



# Projet 2

Analyse des données de systèmes éducatifs

Guillaume LAFON



# Sommaire

- 1 Introduction
- 2 Qualité et pertinence des données
- 3 Sélection des indicateurs
- 4 Analyse exploratoire des indicateurs
- 5 Scoring
- 6 Présentation des résultats
- 7 Conclusion



# 1 Introduction

## 1.1 Contexte

L'entreprise academy a pour objectif de s'implanter à l'étranger

**Mon objectif:** Définir une liste de pays à fort potentiel pour l'installation de l'entreprise

## 1.2 Données

Données éducatives mondiales issues de la Banque mondiale (EdStats All Indicator Query)

## 1.3 Mission

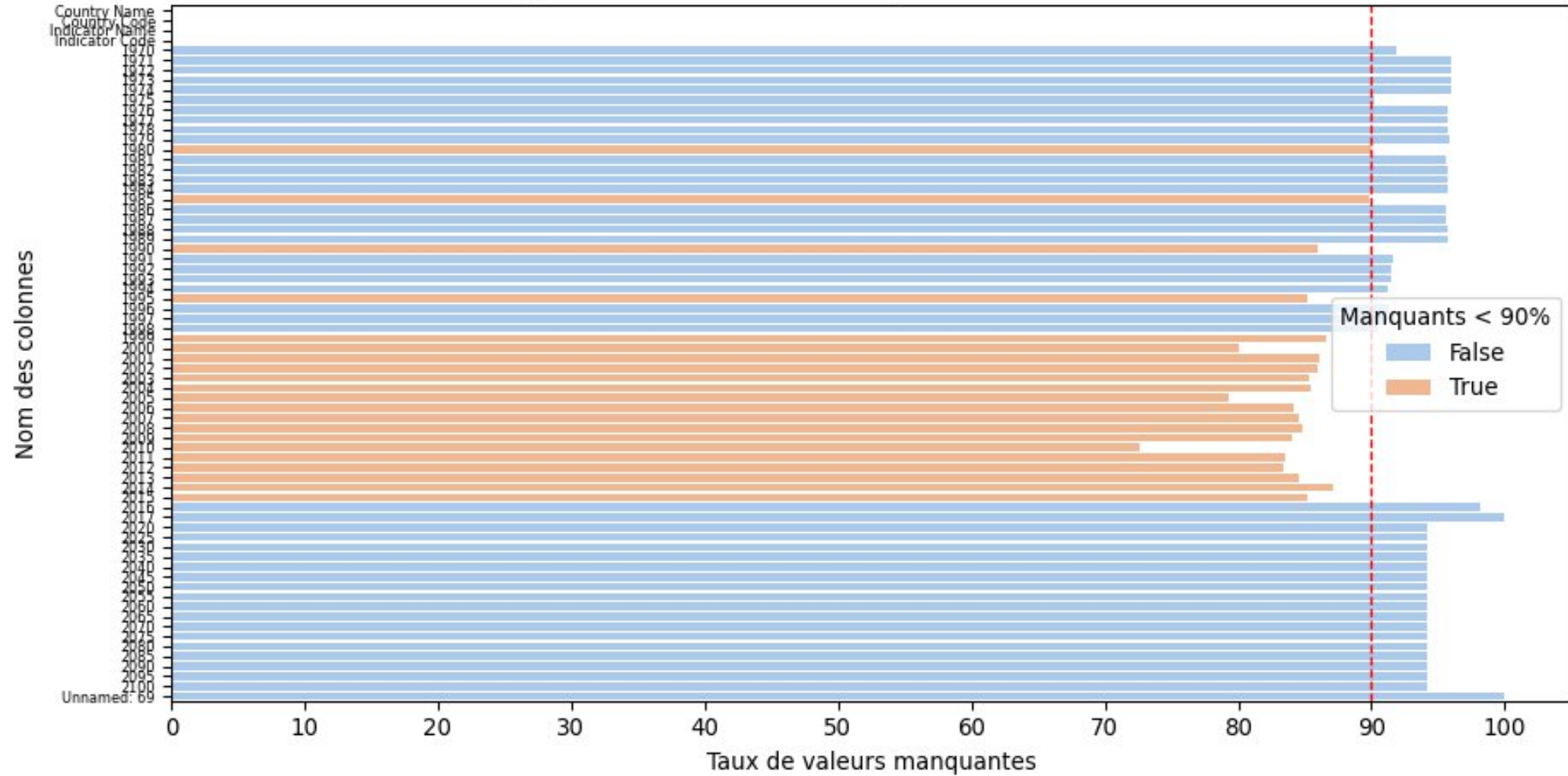
Pré-Analyse exploratoire des données pour:

- Évaluer si les données de la Banque mondiale sont pertinentes pour guider l'expansion internationale de l'entreprise
- Déterminer une liste de 5 pays dans lesquels l'entreprise doit opérer en priorité.

