

ANALYSEZ DES DONNÉES DE SYSTÈMES ÉDUCATIFS

MARWA EL HOURI

PROBLÉMATIQUE

- La start up **academy**, propose des contenus de formation en ligne pour un public de niveau lycée et université.
- L'entreprise **academy** a un projet d'expansion à l'international.
- Objectifs:
 - Pays avec un fort potentiel de clients
 - Evolution de ce potentiel de clients
 - Dans quels pays opérer en priorité ?



PLAN DE TRAVAIL

Etape 1 :
Présentation des
données

Etape 2 :
Premier nettoyage
et regroupement
des données

Etape 3 :
Exploration et
choix des
indicateurs

Etape 4 :
Nettoyage et
études sur les
indicateurs

Etape 5 :
Score des pays a
fort potentiel client

Etape 6 :
Discussions des
résultats

Etape 1 :
Présentation des
données

Etape 2 :
Premier nettoyage
et regroupement
des données

Etape 3 :
Exploration et
choix des
indicateurs

Etape 4 :
Nettoyage et
études sur les
indicateurs

Etape 5 :
Score des pays a
fort potentiel client

Etape 6 :
Discussions des
résultats

JEUX DE DONNÉES



Origine : La Banque mondiale



5 jeux de données

EdStatsCountry.csv

EdStatsCountry-Series.csv

EdStatsData.csv

EdStatsFootNote.csv

EdStatsFootNote.csv

TABLE I EDSTATSCOUNTRY.CSV

- Informations sur la répartition de **241 pays** par **Région** et par **Niveau de ressource**
- Dimensions: (241,32)

1	data_pays.head()														
	Country Code	Short Name	Table Name	Long Name	2-alpha code	Currency Unit	Special Notes	Region	Income Group	WB-2 code	...	IMF data dissemination standard	Latest population census	Latest household survey	So most Incon exper
0	ABW	Aruba	Aruba	Aruba	AW	Aruban florin	SNA data for 2000-2011 are updated from offici...	Latin America & Caribbean	High income: nonOECD	AW	...	NaN	2010	NaN	
1	AFG	Afghanistan	Afghanistan	Islamic State of Afghanistan	AF	Afghan afghani	Fiscal year end: March 20; reporting period fo...	South Asia	Low income	AF	...	General Data Dissemination System (GDDS)	1979	Multiple Indicator Cluster Survey (MICS), 2010/11	Inte hou survey
2	AGO	Angola	Angola	People's Republic of Angola	AO	Angolan kwanza	April 2013 database update: Based on IMF	Sub-Saharan Africa	Upper middle income	AO	...	General Data Dissemination System (GDDS)	1970	Malaria Indicator Survey (MIS), 2011	Inte hou survey

TABLE I EDSTATSCOUNTRY.CSV

- Dimensions: (241,32)
- Valeurs NaN dans colonne **Region** sont des groupements d'entités géographiques ou économiques non-pertinents pour l'étude

Country Code	0
Short Name	0
Table Name	0
Long Name	0
2-alpha code	3
Currency Unit	26
Special Notes	96
Region	27
Income Group	27
WB-2 code	1
National accounts base year	36
National accounts reference year	209
SNA price valuation	44
Lending category	97
Other groups	183
System of National Accounts	26
Alternative conversion factor	194
PPP survey year	96
Balance of Payments Manual in use	60
External debt Reporting status	117
System of trade	41
Government Accounting concept	80
IMF data dissemination standard	60
Latest population census	28
Latest household survey	100
Source of most recent Income and expenditure data	81
Vital registration complete	130
Latest agricultural census	99
Latest industrial data	134
Latest trade data	56
Latest water withdrawal data	62
Unnamed: 31	241

Nombre de valeurs NaN dans chaque colonne

TABLE 2 EDSTATSCOUNTRY-SERIES.CSV

- Informations sur différents indicateurs concernant **211 pays**. On peut identifier **2 indicateurs** pertinents concernant la **population** présents pour tous les pays.
- Dimensions : (6 | 3,4)
- Pas de valeurs NaN a part la dernière colonne

```
1 data_pays_series.head()
```

	CountryCode	SeriesCode	DESCRIPTION	Unnamed: 3
0	ABW	SP.POP.TOTL	Data sources : United Nations World Population...	NaN
1	ABW	SP.POP.GROW	Data sources: United Nations World Population ...	NaN
2	AFG	SP.POP.GROW	Data sources: United Nations World Population ...	NaN
3	AFG	NY.GDP.PCAP.PP.CD	Estimates are based on regression.	NaN
4	AFG	SP.POP.TOTL	Data sources : United Nations World Population...	NaN

TABLE 3 EDSTATSDATA.CSV

- Données de **3665** indicateurs sur **242** pays pour les années entre **1970** a **2100**.

	Country Name	Country Code	Indicator Name	Indicator Code	1970	1971	1972	1973	1974	1975	...	2060	2065	2070	2075	2080	2085
0	Arab World	ARB	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
1	Arab World	ARB	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2.F	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
2	Arab World	ARB	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2.GPI	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
3	Arab World	ARB	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2.M	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN

TABLE 3 EDSTATSDATA.CSV

- Dimensions : (886930 , 70)
- Beaucoup de valeurs manquantes sur les années
- C'est le jeu de données principal a nettoyer

Country Name	0
Country Code	0
Indicator Name	0
Indicator Code	0
1970	814642
1971	851393
1972	851311
1973	851385
1974	851200
1975	799624
1976	849447
1977	849356
1978	849354
1979	850121
1980	797808
1981	848153
1982	849419
1983	848470
1984	848324
1985	796634

2016	870470
2017	886787
2020	835494
2025	835494
2030	835494
2035	835494
2040	835494
2045	835494
2050	835494
2055	835494
2060	835494
2065	835494
2070	835494
2075	835494
2080	835494
2085	835494
2090	835494
2095	835494
2100	835494
Unnamed: 69	886930

TABLE 4 EDSTATSFOOTNOTE.CSV

- Répartition par années d'étude de **1558** **indicateurs** concernant **239 pays**
- **Dimensions** (643638 , 5)
- Pas de valeurs manquantes a part la dernière colonne

	CountryCode	SeriesCode	Year	DESCRIPTION	Unnamed: 4
0	ABW	SE.PRE.ENRL.FE	YR2001	Country estimation.	NaN
1	ABW	SE.TER.TCHR.FE	YR2005	Country estimation.	NaN
2	ABW	SE.PRE.TCHR.FE	YR2000	Country estimation.	NaN
3	ABW	SE.SEC.ENRL.GC	YR2004	Country estimation.	NaN
4	ABW	SE.PRE.TCHR	YR2006	Country estimation.	NaN

TABLE 5 EDSTATSSERIES.CSV

- Informations sur la définition de **3665 indicateurs** repartis sur **37 thèmes**.
- **Dimensions : (3665, 21)**
- Pas de valeurs manquantes sur les colonnes pertinents

	Series Code	Topic	Indicator Name	Short definition	Long definition	Unit of measure	Periodicity	Base Period	Other notes	Aggregation method	...	Notes from original source	General comments
0	BAR.NOED.1519.FE.ZS	Attainment	Barro-Lee: Percentage of female population age...	Percentage of female population age 15-19 with...	Percentage of female population age 15-19 with...	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN
1	BAR.NOED.1519.ZS	Attainment	Barro-Lee: Percentage of population age 15-19 ...	Percentage of population age 15-19 with no edu...	Percentage of population age 15-19 with no edu...	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN
2	BAR.NOED.15UP.FE.ZS	Attainment	Barro-Lee: Percentage of female population age...	Percentage of female population age 15+ with n...	Percentage of female population age 15+ with n...	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN
3	BAR.NOED.15UP.ZS	Attainment	Barro-Lee: Percentage of population age 15+ wi...	Percentage of population age 15+ with no educa...	Percentage of population age 15+ with no educa...	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN

Etape 1 :
Présentation des
données

Etape 2 :
Premier nettoyage
et regroupement
des données

Etape 3 :
Exploration et
choix des
indicateurs

Etape 4 :
Nettoyage et
études sur les
indicateurs

Etape 5 :
Score des pays a
fort potentiel client

Etape 6 :
Discussions des
résultats

I- NETTOYAGE PAR THÈMES

Objectif: Garder les thèmes les plus pertinents pour notre étude.

Après nettoyage :

- De 37 thèmes a 20 thèmes
- De (886930 , 70) a (375826,70)
(dans jeu data)

```
1 data_series=data_series[~data_series['Topic'].isin(not_needed_topics)]  
2 data_series['Topic'].value_counts()
```

Attainment	733
Secondary	256
Population	213
Tertiary	158
Expenditures	93
Literacy	32
Education Management Information Systems (SABER)	23
Post-Secondary/Non-Tertiary	19
Tertiary Education (SABER)	17
Health: Population: Structure	13
Workforce Development (SABER)	12
Economic Policy & Debt: Purchasing power parity	6
Social Protection & Labor: Unemployment	3
Economic Policy & Debt: National accounts: US\$ at current prices: Aggregate indicators	3
Infrastructure: Communications	2
Economic Policy & Debt: National accounts: US\$ at constant 2010 prices: Aggregate indicators	2
EMIS	2
Background	2
Economic Policy & Debt: National accounts: Atlas GNI & GNI per capita	1
Health: Population: Dynamics	1
Name: Topic, dtype: int64	

2- NETTOYAGE PAR ANNÉES

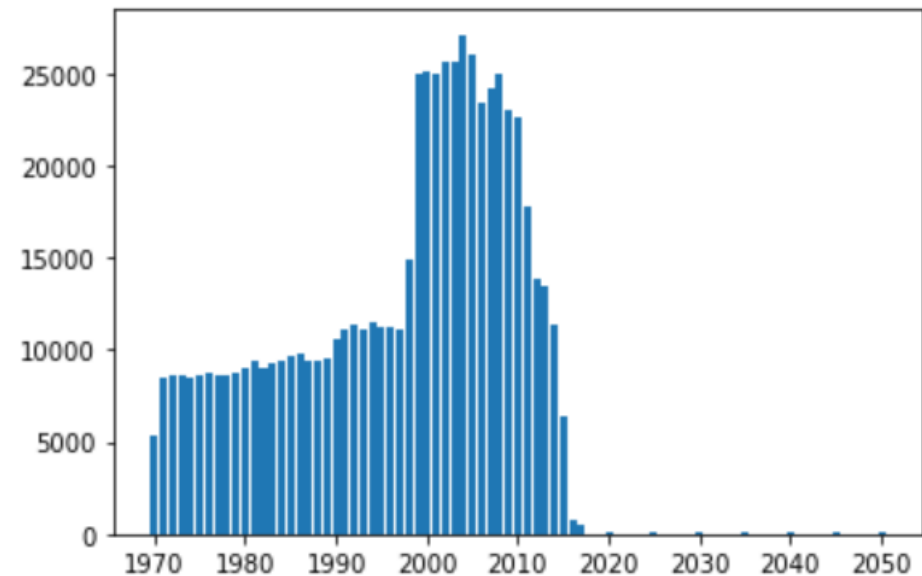
Objectif: Garder les thèmes les années ou il y a le plus d'informations sur les indicateurs.

