Projet 2

Analysez des données de systèmes éducatifs

Rédigé et Présenté par

GASSUC CEDRIC

Mentor : Oussama Elmaataoui

Problématiques

1 - Quels sont les pays avec un fort potentiel de clients pour une formation en ligne ?

2- Pour chacun de ces pays , quelle sera l'évolution ou la projection dans le temps ?

Plan

1. Choix des indicateurs
2. Nettoyage des données
3. Normalisation
4. Calcule des scores
5. Choix des pays favorables
6. Projection

1- Choix des indicateurs

- ✓ **Internet users (per 100 people)**
- ✓ **GDP per capita (current US$)**
- ✓ **Youth literacy rate, population 15-24 years, both sexes (%)**
- ✓ **Population, âges 15-24, total**

2- Nettoyage des données

- Suppression des données(Country Name) suivantes :
  - **Régions**
  - **Continents**
  - **Ville**
  - **Monde**

- Suppressions des années non exploitables
  - **Années : 1970 à 2009 et 2016 à 2100**

- Données exploitables
  - **Années : 2010, 2011, 2012, 2013, 2014, 2015**

2- Nettoyage des données

❖ Gestion des données manquantes NaN

- Supprime les valeurs manquantes(NaN)

```
import numpy as np

# Suppression des lignes avec des valeurs manquantes dans les colonnes numériques
data4 = data3.dropna(subset=data3.select_dtypes(include=[np.number]).columns, thresh=3)
data4
```

- Remplissage des NaN (**ffill, zéro, interpolation linéaire**)

```
data5 = data4.apply(lambda x: x.fillna(method='ffill') if x.dtype in [np.float64, np.int64] else x,axis = 1)
data5
```

❖ Moyenne, Médiane et Ecart-type

```
numeric_colonne = data5.select_dtypes(include='number')

# Calcul de la moyenne,médiane et ecart-type
data5['Moyenne'] = numeric_colonne.mean(axis=1)
data5['Médiane'] = numeric_colonne.median(axis=1)
data5['ecart_type'] = numeric_colonne.std(axis=1)
data6 = data5
data6
```

3- Normalisation

```
1 def min_max_scaling(column):
2     min_val = column.min()
3     scaled_values = (column - min_val) / (column.max() - min_val)
4     return scaled_values
5
6 normalisation = table.apply(min_max_scaling)
7 table1 = normalisation
8 table1
```

Indicator Name Country Name	GDP per capita (current US$)	Internet users (per 100 people)	Population, ages 15-24, total	Youth literacy rate, population 15-24 years, both sexes (%)
Afghanistan	0.002048	0.058906	0.027732	NaN
Albania	0.024895	0.580296	0.002387	NaN
Algeria	0.031763	0.211147	0.028535	NaN
Andorra	0.244899	0.936074	NaN	NaN
Angola	0.025620	0.079894	0.016649	NaN
...	...	...	...	...

Openclassrooms

4- Calcule des scores

```
table2["score"] = 4*table2["GDP per capita (current US$)"]
                + 5*table2["Internet users (per 100 people)"]
                + 3*table2["Youth literacy rate, population 15-24 years, both sexes (%)"]
                + 2*table2["Population, ages 15-24, total"]
table2
```

Indicator Name Country Name	GDP per capita (current US$)	Internet users (per 100 people)	Population, ages 15-24, total	Youth literacy rate, population 15-24 years, both sexes (%)	score
Afghanistan	0.002048	0.058906	0.027732	0.000000	0.358189
Albania	0.024895	0.580296	0.002387	0.000000	3.005835
Algeria	0.031763	0.211147	0.028535	0.000000	1.239856
Andorra	0.244899	0.936074	0.000000	0.000000	5.659967
Angola	0.025620	0.079894	0.016649	0.000000	0.535248
...	...	...	...	...	...
Vietnam	0.009556	0.390648	0.074106	0.000000	2.139676
West Bank and Gaza	0.015665	0.466910	0.003950	0.987259	5.366886
Yemen, Rep.	0.006980	0.194268	0.023625	0.000000	1.046512
Zambia	0.008689	0.149765	0.011952	0.000000	0.807484
Zimbabwe	0.004439	0.142667	0.013978	0.000000	0.759046

5- Choix des pays favorables

```
1 table3 = table2.sort_values(by='score', ascending = False).head(10)
2 table3
```

