



EA Stephen
Data scientist

Analysez des données de systèmes éducatifs



academy

Sommaire

1. Mission
2. Description du jeu de données
 - 2.1 EdStats Country Series
 - 2.2 EdStatsCountry
 - 2.3 EdStatsData
 - 2.4 EdStatsFootNote
 - 2.5 EdStatsSeries
3. Description de la table fusionner
4. Critère sélection des indicateurs
5. Data cleaning
 - 5.1 Recherche des indicateurs pertinents
 - 5.2 Sélection des indicateurs pertinents selon le taux de remplissage
 - 5.3 Sélection des indicateurs selon l'année
 - 5.4 Sélection de l'année et le taux remplissage selon la population
 - 5.5 Vérifier redondance des indicateurs
6. Étude statistique 2014 (même pays)
7. Les outliers selon les indicateurs
8. Répartition des pays
9. Inégalité des Income Group selon les régions
10. Segmentation des pays
11. ACP segmentation
12. Scoring attractivité des pays
13. Evolution des pays cibles
14. Conclusion

Je suis Data Scientist dans une **start-up de la EdTech**, nommée ***academy***, qui propose des contenus de formation en ligne pour un public de niveau lycée et université.

Le besoin métier

Analyse exploratoire pour déterminer si les données sur l'éducation de la banque mondiale permettent d'informer le projet d'expansion :

- Inspecter le dataset
- Extraction des données pertinentes
- Statistiques descriptives
- Segmentation des pays pour l'expansion des contenus de formation en ligne pour un public de niveau lycée et université.

Description de EdStatsCountry_Series

EdStatsCountry_Series

	CountryCode	SeriesCode	DESCRIPTION
0	ABW	SP.POP.TOTL	Data sources : United Nations World Population...

➤ 0 doublon

Le dataset : **EDSTATSCOUNTRY_SERIES**

	CountryCode	SeriesCode	DESCRIPTION
count	613	613	613
unique	211	21	97
top	GEO	SP.POP.TOTL	Data sources : United Nations World Population Prospects
freq	18	211	154

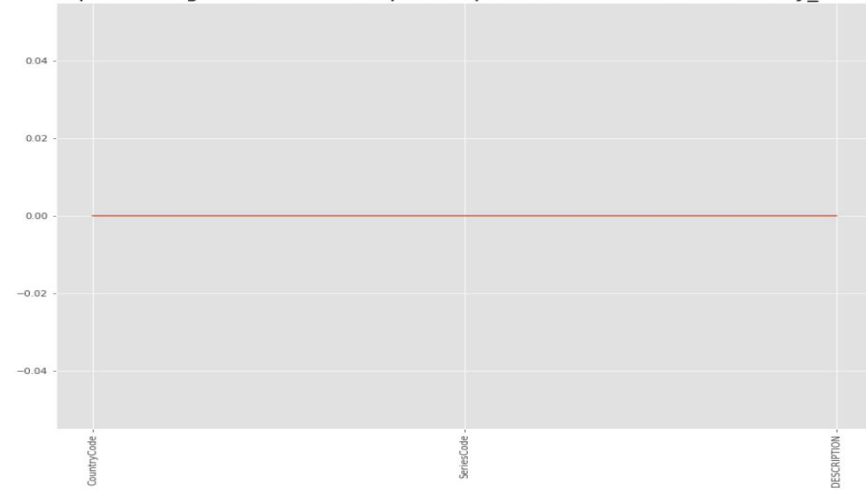
Le dataset : **EDSTATSCOUNTRY_SERIES**

Les dimensions du EdStatsCountry_Series : (613, 3)

Les valeurs manquantes dans chaque colonne :

	Count_NaN	%_NaN_col	Types	Total_NaN_in_dataset	%_NaN_in_dataset
CountryCode	0	0.0	object	0	0.0
SeriesCode	0	0.0	object	0	0.0
DESCRIPTION	0	0.0	object	0	0.0

Le pourcentage de valeurs manquantes pour la colonne EdStatsCountry_Series



Description de EdStatsCountry

Le dataset : **EDSTATSCOUNTRY**
Les dimensions du EdStatsCountry : (241, 31)
Les valeurs manquantes dans chaque colonne :

	Count_NaN	%_NaN_col	Type	Total_NaN_in_dataset	%_NaN_in_dataset
National accounts reference year	209	86.72	float64	2113	28.28
Alternative conversion factor	194	80.5	object	2113	28.28
Other groups	183	75.93	object	2113	28.28
Latest industrial data	134	55.6	float64	2113	28.28
Vital registration complete	130	53.94	object	2113	28.28
External debt Reporting status	117	48.55	object	2113	28.28
Latest household survey	100	41.49	object	2113	28.28
Latest agricultural census	99	41.08	object	2113	28.28
Lending category	97	40.25	object	2113	28.28
Special Notes	96	39.83	object	2113	28.28
PPP survey year	96	39.83	object	2113	28.28
Source of most recent income and expenditure data	81	33.61	object	2113	28.28
Government Accounting concept	80	33.2	object	2113	28.28
Latest water withdrawal data	62	25.73	object	2113	28.28
Balance of Payments Manual in use	60	24.9	object	2113	28.28
IMF data dissemination standard	60	24.9	object	2113	28.28
Latest trade data	56	23.24	float64	2113	28.28
SNA price valuation	44	18.26	object	2113	28.28
System of trade	41	17.01	object	2113	28.28
National accounts base year	36	14.94	object	2113	28.28
Latest population census	28	11.62	object	2113	28.28
Income Group	27	11.2	object	2113	28.28
System of National Accounts	26	10.79	object	2113	28.28
Current	26	10.79	object	2113	28.28
2-alpha code	3	1.24	object	2113	28.28
WB-2 code	1	0.41	object	2113	28.28
Short Name	0	0.0	object	2113	28.28
Long Name	0	0.0	object	2113	28.28
Table Name	0	0.0	object	2113	28.28
Country Code	0	0.0	object	2113	28.28

EdStatsCountry

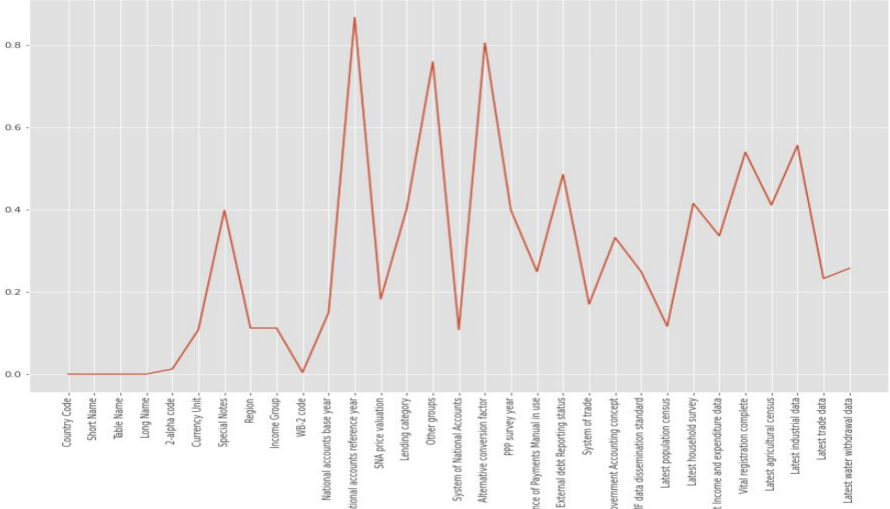
Country Code	Short Name	Table Name	Long Name	2-alpha code	Currency Unit	Special Notes	Region	Income Group
0	ABW	Aruba	Aruba	AW	Aruban florin	SNA data for 2000-2011 are updated from offici...	Latin America & Caribbean	High income: nonOECD

➤ 0 doublon

Le dataset : **EDSTATSCOUNTRY**

	Country Code	Short Name	Region	Income Group
count	241	241	214	214
unique	241	241	7	5
top	SVN	Mauritania	Europe & Central Asia	Upper middle income
freq	1	1	57	55
mean	nan	nan	nan	nan
std	nan	nan	nan	nan
min	nan	nan	nan	nan
25%	nan	nan	nan	nan
50%	nan	nan	nan	nan
75%	nan	nan	nan	nan
max	nan	nan	nan	nan

Le pourcentage de valeurs manquantes pour la colonne EdStatsCountry



Description de EdStatsData

Le dataset : EDSTATSDATA
Les dimensions du EdStatsData : (886930, 69)
Les valeurs manquantes dans chaque colonne :

	Count_NaN	%_NaN_col	Types	Total_NaN_in_dataset	%_NaN_in_dataset
2017	886787	99.98	float64	52568249	85.9
2016	870470	98.14	float64	52568249	85.9
1971	851393	95.99	float64	52568249	85.9
1973	851385	95.99	float64	52568249	85.9
1972	851311	95.98	float64	52568249	85.9
1974	851200	95.97	float64	52568249	85.9
1979	850121	95.85	float64	52568249	85.9
1976	849447	95.77	float64	52568249	85.9
1982	849419	95.77	float64	52568249	85.9
1989	849390	95.77	float64	52568249	85.9
1977	849356	95.76	float64	52568249	85.9
1978	849354	95.76	float64	52568249	85.9
1983	848470	95.66	float64	52568249	85.9
1988	848378	95.65	float64	52568249	85.9
1984	848324	95.65	float64	52568249	85.9
1987	848289	95.64	float64	52568249	85.9
1981	848153	95.63	float64	52568249	85.9
1986	847558	95.56	float64	52568249	85.9
2030	835494	94.2	float64	52568249	85.9
2025	835494	94.2	float64	52568249	85.9
2020	835494	94.2	float64	52568249	85.9
2075	835494	94.2	float64	52568249	85.9
2100	835494	94.2	float64	52568249	85.9
2035	835494	94.2	float64	52568249	85.9
2085	835494	94.2	float64	52568249	85.9
2040	835494	94.2	float64	52568249	85.9
2045	835494	94.2	float64	52568249	85.9
2050	835494	94.2	float64	52568249	85.9
2055	835494	94.2	float64	52568249	85.9
2060	835494	94.2	float64	52568249	85.9
2080	835494	94.2	float64	52568249	85.9
2065	835494	94.2	float64	52568249	85.9

