The Battle of Neighborhoods

Xuebo Ling

April 23, 2021

Contents

1 Abstract

The situation we assumed is that we have a friend who wants to open a coffee shop in the capital city manila of the Philippines. So far, we don't have any information about the market of Manila. Finding out where or which one of neighborhoods has the greatest development potential is our goal of this project.

In order to solve this problem, we try to compare Manila with the cities that we already analyzed before, such as New York city and the city of Toronto. In a sense, the similarity or dissimilarity between cities could reflect the development mode of a city. The more similar city could be regarded as a training set. We could use this training set to build our model. And then we could use this model to predict our market situation in manila.

Another thing we believe is that any one of business venue does not exist alone, the existence of coffee shop tends to be influenced by some other types of shops. Therefore, we chose other types of shops as variables and the number of coffee shops in each neighbourhood as response. In this case, we utilized regression model to predict the number of coffee shops in manila and compare with the existing number of coffee shops currently. Thus, the difference of current number of coffee shops and the predicted number of coffee shops could be regarded as an indicator describing the development potential in each neighbourhood.

In this end, the neighbourhood with maximum difference could be recommended as the location to start our friend's business.

2 Introduction / Business Problem

We assumed that we have a friend who wants to open a coffee shop in the capital city *manila* of the Philippines. So far, we have no idea about the market situation of manila. Our assignment is to help our friend to find out where or which neighborhoods have a greater development potential. In order to answer this question, we have built a model to get some recommendations where to start his/her business.

3 Data acquisition and cleaning

3.1 Data sources

As we analyzed before in New York city and the city of Toronto, one type of data is the table of information about neighborhoods and their postal code, the other type of data is the location information for different demands.

The information about neighborhoods and their postal code we can scrap them from the link as follows:https://en.wikipedia.org/wiki/List_of_ZIP_codes_in_the_Philippines.

The geocoder package provides a convenient API to get the location to get the latitude and the longitude of each point.

As to venues information, the Foursquare API provide a convenient method to explore the venues around on specific location, so we could obtain venue's name and category.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
4	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop

Table 1:Venues information

3.2 Data cleaning

Data downloaded or scraped form multiple sources were combined into one table. There were a lot of outliers in our dataset.

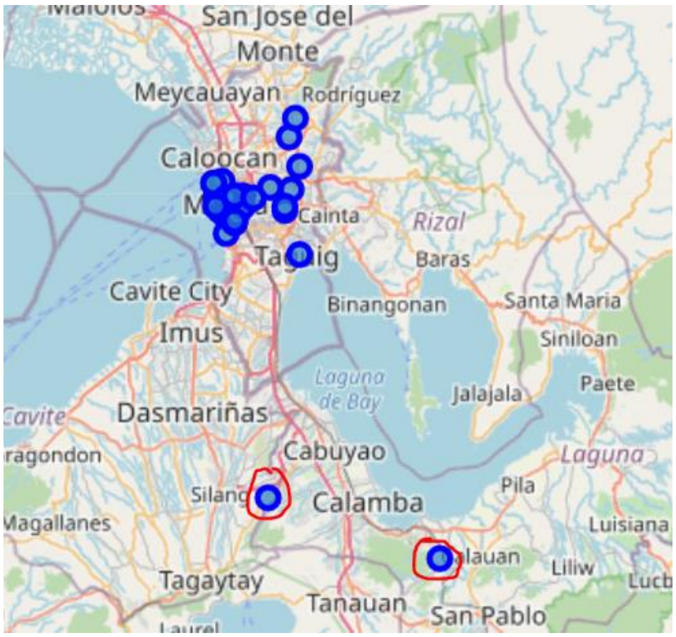

Figure 1: Data outliers

As we can see the figure 1, there are two outliers in our dataset. They are located actually not in manila. Thus, we should remove them from the dataset. The reason of generating outliers could be that the package geocoder could not find precisely the location while inputting only the name of neighborhoods.

4 Methodology

4.1 Chose similar city

The first step to deal with this problem is that finding out which city is much more similar to Manila. We explored New York City and the city of Toronto and segmented and clustered their neighborhoods. Both cities are very diverse and are the financial capitals of their respective countries. As same as Manila, we segmented their neighborhoods into different clusters.

