Capstone Project

Final Submission

Loan Default Prediction

MIT-PE ADSP

Andrew Cameron

April 16, 2025

Contents / Agenda

- Executive Summary
- Problem Definition
- Data Exploration
- Building Models
- Techniques’ Comparison
- Final Solution Design

Executive summary

Introduction

This document proposes an efficient machine learning model to improve loan evaluation, minimizing the bank's risk of bad loans and financial losses. The current loan assessment system relies on manual evaluation and limited automation, making it labor-intensive, error-prone, and susceptible to bias. A more efficient, unbiased process is crucial for banks to stay competitive and uphold customer trust.

The HMEQ dataset includes baseline and loan performance data for 5,960 home equity loans. The target variable (BAD) indicates loan default or severe delinquency, occurring in 1,189 cases (20%). It contains 12 input variables per applicant.

Analysis and classification modeling revealed that total defaulted loans amount to approximately $20 million, highlighting the severe financial impact of incorrectly assessing loan applications. Three classification models—Logistic Regression, Decision Tree, and Random Forest—were considered for evaluating the BAD target variable.

The analysis began with Logistic Regression as a baseline. Model performance was assessed using accuracy, confusion matrix, precision, recall, and F1-score. To improve accuracy, Decision Trees and Random Forests were explored, with hyperparameter tuning (e.g., GridSearchCV) applied for optimization.

Final proposed model specifications

The final proposed model for this classification exercise will be the 'Decision Tree' with a test recall of 0.6942 for class 1 (= client defaulted on loan).   This model was also selected for its interpretability (a benefit for bank stakeholders).

Problem & Solution Summary (Key points for the final proposed solution design)

Model evaluation criteria:

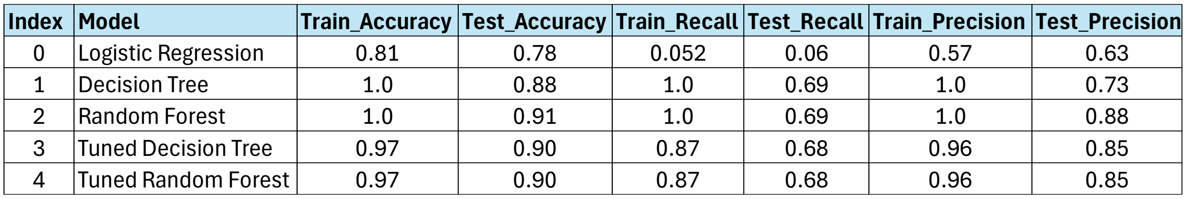
The model can make two types of wrong predictions:

1. Predicting a client will default on a loan when they would not actually default.
2. Predicting a client will not default on a loan when the client defaults.

Predicting that the client will not default on a loan but *does* default on a loan, i.e., resulting in significant losses for the bank would be considered a significant error and hence this would be a more important case of wrong predictions. Thus, a key question will be: How to reduce this loss? i.e., the need to reduce False Negatives.

The bank will want Recall to be maximized. The greater the Recall, the higher the chances of minimizing the false negatives. Therefore, the final proposed model for this classification exercise will be the 'Decision Tree' with a test recall of 0.6942 for class 1 (= client defaulted on loan).

Model Comparison



Index	Model	Train_Accuracy	Test_Accuracy	Train_Recall	Test_Recall	Train_Precision	Test_Precision
0	Logistic Regression	0.81	0.78	0.052	0.06	0.57	0.63
1	Decision Tree	1.0	0.88	1.0	0.69	1.0	0.73
2	Random Forest	1.0	0.91	1.0	0.69	1.0	0.88
3	Tuned Decision Tree	0.97	0.90	0.87	0.68	0.96	0.85
4	Tuned Random Forest	0.97	0.90	0.87	0.68	0.96	0.85

Key Next Steps

The model has achieved a recall of 70% which is an acceptable threshold. However, if the bank wishes to improve the recall even further, the recommendation would be to consider further tuning where appropriate and feasible.

Next steps would be to put the model into production by preparing it in a format that can be easily used and set up a stable environment for deployment. An accessible system such as an API (Application Programming Interface) will enable data exchange and automation to be accessible to the required team members. Performance can be monitored (for things such as drift) and regular updates made available, with security and compliance with all relevant regulations.

The model was selected for its optimal performance metrics, enabling the bank to reduce default risk and financial losses. Since minimizing false negatives is the priority (i.e., making sure to reduce the situation of not catching applications with a true default), Recall is more important than Precision.

Recommendations for implementation: Key Actionable for Stakeholders

- Model Deployment: Prepare and integrate the Decision Tree model into the bank's loan evaluation system.
- Performance Monitoring: Establish tracking mechanisms to assess accuracy, recall, and data drift.
- Regulatory Compliance: Adherence to financial regulations & data privacy standards.
- Stakeholder Training: Educate teams on model interpretation, decision-making, and risk management.

- Continuous Improvement: Explore further optimization through hyperparameter tuning or additional data sources.

Expected Benefit & Costs

- Benefits: Improved risk assessment reduced financial losses, enhanced efficiency, and better customer trust.
- Costs: Implementation expenses, infrastructure setup, ongoing monitoring, and periodic retraining.

Solution Benefits

- Risk Reduction: Higher recall minimizes false negatives, lowering bad loan occurrences.
- Operational Efficiency: Automation reduces manual workload and accelerates loan processing.
- Data-Driven Decision Making: Helps improve loan approval strategies based on empirical insights.

Estimated Costs/Benefits (Assumptions)

- Potential Savings: Reducing bad loans could save the bank $5M–$10M annually, assuming recall improvements lead to a 25-50% reduction in missed default cases.
- Implementation Costs: Model development and deployment could cost $500K–$1M, including infrastructure, compliance, and training.

Key Risks & Challenges

- Model Bias: If training data lacks diversity, the model may reinforce biases.
- Regulatory Constraints: Compliance with financial laws & consumer protection policies.
- Data Drift: Changing economic conditions may reduce model effectiveness over time.

Further Analysis Required

- Economic Impact Modeling: Assess long-term financial gains from improved risk assessment.
- Alternative Model Comparisons: Evaluate if more complex models offer better performance.
- Explainability & Transparency: Ensure the model's decisions can be justified to regulators and customers.

Problem Definition

The context:

Banks rely heavily on the revenue stream from loan interest. Managing loans carefully ensures a steady flow of revenue and helps the bank to grow and keep operations running smoothly. Banks need to evaluate loan applicants carefully to reduce the risk of bad loans and related financial losses. By improving their loan applicant evaluations, banks will also be able to lend to more people and businesses, thereby boosting the economy and further maintaining trust with legitimate customers and investors.

The current system for assessing loan applications is largely based on manual evaluation as well as some automation using heuristics. This process has several drawbacks. It is labor-intensive, error-prone and allows for bias due to human judgement. Designing a loan approval process that is more efficient and freer from bias is essential for banks to remain competitive and maintain strong relationships with customers and reputation.

The objectives:

The goal is to create a classification model that predicts which customers are likely to default on their loans and advise the bank on key factors for loan approval.

An additional benefit to be gained by achieving the objectives can be improved labor productivity and better deployment of human resources. Additionally, the bank will benefit from an improved reputation and strengthened balance statement. This will also protect consumers from entering financially difficult circumstances.

The key questions:

- Which classification model best predicts the probability of loan default?
- Which features are most impactful for building an accurate and reliable classification model to predict loan defaults and reduce bad loans?
- What are characteristics of applicants with BAD=1 vs. BAD = 0?
- Are there any strong correlations between independent variables and the target variable (BAD)?
- Do attributes such as years on the job or type of job influence the likelihood to default?
- How do variables that have information effecting credit history (e.g., DEROG) contribute to predicting loan default?
- What is the significance of missing variables in the data set and how should they be treated?
- What strategies can be used to balance the data (e.g., 20% = 1, 80% =0 for BAD).

The problem formulation:

Using a reliable and relatively robust data set (HMEQ) we are trying to find an accurate, dependable and unbiased way of predicting loan default so that we can streamline decision-making and maintain healthy revenue growth for the bank.

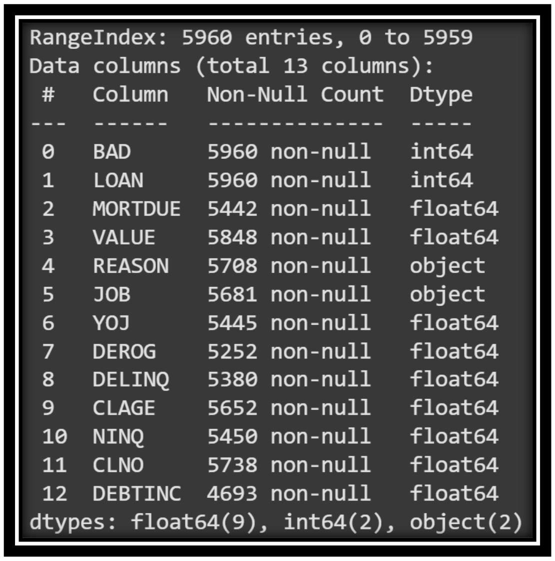
Using data science, we will utilize the data set to train, test and build a classification model that can predict which will predict clients likely to default on their loans (current clients) while also giving insights into what to consider when approving loans (future, potential clients).

Data Exploration

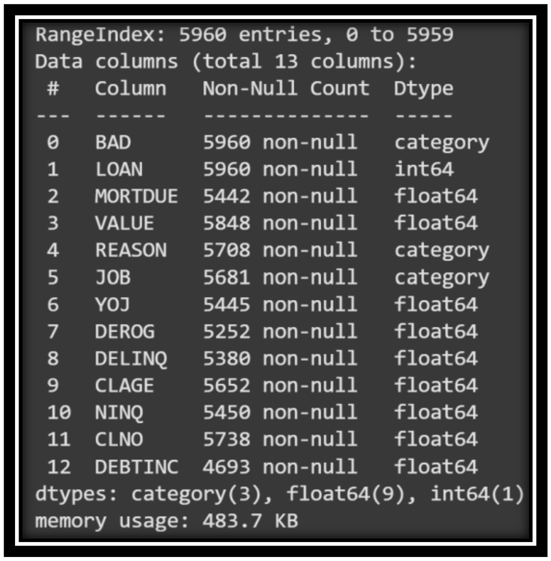
Data Description:

The dataset (HMEQ) contains baseline and loan performance information for 5,960 recent home equity loans. The target (BAD) is a binary variable that indicates whether an applicant has defaulted or has been severely delinquent. This adverse outcome occurred in 1,189 cases (20 percent). Twelve input variables were registered for each applicant.

