

The main purpose behind this study was to analyze the problems faced by big retail banks during business expansion. Typically banks tend to acquire new customers at huge costs rather than leveraging their existing customer base. A bank's customers leave behind a large footprint in terms of the transactions they perform, which can be analyzed to understand their behavior pattern which may be leveraged for selling new products.

This paper analyzes the customers' transaction patterns, product holdings, demographics, past trends, and other attributes to devise an effective strategy for engaging them further. In a retail bank with various product offerings, the focus should be on customer segmentation and profiling to ensure ease of targeting, marketing, and offering personalized products to retain profitable customers and capturing market share across geographies.

Keywords: *customer segmentation, profiling, clustering, business expansion, profitable customers, scorecard, bank, customer, transactions*

Table of Contents

1. Introduction	
2. Literature Review	
3. Data	
4. Profiling To Define Profitable Customers	
◦ Technique Used: Scorecard	
◦ Scorecard Analysis Summary	
◦ Recommendations	
5. Profiling Profitable Customers Into Various Segments To Customize Product Offerings	
◦ Technique Used: K-Means Clustering Algorithm	
◦ Clustering Output and Summary	
◦ Recommendations	
6. Application And Conclusion	

INTRODUCTION

Retail banks deal with various problems during business expansion. Over the years, banks have been trying to expand their customer base without taking into full consideration the value each customer brings now and is able to bring in the future. Customers leave behind a large footprint in terms of the transactions they perform, which can be analyzed to determine who the most valuable customers are and how to nurture and grow the business by leveraging the existing customer base.

The main idea behind taking up the study of Czech Bank is to understand their existing customers – their transaction patterns, product holdings, demographics, past trend, and other attributes and behavior with the bank to devise an effective strategy. Czech Republic Bank is a banking group that offers major retail banking services. The services include managing savings and current accounts, offering loans and credit card services.

The bank is functioning since 1994 and has a large number of customers. The bank has determined that if it can devise any strategy to tap potential around its existing customer base, then it can scale up business quickly in a more cost-effective manner as there is no acquisition cost. These customers are already banking with them; they need attention, service, and hand-holding.

The bank wants to target its services to the selected groups of customer segments, created by differentiating between valuable and non-value-add customers. Currently, the bank works on gut feelings, without a strategy based on analytics, regarding which customer to target (whom to offer an additional service) and who is a potential risk.

To help formulate an effective strategy for business expansion, the following two objectives were taken up.

- Profiling to classify each customer into either profitable or non-profitable buckets
- Profiling profitable customers into various segments to customize product offerings to increase the overall business of the bank

LITERATURE REVIEW

The dataset of the Czech Bank was available in the public domain since 1999. Most of the analytical study in the financial analytics domain has been done around default prediction, fraud risk, preventive forecast, credit card analysis. We have extended this study in customer profiling and segmentation part using the analytical approach – clustering technique and scorecard. RFM (Recency Frequency Measure) being the most frequently used technique in the retail banking domain for customer segmentation.

Customer Profiling and Segmentation play a pivotal role in deriving customer service strategies which in turn enhances customer satisfaction levels as well as to gain market positions. The inability to discover valuable information hidden in the data prevents the organizations from transforming the data into knowledge. Effective customer relationships require an understanding of what the relationship entails and the ability to provide personalized services, a means for building mutual value and respect, and a commitment to the relationship itself. By identifying the associations between products purchased in point of sale transactions, retailers can develop focused promotion strategies.

The clustering technique used for data mining is the key to bringing business intelligence to more varying disciplines and intricate tasks in retail that enables precise insights and patterns by providing an in-depth understanding of the behavioral and demographic patterns and also to identify main characteristics of the customers in each segment to retain the existing profitable customers. Effective communication is very difficult to establish in a retail bank with various product offerings, so it is necessary to divide customers into groups, whose members have similar characteristics to ensure ease of targeting, marketing, and offering personalized products to retain the customers. The paper has segregated the customers into different clusters based on demographic data, product holdings, and transactional behavior patterns as well as classified each customer into either profitable or non-profitable. This has been further used to guide the bank to formulate its business strategy and product mix offerings.

