

# Rapport de projet Analyse de données

---

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

---

### Étudiants :

Aymane Kellaa  
Mohamed Harbili  
Mouaad Ait Ahlal

### Professeur :

Dr. Sakat Abdeljalil

Génie Informatique S8  
École Nationale des Sciences Appliquées de Berrechid

20 Mai 2025

## Résumé

Ce rapport présente une analyse approfondie d'un jeu de données mondial visant à comprendre les similitudes et différences entre pays à travers des indicateurs socio-économiques clés. En utilisant l'Analyse en Composantes Principales (ACP) et des algorithmes de clustering, nous avons identifié les variables les plus discriminantes ainsi que les regroupements naturels entre pays. Une attention particulière a été portée à la position du Maroc dans ce paysage mondial, permettant d'identifier ses similarités avec d'autres nations et sa place au sein des clusters formés. Cette étude démontre l'efficacité des méthodes d'analyse multidimensionnelle pour extraire des informations pertinentes à partir de données complexes et met en lumière les caractéristiques distinctives du développement socio-économique du Maroc par rapport au reste du monde.

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Contexte et Problématique	3
1.2	Description du Projet	3
<b>2</b>	<b>Méthodologie</b>	<b>5</b>
2.1	Présentation du jeu de données	5
2.2	Prétraitement des données	5
2.2.1	Nettoyage et conversion des variables	6
2.2.2	Sélection des variables	6
2.2.3	Gestion des valeurs manquantes	6
2.2.4	Standardisation des données	6
2.3	Analyse des corrélations	6
2.4	Analyse en Composantes Principales (ACP)	7
2.4.1	Principes de l'ACP	7
2.4.2	Application et résultats de l'ACP	8
2.4.3	Interprétation des composantes principales	9
<b>3</b>	<b>Résultats et Analyses</b>	<b>11</b>
3.1	Visualisation des pays dans l'espace des composantes principales	11
3.1.1	Projection en 2D	11
3.1.2	Projection en 3D	12
3.2	Analyse de proximité du Maroc avec d'autres pays	13
3.3	Clustering des pays	13
3.3.1	Position du Maroc dans les clusters	15
3.4	Analyse détaillée du profil du Maroc	15
<b>4</b>	<b>Conclusion et Perspectives</b>	<b>17</b>
4.1	Synthèse des résultats	17
4.2	Implications pour les politiques de développement	17
4.3	Limites de l'étude et perspectives futures	18

# Liste des figures

2.1	Carte thermique de la matrice de corrélation des variables sélectionnées. . . . .	7
2.2	Histogramme de la variance expliquée par chaque composante principale. . . . .	8
2.3	Courbe du coude montrant la variance expliquée cumulée avec seuil à 80%. . . . .	9
2.4	Histogramme des contributions des variables aux deux premières composantes principales. . . . .	10
3.1	Nuage de points 2D des pays sur PC1 vs PC2, mettant en évidence le Maroc et certains pays. . . . .	11
3.2	Nuage de points 3D des pays sur PC1, PC2 et PC3, mettant en évidence le Maroc. . . . .	12
3.3	Histogramme horizontal des 15 pays les plus similaires au Maroc basé sur la distance euclidienne dans l'espace ACP. . . . .	13
3.4	Courbe du score de silhouette pour différentes valeurs de k (nombre de clusters). . . . .	14
3.5	Résultats du clustering K-means avec k=2 dans l'espace des deux premières composantes principales. . . . .	14
3.6	Projection des variables et pays dans l'espace des composantes principales, mettant en évidence la position du Maroc. . . . .	16

# Chapitre 1

## Introduction

### 1.1 Contexte et Problématique

Dans un monde de plus en plus interconnecté et complexe, la compréhension des dynamiques socio-économiques entre pays devient cruciale pour les décideurs politiques, les économistes et les chercheurs [6]. Les données mondiales sur différents indicateurs tels que la santé, l'éducation, l'économie et la démographie offrent une mine d'informations qui, lorsqu'elles sont correctement analysées, peuvent révéler des tendances profondes et des relations significatives entre variables et entre pays.

Cependant, l'analyse de ces données pose plusieurs défis majeurs :

- **Dimensionnalité élevée** : Les jeux de données mondiales contiennent souvent des dizaines voire des centaines de variables pour chaque pays, rendant difficile l'identification des tendances sous-jacentes par simple observation.
- **Multicolinéarité** : De nombreuses variables socio-économiques sont fortement corrélées entre elles, ce qui peut compliquer l'interprétation des résultats et la compréhension des relations causales [2].
- **Données manquantes** : Tous les pays ne disposent pas des mêmes capacités de collecte de données, ce qui entraîne des valeurs manquantes qui doivent être traitées avec soin.
- **Hétérogénéité** : Les pays présentent d'énormes disparités en termes de taille, de population, de PIB et d'autres indicateurs, ce qui peut créer des valeurs aberrantes influençant excessivement les analyses.

Dans ce contexte, les méthodes d'analyse multidimensionnelle comme l'Analyse en Composantes Principales (ACP) et les techniques de clustering deviennent particulièrement pertinentes. Elles permettent de réduire la dimensionnalité des données tout en préservant l'information essentielle, et d'identifier des groupes de pays partageant des caractéristiques similaires.

Pour le Maroc en particulier, qui connaît des transformations socio-économiques importantes depuis plusieurs décennies, se situer dans le panorama mondial représente un enjeu stratégique pour comprendre ses forces, ses faiblesses et identifier des modèles de développement potentiellement inspirants.

### 1.2 Description du Projet

Ce projet s'appuie sur un jeu de données mondial de 2023 [4] contenant des informations détaillées sur plus d'une centaine de pays et comprenant des dizaines d'indicateurs socio-économiques.

L'objectif principal est de réaliser une analyse approfondie de ces données en utilisant des techniques avancées d'analyse statistique multivariée pour :

- Identifier les principales dimensions (facteurs) qui expliquent les variations entre les pays du monde
- Positionner le Maroc dans ce paysage mondial et déterminer les pays qui lui sont les plus similaires
- Découvrir des regroupements naturels (clusters) de pays et analyser à quel groupe appartient le Maroc
- Comprendre les forces et faiblesses relatives du Maroc par rapport aux autres pays

Pour atteindre ces objectifs, nous avons mis en œuvre une méthodologie rigoureuse comprenant :

1. Le prétraitement des données, incluant le nettoyage, la gestion des valeurs manquantes et la standardisation

2. L'analyse des corrélations entre les variables pour comprendre leurs interrelations
3. L'application de l'Analyse en Composantes Principales (ACP) pour réduire la dimensionnalité
4. L'utilisation de l'algorithme K-means pour le clustering des pays
5. L'interprétation des résultats avec un focus particulier sur le Maroc

Ce rapport détaille l'ensemble de cette démarche et présente les résultats obtenus, avec une attention particulière portée à la visualisation des données pour faciliter l'interprétation des phénomènes complexes identifiés.

# Chapitre 2

## Méthodologie

### 2.1 Présentation du jeu de données

Le jeu de données utilisé dans cette étude est le "Global Country Information Dataset 2023" [4], qui rassemble des informations détaillées sur des pays du monde entier. Ce jeu de données contient initialement plus d'une centaine de pays et plusieurs dizaines de variables couvrant différents aspects du développement socio-économique.

