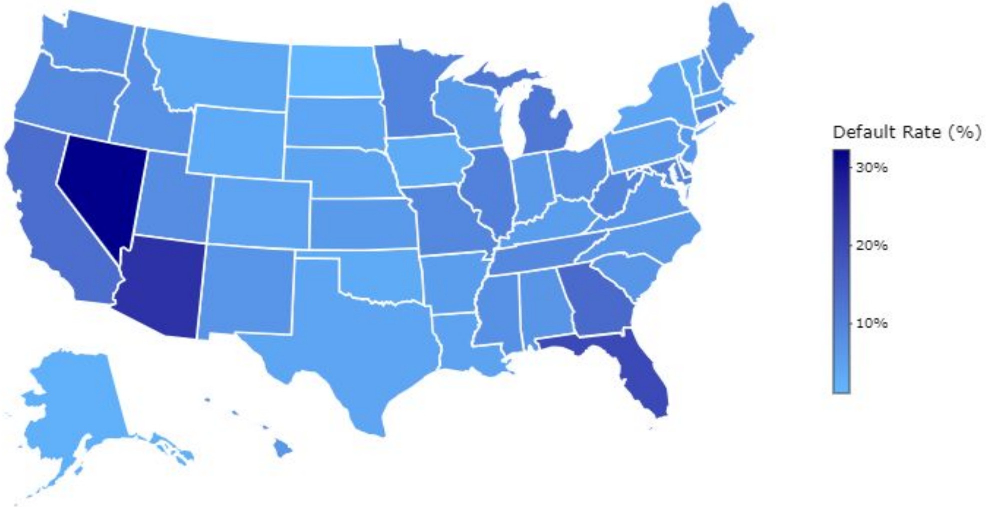
Florida: The American Dream

Loans In Limbo: Florida's Housing Challenge



April 2024: 3rd-Highest Foreclosure Rate
Foreclosure Rate: 1 For Every 2,779 Homes

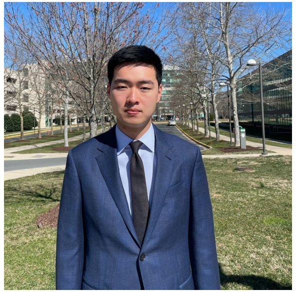
Our Team

Pranav Bhushan

David Dapaah

Chieng-Jui Huang

Tanmay Sakharkar



Henry Liu

Capstone
Final Presentation
Florida 30

Table of **Contents**

01

Executive **Summary**

Executive Summary: Client Overview

FreddieMac
We make home possible®

Freddie Mac:
The Federal Home Loan Mortgage Corporation, also known as Freddie Mac, is a government-sponsored entity (GSE) dedicated to supporting the U.S. housing market.

- **Mission:** Promote stability and affordability in housing by purchasing and securitizing mortgages
- **Impact:** Ensures a steady flow of funds for homebuyers and renters

Stakeholders:

- Freddie Mac
- Fannie Mae
- Corporate Financial Institutions (e.g.: Wells Fargo, Chase Bank, etc.)
- Mortgage Payers
- Market Investors

Executive Summary: Project Overview

Freddie Mac
We make home possible®

Goals:

1. Delinquency Prediction: Predict delinquency rates for home loans in Florida, identifying patterns and trends

2. Payment Class Transition: Predict the probability of loans already 30 days delinquent (Class 1) transitioning to different payment classes: Current (Class 0), 60 day delinquent (Class 2), 90 day delinquent (Class 3), or Repossession (Class RA or 4) over a one-year period

3. Factor Analysis: Identify the key variables/factors contributing to loan delinquency, such as Credit Score or DTI Ratio

4. Model Validation: Plot our delinquency model's prediction against the given data, and minimize the margin of error (MOE)

02

Business Problem & Project Objectives

Business Problem & Project Objectives

Problem:

Florida presents unique challenges with its historically volatile housing market, seasonal population shifts, and natural disaster risks — making it an ideal test case for **developing models to estimate the probability of mortgage delinquency**

Objectives:

1. **Identify Key Predictive Variables**
2. **Develop and Validate Predictive Models**
3. **Provide Data-Driven Insights For Decision Making**

03

Data Overview & Highlights

Data Source

Freddie Mac Single Family Loan-Level Sample Historical Dataset for FL **(2000-2018)**:
- 32 Features (Columns)
- 54,895 Loans (Rows)

Freddie Mac Single Family Loan-Level Sample Performance Dataset for FL **(2000-2018)**:
- 32 Features (Columns)
- 950,000 Monthly Loan Payments (Rows)

