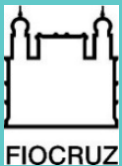




Espace DEV



# Typologie des favelas à partir de variables environnementales multi-sources

Mouad SAMAI

Encadrants : Mr Laurent DEMAGISTRI, Mme Nadine DESSAY et Mr Joris GUERIN

Enseignant référent : Mr Paul PASSY



STAGIAIROTHON MTD - 2025



**HÉTÉROGÉNÉITÉ:** physique, Spatiale, Sociale

## Liste des workpackages

Niveau 1 : Base de connaissance

Niveau 2 : Méthodes de Détection et suivi des favelas

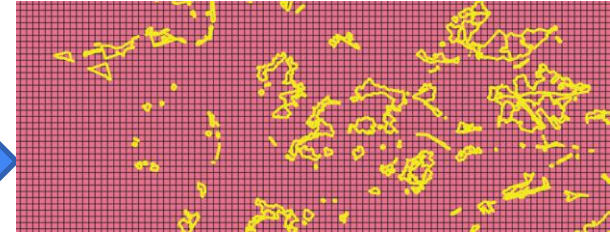
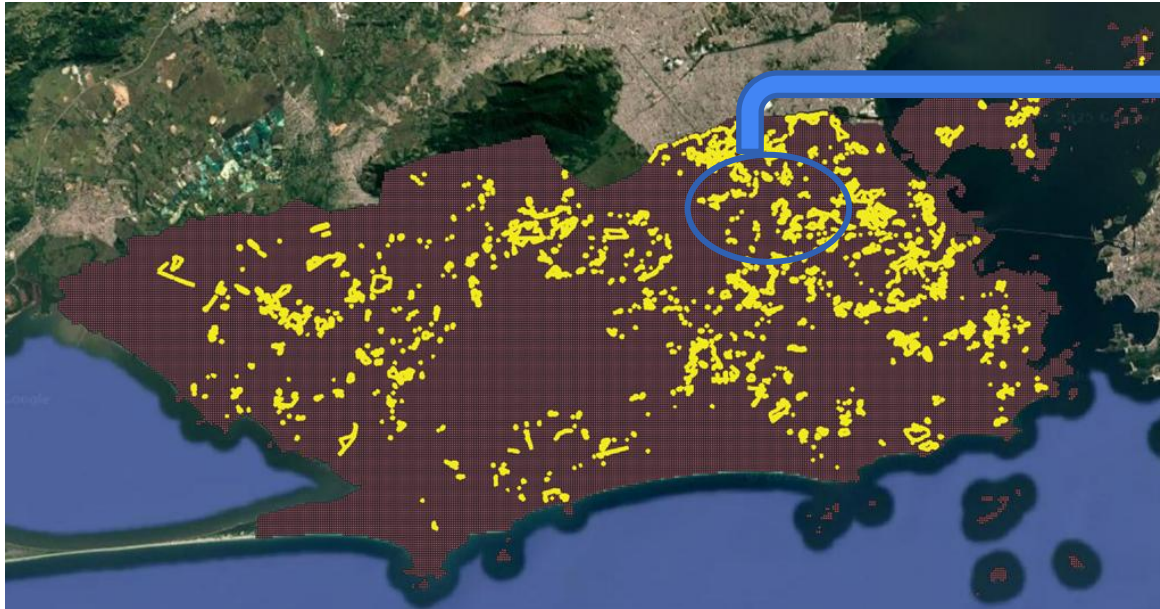
Niveau 3 **Typologie des favelas** (environnement et santé)

