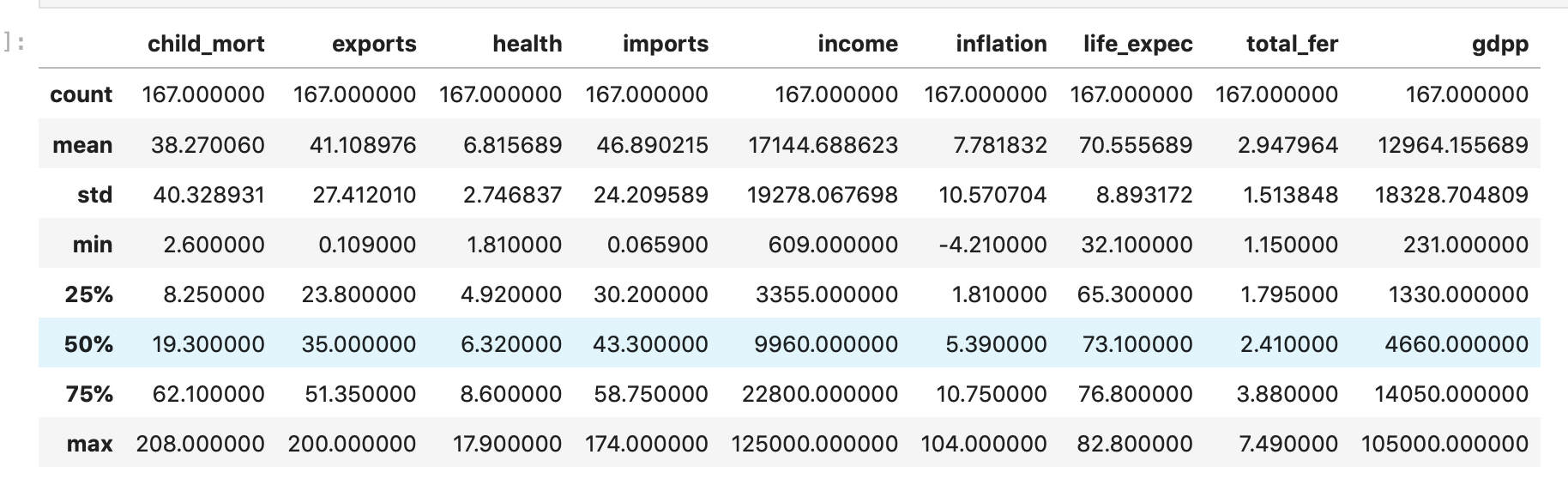
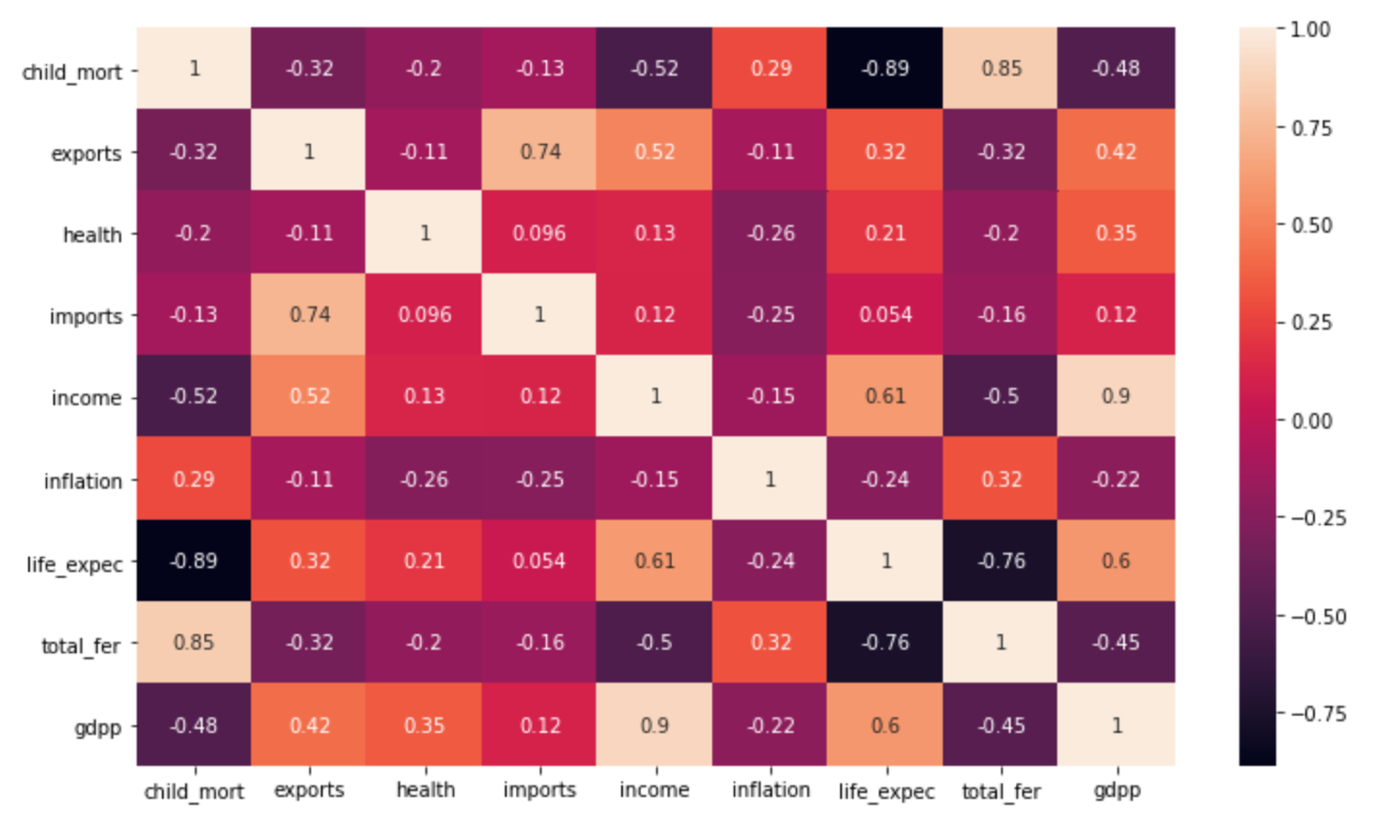
**CLUSTERING**

