

Projet 2 : Analysez des données de systèmes éducatifs

Analyse exploratoire pour le projet d'expansion à l'international de l'entreprise academy



Abdessalem BOUZAYANI

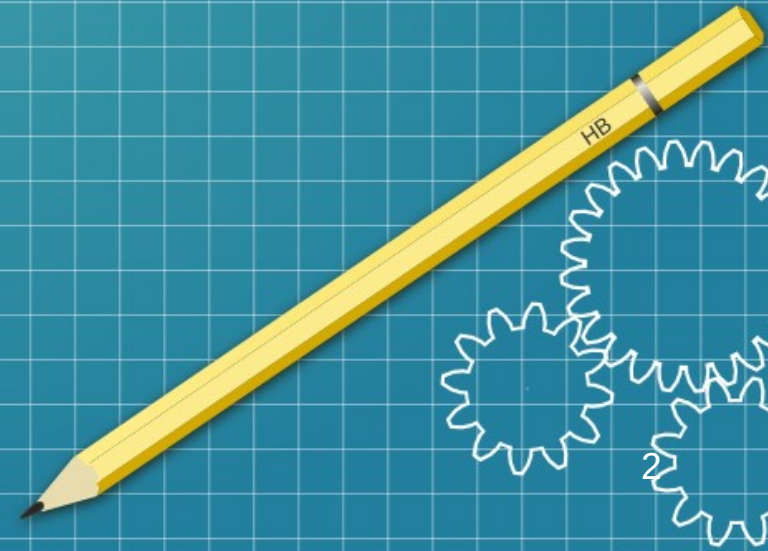


18/10/2022

Sommaire



- **Problématique**
- **Jeu de données**
- **Nettoyage (filtrage)**
- **Choix des indicateurs**
- **Définition d'un score**
- **Conclusion**



Problématique

- Projet d'expansion à l'international de l'entreprise academy
- A partir d'un jeu de données de la banque mondiale, on doit répondre aux questions :
 - Valider la qualité du jeu de données
 - Description des informations contenues dans le jeu de données
 - Sélection des informations pertinentes
 - Calcul des ordres de grandeurs des indicateurs statistiques classiques
 - Déterminer les pays avec un fort potentiel de clients
 - Déterminer l'évolution du potentiel de clients
 - Déterminer les pays en priorité

Jeu de données

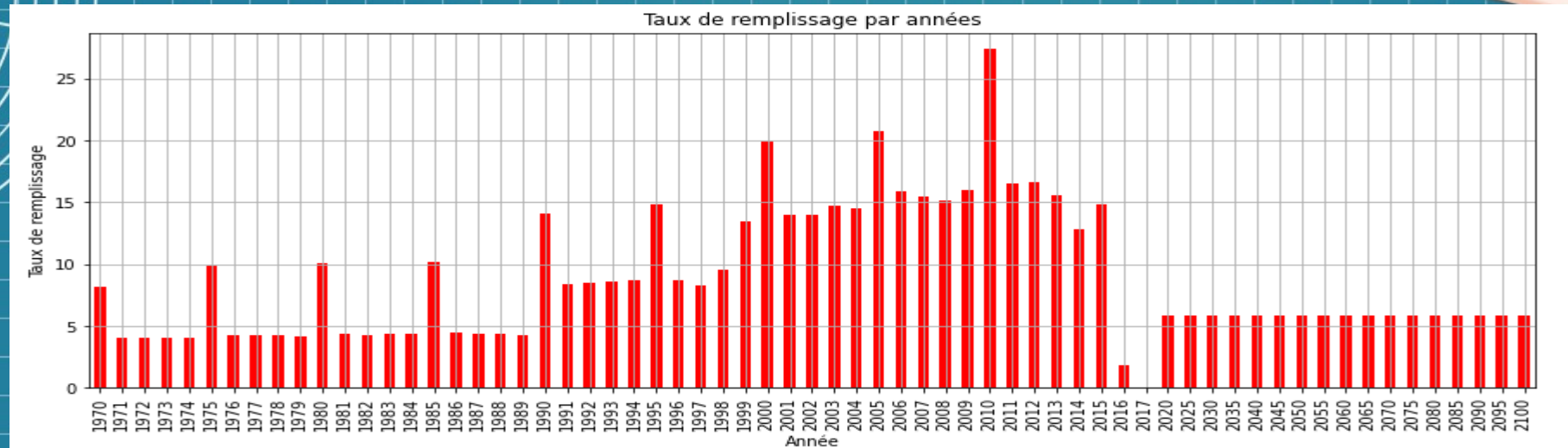
- **EdStatsCountry-Series** : Quelques indicateurs par pays
- **EdStatsFootNote** : Evolution des indicateurs par pays
- **EdStatsCountry** : Informations géographiques et économiques des pays
- **EdStatsData** : Valeurs indicateurs par pays et par année
- **EdStatsSeries** : Description de chaque indicateur



