

Analyse en Composantes Principales et Clustering des Données Mondiales : Focus sur le Maroc

A.Kellaa M.Harbili M.Ait Ahlal

École Nationale des Sciences Appliquées de Berrechid
Génie Informatique S8

20 Mai 2025

Plan de la présentation

- 1 Introduction
- 2 Méthodologie
- 3 Analyse en Composantes Principales
- 4 Résultats et Analyses
- 5 Conclusion

Contexte et Problématique

- **Contexte mondial** : Complexité des données socio-économiques entre pays
- **Défis analytiques** :
 - Dimensionnalité élevée (nombreuses variables)
 - Multicolinéarité entre variables
 - Données manquantes
 - Hétérogénéité des pays
- **Pertinence pour le Maroc** : Comprendre sa position dans le panorama mondial
- **Approche** : Analyse multidimensionnelle (ACP et clustering)

Objectifs du projet

- Identifier les dimensions principales expliquant les variations entre pays
- Positionner le Maroc dans le paysage mondial
- Déterminer les pays similaires au Maroc
- Découvrir des regroupements naturels (clusters) de pays
- Analyser les forces et faiblesses du Maroc par rapport aux autres pays

Présentation du jeu de données

Global Country Information Dataset 2023

- Plus de 100 pays
- Plusieurs dizaines d'indicateurs
- Couvre démographie, santé, économie, éducation...

Type de variables	Exemples
Démographiques	Population, Fécondité
Santé	Espérance de vie, Médecins/1000
Économiques	Chômage, Revenus fiscaux
Éducation	Scolarisation primaire et tertiaire

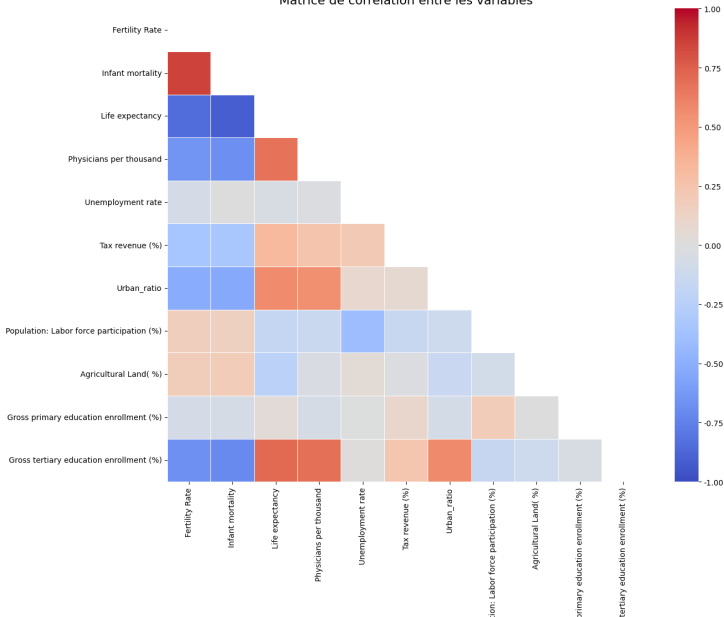
Variables sélectionnées pour l'analyse

- Taux de fécondité
- Mortalité infantile
- Espérance de vie
- Médecins pour 1000 habitants
- Taux de chômage
- Revenus fiscaux (% PIB)
- Population urbaine (%)
- Participation à la force de travail (%)
- Terres agricoles (%)
- Scolarisation primaire (%)
- Scolarisation tertiaire (%)

Prétraitement : Standardisation (moyenne=0, écart-type=1), gestion des valeurs manquantes