Voici un aperçu des premières lignes du jeu de données original :

TABLE 2.1 – Aperçu des données du jeu "Countries of the World 2023"

Pays	Densité (P/Km <sup>2</sup> )	Abréviation	Terres agricoles (%)	Superficie (Km <sup>2</sup> )	Forces ar- mées	Taux de natalité	Capitale
Afghanistan	60	AF	58,10%	652 230	323 000	32,49	Kaboul
Albanie	105	AL	43,10%	28 748	9 000	11,78	Tirana
Algérie	18	DZ	17,40%	2 381 741	317 000	24,28	Alger

TABLE 2.2 – Suite des données (variables supplémentaires)

Pays	Médecins pour 1000	Population	Part. force de travail (%)	Revenus fiscaux (%)	Taux de chômage	Population urbaine	Latitude
Afghanistan	0,28	38 041 754	48,90%	9,30%	11,12%	9 797 273	33,939110
Albanie	1,20	2 854 191	55,70%	18,60%	12,33%	1 747 593	41,153332
Algérie	1,72	43 053 054	41,20%	37,20%	11,70%	31 510 100	28,033886

Parmi les variables disponibles dans ce jeu de données, on trouve :

- Des indicateurs démographiques (population totale, population urbaine, taux de fécondité)
- Des indicateurs de santé (espérance de vie, mortalité infantile, nombre de médecins par habitant)
- Des indicateurs économiques (taux de chômage, revenus fiscaux)
- Des indicateurs d'éducation (taux de scolarisation primaire et tertiaire)
- Des indicateurs territoriaux (pourcentage de terres agricoles)
- Des indicateurs de travail (taux de participation à la population active)

Ce jeu de données riche et varié permet d'avoir une vision globale et multidimensionnelle des différences entre pays, ce qui en fait un excellent candidat pour l'application de méthodes d'analyse statistique multivariée comme l'ACP.

### 2.2 Prétraitement des données

Le prétraitement des données constitue une étape cruciale de notre analyse, car la qualité des résultats dépend directement de la qualité des données d'entrée. Nous avons procédé aux étapes suivantes :

### 2.2.1 Nettoyage et conversion des variables

Plusieurs défis ont dû être relevés lors du nettoyage des données :

- **Conversion des types de données** : De nombreuses variables étaient initialement au format texte, avec des caractères spéciaux comme les pourcentages (%) ou les séparateurs de milliers (,). Nous avons systématiquement converti ces variables en format numérique.
- **Création de variables dérivées** : Nous avons calculé le ratio de population urbaine en pourcentage (Urban\_ratio) à partir des variables de population urbaine et population totale, pour obtenir un indicateur plus pertinent pour les comparaisons entre pays.
- **Gestion des valeurs aberrantes** : Nous avons identifié et supprimé certains pays considérés comme des valeurs aberrantes (outliers) qui pouvaient influencer excessivement les résultats de l'analyse : Somalie, Grèce, Îles Salomon.

### 2.2.2 Sélection des variables

Pour notre analyse, nous avons sélectionné un sous-ensemble de variables jugées pertinentes pour caractériser les pays :

- **Taux de fécondité** (Fertility Rate) : Nombre moyen d'enfants par femme
- **Mortalité infantile** (Infant mortality) : Nombre de décès d'enfants de moins d'un an pour 1000 naissances vivantes
- **Espérance de vie** (Life expectancy) : Nombre moyen d'années qu'un nouveau-né peut espérer vivre
- **Médecins pour mille habitants** (Physicians per thousand) : Densité de médecins dans le pays
- **Taux de chômage** (Unemployment rate) : Pourcentage de la population active sans emploi
- **Revenus fiscaux** (Tax revenue (%)) : Pourcentage du PIB représenté par les recettes fiscales
- **Population urbaine** (Urban\_ratio) : Pourcentage de la population vivant en zone urbaine
- **Taux de participation à la population active** (Population : Labor force participation (%))
- **Pourcentage de terres agricoles** (Agricultural Land( %))
- **Taux brut de scolarisation primaire** (Gross primary education enrollment (%))
- **Taux brut de scolarisation dans le supérieur** (Gross tertiary education enrollment (%))

### 2.2.3 Gestion des valeurs manquantes

Les valeurs manquantes représentent un défi important dans les jeux de données internationaux [1]. Notre approche pour les gérer a consisté à :

- Éliminer les colonnes contenant trop de valeurs manquantes (seuil fixé à 30%)
- Pour les colonnes conservées, remplacer les valeurs manquantes par la médiane de la variable concernée

Ce compromis nous a permis de conserver un maximum d'information tout en évitant les biais excessifs.

### 2.2.4 Standardisation des données

Pour l'application de l'ACP, il est essentiel que toutes les variables soient sur une échelle comparable, sans quoi les variables avec les valeurs numériques les plus élevées domineraient l'analyse. Nous avons donc standardisé toutes les variables (moyenne de 0 et écart-type de 1) en utilisant la formule :

$$z = \frac{x - \mu}{\sigma}$$

où  $x$  est la valeur originale,  $\mu$  est la moyenne de la variable et  $\sigma$  son écart-type.

## 2.3 Analyse des corrélations

Avant de procéder à l'ACP, nous avons examiné les corrélations entre les variables sélectionnées pour mieux comprendre leurs interrelations. La matrice de corrélation est présentée dans la Figure 1.



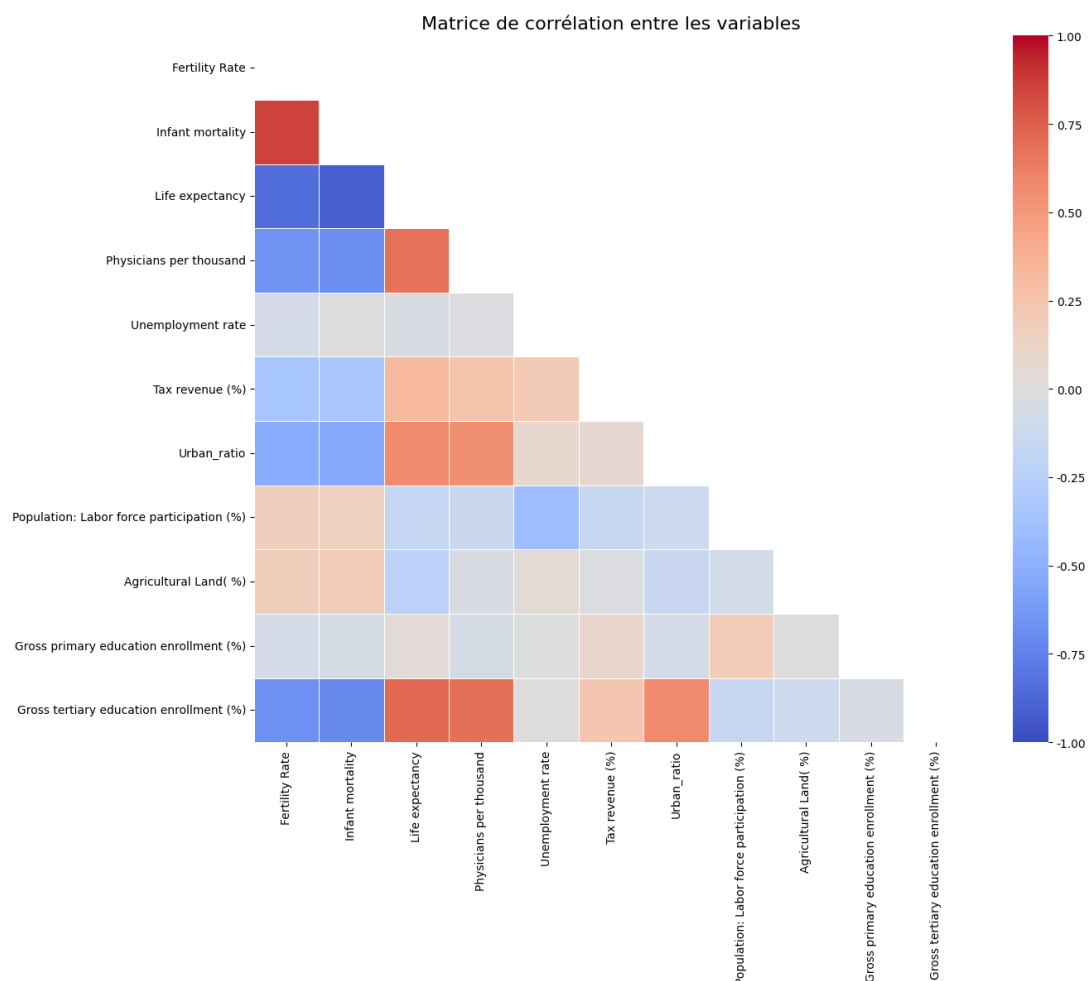


FIGURE 2.1 – Carte thermique de la matrice de corrélation des variables sélectionnées.

