BUSINESS REPORT ON

ALL LIFE BANK CUSTOMER SEGMENTATION ANALYSIS

TABLE OF CONTENTS

LIST OF FIGURES

TITLE **PAGE NO**

1. OBJECTIVE

The primary objective of this project is to identify customer segments based on their financial and interaction data with the bank. This segmentation will help the bank better understand its customer base and develop targeted marketing strategies, optimize product offerings, and improve overall customer satisfaction.

2.BUSINESS CONTEXT

AllLife Bank seeks to improve its customer relationship management by leveraging data-driven insights. By analyzing customer data using clustering algorithms, the bank aims to segment its customers into meaningful groups. These segments will enable the bank to:

1. Tailor products and services to meet the specific needs of each group.
2. Design personalized marketing campaigns for better customer engagement.
3. Allocate resources effectively to improve customer satisfaction and loyalty.

3.DATA DESCRIPTION

The data contains the different factors to analyze for the content. The detailed data dictionary is given below.

Data Dictionary

- Sl_No: Primary key of the records
- Customer Key: Customer identification number
- Average Credit Limit: Average credit limit of each customer for all credit cards
- Total credit cards: Total number of credit cards possessed by the customer
- Total visits bank: Total number of visits that the customer made (yearly) personally to the bank
- Total visits online: Total number of visits or online logins made by the customer (yearly)
- Total calls made: Total number of calls made by the customer to the bank or its customer service department (yearly)

4.EXPLORATORY DATA ANALYSIS (EDA)

The distribution and relationships of variables within the dataset. This analysis helps uncover insights about visa approval patterns that will inform the model-building process.

4.1 UNIVARIATE ANALYSIS

Customer Key

- Distribution is uniform, ensuring a balanced representation of customer IDs across the dataset.

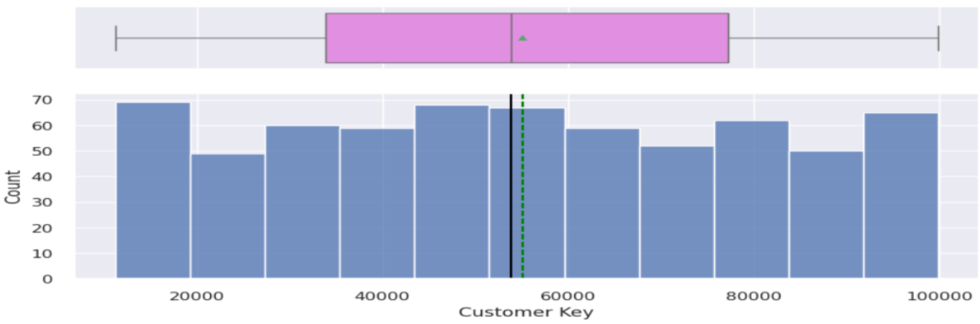


FIG 1 CUSTOMER KEY

Average Credit Limit

- The distribution is skewed to the right, with a long tail. This indicates a few customers with very high credit limits. The boxplot helps identify outliers which could represent high-value customers or potential fraud risks.
- Outliers, as indicated by the boxplot, represent high-value customers who require tailored services.

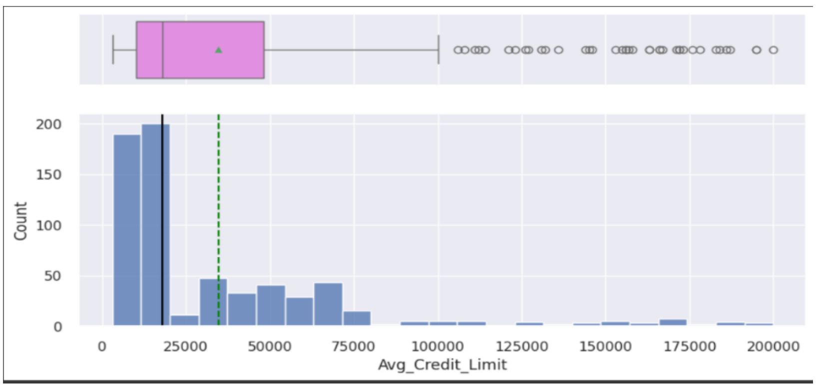


FIG 2 AVG_CREDIT_LIMIT

TOTAL CREDIT CARDS

- The boxplot indicates that the number of total credit cards owned by customers is evenly distributed with no significant outliers.
- A peak around 4–5 credit cards, indicating most customers fall in this range.
- A small proportion of customers owning a very high number of credit cards (7–10), which may represent a specific segment of financially active customers

