



# **Projet d'expansion à l'international**

## **Analyse exploratoire**

Yann Pham-Van

# Sommaire

- **Objectifs**
- **Qualité / description du jeu de données**
- **Sélection des informations pertinentes**
- **Nettoyage du jeu de données**
- **Analyse des données**
  - **Scores**
  - **Conclusion intermédiaire**
  - **Perspectives**
- **Conclusion**

# Objectifs

- **Sur le jeu de données fourni par la Banque Mondiale**
  - Valider la qualité
  - Décrire les informations contenues dans le jeu de données
  - Sélectionner les informations pertinentes
  - Déterminer des indicateurs statistiques classiques pour les différentes zones géographiques et pays du monde
- **Informar les décisions d'ouverture vers de nouveaux pays**
  - Identifier les pays avec un fort potentiel de clients sur lesquels opérer en priorité
  - Apprécier l'évolution de ce potentiel sur 1, 2, 3 et 5 ans

# Qualité / description du jeu de données

5 fichiers dont 2 exploitables pour nos besoins, sans doublon

- ***edStatsCountry*** nommant des pays et des agrégations
  - 241 lignes x 32 colonnes
  - 1 ligne = 1 pays ou agrégation (géographique/revenus)
  - colonnes = variables diverses (monnaie, région,...)
  - manquants variables jusqu'à 87%
- ***edStatsData*** regroupant le cœur des données
  - 886930 lignes x 70 colonnes
  - 1 ligne = 1 couple Pays/Indicateur
  - 1 colonne (hors en-tête) = valeur d'une année
  - manquants nombreux : entre 72% et 100%

# Sélection des informations pertinentes

Les indicateurs à sélectionner :

- description sur le site de la Banque Mondiale
- en lien avec la problématique d'*academy*
  - **cibler** les niveaux **lycée** et **université**  
12 indicateurs directs ou indirects
  - **pouvoir d'achat** des cibles  
1 indicateur : *PIB par habitant*
  - accès technique aux contenus de formation **en ligne**  
1 indicateur : *Internautes (pour 100 personnes)*
  - pérennisation par la **croissance** des cibles  
1 indicateur : *Croissance démographique (% annuel)*

# Nettoyage du jeu de données

Le processus de nettoyage est mis en œuvre :

- réduction à **3630 lignes** de couples pays/indicateur
- réduction à **7 années** d'observation (2010 à 2016)
- passage de 15 à **6 indicateurs** adressant la problématique
- affectation d'une typologie :
  - **217 pays** - taux de manquants de **1% à 13%**
  - **7 continents** - remplissage à **100%**
- suppression du **premier quartile** des pays en terme de **PIB/habitants** et ceux avec **moins de 1M d'habitants**
- abandon du Kosovo avec trop peu d'indicateurs valables
- étude complémentaire pour 4 pays avec bon potentiel

# Nettoyage du jeu de données

Corrélation entre population totale et cibles des pays ?



# Nettoyage du jeu de données

Finalisation du processus de nettoyage :

- calcul du coefficient de corrélation sur les populations
- détermination des droites de régression
- calcul des dernières valeurs manquantes des cibles pour
  - Bosnie-Herzégovine
  - Oman
  - Singapour
  - Emirats arabes unis

**=> jeu de données nettoyé et complet à 100%**



# Analyse des données - Scores

Premier score d'attractivité basé sur la  
**moyenne des indicateurs**

**Manque de cohérence entre  
pays et continents**

## Continent SCORE\_SYNT

North America	0.71
East Asia & Pacific	0.62
Europe & Central Asia	0.60
South Asia	0.48
Latin America & Caribbean	0.43
Middle East & North Africa	0.38
Sub-Saharan Africa	0.27

## Top 15 Pays SCORE\_SYNT

Qatar	0.65
Norway	0.51
Singapore	0.50
United Arab Emirates	0.49
Switzerland	0.46
United States	0.46
Kuwait	0.46
Denmark	0.46
Ireland	0.45
Bahrain	0.45
India	0.45
Hong Kong SAR, China	0.45
Sweden	0.44
United Kingdom	0.44
Netherlands	0.44

