

**Projet réalisé par  
Alain KENFACK**

Samedi, 11 mars 2023

# Rappel de la problématique

## A propos de Academy

- Entreprise de la Edtech
- Publique cible : lycéens et étudiants
- Objectif : s'étendre à l'internationale



BANQUE MONDIALE

## Mon rôle

Analyser les données de la banque mondiale pour informer le projet d'expansion.



**Projet réalisé par  
Alain KENFACK**

Samedi, 11 mars 2023

# Jeu de données

## EdStatsCountry.csv (Country)

- Economie des pays du monde
- Taille : 241 lignes (correspondant aux pays et zone) , 32 colonnes
- Quelques valeurs manquantes
- Aucun doublon

## EdStatsSeries.csv (Series)

- Informations sur les indicateurs disponibles dans EdStatsData.
- 6 colonnes vides.
- 10 colonnes avec plus de 80% de données manquantes

## EdStatsCountry-Series.csv (Country-series)

- Informations sur la source des données contenues dans EdStatsCountry
- La colonne "Unnamed : 3" est composée uniquement de valeur manquante)
- Aucun doublon

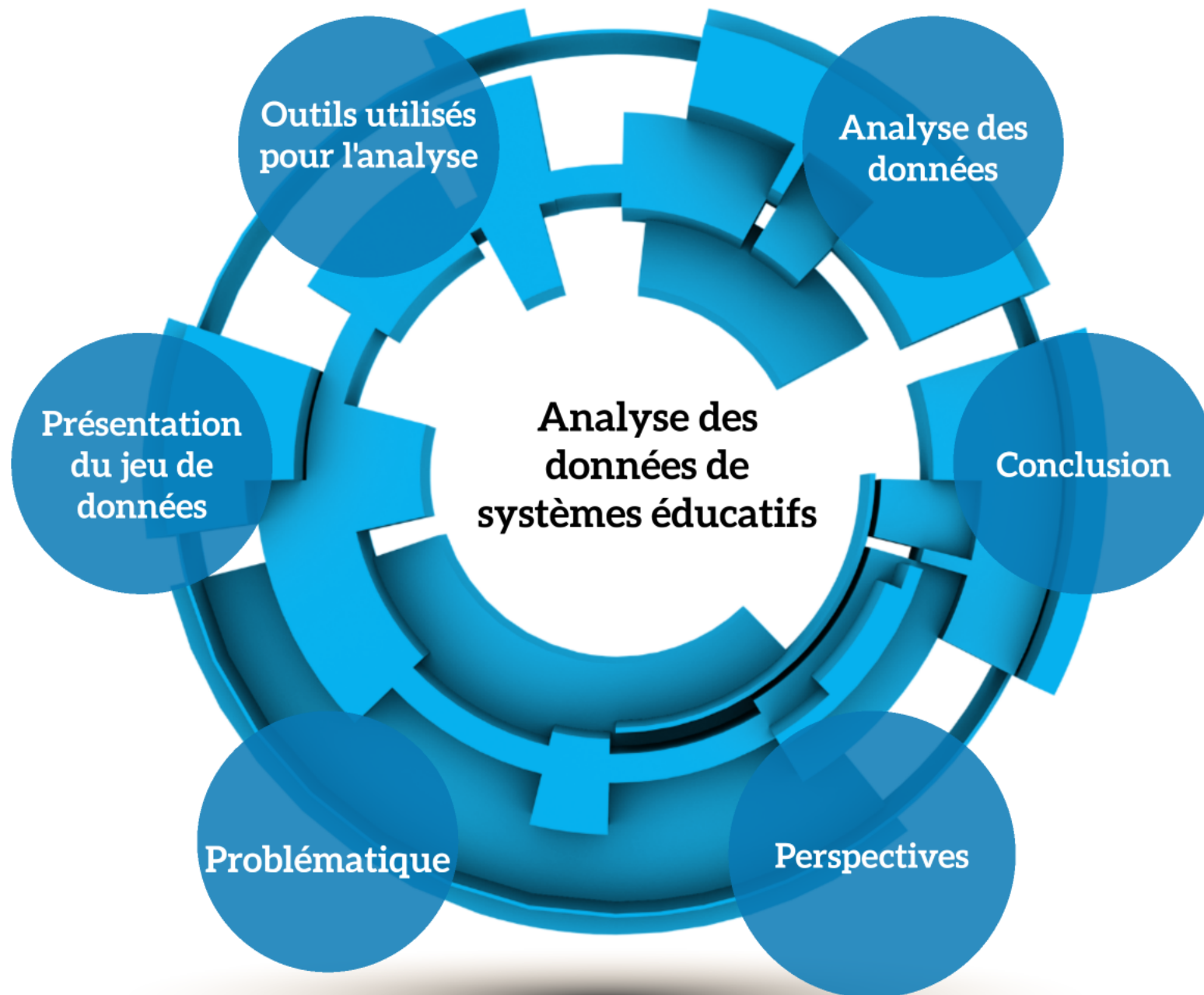
## EdStatsFootNote.csv (FootNote)

- Renseigne sur l'année d'origine des données et les incertitudes sur les données)
- "Unnamed : 4" : colonne renfermant uniquement les valeurs manquantes)
- Aucun doublon

# Jeu de données

EdStatsData.csv (Data)

- Evolution des indicateurs pour tous les pays de 1970 à 2100
- 
- Indicateurs : démographiques, éducatifs, pénétration d'internet etc.
- Taille : 886 930 lignes, 70 colonnes
- Nombreuses valeurs manquantes
- Aucun doublon



