



# Segmenting and Clustering Neighborhoods in Jakarta City

## Introduction

In this project, we are going to analyse the segmenting and clustering neighborhoods in [Jakarta](#) (Capital City of Indonesia). We are going to use ready data from [Biro Pusat Statistik](#) (Statistics Indonesia), convert addresses into their equivalent latitude and longitude values. Also, we will use the [Foursquare API](#) to explore neighborhoods. We will use the **explore** function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. We will use the *k*-means clustering algorithm to complete this task with Silhouette and Elbow Method to determine the number of clusters. We will use the Folium library to visualize the neighborhoods in Jakarta and their emerging clusters. Finally once each of neighbourhood is recognized, then we will propose the possible business to be done based on that Foursquare Venues. This information will be used by an investor to determine correct investment for each neighborhood in Jakarta.

## Table of Contents

1. Download and Explore Dataset 2. Explore Neighborhoods in Jakarta City 3. Analyze Each Neighborhood 4. Cluster Neighborhoods 5. Examine Clusters

Before we get the data and start exploring it, let's download all the dependencies that we will need.

```
In [35]: import numpy as np # Library to handle data in a vectorized manner
import pandas as pd # Library for data analysis

pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # Library to handle JSON files

#!conda install -c conda-forge geopy --yes # uncomment this line if you haven't comp
from geopy.geocoders import Nominatim # convert an address into latitude and longitu

import requests # Library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe
#from pandas.json_normalize import json_normalize

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors
```

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

#!conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you haven
import folium # map rendering library

print('Libraries imported.')
```

Libraries imported.

## 1. Download and Explore Dataset

Jakarta has a total of 6 Kabupaten/Kota ([Regency](#)) and 44 Kecamatan ([District](#)). In order to segment the Kabupaten and explore them, we will essentially need a dataset that contains the 5 Kabupaten and Kecamatan that exist in each Kabupaten as well as the the latitude and logitude coordinates of each Kabupaten.

**Tips:** We got the data from [BPS](#) and save it into file in server to avoid frequent access to Goverment website :)

```
In [154...] data = pd.read_csv("data-jumlah-kecamatan-2019.csv")
column_names = ['Kabupaten', 'Kecamatan', 'Luas']
data.columns = column_names
data.head()
```

```
Out[154...]
Kabupaten      Kecamatan  Luas
0  JAKARTA BARAT  CENGKARENG  26.55
1  JAKARTA BARAT  GROGOL PETAMBURAN  9.99
2  JAKARTA BARAT      KALI DERES  30.23
3  JAKARTA BARAT  KEBON JERUK  17.63
4  JAKARTA BARAT  KEMBANGAN  24.17
```

Thanks to geolocator Nominatim we will get the equivalent of address

```
In [149...] geolocator = Nominatim(user_agent="ny_explorer")
coordinate = []

for index, row in data.iterrows():
    address = row['Kecamatan'] #+ ", " + row['Kabupaten']
    location = geolocator.geocode(address)
    if location is not None:
        coordinate.append([location.latitude, location.longitude])
    else:
        coordinate.append([null, null])

df_geo = pd.DataFrame(coordinate, columns=['Latitude', 'Longitude'])
neighborhoods=pd.merge(data, df_geo, left_index=True, right_index=True)
```

**Tips:** Geolocator tends to consume time, hence we store the data to avoid retrieving the data from internet for further references

```
In [150...] neighborhoods.to_csv(r'..\Coursera_Capstone\neighborhoods.csv')
```

```
In [155...] neighborhoods = pd.read_csv("neighborhoods.csv")
neighborhoods = neighborhoods.drop('Unnamed: 0', 1)
```

```
In [156... neighborhoods.shape
```

```
Out[156... (44, 5)
```

Quickly examine the resulting dataframe and we will have columns as below:

```
In [157... neighborhoods.head()
```

```
Out[157... 
```

	Kabupaten	Kecamatan	Luas	Latitude	Longitude
0	JAKARTA BARAT	CENGKARENG	26.55	-6.149093	106.734781
1	JAKARTA BARAT	GROGOL PETAMBURAN	9.99	-6.164188	106.788317
2	JAKARTA BARAT	KALI DERES	30.23	-6.134300	106.705800
3	JAKARTA BARAT	KEBON JERUK	17.63	-6.192572	106.769725
4	JAKARTA BARAT	KEMBANGAN	24.17	-6.193000	106.742600

```
In [158... print('As mentioned in the introduction The dataframe has {} Kabupaten and {} Kecama
        len(neighborhoods['Kabupaten'].unique()),
        neighborhoods.shape[0]
    )
)
```

As mentioned in the introduction The dataframe has 6 Kabupaten and 44 Kecamatan.

## Use geopy library to get the latitude and longitude values of Jakarta.

In order to define an instance of the geocoder, we need to define a user\_agent. We will name our agent *ny\_explorer*, as shown below.

```
In [44]: address = 'Jakarta'

geolocator = Nominatim(user_agent="ny_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Jakarta are {}, {}'.format(latitude, longitude))
```

The geograpical coordinate of Jakarta are -6.1753942, 106.827183.

## Create a map of Jakarta with neighborhoods superimposed on top.

```
In [45]: # create map of New York using Latitude and Longitude values
map_Jakarta = folium.Map(location=[latitude, longitude], zoom_start=10)

# add markers to map
for lat, lng, borough, neighborhood in zip(neighborhoods['Latitude'], neighborhoods[
    label = '{} {}'.format(neighborhood, borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_Jakarta)

map_Jakarta
```

Out[45]:

+  
—

Leaflet (<https://leafletjs.com>) | Data by © OpenStreetMap (<http://openstreetmap.org>), under ODbL (<http://www.openstreetmap.org/copyright>).

**Folium** is a great visualization library. Feel free to zoom into the above map, and click on each circle mark to reveal the name of Kecamatan and respective Kabupaten

Next, we are going to start utilizing the Foursquare API to explore the neighborhoods and segment them.

## Define Foursquare Credentials and Version

```
In [46]: CLIENT_ID = 'WAFDIUY5OL25UM44YNMRQWZNT5OQ2CVL0RJD4QE4IHFJYS4R' # your Foursquare ID
CLIENT_SECRET = 'X044CFGQSJCIGKLPD2XNZREII0FISRGRZ0CIQQKDRSLQ0BSS' # your Foursquare
VERSION = '20180605' # Foursquare API version

#CLIENT_ID = 'your-client-ID' # your Foursquare ID
#CLIENT_SECRET = 'your-client-secret' # your Foursquare Secret
#VERSION = '20180605' # Foursquare API version
LIMIT = 100 # A default Foursquare API limit value

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET: ' + CLIENT_SECRET)
```

Your credentails:

CLIENT\_ID: WAFDIUY5OL25UM44YNMRQWZNT5OQ2CVL0RJD4QE4IHFJYS4R

CLIENT\_SECRET: X044CFGQSJCIGKLPD2XNZREII0FISRGRZ0CIQQKDRSLQ0BSS

## Let's explore the first Kecamatan in our dataframe.

Let's try and get the Kecamatan's name.

