

The background features abstract geometric shapes in shades of blue and grey at the top and bottom. On the left side, there are several thin, grey circuit-like lines with small circles at their ends, extending horizontally and slightly angled. A similar set of lines is located on the right side, also extending horizontally.

Analysez des données de systèmes éducatifs

Nabil ERRAHMANI

Sommaire

- I. Introduction
- II. Chargement et Exploration des Données
- III. Nettoyage des Données
- IV. Sélection et Analyse des Indicateurs Pertinents
- V. Suppression des régions
- VI. Identification et Visualisation des Valeurs Aberrantes
- VII. Analyse Comparative des Indicateurs de Développement
- VIII. Conclusion



01 INTRODUCTION



02. Chargement et Exploration des Données

Fichier EdStatsData.csv:

Points clés:

- Chargement du fichier principal EdStatsData.csv contenant les données de la Banque mondiale.
- Le fichier contient un ensemble riche d'indicateurs éducatifs mondiaux.

	Country Name	Country Code	Indicator Name	Indicator Code	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
0	Arab World	ARB	Adjusted net enrolment rate, lower secondary, both sexes (%)	UIS.NERA.2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Arab World	ARB	Adjusted net enrolment rate, lower secondary, female (%)	UIS.NERA.2.F	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Arab World	ARB	Adjusted net enrolment rate, lower secondary, gender parity index (GPI)	UIS.NERA.2.GPI	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Arab World	ARB	Adjusted net enrolment rate, lower secondary, male (%)	UIS.NERA.2.M	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	Arab World	ARB	Adjusted net enrolment rate, primary, both sexes (%)	SE.PRM.TENR	54.822121	54.894138	56.209438	57.267109	57.991138	59.36554	60.999962	61.92268	62.69342	64.383186	65.617767	66.085152	66.608139	67.290451	68.510094	69.033211	69.944908	71.04167	71.693779	71.699097	71.995819	72.602837	70.032722	70.464621	72.645683	71.81176	73.903511	74.425201	75.1108

Ofi. Chargement et Exploration des Données

Fichier EdStatsData.csv:

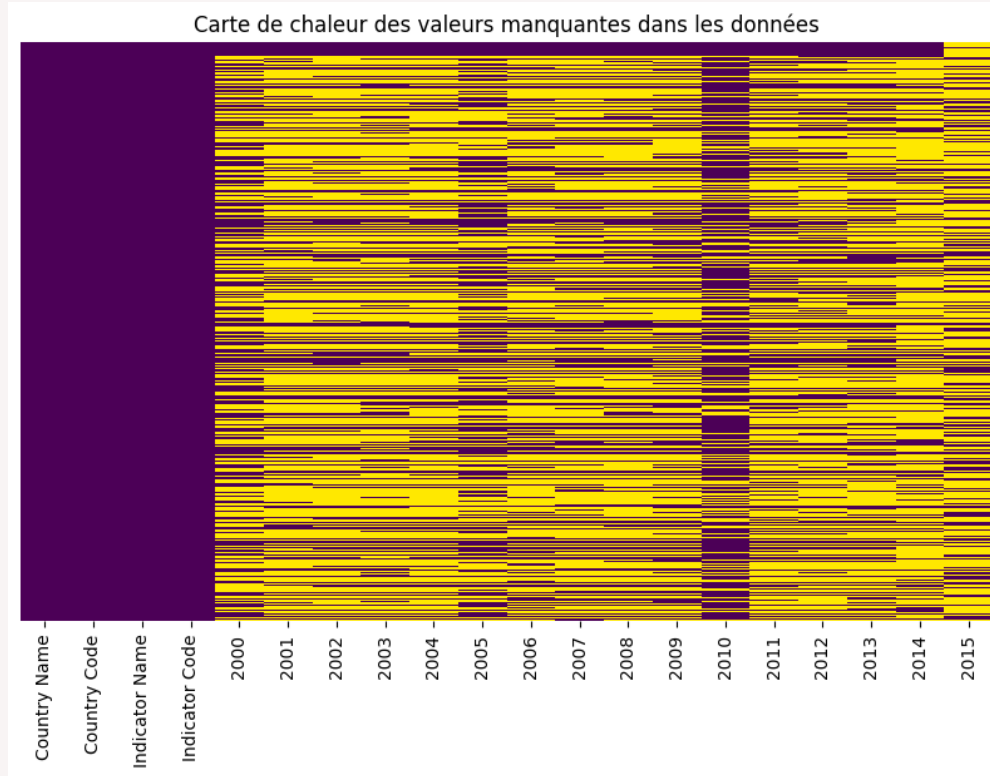
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 886930 entries, 0 to 886929
Data columns (total 70 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Country Name        886930 non-null  object
1   Country Code        886930 non-null  object
2   Indicator Name      886930 non-null  object
3   Indicator Code      886930 non-null  object
4   1970                72288 non-null   float64
5   1971                35537 non-null   float64
6   1972                35619 non-null   float64
7   1973                35545 non-null   float64
8   1974                35730 non-null   float64
9   1975                87306 non-null   float64
10  1976                37483 non-null   float64
11  1977                37574 non-null   float64
12  1978                37576 non-null   float64
13  1979                36809 non-null   float64
14  1980                89122 non-null   float64
15  1981                38777 non-null   float64
16  1982                37511 non-null   float64
17  1983                38460 non-null   float64
18  1984                38606 non-null   float64
19  1985                90296 non-null   float64
...
68  2100                51436 non-null   float64
69  Unnamed: 69         0 non-null       float64
dtypes: float64(66), object(4)
memory usage: 473.7+ MB
```

02. Chargement et Exploration des Données

Fichiers:

- EdStatsCountry-Series.csv
- EdStatsCountry.csv
- EdStatsSeries.csv
- EdStatsFootNote.csv (exemple en image)

	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



03. Nettoyage des données

b. Traitement des doublons

Grâce à la méthode `. duplicated()`, on recherche les doublons:

```
# Vérification des données dupliquées
duplicates = cleaned_df.duplicated().sum()

if duplicates > 0:
    print(f"Le jeu de données contient {duplicates} lignes dupliquées.")
else:
    print("Pas de lignes dupliquées dans le jeu de données.")
```

✓ 0.4s

Pas de lignes dupliquées dans le jeu de données.

04. Sélection et Analyse des Indicateurs Pertinents

- **UIS.E.5.B** : Nombre total d'étudiants inscrits dans l'enseignement supérieur, reflet de l'importance de l'éducation avancée.
- **UIS.E.3** : Nombre total d'étudiants inscrits au niveau secondaire supérieur, indicateur du niveau d'éducation de base.
- **NY.GDP.PCAP.PP.KD** : PIB, Revenu moyen par personne, mesure de la prospérité économique du pays.
- **SE.TER.ENRR** : Pourcentage de la population en âge d'aller à l'université inscrite dans l'enseignement supérieur.
- **SE.SEC.ENRR.UP** : Pourcentage de la population inscrite au niveau secondaire supérieur, information sur l'accessibilité de l'éducation pré-universitaire.
- **IT.NET.USER.P2** : Nombre d'utilisateurs d'Internet pour 100 habitants, indice de la connectivité numérique.

05. Suppression des régions

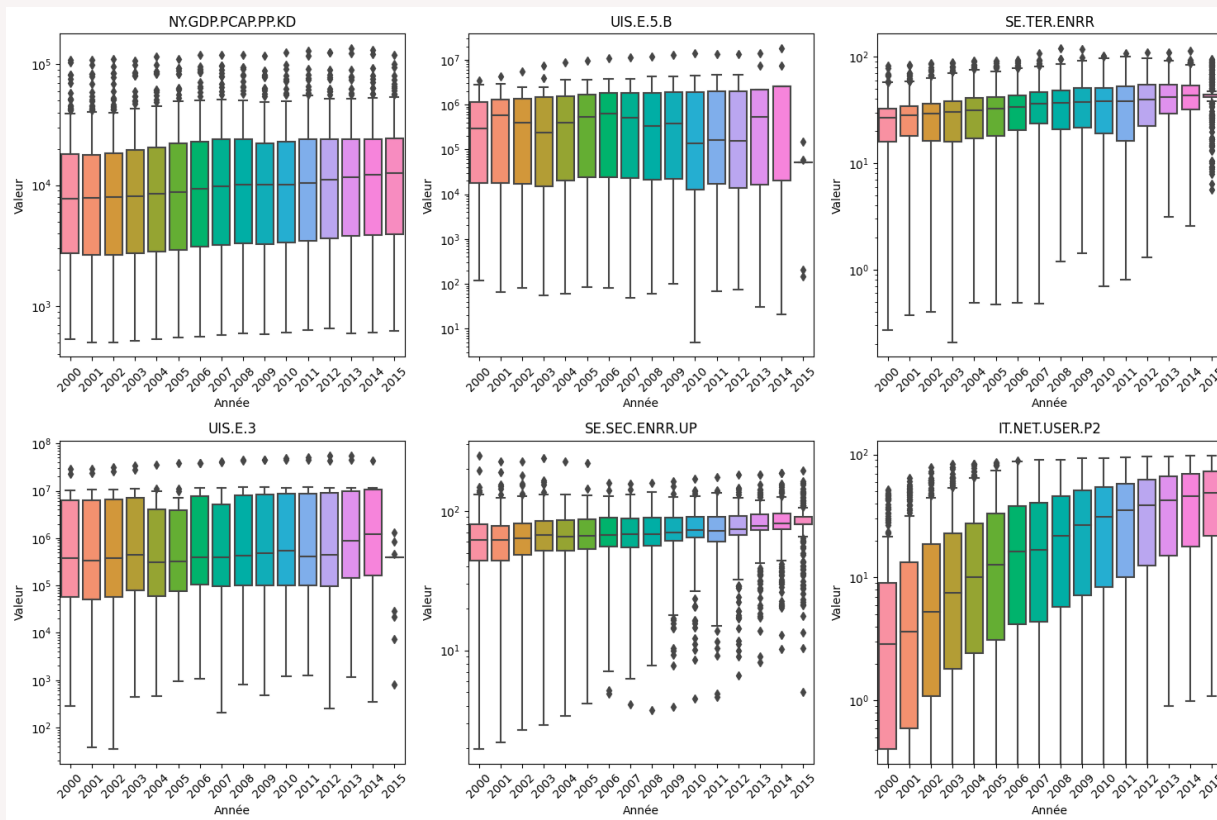
Avant :

