Rendu Python

Notre équipe :

- KLOC Nicolas
- MARCHAND Yohan

ÉTAPE 1 : CHOIX DU DATASET

https://data.world/health/big-cities-health/workspace/file?filename=Big_Cities_Health_Data_Inventory.csv

Après avoir choisi notre fichier sur l'état de la santé aux USA dans les différentes villes, nous nous sommes rendus compte que chaque catégorie avait un axe de lecture différent de l'autre. (Parfois un nombre sur 100.000, parfois un pourcentage, parfois un nombre de personnes atteintes, parfois une mortalité...)

Cette contrainte nous impose un travail par type de catégorie, et il est difficile de pouvoir comparer les différentes catégories entre elles.

ÉTAPE 2 : PRÉPROCESSING

Nous nous sommes rendus compte qu'il y'avait pas mal de colonnes inutiles pour notre axe de recherche, car nous ne sommes pas capables de juger la qualité des différentes enquêtes permettant de créer la data.

Nous avons donc supprimé les ['Notes', 'Methods', 'Source', 'BCHC Requested Methodology'], ainsi que les lignes dont certaines valeurs étaient nulles, avec la **librarie panda.**

ÉTAPE 3 : ANALYSE

Nous avons décidé d'étudier en détail le VIH et la nutrition, activité physique et l'obésité aux USA, en comparants les données parmi les différentes villes, années, sexe et encore "races".

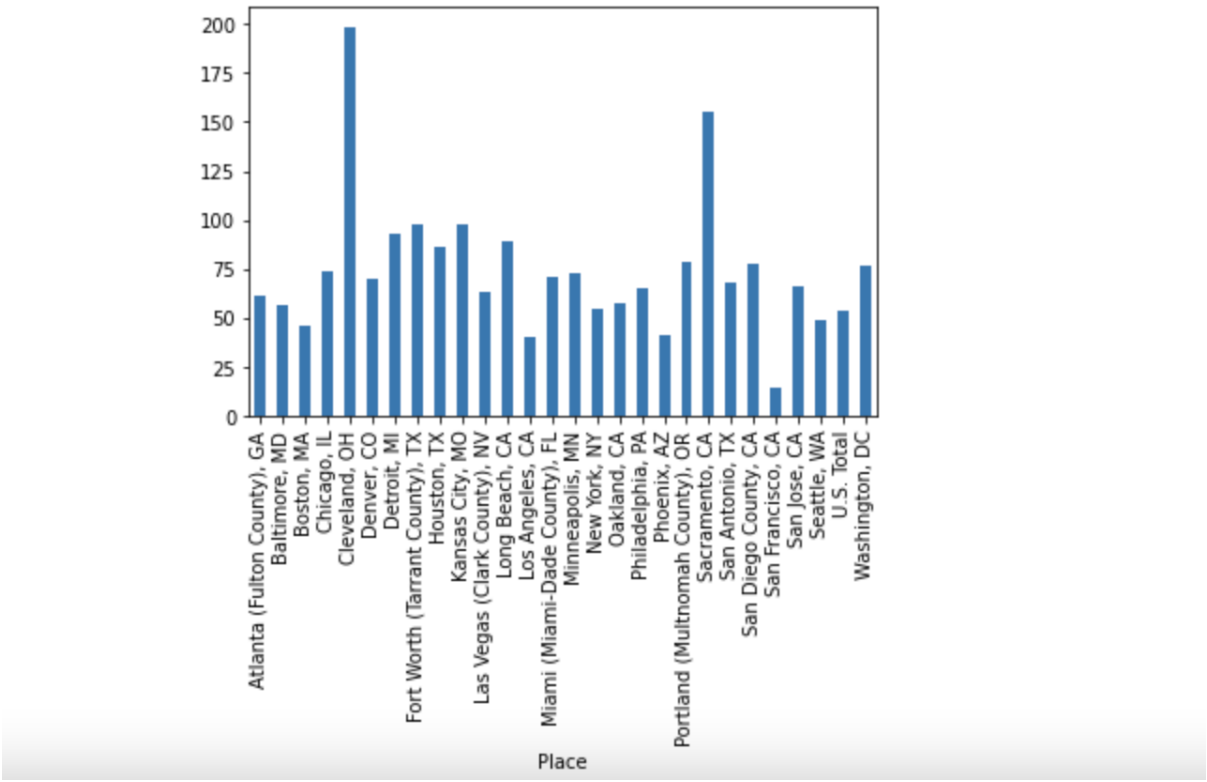
Nous avons donc utilisé les librairies Matplotlib, ainsi que **panda** pour **jouer sur les dataframes, les adapter à nos recherches** et enfin les **visualiser avec**

matplotlib.

Analyse du diabète en Amérique.

Au cours de notre enquête, parsemée de commentaires, nous nous sommes aperçus que Cleveland était la ville avec le plus gros taux de mortalité due au diabète en Amérique (4x plus que la moyenne soit approximativement 199 adultes mort de diabète tous les 100.000 habitants.).

