



# Étude de marché - expansion internationale

## Synthèse

## Executive brief – la filière EdTech française: enjeux & perspectives

### Une croissance accélérée par la pandémie

**438 millions d'euros**

levée de fonds record

**1.3 milliards**

de chiffre d'affaires

**430**

entreprises spécialisées

**25%**

sont positionnées exclusivement  
sur le segment de  
l'enseignement secondaire &  
supérieur

**34**

créations d'entreprises

**10 000**

salariés

### Une filière française concentrée...

**70%**

du chiffre d'affaires de  
la filière est réalisé par  
les 20 premiers  
acteurs

**85%**

des acteurs de la  
filiale se positionnent  
sur la formation  
professionnelle

### ... et en voie d'internationalisation

**57%**

des entreprises du secteur ont une activité  
à l'international

**- de 20% du C.A.**

est réalisé à l'international pour 77% d'entre-elles

Source : [https://www.ey.com/fr\\_fr/strategy/la-filiere-edtech-francaise-l-annee-du-milliard](https://www.ey.com/fr_fr/strategy/la-filiere-edtech-francaise-l-annee-du-milliard) (2022)



Département Marketing Stratégique

## Executive brief – problématique & enjeux pour Academy

- ❖ Objectif : expansion à l'international
- ❖ Catalogue de produits : formations en ligne niveau lycée et université, à l'exclusion donc de la formation professionnelle ou des formations diplômantes, sur lesquelles des barrières à l'entrée réglementaires notamment pourraient s'appliquer
- ❖ Problématique : déterminer les pays
  - Présentant un bon potentiel à court et moyen terme, et qui seront à prioriser
  - Et un potentiel à plus long terme - évolution de la stratégie vers l'«objectif 2030»
- ❖ Challenge : trouver les indicateurs les plus pertinents pour informer ces décisions, et les évaluer historiquement et de façon prospective pour déterminer la liste des pays à cibler pour les différentes phases. Thèmes : télécoms & accès internet, revenu par habitant, taille et croissance de la population étudiante et taux de scolarisation secondaire et universitaire

# Sommaire

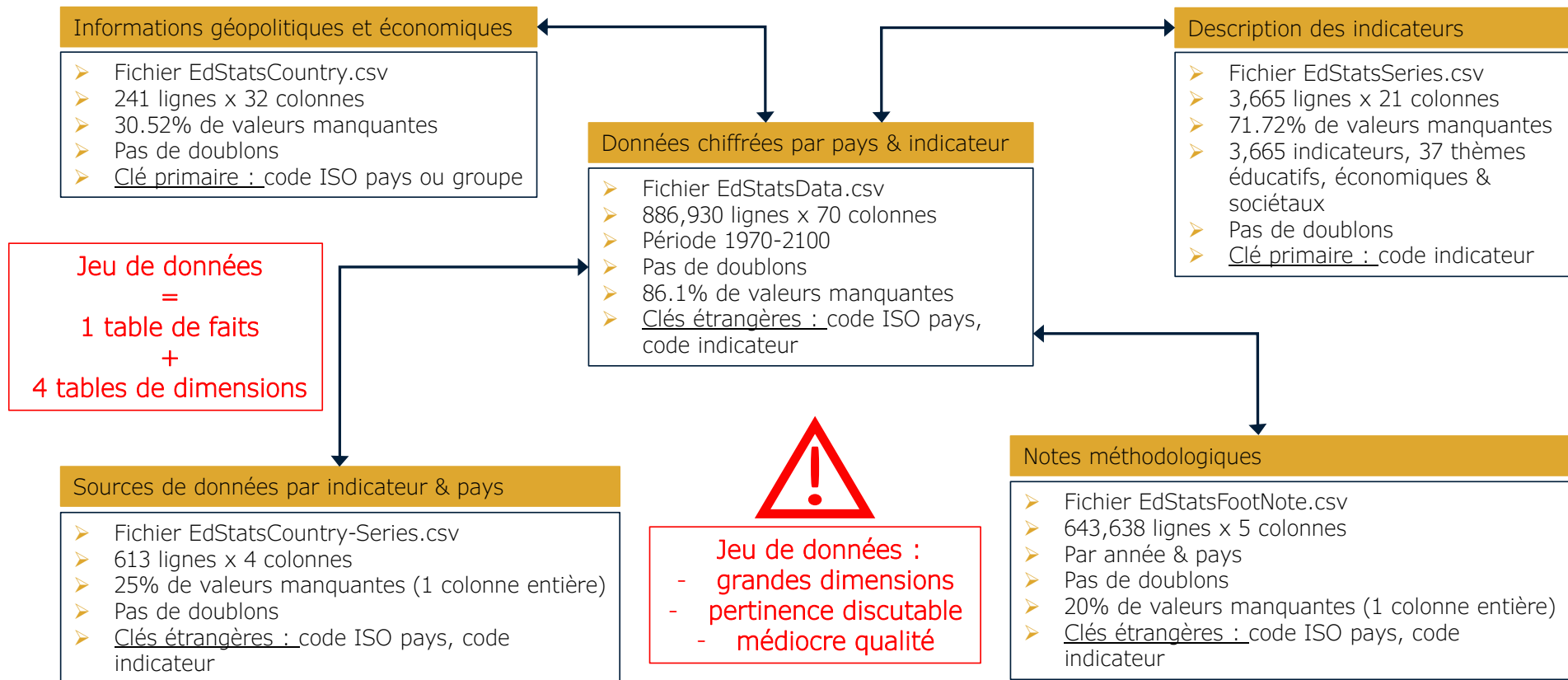
- 1 – Téléchargement, vérification & sélection des données
  - 1.1 – Import & description
  - 1.2 – Exploration & nettoyage
  - 1.3 – Sélection des indicateurs & définitions
- 2 – Analyses statistiques & graphiques
  - 2.1 – Indicateurs historiques
  - 2.2 – Indicateurs prospectifs
  - 2.3 – Imputations & classement des pays
- 3 – Choix des marchés-cibles & recommandations stratégiques
  - 3.1 – Priorisation des marchés-cibles
  - 3.2 – Objectif 2030