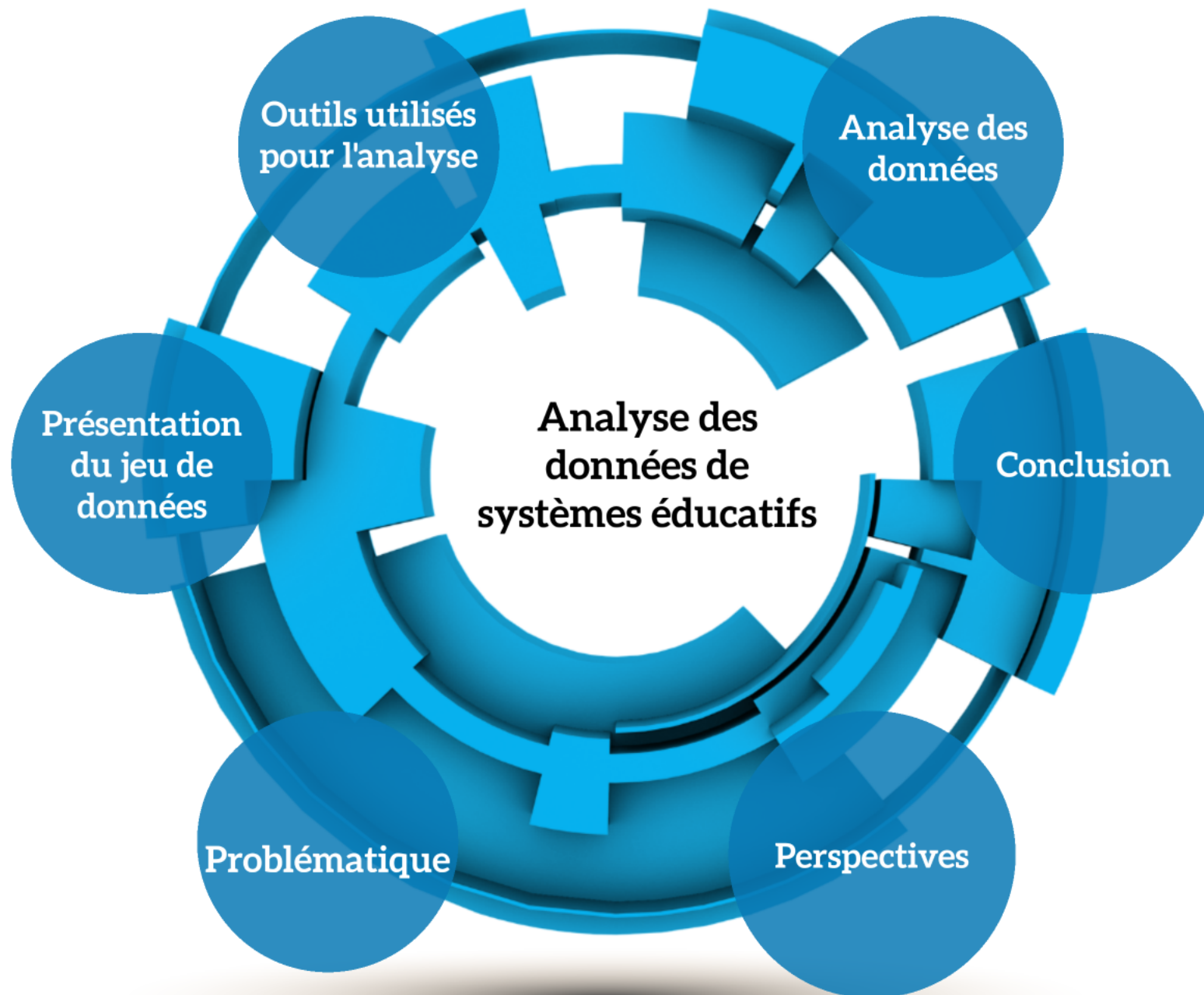
**Projet réalisé par  
Alain KENFACK**

Samedi, 11 mars 2023

# Outils utilisés

Anaconda	Accès à JupyterNotebook
Jupyter Notebook 6.4.12	Exécution du code via Python 3.9.13 et la documentation de mon travail
<ul style="list-style-type: none"><li>• Pandas 1.4.4</li><li>• Numpy 1.21.5</li></ul>	Manipulation et la Représentation des données
<ul style="list-style-type: none"><li>• Matplotlib 3.5.2</li><li>• Seaborn 0.11.2</li></ul>	Génération de graphes





**Projet réalisé par  
Alain KENFACK**

Samedi, 11 mars 2023

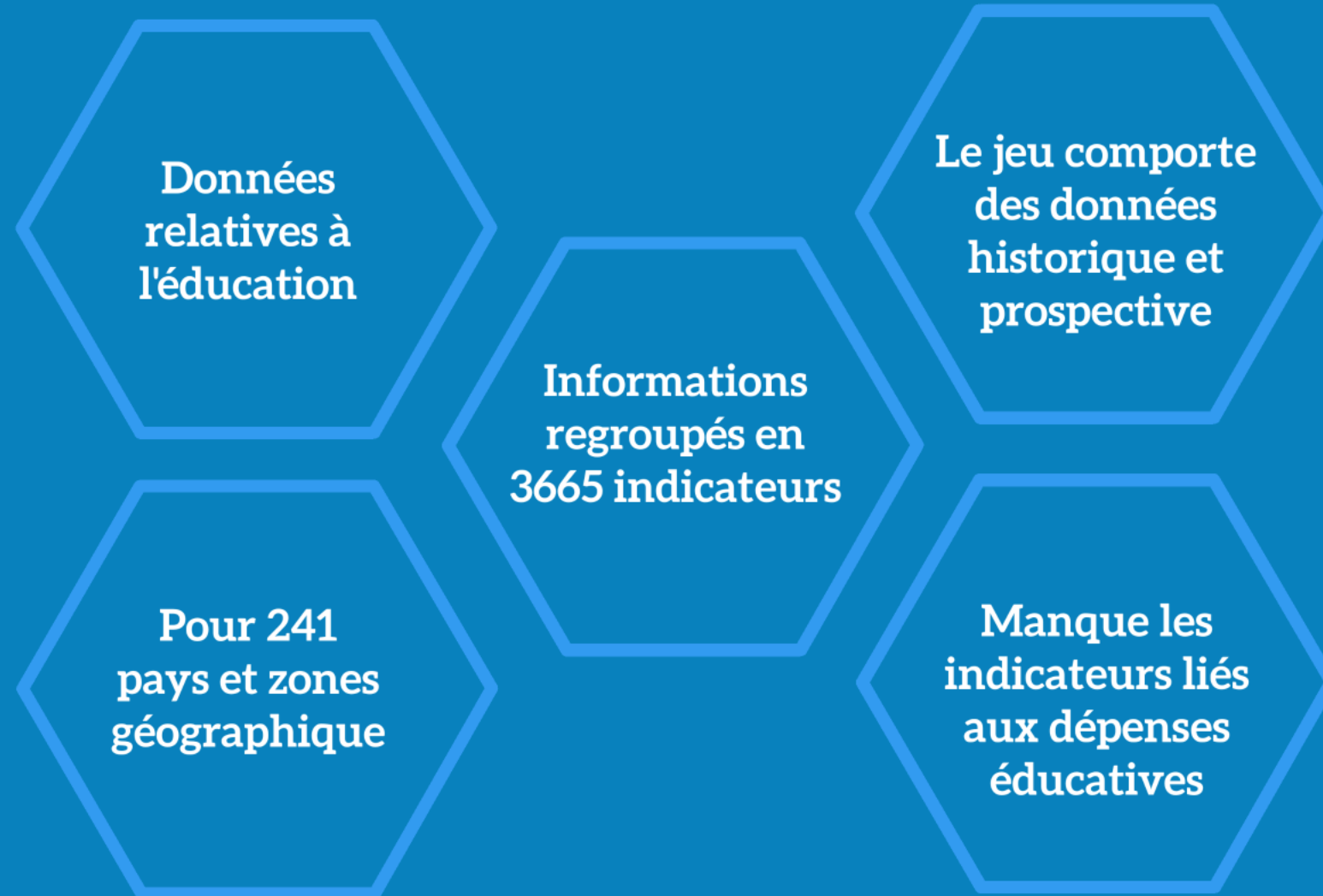


# Processus d'analyse

1. Exploration des données
2. Identification des indicateurs clés
3. Comparaison des pays
4. Identification des pays à fort potentiel

