BANKING CUSTOMER SEGMENTATION

Peerapat Tancharoen • Data scientist • github.com/peerapat-t

EXECUTIVE SUMMARY

Business Problem

- Banking institutions struggle to capture customer attention and loyalty.
- Lack of personalized services limits conversion rates.

Solution

- Implement customer segmentation strategies to understand customer behavior.
- Leverage personalized marketing, tailored products to meet diverse customer needs.

Impact

- Increased customer satisfaction through personalized offerings.
- Optimized resource allocation and improved decision-making with actionable insights.

Challenges:

- Complexities in feature engineering and identifying relevant data attributes.
- Difficulty in interpreting and effectively utilizing segmented customer groups for strategic actions.

1. BUSINESS PROBLEM

1. **Low Customer Engagement:** Difficulty in capturing customer attention with generic offerings.
2. **Weak Customer Loyalty:** Challenges in creating personalized experiences to build lasting relationships.
3. **Limited Conversion Rates:** Struggles in turning prospects into customers due to insufficiently tailored services.
4. **Inefficient Resource Allocation:** Inability to effectively target resources to areas with the highest potential impact.
5. **Limited Customer Insights:** Lack of actionable data to understand customer behavior, hindering informed decision-making and product innovation.

2. SOLUTION

1. Condition-Based Segmentation

Advantages:

- Clear and interpretable criteria.
- Customization to specific business needs.
- Easily aligned with predefined goals.

Disadvantages:

- Rigid and may not adapt to changing behaviors.
- Oversimplifies customer behavior.
- May miss unexpected insights from data.

2. Clustering Models

Advantages:

- Unbiased and data-driven segmentation.
- Flexibility for changing data patterns.
- Discovery of hidden customer patterns.

Disadvantages:

- Less interpretable cluster labels.
- Sensitivity to initial conditions.
- Subjective choice of features.

2. SOLUTION

- This dataset comprises over 1 million transactions involving more than 800,000 customers of an Indian bank.
- The dataset encompasses various details including:

Age

Location

Gender

Average transaction time

Additional features

Average balance amount

Average transaction amount

Number of transaction

+ RFM features

3. RESULT

Customer's persona

Urban man

- Reside in large cities
- Male
- Oldest
- Highest account balance
- Most frequent transactions

30.95% of customers

Urban woman

- Reside in large cities
- Female
- Largest average transaction amount

9.56% of customers

Suburban man

- Reside in suburban areas
- Male
- Higher account balance compared to suburban women
- More frequent transactions than suburban women

45.38% of customers

Suburban woman

- Reside in suburban areas
- Female
- Youngest age
- Larger average transaction amount compared to suburban men

14.11% of customers

3. RESULT

Customer's persona

Suburban man

- Assume that men use bank transactions for daily expenses (smaller transaction amounts but higher frequency), we should consider collaborating with urban-zone shops, such as coffee shops, to provide promotions through bank transactions. Given the affluence of this group, we can also recommend financial asset opportunities.

Suburban woman

- Assume that women use bank transactions for purchases like fashion clothing or cosmetics (less frequent but larger amounts), we should explore collaboration with fashion and cosmetic shops in urban zones to offer promotions through bank transactions.

