

Projet d'expansion à l'international



Sommaire



- ❖ Problématique
- ❖ Analyser les données de la Banque
- ❖ Description des données
- ❖ Le jeux de données principal 'EdStatsData.csv'
- ❖ Pré-traitement des données
- ❖ Indicateurs retenus et leur statistique
- ❖ Scoring
- ❖ Conclusion
- ❖ Suite du Projet



Problématique



Academy souhaite élargir leur marché à l'international pour la formation en ligne de niveau lycée et universitaire.



- 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 ?



Analyser les données de la banque mondiale du portail EdStat

- Informations pertinentes? De qualité? Cohérentes?
- Indicateurs statistiques pour départager les pays?
- Conclusion : pays à fort potentiel ? Évolution? Priorisation?



LA BANQUE MONDIALE
BIRD • IDA

- [Table des Matieres](#)
- [1. Introduction](#)
- [2. Bibliothèques d'analyse Python](#)
- [3. Analyse des jeux de données](#)
 - [3.1. Analyse du jeu de données EdStatsCountry.csv](#)
 - [3.2. Analyse du jeu de données EdStatsData.csv](#)
 - [3.3. Analyse du jeu de données EdStatsCountry-Series.csv](#)
 - [3.4. Analyse du jeu de données EdStatsSeries.csv](#)
 - [3.4. Analyse du jeu de données EdStatsFootNote.csv](#)
- [4. Bilan des jeux de données](#)
- [5. Selection des indicateurs Pertinents](#)
- [6. Extraction du Dataframe final](#)
- [7. Analyse des indicateurs retenus](#)
- [8. Scoring](#)
 - [8.1. Normalisation des données](#)
 - [8.2. Calcul du score](#)
- [9. Conclusion](#)



