

# Analysez des données de systèmes éducatifs



academy

# Sommaire :

- 1 : Objectifs et Missions
- 2 : Le jeu de donnée
- 3 : Pré Analyse des données
- 4 : Analyse des données chiffrées mondiales
- 5 : Analyse orientée entreprise
- 6 : Perspectives et manques

# 1 : Objectifs et Missions



- **Société** : propose des contenus de formation en ligne pour un **public lycéé et université**.
- **Objectif** : Expansion à l'international de l'entreprise.
- **Mission** : Analyse exploratoire, pour déterminer si les données sur l'éducation de la banque mondiale permettent d'informer le projet d'expansion.
- **Exploration** :
  - 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 ?

## 2 : Le jeu de données

- **Les données de la Banque mondiale :**

<https://datacatalog.worldbank.org/dataset/education-statistics>

- **Le jeu de données comprend 5 fichiers :**

- Un fichier Data : données chiffrées pour les différents pays et les différents indicateurs entre 1970 et 2100.
- Un fichier Country : infos supplémentaires pour les différents pays/zones du fichier Data.
- Un fichier Stats Séries : infos supplémentaires pour les différents indicateurs du fichier Data.
- Un fichier Stats Foot Note : infos supplémentaires décrivant les méthodes de calculs et l'incertitude de certains des indicateurs en fonction de l'année et du pays.
- Un fichier Country Séries : infos supplémentaires décrivant la provenance de certains des indicateurs pour certains pays.

### 3 : Pré Analyse des données

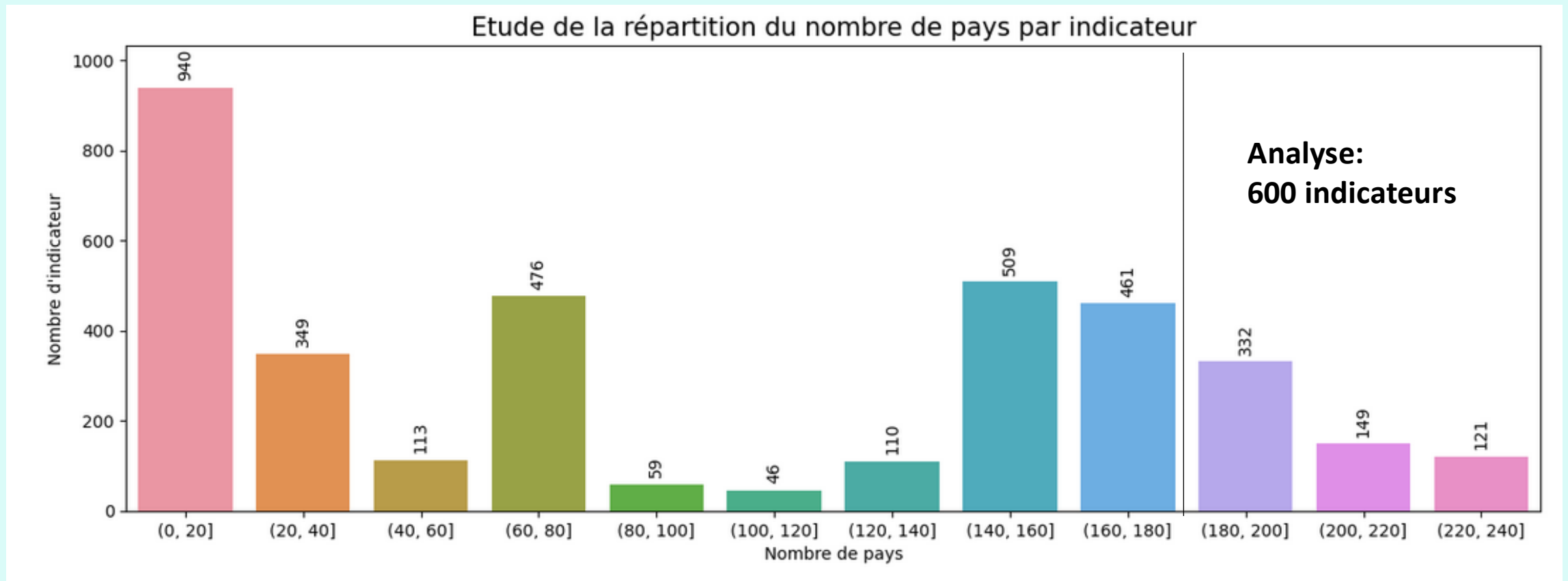
- Valider la qualité de ce jeu de données et décrire les informations contenues dans le jeu de données :

	Nblignes	NbColonnes	NbDuplicatas	NbNaNs	NbNonNaNs	PourcentageNaNs	NbPays/Zones	NbIndicateurs
NomFichier								
Data	357405	69	0	18149124	6511821	74	242	3665
Country	241	32	0	2354	5358	31	241	0
CountrySeries	613	4	0	613	1839	25	211	21
StatsFootNote	643638	5	0	643638	2574552	20	239	1558
StatsSeries	3665	21	0	55203	21762	72	0	3665

#### Analyse :

- Absence de duplicatas mais beaucoup de vide dans ces jeux de données.
- 3665 indicateurs.
- 242 pays.

- Sélectionner les informations qui semblent pertinentes pour répondre à la problématique :
  - Pour le fichier **Data** :
    - Garder les colonnes entre l'an 2000 et l'an 2020:
      - Avant 2000 les données seront trop vieilles pour être exploitées et donner une idée.
      - Après 2020 les données sont des extrapolations vu que nous sommes en 2023.
    - Garder les indicateurs regroupant plus de 180 pays.



- Pour le fichier **Country** :
  - Garder les colonnes de noms des pays.
  - Garder la colonne des groupements de pays par zones Géographiques ['**Region**']:

Short Name	
Region	
East Asia & Pacific	[American Samoa, Australia, Brunei, China, Fiji, Micronesia, Guam, Hong Kong SAR, China, Indonesia, Japan, Cambodia, Kiribati, Korea, Lao PDR, Mac...
Europe & Central Asia	[Albania, Andorra, Armenia, Austria, Azerbaijan, Belgium, Bulgaria, Bosnia and Herzegovina, Belarus, Switzerland, Channel Islands, Cyprus, Czech R...
Latin America & Caribbean	[Aruba, Argentina, Antigua and Barbuda, The Bahamas, Belize, Bolivia, Brazil, Barbados, Chile, Colombia, Costa Rica, Cuba, Curaçao, Cayman Islands...
Middle East & North Africa	[United Arab Emirates, Bahrain, Djibouti, Algeria, Egypt, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Morocco, Malta, Oman, West Bank and ...
North America	[Bermuda, Canada, United States]
South Asia	[Afghanistan, Bangladesh, Bhutan, India, Sri Lanka, Maldives, Nepal, Pakistan]
Sub-Saharan Africa	[Angola, Burundi, Benin, Burkina Faso, Botswana, Central African Republic, Côte d'Ivoire, Cameroon, Dem. Rep. Congo, Congo, Comoros, Cabo Verde, E...

- Garder la colonne des groupements de pays par Richesse ['**Income Group**']:

Short Name	
Income Group	
High income: OECD	[Australia, Austria, Belgium, Canada, Switzerland, Chile, Czech Republic, Germany, Denmark, Spain, Estonia, Finland, France, United Kingdom, Greec...
High income: nonOECD	[Aruba, Andorra, United Arab Emirates, Antigua and Barbuda, Bahrain, The Bahamas, Bermuda, Barbados, Brunei, Channel Islands, Curaçao, Cayman Isla...
Low income	[Afghanistan, Burundi, Benin, Burkina Faso, Bangladesh, Central African Republic, Dem. Rep. Congo, Comoros, Eritrea, Ethiopia, Guinea, The Gambia,...
Lower middle income	[Armenia, Bolivia, Bhutan, Côte d'Ivoire, Cameroon, Congo, Cabo Verde, Djibouti, Egypt, Micronesia, Georgia, Ghana, Guatemala, Guyana, Honduras, I...
Upper middle income	[Angola, Albania, Argentina, American Samoa, Azerbaijan, Bulgaria, Bosnia and Herzegovina, Belarus, Belize, Brazil, Botswana, China, Colombia, Cos...

Hypothèse : A ce stade on peut supposer que l'entreprise devra privilégier les pays peuplés et riches pour se développer.

