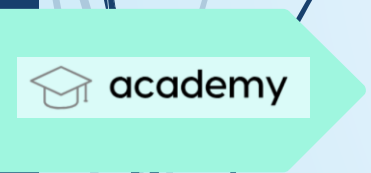


Pour info, sur les slides 11 et 27, il y a des animations qui cachent le texte en dessous donc si possible plutôt regarder le ppt en mode diaporama pour voir le texte sur ces 2 slides.

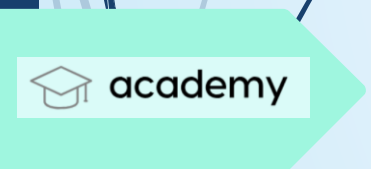


Expansion à l'international

Analyse exploratoire des données de la Banque mondiale sur l'éducation

Sommaire

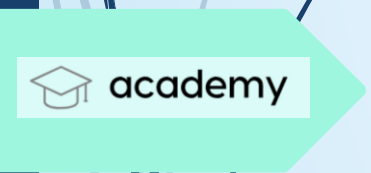
1. Rappel de la problématique
2. Présentation du jeu de données
3. Analyse du jeu de données
 - a. Vision macro
 - b. Indicateurs retenus
 - c. Périmètre d'étude et hypothèses sur les valeurs manquantes
 - d. Analyse des indicateurs
4. Mise en place d'un système de scores
 - a. Méthodologie
 - b. Résultats
 - c. Perspectives d'évolution
5. Conclusion



1. Rappel de la problématique

Rappel de la problématique

- Projet de l'entreprise de s'étendre à l'international
- Différents points à déterminer :
 - Les pays avec un fort potentiel de clients
 - Evolution du potentiel de clients pour chacun de ces pays
 - Classer ces pays par ordre de priorité
- Première analyse à partir des données sur l'éducation fournies par la Banque mondiale



2. Présentation du jeu de données

Présentation du jeu de données (1/3)



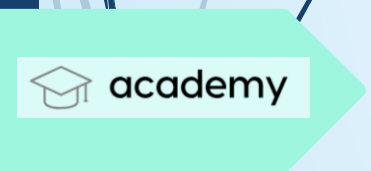
- **Institution financière internationale** qui aide et finance des projets dans des pays en développement pour notamment ***lutter contre la pauvreté*** et ***promouvoir l'éducation***.
- World Bank EdStats (Education Statistics), données sur :
 - L'accès à l'éducation
 - L'obtention de diplômes dans la population
 - Les apprentissages
 - Les dépenses dans l'éducation
 - Les politiques menées et l'égalité dans l'éducation
- Sources des données :
 - UNESCO Institute for Statistics (UIS)
 - Résultats d'études internationales sur l'éducation (PISA...)
 - Résultats d'études réalisées par la Banque mondiale

Présentation du jeu de données (2/3)

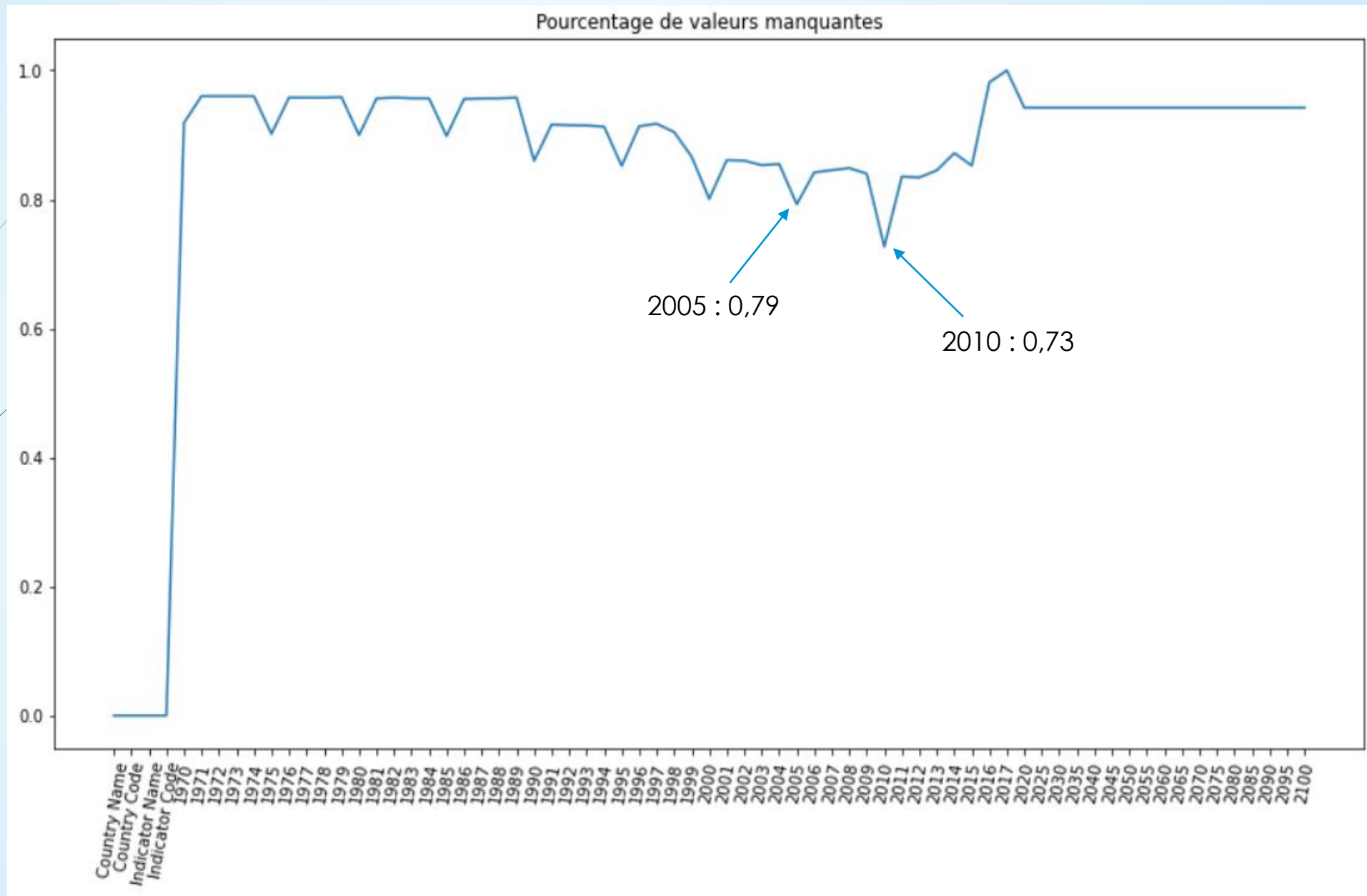
- 5 fichiers disponibles
- **'EdStatsCountry'** : 241 lignes et 32 colonnes
 - Donne des indications sur les pays (région du pays, niveau de revenu, monnaie, source des dernières données etc...)
- **'EdStatsSeries'** : 3 665 lignes et 21 colonnes
 - Donne des informations sur les différents indicateurs du jeu de données (nom, sujet concerné, source, unité de mesure etc...)
- **'EdStatsCountry-Series'** : 613 lignes et 4 colonnes
 - Donne la source des données pour certains indicateurs de certains pays
- **'EdStatsFootNote'** : 643 638 lignes et 5 colonnes
 - Donne des informations supplémentaires sur la source ou la valeur des données de certains indicateurs, pour certains pays et certaines années
- **'EdStatsData'** : **886 930 lignes et 70 colonnes**
 - **Donne les valeurs des différents indicateurs par pays et par année**