Freddie Mac Single Family Loan-Level Cleaned Sample Dataset for FL **(2000-2018)**:
- 32 Features (Columns)
- 9,941 Loans (Rows)

-Used cleaned dataset to do model selection, cross validation, and model training

Data Transformation

Data Transformation Steps:

1) Unpivot The Data To Show Reporting Periods As Columns:

	LOAN_SEQUENCE_NUMBER	02/01/2000	03/01/2000	04/01/2000	05/01/2000	06/01/2000	07/01/2000	08/01/2000	09/01/2000	10/01/2000
0	F00Q10000035	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	F00Q10000049	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	F00Q10000054	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	F00Q10000091	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	F00Q10000094	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

2) Start Tracking For The Next 13 Months When Borrower Misses Their First Payment:

	LOAN_SEQUENCE_NUMBER	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10	Month 11	Month 12	Month 13
0	F00Q10000116	1	0	0	0	0	0	0	0	0	0	NaN	NaN	NaN
1	F00Q10000238	1	0	0	0	0	0	0	0	0	0	0	0	0
2	F00Q10000355	1	2	3	3	3	3	3	3	3	RA	RA	NaN	NaN
3	F00Q10000736	1	1	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	F00Q10000821	1	0	0	0	0	0	0	0	0	0	0	0	0

Data Cleaning & Merging

Data Cleaning Steps:

1) Replace ‘RA’ with ‘4’

2) Drop Rows Where the Delinquency Status For All Reporting Periods Is ‘0’

3) Drop Rows Where There Is No Delinquency Status For Month 13

Data Merging : Utilized Inner Join on The Historical Dataset And The Performance Dataset On ‘Loan Sequence Number’

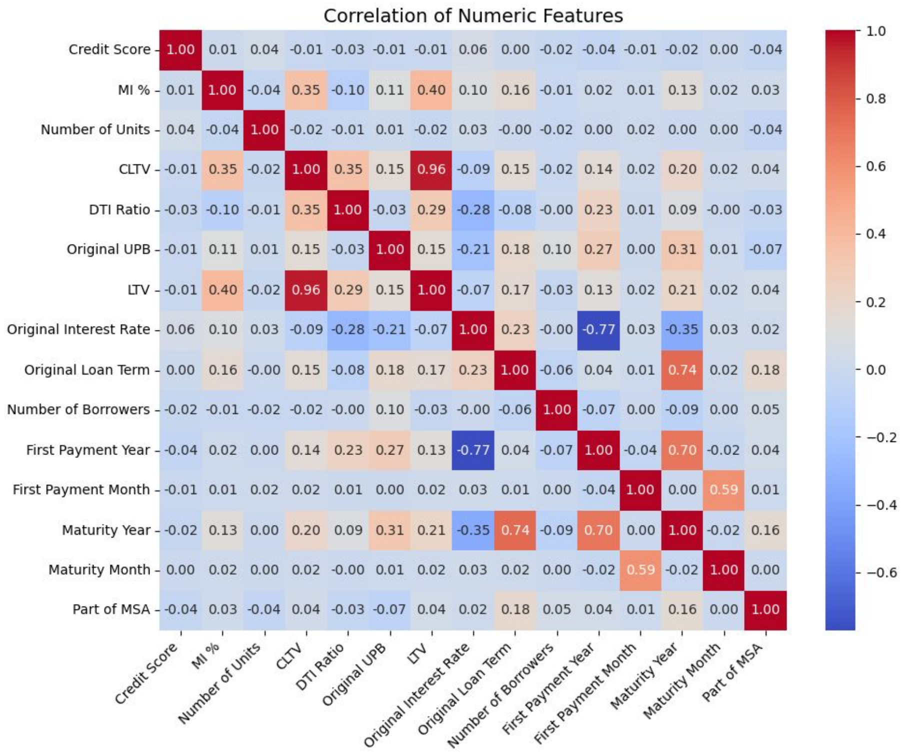
Feature Exploration

LTV (Loan-to-Value) vs CLTV (Combined Loan-to-Value):

- Both measure loan-to-property value, but CLTV includes additional liens
- Frequent refinancing and high home equity loans in Florida cause these metrics to align closely

Maturity Year vs First Payment Year:

- The difference between these variables is the loan term length, which is often fixed
- Florida's housing market trends, such as its preference for traditional fixed-term loans, makes the maturity year and first payment year highly correlated



