Analysez des données de systèmes éducatifs

Félix GANGA

Projet 2

Formation Data Scientist

Du 11/01/2022 Au 10/11/2022

Introduction

Description du projet

Le but de ce projet est d'aider l'entreprise Academy à se développer dans l'international. Nous allons analyser les bases données afin de permettre à Academy de savoir quelle future potentielle clientèle faut-il atteindre, en se basant sur les informations de nos bases des données.

Problématique

- 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 ?

Plan de l’analyse

Etape 1 : Description des bases de données

Etape 2 : Filtrage des Bases de données

Etape 3 : Analyse des 3 indicateurs

Etape 4 : Création des scores

Illustrations by Pixeltrue on icons8

Etape 1 :
Description des données

Etape 1 : Description de EdStatsCountry

-5 lignes du Dataframe
-Dimension du Dataframe (241,42)
-Pas de valeur dupliquée

	Country Code	Short Name	Table Name	Long Name	2-alpha code	Currency Unit	Special Notes	Region	Income Group	WB-2 code	...	IMF data dissemination standard	Latest population census	Latest household survey	Latest
0	ABW	Aruba	Aruba	Aruba	AW	Aruban florin	SNA data for 2000-2011 are updated from offici...	Latin America & Caribbean	High income: nonOECD	AW	...	NaN	2010	NaN	
1	AFG	Afghanistan	Afghanistan	Islamic State of Afghanistan	AF	Afghan afghani	Fiscal year end: March 20; reporting period fo...	South Asia	Low income	AF	...	General Data Dissemination System (GDDS)	1979	Multiple Indicator Cluster Survey (MICS), 2010/11	Integ hous survey
2	AGO	Angola	Angola	People's Republic of Angola	AO	Angolan kwanza	April 2013 database update: Based on IMF data,...	Sub-Saharan Africa	Upper middle income	AO	...	General Data Dissemination System (GDDS)	1970	Malaria Indicator Survey (MIS), 2011	Integ hous survey
3	ALB	Albania	Albania	Republic of Albania	AL	Albanian lek	NaN	Europe & Central Asia	Upper middle income	AL	...	General Data Dissemination System (GDDS)	2011	Demographic and Health Survey (DHS), 2008/09	Stan Measur Study S (I
4	AND	Andorra	Andorra	Principality of Andorra	AD	Euro	NaN	Europe & Central Asia	High income: nonOECD	AD	...	NaN	2011. Population figures compiled from adminis...	NaN	

Valeurs manquantes →

	nb_manquant	pourcentage
Country Code	0	0.000000
Short Name	0	0.000000
Table Name	0	0.000000
Long Name	0	0.000000
2-alpha code	3	1.244813
Currency Unit	26	10.788382
Special Notes	96	39.834025
Region	27	11.203320
Income Group	27	11.203320
WB-2 code	1	0.414938
National accounts base year	36	14.937759
National accounts reference year	209	86.721992
SNA price valuation	44	18.257261
Lending category	97	40.248963
Other groups	183	75.933610
System of National Accounts	26	10.788382
Alternative conversion factor	194	80.497925
PPP survey year	96	39.834025
Balance of Payments Manual in use	60	24.896266
External debt Reporting status	117	48.547718
System of trade	41	17.012448

Typage des colonnes →

```
Country Code                                      object
Short Name                                        object
Table Name                                        object
Long Name                                         object
2-alpha code                                      object
Currency Unit                                     object
Special Notes                                     object
Region                                            object
Income Group                                      object
WB-2 code                                         object
National accounts base year                       object
National accounts reference year                 float64
SNA price valuation                               object
Lending category                                  object
Other groups                                      object
System of National Accounts                       object
Alternative conversion factor                     object
PPP survey year                                   object
Balance of Payments Manual in use                 object
External debt Reporting status                    object
System of trade                                   object
Government Accounting concept                     object
IMF data dissemination standard                   object
Latest population census                          object
Latest household survey                           object
Source of most recent Income and expenditure data object
Vital registration complete                       object
Latest agricultural census                        object
Latest industrial data                           float64
Latest trade data                                float64
Latest water withdrawal data                      object
Unnamed: 31                                      float64
dtype: object
```

Etape 1 : Description de EdStatsData

-5 lignes du Dataframe
-Dimension du Dataframe (886930, 70)
-Pas de valeur dupliquée

