

STAT480 Project Report

Author:

Nattanaï Na Songkhla (nn14),

Ram Sabyrkulov (rs46),

Supanut Wanchai (wanchai2),

Ganesh Prasad D K (ganeshd2)

PNC Bank mortgage portfolio analysis and peer comparison

Introduction

In the dynamic landscape of the financial sector, understanding and analyzing mortgage portfolios is crucial for banks to make informed decisions and maintain a competitive edge. PNC Bank, as a key player in the industry, recognizes the need to assess and benchmark its mortgage portfolio against peer institutions. This study aims to delve into the intricacies of PNC's mortgage portfolio, conducting a comprehensive analysis and drawing comparisons with industry leaders such as JPMorgan Chase & Co, Wells Fargo, Bank of America, and BMO Harris. By examining the approval processes across these institutions, valuable insights can be gained into market trends, risk management practices, and potential areas for strategic improvement.

The significance of this analysis lies in its potential to provide PNC Bank with a strategic advantage in the mortgage market. Through a comparative assessment of approved loans among industry peers, PNC can identify areas of strength and areas that may require enhancement. This comparative analysis will not only aid in understanding the competitive landscape but also serve as a benchmark for evaluating the effectiveness of PNC's mortgage approval processes. Additionally, the insights gained from this study can inform strategic decisions, allowing PNC to optimize its mortgage portfolio, enhance risk management strategies, and ultimately improve overall performance in the highly competitive financial market.

Therefore, our objective of this project is to analyze the mortgage portfolio of PNC bank, and compare among our peers including JPMorgan Chase, Wells Fargo, Bank of America, and BMO Harris bank. Also, we aim to increase PNC's mortgage portfolio by defining interesting opportunities or potential customers we can gain more in our portfolio.

Methodology

1. Data Preparation

a. Data Collection

For our data source, we have used the US Mortgage Market and Lenders Data* website to collect the data for Illinois state and concentrate on the 2022 year. We have decided to focus on five major banks that operate in the state of Illinois and have a huge volume of mortgage loans such as PNC Bank, JPMorgan Chase, Wells Fargo, Bank of America, and BMO Harris Bank.

b. Data Preprocessing

In the initial data preprocessing step, our dataset contained 99 columns and 57,684 rows. It can be categorized into three distinct subgroups: Applicant and Co-applicant demographics (examples include Gender, Income, Ethnicity, etc.), Loan characteristics (examples include Loan purpose, Loan amount, etc.), and the Outcome variable, which is the Action taken (Approval/Denial flag).

During the data preparation phase, we strategically selected variables, guided by business insights, the elimination of duplicate fields, and the clarity of variable definitions. Rows exhibiting multiple null values were excluded to maintain data integrity, leaving our dataset with 25 columns and 42,543 rows. Furthermore, after conducting the Lasso method analysis, we identified and removed three insignificant variables. As a result, our dataset now consists of 22 columns and 42,543 rows, ready for further analysis.

*US Mortgage Market and Lenders Data website: <https://ffiec.cfpb.gov/>

2. Modeling Approach

a. Decision Tree

We employed decision trees to all peers and PNC separately. Then, We identified key criterias in the approval process of all the banks and compared key differences. We built decision trees on specific customer segments if it is needed.

b. Clustering

In order to identify our potential groups of customers, we decided to make a K-means clustering model with the optimal number of clusters from the elbow plot to pinpoint the customer groups of each bank, and compare the difference between each group.

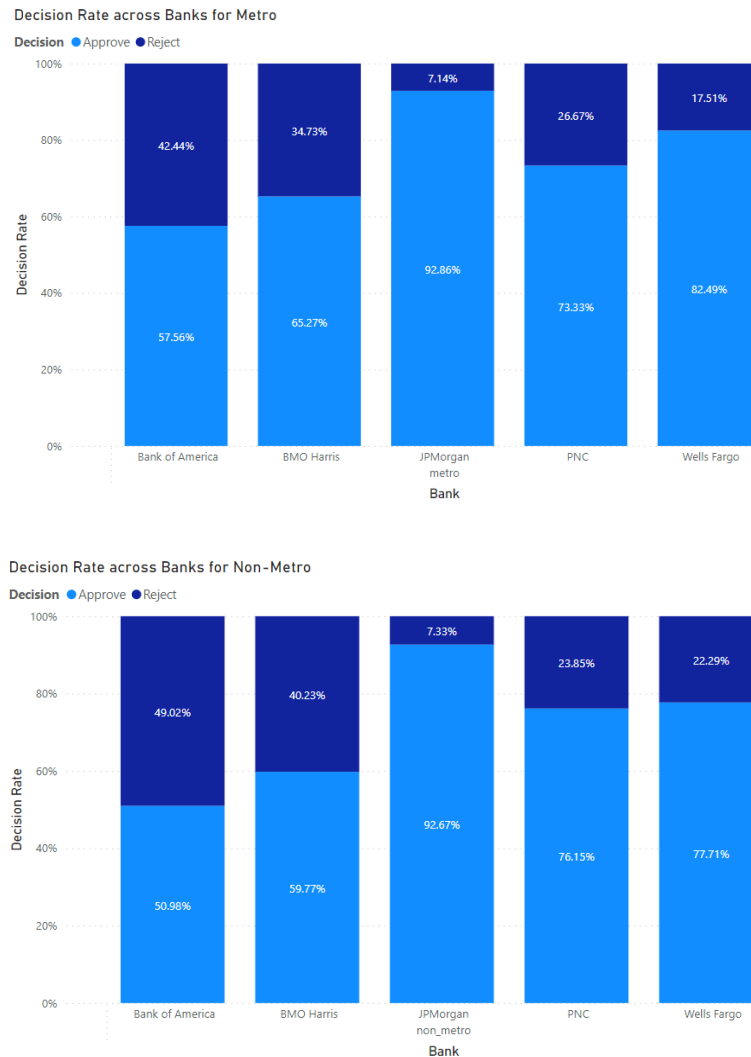
In this research, we focus only on comparing between customer income and property value that is likely to be the most important characteristic of loan approval besides credit score. Additionally, we will consider only PNC Bank and JP Morgan Chase Bank, which have the highest loan and approval rates among all banks.