	Credit Score	MI %	Number of Units	CLTV	DTI Ratio	Original UPB	LTV	Original Interest Rate	Original Loan Term	Number of Borrowers	First Payment Year	First Payment Month	Maturity Year	Maturity Month	Part of MSA
Credit Score	1.00	0.01	0.04	-0.01	-0.03	-0.01	-0.01	0.06	0.00	-0.02	-0.04	-0.01	-0.02	0.00	-0.04
MI %	0.01	1.00	-0.04	0.35	-0.10	0.11	0.40	0.10	0.16	-0.01	0.02	0.01	0.13	0.02	0.03
Number of Units	0.04	-0.04	1.00	-0.02	-0.01	0.01	-0.02	0.03	-0.00	-0.02	0.00	0.02	0.00	0.00	-0.04
CLTV	-0.01	0.35	-0.02	1.00	0.35	0.15	0.96	-0.09	0.15	-0.02	0.14	0.02	0.20	0.02	0.04
DTI Ratio	-0.03	-0.10	-0.01	0.35	1.00	-0.03	0.29	-0.28	-0.08	-0.00	0.23	0.01	0.09	-0.00	-0.03
Original UPB	-0.01	0.11	0.01	0.15	-0.03	1.00	0.15	-0.21	0.18	0.10	0.27	0.00	0.31	0.01	-0.07
LTV	-0.01	0.40	-0.02	0.96	0.29	0.15	1.00	-0.07	0.17	-0.03	0.13	0.02	0.21	0.02	0.04
Original Interest Rate	0.06	0.10	0.03	-0.09	-0.28	-0.21	-0.07	1.00	0.23	-0.00	-0.77	0.03	-0.35	0.03	0.02
Original Loan Term	0.00	0.16	-0.00	0.15	-0.08	0.18	0.17	0.23	1.00	-0.06	0.04	0.01	0.74	0.02	0.18
Number of Borrowers	-0.02	-0.01	-0.02	-0.02	-0.00	0.10	-0.03	-0.00	-0.06	1.00	-0.07	0.00	-0.09	0.00	0.05
First Payment Year	-0.04	0.02	0.00	0.14	0.23	0.27	0.13	-0.77	0.04	-0.07	1.00	-0.04	0.70	-0.02	0.04
First Payment Month	-0.01	0.01	0.02	0.02	0.01	0.00	0.02	0.03	0.01	0.00	-0.04	1.00	0.00	0.59	0.01
Maturity Year	-0.02	0.13	0.00	0.20	0.09	0.31	0.21	-0.35	0.74	-0.09	0.70	0.00	1.00	-0.02	0.16
Maturity Month	0.00	0.02	0.00	0.02	-0.00	0.01	0.02	0.03	0.02	0.00	-0.02	0.59	-0.02	1.00	0.00
Part of MSA	-0.04	0.03	-0.04	0.04	-0.03	-0.07	0.04	0.02	0.18	0.05	0.04	0.01	0.16	0.00	1.00