## Conclusion



# 1 – Téléchargement, vérification & sélection des données

## 1.1 – Import & description



# 1 – Téléchargement, vérification & sélection des données

## 1.2 – Exploration & nettoyage – opérations communes

### ❖ Import des fichiers plats

- `pd.read_csv()`

### ❖ Vérification des en-têtes

- `df.head()`

### ❖ Dimensions, contenu et statistiques descriptives

- `df.shape`
- `df.info()`
- `df.describe()`

### ❖ Distribution des valeurs manquantes

- `df.isna()`
- `msno.matrix(df)`
- `sns.barplot()`

### ❖ Doublons

- `df.duplicated()`

### ❖ Clés primaires et étrangères

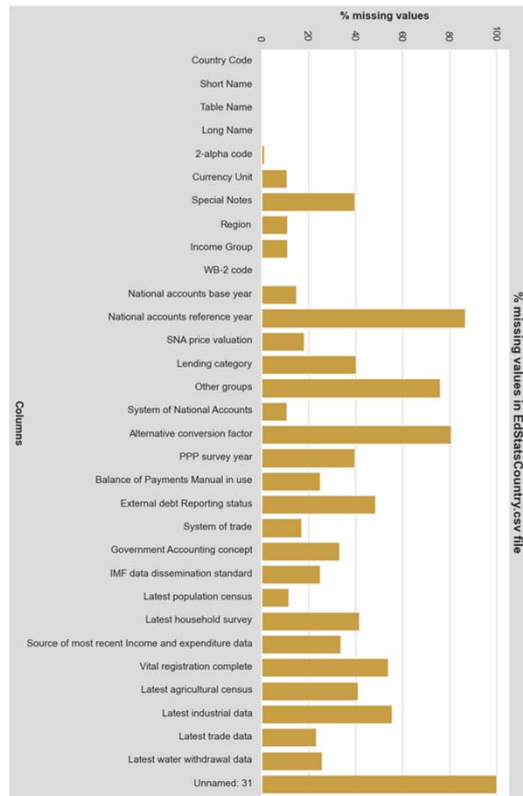
- jointures de tables

### ❖ Type casting & nommage des colonnes

- `.astype()`
- `.rename()`

# 1 – Téléchargement, vérification & sélection des données

## 1.2 – Exploration & nettoyage – table des pays



### ❖ Elimination :

- des valeurs manquantes
- des pays ou groupes non pertinents

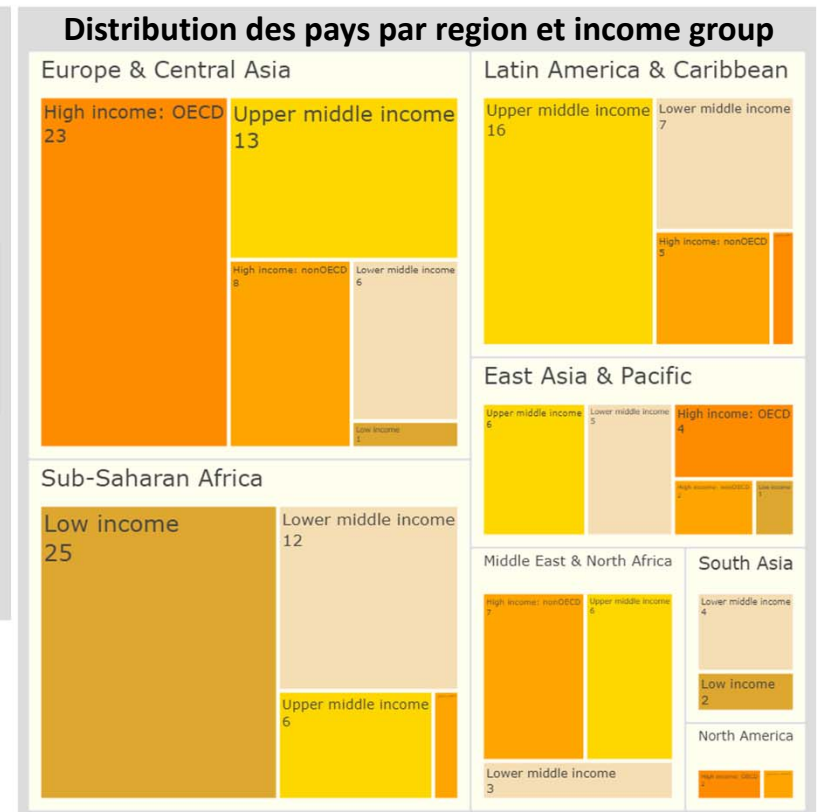
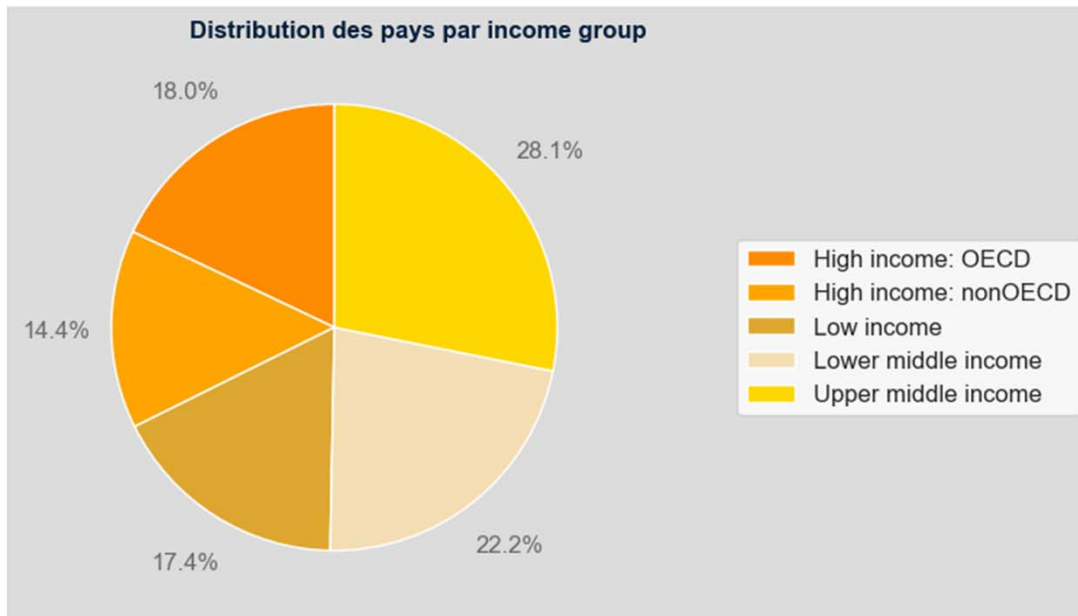
### ❖ Typage & renommage des colonnes

