Final Coursera Capstone Project Summary

A case study of Manila, Philippines

Adeagbo Rapheal B.

May, 2020.

1. Introduction to the Problem

1.1 Background

Manila is the capital of the Philippines and a highly urbanized city. It is the most densely populated city proper in the world as of 2019.[1] This city is known for being the capital city of the Philippines. The country has a rich food culture and is one that distinct it among other countries in the world. Manila encourages this culture with its high count of restaurants, hotels and cafes. Tourists from all over the world visits this Manila to experience and enjoy culture that goes beyond food.

There are restaurants that serves different cuisines with a majority of them delivering to locals and visitors as well. However, there might be indecision of where to buy based on the customer's budget. This kind of problem is one that this project tries to trash out and find a solution to.

1.2 Problem

For someone moving to Manilla, be it for a job or personal reasons, they might want to explore and have an idea of what they are to expect in the city. For a foodie, the choice of restaurant will be an important factor and can even determine where in Manila such person can reside.

For a person with an average salary, a very expensive restaurant might not be the best fit. Hence, there is need for him to find out the best restaurants that he can afford – and of course, provides excellent service. And for Philippines which is known for its rich food culture, it might be hard to the decide and find the right restaurant that fits the person's budgets.

To provide clarity of information, this project is aimed at providing enough information to help people in this situation have sufficient information provided to help in their decision making. With provision of sufficient visualization, it will enable any visitor to take a quick glance and decide what place to visit.

[1] Source: Wikipedia.

1.3 Target audience

The target audience of this project are basically people moving to or visiting the Manila City in Philippines. Also, food lovers looking to try out new places but want to be prudent about their spending can also make use of insights from this project. Anyone on a food or academic research can make use of this project to derive insight and use as a plus to their findings.

Also, a company can use the information contained in this project in the creation of an application or a website. The information is continually updated from the sources and is will be effective for the business. Restaurants looking to establish restaurant where would be profitable might also make use of this information. Hence, this project might be important to those in the food industry.

2. Data acquisition and cleaning

2.1 Data Sources

To get the location and other information about the venues in Manila, a combination of data from the 2 API sources will be used together.

Using the Foursquare's explore API, information of venues in Manila are fetched. This includes their names, categories and locations – using latitude and longitude. Also, using the name, latitude and longitude values, Zomato search API will be used to fetch venues from its database. This API allows us to find venues based on search criteria (usually the name), latitude and longitude values and others.

2.2 Data Cleaning

Data collected from the 2 data sources, will be cleaned to extract vital information and

then be used to solve the problem.

From Foursquare API, we will retrieve the following information

- Name: The name of venues.
- Category: The category type as defined by the API.
- Latitude: The latitude value of the venue.
- Longitude: The longitude value of the venue.

From Zomato API, we will retrieve the following for each venue:

- Name: The name of the venue.
- Address: The complete address of the venue.

- Rating: The ratings as provided by many users.
- Price range: The price range the venue belongs to as defined by Zomato.
- Price for two: The average cost for two people dining at the place.
- Latitude: The latitude value of the venue.
- Longitude: The longitude value of the venue.

The data from multiple resources might not always align. Thus, it is important to combine the data retrieved from multiple resources properly. To see what output I have, I plotted the two data points on the map. I then try to combine data points that have their latitude and longitude values where necessary. From the remaining selected venues, I inspected the venues to ensure that any remaining mismatched venues were also removed from the final dataset of venues before I began any analysis.

I observed that there are many venues identified by both Foursquare and Zomato that showed overlapping near coastline, Malate, Binondo and Sampaloc. To combine the two datasets, I had to check that the latitude and longitude values of each corresponding venue match. So, I rounded both the latitude and longitude values up to 4 decimal places then calculate the difference between the corresponding latitude and longitude values and see if the difference is less than 0.0004. If this is the case, it ideally meant that the two locations are same.

After the merging, it was observed that some venues have the same longitude and latitude values. For these set of venues, I assumed that they are in the same building/complex. This type of scenario is not uncommon in the Phillipines, hence no deletion will be made as sources of data are reliable and up-to-date. Also, venues that have 0.0 rating were dropped as it means it's not been rated yet.

