



academy

Formation en ligne

Etude de marché

Sommaire

1	La mission.	2
2	Les Données.	3
	2.1	EdStatsCountry.csv
	2.2	EdStatsCountry-Series.csv
	2.3	EdStatsData.csv
	2.4	EdStatsFootNote.csv
	2.5	EdStatsSeries.csv
3	Sélection des informations pertinentes.	6
	3.1	Les Critères de sélection
	3.2	Création d'un dataset
	3.2.1	Données par régions
	3.2.2	Données par pays
4	Indicateurs statistiques.	10
	4.1	Indicateurs par régions
	4.2	Indicateurs par pays
5	Graphiques.	16-17
6	Conclusion.	18

1

1. La mission :

Analyse pré-exploratoire d'un jeu de données sur l'éducation de la banque mondiale afin de permettre de déterminer si ces données sont suffisantes pour un projet d'expansion.

Le jeu de données fourni se compose de 5 datasets :

- EdStatsCountry.csv
- EdStatsCountry-Series.csv
- EdStatsData.csv
- EdStatsFootNote.csv
- EdStatsSeries.csv

2

2. Les Données.

2.1 EdStatsCountry.csv

chaque ligne contient des variables relatives a un pays unique.

nb de lignes x nb de colonnes : (241, 32) .

% de données manquantes :

Unnamed: 31	100.000000
National accounts reference year	86.721992
Alternative conversion factor	80.497925
Other groups	75.933610
Latest industrial data	55.601660
Vital registration complete	53.941909
External debt Reporting status	48.547718
Latest household survey	41.493776
Latest agricultural census	41.078838
Lending category	40.248963
PPP survey year	39.834025
Special Notes	39.834025
Source of most recent Income and expenditure data	33.609959
Government Accounting concept	33.195021
Latest water withdrawal data	25.726141
IMF data dissemination standard	24.896266
Balance of Payments Manual in use	24.896266
Latest trade data	23.236515
SNA price valuation	18.257261
System of trade	17.012448
National accounts base year	14.937759
Latest population census	11.618257
Region	11.203320
Income Group	11.203320
Currency Unit	10.788382
System of National Accounts	10.788382
2-alpha code	1.244813
WB-2 code	0.414938
Long Name	0.000000
Short Name	0.000000
Table Name	0.000000
Country Code	0.000000
dtype: float64	

nb de lignes dupliquées : 0 .

2

2.2 EdStatsCountry-Series.csv

Chaque ligne contient des informations relatives aux sources des données ainsi que leur codification.

```
nb de lignes x nb de colonnes : (613, 4) .

% de données manquantes :
  Unnamed: 3      100.0
CountryCode      0.0
SeriesCode       0.0
DESCRIPTION      0.0
dtype: float64

nb de lignes dupliquées : 0 .
```

2.3 EdStatsData.csv

Dataset principal contenant les informations relatives à l'éducation.

```
nb de lignes x nb de colonnes : (886930, 70) .

% de données manquantes :
  Unnamed: 69      100.000000
2017              99.983877
2016              98.144160
1971              95.993258
1973              95.992356
...
2010              72.665036
Country Code      0.000000
Indicator Code    0.000000
Indicator Name    0.000000
Country Name      0.000000
Length: 70, dtype: float64

nb de lignes dupliquées : 0 .
```

2

2.4 EdStatsFootNote.csv

Informations complémentaires relatives aux sources et aux pays.

nb de lignes x nb de colonnes : (643638, 5) .

% de données manquantes :

Unnamed: 4	100.0
CountryCode	0.0
SeriesCode	0.0
Year	0.0
DESCRIPTION	0.0
dtype: float64	

nb de lignes dupliquées : 0 .

nb de lignes x nb de colonnes : (3665, 21) .