- 167 pays \* 4 colonnes

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 167 entries, 2 to 240
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   c_country_code   167 non-null    string
1   c_country_name   167 non-null    string
2   c_region         167 non-null    string
3   c_income_group   167 non-null    string
dtypes: string(4)
memory usage: 6.5 KB
```

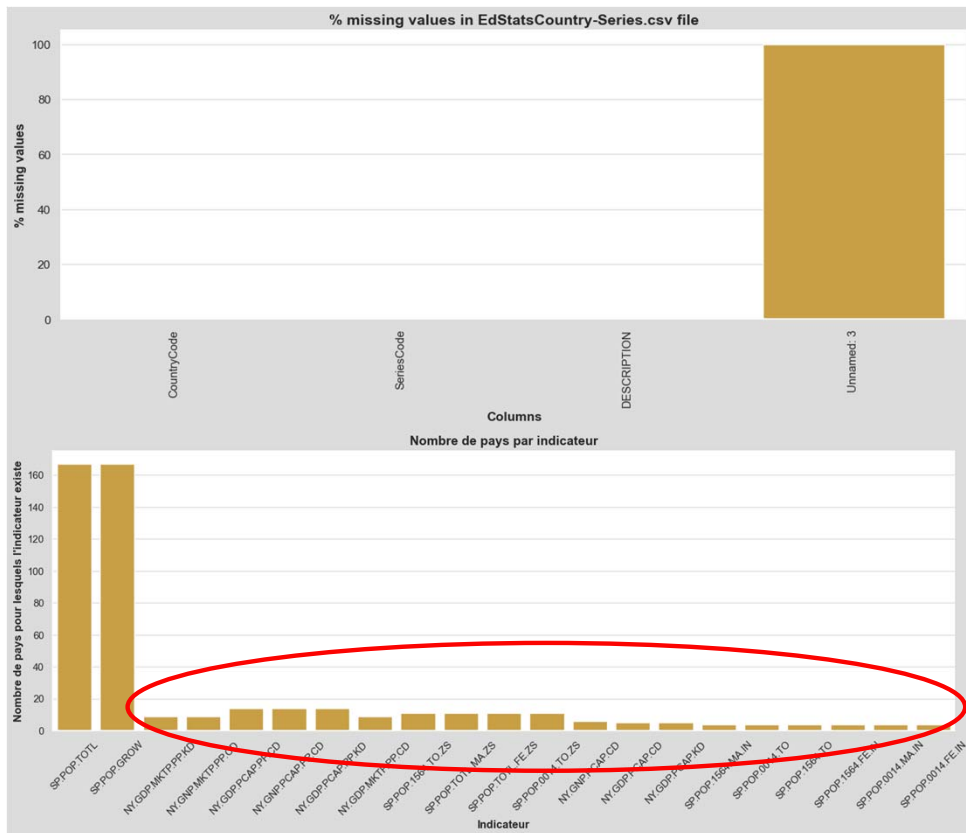
# 1 – Téléchargement, vérification & sélection des données

## 1.2 – Exploration & nettoyage – table des pays (suite)





## 1.2 – Exploration & nettoyage – sources de données par indicateur & pays



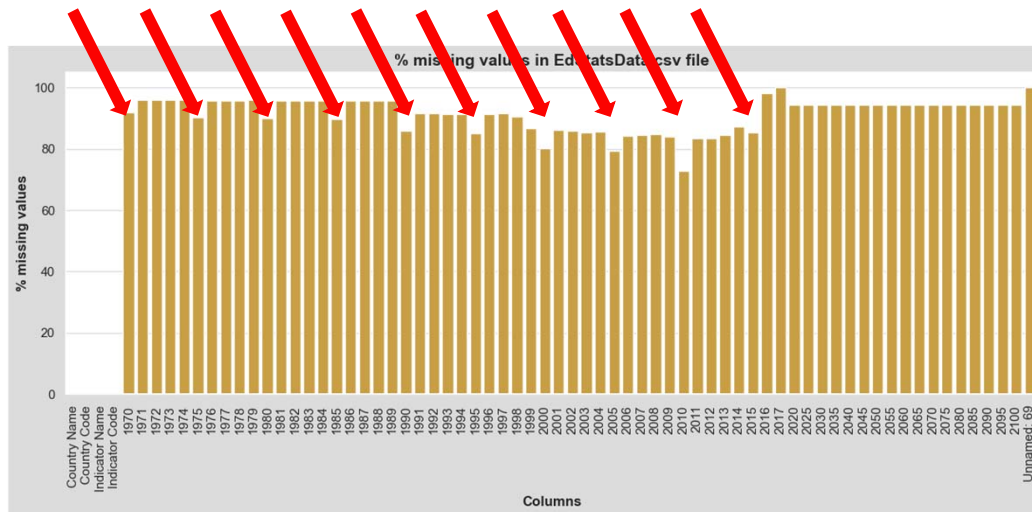
- ❖ Contient les **167 pays** retenus
- ❖ **21 indicateurs** de population et revenu
- ❖ **4 indicateurs retenus** :  
'SP.POP.GROW', 'SP.POP.TOTL',  
'NY.GNP.PCAP.PP.CD',  
'NY.GDP.PCAP.PP.CD'\*

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4 entries, 1664 to 2590
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---  -
0   cs_series_code         4 non-null      string
1   cs_topic               4 non-null      string
2   cs_indicator_name      4 non-null      string
3   cs_long_definition     4 non-null      string
4   cs_source              4 non-null      string
dtypes: string(5)
memory usage: 192.0 bytes
```

\* Voir section 2.1 ci-dessous pour les définitions

# 1 – Téléchargement, vérification & sélection des données

## 1.2 – Exploration & nettoyage – données chiffrées par pays & indicateur

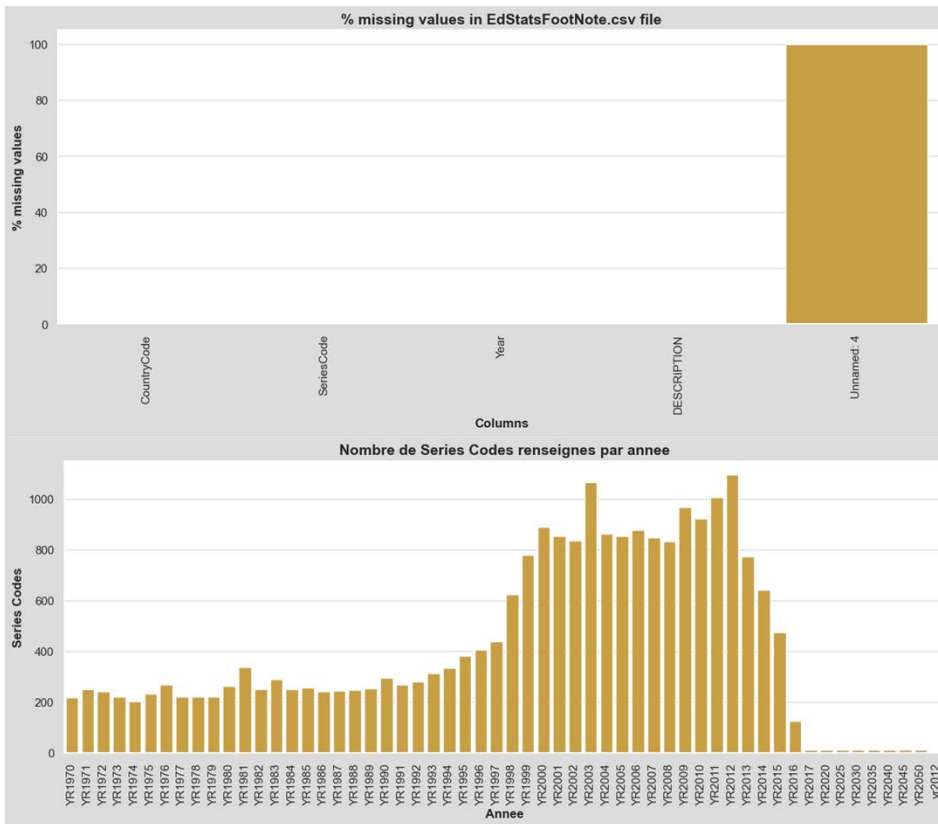


- ❖ Les 167 pays sélectionnés sont présents
- ❖ 3,665 indicateurs
- ❖ Plage de données 1970-2100 avec un "pic" de données tous les 5 ans
- ❖ Données historiques & prévisionnelles

<class 'pandas.core.frame.DataFrame'>  
Int64Index: 612055 entries, 95290 to 886929  
Data columns (total 31 columns):  
# Column Non-Null Count Dtype  
---  
0 d\_country\_name 612055 non-null string  
1 d\_country\_code 612055 non-null string  
2 d\_indicator\_name 612055 non-null string  
3 d\_indicator\_code 612055 non-null string  
4 1970 60948 non-null float64  
...  
30 2100 44352 non-null float64  
dtypes: float64(27), string(4)  
memory usage: 149.4 MB

