



Academy

projet d'expansion internationale

ANALYSE PRÉ-EXPLORATOIRE

Cédric Juillet 2024

Objectifs de l'étude

Permettre de déterminer si le jeu de données à notre disposition peut informer les décisions en vue de notre projet d'expansion vers de nouveaux pays.

Le portail EdStats (portail des statistiques de l'éducation) fourni des indicateurs clés de l'éducation couvrant des sujets tels que l'accès, l'achèvement, les résultats d'apprentissage, les dépenses, les politiques.

EdStats est actuellement en cours de refonte pour améliorer l'expérience utilisateur et la pertinence, de sorte que certains indicateurs peuvent ne pas être à jour.

SOURCE DES DONNÉES : THE WORLD BANK

Fichier EdStatsCountry.csv

Répertorie de tous les pays de l'étude et les informations clés sur ceux-ci

Structure du DataFrame : 31 colonnes pour 241 lignes

Variable	Type	% Valeurs nulles	Clé	Fichiers liés
Country Code	object	0	oui	tous
Short Name	object	0		
...	...			
...	...			

Qualité :
Doublons : aucun

Fichier EdStatsCountry-Series.csv

Relation entre le fichier EdStatsCountry et EdStatsFootNote, cette relation est redondante car elle est déjà présente dans EdStatsFootNote

Ce fichier ne comporte que 6 SeriesCode différentes pour la France contre 211 pour l'autre fichier, nous ne tiendrons donc pas compte de ce fichier.

Structure du DataFrame : 3 colonnes pour 613 lignes

Variable	Type	% Valeurs nulles	Clé	Fichiers liés
CountryCode	object	0	oui	tous
SeriesCode	object	0		
DESCRIPTION	object	0		

Qualité :

Doublons : aucun

Fichier EdStatsFootNote.csv

Véritable clef de voûte des données de l'étude, c'est ici que nous retrouvons toutes les relations entre les différents fichiers.

Structure du DataFrame : 4 colonnes pour 643638 lignes

Variable	Type	% Valeurs nulles	Clé	Fichiers liés
CountryCode	object	0	oui	tous
SeriesCode	object	0	oui	EdStatsData (Indicator Code)
Year	object	0	oui	EdStatsData {année}=YR {année}
DESCRIPTION	object	0		

Qualité :

Doublons : aucun

Fichier `EdStatsData.csv`

Emplacement de la quasi totalité des données, une ligne par indicateur et par pays

Structure du DataFrame : 69 colonnes pour 886930 lignes

Variable	Type	% Valeurs nulles	Clé	Fichiers liés
Country Code	object	0	oui	tous
Country Name	object	0		
Indicator Code	object	0	oui	EdStatsFootNote & EdStatsCountrySeries
Indicator Name	object	0		
{année}	float64	Entre 73% et 93%	oui	EdStatsFootNote[Year]=YR{année}

Qualité :

Doublons : aucun

Valeurs Nulles : nombreuses mais normal au vu de la conception de la base de données
(matrice entre grand nombre d'indicateurs et années)

Méthodologie de l'analyse

Pour simplifier l'approche les régions ont été intégrées dans le jeu de données des pays

Création d'un premier dataframe pour l'étude de la situation actuelle :

- Sélection réalisée sur les années (de 2000 à 2020)
- Filtrage sur la fiabilité (indicateurs avec plus de cinq années de valeurs)
- Elimination par mots clés (Enfance, Primaire, Tests de lecture, répartition par genres, ..)
- Choix des indicateurs

Création d'un deuxième dataframe pour l'étude prévisionnelle :

- Sélection réalisée sur les années (de 2020 à 2100)
- Choix des indicateurs

L'étude s'appuiera sur la liste des 35 pays ayant les PIB les plus importants et définira une 'short' liste en fonction de l'examen des différents graphiques.

