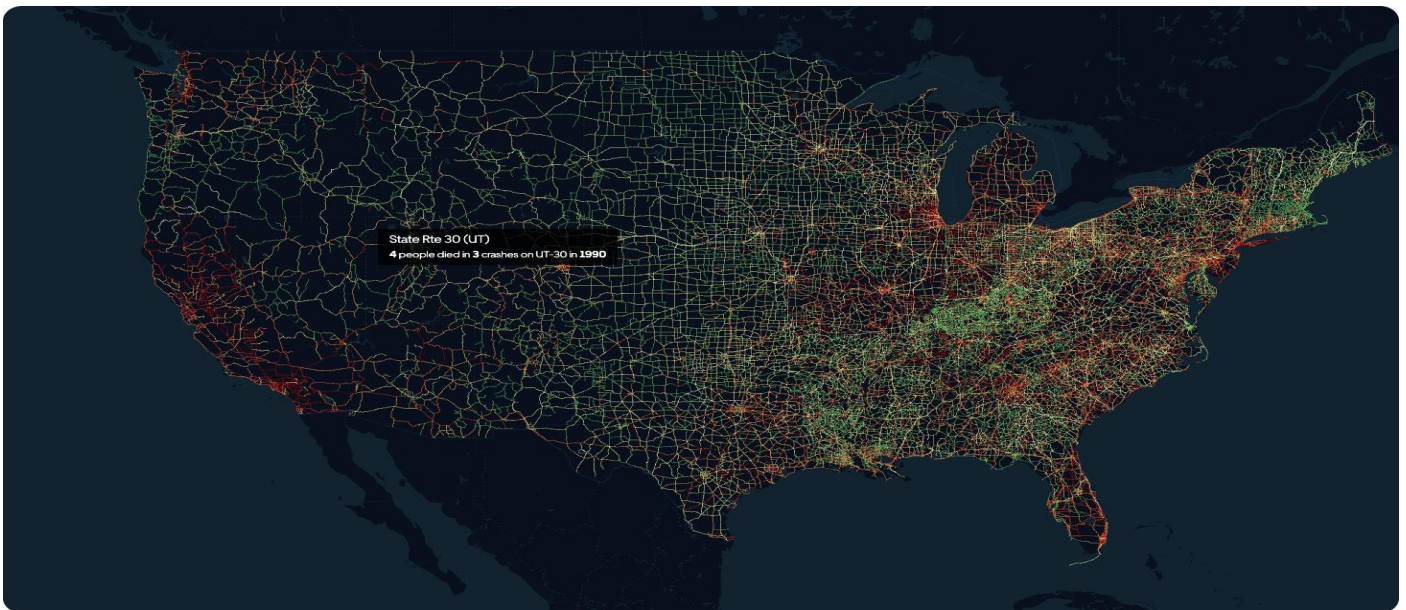


---

# Rapport De Projet Fin Module Deep Learning

## Mapping Historical Sites (Geospatial Visualization)

---



### Réaliser par :

Hiba Dellaji  
Othman Salahi  
Malak Houali

### Encadrer par :

Pr. El Habib Ben Lahmer  
Dr. Oussama Kaich

**Année Universitaire : 2024-2025**

# Remerciements

Je tiens à exprimer ma profonde gratitude à toutes les personnes qui ont contribué à la réalisation de ce projet de Deep Learning intitulé *"Mapping Historical Sites (Geospatial Visualization)"*.

Je remercie tout d'abord **Pr. El Habib Ben Lahmer**, responsable du cours de Deep Learning, pour la qualité de son enseignement, sa pédagogie rigoureuse et sa capacité à rendre accessibles des concepts complexes. Ce cours m'a permis d'acquérir des bases solides et essentielles pour mener à bien ce projet.

Je souhaite également adresser mes plus sincères remerciements à **Dr. Oussama Kaich**, qui a assuré l'encadrement de ce travail. Sa disponibilité, ses conseils techniques avisés, ainsi que ses retours constructifs ont été d'une grande aide tout au long du processus. Son accompagnement attentif a constitué un réel atout dans l'avancement et la structuration du projet.