Model and Results

1. Exploratory Data Analysis (EDA)

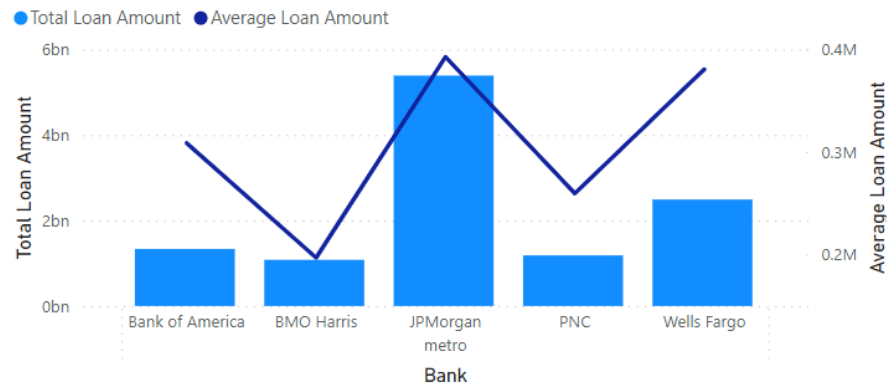
We approached the EDA by analyzing metro (greater Chicago) and non-metro regions separately as we felt the loan operations and target customers for mortgages would be different in the two regions. Using the “metro_area” flag we first looked into the difference in loan approval rate comparing each bank - as the loan approval flag is the outcome variable for modeling. For

both metro region and non-metro region, JP Morgan has the highest loan approval rate of around 93%

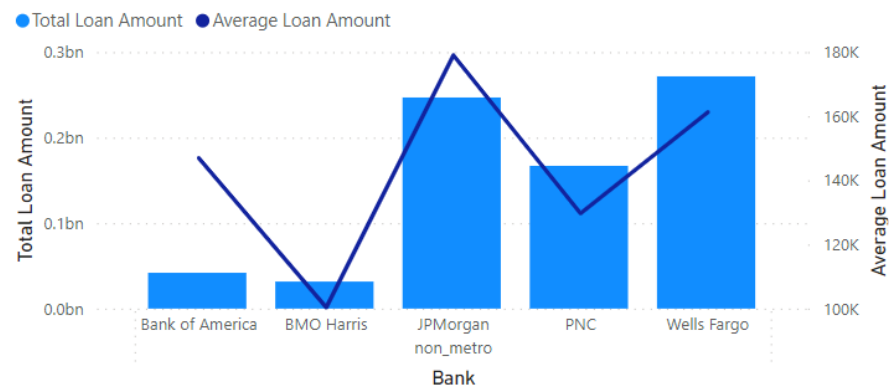


We also wanted to compare the loan size difference across each bank using “Total Loan” and “Average Loan”. This would give us an estimate of lending capacity and market size (among the top 5 banks) for each bank. For the metro region, JP Morgan approved the highest amount of loan and for the non-metro region Wells-Fargo and JP Morgan approved the highest amount of loan.

Loan Amount across Banks for Metro



Loan Amount across Banks for Non-Metro



We also explored ethnicity variables - exploring distribution of different ethnicities across banks, but we did not find any significant differences. Using the results from the above two descriptive analysis - JP Morgan stood out in terms of highest approval rate, market occupancy and loan size.