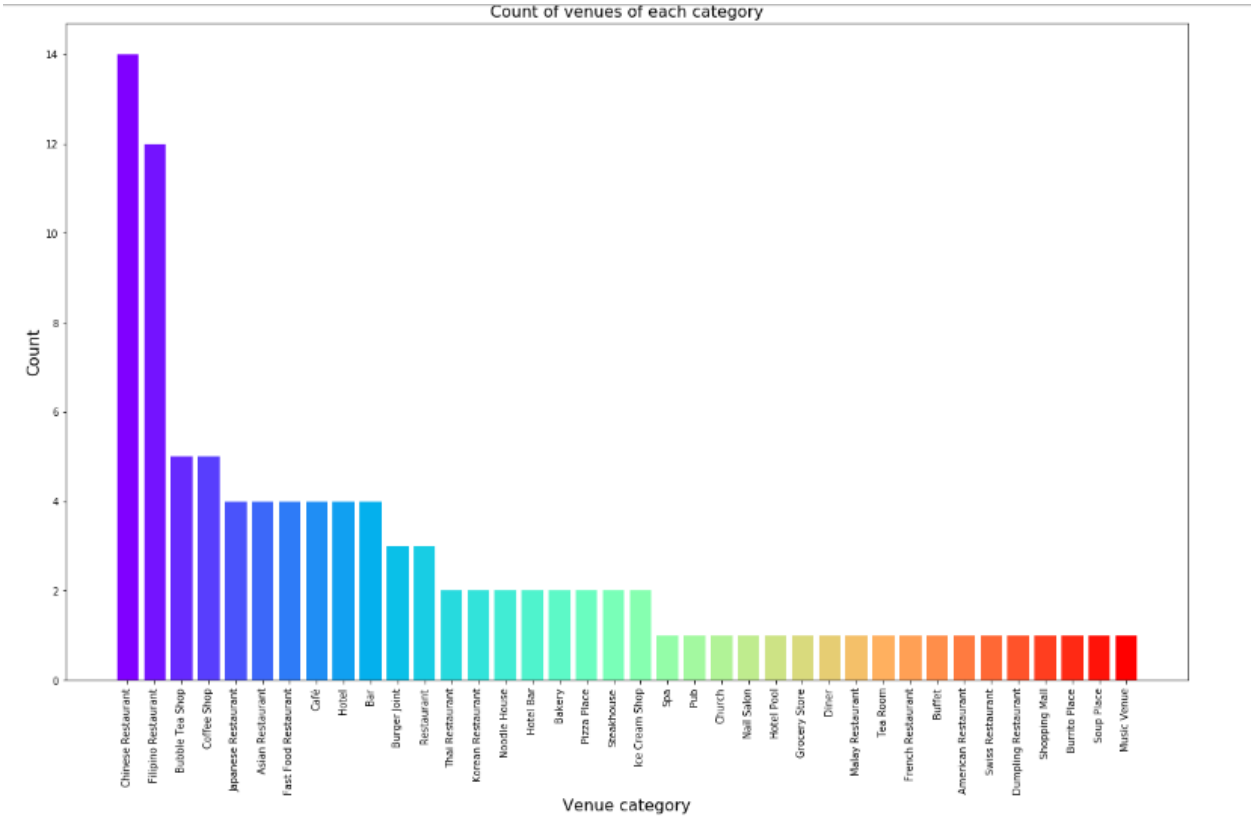
3. Methodology

In order to solve the problem risen in the research work, I analyzed the data that created based on the ratings and price of each venue. I identified places where many venues are located so that any visitor can go to one place and enjoy the option to choose amongst many venue options. I also explored areas that are highly rated and those that are low rated while also plotting the map of high and low priced venues. Lastly, I clustered the venues based on the available information of each venue. This will allow us to clearly identify which venues can be recommended and with what characteristics.

