



Projet 2 : ANALYSEZ DES DONNEES DE SYSTEMES EDUCATIFS

Parcours Data sciences | ABBOUD Marwa | 19 aout 2020



Projet de manipulations de données à partir des données de la banque mondiale.

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

Encadrant :

Bertrand Beaufils

Projet : Formation online

Pour les étudiants en
niveau lycée et université

Dans quels
pays ?

Sur **quels**
critères est basé
le choix de pays ?

Meilleurs pays

Pays	Score finale	Rang
United States	1267	1
Germany	1258	2
China	1232	3
France	1232	3
Russia	1217	4

- Afin de choisir les meilleurs pays pour installer la formation on doit manipuler et nettoyer notre jeu de données.
 - Cela se fait en quatre parties :
 - **Première partie:** Chargement de données
 - **Deuxième partie:** Nettoyage de données
 - **Troisième partie:** Visualisation de données
 - **Quatrième partie :** Classement de pays

Description du jeu de données

- Importer les données par la Librairie **Pandas**
- Afficher les informations contenues dans le jeu de données :

- Supprimer les colonnes non-intéressants :

```
del df['Unnamed: 69']  
del df['Country Code']  
del df['Indicator Code']
```

- Les premières lignes du jeu de données :

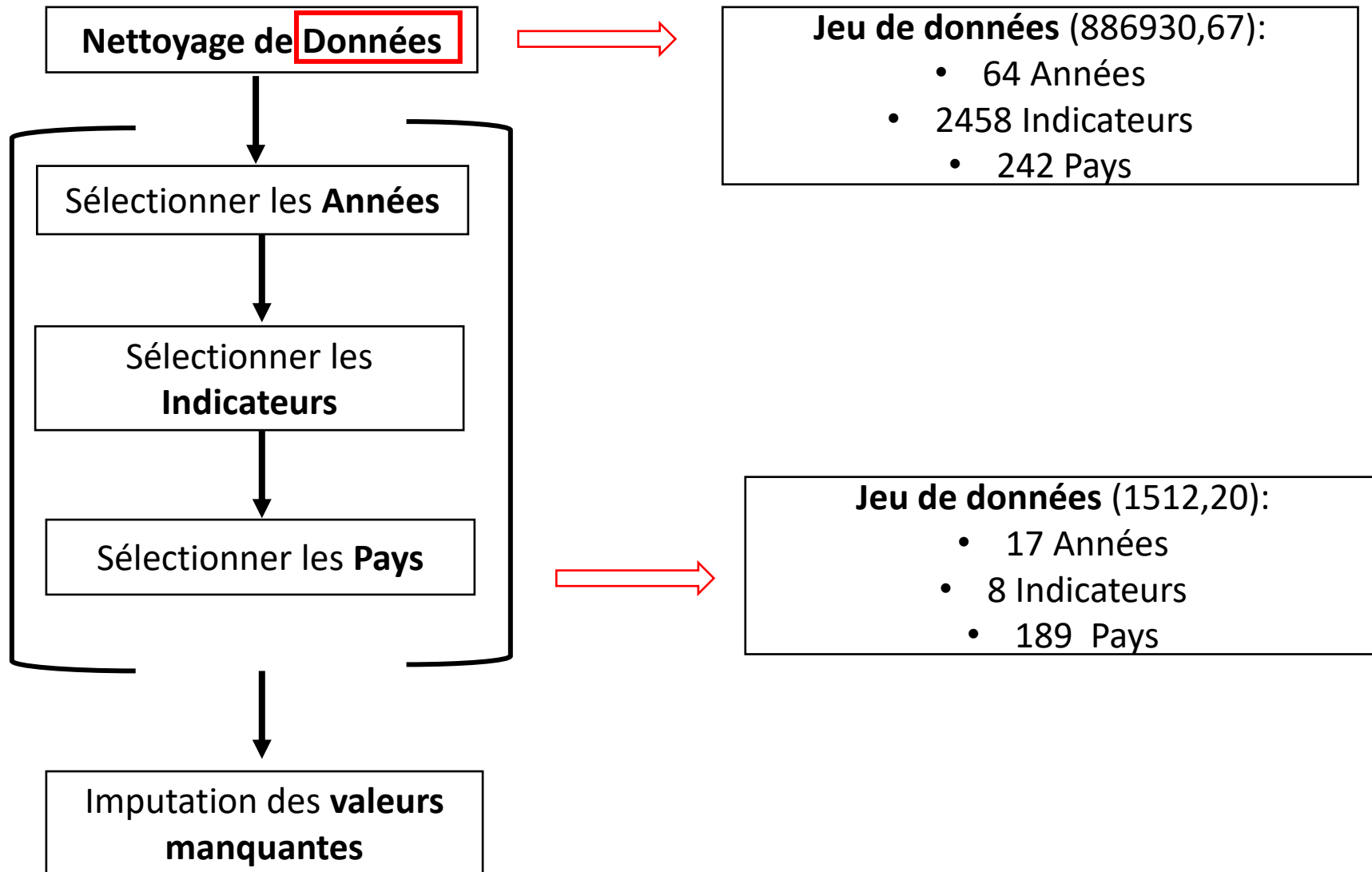
	Country Name	Indicator Name	1970	1971	1972
0	Arab World	Adjusted net enrolment rate, lower secondary, ...	NaN	NaN	NaN
1	Arab World	Adjusted net enrolment rate, lower secondary, ...	NaN	NaN	NaN
2	Arab World	Adjusted net enrolment rate, lower secondary, ...	NaN	NaN	NaN
3	Arab World	Adjusted net enrolment rate, lower secondary, ...	NaN	NaN	NaN

Valeurs manquantes

Dimensions: (886930, 67)

