



# Review report of **selected metrics**

Presented by: Kratika Jain

# TABLE OF CONTENTS

**01**

## INTRODUCTION

Brief summary about the company.

**02**

## Background and description of dataset

Understanding the columns of the data set

**03**

## WHY

Storytelling

**04**

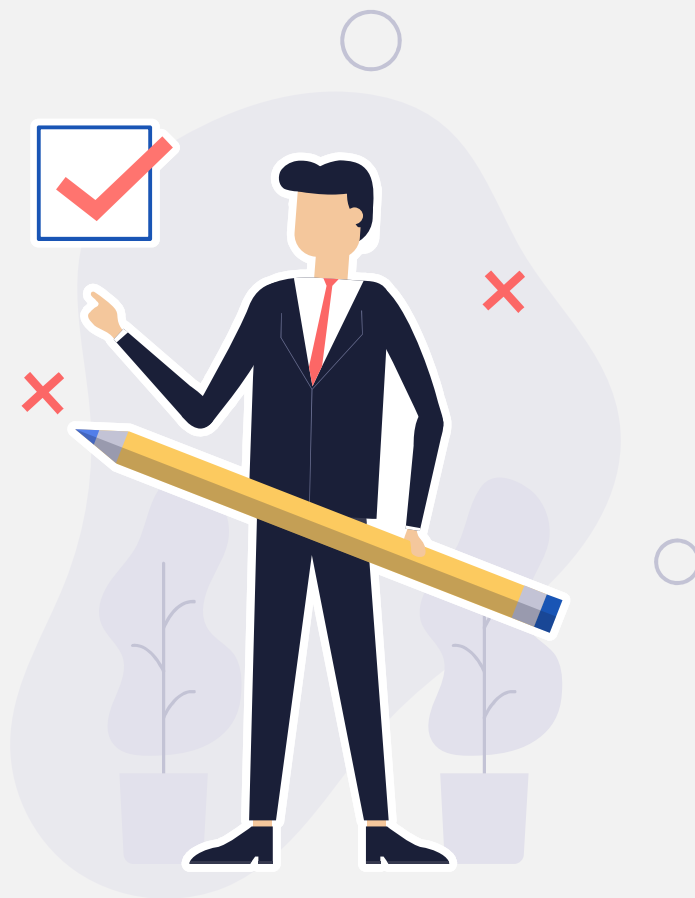
## METRICS

These metrics will decide whether a person's loan should be approved or not

01

# INTRODUCTION

SBA NATIONAL





The Small Business Administration (SBA) is a United States Government Agency formed in 1953 that provides support to entrepreneurs and small businesses. This support comes through loans made through smaller local banks, which are guaranteed up to 90%. In short, local banks give SBA loans to small businesses, and if the loan defaults, the SBA covers up to 90% of the remaining charge off. This helps mitigate risk for the local banks, and helps small businesses get the capital that they need.

02

## Background and description of dataset

Data Dictionary



# THIS IS A TABLE of Data Dictionary

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, N = No
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 charged off = CHGOFF, Paid in full = PIF
ChgOffPrinGr	Currency	Charged-off amount
GrAppv	Currency	Gross amount of loan approved by bank
SBA_Appv	Currency	SBA's guaranteed amount of approved loan

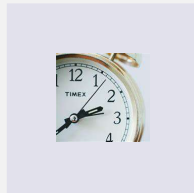
## NEED FOR METRICS ?

Let's assume a scenario in which XYZ person goes to the SBA national and apply for a loan and the bankers check his/her profile before lending him a loan amount so on what metrics they will approve the loan of that person? Therefore, SBA is assuming almost all of the risk in these situations, they have to be very careful in deciding whether or not to approve certain loans for that XYZ person.

# Metrics



**LOCATION**



**TERM**



**NAICS**  
North American industry  
classification code



**NEWEXIST**






Metrics

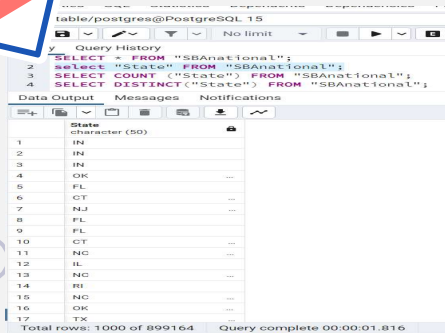
URBAN RURAL

DISBURSEMENT GROSS



# LOCATION (STATE)

It is the primary factor to be considered while lending a loan to the borrower as the state is used as a predictor since different states have different economic environments. 