- Pour le fichier **Stats Séries (Indicateur)** :
  - La colonne Topic permet de ranger les indicateurs (3665) par catégorie (37).
  - La colonne Longue Définition permet de comprendre mieux les indicateurs.
  - La Périodicité est intéressante car elle ne comprend que 99 indicateurs ayant une mesure annuelle dont 54 parlant de la population totale:

```
'Personal computers (per 100 people)',
'Internet users (per 100 people)',
'GDP (current US$)',
'GDP (constant 2010 US$)',
'GDP, PPP (current international $)',
'GDP, PPP (constant 2011 international $)',
'GDP per capita (current US$)',
'GDP per capita (constant 2010 US$)',
'GDP per capita, PPP (current international $)',
'GDP per capita, PPP (constant 2011 international $)',
'GNI (current US$)',
'GNI, PPP (current international $)',
'GNI per capita, Atlas method (current US$)',
'GNI per capita, PPP (current international $)',
'Prevalence of HIV, total (% of population ages 15-49)',
'Mortality rate, under-5 (per 1,000 live births)',
'Labor force with advanced education (% of total working-age population with advanced education)',
'Labor force with basic education (% of total working-age population with basic education)',
'Labor force with intermediate education (% of total working-age population with intermediate education)',
'Labor force, total',
'Share of youth not in education, employment or training, total (% of youth population)',
'Unemployment, total (% of total labor force) (modeled ILO estimate)',
'Population ages 0-14, total',
'Population ages 0-14 (% of total)',
'Population ages 15-64, total',
```

```
'Age population, age 0, total, UNESCO',
'Age population, age 01, total, UNESCO',
'Age population, age 02, total, UNESCO',
'Age population, age 03, total, UNESCO',
'Age population, age 04, total, UNESCO',
'Age population, age 05, total, UNESCO',
'Age population, age 06, total, UNESCO',
'Age population, age 07, total, UNESCO',
'Age population, age 08, total, UNESCO',
'Age population, age 09, total, UNESCO',
'Age population, age 10, total, UNESCO',
'Age population, age 11, total, UNESCO',
'Age population, age 12, total, UNESCO',
'Age population, age 13, total, UNESCO',
'Age population, age 14, total, UNESCO',
'Age population, age 15, total, UNESCO',
'Age population, age 16, total, UNESCO',
'Age population, age 17, total, UNESCO',
'Age population, age 18, total, UNESCO',
'Age population, age 19, total, UNESCO',
'Age population, age 20, total, UNESCO',
'Age population, age 21, total, UNESCO',
'Age population, age 22, total, UNESCO',
'Age population, age 23, total, UNESCO',
'Age population, age 24, total, UNESCO',
'Age population, age 25, total, UNESCO',
'Population growth (annual %)',
'Population, total']
```

## Analyse :

- Un calcul montre que les indicateurs annuels sont remplis à 47% contre 7% pour les autres.
- Ces 54 indicateurs sont présent dans plus de 180 pays de la DF Data.
- Les 45 autres indicateurs sont les mêmes mais axés homme ou femme (utile pour une future démarche commerciale).

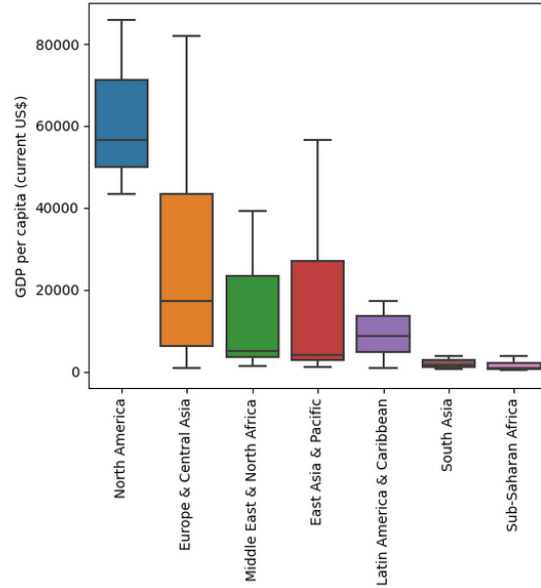


- Pour le fichier **Stats Séries (Indicateur)** :
  - Choix des indicateurs manuels pour l'entreprise :
    - **Nombre d'ordinateurs et accès internet pour 100 personnes** : intéressant pour vendre des cours en ligne.
    - **PIB par habitant** : intéressant car le coût des cours en ligne peut être important.
    - **La force de travail total et le pourcentage de chômeurs** : intéressant si les gens veulent une remise à niveau sur certaines matières pour changer de poste ou d'entreprise voir même de domaine et en l'absence d'aide de l'état c'est la population qui a le plus les moyens financiers.
    - **La population total et la croissance démographie** : taille du pays et développement futur.
    - **Les indicateurs de différents âges** : création d'indicateurs en fonction du niveau d'éducation et savoir la quantité de personnes par niveau et donc les clients actuels et futurs.
    - **Autres indicateurs personnalisés** : nombre d'utilisateurs d'internet et enfants hors système scolaire.
  - Différentes stratégies commerciales :
    - Nombre de clients actuellement disponibles pour ces cours.
    - Nombre de clients pour le futur.
    - Axer les cours pour une remise à niveau (force de travail et chômeurs) ou juste pour les étudiants.
- Les deux autres fichiers sont intéressants mais manque d'informations pour bien les comprendre et surtout ne prennent pas en compte tous les indicateurs, toutes les années et tous les pays.

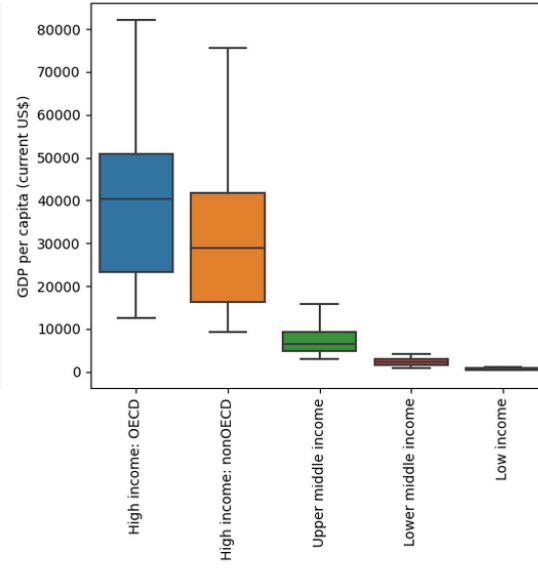
Region	East Asia & Pacific	Europe & Central Asia	Latin America & Caribbean	Middle East & North Africa	North America	South Asia	Sub-Saharan Africa
Indicator Name							
Collegiens	4572561.300000	866627.300000	1411113.000000	1424183.700000	9363668.000000	16977249.100000	2003717.200000
Etudiants	8927181.700000	1741995.300000	2370430.700000	2518463.100000	17899803.000000	28940538.200000	2812054.200000
GDP per capita (current US\$)	15452.400000	28253.900000	10471.000000	15628.000000	61844.300000	2725.500000	2196.200000
Internet users (per 100 people)	45.900000	72.500000	53.100000	55.400000	87.100000	25.600000	18.200000
Labor force, total	42685180.400000	8891865.700000	9716732.100000	6854207.000000	90255019.500000	83628353.500000	8525866.000000
Lyceens	3618886.600000	660198.900000	1045748.900000	1029320.000000	7113575.500000	12670593.000000	1369409.700000
NbUtilisateurInternet	33486858.900000	12144127.700000	9256228.400000	9070581.300000	90340440.300000	50591095.400000	3845841.200000
OutOfCollege	1573749.500000	44141.600000	170758.400000	444521.900000	2533880.000000	6617774.700000	1062460.200000
OutOfFac	8990811.000000	927932.400000	1995482.600000	1713848.500000	13015004.000000	24305765.800000	2379717.200000
OutOfLycee	1189521.300000	50602.000000	249192.700000	277232.700000	640612.500000	4613956.800000	865530.800000
Population growth (annual %)	1.100000	0.400000	0.900000	2.300000	0.600000	1.600000	2.600000
Population, ages 0-14, total	14567832.800000	3275381.300000	4612273.300000	6147492.300000	33690641.500000	64838693.900000	9160693.200000
Population, ages 15-64, total	51216158.100000	12312542.300000	12039644.700000	13308844.900000	118297426.000000	141333867.500000	11470517.300000
Population, total	62714691.200000	15919133.500000	15390836.800000	20426748.800000	118936821.000000	218023838.000000	21279529.000000
Unemployment, total (% of total labor force)	6.100000	10.100000	8.700000	10.300000	6.100000	4.700000	9.500000

# Etude du PIB, et des utilisateurs d'internet par région et par niveau de richesse (Income):

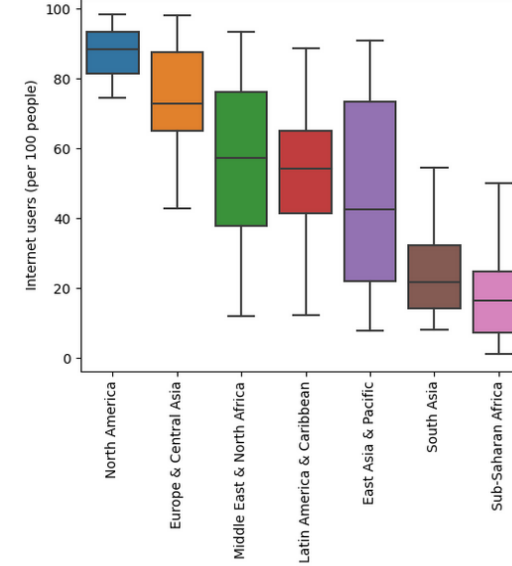
Etude de la répartition de "GDP per capita (current US\$)" en fonction :  
Des Régions.



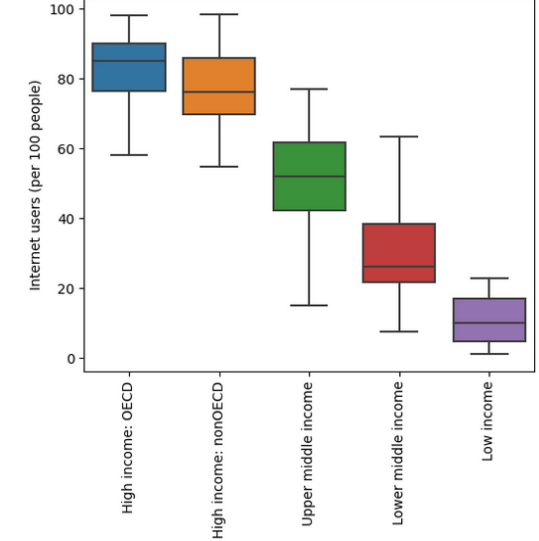
Des Incomes.