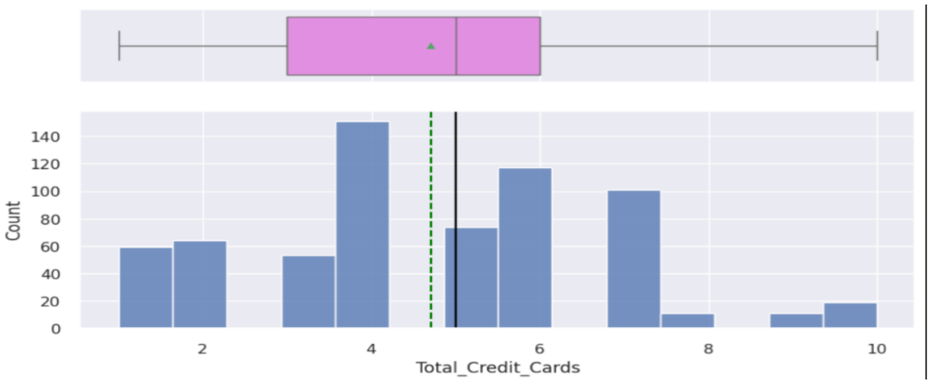


FIG 3 TOTAL CREDIT CARDS

TOTAL VISITS BANK

- The distribution of 'Total_visits_bank' is slightly right-skewed, with a peak around 2-4 visits. There are fewer customers with a high number of bank visits.
- The mean and median appear to be close, indicating that the skew is not very extreme.
- The boxplot suggests the presence of a few outliers on the higher end. These represent customers with unusually high numbers of bank visits. Investigate these customers further to understand the reasons behind the high frequency of visits (e.g., specific banking needs, customer service issues).
- The majority of customers make between 2 and 6 visits to the bank.

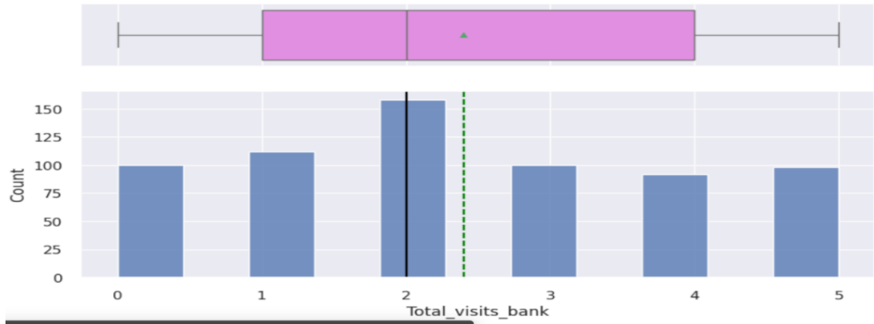


FIG 4 TOTAL VISITS BANKE

TOTAL VISITS ONLINE

- A right-skewed distribution suggests that most customers visit online infrequently,with a few customers making many online visits. The mean will be higher than the median in this case.

- Outliers represent customers with exceptionally high or low numbers of online visits.The outliers are they valid data points or errors? Their presence could significantly affect clustering.

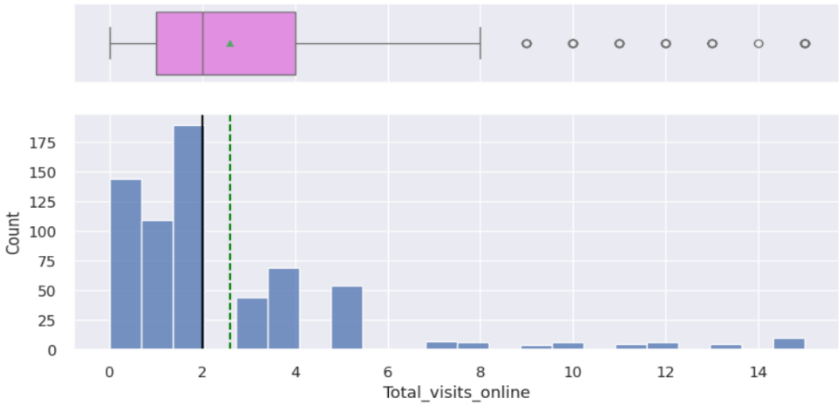


FIG 5 TOTAL VISITS ONLINE

TOTAL CALLS MADE

- The distribution of 'Total_calls_made' is slightly right-skewed,meaning there are more customers who made fewer calls and a smaller number of customers who made a large number of calls. The peak of the histogram seems to be around 2-3 calls.
- The mean and median are relatively close, but the mean is slightly higher than the median due to the right skew. This reinforces the observation of some customers making a higher number of calls, which pulls the mean upwards.
- The boxplot indicates the presence of some outliers on the higher end of the 'Total_calls_made' variable.

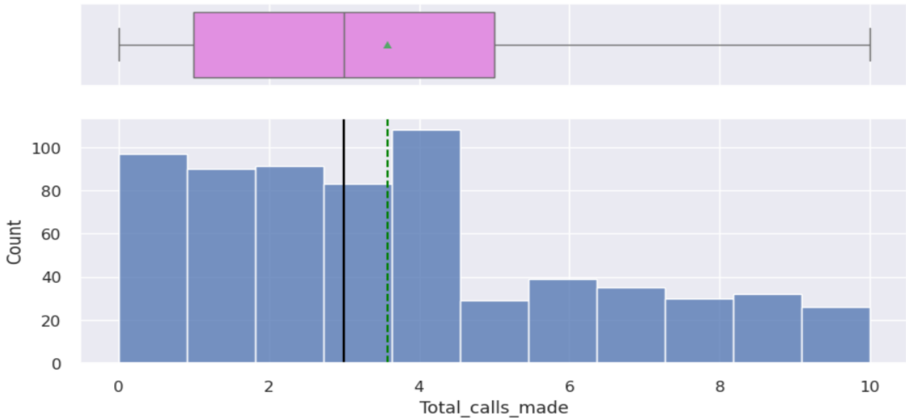


FIG 6 TOTAL CALLS MADE

4.2 BIVARIATE ANALYSIS

- This visualization helps understand how bank visits might influence credit limit and card ownership. Look for clusters or patterns that could indicate distinct customer segments. Do customers with many cards and high credit limits visit the bank more or less often.