Data Types (before and after conversion of object type to categories)



```
RangeIndex: 5960 entries, 0 to 5959
Data columns (total 13 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   BAD      5960 non-null   int64
 1   LOAN     5960 non-null   int64
 2   MORTDUE  5442 non-null   float64
 3   VALUE    5848 non-null   float64
 4   REASON   5708 non-null   object
 5   JOB      5681 non-null   object
 6   YOJ      5445 non-null   float64
 7   DEROG    5252 non-null   float64
 8   DELINQ   5380 non-null   float64
 9   CLAGE    5652 non-null   float64
 10  NINQ     5450 non-null   float64
 11  CLNO     5738 non-null   float64
 12  DEBTINC  4693 non-null   float64
dtypes: float64(9), int64(2), object(2)
```

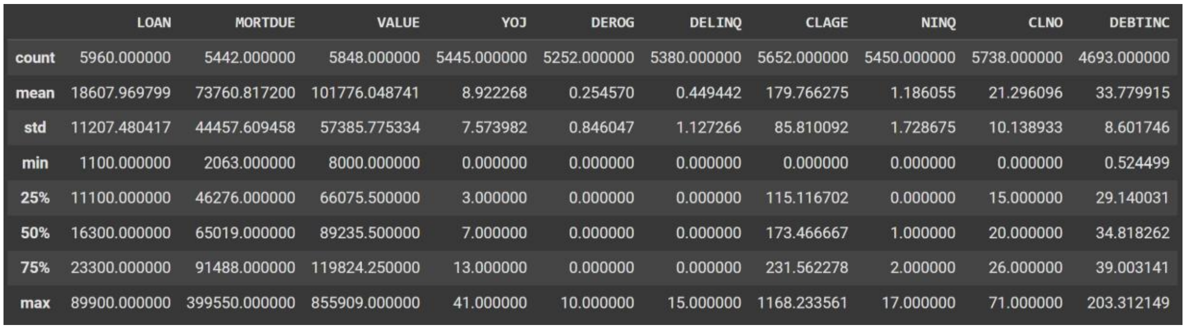


```
RangeIndex: 5960 entries, 0 to 5959
Data columns (total 13 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   BAD      5960 non-null   category
 1   LOAN     5960 non-null   int64
 2   MORTDUE  5442 non-null   float64
 3   VALUE    5848 non-null   float64
 4   REASON   5708 non-null   category
 5   JOB      5681 non-null   category
 6   YOJ      5445 non-null   float64
 7   DEROG    5252 non-null   float64
 8   DELINQ   5380 non-null   float64
 9   CLAGE    5652 non-null   float64
 10  NINQ     5450 non-null   float64
 11  CLNO     5738 non-null   float64
 12  DEBTINC  4693 non-null   float64
dtypes: category(3), float64(9), int64(1)
memory usage: 483.7 KB
```

Initial Observations & Insights:

Key Patterns in the Data

Summary Statistics of the Dataset

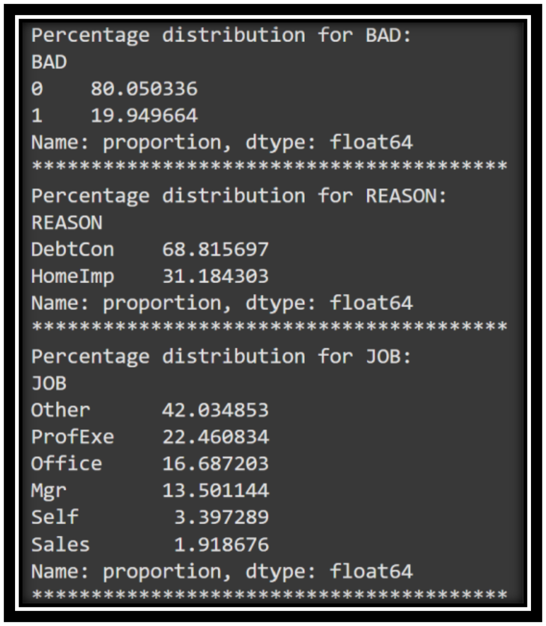


	LOAN	MORTDUE	VALUE	YOJ	DEROG	DELINQ	CLAGE	NINQ	CLNO	DEBTINC
count	5960.000000	5442.000000	5848.000000	5445.000000	5252.000000	5380.000000	5652.000000	5450.000000	5738.000000	4693.000000
mean	18607.969799	73760.817200	101776.048741	8.922268	0.254570	0.449442	179.766275	1.186055	21.296096	33.779915
std	11207.480417	44457.609458	57385.775334	7.573982	0.846047	1.127266	85.810092	1.728675	10.138933	8.601746
min	1100.000000	2063.000000	8000.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.524499
25%	11100.000000	46276.000000	66075.500000	3.000000	0.000000	0.000000	115.116702	0.000000	15.000000	29.140031
50%	16300.000000	65019.000000	89235.500000	7.000000	0.000000	0.000000	173.466667	1.000000	20.000000	34.818262
75%	23300.000000	91488.000000	119824.250000	13.000000	0.000000	0.000000	231.562278	2.000000	26.000000	39.003141
max	89900.000000	399550.000000	855909.000000	41.000000	10.000000	15.000000	1168.233561	17.000000	71.000000	203.312149

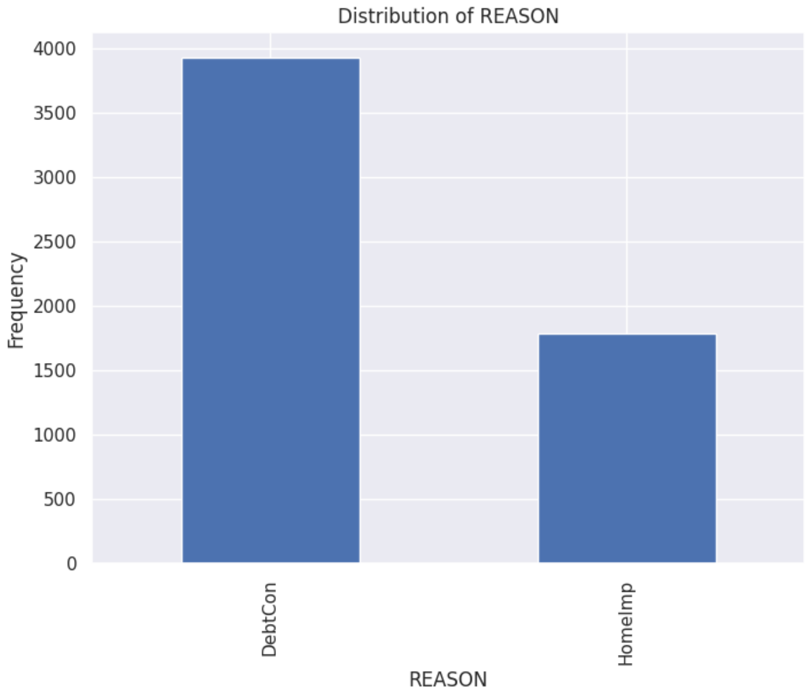
Numerical Variables in the dataset

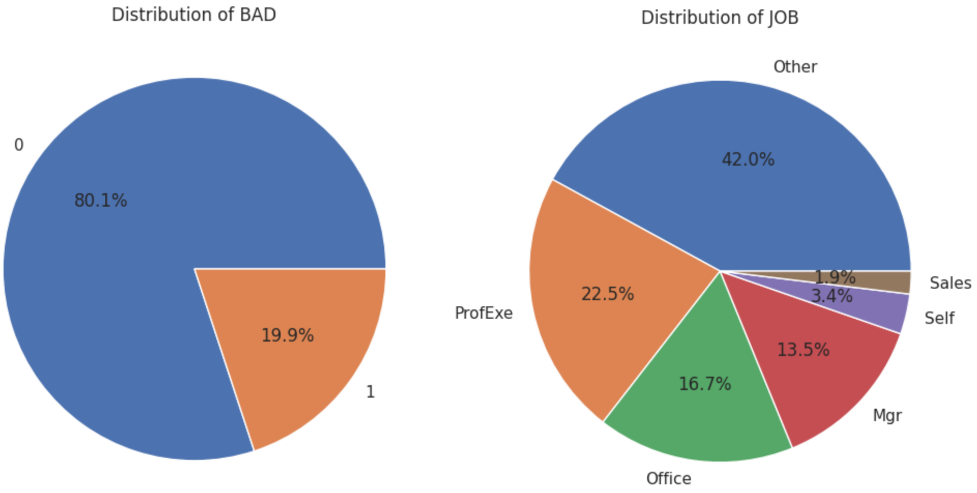
- LOAN: Averages about $18,608, ranging from $1,100 to $89,900, indicating diverse borrowing amounts.
- MORTDUE: Mortgage dues average $73,761, with some reaching up to $399,550
- VALUE: Home values average $101,776, with extremes from $8,000 to $855,909
- YOJ (Years on Job): Most people have been at their job for around 9 years, with a max of 41.
- DEROG and DELINQ (Credit reports): Both mostly hover near zero, but a few serious outliers exist, like 10 and 15 respectively.
- CLAGE (Credit age): Average credit line age is 180 months, maxing out at 1168 months.
- NINQ: Recent inquiries typically fall around 1, but some individuals had up to 17 inquiries.
- CLNO: Averages 21 credit lines, reaching as high as 71.
- DEBTINC: Average debt-to-income ratio is 33.78, with values peaking around 70.1%.

