

# Projet d'expansion à l'international



**Start-up de la EdTech**

**Spécialisé dans la formation en ligne**  
**Public: niveau Lycée et Enseignement Supérieur**

# Ma Mission :

Analyser les données de la banque mondiale sur les statistiques de l'éducation afin de pouvoir répondre aux interrogations d'Academy ... :

- ❖ 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 ?

... Et pour cela on va explorer la pertinence et la qualité des données, afin de l'informer sur son projet de croissance externe.



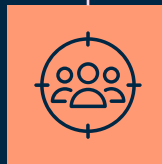
# Projet de stratégie de croissance externe : développement commercial à l'international

01

Importation des données,  
vérification du typage des  
colonnes, étude de la qualité,  
duplicatas

03

Sélection des indicateurs,  
Analyse statistique, jointure de  
fichiers, visualisation



02

Sélection, filtrage, imputation

04

Score, prédiction, conclusion

## 1

## Importation des données, vérification du typage des colonnes, étude de la qualité, duplicatas

|                           | Nb lignes | Nb colonnes (variables) | Typage des colonnes | Taux de "NaN" | Taux remplissage moyen | Doublons | Description  |
|---------------------------|-----------|-------------------------|---------------------|---------------|------------------------|----------|--|
| EdStatsData.csv           | 886930    | 70                      | object et float     | 86,10%        | 13.9%                  | 0        | Informations sur les pays ainsi que l'évolution des indicateurs (nom et code) sur la période de 1970 à 2017 (historique) et de 2020 à 2100 (données prospectives);   |
| EdStatsCountry.csv        | 241       | 32                      | object              | 30,50%        | 69.5%                  | 0        | Données sur les pays (leurs indicatifs, leurs noms, leurs codes de désignation de la norme ISO, leurs appartenances géographiques par région, de son groupe de revenu, etc), des données de leurs systèmes économiques (au niveau agrégé), leurs dernières études, recensements, enquêtes... |
| EdStatsCountry-Series.csv | 613       | 4                       | object              | 25,00%        | 75.0%                  | 0        | Description des données sur le pays (nom et code) et ses sources.  |
| EdStatsSeries.csv         | 3665      | 21                      | object et float     | 71,70%        | 28.3%                  | 0        | Description des indicateurs (code, nom, définitions), unité de mesures et autres données.  |
| EdStatsFootNote.csv       | 643638    | 5                       | object              | 20,00%        | 80.0%                  | 0        | Informations sur le code pays, le code indicateurs, l'année et une description   |

