



Projet d'Expansion à l'International

Yassir Boudda

Sommaire

1) Valider la qualité du jeu de données

2) Décrire les informations contenues

3) Sélectionner les informations pertinentes

4) Déterminer les ordres de grandeurs par zone géographique

Mission évaluer les données EdStats pour l'expansion

Objectif:  Données Banque Mondiale pertinentes pour guider l'expansion.

Analyse: Comparer les pays cibles aux pays similaires avec les indicateurs choisis issue des Données Banque Mondiale(ex: alphabétisation, scolarisation).

Décision: Identifier les marchés potentiels pour l'expansion. Conseil d'administration:
Présentation claire avec graphiques clés

1) Valider la qualité du jeu de données :

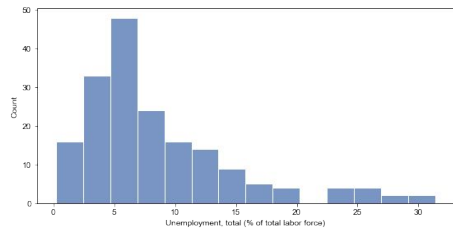
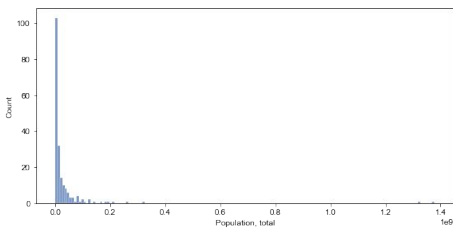
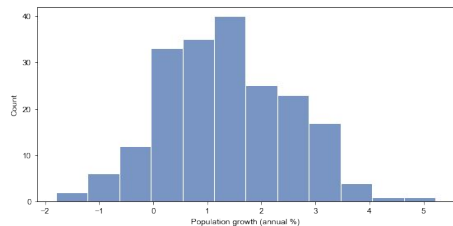
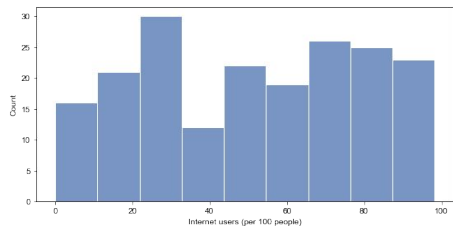
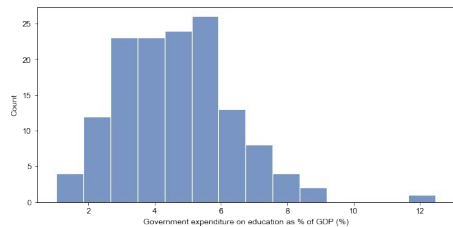
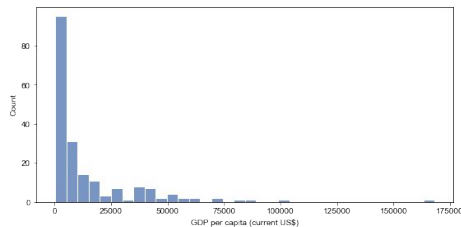
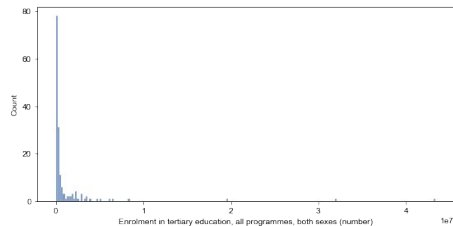
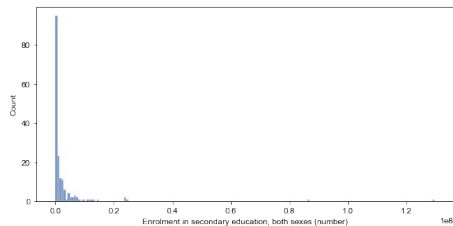
Il y a quelques données manquantes pour certains indicateurs et certains pays, comme le montre l'analyse des valeurs manquantes effectuée précédemment.

