Motivation of the Problem

DebtConsolidate Corp is committed to supporting its customers in achieving financial stability by offering consolidation loans.

Many customers continue to struggle with debt, and their financial well-being remains compromised.

The company developed a mobile app to promote positive financial behaviors, such as improving credit scores, reducing debt, and increasing savings, through challenges and rewards.

“The motivation behind this project is to leverage analytics and AI to gain a deeper understanding of the factors influencing customers' financial health.”

By identifying the key drivers of financial well-being, we aim to help DebtConsolidate Corp improve their services and offer personalized solutions to customers, ultimately helping them achieve long-term financial stability

Project Objectives (SMART)

Specific: To leverage analytics and AI in order to identify the key factors impacting DebtConsolidate Corp's customers' financial well-being by analyzing data from the company's mobile app and other financial sources.

Measurable: The project's success can be measured by the ability to develop accurate predictive models that can help us guide DebtConsolidate Corp in creating effective financial strategies and improving customer financial wellbeing, such as enhancing credit scores or reducing default rates.

Achievable: We have two years of monthly data from 200 clients, which includes financial, demographic, and behavioral information. The data has been analyzed using AI and ML techniques to achieve the desired goals.

Relevant: To be able to help customers manage their debt more efficiently, it is critical for the company to improve customers' financial well-being. By understanding the key drivers, we can help the company offer better services and support customers in reaching their financial goals.

Time-bound: We have analyzed the data of a specified timeframe (two years), and delivered our analysis within the specified timeline of the EMC project delivery. The objective is to deliver actionable insights that can guide DebtConsolidate Corp's future strategies.

Importance of solving the Problem

Customer Financial Stability: Helping customers manage their debt effectively and improve their financial health can lead to long-term financial stability. This aligns with DebtConsolidate Corp's mission to assist customers in overcoming debt challenges.

Business Value: By identifying the key drivers of financial well-being, the company can develop targeted interventions, reduce default rates, and improve customer loyalty, leading to greater business sustainability.

Improved Service Offerings: With insights from data analytics and AI, DebtConsolidate Corp can tailor its loan products and customer support services to meet the unique needs of different customer segments, improving overall service quality.

Proactive Risk Management: Predictive models can help the company anticipate potential risks, such as customer defaults, enabling proactive measures to mitigate these risks

Descriptive Analysis

Summary statistics:

- Overview of the dataset
- Key measures: Count, mean, standard deviation, minimum, and maximum values
- Understand the distribution and central tendencies of the variables.

	id	Loan Amount	Outstanding Balance	Outstanding Principal	Interest Rate	Loan Term (Months)	Stated Income on application	Qualified / Verified\nIncome	Aptitude for change Score	Financial Literacy Score	Self Assessments	Quiz Count	Mood Count	Inspiration Count	Total Activies	Average total activities per month	Average activities per day	average_wellness_score	average_FICO	average_credit_utilisatio
count	211.000000	211.000000	211.000000	211.000000	211.000000	211.000000	211.000000	211.000000	188.000000	188.000000	194.000000	194.000000	194.000000	194.000000	194.000000	194.000000	194.000000	181.000000	208.000000	208.00000
mean	11446.303318	14768.431896	20370.718768	12089.351280	0.284071	56.715640	78128.687204	74712.706161	3.595638	2.792553	8.128866	61.855670	10.175258	5.170103	85.329897	6.842678	0.228088	48.722638	629.453203	0.70973
std	5635.756041	5659.903418	10790.558090	6229.183424	0.017719	6.936768	50793.448599	45380.645756	1.072512	1.556270	6.084797	53.298454	20.252300	10.620975	80.430155	4.793183	0.159774	11.019077	47.367252	0.29306
min	123.000000	3000.000000	0.000000	0.000000	0.249900	26.000000	32570.000000	32000.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.061200	0.002000	21.000000	439.000000	0.00000
25%	11183.500000	10179.620000	12748.005000	7867.910000	0.269900	58.500000	55000.000000	52500.000000	3.170000	2.000000	3.000000	15.250000	0.000000	0.000000	21.500000	1.818025	0.060625	42.636364	601.535714	0.53657
50%	13443.000000	14400.000000	20465.470000	12220.890000	0.289900	60.000000	70000.000000	68600.000000	3.830000	3.000000	7.000000	52.000000	2.000000	1.000000	64.000000	8.105150	0.270150	49.200000	632.625000	0.72932
75%	15533.500000	20000.000000	29365.620000	17550.330000	0.299900	60.000000	90581.500000	88600.000000	4.330000	4.000000	12.000000	98.000000	10.000000	4.000000	125.000000	9.659475	0.322000	55.071429	659.275000	0.90445
max	18063.000000	25000.000000	45848.500000	24927.320000	0.299900	60.000000	700000.000000	605300.000000	5.000000	5.000000	24.000000	196.000000	110.000000	65.000000	383.000000	22.941200	0.764700	78.166667	788.000000	2.43780