3.1 Analysis

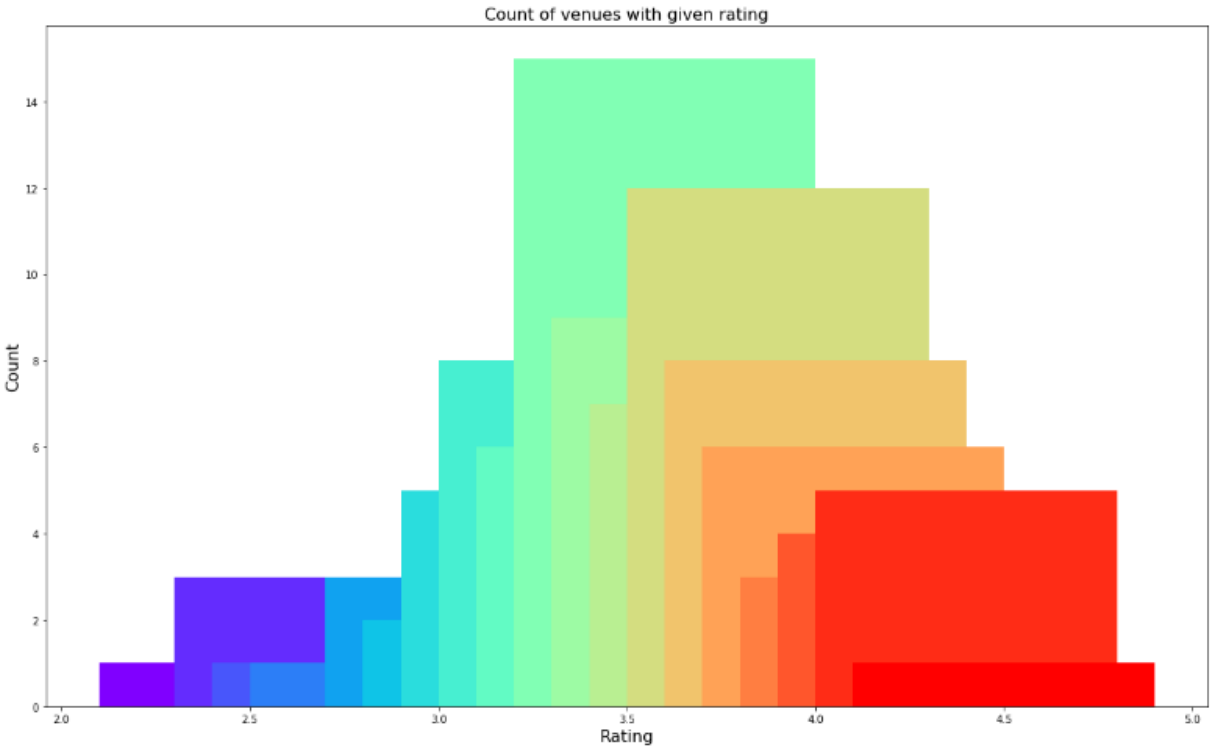
When the complete dataset is now in its final form, I inspected these venues based on their rating. The rating of a venue was based on user reviews and belongs to a range from 1 to 5. I also analyzed the venues based on their price per person as well as the price range.

I had various types of venues in the final dataset in which I checked which the majority venue categories in the list are.

