Customer Lifetime Value models for non-contractual businesses

MS BANA 8083 Summer Capstone

By: Utkarsh K. Singh (M13457333)

5/31/2020

1st Reader: Dr. Yichen Qin

Contents

Introduction.....2

Objective.....3

Data.....3

Exploratory Analysis.....4

- i) Orders per customer.....4
- ii) Average order value.....4
- iii) Purchase cycle.....5
- iv) RFM Metrics.....6

References:.....8

Introduction

Organizations are increasingly developing their arsenal of marketing tactics to drive customer centricity. This has led to an increased spending in marketing analytics and AI research to re-invent strategies for retention, acquisition and customer engagement. The CMO survey results report of February 2020 [1] projects an increase of 8% in annual marketing spend in the Retail Industry. It is currently estimated to be ~ 12% of the total revenue.

At the centre of marketing research is Campaign Budget Optimization (CBO), which focusses on redistributing budget to target the best performing audience. Customer Lifetime Value (CLV) is an important metric used by marketers to identify such an audience. It is defined as the net present value of all the future cash flows generated by the customer over his/her "lifetime". Typically, the "lifetime" duration is considered to be three years due to two reasons: (a) Product Lifecycle and, (b) 80% of profit comes in three years (Gupta and Lehman, 2006). However, this period can be defined as per marketing needs. In order to estimate CLV, the following questions need to be addressed:

- How many times is a customer expected to purchase in a given period in future?
- What will be the worth of these future purchases?
- What will be the total cost of making the sale?

For the scope of this project, the cost component is kept out of the computation due to data unavailability, and CLV is calculated based on expected revenue earned from customers. If data is available, the total cost can simply be subtracted from the expected revenue before adjusting for time value of money, without any change in the CLV modelling techniques.

In a non-contractual setting, a customer can be expected to drop out at any point in time. Therefore, a single definition of churn cannot be used to predict drop-out probability, which depends on, amongst other factors, the individual's own frequency of purchase and the time of the last purchase. This point is illustrated with the plot below:

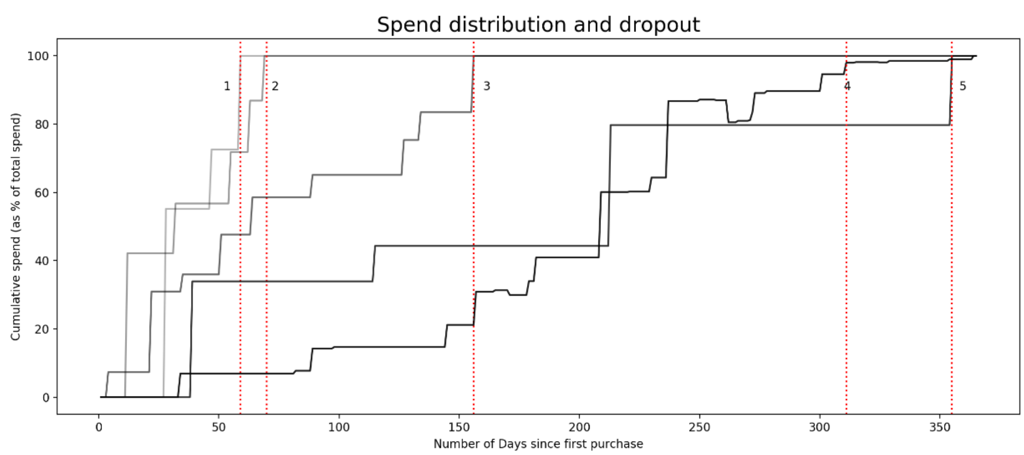


Figure 1: Illustrative: Spend patterns for customers

Customer 1 ceases activity within ~ 60 days from the first shop (cumulative spend reaches total spend), whereas customer 5 takes over 100 days within consecutive purchases and continues to stay alive till the 350th day. At any point beyond this timeline, the drop-out probability of customer 1 will be much higher than that of customer 5. The report discusses in detail the probabilistic techniques that account for these subtleties of customer buying pattern.

Objective

The objective of this project is to demonstrate the application of probabilistic and machine learning models to estimate future customer value. This will involve predicting the expected count and value of transactions at a customer level during a specified interval of time. The models are designed for a non-contractual business, and assume that customer response is unsolicited i.e. no sales triggers are used in the calibration or holdout period. The probabilistic models discussed account for the fact that a customer can drop out at any given point in time. The final report will also compare the performances of pure probabilistic and ML models with a hybrid modelling approach.