Conclusion: On garde les années 1972 et 2017 (35 ans)

```
1 indic_par_annee=data_footnote['Year'].value_counts()
```

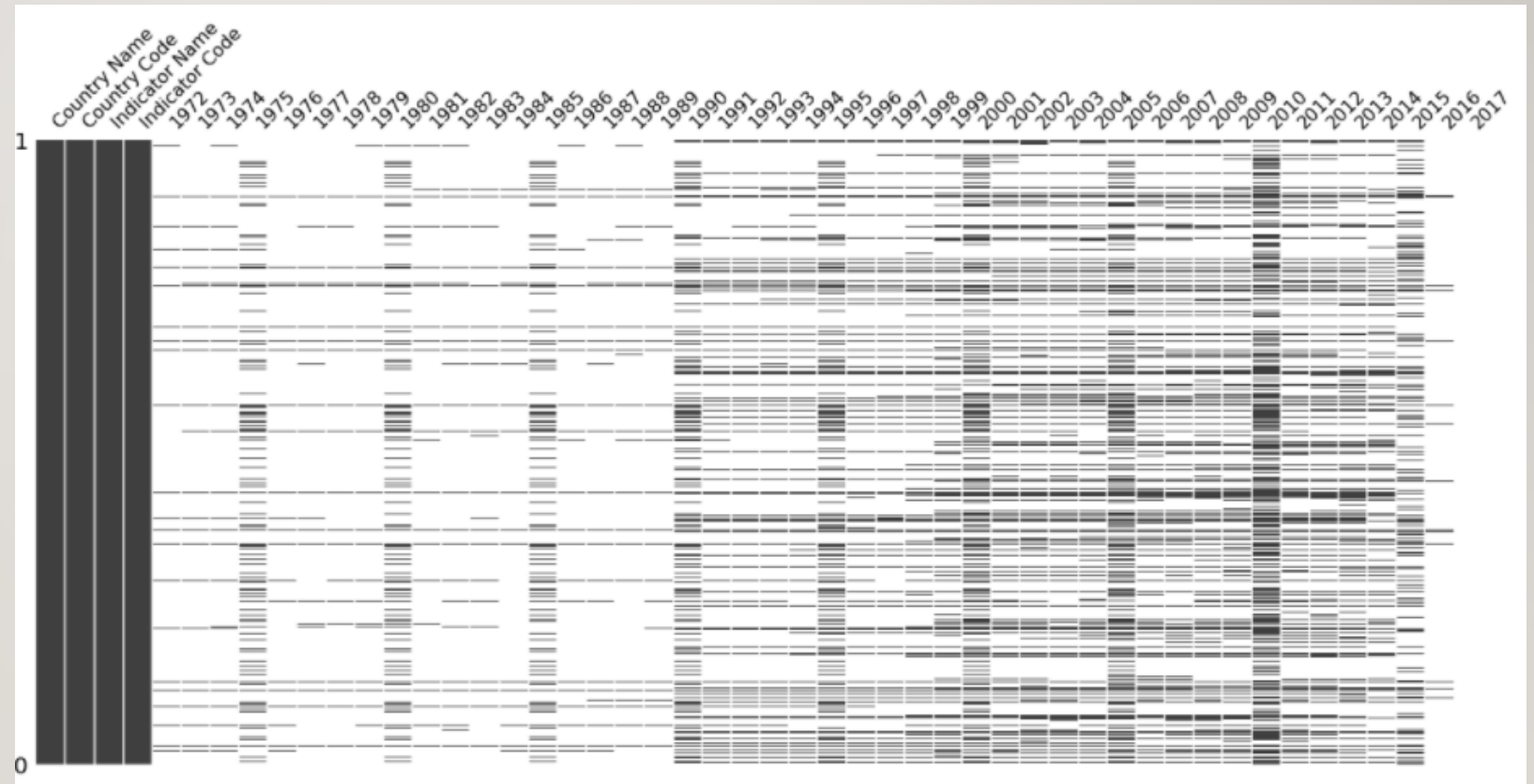
```
1 plt.bar(indic_par_annee.index, indic_par_annee.values)
```

<BarContainer object of 55 artists>



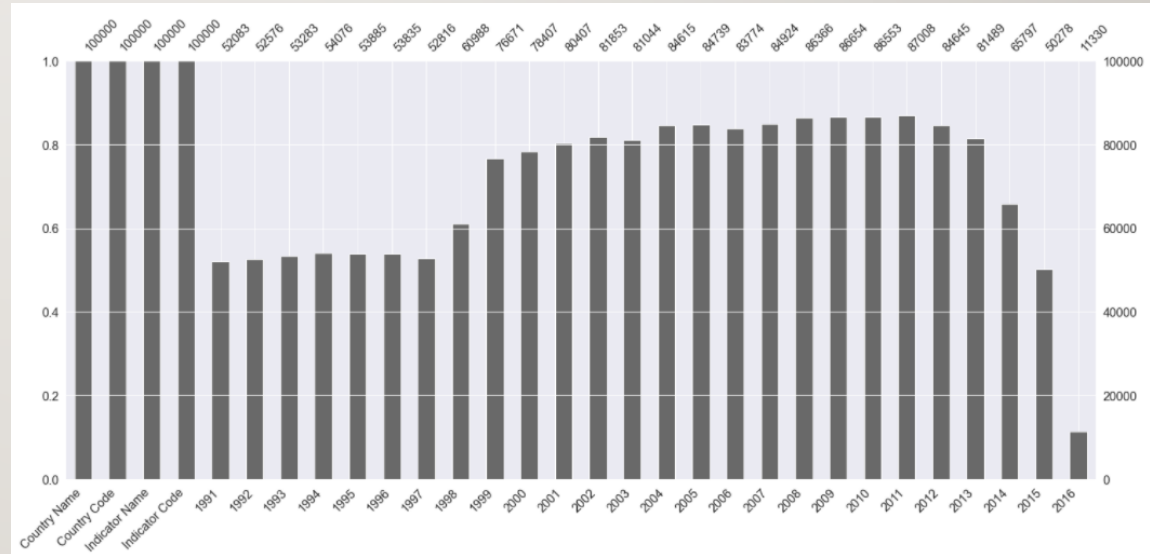
3- NETTOYAGE DES VALEURS NAN

- La matrice de nullité nous montre un grand manque d'information entre 1972 et 1989 par rapport a la période 1990 et 2016



3- NETTOYAGE DES VALEURS NAN

- Garder la période **1991-2016**
- Nettoyer les lignes contenant moins que 5 valeurs non nulles (pour prendre en considérations les indicateurs périodiques sur une période de 5ans)
- Apres nettoyage
- De (375826,70) a (106403, 30)



4- REGROUPEMENT DES DONNÉES

- A la table de données **data** on ajoute

- L'information **Topic** de la table **EdStatsSeries**

```
data_Topic=data_series25ans[['Series Code','Topic']]
```

- L'information **Region** et **Income Group** de la table **EdStatsCountry**

```
pays_Region=data_pays[['Country Code','Region','Income Group']]
```

4- REGROUPEMENT DES DONNÉES

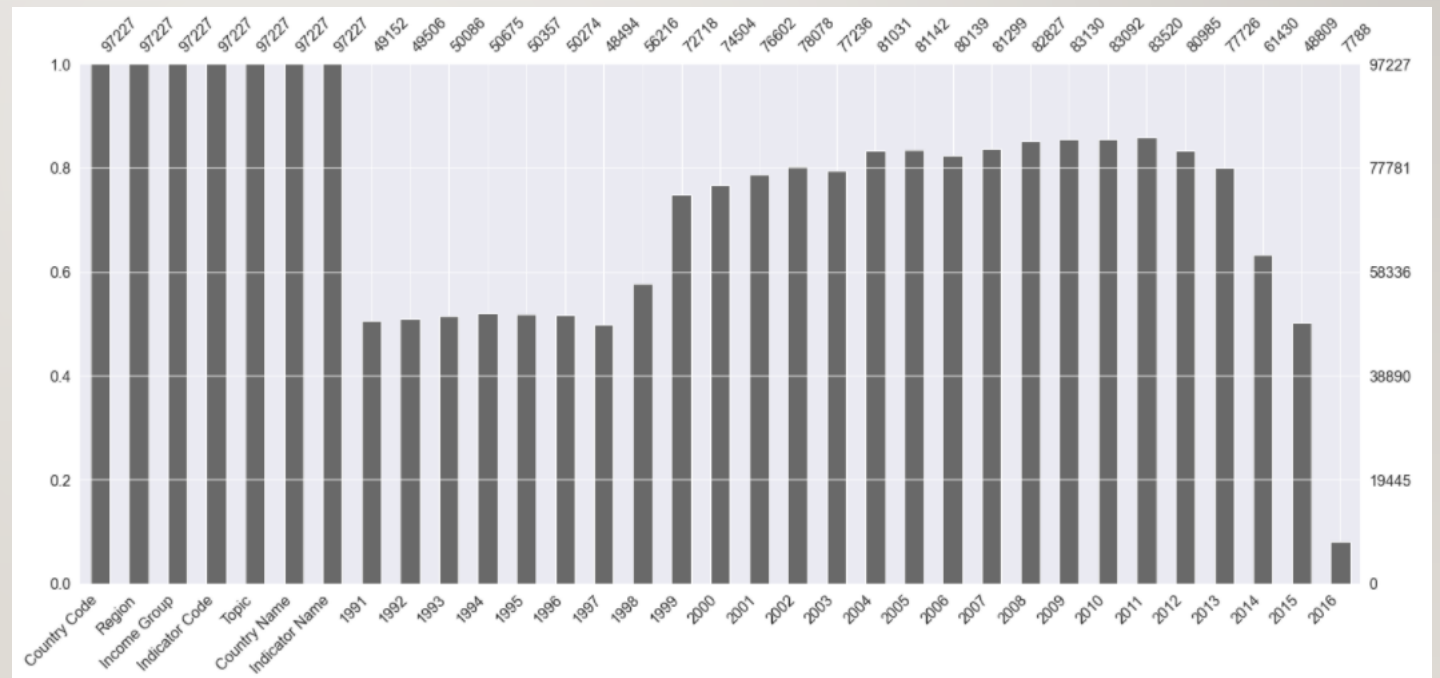
RESULTAT:

- 214 Pays
- 807 Indicateurs
- 15 thèmes

	Country Code	Region	Income Group	Indicator Code	Topic	Country Name	Indicator Name	1990	1991	1992	...	2006	2007	2008	2009	2010	2011	201
0	AFG	South Asia	Low income	BAR.NOED.1519.FE.ZS	Attainment	Afghanistan	Barro-Lee: Percentage of female population age...	84.20	NaN	NaN	...	NaN	NaN	NaN	NaN	59.18	NaN	NaN
1	ALB	Europe & Central Asia	Upper middle income	BAR.NOED.1519.FE.ZS	Attainment	Albania	Barro-Lee: Percentage of female population age...	1.26	NaN	NaN	...	NaN	NaN	NaN	NaN	0.95	NaN	NaN
2	DZA	Middle East & North Africa	Upper middle income	BAR.NOED.1519.FE.ZS	Attainment	Algeria	Barro-Lee: Percentage of female population age...	28.38	NaN	NaN	...	NaN	NaN	NaN	NaN	6.79	NaN	NaN
3	ARG	Latin America & Caribbean	Upper middle income	BAR.NOED.1519.FE.ZS	Attainment	Argentina	Barro-Lee: Percentage of female population age...	1.20	NaN	NaN	...	NaN	NaN	NaN	NaN	0.20	NaN	NaN

4- REGROUPEMENT DES DONNÉES

- Après le nettoyage final:
(886 930 , 70) → (97 227,33)



Etape 1 :
Présentation des
données

Etape 2 :
Premier nettoyage
et regroupement
des données

Etape 3 :
Exploration et
choix des
indicateurs

Etape 4 :
Nettoyage et
études sur les
indicateurs

Etape 5 :
Score des pays a
fort potentiel client

Etape 6 :
Discussions des
résultats

RECHERCHE DES INDICATEURS

Un pays potentiel doit être un pays où

- ✓ internet est largement accessible,
- ✓ le niveau de vie des habitants est bon
- ✓ la population est relativement élevée en particulier la population cible (15-24 ans)
- ✓ le taux de scolarisation est considérable
- ✓ les dépenses du gouvernement dans l'éducation sont importantes

RECHERCHE DES INDICATEURS

- Groupement par « **Topic** » et exploration des différents indicateurs

Theme Infrastructure: Communications

```
1 Infrastructure_data=Topic_group.get_group("Infrastructure: Communications")
2 Infrastructure_data['Indicator Name'].value_counts()
```

```
03]: Internet users (per 100 people)      203
     Personal computers (per 100 people)  182
     Name: Indicator Name, dtype: int64
```

Indicateurs a garder

```
1 Infrastructure_data[Infrastructure_data['Indicator Name']=='Internet users (per 100 people)']['Indicator Code'].head(1)
```

```
04]: 232    IT.NET.USER.P2
     Name: Indicator Code, dtype: object
```

INDICATEURS PERTINENTS

Internet	Pourcentage d'utilisateurs d'internet
GDPPerCapita	$\text{PIB}/(\text{Population Totale})$
Population	Population totale
Population:15-25ans	Pourcentage de la population éligible à l'enseignement secondaire et supérieur $(\text{Population } 15-25)/(\text{Population Totale})$
Scolarisation	Le taux de scolarisation en université(%)
Dépenses	Dépenses du gouvernement sur l'éducation % du PIB

