

# Battle of Neighborhoods - Clustering and Segmenting the Neighborhoods of Jakarta and Surabaya, Where is the Good Place to Open New Restaurant in the Area

December 15, 2020

## 1 Introduction

In this report i will analyze area of cluster between two major metropolitan area in Indonesia that is Jakarta and Surabaya. Jakarta and Surabaya are very popular and most populated metropolitan area in Indonesia. Although Jakarta have approximately 3 times population than Surabaya, Surabaya has it's 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 Surabaya 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 try to help tourist or business owner to open new detinations 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 simmilarity based on destination and places. This also will help people to make decision if they are want to migrate or move into another neighborhood.

## 3 Data Collecting

In this report, we require neighborhood (Kecamatan) for Jakarta and Surabaya. Using the location of the neighborhood we can search most popular venue or places for each categories using **Foursquare API**. We also need the coordinates/geographical location for each neighborhood in Jakarta and Surabaya. Using the coordinates of the neighborhood we can visualize with **OpenStreetMap using Folium API**.

### 3.1 Jakarta

In order to get neighborhood (Kecamatan) in Jakarta we scrape the data from : [https://id.wikipedia.org/wiki/Daftar\\_kecamatan\\_dan\\_kelurahan\\_di\\_Daerah\\_Khusus\\_Ibukota\\_Jakarta](https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Daerah_Khusus_Ibukota_Jakarta)

In this wikipedia page there is several table representing each Town in Jakarta. in each table, there is data about name of neighborhood (Kecamatan) for each town and name of villages (Kelurahan) for each neighborhood.

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

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

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

### 3.2 Surabaya

We scrape neighborhood data in Surabaya also from wikipedia page : [https://id.wikipedia.org/wiki/Daftar\\_kecamatan\\_dan\\_kelurahan\\_di\\_Kota\\_Surabaya](https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Kota_Surabaya)

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 Surabaya.

### 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 Surabaya. 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 collecting data (data scrapping) from wikipedia page in order to get **neighborhood information** for **Jakarta** and **Surabaya**. 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* package. 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**

### 4.0.1 Import library

Before we start collecting and processing data we want to import necessary library that we use in this research notebook.

```
# basic library
import pandas as pd
import requests
import numpy as np
import matplotlib.cm as cm
import matplotlib.colors as colors

# import folium library for map visualization
# !pip install folium # uncomment this if you have not already insalled folium
import folium

# Import Nominatim API from geopy.geocoders.Nominatim package for providing information about
# !pip install geopy # uncomment this if you have not already installed folium
from geopy.geocoders import Nominatim

# import k-means for the clustering stage
from sklearn.cluster import KMeans
```

## 4.1 Data Collection

### 4.1.1 Explore Jakarta

In this part i will do data wrangling from wikipedia page for providing neighborhood data and information in Jakarta URL : [https://id.wikipedia.org/wiki/Daftar\\_kecamatan\\_dan\\_kelurahan\\_di\\_Daerah\\_Khusus\\_Ibukota\\_Jakarta](https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Daerah_Khusus_Ibukota_Jakarta)

From this wikipedia page there is several table that we need so we can use **pandas.read\_html()** function to get a list of table that we need.

	Neighborhood	Town
0	Cempaka Putih	Central Jakarta
1	Gambir	Central Jakarta
2	Johar Baru	Central Jakarta
3	Kemayoran	Central Jakarta
4	Menteng	Central Jakarta

After processing data here is the head of resulting dataframe that we will use for now.

#### 4.1.2 Explore Surabaya

Here are the URL that we will use for data scrapping, URL : [https://id.wikipedia.org/wiki/Daftar\\_kecamatan\\_dan\\_kelurahan\\_di\\_Kota\\_Surabaya](https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Kota_Surabaya)

The approach to get data is pretty much the same from what i did with Jakarta Neighborhood

	Neighborhood	Town
0	Asemrowo	Surabaya
1	Benowo	Surabaya
2	Bubutan	Surabaya
3	Bulak	Surabaya
4	Dukuh Pakis	Surabaya

After processing data here is the head of resulting dataframe that we will use for now.

## 4.2 Nominatim OpenStreetMap API

To get information about latitude and longitude for each neighborhood in Jakarta and Surabaya we can use Nominatim from **geopy.geocoders.Nominatim** package to provide coordinates passing neighborhood keyword as an argument.

First we create Nominatim object. Here is the script code :

```
# Create Nominatim object as 'geolocator'
geolocator = Nominatim(user_agent='explorer')
```

To get latitude and longitude, for each neighborhood i will define some function to apply to corresponding dataframe.