Data

Typically e-commerce transaction-level datasets are proprietary and consequently unavailable in public domain for free. However, The UCI Machine Learning Repository has collected this dataset containing actual transactions from 2010 and 2011 for a UK-based and registered non-store online retailer. The company mainly sells unique all-occasion gifts. A majority of customers are wholesalers and smaller B2C businesses. The dataset is available for public access and can be found on UCI’s website by the title "Online Retail".

This data contains over 25k transactions occurring between 01/12/2010 and 09/12/2011. Information on individual data fields is tabulated below:

Column name	Data type	Description	Non-missing values (% of total records)
InvoiceNo	Char	Unique identifier for a transaction	100
StockCode	Char	Unique identifier for an item in invoice	100
Description	Char	Item description	99.7
Quantity	Int	Item units purchased in invoice	100
InvoiceDate	Char	Date of purchase	100
UnitPrice	Float	Per unit cost of item	100
CustomerID	Float	Unique customer identifier	75
Country	Char	Customer location	100

Table 1: Data Columns description

Exploratory Analysis

i) Orders per customer

Probabilistic models can be used to model the expected number of orders in a given period for repeat buyers. 66% of the customers in the given dataset are repeat buyers. On an average, repeat purchasers have placed ~ 6 orders per customer, but the deviation is very high. The distribution for number of orders per customer, for repeat buyers, can be modelled by a Poisson family distribution (Colombo and Jiang, 1999).

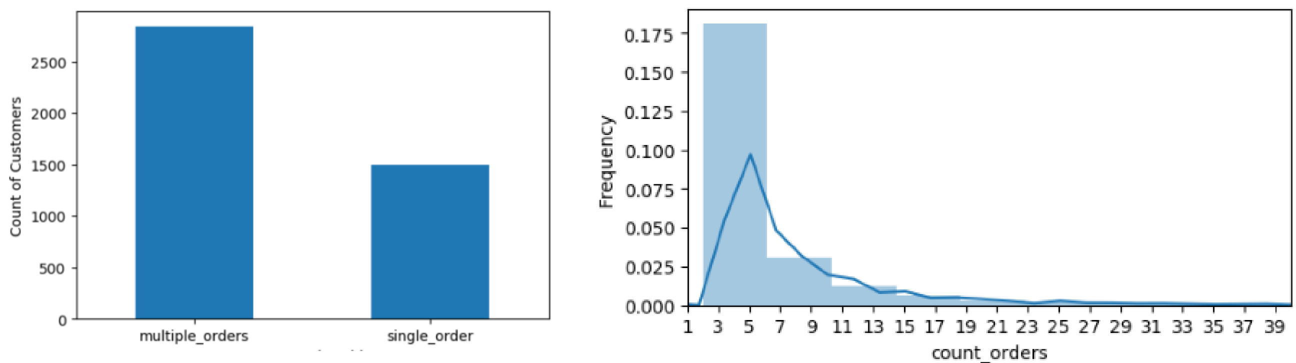


Figure 2: Distribution of orders per customer

Repeat buyers	
Average orders per customers	5.99
Std. Dev	9.05

ii) Average order value

Average order value for repeat buyers typically is in the range of [100,400] GBP (Please note that most customers are B2C businesses). Since spend data is bounded by 0 on the left and tends to be right skewed, probabilistic models typically use the gamma distribution for modelling spend (Colombo and Jiang, 1999).

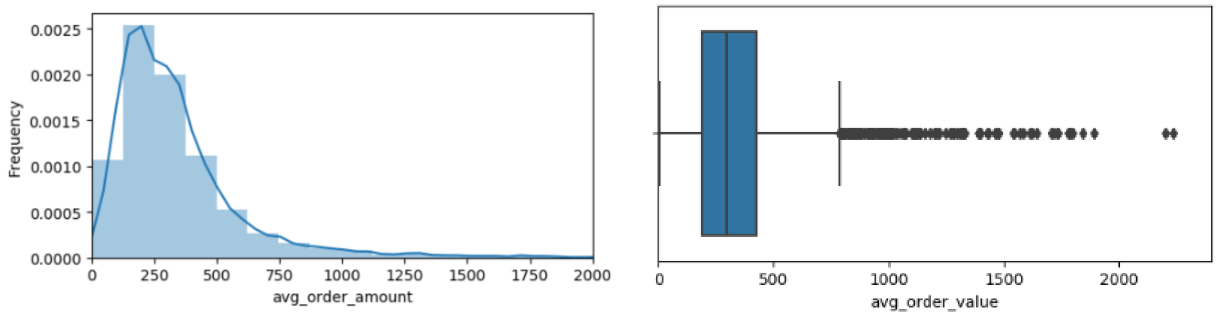


Figure 3: Average order value for customers

The plot below shows that the average value of transactions does not change by a high degree as customers make subsequent transactions. Higher average values are observed on the right end of the plot due to a small base of regular high-volume buyers (<10%) who go beyond 9 orders.