Etape 1 :
Présentation des
données

Etape 2 :
Premier nettoyage
et regroupement
des données

Etape 3 :
Exploration et
choix des
indicateurs

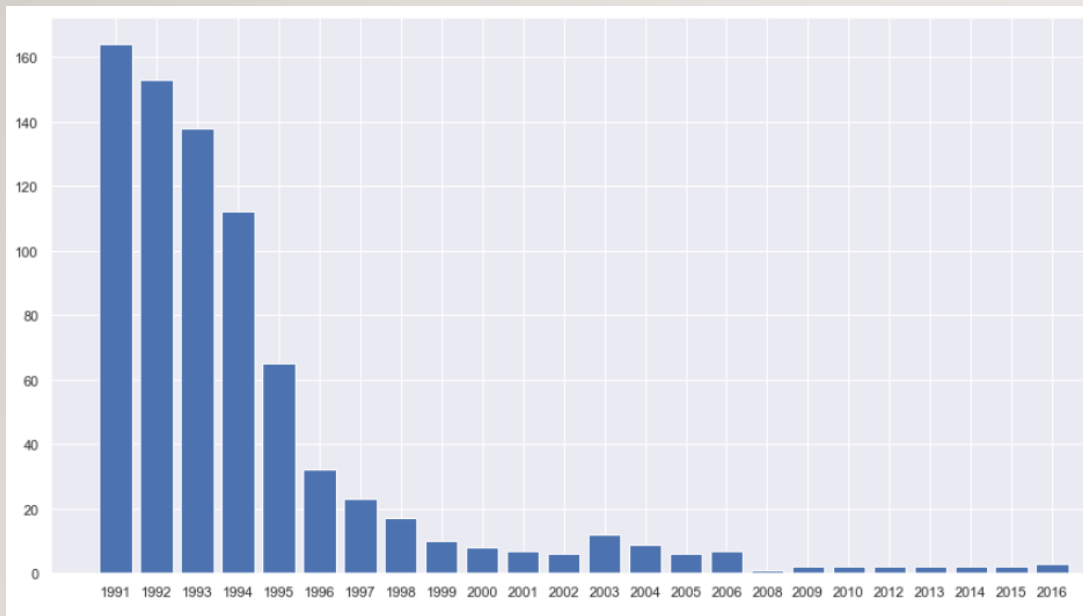
Etape 4 :
Nettoyage et
études sur les
indicateurs

Etape 5 :
Score des pays a
fort potentiel client

Etape 6 :
Discussions des
résultats

ETUDE DES INDICATEURS - INTERNET

- Pourcentage d'utilisation d'internet
- Information sur 229 pays



```
1 internet=dataFinal[dataFinal['Indicator Code']=='IT.NET.USER.P2']
```

```
1 internet.shape
```

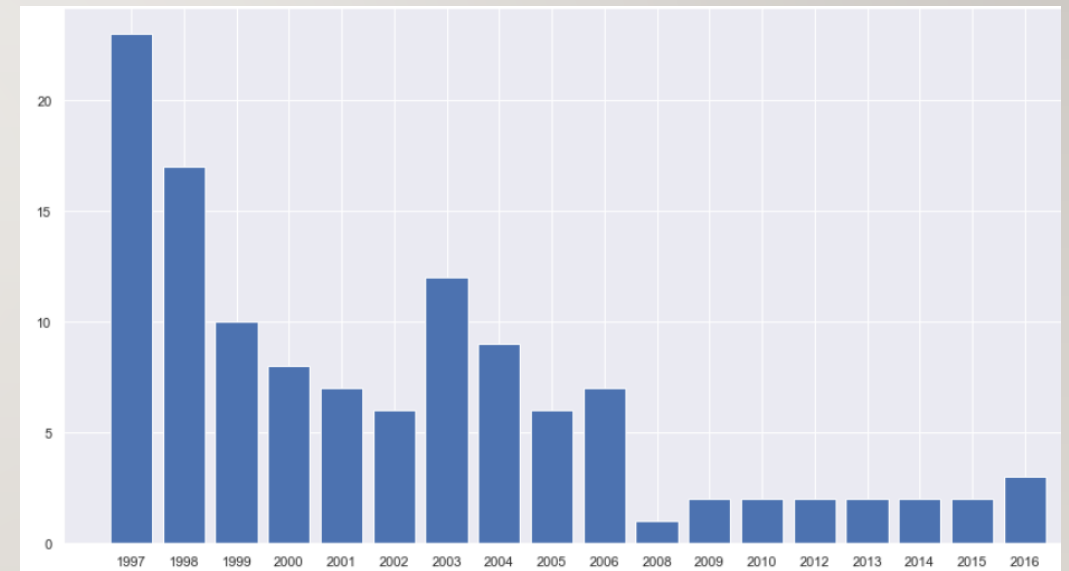
```
(229, 27)
```

- Peu d'informations avant 1996, probablement du au fait que internet n'était pas bien établie avant cette période!
- On continue nos études sur les années **1997 - 20016**

NETTOYAGE DES DONNÉES

Période 1997 - 2016

- On effectue un « backward fill »
- Puis un « forward fill » pour remplacer toutes les valeurs manquantes



Nombre de valeurs NaN par année

STATISTIQUES SUR L'ANNÉE 2016

- Etude du pourcentage d'utilisation d'internet en l'année 2016.
- 50% des pays ont une utilisation internet inférieure à 53.2% en 2016
- Internet est un des critères les plus importants pour l'expansion de « **academy** ».
- On choisit de réduire le nombre de pays à ceux qui ont une utilisation internet supérieure à 50%

1	internet['2016'].describe()
count	203.000000
mean	51.051608
std	28.538088
min	0.000000
25%	25.306275
50%	53.226773
75%	76.143368
max	98.240016

DONNÉES POUR LES ÉTUDES DES INDICATEURS

- 108 pays
- Moyenne d'utilisation d'internet 74.5
- Nous étudierons tous les indicateurs sur cet ensemble de pays

```
1 internet=internet[internet['2016']>50]
2 internet['2016'].describe()
```

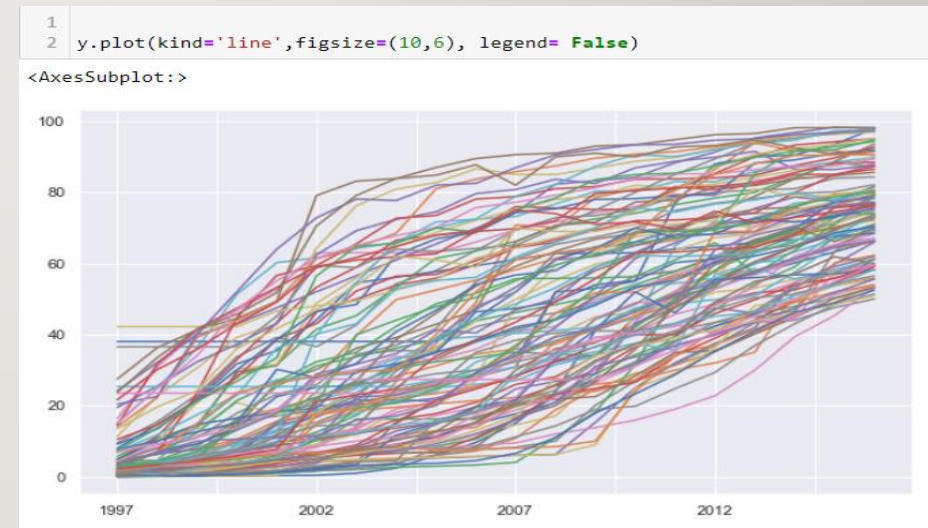
```
count    108.000000
mean      74.520524
std       13.745646
min       50.001637
25%       61.832048
50%       75.249252
75%       85.845775
max       98.240016
Name: 2016, dtype: float64
```

PRÉDICTIONS

- Evolution linéaire dans le temps
- Choix de méthode de prédiction:
Régression linéaire

- Ajouter les informations sur les années 2021, 2026 et 2031 à la table de données de l'indicateur

	2016	2021	2026	2031
	66.363445	85.541568	106.491425	127.441283
	97.930637	139.234060	169.484186	199.734312
	73.000000	92.530608	112.637548	132.744488



% d'utilisation internet par année tous les pays

PRÉDICTION – RÉGRESSION LINÉAIRE

- Calculer la croissance sur les tranches de 5 ans entre 1997 et 2031

$$\frac{\text{Valeur Nouvelle} - \text{Valeur Ancienne}}{\text{Valeur Ancienne}}$$

```
def calculateGrowth(datafr, Growth):  
  
    Growth['1997-2001']=(datafr['2001']-datafr['1997'])/datafr['1997']  
    Growth['2002-2006']=(datafr['2006']-datafr['2002'])/datafr['2002']  
    Growth['2007-2011']=(datafr['2011']-datafr['2007'])/datafr['2007']  
    Growth['2012-2016']=(datafr['2016']-datafr['2012'])/datafr['2012']  
    Growth['2016-2021']=(datafr['2021']-datafr['2016'])/datafr['2016']  
    Growth['2021-2026']=(datafr['2026']-datafr['2021'])/datafr['2021']  
    Growth['2026-2031']=(datafr['2031']-datafr['2026'])/datafr['2026']  
    return(Growth)
```

ETUDE DES INDICATEURS

Pour chaque indicateur

- ✓ Récupérer les informations sur l'indicateur
- ✓ Exploration et nettoyage des données manquantes
- ✓ Utilisation de la régression pour avoir des prédictions pour les années 2021,2026,2031
- ✓ Calcul de la croissance sur des tranches de 5 ans

INDICATEURS

```
1 GDPPerCapita['2021'].describe()
```

```
count      101.000000
mean       26762.632549
std        23401.492855
min        2351.522888
25%        8534.421423
50%       17487.062693
75%       39572.256402
max       118373.318079
Name: 2021, dtype: float64
```

```
1 PopTot['2021'].describe()
```

```
count      1.010000e+02
mean       3.836762e+07
std        1.468556e+08
min        5.663390e+04
25%       1.818143e+06
50%       6.364389e+06
75%       2.538831e+07
max        1.419392e+09
Name: 2021, dtype: float64
```

```
1 PctPop1524['2021'].describe()
```

```
count      95.000000
mean       13.363416
std        4.019556
min        0.711509
25%       10.923299
50%       12.950496
75%       15.930660
max       25.350054
Name: 2021, dtype: float64
```

```
1 scholarisation['2021'].describe()
```

```
count      90.000000
mean       67.640450
std        27.874762
min         6.112482
25%       47.594695
50%       70.699554
75%       87.476786
max       146.059401
Name: 2021, dtype: float64
```

```
1 depenses['2021'].describe()
```

```
count      92.000000
mean       4.736311
std        1.904847
min        1.288312
25%       3.287346
50%       4.800850
75%       5.956576
max       10.057696
Name: 2021, dtype: float64
```

