

OPENCLASSROOM

Parcours data scientist en alternance

10 février 2022

Maxime Dupouy

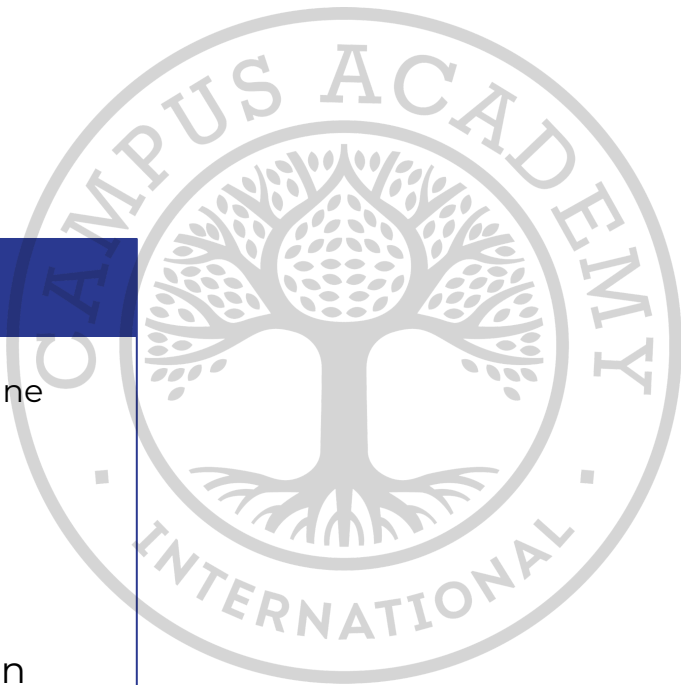
Livrable 2

Objectifs

Academy : start up proposant des contenus de formation en ligne pour un public de niveau lycée et université.

Projet d'expansion à l'international :

- 1) Quels sont les pays avec un fort potentiel de clients pour nos services ?
- 2) Pour chacun de ces pays, quelle sera l'évolution de ce potentiel de clients ?
- 3) Dans quels pays l'entreprise doit-elle opérer en priorité ?



Pré-analyse

La source

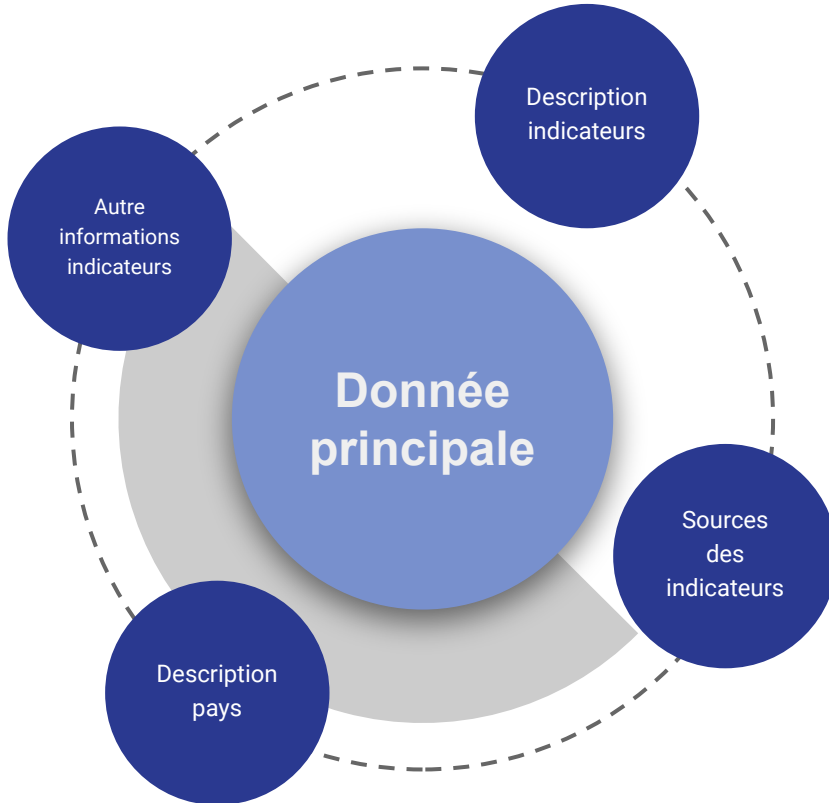


THE WORLD BANK

La banque de données mondiales est une institution financière internationale qui accorde des prêts à effet de levier à des pays en développement pour des projets d'investissements.