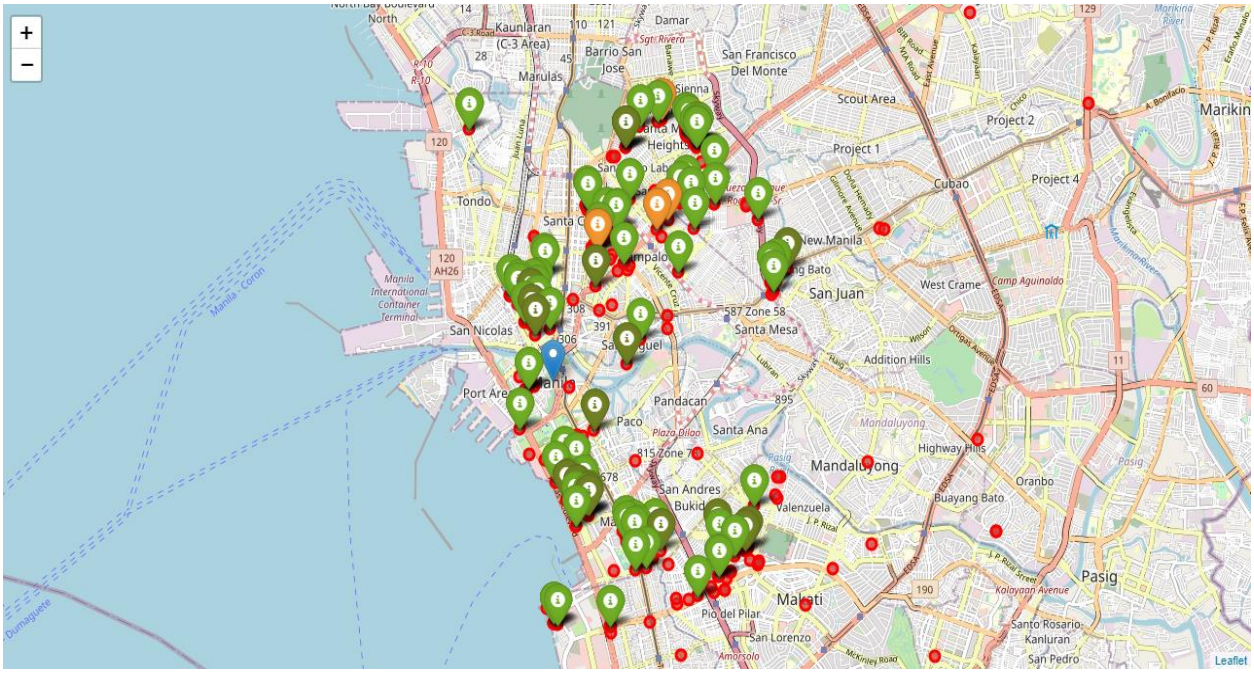


As we can depict from the histogram, the majority venues are Chinese and Filipino Restaurants. Ideally, we can say Manilla is a welcoming place for Chinese tourists and lovers of Chinese food.

Besides the categories classification, the rating of a venue is an important factor on which a visitor decides whether it is worth it to visit the place. To put this in consideration, I found out what the average rating for all the venues in the city are. After that, I plotted the venues on the map and color coded them. I identified the various rating values and plot them as a bar plot with their counts to see the most common rating.

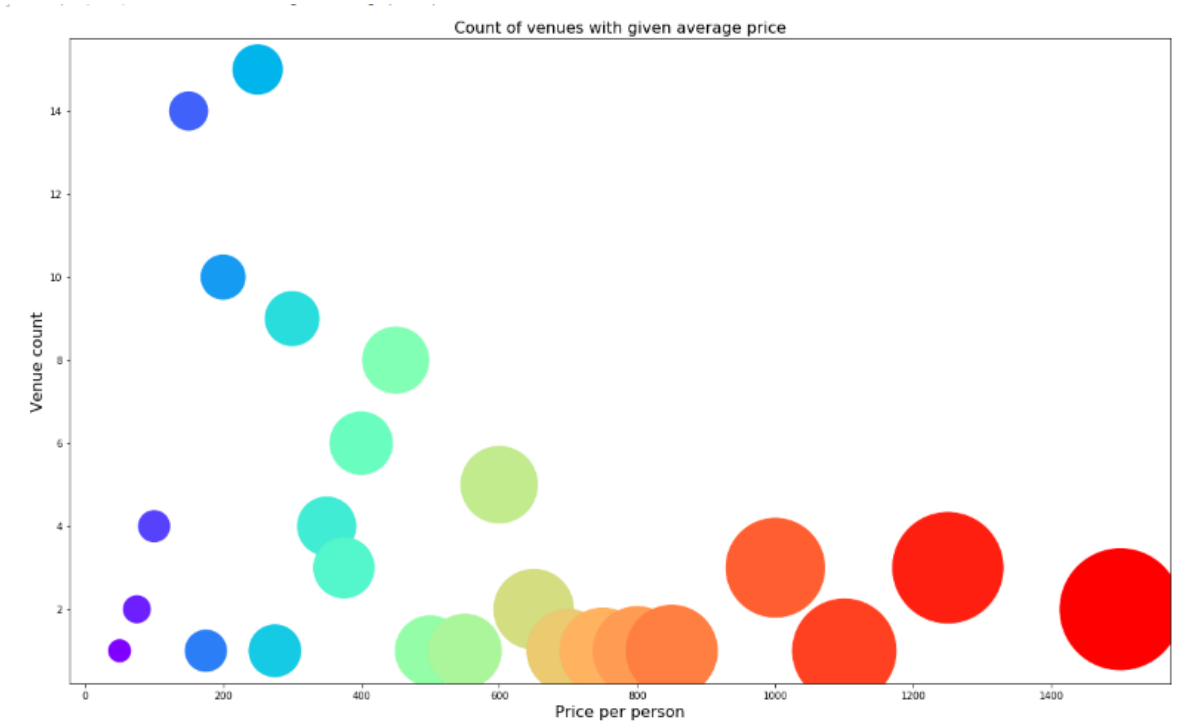


From the plot above, it is clear that majority of the venues have their rating close to 3.5. Also the lowest number of restaurant has a low rating between 2 and 3. Next, I created bins for various ratings and plotted them in different colors on the map. The ratings were divided into 4 bins: 1 to 2, 2 to 3, 3 to 4 and 4 to 5. I then plotted these venues on a map to better represent the data.