- This plot visualizes the relationship between in-person, online, and phone banking preferences. Are customers who use online banking less likely to visit the bank in person? Do customers who make frequent calls also use online banking less.
- 'Avg_Credit_Limit' vs. 'Total_Credit_Cards' provides a visual confirmation of the positive correlation identified in the heatmap. The upward trend of the data points reinforces the observation that higher credit limits are associated with more credit cards. Observing clusters or patterns within this scatter plot could reveal customer segments (e.g., those with high credit limits and many cards might be distinct from customers with lower credit limits and fewer cards). The color coding based on 'Total_visits_bank' adds another dimension to the visualization; investigate if there are any patterns between credit usage and bank visits.

FIG 7 PAIR PLOT

4.3 CORRELATION MATRIX

- a strong positive correlation between 'Avg_Credit_Limit' and 'Total_Credit_Cards' suggests that customers with higher credit limits tend to have more credit cards.
- 'Total_visits_bank', 'Total_visits_online', and 'Total_calls_made'. Do customers who visit the bank frequently also use online services less, and vice-versa? Are call center interactions related to online or in-person visits? Investigate these relationships further.

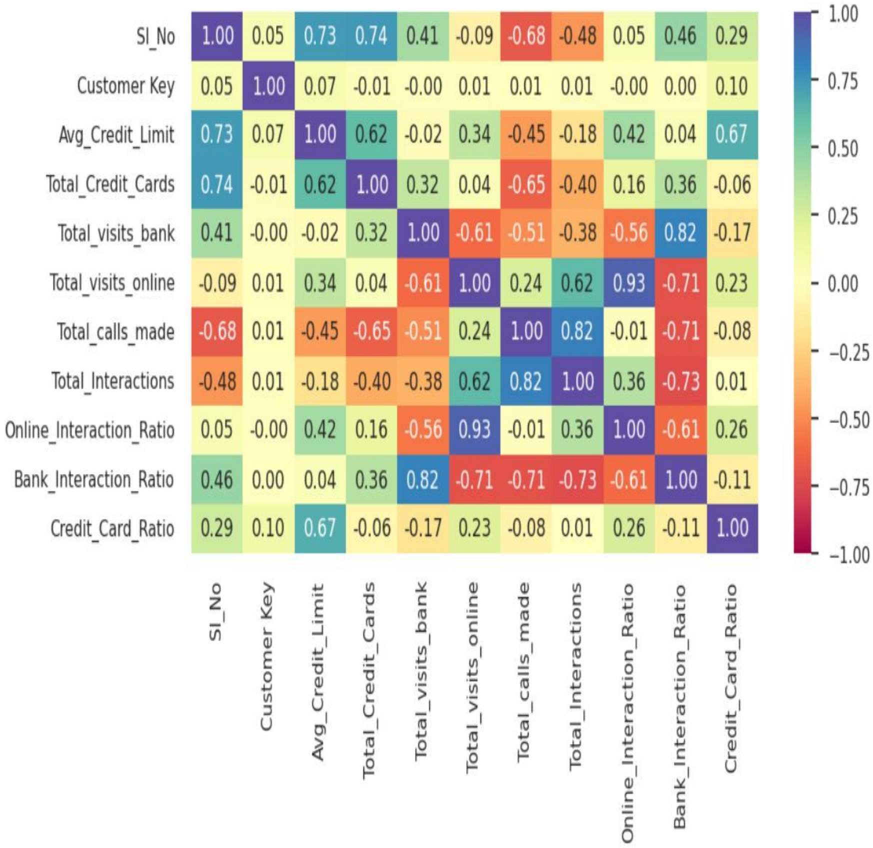


FIG 8 CORRELATION MATRIX

5. DATA PREPROCESSING

The data preparation steps outlined in this section are crucial for ensuring that the dataset is in a suitable format for modeling. This preparation enhances the model's ability to learn and generalize from the data, ultimately leading to better performance and insights.

