

# **Analyses de données de systèmes éducatifs**

**Expansion de *academy***

**Joyce Kuoh Moukouri,  
P2, Soutenance du 13/01/2023**

# Ordre du jour

## **Analyses de données de systèmes éducatifs**

1. La mission
2. Présentation de la base de données
3. Calcul du potentiel de clients
4. Résultats : pays à fort potentiel de clients
5. Évolution du potentiel de clients
6. Conclusion

# 1. La mission

# La mission

## Rappel des objectifs fixés par Mark

- Identifier les pays avec un fort potentiel de clients pour nos services.
- Quelle sera l'évolution de ce potentiel de clients ?
- Dans quels pays l'entreprise doit-elle opérer en priorité ?

# 1. La mission

Échanges fictifs avec Mark : les hypothèses fixées

- (H1) **academy** est une plateforme de soutien scolaire (ex. Maxicours)
- (H2) la plateforme est accessible sur PC, tablette et smartphone

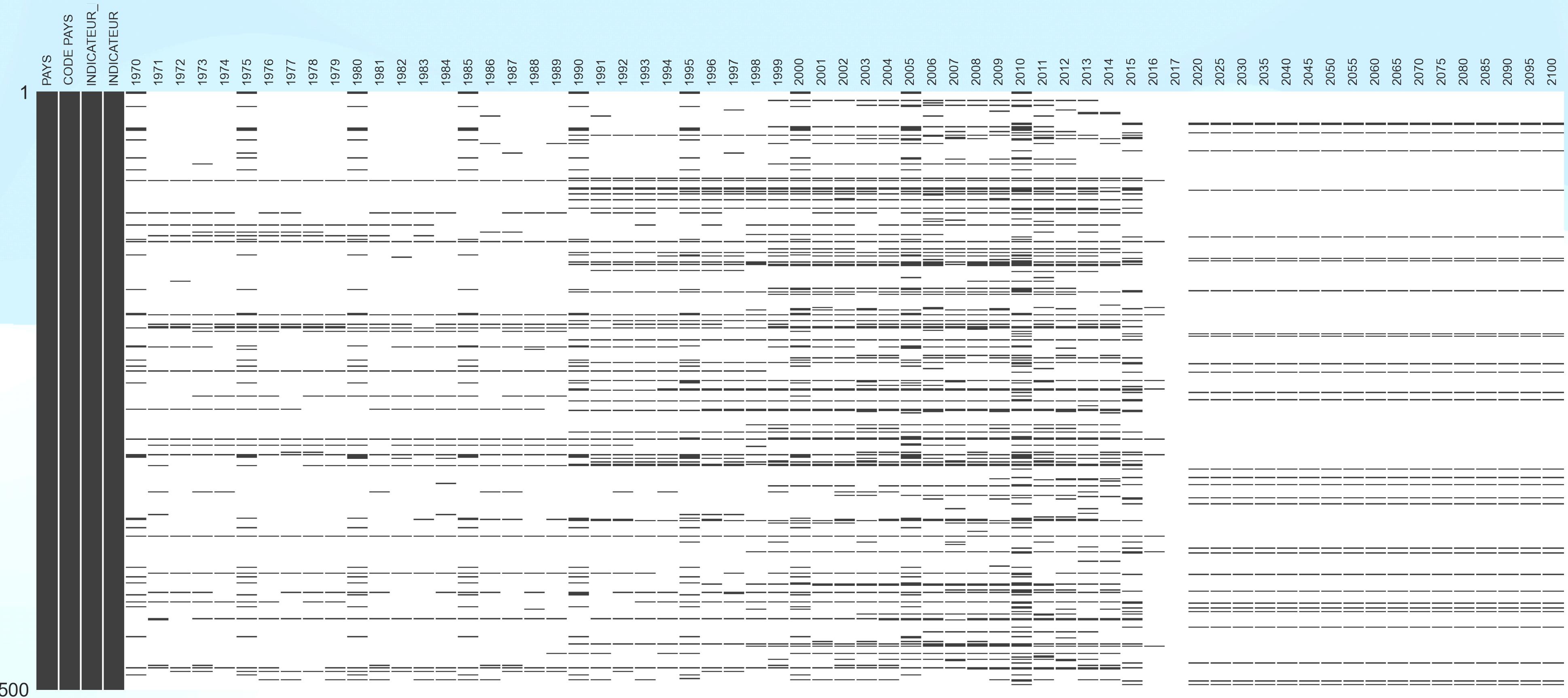
## **2. La base de données**

## 2. La base de données

- Base de données de la banque mondiale, +4000 indicateurs
- À disposition : 3665 indicateurs concernant l'éducation.
- **5 tables**

	Fonction	Clé	Taille	Traitement
<b>Country</b>	Description du pays	Un pays	241 lignes, 32 colonnes	Nettoyage et création de Country_sub, 214 lignes, 4 colonnes
<b>Country_s</b>	Pour chaque pays, répertorie les indicateurs propres à la population et renseigne leur sources	Un pays et un indicateur propre à la population	613 lignes, 4 colonnes	Nettoyage, 613 lignes, 3 colonnes
<b>Data</b>	Pour chaque pays et indicateur, répertorie les valeurs entre 1970 et 2100	Un pays, un indicateur	886 930 lignes, 70 colonnes	Nettoyage et création de Data_sub (voir slide d'après)
<b>Series</b>	Liste les 3665 indicateurs que l'on a disposition dans Data et donne leur définition	Un indicateur	3665 lignes, 21 colonnes	Nettoyage, 3665 lignes et 6 colonnes
<b>Foot_Note</b>	Répertorie les note de bas de page pour 1471 indicateurs	un pays, à un indicateur, l'année de calcul de l'indicateur	643638 lignes, 5 colonnes	Nettoyage, 515752 lignes et 4 colonnes