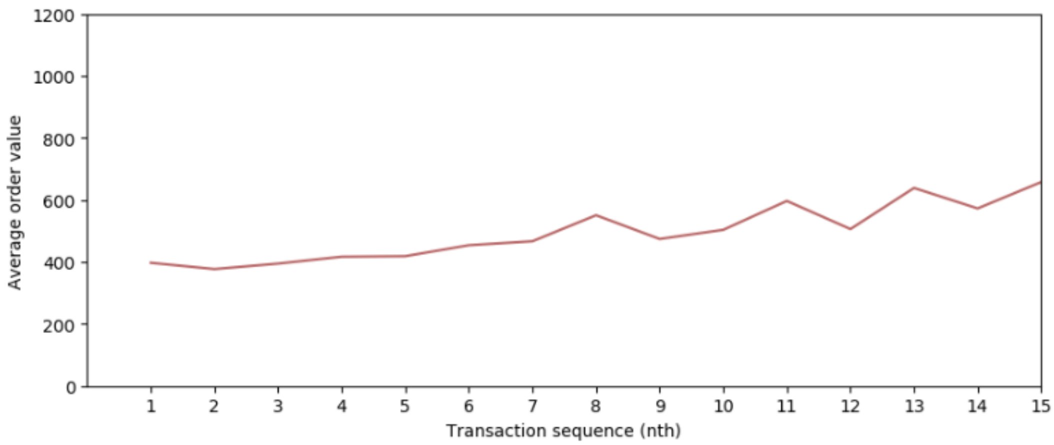


Figure 4: Average order value across subsequent transactions

iii) Purchase cycle

Typically customers place a subsequent order within 30-45 days from the previous purchase, and 75% of the population makes under 1.2 orders per month. The below table shows the statistics of the gap (in days) between orders at a customer level. The plot shows average number of orders per month across customer segments based on total transactions in the given data. Therefore, a period of 1 month can be taken as a suitable unit of time for further modelling and analysis.

Average Gap between orders (Days)	
Mean	45.6
Std dev	53.06
Median	28
3rd Quantile	58

Table 2: Summary - Avg. order gap

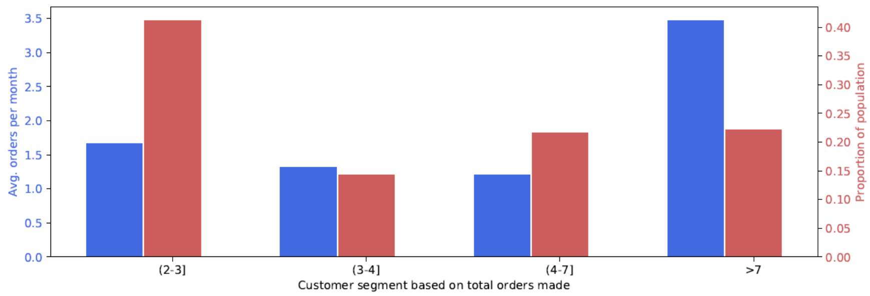


Table 3: Avg. orders per month

iv) RFM Metrics

Recency, Frequency and Monetary metrics are key inputs to predicting future purchases and drop-out probability of a customer at any point in time. Further sections of the report would discuss in detail the theory behind the use of RFM data for probabilistic models. In the context of CLV modelling, the RFM metrics are defined as follows:

a) Recency: Number of time periods elapsed (months as established in the previous section) between acquisition (1st purchase) and the last purchase in the caliberation period
b) Frequency: Number of repeat purchases in the caliberation period
c) Monetary: Avg. spend per transaction

50% of the customers have less than or equal to 1 repeat order in the data. Summary of Recency and Frequency metrics calculated for the entire dataset is tabulated below:

Frequency (# repeat orders)	
1st Quantile	0
Mean	3.2
Median	1
3rd Quantile	4

Recency	
1st Quantile	99
Mean	199.4
Median	208
3rd Quantile	300

Table 4: Recency and Frequency - Summary

The Monetary and Frequency metrics do not show strong linear correlation with each other as can be seen in the plot below. Recency and Frequency display non-linear correlation.