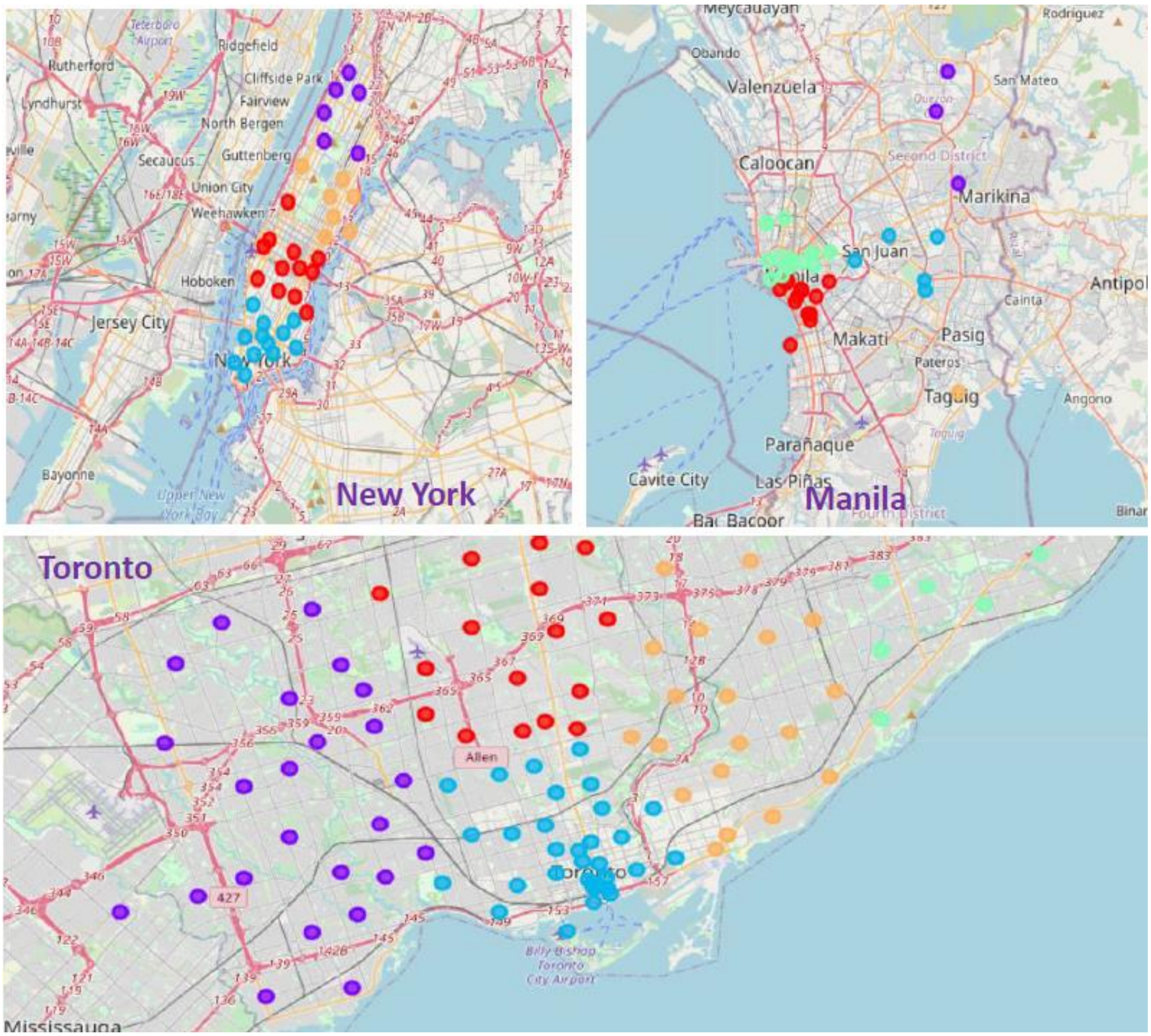

Figure 2: clustering neighborhoods of different cities

The Figure 2 depicted the clusters of different cities, we can find out that Toronto and Manila have radial distribution of neighborhoods, but the neighborhoods in New York city are distributed much more homogeneously. The density of neighborhoods in Toronto and Manila increases as the neighborhoods get close to the city center. Therefore, we could make a conclusion that manila is much more similar to Toronto than the New York city, and we prefer using dataset from Toronto to build our model.

4.2 Build model

As we mentioned before, Foursquare API provides a convenient method to explore the venues around on specific location, we could obtain venue's name and category in this way. Then one-hot encoding for each category was created so that we can count the number of venues and their types in each neighborhood. The further analysis was based on an assumption that any one of business venue does not exist alone, the existence of coffee shop tends to be influenced by some other types of shops. Therefore, we chose other types of shops as variables and the number of coffee shops in each neighborhood as response. In this case, the goal is to predict the number of coffee shops in each neighborhood, which means it is a quantitative prediction. On the other hand, all of the variables are the numbers of different types of venues in neighborhoods which are numeric. Thus, we utilized Regression model to predict the number of coffee shop in manila.