Etape 1 :
Présentation des
données

Etape 2 :
Premier nettoyage
et regroupement
des données

Etape 3 :
Exploration et
choix des
indicateurs

Etape 4 :
Nettoyage et
études sur les
indicateurs

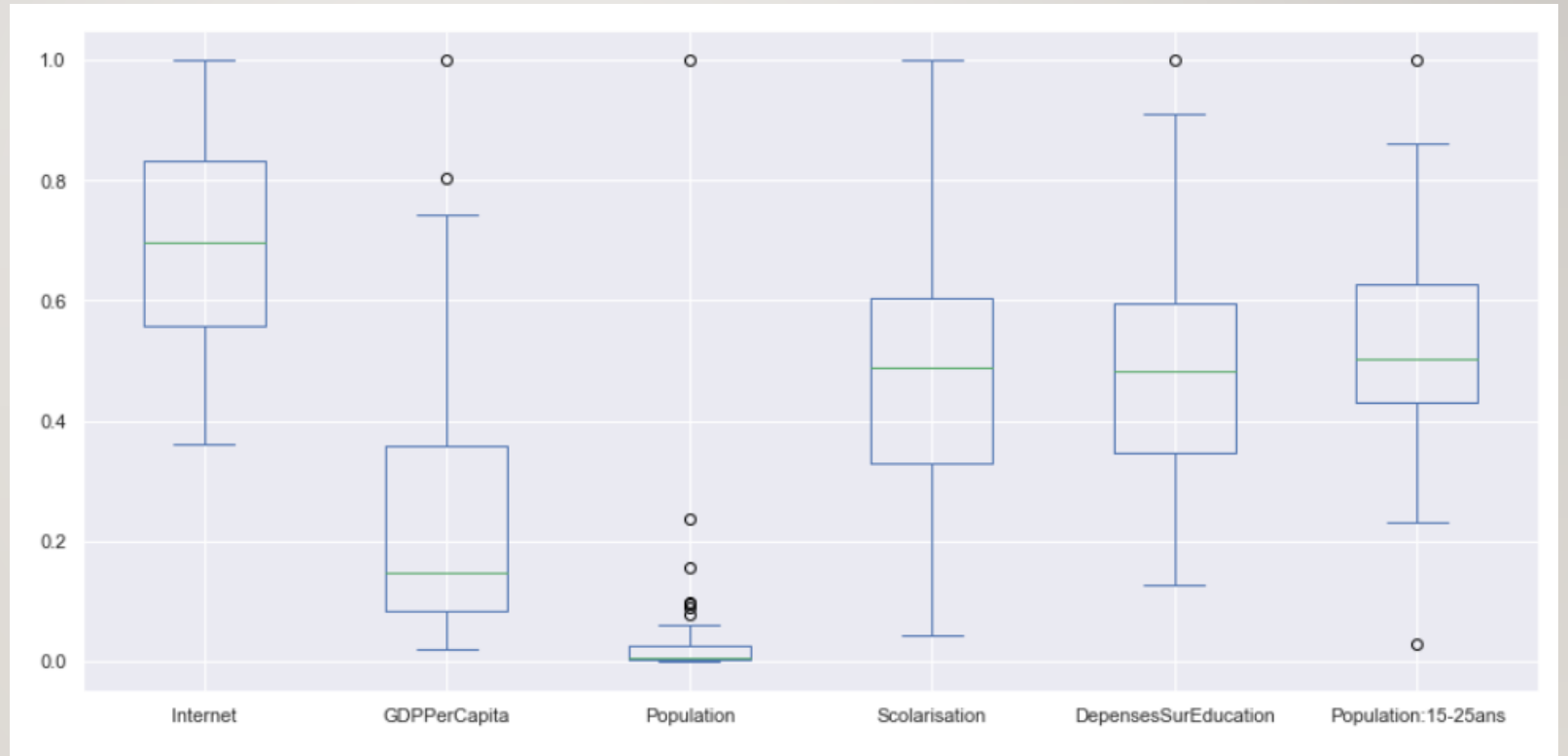
Etape 5 :
Score des pays a
fort potentiel client

Etape 6 :
Discussions des
résultats

NORMALISATION

- Normalisation des valeurs des indicateurs :

$$\frac{\text{valeur Indicateur}}{\text{Max(valeur Indicateur)}}$$



CORRÉLATION ENTRE INDICATEURS

- **GDPPerCapita** est fortement corrélé a **Internet**.
- **Population:15-25ans** est modérément corrélé a **Internet** et **GDPerCapita**
- Corrélation faible entre **Scolarisation** et **DepensesSurEducation**

	Internet	GDPPerCapita	Population	Population:15-25ans	Scolarisation	DepensesSurEducation
Internet	1.000000	0.747835	-0.175967	-0.574172	0.213088	0.184997
GDPPerCapita	0.747835	1.000000	-0.092365	-0.448909	0.078821	0.137143
Population	-0.175967	-0.092365	1.000000	0.056687	-0.039563	-0.135410
Population:15-25ans	-0.574172	-0.448909	0.056687	1.000000	-0.219548	-0.086065
Scolarisation	0.213088	0.078821	-0.039563	-0.219548	1.000000	0.315133
DepensesSurEducation	0.184997	0.137143	-0.135410	-0.086065	0.315133	1.000000

SCORE SYNTHÉTIQUE

- A cause de leur corrélation avec les autres indicateurs on choisit d'omettre de la fonction de score synthétique **Internet** et **Population: 15-25ans**
- Les scores n'ont pas les mêmes unités de mesure, nous choisirons la **multiplication** pour l'agrégation des scores
- Score synthétique:

```
Scores['GDPPerCapita']*Scores['Scolarisation']*Scores['Population']*Scores['DepensesSurEducation']
```

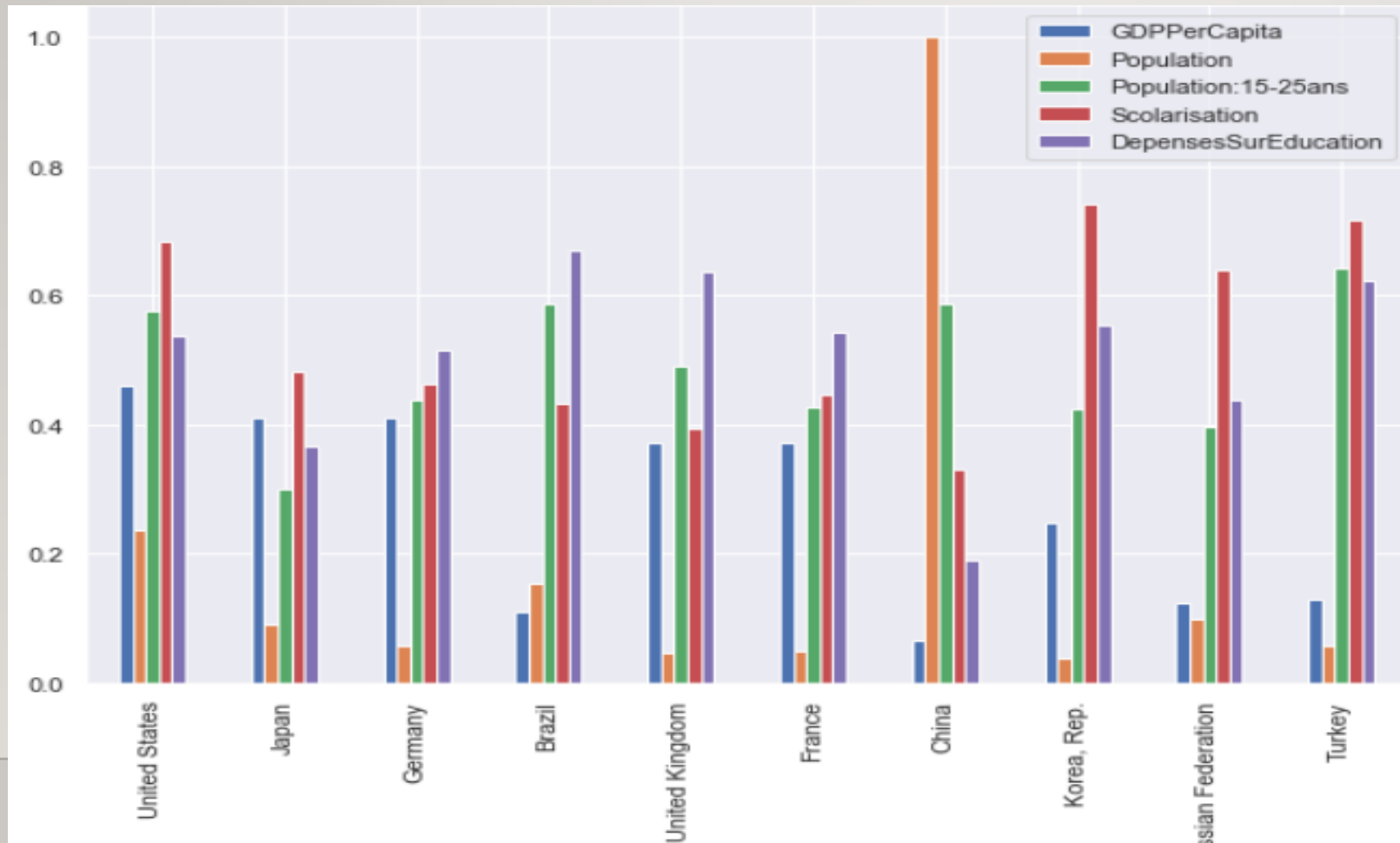
PAYS A FORT POTENTIEL CLIENT TOP 10

Country Name	Region	Internet	GDPPerCapita	Population	Scolarisation	DepensesSurEducation	Population:15-25ans	Score
United States	North America	0.699461	0.461785	0.237935	0.683246	0.538755	0.575854	0.040445
Japan	East Asia & Pacific	0.895673	0.410009	0.090234	0.481721	0.367940	0.300353	0.006557
Germany	Europe & Central Asia	0.896267	0.409657	0.057067	0.463206	0.514560	0.438728	0.005572
Brazil	Latin America & Caribbean	0.570093	0.108903	0.154608	0.432167	0.671126	0.588247	0.004883
United Kingdom	Europe & Central Asia	0.937916	0.373459	0.047361	0.394144	0.638391	0.489972	0.004450
France	Europe & Central Asia	0.883713	0.372336	0.048686	0.446848	0.544591	0.426324	0.004411
China	East Asia & Pacific	0.486087	0.067255	1.000000	0.331826	0.188900	0.587652	0.004216
Korea, Rep.	East Asia & Pacific	0.898743	0.246715	0.037061	0.741946	0.554958	0.424350	0.003765
Russian Federation	Europe & Central Asia	0.687200	0.123430	0.099522	0.639850	0.438815	0.397201	0.003449
Turkey	Europe & Central Asia	0.533210	0.128788	0.058888	0.717403	0.622776	0.641355	0.003388

PAYS A FORT POTENTIEL ÉCONOMIQUE – TOP 10 PAR RÉGION GÉOGRAPHIQUE



PAYS A FORT POTENTIEL ÉCONOMIQUE – TOP 10 PAR INDICATEUR



On voit la décomposition du score synthétique

- Les États-Unis sont premiers par le GDPPerCapita le taux de la population 15-25ans et le taux de scolarisation universitaire
- La position de la Chine est en priorité du a sa population totale et pourcentage entre 15et24 ans

Etape 1 :
Présentation des
données

Etape 2 :
Premier nettoyage
et regroupement
des données

Etape 3 :
Exploration et
choix des
indicateurs

Etape 4 :
Nettoyage et
études sur les
indicateurs

Etape 5 :
Score des pays a
fort potentiel client

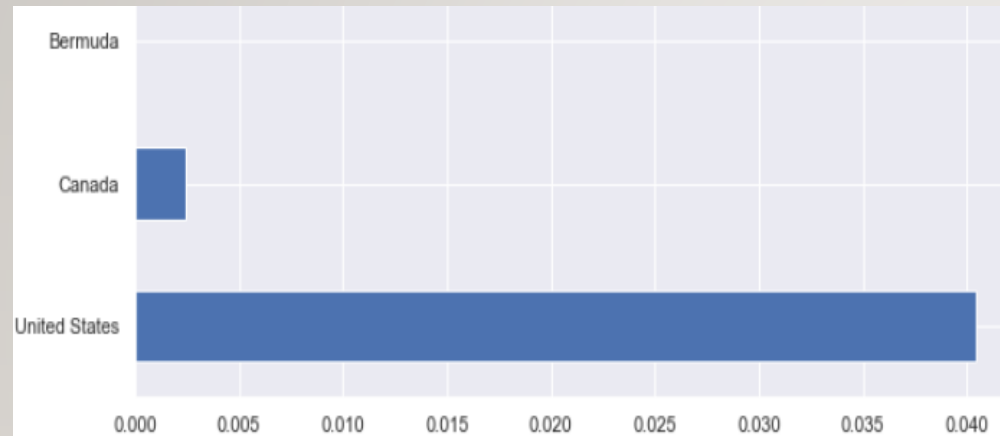
Etape 6 :
Discussions des
résultats

TOP 5 PAR RÉGION GÉOGRAPHIQUE

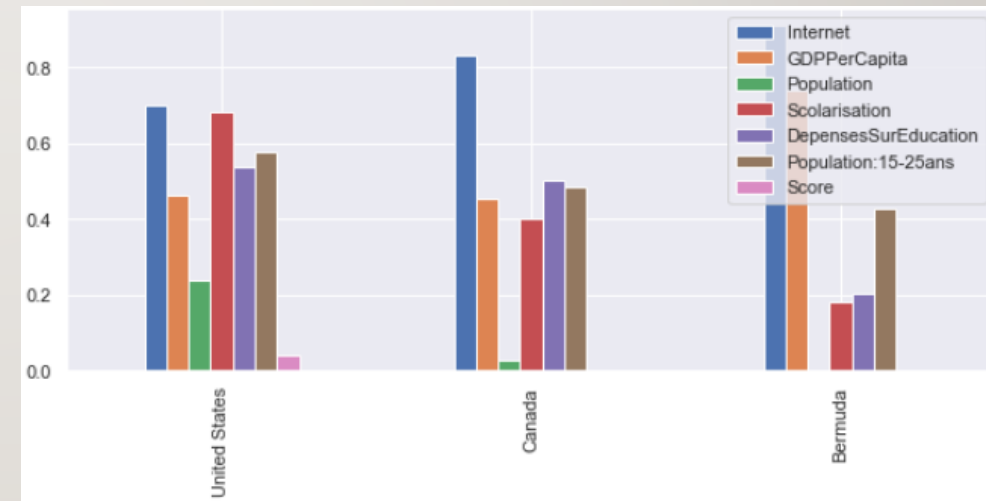
On fait un groupement par région géographique

On étudie le classement des pays dans leur zone géographique

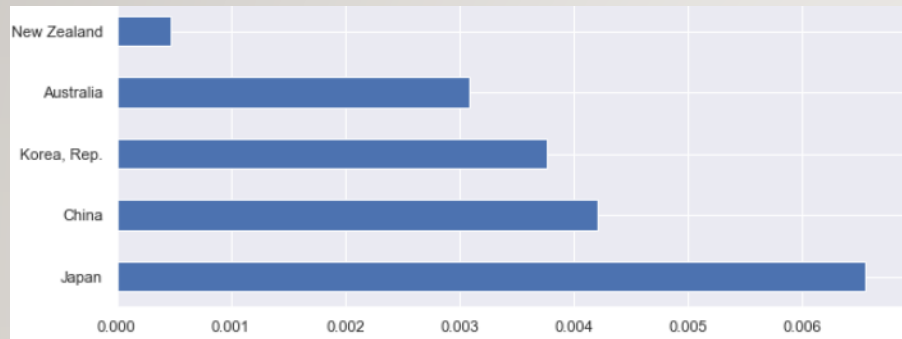
AMÉRIQUE DU NORD



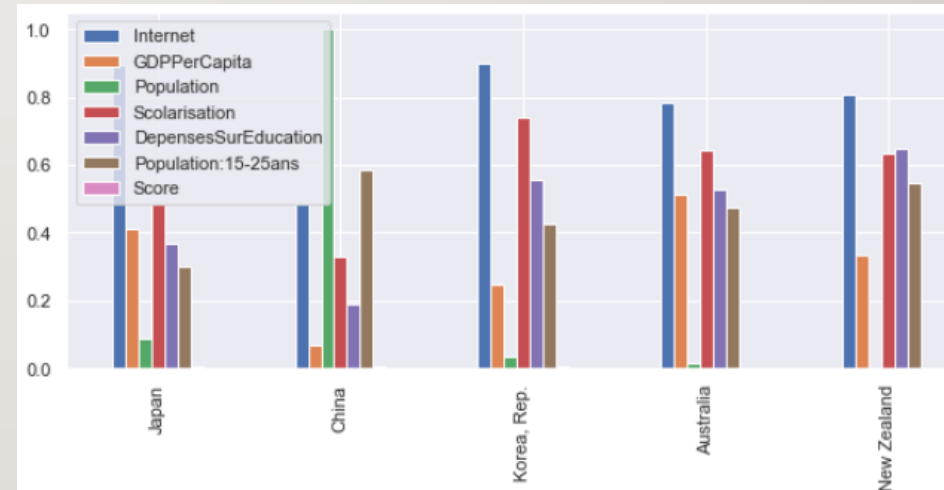
```
North America
count    3.000000
mean     0.014299
std      0.022676
min      0.000001
25%      0.001226
50%      0.002451
75%      0.021448
max      0.040445
```



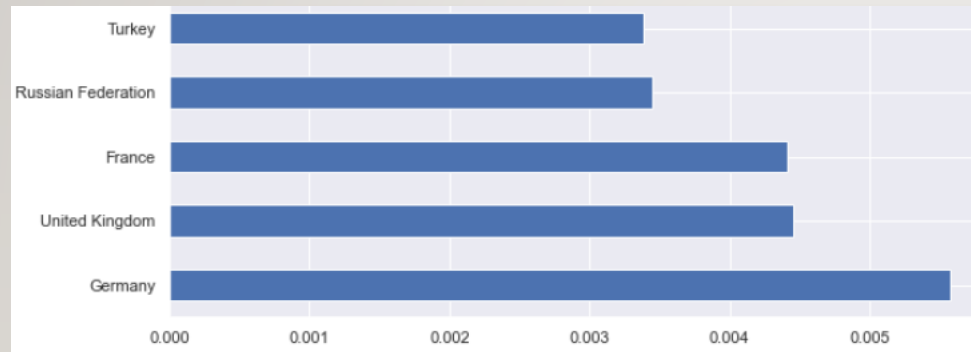
ASIE DE L'EST ET PACIFIQUE



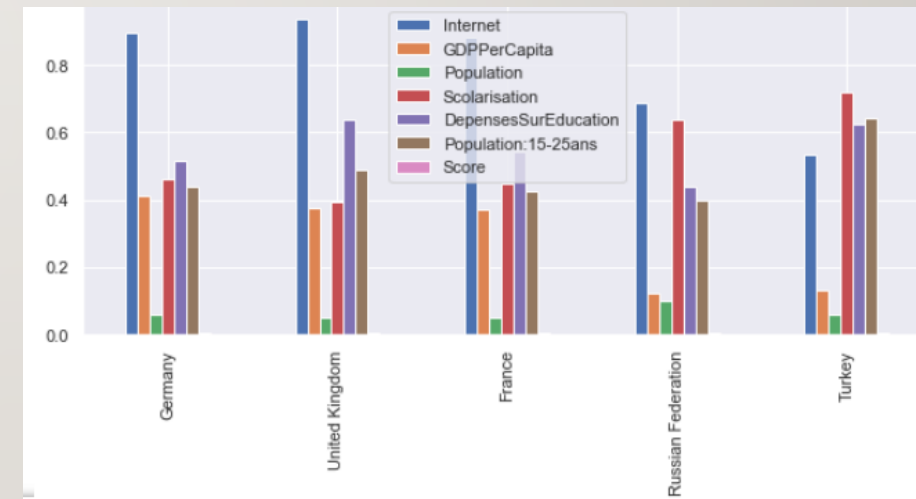
```
East Asia & Pacific
count    10.000000
mean      0.001885
std       0.002342
min       0.000005
25%       0.000145
50%       0.000407
75%       0.003595
max       0.006557
```



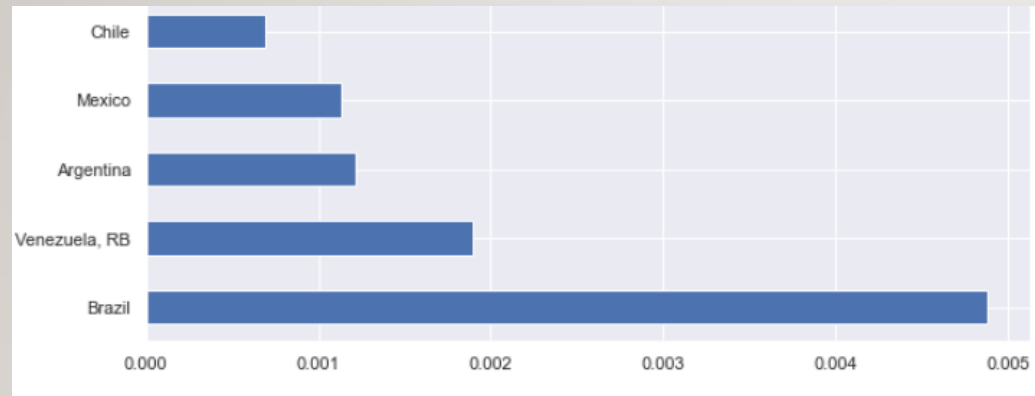
EUROPE ET ASIE CENTRALE



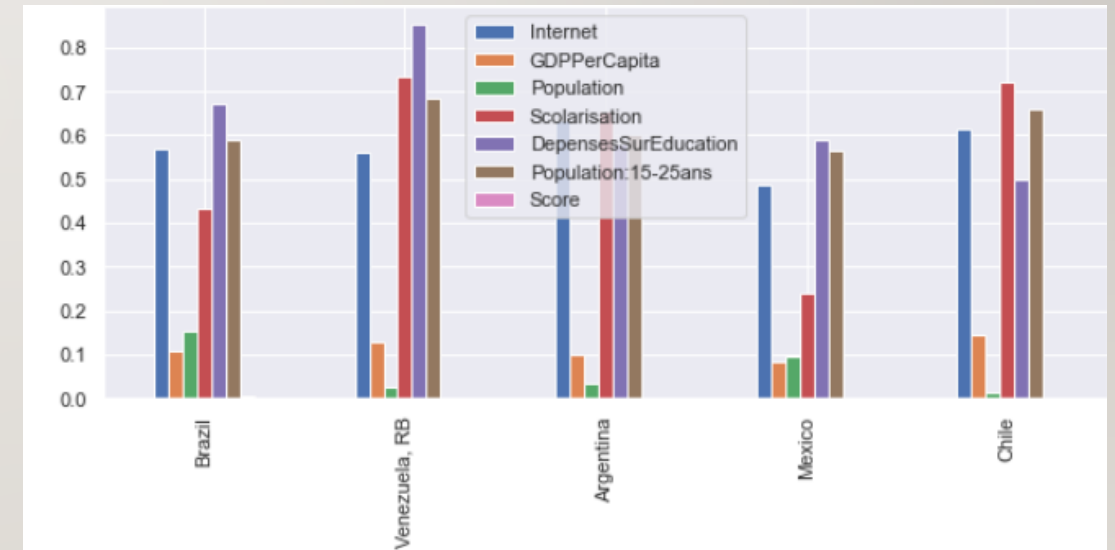
```
Europe & Central Asia
count    41.000000
mean     0.001029
std      0.001437
min      0.000005
25%      0.000061
50%      0.000382
75%      0.001199
max      0.005572
```



AMÉRIQUE LATINE ET CARAÏBE



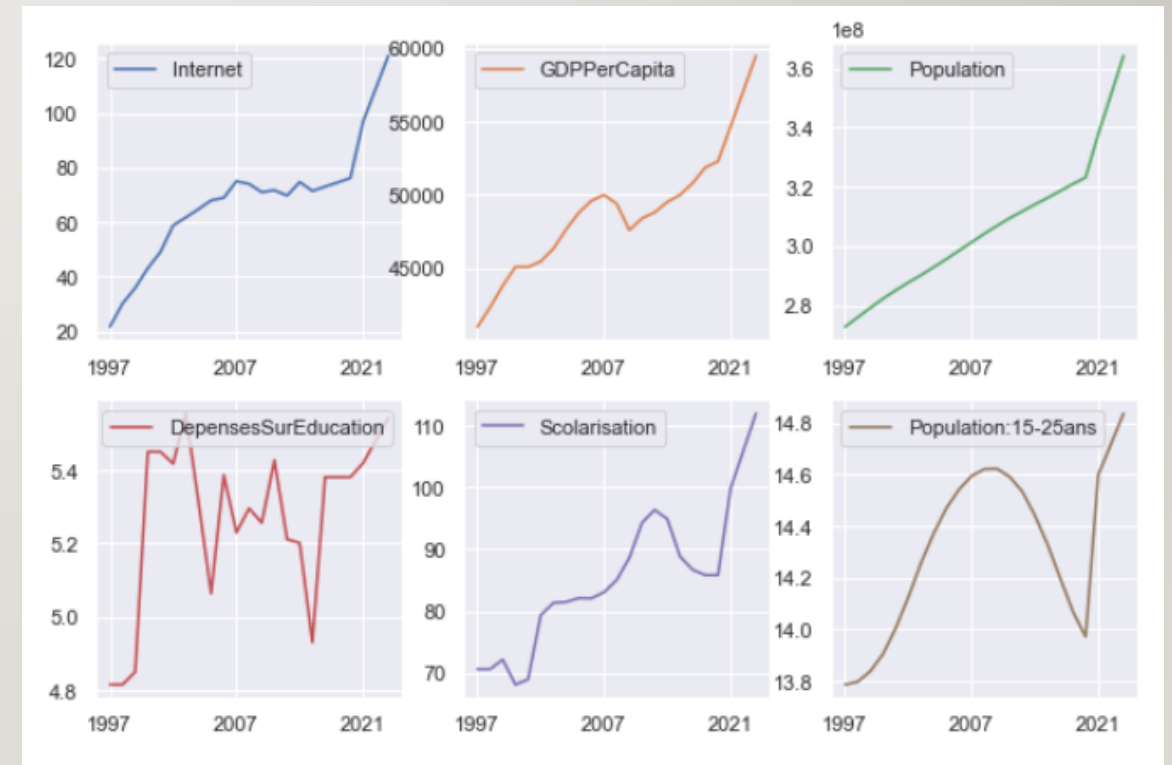
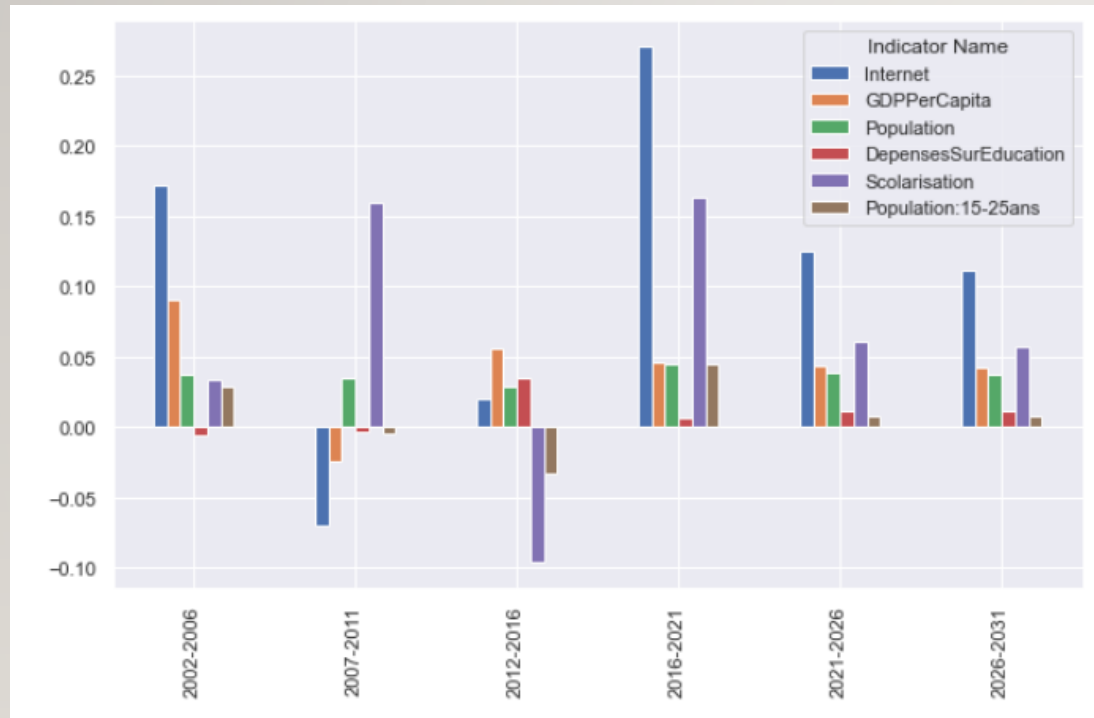
```
Latin America & Caribbean
count    1.500000e+01
mean     7.176597e-04
std      1.288932e-03
min      8.344564e-07
25%      2.446797e-05
50%      1.086597e-04
75%      9.123234e-04
max      4.883449e-03
```