# 1 – Téléchargement, vérification & sélection des données

## 1.2 – Exploration & nettoyage – table de notes méthodologiques



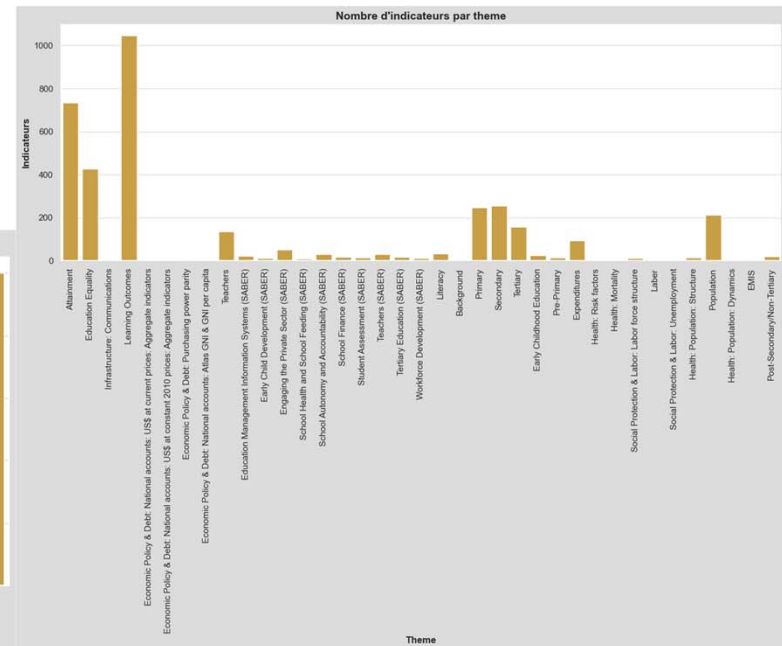
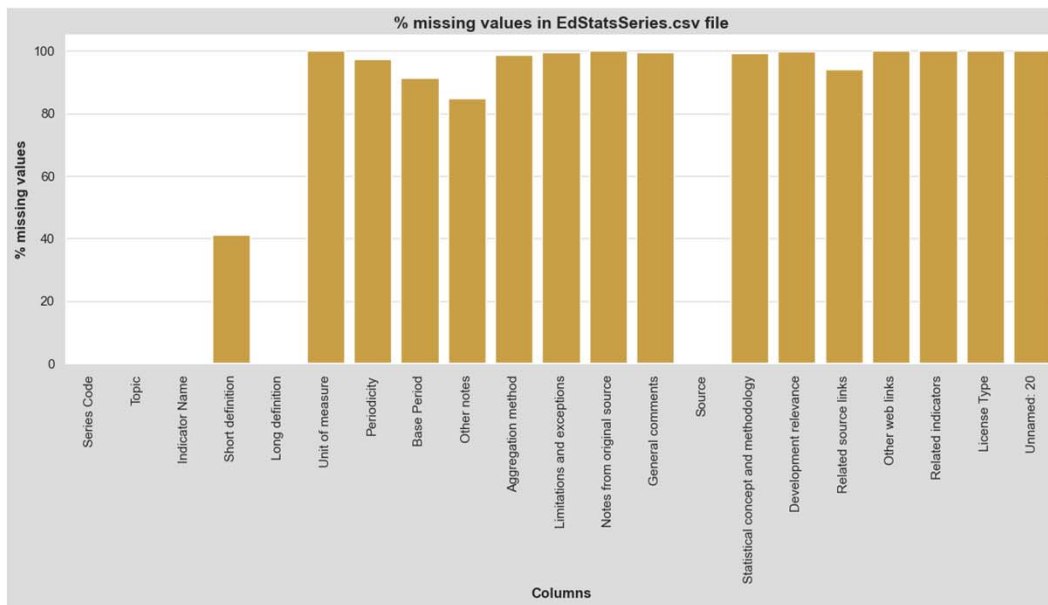
- ❖ Les 167 pays sélectionnés sont présents
- ❖ Grande variabilité du nombre d'indicateurs présents pour chaque année

