

Mapping Historical Sites (Geospatial Visualization)

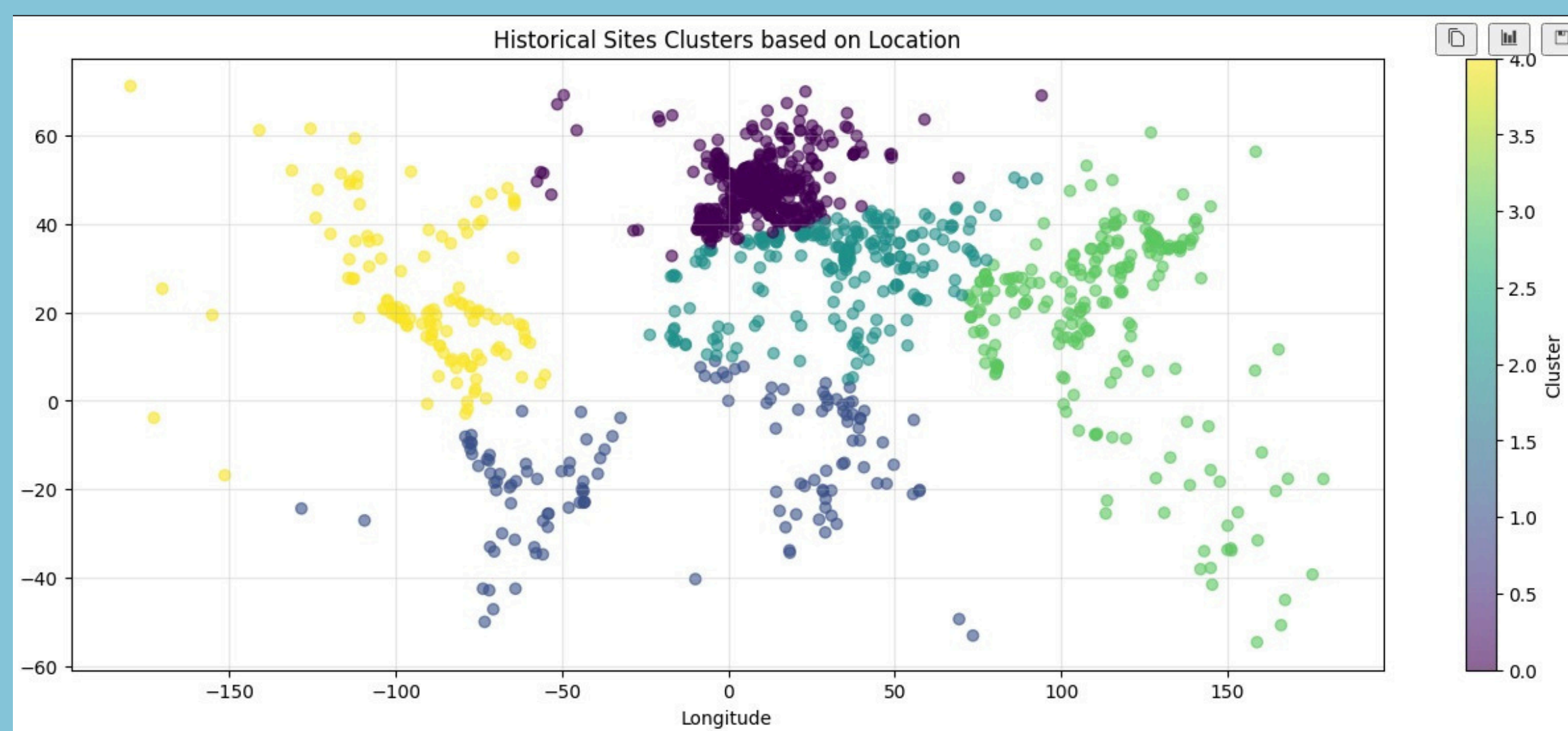
Encadré par:
Pr. EL Habib Ben Lahmar
M. Oussama Kaich



