CLUSTERING CITIES IN THE PHILIPPINES

Applied Data Science Capstone Project

D. Fenix | December 2020

The Philippines amidst a global pandemic

- The year 2020 will be remembered as the year COVID-19 changed the world. As of mid-December, the World Health Organization reported over 71 million cases and over 1.6 million deaths.
- In the Philippines, almost have a million cases have been reported with over 8,000 deaths. Managing the pandemic is a challenge especially in densely populated areas like Metro Manila.
- In May, the Philippine government instituted *Balik Probinsya* program which aims to develop quality of life in rural areas and encourage people from Metro Manila to relocate to the countryside.

Relocation from urban to rural areas

- Quality of life is a key consideration for relocation. Infrastructure, culture and leisure, housing, education are some indicators of quality of life and these are largely determined by the types of venues found in an area.
- Clustering Philippine cities according to most common venues will be done to address the following questions:
  - For an urban area resident considering relocation, which provinces are most similar to the current cities where they live? This will help with transition to the new location.
  - For local governments, what features might potential relocatees look for? This will help decide the types of establishments to prioritize or incentivize for future development.

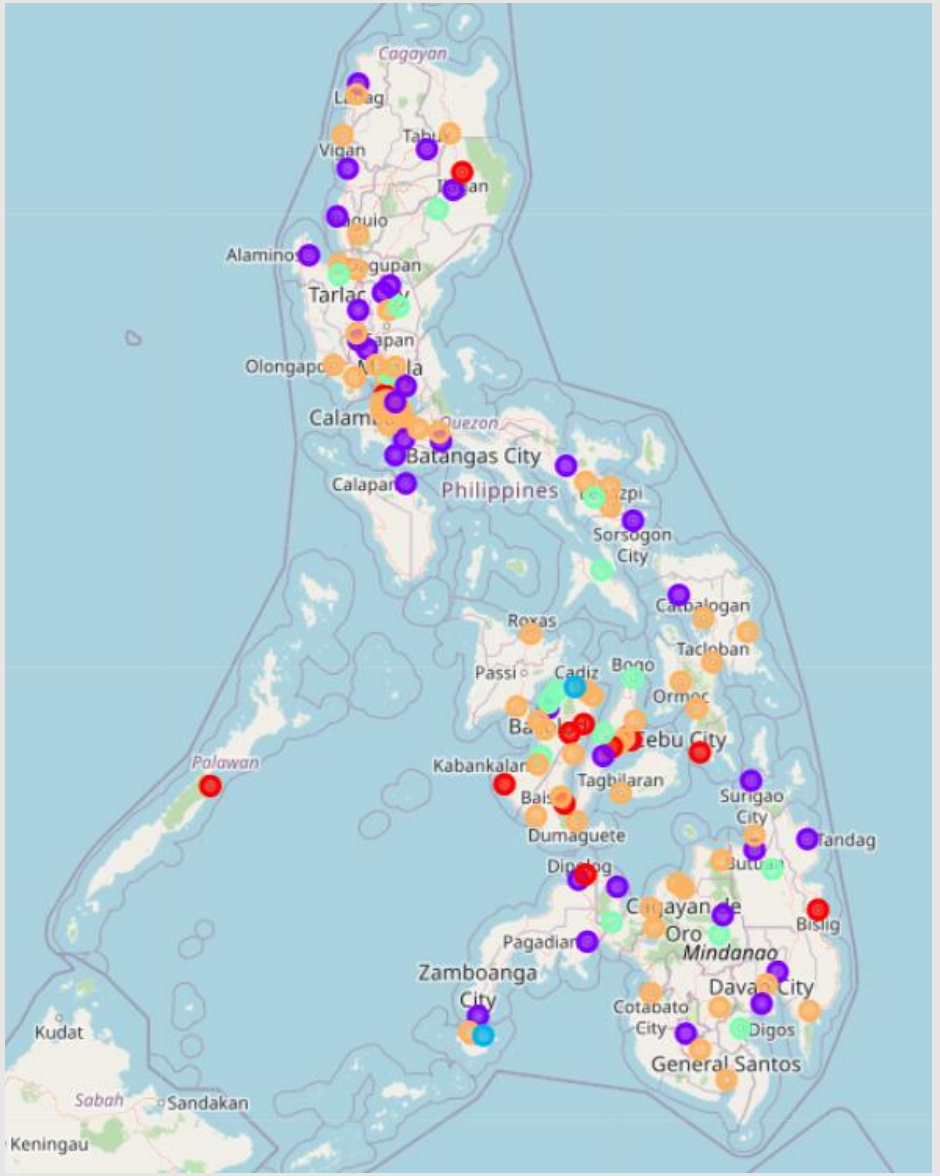
Data and Methodology

- The **list of Philippine cities** is obtained from Wikipedia and converted into a pandas dataframe. There are 146 cities in all, 16 of which are in Metro Manila. The coordinates were available from the Wikipedia backend table.
- **Foursquare API** is used to pull the most common venues from each city as determined by coordinates. A total of 2207 venues were pulled, with 203 unique venue categories. These were aggregated into 18 larger categories.
- The top 10 most common venue categories were determined for each city and **K-means clustering** was done to group similar cities.

