



DATA SCIENCE PROJET 2

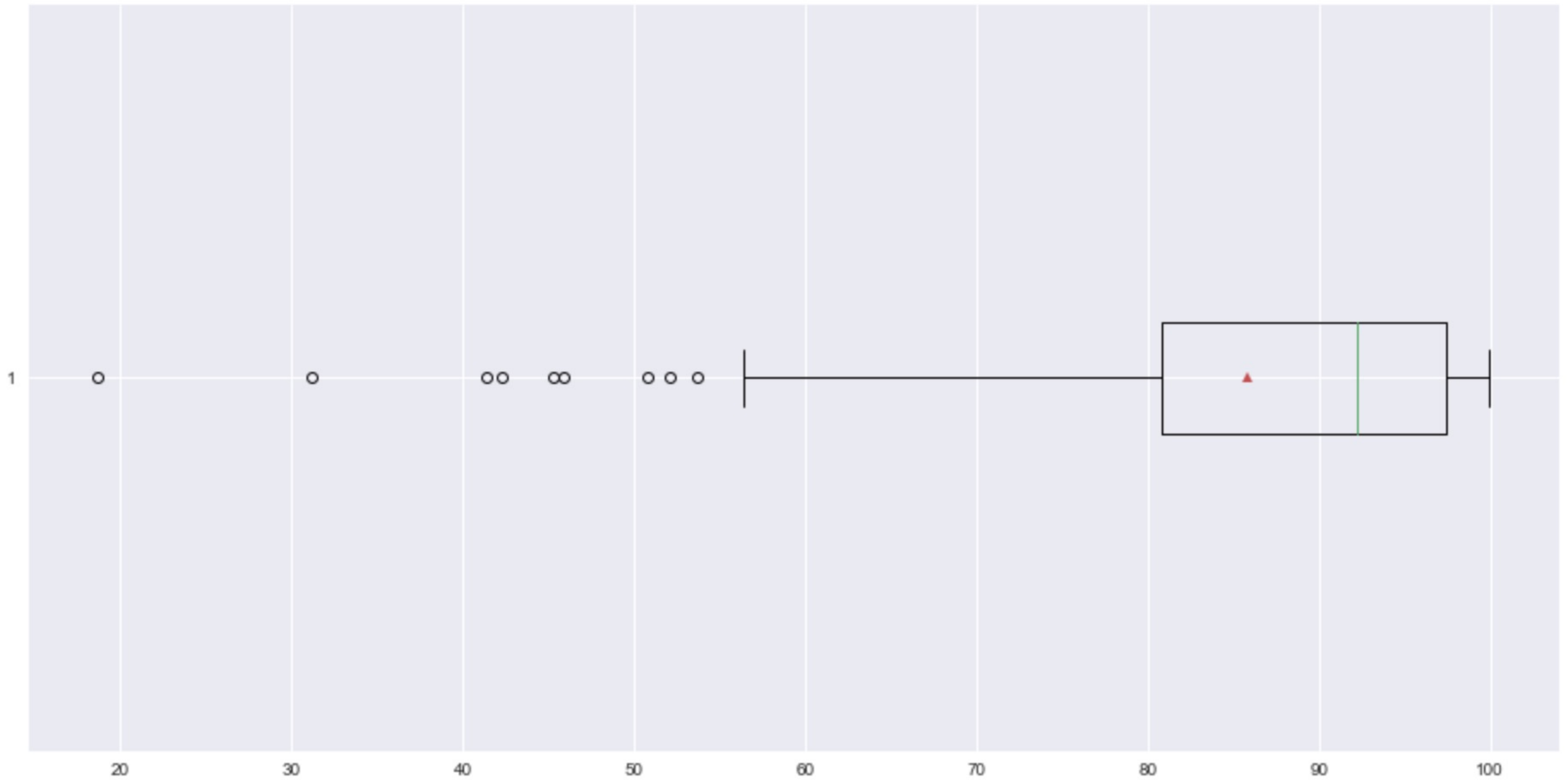
Analysez des données de systèmes éducatifs



