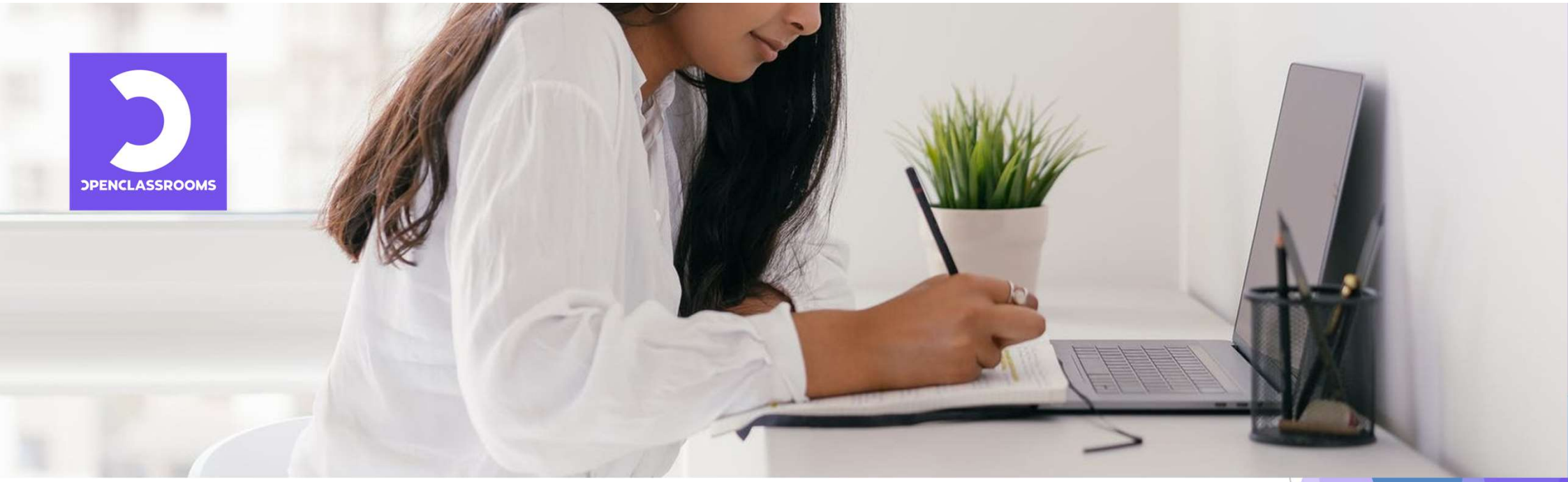


FORMATION DATA SCIENTIST : PROJET 2

Analysez des données de systèmes éducatifs

Sommaire

1. Contexte et Objectifs
2. Exploration des données
3. Sélection des indicateurs
4. Analyse des données
5. Résultats
6. Conclusion

1. Contexte et Objectifs

- Academy

  Start-up de la EdTech, qui fournit des contenus de formation en ligne pour un public de niveau lycée et université

- Sources

  Données provenant de la Banque mondiale

- Etude de marché

  Identifier les pays présentant un potentiel de clients élevé pour son expansion à l'international

2. Exploration des données

Les jeux de données

	Contenu	Taille	Valeurs manquantes	Informations
EdStatsCountry	liste de pays avec : - code - informations économiques - régions et classes de revenus	241 lignes 32 colonnes	30.52%	241 pays
EdStatsSeries	liste d'indicateurs et leur définition	3665 lignes 21 colonnes	71.72%	3665 indicateurs 37 sujets
EdStatsData	données des indicateurs par pays et par année	886930 lignes 70 colonnes	86.1%	3665 indicateurs 242 pays 1970 à 2100
EdStatsCountry-Series	informations sur la récolte des indicateurs par pays	613 lignes 4 colonnes	25%	21 indicateurs 211 pays
EdStatsFootNote	informations pour certains indicateurs	643638 lignes 5 colonnes	20%	1558 indicateurs 239 pays

3. Stratégie de sélection des indicateurs

Pré-sélection

Les sujets

Sélection des sujets les plus pertinants

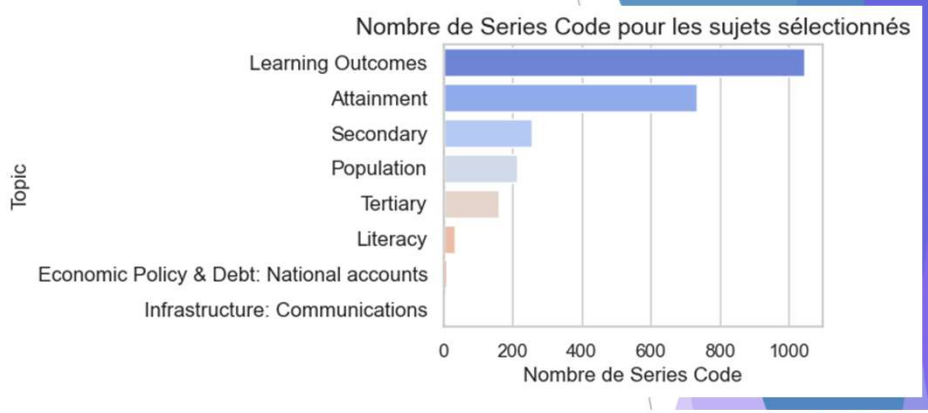
Mots-clefs

Filtre avec exclusion de mots-clefs pour ne conserver que les indicateurs pertinants (bonne tranche d'âge, hommes et femmes, etc.)

Pré-sélection

12 indicateurs pré-selectionnés,
classés par nombre de pays avec des données

4 thèmes :
- technologie (internet et pc)
- taille du marché
- revenus
- alphabétisation



	Indicator Name	count
0	GDP per capita (current US$)	209
1	Internet users (per 100 people)	208
2	Enrolment in tertiary education, all programmes, both sexes (number)	202
3	Population of the official age for tertiary education, both sexes (number)	201
4	Enrolment in upper secondary education, both sexes (number)	200
5	Population of the official age for upper secondary education, both sexes (number)	200
6	Gross enrolment ratio, upper secondary, both sexes (%)	195
7	Gross enrolment ratio, tertiary, both sexes (%)	193
8	Personal computers (per 100 people)	193
9	Wittgenstein Projection: Population in thousands by highest level of educational attainment. Post Secondary. Total	166
10	Youth literacy rate, population 15-24 years, both sexes (%)	164
11	Barro-Lee: Population in thousands, age 15-19, total	144
12	Barro-Lee: Population in thousands, age 20-24, total	144