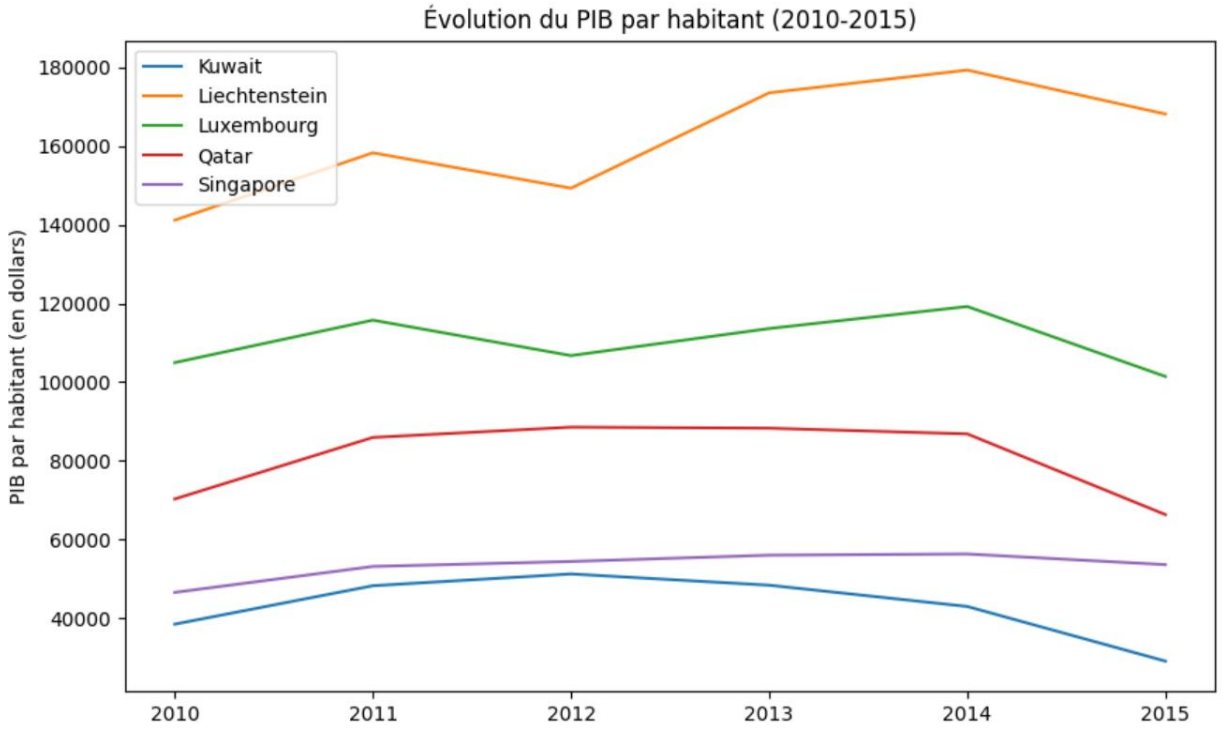
Indicator Name	GDP per capita (current US$)	Internet users (per 100 people)	Population, ages 15-24, total	Youth literacy rate, population 15-24 years, both sexes (%)	score
Country Name					
Qatar	0.528576	0.802048	0.000455	0.975649	9.052397
Liechtenstein	1.000000	0.950461	0.000000	0.000000	8.752307
Singapore	0.329914	0.783439	0.002769	0.997939	8.236206
Kuwait	0.278374	0.756965	0.001811	0.981583	7.846694
Luxembourg	0.674559	0.963528	0.000206	0.000000	7.516290
Spain	0.178859	0.733801	0.018194	0.993777	7.402161
Norway	0.605629	0.984160	0.002611	0.000000	7.348540
Bermuda	0.525263	0.968057	0.000000	0.000000	6.941334
Oman	0.125495	0.654368	0.002352	0.982867	6.727127
Switzerland	0.515841	0.889931	0.003944	0.000000	6.520903

