

**Rapport du projet analyse de données**

**ACP (Analyse des composantes principales)**

**Réalisé par :**

**Khaled Zagoub**

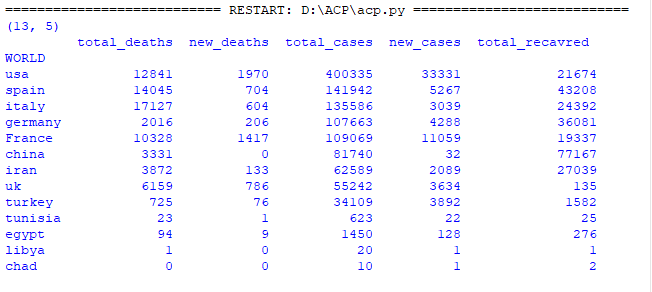
**Bilel Khemissi**

**Feres Sakouhi**

### Lecture des données:‎

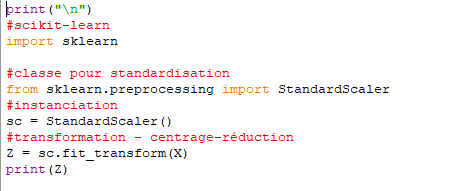
### 

**#affichage :**

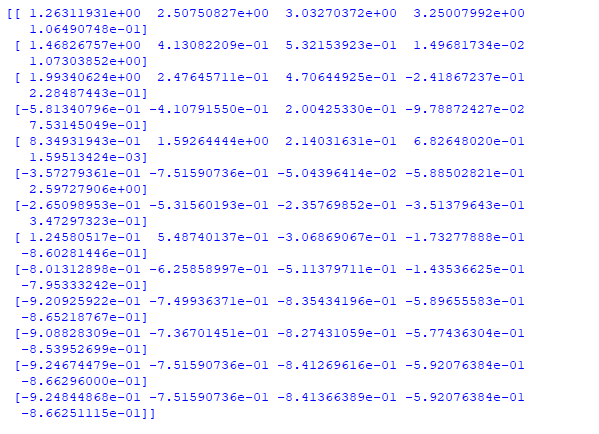
****

**#transformation de la matrice en une matrice centrée –réduite :**

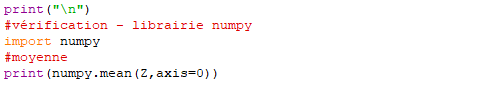
C’est pour donner une importance identique à chaque variables .

****

### #affichage :



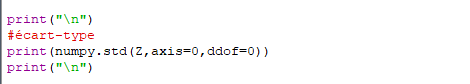
**#calcul de la moyenne des variables :**



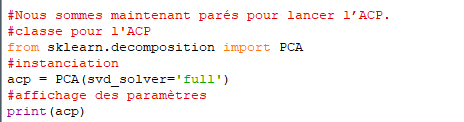
**#affichage :**



**#calcul du l’écart-type :**



**#affichage :**



**Affichage des paramètres :**

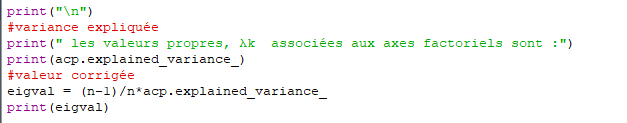


### 



I : le nombre des composants générés doit être égal au nombres des variables.

**#calcul des valeurs propres :**

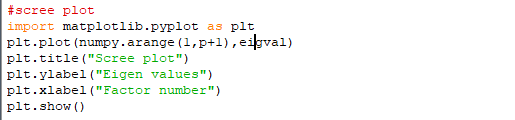


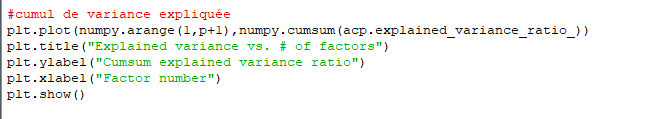
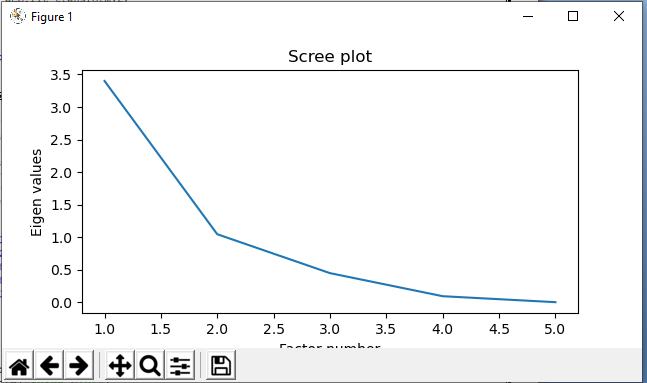


