

Banks Loan



**Project Contributors:**

* Anjali Narala
* Kalyan Chander
* Anil Kumar
* Dibyajyoti Rath

# **INTRODUCTION**

The aim of the project is to predict the given case assigned. To achieve this, we split our project into different stage and assigned this to our team-members for fast execution. We followed scrum methodology to share the updates.

In this project we are using snowflake as out main tool to predict few queries. We are using cleaned data as assigned. The data we used in this project is taken from Kaggle site where they are using data to predict defaulter. They are collecting data by a questioner. We got data of 148647 rows. Which is a large chunk of data, while predicting huge amount of data the process may take longer so we split data accordingly into small chunks.

To split data into small chunks we are using python in google colab which is our supporting platform for execution. In this stage we load full data and split it into equal partition, in which each patriation is of 37162 rows. So, the data is slitted into 4 different files. After we achieved to get data, we need to load data into AWS s3 bucket.

In this stage we integrate our snowflake account with external AWS account. To achieve this, we created a s3 bucket in AWS, created a policy and connected it with our s3 bucket using ARN code. After this we created a policy and edited it with few functions which the platform can understand and make the process easy. After creating the policy, we created a role and connected our policy to the respective roles. Now our bucket is connected to our respective policy and roles. When our s3 bucket is completely created we directly load the split data into it.

Now in this stage we moved to snowflake platform and created a database, stage and a table. After creating required we loaded data into our stage from our s3 and called it into a table where we can perform our predictions.

# **LITERATURE REVIEW**

To achieve this project, we referred through different sites for implementing efficient code. To write code in python we referred through tutorials point, YouTube and google. We executed our python code for splitting data and converting csv file to json by using google colab. We choose google colab to execute our python codes since it is a online compiler in which there is no need to download different software’s and the interface is easy to handle. We used pandas, csv and json libraries to achieve results in python.

To predict data, we use snowflake interface where we referred through snowflake documentation, w3 schools, wiki, YouTube, tutorials point, google, java point and our training reference materials and recordings. We achieved integration of snowflake and AWS by referring throughout class recordings and notes given by our trainer.

# **Project Work Flow**

# **Methodology**

# Collected data set from Kaggle website.

<www.kaggle.com/datasets/qusaybtoush1990/banks-loan>

* Creating s3 bucket at AWS.
* Creating Policies and roles in AWS.
* Creation of Integration object to connect AWS with snowflake.
* Splitted the data-set into 4 chunks using Python.
* Creating the task to do the job done on scheduled time.
* Created an external stage and loaded the data from external stage to table.
* created a stream to perform scd2.
* Performed the given tasks by using SQL commands.
* Applying Column level security for particular columns.
* Did the unit test.
* Push the code into Git-Hub repository.

Git-Hub: - <https://github.com/kalyanchander/Project_Bank_loan>

**CONCLUSION**

From this project we derived different outcomes mentioned below and there is a future scope of predicting data with more accuracy.

1. Calculate the total loan amount for gender = ‘female’ and loan limit = ‘cf’   
2. What is the difference in percentage for the number of loans between different valid genders.  
3. What is the difference in percentage of approval in advance between business and commercial loan.    
4. Is there any lumpsum pay for business loan.   
5. Average credit score for various age groups. 

After performing the given tasks we uploaded the project into git-hub repository

We conclude this report by sharing our knowledge, code and outcomes. We also acknowledge that the report, code and ppt presented are completely done by our efforts referring through different mediums to get familiar with the different platforms and increase our critical thinking and increase our knowledge with different commands used, functions created and libraries used.