Script function :

```
# All of these function will provide information about latitude and longitude for neighborhood
def get_latitude_jakarta(neighborhood):
    location = geolocator.geocode(f'{neighborhood}, Jakarta, Indonesia')
```

```

latitude = location.latitude
return latitude

def get_longitude_jakarta(neighborhood):
    location = geolocator.geocode(f'{neighborhood}, Jakarta, Indonesia')
    longitude = location.longitude
    return longitude

def get_latitude_surabaya(neighborhood):
    location = geolocator.geocode(f'{neighborhood}, Surabaya, Indonesia')
    latitude = location.latitude
    return latitude

def get_longitude_surabaya(neighborhood):
    location = geolocator.geocode(f'{neighborhood}, Surabaya, Indonesia')
    longitude = location.longitude
    return longitude

```

After we define some of the function, then we can apply to the dataframe given the keyword of the name of neighborhood.

Here are the updated dataframes. Jakarta

	Neighborhood	Town	Latitude	Longitude
0	Cempaka Putih	Central Jakarta	-6.181214	106.868548
1	Gambir	Central Jakarta	-6.176684	106.830653
2	Johar Baru	Central Jakarta	-6.183125	106.855332
3	Kemayoran	Central Jakarta	-6.162546	106.856890
4	Menteng	Central Jakarta	-6.195026	106.832224

Surabaya

	Neighborhood	Town	Latitude	Longitude
0	Cempaka Putih	Central Jakarta	-6.181214	106.868548
1	Gambir	Central Jakarta	-6.176684	106.830653
2	Johar Baru	Central Jakarta	-6.183125	106.855332
3	Kemayoran	Central Jakarta	-6.162546	106.856890
4	Menteng	Central Jakarta	-6.195026	106.832224

### 4.3 Map Visualize

Visualizing map using Folium API with OpenStreetMap view with information of neighborhood from both dataframes

### 4.3.1 Jakarta Neighborhood Map View

To visualize OpenStreetMap of Jakarta, first we need to get coordinates for Jakarta. Then we apply each neighborhood coordinates to for mark a circular area representing each neighborhood.

Here are the script for getting coordinates of Jakarta

```
address = 'Jakarta'

location = geolocator.geocode(address)
jakarta_latitude = location.latitude
jakarta_longitude = location.longitude
print(f'Coordinates of Jakarta are {jakarta_latitude}, {jakarta_longitude}')
```

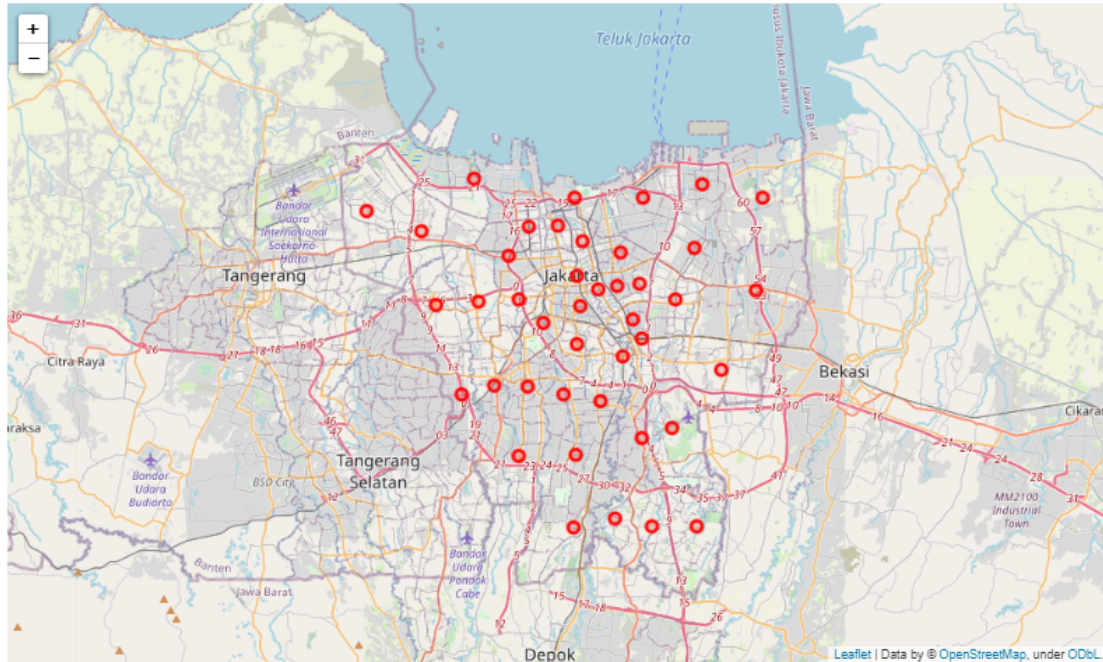
Scrip for map visualizing using Folium python library :

```
jakarta_map = folium.Map(location=[jakarta_latitude, jakarta_longitude], zoom_start=11)

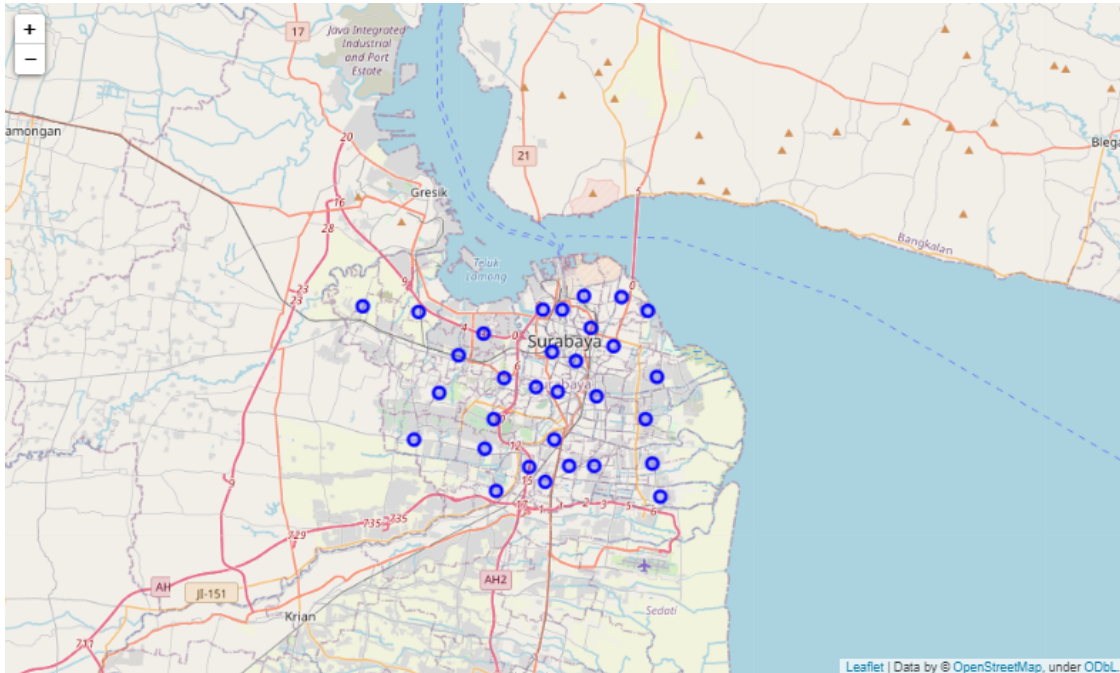
for latitude, longitude, borough, neighborhood in zip(jakarta_df['Latitude'], jakarta_df['Longitude'], jakarta_df['Borough'], jakarta_df['Neighborhood']):
    label = '{} , {}'.format(neighborhood, borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [latitude, longitude],
        radius=5,
        popup=label,
        color='red',
        fill=True
    ).add_to(jakarta_map)

# Show the map
jakarta_map
```

