

Projet n°2 :

**Analysez des données de systèmes éducatifs**

---

Soutenance de Projet  
26 mai 2019

# Programme

I - Rappel de la problématique

II- Présentation du jeu de données

II - Analyse pré-exploratoire

III - Conclusions sur la pertinence du jeu de données

# Rappel de la problématique



---

**Academy:** une start-up de la EdTech, propose des contenus de formation en ligne pour un public de niveau lycée et université.

**Objectif :** Expansion à l'international de l'entreprise

## Mission

1- Faire une analyse pré exploratoire pour déterminer si les données sur l'éducation de la Banque Mondiale permettent d'informer le projet d'expansion.

2- Répondre aux questions suivantes:

- ❖ Quels sont les pays avec un fort potentiel de clients pour nos services ?
- ❖ Pour chacun de ces pays, quelle sera l'évolution de ce potentiel de clients ?
- ❖ Dans quels pays l'entreprise doit-elle opérer en priorité ?

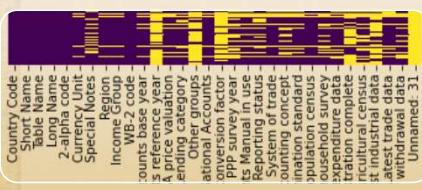
# Présentation du jeu de données

## « Country-Series »



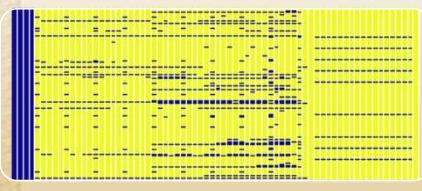
## Informations sur les sources des données contenues dans « EdStatsCountry »,

Taille : 613 lignes, 4 colonnes. Pas de valeur manquante (sauf "Unnamed : 3" qui est une colonne uniquement composée de NaN), Aucun doublon



## « Country »,

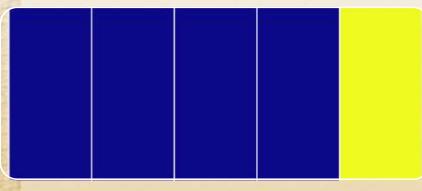
Informations globales sur l'économie de chaque pays, et leur niveau de revenu, ainsi qu'un classement par région. Taille : 241 lignes (1 par pays / zone) , 32 colonnes. Quelques valeurs manquantes, Aucun doublon



« Data »

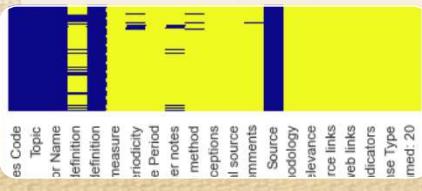
Evolution de nombreux indicateurs pour tous les pays et certains groupes de pays.

Taille : 886930 lignes, 70 colonnes, données depuis 1970, Nombreuses valeurs manquantes,  
Aucun doublon



« FootNote »

Informations sur l'année d'origine des données et des précisions sur le mode de calcul relatif à divers couples Indicateur/Pays.Taille : 643 638 lignes, 4 colonnes. Pas de valeur manquante (sauf "Unnamed : 4" qui est une colonne uniquement composée de NaN), Aucun doublon



## « Series »

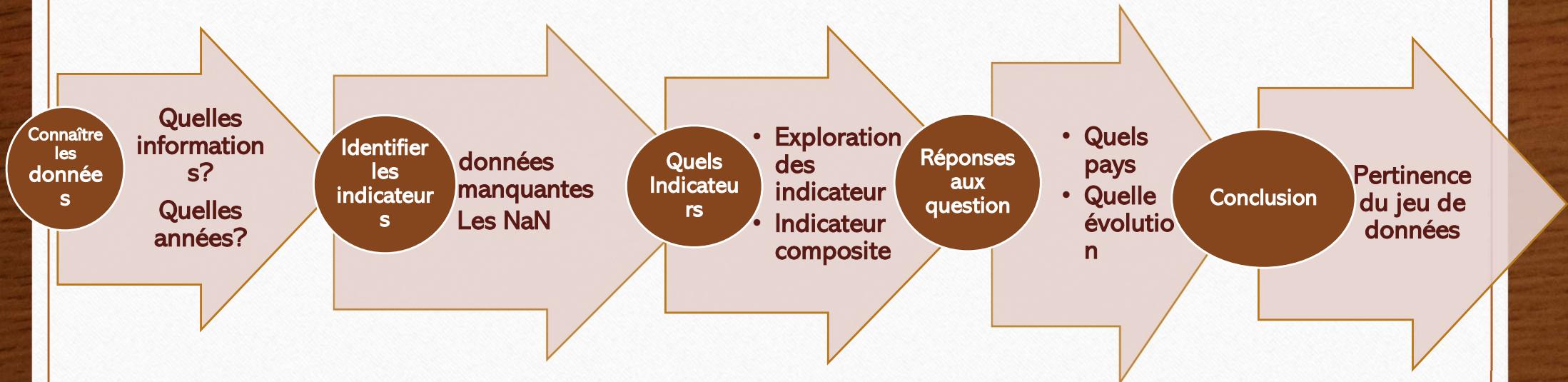
## Données descriptives sur les indicateurs socio économiques disponibles dans "data".

Taille : 3665 lignes, 21 colonnes, 6 colonnes vides pour lesquelles il manque toutes les valeurs. Il manque plus de 80 % des données dans 10 autres colonnes de la table, Aucun doublon