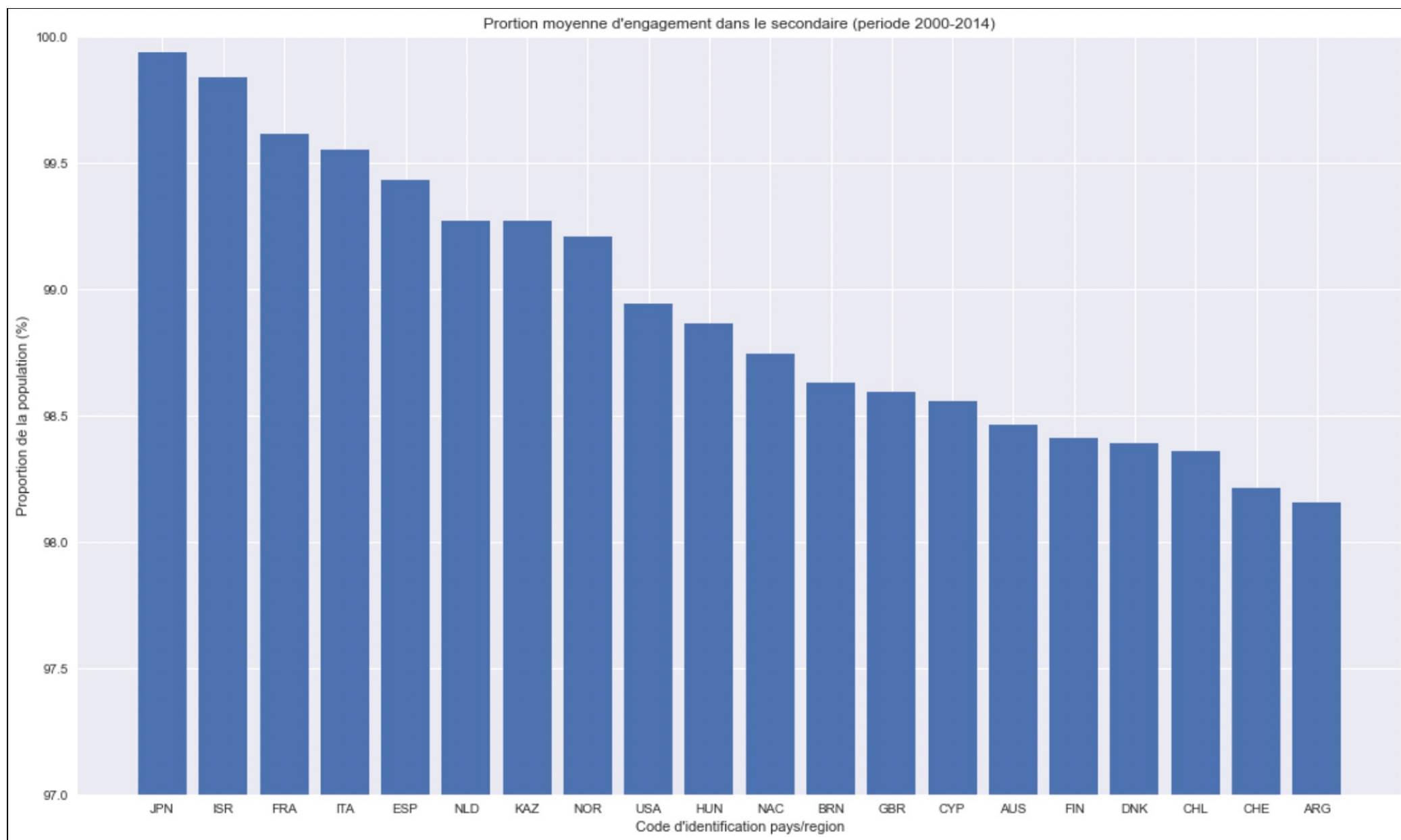
Etat général du dataset et selection des indicateurs

