BANK CUSTOMER SEGMENTATION

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

TABLE OF CONTENT

1. BUSINESS PROBLEM

1.1 Personalized marketing

- In the dynamic landscape of banking, marked by intense competition and diverse customer preferences, personalized banking services offer several benefits:

  - **Heightened Engagement:** Tailored banking services capture customer attention and engagement by presenting relevant solutions and information aligned with individual financial needs.
  - **Strengthened Customer Loyalty:** Personalized experiences and solutions cultivate stronger customer loyalty, fostering enduring relationships with the bank.
  - **Increased Conversion Rates:** Customized service approaches enhance the likelihood of transforming potential clients into active customers through targeted guidance and offerings.
  - **Optimized Resource Allocation:** Tailored banking strategies streamline resource allocation by focusing on segments with the highest likelihood of positive response.
  - **Deeper Insights:** Utilizing clustering models provides profound insights into customer financial behaviors, facilitating informed decision-making and the refinement of banking products and services.
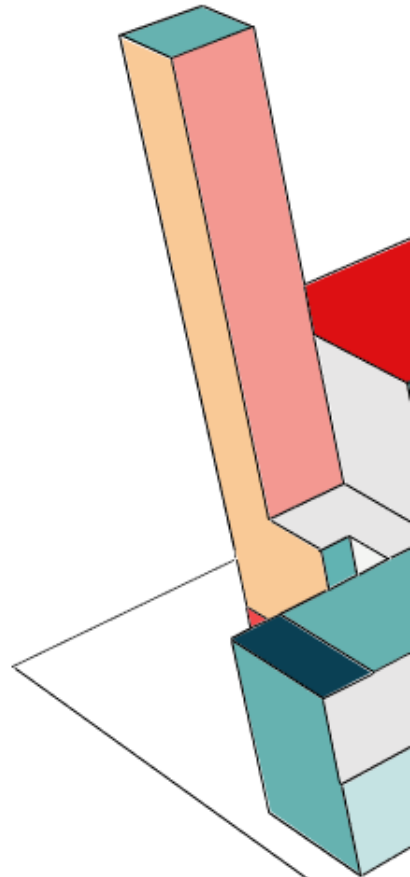
2. HOW TO SOLVE THIS PROBLEM

2.1 Clustering model

- In the realm of banking, segmentation models serve as a crucial tool for comprehending customers by categorizing them into segments founded on common traits and behaviors.

- Leveraging data-driven methodologies, these models pinpoint patterns and resemblances among customers, empowering banks to meticulously personalize their strategies and outreach efforts.
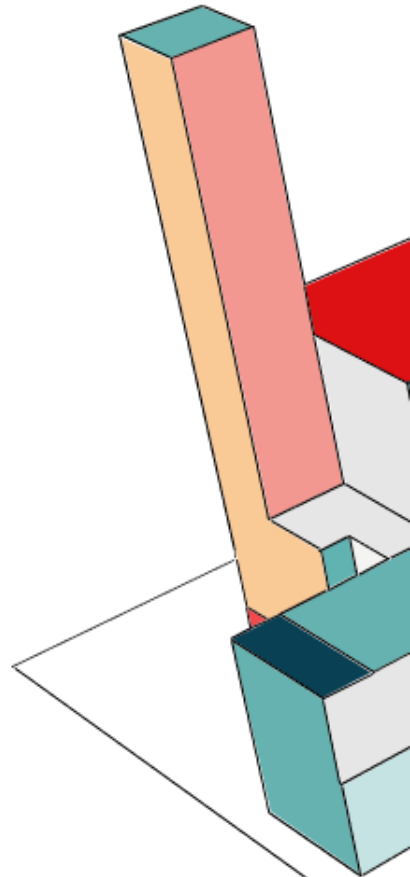
3. DATA

