

Profiling Taguig Philippines Neighborhood by Amenities Food, Recreation, Fitness, Market, Utility and Leisure

July 25, 2020

1 Introduction

Profiling of Boroughs is a big area of research in the real-estate, logistics or even delivery businesses. With this in mind a good and ready available dataset is a must for such research. Fortunately, through webscraping Datasets can be built using various web sources and can be transformed into a relatable dataset.

Philippines, being a country with Barangays inside a municipality rather than Boroughs, makes it difficult to make such profiles. Barangays are not profiled by its capital but by the number of households under it. A good area of research is to profile each barangay and treat it as a Borough and determine business establishments inside it. The idea is to build a dataset that will enable further studies in Maket Research, Capital Research or even Traffic Management by considering key areas and determining features around it.

Various Lifestyle Companies such as Forbes or Machadvisor, tells us that Amenities in a neighborhood attracts people and adds value to the area it is located in so this is a good way to have a profile around.

This study will focus on Taguig City, Philippines and the Barangays contained in this city. By the end of the study we will be able to:

- (1) Build a data set of Barangays inside Taguig with it's coordinates;
- (2) Get all nearby Venues inside Taguig using Foursquare API and Categorize it into Amenities; and
- (3) Get Optimal Neighborhoods and Profile each Neighborhood by the amenities it contains.

2 Methods

In this study we will use Python as the primary tool for Data Analysis

2.1 Data Collection

Data was built using Taguig Barangay Zip codes found in [1]. The coordinates of the barangays were retrieved using google maps by buteforce method. Foursquare API was used to get nearby venues along with their coresponding longitude & latitude and

venue category (see [2] for the full documentation of Dataset).

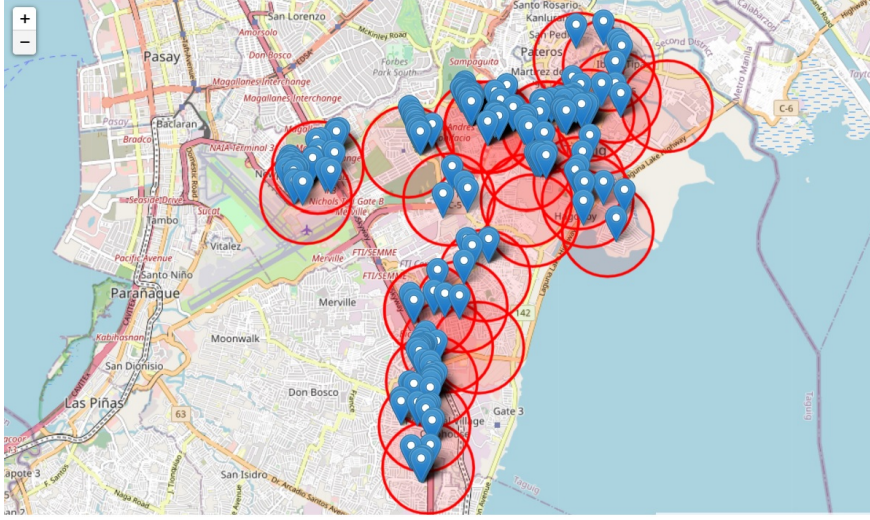


Figure 1. Taguig Folium Map with barangay coordinates marked using Folium.Circlemaker and Venue Coordinates using Folium.Marker