Je n'oublie pas mes camarades, dont le soutien, les échanges et l'entraide ont fortement contribué à créer un climat de travail motivant et enrichissant.

Je remercie enfin ma famille et mes proches pour leur soutien constant, leur patience et leurs encouragements tout au long de cette aventure académique.

## Table des matières

<b>Remerciements .....</b>	<b>2</b>
<b>1. Introduction.....</b>	<b>5</b>
<b>2. Problématique .....</b>	<b>5</b>
<b>1. Choix des Données .....</b>	<b>7</b>
<b>2. Description des données.....</b>	<b>7</b>
<b>4. Prétraitement appliqué.....</b>	<b>8</b>
<b>5. Choix et Entraînement du modèle.....</b>	<b>8</b>
<b>5.1. Détermination du nombre optimal de clusters .....</b>	<b>8</b>
<b>5.2. Clustering des sites historiques.....</b>	<b>9</b>
<b>6. Visualisation interactive sur carte .....</b>	<b>10</b>
<b>Conclusion .....</b>	<b>11</b>

## Liste des figures

Figure 1: Extrait du jeux de données (dataset) .....	7
Figure 2:Elbow Method for Optimal k .....	8
Figure 2:Elbow Method for Optimal k .....	8
Figure 3: Historical Sites Clusters based on Location.....	9
Figure 4: Carte interactive finale des clusters de sites historiques .....	10
Figure 4: Carte interactive finale des clusters de sites historiques .....	10

# 1. Introduction

Le patrimoine historique constitue une richesse culturelle inestimable, témoignant des civilisations passées et façonnant notre compréhension du monde. La cartographie et la visualisation géospatiale des sites historiques jouent aujourd'hui un rôle crucial dans leur préservation, leur étude, et leur valorisation. Grâce aux avancées récentes en intelligence artificielle et en Deep Learning, il est désormais possible d'exploiter de vastes ensembles de données géographiques, textuelles et visuelles pour identifier, localiser et analyser ces sites de manière automatisée et efficace.

Ce projet s'inscrit dans cette dynamique en proposant une solution basée sur le Deep Learning pour la cartographie intelligente des sites historiques. L'objectif principal est de développer un modèle capable d'extraire, de traiter et de visualiser des données géospatiales en lien avec des lieux d'intérêt patrimonial. En s'appuyant sur des données telles que des images satellites, des coordonnées GPS ou encore des bases de données patrimoniales, le projet vise à générer une visualisation interactive et contextualisée des sites historiques, contribuant ainsi à leur accessibilité et à leur préservation numérique.

Pour ce faire, plusieurs étapes ont été nécessaires : la collecte et le prétraitement des données, la mise en place d'un modèle de Deep Learning adapté, ainsi que l'intégration des résultats dans une interface de visualisation géographique. Le projet mobilise des outils modernes tels que [PyTorch/TensorFlow, Folium, Mapbox.], et s'inscrit dans une logique interdisciplinaire à la croisée de l'intelligence artificielle, des humanités numériques et des systèmes d'information géographique .

Ce rapport présente de manière détaillée l'ensemble des étapes du projet, les choix techniques effectués, les résultats obtenus ainsi que les perspectives d'amélioration et d'extension du système proposé.

## 2. Problématique

La réalisation d'un système de cartographie intelligente des sites historiques soulève plusieurs problématiques majeures. Tout d'abord, se pose la question de l'automatisation de l'identification et de la localisation des sites historiques à partir de données hétérogènes, telles que des images satellites, des métadonnées ou encore des coordonnées GPS. Ces données étant souvent bruitées ou incomplètes, il est essentiel de concevoir un processus de traitement robuste.

Une seconde problématique réside dans le choix d'un modèle de Deep Learning adapté à cette tâche. Il s'agit de déterminer quelle architecture (CNN, RNN, ou modèle multimodal) permettrait d'extraire efficacement les

caractéristiques pertinentes pour détecter les sites historiques à partir d'images et de données géospatiales.