Cette analyse des corrélations a révélé plusieurs relations importantes entre les variables :

- Une forte corrélation négative entre le taux de fécondité et l'espérance de vie, indiquant que les pays avec une fécondité élevée ont généralement une espérance de vie plus faible
- Une corrélation positive entre l'espérance de vie et le nombre de médecins pour 1000 habitants, ce qui est cohérent avec l'intuition que l'accès aux soins médicaux améliore la longévité
- Une corrélation négative entre le taux de fécondité et le taux de scolarisation tertiaire, suggérant que l'accès à l'éducation supérieure est associé à une baisse de la natalité
- Une corrélation positive entre le taux d'urbanisation et plusieurs indicateurs de développement comme l'espérance de vie et le taux de scolarisation tertiaire

Ces corrélations justifient pleinement l'utilisation de l'ACP, qui permettra de synthétiser ces relations complexes en un nombre réduit de dimensions.

## 2.4 Analyse en Composantes Principales (ACP)

### 2.4.1 Principes de l'ACP

L'Analyse en Composantes Principales (ACP) est une technique statistique multivariée qui transforme un ensemble de variables possiblement corrélées en un ensemble de variables linéairement non corrélées appelées composantes principales [3]. Les objectifs principaux de l'ACP sont :

- Réduire la dimensionnalité d'un jeu de données contenant un grand nombre de variables interreliées
- Retenir autant d'information que possible de l'ensemble de données original
- Faciliter l'interprétation des données en identifiant les directions de plus grande variance

Dans notre contexte, l'ACP nous permet de synthétiser l'information contenue dans les 11 variables sélectionnées en un nombre plus restreint de composantes principales, facilitant ainsi la visualisation et

l'interprétation des différences entre pays.

## 2.4.2 Application et résultats de l'ACP

Nous avons appliqué l'ACP aux données standardisées et avons obtenu les composantes principales. Pour déterminer le nombre de composantes à retenir, nous avons analysé la variance expliquée par chaque composante et la variance cumulée.

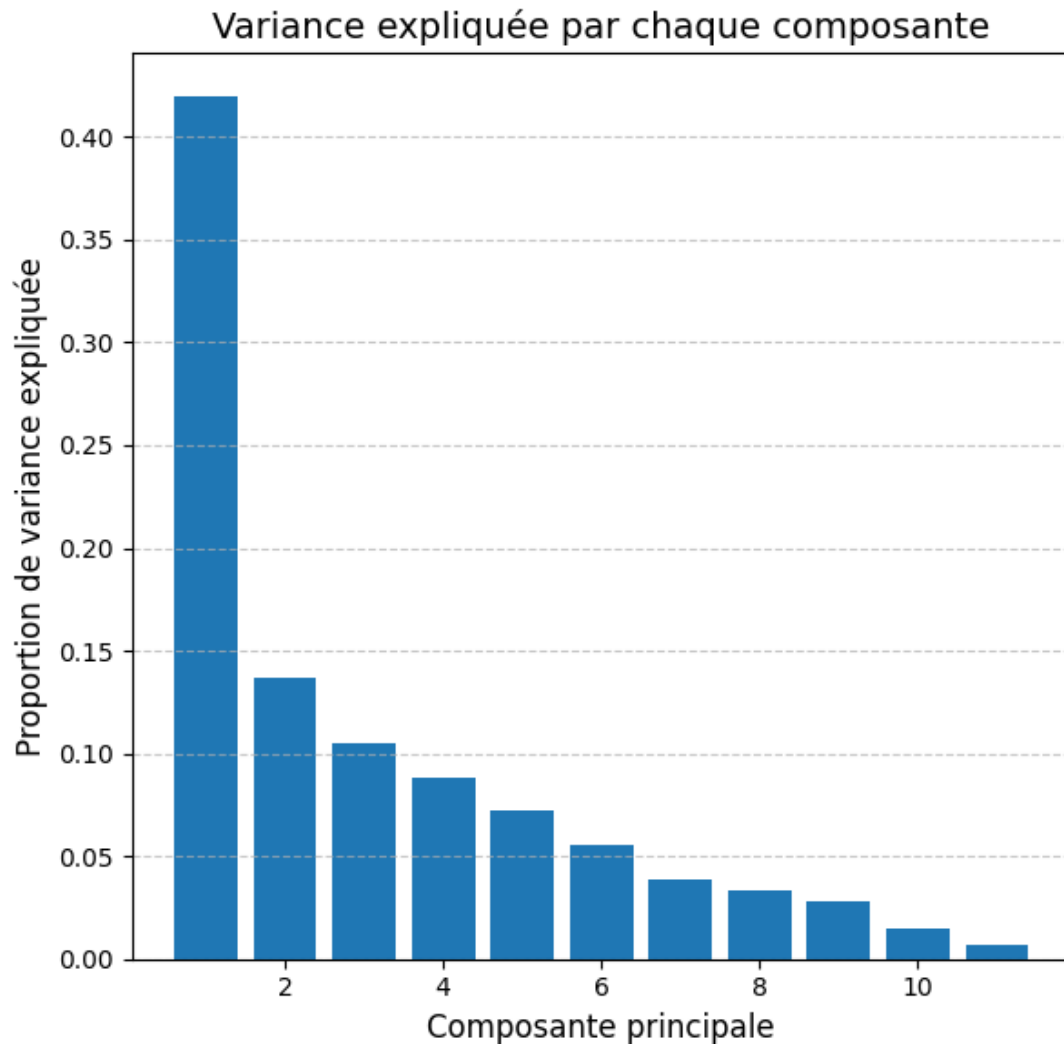


FIGURE 2.2 – Histogramme de la variance expliquée par chaque composante principale.