## 2. La base de données



# Visualisation des valeurs manquantes de la table *Data*

## 2. La base de données

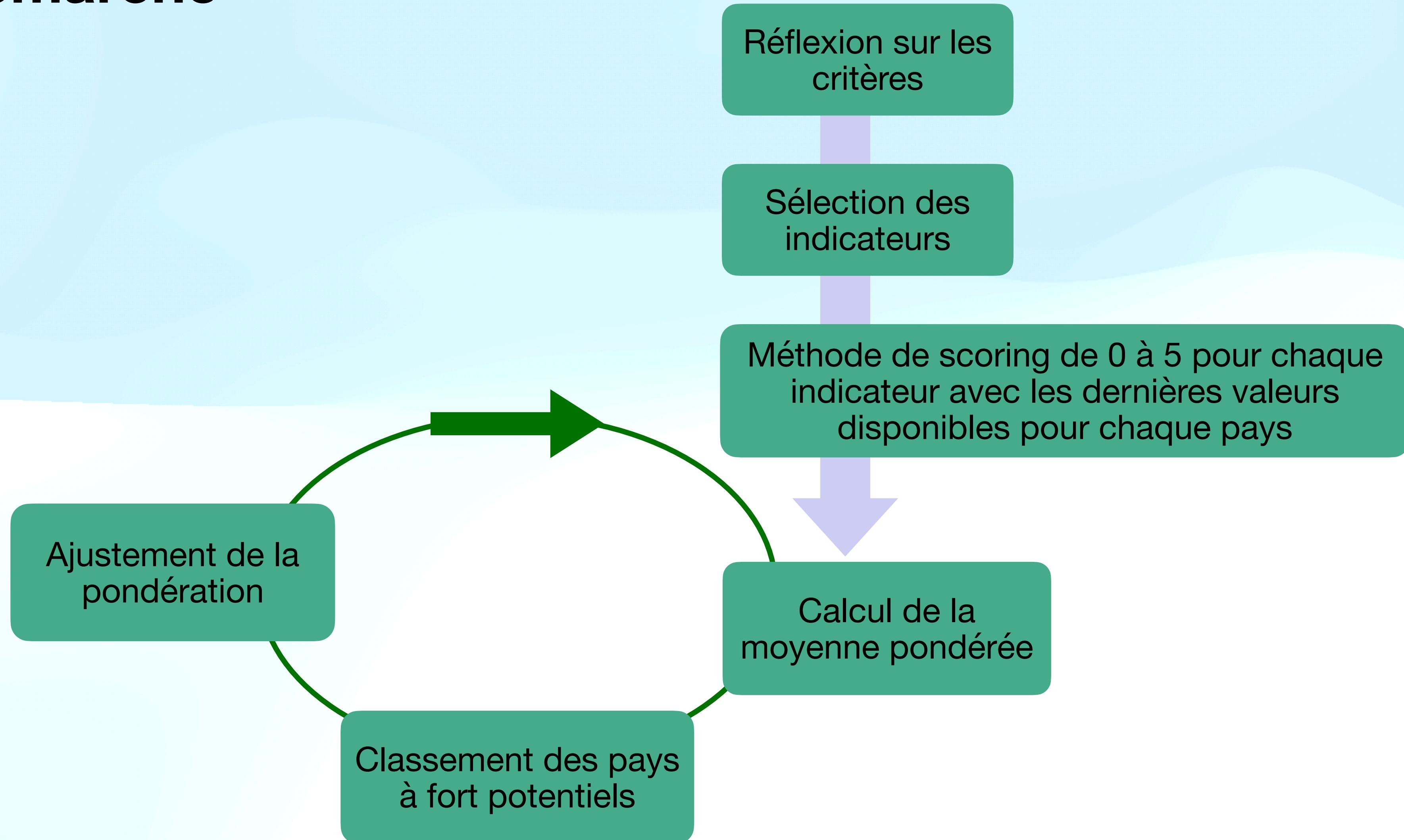
### Qualité du jeu de données, table *Data*

- Les taux de valeurs manquantes sont élevés et en moyenne supérieur à 70% . L'année 2010 est la mieux renseignée.
- Absence de doublons
- Si 30% des d'enregistrement sont correctement renseignés entre 2005 et 2016, il serait possible d'en tirer des conclusions intéressantes.
- Étude faite principalement entre 2005 et 2016

### **3. Calcul du potentiel de clients**

# 3. Calcul du potentiel de clients

## Démarche



### 3. Calcul du potentiel de clients

#### Les critères théorique vs indicateurs de la bdd

- L'accès à une connexion internet **OK**
- Le pouvoir d'achat **NOK - niveau de revenus**
- L'investissement public dans les universités et lycées locaux **OK**
- Le coût de l'instruction **NOK**
- Taux d'inscription aux universités **OK**
- Taux d'inscription aux lycées **OK**
- L'attractivité d'investissement **NOK - cf. Source, étude Forbes**

# 3. Calcul du potentiel de clients

## Les critères accessibles avec la bdd

Indicateurs	Critères	Unité	Description	Taux de vm moyen (entre 2005 et 2015)	Commentaire
INCOME GROUPS	Revenus	Cat	Données qualitative de la table Country_sub, divise les pays en 5 groupes. 'High income: nonOECD', 'Low income', 'Upper middle income', 'Lower middle income', 'High income: OECD'	0 %	Les + : Taux de valeurs manquantes nul Pour analyse préliminaire, bonne indication du pouvoir d'achat.
IT.NET.USER.P2	Accès à internet	Nombre d'utilisateurs pour 100 personnes	Utilisateur d'un réseau internet, qui ont utilisé internet durant les 3 mois précédent l'enregistrement.	6,7 %	Les + : Taux de valeur manquantes faibles. Les - : La facilité d'accès n'est pas précisé et l'intervalle de trois mois est large.
UIS.E.3	Inscription au lycée	Nombre total	Nombre total d'élèves inscrits au lycée en institutions publique ou privée tous âge confondus.	37 %	Les + : Taux de valeur manquantes acceptable. Les - : Indicateur qui ne correspond exactement au critère. Convient à l'offre de soutien scolaire
SE.TER.ENRL	Inscription à l'université	Nombre total	Nombre total d'inscription en enseignement supérieur en institutions publiques ou privées tous âge confondus.	39 %	Les + : Taux de valeur manquantes acceptable. Les - : Indicateur qui ne correspond exactement au critère. Convient à l'offre de soutien scolaire
UIS.XGDP.56.FSGOV	L'investissement public dans les universités locales	(%) du PIB	Total des dépenses du gouvernement pour l'éducation pour le cycle tertiaire	63 %	Les + : Indicateur qui correspond aux critères. Les - : 63% de valeurs manquantes
UIS.XGDP.23.FSGOV	L'investissement public dans les collèges et lycées	(%) du PIB	Total des dépenses du gouvernement pour l'éducation pour le cycle secondaire	64 %	Les + : Indicateur qui correspond aux critères. Les - : 63% de valeurs manquantes

# **3. Calcul du potentiel de clients**

## **Création de nouveaux indicateurs**

- ACC.IT.EL = Proportion ayant accès internet x Nb élèves total
- DIFF.INV.REV = moyenne (score d'Investissement dans l'éducation - score niveau de revenus moyen)