3. Stratégie de sélection des indicateurs

Analyse des indicateurs pré-sélectionnés : méthode

Complétude

Indicateurs avec le plus de données

Ancienneté

Données les plus récentes

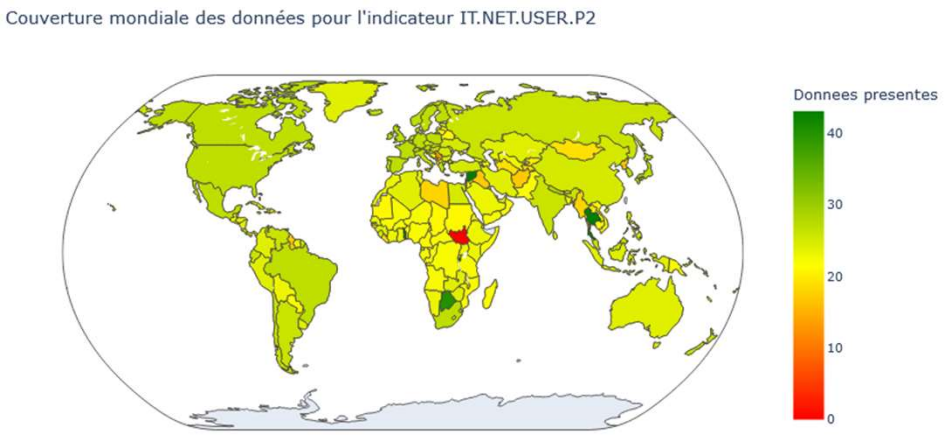
Représentativité mondiale

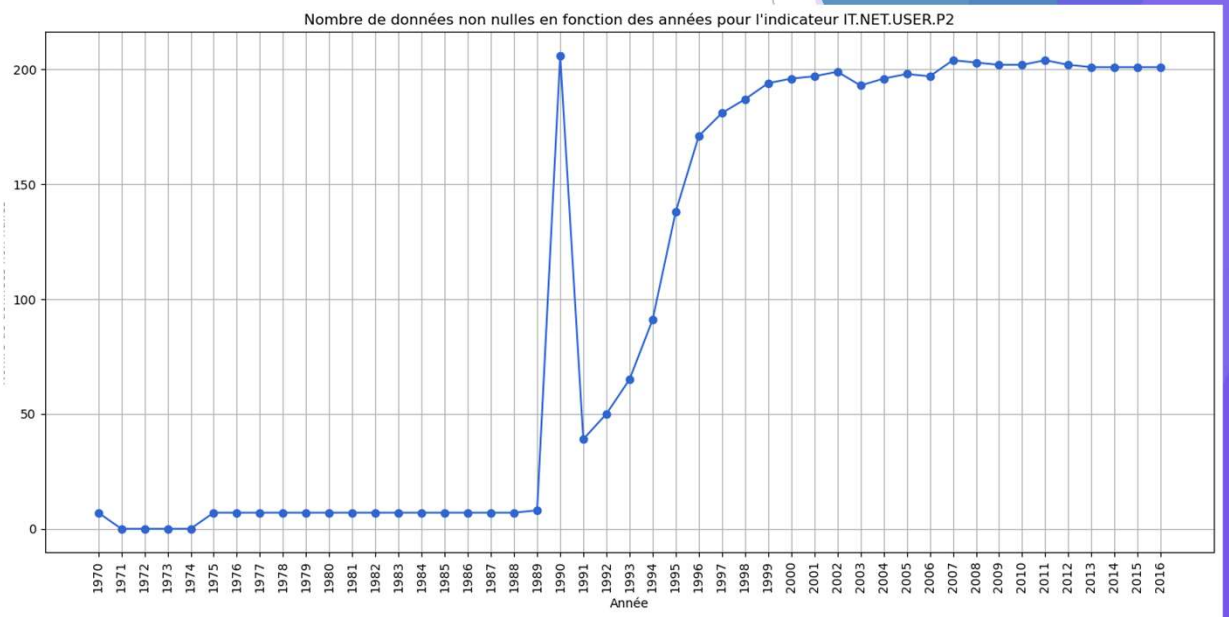
Indicateurs avec le plus de pays prometteurs renseignés

	Indicator Name
0	GDP per capita (current US$)
1	Internet users (per 100 people)
2	Population of the official age for tertiary education, both sexes (number)
3	Population of the official age for upper secondary education, both sexes (number)
4	Wittgenstein Projection: Population in thousands by highest level of educational attainment. Post Secondary. Total

3. Stratégie de sélection des indicateurs

Analyse des indicateurs pré-sélectionnés : Internet

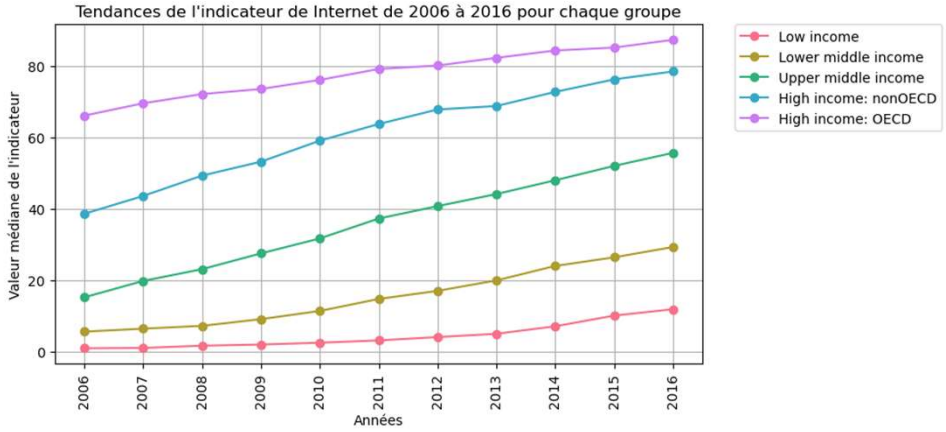


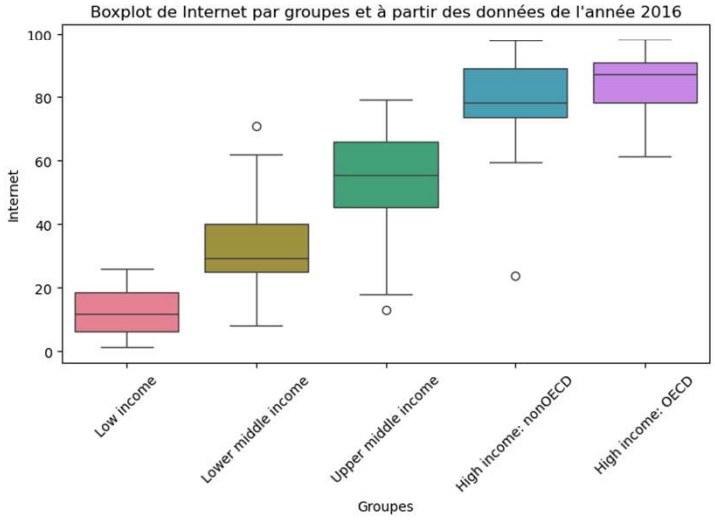


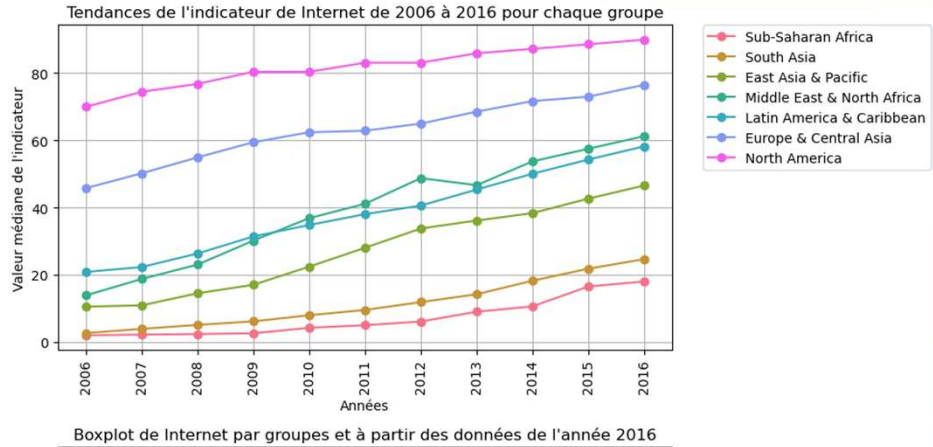
Pays sans données en 2016: ['American Samoa', 'Channel Islands', 'Curacao', 'Isle of Man', 'Korea, Dem. People's Rep.', 'Kosovo', 'Nauru', 'New Caledonia', 'Northern Mariana Islands', 'Palau', 'San Marino', 'Sint Maarten (Dutch part)', 'South Sudan', 'St. Martin (French part)', 'Turks and Caicos Islands']

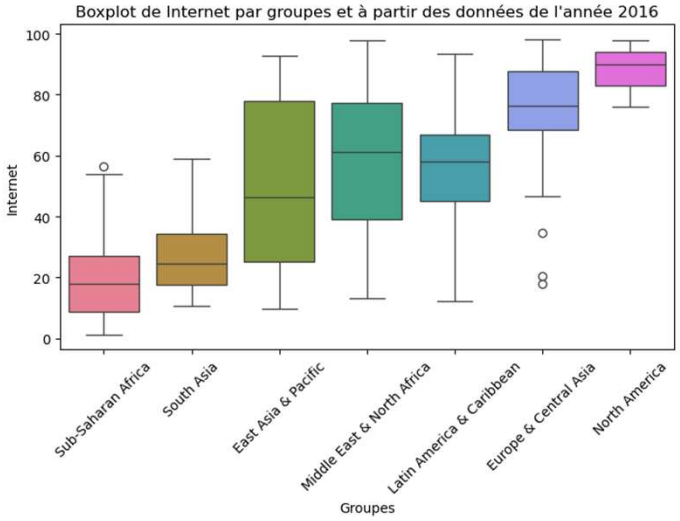
3. Stratégie de sélection des indicateurs

Analyse des indicateurs pré-sélectionnés : Internet (suite)