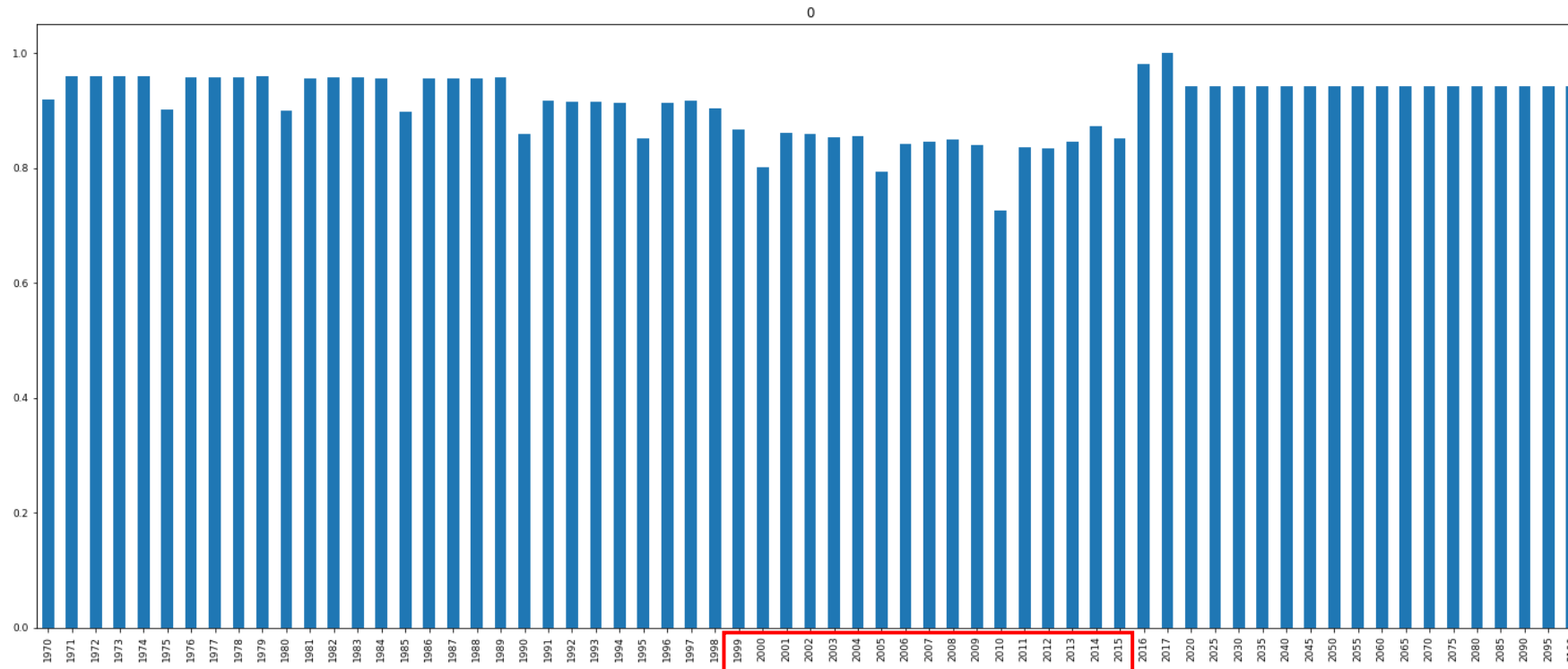
Traiter un ensemble de données avec 800 mille de lignes

Nettoyage de données



Nettoyage de données

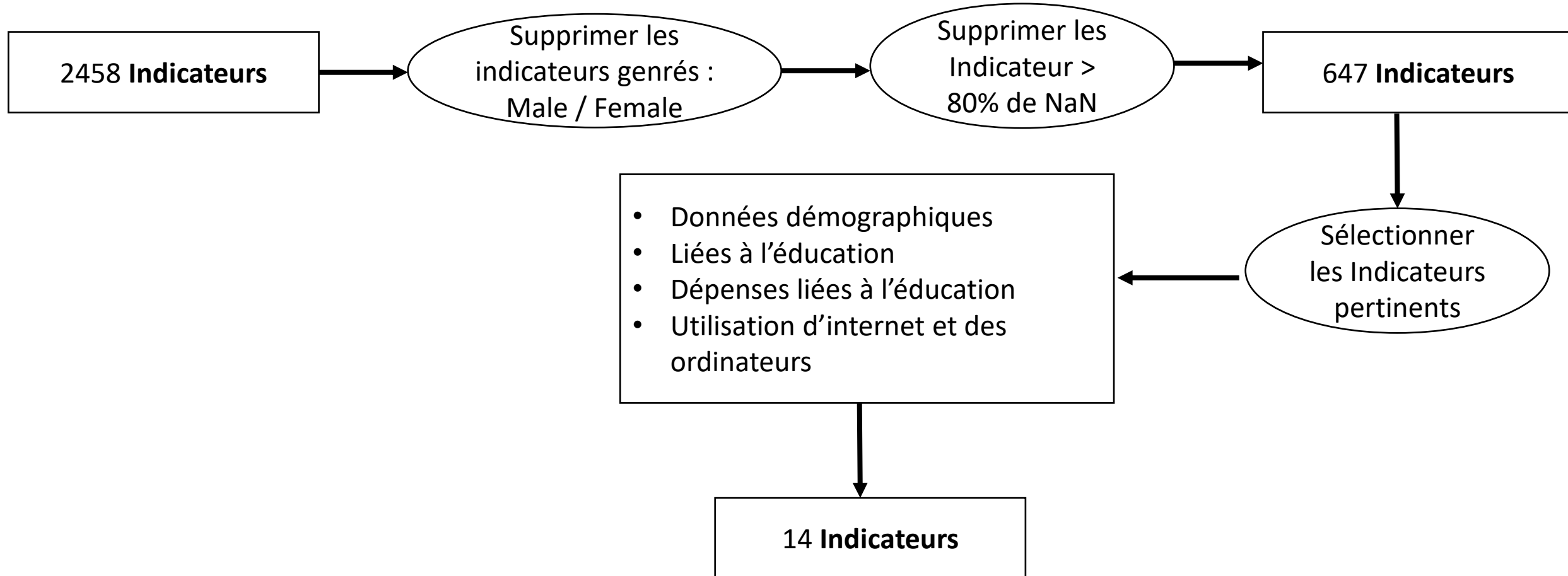
➤ Selon les Années :