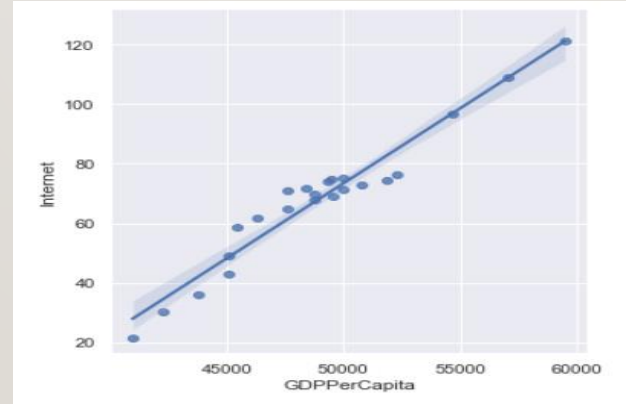
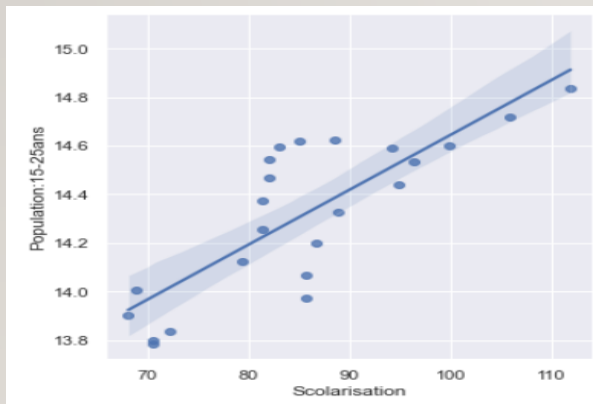
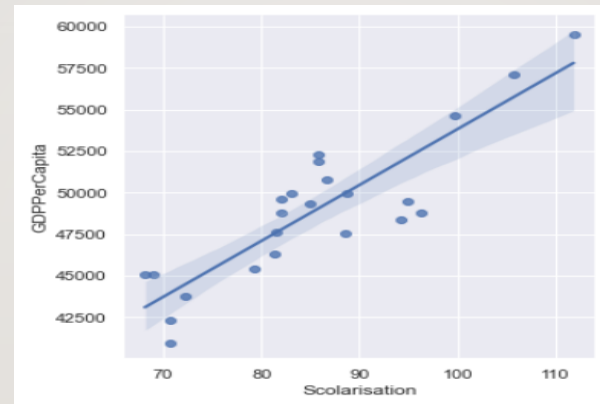
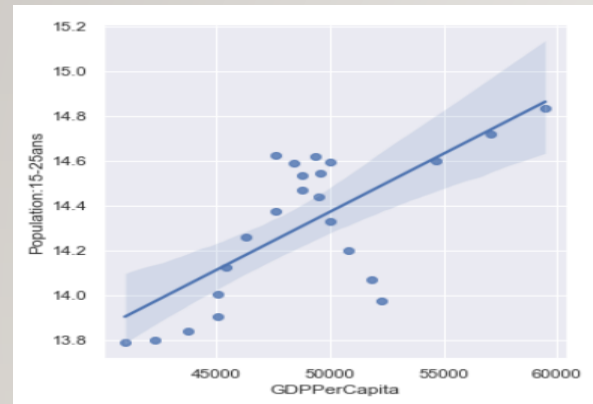
ETUDE PAR PAYS

- ✓ Croissance par indicateurs
- ✓ Courbes d'évolution de chaque indicateur par rapport aux années
- ✓ Relation entre PIB par habitants et internet et avec le taux de scolarisation
- ✓ Relation entre la proportion de la population entre 15 et 24ans avec le PIB par habitants et avec le taux de scolarisation

ETATS UNIS - CROISSANCE

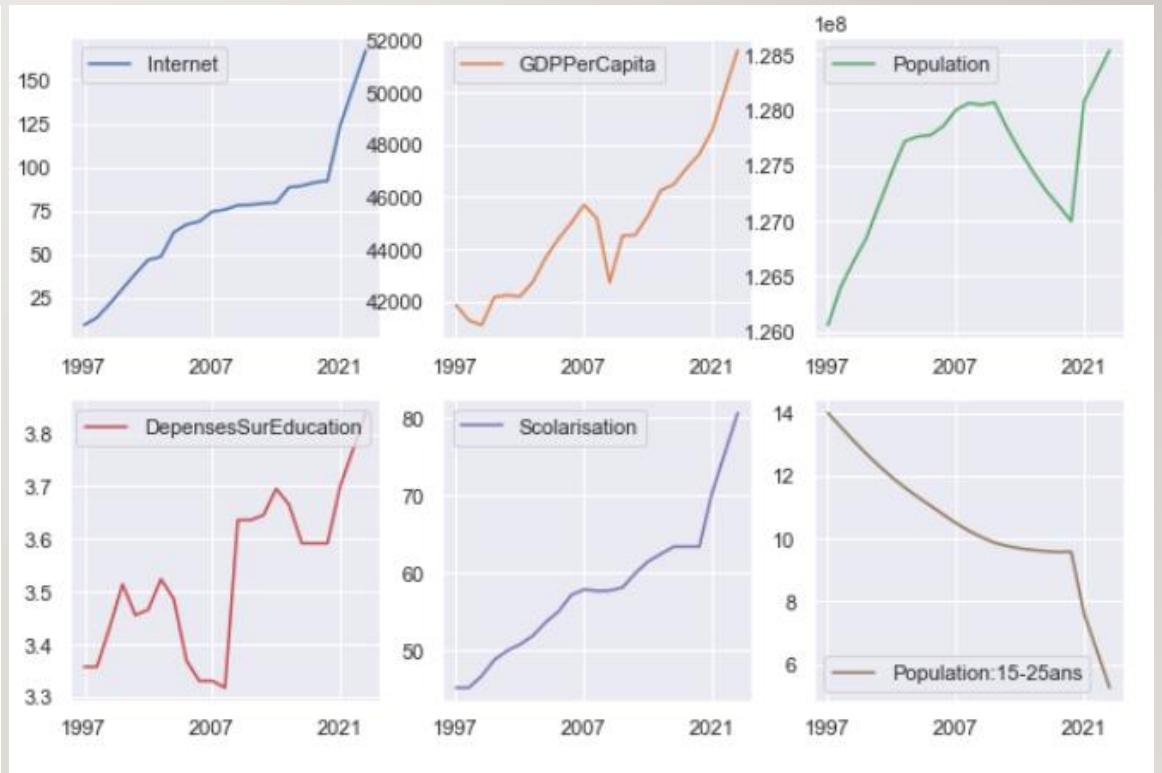
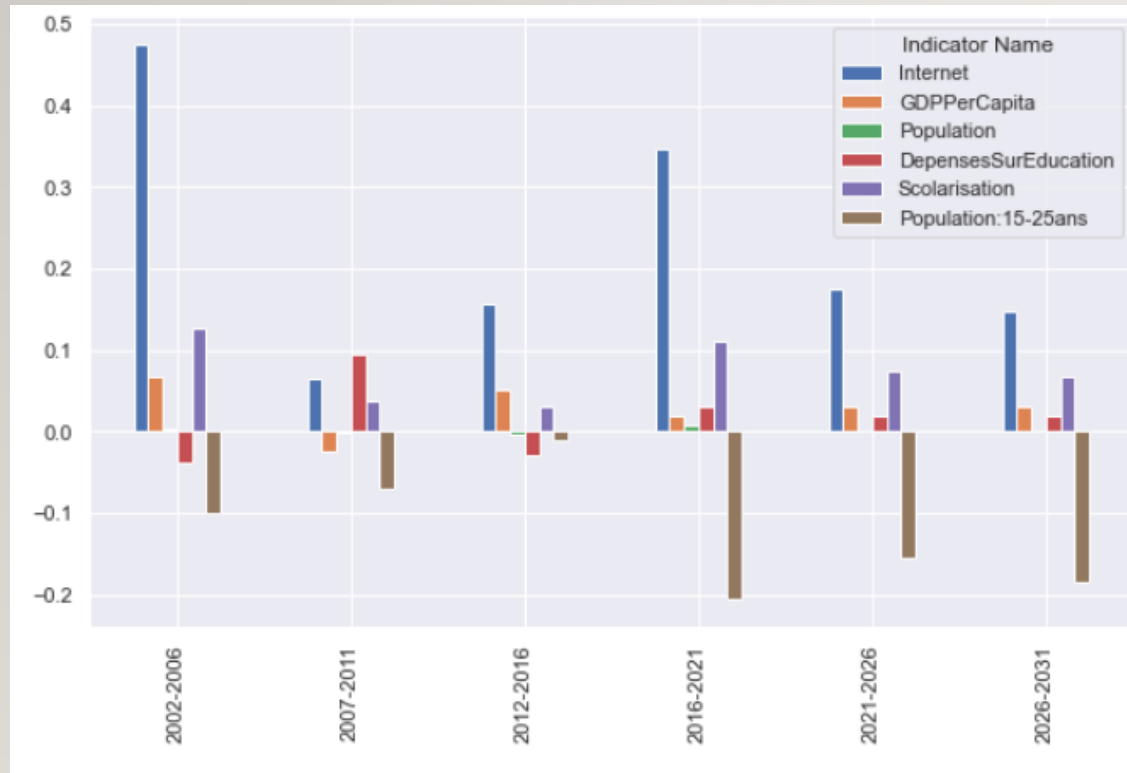


