



CentraleSupélec



## Projet 2 – 60 heures

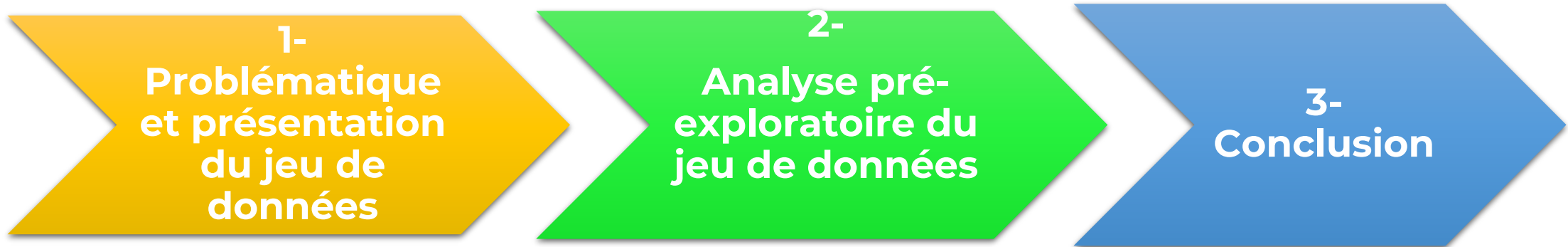
# Analysez des données de systèmes éducatifs

L'entreprise "academy" cherche à s'étendre à l'international. Dans ce premier projet, vous ferez des recommandations stratégiques à partir de données de systèmes éducatifs.

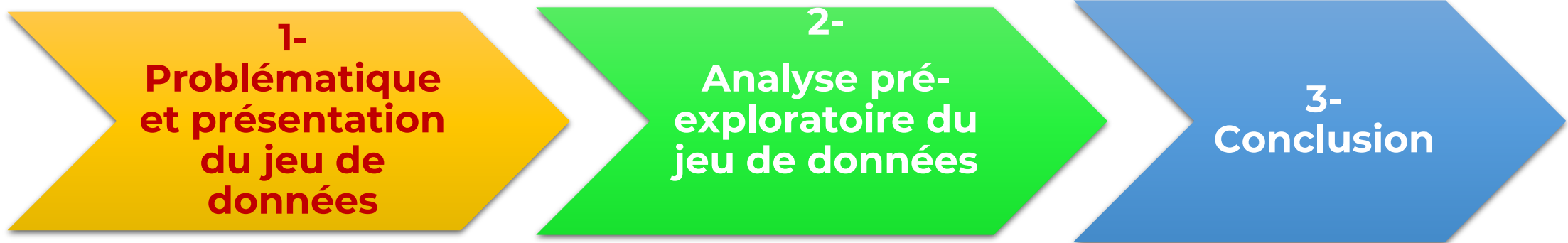
Réalisé par : Sabrina OUANNES

Date de soutenance : 14 déc 2021

# Sommaire



# Sommaire





# academy

## Présentation du projet

- *academy* est une start-up française de la EdTech qui propose des contenus de formation en ligne pour un public de niveau lycée et université.
- L'entreprise souhaite s'étendre à l'international.
- Les données sont fournies par la banque mondiale rubrique « EdStats »