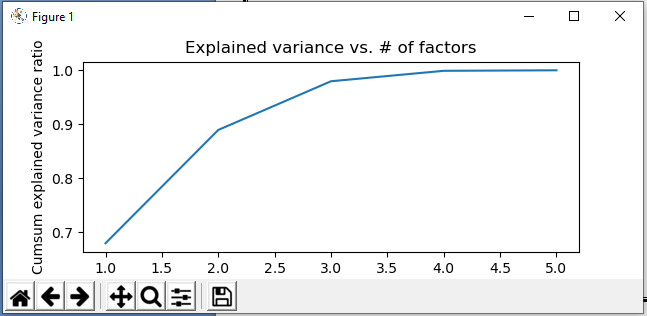
**Interpretation** :

la somme des valeurs propres doit être égale aux nombres des variables

Verification : 3.40008793+ 1.04779488 +0.45061911+ 0.09659226+ 0.00490582=1



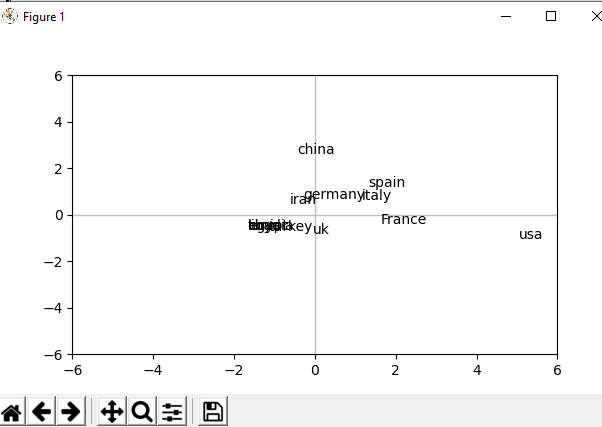
****

****

### #Représentation des individus :

### 

### 

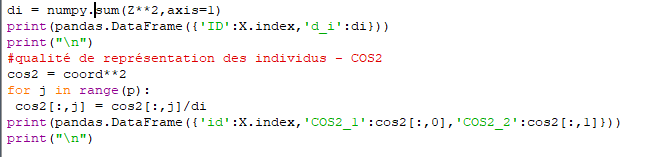


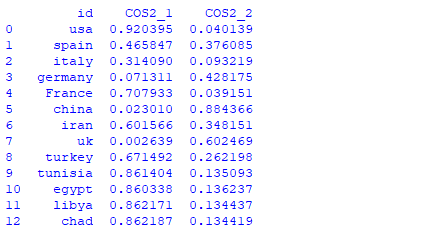
**Interprétation :**

1. Le taux d’infection du coronavirus est très élevé dans **USA** par rapport aux autres pays.
2. **La chine** a un taux de recouvrement le plus élevé malgré qu’elle a un taux d’infection un peu élevé ce qui justifie leur emplacement dans le plan.
3. **Spain** a un taux d’infection du coronavirus élevé et aussi un taux assez élevé de personnes rétablies.
4. Le taux d’infection de la T**unisie et Chad** ainsi que **Turkey** est faible par rapport aux autres pays et le taux de recouvrement est faible aussi .

**#Qualité de représentation – Les COS² (cosinus carré)**

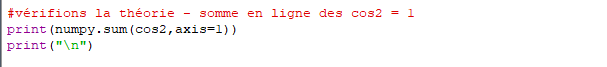
**#contribution des individus dans l'inertie totale :**

****



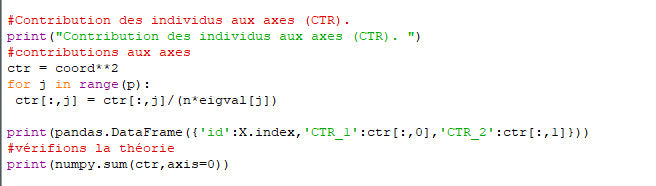
Interprétation :

1. On a ‘**Germany’** et **’Italy’** sont à peu près proche du centre du repère, ses qualités de représentation sont très faibles ( **germany** : 0.49 / **italy** : 0.4 ) c'est-à-dire qu’ils sont subit une déformation lors de la projection .
2. Les pays les mieux représentés sont : **Tunisia** , **Egypte** ,**Tchad** et **Libya** .





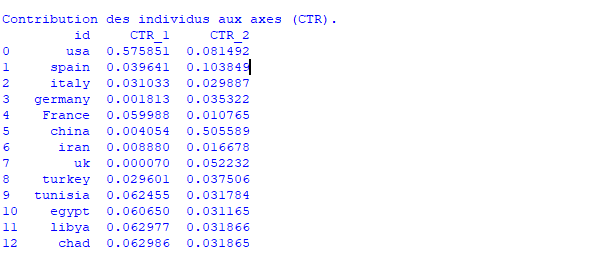
**#Contribution des individus aux axes (CTR) :**



**Interprétation :**

**1-**Le pays qui a la valeur la plus élevée de contribution à la formation du 1ére axe est USA : Avec une valeur de 57.5%.