```
In [166]: jakarta_data=neighborhoods.reset_index(drop=True)
```

```
In [167]: print('Get latitude and longitude values of Kecamatan {}'.format(jakarta_data.loc[0]
```

Get latitude and longitude values of Kecamatan CENGKARENG.

```
In [168]: neighborhood_latitude = jakarta_data.loc[0, 'Latitude'] # neighborhood Latitude valu
neighborhood_longitude = jakarta_data.loc[0, 'Longitude'] # neighborhood Longitude v

neighborhood_name = jakarta_data.loc[0, 'Kecamatan'] # neighborhood name

print('Latitude and longitude values of {} are {}, {}'.format(neighborhood_name,
```

```
neighborhood_latitude  
neighborhood_longitude
```

Latitude and longitude values of CENGKARENG are -6.1490933, 106.73478100000001.

**Now, let's get the top 100 venues that are within a radius of 500 meters.**

First, let's create the GET request URL. Name your URL **url**.

```
In [169... # type your answer here  
LIMIT = 100 # limit of number of venues returned by Foursquare API  
  
radius = 500 # define radius  
  
# create URL  
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v=  
CLIENT_ID,  
CLIENT_SECRET,  
VERSION,  
neighborhood_latitude,  
neighborhood_longitude,  
radius,  
LIMIT)  
url # display URL
```

```
Out[169... 'https://api.foursquare.com/v2/venues/explore?&client_id=WAFDIUY5OL25UM44YNMRQWZNT5O  
Q2CVL0RJ4QE4IHFJYS4R&client_secret=X044CFGQSJCIGKLPD2XNZREII0FISRGRZ0CIQQKDRSLQ0BSS  
&v=20180605&ll=-6.1490933,106.73478100000001&radius=500&limit=100'
```

Send the GET request and examine the results

```
In [170... results = requests.get(url).json()  
results
```

```
Out[170... {'meta': {'code': 200, 'requestId': '6027558741cc7e0fe9587457'},  
'response': {'headerLocation': 'Cengkareng',  
'headerFullLocation': 'Cengkareng, Jakarta',  
'headerLocationGranularity': 'neighborhood',  
'totalResults': 6,  
'suggestedBounds': {'ne': {'lat': -6.1445932954999956,  
'lng': 106.73929859432982},  
'sw': {'lat': -6.153593304500004, 'lng': 106.7302634056702}},  
'groups': [{ 'type': 'Recommended Places',  
'name': 'recommended',  
'items': [{ 'reasons': {'count': 0,  
'items': [{ 'summary': 'This spot is popular',  
'type': 'general',  
'reasonName': 'globalInteractionReason'} ]}],  
'venue': {'id': '50fb9af9e4b0c2329c07e0ca',  
'name': 'XXI Puri Indah Mall',  
'location': {'lat': -6.151018337252608,  
'lng': 106.73398277361841,  
'labeledLatLngs': [{ 'label': 'display',  
'lat': -6.151018337252608,  
'lng': 106.73398277361841}],  
'distance': 231,  
'cc': 'ID',  
'country': 'Indonesia',  
'formattedAddress': ['Indonesia']},  
'categories': [{ 'id': '4bf58dd8d48988d17f941735',  
'name': 'Movie Theater',  
'pluralName': 'Movie Theaters',  
'shortName': 'Movie Theater',  
'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/arts_entertainme  
nt/movietheater_',  
'suffix': '.png'}},
```

```

    'primary': True]],
    'photos': {'count': 0, 'groups': []}},
    'referralId': 'e-0-50fb9af9e4b0c2329c07e0ca-0'},
    {'reasons': {'count': 0,
    'items': [{'summary': 'This spot is popular',
    'type': 'general',
    'reasonName': 'globalInteractionReason'}]}],
    'venue': {'id': '4c1adf28eac020a1191344c2',
    'name': 'Studio 29',
    'location': {'address': 'Jl.Fajar Baru Raya No.10',
    'lat': -6.146642554033565,
    'lng': 106.73227380822982,
    'labeledLatLngs': [{'label': 'display',
    'lat': -6.146642554033565,
    'lng': 106.73227380822982}],
    'distance': 389,
    'postalCode': '11730',
    'cc': 'ID',
    'city': 'Cengkareng',
    'state': 'Jakarta',
    'country': 'Indonesia',
    'formattedAddress': ['Jl.Fajar Baru Raya No.10',
    'Cengkareng',
    'Jakarta 11730',
    'Indonesia']},
    'categories': [{'id': '4bf58dd8d48988d1e5931735',
    'name': 'Music Venue',
    'pluralName': 'Music Venues',
    'shortName': 'Music Venue',
    'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/arts_entertainme
nt/musicvenue_',
    'suffix': '.png'}},
    'primary': True]],
    'photos': {'count': 0, 'groups': []}},
    'referralId': 'e-0-4c1adf28eac020a1191344c2-1'},
    {'reasons': {'count': 0,
    'items': [{'summary': 'This spot is popular',
    'type': 'general',
    'reasonName': 'globalInteractionReason'}]}],
    'venue': {'id': '4c3c9d9917f2ef3b981f81f4',
    'name': 'Laut Dadap',
    'location': {'address': 'Cengkareng',
    'lat': -6.148721908253163,
    'lng': 106.73294036575736,
    'labeledLatLngs': [{'label': 'display',
    'lat': -6.148721908253163,
    'lng': 106.73294036575736}],
    'distance': 207,
    'cc': 'ID',
    'city': 'Jakarta',
    'state': 'Jakarta',
    'country': 'Indonesia',
    'formattedAddress': ['Cengkareng', 'Jakarta', 'Jakarta', 'Indonesia']},
    'categories': [{'id': '4bf58dd8d48988d1e0941735',
    'name': 'Harbor / Marina',
    'pluralName': 'Harbors / Marinas',
    'shortName': 'Harbor / Marina',
    'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/parks_outdoors/h
arbor_',
    'suffix': '.png'}},
    'primary': True]],
    'photos': {'count': 0, 'groups': []}},
    'referralId': 'e-0-4c3c9d9917f2ef3b981f81f4-2'},
    {'reasons': {'count': 0,
    'items': [{'summary': 'This spot is popular',
    'type': 'general',
    'reasonName': 'globalInteractionReason'}]}],
    'venue': {'id': '54550c20498ec939e61fb693',
    'name': 'Family Mart City Park Apartement',

```