<https://datacatalog.worldbank.org/dataset/education-statistics>

<http://datatopics.worldbank.org/education/>



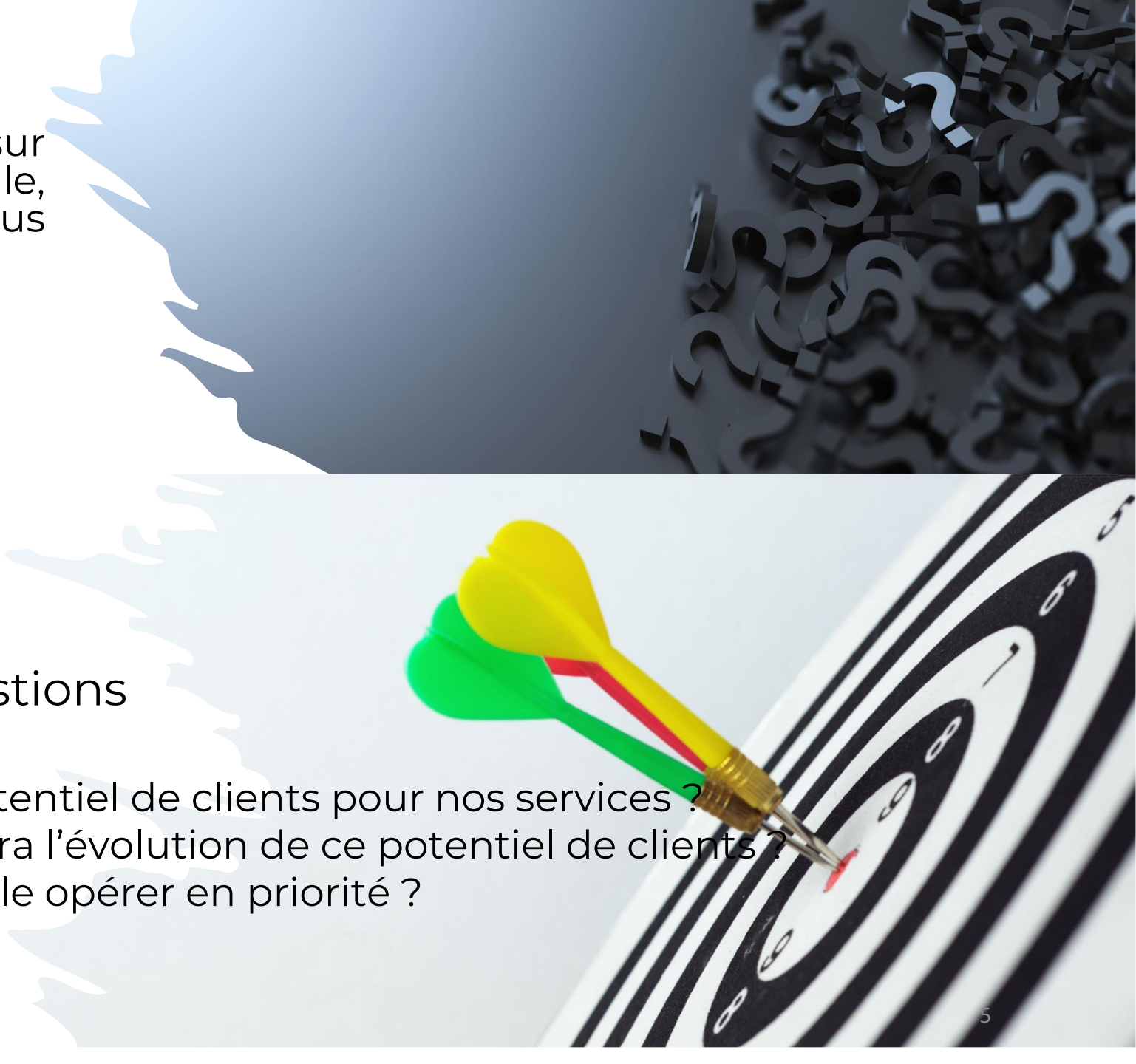
**BANQUE MONDIALE**

## Problématique :

Analyse exploratoire des données sur l'éducation de la banque mondiale, afin de cibler les destinations les plus intéressantes pour l'entreprise.

## Objectifs : Répondre à ces questions

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



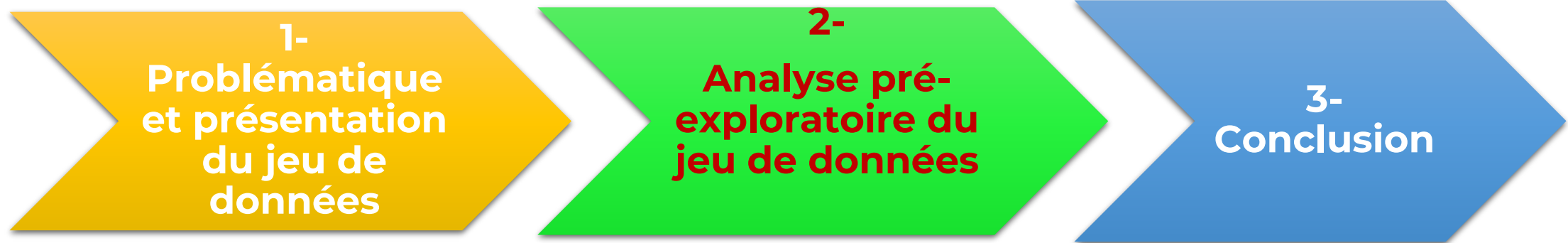


**BANQUE MONDIALE**

## Présentation Générale du Dataset

	Dimension	Taux remplissage moyen	Doublons	Description
<b>EdStatsCountry.csv</b>	(241, 32)	69.5%	0	Informations globales sur l'économie de chaque pays du monde et des zones géographiques
<b>EdStatsCountry-Series.csv</b>	(613, 4)	75.0%	0	Informations sur la source des données contenues dans EdStatsCountry
<b>EdStatsData.csv</b>	(886930, 70)	13.9%	0	Données d'évolution des indicateurs par pays et par région
<b>EdStatsFootNote.csv</b>	(643638, 5)	80.0%	0	Origine des différentes séries de données
<b>EdStatsSeries.csv</b>	(3665, 21)	28.3%	0	Informations sur les indicateurs socio économiques disponibles dans EdStatsData

# Sommaire



# Description du fichier EdStatsData

Dataset principal contenant la plupart des données qui sont utiles à notre analyse

70 colonnes

886930 lignes

	Country Name	Country Code	Indicator Name	Indicator Code	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
0	Arab World	ARB	Adjusted net enrolment rate, lower secondary, both sexes (%)	UIS.NERA.2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Arab World	ARB	Adjusted net enrolment rate, lower secondary, female (%)	UIS.NERA.2.F	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Arab World	ARB	Adjusted net enrolment rate, lower secondary, gender parity index (GPI)	UIS.NERA.2.GPI	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Arab World	ARB	Adjusted net enrolment rate, lower secondary, male (%)	UIS.NERA.2.M	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	Arab World	ARB	Adjusted net enrolment rate, primary, both sexes (%)	SE.PRM.TENR	54.82212	54.89414	56.20944	57.26711	57.99114	59.36554	60.99996	61.92268	62.69342	64.38319	65.61777

