

Analysez des données de systèmes éducatifs

Etudiant: Etienne NGENZI
Mentor : Julia Wabant

December 20, 2022

Introduction

Introduction générale

- 1 Une start-up de la EdTech, nommée academy, qui propose des contenus de formation en ligne pour un public de niveau lycée et université.
- 2 Un project d'Expansion à l'international.
- 3 Se servir des données sur l'éducation de la banque mondiale d'informer le projet d'expansion.



- Cet ensemble de données contient-il les indicateurs pertinents ?

La qualité de données

La qualité de jeu de données

Analyser fichier par fichier

- Analyser les fichiers donnés pour vérifier s'il ya les valeurs manquantes ou les doublons.

La qualité de jeu de donnée

Nom du fichier	Valeur manquante	Doublons (localement)	Doublons (globalement)
EdStatsCountry-Series	Non	Oui	Non
EdStatsCountry	Oui	Non	Non
EdStatsData	Oui	oui	non
EdStatsFootNote	Non	oui	Non
EdStatsSeries	Non	oui	non

- Le fichier **EdStatsData.csv** contient beaucoup de valeurs manquantes. Pour mieux l'analyser demander de faire en sorte que le fichier reste uniquement avec les colonnes les plus importantes.

Déscriptions des informations contenues dans le jeu de données

- Par l'utilisation des méthodes **NombredेRange = len(dataframe.axes[0]) et NombredеColonne = len(dataframe.axes[1])** permettent de savoir Nombre de rangées et de colonne respectivement.
- Pour cela le fichier **EdStatsData.csv** contient 886930 rangées et 70 colonnes.

Les données suivent cette distribution statistique

	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	...
count	7.228800e+04	3.553700e+04	3.561900e+04	3.554500e+04	3.573000e+04	8.730600e+04	3.748300e+04	3.757400e+04	3.757600e+04	3.680900e+04	...
mean	1.974772e+09	4.253638e+09	4.592365e+09	5.105006e+09	5.401493e+09	2.314288e+09	5.731808e+09	6.124437e+09	6.671489e+09	7.436724e+09	...
std	1.211687e+11	1.804814e+11	1.914083e+11	2.059170e+11	2.112150e+11	1.375059e+11	2.215546e+11	2.325489e+11	2.473986e+11	2.660957e+11	...
min	-1.435564e+00	-1.594625e+00	-3.056522e+00	-4.032582e+00	-4.213563e+00	-3.658569e+00	-2.950945e+00	-3.174870e+00	-3.558749e+00	-2.973612e+00	...
25%	8.900000e-01	8.853210e+00	9.240920e+00	9.595200e+00	9.861595e+00	1.400000e+00	9.312615e+00	9.519913e+00	1.000000e+01	1.000000e+01	...
50%	6.317724e+00	6.316240e+01	6.655139e+01	6.969595e+01	7.087760e+01	9.677420e+00	7.101590e+01	7.133326e+01	7.290512e+01	7.510173e+01	...
75%	6.251250e+01	5.655200e+04	5.863650e+04	6.202900e+04	6.383675e+04	7.854163e+01	5.682800e+04	5.739175e+04	5.940425e+04	6.411500e+04	...
max	1.903929e+13	1.986457e+13	2.100916e+13	2.238367e+13	2.282991e+13	2.300634e+13	2.424128e+13	2.521383e+13	2.622101e+13	2.730873e+13	...

- Cette description est basée sur les données non cleanée ce qui demander de faire la même chose après avoir nettoyer le jeu de donnée pour éliminer les valeurs manquantes.

La qualité de jeu de données

Nettoyage du jeu de données

- Nous avons effacé toutes les lignes contenant des valeurs manquantes supérieures à 30.
- Pour cela le nouveau fichier contient 13294 rangées et 49 colonnes.
- Globalement, le jeux de donnée cleanée ne contient ni les valeur manquante ni les doublons.