3. RESULT

Customer's persona

Suburban man

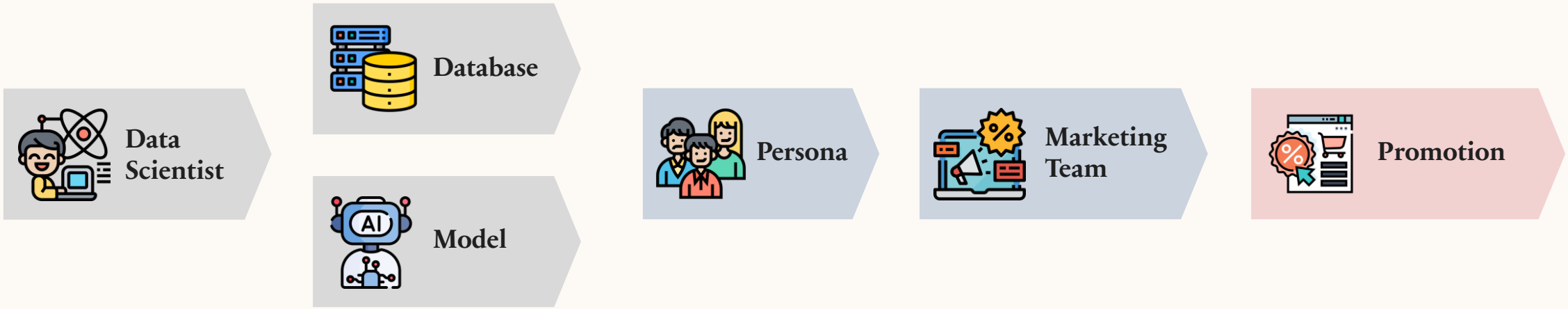
- Assume that men use bank transactions for daily expenses (smaller transaction amounts but higher frequency), a collaboration strategy could involve suburban-zone shops, such as coffee shops, to provide promotional offers using bank transactions.

Suburban woman

- Assume that women use bank transactions for purchases like fashion clothing or cosmetics (less frequent but larger amounts), collaborating with fashion and cosmetic shops in suburban zones to provide promotions through bank transactions could be effective.

4. APPLICATION

Collaboration framework



Step1: Data scientist team will focus on creating a clustering model. This model identifies different customer personas based on patterns in the data.

Step2: The model assigns each customer to a specific persona cluster.

Step3: Marketing team uses the personas to tailor campaigns. For example, each persona receives customized promotions based on their characteristics.

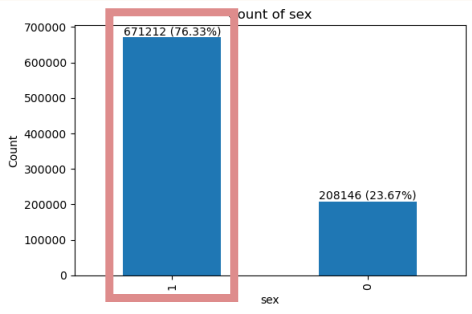
5. APPENDIX

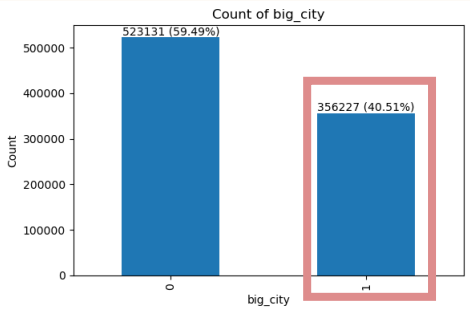
Technique

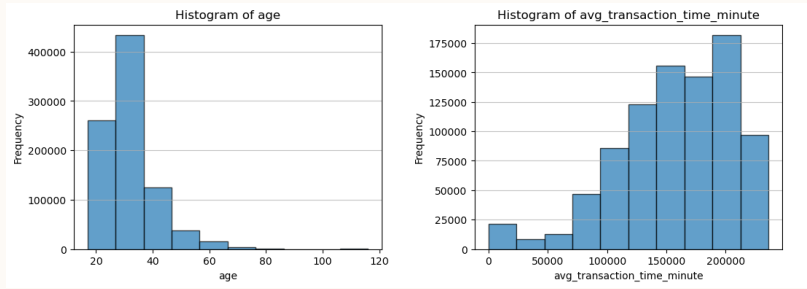
Step	Topic	Method
1	Normalization	• Use Min-Max scaling to normalize the data.
2	Clustering	• Implement K-Means clustering to group similar data points.
3	Selecting K	• Determine the optimal number of clusters using the elbow method. • Analyze attribute means to select a meaningful value for K in the context of marketing segmentation.
4	Visualization	• Utilize PCA to visualize the clustered data.

