

Projet n°2 : Analyse des données de systèmes éducatifs





Sommaire

Problématique et jeu de données

Analyse pré-exploratoire

Conclusions



Problématique

Projet d'expansion à l'international de Start-up de la EdTec (Academy)

Objectif : Étude de marché pour ouverture à l'international

- 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 ?

Jeu de données

EdStatsCountry.csv	Information globales sur l'économie de chaque pays du monde. 241 lignes, 32 colonnes : Quelques valeurs manquantes, aucun doublon
EdStatsCountrySeries.csv	Informations sur les données contenues dans EdStatsCountry.csv 613 lignes, 4 colonnes : Pas de données manquantes, aucun doublon (sauf dernière colonne).
EdStatsData.csv	Information sur l'évolution de nombreux indicateurs pour les pays et zone géographique 886930 lignes, 70 colonnes : Beaucoup de données manquantes, aucun doublon
EdStatsFootNote.csv	Information sur les années d'origine des données. 643 638 lignes, 5 colonnes : Pas de valeurs manquante (sauf dernière colonne) et aucun doublon.
EdStatsSeries.csv	Informations sur les indicateurs 3665 lignes, 21 colonnes : Beaucoup de données manquantes, aucun doublon.



Analyse pré-exploratoire

Familiarisation et prétraitement des données

Identification des variables et indicateurs pertinents

Comparaison entre pays/données



Analyse pré-exploratoire

2 datasets sont intéressants :

EdStatsCountry, EdStatsData

Plusieurs variables seront intéressantes :

Pays, Région, Indicateurs et les différentes années

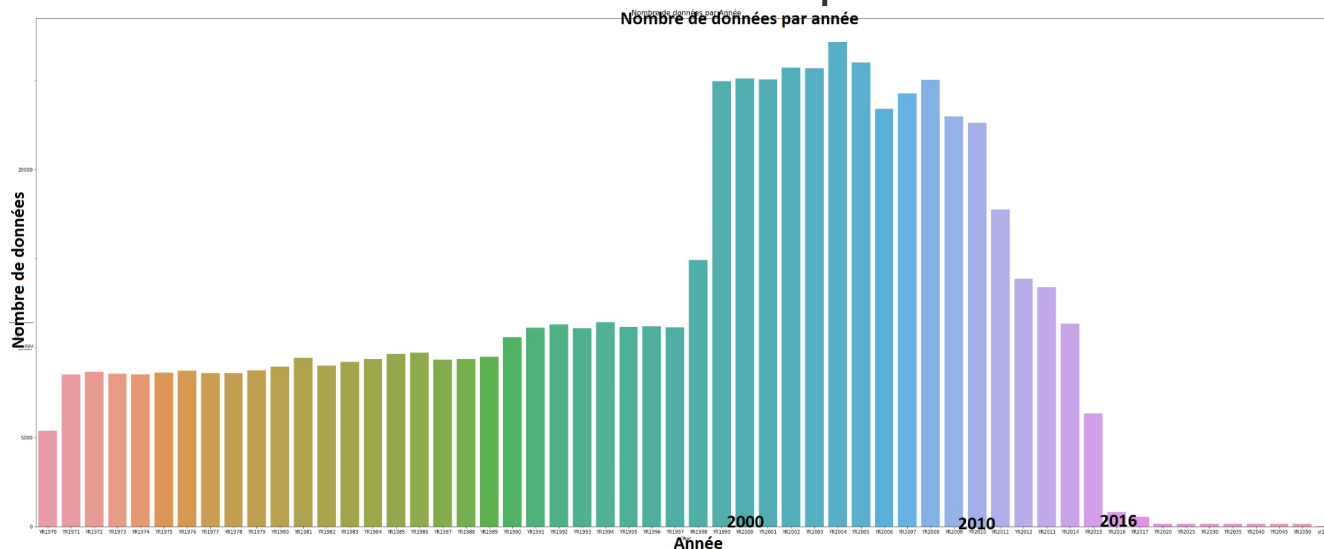
Analyse pré-exploratoire

Séries temporelles (1970-2016) par pays, pour chaque indicateurs

3665 indicateurs

241 pays

Nombre de données par année :