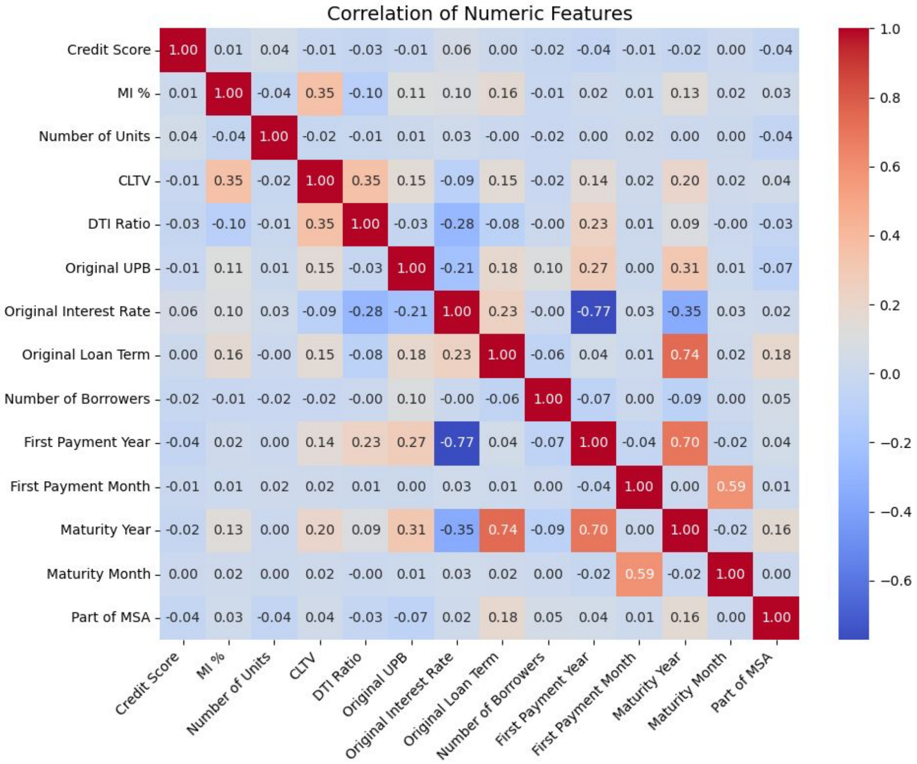
Feature Exploration

Original Loan Term vs Maturity Year:

- The maturity year is directly determined by the original loan term and the loan start date
- In Florida, the prevalence of standardized loan terms (e.g., 15- or 30-year mortgages) creates a direct relationship, leading to high collinearity

Address Multicollinearity:

- **Reduce Redundant Information**: Eliminate or combine variables with overlapping information
- **Set Threshold**: Remove variables with correlation coefficients exceeding 80% to ensure model stability



04

Model Description

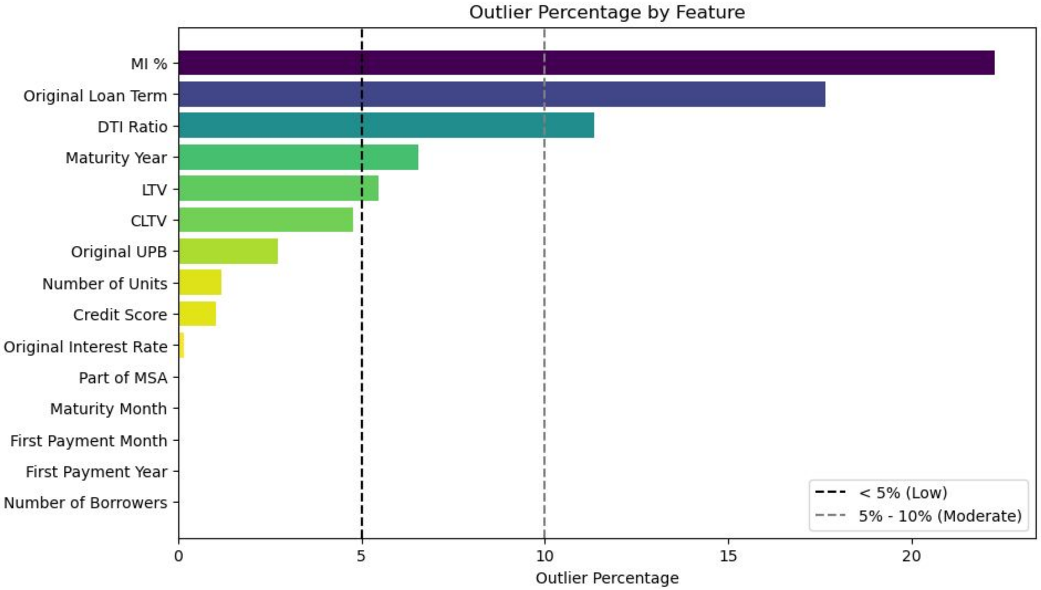
Feature Exploration

Address Outliers To Improve Model Quality:

- Replace outliers with mean values
- Focus on features where outliers exceed a threshold of >5% of the data