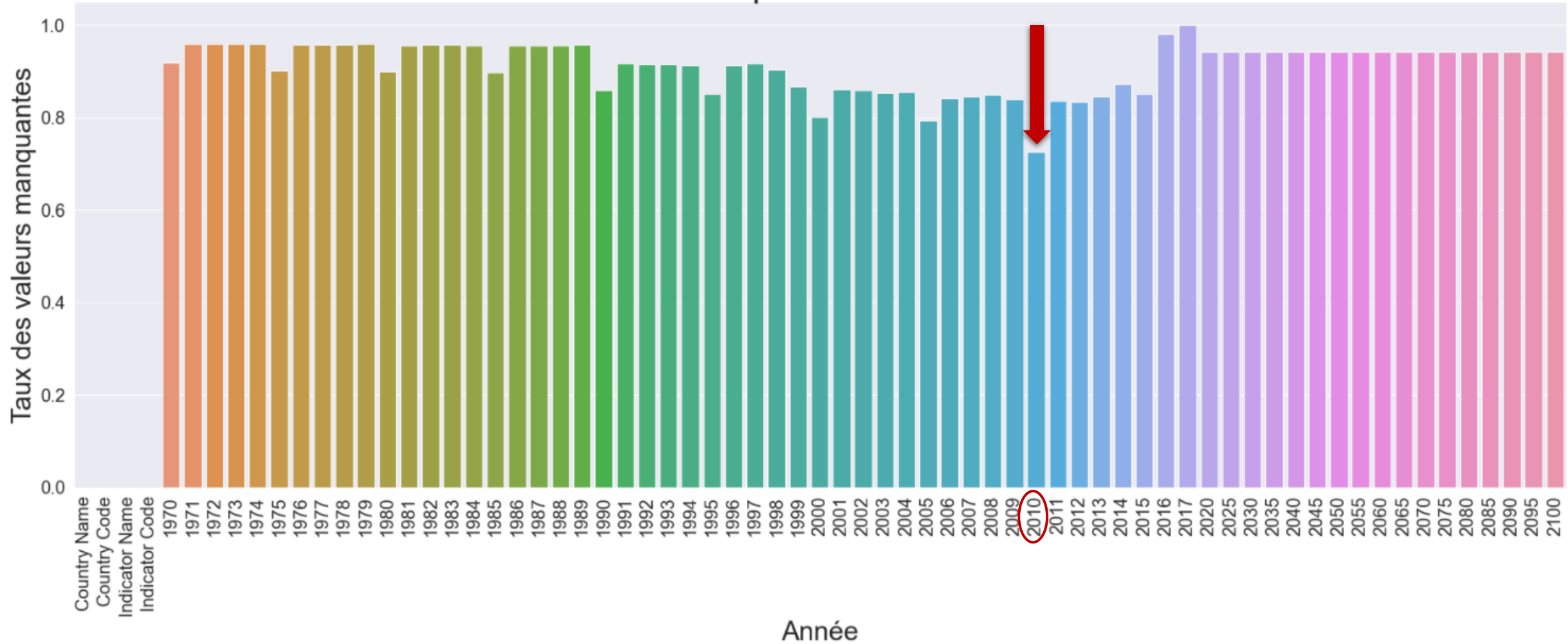
242 régions et pays  
3665 indicateurs



# Nettoyage des données

Taux global des valeurs manquantes : ~ 86%

Taux des valeurs manquantes en fonction des années



# Indicateurs sélectionnés



- Population, total
- Population growth (annual %)',
- Population of the official age for tertiary education, both sexes (number)
- Population of the official age for upper secondary education, both sexes (number)
- Population, ages 15-64, total
- Population, ages 15-24, total



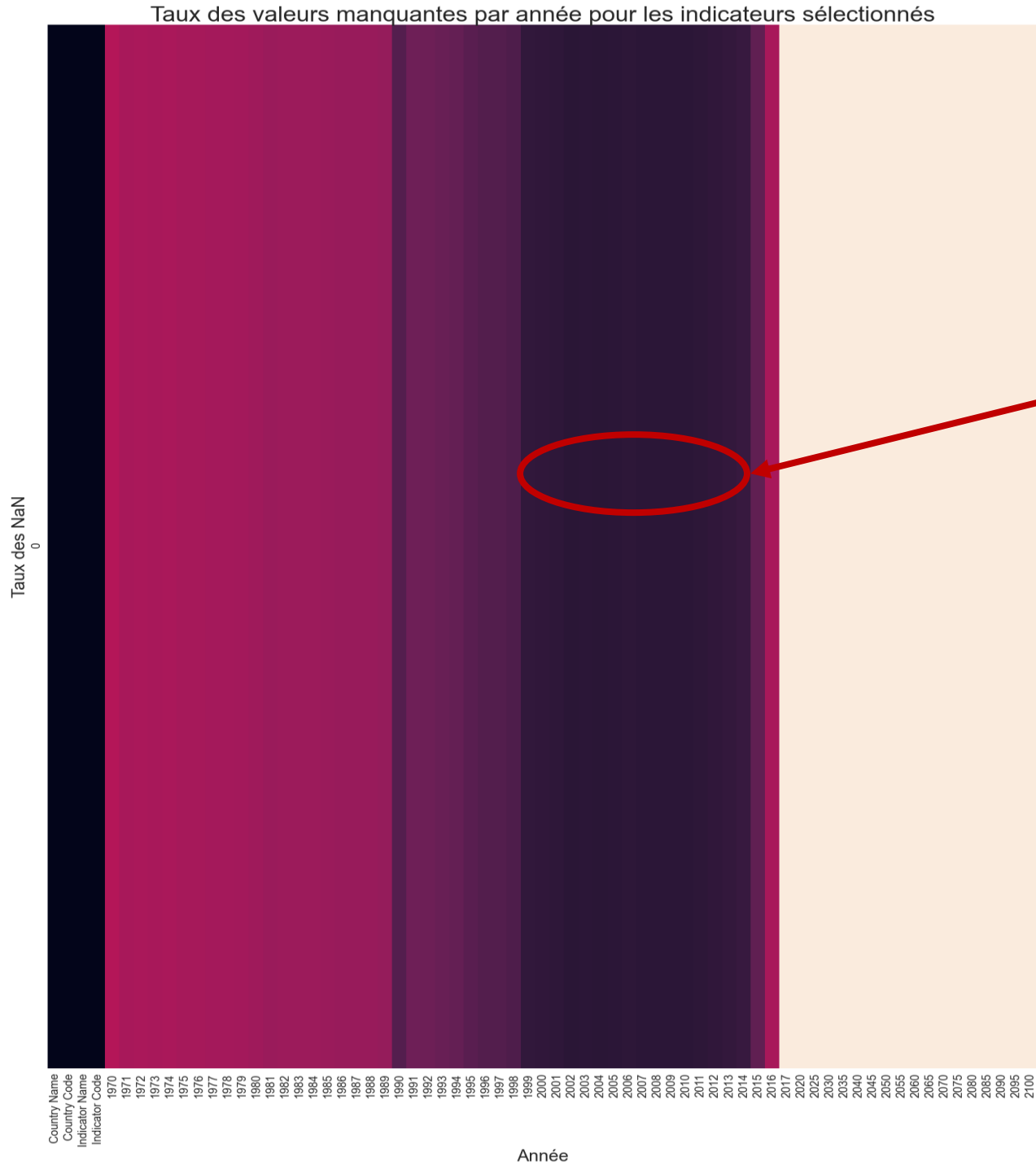
- Enrolment in tertiary education, all programmes, both sexes (number)
- Enrolment in upper secondary general, both sexes (number)



- GDP per capita (current US\$)
- GDP at market prices (current US\$)
- GNI (current US\$)



