

# Plan de développement Academy

Laurent Cagniard

# Énoncé de mission

Première mission d'analyse exploratoire pour déterminer si les données sur l'éducation de la banque mondiale permettent d'informer le projet d'expansion à l'international de l'entreprise.



# Jeu de données (1/2)

**THE WORLD BANK** | Data Catalog  
IBRD · IDA

HOME DATA COLLECTIONS GETTING STARTED FAQs LOGIN

Home / Search Results / Details

## Education Statistics

Metadata last updated on - Jul 16, 2021 [Compare](#) Version: 2 (Latest) ▾

The World Bank EdStats All Indicator Query holds over 4,000 internationally comparable indicators that describe education access, progression, completion, literacy, teachers, population, and expenditures. The indicators cover the education cycle from pre-primary to vocational and tertiary education. The query also holds learning outcome data from international and regional learning assessments (e.g. PISA, TIMSS, PIRLS), equity data from household surveys, and projection/attainment data to 2050. For further information, please visit the EdStats website.

[Overview](#) [Indicators](#) [Citations](#)

[Databank](#) [DATABANK](#) [METADATA](#)

[API](#) [HTML](#) [METADATA](#)

[CSV file](#)

### Chargement des fichiers sous forme de dataframe

```
Entrée [6]: statsData = pd.read_csv("EdStatsData.csv")
Entrée [7]: statsData.head()
Out[7]:
```

	Country Name	Country Code	Indicator Name	Indicator Code	1970	1971	1972	1973	1974	1975	...	2060	2065	2070	2080	2085	2090
0	Arab World	ARB	Adjusted net enrollment rate, lower secondary	UIS.NERA.2	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
1	Arab World	ARB	Adjusted net enrollment 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 enrollment 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 enrollment rate, lower secondary	UIS.NERA.2.M	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN

Pas de doublons identifiés

Les données chiffrées (float) sont exploitables

3665 indicateurs individuels identifiés

## Jeu de données (2/2)

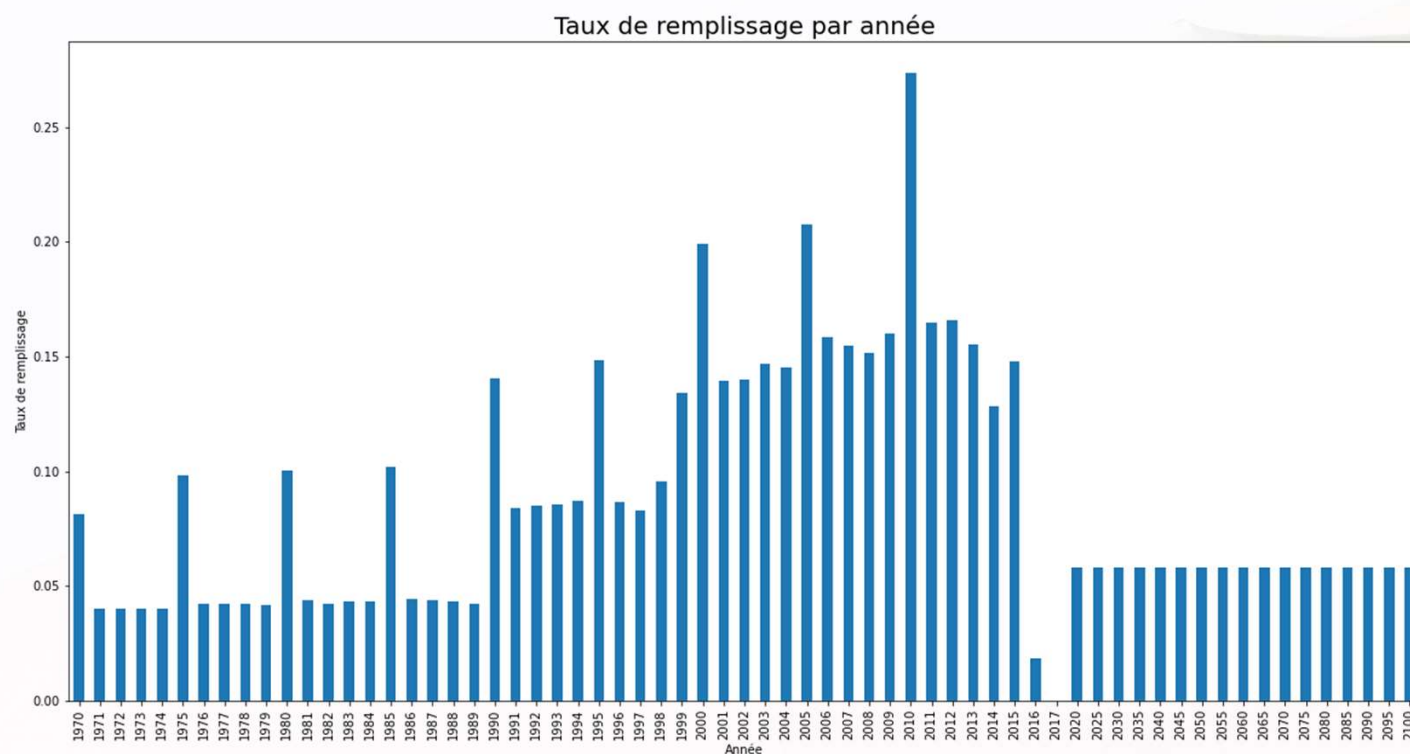
	Data	Country	CountrySeries	Series	FootNote
<b>lignes</b>	886 930	241	613	3 665	643 638
<b>colonnes</b>	70	32	4	21	5

