Financial Programming

Client Evaluation for Banking Client

Theau Bruno | Shruti Bajpai

Table of Contents

OBJECTIVE

The banking client currently offers account management and loan opportunities (among several other service offerings) to their customers. As the banking client expands services, a need arises to understand which clients to target. Data has been collected based on the banking client for customers from 1999. Customers will be segregated as “good” vs. “bad” to evaluate for whom to offer additional services or to whom minimize the bank loses. The following PKDD’99 dataset been utilized to provide actionable insight that will enhance serviceability.

INTRODUCTION | Data Marts Data Cleansing

The world renown data leading organization, Oracle, defines a data-warehouse role as: “centraliz[ing] and consolidate[ing] large amounts of data from multiple sources.” In spanning across 8 different datasets, the task to create a singular datasource was to utilize the join function that would combine to have over 5700 observations across 62 rows. In order to ensure that only relevant data was included for segmenting good vs. bad customers, tables were simplified by selecting only relevant columns to be used during analyses. In parallel to joining and selecting specific columns from each datasets, data cleansing was also conducted. Given PKDD’99 data was provided partially in English and partially in Czech, the data was translated utilizing the .map function. Upon completion of these steps, all dates were corrected to consist of the correct format (example: date), thus allowing ease of analysis. Once the tables were cleaned up, pivot table were created by grouping the account_id information to reduce the number of rows. Instead of having one row per transaction type, one column was used per transaction type where the sum was accumulated from each account id.

To simplify our interpretation, the datasets were manipulated by adding and simplifying certain columns. For example, a seniority column has been created to calculate how many days ago the client's account was created. Other columns were created using ‘if loops’ to determine in which age range the customer is in and/or which loan size range the customer is in.

Most of the charts and analyses have been based on the column: 'loan_client_type'. This is a simplification of the result of the loan: if the customer has repaid the loan then he/she is a good customer, if not then he/she can be classified a bad customer or at least is classified as a bad borrower.

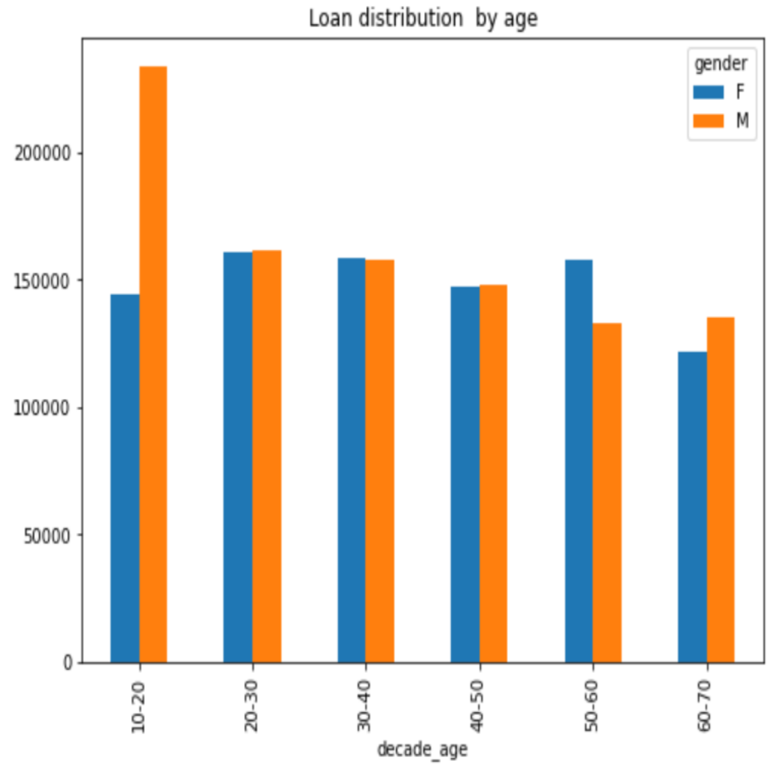
The compilation of all the individual table has been merged in one big table to create a general database. On the database, there is one row per client consisting of all the information regarding the client in the columns. This data mart has been labeled as "database" for the purpose of all charts and analyses.