Graphique représentant le taux de valeurs manquantes pour chaque année

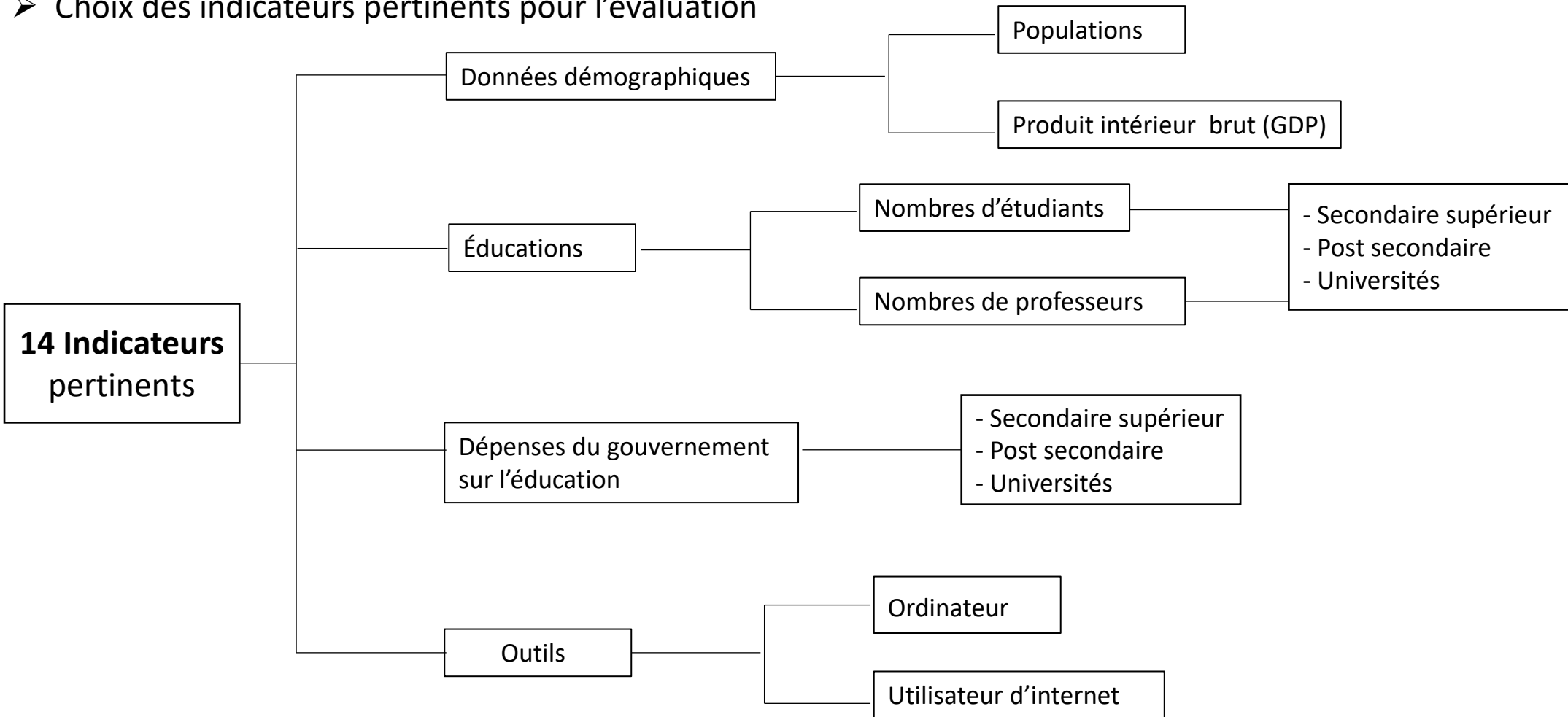
Nettoyage de données

➤ Selon les indicateurs :



Nettoyage de données

➤ Choix des indicateurs pertinents pour l'évaluation



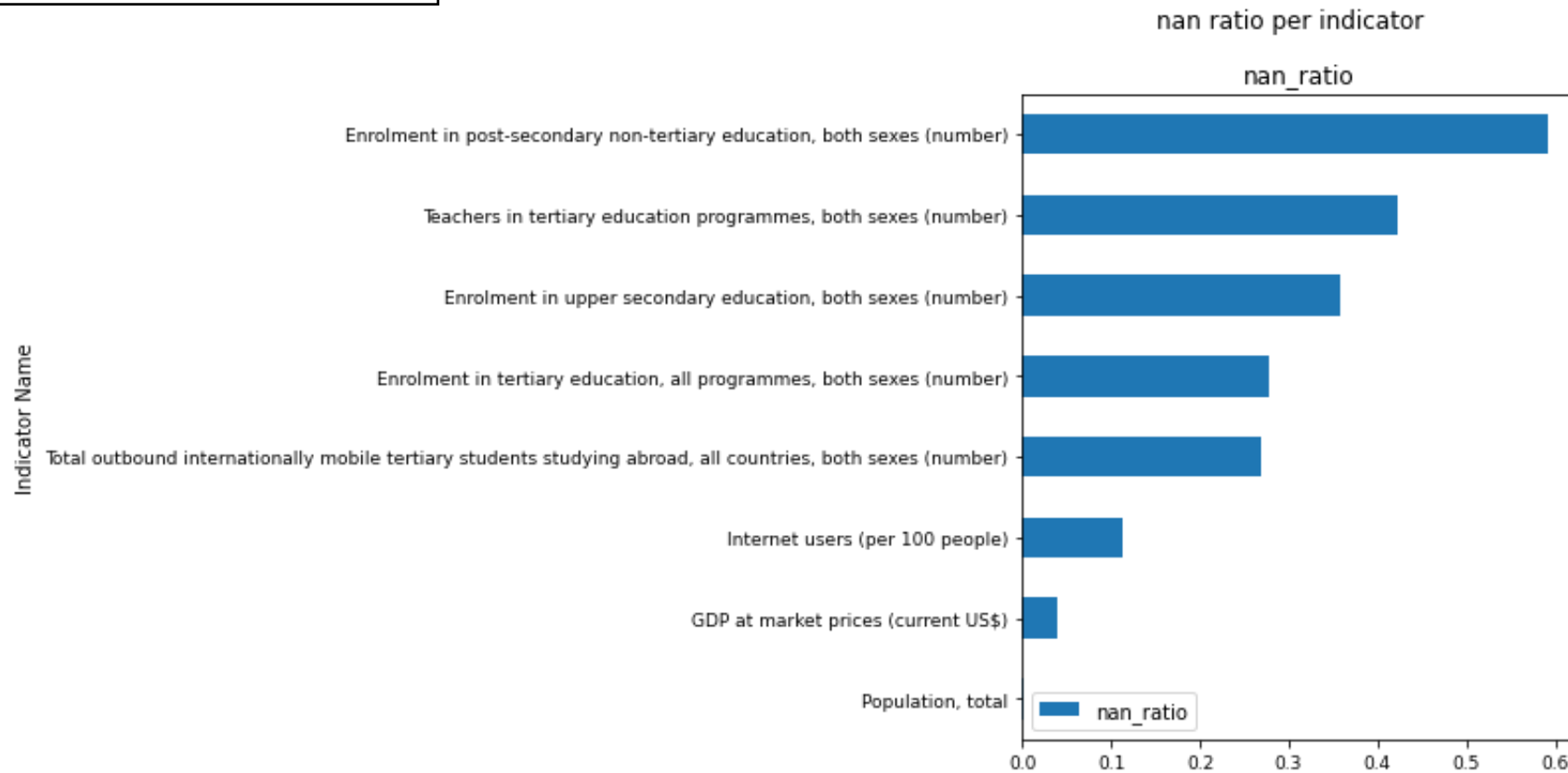
Nettoyage de données

➤ Selon les indicateurs :

