

Advanced Customer Analytics for Banking Campaigns: Market Basket Analysis & Customer Segmentation

Introduction:-

This project applies **Market Basket Analysis** (Apriori algorithm) and **Customer Segmentation** (KMeans clustering) to the Bank Marketing dataset. Market Basket Analysis helps identify strong associations among customer attributes and behaviors, revealing which profiles are more likely to respond to specific offers. Customer Segmentation groups customers into distinct personas based on demographics (age, marital status, job), financial history, and campaign interactions, enabling more targeted marketing strategies.

i) Market Basket Analysis:-

Objective:-

The goal of this analysis was to uncover hidden patterns in customer attributes and behaviors that can guide targeted marketing strategies for future banking campaigns. I applied the Apriori algorithm on the dataset to identify the most significant association rules, focusing on rules sorted by Lift, which highlights strong relationships.

Methodology:-

1. Data preparation:-

- Selected relevant features which can represent customer behaviour
- Encoded categorical variables using one-hot encoding

2. Algorithm used:-

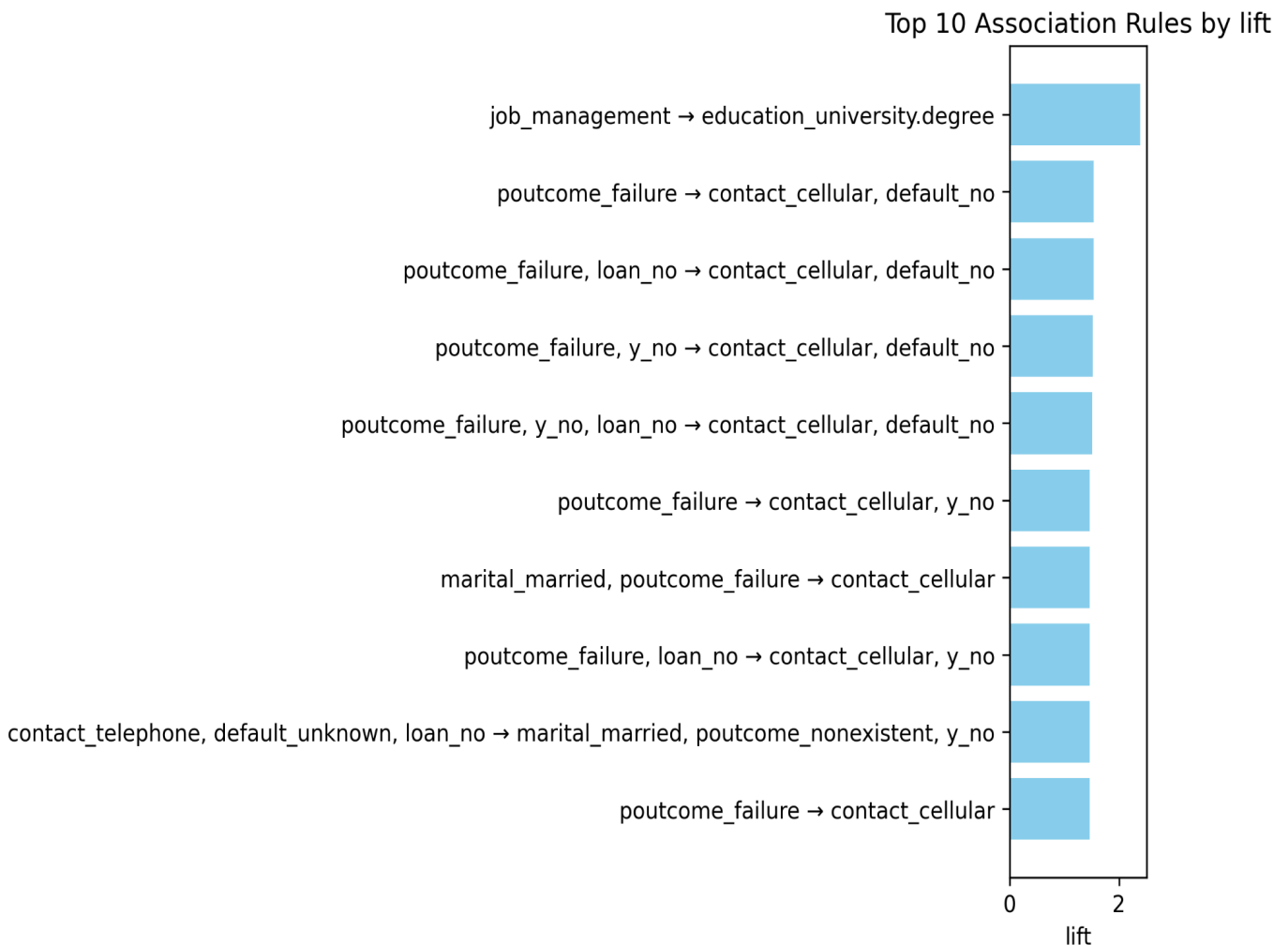
- Apriori Algorithm is used which generates the items and combinations of items which exceed minimum support (user defined)
- Derived association rules (antecedent → consequent) with metrics:
 - Support: frequency of occurrence in dataset