Sa nouvelle distribution statique

	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	...
count	1.329400e+04	...									
mean	8.790961e+09	9.311324e+09	1.007161e+10	1.117070e+10	1.190592e+10	1.243138e+10	1.323111e+10	1.416613e+10	1.538324e+10	1.676860e+10	...
std	2.509449e+11	2.622878e+11	2.785364e+11	2.993312e+11	3.081776e+11	3.135374e+11	3.313535e+11	3.482152e+11	3.695997e+11	3.930003e+11	...
min	-1.435564e+00	-1.594625e+00	-3.056522e+00	-4.032582e+00	-4.213563e+00	-3.658569e+00	-2.950945e+00	-3.174870e+00	-3.558749e+00	-2.973612e+00	...
25%	4.881782e+01	4.884217e+01	4.886966e+01	4.889246e+01	4.893052e+01	4.891297e+01	4.901705e+01	4.901154e+01	4.900841e+01	4.912048e+01	...
50%	5.874750e+04	5.975636e+04	6.152250e+04	6.294000e+04	6.444250e+04	6.516750e+04	6.604200e+04	6.697100e+04	6.859200e+04	6.961700e+04	...
75%	1.200298e+06	1.188987e+06	1.214774e+06	1.246351e+06	1.273535e+06	1.300758e+06	1.332663e+06	1.359193e+06	1.396706e+06	1.426938e+06	...
max	1.903929e+13	1.986457e+13	2.100916e+13	2.238367e+13	2.282991e+13	2.300634e+13	2.424128e+13	2.521383e+13	2.622101e+13	2.730873e+13	...

- Il y a une amélioration remarquable par rapport aux données non nettoyées.

Monde Arabe

Ordres de grandeurs des indicateurs statistiques dans le monde Arabe

- Faisons un tour d'horizon du monde arabe.

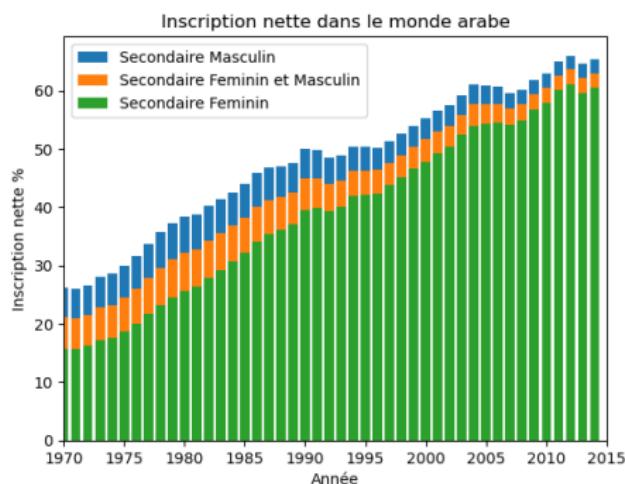
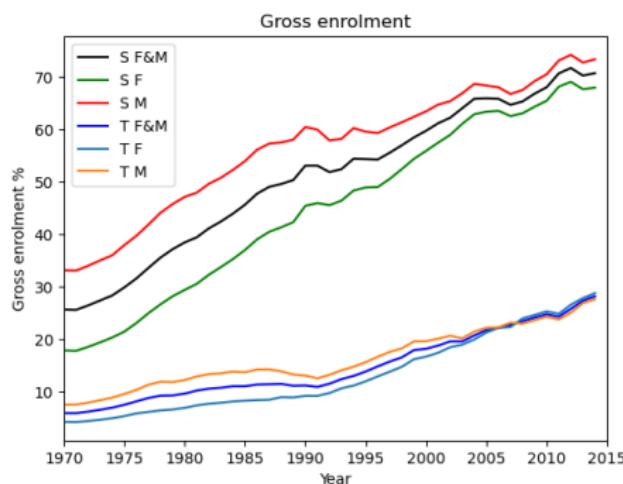
	index	1970	1971	1972	1973	1974	1975	1976	1977	1978	...
count	14.0000	14.000000	14.000000	14.000000	14.000000	14.000000	14.000000	14.000000	14.000000	14.000000	...
mean	52.5000	17.199850	17.173103	17.673824	18.349863	18.806719	19.703771	20.710240	21.861094	23.032889	...
std	4.1833	15.573100	15.575986	15.957483	16.340001	16.566217	17.099978	17.669348	18.248529	18.913956	...
min	46.0000	0.538990	0.537230	0.545050	0.553610	0.555900	0.557770	0.565900	0.542400	0.539120	...
25%	49.2500	4.643845	4.619408	4.822425	5.100525	5.423670	5.857510	6.412783	6.787303	7.108217	...
50%	52.5000	16.818619	16.751579	17.355227	18.262335	18.971873	20.084704	21.543420	23.286285	24.904601	...
75%	55.7500	25.999037	25.913503	26.495881	27.941299	28.509172	29.949754	31.429029	33.360433	35.620881	...
max	59.0000	54.792889	54.865051	56.180119	57.237358	57.960880	59.334751	60.968681	61.890991	62.661449	...