# 3. Calcul du potentiel de clients

## Choix des dernières données disponibles

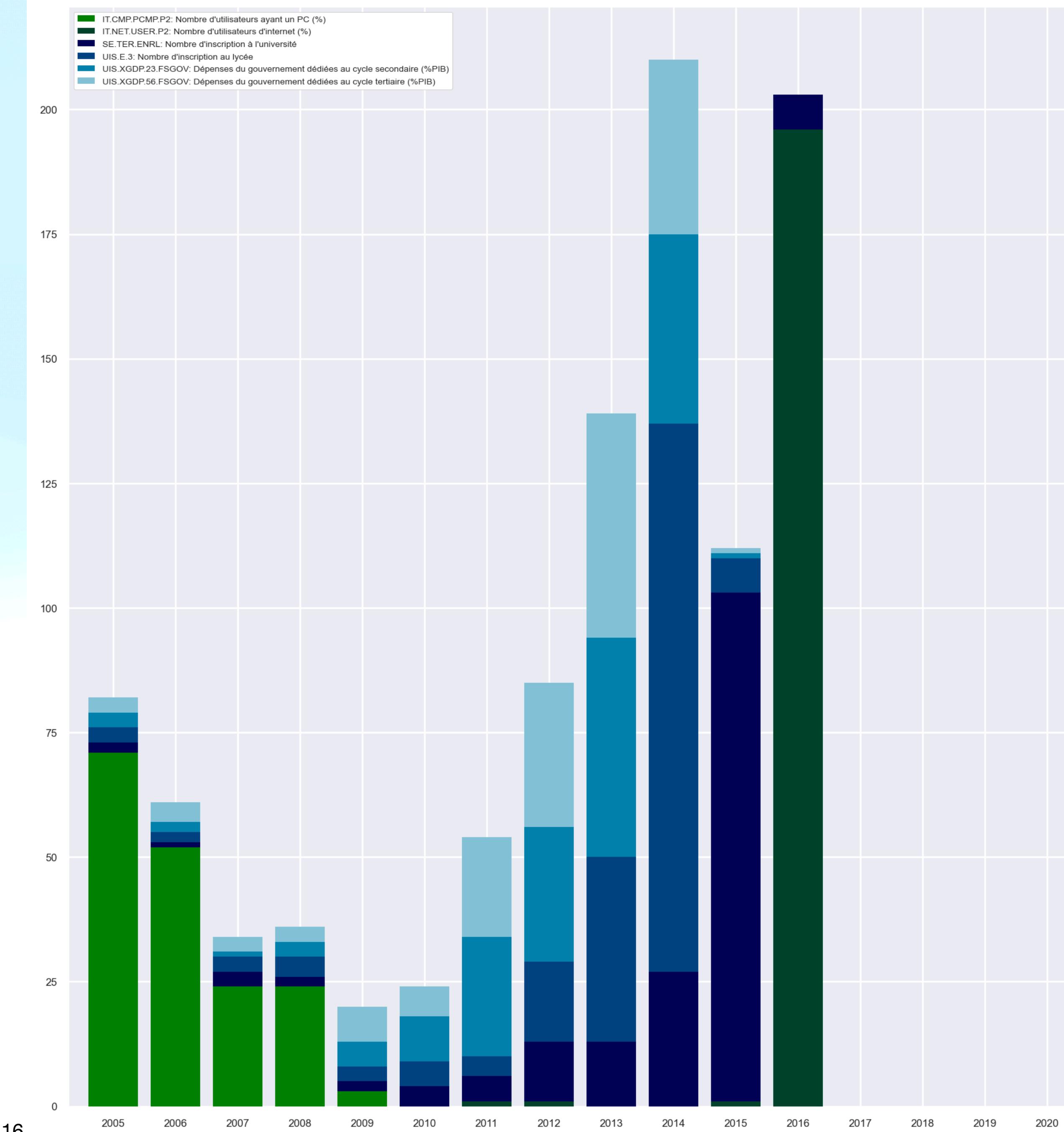
Data_update							
	PAYS	CODE PAYS	INDICATEUR_N	INDICATEUR	MAJ	ANNÉE_MAJ	NB_NAN
0	Afghanistan	AFG	Nombre d'inscription à l'université	SE.TER.ENRL	262874.0	2014	11
1	Afghanistan	AFG	Nombre d'inscription au lycée	UIS.E.3	968769.0	2014	4
2	Afghanistan	AFG	Nombre d'utilisateurs d'internet (%)	IT.NET.USER.P2	10.595726	2016	2
3	Afghanistan	AFG	Nombre d'utilisateurs ayant un PC (%)	IT.CMP.PCMP.P2	0.390148	2006	12
4	Albania	ALB	Nombre d'inscription à l'université	SE.TER.ENRL	160527.0	2015	3
...	...	...	...	...	...	...	...
1055	Zimbabwe	ZWE	Nombre d'inscription au lycée	UIS.E.3	490522.0	2013	12
1056	Zimbabwe	ZWE	Dépenses du gouvernement dédiées au cycle seco...	UIS.XGDP.23.FSGOV	0.50527	2010	13
1057	Zimbabwe	ZWE	Dépenses du gouvernement dédiées au cycle tert...	UIS.XGDP.56.FSGOV	0.45021	2010	12
1058	Zimbabwe	ZWE	Nombre d'utilisateurs d'internet (%)	IT.NET.USER.P2	23.119989	2016	2
1059	Zimbabwe	ZWE	Nombre d'utilisateurs ayant un PC (%)	IT.CMP.PCMP.P2	7.43114	2008	11

1060 rows × 7 columns

Table *Data\_update*

### 3. Calcul du potentiel de clients

- Les années 2017 et 2020 ne renferment aucune données relatives aux indicateurs sélectionnés.
- Les valeurs les plus récentes, datent de l'année 2016.
- L'indicateur IT.CMP.PCMP.P2, relatif à l'accès à un ordinateur personnel n'est pas renseigné au delà de l'année 2009. Entre temps de nombreux moyens et appareils ont permis d'accéder à un internet.



### 3. Calcul du potentiel de clients

- Choix des dernières données disponibles 
- Calcul d'un score sur 5 pour chaque indicateur sélectionné 
- Calcul de la moyenne pondérée des scores 

Indicateur	Description	Pondération
ACC.IT.EL	Nombre d'élèves inscrit ayant accès à internet	1.5
UIS.E.3	Nombre total d'inscrit au lycée	1.5
SE.TER.ENRL	Nombre total d'inscrit à l'université	1.5
INCOME GROUPS	Niveau de revenu moyen	0.1
DIFF.INV.REV	Différence entre l'investissement public et le niveau de revenu	0.1

# 4. Pays à fort potentiel de clients

## Résultats

	PAYS	CODE PAYS	REGION	SCORE REVENUS	SE.TER.ENRL	UIS.E.3	ACC.IT.EL	DIFF.INV.REV	SCORE TOTAL
0	China	CHN	East Asia & Pacific	3	5.00	3.96	5.00	NaN	4.52
1	India	IND	South Asia	2	3.70	5.00	2.79	0.07	3.72
2	United States	USA	North America	5	2.25	1.06	2.57	0.35	2.06
3	Brazil	BRA	Latin America & Caribbean	3	0.96	0.90	1.17	0.12	1.06
4	Russian Federation	RUS	Europe & Central Asia	5	0.76	0.26	0.78	0.42	0.77
5	Turkey	TUR	Europe & Central Asia	3	0.70	0.45	0.70	0.21	0.70
6	Indonesia	IDN	East Asia & Pacific	2	0.59	0.90	0.41	0.14	0.68
7	Japan	JPN	East Asia & Pacific	5	0.45	0.33	0.75	0.40	0.68
8	United Kingdom	GBR	Europe & Central Asia	5	0.27	0.38	0.67	0.32	0.60
9	Iran, Islamic Rep.	IRN	Middle East & North Africa	3	0.55	0.33	0.49	0.21	0.55
10	Mexico	MEX	Latin America & Caribbean	3	0.39	0.42	0.52	0.18	0.53
11	Germany	DEU	Europe & Central Asia	5	0.34	0.23	0.54	0.33	0.53
12	Korea, Rep.	KOR	East Asia & Pacific	5	0.38	0.17	0.52	0.37	0.53
13	France	FRA	Europe & Central Asia	5	0.28	0.24	0.46	0.32	0.48
14	Italy	ITA	Europe & Central Asia	5	0.21	0.25	0.30	0.38	0.43
15	Spain	ESP	Europe & Central Asia	5	0.23	0.15	0.32	0.38	0.41
16	Poland	POL	Europe & Central Asia	5	0.20	0.14	0.27	0.37	0.38
17	Argentina	ARG	Latin America & Caribbean	3	0.33	0.16	0.35	0.14	0.36
18	Saudi Arabia	SAU	Middle East & North Africa	5	0.18	0.15	0.26	0.31	0.36
19	Philippines	PHL	East Asia & Pacific	2	0.41	0.14	0.31	0.15	0.35

## **4. Résultats : pays à fort potentiel de clients**

## 4. Pays à fort potentiel de clients

Rapport *Forbes*, 2021

The world's most attractive spots for VC/PE investors		
Country	Rank	Score
United States	1	100.0
United Kingdom	2	90.3
Japan	3	87.4
Germany	4	87.3
Canada	5	87.2
Singapore	6	85.0
China	7	84.7
Australia	8	84.0
Korea, South	9	83.8
France	10	83.6

The report was prepared by [Alexander Groh](#) (emlyon business school), [Heinrich Liechtenstein](#) (IESE Business School), [Karsten Lieser](#) (eXapital) and IESE guest researcher Markus Biesinger.