Leur objectif : mettre fin à l'extrême pauvreté et promouvoir une prospérité partagée.

Les données



5 bases de données liées

A première vue :

- Beaucoup de données manquantes
- Beaucoup d'indicateurs

Donnée principale

- **886 930 lignes :**
 - 242 pays/régions
 - 3665 indicateurs
- **69 colonnes :**
 - 65 années de 1970 à 2100

86% de valeurs manquantes

Pas de doublons dans ce dataset

	Country Name	Indicator Name	Indicator Code	1970	1971	1972	1973	1974	1975	1976	...	2055	2060	2065	2070	2075	2080	2085	2090	2095	2100
Country Code																					
ABW	3665	3665	3665	49	49	49	49	49	49	49	...	308	308	308	308	308	308	308	308	308	308
AFG	3665	3665	3665	523	139	158	236	351	537	164	...	0	0	0	0	0	0	0	0	0	0
AGO	3665	3665	3665	49	95	114	65	55	49	49	...	0	0	0	0	0	0	0	0	0	0
ALB	3665	3665	3665	409	78	74	50	50	410	105	...	308	308	308	308	308	308	308	308	308	308
AND	3665	3665	3665	41	63	63	63	63	99	63	...	0	0	0	0	0	0	0	0	0	0
...
XXC	3665	3665	3665	2	2	2	2	2	2	2	...	0	0	0	0	0	0	0	0	0	0
YEM	3665	3665	3665	410	50	50	50	50	410	50	...	0	0	0	0	0	0	0	0	0	0
ZAF	3665	3665	3665	471	55	55	59	56	419	56	...	308	308	308	308	308	308	308	308	308	308
ZMB	3665	3665	3665	518	117	259	216	270	591	310	...	308	308	308	308	308	308	308	308	308	308
ZWE	3665	3665	3665	481	127	127	135	135	491	135	...	308	308	308	308	308	308	308	308	308	308

Sélection des indicateurs :

- en fonction des données manquantes/présentes
- pertinence de l'indicateur pour les objectifs
- (source indicateur)

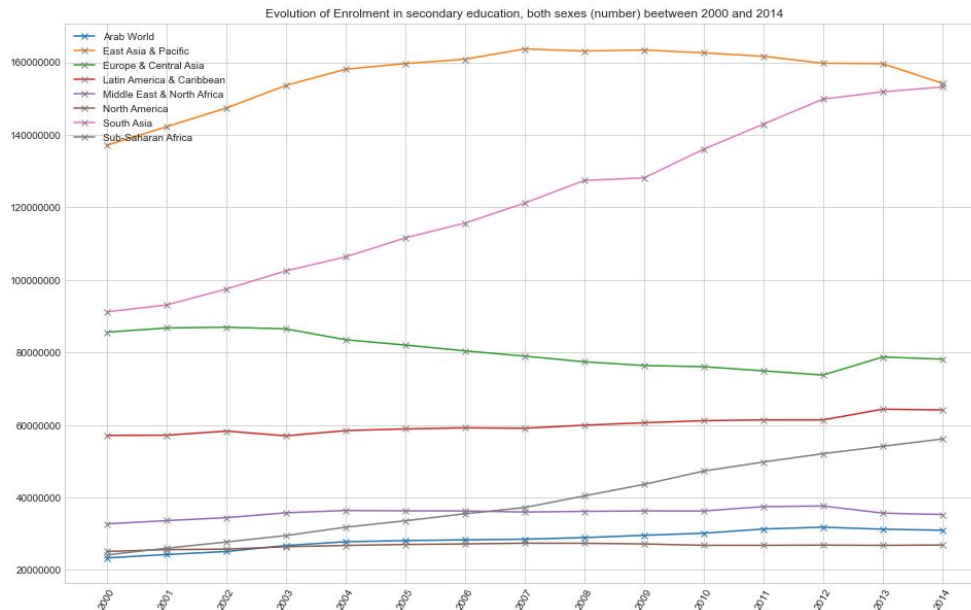
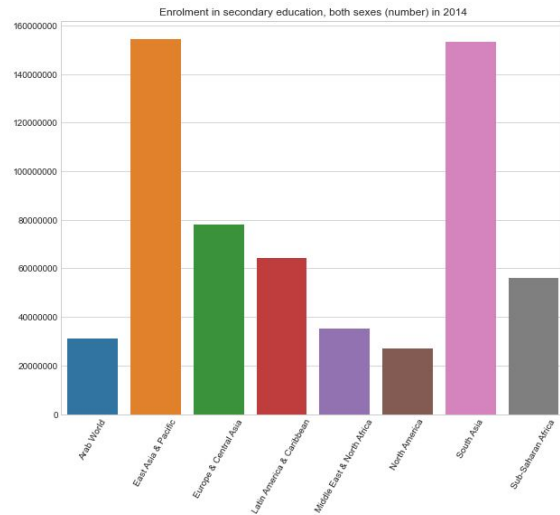
Sélection des années