# 2

## Sélection, filtrage, imputation : Le jeu de données EdStatsData.csv

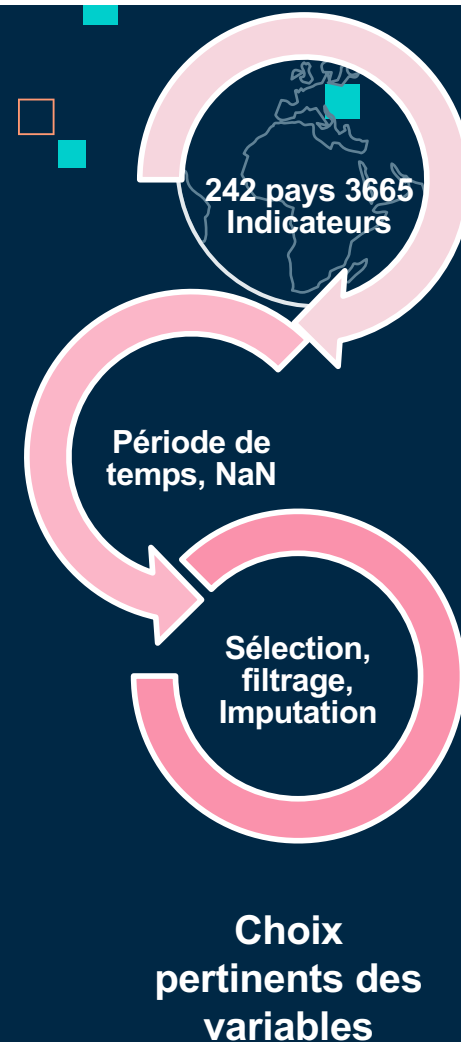
70 variables



Données  
complètes

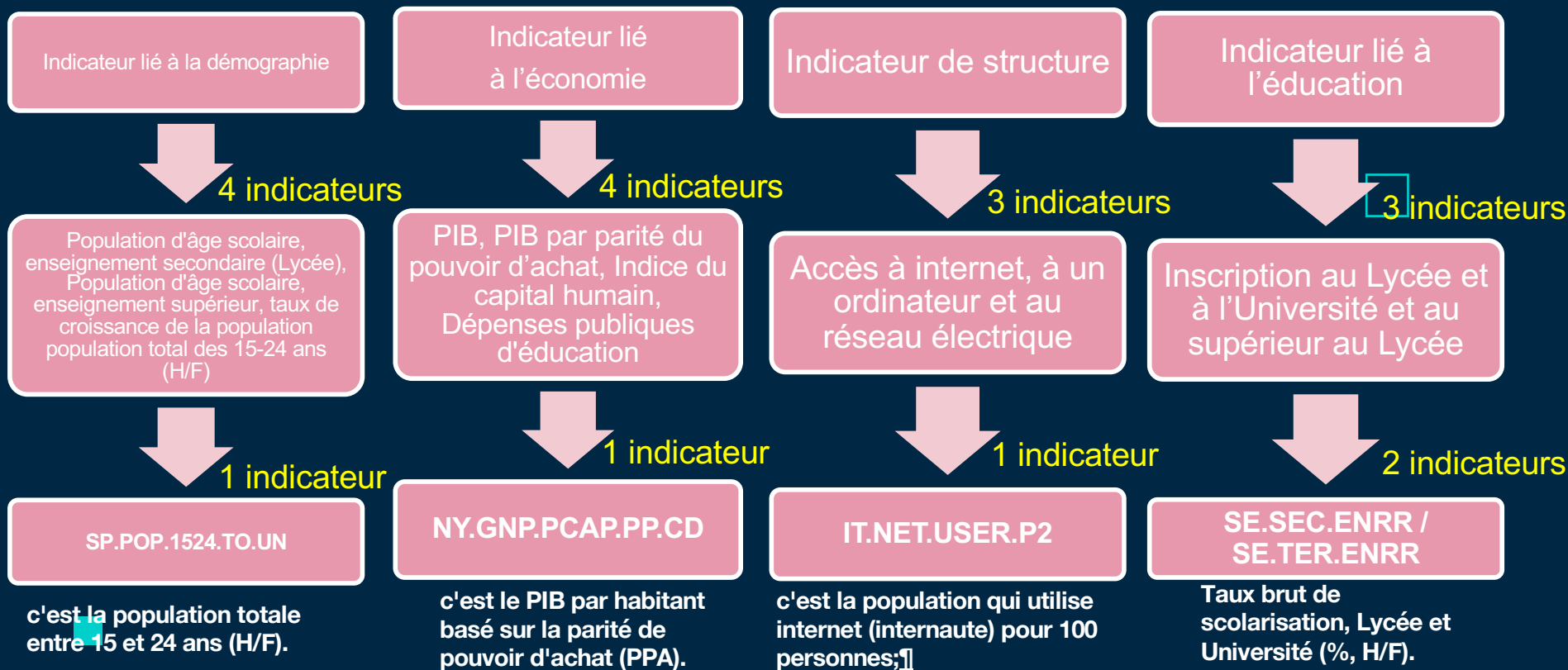
Années d'observations  
: 1970 à 2019

Années de prédictions:  
2020-2100, on a 5,8% de  
données;