Choix des données actuelles

Après un balayage des 568 indicateurs restants il semble pertinent d'étudier les chiffres suivants :

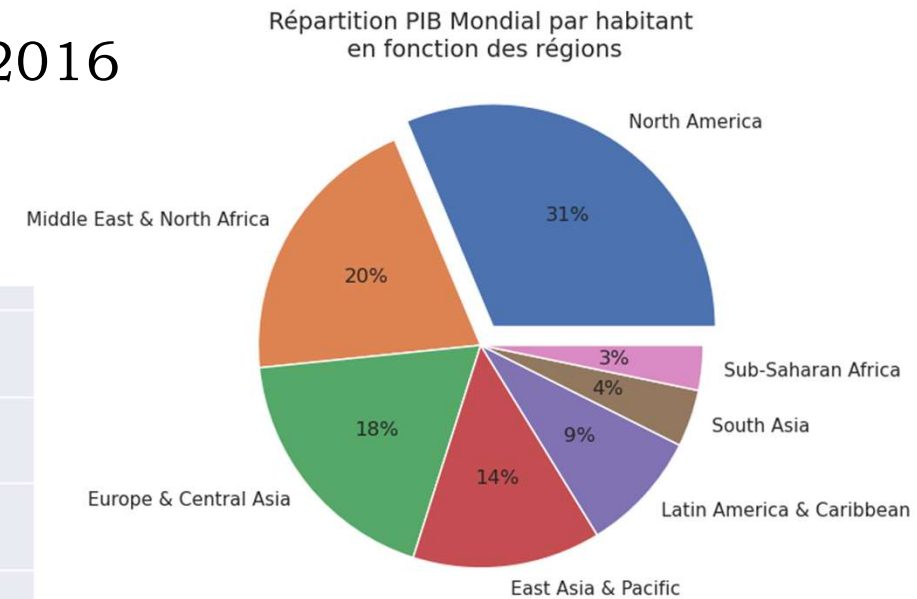
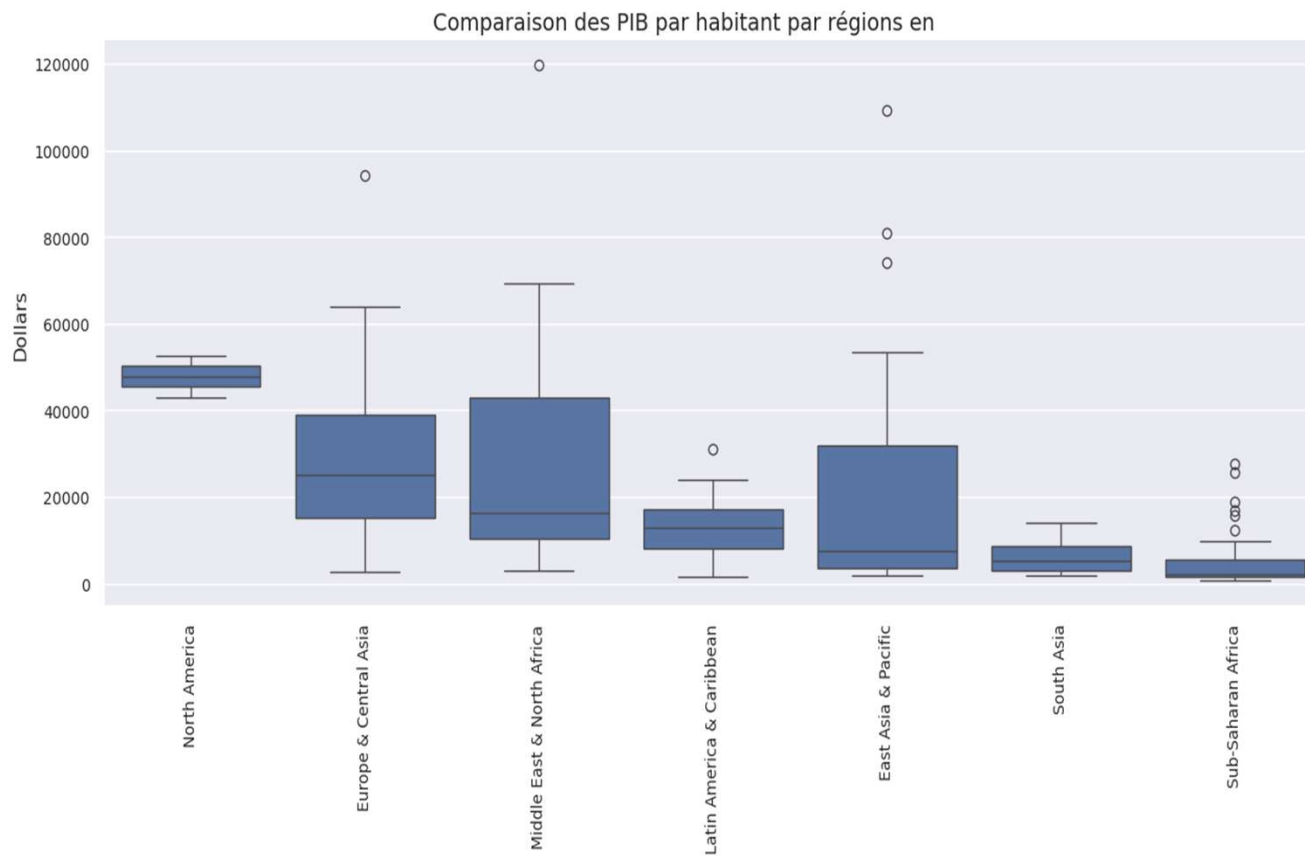
1. PIB par habitant (en \$ constants de 2011)
2. Ordinateurs personnels (pour 100 personnes)
3. Utilisateurs Internet (pour 100 personnes)
4. Population âgée de 15 à 24 ans (Total)
5. Croissance démographique (% annuel)
6. Dépenses du secondaire (en % des dépenses publiques d'éducation)
7. Dépenses d'enseignement supérieur (en % des dépenses publiques d'éducation)