% de données manquantes :

Unnamed: 20	100.000000
Notes from original source	100.000000
License Type	100.000000
Related indicators	100.000000
Other web links	100.000000
Unit of measure	100.000000
Development relevance	99.918145
General comments	99.618008
Limitations and exceptions	99.618008
Statistical concept and methodology	99.372442
Aggregation method	98.717599
Periodicity	97.298772
Related source links	94.133697
Base Period	91.432469
Other notes	84.938608
Short definition	41.173261
Topic	0.000000
Source	0.000000
Long definition	0.000000
Indicator Name	0.000000
Series Code	0.000000
dtype: float64	

nb de lignes dupliquées : 0 .

2.5 EdStatsSeries.csv

Dataset contenant les informations détaillés relatives a chacun des indicateurs d'éducation contenues dans EdStatsData.csv.

3

3. Sélection des informations pertinentes.

3.1 Critères de sélection

- Tranches d'âges
- Niveau d'études achevé ou non
- Financement public lié à l'éducation
- Pib par habitant
- Disponibilité internet
- Disponibilité ordinateur
-

sur

<https://datatopics.worldbank.org/education/indicators>

Et vérification leur disponibilité sur le dataset principal EdStatsData.csv .

3

3.2 Création d'un dataset.

A partir des datasets EdStatsData.csv et EdstatCountry.csv, création de 'df' contenant les données présélectionnées

- Colonne 'Region'
- Colonne 'Year'
- Colonne 'Country Name'.
- 22 colonnes représentants les 22 indicateurs pertinents sélectionnés.

	Year	Region	Country Name	Barro-Lee: Percentage of population age 15-19 with secondary schooling. Completed Secondary	Barro-Lee: Percentage of population age 15-19 with secondary schooling. Total (Incomplete and Completed Secondary)	Barro-Lee: Percentage of population age 20-24 with tertiary schooling. Completed Tertiary	Barro-Lee: Percentage of population age 20-24 with tertiary schooling. Total (Incomplete and Completed Tertiary)	Barro-Lee: Percentage of population age 25+ with tertiary schooling. Completed Tertiary	Barro-Lee: Percentage of population age 25+ with tertiary schooling. Total (Incomplete and Completed Tertiary)	Barro-Lee: Percentage of population age 25-29 with tertiary schooling. Completed Tertiary	...	Expenditure on education as % of total government expenditure (%)	GDP per capita (current US\$)
0	1970	NaN	Arab World	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	250.488643
1	1970	NaN	East Asia & Pacific	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	316.225712
2	1970	NaN	East Asia & Pacific (excluding high income)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	114.138515
3	1970	NaN	Euro area	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	2231.949573
4	1970	NaN	Europe & Central Asia	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	1365.643199
...
11369	2016	Latin America & Caribbean	Virgin Islands (U.S.)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN
11370	2016	Middle East & North Africa	West Bank and Gaza	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	2943.404534
11371	2016	Middle East & North Africa	Yemen, Rep.	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	990.334774
11372	2016	Sub-Saharan Africa	Zambia	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	1269.573537
11373	2016	Sub-Saharan Africa	Zimbabwe	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	1029.076649

11374 rows × 25 columns

3

3.2.1 Données par régions.

Pourcentage des régions disponibles par variable et par années (6 dernières années sur les 7 régions contenues dans 'Country Name').

	GDP per capita (current US\$)	Internet users (per 100 people)	Population growth (annual %)	Unemployment, total (% of total labor force)
Year				
2011	100.0	100.0	100.0	100.0
2012	100.0	100.0	100.0	100.0
2013	100.0	100.0	100.0	100.0
2014	100.0	100.0	100.0	100.0
2015	100.0	100.0	100.0	100.0
2016	100.0	100.0	100.0	100.0

4 variables complètes

3

3.2.1 Données par régions.

Pourcentage des régions disponibles par variable et par années (6 dernières années sur les 7 régions contenues dans la colonne 'Region').