OUTLIER

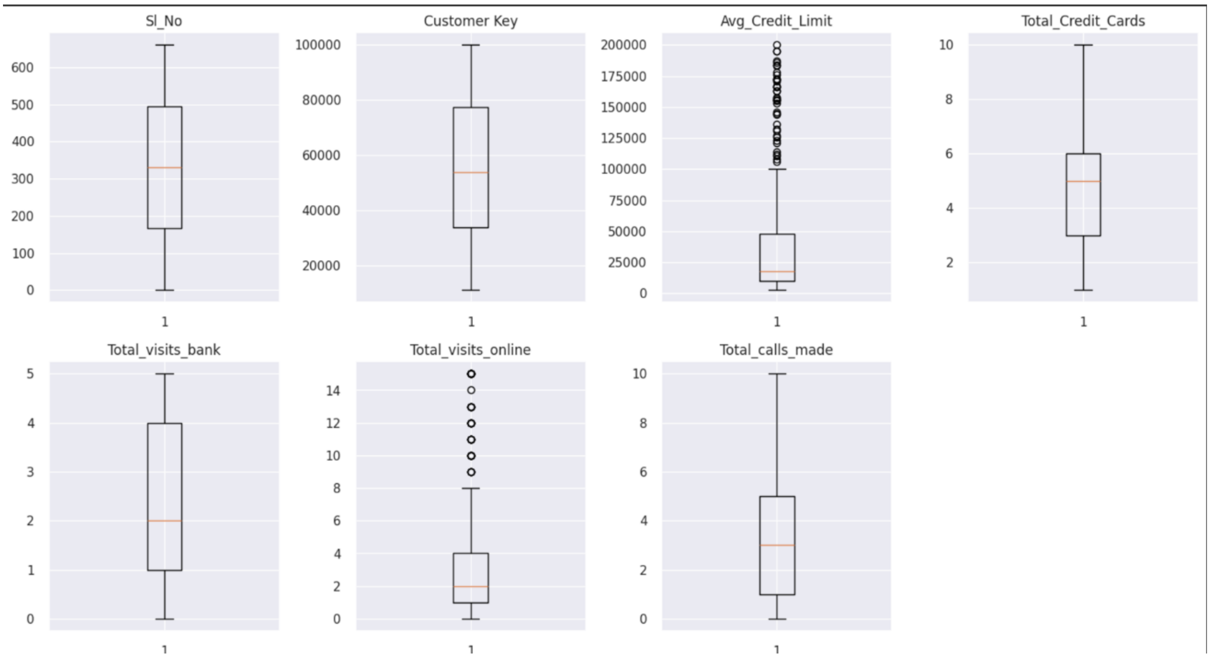


FIG 9 OUTLIER TREATMENTS

Feature Engineering and Analysis

To enrich the dataset and derive meaningful insights, the following interaction features were created:

Total_Interactions

- Captures the overall engagement level of a customer across different channels, summing up visits to the bank, online interactions, and calls made.
- This metric helps identify highly engaged customers versus those with minimal interaction, which can assist in segmenting active vs. dormant customers.

Online_Interaction_Ratio

- Represents the proportion of interactions made online relative to total interactions.
- Customers with higher values may prefer digital channels, indicating potential targets for online-specific campaigns or services.

Bank_Interaction_Ratio

- Represents the proportion of interactions made in person at the bank relative to total interactions.

- Low ratios may indicate less reliance on physical branches, which could inform the allocation of branch resources.

Credit_Card_Ratio

- Reflects the average credit limit per card. This feature may help assess a customer's spending power and financial stability.
- Higher ratios may indicate premium customers who might benefit from exclusive card offers or higher credit limits.

Data Scaling

The following features were scaled:

1. Avg_Credit_Limit: Average credit limit of a customer.
2. Total_Credit_Cards: Total number of credit cards held.
3. Total_visits_bank: Number of visits made to the bank.
4. Total_visits_online: Number of online visits.
5. Total_calls_made: Number of calls made to the bank.
6. Total_Interactions: Sum of all interactions.
7. Online_Interaction_Ratio: Proportion of online interactions.
8. Bank_Interaction_Ratio: Proportion of bank interactions.
9. Credit_Card_Ratio: Average credit limit per card.

The transformed values now represent z-scores:

- **Positive Values**: Indicate data points above the mean.
- **Negative Values**: Indicate data points below the mean.
- **Close to Zero**: Data points near the mean.

For example:

- **Avg_Credit_Limit** (Row 1): A z-score of 2.398942 indicates this customer has a significantly higher credit limit compared to the mean.

- **Total_visits_online** (Row 4): A value of 2.705813 highlights a strong preference for online interactions.

6. K-MEANS CLUSTERING

The K-Means Clustering algorithm has been applied to segment customers based on scaled features.

Average Distortion and the Elbow Method

- The **average distortion** (inertia) measures the sum of squared distances between each data point and its assigned cluster centroid. Lower distortion indicates better clustering.
- As the number of clusters increases, the distortion decreases. This is expected because more clusters mean data points are closer to their centroids
- .The **Elbow Method** helps identify the optimal number of clusters by locating the "elbow point," where distortion starts decreasing more slowly.
- Based on the results, the **elbow point** is at **K = 5**, where the average distortion is **1.5109991098296305**.
- This suggests that dividing the customers into 5 clusters captures the most meaningful segments while balancing complexity.
- It balances the trade-off between reducing distortion and overcomplicating segmentation.
- Five clusters provide meaningful groupings that can be analyzed for business insights.

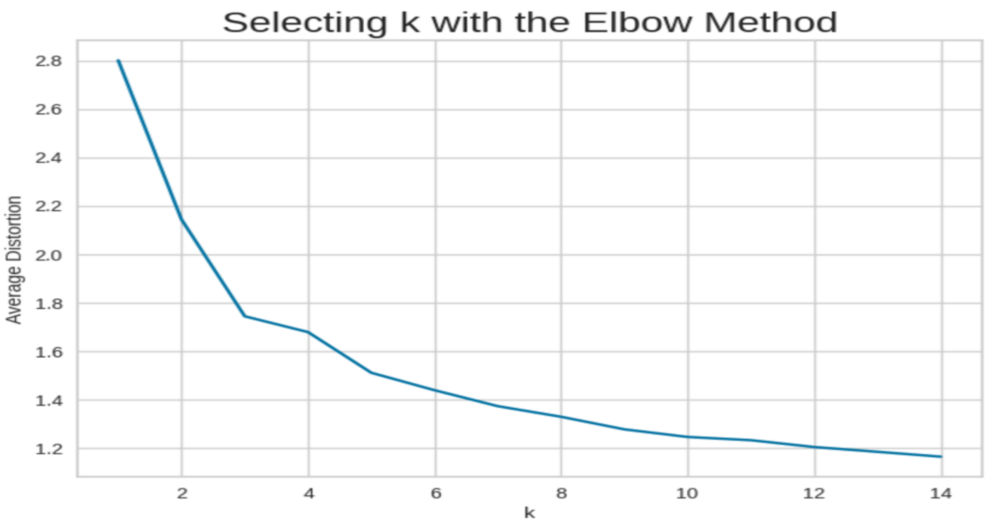


FIG 10 Distortion Score Elbow for KMeans Clustering

THE SILHOUETTE SCORES

The Silhouette Score is used to measure the quality of clustering. It ranges between -1 and 1:

- A score close to 1 indicates that the data points are well-matched within their clusters and poorly matched to neighboring clusters.
- A score around 0 suggests overlapping clusters.
- A negative score indicates that data points might be assigned to the wrong cluster.
- For **n_clusters = 3**, the **silhouette score is 0.464**, the highest among all tested values of K.
- This suggests that a **3-cluster solution** provides the most distinct and compact clusters compared to other cluster numbers.
- Fewer clusters (compared to K=5) offer better interpretability while still maintaining high-quality segmentation.

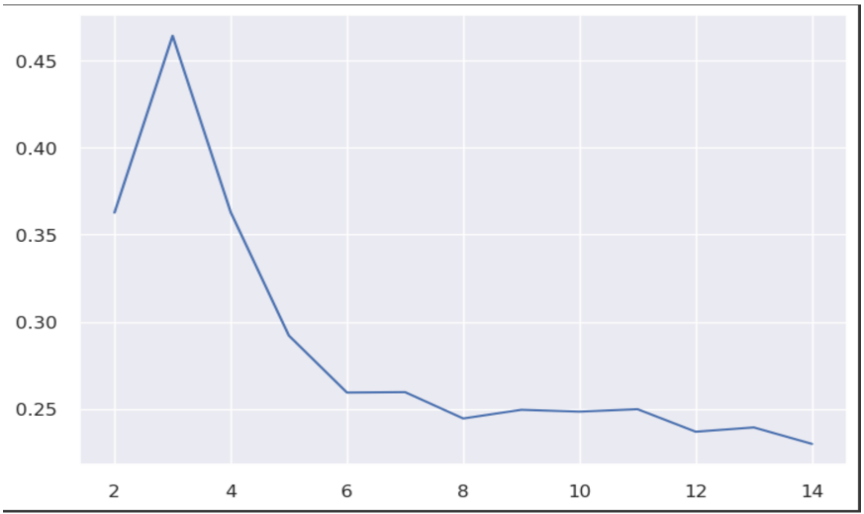


FIG 11 Silhouette Score Elbow for KMeans Clustering

- **optimal no. of clusters with silhouette coefficients**

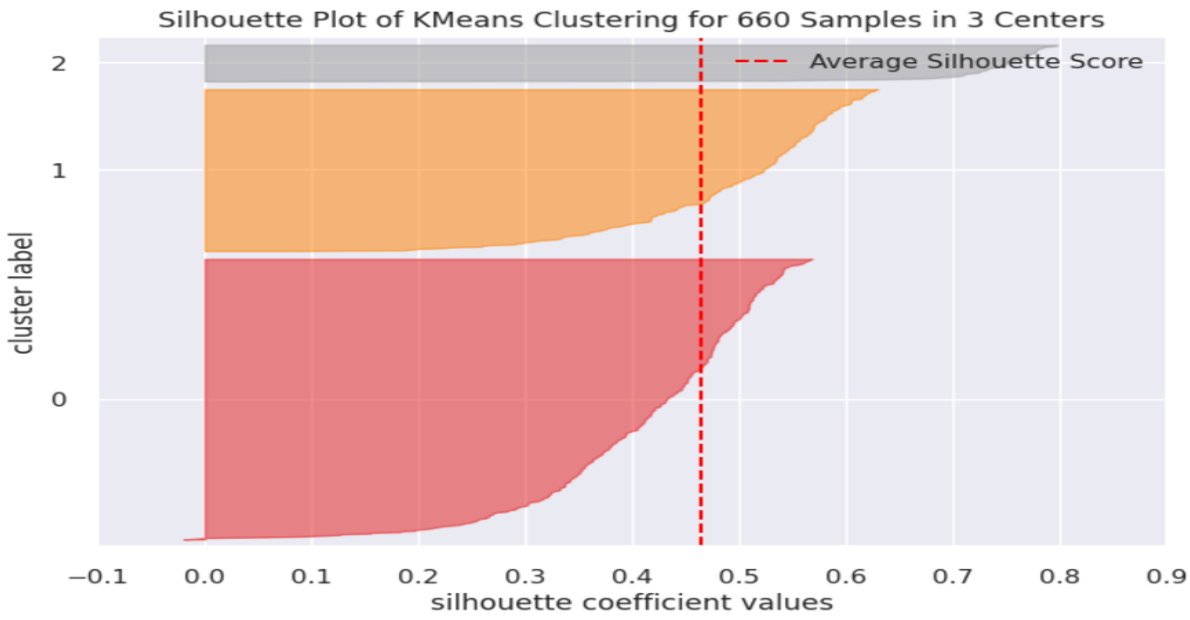


FIG 12 SILHOUETTE COEFFICIENT VALUE

Cluster Profiling Insights

Cluster 0: General Overview

- Size: Largest group with 386 customers.
- Average Credit Limit: ₹54,752 (Moderate range).
- Total Credit Cards: 0.066 (Minimal).
- Bank Interactions: Higher focus on online interactions compared to in-person bank visits or calls.
- Customers exhibit a balanced use of online platforms and other channels but lean more towards online interactions.
- Moderate engagement overall with no significant preference for direct bank visits or calls.
- Credit Card Usage: Customers in this segment generally have limited reliance on credit cards.
- Potential for digital engagement campaigns to promote bank services online.
- Focus on offering low-maintenance financial products such as prepaid cards or fixed deposits.

