

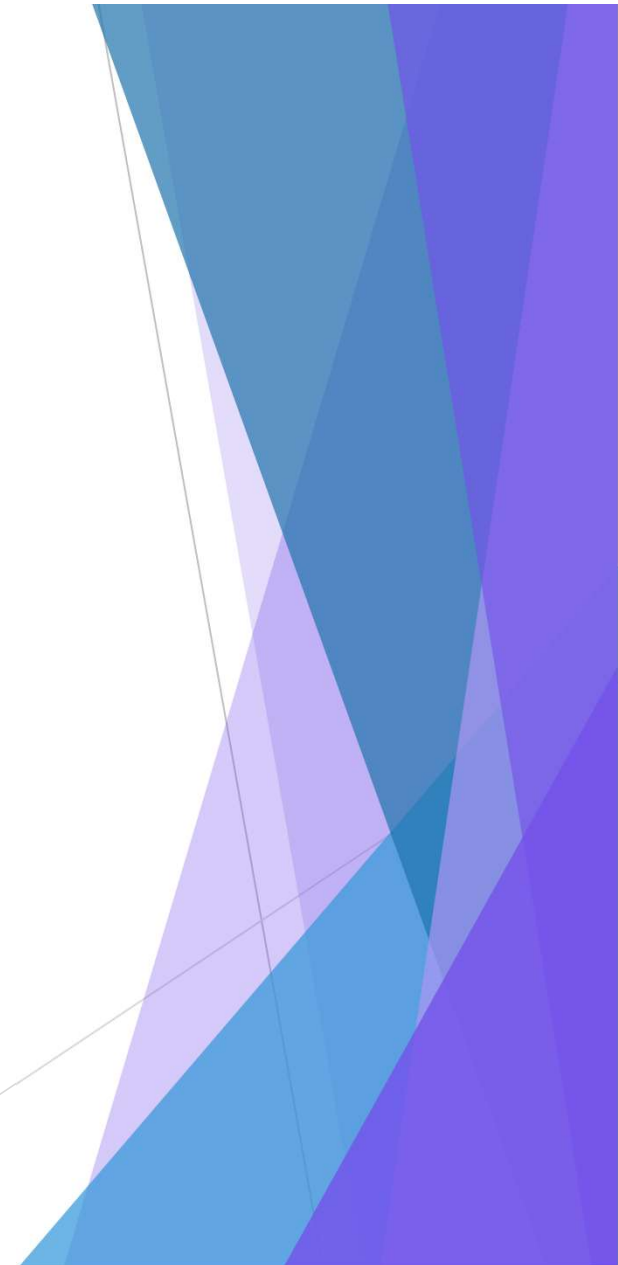
FORMATION DATA SCIENTIST : PROJET 2

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

- **Etude de marché**

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

- **Sources**

Données provenant de la Banque mondiale





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 pertinents



Mots-clefs

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



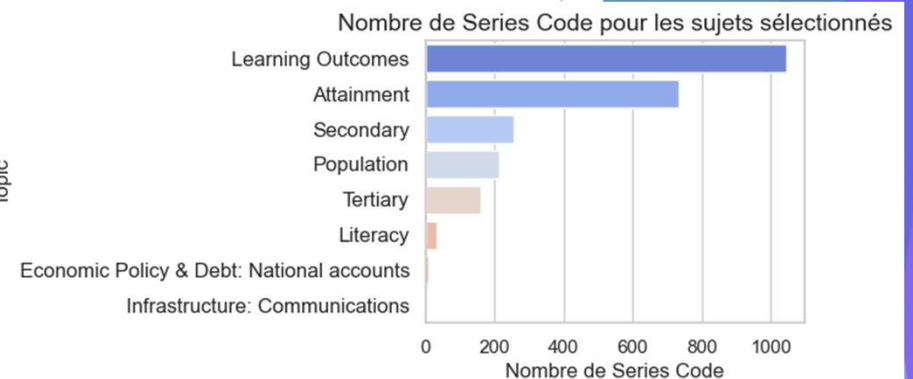
Pré-sélection

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

4 thèmes :

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

Topic



	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	Youth literacy rate, population 15-24 years, both sexes (%)	164
10	Barro-Lee: Population in thousands, age 15-19, total	144
11	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