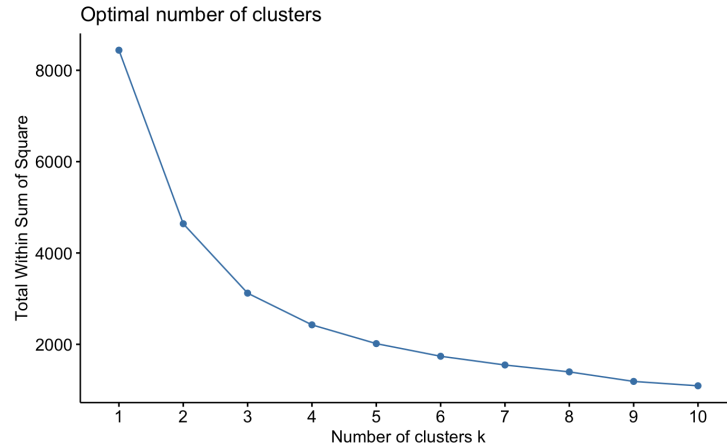
2. Decision Tree

First, we built decision trees on all customers of each bank separately to identify key differences but there is no significant difference. According to the decision tree structure shown in Appendix B, it can be observed that every bank has similar criteria which are the applicant's credit score type and income. Although the banks have different minimum income criteria, they all tend to reject customers whose credit types are 8 or 11. Credit score type 8 means other credit scoring models than commonly used credit score models including Equifax, Experian, FICO, and Vantage. Credit score type 11 means credit score exempt. Thus, we dug into a customer group with commonly used credit scores.

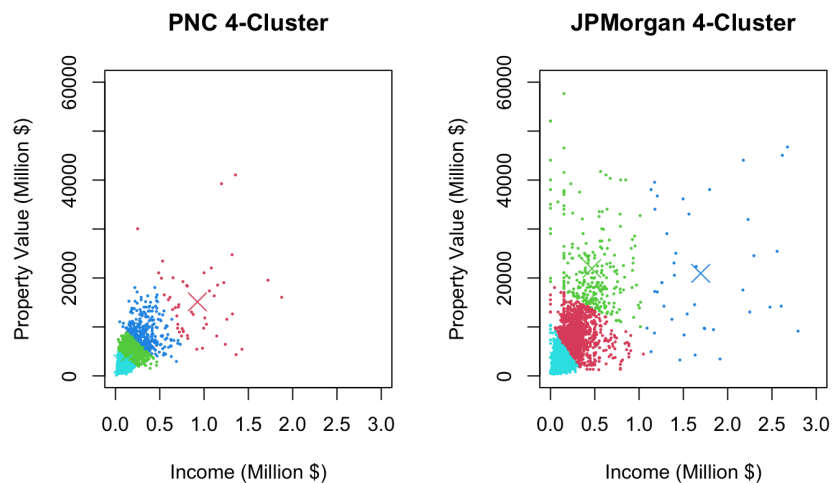
We then built decision trees on only customers with commonly used credit scores (credit score type not 8 and 11). They can be found in Appendix C. We observed that almost all decision trees are complex but JPMorgan Chase's one. Also, we have shown that JP Morgan Chase has the highest mortgage loan amount. Hence, our analysis focused on the comparison between PNC and JPMorgan Chase. The obvious observation from these decision trees is that JPMorgan Chase tends to approve mortgages to customers with lower incomes. Its income threshold is 28 thousand dollars while PNC's is 36 thousand dollars. To further analyze differences in customer segmentation, we employed unsupervised learning techniques for customers who were approved for mortgages by JPMorgan Chase and PNC.

3. Clustering

In the beginning, we clean out some outliers of customers with too much income (greater than 3 million dollars) then standardize all parameters before training the model, in order to increase the model stability.



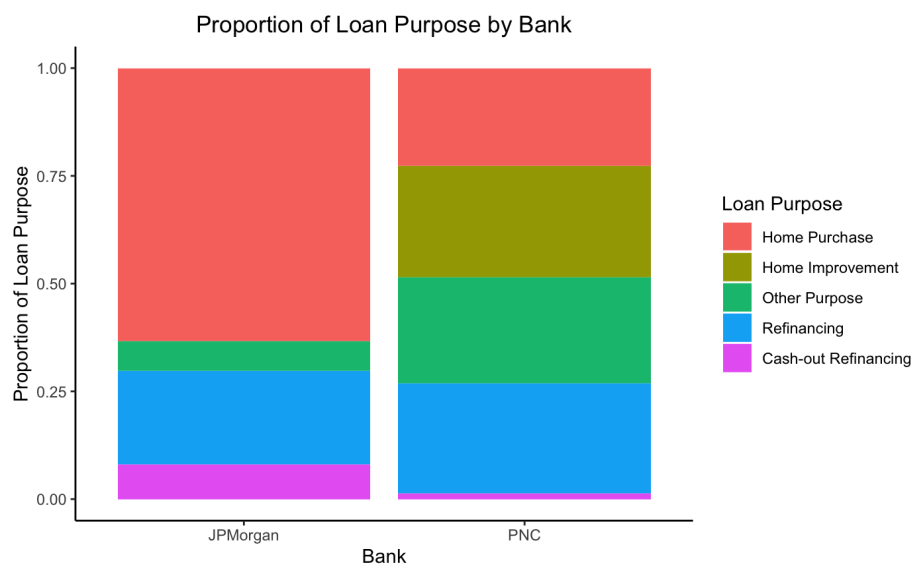
According to the elbow plot using total within sum of squares as a criteria, we decide to choose k equals to 4 as an optimal number of clusters for our k-means clustering model. The result of clustering is shown in figure below, by separating each group by color with marks as a center of the group.