3.1 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 (date of birth)
  - Location
  - Gender
  - Account balance during the transaction
  - Transaction amount, and more.
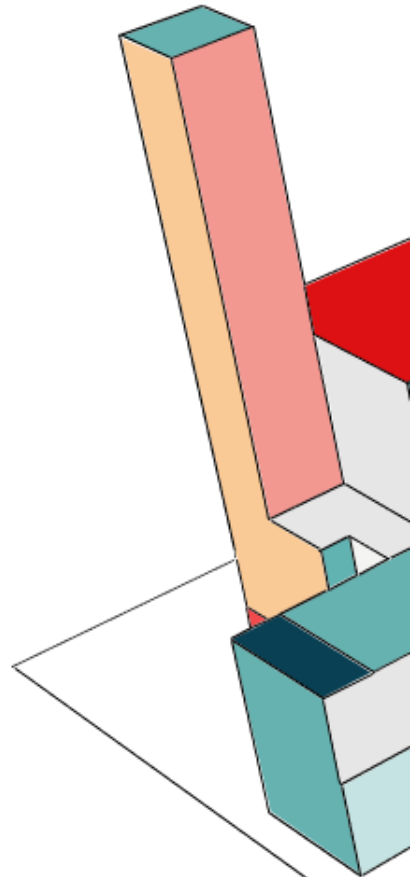
4. SOLUTION