# Analyse Pré-Exploratoire

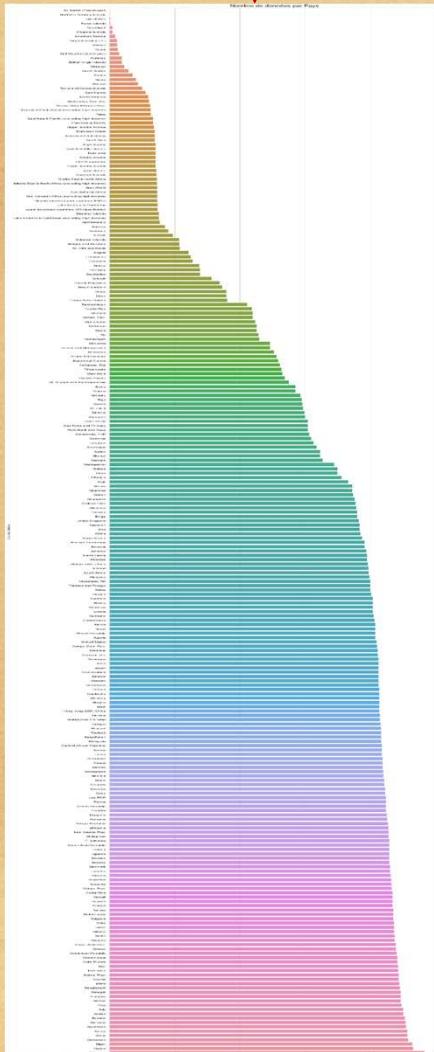
---

## Processus d'analyse pré-exploratoire

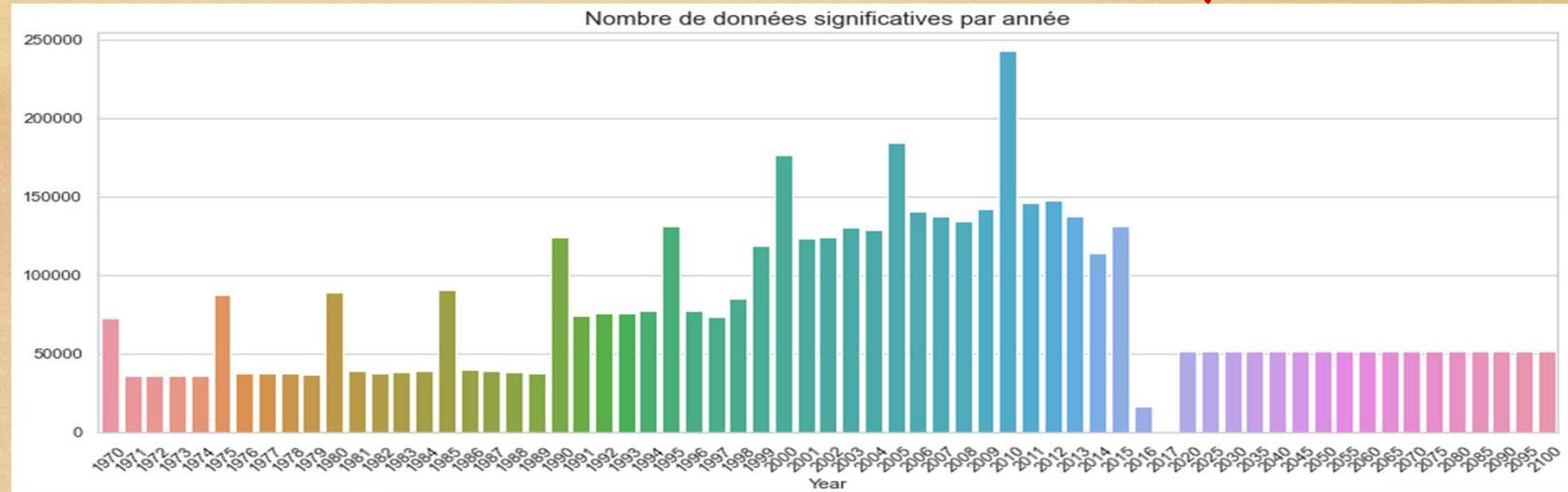


# Connaître les données

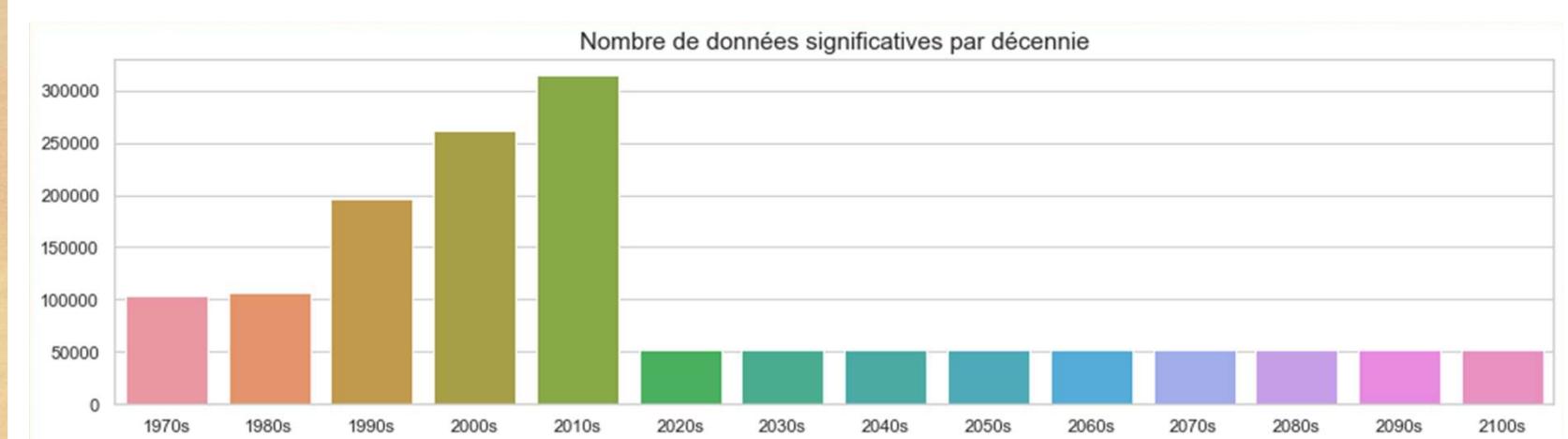