# 3

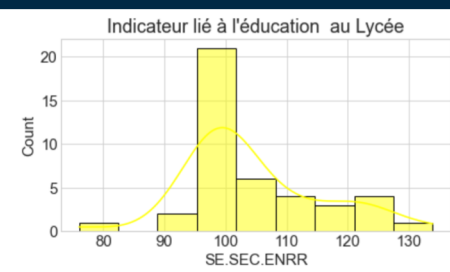
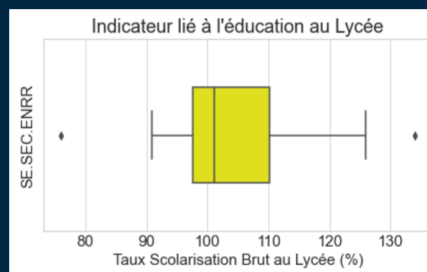
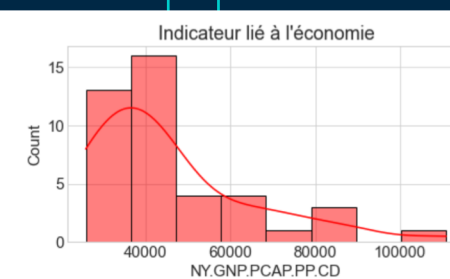
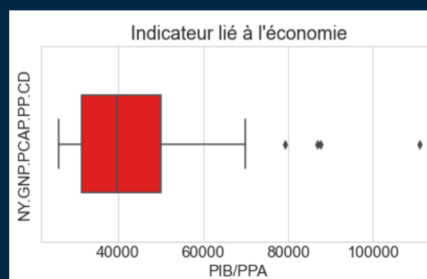
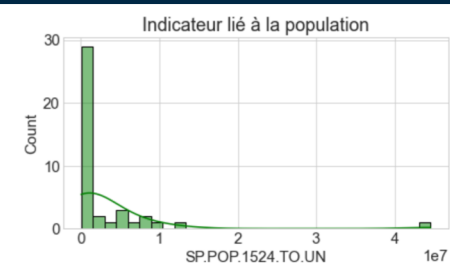
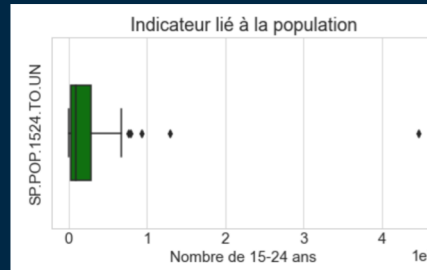
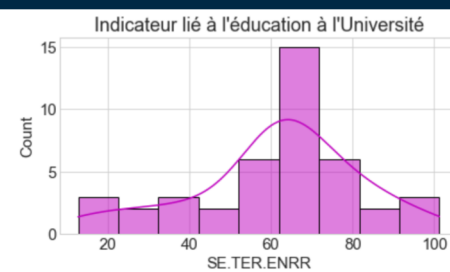
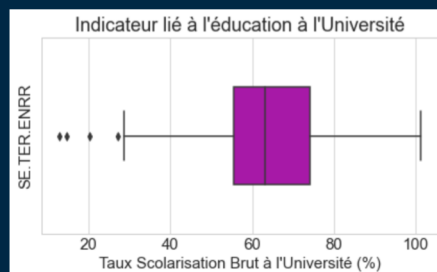
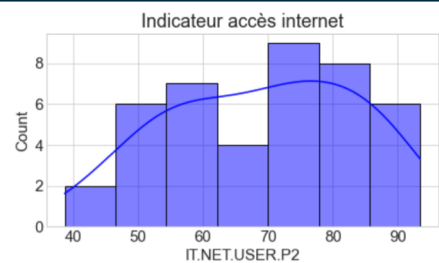
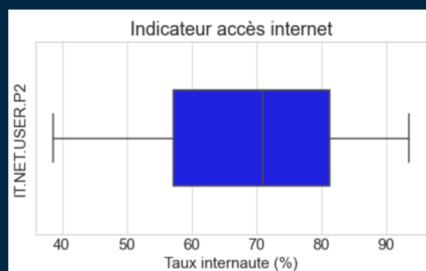
## Analyse statistique, Sélection des indicateurs, jointure de fichiers, visualisation