	Expenditure on education as % of total government expenditure (%)	GDP per capita (current US\$)	Government expenditure on education as % of GDP (%)	Internet users (per 100 people)	MICS: Secondary completion rate	Population growth (annual %)	Population, ages 14-19, total	Population, ages 15-24, total	Unemployment, total (% of total labor force)
Year									
2011	100.0	100.0	100.00	100.0	28.57	100.0	100.0	100.0	100.0
2012	100.0	100.0	100.00	100.0	0.00	100.0	100.0	100.0	100.0
2013	100.0	100.0	100.00	100.0	0.00	100.0	100.0	100.0	100.0
2014	100.0	100.0	100.00	100.0	42.88	100.0	100.0	100.0	100.0
2015	100.0	100.0	100.00	100.0	0.00	100.0	100.0	100.0	100.0
2016	0.0	100.0	42.88	100.0	0.00	100.0	0.0	0.0	100.0

4 variables complètes + 5 variables incomplètes

3

3.2.2 Données par pays.

Pourcentage de pays disponibles par variable et par années (6 dernières années des 217 pays contenue dans 'Country Name').

Year	Expenditure on education as % of total government expenditure (%)	GDP per capita (current US\$)	Government expenditure on education as % of GDP (%)	Internet users (per 100 people)	MICS: Secondary completion rate	Population growth (annual %)	Population, ages 14-19, total	Population, ages 15-24, total	Unemployment, total (% of total labor force)
2011	54.38	93.66	56.78	93.55	0.92	98.62	83.41	83.41	85.71
2012	51.15	91.71	52.07	92.63	0.00	98.62	83.41	83.41	85.71
2013	53.00	92.17	54.38	92.63	0.00	98.62	83.41	83.41	85.71
2014	40.65	90.78	41.01	92.63	1.84	96.77	83.41	83.41	85.71
2015	18.43	90.32	20.74	92.63	0.00	96.77	83.41	83.41	85.71
2016	0.00	87.10	1.84	92.63	0.00	96.77	0.00	0.00	85.71

9 variables incomplètes

4

4. Indicateurs statistiques.

4.1 Indicateurs par régions ('Country Name').

Moyennes

Country Name	GDP per capita (current US\$)	Internet users (per 100 people)	Personal computers (per 100 people)	Population growth (annual %)	Unemployment, total (% of total labor force)
East Asia & Pacific	3654.092315	18.813058	3.890438	1.293952	4.439740
Europe & Central Asia	11922.191981	31.121074	13.904042	0.468166	9.073022
Latin America & Caribbean	4007.676980	19.603180	3.608141	1.765886	8.776970
Middle East & North Africa	3261.098802	15.266437	3.749321	2.526329	12.023846
North America	28021.787082	47.380199	31.406071	1.013922	6.261544
South Asia	535.120983	5.470289	0.622202	1.998757	4.169711
Sub-Saharan Africa	758.605457	5.388914	1.050565	2.777192	7.846521

Ecart-types

Country Name	GDP per capita (current US\$)	Internet users (per 100 people)	Personal computers (per 100 people)	Population growth (annual %)	Unemployment, total (% of total labor force)
East Asia & Pacific	2674.724000	18.133030	2.742755	0.538734	0.235561
Europe & Central Asia	8078.412712	26.452311	9.693711	0.214042	0.827046
Latin America & Caribbean	2869.781764	19.424099	3.227705	0.477205	1.425947
Middle East & North Africa	2442.522107	15.811009	1.814005	0.556232	1.210453
North America	16084.481703	29.546398	24.133553	0.159483	1.483280
South Asia	437.461834	7.647880	0.885822	0.383857	0.255123
Sub-Saharan Africa	443.535524	6.222082	0.350875	0.079128	0.349642

