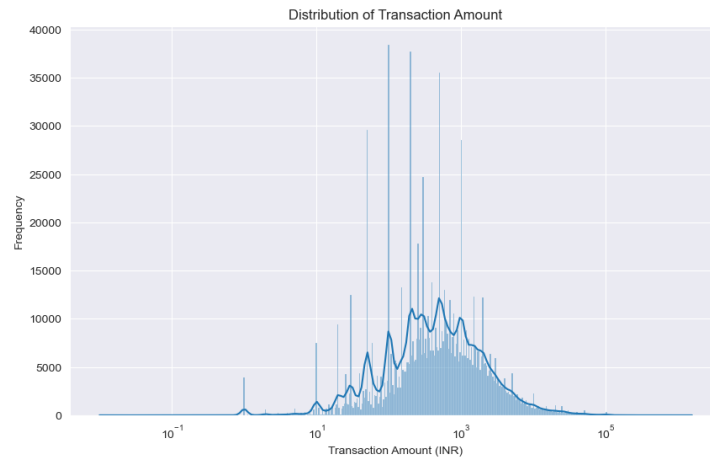
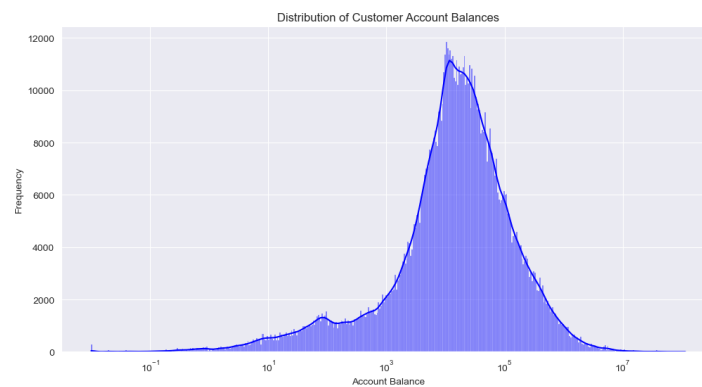


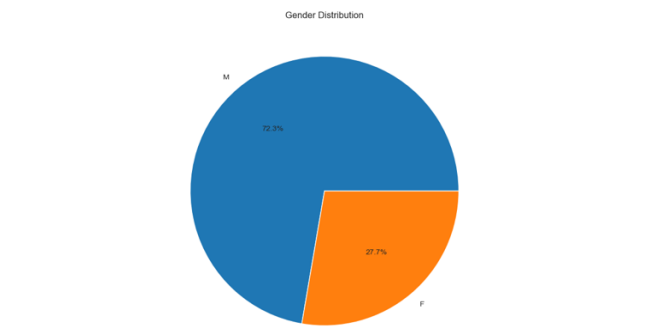
DSAA 5002 - Data Mining and Knowledge Discovery in Data Science**Final Exam Report – Q6 Bank Customer Clustering****50015940 Jiaxiang Gao****1 Visualization****1.1 Distribution of Transaction Amount (INR)**

This figure shows the distribution of transaction amounts in Indian Rupees (INR) for a set of banking transactions. The distribution is plotted on a logarithmic scale for the transaction amount. From the histogram, it is evident that most transactions are of a smaller amount, as indicated by the higher frequency of transactions on the lower end of the scale. The long tail to the right suggests that there are relatively few very large transactions.

1.2 Distribution of Customer Account Balances

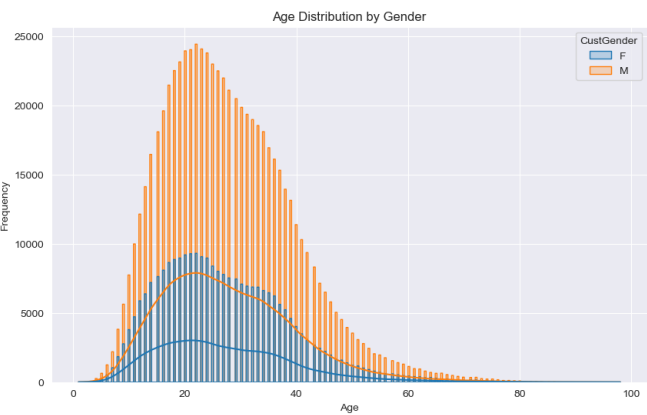
This figure shows that a large majority of customers have lower account balances. The frequency of customers gradually decreases as the account balance increases.