## 2 Qualité et pertinence des données



## Répartition des valeurs manquantes dans EdStatsData



## Valeurs manquantes



- 3365 indicateurs
- 215 pays répartis dans 5 régions
- [2010 - 2015]



# 3 Sélection des indicateurs

Sélection de 6 indicateurs parmi les 3365.



## 3.1 Accès à la technologie :

- *Internet users (per 100 people)*





## 3.2 Répartition de la population :

- *Population, ages 15-64 (% of total)*
- *Population, total*



## 3.3 Capacité financière des pays :

- *GNI per capita, PPP (current international \$)*

## 3.4 Investissement public dans l'éducation :

- *Government expenditure on education as % of GDP (%)*



# 4 Analyse exploratoire des indicateurs

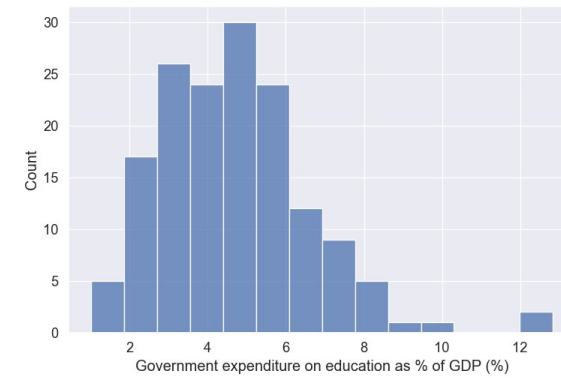
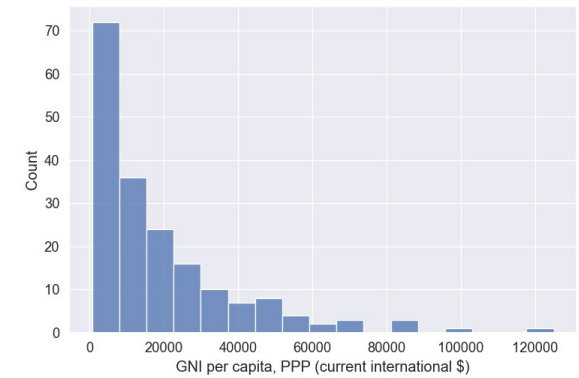
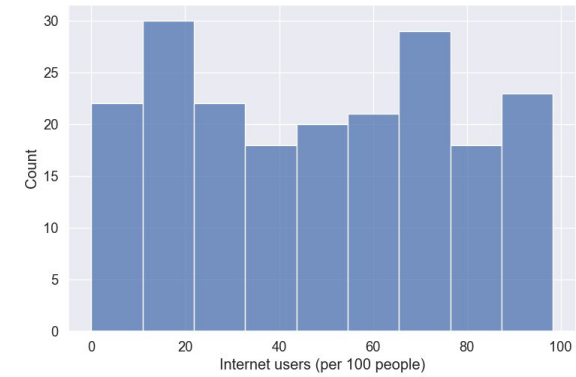
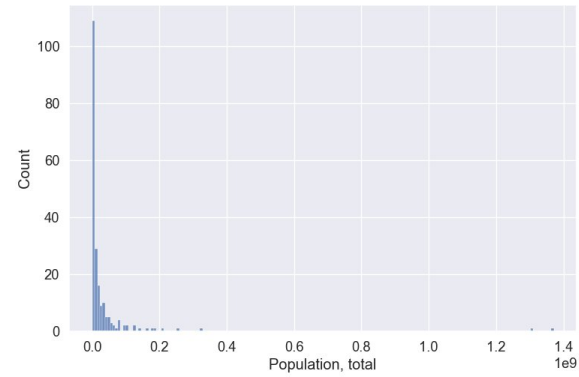
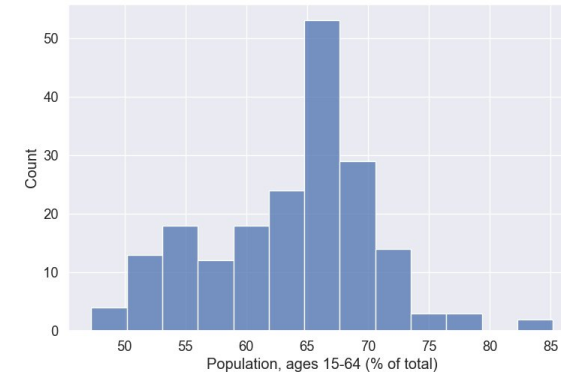
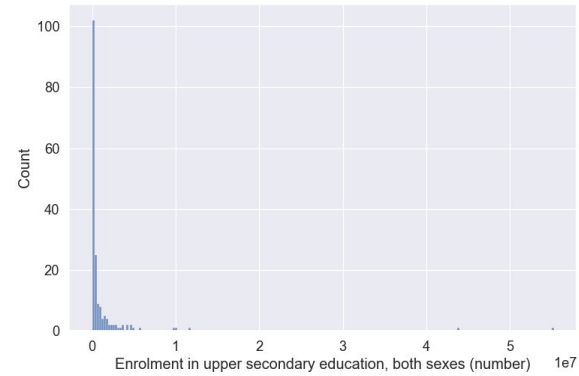