X_train (dropped Neighborhood and Coffee shop columns)

	Yoga Studio	Accessories Store	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant
0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	1
5	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0
8	0	0	1	1	1	2	2	1	0
9	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0
11	1	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0
13	2	0	0	0	0	0	0	0	1
14	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	1
17	0	0	0	0	0	0	0	0	3
18	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0

Index: Neighborhoods

Y_train

Coffee Shop
0
1
2
0
2
5
0
2
0
0
0
11

Figure 3: data x and y

The figure 3 shows the format of training dataset x and the corresponding response y. The index represents each neighborhood in Toronto. The dataset x contains all kinds of venues in Toronto except Coffee shop, and the number of coffee shop made up the dataset y.

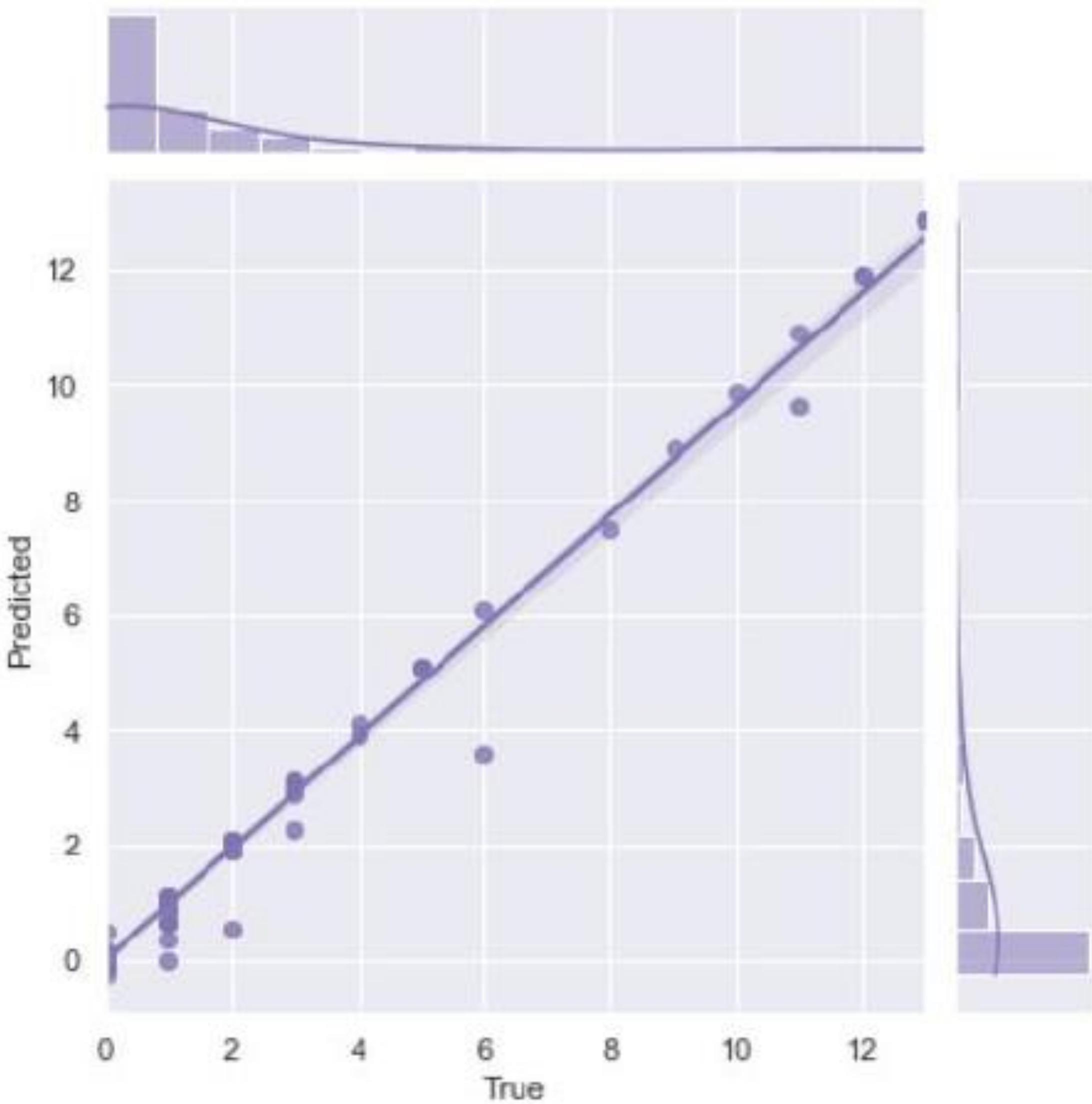

Figure 4:model results

As shown in the figure 4, the predicted value matches well to the true value. Afterwards we could use this trained model to predict coffee shop in Manila.