Analyse et résultats

Indicateurs sélectionnés

01	Enrolment in secondary education, both sexes (number)	• Source : UNESCO Institute for Statistics
02	Population of the official age for upper secondary education, both sexes (number)	• Source : UNESCO Institute for Statistics
02	Enrolment in tertiary education, all programmes, both sexes (number)	• Source : UNSECO Institute for Statistics
03	Population of the official age for tertiary education, both sexes (number)	• Source : UNSECO Institute for Statistics
03	Internet users (per 100 people)	• Source : International Telecommunication Union, World Telecommunication/ICT development Report and database

Population au lycée en 2014

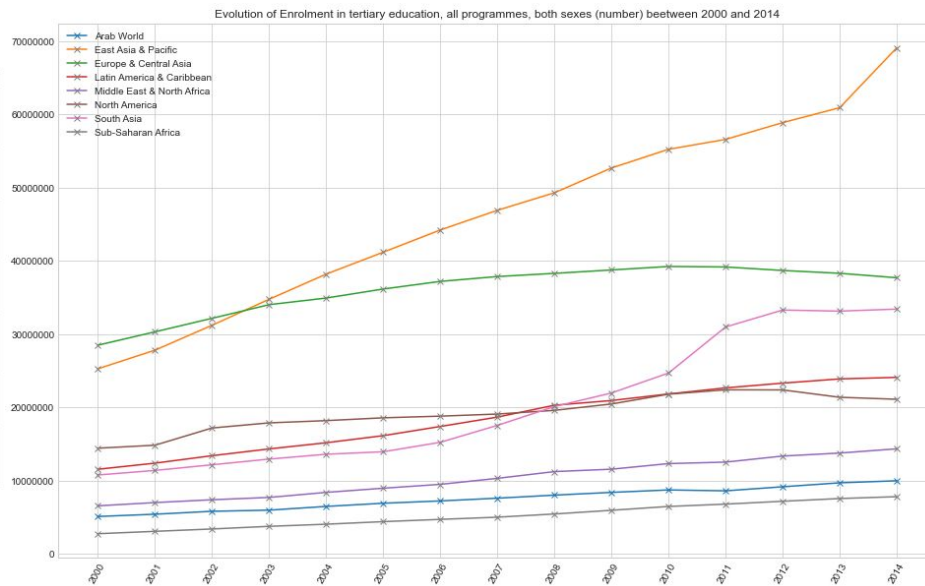
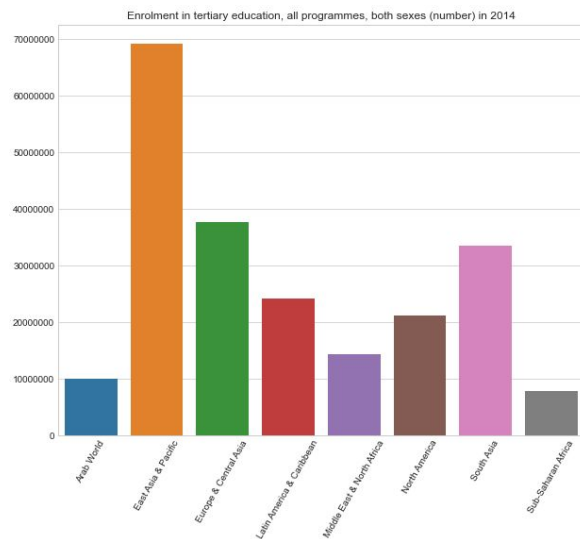