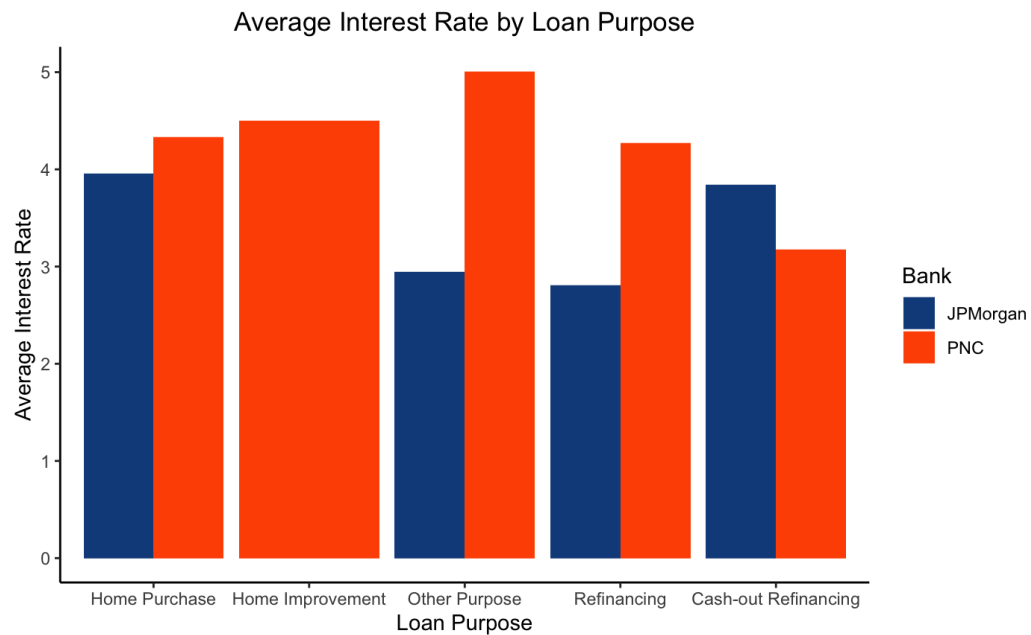
We found that these 4 groups of each bank seem to have a similar characteristic, indicating in the table below.

Cluster number	Characteristic
1	Low income, low property value
2	Low income, medium property value
3	Low income, high property value
4	High income

After we got the result from clustering, we found that cluster 3 is our most interesting group as it showed the most difference between PNC and JP Morgan Chase bank. While JP Morgan Chase bank has significantly more customers in these groups, PNC tends to approve lower numbers of customers with low income, but want to apply for high property. Therefore, we further investigated the loan purpose of these customers and found that home purchase is the largest proportion in JP Morgan Chase bank, which is significantly different from PNC as can be seen in the plot below.



Moreover, we wanted to know more about why the number of customers with the purpose of home purchase is significantly different between two banks, so we further investigated on others characteristics, and found some interesting insight that these 2 banks offer different interest rates, which could result in a different number of customers applying for the loan.



According to the average interest rate by loan purpose comparison, JP Morgan Chase bank seems to offer lower interest rate for most of the purpose, with slightly lower for home purchase. Therefore, we draw an assumption that this lower rate of interest in home purchase resulted in a higher number of customers applying for the loan at JP Morgan Chase bank than PNC, if they want to buy a house.

Summary and Recommendation

To gain the mortgage portfolio, our potential customer segment is the group of low income but applying for high property value, which is the segment that differentiates PNC and JPMorgan Chase. While the customers in this segment have similar income, JPMorgan Chase tended to approve much greater loan values than PNC. Thus, we should perform some feasibility studies on this group of customers to better understand the behavior and characteristics.

In addition, the majority loan purpose of this segment of JPMorgan Chase was home purchase, around 60%. In contrast, PNC's loan purposes are split equally among home purchase, home improvement, refinancing, and other purposes. We suggested that PNC should focus more on acquiring new customers who want to purchase homes than retention of old customers through refinancing and home improvement.

Finally, we recommended studying more on the interest of both the lending and funding sides. PNC should lower interest rates to compete with JPMorgan Chase if it is possible. Moreover, reducing funding costs such as saving interest rates would also help increase the possibility to lower mortgage interest.

Discussion

Our analysis and the resulting recommendations are firmly rooted in data-driven insights. However, it's crucial to recognize that the mortgage loan sector is influenced by various external factors beyond our dataset. To enhance PNC's mortgage portfolio, we would like to discuss the following assumptions:

JP Morgan Bank is known for having the largest client base, which provides direct access to their clients' card transactions and payment behaviors. This dominance is not just a marker of scale but a critical predictor of an applicant's ability to manage and allocate finances effectively. Understanding these dynamics could provide invaluable insights, such as internal credit scores, for strategic decision-making in the mortgage sector.