4.1 Customer's 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
(45.38% of customers)**

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

**3. Suburban man
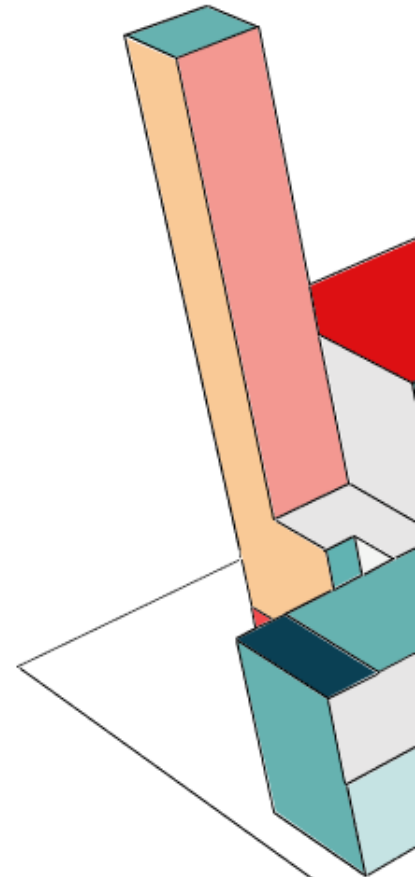(14.11% of customers)**

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
(9.56% of customers)**

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

4. SOLUTION

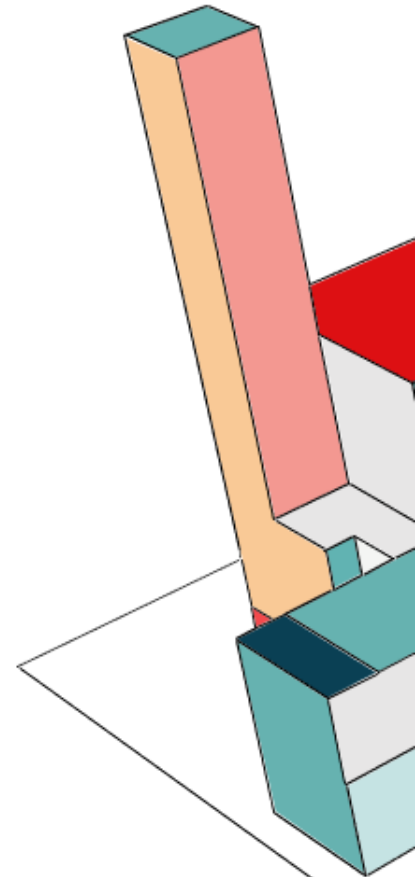## 4.2 Promotion and campaign related

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. SOLUTION

## 4.2 Promotion and campaign related

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.
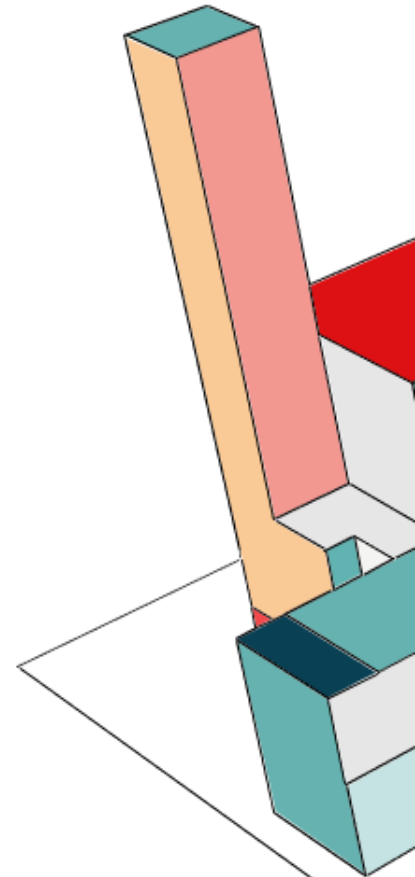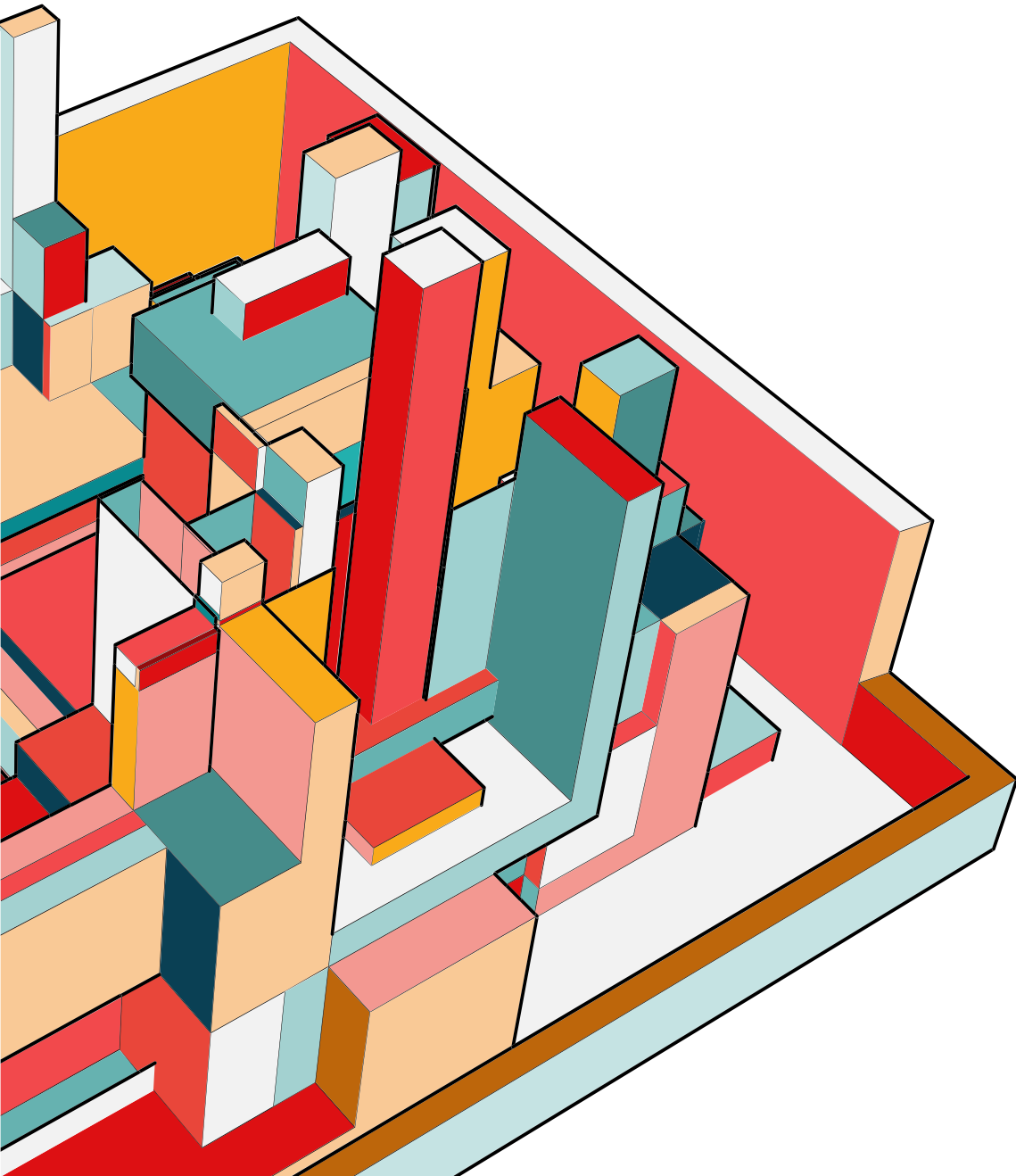
4. SOLUTION