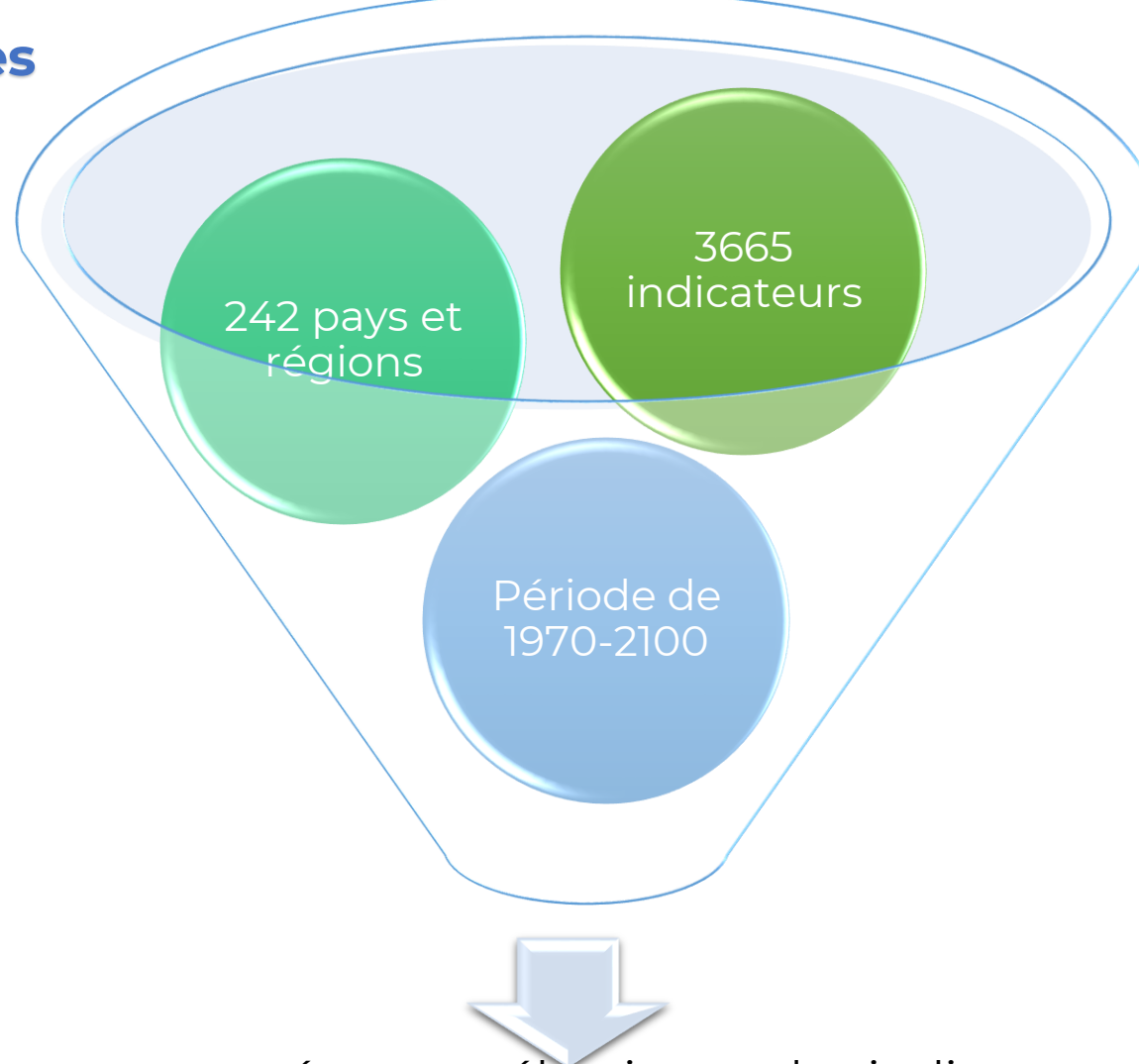
- Internet users (per 100 people)