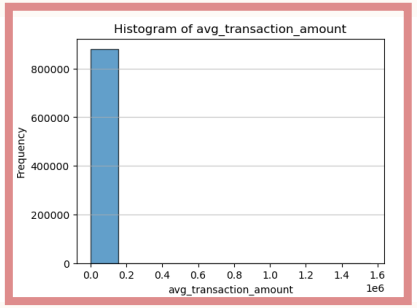
5. APPENDIX

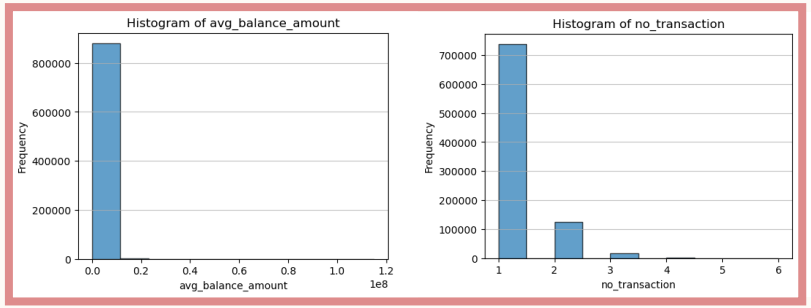
Exploratory data analysis (EDA)