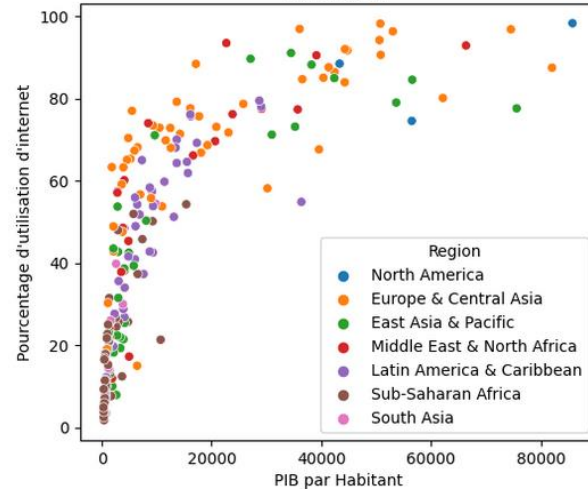
Etude de la répartition de "Internet users (per 100 people)" en fonction :  
Des Régions.



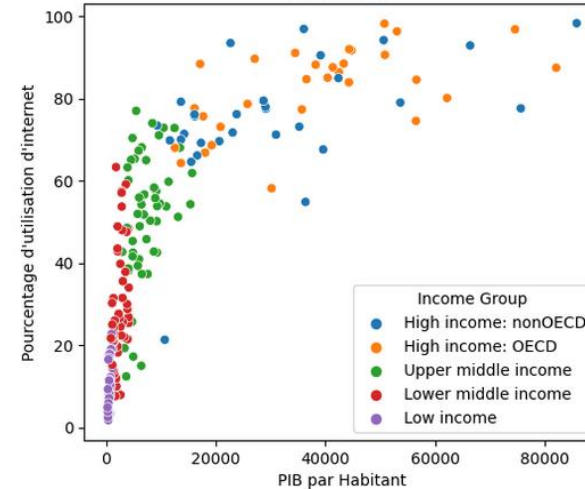
Des Incomes.



Etude de la corrélation entre les utilisateur d'internet(100 habitants) et le PIB par habitant  
Par Region



Par Income



## Analyse :

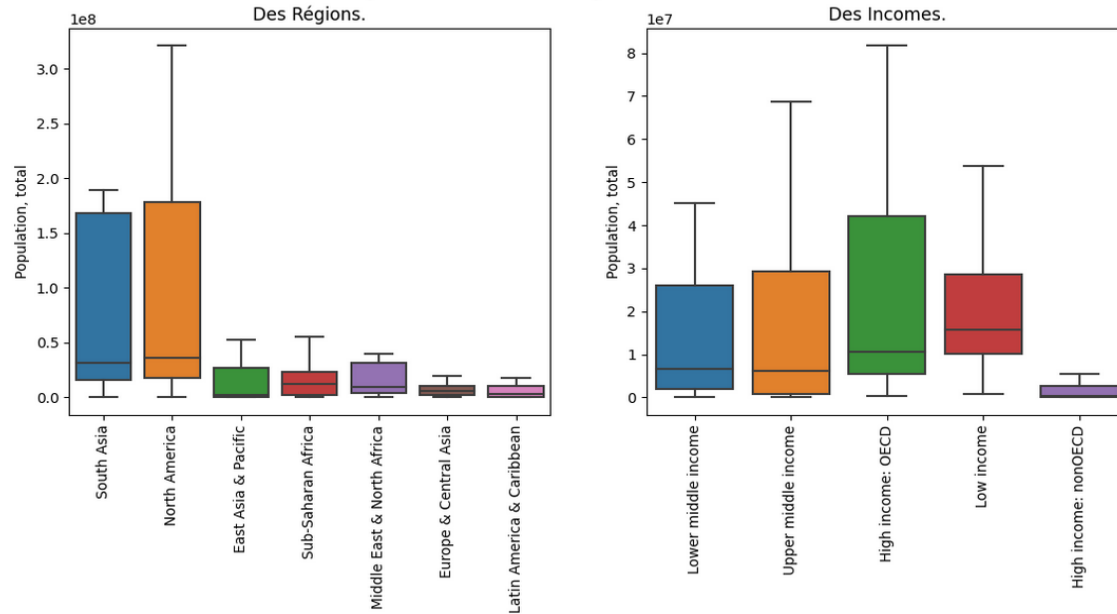
- Comme attendu les zones les plus riches sont les zones OECD et non OECD.
- Les zones géographique et Income utilisant le plus internet sont les mêmes zones que pour le PIB.
- On observe une belle corrélation (0,7) entre ces deux indicateurs.

## Conclusion :

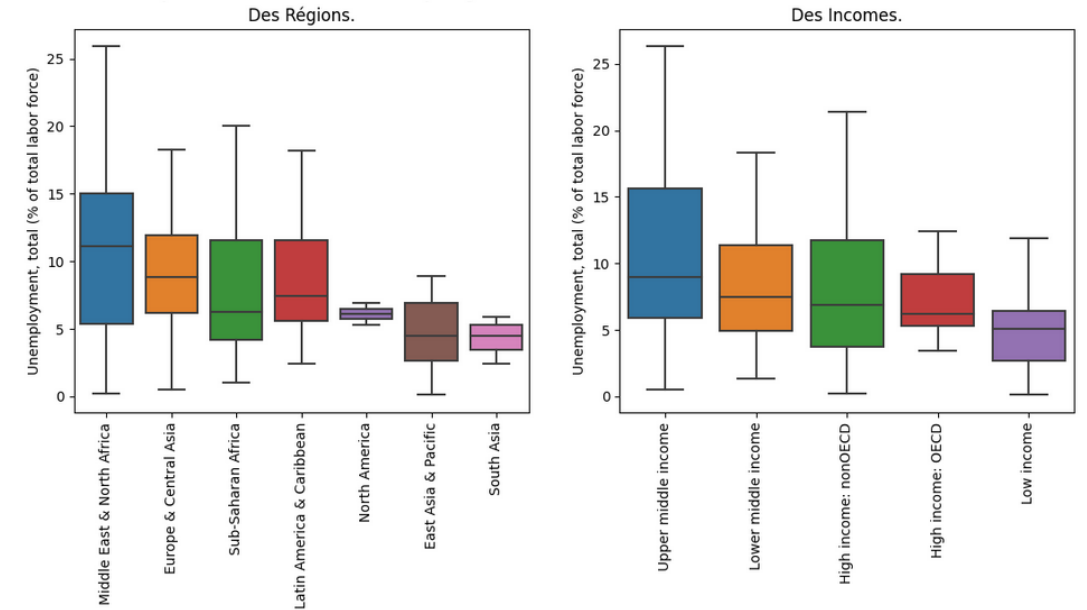
- L'entreprise devra donc s'orienter de préférence vers les pays riches pour son business plan.

# Etude la population totale, du taux de chômage et de la croissance démographique :

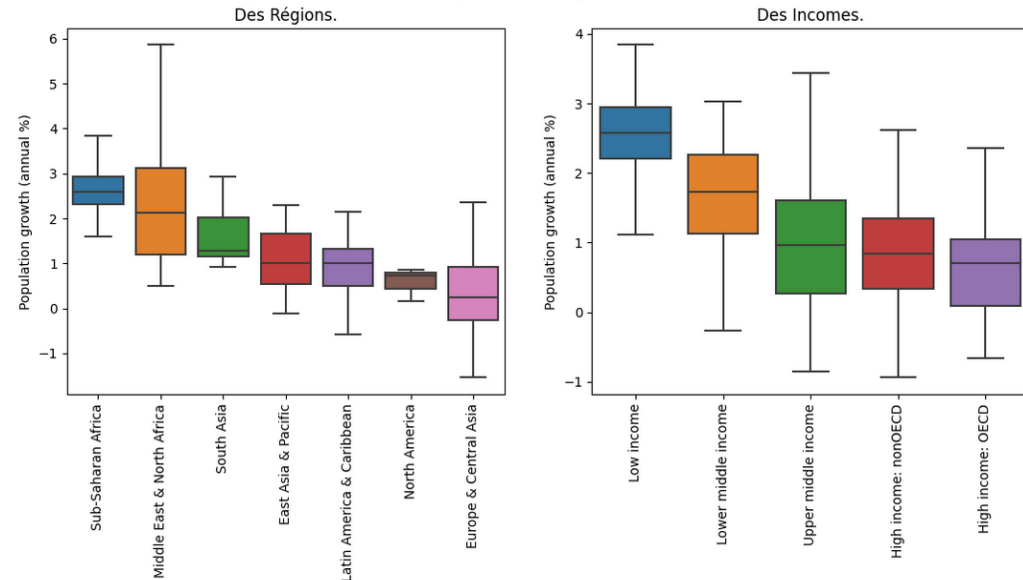
Etude de la répartition de "Population, total" en fonction :



Etude de la répartition de "Unemployment, total (% of total labor force)" en fonction :



Etude de la répartition de "Population growth (annual %)" en fonction :



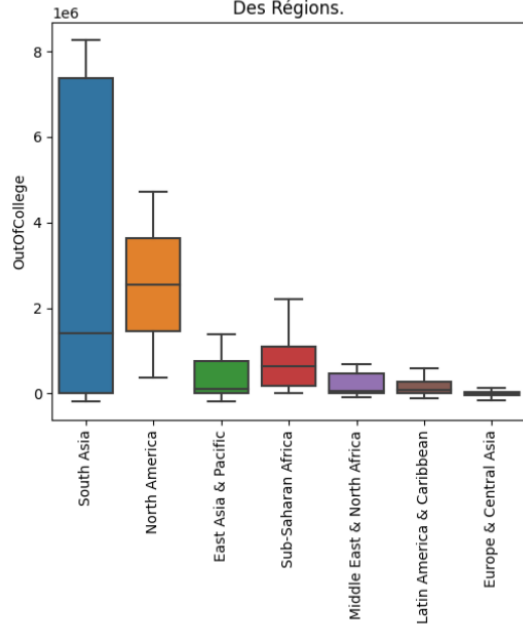
## Analyse :

- Les zones les plus peuplées sont l'Asie et l'Amérique du nord et les zones moyennes riches.
- Les zones les plus touchés par le chômage sont l'Afrique, l'Asie centrale et l'Europe et les zones moyennement riches
- Les zones avec la plus forte croissance démographique sont l'Afrique et les pays des zones pauvres.