1.3 Distribution of Customer Gender



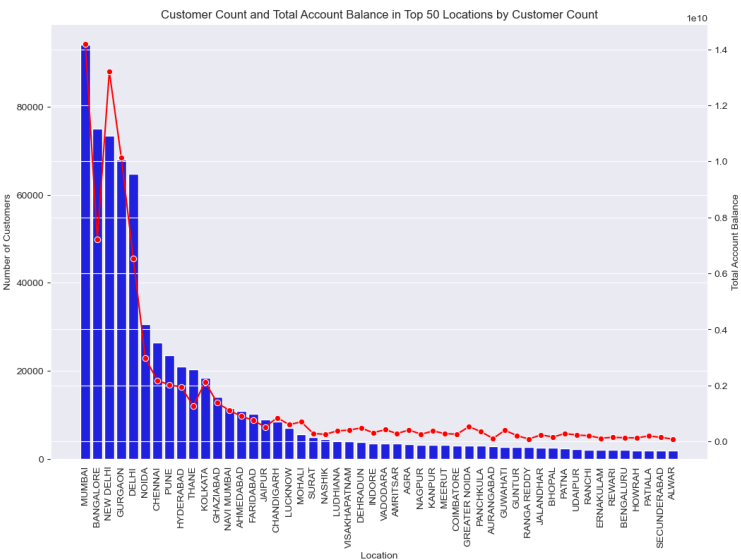
This figure shows that 72.3% of customer are male and 27.7 % of customer is female.

1.4 Distribution of Customer Age



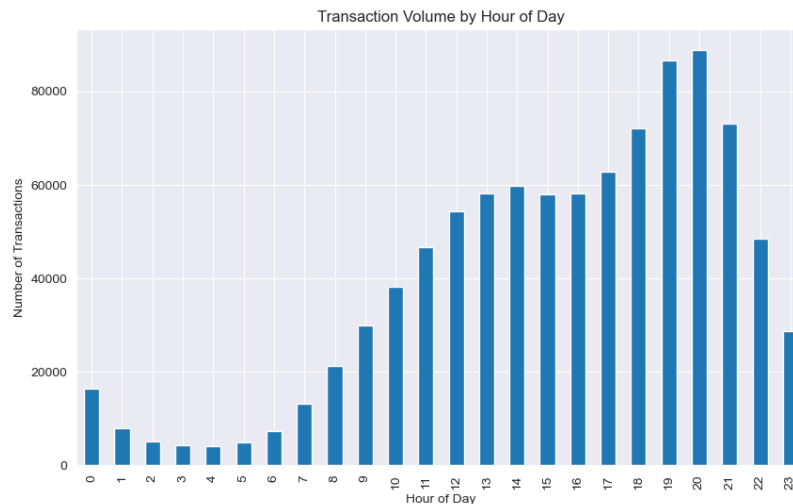
This Figure shows that most customer’s age is between 18 – 40.

1.5 Customer Count and Total Account Balance in Top 50 Locations by Customer Count



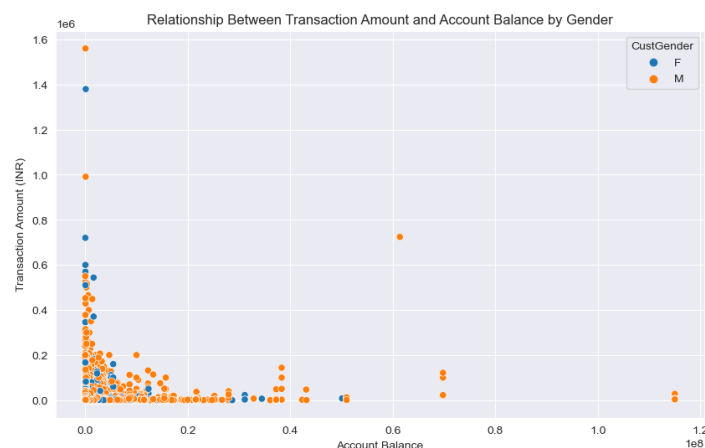
This figure shows that the total balance in each region is generally positively correlated with the number of accounts. However, the number of accounts in BANGALORE region is relatively large, but the balance is relatively small. This may indicate that customers in this region are more inclined to use other bank services, such as loans, rather than savings.

1.6 Transaction Volume by Hour of Day



This figure shows that most transactions occur between 18:00 and 21:00.

