



PROJET 2

ALEXIS MARCEAU

PROBLÉMATIQUE ET PRESENTATION DU JEU DE DONNEES



Formation en ligne pour un public de niveau lycée et université.



Projet d'**expansion** à l'**international** de l'entreprise.



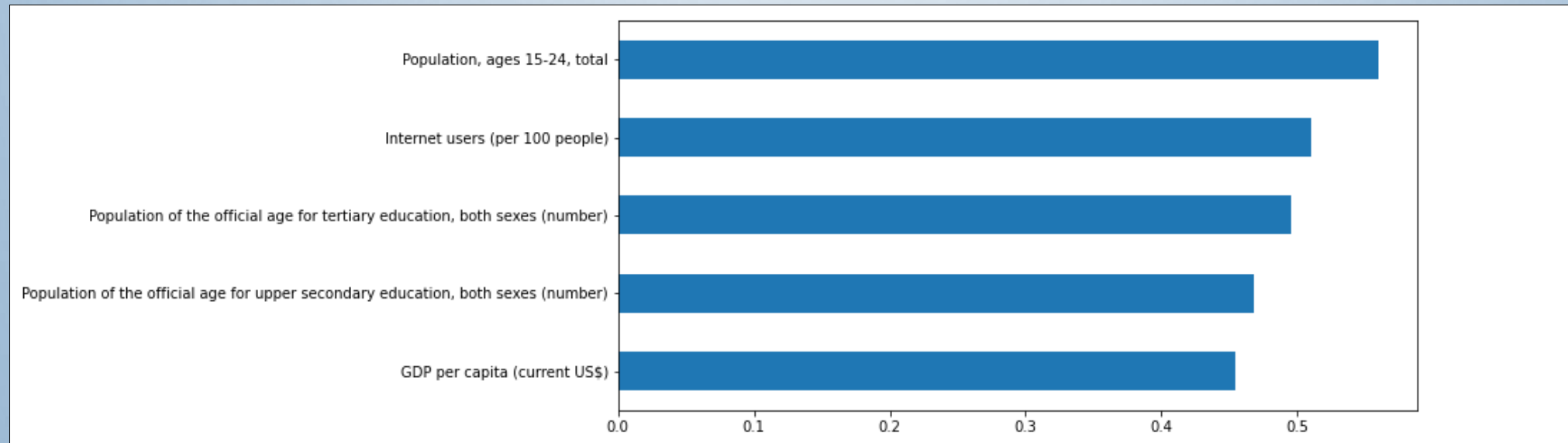
Déterminer les **pays** et les **régions** où se développer à l'aide des jeux de données.

5 fichiers de données :

- *EdStatsFootNote* et *EdStatsCountry-Series* : indication sur la source des données.
- *EdStatsCountry* : contient toutes les régions du monde utiles pour notre étude.
- *EdStatsSeries* : description de tous les indicateurs.
- *EdStatsData* : valeur des indicateurs par pays et par an.