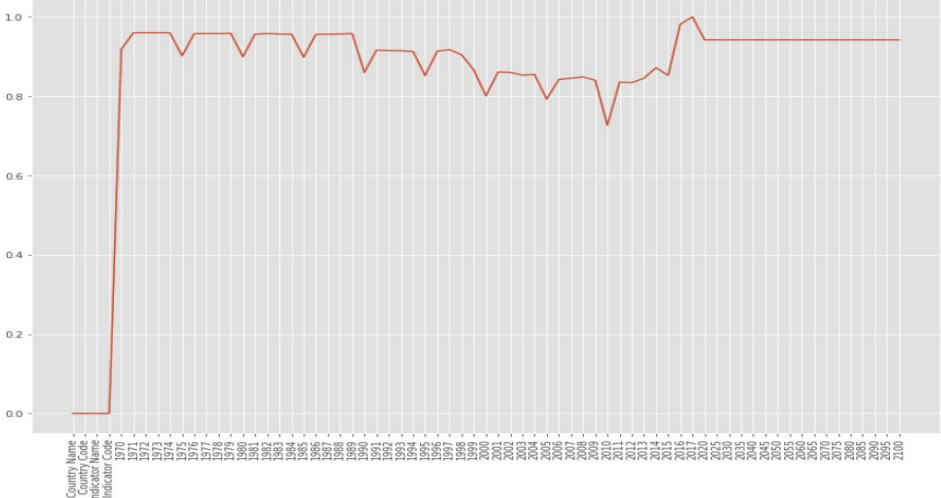
Country Name	Country Code	Indicator Name	Indicator Code	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979
0	Arab World	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

➤ 0 doublon

Le dataset : EDSTATSDATA

	Country Name	Country Code	Indicator Name	Indicator Code	1970
count	886930	886930	886930	886930	72288.0
unique	242	242	3665	3665	nan
top	Brunei Darussalam	BFA	PIAAC: Distribution of Adult Numeracy Scores: 50th Percentile Score	SP.POP.AG02.MA.UN	nan
freq	3665	3665	242	242	nan
mean	nan	nan	nan	nan	1974772449.58550
std	nan	nan	nan	nan	121168685352.022
min	nan	nan	nan	nan	-1.4355636441178
25%	nan	nan	nan	nan	0.89
50%	nan	nan	nan	nan	6.3177236012718
75%	nan	nan	nan	nan	62.5125
max	nan	nan	nan	nan	19039286948196.

Le pourcentage de valeurs manquantes pour la colonne EdStatsData



Description de EdStatsFootNote

EdStatsFootNote

	CountryCode	SeriesCode	Year	DESCRIPTION
0	ABW	SE.PRE.ENRL.FE	YR2001	Country estimation.

➤ 0 doublon

Le dataset : **EDSTATSFOOTNOTE**

	CountryCode	SeriesCode	Year	DESCRIPTION
count	643638	643638	643638	643638
unique	239	1558	56	9102
top	LIC	SH.DYN.MORT	YR2004	Country Data
freq	7320	9226	27128	191188

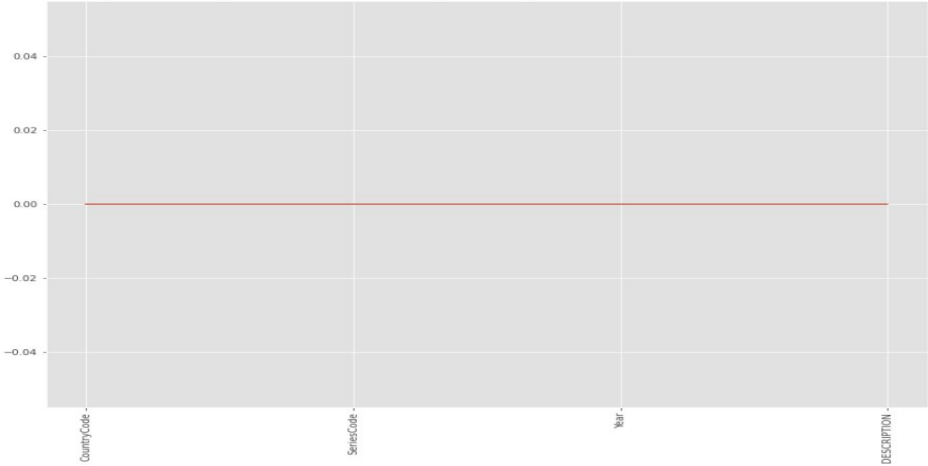
Le dataset : **EDSTATSFOOTNOTE**

Les dimensions du EdStatsFootNote : (643638, 4)

Les valeurs manquantes dans chaque colonne :

	Count_NaN	%_NaN_col	Types	Total_NaN_in_dataset	%_NaN_in_dataset
CountryCode	0	0.0	object	0	0.0
SeriesCode	0	0.0	object	0	0.0
Year	0	0.0	object	0	0.0
DESCRIPTION	0	0.0	object	0	0.0

Le pourcentage de valeurs manquantes pour la colonne EdStatsFootNote



Description de EdStatsSeries

EdStatsSeries

	Series Code	Topic	Indicator Name	Short definition	Long definition	Periodicity	Base Period	Other notes
0	BAR.NOED.1519.FE.ZS	Attainment	Barro-Lee: Percentage of female population age...	Percentage of female population age 15-19 with...	Percentage of female population age 15-19 with...	NaN	NaN	NaN

➤ 0 doublon

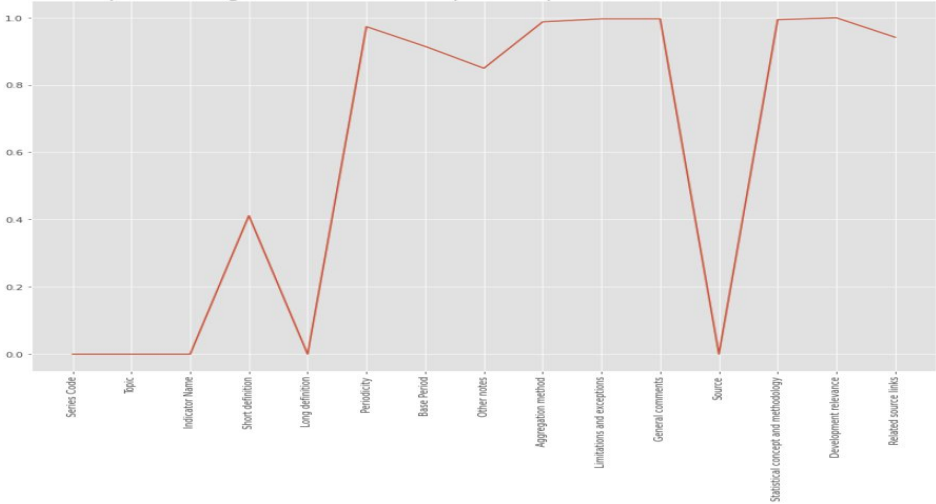
Le dataset : **EDSTATSSERIES**

Les dimensions du EdStatsSeries : (3665, 15)

Les valeurs manquantes dans chaque colonne :

	Count_NaN	%_NaN_col	Types	Total_NaN_in_dataset	%_NaN_in_dataset
Development relevance	3662	99.92	object	33213	60.41
Limitations and exceptions	3651	99.62	object	33213	60.41
General comments	3651	99.62	object	33213	60.41
Statistical concept and methodology	3642	99.37	object	33213	60.41
Aggregation method	3618	98.72	object	33213	60.41
Periodicity	3566	97.3	object	33213	60.41
Related source links	3450	94.13	object	33213	60.41
Base Period	3351	91.43	object	33213	60.41
Other notes	3113	84.94	object	33213	60.41
Short definition	1509	41.17	object	33213	60.41
Series Code	0	0.0	object	33213	60.41
Topic	0	0.0	object	33213	60.41
Indicator Name	0	0.0	object	33213	60.41
Long definition	0	0.0	object	33213	60.41
Source	0	0.0	object	33213	60.41

Le pourcentage de valeurs manquantes pour la colonne EdStatsSeries



Le dataset : **EDSTATSSERIES**

	Series Code	Topic	Indicator Name
count	3665	3665	3665
unique	3665	37	3665
top	LO.PASEC.WAT.6.PP	Learning Outcomes	SABER: (Engaging the Private Sector, Government funded) Policy Goal 7 Lever 1: Information
freq	1	1046	1

Description de la table fusionner

- `df_merged1 = EdStatsData.merge(EdStatsSeries, how='left', on='Indicator Name')`
- `df_fusion = df_merged1.merge(EdStatsCountry, how='left', on='Country Code')`