	Country Name	Country Code	Indicator Name	Indicator Code	1970	1971	1972	1973	1974	1975	...	2060	2065	2070	2075	2080	2085	2
0	Arab World	ARB	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	N
1	Arab World	ARB	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2.F	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	N
2	Arab World	ARB	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2.GPI	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	N
3	Arab World	ARB	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2.M	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	N
4	Arab World	ARB	Adjusted net enrolment rate, primary, both sex...	SE.PRM.TENR	54.822121	54.894138	56.209438	57.267109	57.991138	59.36554	...	NaN	NaN	NaN	NaN	NaN	NaN	N

5 rows × 70 columns

Valeurs manquantes →

	nb_manquant	pourcentage
Country Name	0	0.000000
Country Code	0	0.000000
Indicator Name	0	0.000000
Indicator Code	0	0.000000
1970	814642	91.849639
...	...	...
2085	835494	94.200670
2090	835494	94.200670
2095	835494	94.200670
2100	835494	94.200670
Unnamed: 69	886930	100.000000

70 rows × 2 columns

Typage des colonnes →

```
Country Name        object
Country Code        object
Indicator Name      object
Indicator Code      object
1970               float64
                       ...
2085               float64
2090               float64
2095               float64
2100               float64
Unnamed: 69        float64
Length: 70, dtype: object
```

Etape 1 : Description de EdStatsSeries

-5 lignes du Dataframe
-Dimension du Dataframe (3665, 21)
-Pas de valeur dupliquée

	Series Code	Topic	Indicator Name	Short definition	Long definition	Unit of measure	Periodicity	Base Period	Other notes	Aggregation method	...	Notes from original source	General comments	Sou
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	NaN	NaN	...	NaN	NaN	Rober Barro a Jong-W L http://www.
1	BAR.NOED.1519.ZS	Attainment	Barro-Lee: Percentage of population age 15-19 ...	Percentage of population age 15-19 with no edu...	Percentage of population age 15-19 with no edu...	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	Rober Barro a Jong-W L http://www.
2	BAR.NOED.15UP.FE.ZS	Attainment	Barro-Lee: Percentage of female population age...	Percentage of female population age 15+ with n...	Percentage of female population age 15+ with n...	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	Rober Barro a Jong-W L http://www.
3	BAR.NOED.15UP.ZS	Attainment	Barro-Lee: Percentage of population age 15+ wi...	Percentage of population age 15+ with no educa...	Percentage of population age 15+ with no educa...	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	Rober Barro a Jong-W L http://www.
4	BAR.NOED.2024.FE.ZS	Attainment	Barro-Lee: Percentage of female population age...	Percentage of female population age 20-24 with...	Percentage of female population age 20-24 with...	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	Rober Barro a Jong-W L http://www.

Valeurs manquantes →

Series Code	0	0.000000
Topic	0	0.000000
Indicator Name	0	0.000000
Short definition	1509	41.173261
Long definition	0	0.000000
Unit of measure	3665	100.000000
Periodicity	3566	97.298772
Base Period	3351	91.432469
Other notes	3113	84.938608
Aggregation method	3618	98.717599
Limitations and exceptions	3651	99.618008
Notes from original source	3665	100.000000
General comments	3651	99.618008
Source	0	0.000000
Statistical concept and methodology	3642	99.372442
Development relevance	3662	99.918145
Related source links	3450	94.133697
Other web links	3665	100.000000
Related indicators	3665	100.000000
License Type	3665	100.000000
Unnamed: 20	3665	100.000000

Typage des colonnes →

```
Series Code                          object
Topic                                object
Indicator Name                       object
Short definition                     object
Long definition                      object
Unit of measure                     float64
Periodicity                          object
Base Period                          object
Other notes                          object
Aggregation method                   object
Limitations and exceptions           object
Notes from original source          float64
General comments                     object
Source                               object
Statistical concept and methodology  object
Development relevance                object
Related source links                 object
Other web links                     float64
Related indicators                  float64
License Type                        float64
Unnamed: 20                         float64
dtype: object
```

Etape 1 : Description de EdStatsFootNotes

-5 lignes du Dataframe
-Dimension du Dataframe (643638, 5)
-Pas de valeur dupliquée

	CountryCode	SeriesCode	Year	DESCRIPTION	Unnamed: 4
0	ABW	SE.PRE.ENRL.FE	YR2001	Country estimation.	NaN
1	ABW	SE.TER.TCHR.FE	YR2005	Country estimation.	NaN
2	ABW	SE.PRE.TCHR.FE	YR2000	Country estimation.	NaN
3	ABW	SE.SEC.ENRL.GC	YR2004	Country estimation.	NaN
4	ABW	SE.PRE.TCHR	YR2006	Country estimation.	NaN