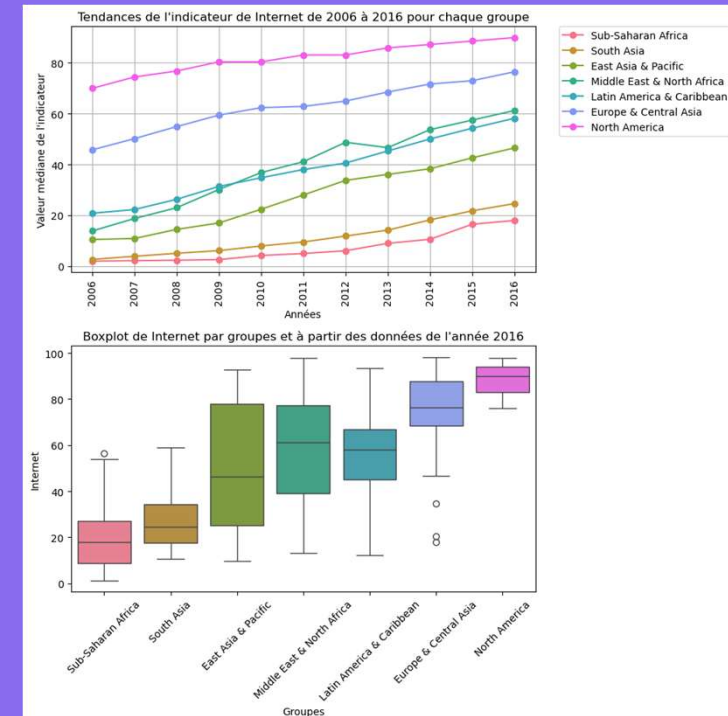
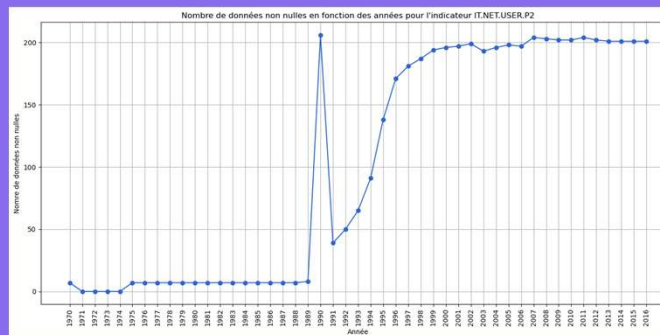
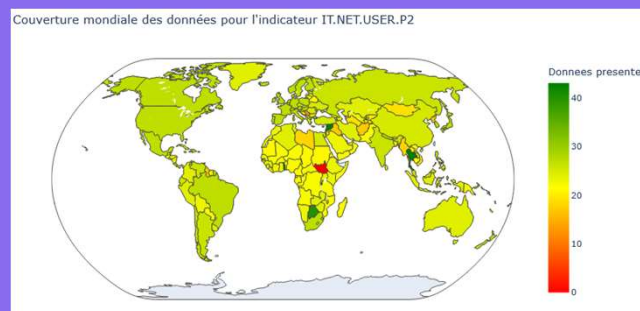
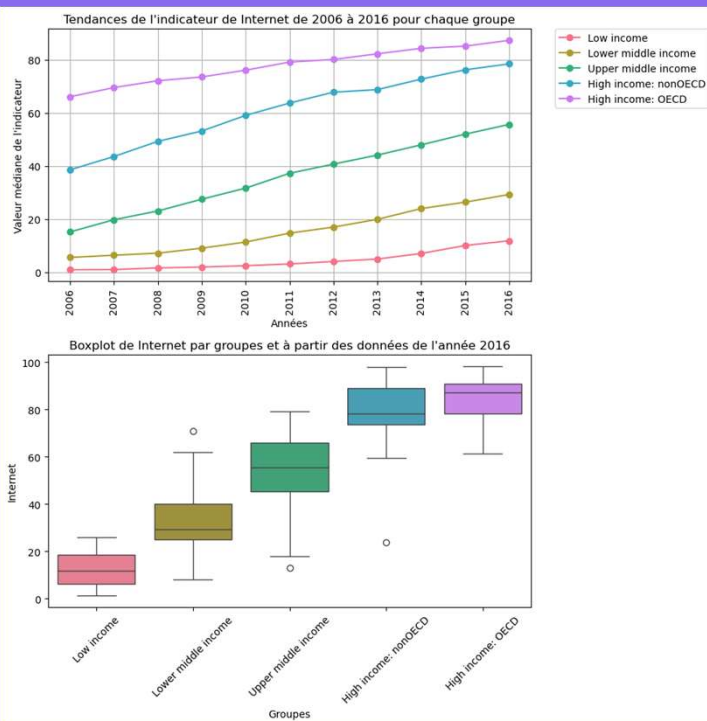
Représentativité mondiale