Après une vérification graphique du nombre d'occurrences par indicateur, il semblerait que les chiffres portants sur les ordinateurs personnels ne sont pas utilisables.

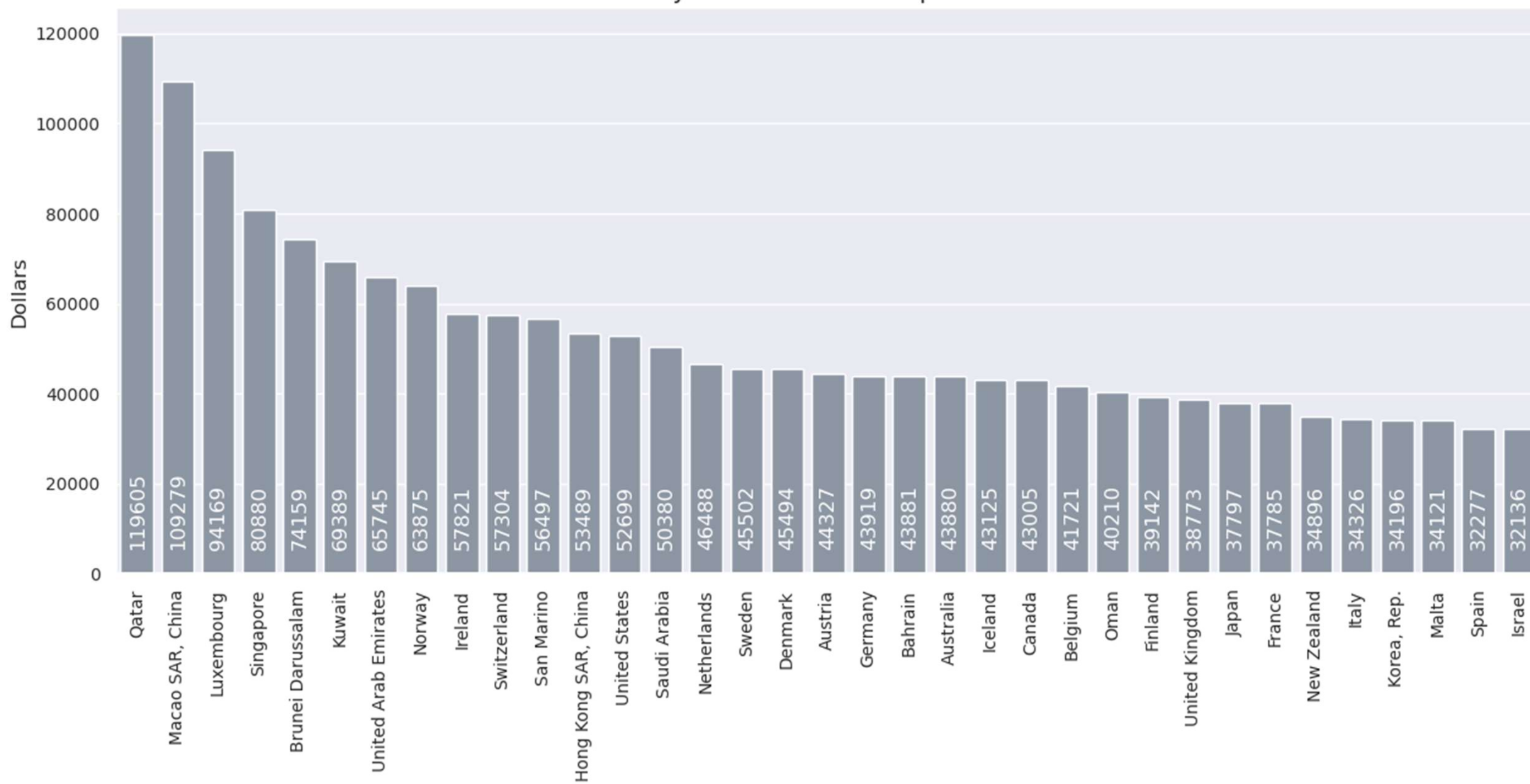
Produit intérieur brut de 2016

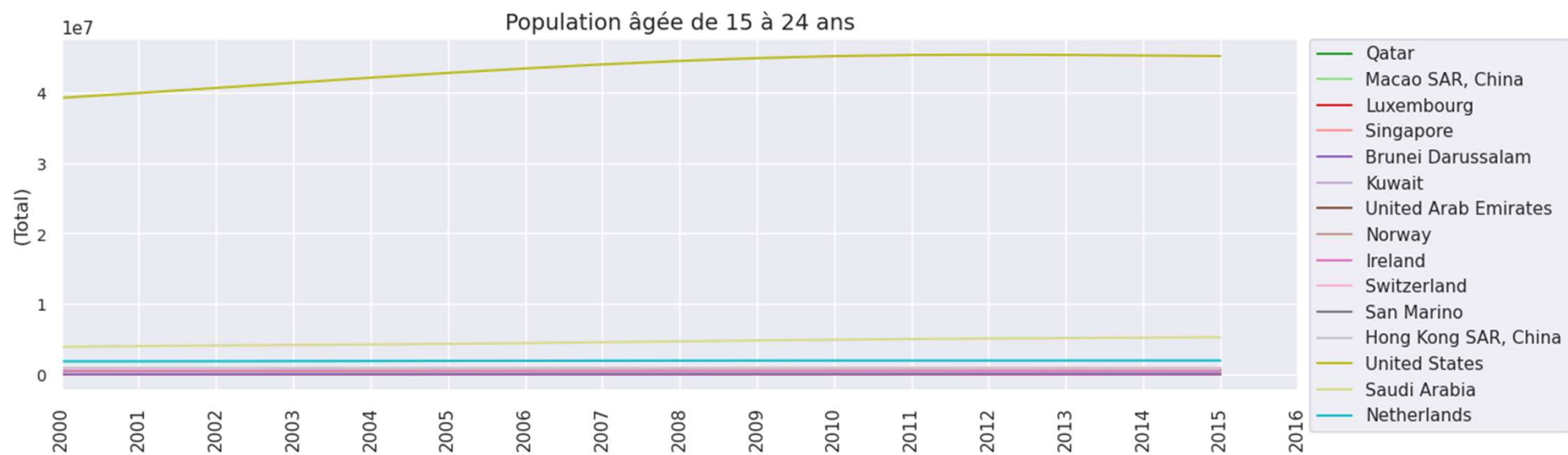
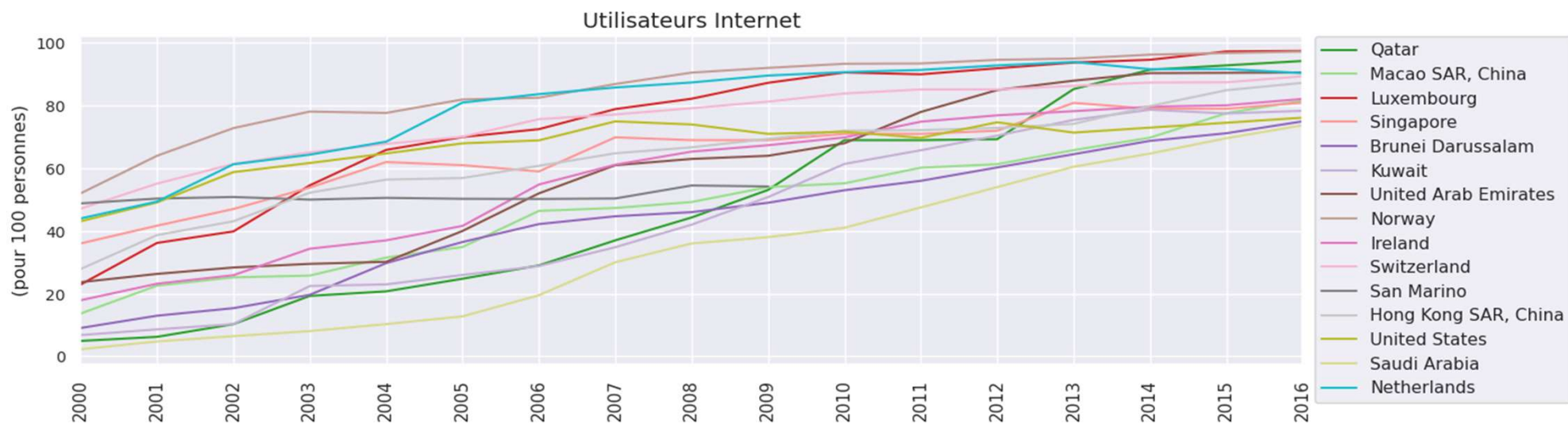
Région	Nb Pays	Moyenne	Ecart type	Min	25%	Médiane	75%	Max
Amérique du Nord	2	48214	7250	43087	45651	48214	50778	53341
Moyen Orient & Afrique du Nord	15	31594	32140	2325	10535	16086	43100	118207
Europe & Asie Centrale	50	28767	18504	2762	15667	25852	39449	94764
Asie de l'est & Pacifique	30	20683	25650	1955	3767	7841	32656	96565
Amérique Latine & Caraïbes	31	13463	6633	1653	8084	13124	17660	30464
Asie du Sud	8	6569	4564	1739	3063	5473	9044	14581
Afrique Sub-Saharienne	45	4933	6214	647	1491	2379	5301	26319