Le fichier csv "statsData" servira de base de travail pour réaliser les analyses demandées  
(70 colonnes pour 886.930 lignes)

Les autres fichiers fournis ne seront pas exploités



# Analyse de la qualité du jeu de données (1/5)



Globalement, on voit que les taux de remplissage sont relativement bas, ce qui nécessitera une analyse plus approfondie pour déterminer les indicateurs les plus pertinents..

## Analyse de la qualité du jeu de données (2/5)



On peut remarquer à vue d'oeil une périodicité de 5 ans sur certains indicateurs depuis les années 1970 ainsi que globalement un niveau d'informations inférieur avant les années 90. La dernière année la plus exploitable semble être 2015.

La moyenne du taux de remplissage sur la période couverte est de : 8,8% (écart-type 5,4%)

La moyenne du taux de remplissage sur les années 2000 est de : 16,5% (écart-type 3,5%)

Les années 2020 à 2100 correspondant à un indicateur prospectif, celles-ci ne seront pas intégrées dans nos analyses ainsi que l'année 2017.

Nous nous concentrerons donc sur les années 2000 jusqu'à 2016

## Analyse de la qualité du jeu de données (3/5)

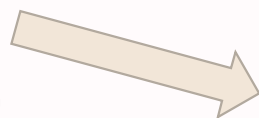
Tout d'abord, on va fixer un critère éliminatoire sur l'indicateur « Population, total » pour éliminer de l'analyse tous les pays dont la population est inférieure à 5 millions d'habitants



-94 pays

Ensuite, on analyse le taux de remplissage par indicateur en 2010 et 2015 (nb de pays avec des données valides / nb de pays) puis on laisse de côté les indicateurs pour lesquels le taux est inférieur à 50 %

3665



1326

# Analyse de la qualité du jeu de données (4/5)

Sur cette base, on déterminera une liste d'indicateurs permettant de répondre à plusieurs catégories :

## Démographie :

Population, total; Gross enrolment ratio, primary, both sexes (%) (idem pour secondary and tertiary)

## Economie :

GDP per capita (current US\$), Government expenditure on education as % of GDP (%), Internet users (per 100 people)

## Emploi :

Labor force, total ; Unemployment, total (% of total labor force)



## Analyse de la qualité du jeu de données (5/5)

Indicator Name	nb total de pays	nb de pays rempli	taux de remplissage
Population, total	123	121	98,4%
Unemployment, total (% of total labor force)	123	117	95,1%
Labor force, total	123	117	95,1%
Internet users (per 100 people)	123	116	94,3%
GDP per capita (current US\$)	123	115	93,5%
Gross enrolment ratio, primary, both sexes (%)	123	100	81,3%
Gross enrolment ratio, secondary, both sexes (%)	123	90	73,2%
Gross enrolment ratio, tertiary, both sexes (%)	123	86	69,9%
Government expenditure on education as % of GDP (%)	123	83	67,5%

Même si les taux de remplissage sont moins élevés pour les 4 derniers indicateurs, il reste intéressant de les conserver pour avoir une analyse plus fine du marché potentiel

# Détermination du scoring (1/3)

On va donner une **valeur de 0 à 100** pour chaque indicateur retenu afin de déterminer un scoring par catégorie puis global

Exemple :

$$\text{Scoring "Population"} = \text{Valeur} / \text{Valeur Max} * 100$$

Remarque : après essai, la méthode de standardisation  $(x - \text{moy}) / \text{std}$  a rendu des résultats identiques sur le top 10 pour ce scoring.