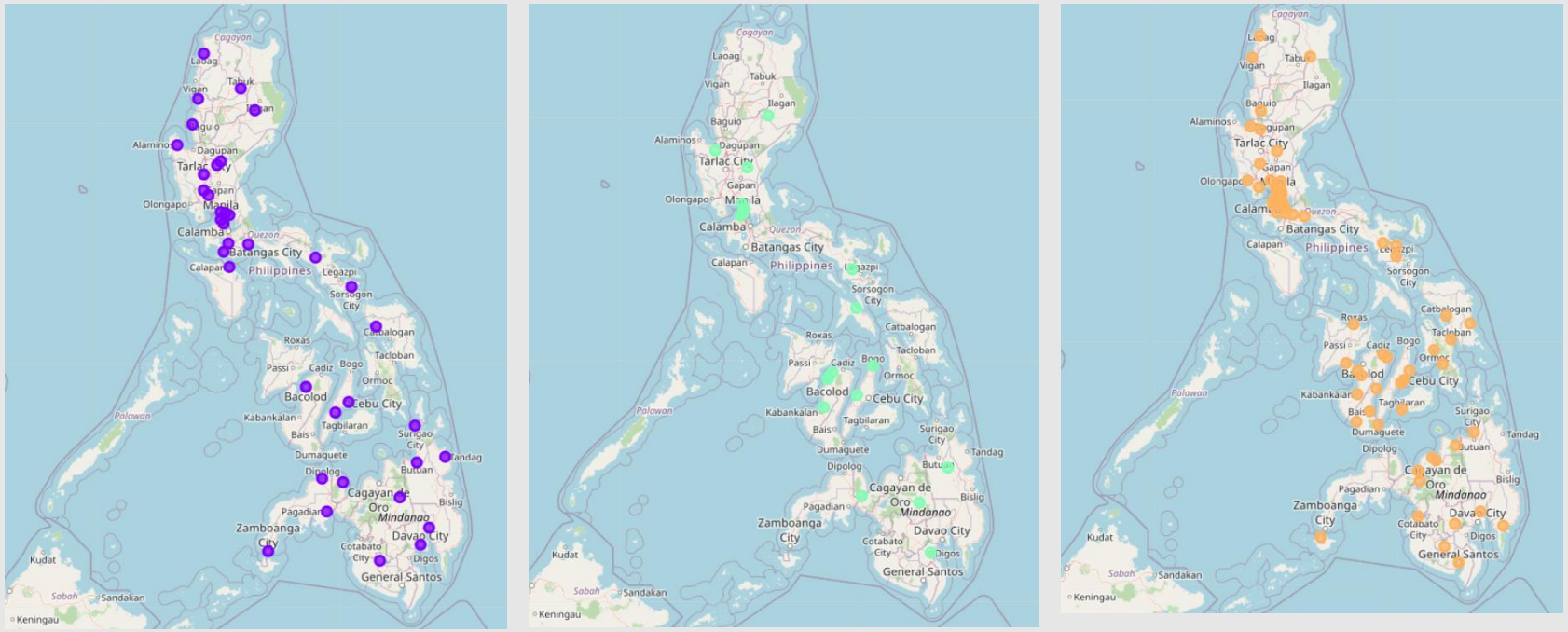
Results and Discussion

From the list of 146 cities, five were not assigned clusters due to missing data from Foursquare. The 141 cities are grouped into 5 clusters. Only 3 clusters had Metro Manila cities.

Cluster	Total Number of Cities	Metro Manila Cities
1	12	0
2	37	4
3	2	0
4	19	2
5	71	10
Total	141	16

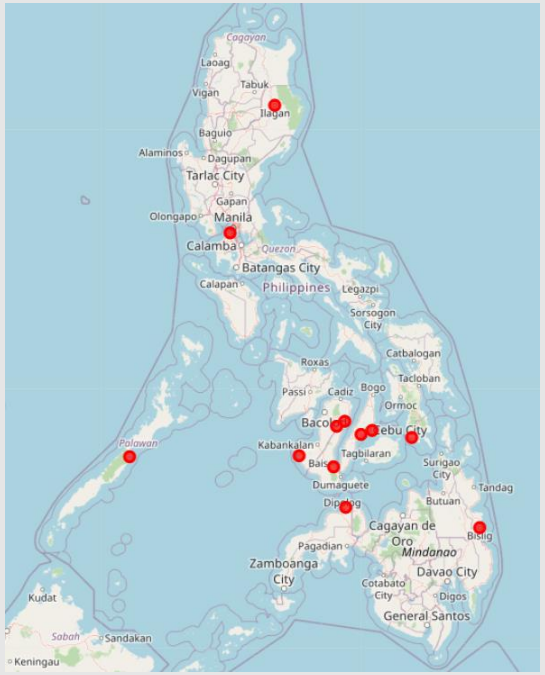


Clusters with Metro Manila cities



Clusters 2, 4 and 5 contain Metro Manila cities. For a Metro Manila resident, rural cities in the same can be considered possible options. The cities are also distributed throughout the country which provides flexibility for other criteria to be considered such as proximity to family, employment, et cetera.

Clusters with no Metro Manila cities



Clusters 1 and 3 contain no Metro Manila cities and represent a total of 14 rural cities. While these places might not be considered similar to Metro Manila, the output is still useful for others who want to consider a lifestyle change and move away from big city environments.

Cluster 5 Venues

Cluster 5 is the largest cluster with 71 cities overall and 10 located in Metro Manila.

An examination of the cities in this cluster shows that the top venue categories are Restaurants, Athletics and Sports, Food Shops, Shopping and Transportation.

Governments in rural cities can use this as a guide for the kind of venues that relocatees would look for and thus prioritize for future development.

1st Most Common Venue	2nd Most Common Venue
Restaurant	Food shop
Restaurant	Athletics or Sports
Restaurant	Food shop
Restaurant	Food shop
Restaurant	Athletics or Sports

Conclusion and Recommendation

- Balik Probinsya encourages Philippine residents in densely populated urban areas to relocate to rural areas. Clustering Philippine cities can help Metro Manila residents identify similar rural cities for relocation. Common venues in clusters can help local governments prioritize the types of establishments to prioritize for future development.
- A possible extension of this project would be to further segment large clusters. Other criteria can also be added in clustering such as density or area.