- Une étude préliminaire des données montre un taux de remplissage très faible, même en affinant la période sur 2000-2020 : 69.74% des valeurs manquent.
- La plus grande partie des données pertinente se situent entre 2000 et 2020 (en réalité 2015, puisque la plupart des indicateurs ne dépassent pas cette année)
- Les indicateurs étudiés sont :
 - *Général : Le pouvoir d'achat par habitant, la démographie cible (15-25ans avec accès à internet).*
 - *Secondaire : Le taux d'inscription au secondaire, l'étude Baro Lee montrant par tranche d'âge le taux d'inscription et le taux de réussite dans le secondaire(*).*
 - *Supérieur : Taux d'inscription présent selon le site mais absent du dataset, l'étude Baro Lee montrant par tranche d'âge le taux d'inscription et le taux de réussite dans le supérieur(*).*

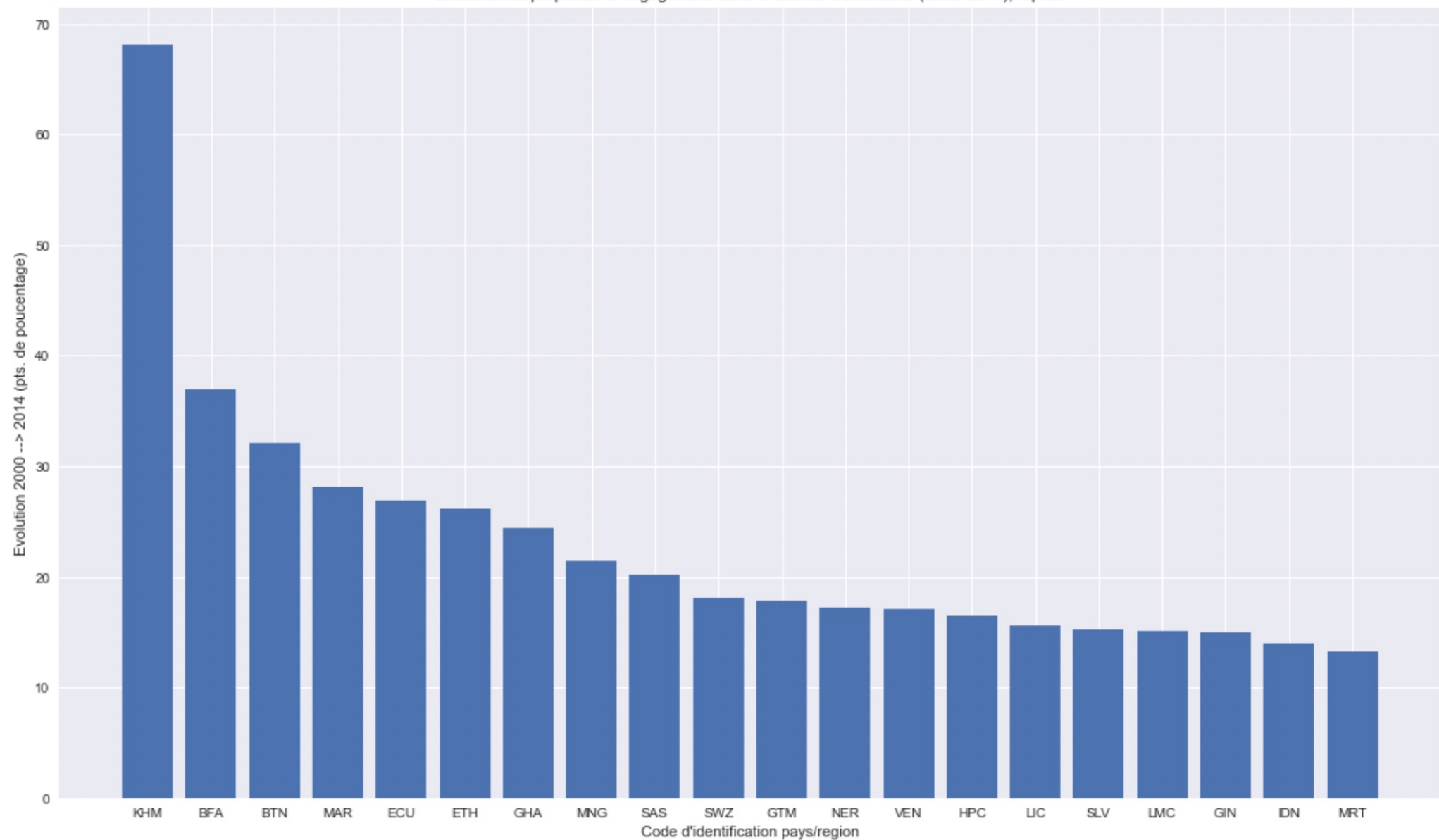
Etudes des données (cf. notebook)

