

Crescent Bank Project

By: Daniel Fox, Nancy Kohut, Amanda Barry,
Carter Mitchell, and Andrew Possi



Agenda

1. Company background
2. Project background
3. Dataset Overview
4. Data cleaning, preparation, and exploratory data analysis (EDA)
5. Correlation Matrix
6. Partial Dependence Plots
7. Feature Importance
8. Initial Models
9. Model Evaluation Metrics
10. Conclusion

Crescent Bank Background

- Founded in New Orleans in 1991
- Currently nationwide with over 450 employees
- \$1,121,005,000 in assets
- Business mainly consists of:
 - Auto loans
 - Local banking services
- Sourced from over 6,000 car dealerships across 34 states
- Strengthen relationships to expand market share
- Looking to find new opportunities for partnerships



Project Background

- Largely subprime loans
- Looking to predict which customers are likely to default in the next 180 days
 - Reduce Losses
 - Improve Profitability
- Dataset contains information on loans that are 30-59 days past due on payments
- Understanding what factors can predict a default before it occurs
- More efficient allocation of resources to prevent defaults
- Reducing loan loss rates while maintaining customer satisfaction
- Models in Python



Dataset Overview



| Column Name | Description |
|------------------------|---------------------------------------|
| BUS_DATE | Business Date |
| ACCOUNT | Account Number |
| DPD | Days Past Due |
| NUM_PMTS_LIFE | Number of Payments - Lifetime |
| NUM_PMTS_6M | Number of Payments - Last 6 months |
| LAST_PMT_DATE | Last Payment Date |
| LAST_RPC_DATE | Last Right Person Contact Date |
| LAST_OB_RPC | Last OutBound RPC Date |
| RPC_COUNT_6M | Number of RPCs in the last 6 months |
| NUM_EXTENSIONS | Number of Extensions - Lifetime |
| LAST_EXTENSION_DATE | Last Extension Date |
| AVG_PMT_DPD | Avg DPD when a payment is made |
| NUM_NSF_CHECKS | Number of Not Sufficient Funds checks |
| NSF_PMTS_PERC | NSF count/ PMT Count |
| PAST_REPO | Past Repossession |
| DARv3 | Dollars at Risk Score |
| PROMISE_KEPT_RATE | Ratio of Promises Kept |
| LAST_PROMISE_DATE | Last Promise Date |
| LAST_PROMISE_KEPT_DATE | Last Promise Kept Date |
| MOB | Month on Book |
| AMOUNT_FINANCED | Amount Financed |
| PRINCIPAL_BALANCE | Principal Balance |
| INTEREST_BALANCE | Interest Balance |
| TARGET_VAR | Target Variable |

Data Cleaning, Preparation, EDA

- Initial Data Assessment
 - Dataset contained over 75,000 records
 - Identified issues with missing values and unbalanced data
- Data Cleaning Steps
 - Corrected date timestamps
- Variable Creation
 - Created new variables to use in model
 - Payment ratio, time-based, financial ratio variables
- Data Transformation
- Final Dataset Structure
 - Target Variable based on 180 days delinquency
 - Split 80% training 20% test

Exploratory Data Analysis

- Distribution analysis of key variables
- Correlation analysis between features
- Partial dependence analysis to understand feature relationships with default probability

Feature Importance

Top Predictive Features:

1. MOB (Months on Book) - 36.7%
2. Interest/financed ratio - 35.1%
3. Days since last payment - 8.7%
4. Number of payments in life of loan - 5.8%
5. Past repo indicator - 3.2%
6. Principal balance - 2.4%
7. Days since last extension - 2.0%
8. Days since last RPC - 1.7%
9. DPD - 1.6%
10. Average payment DPD - 1.6%