BANQUE MONDIALE

	Colonnes	Lignes	Nbre donnees manquantes	% donnees manquantes	Nbre doublons
EdStatsCountry	32	241	2354	30.52	0
EdStatsCountry-Series	4	613	613	25.00	0
EdStatsData	70	886930	53455179	86.10	0
EdStatsFootNote	5	643638	643638	20.00	0
EdStatsSeries	21	3665	55203	71.72	0

Nettoyage : sélection des années



Données après 2016 : par plage de 5 ans avec un taux de remplissage faible

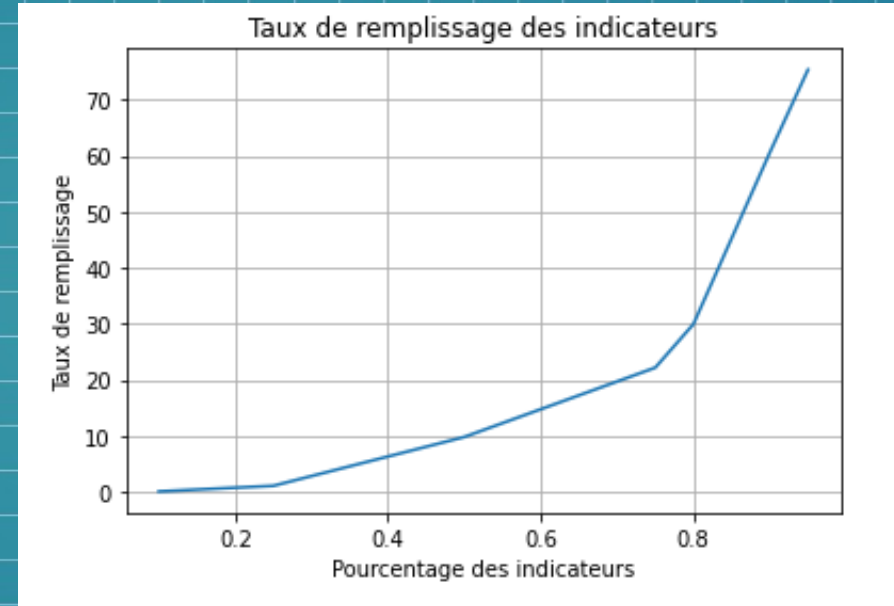
Les données avant 2005 : taux de remplissage faible

Conclusion : garder les données de 2005 à 2015

Nettoyage : remplissage des indicateurs

- 3665 indicateurs
- 20% ont un taux de remplissage supérieur à 30%
- 10% ont un taux de remplissage supérieur à 60%
- Choisir un seuil de remplissage des indicateurs : 50 %

