

# **A LARGE AND RICH DATASET OF U.S. SMALL BUSINESS ADMINISTRATION (SBA)**

**(Analyze by Faroye Olumide )**

## **About Dataset :**

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 Fed Ex 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
NAICS	North American industry classification system code
Approval Date	Date SBA commitment issued

Variable Name	Description
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
Retained Job	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

Variable Name	Description
LowDoc	LowDoc Loan Program: Y = Yes, N = No
ChgOffDate	The date when a loan is declared to be in default
DisbursementDate	Disbursement date
Disbursement Gross	Amount disbursed
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

<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

<i>Sector</i>	<i>Description</i>
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 re- mediation services
61	Educational services
62	Health care and social assistance
71	Arts, entertainment, and recreation
72	Accommodation and food services

## Sector

## Description

81 Other services (except public administration) 92 Public administration

### TASK :

The task is to analyse and make a suggestion if some loans should be approved or Denied?

### PROCESS:

**Loading packages in Excel sheets and cleaning data.**

The COUNT function was used to determine the total number of people who applied for loan in the datasets. =COUNT(B2:B899166) = 899164

A1 fx =COUNT(B1:B899166)										
	A	B	C	D	E	F	G	H	I	J
1	899164	LoanNr_ChkDgt	Name	City	State	Zip	Bank	BankState	NAICS	ApprovalC
2		1000014003	ABC HOBBYCRAFT	EVANSVILLE	IN	47711	FIFTH THIRD BANK	OH	451120	28-Feb-9
3		1000024006	ANDMARK BAR & GRILLE (THI	NEW PARIS	IN	46526	1ST SOURCE BANK	IN	722410	28-Feb-9
4		1000034009	WHITLOCK DDS, TODD M.	BLOOMINGTON	IN	47401	VT COUNTY STATE E	IN	621210	28-Feb-9
5		1000044001	3 BUCKS PAWN & JEWELRY, L	BROKEN ARROW	OK	74012	TLBK & TR CO OF B	OK	0	28-Feb-9
6		1000054004	NASTASIA CONFECTIONS, IN	ORLANDO	FL	32801	RIDA BUS. DEVEL C	FL	0	28-Feb-9
7		1000084002	SCREW MACHINE COMPANY	PLAINVILLE	CT	6062	Y, NATIONAL ASSO	DE	332721	28-Feb-9
8		1000093009	DDLE ATLANTIC SPORTS CO I	UNION	NJ	7083	FARGO BANK NATL	SD	0	02-Jun-8
9		1000094005	WEAVER PRODUCTS	SUMMERFIELD	FL	34491	REGIONS BANK	AL	811118	28-Feb-9
10		1000104006	TURTLE BEACH INN	PORT SAINT JOE	FL	32456	CENTENNIAL BANK	FL	721310	28-Feb-9
11		1000124001	INTEXT BUILDING SYS LLC	GLASTONBURY	CT	6073	STER BANK NATL A	CT	0	28-Feb-9
12		1000134004	MERCIAL TRUCKING MAINTEN	CHARLOTTE	NC	28256	SUNTRUST BANK	GA	811111	28-Feb-9
13		1000144007	PROFESSIONAL ELEVATOR SERVI	CHICAGO	IL	60605	OF AMERICA NATL	OR	235950	28-Feb-9
14		1000146010	CARVEL	APEX	NC	27502	ARNS BK NATL ASS	MN	445299	07-Feb-04
15		1000154010	RICHARD CAFE & BAKERY, INC	SLATERSVILLE	RI	2876	ENS BANK NATL A	RI	0	28-Feb-9
16		1000214004	EBC INVESTMENTS LLC	WINSTON-SALEM	NC	27106	EST PIEDMONT DEV	NC	0	28-Feb-9
17		1000244002	RONMENTAL ROOFING SYST	OKLAHOMA CITY	OK	73112	OF AMERICA NATL	NC	421330	28-Feb-9
18		1000254005	MAMAGEMENT ENTERPRISES	MIDLAND	TX	79701	FARGO BANK NATL	TX	0	28-Feb-9
19		1000264008	AIRFAX COURT SUBWAY, INC	CENTREVILLE	VA	20120	OF AMERICA NATL	MD	0	28-Feb-9
20		1000284003	FANTASTIC SAMS	PLANO	TX	75093	SMALL BUS. FINAN	NY	0	28-Feb-9
21		1000294006	SIR GOONY'S GOLF	KNOXVILLE	TN	37922	ZENS NATIONAL BA	TN	0	28-Feb-9
22		1000324002	ECONOLOGDE OF DUMAS	DUMAS	TX	79029	NESS LOAN CENTER	SC	0	28-Feb-9
23		1000344008	ING ACHIEVERS PRESCHOOL	CORAL SPRINGS	FL	33065	BANESCO USA	FL	624410	28-Feb-9
24		1000394001	NICOLES RESTAURANT	JOHNSTON	RI	2919	OF AMERICA NATL	RI	0	28-Feb-9