- Confidence: likelihood of consequent given antecedent
- Lift: strength of rule compared to random chance

3. Rule selection:-

- Top 10 rules are selected which are sorted by Lift.
- The reason behind sorting by descending order of Lift is that we get the combinations which are high-value and not just occurring at random chance. If sorted by confidence, we might get rules that are a bit too obvious to occur, which the bank may already know.

Data and Business insights for the top 10 rules



Rule 1:-

antecedent:- job = Management

consequent:- Education = University Degree

confidence:- 70.55%

lift:- 2.39

insights:- Customers that are having jobs in the field of management are 70% likely to hold a university degree. This group is 2.4x more likely to hold a university degree than average customers. Due to high financial literacy, this group is suitable for being targeted for premium banking products.

Rule 2:-

antecedent:- poutcome = Failure

consequent:- Default = No & Contact = Cellular

confidence:- 81.51%

lift:- 1.54

insights:- Customers who rejected the previous campaign are 81% likely to have no default status and are easily reachable via cellular contacts. This group is 1.54x more likely than a random chance. This group should be re-targeted as it's readily reachable and has no default customers, maintaining the low risk.

Rule 3:-

antecedent:- Loan = No & poutcome = Failure

consequent:- Default = No & Contact = Cellular

confidence:- 81.12%

lift:- 1.53

Customers with no loans and rejection to previous campaigns are 81% likely to have no default status and are easily reachable. This group is 1.53x more likely than a random chance. As this group has customers with no loans taken and no default, it should be considered for re-targeting as it provides less risk along with less financial burden.

Rule 4:-

antecedent:- y = No & Poutcome = Failure

consequent:- Default = No & Contact = Cellular

confidence:- 80.20%

lift:- 1.52

customers from this group rejected this as well as previous campaigns and are 80% likely to have no default and are easily reachable. This group is 1.52x more likely than a random chance. Even though this group provides less risk due to no defaults, it shall not be targeted because of its repetitive rejections suggesting lack of interest. For this group, the marketing budget shall not be allocated as the majority of it shall go to waste.

Rule 5:-

antecedent:- y = No & Loan = No & Poutcome = Failure

consequent:- Default = No & Contact = Cellular

confidence:- 79.85%

lift:- 1.51

Customers who have rejected this as well as previous campaigns and have no loan are 80% likely to have default status and are easily reachable. This group is 1.51x more likely than a random chance. Even though this group is seen with consistent rejections, it should still be re-targeted with a good offer later, as this group brings a good credit profile aligning with positive customer behaviour.

Rule 6:-

antecedent:- poutcome = Failure

consequent:- y = No & Contact = Cellular

confidence:- 79.51%

lift:- 1.47

customers with previous rejections are 80% likely to reject this campaign and are easily reachable. This group is 1.47x more likely than a random chance. Either this group shall not be re-targeted or the bank should come up with a more exciting offer the next time.

Rule 7:-

antecedent:- Marital = Married & poutcome = Failure

consequent:- Contact = Cellular

confidence:- 93.12%

lift:- 1.47

customers who are married and have rejected previously are 93% likely to be easily reachable again. This group is 1.47x more likely than a random chance.. previous rejection doesn't necessarily mean that he/she will reject everytime. as this customer group is answering the calls 93% of times, it should be re-targeted.

