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

# The dataset is from the U.S  Small Business Administration (SBA). Now in this document we will look in to the problem of “whether to accept or reject the loan” for particular owner. We will look in to various parameters to predict the defaulters and to whom to accept the loan. And why we choose these parameters.

Here we are using the postgresql to predict the data parameters.

First we will create a table and then import the data in to the table created in the database

**Create Table:**

CREATE TABLE Loan(LoanNr\_ChkDgt VARCHAR(50) PRIMARY KEY,Name VARCHAR(50),City VARCHAR(50),State VARCHAR(50),Zip VARCHAR(50),Bank VARCHAR(40),BankState VARCHAR(50),NAICS VARCHAR(50),ApprovalDate DATE,ApprovalFY VARCHAR(4), Term INT,NoEmp INT,NewExist INT,CreateJob INT,RetainedJob INT,FranchiseCode INT,UrbanRural INT,RevLineCr VARCHAR(5),LowDoc VARCHAR(5),ChgOffDate DATE,DisbursementDate DATE,DisbursementGross MONEY,BalanceGross MONEY,MIS\_Status VARCHAR(20),ChgOffPrinGr MONEY,GrAppv MONEY, SBA\_Appv MONEY);

**Copy table:** Copy table from the csv file that is downloaded in to system using the below query

COPY Loan(LoanNr\_ChkDgt,Name,City,State,Zip,Bank,BankState,NAICS,ApprovalDate,ApprovalFY,Term,NoEmp,NewExist,CreateJob,RetainedJob,FranchiseCode,UrbanRural,RevLineCr,LowDoc,ChgOffDate,DisbursementDate,DisbursementGross,BalanceGross,MIS\_Status,ChgOffPrinGr,GrAppv,SBA\_Appv

)

FROM ‘C:\Users\Admin\Documents\Internship\SBAnational.csv’

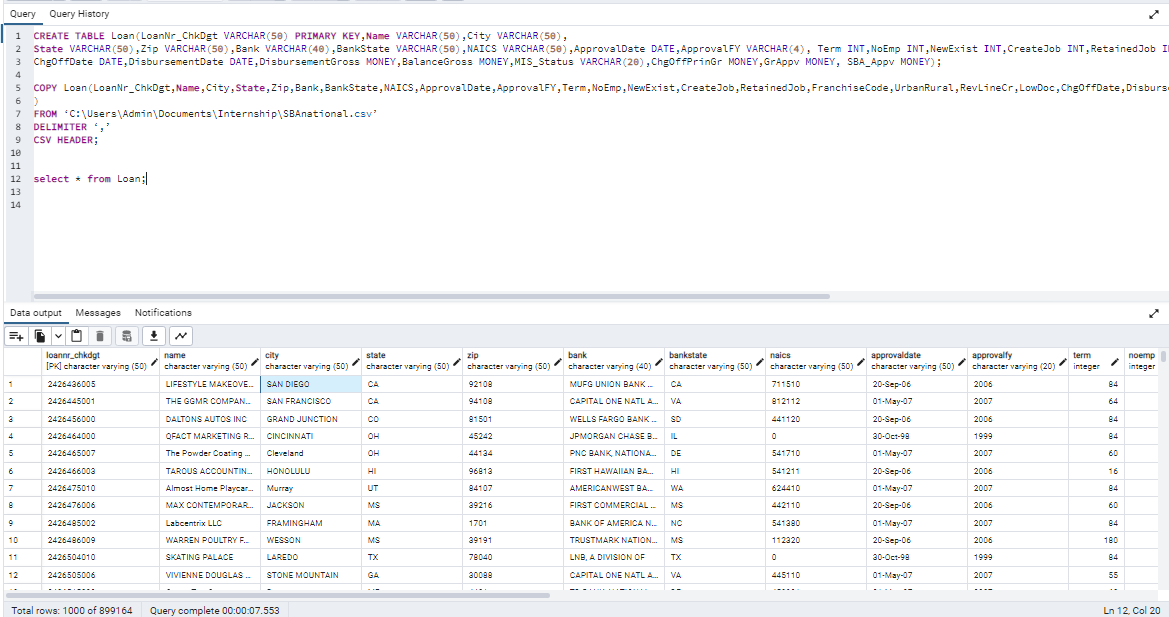
DELIMITER ‘,’

CSV HEADER;

**View table**: View the table using the below query

select \* from Loan;

The table will be displayed



In order to build a data report for loan approval metrics using SQL, we first need to determine which metrics to include in the report. Some possible metrics to consider include:

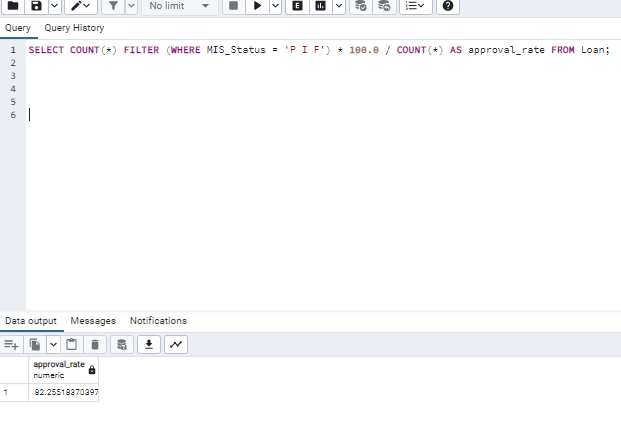
1. Loan approval rate: The percentage of loan applications that are approved.
2. Average loan amount: The average amount of money borrowed by borrowers.
3. Loan default rate: The percentage of loans that go into default.
4. Time to approval: The average time it takes for a loan application to be approved.
5. Loan term: The average length of time for which loans are approved.
6. Loan interest rate: The average interest rate charged on loans.

Once we have determined which metrics to include in the report, we can write SQL queries to calculate each of these metrics. Here are some example SQL queries:

1. Loan approval rate:

SELECT COUNT(\*) FILTER (WHERE MIS\_Status = 'P I F') \* 100.0 / COUNT(\*) AS approval\_rate FROM Loan;

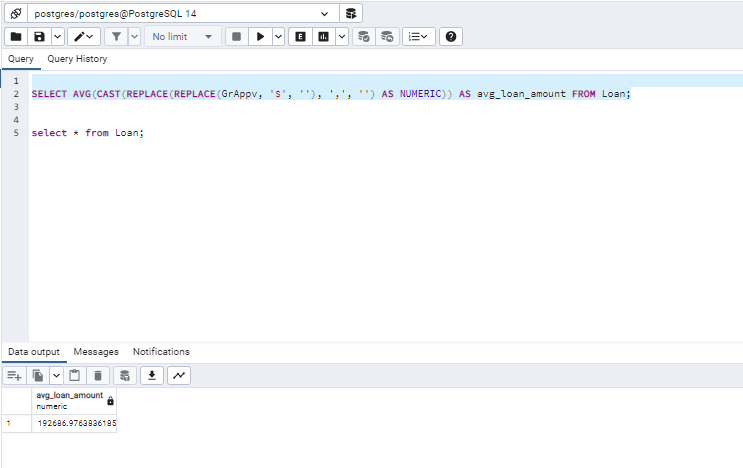
This query calculates the loan approval rate by counting the number of loans that have a status of "P I F" (paid in full) and dividing it by the total number of loans in the "Loan" table.



1. Average loan amount:

SELECT AVG(CAST(REPLACE(REPLACE(GrAppv, '$', ''), ',', '') AS NUMERIC)) AS avg\_loan\_amount FROM Loan;

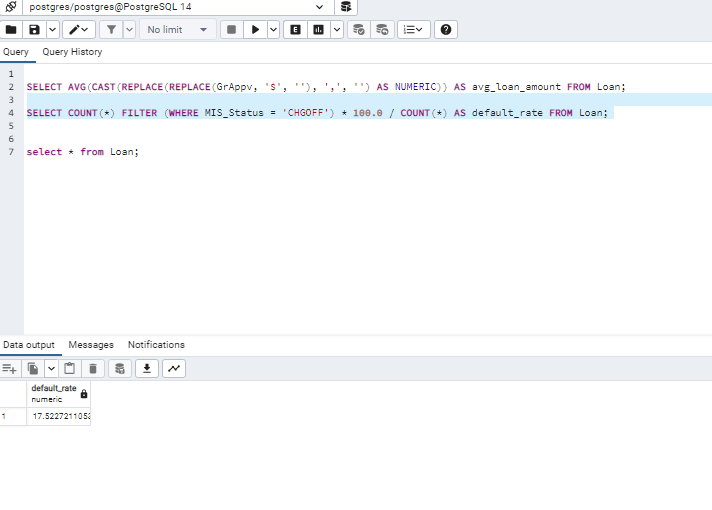
This query calculates the average loan amount by taking the average of the "GrAppv" column in the "Loan" table.



