**Clustering Philippine Cities**  
Applied Data Science Capstone  
by D. Fenix

**Introduction/Business Problem**

The year 2020 will be remembered as the year when the COVID-19 pandemic changed the world. Its impact covers all levels and aspects of society. As of December 6, 2020, the World Health Organization reports that over 65 million people have contracted the virus and over 1.5 million have died. Governments continue to impose restrictions on human activity in order to limit the spread of the virus and they are looking for ways to address its economic and societal impact.

On May 6, 2020, Philippine President Rodrigo Duterte issued Executive Order No. 114, instituting the *Balik Probinsya* program. This goal of program is to “develop the quality of life in the rural areas, in effect decongesting the densely populated areas of the country such as Metro Manila by encouraging people to move to the countryside once COVID-19-related quarantine measures are lifted.” (<https://en.wikipedia.org/wiki/Balik_Probinsya>)

While most migrants to the National Capital Region (NCR) of the Philippines might consider going back to their home cities or provinces as a first choice, a presentation of other options could be invaluable. This is where a clustering exercise of Philippine cities can help.

The audience that can benefit from this project are:

* Metro Manila residents who want to identify which rural areas/provinces are potentially good options for relocation.
* Provincial governments that can use this as a reference for what establishments they can promote and incentivize to make locations more attractive for relocation.

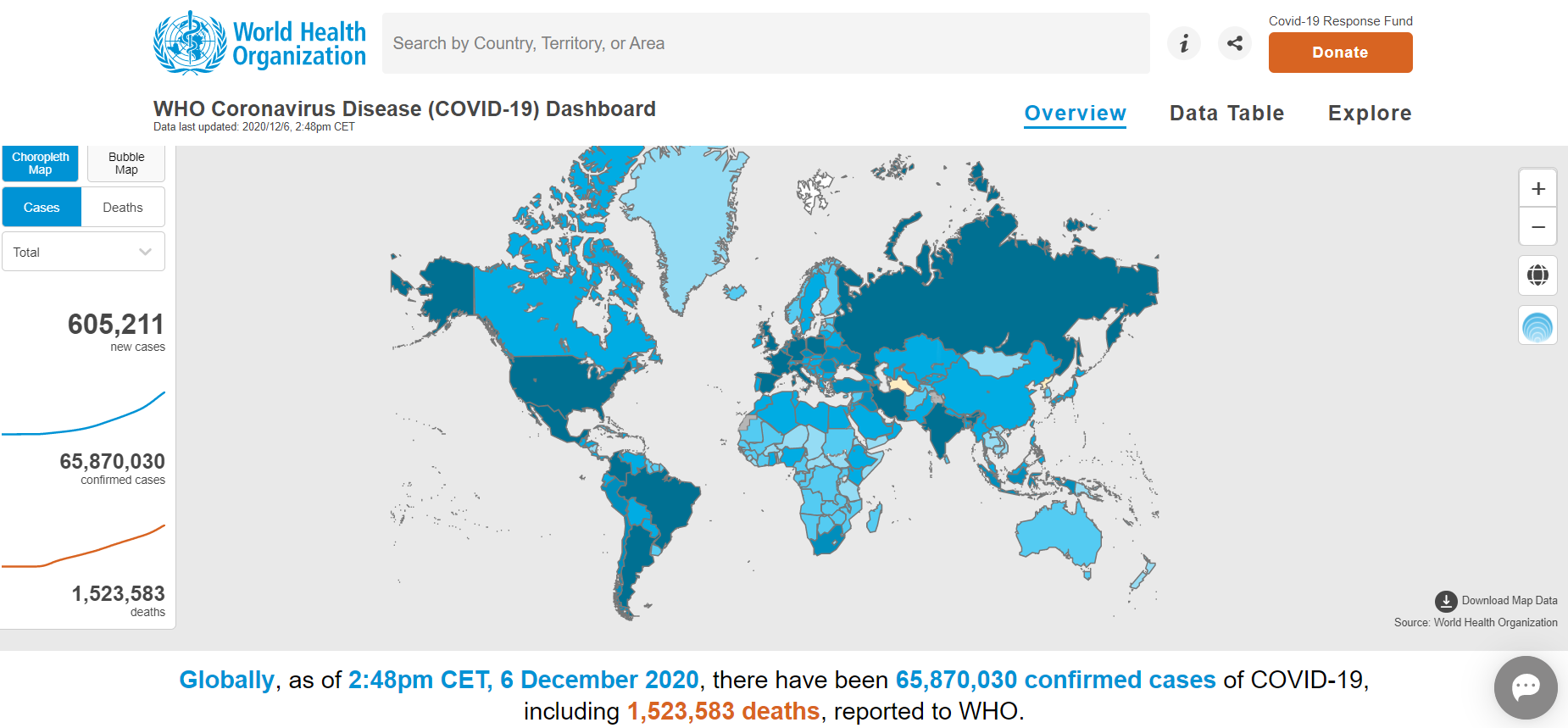
Being originally from the Philippines, this project is personally interesting to me as I want to see which areas are similar to my home town in Iligan City.

**Data**

The following data will be used in clustering Philippine cities.

* **List of Philippine cities** from <https://en.wikipedia.org/wiki/List_of_cities_in_the_Philippines>. This data will be read into Python using pandas. This is the list that will be clustered based on common venues.
* **Forsquare API** will be used to get the most common venues of each city. From doing data exploration of the Foursquare data, it might be necessary to group the venue types into larger buckets. For example, venue types like American Restaurant and Japanese Restaurant can be grouped into Restaurant.
* **Google Maps** will be used to get the coordinates of each city. The longitude and latitude are inputs required in the Foursquare API.

**References**

1. Coronavirus cases and deaths from WHO Coronavirus Disease (COVID-19) Dashboard (<https://covid19.who.int/>) 
2. List of Philippine cities snapshot from Wikipedia (<https://en.wikipedia.org/wiki/List_of_cities_in_the_Philippines>) 