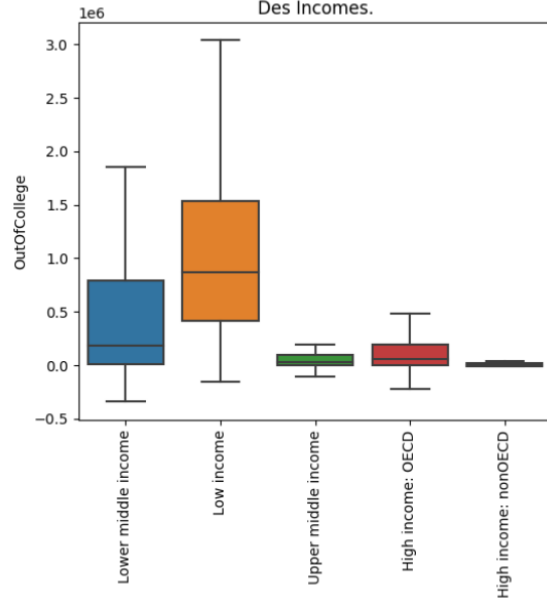
# Etude des enfants/jeunes en dehors du système scolaire :

Etude de la répartition de "OutOfCollege" en fonction :

Des Régions.

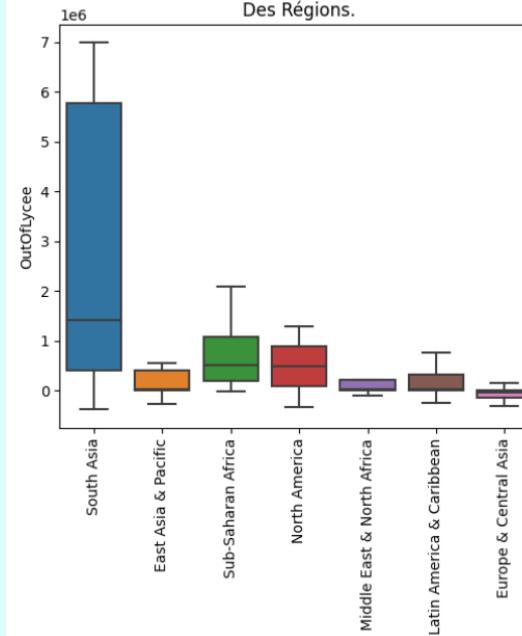


Des Incomes.

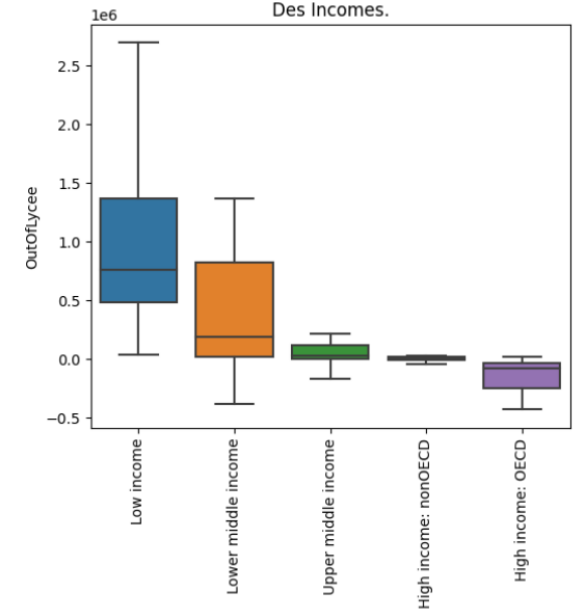


Etude de la répartition de "OutOfLycee" en fonction :

Des Régions.

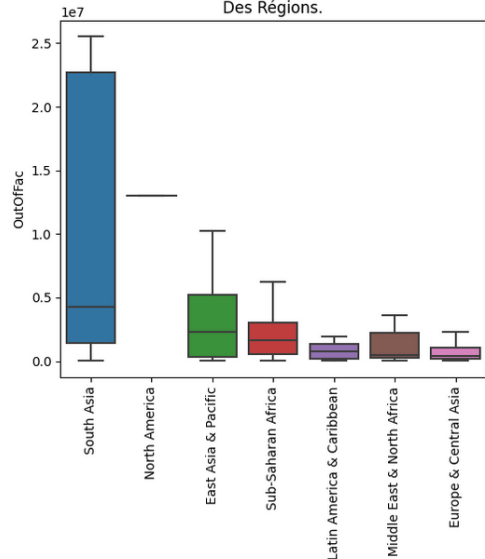


Des Incomes.

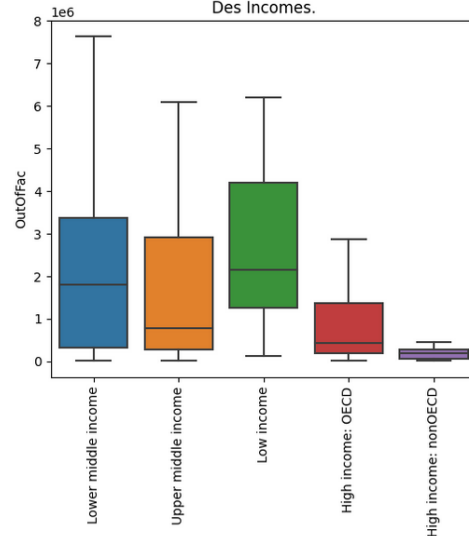


Etude de la répartition de "OutOfFac" en fonction :

Des Régions.



Des Incomes.



## Analyse :

- Les enfants allant le moins à l'école sont surtout situés en Asie du Sud en Amérique du Nord (en l'absence des données du Canada) et en Afrique.
- Les zones les plus pauvres sont les plus déscolarisés.

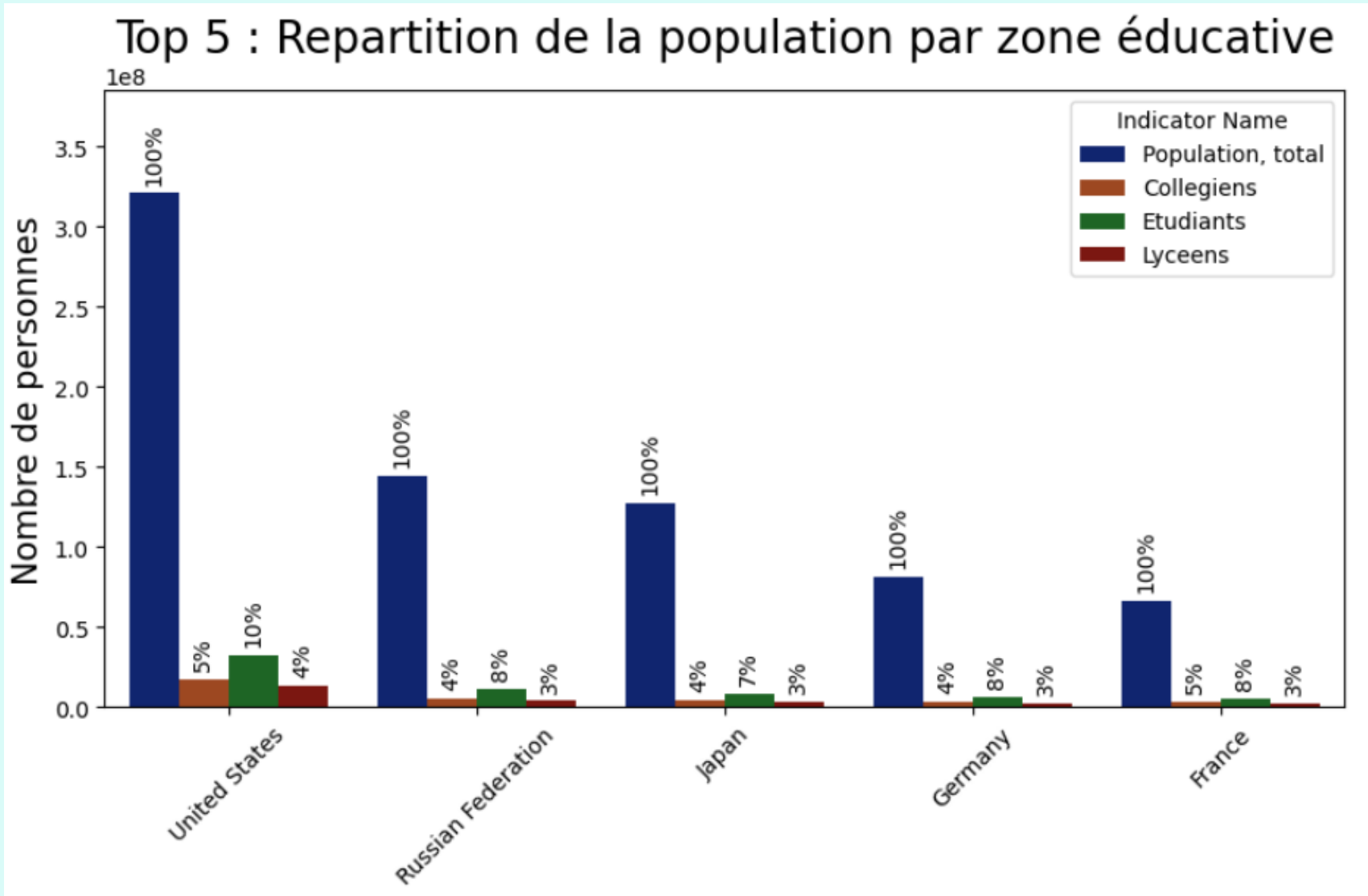
## 5 : Analyse des pays utiles pour l'entreprise

