

## Reflection

Ian Kollipara

CS 392 Seminar in Data Science

## Introduction

This dataset is collection of countries, with certain statistics to help indicate status in terms of socioeconomic standing.

## Data and Methods

### Data Description

Name	Type	Description	Measurement
country	object	The name of the country	Nominal
child_mort	float	The death rate for children < 5 per 1000 live births	Ratio
exports	float	Percentage of the GDPP for exports of goods and services	Ratio
health	float	Percentage of the GDPP spent on healthcare	Ratio
imports	float	Percentage of the GDPP for imports of goods and services	Ratio
income	integer	Net Income per person	Ratio
inflation	float	Annual growth rate of total GDPP	Ratio
life_expec	float	Expected years that a person should live given current mortality patterns	Ratio
total_fer	float	Expected number of children born given current fertility patterns	Ratio
gdpp	integer	Nominal GDP per Capita	Ratio

### Relationships

Many of the columns had relationships, but the most common relationship was with GDPP. After doing some research (noted in the notebook), I found that it was the Nominal GDPP, as opposed to the PPP GDPP. This is an important distinction, as Nominal GDPP is a good estimate of a Country's Standard of Living. The GDPP is measured in US Dollars (\$).

Child Mortality also holds plenty of relationships (also noted in the notebook). My fiancée noted that this relationship is a noted phenomena, and quite indicative of a country's success.

These two relationships I consider the foundational ones of the whole dataset, as they help define all the other columns of data.

## Quality Issues

I didn't encounter any outright issues, but I did some annoyances in the data. The percentage columns were hard to work with in connection to other data, so I converted them to their monetary value via this formula

$$\frac{X_{\text{percent column}}}{100} X_{gdp}$$

This made the correlations (see notebook) far more apparent.

## Methods

I chose to use only 3 columns, two of which are the aforementioned foundational columns. The other was income. I chose to add income, as its a good metric to measure the individual citizen in comparison to the whole country.

The 2 clustering algorithms I chose were **K-Means** and **Agglomerative**. In the case of **K-Means** the algorithm is iterative, putting all values into cluster based on a centroid.

In the case of **Agglomerative** the clustering is hierarchical, where every data object is clustered together - over the course of time - into one large cluster. From there you can pinpoint the ideal clustering in its history. This history is represented in a dendrogram.

For the distance I chose to use Euclidean. There are other options, but I wasn't convinced there was a use case for another distance. Much of this data can be accurately summed with Euclidean, so I did not see a reason to use another metric.

## Results

### Clustering Results

My clustering results yielded clusters that represented socioeconomic groups. This makes sense based off my metrics. The saddest part of this, noted in the notebook, was the fact that many were African countries.

Overall, the results were clustered in ways that made complete sense when thinking about the dataset.

### Clustering Algorithms

**K-Means** provided more groups than **Agglomerative**, but lower groups of both overlapped heavily. The end result being that both gave similar countries (this is noted more in the notebook).

In this instance the choice didn't have too much of an effect on the end conclusion, but that's part of the use of having 2 different algorithms. With both yielding similar results, the groups are really strong cluster groups.

## Recommendations

This section is detailed at the end of the notebook. I have copy pasted it here for convince.

---

Based on all the above analysis, I recommend 3 countries in, and to a lesser extent the continent of, Africa.

### Central African Republic



Figure 1: (“Central African Republic,” 2022)

CAR is landlocked, with an ongoing civil that started over a decade ago in 2012. The country has faced severe strife, including numerous human rights abuses, “The civil war perpetuated the country’s poor human rights record: it was characterized by widespread and increasing abuses by various participating armed groups, such as arbitrary imprisonment, torture, and restrictions on

freedom of the press and freedom of movement” (“Central African Republic,” 2022).

This country is in need of major aid, much beyond financially. It first needs some stability before we can truly help it. However, if that were to occur, funding ways to capitalize on its massive natural resources could catapult this country to a powerhouse of the continent.

### **Democratic Republic of the Congo**



Figure 2: (“Democratic Republic of the Congo,” 2022)

DRC is another landlocked African country, located directly south of CAR it shares many of the same issues. DRC has been home to ongoing conflict on its eastern half for quite a long time. The large country, with the world’s 11<sup>th</sup> in population. > “The Democratic Republic of the Congo is extremely rich in natural resources but has suffered from political instability, a lack of infrastructure, corruption, and centuries of both commercial and colonial extraction and

exploitation with little widespread development. Besides the capital Kinshasa, the two next largest cities, Lubumbashi and Mbuji-Mayi, are both mining communities. The DRC's largest export is raw minerals, with China accepting over 50% of its exports in 2019. In 2019, DR Congo's level of human development was ranked 175th out of 189 countries by the Human Development Index. As of 2018, around 600,000 Congolese have fled to neighbouring countries from conflicts in the centre and east of the DRC. Two million children risk starvation, and the fighting has displaced 4.5 million people. ("Democratic Republic of the Congo," 2022)

Much of my concerns with immediately aiding the CAR are present here in the DRC. I would look for stability, then provide strong financial aid. With a large population, may resources, and a location that can breed strength, the DRC is a vital country for the success of the continent. But for that to be so, we must find that stability.

## Niger

Niger is another landlocked country, but doesn't share the same abundance of resources that the previous two have. The biggest struggle is the Sahara, which covers most of the eastern half of the country. This limits the ability to grow food and thrive, and with the growth of desertification, it becomes even more dangerous. Niger also faces other issues, with low education rates and a rural community. The population is predominately Mulsim, making aid from the west potentially difficult.

Out of all the countries, I believe Niger is in the most need. Its is relatively stable, but lacks in ways to build a strong populus. The need for education can be solved through funding. Funding can also lead to new ways of farming, or potentially reclaiming the desert. Niger is, by far, the country with the most immediate potential.

## Resources

Central African Republic. (2022). In Wikipedia. [https://en.wikipedia.org/w/index.php?title=Central\\_African\\_Republic](https://en.wikipedia.org/w/index.php?title=Central_African_Republic)

Democratic Republic of the Congo. (2022). In Wikipedia. [https://en.wikipedia.org/w/index.php?title=Democratic\\_Republic\\_of\\_the\\_Congo](https://en.wikipedia.org/w/index.php?title=Democratic_Republic_of_the_Congo)

Implementing Agglomerative Clustering using Sklearn. (2019, June 7). GeeksforGeeks. <https://www.geeksforgeeks.org/implementing-agglomerative-clustering-using-sklearn/>

List of countries by GDP (nominal) per capita. (2022). In Wikipedia. [https://en.wikipedia.org/w/index.php?title=List\\_of\\_countries\\_by\\_GDP\\_\(nominal\)\\_per\\_capita&oldid=1071744444](https://en.wikipedia.org/w/index.php?title=List_of_countries_by_GDP_(nominal)_per_capita&oldid=1071744444)

Niger. (2022). In Wikipedia. <https://en.wikipedia.org/w/index.php?title=Niger&oldid=1070263403>



Figure 3: (“Niger,” 2022