Opening a new food venue in Surabaya, Indonesia

Introduction

Surabaya is the second largest city in Indonesia and the third-largest metropolitan area in Indonesia [1]. As a metropolitan city, all types of Indonesian cuisine and other international restaurants have a presence in Surabaya. Indonesian usually don't hesitate to try different cuisines and experience with different flavors, anyone who interested in culinary business can take this opportunity. In many cases, one of the strong consideration to open a new food venue (restaurant, coffe shop, etc) is about the competitors around the location.

We will make cluster based on food venue categories. Surabaya has 31 districs and each district will be clustered using unsupervised machine learning K-Means clustering algorithm. For the result, we prefer locations that are not already crowded with food venue that has similar categories.

This project can be useful for business owners or any other stakeholders who are looking to open a new food venue in Surabaya. The aim of this project is to carefully analyze appropriate data and find recommendations for the stakeholders.

Data

You can find the list of districts in Surabaya [2].

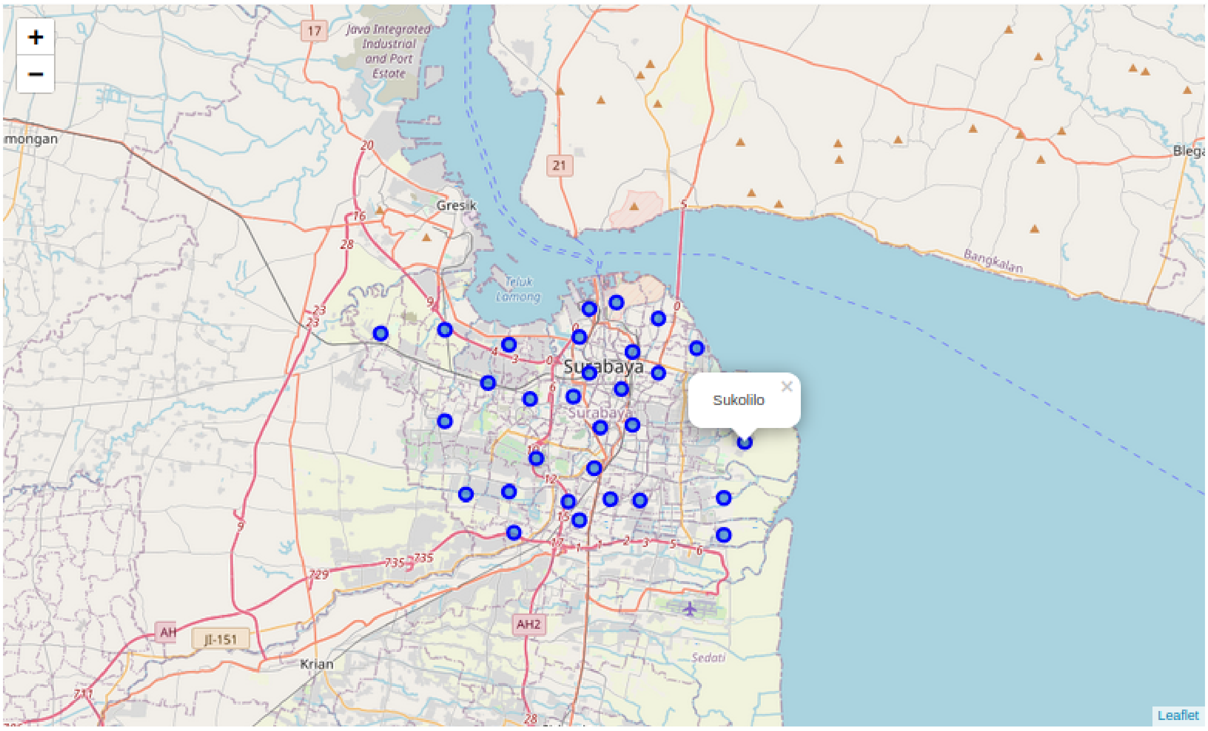
There is no public dataset to retrieve the coordinates of each distric in Surabaya, so I collect them manually. After that I create the clean dataset that contain latitude and longitude features. You can googling the exact coordinate but it returns with WGS84 datum format. You can get the latitude and logitude by convert it in [3].

We can use Forsquare API to get the most common food venues of given district of Surabaya. For each district, we have chosen the radius to be 500 meter.

Methodology

As a database, I used GitHub repository in my study. My master data which has the main components Neighborhood, Latitude and Longitude informations of the city.

I used python folium library to visualize geographic details of Surabaya and its district and I created a map of Surabaya with districtss superimposed on top. I used latitude and longitude values to get the visual as below:



I utilized the Foursquare API to explore the districts and segment them. I designed the limit as 100 venue and the radius 500 meter for each borough from their given latitude and longitude informations.

