**World bank -Loan Case study**

****

| **Data sources:** Flat files , RDB , S3 ,JSON,parquet  Scheduler : Task  Streaming : Streams / Snowpipes  **Visualization :** PowerBI  **Language :** Javascript (Stored procedure) |
| --- |

**Problem Statement :**

The International Bank for Reconstruction and Development ([IBRD](https://www.worldbank.org/en/who-we-are/ibrd)) is the world's largest development cooperative working with the World Bank to service middle or low-income countires. One of the main services is loans for the countries to carry out diverse projects such as construction, energy, road, or education. Historically there have been about 758K cases since the group was formed in 1947 for redevelopment after WWII. Most of these were disbursed and repaid, however, some of them were cancelled. The loan cancellation is a loan that was ready to begin but widthdrawn before disbursement. Considering the scale of loan amounts, cancellations would cause a tremendous waste of time as well as financial cost. For that reason, this project focused on analyzing cancelled loans and building a predictive model to prevent or prepare for such loans.

While there are officially 33 columns, the project took steps to filter important features and cases mostly in the first half. The data dictionary can be found at [World Bank Finance](https://finances.worldbank.org/api/assets/6B259BEE-3B23-4BDF-B37C-0A8D36469060?download=true), which could help to understand the overall dataset structure.

As a snowflake Data Engineer, you are helping the data science team to deal with a volume

amount of data and fast query processing .

**Tasks :**

1. **Create database loan\_db and schema schema**
2. **Create loan history tables and apply cluster keys on most used columns for querying.**
3. **Create internal stage and load the historical data from local.**
4. **Perform basic transformation (cast, change date format) during data load**
5. **Upload all current data into AWS s3 bucket snow\_extstage and perform load**
6. **Load the data from external stage**
7. **Perform bad records filtering on loading ,file format and compression**
8. **Create snowpipes to continuously load new US covid data from external stage from s3 bucket to the staging table ,maintain the change data capture and merge the data to the consumption table**
9. **Share your table to new non-snowflake user**
10. **create clone of the table with time travel before one day and write query to get history data using particular timestamp**
11. **Create stored procedure to insert the data into table after typecasting date format**
12. **Create a visualization for loan repayment analysis using powerBI and prepare reports /dashboard for business queries**

**Business questions :**

1.How many days taken to sign the loan and the other was how many days taken for repayment.

2. Find the top three countries had huge amount of loans and it quickly dropped.

3. Which countries that had at least one cancellation with the assumption that they had borrowed enough number of loan

4. Find the average repayment period for a country

**Datasets ;**

IBRD : historical loan datasets:

<https://finances.worldbank.org/Loans-and-Credits/IBRD-Statement-Of-Loans-Historical-Data/zucq-nrc3>

IBRD latest loan data :

<https://finances.worldbank.org/Loans-and-Credits/IBRD-Statement-of-Loans-Latest-Available-Snapshot/sfv5-tf7p>