

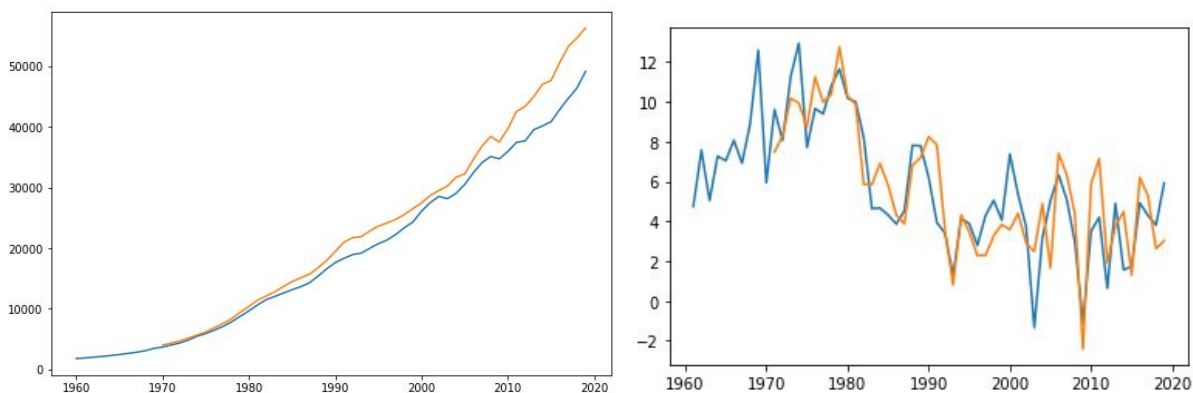
Projet DATA 731

Introduction:

Ce projet s'intéresse aux données sur le PIB par habitant de l'OCDE et veut prédire les valeurs qu'il pourrait prendre. Pour le prédire, nous allons implémenter différentes méthodes qui seront explicitées dans ce rapport. Nous allons par exemple comparer le PIB à d'autres facteurs socio-économiques, le but étant d'identifier quels sont les points à améliorer dans une société pour avoir un meilleur développement économique.

Nos données:

Pour les données liées au PIB, nous avons préféré prendre les données du PIB par habitant brut plutôt que la croissance. La raison est que l'on peut retrouver la croissance par une simple division comme nous pouvons le voir ci-dessous. Les courbes de gauche représentent le PIB par habitant de la France (en bleu) et de l'Allemagne (en orange) avec le PIB en dollars en ordonnée et l'année en abscisse. Les courbes de droite représentent la croissance en pourcentage de ces deux mêmes pays sur la même période. Comme nous pouvons le voir sur l'abscisse, nous avons pris les données annuelles de PIB de 1960 à 2019.



Ces données (comme toutes celles du TP) ont été téléchargées au format csv et manipulées sur python grâce au module pandas.

```
Pib = pd.read_csv("PIB.csv")
```

Voici par exemple notre fonction qui permet de récupérer le PIB pour chaque année d'un pays "country" dans un dataframe pandas "df".

```
def data_country(country, df):  
    data1 = []  
    data2 = []  
    for k in range(len(df)):  
        if str(df["LOCATION"][k]) == country:  
            data1.append(float(df["TIME"][k]))  
            data2.append(float(df["Value"][k]))  
    return data1, data2
```

Corrélation du PIB de différents pays:

Nous avons voulu nous intéresser à la corrélation entre la croissance du PIB de certains pays de différentes parties du monde. Ces pays sont ajoutés à une liste que le programme va utiliser:

```
list_countries = ["FRA", "DEU", "USA", "CAN", "SGP", "JPN"]
```

Pour connaître la corrélation entre la croissance de ces différents pays, nous avons calculé la corrélation de chaque couple de pays un à un puis nous avons consigné ces valeurs dans une matrice. Certains pays avaient des données sur plus d'années que d'autres, nous avons uniquement comparé les années où les deux pays ont des données. Pour ce faire, nous avons tout d'abord récupéré les données de tous les pays listés grâce à la fonction vue plus haut. Nous avons ensuite comparé tous les couples de pays de cet ensemble de données. Nous obtenons alors une matrice de corrélation que voici:

	FRA	DEU	USA	CAN	SGP	JPN
FRA	1.0	0.51	0.49	0.33	0.03	0.5
DEU	0.51	1.0	0.74	0.76	0.14	0.79
USA	0.49	0.74	1.0	0.89	-0.06	0.82
CAN	0.33	0.76	0.89	1.0	0.05	0.72
SGP	0.03	0.14	-0.06	0.05	1.0	-0.02
JPN	0.5	0.79	0.82	0.72	-0.02	1.0

Corrélation du PIB avec d'autres indices:

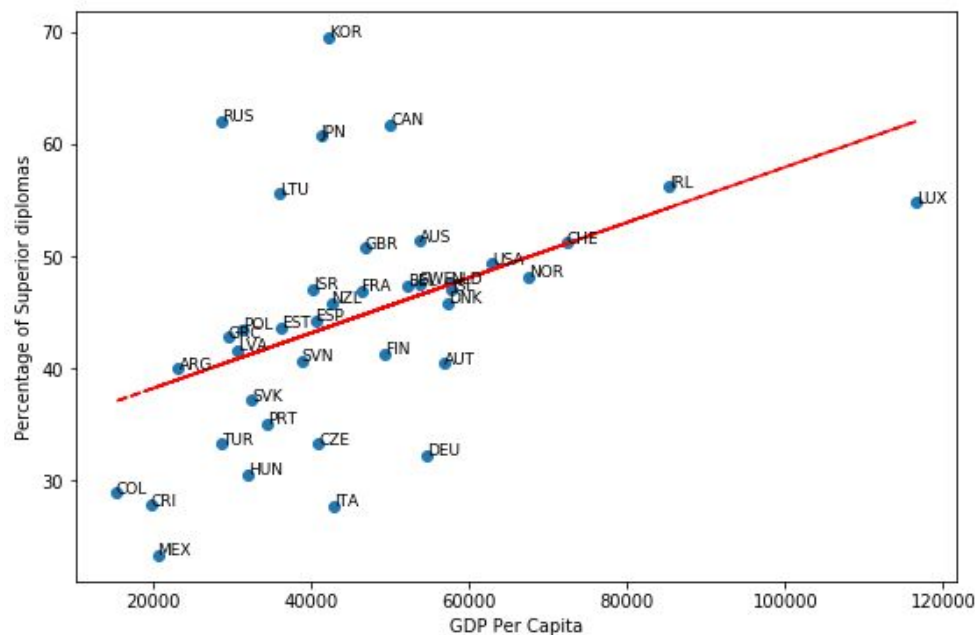
Pour tenter de prédire le PIB d'un pays, nous avons voulu comparer le PIB, que ce soit sa valeur absolue ou sa croissance, avec d'autres indices. Le but étant de trouver un indice facile à prédire et fortement corrélé au PIB.

Le premier indice que nous avons voulu comparer est le taux de personnes ayant un diplôme supérieur et le PIB par habitant car il semble logique que plus une population est éduquée plus elle est riche. Pour voir la corrélation, nous avons voulu afficher un graphique avec le PIB en abscisse et le taux d'éducation en ordonnée. Nous avons aussi ajouté en rouge une droite de corrélation grâce à cette fonction:

```
z = np.polyfit(pib_list, dip_list, 1)
p = np.poly1d(z)
plt.plot(pib_list, p(pib_list), "r--")
```

pib_list est la liste des pib des pays et dip_list est le taux de personnes ayant fini leurs études supérieures dans ces pays.

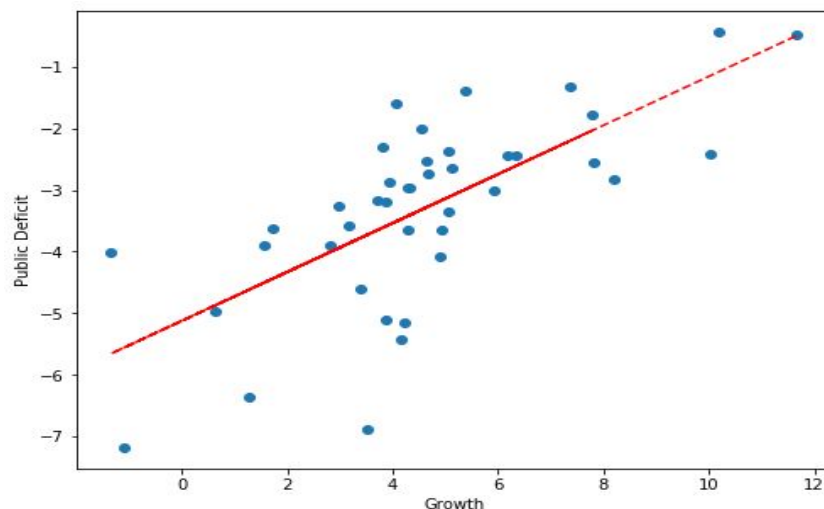
Le graphique en lui même ressemble à ceci:



On obtient un coefficient de corrélation relativement élevé, mais pas assez élevé pour les prédictions que l'on veut faire.

Coefficient de corrélation: 0.4493238861545058

Dans le même esprit, nous avons voulu faire la même analyse entre la croissance et le déficit public d'un pays. Ce graphique comprend un point pour chaque année et ses coordonnées sont la croissance et le déficit public de la France:



Cet autre indice est très intéressant car il est fortement corrélé au PIB:

Coefficient de corrélation: 0.6982129549938035

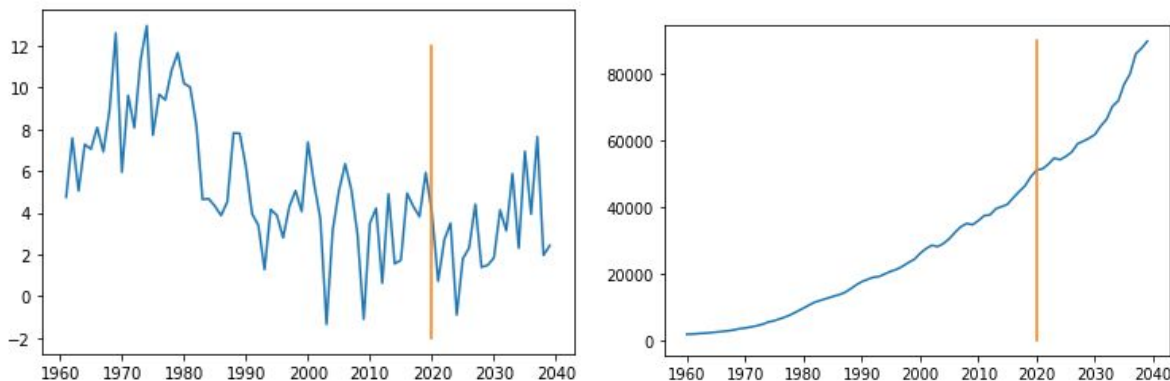
De plus, contrairement au PIB, il est facile à prédire. En regardant les politiques économiques d'un pays, on peut facilement prédire si les coffres du pays se rempliront ou se videront cette année. Ces deux indices ayant une forte corrélation, prédire l'un aide à prédire l'autre.

Prédiction du PIB par ses valeurs précédentes:

Pour tenter de prévoir le PIB, nous avons calculé la croissance moyenne et l'écart-type du PIB des 20 dernières années. Nous avons ensuite simulé la croissance en utilisant une fonction numpy qui simule une distribution normale avec les deux valeurs calculées précédemment:

```
dates, values = data_country(country, df)
grow = growth(values)
sigma = np.std(data[len(data)-20:])
mu = np.mean(data[len(data)-20:])
for k in range(20):
    new_growth = np.random.normal(mu, sigma)
    grow.append(new_growth)
    values.append(values[-1] * (1+ (new_growth/100)))
    dates.append(dates[-1] + 1)
```

On obtient alors ces courbes pour la croissance et le PIB total, la barre orange représente à partir où notre prédiction commence:



Conclusion:

Nous avons utilisé différentes manières pour tenter de prédire le PIB, mais ce qui est ressorti est que le PIB n'est pas quelque chose de simple à prédire. Surtout à cause du fait qu'à tout moment une crise économique peut faire tomber toutes nos prédictions à l'eau. Cependant, nous avons tenté de faire de notre mieux pour le prédire de différentes manières et en apprendre le plus possible.