- Orientation de l'analyse vers les pays les plus riches pour lesquels on a des données.
- Ces pays ayant une forte variété de langue il est préférable de ne garder que les pays qui ont une population suffisante pour investir dans la création et la traduction d'un site web.
- Mesures montrant que sur les 35 pays riches la moitié à plus de 3 millions d'habitants et dont la moyenne est à 17 millions d'habitants.
- Les analyses suivantes sont donc réalisées sur les pays avec plus de 3.5 millions d'habitants.

count	75.0
mean	17318900.0
std	44617004.0
min	31264.0
25%	162778.0
50%	2904910.0
75%	10078631.0
max	320896618.0

'Australia', 'Austria', 'Belgium', 'Chile', 'Croatia', 'Czech Republic', 'Denmark', 'Finland', 'France', 'Germany', 'Greece', 'Hong Kong SAR, China', 'Ireland', 'Israel', 'Italy', 'Japan', 'Korea, Rep.', 'Kuwait', 'Lithuania', 'Netherlands', 'New Zealand', 'Norway', 'Oman', 'Poland', 'Portugal', 'Russian Federation', 'Saudi Arabia', 'Slovak Republic', 'Spain', 'Sweden', 'Switzerland', 'United Arab Emirates', 'United Kingdom', 'United States', 'Uruguay'

# Etude de la répartition de la population par niveau éducatif

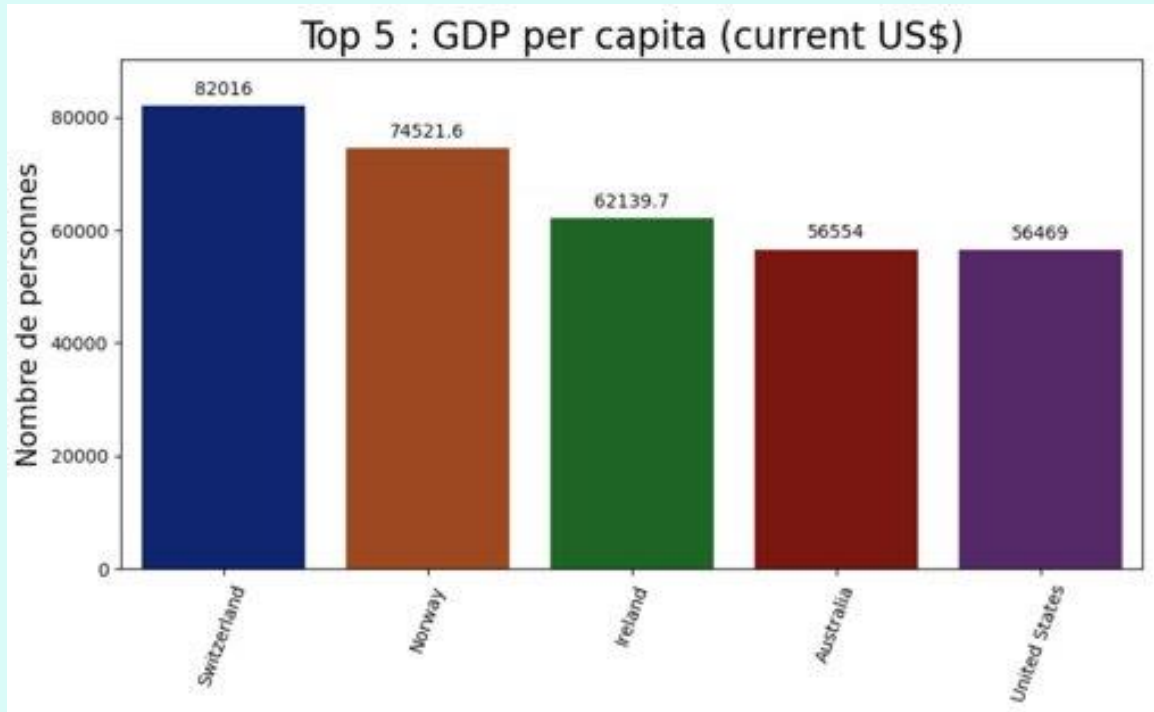


## Analyse :

- On remarque une proportion équivalente de collégiens, lycéens et étudiants entre les différents pays.

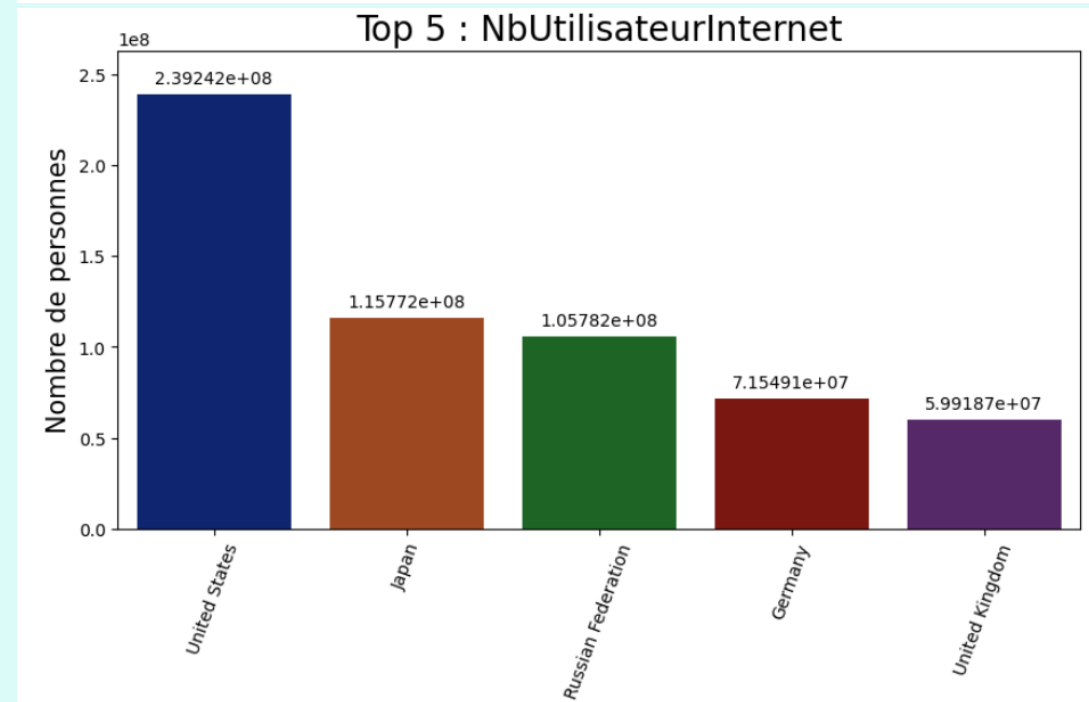
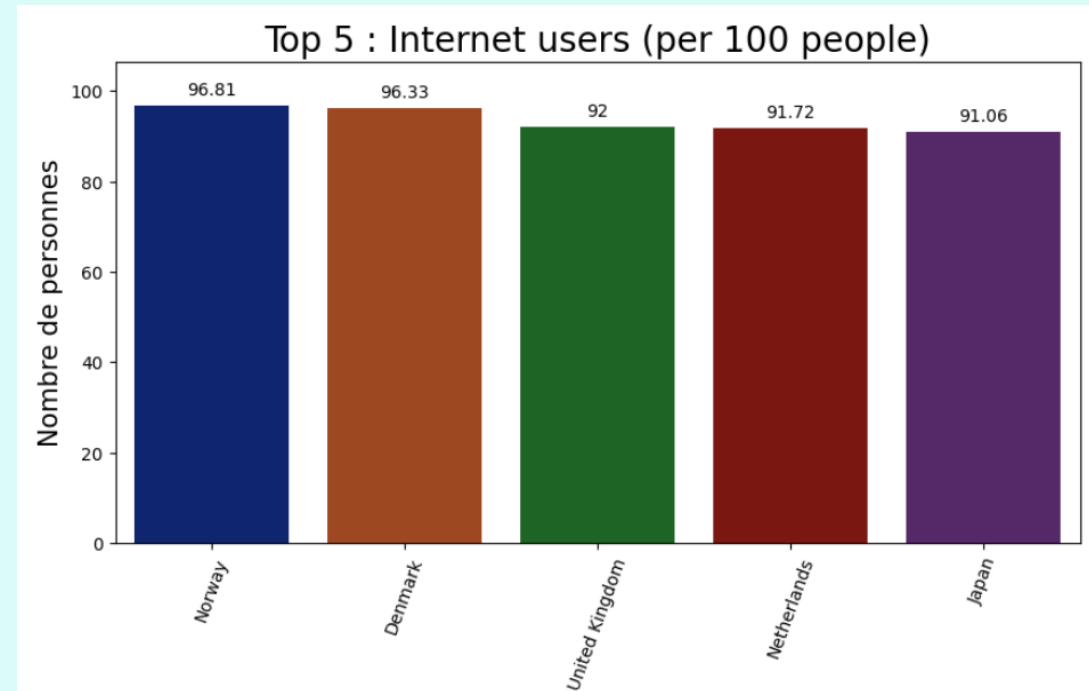