Cependant, le taux de données manquantes semble raisonnable et nous avons réussi à prédire une grande partie de ces valeurs manquantes.

Pour remplir les données manquantes, nous avons utilisé la méthode de l'iterative imputer. Cet algorithme effectue une régression linéaire pour chaque variable en fonction de toutes les autres afin de prédire les valeurs manquantes. Cela nous a permis de remplir les données manquantes de manière robuste.

Visualization de certain indicateurs en histogramme

Ces histogrammes illustrent la progression générale des taux d'alphabétisation et d'inscription à l'école, mais également les disparités qui persistent entre les pays.

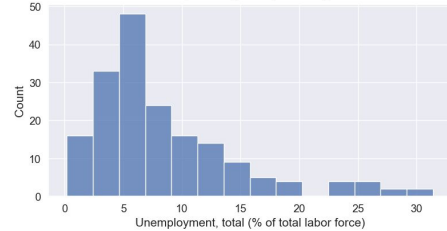
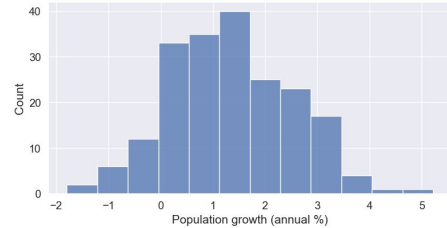
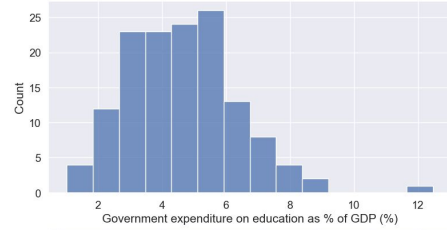
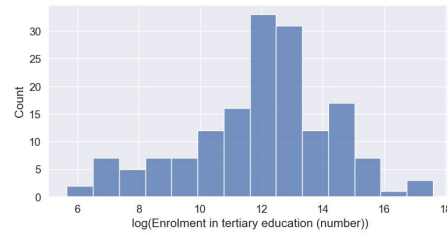
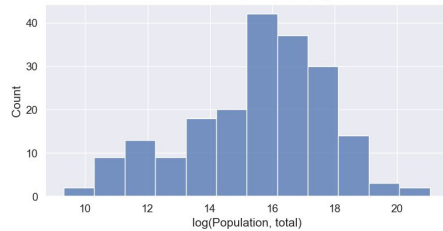
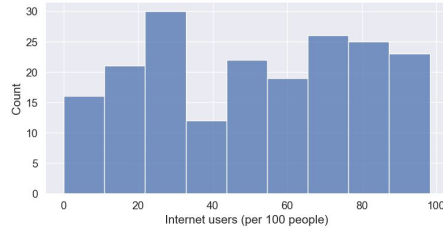
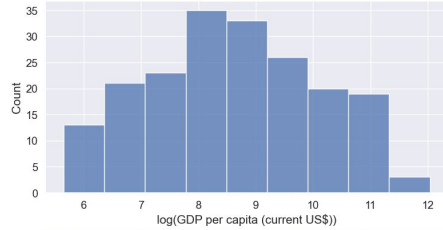
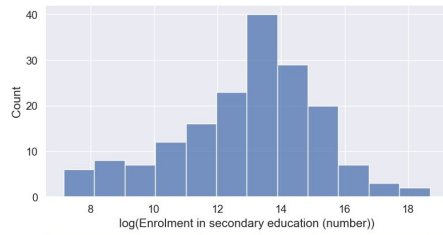


2) Décrire les informations contenues :

- Le jeu de données contient 199 pays et 8 indicateurs sélectionnés (variables).
- Pour chaque pays, on dispose de la valeur de ces 8 indicateurs sur différentes années (de 2012 à 2016).