Descriptive Analysis

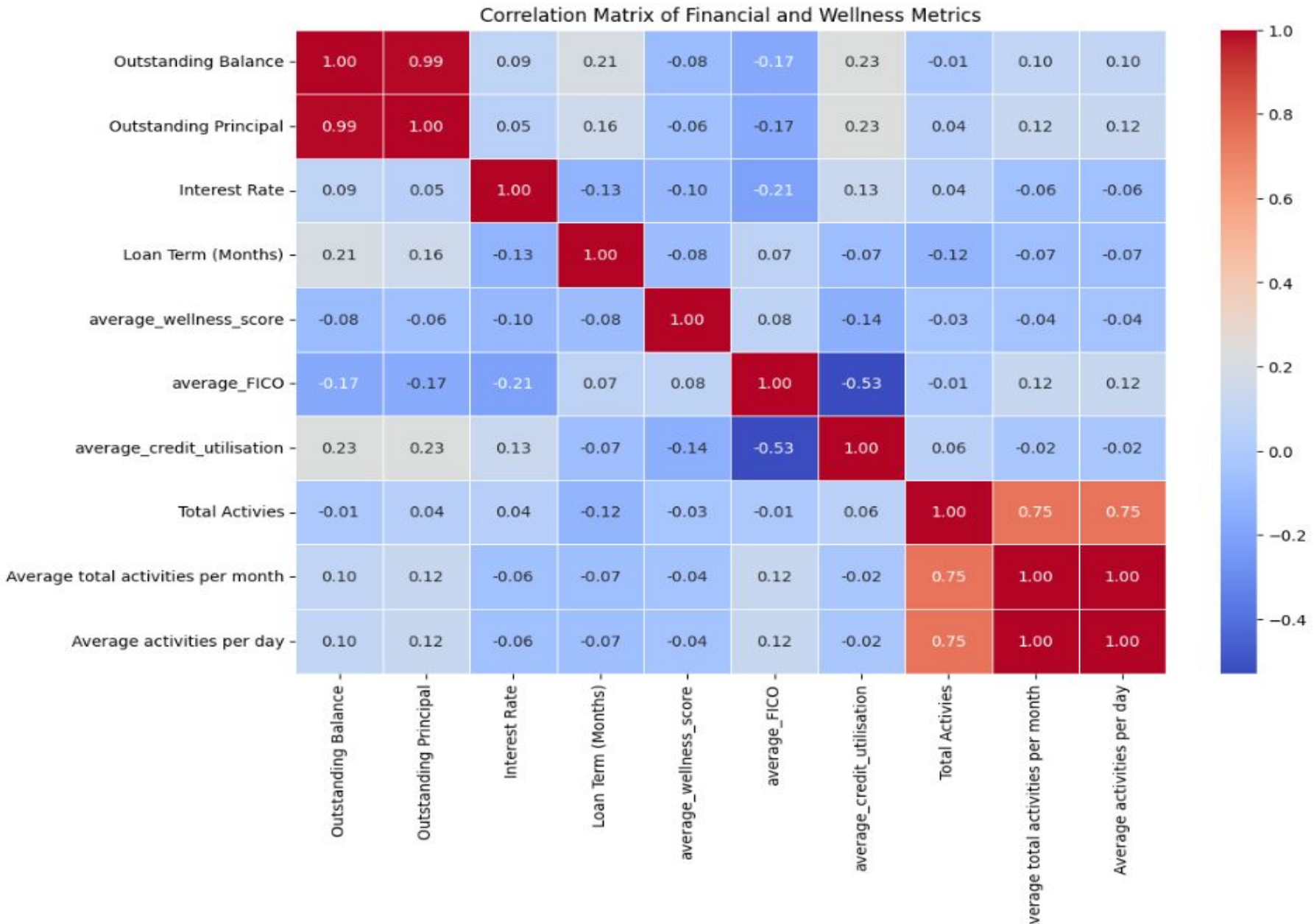
Summary statistics: We obtained overview of the dataset, including key measures like the count, mean, standard deviation, minimum, and maximum values for each variable to understand the distribution and central tendencies of the variables.

Correlation Matrix: Provided relationship between key financial metrics (e.g., outstanding balance, interest rate, loan term) and wellness metrics (e.g., average wellness score, total activities)

Heatmap Visualization:

- Credit utilization & FICO: Inverse correlation. Higher credit utilization negatively impacts the FICO score.
- Outstanding balance & outstanding principal: Strong correlation
- Average wellness score, total activities showed weak correlation with financial metrics. Further analysis was done on it.

Descriptive Analysis: Heatmap



Feature Engineering & Segmentation

New Features engineered:

- Loan-to-Income Ratio: Derived by dividing the loan amount by the customer's stated income. Reflects the financial burden a customer faces in relation to their income.
- Financial Stress Score: Derived by dividing the outstanding balance by the original loan amount. Represents the level of financial stress. A higher outstanding balance relative to the original loan can indicate financial difficulties.

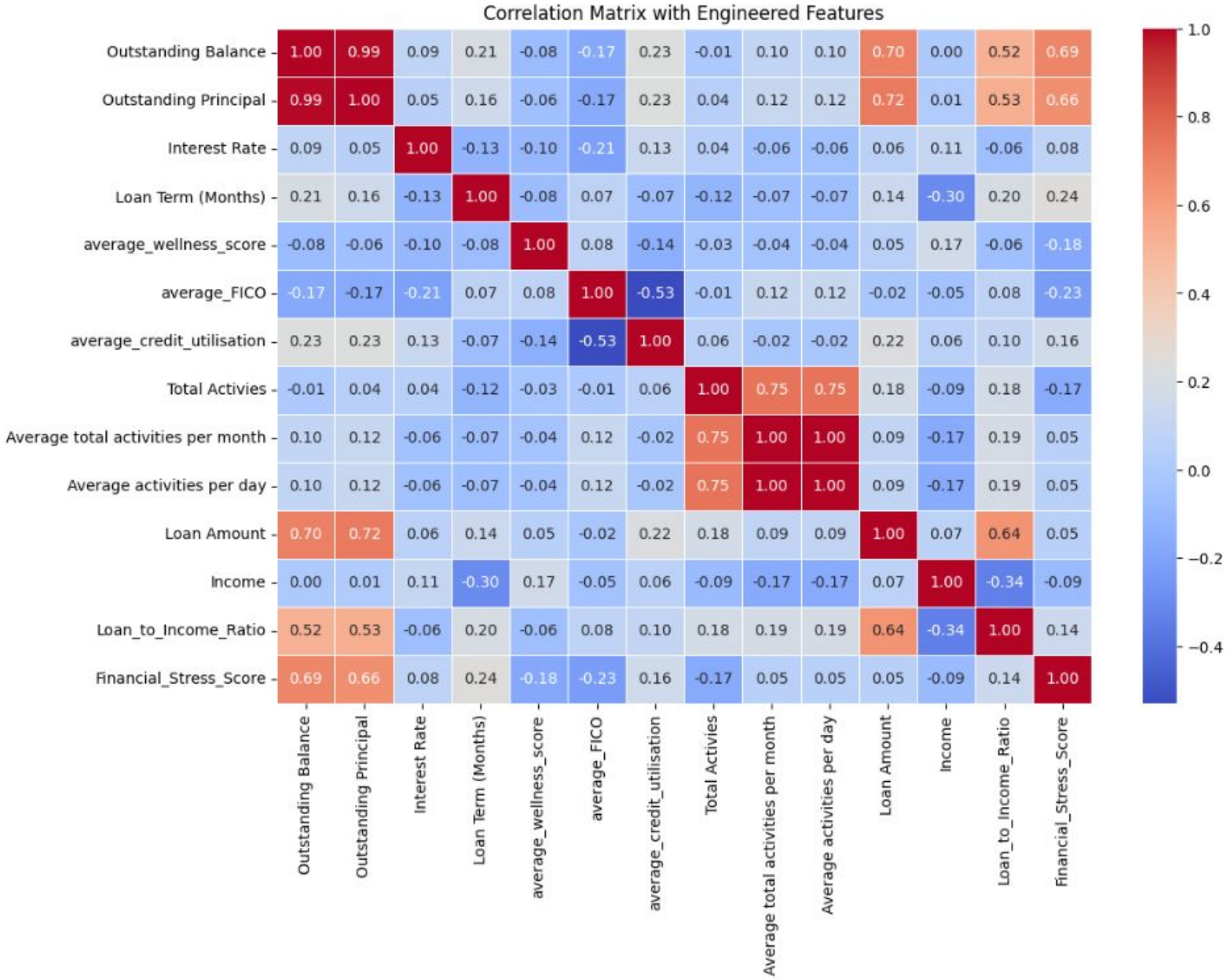
Customer Segmentation by Repayment Behaviour:

- “Good Standing" (loans repaid), "Defaulted" (loans written off), and "At Risk" (loans past due)

Customer categorization by Activity Level:

- "Low," "Medium," and "High" activity levels based on the total activities completed. Reflects correlation of customer engagement with financial well-being.