Summary of categorical data



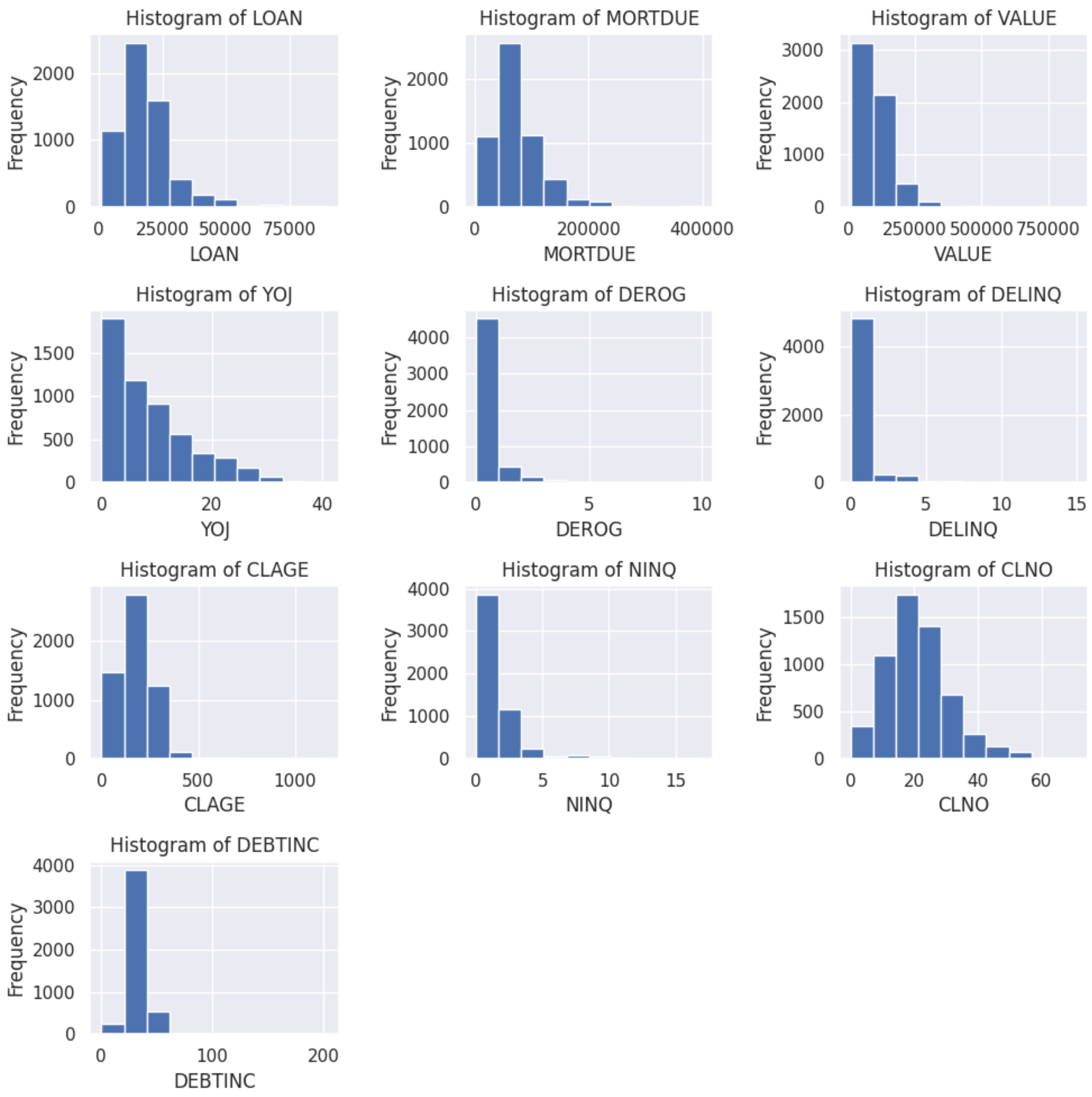
```
Percentage distribution for BAD:
BAD
0    80.050336
1    19.949664
Name: proportion, dtype: float64
****************************************
Percentage distribution for REASON:
REASON
DebtCon    68.815697
HomeImp    31.184303
Name: proportion, dtype: float64
****************************************
Percentage distribution for JOB:
JOB
Other      42.034853
ProfExe    22.460834
Office     16.687203
Mgr        13.501144
Self        3.397289
Sales       1.918676
Name: proportion, dtype: float64
****************************************
```





Univariate Analysis – Numerical Data

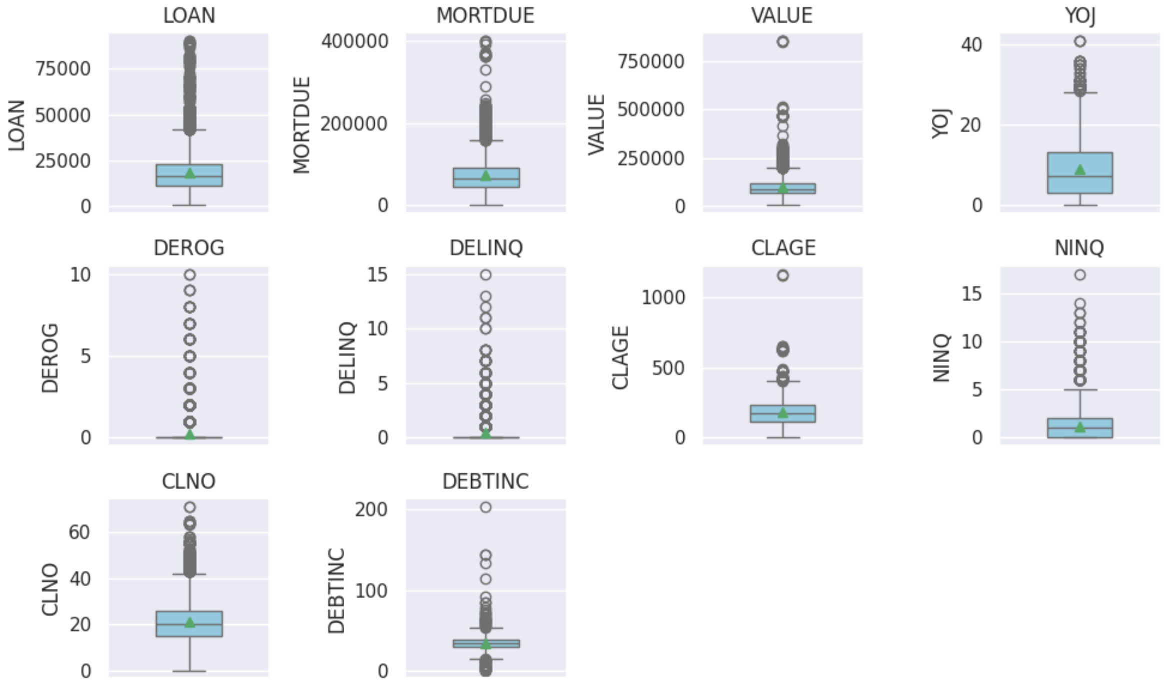
The following visualizations give a view into the financial health and behaviors of the dataset.
Histograms



- LOAN & MORTDUE: Most loans and mortgages seem to concentrate in the lower ranges, suggesting a preference for smaller, manageable amounts.
- VALUE: Property values show a notable spread, but many cluster under $250,000, hinting at a market dominated by mid-range properties.
- YOJ (Years on Job): Many applicants have under 20 years of job tenure, with a dip as the years increase.
- DEROG & DELINQ: These low-frequency peaks indicate that most individuals maintain clean records, with very few experiencing serious delinquencies.
- CLAGE (Credit Line Age): A maturity-heavy distribution, as many credit lines are older, reflecting stable credit usage.
- NINQ (Recent Credit Inquiries): The histogram shows that recent credit checks are minimal for most, suggesting cautious borrowing.

- DEBTINC: A majority fall under the ideal debt-to-income ratio of 36%, but there are outliers with significantly higher values, likely under stress.

Univariate Analysis – Boxplots of Numerical Data



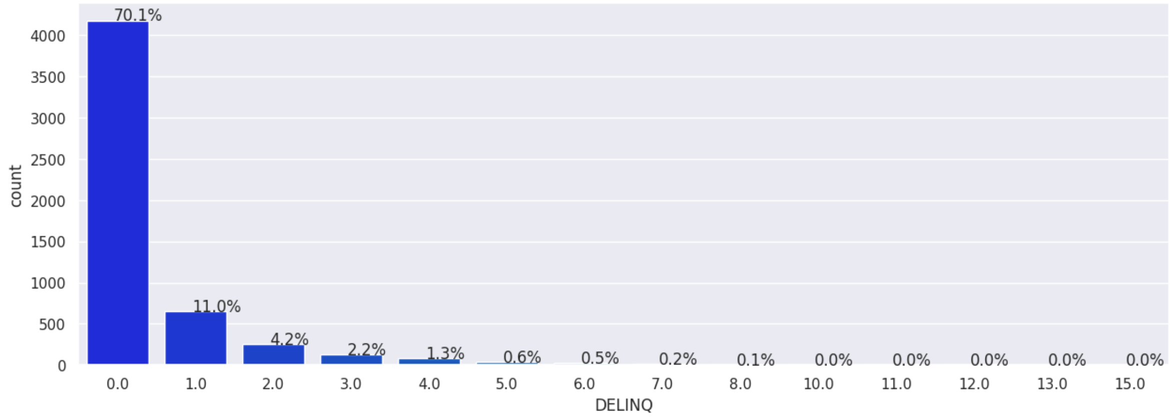
- LOAN: Most loans cluster below $25,000, but there are significant outliers reaching beyond $75,000.
- MORTDUE: Mortgage dues show a widespread pattern, with a median around $100,000, indicating diverse borrowing patterns. Outliers extend significantly.
- VALUE: Property values appear quite varied, with most centered around $150,000.
- YOJ: Most applicants have job tenures below 20 years, with a median around 10 years.
- DEROG & DELINQ: Many applicants maintain clean records (both medians at 0), though a few have notable derogatory reports and delinquencies as outliers.
- CLAGE: Credit line age reflects financial maturity, with a median near 200 months (~16.7 years). Some outliers suggest longstanding credit histories.
- NINQ: Most individuals have few recent inquiries (median close to 0), but outliers exceed 10.
- CLNO: Credit lines often hover below 40, but a few outliers boast higher numbers.
- DEBTINC: The debt-to-income ratio showcases a healthy concentration around 30-40%, though extreme outliers highlight some who may face financial strain.

