OC – Data scientist – World Bank Education Statitstics analysys

- **PROBLEMATIQUE:**
- ***EdTech*** propose des contenus de formation en ligne pour un public de niveau ***lycée*** et ***université***.

- **1. Pays avec un fort potentiel de clients** pour nos services ?

- 2. Quelle sera **l'évolution** de ce potentiel de clients ?

- 3. Dans quels pays l'entreprise doit-elle opérer **en priorité** ?

academy

OC – Data scientist – World Bank Education Statitstics analysys

- **Données utiles : Bases de données de la banque mondiale recensant plus de 4000 indicateurs pour l'ensemble des pays et grandes régions du monde.**

Un jeu de 4 datasets principaux :

a. **EdStatsData**

Table principale contient 3665 indicateurs pour 242 pays et régions du monde *« renseignés »* depuis 1970.

B. **EdStatsSeries**

Détails concernant les indicateurs : description, source, méthode…

b. **EdStatsCountry**

Informations générales sur les pays : région, monnaie, richesse,…

d. **EdStatsCountry-Series**

Source ou mode de calcul de quelques séries (pays/indicateur)

academy

OC – Data scientist – World Bank Education Statitstics analysys

- C’est le dataset *EdStatsData.csv* qu’on utilisera principalement :

  - Taille = 886 930 lignes (1 ligne = une paire 1 pays + 1 indicateur
  - 217 pays et 25 régions
  - 3665 indicateurs
  - Pas de doublons
  - Pas de ligne manquantes

- Exemples d’indicateur :
  - Population pour un âge donné.
  - Taux de chômage
  - Taux d’élèves en primaire, au collège, au lycée, à l’université.
  - Le niveau en lecture, en mathématiques
  - Les soutiens gouvernementaux, etc.

academy

OC – Data scientist – World Bank Education Statitstics analysys

- Méthode :

1. Choix des années d'étude à considérer : **les années 2000** suffisent pour répondre aux questions posées.

1. Choix des indicateurs pertinents selon :
   - **Des critères qualitatifs**
   - **Taux de remplissage pour les années considérées**

2. Définir pour chaque indicateur ***un seuil*** souhaité

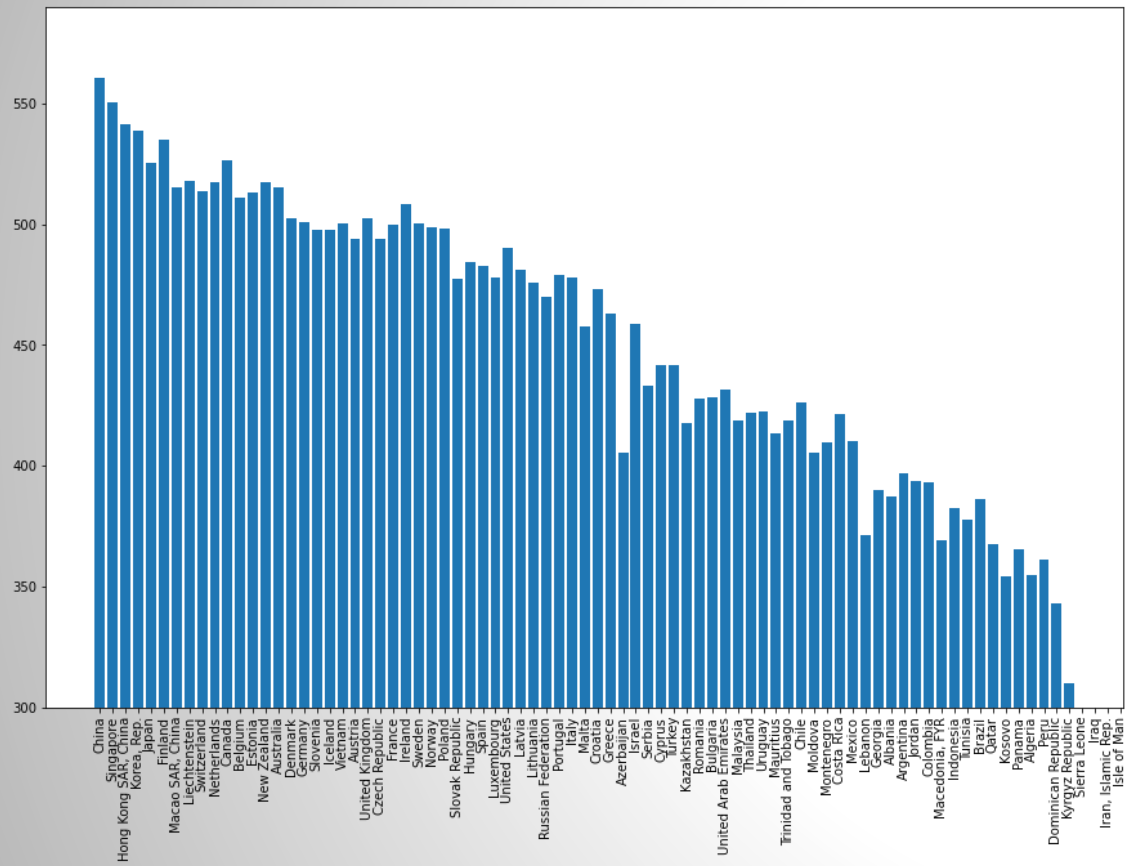
Indicateurs retenus :