Régions du monde

Asie du Sud affiche le meilleur nombre de population dans le secondaire et la meilleure croissance.

Asie de l'Est et Pacifique est plutôt sur une légère décroissance

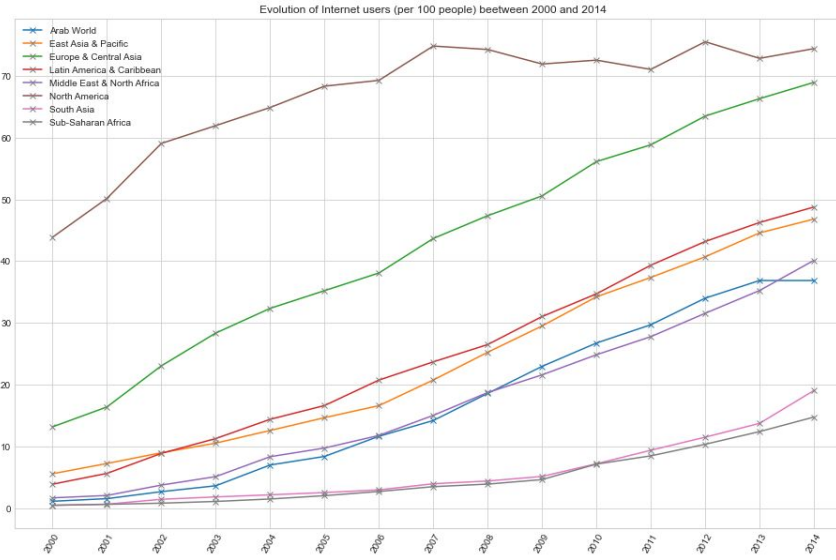
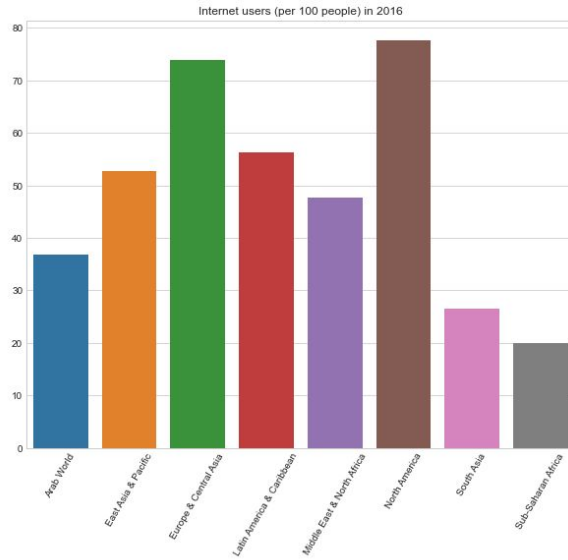
Population en étude universitaire en 2014



Régions du monde

Asie de l'Est et Pacifique domine largement le terrain universitaire par la population la plus importante et une croissance sur les dernières années.

Population ayant accès à internet (%) en 2014



Régions du monde

L'Amérique du Nord et l'Europe & Asie du centre présentent l'accès à internet le plus important.