# 4. Pays à fort potentiel de clients

## Résultats nuancés par l'attractivité selon *Forbes*



Tableau de score

# **5. Évolution du potentiel**

# 5. Évolution du potentiel

## Démarche

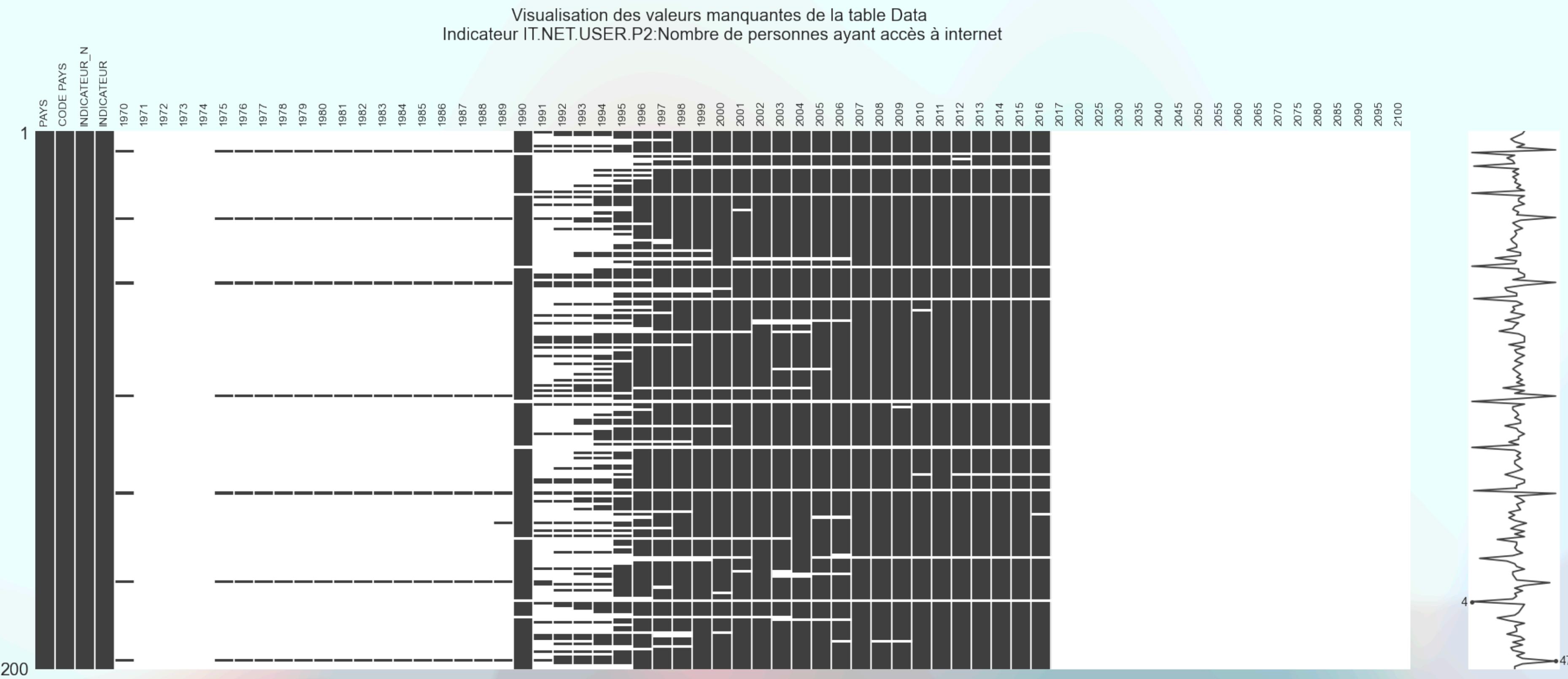
Traitement des  
valeurs manquantes  
de *Data\_sub*