Présentation du jeu de données (3/3)

- 3 665 indicateurs
- 217 pays/territoires (ex : St Martin, Polynésie Française)
- 25 groupements de pays (ex : 'Arab World', 'East Asia & Pacific (developing only)' etc...)
- Temporalité : 1970 – 2100
 - 1970-2017 : données annuelles
 - 2020-2100 : données tous les 5 ans



3. Analyse du jeu de données



Indicateurs retenus pour l'analyse (1/2)

➤ Besoins métier :

- Marché potentiel (=personne avec un niveau lycée minimum)
- Niveau d'accès à internet (formation en ligne)
- Niveau économique du marché
- Taux de chômage
- Croissance potentielle du marché

➤ Marché potentiel :

- "Barro-Lee: Percentage of population age X with secondary schooling. Completed Secondary"
- "Barro-Lee: Percentage of population age X with tertiary schooling. Total (Incomplete and Completed Tertiary)"
- "Barro-Lee: Population in thousands, age X, total"

avec X=['20-24', '25-29', '30-34', '35-39', '40-44', '45-49', '50-54', '55-59']

Indicateurs retenus pour l'analyse (2/2)

- Niveau d'accès à internet :
 - "Internet users (per 100 people)"
- Niveau économique du marché :
 - "GDP per capita, PPP (current international \$)"
 - Colonne "Income Group" du fichier 'EdStatsCountry'
- Taux de chômage :
 - "Unemployment, total (% of total labor force)"
 - "Labor force, total"
- Croissance potentielle du marché :
 - "Wittgenstein Projection: Population in thousands by highest level of educational attainment. Upper Secondary. Total"
 - "Wittgenstein Projection: Population in thousands by highest level of educational attainment. Post Secondary. Total"

Périmètre d'étude et hypothèses sur les valeurs manquantes (1/3)

- Marché potentiel :
 - Année de référence : 2010
 - 141 pays dans l'analyse
- Pays non inclus dans l'analyse :

Country Name	Indicator Name	2016	Income Group
Nigeria	Population, total	185,989,640	Lower middle income
Ethiopia	Population, total	102,403,196	Low income
Uzbekistan	Population, total	31,848,200	Lower middle income
Angola	Population, total	28,813,463	Upper middle income
Korea, Dem. People's Rep.	Population, total	25,368,620	Low income
Madagascar	Population, total	24,894,551	Low income
Burkina Faso	Population, total	18,646,433	Low income
Chad	Population, total	14,452,543	Low income
Somalia	Population, total	14,317,996	Low income
Guinea	Population, total	12,395,924	Low income
Azerbaijan	Population, total	9,762,274	Upper middle income
Belarus	Population, total	9,507,120	Upper middle income
Lebanon	Population, total	6,006,668	Upper middle income
Turkmenistan	Population, total	5,662,544	Upper middle income
West Bank and Gaza	Population, total	4,551,566	Lower middle income

Périmètre d'étude et hypothèses sur les valeurs manquantes (2/3)

- Niveau d'accès à internet :
 - Années de référence : 2014, 2015, 2016
 - Moyenne de ces 3 années

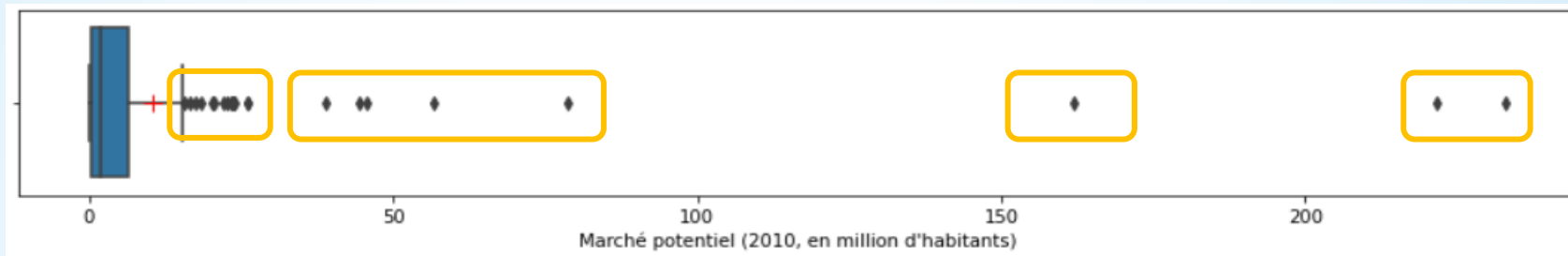
- Niveau économique du marché :
 - Années de référence : 2014, 2015, 2016
 - Moyenne de ces 3 années
 - Valeurs manquantes :
 - Bahreïn 2016 : hypothèse → moyenne de 2013 à 2015
 - Venezuela 2015 : hypothèse → moyenne de 2012 à 2014
 - Venezuela 2016 : hypothèse → moyenne de 2013 à 2015

Périmètre d'étude et hypothèses sur les valeurs manquantes (3/3)