Vient ensuite la question de la visualisation : comment représenter de manière claire, interactive et compréhensible les résultats du modèle sur une carte géographique ? Cela implique l'intégration avec des outils SIG et des bibliothèques de cartographie dynamiques tout en veillant à l'ergonomie de l'interface.

Enfin, une problématique transversale concerne la fiabilité des résultats : comment garantir que le modèle est suffisamment précis pour être utilisé sur des données réelles, parfois anciennes ou imprécises, dans un contexte historique aux références parfois floues ?

# 1. Choix des Données

Les données exploitées dans le cadre de ce projet proviennent d'un fichier CSV nommé whc-sites.csv. Ce fichier contient une liste exhaustive des sites inscrits au patrimoine mondial de l'UNESCO. Il s'agit de données ouvertes, accessibles au public via les portails officiels de l'UNESCO ou à travers des plateformes collaboratives telles que Kaggle. Leur origine institutionnelle garantit la fiabilité et la qualité des informations qu'elles contiennent.

## 2. Description des données

Le fichier regroupe des enregistrements correspondant à différents sites culturels, naturels ou mixtes reconnus mondialement. Chaque ligne du dataset représente un site et comprend diverses colonnes telles que : le nom du site (Name), sa localisation géographique à travers les coordonnées (latitude, longitude), le pays concerné (Country name), la région géographique (Region), la catégorie du site (category\_long), l'année d'inscription (date\_inscribed), ainsi qu'une courte description historique (short\_description). Ce niveau de détail permet d'analyser les sites sous plusieurs angles : géographique, temporel et culturel.

1	Name	short_descri	date_inscrib	danger	date_end	longitude	latitude	area_hectare	category_lon	category_shc	Country nam	Region	iso_code	transbound	rev_bis
2	Lâ€"Anse au <p>At the tip		1978	0		-55.6166667	51.4666667	7991	Cultural	C	Canada	Europe and f	ca		0
3	Nahanni Nat <p>Located a		1978	0		-125.589444	61.5472222	476560	Natural	N	Canada	Europe and f	ca		0
4	Galâ"pagos I <p>Situat		1978	0	2010	-90.501319	-0.68986	14066514	Natural	N	Ecuador	Latin Americ	ec	0 Bis	
5	City of Quito <p>Quito, th		1978	0		-78.5120833	-0.22	70.43	Cultural	C	Ecuador	Latin Americ	ec		0
6	Simien Natic <p>Massive i		1978	0	2017	38.0666667	13.1833333	13600	Natural	N	Ethiopia	Africa	et		0
7	Rock-Hewn C <p>The 11 m		1978	0		39.04042	12.02935	0	Cultural	C	Ethiopia	Africa	et		0
8	Aachen Cath <p>Construc		1978	0		6.08444444	50.7744444	0.2	Cultural	C	Germany	Europe and f	de		0 bis
9	Historic Cent <p>The Histic		1978	0		19.9372222	50.0613889	149.65	Cultural	C	Poland	Europe and f	pl		0 bis
10	Wieliczka an <p>The depo		1978	0	1998	20.0638889	49.9791667	1104.947	Cultural	C	Poland	Europe and f	pl		0 Ter
11	Island of Gor <p>The islan		1978	0		-17.40083	14.66722	0	Cultural	C	Senegal	Africa	sn		0
12	Mesa Verde I <p>A great c		1978	0		-108.485556	37.2616667	21043	Cultural	C	United State	Europe and f	us		0
13	Yellowstone <p>The vast		1978	0	2003	-110.82778	44.46056	898349	Natural	N	United State	Europe and f	us		0
14	Boyana Chur <p>Located c		1979	0		23.2666667	42.65	0.68	Cultural	C	Bulgaria	Europe and f	bg		0
15	Madara Ride <p>The Mad		1979	0		27.15	43.3	1.2	Cultural	C	Bulgaria	Europe and f	bg		0
16	Thracian Tor <p>Discover		1979	0		25.4	42.6166667	0.0155	Cultural	C	Bulgaria	Europe and f	bg		0
17	Rock-Hewn C <p>In the val		1979	0		25.9666667	43.7166667	171.9	Cultural	C	Bulgaria	Europe and f	bg		0
18	Dinosaur Prc <p>In additi		1979	0		-111.492222	50.7677778	7825	Natural	N	Canada	Europe and f	ca		0
19	Old City of Di <p>The 'Pear		1979	0	1998	18.1088611	42.6414211	96.7	Cultural	C	Croatia	Europe and f	hr		0
20	Historical Cc <p>The ruins		1979	0		16.44333	43.50944	20.8	Cultural	C	Croatia	Europe and f	hr		0
21	Plitvice Lake <p>The wate		1979	0	1997	15.61444	44.87778	29630.77	Natural	N	Croatia	Europe and f	hr		0 Bis
22	Virunga Nati <p>Virunga f		1979	1		29.1666667	0.9166667	800000	Natural	N	Democratic f	Africa	cd		0

