
Small Business Loan Analysis Report

Prepared by: Afifi Yasser Afifi

Project Overview

:This project consists of two main Jupyter notebooks

Data Cleaning Notebook: Handles missing values, data normalization, and feature .1
.engineering

Data Analysis Notebook: Performs statistical and categorical analysis to identify .2
.trends, patterns, and risk indicators

Data Cleaning Summary

Rows containing missing values in columns with fewer than 5,000 missing entries •
.were removed to eliminate minor gaps without reducing data size significantly

The NA/CS column was partially imputed based on company names. Remaining •
.missing values were completed using the **KNN Imputer**

The *ChgOffDate* column was removed because its meaning overlaps with •
.MIS_Status and ChgOffPrinGr

:Additional derived columns were created to enhance analysis •

(**Region** (grouped U.S. states into Northeast, Midwest, South, and West) ○

(**ApprovalMonth** (derived from ApprovalDate ○

Emp_Category, **Area_Type**, **Is_Franchise**, **Disbursement_Category**, and ○
.other engineered features

Key Analysis Findings

Regional Performance .1

The *Northeast* region had the lowest default rate (20.36%), while the *South* had the highest (25.77%).

Recommendation: Strengthen the role of local banks in southern states by providing easier financing programs for small businesses

State-Level Loan Behavior .2

Analysis of *BankState* and *MIS_Status* showed that defaults are distributed across most states, but **AK, AL, CA, DE, FL, ID, KS, IL, NC, PR, RI, SC, and VA** have higher-than-average default rates.

Recommendation: Improve repayment facilitation systems and borrower monitoring in these states

(Industry Analysis (NAICS .3

About 200,000 records had NAICS = 0, mostly unclassified sectors. The highest default rates were in industries such as **111422, 112310, 112320, 112920, 112990, 113310**, and **0** — mainly agriculture, aquaculture, and forestry.

Recommendation: Conduct deeper credit risk assessments for agricultural sectors and provide tailored financial support

Temporal Analysis .4

Loan approval month (*ApprovalMonth*) showed no significant effect on loan outcome. Fiscal year (*ApprovalFY*) analysis indicated high default rates from **1972–1982**, a sharp decline around **1997–1998**, then a gradual increase until **2006–2008**, likely due to economic fluctuations

Loan Term .5

Loans with terms under 7 years showed high default rates, which decreased as loan duration increased. However, the relationship is not strictly linear.

Recommendation: Reassess short-term loan policies and adjust repayment periods for small businesses

(Business Size (Emp_Category .6

:Default rates decrease as company size increases

Small (1–10 employees): 19.9% •

Medium (11–50 employees): 11.1% •

Large (>50 employees): 6.3% •

Interpretation: Larger businesses exhibit greater financial stability and repayment reliability

Franchise Status .7

Franchise businesses show a lower default rate (12.6%) than independent ones (34.3%).

Recommendation: Encourage small independent businesses to adopt franchising or partnership models to reduce risk

Urban vs. Rural .8

Rural loans have lower default rates (18.8%) than urban ones (24.5%).

Recommendation: Investigate what factors in rural financing models contribute to higher repayment stability

(Revolving Credit (RevLineCr .9

Borrowers with revolving credit lines show higher default rates (26.8%) than those without (14.7%).

Recommendation: Review the risk exposure of borrowers relying heavily on revolving credit facilities

(Low Documentation Loans (LowDoc .10

LowDoc loans (Y) show better repayment rates (~91%) compared to regular loans (N) at 18.8%.

Recommendation: Streamline small-scale loan approval processes for reliable applicants while keeping oversight mechanisms

Loan Amount Analysis .11

Disbursement, Gross Approved, and SBA Approved Amounts all show the same pattern: smaller loans have higher default rates (24–29%) than larger loans (~9–10%). •

Interpretation: Smaller loans likely go to higher-risk borrowers or less stable businesses

General Recommendations

- .Strengthen local bank support programs for small businesses .1
 - .Implement specialized monitoring for high-risk sectors and states .2
 - .Develop risk-based pricing models reflecting company size and industry type .3
 - Provide training and guidance to small independent businesses to improve financial .4 literacy
 - .Enhance data completeness and standardization for NAICS and LowDoc attributes .5
-

Prepared by:
Afifi Yasser Afifi
Data Scientist