Introduction

L'explosion des données géographiques et la nécessité croissante de mieux comprendre l'organisation spatiale de certains phénomènes ont conduit au développement de nombreuses techniques d'analyse spatiale. Parmi ces techniques, le clustering (ou regroupement) est particulièrement utile pour identifier des schémas et des tendances non visibles à l'œil nu. Dans ce projet, nous nous intéressons à l'algorithme K-Means, une méthode d'apprentissage non supervisé qui permet de regrouper automatiquement des données en fonction de leur proximité. L'objectif principal est de regrouper des points géographiques (latitude et longitude) en clusters homogènes afin d'identifier des zones de forte densité, des regroupements spatiaux naturels, ou encore des zones isolées. Cette analyse peut servir de base à de nombreuses applications, notamment dans les domaines de la planification urbaine, de l'environnement, de la logistique, ou encore de la gestion de ressources naturelles.

Results



L'algorithme K-Means a réussi à regrouper les points géographiques en 5 clusters distincts, révélant des zones de regroupement naturel.

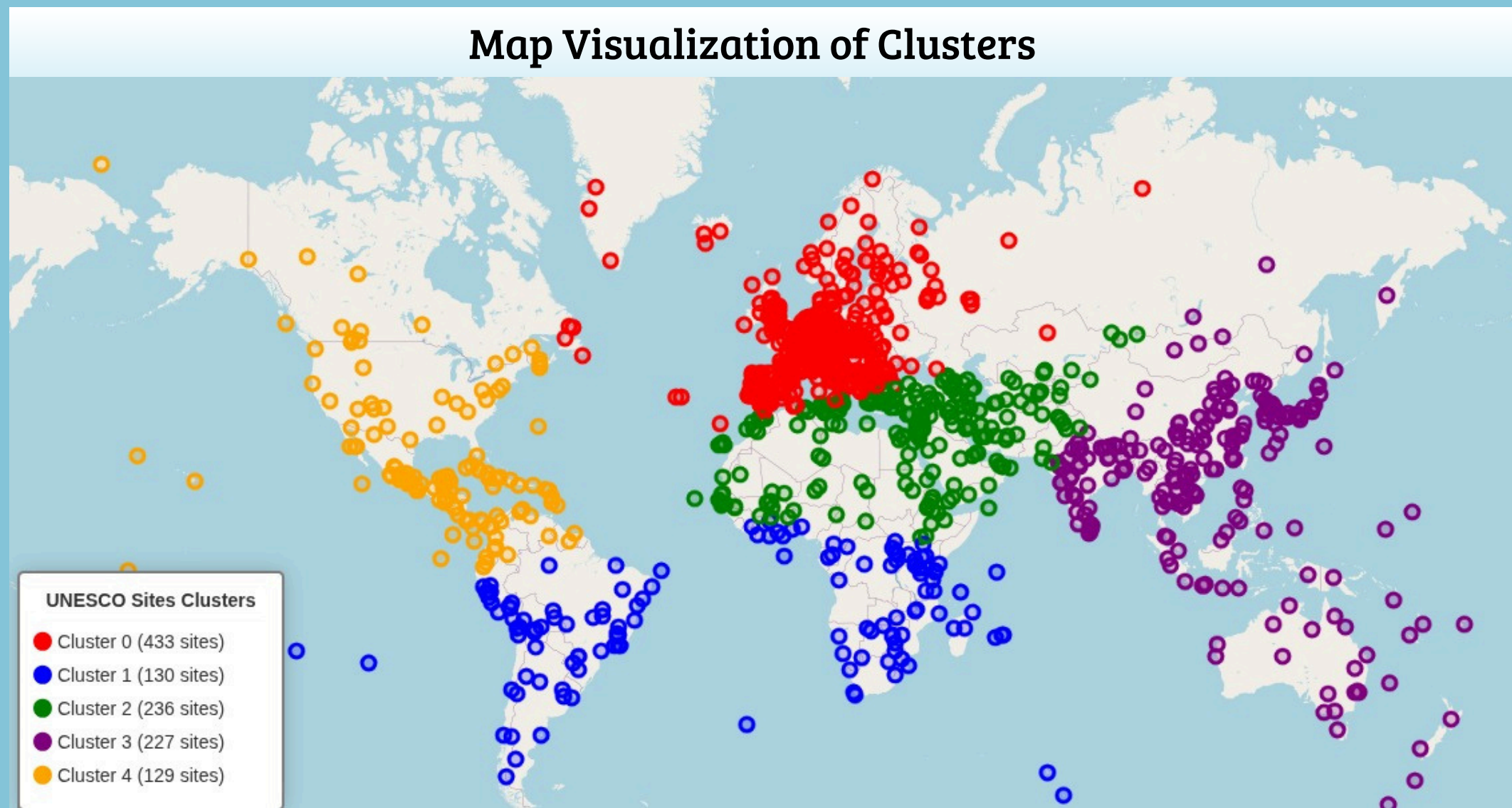
Cluster 0 (433 points) :
Majoritairement localisé en Europe et en Afrique du Nord, avec des coordonnées moyennes proches de 48°N et 9°E.

