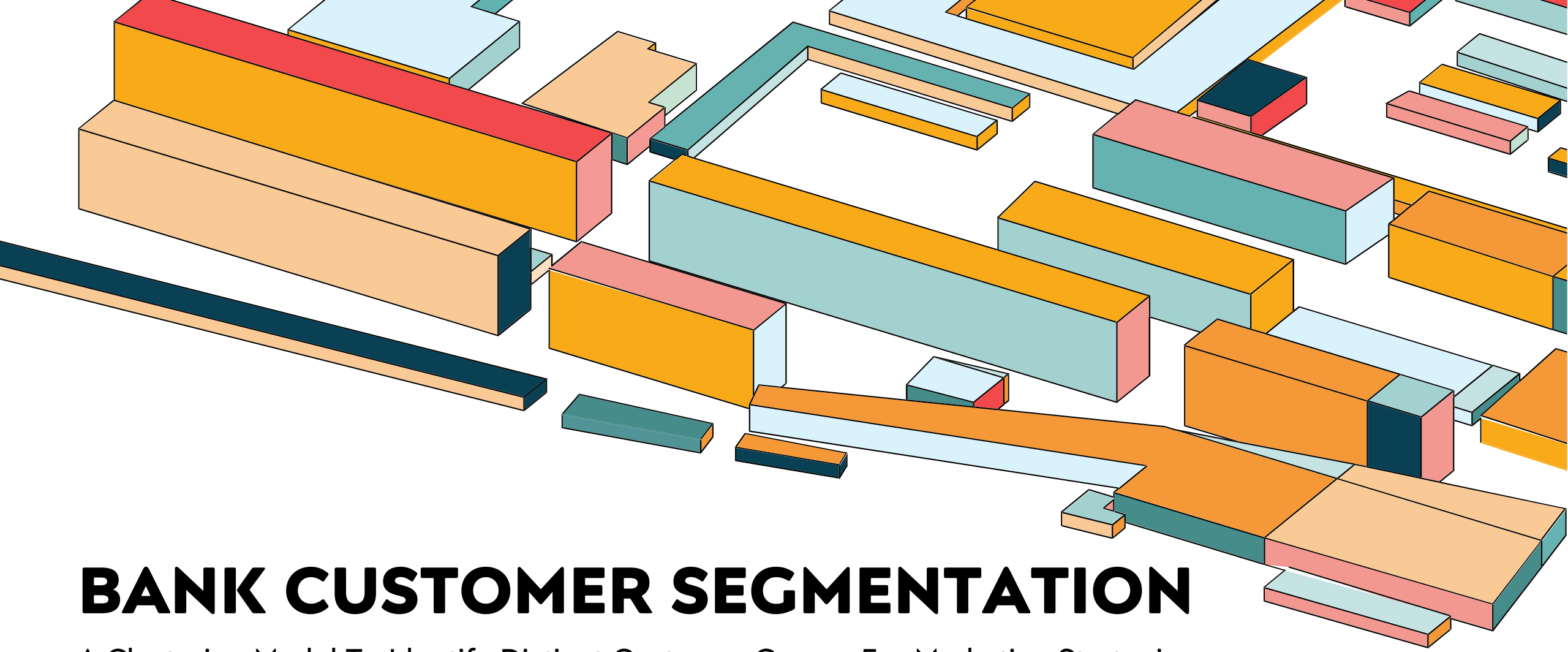


BANK CUSTOMER SEGMENTATION

A Clustering Model To Identify Distinct Customer Groups For Marketing Strategies.

Peerapat.t
For project's material please visit : github.com/peerapat-t

BANK CUSTOMER SEGMENTATION

Problem context

- In the ever-changing world of banking, personalized services capture customer attention, foster loyalty, and increase conversion rates.
- Additionally, they optimize resource allocation and provide valuable insights for informed decision-making and service enhancement.