## Nettoyage des données

***1999 à 2014 représente le plus de données pour étudier l'évolution de quelques indicateurs.***

## Nettoyage des données

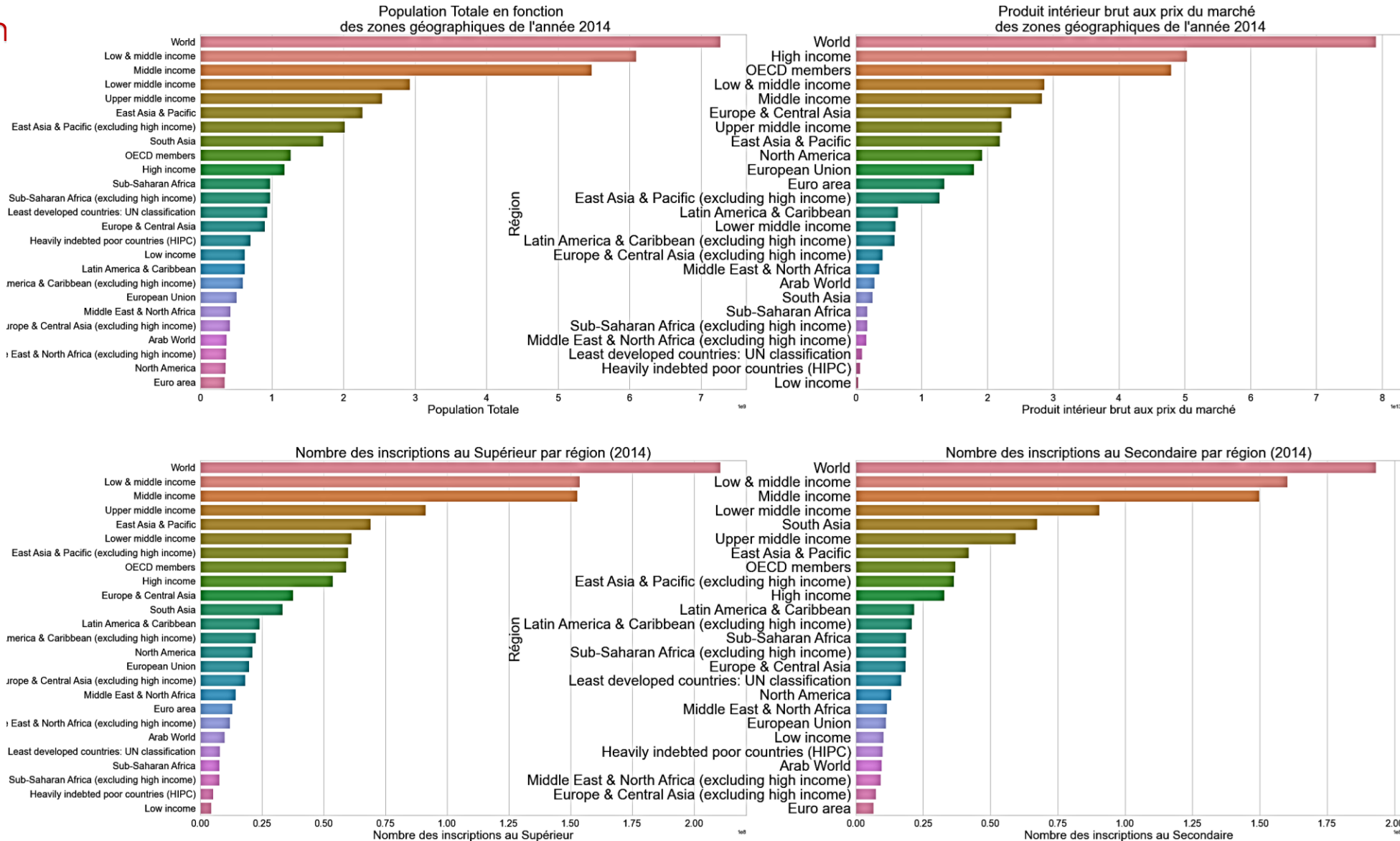


- ➡ 2010 comme année pour sélectionner les indicateurs pertinents.
- ➡ 2014 pour étudier les différents indicateurs sélectionnés.
- ➡ 1999 à 2014 représente le plus de données pour étudier l'évolution de quelques indicateurs.

# Analyse

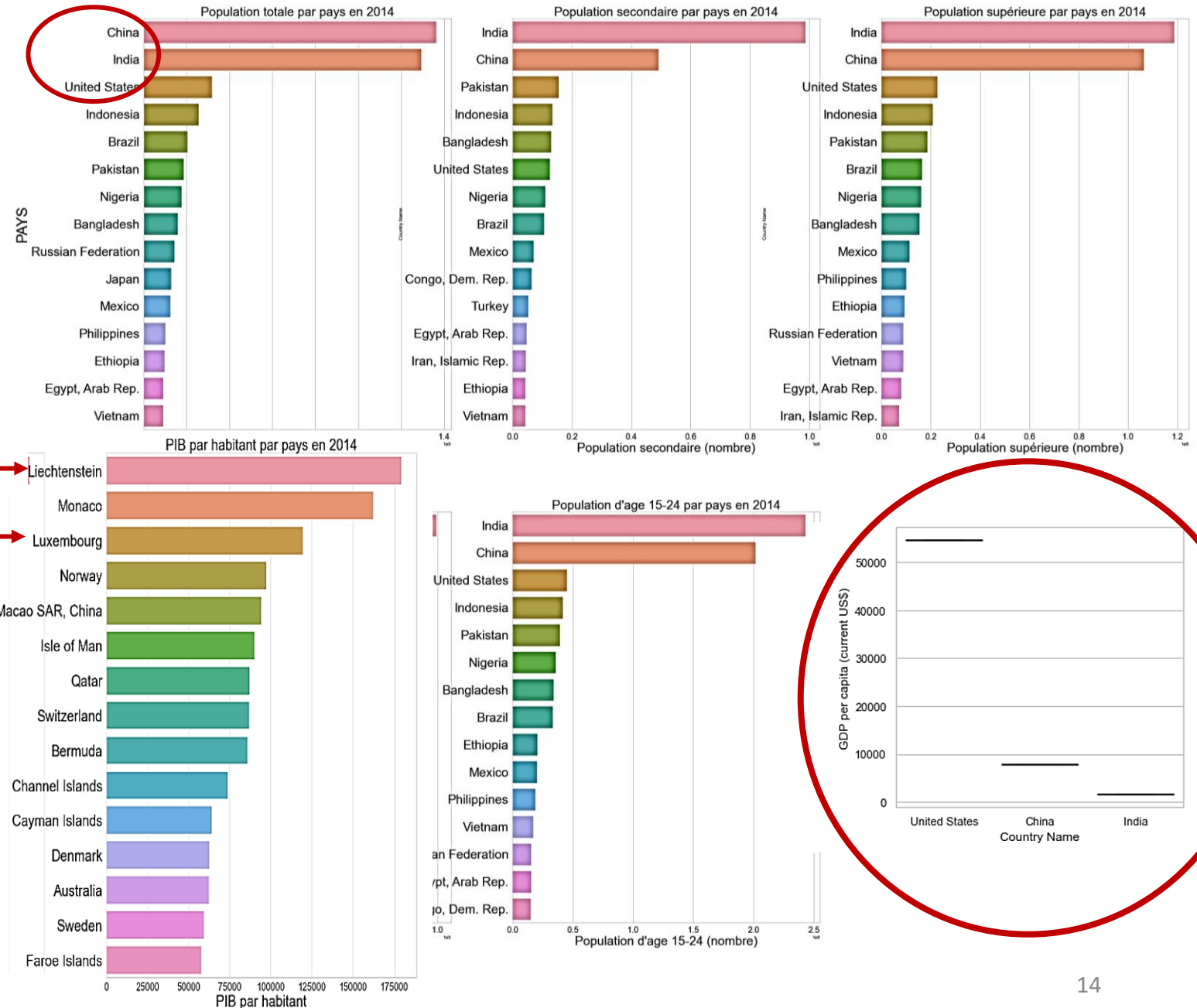
## Classement des région

- Bien que l'Europe n'est pas parmi les premiers en tenant compte de la population et du nombre des inscriptions au lycée et au supérieur, elle représente un facteur économique important.
- Il y a une forte corrélation entre la population et le nombre des inscriptions