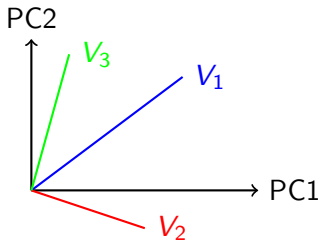
Analyse des corrélations

Matrice de corrélation entre les variables



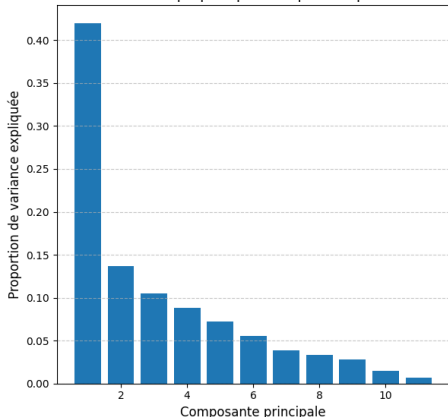
Principe de l'ACP

- **Objectif** : Réduire la dimensionnalité tout en préservant l'information essentielle
- **Principe** : Transformer des variables corrélées en composantes principales non corrélées
- **Avantages** :
 - Simplification des données (de 11 variables à quelques composantes)
 - Visualisation facilitée
 - Identification des tendances principales

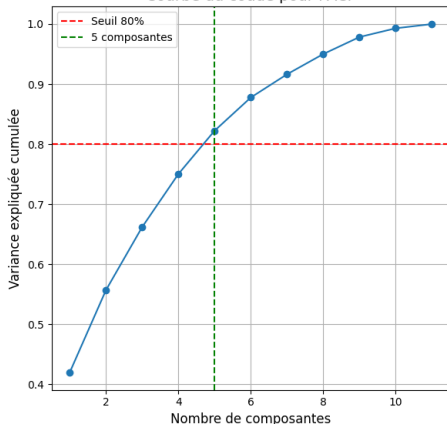


Variance expliquée

Variance expliquée par chaque composante



Courbe du coude pour l'ACP

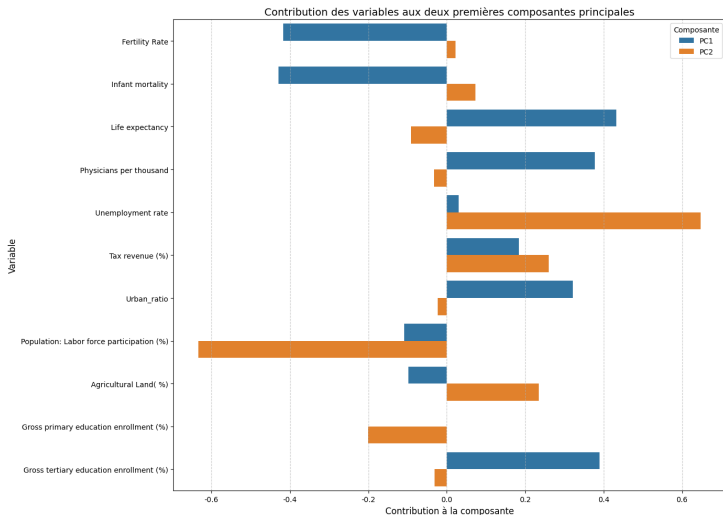


Variance expliquée par composante

- PC1 : **42%** de la variance totale
- PC2 : **14%** de la variance
- Les 5 premières composantes : **>80%** de l'information

Variance cumulée

Interprétation des composantes principales



des variables aux composantes principales

Contributions

Interprétation des composantes principales (suite)

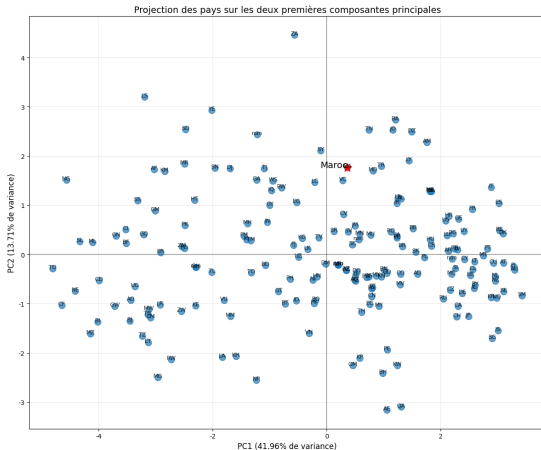
PC1 (42%) Niveau de développement socio-économique