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)

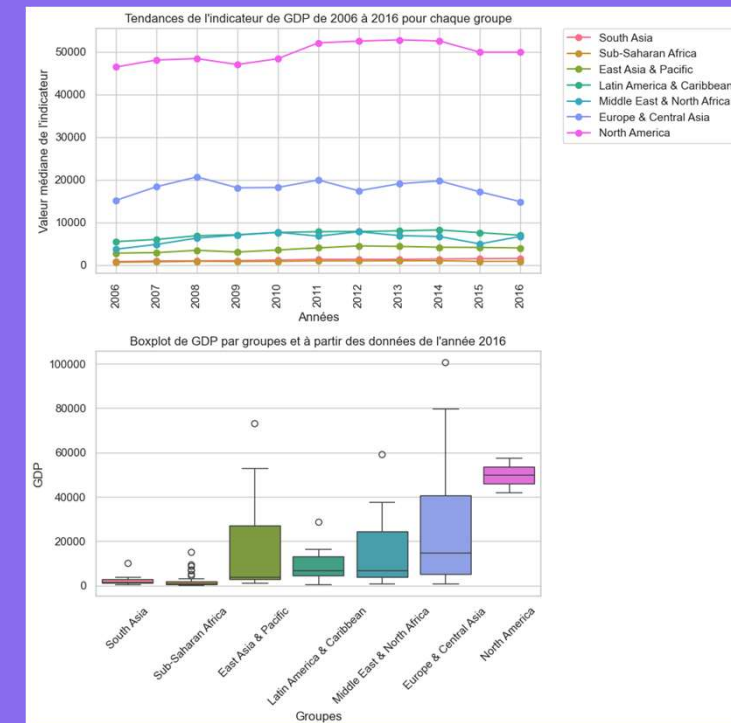
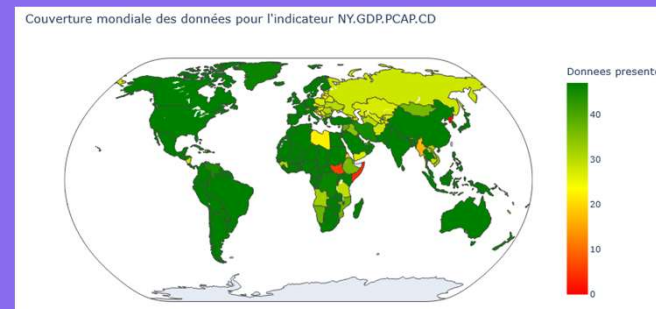
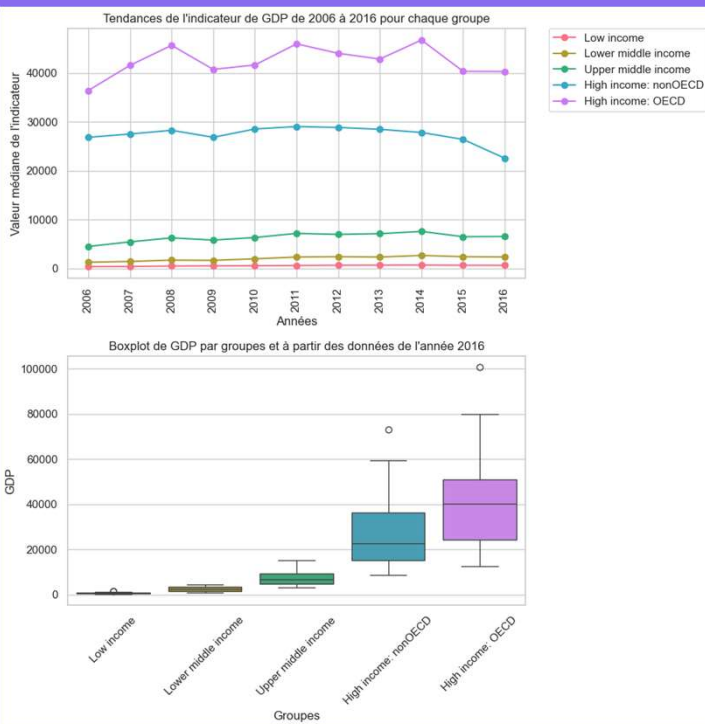
3. Stratégie de sélection des indicateurs