Cluster 1: General Overview

- Size: Moderate group with 223 customers.
- Average Credit Limit: ₹55,319 (Moderate range).
- Total Credit Cards: Negative trend (-0.685), suggesting these customers rarely utilize credit cards.
- Higher emphasis on total interactions overall.
- Greater preference for online interaction ratio than physical visits or calls.
- Digitally active customers who are less inclined towards using traditional banking products, such as credit cards.
- Their online interaction ratio is slightly higher than others.
- A good fit for digital-first banking strategies such as personal finance management apps or virtual consultations.
- Opportunity to increase product penetration by promoting low-risk credit or EMI options.

Cluster 2: General Overview

- Size: Smallest group with 51 customers.
- Average Credit Limit: ₹57,304 (Highest).
- Total Credit Cards: Positive and high utilization (2.49).
- Highest total bank visits, indicating a preference for in-person services.
- Substantial online interaction ratio and notable overall interaction score.
- These customers are highly active across multiple channels.
- These are premium customers who actively interact with the bank both online and in-person.
- Likely to have sophisticated financial needs, including multiple credit cards.

- This is a high-value customer segment; they should be prioritized for personalized services.
- Tailor financial products such as premium credit cards, investment advisory, and loyalty programs.
- Enhance their in-bank experience while maintaining robust online service options.

Credit Limit Group Analysis

Low Credit Limit Customers:

Majority (191 customers) in Cluster 0.

Minor presence in Clusters 1 and 2.

Medium, High, Very High Credit Limit Groups:

No representation in the data provided, indicating the customer base might not yet include such segments.

Opportunity exists to attract and retain higher-net-worth individuals through targeted campaigns.

7. HIERARCHICAL CLUSTERING

The Cophenetic Correlation is a metric that measures how well the hierarchical clustering preserves the pairwise distances between the data points. A higher cophenetic correlation indicates that the clustering method has done a better job of preserving the original distance structure.

The highest cophenetic correlation of **0.8604** was obtained with **Euclidean distance** and **average linkage**. This indicates that this combination of distance measure and linkage method best preserves the relationships between the data points in the hierarchical clustering process.

- **Euclidean distance and complete linkage**: Correlation of **0.7602**.
- **Euclidean distance and ward linkage**: Correlation of **0.7289**.
- **Cityblock distance and average linkage**: Correlation of **0.8513**.
- **Euclidean distance and average linkage** performed the best, suggesting it captures the hierarchical structure of the data more effectively than other combinations.

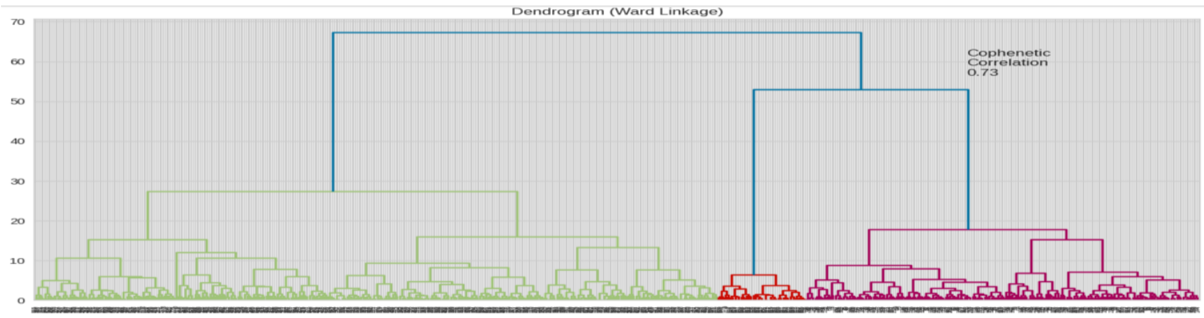
- **Average linkage** is often preferred when the data exhibits roughly equal dispersion across clusters, making it a good fit for this dataset.
- **Ward linkage**, which minimizes the variance within clusters, had a relatively lower cophenetic correlation, suggesting it may not preserve the distances as well as other methods in this case.