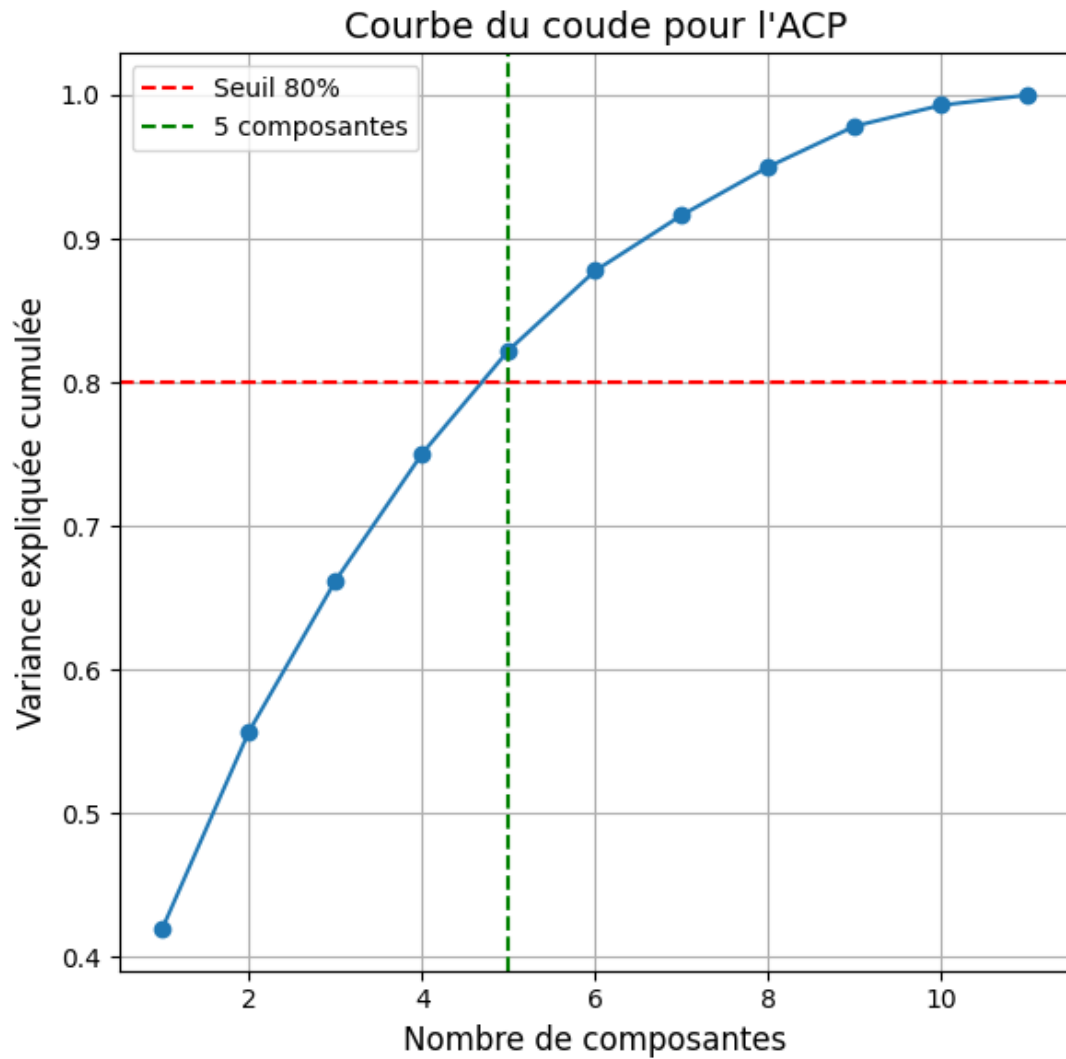


FIGURE 2.3 – Courbe du coude montrant la variance expliquée cumulée avec seuil à 80%.

L'analyse de la variance expliquée montre que :

- La première composante principale explique environ 42% de la variance totale
- La deuxième composante explique environ 14% de la variance
- Les 5 premières composantes expliquent plus de 80% de la variance totale

Ce résultat est très satisfaisant, car il indique que nous pouvons réduire considérablement la dimensionnalité du jeu de données (de 11 variables à 5 composantes principales) tout en conservant plus de 80% de l'information originale.

### 2.4.3 Interprétation des composantes principales

Pour comprendre la signification des composantes principales, nous avons analysé la contribution des variables originales à chacune d'elles.

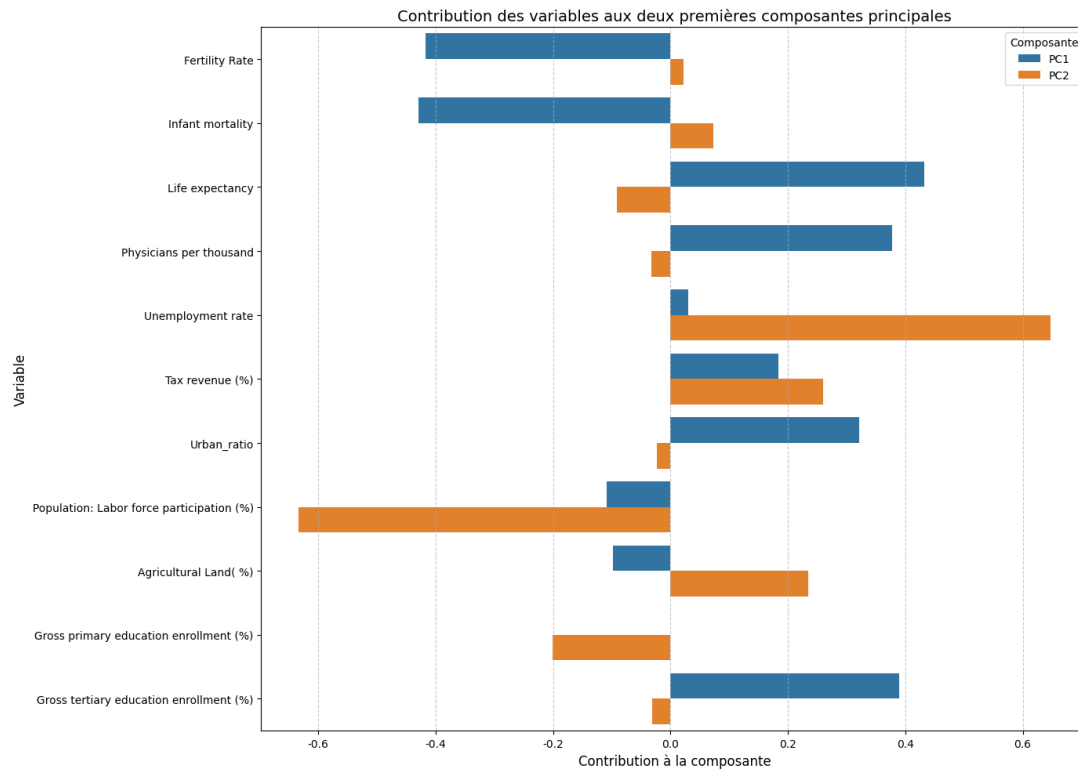


FIGURE 2.4 – Histogramme des contributions des variables aux deux premières composantes principales.

L'analyse des contributions nous permet d'interpréter les composantes principales comme suit :

- **Première composante principale (PC1)** : Fortement corrélée positivement avec l'espérance de vie, le taux de médecins pour 1000 habitants, le taux de scolarisation tertiaire et le taux d'urbanisation, et négativement avec le taux de fécondité et la mortalité infantile. Cette composante semble représenter le niveau global de développement socio-économique et sanitaire.
- **Deuxième composante principale (PC2)** : Corrélée positivement avec le taux de chômage et négativement avec le taux de participation à la population active. Cette composante semble capturer les aspects liés au marché du travail et à l'emploi.

Cette interprétation des composantes principales nous permet de donner un sens aux projections des pays dans l'espace réduit de l'ACP.

## Chapitre 3

# Résultats et Analyses

### 3.1 Visualisation des pays dans l'espace des composantes principales

#### 3.1.1 Projection en 2D

La Figure 5 présente la projection des pays sur les deux premières composantes principales, avec une mise en évidence particulière du Maroc.