CHOIX DES INDICATEURS

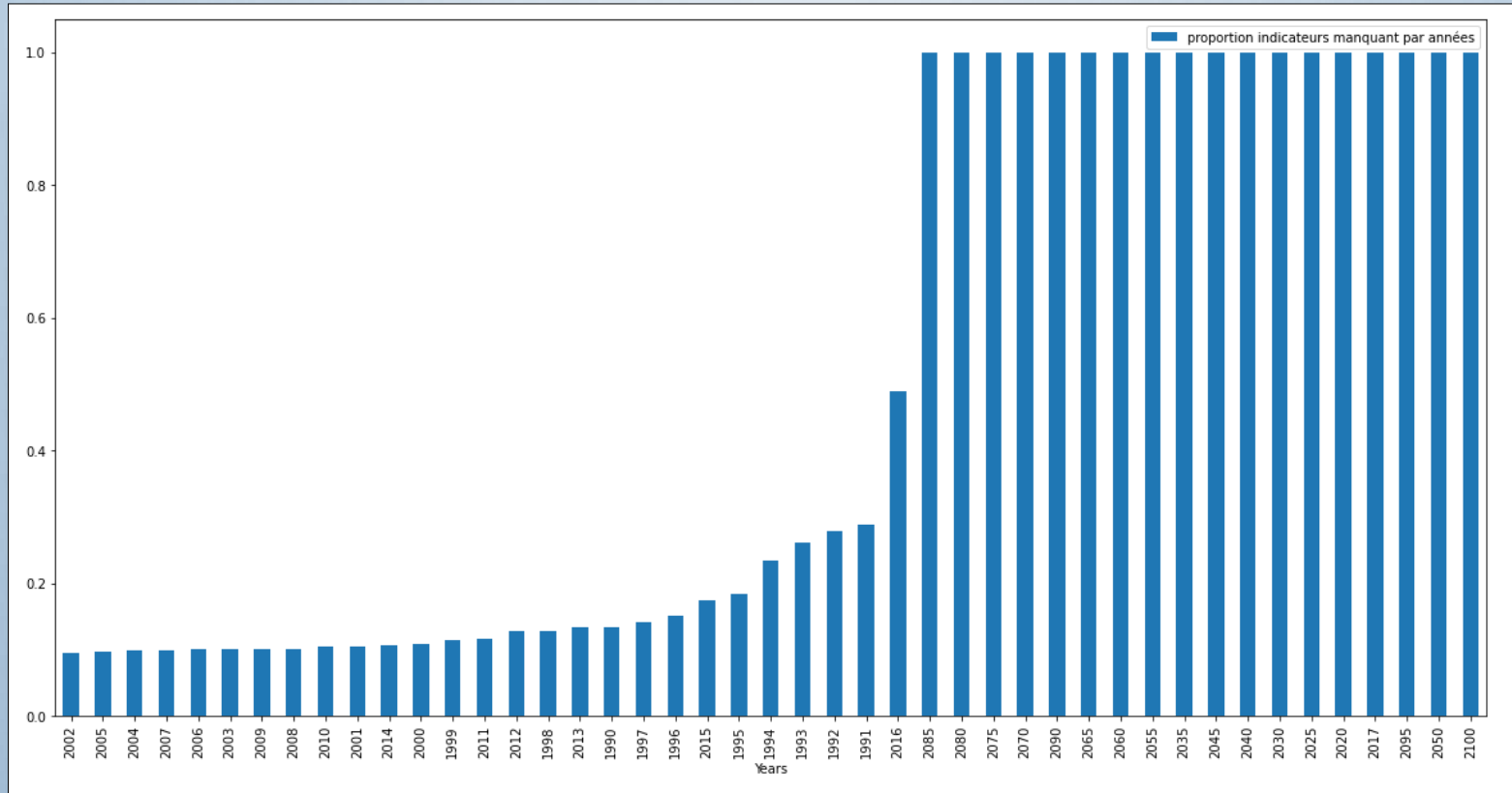
Indicateurs avec le moins de valeur manquantes,
Je choisis les indicateurs suivants qui possèdent tous moins de 60% de valeurs manquantes et qui sont cohérent avec l'étude :



- Je choisis une approche où l'on souhaite s'implanter et dégager des revenus rapidement sur une population élevée et qui dispose dès à présent d'un PIB élevé et un accès internet élevé.

TAUX DE REMPLISSAGE DES INDICATEURS PAR ANNÉES

Observation sur le jeu de données complet :



Je supprime les années avec 100% de valeurs manquantes

SCORE POUR LES RÉGIONS

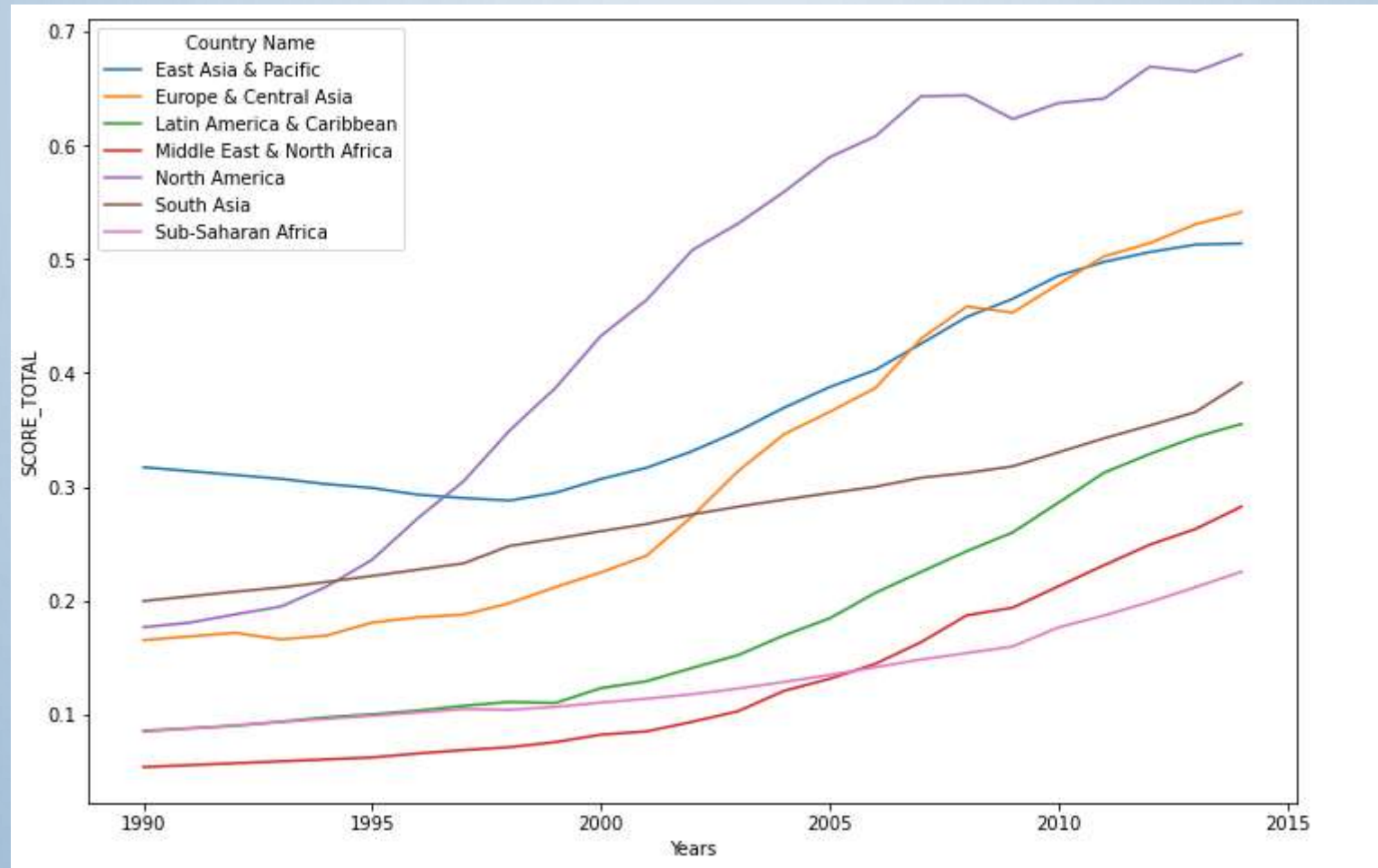
Pour chaque indicateur :