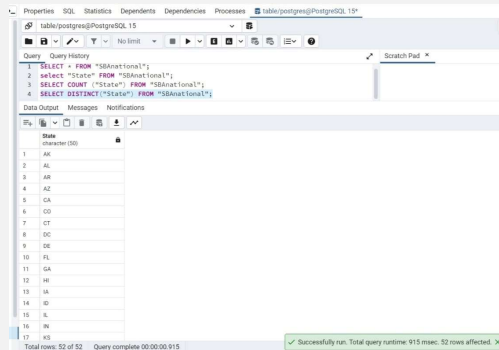
Query History

```
1 SELECT * FROM "SBAnational";  
2 select "State" FROM "SBAnational";  
3 SELECT COUNT ("State") FROM "SBAnational";  
4 SELECT DISTINCT ("State") FROM "SBAnational";
```

Data Output Messages Notifications

State character (50)
1 IN
2 IN
3 IN
4 OK
5 FL
6 CT
7 NJ
8 FL
9 DE
10 CT
11 NC
12 IL
13 NC
14 RI
15 NC
16 OK
17 TX

Total rows: 1000 of 899164 Query complete 00:00:01.816



Query History

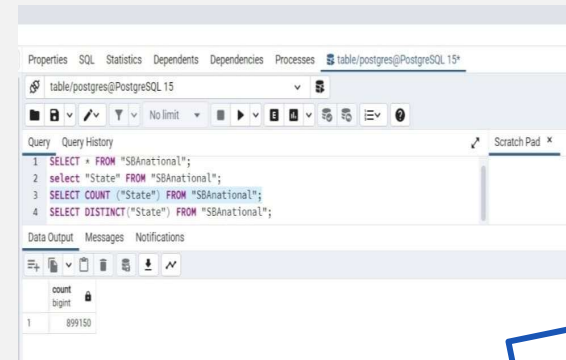
```
1 SELECT * FROM "SBAnational";  
2 select "State" FROM "SBAnational";  
3 SELECT COUNT ("State") FROM "SBAnational";  
4 SELECT DISTINCT ("State") FROM "SBAnational";
```

Data Output Messages Notifications

State character (50)
1 AK
2 AL
3 AR
4 AZ
5 CA
6 CO
7 CT
8 DC
9 DE
10 FL
11 GA
12 HI
13 IA
14 ID
15 IL
16 IN
17 KS

Total rows: 52 of 52 Query complete 00:00:00.915

Successfully run. Total query runtime: 915 msec; 52 rows affected.



Query History

```
1 SELECT * FROM "SBAnational";  
2 select "State" FROM "SBAnational";  
3 SELECT COUNT ("State") FROM "SBAnational";  
4 SELECT DISTINCT ("State") FROM "SBAnational";
```

Data Output Messages Notifications

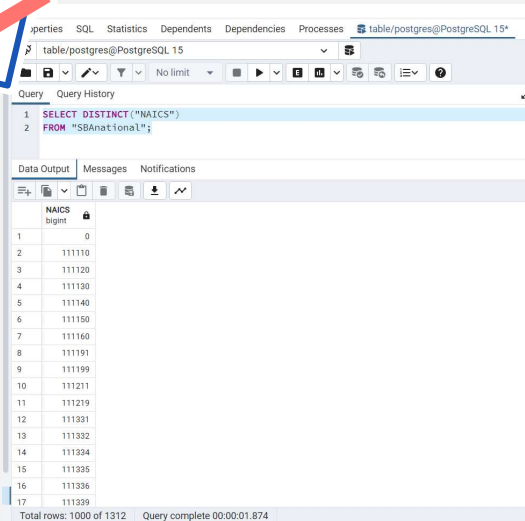

count bigint
1 899150



# NAICS

It is the second metric we can take into account as it is the **North American industry classification code** which depicts that the borrower belongs to which occupation or industry as the first two digits of that number describe the industry for example 11= agriculture, forestry, etc.