Visualization Indicateurs lissé en histogramme

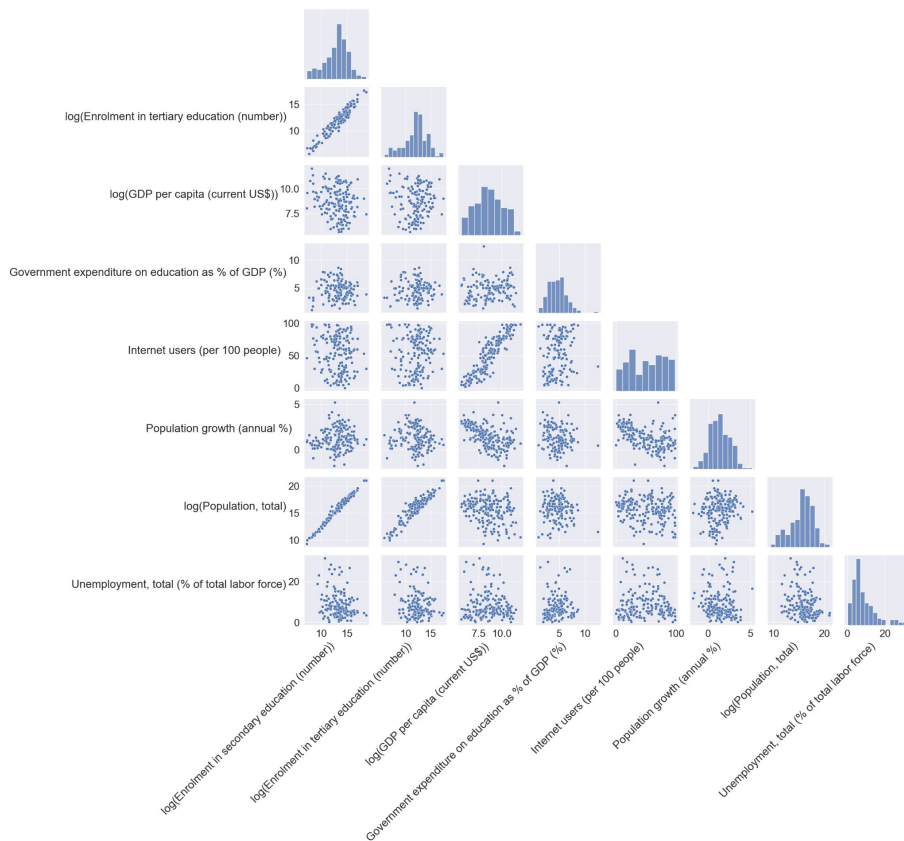
Les données ont été lissées à l'aide de la transformation logarithmique. Cela permet de réduire la volatilité des données et de mettre en évidence les tendances générales.



Indicateurs et corrélation

Ces graphiques présentent la relation entre le taux d'alphabétisation et le PIB par habitant, avec les données lissées grâce à une transformation logarithmique.

On observe une corrélation positive entre le taux d'alphabétisation et le PIB par habitant. En d'autres termes, les pays ayant des taux d'alphabétisation plus élevés ont tendance à afficher un PIB par habitant plus élevé.



3) Sélectionner les informations pertinentes :

- Les indicateurs économiques comme le PIB par habitant semblent pertinents pour évaluer le potentiel d'un marché.
- Les indicateurs démographiques comme la population et son évolution sont aussi importants.
- Le taux d'utilisation d'internet peut renseigner sur le niveau de digitalisation d'un pays.