#### 4.3.2 Here are the map for Jakarta :



#### 4.3.3 We do the same for visualizing Surabaya OpenStreetMap :



#### 4.4 Foursquare API

Defining Foursquare API Credentials and Version. You can get your Foursquare credentials by signing up to Foursquare Developer, More : [Foursquare Developer API](#)

We create a function to get nearby venues and places around the neighborhood. Here are the script function :

```
# Function that return latitude, longitude, venues, and venue_categories in neighborhood_df
def get_nearby_venues(names, latitudes, longitudes, radius=500):

    # create an empty list
    venues_list=[]

    # for loop that iterate through dataframe
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}'
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius
            )

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['categories'][0]['name']) for v in results])

    # Create pandas dataframe from venues_list
    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['Neighborhood',
                            'Neighborhood Latitude',
                            'Neighborhood Longitude',
                            'Venue',
                            'Venue Category']

    return(nearby_venues)
```

Then, apply it to the dataframe :

```
jakarta_venues = get_nearby_venues(jakarta_df['Neighborhood'], jakarta_df['Latitude'], jakarta,
```

Here are the resulting dataframe for Jakarta and Surabaya.



## Jakarta DataFrame

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Category
0	Cempaka Putih	-6.181214	106.868548	Mie Aceh Bungong Cempaka	Acehnese Restaurant
1	Cempaka Putih	-6.181214	106.868548	Arcici Swiming Pool™	Pool
2	Cempaka Putih	-6.181214	106.868548	Pizza Hut	Pizza Place
3	Cempaka Putih	-6.181214	106.868548	Pizza Hut	Pizza Place
4	Cempaka Putih	-6.181214	106.868548	Bebek Bentu	BBQ Joint

## Surabaya DataFrame

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Category
0	Asemrowo	-7.241740	112.688802	CANTEEN	Wine Bar
1	Benowo	-7.229055	112.649775	Benowo Trade Centre [ BTC ], Benowo, Surabaya	Shoe Store
2	Benowo	-7.229055	112.649775	Stadion GBT Benowo	Soccer Field
3	Benowo	-7.229055	112.649775	Pecel B.Yatin ketabang kali	Food Court
4	Bubutan	-7.252671	112.730062	CGV Cinemas	Multiplex

## 4.5 One Hot Encoding

In order to find top 5 most common venue, we need to transform each categorical data into number with One Hot Encoding using `pandas.get_dummies()` function. After get the one hot encoded dataframe, then we can count the average for each venue in each neighborhood. Here are the result after apply the average most common venue in neighborhood.