Description



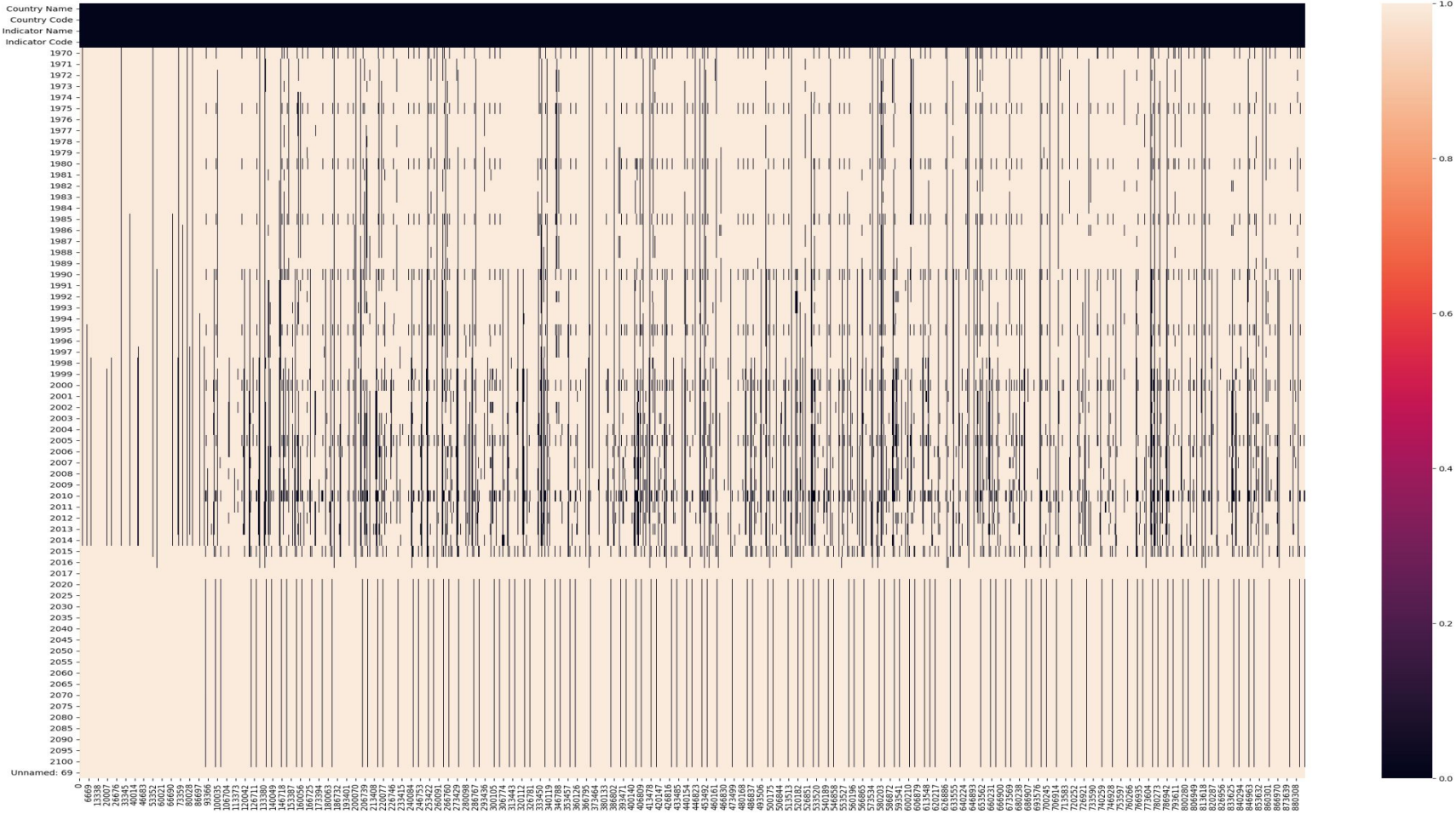
LA BANQUE MONDIALE
BIRD • IDA



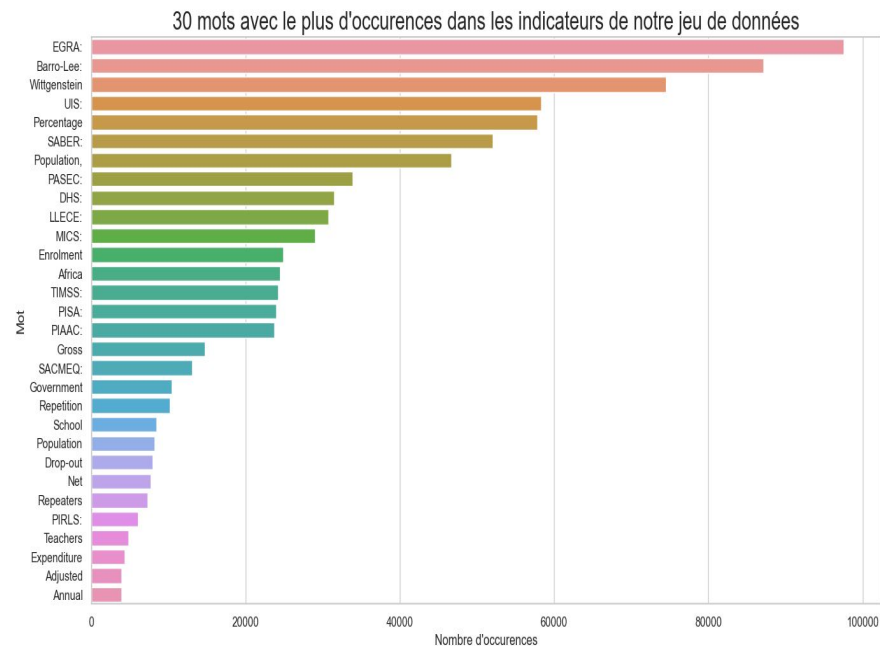
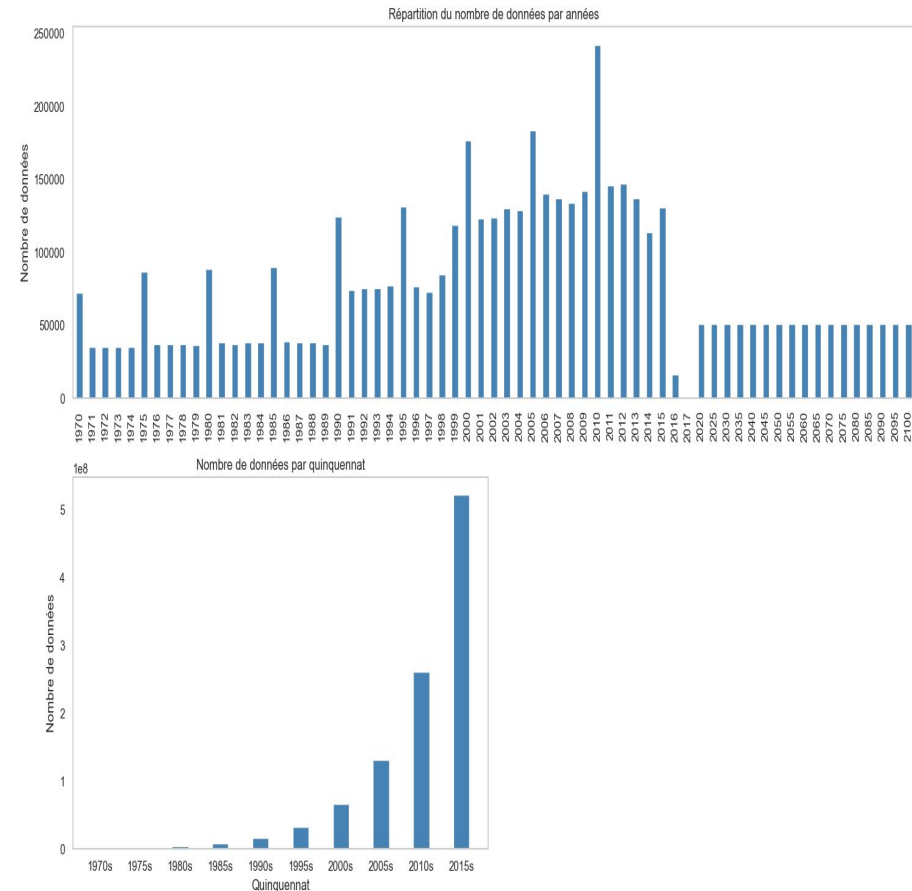
BANQUE MONDIALE