ETATS UNIS

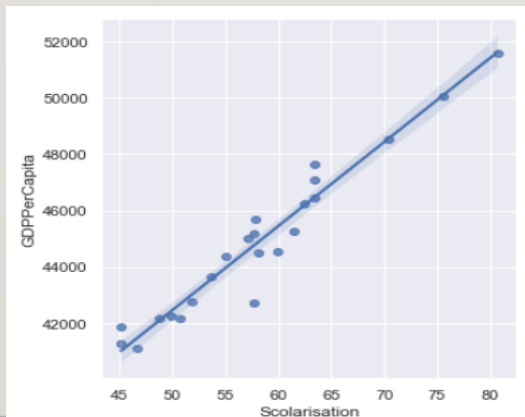
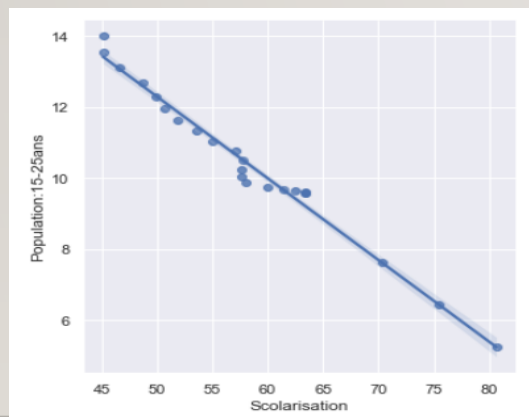
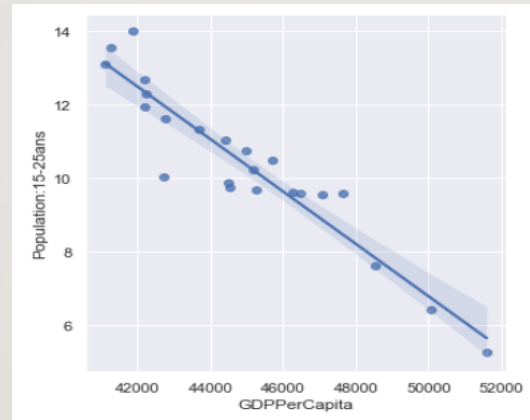
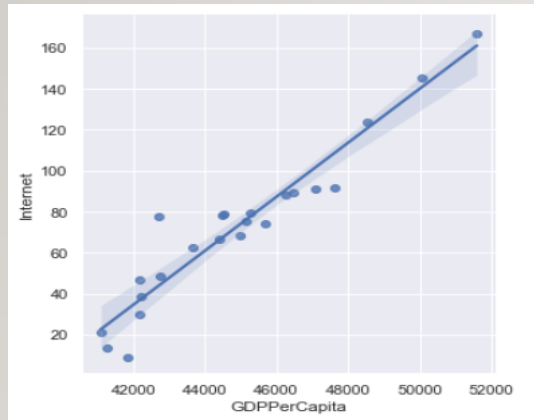


- Pas de relation entre la proportion de la population 15-25 ans et le taux de scolarisation ou le PIB

JAPON

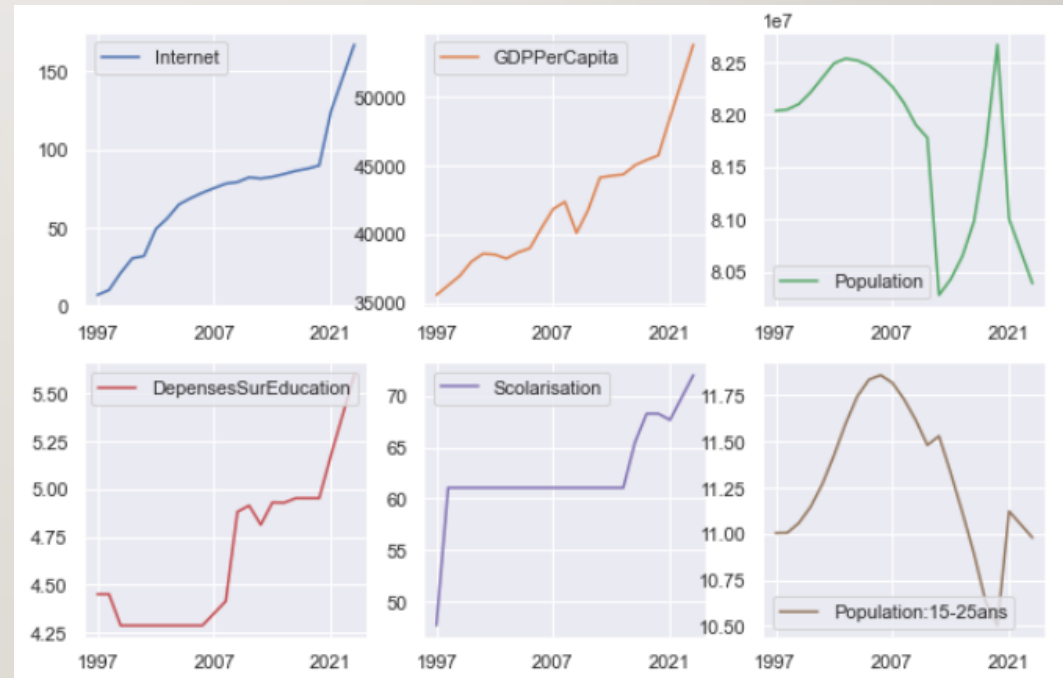
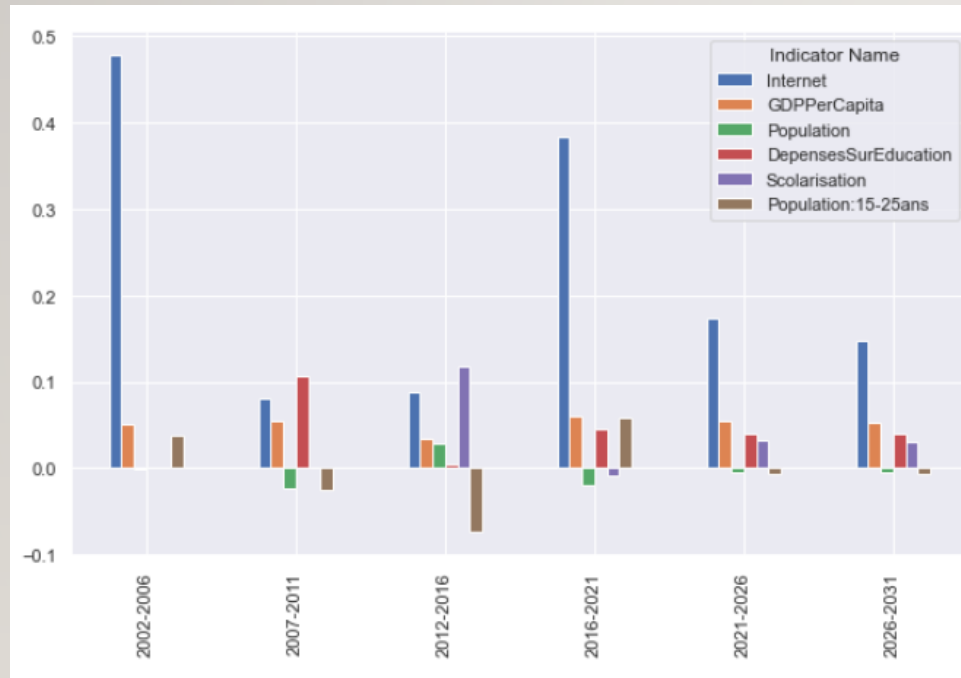


JAPON

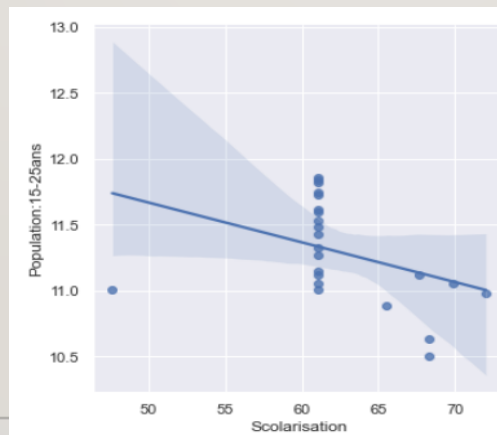
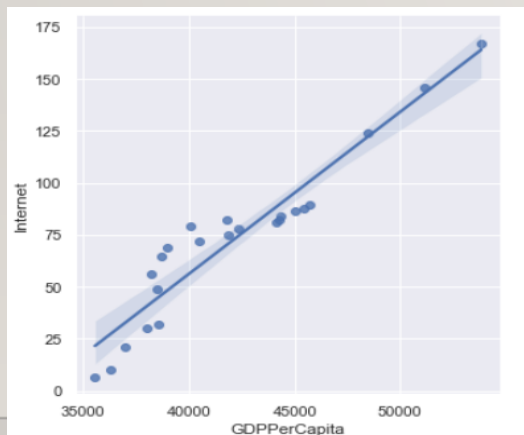
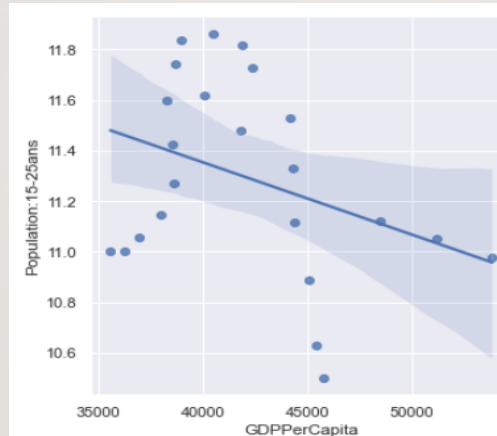
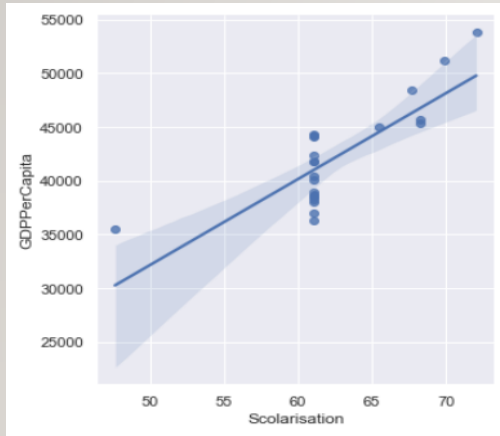


- La proportion de la population 15-25 ans est négativement corrélée au taux de scolarisation et au PIB

ALLEMAGNE

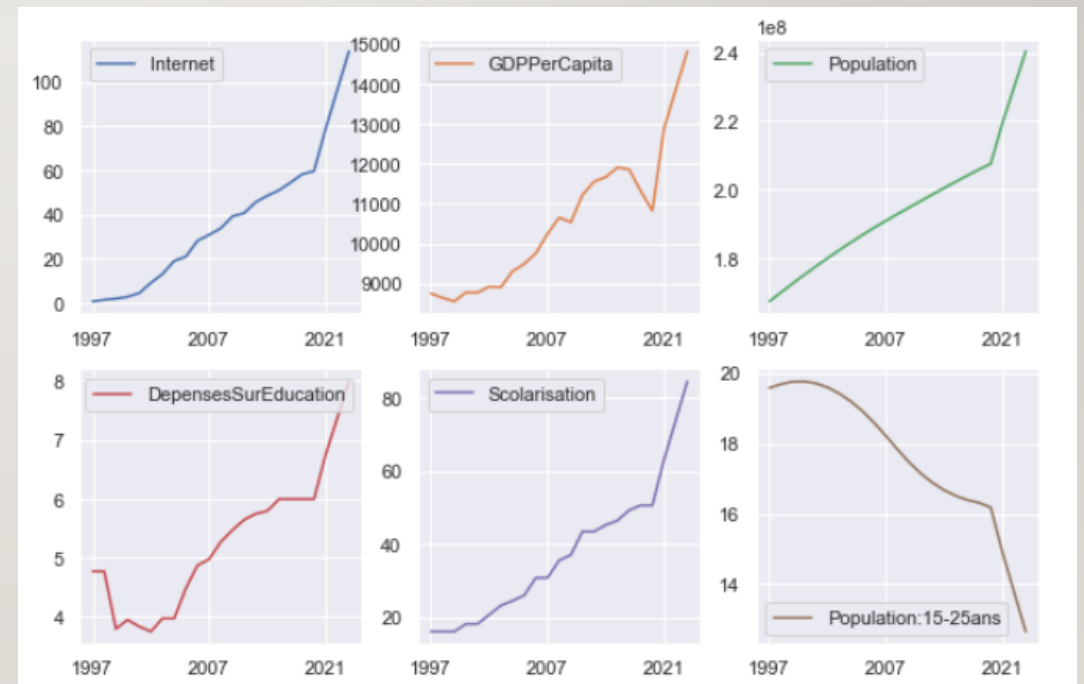
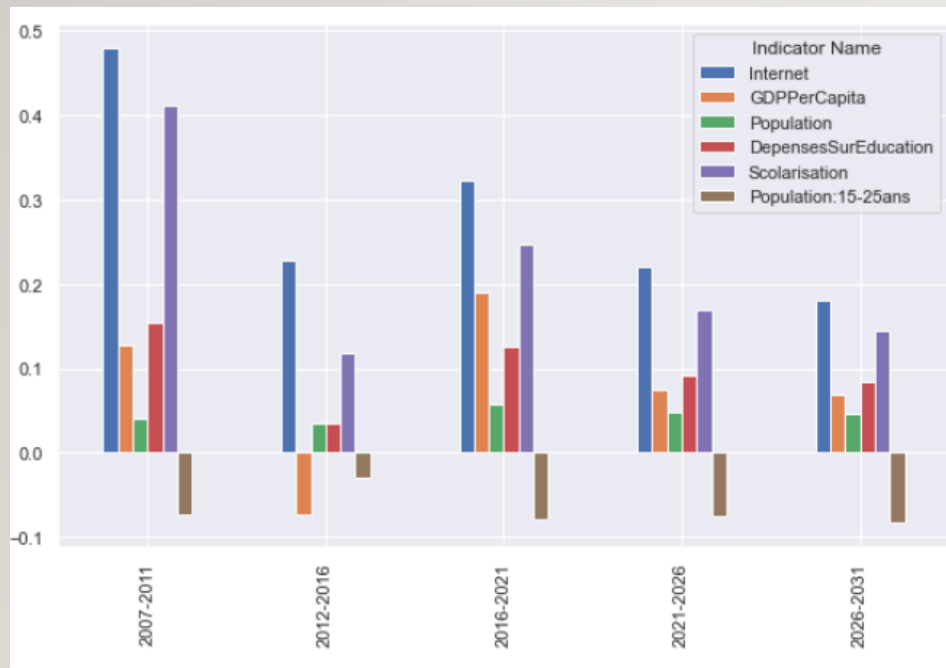