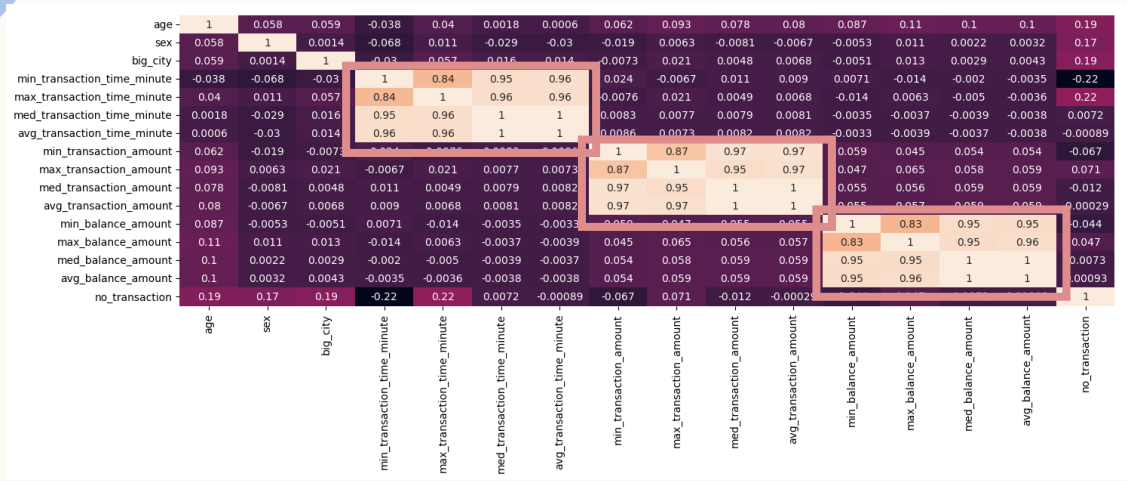


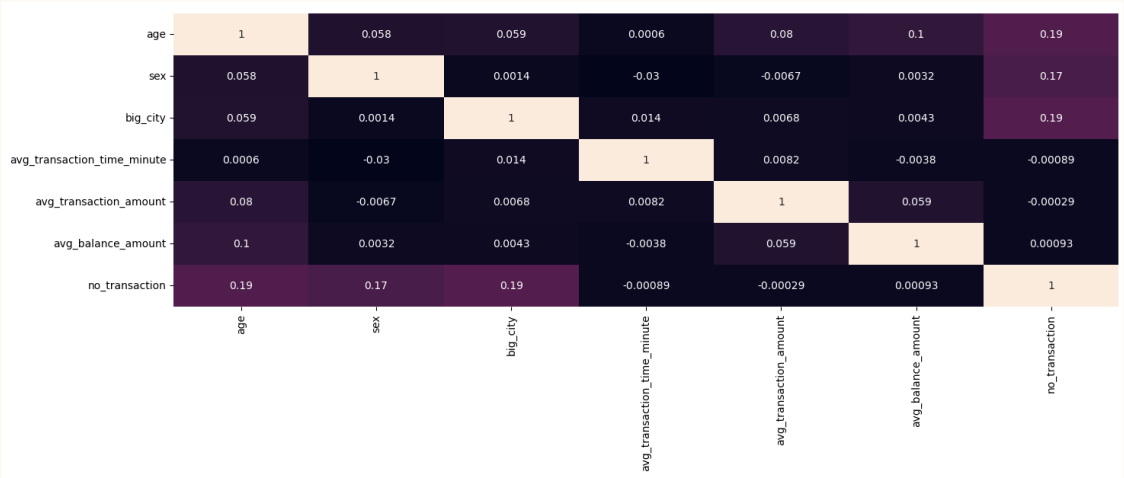
- Both attributes appear significantly different, making them important features for distinguishing each cluster. 76% of customers are men, and 59% of customers live in big cities.

- However, transaction amounts and balance amounts are quite similar among customers.

5. APPENDIX

Exploratory data analysis (EDA)

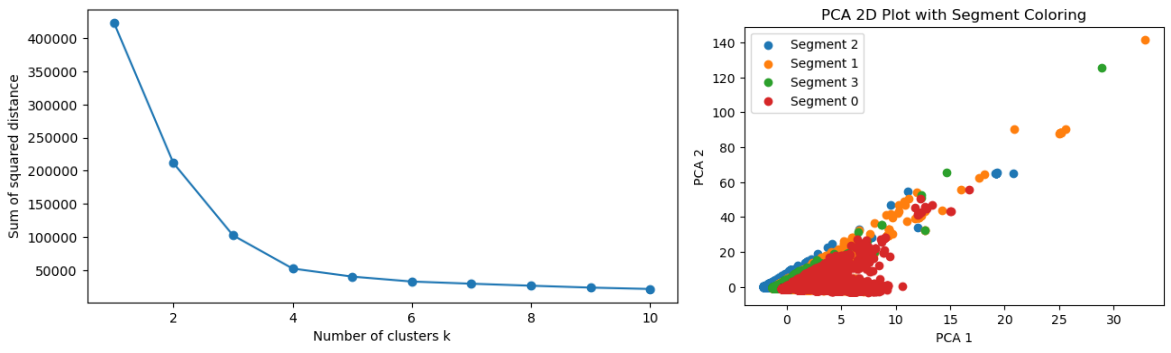




- We create Max, Min, Med, and Avg for each feature, but they have a high correlation among them, so I'm decide to drop it.

5. APPENDIX

Select number of cluster



segment	0	1	2	3
age	32.818505	31.500730	30.584142	31.041394
avg_balance_amount	121512.315040	112508.162380	109794.062256	110448.322763
avg_transaction_amount	1597.267205	1505.809306	1611.175622	1692.120499
avg_transaction_time_minute	157301.190126	155648.438461	159508.739854	160188.234572
big_city	1.000000	0.000000	0.000000	1.000000
no_transaction	1.354916	1.139888	1.027040	1.077138
sex	1.000000	1.000000	0.000000	0.000000

- The elbow method is a technique used in cluster analysis to determine the optimal number of clusters for a given dataset. The basic idea is to plot the within-cluster sum of squares (WCSS) against the number of clusters.
- WCSS is a measure of how spread out the data points within each cluster are, and it quantifies the compactness of the clusters.
- K = 4 is chosen as the optimal number of clusters based on the elbow method because at that point in the plot, adding more clusters doesn't significantly reduce the within-cluster sum of squares.