Skewness

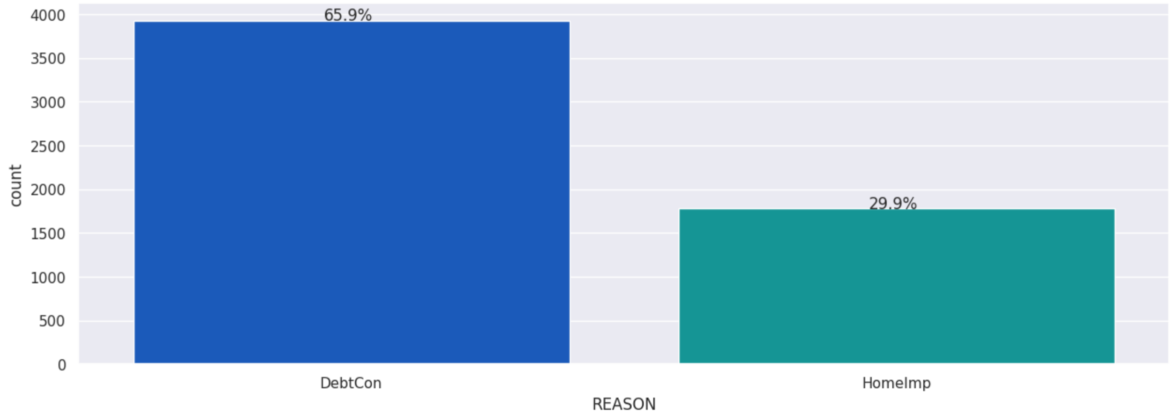
- LOAN and MORTDUE: Both exhibit positive skewness, meaning most values concentrate towards the lower ranges, with some outliers pulling the tail towards higher loan and mortgage amounts.
- VALUE: While fairly spread out, there's moderate skewness to the right, highlighting a cluster of mid-range property values with fewer high-value outliers.
- YOJ (Years on Job): Mild positive skewness, with most applicants having shorter tenures and fewer having extremely long durations.
- DEROG & DELINQ: These are heavily skewed to the right, as many applicants have clean credit records, but a small fraction shows notable derogatory reports and delinquencies.
- CLAGE: Credit line age leans slightly towards the positive side, indicating more mature credit lines but few stretching significantly.
- DEBTINC: A healthy skew to the right, reflecting that most applicants maintain manageable debt-to-income ratios, while outliers face financial strain.

Univariate Data Analysis – Categorical Data

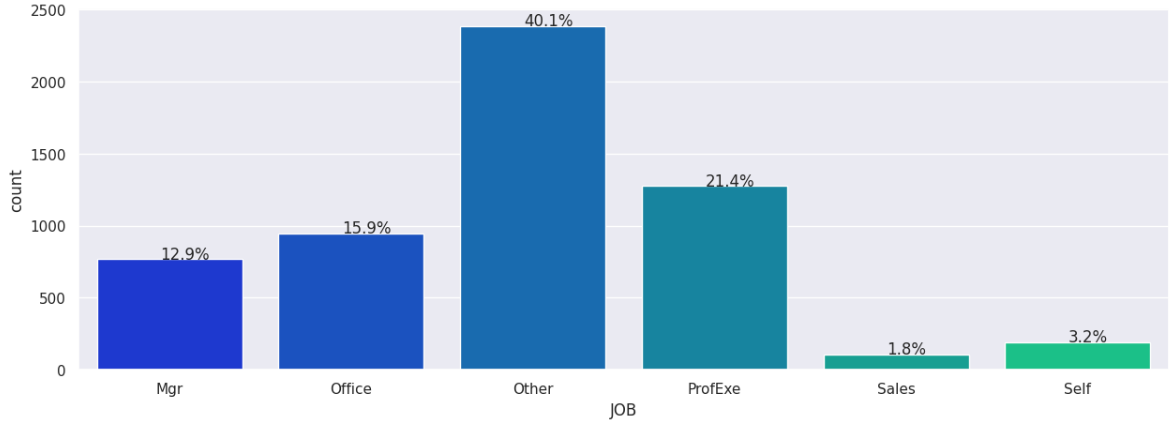
Barplot for DELINQ



Barplot for Loan Reason

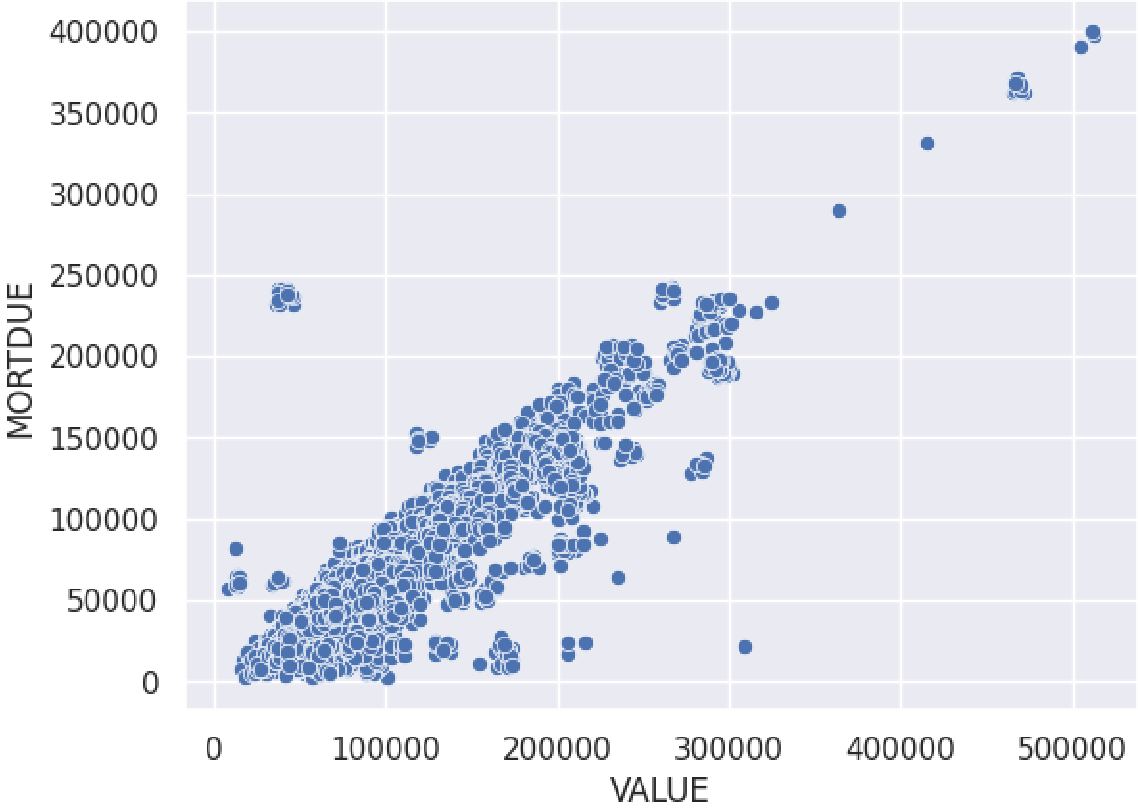


Barplot for Job Type

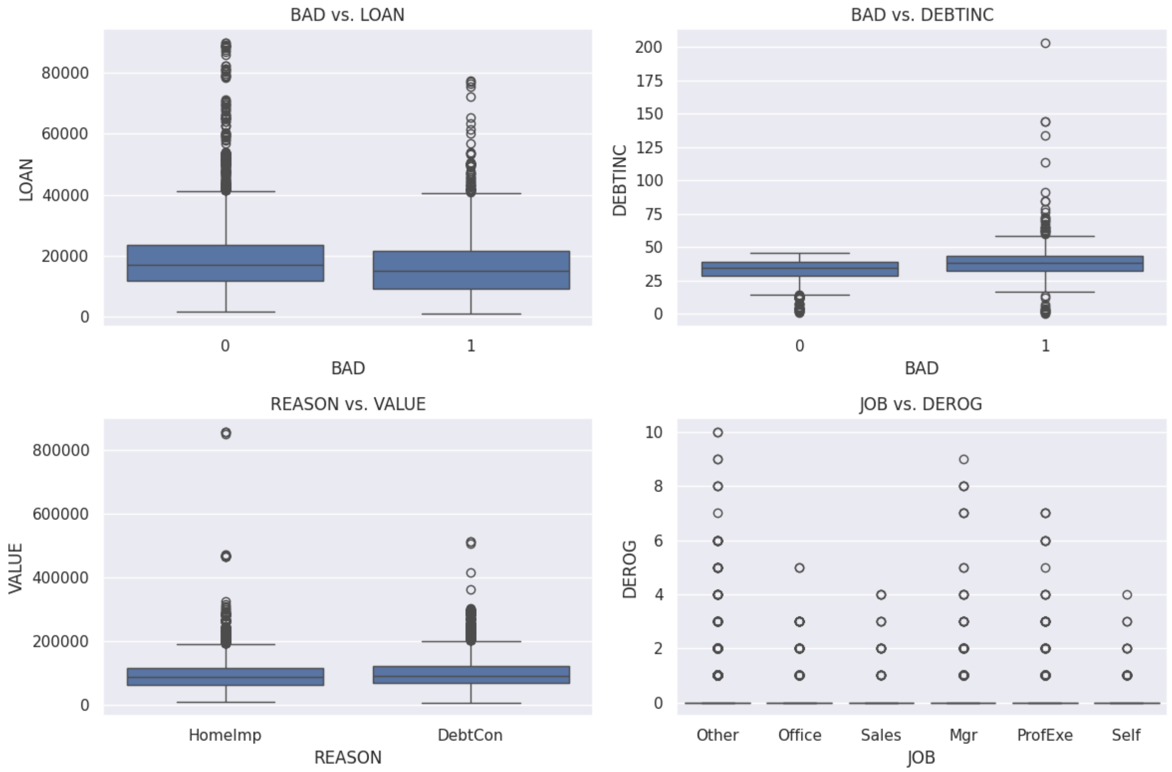


Bivariate Analysis

This scatterplot illustrates a positive linear relationship between VALUE (property value) and MORTDUE (mortgage due)