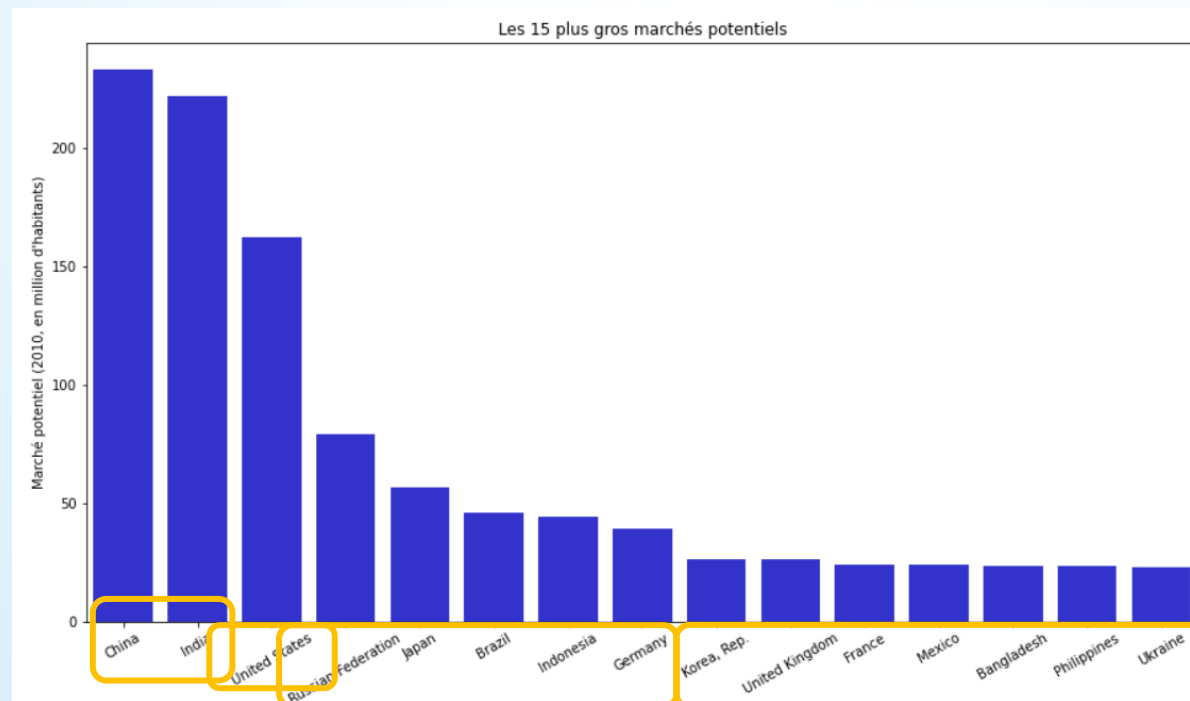
- Taux de chômage :
 - Années de référence : 2014, 2015, 2016
 - Moyenne de ces 3 années
- Croissance potentielle du marché :
 - Données sur le 'upper secondary'
 - Données sur le 'post secondary' (qui regroupe tout le 'tertiary')

Analyse des indicateurs (1/6)

Marché potentiel : Médiane = 2M ; Moyenne = 10,6M ; Ecart-type = 31,3M



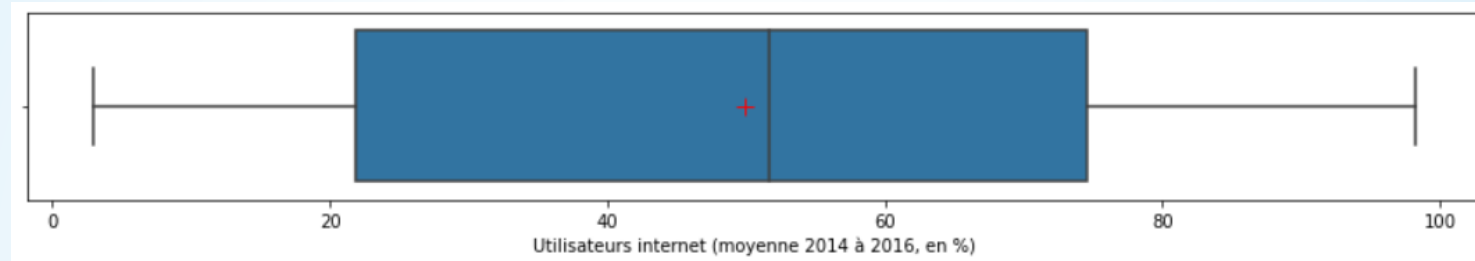
Min = 0,018 M
Q1 = 0,47 M
Q3 = 6,6 M
Max = 233 M



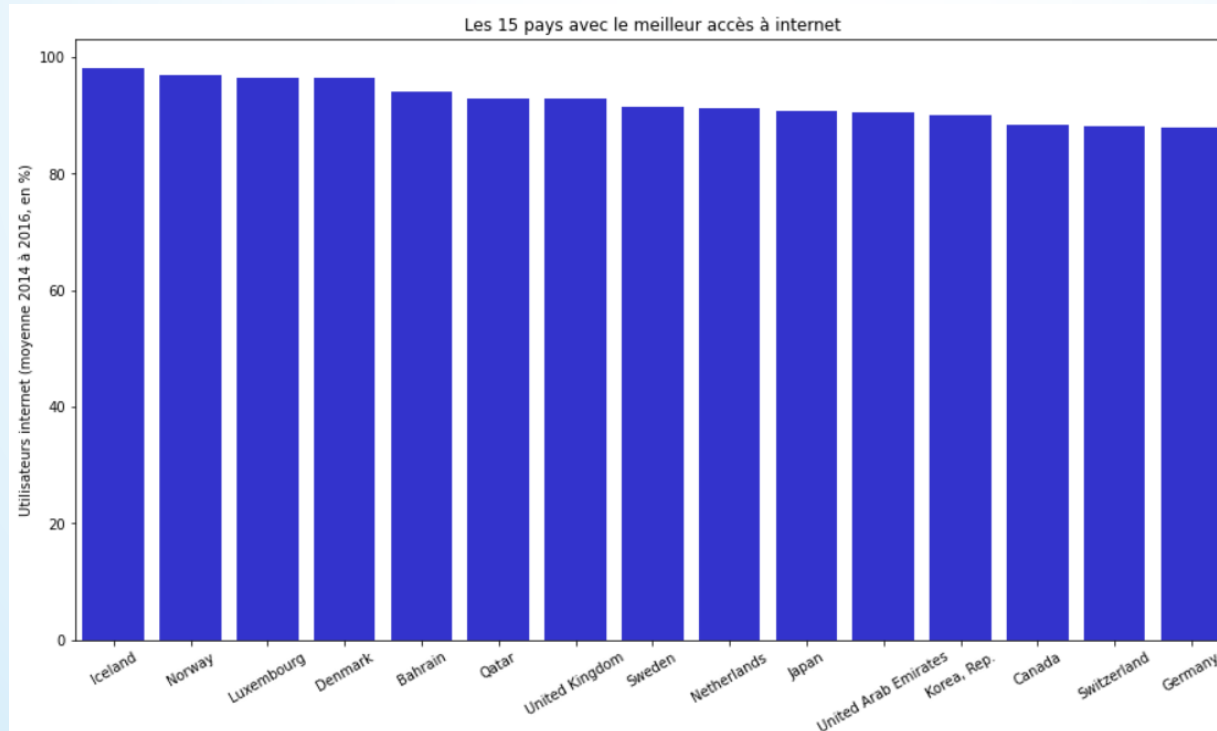
Country Name	Marché potentiel (2010, en million d'habitants)
China	232.8
India	221.6
United States	162.0
Russian Federation	78.8
Japan	56.9
Brazil	45.8
Indonesia	44.4
Germany	39.0
Korea, Rep.	26.2
United Kingdom	26.0
France	24.0
Mexico	23.8
Bangladesh	23.6
Philippines	23.5
Ukraine	22.7