Figure 1: Extrait du jeu de données (dataset))

Le choix de ce jeu de données se justifie par sa richesse informative et sa pertinence par rapport aux objectifs du projet. Il offre une combinaison idéale de données géospatiales et descriptives, permettant ainsi à la fois un traitement algorithmique (clustering) et une représentation visuelle interactive. Sa diversité, incluant plusieurs régions du monde et différentes typologies de sites, rend les résultats plus représentatifs. En outre, la disponibilité de ces données en open data facilite leur réutilisation dans un cadre pédagogique et technique.

## 4. Prétraitement appliqué

Avant d'entamer l'analyse, un ensemble d'opérations de nettoyage et de préparation des données a été réalisé. Les entrées présentant des valeurs manquantes dans les colonnes essentielles (latitude, longitude) ont été éliminées. Ensuite, un processus de standardisation a été appliqué sur les variables de localisation pour assurer une mise à l'échelle adéquate avant l'application de l'algorithme de clustering. Enfin, seules les colonnes utiles à l'analyse spatiale ont été conservées, ce qui a permis d'optimiser les performances des modèles tout en réduisant la complexité du traitement.

## 5. Choix et Entraînement du modèle

### 5.1. Détermination du nombre optimal de clusters

Avant d'appliquer l'algorithme K-Means, une étape essentielle consiste à déterminer le nombre optimal de clusters (**k**) à utiliser. Pour cela, nous avons recours à la méthode du coude, qui consiste à faire varier la valeur de  $k$  sur un intervalle défini et à calculer pour chaque cas l'inertie intra-cluster, c'est-à-dire la somme des distances au carré entre les points d'un cluster et leur centre.

Les résultats sont ensuite représentés graphiquement sous la forme d'une courbe d'inertie. Cette courbe présente généralement une baisse rapide de l'inertie lorsque  $k$  augmente, puis une stabilisation progressive. Le point où la courbe commence à "se plier", formant un coude, correspond à une valeur de  $k$  au-delà de laquelle les gains deviennent marginaux. Dans notre cas, l'observation du graphique a permis d'identifier un coude visible pour **k = 5**, que nous avons retenu comme valeur optimale pour le regroupement.

