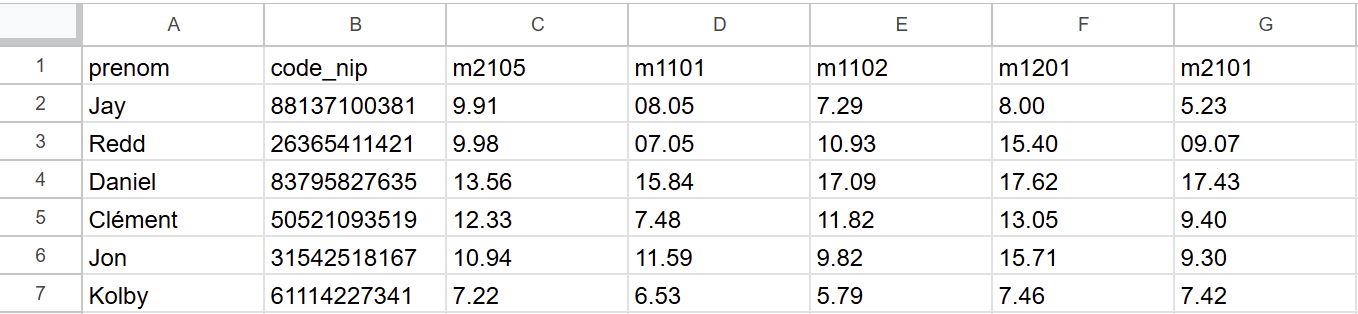
**Rapport Livrable 3**

# I - Présentation des données

## Présentation des données

Le fichier vue\_voyelles.csv contient plusieurs séries statistiques extraites de la base de données des étudiants en DUT Informatique :

* La population étudiée correspond aux étudiants de première année de DUT, identifiés par leur code\_nip.
* La 1ʳᵉ variable est le prénom de l’étudiant.
* La 2ᵉ est la note obtenue au module M2105, qui porte sur les statistiques (au semestre 2). Cette variable est celle qu’on souhaite expliquer.
* Les autres variables sont les notes de modules issus du semestre 1 et début du semestre 2 :
  + m1101 : Algorithmique et programmation
  + m1102 : Mathématiques discrètes
  + m1201 : Projet tutoré
  + m2101 : Système



## Problématique

En utilisant ces données, on va essayer de répondre à la problématique suivante :

*Est- ce qu' avoir une voyelle dans son prénom permet d’avoir une meilleure moyenne.*

# 

# II - Import des données, mise en forme

## Importer les données en Python

Les données sont importées en Python sous forme de DataFrame à l’aide de la commande suivante :

| Voyelle\_df = pd.read\_csv("vue\_voyelles.csv") |
| --- |

## Mise en forme

On a besoin de supprimer les cases vides (qui contiennent nan en Pythons), puis on transforme notre DataFrame en Array :

| Voyelle\_df = Voyelle\_df.dropna() Voyelle\_ar = Voyelle\_df.to\_numpy() |
| --- |

## Normalisation

On conserve uniquement les colonnes contenant des données numériques. Ces données sont ensuite centrées-réduites :

from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

data\_scaled = scaler.fit\_transform(df.select\_dtypes(include=np.number))