Conclusion :

- **L'Amérique du Nord + Europe & Asie du centre** (internet)
- Mais peut s'inverser dans les années à venir

Analyse des pays

Sélection des
indicateurs

Déjà réalisée

Pré-sélection des
pays

*Par rapport à la médiane des
populations total des pays (8 741 154)
Donnée présente sur les dernières
années*

Score attractivités

*Attribution de scores, pour chaque
pays, selon indicateurs
sélectionnés*

Classement pays	Score	Score internet
avant écart interquartile 1	-50	-150
entre interquartile 1 et 2	-25	-50
entre interquartile 2 et 3	25	50
après écart interquartile 3	75	100



Analyse des pays

Sélection des indicateurs

Déjà réalisée

Pré-sélection des pays

*Par rapport à la médiane des populations total des pays (8 741 154)
Donnée présente sur les dernières années*

Score attractivités

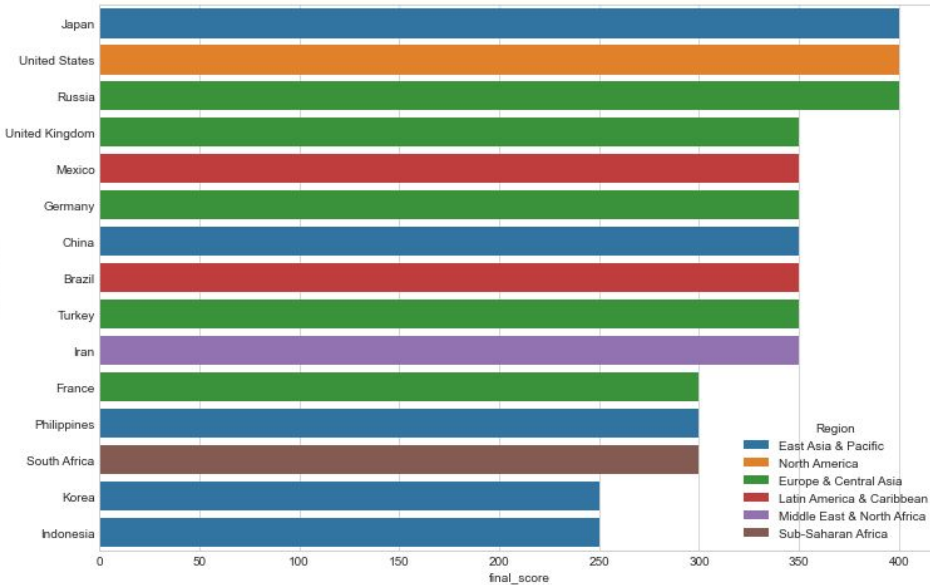
Attribution de scores, pour chaque pays, selon indicateurs sélectionnés

	Country Code	score_SE.SEC.ENRL	score_SP.SEC.UTOT.IN	score_SE.TER.ENRL	score_SP.TER.TOTL.IN	score_IT.NET.USER.P2
0	AFG	25	25	-25	25	-150
1	AGO	-50	25	-25	-25	-150
2	ARG	25	25	75	25	100
3	AUS	25	-50	25	-25	100
4	AUT	-50	-50	-25	-50	100



