Credit Risk Report

By: Laymoni Morrison

Project Overview

This report outlines the full process of building a credit risk analysis pipeline, from data preparation and model development in Google Colab to interactive data storytelling through Power BI. The goal was to evaluate risk and profitability in peer-to-peer loans, using historical loan data to identify trends, predict default likelihood, and present insights for informed decision-making.

Methods

The project began with importing the LendingClub loan dataset into Python, where extensive data cleaning and preprocessing were performed. Columns with over 40% missing values were removed to preserve data quality. Categorical variables such as grade, sub_grade, emp_length, and home_ownership were encoded into appropriate formats. For handling missing values, median imputation was used for numerical fields and the mode for categorical ones.

Feature engineering played a crucial role. A binary loan status variable was created to represent default versus paid loans. Interest income was computed as the product of loan amount and interest rate. Estimated profit was calculated by subtracting losses from defaults from the interest income.

Three models were trained to predict defaults: Logistic Regression, Random Forest, and XGBoost. XGBoost was selected as the final model due to its strong performance. The final predictions and metrics were exported to a CSV file and later visualized in Power BI.

Power BI Dashboard Breakdown

Page 1: Credit Risk Dashboard - Overview

The filters allow users to segment the data by borrower features such as home ownership, employment length, and loan grade. This enables exploration of risk and profitability across borrower profiles. Loan grade is a key metric representing the borrower's creditworthiness, ranging from A (highest) to G (lowest).

The KPIs at the top of the page provide a quick summary: the total loan amount issued was $18B, projected interest income is $16.76M, and estimated profit is negative $1.40B, indicating losses from predicted defaults. About 34,000 loans are expected to default, and the average interest rate across all loans is 13.24%.

A bar chart displays loan volume by credit grade, showing that Grade A receives the highest volume. A stacked column chart reveals how home ownership type impacts loan volume across different grades. A tooltip page is activated on hover for deeper insights on KPIs and bar charts.

Page 2: Credit Risk Dashboard – Deeper Insights & Model Evaluation

The second page offers a closer look at individual predictions and model performance. A table lists top borrowers at risk of default, sorted by predicted probability, with conditional formatting highlighting levels of risk from green to red. This provides a clear way to identify high-risk individuals.

A bar chart compares estimated profit and average interest rate across loan grades. While higher grades yield lower interest rates, they are more profitable due to reduced risk. Another chart shows predicted default rates by grade, confirming that default risk increases as credit grade decreases.

Page 3: Credit Risk Dashboard – Geographic View

The third page contains two maps. The first shows total loan amounts issued across U.S. states, with California, Texas, and New York leading in loan volume. The second map illustrates average interest rates by state, with southern and midwestern regions tending to have higher interest rates due to greater credit risk.

Dashboard Questions & Insights

Are we making money overall and which types of loans are helping or hurting?

Overall, we are not making a profit. The projected interest income is significantly outweighed by expected losses due to defaults, leading to an overall estimated loss of $1.40 billion. Grade A loans are helping the most, as they provide steady returns with low risk. Lower-grade loans, especially F and G, contribute the most to losses due to high default rates.

Which borrower segments are driving the highest risk or returns?

Segments with low credit grades, short employment length, and renters tend to drive the highest risk. In contrast, borrowers with stable employment, homeownership, and higher grades contribute more positively to returns. These segments show stronger repayment trends and more predictable profits.

Which credit grades are getting the most capital and are they worth the risk?

Grades A through C receive the majority of issued capital. Grade A in particular receives the highest loan volume and demonstrates the most favorable risk-return profile. These grades are generally worth the risk, while capital invested in Grades E to G yields significantly less return and more defaults.

How does housing status influence loan profitability?

Borrowers who own homes tend to have better credit grades and lower default rates. As a result, homeownership correlates with higher profitability. Renters, in contrast, are often assigned lower credit grades and have higher chances of default, reducing their contribution to overall profits.

Which purposes dominate loan demand and do they drive gains or losses?

Debt consolidation and credit card refinancing dominate loan purposes. These categories show mixed profitability; while they account for a large portion of volume, they also come with varying levels of risk. Debt consolidation tends to yield lower profits due to higher associated default rates.

Which borrower segments are receiving the most capital and are they also the most risky?

The borrower segments receiving the most capital are those with higher grades, especially A and B, who are not the most risky. This is a positive indicator that capital allocation is aligned with lower-risk profiles. However, some capital is still directed toward lower-grade segments, increasing exposure to risk.

Which model performs best at predicting loan defaults?

Among the models tested, Logistic Regression, Random Forest, and XGBoost, the XGBoost model performed the best. It delivered the highest accuracy and the most reliable predictions when evaluating metrics such as AUC, MAE, and confusion matrix outputs.

How does borrower grade impact profitability?