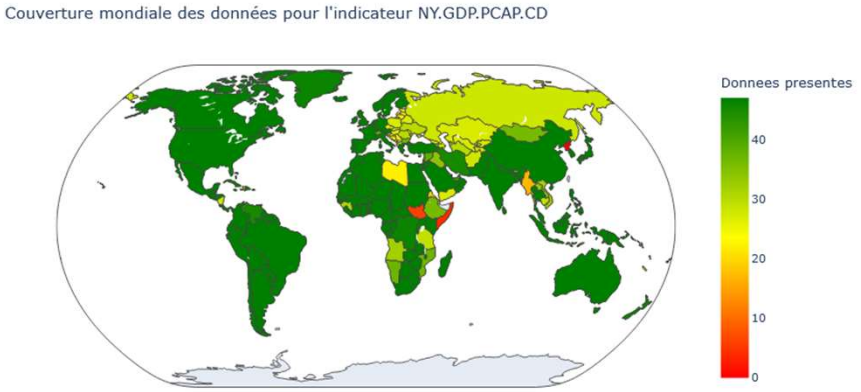


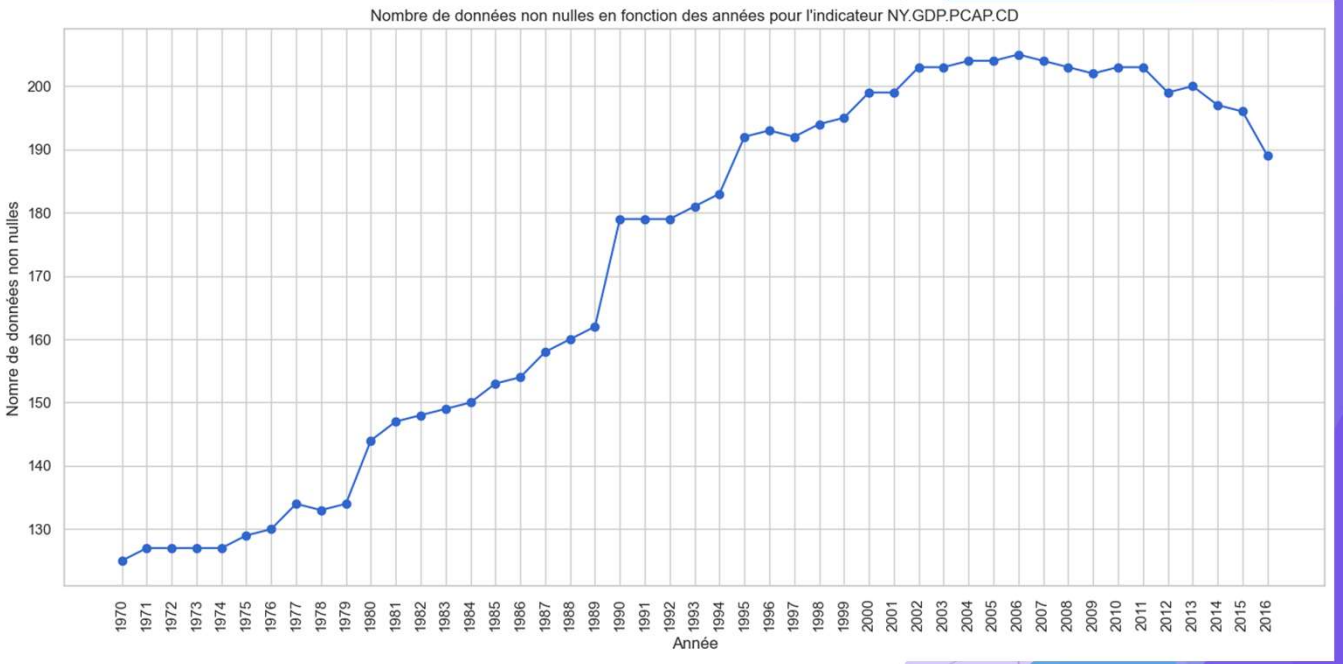




3. Stratégie de sélection des indicateurs

Analyse des indicateurs pré-sélectionnés : GDP

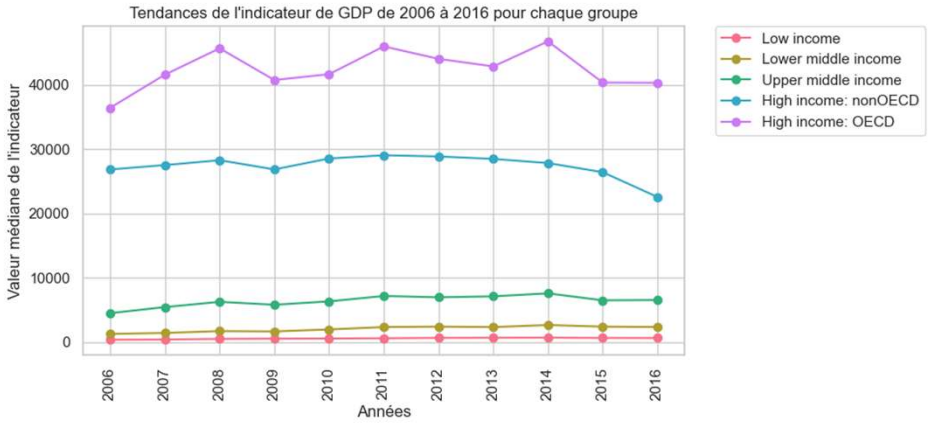


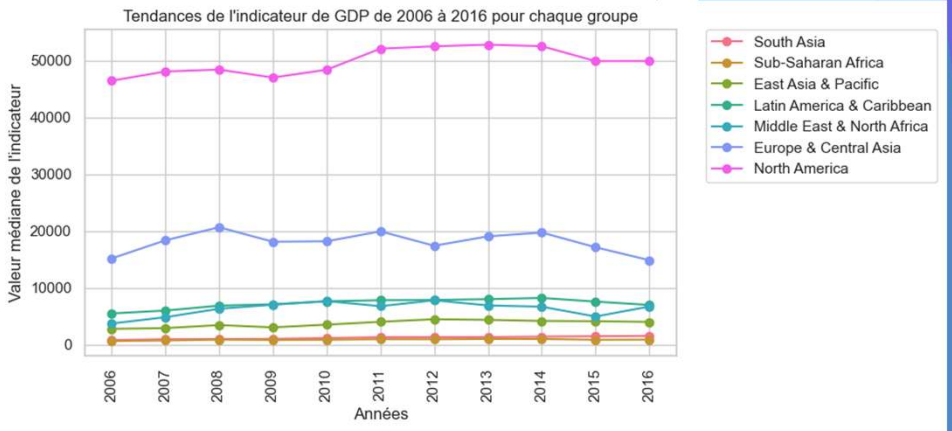


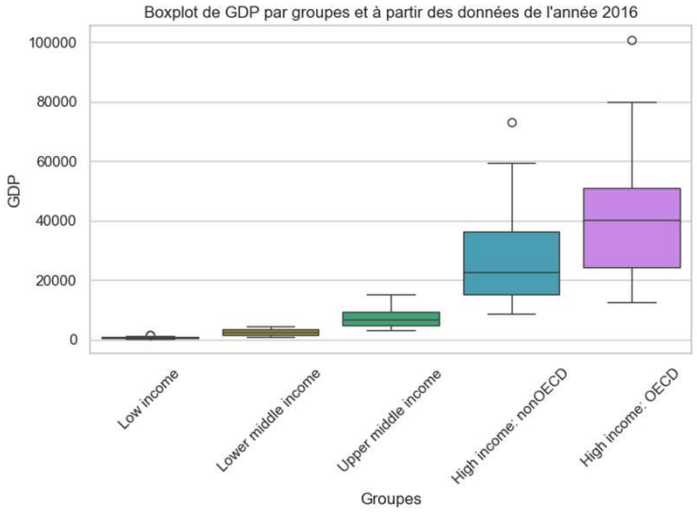
Pays sans données en 2016: ['Aruba', 'Bermuda', 'Cayman Islands', 'Channel Islands', 'Cuba', 'Curacao', 'Djibouti', 'Eritrea', 'Faroe Islands', 'French Polynesia', 'Gibraltar', 'Greenland', 'Isle of Man', 'Korea, Dem. People's Rep.', 'Libya', 'Liechtenstein', 'Monaco', 'Nauru', 'New Caledonia', 'Puerto Rico', 'Sint Maarten (Dutch part)', 'South Sudan', 'St. Martin (French part)', 'Syrian Arab Republic', 'Turks and Caicos Islands', 'Venezuela, RB', 'Virgin Islands (U.S.)']