Liste finale des **Indicateurs**

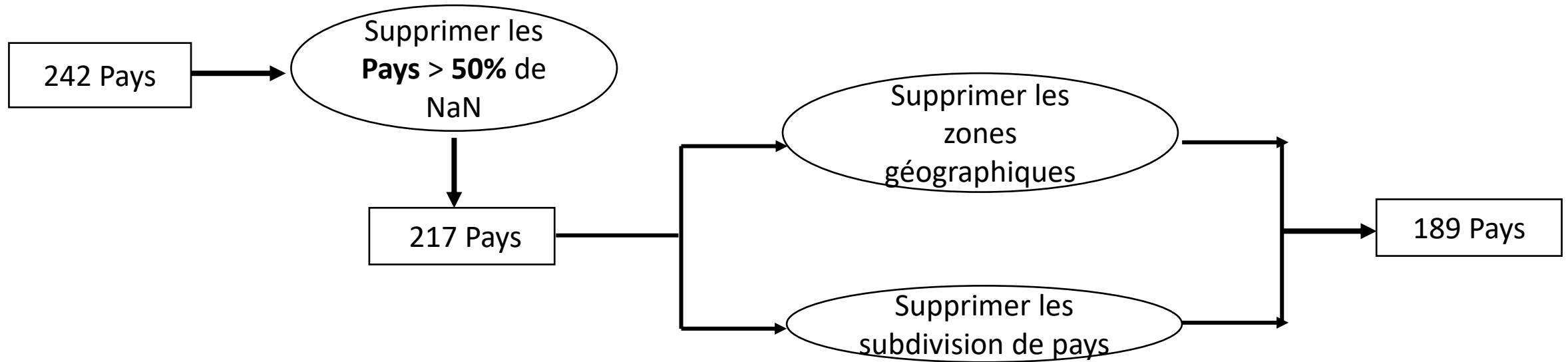


8 Indicateurs

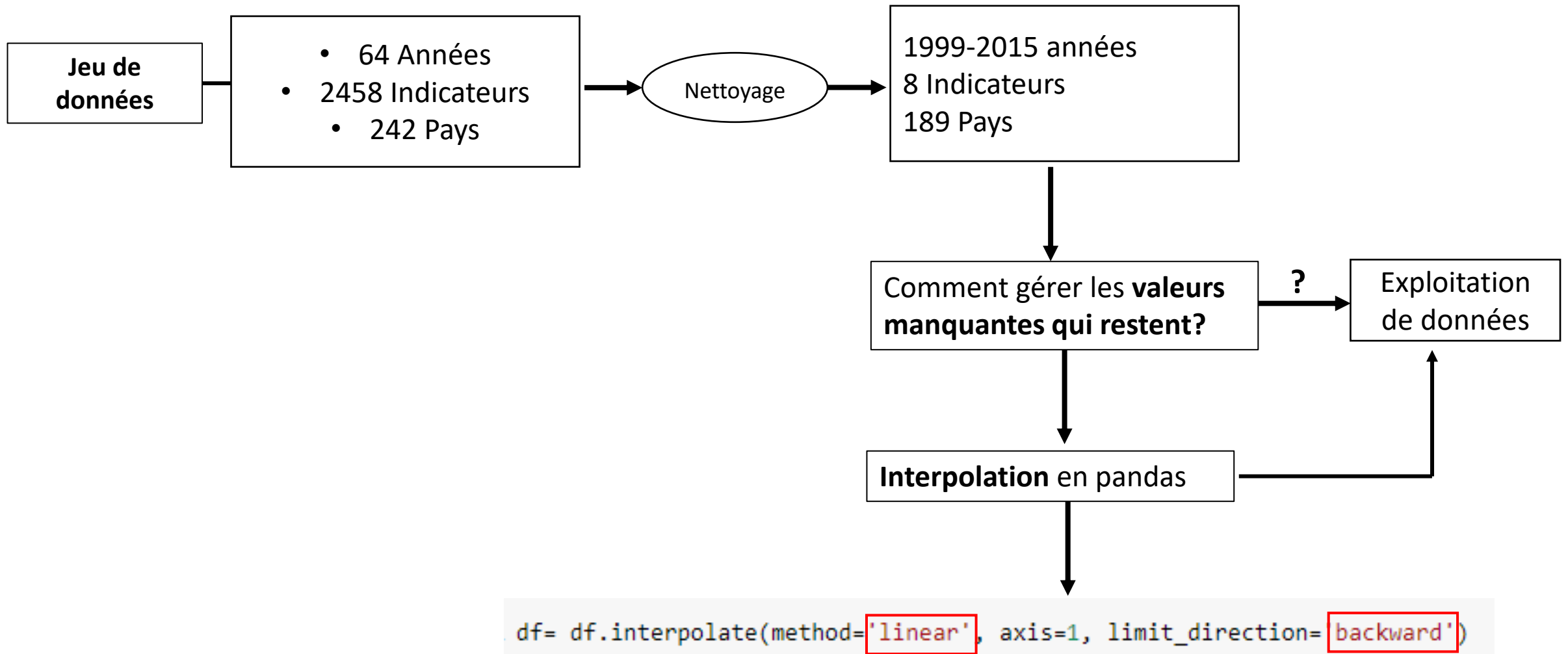


Nettoyage de données

➤ Selon les pays :

