

Clustering Based on Healthcare System Grade

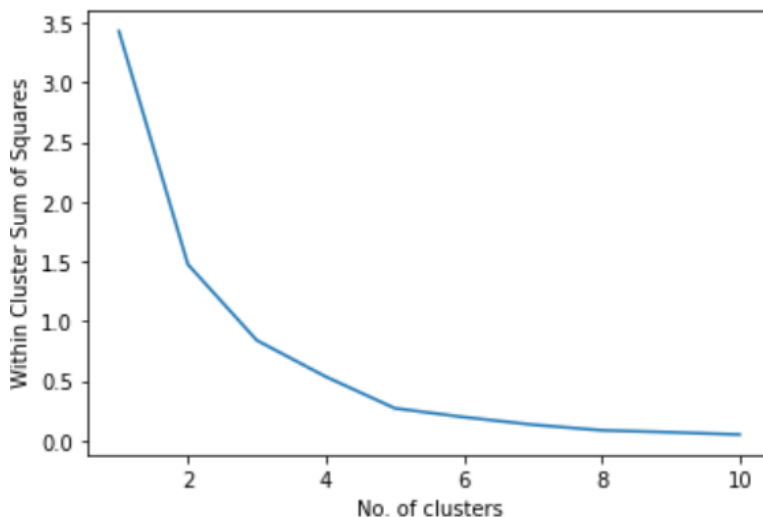
One of the most important field in this world is healthcare. Healthcare has a significant impact on how the country operates. If the country itself can't handle and accommodate its citizen leading into good healthcare system and healthy lifestyle, It cannot sustain and at some point, the country will collapse and people start dying. In this day and age, technology have improved in a fast pace than before and healthcare is also improved.

One of the example we can use to test each country healthcare is COVID-19 pandemic from early 2020. People dying, people couldn't act accordingly and limping through their daily life because layoffs, regulation, panic, and fear surrounded them. A lot of people also didn't believe COVID-19 is real and they tended to ease one of the most important method to prevent COVID-19 spreading is using a mask in public places. Hence, people started dying in the hospital because they got infected. Hospital started to fail, people that need medication had to wait in front of the hospital because the hospital couldn't accommodate for each patient, health workers got infected, and medications were scarce and expensive since supply was low and demand was high.

Analysis for a country to improve its healthcare systems is to see their positions worldwide. Gathering data of COVID-19 for each country and data science method to get insights. Data that are collected are population density, population density, people using basic sanitation services, the prevalence of malnourishment, current health expenditure per capita, access to healthcare, physician density, the population in urban areas, maternal mortality rate, number of COVID-19 tests, confirmed cases, deaths, recovered cases, and active cases, etc. Cleaning data also needed to get accurate representation. These datasets are from WHO, Harvard, John Hopkins, etc.