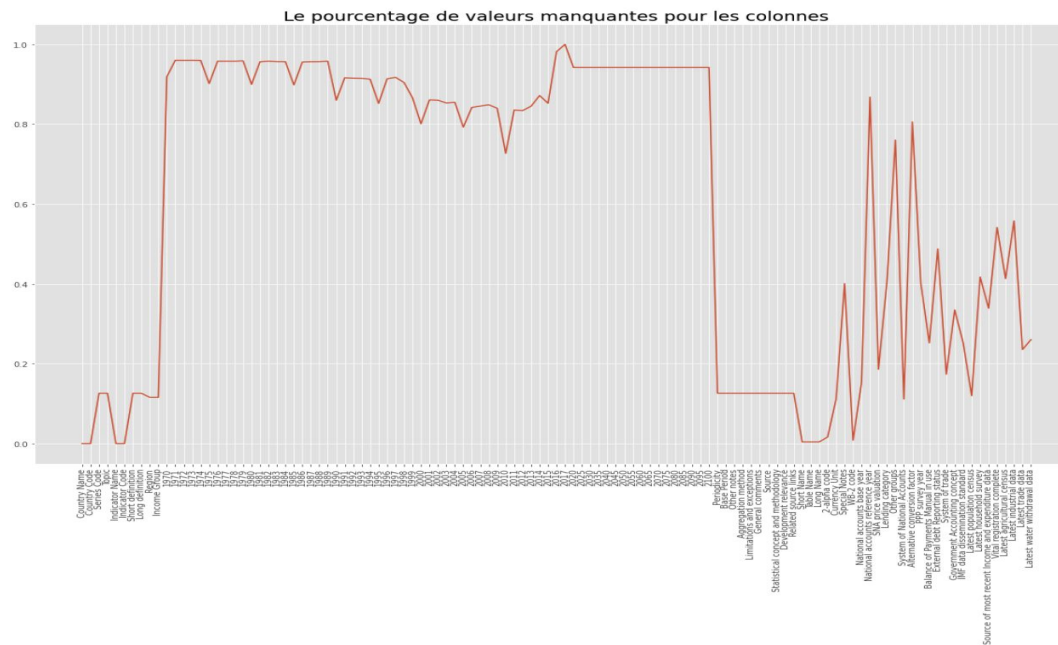
	Country Name	Country Code	Series Code	Topic	Indicator Name	Indicator Code	Short definition	Long definition	Region	Income Group	1970	1971	1972
0	Arab World	ARB	UIS.NERA.2	Secondary	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2	NaN	Total number of students of the official lower...	NaN	NaN	NaN	NaN	NaN

- Les dimensions du `df_fusion`: (886930, 113)
- dtypes: float64(68), object(45)
- 0 doublon

	Country Name	Country Code	Series Code	Topic	Indicator Name	Indicator Code	Short definition	Long definition	Region	Income Group	1970
count	886830	886830	775126	775126	886830	886830	775126	775126	784310	7228800e+04	
unique	242	242	3203	32	3665	3665	1125	1977	7	5	NaN

	Country Name	Country Code	Series Code	Topic	Indicator Name	Indicator Code	Short definition	Long definition	Region	Income Group	1970
top	Syrian Arab Republic	VIR	UIS.EA.778.AG25T99.M	Learning Outcomes	PIAAC: Male adults by numeracy proficiency lev...	UIS.EA.778.AG25T99.M	NaN	Data Interpretation: 1=Latent; 2=Emerging; 3=E...	Europe & Central Asia	Upper middle income	NaN
freq	3665	3665	242	242484	242	242	272250	44528	208905	201575	NaN

	Count_NaN	%_NaN_col	Total_NaN_in_dataset	%_NaN_in_dataset	Types
2017	886787	99.98	61987600	61.85	float64
2016	870470	98.14	61987600	61.85	float64
1971	851393	95.99	61987600	61.85	float64
1973	851385	95.99	61987600	61.85	float64
1972	851311	95.98	61987600	61.85	float64
1974	851200	95.97	61987600	61.85	float64
1979	850121	95.85	61987600	61.85	float64
1976	849447	95.77	61987600	61.85	float64
1982	849419	95.77	61987600	61.85	float64
1989	849390	95.77	61987600	61.85	float64
1977	849356	95.76	61987600	61.85	float64
1978	849354	95.76	61987600	61.85	float64
1983	848470	95.66	61987600	61.85	float64
1988	848378	95.65	61987600	61.85	float64
1984	848324	95.65	61987600	61.85	float64
1987	848289	95.64	61987600	61.85	float64
1981	848153	95.63	61987600	61.85	float64
1986	847558	95.56	61987600	61.85	float64
2055	835494	94.2	61987600	61.85	float64
2045	835494	94.2	61987600	61.85	float64



Critère sélection des indicateurs

- Assurer des sessions de vidéos conférences (accès à internet, accès à un ordinateur)
- Niveau activité économique
- La population visée (Secondary, tertiaire)

Supprimer les colonnes :

- Supprimer colonnes 100% NaN
- Supprimer 39 colonnes non pertinentes

Supprimer les lignes :

- Supprimer les lignes contenant 100% NaN sur les années 1970 à 2100

Supprimer les Nan colonne Region :

- Supprimer les NaN dans colonne Region qui correspondent à des zones géographiques , Income Group dans la colonne Country Name (28)

Les dimensions après nettoyage : (347348, 73)

Le nombre de pays uniquement est de : 214

Le nombre de Régions est de : 7

Le nombre de Topics est de : 32

Le nombre de Income Group est de : 5

Le nombre de Indicator Name est de : 3665

	index	Country Name	Country Code	Topic	Indicator Name	Long definition	Region	Income Group	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981		1982	1983	1984	1985	1986	1987
91625	91625	Afghanistan	AFG	Secondary	Adjusted net enrolment rate, lower secondary, ...	Total number of students of the official lower...	South Asia	Low income	NaN	NaN	NaN	NaN	7.05911	NaN	NaN	NaN	NaN	NaN	NaN	NaN		NaN	NaN	NaN	NaN	NaN	NaN

Recherche des indicateurs pertinents

World bank :

Il y a 7 TOPICS et 42 INDICATEURS (WORLD BANK):

Le nombre d'indicateurs trouvés sur la world bank pour **Expenditures** : 8

Le nombre d'indicateurs trouvés sur la world bank pour **Background** : 4

Le nombre d'indicateurs trouvés sur la world bank pour **Secondary** : 8

Le nombre d'indicateurs trouvés sur la world bank pour **Tertiary** : 6

Le nombre d'indicateurs trouvés sur la world bank pour **Attainment** : 6

Le nombre d'indicateurs trouvés sur la world bank pour **Population** : 8

Le nombre d'indicateurs trouvés sur la world bank pour **Education_Equality** : 2

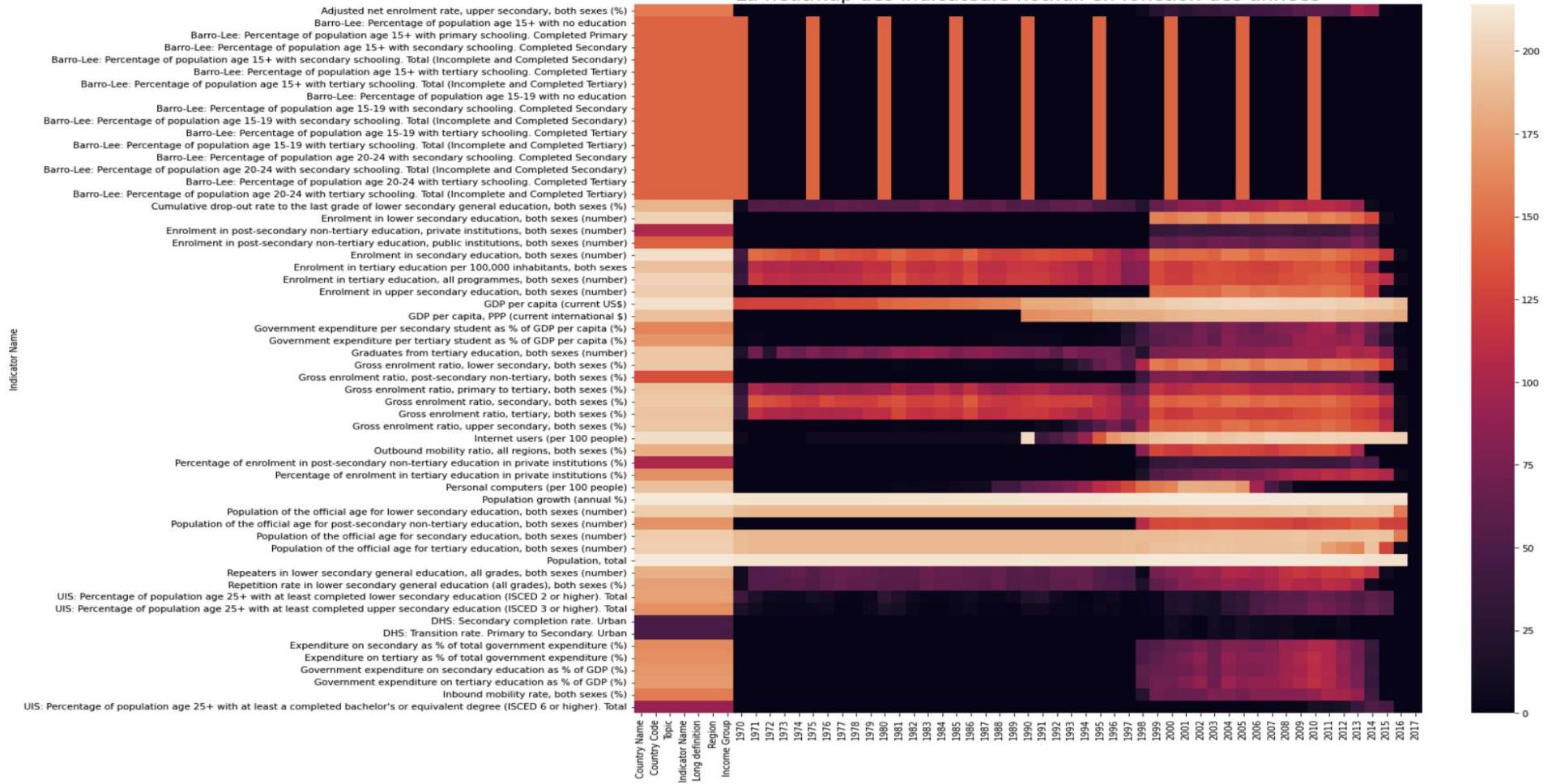
- 20 indicateurs trouvés dans le dataset
- 22 indicateurs non trouvés dans le dataset