**Objective:** To categorize countries using 9 socio-economic and health factors associated with 167 countries

**Problem Statement:** After raising $ 10 million, CEO of an NGO needs to decide how to use the money strategically and effectively on countries that are in direct need. Based on socio economic and health factors data collected for 167 countries, the CEO needs to decide which countries should be given more focus.

**Exploratory Data Analysis:**





**child\_mort** : There seems to be huge difference between min and max values of child mortality factor, along with large difference in mean and median

**health**: Mean and median of health spend as % of gdpp is similar

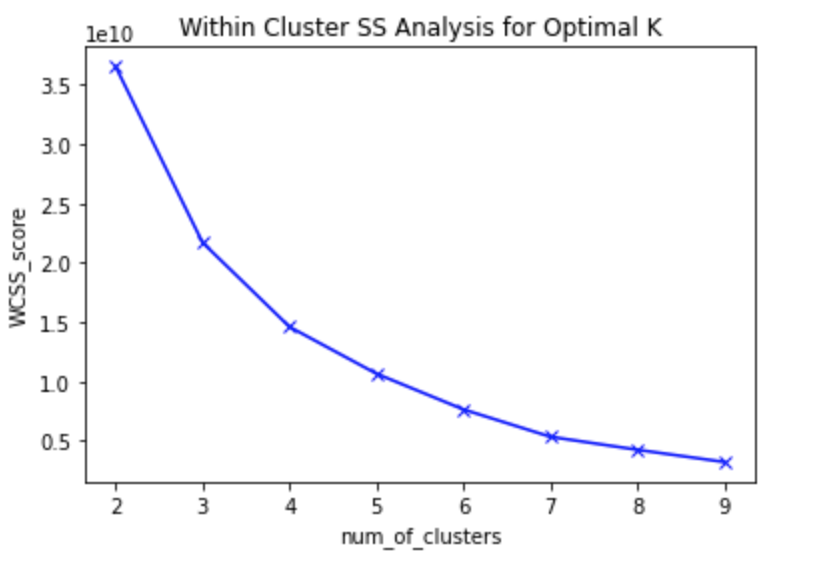
**income**: Spread is huge from the mean income value