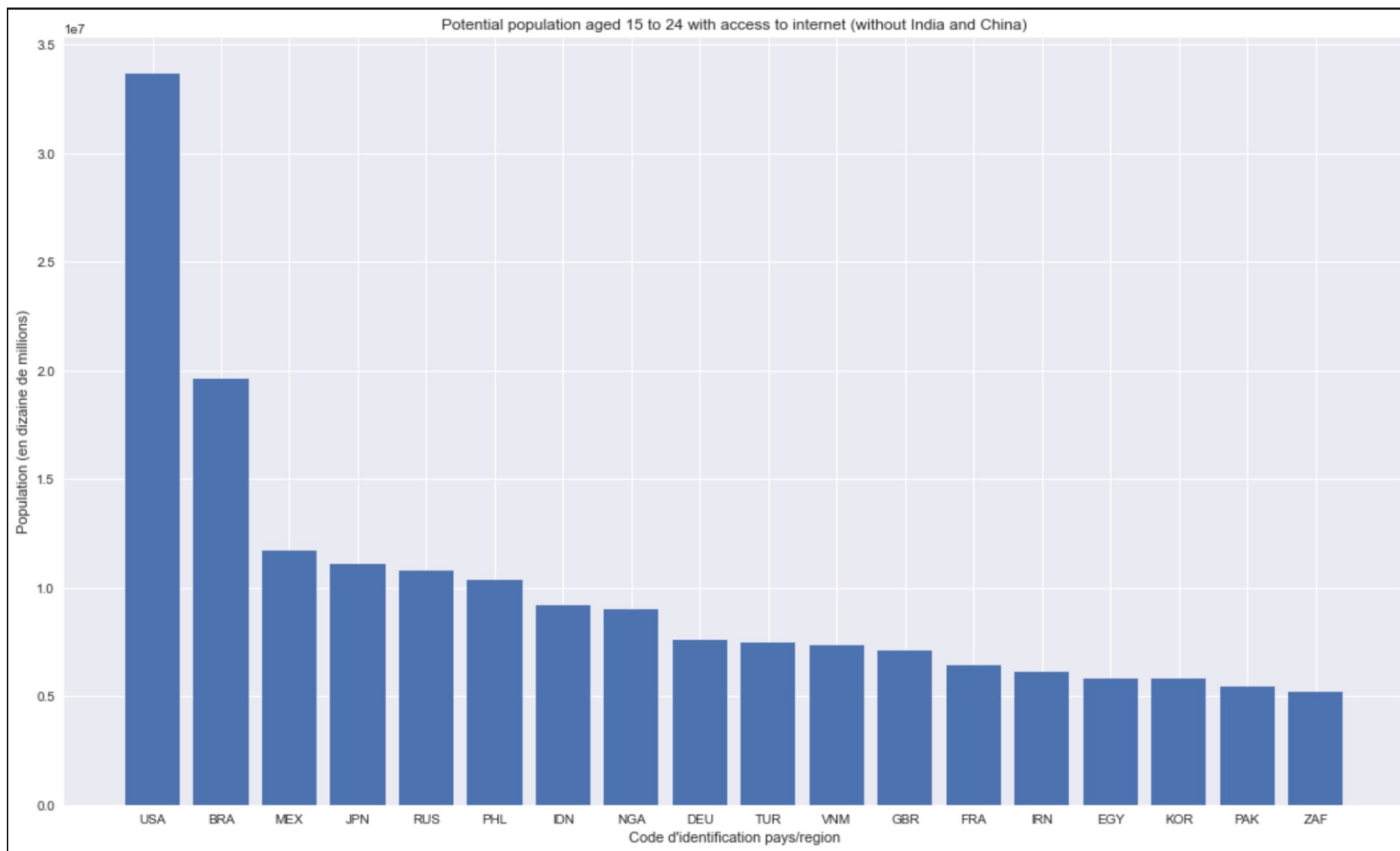


Graphique (boxplot) représentant la moyenne du taux d'inscription dans le secondaire entre 2000 et 2015