Additional Heatmap with Engineered Features



- ➔ Very strong: Outstanding Balance & Outstanding Principal
- ➔ Strong: Original Loan Amount & Outstanding Balance
- ➔ Little: Interest Rate & Financial Metrics
- ➔ Low: Average Wellness Score & Financial Metrics
- ➔ Negative: Credit Utilization & FICO Score

Key Insight:

Customers with a high loan-to-income ratio and financial stress score were more likely to be at risk of financial issues.

Methodologies of Analysis

Predictive Models:

- Case 1: Random Forest Classifier for Repayment Prediction
- Case 2: Random Forest Regressor for FICO Score Prediction
- Case 3: XGBoost Regressor for FICO Score Prediction
- Case 4: Ridge Regression for FICO Score Prediction
- Case 5: Lasso Regression for FICO Score Prediction

Predictive Model# 1: Random Forest Classifier

Target Variable: **Repayment_Category** (whether the loan is in good standing, defaulted, or at risk).

This is a classification problem.

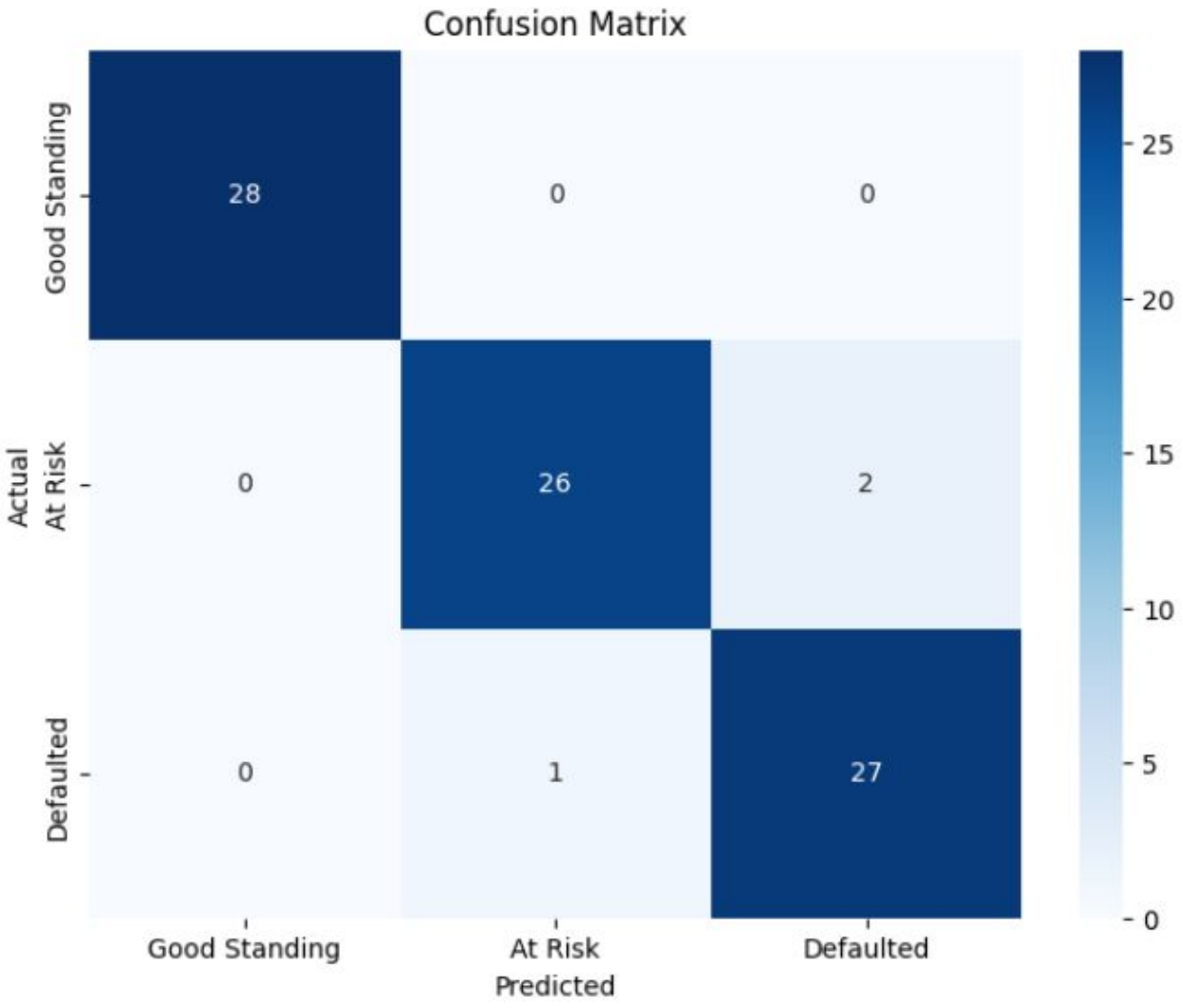
Modeling Selection: Random Forest Classifier

Justification of model selection: Robustness in handling both categorical and numerical data and its ability to mitigate overfitting by averaging multiple decision trees. Given the imbalanced classes in the dataset (with fewer instances of "Good Standing" and "Defaulted"), Random Forest works well with resampling techniques like SMOTE (Synthetic Minority Over-sampling Technique) to balance the class distribution.

Description of the model:

- Train-Test Split: We split the data into training and test sets to evaluate model performance.
- Random Forest creates an ensemble of decision trees, each trained on a random subset of data.
- Evaluation Metrics: For classification, we used Accuracy, F1-score, and Confusion Matrix.

Predictive Model# 1: Evaluation



Classification Report:	precision	recall	f1-score	support
Good Standing	1.00	1.00	1.00	28
At Risk	0.96	0.93	0.95	28
Defaulted	0.93	0.96	0.95	28
accuracy			0.96	84
macro avg	0.96	0.96	0.96	84
weighted avg	0.96	0.96	0.96	84

Accuracy: 96.43%. Shows strong predictive power for classifying ‘loan repayment status’.

F1-score: 1.00 for Good Standing, and 0.95 for both At Risk and Defaulted. Indicates **high precision** and **recall** across all categories.

