**CUSTOMER LIFE CYCLE MANAGEMENT**

**Table of Contents**

**1.Problem Statement**

**2.Data Loading and Preprocessing**

**3.Exploratory Data Analysis EDA**

**4.Customer Segmentation using RFM Analysis**

**5.CLTV Calculation using Formulabased Approach**

**6.CLTV Prediction using Regression Modelling**

**7.CLTV Modelling using Probabilistic Models Lifetimes library**

**8.Conclusion and Results**

**1. Problem Statement**

The main objective of this project is to analyse customer behaviour and calculate Customer Lifetime Value CLTV for a Brazilian ecommerce company. By understanding the value of each customer, the business can make data driven decisions to optimize marketing strategies, improve customer retention, and maximize profitability. This project will explore different methods for calculating and predicting CLTV, including formula-based approaches, regression modelling, and probabilistic models.

**2. Data Loading and Preprocessing**

In this step, we loaded the necessary datasets: `customers`, `orders`, `payments`, and `order\_items`. We then merged these datasets into a single master dataframe to create a comprehensive view of each customer's transaction history. We also performed data cleaning and preprocessing tasks, such as converting date columns to the correct format, handling missing values, and engineering new features like `delivery\_against\_estimated`, `order\_purchase\_year`, and `order\_purchase\_month`.

**3. Exploratory Data Analysis EDA**

The EDA phase involved visualizing the data to gain insights into customer behaviour and sales trends. We analysed:

Daily and monthly revenue: To understand the overall sales performance over time.

Monthly active customers: To track customer engagement.

Monthly order count: To monitor the volume of transactions.

Average revenue per order: To assess the value of each transaction.

New vs. repeat customer revenue: To understand the contribution of different customer segments to the total revenue.

Repeat customer ratio and retention rate: To measure customer loyalty.

**4. Customer Segmentation using RFM Analysis**

We performed customer segmentation using the RFM Recency, Frequency, Monetary framework. This involved:

Calculating Recency, Frequency, and Monetary values for each customer.

Using K-Means clustering to group customers into distinct segments based on their RFM scores.

Analysing the characteristics of each segment to understand their behaviour and value to the business.

**5. CLTV Calculation using Formula-based Approach**

We calculated CLTV using a formula-based approach. This method provides a simple and intuitive way to estimate the value of each customer based on their past purchasing behaviour. The formula we used is:

CLTV = Average Order Value Purchase Frequency / Churn Rate

**6. CLTV Prediction using Regression Modelling**

We framed the CLTV prediction problem as a regression task. We used the past 12 months of sales data as features to predict the total purchase value of each customer. We experimented with different approaches, such as including only repeat customers and customers with a minimum number of transactions, to improve the model's performance.

**7. CLTV Modelling using Probabilistic Models Lifetimes library**

We used the `lifetimes` library to build probabilistic models for CLTV prediction. This library provides advanced models like Beta Geometric/Negative Binomial Distribution BG/NBD and Gamma Gamma to model customer purchasing behaviour.

**BG/NBD Model:** We used this model to predict the number of future transactions for each customer.

**Gamma Gamma Model:** We used this model to predict the average monetary value of each customer's transactions.

**CLTV Calculation:** We combined the predictions from both models to calculate the final CLTV for each customer.

**8. Conclusion and Results**

This project provided a comprehensive analysis of customer lifetime value for a Brazilian ecommerce company. We explored various methods for calculating and predicting CLTV, each with its own strengths and weaknesses.

RFM analysis helped us to segment customers into different groups based on their purchasing behaviour, allowing for targeted marketing campaigns.

The formula based approach provided a quick and easy way to estimate CLTV, but it may not be accurate for all customers.

Regression modelling allowed us to predict CLTV based on past purchasing behaviour, but the model's performance was limited by the small number of repeat customers.

Probabilistic models from the `lifetimes` library provided a more sophisticated approach to CLTV modelling, taking into account the stochastic nature of customer purchasing behaviour.

By using these insights, the business can make more informed decisions about customer acquisition, retention, and marketing strategies, ultimately leading to increased profitability and longterm growth.