Benefits of customer profiling and segmentation:

- More customer retention
- Enhances competitiveness
- Establishes brand identity
- Better customer relationship
- Leads to price optimization
- Best economies to sale
- Improves channel of distribution
- Increase profit by keeping costs down
- Identify potential customers
- Improves Customer Engagement and Brand Loyalty

DATA

The data for the project has been sourced from the internet; a real anonymized banking transactional dataset of Czech Bank from 1st Jan1993 to 31st Dec 1998. It's based on the 5 years' data – approximately data volume is about 1 million transaction records comprising of 4,500 unique customers. Please refer to the below link to access data: http://sorry.vse.cz/~berka/challenge/pkdd1999/data_berka.zip

There were also some interesting results of classification according to loans (running loans with no problems, running loans with the client in debt, finished contract with the loan paid off, finished contract with the loan not paid) and according to credit cards (does not own credit card, owns the junior card, owns the classic card, owns gold card). Only 15% of clients had loan contracts, out of them only 11% of loans were with problems (running or finished). Similar proportions hold for the credit cards as only 20% of clients used credit cards, out of them only 10% (2% of all clients) used gold cards.

PROFILING TO DEFINE PROFITABLE CUSTOMERS

Technique Used: Scorecard

We formulated a customer-based Scorecard by providing numeric scores by using a box plot and weightages to each customer based on their product holdings and transactional history to analyze customer persona for the bank, so that specific strategies may be chalked out.

To determine customer profile through a Scorecard, we have covered the below data points:

- loan account with bank
- loan duration
- loan amount given to customer
- whether customer is having credit card
- whether customer is having multiple accounts
- credit quality of customer
- whether customer created any standing order with bank
- relationship tenure with bank
- average monthly balance in customer account

The following rule would help to understand the scoring methodology for continuous and categorical variables.

Continuous variables: Box plot statistics have been used to score them between 1 to 4. If the customer falls into the 1st quartile, the customer has scored 1, for quartile 2 score would be 2 and for quartile 3 and 4 scores would be 3 and 4 respectively.

Categorical variables: The answer is either Yes or No. If the customer is having a credit card, that customer would be given 1 else 0. After providing a score to each customer against each variable, we have then assigned the weightages to each variable between -3 to +3. Following are details on the rationale for the weightages attained, against each criterion:

Assigning Weights:

(having_loan weightage +2)

The bank considers any customer holding Loan Accounts to be more profitable as the probability to earn a higher income is more through loan interest charged and processing fees.

(loan_duration weightage +1)

Loan Duration is one of the considerations while weighing the profitability as long-term loans ensure that banks earn more interest and ensure longer customer stickiness.

(loan_amount weightage +3)

Loan Amount is an important aspect as banks will generally take bigger loan exposure only on those clients who are creditworthy, have good cash flow, and maintain higher AMB. Hence weightage of +3 is

taken, as this will be critical to determine profitability.

(having_creditcard weightage +3)

Credit Card is an unsecured lending product, as this is extended to only that customer set who has established DEMOG, KYC, income flow, or satisfactory associations with the bank. The earnings by the bank in case of delayed payment and minimum amount due pay is high. Hence bank categorizes them as profitable clients and the weightage of +3 suits this category.

(havingmultipleaccount weightage +1)

Multiple accounts reflect multiple associations of the client with the bank, thus establishing higher product holding. This aspect will determine the profitability, hence weightage of +1 is taken.

(credit_quality weightage -3)

Credit Quality is one of the primary tasks for determining the lending eligibility of the client. The customers score is 0 in case of no defaults and 1 in case of default. We have kept the weightage as -3 to penalize the defaulters in the scorecard.

(standing_order weightage +2)

The Standing Order or SI/ECS in a bank account reflects that the customer holds a loan or card product or any registered Bill pay. This reflects that the trust level of the customer on the bank is high, as such accounts have higher balances than normal accounts. This leads to profitability and income for the bank.