Challenges

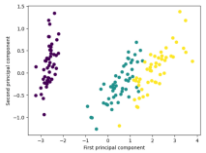
- Feature engineering
- Interpret the segment

Tools

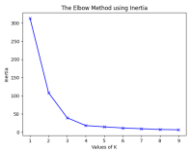
python

pandas

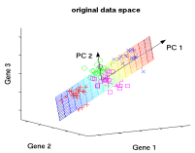
Methods



KMeans



Elbow method



PCA

Business impact

Customer segmentation can significantly impact business by allowing targeted marketing efforts, tailored products/services, and improved customer satisfaction.

TABLE OF CONTENT

1. PROBLEM STATEMENT

- In the competitive banking industry, the challenge is to implement effective personalized services by:
  - **Increasing Engagement:** Deliver tailored solutions to capture customer attention.
  - **Building Loyalty:** Create personalized experiences to foster enduring relationships.
  - **Boosting Conversions:** Offer customized services to turn prospects into customers.
  - **Efficient Resource Allocation:** Target resources where they'll yield the best results.
  - **Gaining Insights:** Use data analysis to understand customer behavior for informed decision-making and product refinement.

2. BUSINESS VALUE

- Implementing personalized banking services can provide significant business value, including:
  - **Enhanced Engagement:** Tailored services boost customer satisfaction and loyalty.
  - **Improved Retention**: Personalization reduces churn rates and extends customer lifetime value.
  - **Increased Revenue:** Customization drives customer acquisition and cross-selling success.
  - **Efficient Resources:** Targeted strategies optimize resource allocation, reducing waste.
  - **Informed Decisions:** Data-driven insights inform product development and marketing, enhancing competitiveness.

3. METHODOLOGY

Condition-Based Segmentation

- **Advantages:**
  1. Clear and interpretable criteria.
  2. Customization to specific business needs.
  3. Easily aligned with predefined goals.
- **Disadvantages:**
  1. Rigid and may not adapt to changing behaviors.
  2. Oversimplifies customer behavior.
  3. May miss unexpected insights from data.

Clustering Models

- **Advantages:**
  - Unbiased and data-driven segmentation.
  - Automation and scalability.
  - Flexibility for changing data patterns.
  - Discovery of hidden customer patterns.
- **Disadvantages:**
  - Less interpretable cluster labels.
  - Sensitivity to initial conditions.
  - Subjective choice of features.

3. METHODOLOGY

Features

- This dataset comprises over 1 million transactions involving more than 800,000 customers of an Indian bank. The dataset encompasses various details including:
  - Customer age
  - Location
  - Gender
  - Average transaction time
  - Average account balance amount
  - Average transaction amount
  - Number of transaction

} RFM

4. RESULT

Customer segment

- Following the model's results, we have 4 groups of customers.

1. Urban man
(30.95% of customers)

1. Reside in large cities
2. Male
3. Oldest
4. Highest account balance
5. Most frequent transactions

2. Urban woman
(9.56% of customers)

1. Reside in large cities
2. Female
3. Largest average transaction amount

3. Suburban man
(45.38% of customers)

1. Reside in suburban areas
2. Male
3. Higher account balance compared to suburban women
4. More frequent transactions than suburban women

4. Suburban woman
(14.11% of customers)

1. Reside in suburban areas
2. Female
3. Youngest age
4. Larger average transaction amount compared to suburban men

4. RESULT

Customer segment

1. Urban 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.

2. Urban 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.

4. RESULT

Customer segment

3. 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.

4. 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.

5. CONCLUSIONS/RECOMMENDATIONS

Conclusions

- Based on the results from the model, we have identified four distinct customer groups: urban men, urban women, suburban men, and suburban women. These groups exhibit differences in various attributes, including their account balances and transaction frequencies.