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 445607 entries, 3288 to 643637
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   f_country_code  445607 non-null  string
1   f_series_code   445607 non-null  string
2   f_year          445607 non-null  int32
3   f_description   445607 non-null  string
dtypes: int32(1), string(3)
memory usage: 15.3 MB
```

# 1 – Téléchargement, vérification & sélection des données

## 1.3 – Sélection des indicateurs – table EdStatsSeries.csv

- ❖ 21 colonnes, 3,665 indicateurs et beaucoup de données manquantes (>70%)



- ❖ Indicateurs très inégalement répartis entre 37 catégories

# 1 – Téléchargement, vérification & sélection des données

## 1.3 – Sélection des indicateurs – table EdStatsSeries.csv (suite)

### ❖ Stratégie de sélection :

1. Sélection des thèmes pertinents : Economic Policy & Debt: National accounts: Atlas GNI & GNI per capita, Infrastructure: Communications, Tertiary, Population, Secondary, Education Equality, Attainment, Learning Outcomes, Literacy
2. Évaluation manuelle des indicateurs disponibles dans chaque catégorie, parfois avec des mots-clés

### ❖ Résultat : 13 indicateurs historiques...

	s_topic	indicators
0	Attainment	[BAR.POP.1519, BAR.POP.2024, BAR.SCHL.1519, BAR.SCHL.2024]
1	Economic Policy & Debt: National accounts: Atlas GNI & GNI per capita	[NY.GNP.PCAP.CD]
2	Infrastructure: Communications	[IT.CMP.PCMP.P2, IT.NET.USER.P2]
3	Literacy	[SE.ADT.1524.LT.ZS]
4	Population	[SP.POP.1524.TO.UN]
5	Secondary	[SE.SEC.ENRL, SE.SEC.ENRR]
6	Tertiary	[SE.TER.ENRL, SE.TER.ENRR]

### ❖ ...et 7 indicateurs prévisionnels sélectionnés

	s_topic	indicators
0	Attainment	[PRJ.POP.1519.2.MF, PRJ.POP.1519.3.MF, PRJ.POP.1519.NED.MF, PRJ.POP.2024.2.MF, PRJ.POP.2024.3.MF, PRJ.POP.2024.4.MF, PRJ.POP.2024.NED.MF]

## 2 – Analyses statistiques & graphiques

### 2.1 – Indicateurs historiques - définitions

s_series_code	s_topic	s_indicator_name	s_source
BAR.POP.1519	Attainment	Barro-Lee: Population in thousands, age 15-19, total	Robert J. Barro and Jong-Wha Lee: <a href="http://www.barrolee.com/">http://www.barrolee.com/</a>
BAR.POP.2024	Attainment	Barro-Lee: Population in thousands, age 20-24, total	Robert J. Barro and Jong-Wha Lee: <a href="http://www.barrolee.com/">http://www.barrolee.com/</a>
BAR.SCHL.1519	Attainment	Barro-Lee: Average years of total schooling, age 15-19, total	Robert J. Barro and Jong-Wha Lee: <a href="http://www.barrolee.com/">http://www.barrolee.com/</a>
BAR.SCHL.2024	Attainment	Barro-Lee: Average years of total schooling, age 20-24, total	Robert J. Barro and Jong-Wha Lee: <a href="http://www.barrolee.com/">http://www.barrolee.com/</a>
IT.CMP.PCMP.P2	Infrastructure: Communications	Personal computers (per 100 people)	International Telecommunication Union, World Telecommunication/ICT Development Report and database.
IT.NET.USER.P2	Infrastructure: Communications	Internet users (per 100 people)	International Telecommunication Union, World Telecommunication/ICT Development Report and database.