Calcul du potentiel pour chaque  
pays et chaque années entre  
2005 et 2016

Classement des pays à fort  
potentiels en fonction de leur  
score moyen

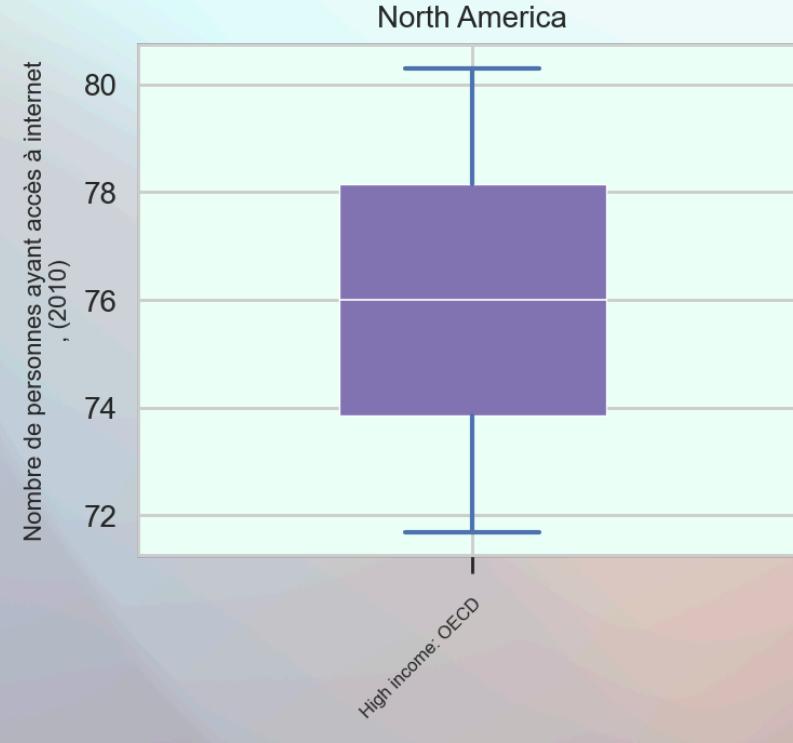
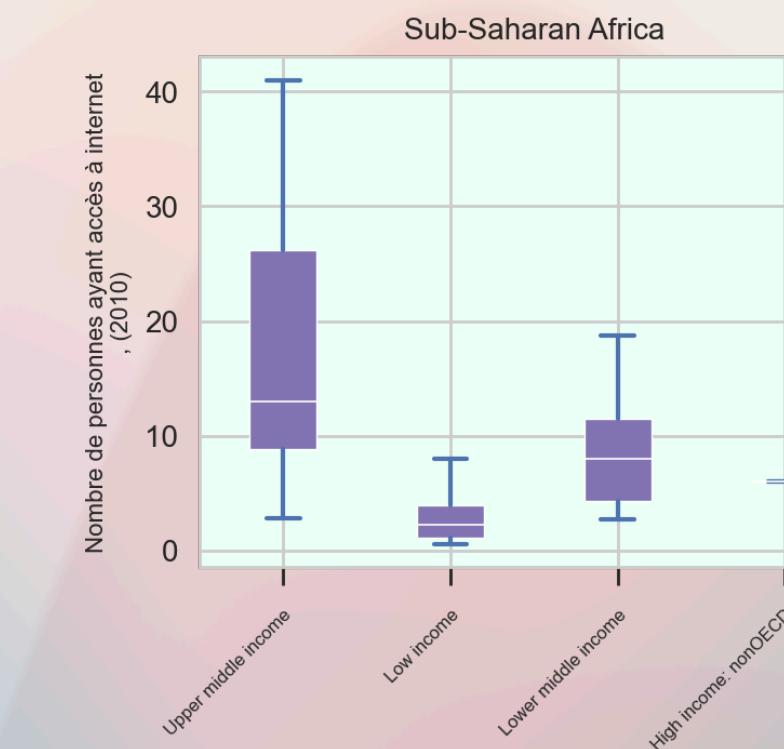
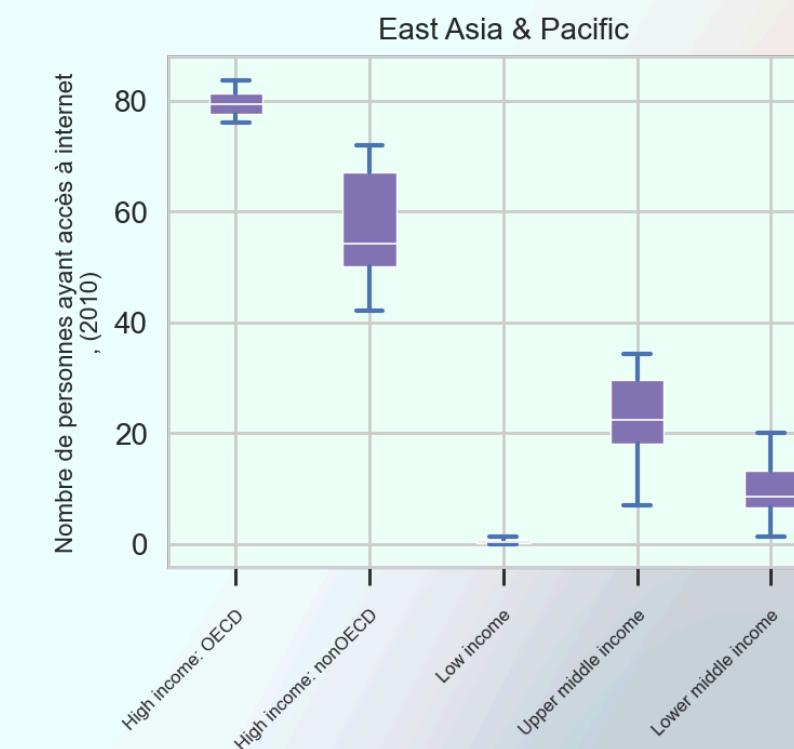
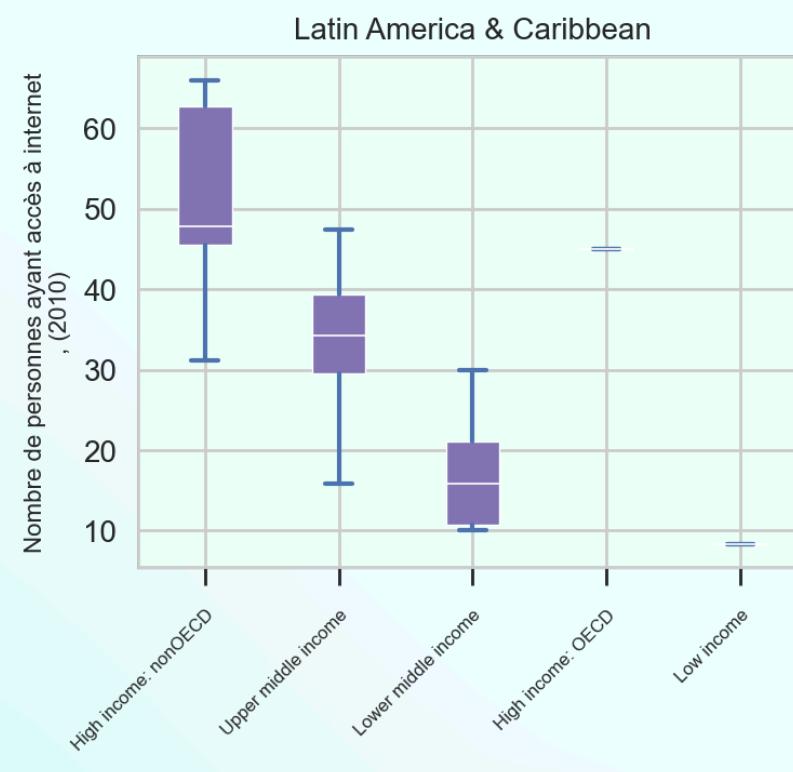
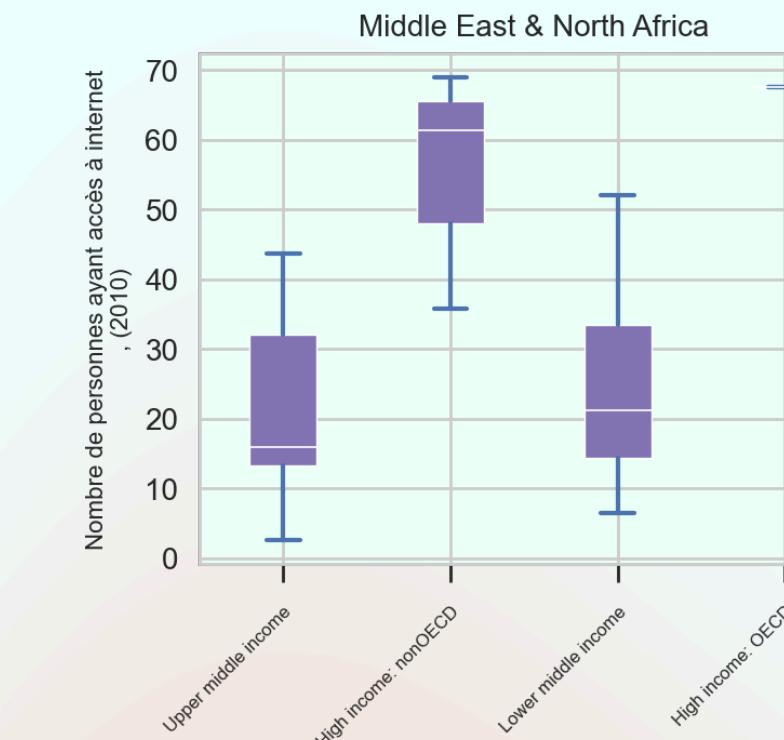
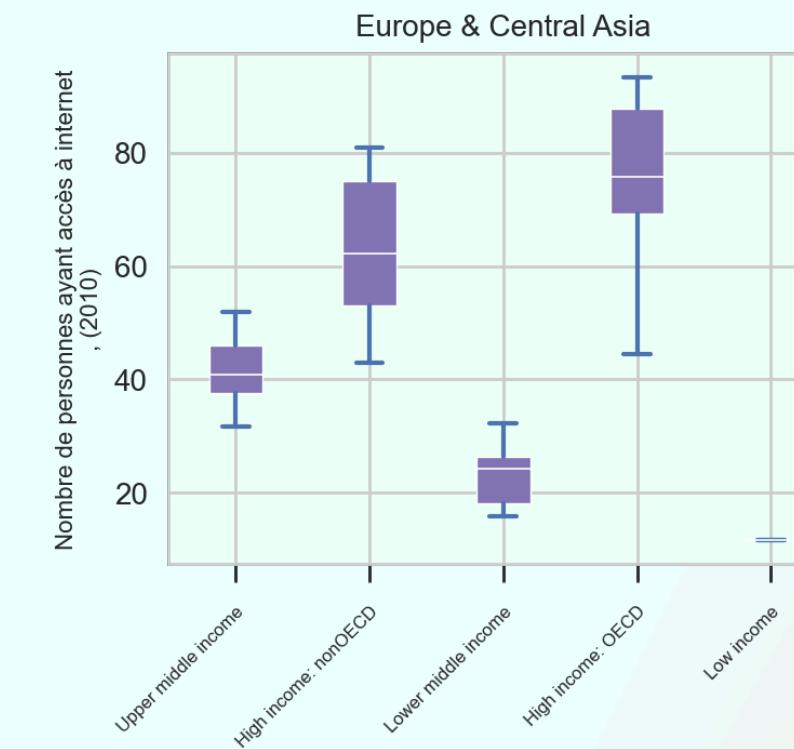
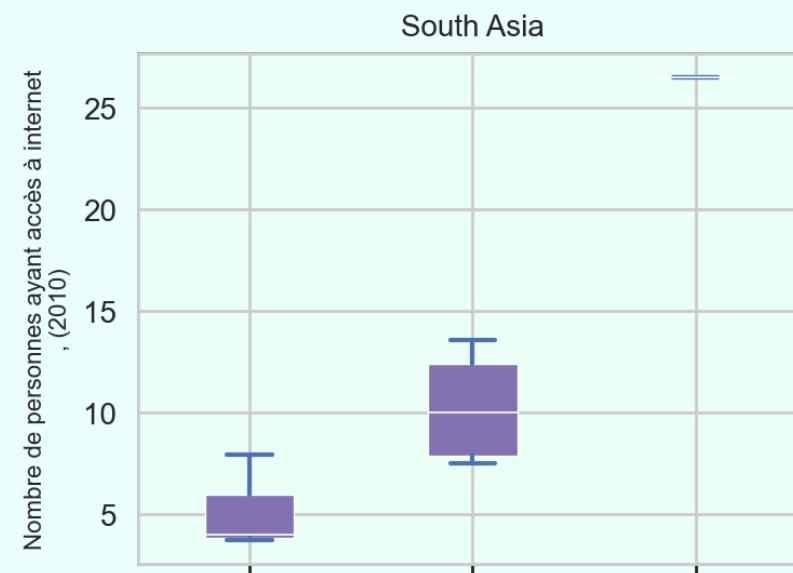
Régression linéaire et  
conclusion