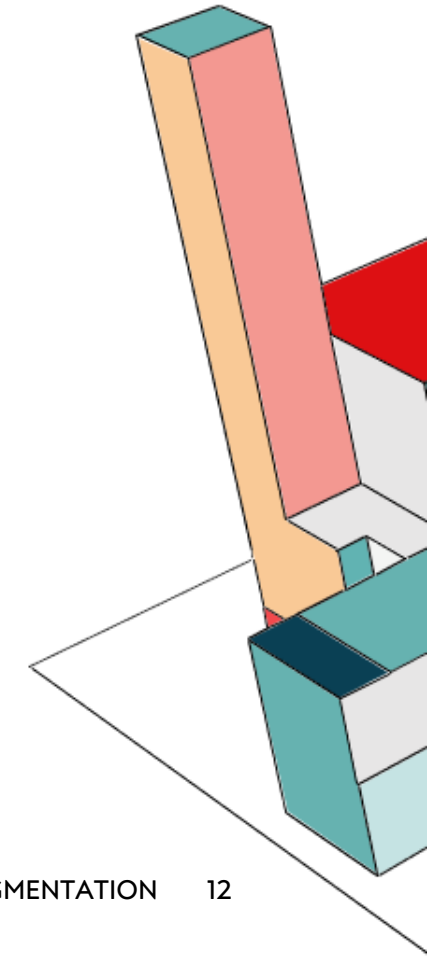
Recommendation

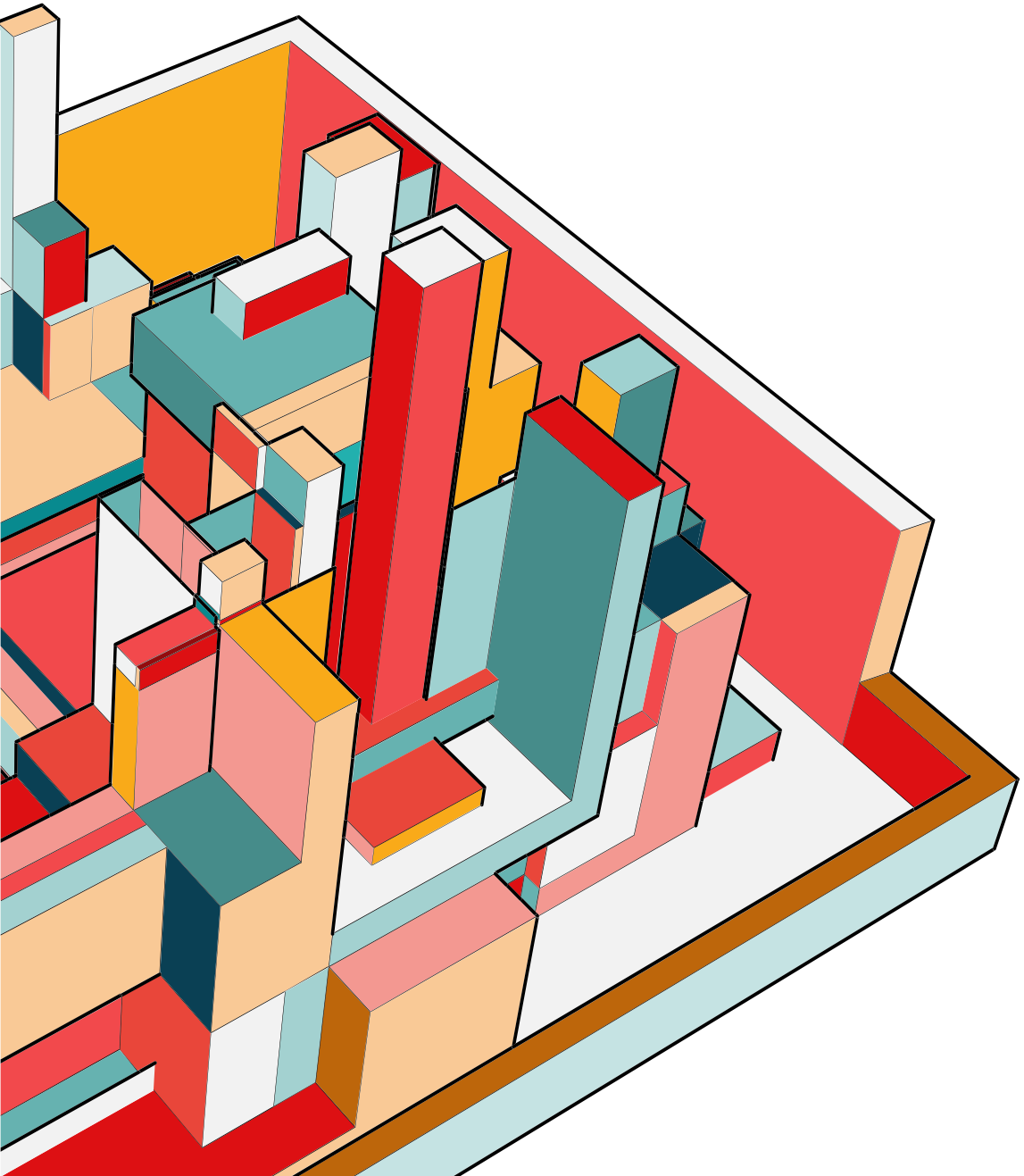
- Leveraging these insights, we can develop personalized marketing strategies aimed at enhancing customer relationships and boosting the company's revenue.