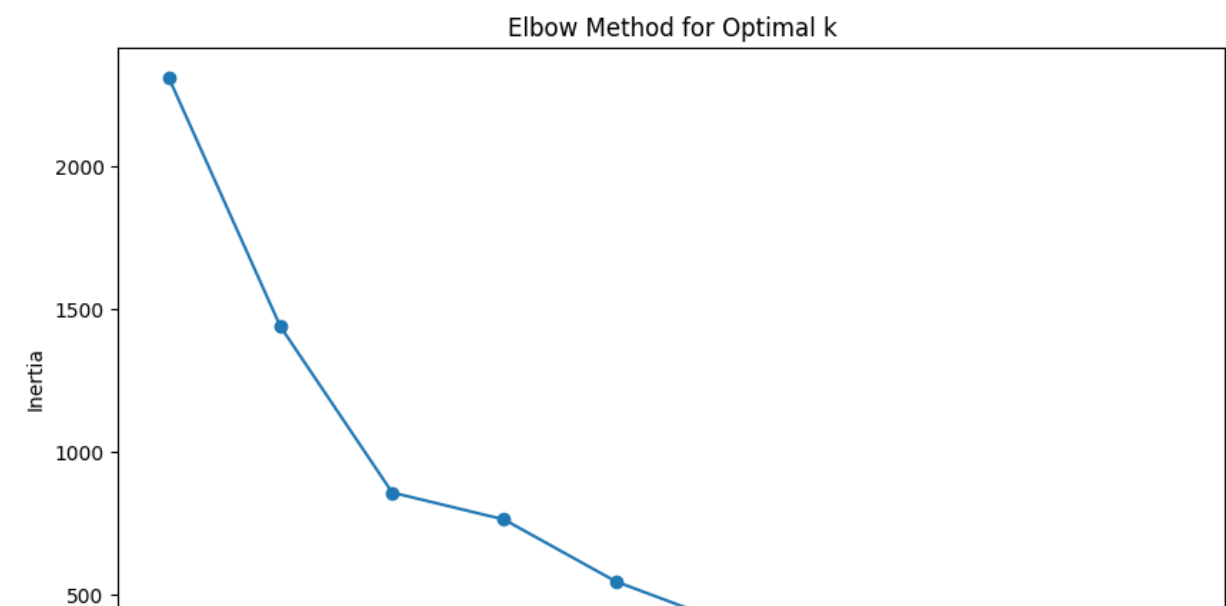


Figure 2: Elbow Method for Optimal k

Figure 3: Elbow Method for Optimal k



## 5.2. Clustering des sites historiques

Une fois le nombre optimal de clusters déterminé, nous avons appliqué l'algorithme K-Means avec  $k = 5$  sur les coordonnées géographiques normalisées des sites. Cette méthode consiste à répartir les points dans l'espace selon leur proximité à des centres calculés itérativement de manière à minimiser la distance moyenne intra-cluster.

Chaque site a ainsi été affecté à un groupe, formant cinq clusters principaux qui traduisent une logique de proximité géographique. L'analyse de ces regroupements permet de révéler des concentrations régionales de sites du patrimoine mondial. Une visualisation statique de ces clusters a été produite à l'aide de Matplotlib, illustrant la répartition spatiale des groupes sur un plan bidimensionnel.

Enfin, une analyse descriptive de chaque cluster a été menée : nombre total de sites, coordonnées moyennes, et régions les plus représentées. Ces informations permettent de comprendre les zones géographiques à forte densité patrimoniale, et d'identifier des tendances selon les continents ou les régions culturelles.