## Etude du PIB et des utilisateurs d'internet



### Analyse :

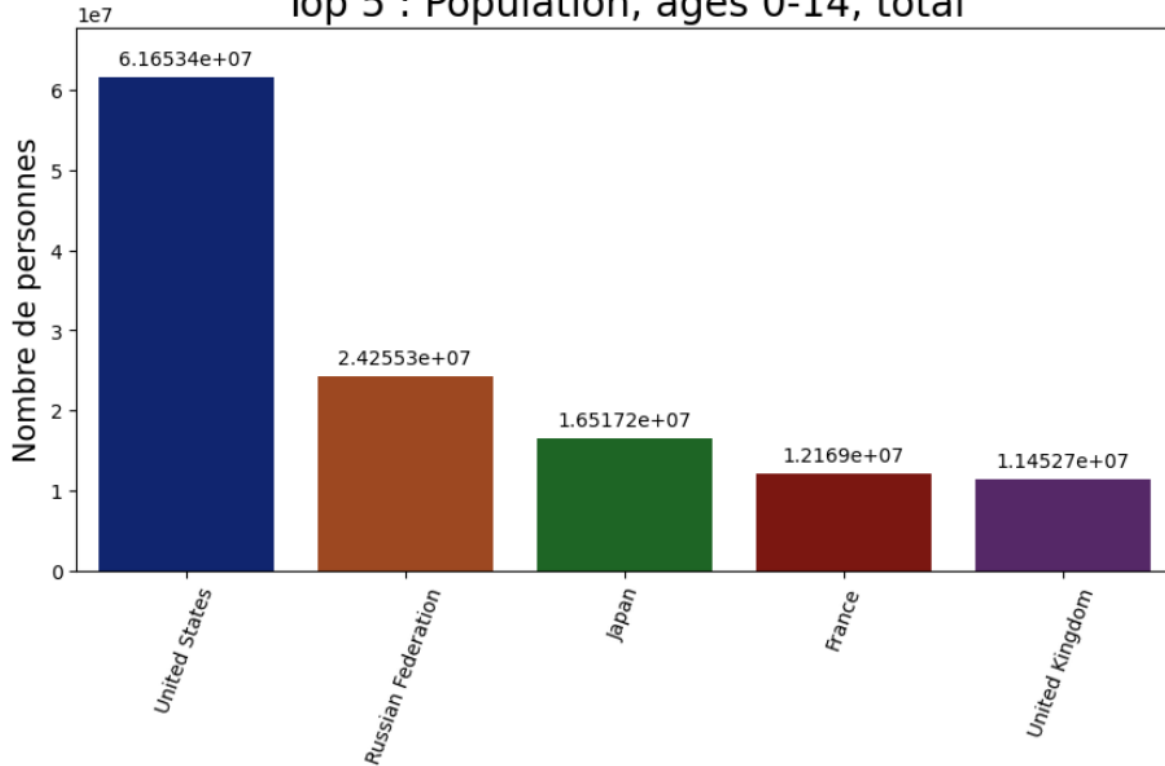
- Le pays le plus riche est la suisse suivi par la Norvège.
- Les plus gros utilisateurs d'internet sont la Norvège et le Danemark.
- Mais il y a plus d'utilisateurs aux USA et au Japon (pays les plus peuplés).



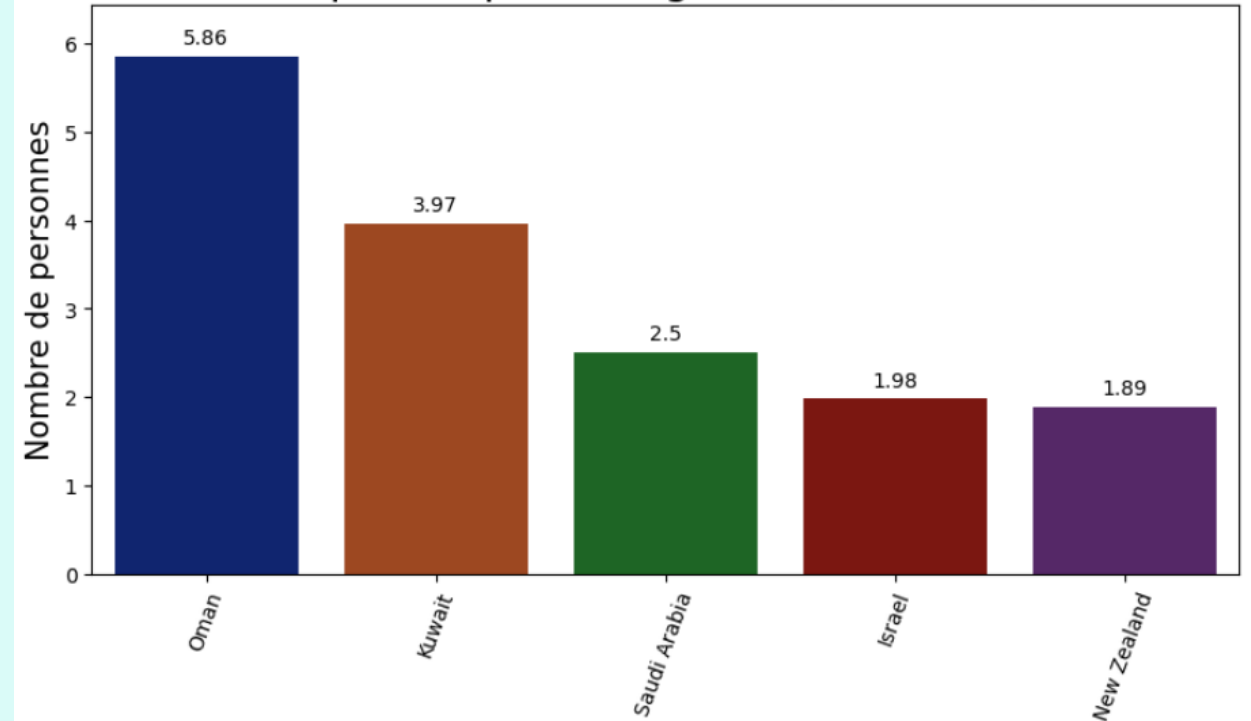


## Etude des indicateurs de futurs clients

Top 5 : Population, ages 0-14, total



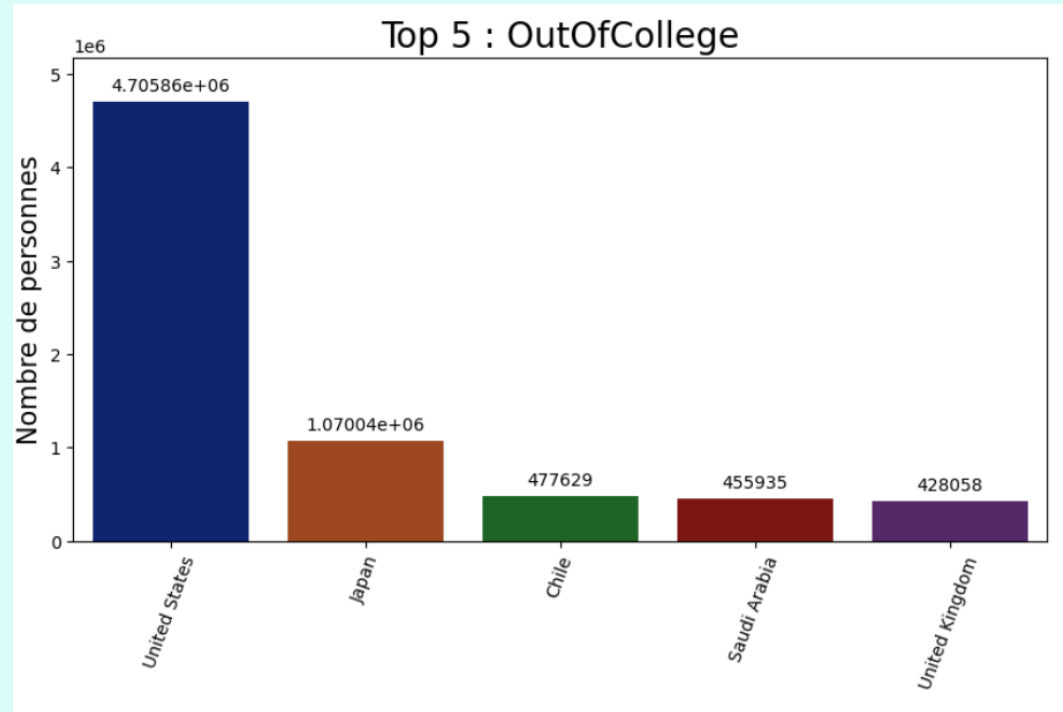
Top 5 : Population growth (annual %)