Données par pays



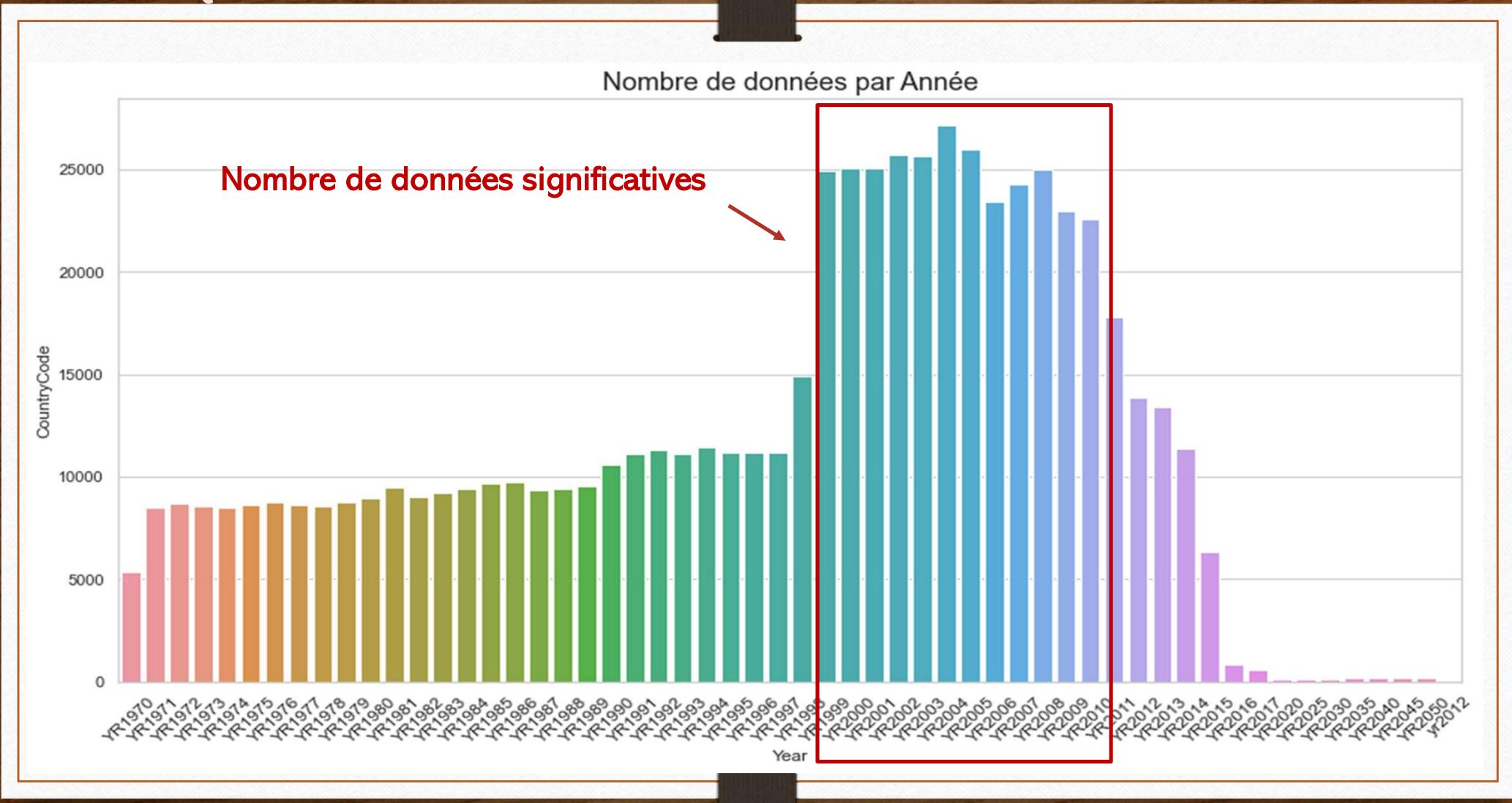
Données par année



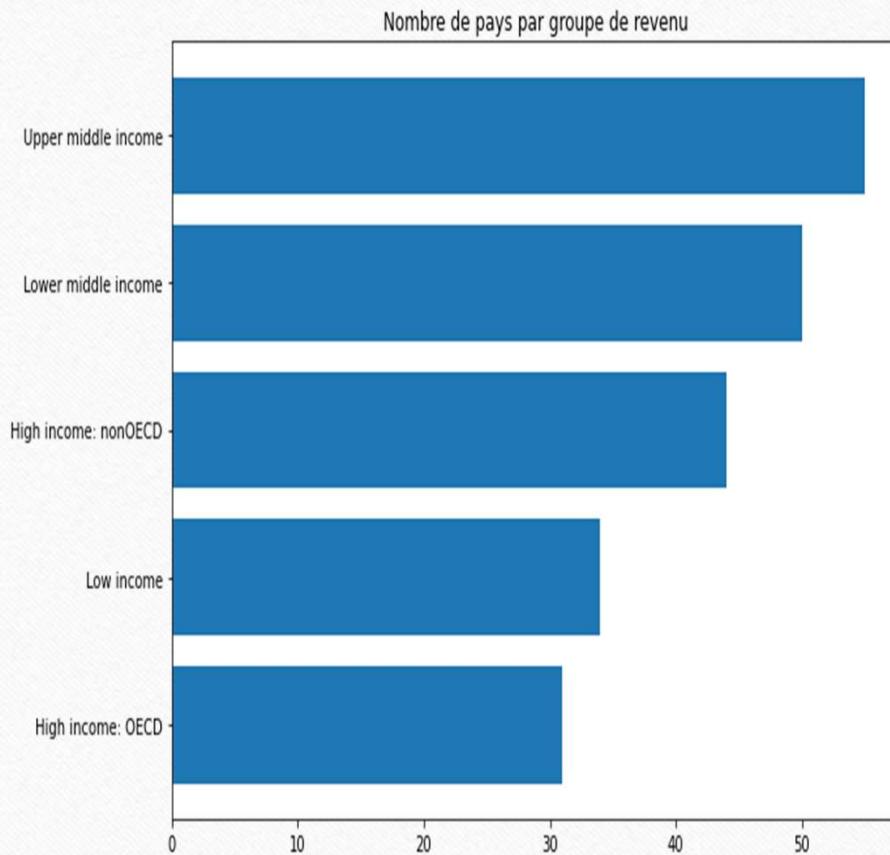
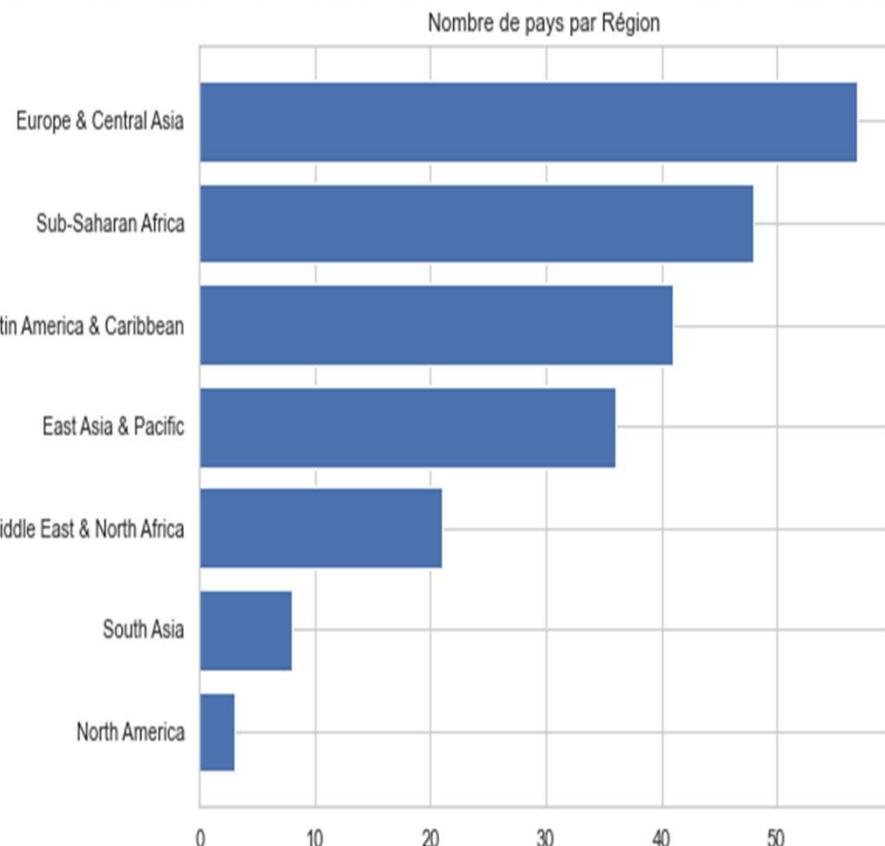
Nombre de données significatives par décennie