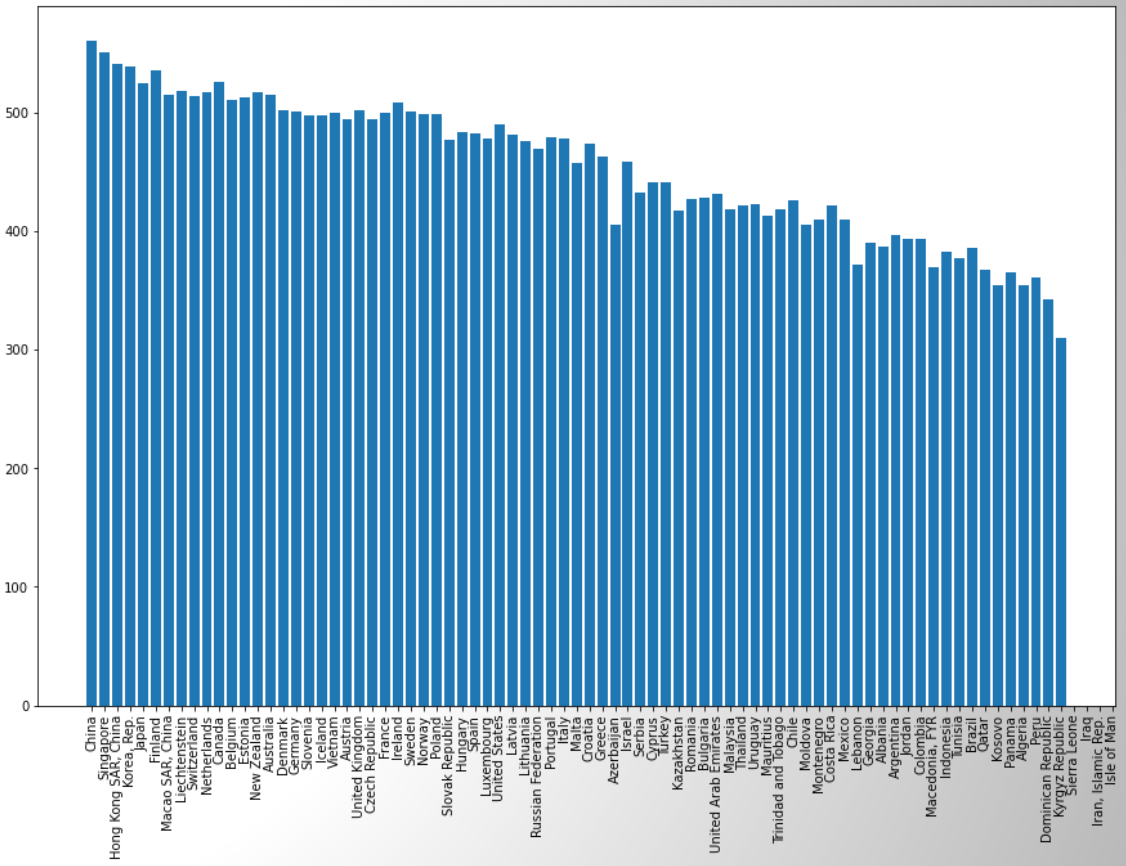
- Démographiques :
  - La population des personnes de 15 à 25 ans (cibles de EdTech)
- Technologiques
  - L'accès à internet
- Economiques
  - (PIB par habitant)
- Sociologiques
  - Résultat PISA + Taux d'alphabétisation (peu pertinent)
  - L'appétence pour les institutions privés (dans le secondaire et le supe