Par pays et pour chaque indicateur, je conserve la dernière valeur connue.

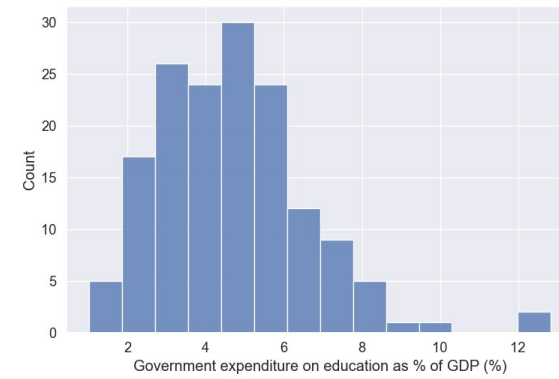
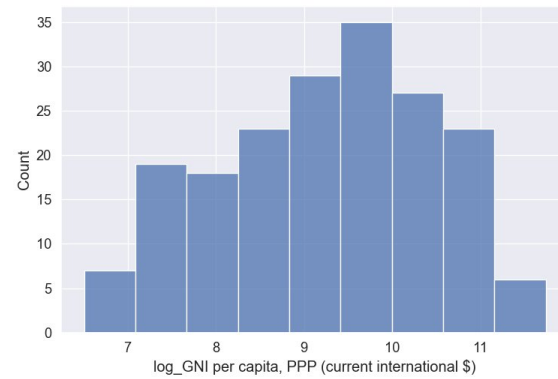
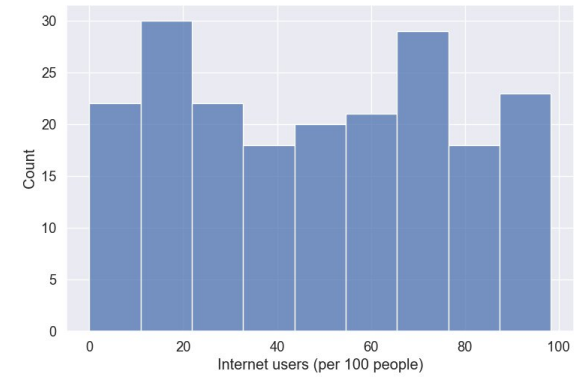
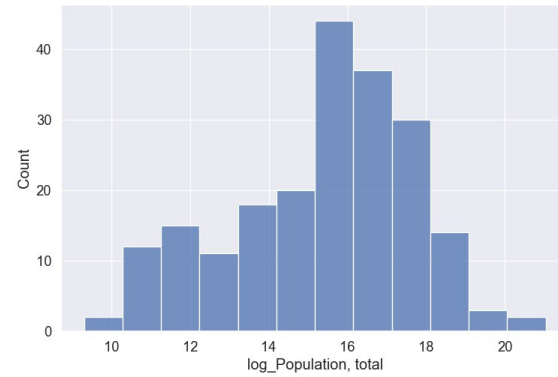
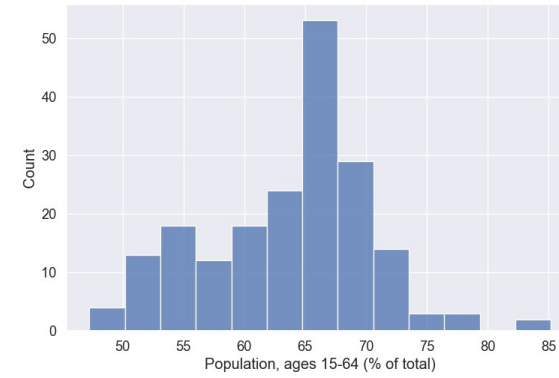
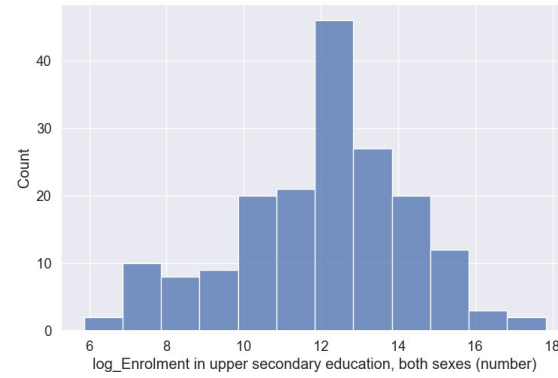