This rationale(max number of people from public administration industry borrowed the loan) is confirmed by looking below in the query output:

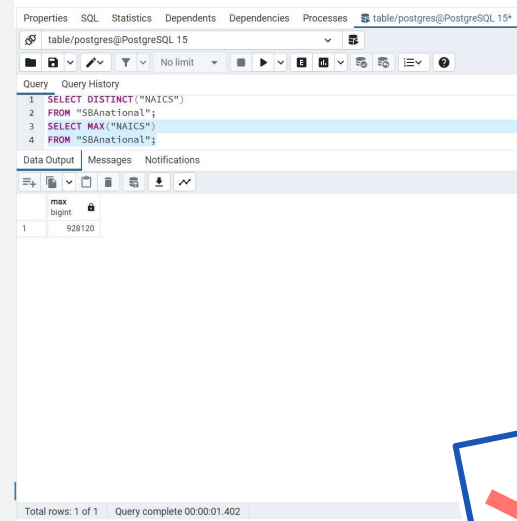



The screenshot shows a PostgreSQL query window with the following query:

```
1 SELECT DISTINCT ("NAICS")
2 FROM "SBAnational";
```

The Data Output tab displays a list of NAICS codes. The first row is 0, and the subsequent rows are 111110, 111120, 111130, 111140, 111150, 111160, 111191, 111199, 111211, 111219, 111331, 111332, 111324, 111335, 111336, and 111339. The total number of rows is 1000 of 1312. The query is complete in 00:00:01.874.

NAICS
0
111110
111120
111130
111140
111150
111160
111191
111199
111211
111219
111331
111332
111324
111335
111336
111339



The screenshot shows a PostgreSQL query window with the following query:

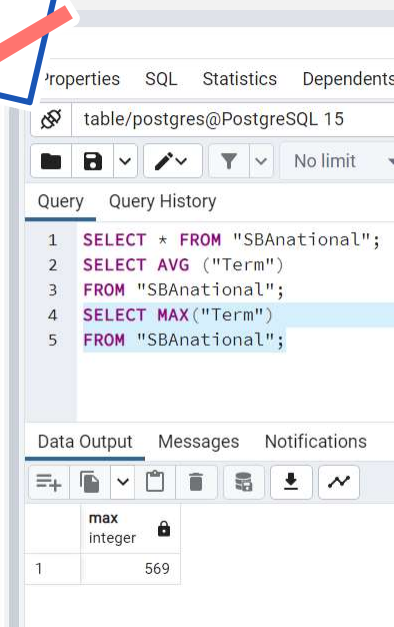

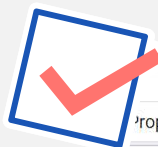
```
1 SELECT DISTINCT ("NAICS")
2 FROM "SBAnational";
3 SELECT MAX ("NAICS")
4 FROM "SBAnational";
```

The Data Output tab displays the maximum NAICS code. The first row is 928120. The total number of rows is 1 of 1. The query is complete in 00:00:01.402.

max
928120

# TERM

Term is another risk indicator that is identify as a key metrics to consider. The rationale behind selecting Term is that it is the loan terms in month called as duration which the borrower took from the bank . This logic can be seen below in the query table:



Properties SQL Statistics Dependents

table/postgres@PostgreSQL 15

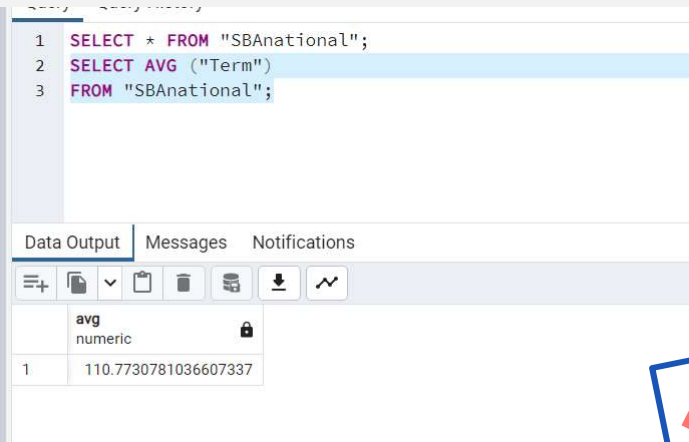

No limit

Query Query History

```
1 SELECT * FROM "SBAnational";
2 SELECT AVG ("Term")
3 FROM "SBAnational";
4 SELECT MAX ("Term")
5 FROM "SBAnational";
```

Data Output Messages Notifications


	max	integer
1	569	



```
1 SELECT * FROM "SBAnational";
2 SELECT AVG ("Term")
3 FROM "SBAnational";
```

Data Output Messages Notifications

	avg	numeric
1	110.7730781036607337	



# URBAN RURAL

It is also the mandatory factor which is to be considered ,since Urban areas are more advanced and they generates employment as well so it can be one of the most apt reason for borrowing a loan from the bank.