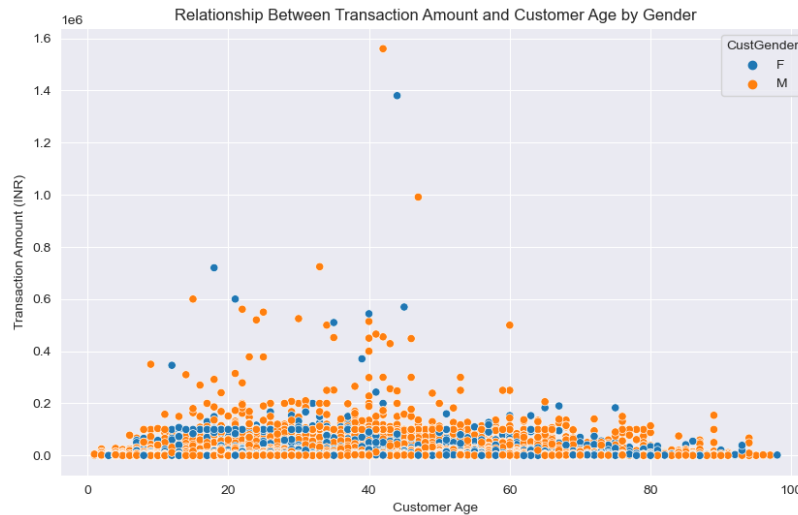
1.7 Relationship Between Transaction Amount and Account Balance by Gender



The figure shows that:

1. There is a dense concentration of data points towards the lower end of both account balances and transaction amounts for both genders, indicating that most transactions are of smaller amounts and occur in accounts with smaller balances.
2. Fewer data points appear as the transaction amount and account balance increase, indicating that larger transactions are less common and typically involve customers with higher account balances.
3. There are a few points that stand out from the main concentration, representing individuals with higher account balances and transaction amounts. These could be outliers or high net-worth individuals who could be targeted for specialized banking services.

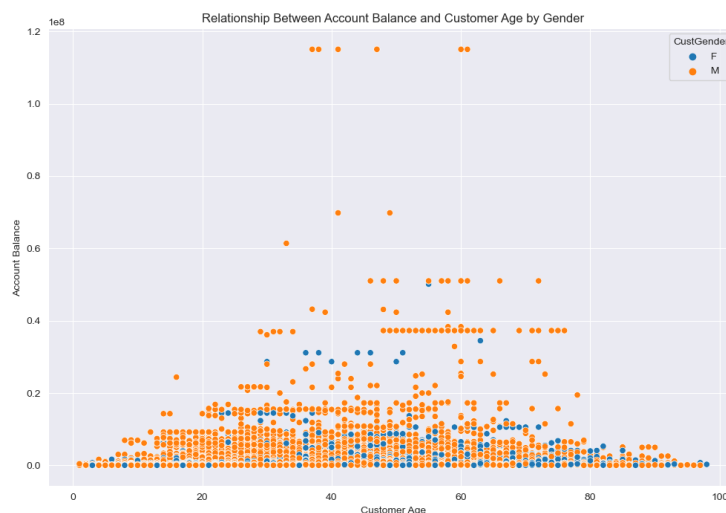
1.8 Relationship Between Transaction Amount and Customer Age by Gender



The figure shows that:

1. Transactions are spread across a wide range of customer ages, from young adults to the elderly.
2. There is a wide range of transaction amounts across all ages, but most transactions, for both genders, appear to be lower in value, with the distribution becoming sparser as the transaction amount increases.
3. A small number of high-value transactions are evident across various age groups. These could represent major financial activities such as property purchases, investments, or large personal expenditures.

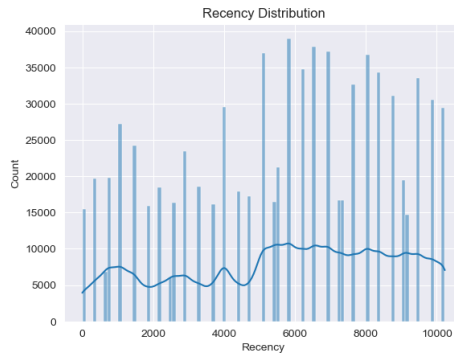
1.9 Relationship Between Account Balance and Customer Age by Gender



The figure shows that:

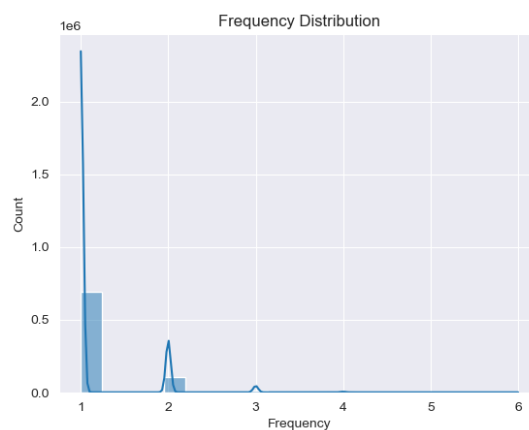
1. The balance is relatively concentrated between the ages of 20 and 80
2. Compared to female customer, male customer have higher balances

1.10 Recency Distribution from RFM Analysis



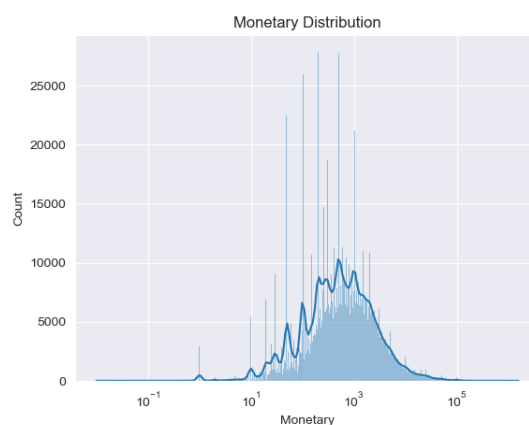
This histogram shows how recently customers have engaged with the business. A lower recency value indicates more recent activity. The distribution might suggest several peaks and valleys, indicating possible periodic trends in customer engagement or reflecting specific marketing campaigns or seasonal effects.

1.11 Frequency Distribution from RFM Analysis



This graph depicts the frequency of transactions per customer. Most customers seem to have a low frequency of transactions, with a steep drop-off as frequency increases.

1.12 Monetary Distribution from RFM Analysis



This histogram represents the total monetary value spent by customers, displayed on a logarithmic scale. It shows that many customers spend relatively small amounts, with fewer customers making up the higher-end spenders.

2 Business insights:

Based on the visualization of the various distribution figures and histograms, here are business insights that can be derived:

- 2.1 **Focus on Small Transactions:** Since most banking transactions are of a smaller amount, the bank should tailor its services and marketing strategies towards facilitating and encouraging small-scale transactions.
- 2.2 **Risk Management:** The transactions with very high values, although less frequent, could require additional scrutiny for risk management, as they could represent higher risk for fraudulent activity
- 2.3 **Product Offerings for Lower Account Balances:** With most customers having lower account balances, the bank should consider financial products that are accessible and beneficial to this demographic, such as low-balance savings accounts or micro-investment opportunities.
- 2.4 **Gender-Specific Financial Products:** Given the significant difference in the proportion of male and female customers (72.3% male and 27.7% female), the bank could consider developing gender-specific financial products or services to better cater to the needs of both groups.
- 2.5 **Youth-Oriented Services:** With most customers aged between 18-40, the bank could focus on services and products that appeal to younger demographics, like student loans, first-time home buyer programs, or digital banking solutions.
- 2.6 **Regional Strategy for Bangalore:** The observation that Bangalore has a large number of accounts but relatively small total balances suggest a potential market for services other than savings, like personal loans or credit facilities.
- 2.7 **Peak Transaction Hours Targeting:** The high volume of transactions between 18:00 and 21:00 indicates these are peak banking hours. The bank could optimize staffing or enhance digital service capabilities during these hours for better customer service.
- 2.8 **Tailored Services for High Net-worth Individuals:** The presence of outliers with high account balances and transaction amounts suggests a segment of high net-worth individuals who could be targeted with specialized banking services like wealth management or private banking.
- 2.9 **Age-Diverse Transaction Patterns:** The wide range of transaction amounts across all ages implies that the bank's marketing and product development should be inclusive of all age groups, from young adults to the elderly.
- 2.10 **Encouraging Increased Transaction Frequency:** With most customers having a low frequency of transactions, there is an opportunity to develop strategies or incentives to encourage customers to engage more frequently with the bank's services.