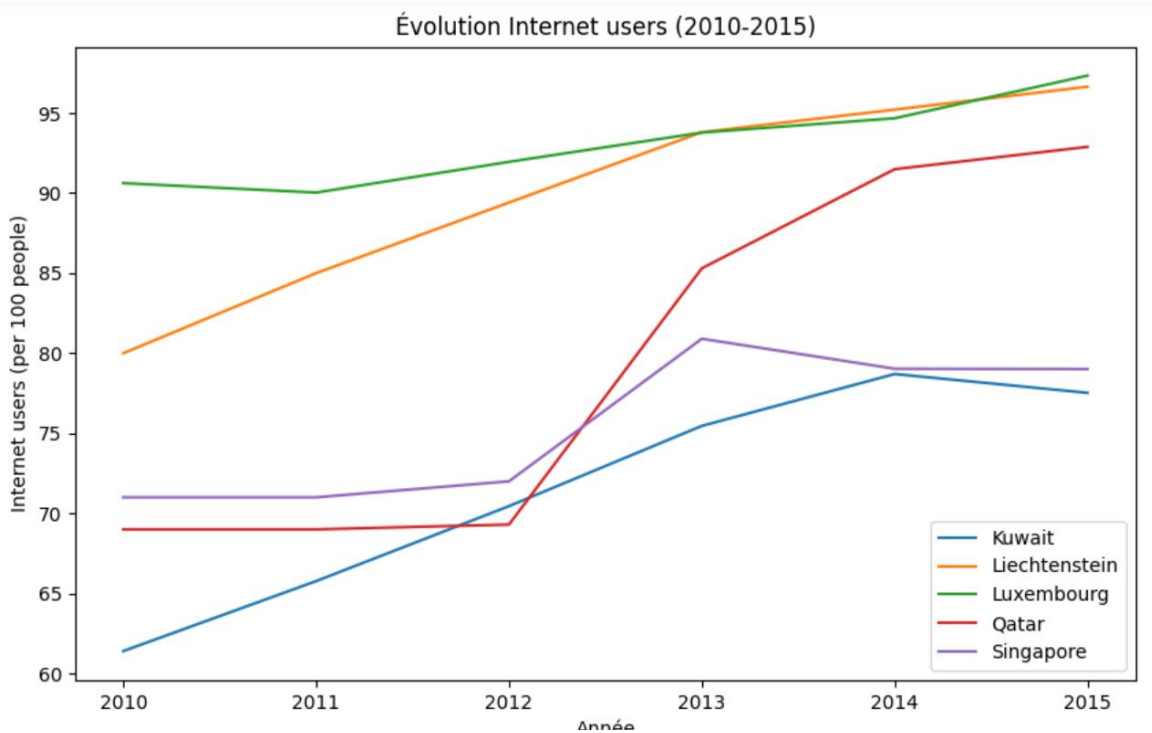
1er Cas

Top 5 des pays favorables à la distribution des cours en ligne

1* Qatar
2* Liechtenstein
3* Singapore
4* Kuwait
5* Luxembourg

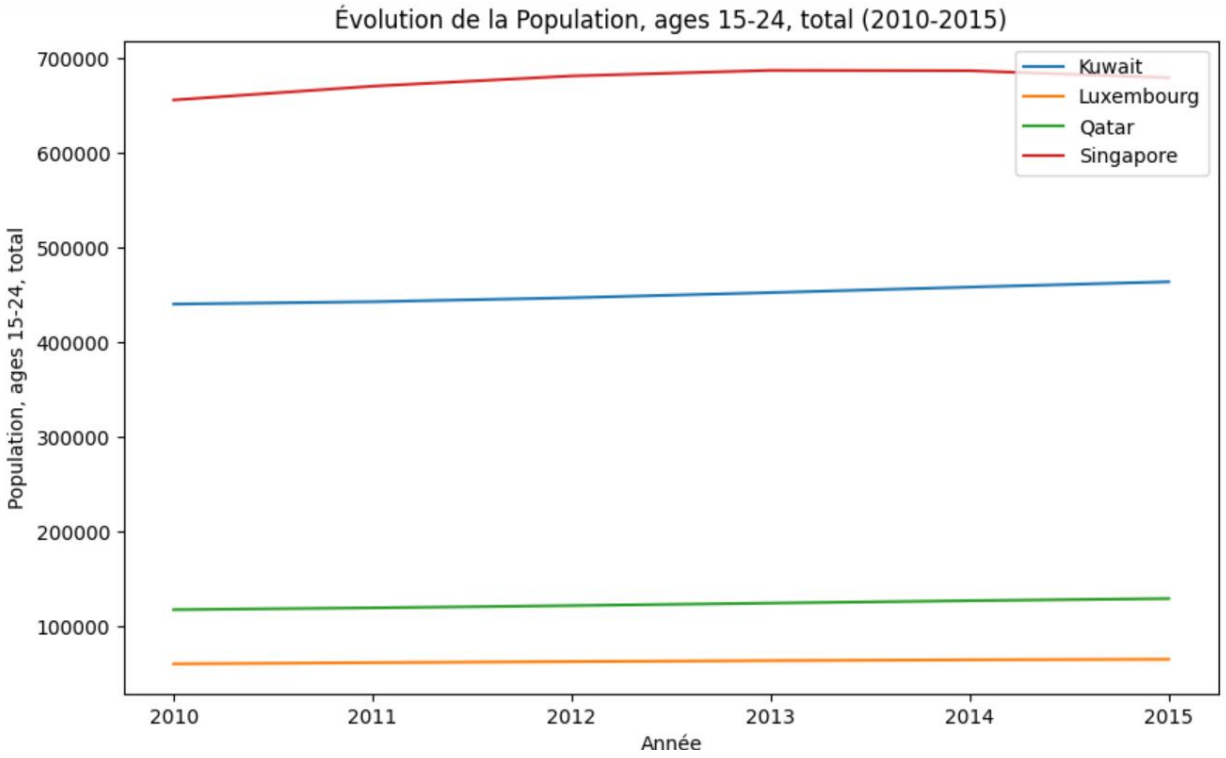