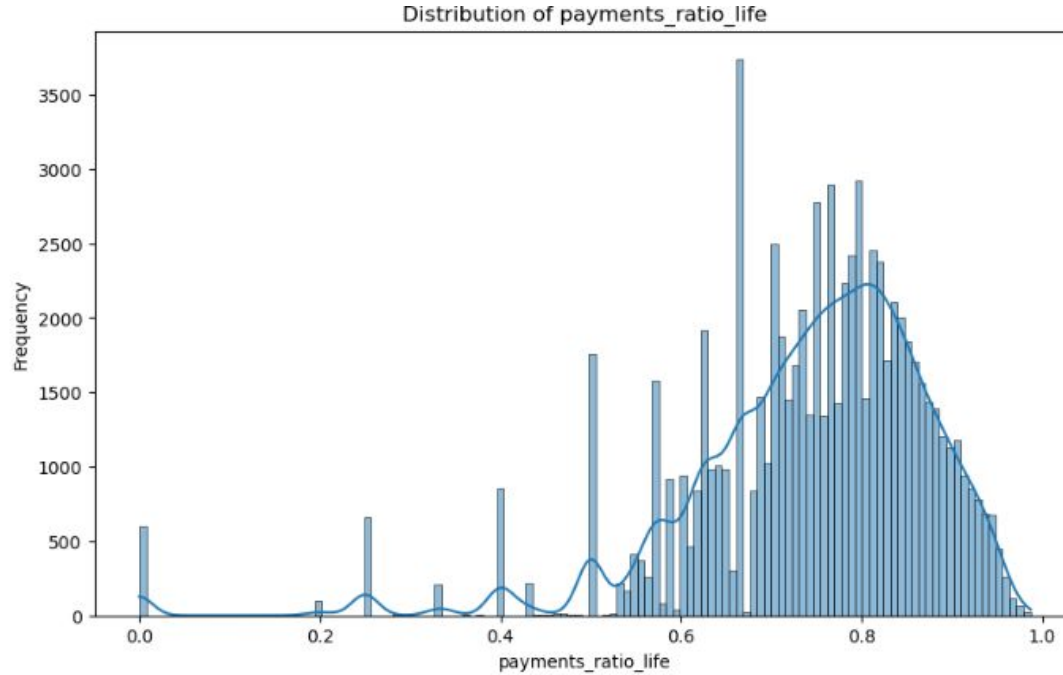


Data Dictionary for Selected Variables

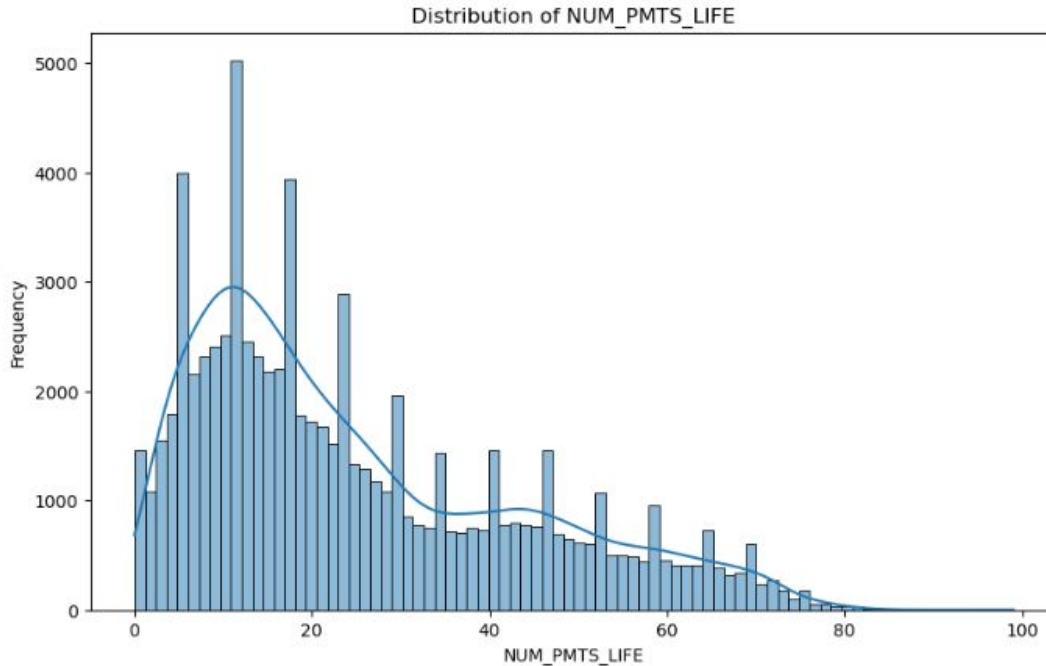


| Variable Name | Definition |
|-------------------------|--|
| payments_ratio_life | Ratio of payments made to payments due over the lifetime of the loan |
| interest/financed | Current Interest balance divided by total amount financed |
| days_since_last_payment | Days since last payment was received |
| PAST_REPO_B | Binary indicator if the customer has had a car repossessed in the past |
| NUM_PMTS_LIFE | Total number of payments the customer has made over the life of the loan |
| AVG_PMT_DPD | Average days past the due date for a customer's payments |

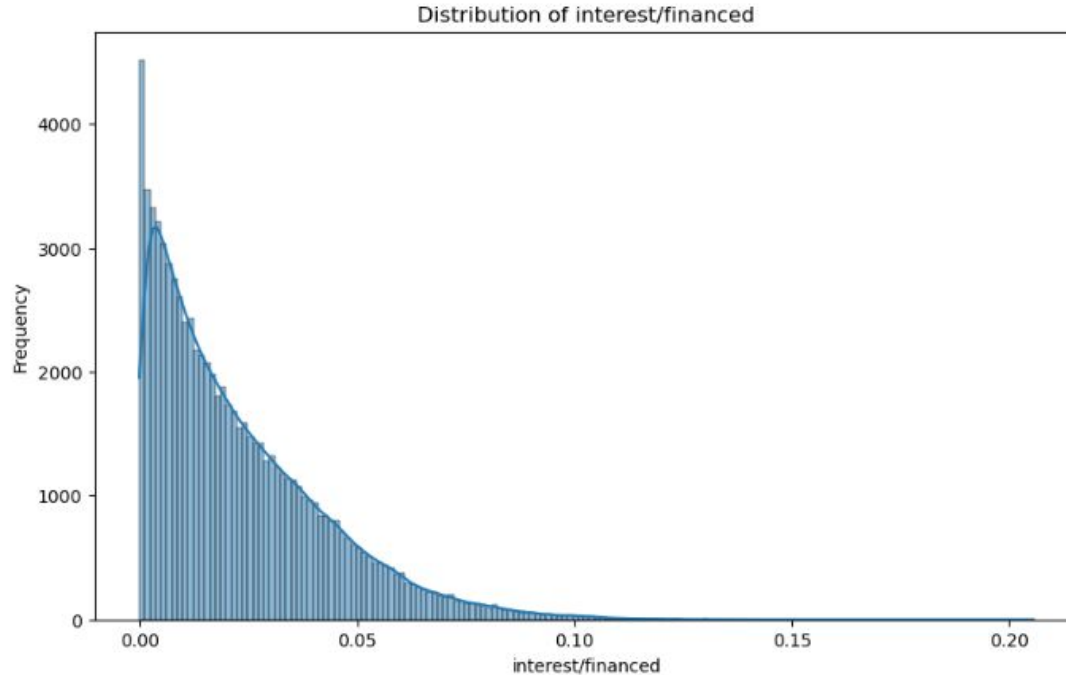
Exploratory Data Analysis: Payments Ratio Life