### 4.5.1 One hot encoding for Jakarta venues

Neighborhood	Wings Joint	Accessories Store	Acehnese Restaurant	Airport Terminal	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	...	Track	Track Stadium	Trail	Train	Train Station	Udo Restaurar
0	Cakung	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.
1	Cempaka Putih	0.0	0.0	0.166667	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.
2	Cengkareng	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.
3	Cilandak	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.
4	Cilincing	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.

5 rows × 141 columns

### 4.5.2 One hot encoding for Surabaya venues

Neighborhood	Arcade	Asian Restaurant	Australian Restaurant	Bakery	Balinese Restaurant	Basketball Court	Batik Shop	Beach	Bed & Breakfast	...	Soccer Field	Soccer Stadium	Soup Place	Sporting Goods Shop	Steakhouse
0	Asemrowo	0.0	0.0	0.0	0.0	0.0	0.0	0.000	0.0	...	0.000000	0.0	0.0	0.0	0
1	Benowo	0.0	0.0	0.0	0.0	0.0	0.0	0.000	0.0	...	0.333333	0.0	0.0	0.0	0
2	Bubutan	0.0	0.0	0.0	0.0	0.0	0.0	0.000	0.0	...	0.000000	0.0	0.0	0.0	0
3	Bulak	0.0	0.0	0.0	0.0	0.0	0.0	0.375	0.0	...	0.000000	0.0	0.0	0.0	0
4	Dukuh Pakis	0.0	0.0	0.0	0.0	0.0	0.0	0.000	0.0	...	0.000000	0.0	0.0	0.0	0

5 rows × 77 columns

### 4.5.3 Find 5 most common venue

To find 5 most common venue in area we can use loop to see what is the most venue categories in neighborhood.

Script :

```
num_top_venues = 5

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))
```

Then apply it to create new dataframe for Jakarta and Surabaya. Script for Jakarta :

```
# create a new dataframe
jakarta_neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
jakarta_neighborhoods_venues_sorted['Neighborhood'] = jakarta_grouped['Neighborhood']

for ind in np.arange(jakarta_grouped.shape[0]):
    jakarta_neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(jakarta_grouped, ind+1, num_top_venues)
```

#### *Jakarta DataFrame*

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Cakung	Lounge	Gas Station	Wine Bar	French Restaurant	Food Stand
1	Cempaka Putih	Pizza Place	Acehnese Restaurant	Pool	Indonesian Meatball Place	BBQ Joint
2	Cengkareng	Music Venue	Pet Store	Restaurant	Movie Theater	Electronics Store
3	Cilandak	Gym	Indonesian Restaurant	Food Truck	Convenience Store	Pizza Place
4	Cilincing	Park	Diner	Shopping Mall	Wine Bar	Farmers Market

#### *Surabaya DataFrame*

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Asemrowo	Wine Bar	Fish & Chips Shop	Dessert Shop	Diner	Donut Shop
1	Benowo	Shoe Store	Food Court	Soccer Field	Wine Bar	Dessert Shop
2	Bubutan	Indonesian Restaurant	Hotel	Multiplex	Donut Shop	Pizza Place
3	Bulak	Beach	Indonesian Restaurant	Food Truck	Fish & Chips Shop	Dessert Shop
4	Dukuh Pakis	Convenience Store	Mobile Phone Shop	Boutique	Fast Food Restaurant	Diner

## 5 Modeling

After we get data about top 10 most common venue for each neighborhood in Jakarta and Surabaya we can begin create a clustering model using **K-Means Clustering** library from Scikit-Learn

We will run the K-Means Clustering to cluster and segment the neighborhood into 5 different clusters based on type of venues and places.

```
# set number of clusters
kclusters = 5

# instantiate kmeans model
kmeans = KMeans(n_clusters=kclusters, random_state=0)
```

## 5.1 Prepare the data (features) for modeling

We will use grouped dataframe for Jakarta and Surabaya that is containing values of one hot encoded venues and places and drop 'Neighborhood' column that is contain Neighborhood name (string dtypes)

## 5.2 Begin modeling

### 5.2.1 Clustering in Jakarta Neighborhood

```
# fit the data
jakarta_kmmeans = kmeans.fit(jakarta_cluster)

# add clustering labels
jakarta_neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

jakarta_merged = jakarta_df

# merge jakarta_grouped with neighborhood_df to add latitude/longitude for each neighborhood
jakarta_merged = jakarta_merged.join(jakarta_neighborhoods_venues_sorted.set_index('Neighborhood', drop=True))

jakarta_merged.head()
```