NY.GNP.PCAP.CD	Economic Policy & Debt: National accounts: Atlas GNI & GNI per capita	GNI per capita, Atlas method (current US\$)	World Bank national accounts data, and OECD National Accounts data files.
SE.ADT.1524.LT.ZS	Literacy	Youth literacy rate, population 15-24 years, both sexes (%)	UNESCO Institute for Statistics
SE.SEC.ENRL	Secondary	Enrolment in secondary education, both sexes (number)	UNESCO Institute for Statistics
SE.SEC.ENRR	Secondary	Gross enrolment ratio, secondary, both sexes (%)	UNESCO Institute for Statistics
SE.TER.ENRL	Tertiary	Enrolment in tertiary education, all programmes, both sexes (number)	UNESCO Institute for Statistics
SE.TER.ENRR	Tertiary	Gross enrolment ratio, tertiary, both sexes (%)	UNESCO Institute for Statistics
SP.POP.1524.TO.UN	Population	Population, ages 15-24, total	UNESCO Institute for Statistics (Derived)

N.B. : la méthodologie de calcul de l'indicateur SE.SEC.ENRR peut conduire à un taux supérieur à 100%.  
 Les indicateurs historiques ci-dessus s'ajoutent à la liste 'SP.POP.GROW', 'SP.POP.TOTL', 'NY.GNP.PCAP.PP.CD' et 'NY.GDP.PCAP.PP.CD' déjà définie à la section 1.2 plus haut.  
 En pratique, le taux de remplissage de ces indicateurs s'avèrera très inégal voire insuffisant et amènera à des restrictions supplémentaires de la liste au fur et à mesure de l'analyse.

## 2 – Analyses statistiques & graphiques

### 2.1 – Indicateurs historiques – définitions (suite)

cs_series_code	cs_topic	cs_indicator_name	cs_source
NY.GDP.PCAP.PP.CD	Economic Policy & Debt: Purchasing power parity	GDP per capita, PPP (current international \$)	World Bank, International Comparison Program database.
NY.GNP.PCAP.PP.CD	Economic Policy & Debt: Purchasing power parity	GNI per capita, PPP (current international \$)	World Bank, International Comparison Program database.
SP.POP.GROW	Health: Population: Dynamics	Population growth (annual %)	Derived from total population. Population source: (1) United Nations Population Division. World Population Prospects, (2) Census reports and other statistical publications from national statistical offices, (3) Eurostat: Demographic Statistics, (4) United Nations Statistical Division. Population and Vital Statistics Reprot (various years), (5) U.S. Census Bureau: International Database, and (6) Secretariat of the Pacific Community: Statistics and Demography Programme.
SP.POP.TOTL	Health: Population: Structure	Population, total	(1) United Nations Population Division. World Population Prospects, (2) Census reports and other statistical publications from national statistical offices, (3) Eurostat: Demographic Statistics, (4) United Nations Statistical Division. Population and Vital Statistics Report (various years), (5) U.S. Census Bureau: International Database, and (6) Secretariat of the Pacific Community: Statistics and Demography Programme.

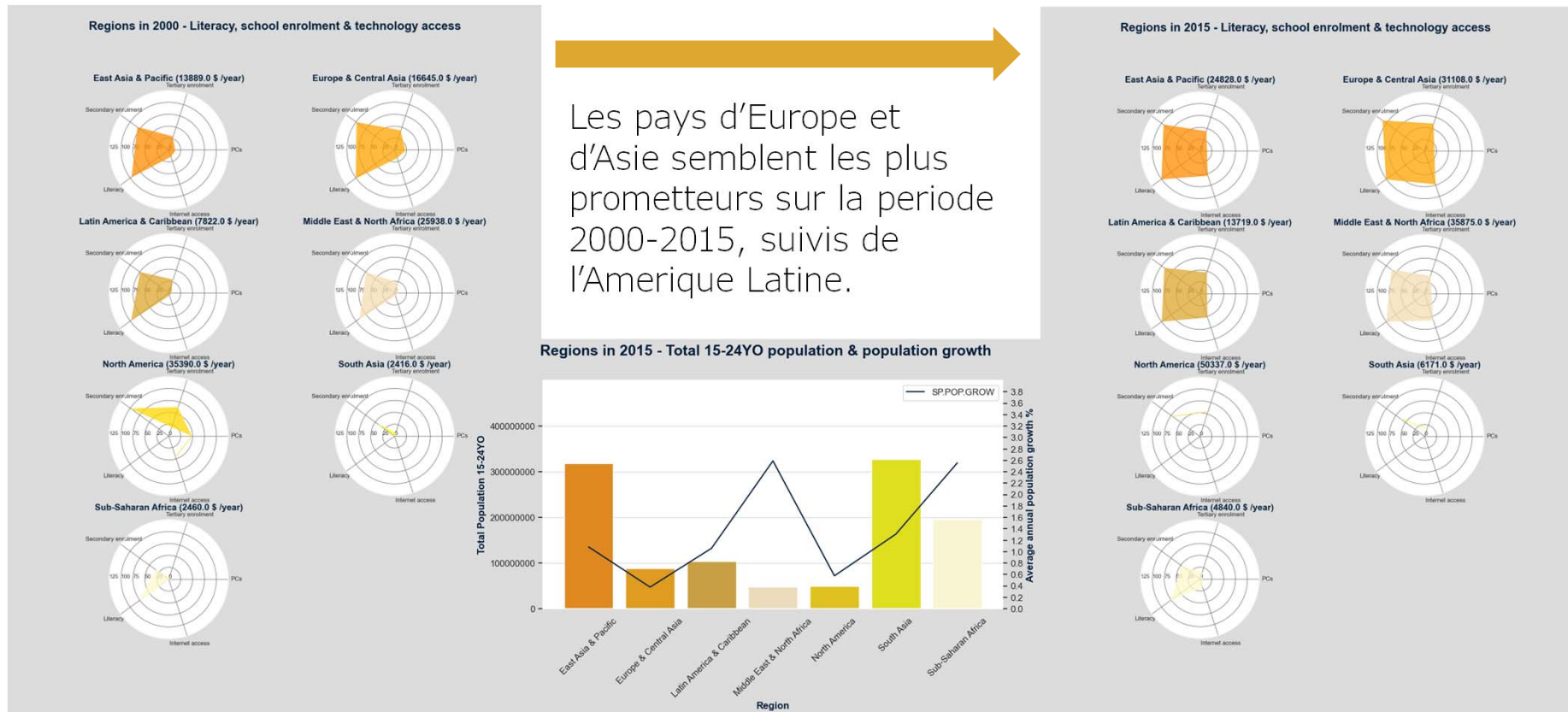
N.B. : Ces indicateurs sont ceux définis à la section 1.2 plus haut.

En pratique, le taux de remplissage de ces indicateurs s'avèrera très inégal voire insuffisant et amènera à des restrictions supplémentaires de la liste au fur et à mesure de l'analyse.