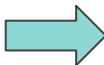
	Nb lignes	Nb colonnes	Taux remplissage moyen	Doublons	Description
EdStatsCountry.csv	241	32	69.8%	0	Liste des pays avec leurs données principales
EdStatsSeries.csv	3665	21	28.3%	0	Liste des indicateurs avec description, unité, période, etc...
EdStatsData.csv	886930	70	13.9%	0	Données de chaque indicateur par pays et par année
EdStatsCountry-Series.csv	613	4	75.0%	0	Description des différentes séries de données (majoritairement provenance)
EdStatsFootNote.csv	643638	5	80.0%	0	Commentaire pour chaque couple série de données / pays

Le jeu de données principal EdStatsData.csv



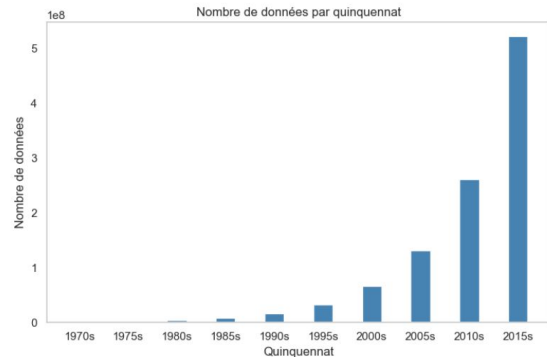
Pré-traitement des jeux de données



	count	mean	std	min	25%	50%	75%	max	
2005	3665.0	47.932060	65.292201	0.0	0.0	12.0	97.0	215.0	
2006	3665.0	35.979536	56.229376	0.0	0.0	1.0	55.0	215.0	
2007	3665.0	35.143520	57.131142	0.0	0.0	0.0	54.0	215.0	
2008	3665.0	34.347340	58.223390	0.0	0.0	0.0	58.0	215.0	
2009	3665.0	36.440928	58.455577	0.0	0.0	0.0	67.0	215.0	
2010	3665.0	63.735334	70.015971	0.0	0.0	13.0	144.0	215.0	
2011	3665.0	37.499591	57.961491	0.0	0.0	2.0	63.0	215.0	
2012	3665.0	37.773533	57.387026	0.0	0.0	6.0	64.0	214.0	
2013	3665.0	35.112688	57.002376	0.0	0.0	1.0	55.0	214.0	
2014	3665.0	28.825102	51.850513	0.0	0.0	1.0	33.0	210.0	
2015	3665.0	35.271760	63.345214	0.0	0.0	1.0	34.0	210.0	
TOUT	3665.0	428.061392	575.813483	0.0	26.0	259.0	570.0	2353.0	961 indicateurs

