**JIBU**

**Optimizing Customer Lifetime Value**

**Abstract**

**This report examines the optimisation of Customer Lifetime Value (CLV) at Jibu, highlighting the use of data analytics to enhance customer satisfaction, and market share in the competitive water supply sector in Africa.**

**Khushboo Surana**

**Table of Contents**

[**Introduction to Jibu 3**](#_sgum05fxcr3i)

[Addressing Systemic Challenges 3](#_ehxmpuc4llm5)

[Market and Competitive Landscape 3](#_13e6ow136iib)

[Monthly Meters Reading Dataset 3](#_ubkr36xidh81)

[Health Dataset 4](#_4qw4vkvdvya9)

[**Situation Analysis 4**](#_vm6xgr54iy7i)

[Meters Reading data [Figure 1, 2 and 3 in Appendix] 4](#_ordiso30b51e)

[RFM data [Figure 5] 4](#_biw4i9pyk84r)

[Health data 4](#_uptj1p9wh3bu)

[The Problem Statement: Optimizing Customer Lifetime Value 5](#_2fz92g6r74th)

[RFM Analysis and Customer Clustering 5](#_1d5ay5uobqc4)

[Calculation of RFM Metrics 5](#_d6j2gqxim6vw)

[Data Transformation and Visualization 5](#_xhvlc9xtj1sp)

[Analysis of Data Distribution 5](#_l6fnytrz1gh9)

[Modelling Techniques and Assessment 5](#_10468xi0cmme)

[Clustering 5](#_cm4wxhft4box)

[**CLV Calculation and Analysis: 6**](#_d3jmpugmzktt)

[CLV Calculation and Cluster Mapping 6](#_rikjg9jyl0y)

[Summary [Figure 9] 7](#_krbvcmf1w9lh)

[**Regression of CLV and Feature Selection 7**](#_421dwf8wgk9s)

[**Conclusions and Recommendations 7**](#_r0ljuye1gq45)

[**Appendix 9**](#_tphw3aosaca8)

# 

# 

# Introduction to Jibu

Founded with a vision in the United States, Jibu is a pioneering company at the forefront of revolutionizing the access to clean, safe, and locally-sourced drinking water in various regions across Africa, including Rwanda, Uganda, Kenya, Tanzania, DR Congo, Burundi, Zambia, and Ghana. More than just a water supplier, Jibu stands as a pillar of opportunity within the communities it reaches. By providing seed financing for business-in-a-box franchises, Jibu empowers local entrepreneurs to meet the critical need for affordable and accessible drinking water in underserved areas. This innovative model extends beyond water sales, aiming to establish sustainable business opportunities that encourage local ownership and yield long-term benefits for communities.

## Addressing Systemic Challenges

Aware of the fact that approximately half of all water projects in developing countries fail within two years due to a lack of local engagement and proper incentives, Jibu directly confronts these systemic challenges. Through enabling local entrepreneurs to establish their own safe drinking water franchises, Jibu ensures its initiatives are deeply embedded within communities, significantly enhancing the prospects of success and sustainability. This strategy not only tackles the pressing global issue of water accessibility but also fosters entrepreneurship and generates employment opportunities in emerging markets. In the competitive realm of water supply, Jibu is faced with the crucial task of effectively distributing its marketing and operational resources to maximize the long-term value of its customer relationships.

## Market and Competitive Landscape

Many NGO, government agencies, and international aid organizations have collaborated to provide sustainable water solutions, emphasizing community engagement and capacity-building. Similar to Jibu, there are other companies which provide safe drinking water in Africa. Some of these are SolarDew, OpenVersum, Nazava Water Filters, Boon, PakVitae and Oasis Water. Oasis Water, SolarDew, and Openversum share a common goal of providing clean and safe drinking water solutions to communities. They all emphasize innovation, sustainability, and entrepreneurial opportunities in their respective approaches to addressing water purification challenges. Oasis Water primarily serves the retail and franchise sectors in South Africa, offering refill water services, bottled water, and water dispenser rentals. In contrast, SolarDew focuses on global water scarcity issues, leveraging solar energy for desalination systems. Openversum, based in Europe, specializes in membrane-based water filtration technology and micro-franchise initiatives, targeting a diverse range of markets including households, workplaces, and educational institutions.