### Analyse :

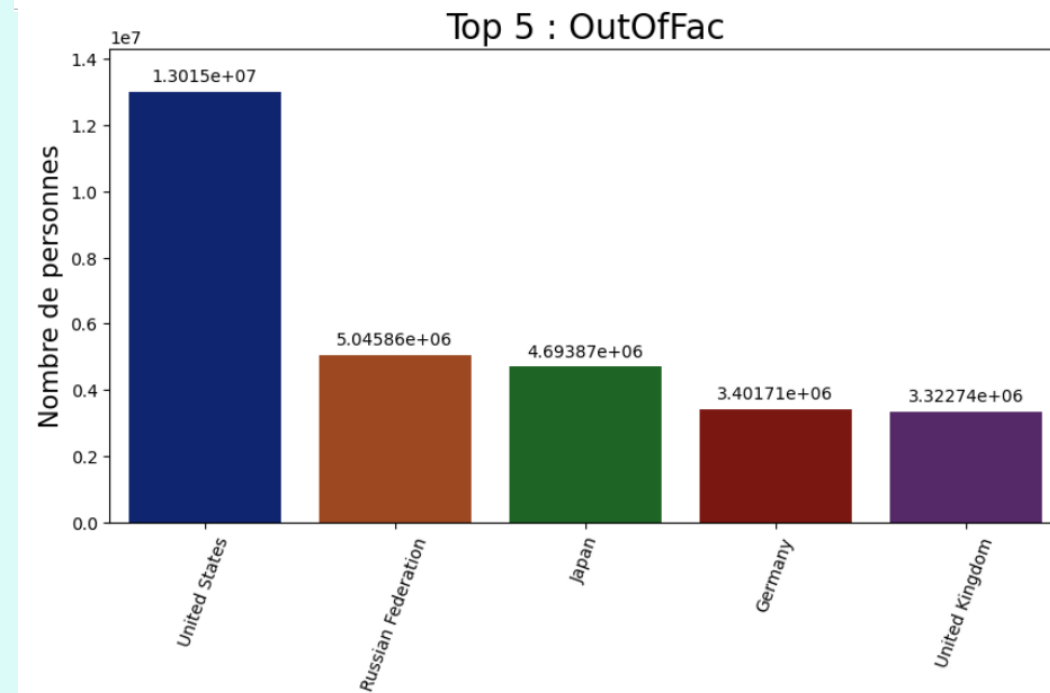
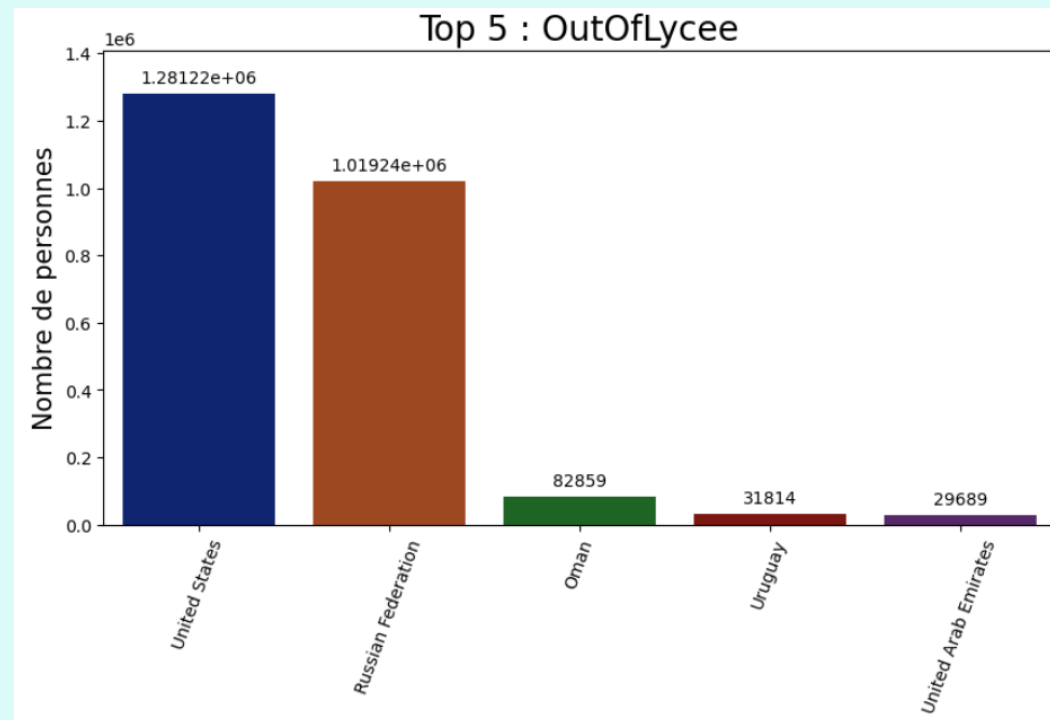
- On retrouve la plupart des pays les plus peuplés pour l'étude de la population 0-14ans.
- Les pays ayant la plus forte croissance démographique sont Oman et le Koweït.

# Etude des indicateurs OutOfEducation



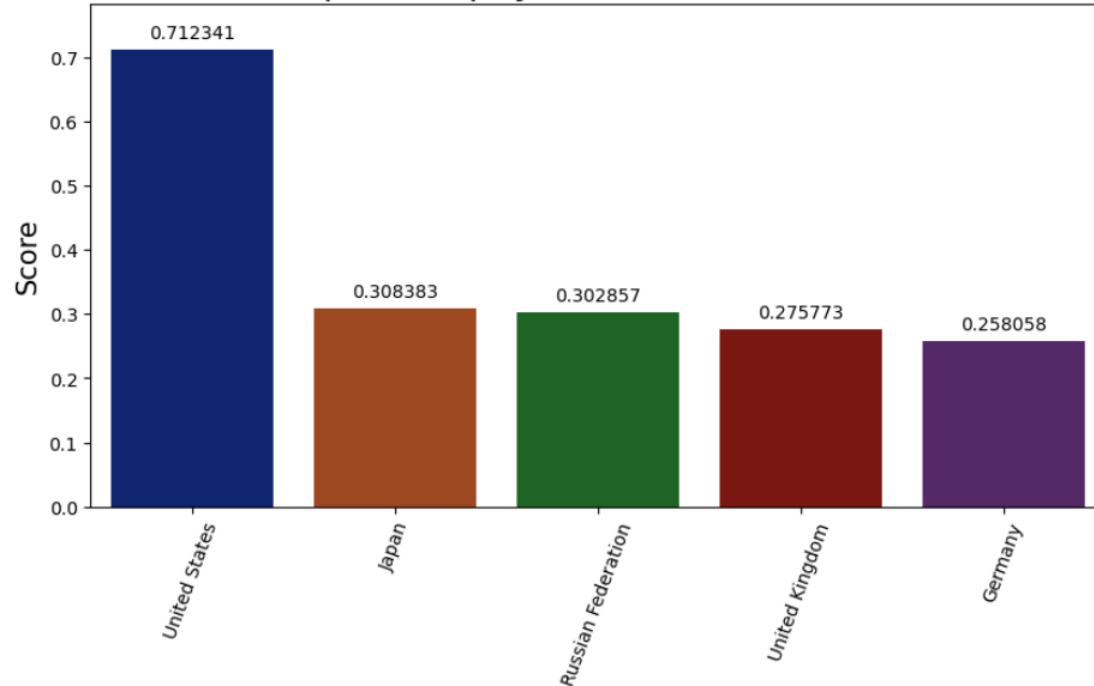
## Analyse :

- Les USA ont le plus grand nombre d'enfants et de jeunes en dehors du système scolaire.
- Ensuite nous avons la Russie et le Japon.



