Battle of Neighborhoods - Clustering and Segmenting the Neighborhoods of Jakarta and Kuala Lumpur, to Find a Recommended Place to Open a New Restaurant or a New Business Venue in the Neighborhood.

February 3rd, 2021

1 Introduction

In this report I will analyze area of cluster between two major metropolitan area in Indonesia that is Jakarta and Kuala Lumpur. Jakarta and Kuala Lumpur are very popular and most populated metropolitan area in Indonesia. Although Jakarta have approximately 3 times population than Kuala Lumpur, Kuala Lumpur has its own destination, unique places, and landmark to go to.

1.0.1 Disclaimer:

We will use Foursquare API to get data about venues and places in the neighborhood. However, there is another popular venues and places in Jakarta or Kuala Lumpur neighborhood that is not in the data of Foursquare API. That being said, this article is not the best result for venues and places in the neighborhood, although this article is not perfect I hope this will help you or others to know about my analysis.

2 Business Problem

The aim of this report is to help tourist or business owner to open new destinations or places in the neighborhood depending on the experiences that neighborhood have. Once the data is obtained, the cluster and segmentation between neighborhood is created to see which neighborhood has the same similarity based on destination and places. This also will help people to make decision if they want to migrate or move into another neighborhood.

3 Data Collecting

In this report, we require neighborhood for Jakarta and Kuala Lumpur. Using the location of the neighborhood we can search most popular venue or places for each category using **Foursquare API**. We also need the coordinates/geographical location for each neighborhood in Jakarta and Kuala Lumpur. Using the coordinates of the neighborhood we can visualize with **Open-StreetMap using Folium API**.

3.1 Jakarta

In order to get neighborhood in Jakarta we scrape the data from:

https://id.Wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Daerah_Khusus_Ibukota_Jakarta

In this Wikipedia page there is several tables representing each Town in Jakarta. in each table, there is data about name of neighborhood for each town and name of borough for each neighborhood.

After doing data processing we limit the data and concatenate 5 table into 1 table containing information about:

1. *Neighborhood:* we call this neighborhood to make it easy for report.
2. *Borough:* Name of Administrative Town for each neighborhood.

At the end, we obtained 48 rows of data each representing its neighborhood.

3.2 Kuala Lumpur

We scrape neighborhood data in Kuala Lumpur also from Wikipedia page:

https://en.wikipedia.org/wiki/Kuala_Lumpur

In this Wikipedia page there is just containing 1 table with the same information from Wikipedia page in Jakarta. Because the table contains some data that we do not need so we can keep the same information we got from table Jakarta.

3.3 Nominatim OpenStreetMap

The data scraping from Wikipedia page does not give information about the coordinates for each neighborhood. So, we can use Nominatim OpenStreetMap API in order to get *latitude* and *longitude* for each neighborhood.

Using Nominatim OpenStreetMap API in python we can use **geopy** library and import

geopy.geocoders.Nominatim package into notebook.

Using nominatim we can pass neighborhood keyword into nominatim object and get the representing latitude and longitude so we can add this information into neighborhood table for Jakarta and Kuala Lumpur.

3.4 Foursquare API

Foursquare is a company focusing on social media services. One of their products is Foursquare City Guide commonly called Foursquare is a product that give information about venues, places, or events within an area of interest. This app also provides personalized recommendations of places to go in near the user's current location based on other user's rating for the places. Using Foursquare API, we can find data about different venues for different neighborhood. With Foursquare API we can make a call containing neighborhood information so we can gain information about the places or venues.

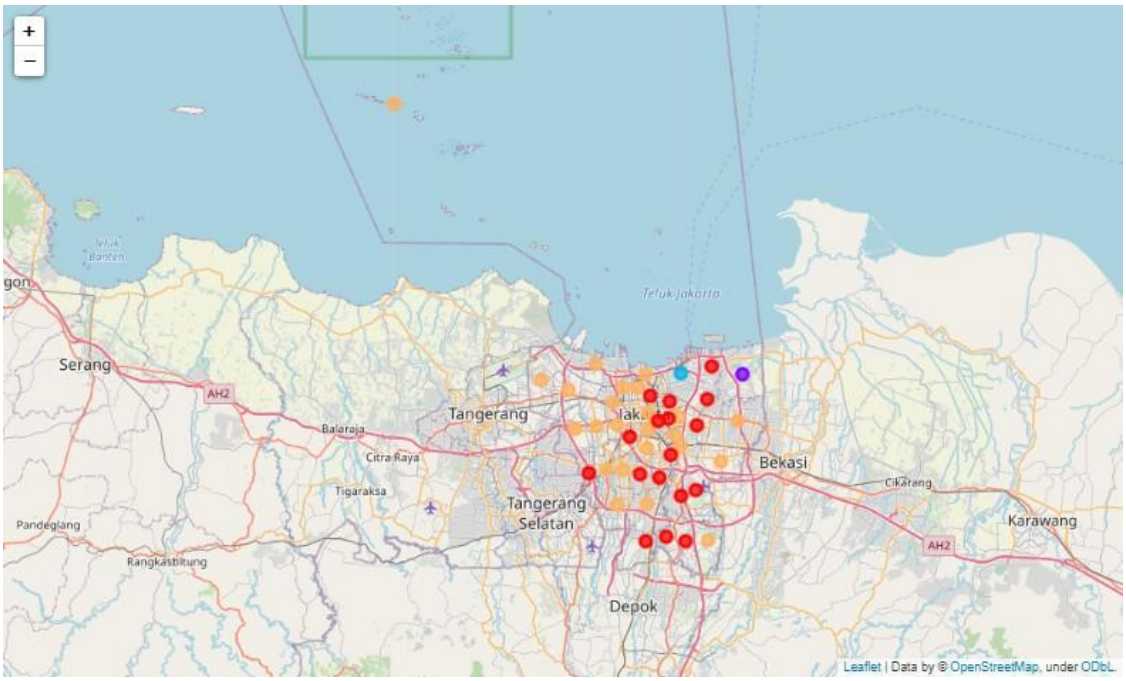
After using Foursquare API, we can find data about venues for each neighborhood and we can create a **Pandas Dataframe** object for information about Jakarta and Kuala Lumpur. After this, the information we obtained as follows:

1. *Neighborhood* : Name of kecamatan, we call this neighborhood to make it easy for report.
2. *Town* : Name of Administrative Town for each neighborhood.
3. *Latitude* : Latitude coordinates of the neighborhood.
4. *Longitude* : Longitude coordinates of the neighborhood.
5. *Venue* : Name of the venue.
6. *Venue Category* : Category of the venue.
7. *Venue Latitude* : Latitude coordinates of the venue.
8. *Venue Longitude* : Longitude coordinates of the venue.

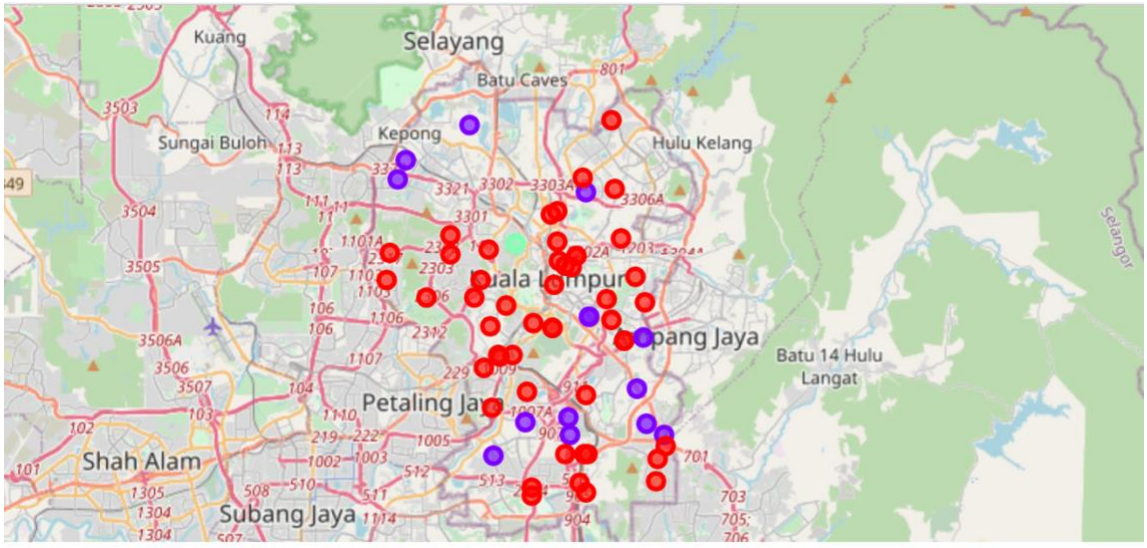
4 Methodology

In this part of the section, I will be collecting data (data scrapping) from Wikipedia page in order to get **neighborhood information** for **Jakarta** and **Kuala Lumpur**. After getting that information, I will use name of the neighborhood as a keyword to providing information about **neighborhood coordinates** (latitude and longitude) using **Nominatim** with *geopy.geocoders.Nominatim* pack- age. Using coordinates for each neighborhood I will use **Foursquare API** to get relevant venues and places near the given **latitude** and **longitude**. Using that information, we create a pandas dataframe to sort **5 most popular venues (categories) for each neighborhood.**

5.3.1 Jakarta Clusters



5.3.2 Kuala Lumpur Clusters



5 Results and Discussion

In this section we will see the clusters results from Jakarta Neighborhood and Kuala Lumpur Neighborhood.

5.1 Discussion

From the resulting cluster in Jakarta and Kuala Lumpur we can see that Food Place or Restaurant is the most common venue with the most neighborhood in the cluster. Although the area of Jakarta is much bigger than Kuala Lumpur and population of Jakarta is about 3 times than Kuala Lumpur, the neighborhood is relatively similar with the most common venue is the Restaurant and Bars. There is also different type for leisure and hangout because there are couple of Parks, Movies, Golf Course, Resort, even Soccer Stadium.

6 Conclusion

After we create a cluster for neighborhood in Jakarta and Kuala Lumpur, there is several clusters and segmentation based on venues and places from Foursquare API. But there is a cluster in Jakarta and Kuala Lumpur with restaurant being the first most common venue in the area. Back to the first question from beginning of these articles and you want to open new restaurant hopefully this will help you for consideration to decide if you want to open a restaurant in the neighborhood.

Link to original notebook : notebook