2.2 Data Cleaning

Data was categorized as Food, Recreation, Fitness, Utility and Leisure Amenities according to their venue category provided by Foursquare API. As observed, these venue categories have various inconsistencies such as M lhuillier categorized as a jewellery shop instead of pawnshop so Data Cleaning was required by first Categorizing Venues with accurate venue category then checking each uncategorized amenity. Furthermore, with Foursquare lacking Data on Transportation, Government Building, etc, some venues were uncategorized and dropped as well. See figure 1 for the first 10 rows of the Dataset and [2] for the full documentation.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Flag Food	Flag Recreation	Flag Fitness	Flag Market	Flag Utility	Flag Leisure
0	Bagumbayan; Pinagsama (including EP Village)	14.517618	121.048676	McDonald's	14.516306	121.051504	Fast Food Restaurant	1	0	0	0	0	0
1	Bagumbayan; Pinagsama (including EP Village)	14.517618	121.048676	Heritage Park Jogging Oval	14.519163	121.049195	Park	0	1	0	0	0	0
2	Bagumbayan; Pinagsama (including EP Village)	14.517618	121.048676	Philippine Veterans Museum	14.514209	121.047617	History Museum	0	0	0	0	0	1
3	Bagumbayan; Pinagsama (including EP Village)	14.517618	121.048676	HBC	14.515108	121.052174	Cosmetics Shop	0	0	0	0	0	1
4	Bay Breeze Executive Village	14.512020	121.078615	NAIA Canteen	14.509831	121.080218	Cafeteria	1	0	0	0	0	0
5	Bay Breeze Executive Village	14.512020	121.078615	Siomai House	14.514723	121.081629	Food Truck	1	0	0	0	0	0
6	Bay Breeze Executive Village	14.512020	121.078615	Phines Bakery	14.516211	121.077801	Bakery	1	0	0	0	0	0
7	Bay Breeze Executive Village	14.512020	121.078615	All Day Supermarket	14.512756	121.074039	Market	0	0	0	1	0	0
8	Maharlika Village, Taguig	14.498195	121.051167	Ayala Malls Arca South	14.496085	121.047928	Shopping Mall	0	0	0	0	0	1
9	Maharlika Village, Taguig	14.498195	121.051167	McDonald's	14.502004	121.051440	Fast Food Restaurant	1	0	0	0	0	0

Figure 2. First 10 Rows of Taguig Venues using foursquare API categorized into Amenities

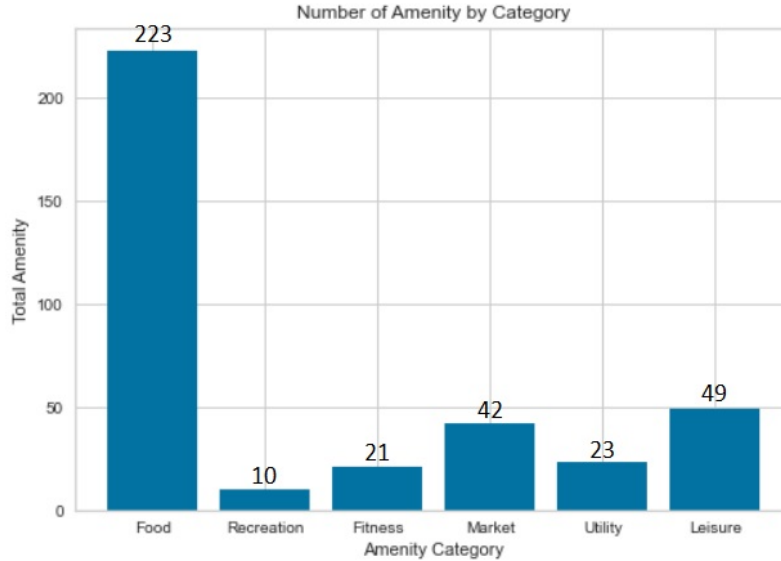
2.3 KMeans Algorithm & Elbow Method

Kmeans algorithm will be used for the clustering Venues in Taguig City since the main concern of this clustering will be the distance of each venues. This is the most commonly used algorithm in geospatial Data because it is concerned with minimizing Distances of the points to be considered together with the predetermined clusters.

Finding the optimal number of clusters in this study is tricky. This is because a lower number of cluster will result to a huge error/distances but increasing it will definitely minimize the error but a zero error/distance would mean no clusters at all. To overcome this problem, the elbow method will be used. Cluster analysis makes use of a heuristic called the Elbow method in determining the amount of distortion a model has with varying number of cluster [3].

2.4 Exploratory Analysis

Before segmenting These venues into clusters, the total amenities in taguig were first Determined. This will help in seeing if the number of clusters to be determined is realistic with respect to the optimal number of clusters to be determined using Kmeans. There are a total of 368 total of amenities in Taguig and figure 2 summarizes the number of Amenities in Taguig City per Category.



gory.

Figure 3. Total Amenities in Taguig per Category

In order for us to use determine the optimal number of clusters in Taguig, the yellowbrick.cluster package in python was used. Using this package together with the elbow method suffices as the way to determine the optimal number of clusters to be used in KMeans Algorithm since the Data being considered to clustered are longitude and latitudes which requires

no distribution analysis. From figure 3 we see that the optimal number of clusters is $K=3$, this is because it has the least propagated error determined from the running time and least overfitting from the elbow method.