Valeurs manquantes →

	nb_manquant	pourcentage
CountryCode	0	0.0
SeriesCode	0	0.0
Year	0	0.0
DESCRIPTION	0	0.0
Unnamed: 4	643638	100.0

Typage des colonnes →

```
CountryCode      object
SeriesCode       object
Year             object
DESCRIPTION      object
Unnamed: 4      float64
dtype: object
```

Etape 1 : Description de EdStatsFootNotesEdStatsCountry-Series

-5 lignes du Dataframe
-Dimension du Dataframe (613, 4)
-Pas de valeur dupliquée

Valeurs manquantes →

	nb_manquant	pourcentage
CountryCode	0	0.0
SeriesCode	0	0.0
DESCRIPTION	0	0.0
Unnamed: 3	613	100.0

	CountryCode	SeriesCode	DESCRIPTION	Unnamed: 3
0	ABW	SP.POP.TOTL	Data sources : United Nations World Population...	NaN
1	ABW	SP.POP.GROW	Data sources: United Nations World Population ...	NaN
2	AFG	SP.POP.GROW	Data sources: United Nations World Population ...	NaN
3	AFG	NY.GDP.PCAP.PP.CD	Estimates are based on regression.	NaN
4	AFG	SP.POP.TOTL	Data sources : United Nations World Population...	NaN

Typage des colonnes →

```
CountryCode      object
SeriesCode       object
DESCRIPTION      object
Unnamed: 3      float64
dtype: object
```

Etape 1 : Description

Nous allons nous concentrer sur EdStatsCountry et EdStatsData qui les Dataframe principales pour l’analyse.

Comme constaté les autres Dataframe : EdStatsSeries ; EdStatsFootNote et EdStatsCountry-Series constituent des compléments d’information.

Etape 2 :
Filtrage des données et sélection des indicateurs pertinents

Etape 2 :

Filtrage dans EdStatsData pour avoir les années de recensement ≥ 2010

	Country Code	Short Name	Table Name	Long Name	2-alpha code	Currency Unit	Special Notes	Region	Income Group	WB-2 code	...	IMF data dissemination standard	Latest population census	Latest household survey	Source of most recent Income and expenditure data
0	ABW	Aruba	Aruba	Aruba	AW	Aruban florin	SNA data for 2000-2011 are updated from offici...	Latin America & Caribbean	High income: nonOECD	AW	...	NaN	2010	NaN	NaN
3	ALB	Albania	Albania	Republic of Albania	AL	Albanian lek	NaN	Europe & Central Asia	Upper middle income	AL	...	General Data Dissemination System (GDDS)	2011	Demographic and Health Survey (DHS), 2008/09	Living Standards Measurement Study Survey (LSM...
4	AND	Andorra	Andorra	Principality of Andorra	AD	Euro	NaN	Europe & Central Asia	High income: nonOECD	AD	...	NaN	2011. Population figures compiled from adminis...	NaN	NaN
6	ARE	United Arab Emirates	United Arab Emirates	United Arab Emirates	AE	U.A.E. dirham	April 2013 database update: Based on data from...	Middle East & North Africa	High income: nonOECD	AE	...	General Data Dissemination System (GDDS)	2010	NaN	NaN
7	ARG	Argentina	Argentina	Argentine Republic	AR	Argentine peso	NaN	Latin America & Caribbean	Upper middle income	AR	...	Special Data Dissemination Standard (SDDS)	2010	Multiple Indicator Cluster Survey (MICS), 2011/12	Integrated household survey (IHS), 2012

Ensuite

Jointure entre ed_stats_country et ed_stats_data pour avoir les valeurs de chaque indicateur par date de 1970 à 2100

	Country Code	Short Name	Table Name	Long Name	2-alpha code	Currency Unit	Special Notes	Region	Income Group	WB-2 code	...	2060	2065	2070	2075	2080	2085	2090	2095	2100	Unname 6
0	ABW	Aruba	Aruba	Aruba	AW	Aruban florin	SNA data for 2000-2011 are updated from offici...	Latin America & Caribbean	High income: nonOECD	AW	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
1	ABW	Aruba	Aruba	Aruba	AW	Aruban florin	SNA data for 2000-2011 are updated from offici...	Latin America & Caribbean	High income: nonOECD	AW	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
2	ABW	Aruba	Aruba	Aruba	AW	Aruban florin	SNA data for 2000-2011 are updated from offici...	Latin America & Caribbean	High income: nonOECD	AW	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
3	ABW	Aruba	Aruba	Aruba	AW	Aruban florin	SNA data for 2000-2011 are updated from offici...	Latin America & Caribbean	High income: nonOECD	AW	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
4	ABW	Aruba	Aruba	Aruba	AW	Aruban florin	SNA data for 2000-2011 are	Latin America &	High income:	AW	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na