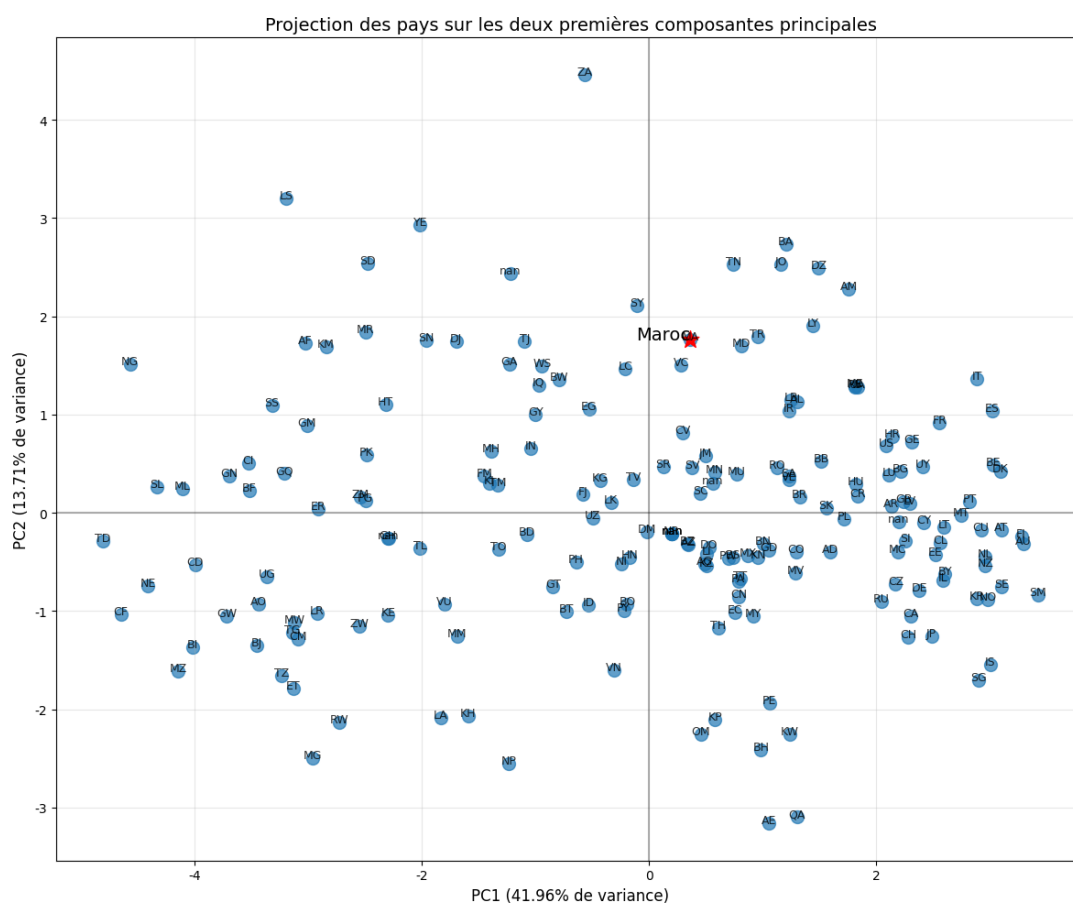


FIGURE 3.1 – Nuage de points 2D des pays sur PC1 vs PC2, mettant en évidence le Maroc et certains pays.

Cette visualisation nous permet de tirer plusieurs observations :

- Le Maroc se situe dans une position intermédiaire sur la première composante principale (PC1),

ce qui suggère un niveau de développement moyen en termes d'indicateurs socio-économiques et sanitaires.

- Sur la deuxième composante principale (PC2), le Maroc se trouve dans la partie inférieure, ce qui pourrait indiquer un taux de chômage relativement élevé par rapport à d'autres pays de niveau de développement comparable.
- On observe que le Maroc se trouve relativement proche de certains pays d'Amérique latine et d'autres pays du Moyen-Orient et d'Afrique du Nord, suggérant des similarités dans les profils socio-économiques.
- Les pays les plus développés (Europe occidentale, Amérique du Nord, Australie, Japon) se trouvent généralement à droite du graphique (valeurs élevées de PC1), tandis que les pays les moins développés se trouvent à gauche.

### 3.1.2 Projection en 3D

Pour compléter notre analyse, nous avons également examiné la projection des pays dans l'espace des trois premières composantes principales, ce qui permet de capturer environ 66% de la variance totale.

Représentation 3D après ACP (3 premières composantes)

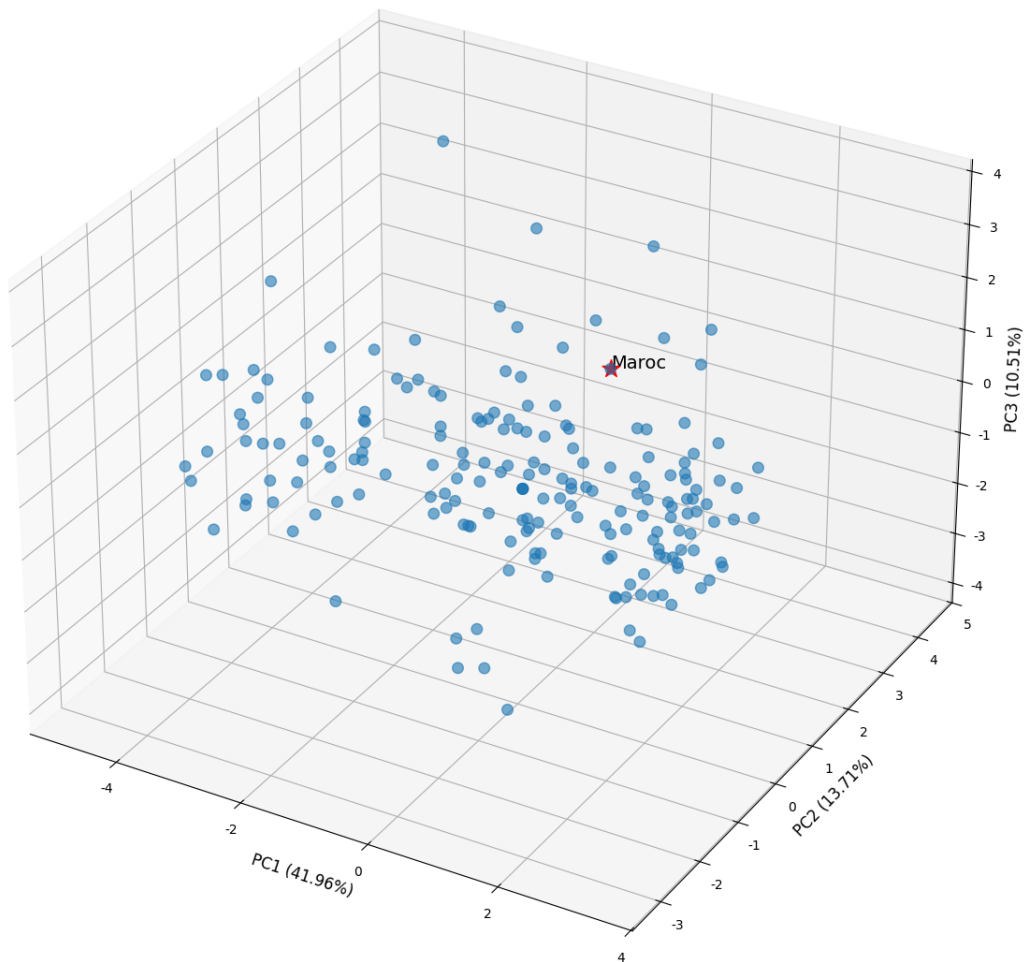


FIGURE 3.2 – Nuage de points 3D des pays sur PC1, PC2 et PC3, mettant en évidence le Maroc.

Cette visualisation en 3D confirme nos observations précédentes et apporte des informations supplémentaires sur la position du Maroc par rapport à la troisième composante principale. Le Maroc semble avoir une valeur plutôt moyenne sur cette composante, ce qui suggère une structure économique équilibrée entre agriculture et autres secteurs, ainsi qu'une capacité fiscale modérée.