327 indicateurs

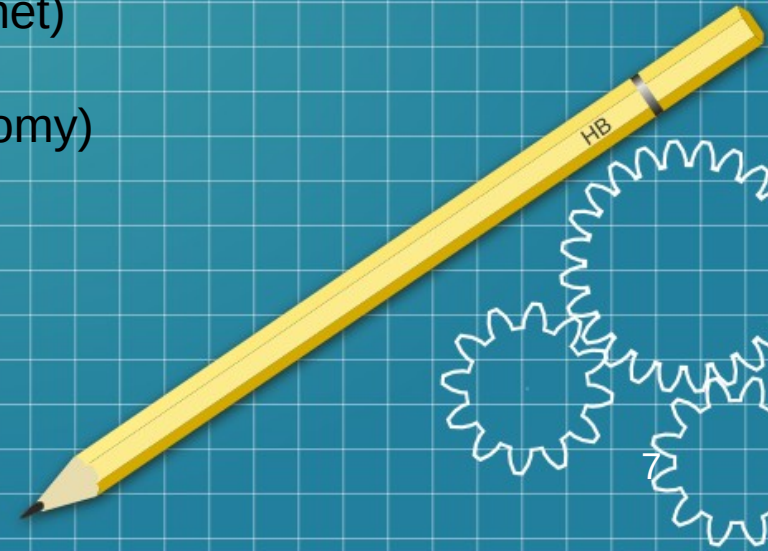


Nettoyage : recherche des indicateurs par mots-clés



- 327 indicateurs
- Activité de l'entreprise : formations en ligne pour les lycéens et les étudiants.
- Choisir la bonne tranche d'age (mots-clés : school, education, 15, 20 et 24)
- Choisir une bonne connexion Internet (mots-clés : Internet)
- Choisir un bon pouvoir d'achat (mots-clés : gdp et economy)
- Choisir une population élevée (mots-clés: demography)