## Quelle année choisir ?



# Connaitre les données



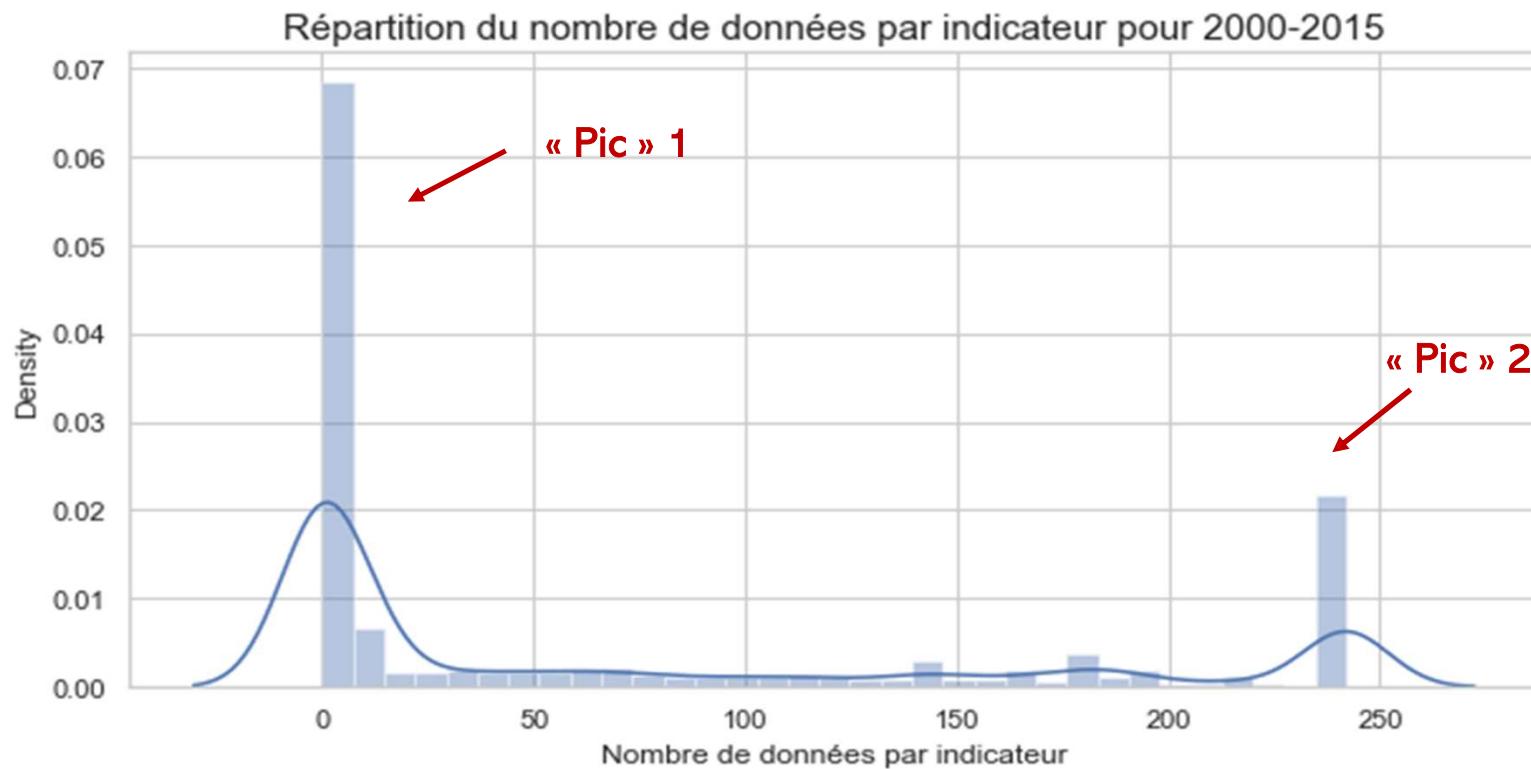
98] :

	Country Name	Country Code	Indicator Name	Indicator Code	2000	2001	2002	2003	2004
0	Arab World	ARB	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2	NaN	NaN	NaN	NaN	NaN
1	Arab World	ARB	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2.F	NaN	NaN	NaN	NaN	NaN
2	Arab World	ARB	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2.GPI	NaN	NaN	NaN	NaN	NaN

# Quelles informations conserver?

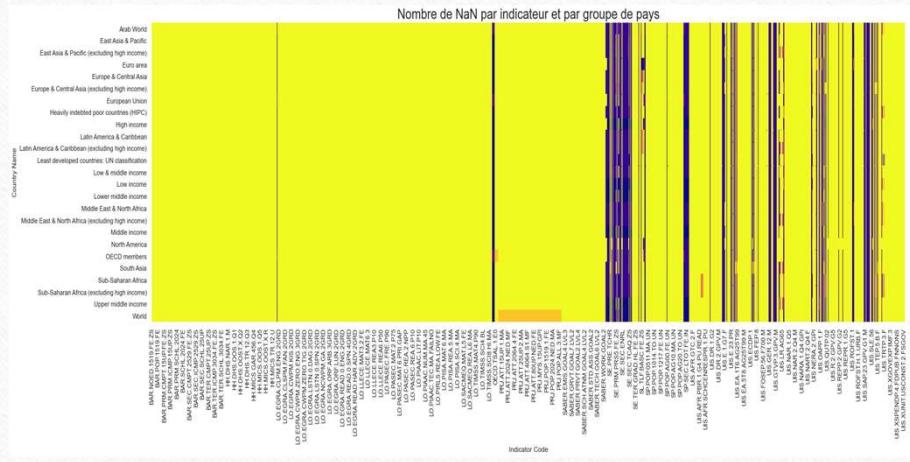
## **2 - Identifier les indicateurs exploitables**

## Quels sont les indicateurs avec le plus de données ?

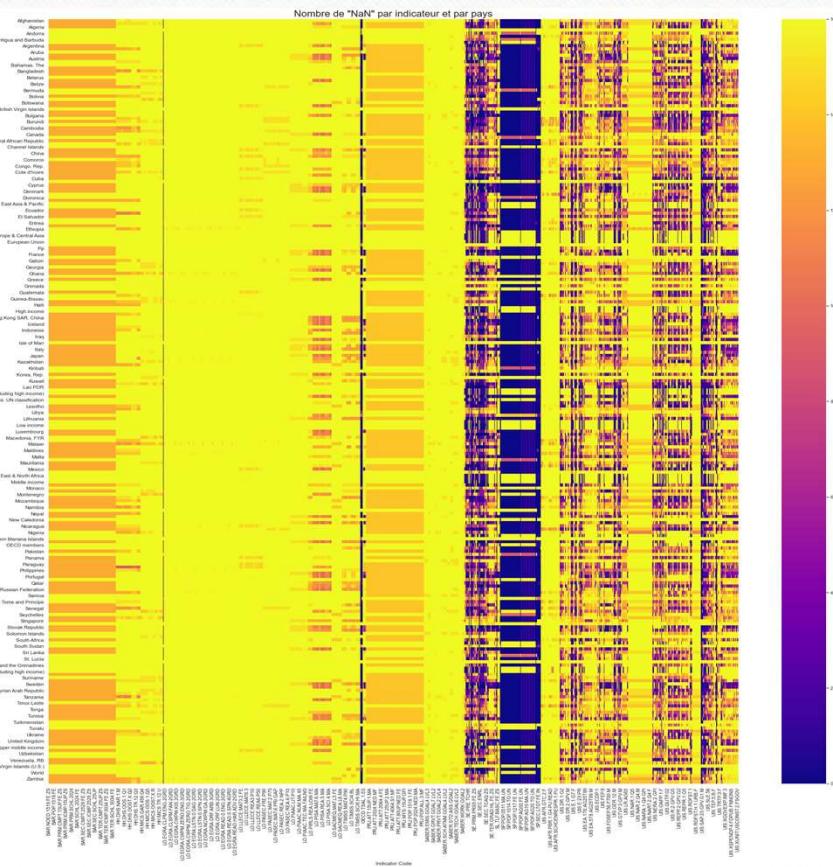


## Identifier les NaN graphiquement (Bleu = donnée manquante)

## Par zone



## Par pays



NB : 197 États reconnus par l'ONU, 22 “zones” dans notre data set

# Sélection des indicateurs

[313] :