**MATHIS (Projet CNES / TOSCA)**  
Méthodes Automatiques de Télédétection  
des Habitats Informels pour la Santé



## Zone d'étude : Rio de Janeiro

Grille globale d'analyse : mailles 150m x 150m



**Polygones de référence -  
FAVELAS**  
(Source IBGE - Institut Brésilien  
de Géographie et de Statistique)

# Variables pour chaque maille

## 1. dérivées de l'**IMAGE SATELLITE** (Sentinel-2) :

- Végétation (NDVI  $\geq$  0.6)
- Entropy

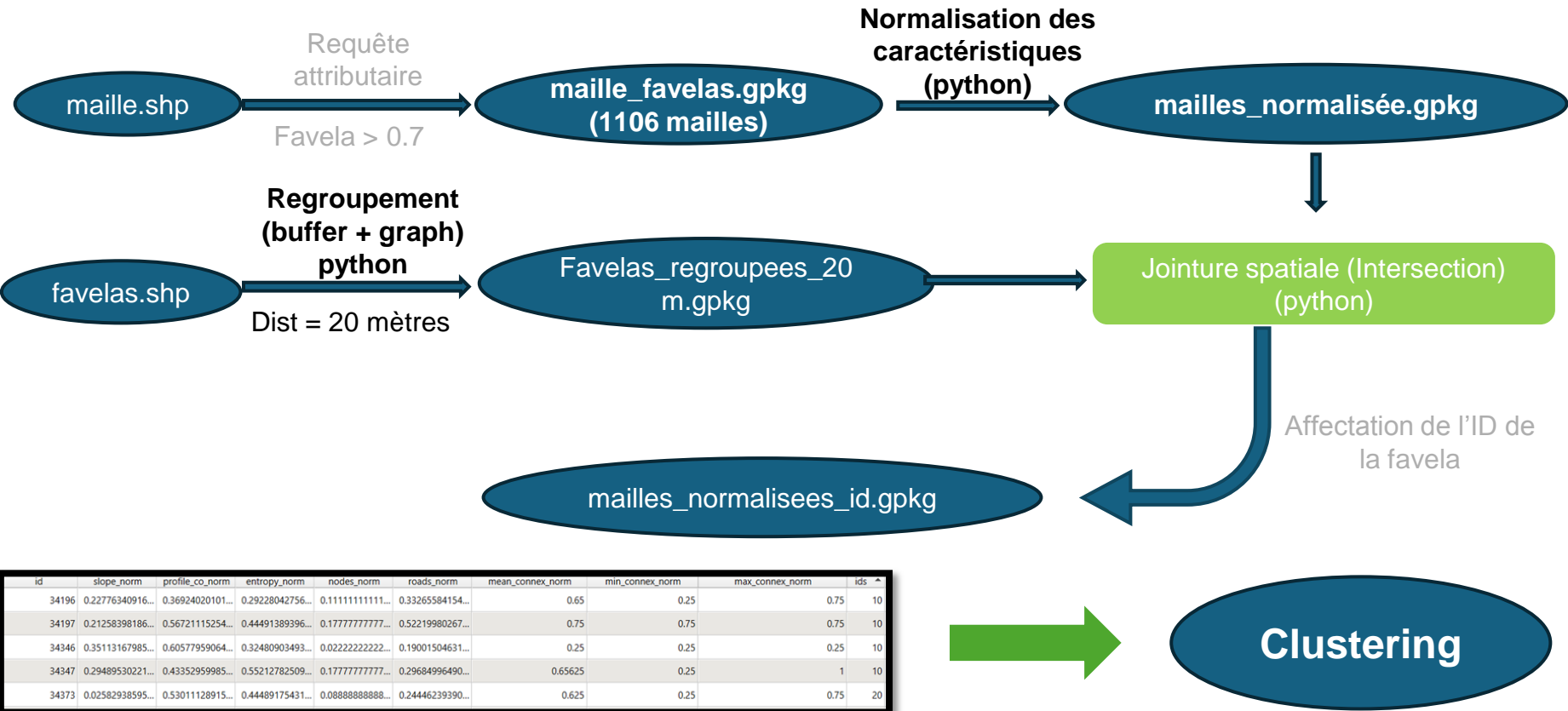
## 2. dérivées d'un **MNE** (modèle numérique d'élévation) :