Analyse des pays

Top 15 des pays, séparés par régions, selon mon scoring

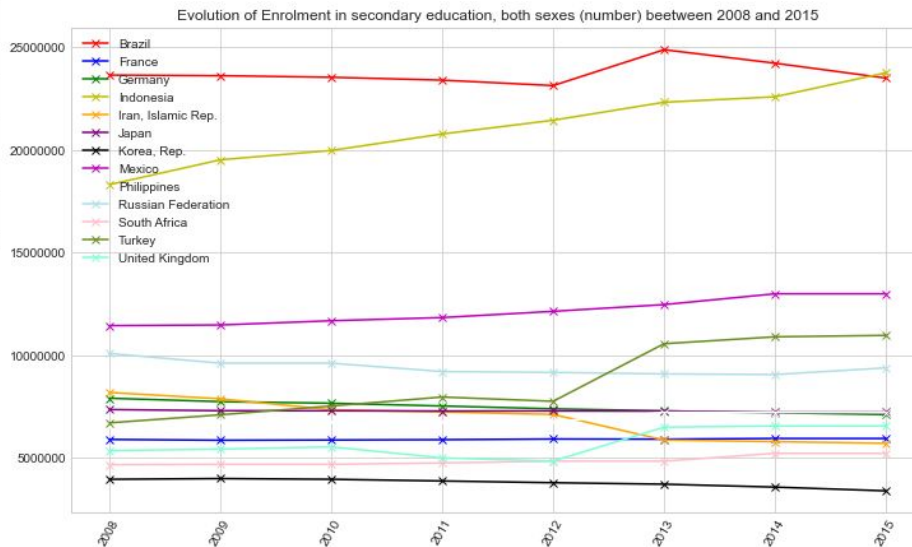
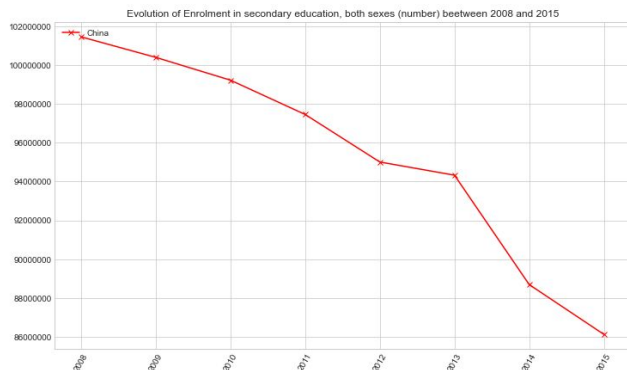


Et l'évolution ?

- Top 3 : Japon, US, Russie
- Région la plus attractive : Europe & Asie du centre



Évolution de la population au lycée (2008-2015)



Pays en croissance :

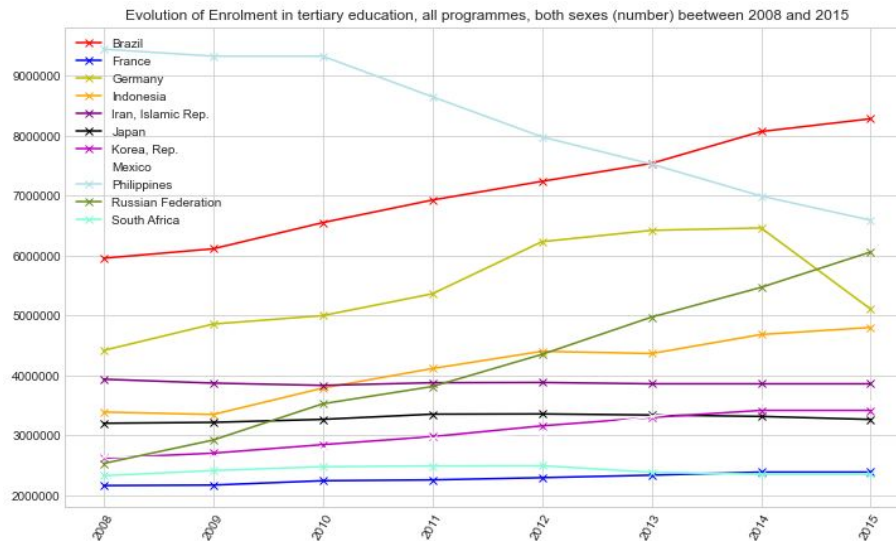
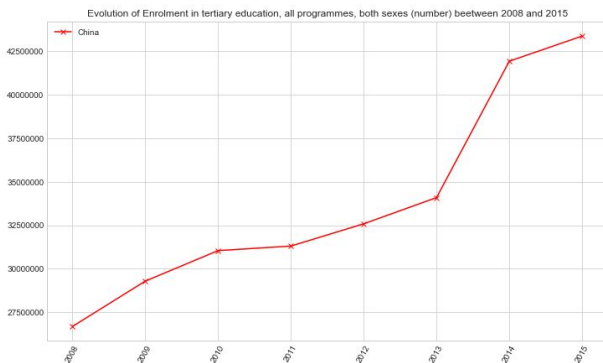
- Indonésie, Turquie, Japon