Taux de remplissage des indicateurs potentiellement pertinents par année. (sur 217 pays)

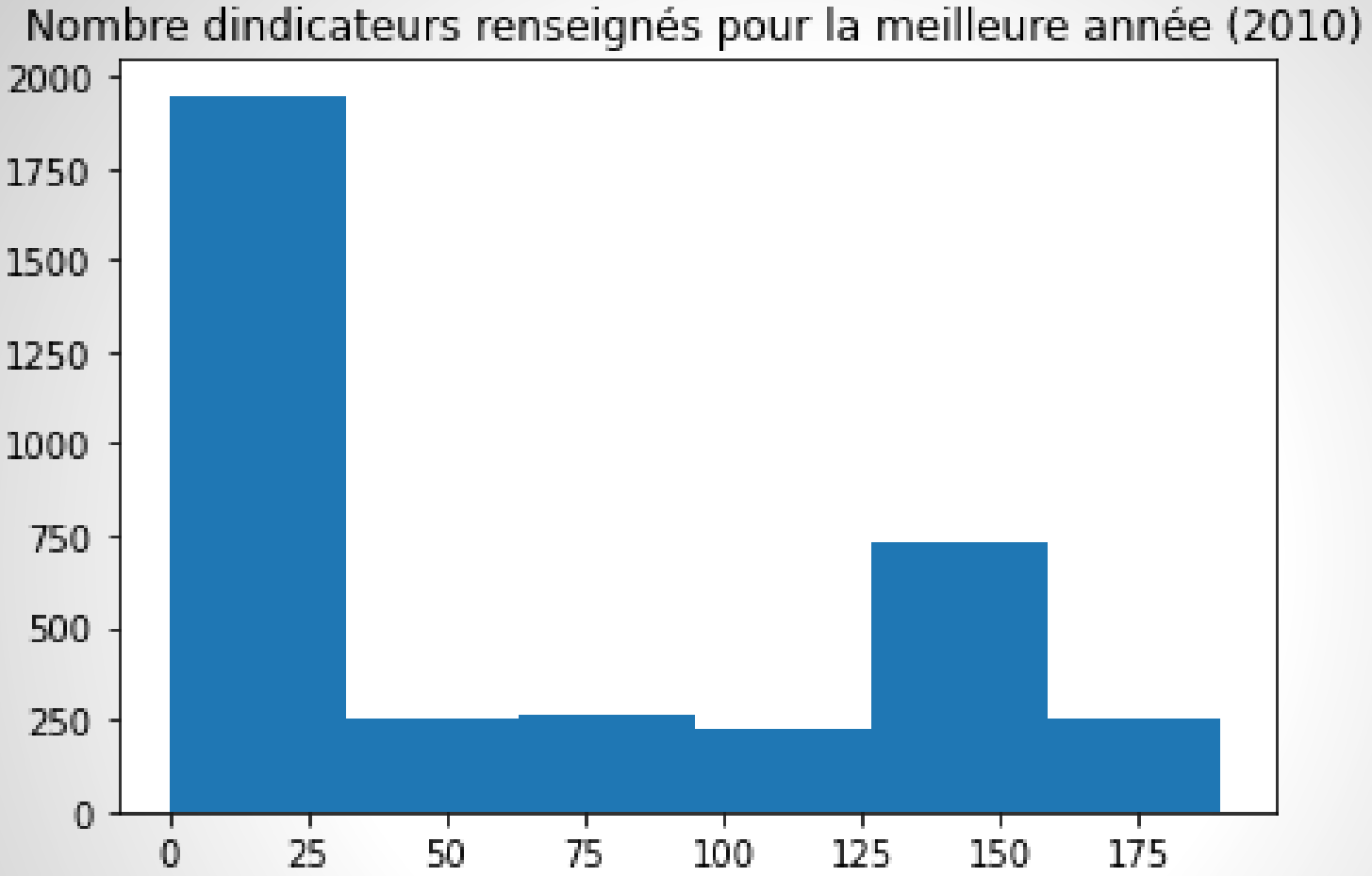
Indicator Name	Country Name	Indicator Code	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
GDP per capita (current US$)	217	217	199	199	203	203	204	204	205	204	203	202	203	203	199	200	197	196	189	0
Internet users (per 100 people)	217	217	196	197	199	193	196	198	197	204	203	202	202	204	202	201	201	201	201	0
Population, ages 0-14, total	217	217	194	194	194	194	194	194	194	194	194	194	194	194	193	193	191	191	191	0
Population, ages 10-18, total	217	217	190	191	192	192	191	191	187	181	181	181	181	181	181	181	181	181	0	0
Enrolment in secondary education, both sexes (number)	217	217	149	151	150	146	158	158	149	157	154	153	151	156	150	141	141	123	8	0
Barro-Lee: Percentage of population age 15+ with secondary schooling. Completed Secondary	217	217	144	0	0	0	0	144	0	0	0	0	144	0	0	0	0	0	0	0
Barro-Lee: Population in thousands, age 15-19, total	217	217	144	0	0	0	0	144	0	0	0	0	144	0	0	0	0	0	0	0