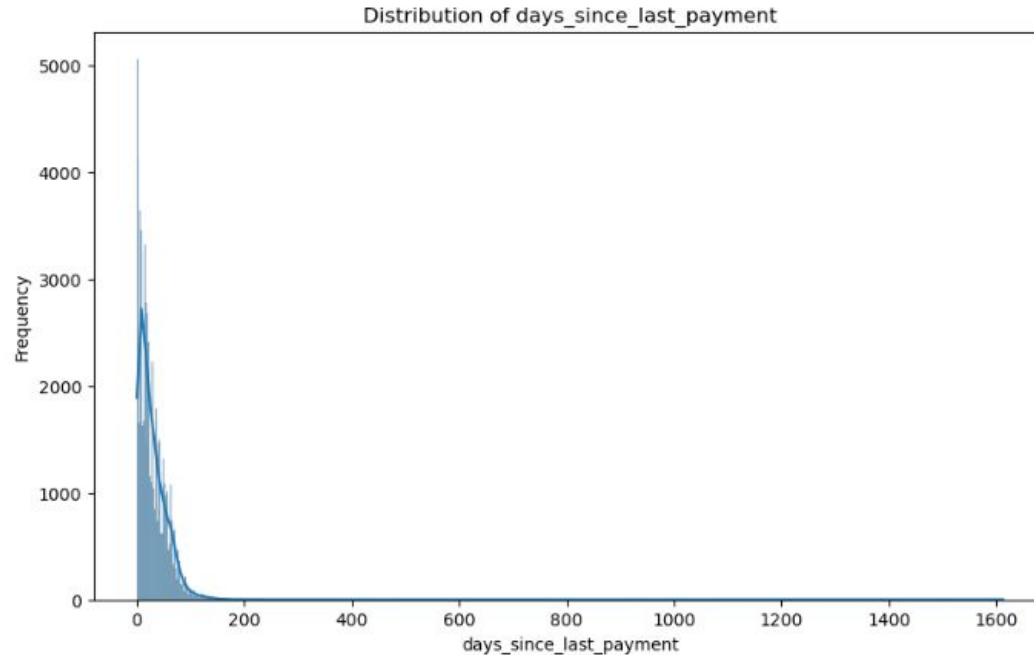
Exploratory Data Analysis: Num Pmts Life



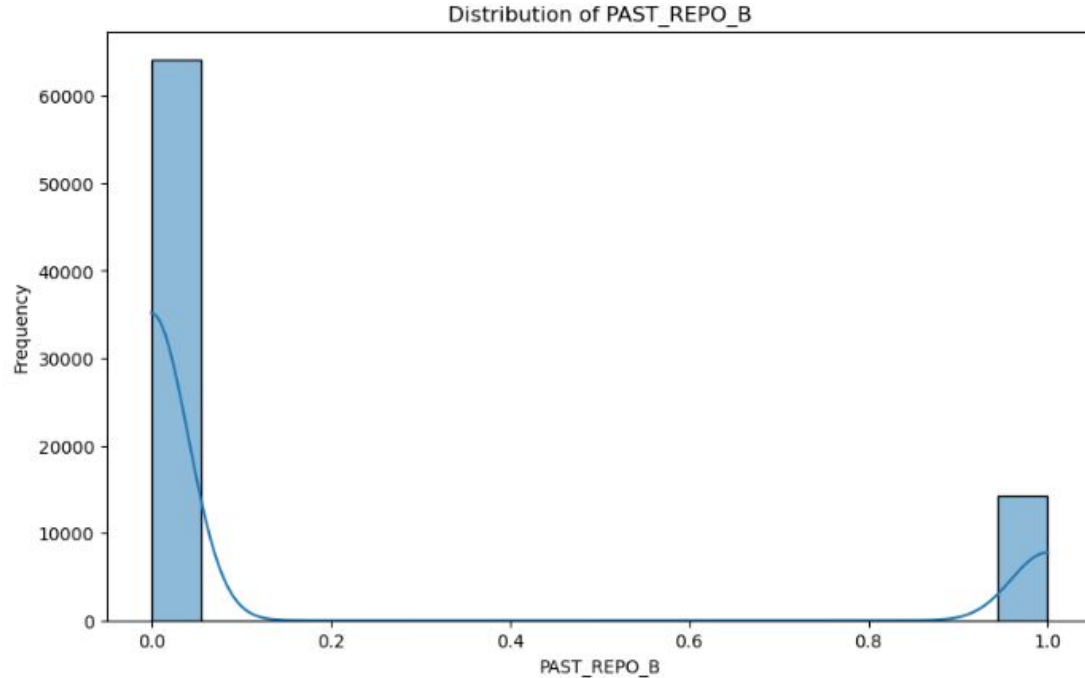
Exploratory Data Analysis: interest/financed



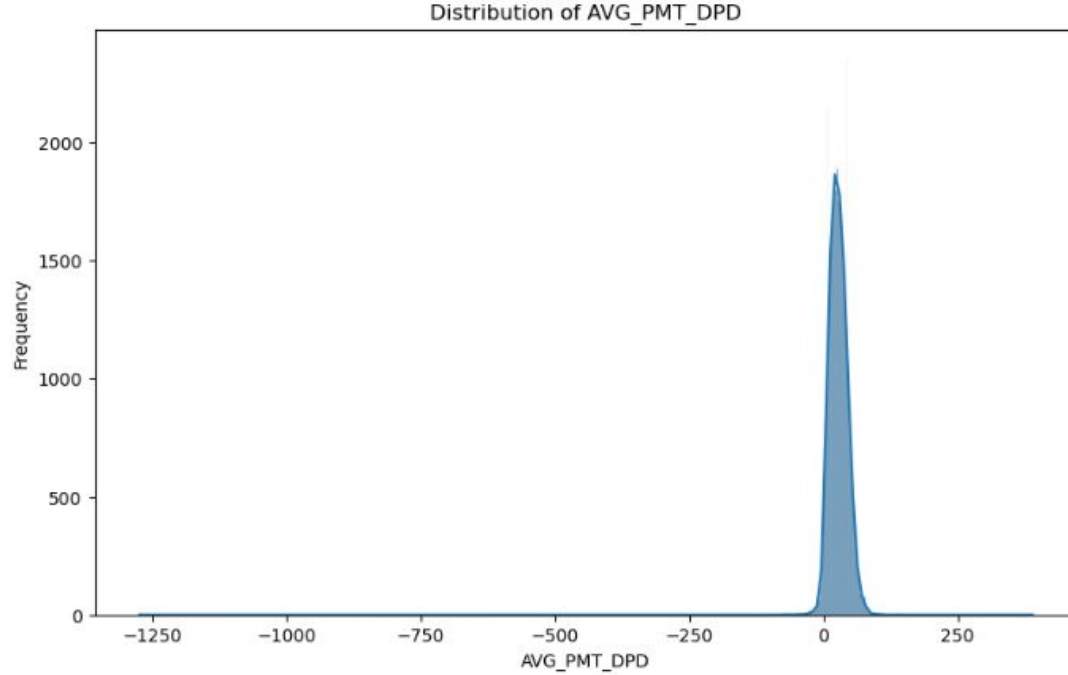
Exploratory Data Analysis: days since last payment