Etape 2 :

Recherche des indicateurs potentiellement intéressantes pour le besoin de notre étude :

- Démographie : population totale, par tranches d'âge 20-29 ans, voire autres tranches d'âge :
- Niveau éducation : nombre et taux étudiants dans le secondaire et tertiaire
- Déploiement Internet : Taux utilisation internet
- Économique : PIB, PIB / habitant, niveau de revenu

	Indicator Code	Indicator Name	2010	2011	2012	2013	2014	2015	2016	2017	2020
0	BAR.SCHL.1519	Barro-Lee: Average years of total schooling, a...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	BAR.SCHL.2024	Barro-Lee: Average years of total schooling, a...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	BAR.NOED.1519.ZS	Barro-Lee: Percentage of population age 15-19 ...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	BAR.NOED.2024.ZS	Barro-Lee: Percentage of population age 20-24 ...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NY.GDP.PCAP.KD	GDP per capita (constant 2005 US$)	24271.940421	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...
2219	UIS.XUNIT.GDPCAP3.FSGOV	Government expenditure per upper secondary stu...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2220	IT.NET.USER.P2	Internet users (per 100 people)	6.400000	8.400000	12.000000	15.500000	16.364740	22.742818	23.119989	NaN	NaN
2221	IT.CMP.PCMP.P2	Personal computers (per 100 people)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2222	SP.POP.GROW	Population growth (annual %)	1.976756	2.109672	2.228306	2.309366	2.344799	2.345643	2.336070	NaN	NaN
2223	SL.UEM.NEET.ZS	Share of youth not in education, employment, o...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

2224 rows × 11 columns

Problème de remplissage pour l'année 2017 et 2020 donc supprimer

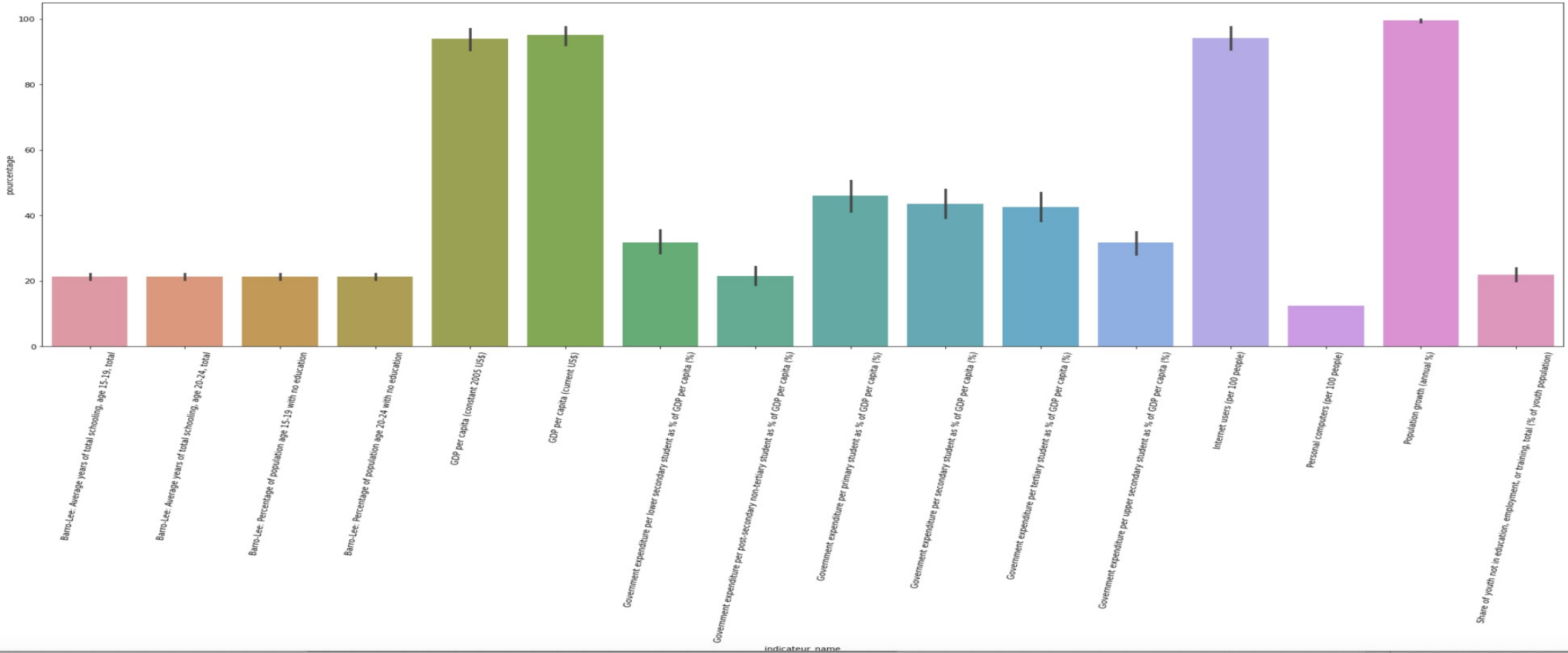
	nb_manquant	pourcentage
Indicator Code	0	0.000000
Indicator Name	0	0.000000
2010	885	39.793165
2011	1297	58.318345
2012	1360	61.151079
2013	1398	62.859712
2014	1511	67.940647
2015	1633	73.426259
2016	1704	76.618705
2017	2224	100.000000
2020	2224	100.000000