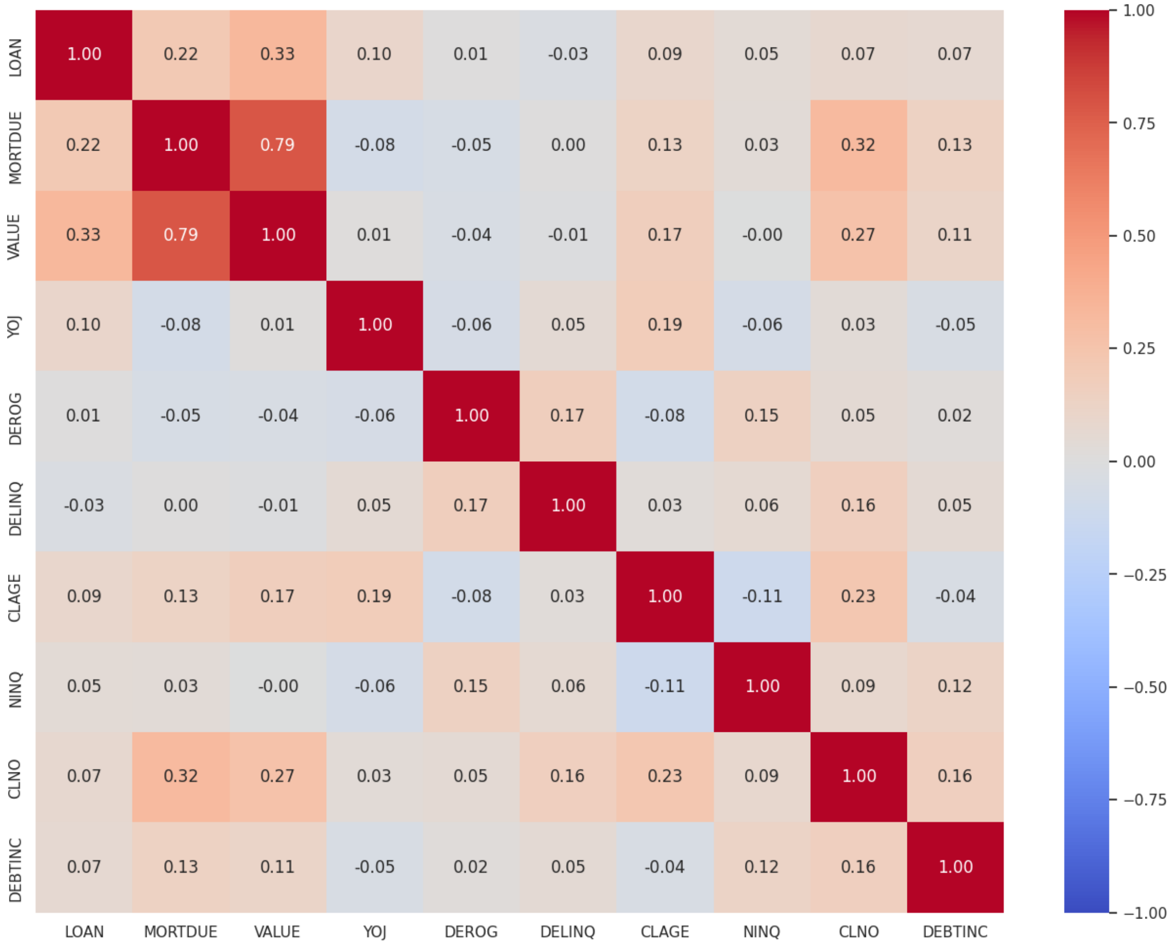
Bivariate Analysis: Continuous and Categorical Variables



- BAD vs. LOAN: Non-defaulters (BAD = 0) have slightly higher median loan amounts compared to defaulters. This could indicate that applicants with better financial standings are granted larger loans.
- BAD vs. DEBTINC (Debt-to-Income Ratio): Defaulters (BAD = 1) have significantly higher debt-to-income ratios, which may contribute to their financial strain.
- REASON vs. VALUE: Property values are higher for loans taken out for home improvement compared to debt consolidation. This suggests that home improvement loans are associated with higher-value assets.
- JOB vs. DEROG: Interestingly, derogatory reports appear consistent across job types, with no major outliers standing out.
- BAD and DEBTINC highlights a key factor contributing to loan defaults.

Multivariate Analysis

Correlation Heatmap



	LOAN	MORTDUE	VALUE	YOJ	DEROG	DELINQ	CLAGE	NINQ	CLNO	DEBTINC
LOAN	1.00	0.22	0.33	0.10	0.01	-0.03	0.09	0.05	0.07	0.07
MORTDUE	0.22	1.00	0.79	-0.08	-0.05	0.00	0.13	0.03	0.32	0.13
VALUE	0.33	0.79	1.00	0.01	-0.04	-0.01	0.17	-0.00	0.27	0.11
YOJ	0.10	-0.08	0.01	1.00	-0.06	0.05	0.19	-0.06	0.03	-0.05
DEROG	0.01	-0.05	-0.04	-0.06	1.00	0.17	-0.08	0.15	0.05	0.02
DELINQ	-0.03	0.00	-0.01	0.05	0.17	1.00	0.03	0.06	0.16	0.05
CLAGE	0.09	0.13	0.17	0.19	-0.08	0.03	1.00	-0.11	0.23	-0.04
NINQ	0.05	0.03	-0.00	-0.06	0.15	0.06	-0.11	1.00	0.09	0.12
CLNO	0.07	0.32	0.27	0.03	0.05	0.16	0.23	0.09	1.00	0.16
DEBTINC	0.07	0.13	0.11	-0.05	0.02	0.05	-0.04	0.12	0.16	1.00

Strong Positive Correlations:

- MORTDUE (Mortgage Due) and VALUE (Property Value) are closely linked, flowing together with a strong positive correlation. This suggests that higher property values often come with larger mortgages.

- LOAN and VALUE also show a solid connection, as larger loans align with higher-valued properties.

EDA Key Insights

- Loan Default Rate: Approximately 20% of loans defaulted or were severely delinquent, meaning there's a significant minority of applicants with repayment issues.

Categorical Variables:

- REASON: Most loans were for Debt Consolidation (about 68%), with Home Improvement following.
- JOB: Applicants' occupations vary, with 'Other' being the most frequent (42%), followed by Professional/Executive roles.

Numerical Variables:

- Variables like LOAN, VALUE, and MORTDUE show positive correlations, indicating larger loans for high-value properties and mortgages.
- DEBTINC (Debt-to-Income Ratio): Missing data is significant, highlighting challenges in assessing applicants' financial health.

Outlier and Outlier treatment

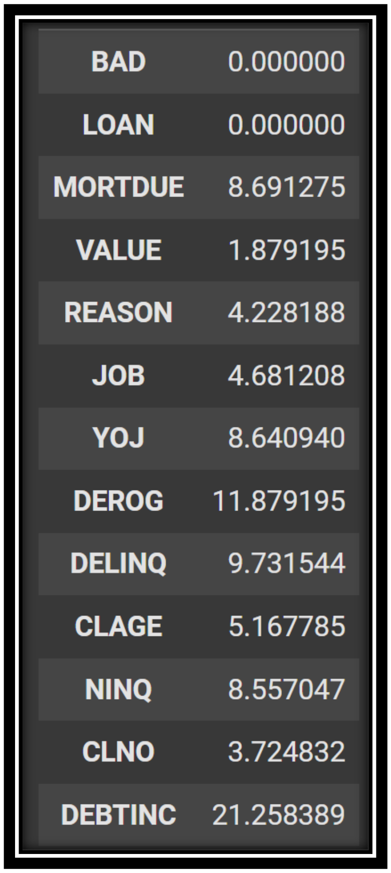
Outliers:

- LOAN: Some very high loan values compared to the majority.
- VALUE: A few extreme property values stand out.
- DEROG and DELINQ: Both have outliers indicating higher-than-usual derogatory or delinquent credit lines.

Outlier Treatment:

Outliers can be treated by capping/flooring (adjust extreme values to a fixed threshold) or potentially through transformations (logarithmic or other scaling methods to reduce their impact).

Missing Values



BAD	0.000000
LOAN	0.000000
MORTDUE	8.691275
VALUE	1.879195
REASON	4.228188
JOB	4.681208
YOJ	8.640940
DEROG	11.879195
DELINQ	9.731544
CLAGE	5.167785
NINQ	8.557047
CLNO	3.724832
DEBTINC	21.258389

Missing Value Treatment

Missing Values: Variables like YOJ and DEROG have missing data that could affect analysis. Imputation or binary flags may help mitigate this.

It may be essential to address missing values in the dataset, as missing values may impact model performance (and predictions) and columns such as DEBTINC and DEROG (21% and 11.8% respectively) could strongly influence the model's ability to predict defaults.

Potential strategies to treat missing values may include imputing missing numerical values with median or mean values to avoid distortion due to outliers and mode imputation for missing categorical variables. Another option may be to resort to flagging – via adding binary flags to indicate which rows have missing values.

However, caution must be exercised as on some occasions missing information itself may be predictive of loan default risk and this missing data may also become an important signal in and of itself. In such an instance, it may be beneficial to add binary flag features that can allow models to learn the relationship between missing information and default risk.

Building Models

Three models were considered for this exercise:

1. Logistic Regression
2. Decision Tree
3. Random Forest

As this is a binary classification exercise, certain models should be considered at the outset. In this analysis, we started with Logistic Regression. This technique offers a way to predict whether something belongs to one category or another, so it is a good starting point for our BAD target variable.

