

Projet 2 :

Analysez des données de systèmes éducatifs

Mentor : Morgan MOISON
Elève : Marin DUCHEMIN

Plan de la soutenance :

- I~Presentation des jeux de données
 - Rappel de la problématique
 - Description des jeux de données
 - Présentation de la démarche
- II~Présentation des pré-analyses
 - Données manquantes
 - Tri des données
 - Nouveaux tableaux
- III~Conclusions
 - Point sur les informations des différents indicateurs
 - Prépondérance des Regions
 - Top 10 des Pays



I~Presentation des jeux de données

Rappel de la problématique :

Travaillant pour une entreprise de cours à distance pour le secondaire et le tertiaire, nous souhaitons connaître les pays dans lesquels s'implanter en priorité.

Pour cela nous chercherons dans notre base de données les indicateurs relevant de l'éducation, de l'accès à internet et de l'économie de chaque pays pour avoir une idée de l'ordre dans lequel établir les nouvelles antennes de l'entreprise.



I~Presentation des jeux de données

Description des jeux de données:

Jeux de données	Nbr lignes	Nbr colonnes	Description succincte
EdStatsCountry-Series.csv	613	4	Pays et indicateurs faits
EdStatsCountry.csv	241	32	Pays et informations diverses (population, réserve d'eau, etc)
EdStatsData.csv	886930	70	Pays et résultats d'indicateurs
EdStatsFootNote.csv	643638	5	Pays et années de dernière observation d'indicateur
EdStatsSeries.csv	3665	21	Indicateurs et ce qu'ils observent

I~Presentation des jeux de données

Description des jeux de données:

Exemple du jeu de données EdStatsData.csv pour les types de données retrouvées dans les colonnes.

	Country Name	Country Code	Indicator Name	Indicator Code	1970	1971	...	2080	2085	2090	2095	2100	Unnamed: 69
0	Arab World	ARB	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
1	Arab World	ARB	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2.F	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
2	Arab World	ARB	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2.GPI	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
3	Arab World	ARB	Adjusted net enrolment rate, lower secondary, ...	UIS.NERA.2.M	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
4	Arab World	ARB	Adjusted net enrolment rate, primary, both sex...	SE.PRM.TENR	54.822121	54.894138	...	NaN	NaN	NaN	NaN	NaN	NaN



I~Presentation des jeux de données

Présentation de la démarche :

- Information contenues dans chaque ligne et colonnes.

- Lignes et colonnes à garder/filtrer.

- Connexions entre les jeux de données.

- Ordre de traitement :

1) Trier les indicateurs d'intérêt dans EdStatsSeries.csv

2) Relever les résultats des indicateurs retenus pour chaque pays et la dernière année de relevé dans EdStatsFootNote.csv