# Analyse des données - Scores

Second score d'attractivité basé sur la  
**multiplication des indicateurs**

**Correction du défaut mais  
d'autres surviennent**

## Continent SCORE\_SYNT

East Asia & Pacific	5.4
Europe & Central Asia	6.0
North America	7.5
Latin America & Caribbean	17.2
Middle East & North Africa	22.9
South Asia	27.2
Sub-Saharan Africa	114.2

## Top 15 Pays SCORE\_SYNT

United States	17.8
China	22.7
India	64.3
Japan	74.5
Germany	94.1
United Kingdom	96.2
Brazil	112.2
France	129.1
Russian Federation	130.0
Mexico	135.0
Korea, Rep.	144.7
Turkey	154.9
Saudi Arabia	171.3
Canada	181.2
Italy	208.2

# Analyse des données - Scores

Score basé sur la **multiplication des indicateurs** avec **accentuation du PIB**

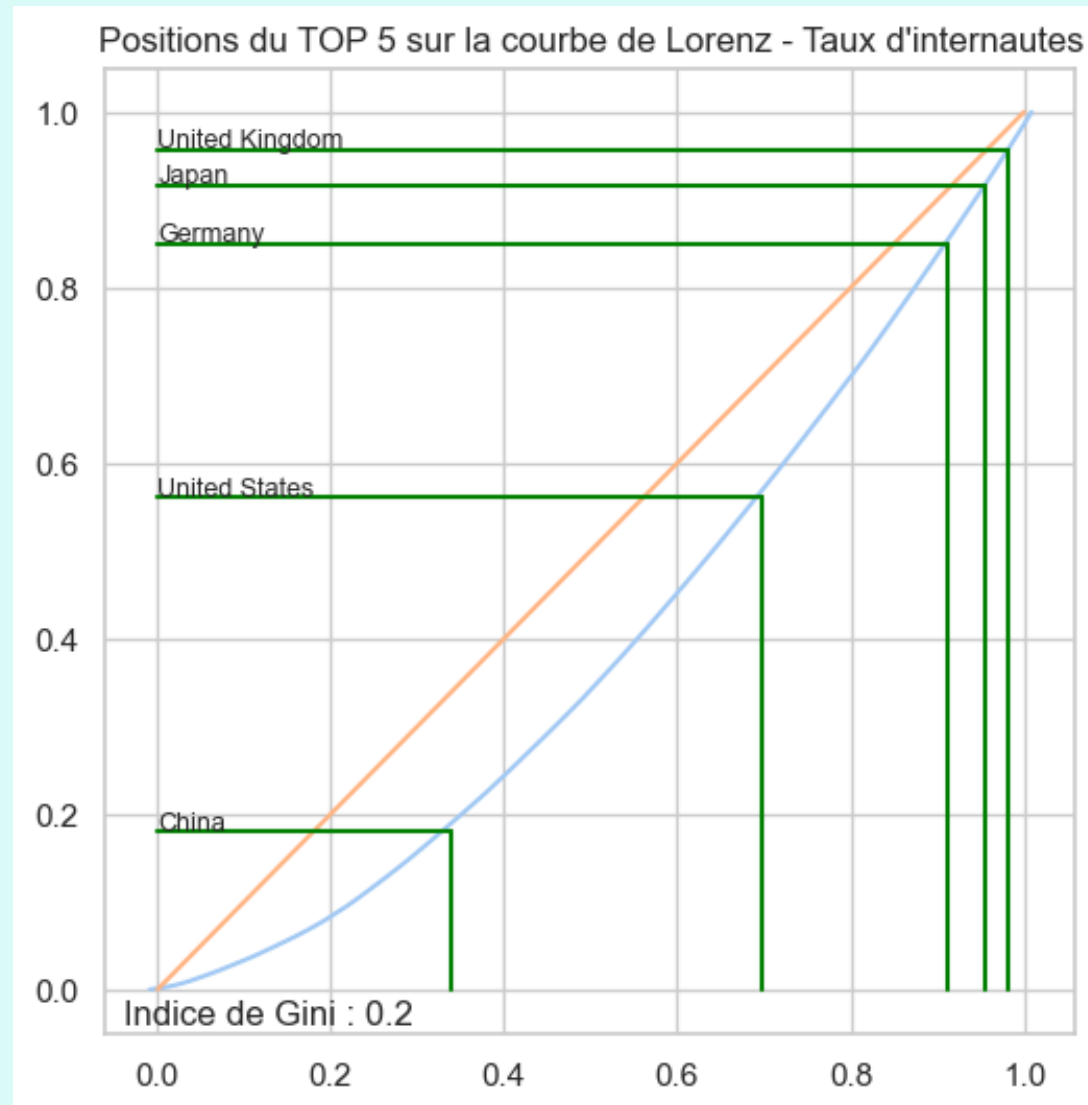
## Classements en cohérence

Continent	SCORE_SYNT
North America	7.5
Europe & Central Asia	10.9
East Asia & Pacific	18.0
Latin America & Caribbean	62.8
Middle East & North Africa	66.3
South Asia	253.7
Sub-Saharan Africa	1730.1