```
... Country Name Country Code \
1206 Arab World ARB
1214 Arab World ARB
1246 Arab World ARB
1339 Arab World ARB
1343 Arab World ARB
1375 Arab World ARB
2533 Arab World ARB
4871 East Asia & Pacific EAS
4879 East Asia & Pacific EAS
4911 East Asia & Pacific EAS
5004 East Asia & Pacific EAS
5008 East Asia & Pacific EAS
5040 East Asia & Pacific EAS
6198 East Asia & Pacific EAS
8536 East Asia & Pacific (excluding high income) EAP
8544 East Asia & Pacific (excluding high income) EAP
8576 East Asia & Pacific (excluding high income) EAP
8669 East Asia & Pacific (excluding high income) EAP
8673 East Asia & Pacific (excluding high income) EAP
8705 East Asia & Pacific (excluding high income) EAP
9863 East Asia & Pacific (excluding high income) EAP
12201 Euro area EMU
12209 Euro area EMU
12241 Euro area EMU
...
884604 5.871750e+00 4.363872e+01 8.433270e+00
884608 3.696362e+01 8.067067e+01 7.991996e+01
884640 1.550000e+01 1.636474e+01 2.274282e+01
885798 1.505451e+07 1.541168e+07 1.577745e+07
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

Après:

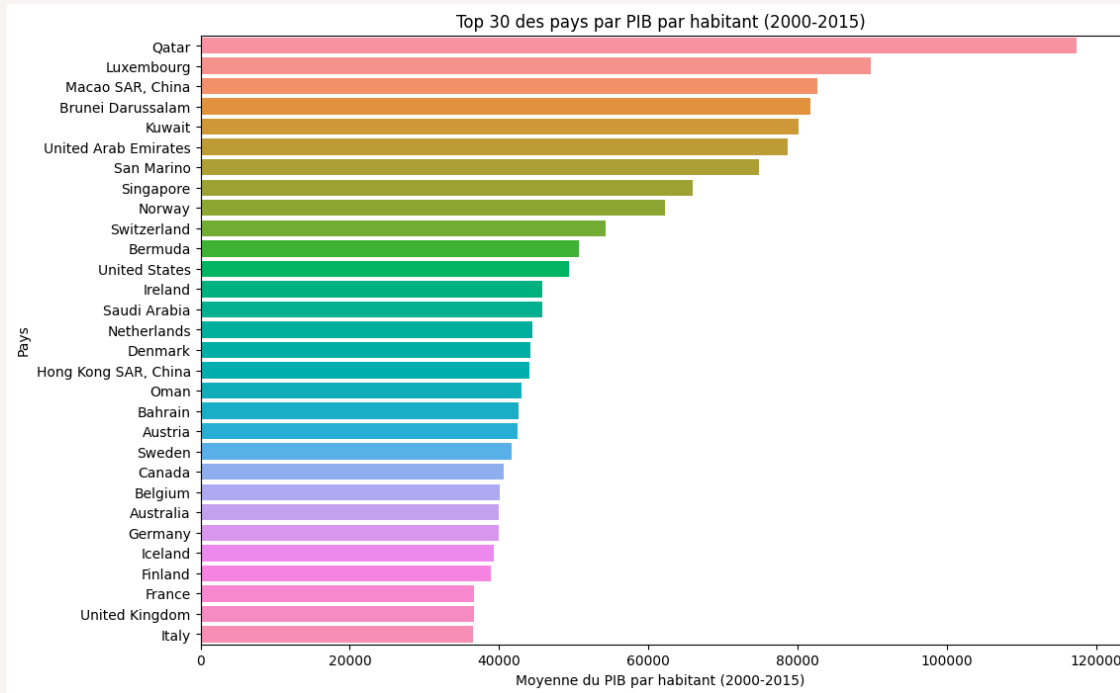
```
Country Name Country Code \
92831 Afghanistan AFG
92839 Afghanistan AFG
92871 Afghanistan AFG
92964 Afghanistan AFG
92968 Afghanistan AFG
93000 Afghanistan AFG
94158 Afghanistan AFG
96496 Albania ALB
96504 Albania ALB
96536 Albania ALB
96629 Albania ALB
96633 Albania ALB
96665 Albania ALB
97823 Albania ALB
100169 Algeria DZA
100201 Algeria DZA
100294 Algeria DZA
100298 Algeria DZA
100330 Algeria DZA
101488 Algeria DZA
105153 American Samoa ASM
107491 Andorra AND
107499 Andorra AND
107660 Andorra AND
...
884604 5.871750e+00 4.363872e+01 8.433270e+00
