Credit Card Default Risk Analysis Report

Executive Summary

This project develops a complete end-to-end credit risk analytics pipeline using Python. The goal is to predict 12-month credit card default, analyze behavioral and financial risk factors, model expected losses, and optimize portfolio profitability through PD cutoff strategies.

Data Description

The dataset contains 1,500+ simulated customer credit profiles. Key variables include age, income, credit limit, utilization rate, late payments, delinquency history, months on book, and the target variable (default_12m).

Methods

The workflow includes data cleaning, feature engineering (utilization, exposure, LGD), exploratory data analysis, logistic regression, random forest modeling, ROC analysis, expected loss calculations, and PD cutoff optimization.

Exploratory Data Analysis

EDA revealed higher default rates among customers with high utilization, repeated late payments, and prior delinquency. Utilization rate and delinquency history displayed strong correlation with default behavior.

Modeling

Both Logistic Regression and Random Forest were used. Random Forest achieved stronger performance with higher AUC. Models generated probability of default estimates (PD) for every account.

Risk Analysis

Expected Loss (EL) was computed using PD × LGD × Exposure. Profit modeling accounted for annual interest revenue offset by potential losses. Risk metrics were evaluated by score bands and features influencing credit behavior.

Optimization

A PD cutoff sweep from 1% to 30% identified the approval threshold that maximizes portfolio profit. The optimal cutoff balances approval volume, expected loss, and revenue generation.

Recommendations

Adopt the optimal PD cutoff as the credit approval threshold. Monitor model drift, retrain quarterly, and integrate additional behavioral/time■series features for more stable predictions. Implement risk segmentation strategies to improve portfolio stability.

Limitations

The dataset is simulated and lacks macroeconomic variables. Additional real-world behavioral trends, transaction patterns, or bureau data would strengthen predictive power.

Conclusion

The project demonstrates a full banking■style credit risk workflow. Predictive modeling, risk quantification, and profit■based optimization come together to support data■driven lending decisions.