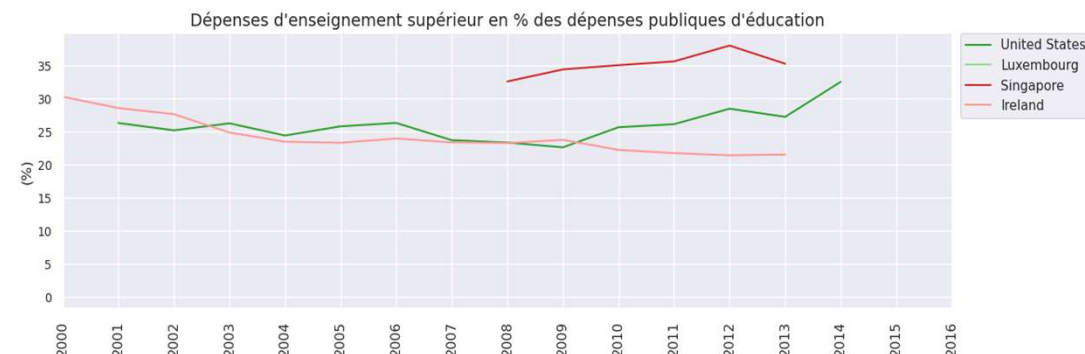
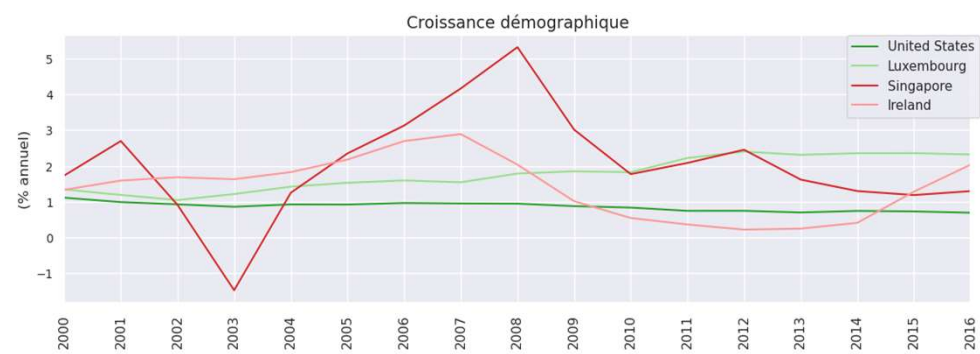
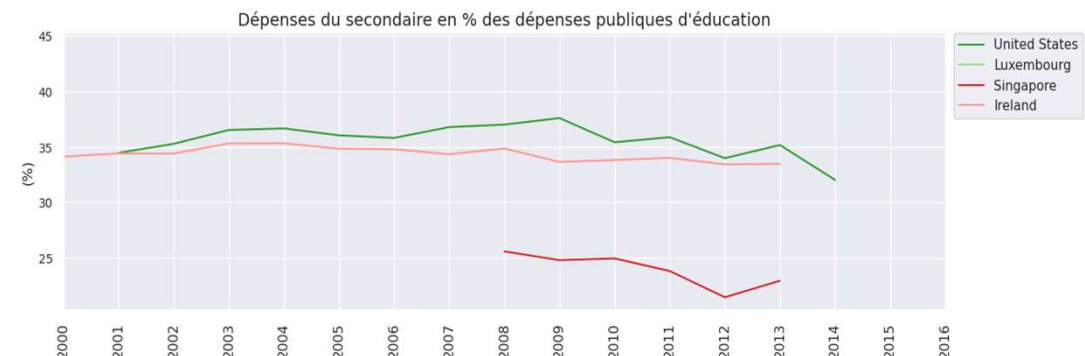
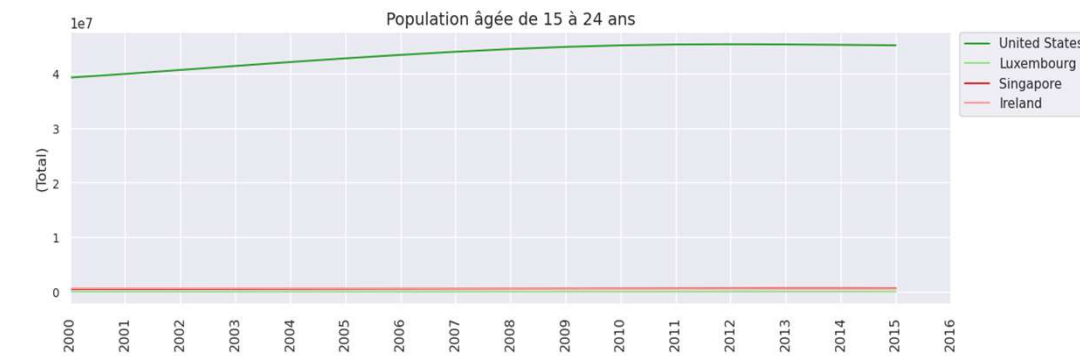