4

4.1 Indicateurs par régions ('Country Name').

Médianes

Country Name	GDP per capita (current US\$)	Internet users (per 100 people)	Personal computers (per 100 people)	Population growth (annual %)	Unemployment, total (% of total labor force)
East Asia & Pacific	3423.174263	12.596964	2.982600	1.242380	4.370391
Europe & Central Asia	11000.376220	30.361230	11.471852	0.471662	9.062414
Latin America & Caribbean	3331.977867	14.402387	2.683911	1.779383	8.483110
Middle East & North Africa	2259.919451	9.753088	3.597743	2.315994	12.070234
North America	25822.080388	61.939741	24.814917	0.996637	5.909534
South Asia	345.761577	2.164376	0.236455	2.145093	4.154114
Sub-Saharan Africa	583.329758	2.724794	0.918649	2.780162	7.811388

4

4.2 Indicateurs par pays.

Moyennes

Country Name	Barro-Lee: Percentage of population age 15-19 with secondary schooling. Completed Secondary	Barro-Lee: Percentage of population age 15-19 with secondary schooling. Total (Incomplete and Completed Secondary)	Barro-Lee: Percentage of population age 20-24 with tertiary schooling. Completed Tertiary	Barro-Lee: Percentage of population age 20-24 with tertiary schooling. Total (Incomplete and Completed Tertiary)	Barro-Lee: Percentage of population age 25+ with tertiary schooling. Completed Tertiary	Barro-Lee: Percentage of population age 25+ with tertiary schooling. Total (Incomplete and Completed Tertiary)	Barro-Lee: Percentage of population age 25-29 with tertiary schooling. Completed Tertiary	Barro-Lee: Percentage of population age 25-29 with tertiary schooling. Total (Incomplete and Completed Tertiary)	Barro-Lee: Population in thousands, age 15-19, total	Barro-Lee: Population in thousands, age 20-24, total	...	Expenditure on education as % of total government expenditure (%)
Afghanistan	4.166667	13.773333	3.451111	8.535556	3.431111	5.170000	4.747778	7.834444	2074.666667	1723.777778	...	14.091753
Albania	19.003333	76.648889	2.594444	6.443333	3.097778	4.667778	3.875556	6.438889	289.000000	256.888889	...	10.661215
Algeria	14.858889	34.171111	4.088889	9.997778	2.592222	3.887778	4.451111	7.275556	2725.666667	2405.666667	...	11.429410
American Samoa	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Andorra	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...
Virgin Islands (U.S.)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
West Bank and Gaza	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Yemen, Rep.	4.336667	16.650000	0.931111	2.264444	0.542222	0.814444	1.057778	1.720000	1524.000000	1232.555556	...	24.433267
Zambia	4.446667	25.137778	0.185556	0.461111	0.703333	1.062222	0.400000	0.660000	942.666667	786.666667	...	7.709410
Zimbabwe	8.053333	42.558889	1.300000	3.264444	1.784444	2.731111	2.118889	3.533333	1173.888889	994.333333	...	24.147020