Transposition de la Dataframe pour avoir les année de 2010 à 2016 en ligne et les indicateurs choisis en colonne

Indicator Code	BAR.SCHL.1519	BAR.SCHL.2024	BAR.NOED.1519.ZS	BAR.NOED.2024.ZS	NY.GDP.PCAP.KD	NY.GDP.PCAP.CD	UIS.XUNIT.GDPCAP2.FSGOV	UIS.XUNIT
Indicator Name	Barro-Lee: Average years of total schooling, a...	Barro-Lee: Average years of total schooling, a...	Barro-Lee: Percentage of population age 15-19 ...	Barro-Lee: Percentage of population age 20-24 ...	GDP per capita (constant 2005 US$)	GDP per capita (current US$)	Government expenditure per lower secondary stu...	Governm pc
2010	NaN	NaN	NaN	NaN	24271.940421	24271.940421	NaN	
2011	NaN	NaN	NaN	NaN	NaN	25324.720362	NaN	
2012	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2013	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2014	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2015	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2016	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

Etape 2 :

Diagramme en bâton du taux de remplissage par indicateurs



Etape 2 :

Nous avons donc décider de garder par l’analyse du diagramme en bâton trois indicateur :
-GDP per capita (current US$)
-internet users (per 100 people)
-Population growth (annual%)

Puis finalement création d’une Dataframe avec seulement les indicateur choisis

Etape 3 :
Analyse des 3 indicateurs sélectionnés

Etape 3 :
Filtrage PIB/Habitant

Dataframe en fonction de GDP

Country Code	ABW	ALB	AND	ARE	ARG	ARM	ASM	ATG	AUS	AUT
2010	24271.940421	4094.358832	39736.354063	35049.148317	10276.260498	3218.381655	10352.822762	12174.697859	51874.080482	46858.043273
2011	25324.720362	4437.178068	41098.766942	40462.312089	12726.908359	3526.978143	10375.994215	11931.204107	62245.100292	51374.958407
2012	NaN	4247.614308	38391.080867	42112.235647	12969.707124	3684.80481	11660.329531	12517.557929	67677.634766	48567.695286
2013	NaN	4413.081697	40619.711298	43350.642681	12976.636425	3843.591213	11589.853002	12194.608761	67792.303862	50719.388277
2014	NaN	4578.666728	42294.994727	44449.740349	12245.256449	3994.712355	11598.751736	12946.986936	62214.609121	51733.481917
2015	NaN	3934.895394	36038.267604	39101.74689	13467.102357	3617.935746	11865.963232	13659.147914	56554.038761	44255.583356
2016	NaN	4124.98239	36988.62203	37622.207456	12440.320982	3614.688357	11834.745229	14462.176279	49927.819509	44676.345835

Valeurs Manquantes

Country Code	nb_manquant	pourcentage
GIB	7	100.0
CYM	7	100.0
CHI	7	100.0
SXM	7	100.0
TCA	7	100.0
...	...	...
FIN	0	0.0
EST	0	0.0
ESP	0	0.0
ECU	0	0.0
ZWE	0	0.0