Confusion Matrix:

- Good Standing: Perfect classification.
- At Risk: 2 instances classified as "Defaulted"
- Defaulted: 1 instance classified as "At Risk."

Predictive Model# 1: Important Highlights

- **Accuracy:** 96.43% after applying SMOTE to balance the dataset, indicating strong overall performance.
- **F1-scores:** 1.00 for "Good Standing" and 0.95 for both "At Risk" and "Defaulted."
- **Confusion matrix:** The model perfectly classifies all "Good Standing" instances, but misclassifies 2 "At Risk" cases as "Defaulted" and 1 "Defaulted" case as "At Risk."
- Despite these minor errors, the precision and recall for all classes remain high, showing that the model handles the previously imbalanced data effectively. The model's performance across all metrics suggests it can reliably predict financial outcomes across the three categories after balancing the dataset.

In summary, this Random Forest Classifier model provides strong Strong classification performance across all metrics, with a few minor misclassifications between the "At Risk" and "Defaulted" categories.

Predictive Model# 2: Random Forest Regressor

Model: **Random Forest Regressor**

Target Variable: Predict **FICO Score**

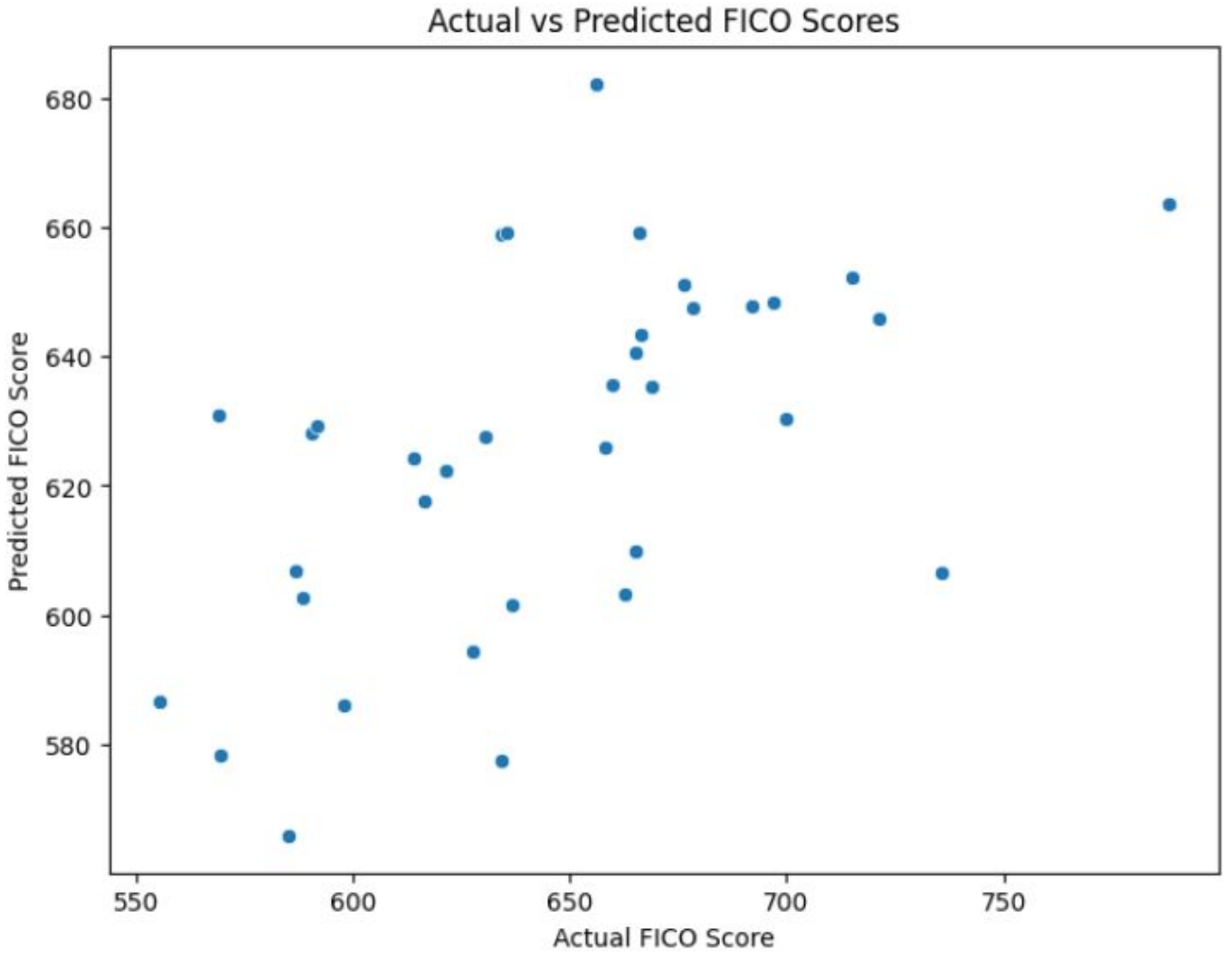
Justification: Its ability to capture complex non-linear relationships, which might exist between financial metrics and FICO scores. It aggregates the predictions of individual decision trees to reduce variance and improve predictive accuracy.

Description: Like the classifier, the Random Forest Regressor builds an ensemble of decision trees. The model outputs the average prediction from these trees, which helps in dealing with noisy data.

Evaluation Metrics: Mean Squared Error (MSE), R-squared (R^2) score

Predictive Model# 2: Evaluation

FICO score distribution:
count 179.000000
mean 627.835965
std 48.926387
min 439.000000
25% 598.798701
50% 631.444444
75% 658.104412
max 788.000000
Name: average_FICO, dtype: float64
Mean Squared Error: 2207.788277079987
R-squared: 0.1507147309050565



R-squared (R^2): 0.15, meaning only 15% of the variance in FICO scores was explained by the model.

Mean Squared Error (MSE): 2207.79, reflecting considerable differences between predicted and actual FICO scores.

Scatterplot: Noticeable spread, indicating the model's difficulty in accurately predicting scores

Predictive Model# 2: Important Highlights

The model underperformed in predicting FICO scores. This suggests that the features used may not be sufficient to capture the variability in FICO scores.

The low R^2 indicates poor predictive power.

In summary, the results suggest that the current features used for prediction may not be strong enough to capture the patterns in FICO scores. Further feature engineering, the inclusion of more relevant predictors (such as loan repayment behavior, credit utilization), or the use of more advanced models like XGBoost or neural networks may help improve predictive accuracy.

Predictive Model# 3: XGBoost Regressor

To improve the model# 2, we introduced ‘credit utilization’ as a feature and use the XGBoost algorithm, which often performs well with structured data.