	<b>Indicator Name</b>	<b>Indicator Code</b>
<b>0</b>	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2
<b>1</b>	Adjusted net enrolment rate, upper secondary, ...	UIS.NERA.3
<b>2</b>	GDP per capita, PPP (current international \$)	NY.GDP.PCAP.PP.CD
<b>3</b>	Government expenditure on secondary education ...	UIS.XGDP.23.FSGOV
<b>4</b>	Government expenditure on tertiary education a...	UIS.XGDP.56.FSGOV
<b>5</b>	Gross enrolment ratio, tertiary, both sexes (%)	SE.TER.ENRR
<b>6</b>	Internet users (per 100 people)	IT.NET.USER.P2
<b>7</b>	Population, ages 15-24, total	SP.POP.1524.TO.UN

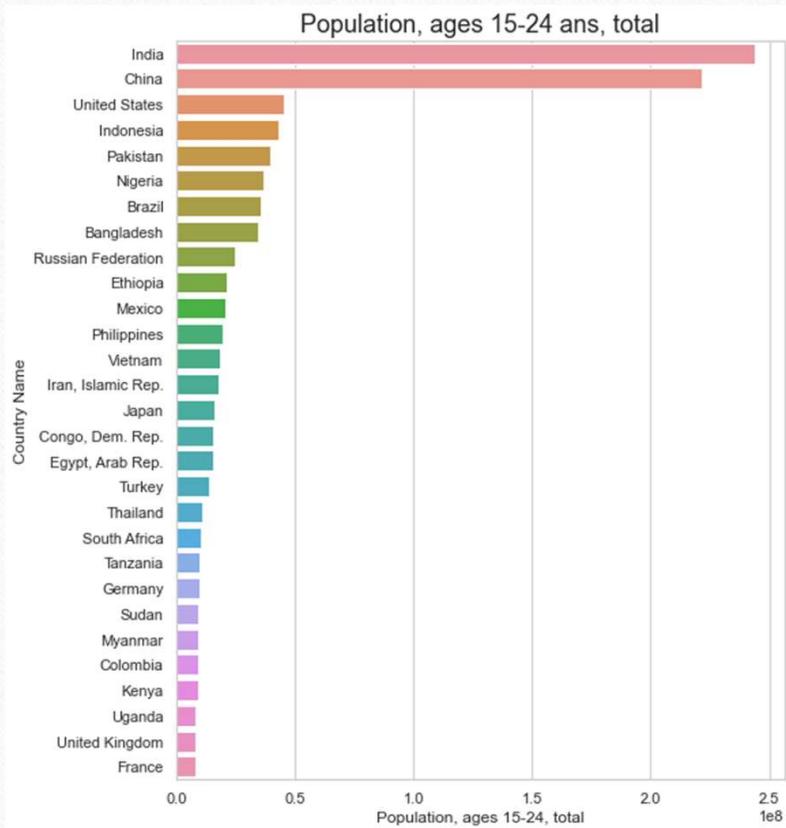
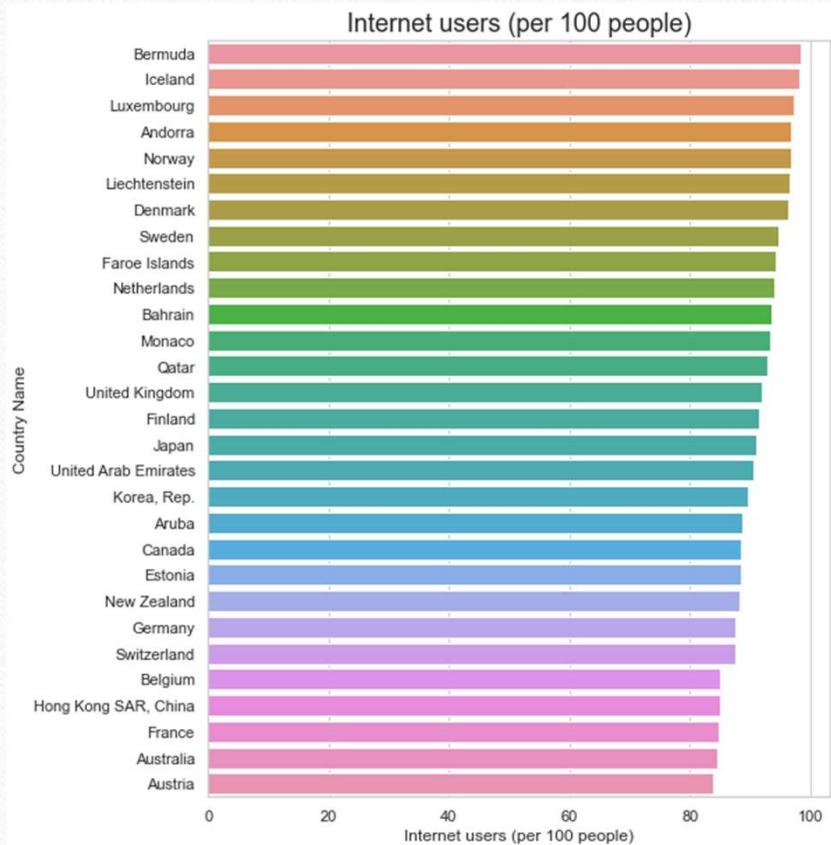
[ ] :