Filtrage des valeurs manquantes avec un remplissage de plus de 60 %

Country Code	2010	2011	2012	2013	2014	2015	2016	nb_manquant
PRI	26435.740187	27278.874895	27811.548147	28703.748347	NaN	NaN	NaN	3
BMU	88207.32756	85973.158416	85458.455508	85748.065414	NaN	NaN	NaN	3
VEN	13545.205838	10741.576384	12755.000082	12237.224064	15692.412878	NaN	NaN	2
CUB	5676.141403	6075.924306	6425.941119	6760.15344	7050.50199	7602.261218	NaN	1
LIE	141165.082857	158283.07626	149295.646044	173528.150454	179308.075616	168146.015281	NaN	1
...	...	...	...	...	...	...	...	...
FIN	46202.415163	50790.724145	47415.559871	49638.07713	49914.618642	42419.565567	43402.863264	0
EST	14638.604817	17454.843425	17421.890223	19072.238518	19949.581377	17155.874176	17727.492929	0
ESP	30736.627853	31835.288801	28563.859038	29211.773746	29623.164445	25787.928792	26639.740588	0
ECU	4657.302361	5223.351763	5702.104313	6074.090829	6396.630483	6150.155955	6018.526868	0
ZWE	719.979517	840.94988	968.163871	1026.388289	1031.104616	1033.415842	1029.076649	0

Country Code	ALB	AND	ARE	ARG	ARM	ASM	ATG	AUS	AUT	BEL
2010	4094.358832	39736.354063	35049.148317	10276.260498	3218.381655	10352.822762	12174.697859	51874.080482	46858.043273	44380.176633
2011	4437.178068	41098.766942	40462.312089	12726.908359	3526.978143	10375.994215	11931.204107	62245.100292	51374.958407	47702.811894
2012	4247.614308	38391.080867	42112.235647	12969.707124	3684.804810	11660.329531	12517.557929	67677.634766	48567.695286	44740.583248
2013	4413.081697	40619.711298	43350.642681	12976.636425	3843.591213	11589.853002	12194.608761	67792.303862	50719.388277	46582.669550
2014	4578.666728	42294.994727	44449.740349	12245.256449	3994.712355	11598.751736	12946.986936	62214.609121	51733.481917	47379.173917
2015	3934.895394	36038.267604	39101.746890	13467.102357	3617.935746	11865.963232	13659.147914	56554.038761	44255.583356	40375.388639
2016	4124.982390	36988.622030	37622.207456	12440.320982	3614.688357	11834.745229	14462.176279	49927.819509	44676.345835	41236.266589

Remplissage des valeurs manquantes par la médiane

Etape 3 : Filtrage utilisateur internet

Dataframe en fonction de Net

Country Code	ABW	ALB	AND	ARE	ARG	ARM	ASM	ATG	AUS	AUT	...	URY	USA	VCT	VEN
2010	62.0	45.0	81.0	68.0	45.0	25.0	NaN	47.0	76.0	75.17	...	46.4	71.69	33.7	37.37
2011	69.0	49.0	81.0	78.0	51.0	32.0	NaN	52.0	79.487698	78.739993	...	51.404661	69.729461	36.7	40.22
2012	74.0	54.655959	86.434425	84.999991	55.8	37.5	NaN	58.0	79.0	80.029994	...	54.453769	74.7	40.0	49.050083
2013	78.9	57.2	94.0	88.0	59.9	41.9	NaN	63.4	83.453497	80.6188	...	57.69	71.4	43.5	54.9
2014	83.78	60.1	95.9	90.4	64.7	54.622806	NaN	67.78	84.0	80.995825	...	61.46	73.0	47.4	57.0
2016	93.542454	66.363445	97.930637	90.600007	70.150764	62.0	NaN	73.0	88.238658	84.323331	...	66.4	76.176737	55.574619	60.0

Valeurs Manquantes

Country Code	nb_manquant	pourcentage
SXM	6	100.0
XKX	6	100.0
ASM	6	100.0
PLW	6	100.0
CHI	6	100.0
...	...	...
EST	0	0.0
ESP	0	0.0
ECU	0	0.0
DOM	0	0.0
ZWE	0	0.0

Filtrage des valeurs manquantes avec un remplissage de plus de 60 %

Country Code	2010	2011	2012	2013	2014	2016	nb_manquant
GIB	65.0	65.0	65.017	NaN	NaN	94.444472	2
ABW	62.0	69.0	74.0	78.9	83.78	93.542454	0
PRY	19.8	24.763516	29.34	36.9	43.0	51.349874	0
PRI	45.3	48.0	68.999983	68.99999	76.13386	80.32	0
POL	62.32	61.949999	62.309997	62.8492	66.6	73.3007	0
...	...	...	...	...	...	...	...
ESP	65.8	67.089999	69.81	71.635	76.19	80.561333	0
ECU	29.03	31.366808	35.135146	40.279122	45.590391	54.062925	0
DOM	31.4	38.0	42.32	45.9	49.58	61.32819	0
DNK	88.72	89.810013	92.260012	94.6297	95.99	96.967785	0
ZWE	6.4	8.4	12.0	15.5	16.36474	23.119989	0