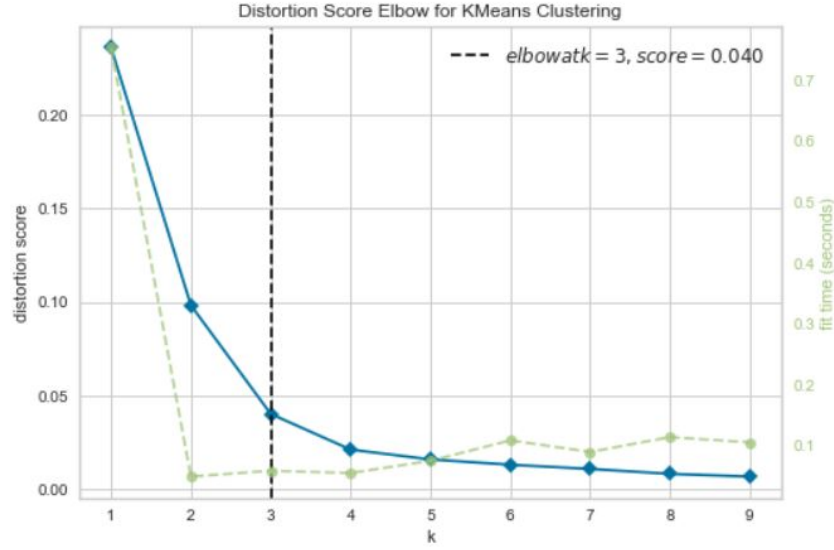


Figure 4. Line Plot of Variance and Operation Count of $K=n$, $n=0,1,2,...,9$

3 Results and Discussion

3.1 KMeans Algorithm with $K=3$

As seen in figure 5, the Kmeans algorithm estimated this 3 clusters properly.

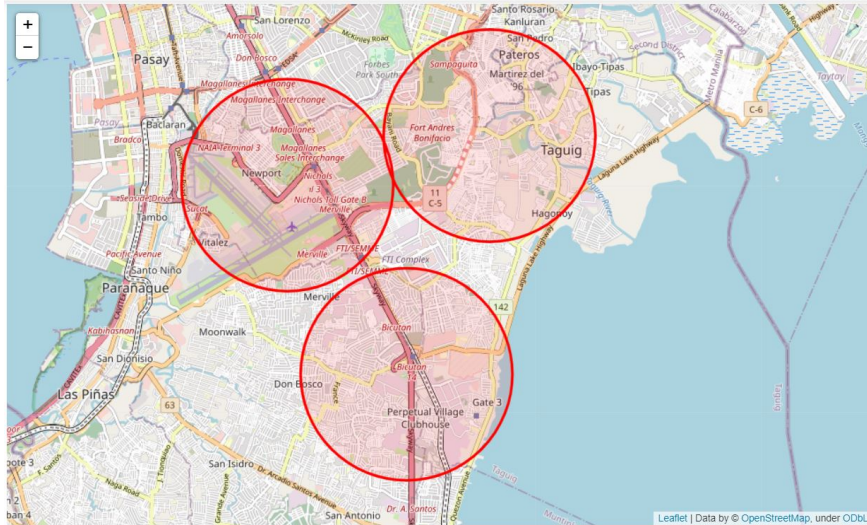


Figure 5. Taguig City Clustered determined from KMeans Algorithm with $K=3$ roughly visualized by folium.circlemaker

Amenities are then plotted into clusters with cluster 1,2,3 markers plotted with colors blue, green and purple respectively. Cluster 2 is expected to have the most number of venues being the cluster covering many barangays. Moreover, Cluster 1 and 3 are borders of Pasay City and Makati respectively. Results can visually be summarized by Figures 6-8.