# 4.1 Analyses bivariées

## Répartition empirique des indicateurs



## Répartition empirique des indicateurs



# Matrice de corrélation entre les indicateurs



log\_Enrolment in upper secondary education, both sexes (number)

Population, ages 15-64 (% of total)

log\_Population, total

Internet users (per 100 people)

log\_GNI per capita, PPP (current international \$)

Government expenditure on education as % of GDP (%)



log\_Enrolment in upper secondary education, both sexes (number)  
Population, ages 15-64 (% of total)  
log\_Population, total  
Internet users (per 100 people)  
log\_GNI per capita, PPP (current international \$)  
Government expenditure on education as % of GDP (%)

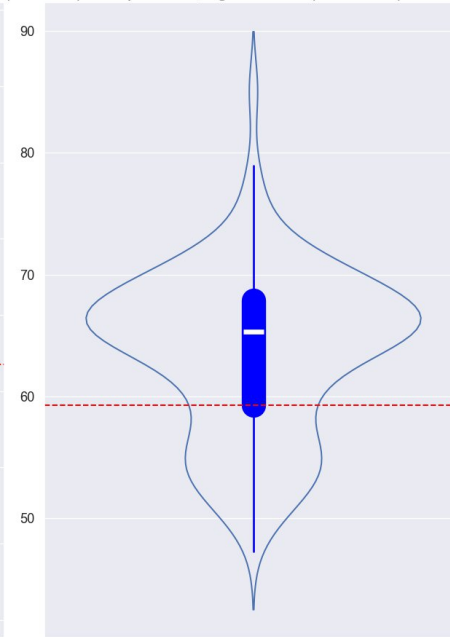
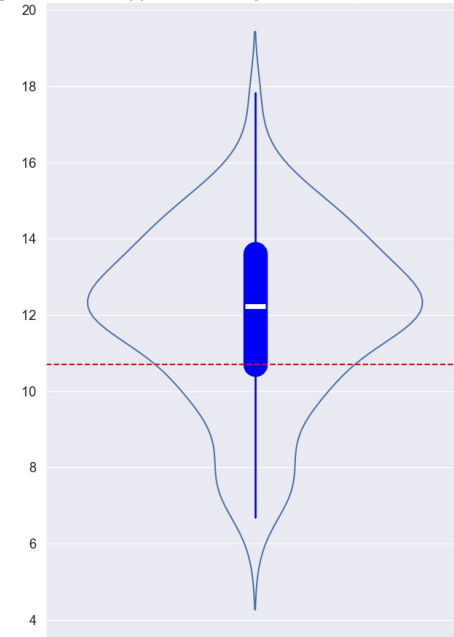