Le nombre d'indicateurs trouvés dans le dataset :

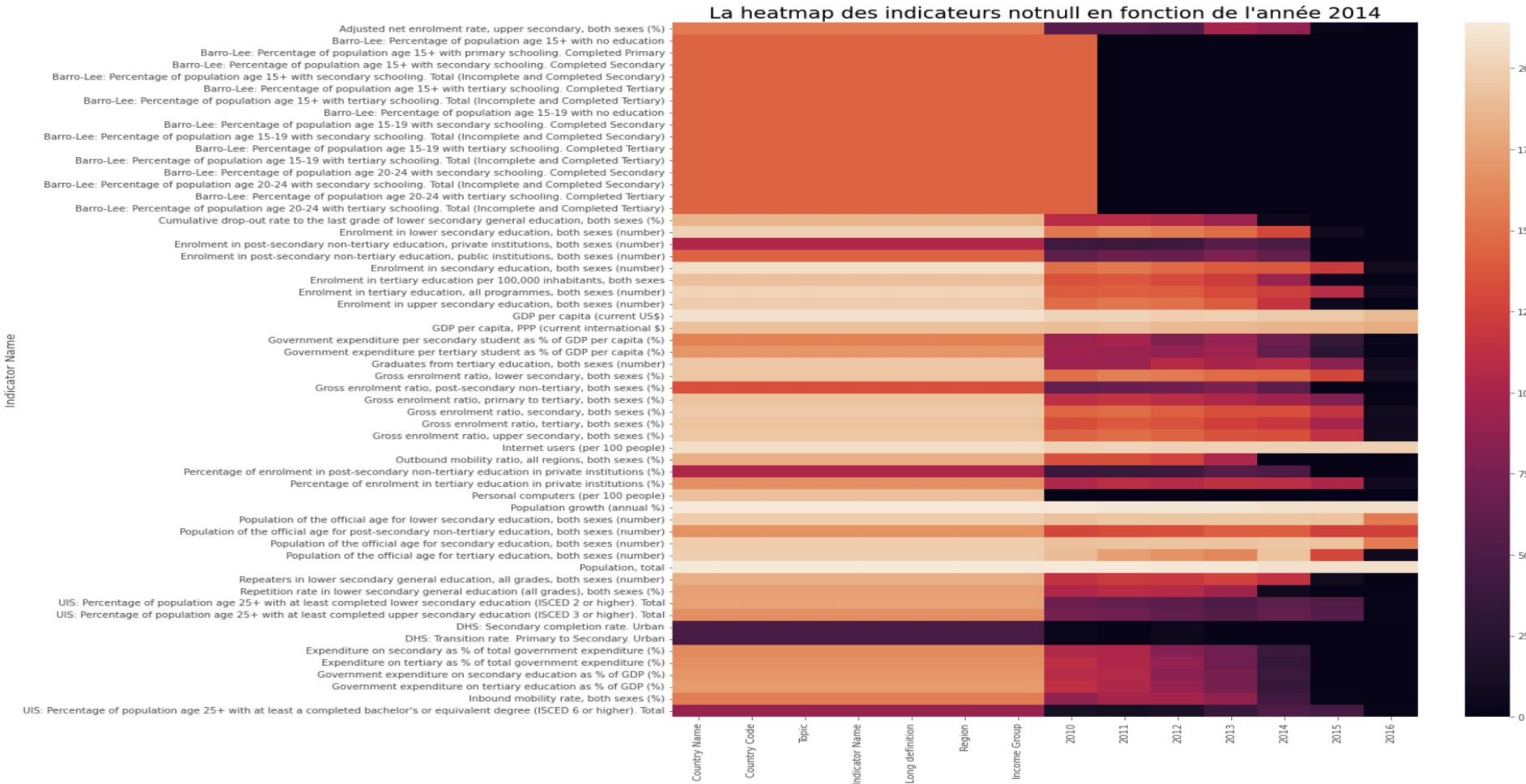
- 58 indicateurs

Sélection des indicateurs selon le taux de remplissage

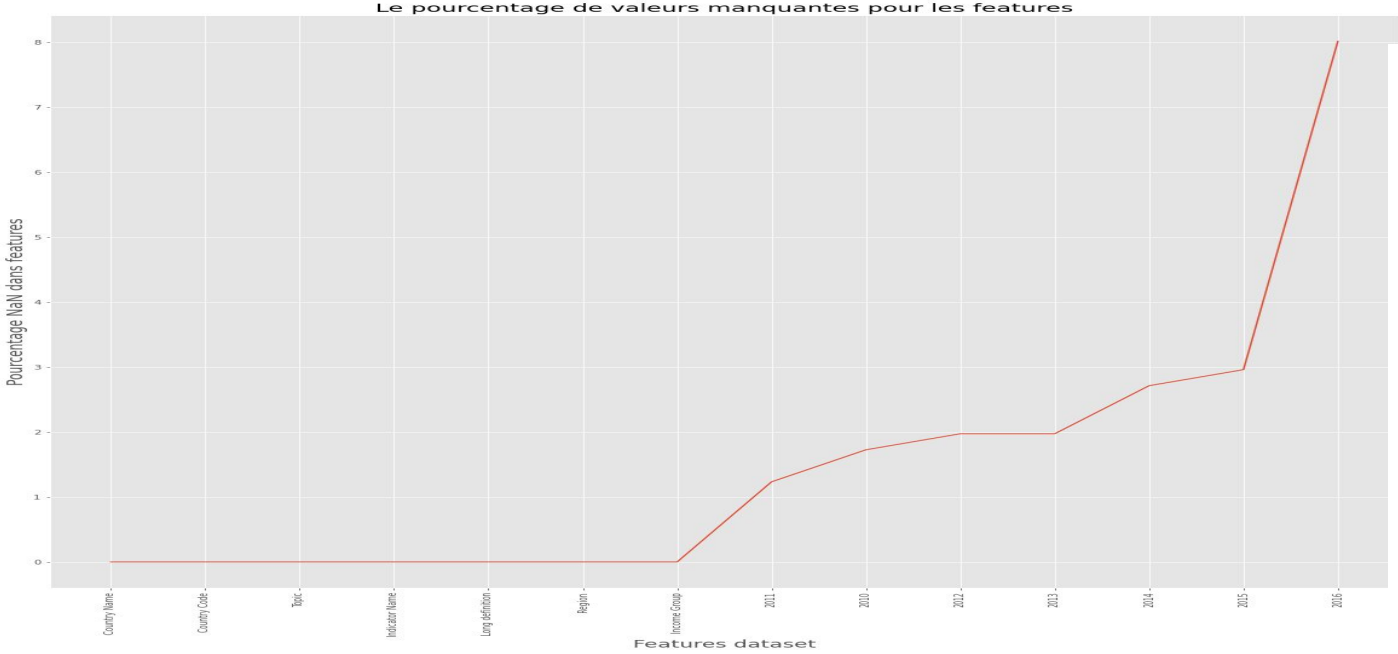
La heatmap des indicateurs notnull en fonction des années



Sélection des indicateurs selon l'année



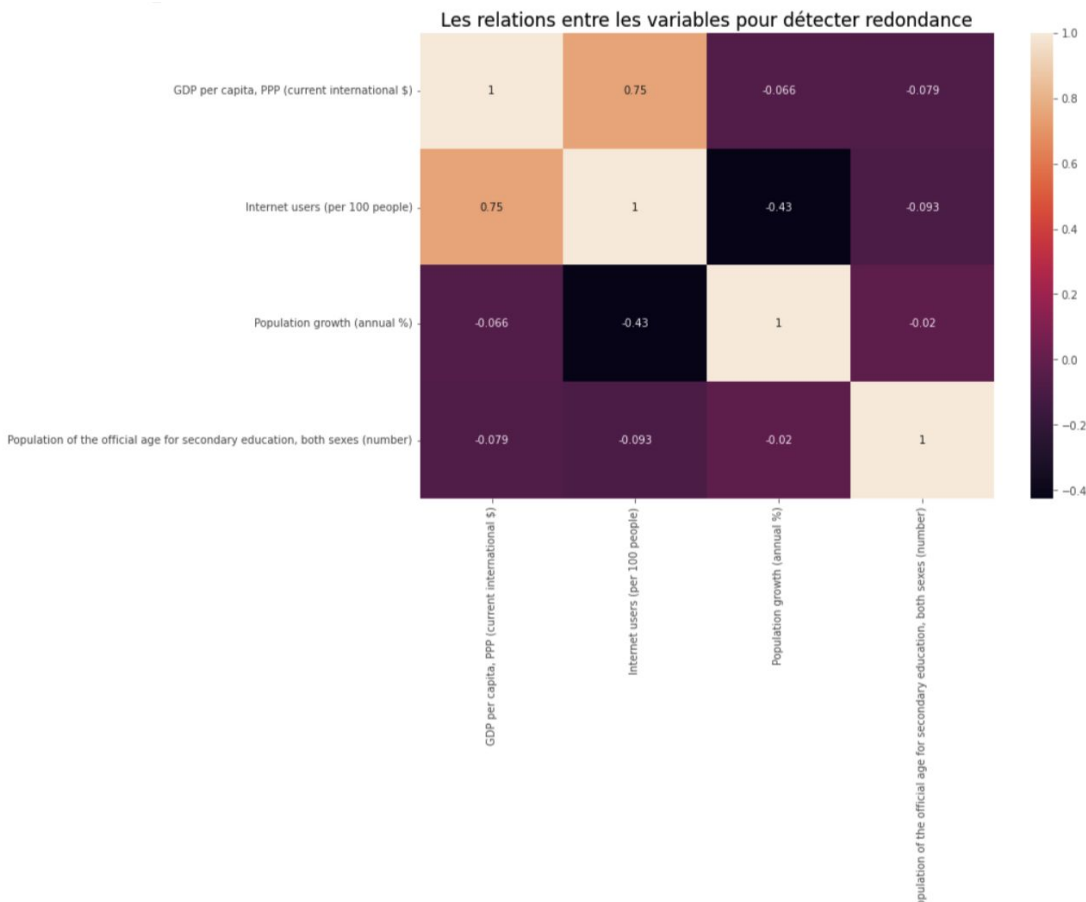
Sélection de l'année et le taux remplissage selon la population