Model: **XGBoost Regressor**

Target Variable: Predict **FICO Score**

Feature addition: Credit Utilization

Justification: XGBoost was selected for its effectiveness in handling structured data and its ability to model complex relationships through boosting, which sequentially builds models that correct the errors of the previous ones. XGBoost typically outperforms simpler models, especially when there are intricate patterns in the data.

Description: XGBoost uses gradient boosting to improve predictions iteratively. Each new tree corrects the errors made by the previous trees, and the final model is a combination of these boosted trees.

Evaluation Metrics: Mean Squared Error and R-squared

Predictive Model# 3: Evaluation

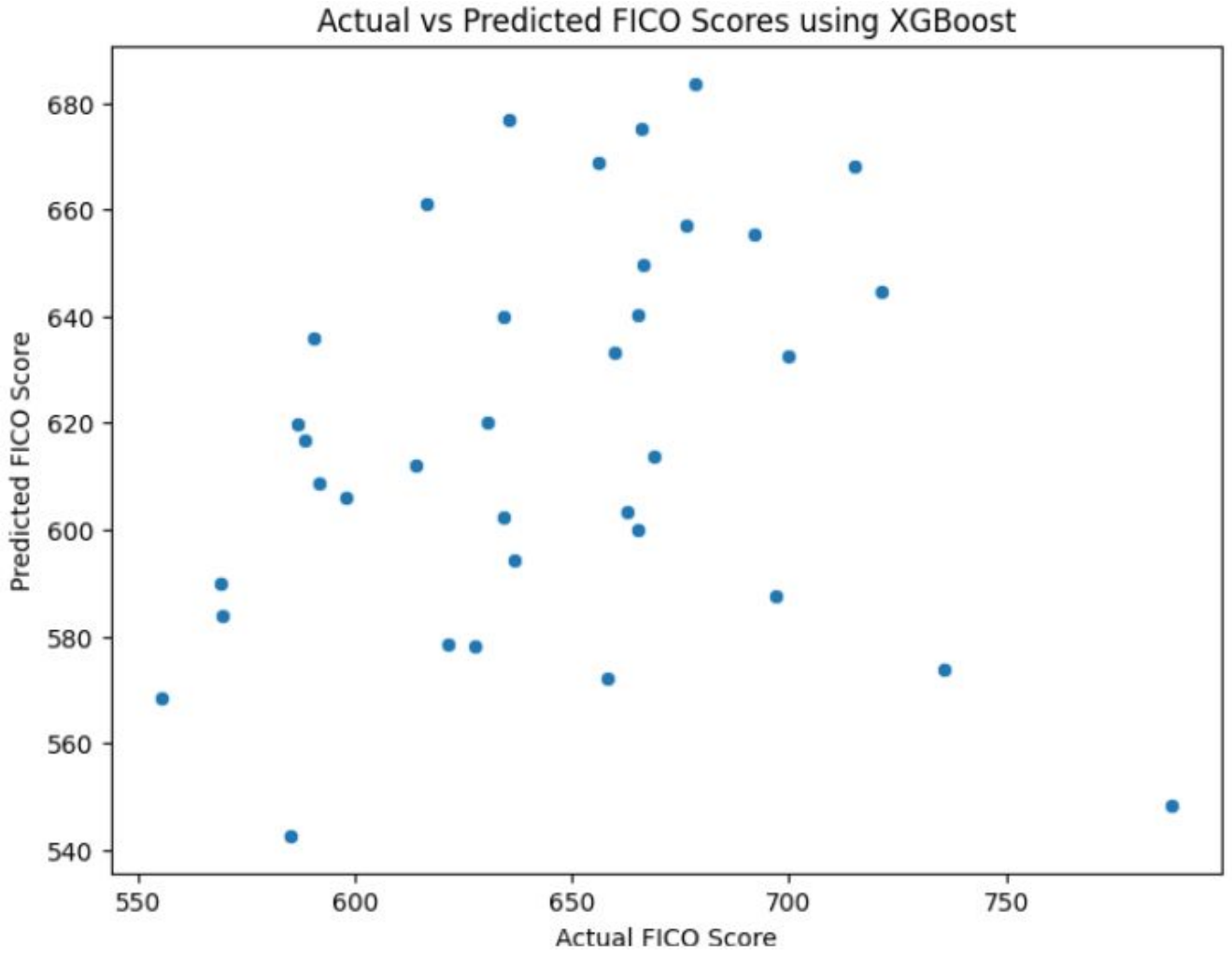
Mean Squared Error: 4109.909277255406

R-squared: -0.5809873812294022

R-squared (R^2): -0.58, a negative value indicating that the model performed worse than a baseline model predicting the mean FICO score for all observations.

Mean Squared Error (MSE): 4109.91, showing a high degree of error between the predicted and actual values.

Scatterplot: The Predicted vs Actual FICO scores are widely scattered. Model struggles with FICO scores> 650 with several large deviations, and no clear pattern for capturing the relationships in the data



Predictive Model# 3: Important Highlights

In summary, with a very high MSE and a negative R-squared score, the XGBoost model struggled significantly, performing worse than Random Forest. The negative R^2 indicates that the model was not able to capture any meaningful relationship between the features and the FICO score. This suggests a mismatch between the model's complexity and the data, or insufficient feature engineering.

The poor performance of XGBoost in this case may be due to insufficient feature engineering, lack of relevant predictors, or improper tuning of the model's hyperparameters. This model is unable to generalize well, suggesting that either more advanced feature extraction is required or a different model may be better suited for this task.

Our suggestion on other Algorithms: Explore other machine learning models, such as neural networks or ridge regression, which might perform better for this kind of regression problem.

Predictive Model# 4: Ridge Regression

As our model# 4, we decided to use **Ridge Regression** to attempt improving the prediction of FICO scores. Ridge regression is a regularized linear regression technique that can help prevent overfitting, especially when the model has many features or when features are highly correlated.

Model: **Ridge Regression**

Target Variable: Predict **FICO Score**

Justification: We chose Ridge Regression, which uses L2 regularization, to help reduce the impact of multicollinearity and prevent overfitting.

Description: Ridge regression is a linear model that minimizes a loss function penalized by the squared magnitude of the coefficients. This regularization shrinks the coefficients, reducing model complexity while still fitting the data.