- En observant ce tableau on s'aperçoit qu'il y a une tendance générale.
- On voit qu'en moyenne les valeurs s'améliorent avec le temps.

Informations pertinentes

Ecoles secondaire et supérieures

- Inscription des élèves dans les écoles secondaire et supérieures.

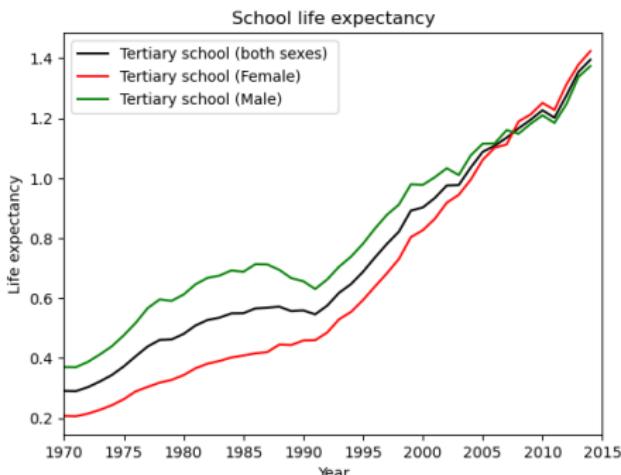
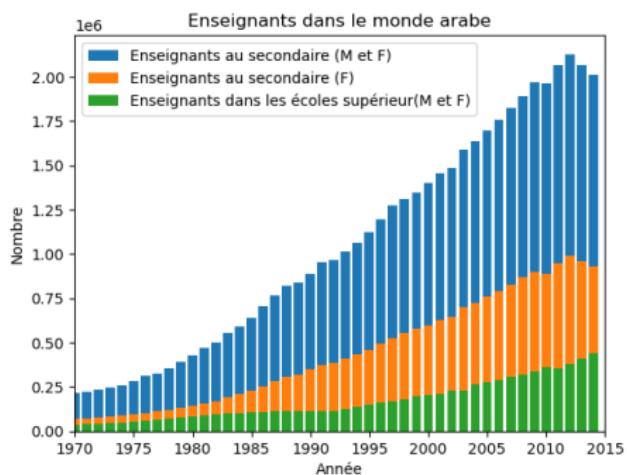


- Au fil des années, le nombre d'inscriptions d'étudiants pour les hommes et les femmes augmente proportionnellement.
- On voit aussi que les inscriptions brutes augmentent proportionnellement avec les inscriptions nettes.

Informations pertinentes

Sélection des informations pertinentes pour le monde arabe continue

■ Comparaison des espérances de vie scolaire entre garçons et filles.



- Augmentation de nombre des enseignements dans les écoles secondaires et supérieures.
- Augmentation de nombre des enseignements pourra signaler l'augmentation des étudiants aussi.

Asie de l'Est Pacifique

Ordres de grandeurs des indicateurs statistiques

- Faisons un tour d'horizon de la zone de l'Asie de l'Est Pacifique.

