Condominium Development Insights for Metro Manila

IBM Data Science Capstone Project: Battle of Neighborhoods

By: Karl Baliton

June 14, 2020

Table of Contents

List of Figures

I. Introduction

Metro Manila is the busiest area in the Philippines. Although it does not host the beautiful natural landscapes and beaches which Philippines is known for, this location gives a multicultural experience from lavish shopping malls, to historical parks, and to exotic street food, that reflects that Philippine culture. Filipinos from the different provinces and even foreigners flock to Metro Manila to find leisure, recreational activities, education, work and many more. With the constant influx of population, no wonder this area has been bustling with different business opportunities, and one of which is the real estate business. The demand for housing has been increasing every year and for many real estate developers, the question in mind is "where do we develop our next project?"

II. Problem and Objective

With the growing demand in housing in Metro Manila, there is a huge opportunity for real estate developers to capture existing and even build new markets. This case study will help recommend the most suitable type of condominiums to develop around Metro Manila. The establishment can be tailored fit to different market profiles and this study will include a description of each clustered neighborhood that will be identified with the machine learning technique that will be used.

III. Data

A. Data Sources

The following data were used to come up with the recommendation for the case study:

- Population data from the Philippine Statistics Authority (PSA) indicating the population per city in Metro Manila in 2000, 2010 and 2015 were used where the population growth rate shall be derived. In addition, 2010 population demographics by sex and age group were utilized to describe the population. (http://openstat.psa.gov.ph/)
- Latitudes and longitudes of each city were generated using the Geopy library.
- Foursquare API keys were utilized to call the establishments near the cities that have been identified.
- A map was generated through the use of Folium library to visualize the clustered locations.
- Clustered locations was described according to venue categories as provided by the Foursquare API.

B. Data Pre-Processing

The population data was directly downloaded from the PSA online portal which was pre-processed in excel to simplify the format of the dataset. The excel file was read to a dataframe using Pandas library. With the limited choice of dataset, 2010 census on the population was utilized for the interest of the study. This is considered a limitation and should be updated in future studies.

The data frame generated had 37 features that was composed of eighteen age groups, one set for each gender. The first feature contained the list of cities in Metro Manila. The columns were renamed in preparation of the data exploration. Since the dataset has separate data for male and female, the dataset was split into two data frames. The total population for each age group was generated using the sum function.

To aid in retrieving data through Foursquare API, the same population dataset was used where a new data frame was generated using the city column. Geographical coordinates were determined for each city using the Geopy library.

IV. Methodology

This study will identify the most suitable types of condominiums to develop in the different cities in Metro Manila.

First, Metro Manila population was describe using the dataset from the Philippine Statistics Authority (PSA). Visualization techniques were leveraged to understand the demographics. An article by Lamudi was used as reference to support the discussion. This will explore the current market situation and sentiments of real estate consumers.

Second, geographical coordinate of each city will be generated using Geopy library. Foursquare API will be utilized to identifiy the venues around the cities.

Third and last, the cities in Metro Manila was mapped out using Folium library. These were clustered based on the most common venue categories around them using k-Means clustering algorithm. By doing this, the relationship between the different cities can be identified where conclusions can be drawn up as recommendation to condominium developers.

V. Data Analysis

A. Descriptive Analysis

Before the data from the Philippine Statistics Authority is processed, a related literature is discussed regarding the behavior of condominium consumers.

Lamudi is an online real estate classified website focusing exclusively on the emerging markets. They published an online article on their analysis on customer behavior who were using their website. The findings suggests that the top 3 age groups that uses their platform are 25-34, 35-44, and 45-54 with share views of condominiums for sale at 37.04%, 22.64%, and 15.21%, respectively. In addition, they found out that 55% of views are comprised of femailes. We can look in the the population by age group and sex and see the potential market for condominiums.

In another online article, they were able to conclude that location is the best indicator of the value of the condominium regardless of the age of the customer and regardless of the city where the establishment is located.

With the insights from Lamudi's analysis, the demographics of the population by age group plays a role in determining the market potential of condominiums in Metro Manila.

Sources:

Millennials Fueling Surge in Condominium Development in the Philippines, Lamudi 2018
(https://www.lamudi.com.ph/journal/millennials-fueling-surge-in-condominium-development-in-the-philippines/)

The Outlook 2018's Property Seeker Survey: Revealing the Pulse of the Filipino Property Seeker, Lamudi 2018
(https://www.lamudi.com.ph/journal/the-outlook-2018s-property-seeker-survey-revealing-the-pulse-of-the-filipino-property-seeker/)

A bar graph was generated with the use of Matplotlib library from the pre-processed population demographics dataset.