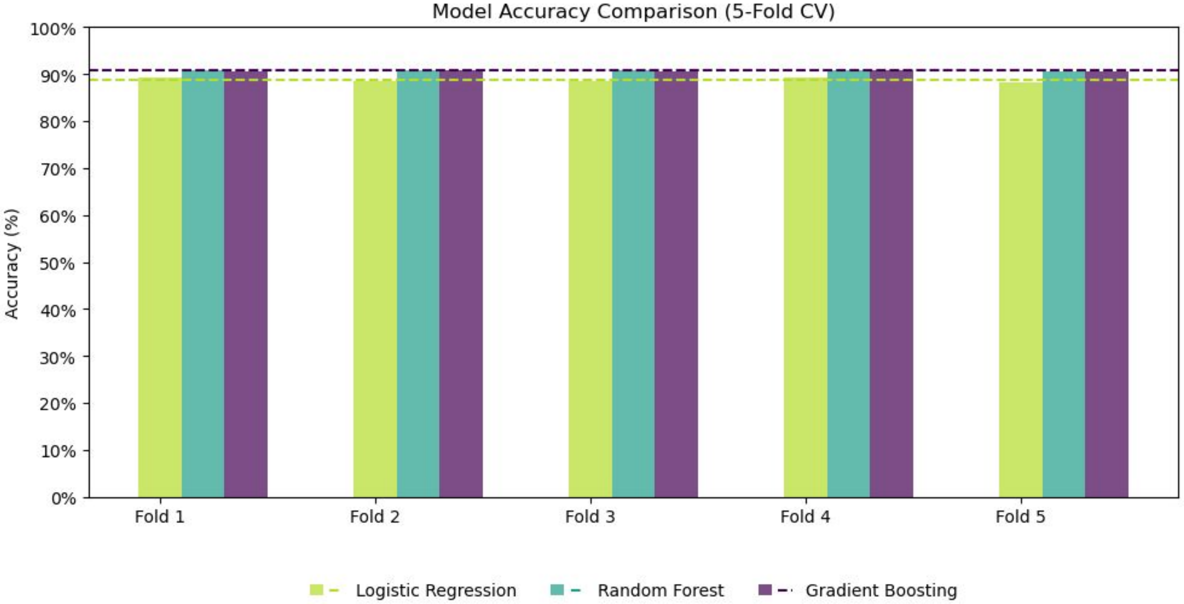
Actions Taken:

- Features Removed Due to Excessive Outliers:
  - Mortgage Insurance %
  - Original Loan Term
  - Debt-to-Income (DTI) Ratio
  - Maturity Year
  - Loan-to-Value (LTV)



Determine Model Specification

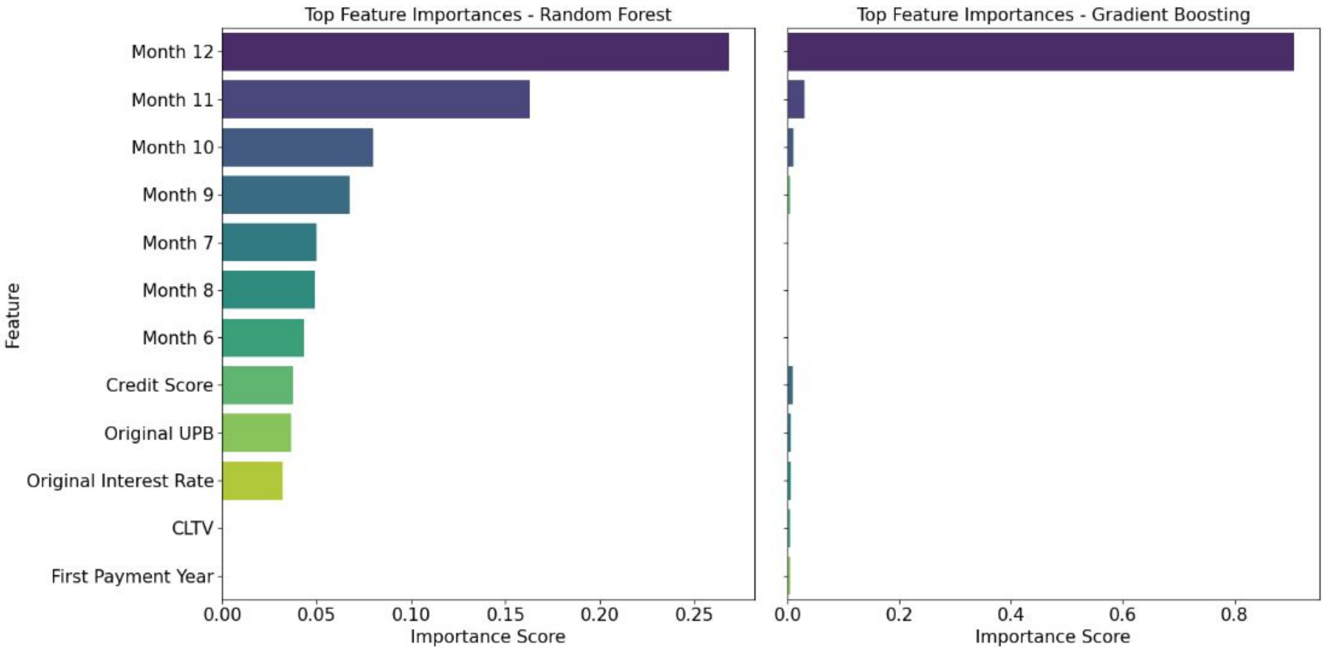
Goal	**Find The Best Model**
How	5-Fold Cross Validation
Why	Make Use of Limited Data
What	Accuracy as Measurement