```
'location': {'address': 'Apartmeen City Park',
'lat': -6.146726144996763,
'lng': 106.73564142165156,
'labeledLatLngs': [{'label': 'display',
'lat': -6.146726144996763,
'lng': 106.73564142165156}]},
'distance': 280,
'cc': 'ID',
'city': 'Jakarta',
'state': 'Jakarta',
'country': 'Indonesia',
'formattedAddress': ['Apartmeen City Park',
'Jakarta Barat',
'Jakarta',
'Indonesia']},
'categories': [{'id': '53e510b7498ebcb1801b55d4',
'name': 'Night Market',
'pluralName': 'Night Markets',
'shortName': 'Night Market',
'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/shops/fleamarket',
'suffix': '.png'},
'primary': True}],
'photos': {'count': 0, 'groups': []},
'referralId': 'e-0-54550c20498ec939e61fb693-3'},
{'reasons': {'count': 0,
'items': [{'summary': 'This spot is popular',
'type': 'general',
'reasonName': 'globalInteractionReason'}]},
'venue': {'id': '4ca0723d46978cfadf4eb37f',
'name': 'Mie Baso',
'location': {'address': 'Jl Nurul Amal',
'lat': -6.149799,
'lng': 106.730642,
'labeledLatLngs': [{'label': 'display',
'lat': -6.149799,
'lng': 106.730642}]},
'distance': 464,
'cc': 'ID',
'city': 'Cengkareng',
'state': 'Jakarta',
'country': 'Indonesia',
'formattedAddress': ['Jl Nurul Amal',
'Cengkareng',
'Jakarta',
'Indonesia']},
'categories': [{'id': '4bf58dd8d48988d1c4941735',
'name': 'Restaurant',
'pluralName': 'Restaurants',
'shortName': 'Restaurant',
'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/default_',
'suffix': '.png'},
'primary': True}],
'photos': {'count': 0, 'groups': []},
'referralId': 'e-0-4ca0723d46978cfadf4eb37f-4'},
{'reasons': {'count': 0,
'items': [{'summary': 'This spot is popular',
'type': 'general',
'reasonName': 'globalInteractionReason'}]},
'venue': {'id': '546b16af498edfac405e64c2',
'name': 'bebek penyet ayu ros',
'location': {'lat': -6.152406215667725,
'lng': 106.73737335205078,
'labeledLatLngs': [{'label': 'display',
'lat': -6.152406215667725,
'lng': 106.73737335205078}]},
'distance': 467,
'cc': 'ID',
'country': 'Indonesia',
```

```

        'formattedAddress': ['Indonesia']],
        'categories': [{ 'id': '4bf58dd8d48988d147941735',
            'name': 'Diner',
            'pluralName': 'Diners',
            'shortName': 'Diner',
            'icon': { 'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/diner_',
                'suffix': '.png' },
            'primary': True}],
        'photos': { 'count': 0, 'groups': []}],
        'referralId': 'e-0-546b16af498edfac405e64c2-5'}}]]}}

```

From the Foursquare lab in the previous module, we know that all the information is in the *items* key. Before we proceed, let's borrow the **get\_category\_type** function from the Foursquare lab.

```

In [171... # function that extracts the category of the venue
def get_category_type(row):
    try:
        categories_list = row['categories']
    except:
        categories_list = row['venue.categories']

    if len(categories_list) == 0:
        return None
    else:
        return categories_list[0]['name']

```

Now we are ready to clean the json and structure it into a *pandas* dataframe.

```

In [172... venues = results['response']['groups'][0]['items']

nearby_venues = pd.json_normalize(venues) # flatten JSON

# filter columns
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.l
nearby_venues = nearby_venues.loc[:, filtered_columns]

# filter the category for each row
nearby_venues['venue.categories'] = nearby_venues.apply(get_category_type, axis=1)

# clean columns
nearby_venues.columns = [col.split(".")[1] for col in nearby_venues.columns]

nearby_venues.head()

```

```

Out[172...

```

	name	categories	lat	lng
0	XXI Puri Indah Mall	Movie Theater	-6.151018	106.733983
1	Studio 29	Music Venue	-6.146643	106.732274
2	Laut Dadap	Harbor / Marina	-6.148722	106.732940
3	Family Mart City Park Apartement	Night Market	-6.146726	106.735641
4	Mie Baso	Restaurant	-6.149799	106.730642

And how many venues were returned by Foursquare?

```

In [173... print('{} venues were returned by Foursquare.'.format(nearby_venues.shape[0]))

6 venues were returned by Foursquare.

```

## 2. Explore Neighborhoods in Jakarta



## Let's create a function to repeat the same process to all the neighborhoods in Jakarta

In [174...

```
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    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_id={}&version={}&lat={}&lng={}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # 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']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name'] for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['Kecamatan',
                            'Kecamatan Latitude',
                            'Kecamatan Longitude',
                            'Venue',
                            'Venue Latitude',
                            'Venue Longitude',
                            'Venue Category']

    return(nearby_venues)
```

In [175...

```
jakarta_venues = getNearbyVenues(names=jakarta_data['Kecamatan'],
                                  latitudes=jakarta_data['Latitude'],
                                  longitudes=jakarta_data['Longitude']
                                  )
```

```
CENGKARENG
GROGOL PETAMBURAN
KALI DERES
KEBON JERUK
KEMBANGAN
PALMERAH
TAMAN SARI
TAMBORA
CEMPAKA PUTIH
GAMBIR
JOHAR BARU
KEMAYORAN
MENTENG
SAWAH BESAR
SENEN
TANAH ABANG
CILANDAK
JAGAKARSA
```

KEBAYORAN BARU  
 KEBAYORAN LAMA  
 MAMPANG PRAPATAN  
 PANCORAN  
 PASAR MINGGU  
 PESANGGRAHAN  
 SETIA BUDI  
 TEBET  
 CAKUNG  
 CIPAYUNG  
 CIRACAS  
 DUREN SAWIT  
 JATINEGARA  
 KRAMAT JATI  
 MAKASAR  
 MATRAMAN  
 PASAR REBO  
 PULO GADUNG  
 CILINCING  
 KELAPA GADING  
 KOJA  
 PADEMANGAN  
 PENJARINGAN  
 TANJUNG PRIOK  
 KEPULAUAN SERIBU SELATAN  
 KEPULAUAN SERIBU UTARA

Let's check the size of the resulting dataframe

```
In [176... print(jakarta_venues.shape)
jakarta_venues.head()
```

(719, 7)

```
Out[176... 
```

	Kecamatan	Kecamatan Latitude	Kecamatan Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	CENGKARENG	-6.149093	106.734781	XXI Puri Indah Mall	-6.151018	106.733983	Movie Theater
1	CENGKARENG	-6.149093	106.734781	Studio 29	-6.146643	106.732274	Music Venue
2	CENGKARENG	-6.149093	106.734781	Laut Dadap	-6.148722	106.732940	Harbor / Marina
3	CENGKARENG	-6.149093	106.734781	Family Mart City Park Apartement	-6.146726	106.735641	Night Market
4	CENGKARENG	-6.149093	106.734781	Mie Baso	-6.149799	106.730642	Restaurant

**Tips:** To avoid frequent access to Foursquare API since we only use Developer Free License, then we store the data into file locally

```
In [178... jakarta_venues.to_csv(r'..\Coursera_Capstone\jakarta_venues.csv')
```

```
In [179... #to import from file
jakarta_venues = pd.read_csv("jakarta_venues.csv")
jakarta_venues = jakarta_venues.drop('Unnamed: 0', 1)
```

Let's check how many venues were returned for each Kecamatan