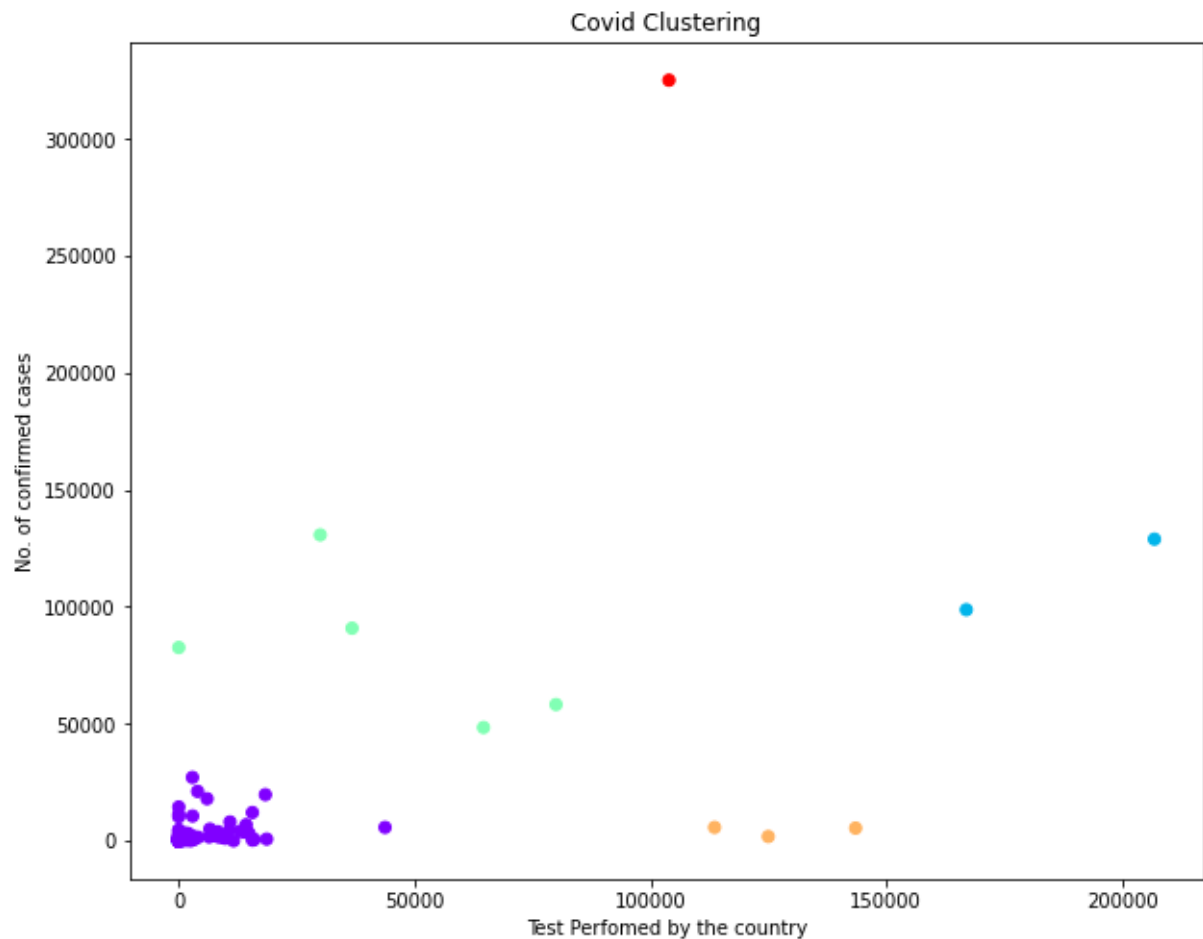
Method that is used to classify each country is clustering. One of the algorithms of clustering is K-Means clustering. K-means algorithm is an unsupervised machine learning algorithm that make clustering through a similar data and iteratively improve their clustering. K-means clustering algorithm doesn't need a labelled dataset and it is easier to implement to a numerical dataset but not so much as a Boolean dataset.

The only thing we need to determine is number of cluster of datasets. Elbow method can be used to determine number of cluster. From this graph, we can see x-axis as number of



clusters, y-axis are SSE of distance of each plot to centroid of the cluster. We can conclude that an optimal number of cluster is the "elbow" of the graph, which is 5.

From this algorithm, we get countries with similar characters within cluster. There are 5 types of cluster. From these cluster, we have certain characteristics which we can



describe in this table.

Cluster	Characteristics
1	<ul style="list-style-type: none"> • Low confirmed cases • Low number of tests performed • Least deaths • High population density • Malnourishment and low sanitation
2	<ul style="list-style-type: none"> • Highest confirmed cases • Highest number of tests performed • Highest number of deaths • Better healthcare system • Highest population density
3	<ul style="list-style-type: none"> • Relatively high confirmed cases • Relatively high number of tests performed • Relatively high number of deaths • Better healthcare system • Relatively high population density

4	<ul style="list-style-type: none"> • High number of tests performed • Least number of confirmed cases • Lowest number of deaths • Low population density • Better healthcare system
5	<ul style="list-style-type: none"> • High number of tests performed • High number of confirmed cases • Low population density • High number of deaths • Better healthcare system

From this cluster, we can conclude that healthcare system does help to lower number of deaths. Also, better to do test to verify cases for COVID-19 to prevent and mitigate deaths. Population density also a good indicator for predicting number of cases. In these datasets, cluster 1 shows that they're in phase 1 where COVID-19 hasn't been spreading for a good amount of time and healthcare system hasn't started to begin tracing tests even though population density is pretty high.

Deployment for this case is articles, website informing which country you are in, an updated datasets from and redo calculation to get a newest insight. We can do this method for other disease, not only COVID-19. One of them is heart attack cases. We can identify which country is prone to have its people get a heart attack based on certain parameters such as food, culture, lifestyle, and a lot more.

Github Repository

[EvanZJ/covid-19-k-means \(github.com\)](https://github.com/EvanZJ/covid-19-k-means)