	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	...
count	8.600000e+01	...									
mean	4.248305e+10	4.522184e+10	5.035141e+10	5.736585e+10	6.013764e+10	6.344201e+10	6.737068e+10	7.420074e+10	8.419871e+10	9.007871e+10	...
std	3.108894e+11	3.275451e+11	3.544435e+11	3.868580e+11	3.934955e+11	4.098203e+11	4.299962e+11	4.603581e+11	5.021468e+11	5.342898e+11	...
min	1.445604e-01	1.406968e-01	1.503386e-01	1.609799e-01	1.766202e-01	1.925374e-01	2.059662e-01	2.121178e-01	2.240313e-01	2.412815e-01	...
25%	5.213744e+01	5.212930e+01	5.212852e+01	5.505161e+01	5.366293e+01	5.439682e+01	5.759412e+01	5.829665e+01	5.814709e+01	5.647852e+01	...
50%	1.364909e+07	6.483867e+06	6.775686e+06	6.930774e+06	7.281620e+06	7.776966e+06	7.810960e+06	7.677862e+06	7.650738e+06	7.681853e+06	...
75%	6.213945e+07	6.379551e+07	6.506097e+07	6.685456e+07	6.827744e+07	6.924903e+07	7.222367e+07	8.500663e+07	9.722104e+07	9.958238e+07	...
max	2.835822e+12	2.981554e+12	3.204914e+12	3.452225e+12	3.469241e+12	3.591370e+12	3.744744e+12	3.940676e+12	4.163032e+12	4.405500e+12	...

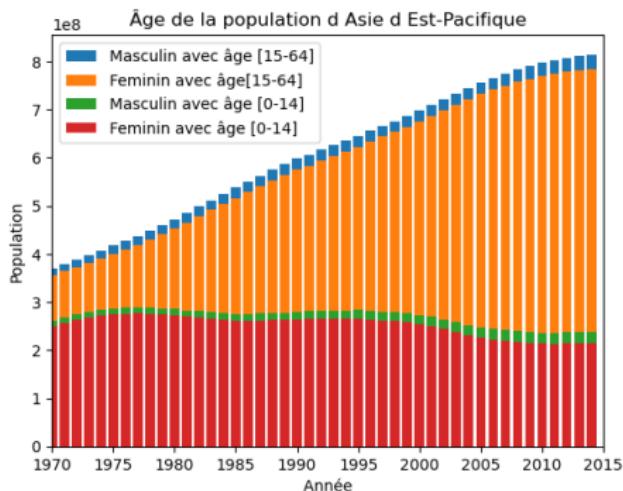
- En observant ce tableau on s'aperçoit qu'il y a une tendance générale.
- On voit qu'en moyenne les valeurs s'améliorent avec le temps.

OO OOO OOOO OOO●O OOO OOO OOOO OOO

Analysons la population

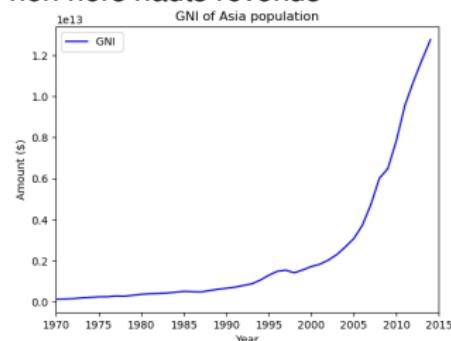
La population de l'Asie de l'Est Pacifique

■ La population de l'Asie de l'Est Pacifique par tranche d'âge

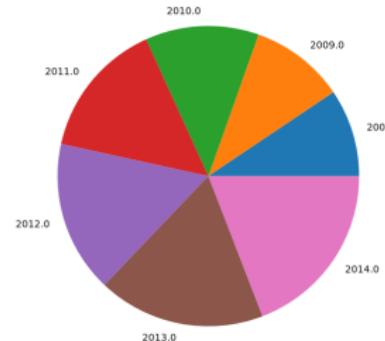


■ Le nombre de la population ayant l'âge officiel pour accéder à l'enseignement supérieur et le pouvoir économique de la population augmentent simultanément.