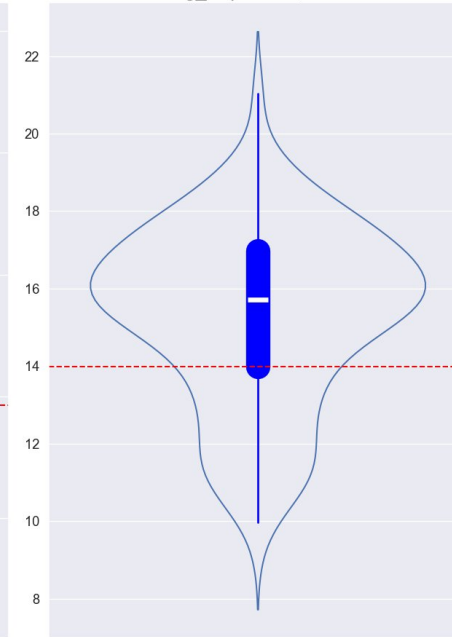
## 4.2 Analyses multivariées



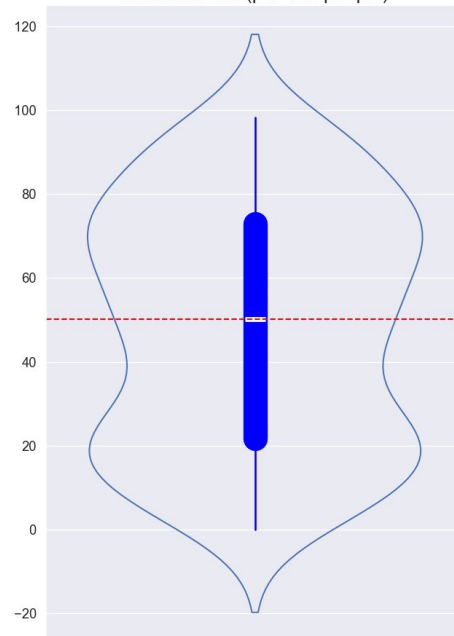
\_Enrolment in upper secondary education, both sexes (number)    Population, ages 15-64 (% of total)



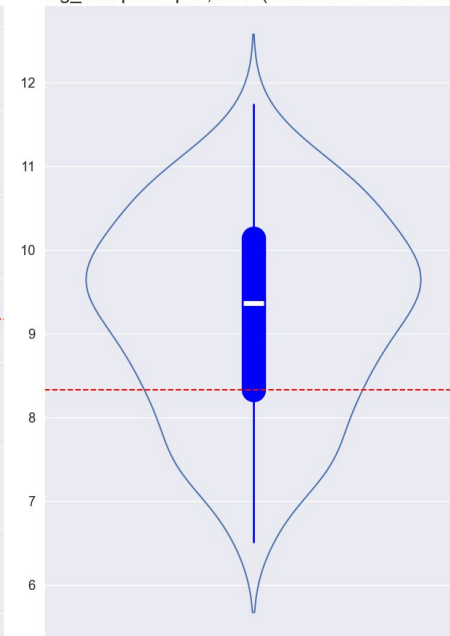
log\_Population, total



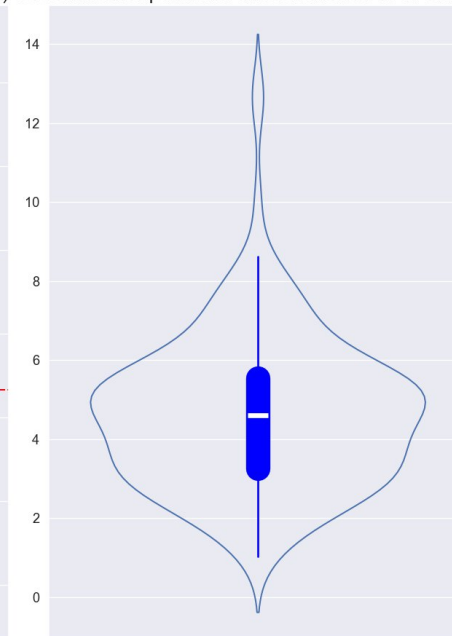
Internet users (per 100 people)



log\_GNI per capita, PPP (current international \$)



Government expenditure on education as % of GDP (%)