Another factor that enables JP Morgan to offer lower interest rates and lend more money to clients with lower household incomes is their large cash flow, stemming from their substantial client base. In contrast, PNC Bank must be more selective in its approval decisions due to its relatively lower cash flow and client base. This indicates that PNC faces greater risk, and as a result, its interest rates might be higher and need to be more selective to approve the loan.

Lastly, we believe that marketing campaigns play a significant role in JP Morgan's brand awareness. The JP Morgan brand has a strong reputation among lending banks in the country, making it easier for them to attract more applicants compared to PNC Bank.

Appendix

1. Appendix A: List of variables in the model

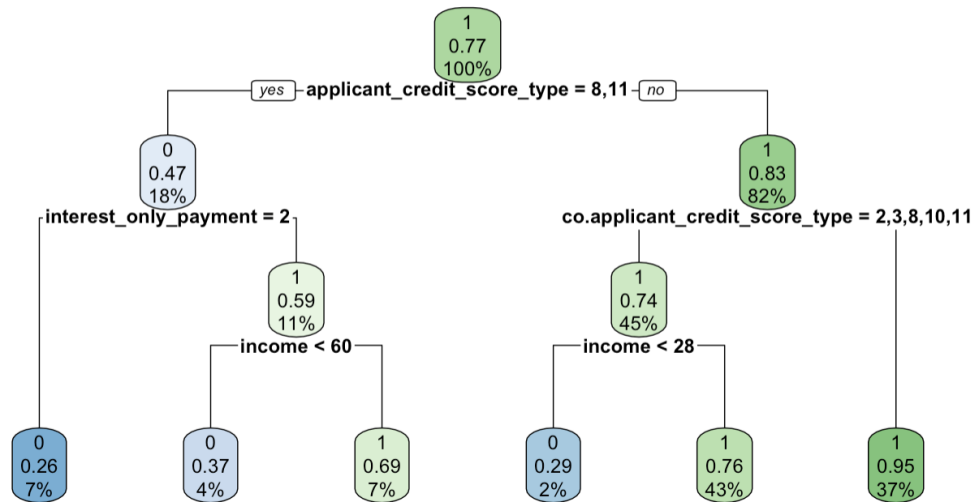
Final 22 variables after preprocessing that were used in the model:

"derived_ethnicity", "derived_race", "derived_sex", "loan_type", "loan_purpose", "lien_status",
"open.end_line_of_credit", "business_or_commercial_purpose", "loan_amount",
"interest_only_payment", "balloon_payment", "other_nonamortizing_features",
"property_value", "occupancy_type", "manufactured_home_secured_property_type",
"manufactured_home_land_property_interest", "metro" "income",
"applicant_credit_score_type", "co.applicant_credit_score_type", "Bank", "approve_flag"

The definitions for these variables can be found in the [dataset documentation](#).

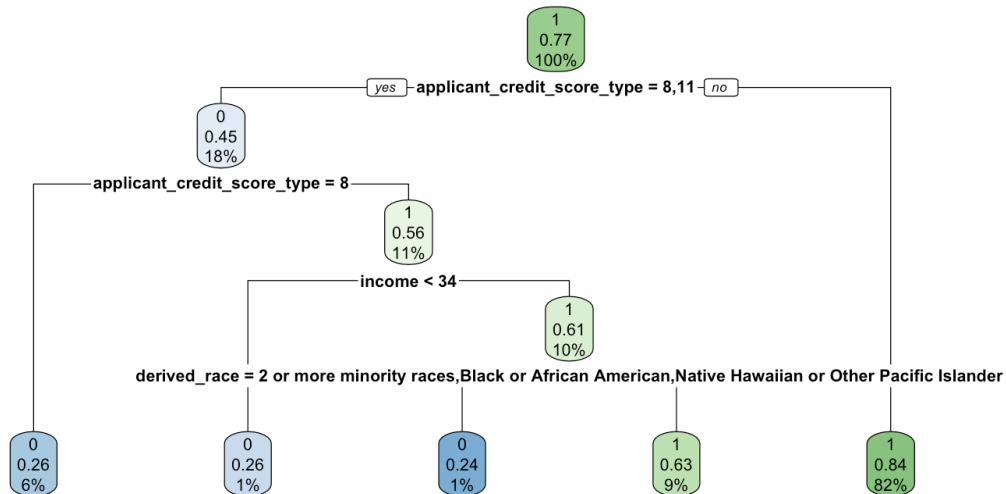
2. Appendix B: Decision trees on all customers by bank