Pouvoir économique de la population non hors hauts revenus



GDP per capita de la population

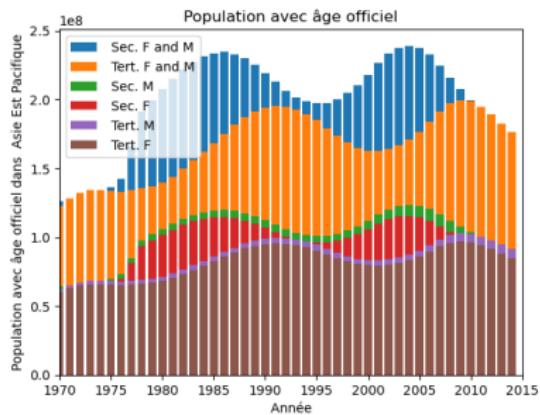


Analysons la population

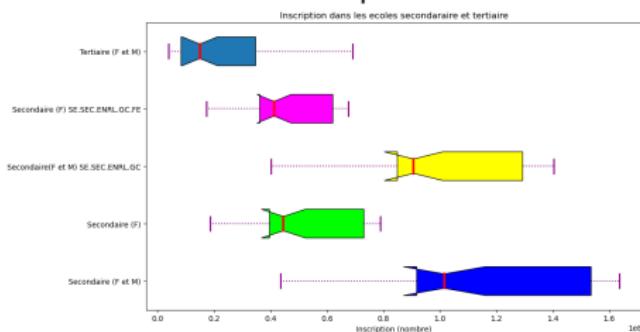
Capacité d' aller à l'école

- Analyser la population de l'Asie de l'Est Pacifique pour sa capacité à aller à l'école

Population d'âge officiel pour l'école secondaire et supérieur



Inscription des élèves dans les écoles secondaire et supérieures.



- Il y a une diminution non justifiée de la population en âge officiel d'aller à l'école.

Zone euro

Ordres de grandeurs des indicateurs statistiques

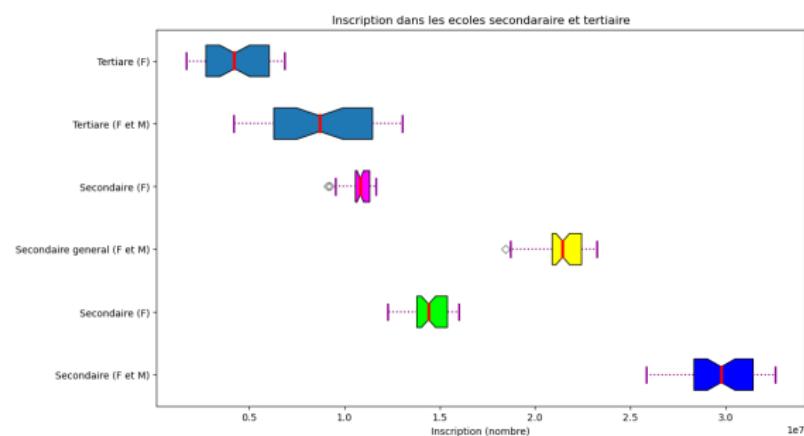
- Faisons un tour d'horizon de la zone Euro.

	index	1970	1971	1972	1973	1974	1975	1976	1977	1978	...
count	166.000000	1.660000e+02	1.660000e+02	1.660000e+02	1.660000e+02	1.660000e+02	1.660000e+02	1.660000e+02	1.660000e+02	1.660000e+02	...
mean	378.500000	3.8110986e+10	4.030388e+10	4.368422e+10	4.889098e+10	5.184814e+10	5.407735e+10	5.667413e+10	6.043002e+10	6.648584e+10	...
std	48.064193	3.965740e+11	4.129082e+11	4.357636e+11	4.680452e+11	4.865178e+11	4.901118e+11	5.143658e+11	5.364366e+11	5.678077e+11	...
min	296.000000	5.461324e-01	5.623531e-01	6.947973e-01	6.600723e-01	5.897176e-01	5.277760e-01	4.483749e-01	3.917042e-01	3.917605e-01	...
25%	337.250000	2.231199e+01	2.229884e+01	2.226551e+01	2.262282e+01	2.310286e+01	2.382849e+01	2.396815e+01	2.377323e+01	2.349929e+01	...
50%	378.500000	1.071514e+02	1.071750e+02	1.028977e+02	1.011118e+02	1.014398e+02	1.011844e+02	1.020553e+02	1.011576e+02	1.020365e+02	...
75%	419.750000	5.241094e+06	5.168972e+06	5.122658e+06	5.131104e+06	5.174924e+06	5.186930e+06	5.157870e+06	5.045146e+06	4.980661e+06	...
max	461.000000	5.035877e+12	5.228214e+12	5.485623e+12	5.824099e+12	6.011488e+12	5.965894e+12	6.265365e+12	6.458288e+12	6.660507e+12	...