Cluster 1 (130 points) :
Regroupe des points situés principalement dans l'hémisphère sud, notamment en Afrique australe, en Océanie, et en Amérique du Sud.

Cluster 2 (236 points) :
Comprend des points localisés dans les régions du Moyen-Orient, en Afrique et certaines zones d'Asie.

Cluster 3 (227 points) :
Correspond à des localisations en Asie du Sud-Est, Asie centrale, et Océanie.

Cluster 4 (129 points) :
Points situés en grande majorité dans l'Amérique centrale, les Caraïbes, et le nord de l'Amérique du Sud.



Interprétation

- Chaque cluster présente une cohérence géographique, ce qui montre la capacité du K-Means à capturer des structures spatiales.
- Certains clusters sont très denses (comme le cluster 0), tandis que d'autres sont plus dispersés, ce qui peut indiquer des différences de distribution.
- La carte interactive permet de mieux visualiser l'emplacement de chaque cluster et d'interpréter visuellement les regroupements.

Conclusion

Ce projet a montré que l'algorithme K-Means est un outil puissant pour regrouper des données géographiques et identifier des tendances spatiales. Les résultats obtenus confirment que :

- Les clusters formés sont géographiquement cohérents
- Il est possible de mettre en évidence des zones à forte concentration
- Cette méthode peut être utilisée dans des domaines très concrets, tels que l'urbanisme, l'environnement, ou la logistique.

Pistes futures

- Tester d'autres algorithmes comme DBSCAN, qui tient compte de la densité.
- Ajouter d'autres variables contextuelles (altitude, climat, population...).
- Développer une application web interactive pour explorer dynamiquement les clusters.

Methodology

Pour mener cette étude, nous avons suivi une démarche structurée reposant sur les étapes suivantes :

STEP
01

Prétraitement des données

Nous avons travaillé sur un jeu de données contenant uniquement des coordonnées géographiques (latitude et longitude). Afin d'assurer une bonne performance de l'algorithme K-Means, nous avons appliqué une normalisation des valeurs à l'aide de la méthode StandardScaler de scikit-learn.

STEP
02

Détermination du nombre de clusters (k)

Nous avons utilisé la méthode du coude (Elbow Method), qui consiste à mesurer l'inertie (la somme des distances intra-cluster) pour plusieurs valeurs de k. Le point où l'inertie commence à diminuer moins rapidement correspond au "coude". Dans notre cas, la valeur optimale était k = 5.