Expansion d'Academy : tendances et défis globaux

Les indicateurs pris en compte pour le calcul du score sont:

- *PIB par habitant*
- *Dépenses d'éducation en % du PIB*
- *Taux d'utilisateurs d'internet*
- *Croissance de la population*
- *Population totale*
- *Nombre d'étudiants au secondaire et supérieur*

La méthode qui a été utilisée pour calculer le score des pays et déterminer les 6 pays ayant obtenu le meilleur score:

1) Standardisation des données avec *MinMaxScaler* afin de les mettre à une même échelle:


```
0 et 1: scaler = MinMaxScaler()  
DataPivot_imp_scaled =  
pd.DataFrame(scaler.fit_transform(DataPivot_imp_F))
```

2) Calcul de la moyenne des indicateurs standardisés pour obtenir un score par pays:

```
DataPivot_imp_scaled['Score'] = DataPivot_imp_scaled.mean(axis=1)
```

3) Tri des pays par score décroissant pour identifier les 6 meilleurs:

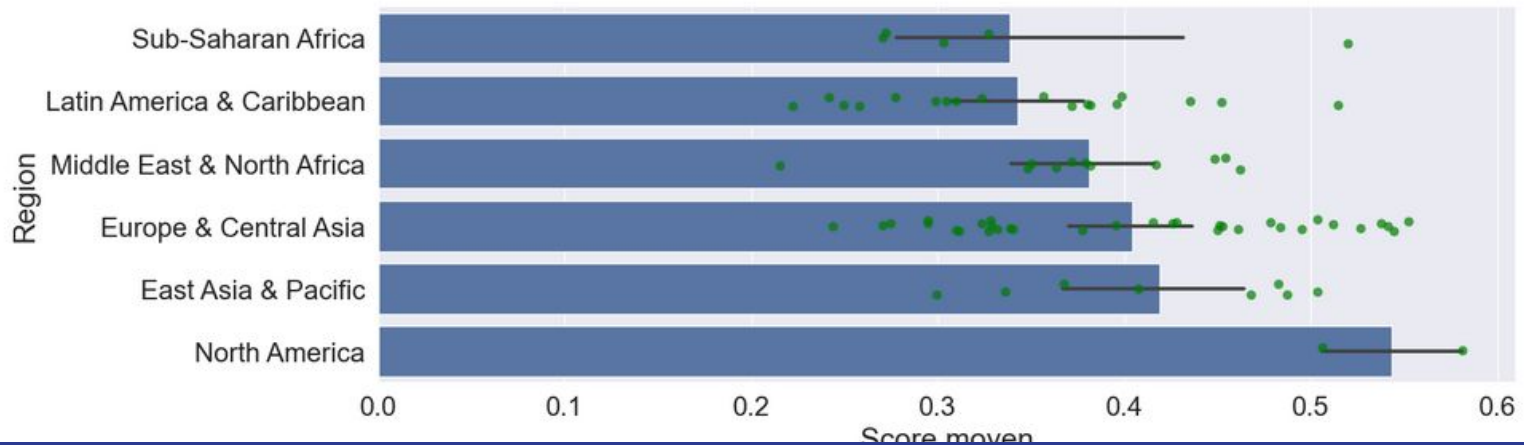
```
DataPivot_imp_F.sort_values('Score', ascending=False).head
```



4) Déterminer les ordres de grandeurs par zone géographique:

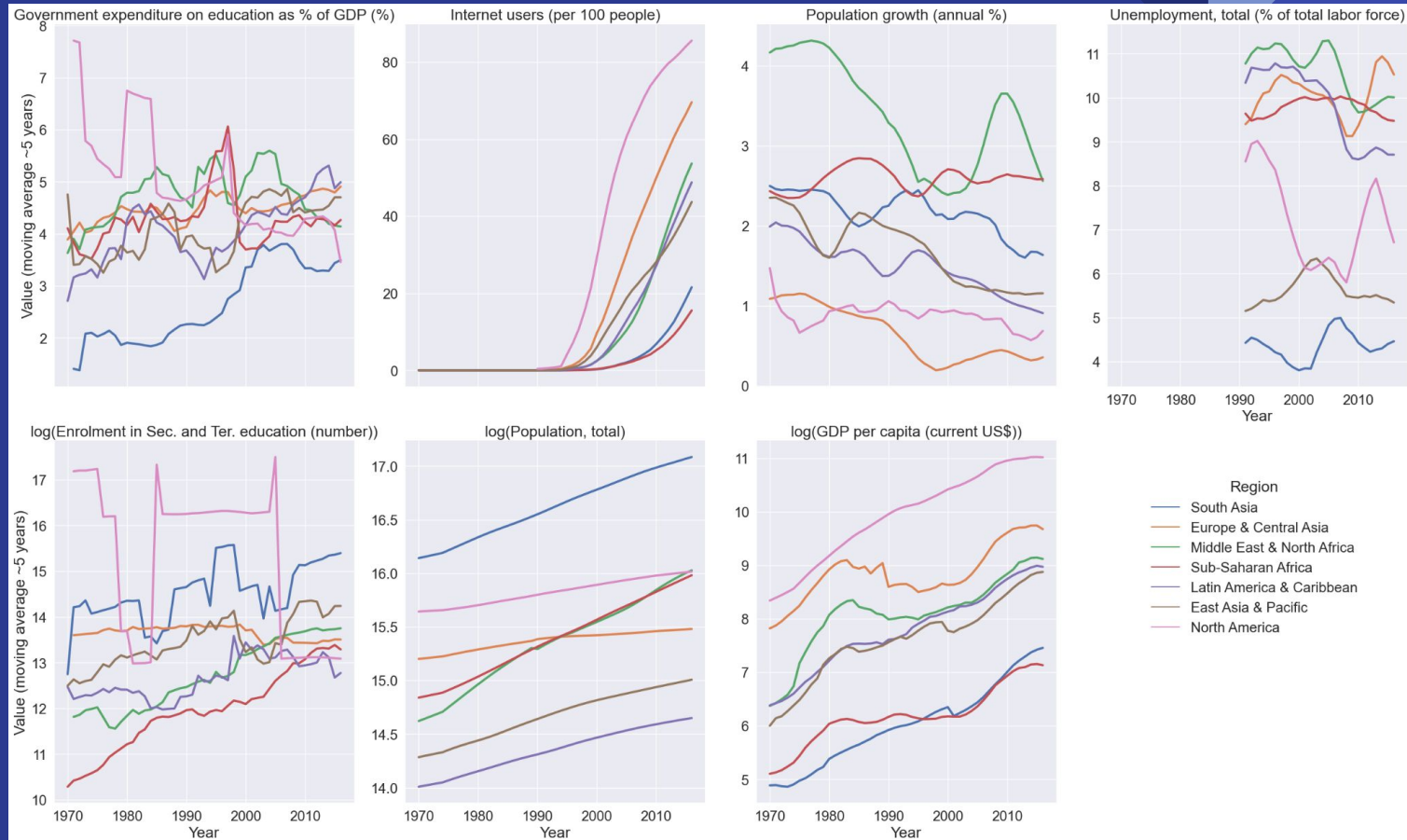
- Une analyse descriptive a été effectuée et donne pour chaque indicateur et chaque région : la moyenne, la médiane, l'écart-type, le minimum, le premier quartile, le troisième quartile et le maximum.
- Cela permet d'avoir un aperçu des ordres de grandeurs et de la dispersion pour chaque région