Analyse des indicateurs (2/6)

➡ Niveau d'accès à internet : Médiane = 52 % ; Moyenne = 50 % ; Ecart-type = 28 %



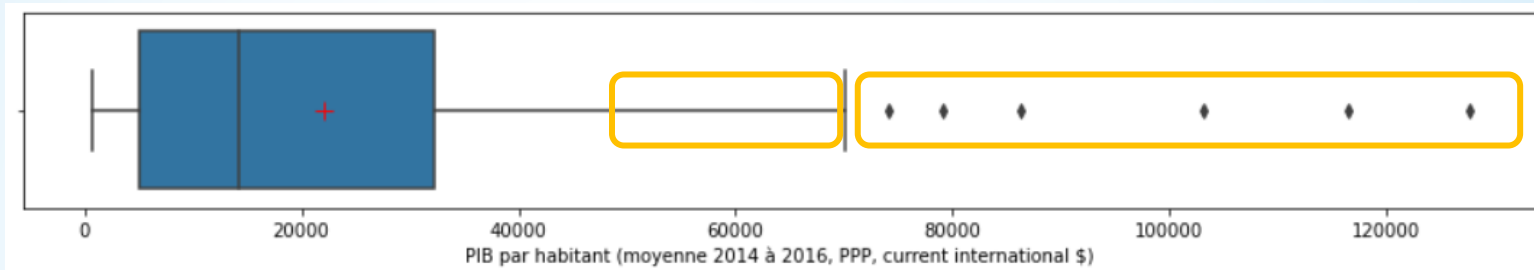
Min = 3 %
Q1 = 22 %
Q3 = 75 %
Max = 98 %



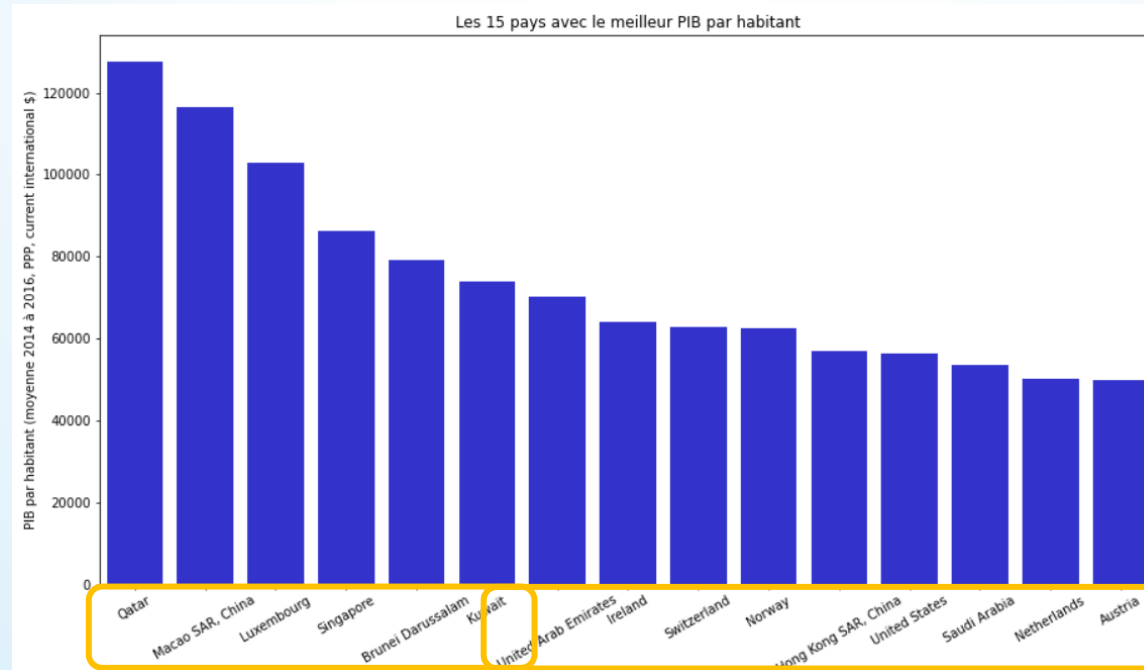
Country Name	Utilisateurs internet (moyenne 2014 à 2016, en %)
Iceland	98.2
Norway	96.8
Luxembourg	96.5
Denmark	96.4
Bahrain	94.0
Qatar	92.9
United Kingdom	92.8
Sweden	91.5
Netherlands	91.3
Japan	90.7
United Arab Emirates	90.5
Korea, Rep.	90.0
Canada	88.5
Switzerland	88.1
Germany	87.8

Analyse des indicateurs (3/6)

- Niveau économique du marché : Médiane = 14 272 \$; Moyenne = 22 134 \$; Ecart-type = 23 583 \$