217 rows × 22 columns

4

4.2 Indicateurs par pays.

Ecart-types

Country Name	Barro-Lee: Percentage of population age 15-19 with secondary schooling. Completed Secondary	Barro-Lee: Percentage of population age 15-19 with secondary schooling. Total (Incomplete and Completed Secondary)	Barro-Lee: Percentage of population age 20-24 with tertiary schooling. Completed Tertiary	Barro-Lee: Percentage of population age 20-24 with tertiary schooling. Total (Incomplete and Completed Tertiary)	Barro-Lee: Percentage of population age 25+ with tertiary schooling. Completed Tertiary	Barro-Lee: Percentage of population age 25+ with tertiary schooling. Total (Incomplete and Completed Tertiary)	Barro-Lee: Percentage of population age 25-29 with tertiary schooling. Completed Tertiary	Barro-Lee: Percentage of population age 25-29 with tertiary schooling. Total (Incomplete and Completed Tertiary)	Barro-Lee: Population in thousands, age 15-19, total	Barro-Lee: Population in thousands, age 20-24, total	...	Expenditure on education as % of total government expenditure (%)
Afghanistan	3.251946	9.641605	0.420786	0.908104	1.764684	2.596878	1.509286	2.400157	910.346225	731.788183	...	2.419881
Albania	7.591049	15.779808	0.559332	1.478445	1.284005	1.874852	0.921264	1.595991	34.084454	48.029794	...	0.746562
Algeria	5.116941	10.951179	3.329645	8.061454	2.325058	3.464436	3.777110	6.129835	898.699199	991.346559	...	NaN
American Samoa	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Andorra	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...
Virgin Islands (U.S.)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
West Bank and Gaza	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Yemen, Rep.	3.905803	14.777690	0.774862	1.879941	0.615808	0.925852	1.050366	1.710088	764.438519	647.929609	...	10.344172
Zambia	3.600479	21.119953	0.157092	0.384722	0.518411	0.770564	0.253574	0.409329	392.265854	342.082227	...	1.943688
Zimbabwe	7.459636	18.898248	1.117542	2.829766	0.877042	1.327944	1.600246	2.648783	460.528621	455.548296	...	10.298967