4.3 Use of model

- Save centroids and re-fit the model in the future. If the data shifts too much, retraining will be required.
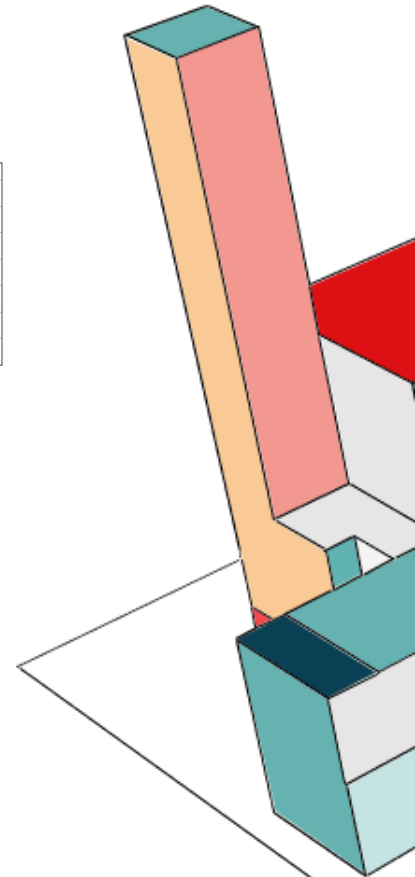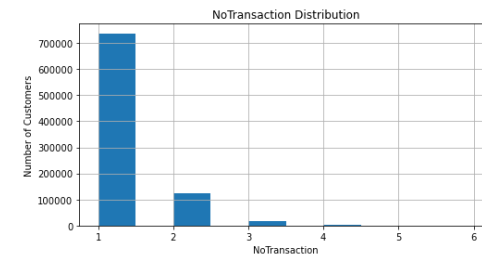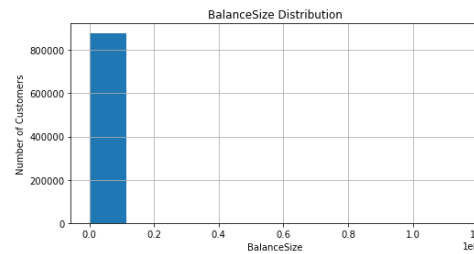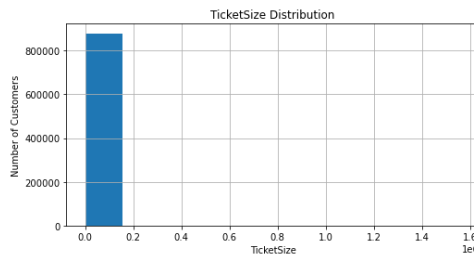
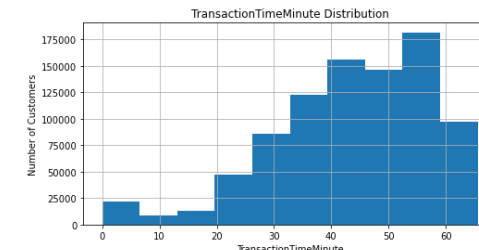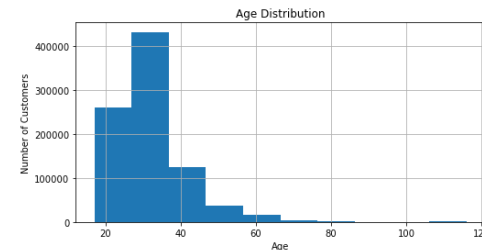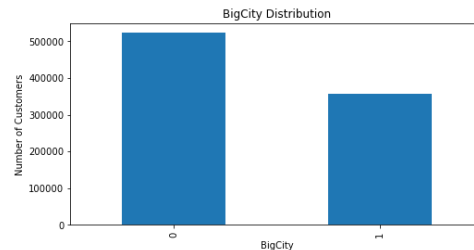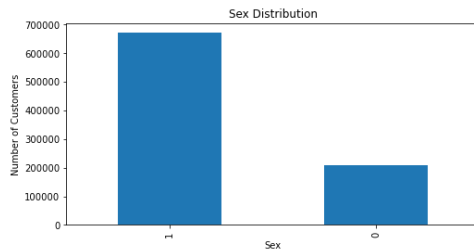| Train	→	Fit
		Using saved centroid
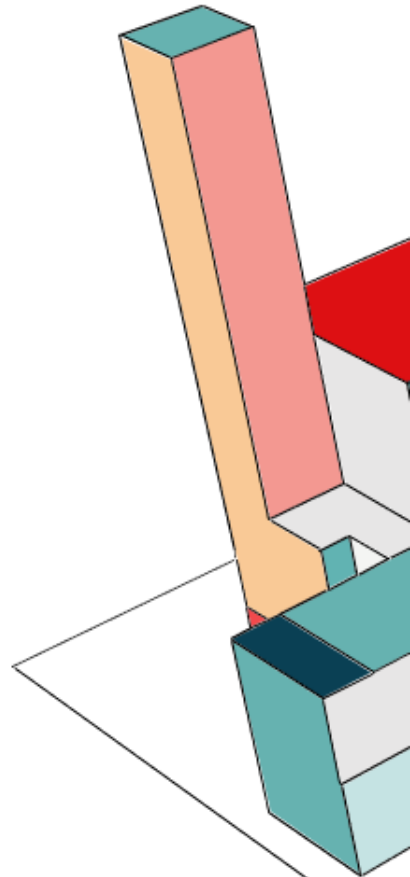
END OF BUSINESS SESSION

