



# Etude Observationnelle en épidémiologie clinique



## Contexte

Dans ce projet, j'ai enrichi la base de données SIRENE avec des coordonnées géographiques et des détails supplémentaires sur les établissements en utilisant les données de Google Maps.

Objectif était d'enrichir la base de données SIRENE avec des informations géographiques et des détails complémentaires pour améliorer sa précision et sa valeur informative.

Ce projet a été réalisé en binôme sur une période de 2 mois.

## La démarche suivie

### Nettoyage de la base des données

J'ai utilisé l'outil Python et la bibliothèque pandas pour importer, nettoyer, et transformer les données du fichier. L'objectif était d'assurer la qualité et la lisibilité des données avant leur enrichissement.

### Collecte des données

J'ai utilisé l'API pour obtenir des coordonnées géographiques précises. Puis j'ai intégré des coordonnées obtenues dans le DataFrame existant.



### Extraction d'Informations

J'ai employé Selenium et Chromedriver pour automatiser la navigation sur Google Maps.

J'ai ensuite extrait les détails supplémentaires, tels que les avis et les heures d'ouverture des établissements.

### Enrichissement de la Base de Données

Le projet a permis d'ajouter des coordonnées géographiques et des informations web aux entrées existantes, améliorant ainsi la précision et l'utilité de la base de données SIRENE.

## Conclusion



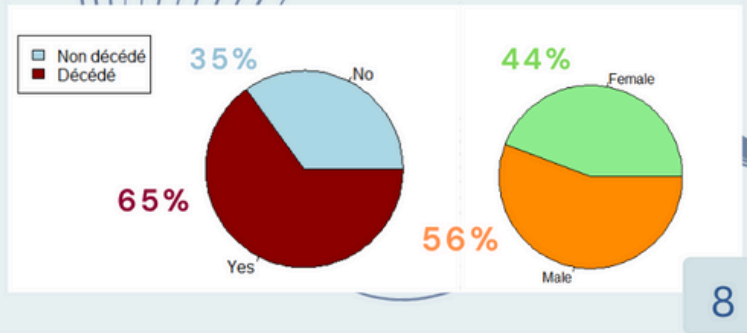
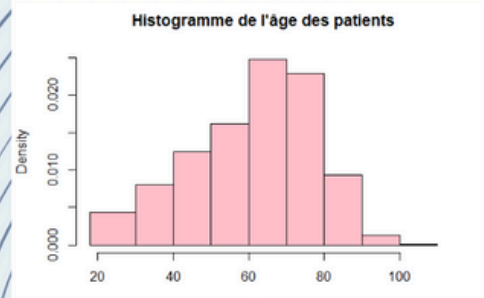
Ce projet a démontré l'importance de l'enrichissement des bases de données pour des applications pratiques. J'ai acquis des compétences en manipulation de données avec Python, en utilisation des API, et en automatisation de la navigation web avec Selenium. Le travail en binôme a également renforcé mes compétences en collaboration et en gestion de projet.

# Exemples

Page de la présentation Orale:

## ANALYSE UNIVARIÉE

- La plupart des variables comme l'âge sont des variables quantitatives continues
  - Les variables "DEATH" et "SEX" sont qualitatives nominale
  - Définition des variables
  - Analyse descriptives de ces variables
1. Minimum
  2. 1er quartile
  3. Médiane
  4. Moyenne
  5. Maximum
  6. Nombre de NA



Extrait du programme avec graphique statistique:

```

146 #Personnes pas decede et decede au bout de 2 mois
147 epi$DECES_2mois = ifelse(epi$Temps_Décès < 62, 1, 0)
148 epi$DECES_2mois = ifelse(is.na(epi$DECES_2mois), 0, epi$DECES_2mois)
149
150 mean(epi$SURVIE À 2 MOIS[epi$DECES_2mois == 1])
151 mean(epi$SURVIE À 2 MOIS[epi$DECES_2mois == 0])
152
153
154 x = c(mean(epi$SURVIE À 2 MOIS[epi$DECES_2mois == 1]),
155       mean(epi$SURVIE À 2 MOIS[epi$DECES_2mois == 0]))
156 barplot(table(epi$SURVIE À 2 MOIS, epi$DECES_2mois))
157 table(x[epi$DECES_2mois])
158 barplot(x, main = "Graphique de la moyenne de la probabilité de survie à 2 mois
159           en fonction de si le patient est décédé ou non",
160         , ylab = "Probabilité moyenne de survie à 2 mois"
161         , xlab = "Etat du patient au bout de 2 mois"
162         , col = c("orange", "red"))
163
164 legend("topleft", legend = c("Non décédé", "Décédé"), fill = c("red", "orange"))
165 table(epi$DECES_2mois)
166
167
168.1 (Untitled)
R Script

```

```

> table(x[epi$DECES_2mois])
0.496593106535363
2234
> barplot(x, main = "Graphique de la moyenne de la probabilité de survie à 2 mois
+           en fonction de si le patient est décédé ou non",
+         , ylab = "Probabilité moyenne de survie à 2 mois"
+         , xlab = "Etat du patient au bout de 2 mois"
+         , col = c("orange", "red"))
+ legend("topleft", legend = c("Non décédé", "Décédé"), fill = c("red", "orange"))
> table(epi$DECES_2mois)
0      1
3274 2234
>

```