Enrolment in tertiary education per 100,000 inhabitants, both sexes	217	217	121	121	130	133	131	128	126	124	130	135	136	138	132	123	93	4	0	0
Percentage of enrolment in secondary education in private institutions (%)	217	217	95	97	111	105	110	115	118	118	120	121	130	135	131	131	131	115	6	0
Percentage of enrolment in tertiary education in private institutions (%)	217	217	59	61	70	72	79	75	76	86	96	102	104	111	107	114	111	103	7	0
Youth literacy rate, population 15-24 years, both sexes (%)	217	217	43	28	18	10	18	20	26	34	31	32	47	57	41	29	36	32	17	0
PIAAC: Young adults by proficiency level in problem solving in technology-rich environments (%). No computer experience or failed the ICT core test	217	217	0	0	0	0	0	0	0	0	0	0	0	0	22	0	0	8	0	0
PISA: 15-year-olds by reading proficiency level (%). Level 6	217	217	41	0	0	40	0	0	55	0	0	70	0	0	64	0	0	71	0	0
PISA: Mean performance on the mathematics scale	217	217	42	0	0	41	0	0	56	0	0	69	0	0	62	0	0	71	0	0
PISA: Mean performance on the reading scale	217	217	42	0	0	41	0	0	55	0	0	70	0	0	64	0	0	71	0	0
Personal computers (per 100 people)	217	217	162	182	182	179	179	171	99	48	27	3	0	0	0	0	0	0	0	0