Evaluation Metrics: Mean Squared Error and R-squared

Predictive Model# 4: Evaluation

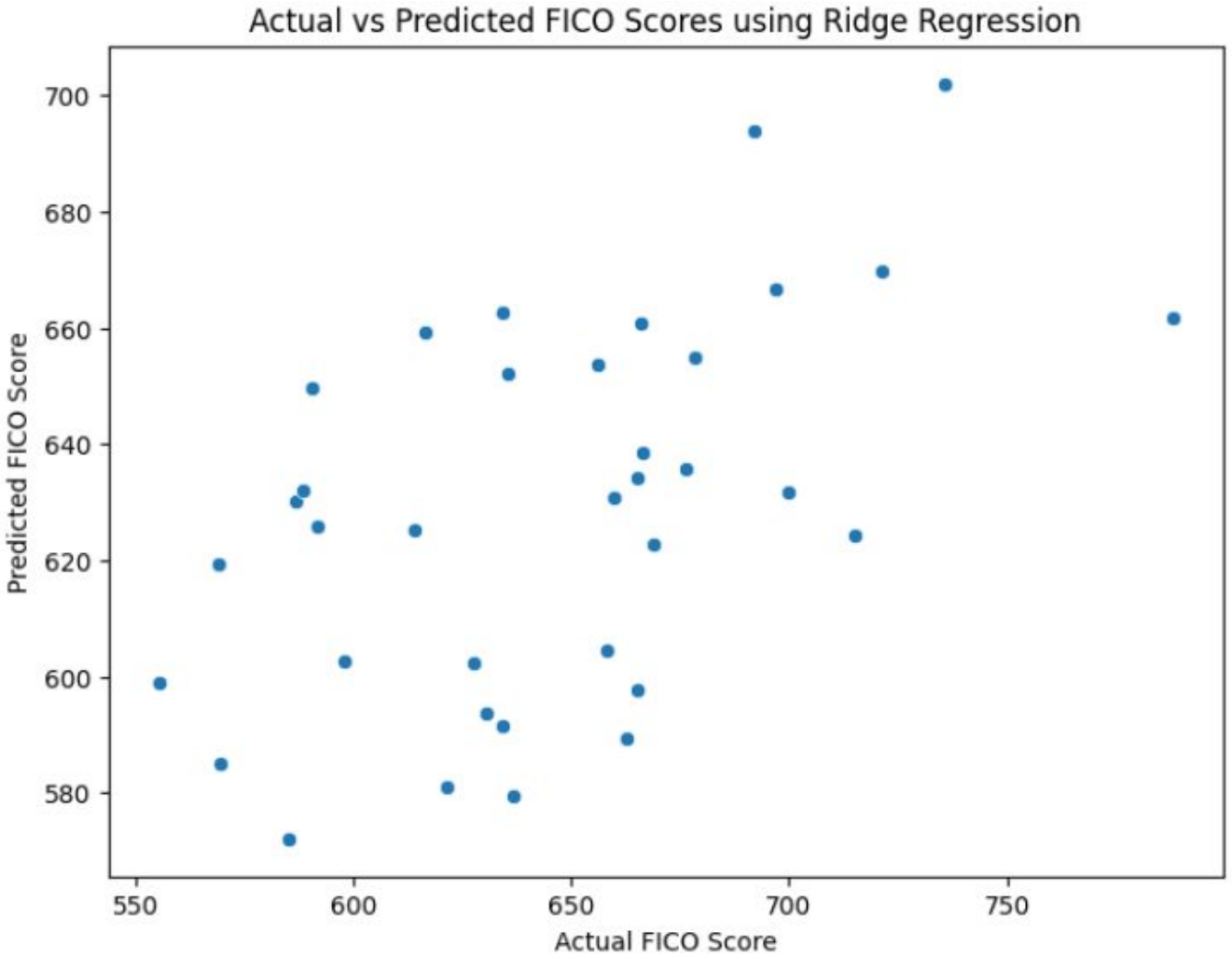
Mean Squared Error: 2178.5809886759907

R-squared: 0.1619501016374932 4

R-squared (R^2): 0.16, meaning the model explained 16% of the variance in FICO scores. Likelihood of important factors influencing FICO scores are not captured by the current features in the model.

Mean Squared Error (MSE): 2178.58, indicating moderate predictive performance. while it is better than XGBoost's performance, there is still significant room for improvement

Scatterplot: Noticeable spread, with several points deviating from the diagonal line, particularly for scores > 650, where the model struggles the most.



Predictive Model# 4: Important Highlights

Ridge Regression performed slightly better than Random Forest and XGBoost, but the low R^2 still suggests that the features were not strong predictors of FICO scores. This model may benefit from further feature engineering.

The Ridge Regression model performs reasonably well for FICO scores between 600 and 650, where the predictions are more tightly clustered around the actual values, but it still shows variability. This model benefits from the regularization offered by Ridge Regression, but the low R^2 suggests that the linear approach may not be fully capturing the relationships in the data, and additional feature engineering or more complex models may be required to improve the predictions.

Our suggestion is to try other models such as Lasso Regression (L1 regularization) or ElasticNet (a combination of L1 and L2 regularization) to further improve predictions.

Predictive Model# 5: Lasso Regression

As our final model, we have used Lasso Regression, which is similar to Ridge Regression but uses L1 regularization. This type of regularization tends to shrink less important feature coefficients to zero, effectively performing feature selection as well as regression.

Model: **Lasso Regression**

Target Variable: Predict **FICO Score**

Justification: Lasso regression, which uses L1 regularization, was selected for its ability to shrink irrelevant feature coefficients to zero, effectively performing feature selection. This can improve model interpretability and focus on the most important predictors.

Description: Lasso regression minimizes a loss function with an added penalty proportional to the absolute magnitude of the coefficients. This forces some coefficients to be exactly zero, reducing the complexity of the model.