- Pente (slope)
- Convexité de profil (profile\_co)

## 3. dérivées **OSM** (OpenStreetMap - **Réseau routier**) :

- Nombre de noeuds de rue (nodes)
- Nombre totale des rues (roads)
- Nombre de Connexions (moyennes, min, max)

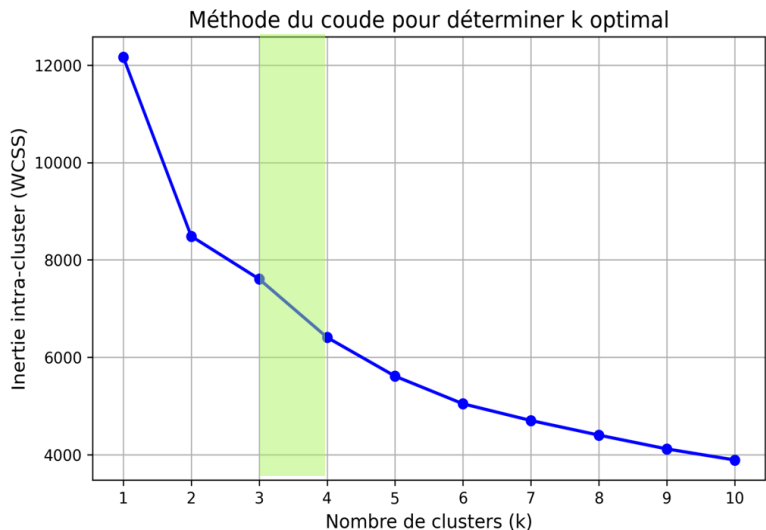
# Traitement des données



id	slope_norm	profile_co_norm	entropy_norm	nodes_norm	roads_norm	mean_connex_norm	min_connex_norm	max_connex_norm	ids	ids
34196	0.22776340916...	0.36924020101...	0.29228042756...	0.111111111111...	0.33265584154...	0.65	0.25	0.75	10	10
34197	0.21258398186...	0.56721115254...	0.44491389396...	0.17777777777...	0.52219980267...	0.75	0.75	0.75	10	10
34346	0.35113167985...	0.60577959064...	0.32480903493...	0.02222222222...	0.19001504631...	0.25	0.25	0.25	10	10
34347	0.29489530221...	0.43352959985...	0.55212782509...	0.17777777777...	0.29684996490...	0.65625	0.25	1	10	10
34373	0.02582938595...	0.53011128915...	0.44489175431...	0.08888888888...	0.24446239390...	0.625	0.25	0.75	20	20

# Détermination de nombre de clusters

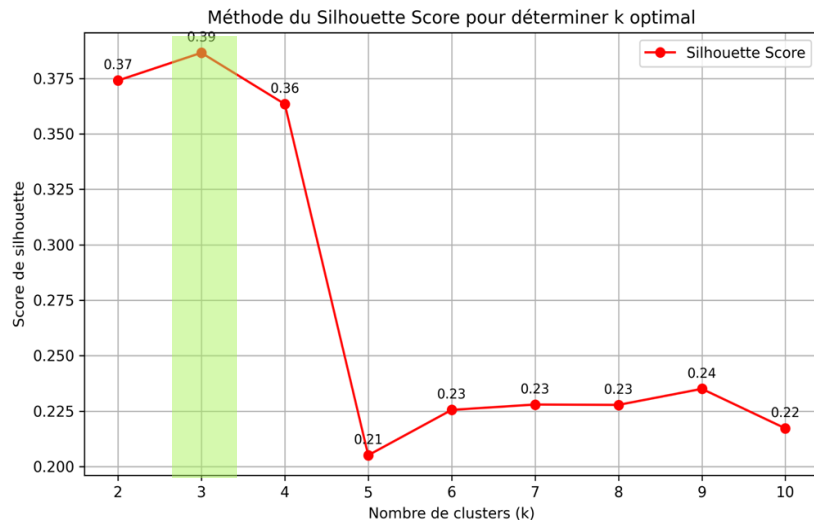
## Méthode du coude



La **méthode du coude** est une technique utilisée pour **déterminer le nombre optimal de clusters** dans une analyse de regroupement non supervisée, comme **K-means**.