Data Analyses | Customer Evaluation & Segmentation

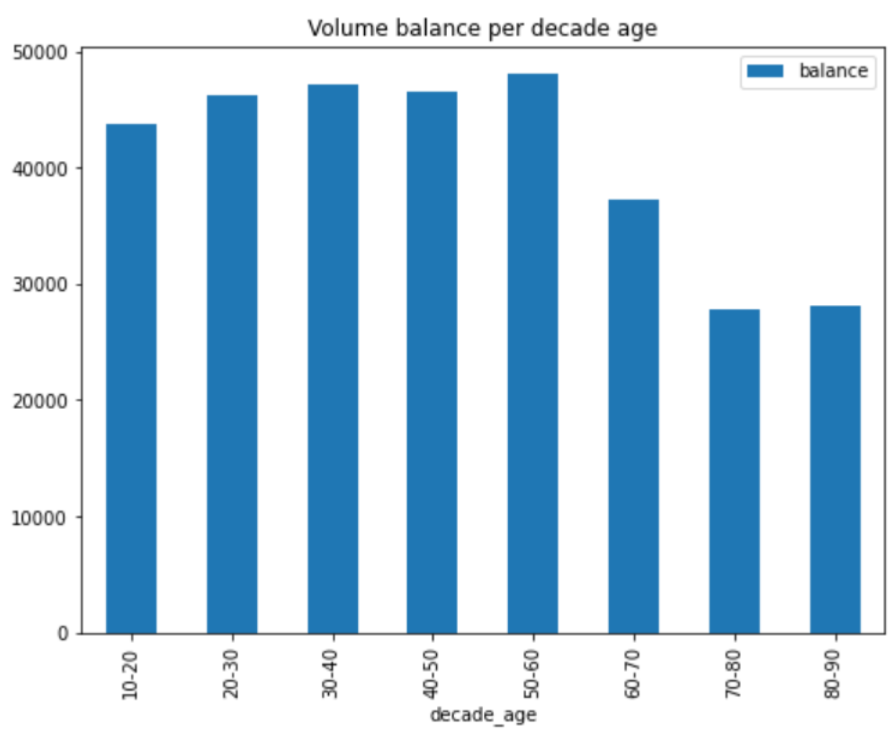
There were several factors that the banking client could analyze their customers and the customer loans. We focused on three primary variables: age, region, and transaction/ loan type. This information was then also helpful to translate the customer insight and provide a potential method to create variables such as credit scores which would help analyze the dataset. Good customers were classified as clients with a positive account balance with all loans paid and zero interest resulting from a negative account balance. Opposingly, bad customers were classified as customers with negative account balances, loans not paid and having an interest rate resulting from a negative account balance.

i. Age

When evaluating customers based on age range, it can be easily noted that men tend to request for higher loan values than women, particularly at the age range between 10-20. This potentially might indicate that the highest loan age range (10-20 years old) is as a result of young students requesting loans for colleges. The highest age range for total good customers or customers who have repaid their loan tend to be approximately between 10-20 years of age (image on next page).



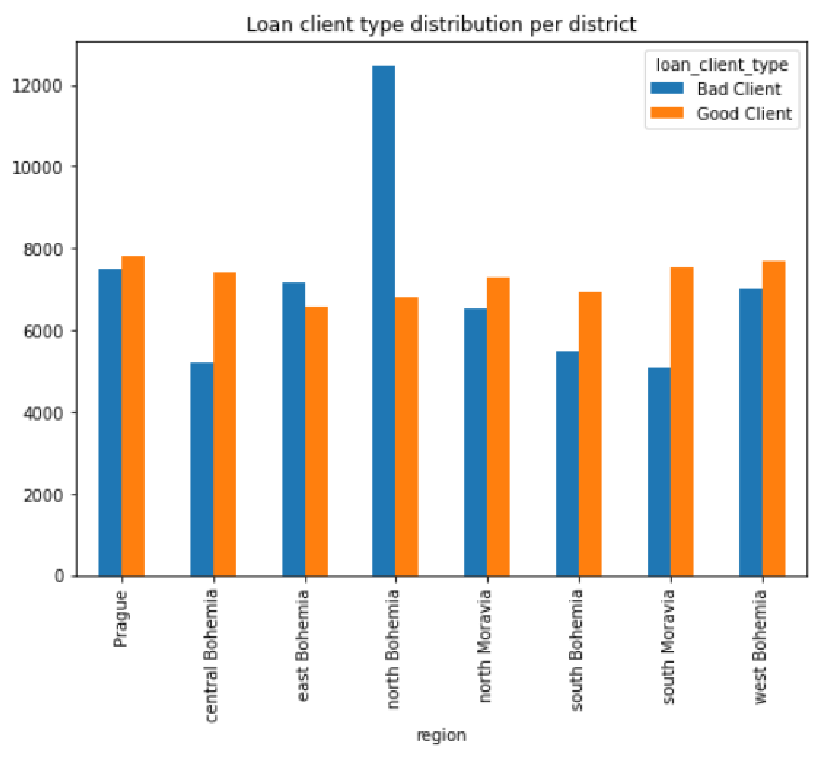
Ironically, a high rate of negative interest rate balance is between 50-60 years of age who also are among those that have the most balance as well (in terms of volume, image below).