(tenure_score weightage +3)

The tenure score or the conduct of the client during the entire loan period is how his/her EMI payment history, any miss outs, defaults, AMB is. This is important for the bank from the income and risk mitigation point of view. This predicts largely how exposure taken by the bank will turn out into a loss-making or profitable one. The bank has given this high weightage of +3 to determine the profitability.

(average_monthlybalance weightage +3)

The Average Monthly Balance (AMB) is a clear reflection of the strength of the relationship between the bank and the client. The transacting banks have low AMB whereas the primary banks enjoy higher AMB. The banks become primary by higher product holding ratio, loan offerings, and service levels. The banks earn higher Net Interest Income (NII) on the money kept in accounts, hence we have given it a weightage of +3.

-

SCORECARD RESULTS

As per the total score bracket (7 to 36), we have divided them into 4 quartiles. Q4 has scored the highest and Q1 is having the least total score bracket. Q2 and Q3 remain in the middle portion. If a customer's is falling in Q4 bracket, then the customer would be a profitable one. This would help the bank in the segregation of its base in various segments as per the profitability generated by client relationships.

After analyzing the final score, we separated the customer base into 3 brackets – the most profitable, least profitable, and profitable customers. The assigned weightages/scores become the determinant of the customer; that is whether a customer is profitable for the bank or the least profitable one. It is also helpful in deciding a base of customer who is neither profitable nor non-profitable.

Hence we have drawn an inference, suggesting that the bank has a strong customer base where it can focus and create some strategies for these profitable customers to convert them to the most profitability bracket and at the same time retain the most profitable customers to generate more revenue. The least profitable customers do not add significant value to the growth of the bank's business and would be better if these can be let go off.

RECOMMENDATIONS

From the scorecard analysis results, out of the base of 4500 customers, we could identify 3184 customers (falling in Q2, Q3, and Q4) are the potential and profitable customers who add value to the bank's profit.

- The bank should target these customers with customized offerings to further increase its revenue.
- The bank should further weed out the base of non-profitable 1316 customers (falling in Q1) to reduce the cost that incurs in their retention.
- The bank should further segment the profitable customers to move them up to the higher profitability bands, for example from Q2 to Q3 and from Q3 to Q4 by suitably nurturing them.

With this scorecard result, to increase the revenue from these profitable customers, we will further proceed to define customer segments to customize offerings using the clustering technique as executed below.

PROFILING PROFITABLE CUSTOMERS INTO VARIOUS SEGMENTS TO CUSTOMIZE PRODUCT OFFERINGS

Technique Used: K-Means Clustering Algorithm

The purpose is to segregate the Profitable bank customer base into different customer segments, thus ensuring ease of targeting and communication so that the bank can offer the bundle of products or services to the different band of customers that is most likely to buy from the bank. For the customer segmentation and to study the behavioral data based on customer's transactions and their demographics, we have done feature selection for the available data.

We segregated the customer base into 3 different segments on the basis of their product holdings, traits, and transactional patterns. Our approach to arrive at a solution with 3 clusters was mainly focused on identifying different customer segments with common traits and holdings, demographics, and transaction behavior, so that the bank may understand the trend and customize offerings accordingly. In addition, if we are able to identify products that have similar customer attributes, then we can highlight cross-sell or up-sell opportunities towards targeted customers.

For modeling this objective, we have used K-Means clustering. In this, we have found out the best value of parameter K, i.e. K=3 with the help of the Elbow method. After clustering, we found that the below mentioned 10 variables play the most significant role in clustering:

- household_payment
- having_loan
- oldagepension_payment
- average_salary
- age
- having_creditcard
- numberof_credit_transactions
- numberof_debit_transactions
- average_monthly_balance
- insurance_payment

CLUSTERING RESULTS

Out of the base of 3184 Profitable customers, Cluster 1 is having the highest population of customer concentration with a total number of 2199 customers and on the other side Cluster 2 is having the least number of customers with a total of 462 customers. Cluster 3 is having 523 customers.