Exemple des données du test PISA.

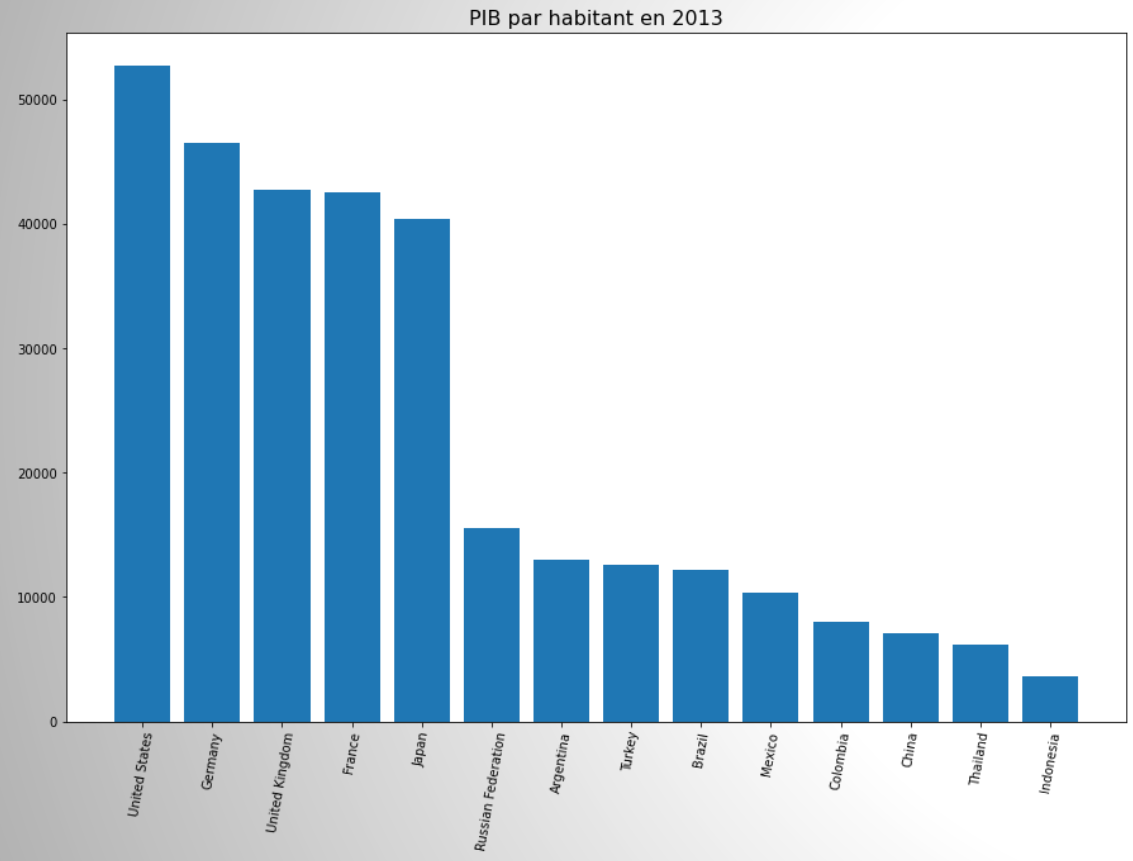


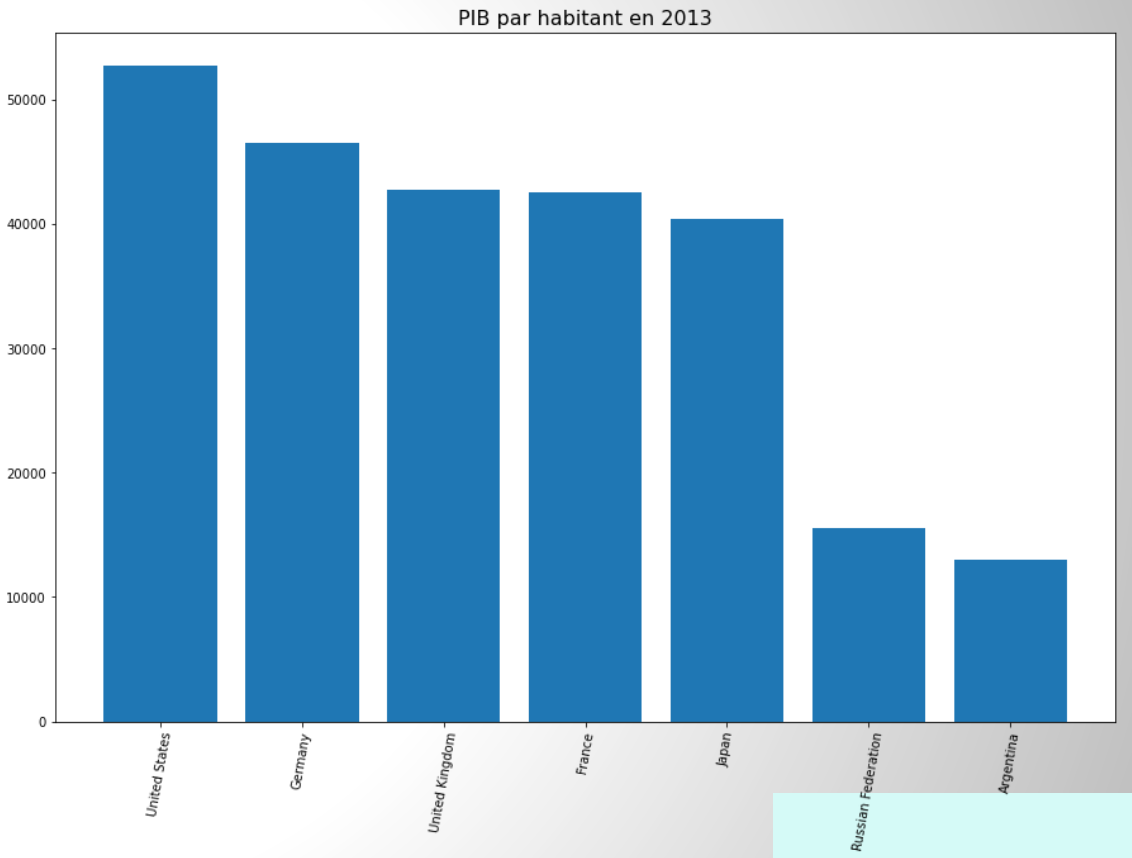


C’est le même graphique, mais l’échelle en ordonnée est différente. Cela permet de prendre des décisions.