Le WCSS est la somme des distances au carré entre chaque point et le centroïde de son cluster.

## Méthode du Silhouette Score



Il mesure à quel point un point est bien associé à son propre cluster (cohésion) et mal associé aux clusters voisins (séparation).

La différence entre la distance moyenne entre le point  $i$  et les points du cluster le plus proche et la distance moyenne entre  $i$  et les autres points de son propre cluster divisée par le maximum de ces deux moyennes.

# Clustering des favelas

## 1. Kmeans classique

*Pas satisfaisant : les favelas voisines sont souvent séparées en différents clusters*

## 2. K-means avec variables spatiales (x,y)

*Clusters trop influencés par la position géographique*

## 3. K-means + lissage spatiale

*Meilleure cohérence spatiale, mais certains favelas voisines sont séparées en différents clusters.*

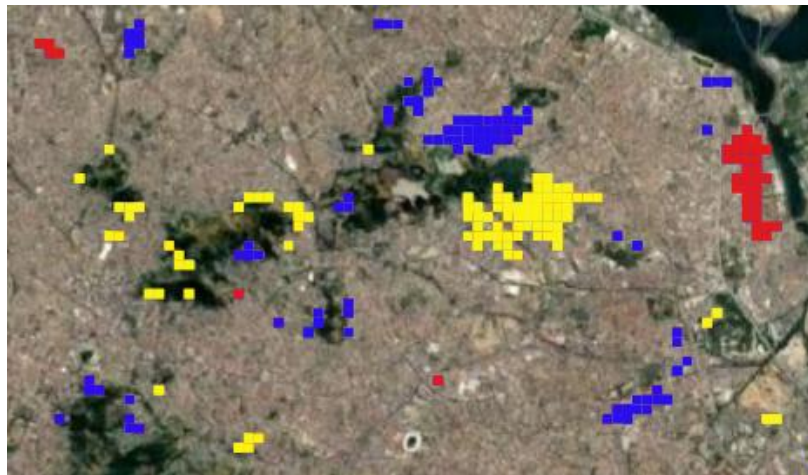
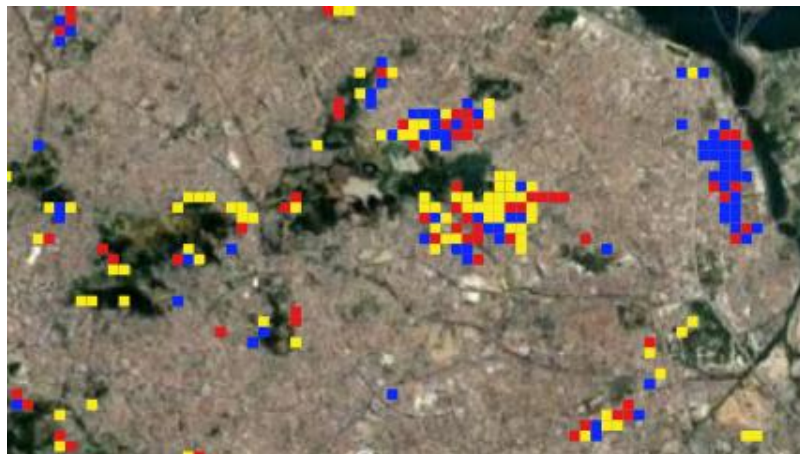
## 4. COP-Kmeans (semi-supervisé avec contraintes)