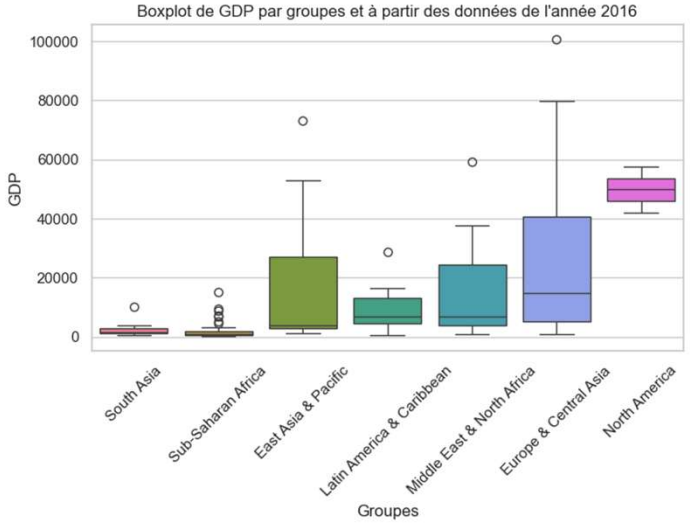
3. Stratégie de sélection des indicateurs

Analyse des indicateurs pré-sélectionnés : GDP (suite)



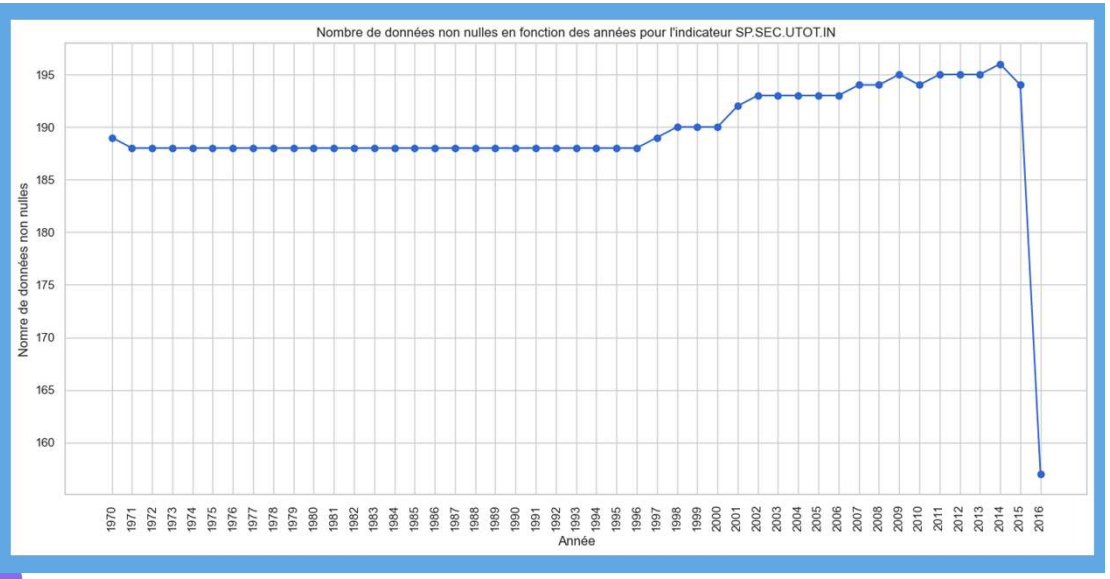


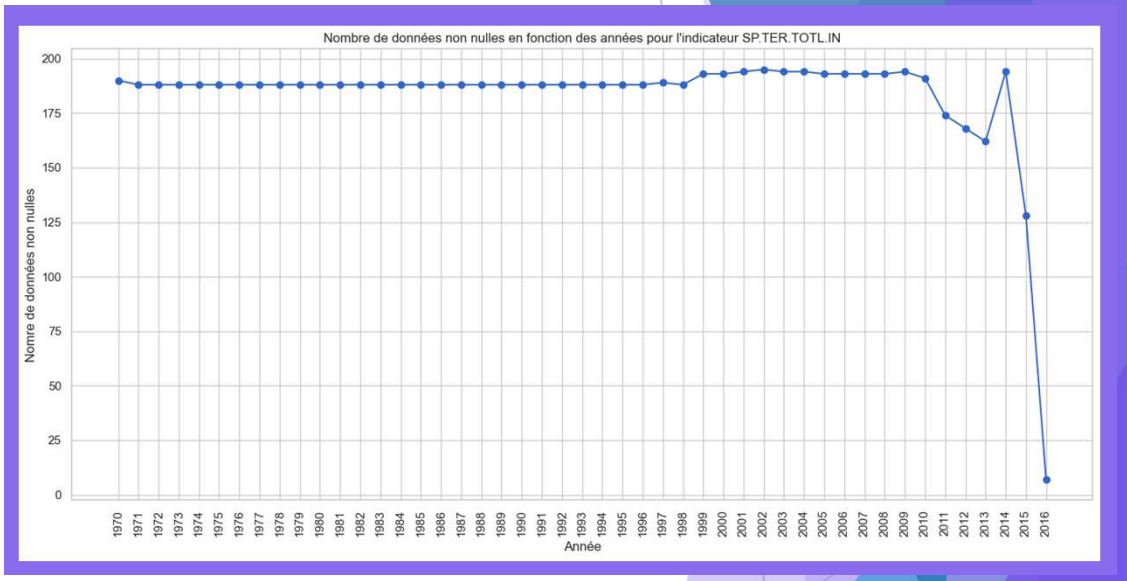




3. Stratégie de sélection des indicateurs

Analyse des indicateurs pré-sélectionnés :
Population of the official age for upper secondary and tertiary education