- Normalisation des valeurs entre 0 et 1 en divisant par la valeur max

Le score final est la moyenne des valeurs des indicateurs normés (cf annexe)

On peut tracer un graphique montrant l'évolution du score final au cours des années pour chaque région en utilisant la librairie seaborn.

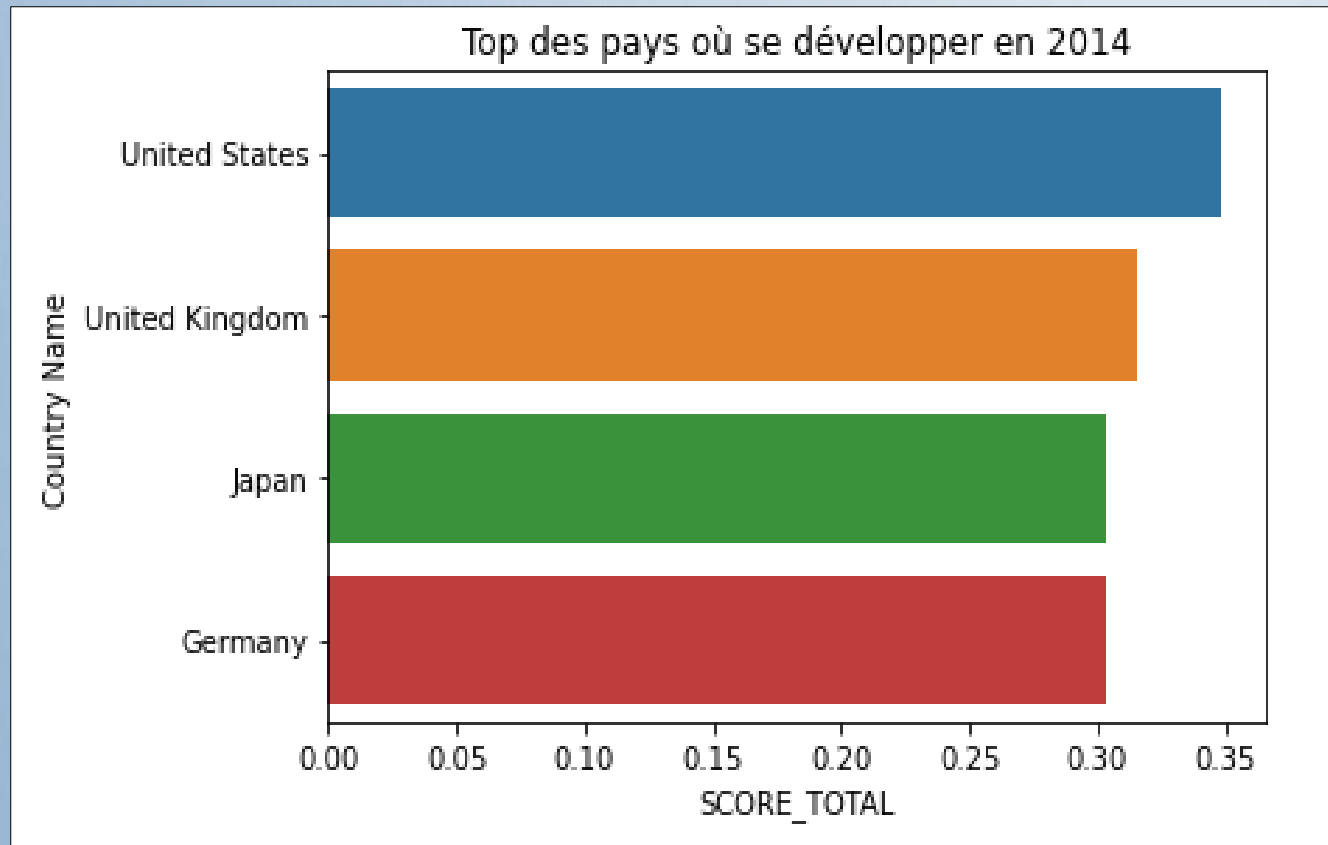
TOP RÉGIONS



Les scores sont directement issus des indicateurs, ils sont bornés entre 0 et 1.

3 régions en tête : Amérique du nord, Europe et Asie de l'Est/ Pacifique.

FOCUS SUR L'ANNÉE LA PLUS RÉCENTE AVEC LE MOINS DE VALEURS MANQUANTES : 2014



Calcul du score identique aux régions:

- Normalisation des valeurs entre 0 et 1 en divisant par la valeur max

Le score final est la moyenne des valeurs des indicateurs normés (cf annexe)

Seuillage sur les pays basé sur les valeurs moyennes (à discuter) :

