**LES DÉTERMINANTS DE LA SOLVABILITÉ DES CLIENTS :**

**Problématique :**

Le poste client est un actif majeur du bilan des banques. Etudier la solvabilité des clients et sélectionner ceux capables d'honorer leurs engagements est essentiel pour [se protéger contre les impayés](https://fiches-pratiques.chefdentreprise.com/Thematique/gestion-finance-1050/Guide/Comment-proteger-contre-impayes-300383.htm), et donc, optimiser sa trésorerie.

La connaissance des déterminants de la solvabilité des clients va permettre de prédire la situation du client et par la suite décider de lui accorder des crédits ou pas.

**Solution :**

Etablir un modèle de classification basé sur une régression logistique permettant de prédire la situation de solvabilité du client.

**Démarche :**

*1-régression logistique :*

La régression logistique est l’un des modèles d’analyse multivariée elle permet de modéliser des variables binaires ou des sommes de variables binaires.

Le principe du modèle de la régression logistique est de relier la survenance ou la non survenance d'un événement au niveau de variables explicatives.

*2- Application :*

**2-1:apreçu**

Pour notre cas d’étude nous avons une base de donnée avec 77 431 observations(soient 77431 clients), pour chaque client nous avons associé 23 variables(7 quantitatives, 16 qualitatives).

Pour pouvoir modéliser la solvabilité avec une régression logistique nous aurons besoin d’expliquer la solvabilité (étant une variable catégorielle ‘solvable’=1 ou ‘non solvable’=0) par des variables quantitatives.

Pour ce faire, nous allons nous contenter de faire un modèle de régression logistique multivariée de 7 variables quantitatives.

**2-2: variables:**

Les variables explicatives choisies:

1-'Age'

2-'Nombre\_Enfant\_Emprunteur'

3-'Revenu\_Emprunteur'

4-'Montant\_prêt'

5-'Mensualité'

6-'Durée\_crédit'

7-'Taux\_interet'

La varible à expliquer ou la variable de classification:

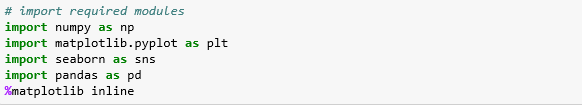
‘SOLVABILITE’

**2-3: Outil utilisé:**

Pour implémenter notre modèle, nous avons utilisé le langage PYTHON.

**2-4: Modélisation**

* Dans un premier temps nous importons les bibliothèques et les librairies nécessaires.

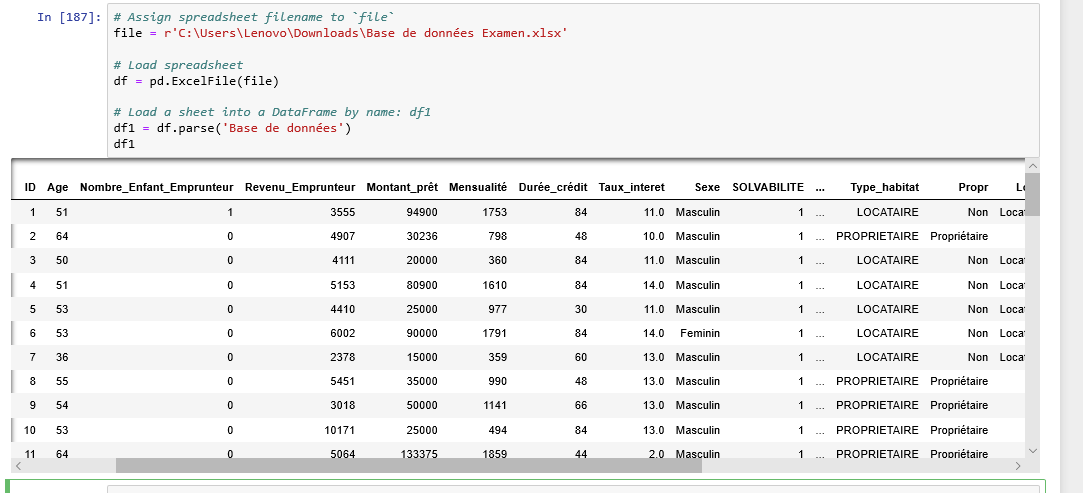


numpy: pour le calcul mathématique

matplotlib: pour dessiner des graphes

seaborn et pandas : pour l’analyse statistique

* Par la suite nous faisons appel à la base de donnée:



* On précise par la suite les variables explicatives et la variables dépendante:

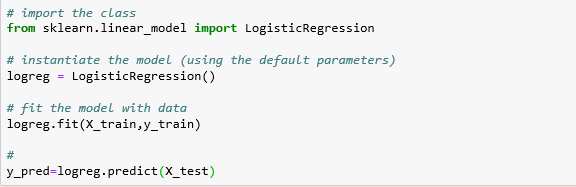


* Ensuite nous divisons nos données en deux parties 70% pour

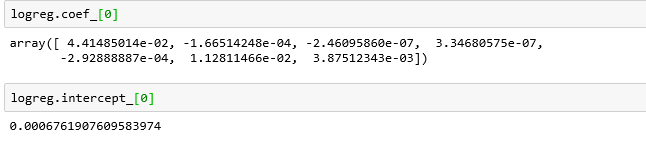
l’entraînement du modèle et 30% pour le test.



* Nous faisons appel à un module de régression logistique prédéfini:



* le résultat de la régression est le suivant:

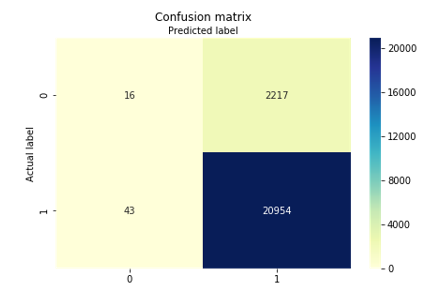


la première commande renvoie l’ensemble des coefficients attachés aux variables et la deuxième renvoie la valeur de la constante.

en arrivant à cette étape notre modèle est constriut, mais reste à le tester pour le valider.

**2-5: Test du modèle:**

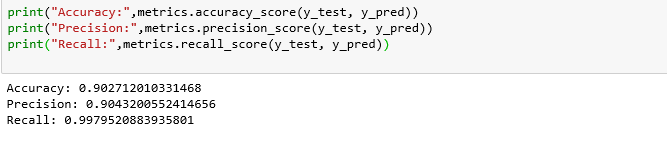
La matrice de confusion est une [matrice](https://fr.wikipedia.org/wiki/Matrice_(math%C3%A9matiques)) qui mesure la qualité d'un système de [classification](https://fr.wikipedia.org/wiki/Classement_automatique). Chaque ligne correspond à une classe réelle, chaque colonne correspond à une classe estimée. La cellule ligne L, colonne C contient le nombre d'éléments de la classe réelle L qui ont été estimés comme appartenant à la classe C.



20954 observations qui sont égales à 1 ont été bien prédites par le modèle des 1.

43 observations étant égales à 1 ont été prédites par le modèle des 0…

* Calculons maintenant le degré de précision et l’accuracy du modèle:



Notre modèle est de 90% de précision ce qui est jugé un bon résultat.

**Conclusion:**

Notre modèle précise que les variables choisies détermine bien la solvabilité, et avec une précision de 90%, notre modèle est capable de prédire si un client est solvable ou pas.