	Neighborhood	Town	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Cempaka Putih	Central Jakarta	-6.181214	106.868548	3	Pizza Place	Acehnese Restaurant	Pool	Indonesian Meatball Place	BBQ Joint
1	Gambir	Central Jakarta	-6.176684	106.830653	3	Indonesian Restaurant	Park	Coffee Shop	Fast Food Restaurant	Food Court
2	Johar Baru	Central Jakarta	-6.183125	106.855332	4	Food Truck	Convenience Store	Arcade	Indonesian Restaurant	Italian Restaurant
3	Kemayoran	Central Jakarta	-6.162546	106.856890	4	Diner	Arcade	Indonesian Restaurant	Noodle House	Wine Bar
4	Menteng	Central Jakarta	-6.195026	106.832224	3	Indonesian Restaurant	Breakfast Spot	Coffee Shop	Park	Sushi Restaurant

The resulting dataframe :

### 5.2.2 Clustering in Surabaya Neighborhood

For clustering in Surabaya, we use the same script we use for Jakarta using dataframe from neighborhood in Surabaya.

	Neighborhood	Town	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Asemrowo	Surabaya	-7.241740	112.688802	4.0	Wine Bar	Fish & Chips Shop	Dessert Shop	Diner	Donut Shop
1	Benowo	Surabaya	-7.229055	112.649775	1.0	Shoe Store	Food Court	Soccer Field	Wine Bar	Dessert Shop
2	Bubutan	Surabaya	-7.252671	112.730062	1.0	Indonesian Restaurant	Hotel	Multiplex	Donut Shop	Pizza Place
3	Bulak	Surabaya	-7.228354	112.787631	1.0	Beach	Indonesian Restaurant	Food Truck	Fish & Chips Shop	Dessert Shop
4	Dukuh Pakis	Surabaya	-7.293024	112.695125	0.0	Convenience Store	Mobile Phone Shop	Boutique	Fast Food Restaurant	Diner

The resulting dataframe :

**5.2.3** The column ‘Cluster Labels’ indicates the resulting cluster for each neighborhood in Jakarta and Surabaya

### 5.3 Visualizing the clusters

Using OpenStreetMap from Folium we can create a cluster from coordinates of the neighborhood with different color for each color representing the cluster.

Here are the script for Jakarta Clusters and with using different dataframe we also use it for Surabaya Clusters :