**Data Cleaning and Merging**

## Monthly Meters Reading Dataset

1. **Clean Names Column:** Standardizes column names by converting to lowercase, removing special characters, and replacing spaces with underscores for consistency.
2. **Change Month Dates to 25th of the Month:** Uniformly sets all month dates to the 25th to ensure consistent reporting and ease of analysis.
3. **Data for Rolling Three Years:** Filters dataset to include only records from December 2020 to November 2023 for trend analysis over a recent period.
4. **Filter Data for Selected Countries and Positive Water Production**: Retains data entries from Kenya, Rwanda, Uganda, and Goma while excluding negative water production values for meaningful analysis.

## Health Dataset

1. Clean **franchise names** to ensure consistency and accuracy in analysis.
2. Convert **sales and expenses** to USD to facilitate uniform currency comparison.
3. Perform outlier elimination on **hygiene audit score, P&Q score, retailer satisfaction score, and retailer branding score** to enhance data reliability and model performance.
4. Merged both files using Franchise ID

# Situation Analysis

## Meters Reading data [Figure 1, 2 and 3 in Appendix]

* The average number of liters produced per month across all franchises is approximately 66,669, with a median of 53,823 liters.
* There is notable variation in production levels, with minimum production as low as 20 liters and maximum production as high as 386,162 liters.
* On average, franchises produce around 2,564 liters per day, with considerable variation observed in daily production levels.
* Frequency of water produced is highest in Rwanda contributing 44% of total followed by Uganda contributing 26%.

## RFM data [Figure 5]

* The median recency value is 1, indicating that a significant portion of franchises made purchases recently.
* The mean recency value is slightly higher at 2.026, suggesting a slight skew towards more recent purchases.
* The median frequency is 35, indicating that many franchises produce water frequently.
* The mean frequency is 29.39, suggesting a generally high level of water production activity across franchises.
* Monetary value and Frequency shows high correlation with a coefficient of 0.56

## Health data

* Retailer satisfaction scores range from 0 to 65.222, with a median score of 6.583 and a mean score of 11.439, indicating a generally positive satisfaction level among retailers.
* Retailer branding scores vary from 0 to 85.00, with a median score of 31.44 and a mean score of 35.11, suggesting a moderate to high level of branding effectiveness.
* Adjusted expenses range from 0 to $817,105, with a median value of $3,900 and a mean value of $38,047, indicating significant variability in expenditure levels across franchises.

## The Problem Statement: Optimizing Customer Lifetime Value

Despite Jibu's noble mission and innovative operational model, the company confronts the significant challenge of optimizing Customer Lifetime Value (CLV). There exists a vast potential to enrich the understanding of Jibu's diverse customer bases and tailor strategies to meet their specific needs, an opportunity that remains largely untapped. Enhancing CLV is essential for Jibu to efficiently allocate its marketing efforts, customize its product offerings, and formulate effective customer retention strategies. The present underutilization of real-time data on water volumes, customer locations, and franchise performance restricts Jibu's ability to calculate and improve CLV. This limitation not only impedes Jibu's capacity to maximize revenue from its existing customer base but also affects its strategic allocation of customer acquisition resources. By addressing this challenge and developing a deeper comprehension of its customers' lifetime value, Jibu can refine its marketing approaches, enhance customer satisfaction, and significantly increase its market share in the competitive landscape.

# RFM Analysis and Customer Clustering

RFM analysis stands as a pivotal technique in marketing analytics, facilitating the segmentation of customers based on their transaction history.

## Calculation of RFM Metrics

Aggregate data by franchise\_id to calculate Recency (time since last transaction), Frequency (count of transactions), and Monetary Value (total liters produced).

## Data Transformation and Visualization

Convert RFM metrics to numeric form and use correlation matrices for visual analysis of the relationships between the metrics.

## Analysis of Data Distribution

Assess skewness of RFM metrics using histograms and density plots, applying logarithmic transformations where needed to normalize distributions.

## Modelling Techniques and Assessment

## Clustering

Segment customers using K-means clustering to identify distinct customer segments for targeted marketing strategies. The clustering process unfolds as follows: Identification of Optimal Cluster Count employing both silhouette [Figure 6] and within-cluster sum of squares (wss) [Figure 7] methods, a rigorous evaluation ensues to discern the optimal number of clusters as 8 clusters.

Implementation of K-means Clustering using the standardized RFM data segregates franchises into clusters characterized by analogous RFM profiles. Subsequent to the assignment of cluster labels to each franchise, the computation of cluster centers furnishes further insights into the distinctive attributes of each segment [Figure 8].