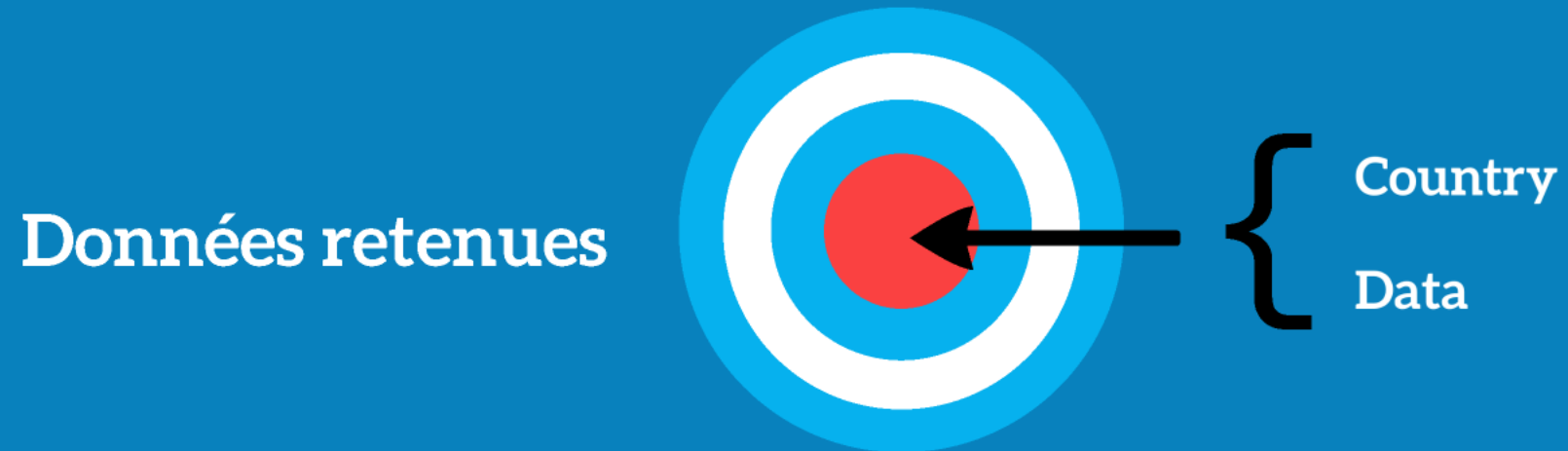
# 1. Exploration des données

## Informations contenues dans le jeu de données

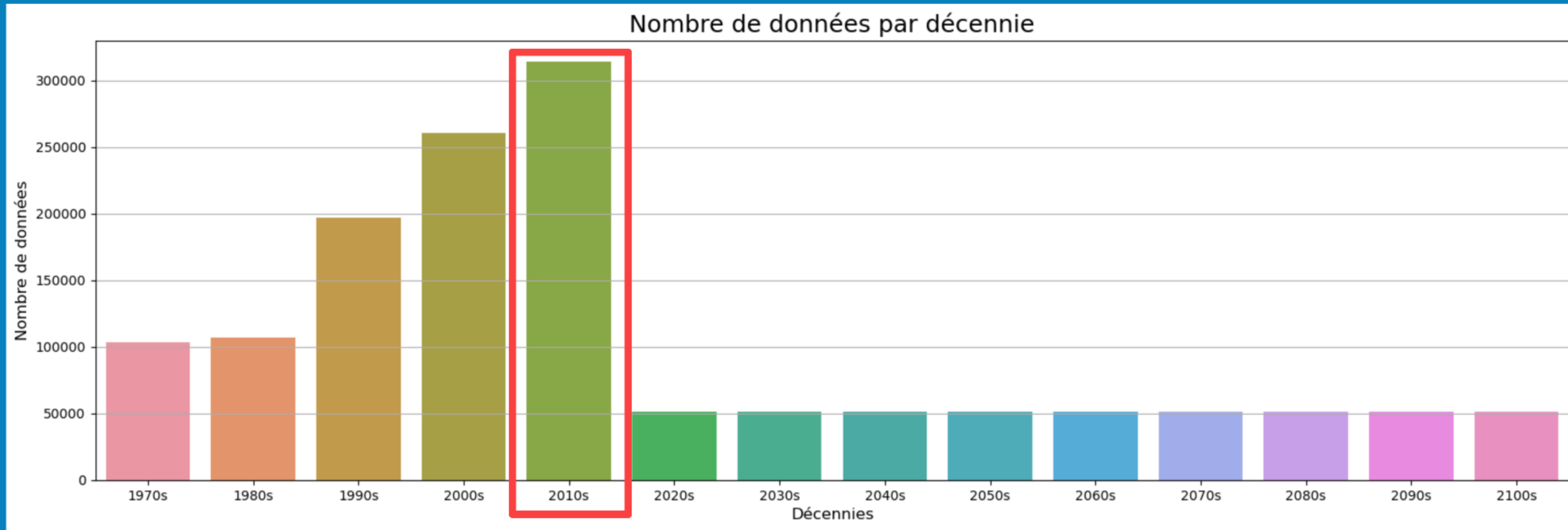


# 1. Exploration des données

Choix des fichiers à conserver



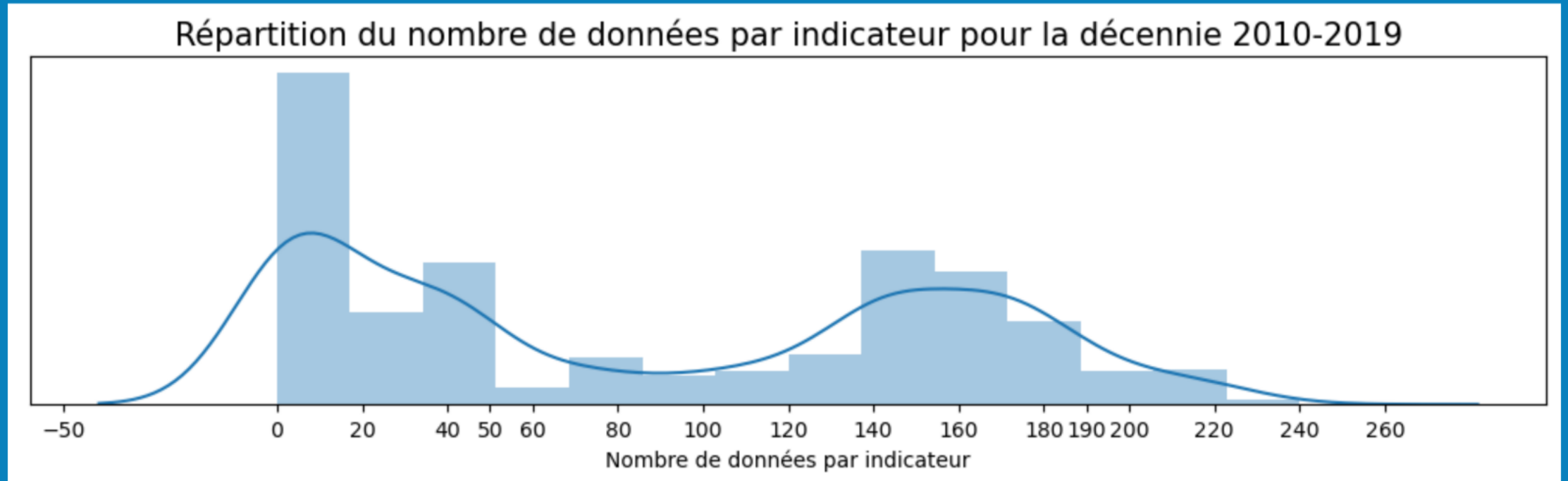