Analyse des indicateurs pré-sélectionnés : Internet



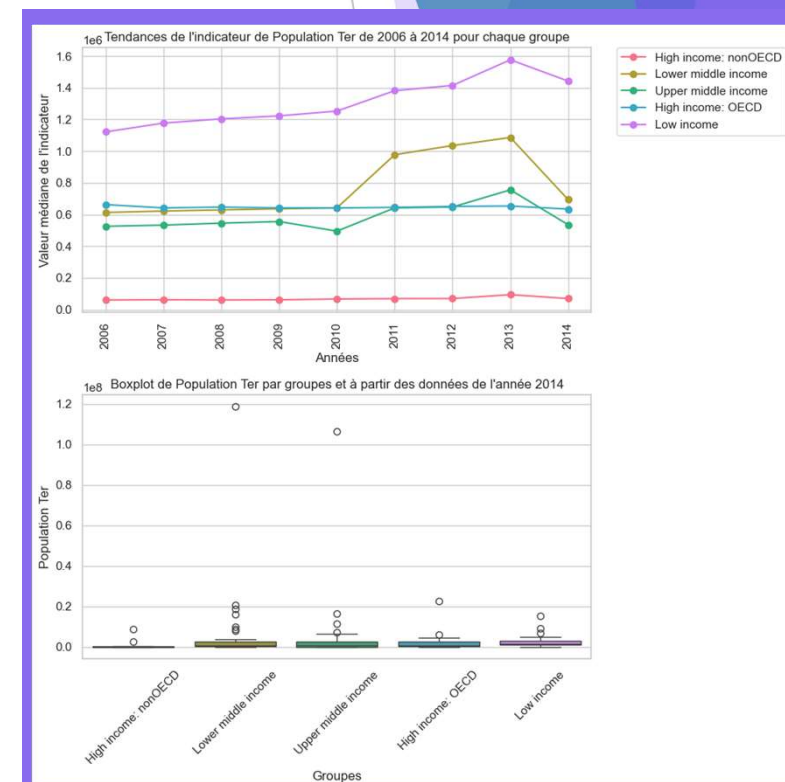
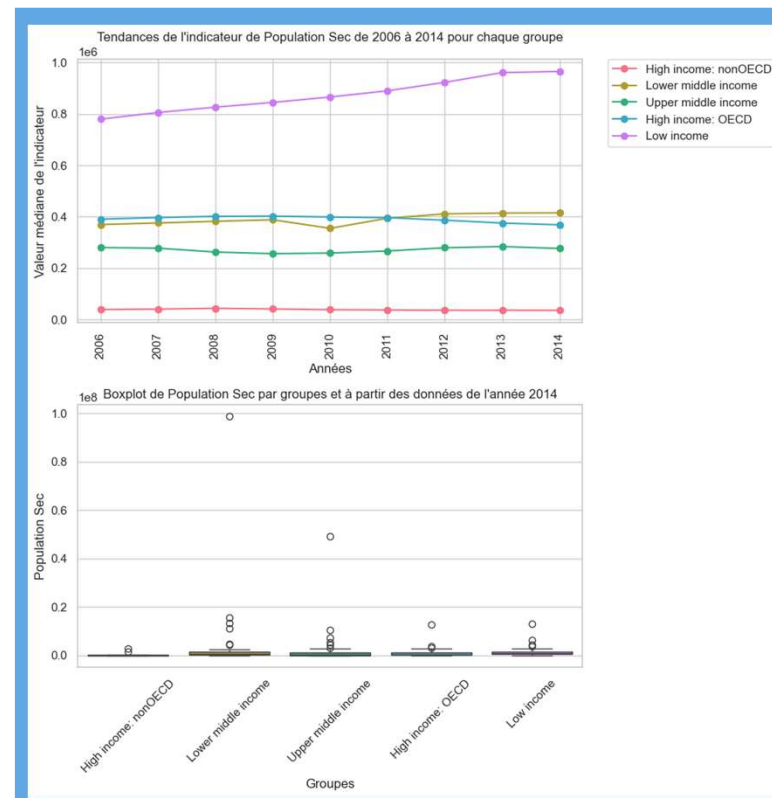
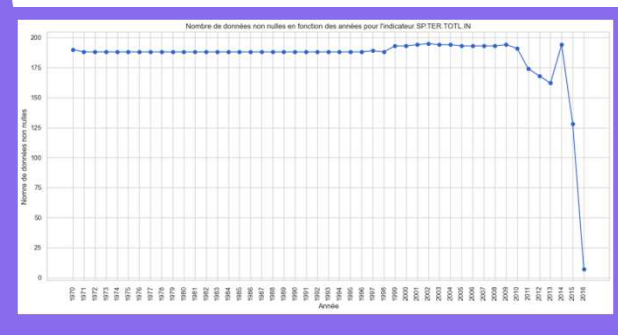
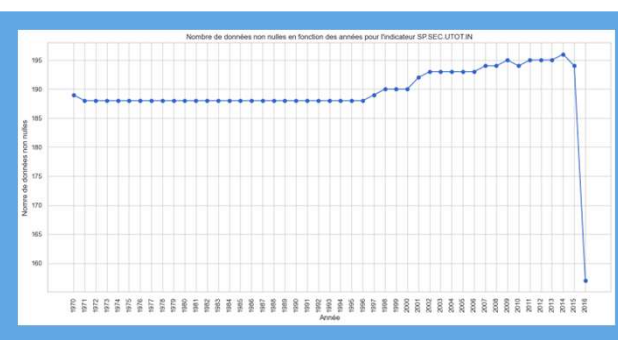
3. Stratégie de sélection des indicateurs