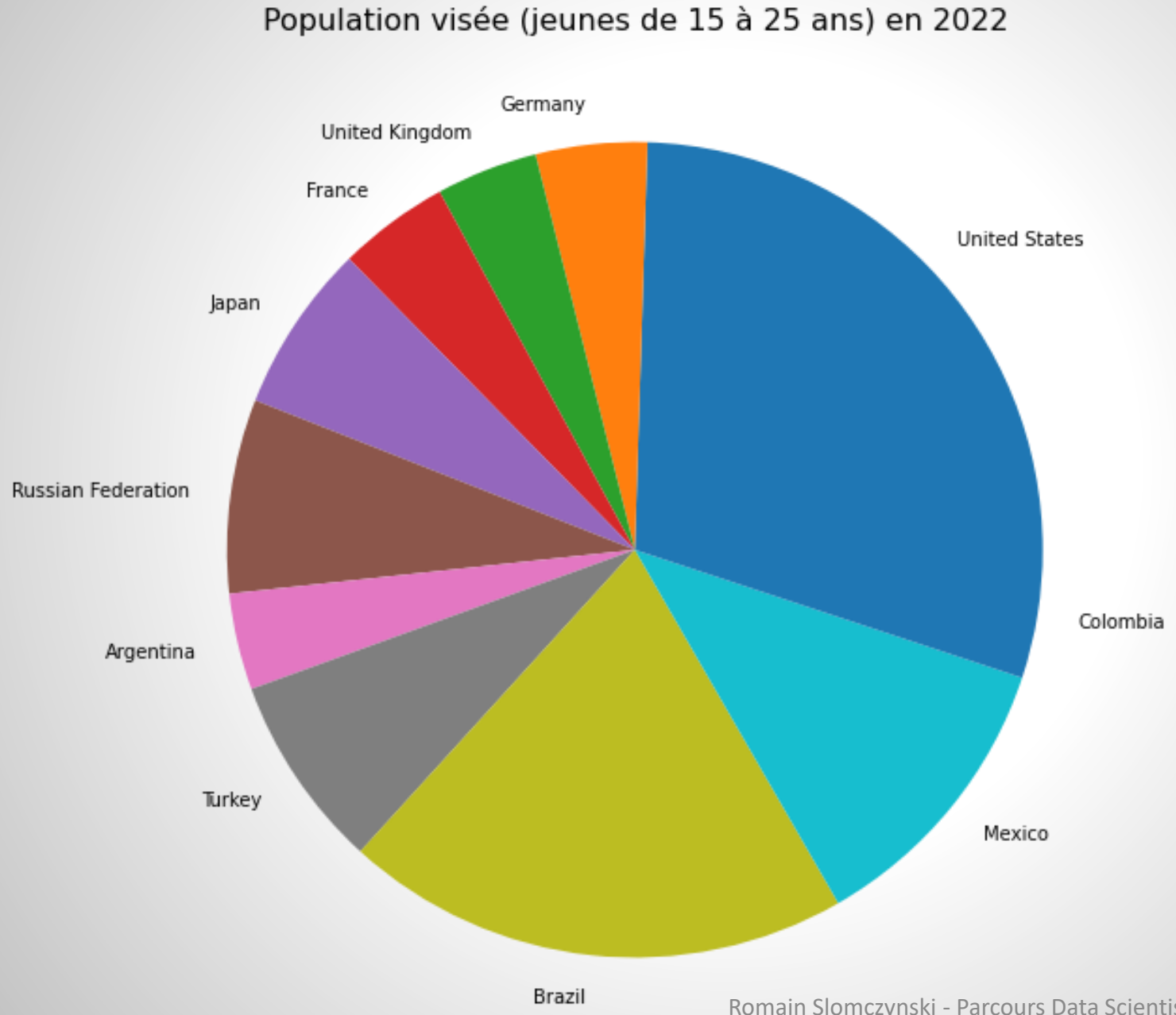

Nombre dindicateurs renseignés pour la meilleure année (2010)
2000
1750
1500
1250
1000
750
500
250
0
0
25
50
75
100
125
150