## 5. Évolution du potentiel : traitement des valeurs manquantes



- Pour les indicateurs sélectionnés aucune valeur n'est disponible au delà de l'année 2016.

## 5. Évolution du potentiel : traitement des valeurs manquantes



- Les disparités sont trop importantes pour remplacer les valeurs manquantes de Data par la valeur médiane par région et income groups
- Stratégie : remplacer les valeurs manquantes de proche en proche.

Nombre de personnes ayant accès à internet

## 5. Évolution du potentiel de clients: traitement des valeurs manquantes

Stratégie : remplacer les valeurs manquantes de proche en proche.

Entrée [249]: `Data[Data['CODE PAYS']=='BRA'].loc[Data['INDICATEUR']=='UIS.E.3', col]`

Out [249]:

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
188129	9793988.0	NaN	9073330.0	9169279.0	9200656.0	9284000.0	9395813.0	9442334.0	9949583.0	NaN	NaN	NaN

Entrée [250]: `Data_sub[Data_sub['CODE PAYS']=='BRA'].loc[Data_sub['INDICATEUR']=='UIS.E.3', col]`

Out [250]:

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
110	9793988.0	9073330.0	9073330.0	9073330.0	9169279.0	9200656.0	9284000.0	9395813.0	9442334.0	9949583.0	9949583.0	9949583.0

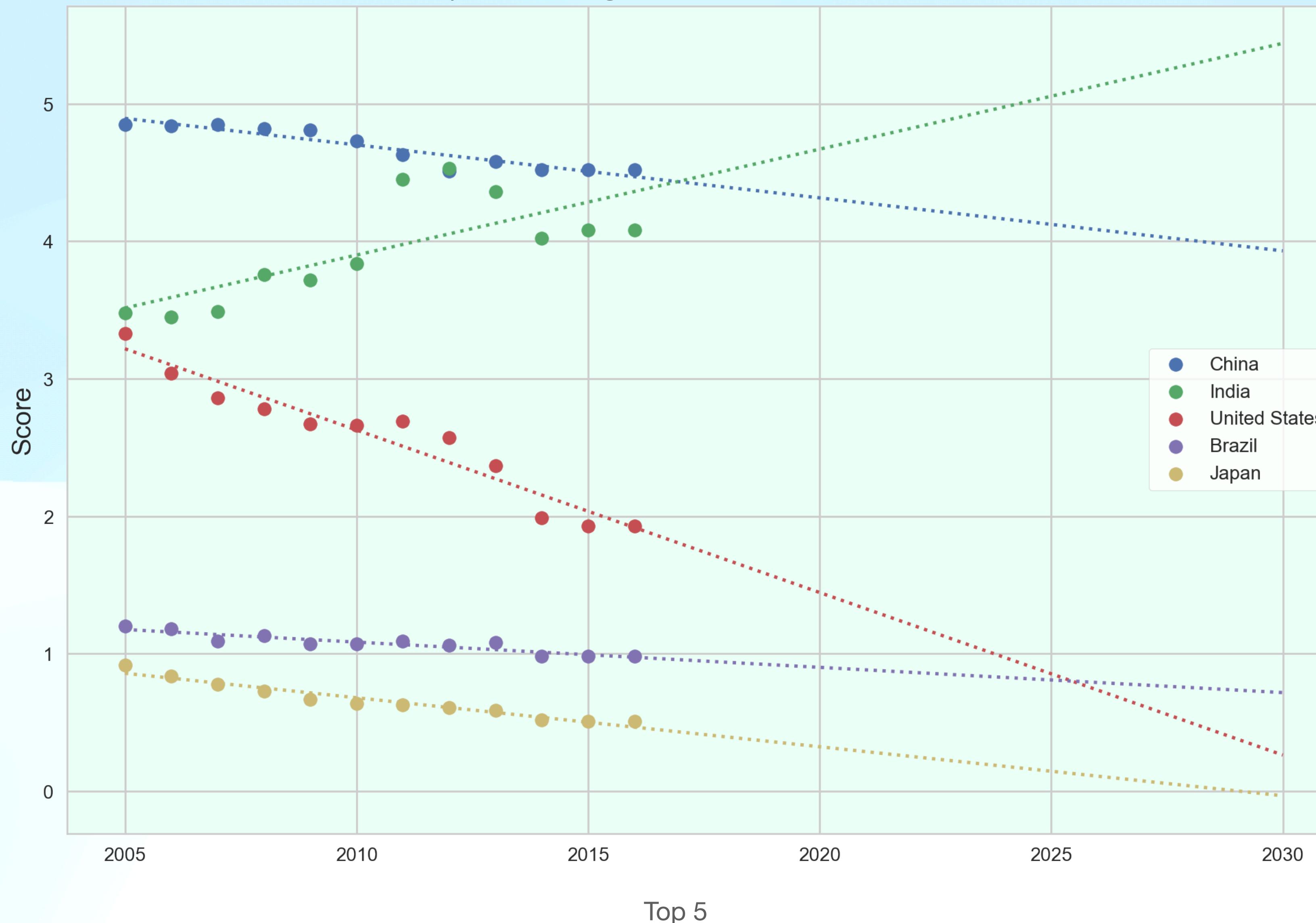
Exemple : Brésil, indicateur UIS.E.3, nombre d'inscriptions au lycée