- + Espérance de vie, médecins/1000, scolarisation tertiaire
- - Taux de fécondité, mortalité infantile

PC2 (14%) Marché du travail

- + Taux de chômage
- - Participation à la force de travail

Projection des pays (2D)



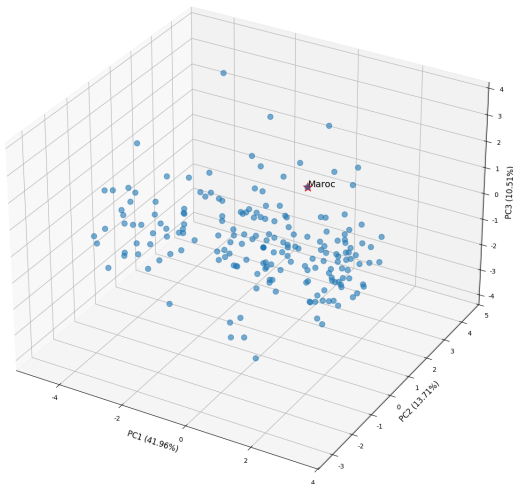
Position du Maroc et
d'autres pays sur PC1 vs PC2

Projection des pays (2D, suite)

- Le Maroc occupe une position **intermédiaire** sur PC1 (développement)
- Position **basse** sur PC2 (problèmes de chômage)
- Proximité avec certains pays d'Amérique latine et MENA

Projection des pays (3D)

Représentation 3D après ACP (3 premières composantes)



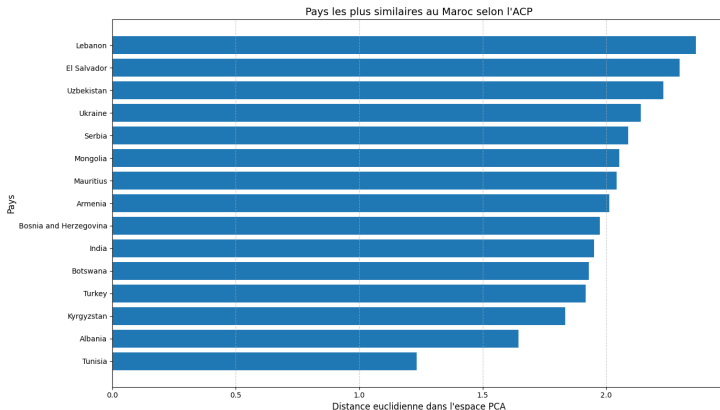
(PC1, PC2, PC3)

Projection 3D des pays

Projection des pays (3D, suite)

- La visualisation 3D confirme la position intermédiaire du Maroc
- PC3 : Position moyenne suggérant une structure économique équilibrée

Pays similaires au Maroc



Les 15 pays

les plus similaires au Maroc (distance euclidienne)

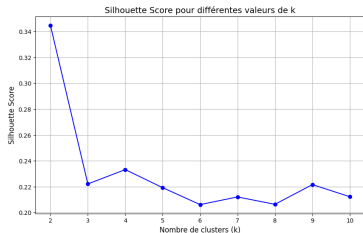
Région MENA Liban, Tunisie

Amérique latine El Salvador

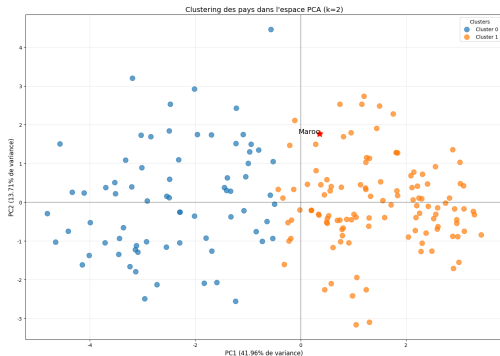
Autres régions Inde, Botswana

Europe de l'Est/Asie centrale Ukraine, Serbie, Arménie, Albanie

Clustering des pays



Score de silhouette



Résultats du clustering K-means (k=2)