3 Customer Clustering Using Three Distinct Algorithms

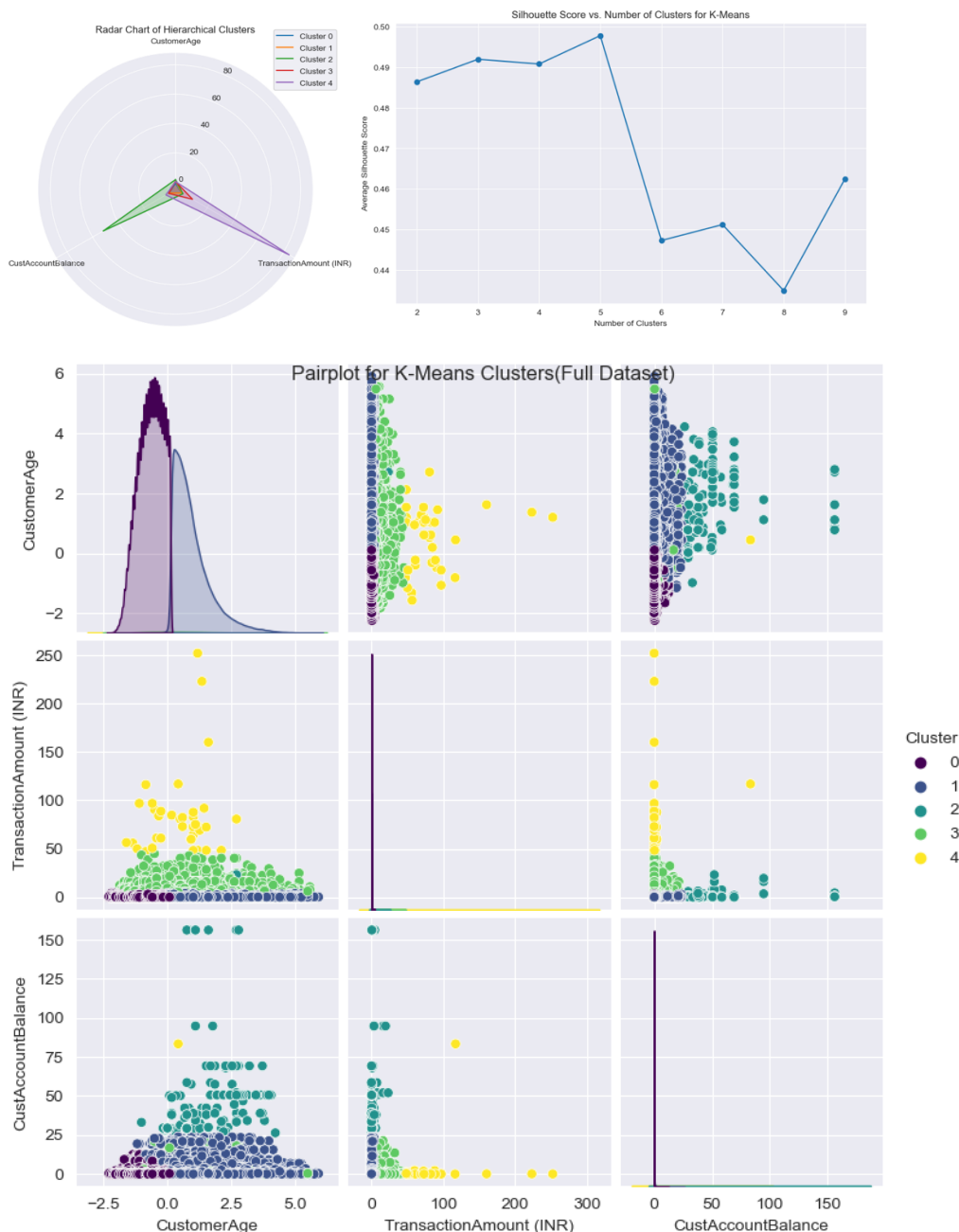
For this task, I use three features: 'CustomerAge', 'TransactionAmount (INR)' and 'CustAccountBalance'.

For K-Means and Hierarchical clustering, we used sample data to calculate the silhouette score, and then use the silhouette score to determine the number of clusters.

For K-Means, We Use the entire dataset to calculate the clusters.

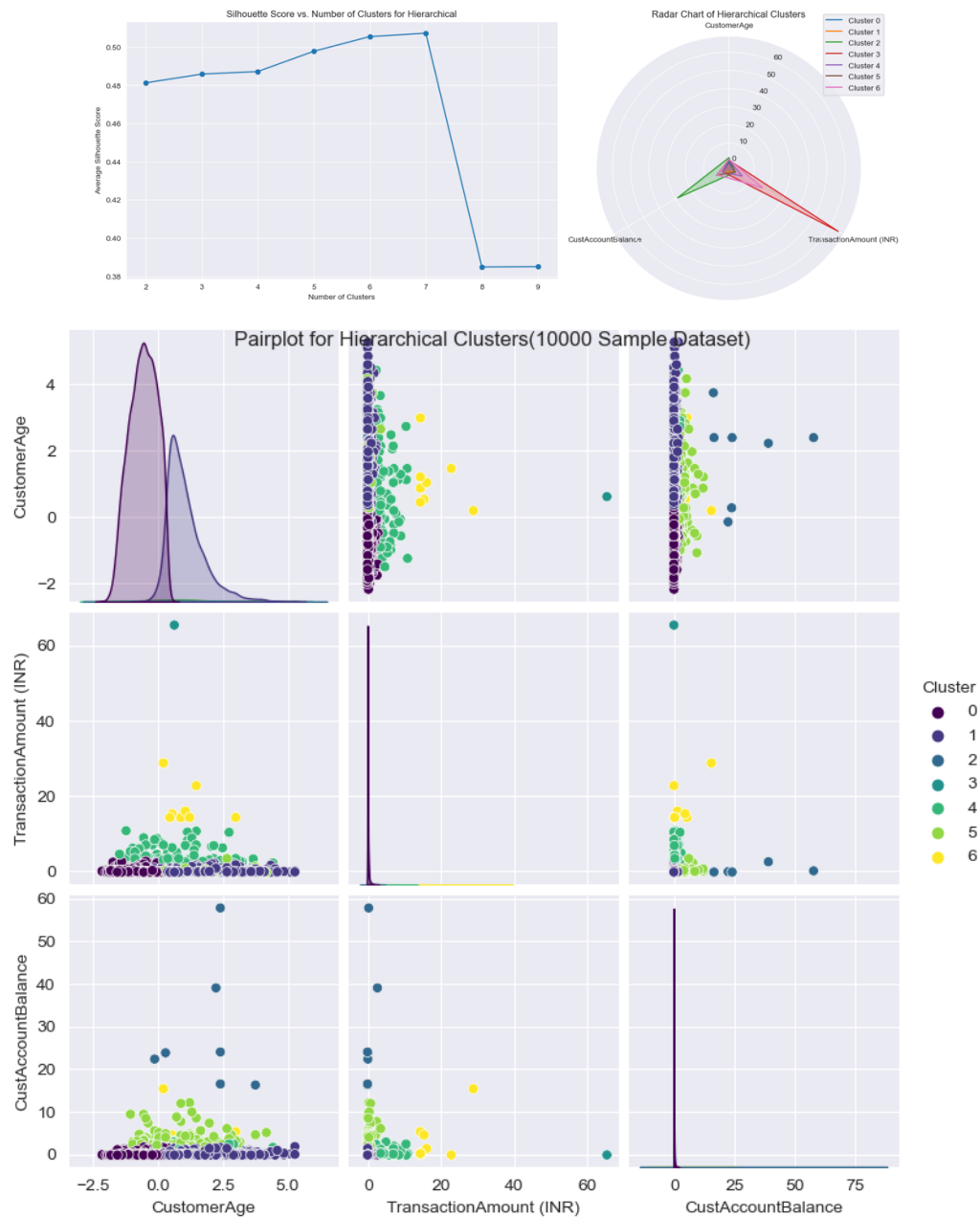
For Hierarchical and DBSCAN, we use the sample data to calculate the clusters, since the entire dataset is too large to calculate.