6. FUTURE WORK

- Try another algorithm, such as DBSCAN or hierarchical clustering, and compare the results of them.
-
- Create new features to better understand customer behavior. However, this will require re-clustering.

- Monitor the distribution of new data; if the distribution changes significantly, re-clustering should be considered





END OF PRESENTATION

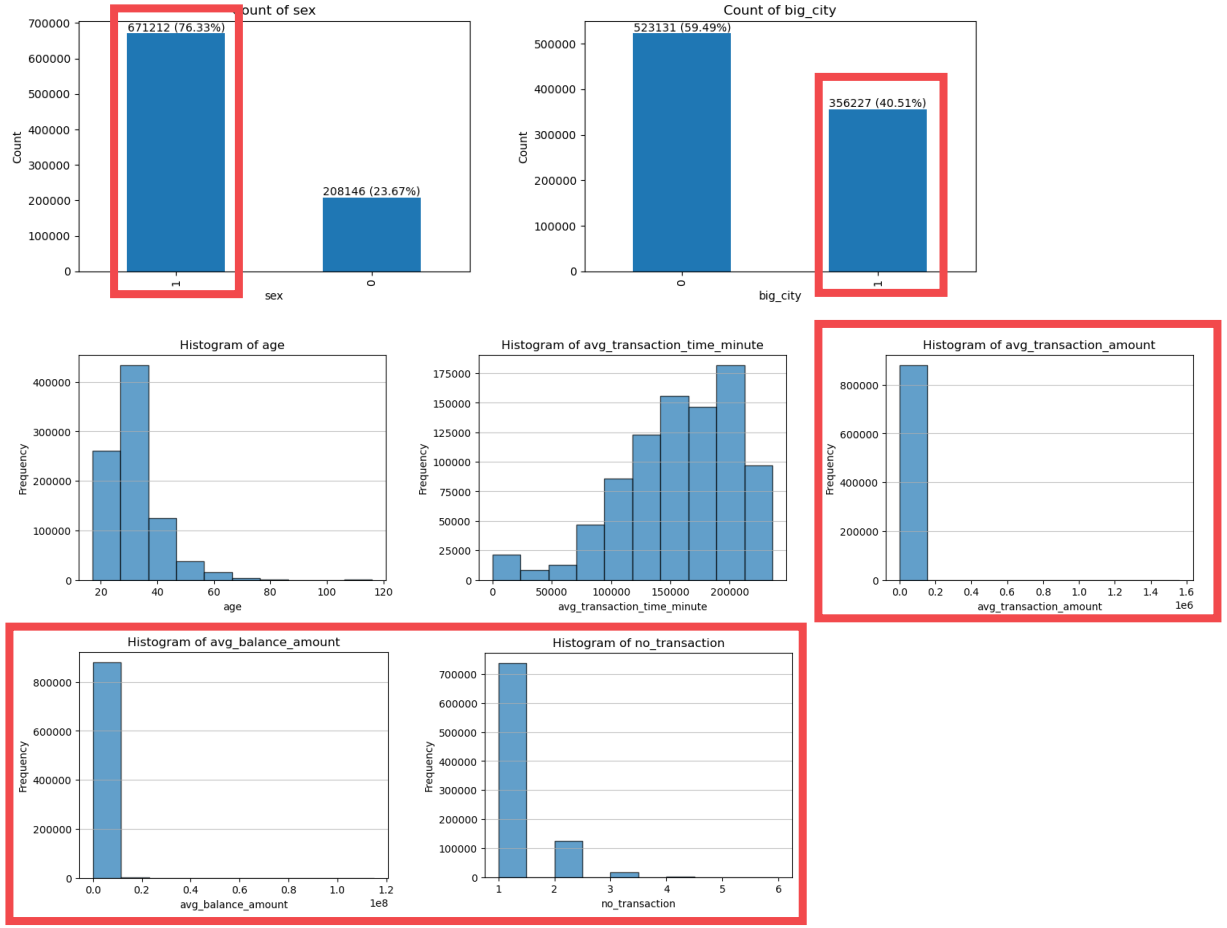
7. APPENDIX

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.

7. APPENDIX

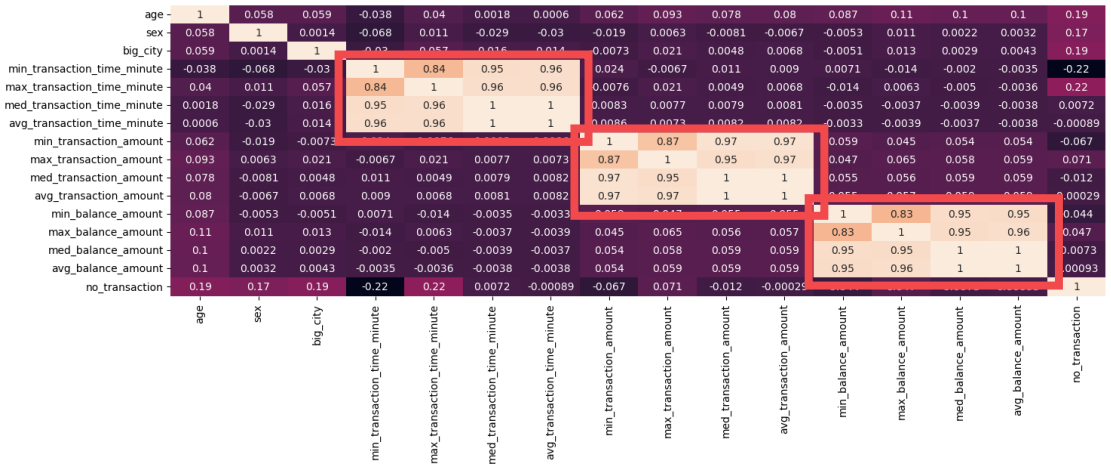
Exploratory data analysis (EDA)