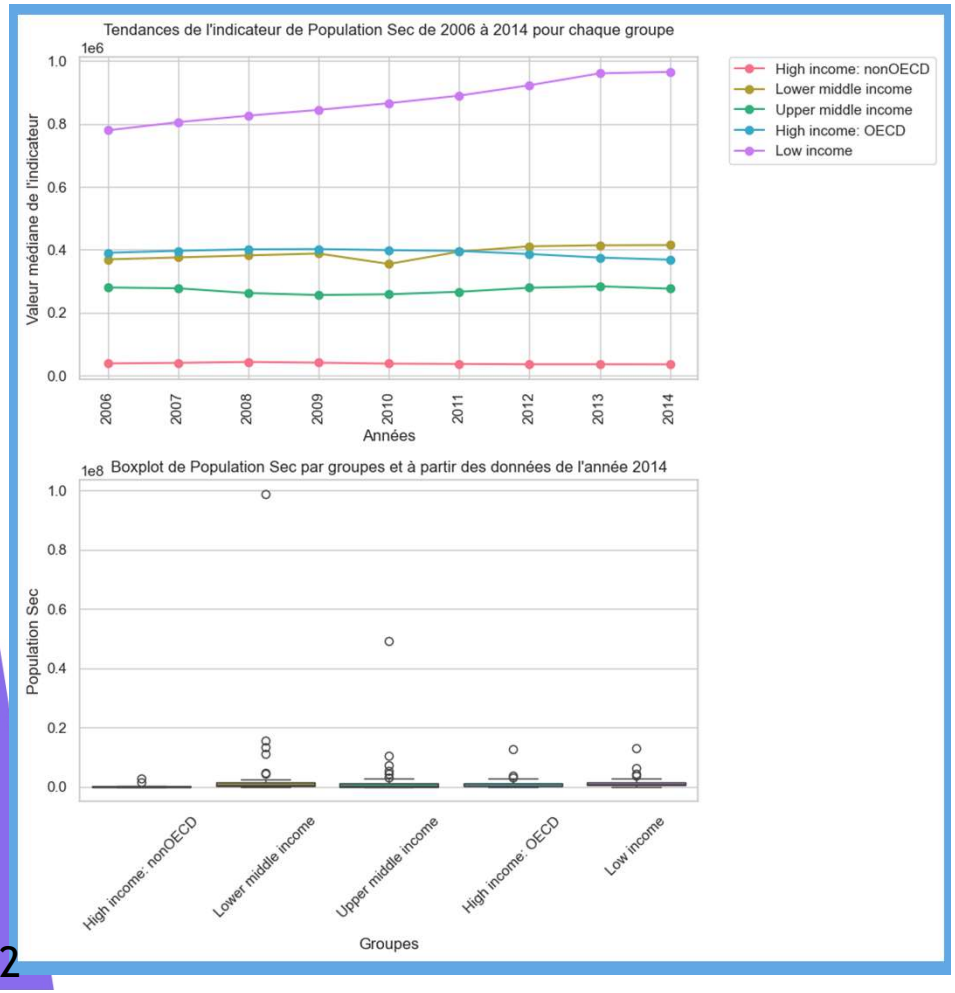


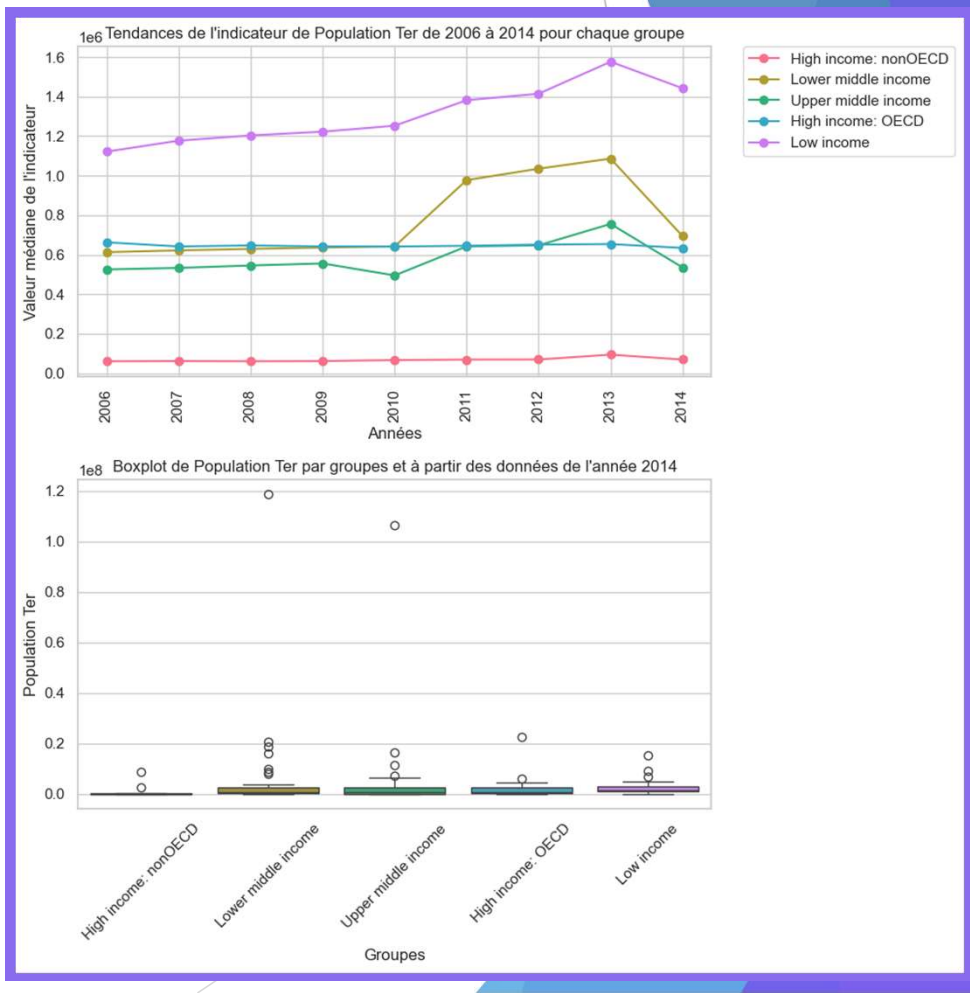


Pays sans données en 2014: ['American Samoa', 'Andorra', 'Bosnia and Herzegovina', 'Cayman Islands', 'Channel Islands', 'Faroe Islands', 'French Polynesia', 'Gibraltar', 'Greenland', 'Guam', 'Isle of Man', 'Kosovo', 'Monaco', 'New Caledonia', 'Northern Mariana Islands', 'Oman', 'Singapore', 'Sint Maarten (Dutch part)', 'St. Martin (French part)', 'Turks and Caicos Islands', 'Virgin Islands (U.S.)']

3. Stratégie de sélection des indicateurs

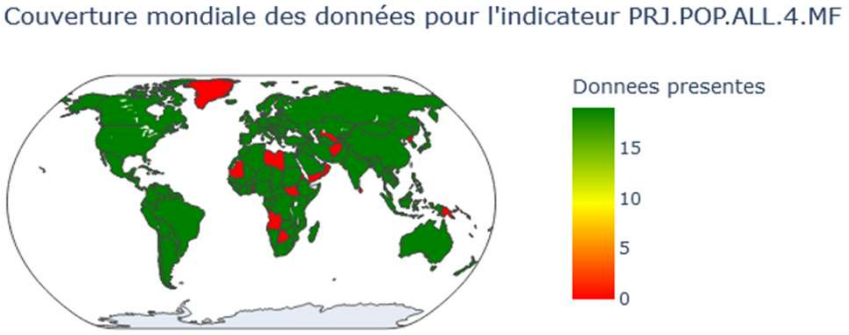
Analyse des indicateurs pré-sélectionnés :
Population of the official age for upper secondary and tertiary education (suite)