175

PIB par habitant: une grande disparité qui aide à la prise de décision





academy



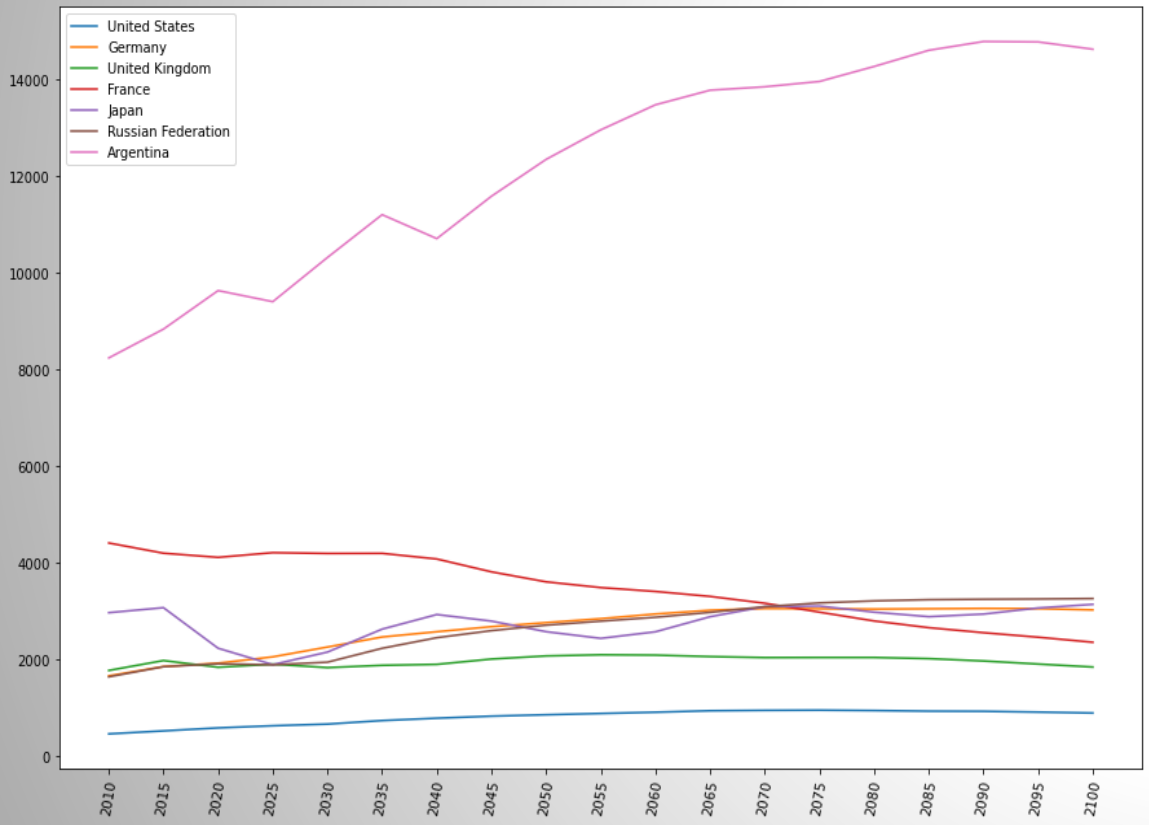
Un pays c'est aussi une langue.