Pays en décroissance :

- Chine, Brésil

Analyse
évolution
méthode 1

Évolution population en études universitaire (2008-2015)



Pays en croissance :

- **Chine, Brésil, Russie**

Pays en décroissance :

- **Philippines, Allemagne**

Analyse
évolution
méthode 1

Analyse évolution 2.0

- Analyse des colonnes "projection" 2020 -> 2040
- Sélection des indicateurs

01

"Wittgenstein Projection: Population in thousands by highest level of educational attainment. Upper Secondary. Total

- Source : Wittgenstein Centre for demography and Global Human Capital
- Donnée : 167 pays

02

Wittgenstein Projection: Population in thousands by highest level of educational attainment. Post Secondary. Total

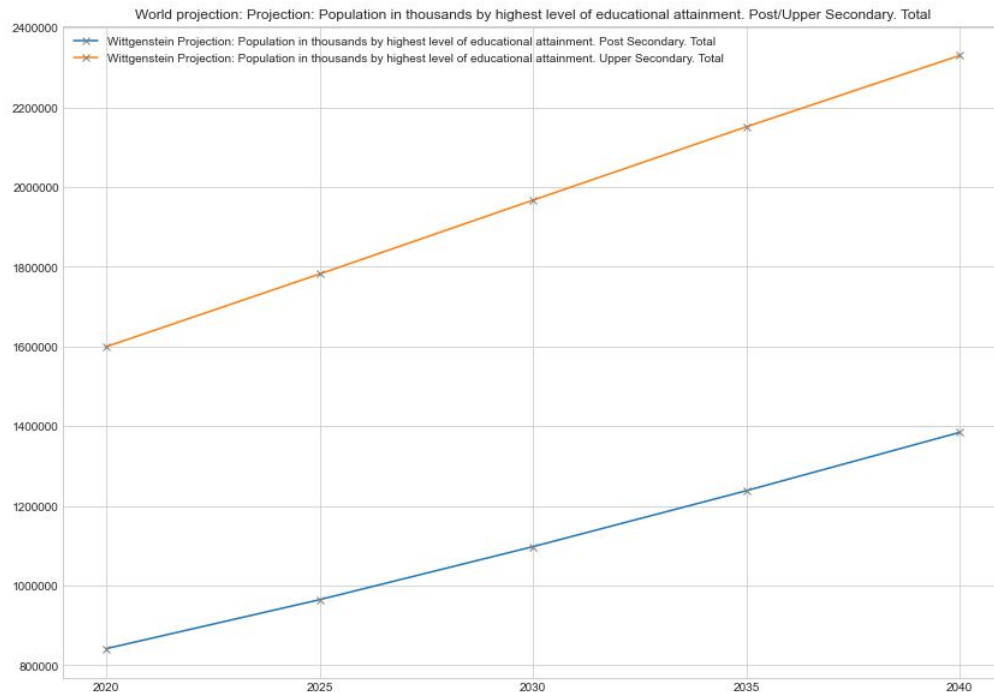
- Source : Wittgenstein Centre for demography and Global Human Capital
- Donnée : 167 pays

Pas de critère internet, mais ce critère sera probablement moins important dans les années à venir.