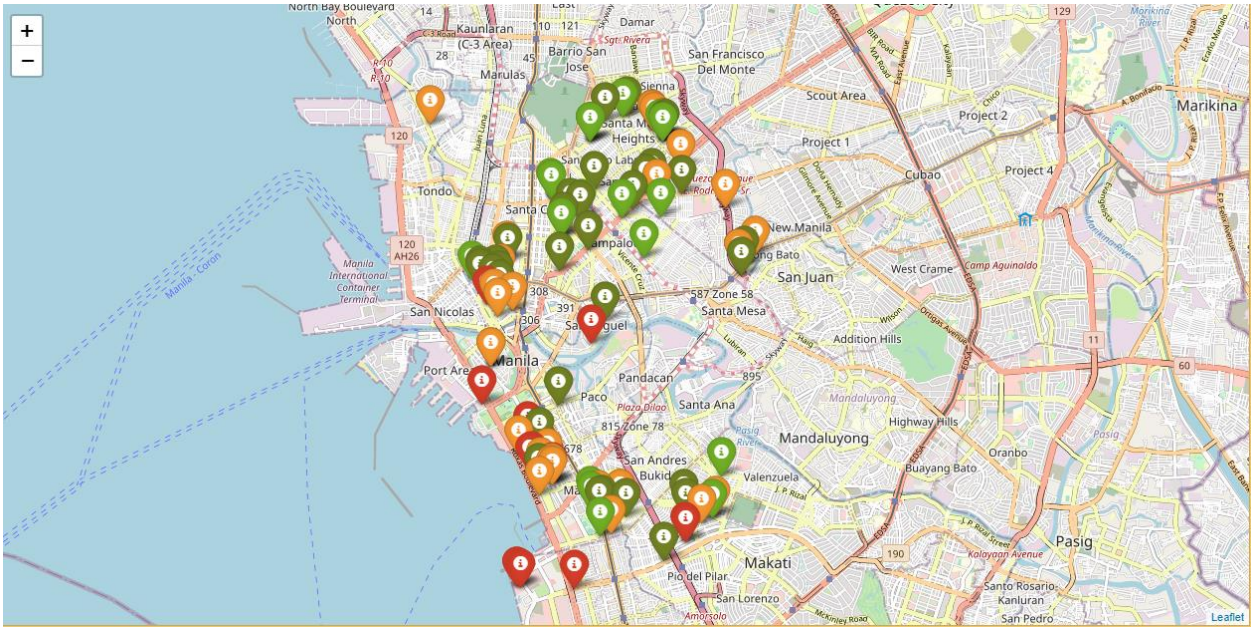
The map has the location of all the venues. It appears that most of the venues have ratings above 3.0. However, 2 restaurants have rating values that are lesser than 3.0 in San Teresita and one in Santa Cruz.

The venues were also classified based on the price values. I explored the price using a scatter plot between the price and the count of venues with that average price. I hence size the points based on the price to highlight the venues' price.