4.3 Predict the model

Before the prediction action was taken into account, we used Foursquare API again to obtain the current coffee shop distribution in Manila.

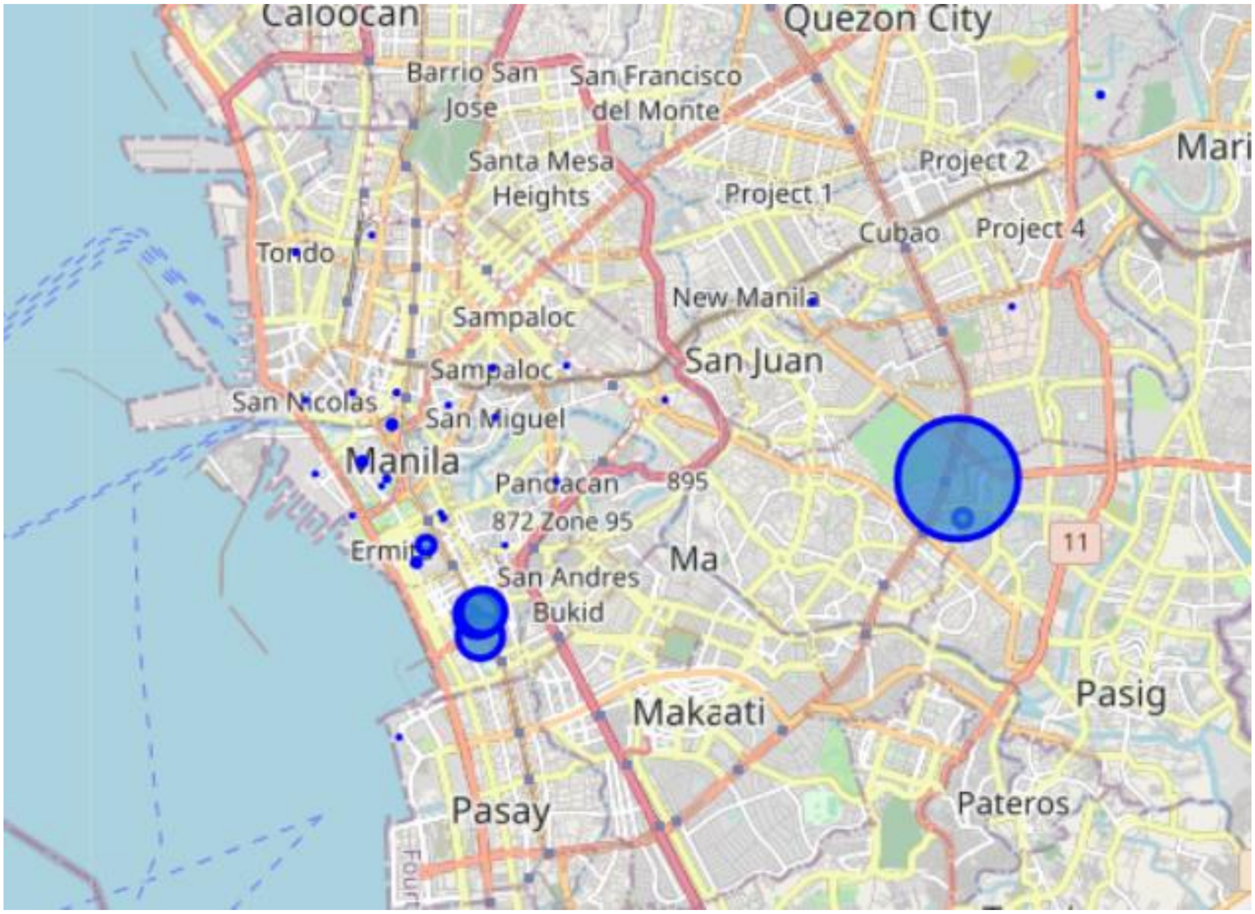

Figure 5: Distribution current coffee shops

The figure 5 depicts the distribution of coffee shops in different neighborhoods. The bigger blue circle represents the larger number of coffee shops in these neighborhoods. According to the figure we could see that the most of coffee shops are distributed randomly and homogenously in different neighborhoods. Only in three of these neighborhoods, coffee shops clustered together and have large scales there.

