Study on Digital Transformation of Small and Medium-Sized Rural Commercial Banks

Wenting Han*

Yingcheng No. 1 High School, Changchun, 130502, China

* Corresponding *author. Email: hwt050312@163.com*

Abstract. With the rapid development of the global technology economy and the increasingly personalized demand for capital and services, the old marketing model can no longer adapt to the changes of the times. How to successfully achieve "digital transformation", change development strategies, and business management models, improve customer experience, and enhance the core competitiveness of the industry has become an important problem for small and medium-sized rural commercial banks to solve. This study achieves a breakthrough in one part of the digital transformation process based on the real-life characteristics of small and medium-sized agricultural and commercial banks. According to the customer characteristics, the collection of government big data and customer behavior and personality preference data, and further refinement of customer classification will be carried out. On this basis, carry out digital transformation while maintaining local advantages, and explore the value of long-tail customers to realize the transformation of rural commercial banks.

Keywords: Digital Transformation, Commercial Bank, Customer Hierarchical Classification, Market Segmentation, Cluster Analysis, The Long Tail business mode

P. Dou and K. Zhang (eds.), *Proceedings of the 2023 International Conference on Economic Management, Financial Innovation and Public Service (EMFIPS 2023)*, Advances in Economics, Business and Management Research 287,
https://doi.org/10.2991/978-94-6463-441-9_37

1 Introduction

As digitalization and intelligent technology have penetrated into all aspects of the financial industry, the banking business structure at home and abroad is undergoing earth-shaking changes. At the same time, as large and medium-sized banks sink into counties and cities, small and medium-sized banks are generally under pressure to survive, and the old marketing model can no longer adapt to the changes of the times. In the context of increasingly personalized demand for capital and services, how to successfully realize "digital transformation" by changing development strategies, operation and management modes, and improving customer experience, in order to attract customers and enhance the core competitiveness of the industry has become an important issue for small and medium-sized commercial banks to address.

Based on its many advantages such as low cost, risk diversification and numerous channels, retail banking is becoming a major growth point for the banking industry. Although retail banking is traditionally an area of strength for commercial banks, competition in this business area is increasingly intensified. In order to be competitive in the retail business, small and medium-sized commercial banks need to transform digitally while maintaining their local advantages, especially to explore the value of long-tail customers and shift from focusing on high-end customers and large customers to focusing on long-tail customers as well. Taking Wells Fargo as an example, One of its retail core competencies is a personal finance, which provides personalized, differentiated, and confidential financial services to high-value individuals through product portfolio, business consulting, and investment advisory to achieve the bank's goal of improving operational efficiency and preserving and creating wealth for high-value customers. For example, it designs one-to-one customized financial solutions for each customer based on indicators such as "financial status", "credit rating" and "risk appetite".

Based on the practice of a local agricultural and commercial bank in China, this study starts from the more easily available customer deposit and transaction data, and pre-stratifies each indicator data, and based on this, uses clustering algorithms to stratify and classify customers. Combined with business experience, the stratification results can well show the characteristics of different categories of customers and provide a basis for the development of marketing strategies to achieve retail business for different customer segments.

This study achieves a breakthrough in one part of the digital transformation process based on the real-life characteristics of small and medium-sized agricultural and commercial banks. As the work progresses, more customer characteristics will be collected, including the collection of government big data and customer behavior and personality preference data, and further refinement of customer classification will be carried out.

2 Material

In this paper, 30% of the local population of a small and medium-sized bank in a county-level city in China with a moderate to low economic capacity is taken for the last three months. By filtering and removing customers with missing and erroneous data, 230,975 more valid data are screened from 377,110 customers. By collecting information from different categories of each customer, a table of all customer data sets was compiled:

Table 1. Customer Data (Excerpt)

Customer ID	age	DDA (￥)	Short Term FD (￥)	Long Term FD (￥)	Transaction Number (times)	Transaction (￥)
100369307#	49	1356.3	0	78002.5	1158.66	6
100597743#	69	11903.8	25000	95000	4769.0	17000
100320810#	33	5879.8	0	0	53662.8	49312.1
100344576#	58	147.7	234514.2	50000	878.3	3600
E.t.c						

A brief statistical analysis of the distribution of the 230,975 sample data reveals the following significant features in the data:

(1) In demand deposits, the data are roughly normally distributed, with the peak of the data occurring between 1000 and 1260 yuan with 13,298 households.