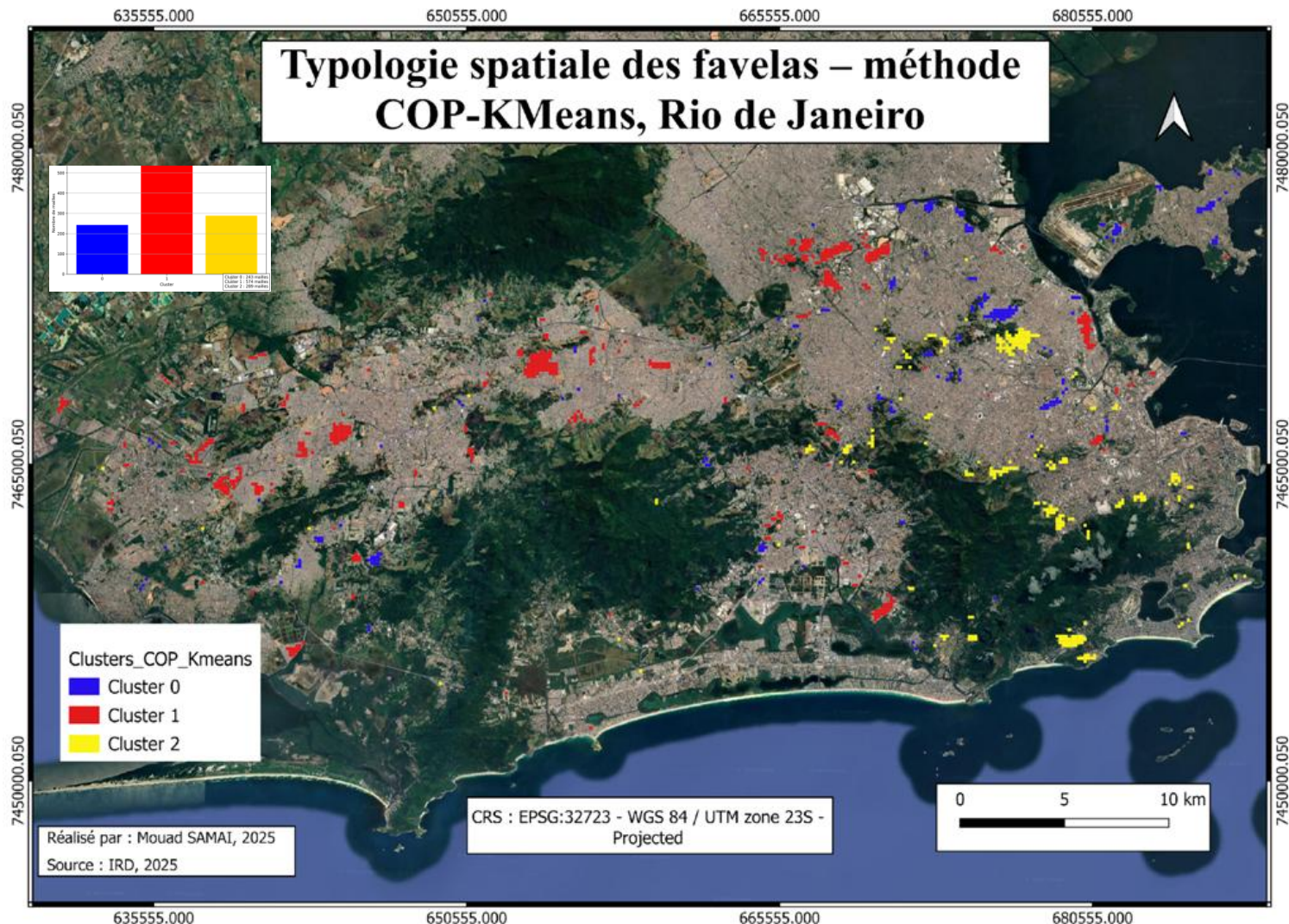
*Résultats plus pertinents : les favelas liées (par ID) sont regroupées, meilleure cohérence spatiale et thématique.*

**COP-KMeans** (Constrained KMeans) : extension du KMeans classique intégrant des **contraintes supervisées** (*Must-Link* et/ou *Cannot-Link*).