	Neighborhood	Real	Predicted	difference
0	Asian Development Bank,Manila	7	8	1
1	Ateneo de Manila University,Manila	3	5	2
2	Binondo,Manila	2	8	6
3	Bureau of Plant Industry,Manila	5	5	0
4	Chinabank,Manila	0	0	0

Table 2:prediction result

In the end, the model came into use to predict the coffee shop in Manilla. the difference of current number of coffee shops and the predicted number of coffee shops could be regarded as an indicator describing the development potential in each neighbourhood.

5 Results and Conclusion

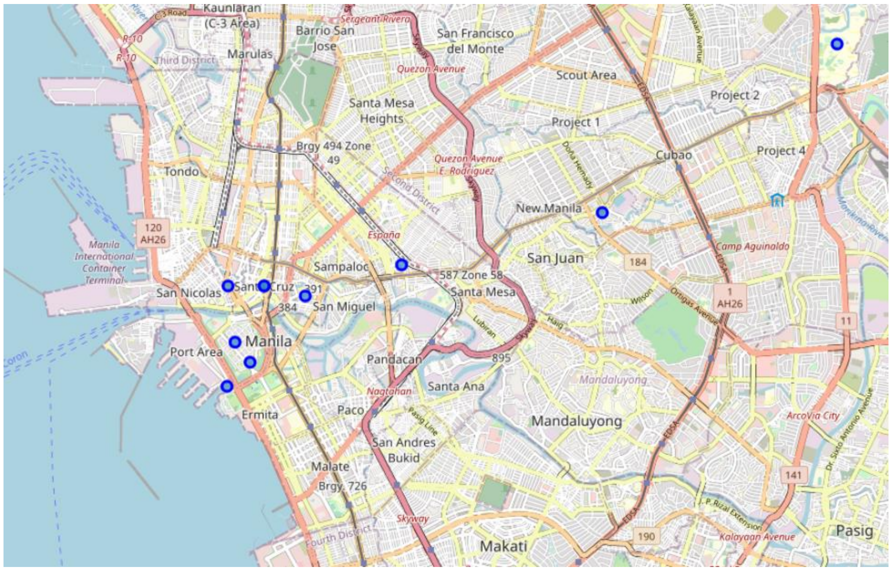

Figure 6: location of reommended neighborhoods for coffee shops

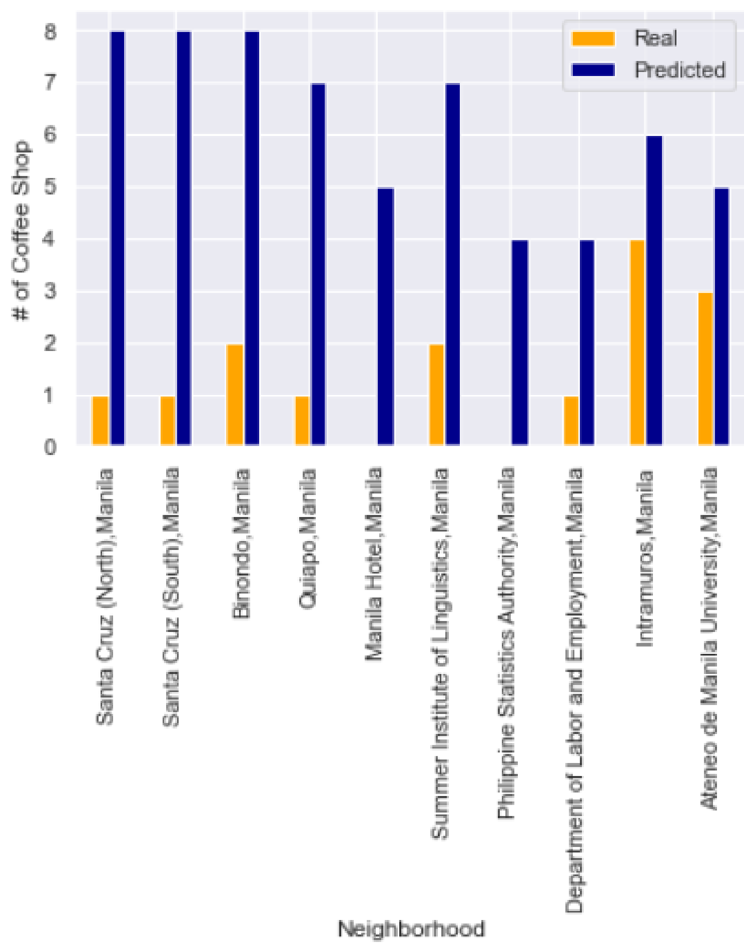

Figure 7: Top 10 recommended neighborhoods for coffee shops

Figure 6 and figure 7 shows the top 10 recommended location for opening a coffee shop. Don’t wait, start your business!