

Exploring Places and Area Region in Surabaya

Muhammad Sifa'ul Rizky

April 17th 2019

1. Introduction

Surabaya is a big city of Indonesia. Instead of Jakarta, capital cities of Indonesia, Surabaya are second biggest cities in Indonesia. With big city, of course we should expect high about the developing program. With 326.81 km squared of areas region of Surabaya, actually we should hoping that building development is quite fast [1]. Since there are so many building we can develop, such as restaurant, coffee shops, supermarket, etc that can helping for financial of Surabaya, because with growing businesses, of course many people would get a job, and decrease people who don't have job. My problem is how we can know which neighborhood that is very crowded with building, and which is less crowded, so that we can create more and more building. I think for businessman who want to build something like restaurant, coffee shops, etc and maybe for government is important to know that in which area should they have to build, is it too dense to build some venue or less dense.

2. Data

Based on problem above, we need so much data about Surabaya itself and Foursquare API to helping us know what is nearby that area. So data sources we needed is:

- Data of areas region of every neighborhood on Surabaya, for this data I am scraping the data from BPS Kota Surabaya[1].
- Data of latitude and longitude of every neighborhood on Surabaya[2], for this data I am using data longitude and latitude of every districts in Indonesia from Github.
- Data of Foursquare API for understanding what building in radius of every neighborhood on Surabaya[3].

3. Methodology

In this project, we want to know what kind of places near of every neighborhood, so that because Surabaya is so big, I decide to create radius 1 km for every location, to make sure that everything is gathered. After that I want to know how many places are gathered, and can create some visualization of it, and then linked it with areas region, it is quite dense or not. Then, using K-Means to helping us understand actually it can diversified into how many cluster to understand that maybe we can build something on this neighborhood, just like that.