The COUNTBLANK function was used to determine the total number of blank cells in the datasets. =COUNTBLANK(B2:Ac899166) = 1650410

A2    fx    =COUNTBLANK(B2:Ac899164)

	A	B	C	D	E	F	G	H	I	J
1		LoanNr_ChkDgt	Name	City	State	Zip	Bank	BankState	NAICS	ApprovalC
2	1650410	1000014003	ABC HOBBYCRAFT	EVANSVILLE	IN	47711	FIFTH THIRD BANK	OH	451120	28-Feb-9
3		1000024006	ANDMARK BAR & GRILLE (TH	NEW PARIS	IN	46526	1ST SOURCE BANK	IN	722410	28-Feb-9
4		1000034009	WHITLOCK DDS, TODD M.	BLOOMINGTON	IN	47401	VT COUNTY STATE E	IN	621210	28-Feb-9
5		1000044001	5 BUCKS PAWN & JEWELRY, L	BROKEN ARROW	OK	74012	TL BK & TR CO OF B	OK	0	28-Feb-9
6		1000054004	NASTASIA CONFECTIONS, IN	ORLANDO	FL	32801	RIDA BUS. DEVEL C	FL	0	28-Feb-9
7		1000084002	SCREW MACHINE COMPANY	PLAINVILLE	CT	6062	Y, NATIONAL ASSO	DE	332721	28-Feb-9
8		1000093009	DDLE ATLANTIC SPORTS CO I	UNION	NJ	7083	FARGO BANK NATL	SD	0	02-Jun-8
9		1000094005	WEAVER PRODUCTS	SUMMERFIELD	FL	34491	REGIONS BANK	AL	811118	28-Feb-9
10		1000104006	TURTLE BEACH INN	PORT SAINT JOE	FL	32456	CENTENNIAL BANK	FL	721310	28-Feb-9
11		1000124001	INTEXT BUILDING SYS LLC	GLASTONBURY	CT	6073	STER BANK NATL A	CT	0	28-Feb-9
12		1000134004	MERCIAL TRUCKING MAINTEN	CHARLOTTE	NC	28256	SUNTRUST BANK	GA	811111	28-Feb-9
13		1000144007	PROFESSIONAL ELEVATOR SERVI	CHICAGO	IL	60605	OF AMERICA NATL	OR	235950	28-Feb-9
14		1000146010	CARVEL	APEX	NC	27502	JARNS BK NATL ASS	MN	445299	07-Feb-0
15		1000154010	RICHARD CAFE & BAKERY, INC	SLATERSVILLE	RI	2876	ENS BANK NATL AS	RI	0	28-Feb-9
16		1000214004	EBC INVESTMENTS LLC	WINSTON-SALEM	NC	27106	EST PIEDMONT DEV	NC	0	28-Feb-9
17		1000244002	RONMENTAL ROOFING SYST	OKLAHOMA CITY	OK	73112	OF AMERICA NATL	NC	421330	28-Feb-9
18		1000254005	MAMAGEMENT ENTERPRISES	MIDLAND	TX	79701	FARGO BANK NATL	TX	0	28-Feb-9
19		1000264008	AIRFAX COURT SUBWAY, INC	CENTREVILLE	VA	20120	OF AMERICA NATL	MD	0	28-Feb-9
20		1000284003	FANTASTIC SAMS	PLANO	TX	75093	SMALL BUS. FINAN	NY	0	28-Feb-9
21		1000294006	SIR GOONY'S GOLF	KNOXVILLE	TN	37922	ZENS NATIONAL BA	TN	0	28-Feb-9
22		1000324002	ECONOLOGDE OF DUMAS	DUMAS	TX	79029	NESS LOAN CENTER	SC	0	28-Feb-9
23		1000344008	ING ACHIEVERS PRESCHOOL	CORAL SPRINGS	FL	33065	BANESCO USA	FL	624410	28-Feb-9
24		1000394001	NICOLES RESTAURANT	JOHNSTON	RI	2919	OF AMERICA NATL	RI	0	28-Feb-9