# CLV Calculation and Analysis:

**Data Chronology and Time Indexing**

* **Sorting by Date:** The dataset mydata2 is arranged in chronological order based on the mm\_yy column to ensure accurate analysis.
* **MonthPresent Calculation:** We calculate the MonthPresent for each entry as the number of months since a start\_date (2020-11-25), adding 1 to start the count from 1 instead of 0.
* **Interest Rates Definition:** We created an interest\_rates vector, mapping specific dates to their respective interest rates. This step is crucial for assigning the correct interest rate to each data entry based on its date.
* **Mapping Interest Rates:** We map these interest rates to the corresponding mm\_yy dates in mydata2, storing them in a new column called USInterestRate.

## CLV Calculation and Cluster Mapping

With the USInterestRate and MonthPresent calculated, we proceed to calculate the CLV for each entry. The formula used here takes into account the amount of liters\_produced and adjusts it by the compounded effect of the interest rate over the number of months since the start date. This step yields a time-adjusted value of the liters produced, representing the CLV.

1. **CLV is calculated using this following formula:**

**CLV <- liters\_produced / ((1 + USInterestRatebyMonth) ^ $MonthPresent)**

1. **Aggregating CLV by Franchise:** We then group the data by franchise\_id and summarize it to compute the total CLV (TotalCLV) and count of entries (count) for each franchise. This aggregation provides a comprehensive view of each franchise's value over time.
2. **Joining Total CLV:** The total CLV for each franchise, along with the count, is then joined back to the original dataset. We then have mydata2 with the total CLV value for each franchise, facilitating further analysis.
3. **Merging with Cluster Labels:** The dataset df\_rfm\_log is updated with cluster labels (cluster\_labels), and these labels are merged with df\_clv and an RFM clustering dataset (rfm\_clust) to produce df\_final. This step categorizes franchises into clusters based on similar characteristics.
4. **Average CLV by Cluster:** The script calculates the average TotalCLV for each cluster by grouping df\_final by Cluster and computing the mean of TotalCLV. This provides insights into the average value of franchises within each cluster, highlighting which clusters are more valuable on average.
5. **Merge with RFM Values:** Finally, this average CLV is merged with a summary of RFM clustering metrics (rfm\_clust\_summary), combining the financial valuation (CLV) with behavior-based segmentation (RFM). This comprehensive dataset (final\_cluster\_summary) offers a view of the customer base, blending their lifetime value with their recency, frequency, and monetary behaviors.

## Summary [Figure 9]

These clusters help the business to understand various segments of their franchise operations. Here’s what the business might infer:

* Clusters with higher average Total CLV (like 2, 3, and 8) are very valuable and consist of franchises with high sales volume, repeat purchases over short periods. 77% of franchises are in these clusters.
* Clusters with lower counts (like 4, 6, and 7) could be niche franchises or new ones that haven't had the time to build up their customer base or sales volume.
* Clusters with lower Total CLV (like 1, 5, and 7) might need strategies to boost their sales.

The business can use these insights to tailor specific strategies for each cluster. For example, they might focus on customer retention programs for clusters with high CLV or devise ways to increase the transaction frequency and monetary value for those with lower scores.

# Regression of CLV and Feature Selection

We aim to identify which factors are significant predictors of CLV, thereby providing actionable insights for strategic business decisions to enhance customer value. Variables showing significant multicollinearity were excluded to ensure the model's interpretability and accuracy. The final model focused on variables that were both statistically significant and offered practical insights into how operational adjustments could influence CLV.

The built model has an adjusted R-squared of 85% with statistically significant features that can be specified in a general linear regression form as follows: **CLV = 13870 + 0.51 (Liters Produced) + 56.47 (Hygiene Audit Score) -174.18 (PQ Audit Score) + 1679 (Staff) -93.80 (Retailer Satisfaction Score) + 0.01 (Adjusted Sales) + ϵ**

**Findings and Implications** The regression analysis revealed key predictors of CLV, highlighting the importance of staff and product quality audits, which directly impact customer satisfaction and loyalty. The positive association between staff count and CLV underscores the value of investment in human resources to enhance customer service. Retailer Satisfaction Score has a significant impact on Adjusted Sales. There is also some impact observed from the Hygiene Audit Score, although it might not be as pronounced as the effect of Retailer Satisfaction Score on sales. This insight underscores the importance of maintaining high levels of retailer satisfaction to drive sales performance within the franchise network.