For the classification models, we will interpret results through several specific metrics – accuracy, confusion matrix and via classification report (precision, recall and F1-score).

As the process is often one of trial and error, we will also attempt to progress to different models, such as Decision Trees and more complex models, such as Random Forest.
We will also attempt to further enhance the accuracy of the models via hyperparameter tuning such as via methods such as GridSearchCV.

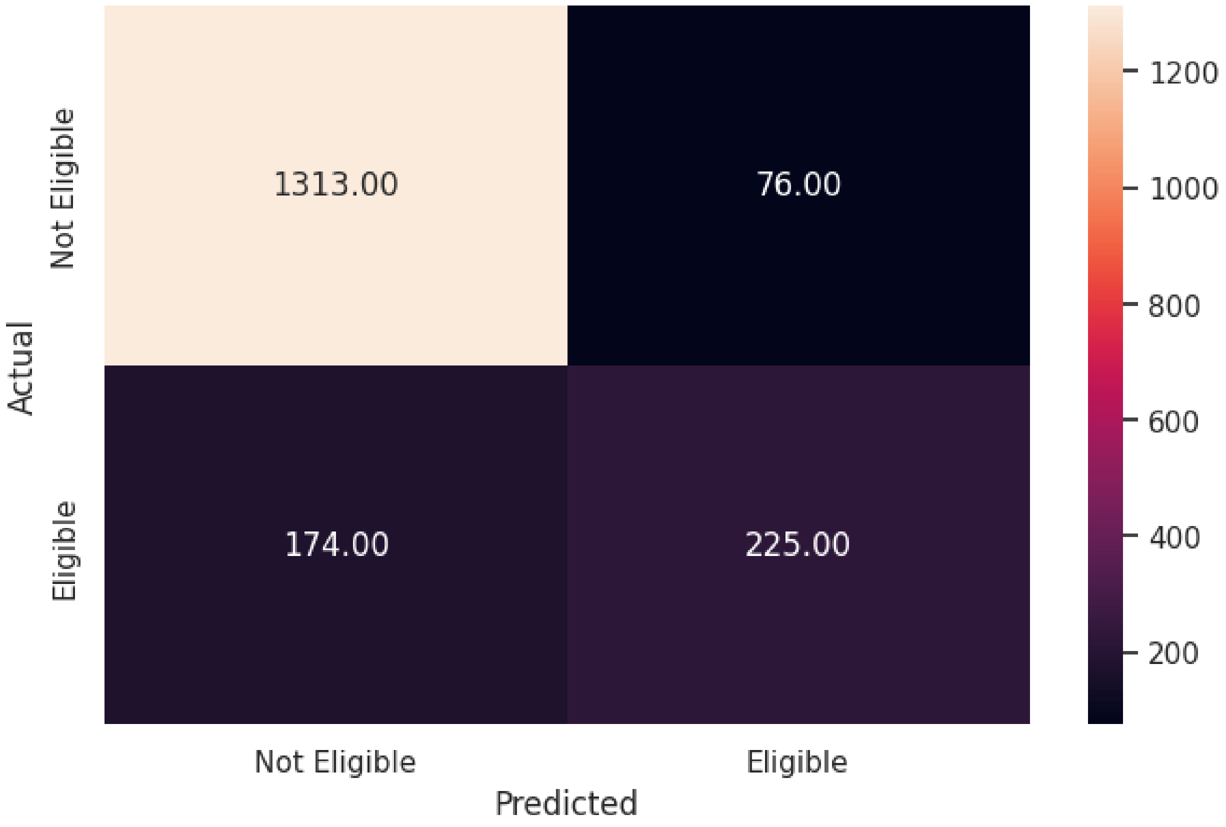
Comparison of various techniques and their relative performance based on chosen Metric (Measure of success)

Key Metrics utilized.

Metric	Description
Train Accuracy	Measures how well the model fits the training data.
Test Accuracy	Reflects model performance on unseen data, highlighting its ability to generalize.
Train Recall	Focuses on true positives in the training set, emphasizing the model's sensitivity.
Test Recall	Like Train Recall but tested on new data—essential for catching actual defaults.
Train Precision	Shows the accuracy of predictions labeled as positive during training
Test Precision	Applies precision to unseen data, ensuring the model stays precise in real-world scenarios.

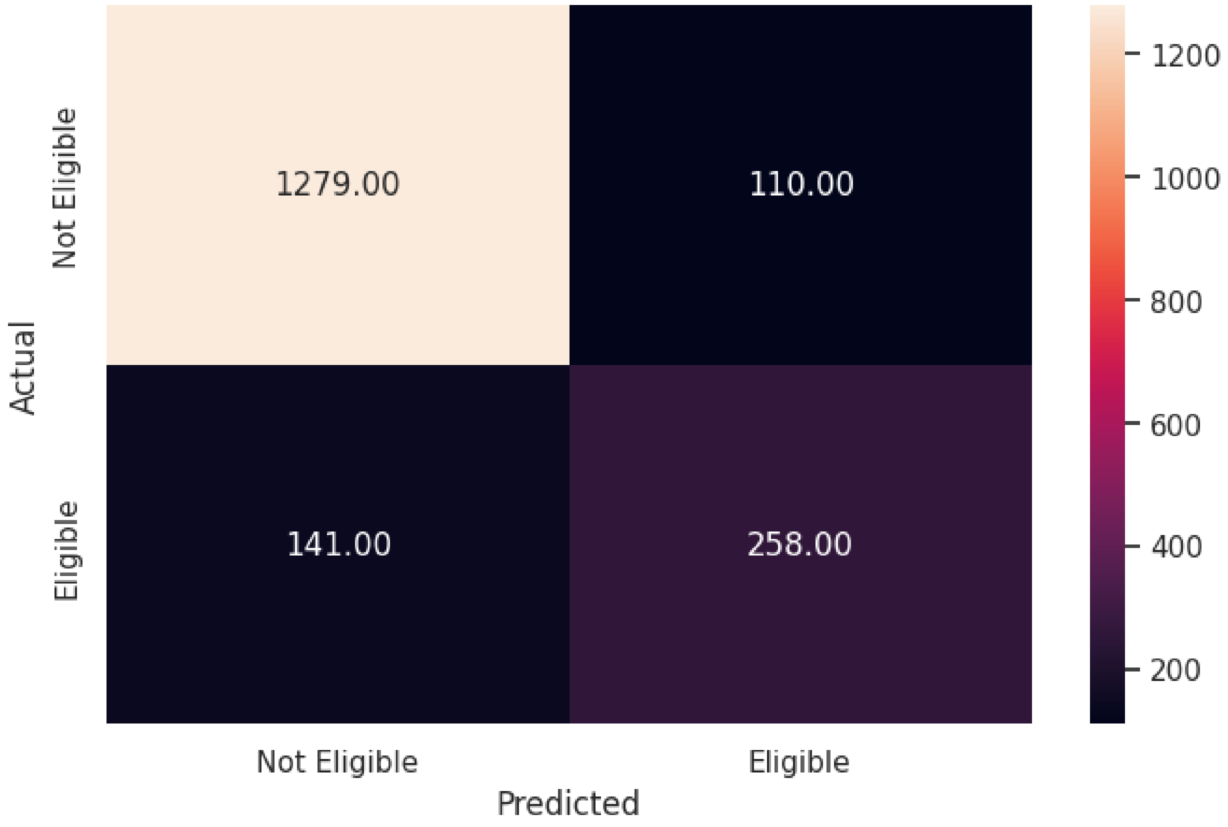
Logistic Regression for Test Dataset

	precision	recall	f1-score	support
0	0.88	0.95	0.91	1389
1	0.75	0.56	0.64	399
accuracy			0.86	1788
macro avg	0.82	0.75	0.78	1788
weighted avg	0.85	0.86	0.85	1788



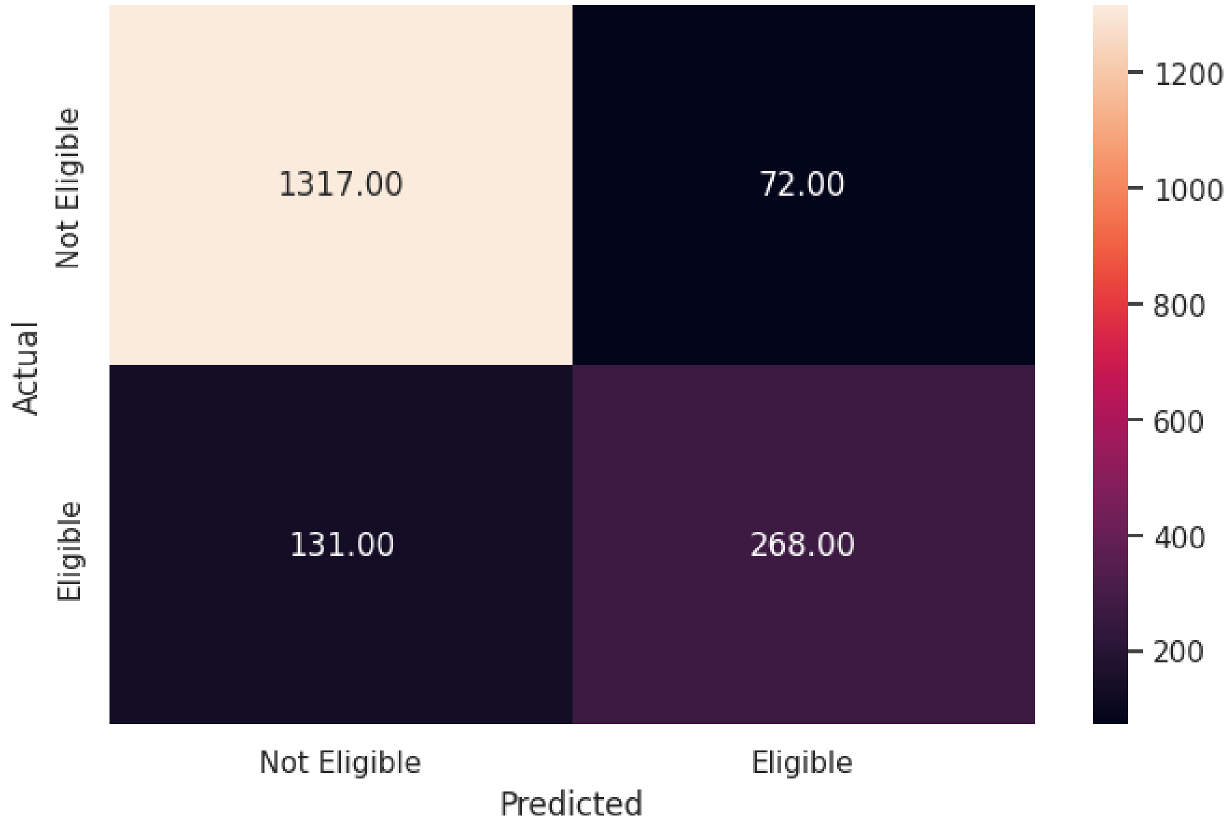
Decision Tree Test Data Set

Class	Precision	Recall	F1-Score	Support
0	0.90	0.92	0.91	1389
1	0.70	0.65	0.67	399
Accuracy			0.86	1788
Macro Avg	0.80	0.78	0.79	1788
Weighted Avg	0.86	0.86	0.86	1788