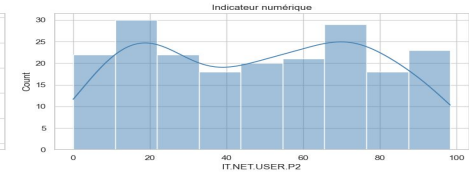
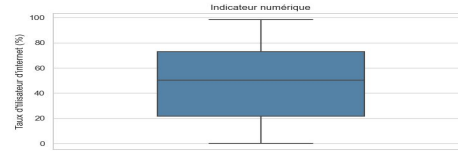
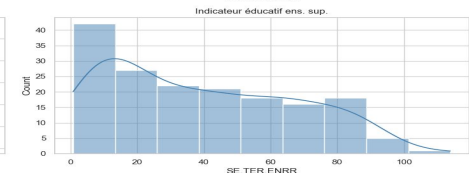
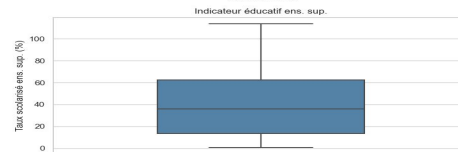
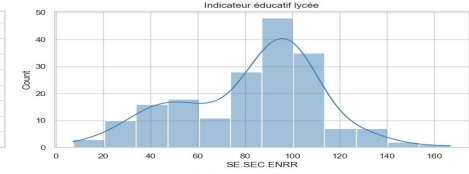
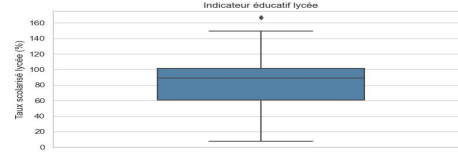
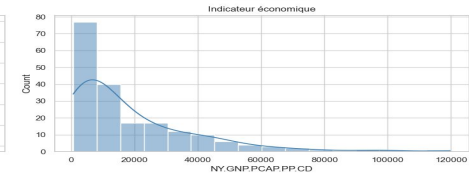
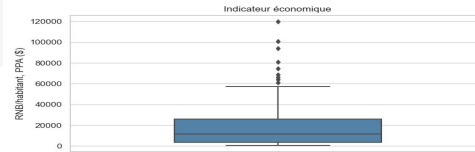
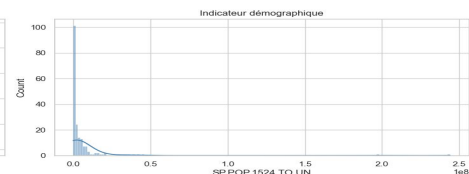
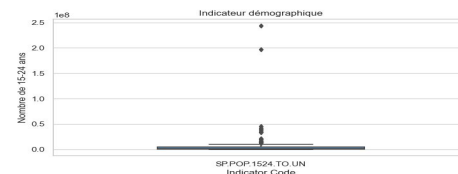
Indicateurs retenus et leur statistique

Démographique	<ul style="list-style-type: none">15 : pour la cible 15-19ans20 : pour la cible 20-24ans	SP.POP.1524.TO.UN
Economique	<ul style="list-style-type: none">NY : Nation Accounts	NY.GDP.PCAP.PP.KD
Educatif	<ul style="list-style-type: none">SEC : pour les lycéensTER : pour le niveau universitaire	<ul style="list-style-type: none">SE.SEC.ENRRSE.TER.ENRR
Numérique	<ul style="list-style-type: none">IT : pour l'accès aux infrastructures techniques	IT.NET.USER.P2



