

# Overview

## Assumptions:

1. Raw files are in correct format. No need to check.
2. Use of PostgreSQL database
3. Standardized columns along L0 and L1 schema.
4. Power BI dashboard will be set for daily sync for updates

## Interesting datasets:

[Link1](#)  
[Link2](#)

Will create own dataset with dummy values in banking.

## Key Questions:

1. What kind of problem are you looking for?  
Ans. In the area of ~~social media~~ or ~~social networks~~ banking.

2. What kind of models are you looking to apply?  
Ans. ~~Online models so it makes sense to have all the data in the DB.~~ We can try a bunch of ML models.

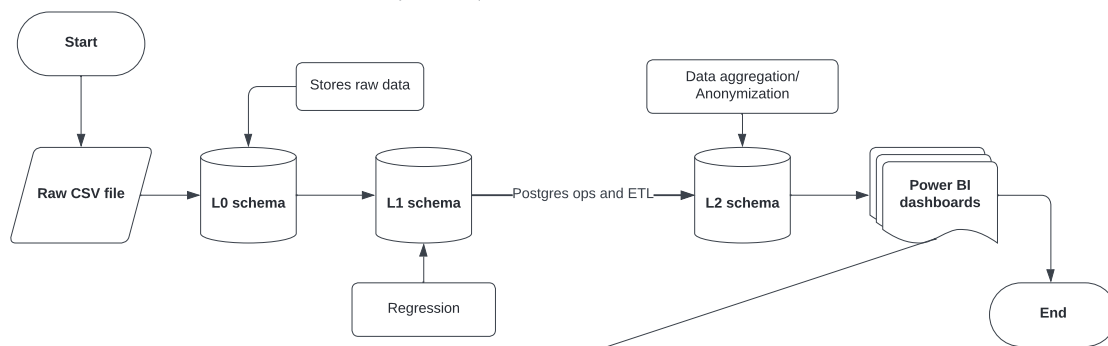
3. How does Postgres functions or ops come into play?  
Ans. We will store customer info in different tables. PG can be used for joins & computations.

## Idea

Data  
Banking data

## Dashboard

1. Credit risk analysis
2. Loan performance
3. NPA
4. Customer Insights
5. Executive summary and trends



**Credit Risk Analysis Dashboard**

1. Credit risk score distribution
2. High Risk customers
3. Delinquency rates
4. Risk score by industry
5. Geographical distribution
6. Risk trend MoM

**Non-Performing Assets Dashboard**

1. YoY trend in NPA held by bank
2. NPA by sector
3. NPA by resolution status
4. NPA by default period
5. Customers responsible for NPA

**Executive Summary Dashboard**

1. Aggregate Metrics
  - a. Total loans
  - b. NPAs
  - c. Revenue
  - d. Profit Margins
2. Default rates
3. Credit exposure
4. Capital Adequacy Ratio
5. Loan breakdown by
  - a. Sectors
  - b. Type
  - c. Location
6. Net interest income
7. Operating expenses

**Slicers**

1. Year
2. Month
3. Regions (States)
4. Customer ID
5. Industry
6. Customer type
7. Credit Type
8. Loan Type
9. Loan Tenure

**Loan Performance Dashboard**

1. Loan by types
2. Loan by amount
3. Default rates by loan types
4. Top borrowers
5. Return on Assets by loan types
6. Loans by tenures

**Customer Insights Dashboard**

1. Breakdown by demographics
2. Breakdown by income levels
3. Breakdown by loan preference
4. Enrolment/dis-enrollment count
5. Credit Utilization ratios
6. Customers with potential up-selling

## L2 Schema

**credit\_risk**

1. customer\_id
2. risk\_score -> Regression
3. industry
4. geography
5. delinquency status
6. month
7. year
8. customer\_type
9. credit\_type

**loan**

1. loan\_id
2. loan\_type
3. loan\_amount
4. ongoing\_interest\_rate
5. borrower\_id [ <----> customer\_id ]
6. loan\_tenure
7. loan\_start\_date
8. loan\_end\_date
9. month
10. year
11. default\_flag
12. risk\_score

**customer**

1. customer\_id
2. age
3. gender
4. income\_level
5. geography
6. month
7. year

**enrollment\_status**

1. month
2. year
3. geography
4. age
5. gender
6. income\_level
7. industry
8. customer\_type
9. count\_of\_new\_enrollment
10. count\_of\_disenrollment

risk\_score -> Regression  
borrower\_id = customer\_id  
customer\_id, month, year -> PK for **customer** table  
one customer can have one type and industry

Assumptions