SBAnational

A1    fx    =COUNTBLANK(C2:C899165)

	A	B	C	D	E	F	G	H	I	J
1		LoanNr_ChkDgt	Name	City	State	Zip	Bank	BankState	NAICS	ApprovalC
2	3	1000014003	ABC HOBBYCRAFT	EVANSVILLE	IN	47711	FIFTH THIRD BANK	OH	451120	28-Feb-9
3		1000024006	ANDMARK BAR & GRILLE (TH	NEW PARIS	IN	46526	1ST SOURCE BANK	IN	722410	28-Feb-9
4		1000034009	WHITLOCK DDS, TODD M.	BLOOMINGTON	IN	47401	VT COUNTY STATE E	IN	621210	28-Feb-9
5		1000044001	5 BUCKS PAWN & JEWELRY, L	BROKEN ARROW	OK	74012	TL BK & TR CO OF B	OK	0	28-Feb-9
6		1000054004	NASTASIA CONFECTIONS, IN	ORLANDO	FL	32801	RIDA BUS. DEVEL C	FL	0	28-Feb-9
7		1000084002	SCREW MACHINE COMPANY	PLAINVILLE	CT	6062	Y, NATIONAL ASSO	DE	332721	28-Feb-9
8		1000093009	DDLE ATLANTIC SPORTS CO I	UNION	NJ	7083	FARGO BANK NATL	SD	0	02-Jun-8
9		1000094005	WEAVER PRODUCTS	SUMMERFIELD	FL	34491	REGIONS BANK	AL	811118	28-Feb-9
10		1000104006	TURTLE BEACH INN	PORT SAINT JOE	FL	32456	CENTENNIAL BANK	FL	721310	28-Feb-9
11		1000124001	INTEXT BUILDING SYS LLC	GLASTONBURY	CT	6073	STER BANK NATL A	CT	0	28-Feb-9
12		1000134004	MERCIAL TRUCKING MAINTEN	CHARLOTTE	NC	28256	SUNTRUST BANK	GA	811111	28-Feb-9
13		1000144007	PROFESSIONAL ELEVATOR SERVI	CHICAGO	IL	60605	OF AMERICA NATL	OR	235950	28-Feb-9
14		1000146010	CARVEL	APEX	NC	27502	JARNS BK NATL ASS	MN	445299	07-Feb-0
15		1000154010	RICHARD CAFE & BAKERY, INC	SLATERSVILLE	RI	2876	ENS BANK NATL AS	RI	0	28-Feb-9
16		1000214004	EBC INVESTMENTS LLC	WINSTON-SALEM	NC	27106	EST PIEDMONT DEV	NC	0	28-Feb-9
17		1000244002	RONMENTAL ROOFING SYST	OKLAHOMA CITY	OK	73112	OF AMERICA NATL	NC	421330	28-Feb-9
18		1000254005	MAMAGEMENT ENTERPRISES	MIDLAND	TX	79701	FARGO BANK NATL	TX	0	28-Feb-9
19		1000264008	AIRFAX COURT SUBWAY, INC	CENTREVILLE	VA	20120	OF AMERICA NATL	MD	0	28-Feb-9
20		1000284003	FANTASTIC SAMS	PLANO	TX	75093	SMALL BUS. FINAN	NY	0	28-Feb-9
21		1000294006	SIR GOONY'S GOLF	KNOXVILLE	TN	37922	ZENS NATIONAL BA	TN	0	28-Feb-9
22		1000324002	ECONOLOGDE OF DUMAS	DUMAS	TX	79029	NESS LOAN CENTER	SC	0	28-Feb-9
23		1000344008	ING ACHIEVERS PRESCHOOL	CORAL SPRINGS	FL	33065	BANESCO USA	FL	624410	28-Feb-9
24		1000394001	NICOLES RESTAURANT	JOHNSTON	RI	2919	OF AMERICA NATL	RI	0	28-Feb-9

The COUNTBLANK function was used to determine the total number of blank cells in the datasets of name. =COUNTBLANK(c2:c899166) =3. this shows three names was not recorded on the datasets.