```
In [181... jakarta_venues.groupby('Kecamatan').count()
```

```
Out[181... 
```

	Kecamatan Latitude	Kecamatan Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
--	-----------------------	------------------------	-------	-------------------	--------------------	-------------------

Kecamatan	Kecamatan Latitude	Kecamatan Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Kecamatan						
CAKUNG	3	3	3	3	3	3
CEMPAKA PUTIH	7	7	7	7	7	7
CENGKARENG	6	6	6	6	6	6
CILANDAK	21	21	21	21	21	21
CILINCING	2	2	2	2	2	2
CIPAYUNG	2	2	2	2	2	2
CIRACAS	2	2	2	2	2	2
DUREN SAWIT	5	5	5	5	5	5
GAMBIR	15	15	15	15	15	15
GROGOL PETAMBURAN	35	35	35	35	35	35
JAGAKARSA	13	13	13	13	13	13
JATINEGARA	16	16	16	16	16	16
JOHAR BARU	4	4	4	4	4	4
KALI DERES	8	8	8	8	8	8
KEBAYORAN BARU	33	33	33	33	33	33
KEBAYORAN LAMA	84	84	84	84	84	84
KEBON JERUK	23	23	23	23	23	23
KELAPA GADING	21	21	21	21	21	21
KEMAYORAN	4	4	4	4	4	4
KEMBANGAN	14	14	14	14	14	14
KEPULAUAN SERIBU SELATAN	2	2	2	2	2	2
KEPULAUAN SERIBU UTARA	1	1	1	1	1	1
KOJA	4	4	4	4	4	4
KRAMAT JATI	4	4	4	4	4	4
MAKASAR	4	4	4	4	4	4
MAMPANG PRAPATAN	5	5	5	5	5	5
MATRAMAN	5	5	5	5	5	5
MENTENG	29	29	29	29	29	29
PADEMANGAN	13	13	13	13	13	13
PALMERAH	16	16	16	16	16	16
PANCORAN	16	16	16	16	16	16
PASAR MINGGU	11	11	11	11	11	11

	Kecamatan Latitude	Kecamatan Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Kecamatan						
PASAR REBO	4	4	4	4	4	4
PENJARINGAN	4	4	4	4	4	4
PESANGGRAHAN	5	5	5	5	5	5
PULO GADUNG	33	33	33	33	33	33
SAWAH BESAR	19	19	19	19	19	19
SEZEN	14	14	14	14	14	14
SETIA BUDI	43	43	43	43	43	43
TAMAN SARI	76	76	76	76	76	76
TAMBORA	9	9	9	9	9	9
TANAH ABANG	18	18	18	18	18	18
TANJUNG PRIOK	3	3	3	3	3	3
TEBET	63	63	63	63	63	63

Let's find out how many unique categories can be curated from all the returned venues

```
In [182... print('There are {} uniques categories.'.format(len(jakarta_venues['Venue Category'])))
```

There are 156 uniques categories.

### 3. Analyze Each Kecamatan

```
In [183... # one hot encoding
jakarta_onehot = pd.get_dummies(jakarta_venues[['Venue Category']], prefix="", prefix_sep=""

# add neighborhood column back to dataframe
jakarta_onehot['Kecamatan'] = jakarta_venues['Kecamatan']

# move neighborhood column to the first column
fixed_columns = [jakarta_onehot.columns[-1]] + list(jakarta_onehot.columns[:-1])
jakarta_onehot = jakarta_onehot[fixed_columns]

jakarta_onehot.head()
```

[illegible]

And let's examine the new dataframe size.

```
In [184... jakarta_onehot.shape
```

```
Out[184... (719, 157)
```

Next, let's group rows by Kecamatan and by taking the mean of the frequency of occurrence of each category

```
In [185... jakarta_grouped = jakarta_onehot.groupby('Kecamatan').mean().reset_index()  
jakarta_grouped
```

```
Out[185...
```

	Kecamatan	Accessories Store	Acehnese Restaurant	Airport Lounge	American Restaurant	Arcade	Art Gallery	Art Museum	Art Studio
0	CAKUNG	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
1	CEMPAKA PUTIH	0.000000	0.142857	0.00	0.000000	0.000000	0.000000	0.000000	0.0
2	CENGKARENG	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
3	CILANDAK	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
4	CILINCING	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
5	CIPAYUNG	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
6	CIRACAS	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
7	DUREN SAWIT	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
8	GAMBIR	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
9	GROGOL PETAMBURAN	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
10	JAGAKARSA	0.000000	0.000000	0.00	0.000000	0.076923	0.000000	0.000000	0.0
11	JATINEGARA	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
12	JOHAR BARU	0.000000	0.000000	0.00	0.000000	0.250000	0.000000	0.000000	0.0
13	KALI DERES	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
14	KEBAYORAN BARU	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
15	KEBAYORAN LAMA	0.000000	0.000000	0.00	0.011905	0.011905	0.011905	0.000000	0.0
16	KEBON JERUK	0.000000	0.000000	0.00	0.000000	0.043478	0.000000	0.043478	0.0
17	KELAPA GADING	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
18	KEMAYORAN	0.000000	0.000000	0.00	0.000000	0.250000	0.000000	0.000000	0.0
19	KEMBANGAN	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
20	KEPULAUAN SERIBU SELATAN	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
21	KEPULAUAN SERIBU UTARA	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
22	KOJA	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
23	KRAMAT JATI	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0

	Kecamatan	Accessories Store	Acehnese Restaurant	Airport Lounge	American Restaurant	Arcade	Art Gallery	Art Museum	Cr St
24	MAKASAR	0.000000	0.000000	0.25	0.000000	0.000000	0.000000	0.000000	0.0
25	MAMPANG PRAPATAN	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
26	MATRAMAN	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
27	MENTENG	0.034483	0.000000	0.00	0.034483	0.000000	0.000000	0.000000	0.0
28	PADEMANGAN	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
29	PALMERAH	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
30	PANCORAN	0.000000	0.062500	0.00	0.000000	0.000000	0.000000	0.000000	0.0
31	PASAR MINGGU	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
32	PASAR REBO	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
33	PENJARINGAN	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
34	PESANGGRAHAN	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
35	PULO GADUNG	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
36	SAWAH BESAR	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
37	SENEN	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
38	SETIA BUDI	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
39	TAMAN SARI	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
40	TAMBORA	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
41	TANAH ABANG	0.000000	0.000000	0.00	0.000000	0.055556	0.000000	0.000000	0.0
42	TANJUNG PRIOK	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
43	TEBET	0.000000	0.000000	0.00	0.000000	0.000000	0.015873	0.000000	0.0

Let's confirm the new size

In [186...] `jakarta_grouped.shape`

Out[186...] `(44, 157)`

Let's print each neighborhood along with the top 3 most common venues

We use only 3 to limit the research for this to, however it can be extended :)