Avons-nous les moyens de produire du contenu pédagogique en japonais ?

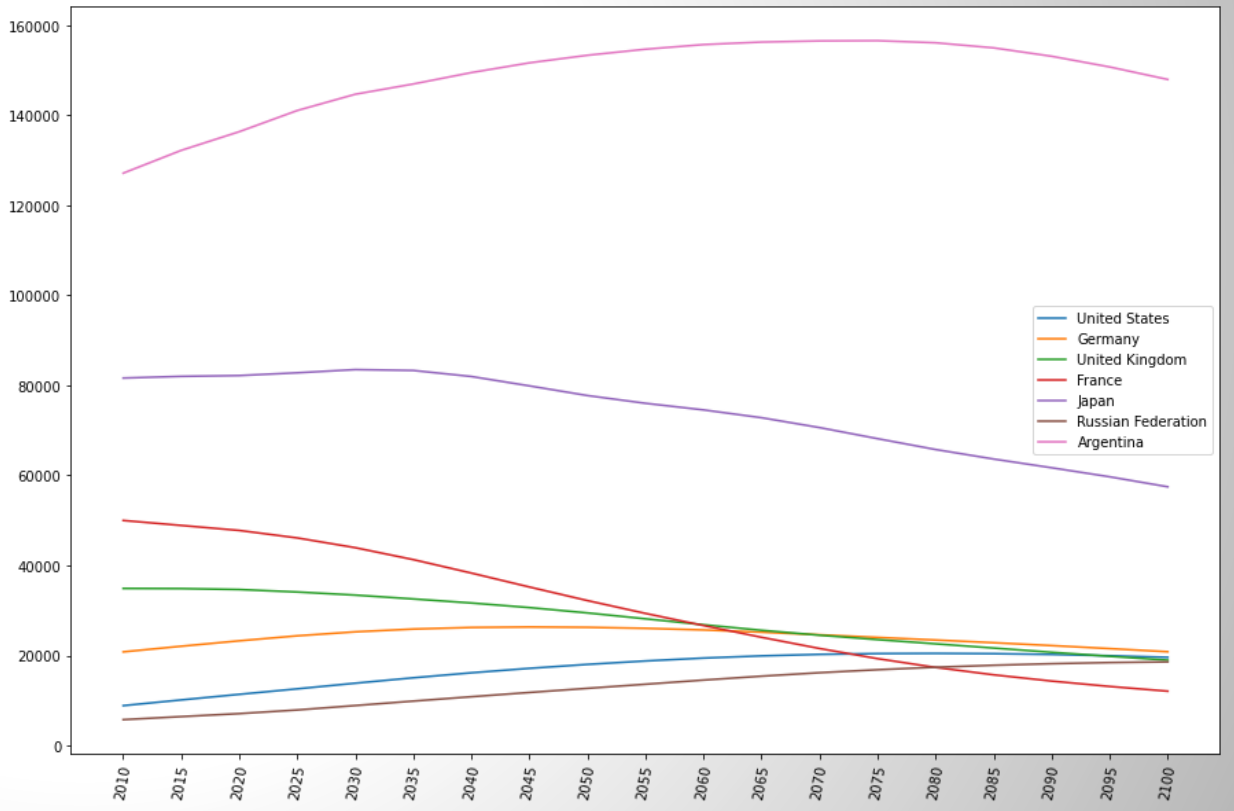
academy

OC – Data scientist – Projet 2

Population âgée de 25 à 29 ans en milliers ayant atteint un niveau de postsecondaire.



Population en milliers ayant atteint un niveau de secondaire supérieur.



A vous de jouer ! Rendez-vous au 3.8 du notebook et choisissez le seuil des indicateurs de références :

- L'accès à internet
- Le PIB par personne
- Le nombre d'individus potentiellement ciblés
- La capacité d'apprentissage des étudiants
- L'appétence pour les institutions privés pour l'enseignement secondaire

Vous avez des questions ?

academy