217 rows × 22 columns

4

4.2 Indicateurs par pays.

médianes

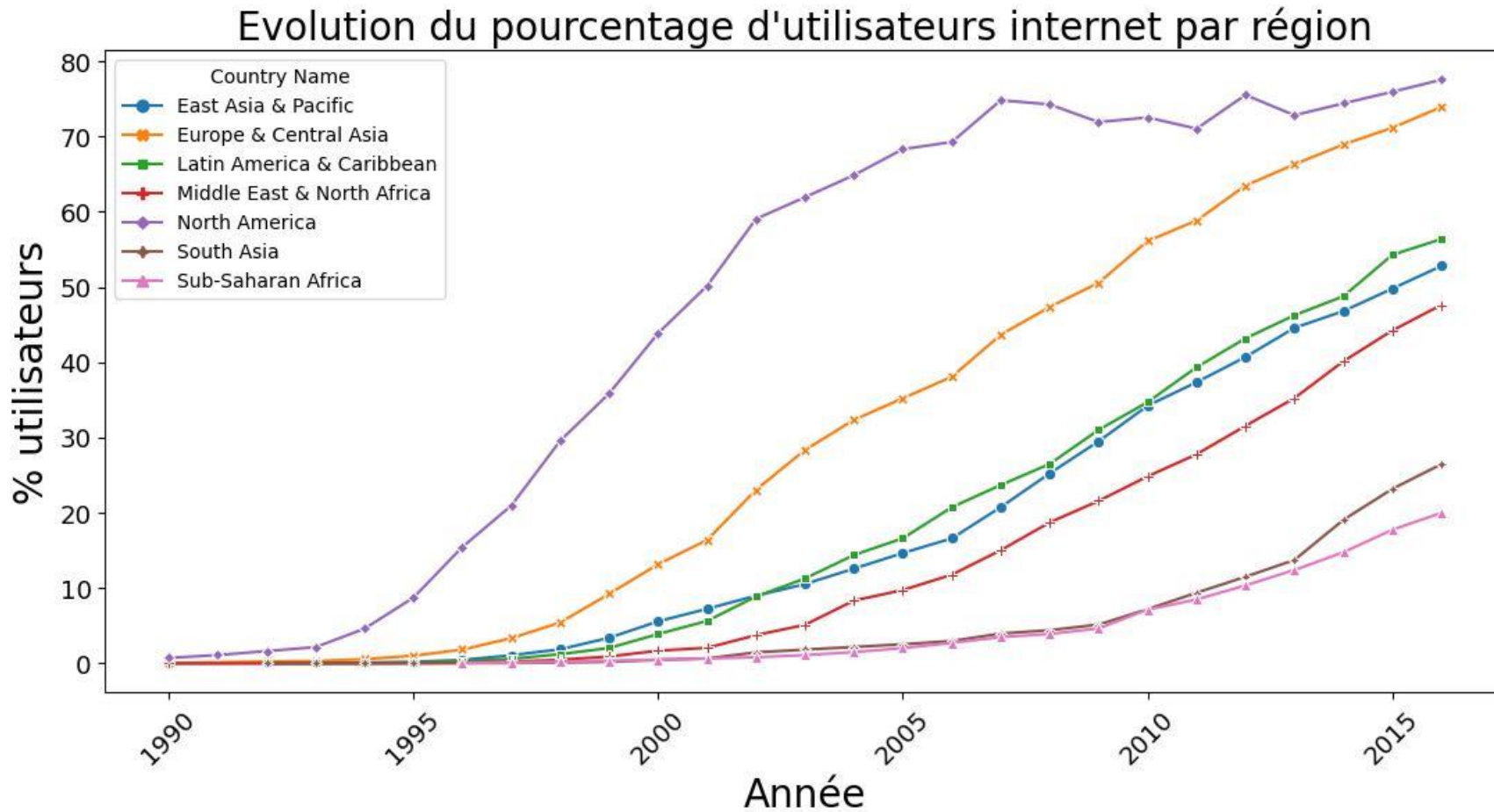
Country Name	Barro-Lee: Percentage of population age 15-19 with secondary schooling. Completed Secondary	Barro-Lee: Percentage of population age 15-19 with secondary schooling. Total (Incomplete and Completed Secondary)	Barro-Lee: Percentage of population age 20-24 with tertiary schooling. Completed Tertiary	Barro-Lee: Percentage of population age 20-24 with tertiary schooling. Total (Incomplete and Completed Tertiary)	Barro-Lee: Percentage of population age 25+ with tertiary schooling. Completed Tertiary	Barro-Lee: Percentage of population age 25+ with tertiary schooling. Total (Incomplete and Completed Tertiary)	Barro-Lee: Percentage of population age 25-29 with tertiary schooling. Completed Tertiary	Barro-Lee: Percentage of population age 25-29 with tertiary schooling. Total (Incomplete and Completed Tertiary)	Barro-Lee: Population in thousands, age 15-19, total	Barro-Lee: Population in thousands, age 20-24, total	...	Expenditure on education as % of total government expenditure (%)
Afghanistan	2.09	7.68	3.54	8.80	3.70	5.55	5.35	8.79	1537.0	1288.0	...	14.284365
Albania	18.39	81.51	2.70	6.62	3.06	4.72	4.05	6.62	303.0	268.0	...	10.642545
Algeria	15.48	34.85	3.41	8.34	2.12	3.19	4.07	6.66	2769.0	2390.0	...	11.429410
American Samoa	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN
Andorra	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN
...
Virgin Islands (U.S.)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN
West Bank and Gaza	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN
Yemen, Rep.	4.63	18.30	0.90	2.20	0.31	0.47	0.85	1.39	1258.0	1035.0	...	30.321280
Zambia	3.99	20.67	0.13	0.33	0.39	0.60	0.29	0.50	943.0	741.0	...	7.729730
Zimbabwe	4.10	51.77	1.18	2.89	1.77	2.80	1.97	3.33	1181.0	976.0	...	28.929774

