**Rapport Livrable 3**

# I - Présentation des données

## Présentation des données

* Population étudiée : Étudiant ayant fourni leur prénom, leur moyenne générale, leur code postal, leur niveau d'études et leur mention au bac.
* Description des variables :
  + *‘prenom’* : Prénom de l’étudiant (ex : Maurice)
  + *‘moyenne’* : Moyenne générale obtenue de l'étudiant (ex : 12.75)
  + *‘code\_postal’* : Code postal de résidence de l’étudiant (ex : 22300)
  + *‘niveau\_etude’* : Niveau d’étude de l’étudiant (ex : Terminale)
  + *‘mention\_bac’* : Mention au bac de l’étudiant (ex : TB)

## Problématique

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

*Existe-t-il une influence du niveau d'études, de la mention obtenue au bac, du code postal et de la première lettre du prénom sur la moyenne générale des étudiant·es ?*

# 

# 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 :

| cheminFichier = "vue.csv" VueDf = pd.read\_csv(cheminFichier, sep=";") |
| --- |

## 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 change toutes les données qui étaient des chaînes de caractère par des int. Pour cela on quantifie les variables.

| VueDf['initiale'] = VueDf['prenom'].str[0]  alphabet = list("ABCDEFGHIJKLMNOPQRSTUVWXYZ") mapping\_initiale = {lettre: i+1 for i, lettre in enumerate(alphabet)} VueDf['initiale\_num'] = VueDf['initiale'].map(mapping\_initiale)  VueDf['mention\_bac\_num'] = VueDf['mention\_bac'].map({  'P': 1, 'AB': 2, 'B': 3, 'TB': 4 })  VueDf['niveau\_etude\_num'] = VueDf['niveau\_etude'].map({  "Terminale": 1, "Année préparatoire aux études supérieures": 2, "1ère année d'études supérieures": 3, "2nd année d'études supérieures": 4 }) |
| --- |

# III - Représentations graphiques

# 

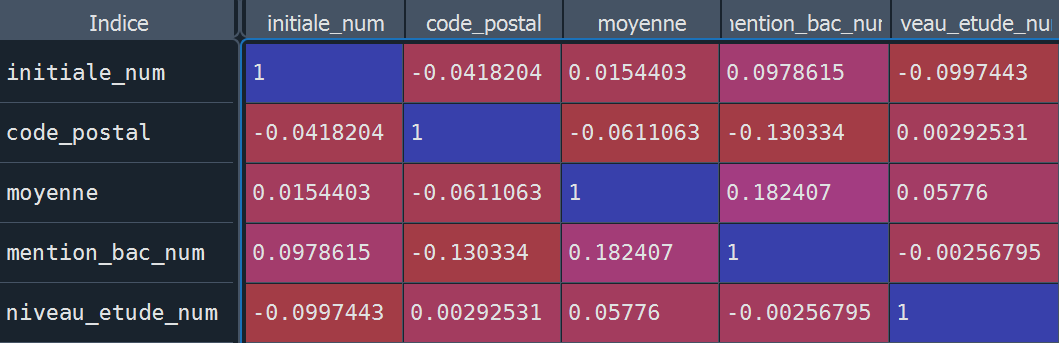
# IV - Matrice des coefficients de corrélation

## Démarche

| df\_corr = VueDf[['initiale\_num', 'code\_postal', 'moyenne', 'mention\_bac\_num', 'niveau\_etude\_num']]  corr\_matrix = df\_corr.corr() |
| --- |

## Matrice des corrélations

On obtient la matrice suivante :

****

# V - Régression linéaire multiple