```
In [188...] num_top_venues = 3

for hood in jakarta_grouped['Kecamatan']:
    print("-----"+hood+"-----")
    temp = jakarta_grouped[jakarta_grouped['Kecamatan'] == hood].T.reset_index()
    temp.columns = ['venue', 'freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
    print('\n')
```

---

----CAKUNG----

	venue	freq
0	Gas Station	0.33
1	Neighborhood	0.33
2	Lounge	0.33

----CEMPAKA PUTIH----

	venue	freq
0	Pizza Place	0.29
1	BBQ Joint	0.14
2	Fast Food Restaurant	0.14

----CENGKARENG----

	venue	freq
0	Music Venue	0.17
1	Diner	0.17
2	Night Market	0.17

----CILANDAK----

	venue	freq
0	Indonesian Restaurant	0.1
1	Gym	0.1
2	Convenience Store	0.1

----CILINCING----

	venue	freq
0	Park	0.5
1	Shopping Mall	0.5
2	Accessories Store	0.0

----CIPAYUNG----

	venue	freq
0	Shop & Service	0.5
1	Restaurant	0.5
2	Accessories Store	0.0

----CIRACAS----

	venue	freq
0	Playground	0.5
1	Department Store	0.5
2	Office	0.0

----DUREN SAWIT----

	venue	freq
0	Indonesian Meatball Place	0.4
1	Convenience Store	0.2
2	Coffee Shop	0.2

----GAMBIR----

	venue	freq
0	Hotel	0.27
1	Indonesian Restaurant	0.20
2	Food Truck	0.13

----GROGOL PETAMBURAN----

	venue	freq
0	Asian Restaurant	0.11
1	Noodle House	0.11
2	Hotel	0.09

----JAGAKARSA----

	venue	freq
0	Indonesian Restaurant	0.31
1	Asian Restaurant	0.15
2	Noodle House	0.08

----JATINEGARA----

	venue	freq
0	Jewelry Store	0.25
1	Asian Restaurant	0.19
2	Indonesian Restaurant	0.06

----JOHAR BARU----

	venue	freq
0	Indonesian Restaurant	0.25
1	Convenience Store	0.25
2	Arcade	0.25

----KALI DERES----

	venue	freq
0	Noodle House	0.25
1	Fried Chicken Joint	0.12
2	Supermarket	0.12

----KEBAYORAN BARU----

	venue	freq
0	Coffee Shop	0.18
1	Japanese Restaurant	0.15
2	Sushi Restaurant	0.09

----KEBAYORAN LAMA----

	venue	freq
0	Steakhouse	0.07
1	Coffee Shop	0.07
2	Japanese Restaurant	0.06

----KEBON JERUK----

	venue	freq
0	Noodle House	0.13
1	Indonesian Restaurant	0.09
2	Concert Hall	0.09

----KELAPA GADING----

	venue	freq
0	Indonesian Restaurant	0.19
1	Asian Restaurant	0.14
2	Korean Restaurant	0.10

----KEMAYORAN----

	venue	freq
0	Hotel	0.25
1	Arcade	0.25
2	Noodle House	0.25

----KEMBANGAN----

	venue	freq
0	Asian Restaurant	0.21
1	Seafood Restaurant	0.14
2	Pizza Place	0.07



----KEPULAUAN SERIBU SELATAN----

	venue	freq
0	Boat or Ferry	0.5
1	Clothing Store	0.5
2	Noodle House	0.0

----KEPULAUAN SERIBU UTARA----

	venue	freq
0	Resort	1.0
1	Accessories Store	0.0
2	Noodle House	0.0

----KOJA----

	venue	freq
0	Pizza Place	0.50
1	Indonesian Restaurant	0.25
2	Bookstore	0.25

----KRAMAT JATI----

	venue	freq
0	Noodle House	0.25
1	Hospital	0.25
2	Seafood Restaurant	0.25

----MAKASAR----

	venue	freq
0	Indonesian Restaurant	0.50
1	Airport Lounge	0.25
2	Asian Restaurant	0.25

----MAMPANG PRAPATAN----

	venue	freq
0	Indonesian Restaurant	0.4
1	Fast Food Restaurant	0.4
2	Noodle House	0.2

----MATRAMAN----

	venue	freq
0	Fast Food Restaurant	0.2
1	College Cafeteria	0.2
2	Dog Run	0.2

----MENTENG----

	venue	freq
0	Indonesian Restaurant	0.10
1	Coffee Shop	0.07
2	Park	0.07

----PADEMANGAN----

	venue	freq
0	Hotel	0.31
1	Seafood Restaurant	0.15
2	Toy / Game Store	0.08

----PALMERAH----

	venue	freq
0	Fast Food Restaurant	0.19
1	Pizza Place	0.12

2 Coffee Shop 0.12

----PANCORAN----

	venue	freq
0	Indonesian Restaurant	0.19
1	Steakhouse	0.06
2	Restaurant	0.06

----PASAR MINGGU----

	venue	freq
0	Japanese Restaurant	0.18
1	Convenience Store	0.18
2	Music Store	0.09

----PASAR REBO----

	venue	freq
0	Cosmetics Shop	0.25
1	Indonesian Restaurant	0.25
2	Bakery	0.25

----PENJARINGAN----

	venue	freq
0	Pharmacy	0.25
1	Spa	0.25
2	Theme Park	0.25

----PESANGGRAHAN----

	venue	freq
0	Noodle House	0.6
1	Food Truck	0.2
2	Indonesian Restaurant	0.2

----PULO GADUNG----

	venue	freq
0	Indonesian Restaurant	0.15
1	Convenience Store	0.12
2	Café	0.06

----SAWAH BESAR----

	venue	freq
0	Hotel	0.11
1	Indonesian Restaurant	0.11
2	Convenience Store	0.11

----SENEN----

	venue	freq
0	Hotel	0.29
1	Indonesian Restaurant	0.21
2	Dance Studio	0.07

----SETIA BUDI----

	venue	freq
0	Coffee Shop	0.21
1	Hotel	0.12
2	Pizza Place	0.07

----TAMAN SARI----

	venue	freq
0	Chinese Restaurant	0.26

```

1   Asian Restaurant  0.12
2       Noodle House  0.11

```

----TAMBORA----

```

          venue  freq
0  Italian Restaurant  0.11
1           Garden  0.11
2  Convenience Store  0.11

```

----TANAH ABANG----

```

          venue  freq
0  Indonesian Restaurant  0.28
1           Coffee Shop  0.11
2   Seafood Restaurant  0.11

```

----TANJUNG PRIOK----

```

          venue  freq
0   Cosmetics Shop  0.33
1  Convenience Store  0.33
2   Asian Restaurant  0.33

```

----TEBET----

```

          venue  freq
0  Indonesian Restaurant  0.13
1   Asian Restaurant  0.08
2   Convenience Store  0.06

```

## Let's put that into a *pandas* dataframe

First, let's write a function to sort the venues in descending order.

```

In [189... def return_most_common_venues(row, num_top_venues):
            row_categories = row.iloc[1:]
            row_categories_sorted = row_categories.sort_values(ascending=False)

            return row_categories_sorted.index.values[0:num_top_venues]

```

Now let's create the new dataframe and display the top 10 venues for each neighborhood.