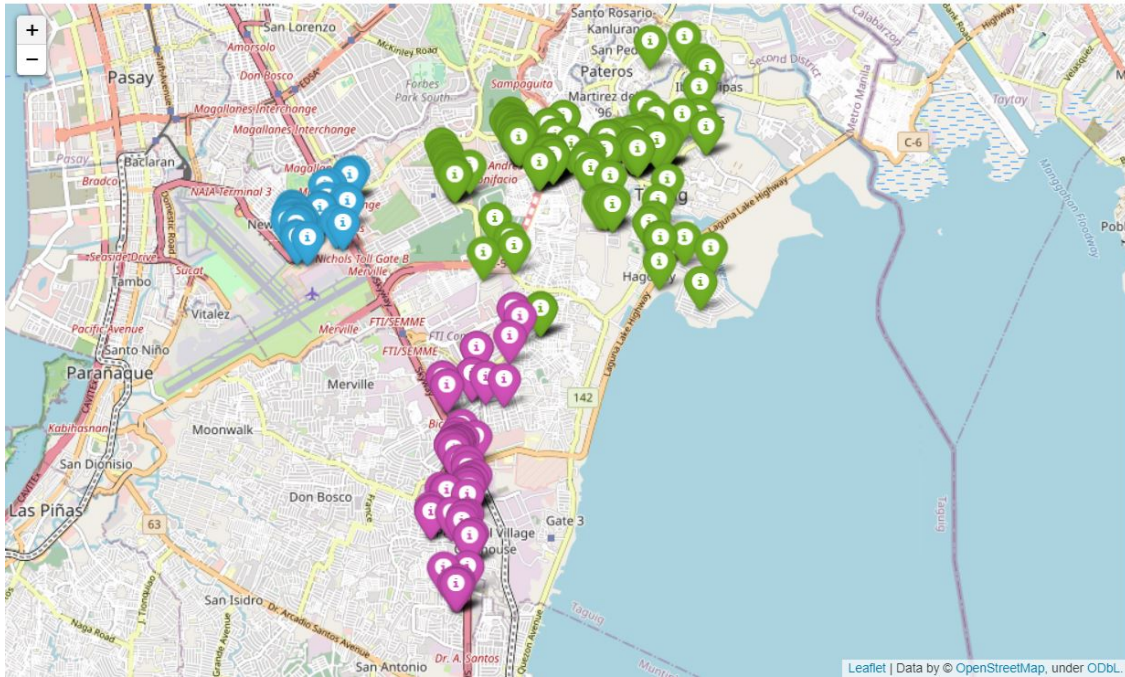


Figure 6. Taguig City Venues determined from KMeans Algorithm with K=3 roughly color visualized by folium.markers