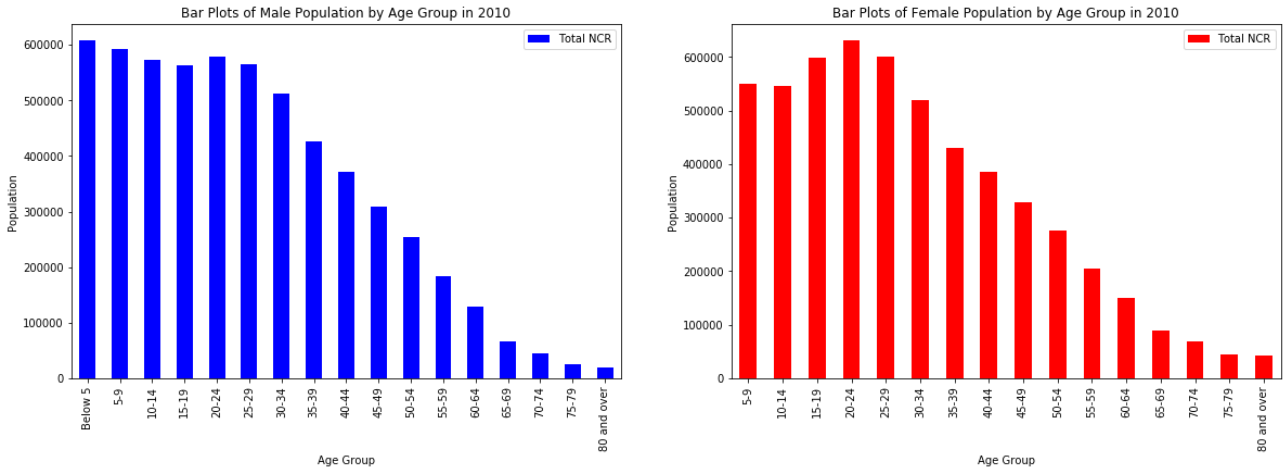

Figure 1. Bar graph of the population in Metro Manila by age group. Blue graphs represents the male population while the red graphs represents the female population.

Both of the bar plots for male and female population shows a similar trend wherein the whole population is majorly comprised of younger age groups. As the findings of Lamudi seemingly suggest, the older the age group, the less likely that they will consider purchasing condominium units. This can signify that there is a very good opportunity for developers to dwell on this business.

B. Foursquare API

Using the pre-processed dataset with the geographical coordinates of each city, Foursquare API keys were called to collect information on the nearest venues. One hot coding was done to convert the venue categories to a Boolean form and frequency for the categories per city was calculated and sorted from the most common to the least common in the area.

C. k-Means Clustering Algorithm

The dataset with the frequency of each category per city was fit into k-Means clustering algorithm with four (4) clusters and then data frame was updated with the cluster labels. Resulting dataset was mapped using the Folium library.

VI. Results and Discussion

Below are the maps generated using Folium library.