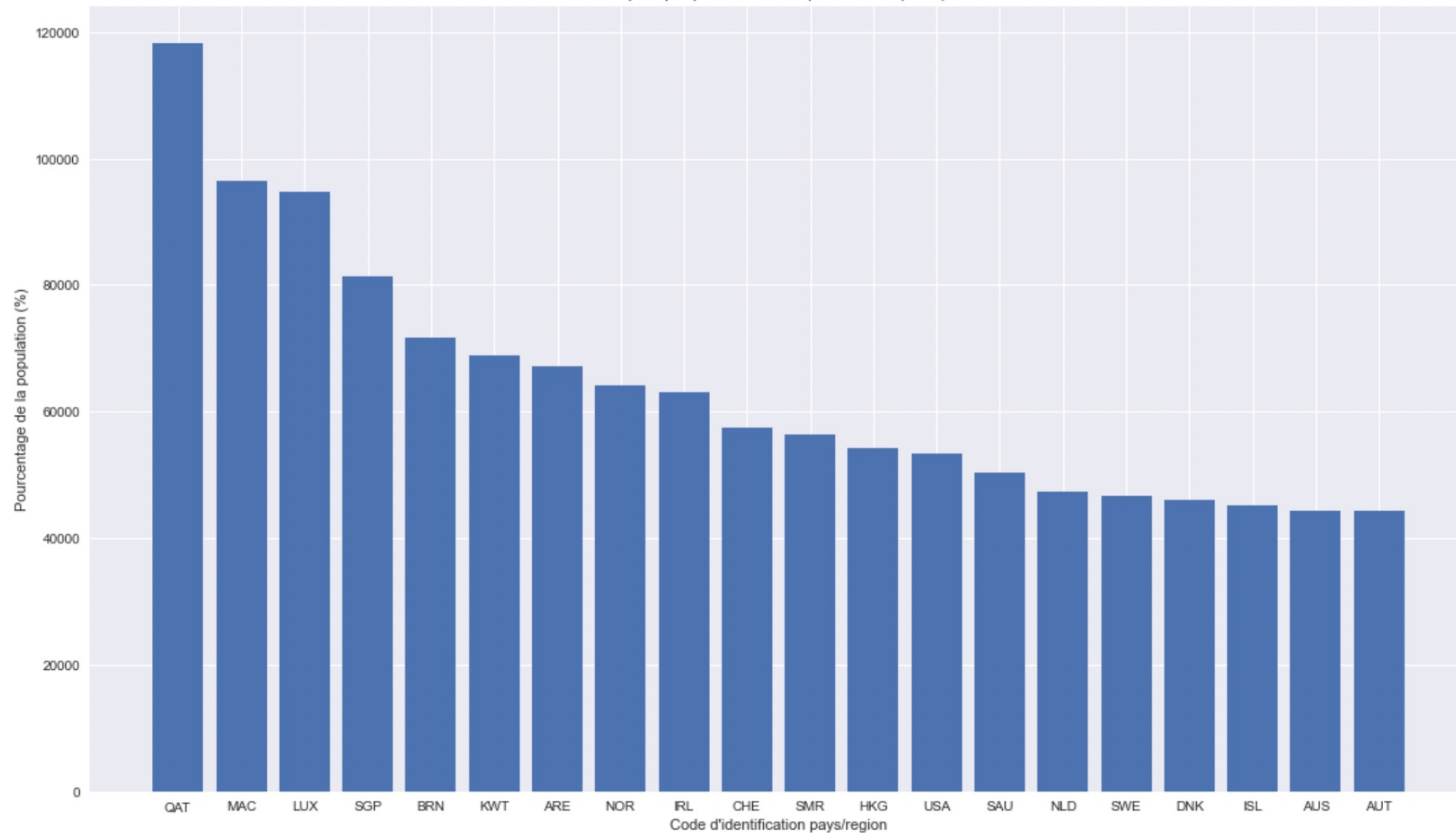


Evolution de proportion d'engagement dans le secteur secondaire (2000-2014), top 20





Top 20 par pouvoir d'achat par habitant (2016)



Le système de notation

- Les indicateurs notés dans la sélection ci dessus permettront de classer les différents pays et zones selon la valeur de leur indicateur → Chaque indicateur ou statistique sélectionné est noté sur 100.
- Les indicateurs et statistiques calculées à partir de ces indicateurs auront un poids différent.
- Note finale sur 100 en faisant une moyenne avec coef. de toutes les statistiques
- Si un pays ne figure pas dans une étude par manque de donnée, son résultat est ignoré (pas pénalisé), s'il manque un indicateur à un pays, il sera ignoré : le résultat ne serait pas fiable

→ Coefficients

- Taux d'inscription moyen aux études secondaires : 1
- Evolution du taux d'inscription aux études secondaires : 3
- Evolution ratio inscription/sucés (Baro Lee) : /// Non comptabilisée : voir étude de l'indicateur
- Pourcentage de la population ayant accès à internet * Population 15-24 ans : 2
- Pouvoir d'achat par habitant * Population 15-24 ans : 1

	country_code	average_2	evo_2	target_dem	ppp	combined_rating
