**BANKING PROJECT**

**OBJECTIF:**

V1 : Analyser les demandes de prêts pour comprendre les facteurs qui influencent l’approbation ou le rejet

V2 : construire un modèle de machine learning pour prédire la probabilité d'acceptation d’un prêt.

V3 : automatisation (de la collecte a l’analyse et visualisation)

**DONNÉE:**

Dataset : <https://www.kaggle.com/datasets/architsharma01/loan-approval-prediction-dataset?utm_source=chatgpt.com>  
  
Fichier: loan\_approval\_dataset.csv

Taille : 385 ko

Lignes : 4269

Colonnes : 13

Dtypes : -------

|  |  |
| --- | --- |
| **Nom de la colonne** | **Signification** |
| **loan\_id** | Identifiant unique du prêt (utilisé pour distinguer chaque ligne dans le dataset). |
| **no\_of\_dependents** | Nombre de personnes à charge du demandeur (enfants, parents, etc.). |
| **education** | Niveau d’éducation du demandeur (ex. : diplômé, non diplômé). |
| **self\_employed** | Indique si le demandeur est travailleur indépendant (Oui ou Non). |
| **income\_annum** | Revenu annuel brut du demandeur (devise locale). |
| **loan\_amount** | Montant total du prêt demandé. |
| **loan\_term** | Durée du prêt, généralement exprimée en mois. |
| **cibil\_score** | Cote de crédit (CIBIL) du demandeur, qui reflète sa solvabilité. Plus le score est élevé, mieux c’est. |
| **residential\_assets\_value** | Valeur des biens résidentiels détenus (maisons, appartements). |
| **commercial\_assets\_value** | Valeur des biens commerciaux détenus (bureaux, locaux professionnels). |
| **luxury\_assets\_value** | Valeur des biens de luxe (voitures de luxe, bijoux, œuvres d’art). |
| **bank\_asset\_value** | Valeur des actifs bancaires (épargne, dépôts à terme, investissements). |
| **loan\_status** | État du prêt (Approuvé, Rejeté, etc.). C’est souvent la variable cible pour un modèle prédictif. |

Debt-to-Income Ratio **= loan\_amount / income\_annum 🡪** Mesure la capacité à gérer ses dettes

Ce ratio mesure **la part du revenu annuel** que représente la dette (prêt demandé).  
Plus il est **élevé**, plus le client consacre une portion importante de son revenu au remboursement → donc plus le risque est **élevé**.

Mesure la capacité à gérer ses dettes

Asset-to-Loan Ratio (ATL) **= total\_assets\_value / loan\_amount 🡪** Mesure la couverture des dettes par les actifs

Ce ratio mesure **la capacité du client à couvrir le montant emprunté** grâce à ses biens (immobiliers, bancaires, de luxe...).

**ENVIRONEMENT ET SUIVI: github**

**To do: organiser le git en dossier**

* **Dossier dataset**
* **Dossier python**
* **Dossier sql**
* **Dossier tableau**

**CONNECTION AU DATASET: via python**

**TRANSFORMATION DES DONNÉES (ETL) : via python**

**Creation des colones :**

**total\_asset\_value** = residential\_assets\_value + commercial\_assets\_value + bank\_asset\_value + luxury\_assets\_value

**asset\_to\_loan\_ratio** = total\_asset\_value / loan\_amount

**risk\_category** : cibil score

Low : < 600

Medium : 600–750

High : > 750

### **is\_low\_income :**

### 1 si income\_annum < 300000, sinon 0

### **Loan\_term\_half\_years:**

Short < 6 mois

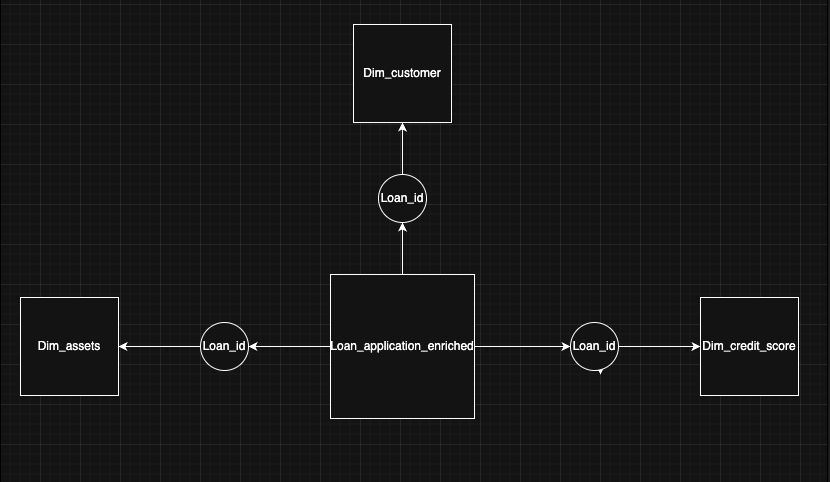
### Medium < 12 mois(1 an)

### Else: long

**CREATION DE MODELE: via SQl**

**To do:**

**MODELE EN ETOILE**

****

**Fact table :** loan\_application\_enriched

Colonnes:

loan\_id

no\_of\_dependents

education

education

loan\_amount

loan\_term

loan\_term\_years

cibil\_score

risk\_category

residential\_assets\_value

commercial\_assets\_value

luxury\_assets\_value

bank\_asset\_value

bank\_asset\_value

asset\_to\_loan\_ratio

has\_dependents

is\_low\_income

loan\_status

loan\_status\_code

**Dim:** dim\_customer

Colonnes:

loan\_id

education

self\_employed

no\_of\_dependents

**Dim:** dim\_assets

Colonnes:

loan\_id

cibil\_score

risk\_category

**Dim:** dim\_credit\_score

Colonnes:

loan\_id

residential\_assets\_value

commercial\_assets\_value

luxury\_assets\_value

bank\_asset\_value

total\_asset\_value

**VISUALISATION : via tableau**

**To do:**

* **Integer les filtres interactifs (education, no\_of depedent)**
* **Modifier le nom des feuilles et des tableaux**
* **Refaire toutes les visualisations en français (cohérence)**
* **Masquer les feuilles et laisser juste les tableaux de visualisations**