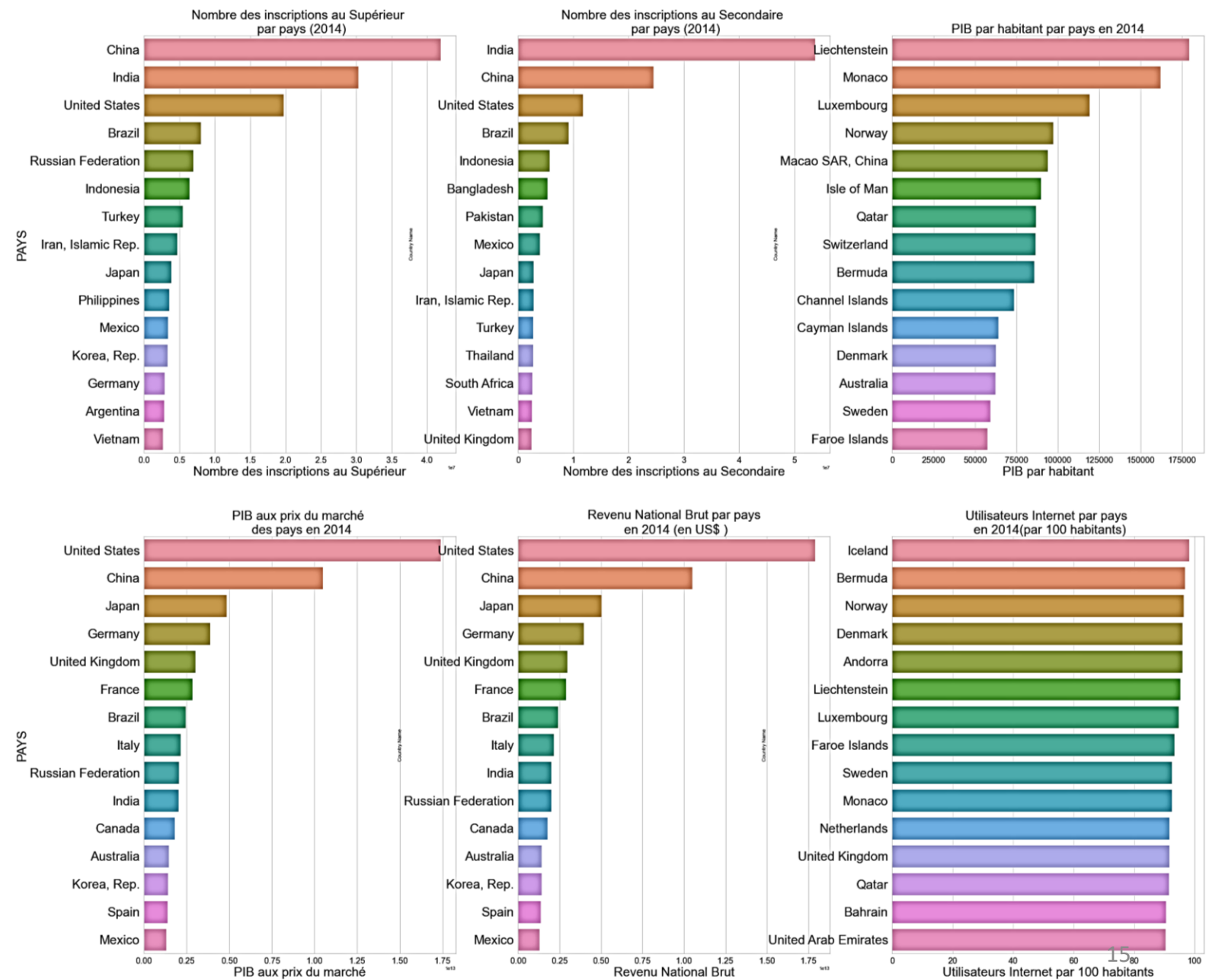
# Classement des pays par population

- On s'aperçoit que les pays les plus riches ne sont pas les pays les plus peuplés.
- Le Chine, l'Inde et l'USA présente le TOP 3 pour la population ciblée de tranche d'âge 15-24.
- L'inde et la Chine ont une très forte population en enseignement supérieur et secondaire mais leur PIB par habitant est très faible.
- Bien que Liechtenstein et Luxembourg sont riches mais sont des très petits pays. Ils ne représentent pas un gros potentiel.