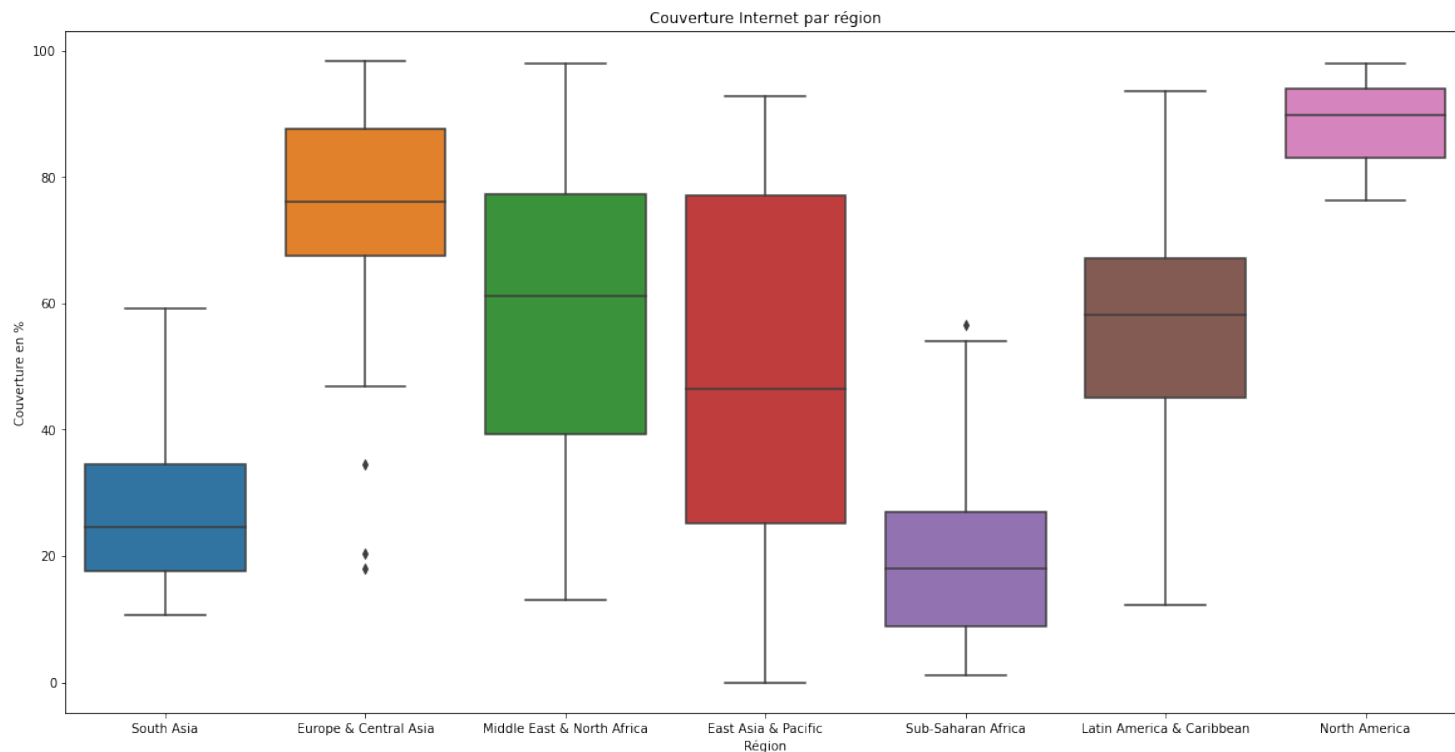
Analyse pré-exploratoire

Les indicateurs gardés :

```
listeIndicateurGardés = ['Population, ages 15-24, total',  
                        'Internet users (per 100 people)',  
                        'Government expenditure on education as % of GDP (%)',  
                        'Enrolment in tertiary education, all programmes, both sexes (number)',  
                        'Enrolment in upper secondary education, both sexes (number)',  
                        'GDP per capita (current US$)',  
                        'Unemployment, total (% of total labor force)']
```


Analyse pré-exploratoire

Couverture internet par région



Analyse pré-exploratoire

Première méthode

	deniereValeur
0,25	25,336288
0,50	53,613386
0,75	76,234824

['Andorra', 'Aruba', 'Australia', 'Austria', 'Azerbaijan', 'The Bahamas', 'Bahrain', 'Barbados', 'Belgium', 'Bermuda', 'Canada', 'Cayman Islands', 'Czech Republic', 'Denmark', 'Estonia', 'Faeroe Islands', 'Finland', 'France', 'Germany', 'Gibraltar', 'Guam', 'Hong Kong SAR, China', 'Hungary', 'Iceland', 'Ireland', 'Israel', 'Japan', 'Kazakhstan', 'Korea', 'Kuwait', 'Latvia', 'Liechtenstein', 'Luxembourg', 'Macao SAR, China', 'Malaysia', 'Malta', 'Monaco', 'Netherlands', 'New Zealand', 'Norway', 'Puerto Rico', 'Qatar', 'Russia', 'Singapore', 'Slovak Republic', 'Spain', 'St. Kitts and Nevis', 'Sweden', 'Switzerland', 'United Arab Emirates', 'United Kingdom']