170 indicateurs

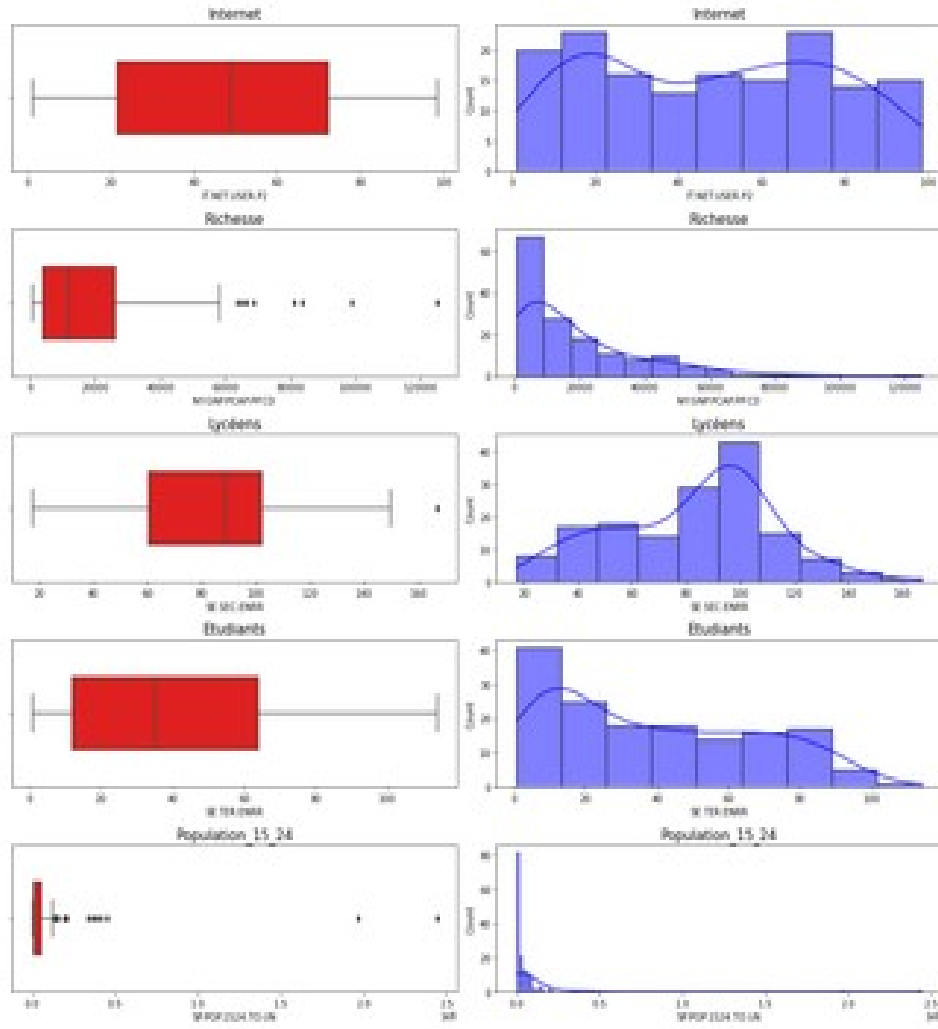


Nettoyage : indicateurs retenus

5 Indicateurs :

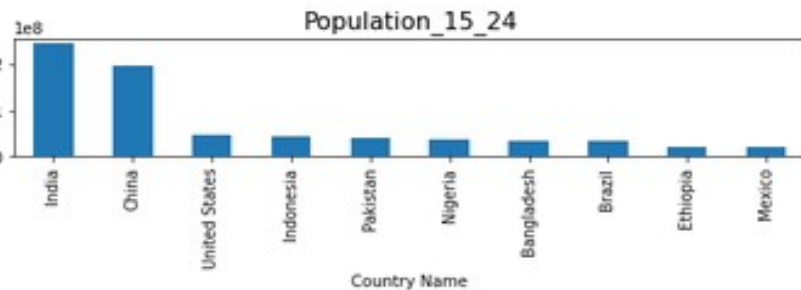
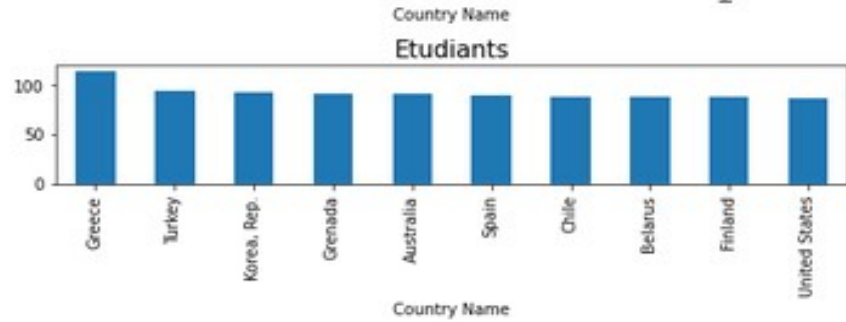
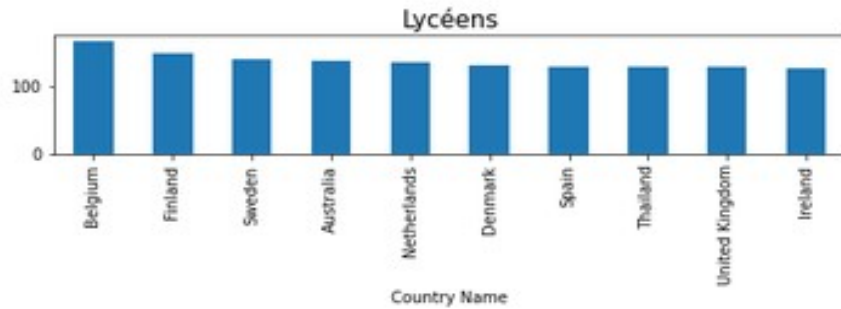
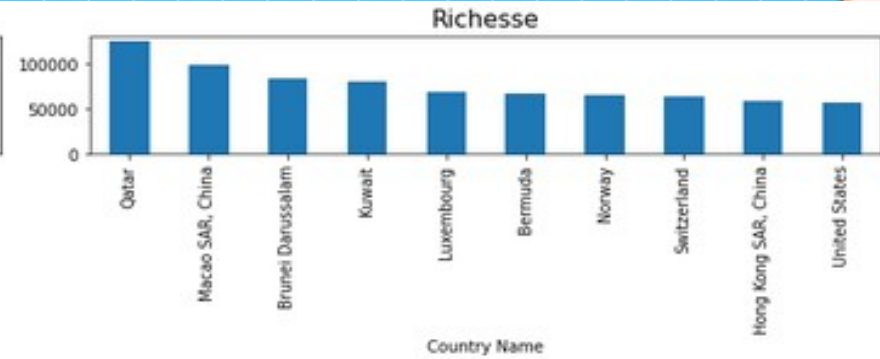
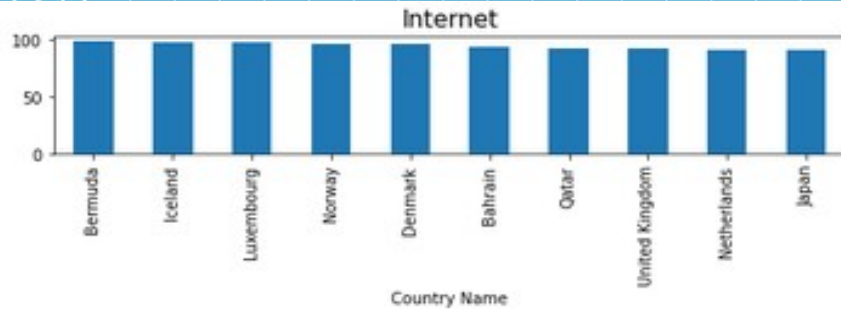
- GNI per capita, PPP (current international \$) : NY.GNP.PCAP.PP.CD
- Internet users (per 100 people) : IT.NET.USER.P2
- Population, ages 15-24, total : SP.POP.1524.TO.UN
- Gross enrolment ratio, secondary, both sexes (%) : SE.SEC.ENRR
- Gross enrolment ratio, tertiary, both sexes (%) : SE.TER.ENRR