ALLEMAGNE

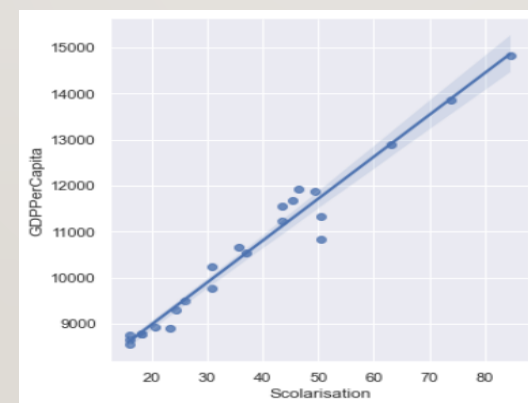
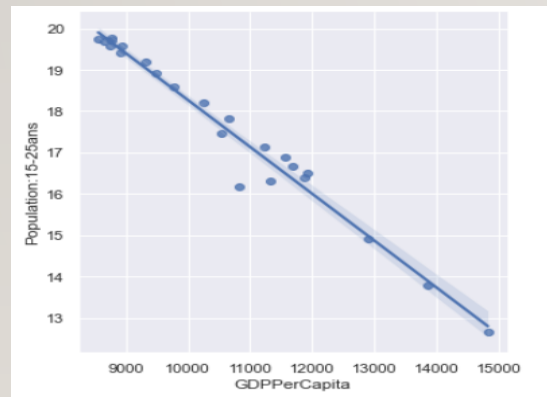
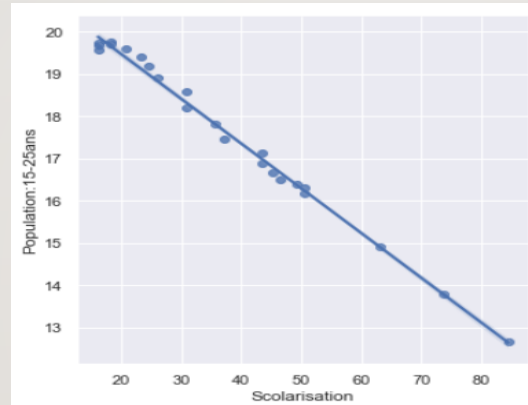
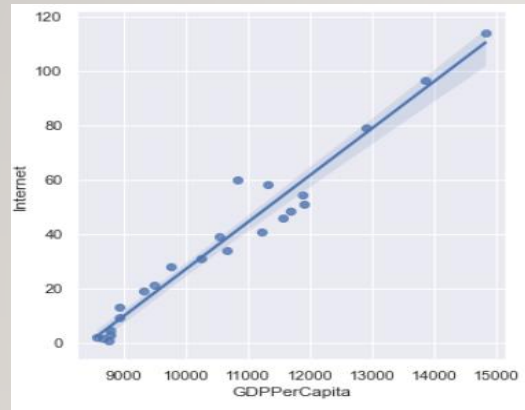


- Le taux de scolarisation est indépendant du PIB et de la population 15-25 ans
- La population 15-25 ans n'a pas d'influence sur le PIB du pays

BRÉSIL

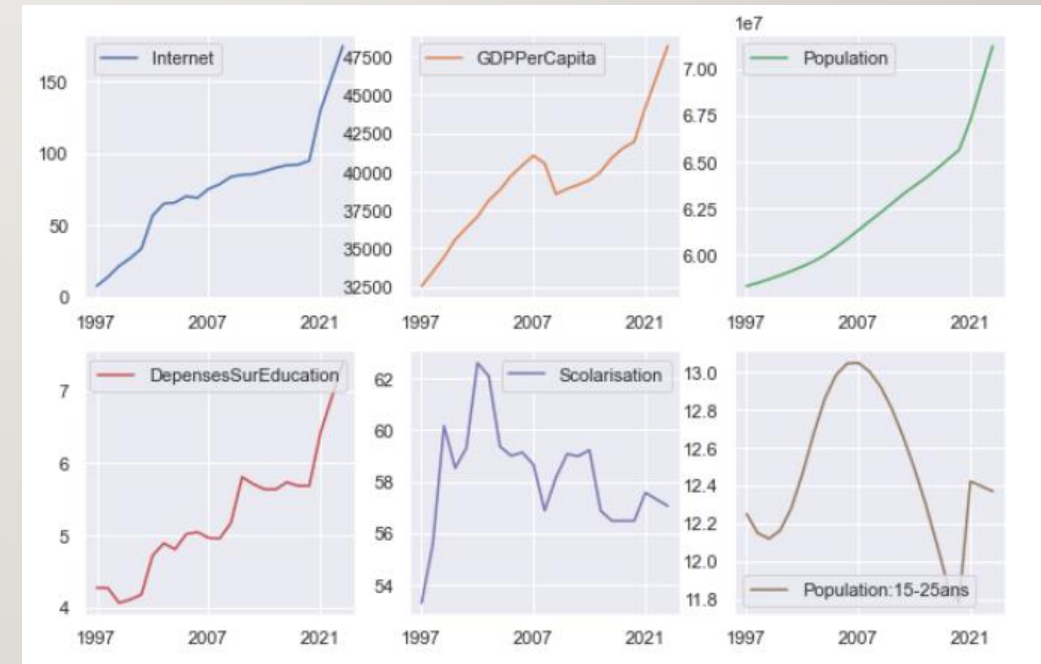
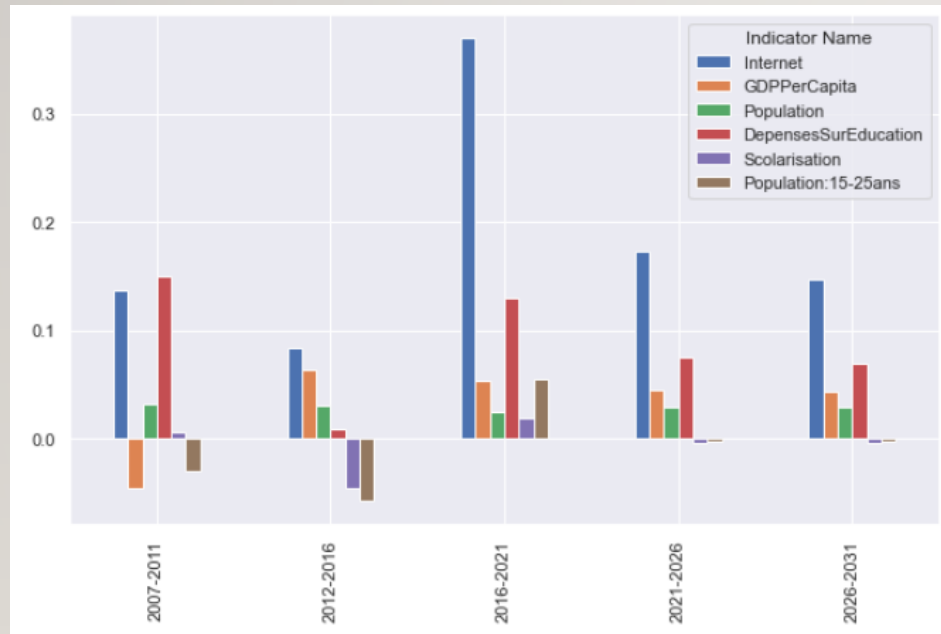


BRÉSIL

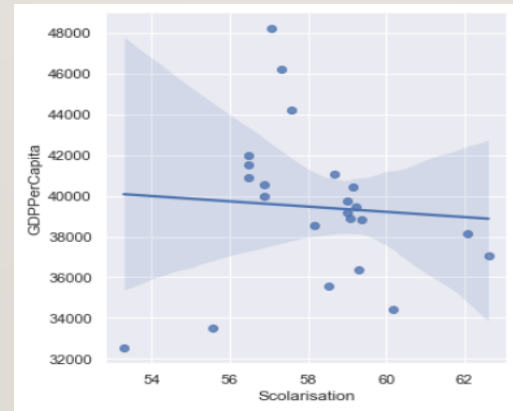
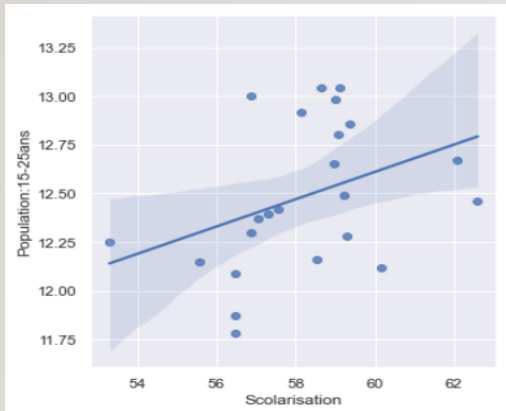
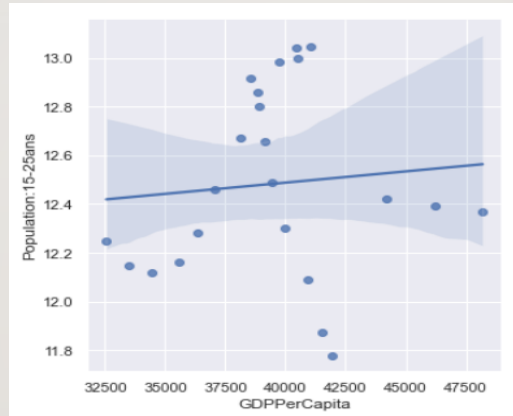
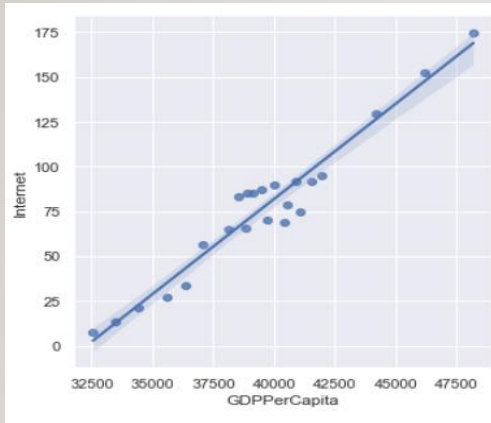


- La proportion de la population 15-25 ans est négativement corrélée au taux de scolarisation et au PIB

ROYAUME-UNI



ROYAUME-UNI



- Le taux de scolarisation est indépendant du PIB et de la population 15-25 ans
- La population 15-25 ans n'a pas d'influence sur le PIB du pays

CONCLUSION

- La croissance négative de la population au Japon et au Brésil peut être considéré inhibiteur pour l'exploitation du pays
- Un premier choix serait une expansion vers les **Etats unis**
- Puis une expansion vers l'Europe, notamment l'**Allemagne** et le **Royaume-Uni**

Merci!