```

In [190... num_top_venues = 3

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

# create columns according to number of top venues
columns = ['Kecamatan']
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))

# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Kecamatan'] = jakarta_grouped['Kecamatan']

for ind in np.arange(jakarta_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(jakarta_gr

neighborhoods_venues_sorted.head()

```

Out[190...

	Kecamatan	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	CAKUNG	Lounge	Gas Station	Neighborhood
1	CEMPAKA PUTIH	Pizza Place	BBQ Joint	Acehnese Restaurant
2	CENGKARENG	Restaurant	Night Market	Movie Theater
3	CILANDAK	Convenience Store	Gym	Indonesian Restaurant
4	CILINCING	Park	Shopping Mall	Wine Bar

## 4. Cluster Neighborhoods

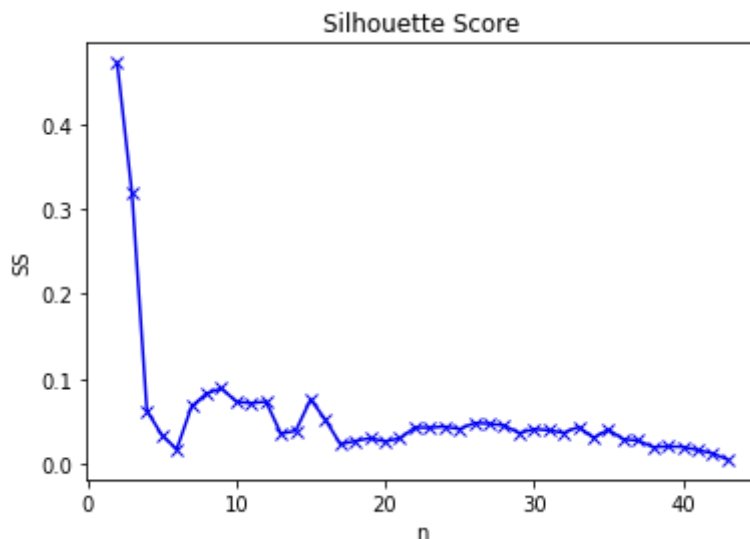
We will use **Silhouette Score** and **Elbow Method** to [determine the appropriate number of clusters](#).

We use range 2 to 44 since the maximum number of Kecamatan is 44. However we are not going to use 44 since there will be no point. We will find whether any specific segmentation in between of 2 and 44 by using Silhouette and Elbow Method.

```
In [192... from sklearn.metrics import silhouette_score
Sum_of_squared_distances = []
sos=[]
jakarta_grouped_clustering = jakarta_grouped.drop('Kecamatan', 1)

N = range(2,44)
for n in N:
    kmeans = KMeans(n_clusters=n, random_state=123).fit(jakarta_grouped_clustering)
    label = kmeans.predict(jakarta_grouped_clustering)
    Sum_of_squared_distances.append(kmeans.inertia_)
    sos.append(silhouette_score(jakarta_grouped_clustering, label))
    #print('Number Cluster {} = {} inertia after {} iteration'.format(n,kmeans.inert
    #print(f'Silhouette Score(n={n}): {silhouette_score(jakarta_grouped_clustering,
```

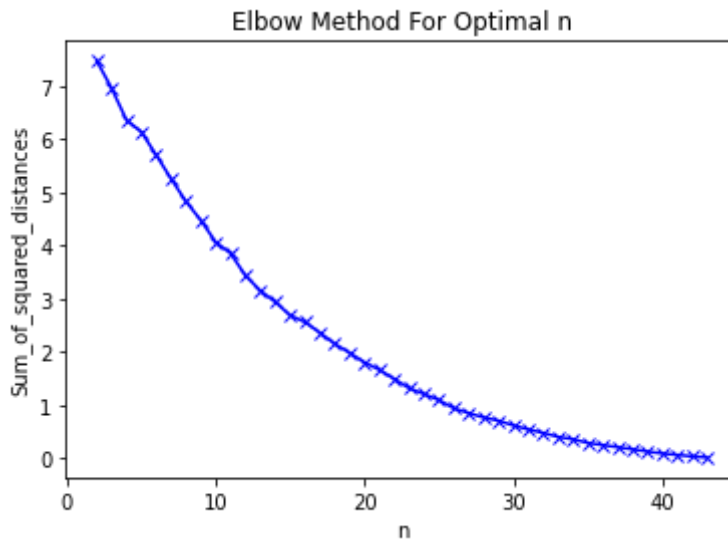
```
In [193... import matplotlib.pyplot as plt
plt.plot(N, sos, 'bx-')
plt.xlabel('n')
plt.ylabel('SS')
plt.title('Silhouette Score')
plt.show()
```



The best silhouette score is the one close to 1. From the graph above the best will be 2, 3, and 9.

We will omit 2 and 3 since it will be too general

```
In [195... import matplotlib.pyplot as plt
plt.plot(N, Sum_of_squared_distances, 'bx-')
plt.xlabel('n')
plt.ylabel('Sum_of_squared_distances')
plt.title('Elbow Method For Optimal n')
plt.show()
```



From Elbow method above it is not too obvious, but if we see closely we can see slope between from 9 to 10 which is something we are looking for.

Run *k*-means to cluster the neighborhood into 9 clusters.

```
In [197... # set number of clusters
kclusters = 9

jakarta_grouped_clustering = jakarta_grouped.drop('Kecamatan', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(jakarta_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

```
Out[197... array([6, 0, 6, 6, 7, 3, 2, 6, 6, 6])
```

Let's create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

```
In [198... #neighborhoods_venues_sorted = neighborhoods_venues_sorted.drop('Cluster Labels', 1)
#jakarta_merged=jakarta_merged.drop('Cluster Labels', 1)

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

jakarta_merged = jakarta_data

# merge Toronto_grouped with Toronto_data to add Latitude/Longitude for each neighbo
jakarta_merged = jakarta_merged.join(neighborhoods_venues_sorted.set_index('Kecamata

#df = df.join(df_geo.set_index('Postal Code'), on='PostalCode')

jakarta_merged.head(111) # check the last columns!
```

	Kabupaten	Kecamatan	Luas	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
0	JAKARTA BARAT	CENGKARENG	26.55	-6.149093	106.734781	6	Restaurant	Night Market
1	JAKARTA BARAT	GROGOL PETAMBURAN	9.99	-6.164188	106.788317	6	Noodle House	Asian Restaurant
2	JAKARTA BARAT	KALI DERES	30.23	-6.134300	106.705800	6	Noodle House	Supermarket
3	JAKARTA BARAT	KEBON JERUK	17.63	-6.192572	106.769725	6	Noodle House	Convenience Store
4	JAKARTA BARAT	KEMBANGAN	24.17	-6.193000	106.742600	6	Asian Restaurant	Seafood Restaurant
5	JAKARTA BARAT	PALMERAH	7.51	-6.191002	106.794363	6	Fast Food Restaurant	Pizza Place
6	JAKARTA BARAT	TAMAN SARI	7.74	-6.146142	106.818499	6	Chinese Restaurant	Asian Restaurant
7	JAKARTA BARAT	TAMBORA	5.40	-6.146614	106.801046	6	Convenience Store	Italian Restaurant
8	JAKARTA PUSAT	CEMPAKA PUTIH	4.70	-6.181214	106.868548	0	Pizza Place	BBQ Joint
9	JAKARTA PUSAT	GAMBIR	7.59	-6.170300	106.814800	6	Hotel	Indonesian Restaurant
10	JAKARTA PUSAT	JOHAR BARU	2.37	-6.183125	106.855332	1	Convenience Store	Arcade
11	JAKARTA PUSAT	KEMAYORAN	7.25	-6.162546	106.856890	6	Hotel	Ice Cream Shop
12	JAKARTA PUSAT	MENTENG	6.53	-6.195026	106.832224	6	Indonesian Restaurant	Park
13	JAKARTA PUSAT	SAWAH BESAR	6.16	-6.155891	106.833580	6	Convenience Store	Noodle House
14	JAKARTA PUSAT	SENEN	4.23	-6.184971	106.843235	6	Hotel	Indonesian Restaurant
15	JAKARTA PUSAT	TANAH ABANG	9.30	-6.205258	106.809500	1	Indonesian Restaurant	Seafood Restaurant
16	JAKARTA SELATAN	CILANDAK	18.16	-6.286898	106.794421	6	Convenience Store	Gym
17	JAKARTA SELATAN	JAGAKARSA	24.88	-6.330008	106.828191	1	Indonesian Restaurant	Asian Restaurant
18	JAKARTA SELATAN	KEBAYORAN BARU	12.92	-6.244146	106.800434	6	Coffee Shop	Japanese Restaurant
19	JAKARTA SELATAN	KEBAYORAN LAMA	16.74	-6.243886	106.779859	6	Steakhouse	Coffee Shop
20	JAKARTA SELATAN	MAMPANG PRAPATAN	7.74	-6.249374	106.821860	4	Fast Food Restaurant	Indonesian Restaurant
21	JAKARTA SELATAN	PANCORAN	8.53	-6.253298	106.844977	1	Indonesian Restaurant	Asian Restaurant