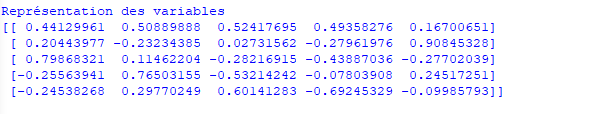
2- Le pays qui la plus élevée de contribution la formation du 2éme axe est CHINA : Avec une valeur de 50.5%



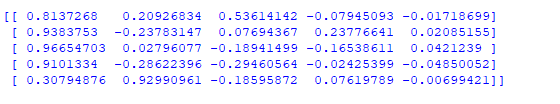




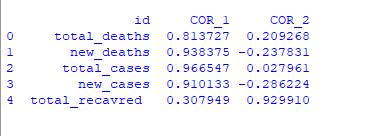
**#Représentation des variables :**

****

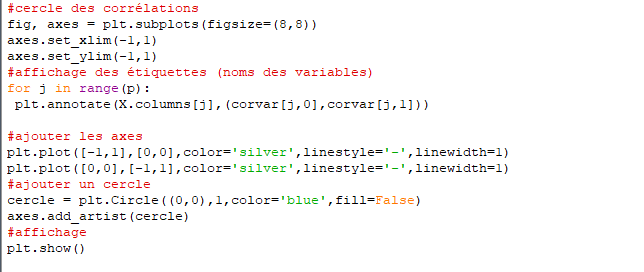


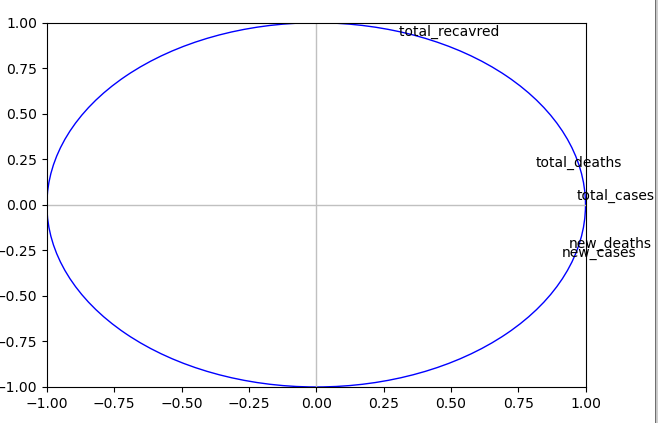






**#cercle des corrélations :**

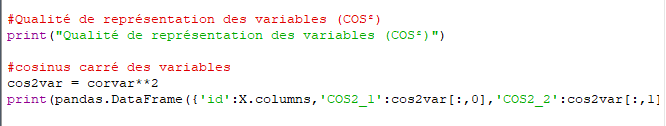
****

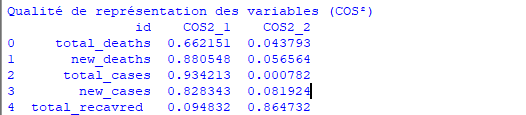


**Interprétation** :

1. Le premier axe est très corrélées positivement avec les variable ( « total\_death » ‎‎« total\_cases ‎»‎‎«‎ new\_death ‎ ‎»‎ ‎«‎‎ new cases ‎ ‎»‎ ‎)‎ donc on peut dire que ces variables représentent le TAUX\_D'INFECTION ‎dans les différents pays .‎
2. Le 2éme axe est fortement corrélée‎ avec la seul variable "total\_recovered" donc il représente le TAUX DE RECOUVREMENTS dans les differents pays.
3. Les variables ‎ « ‎new\_death ‎ ‎»‎ et ‎«‎‎ new cases ‎ ‎»‎ ‎sont fortement corrolées positivement c-à-d ces deux variables sont proportionnelles .
4. Toutes les variables sont corrélées positivement avec le premier facteur principal.

**#Qualité de représentation des variables (COS²)**





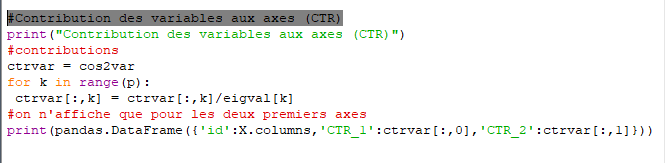
Interprétation :

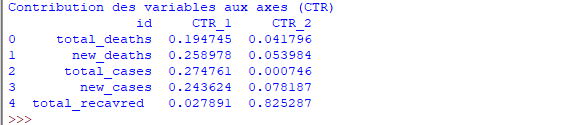
1. Tous les variables sont très bien représentés dans le plan sauf la variable

«  total\_death » car d’après les valeurs ci-dessus la somme des variables de ‎qualité cos² du 1ere et 2éme axes sont très proche à 1 mais ce n’est pas le cas ‎de la variable « ‎total\_death ‎ » ‎.

1. La variable ’total\_cases’ est mieux représentée sur le premier plan factorielle avec une valeur de 0.934213 .
2. La variable ’total\_recovred ‘ est mieux représentée sur le deuxième plan factorielle avec une valeur de 0.864732 .

**#Contribution des variables aux axes (CTR) :**

****

****