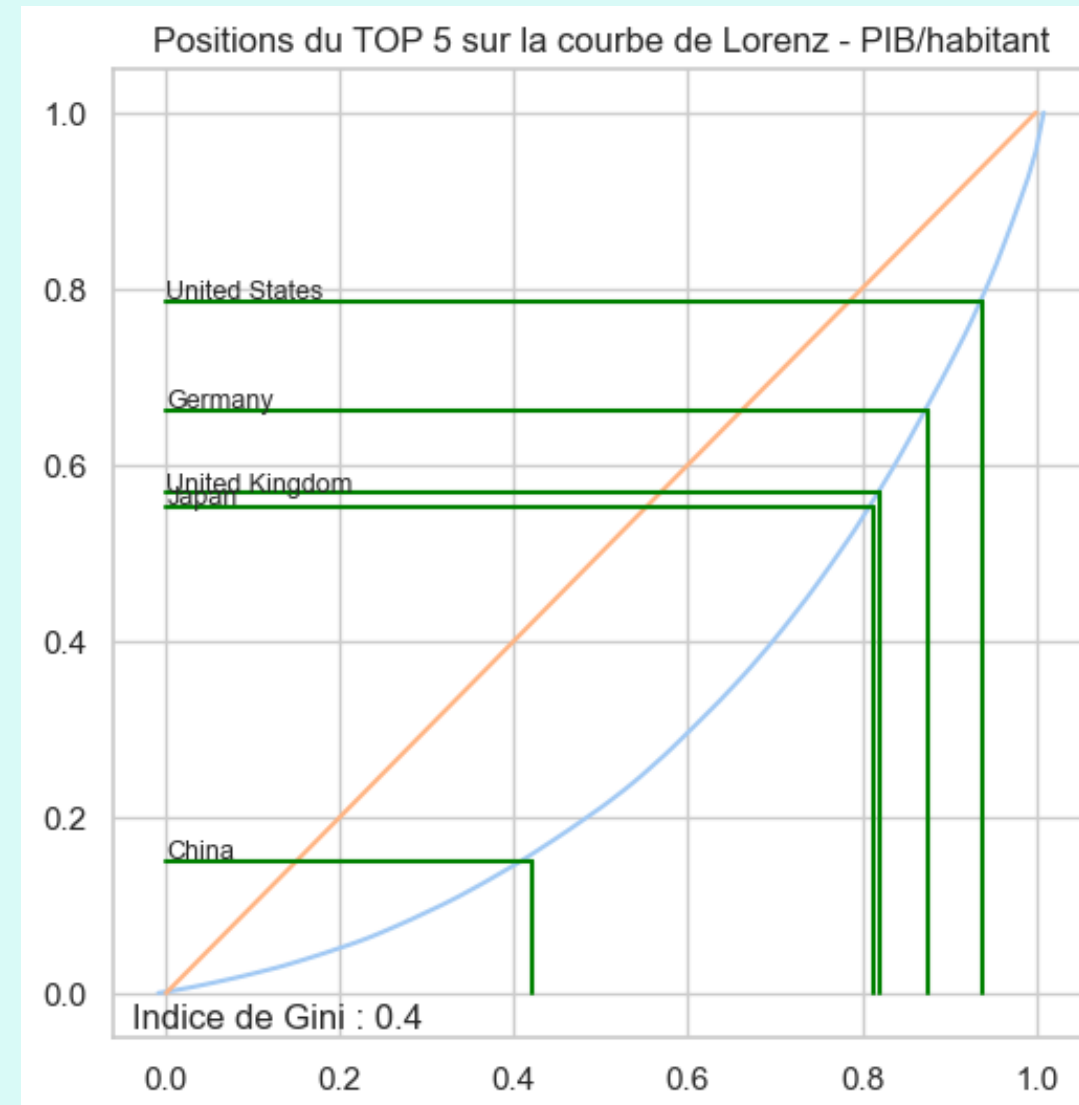
### Top 15 Pays SCORE\_SYNT

United States	39.5
China	187.0
Japan	230.4
Germany	251.3
United Kingdom	290.0
France	401.1
Saudi Arabia	401.4
Korea, Rep.	489.0
Canada	497.3
Russian Federation	639.8
Italy	708.9
Turkey	771.1
Australia	785.7
United Arab Emirates	839.6
Netherlands	919.3