4. MODEL

4.1 Data understanding

- Sex, BigCity, Age, and TransactionTimeMinute may contain information for clustering, while other values are too similar.

4. MODEL

4.2 Technique used

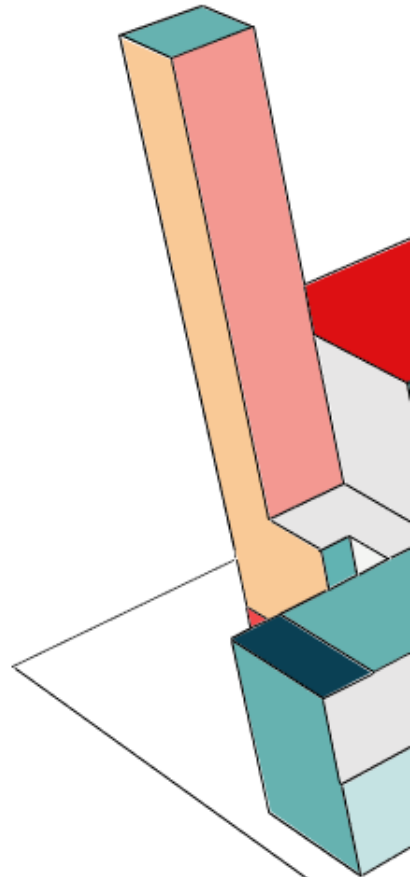| Step	Topic	Method
1	Normalization	• Use Min-Max scaling to normalize the data.
2	Clustering	• Implement K-Means clustering to group similar data points.
3	Selecting K	• Determine the optimal number of clusters using the elbow method. • Analyze attribute means to select a meaningful value for K in the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

4. MODEL

4.3 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.