217 rows × 22 columns

5

5. Graphiques.

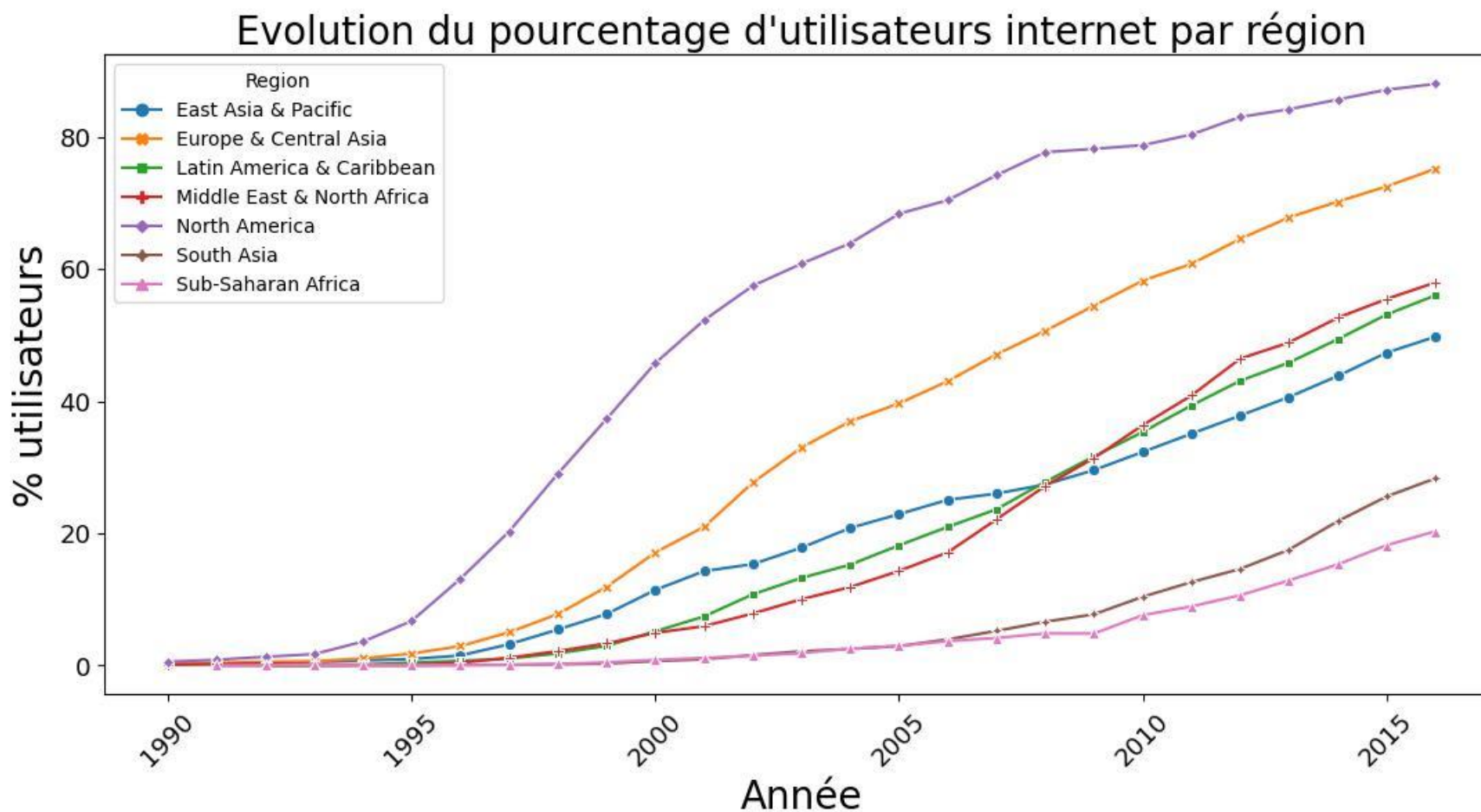
Obtenu avec les données des 7 régions contenues dans la colonne 'Country Name'.



5

5. Graphiques.

Obtenu avec les données des 7 régions contenues dans la colonne 'Region'.



6

6. Conclusion.

Les données 2011 à 2016 qui sont disponibles :

- 4 variables complètes et 5 incomplètes en ce qui concerne les données région.
- 9 variables incomplètes en ce qui concerne les données pays.

Les 2 graphiques laissent penser que les données région pourraient parfois différer en fonction de la manière dont elles sont obtenues.

Il reste donc à savoir si les variables disponibles sont suffisamment pertinentes et en quantités suffisantes pour une étude de marché complète.