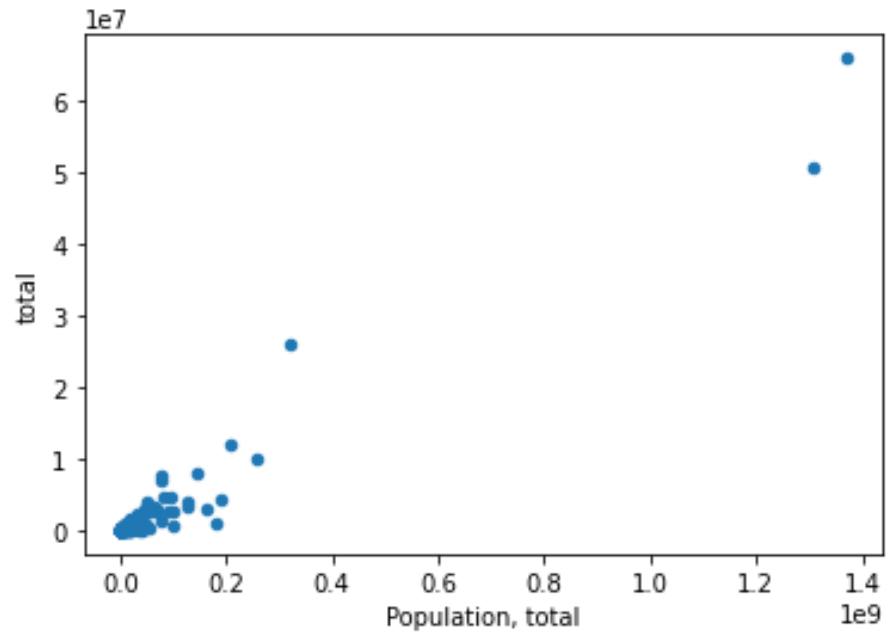


Nettoyage de données



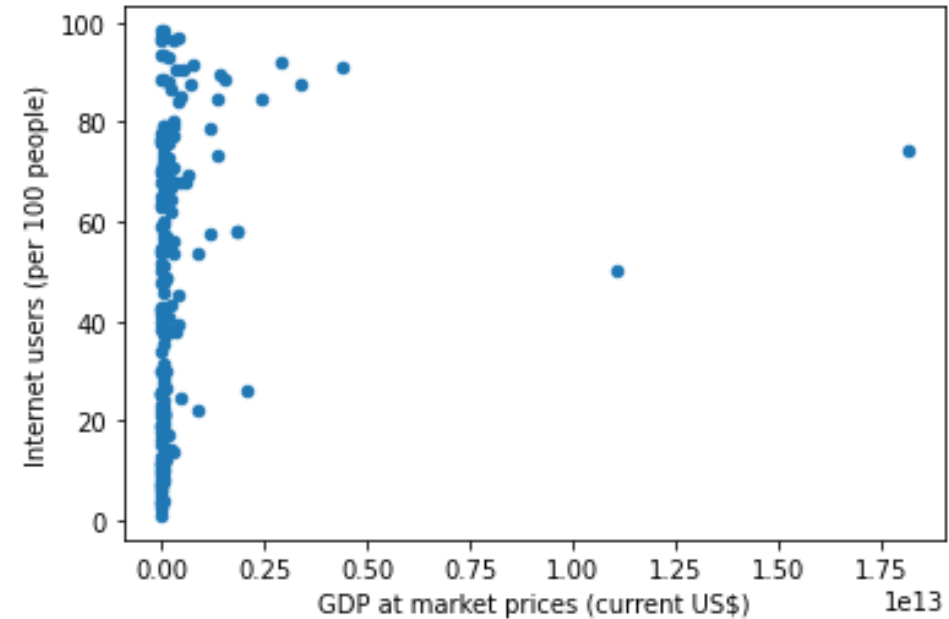
Visualisation de données

Graphe représente la distribution d'étudiants selon la population



- Nombre de population augmente → le nombre total d'étudiants augmente

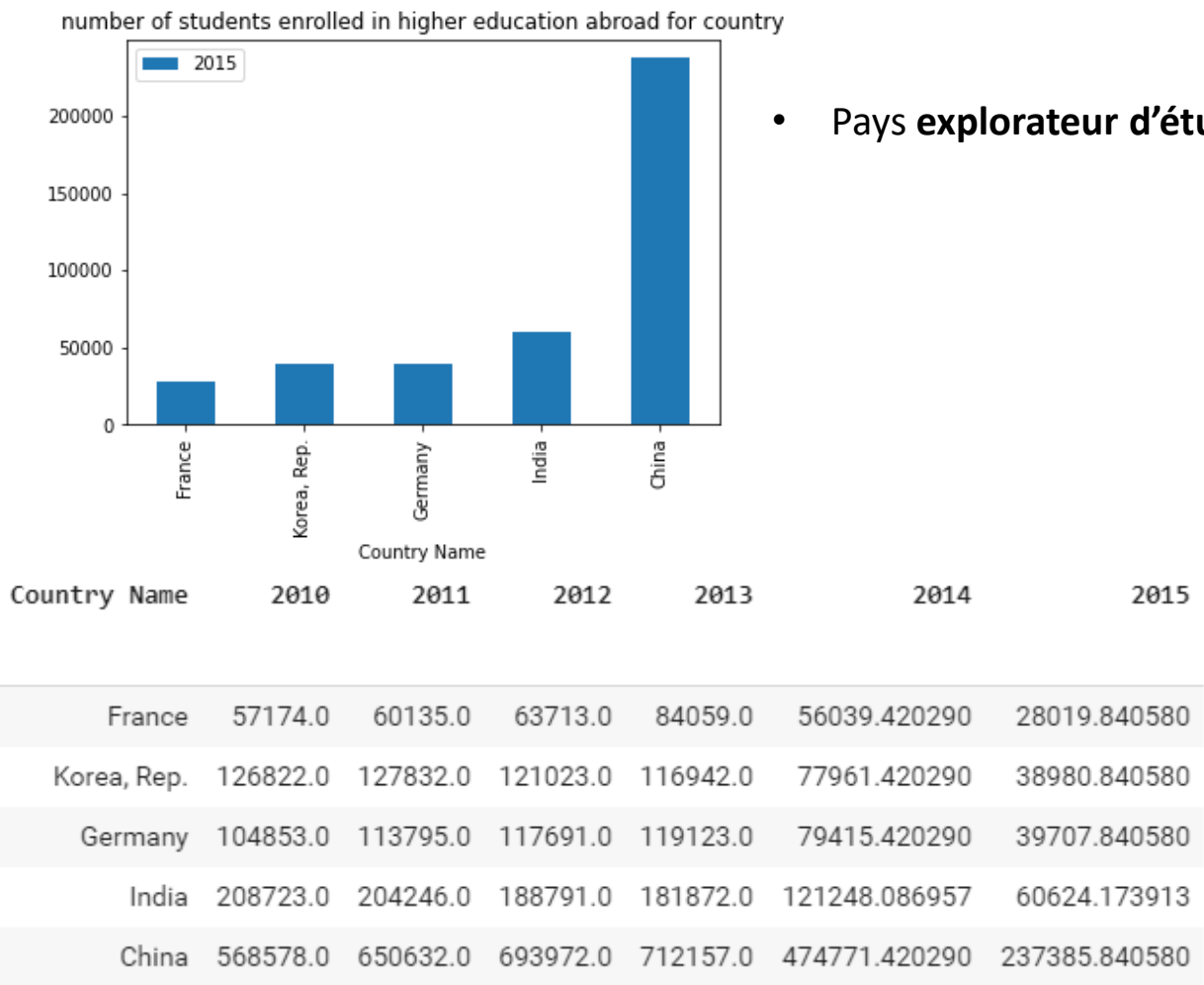
Graphe représente la distribution d'étudiants selon la population



- Pas de corrélations

Visualisation de données

Graphe représente les pays qui ont le plus grand nombre **d'étudiants scolarisés à l'étranger** pour l'année 2015.



- Pays **explorateur d'étudiants** à l'étranger :

- 1- Chine
- 2- India
- 3- Germany
- 4- Korea
- 5- France



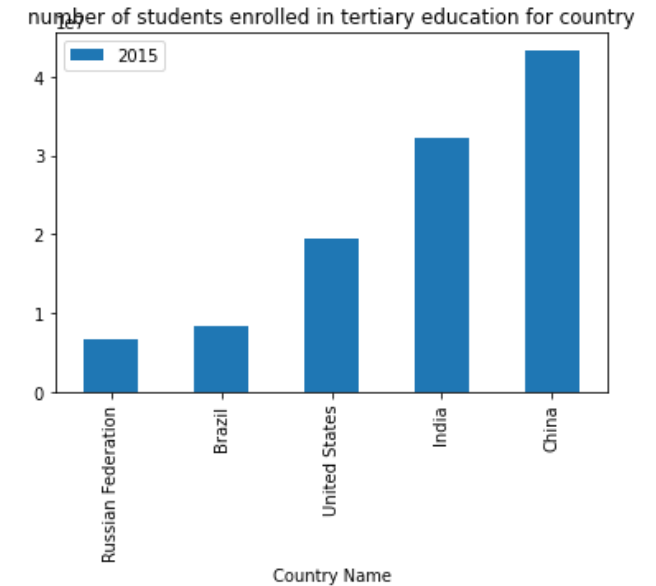
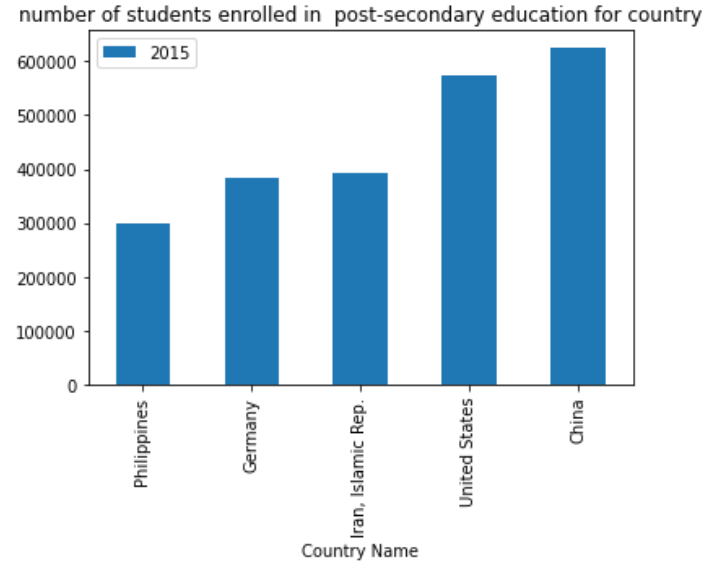
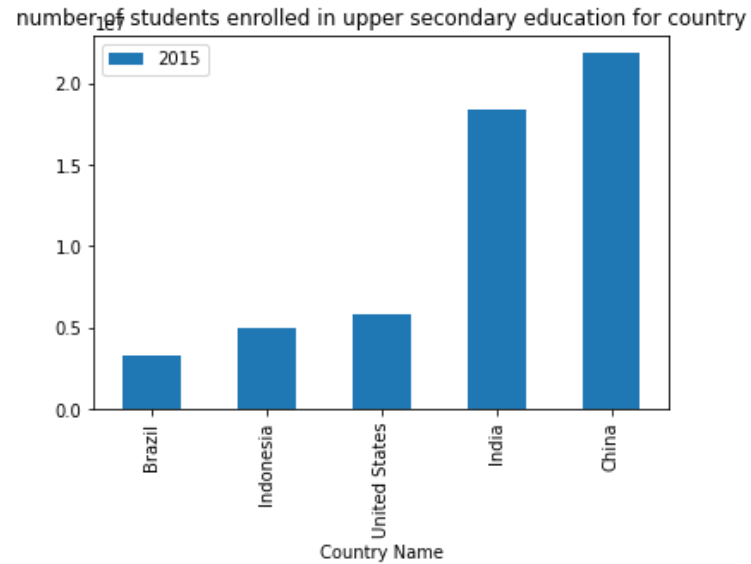
Liste de clients probable

- Le nombre d'étudiants à l'étranger a diminué au cours des 6 dernières années.