## 5. Évolution du potentiel de clients : résultats, nuancés par l'attractivité selon *Forbes*

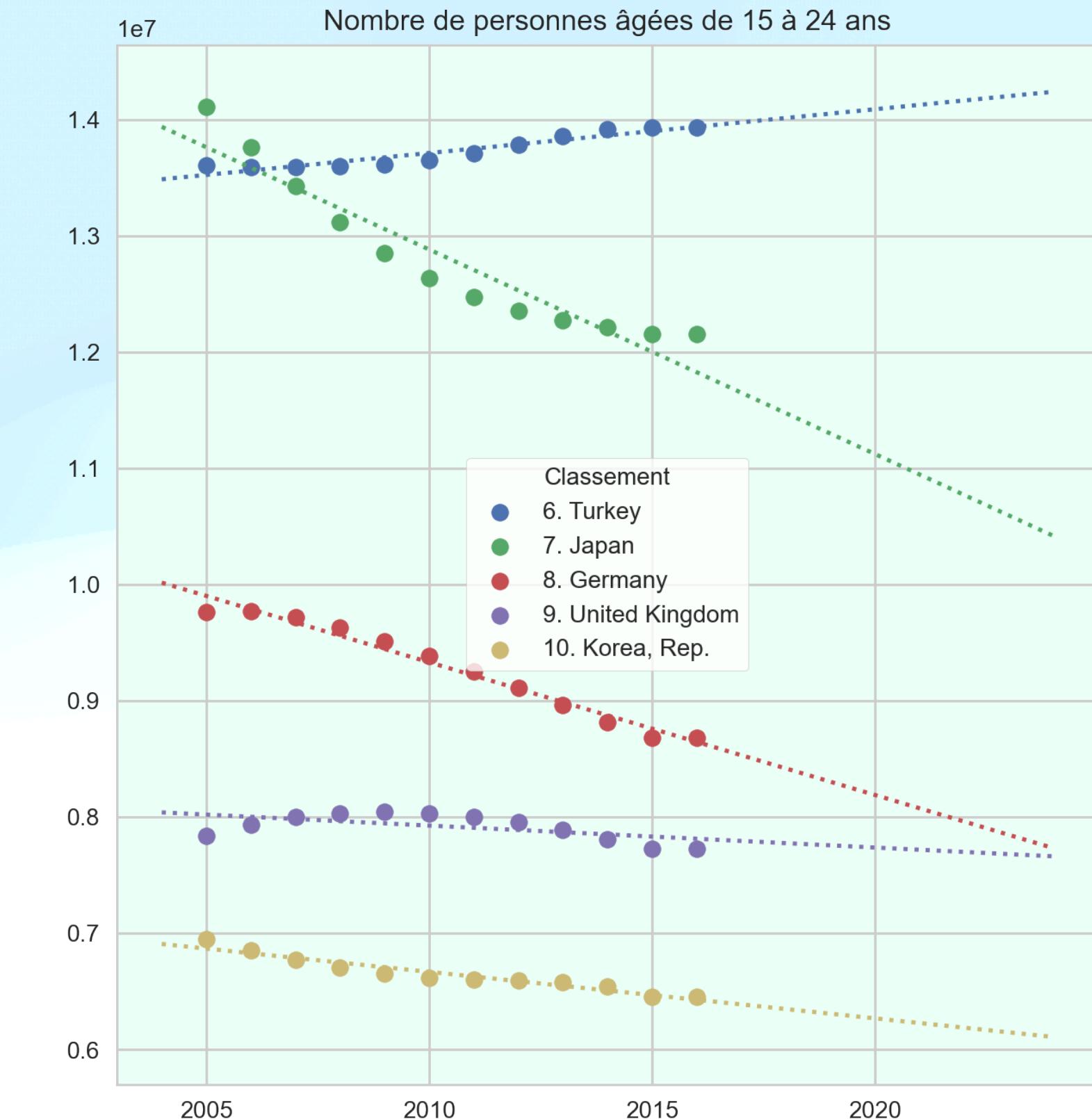
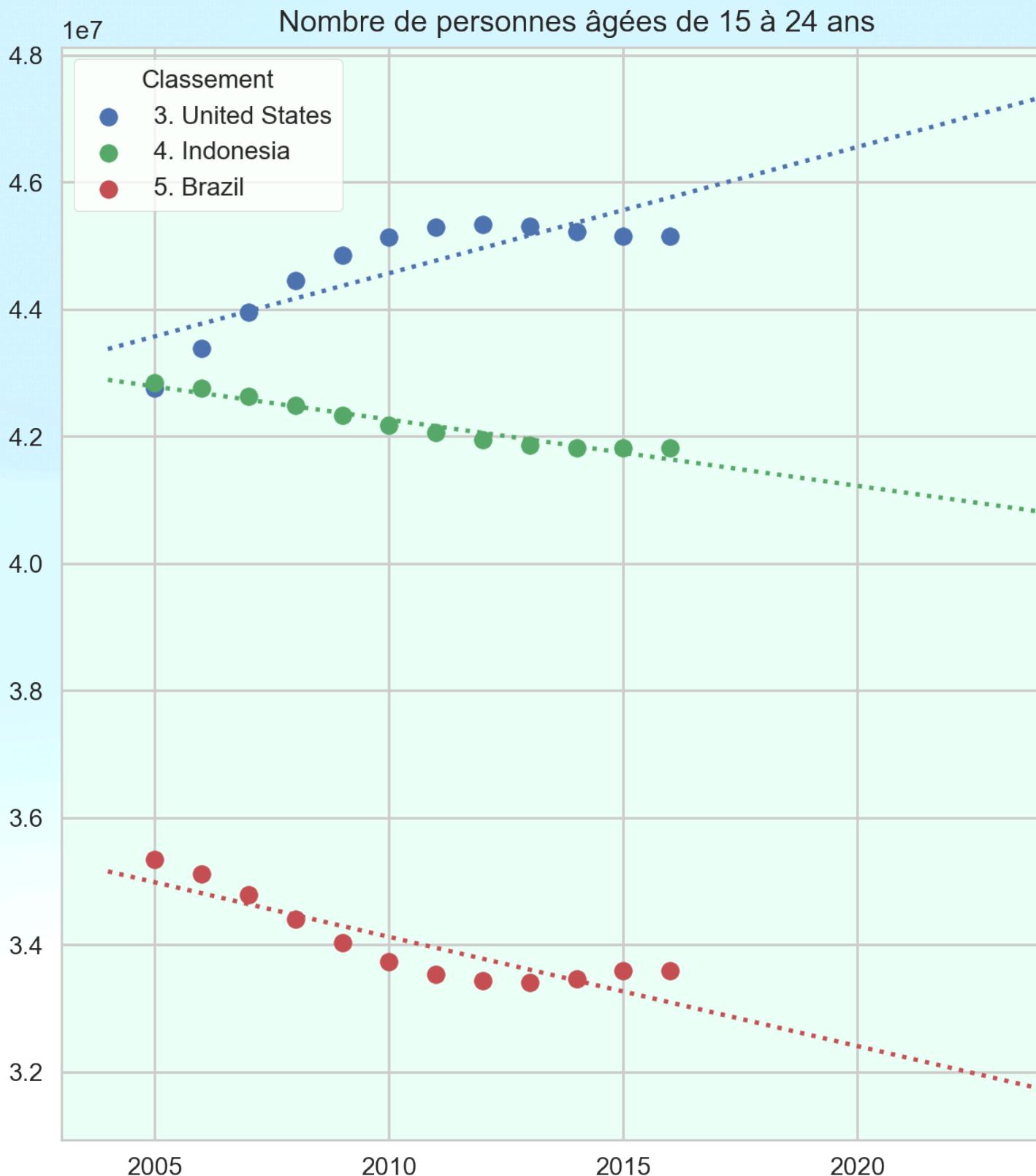
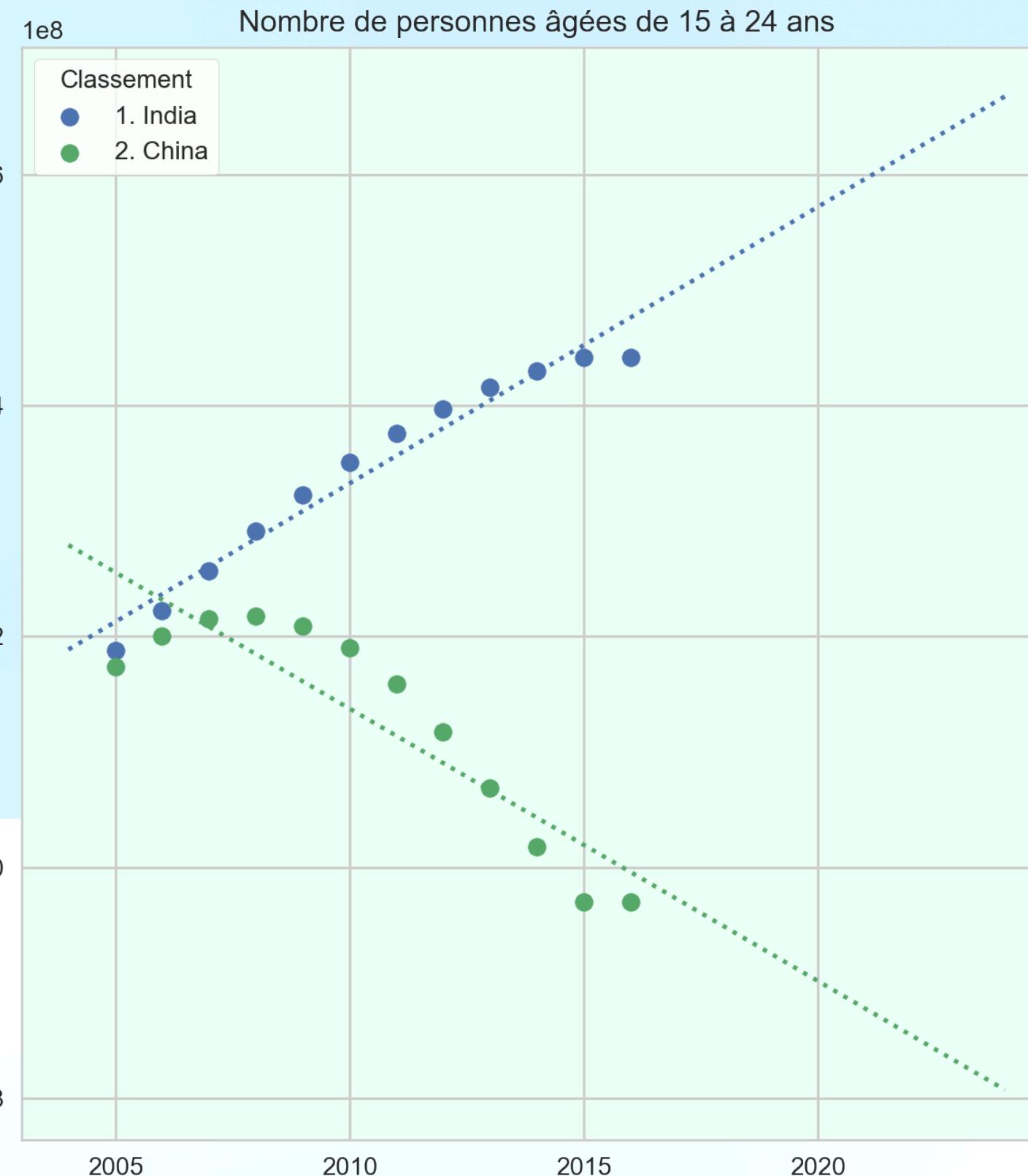
	PAYS	CODE PAYS	RÉGION	REVENUS	INDICATEUR_N	INDICATEUR	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	SCORE MOYEN
0	China	CHN	East Asia & Pacific	Upper middle income	Score annuel, moyenne pondérée des scores de c...	SCORE ANNUEL	4.85	4.84	4.85	4.82	4.81	4.73	4.63	4.51	4.58	4.52	4.52	4.68	
1	India	IND	South Asia	Lower middle income	Score annuel, moyenne pondérée des scores de c...	SCORE ANNUEL	3.48	3.45	3.49	3.76	3.72	3.84	4.45	4.53	4.36	4.02	4.08	4.08	3.94
2	United States	USA	North America	High income: OECD	Score annuel, moyenne pondérée des scores de c...	SCORE ANNUEL	3.33	3.04	2.86	2.78	2.67	2.66	2.69	2.57	2.37	1.99	1.93	1.93	2.57
3	Russian Federation	RUS	Europe & Central Asia	High income: nonOECD	Score annuel, moyenne pondérée des scores de c...	SCORE ANNUEL	1.74	1.58	1.48	1.4	1.27	1.12	1.11	1.0	0.92	0.74	0.69	0.69	1.14
4	Brazil	BRA	Latin America & Caribbean	Upper middle income	Score annuel, moyenne pondérée des scores de c...	SCORE ANNUEL	1.2	1.18	1.09	1.13	1.07	1.07	1.09	1.06	1.08	0.98	0.98	0.98	1.08
5	Indonesia ↓	IDN	East Asia & Pacific	Lower middle income	Score annuel, moyenne pondérée des scores de c...	SCORE ANNUEL	0.89	0.82	0.81	0.85	0.87	0.85	0.88	0.94	0.95	0.84	0.72	0.72	0.85
6	Japan ↑	JPN	East Asia & Pacific	High income: OECD	Score annuel, moyenne pondérée des scores de c...	SCORE ANNUEL	0.92	0.84	0.78	0.73	0.67	0.64	0.63	0.61	0.59	0.52	0.51	0.51	0.66