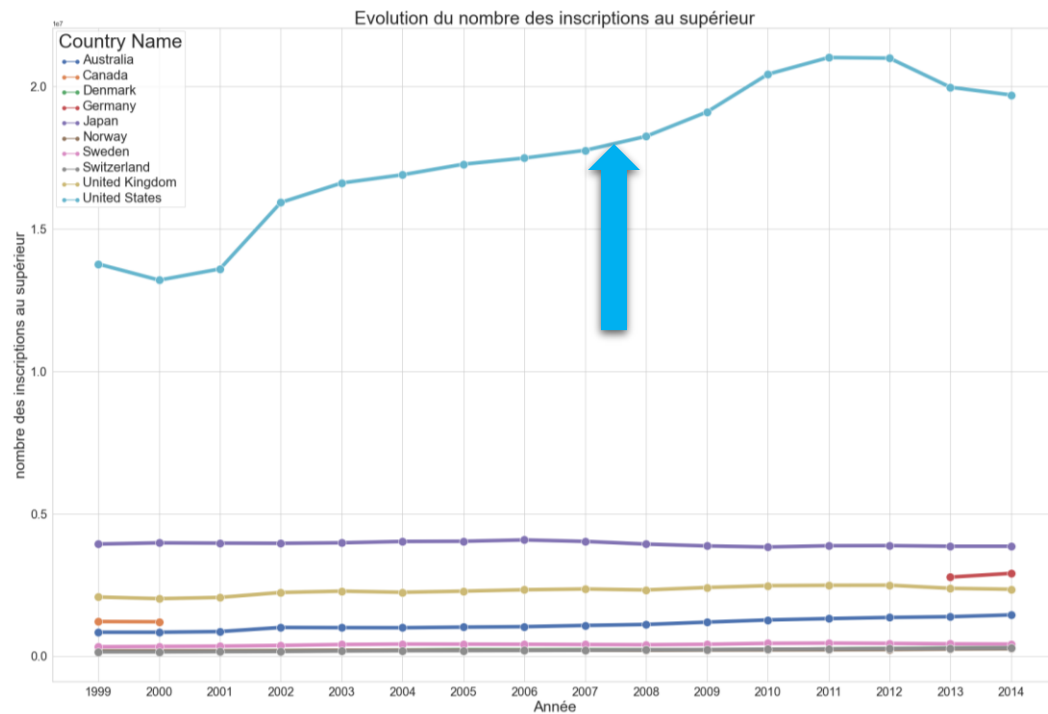


# Classement des pays Par Éducation, Internet, Économie

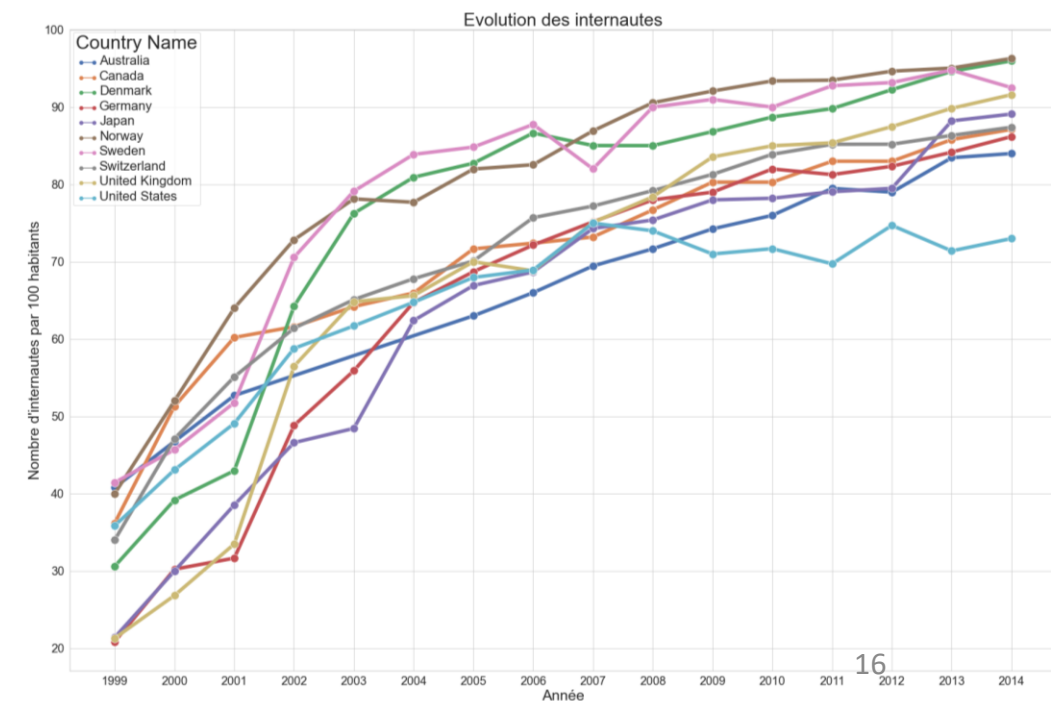
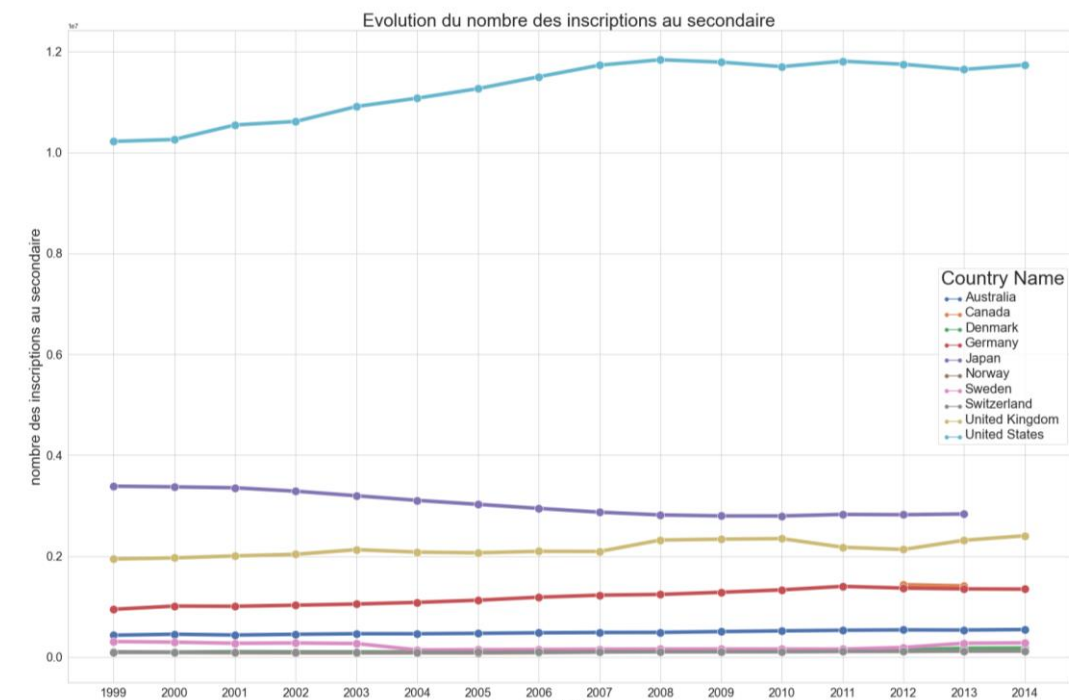
- Qatar et Norvège pourraient être un pays intéressants de point de vue PIB et nombre d'internautes avec une population des lycéens et étudiants plus ou moins très moyenne (245050).





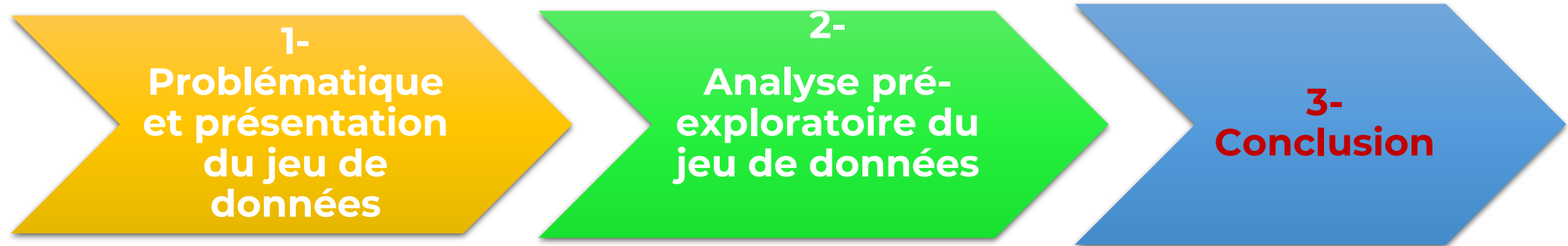


- Les Etats-Unis représente un fort potentiel pour l'entreprise avec un nombre d'inscriptions élevé et taux d'accès à internet qui a une belle courbe de croissance.
- On remarque que l'évolution du taux d'accès à internet suit une courbe croissante pour le TOP 10 de point de vue nombre d'internautes.