1=URBAN

2=RURAL

0=UNDEFINED(Data anomaly as Data outliers)

This condition is confirmed by looking below in the query output:

```
6 SELECT MAX("UrbanRural")
7 FROM "SBAnational";
```

max
integer
1
2

Total rows: 1 of 1 Query complete 00:00:01.427

```
1 SELECT * FROM "SBAnational";
2 SELECT DISTINCT("UrbanRural");
3 FROM "SBAnational";
4 SELECT SUM("UrbanRural");
```

UrbanRural
integer
1
2
2

Total rows: 3 of 3 Query complete 00:00:00.830

```
4 SELECT SUM("UrbanRural")
5 FROM "SBAnational";
6 SELECT MIN("NAICS")
7 FROM "SBAnational";
```

sum
bigint
1
681340

Total rows: 1 of 1 Query complete 00:00:00.910

```
1 SELECT * FROM "SBAnational";
2 SELECT COUNT("UrbanRural")
3 FROM "SBAnational";
```

count
bigint
1
899164

# NEWEXIST

This metrics gives the idea of loans borrowers as they are taking loan for their new venture or existing venture.

As 1= existing business

2=new business

This depicts the risk the security of returning the loan amount.

This logic is confirmed by looking below queries :

```
2 SELECT COUNT("NewExist")
3 FROM "SBAnational";
4 SELECT DISTINCT("NewExist")
5 FROM "SBAnational";
6 SELECT SUM("NewExist")
7 FROM "SBAnational";
8 SELECT MAX("NewExist")
9 FROM "SBAnational";
```

Data Output Messages Notifications

	max integer
1	2

```
1 SELECT * FROM "SBAnational";
2 SELECT COUNT("NewExist")
3 FROM "SBAnational";
4 SELECT DISTINCT("NewExist")
5 FROM "SBAnational";
6 SELECT SUM("NewExist")
7 FROM "SBAnational";
8 SELECT MAX("NewExist")
```

Data Output Messages Notifications

	sum bigint
1	1151119

Query Query History

```
1 SELECT * FROM "SBAnational";
2 SELECT COUNT("NewExist")
3 FROM "SBAnational";
4 SELECT DISTINCT("NewExist")
5 FROM "SBAnational";
6 SELECT SUM("NewExist")
7 FROM "SBAnational";
8 SELECT MAX("NewExist")
```

Data Output Messages Notifications

	NewExist integer
1	0
2	1
3	2
4	[null]

Properties SQL Statistics Dependents

table/postgres@PostgreSQL 15

Query Query History

```
1 SELECT * FROM "SBAnational";
2 SELECT COUNT("NewExist")
3 FROM "SBAnational";
4 SELECT DISTINCT("NewExist")
5 FROM "SBAnational";
6 SELECT SUM("NewExist")
7 FROM "SBAnational";
8 SELECT MAX("NewExist")
```

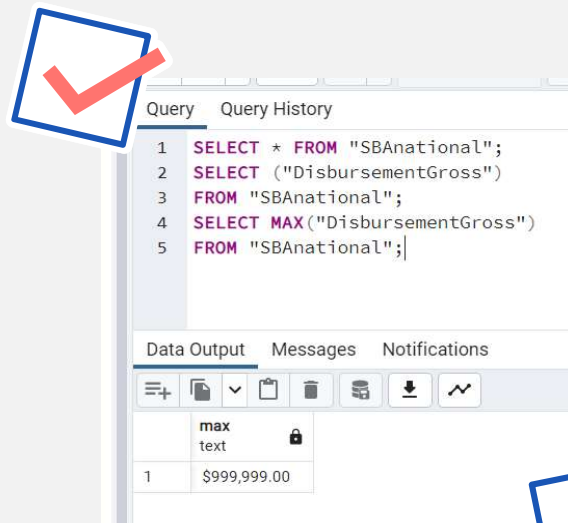
Data Output Messages Notifications

	count bigint
1	899028

## Disbursement Gross

Disbursement Gross is another risk indicator that is identify as a key metrics to consider.

The rationale behind selecting “Disbursement Gross” is that the larger the loan size, the more likely the underlying business will be established and expanding (i.e., purchasing assets that have some resale value), thereby increasing the likelihood of paying off the loan.



**THANKYOU**