884608 3.696362e+01 8.067067e+01 7.991996e+01
884640 1.550000e+01 1.636474e+01 2.274282e+01
885798 1.505451e+07 1.541168e+07 1.577745e+07
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

06. Identification des valeurs aberrantes



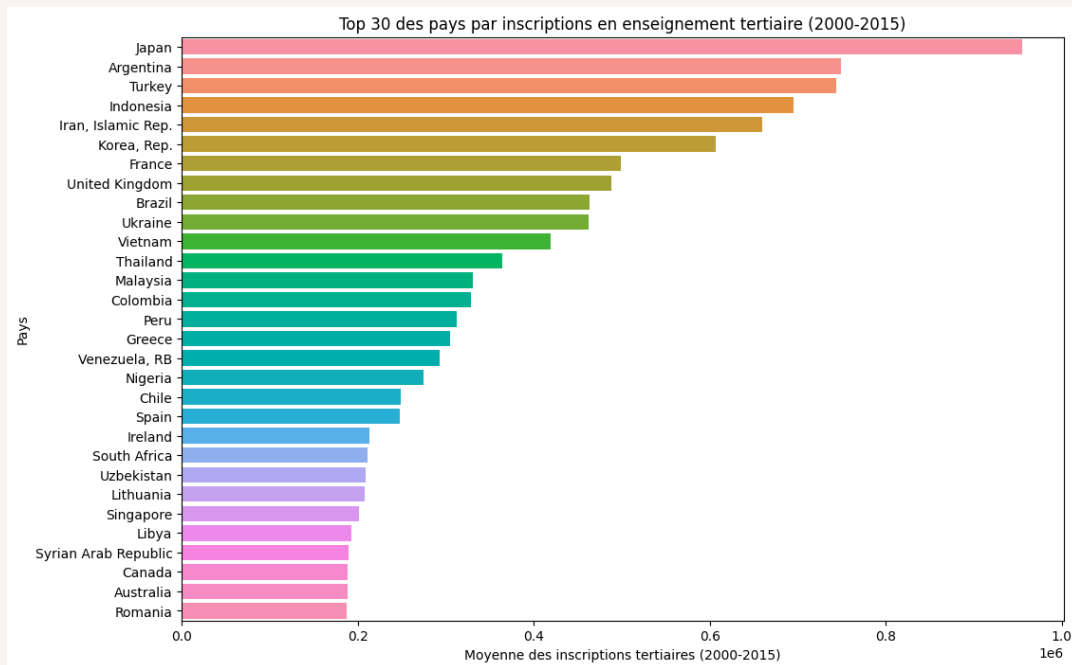
07. Analyse Comparative des Indicateurs de Développement

PIB par Habitant : Mesure de la prospérité économique par personne.



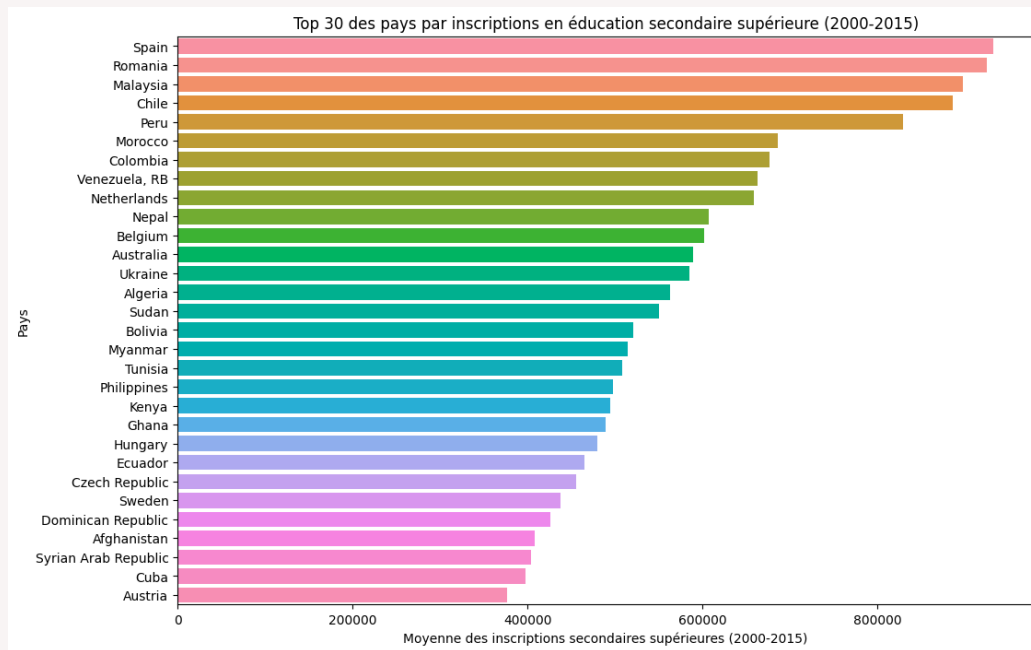
07. Analyse Comparative des Indicateurs de Développement

Inscriptions Universitaires : Nombre d'étudiants inscrits dans l'enseignement supérieur.