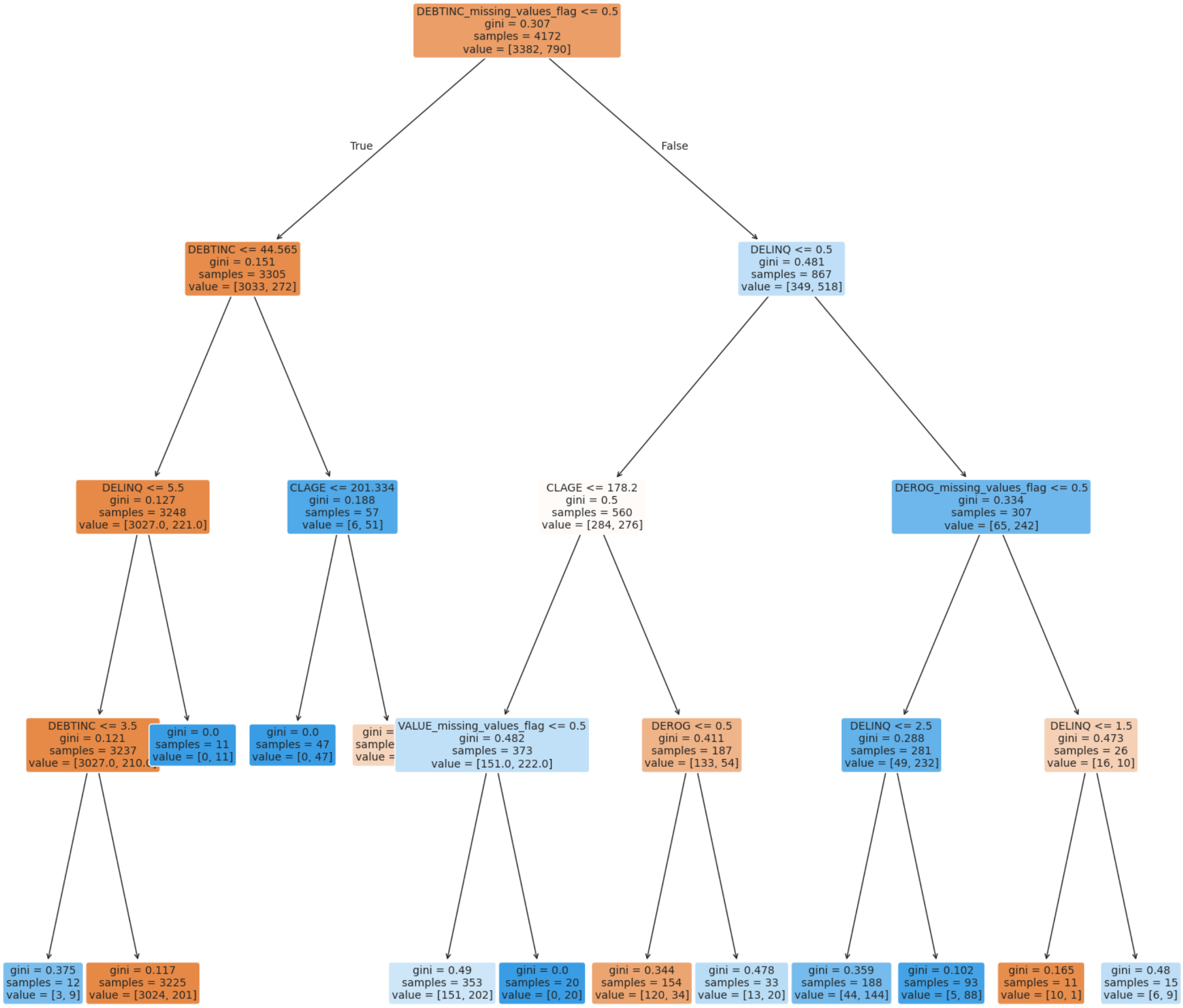


Post use of GridSearchCV (Test Data Set)

	precision	recall	f1-score	support
0	0.91	0.95	0.93	1389
1	0.79	0.67	0.73	399
accuracy			0.89	1788
macro avg	0.85	0.81	0.83	1788
weighted avg	0.88	0.89	0.88	1788



Visualization of the Decision Tree

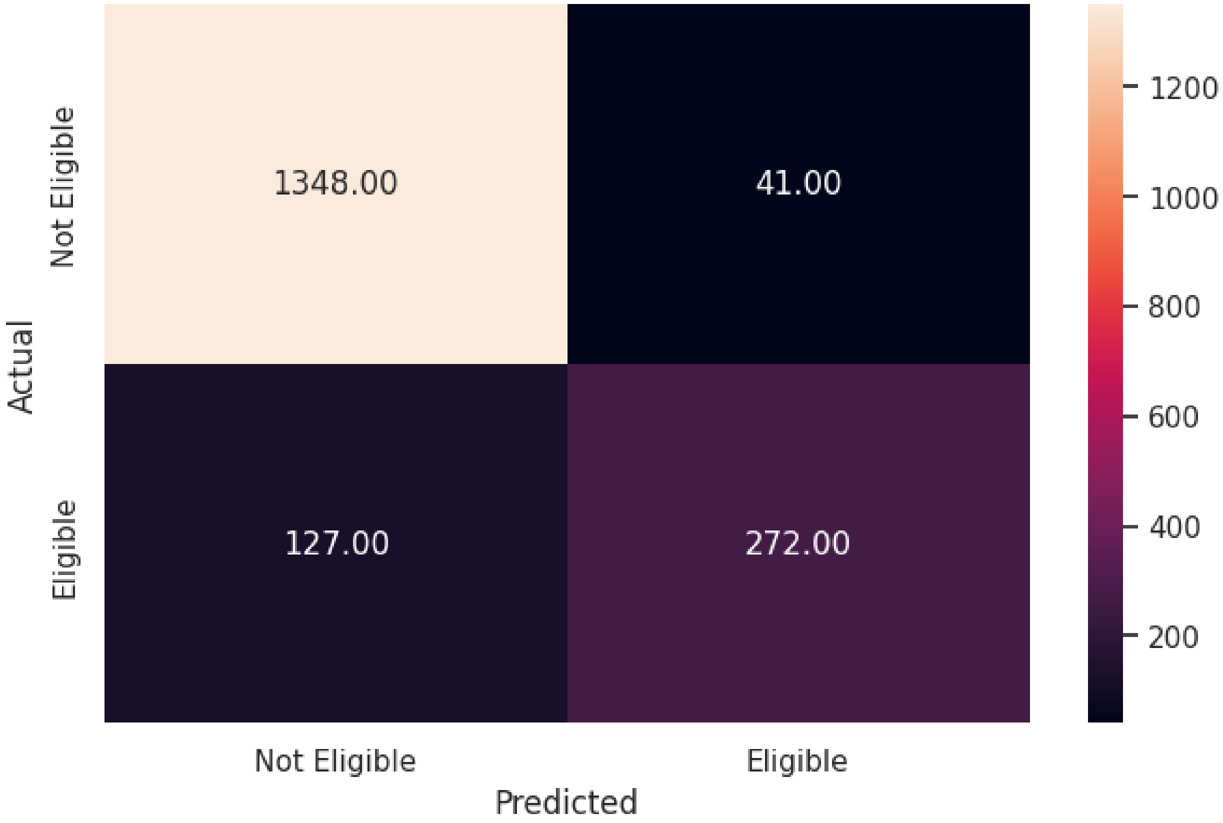


Key Insights:

- The Decision tree shows that low debt-to-income ratios (e.g., DEBTINC) correlate with higher repayment likelihood.