STEP
03

Application de l'algorithme K-Means

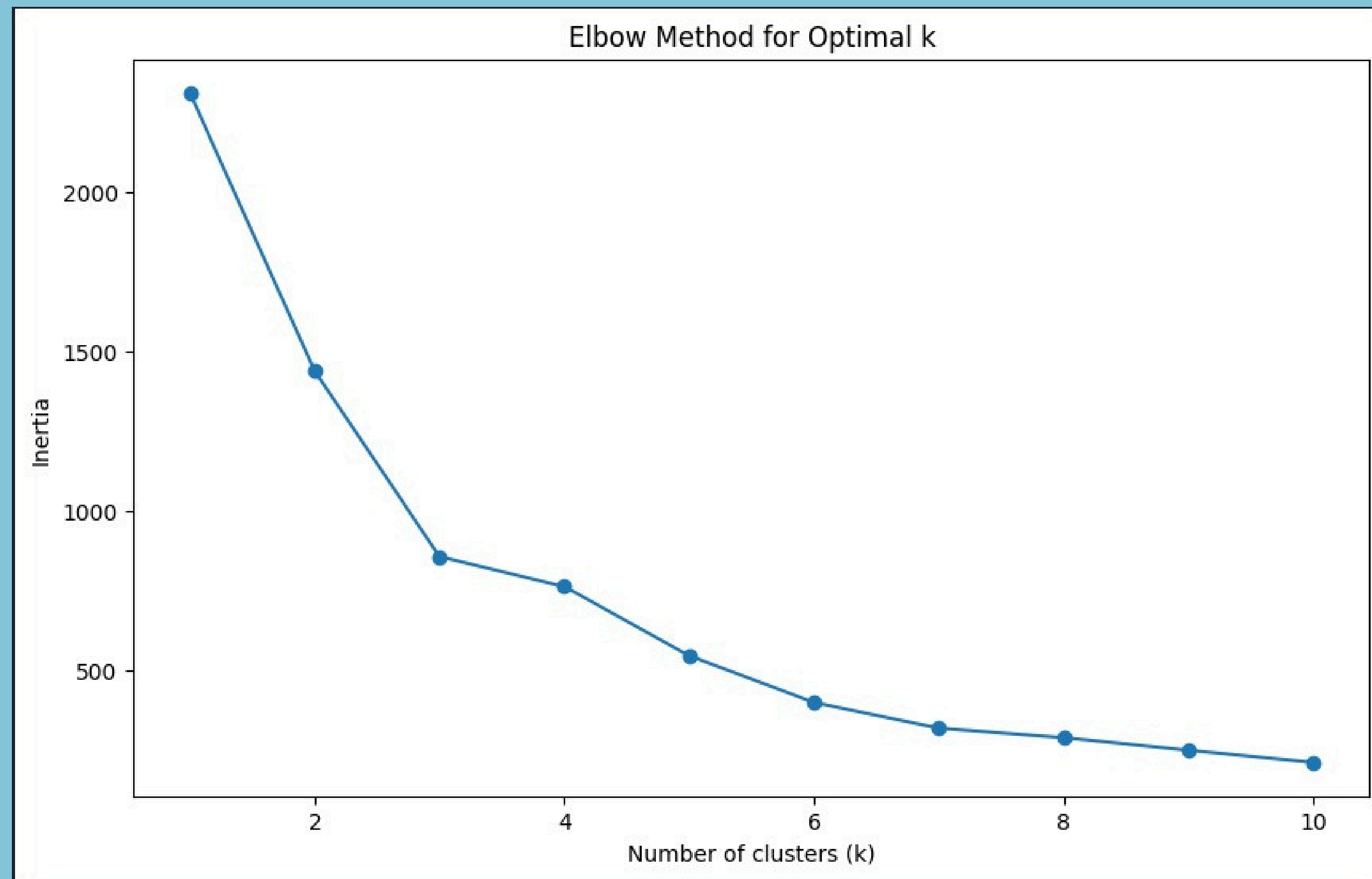
L'algorithme a été appliqué sur les données normalisées. Chaque point a été assigné à un cluster en fonction de sa distance au centroïde le plus proche.

STEP
04

Visualisation des résultats

Nous avons utilisé :

- des graphiques de dispersion pour observer les clusters dans l'espace normalisé
- une carte interactive (avec la bibliothèque folium) pour visualiser les clusters sur le globe.



réalisé par:

OTHMAN SALAHI : oothmansalahi.24@gmail.com

MALAK HOUALI : malak.houali05@gmail.com

HIBA DELLAJI : hibadlj3@gmail.com