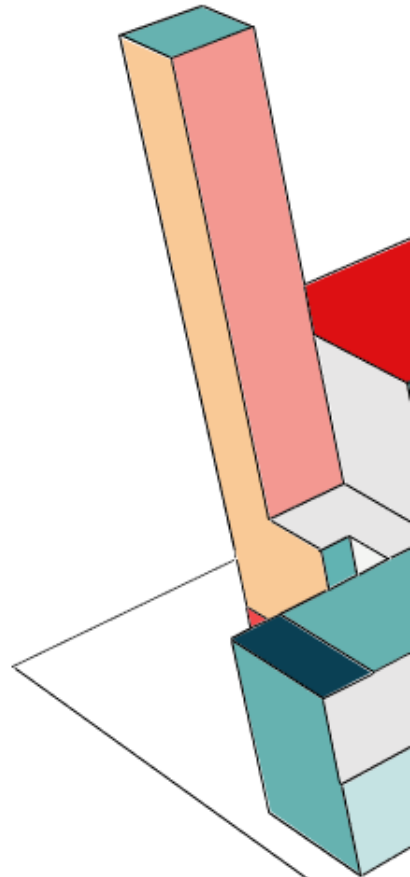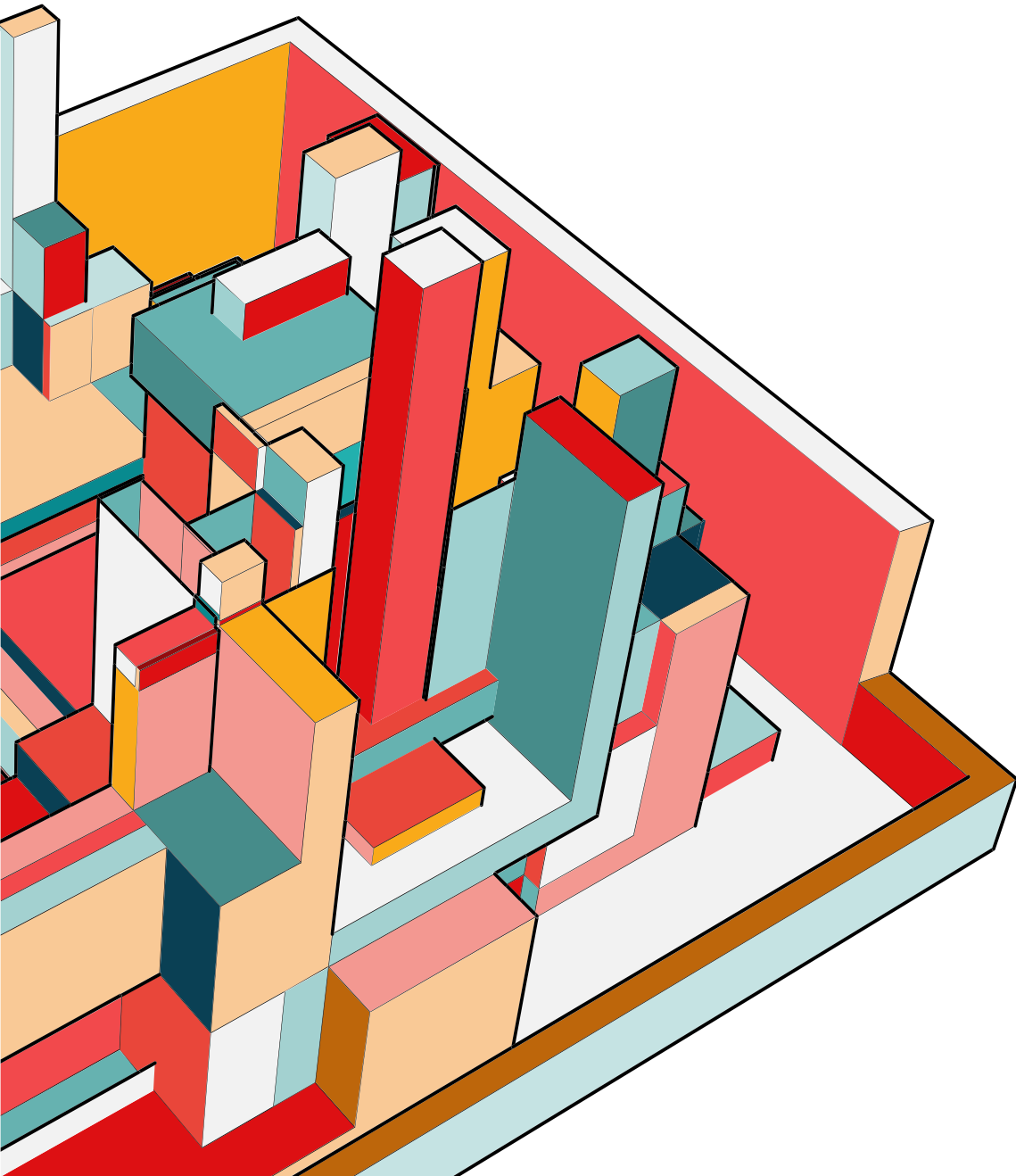| segment	0	1	2	3
Age	32.932420	31.672511	30.653708	31.127935
BalanceSize	121512.315040	112508.162380	109794.062256	110448.322763
BigCity	1.000000	0.000000	0.000000	1.000000
NoTransaction	1.354916	1.139888	1.027040	1.077138
Sex	1.000000	1.000000	0.000000	0.000000
TicketSize	1597.267205	1505.809306	1611.175622	1692.120499
TransactionTimeMinute	43.694775	43.235677	44.307983	44.496732

5. FUTURE WORK

- Try another algorithm, such as DBSCAN or hierarchical clustering.

- 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

6. 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/

END OF PRESENTATION

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