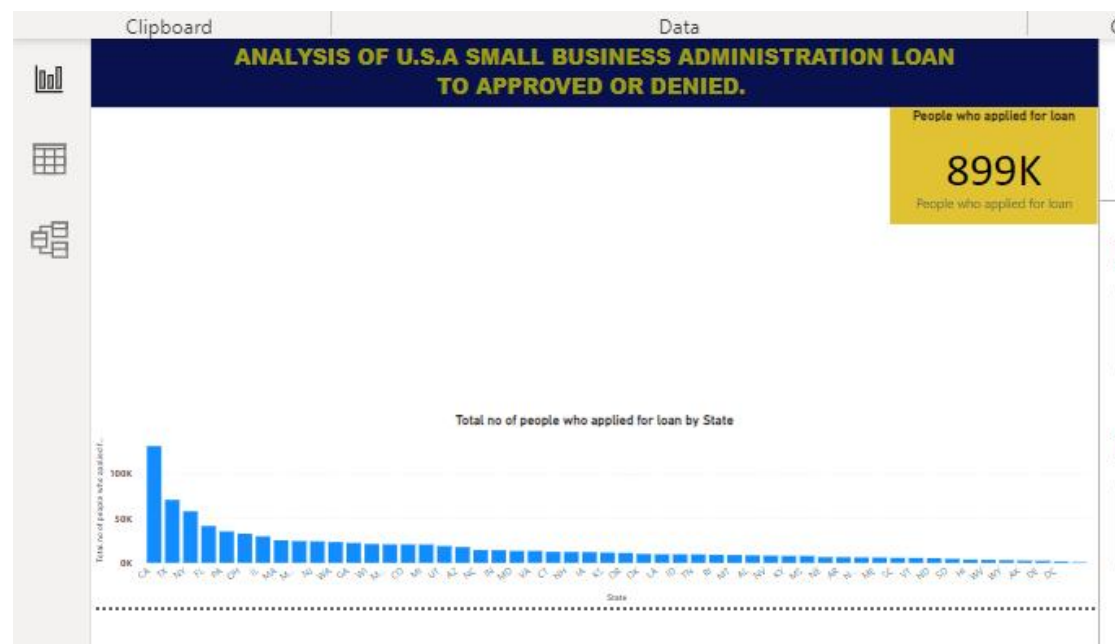
## PROCESS WITH POWERBI:

Loading data packages into PowerBI and Analysing .

Queries [1]

Table.TransformColumnTypes(\*Promoted Headers\*,{{"LoanNr\_ChkDgt", Int64.Type}, {"Name", type text}, {"City", type text}, {"State", type text}, {"Zip", type text}}

	LoanNr_ChkDgt	Name	City	State	Zip
1	1000014003	ABC HOBBYCRAFT	EVANSVILLE	IN	47711
2	1000024006	LANDMARK BAR & GRILLE (THE)	NEW PARIS	IN	46526
3	1000034009	WHITLOCK DDS, TODD M.	BLOOMINGTON	IN	47401
4	1000044001	BIG BUCKS PAWN & JEWELRY, LLC	BROKEN ARROW	OK	74012
5	1000054004	ANASTASIA CONFECTIONS, INC.	ORLANDO	FL	32801
6	1000084002	B&T SCREW MACHINE COMPANY, ...	PLAINVILLE	CT	6062
7	1000093009	MIDDLE ATLANTIC SPORTS CO INC	UNION	NJ	7083
8	1000094005	WEAVER PRODUCTS	SUMMERFIELD	FL	34491
9	1000104006	TURTLE BEACH INN	PORT SAINT JOE	FL	32456
10	1000124001	INTEXT BUILDING SYS LLC	GLASTONBURY	CT	6073
11	1000134004	COMMERCIAL TRUCKING MAINTENANCE	CHARLOTTE	NC	28256
12	1000144007	PROFESSIONAL ELEVATOR SERVICES	CHICAGO	IL	60605
13	1000146010	CARVEL	APEX	NC	27502
14	1000154010	ORCHARD CAFE & BAKERY, INC.	SLATERSVILLE	RI	2876
15	1000214004	EBC INVESTMENTS LLC	WINSTON-SALEM	NC	27106
16	1000244002	ENVIRONMENTAL ROOFING SYSTEMS	OKLAHOMA CITY	OK	73112
17	1000254005	ARK MANAGEMENT ENTERPRISES...	MIDLAND	TX	79701
18	1000264008	FAIRFAX COURT SUBWAY, INC.	CENTREVILLE	VA	20120
19	1000284003	FANTASTIC SAMS	PLANO	TX	75093
20	1000294006	SIR GOONY'S GOLF	KNOXVILLE	TN	37922
21	1000324002	ECONOLODGE OF DUMAS	DUMAS	TX	79029
22	1000344008	YOUNG ACHIEVERS PRESCHOOL, L...	CORAL SPRINGS	FL	33065
23	1000394001	NICOLES RESTAURANT	JOHNSTON	RI	2919