6- Projection

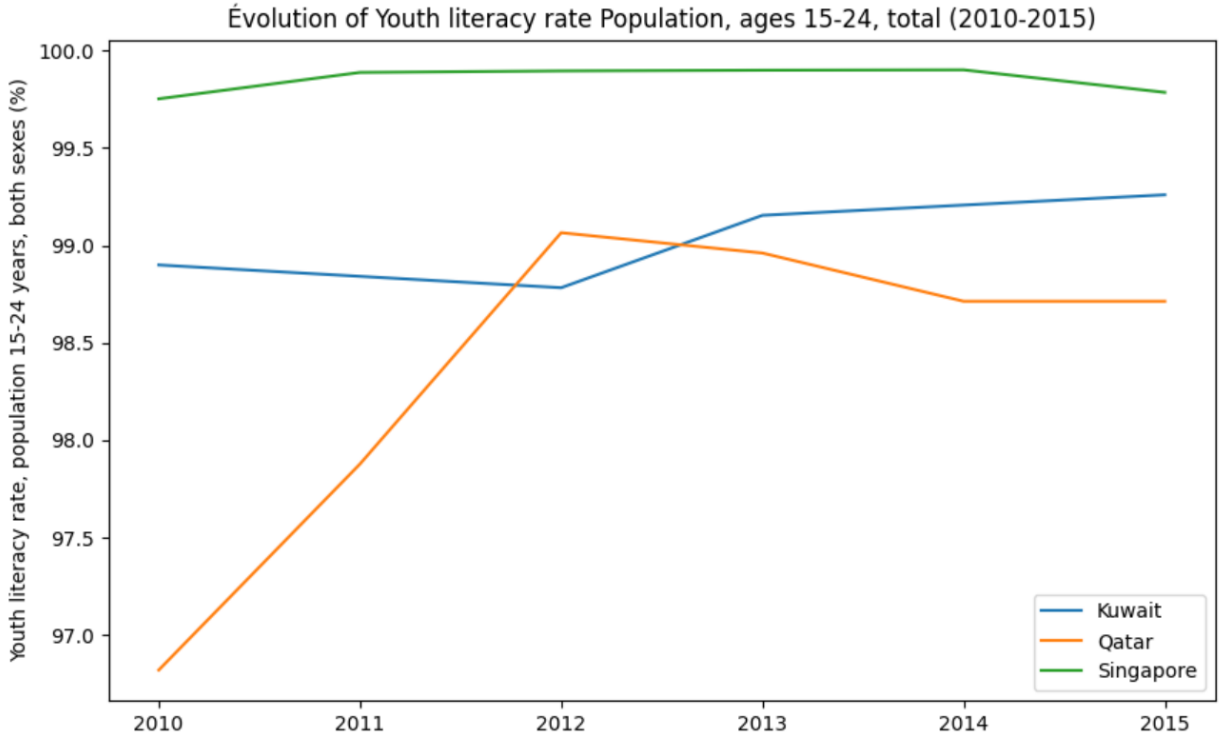




Openclassrooms

6- Projection





2ième Cas

Top 5 des pays ayant des données sur tous les 4 indicateurs

1* Qatar
2* Singapore
3* Kuwait
4* Spain
5* Oman