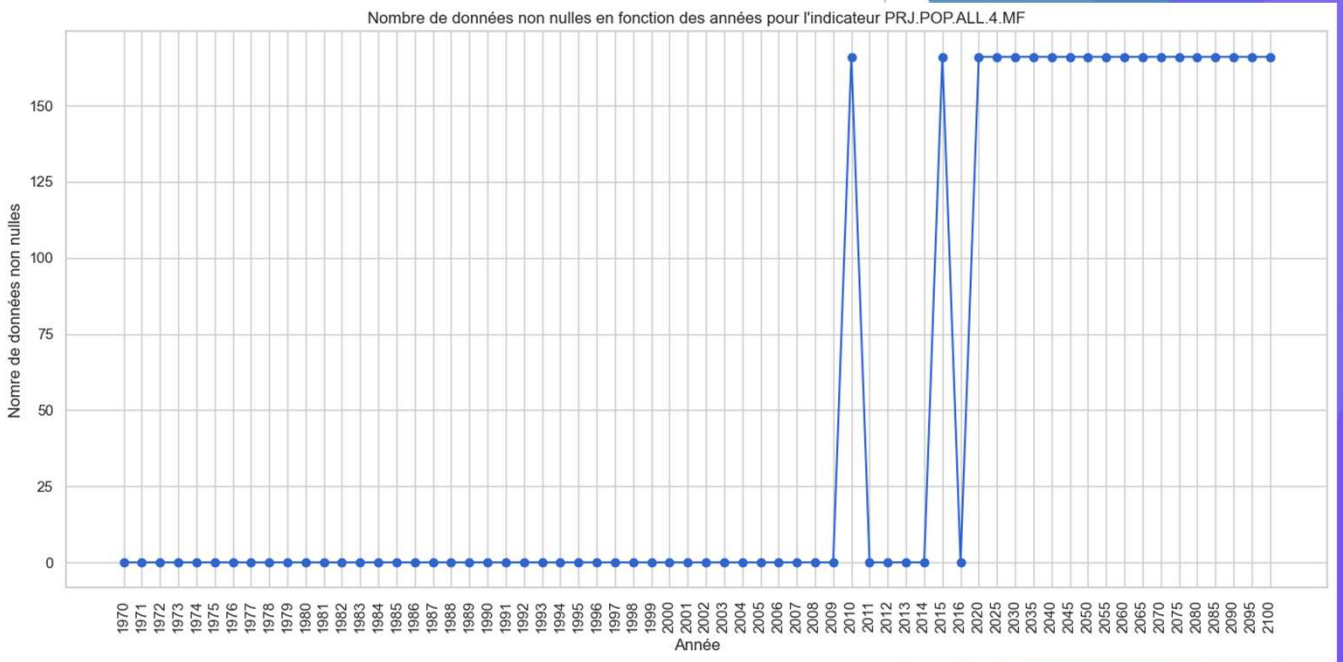




3. Stratégie de sélection des indicateurs

Analyse des indicateurs pré-sélectionnés : Projection du Marché

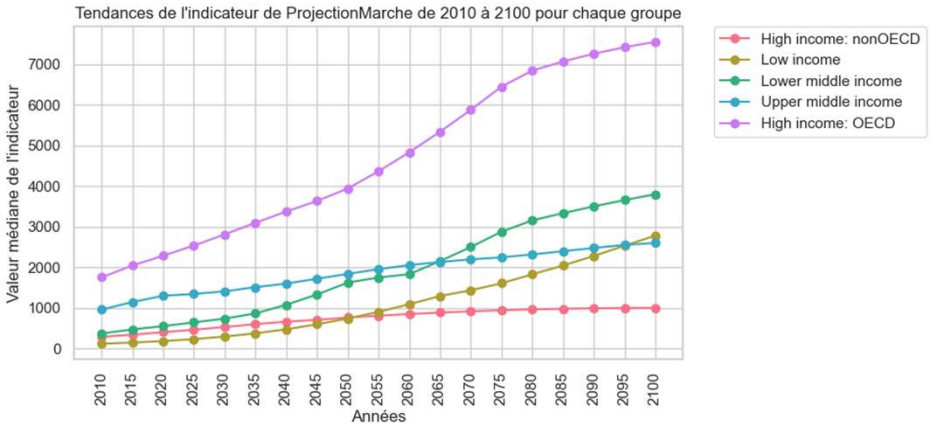


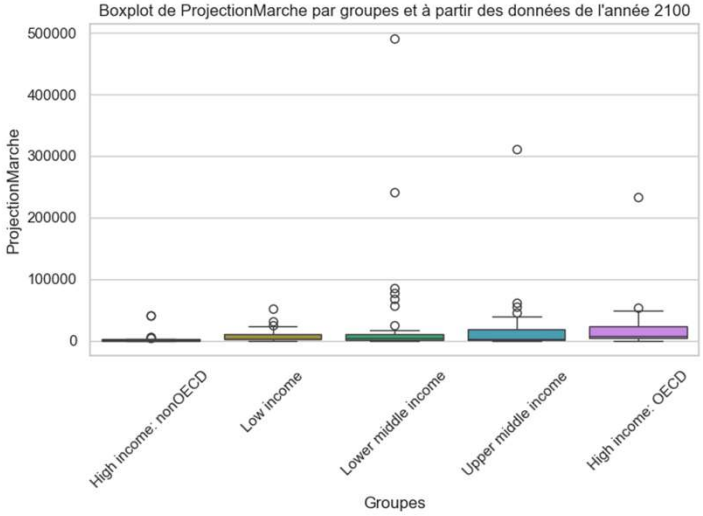


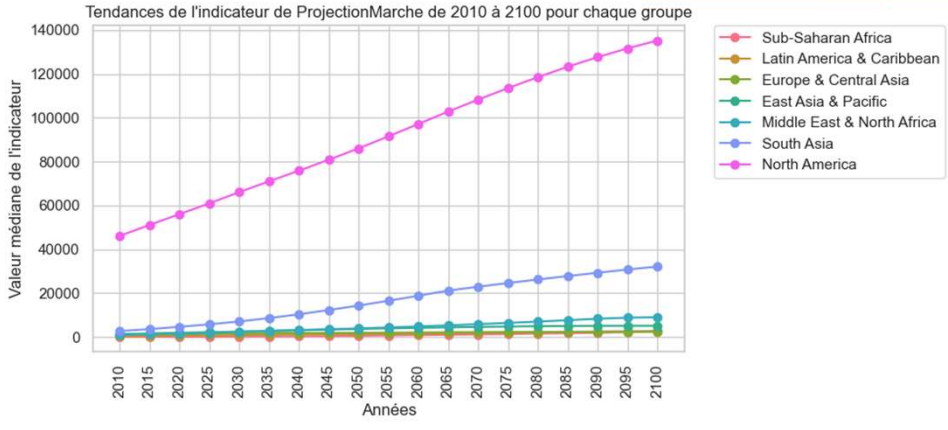
Pays sans données en 2100: ['Afghanistan', 'American Samoa', 'Andorra', 'Angola', 'Antigua and Barbuda', 'Barbados', 'Bermuda', 'Botswana', 'Brunei Darussalam', 'Cayman Islands', 'Channel Islands', 'Curacao', 'Djibouti', 'Dominica', 'Eritrea', 'Faroe Islands', 'Fiji', 'Gibraltar', 'Greenland', 'Grenada', 'Guam', 'Isle of Man', 'Kiribati', 'Korea, Dem. People's Rep.', 'Kosovo', 'Libya', 'Liechtenstein', 'Marshall Islands', 'Mauritania', 'Micronesia, Fed. Sts.', 'Monaco', 'Nauru', 'Northern Mariana Islands', 'Oman', 'Palau', 'Papua New Guinea', 'San Marino', 'Seychelles', 'Sint Maarten (Dutch part)', 'Solomon Islands', 'South Sudan', 'Sri Lanka', 'St. Kitts and Nevis', 'St. Martin (French part)', 'Togo', 'Turks and Caicos Islands', 'Tuvalu', 'Uzbekistan', 'Virgin Islands (U.S.)', 'Yemen, Rep.']

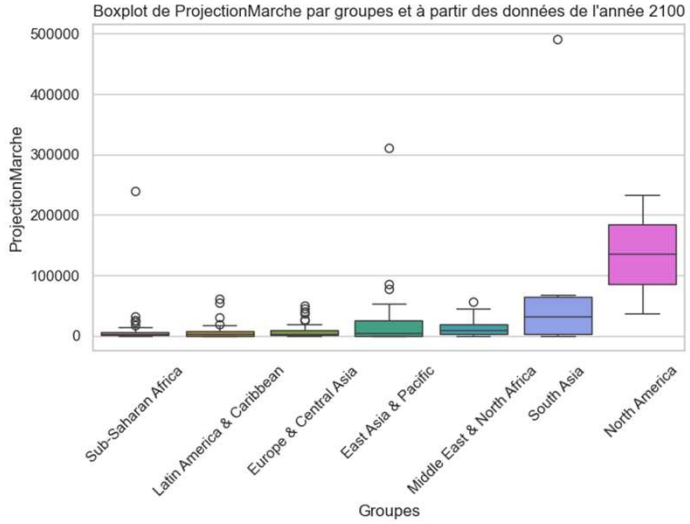
3. Stratégie de sélection des indicateurs

Analyse des indicateurs pré-sélectionnés : Projection du Marché









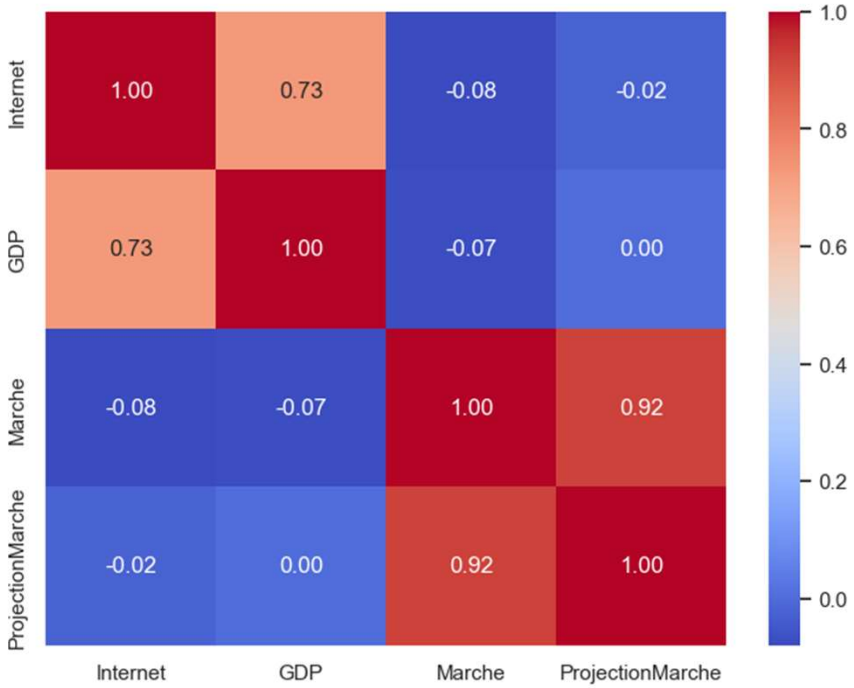
3. Stratégie de sélection des indicateurs

Les indicateurs manquants

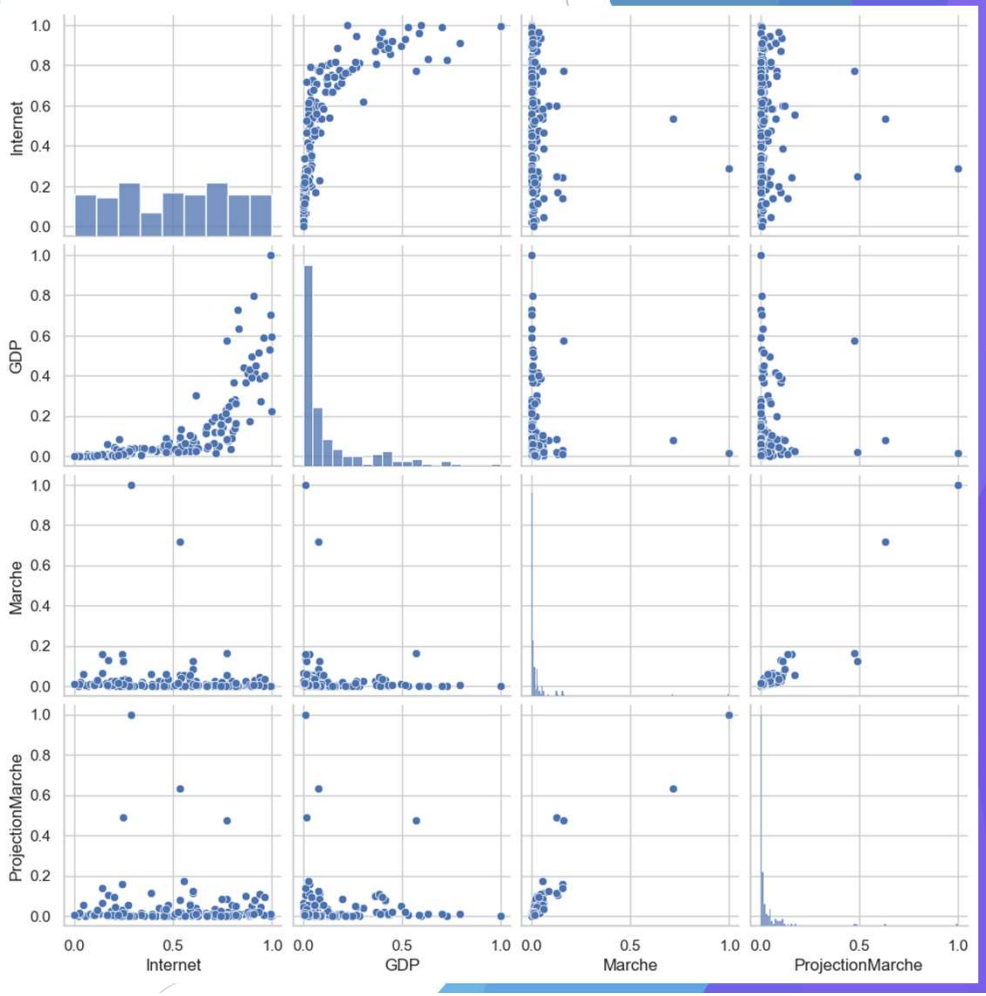
- Ordinateurs
- Langues
- Concurrence
- Alphabétisation
- Utilisation de MOOC