# Les Indicateurs retenus

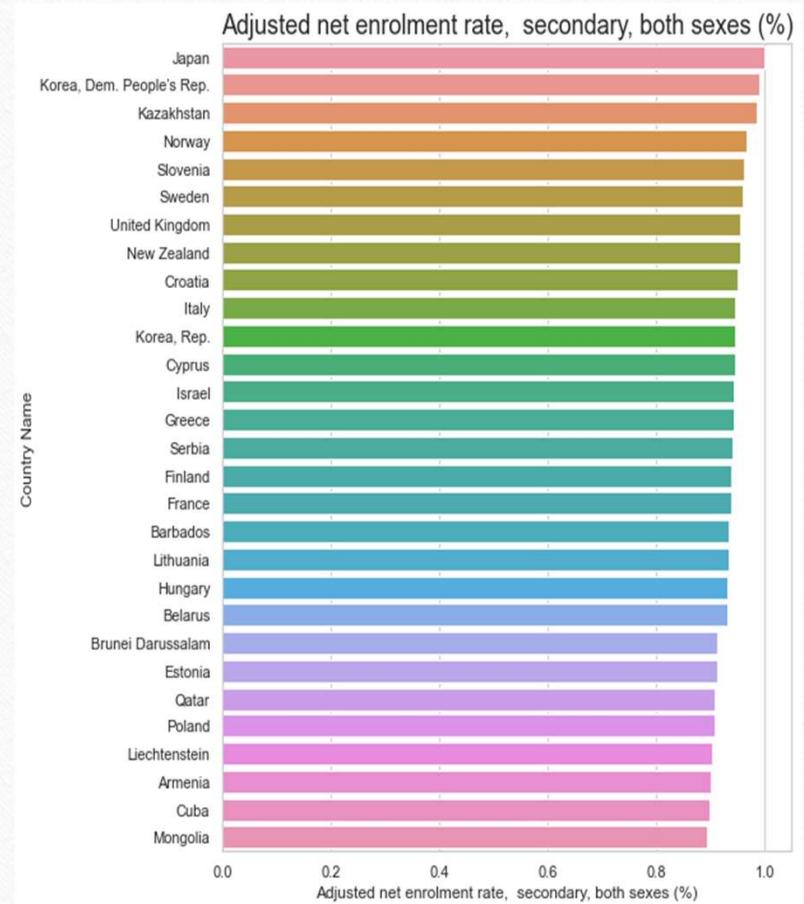
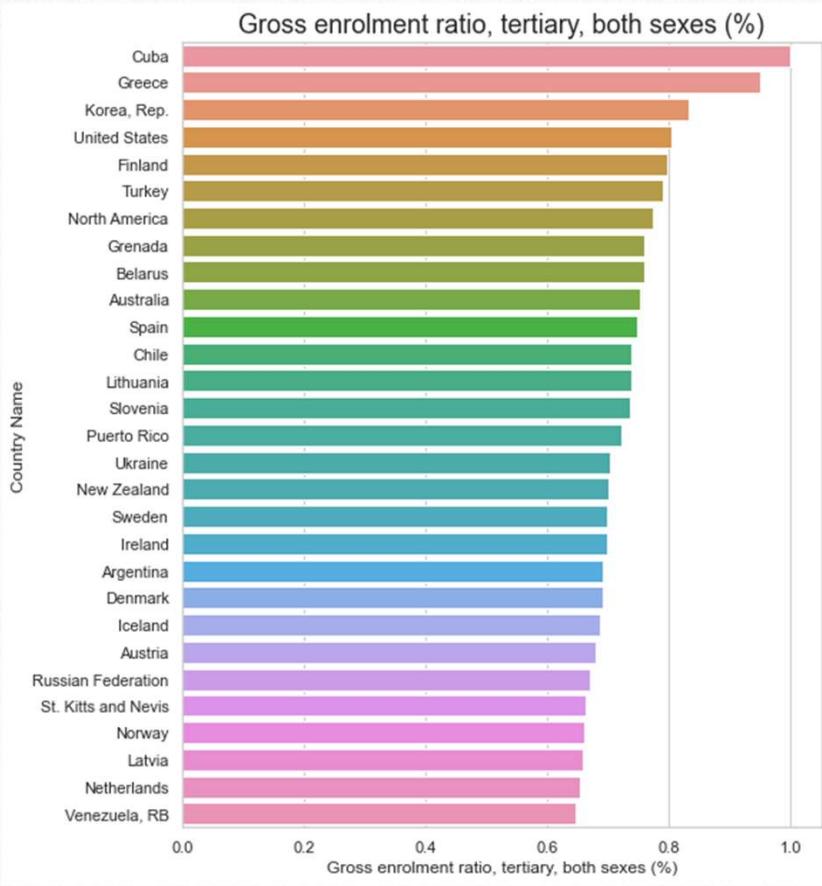
---