The Average Age of customers in Cluster 1 is relatively higher than the Age of customers in other clusters as well as the oldest one. The bank will consider them suitably eligible for insurance products as increasing age will lead to increased medical expenses, also the cluster needs to be targeted for investment plans with moderate risk exposure and pension coverage. Cluster 3 is having the youngest customer age group as compared to other clusters. Cluster 2 is characterized by the account holders of the middle as well as the old age group.

This cluster analysis revealed specific characteristics and insights that will help the bank to understand their customers need and requirement so that the bank can create customized offers and custom plans to attract potential and profitable customers and also cross-sell or up-sell their products and services to those holding few products that will lead to higher product penetration, higher stickiness, and lower base erosion. This will empower banks to nurture by targeting specific segments with suitable products and services, thus providing a more personalized approach that might lead the bank with appropriate marketing propositions, growth, and profitability.

We have profiled these clusters descriptively:

RECOMMENDATIONS

Cluster 1: This is the largest cluster for the bank hence needs the most appropriate targeting and product offering. The bank may offer traditional banking products like fixed deposits, term insurance, medical insurance, general insurance, and debt investment plans as this group does not seem to be willing to take any higher risk equity products.

The bank must consider them suitably eligible for insurance products as increasing age will lead to increased medical and health-related expenses. The bank has a wonderful opportunity to propose them investment plans with moderate risk exposure, pension plans coverage, and also approach them to create an investment corpus to ensure a stress-free retirement period.

Cluster 2: The bank should offer them some promo-based cashback offers and discounts to further increase the usage of credit cards for household payments. The bank may target this cluster for credit card up-gradation like Silver to Gold, or Platinum variants with higher credit limits. This group also can be offered premium concierge services on a chargeable basis. The salary account holders may be upgraded to Wealth and Private banking platforms so that customers feel more important.

The bank should target them for household durables, easy EMI products, and purchases related to children. As the profile suggests, these are least likely to default and must be targeted for loan cross-sell as this will help to increase profitability and stickiness. Old-age pension plans must be adequately sold to this cluster.

Cluster 3: The bank must target this cluster for credit card upgrade schemes along with lifestyle-based offers on cards. As the profile suggests, this cluster has not started the retirement planning yet so the bank must sensitize this cluster to start pension planning immediately and must offer related product solutions as well. The bank should offer insurance plans related to life, health, and general categories. The bank may also offer attractive interests on loans or unsecured loans with a higher rate of interest and processing fees. This cluster also has a very high average monthly balance that means funds remain idle in a savings account at a low-interest rate. The bank must offer better interest-yielding products like fixed deposits, mutual systematic investment plans, and overdraft products to this cluster.

APPLICATION AND CONCLUSION

The outcome of this study is based on a data-driven analytical approach that will empower the bank to devise an effective marketing strategy to increase its profitability by targeting potential customers from its existing customer base, thus ensuring optimization of resources. This will enable the bank to target and sell to those customers competitively and economically that are most likely to buy their products or services as the bank now understands its customer's requirements very well. With the help of scorecard and clustering output, the bank can identify the most profitable and potential customers along with their characteristics and devise various strategies to move them up to the higher profitability bands.

From the scorecard analysis results, out of the base of 4500 customers, we identified 3184 customers (falling in Q2, Q3, and Q4) are the potential and profitable customers that add value to the bank's profit. The bank should target this segment further to increase its revenue. In addition out of 3184 customers, 1007 customers (falling in Q4) are the most profitable ones for the bank. The remaining customers (falling in Q1) do not add value to the growth of the bank's business and would be better if these can be let go off and this will also reduce the cost that incurs in their retention.

To further define the strategies to increase the revenue from these profitable customers, we profiled them into 3 separate clusters. Each cluster uniquely provides insights into their needs and requirements. The bank could use these insights to create customized offers and custom plans to cross-sell or up-sell more products and services to those currently having low product holdings, and this will lead the bank towards higher product penetration, profitability, and capturing appropriate market share. This will also increase customer loyalty towards the bank. More loyal customers will help in improving revenues and profits for the bank.

End Notes:

These analytical models or techniques (scorecard and clustering) used in this study are generic in nature and not specific to the case in point. These models can find a wide application across the financial services industry.