In reviewing age as a variable, this data might indicate that a negative interest rate might occur with the older (50-60 age range) generation that has accumulated the most amount of transactions and therefore has a more biased interest rate, while the youngest generation has yet to accumulate such large transactions and has a minimal transaction history, therefore a much lesser negative interests rate (i.e. starting with a clean slate).

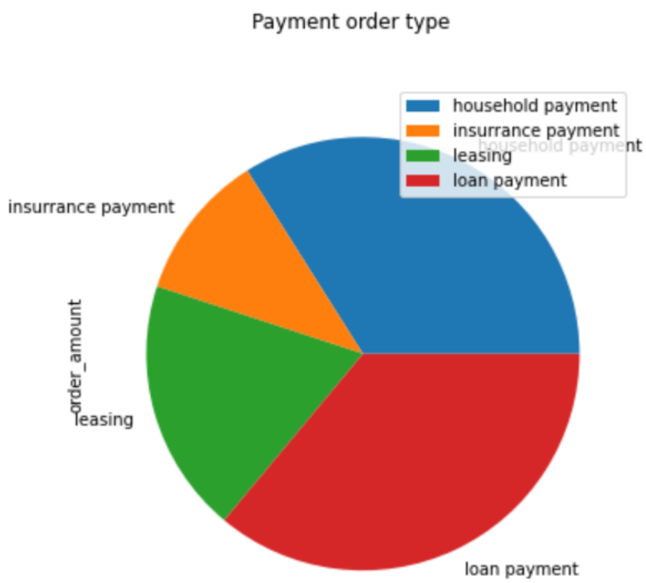
ii. Market Region

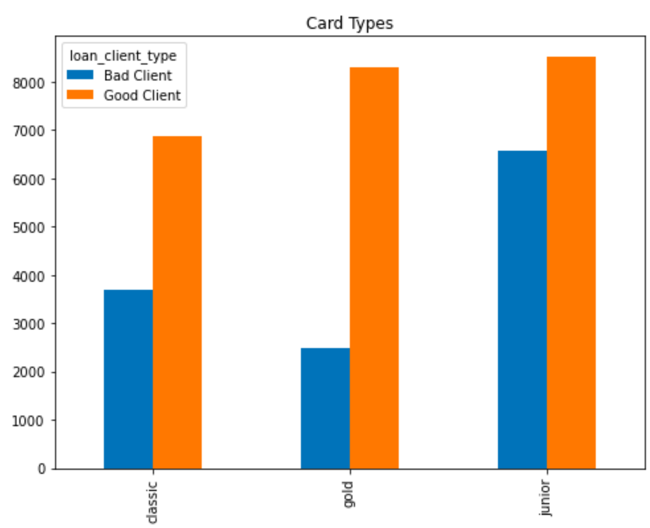
The data indicated eight regions, each having at least 3000 customers for each region. Among those regions, the top three regions with the worst loan payback rate were est Bohemia, central Bohemia, and Prague; ironically, Prague was found to have been listed as the top region to have paid most of their loans and been listed as a good borrowing region. Contrastingly, north Bohemia and west Bohemia had the best loan payback rate. The image below provides an overall perspective on good vs. bad clients by region to help the banking client for a more detailed review by region.