Analyse pré-exploratoire

Pays potentiellement intéressants

['Cayman Islands', 'Kazakhstan', 'Germany', 'Ecuador', 'Lesotho', 'Mexico', 'Austria', 'Estonia', 'Morocco', 'Tunisia', 'Cabo Verde', 'Slovak Republic', 'Ethiopia', 'Malta', 'Uruguay', 'Venezuela', 'Argentina', 'Egypt', 'Mauritius', 'Serbia', 'Iceland', 'Iraq', 'Moldova', 'France', 'Afghanistan', 'Aruba', 'Peru', 'Dem. Rep. Congo', 'Kenya', 'Gabon', 'Colombia', 'Sweden', 'Portugal', 'Maldives', 'Iran', 'Sri Lanka', 'The Bahamas', 'Oman', 'Ukraine', 'St. Vincent and the Grenadines', 'New Caledonia', 'Bhutan', 'Costa Rica', 'Hungary', 'Isle of Man', 'Tanzania', 'Côte d'Ivoire', 'Switzerland', 'Togo', 'Lithuania', 'Poland', 'Nigeria', 'St. Kitts and Nevis', 'Greece', 'Solomon Islands', 'Tajikistan', 'Japan', 'China', 'Micronesia', 'Armenia', 'Greenland', 'Pakistan', 'Romania', 'Croatia', 'Mongolia', 'Cuba', 'Uzbekistan', 'Curaçao', 'Ireland', 'Belize', 'Finland', 'United Kingdom', 'Turkey', 'Vanuatu', 'Georgia', 'Yemen', 'Timor-Leste', 'Monaco', 'Vietnam', 'Canada', 'Dominican Republic', 'Trinidad and Tobago', 'Haiti', 'United Arab Emirates', 'Jamaica', 'Malaysia', 'Guatemala', 'Botswana', 'Burkina Faso', 'Niger', 'Kyrgyz Republic', 'Myanmar', 'Netherlands', 'The Gambia', 'Zimbabwe', 'Paraguay', 'Latvia', 'Puerto Rico', 'Guyana', 'Cameroon', 'Faeroe Islands', 'Dem. People's Rep. Korea', 'Guam', 'Qatar', 'Algeria', 'Libya', 'New Zealand', 'Jordan', 'Benin', 'Israel', 'Nicaragua', 'Northern Mariana Islands', 'Ghana', 'South Africa', 'Sudan', 'Slovenia', 'Gibraltar', 'Uganda', 'Cyprus', 'Djibouti', 'Bahrain', 'Singapore', 'Virgin Islands', 'Australia', 'Senegal', 'Congo', 'Brazil', 'Chile', 'Spain', 'Malawi', 'Syrian Arab Republic', 'Macedonia', 'Belgium', 'Comoros', 'Fiji', 'Burundi', 'Thailand', 'Liechtenstein', 'Mozambique', 'West Bank and Gaza', 'Korea', 'Indonesia', 'Czech Republic', 'Denmark', 'India', 'Bangladesh', 'Barbados', 'Bosnia and Herzegovina', 'Italy', 'Nepal', 'Macao SAR, China', 'United States', 'Bermuda', 'Luxembourg', 'Namibia', 'Kuwait', 'Hong Kong SAR, China', 'Philippines', 'Albania', 'São Tomé and Príncipe', 'St. Lucia', 'French Polynesia', 'Norway', 'Azerbaijan', 'Montenegro', 'Bolivia', 'Brunei', 'Russia', 'Saudi Arabia', 'Swaziland', 'Bulgaria', 'Belarus', 'Honduras', 'San Marino', 'Andorra']

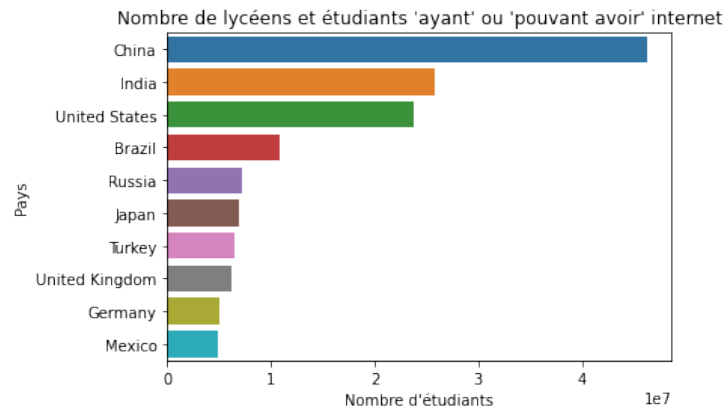
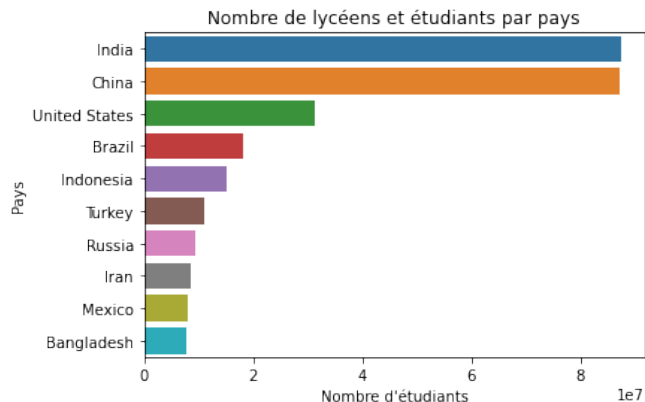
```
population1524_075 = 4945440.0
internetUsers_075 = 76.409085
gdpE_075 = 3.819358
tertiaryEducation_075 = 256385.5
uSecondaryEducation = 659084.0
gdpCapita_075 = 15891.626549
unemployment_075 = 11.375
```