## 2 – Analyses statistiques & graphiques

### 2.1 – Indicateurs historiques – analyses graphiques





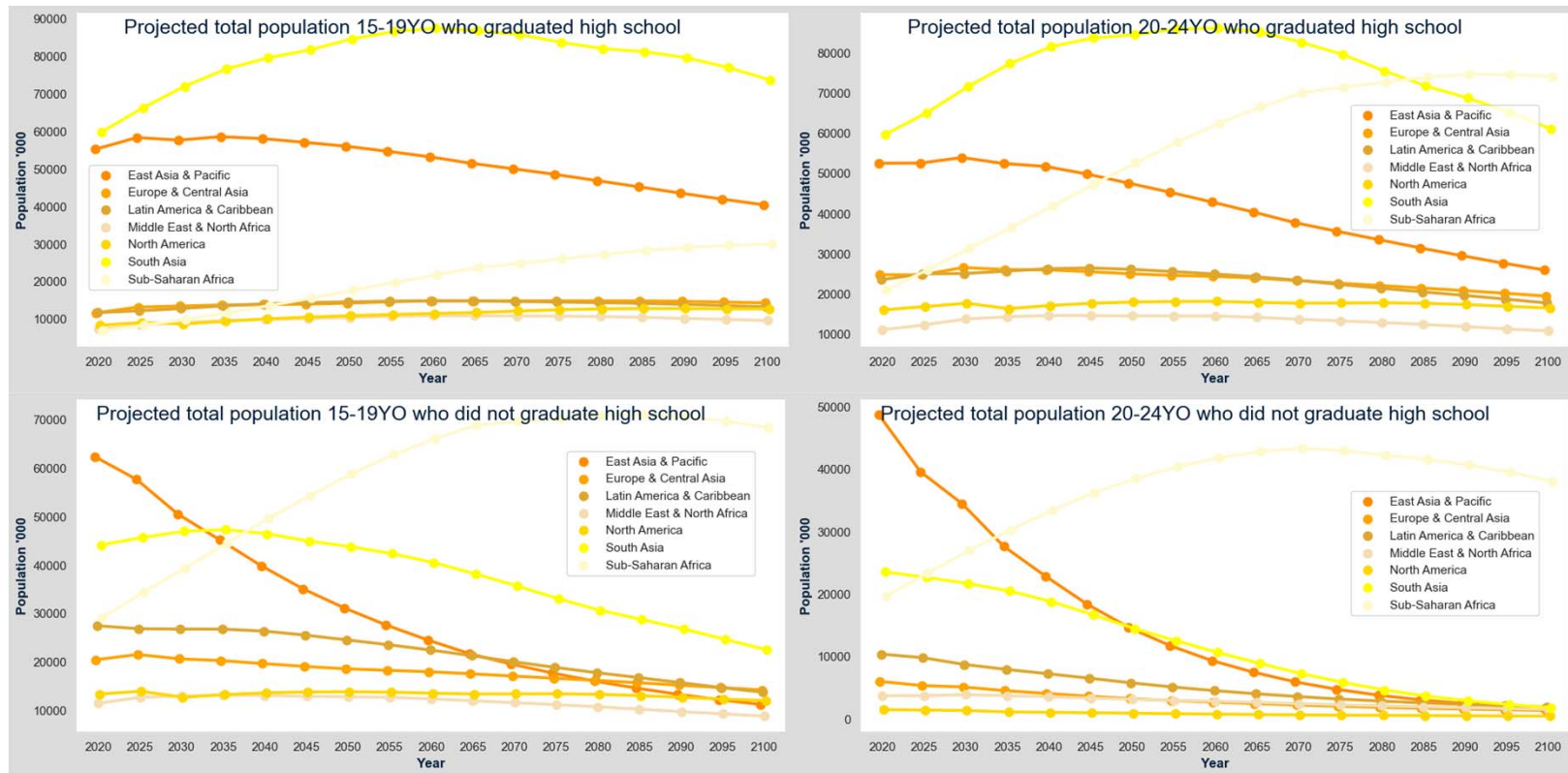
## 2 – Analyses statistiques & graphiques

### 2.2 – Indicateurs prospectifs - définitions

s_series_code	s_topic	s_indicator_name	s_source
PRJ.POP.1519.2.MF	Attainment	Projection: Population age 15-19 in thousands by highest level of educational attainment. Lower Secondary. Total	Wittgenstein Centre for Demography and Global Human Capital: <a href="http://www.oeaw.ac.at/vid/dataexplorer/">http://www.oeaw.ac.at/vid/dataexplorer/</a>
PRJ.POP.1519.3.MF	Attainment	Projection: Population age 15-19 in thousands by highest level of educational attainment. Upper Secondary. Total	
PRJ.POP.1519.NED.MF	Attainment	Projection: Population age 15-19 in thousands by highest level of educational attainment. No Education. Total	
PRJ.POP.2024.2.MF	Attainment	Projection: Population age 20-24 in thousands by highest level of educational attainment. Lower Secondary. Total	
PRJ.POP.2024.3.MF	Attainment	Projection: Population age 20-24 in thousands by highest level of educational attainment. Upper Secondary. Total	
PRJ.POP.2024.4.MF	Attainment	Projection: Population age 20-24 in thousands by highest level of educational attainment. Post Secondary. Total	
PRJ.POP.2024.NED.MF	Attainment	Projection: Population age 20-24 in thousands by highest level of educational attainment. No Education. Total	

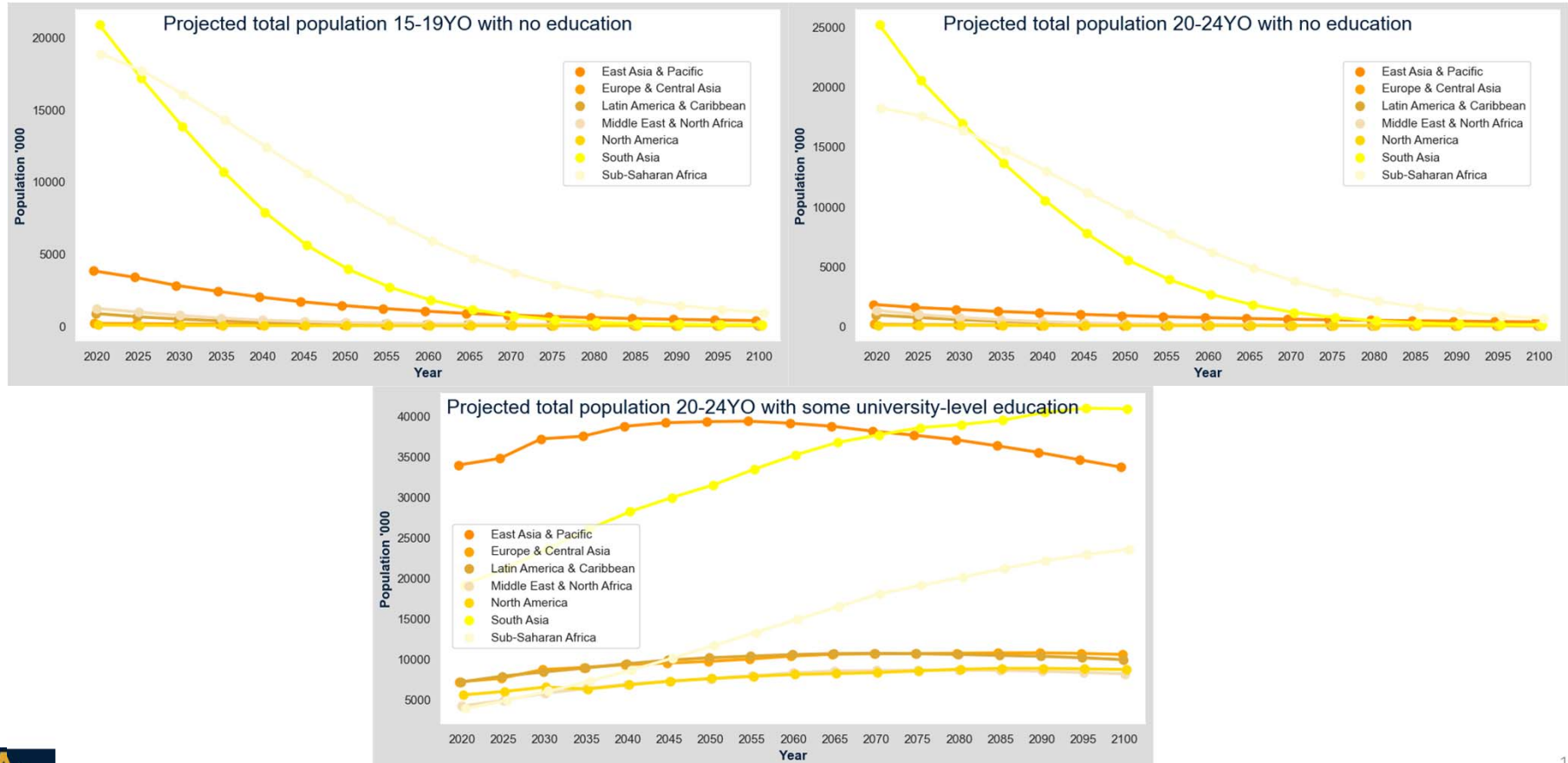
## 2 – Analyses statistiques & graphiques

### 2.2 – Indicateurs prospectifs – analyses graphiques



## 2 – Analyses statistiques & graphiques

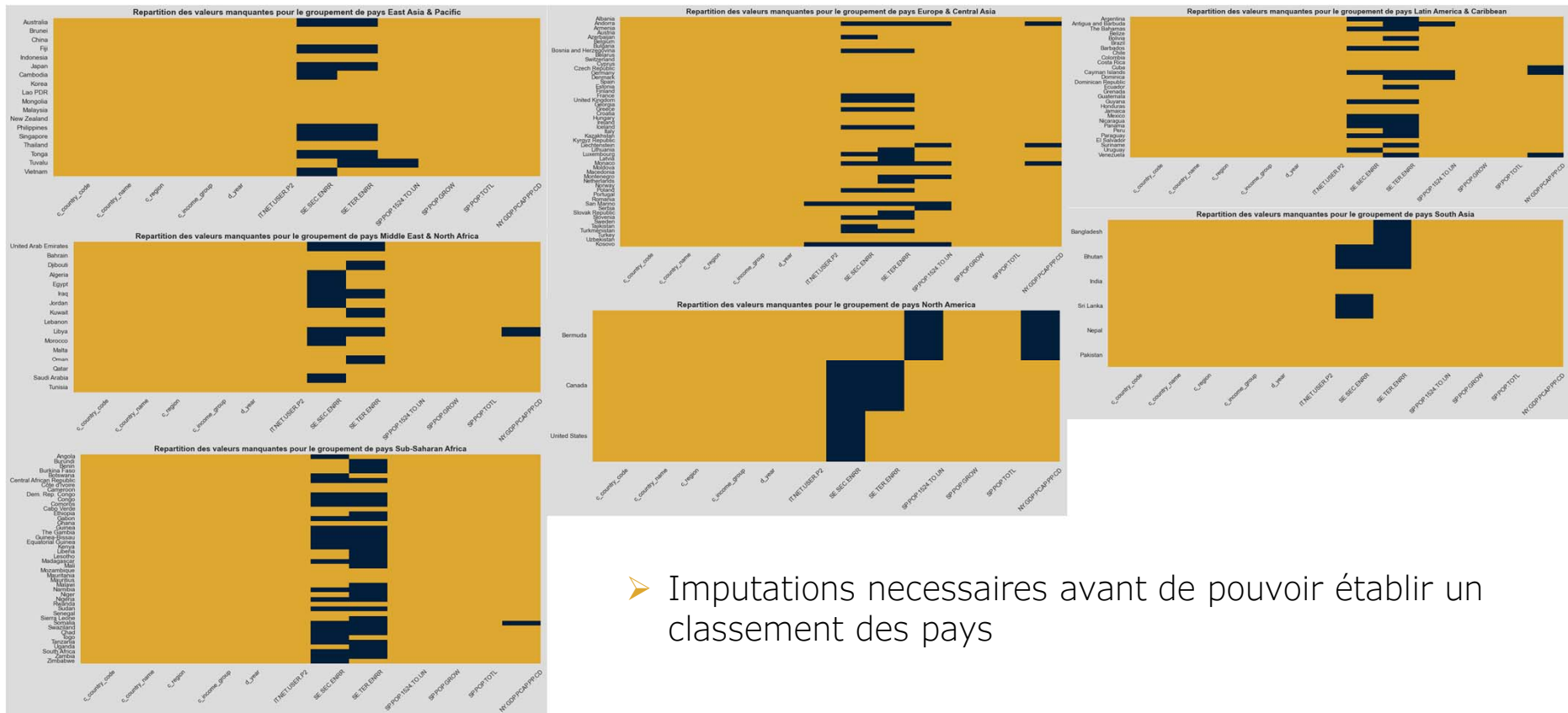
### 2.2 – Indicateurs prospectifs – analyses graphiques



Département Marketing Stratégique

## 2 – Analyses statistiques & graphiques

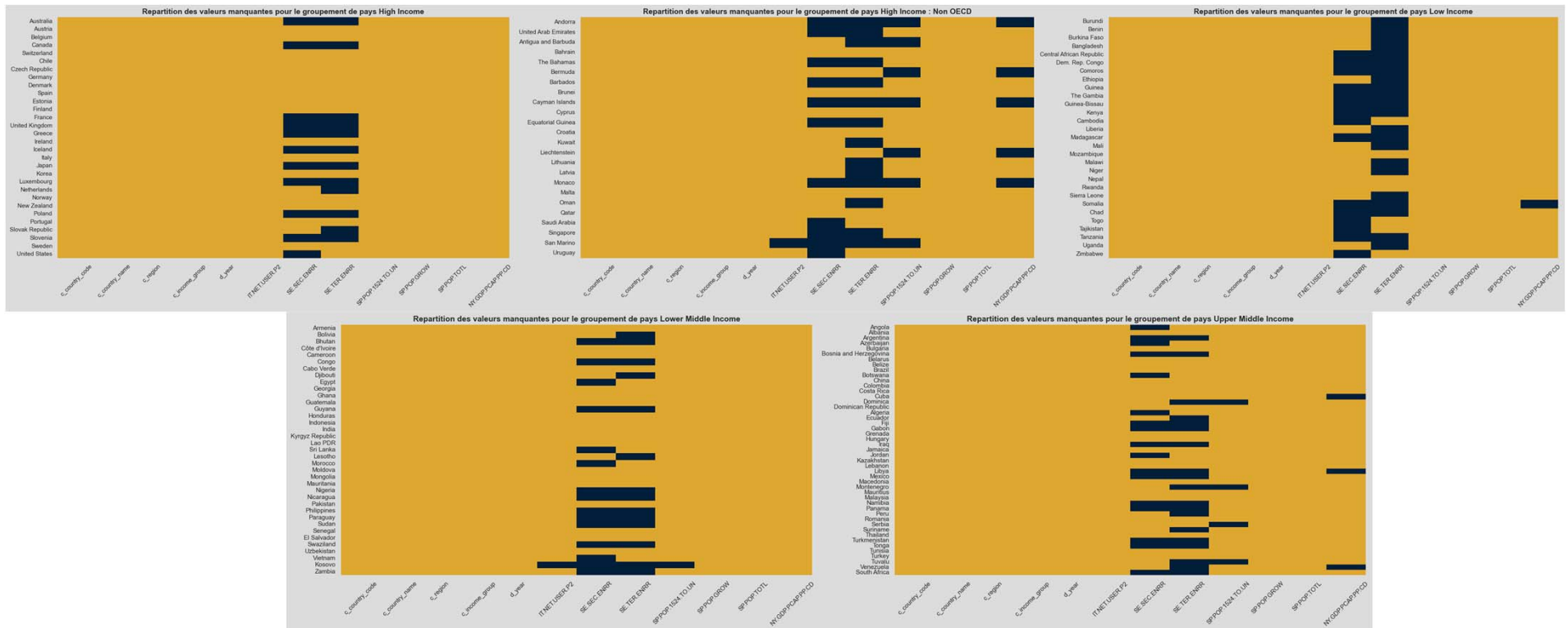
### 2.3 – Imputations & classement des pays – valeurs manquantes par région



- Imputations nécessaires avant de pouvoir établir un classement des pays

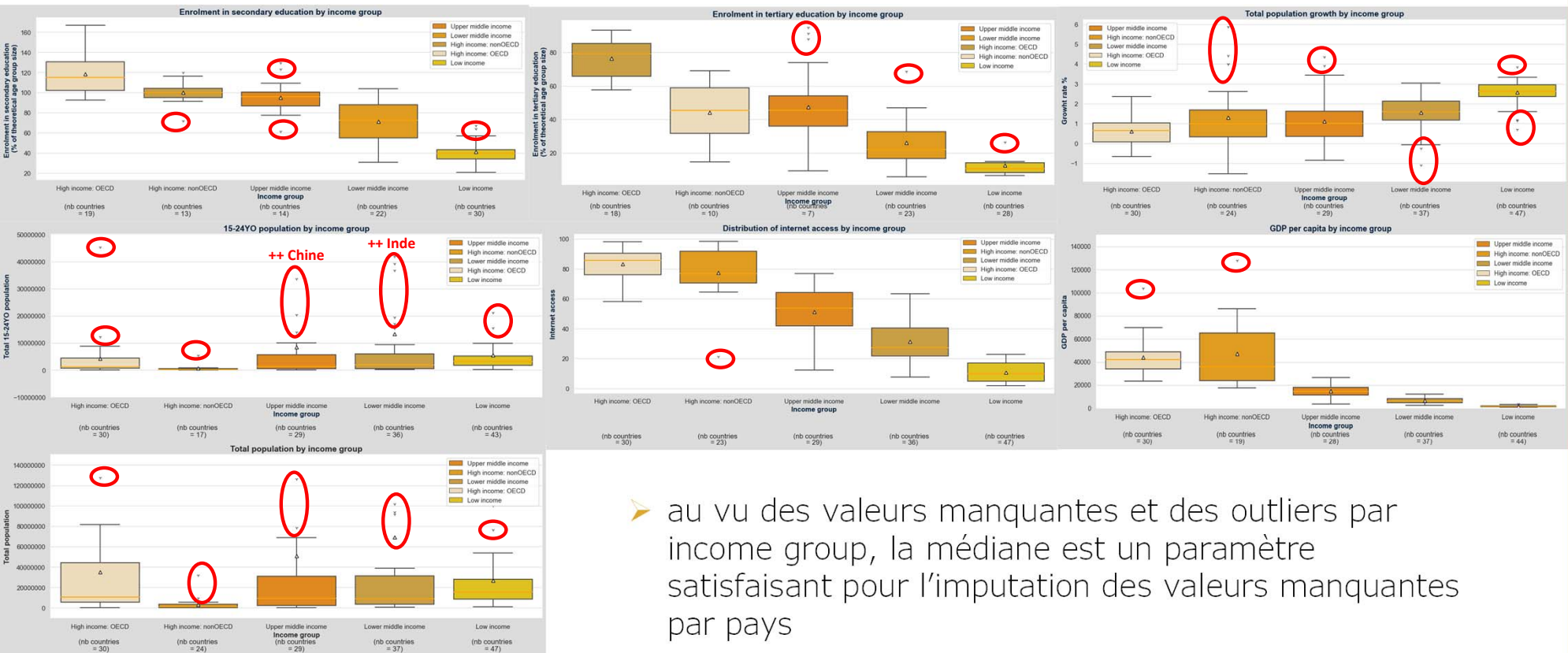
## 2 – Analyses statistiques & graphiques

### 2.3 – Imputations & classement des pays – valeurs manquantes par income group



## 2 – Analyses statistiques & graphiques

### 2.3 – Imputations & classement des pays – outliers par income group



➤ au vu des valeurs manquantes et des outliers par income group, la médiane est un paramètre satisfaisant pour l'imputation des valeurs manquantes par pays



## 2 – Analyses statistiques & graphiques

### 2.3 – Imputations & classement des pays

#### Méthodologie:

- ❖ Valeurs manquantes imputées pour chaque indicateur
  - valeur médiane par Income Group en 2015
- ❖ Classement de chaque pays en fonction de 6 des 7 indicateurs historiques retenus (à l'exclusion de la population totale)
  - dense ranking (pas d'ex-æquo)
- ❖ Calcul du **rang moyen** de chaque pays sur les 6 indicateurs
  - ex-æquo possibles
  - classement décroissant Top 20
- ❖ On calcule également le **percentile rank** de chaque pays vis-à-vis des 6 indicateurs retenus
  - vérifications de cohérence & explication du classement

