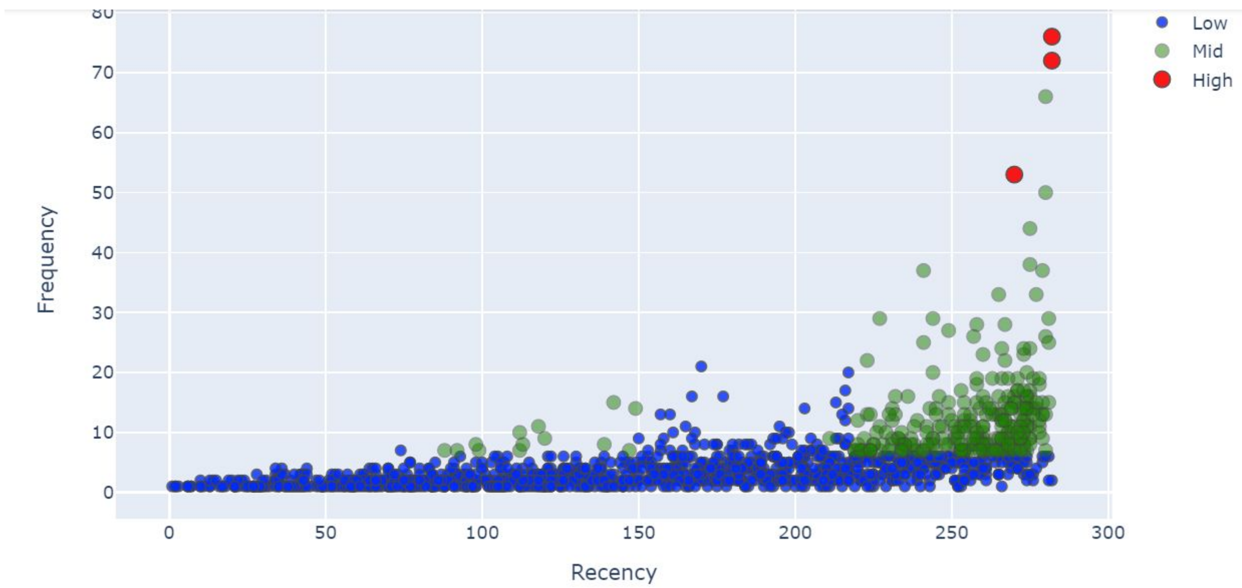
- Revenue with Frequency
- Revenue with Recency
- Recency with Frequency



Frequency vs Monetary



Recency vs Monetary



Recency vs Frequency

Our analysis of these plots are as follows -

- High-Value Customers (HVC)
    - Who tend to buy more frequently but their purchases are not very high revenue-focused, therefore we need to focus more on their revenue increase *(Revenue v/s Frequency)*
    - Who tend to buy high revenue items but haven't bought recently *(Revenue v/s Recency)*
    - They maintain their consistency, with high-frequency purchases but haven't purchased recently (*Frequency v/s Recency*)

- Mid Value Customers (MVC)
    - Few of them tend to buy less frequently but are high revenue buyers *(Revenue v/s Frequency)*
    - Majority of them have not bought recently *(Revenue v/s Recency)*
    - They clearly range from low to the high frequency with a few of them who have bought recently but have low numbers (*Frequency v/s Recency*)

- Low-Value Customers (LVC)
    - They don't buy very frequency and even when they do, it is low revenue purchases *(Revenue v/s Frequency)*
    - They clearly range considerably with respect to recency but staying consistent with low revenue purchases *(Revenue v/s Recency)*
    - They maintain their consistency with a low number of purchases but have a spread-out range with respect to recency (*Frequency v/s Recency*)

4. Marketing Strategy Recommendations, Limitations, and Future Research Directions

We have used the Pareto/NBD and BG/NBD models to predict the Customer Lifetime Value. Furthermore, we have performed customer segmentation on RFM values to get 3 major groups as mentioned and have analyzed them individually with respect to

- Revenue with Frequency
- Revenue with Recency
- Recency with Frequency

Now we create individual strategies for these brackets and sometimes different strategies within one major bracket. Our strategies based on our analyzed hypothesis is as follows

- For high-value customers, we know that these customers buy less frequency but are high revenue purchasers. They also haven't bought any products lately. Therefore we need to inquire if they are sleeping or dead customers. They could also be groups that purchase products based on certain period gaps, for instance, customers who shop based on season or customers who shop based on quarterly bonuses

- For our mid-value customers, they haven't bought recently but have a considerable range with respect to frequency and revenue. They could be potential brand loyalists. They could also be high revenue generators that either buy large quantities or have an affinity towards expensive items. We need to dig a little deeper into this but overall need to increase their recency.
- For our low value, we know that they buy less frequency and have low revenue purchases but they have sporadic buying patterns, which is an attribute the business could leverage and improve

A few recommendations that we have are :

1. Further segregating on bought items analysis - affinity to certain products, affordability, brand-conscious, etc. and sending out advertisements, emails, coupons on this analysis
2. For customers with sporadic buying patterns, offering options of buy now and pay later schemes or pay in installments
3. For customers that are potential brand loyalists, converting them to brand ambassadors would be cost-effective and highly profitable in the long run. This could be achieved through elite memberships, exclusive coupons, special offers
4. For customers who haven't bought lately, sending emails about upcoming sales, offering them exclusive offers for their continuous love towards the brand, and staying in touch through reminders. Continuous emails beyond a certain point might also alienate the customer so the number needs to be set on a combination of surveys, past experiences and domain knowledge

In the future, we can combine the customer segmentation model with the customer lifetime value model to predict the value of customers in each segment. This approach can also help us with segmented targeting i.e. identifying which of the inactive or sleeping customers should be targeted with a reactivation trigger, or which of the average customers can be pushed into the loyal customer category, etc. We can further perform a/b testing, surveying, etc. to understand the feasibility of our recommendations through quantitative metrics.

5. Appendix

Technologies Used

Link to Data

- http://archive.ics.uci.edu/ml/datasets/Online+Retail

Code

- https://colab.research.google.com/drive/1E2znAaldnEpfG8JtQzVPPxWcUNlFlGRv

References

- https://www.shopify.com/encyclopedia/customer-lifetime-value-clv
- https://www.crazyegg.com/blog/customer-lifetime-value/
- https://cran.r-project.org/web/packages/BTYD/vignettes/BTYD-walkthrough.pdf
- https://pypi.org/project/Lifetimes/