Borrower grade significantly impacts profitability. Grade A borrowers are the most profitable due to their low default rate, despite having lower interest rates. As grades descend toward G, profitability declines sharply due to increased risk and likelihood of non-payment.

Are we lending too much to high-risk borrowers, and missing out on safer, more profitable ones?

While much of the capital is directed toward safer borrowers, there is still a concerning portion being lent to high-risk borrowers in lower grades. Redirecting more capital to Grade A and B borrowers could improve profitability and reduce overall risk exposure.

Are we focusing too heavily on high-volume but low-profit states?

Yes, some states like California and Texas have high loan volumes but mixed profitability. While these states dominate in volume, the profit margins vary. More analysis is needed to ensure we are not sacrificing profit for scale in these regions.

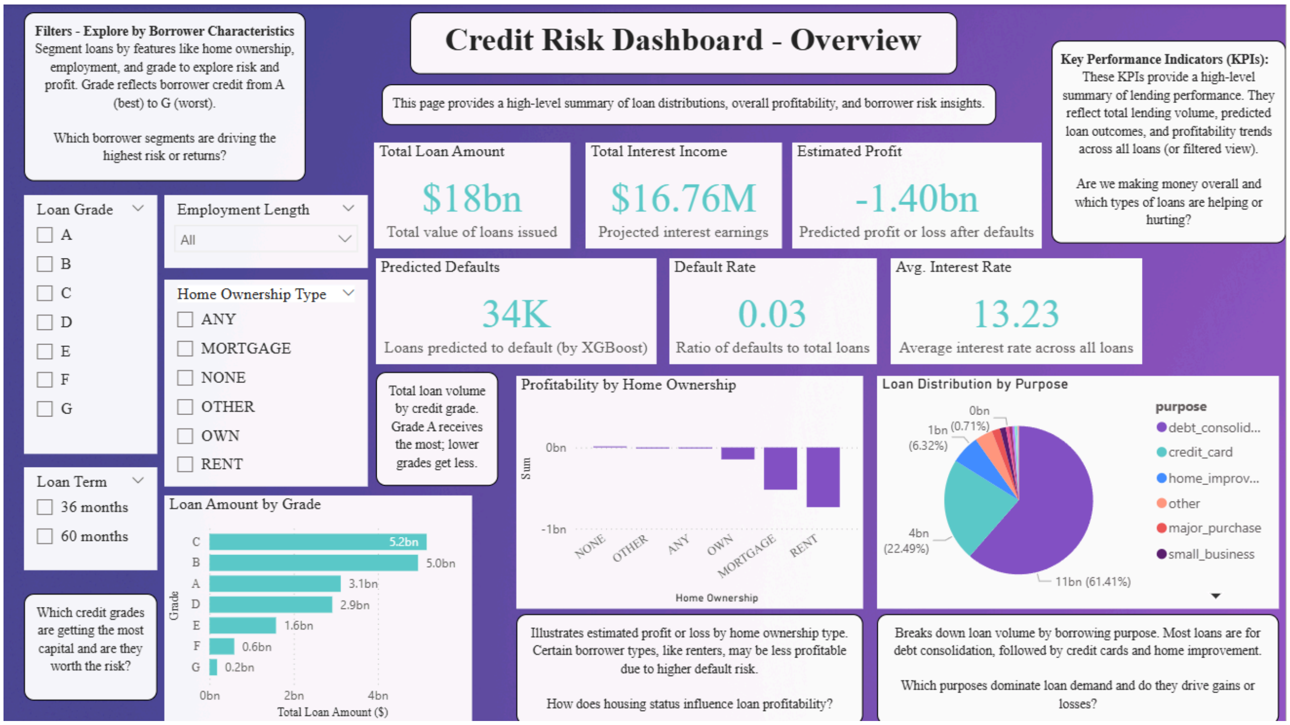
Future Recommendations

In future iterations, incorporating more recent and diverse datasets would enhance model robustness and relevance. Exploring SHAP values could offer deeper interpretability into model predictions. Adding variables like FICO scores, debt-to-income ratios, and external economic factors would also provide a more comprehensive risk analysis.

Reflections on Challenges

The project presented multiple challenges. Missing values required careful consideration to avoid skewing results. The dataset exhibited class imbalance, making accuracy alone a misleading metric, AUC and MAE were used instead. Hyperparameter tuning for models took several iterations to avoid overfitting. In Power BI, visual limitations such as tooltip interactivity and space management required strategic layout planning and creative formatting.