## 3 – Choix des marchés cibles & recommandations stratégiques

### 3.1 – Priorisation des marchés-cibles

❖ Top 20 des pays :

c_country_code	c_country_name	c_region	c_income_group	avg_rank	it_pct_rank	sec_pct_rank	ter_pct_rank	pop_1524_pct_rank	pop_tot_pct_rank	gdp_pct_rank
USA	United States	North America	High income: OECD	32	23%	14%	9%	2%	2%	7%
SAU	Saudi Arabia	Middle East & North Africa	High income: nonOECD	32	30%	34%	25%	25%	22%	8%
GBR	United Kingdom	Europe & Central Asia	High income: OECD	34	7%	14%	15%	16%	12%	16%
AUS	Australia	East Asia & Pacific	High income: OECD	34	15%	14%	15%	37%	28%	13%
DEU	Germany	Europe & Central Asia	High income: OECD	36	12%	26%	22%	15%	9%	10%
CAN	Canada	North America	High income: OECD	37	10%	14%	15%	29%	20%	14%
LUX	Luxembourg	Europe & Central Asia	High income: OECD	37	2%	14%	15%	96%	89%	1%
NZL	New Zealand	East Asia & Pacific	High income: OECD	39	12%	12%	10%	71%	65%	18%
FRA	France	Europe & Central Asia	High income: OECD	39	15%	14%	15%	17%	11%	16%
TUR	Turkey	Europe & Central Asia	Upper middle income	39	48%	27%	1%	10%	10%	29%
JPN	Japan	East Asia & Pacific	High income: OECD	39	8%	14%	15%	10%	5%	17%
SWE	Sweden	Europe & Central Asia	High income: OECD	40	9%	3%	27%	56%	47%	11%
KOR	Korea	East Asia & Pacific	High income: OECD	40	10%	38%	2%	20%	14%	20%
NLD	Netherlands	Europe & Central Asia	High income: OECD	40	7%	4%	15%	45%	34%	9%
IRL	Ireland	Europe & Central Asia	High income: OECD	41	16%	8%	11%	72%	64%	4%
NOR	Norway	Europe & Central Asia	High income: OECD	42	3%	15%	16%	69%	62%	6%
OMN	Oman	Middle East & North Africa	High income: nonOECD	43	35%	22%	47%	72%	67%	14%
KWT	Kuwait	Middle East & North Africa	High income: nonOECD	43	20%	48%	47%	79%	69%	3%
DNK	Denmark	Europe & Central Asia	High income: OECD	44	4%	5%	12%	67%	59%	9%
BRA	Brazil	Latin America & Caribbean	Upper middle income	53	43%	34%	28%	5%	3%	45%



# 3 – Choix des marchés cibles & recommandations stratégiques

## 3.2 – Objectif 2030



# Conclusion

Le jeu de données fourni peut en effet, dans une certaine mesure, informer les décisions stratégiques d'expansion internationale de l'entreprise, mais...

❖ ...pas en première analyse :

- Besoin d'analyser le positionnement concurrentiel prix-produit pour déterminer la stratégie en amont
- Et de penser la gamme de produits dans un contexte international : où sont les besoins et quels sont-ils? La gamme de cours actuelle y est-elle adaptée?
- Besoin de repenser également le marché domestique (formation CPF? Diplômantes?)
- Et d'évaluer les barrières à l'entrée réglementaires des différents marchés (l'éducation étant la prérogative des états – conditions de faisabilité hors-EU à déterminer)

❖ ...pas en unique source de données chiffrées :

- Besoin d'intégrer d'autres indicateurs comme les classements PISA et les pratiques éducatives extra-scolaires (cf. Singapour par exemple), ainsi que d'évaluer le champ concurrentiel domestique & international

❖ ...et pas intégralement :

- Multiplicité d'indicateurs inutiles et de faible densité informationnelle – mieux cibler la collecte des données