Finalize Model Specification

Random Forest is Better Due to XGBoost Having The Following Issues:

- **Overfitting:** Limits learning from all features & increases dependency on one feature
- **Consequence:** Feature Bias & reduced robustness & lack of generalization



Features in the Random Forest Model

Payment History	Categorical Features	Numeric Features
• From Month 2 to Month 12	• First Time Buyer • Property Valuation Method • Metropolitan Statistical Area	• Credit Score • Original Combined Loan-to-Value • First Payment Year • Maturity Month • First Payment Month • Number of Borrowers • Number of Property Units

05

Findings & **Insights**

Model Evaluation

Random Forest Classifier

<u>Key Attribute: Subsampling</u>

- Random subset of features for every split
- Lower risk of overreliance on specific features

Less Prone to Overfitting: spreads importance across a variety of features

Accuracy

91%

(Average from 5-Fold Cross-Validation)

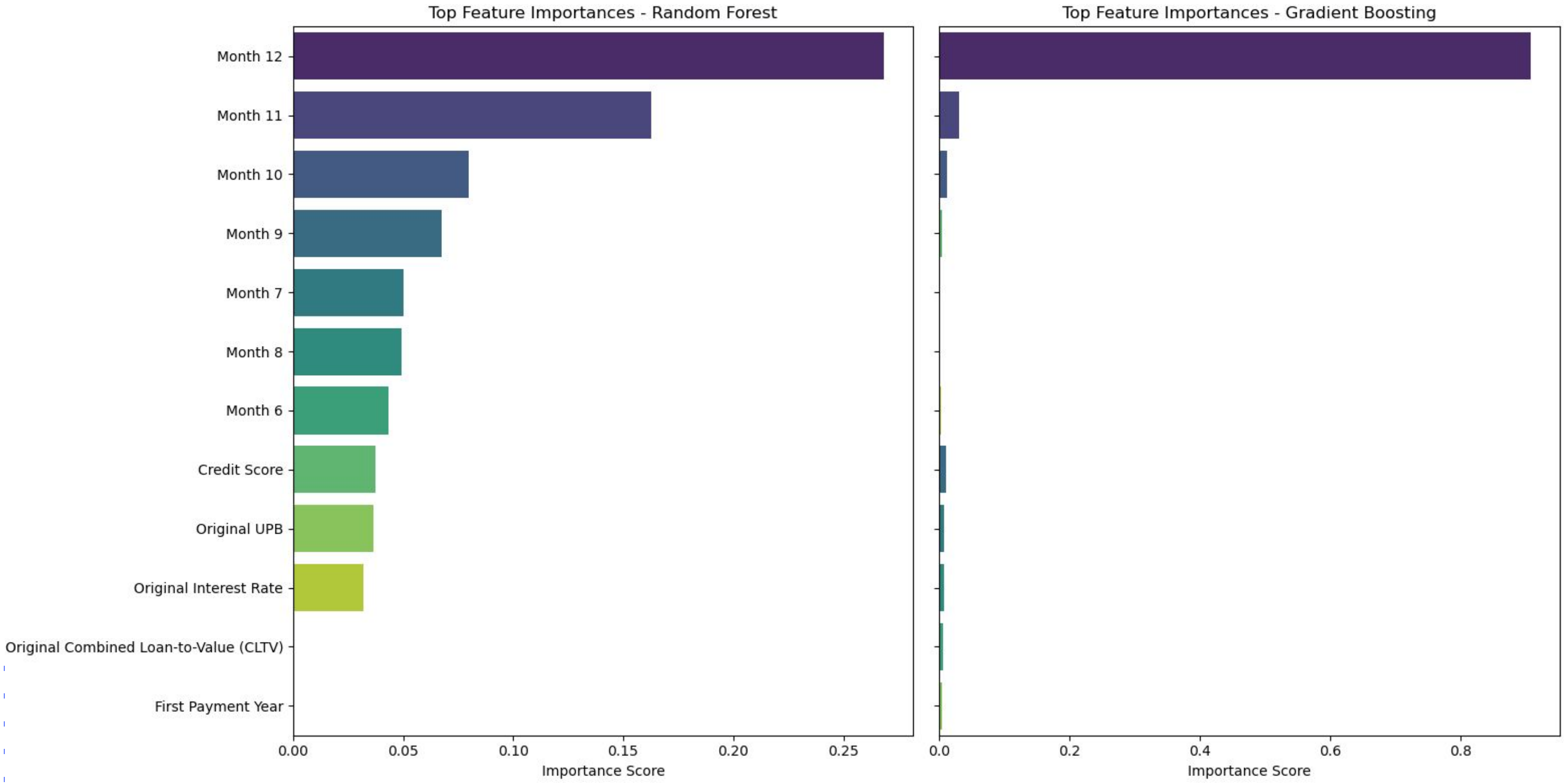
Gradient Boosting Classifier

<u>Key Attribute: Sequentiality</u>

- Sequential trees attempt to correct errors of prior trees
- Higher risk of emphasizing dominant features

More Prone to Overfitting: specific features may be excessively emphasized

Revisiting Features



Actual vs Predicted Probability Distribution of Class 1

	0	1	2	3	RA
Predicted	67.03	5.81	2.08	24.5	0.56
Actual	66.08	6.66	2.61	23.98	0.64

*In Percentage (%)

Takeaways

Payment History is the Most Important Predictor:

- Consistent across both models
- Recent payment history (Months 6-12) is the most valuable predictor

Additional Important Features:

- Credit Score
- Original Unpaid Balance (Amount Borrowed)
- Original Interest Rate

