

sba.csv: SBA Loan Data

699,153 observations, 28 variables

Summary:

The United States Small Business Administration (SBA) was founded in 1953 on the principle of promoting and assisting small enterprises in the US credit market (SBA Overview and History, <https://www.sba.gov/about-sba/what-we-do/history>). To encourage banks to provide loans to small businesses, the SBA has a loan guarantee program. SBA acts much like an insurance provider by reducing the risk for a bank by taking on some of the risk through guaranteeing a portion of the loan. In the case that a loan goes into default, the SBA then covers the amount they guaranteed. Despite this guarantee, banks are often faced with a difficult choice as to whether or not they should grant such a loan because of the high risk of default. One way to inform their decision making is through analyzing relevant historical data such as the data set provided here.

While there have been many success stories of start-ups receiving SBA loan guarantees such as FedEx and Apple Computer, there have also been stories of small businesses and/or start-ups that have defaulted on their SBA-guaranteed loans. The rate of default on these loans has been a source of controversy for decades. Conservative economists believe that credit markets perform efficiently without government participation and want to eliminate the SBA, while others believe that government should help deserving businesses that would be unable to get a loan without the SBA loan guarantee program.

This data set is from the U.S. Small Business Administration (SBA) and provides historical data from 1987 through 2014. This large data set contains 28 variables and 699,153 observations. Each observation represents a loan that was guaranteed to some degree by the SBA. Included is a variable [MIS_Status] which indicates if the loan was paid-in-full or defaulted/charged-off.

SOURCE: United States Small Business Administration

The data reside in a csv file. A header line contains the name of the variables. There are no missing values. For most of the variables, the description is self-evident and are included in the table below. The variables needing further explanation include: *NewExist*, *LowDoc*, and *MIS_Status*.

NewExist (1 = Existing Business, 2 = New Business): This represents whether the business is an existing business (in existence for more than 2 years) or a new business (in existence for less than or equal to 2 years).

LowDoc (Y = Yes, N = No): In order to process more loans efficiently, a “LowDoc Loan” program was implemented where loans under \$150,000 can be processed using a one-page application. “Yes” indicates loans with a one-page application, and “no” indicates loans with more information attached to the application.

MIS_Status: There are a number of categories for this variable in the data provided by SBA. Rather than consider loans that are currently active, we restrict this variable to just loans that have already defaulted/charged-off (CHGOFF) or have been successfully paid-in-full (PIF).

<i>Variable Name</i>	<i>Data Type</i>	<i>Description of variable</i>
LoanNr_ChkDgt	Text	Identifier -- Primary Key
Name	Text	Borrower Name
City	Text	Borrower City
State	Text	Borrower State
Zip	Text	Borrower Zip Code
Bank	Text	Bank Name
BankState	Text	Bank State
NAICS	Text	North American Industry Classification System code
2DIG_NAICS	Text	First two digits of NAICS Code
ApprovalDate	Date/Time	Date SBA Commitment Issued
ApprovalFY	Text	Fiscal Year of Commitment
Term	Text	Loan term in months
NoEmp	Number	Number of Business Employees
NewExist	Text	1 = Existing Business, 2 = New Business
CreateJob	Number	Number of jobs created
RetainedJob	Number	Number of jobs retained
FranchiseCode	Text	Franchise Code 0000 = No Franchise, 0001 =
UrbanRural	Text	1= Urban, 2= Rural, o = Undefined
RevLineCr	Text	Revolving Line of Credit : Y = Yes
LowDoc	Text	LowDoc Loan Program: Y = Yes
ChgOffDate	Date/Time	The date when a loan is declared to be in default
DisbursementDate	Date/Time	Disbursement Date
DisbursementGross	Currency	Amount Disbursed
BalanceGross	Currency	Gross amount outstanding
MIS_Status	Text	Loan Status
ChgOffPrinGr	Currency	Charged-off Amount
GrAppv	Currency	Gross Amount of Loan
SBA_Appv	Currency	SBA's Guaranteed Portion of Approved Loan

NAICS (North American Industry Classification System): This is a 2- through 6-digit hierarchical classification system used by Federal statistical agencies in classifying business establishments for the collection, analysis, and presentation of statistical data describing the U.S. economy.

2DIG_NAICS: The first two digits of the NAICS classification. These two digits represent the economic sector. Table 2 below shows the 2-digit sectors and a corresponding description for each sector.

Table 2. Description of the first two digits of NAICS.

<i>Sector</i>	<i>Description</i>
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
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

Variable Name	Data Type	Description of variable
LoanNr_ChkDgt	Text	Identifier -- Primary Key
Name	Text	Borrower Name
City	Text	Borrower City
State	Text	Borrower State
Zip	Text	Borrower Zip Code
Bank	Text	Bank Name
BankState	Text	Bank State
NAICS	Text	North American Industry Classification System code
2DIG_NAICS	Text	First two digits of NAICS Code
ApprovalDate	Date/Time	Date SBA Commitment Issued
ApprovalFY	Text	Fiscal Year of Commitment
Term	Text	Loan term in months
NoEmp	Number	Number of Business Employees
NewExist	Text	1 = Existing Business, 2 = New Business
CreateJob	Number	Number of jobs created
RetainedJob	Number	Number of jobs retained
FranchiseCode	Text	Franchise Code 0000 = No Franchise, 0001 =
UrbanRural	Text	1= Urban, 2= Rural, o = Undefined
RevLineCr	Text	Revolving Line of Credit : Y = Yes
LowDoc	Text	LowDoc Loan Program: Y = Yes
ChgOffDate	Date/Time	The date when a loan is declared to be in default
DisbursementDate	Date/Time	Disbursement Date
DisbursementGross	Currency	Amount Disbursed
BalanceGross	Currency	Gross amount outstanding
MIS_Status	Text	Loan Status
ChgOffPrinGr	Currency	Charged Off Amount
GrAppv	Currency	Gross Amount of Loan
SBA_Appv	Currency	SBA's Guaranteed Portion of Approved Loan