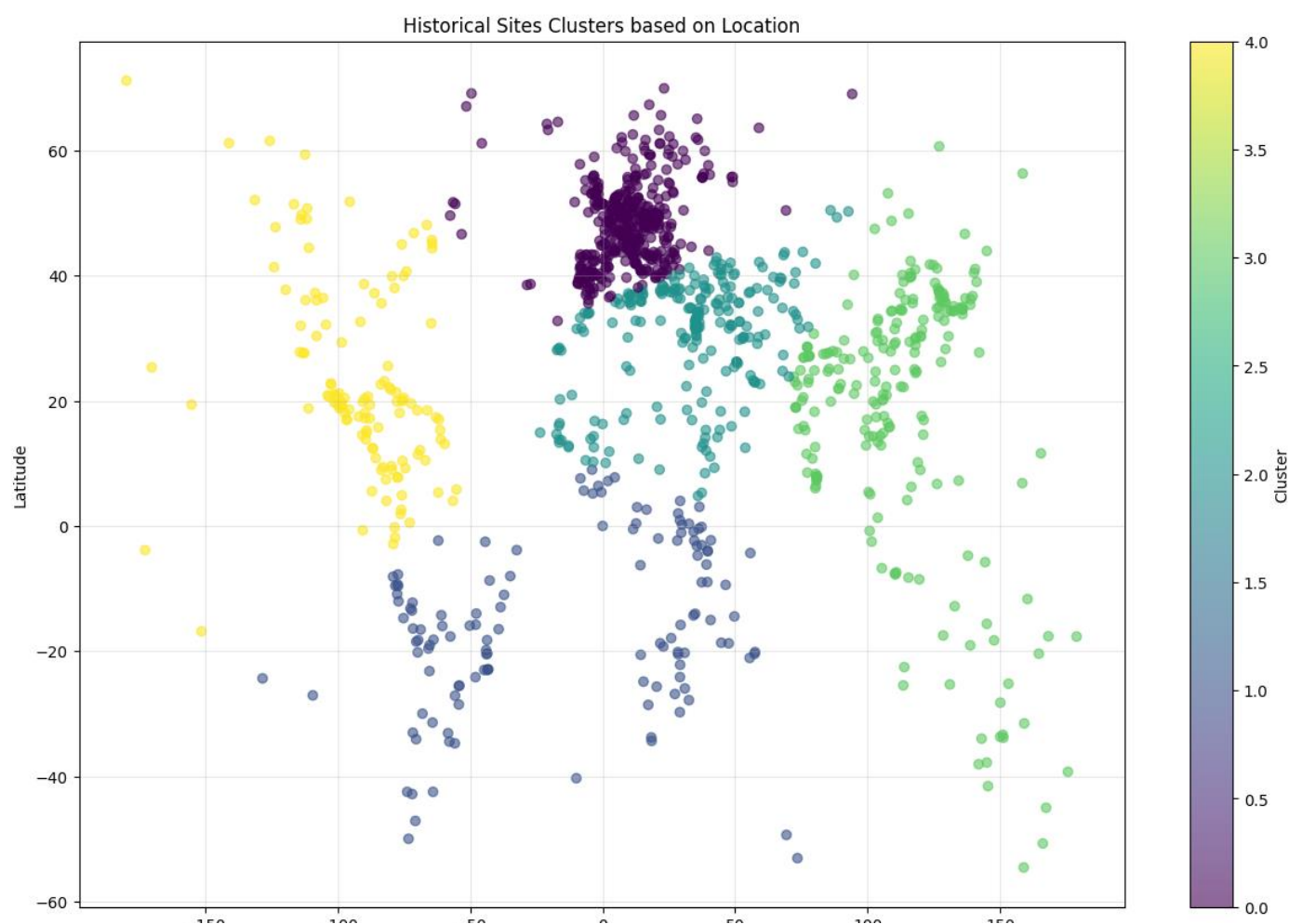


Figure 4: Historical Sites Clusters based on Location

## 6. Visualisation interactive sur carte

Pour rendre l'analyse plus accessible et dynamique, une visualisation interactive a été développée à l'aide de la bibliothèque Folium, spécialisée dans la cartographie web. Une première carte a été construite en positionnant les sites selon leurs coordonnées géographiques, avec un code couleur pour distinguer les clusters. Chaque site a été représenté par un cercle coloré, et une légende a été ajoutée afin d'interpréter facilement la répartition des groupes.

Une version enrichie de la carte a ensuite été créée, intégrant plusieurs fonctionnalités interactives. Chaque site affiche une fenêtre contextuelle (popup) contenant des informations détaillées telles que le nom, la catégorie, le pays, la région, l'année d'inscription et une courte description. De plus, un système de regroupement dynamique a été mis en place grâce au module MarkerCluster, permettant de condenser visuellement les points lorsque l'on zoome en arrière. D'autres éléments d'interface ont été ajoutés pour améliorer la lisibilité : une mini-carte, un bouton plein écran, un contrôle des calques, et une légende dynamique indiquant le nombre de sites par cluster.

Cette carte interactive constitue un outil à la fois esthétique, informatif et fonctionnel. Elle permet à l'utilisateur d'explorer le patrimoine mondial à travers une interface intuitive, tout en mettant en évidence les regroupements géographiques issus du clustering.