# 3

## Analyse statistique, Sélection des indicateurs, jointure de fichiers, visualisation

|   | La mesure statistique | IT.NET.USER.P2 | NY.GNP.PCAP.PP.CD | SE.SEC.ENRR | SE.TER.ENRR | SP.POP.1524.TO.UN |
|---|-----------------------|----------------|-------------------|-------------|-------------|-------------------|
| 0 | mean                  | 69.108504      | 4.562123e+04      | 104.355662  | 61.101317   | 3.182335e+06      |
| 1 | median                | 70.942920      | 3.980545e+04      | 101.136694  | 63.135774   | 8.846617e+05      |
| 2 | var                   | 214.868211     | 3.781945e+08      | 127.055453  | 422.180641  | 5.302659e+13      |
| 3 | std                   | 14.658384      | 1.944722e+04      | 11.271888   | 20.547035   | 7.281936e+06      |



# 3

## Analyse statistique, Sélection des indicateurs, jointure de fichiers, visualisation

Suppression des variables « Country Name », "Indicator Name" et "Unnamed:69"

Sélection des indicateurs sélectionnés (1<sup>ère</sup> sélection 8 indicateurs)

Joindre la variable "Short Name"

Filtrer sur la période de 2005 à 2015 et sur le PIB PPA

|              | Indicator Code       | 2005      | 2006      | 2007      | 2008      | 2009      | 2010      | 2011         | 2012      | 2013     | 2014          | 2015   |
|--------------|----------------------|-----------|-----------|-----------|-----------|-----------|-----------|--------------|-----------|----------|---------------|--------|
| Country Code |                      |           |           |           |           |           |           |              |           |          |               |        |
| ARB          | UIS.NERA.2           | NaN       | NaN       | NaN       | NaN       | NaN       | NaN       | NaN          | NaN       | NaN      | NaN           | NaN    |
| ARB          | UIS.NERA.2.F         | NaN       | NaN       | NaN       | NaN       | NaN       | NaN       | NaN          | NaN       | NaN      | NaN           | NaN    |
| ARB          | UIS.NERA.2.GPI       | NaN       | NaN       | NaN       | NaN       | NaN       | NaN       | NaN          | NaN       | NaN      | NaN           | NaN    |
| ARB          | UIS.NERA.2.M         | NaN       | NaN       | NaN       | NaN       | NaN       | NaN       | NaN          | NaN       | NaN      | NaN           | NaN    |
| ARB          | SE.PRIM.TENR         | 82.489487 | 82.685509 | 83.280342 | 84.011871 | 84.195961 | 85.211998 | 85.24514     | 86.101669 | 85.51194 | 85.320152     | NaN    |
| ...          | ...                  | ...       | ...       | ...       | ...       | ...       | ...       | ...          | ...       | ...      | ...           | ...    |
| ZWE          | UIS.LPAG15T24.M      | NaN       | NaN       | NaN       | NaN       | NaN       | NaN       | 166679.00000 | NaN       | NaN      | 199464.000000 | NaN Zi |
| ZWE          | SE.ADT.1524.LT.ZS    | NaN       | NaN       | NaN       | NaN       | NaN       | NaN       | 90.93070     | NaN       | NaN      | 90.428120     | NaN Zi |
| ZWE          | SE.ADT.1524.LT.FE.ZS | NaN       | NaN       | NaN       | NaN       | NaN       | NaN       | 92.12456     | NaN       | NaN      | 93.188350     | NaN Zi |
| ZWE          | SE.ADT.1524.LT.FM.ZS | NaN       | NaN       | NaN       | NaN       | NaN       | NaN       | 1.02828      | NaN       | NaN      | 1.063890      | NaN Zi |
| ZWE          | SE.ADT.1524.LT.MA.ZS | NaN       | NaN       | NaN       | NaN       | NaN       | NaN       | 89.59058     | NaN       | NaN      | 87.591860     | NaN Zi |

