**NAME:** National SBA

**TYPE:** Census

**SIZE:** 899,164 observations, 27 variables

**ARTICLE TITLE:** “Should This Loan Be Approved or Denied?”: A ‘Big’ Data Set with Class Assignment Guidelines

**SOURCE:** United States Small Business Administration

**STORY BEHIND THE DATA:** 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 27 variables and 899,164 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.

**VARIABLE DESCRIPTIONS:** The data reside in a comma-separated values (csv) file. A header line contains the name of the variables.

| **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 |
| ApprovalDate | Date/Time | Date SBA Commitment Issued |
| ApprovalFY | Text | Fiscal Year of Commitment |
| Term | Number | 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 00000 or 00001 = No Franchise |
| UrbanRural | Text | 1= Urban, 2= Rural, 0 = Undefined |
| RevLineCr | Text | Revolving Line of Credit : Y = Yes |
| LowDoc | Text | LowDoc Loan Program: Y = Yes, N = No |
| 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 Approved by Bank |
| SBA_Appv | Currency | SBA’s Guaranteed Amount of Approved Loan |

**PEDAGOGICAL NOTES:** These data provide educators the opportunity to create assignments that are aligned with GAISE's 2016 recommendations. The authors have used the data set to illustrate how logistic regression can be used to classify a loan application as a "lower risk" (approve) or "higher risk" (deny).

**SUBMITTED BY:**
Name: Min Li, Amy Mickel and Stanley Taylor
Affiliation: California State University, Sacramento
Address: College of Business Administration, CSUS, Sacramento, CA 95819-6088, USA
Email: limin@csus.edu mickela@csus.edu satay lor@csus.edu