4. Analysis

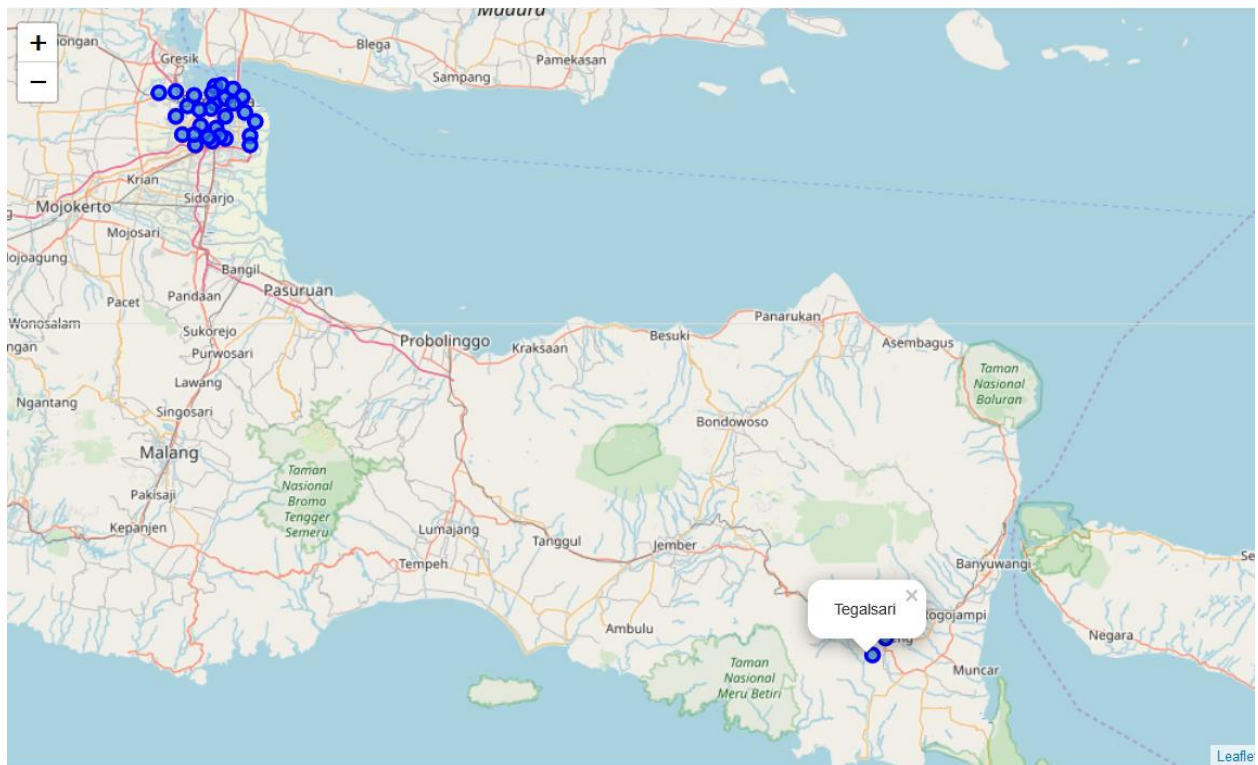


Figure 1: Before Fixed

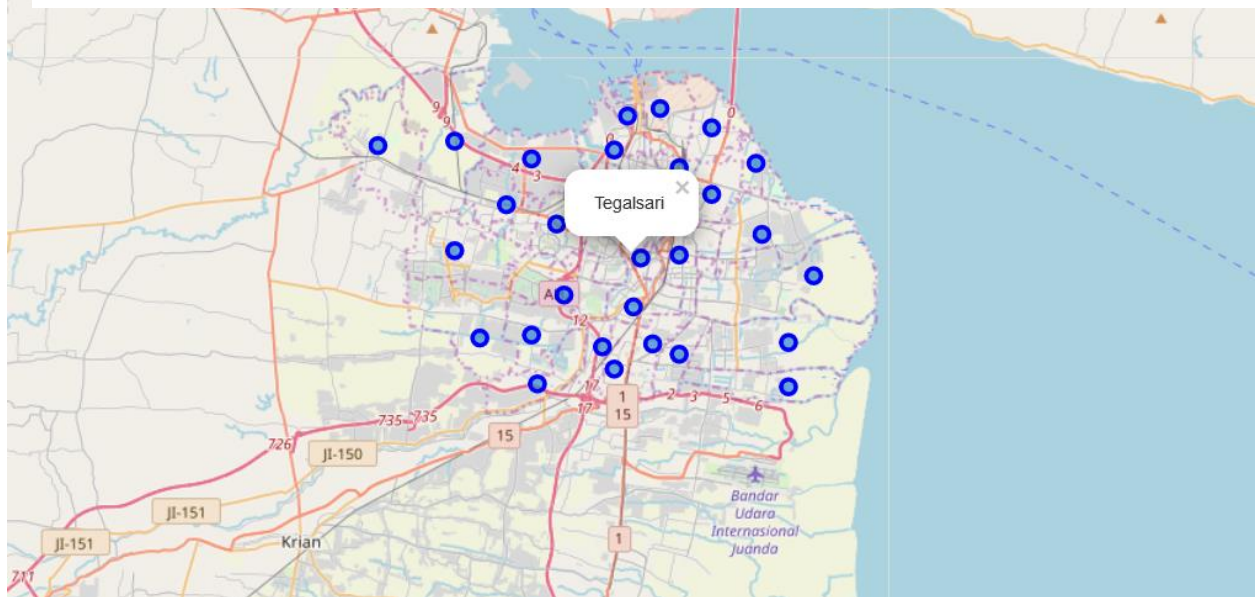


Figure 2 After Fixed

After fixed, I want to know how was the statistical value of Area Region in Surabaya, for now I am using Box Plot to understand it.

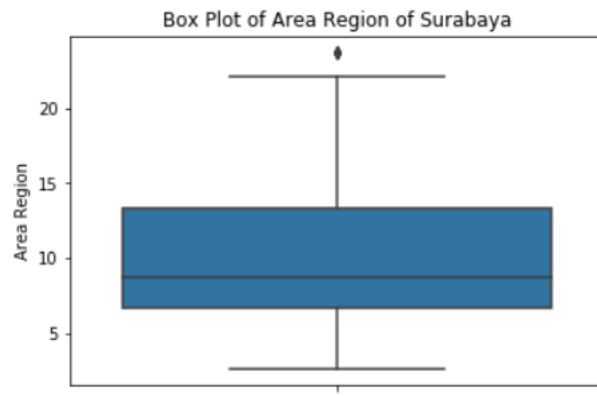


Figure 3: Box Plot of Area Region

From this data we know that majority of neighborhood on Surabaya has Area Region between approx. 7 until approx. 13 km squared, while there are some outlier outside 20 km squared.

Now, it's time to gathered data from Foursquare API to know what kind of places in every neighborhood of Surabaya. I am using radius 1 km since that Surabaya is a big city.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Tegalsari	-7.279848	112.736069	Völks Coffee	-7.281788	112.736857	Coffee Shop
1	Tegalsari	-7.279848	112.736069	Soto daging imam bonjol	-7.276836	112.735557	Indonesian Restaurant
2	Tegalsari	-7.279848	112.736069	Bon Ami	-7.283387	112.735248	Bakery
3	Tegalsari	-7.279848	112.736069	La Rucola Ristorante Italiano	-7.282881	112.737737	Italian Restaurant
4	Tegalsari	-7.279848	112.736069	Cafesera Kartini	-7.279219	112.738307	Food Court

Figure 4: Data from Foursquare

We get 159 unique categories, with spread of every data is like this visualization.

