

COURSERA'S IBM APPLIED DATA SCIENCE CAPSTONE

Opening a new restaurant in Tagaytay City, Cavite, Philippines

By: Robert Neil Ambion

May 2020

INTRODUCTION

Tagaytay City has been one of the most visited cities in Luzon during weekends and holidays. Most people come to Tagaytay to enjoy the breeze in the city. While enjoying the breeze of Tagaytay, most of the people took their time in spending their moments in Tagaytay on restaurants. Because of this, business owners are taking this opportunity to build more restaurants in the city. The question that lies here is where would be the ideal location if a business owner would try to open a new restaurant in the city where there would be a minimal competition.

BUSINESS PROBLEM

The objective of this capstone project is to know where would be the ideal location to open up a new restaurant in Tagaytay City using data science and machine learning technique such as clustering. This also answer to a business problem such as: where would be the ideal location to open up a restaurant with minimal competitions?

DATA

To solve the business problem, we shall need the following data:

1. List of barangays (suburbs) in Tagaytay City.
2. Latitude and Longitude coordinates of the barangays. This is to plot the map and get the venue data.
3. Venue data. This is the data of all landmarks or commercial establishments near a particular location.

Sources of data and methods to get them

1. Wikipedia page of Tagaytay City consists the list of all barangays in it. We can extract them from the website using python and beautifulsoup packages.
2. Latitude and Longitude coordinates of the barangays can be extracted from the Python geocoder package.
3. The venue data for a particular location can be extracted using the Foursquare api.

METHODOLOGY

To start this project, first we need to find and list all the barangays in Tagaytay City. To do this, we will be using web scraping techniques to get the data from the Wikipedia page of Tagaytay city (<https://en.wikipedia.org/wiki/Tagaytay>). We extracted the data using the python and beautifulsoup packages. After getting the list of barangays, we then get their corresponding coordinates using the geocoder package. Getting the coordinates of the barangays is needed as these coordinates will be used by the foursquare api to get the venues near them. After getting all the necessary initial data of the

barangays, we then converted all of it to a pandas dataframe and then visualize it using the folium package. This is to make sure we have the correct location of the barangays.

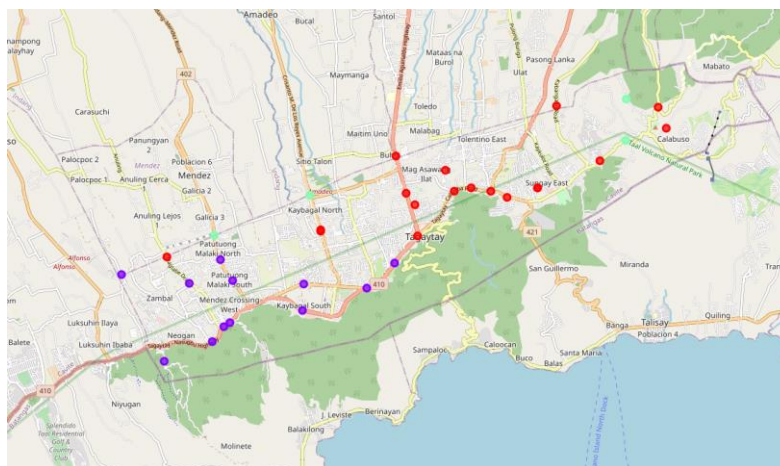
Next, we used the foursquare API to collect the venues data. Since we are only using the free version of the api, we are only limited to the top 100 venues that are within a radius of 2000 meters. We then make API calls to Foursquare passing in the geographical coordinates of the barangays. Then the API will return the venue data the we needed. After getting the data, we then merge all types of restaurant into one venue category as the goal of this project is to find the any type of restaurant within the barangay. Then we analyse each barangay by grouping the rows by barangay and taking the mean of the frequency of occurrence per venue category. With this, we are preparing the for clustering use. Since we are analysing the restaurant data, we will filter the 'restaurant' as venue category for the barangay.

Lastly, we will perform clustering on the data by using k-means algorithm. K-means Algorithm identifies k number of centroids and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. We will cluster the barangays into 3 clusters based on their frequency of occurrence for 'Restaurant'. The results will help us to identify which baraggay has the lowest frequency of having a restaurant near them. It will help us to answer where would be the ideal location to open up a restaurant with minimal competitions.

RESULTS

The results from the k-means cluster algorithm shows us that we can categorized the barangay into 3 cluster based on the frequency of occurrence for 'Restaurant'

- Cluster 0: Barangay with high concentration of restaurants
- Cluser 1: Barangay with moderate concentration of restaurants
- Cluster 2: Barangay with low concentration of restaurants.



Cluster 0: red, Cluster 1: Violet, Cluster 2: Green

DISCUSSION

Based from the clusters formed, majority of restaurants are located on the central part of the city. It can be seen that majority of restaurant are in clusters 0, this means that the barangays in this cluster has a high number of restaurants within its vicinity. Basically, it is not recommended to open up another restaurant here as there would a lot of competitions already. Cluster 1 has a moderate number of restaurants here. It is not also wise to open up some restaurant business here as there would also be a lot of competitions. The last cluster, cluster 2 shows a low number of restaurants within the vicinity of the barangay. This means that there is a high potential to open a new restaurant on barangays listed on cluster 2 as it has the minimum number of restaurant or potential competitions in its vicinity.

LIMITATIONS

In this project, we only considered one factor which is the frequency of occurrence of restaurants within a barangay. There could be other factor that could affect on where to establish new restaurant business such as the population per barangay and the buying capability of the people per barangay. Given that the researcher has limited access to those data, they are not included in the project.

CONCLUSIONS

It can be concluded that it is not recommended to establish new restaurants in the barangays the belongs to cluster 0 and 1, as it shows that it has high frequency of having restaurants near them that could mean business competitions. Meanwhile, it can also be concluded that it recommended to open up new restaurants on barangays that belongs to cluster 2, as it has the lowest frequency of having a restaurant near them meaning that it has the lowest chance of having a business competition. This project also answers the business problem question: where would be the ideal location to open up a restaurant with minimal competitions and that is on barangay on cluster 2.