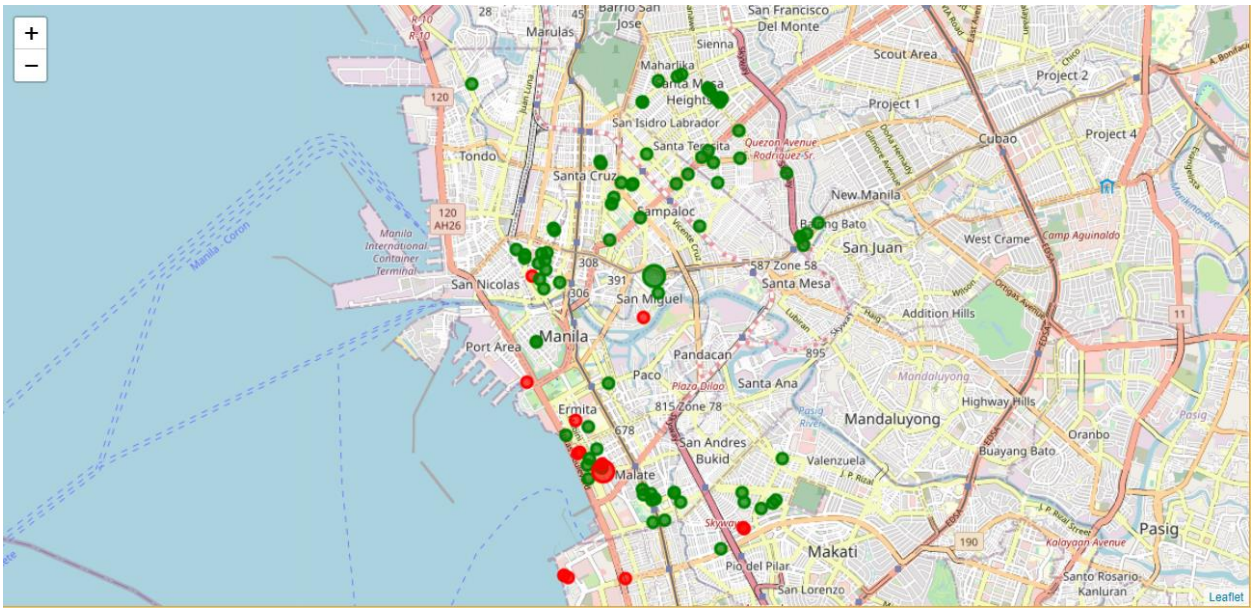
From the plot above we can see that a large number of venues have an average price between PHP400 and PHP700.

To further strengthened this, I used the price range column to plot the venues on a map. I represented the venues with lower price in green and move towards red as the price increases.



From the map, the expensive venues are along the coast line or close to the sea. The middle price ranges are seen to be clustered around Sampaloc. Other places around this location have moderate prices also.

Also, I used the K-means clustering to identify similar venues and the relationship among them. I clustered the venues into two groups.



From the map, we can see the two clusters:

- The first cluster is spread across the whole city and includes the majority venues. It consists of venues with relatively low prices and price range.

- The second cluster is very sparsely spread and has very limited venues. This cluster consists of areas with very high price and price range.

4. Results and Discussion

Based on our analysis above, we can draw a number of conclusions that will be useful to aid any visitor visiting the city of Manila, Philippines.

After collecting data from the Foursquare and Zomato APIs, we got a list of 232 different venues. However, not all venues from the two APIs were identical. Hence, we had to inspect their latitude and longitude values as well as names to combine them and remove all the outliers. This resulted in a total venue count of 100.

We identified that from the total set of venues, majority of them were Chinese Restaurants and then Fillipino Restaurants follows. A visitor who loves Chinese and Filipino foods would surely want Manila as a resident city.

While the complete range of ratings range from 1 to 5, the majority venues have ratings close to 4. This means that most restaurants provide good quality food which is liked by the people of the city, thus indicating the high rating. When we plot these venues on the map, we discover that there are clusters of venues close to the sea. This means that anyone that gets a residential apartment close to the water body will definitely enjoy good food. However, the budget of the person might be a limiting factor which takes us to the next insight.

When observe the price values of each venue, we explore that many venues have prices which are in the range of PHP400 to PHP700 for one person. However, the variation in prices is very large, given the complete range starts from PHP100 and goes up to PHP1600. On plotting the venues based on their price range on the map, we discovered that venues located near the coastline are relatively priced higher than those far from it.

Finally, through clusters we identified that there are many venues which are relatively lower priced but have an average rating of 3.69. On the other hand, there are few venues which are high priced and have average rating of 3.76.

If you're looking for cheap places with relatively high rating, you should check San Isidro Labrador. If you're looking for the best places, with the highest rating but might also carry a high price tag, you should visit Binando and Malate A company can use this information to build up an online website/mobile application, to provide users with up to date information about various venues in the city based on the search criteria (name, rating and price).

5. Conclusion

The purpose of this project was to explore the places that a person visiting Manila could visit or stay. The venues have been identified using Foursquare and Zomato API and have been plotted on the map. The map reveals that there are three major areas a person can visit or stay: San Isidro Labrador, Binondo and Malate. Based on the visitor's venue rating and price requirements, he/she can choose amongst the three places.