Internet users > 46%

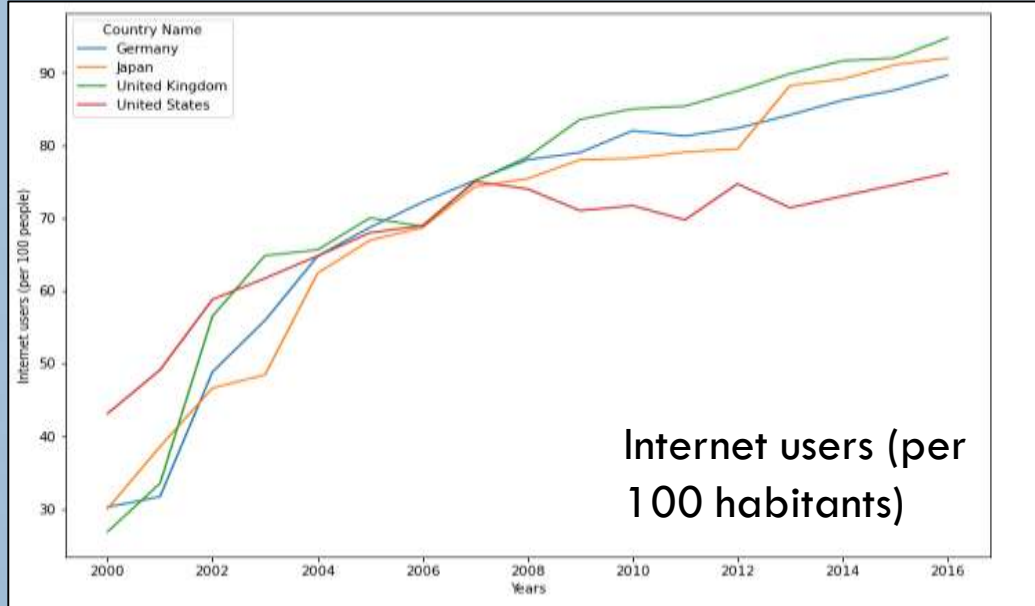
GDP per capita > 16 000 \$

Pop age 15-24 > 7 000 000 personnes

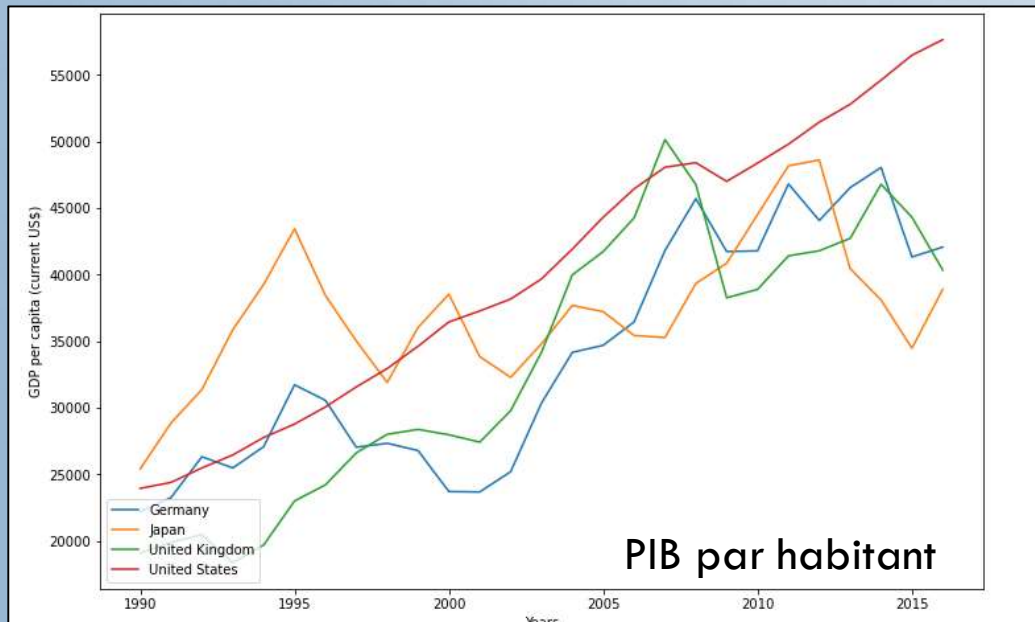
Pop age upper education > 2 000 000 personnes

Pop age tertiary education > 3 000 000 personnes