Analyse des indicateurs pré-sélectionnés : GDP



3. Stratégie de sélection des indicateurs

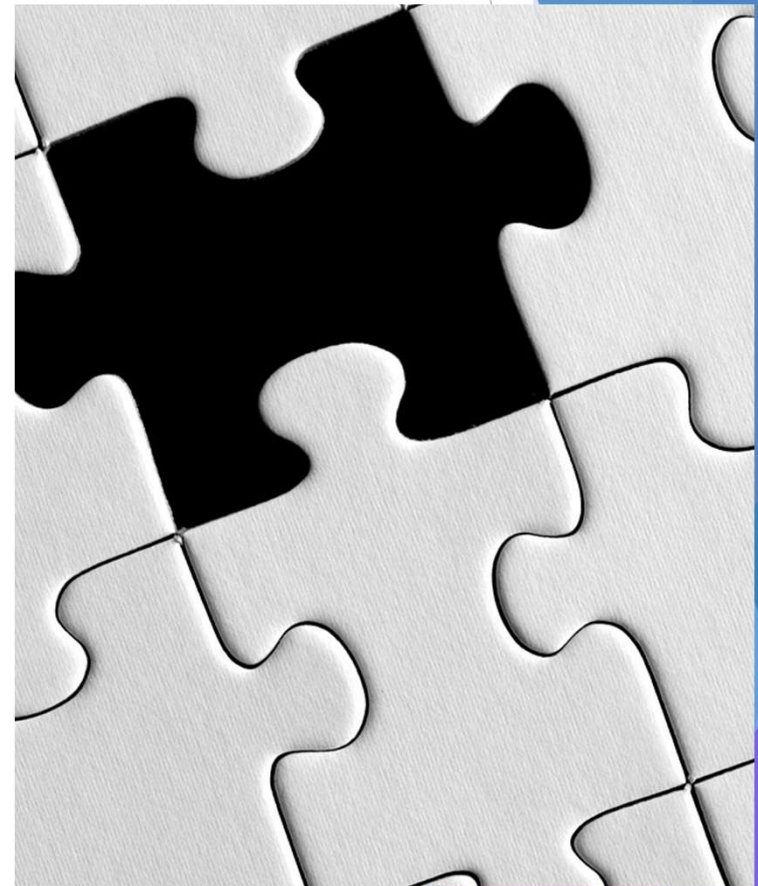
Analyse des indicateurs pré-sélectionnés : Population of the official age for upper secondary and tertiary education