Evaluation Metrics: Mean Squared Error and R-squared

Predictive Model# 5: Evaluation

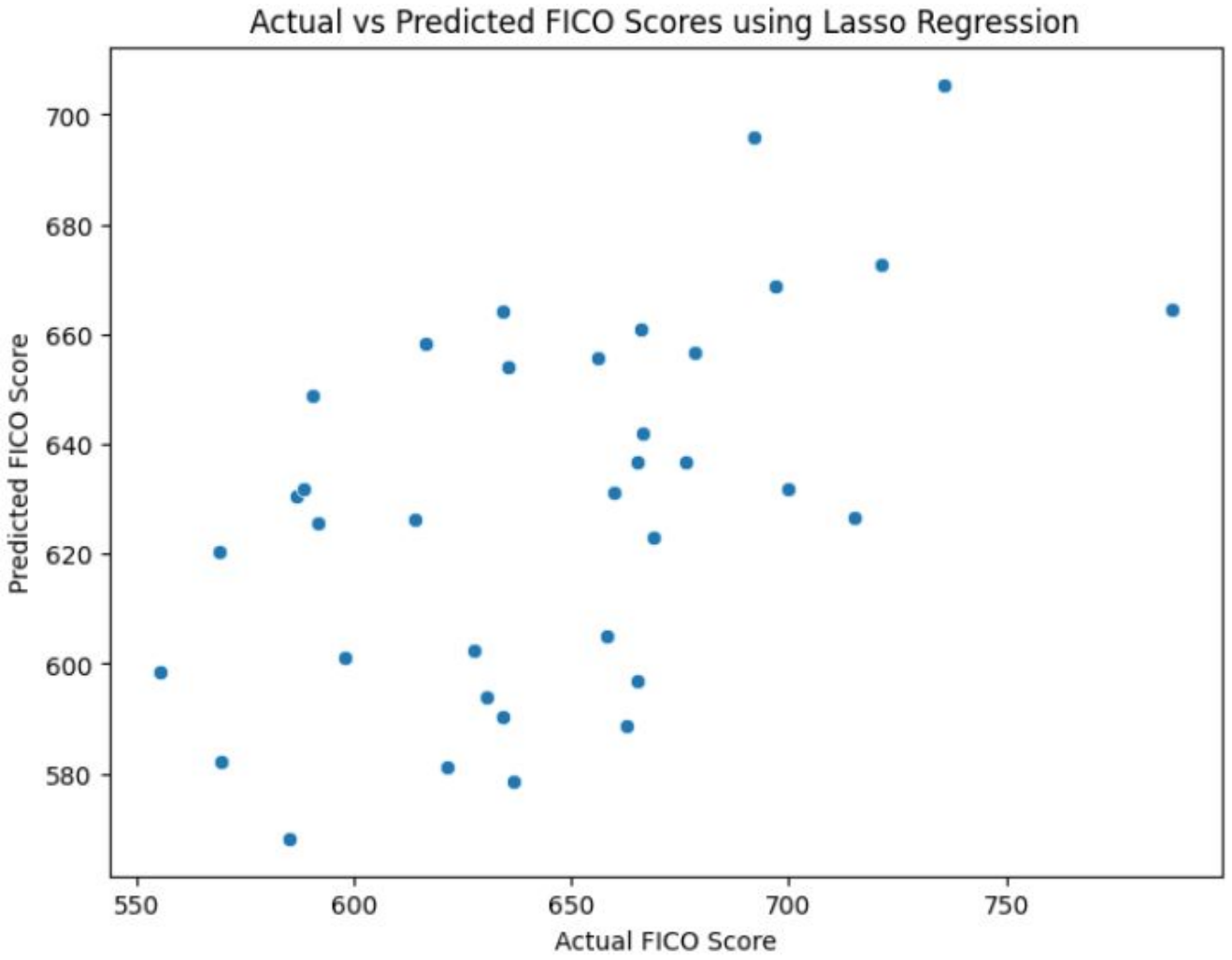
Mean Squared Error: 2125.2706485964477

R-squared: 0.18245736086611786

R-squared (R^2): 0.18, indicating that 18% of the variance in FICO scores was explained by the model.

Mean Squared Error (MSE): 2125.27, a slight improvement over Ridge Regression; however, there is still room for refinement.

Scatterplot: Still noticeable spread for scores > 650. The model performs slightly better for customers with scores between 600 and 650, where the points are more clustered, but still shows variability in predictions.



Predictive Model# 5: Important Highlights

The Mean Squared Error (MSE) of 2125.27 reflects the average squared difference between the actual and predicted FICO scores, and while it is an improvement over previous models, it suggests there is still room for refinement.

The R-squared (R^2) value of 0.18 means that the model only explains about 18% of the variance in the FICO scores, which indicates that many important factors affecting FICO scores are not captured by the current feature set.

"While Lasso Regression performed slightly better than Ridge in terms of R^2 and MSE, the low R^2 score and significant scatter indicate that the model struggled to predict FICO scores accurately."

Additional feature engineering or more complex models may be needed to improve prediction accuracy.

Results & Recommendations

(all five cases)

Summary of Model Results

Model Case	Model Results	Key Recommendations / Actionable Implications
Case #1: Random Forest Classifier (Loan Repayment Prediction)	- Accuracy: **96.43%** - F1-score: **1.00 for Good Standing**, **0.95 for At Risk and Defaulted** - Minor misclassification between “At Risk” and “Defaulted” classes	- **Risk-Based Segmentation**: Implement proactive interventions for at-risk customers (flexible repayment plans, financial counseling) - **Targeted Marketing**: Engage “Good Standing” customers with loyalty programs, promotions, and referrals to attract new clients.
Case #2: Random Forest Regressor (FICO Score Prediction)	- R^2: **0.15** (explained 15% variance) - MSE: **2207.79** - Poor performance in predicting FICO scores	- **Feature Engineering**: Improve the feature set by incorporating more behavioral and financial variables - **Customer Education**: Introduce educational content to improve financial literacy and help customers manage credit utilization.
Case #3: XGBoost Regressor (FICO Score Prediction)	- R^2: **-0.58** (negative performance) - MSE: **4109.91** - Underperformed significantly, worse than predicting mean value	- **Revise Model**: Re-evaluate the model choice and improve feature engineering - **Data Enrichment**: Incorporate richer financial and customer engagement data to enhance model accuracy.
Case #4: Ridge Regression (FICO Score Prediction)	- R^2: **0.16** (explained 16% variance) - MSE: **2178.58** - Slightly better performance than Random Forest	- **Marketing Outreach**: Use insights from financial stress scores to segment and tailor marketing messages, encouraging improved financial behaviors (e.g., debt reduction programs).
Case #5: Lasso Regression (FICO Score Prediction)	- R^2: **0.18** (explained 18% variance) - MSE: **2125.27** - Slight improvement over Ridge regression	- **Personalized Financial Tools**: Offer personalized budgeting and credit improvement tools based on risk factors like high loan-to-income ratios and financial stress scores.

Interpretation of Results Using Relevant Marketing Metrics

Customer Segmentation: The Random Forest Classifier (Case #1) successfully classified customers into groups like "Good Standing," "At Risk," and "Defaulted." This segmentation is valuable from a marketing standpoint as it helps tailor communication strategies to each group.

- Customers in Good Standing can be targeted for loyalty programs,
- Customers at Risk can receive targeted financial guidance or offers to consolidate debt.

FICO Score Prediction: The regression models (Cases #2 to #5) showed challenges in accurately predicting FICO scores, which can be tied to customer lifetime value (CLV) and churn rates. Although these models underperformed, knowing which factors (like credit utilization) most affect FICO scores helps refine customer retention strategies by predicting financial health declines and offering timely interventions to improve retention.