JPMorgan Chase & Co



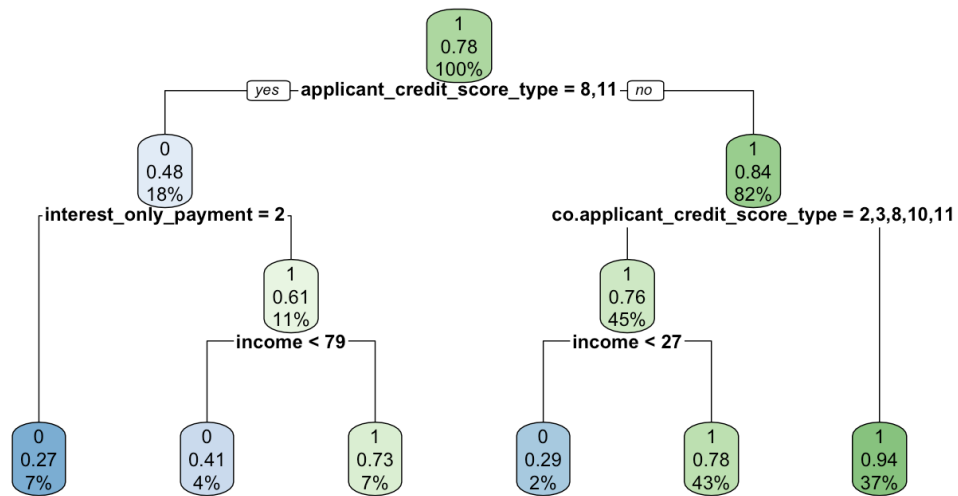
(Decision tree of all customers for JPMorgan Chase)

Bank of America



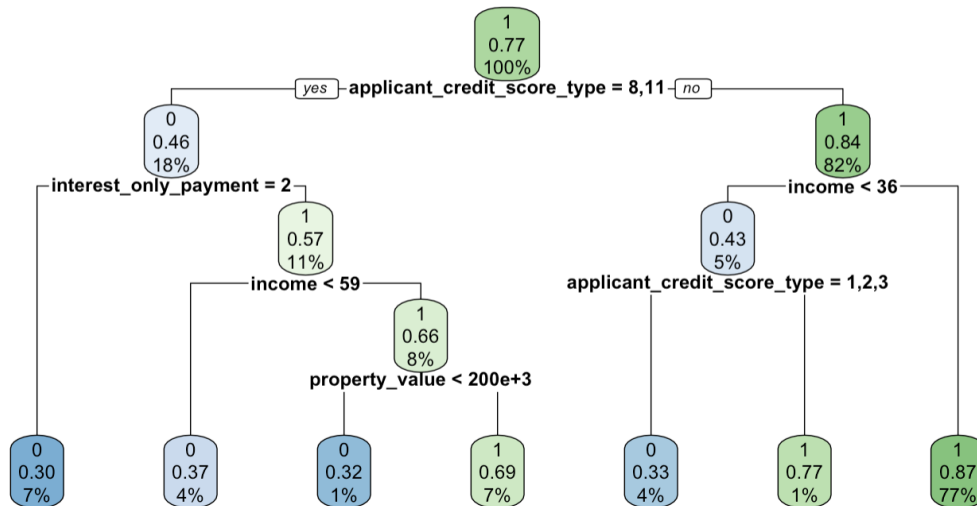
(Decision tree of all customers for Bank of America)

BMO Harris



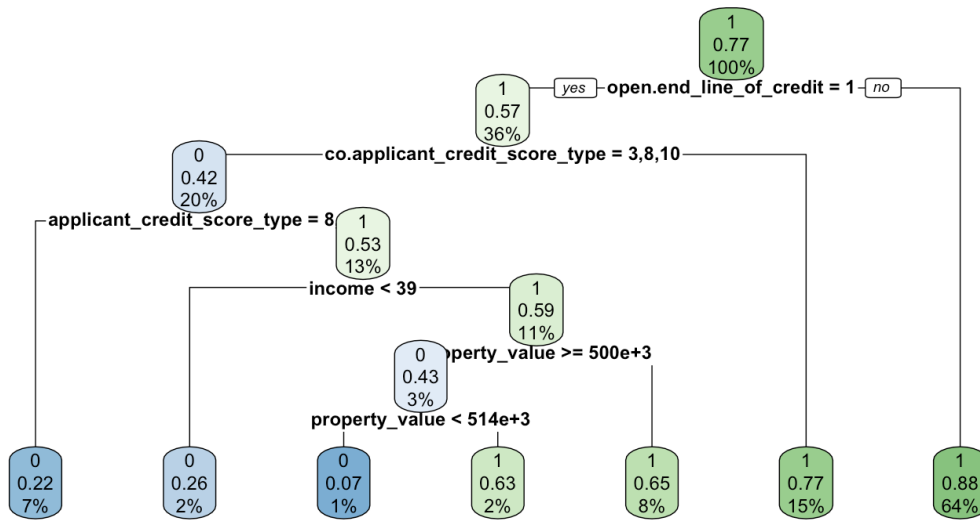
(Decision tree of all customers for BMO Harris)

PNC



(Decision tree of all customers for PNC)