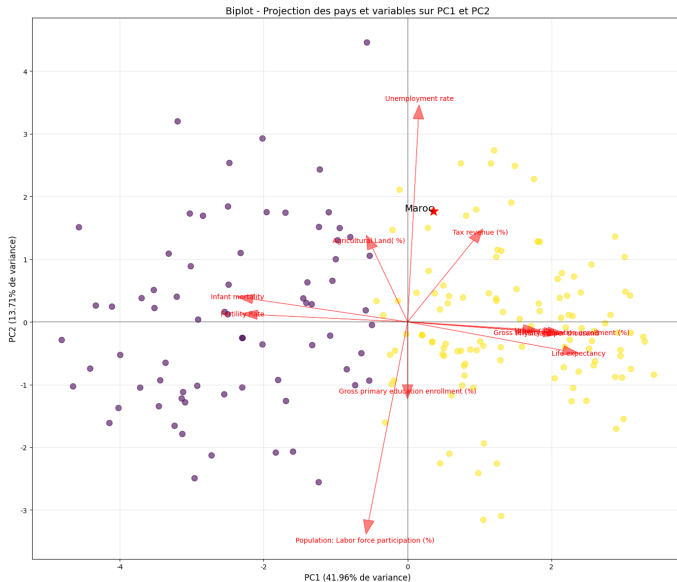
Cluster 0 (bleu) Pays en développement

- Fécondité élevée, espérance de vie plus faible
- Afrique subsaharienne, Asie du Sud, certains pays MENA

Cluster 1 (orange) Pays développés/intermédiaires

- Forte espérance de vie, faible fécondité
- **Le Maroc appartient à ce cluster !**

Position du Maroc - Détail



dans l'espace des composantes principales

Position du Maroc - Détail (suite)

- Position **périphérique** dans son cluster (près de la frontière)
- Proximité avec le vecteur **taux de chômage**
- Plus éloigné des vecteurs **espérance de vie** et **revenus**
- Position reflétant une **transition** démographique et économique

Synthèse des résultats

- **Réduction efficace** : 5 composantes principales expliquent $>80\%$ de la variance
- **Position du Maroc** :
 - Développement intermédiaire (entre pays avancés et en développement)
 - Classé dans le cluster des pays à développement intermédiaire
 - Proximité avec des pays comme Tunisie, Jordanie, Algérie
- **Forces du Maroc** :
 - Transition démographique avancée
 - Scolarisation en progression
- **Défis du Maroc** :
 - Taux de chômage élevé
 - Accès à l'éducation supérieure à renforcer

Implications pour les politiques de développement

- **Priorités recommandées :**

- **Création d'emplois** : Réduire le taux de chômage élevé
- **Enseignement supérieur** : Augmenter l'accès et la qualité
- **Santé publique** : Améliorer l'espérance de vie

- **Apprentissage des pays similaires :**

- Étudier les succès de la Tunisie (éducation)
- Analyser les stratégies de la Colombie (diversification économique)

Perspectives futures

- **Analyse temporelle** : Étudier l'évolution de la position du Maroc au fil du temps
- **Variables supplémentaires** : Intégrer des indicateurs
 - Inégalités
 - Gouvernance
 - Innovation
 - Durabilité environnementale
- **Analyse régionale** : Focus sur la région MENA pour comparaisons affinées

Merci de votre attention

Questions ?

Aymane Kellaa
Mohamed Harbili
Mouaad Ait Ahlal

École Nationale des Sciences Appliquées de Berrechid
Génie Informatique S8
20 Mai 2025