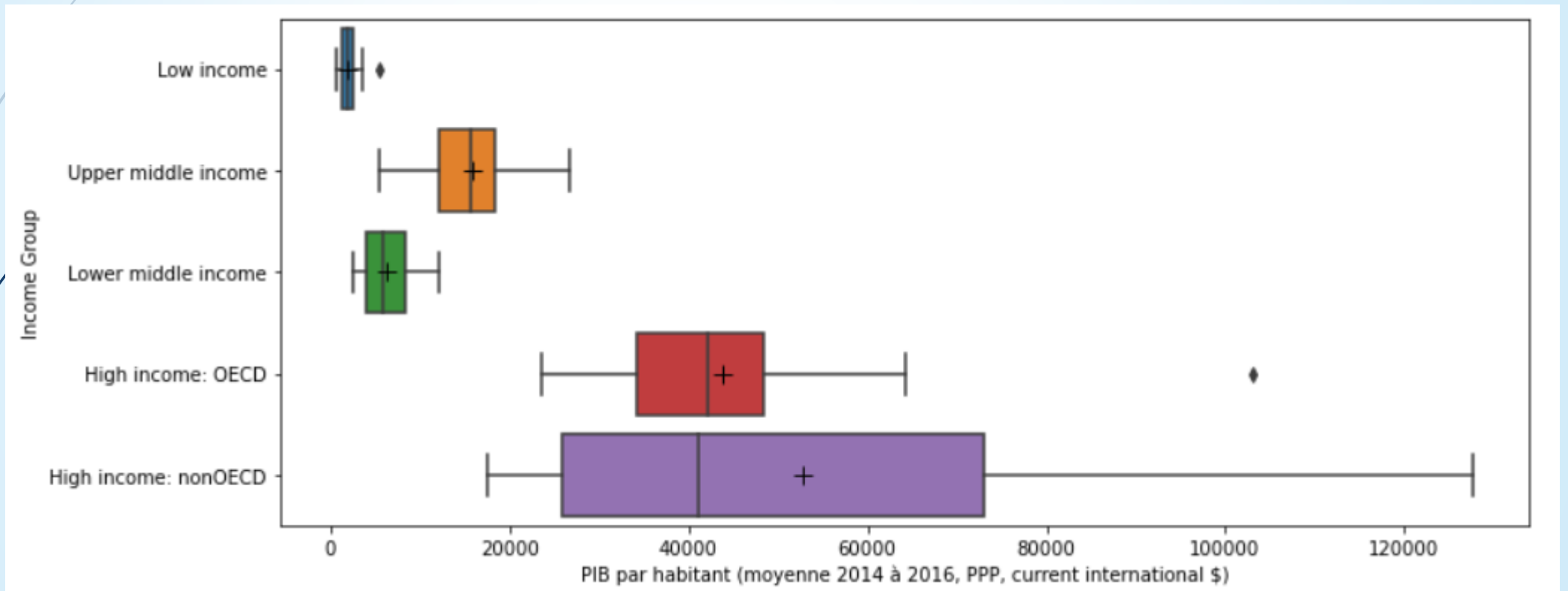
Min = 667 \$
 Q1 = 5 025 \$
 Q3 = 32 279 \$
 Max = 127 602 \$



Country Name	PIB par habitant (moyenne 2014 à 2016, PPP, current international \$)
Qatar	127,602.0
Macao SAR, China	116,442.1
Luxembourg	103,055.3
Singapore	86,300.9
Brunei Darussalam	79,105.4
Kuwait	74,027.2
United Arab Emirates	70,160.8
Ireland	64,221.2
Switzerland	62,941.8
Norway	62,412.9
Hong Kong SAR, China	57,078.4
United States	56,235.2
Saudi Arabia	53,755.7
Netherlands	50,128.6
Austria	49,912.4

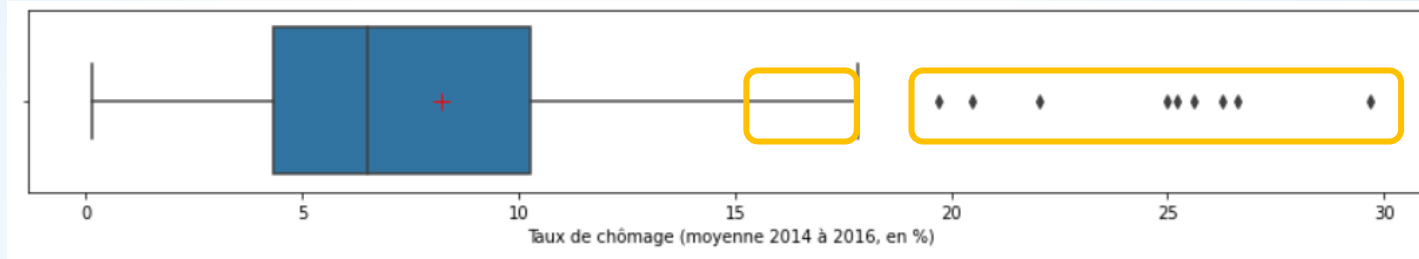
Analyse des indicateurs (4/6)

- Corrélation entre le PIB par habitant et “l’income group” :

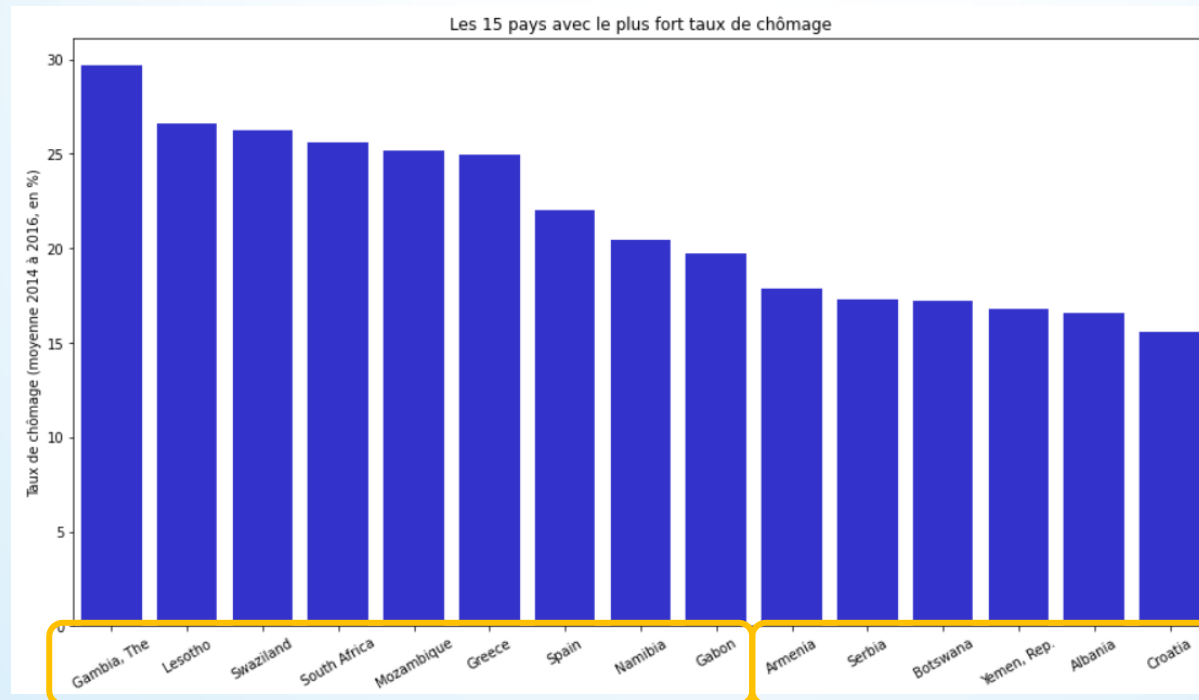


Analyse des indicateurs (5/6)