Observations sur les indicateurs retenus



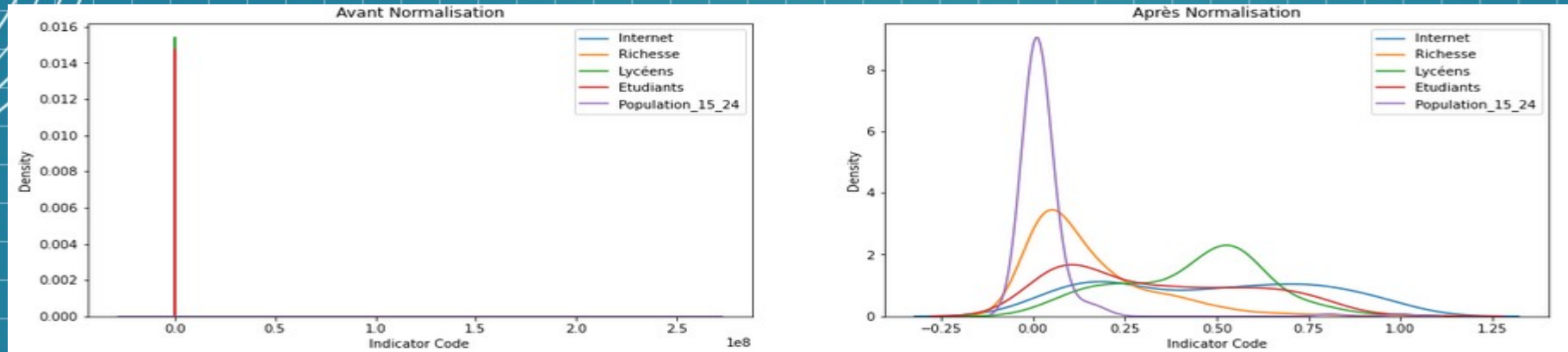
	Mesure	Internet	Richesse	Lycéens	Etudiants	Population_15_24
0	moyenne	47.60	1.953142e+04	82.75	38.85	7.449137e+06
1	médiane	48.88	1.190000e+04	88.67	35.08	1.493722e+06
2	variance	816.00	4.490851e+08	863.05	825.70	6.692897e+14
3	écart type	28.57	2.119163e+04	29.38	28.74	2.587063e+07
4	skewness	0.05	1.920000e+00	-0.19	0.44	7.690000e+00
5	kurtosis	-1.29	4.730000e+00	-0.33	-1.06	6.379000e+01