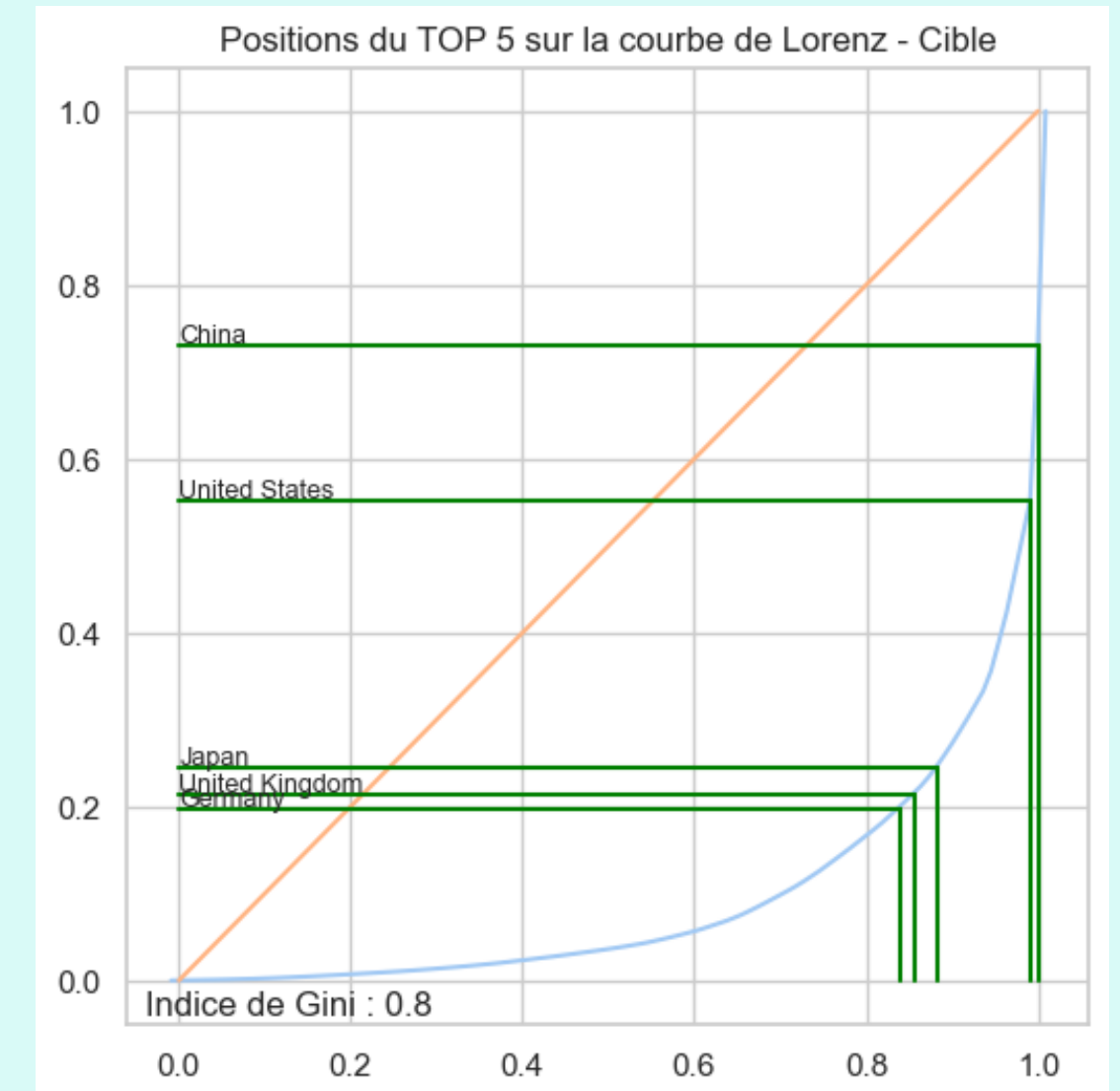
# Conclusions sur les classements



Moyenne : 61.1 %  
Médiane : 66.0 %  
Ecart-type : 23.2 %



Moyenne : 23.9 k\$  
Médiane : 17.3 k\$  
Ecart-type : 19.0 k\$



Moyenne : 7.2 M  
Médiane : 1.0 M  
Ecart-type : 25.0 M

Pays	Taux d'internautes	PIB/habitant	Cible
United States	76.1 %	53.3 k\$	34.9 M
China	53.2 %	14.3 k\$	145.0 M
Japan	92.0 %	38.2 k\$	9.6 M
Germany	89.6 %	44.2 k\$	6.7 M
United Kingdom	94.7 %	39.2 k\$	7.0 M

- Pays du **Top 5** globalement bien classés en scores **individuels**
- Rattrapage de la **Chine** grâce au score de la population **cible**
- Utilité de l'**interpolation** sur pays avec manquants
- **Taux d'internautes** à **relativiser** en cas d'étude approfondie (critère peu discriminatoire, prise en compte des mobiles)

# Analyse des données - Perspectives

Influence de la croissance démographique - Continents

Année_N	Année_N+1	Année_N+2	Année_N+3	Année_N+5
North America	North America	North America	North America	North America
Europe & Central Asia	Europe & Central Asia	Europe & Central Asia	Europe & Central Asia	Europe & Central Asia
East Asia & Pacific	East Asia & Pacific	East Asia & Pacific	East Asia & Pacific	East Asia & Pacific
Latin America & Caribbean	Latin America & Caribbean	Latin America & Caribbean	Latin America & Caribbean	Latin America & Caribbean
Middle East & North Africa	Middle East & North Africa	Middle East & North Africa	Middle East & North Africa	Middle East & North Africa