# Exploration des indicateurs

# Exploration des indicateurs

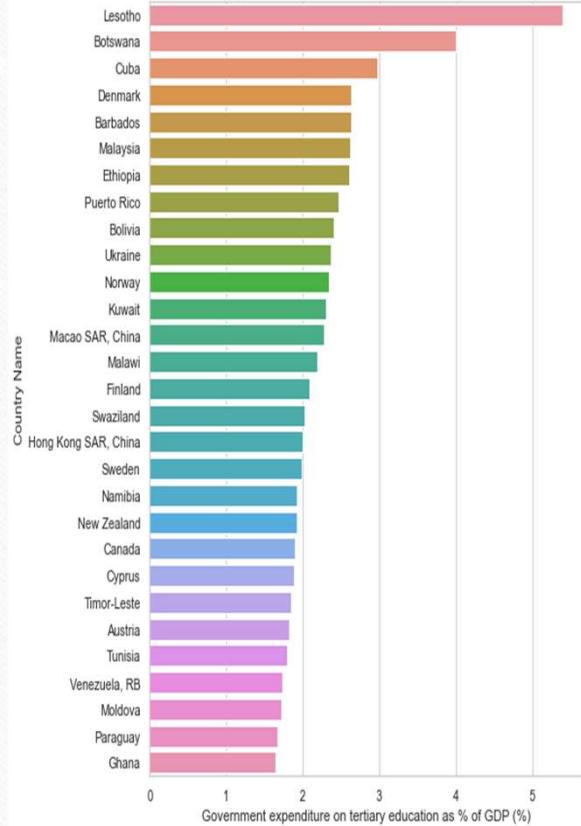


# Exploration des indicateurs

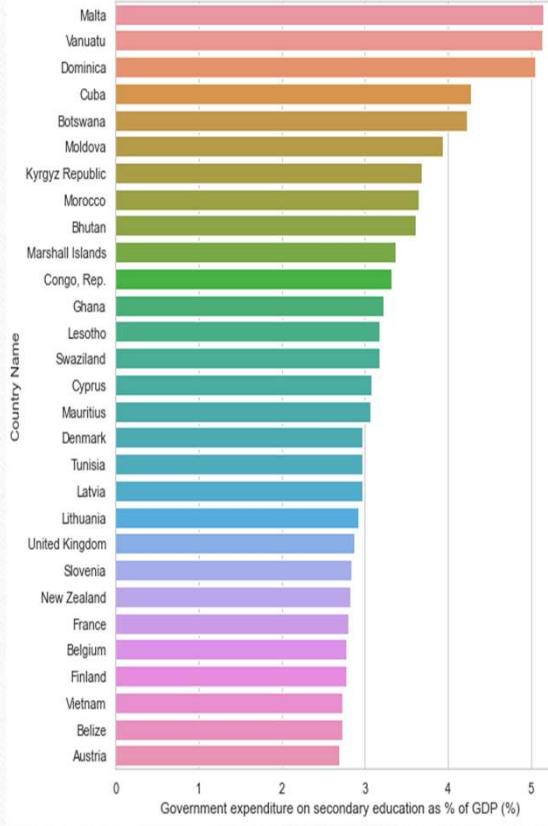


# Exploration des indicateurs

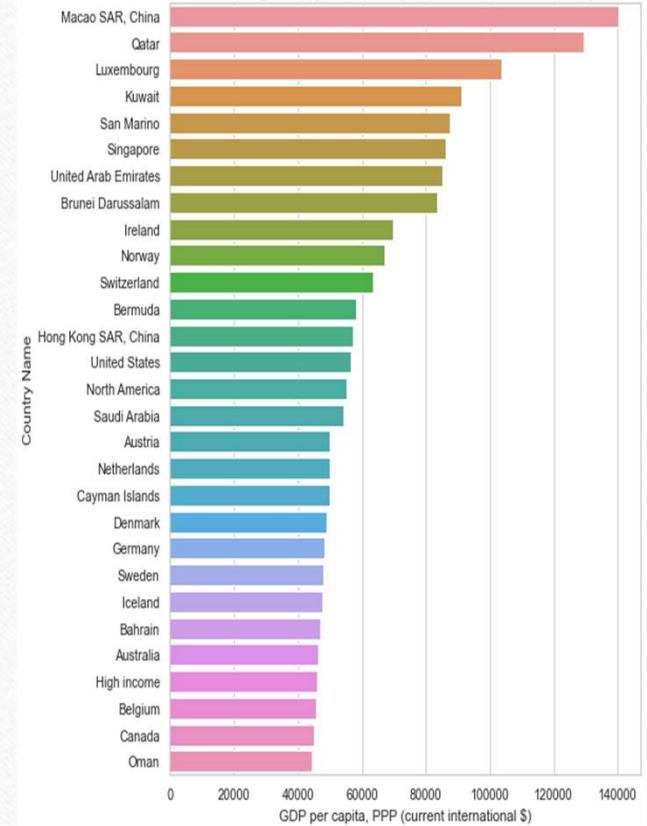
Government expenditure on tertiary education as % of GDP (%)



Government expenditure on secondary education as % of GDP (%)



GDP per capita, PPP (current international \$)