# 1. Exploration des données



## 2. Identification des indicateurs clés

### Identification des indicateurs exploitables

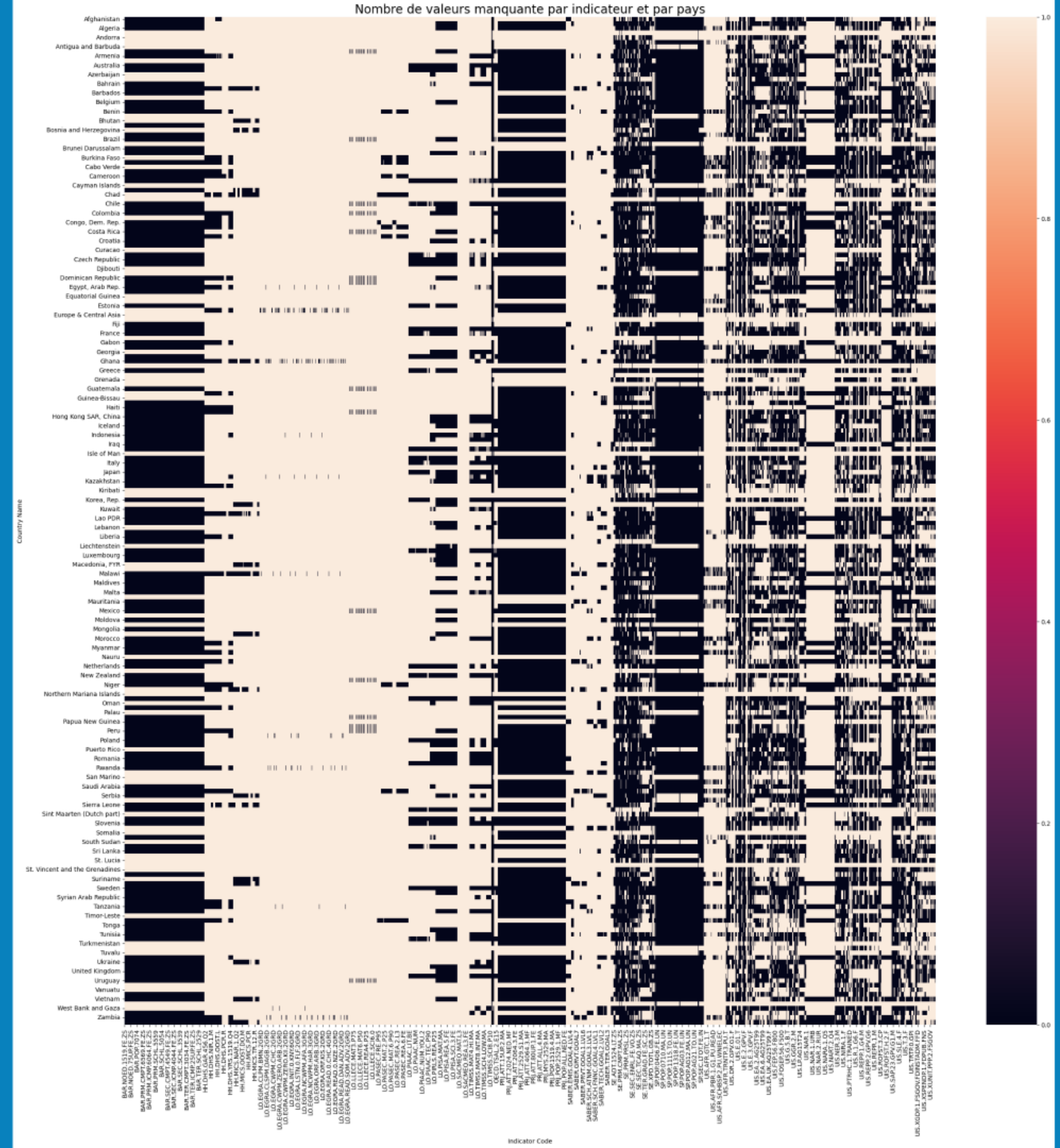
- Aggrégation des deux données retenues (Data et country)
- Recherche des indicateurs pertinents en identifiant ceux avec le plus de données