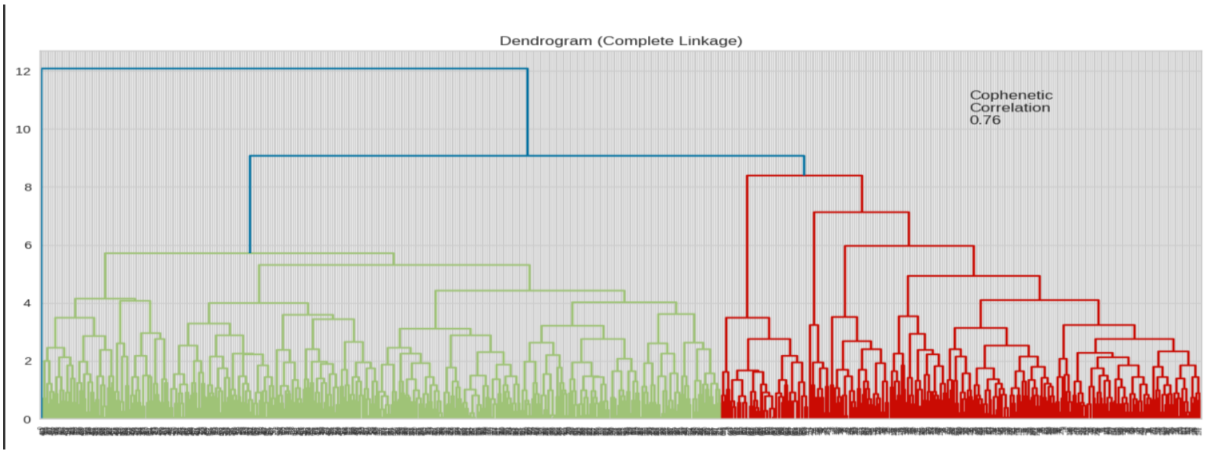
DENDROGRAM ANALYSIS

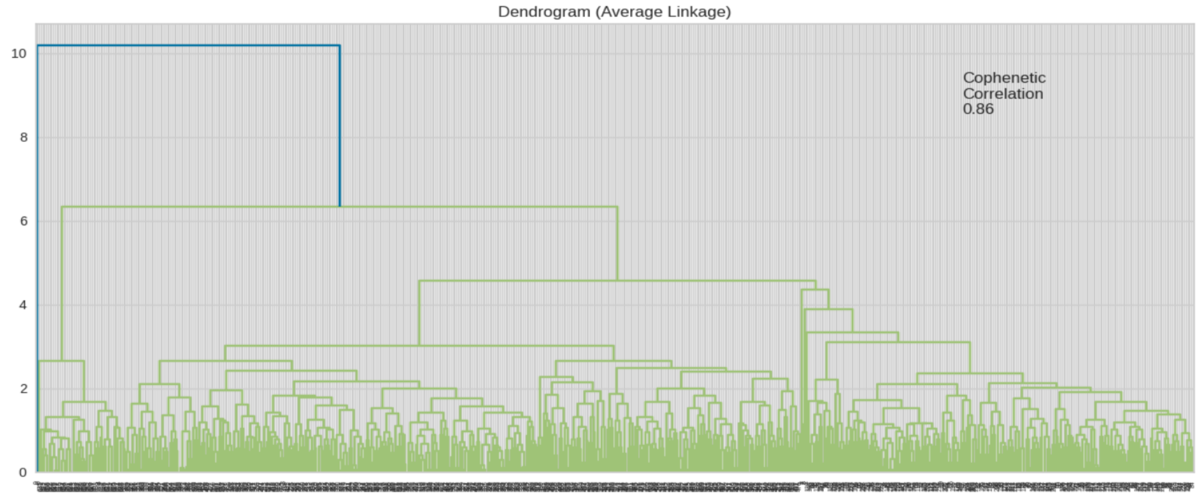
cophenetic correlation coefficients for various linkage methods in hierarchical clustering, and based on the results:

- **Single linkage**: 0.6499
- **Ward linkage**: 0.7289
- **Complete linkage**: 0.7602
- **Average linkage**: 0.8604

average linkage provides the highest cophenetic correlation, suggesting that it best preserves the original pairwise distances between the data points.







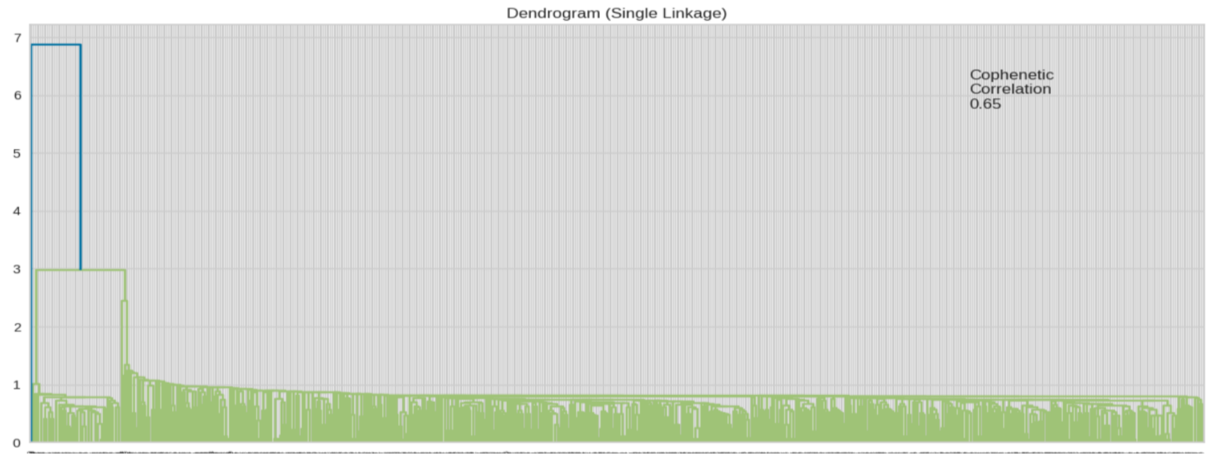


FIG 13 DENDROGRAM

CLUSTER PROFILING

Cluster 0:

- **Average Credit Limit**: $54925.97
- **Total Visits to Bank**: 33713.18
- **Total Visits Online**: 5.51
- **Total Calls Made**: 3.49
- **Count in Cluster**: 387 customers

Credit Limit Distribution:

- **Low**: 71 customers
- **Medium**: 98 customers

- **High**: 116 customers
- **Very High**: 102 customers

This cluster contains customers with higher-than-average credit limits and frequent interactions with the bank (both in-person and online). The high number of visits to the bank and online suggests that these customers might have active accounts or a high level of engagement with the bank's services.