Rule 8:-

antecedent:- Loan = No & poutcome = Failure

consequent:- y = No & Contact = Cellular

confidence:- 79.32%

lift:- 1.47

Customers with previous rejections and no loans taken are 79% likely to reject this campaign and are easily reachable. This group is 1.47x more likely than a random chance. Either this group shall not be re-targeted or the bank should come up with a more exciting offer the next time.

Rule 9:-

antecedent:- Contact = Telephone, Loan = No, Default = Unknown

consequent:- Marital = Married, y = No, Poutcome = Nonexistent

confidence:- 71.47%

lift:- 1.46

Customers who are contacted via telephone, have no loans taken, and their default status is unknown are 71% likely to be married, have rejected this campaign and were not reached in the previous campaign. This group is 1.46x more likely than a random chance. As this group has not taken any loans and default status is unknown, then despite 70% customers rejecting the campaign for the first time, the credit profile of the customers is still considerably good. Most of them are targeted first time so chances shall be taken for re-targeting where customers next time could be converted.

Rule 10:-

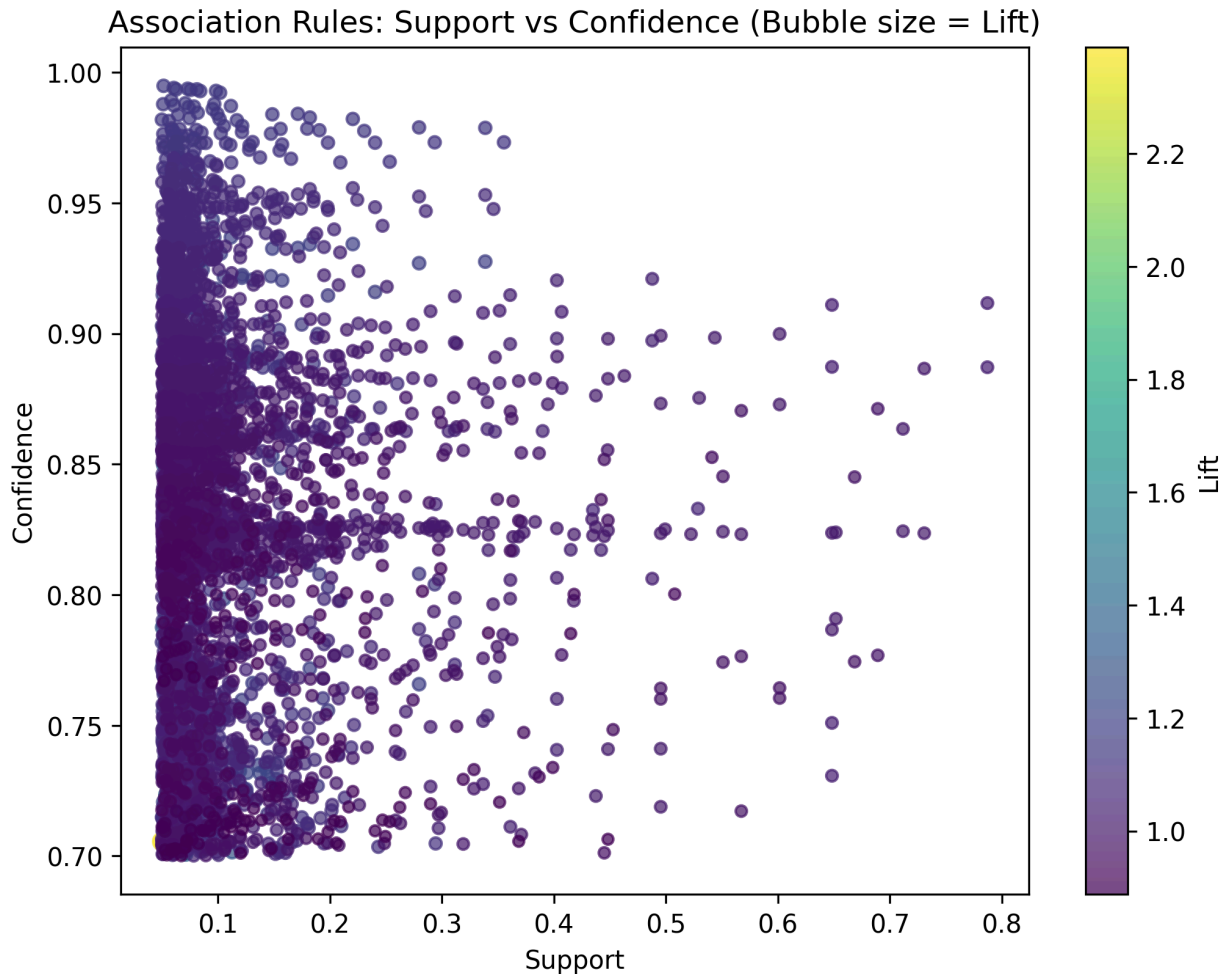
antecedent:- poutcome = Failure

consequent:- Contact = Cellular

confidence:- 71.47%

lift:- 1.46

customers who rejected the previous campaign are 70% easily reachable. This group is 1.46x more likely than a random chance. As just the previous rejection can not tell us certainly that they will reject again, and that 71% of them actively pick up the calls, re-targeting shall be done.



The graph above shows the scatter plot where each dot represents a rule. Looking at the graph, we can clearly see that most of the rules have a support less than or equal to 0.1, indicating that most of them are rare, uncovering some or the other hidden pattern. There are very few common rules, which the banking company might be well versed with. The confidence ranges from 0.7 to 1.0 which suggests that each rule has a strong likelihood of occurrence of the consequent given the antecedent.

Recommendations:-

- Use your insights to make actionable suggestions:
- Prioritize high-confidence, high-lift segments for campaigns.
- Avoid segments with repeated rejections, unless offering a new incentive.
- Consider channel strategy: cellular contacts respond more reliably than telephone.
- Suggest tailored offers for premium segments (management/university-educated) vs. safe segments (no loans, no default).

Conclusion:-

- Several high lift associations were uncovered which mostly gave us patterns that aren't too obvious.
- Cellular contact most the most preferable when customers had no loans and defaults, which indicated that group had low risk and was easily reachable. This group was identified to be beneficial for the bank and having potential to get converted.
- At the same time, a group of customers was also identified which repeatedly rejected the campaigns thereby wasting campaign resources. This group is identified to have least conversion rate
- Many management professionals were also identified and were strong candidates to cross-sell advanced products.