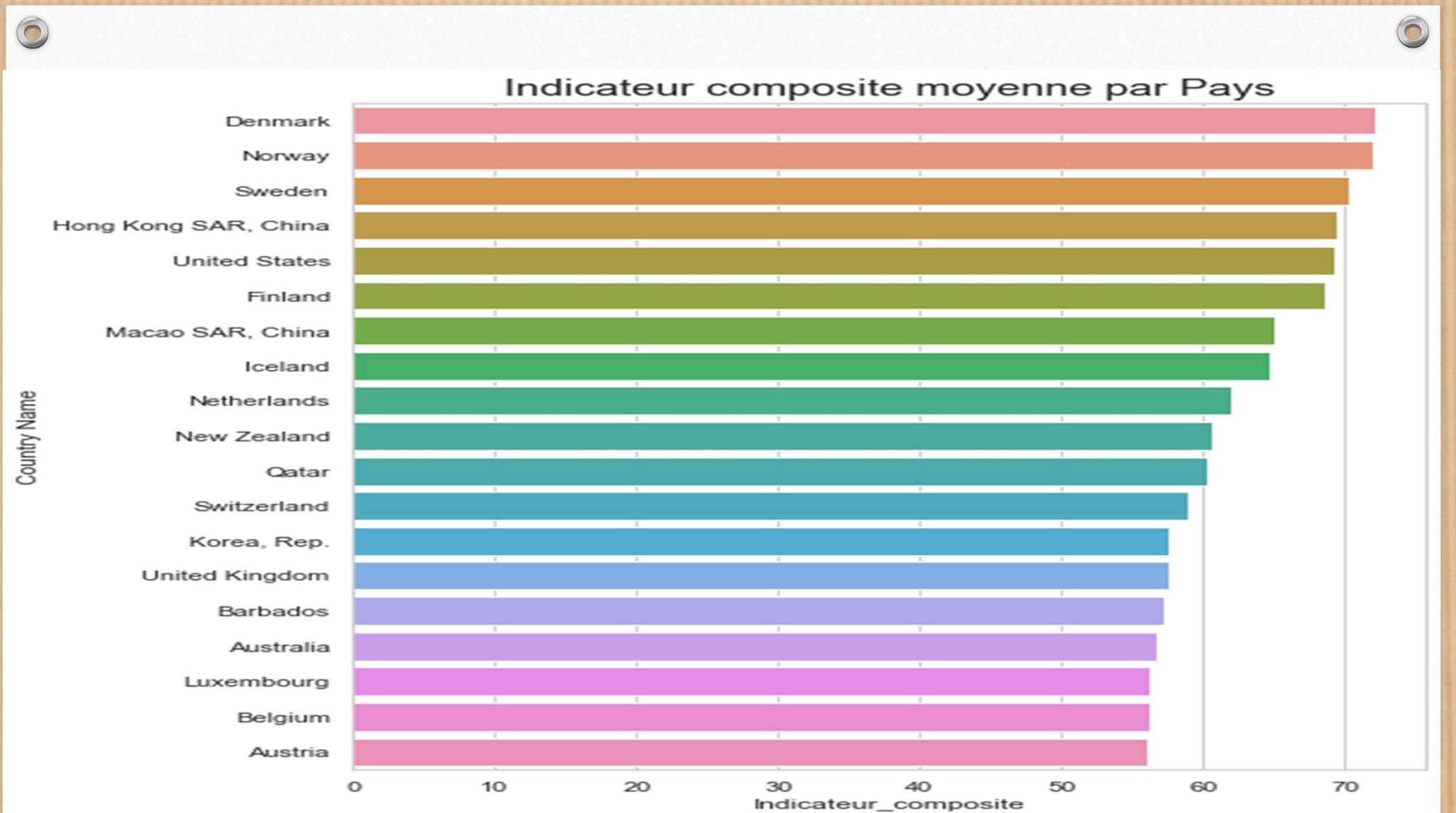


# Réponses aux questions

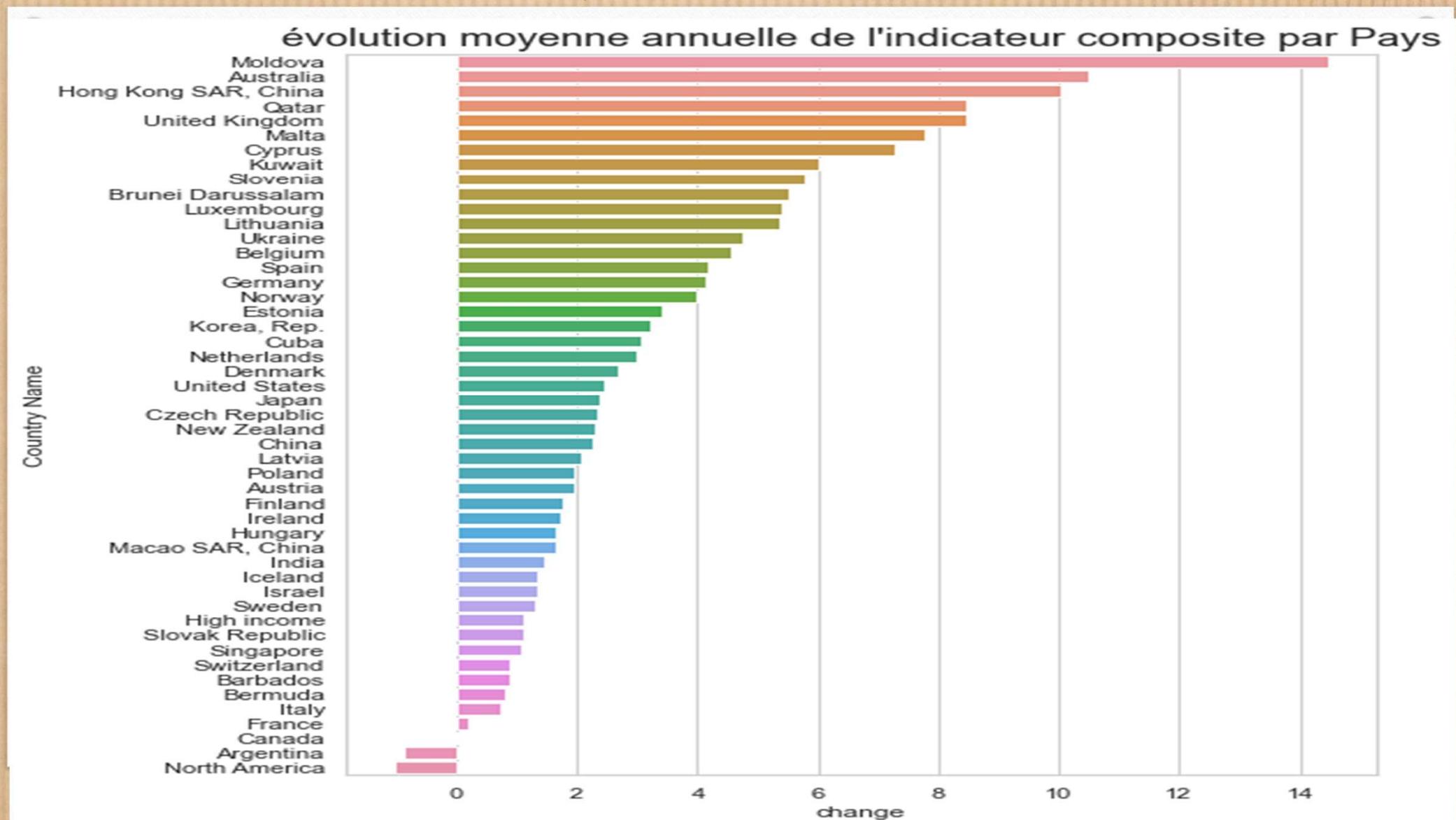
# Création de l'indicateur composite

Indicateurs	Les coefficients de pondération
1- GDP per capita, PPP (current international \$)	% 15
2- Government expenditure on secondary education as % of GDP (%)'	% 10
3- Government expenditure on tertiary education as % of GDP (%)'	% 10
4- Gross enrolment ratio, tertiary, both sexes (%)'	% 10
5 - Adjusted net enrolment rate, secondary, both sexes (%)']	% 15
6- Internet users (per 100 people)'	% 20
7- Population, ages 15-24, total'	% 20

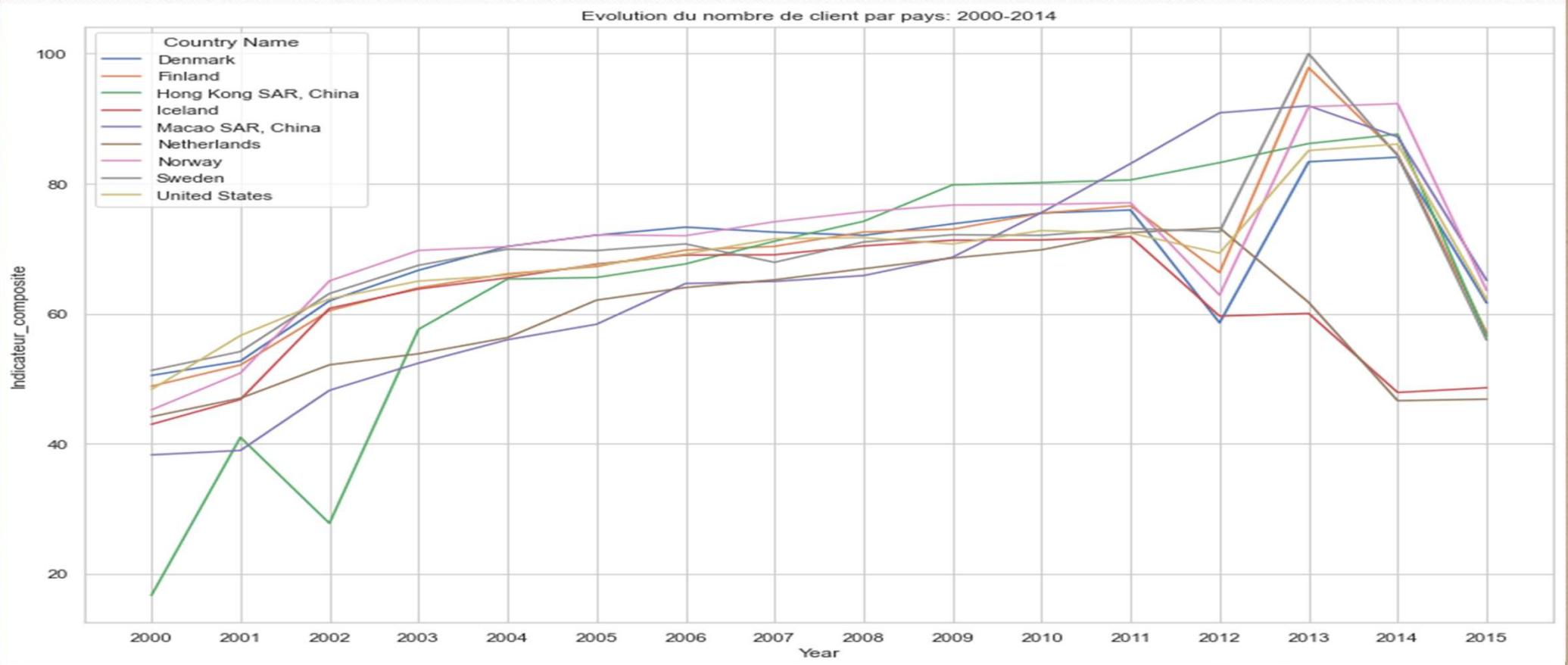
## Question 1: Quels sont les pays avec un fort potentiel de clients pour l'entreprise ?



## Question 2: Pour chacun de ces pays, quelle sera l'évolution de ce potentiel de clients ?



# Evolution du nombre de client par pays : 2000-2014



# Evolution du nombre de client par pays : 2000-2014



# Question 3:

## Dans quels pays l'entreprise doit-elle opérer en priorité ?

- 
- Australia
  - Hong Kong SAR, China
  - United Kingdom
  - Luxembourg
  - Korea, Rep.
  - Belgium
  - Qatar
  - Norway
- 

## Conclusions

---

Le jeu de données permet-il de répondre aux attentes de ACADEMY ?

---

- Merci de votre attention