PERSPECTIVES D'EVOLUTION : POTENTIEL D'ACHAT

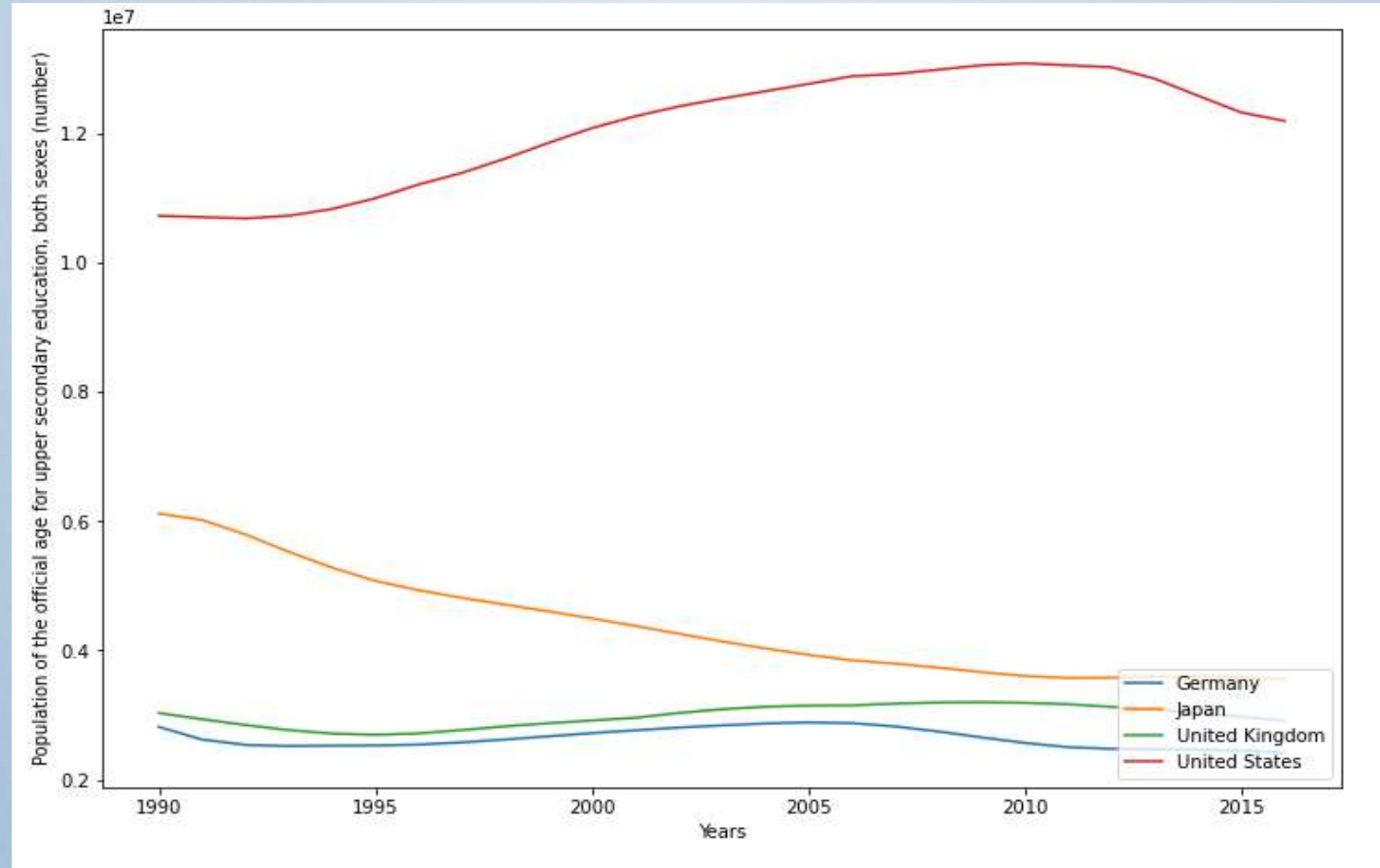


Couverture internet « relativement faible » pour les USA comparé aux autre pays en tête. PIB par habitant élevé ce qui pourrait être plus facile pour convaincre la population à payer des cours.