ii) Customer Segmentation:-

Objective:-

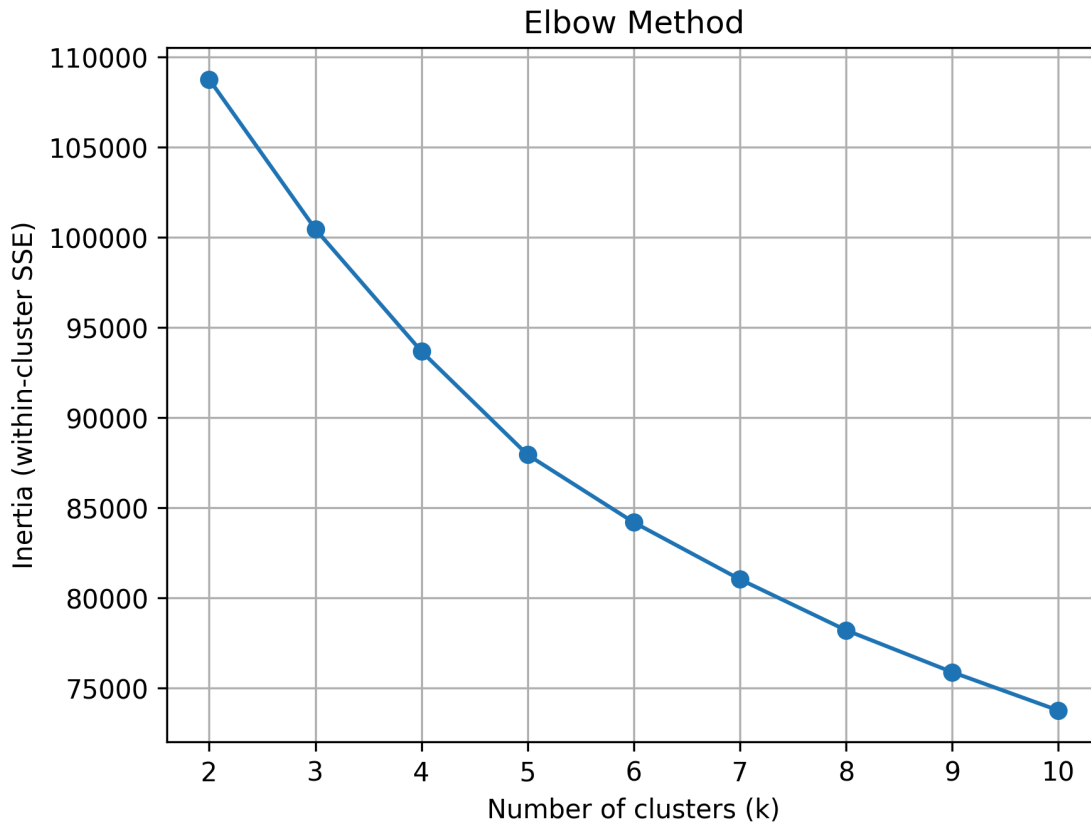
The goal is to group customers with similar demographics, financial status, and campaign behavior into distinct segments. This helps the bank design targeted campaigns, tailor product offers, and prioritize low-risk, high-response groups to improve conversions and resource efficiency.

Data preparation:-

- Selected relevant customer attributes
- Encoded categorical variables (One-Hot Encoding).
- Scaled numerical variables (standardization) to ensure equal weight in clustering.

Clustering approach:-

- Applied KMeans clustering to group customers into distinct segments.
- Used the Elbow Method to identify the optimal number of clusters (k=5 chosen).



Elbow method graph which helped me visually check the optimal number of clusters.