# Etude via des scores regroupant les indicateurs

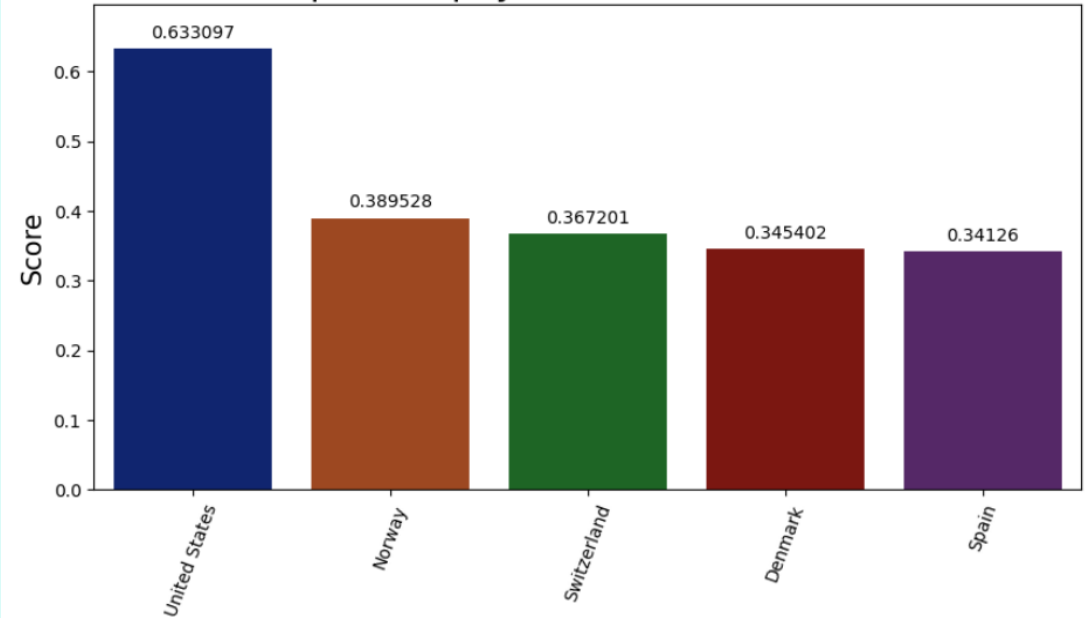
Top 5 des pays ou investir : Global



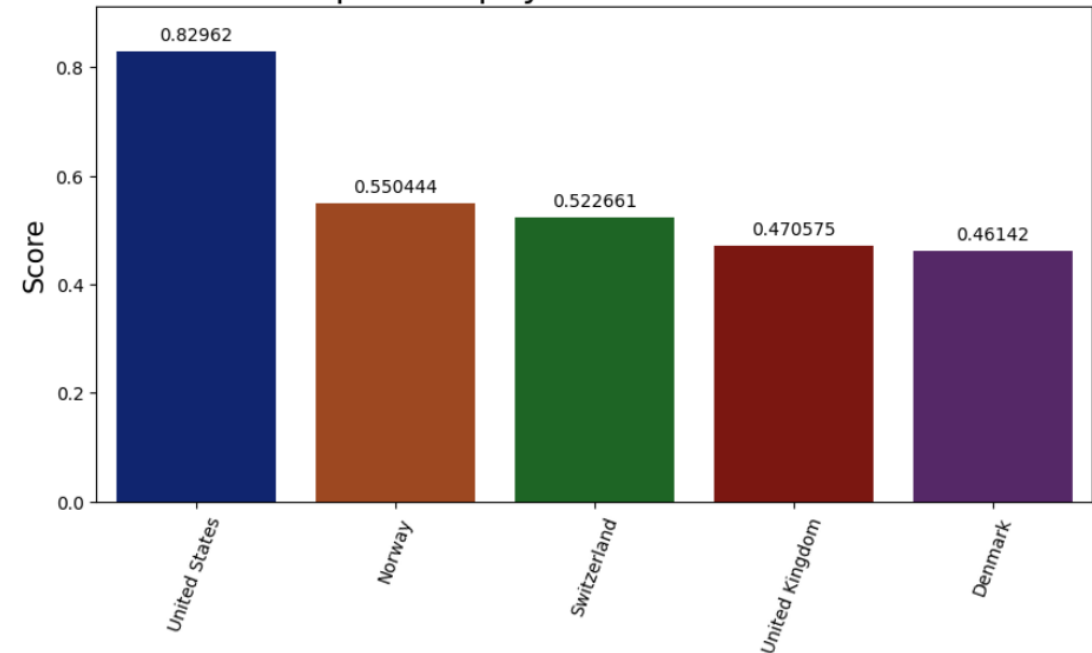
## Analyse :

- De manière générale les pays prioritaire ou investir sont, les USA, la Norvège, la Suisse, l'Angleterre, le Japon et le Danemark.

Top 5 des pays ou investir : Present

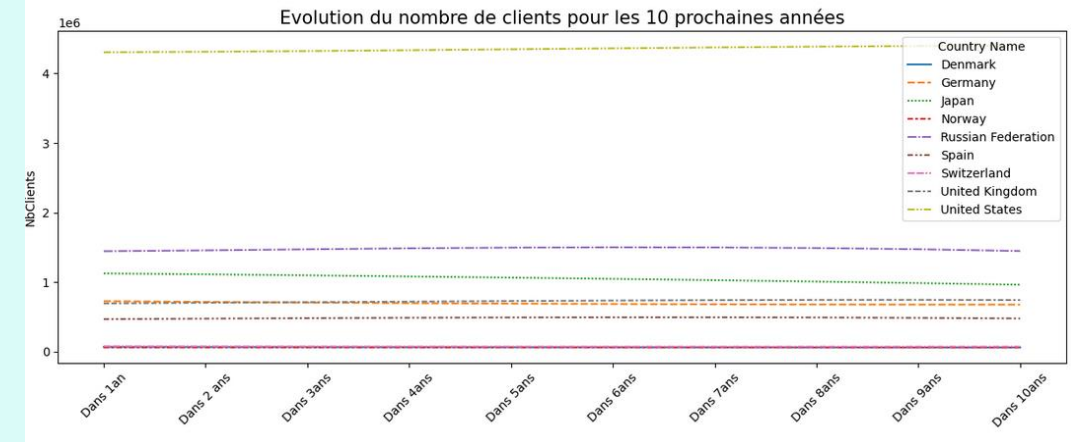


Top 5 des pays ou investir : Futur



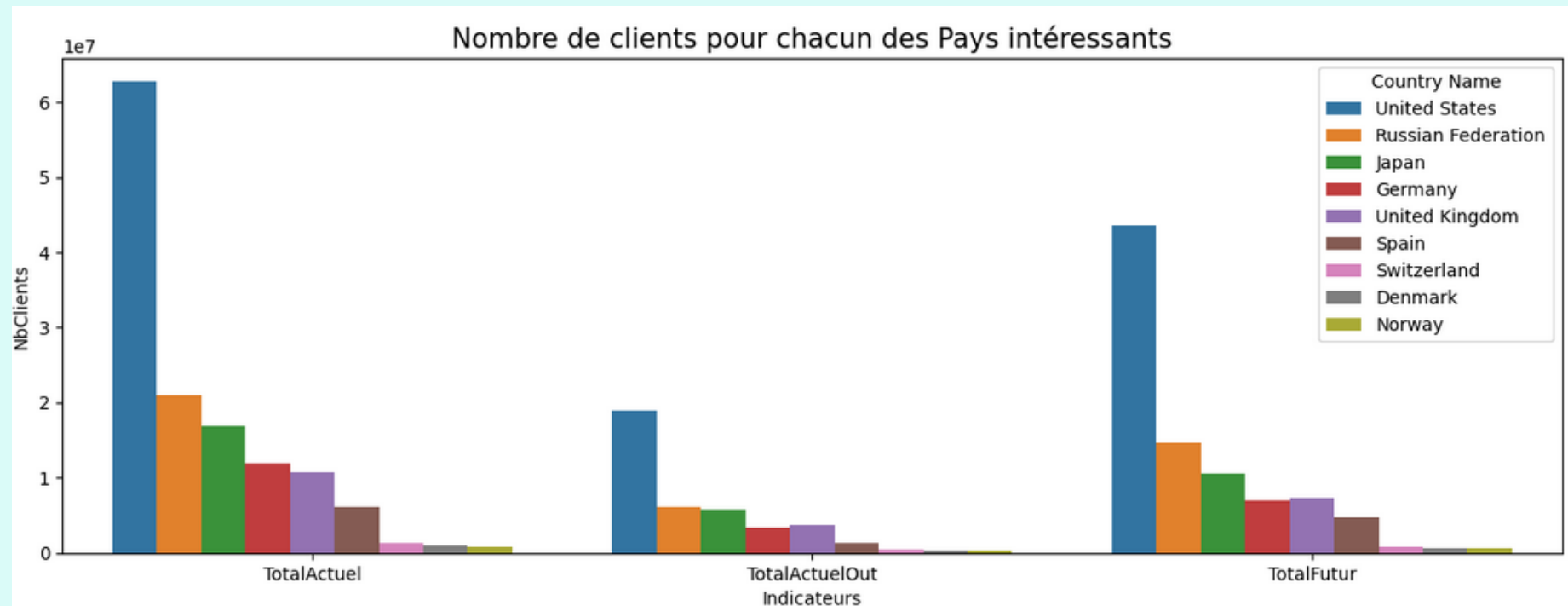
## Pays ou investir et clients futurs

Country Name	TotalActuel	TotalActuelOut	TotalFutur
United States	62763595.0	19002090.0	43546263.0
Russian Federation	20926472.0	6065103.0	14751594.0
Japan	16897623.0	5763909.0	10500115.0
Germany	11934828.0	3401706.0	6918085.0
United Kingdom	10658551.0	3750801.0	7249091.0
Spain	6099219.0	1277354.0	4829252.0
Switzerland	1302862.0	473973.0	729040.0
Denmark	979127.0	214007.0	617827.0
Norway	891979.0	243604.0	580120.0



### Analyse :

- Les clients potentiels sont proportionnels à la population totale des pays.



## 6 : Perspectives et manques

**Pour approfondir l'analyse et la pertinence des données il manque :**

- Des données récentes et non extrapolées au moins jusqu'à 2021 (surtout post Covid19).
- Des informations sur la langue du pays, pour regrouper les pays par langue (traduction des cours en ligne).
- Sur les potentiels concurrents.
- Sur le niveau d'éducation des pays et sur le coût et le programme éducatif.
- Sur les potentielles aides de l'état ou des collectivités des différents pays.