Key Insight:
Freddie Mac should place much higher emphasis on evaluating the most recent history of how a loan has been performing compared to initial attributes of the loan.

06

Challenges & Workarounds

Challenges

Data Issues

Incomplete Data: Missing values in critical variables

Class Imbalance: Unequal distribution across payment classes

Historical Data Size: Large and comprehensive, though computationally intensive

Modeling Constraints

Random Forest Limitations: Computationally intensive and less interpretable

Feature Importance: Difficulty in determining the most impactful features without over-reduction in dimensionality

Operational Constraints

Time: Project timeline constraints restricted model exploration and fine-tuning for deeper analysis

Computational Power: Insufficient tools for processing large datasets efficiently

Workarounds

Data Handling

Imputation:
Replaced missing values in critical features

Outlier Handling:
Dropped features with more than 5% outliers

Dimension Reduction:
Removed irrelevant columns (e.g., "Postal Code")

Improved Modeling

Replaced XGBoost With Random Forest:
Replace the model for better generalization and to mitigate overfitting

Feature Elimination:
Eliminated certain features to identify the most predictive variables while avoiding over-reduction

Workflow Optimization

Streamlined Data:
Used sample dataset instead of historical dataset in order to reduce total code output generation time

Created ETL Pipeline:
Created an ETL pipeline to load, clean, and transform the data in a sequential manner

07

Recommendations & Opportunities

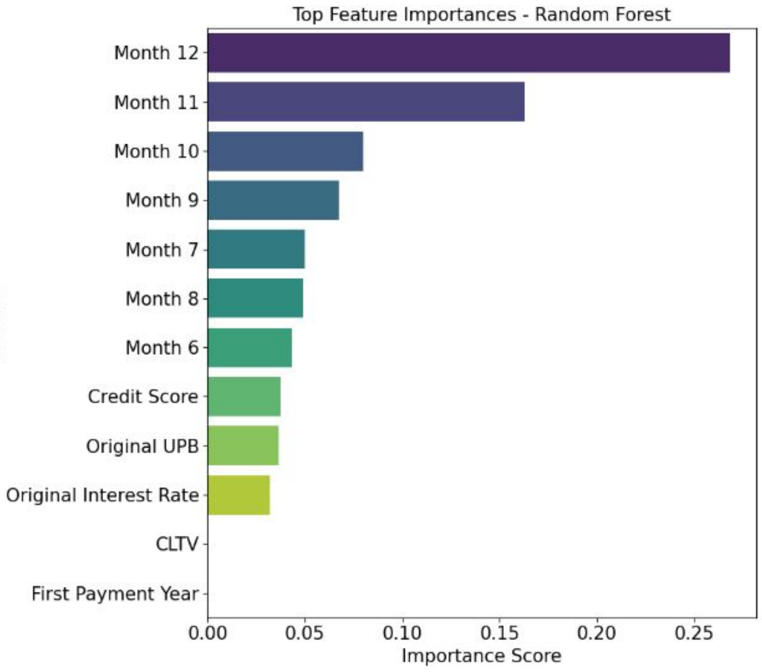
Recommendations

Incorporate Recent Payment History:

- Focus on analyzing the last six months of payment data to assess trends in financial stability
- Use timely payments as a positive indicator of recovery and missed payments as a warning sign for further delinquency

Leverage Credit Score Insights:

- Prioritize borrowers with high credit scores for recovery programs or retention efforts
- Develop targeted interventions for borrowers with low credit scores to mitigate delinquency risks



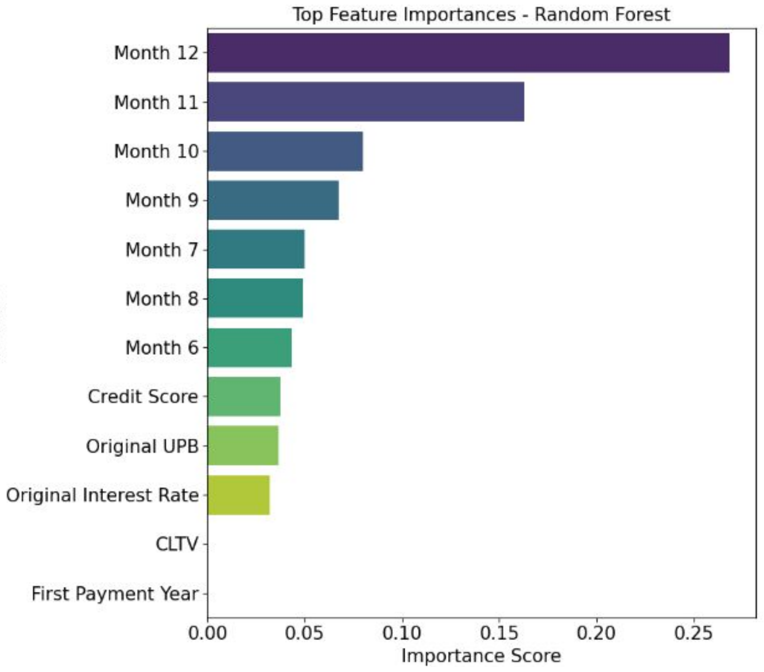
Recommendations

Consider The Impact of Original UPB:

- Pay closer attention to borrowers with higher UPB, as larger loan sizes may indicate a higher risk of financial strain
- Tailor repayment plans or refinancing options for borrowers with lower UPB to ensure affordability

Factor In Original Interest Rates:

- Identify high-interest loans as potential stress points and consider offering rate modifications or consolidation options
- Use low-interest loans as indicators of borrowers with higher recovery potential and less financial burden



Recommendations

	0	1	2	3	RA
Predicted	67.03	5.81	2.08	24.5	0.56
Actual	66.08	6.66	2.61	23.98	0.64

*In Percentage (%)

Focus on High-Risk Borrowers (Class 3):

- Allocate resources to borrowers in worsening conditions (Class 3) through targeted loan restructuring and intensive outreach programs to minimize financial losses

Implement Early Interventions (Class 1 and 2):

- Offer forbearance, repayment plans, or financial counseling to borrowers who are 30-60 days delinquent to prevent escalation to more severe delinquency

Recommendations

	0	1	2	3	RA
Predicted	67.03	5.81	2.08	24.5	0.56
Actual	66.08	6.66	2.61	23.98	0.64

*In Percentage (%)

Maintain Positive Status for Current Borrowers (Class 0):

- Introduce incentives like interest rate reductions or rewards for consistent payments to ensure borrowers stay current

Strengthen Communication and Support:

- Provide clear repayment options, personalized assistance, and financial counseling to enhance borrower engagement and satisfaction

Opportunities

Expand Predictive Modeling Beyond Current Use Cases:

- Apply prediction models for other scenarios such as **COVID-19 impact analysis**, identifying trends in payment behavior, or forecasting recovery rates for economic shocks
- Use these models to address emerging challenges beyond traditional delinquency management

Validate Models With Historical Data:

- Test and refine models using actual historical data to ensure robustness and accuracy
- Showcase the accuracy of these models in predicting key outcomes, building confidence in their application

Opportunities

Drive Business Efficiency Through Insights:

- Use insights from predictive models to optimize resource allocation for high-risk borrowers and tailor interventions
  - This, in-turn, helps reduce delinquency rates and financial losses while improving overall portfolio performance

Leverage Data For Adjacent Business Areas:

- Apply similar strategies in adjacent business areas like auto loan or credit card loan product design to open new growth avenues
- Use data-driven recommendations to scale programs that work effectively, such as early interventions or rewards for positive borrower behavior

Thank You!

Questions?