	Kabupaten	Kecamatan	Luas	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
22	JAKARTA SELATAN	PASAR MINGGU	21.71	-6.285642	106.829735	6	Convenience Store	Japanese Restaurant
23	JAKARTA SELATAN	PESANGGRAHAN	13.46	-6.248830	106.759631	4	Noodle House	Food Truck
24	JAKARTA SELATAN	SETIA BUDI	8.86	-6.219600	106.832600	6	Coffee Shop	Hotel
25	JAKARTA SELATAN	TEBET	9.03	-6.226016	106.858396	6	Indonesian Restaurant	Asian Restaurant
26	JAKARTA TIMUR	CAKUNG	42.27	-6.185562	106.940109	6	Lounge	Gas Station
27	JAKARTA TIMUR	CIPAYUNG	28.46	-6.329399	106.903739	3	Shop & Service	Restaurant
28	JAKARTA TIMUR	CIRACAS	16.08	-6.329635	106.876604	2	Playground	Department Store
29	JAKARTA TIMUR	DUREN SAWIT	22.66	-6.234138	106.919247	6	Indonesian Meatball Place	Convenience Store
30	JAKARTA TIMUR	JATINEGARA	10.26	-6.214976	106.870340	6	Jewelry Store	Asian Restaurant
31	JAKARTA TIMUR	KRAMAT JATI	13.30	-6.275477	106.870376	6	Hospital	Chinese Restaurant
32	JAKARTA TIMUR	MAKASAR	21.86	-6.269341	106.888818	1	Indonesian Restaurant	Airport Lounge
33	JAKARTA TIMUR	MATRAMAN	4.88	-6.203624	106.864579	0	Pizza Place	Dog Run
34	JAKARTA TIMUR	PASAR REBO	12.97	-6.324973	106.853376	4	Cosmetics Shop	Indonesian Restaurant
35	JAKARTA TIMUR	PULO GADUNG	15.60	-6.191109	106.890605	6	Indonesian Restaurant	Convenience Store
36	JAKARTA UTARA	CILINCING	37.70	-6.129015	106.944454	7	Park	Shopping Mall
37	JAKARTA UTARA	KELAPA GADING	16.12	-6.159938	106.902483	1	Indonesian Restaurant	Asian Restaurant
38	JAKARTA UTARA	KOJA	13.21	-6.120750	106.907362	0	Pizza Place	Indonesian Restaurant
39	JAKARTA UTARA	PADEMANGAN	9.91	-6.129052	106.828972	6	Hotel	Seafood Restaurant
40	JAKARTA UTARA	PENJARINGAN	34.48	-6.117265	106.767433	6	Pharmacy	Boutique
41	JAKARTA UTARA	TANJUNG PRIOK	25.12	-6.128858	106.870793	1	Convenience Store	Asian Restaurant
42	KABUPATEN ADMINISTRATIF KEPULAUAN SERIBU	KEPULAUAN SERIBU SELATAN	3.05	-5.803887	106.525306	8	Boat or Ferry	Clothing Store

	Kabupaten	Kecamatan	Luas	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
43	KABUPATEN ADMINISTRATIF KEPULAUAN SERIBU	KEPULAUAN SERIBU UTARA	5.65	-5.654443	106.568084	5	Resort	Wine Bar

Finally, let's visualize the resulting clusters

In [199...

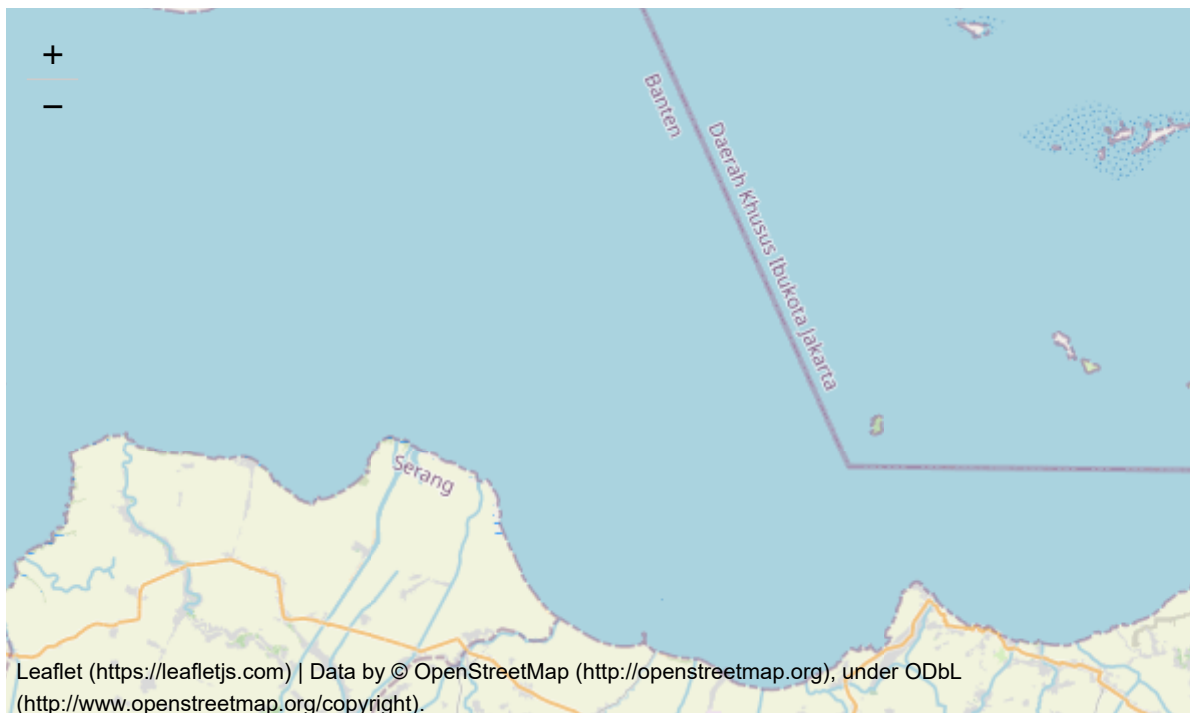
```
# create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

# 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[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

Out[199...



## 5. Examine Clusters

Now, we examine each cluster and determine the discriminating venue categories that



distinguish each cluster. Based on the defining categories, we can then assign a name to each cluster.

## Cluster 1. Pizza Place

In this Cluster we find there are alot of Pizza Place in this Kecamatan

