**Customer Lifetime Value Modeling Using Python**

**Objectives**

In our internship project, we are given a dataset of Online retail sales of an organization. We are required to develop an optimized Customer Lifetime Value (CLV) model using

Probabilistic or Machine Learning or Deep Learning or Forecasting approach to increase customer experience, customer lifetime value and revenue. Customer lifetime value is the total worth to a business of a customer over the whole period of their relationship with this our model will help to increase customer experience, customer lifetime value and revenue.

**Attribute Information :-**

* InvoiceNo: Invoice number. Nominal, a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter ‘c’, it indicates a cancellation.
* StockCode: Product (item) code. Nominal, a 5-digit integral number uniquely assigned to each distinct product.
* Description: Product (item) name. Nominal.
* Quantity: The quantities of each product (item) per transaction. Numeric.
* InvoiceDate: Invoice Date and time. Numeric, the day and time when each transaction was generated.
* UnitPrice: Unit price. Numeric, Product price per unit in sterling.
* CustomerID: Customer number. Nominal, a 5-digit integral number uniquely assigned to each customer.
* Country: Country name. Nominal, the name of the country where each customer resides.

**Methodologies**

**Step 1:- To filter the contents based on the key assumptions that is preprocessing of the data**

**The first step is to filter the dataset based on key assumptions which are as follows :-**

* 2-3 years customers transaction data
* Take only positive sale transactions
* Take only last 1-year active customers
* To split the invoice data into data and time
* Further preprocessing steps will be added based on the need of the dataset which we will explore later.

**Step 2:- Exploratory Data Analysis (EDA)**

Exploratory Data Analysis (EDA) is an approach to analyze datasets to summarize their main characteristics, often with visual methods. EDA is used for seeing what the data can tell us before the modeling task. With the help of different graphs, we will be able to make analysis about the customers and also the ways to increase the customer experience which will be explored later.

**Step 3:- To increase the customer experience**

* Market basket analysis is a data mining strategy that merchants apply to increase profits by better understanding client buying habits. Large data sets, such purchase records, must be analyzed to identify product groups and those that are most likely to be bought together.
* We will use apriori algorithm to implement the concept of market basket analysis.
* We will find out the frequently bought items and also the items which are bought together to increase the customer experience.
* Association rule is used to find the relationship between different items. Association rules are used in market basket analysis to determine the possibility that two products will be bought together. By counting how often certain things occur together, association rules look for associations that happen far more frequently than they should.
* By using these concepts customers need not have to waste time in finding items hence helping customers save their time. This way customer experience can be enhanced.

**Step 4:- Customer lifetime value**

* Customer lifetime value is the total worth to a business of a customer over the whole period of their relationship. It’s an important metric as it costs less to keep existing customers than it does to acquire new ones, so increasing the value of your existing customers is a great way to drive growth. Knowing the CLV helps businesses develop strategies to acquire new customers and retain existing ones while maintaining profit margins.
* Customer lifetime value or CLV helps us understand how much a customer likes the product or services of a specific company thereby giving an opportunity to the company to understand and analyze their faults and shortcomings in a specific area and hence improving it later on. It basically distinguishes active customers from inactive ones as it is not an optimal choice to spend marketing budget on customers who are not profitable at all.
* Higher CLV implies more satisfied customers. Through high CLV customers, companies can generate more revenue during their lifetime rather than focusing on hiring new customers. So, CLV plays a vital role in increasing the revenue of the company and for achieving this goal we need to find ways to improve it.
* There are various ways to improve CLV. One is to create customer segments and the other one is to send out promotional offers, discounts and coupon codes. We will deal with customer segmentation in our project. Now what exactly is Customer segmentation. It is basically differentiating or segmenting the customers and finding who all are profitable and hence giving special discounts and offers to them thereby increasing the CLV. So we will segment the customers using two models:-
* **RFM clustering**
* **BG/ NBD nad Gamma Gamma model**
* **Finding CLV using RFM Clustering**

RFM (recency, frequency, monetary) is an effective customer segmentation technique where it will be very helpful for marketers, to make strategic choices in the business. It engages marketers to rapidly distinguish and segment customers into similar clusters and target them with separated and personalized promoting methodologies.

There are various steps involved in RFM Clustering which are as follows:-

Step 1 :- Find out recency, frequency and monetary value of customers.

Step 2 :- Divide them into clusters for each of RFM (group similar types of customers together). Give a score to customers for each of RFM.

Step 3 :- Get an overall score for each customer

Step 4 :- Overall score is based on recency, frequency and monetary value scores

Step 5 :- Label overall scores as High, Low and Mid CLVbased on the number of clusters obtained using python code

Step 6 :- Train and fit a machine learning model using different algorithms

Step 7 :- Predict the future CLV based on the ML model.

* **Finding CLV using BG/ NBD model and Gamma Gamma Model**

**Formula used :-**

CLV = Expected number of Transactions \* Revenue per Transaction \* Margin

* Expected number of transactions will be calculated using the BG / NBD model.
* Revenue per transaction will be calculated using the gamma-gamma model.
* Margin is proposed by the company but mostly we take 5%.
* BG / NBD model predicts whether the customer is alive that means active and has interacted with the company or are they dead means inactive.
* Model can predict the probability of a customer visiting a website for a product during a specific period of time as well as predict the expected number of transactions in future.
* A package called lifetimes will be used which contains function and metric to estimate CLV.
* Like the previous model, with this again we can find out the most loyal customer of the company as well as the ones who are on the verge of churning out.