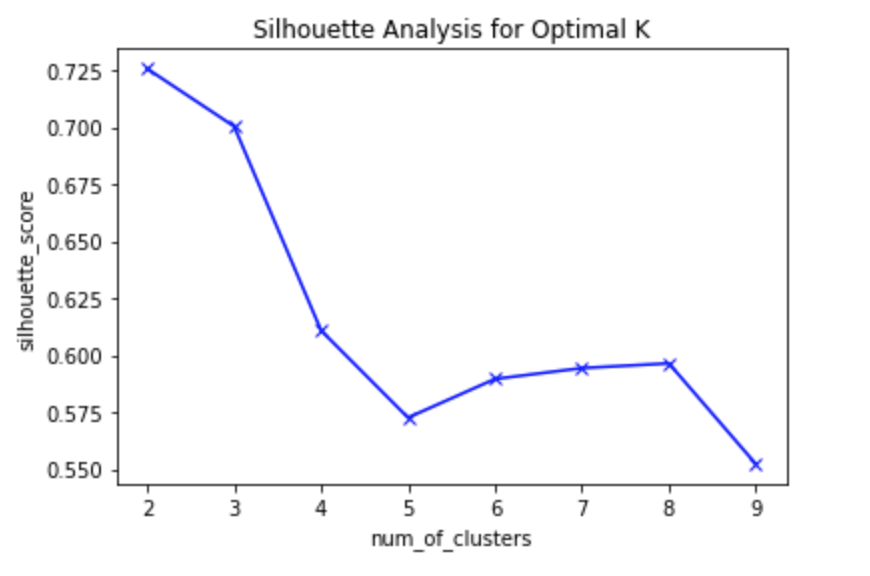
**gdpp**: Huge difference between mean and median value

The correlation matrix shows that most the factors taken into consideration are not correlated with each other.