	Pct_NaN_colonne
2016	8.014797
2015	2.959309
2014	2.712700
2013	1.972873
2012	1.972873
2010	1.726264
2011	1.233046
Income Group	0.000000
Region	0.000000
Long definition	0.000000
Indicator Name	0.000000
Topic	0.000000
Country Code	0.000000
Country Name	0.000000

Indicator Name	Country Name	Country Code	Topic	Indicator Name	Long definition	Region	Income Group	2010	2011	2012	2013	2014	2015	2016
GDP per capita, PPP (current international \$)	0	0	0	0	0	0	0	2	0	4	4	7	8	11
Internet users (per 100 people)	0	0	0	0	0	0	0	6	4	6	6	6	6	7
Population growth (annual %)	0	0	0	0	0	0	0	0	1	1	1	5	5	5
Population of the official age for secondary education, both sexes (number)	0	0	0	0	0	0	0	6	5	5	5	4	5	42

Vérifier redondance des indicateurs



feature	VIF
0 GDP per capita, PPP (current international \$)	3.895448
1 Internet users (per 100 people)	4.028225
2 Population growth (annual %)	1.348500
3 Population of the official age for secondary education, both sexes (number)	1.052095

Année 2014

- 179 pays conservés
- 33 pays supprimés

Etude statistique 2014

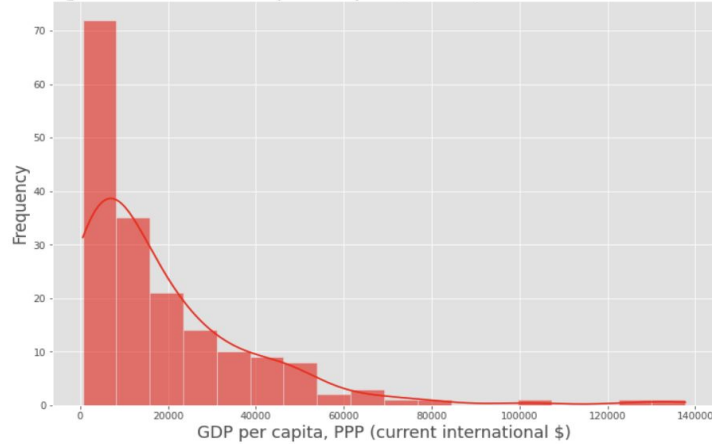
Les Income Group et les régions sont hétérogènes par rapport aux indicateurs

Region	GDP per capita, PPP (current international \$)			Internet users (per 100 people)			Population growth (annual %)			Population of the official age for secondary education, both sexes (number)		
	mean	std	median	mean	std	median	mean	std	median	mean	std	median
East Asia & Pacific	20596.596743	30127.281851	6371.496100	40.098294	27.642454	35.422220	1.215194	0.702827	1.237154	6169287.035714	18092967.950050	415892.500000
Europe & Central Asia	30148.082293	18869.355094	27358.060500	68.716714	20.208567	71.055000	0.389662	0.874926	0.226035	1589349.521739	2388667.992457	624480.000000
Latin America & Caribbean	14063.913759	6936.190576	12997.621000	47.298620	16.612930	48.490000	1.034942	0.528544	1.067498	2091350.343750	4769272.268499	510296.000000
Middle East & North Africa	32612.126667	32211.436719	16654.631500	55.866063	26.606778	60.756825	2.833274	1.847549	2.104405	2229650.444444	2683327.459333	831162.000000
North America	49840.279000	6729.211382	49840.279000	80.060000	9.984348	80.060000	0.921940	0.252883	0.921940	13628093.500000	15848359.537997	13628093.500000
South Asia	6450.500187	4547.426243	5249.140800	21.840000	13.426877	18.220000	1.652873	0.748489	1.345390	29553958.375000	59392527.581898	4704007.500000
Sub-Saharan Africa	5293.642026	7039.879907	2337.013400	15.916981	13.058488	10.700000	2.532641	0.774087	2.606212	2842619.666667	4159771.539774	1763278.000000

Income Group	GDP per capita, PPP (current international \$)			Internet users (per 100 people)			Population growth (annual %)			Population of the official age for secondary education, both sexes (number)		
	mean	std	median	mean	std	median	mean	std	median	mean	std	median
High income: OECD	42546.647484	15117.807814	41034.918000	81.585677	11.088296	84.241530	0.600659	0.694962	0.507053	2502143.709677	4621716.030962	734239.000000
High income: nonOECD	46423.448818	33243.487376	33254.691500	71.239684	14.333042	70.370000	1.576504	2.001562	1.002182	699176.181818	1968428.893313	102421.000000
Low income	1865.551187	893.211069	1724.698700	8.685108	4.955856	7.000000	2.476594	0.795888	2.606212	3774226.516129	4654730.959217	2195237.000000
Lower middle income	5649.226500	2613.053746	5407.388700	25.310591	12.066501	22.550000	1.641243	0.910401	1.733953	7286559.444444	26281946.846825	906451.000000
Upper middle income	14755.384800	5354.266394	14651.185500	46.986525	15.521094	48.450000	1.223376	1.261530	1.124486	4133364.900000	13632913.117129	603255.500000

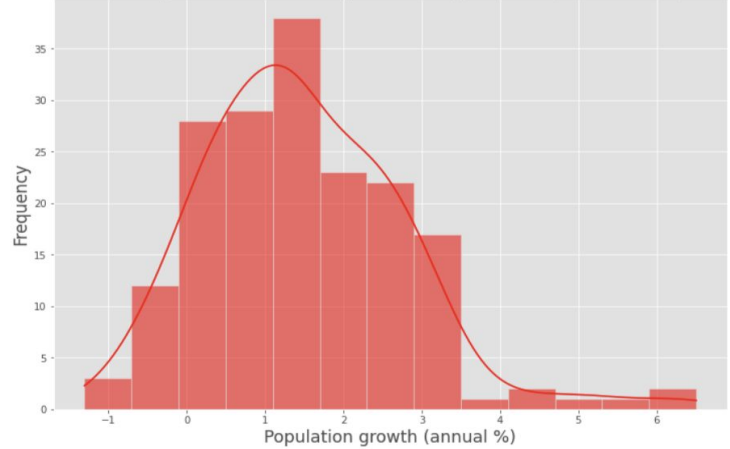
Les outliers selon les indicateurs

Histogramme de GDP per capita, PPP (current international \$)



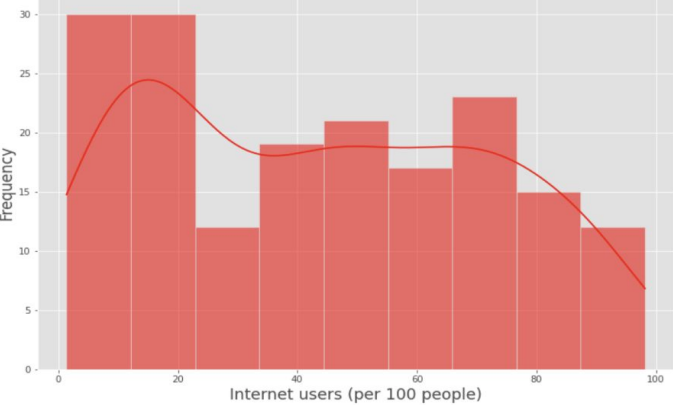
- Macao SAR, China
- Qatar
- Luxembourg
- Brunei
- Darussalam
- Kuwait
- United Arab Emirates
- Norway
- Switzerland
- Hong Kong SAR, China
- United States

Histogramme de Population growth (annual %)



- Oman
- Lebanon
- Qatar
- Kuwait
- Jordan
- Equatorial Guinea
- Niger
- Angola
- Uganda
- Congo, Dem. Rep.