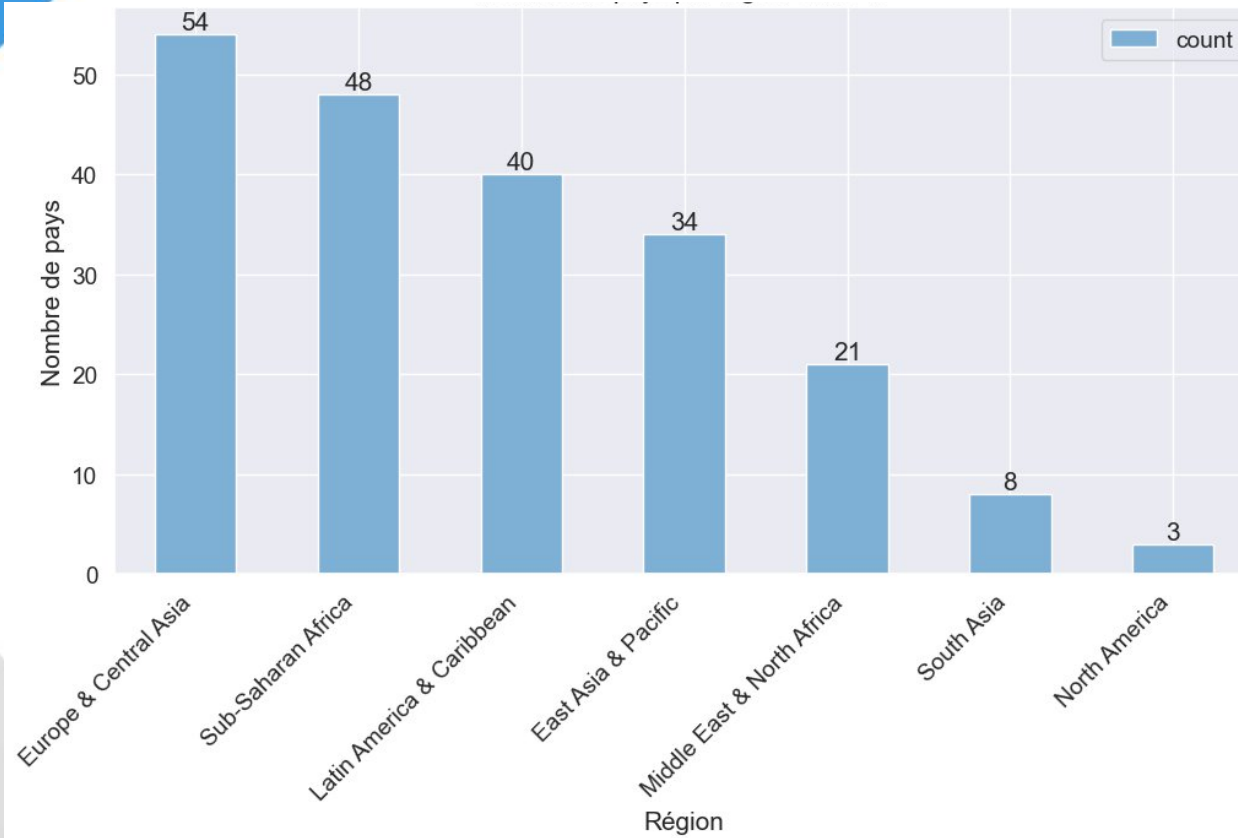
# 5 Scoring (Min-Max scaling)

Le min-max scaling permet de normaliser les données entre 0 et 1 en soustrayant le minimum et en divisant par l'étendue des valeurs.

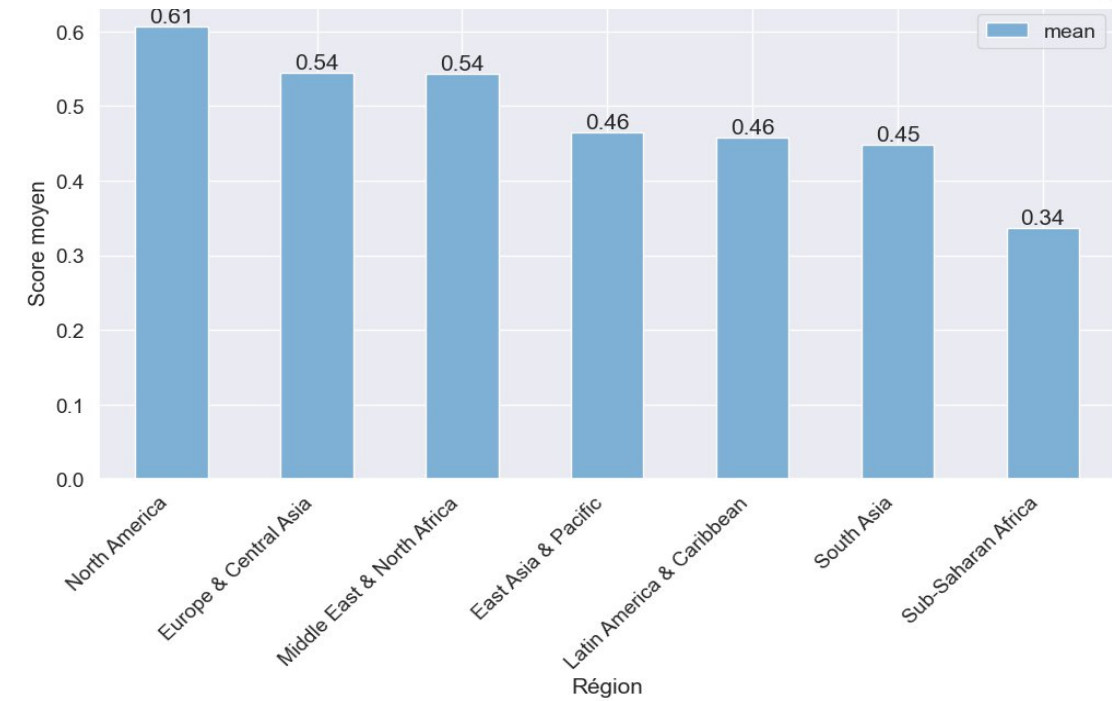
$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