**Clustering Analysis:**

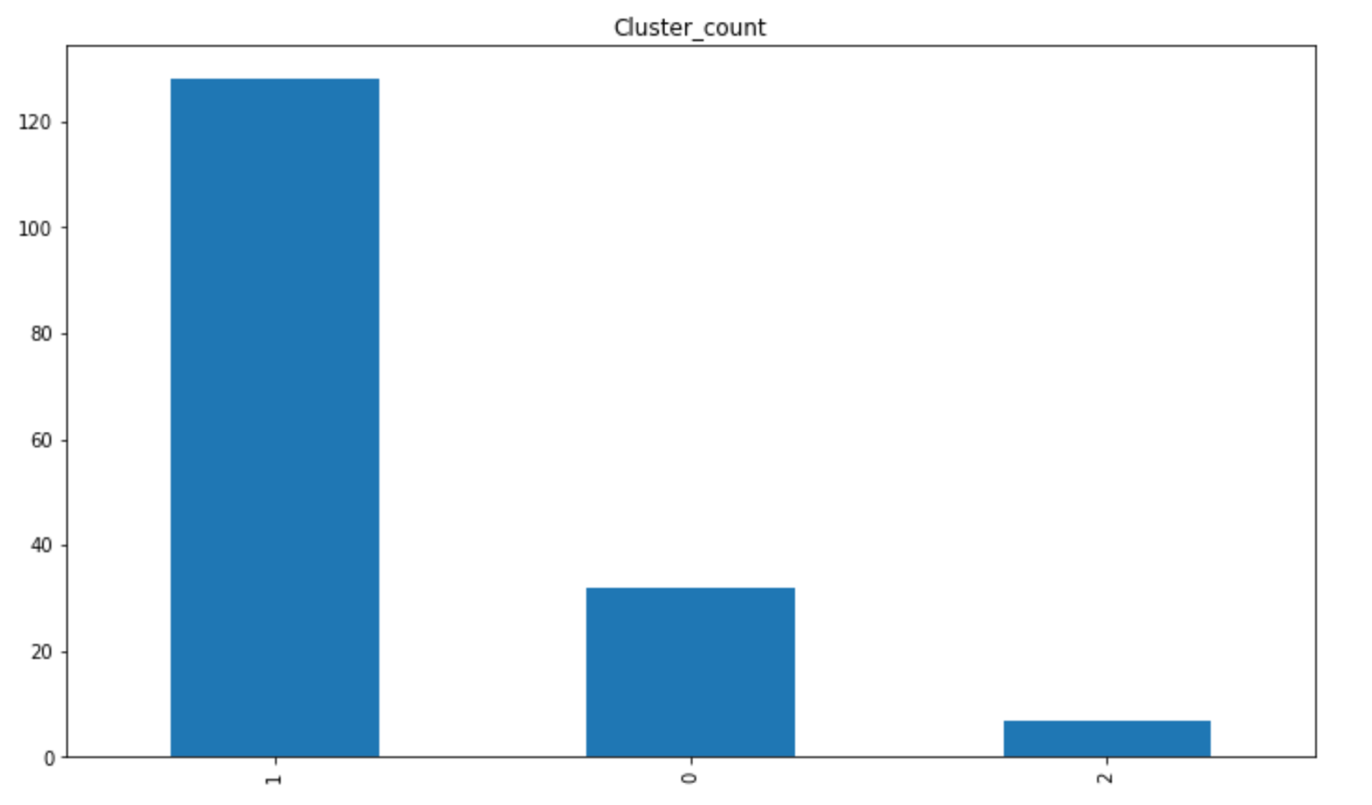
To find the optimal number of clusters ‘k’, WCSS and silhouette methods are used and got the below outputs:





From the above two analyses, the optimal ‘k’ value is found to be **k = 3**

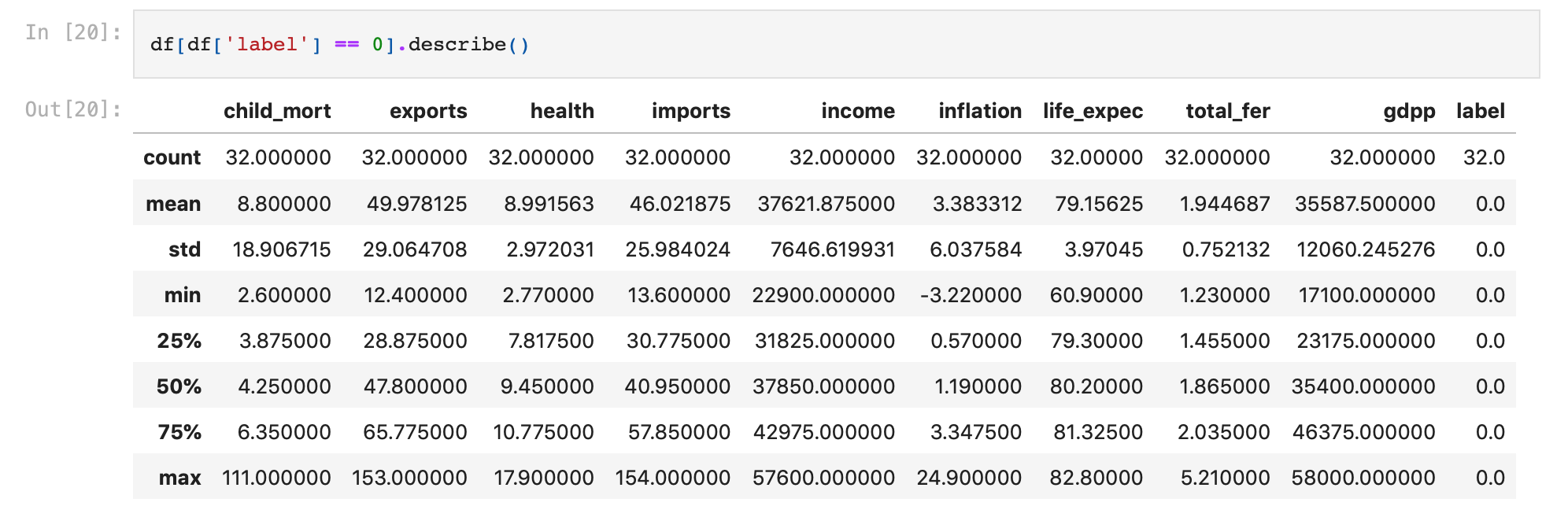
Using k-means clustering and k=3, the below output is obtained:



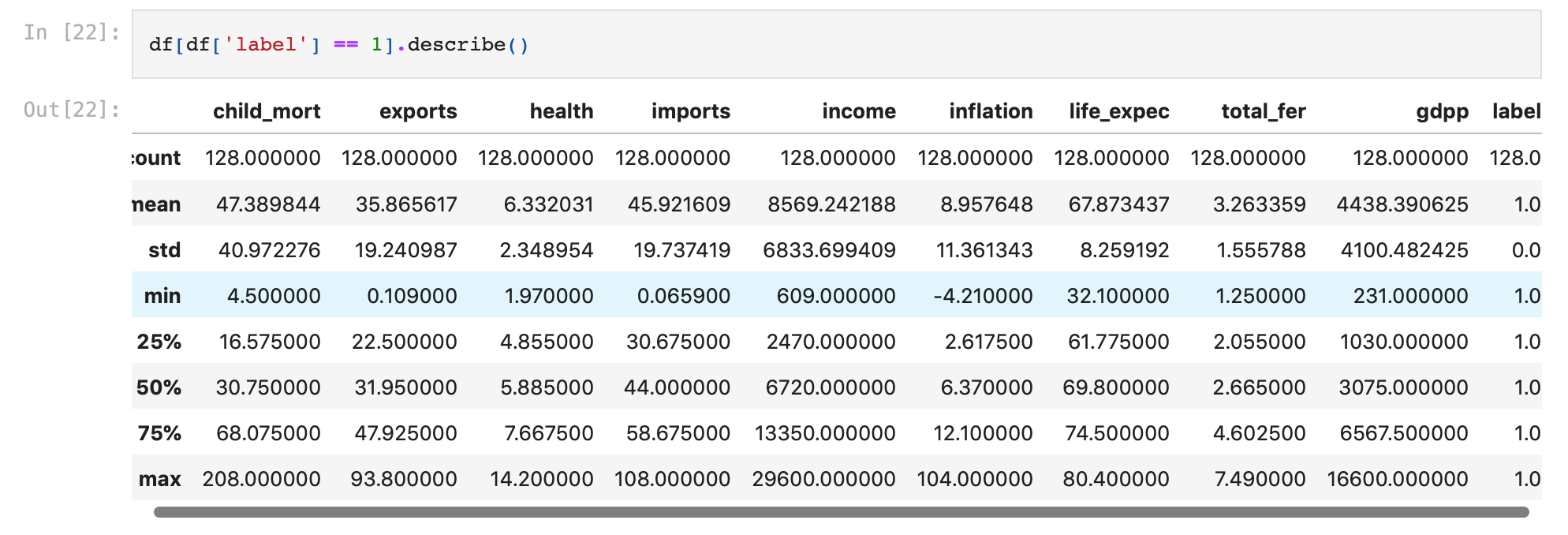
Huge difference in cluster count can be understood from the above frequency count chart.