# Détermination du scoring (2/3)

## DEMOGRAPHIE :

Scoring « Démographie » = Moyenne non pondérée des Scorings « Population », « Taux de scolarisation Primaire », « Taux de scolarisation Secondaire », « Taux de scolarisation Tertiaire »

## ECONOMIE :

Scoring « Economie » = Moyenne pondérée des Scorings « GDP per capita », « Dépenses en éducation » et « utilisateurs Internet »

Nous avons introduit une pondération pour donner plus de poids au revenu par habitant (\*3), puis aux dépenses en éducation (\*2) et enfin de l'utilisation d'internet (\*1)

## EMPLOI :

Scoring « Emploi » = Moyenne non pondérée des Scorings « Population active » et « chômage »

# Détermination du scoring (3/3)

## SCORING GLOBAL :

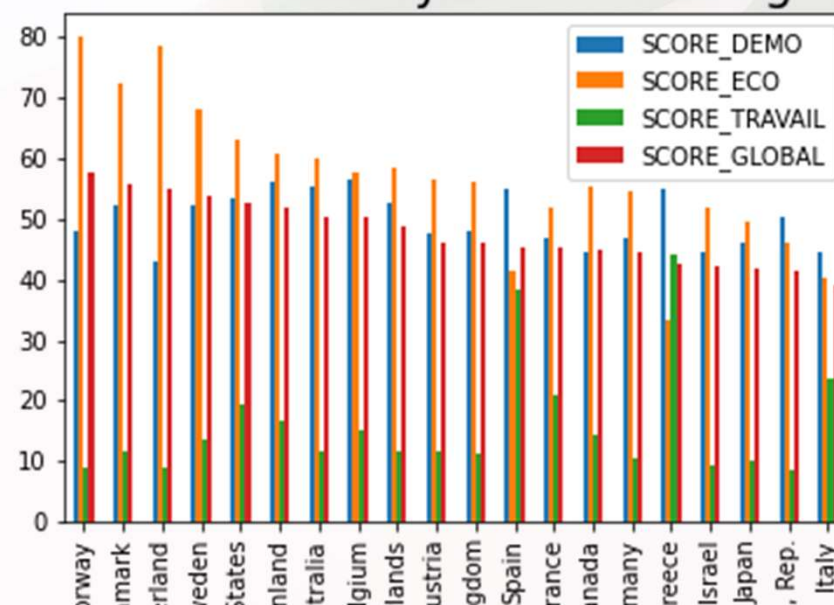
Scoring Global = Moyenne pondérée des Scorings « Démographie », « Economie » et « Emploi »

Nous avons introduit une pondération pour donner plus de poids au scoring Economie (\*3), puis Démographie (\*2) et enfin au scoring Emploi (\*1)

## Analyse du scoring (1/3)

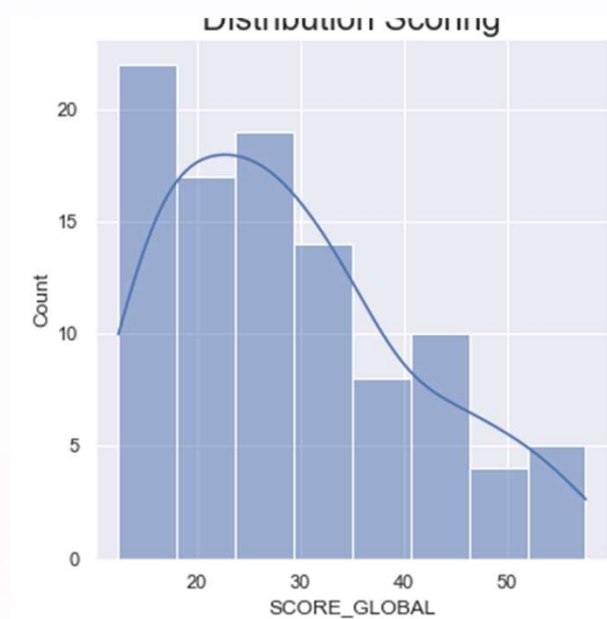
Country Name	SCORE_DEMO	SCORE_ECO	SCORE_TRAVAIL	SCORE_GLOBAL
Norway	48	80	9	58
Denmark	52	73	12	56
Switzerland	43	79	9	55
Sweden	52	68	13	54
United States	54	63	20	53
Finland	56	61	17	52
Australia	55	60	11	50
Belgium	57	58	15	50
Netherlands	53	58	12	49
Austria	48	57	12	46
United Kingdom	48	56	11	46
Spain	55	42	38	45
France	47	52	21	45
Canada	44	55	14	45
Germany	47	55	10	45
Greece	55	33	44	42
Israel	45	52	9	42
Japan	46	49	10	42
Korea, Rep.	50	46	9	41
Italy	45	40	24	39