# 6 Présentation des résultats

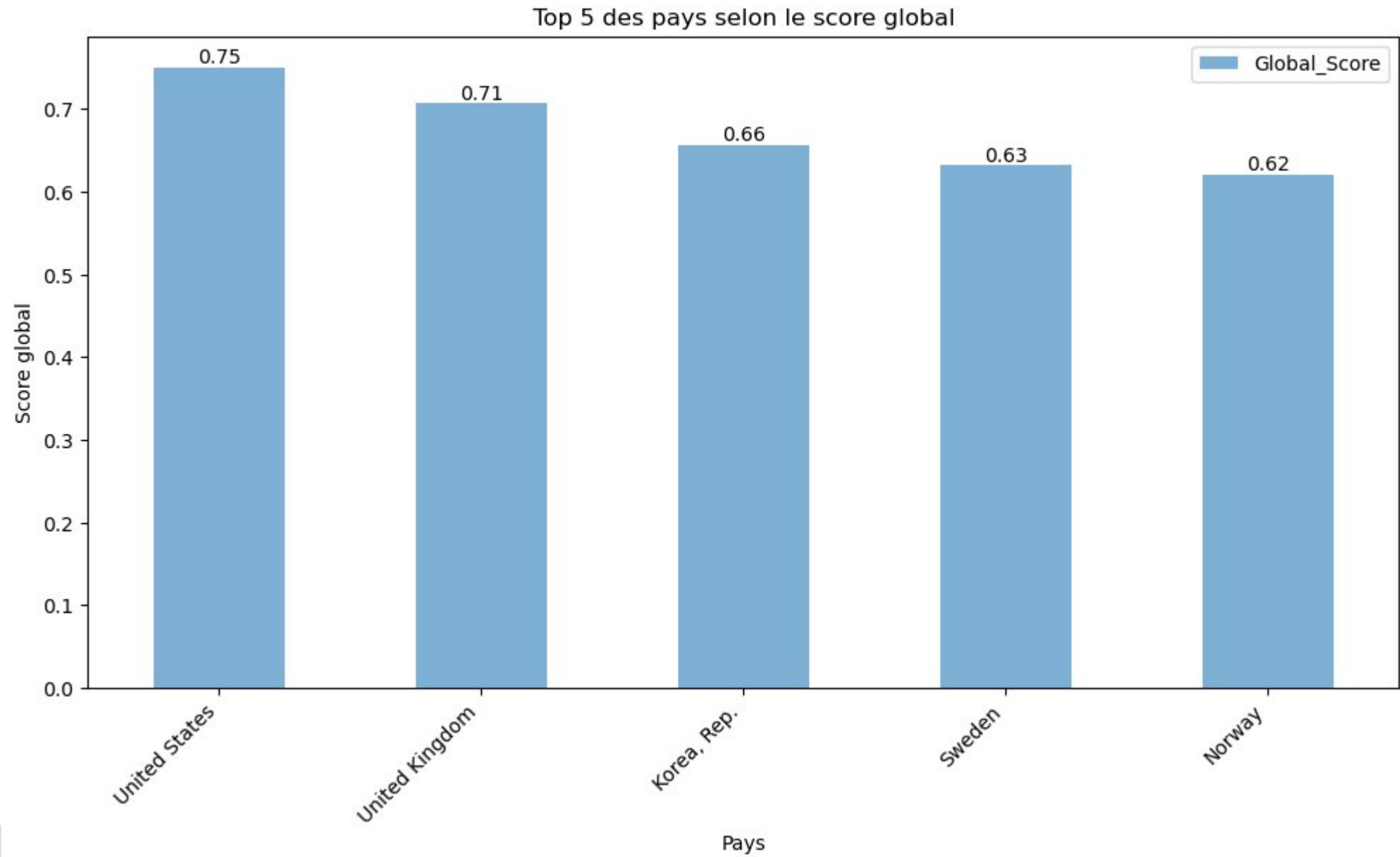




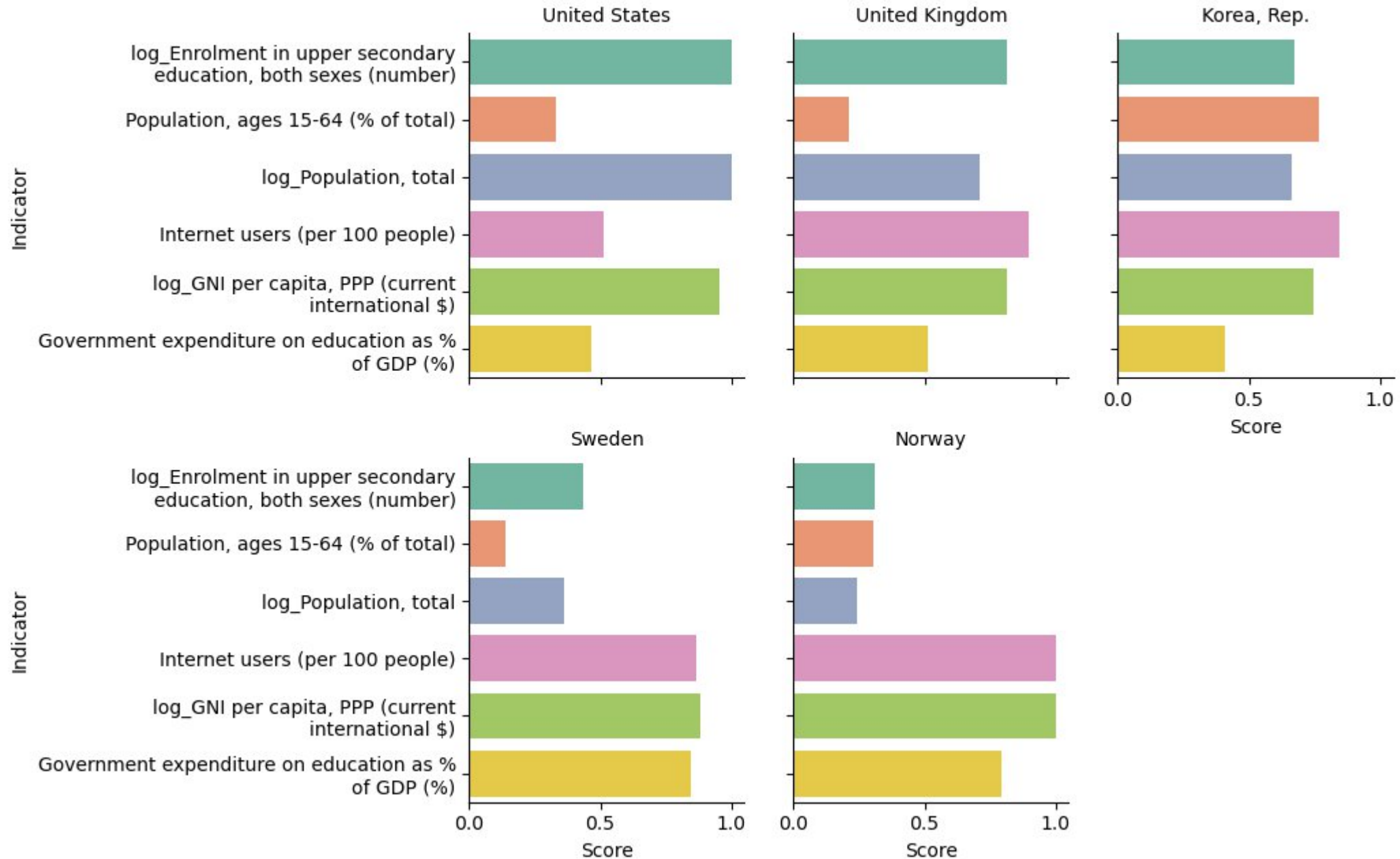
Nombre de pays par région



Score moyen par région



# Score par indicateur du top 5 des pays





## Top 5 des pays correspondants au mieux à nos critères







# 7 Conclusion

Le jeu de données m'a permis d'identifier 5 pays qui correspondent au mieux à nos critères.

- United States
- United Kingdom
- Korea Rep.
- Sweden
- Norway

En revanche, je n'ai pas de données suffisantes pour établir une liste potentielle pour 2025. Un autre jeu de données est nécessaire pour cela.