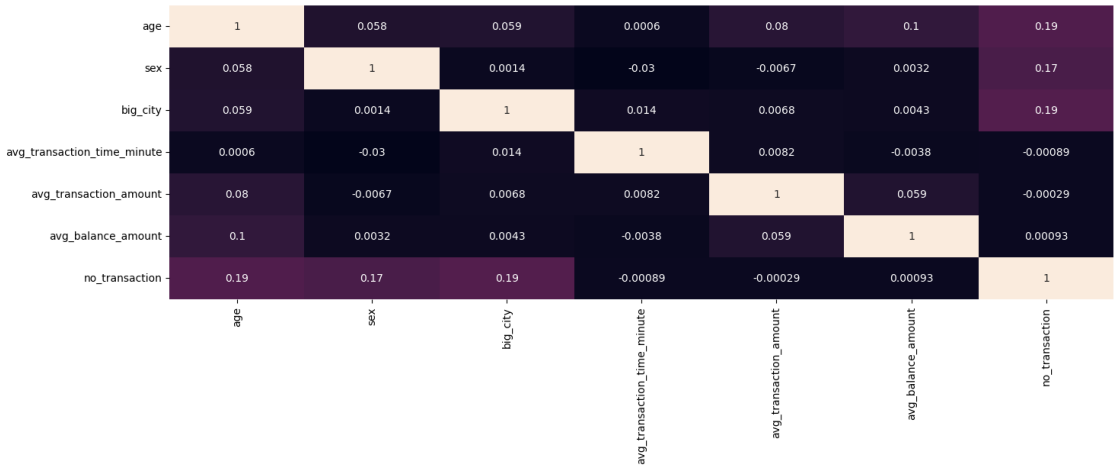


- Both attributes appear significantly different, making them important features for distinguishing each cluster. 76% of customers are men, and 59% of customers live in big cities.
- However, transaction amounts and balance amounts are quite similar among customers.

7. APPENDIX

Exploratory data analysis (EDA)





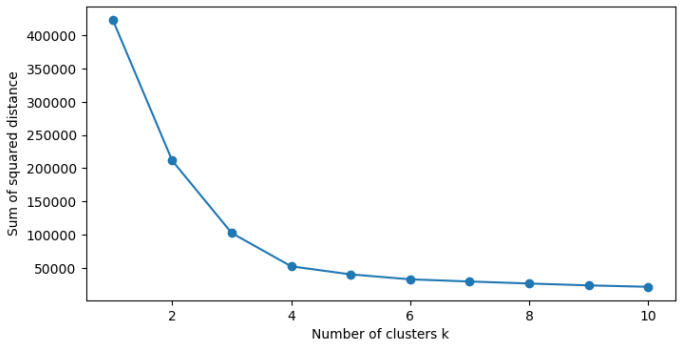
- We create Max, Min, Med, and Avg for each feature, but they have a high correlation among them, so I'm decide to drop it.

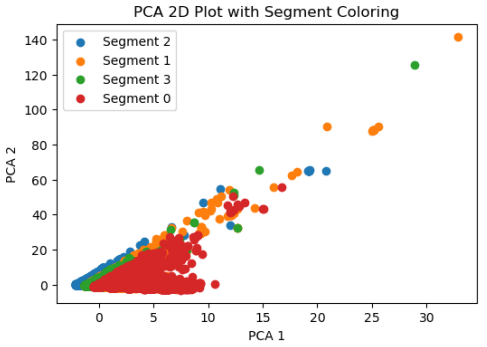
7. APPENDIX

Select number of cluster

- 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.





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

7. APPENDIX

Reference

- https://medium.com/grabngoinfo/4-clustering-model-algorithms-in-python-and-which-is-the-best-7f3431a6e624

- https://grabngoinfo.com/k-means-clustering-example-code-using-python-scikit-learn/

- https://grabngoinfo.com/5-ways-for-deciding-number-of-clusters-in-a-clustering-model/

8. ABOUT AUTHOR

- Peerapat Tancharoen holds a Bachelor's degree in Economics from Srinakharinwirot University, graduating with first honors and a GPA of 3.67. He also earned a Master's degree in Economics from Thammasat University, achieving a GPA of 3.98.

Contact me on

- linkedin.com/in/peerapat-tancharoen-664759220/
- peerapat.tcr@gmail.com