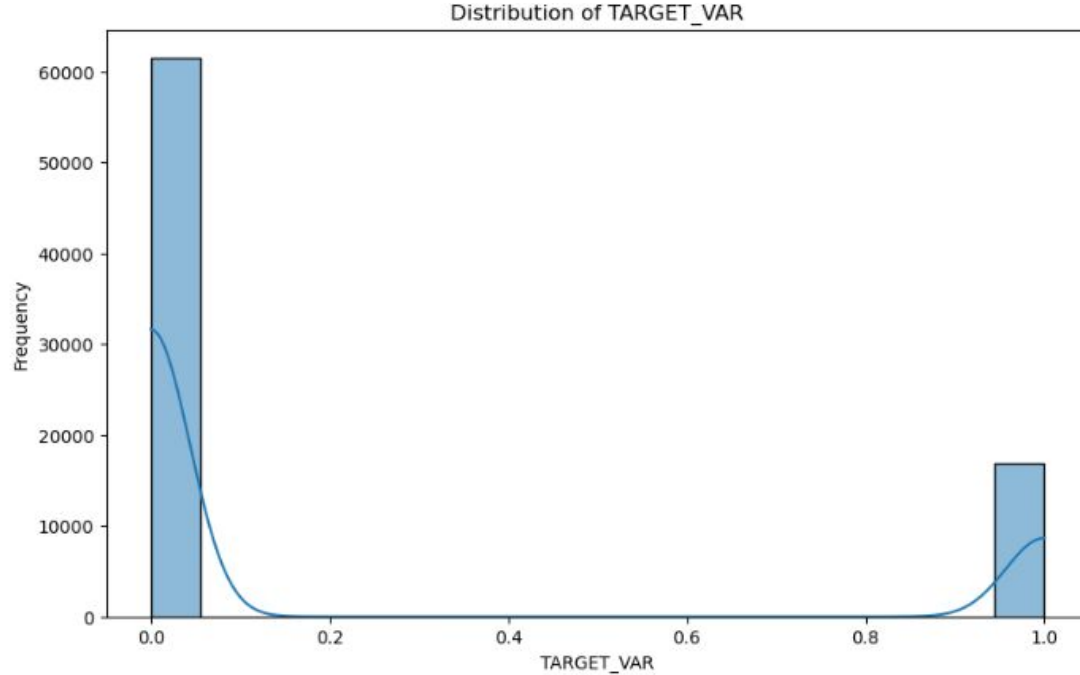
Exploratory Data Analysis: Past Repo B



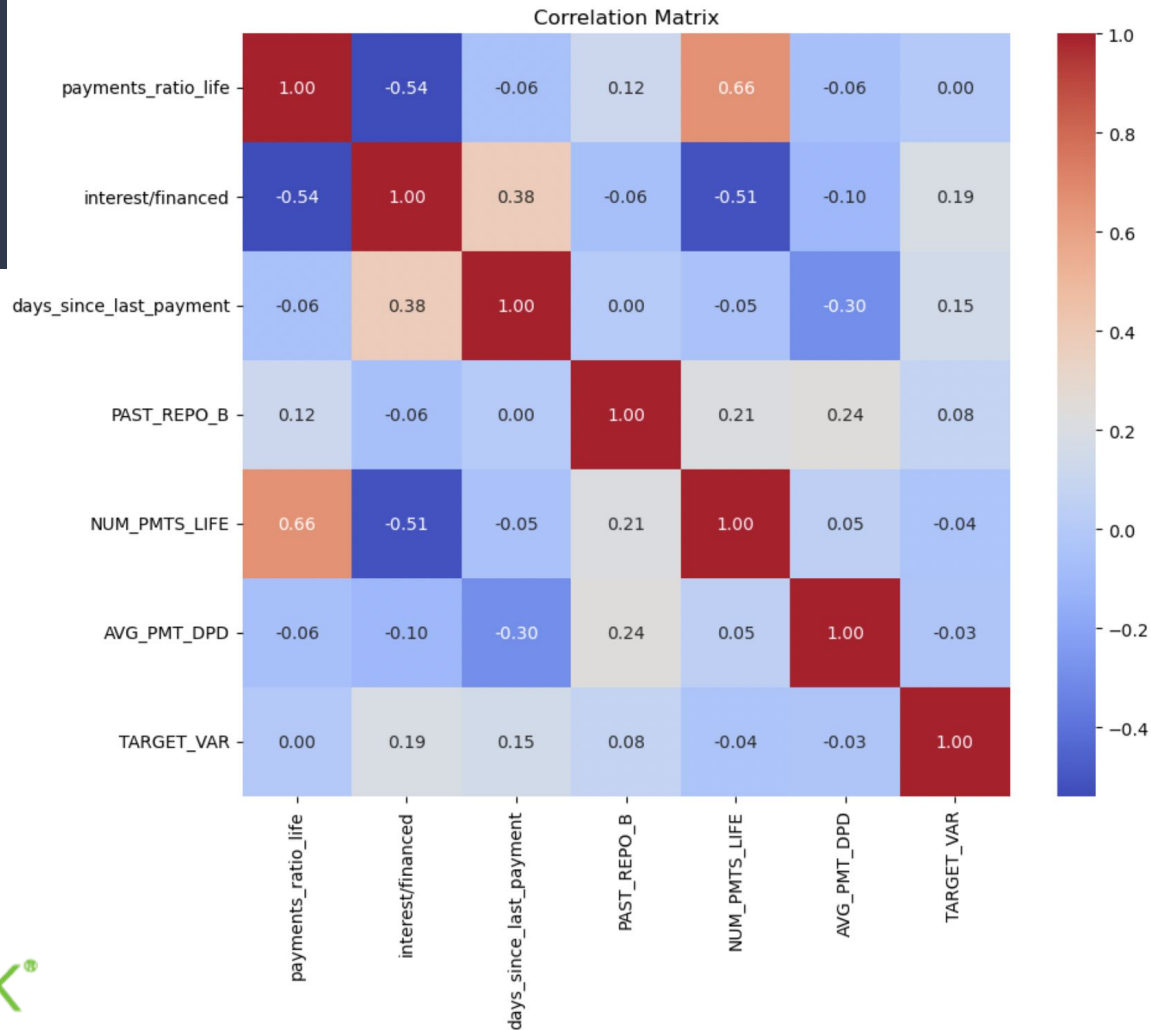
Exploratory Data Analysis: Avg Pmt DPD



Exploratory Data Analysis: Target Variable



Correlation Matrix