## 3.2 Analyse de proximité du Maroc avec d'autres pays

Pour identifier les pays les plus similaires au Maroc dans l'espace des composantes principales, nous avons calculé les distances euclidiennes entre le Maroc et tous les autres pays dans l'espace des 5 premières composantes principales.

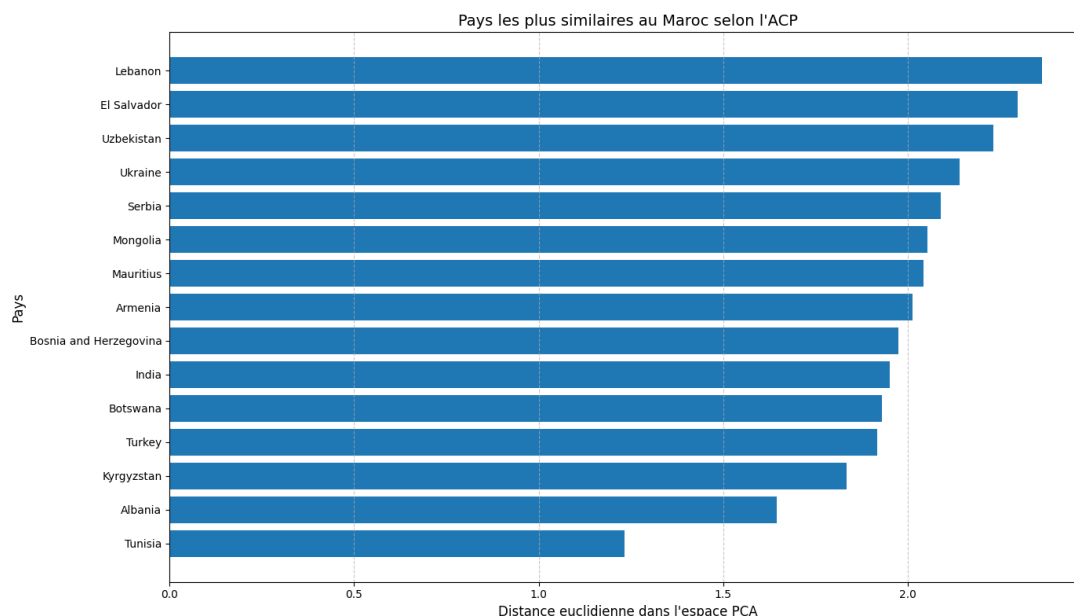


FIGURE 3.3 – Histogramme horizontal des 15 pays les plus similaires au Maroc basé sur la distance euclidienne dans l'espace ACP.

Cette analyse révèle que les pays les plus similaires au Maroc sont principalement :

- Des pays du Moyen-Orient et d'Afrique du Nord comme le Liban et la Tunisie, reflétant une proximité géographique et des similitudes culturelles et socio-économiques.
- Des pays d'autres régions, comme El Salvador, l'Inde et le Botswana, qui partagent avec le Maroc un niveau de développement intermédiaire et des caractéristiques socio-économiques similaires.
- Des pays d'Europe de l'Est et d'Asie centrale, comme l'Ukraine, la Serbie, l'Arménie, l'Albanie et l'Ouzbékistan, présentant des indicateurs de développement comparables.

Cette analyse de proximité est particulièrement intéressante car elle permet d'identifier des pays qui, malgré des différences géographiques, culturelles ou historiques, partagent avec le Maroc des défis et des caractéristiques similaires en termes de développement socio-économique.

## 3.3 Clustering des pays

Pour compléter notre analyse, nous avons appliqué un algorithme de clustering (K-means) aux données projetées dans l'espace des 5 premières composantes principales [5]. Pour déterminer le nombre optimal de clusters, nous avons utilisé la méthode du score de silhouette.

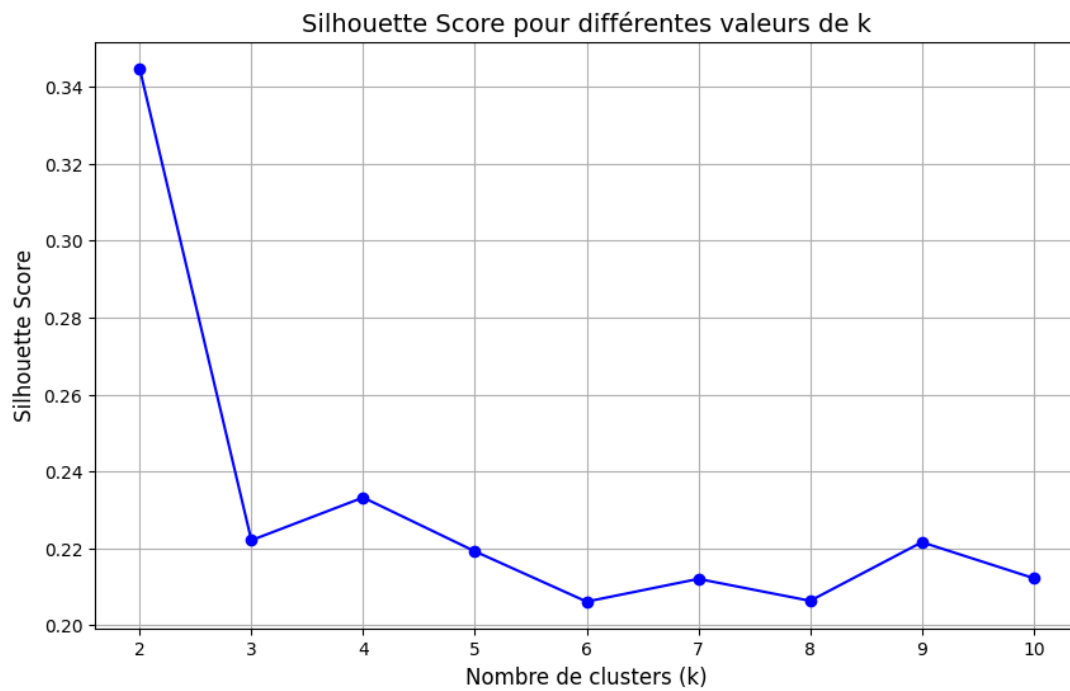


FIGURE 3.4 – Courbe du score de silhouette pour différentes valeurs de  $k$  (nombre de clusters).

L'analyse du score de silhouette a indiqué que le nombre optimal de clusters était  $k=2$ . Nous avons donc procédé à un clustering K-means avec 2 clusters et avons visualisé les résultats dans l'espace des deux premières composantes principales.