(2) Most of the customers are middle-aged people between the ages of 50-54.

(3) In the transaction amount, the overall total debit transaction amount is average with the total credit transaction amount. The loan transaction amount of 88,385 households is less than 10 yuan, the number of people accounted for about 38.27%, the cumulative amount accounted for less than 0.01%; 136,375 households have debit transaction amount less than 10 yuan, the number of people accounted for about 59%, the cumulative amount accounted for less than 0.01%; the highest transaction amount of 35 households, the number of people accounted for was 0.01%, but the cumulative transaction amount accounted for about 8%.

3 Method

Based on the market user data extracted from a small and medium-sized agricultural and commercial bank in China, this paper analyzes customer characteristics and classifies them in a refined way using Market Segmentation and Cluster Analysis based on k-means algorithm.

3.1 Market Segmentation

Market segmentation is the process of dividing a broad consumer or business market, normally consisting of existing and potential customers, into sub-groups of consumers (known as segments) based on shared characteristics. In dividing or segmenting markets, researchers typically look for common characteristics such as shared needs, common interests, similar lifestyles, or even similar demographic profiles.[1] This allows companies to target different categories of consumers and tailor their marketing approach by the characteristics of different consumer groups, thereby increasing sales and customer loyalty.

3.1.1Types of Market Segmentation.

Market segmentation is based on four main variables, namely geographic, demographic, psychographic, and behavioral segmentation. For the customers of small and medium-sized agribanks, segmenting the group by demographic and behavioral variables to construct individual customer portraits is the most effective and reliable method. Demographic segmentation is one of the simple, common methods of market segmentation. It involves breaking the market into customer demographics as age, income, gender, race, education, or occupation. This market

segmentation strategy assumes that individuals with similar demographics will have similar needs. Behavioral segmentation, on the other hand, relies heavily on market data, consumer actions, and decision-making patterns of customers. This approach groups consumers based on how they have previously interacted with markets and products. This approach assumes that consumers prior spending habits are an indicator of what they may buy in the future.[2] In order to make the data classification more valid and reliable, the customer information that can be derived from banking statistics is classified in the following table:

Table 2. Customer Segmentation Variables

	Variable
Demographic segmentation	Age
	Demand Deposit (per day)
	Short-Term Fixed Deposit
Behavioral segmentation	Long-Term Fixed Deposit
	Transaction Activity
	Credit Transaction
	Debit Transaction

3.1.2 Determine Market Segment

First of all, according to the Pareto Principle, in order to cater to the different needs of customers of different amounts, 20% of high-dollar users and the remaining 80% of general users are divided, and different marketing methods and countermeasures are adopted.[3]

In addition, by organizing the data corresponding to different information of customers, we can obtain visual charts such as cumulative distribution charts and

histograms. Among them, the cumulative distribution chart is similar to the 80-20 distribution of the Pareto Distribution trend. According to the gentle steepness of the data trend of the cumulative distribution chart, the position with a relatively large slope is used as the division point in combination with the business scenario, and the customers with different characteristics can be most effectively grouped in one group. Take the data of Demand Deposit Accounts (DDA) in the following figure as an example:

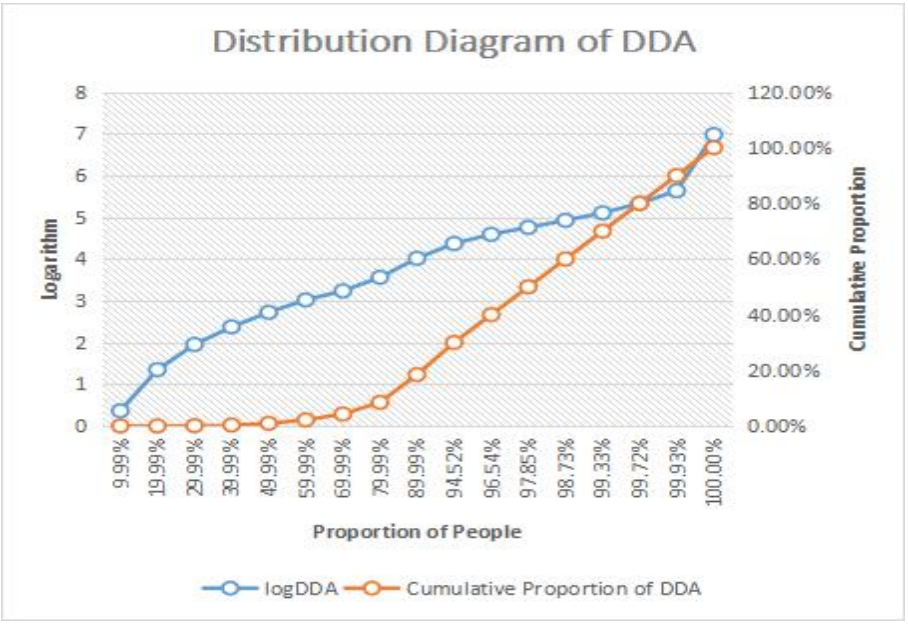

Fig. 1. Distribution Diagram of Demand Deposit Accounts

According to the gentle and steep degree of the trend of the cumulative distribution map data, the position with a large slope is used as the division point, which can most effectively group customers with different characteristics into one group.

According to the above two points and the actual sales experience, we can roughly divide the customers into eight levels, rank 1 is the most inactive customer, rank 8 is the most active customer, so as to increase. The customer 's age characteristics are generalized according to age structure.

Table 3. Customer Classification

Rank	DDA (￥)	Short Term FD (￥)	Long Term FD (￥)	Transaction Number (times)	Transaction (￥)
1	[1,500)	0	0	0	[1,10)
2	[500,1000)	[1,10000)	[1,20000)	[1,3)	[10,400)
3	[1000,1700)	[10000,25000)	[20000,50000)	[3,6)	[400,1250)
4	[1700,3600)	[25000,60000)	[50000,100000)	[6,10)	[1250,5400)
5	[3600,10000)	[60000,100000)	[100000,200000)	[10,14)	[5400,30000)
6	[10000,23000)	[100000,200000)	[200000,320000)	[14,30)	[30000,100000)
7	[23000,40000)	[200000,300000)	[320000,550000)	[30~103)	[100000，240000)
8	≥40000	≥300000	≥550000	≥103	≥240000

3.2 Cluster Analysis

Cluster analysis is a useful tool for processing bank customer data to develop a customer profile. By aggregating similar characteristics, customers are grouped into different clusters, where the objects within the clusters are similar to each other, while the objects in different clusters are different. The data statistics within the clusters are analyzed and the corresponding customer profiles are developed as a way to support banking and improve customer satisfaction. There are many clustering methods to choose from. In this paper, K-means algorithm is chosen as a tool to process the data.

K-means clustering algorithm can divide customer data into multiple clusters, and each cluster represents a group of customers with similar characteristics. Based on the clustering results, banks can develop corresponding customer profiles to understand their customers' needs, preferences, etc., in order to better provide personalized services and products, thus improving customer satisfaction and business revenue.

The following are the steps to be implemented in the data processing of K-means clustering algorithm.

3.2.1Model Preprocessing.

Cluster analysis is mainly based on the classification of distances between classified objects, which is easily influenced by the measurement units of clustering variables. The larger the order of magnitude, the greater the influence on the distance calculation results, and the more dominant they will be in the clustering process, thus overshadowing other variables of a smaller order of magnitude and leading to biased clustering results. Therefore, before the clustering analysis, the data needs to be processed so that the data can be compared under the same criteria. In order for the algorithm to obtain higher quality results, a large range of data in each level needs to be given a specific representative value data. For this purpose, we took the median value of the respective range for each of the eight levels as the representative value for that level.