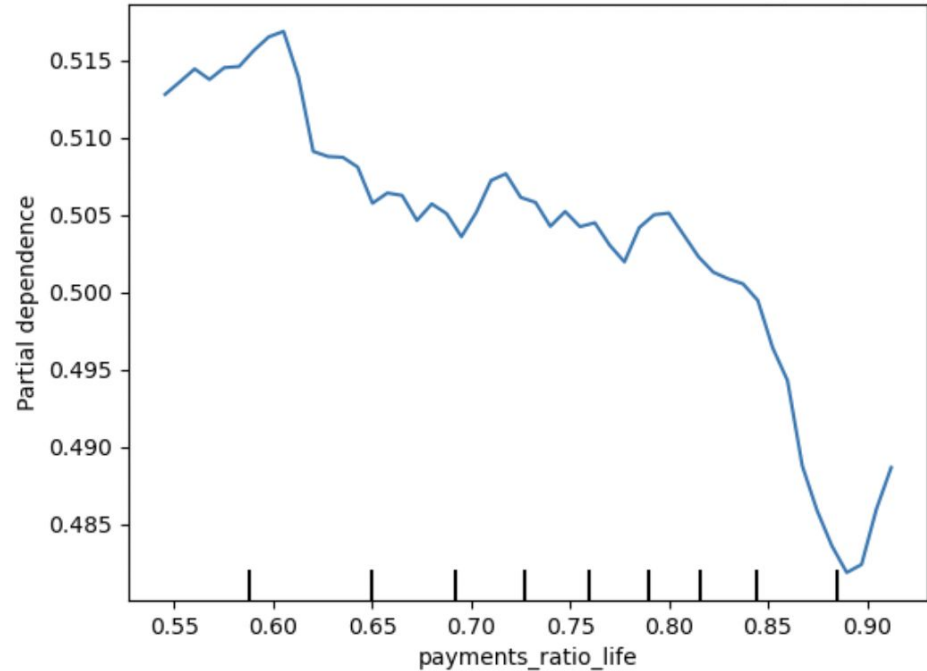


Partial Dependence Plots

Payments_ratio_life

Ratio of payments made to payments due over the lifetime of the loan

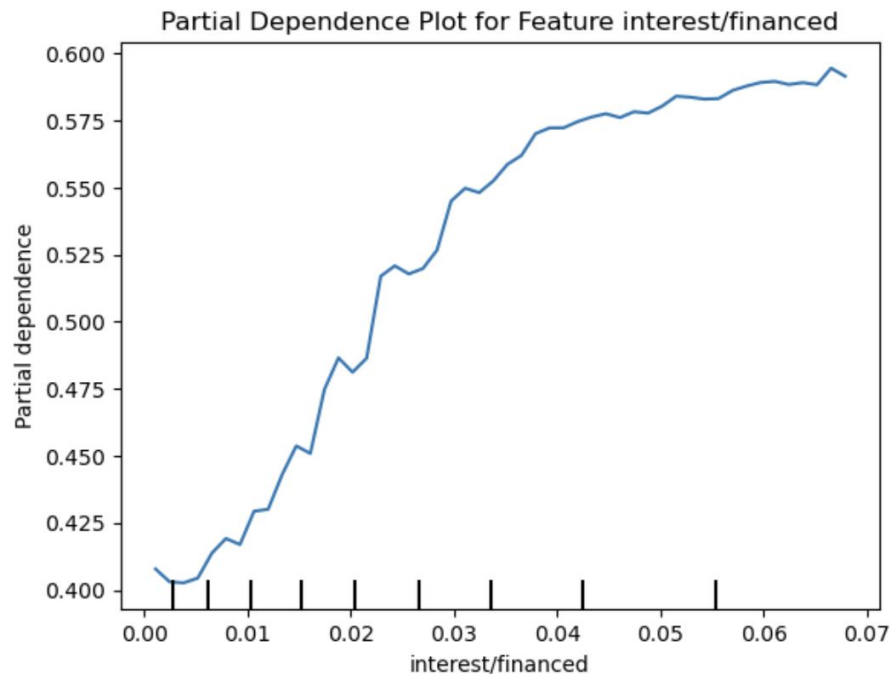
Partial Dependence Plot for Feature payments_ratio_life



