DefaultShield: Business Loan Default Prediction

By: Dhevina Tailor

Addressing Loan Defaults with Data Science

Loan defaults occur when people or businesses can't repay borrowed money, causing financial challenges.

We see an opportunity to use data science to predict and manage loan defaults effectively.

Our mission is to make borrowing and lending money safer and more efficient for everyone.

Vision: Making Lending Safer with Data Science

Our vision is to use data science to predict loan defaults by analyzing relevant data.

Factors like a borrower's history, financial situation, and industry trends will guide the analysis.

Aim to build a system that reduces lending risks and increases efficiency for banks.

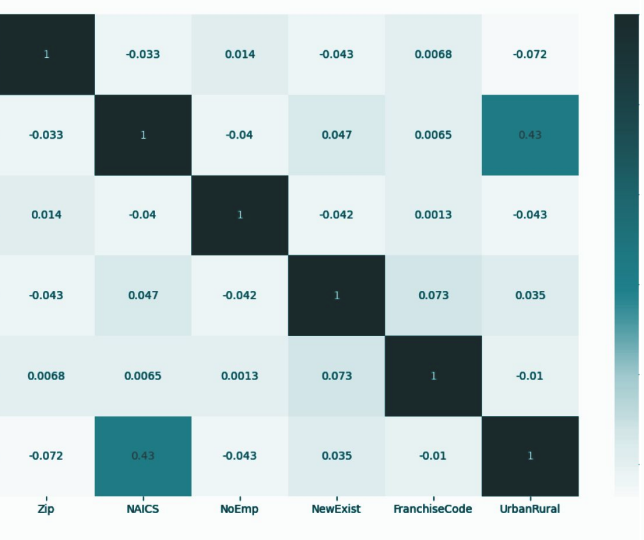
Create a financial world where loans are safer and more accessible for all.

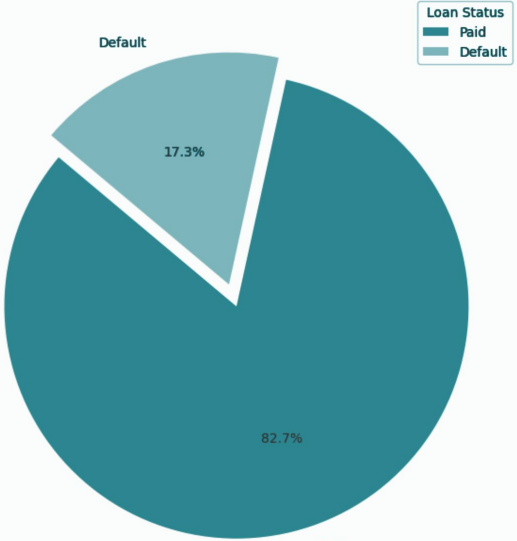
Estimating the Potential Impact

First: For Lending Institutions

Second: For Borrowers

Last But Not Least: For the Whole Economy





6	Bank	Bank name
7	BankState	Bank state
8	NAICS	North American industry classification system code
9	ApprovalDate	Date SBA commitment issued
10	ApprovalFY	Fiscal year of commitment
11	Term	Loan term in months
12	NoEmp	Number of business employees
13	NewExist	1 = Existing business, 2 = New business
14	CreateJob	Number of jobs created
15	RetainedJob	Number of jobs retained
16	FranchiseCode	Franchise code, (00000 or 00001) = No franchise
17	UrbanRural	1 = Urban, 2 = rural, 0 = undefined
18	RevLineCr	Revolving line of credit: Y = Yes, N = No
19	LowDoc	LowDoc Loan Program: Y = Yes, N = No
20	ChgOffDate	The date when a loan is declared to be in default
21	DisbursementDate	Disbursement date

Exploring the Data: Quality and Early Insights

Moving Forward: Data Enhancement and Baseline Modeling

Data Processing: Address data quality concerns, including cleaning, imputing missing values, and handling outliers.

Feature Engineering: Create new features and modify existing ones to capture important relationships in the data.

Baseline Modeling: Begin with a baseline model, typically a logistic regression model, to establish a starting point for predictions.

Thank You!