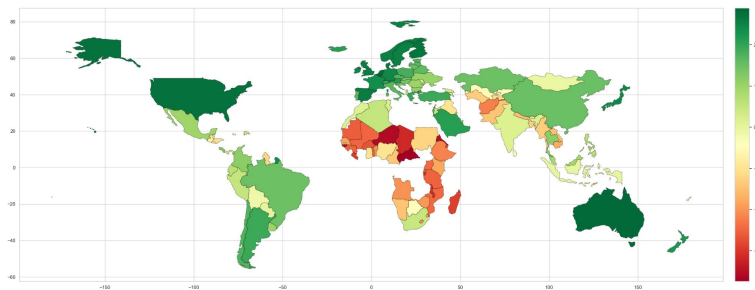
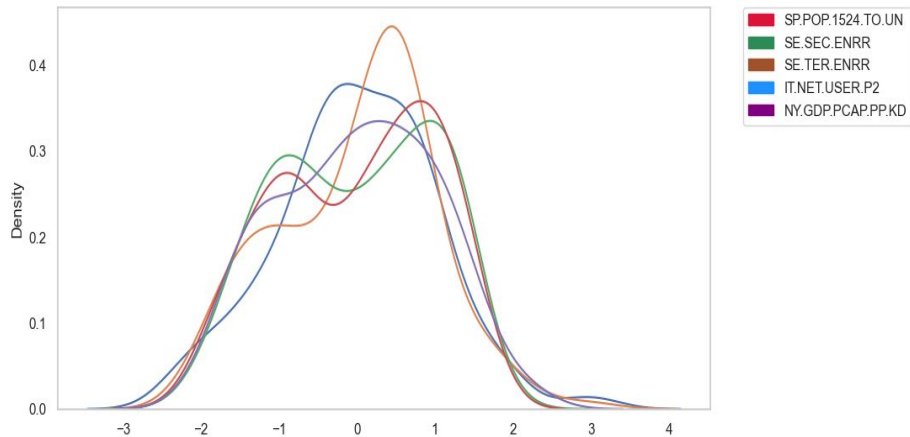
Plage 2005-2015

Indicator Code		Topic	Long definition	Pourcentage
0	IT.NET.USER.P2	Internet users (per 100 people)	Internet users are individuals who have used the Internet (from any location) in the last 3 months. The Internet can be used via a computer, mobile phone, personal digital assistant, games machine, digital TV etc.	83.21
8	NY.GDP.PCAP.PP.KD	GDP per capita, PPP (constant 2011 international \$)	GDP per capita based on purchasing power parity (PPP). PPP GDP is gross domestic product converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GDP as the U.S. dollar has in the United States. GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in constant 2011 international dollars.	77.91
104	SE.SEC.ENRR	Gross enrolment ratio, secondary, both sexes (%)	Total enrollment in secondary education, regardless of age, expressed as a percentage of the population of official secondary education age. GER can exceed 100% due to the inclusion of over-aged and under-aged students because of early or late school entrance and grade repetition.	59.20
129	SE.TER.ENRR	Gross enrolment ratio, tertiary, both sexes (%)	Total enrollment in tertiary education (ISCED 5 to 8), regardless of age, expressed as a percentage of the total population of the five-year age group following on from secondary school leaving.	51.95
296	SP.POP.1524.TO.UN	Population, ages 15-24, total	Population, ages 15-24, total is the total population age 15-24.	75.39

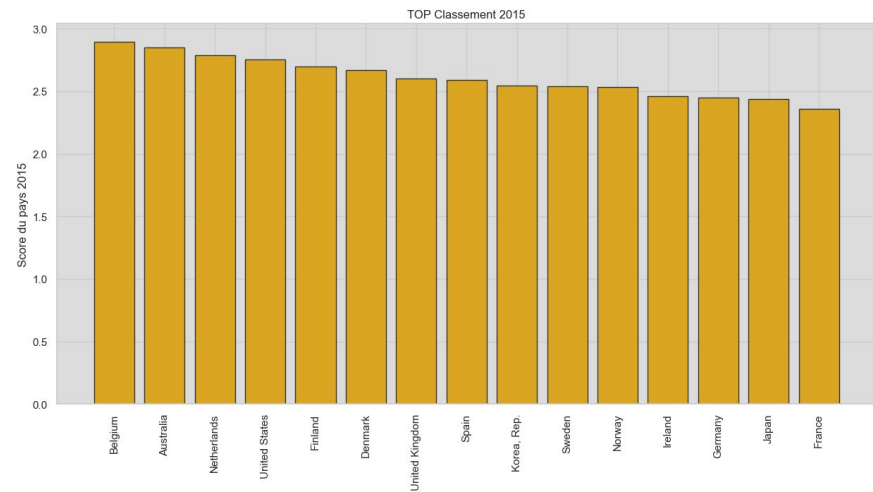


Scoring

Normalisation :
PowerTransform()