PERSPECTIVES D'EVOLUTION : VOLUME DU POTENTIEL D'AFFAIRES

Population en âge pour le lycée

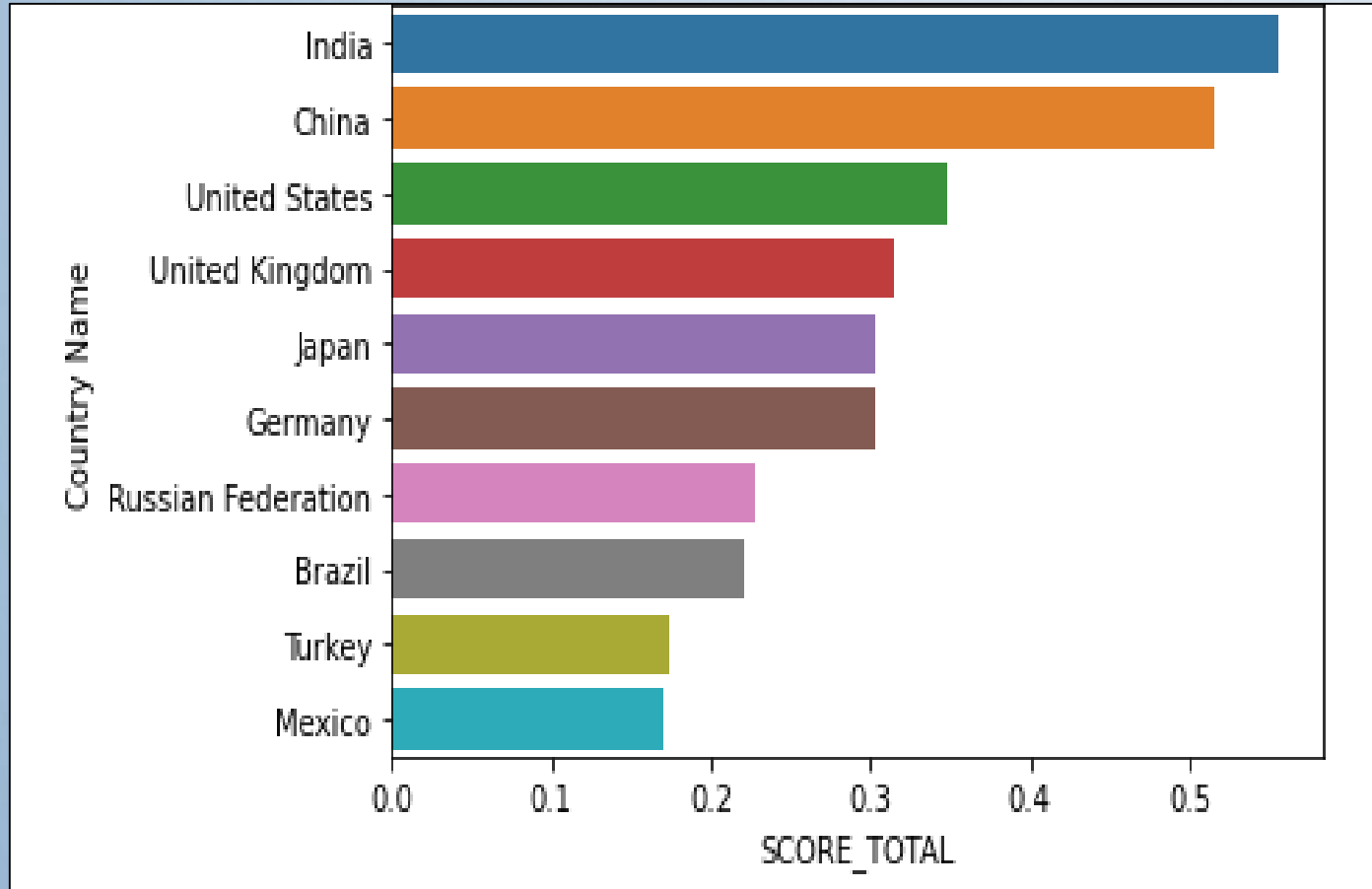


USA en tête, dispose d'un volume de marché le plus élevé mais tendance à la baisse. Baisse également pour les autres pays.