Meilleurs pays selon chaque indicateur



Choix d'un score

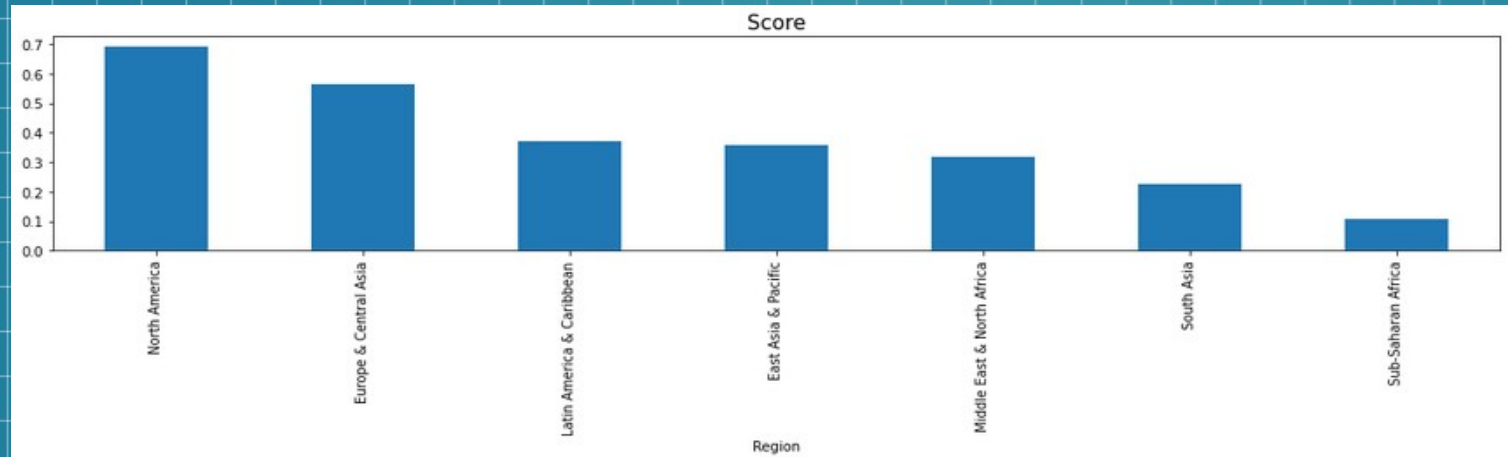
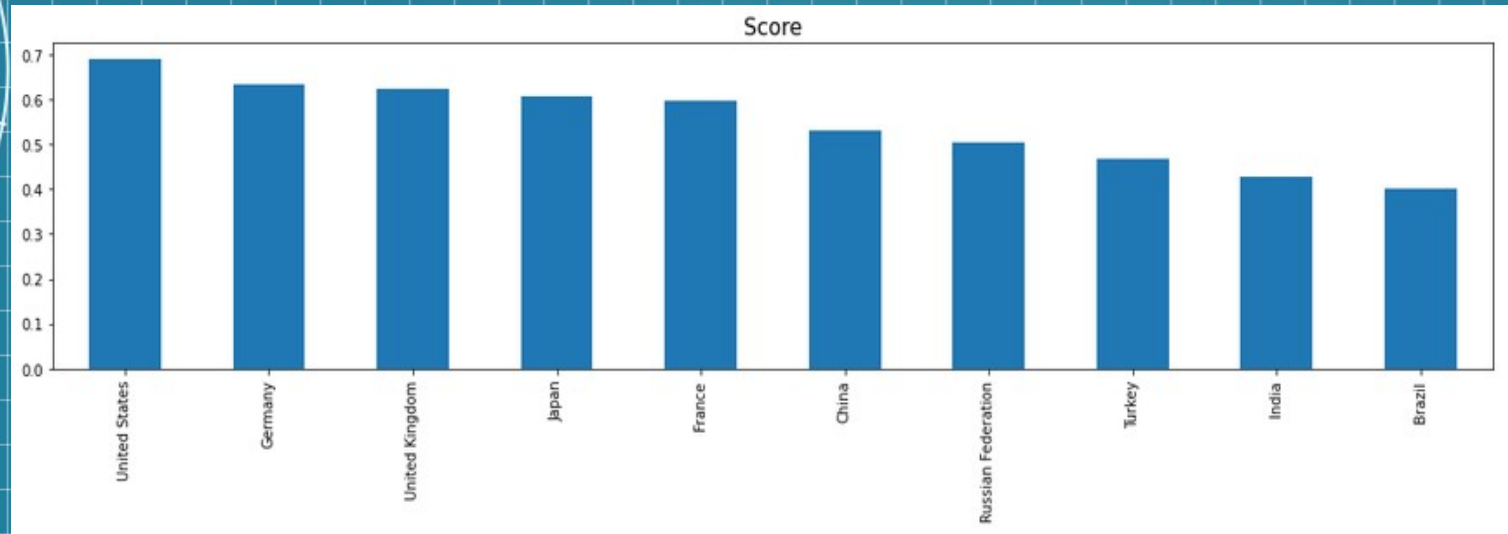
- Normalisation