3) Garder les années et indicateurs d'intérêt pour chaque pays dans EdStatsData

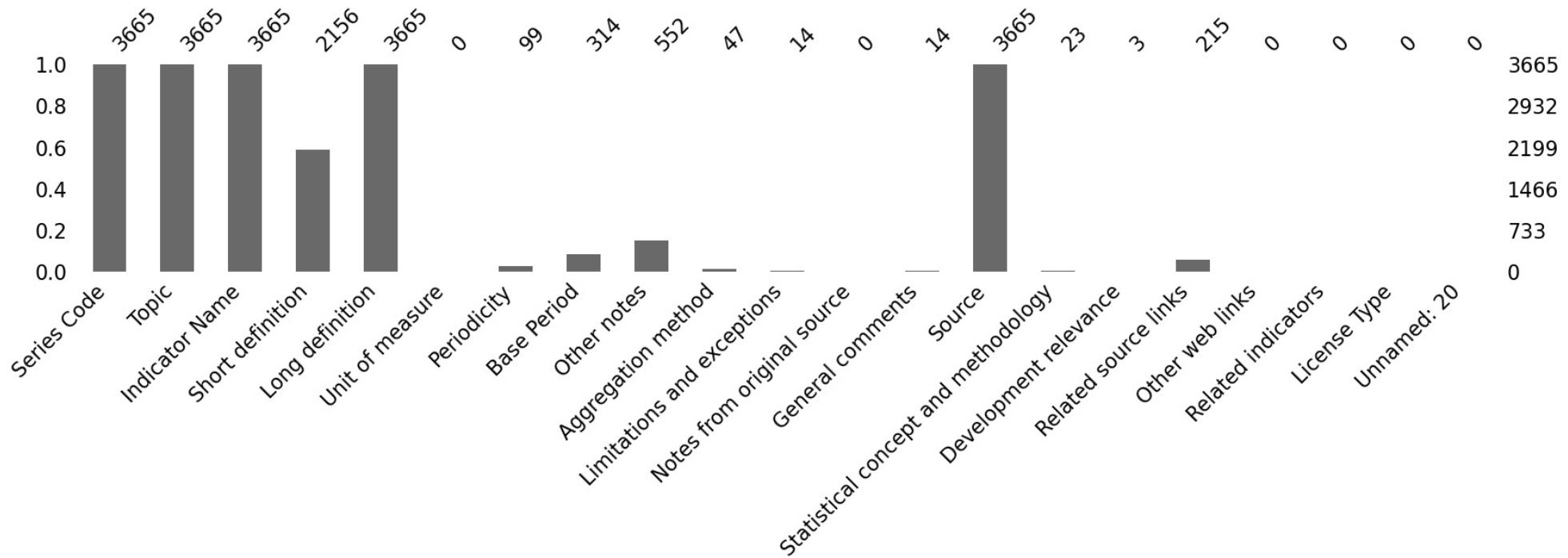
4) Aller chercher le nom complet et la région de chaque pays dans EdStatsCountry

5) Créer un nouveau jeu de données avec les informations d'intérêt uniquement.

II~Présentation des pré-analyses

Données manquantes :