Remplissage des valeurs manquantes par la médiane

Country Code	ABW	ALB	AND	ARE	ARG	ARM	ATG	AUS	AUT	BEL	...	TZA	URY	USA
2010	62.000000	45.000000	81.000000	68.000000	45.000000	25.000000	47.00	76.000000	75.170000	75.000000	...	2.90	46.400000	71.690000
2011	69.000000	49.000000	81.000000	78.000000	51.000000	32.000000	52.00	79.487698	78.739993	81.609996	...	3.20	51.404661	69.729461
2012	74.000000	54.655959	86.434425	84.999991	55.800000	37.500000	58.00	79.000000	80.029994	80.719991	...	3.95	54.453769	74.700000
2013	78.900000	57.200000	94.000000	88.000000	59.900000	41.900000	63.40	83.453497	80.618800	82.170200	...	4.40	57.690000	71.400000
2014	83.780000	60.100000	95.900000	90.400000	64.700000	54.622806	67.78	84.000000	80.995825	85.000000	...	7.00	61.460000	73.000000
2016	93.542454	66.363445	97.930637	90.600007	70.150764	62.000000	73.00	88.238658	84.323331	86.516500	...	13.00	66.400000	76.176737

Etape 3 :
Filtrage croissance de la population

Dataframe en fonction de Growth

Country Code	ABW	ALB	AND	ARE	ARG	ARM	ASM	ATG	AUS	AUT	...	URY	USA	VCT
2010	0.21268	-0.496462	-0.015393	7.587098	1.035037	-0.391024	-1.054862	1.147472	1.55549	0.240394	...	0.34614	0.836422	0.056733
2011	0.376985	-0.269017	-0.829969	4.7437	1.04486	-0.060144	-0.571394	1.111473	1.389527	0.337081	...	0.331626	0.745614	0.023782
2012	0.512145	-0.165151	-1.588653	2.594796	1.050376	0.220269	-0.162822	1.099255	1.722895	0.455937	...	0.328881	0.74642	-0.01189
2013	0.592914	-0.183211	-2.013314	1.181805	1.047277	0.401252	0.13932	1.076058	1.697473	0.584104	...	0.330004	0.700262	-0.007318
2014	0.587492	-0.207047	-1.956178	0.714763	1.032709	0.438332	0.234776	1.068648	1.474288	0.730867	...	0.338072	0.743124	0.03384
2015	0.524658	-0.291206	-1.537836	0.915608	1.009855	0.368528	0.180222	1.054346	1.391107	1.066623	...	0.350484	0.729732	0.089575
2016	0.459929	-0.15988	-0.944017	1.251759	0.984742	0.269302	0.111575	1.035422	1.410064	1.314007	...	0.362269	0.692801	0.171613

Valeurs Manquantes

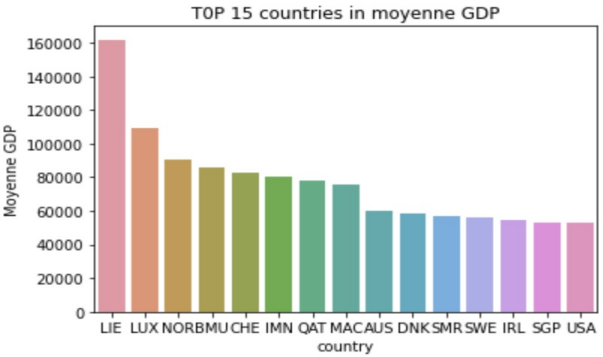
Country Code	nb_manquant	pourcentage
SXM	3	42.857143
HRV	1	14.285714
ABW	0	0.000000
MYS	0	0.000000
NER	0	0.000000
...	...	...
ESP	0	0.000000
ECU	0	0.000000
DOM	0	0.000000
DNK	0	0.000000
ZWE	0	0.000000