ETUDE ANNEXE

- On peut renouveler l'étude en cherchant les pays avec un PIB par habitant faible et un accès à internet faible mais qui pourrait augmenter, et où la concurrence est peut être moins élevée.

ETUDE ANNEXE : PIB/HAB ET INTERNET FAIBLE



Calcul du score identique aux régions:

- Normalisation des valeurs entre 0 et 1 en divisant par la valeur max

Le score final est la moyenne des valeurs des indicateurs normés (cf annexe)

Seuillage :

Internet users > **1 %**

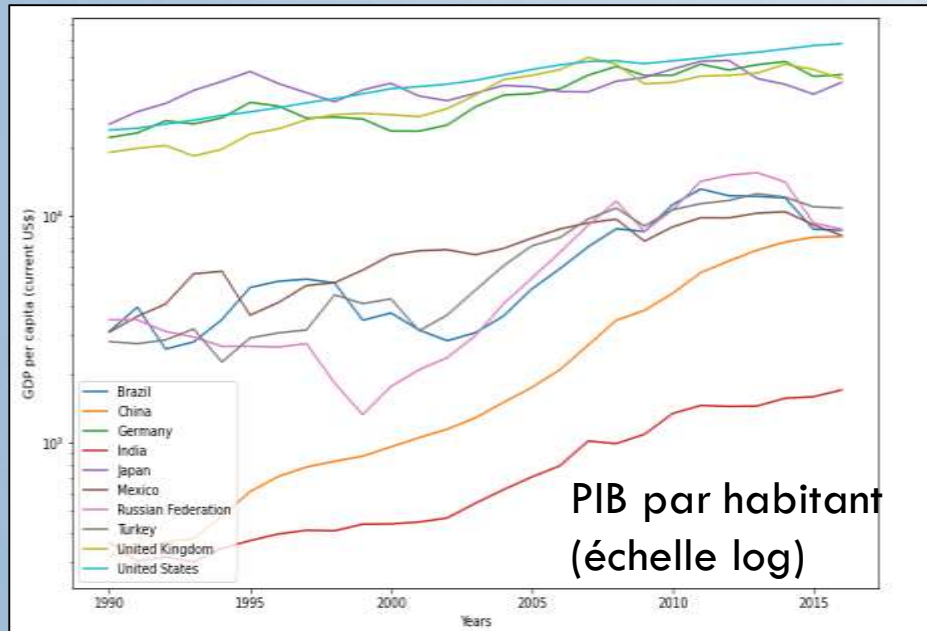
GDP per capita > **300 \$**

Pop age 15-24 > 7 000 000 personnes

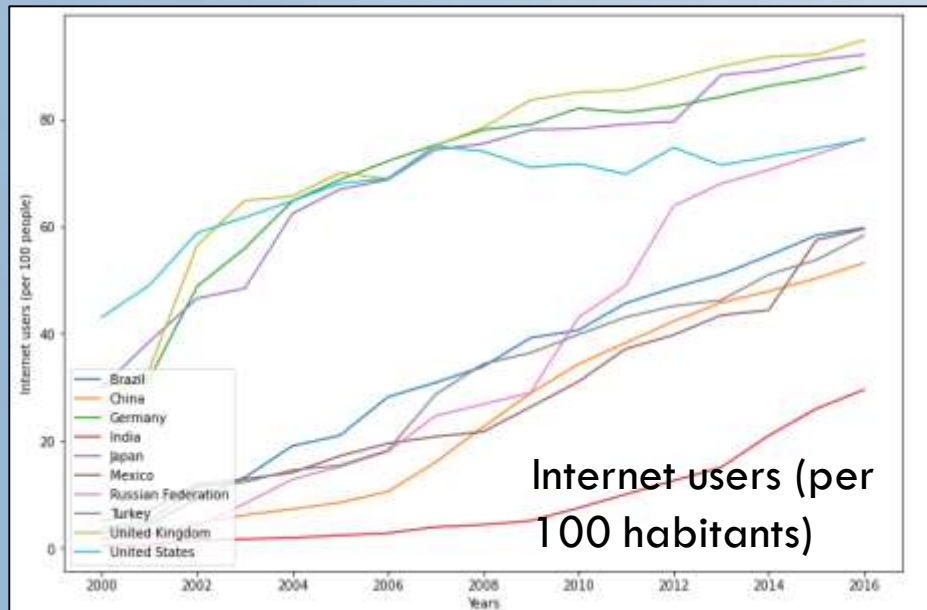
Pop age upper education > 2 000 000 personnes

