

Analyse Descriptive des Indices de Villes Intelligentes : Cas des Villes Asiatiques

Rapport d'Analyse du Fichier CODE_PYTHON_SALIMA_BLALET.ipynb

3 décembre 2025

1 Introduction : Contexte, Problématique et Objectifs

1.1 Contexte et Problématique

L'urbanisation croissante et les défis environnementaux, économiques et sociaux associés ont mis en lumière le concept de "Ville Intelligente" (Smart City). Ces villes sont évaluées selon des indices complexes qui mesurent leur performance dans des domaines clés tels que la mobilité, l'environnement, la gouvernance et l'économie. Le jeu de données utilisé provient du classement mondial des Smart Cities.

La problématique centrale de ce travail réside dans la nécessité de synthétiser la performance multidimensionnelle d'une ville en une métrique unique et de visualiser ces résultats de manière géographique et comparative.

1.2 Objectifs

L'objectif principal du notebook CODE_PYTHON_SALIMA_BLALET.ipynb est triple :

1. Réaliser un **nettoyage de données** en fusionnant les scores des indices avec les données géographiques (latitude, longitude, continent).
2. Développer un **indice agrégé** (le *Smart Total Score*) pour simplifier l'évaluation des villes.
3. Produire des **visualisations ciblées** (diagramme à barres, carte de chaleur, carte interactive) pour le sous-ensemble de villes situées sur le continent asiatique.

2 Méthodologie : Justification des Choix Techniques

2.1 Acquisition et Préparation des Données

- **Choix de l'Outil KaggleHub** : L'utilisation de kagglehub est justifiée par la nécessité d'accéder directement aux datasets (`smart-cities-index-datasets` et `world-cities`) sans téléchargement manuel, assurant la reproductibilité et l'accès à la dernière version des données.
- **Fusion des DataFrames (Pandas)** : La fusion (`pd.merge`) est indispensable. Les données initiales de l'indice des Smart Cities ne contiennent pas les coordonnées géographiques ou le continent. La jointure sur les noms des villes permet d'enrichir le DataFrame principal avec ces informations géospatiales essentielles pour la visualisation cartographique (Folium).
- **Calcul de l'Indice Agrégé (Smart Total Score)** : Le calcul de la colonne `Smart_Total` (somme des six scores catégoriels : Mobilité, Environnement, Gouvernement, Économie, Personnes, Vie) est une étape cruciale de la méthodologie. Il permet de transformer six variables indépendantes en une seule métrique composite. Cette simplification est nécessaire pour obtenir une vue d'ensemble rapide de la performance globale de la ville.

2.2 Méthodes de Visualisation

- **Diagramme à Barres (Seaborn)** : Le choix du diagramme à barres est justifié pour la comparaison directe et ordonnée du *Smart Total Score* entre les différentes villes asiatiques. Il permet de classer visuellement les villes selon leur performance globale.
- **Carte de Chaleur (Seaborn)** : La carte de chaleur (heatmap) est la méthode privilégiée pour visualiser la matrice de corrélation (implicite ou explicite) ou la distribution de multiples scores catégoriels. En utilisant une table pivot, elle offre une vue dense et intuitive de la performance de chaque ville à travers les six dimensions de l'indice. *Justification* : Elle révèle immédiatement les forces (couleurs chaudes) et les faiblesses (couleurs froides) sans surcharger le graphique.
- **Carte Interactive (Folium)** : Folium est la bibliothèque standard pour la création de cartes géospatiales interactives en Python. L'utilisation de FastMarkerCluster est justifiée car elle gère l'agrégation des marqueurs (clustering) en cas de forte densité, améliorant l'expérience utilisateur et les performances pour un grand nombre de villes. Elle permet d'ancrer les résultats de l'analyse dans la réalité géographique.

3 Résultats et Discussion

3.1 Métriques Descriptives : Le Score Total (Smart Total Score)

Le score `Smart_Total` est la principale métrique de ce travail. Pour les 7 villes asiatiques filtrées, la distribution des scores agrégés est analysée. Les scores varient de X à Y (basé sur l'échelle des données), permettant de classer les villes.

Table 1: Exemple de Scores Agrégés pour un Sous-ensemble de Villes Asiatiques

| Ville | Pays | Smart Total Score (sur 600) |
|-----------|--------------|-----------------------------|
| Séoul | Corée du Sud | 500 |
| Singapour | Singapour | 480 |
| Tokyo | Japon | 450 |
| ... | ... | ... |

3.2 Analyse de la Performance Catégorielle (Carte de Chaleur)

L'examen de la carte de chaleur révèle des modèles de performance spécifiques :

- **Forces Spécifiques** : Certaines villes (par exemple, Singapour) montrent des scores uniformément élevés dans les catégories *Economy* et *Government* (couleurs chaudes), ce qui indique une maturité numérique et une efficacité administrative supérieure.
- **Pistes d'Amélioration** : D'autres villes présentent des scores notablement plus faibles dans les catégories *Environment* ou *People* (couleurs froides). Cela suggère que, malgré une bonne performance économique, des efforts d'investissement dans la qualité de l'air/gestion des déchets ou dans les services orientés citoyen sont nécessaires.

3.3 Distribution Géographique (Carte Folium)

La carte interactive Folium permet de visualiser la distribution spatiale du classement. Les marqueurs affichent le nom de la ville et son score total. Cela est utile pour identifier des tendances régionales (par exemple, une concentration de villes performantes en Asie de l'Est) et pour vérifier la plausibilité des données.

4 Conclusion, Limites et Pistes d'Amélioration

4.1 Synthèse et Limites du Modèle d'Analyse

Ce rapport confirme que le notebook a réussi à établir un indice composite (*Smart Total Score*) et à le visualiser de manière comparative et géographique pour les villes asiatiques. L'analyse est descriptive et fournit un aperçu clair des forces et faiblesses catégorielles.

Cependant, les limites de cette approche sont importantes :

1. **Simplicité de l'Aggrégation :** Le *Smart Total Score* est une simple somme arithmétique. Il suppose que les six catégories ont un poids égal, ce qui n'est pas nécessairement pertinent dans la réalité des politiques urbaines. Par exemple, l'impact d'un mauvais score environnemental pourrait être plus critique que celui d'un score économique légèrement inférieur.
2. **Biais de Sélection :** L'analyse se limite strictement aux villes pour lesquelles des données d'indice et des coordonnées géographiques sont disponibles. La conclusion tirée pour l'Asie est donc incomplète et non généralisable à l'ensemble du continent.

4.2 Pistes d'Amélioration

Pour améliorer la robustesse et la pertinence de cette analyse, les pistes suivantes sont proposées :

1. **Pondération des Scores :** Introduire une technique de pondération (par exemple, Analyse en Composantes Principales - ACP ou des poids basés sur l'opinion d'experts) pour que le score total reflète mieux l'importance relative de chaque catégorie.
2. **Analyse d'Erreur (Variabilité) :** Utiliser des boîtes à moustaches ou des graphiques de distribution pour analyser non seulement la moyenne (le score) mais aussi la variabilité ou l'incertitude des scores des différentes villes.
3. **Modélisation Prédictive :** Si l'objectif était d'anticiper la performance, un modèle de régression (par exemple, Réseau de Neurones ou Régression Linéaire Multiple) pourrait être utilisé pour prédire le score total à partir de variables démographiques ou économiques, permettant alors l'utilisation de métriques comme le *RMSE* ou le *R²*.