**Results:**

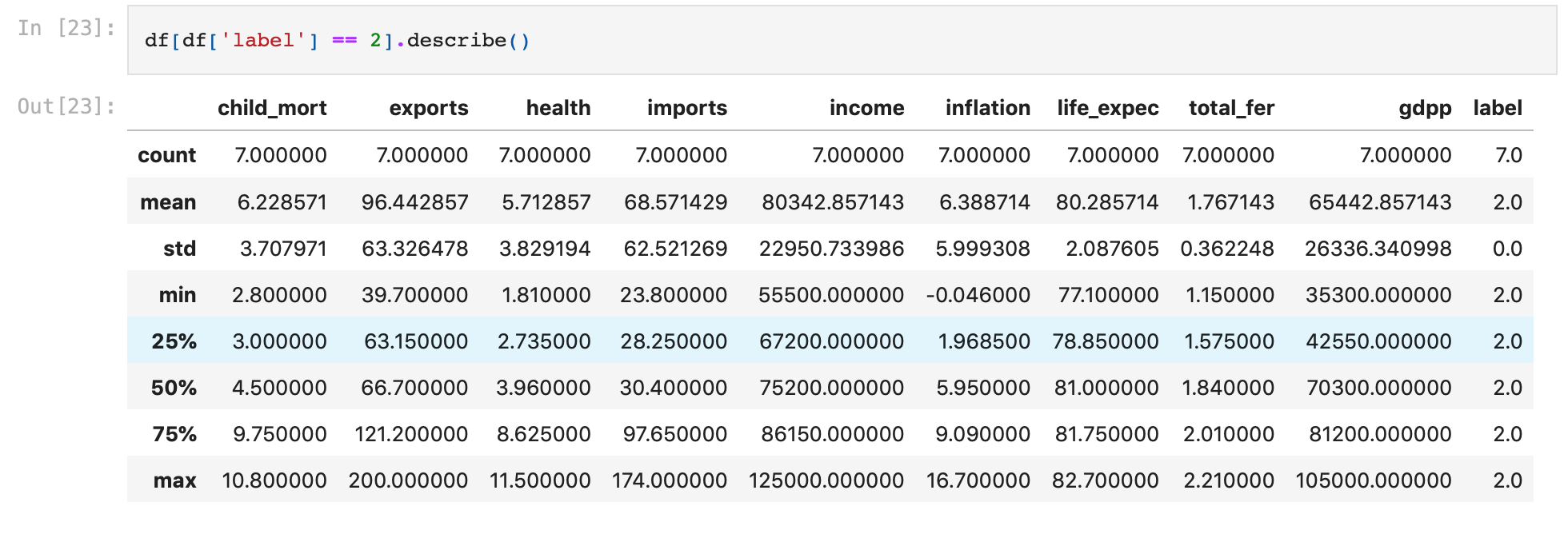
Cluster 0:



Cluster 1:



Cluster 2:



From the above three clusters, it can be found that:

* Allocation of fund needs to be concentrated towards country in Cluster 1, as their socio economic and health situation is bad as compared to countries in other clusters
* Countries in cluster 2 seems to be rich countries with the socio economic and health status being the best among all clusters

**References:**

[**https://www.kaggle.com/datasets/rohan0301/unsupervised-learning-on-country-data**](https://www.kaggle.com/datasets/rohan0301/unsupervised-learning-on-country-data)

[**https://www.analyticsvidhya.com/blog/2021/05/k-mean-getting-the-optimal-number-of-clusters/#:~:text=The%20optimal%20number%20of%20clusters%20k%20is%20one%20that%20maximizes,Optimal%20of%202%20clusters**](https://www.analyticsvidhya.com/blog/2021/05/k-mean-getting-the-optimal-number-of-clusters/#:~:text=The%20optimal%20number%20of%20clusters%20k%20is%20one%20that%20maximizes,Optimal%20of%202%20clusters)