



# COUNTRY CLASSIFICATION FOR FINANCIAL AID

HELP INTERNATIONAL

---







# PROFILE



***Help International*** is an nongovernmental organization (NGO) that focused in reducing poverty, provide facility and basic needs for people in third-world country in case of catastrophe.



---

## BACKGROUND

---

*Help International* will give financial aid from the collected donation (USD 10mio) for countries in need.

The criteria to be fulfilled is the **health condition and socioeconomic** in each countries.



# DATA

The data has 167 rows and 10 columns, as seen here.

- **Country** : name of the country.
- **Child Mortality**: child mortality under 5 years old per 1,000 births.
- **Export** : good and services export rate per capita.
- **Health Cost**: total of health cost per capita.
- **Import**: good and services import rate per capita.
- **Income Per Capita**: individual net income.
- **Inflation**: inflation rate as per GDP per year.
- **Life Expectancy**: the average number of years a newborn child can live after birth (with the same child mortality rate).
- **Fertility Rate**: the total of childbirth per woman (with the same fertility rate).
- **GDP Per Capita**: GDP per capita (Total of GDP divided by the total population).

	Country	Child Mortality	Export	Health Cost	Import	Income Per Capita	Inflation	Life Expectancy	Fertility Rate	GDP Per Capita
0	Afghanistan	90.2	10.0	7.58	44.9	1610	9.44	56.2	5.82	553
1	Albania	16.6	28.0	6.55	48.6	9930	4.49	76.3	1.65	4090
2	Algeria	27.3	38.4	4.17	31.4	12900	16.10	76.5	2.89	4460
3	Angola	119.0	62.3	2.85	42.9	5900	22.40	60.1	6.16	3530
4	Antigua and Barbuda	10.3	45.5	6.03	58.9	19100	1.44	76.8	2.13	12200



# DATA

There is 167 data, and each column has values (non-null).

```
RangeIndex: 167 entries, 0 to 166
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Country               167 non-null   object
1   Child Mortality       167 non-null   float64
2   Export                167 non-null   float64
3   Health Cost           167 non-null   float64
4   Import                167 non-null   float64
5   Income Per Capita     167 non-null   int64
6   Inflation             167 non-null   float64
7   Life Expectancy       167 non-null   float64
8   Fertility Rate        167 non-null   float64
9   GDP Per Capita        167 non-null   int64
dtypes: float64(7), int64(2), object(1)
```



# GENERAL ANALYSIS

Categorize each column to the required parameter (health and socioeconomic).

Country	Child Mortality	Export	Health Cost	Import	Income Per Capita	Inflation	Life Expectancy	Fertility Rate	GDP Per Capita
---------	-----------------	--------	-------------	--------	-------------------	-----------	-----------------	----------------	----------------

## Health

- Health Cost
- Child Mortality
- Fertility Rate
- Life Expectancy

## Socioeconomic

- GDP Per Capita
- Income Per Capita
- Import
- Export
- Inflation





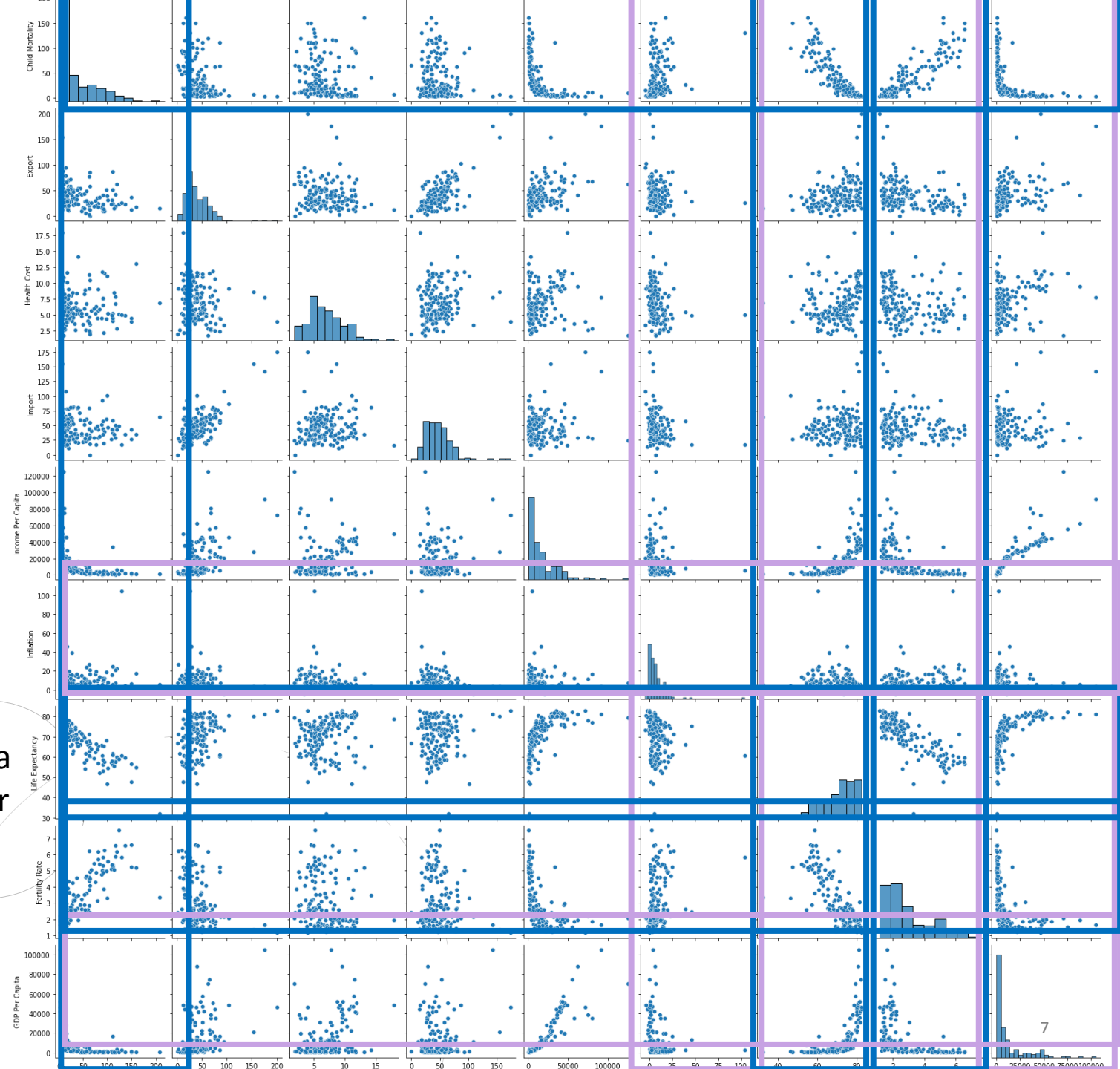
# GENERAL ANALYSIS

Looking for correlation among values.

Data that could be recognized easily as correlated are:

- **Socioeconomic**: GDP Per capita and Inflation.
- **Health**: Fertility Rate, Life Expectancy, and Child Mortality.

As seen on graphic, the correlation is most likely **negative correlation**, e.g: when the GDP per capita getting higher, the Child Mortality getting lower, or the lower the Inflation, the higher the Life Expectancy.



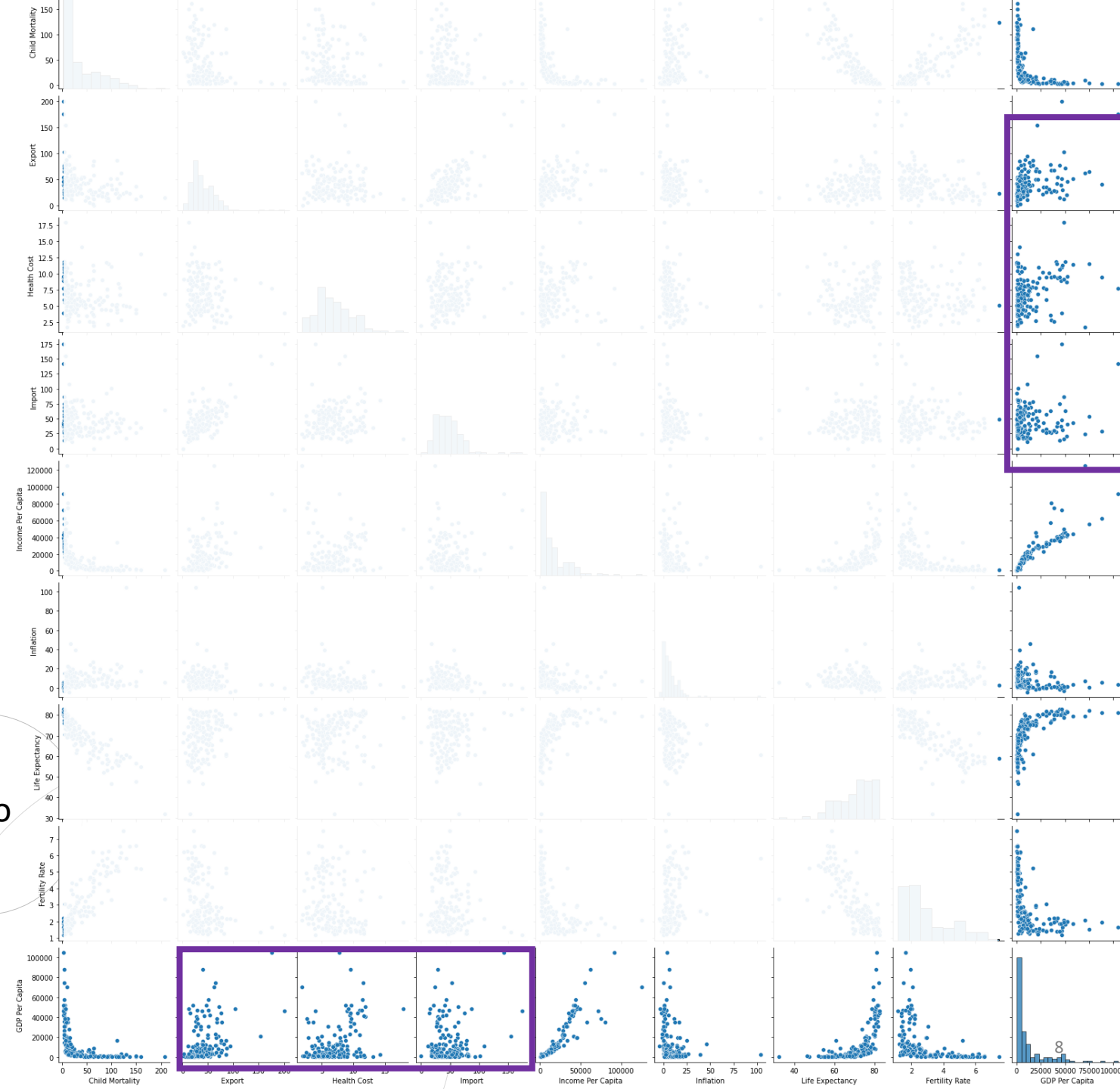


# GENERAL ANALYSIS

Based on *World Bank*, country's **economy** and **wellbeing** could be measured through GDP per capita.

If we see the correlation between **GDP Per capita** and other column, there are 3 columns that have different correlation from before, those are: **Import**, **Export**, and **Health**.

Because Export and Import are both in socio-economic parameter. The data that we are going to focus on is **Health**.





# GENERAL ANALYSIS

5 Countries with the highest GDP per capita are all developed countries.

On the other hand, if we see the countries with the highest health cost, there are **2 underdeveloped countries**. It should be questioned that there is **significant difference** between the GDP per capita and the health cost, or **inequality**.

Data from GDP Per Capita–Health Cost.

## Country List (Highest GDP Per Capita)

	Country	GDP Per Capita	Health Cost
91	Luxembourg	105000	7.77
114	Norway	87800	9.48
145	Switzerland	74600	11.50
123	Qatar	70300	1.81
44	Denmark	58000	11.40
...	...	...	...
132	Sierra Leone	399	13.10
112	Niger	348	5.16
37	Congo, Dem. Rep.	334	7.91
88	Liberia	327	11.80
26	Burundi	231	11.60

167 rows × 3 columns

## Country List (Highest Health Cost)

	Country	GDP Per Capita	Health Cost
159	United States	48400	17.90
101	Micronesia, Fed. Sts.	2860	14.20
132	Sierra Leone	399	13.10
110	Netherlands	50300	11.90
54	France	40600	11.90
...	...	...	...
154	Turkmenistan	4440	2.50
38	Congo, Rep.	2740	2.46
116	Pakistan	1040	2.20
107	Myanmar	988	1.97
123	Qatar	70300	1.81

167 rows × 3 columns



# OUTLIERS

As seen on the graphic, there are some data that are beyond the range of other values. Those are outliers of the **GDP Per capita**.

So, the outliers filtering focused on **GDP Per capita**.

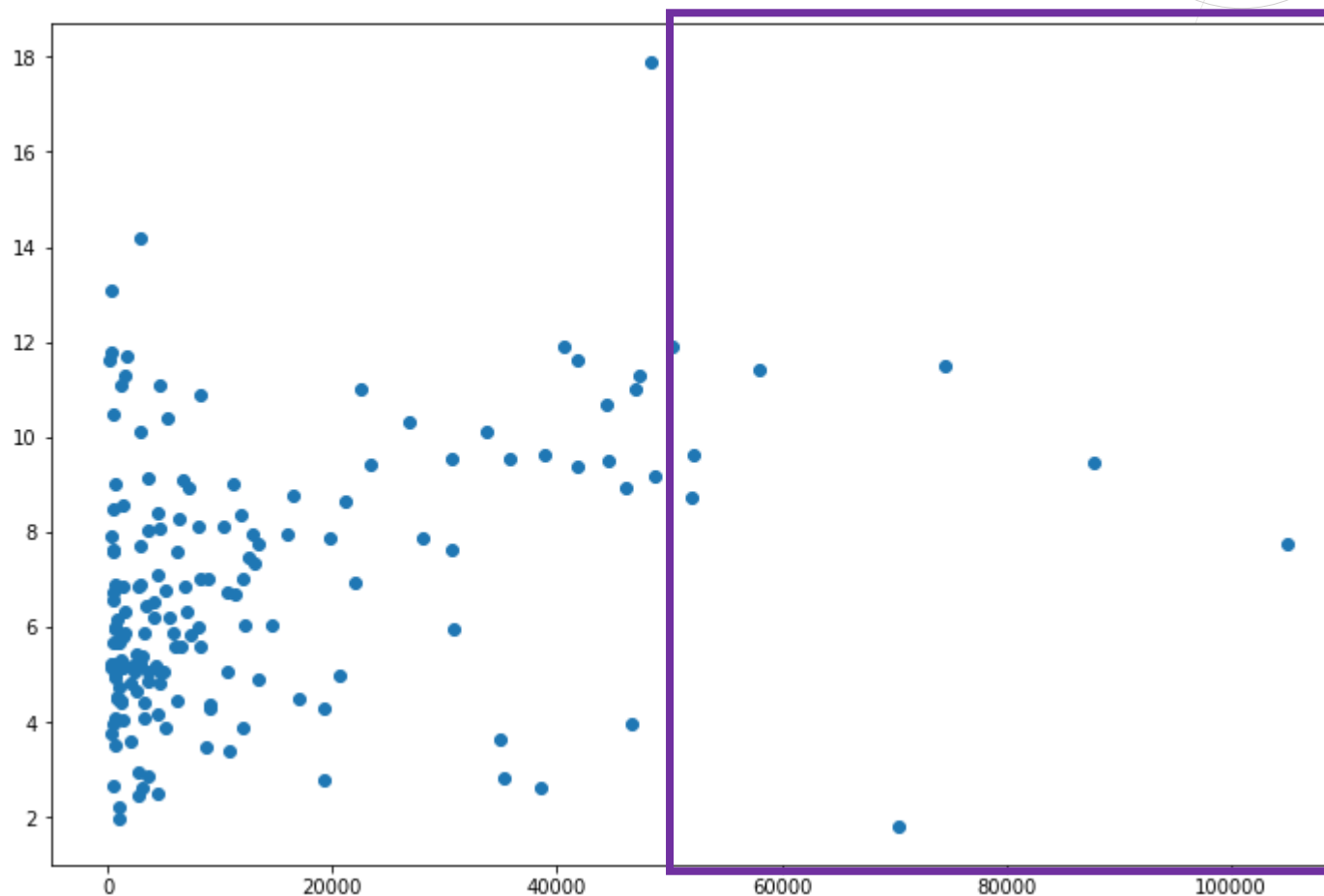
With the result:

**Q1 = 1330**

**Q3 = 14050**

**IQR = 12445**

## Identifying outliers from GDP Per capita–Health





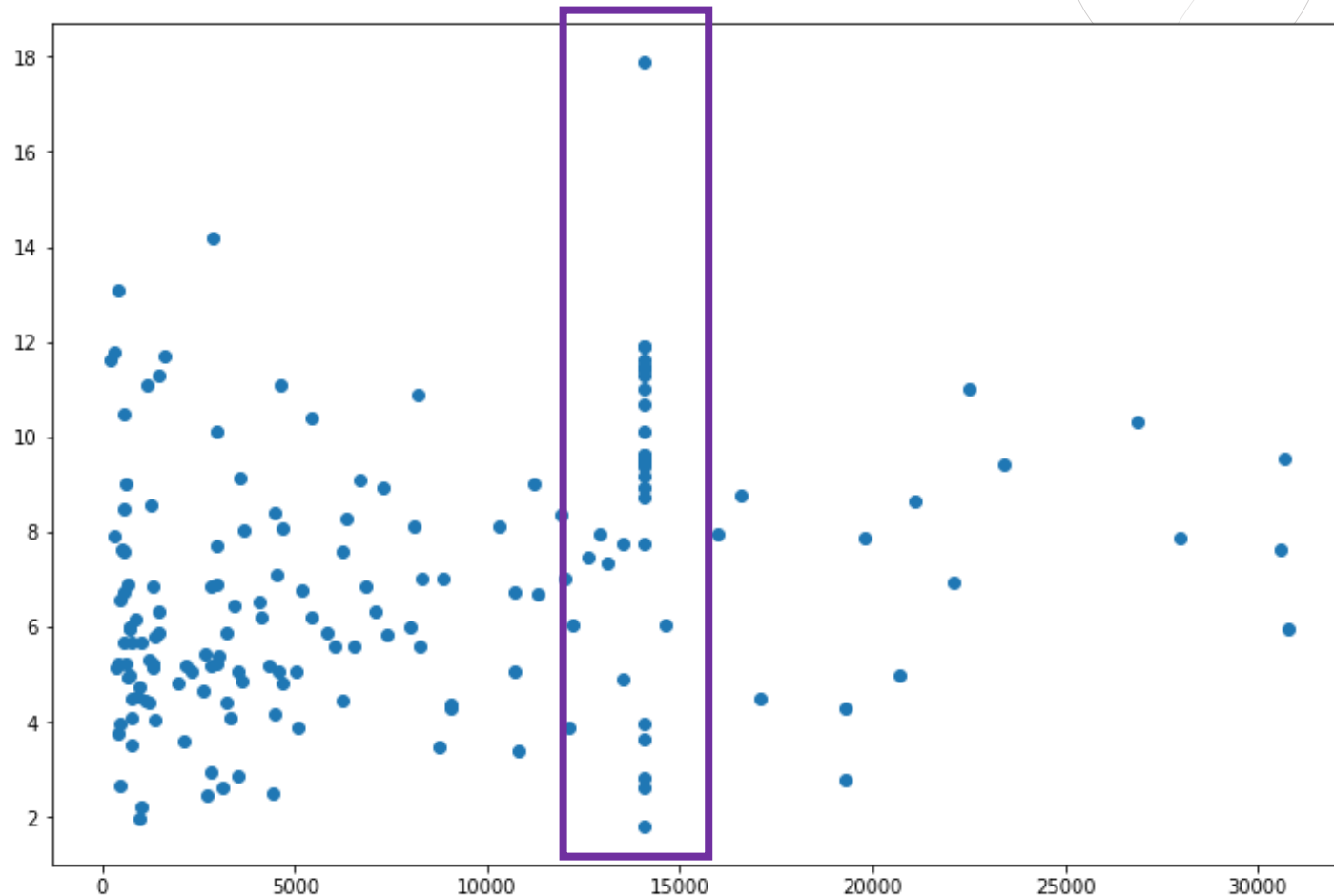
# OUTLIERS

The outliers are turned into the base-range.

So, if it is higher than Q3, it is changed to Q3. The same for those values under Q1.

**There are still 167 data.**

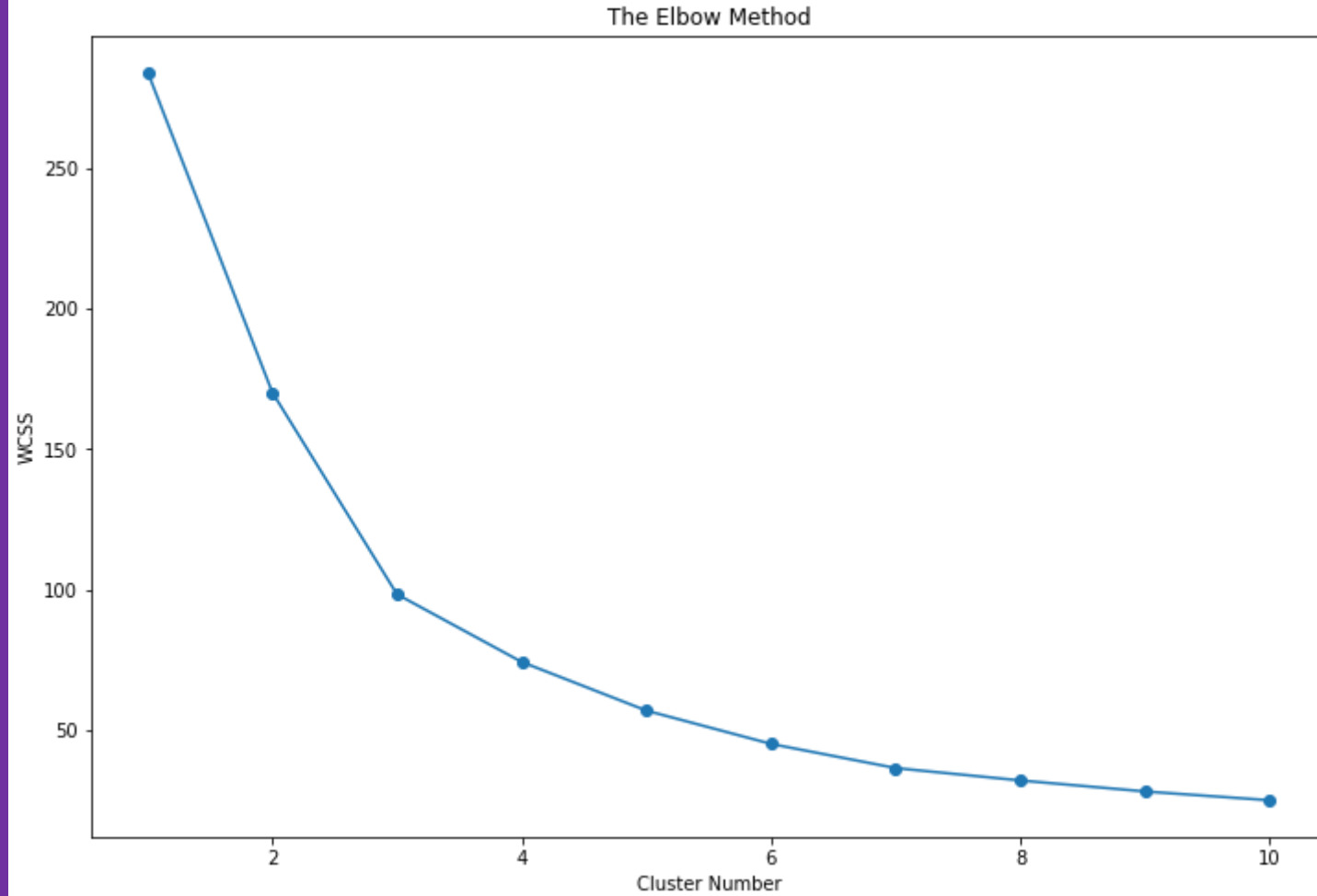
Scatterplot after handling outliers



# CLUSTERING

For clustering, **elbow method** is used. The number of cluster is based on which number has an acute angle ( $<60^\circ$ )

From the graphic, **the cluster number = 4**, with consideration the degree created after 4 is already an obtuse angle and there is no significant difference.







# CLUSTERING

There are 4 clusters with criteria as followed:

- **Cluster 0**

GDP Per capita < 10000; Health > 8

- **Cluster 1**

GDP Per capita < 10000; Health < 8

- **Cluster 2**

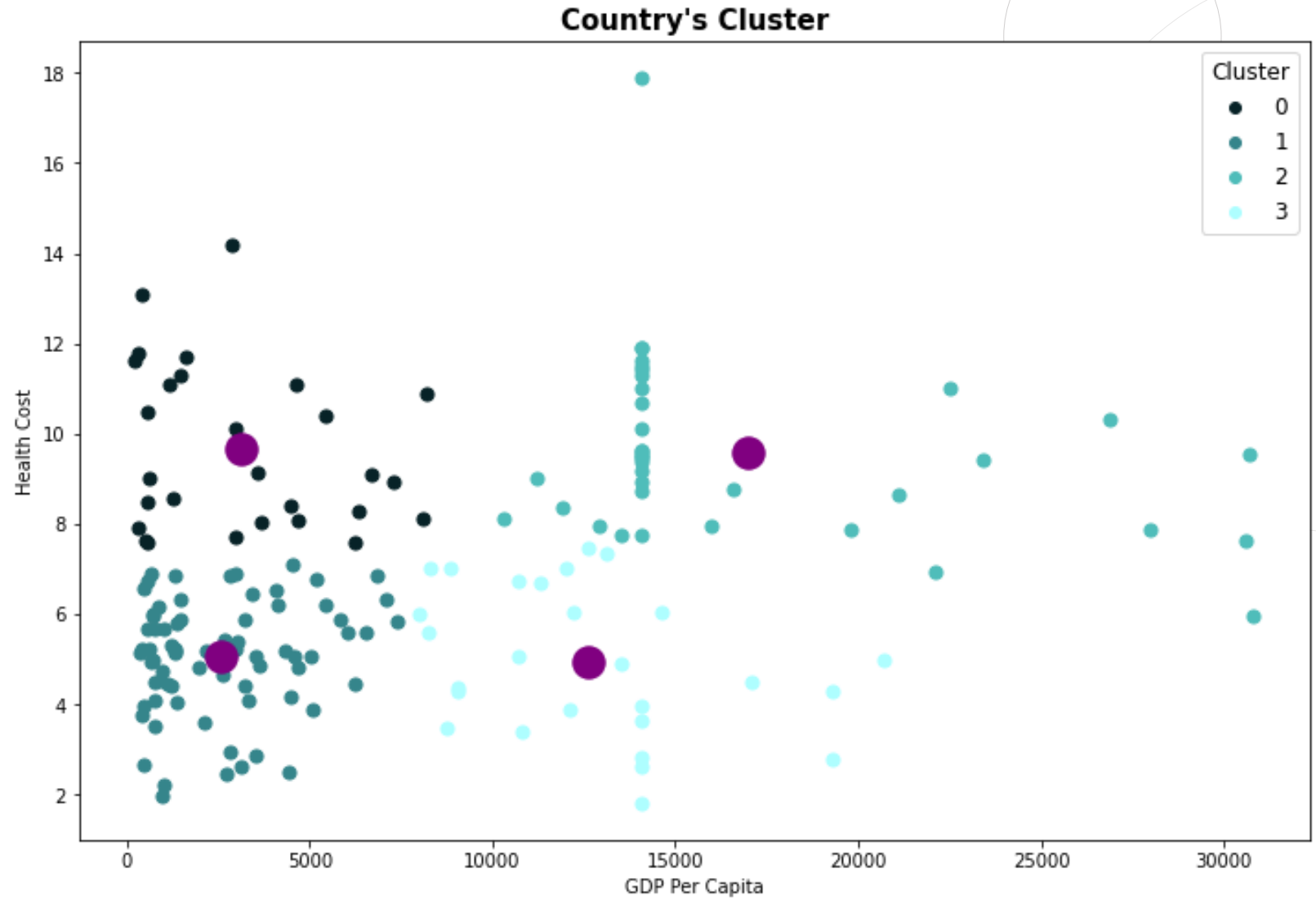
GDP Per capita between > 10000;

Health > 8

- **Cluster 3**

GDP Per capita between > 10000;

Health < 8





# CLUSTERING

Based on the clustering result, there are still some data that overlapping between each clusters. Therefore, the cluster choice is based on a new parameter. This parameter is from dividing the **'Health' with 'GDP Per capita'**.

```
df['Ratio'] = (df['Health Cost']/df['GDP Per Capita'])*100
```

## Why?

So, it is more obvious to see how different is the health cost from a country to its GDP per capita.

If the **parameter is higher**, it means **the health cost is higher and not balanced with a low GDP per capita**.

	Country	GDP Per Capita	Health Cost	Cluster	Ratio
0	Afghanistan	553.0	7.58	0	1.370705
1	Albania	4090.0	6.55	1	0.160147
2	Algeria	4460.0	4.17	1	0.093498
3	Angola	3530.0	2.85	1	0.080737
4	Antigua and Barbuda	12200.0	6.03	2	0.049426



# CLUSTERING

- Create ratio for each clusters.
- Identify cluster with higher ratio.

## CLUSTER 0 28 countries

Country	GDP Per Capita	Health Cost	Cluster	Ratio
Burundi	231.0	11.60	0	5.021645
Liberia	327.0	11.80	0	3.608563
Sierra Leone	399.0	13.10	0	3.283208
Congo, Dem. Rep.	334.0	7.91	0	2.368263
Rwanda	563.0	10.50	0	1.865009

## CLUSTER 1 75 countries

Country	GDP Per Capita	Health Cost	Cluster	Ratio
Niger	348.0	5.16	1	1.482759
Malawi	459.0	6.59	1	1.435730
Mozambique	419.0	5.21	1	1.243437
Burkina Faso	575.0	6.74	1	1.172174
Haiti	662.0	6.91	1	1.043807

## CLUSTER 2 27 countries

Country	GDP Per Capita	Health Cost	Cluster	Ratio
Suriname	8300.0	7.01	2	0.084458
Lebanon	8860.0	7.03	2	0.079345
Mauritius	8000.0	6.00	2	0.075000
Romania	8230.0	5.58	2	0.067801
Turkey	10700.0	6.74	2	0.062991

## CLUSTER 3 37 countries

Country	GDP Per Capita	Health Cost	Cluster	Ratio
United States	14050.0	17.90	3	0.127402
Netherlands	14050.0	11.90	3	0.084698
France	14050.0	11.90	3	0.084698
Germany	14050.0	11.60	3	0.082562
Switzerland	14050.0	11.50	3	0.081851



## CLUSTER 0

Here is the table if choosing the country only based on 'GDP Per capita' or 'Health'.

From this two table, there are **4 countries** in both tables.

What lacking from this method is there is a bias between GDP Per capita or Health cost that has small difference.

*E.g:* Compared to Rwanda, GDP per capita of Congo, Dem. Rep. is lower. But the health cost is higher in Rwanda.

*Which one should be chosen, Rwanda or Congo, Dem. Rep?*

If choosing the country only based on one value.

Country	GDP Per Capita	Health Cost	Cluster	Ratio
Burundi	231.0	11.60	0	5.021645
Liberia	327.0	11.80	0	3.608563
Congo, Dem. Rep.	334.0	7.91	0	2.368263
Sierra Leone	399.0	13.10	0	3.283208
Togo	488.0	7.65	0	1.567623
Guinea-Bissau	547.0	8.50	0	1.553931
Afghanistan	553.0	7.58	0	1.370705
Rwanda	563.0	10.50	0	1.865009
Uganda	595.0	9.01	0	1.514286
Lesotho	1170.0	11.10	0	0.948718

List of Country  
(10 Lowest GDP Per capita)

Country	GDP Per Capita	Health Cost	Cluster	Ratio
Micronesia, Fed. Sts.	2860.0	14.2	0	0.496503
Sierra Leone	399.0	13.1	0	3.283208
Liberia	327.0	11.8	0	3.608563
Moldova	1630.0	11.7	0	0.717791
Burundi	231.0	11.6	0	5.021645
Kiribati	1490.0	11.3	0	0.758389
Lesotho	1170.0	11.1	0	0.948718
Bosnia and Herzegovina	4610.0	11.1	0	0.240781
Costa Rica	8200.0	10.9	0	0.132927
Rwanda	563.0	10.5	0	1.865009

List of Country  
(10 Highest Health cost)



# FINAL RECOMMENDATION

From cluster 0, there are 4 countries with **ratio > 2**, they are:

**Burundi, Liberia, Sierra Leone, and Congo, Dem. Rep.**

Country	GDP Per Capita	Health Cost	Cluster	Ratio
Burundi	231.0	11.60	0	5.021645
Liberia	327.0	11.80	0	3.608563
Sierra Leone	399.0	13.10	0	3.283208
Congo, Dem. Rep.	334.0	7.91	0	2.368263
Rwanda	563.0	10.50	0	1.865009
Togo	488.0	7.65	0	1.567623
Guinea-Bissau	547.0	8.50	0	1.553931
Uganda	595.0	9.01	0	1.514286
Afghanistan	553.0	7.58	0	1.370705
Lesotho	1170.0	11.10	0	0.948718



**THANK YOU**