Partial Dependence Plots

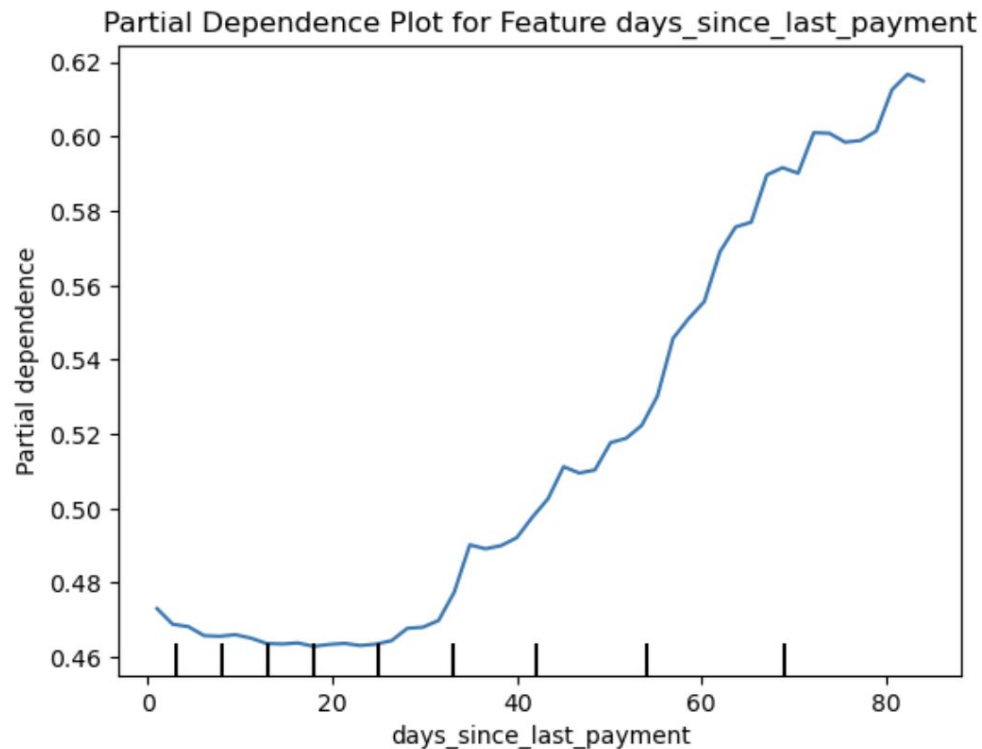
interest/financed

Current Interest balance divided by total amount financed



Partial Dependence Plots

Days_since_last_payment
Days since last payment was received

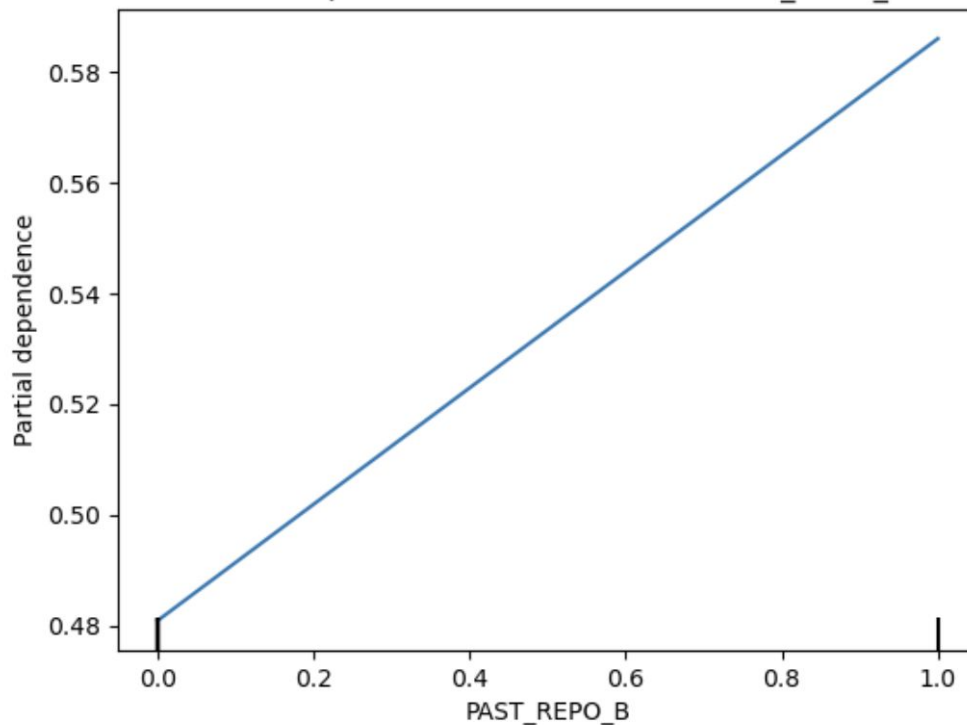


Partial Dependence Plots

PAST_REPO_B

Binary indicator if the customer has had a car repossessed in the past or not

Partial Dependence Plot for Feature PAST_REPO_B

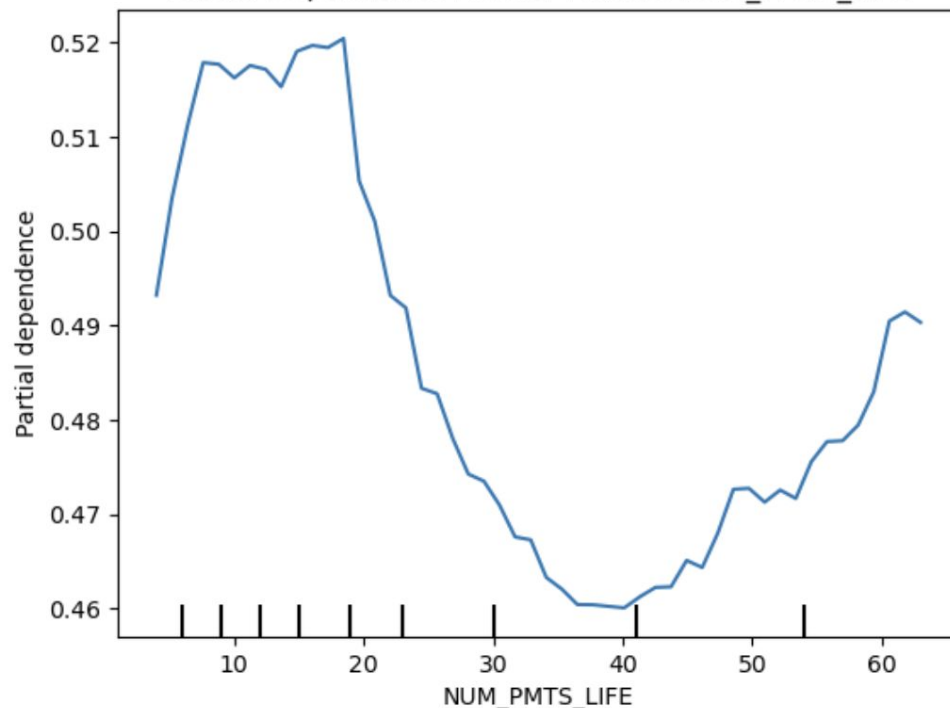


Partial Dependence Plots

NUM_PMTS_LIFE

Total number of payments the customer has made over the life of the loan

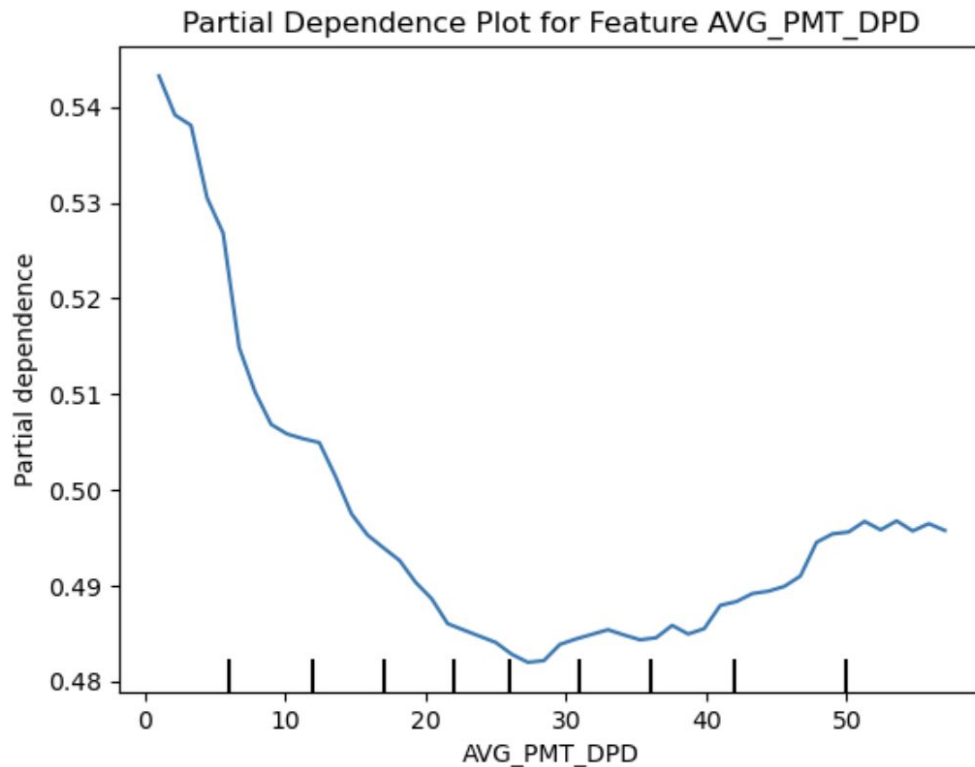