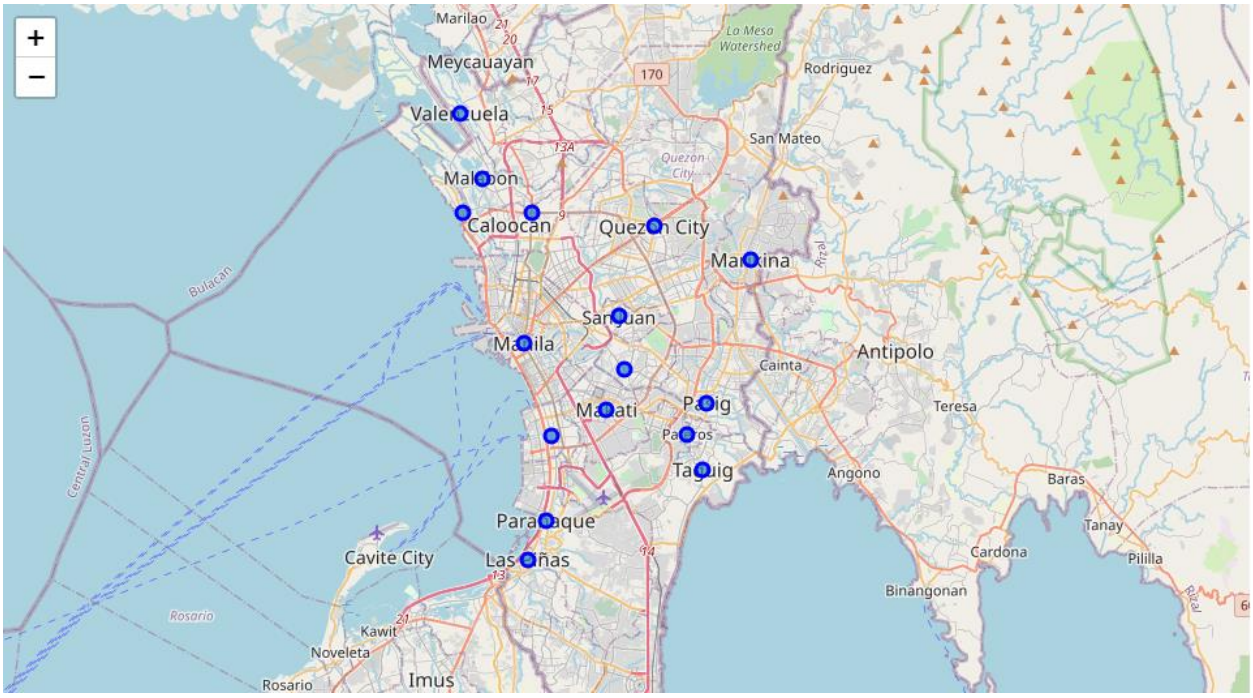

Figure 2. Cities in Metro Manila.

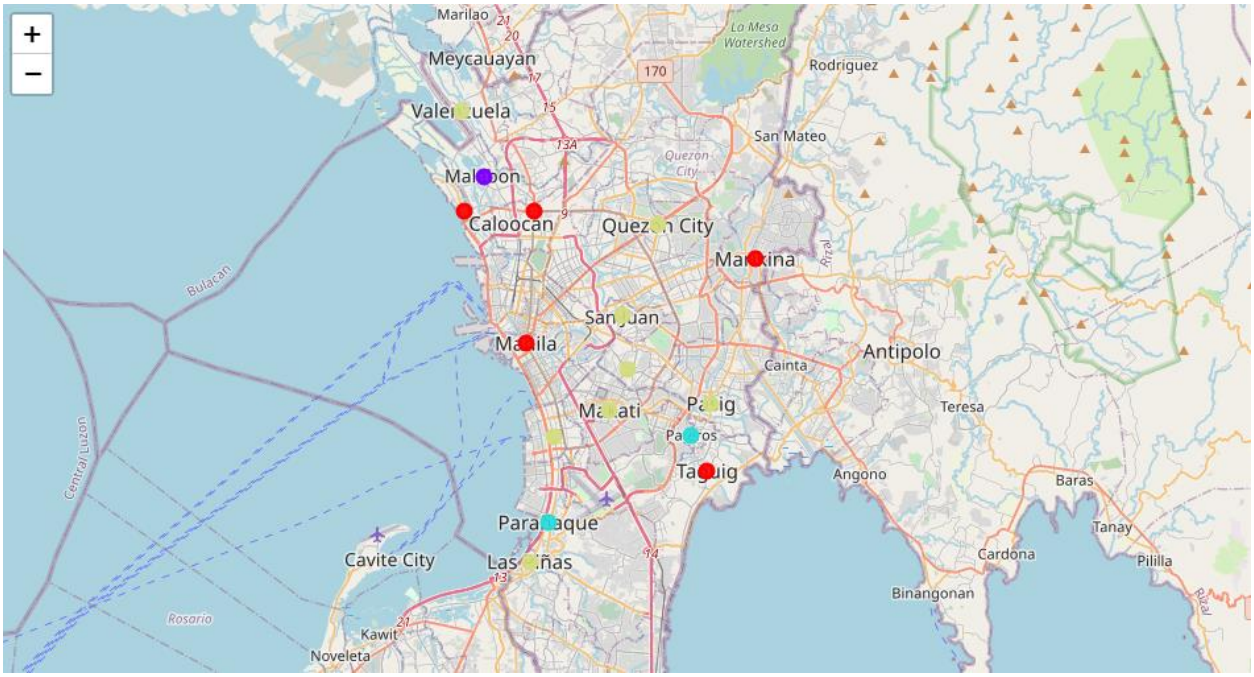

Figure 3. Clustered Cities in Metro Manila.

Below are the resulting clusters from k-Means algorithm.