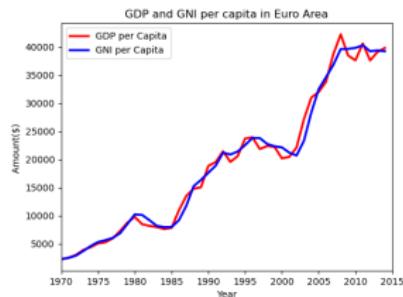
- En observant ce tableau on s'aperçoit qu'il y a une tendance générale.
- On voit qu'en moyenne les valeurs s'améliorent avec le temps.

Inscription dans les écoles secondaires et supérieures

Inscription à l'école secondaire et tertiaire



Le pouvoir d'achat de la population



- Bien que le pouvoir d'achat de la population augmente, les inscriptions dans les écoles secondaires diminuent alors que dans les écoles tertiaires augmentent.

Amérique latine et caraïbes

Ordres de grandeurs des indicateurs statistiques dans la zone Amérique latine et caraïbes

■ Faisons un tour d'horizon de la zone Amérique latine et Caraïbes

	index	1970	1971	1972	1973	1974	1975	1976	1977	1978	...
count	176.000000	1.760000e+02	...								
mean	1365.500000	8.981676e+09	9.674105e+09	1.051793e+10	1.196984e+10	1.354021e+10	1.425500e+10	1.532889e+10	1.625912e+10	1.733292e+10	...
std	50.950957	9.690078e+10	1.035581e+11	1.116156e+11	1.228331e+11	1.327694e+11	1.388156e+11	1.479679e+11	1.555044e+11	1.625751e+11	...
min	1278.000000	4.657030e-01	4.759441e-01	4.922631e-01	5.190182e-01	5.491334e-01	5.916095e-01	6.114561e-01	6.231407e-01	6.427446e-01	...
25%	1321.750000	1.801597e+01	1.760373e+01	1.707108e+01	1.635453e+01	1.570596e+01	1.555048e+01	1.508687e+01	1.485455e+01	1.523154e+01	...
50%	1365.500000	1.175403e+02	1.215507e+02	1.222866e+02	1.230331e+02	1.231421e+02	1.229001e+02	1.207864e+02	1.224767e+02	1.234270e+02	...
75%	1409.250000	5.903905e+06	5.949162e+06	6.068052e+06	6.456447e+06	6.522467e+06	6.785675e+06	6.766144e+06	6.795455e+06	7.063421e+06	...
max	1453.000000	1.268525e+12	1.353875e+12	1.456908e+12	1.591848e+12	1.698179e+12	1.771236e+12	1.881888e+12	1.970540e+12	2.042588e+12	...