Key insights

	average_credit_utilisation	average_FICO \
0	0.947119	565.571429
3	0.938035	612.571429
4	0.897727	524.000000
5	0.636450	748.500000
6	1.757357	599.142857

	Risk_Segment
0	High Risk - At Risk of Default
3	Medium Risk - Needs Financial Guidance
4	High Risk - At Risk of Default
5	Low Risk - Good Standing
6	High Risk - At Risk of Default

	average_credit_utilisation	average_wellness_score \
0	0.947119	21.000000
3	0.938035	48.200000
4	0.897727	44.857143
5	0.636450	50.000000
6	1.757357	36.285714

	Behavioral_Trigger
0	Trigger - High Credit Utilization
3	Trigger - High Credit Utilization
4	Trigger - High Credit Utilization
5	No Trigger
6	Trigger - High Credit Utilization

	Risk_Segment	Behavioral_Trigger \
0	High Risk - At Risk of Default	Trigger - High Credit Utilization
3	Medium Risk - Needs Financial Guidance	Trigger - High Credit Utilization
4	High Risk - At Risk of Default	Trigger - High Credit Utilization
5	Low Risk - Good Standing	No Trigger
6	High Risk - At Risk of Default	Trigger - High Credit Utilization

	Personalized_Recommendation
0	Recommendation: Focus on reducing credit utilization and exploring debt consolidation
3	Recommendation: Keep up with healthy financial habits
4	Recommendation: Focus on reducing credit utilization and exploring debt consolidation
5	Recommendation: Keep up with healthy financial habits
6	Recommendation: Focus on reducing credit utilization and exploring debt consolidation

Summary from key insights

1. Customer Segmentation (Risk Segment) based on their credit utilization and FICO score:

- **High Risk - At Risk of Default:** Customers like row 0, 4, and 6 have high credit utilization (e.g., 0.947, 0.897, and 1.757) and lower FICO scores (e.g., 565, 524, and 599). These customers are considered at high risk for default, needing urgent attention.
- **Medium Risk - Needs Financial Guidance:** Row 3 shows a customer with medium credit utilization (0.938) and a moderate FICO score (612). This customer might not be in immediate danger but could benefit from financial guidance to avoid further deterioration.
- **Low Risk - Good Standing:** Row 5 represents a customer with lower credit utilization (0.636) and a high FICO score (748). This customer is in good financial standing and does not require immediate intervention.

2. Behavioral Triggers:

- **High Credit Utilization Trigger (>0.8):** These customers might be over-relying on credit and could be at risk of financial stress. Early intervention can help these customers manage their credit better.
- **No Trigger/ lower credit utilization (0.636):** Indicates stable financial behavior.

3. Personalized Recommendations:

- **High Risk Customers (e.g., Rows 0, 4, and 6):** Needs immediate intervention to prevent further financial decline.
- **Medium Risk (e.g., Row 3):** Maintain healthy financial habits while focusing on credit utilization reduction to avoid falling into higher risk
- **Low Risk (e.g., Row 5):** Continue with the healthy financial habits, as these customers are in good financial standing and do not need any drastic changes.

Strategic Action Points:

- **High Risk Customers** should be prioritized for interventions like debt consolidation or financial restructuring to prevent default.
- **Medium Risk Customers** need targeted financial guidance to lower their credit utilization and improve their overall financial health.
- **Low Risk Customers** should be encouraged to maintain their healthy financial behaviors, with minimal intervention required.

Linking Results to Actionable Implications

Personalized Risk Management: The Random Forest Classifier (Case #1) allows DebtConsolidate to predict repayment risks with over 96% accuracy. This insight enables the company to offer customized financial solutions based on risk levels, reducing default rates and improving financial stability. For example, customers identified as At Risk can be offered more flexible payment plans or financial counseling.

Behavioral Engagement: Although the predictive models (Cases #2 to #5) for FICO scores were less successful, they still identified factors such as loan-to-income ratio and financial stress score as important. This can inform marketing strategies, such as developing educational content or tools that help customers manage their credit utilization, improving both financial literacy and overall customer engagement.

Targeted Campaigns: For Good Standing customers, a rewards program or referral marketing campaign could encourage them to recommend DebtConsolidate services to friends and family, leveraging their positive experiences to drive customer acquisition.

Key message of taking the analytical approach to address the defined problem

By using AI and data-driven techniques, DebtConsolidate Corp can move from reactive to proactive customer management. The art lies in how these insights are communicated and applied to real-world marketing strategies. By weaving the story of risk-based segmentation into a customer-centric narrative, the company can show how it is uniquely positioned to help customers improve their financial health.

On the science side, the models provide evidence-backed predictions. The Random Forest Classifier clearly demonstrates that machine learning can accurately predict repayment risks, offering a high level of confidence in identifying which customers need attention. Meanwhile, the less successful regression models underscore the need for continuous improvement in data collection and feature engineering, reinforcing the importance of data-driven decision-making in improving customer outcomes.

Quantifying the Value Created for the Firm

- **Increased Retention:** Identifying at-risk customers early enables proactive outreach, reducing churn rates. Even a 5-10% reduction in customer churn could lead to substantial increases in customer lifetime value (CLV).
- **Return on Investment (ROI):** By quantifying the value of AI and analytics, DebtConsolidate can demonstrate a clear ROI from its investment in predictive modeling, not only enhancing its financial performance but also improving customer outcomes.
- **Reduced Defaults:** By using the predictive capabilities of the Random Forest Classifier, DebtConsolidate Corp can lower the number of defaults, which directly improves its bottom line. For instance, a 10% decrease in default rates could save the company significant resources in debt collection and improve overall financial performance.
- **Marketing Efficiency:** AI-driven insights allow for smarter allocation of marketing spend. By focusing marketing efforts on high-value, low-risk customers while providing targeted assistance to at-risk customers, the company can see improved Return on Marketing Investment (ROMI). A 15-20% improvement in marketing efficiency due to targeted campaigns could result in significant cost savings and revenue growth.

Conclusion

Summary

Model Performance:

- **Random Forest Classifier (Case #1):** Achieved 96.43% accuracy in predicting loan repayment status (Good Standing, At Risk, Defaulted). This high performance indicates strong predictive capability, especially for risk-based segmentation.
- **FICO Score Prediction Models (Cases #2 to #5):** Regression models (Random Forest, XGBoost, Ridge, Lasso) struggled to accurately predict FICO scores, with R^2 values ranging from -0.58 to 0.18, showing limited predictive power due to insufficient features or model complexity.

Actionable Implications:

- **Target At-Risk Customers:** Use segmentation to offer personalized repayment plans, reducing defaults.
- **Engagement Programs:** Improve customer wellness engagement with tailored communication and rewards.

Value Creation (ROI of AI):

- **Increased Customer Retention:** Identify and address the needs of at-risk customers early. Reduce defaults by 10-15%, increasing customer lifetime value (CLV) and reducing costs associated with collections.
- **Marketing Efficiency:** Our AI-driven segmentation could improve the Return on Marketing Investment (ROMI) by 15-20%, optimizing resources toward high-value customers and reducing marketing waste.