Tableau de synthèse Scoring





## Analyse du scoring (2/3)



	Moyenne	Médiane	Ecart-type
Scoring Global	29,0	27,0	11,7

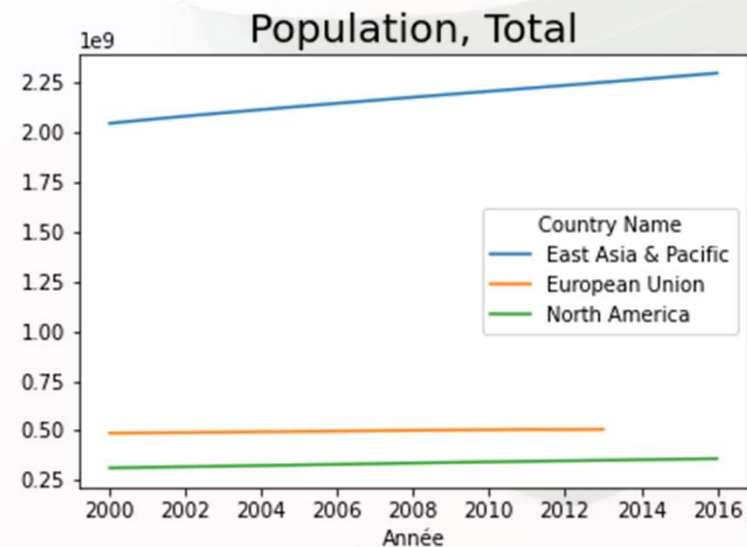
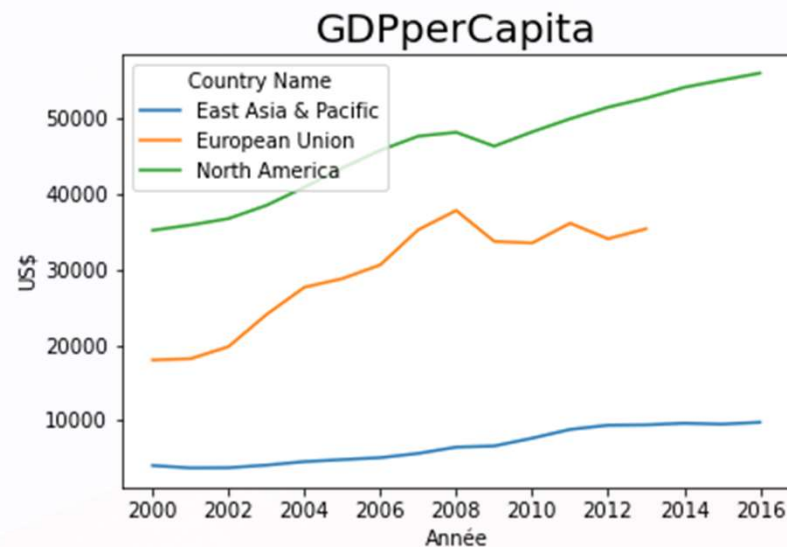
Malgré un taux de remplissage, au départ, inférieur à 20% en moyenne, l'analyse exploratoire a permis d'identifier un Top20 (score > 39, env. moyenne + 1 écart-type) de pays suffisamment robuste

## Analyse du scoring (3/3)



On remarque que notre Top 20 se situe essentiellement sur la zone Europe, Amérique du Nord, Asie de l'Est (Japon, Corée du Sud) + Australie

## Evolution de 2 indicateurs clés par zone géo



Les courbes sont plutôt en croissance, malgré une légère stagnation en Union Européenne

Les conditions semblent ainsi donc favorables à notre développement à l'international

## Pistes futures

Sur la base du top 20 identifié, il est envisageable d'affiner l'analyse en intégrant des données liées aux indicateurs de niveau scolaire de type « test PISA », ou des ratios d'encadrement dans l'éducation : des données que nous avons exclues dans un 1<sup>er</sup> temps par manque de pays concernés afin de se concentrer potentiellement sur moins de pays (Top 10 voir top 5)

Nous pourrions également recroiser avec des informations liées à la langue (notamment la maîtrise de la langue anglaise). En effet, la traduction en langue étrangère (ex. japonais ou coréen) pour un marché spécifique pourrait ne pas s'avérer suffisamment rentable.

De même, pour les analyses ultérieures, nous pourrions utiliser des benchmarks pour examiner l'état de la concurrence et de maturité des marchés



# Merci

Laurent Cagniard