South Asia	South Asia	South Asia	South Asia	South Asia
Sub-Saharan Africa	Sub-Saharan Africa	Sub-Saharan Africa	Sub-Saharan Africa	Sub-Saharan Africa

# Analyse des données - Perspectives

Influence de la croissance démographique - Pays

Année_N	Année_N+1	Année_N+2	Année_N+3	Année_N+5
United States	United States	United States	United States	United States
China	China	China	China	China
Japan	Japan	Japan	Japan	Japan
Germany	Germany	Germany	Germany	Germany
United Kingdom	United Kingdom	United Kingdom	United Kingdom	United Kingdom
France	Saudi Arabia	Saudi Arabia	Saudi Arabia	Saudi Arabia
Saudi Arabia	France	France	France	France
Korea, Rep.	Korea, Rep.	Korea, Rep.	Canada	Canada
Canada	Canada	Canada	Korea, Rep.	Korea, Rep.
Russian Federation	Russian Federation	Russian Federation	Russian Federation	Russian Federation
Italy	Italy	Italy	Italy	Turkey
Turkey	Turkey	Turkey	Turkey	Italy
Australia	Australia	Australia	Australia	Australia
United Arab Emirates	United Arab Emirates	United Arab Emirates	United Arab Emirates	Qatar
Netherlands	Qatar	Qatar	Qatar	United Arab Emirates

# Conclusions

- Shortliste de **5 pays**
  - United States
  - China
  - Japan
  - Germany
  - United Kingdom
- **Choix** et investissements consécutifs **pérennes** à court et moyen termes
- 2 Pays avec une **langue commune** pour **optimiser** la diffusion de notre offre

Merci de votre attention !  
Avez-vous des questions ?