Verdict and Overview

In this project, we undertook the task of segmenting customers based on several features using clustering techniques. The goal was to identify distinct customer segments that could be leveraged to improve marketing strategies. We employed K-means clustering, using methods like the elbow method and silhouette analysis to determine the optimal number of clusters (k). The best segmentation was achieved with K-means clustering with k=3, which was validated through Principal Component Analysis (PCA) and t-SNE visualizations (see appendix).

Analysis Description

We began our analysis by preprocessing the data and selecting the relevant features for clustering. The features chosen for segmentation were Age, Average unit cost per customer, Recency, Frequency, and Customer Lifetime Value (CLV).

- *Recency:* This metric was calculated as the number of days since the last purchase.
- *Frequency:* This was determined by counting the number of transactions made by each customer.
- *Customer Lifetime Value (CLV):* This was calculated as the total revenue generated by each customer, derived from the initial dataset.

We then applied the K-means clustering algorithm, which partitions the data into k clusters, each represented by the centroid of the cluster. To determine the optimal number of clusters, we used the elbow method and silhouette analysis.

- *Elbow Method:* The elbow method involves plotting the within-cluster sum of squares (WCSS) against the number of clusters. The point where the curve starts to flatten (forming an 'elbow') indicates the optimal number of clusters.
- *Silhouette Analysis:* Silhouette analysis measures how similar each point is to its own cluster compared to other clusters. A higher silhouette score indicates better-defined clusters.

For our dataset, both methods suggested that k=3 was the optimal number of clusters.

Insights Identified

By segmenting the customers into three clusters, we gained several key insights:

Cluster 1: This cluster consisted of customers with high frequency and high CLV, indicating a group of loyal and high-value customers.

Cluster 2: This cluster had customers with moderate frequency and CLV, representing a group of regular customers.

Cluster 3: This cluster included customers with low frequency and CLV, identifying them as low-value or infrequent customers.

These clusters were clearly visible in the PCA and t-SNE visualizations, which reduced the dimensionality of the data and provided a clear distinction between the clusters. *(fig. 1-2)*

Best Method for Determining clusters

Through our analysis, the combination of the elbow method and silhouette analysis provided a robust means of determining the optimal number of clusters. Both methods consistently pointed to k=3, and this was further supported by the clarity of the clusters in the PCA and t-SNE visualizations.

Recommendations for Improving Marketing Efficiency

With the customer segments identified, we can tailor our marketing strategies to each specific group to improve efficiency and effectiveness.

For high-value customers in Cluster 1, implementing loyalty programs and exclusive offers will help retain them, while personalized marketing can increase their engagement.

Regular customers in Cluster 2, who buy more frequently but at a lower cost per purchase, can benefit from upselling and cross-selling opportunities. Additionally, engagement campaigns can be designed to boost their purchase frequency and transition them into the high-value segment.

Low-value customers in Cluster 3 should be targeted with reactivation campaigns to re-engage them and offered discounts or incentives to prompt more frequent purchases and enhance loyalty. By leveraging these clusters, the company can allocate marketing resources more effectively, ensuring that each customer group receives the most relevant and impactful marketing efforts. This targeted approach will likely lead to increased customer satisfaction, higher retention rates, and ultimately, improved revenue.

APPENDIX *(fig. 1-2)*