## 5. Évolution du potentiel de clients : visualisation

Évolution du potentiel de clients  
Extrapolation et régression linéaire entre 2005 et 2030



## 5. Évolution du potentiel de clients : visualisation



# Conclusion

## Réponse à Mark

- Nettoyage et exploration des données de la banque mondiale, données de bonne qualité entre 2005 et 2016, aucun indicateur sélectionnés n'est disponible après 2016.
- Les pays à fort potentiels de clients ont été identifiés. Top 5 : Chine, Inde, USA, Brésil et Japon
- Évolution du potentiel de clients :
  - aucune données au-delà de 2016,
  - Aucune projection disponible
  - visualisation des tendances entre 2005 et 2030 par régression linéaire
  - Le taux de croissance de la population pourraient avoir une incidence sur le potentiels de clients, données de population non disponible après 2016
- Les pays où opérer en priorité : Chine, Inde, USA, Brésil et Japon

# Conclusion

## Quelques réserves

- Pour répondre à Mark, des données récentes sont nécessaires (les avancées en terme de connectivités sont exponentielles). Ex : le potentiel de clients de l'Afrique sub-saharienne n'est pas représentatif de la réalité d'aujourd'hui.
- La BDD ne permet pas de prendre en compte l'attractivité des pays à fort potentiels de client, utilisation de la source *Forbes*.

# Indices de démographie, année 2010

Jeunesse de la population par continents