3. Stratégie de sélection des indicateurs

Les indicateurs manquants

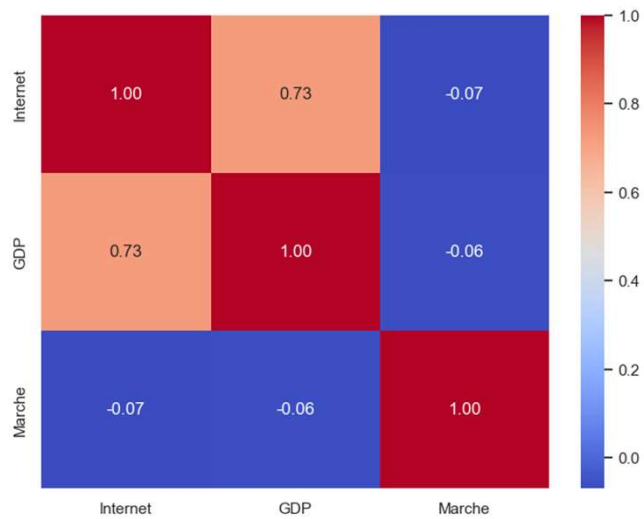
- Ordinateurs
- Langues
- Concurrence
- Alphabétisation
- Utilisation de MOOC



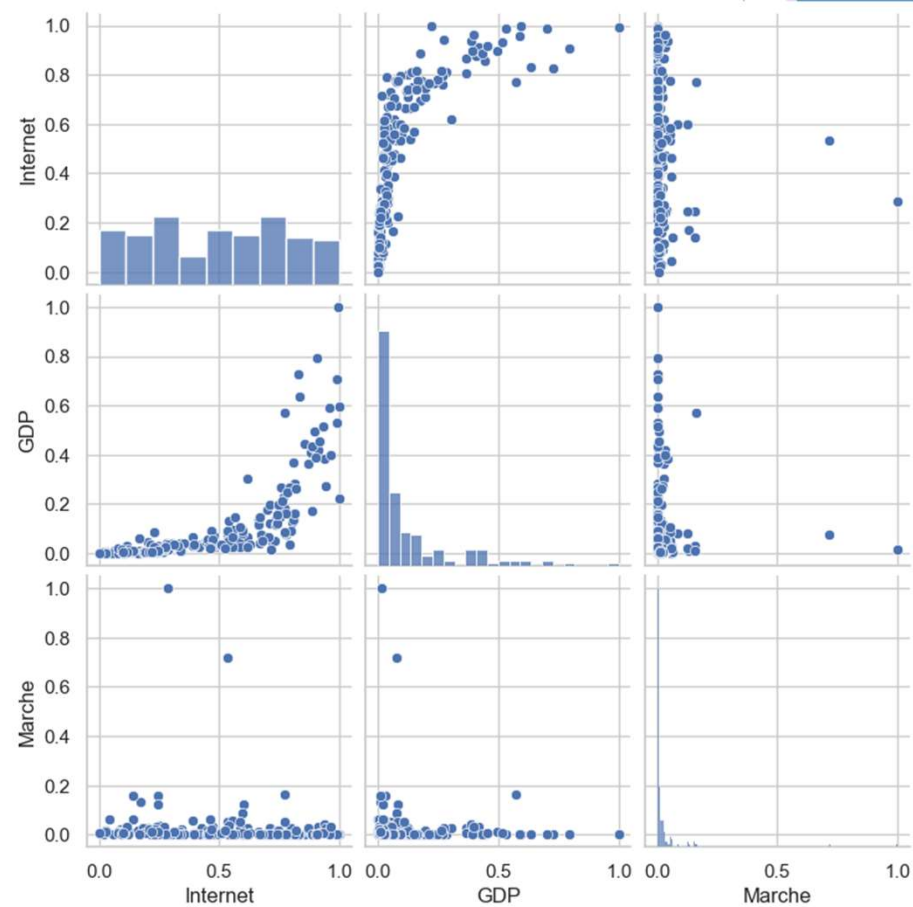
4. Analyse des données



Corrélations entre les indicateurs

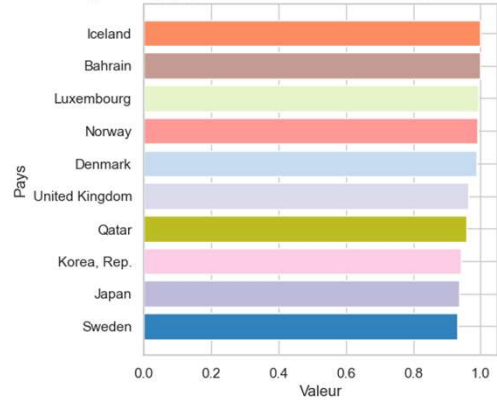


Analyse bivariable

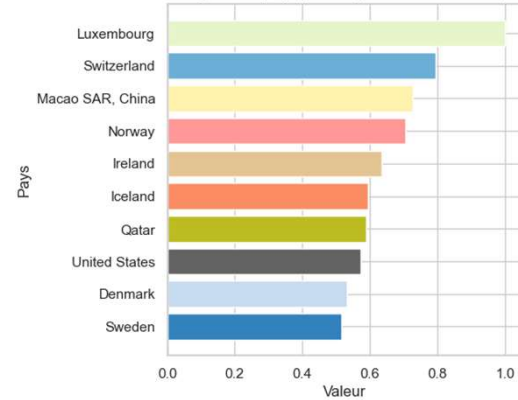


5. Résultats

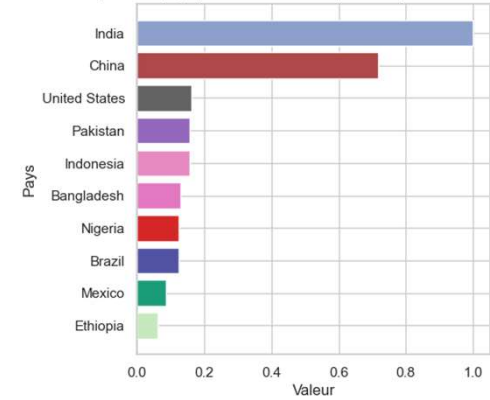