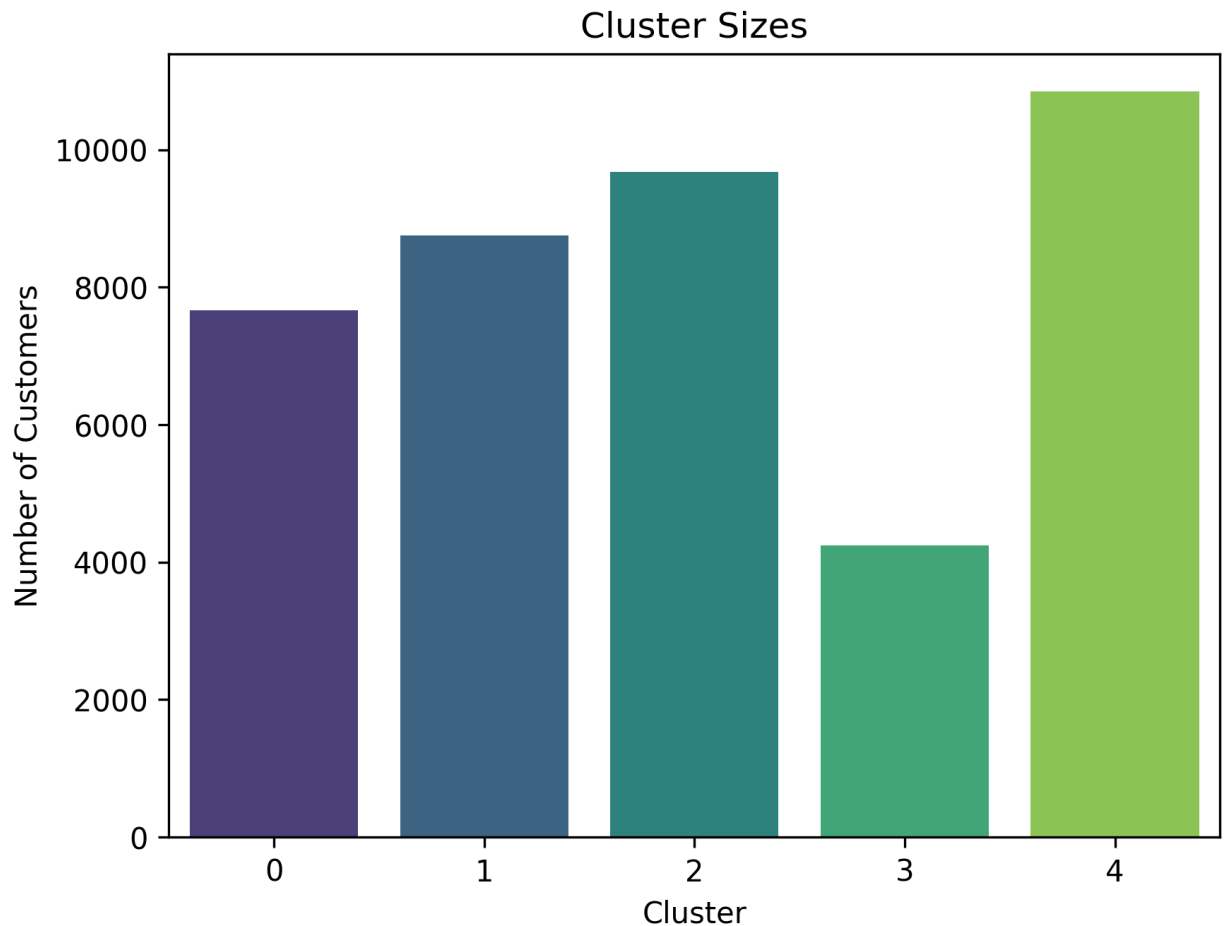
Cluster Validation:-

- Silhouette Score (0.166 → weak separation, expected in overlapping behaviors).
- Davies-Bouldin Index (1.906 → moderate overlap).
- The relatively low silhouette score and high Davies-Bouldin Index reflects natural overlap in customer behaviors (e.g., many segments share loan and education characteristics). To balance interpretability with business value, k=5 was retained as it provided the most meaningful personas for campaign design

Cluster Profiling:-

- For Categorical features, grouped by Cluster and showed percentage contribution of each category
- For Numerical features, grouped by Cluster and showed the mean.

Data and Business insight for each cluster:-



The graph above represents the data distribution in each cluster. The 5th cluster (named 4) represents the highest number of customers which represents the cluster having diversity in age, jobs and is one of the clusters having highest literacy. The 4th cluster (named 3) has the lowest number of customers. This cluster also has highly literate people in it but most of them are doing a professional job unlike cluster 4 where job types vary.

Cluster 0:-

Data insights:-

- People in this group are in their early 30s on average.
- 82% of them haven't taken any personal loan, indicating less financial burden.
- 86% don't have a history of default indicating a low risk profile.
- 51% of them have taken out a housing loan indicating financial burden lies here.
- 35% are either working as admin and 30% of them are giving various services, indicating that they could fall in a moderate income category
- 97% of them have just received education till high school suggesting that most of them have moderate literacy
- 61% of them are easily reachable by cellular means
- 84% of them have never undergone a campaign, indicating most of them are first-timers

Business insights:-

This group shall be targeted because despite having 51% financial burden, 86% of them have no history of defaults and hence show low risk. Also as most of them are easily reachable and are literate, it indicates that they must be having basic understanding about the offers given to them, and hence should be targeted with cross sell offers related to their existing commitments. Also as the majority are first-timers, they should be given joining incentives which could help increase the conversion rate.

Cluster 1:-

Data insights:-

- People in this group are in their mid 30s on average.
- 82% of them have no personal loans, indicating less financial burden
- 67% of them have no default history and for the rest, nothing can be said as default history is unknown. Nonetheless, majority of them have low risk profile
- 49% of them have housing loans, indicating significant financial burden
- 73% of them have blue collar jobs, indicating the majority of them are from the labor class and have stable but moderate incomes.
- 53% of them have received basic 9 year education aligning with most having blue collar jobs. This indicates mild to moderate literacy
- Nearly half of them are reachable via cellular means and the other half are reachable through landline. This indicates that half of them are reachable via quicker means, and the other half may or may not be reachable at times.