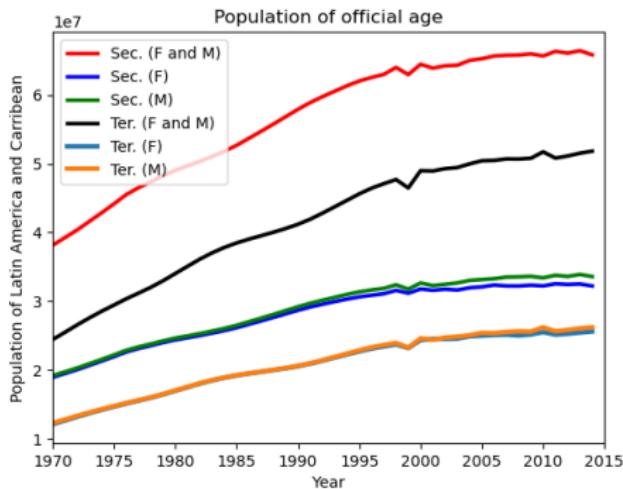
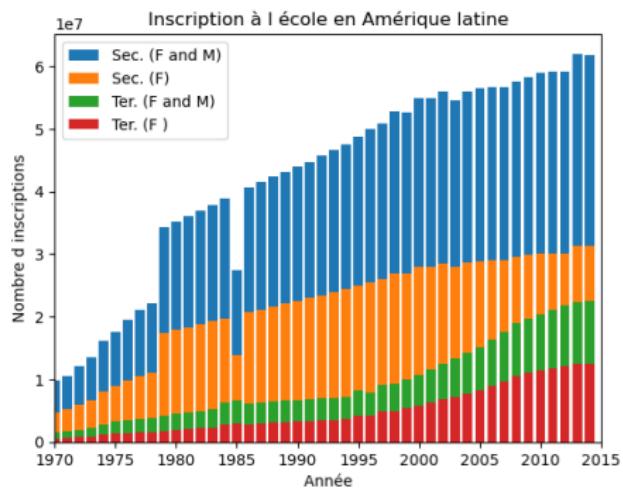
■ En observant ce tableau on s'aperçoit qu'il y a une tendance générale.

■ On voit qu'en moyenne les valeurs s'améliorent avec le temps.

Inscription dans les écoles secondaires et supérieures

Inscription à l'école secondaire et tertiaire

- Les inscriptions dans les écoles secondaires et supérieures augmentent chaque année.



- L'augmentation des inscriptions est justifiée par l'augmentation du nombre de la population en âge officiel de s'inscrire dans les écoles secondaires et supérieures

Afrique sub-saharienne

Ordres de grandeurs des indicateurs Statistiques dans la zone Afrique sub-saharienne

- Faisons un tours d'horizon de la zone Afrique Sub-Saharienne.

	index	1970	1971	1972	1973	1974	1975	1976	1977	1978	...
count	124.000000	1.240000e+02	...								
mean	2958.500000	4.212415e+09	4.463991e+09	4.687397e+09	5.169118e+09	5.934418e+09	6.133542e+09	6.527595e+09	6.871930e+09	7.146079e+09	...
std	35.939764	3.600787e+10	3.858746e+10	3.978996e+10	4.206055e+10	4.622301e+10	4.674887e+10	4.941357e+10	5.115002e+10	5.196728e+10	...
min	2897.000000	5.519595e-02	5.507142e-02	5.522100e-02	5.573494e-02	5.779811e-02	5.928419e-02	6.096249e-02	6.353772e-02	6.640881e-02	...
25%	2927.750000	1.737391e+01	1.758918e+01	1.760791e+01	1.793857e+01	1.825818e+01	1.866445e+01	1.904025e+01	1.956957e+01	2.007058e+01	...
50%	2958.500000	5.626091e+05	5.855985e+05	6.172025e+05	6.593080e+05	7.037991e+05	7.540530e+05	7.917827e+05	8.522632e+05	9.455450e+05	...
75%	2989.250000	9.181925e+06	9.410158e+06	9.622299e+06	9.854444e+06	1.010497e+07	1.039071e+07	1.066959e+07	1.084667e+07	1.114600e+07	...
max	3020.000000	3.915956e+11	4.207424e+11	4.319465e+11	4.507241e+11	4.862904e+11	4.881537e+11	5.117482e+11	5.238806e+11	5.227152e+11	...

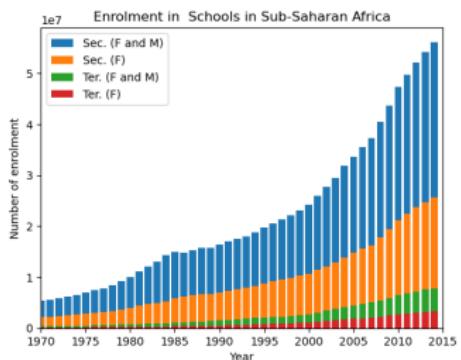
- En observant ce tableau on s'aperçoit qu'il y a une tendance générale.

- On voit qu'en moyenne les valeurs s'améliorent avec le temps.