EdStatsSeries.csv

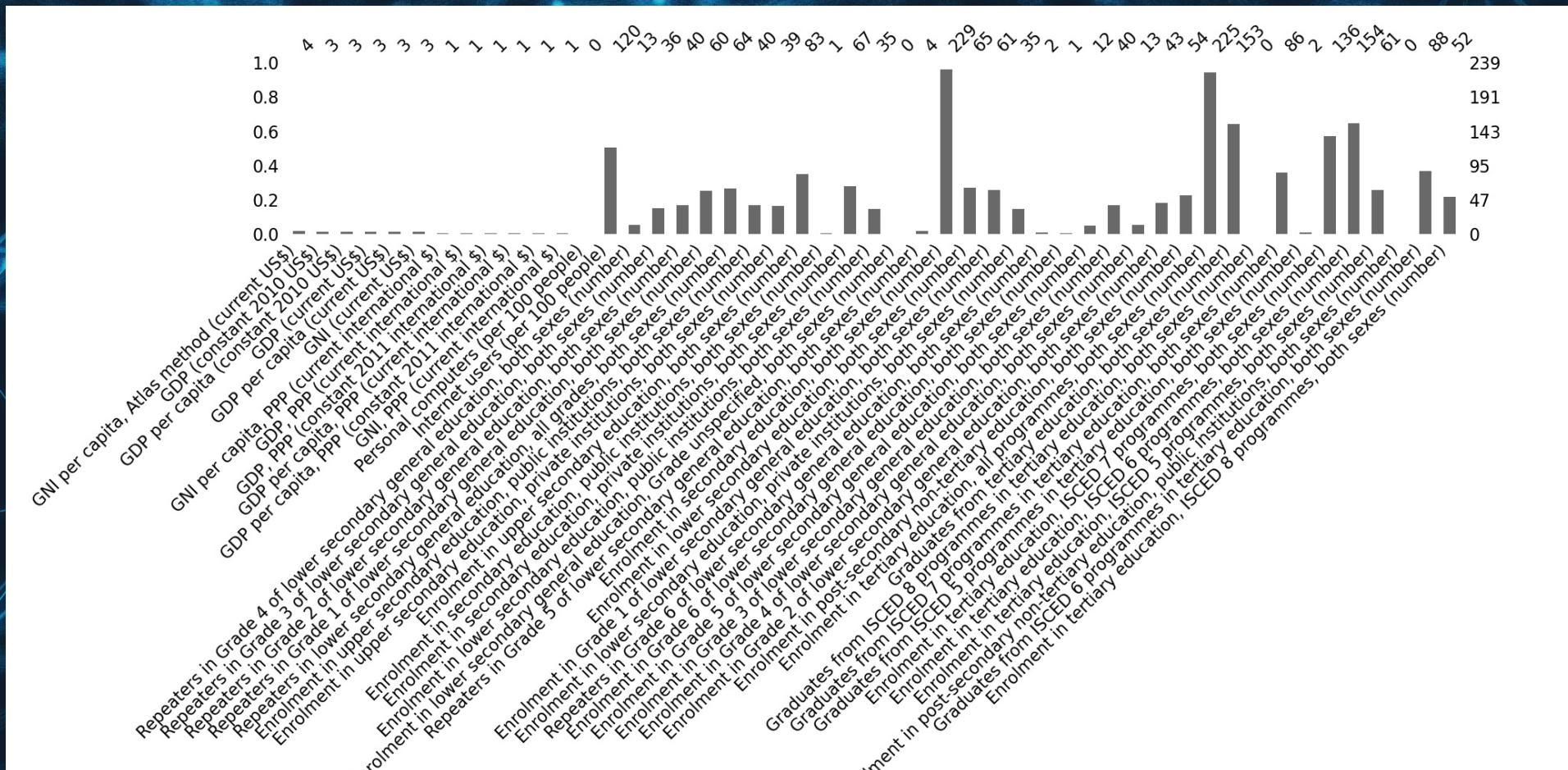


II~Présentation des pré-analyses

Tri des données :

EdStatsSeries : Tri des indicateurs, par sujet, puis par nom.

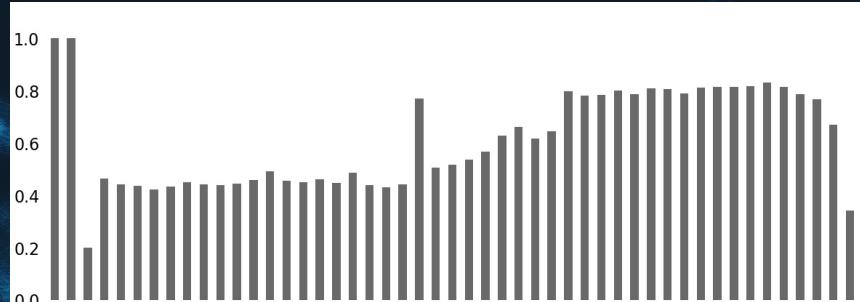
EdStatsFootNote : Filtre des indicateurs, tri des indicateurs et des pays sans infos.



II~Présentation des pré-analyses

Tri des données :