Partial Dependence Plot for Feature NUM_PMTS_LIFE



Partial Dependence Plots

AVG_PMT_DPD

Average days past the due date for a customer's payments



Final Models

Random Forest Model

- Cross Validation Accuracy - 64.24%
- Test Accuracy - 64.94%
- Cross Validation AUC Score - 0.7009
- Test AUC Score - 0.718
- Hyperparameter tuning revealed an ideal depth of 10
- Feature importance consistent with decision tree model

Final Models

XGBoost Model

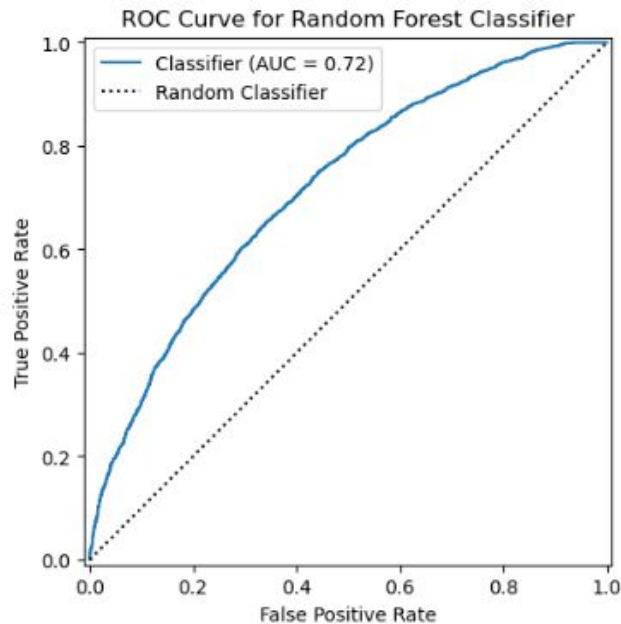
- Cross Validation Accuracy - 64.47%
- Test Accuracy - 65.39%
- Cross Validation AUC Score - 0.7001
- Test AUC Score - 0.7096
- Hyperparameter tuning revealed an ideal depth of 6