iii. Transaction & Loan Variables

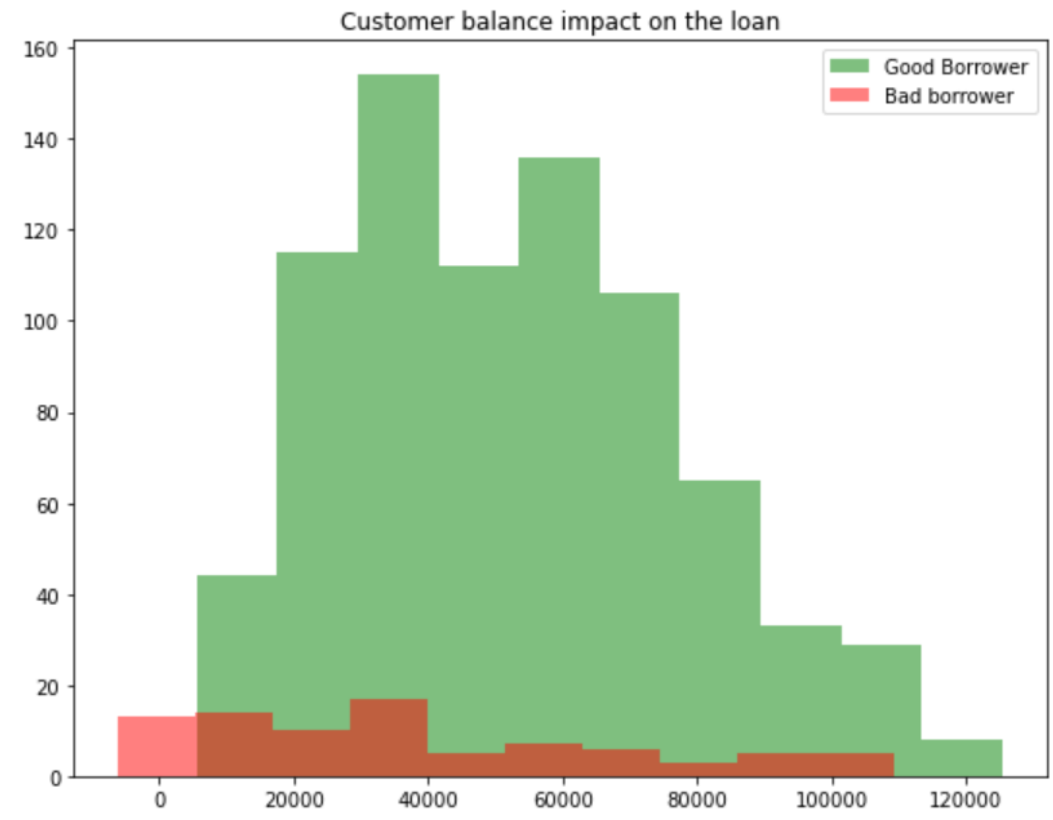
Among the five different payment order types, the highest transactions were classified between household loans and loan payments (image below). When reviewing such transaction types, old-age pension, loan payments and insurance payments were among the top transactions resulting in bank revenue. Interestingly, insurance payment has been the only transaction type to have greater good customers than bad. As you note further for other transactions, card types tend to have a very high ratio of good customers over bad customers.





Conclusion | Final Thoughts & Recommendations

The overall information showcases that there are more good customers than bad (image below). The recommendation based on the three variables conducted showcase that the bank should target between 10-50 age range, with a high focus on college students who have high loan payback rates. The demographic also showed that while women had more loans, they also had more bad loans; therefore, new services the banking client provides should target a more male centric population. Geographically, the bank should be more preventive in north Bohemia there is a real problem, the quantity of loan is way higher and there is almost the double of bad repayment. The team in charge of this place should be have an investigation.



To further hone in on the objective by focusing on specific clients, a simulation of credits score was conducted. The credit score took into account the following calculated field: loan scores, card scores, unemployment rates, seniority, collection from other banks, old-age pensions, and negative interest rates with different coefficient impact (positive and negative). This compilation of calculation resulted in the following top five best and worst clients (below).

	client_id	client_ranking
1	2719	43159.37
2	9717	39824.12
3	9718	39724.12
4	13608	39366.28
5	10269	39045.43
...	...	...
5365	6481	-30682.16
5366	6183	-32070.61
5367	2789	-33315.96
5368	12859	-34394.26
5369	4433	-39357.13

As the values within these variables can easily be manipulated, the recommendation is to work more closely with the banking client to calculate and analyze these values. With more time, we would have liked to see machine learning algorithms provide a better insight as to how the customers in the 1999 banking clients' data would have become with a new additional service such as loan forgiveness for certain clients (as this is being implemented in several countries currently) and whether such a variable could actually provide revenue in the long-term.