➤ Taux de chômage : Médiane = 6,5 % ; Moyenne = 8,2 % ; Ecart-type = 5,9 %



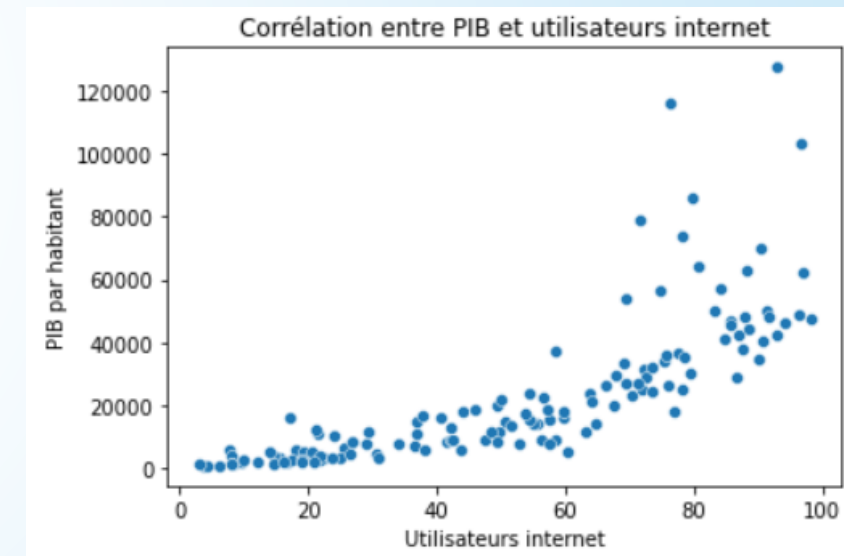
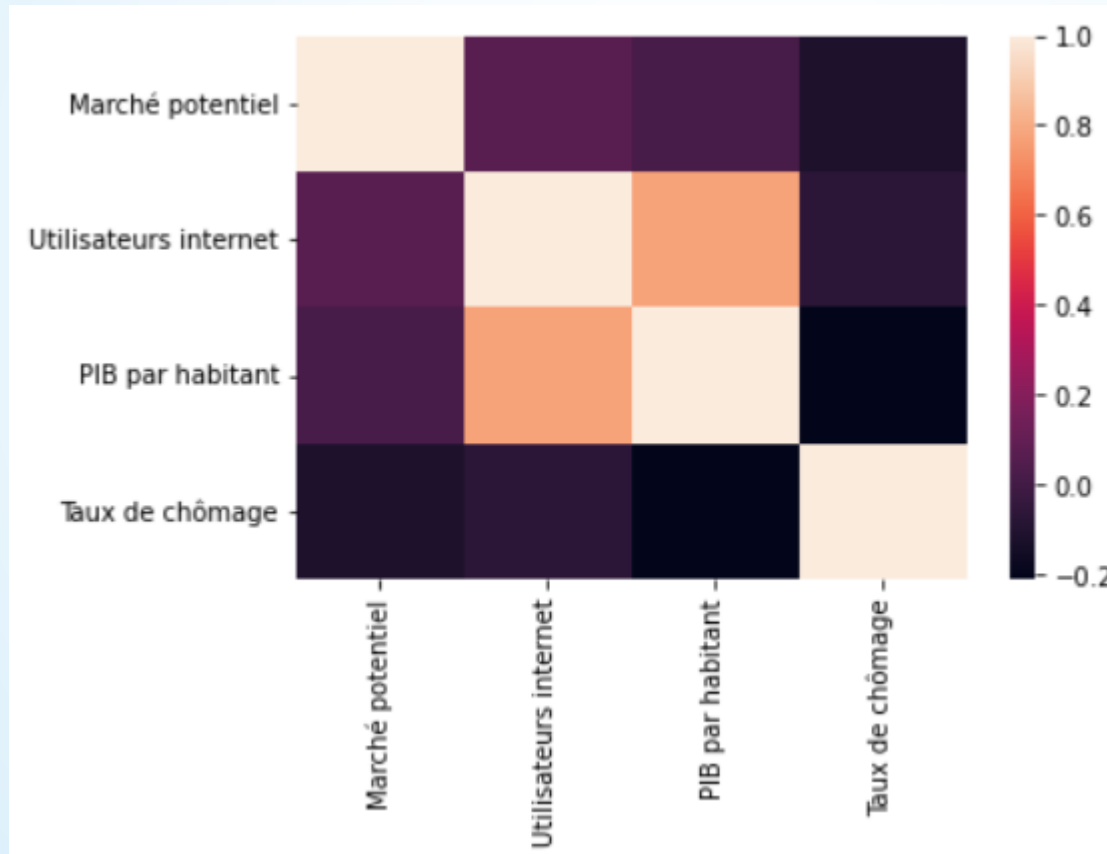
Min = 0,1 %
Q1 = 4,3 %
Q3 = 10,3 %
Max = 29,7%



Country Name	Taux de chômage (moyenne 2014 à 2016, en %)
Gambia, The	29.7
Lesotho	26.6
Swaziland	26.3
South Africa	25.6
Mozambique	25.2
Greece	25.0
Spain	22.0
Namibia	20.5
Gabon	19.7
Armenia	17.8
Serbia	17.3
Botswana	17.2
Yemen, Rep.	16.8
Albania	16.6
Croatia	15.6

Analyse des indicateurs (6/6)

- Analyse des corrélations linéaires entre les variables :