Cluster 1

	City	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Manila	0	Fast Food Restaurant	Filipino Restaurant	Pizza Place	Café	Coffee Shop
2	Marikina	0	Fast Food Restaurant	BBQ Joint	Pizza Place	Chinese Restaurant	Fried Chicken Joint
6	Caloocan City	0	Fast Food Restaurant	Chinese Restaurant	Bookstore	Japanese Restaurant	Café
8	Navotas	0	Fast Food Restaurant	Market	Restaurant	Café	Gym
12	Muntinlupa	0	Fast Food Restaurant	Pizza Place	Bakery	Coffee Shop	Sandwich Place
16	Taguig City	0	Fast Food Restaurant	Fried Chicken Joint	Bakery	Shopping Mall	Steakhouse

Figure 4. Cluster 1 composed of Manila, Marikina, Caloocan, Navotas, Muntinlupa and Taguig Cities.

Cities in Cluster 1 have a lot of surrounding fast food restaurants. The volume of casual diners around the area probably causes this. It can be attributed to these places being major residential areas where daily food consumption demand is in constant high. Congestion around these areas is very probably. This area could be very attractive to people with fast-paced activities in the metro.

Cluster 2

	City	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
7	Malabon	1	Park	Wine Bar	Creperie	Dance Studio	Department Store

Figure 5. Cluster 2 composed of Malabon City.

This cluster is only composed of one city that is Malabon. In contrast to cluster 1, the city hosts venues that caters leisure activities. This area is likely suitable to lain back population where relaxing time is prioritized.

Cluster 3

	City	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
13	Paranaque	2	Convenience Store	Snack Place	Dessert Shop	Seafood Restaurant	Fast Food Restaurant
15	Pateros	2	Convenience Store	Fast Food Restaurant	Plaza	Tea Room	Diner

Figure 6. Cluster 3 composed of Paranaque and Pateros Cities.

Paranaque and Pateros compose this cluster where convenience stores and casual dining places are the most common venues. Essentials are easily accessible and these areas can definitely cater basic daily needs which is suitable for people who are always on-the-go

Cluster 4

	City	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Mandaluyong	3	Filipino Restaurant	Pizza Place	Gym / Fitness Center	Bakery	Convenience Store
3	Pasig	3	Convenience Store	Pizza Place	Café	Donut Shop	Chinese Restaurant
4	Quezon City	3	Filipino Restaurant	Cosmetics Shop	National Park	Bike Rental / Bike Share	Japanese Restaurant
5	San Juan	3	Convenience Store	Pet Store	Diner	Grocery Store	Basketball Stadium
9	Valenzuela	3	Food & Drink Shop	Chinese Restaurant	Pharmacy	Bar	Tea Room
10	Las Pinas	3	American Restaurant	BBQ Joint	Park	Bubble Tea Shop	Snack Place
11	Makati	3	Restaurant	Coffee Shop	Filipino Restaurant	Café	Clothing Store
14	Pasay City	3	Hotel	Breakfast Spot	Café	Art Gallery	Asian Restaurant

Figure 7. Cluster 4 composed of Mandaluyong, Pasig, Quezon, San Juan, Valenzuela, Las Pinas, Makati, and Pasay Cities.

The cities in this cluster is composed of various venues that cater to a broad array of needs from basic needs to leisure activities. It is very likely that these areas could become high-demand.

As seen in the map, cities in Cluster 4 are located in the center of Metro Manila. This somehow explains why the venues are very mixed on those areas as those are very strategic for businesses. There is better reach to consumers in these areas compared to the ones at the outer locations. In comparison, cities in Cluster 1 are located at the outer areas of Metro Manila. Though these are not in the center of the metropolitan, this suggests that there is still a lot of foot-traffic around edges.

These findings may help condominium developers in conceptualizing plans on where to do their next project. If the target market is for people who are always on the go such as students and young professionals, cities in Cluster 4 or Cluster 3 may be recommended with the completeness of the services around the area. If the idea is to have a budget-friendly condominium to target consumers who have limited spending behaviors, cities in Cluster 1 can be recommended as the neighboring establishments can very well compliment with the needs of the customers. Cluster 3 may be for customers who are leaning towards a laid-back neighborhood where they can utilize leisure establishments. Likely customers would be middle-aged and older population who wants to settle down.

VII. Conclusion and Recommendations

Recommendations on the type of condominium to be developed were established where Clusters 3 and 4 can leverage on the completeness of the different services around the area, Cluster 1 for budget-friendly customers, and Cluster 3 for customers seeking leisure activities.

To further the findings in this study, it is recommended to do another k-Means clustering algorithm for specific cities of interest. Generating a heat map for each city factoring the venue categories may give more insights of the areas. Population data should also be updated and land value or land price could be factored in the analysis.