Analyse des indicateurs

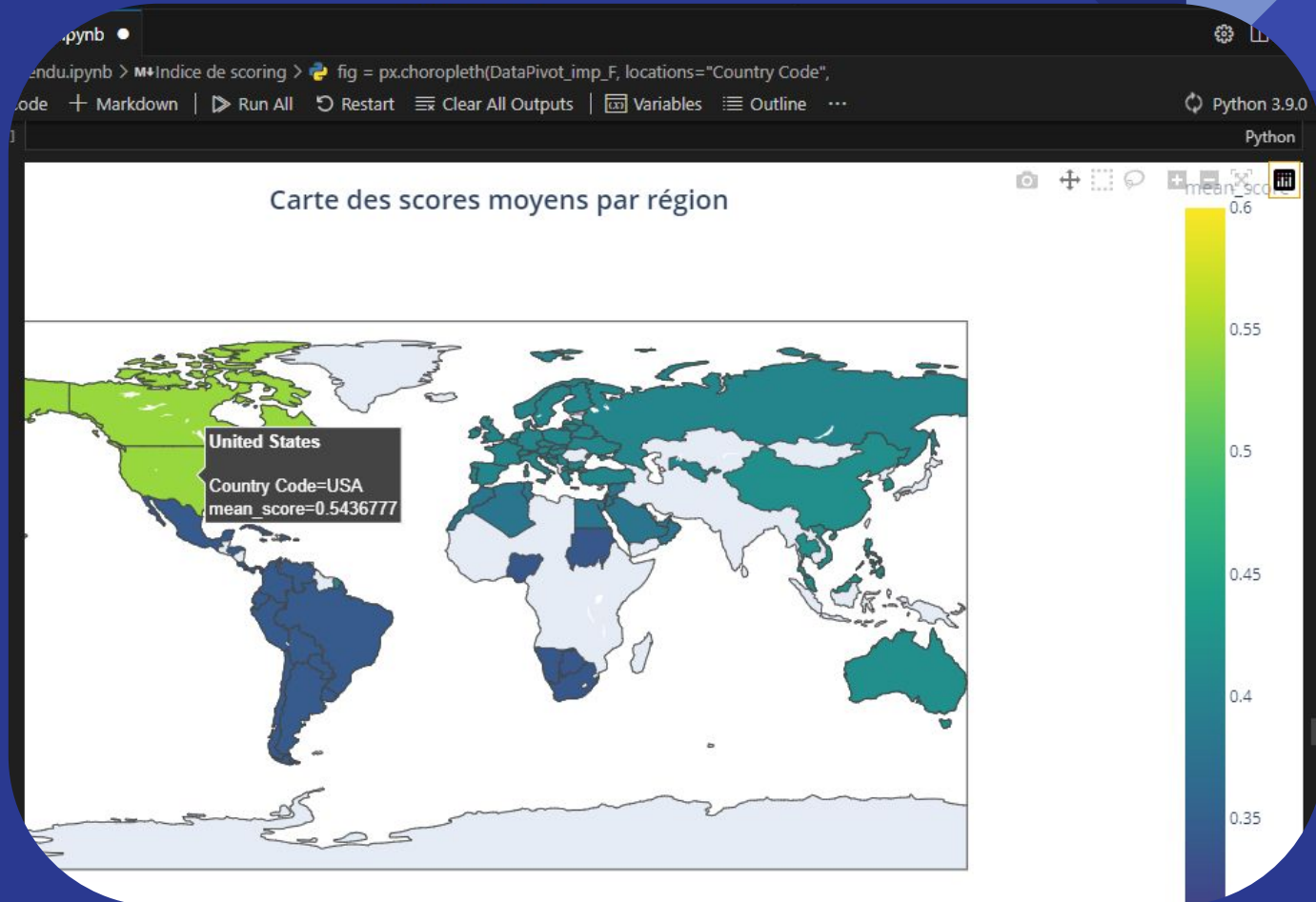


L'étude des indicateurs sélectionnés, après transformation de certaines distributions non-normales, fait ressortir des pays à plus fort potentiel, notamment en Europe et Amérique du Nord.

Expansion d'Academy : tendances et défis globaux



Expansion d'Academy : tendances et défis globaux



Les 6 pays à fort potentiel

.Les 6 pays ayant obtenu le meilleur score sont:

1. Etats-Unis 2. Danemark 3. Royaume-Uni 4. Suède 5. France 6. Allemagne

- Le Danemark et la Suède se distinguent par de forts investissements dans l'éducation
- Les Etats-Unis ont une forte population totale et nombre d'étudiants
- La France, le Royaume-Uni et l'Allemagne ont des indicateurs globalement similaires