1. Loan default rate:

SELECT COUNT(\*) FILTER (WHERE MIS\_Status = 'CHGOFF') \* 100.0 / COUNT(\*) AS default\_rate FROM Loan;

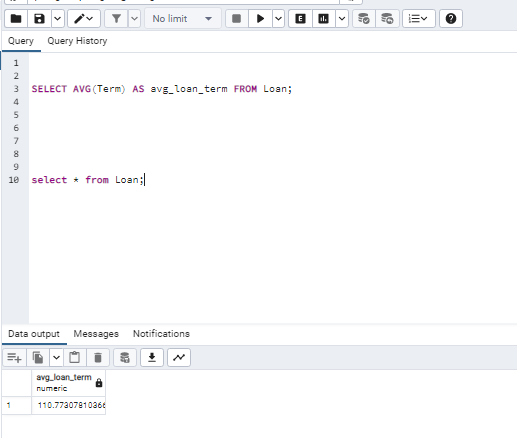
This query calculates the loan default rate by counting the number of loans that have a status of "CHGOFF" (charged off) and dividing it by the total number of loans in the "Loan" table.



1. Loan term:

SELECT AVG(Term) AS avg\_loan\_term FROM Loan;

This query calculates the average loan term by taking the average of the "Term" column in the "Loan" table.

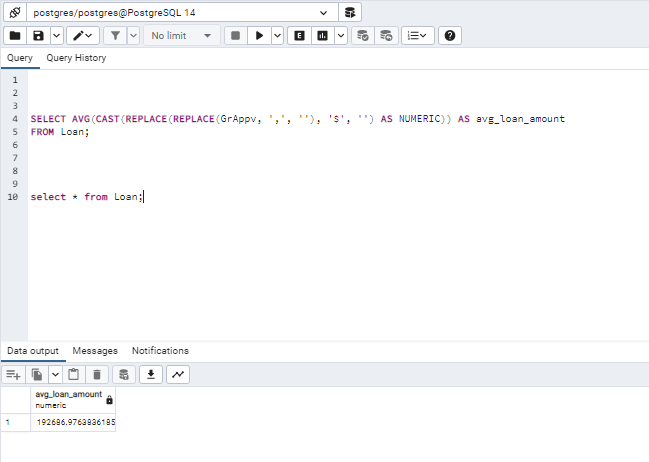


1. Loan interest rate:

SELECT AVG((GrAppv - SBA\_Appv) / SBA\_Appv) AS avg\_interest\_rate FROM Loan;

This query calculates the average interest rate by subtracting the "SBA\_Appv" from the "GrAppv" and dividing the result by "SBA\_Appv". The average of these values is then calculated for all loans in the "Loan" table.

These are just a few examples of the types of metrics that can be included in a loan approval data report. Depending on the specific needs of the report, additional or different metrics may be required.



In conclusion I would choose Average loan amount and Loan default rate as the two metrics to include in the data report for loan approval. Here is the reasoning behind the selection:

1. Average loan amount: This metric is important as it helps us understand the typical loan amount that borrowers are seeking. This information is crucial for banks and lenders as it helps them manage their loan portfolios and ensure that they have enough funds to lend to borrowers. By tracking this metric over time, lenders can also identify trends and changes in borrowing behavior and adjust their lending policies accordingly.
2. Loan default rate: This metric is also important as it measures the percentage of loans that go into default. Defaults are a major concern for lenders as they can result in significant financial losses. By tracking this metric, lenders can identify high-risk borrowers and adjust their lending policies to reduce the likelihood of defaults. They can also use this metric to monitor the performance of their loan portfolio and take action to address any issues that may arise.

Overall, these two metrics provide a good balance of information for lenders to make informed decisions about loan approvals. While loan approval rate is also an important metric, it can be influenced by a variety of factors outside of the lender's control (such as changes in the overall economy). Therefore, it may not be as useful for tracking the performance of the lender's loan portfolio over time. Similarly, while time to approval, loan term, and loan interest rate are all important metrics to consider, they may be less critical for lenders to track on a regular basis compared to loan amount and default rate.

**Relation between two metrics:**

There could be several possible relationships between the average loan amount and loan default rate. It's important to note that correlation does not necessarily imply causation, and there could be other factors at play.

One possibility is that a higher average loan amount could lead to a higher default rate. This could happen if borrowers are taking out larger loans than they can afford to repay, leading to a higher likelihood of default. Additionally, larger loans may be riskier for lenders to approve, leading to a higher default rate overall.

However, it's also possible that there is no significant relationship between the average loan amount and loan default rate. Other factors, such as borrower creditworthiness or economic conditions, could have a bigger impact on default rates than the size of the loans themselves.

Ultimately, further analysis would be needed to determine the specific relationship between these two metrics in a given dataset.