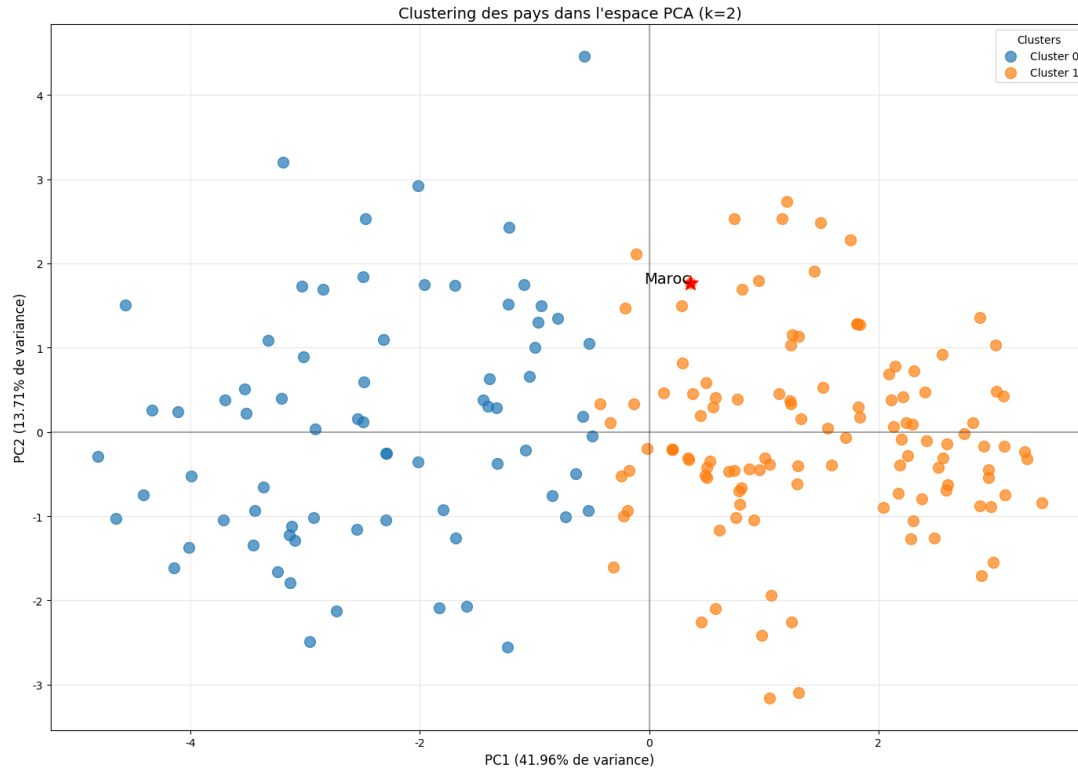


FIGURE 3.5 – Résultats du clustering K-means avec  $k=2$  dans l'espace des deux premières composantes principales.

Les deux clusters identifiés peuvent être interprétés comme suit :



- **Cluster 0 (bleu)** : Pays en développement ou à développement intermédiaire, caractérisés par des taux de fécondité plus élevés, une espérance de vie plus faible, et un accès plus limité aux soins de santé et à l'éducation supérieure. Ce groupe comprend principalement des pays d'Afrique subsaharienne, d'Asie du Sud et certains pays d'Amérique latine et du Moyen-Orient.
- **Cluster 1 (orange)** : Pays développés et à développement avancé, caractérisés par une forte espérance de vie, un faible taux de fécondité, un taux élevé de médecins par habitant et un fort taux de scolarisation tertiaire. Ce groupe comprend les pays d'Europe occidentale, d'Amérique du Nord, ainsi que le Japon, l'Australie, la Nouvelle-Zélande, et certains pays émergents dont le Maroc. L'inclusion du Maroc dans ce cluster témoigne des progrès significatifs réalisés par le pays en matière de développement socio-économique.

### 3.3.1 Position du Maroc dans les clusters

Le Maroc a été assigné au cluster 1 (orange), ce qui indique une tendance vers des indicateurs de développement plus avancés. L'analyse de la position exacte du Maroc dans l'espace des composantes principales révèle toutefois une position nuancée.

Le Maroc se situe à une distance considérable des pays hautement développés comme l'Allemagne, le Royaume-Uni et les États-Unis, particulièrement en ce qui concerne les variables liées à l'espérance de vie et au revenu brut. Il est également relativement éloigné de l'axe représentant le taux de scolarisation primaire.

Au sein de son cluster, le Maroc se distingue par sa proximité avec le vecteur du taux de chômage, indiquant des défis importants dans ce domaine. Il est également parmi les pays du cluster 1 les plus éloignés de l'axe de participation au marché du travail.

Le Maroc occupe une position relativement périphérique dans son cluster, se rapprochant de la frontière avec le cluster 0. Cette position "frontalière" suggère que, bien qu'ayant réalisé des progrès significatifs justifiant son inclusion dans le cluster des pays plus développés, le Maroc conserve certaines caractéristiques des pays en développement.

D'autres pays du Moyen-Orient comme le Qatar, les Émirats arabes unis et Bahreïn se trouvent également dans le cluster 1, mais sont positionnés différemment, plus proches de l'axe de participation au marché du travail. Ces pays présentent des taux de chômage moins élevés tout en faisant face à leurs propres défis en matière d'espérance de vie et de revenus.

Les pays géographiquement ou économiquement similaires au Maroc, tels que la Jordanie, la Tunisie et la Serbie, présentent les positions les plus comparables dans l'espace des composantes principales.

## 3.4 Analyse détaillée du profil du Maroc

Pour approfondir notre compréhension de la position du Maroc dans ce paysage mondial, nous avons examiné la répartition des principales variables socio-économiques dans l'espace des composantes principales, illustrée par les vecteurs directionnels dans la Figure 3.6.

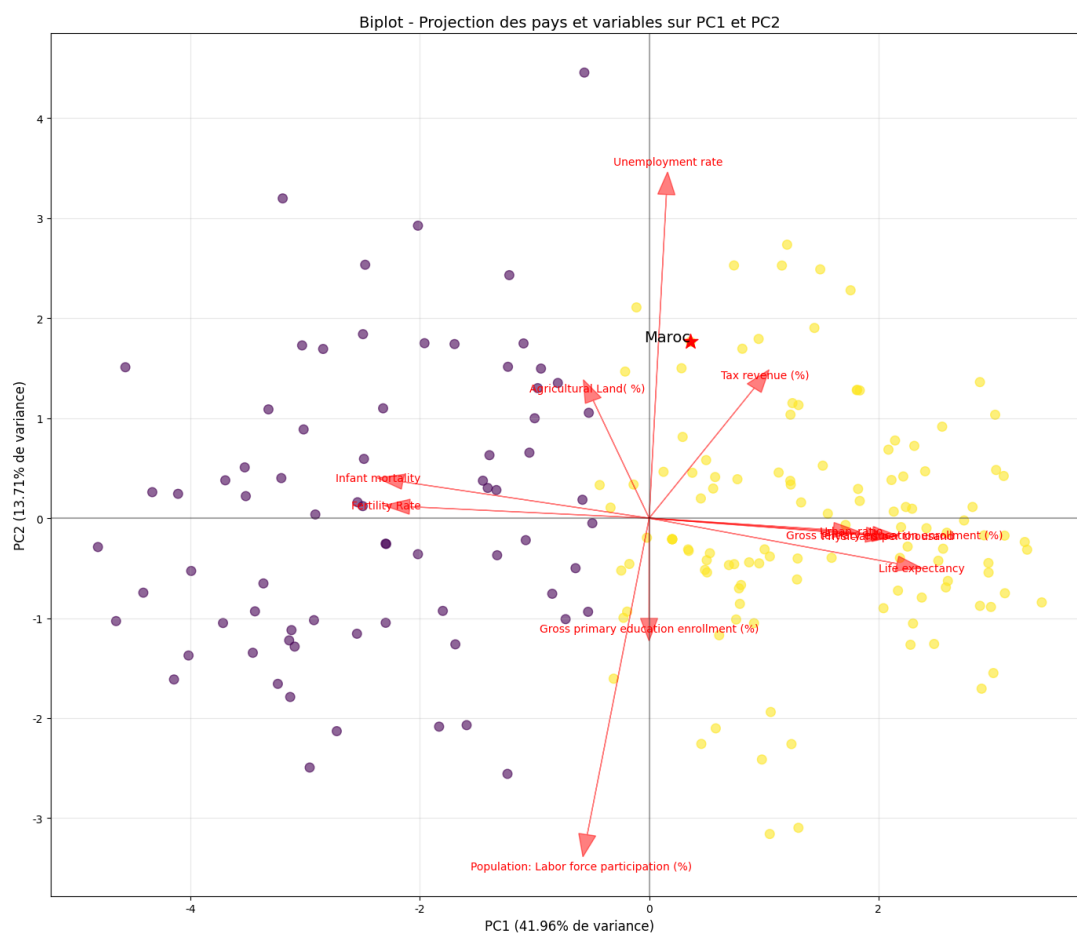


FIGURE 3.6 – Projection des variables et pays dans l'espace des composantes principales, mettant en évidence la position du Maroc.