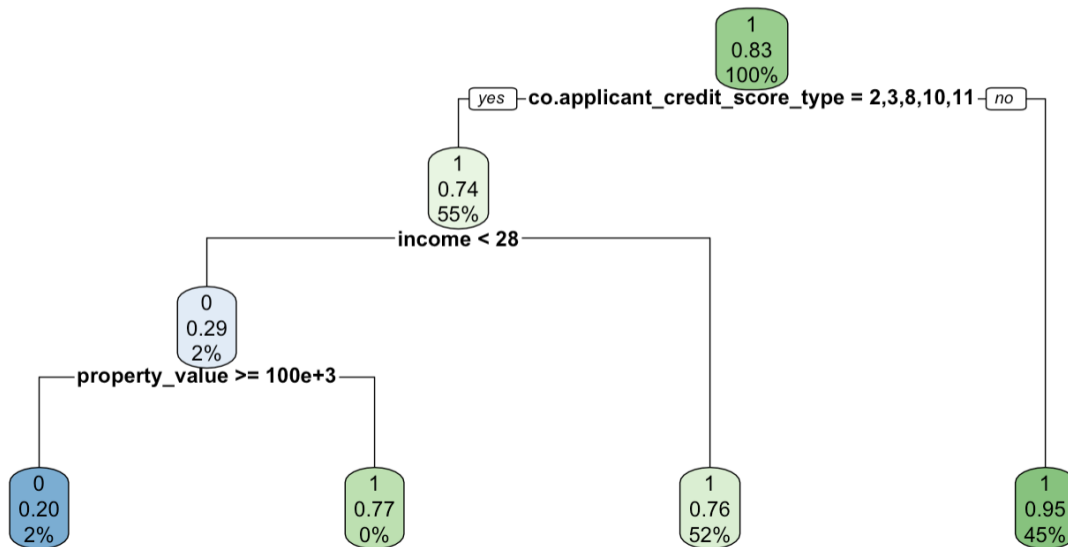
Wells Fargo



(Decision tree of all customers for Wells Fargo)

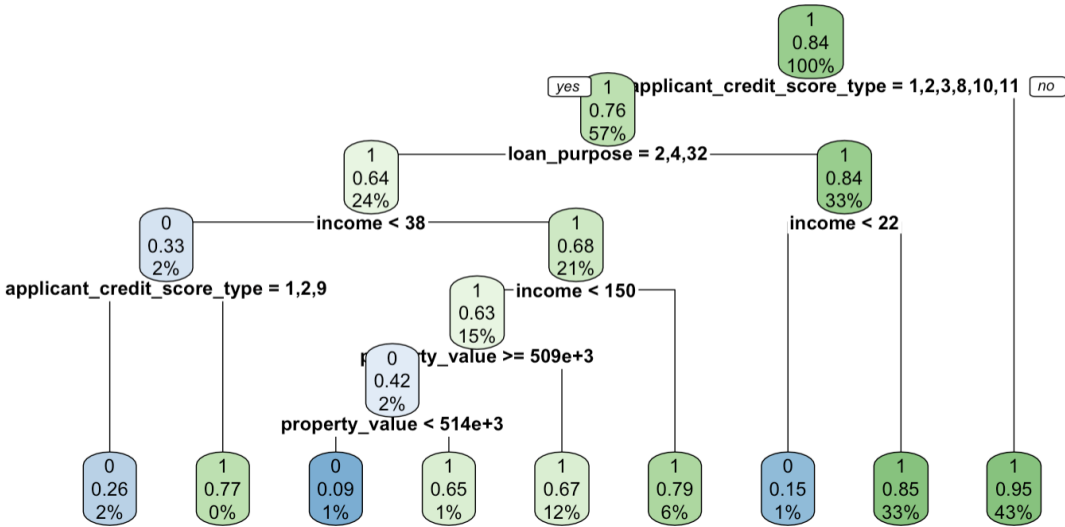
3. Appendix C: Decision trees on customers with commonly used credit score by bank

JPMorgan Chase & Co (credit type not 8,11)



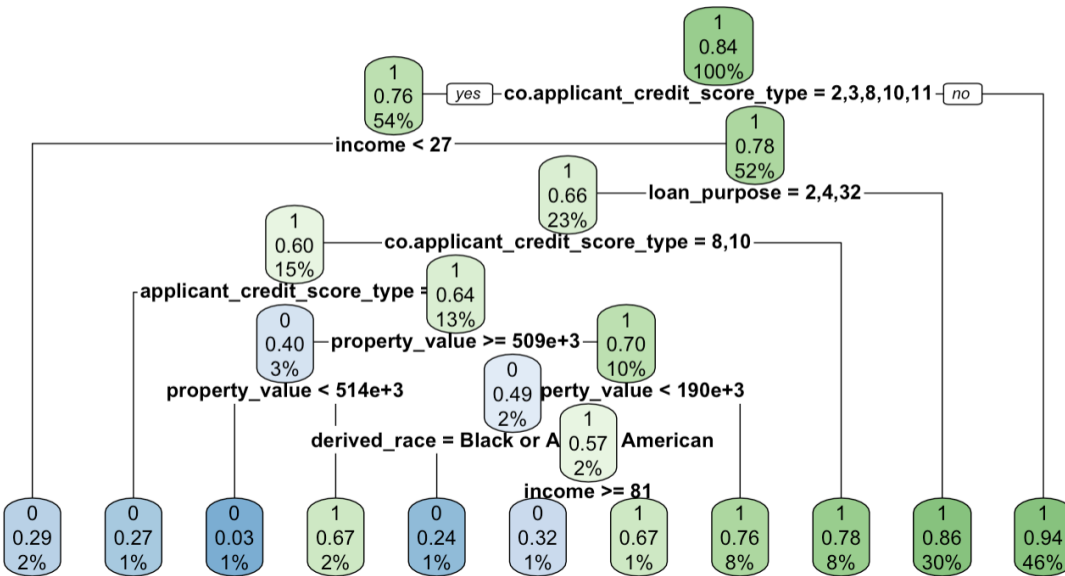
(Decision tree of customers with commonly used credit score type for JPMorgan Chase)