## 2. Identification des indicateurs clés

### Indicateurs avec le moins de NaN par pays :

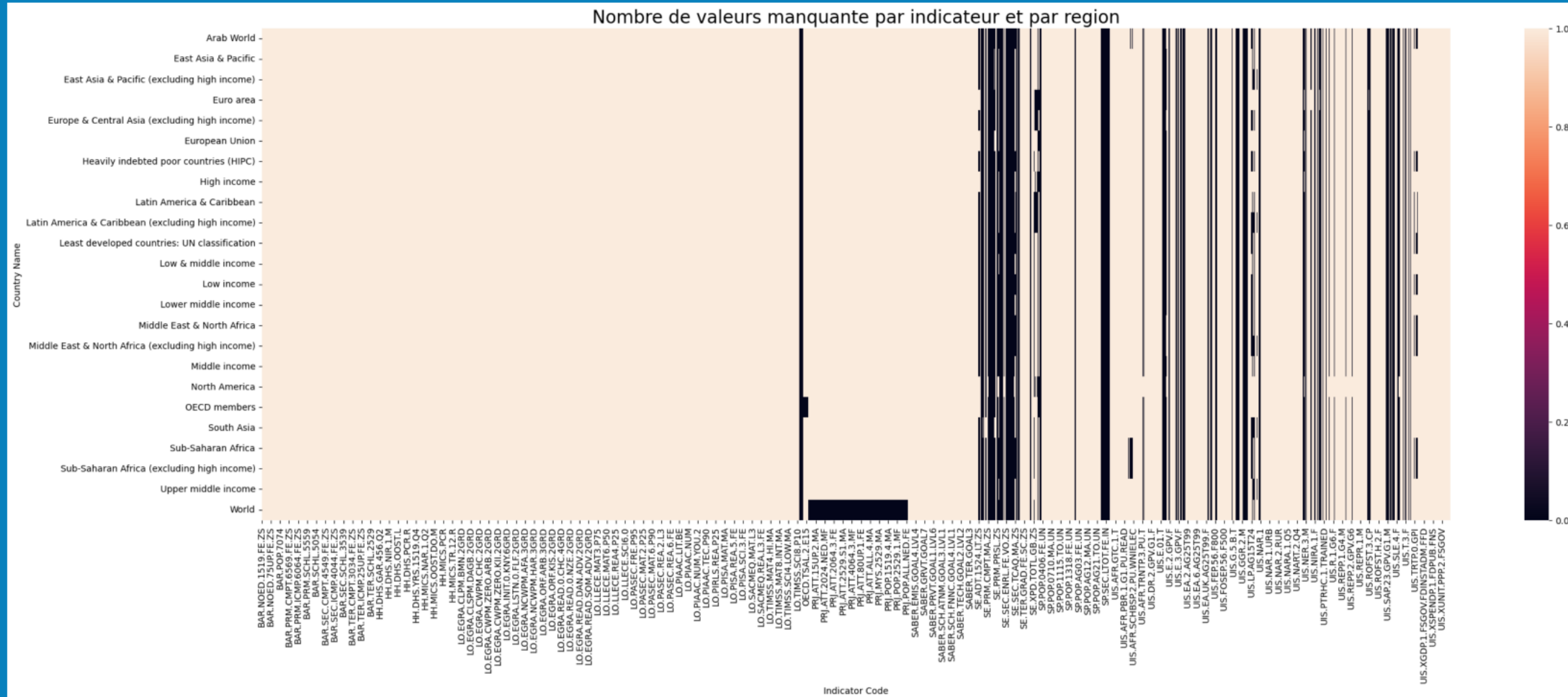
- HH,
- LO,
- SABER,
- UIS.AFR,
- UIS.NAR



# 2. Identification des indicateurs clés

Indicateurs avec le moins de NaN par région :

- BAR
- HH
- LO
- PRJ
- SABER
- SE.TER
- SP.
- UIS





## 2. Identification des indicateurs clés

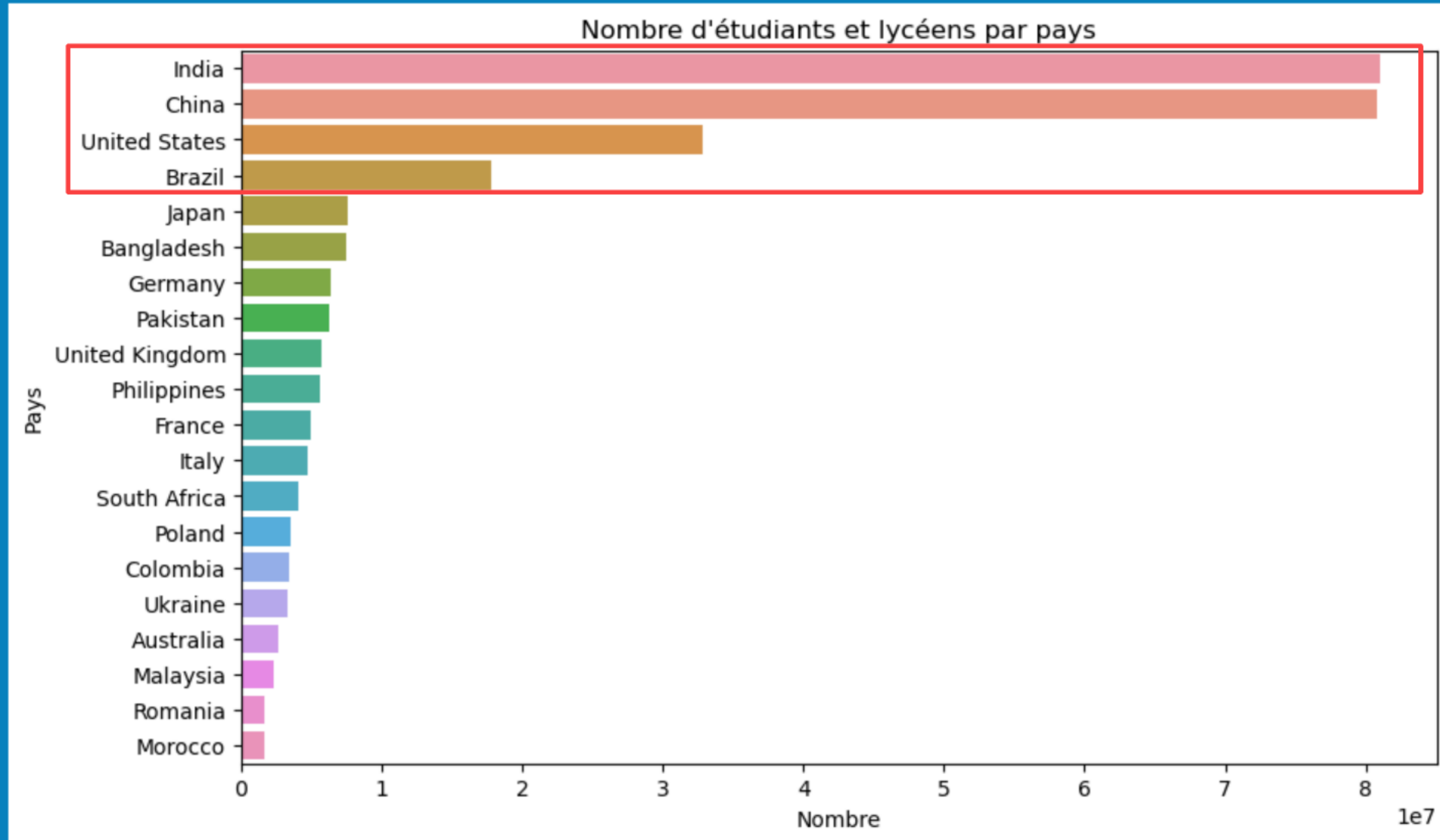
Critères de sélection des indicateurs :