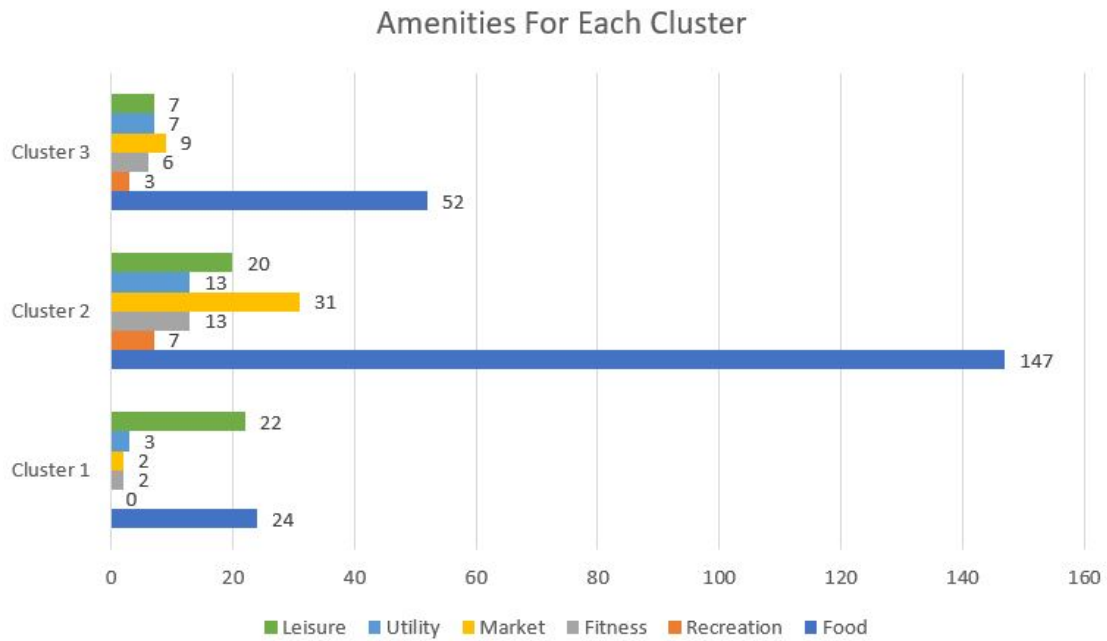


Figure 7. Number of Amenities per Cluster

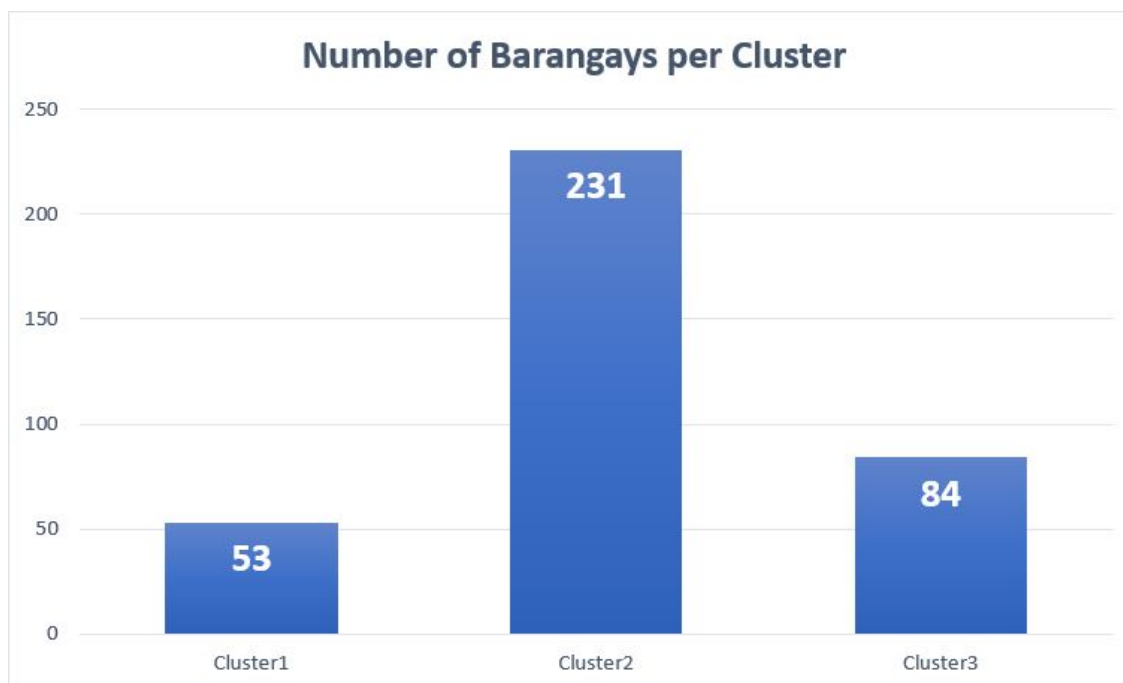


Figure 8. Number of Barangays per Cluster

3.2 Observations and Recommendations

Further Market research can be done through some observations of the results. For example, it can be observed that most of Businesses in Cluster 1 & 3 are focused only on the main roads unlike in Cluster 2 which is very spread. This means a bigger opportunity for small businesses like minigrocery stores can still be placed in the streets (fine roads as seen in the folium map) of cluster 1 & 3 than in Cluster 2. But on the contrary better GDP will result from a well spread establishments in Cluster thru a higher population target by the said spread.

This study can also be useful as profile when doing strategic government municipality planning and can still be further developed according to the government needs.

As for amenities they are profiled into 6 given amenities but by the weakness of foursquare data in the Philippines there is a demand for Foursquare API developers. Developers can still increase accuracy and number of Data in the Philippines. Furthermore, when doing machine learning it is quite useful to have these types of data ready when dealing with geospatial Data.

4 Conclusion

Using Foursquare and Google Data we have successfully profiled Taguig Barangays into Clusters. Moreover, Amenities in each cluster are categorized and counted.

5 Reference

- [1] <https://purlp.com/taguig/>
- [2] https://github.com/alfredo3robles/Coursera_Capstone/blob/master/Building%20Taguig%20Dataset.ipynb
- [3] [https://en.wikipedia.org/wiki/Elbow_method_\(clustering\)](https://en.wikipedia.org/wiki/Elbow_method_(clustering))
- [4] <https://realestate.usnews.com/real-estate/articles/8-neighborhood-amenities-to-look-for-even-if-you-dont-use-them>
- [5] <https://www.mashvisor.com/blog/neighborhood-amenities-residential-investment-property/>
- [6] <https://www.forbes.com/sites/trulia/2014/11/29/12-things-that-make-a-neighborhood-truly-great/#31447fcb35f6>