Bank of America (credit type not 8,11)



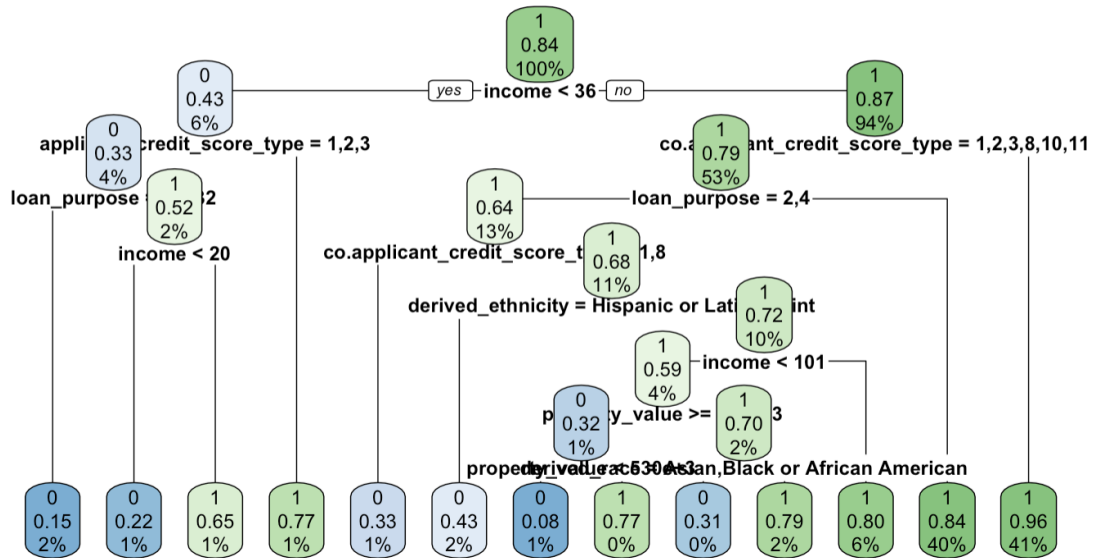
(Decision tree of customers with commonly used credit score type for Bank of America)

BMO (credit type not 8,11)



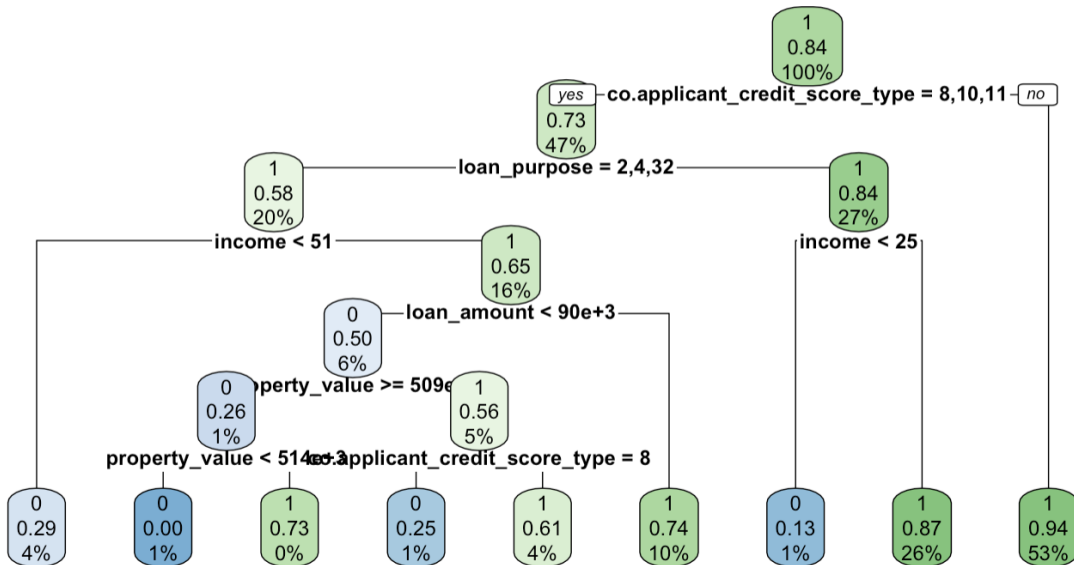
(Decision tree of customers with commonly used credit score type for BMO Harris)

PNC (credit type not 8,11)



(Decision tree of customers with commonly used credit score type for PNC)

Wells Fargo (credit type not 8,11)



(Decision tree of customers with commonly used credit score type for Wells Fargo)