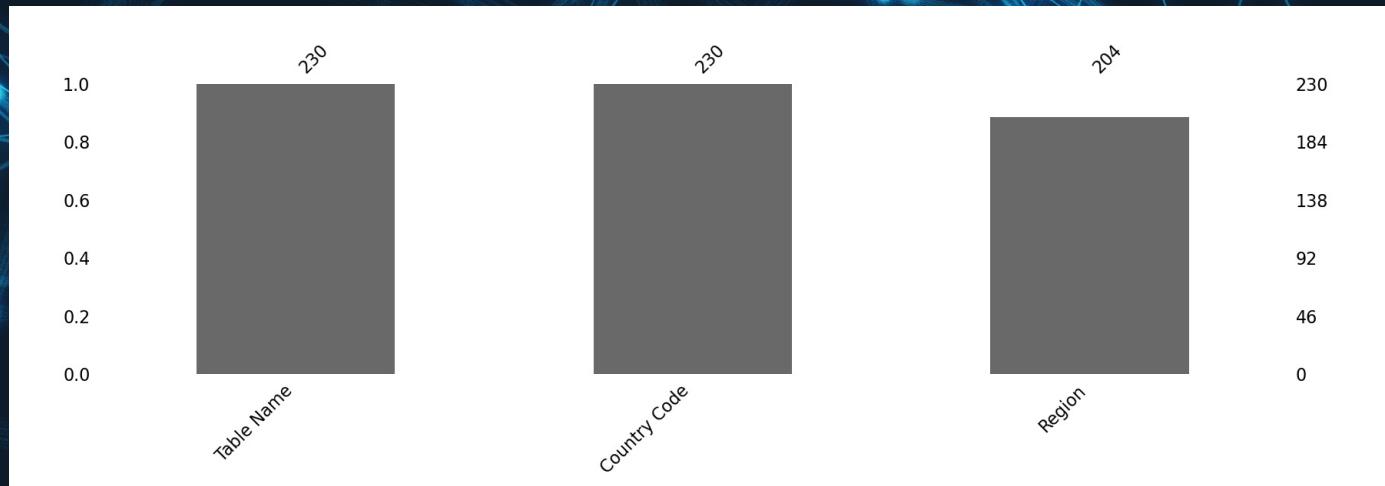
EdStatsData : Filtre des indicateurs, tri des années. (0 info post 2017)



En abscisse :
Pays, indicateurs, puis les résultats
par année.

En ordonnée :
Pourcentage de données manquantes.

EdStatsCountry : Filtre des pays, des codes et des régions.



II~Présentation des pré-analyses

Création de nouveaux tableaux :

Après avoir combiner les informations des jeux de données précédents en un seul jeu de donnée, on peut obtenir des tableaux avec les différents indicateurs en fonction des régions.

Secondaire

	Region	Moyenne (millions d'étudiants)	Mediane (millions d'étudiants)	Ecart-type (millions d'étudiants)	Minimum (millions d'étudiants)	Maximum (millions d'étudiants)
0	Toutes	2.84998	0.520733	11.3078	0.001289	129.542
1	Latin America & Caribbean	1.65764	0.278941	4.30436	0.002013	23.5018
2	South Asia	20.6089	2.98802	44.3464	0.028878	129.542
3	Sub-Saharan Africa	1.23025	0.567838	2.10411	0.007394	12.5328
4	Europe & Central Asia	1.48792	0.553356	2.42781	0.002407	10.9692
5	Middle East & North Africa	1.69289	0.749134	2.15981	0.029434	8.20833
6	East Asia & Pacific	4.45309	0.232003	14.8795	0.001289	86.1272
7	North America	8.97711	2.69751	13.2777	0.00404	24.2298

Tertiaire

	Region	Moyenne (millions d'étudiants)	Mediane (millions d'étudiants)	Ecart-type (millions d'étudiants)	Minimum (millions d'étudiants)	Maximum (millions d'étudiants)
0	Toutes	1.06604	0.162022	4.15526	0.000286	43.3674
1	Latin America & Caribbean	0.657886	0.112369	1.52363	0.000286	8.28547
2	South Asia	4.63477	0.376662	11.1303	0.006089	32.1074
3	Sub-Saharan Africa	0.165369	0.051573	0.290087	0.001003	1.51337
4	Europe & Central Asia	0.753204	0.266955	1.35106	0.000501	6.59242
5	Middle East & North Africa	0.71874	0.31275	1.15523	0.004705	4.80272
6	East Asia & Pacific	2.14977	0.154865	7.64566	0.000301	43.3674
7	North America	6.91495	1.21216	10.9432	0.000973	19.5317

II~Présentation des pré-analyses

Création de nouveaux tableaux :

Après avoir combiner les informations des jeux de données précédents en un seul jeu de donnée, on peut obtenir des tableaux avec les différents indicateurs en fonction des régions.