Final Models

Decision Tree Model

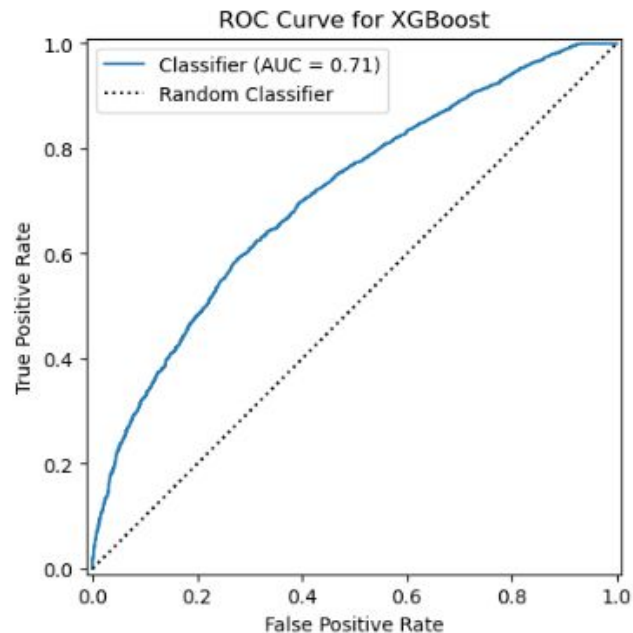
- Cross Validation Accuracy - 63.35%
- Test Accuracy - 63.47%
- Cross Validation AUC Score - 0.6818
- Test AUC Score - 0.6872
- Balanced class approach used to address data imbalance

AUC Graph: Random Forest (Best)



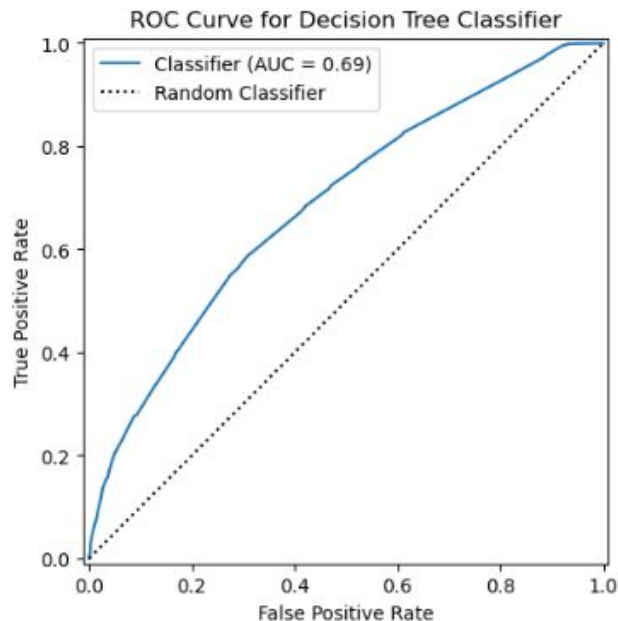
The testing ROC AUC is: 0.718

AUC Graph: XGBoost



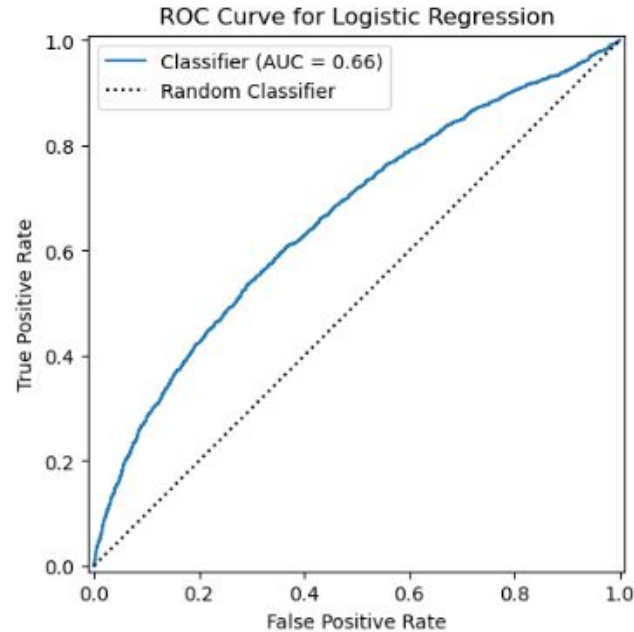
The testing ROC AUC is: 0.7096

AUC Graph: Decision Tree



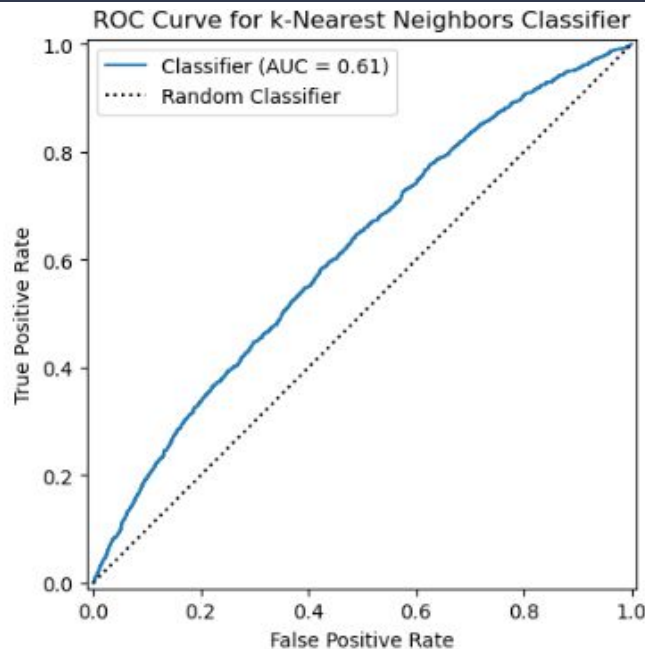
The testing ROC AUC is: 0.6872

AUC Graph: Logistic Regression



The testing ROC AUC is: 0.6612

AUC Graph: k-Nearest Neighbors (worst)



The testing ROC AUC is: 0.6131

Conclusion

- Developed predictive models for loan default within 6 months
 - Best Metric: AUC which measures the model's performance across all possible classification thresholds
 - Best Accuracy (AUC) score was from the Random Forest model of 71.8%
- Key predictor identified
 - Interest/financed ratio: Current Interest balance divided by total amount financed

Moving Forward

- In conclusion, our team is grateful for the opportunity to contribute to Crescent Bank's ongoing success and growth by identifying key trends, inefficiencies, and opportunities in the data.
- We believe the findings and our conclusions will benefit Crescent Bank for many years to come!



Thank You!