Tableau représente l'évolution des pays qui ont le plus grand nombre d'étudiants scolarisés à l'étranger pour les derniers années.

Visualisation de données

Graphes représentent les pays qui ont le plus grand nombre d'étudiants pour l'année 2015



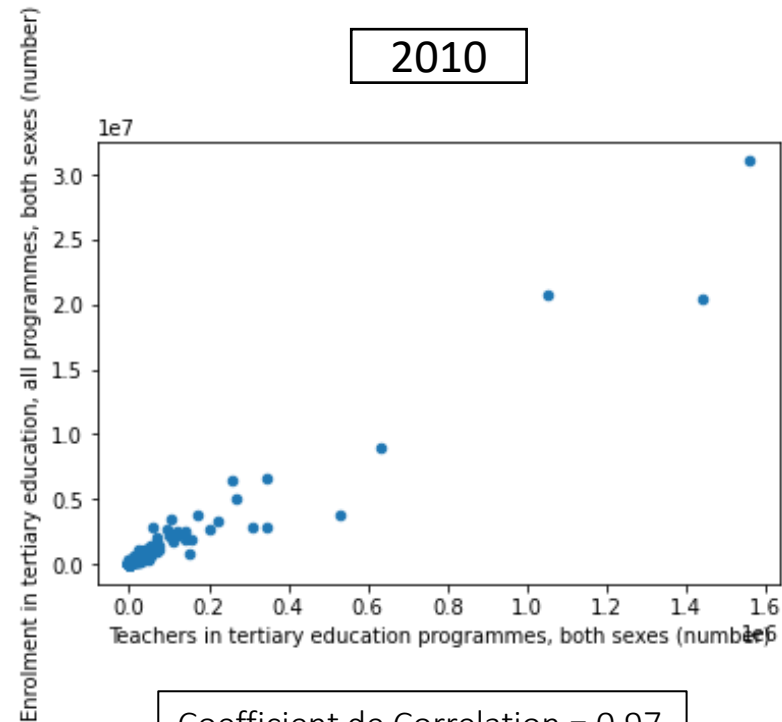
- Très important nombre de lycéens et d'étudiants dans ces pays:

1- Chine
2- India
3- United States
4- Brazil
5- Russia

Liste de clients probable

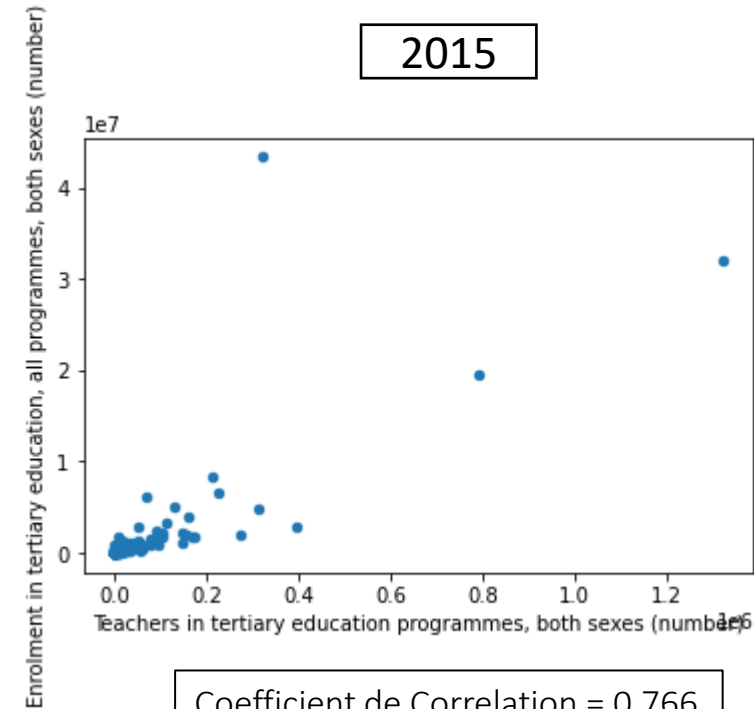
Visualisation de données

La corrélation entre le nombre d'étudiants à l'université et leur professeur pour l'ensemble de pays



Coefficient de Correlation = 0.97

Ratio étudiants par professeur: 22



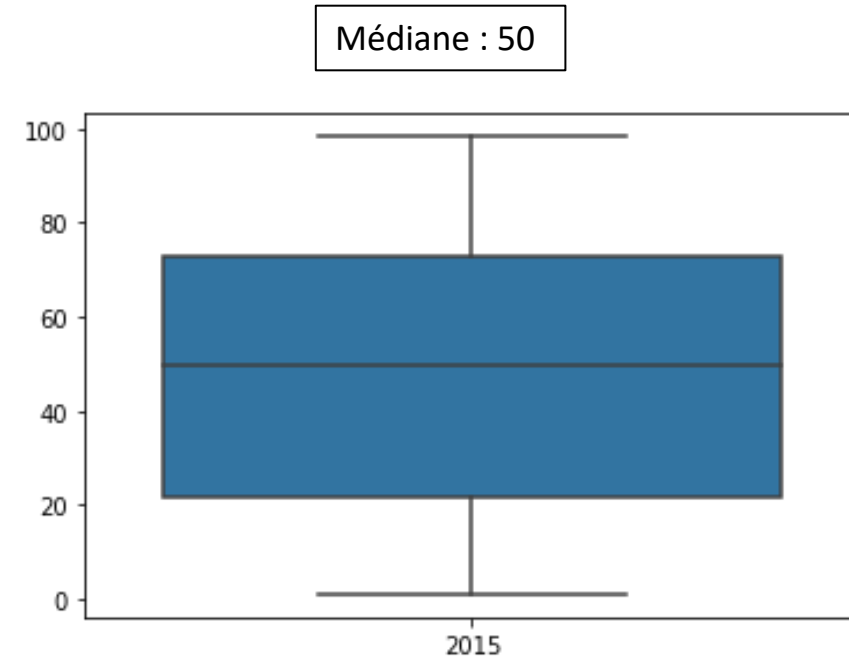
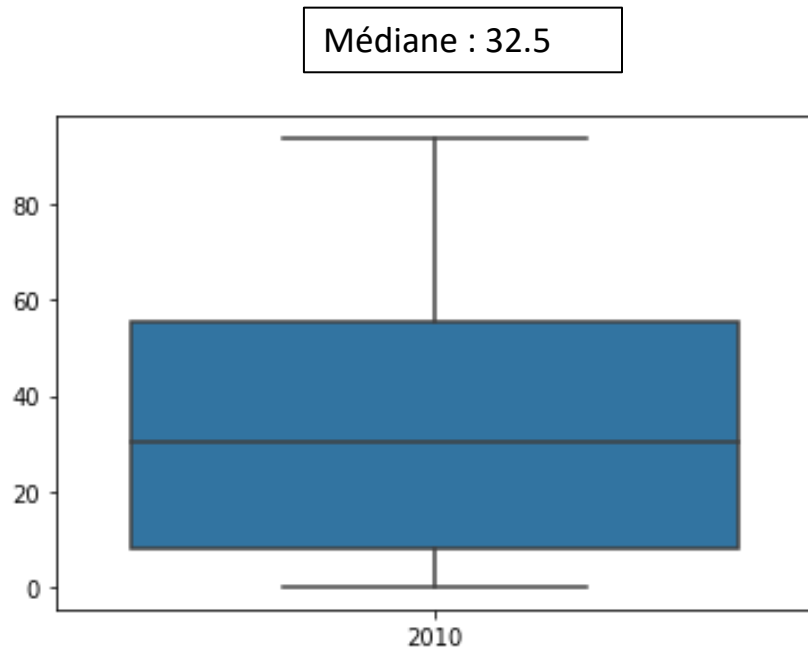
Coefficient de Correlation = 0.766