4. Analyse des données

Corrélations entre les indicateurs

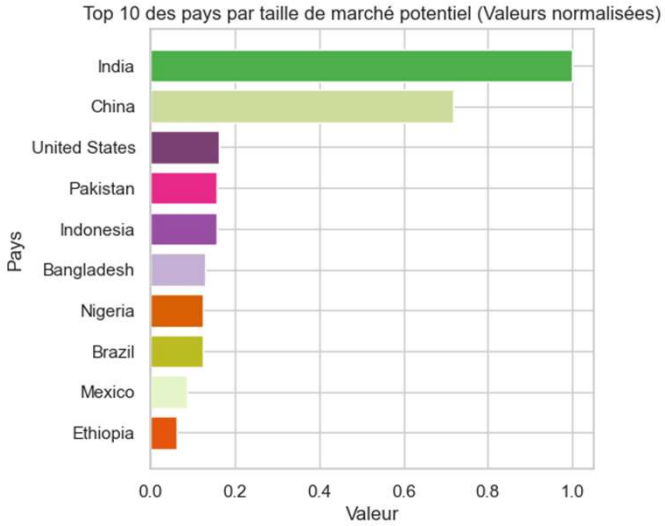


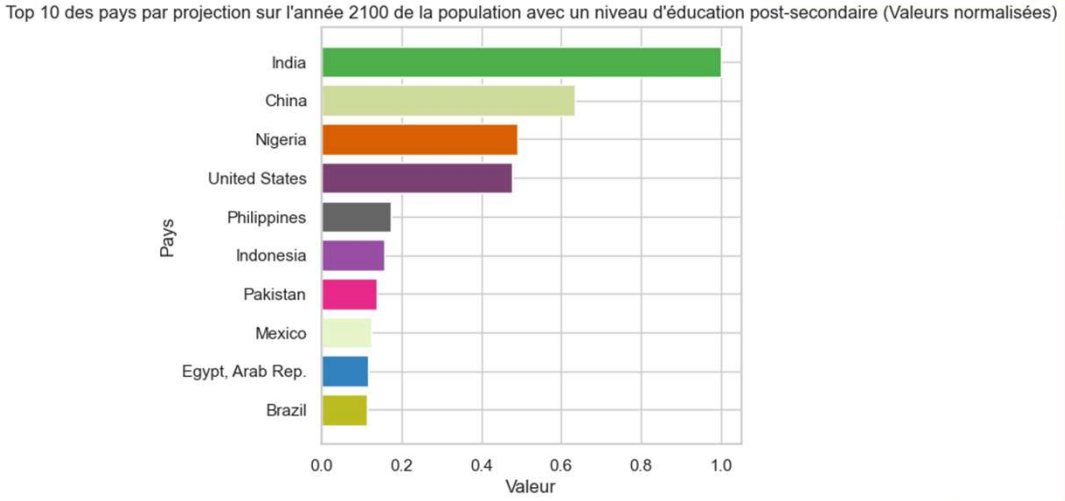
Analyse bivariée

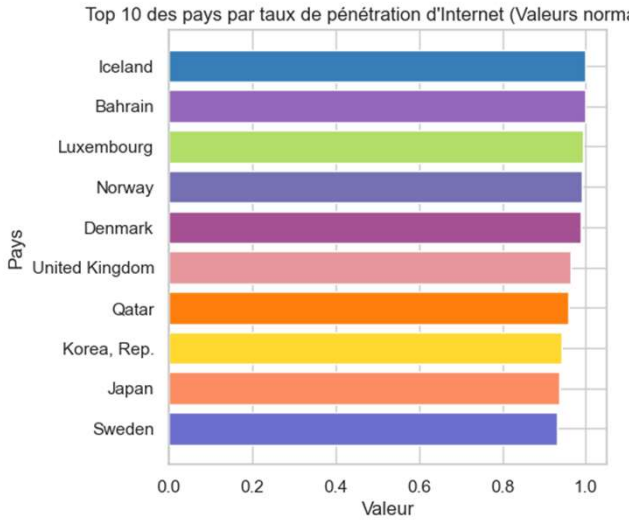


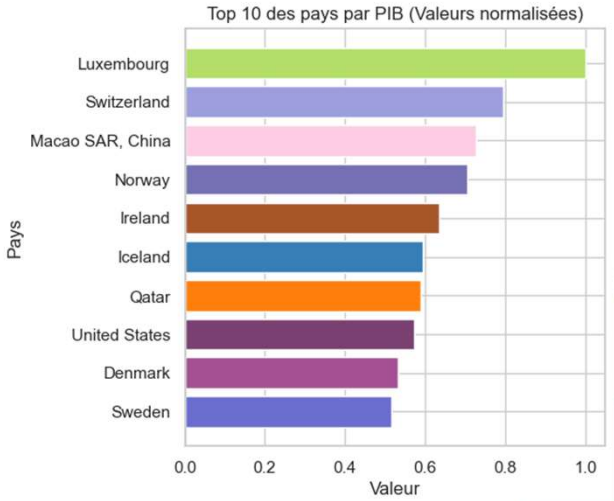
5. Résultats

Top 10 des pays par indicateur

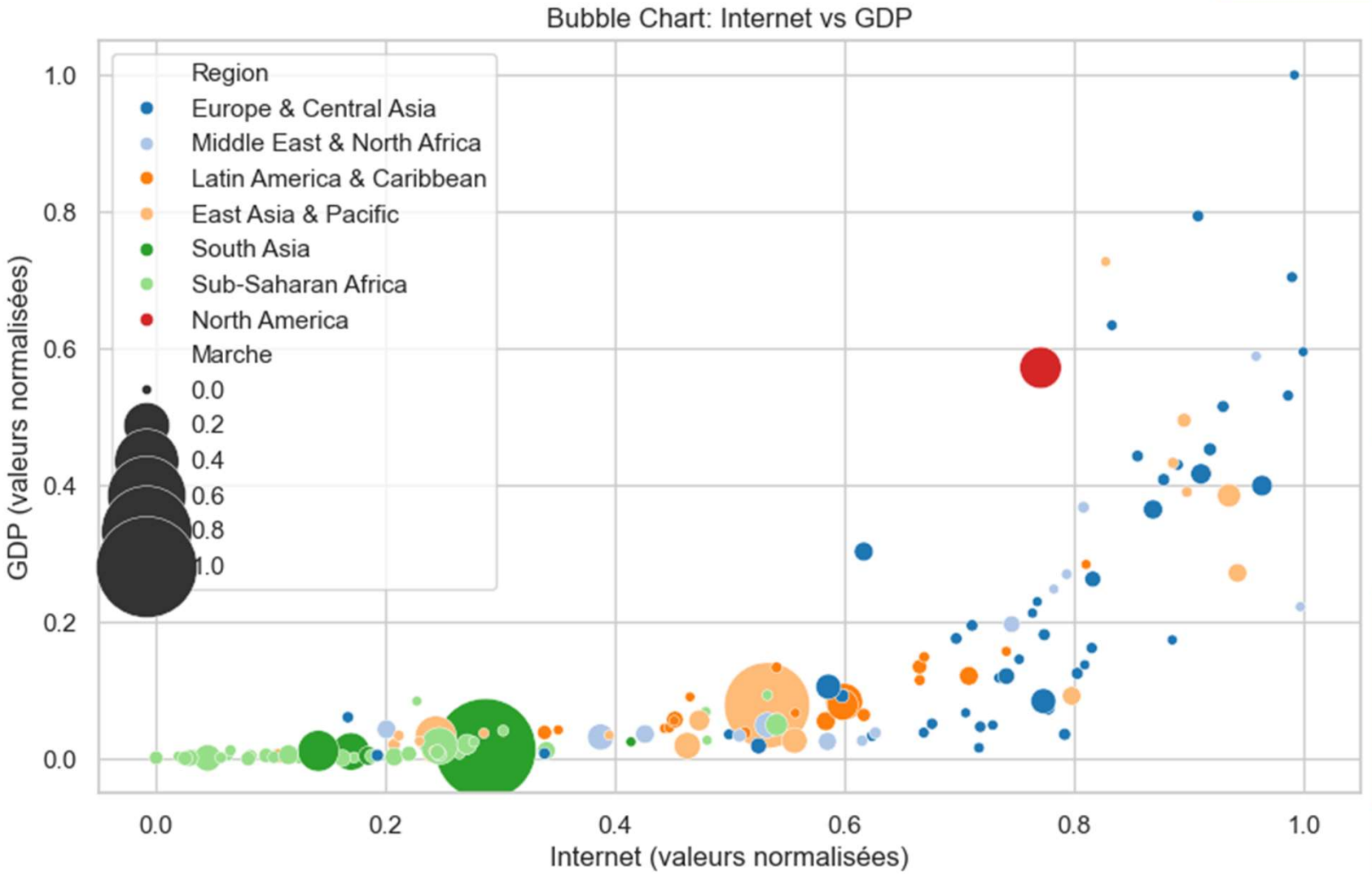




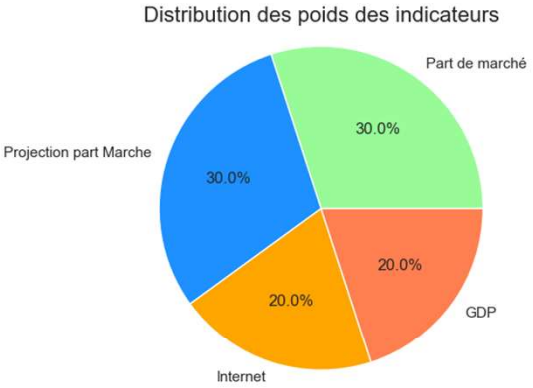


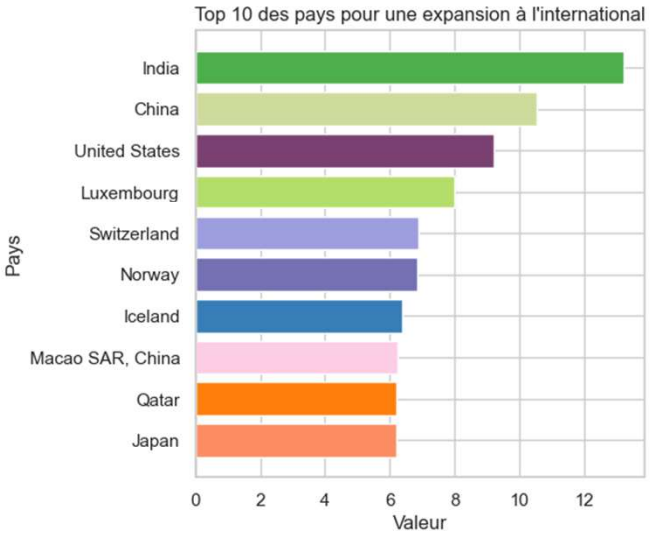


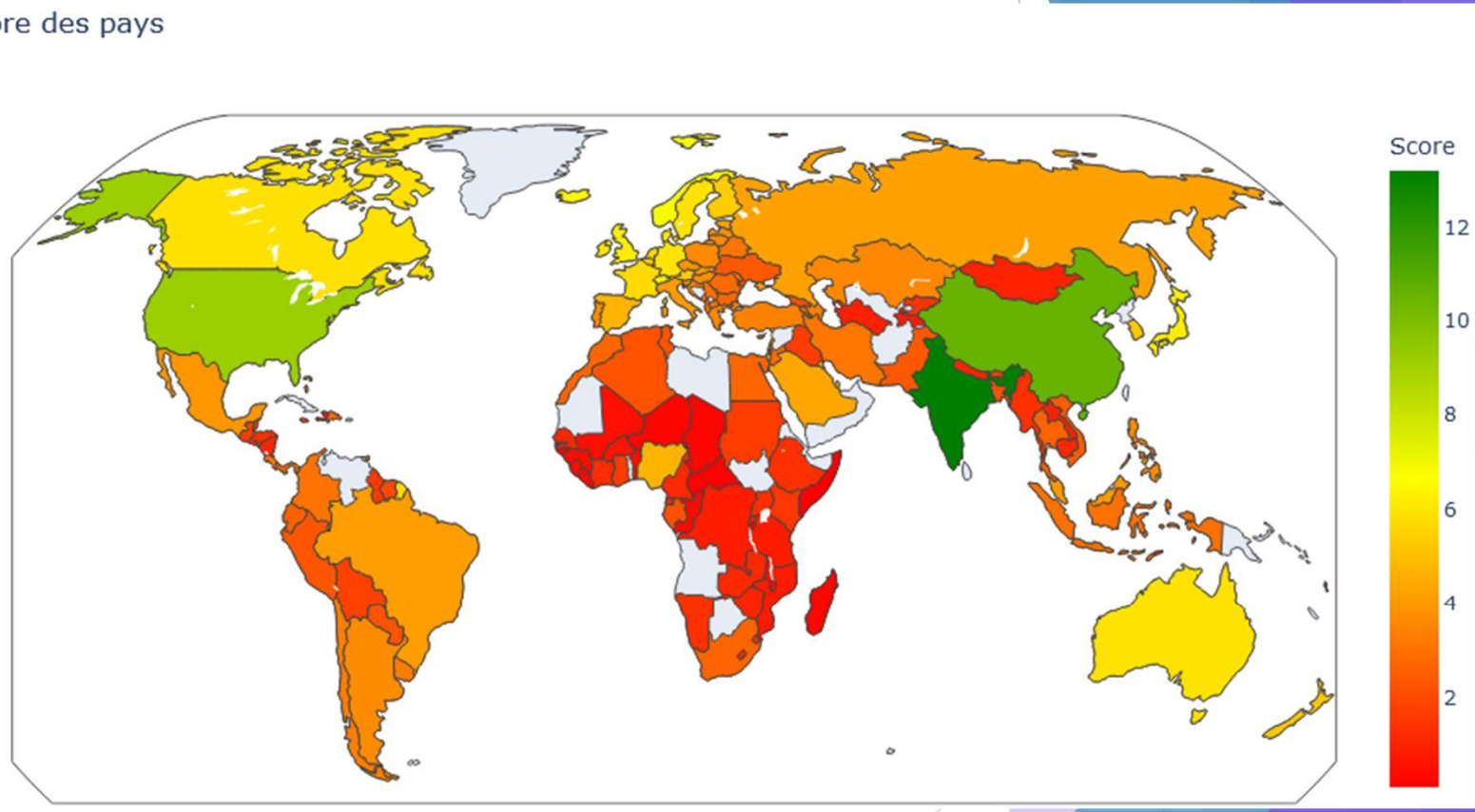
5. Résultats


Bubble Chart: Internet vs GDP
Region
Europe & Central Asia
Middle East & North Africa
Latin America & Caribbean
East Asia & Pacific
South Asia
Sub-Saharan Africa
North America
Marche
0.0
0.2
0.4
0.6
0.8
1.0
GDP (valeurs normalisées)
Internet (valeurs normalisées)
0.0
0.2
0.4
0.6
0.8
1.0

5. Résultats


Distribution des poids des indicateurs
Part de marché
30.0%
Projection part Marche
30.0%
20.0%
20.0%
GDP
Internet


Top 10 des pays pour une expansion à l'international
India
China
United States
Luxembourg
Switzerland
Norway
Iceland
Macao SAR, China
Qatar
Japan
Pays
0
2
4
6
8
10
12
Valeur


Score des pays
Score
12
10
8
6
4
2

6. Conclusion

Adéquation du jeu de données

- Représentation mondiale (géographique et économique)
- Indicateurs pertinents
- Source fiable

Limitations de l'étude

- Ancienneté des données
- Beaucoup de données manquantes
- Manque d'informations (langue, concurrence, …)

Merci pour votre attention

Des questions ?