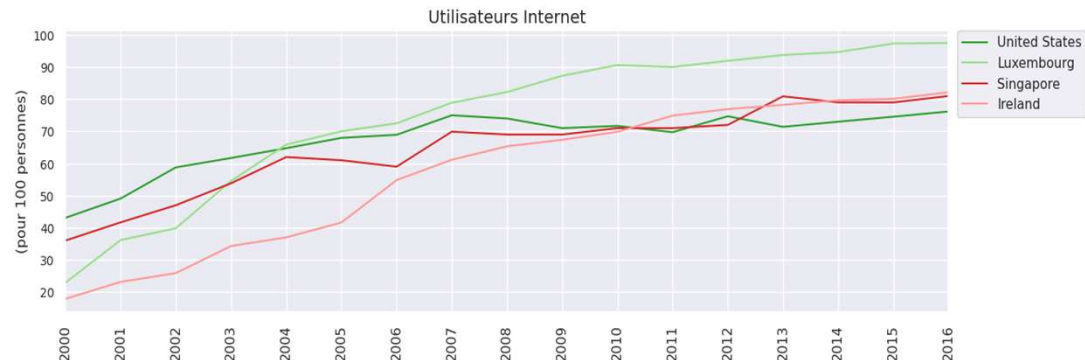
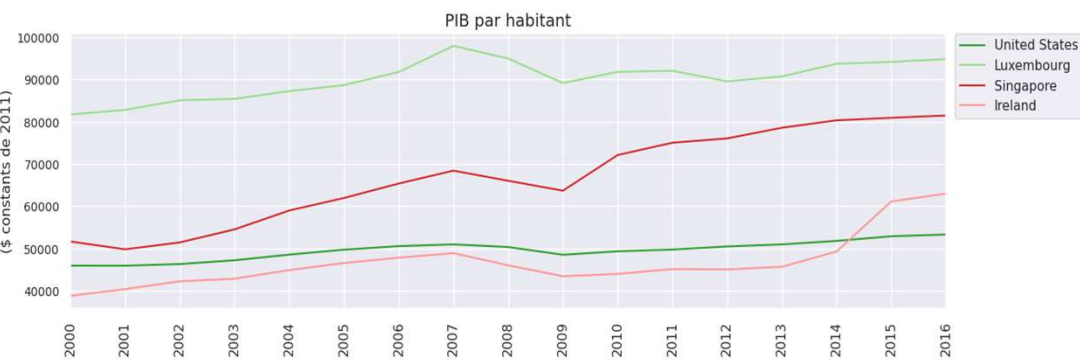
Produit intérieur brut moyen entre 2014 et 2016