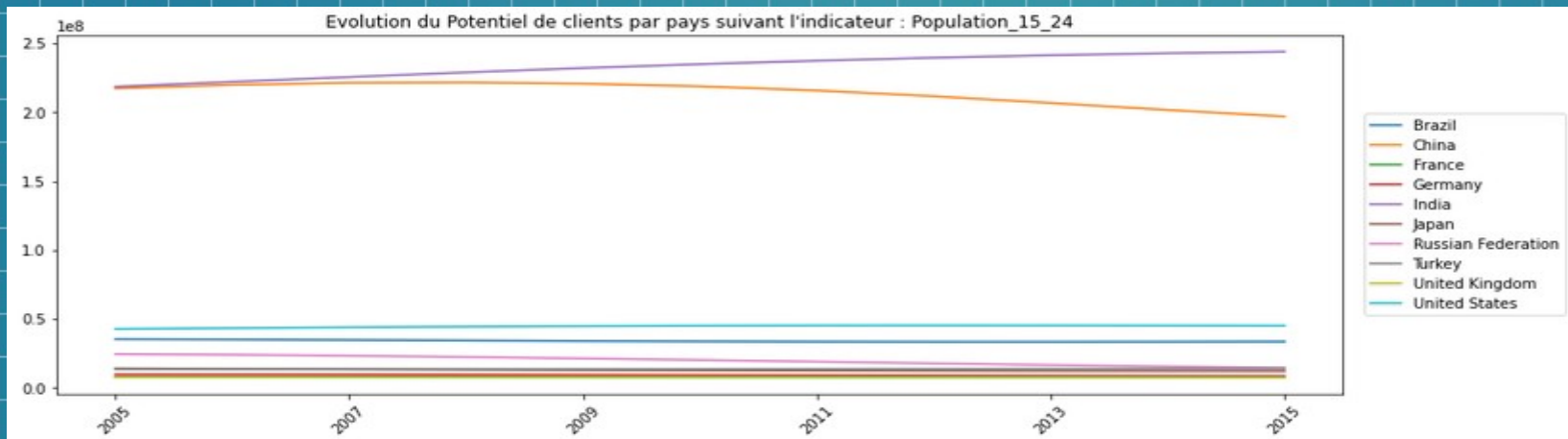
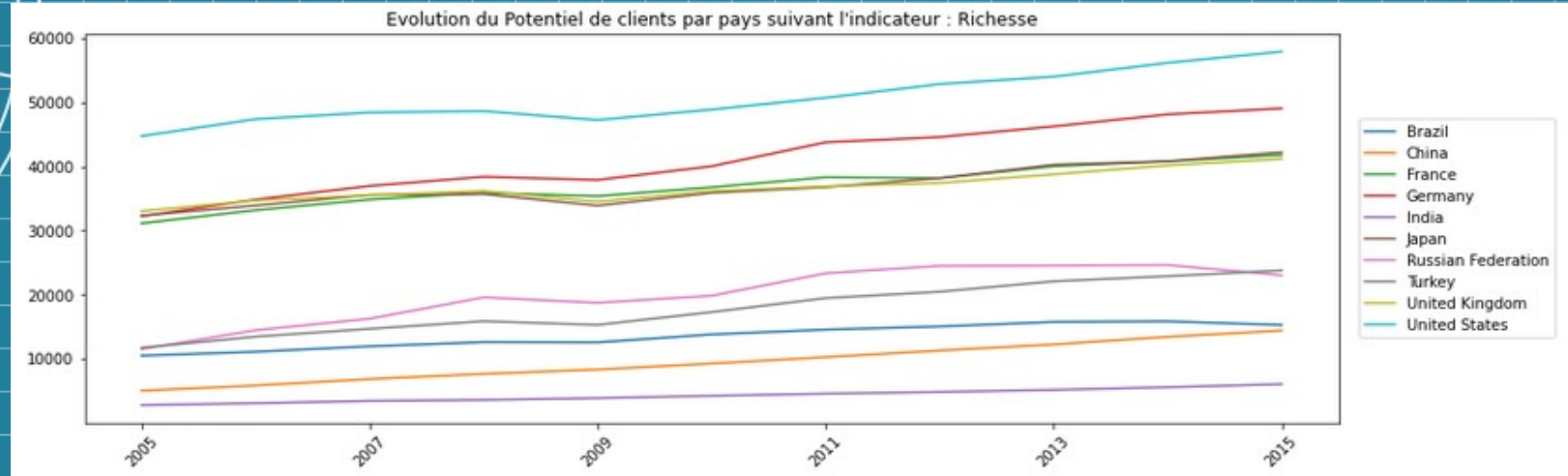
- Formule score

