Customer Lifetime Value & Customer segmentation

By: Anandha Sayanam S., Professional – Lean six sigma black belt and Data Scientist

1 Problem statement

The international on-line retail store wanted to know how they earn from each customer during their existence with them. Also, they wanted to work on different marketing strategies to increase and retain the customers.

Hence, the goal of the problem is to derive their customer life time value (CLV) and segment the customers based on their business value.

2 About the dataset

The international on-line retail store given their data based on the invoice transactions.

1. The excel file (online_retail_II.xlsx) has their customer transactions in terms of invoice number, stock number, description, quantity, price per unit, invoice date, customer ID and Country.
2. Data – has 8 features with 525,461 observtions.

3 Approach towards the problem

The following steps applied to get the relevant solution for the problem.

1. Understood the dataset and found that file has cancelled invoices indicated with prefix "C". So, dropped the cancelled invoices.
2. The customer lifetime value is calculated using the below formula to understand how much each customer provides business value.

   CLTV = (Customer_Value / Churn_Rate) x Profit_margin

   Customer_Value(CV) = Average_Order_Value * Purchase_Frequency

   Churn_Rate = 1 - Repeat_Rate

   Profit_margin – 10%

   Average_Order_Value = Total_Revenue / Total_Number_of_Orders

   Purchase_Frequency = Total_Number_of_Orders / Total_Number_of_Customers

   Churn_Rate = 1 - Repeat_Rate
3. The customers are segmented using K-Means algorithm and validated through Silhouette score and by using in-build function "qcut" in python.

3.1 Tools & Techniques

Tools	Python
Packages	Numby, Pandas, Matplotlib, MinMax Scaler, Standard Scaler, KMeans, silhouette_score

1. Identified the dataset has 8 key features like invoice number, stock number, description, quantity, price per unit, invoice date, customer ID and country.

2. Observed around 20% of the data (107,927 out off 525,461) has missing value. So, we decided to drop those based on the given business rule.
3. Also, dropped 9,839 observations since these were cancelled invoices.
4. Calculated the CLV based on the formula mentioned above and scaled using MinMax method.
5. Customer has been segmented based on their CLV values by two methods as
   a. Using in-build function "qcut"
   b. K-Means algorithm and validated through silhouette score.
6. The segmented customer were named as "Gold", "Silver", "Bronze" and "Lower Bronze".

4 Key observations

1. 90% of their customer were based out of UK, and few were in EIRE, Germany, France and Netherlands.
2. Their least customers were based out of Nigeria, Bahrain, Korea, West Indies and Brazil.
3. The customer segment based on the statistical function is as follows:

Segment	Transactions	Quantity
Gold	397,076	5,066,837
Silver	10,619	472,387

4. The customer segment based on the in-build function is as follows:

Segment	Transactions	Quantity
Gold	259,007	4,383,262
Silver	87,955	733,211
Bronze	42,239	305,016
Lower Bronze	18,494	117,735

5.

5 Proposed Solutions

1. The retailer can provide individual discounts on products and combo combination to move the lower segments to Gold.

==========End of the Document==========