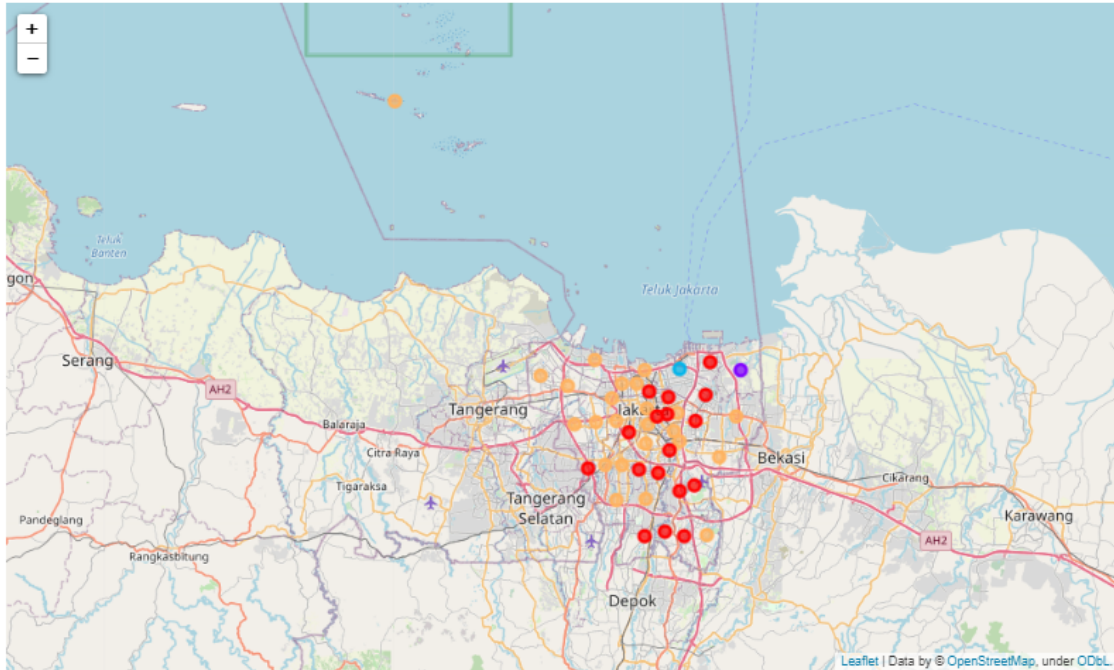
```
# create map
jakarta_map_clusters = folium.Map(location=[jakarta_latitude, jakarta_longitude], zoom_start=13)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

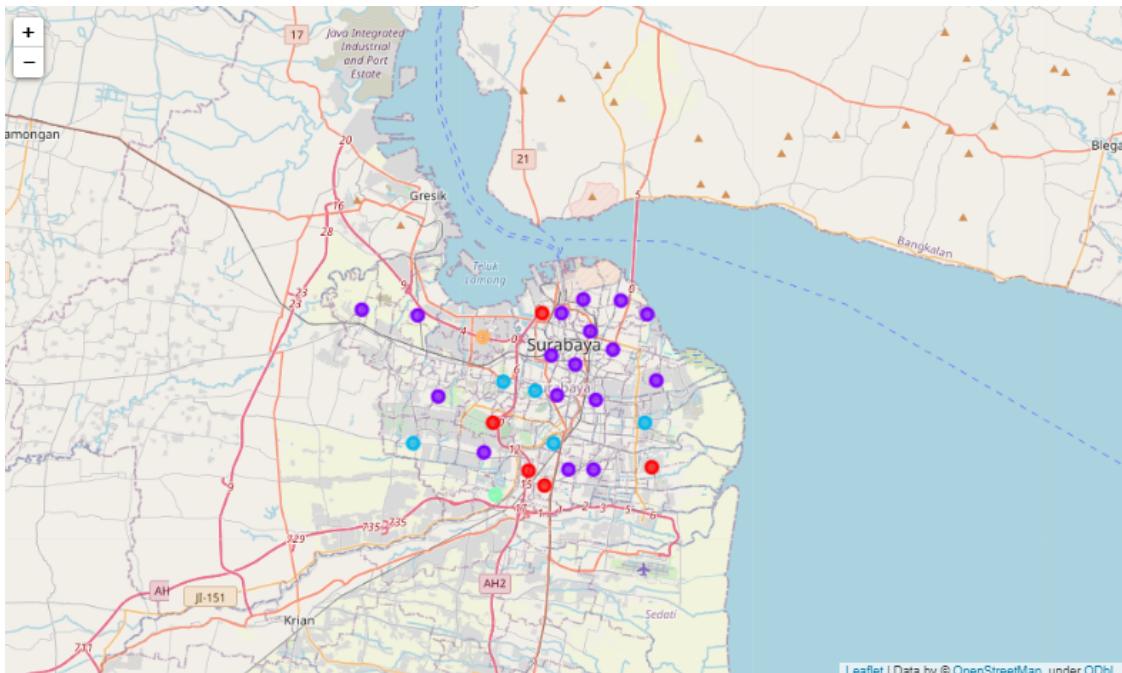
# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(jakarta_merged['Latitude'], jakarta_merged['Longitude'], jakarta_merged['poi'], jakarta_merged['cluster']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[int(cluster)],
        fill=True,
        fill_color=rainbow[int(cluster)],
        fill_opacity=0.7).add_to(jakarta_map_clusters)

# Show the map
jakarta_map_clusters
```

### 5.3.1 Jakarta Clusters



### 5.3.2 Surabaya Clusters



## 6 Results and Discussion

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

## 6.1 Results in Jakarta Neighborhood

- **Cluster 1** : In this cluster we can see this area most common venue is convenience store because of the result being first and second most common venue in the area.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
9 Cilincing	Park	Diner	Shopping Mall	Wine Bar	Farmers Market

- **Cluster 2** : In this cluster there is also only one neighborhood with the most common venue being some kind of Desert Places in the area.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
14 Tanjung Priok	Bakery	Wine Bar	Frozen Yogurt Shop	French Restaurant	Food Truck

- **Cluster 3** : In this cluster also resulting only one neighborhood with Resort as the most common venue in the area.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
47 Kepulauan Seribu Utara	Resort	Wine Bar	Electronics Store	Food Stand	Food Court