Secondaire et Tertiaire

	Region	Moyenne (millions d'étudiants)	Mediane (millions d'étudiants)	Ecart-type (millions d'étudiants)	Minimum (millions d'étudiants)	Maximum (millions d'étudiants)
0	Toutes	3.88989	0.69895	15.1621	0.001289	161.649
1	Latin America & Caribbean	2.31553	0.436891	5.79306	0.002299	31.7873
2	South Asia	25.2437	3.34212	55.4656	0.034967	161.649
3	Sub-Saharan Africa	1.39562	0.61719	2.37892	0.008429	14.0461
4	Europe & Central Asia	2.21215	0.854782	3.71355	0.002776	17.0321
5	Middle East & North Africa	2.41163	1.10764	3.20975	0.04265	11.0772
6	East Asia & Pacific	6.4186	0.388308	22.1397	0.001289	129.495
7	North America	15.8921	3.90967	24.2145	0.005013	43.7615

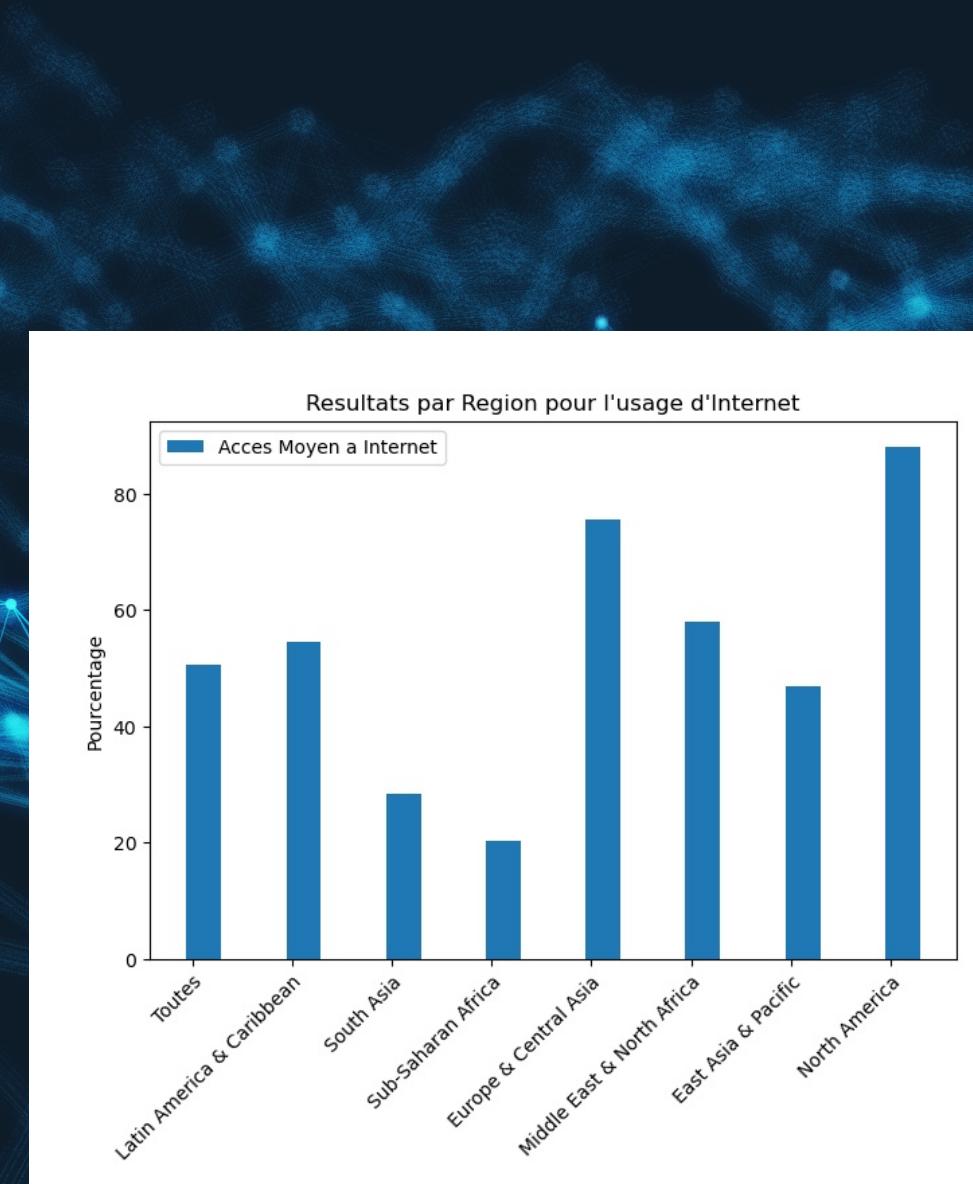
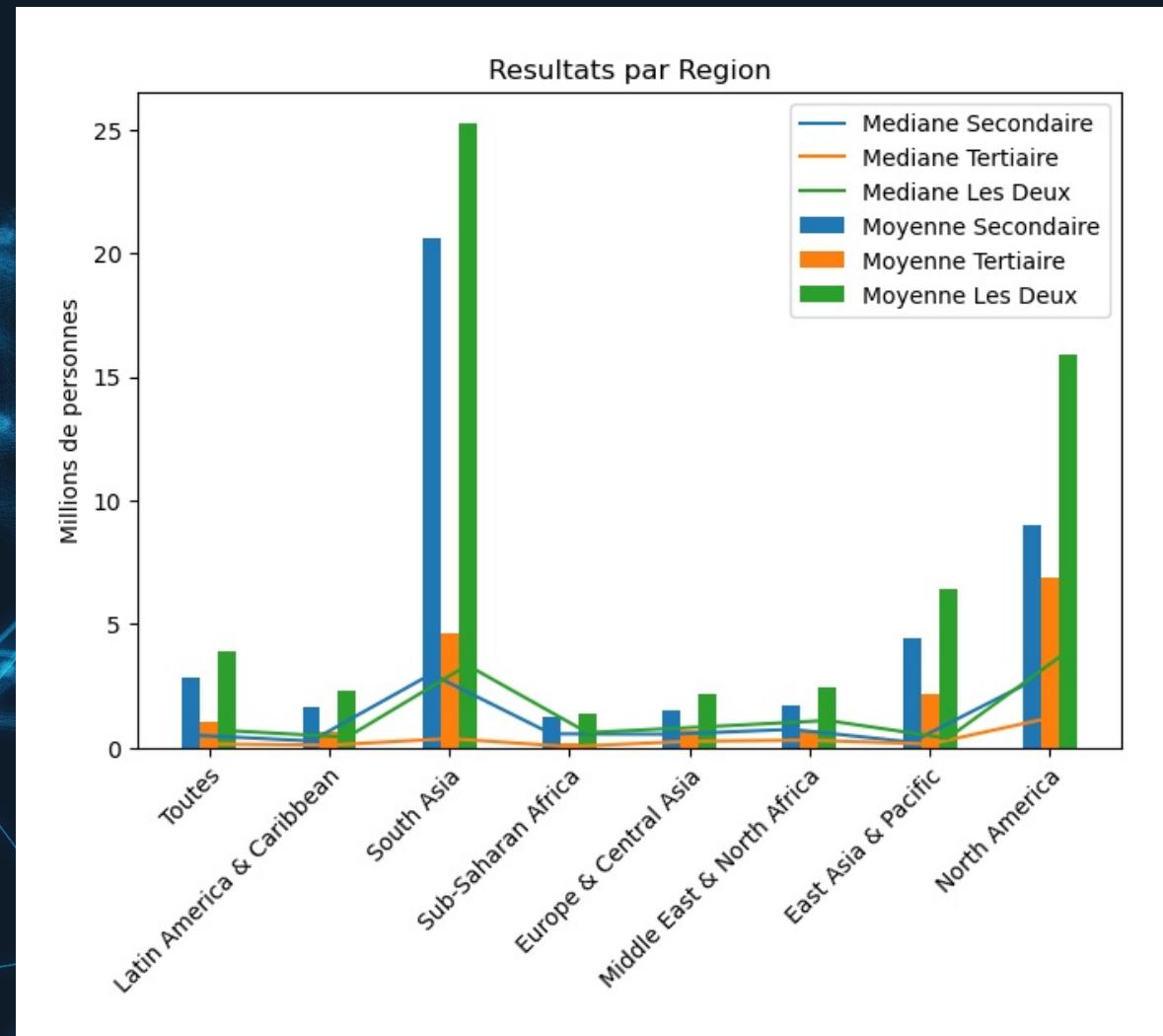
Internet