Pop age tertiary education > 3 000 000 personnes

ETUDE ANNEXE : POTENTIEL D'ACHAT

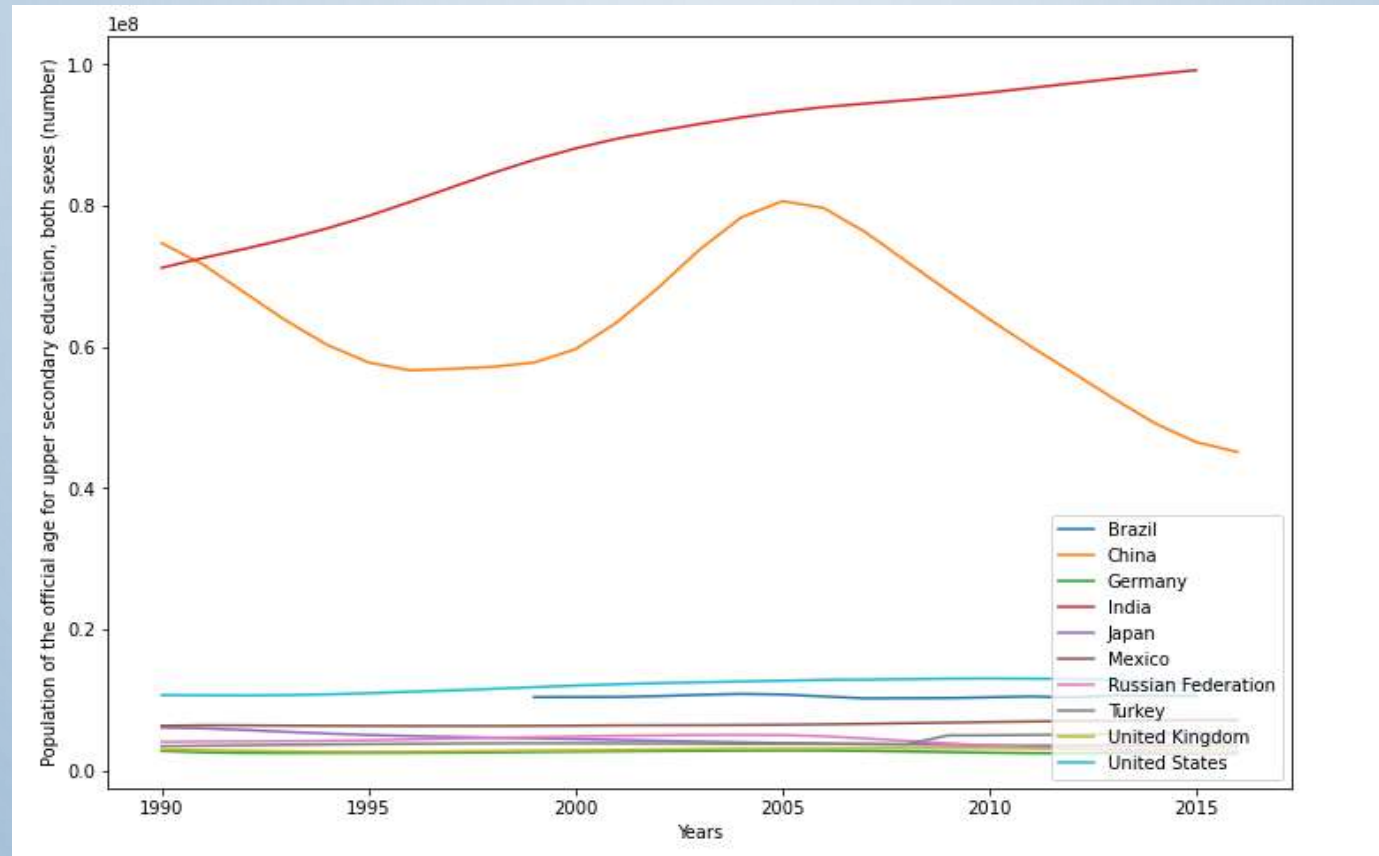


- PIB et accès à internet faible pour le moment, mais la tendance est à la hausse pour la Chine et l'Inde.
















ETUDE ANNEXE : VOLUME DE MARCHÉ

Population en âge pour le lycée



- Chine et Inde en tête, disposent d'un volume de marché le plus élevé. L'Inde dispose d'une tendance à la hausse contrairement à la Chine et aux autres pays.

CONCLUSION

Pays	PIB/hab		Internet User (%)		Pop âge 15-24		Pop lycéenne et université		Commentaires
	Valeur	Tendance	Valeur	Tendance	Valeur	Tendance	Valeur	Tendance	
USA	> 55 000\$		> 60		> 40 M	=	> 30 M	=	Très facile
Japon Allemagne UK	> 30 000\$	=	> 80		≈ 10 M		≈ 8 M		Facile
Inde	< 10 000\$		< 60		> 220 M		> 200 M		Proche avenir
Chine	< 10 000 \$		< 60		> 150 M		> 140 M		Proche avenir