- 89% of people from this group have never undergone a campaign indicating that majority of people from this group are first-timers

Business insights:-

82% of them have no personal loans taken but 49% of them comprise nearly half of the group having taken housing loans. But as the 67% majority does not have a default history for sure, the majority of the group comes under low risk profile and therefore this group shall be targeted with cross-sell offers related to their existing commitments. Also as majority are first-timers, they should be given joining incentives which could help increase the conversion rate

Cluster 2:-

Data insights:-

- Average age of this group is 34
- 81% of them have no personal loans taken
- 91% of them have no default history, indicating that this group has people holding low risk profile
- 54% of them have taken housing loans, indicating this group has significant financial burden
- Nearly a half (48%) of them work as an admin followed by 16% of them working as technicians. This indicates a stable but moderate income group
- 95% of them hold university degree, indicating high literacy levels of this group, aligning with 64% of them (admins and technicians) doing a professional job.
- 73% of them are easily reachable via cellular means
- 89% of people from this group have never undergone a campaign indicating that majority of people from this group are first-timers
- 85% of people from this group have never undergone a campaign indicating that majority of people from this group are first-timers

Business insights:-

Despite 54% of them having housing loans, 91% of them have no default history indicating that most of them have a strong repayment background and hence belong to the low risk segment. With 95% of them being highly literate, they would likely respond to complex products like investment, insurance and education offers. Also as 85% of them are first-timers, they should be given joining incentives which could help increase the conversion rate

Cluster 3:-

Data insights:-

- people are in mid 30s on average
- 82% of them haven't taken any personal loans
- 88% of them have no default records, with 11% having unknown status. With the majority not having default history, people in this segment have a low risk profile.
- 54% of them have taken housing loans, indicating this group has significant financial burden
- 75% of them are technicians indicating that their income is stable but moderate
- 92% of them have done a professional course showing inclination towards studies and hence this segment has a majority of people with high literacy. It also aligns with 75% of them being technicians, a job considered to be professional.
- 69% of them are easily reachable via cellular means
- 85% of people from this group have never undergone a campaign indicating that majority of people from this group are first-timers

Business insights:-

Despite 54% of them having housing loans, 88% of them have no default history indicating that most of them have a strong repayment background and hence belong to the low risk segment. With 92% of them being highly literate, they would likely respond to complex products like investment, insurance and education offers. Also as 85% of them are first-timers, they should be given joining incentives which could help increase the conversion rate

Cluster 4:-

Data insights:-

- Average age of people in this segment is 53
- 82% of them have not taken any personal loans
- 68% of them are not having any default history with the rest 31% having unknown status. The majority is lesser than other clusters indicating that there might be more high risks profiles than other segments
- 52% of them took housing loans indicating financial burden
- Jobs are more distributed with some doing professional jobs like technician (10%) and some doing blue collar jobs (17%). 20% found to be having an admin job which is less professional but not blue collar. It indicates varying incomes in this segment.
- No education category dominates here, with some holding university degrees (26%), followed by some who have just pursued basic 4 year education (21%) and 18% of

them have just completed high school. It indicates that this segment has varying literacy levels

- 64% of them are reachable via cellular means
- 85% of people from this group have never undergone a campaign indicating that majority of people from this group are first-timers

Business insights:-

This segment could have more high risk profile people due to 31% of them having unknown default status. 52% of them are carrying housing loans which indicates many people in this segment are having financial burdens. Due to varying literacy and incomes, and majority of them being first-timers, and average age being 53, they should be targeted with literacy specific offers like awareness/education campaigns and income specific offers which also could be age specific like retirement benefit offers.

Conclusion:-

- While statistically the clustering showed a natural overlap in customer behaviour, they were still highly interpretable.
- 2 large clusters emerged to be prime candidates to be targeted with cross sell and incentive driven offers to convert.
- 2 clusters showed high literacy hence showing potential to respond to advanced products.
- One of the clusters showed diversity in literacy and higher mean age, showing potential in being targeted with education and awareness campaigns and then with cross sell some other time. Due to higher average age, this group may also require retirement driven incentives to convert.