- Avec plus de 90 % de données significantes;
- Liés aux nombres d'élèves et étudiants;
- Liés aux données démographiques;
- Liés à l'accès à internet;

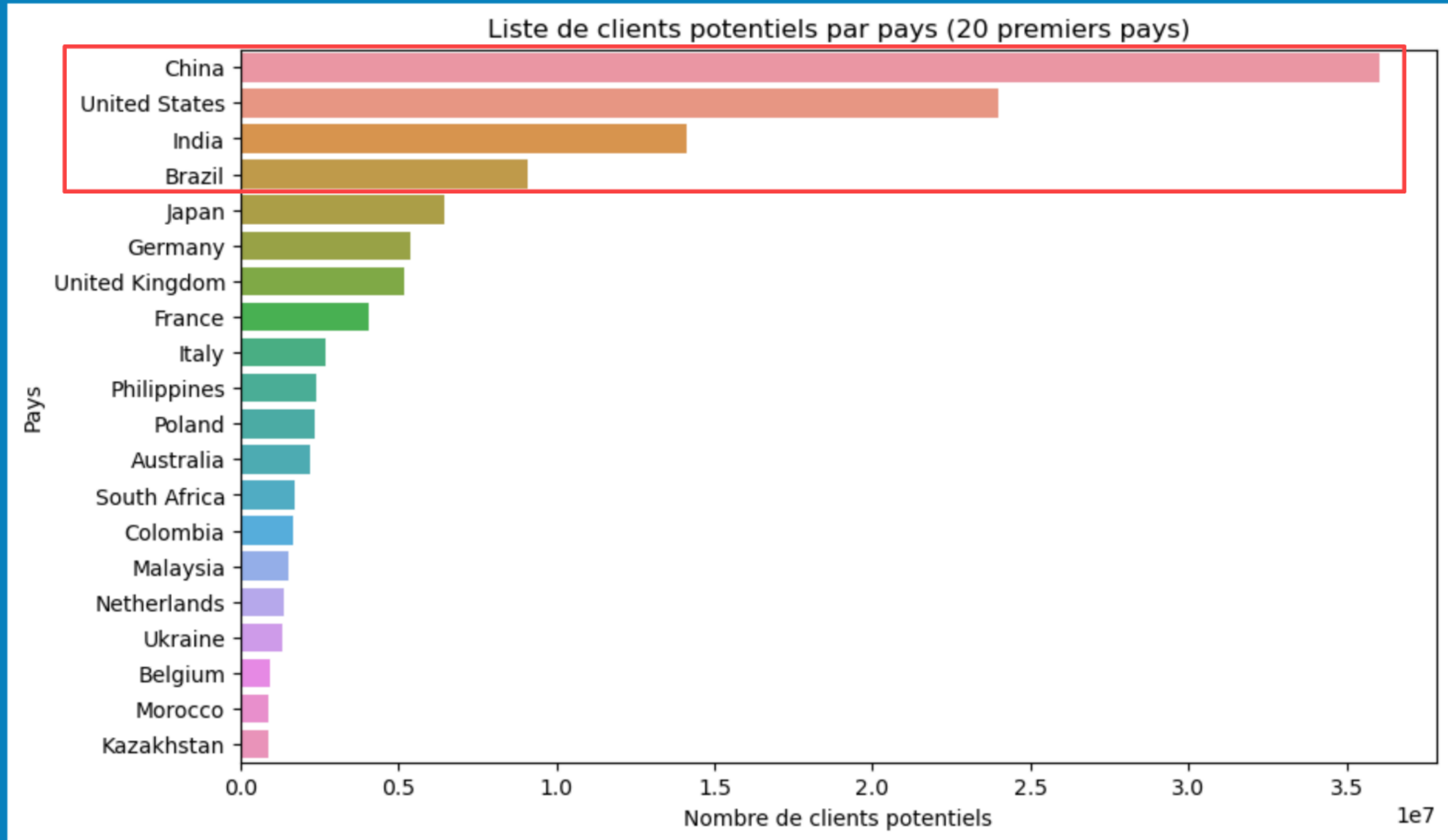
### *Indicateurs sélectionnés*

Indicator Name		Indicator Code	2010s
5	Population, total	SP.POP.TOTL	240
3	Internet users (per 100 people)	IT.NET.USER.P2	229
2	Enrolment in upper secondary education, both s...	UIS.E.3	206
1	Enrolment in tertiary education, all programme...	SE.TER.ENRL	197
4	Population, ages 15-24, total	SP.POP.1524.TO.UN	181
0	Enrolment in post-secondary non-tertiary educa...	UIS.E.4	137