Cette analyse met en évidence plusieurs caractéristiques intéressantes du profil socio-économique du Maroc :

- Le Maroc présente un taux de chômage significativement élevé par rapport aux autres pays de son cluster, comme l'indique sa proximité avec le vecteur représentant cette variable.
- Sa position relativement éloignée des vecteurs de l'espérance de vie et du revenu brut signale des écarts persistants avec les pays les plus développés de son cluster.
- Le taux de scolarisation tertiaire du Maroc, bien que s'améliorant, reste inférieur à celui de nombreux autres pays du cluster 1.
- La position intermédiaire du Maroc entre les deux clusters reflète sa transition démographique et économique en cours, avec un taux de fécondité en baisse mais encore supérieur à celui des pays les plus développés.
- Le Maroc se distingue par un pourcentage de terres agricoles relativement important, soulignant le poids continu de ce secteur dans son économie.

Ces observations confirment la position du Maroc comme un pays en transition, qui a réalisé des progrès significatifs dans plusieurs domaines du développement socio-économique, mais qui fait encore face à des défis importants, notamment en matière d'emploi, d'accès à l'éducation supérieure et d'amélioration du niveau de vie.

## Chapitre 4

# Conclusion et Perspectives

### 4.1 Synthèse des résultats

Cette étude a permis d'analyser en profondeur la position du Maroc dans le contexte mondial à travers l'application de techniques d'analyse multidimensionnelle à un jeu de données comprenant de multiples indicateurs socio-économiques pour plus d'une centaine de pays.

Les principaux résultats de notre analyse peuvent être résumés comme suit :

1. **Réduction de dimensionnalité efficace** : L'Analyse en Composantes Principales a permis de réduire efficacement la dimensionnalité des données, avec les 5 premières composantes principales expliquant plus de 80% de la variance totale. Cette réduction facilite considérablement l'interprétation des données et la visualisation des similitudes et différences entre pays.
2. **Position intermédiaire du Maroc** : Dans l'espace des composantes principales, le Maroc occupe une position intermédiaire en termes de développement socio-économique global, se situant entre les pays hautement développés et les pays les moins développés.
3. **Similitudes avec d'autres pays** : Notre analyse a identifié les pays les plus similaires au Maroc, notamment la Tunisie, la Jordanie, l'Algérie, le Pérou et la Colombie. Ces similitudes offrent des opportunités d'apprentissage mutuel et de coopération dans le domaine du développement.
4. **Classification dans un cluster de développement intermédiaire** : Le clustering K-means a classé le Maroc dans un groupe de pays à développement intermédiaire, caractérisés par des taux d'urbanisation moyens à élevés et des défis en matière d'emploi.
5. **Forces et faiblesses relatives** : L'analyse comparative des indicateurs du Maroc avec les moyennes de son cluster et les moyennes mondiales a mis en évidence ses forces relatives (espérance de vie, transition démographique avancée) et ses défis (taux de chômage élevé, accès à l'éducation supérieure).

Ces résultats permettent de dresser un portrait nuancé du développement socio-économique du Maroc, soulignant à la fois les progrès réalisés et les défis restants.

### 4.2 Implications pour les politiques de développement

Les résultats de cette analyse ont plusieurs implications importantes pour les politiques de développement au Maroc :

- **Priorité à la création d'emplois** : Le taux de chômage relativement élevé du Maroc par rapport à des pays similaires suggère que les politiques de création d'emplois devraient rester une priorité nationale.
- **Développement de l'enseignement supérieur** : Bien que le taux de scolarisation tertiaire du Maroc soit supérieur à la moyenne mondiale, il reste inférieur à celui des pays de son cluster, suggérant la nécessité de poursuivre les efforts pour développer l'accès à l'enseignement supérieur.
- **Apprentissage des pays similaires** : Les pays identifiés comme les plus similaires au Maroc peuvent servir de points de référence utiles pour l'élaboration de politiques. Par exemple, étudier les succès de la Tunisie en matière d'éducation ou de la Colombie en matière de diversification économique pourrait offrir des enseignements précieux.

- **Amélioration des indicateurs de santé** : L'espérance de vie relativement faible du Maroc par rapport à son cluster souligne la nécessité de renforcer les efforts en matière de santé publique pour combler cet écart et améliorer les résultats.

### 4.3 Limites de l'étude et perspectives futures

Comme toute analyse statistique, notre étude présente certaines limites qu'il convient de reconnaître :

- **Sélection des variables** : Notre analyse se base sur un sous-ensemble de variables disponibles dans le jeu de données. D'autres variables pertinentes comme les inégalités de revenus, les infrastructures, ou les indicateurs environnementaux n'ont pas été incluses faute de données complètes.
- **Données manquantes** : Malgré nos efforts pour gérer les valeurs manquantes de manière appropriée, leur présence peut avoir introduit certains biais dans l'analyse.
- **Dynamique temporelle** : Notre analyse offre une image statique de la situation en 2023. Une analyse de l'évolution temporelle des indicateurs permettrait de mieux comprendre les trajectoires de développement des différents pays, y compris le Maroc.

Pour les recherches futures, plusieurs pistes s'avèrent prometteuses :

- **Analyse temporelle** : Étudier l'évolution de la position du Maroc dans l'espace des composantes principales au fil du temps permettrait de mesurer sa trajectoire de développement et d'évaluer l'efficacité des politiques mises en œuvre.
- **Inclusion de variables supplémentaires** : Intégrer des indicateurs d'inégalités, de gouvernance, d'innovation ou de durabilité environnementale enrichirait l'analyse et offrirait une vision plus complète du développement.
- **Analyse régionale** : Une analyse plus fine focalisée sur la région MENA (Moyen-Orient et Afrique du Nord) pourrait révéler des nuances importantes dans les profils de développement des pays de cette région.
- **Application d'autres méthodes** : L'utilisation de méthodes complémentaires comme l'analyse discriminante ou les modèles de mélange gaussien pourrait offrir des perspectives additionnelles sur les regroupements de pays.

En conclusion, cette étude a démontré la puissance des méthodes d'analyse multidimensionnelle pour comprendre la position d'un pays comme le Maroc dans le contexte mondial. En identifiant les dimensions clés du développement socio-économique et en positionnant le Maroc par rapport à ces dimensions, nous avons pu dresser un portrait nuancé de ses forces et défis, ouvrant ainsi des pistes de réflexion pour les politiques futures de développement.

# Bibliographie

# Bibliographie

- [1] García-Laencina, P. J., Sancho-Gómez, J. L., & Figueiras-Vidal, A. R. (2010). *Pattern classification with missing data : a review*. Neural Computing and Applications, 19(2), 263-282.
- [2] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York : Springer.
- [3] Jolliffe, I. T., & Cadima, J. (2016). *Principal component analysis : a review and recent developments*. Philosophical Transactions of the Royal Society A : Mathematical, Physical and Engineering Sciences, 374(2065), 20150202.
- [4] Nelgiryewithana, L. (2023). *Global Country Information Dataset 2023*. Kaggle. <https://www.kaggle.com/datasets/nelgiryewithana/countries-of-the-world-2023>
- [5] Ng, A. Y., Jordan, M. I., & Weiss, Y. (2002). *On spectral clustering : Analysis and an algorithm*. Advances in neural information processing systems, 14.
- [6] Shmueli, G. (2017). *Research Dilemmas with Behavioral Big Data*. Big data, 5(2), 98-119.