Ratio étudiants par professeur: 28

- Le nombre d'étudiants en enseignement supérieur augmente d'une manière plus rapide que celui du professeur

Visualisation de données

Graphes représentent la distribution d'utilisateurs d'internet pour l'ensemble des pays pour les années **2010** et **2015**



- Nombre d'utilisateurs d'internet pour la plupart des pays a augmenté → on peut estimer que le nombre d'étudiants qui ont accès à l'internet va augmenter

Calcul du score

➤ Le **Score** pour chaque pays est calculé selon l'ensemble des indicateurs retenus de la façon suivante:

$$\text{Score} = \sum_{i=1}^8 \text{index}_i$$

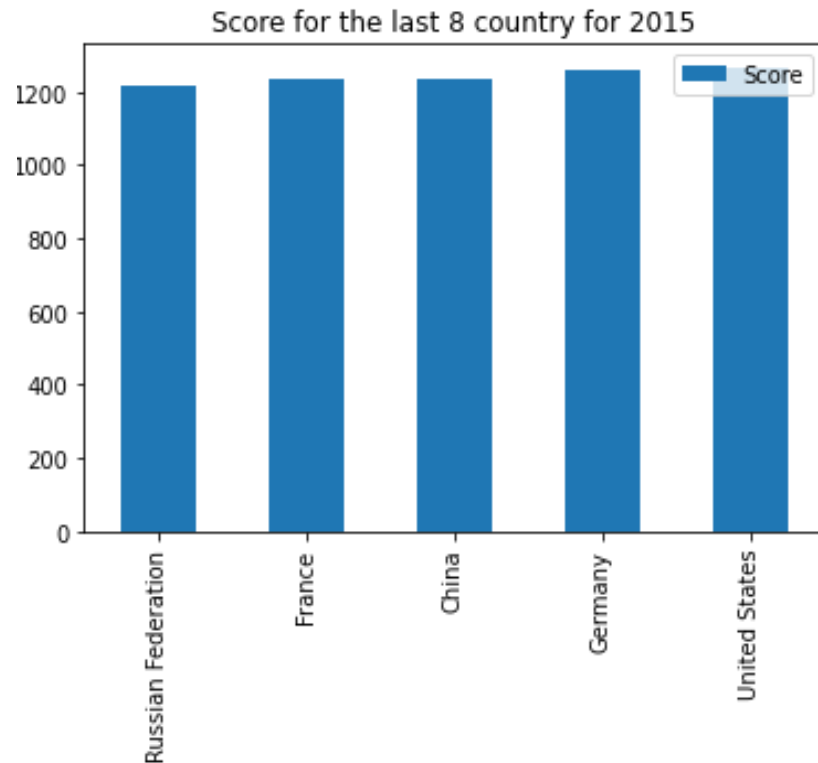
- Pour chaque Indicateur on attribue un index selon son **rang**
- Ordre d'index dépend de l'influence de l'indicateur sur le score :
 - positive : ordre croissant
 - négative : ordre décroissant

Country	Total outbound internationally mobile tertiary students studying abroad, all countries, both sexes (number)	index1	Enrolment in upper secondary education, both sexes (number)	index2	Internet users (per 100 people)	index3	Enrolment in tertiary education, all programmes, both sexes (number)	index4	Enrolment in post-secondary non-tertiary education, both sexes (number)	index5	Population, total	index6	Teachers in tertiary education programmes, both sexes (number)	index7	GDP at market prices (current US\$)	index8	Score
Sudan	0.913043	0	7.539880e+05	148	16.7	74	5.227740e+05	145	0.913043	50	3.438596e+07	156	1.319900e+04	74	6.563411e+10	125	772
Puerto Rico	0.913043	1	1.345040e+05	88	45.3	128	2.493720e+05	122	70013.000000	168	3.721525e+06	67	1.619000e+04	62	9.838127e+10	131	762
Marshall Islands	245.000000	2	2.609217e+03	15	7.0	40	1.244600e+03	17	244.092072	64	5.242500e+04	4	6.400000e+01	166	1.649691e+08	1	309

➤ A condition que : Les pays qui ont le score le plus élevés sont les meilleurs pays pour notre projet

Classement des pays par le score

- Classement pour les pays qui ont le score le plus élevé en 2015 :