- **Cluster 4** : In this cluster we can see there is more Neighborhood with in the cluster. The most common venue in this area is pretty diverse we can see from the variety of restaurant like Japanese, Chinese, Vegan, Mediterranean, Asian, and other type of Restaurant. With the variety of places and venues being this is a good place for a hangout because venues like Restaurant, Coffee Shop, Theme Park, Movie, or Clothing Store.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Cempaka Putih	Pizza Place	Acehnese Restaurant	Pool	Indonesian Meatball Place	BBQ Joint
1	Gambir	Indonesian Restaurant	Park	Coffee Shop	Fast Food Restaurant	Food Court
4	Menteng	Indonesian Restaurant	Breakfast Spot	Coffee Shop	Park	Sushi Restaurant
12	Pademangan	Hotel	Pool	Bowling Alley	Theme Park	Vegetarian / Vegan Restaurant
13	Penjaringan	Boutique	Theme Park	Gift Shop	Grocery Store	Food Truck
16	Cakung	Lounge	Gas Station	Wine Bar	French Restaurant	Food Stand
17	Cipayung	Food	Asian Restaurant	Wine Bar	Frozen Yogurt Shop	French Restaurant
19	Duren Sawit	Indonesian Meatball Place	Convenience Store	Coffee Shop	Mediterranean Restaurant	Wine Bar
20	Jatinegara	Jewelry Store	Asian Restaurant	Arts & Crafts Store	Donut Shop	Food Truck
23	Matraman	Convenience Store	Pizza Place	Electronics Store	Food Stand	Food Court
27	Cilandak	Gym	Indonesian Restaurant	Food Truck	Convenience Store	Pizza Place
29	Kebayoran Baru	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Food Truck	Hotel
30	Kebayoran Lama	Coffee Shop	Bakery	Clothing Store	Shopping Mall	Satay Restaurant
33	Pasar Minggu	Japanese Restaurant	Asian Restaurant	Bus Station	Music Store	Fast Food Restaurant
35	Setiabudi	Coffee Shop	Hotel	Pizza Place	Vietnamese Restaurant	Café
38	Cengkareng	Music Venue	Pet Store	Restaurant	Movie Theater	Electronics Store
39	Grogol Petamburan	Noodle House	Chinese Restaurant	Indonesian Restaurant	Hotel	Asian Restaurant
40	Taman Sari	Chinese Restaurant	Asian Restaurant	Hotel	Karaoke Bar	Restaurant
41	Tambora	Food Truck	Convenience Store	Tanning Salon	Department Store	Train Station
42	Kebon Jeruk	Noodle House	Concert Hall	Asian Restaurant	Park	Indonesian Restaurant
43	Kalideres	Convenience Store	Indonesian Restaurant	Soccer Field	Fried Chicken Joint	Asian Restaurant
44	Palmerah	Pizza Place	Asian Restaurant	Noodle House	Coffee Shop	Medical Center
45	Kembangan	Pizza Place	Asian Restaurant	Park	Indonesian Meatball Place	Bubble Tea Shop
48	Kepulauan Seribu Selatan	Boat or Ferry	Clothing Store	Fast Food Restaurant	French Restaurant	Food Truck

- **Cluster 5** : In this cluster we can see most of the neighborhood have Indonesian Restaurant as the first most common venue. We can say that this cluster an area for having a meal because the variety of the restaurant places. That being said, if you want to open new restaurant in

these neighborhood this is the area that has more restaurant variety and possibly being one of the competitors for your restaurant.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	Johar Baru	Food Truck	Convenience Store	Arcade	Indonesian Restaurant	Italian Restaurant
3	Kemayoran	Diner	Arcade	Indonesian Restaurant	Noodle House	Wine Bar
5	Sawah Besar	Indonesian Restaurant	Pet Store	Convenience Store	Noodle House	Asian Restaurant
6	Senen	Hotel	Indonesian Restaurant	Grocery Store	History Museum	University
7	Tanah Abang	Indonesian Restaurant	Coffee Shop	Seafood Restaurant	Pizza Place	Noodle House
10	Kelapa Gading	Indonesian Restaurant	Asian Restaurant	Steakhouse	Korean Restaurant	Japanese Restaurant
11	Koja	Pizza Place	Indonesian Restaurant	Restaurant	Grocery Store	Donut Shop
18	Ciracas	Noodle House	Indonesian Restaurant	Golf Course	Farmers Market	Food Stand
21	Kramat Jati	Hospital	Indonesian Restaurant	Seafood Restaurant	Chinese Restaurant	Noodle House
22	Makasar	Indonesian Restaurant	Airport Terminal	Asian Restaurant	Wine Bar	Farmers Market
24	Pasar Rebo	Karaoke Bar	Burger Joint	Bakery	Indonesian Restaurant	Grocery Store
25	Pulo Gadung	Indonesian Restaurant	Convenience Store	Pizza Place	Café	Athletics & Sports
28	Jagakarsa	Indonesian Restaurant	Asian Restaurant	Food Court	Arcade	Gym
31	Mampang Prapatan	Indonesian Restaurant	Noodle House	Park	Electronics Store	Food Stand
32	Pancoran	Indonesian Restaurant	Padangnese Restaurant	Asian Restaurant	Fast Food Restaurant	Restaurant
34	Pesangrahan	Convenience Store	Pier	Indonesian Restaurant	Gym	Restaurant
36	Tebet	Indonesian Restaurant	Café	Bakery	Asian Restaurant	Restaurant

## 6.2 Results in Surabaya Neighborhood

- **Cluster 1** : In this cluster we can see this area most common venue is convenience store because of the result being first and second most common venue in the area.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Dukuh Pakis	Convenience Store	Mobile Phone Shop	Boutique	Fast Food Restaurant	Diner
5	Gayungan	Asian Restaurant	Convenience Store	Hotel	Café	Boutique
9	Jambangan	Coffee Shop	Convenience Store	Asian Restaurant	Food Truck	Fish & Chips Shop
12	Krembangan	Convenience Store	Farmers Market	Museum	Coffee Shop	Food Truck
17	Rungkut	Convenience Store	Coffee Shop	Grocery Store	Fast Food Restaurant	Dessert Shop

- **Cluster 2** : In this cluster we can see this area is pretty popular for restaurant and other food places because of variety of the restaurant and other food places. This also being the cluster that if you want to open a new restaurant there is already a variety of restaurant in the neighborhood.