# Conclusions and Recommendations

**Understanding Dual Customer Relationships:** Jibu operates within a unique framework where it serves two distinct sets of customers: franchise owners and end-users. Franchise owners directly engage with Jibu as customers, while end-users, who consume the water provided by the franchises, represent the ultimate beneficiaries of Jibu's services. This understanding is crucial for both high and low CLV franchises, as catering to the needs of both sets of customers is essential for sustained growth and profitability.

**Operational Focus on Product Quality and Customer Service:** Jibu should prioritize maintaining high standards of product quality and customer service to satisfy both franchise owners and end-users. Regular audits should be conducted to ensure consistency in product quality, benefiting both sets of customers. High CLV franchises can capitalize on their existing customer base by maintaining superior product quality and enhancing customer service, leading to increased loyalty and profitability. Low CLV franchises, on the other hand, can leverage improvements in product quality and customer service to attract and retain more customers, thereby boosting their CLV over time.

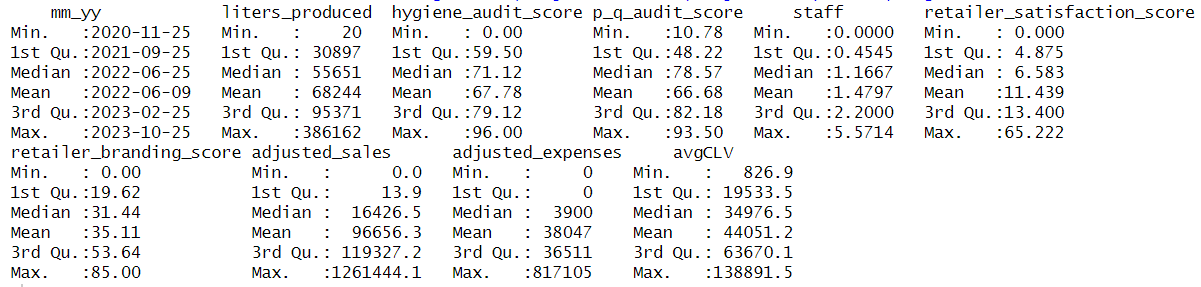
**Strategic Marketing Initiatives:** Tailored marketing campaigns should emphasize product quality, customer service, and brand reputation to attract new franchise owners and end-users alike. Highlighting these aspects in marketing communications can reinforce loyalty among existing franchise owners and build trust among end-users, ultimately driving long-term value for the business. High CLV franchises can utilize strategic marketing initiatives to solidify their position in the market and further enhance customer loyalty. Meanwhile, low CLV franchises can leverage targeted marketing efforts to attract new customers and improve their CLV.

**Continuous Monitoring and Adaptation:** Continuous monitoring of key metrics related to both franchise owners and end-users is essential for Jibu to adapt to evolving market dynamics and changing consumer preferences. By staying agile and adaptive, Jibu can ensure sustained growth and profitability while effectively serving the needs of both customer segments. High CLV franchises can benefit from continuous monitoring by identifying areas for further improvement and capitalizing on emerging opportunities in the market. Similarly, low CLV franchises can improve these key metrics to refine their strategies and enhance their competitiveness, which will drive higher CLV over time.

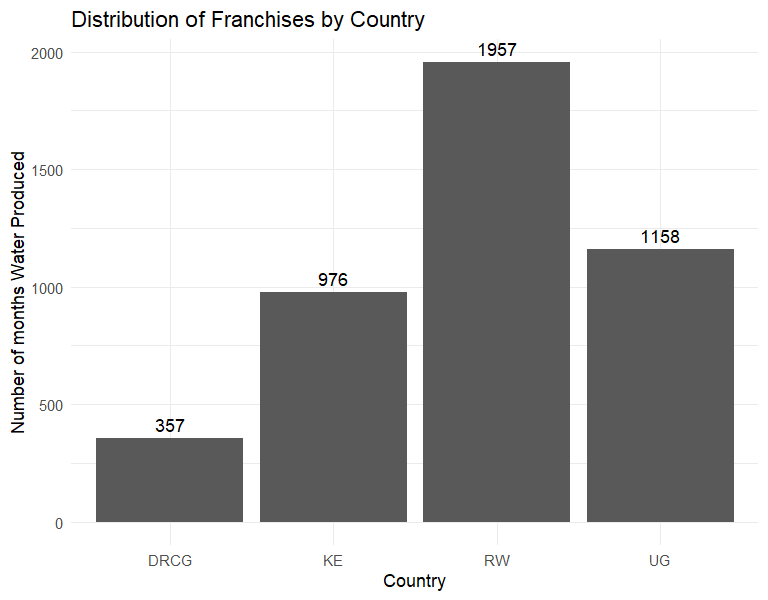
**Investment in Franchisee Relationships:** Fostering strong relationships with franchise owners is paramount. Jibu should involve franchise owners in decision-making processes, solicit their feedback on product quality and customer service initiatives, and provide them with the necessary support and resources to succeed which will enhance their operational efficiency and end-user satisfaction. High CLV franchises can benefit from strengthened franchisee relationships by leveraging their expertise and insights to drive further growth and innovation. For low CLV franchises, investing in franchisee relationships can be instrumental in overcoming operational challenges, unlocking untapped potential and learning from high CLV franchises, leading to improved CLV and long-term sustainability.