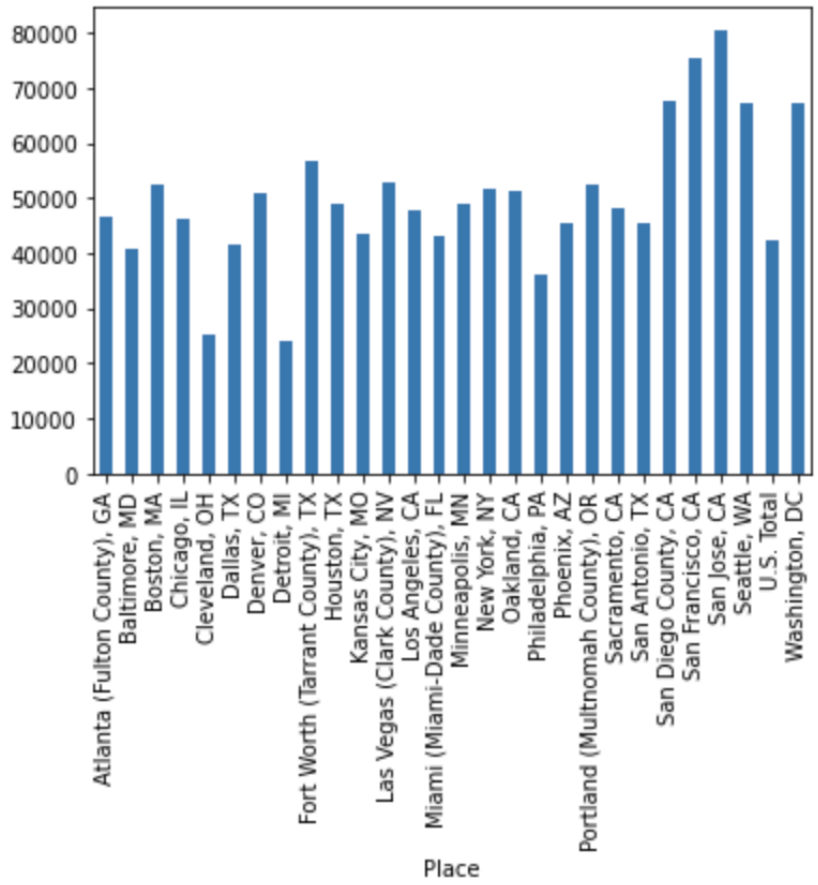
```
Out[5]: <AxesSubplot:xlabel='Place'>
```



Taux de mortalité du au diabète par ville pour 100.000 habitants

Nous avons donc cherché à comprendre cette disparité en étudiant les **critères démographiques** (pauvreté...) qui peuvent justifier une population obèse suite à la malnutrition(trop de fast-food, ou de nourriture riche en graisse/sucre)

```
Out[12]: <AxesSubplot:xlabel='Place'>
```



Revenu médian des habitants par ville

Tout comme les **critères liés aux comportements** (Alcoolisme, drogue.) qui sont souvent liées à des syndromes de dépression (qui peut amener les gens à manger pour se réconforter) ou qui font tout simplement grossir comme une consommation excessive d'alcool.

Critères démographiques :

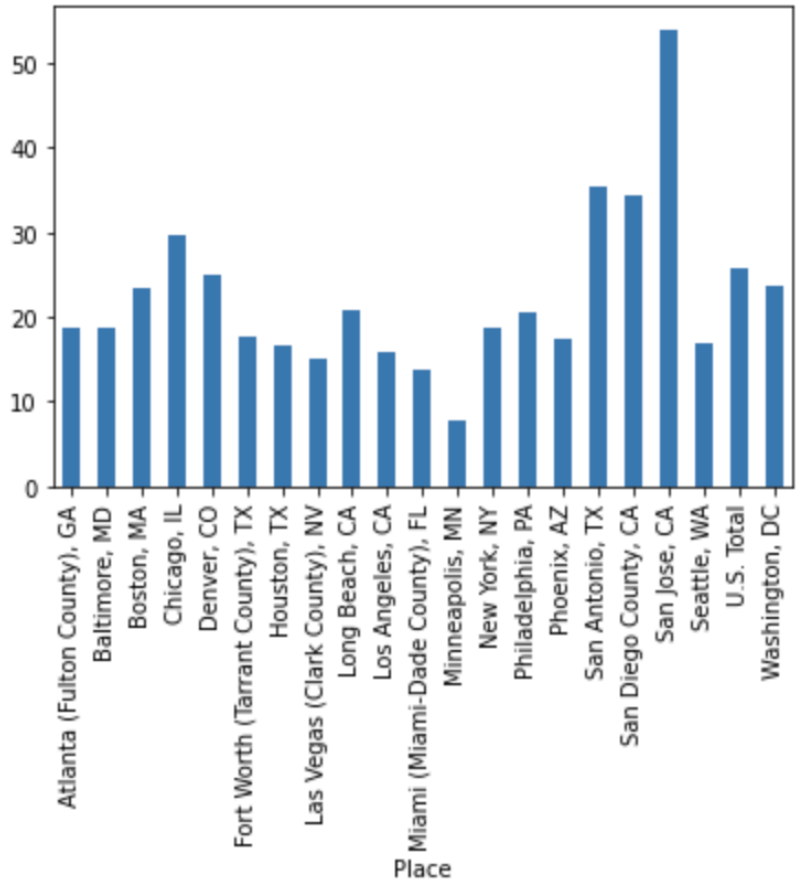
En étudiant le seuil de pauvreté et les revenus médians par ville, on se rend compte que Cleveland est une ville très pauvre, cependant Detroit est encore plus pauvre et n'a pas un taux de mortalité du au diabète aussi élevé que celui de Cleveland. Les villes en dessous du revenu moyen médian aux USA ont tout de même une mortalité due au diabète plus élevée que le reste.