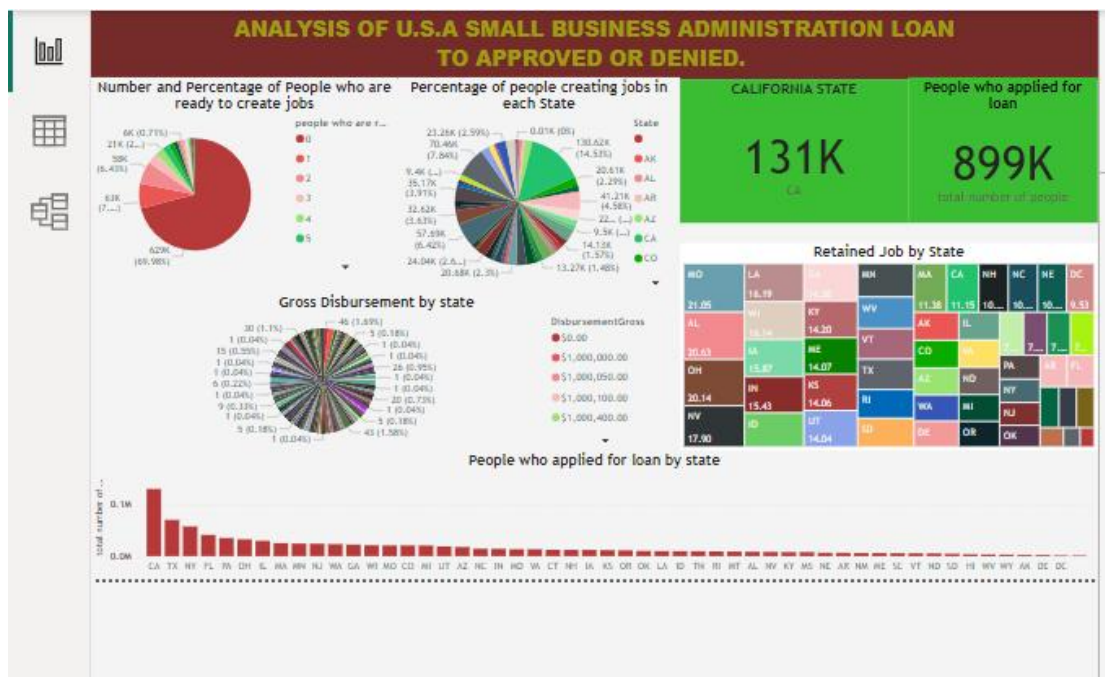
Showing the total number of people who applied for loan which is 899,000 and above and also showing the total number of people who applied for loan by states. With this we can see that CA which is California are the most people who applied with over 100,000 people followed by Texas and others.





Showing the percentage of retained job by each state and the percentage of gross disbursement of loan to states.

## The Final Visualization with PowerBI:



**ACT :**  
**Summary of key findings and recommendations**

(1) Since Small businesses have been a primary source of job creation in the United States I will suggest that those who have the plan to create jobs should be given loans, I mean their loan should be approve.

(2) Those who are retaining their jobs loan should be approve as well.

(3) The percentage of people who are not ready to create jobs with the loan is higher so I will suggest their should be an on-line training with this set of people explaining to them the purpose of giving out this loan and enlighten them on ways to improve their business which will bring about employment of labour.

## **REFERENCES**

Original data set is from the U.S SBA datasets "Should This Loan be Approved or Denied?": A Large Dataset with Class Assignment Guidelines.  
by: Min Li, Amy Mickel & Stanley Taylor

To link to this article: <https://doi.org/10.1080/10691898.2018.1434342>.