# Appendix

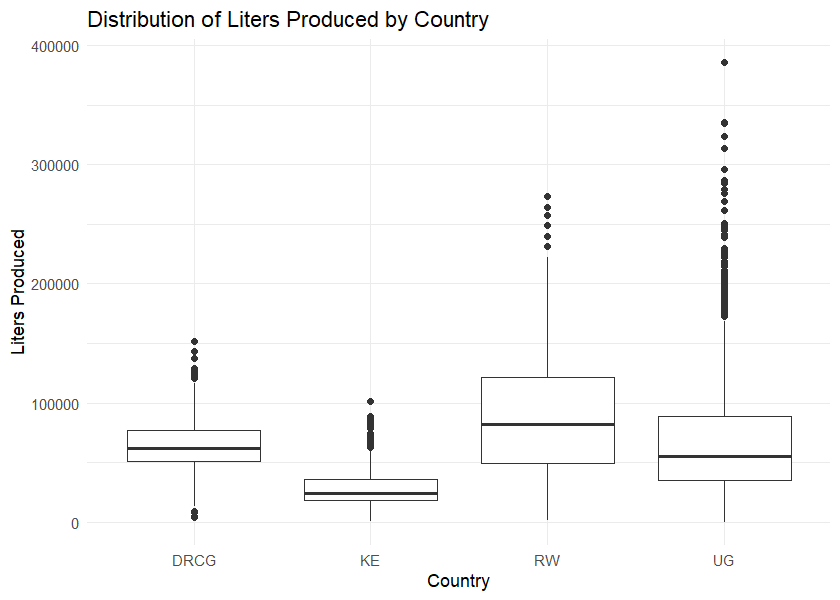
**Figure 1: Data Summary**



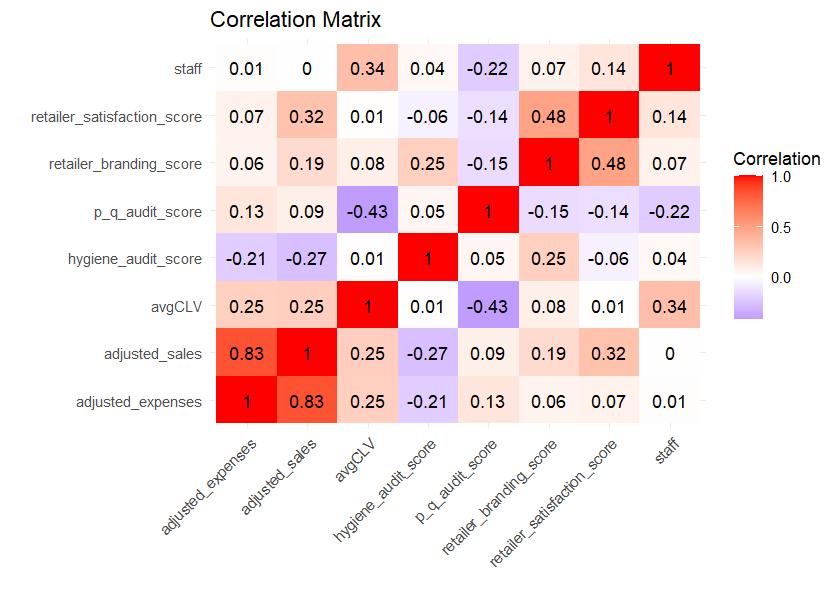
**Figure 2: Distribution of frequency by country**



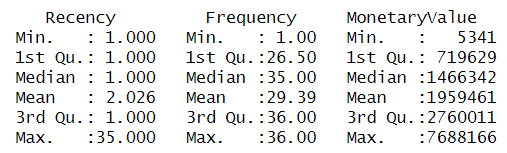
**Figure 3: Distribution of liters produced by country**