Histogramme de Internet users (per 100 people)



- Iceland
- Norway
- Denmark
- Luxembourg
- Sweden
- Netherlands
- United Kingdom
- Qatar
- Bahrain
- United Arab Emirates

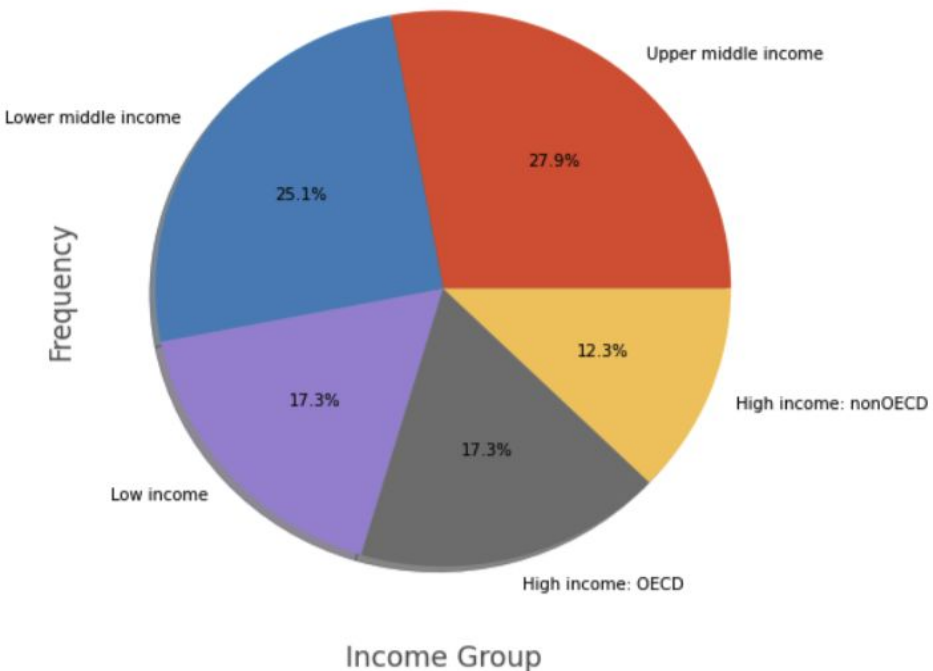
Histogramme de Population of the official age for secondary education, both sexes (number)



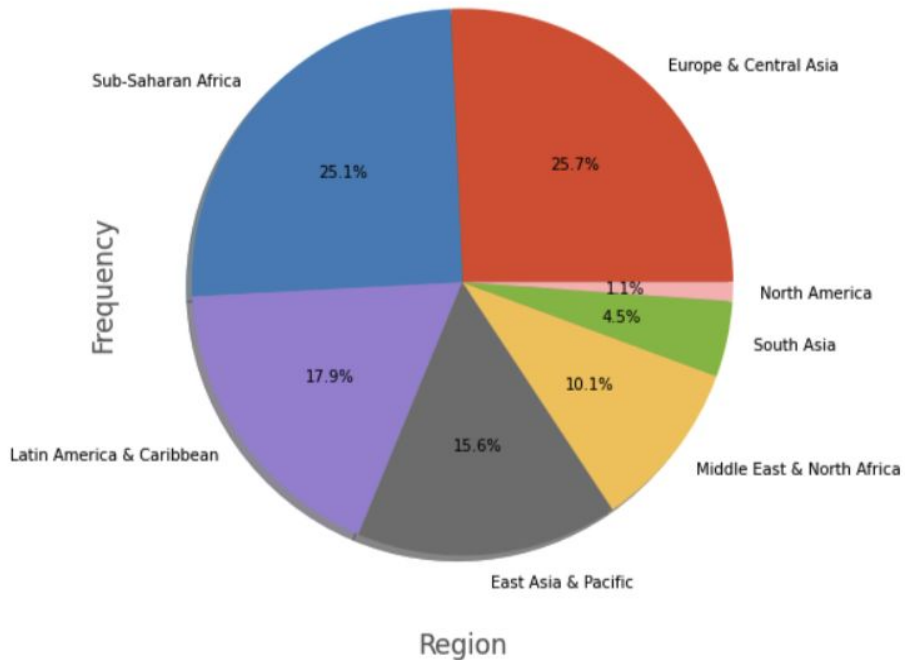
- India
- China
- Indonesia
- Pakistan
- United States
- Brazil
- Nigeria
- Bangladesh
- Mexico
- Ethiopia

