



## Projet 2: Analyse des données de systèmes éducatifs

Maodo FALL

OpenClassrooms

Soutenance du projet

29 février 2024

# Sommaire

- 1 Rappel de la problématique et présentation des datasets
  - Rappel de la problématique
  - Présentation du jeu de données
  
- 2 Analyse pré-exploratoire des données
  - Le jeu de données EdStatsCountry.csv
  - Le jeu de données EdStatsData.csv

# Sommaire

- 1 Rappel de la problématique et présentation des datasets
  - Rappel de la problématique
  - Présentation du jeu de données
- 2 Analyse pré-exploratoire des données
  - Le jeu de données EdStatsCountry.csv
  - Le jeu de données EdStatsData.csv

# Academy



# Rappel de la problématique

- Projet d'expansion à l'international de la start-up Academy qui évolue dans l'EdTech.
- Missions :
  - Analyse exploratoire des données
  - Déterminer si les données sur l'éducation de la banque mondiale permettent d'informer le projet d'expansion.



**BANQUE MONDIALE**

# Le jeu de données

Les données sont issues de la banque mondiale. Nous avons 05 fichiers sous format csv :

- EdStatsCountry.csv : Données économiques des pays.
- EdStatsCountry-Series.csv : Source des indicateurs.
- EdStatsData.csv : L'évolution des indicateurs par pays.
- EdStatsFootNote.csv : Années d'estimation des indicateurs.
- EdStatsSeries.csv : Description détaillée des indicateurs.

Nous explorons "EdStatsCountry.csv" et "EdStatsData.csv".

# Sommaire

- 1 Rappel de la problématique et présentation des datasets
  - Rappel de la problématique
  - Présentation du jeu de données
- 2 Analyse pré-exploratoire des données
  - Le jeu de données EdStatsCountry.csv
  - Le jeu de données EdStatsData.csv

# EdStatsCountry

- 1 Nous avons 241 pays et zones géographiques du monde.
- 2 Zéro doublon de pays.



# EdStatsCountry

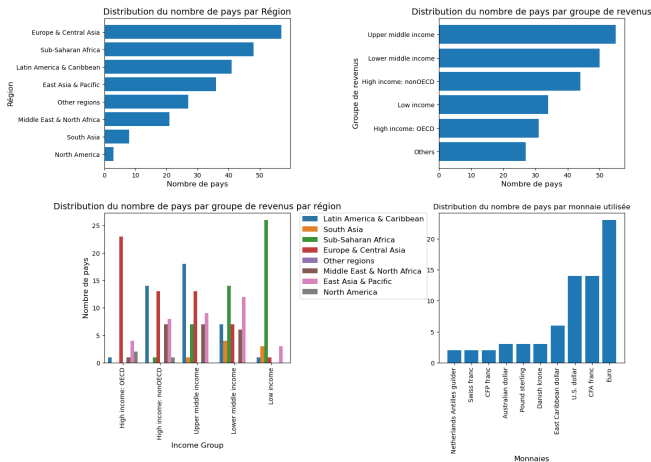


Figure – Distribution des pays par Régions/Revenus/Monnaies



# Infos sur EdStatsData

- EdStatsData contient 886930 lignes et 70 colonnes.
- 3665 indicateurs uniques.
- Pas de doublon d'indicateurs pour un pays donné.

# EdStatsData - Choix des années

- 1998 – 2013 est l'intervalle d'années gardé.

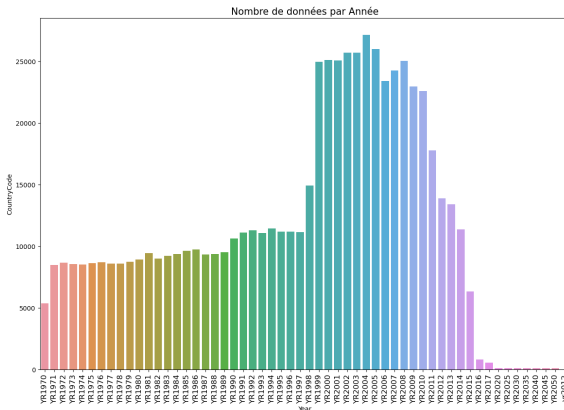
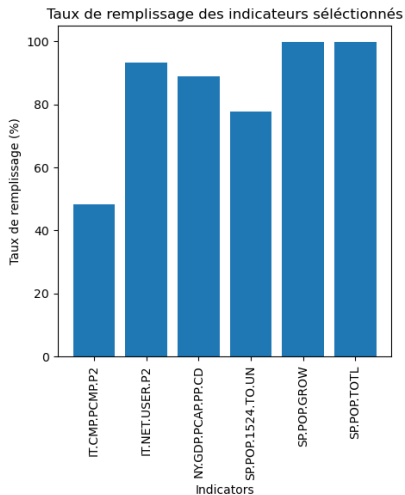


Figure – Taux de remplissage des indicateurs par années

# EdStatsData - Choix des indicateurs

- Après suppression des lignes vides, on a 342811 lignes.
- Critère de choix des indicateurs pertinents :
  - Données sur la population des pays
  - Informations sur l'accès à l'internet
  - Informations sur le niveau de vie des pays.
- Après ce choix des indicateurs, il reste 1335 lignes.