3.1 K-Means (Full Dataset)



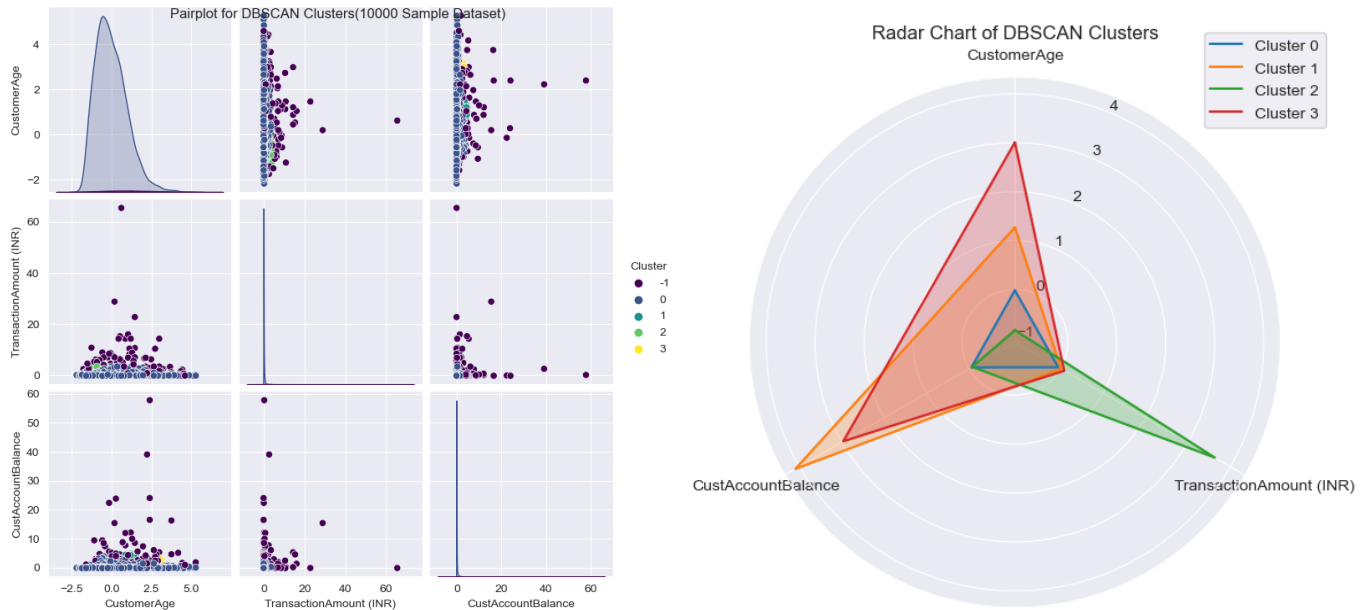
For K-Means, we use 5 as the number of clusters.

3.2 Hierarchical (10000 Sample Dataset)

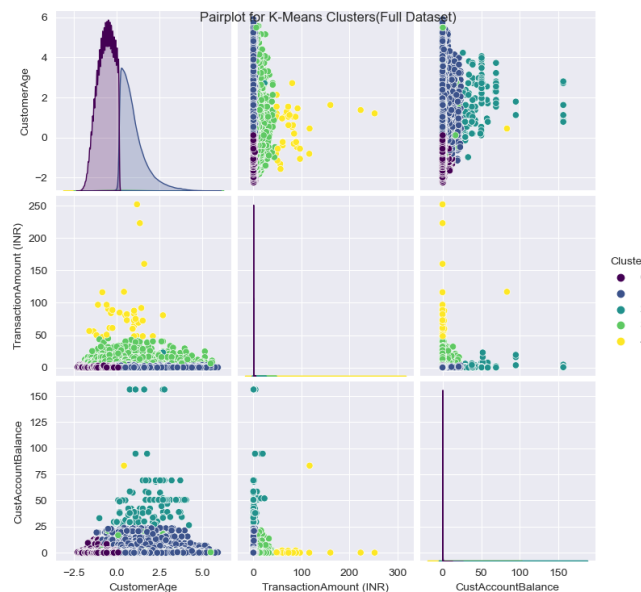


For Hierarchical, we use 7 as the number of clusters

3.3 DBSCAN (10000 Sample Dataset)



4 Understanding Customer Clusters: Common Characteristics and Differences



From the K-Means pairplot, we can observe the following:

For Cluster 0, it aggregates a group characterized by lower customer age, lower transaction amounts, and lower customer account balances.

For Cluster 1, it clusters a group with lower transaction amounts and moderate-to-low customer account balances.

For Cluster 2, it brings together a group with lower transaction amounts and higher customer account balances.

For Cluster 3, it groups a demographic with moderate transaction amounts and moderate-to-low customer account balances.

For Cluster 4, it collects a group marked by moderately high transaction amounts and lower customer account balances.

Reference

Connectif. (2023, July 3). *What are RFM scores and how to calculate them*. Connectif Marketing Automation.
<https://connectif.ai/en/blog/what-are-rfm-scores-and-how-to-calculate-them/>