$$\text{Score} = (2 * \text{Richesse} + 2 * \text{Internet} + 2 * \text{Population} + \text{Lycéens} + \text{Etudiants}) / 8$$

Fort potentiel de clients par pays et par région



Evolution du potentiel de clients



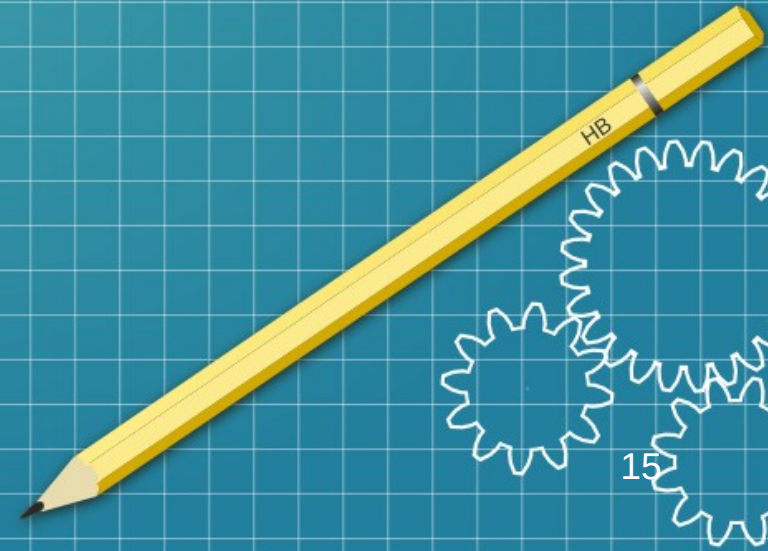
Conclusion

- 1- USA
- 2- Allemagne
- 3- Royaume Uni
- 4- Japon
- 5- France



Perspectives

- Explorer d'autres indicateurs
- Analyse multivariée
- Imputation des données manquantes
- Données plus récentes



Boîte à outils

Environnement



Bibliothèques



Visualisation



Merci de votre attention



Contact : bouzaieni@gmail.com