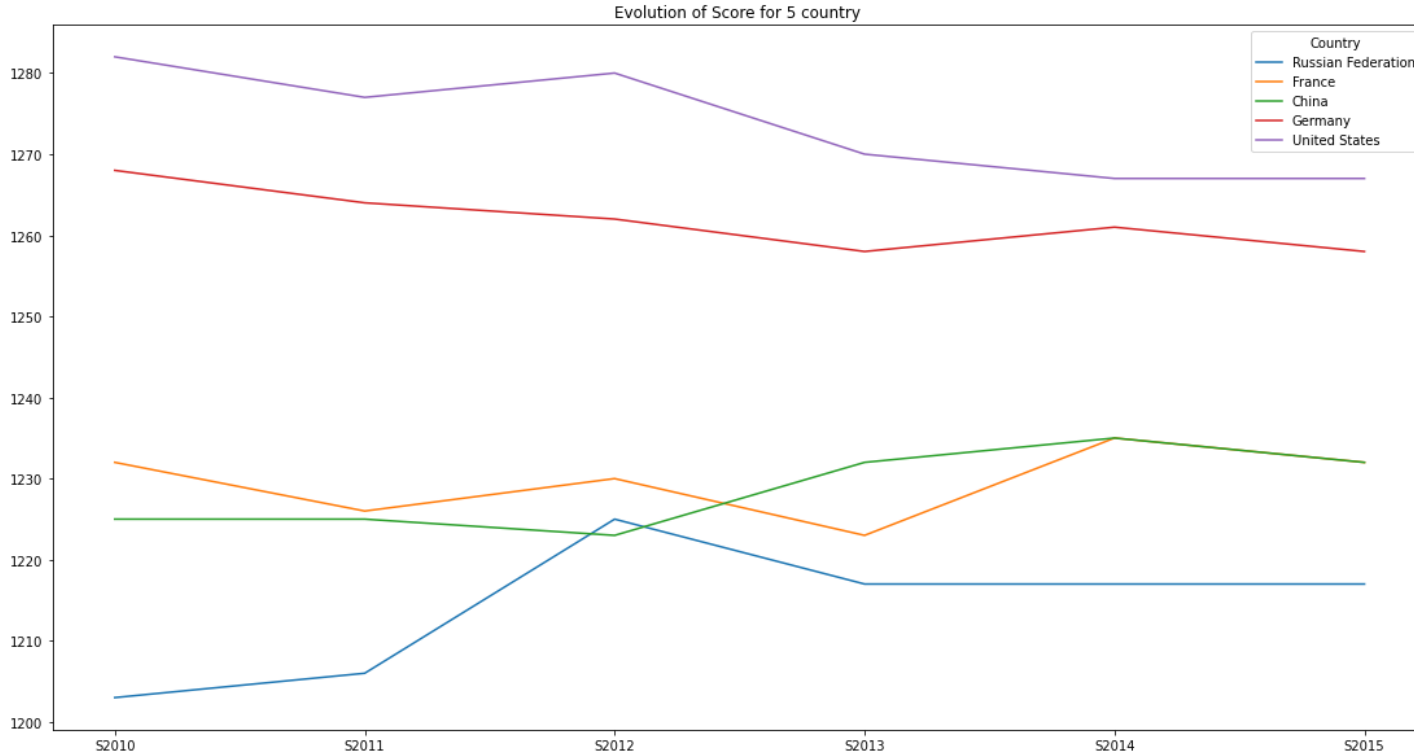
- A partir de score finale, les pays sont classés comme suit:

Pays	Score finale	Rang
United States	1267	1
Germany	1258	2
China	1232	3
France	1232	3
Russia	1217	4

- Les pays du tableau ci-dessous seraient de bons candidats pour installer la formation

Classement des pays

➤ Evolution de Classement pour les dernières années pour les pays qui ont le plus haut score :



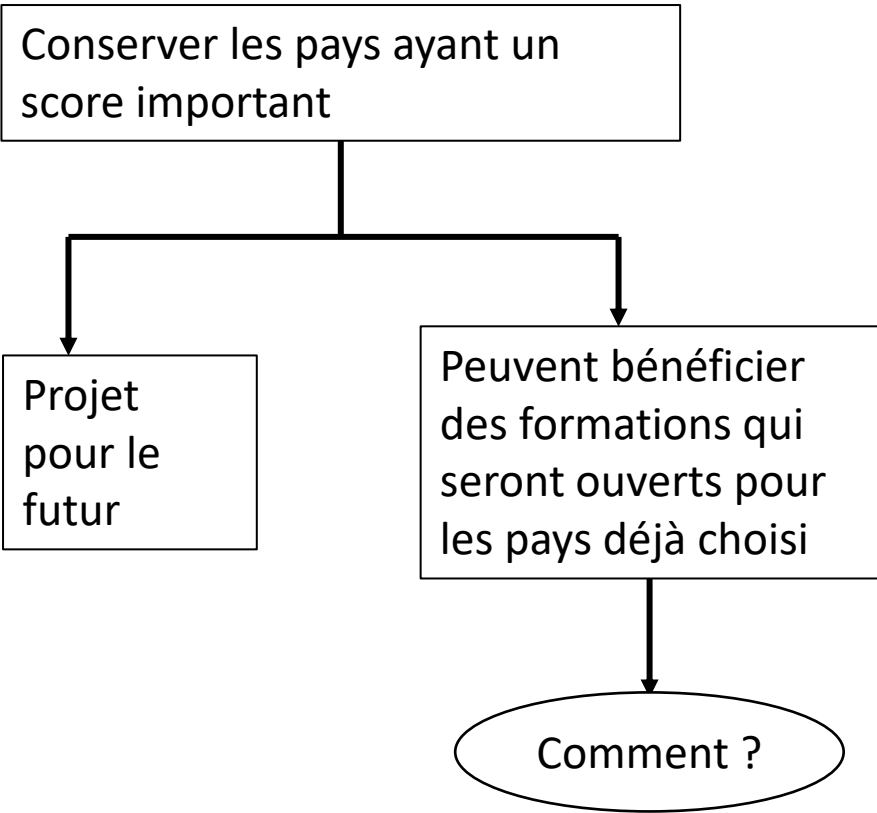
	S2010	S2011	S2012	S2013	S2014	S2015
Country						
Russian Federation	1203	1206	1225	1217	1217	1217
France	1232	1226	1230	1223	1235	1232
China	1225	1225	1223	1232	1235	1232
Germany	1268	1264	1262	1258	1261	1258
United States	1282	1277	1280	1270	1267	1267

- Une légère baisse de score pour Germany et United states ne les empêchent pas de rester les premiers pays du classement
- Même avec une légère variation de score ➡ le classement de pays ne varie pas

Classement des pays

Tableau représente l'évolution du Classement pour les dernières années pour les pays ayant un score important par rapport au cinq premières pays

	S2010	S2011	S2012	S2013	S2014	S2015
Country						
Malaysia	1147	1143	1095	1057	1074	1075
Korea, Rep.	1094	1093	1091	1059	1084	1092
Poland	1153	1137	1119	1135	1130	1116
Egypt, Arab Rep.	1082	1072	1081	1097	1095	1117
Italy	1183	1170	1164	1140	1123	1125
Spain	1117	1110	1116	1116	1122	1126
Israel	1108	1113	1115	1115	1130	1137
Australia	1133	1134	1133	1138	1143	1144
United Kingdom	1164	1162	1153	1156	1153	1144
Iran, Islamic Rep.	1150	1155	1153	1147	1152	1153
Canada	1207	1200	1189	1173	1160	1163
India	1163	1165	1165	1164	1171	1177
Japan	1214	1209	1203	1211	1201	1197
Brazil	1208	1210	1207	1208	1204	1204

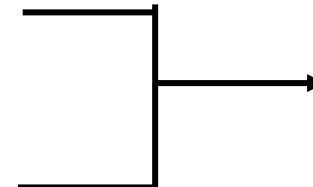


Etudes des langues

- Etude des langues pour les cinq premiers pays afin de créer une **seule formation pour plusieurs pays**

Rang	Pays	Langues d'enseignement
1	United States	Anglais
2	Germany	Allemand , Anglais
2	China	Chinois
3	France	France, Anglais
4	Russia	Russe

- Sélectionner les pays ayant :
- un **score important**
 - une **langues d'enseignement commune** entre les cinq premières pays



Liste 2 des pays :

Brazil
Japan
India
Canada
Untied Kingdom
Australia
Israël
Spain
Italy
Egypt
Korea
Malysia

Conclusion

- Jeu de données manquent des indicateurs qui pourraient être utiles pour répondre à la problématique :
 - l'argent dépensé sur internet
 - l'utilisation des e-learning dans les pays

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

- Les pays ayant un score élevé dans les dernières années seraient de bons candidats pour installer la formation : **United States**
Germany
China
France
Russia
- Les pays ayant un score important et qui parlent la même langue des pays choisis peuvent bénéficier des formations online.