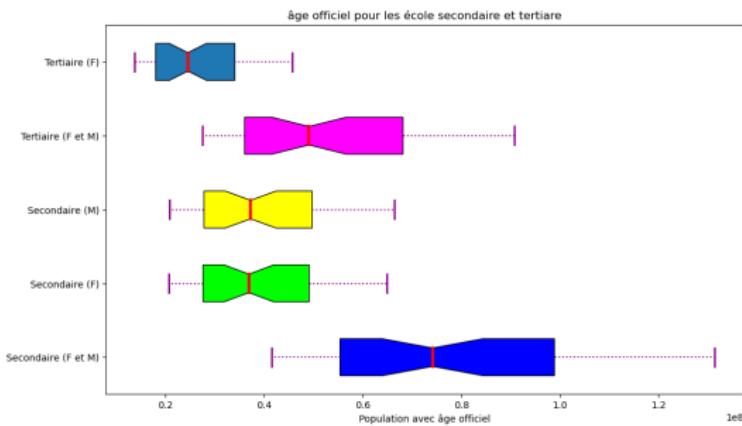
Inscription dans les écoles secondaires et supérieurs

Inscription à l'école secondaire et tertiaire

Augmentation des inscriptions



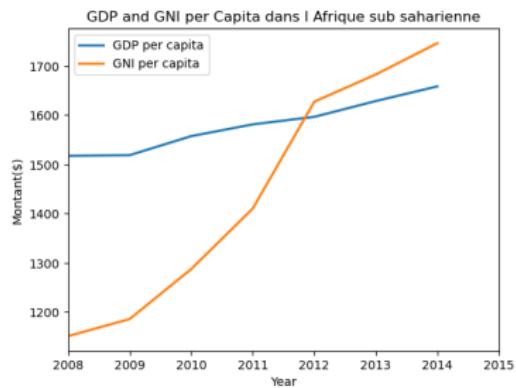
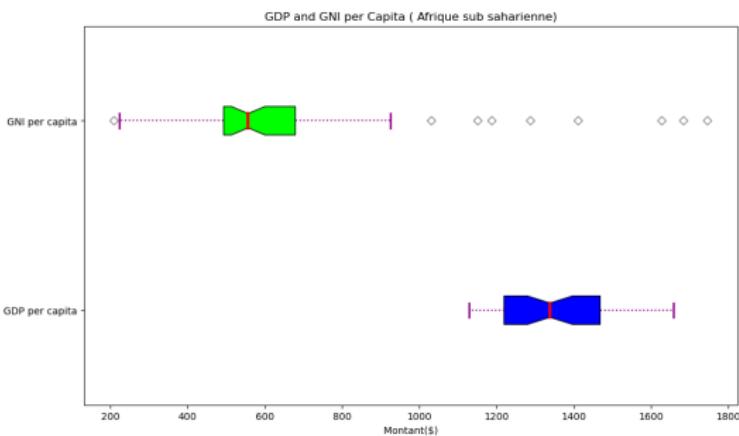
Augmentation de la population



- Il existe une forte corrélation entre l'augmentation des inscriptions et le nombre de personnes ayant l'âge officiel pour fréquenter les écoles secondaires et supérieures

Puissance économique

Augmentation du pouvoir économique en afrique sous saharienne



- GNI per capita contient beaucoup des outliers si compare avec le GDP per capita.
- Il est intéressant de noter que le GGP et le RNB augmentent fortement

Conclusion

Conclusions

- ① En Afrique sub-saharienne, il y a une augmentation exponentielle des inscriptions d'élèves dans les écoles secondaires et supérieures.
- ② Il y a une augmentation exponentielle du nombre de populations ayant l'âge officiel pour fréquenter les écoles secondaires et supérieures.
- ③ On voit qu'au fil du temps, le pouvoir économique des populations augmente fortement.
- ④ Pour toutes ces raisons, je peux fortement conseiller l'entreprise d'investir en Afrique sub-saharienne.

Les données manquantes

- 1 Manque des données sur les compétences de la population sur les technologies.



- 2 Manque des données détaillées sur le taux d'alphabétisation.