

# Automate the business loan approval system for a central bank using historical data of borrowers

## About the data

The dataset is from the U.S. Small Business Administration (SBA)

The U.S. SBA was founded in 1953 on the principle of promoting and assisting small enterprises in the U.S. credit market (SBA Overview and History, US Small Business Administration (2015)). Small businesses have been a primary source of job creation in the United States; therefore, fostering small business formation and growth has social benefits by creating job opportunities and reducing unemployment.

There have been many success stories of start-ups receiving SBA loan guarantees such as FedEx and Apple Computer. However, there have also been stories of small businesses and/or start-ups that have defaulted on their SBA-guaranteed loans.

## Content

Shape of the data: 899164 rows and 27 columns

## Data Dictionary

Variable Name	Description
LoanNr_ChkDgt	Identifier Primary key
Name	Borrower name
City	Borrower city
State	Borrower state
Zip	Borrower zip code
Bank	Bank name
BankState	Bank state

Variable Name	Description
NAICS	North American industry classification system code
ApprovalDate	Date SBA commitment issued
ApprovalFY	Fiscal year of commitment
Term	Loan term in months
NoEmp	Number of business employees
NewExist	1 = Existing business, 2 = New business
CreateJob	Number of jobs created
RetainedJob	Number of jobs retained
FranchiseCode	Franchise code, (00000 or 00001) = No franchise
UrbanRural	1 = Urban, 2 = rural, 0 = undefined
RevLineCr	Revolving line of credit: Y = Yes, N = No
LowDoc	LowDoc Loan Program: Y = Yes, N = No
ChgOffDate	The date when a loan is declared to be in default
DisbursementDate	Disbursement date
DisbursementGross	Amount disbursed

Variable Name	Description
BalanceGross	Gross amount outstanding
MIS_Status	Loan status charged off = CHGOFF, Paid in full =PIF
ChgOffPrinGr	Charged-off amount
GrAppv	Gross amount of loan approved by bank
SBA_Appv	SBA's guaranteed amount of approved loan

**Description of the first two digits of NAICS.**

Sector	Description
11	Agriculture, forestry, fishing and hunting
21	Mining, quarrying, and oil and gas extraction
22	Utilities
23	Construction
31–33	Manufacturing
42	Wholesale trade
44–45	Retail trade
48–49	Transportation and warehousing

Sector	Description
51	Information
52	Finance and insurance
53	Real estate and rental and leasing
54	Professional, scientific, and technical services
55	Management of companies and enterprises
56	Administrative and support and waste management and remediation services
61	Educational services
62	Health care and social assistance
71	Arts, entertainment, and recreation
72	Accommodation and food services
81	Other services (except public administration) 92 Public administration

## Review report of selected metrics

### 1. Mis\_status

The MIS\_Status column in the sbaNational table represents the loan status and whether it has been paid in full or paid off. This column plays an important role in evaluating the success or failure of a loan. The two main values in this column are "P I F" (paid in full) and "CHGOFF" (debited). "P I F" status means that the loan has been fully repaid by the borrower with successful and satisfactory results. On the other hand, the "CHGOFF" status indicates that the loan has been written off. H. The lender has declared it a loss due to the borrower's non-payment or default. Monitoring and analysing the distribution of loan status provides valuable insight into borrower repayment behaviour, creditworthiness and overall performance. By examining the ratio of paid-off loans to loans written off, lenders and analysts can assess the quality of their loan portfolios, identify potential risk factors, and make informed decisions regarding loan approvals and loan management strategies. You can do.

### 2. Gr\_apprv

The Gr\_Apprv column in the sbaNational table represents the approved gross loan amount for each loan. This column is an essential metric in evaluating the size and extent of financial support provided to borrowers. It indicates the initial amount approved by the lender before any deductions or adjustments. The values in this column reflect the financial commitment made by the lender towards supporting the borrowers' business or project. Analysing the distribution and trends of the loan amounts can provide insights into the lending patterns, preferences, and risk appetite of the lending institution. Higher loan amounts may indicate a greater level of confidence in the borrower's creditworthiness or the significance of the project. Conversely, smaller loan amounts could signify more cautious lending practices or smaller-scale ventures. Understanding the distribution of approved loan amounts can help lenders assess their lending strategies, identify target markets, and ensure appropriate allocation of resources to meet the financial needs of borrowers.

### 3. Chgoff\_pringr

The ChgOff\_PrinGr column in the sbaNational table represents the charged-off principal amount for each loan. This column captures the portion of the loan principal that has been declared as a loss by the lender. A charge-off occurs when the lender determines that the borrower is unlikely to repay the loan and writes off the outstanding balance as a loss. The values in this column reflect the extent of the

financial loss incurred by the lender due to non-payment or default by the borrower. Analyzing the distribution and trends of the charged-off principal amounts can provide insights into the credit quality of the loan portfolio, the effectiveness of the lender's risk management practices, and the overall financial health of the lending institution. Higher values in this column indicate larger losses incurred by the lender, while lower values suggest more successful loan recoveries or lower default rates. By monitoring the charged-off principal amounts and comparing them to the total loan portfolio, lenders can assess the effectiveness of their underwriting standards, collections procedures, and loan monitoring practices to mitigate credit risk and maintain a healthy loan portfolio.

#### 4. **Lowdoc**

The **lowdoc** column in the **sbaNational** table represents whether the loan was processed under the Low Documentation Program. The Low Documentation Program, also known as "lowdoc," is a loan program designed to simplify the application process for small businesses by reducing the amount of documentation required. In the **lowdoc** column, a value of "Y" indicates that the loan was processed under the Low Documentation Program, while a value of "N" suggests that the loan was not processed under this program. Analysing the distribution of **lowdoc** values can provide insights into the credit profile and risk assessment of small business borrowers. Loans processed under the lowdoc program typically require less documentation, which may indicate a streamlined evaluation process for borrowers with limited financial records or less extensive paperwork. However, it is important to note that the availability of the lowdoc program does not necessarily imply a borrower's creditworthiness or loan performance. Lenders and analysts can consider the **lowdoc** column as a factor in evaluating the loan portfolio and assessing the effectiveness of the low documentation program in facilitating access to credit for small businesses.

Thank You

Aseem Pasha