In summary of this data 216 venues were returned by Foursquare. Here is a merged table of districts and venues.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Benowo	-7.226500	112.64970	Benowo Trade Centre [BTC], Benowo, Surabaya	-7.225492	112.648972	Shoe Store
1	Benowo	-7.226500	112.64970	Warung Belut Benowo	-7.222692	112.649805	Breakfast Spot
2	Bubutan	-7.249962	112.73011	Kampoeng Ilmu	-7.251933	112.728690	Bookstore
3	Bubutan	-7.249962	112.73011	Dunkin'	-7.248629	112.730910	Donut Shop
4	Bubutan	-7.249962	112.73011	Pasar Turi	-7.246576	112.732809	Market

In summary of this 82 unique categories were returned by Foursquare, then I created a table which shows list of top 10 venue category for each borough in below table.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Benowo	Shoe Store	Breakfast Spot	Flea Market	Dim Sum Restaurant	Diner	Discount Store	Donut Shop	Dumpling Restaurant	Electronics Store	Fast Food Restaurant
1	Bubutan	Coffee Shop	Convenience Store	Fast Food Restaurant	Discount Store	Market	Food Truck	Food Court	Donut Shop	Shopping Mall	Bookstore
2	Bulak	Gift Shop	Vegetarian / Vegan Restaurant	Flea Market	Dim Sum Restaurant	Diner	Discount Store	Donut Shop	Dumpling Restaurant	Electronics Store	Fast Food Restaurant
3	Dukuh Pakis	Mobile Phone Shop	Vegetarian / Vegan Restaurant	Cosmetics Shop	Dim Sum Restaurant	Diner	Discount Store	Donut Shop	Dumpling Restaurant	Electronics Store	Fast Food Restaurant
4	Gayungan	Convenience Store	Steakhouse	Food Truck	Bakery	Café	Seafood Restaurant	Indonesian Restaurant	Boutique	Pizza Place	Vegetarian / Vegan Restaurant

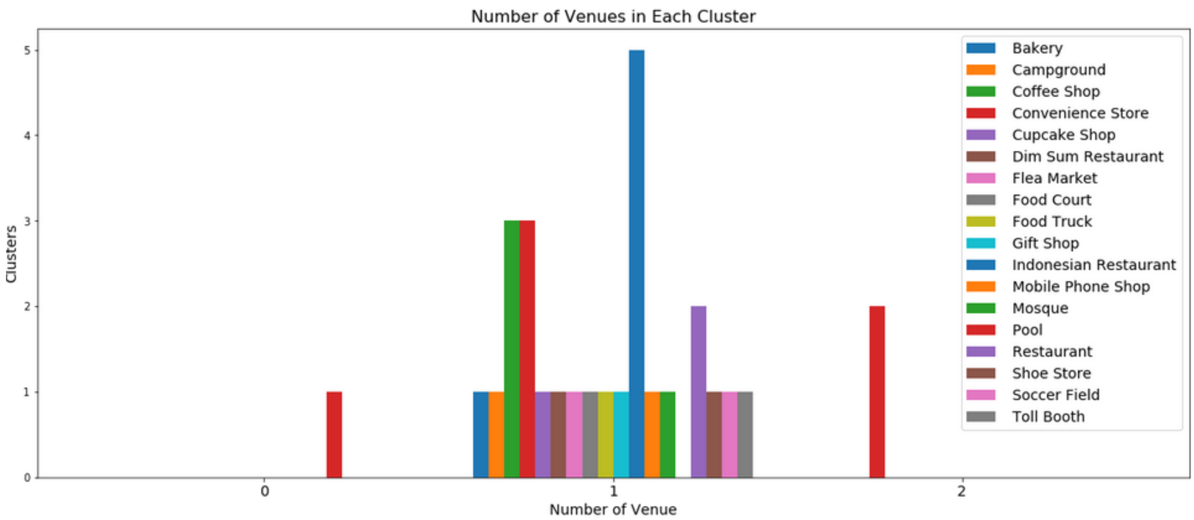
We have some common venue categories in boroughs. In this reason I used unsupervised learning K-means algorithm to cluster the boroughs. K-Means algorithm is one of the most common cluster method of unsupervised learning.

Here is my merged table with cluster labels for each borough.

	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	Benowo	-7.226500	112.649700	1	Shoe Store	Breakfast Spot	Flea Market	Dim Sum Restaurant	Diner	Discount Store	Donut Shop	Dumpling Restaurant
1	Bubutan	-7.249962	112.730110	1	Coffee Shop	Convenience Store	Fast Food Restaurant	Discount Store	Market	Food Truck	Food Court	Donut Shop
2	Bulak	-7.236800	112.789700	1	Gift Shop	Vegetarian / Vegan Restaurant	Flea Market	Dim Sum Restaurant	Diner	Discount Store	Donut Shop	Dumpling Restaurant
3	Dukuh Pakis	-7.296800	112.700300	1	Mobile Phone Shop	Vegetarian / Vegan Restaurant	Cosmetics Shop	Dim Sum Restaurant	Diner	Discount Store	Donut Shop	Dumpling Restaurant
4	Gayungan	-7.330800	112.724200	1	Convenience Store	Steakhouse	Food Truck	Bakery	Café	Seafood Restaurant	Indonesian Restaurant	Boutique
5	Genteng	-7.259088	112.747986	1	Coffee Shop	Multiplex	Fried Chicken Joint	Indonesian Restaurant	Soup Place	Bed & Breakfast	Furniture / Home Store	Noodle House
6	Gubeng	-7.278906	112.753945	1	Indonesian Restaurant	Food Truck	Bakery	Chinese Restaurant	Multiplex	Padangnese Restaurant	Electronics Store	Juice Bar
7	Gunung Anyar	-7.339200	112.804600	2	Convenience Store	Cosmetics Shop	Dim Sum Restaurant	Diner	Discount Store	Donut Shop	Dumpling Restaurant	Electronics Store

Results

We can also estimate the number of 1st Most Common Venue in each cluster. Thus, we can create a bar chart which may help us to find proper labels for each cluster.