Analyse pré-exploratoire

Comparaison entre pays :

En fonction des indicateurs gardés et des données des différentes années, nous pouvons définir les pays qui sont à prioriser

	Nom Pays	Nombre Etudiant	Etudiant_Internet
239439	China	87076616.00	4.632476e+07
411694	India	87336096.00	2.580534e+07
844164	United States	31268043.00	2.381897e+07
188129	Brazil	18235058.00	1.088318e+07
675574	Russia	9415420.00	7.194236e+06
444679	Japan	7545380.00	6.941750e+06
814844	Turkey	11058509.00	6.452389e+06
840499	United Kingdom	6548014.25	6.205933e+06
353054	Germany	5557733.25	4.982347e+06
554629	Mexico	8101727.00	4.823804e+06



Analyse pré-exploratoire

Deuxième méthode

	deniereValeur
0,25	25,336288
0,50	53,613386
0,75	76,234824

```
['Australia',  
'France',  
'Germany',  
'Japan',  
'Korea',  
'Russia',  
'Spain',  
'United Kingdom']
```

Analyse pré-exploratoire

Pays à fort potentiel pouvant répondre à la problématique :

['Argentina', 'Australia', 'Brazil', 'China', 'Colombia', 'Egypt', 'France', 'Germany', 'India', 'Indonesia', 'Iran', 'Italy', 'Japan', 'Korea', 'Mexico', 'Philippines', 'Poland', 'Russia', 'Saudi Arabia', 'South Africa', 'Spain', 'Thailand', 'Turkey', 'United Kingdom', 'United States']

	Nom Pays	Nombre Etudiant	Etudiant_Internet
239439	China	87076616.00	4.632476e+07
411694	India	87336096.00	2.580534e+07
844164	United States	31268043.00	2.381897e+07
188129	Brazil	18235058.00	1.088318e+07
675574	Russia	9415420.00	7.194236e+06
444679	Japan	7545380.00	6.941750e+06

Mais quid de l'avenir ?

Analyse pré-exploratoire

Prévoir l'avenir ?

	Nom Pays	2010- 2011_NombreEtudiant	2011- 2012_NombreEtudiant	2012- 2013_NombreEtudiant	2013- 2014_NombreEtudiant	2014- 2015_NombreEtudiant	2010- mean_NombreEtudiant
411855	India	9315764.0	6387928.0	-1306484.0	NaN	NaN	4.799069e+06
239600	China	112450.0	957738.0	3244428.0	5578904.0	9893520.0	2.473380e+06
815005	Turkey	619048.0	382077.0	1801753.0	NaN	NaN	9.342927e+05
415520	Indonesia	1028477.0	1300826.0	696822.0	159713.0	3185838.0	7.964595e+05
188290	Brazil	488430.0	358602.0	806956.0	NaN	NaN	5.513293e+05

	Nom Pays	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016	2010-2016	mean_InternetUsers
419185	Iran	3.10	3.7300	7.220000	9.400000	5.984976	7.891797	37.326773	6.221129
675735	Russia	6.00	14.8000	4.170000	2.550000	2.890000	2.999085	33.409085	5.568181
694060	Saudi Arabia	6.50	6.5000	6.500000	4.213657	4.902578	4.134668	32.750904	5.458484
653745	Philippines	4.00	7.2351	11.863924	1.500976	4.100000	1.800000	30.500000	5.083333
734375	South Africa	9.97	7.0300	5.500000	2.500000	2.919116	2.080884	30.000000	5.000000



Conclusion

Tous les pays sont traités.

Données utiles pour notre problématique.

Indicateurs et données manquants.

Merci de votre attention