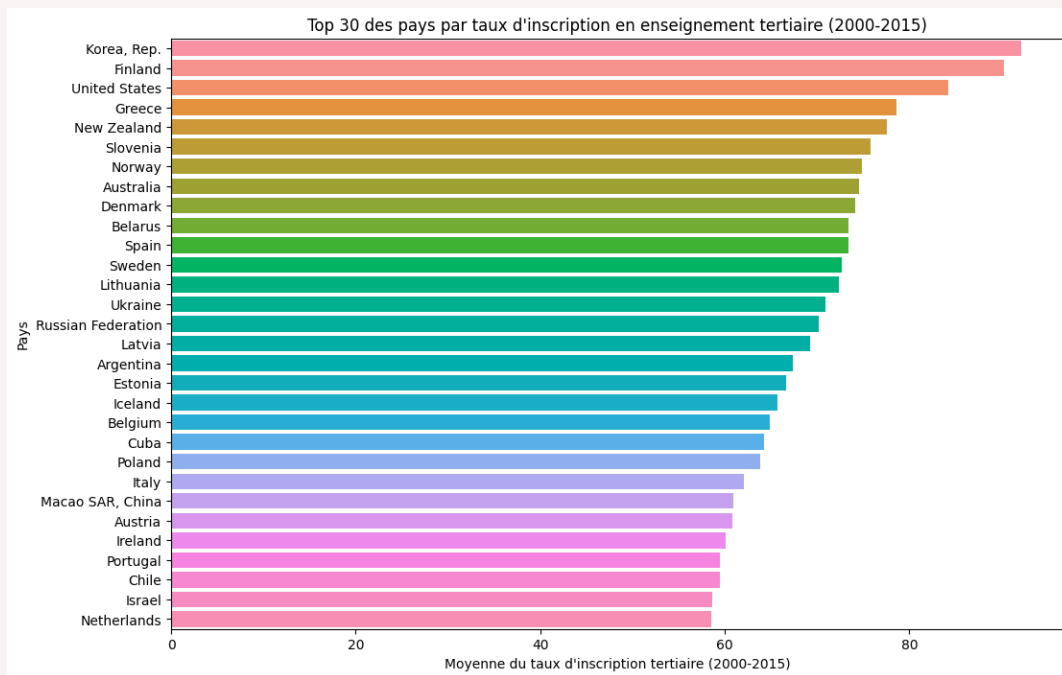
07. Analyse Comparative des Indicateurs de Développement

Éducation Secondaire Supérieure : Nombre d'élèves au lycée.



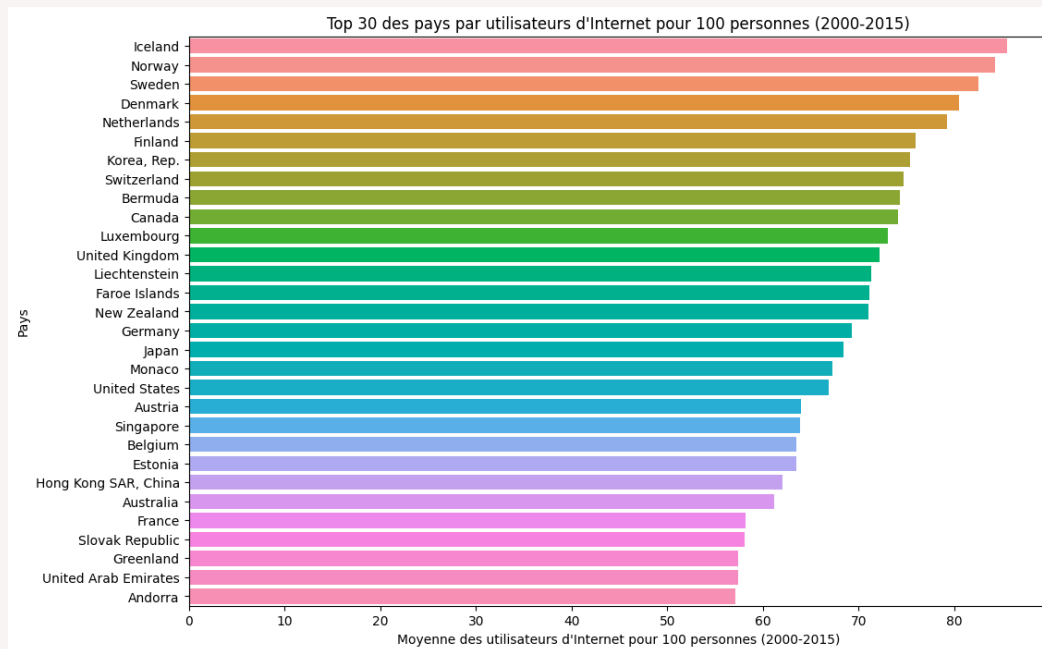
07. Analyse Comparative des Indicateurs de Développement

Accès à l'Université : Part des individus inscrits en enseignement supérieur par rapport à la population totale.



07. Analyse Comparative des Indicateurs de Développement

Accès à Internet: Nombre d'utilisateurs d'Internet pour 100 habitants



08. Conclusion

Classement des pays recommandés à l'exportation:

