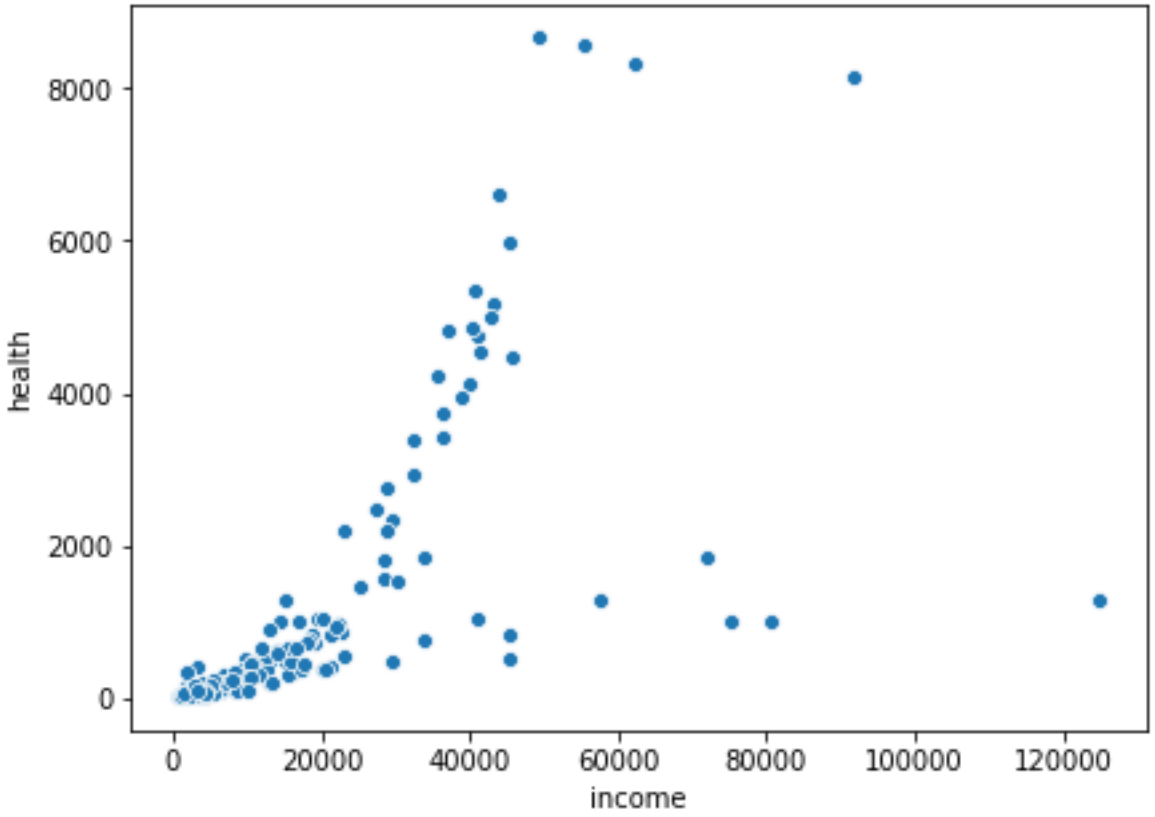
Clustering Assignment

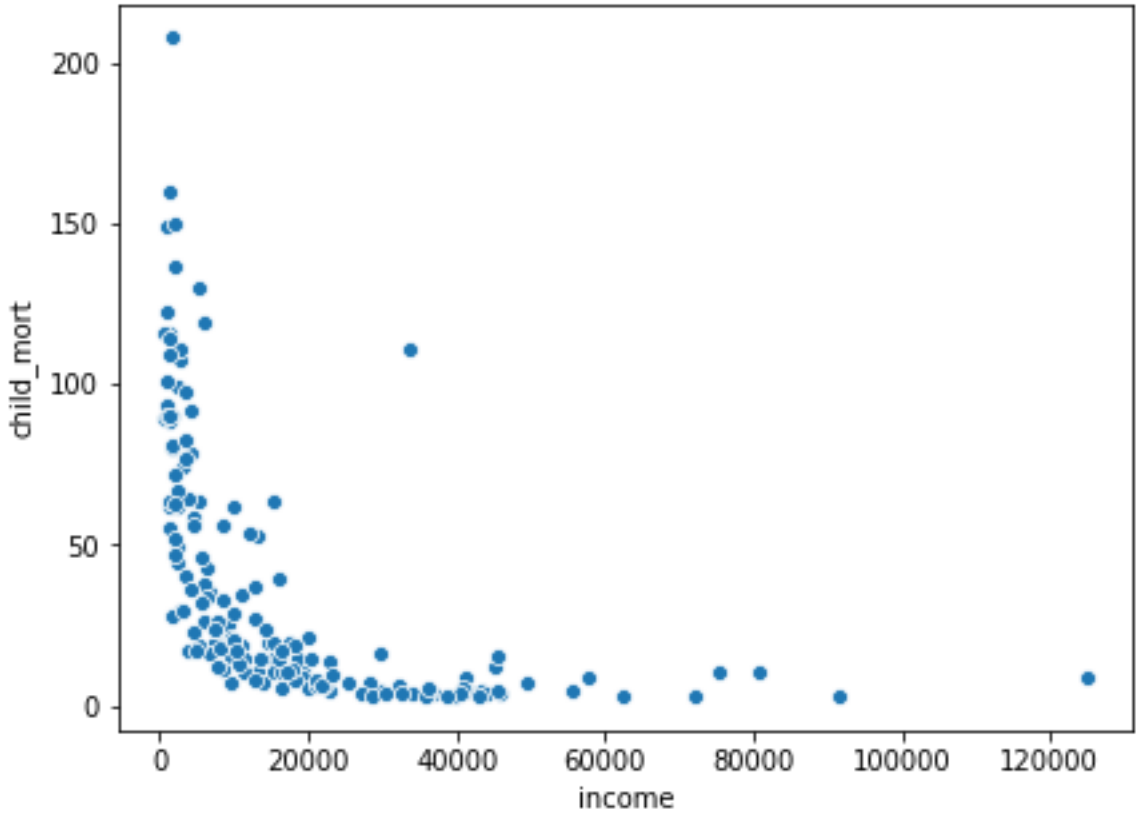
Alaknanda Agarwal

Problem Statement and analysis approach

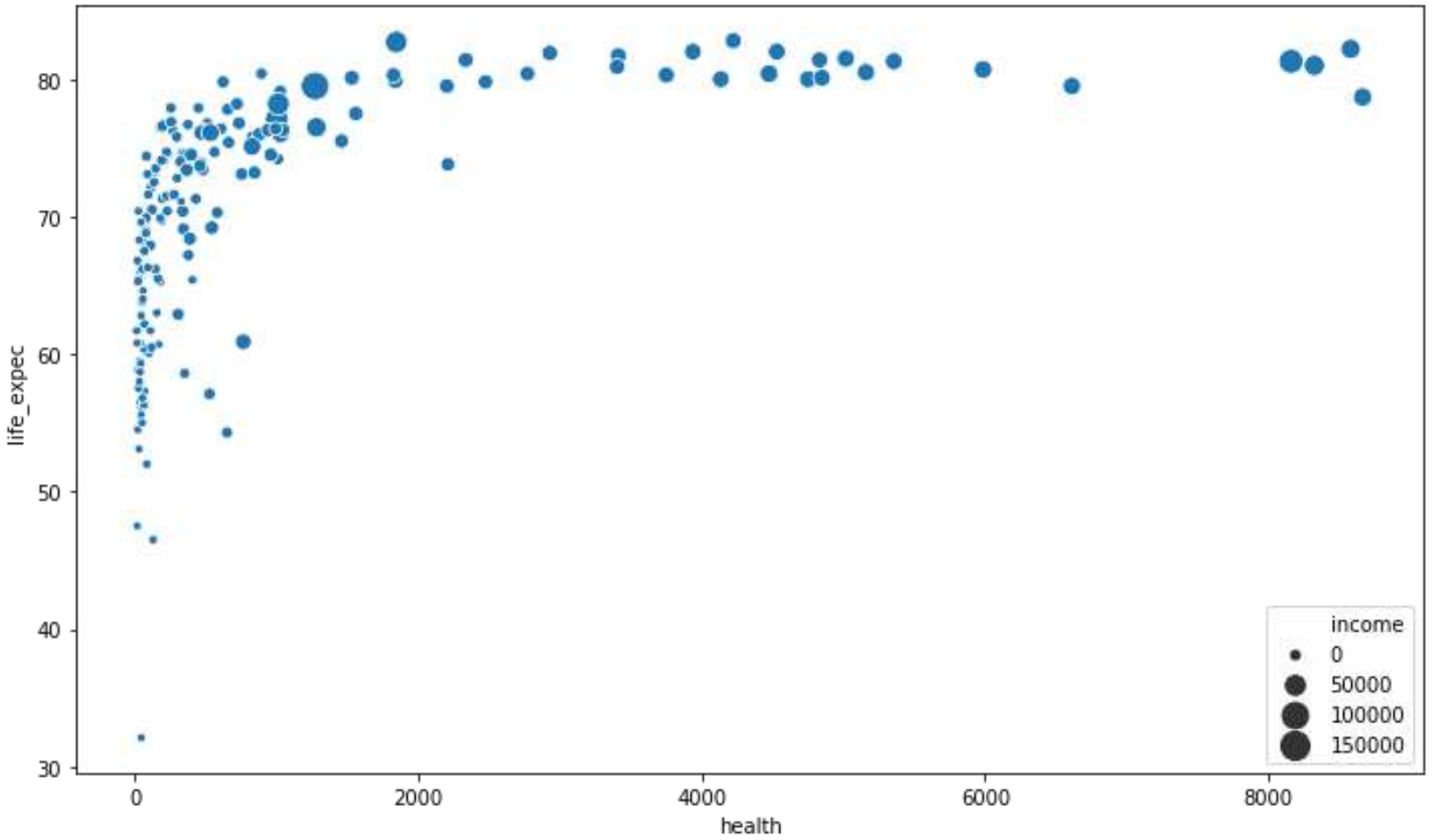
- We have an aid package which will be most effectively used if it is given to countries most in need of it.
- We are given a list of countries, from which we have to select the ones that are most in need of aid.
- We have data on these countries like their GDP, income per person, child mortality rates, etc.
- We will analyze this data, and sort the countries that are in urgent need of the aid.

- These will be countries with the lowest GDPs, highest child mortality rates or lowest income per person etc.
- To identify these countries, we will cluster our countries based on their socio-economic and health parameters.
- The countries having low income and high mortality rates will form one cluster, average income and mortality rates another cluster, high income and low mortality rates a third cluster.



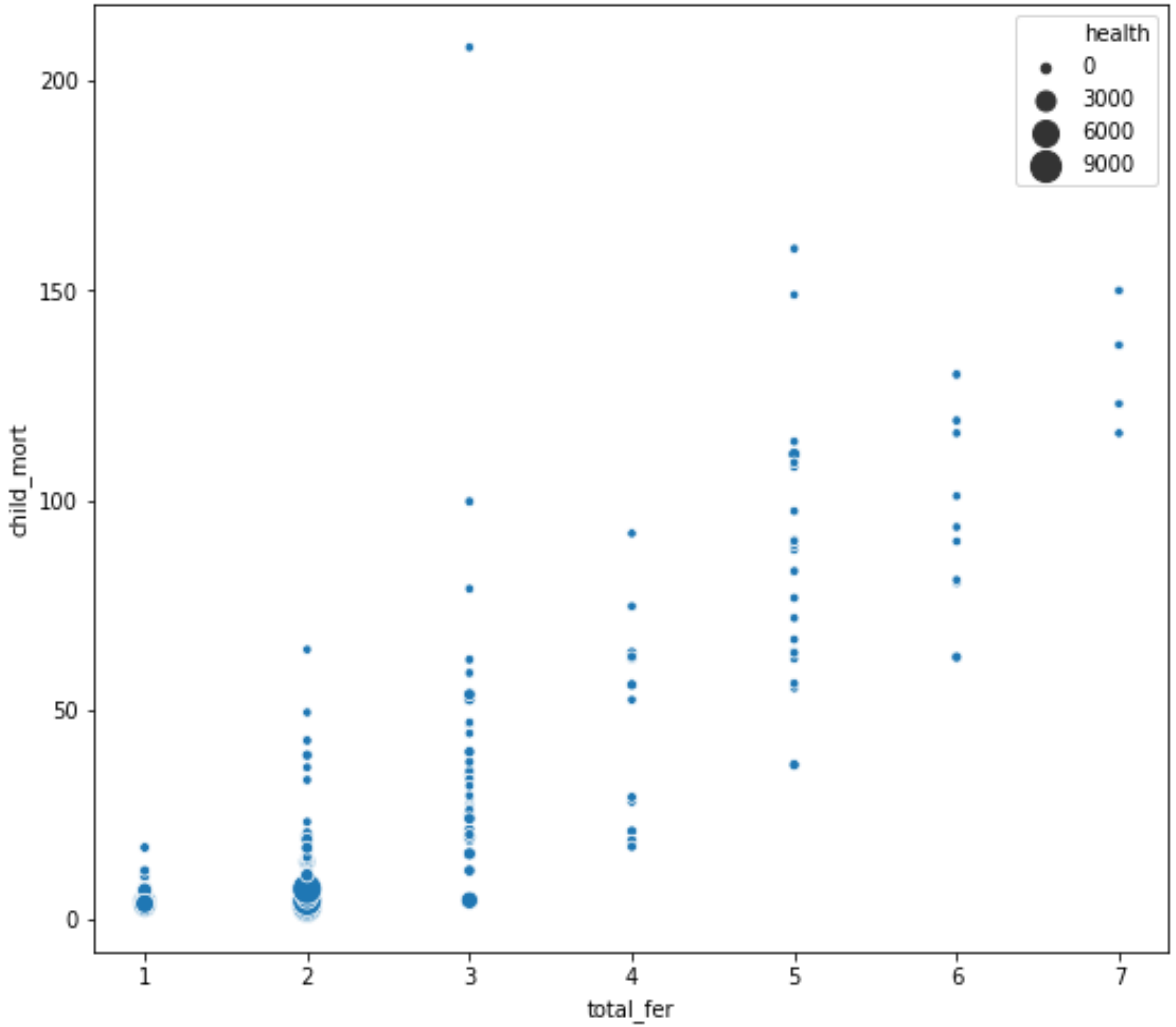


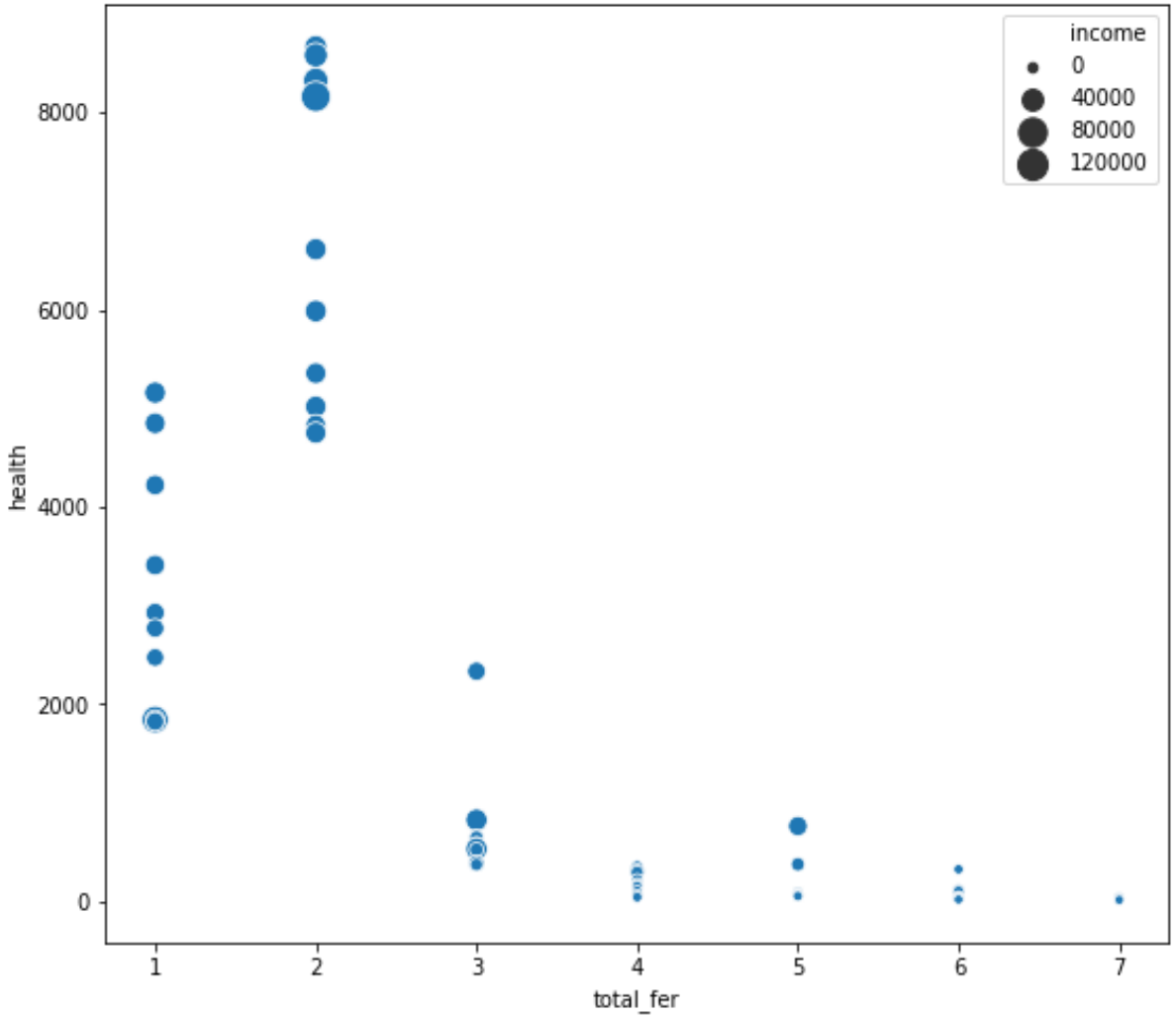
- The expenditure on health is increasing as the income per person of a country is increasing.
- The child mortality is seen to be quite high for many countries with low income per person, and it gradually decreases as the income per person increases.



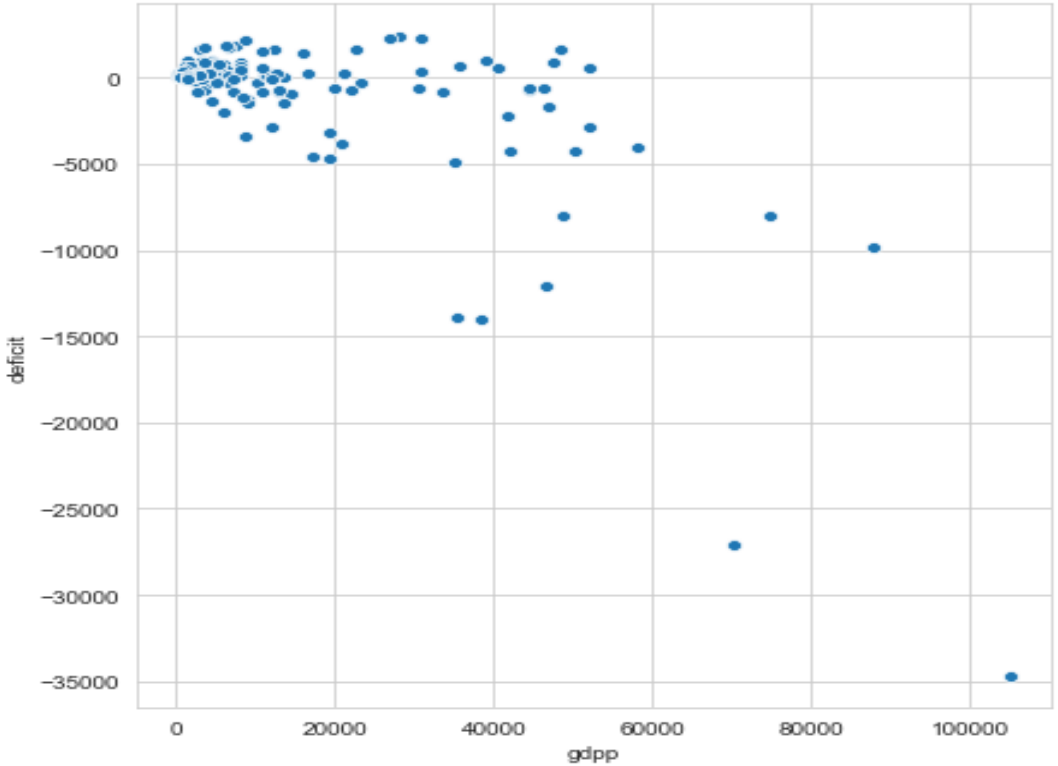
- The life expectancy of a country increases up to a limit, with increase in income and increase in health spend.
- Countries with highest life expectancy, and highest health spending are USA, Switzerland, Norway, Luxembourg and Denmark.

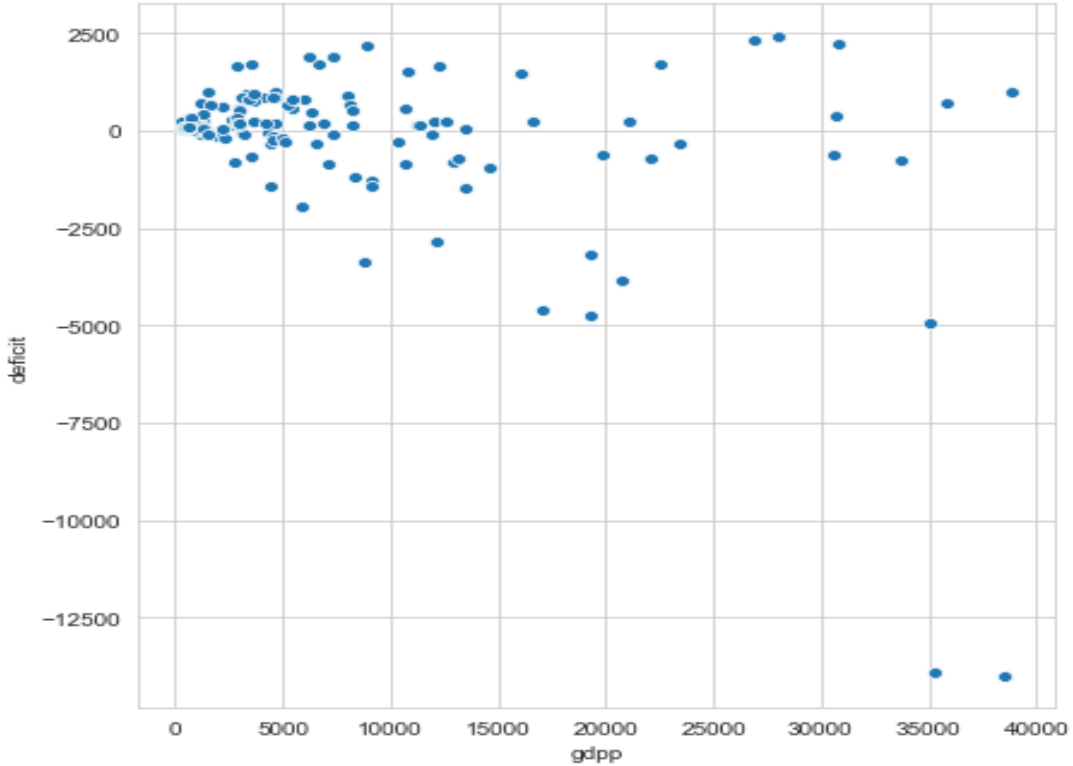
- We can see a country with very low life expectancy- that country is Haiti. The extremely high child mortality rate of Haiti is dragging down it's life expectancy.
- Other countries with low life expectancy and low health spend per person are Lesotho, Central African Republic, Zambia and Malawi.



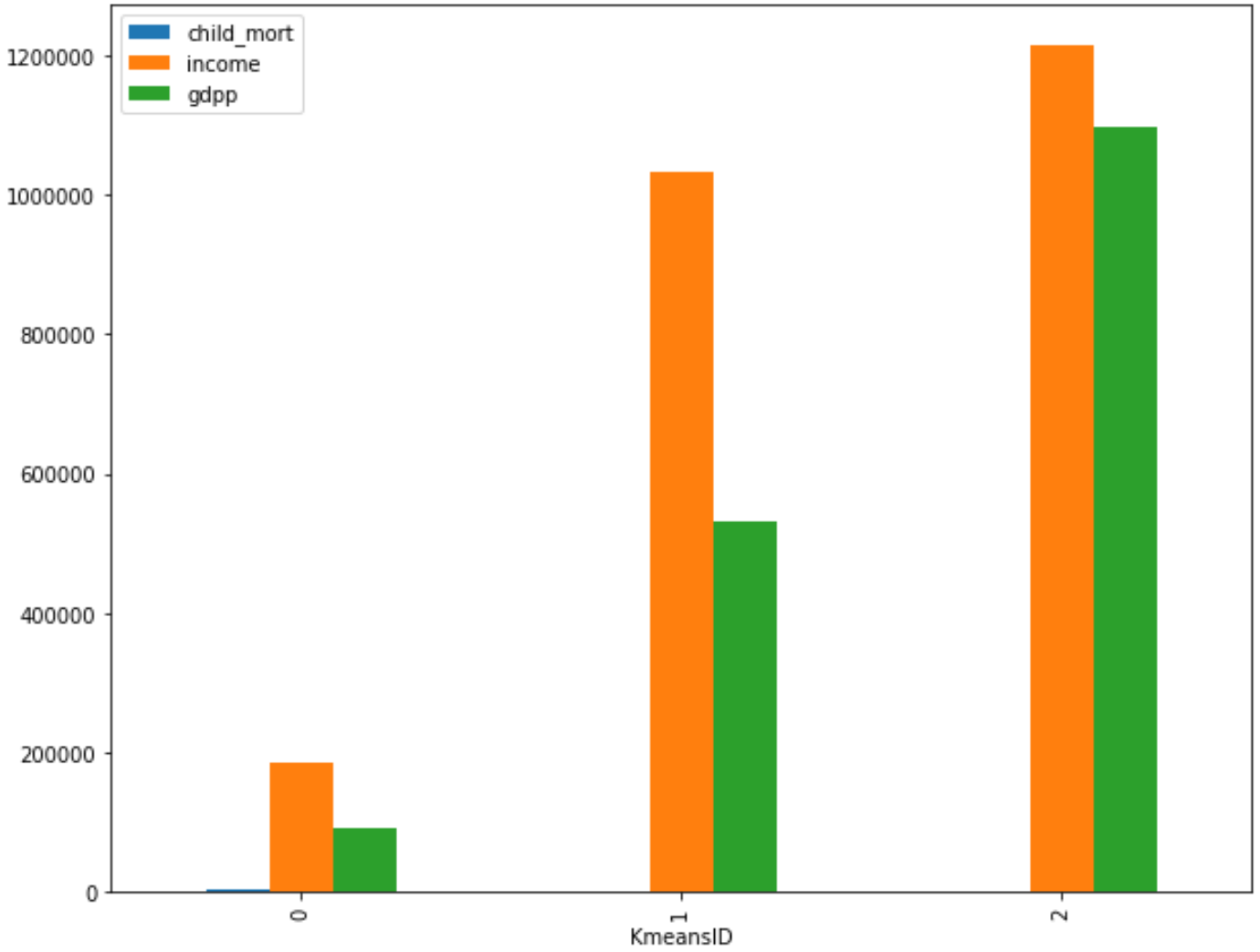


- The child_mortality is increasing as the fertility rate(average number of children per mother) is increasing. Also, the countries with highest health spends are seen to have 1-3 children.
- The fertility rate is 2 in countries with maximum spends on health.

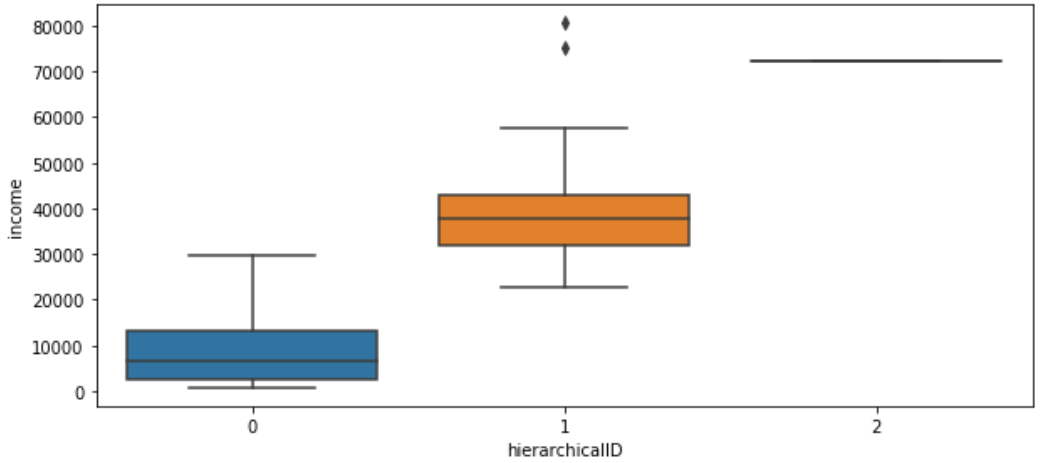


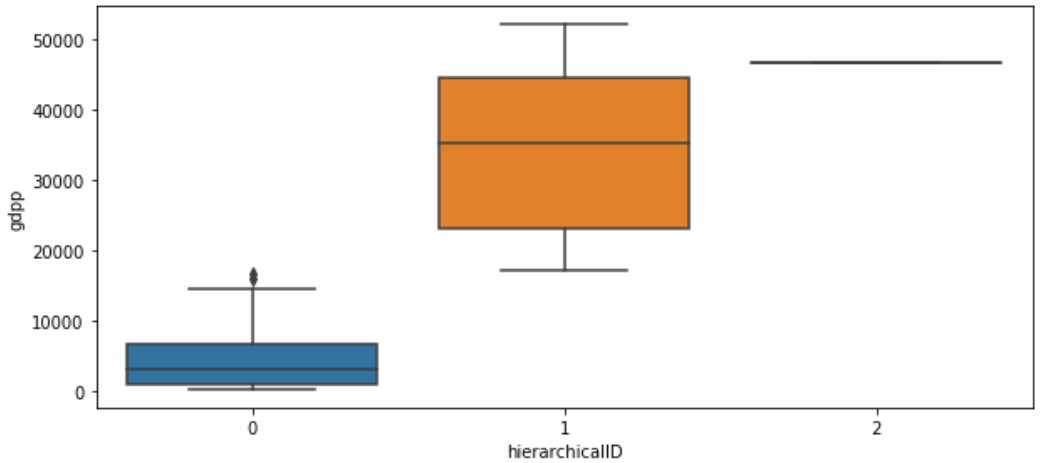


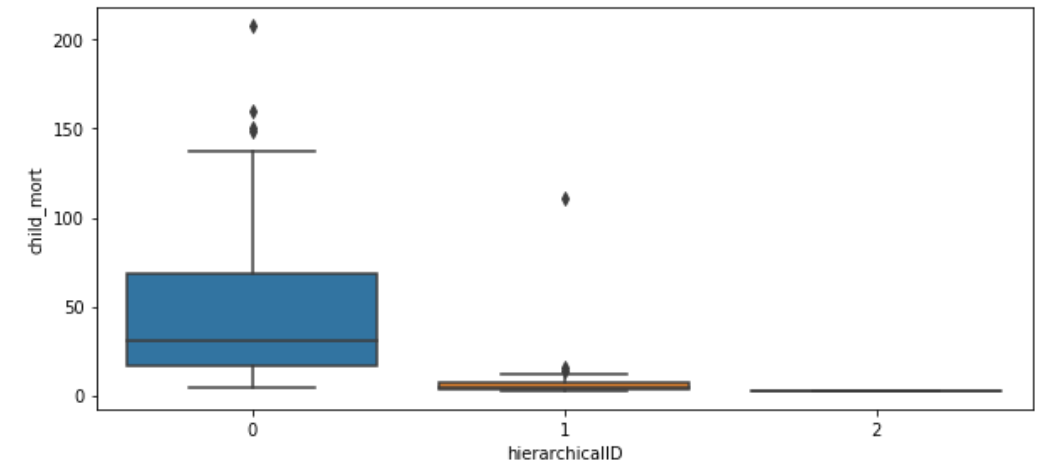
- Here I have plotted trade deficit against GDP. Trade deficit is total_imports - total_exports.
- A negative trade deficit means a country exports more than it imports, and vice versa. Hence, a low trade deficit is perceived to be better.
- There are two countries with extremely low trade deficits and good GDPs. They are Luxembourg and Qatar.
- The countries having lowest GDP along with high trade deficits are- Bahamas, Greece, Cyprus, Lebanon and St. Vincent and the Grenadines.



- Here I have plotted three variables- child mortality, income and GDP, of the three resultant clusters for comparison.
- The values of child mortality variables are so less in comparison to the other variables that the bar is barely visible on the plot. We can see, however, that the cluster number 0 has the highest child mortality since it's bar is visible.
- The cluster 0 is also the one containing countries with the lowest income and GDP. Thus, we can deduce that this is the cluster in which we will find the countries with urgent need of aid.







- The adjoining figure too presents a comparison of the different clusters of our data, using the hierarchical clustering method.
- The hierarchical clustering method builds a model from the top down, it will consider all the data points as one cluster, then go on dividing it into more and more clusters till every point becomes a separate cluster.
- From the graphs we can see that the cluster 0 is the one with lowest income, lowest GDP and highest child mortality.
- Thus, cluster 0 is the one which holds our countries in dire need of aid here too.
- There were quite a few countries that were identified in need of aid. Five of the most needy ones are listed in the next slide.

Analysis Results

On some further checking, the five countries that have been identified in most urgent need of aid are:

1. Haiti
2. Burundi
3. Seirra Leone
4. Liberia
5. Chad

Thank You