Répartition des pays

Répartition de la variable Income Group

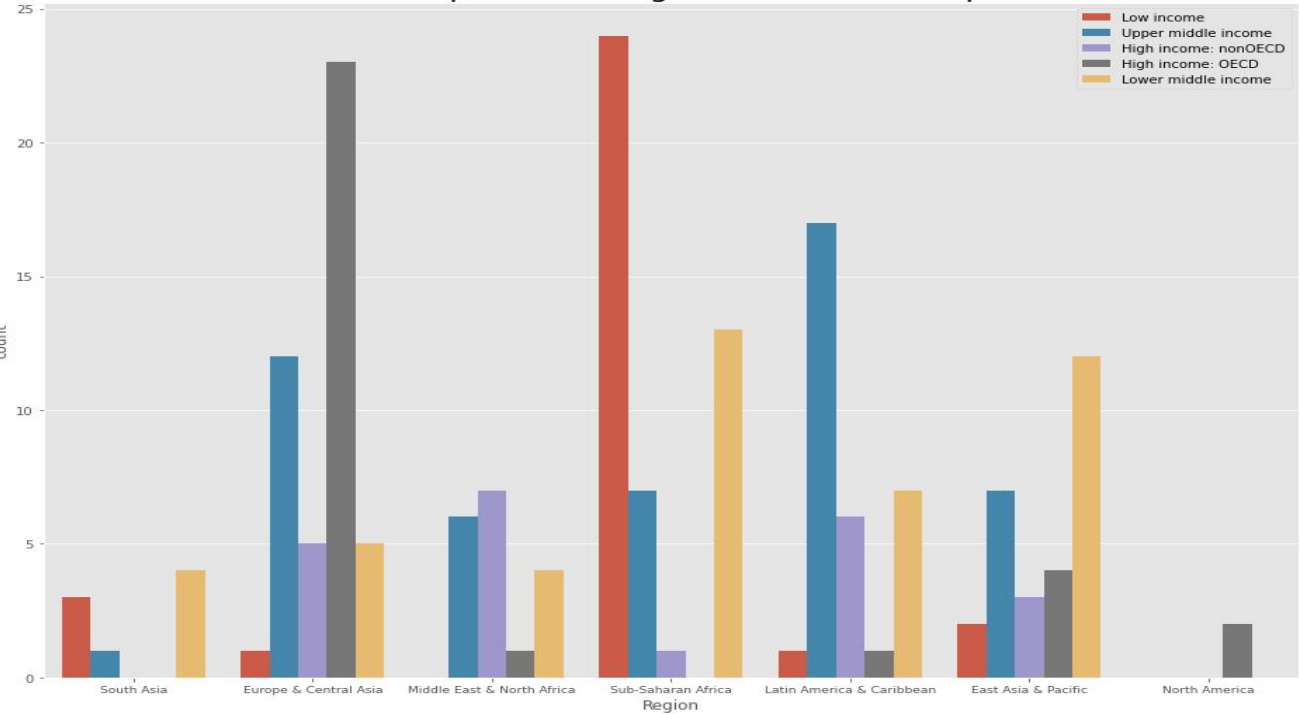


Répartition de la variable Region

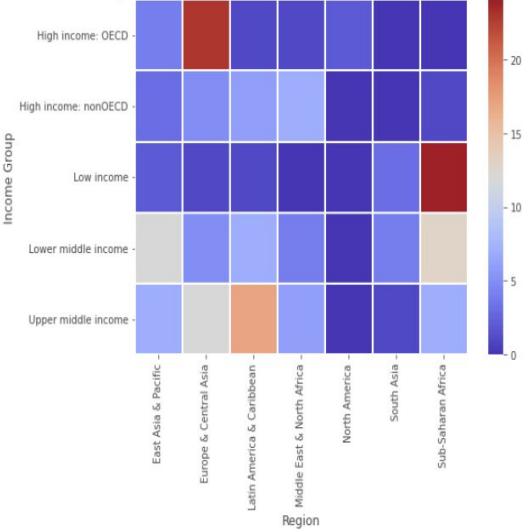


Inégalité des income Group selon les régions

Countplot entre Region et Income Group

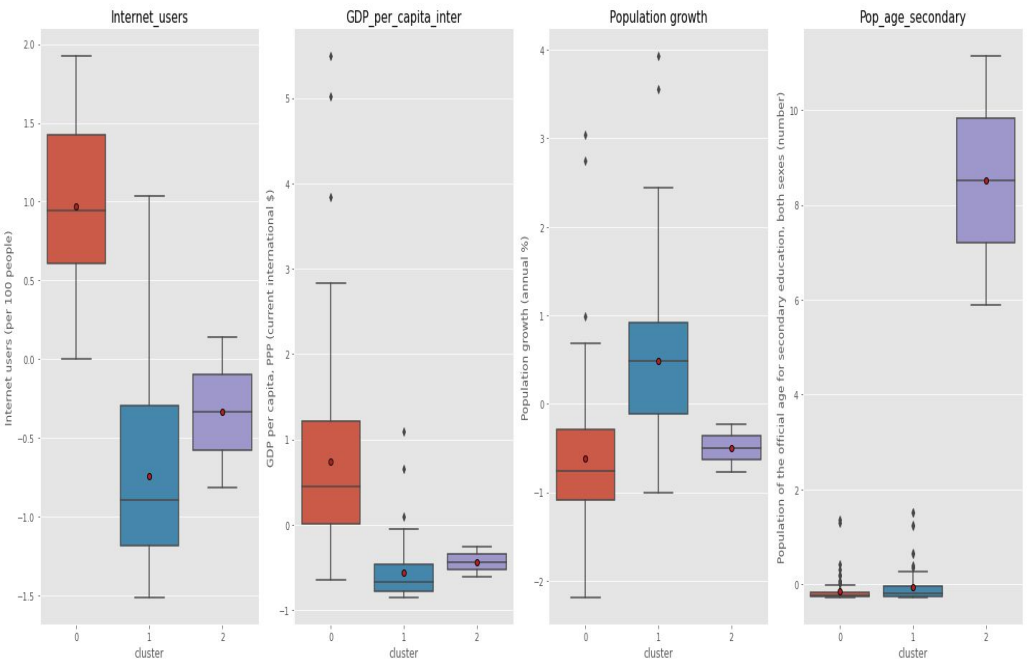


La fréquence entre les incomes Group et les régions



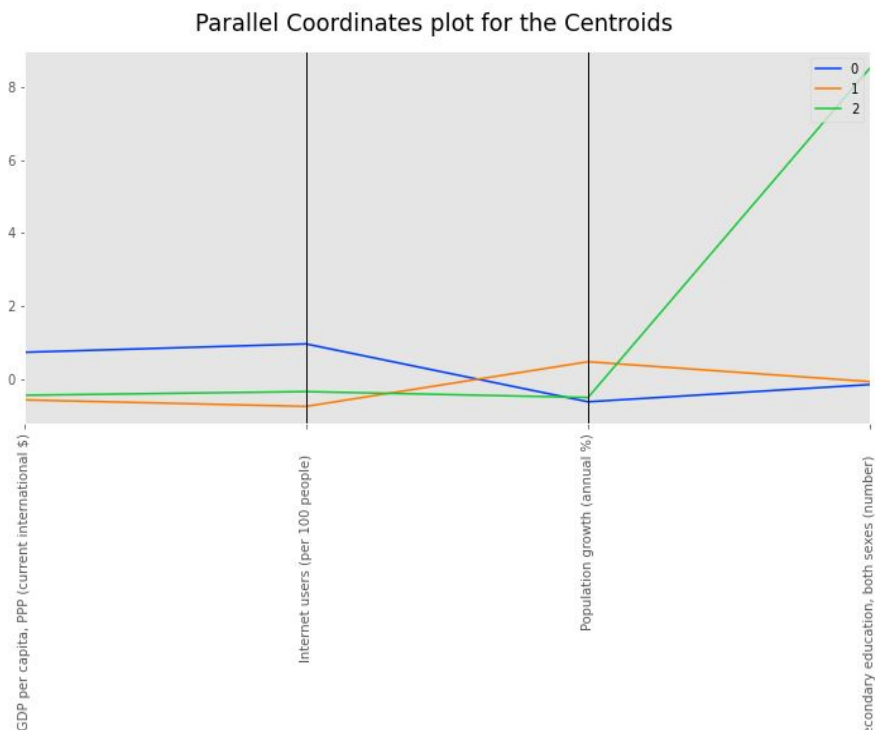
Income Group	East Asia & Pacific	Europe & Central Asia	Latin America & Caribbean	Middle East & North Africa	North America	South Asia	Sub-Saharan Africa	All
High income: OECD	4	23	1	1	2	0	0	31
High income: nonOECD	3	5	6	7	0	0	1	22
Low income	2	1	1	0	0	3	24	31
Lower middle income	12	5	7	4	0	4	13	45
Upper middle income	7	12	17	6	0	1	7	50
All	28	46	32	18	2	8	45	179

Segmentation des pays

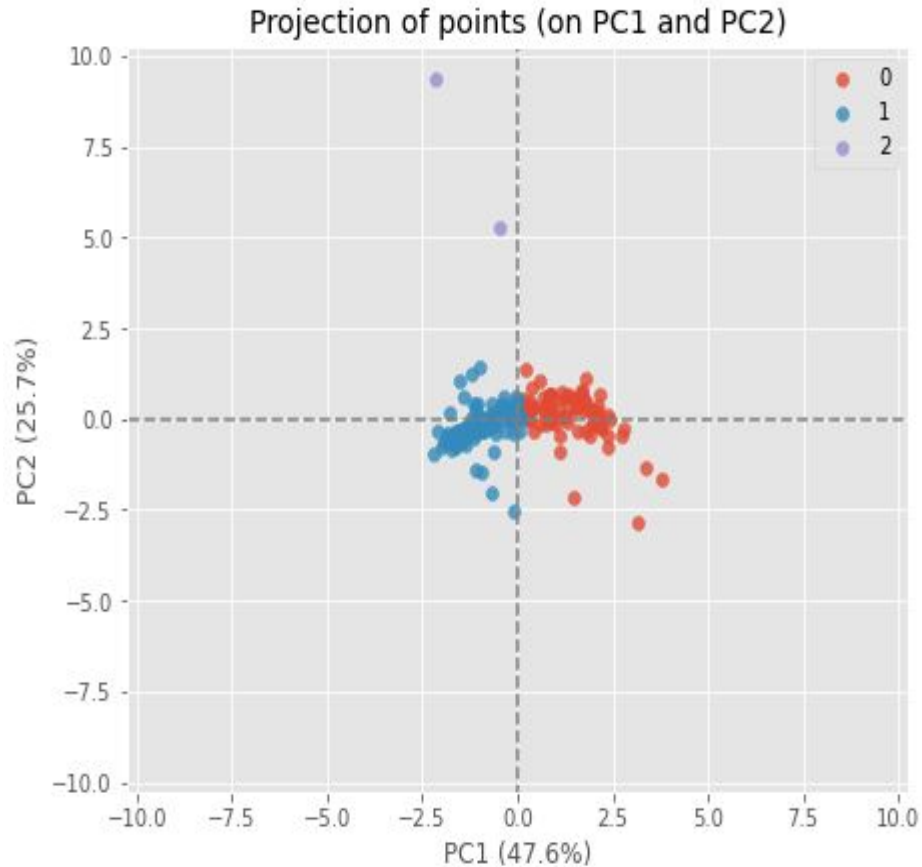
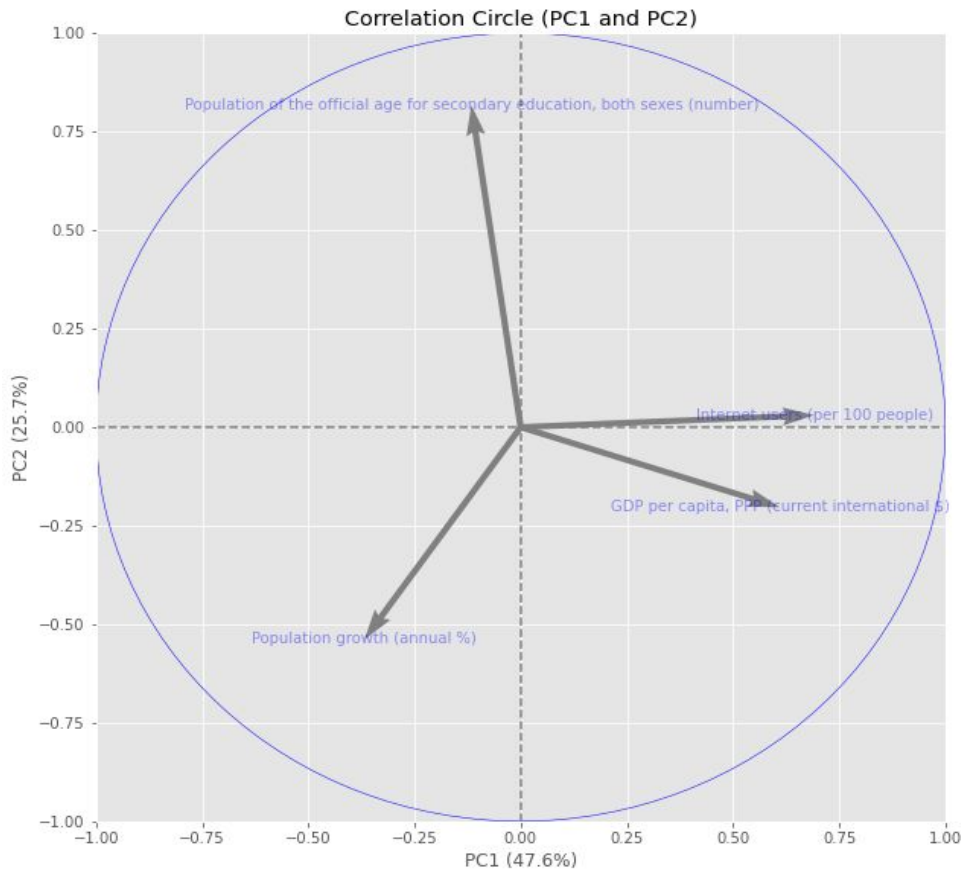


Les pays du cluster 0 :

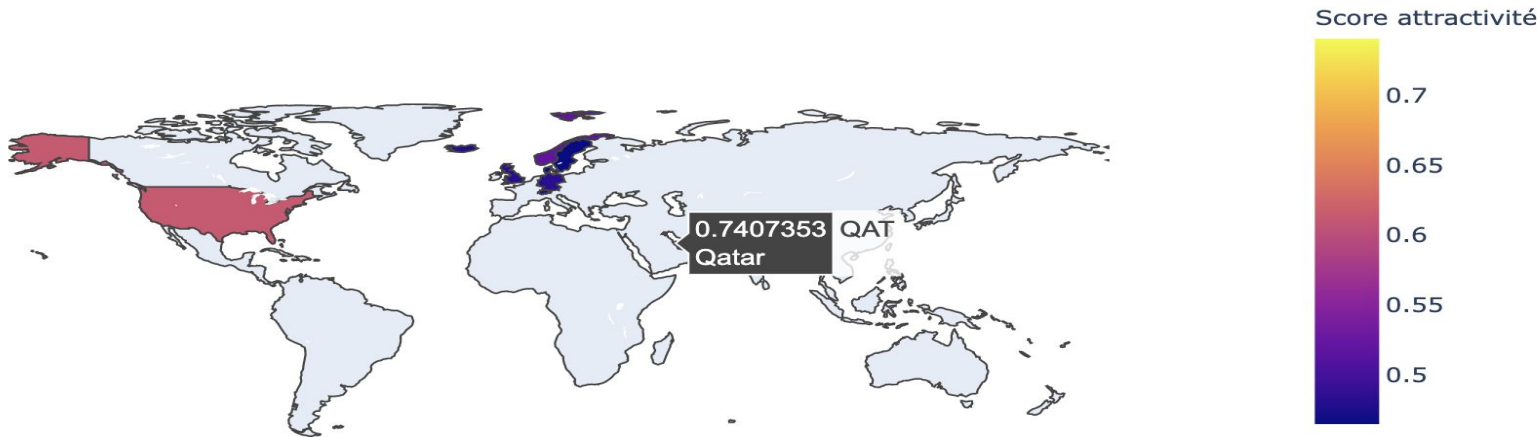
Albania, Antigua and Barbud', Argentina, Australie', Austria, Azerbaijan, Bahamas, The', 'Bahrain', Barbados, Belarus,Belgium, Brazil, Brunei Darussalam, Bulgaria, Canada, Chile, Colombia, Costa Rica, Croatia, Cyprus, Czech Republic, Denmark, Dominica, Estonia, Finland, France, Georgia,Germany, Greece, Grenada, Hong Kong SAR, China, Hungary, Iceland', Ireland, Israel, Italy, Japan, Kazakhstan, Korea, Rep., Kuwait, Latvia, Lithuania, Luxembourg,Macao SAR, China, Macedonia, FYR, Malaysia, Malta, Mauritius, Moldova, Montenegro, Netherlands, New Zealand, Norway, Poland, Portugal, Qatar, Romania, Russian Federation, Saudi Arabia, Serbia, Seychelles, Slovak Republic, Slovenia, Spain, St. Kitts and Nevis, St. Lucia,St. Vincent and the Grenadines, Sweden, Switzerland,Trinidad and Tobago, Turkey,'Ukraine,United Arab Emirates, United Kingdom, United States, Uruguay, Venezuela, RB