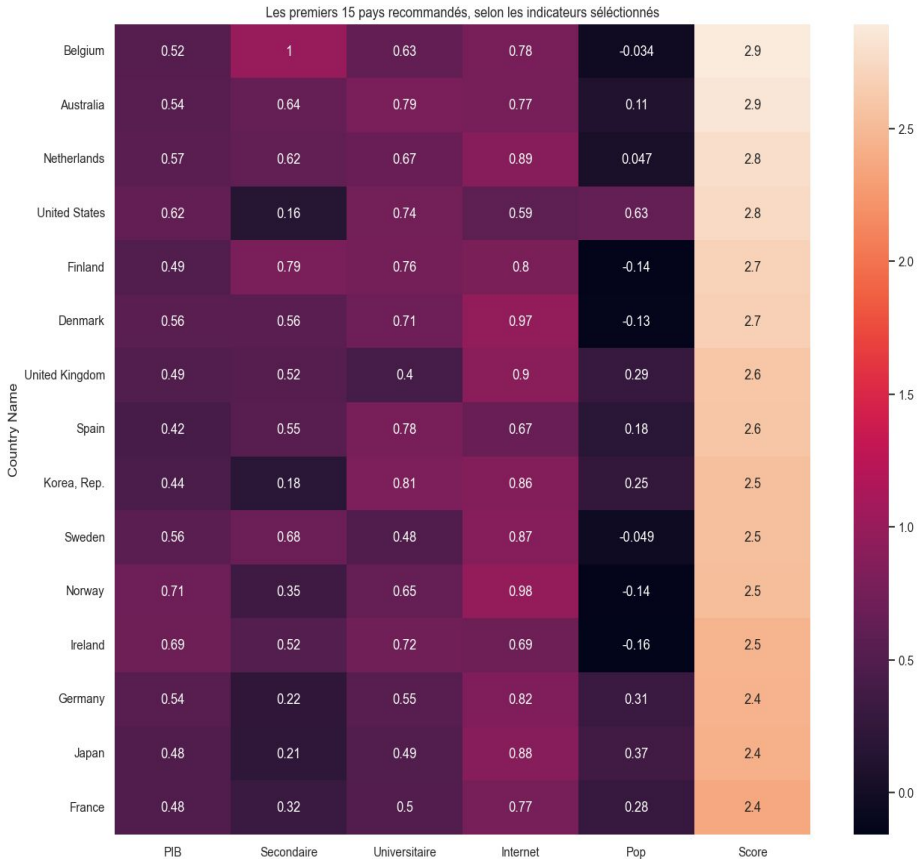


- On divise chaque valeur par le maximum de sa colonne de façon à avoir des valeurs entre 0 et 1.
- On va ensuite calculer la somme pondérée des colonnes pour avoir le score du pays.



Conclusion

	Pays	Score
Priorité		
1	Belgium	2.893017
2	Australia	2.851098
3	Netherlands	2.787149
4	United States	2.755184
5	Finland	2.693621
6	Denmark	2.670662
7	United Kingdom	2.601654
8	Spain	2.588654
9	Korea, Rep.	2.541621
10	Sweden	2.538684
11	Norway	2.535101
12	Ireland	2.462142
13	Germany	2.448108
14	Japan	2.437168
15	France	2.356894





Suite du projet

- Proposer un modèle de prédictions pour consolider notre travail
- Vérifier l'hypothèse de travail avec les décideurs
- Stratégie d'entreprise : langue, marché cible, proximité géographique d'implantation
- Autres indicateurs, Concurrence
- Forte impression des pays d'Asie de l'Est et du Pacifique