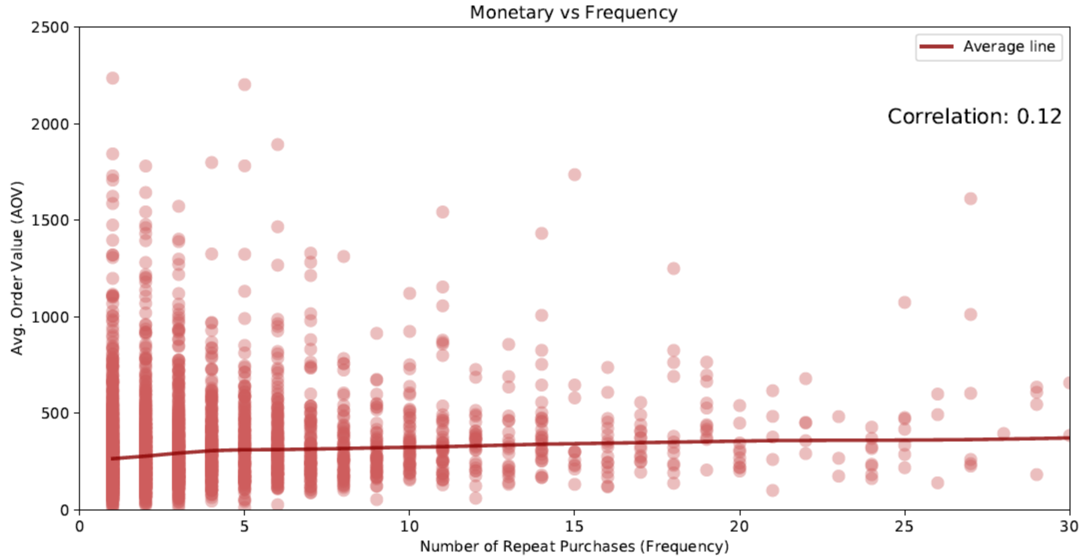
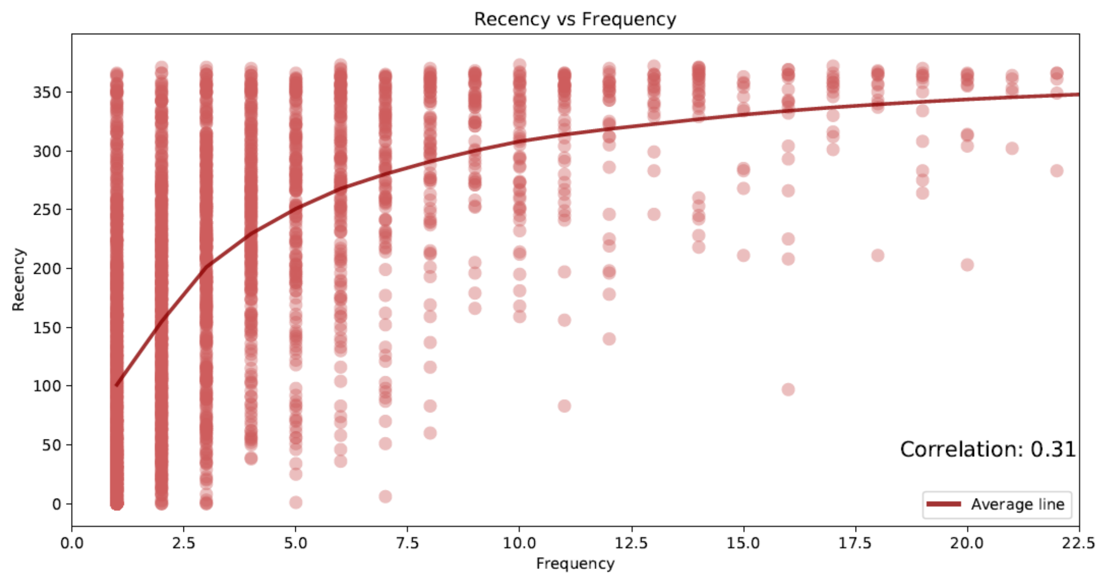


Figure 5: Monetary (AOV) vs Frequency correlation



Please note that higher recency values indicate that the customer has made a purchase more recently (with reference to 09/12/2011 as end date for the EDA). Customers with high frequency and low recency values have a higher drop-out probability.

References:

1. https://cmosurvey.org/results/
2. Sunil Gupta PhD & Donald R. Lehmann PhD (2006) Customer Lifetime Value and Firm Valuation, Journal of Relationship Marketing, 5:2-3, 87-110, DOI: 10.1300/J366v05n02_06
3. Richard Colombo, Weina Jiang (1999) A Stochastic RFM Model, Copyright John Wiley & Sons
4. https://archive.ics.uci.edu/ml/datasets/Online+Retail