PIB moyen de 2014 à 2015 par habitant







Choix des données prévisionnelles

Après un balayage des indicateurs sur les années 2020 – 2100 il semble intéressant d'étudier les données suivantes :

1. Projection des années moyennes de scolarité des 15-19 ans (Années moyennes)
2. Projection des années moyennes de scolarité. 20-24 ans (Années moyennes)
3. Projection de la population âgée de 20 à 24 ans, ayant le niveau Secondaire (en milliers)
4. Projection de la population âgée de 20 à 24 ans, ayant le niveau Postsecondaire (en milliers)
5. Projection du pourcentage de la population âgée de 25 ans et plus ayant le niveau Secondaire (% Population)
6. Projection du pourcentage de la population âgée de 25 ans et plus ayant le niveau Postsecondaire (% Population)
7. Projection du pourcentage de la population totale ayant le niveau Secondaire (% Population)
8. Projection du pourcentage de la population totale ayant le niveau Postsecondaire (% Population)

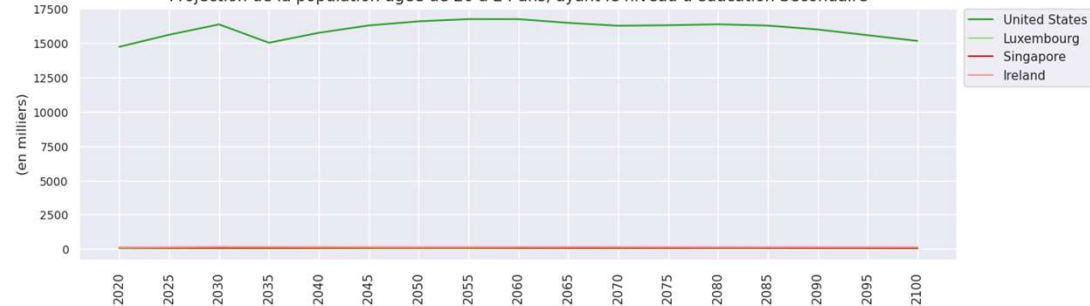
Projection des années moyennes de scolarité. 15-19 ans



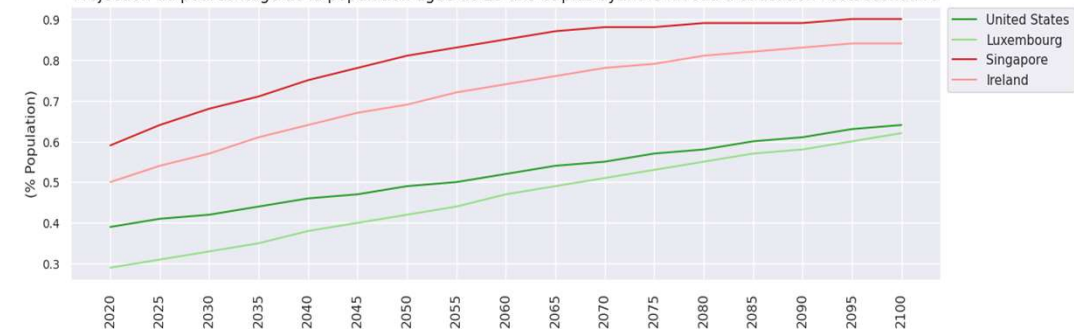
Projection des années moyennes de scolarité. 20-24 ans



Projection de la population âgée de 20 à 24 ans, ayant le niveau d'éducation Secondaire



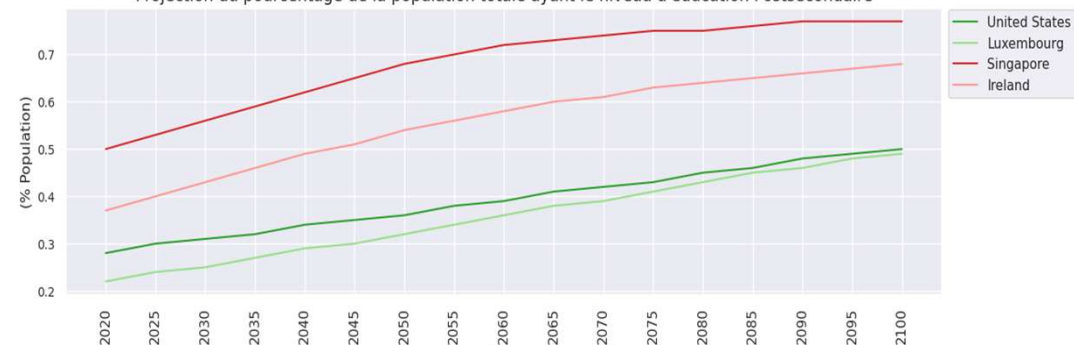
Projection du pourcentage de la population âgée de 25 ans et plus ayant le niveau d'éducation Postsecondaire



Projection de la population âgée de 20 à 24 ans, ayant le niveau d'éducation Postsecondaire



Projection du pourcentage de la population totale ayant le niveau d'éducation Postsecondaire



Résultats d'analyse

Jeu données fiable
Très intéressant pour notre projet

Évolution
Secondaire diminue
Supérieur augmente

Les Etats-Unis sont un bon candidat
Entre 6 et 10 fois plus de volumétrie pour notre implantation

Merci pour votre attention