883265 rows x 13 columns

883265 lignes et 13 variables

|              | Indicator Code       | IT.NET.USER.P2 | NY.GNP.PCAP.PP.CD | SE.SEC.ENRR | SE.TER.ENRR | SPPOP.1524.TO.UN |
|--------------|----------------------|----------------|-------------------|-------------|-------------|------------------|
| Country Code | Short Name           |                |                   |             |             |                  |
| ARE          | United Arab Emirates | 70.899999      | 68392.727273      | 101.209136  | 63.926478   | 6.966785e+05     |
| AUS          | Australia            | 75.533792      | 38960.909091      | 133.922543  | 79.264152   | 2.893898e+06     |
| AUT          | Austria              | 74.253159      | 43181.818182      | 99.368819   | 65.388568   | 9.985681e+05     |
| BEL          | Belgium              | 73.230281      | 40486.363636      | 122.739548  | 67.396297   | 1.257487e+06     |
| BHR          | Bahrain              | 61.944466      | 38127.272727      | 95.384613   | 36.857870   | 1.260156e+05     |
| BMU          | Bermuda              | 84.501475      | 69881.428571      | 76.010968   | 28.733789   | 7.102000e+03     |
| BRN          | Brunei               | 53.825073      | 79388.181818      | 99.413604   | 20.041946   | 7.271827e+04     |
| CAN          | Canada               | 80.177273      | 40195.454545      | 103.891932  | 63.926478   | 4.436876e+06     |
| CHE          | Switzerland          | 81.728371      | 54743.636364      | 96.852188   | 51.987716   | 9.425225e+05     |
| CYP          | Cyprus               | 52.597315      | 31182.727273      | 96.791836   | 45.352923   | 1.287750e+05     |
| DEU          | Germany              | 79.690891      | 41110.909091      | 103.019220  | 64.931936   | 9.329035e+06     |
| DNK          | Denmark              | 89.456384      | 43380.909091      | 123.231315  | 78.481056   | 6.623500e+05     |
| ESP          | Spain                | 64.052236      | 31760.909091      | 123.972306  | 78.563771   | 4.677938e+06     |
| FIN          | Finland              | 84.638811      | 39092.727273      | 119.134257  | 92.366080   | 6.519545e+05     |
| FRA          | France               | 71.363118      | 36945.454545      | 111.478909  | 57.687309   | 7.650038e+06     |
| GBR          | United Kingdom       | 82.470400      | 36789.090909      | 103.614897  | 58.243002   | 7.935112e+06     |

32 pays et 5 indicateurs

|              | Indicator Code       | IT.NET.USER.P2 | NY.GNP.PCAP.PP.CD | SE.SEC.ENRR | SE.TER.ENRR | SP.POP.1524.TO.UN |
|--------------|----------------------|----------------|-------------------|-------------|-------------|-------------------|
| Country Code | Short Name           |                |                   |             |             |                   |
| ARE          | United Arab Emirates | 70.899999      | 68392.727273      | NaN         | NaN         | 6.966785e+05      |
| AUS          | Australia            | 75.533792      | 38960.909091      | 133.922543  | 79.264152   | 2.893898e+06      |
| AUT          | Austria              | 74.253159      | 43181.818182      | 99.368819   | 65.388568   | 9.985681e+05      |
| BEL          | Belgium              | 73.230281      | 40486.363636      | 122.739548  | 67.396297   | 1.257487e+06      |
| BHR          | Bahrain              | 61.944466      | 38127.272727      | 95.384613   | 36.857870   | 1.260156e+05      |
| BMU          | Bermuda              | 84.501475      | 69881.428571      | 76.010968   | 28.733789   | 7.102000e+03      |
| BRN          | Brunei               | 53.825073      | 79388.181818      | 99.413604   | 20.041946   | 7.271827e+04      |
| CAN          | Canada               | 80.177273      | 40195.454545      | 103.891932  | NaN         | 4.436876e+06      |
| CHE          | Switzerland          | 81.728371      | 54743.636364      | 96.852188   | 51.987716   | 9.425225e+05      |