4. Mise en place d'un système de scores

Méthodologie

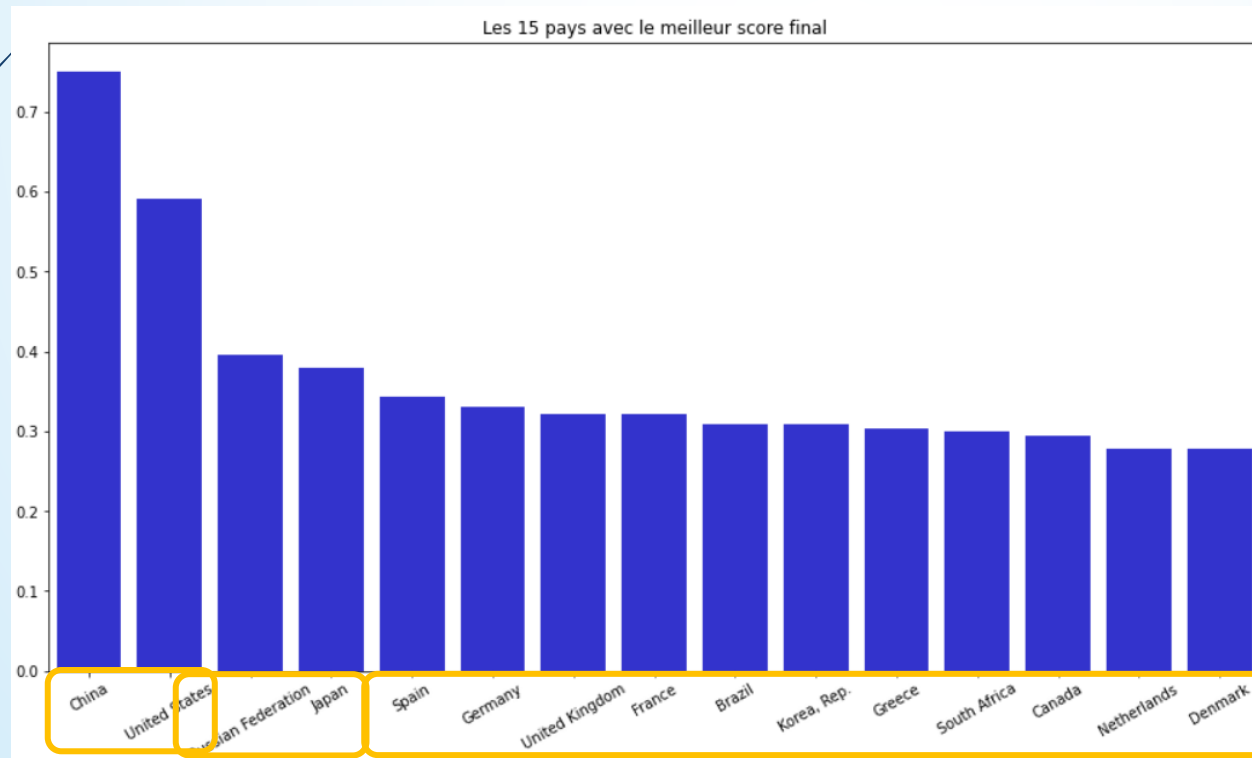
- Indicateurs ramenés entre 0 et 1
- Poids plus important donné à certains indicateurs :
 - Marché potentiel : 60
 - Niveau d'accès à internet : 25
 - Taux de chômage : 15
- Pour le niveau économique : filtre sur les pays avec un 'income group' supérieur ou égal à 'Upper middle income'
- Pour le marché potentiel : poids plus important donné au 20-44 ans qu'au 45-59 (70/30)

Résultats

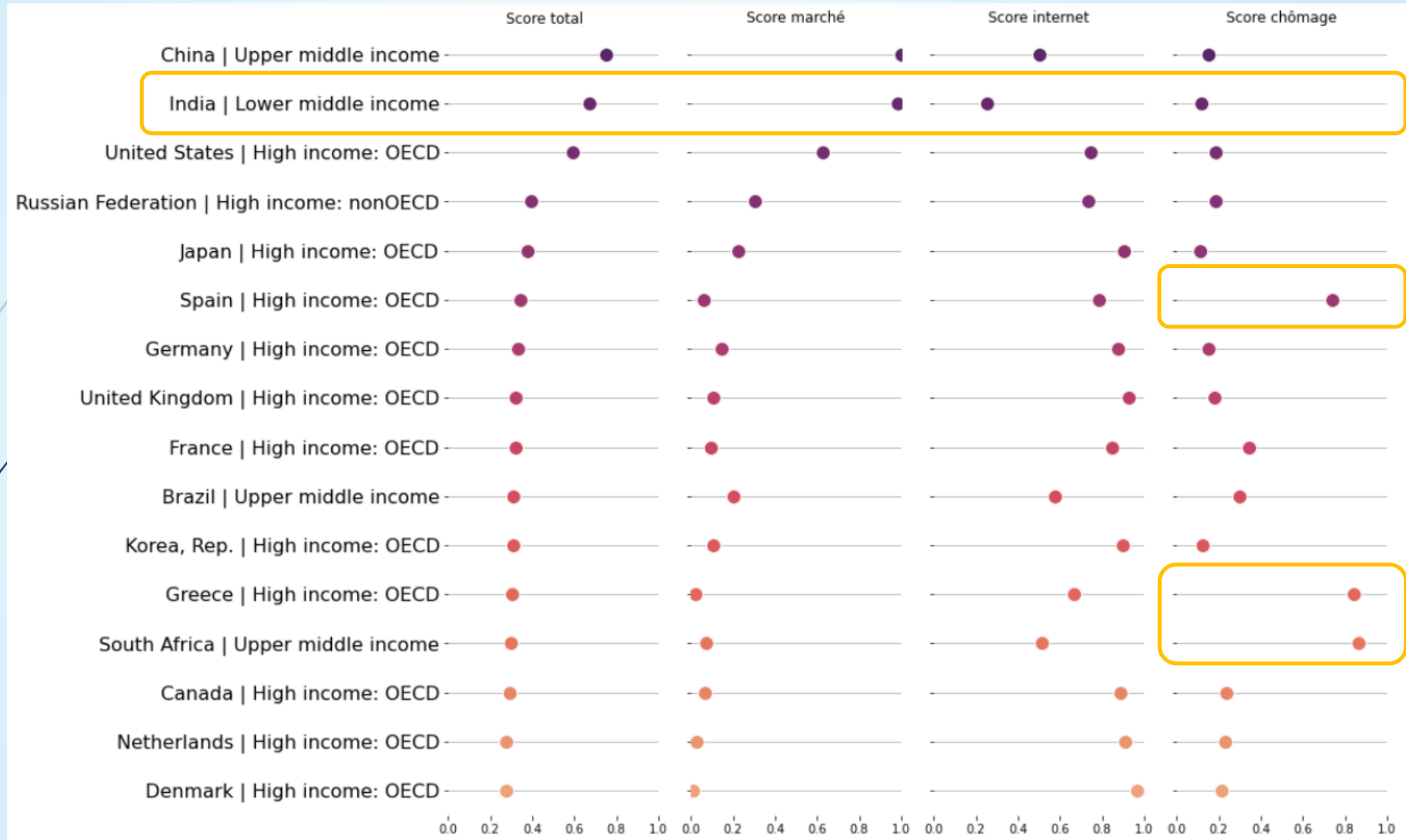
84 pays dans l'analyse : Médiane = 0,23 ; Moyenne = 0,24 ; Ecart-type = 0,09



