**Objectif :**

* Comprendre du mieux possible nos données (un petit pas en avant vaut mieux qu'un grand pas en arrière)
* Développer une première stratégie de modélisation

**Checklist de base**

**Analyse de Forme :**

* **variable target** : diabetes
* **lignes et colonnes** : 100000 , 9
* **types de variables** : 3 variables qualitatives / variables quantitatives
* **Analyse des valeurs manquantes** :
  + Pas de valeurs manquantes

**Analyse de Fond :**

* **Visualisation de la target** :
  + 8.5% de positifs (8500 / 100000)
* **Signification des variables** :
  + variables continues -> age de 0 à 80 ans ( pic à 80 ans ) / HbA1c\_level / bmi ( imc )/ blood\_glucose\_level
  + variables qualitatives : hypertension / gender / heart\_disease / diabetes / smoking\_history
* **Relation Variables / Target** :
  + HbA1c\_level et blood\_glucose\_level corrélation avec diabètes de 0.4 et 0.42
  + Age 0.26, hypertension 0.2, heart\_disease 0.17, bmi 0.21

**Hypothèse :**

* Diabète lié au blood\_glucose\_level et age
* Hypertension lié à age