Screenshots of Dashboard



Credit Risk Dashboard - Deeper Insights & Model Evaluation

Filters - Explore by Borrower Characteristics

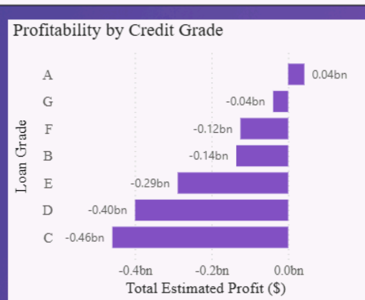
This page explores how borrower characteristics impact profit and risk, and compares model performance to evaluate prediction accuracy.

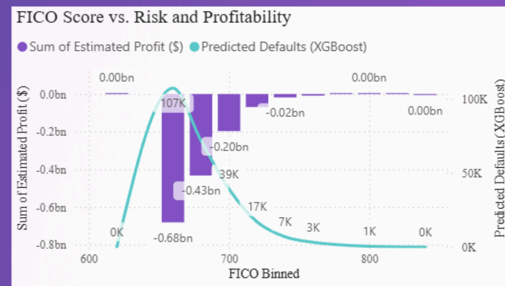
Loan Grade
- ☐ A
- ☐ B
- ☐ C
- ☐ D
- ☐ E
- ☐ F
- ☐ G

Employment Length
All

Home Ownership Type
- ☐ ANY
- ☐ MORTGAGE
- ☐ NONE
- ☐ OTHER
- ☐ OWN
- ☐ RENT

Loan Term
- ☐ 36 months
- ☐ 60 months

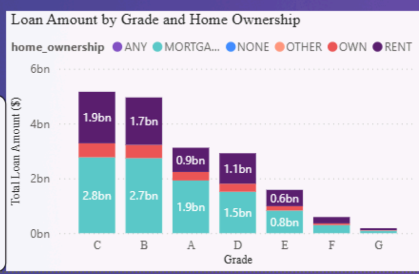




Borrowers with lower FICO scores (<680) received the most loans but had higher predicted defaults and lower profits. Higher FICO borrowers were underrepresented but yielded better profitability.

Are we lending too much to high-risk borrowers, and missing out on safer, more profitable ones?

Compares loan volume by both credit grade and home ownership type. Some lower-grade borrowers (like Grade C) still receive high loan amounts, especially those with mortgages or no homeownership.



Lower credit grades (like C and D) show the biggest losses, while higher grades (A and B) are closer to breaking even or profitable.

How does borrower grade impact profitability?

Which borrower segments are receiving the most capital and are they also the most risky?

Model Performance Comparison

Model	Accuracy	MAE
RF	0.95	0.15
XGB	0.81	0.28
LR	0.81	0.29
Total	**2.57**	**0.72**

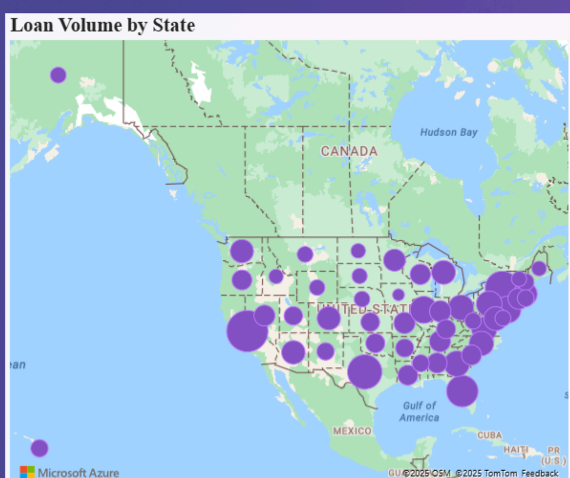
Lower MAE and higher accuracy suggest better prediction quality. This table helps identify the most reliable model.

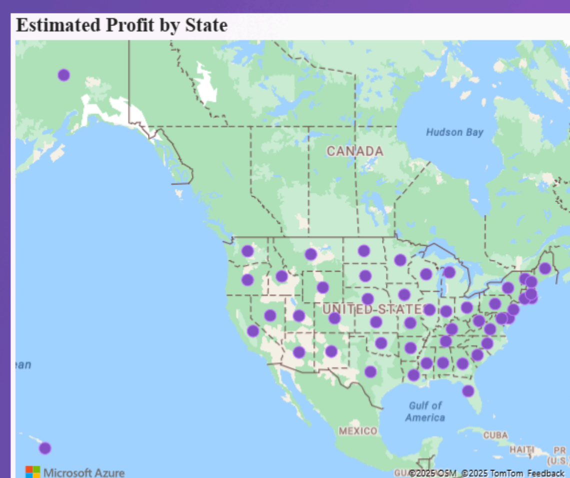
Which model performs best at predicting loan defaults?

Credit Risk Dashboard - Maps

This page highlights how loan volume and profitability vary by state.

Are we focusing too heavily on high-volume but low-profit states?





Larger loan volumes are concentrated in populous states like California, Texas, and New York.

Some states with high loan volume still yield poor profit performance, indicating potential risk concentration.