Cluster 1:

- **Average Credit Limit**: $55163.97
- **Total Visits to Bank**: 12197.31
- **Total Visits Online**: 2.40
- **Total Calls Made**: 0.93
- **Count in Cluster**: 223 customers

Credit Limit Distribution:

- **Low**: 95 customers
- **Medium**: 127 customers
- **High**: 1 customer
- **Very High**: 0 customers

This cluster represents a group of customers with relatively lower engagement, with fewer visits to both the bank and online platforms. The distribution of credit limits here leans more toward the lower or medium range, indicating that these customers might be new or less active.

Cluster 2:

- **Average Credit Limit**: $56708.76
- **Total Visits to Bank**: 141040.00
- **Total Visits Online**: 8.74
- **Total Calls Made**: 10.90
- **Count in Cluster**: 50 customers

Credit Limit Distribution:

- **Low**: 0 customers
- **Medium**: 0 customers
- **High**: 0 customers
- **Very High**: 50 customers

Cluster 2 is comprised of customers with extremely high credit limits. This cluster shows a very high level of engagement, as indicated by both the frequent visits to the bank and the high number of calls made. All these customers fall into the "Very High" category for credit limit, suggesting that they might be high-net-worth individuals or VIP clients.

Cluster 0: These customers are relatively well-established with medium to high credit limits and a balanced mix of in-person and online banking behavior. Targeting this segment with offers related to financial services like loans or premium services could be beneficial.

Cluster 1: This group contains customers with low to medium credit limits who are less engaged with both physical and online banking. This segment could benefit from targeted engagement campaigns such as account upgrades, loyalty programs, or reminders about available services.

Cluster 2: This is a highly engaged, high-credit group. These customers could be ideal targets for premium financial products, investment opportunities, or personalized banking services.

8. K-MEANS VS HIERARCHICAL CLUSTERING

Execution Time: K-means is generally faster and better suited for large datasets, whereas hierarchical clustering is slower but more flexible.

Distinct Clusters: Hierarchical clustering can give more distinct clusters, especially in cases where a finer distinction between groups is required.

Number of Clusters: K-means requires you to specify the number of clusters, while hierarchical clustering offers flexibility in choosing the number of clusters based on a dendrogram.

Cluster Profiles: The cluster profiles are similar in both methods, but hierarchical clustering might allow for more granular or flexible groupings.

9. ACTIONABLE INSIGHTS & RECOMMENDATIONS

- Targeted Marketing Campaigns: Cluster 0 (High-Value Customers): These customers have high credit limits, frequent bank visits, and high overall engagement. Focus on premium products and services, exclusive offers, and personalized financial advice. Consider loyalty programs to retain these valuable customers. Cluster 1 (Medium-Value Customers): These customers show moderate engagement and credit limits. Targeted promotions and incentives can encourage increased usage of online banking and other services. Offer products and services that cater to their moderate spending habits and financial goals. Cluster 2 (High-Credit, High-Engagement Customers): These customers have very high credit limits and are highly engaged. Offer them exclusive financial products and services, personalized wealth management options, and premium rewards. Prioritize maintaining their satisfaction to reduce churn risk.
- Customer Service Optimization: Cluster 0 (High-Value Customers): Provide dedicated customer service channels and priority support to ensure quick resolution of issues. Proactive outreach to understand their financial needs and provide personalized guidance. Cluster 1 (Medium-Value Customers): Offer convenient online support channels and self-service options to cater to their moderate engagement. Address any pain points in the online banking or service experience to improve engagement. Cluster 2 (High-Credit, High-Engagement Customers): Provide dedicated wealth management consultants or personalized financial advisors to cater to their unique needs. Ensure that high-touch customer service is available for complex financial inquiries.
- Product Development: Analyze the specific needs of each cluster to identify product gaps and develop new offerings tailored to their profiles. For example, Cluster 0 might benefit from premium credit cards or investment products, while Cluster 1 might benefit from user-friendly financial planning tools or educational resources.
- Channel Optimization: Analyze the interaction patterns of each cluster to understand their preferred channels (online vs. bank visits vs. phone calls). Optimize resources and service offerings based on these preferences. For instance, encourage customers in Cluster 1 to use online banking more often through incentives, improving the user experience, and offering online-only discounts.
- Risk Management: Monitor the credit limit utilization and spending patterns of each cluster. Implement appropriate risk management measures based on the identified patterns to minimize potential risks.

- Customer Segmentation Refinement: Periodically re-evaluate customer segmentation using updated data to ensure the clusters remain relevant and representative. Refine the segmentation criteria based on changing customer behavior and market trends.

10. CONCLUSION

The clustering analysis has provided valuable insights into customer behavior at AllLife Bank, highlighting the need for personalized marketing strategies, product offerings, and risk management techniques for each customer segment. By acting on these insights, the bank can foster stronger customer relationships, enhance engagement, optimize credit management, and ultimately improve profitability. The recommendations outlined here are designed to maximize customer retention, satisfaction, and growth by aligning the bank's offerings with the unique needs of each cluster.