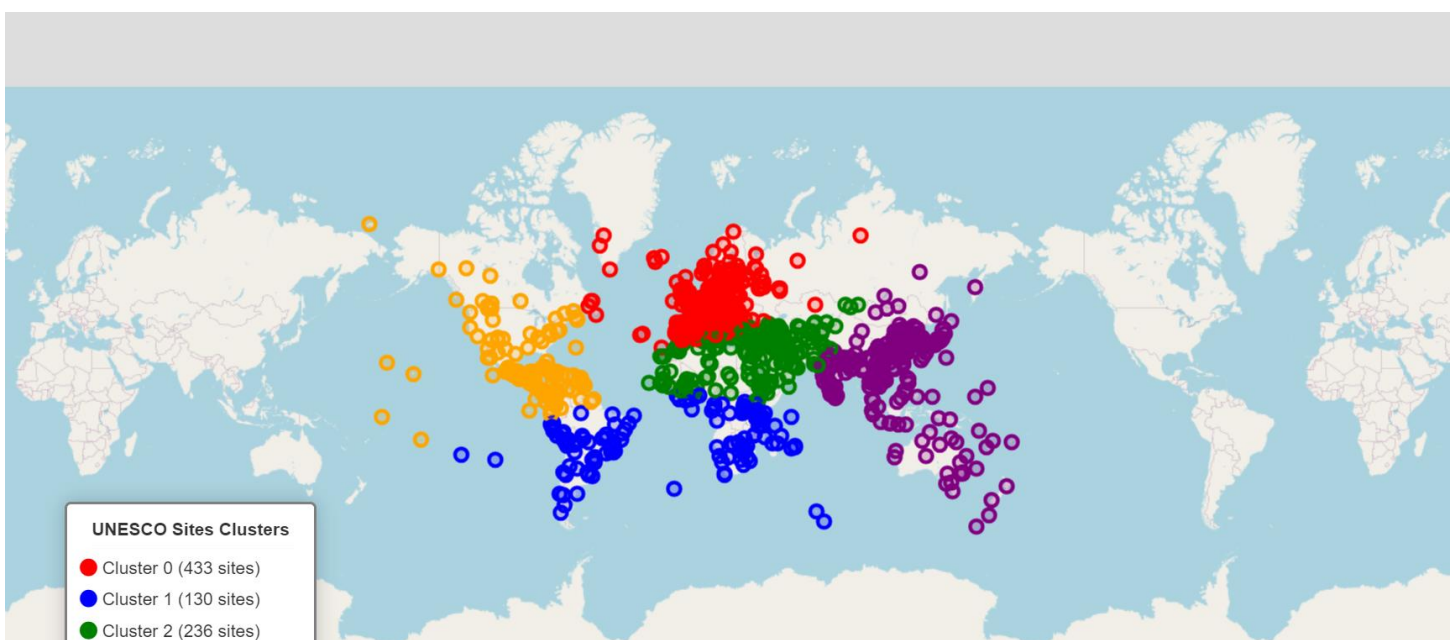


Figure 5: Carte interactive finale des clusters de sites historiques

Figure 6: Carte interactive finale des clusters de sites historiques

## Conclusion

Dans ce projet, nous avons exploré l'application des techniques de clustering non supervisé, en particulier l'algorithme K-Means, pour analyser et visualiser la répartition géographique des sites historiques inscrits au patrimoine mondial de l'UNESCO. Après un travail de prétraitement des données géographiques et une analyse exploratoire, nous avons pu déterminer un nombre optimal de clusters ( $k = 5$ ) à l'aide de la méthode du coude, ce qui a permis d'effectuer un regroupement pertinent des sites selon leur localisation.

Les résultats obtenus ont été visualisés à travers des cartes interactives réalisées avec la bibliothèque Folium, offrant une interface claire et intuitive pour observer la distribution spatiale des différents groupes. Ces visualisations permettent non seulement de détecter des concentrations régionales de sites patrimoniaux, mais aussi de faciliter l'interprétation des zones géographiques les plus riches en patrimoine culturel ou naturel.

Ce travail met en lumière l'apport du Deep Learning géospatial et des outils de Data Science dans la valorisation du patrimoine mondial. Il constitue une base solide pour de futures explorations, comme l'intégration d'autres variables ou encore l'utilisation d'algorithmes plus avancés pour des regroupements plus fins et contextualisés.