Filtrage inutile des valeurs manquantes avec un remplissage de plus de 60 %

Country Code	ABW	ALB	AND	ARE	ARG	ARM	ASM	ATG	AUS	AUT	...	URY	USA	VCT
2010	0.212680	-0.496462	-0.015393	7.587098	1.035037	-0.391024	-1.054862	1.147472	1.555490	0.240394	...	0.346140	0.836422	0.056733
2011	0.376985	-0.269017	-0.829969	4.743700	1.044860	-0.060144	-0.571394	1.111473	1.389527	0.337081	...	0.331626	0.745614	0.023782
2012	0.512145	-0.165151	-1.588653	2.594796	1.050376	0.220269	-0.162822	1.099255	1.722895	0.455937	...	0.328881	0.746420	-0.011890
2013	0.592914	-0.183211	-2.013314	1.181805	1.047277	0.401252	0.139320	1.076058	1.697473	0.584104	...	0.330004	0.700262	-0.007318
2014	0.587492	-0.207047	-1.956178	0.714763	1.032709	0.438332	0.234776	1.068648	1.474288	0.730867	...	0.338072	0.743124	0.033840
2015	0.524658	-0.291206	-1.537836	0.915608	1.009855	0.368528	0.180222	1.054346	1.391107	1.066623	...	0.350484	0.729732	0.089575
2016	0.459929	-0.159880	-0.944017	1.251759	0.984742	0.269302	0.111575	1.035422	1.410064	1.314007	...	0.362269	0.692801	0.171613

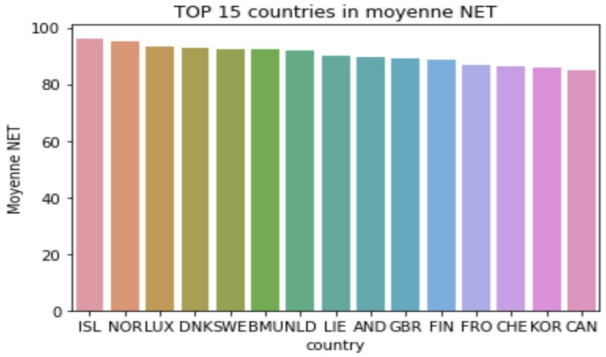
Remplissage des valeurs manquantes par la médiane

Etape 3 :



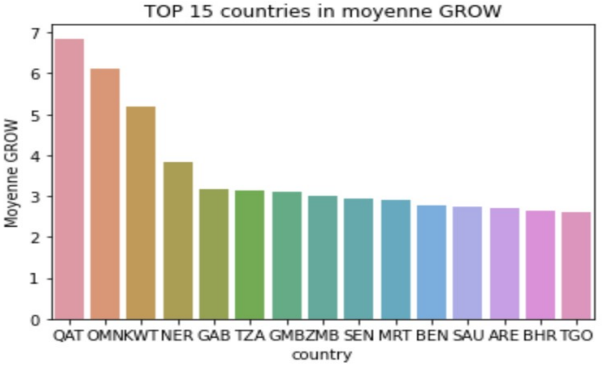
Déduction top5
PIB/Habitant

Country Code	Moyenne	Country Code
LIE	161848.656040	LIE
LUX	108906.609516	LUX
NOR	90834.593814	NOR
BMU	86138.406092	BMU
CHE	82883.450322	CHE



Déduction top5
Utilisateur
Internet

Country Code	Moyenne	Country Code
ISL	96.227717	ISL
NOR	95.030261	NOR
LUX	93.090035	LUX
DNK	93.062918	DNK
SWE	92.460067	SWE



Déduction top5
Croissance de la
population

Country Code	Moyenne	Country Code
QAT	6.851502	QAT
OMN	6.120074	OMN
KWT	5.185630	KWT
NER	3.827682	NER
GAB	3.161405	GAB

Etape 4 :
Scoring

Etape 4

L’union des TOP 5 des 3 indicateurs pour avoir l’échantillon des pays

	Country Code
0	BMU
1	CHE
2	DNK
3	GAB
4	ISL
5	KWT
6	LIE
7	LUX
8	NER
9	NOR
10	OMN
11	QAT
12	SWE

Etape 4

Country Code	Country Code	Score GDP	Score NET	Score GROW	Note Score
LUX	LUX	1	1	0	0.666667
NOR	NOR	1	1	0	0.666667
BMU	BMU	1	0	0	0.333333
CHE	CHE	1	0	0	0.333333
DNK	DNK	0	1	0	0.333333
GAB	GAB	0	0	1	0.333333
ISL	ISL	0	1	0	0.333333
KWT	KWT	0	0	1	0.333333
LIE	LIE	1	0	0	0.333333
NER	NER	0	0	1	0.333333
OMN	OMN	0	0	1	0.333333
QAT	QAT	0	0	1	0.333333
SWE	SWE	0	1	0	0.333333

Conclusion

Après avoir sélectionné les 3 indicateurs les plus pertinent.

Nous constatons que les 2 pays qui ont le meilleur score sont le Luxembourg et la Norvège avec une note de 2/3

Il semblerait que ce soit ces pays que l’entreprise Academy doit opérer en priorité