ACP Segmentation



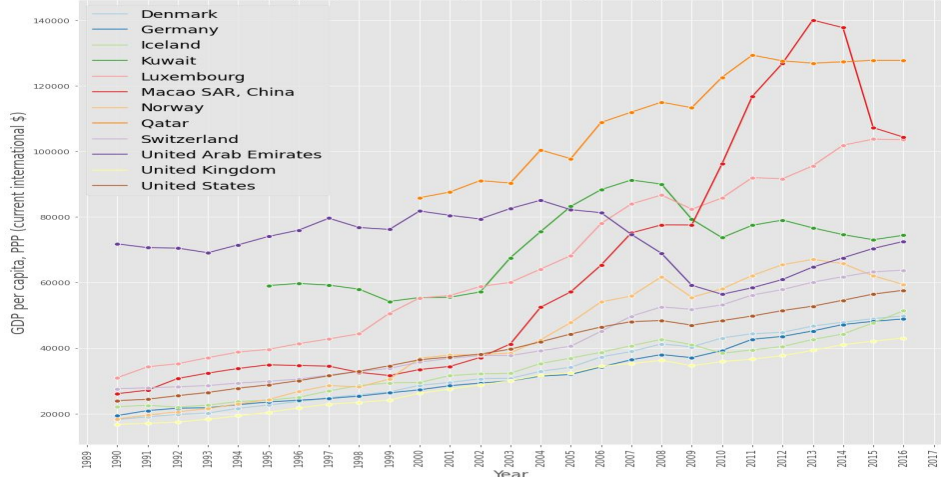
Scoring attractivité des pays



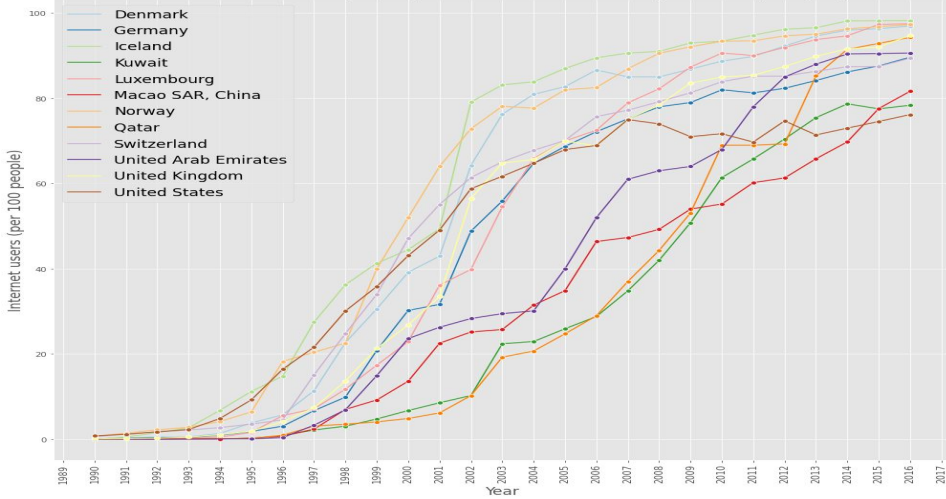
Country Name	GDP per capita, PPP (current international \$)	Internet users (per 100 people)	Population growth (annual %)	Population of the official age for secondary education, both sexes (number)	scoring
Qatar	0.9243234295517652	0.9320497147514262	1.0	0.004250767156242168	0.7407353281023513
Luxembourg	0.7396470591224794	0.9644458027709862	0.43963345392741526	0.0018456123430494647	0.6189729282306305
United States	0.3963972630437947	0.7436837815810922	0.1386106207598397	1.0	0.6129449429714514
Macao SAR, China	1.0	0.710880195599022	0.4145073449480679	0.0013435703512715497	0.5991033127817901
Kuwait	0.5417133891813344	0.8017522412387939	0.9305095504096443	0.01217741748439756	0.5766547225851278
Norway	0.47762818909522825	0.9810513447432763	0.21035013194718533	0.015708542799361663	0.5202568123135536
United Arab Emirates	0.49043484700365436	0.9209453952730238	0.1333204693842987	0.021816116369371478	0.4927539751611749
Switzerland	0.4486253906449939	0.8903830480847597	0.22757886427681978	0.024866581172958216	0.4840124482197349
Germany	0.34265024744630973	0.8780562347188264	0.07775769617304666	0.28308302908010186	0.48326047449755505
United Kingdom	0.2979223657849255	0.9332722086389569	0.1405178768478015	0.20657542129972342	0.48277232313780427
Iceland	0.32155834000970684	1.0	0.20750899605986378	0.001277130722908255	0.47893448173766284
Bahrain	0.3352628325224277	0.9219960880195599	0.2952308385603618	0.0036683130808239506	0.4680355990393576
Denmark	0.3477651421786238	0.9778932355338223	0.09457767947540455	0.017170657542509906	0.465799576999789
Sweden	0.33690844758795657	0.9425427872860636	0.18507294773070984	0.02504536417291763	0.46468783676873693

Evolution des pays cibles

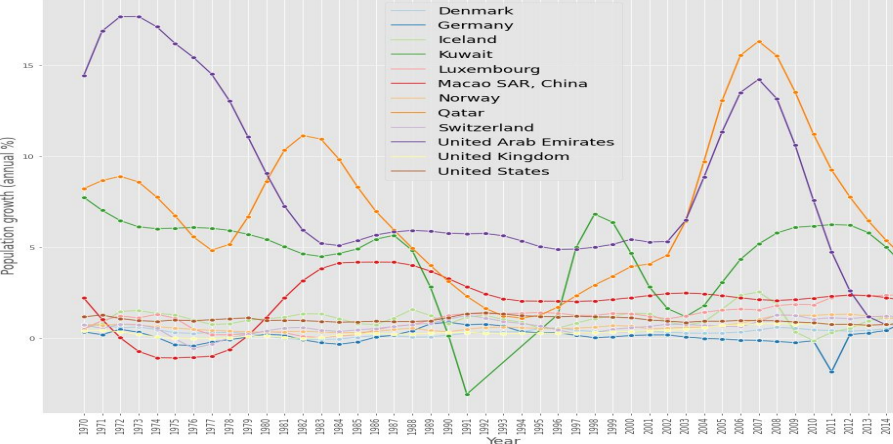
L'Évolution des pays pour GDP per capita, PPP (current international \$)



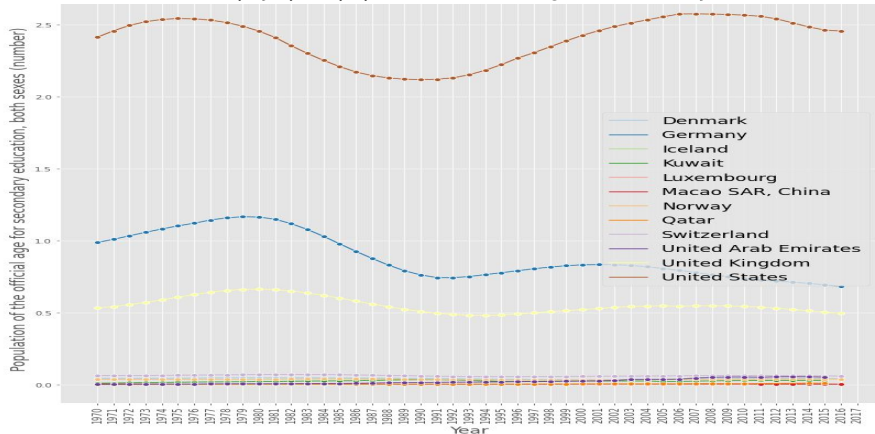
L'Évolution des pays pour Internet users (per 100 people)



L'Évolution des pays pour Population growth (annual %)



L'Évolution des pays pour population of official age for secondary, both sexe



Conclusion

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

- ❑ Cela dépend de la politique de développement
 - Prix des cours en ligne ?
 - Traduction des cours ou animation du cours (langues)

Etat-Unis (Langue anglais, population importante, bon PIB, bonne infrastructure)

Recommandation par groupe :

- **Les pays nordiques** (Norway, Iceland, Danemark, Sweden)

(Bon PIB, Meilleur pays où l'on parle le mieux anglais, bonne infrastructure)

- **Les pays du golfe** (Qatar, United Arab Emirates, Kuwait)

(Excellent PIB, 2eme langue anglais, , bonne infrastructure, croissance growth)

- **Les pays émergents** (Chine, Inde (outliers), Brazil)

(PIB augmentation, Population très importante, modernisation infrastructure, ne parle pas très bien anglais)