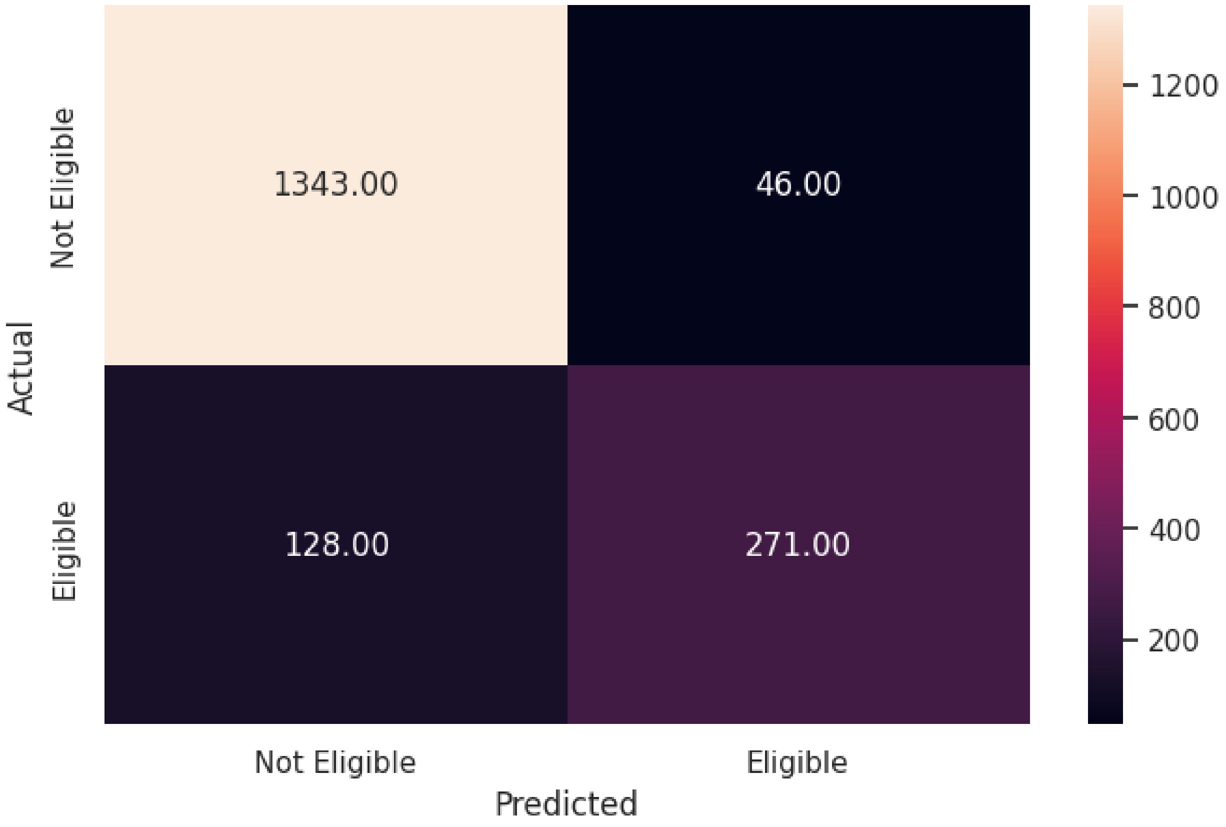
Random Forest Classifier

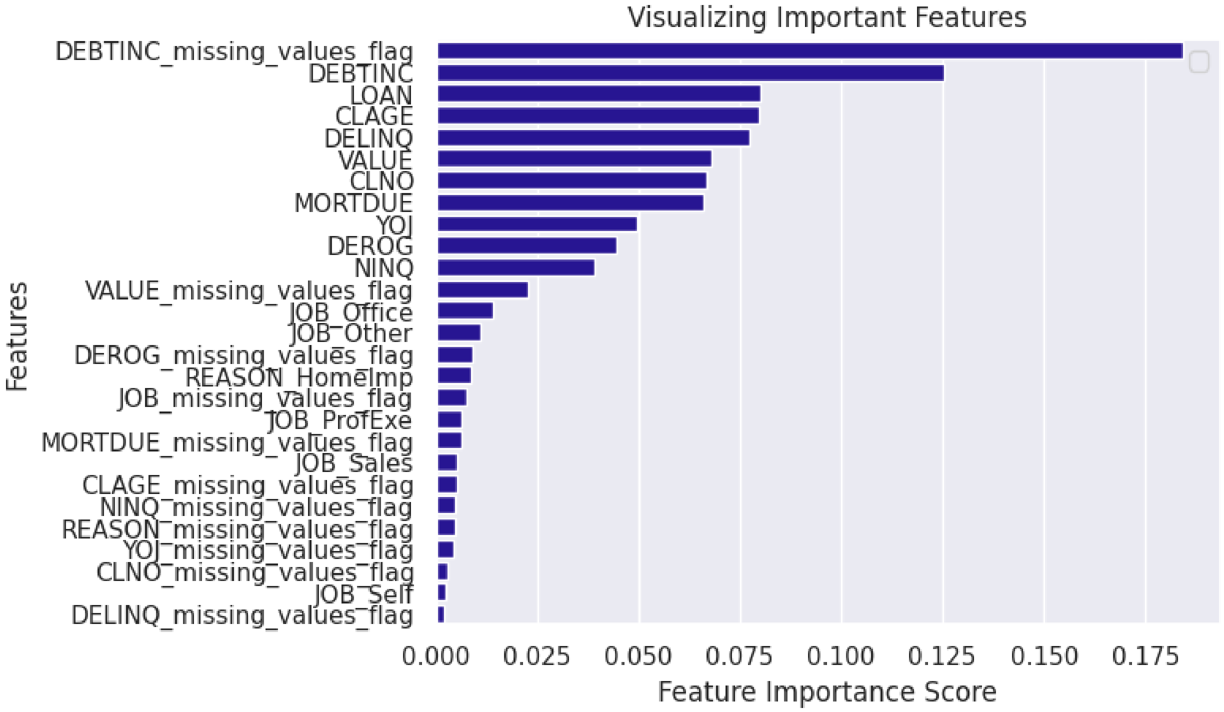
Class	Precision	Recall	F1-Score	Support
0	0.91	0.97	0.94	1389
1	0.87	0.68	0.76	399
Accuracy			0.91	1788
Macro Avg	0.89	0.83	0.85	1788
Weighted Avg	0.90	0.91	0.90	1788



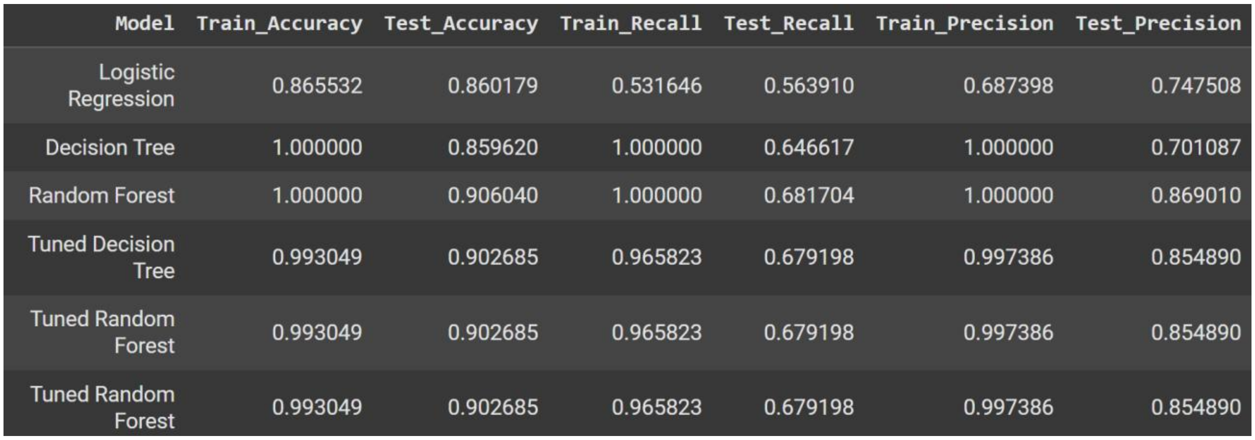
Tuned Random Forest (Test Data Set)

Label	Precision	Recall	F1-Score	Support
0	0.91	0.97	0.94	1389
1	0.85	0.68	0.76	399
accuracy			0.90	1788
macro avg	0.88	0.82	0.85	1788
weighted avg	0.90	0.90	0.90	1788




Visualizing Important Features
DEBTINC_missing_values_flag
DEBTINC
LOAN
CLAGE
DELINQ
VALUE
CLNO
MORTDUE
YOJ
DEROG
NINQ
VALUE_missing_values_flag
JOB_Office
JOB_Other
DEROG_missing_values_flag
REASON_HomeImp
JOB_missing_values_flag
JOB_ProfExe
MORTDUE_missing_values_flag
JOB_Sales
CLAGE_missing_values_flag
NINQ_missing_values_flag
REASON_missing_values_flag
YOJ_missing_values_flag
CLNO_missing_values_flag
JOB_Self
DELINQ_missing_values_flag
Features
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
Feature Importance Score

Model Evaluation Matrix



Model	Train_Accuracy	Test_Accuracy	Train_Recall	Test_Recall	Train_Precision	Test_Precision
Logistic Regression	0.865532	0.860179	0.531646	0.563910	0.687398	0.747508
Decision Tree	1.000000	0.859620	1.000000	0.646617	1.000000	0.701087
Random Forest	1.000000	0.906040	1.000000	0.681704	1.000000	0.869010
Tuned Decision Tree	0.993049	0.902685	0.965823	0.679198	0.997386	0.854890
Tuned Random Forest	0.993049	0.902685	0.965823	0.679198	0.997386	0.854890
Tuned Random Forest	0.993049	0.902685	0.965823	0.679198	0.997386	0.854890

Final Solution Design

Proposal for the final solution design

The top-performing models for loan default prediction are the Tuned Decision Tree and Tuned Random Forest, as they both achieve high performing metrics.

Accuracy: Test accuracy is 0.902685, ensuring consistent and reliable predictions.
Recall: Captures Loan defaults well, with a Test Recall of 0.679198.
Precision: Maintains a strong Test Precision of 0.854890, minimizing false alarms.

The best model for predicting loan defaults appears to be the **Tuned Random Forest**, as it strikes the best balance between accuracy, recall, and precision. It maintains high Test Precision (0.848875) and Test Accuracy (0.898210), meaning it predicts loan defaults more reliably without overfitting. Its recall is equally strong, capturing most defaults.

Modeling Considerations:

- The primary business objective is to maximize recall to minimize the number of Unpaid Loans (Class 1).
- Balancing the 80:20 split of non-defaults vs. defaults is vital due to class imbalance.
- Noteworthy features like DEBTINC, DEROG, and VALUE could heavily influence default prediction.

- To further refine the model, hyperparameter tuning or feature engineering could be explored.

Model Interpretability and trade-offs (in terms of interpretability and accuracy)

- A logistic regression model is very interpretable – because you have coefficients of each specific factor.
- A Decision Tree is very interpretable – you can see the full “story” of how a model makes a certain prediction.
- A Random Forest is not as interpretable – because it is an ensemble of many decision trees – making it more difficult to trace why a certain prediction was made.