- Scoring basé sur la différence entre décennie et double décennie

**Analyse
évolution
méthode 2**

Evolution des indicateurs méthode 2



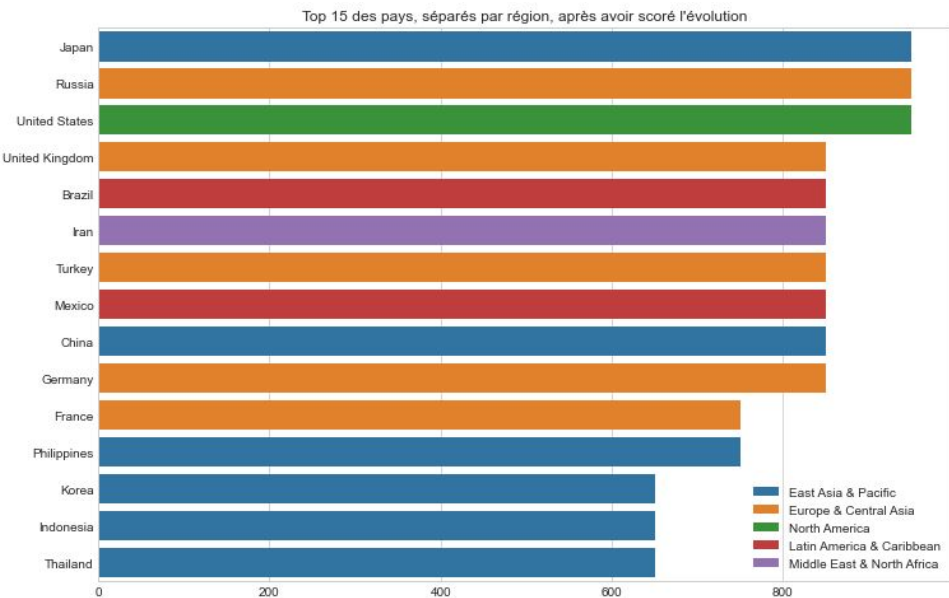
+68% nombre de personnes ayant un niveau upper secondary entre 2020-2040

+60% nombre de personnes ayant un post secondary entre 2020-2040

--> Prometteur pour l'avenir d'Academy 

Analyse
Mondial

Conclusion



Après scoring de l'évolution et analyse globale :

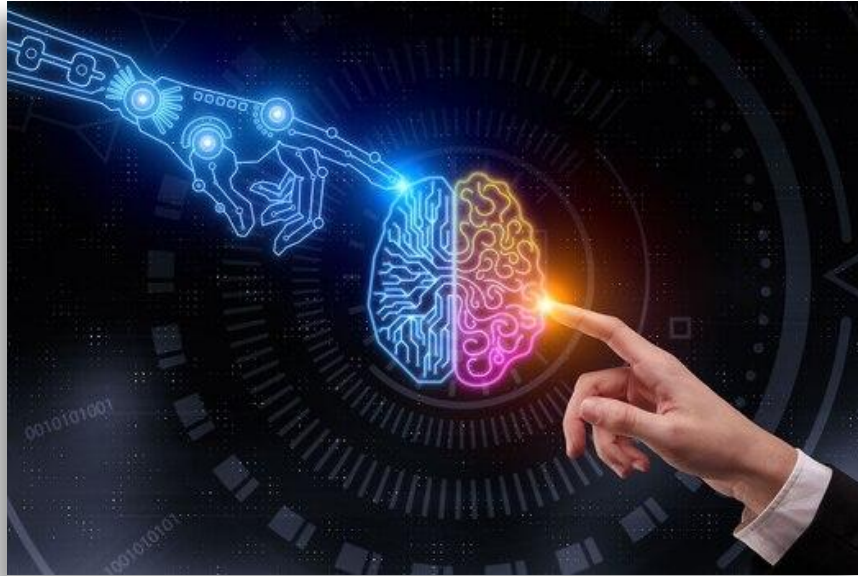
- 1) **Europe** -> Angleterre, Russie, Turquie, Allemagne, France
- 2) **Asie** -> Japon, Chine, Philippines, Indonésie, Corée
- 3) **USA**

Les limites



© CanStockPhoto.com

- Métier (comprendre indicateurs, donnée)
- Scoring : justifiable mais arbitraire
- Travail sur 9 indicateurs sur 3665 (dont aucun sur PIB/habitant, ...)
- Aucune donnée pour étude de marché
- Sélection parfois drastique (-> perte de données)



Je vous remercie pour votre attention