On peut donc partir du principe que les deux critères sont corrélés, et qu'une population pauvre disposera de plus de personnes en état de malnutrition et donc possiblement à risque de diabète.

Critères liés aux comportements :

On aimerait étudier le pourcentage d'adultes en situation d'alcoolisme dans les différentes villes, malheureusement le dataset ne dispose pas des données pour Cleveland.

```
Out[22]: <AxesSubplot:xlabel='Place'>
```



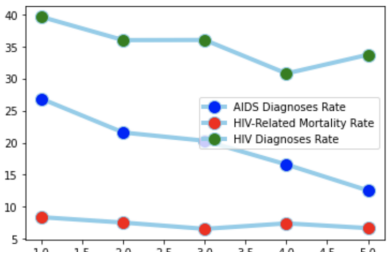
Taux d'adultes alcooliques par ville en % de population

Analyse du SIDA/VIH en Amérique.

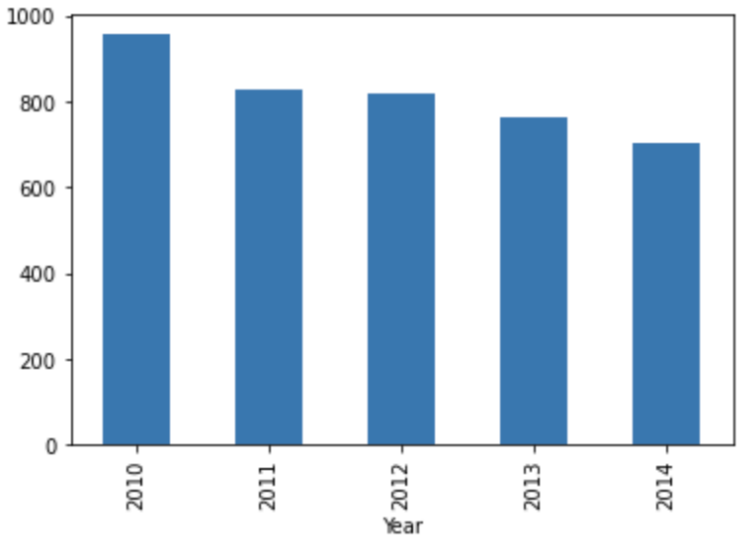
Nous nous sommes également intéressé à l'évolution du SIDA / VIH aux états-unis.

Entrée [7]:
```
# libraries
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

# Data
df=pd.DataFrame({'x_values': range(1,6), 'y1_values': file4, 'y2_values': file2, 'y3_values': file3 })

# multiple line plots
plt.plot( 'x_values', 'y1_values', data=df, marker='o', markerfacecolor='blue', markersize=12, color='skyblue', linewi
plt.plot( 'x_values', 'y2_values', data=df, marker='o', markerfacecolor='red', markersize=12, color='skyblue', linewic
plt.plot( 'x_values', 'y3_values', data=df, marker='o', markerfacecolor='green', markersize=12, color='skyblue', linew

# show legend
plt.legend()

# show graph
```

Out[7]: <matplotlib.legend.Legend at 0x7f8a80ea8880>



Out[2]: <AxesSubplot:xlabel='Year'>



Ce dernier graphique représente le nombre de personnes vivant avec le SIDA/VIH pour 100.000 habitants. On remarque que ce chiffre est en baisse.

Quand au premier graphique, il symbolise le nombre de personnes attrapants le VIH, SIDA ou encore décédant à cause du HIV, pour 100.000 habitants.

On remarque que le taux de mortalité est stable, alors que le taux de personne diagnostiqué avec le SIDA est en baisse.

Pourtant, le taux de personne diagnostiqué avec le VIH n'a pas changé, ce qui veut dire qu'il y'a moins de personnes ayant le VIH qui atteignent la phase 3 du virus, le SIDA.

On peut donc supposer que les médicaments sont donc plus efficaces de nos jours.