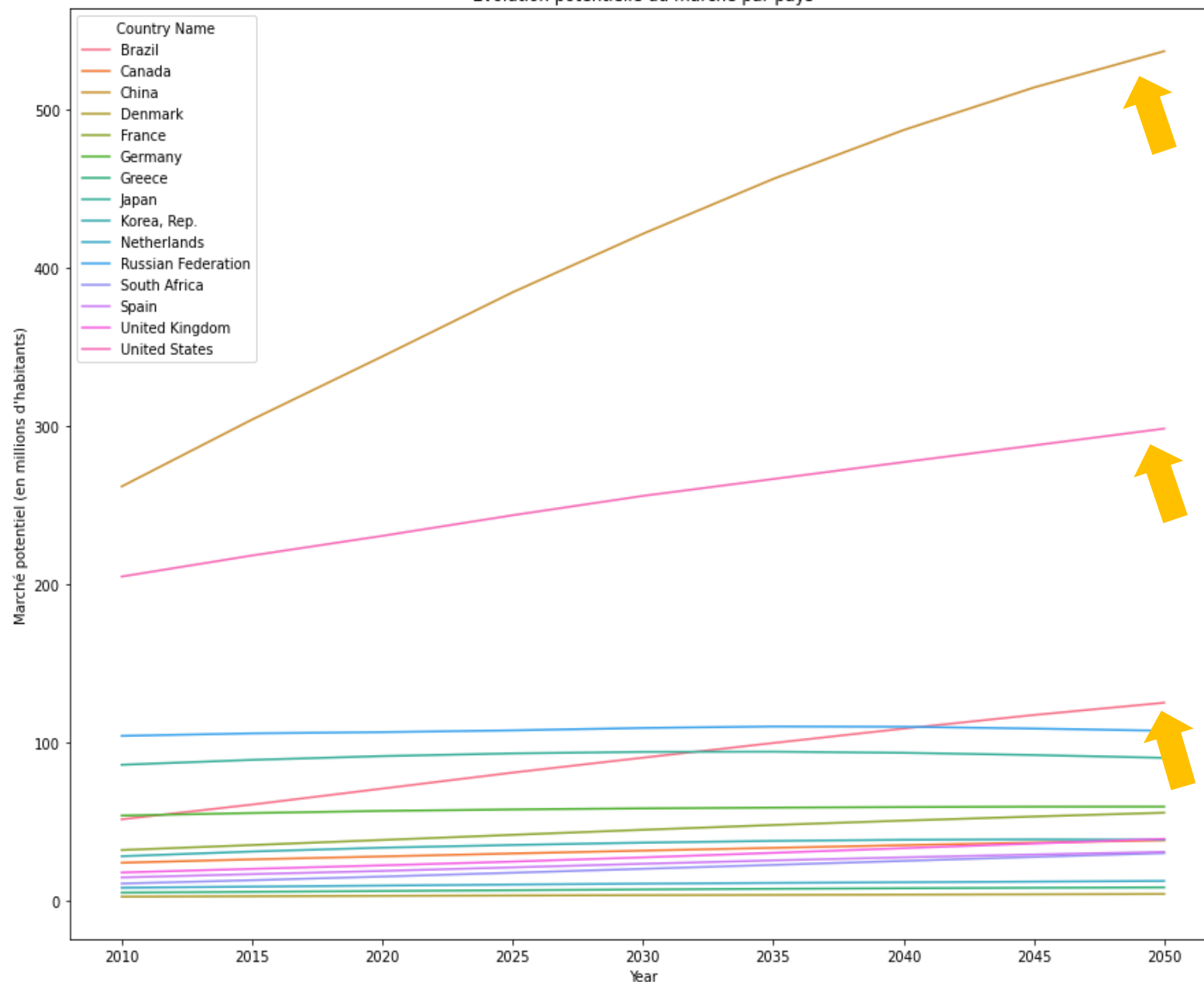
Min = 0,12
Q1 = 0,2
Q3 = 0,27
Max = 0,75



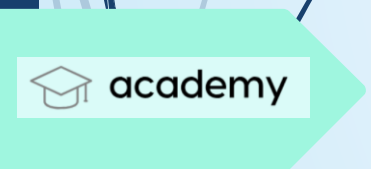
Country Name	Score total	Score total
China	0.75	0.75
United States	0.59	0.59
Russian Federation	0.40	0.40
Japan	0.38	0.38
Spain	0.34	0.34
Germany	0.33	0.33
United Kingdom	0.32	0.32
France	0.32	0.32
Brazil	0.31	0.31
Korea, Rep.	0.31	0.31
Greece	0.30	0.30
South Africa	0.30	0.30
Canada	0.29	0.29
Netherlands	0.28	0.28
Denmark	0.28	0.28



Evolution potentielle du marché par pays



Country Name	2010	2050	Evolution 2050 vs 2010 (en million)	Evolution 2050 vs 2010 (en %)
China	262.2	537.3	275.1	104.9
United States	205.3	298.7	93.5	45.5
Russian Federation	104.6	107.8	3.2	3.1
Japan	86.3	90.7	4.4	5.1
Germany	54.2	59.8	5.6	10.4
Brazil	51.9	125.5	73.7	142.1
France	32.4	56.0	23.6	72.8
Korea, Rep.	28.5	39.0	10.4	36.6
Canada	24.4	38.5	14.1	57.9
United Kingdom	18.2	39.5	21.3	116.5
Spain	15.1	31.2	16.1	106.9
South Africa	11.3	30.5	19.2	170.5
Netherlands	8.6	12.9	4.3	49.8
Greece	5.5	8.8	3.3	59.7
Denmark	3.0	4.7	1.7	56.1



5. Conclusion

Conclusion

- ▶ Jeu de données qui permet de donner un premier aperçu des pays potentiellement intéressants
- ▶ Données sur le marché potentiel pas très récente (2010)
- ▶ Données sur le taux de chômage qui peuvent avoir fortement évoluées depuis 2016. Ex : Espagne 2016 = 20 % ; 2022 = 13 %
- ▶ Compléter si possible le taux de chômage avec un détail sur les personnes ayant un niveau lycée minimum pour affiner le classement
- ▶ Nécessité de compléter l'analyse avec par exemple une étude de la concurrence dans les différents pays ciblés