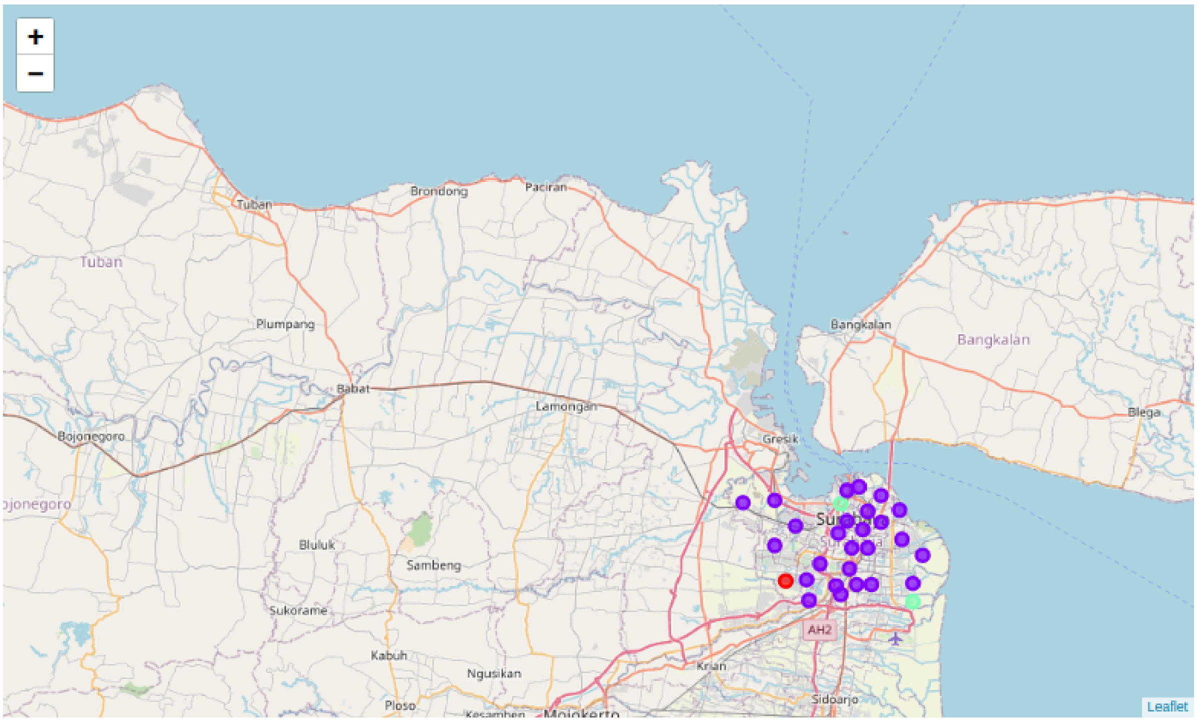
When we examine above graph we can label each cluster as follows:

Cluster 0 : Pool

Cluster 1 : Indonesian Restaurant

Cluster 2 : Convenience Store

You can also see a clustered map boroughs of Istanbul in the below.



Discussion

As I mentioned before, we can see there are 3 cluster where cluster 0 and 2 is has similar characteristics where most common venue in these clustered area is not food venue so we can give the recommendation these area to open a new any food venues. In cluster 1 the most common venue is related to food venues categories so the type of restaurant or caffee determine the recomendation area. We strict with the rule that we recommend the area with less spesific competitor around the area that we are choosen. Spesific competitors means the type of food venues, for instance we recommend open a coffee shop in Sukolilo because that type of food venue not in top of most common venue categories.

I ended the study by visualizing the data and clustering information on the Surabaya map. In future studies, web or telephone applications can be carried out to end-user.

Conclusion

In this project, we have gone through the process of identifying business problem, specifying the requirement of the data, preparing and exctracting the data, performing the machine learning by utilizing K-Means clustering algorithm and providing recommendation to the stakeholders.

In future project I hope there is enough data to be process so we can use the result as a reference because they are more reliable. There are too little amount of the data in Foursquare about any venues in Surabaya. Indonesian venues owner isn’t familiar with Foursquare so they don’t register the venue into it. We cannot generalize the mapping clustered area is to be recommended for location opening a new food venues but it’s give us a good insight a little bit about Surabaya city because in real life the most venue in there is food categories.

References

[1] https://en.wikipedia.org/wiki/Surabaya

[2] https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Kota_Surabaya

[3] https://www.earthpoint.us/Convert.aspx