Top 10 des pays par taux de pénétration d'Internet (Valeurs normalisées)



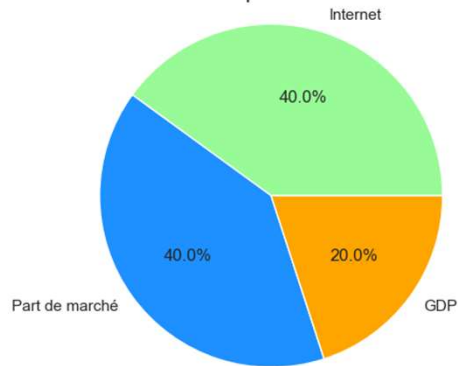
Top 10 des pays par PIB (Valeurs normalisées)



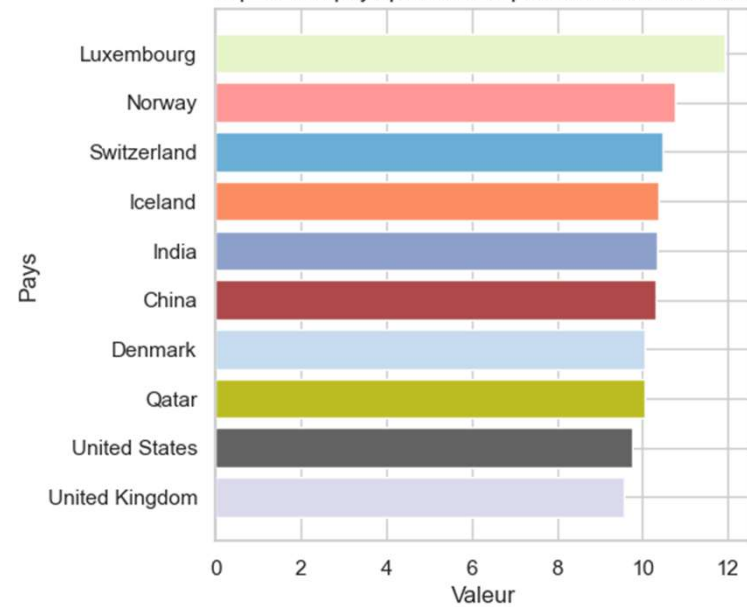
Top 10 des pays par taille de marché potentiel (Valeurs normalisées)



Distribution des poids des indicateurs



Top 10 des pays pour une expansion à l'international



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 ?