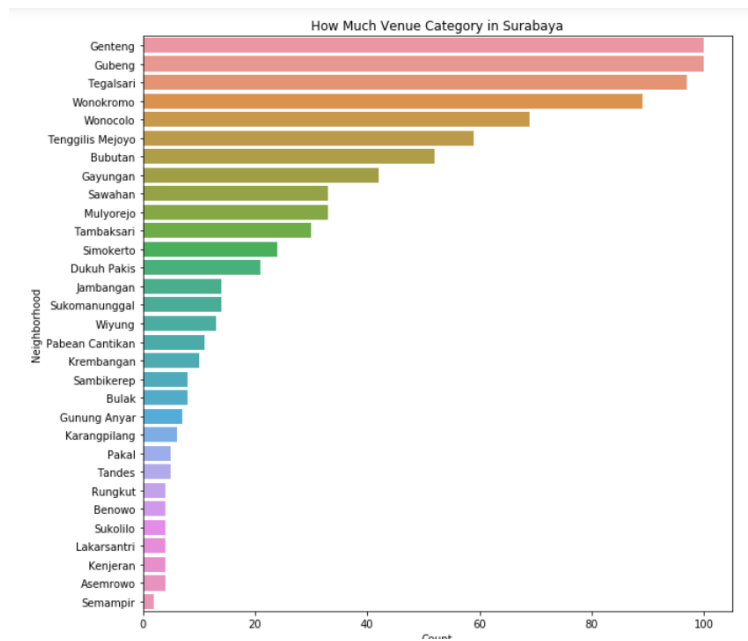


Figure 5: Count of Places Gathered

We can see that Genteng and Gubeng have so many venues or places gathered with 100 each. That's so much, let's extract to what is most common venues on Surabaya.

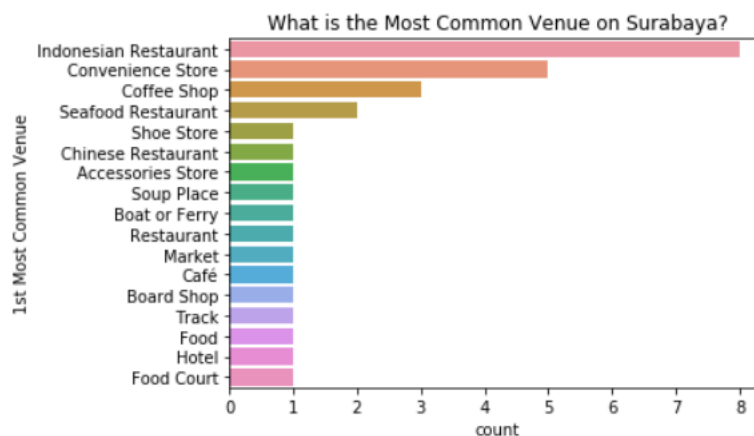


Figure 6: Most Common Places

That's interesting since that 8 of 31 neighborhoods has same places. Indonesian Restaurant. Let's create cluster for helping us how many we can use with K-Means. Using K-Mean to know which best-k we should use.

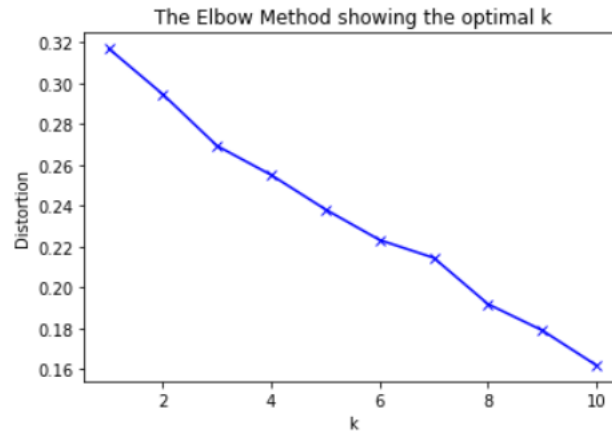


Figure 7: Select Best K

It was tough choice since that there are no significant difference, so I choose to create 3 cluster instead of 7 or 8 because if we create more cluster it takes more complicated. So I am choosing 3 cluster. And merge data from data gathered from Foursquare, get the result. I am using additional geojson[5] to helping understanding shape of Surabaya.