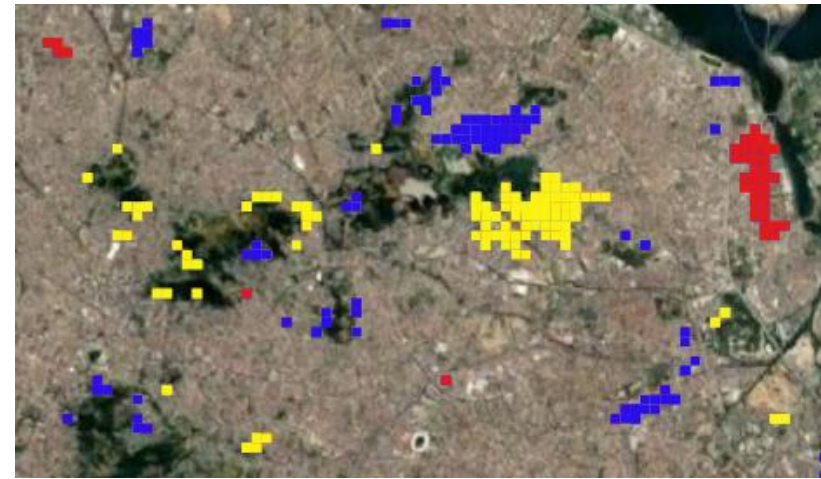
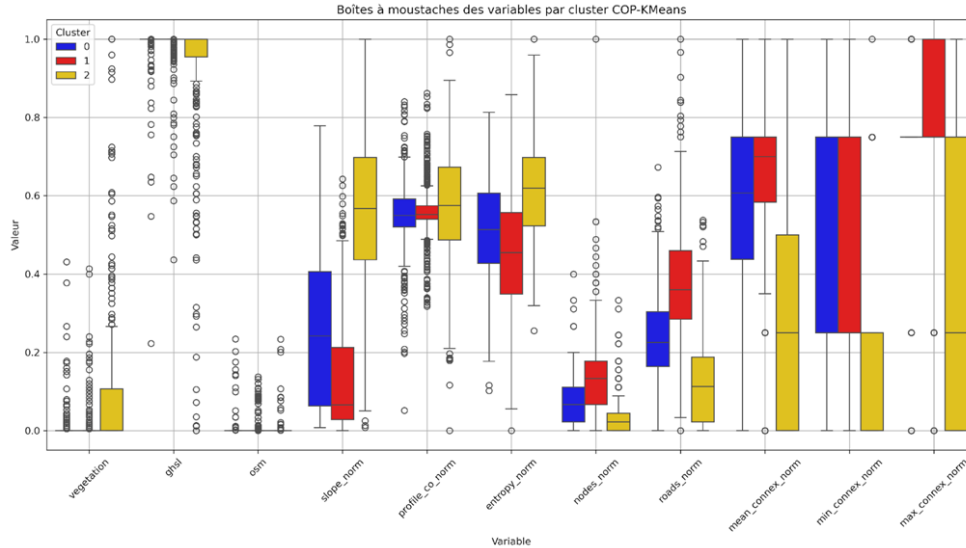
⇒ permettent de guider la classification en imposant que **certains points soient obligatoirement dans le même cluster** (*Must-Link*),

⇒ améliore la **cohérence thématique ou spatiale** des groupes.





# Explication / Caractérisation des clusters de Favelas



## 1. Interprétation des clusters

## 2. Analyse des corrélations avec :

- **Données de Santé** (fichier COVID de la FIOCRUZ)
- **Images de Températures de Surface**