	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Benowo	Shoe Store	Food Court	Soccer Field	Wine Bar	Dessert Shop
2	Bubutan	Indonesian Restaurant	Hotel	Multiplex	Donut Shop	Pizza Place
3	Bulak	Beach	Indonesian Restaurant	Food Truck	Fish & Chips Shop	Dessert Shop
6	Genteng	Soup Place	Furniture / Home Store	Park	Breakfast Spot	Bed & Breakfast
7	Gubeng	Indonesian Restaurant	Food Truck	Electronics Store	Chinese Restaurant	Café
11	Kenjeran	Food Court	Pizza Place	Mosque	Mobile Phone Shop	Wine Bar
14	Mulyorejo	Indonesian Restaurant	Coffee Shop	Convenience Store	Pizza Place	Mobile Phone Shop
15	Pabean Cantian	Indonesian Restaurant	History Museum	Food Truck	Café	Farmers Market
16	Pakal	Campground	Wine Bar	Fish & Chips Shop	Diner	Donut Shop
18	Sambikerep	Food Truck	Indonesian Restaurant	Wine Bar	Fast Food Restaurant	Dessert Shop
20	Semampir	Noodle House	Food	Wine Bar	Fast Food Restaurant	Dessert Shop
21	Simokerto	Seafood Restaurant	Basketball Court	Arcade	Coffee Shop	Market
24	Tambaksari	Convenience Store	Indonesian Restaurant	Asian Restaurant	Bakery	Noodle House
26	Tegalsari	Convenience Store	Indonesian Restaurant	Pool Hall	Noodle House	Convention Center
27	Tenggilis Mejoyo	Indonesian Restaurant	Food Truck	Grocery Store	Sporting Goods Shop	Hotel
28	Wiyung	Arcade	Diner	Hospital	Café	Breakfast Spot
29	Wonocolo	Convenience Store	Bakery	Café	Dumpling Restaurant	Pool Hall

- **Cluster 3** : In this cluster we have the area that is popular for cafe and coffee shop. This neighborhood is perfect for people to hangout or just to enjoy a cup of coffee.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
13	Lakarsantri	Café	Wine Bar	Fish & Chips Shop	Diner	Donut Shop
19	Sawahan	Karaoke Bar	Café	Seafood Restaurant	Convention Center	Diner
22	Sukolilo	Café	Supermarket	Steakhouse	Gym / Fitness Center	Asian Restaurant
23	Sukomanunggal	Noodle House	Café	Hardware Store	Soup Place	Juice Bar
30	Wonokromo	Coffee Shop	Café	Wine Bar	Convention Center	Diner

- **Cluster 4** : In this cluster there is only one neighborhood with Soccer Stadium as the first most common venue.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
10	Karang Pilang	Soccer Stadium	Wine Bar	Fish & Chips Shop	Dessert Shop	Diner

- **Cluster 5** : In this cluster there is also just one neighborhood with Wine Bar as the first most common venue in the area.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Asemrowo	Wine Bar	Fish & Chips Shop	Dessert Shop	Diner	Donut Shop

### 6.3 Discussion

From the resulting cluster in Jakarta and Surabaya 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 Surabaya and population of Jakarta is about 3 times than Surabaya, 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.



## 7 Conclusion

After we create a cluster for neighborhood in Jakarta and Surabaya, there is several clusters and segmentation based on venues and places from Foursquare API. But there is a cluster in Jakarta and Surabaya with restaurant being the first most common venue in the area. Back to the first question from beginning of these article 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](#)

Thanks for reading this article that i have made. Any suggestion or feedback will be appreciated. If you want to reach me, feel free sent a message trough linkedin and if you want to see more of my portfolio you can see trough my github account.

- [Linkedin](#)
- [GitHub](#)

## 8 References :

- [Coursera Applied Data Science Capstone Course](#)

### 8.1 Thanks to :

- [Foursquare Developer API](#)
- [Indonesian Ministry of Internal Affairs \(Kemendagri\)](#) Accessed via Wikipedia page [neighborhood Jakarta](#) and [neighborhood Surabaya](#).
- Python library package (pandas, numpy, matplotlib, folium, scikit-learn, and geopy)

### 8.2 Tools :

- Jupyter Notebooks using Visual Studio Code
- GitHub (version control)

Muhammad Adisatriyo Pratama - December 2020