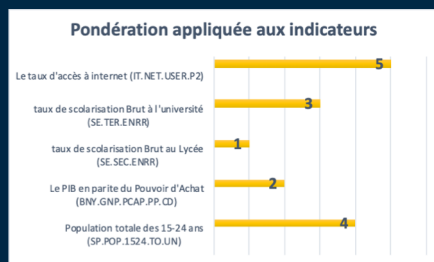
Imputation



# 4

## Score, prédiction, conclusion,

Le classement des 10 premiers pays ayant les plus forts potentiels (avec pondération) :



| Indicator Code |               | score_synt |
|----------------|---------------|------------|
| Country Code   | Short Name    |            |
| USA            | United States | 12.333625  |
| NOR            | Norway        | 9.303382   |
| FIN            | Finland       | 9.179314   |
| DNK            | Denmark       | 9.094821   |
| ISL            | Iceland       | 8.982192   |
| NLD            | Netherlands   | 8.837828   |
| SWE            | Sweden        | 8.832665   |
| DEU            | Germany       | 8.717475   |
| JPN            | Japan         | 8.713137   |
| AUS            | Australia     | 8.575843   |

Période: 2005 à 2015

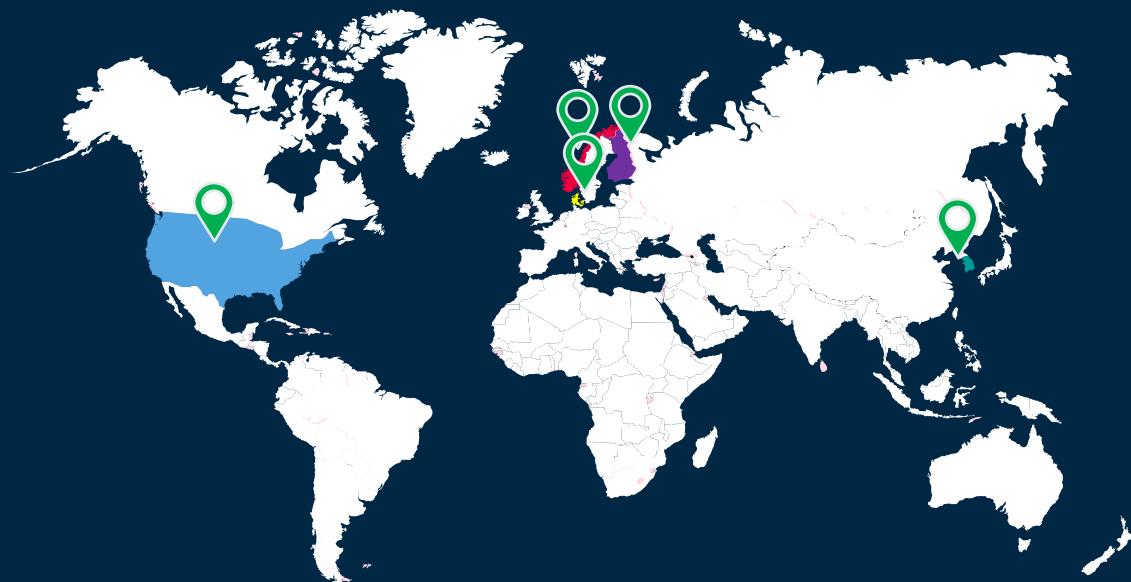
| Indicator Code |               | score_synt |
|----------------|---------------|------------|
| Country Code   | Short Name    |            |
| USA            | United States | 12.082750  |
| NOR            | Norway        | 9.089308   |
| KOR            | Korea         | 9.088403   |
| FIN            | Finland       | 8.915835   |
| DNK            | Denmark       | 8.870949   |
| ISL            | Iceland       | 8.764772   |
| NLD            | Netherlands   | 8.650086   |
| SWE            | Sweden        | 8.629053   |
| JPN            | Japan         | 8.544710   |
| DEU            | Germany       | 8.532253   |

Période: 2020 à 2100

# 4

Score, prédiction, conclusion

Conclusion:



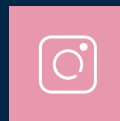
■ Etats Unis   ■ Norvège   ■ Corée du Sud  
■ Finlande   ■ Danemark

Academy doit opérer en  
priorité au:

- Etats Unis;
- Norvège;
- Corée du Sud;
- Finlande;
- Danemark;

Est-ce que vous avez  
des questions?

# THANKS



entreprise@academy.com  
+91 620 421 83  
www.academy.com