0	TUR	61.111111	72.222222	93.333333	72.375691	76.348680
1	COL	65.555556	74.444444	88.333333	53.038674	74.692756
2	ESP	94.444444	53.333333	83.333333	82.320442	73.454266
3	FRA	96.666667	44.444444	91.666667	85.082873	73.119398
4	USA	90.000000	37.777778	98.333333	92.817680	71.892879
5	IDN	18.888889	83.333333	95.000000	48.066298	70.417741
6	MAR	13.333333	95.555556	84.444444	36.464088	69.868631
7	PHL	42.222222	70.000000	95.555556	35.911602	68.702271
8	ECU	21.111111	94.444444	73.333333	46.408840	68.640884
9	JPN	98.888889	24.444444	96.666667	85.635359	67.119091

Selon ce système de scoring,
les 10 premiers pays sont

- La Turquie
- La Colombie
- L'Espagne
- La France
- Les Etats Unis
- L'Indonésie
- Le Maroc
- Les Philippines
- L'Equateur

*Note : Tolérance sur le manque de
note = 0 (appliquer le paramètre
thresh sur la fonction
'DataFrame.dropna()')*

Amélioration potentielles

- Inclusion de plus d'indicateurs
- Utilisation de données plus actuelles (en utilisant l'API de worldblank par exemple)
- Adaptation du modèle de scoring pour refléter plus précisément les besoin d'*Academy*
- Étude des marges | Coûts d'établissements dans une nouvelle région / Formation de l'équipe locale / traduction des cours etc.