# Sommaire



# Conclusion

En regardant tous les paramètres, chacun à part, il est difficile de classer les pays à fort potentiel.

Pour une expansion à l'international, **academy** possède plusieurs possibilités :

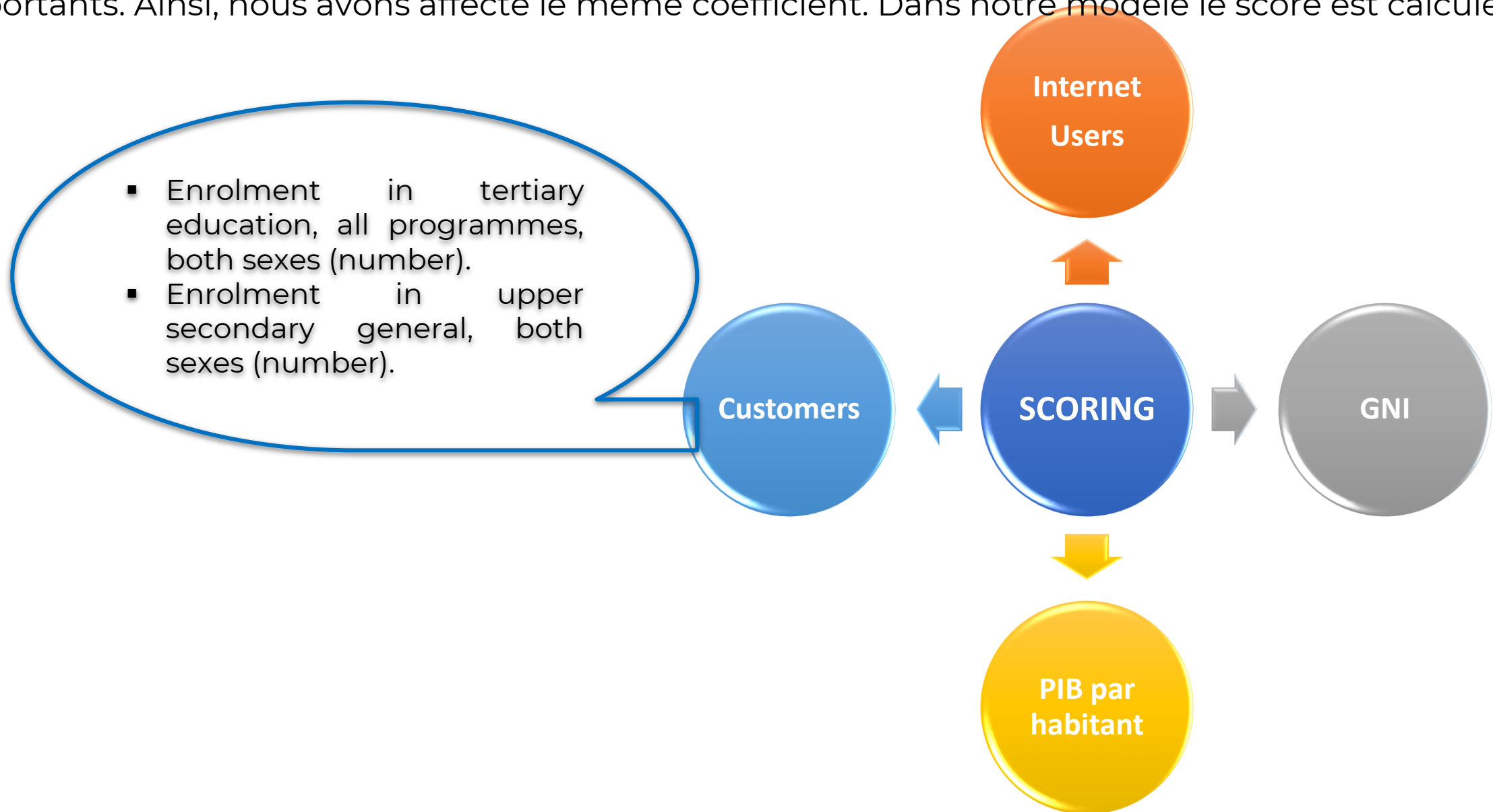
- S'implanter dans un pays où les infrastructures sont développées mais une population faible en enseignement supérieur et secondaire comme Qatar.
- S'implanter dans un pays où les infrastructures sont développées et avec un nombre d'inscriptions au supérieur et au secondaire intéressant comme les Etats-Unis et le Norvège.
- S'implanter dans un pays plus pauvre mais avec une population en enseignement supérieur forte comme l'Inde et la Chine à condition de chercher des Financements par l'état par exemple.



En essayant de prendre en considération différents types d'indicateurs (population, éducation, économie, infrastructure), on peut implémenter un modèle de scoring et classer les pays autrement.

## Modèle de Scoring

Pour implémenter un modèle de scoring, nous avons considéré que tous les indicateurs choisis sont importants. Ainsi, nous avons affecté le même coefficient. Dans notre modèle le score est calculé sur 100.



# Scoring

## TOP 10 Pays à fort potentiels

Rang	Country Name	SCORE
1	United States	82.0
2	Norway	49.8
3	Switzerland	45.1
4	Japan	44.5
5	United Kingdom	43.6
6	Germany	41.7
7	Denmark	40.5
8	Australia	39.7
9	Canada	39.6
10	Sweden	39.2
11	Brazil	38.8
12	France	38.5
13	Netherlands	37.5
14	Finland	34.5
15	Singapore	34.3
16	United Arab Emirates	34.2
17	Ireland	34.2
18	Belgium	34.2
19	Austria	33.8
20	Korea, Rep.	33.6



On peut faire d'autres modèles de scoring en donnant plus d'importance au indicateur économique ou la population ou autre. Ceci dépendra de la vision de l'entreprise.