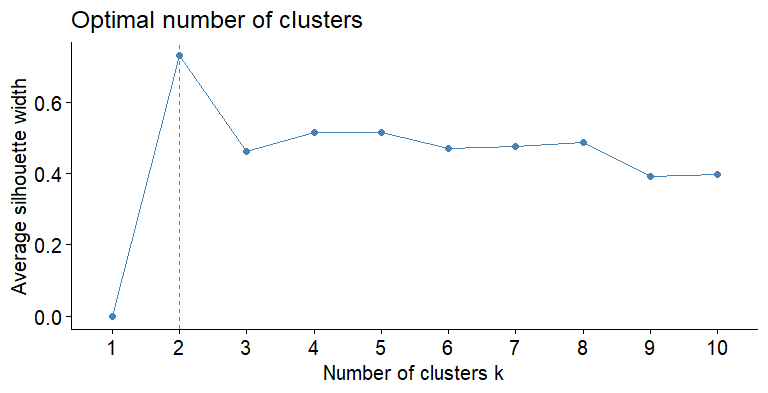
**Figure 4: Correlation between predictor variables and CLV**

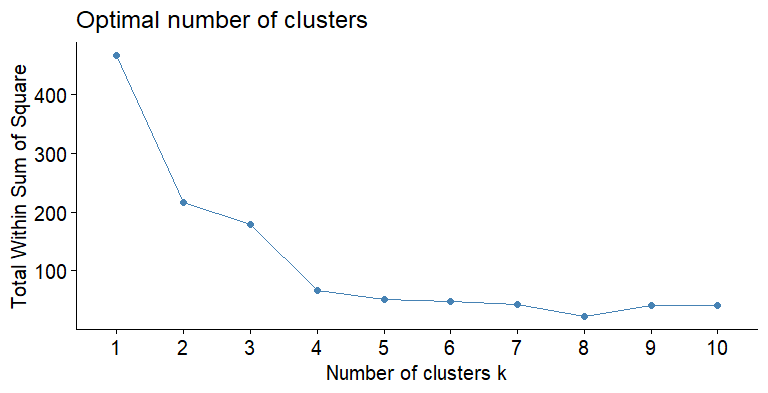
****

**Figure 5: RFM Summary**

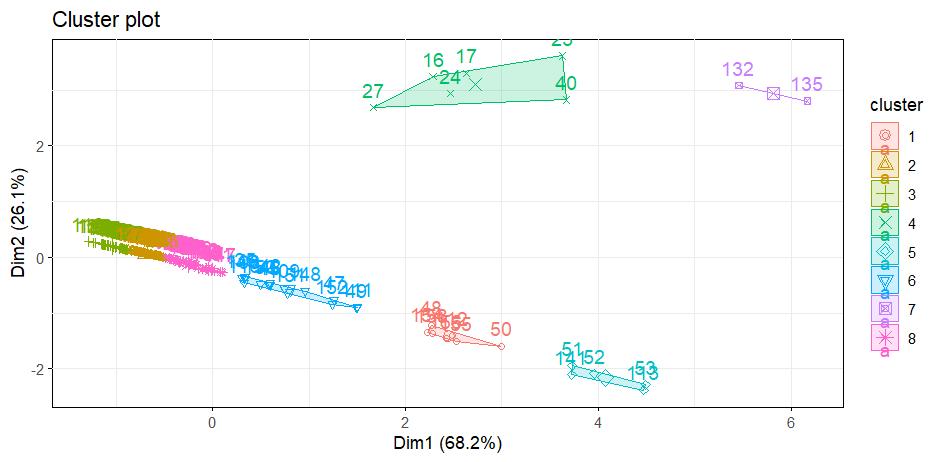


**Figure 6: Silhouette Scores by Cluster**

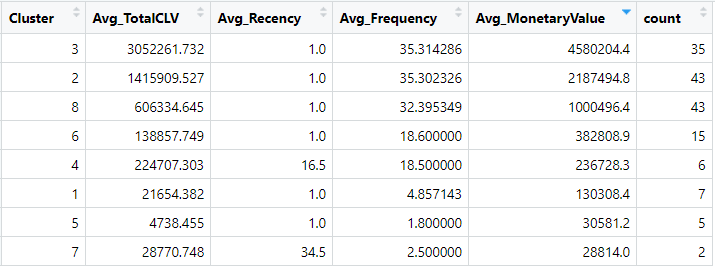
****

**Figure 7: Total WSS Scores by Cluster**

**Figure 8: Final Clusters**



**Figure 9: CLV summary**

****