# Indicateurs choisis



# EdStatsData - Critères de sélection des pays

- Transformation du jeu de données :
  - Les pays en lignes uniques et les indicateurs en colonnes.
- Population minimale fixée : 2000000 d'habitants.
- PIB par habitant supérieur ou égal à 80% de France.
- Il n'y a plus que 27 pays.
- Les données sont mis à l'échelle 0 – 1.

# Pays avec un bon potentiel

Country Name		Internet users (%)	PIB per habitat (PPA)	Population 15-24 years	Growth pop. (%)	Computer owners (%)
Country Code						
USA	United States	0.792245	0.390269	1.000000	0.156097	0.906003
IRL	Ireland	0.528024	0.305583	0.004727	0.246063	0.520524
NOR	Norway	0.972709	0.499399	0.003305	0.125620	0.643750
CHE	Switzerland	0.850357	0.340885	0.010878	0.126311	1.000000
CAN	Canada	0.838997	0.227623	0.092069	0.167056	0.605897
DNK	Denmark	0.993732	0.201393	0.004399	0.061105	0.701056
JPN	Japan	0.783803	0.135295	0.316610	0.009934	0.417445
ESP	Spain	0.539907	0.074691	0.111655	0.133144	0.246292
AUS	Australia	0.805307	0.157120	0.056038	0.224489	0.626633
GBR	United Kingdom	0.804785	0.157163	0.174245	0.114372	0.498569

Figure – Quelques pays parmi les plus pertinents qu'on a gardé



# Accès à l'Internet

Taux normalisé d'utilisation d'internet parmi les pays les plus développés

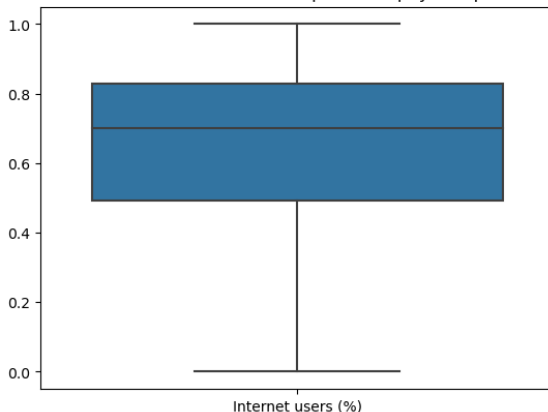


Figure – Taux normalisé d'utilisation d'internet parmi les pays les plus développés

# EdStatsData - Choix définitifs

	Country Name	Internet users (%)	PIB per habitat (PPA)	Population 15-24 years	Growth pop. (%)	Computer owners (%)
Country Code						
AUS	Australia	0.805307	0.157120	0.056038	0.224489	0.626633
CAN	Canada	0.838997	0.227623	0.092069	0.167056	0.605897
CHE	Switzerland	0.850357	0.340885	0.010878	0.126311	1.000000
DEU	Germany	0.818169	0.150079	0.210392	0.000000	0.549315
DNK	Denmark	0.993732	0.201393	0.004399	0.061105	0.701056
FIN	Finland	0.904808	0.151512	0.005153	0.061486	0.516725
GBR	United Kingdom	0.804785	0.157163	0.174245	0.114372	0.498569
JPN	Japan	0.783803	0.135295	0.316610	0.009934	0.417445
NLD	Netherlands	0.973754	0.264282	0.035657	0.079298	0.594693
NOR	Norway	0.972709	0.499399	0.003305	0.125620	0.643750
NZL	New Zealand	0.758431	0.024811	0.003679	0.153278	0.502285
SGP	Singapore	0.701497	0.629599	0.002947	0.372816	0.721717
SWE	Sweden	1.000000	0.200802	0.016270	0.081328	0.820716
USA	United States	0.792245	0.390269	1.000000	0.156097	0.906003

Figure – les pays avec un meilleur potentiel de clients

# MERCI BEAUCOUP !