Table 4. representative value of customer classification

Rank	DDA (￥)	Short Term FD (￥)	Long Term FD (￥)	Transaction Number (times)	Transaction (￥)
1	50	0	0	0	5
2	550	5000	10000	2	200
3	1350	17500	35000	4	800
4	2650	42500	75000	8	3300
5	6800	80000	150000	12	17700
6	16500	150000	260000	22	65000
7	31500	250000	435000	76	1700000
8	45000	350000	600000	110	3000000

3.2.2Mini-Batch.

Although K-Means performs very well in terms of algorithm stability, efficiency and accuracy, and still does so when dealing with large amounts of data, each iteration requires traversing the full amount of data. Once the amount of data is too large, the number of iterations is too large due to the computational complexity, which can lead to very slow convergence.

To address this problem of K-means, the Mini Batch algorithm allows a portion of samples from different categories to be taken as representatives in the clustering algorithm process, thereby artificially reducing the size of the sample. Compared to the K-means algorithm, the Mini-batch K-means algorithm has some advantages in terms of speed and efficiency because it uses only a portion of the data set rather than the entire data when dealing with large data sets. The iterative running time is correspondingly reduced due to the small number of computational samples.

3.2.3Inertia.

K-Means always pursue "small intra-cluster variation and large inter-cluster variation" in the process of category classification and final results, where the variation is measured by the distance from the sample point to the center of mass μ of the cluster it is in. According to the Euclidean distance can be obtained[4]:

$$\mathrm{d}(\mathrm{x}, \mu) = \sqrt{\sum_{i=1}^{n} (x_i - \mu_i)^2} \tag{1}$$

The sum of the squares of the distances of all sample points in a cluster to the center of mass is：

$$\sum_{j=0}^{m} \sum_{i=1}^{n} (x_i - \mu_i)^2 \tag{2}$$

Where m is the number of samples in a cluster and j is the number of each sample. This formula is called Sum of Square, also known as Inertia, and when the sum of the intra-cluster squares of all clusters in a dataset is added, the Total Cluster Sum of Square, also known as Total Inertia, is obtained. The smaller the total Inertia, the more similar the samples in each cluster, the better the clustering effect.[5]

A good model is one with low inertia AND a low number of clusters (K). However, this is a tradeoff because as K increases, inertia decreases. In order to find the optimal value for the dataset, the point at which the inertia starts to slow down is

the optimal value k according to the Elbow method.[6] As shown in the figure below, the optimal value k = 9.

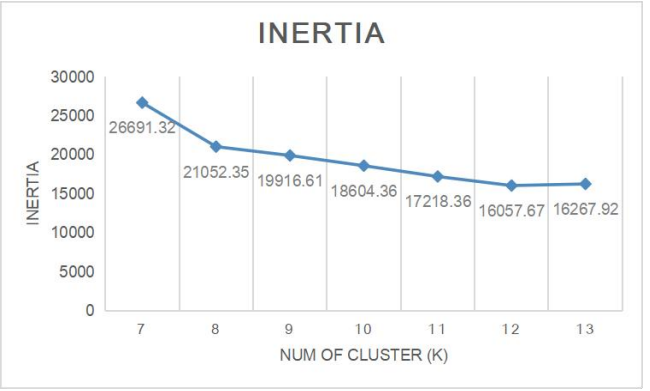

Fig. 2. Optimal Number of Clusters

Therefore, clustering the dataset into nine clusters overall works best.

4 Result

From the above market segmentation and clustering analysis algorithm for customer stratification, combined with business experience, the stratification results can well show the characteristics of different categories of customers and analyze customer portraits. In general, we can divide the bank customer groups into nine categories.

Based on the transaction activity and transaction amount size of different cluster categories, the following analysis is made:

A. The fourth and fifth are clusters with high transaction activity and transaction amounts, of which the fifth category is the most active, and the two categories together account for only 12.57% of the population in total, but the transaction amounts have totaled about 75% of the total amount;

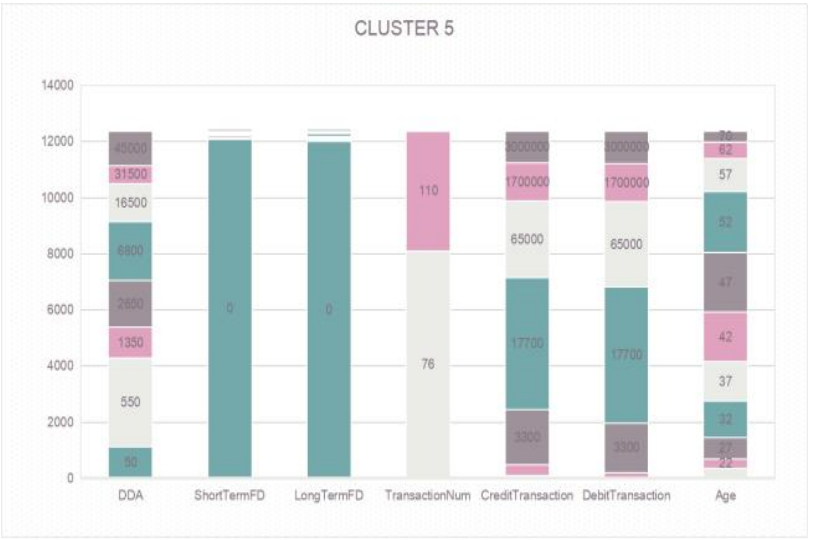

Fig. 3. Stacked Diagram of the Personnel Proportion in Clusters of type a

B. The second, third and ninth categories are groups of customers with moderately active transactions, of which the ninth category has the largest amount of transactions, and none of the three groups have time deposits;

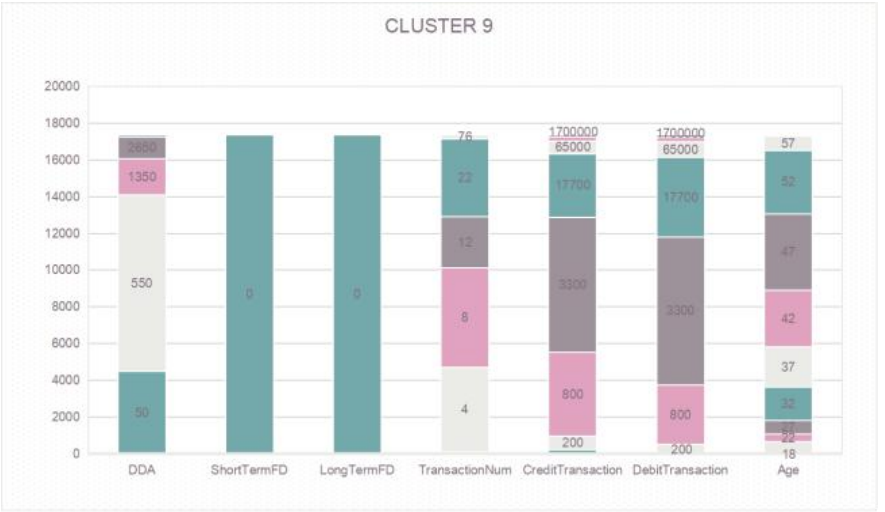

Fig. 4. Stacked Diagram of the Personnel Proportion in Clusters of type b

C. The first and seventh categories have the lowest transaction activity and transaction amounts, concentrated in the younger and middle-aged and older categories;

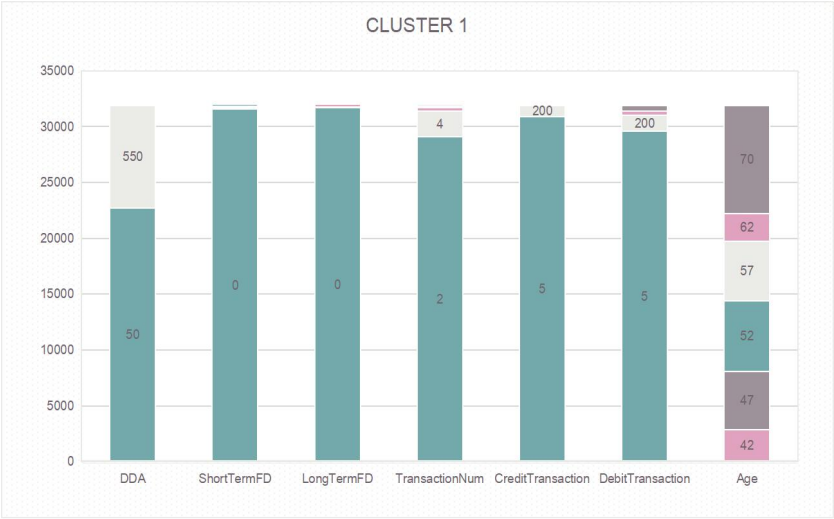

Fig. 5. Stacked Diagram of the Personnel Proportion in Clusters of type c

D. The sixth and eighth categories have average transaction activity and transaction amounts, but the most short-term and long-term time deposits.

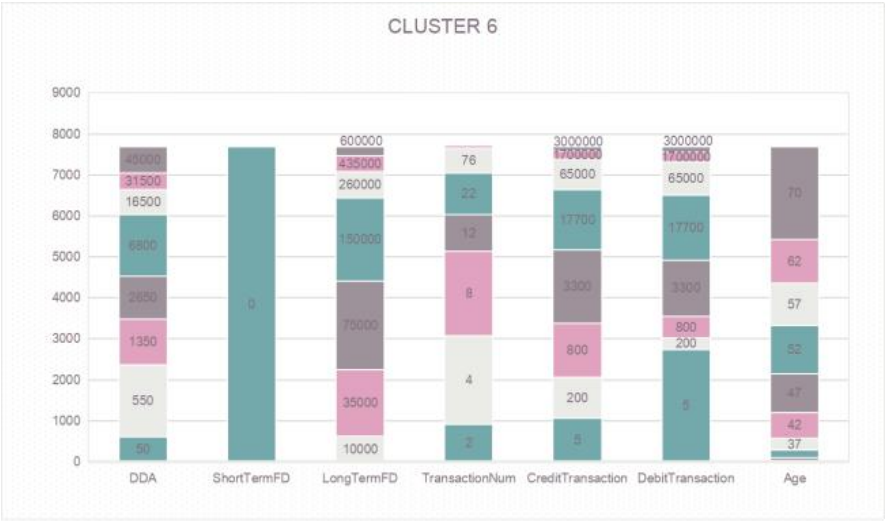

Fig. 6. Stacked Diagram of the Personnel Proportion in Clusters of type d

5 Discussion

The above figures are roughly in line with the Pareto Principle, which states that 20% of customers generate 80% of profits. Nevertheless, Chris Anderson suggests that when the storage and distribution channels for a product are large enough, a broad sales dimension gives 98% of the products a chance to sell, and no longer relies on only 20% of the main products. The market share of products with poor demand or sales can match or even surpass those of a few popular products.[7] As a result, providing standardized, generic service to the average customer, the long tail, can also be extremely rewarding. Long-tail customers have a very high probability of becoming strategic and key customers after a few years of growth. Through market segmentation and cluster analysis, we can summarize the characteristics and customer profiles of long-tail customers, so that commercial banks can better explore the personal value of long-tail customers, maintain, serve and match their needs, and enhance customer stickiness, so that they can become loyal customers and bring high returns and profits to banks.

By analyzing and organizing the data from the clustering algorithm, we can summarize the information about the characteristics of customers in different clusters

and organize them into three major categories, namely high value customers, potential customer and low value customers.

In this data, long-tail customers account for about 27.56% of the overall customers, which are mainly divided into two categories.

(1) Such customers do not have personal time deposits and have a small amount of transactions, but have some demand deposits and active transactions that can have the opportunity to be developed, accounting for 20.90 % of the total.

Table 5. Clustering Customer Analysis Table

Cluster 3				
Description	The demand deposit amount is small; no fixed deposits; the transaction activity is relatively high; the transaction amounts are moderate; elderly people			
Personnel Proportion	13.39% ($\frac{30936}{230975}$)			
	Total Amount	Median	Mean	Proportion
DDA (￥)	28057050	550	906.94	3.97%
Short Term FD (￥)	0	0	0.00	0.00%
Long Term FD (￥)	0	0	0.00	0.00%
Transaction Number	265618	8	9	11.36%
Credit Transaction (￥)	85284955	800	2756.82	0.76%
Debit Transaction (￥)	148459700	3300	4798.93	1.25%
Age		62	64	

Table 6. Clustering Customer Analysis Table

Cluster 9				
Description	The demand deposit amount is small; no fixed deposit; the transaction activity is moderate; the transaction amount is moderate; middle-aged people			
Personnel Proportion	7.51% ($\frac{17353}{230975}$)			
	Total Amount	Median	Mean	Proportion
DDA (￥)	11996550	550	691.32	1.70%
Short Term FD (￥)	50000	0	2.88	0.01%
Long Term FD (￥)	10000	0	0.58	0.00%
Transaction Number	203988	8	12	8.73%
Credit Transaction (￥)	829671810	3300	47811.43	7.44%
Debit Transaction (￥)	926486110	3300	53390.54	7.82%
Age		42	45	

(2) This category of customers has average transaction activity and transaction amounts, but short-term and long-term deposits are the most numerous, accounting for a total of 6.66% of the total.

Table 7. Clustering Customer Analysis Table

Cluster 6	
Description	There is no short-term fixed deposit, but the long-term fixed deposit amount is the most
Personnel Proportion	3.33% ($\frac{7692}{230975}$)

	Total Amount	Median	Mean	Proportion
DDA (￥)	67355300	2650	8756.54	9.52%
Short Term FD (￥)	30000	0	3.90	0.00%
Long Term FD (￥)	1000320000	75000	130046.80	67.64%
Transaction Number	110642	8	14	4.73%
Credit Transaction (￥)	755627605	3300	98235.52	6.78%
Debit Transaction (￥)	923762900	3300	120093.98	7.80%
Age		57	57	

Table 8. Clustering Customer Analysis Table

Cluster 8				
Description	the short-term fixed deposits are the most and the long-term fixed deposits are large; the transaction activity is moderate.			
Personnel Proportion	3.33% ($\frac{7704}{230975}$)			
	Total Amount	Median	Mean	Proportion
DDA (￥)	72401900	2650	9397.96	10.23%
Short Term FD (￥)	565012500	42500	73340.15	93.91%
Long Term FD (￥)	412595000	0	53555.94	27.90%
Transaction Number	133656	8	17	5.72%
Credit Transaction (￥)	981932420	3300	127457.48	8.81%

Debit Transaction (￥)	1156463265	3300	150112.05	9.77%
Age		57	58	

6 Conclusion

Banks must target their customer base in the right way if they want to get the best return on their investment. If these customers are neglected, it will take a toll on the bank's overall bottom line and lose customer trust. Therefore, market segmentation and cluster analysis, which divides customers into groups with similar characteristics (demand, capital, spending habits, etc.) and then carry out a personalized marketing strategy, contribute a lot to the target and personalized business. Studying the specific characteristics of the different clusters and customizing the services corresponding to them also helps to attract more people with similar behaviors.

References

1. Wikipedia Contributors. Market segmentation. Wikipedia. Published June 26, 2023. Accessed June 27, 2023. https://en.wikipedia.org/wiki/Market_segmentation
2. Market Segmentation: Definition, Example, Types, Benefits. Investopedia. Published 2023. Accessed June 27, 2023. https://www.investopedia.com/terms/m/marketsegmentation.asp
3. Arnold, Barry C. "Pareto distribution." Wiley StatsRef: Statistics Reference Online (2014): 1-3.
4. Singh, A., Yadav, A., & Rana, A. (2013). K-means with Three different Distance Metrics. International Journal of Computer Applications, 67(10).
5. Clustering algorithm KMeans. Programmer.group. Published 2021. Accessed June 28, 2023. https://programmer.group/clustering-algorithm-kmeans.html
6. to I. Intro to Machine Learning: Clustering: K-Means Cheatsheet | Codecademy. Codecademy. Published 2023. Accessed June 27, 2023. https://www.codecademy.com/learn/machine-learning/modules/dspath-clustering/cheatsheet
7. Anderson, Chris. The long tail: Why the future of business is selling less of more. Hachette UK, 2006.