### 3. Comparaison des pays



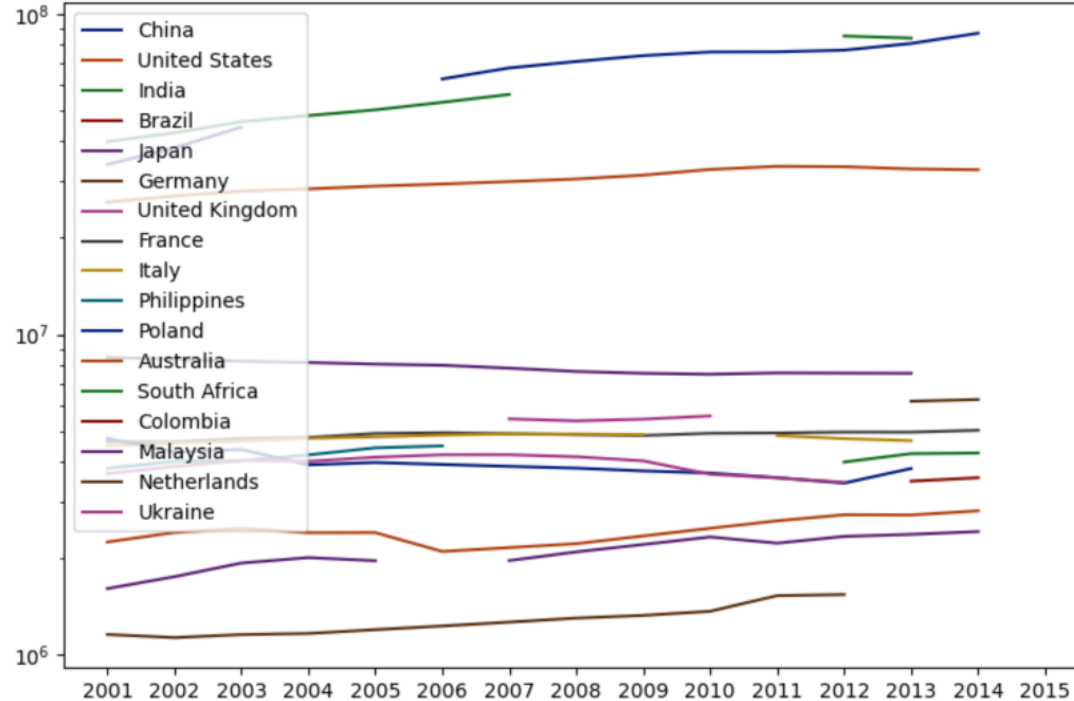
### 3. Comparaison des pays



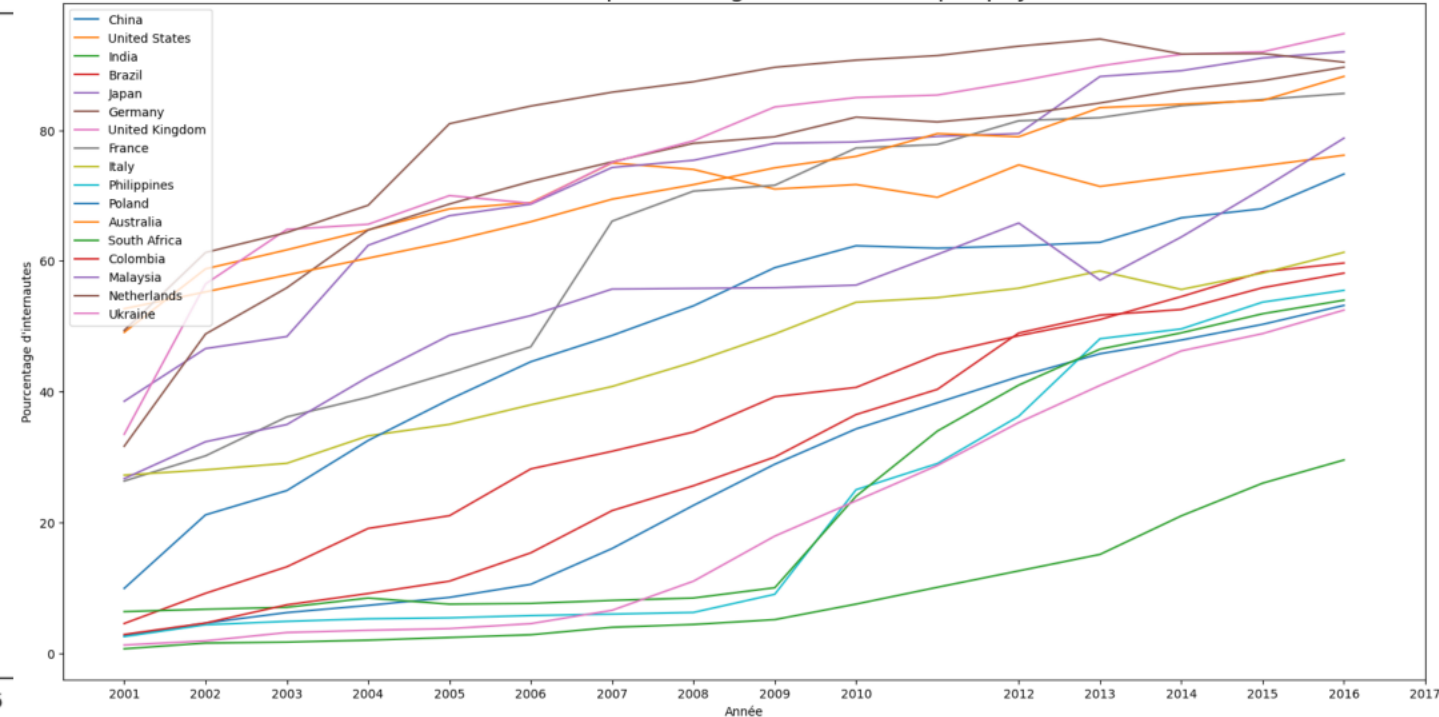
En se basant sur le taux de pénétration d'internet

# 4. Identification des pays à fort potentiel

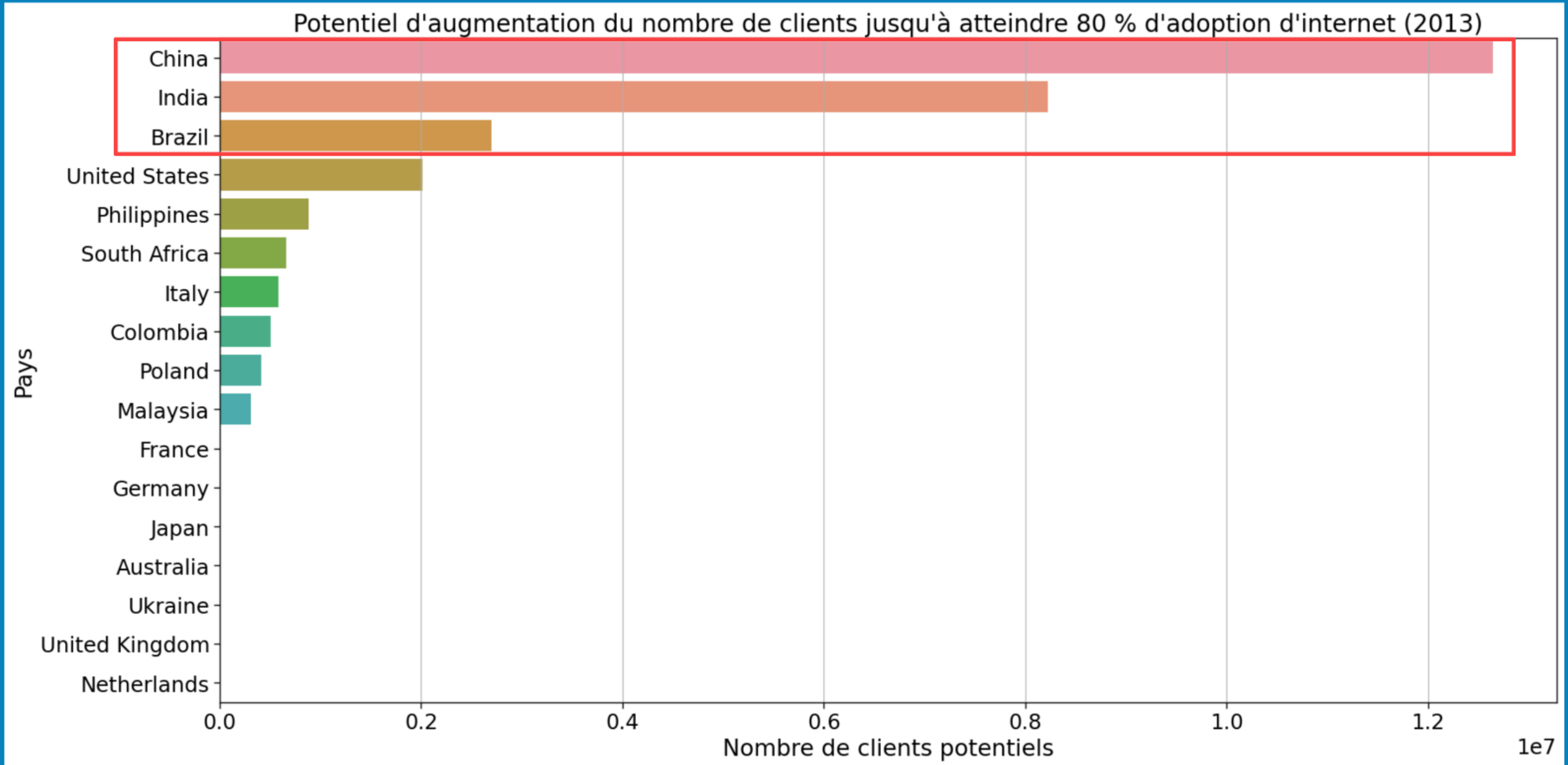
Evolution du nombre d'étudiants par pays



