# Executive Report: LendSmart Credit Risk Analysis

Daniel Alejandro López Martínez
A01770442
Diego Colín Reyes
A01666354
Eduardo Ramírez Almanza
A01660118

November 2025

## 1. Introduction

LendSmart is a FinTech company focused on providing personal and small business loans. In recent months, the management team noted that the default rate had reached 28%, representing a significant financial risk. The goal of this project was to analyze historical customer data and build a model capable of predicting whether a new applicant is likely to repay their loan or default.

To achieve this, two statistical classification methods were compared: Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA). The objective was to identify which method offers the best balance of accuracy, simplicity, and usefulness for credit decision-making.

## 2. Exploratory Analysis and Key Findings

The initial data exploration revealed clear patterns related to customers' payment behavior. Applicants with a strong payment history, stable employment, and higher credit scores were significantly more likely to repay their loans. In contrast, those with high credit utilization or a high debt-to-income ratio showed a considerably higher probability of default.

Personal factors also played an important role. Married clients and those with higher levels of education had lower default rates, while single or widowed customers were more likely to miss payments. These trends align with real-world financial behavior, where economic and social stability often translate into more reliable payment habits.

The correlation analysis showed strong relationships among several financial variables, especially between *payment_history_score*, *credit_utilization*, and *job_stability_score*. These strong correlations contributed to the models' perfect performance, although such results are atypical in real-world datasets.

## 3. Model Evaluation and Results

Both models were trained on standardized data and evaluated using a separate test set. The results were remarkably strong: LDA and QDA both achieved 100% accuracy. Overall, the models correctly identified all 367 customers who repaid and all 133 who defaulted, without a single misclassification.

As requested, special attention was given to false negatives—cases where a risky applicant is mistakenly approved. In this analysis, both LDA and QDA produced zero false negatives. Although this outcome is ideal from a business perspective, it is likely attributable to the academic nature of the dataset, which contains variables that are strongly predictive of the final loan outcome.

## 4. Conclusions

This project demonstrated that discriminant analysis techniques can be powerful tools for evaluating credit risk. Although the dataset produced perfect results, the process made it possible to clearly identify the financial characteristics most associated with default. The most influential variables were payment history, employment stability, credit utilization, and the debt-to-income ratio.

Since both methods achieved identical performance, LDA is recommended because it is simpler and easier to interpret. Its structure allows credit analysts to understand how each variable contributes to the final outcome, improving transparency in loan evaluations.

In conclusion, the LDA model provides a solid foundation for strengthening LendSmart's credit decision process. Its use can help reduce losses from unpaid loans, improve the company's profitability, and make the approval process both more efficient and fairer. For future iterations, it is recommended to test the model with larger and more realistic datasets to validate its performance under real-world conditions.