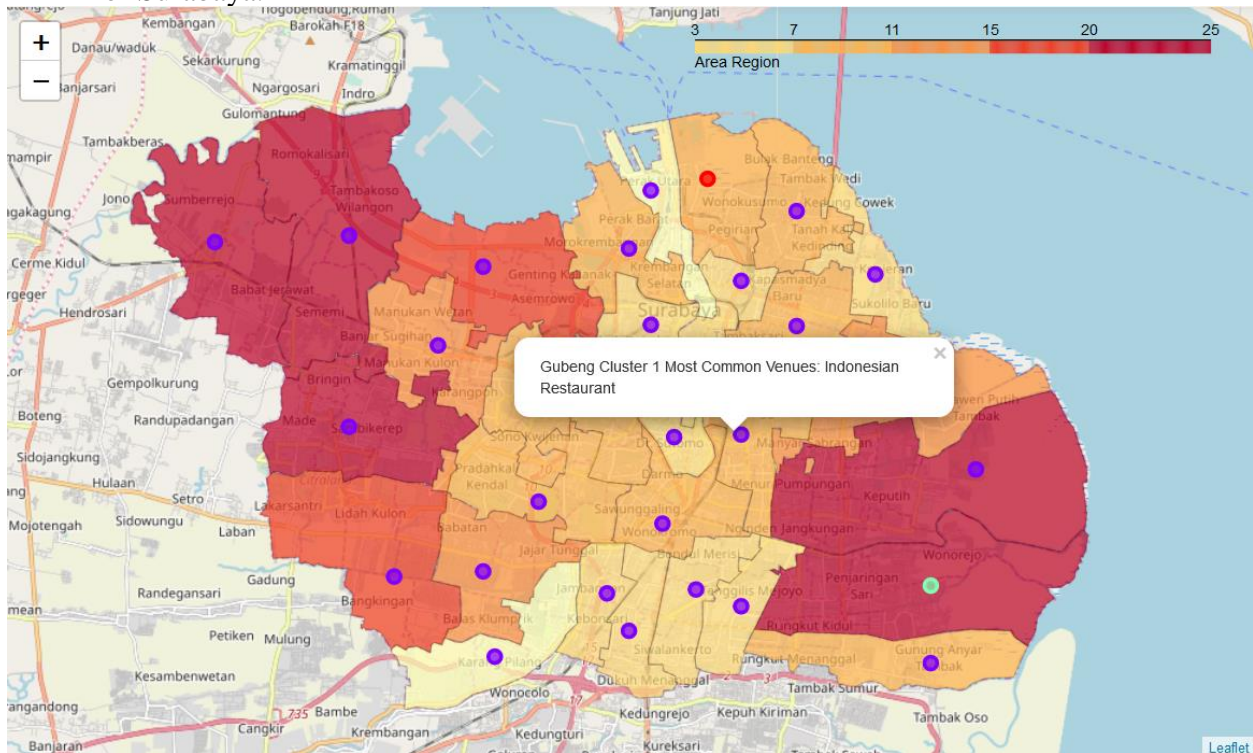


Figure 8: Final Choropleth of Surabaya

For now we can see that actually in cluster, majority in cluster colored blue, when just one each is clustered red and green. And of course color of choropleth indicates that how big that area was. We can see that if area is bigger, the color is more darker.

5. Result and Discussion

After so much time consuming, finally we can get the result. Our analysis show that it's great to know that with how much your area is, it is still doesn't have any impact, except for Indonesian Restaurant that they are spread out to 8 in 31 neighborhood to make it top places in Surabaya.

	Count	Area Region
Count	1.000000	-0.494079
Area Region	-0.494079	1.000000

Figure 9: Correlation of Area and How Many Venues

	Neighborhood	Count	Area Region	Average density
0	Genteng	100	4.05	0.04
2	Tegalsari	97	4.29	0.04
6	Bubutan	52	3.86	0.07
1	Gubeng	100	7.99	0.08
5	Tenggiling Mejoyo	59	5.52	0.09

Figure 10: Top 5 Dense Area

27	Lakarsantri	4	18.99	4.75
24	Rungkut	4	21.08	5.27
26	Sukolilo	4	23.68	5.92
25	Benowo	4	23.73	5.93

Figure 11: Top 4 Less Dense

And then we absolutely know that in **Genteng** and **Tegalsari** is very crowded with just 0,04 km squared per each venue. While only 12 from 31 has density above 1 km squared. In some neighborhood is so dense, some of them not.

For discussion in future, maybe you can create with another method of clustering or using K-Means with different k values. Or you can analyze why Indonesian Restaurant is so dominant in Surabaya, you can explore anywhere. I hope that my report can open your mind that so many thing we can explore in this data.

6. Conclusion

So after looking at the maps above and the data itself. It is clear that Surabaya is dense, when you sample just 2 or more neighborhood, such as **Tegalsari** and **Genteng**. But if you see in depth, there are some land that we can build something. If you want to start your businesses, go ahead for Indonesian Restaurant. You can build something that can

improve economy of each neighborhood. Maybe in some places there are few venues gathered because the data is not complete or what kind of else problem. But in this I assume that we need more venues, but not in dense neighborhood. Take a look at **Lakarsantri, Rungkut, Sukolilo, and Benowo**. Thank you for reading until the end.

7. References

1. [BPS Kota Surabaya](#)
2. [Data Districts in Indonesia](#)
3. [Foursquare API](#)
4. [Folium Package](#)
5. [Surabaya GeoJSON](#)