Evolution du pourcentage d'internautes par pays



# 4. Identification des pays à fort potentiel



L'entreprise doit opérer en priorité en :

- Chine
- Inde
- Et au Brésil





**Projet réalisé par  
Alain KENFACK**

Samedi, 11 mars 2023



# Conclusion

## Rappel de la mission

- Identifier les pays à fort potentiel de clients ;
- Déterminer l'évolution de la clientèle pour ces pays ;
- Identifier les pays cible prioritaire

# Conclusion

## Validité du jeu de données

- Des informations pour tous les pays du monde ;
- Les indicateurs sont référencés
- Le jeu de données a permis d'informer le projet de l'entreprise;

## Les limites

- Pas de données liées aux dépenses internet, au taux de décrochage scolaire/académique
- Manque d'informations sur l'entreprise Academy



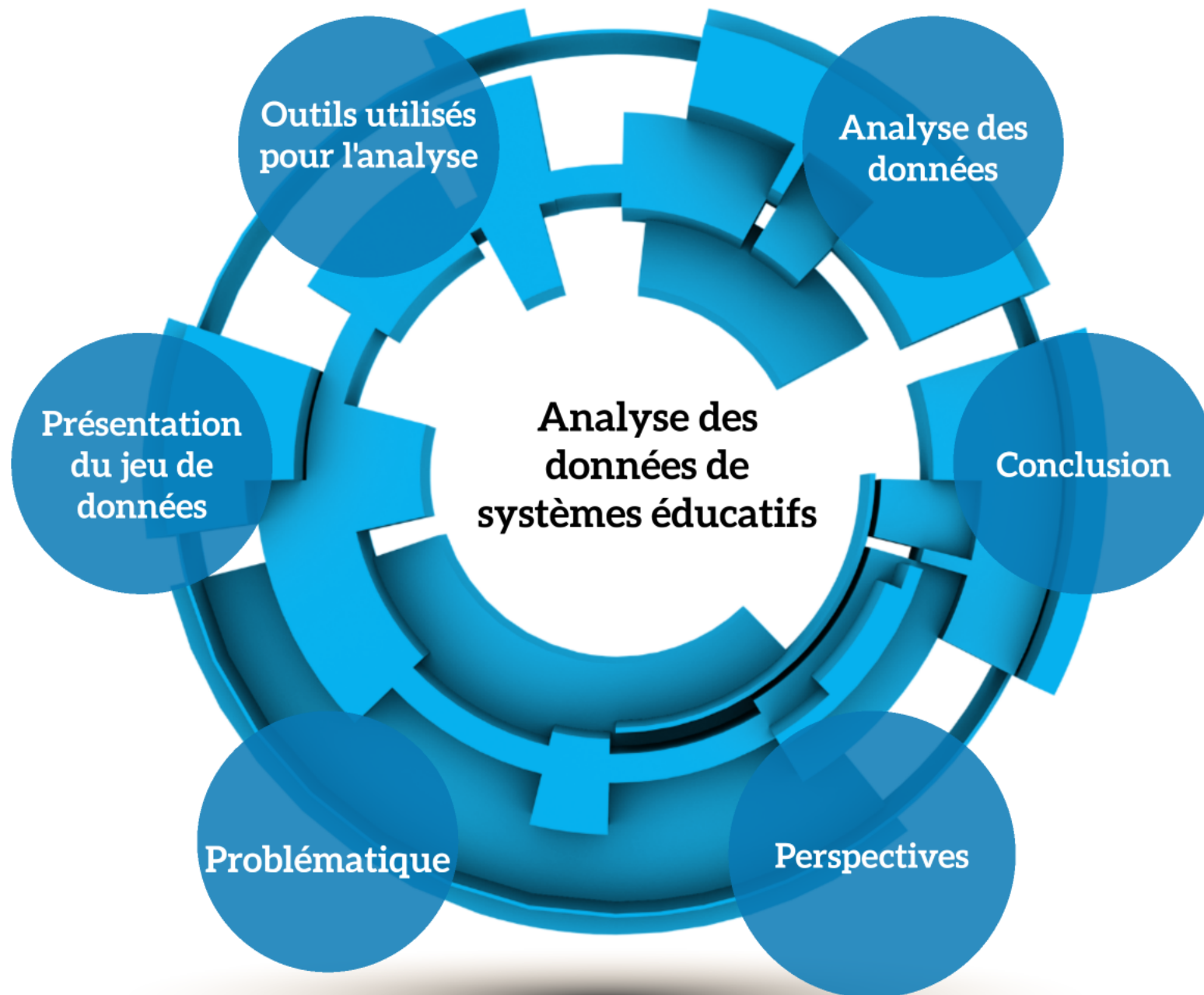
**Projet réalisé par  
Alain KENFACK**

Samedi, 11 mars 2023

# Perspective

Réaliser des études sur la concurrence déjà installée dans ces pays.

Merci pour  
votre attention



**Projet réalisé par  
Alain KENFACK**

Samedi, 11 mars 2023