	Region	Moyenne (pourcentage)	Mediane (pourcentage)	Ecart-type (pourcentage)	Minimum (pourcentage)	Maximum (pourcentage)
0	Toutes	50.5096	53.2131	28.8403	0	98.24
1	Latin America & Caribbean	54.5233	56.9965	19.9942	0	93.5425
2	South Asia	28.3137	24.618	16.0243	10.5957	59.0926
3	Sub-Saharan Africa	20.3121	18	14.5906	1.17712	56.5147
4	Europe & Central Asia	75.5148	76.4451	16.9276	20.47	98.24
5	Middle East & North Africa	57.9574	61.1784	25.7162	13.1349	98
6	East Asia & Pacific	46.9428	46.0092	29.0181	0	92.7165
7	North America	88.0056	89.84	11.0267	76.1767	98



III~Conclusions

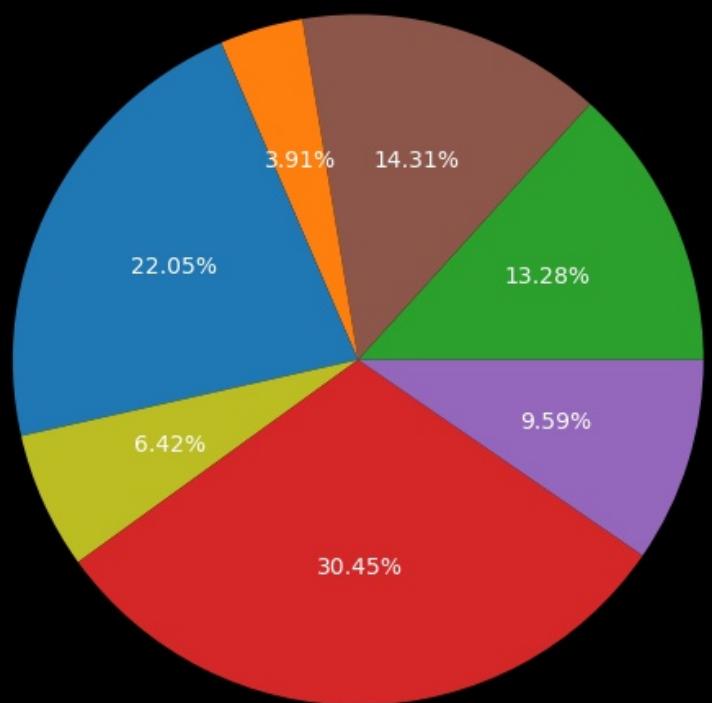
Information des différents indicateurs:



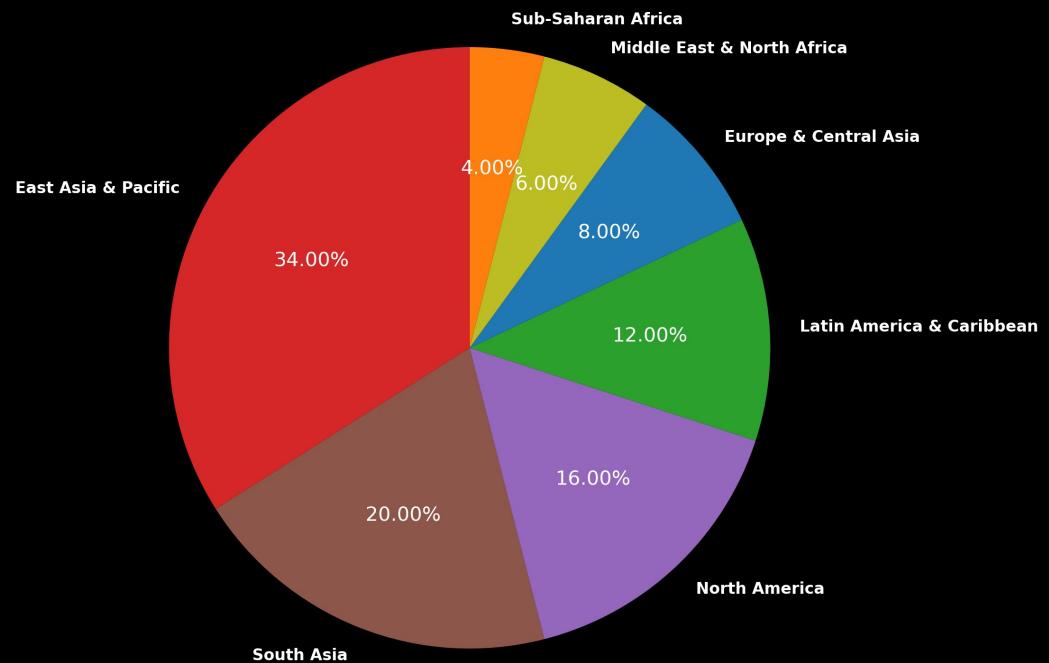
III~Conclusions

Prépondérance des Regions :

Etudiant ayant Internet en fonction des Regions
(en millions d'étudiants)



Representation des Regions dans le top 50



III~Conclusions

Top 10 des pays:

Secondaire
et Tertiaire

	Pays	Region	Resultat (millions d'étudiants)	Accès à Internet
Top 10				
1	India	South Asia	161.649476	29.547163
2	China	East Asia & Pacific	129.494592	53.200000
3	United States	North America	43.761504	76.176737
4	Brazil	Latin America & Caribbean	31.787259	59.682747
5	Indonesia	East Asia & Pacific	28.864315	25.366301
6	Turkey	Europe & Central Asia	17.032064	58.347734
7	Bangladesh	South Asia	16.635126	18.246938
8	Mexico	Latin America & Caribbean	16.412713	59.540446
9	Russian Federation	Europe & Central Asia	15.977026	76.409085
10	Nigeria	Sub-Saharan Africa	14.046124	25.670000

	Pays	Region	Resultat (millions d'étudiants)
Top 10			
1	China	East Asia & Pacific	68.891123
2	India	South Asia	47.762834
3	United States	North America	33.336086
4	Brazil	Latin America & Caribbean	18.971509
5	Russian Federation	Europe & Central Asia	12.207899
6	Japan	East Asia & Pacific	10.202749
7	Turkey	Europe & Central Asia	9.937823
8	Mexico	Latin America & Caribbean	9.772203
9	Germany	Europe & Central Asia	9.046047
10	United Kingdom	Europe & Central Asia	8.444452

Prise en compte de
l'accès à Internet



Merci de votre Attention