```
In [210... jakarta_merged.loc[jakarta_merged['Cluster Labels'] == 0, jakarta_merged.columns[['1]
```

```
Out[210...
```

	Kecamatan	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
8	CEMPAKA PUTIH	0	Pizza Place	BBQ Joint	Acehnese Restaurant
33	MATRAMAN	0	Pizza Place	Dog Run	College Cafeteria
38	KOJA	0	Pizza Place	Indonesian Restaurant	Bookstore

## Cluster 2. Indonesian Restaurant and Asian Restaurant

In this Cluster we find there are alot of Pizza Place in this Kecamatan

```
In [201... jakarta_merged.loc[jakarta_merged['Cluster Labels'] == 1, jakarta_merged.columns[['1]
```

```
Out[201...
```

	Kecamatan	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
10	JOHAR BARU	1	Convenience Store	Arcade	Indonesian Restaurant
15	TANAH ABANG	1	Indonesian Restaurant	Seafood Restaurant	Coffee Shop
17	JAGAKARSA	1	Indonesian Restaurant	Asian Restaurant	Soccer Stadium
21	PANCORAN	1	Indonesian Restaurant	Asian Restaurant	Food Stand
32	MAKASAR	1	Indonesian Restaurant	Airport Lounge	Asian Restaurant
37	KELAPA GADING	1	Indonesian Restaurant	Asian Restaurant	Steakhouse
41	TANJUNG PRIOK	1	Convenience Store	Asian Restaurant	Cosmetics Shop

## Cluster 3. Playground

It only one Kecamatan that has Playground

```
In [202... jakarta_merged.loc[jakarta_merged['Cluster Labels'] == 2, jakarta_merged.columns[['1]
```

```
Out[202...
```

	Kecamatan	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
28	CIRACAS	2	Playground	Department Store	Dog Run

## Cluster 4. Shop and Service

This Kecamatan famous for service store and workshop

```
In [211... jakarta_merged.loc[jakarta_merged['Cluster Labels'] == 3, jakarta_merged.columns[['1]
```

Out[211...

	Kecamatan	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
27	CIPAYUNG	3	Shop & Service	Restaurant	Wine Bar

Cluster 5. Fast Food and Cosmetic Shop

In [204...

```
jakarta_merged.loc[jakarta_merged['Cluster Labels'] == 4, jakarta_merged.columns[[1]]
```

Out[204...

	Kecamatan	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
20	MAMPANG PRAPATAN	4	Fast Food Restaurant	Indonesian Restaurant	Noodle House
23	PESANGGRAHAN	4	Noodle House	Food Truck	Indonesian Restaurant
34	PASAR REBO	4	Cosmetics Shop	Indonesian Restaurant	Bakery

Cluster 6. Resort

This area very famous for vacation and leasure, it has lots of Resort since it is an Island. Similar with Cluster 9

In [205...

```
jakarta_merged.loc[jakarta_merged['Cluster Labels'] == 5, jakarta_merged.columns[[1]]
```

Out[205...

	Kecamatan	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
43	KEPULAUAN SERIBU UTARA	5	Resort	Wine Bar	Dog Run

Cluster 7. Noodle House, Chinese Restaurant and Hotel

It seems there is alot of Noodle House and Restaurant and i proves that these Kecamatan are very lively

In [206...

```
jakarta_merged.loc[jakarta_merged['Cluster Labels'] == 6, jakarta_merged.columns[[1]]
```

Out[206...

	Kecamatan	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	CENGKARENG	6	Restaurant	Night Market	Movie Theater
1	GROGOL PETAMBURAN	6	Noodle House	Asian Restaurant	Hotel
2	KALI DERES	6	Noodle House	Supermarket	Indonesian Restaurant
3	KEBON JERUK	6	Noodle House	Convenience Store	Indonesian Restaurant
4	KEMBANGAN	6	Asian Restaurant	Seafood Restaurant	Bubble Tea Shop
5	PALMERAH	6	Fast Food Restaurant	Pizza Place	Coffee Shop
6	TAMAN SARI	6	Chinese Restaurant	Asian Restaurant	Noodle House
7	TAMBORA	6	Convenience Store	Italian Restaurant	Fast Food Restaurant

	Kecamatan	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
9	GAMBIR	6	Hotel	Indonesian Restaurant	Food Truck
11	KEMAYORAN	6	Hotel	Ice Cream Shop	Arcade
12	MENTENG	6	Indonesian Restaurant	Park	Coffee Shop
13	SAWAH BESAR	6	Convenience Store	Noodle House	Hotel
14	SENEN	6	Hotel	Indonesian Restaurant	Grocery Store
16	CILANDAK	6	Convenience Store	Gym	Indonesian Restaurant
18	KEBAYORAN BARU	6	Coffee Shop	Japanese Restaurant	Sushi Restaurant
19	KEBAYORAN LAMA	6	Steakhouse	Coffee Shop	Japanese Restaurant
22	PASAR MINGGU	6	Convenience Store	Japanese Restaurant	Breakfast Spot
24	SETIA BUDI	6	Coffee Shop	Hotel	Café
25	TEBET	6	Indonesian Restaurant	Asian Restaurant	Convenience Store
26	CAKUNG	6	Lounge	Gas Station	Neighborhood
29	DUREN SAWIT	6	Indonesian Meatball Place	Convenience Store	Mediterranean Restaurant
30	JATINEGARA	6	Jewelry Store	Asian Restaurant	Donut Shop
31	KRAMAT JATI	6	Hospital	Chinese Restaurant	Seafood Restaurant
35	PULO GADUNG	6	Indonesian Restaurant	Convenience Store	Athletics & Sports
39	PADEMANGAN	6	Hotel	Seafood Restaurant	Asian Restaurant
40	PENJARINGAN	6	Pharmacy	Boutique	Spa

### Cluster 8. Park

Actually this area is an Industrial area, that is why being clustered specifically.

```
In [207...] jakarta_merged.loc[jakarta_merged['Cluster Labels'] == 7, jakarta_merged.columns[[1]]
```

	Kecamatan	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
36	CILINCING	7	Park	Shopping Mall	Wine Bar

### Cluster 9. Boat and Fery

This area very famous for vacation and leasure. Similar with Cluster 6

```
In [208...] jakarta_merged.loc[jakarta_merged['Cluster Labels'] == 8, jakarta_merged.columns[[1]]
```

	Kecamatan	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
--	-----------	----------------	-----------------------	-----------------------	-----------------------

	Kecamatan	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
42	KEPULAUAN SERIBU SELATAN	8	Boat or Ferry	Clothing Store	Dumpling Restaurant

## Result and Conclusion

From above 9 specific clusters we can determine each neighborhood segmentation and specific economical business potential that can be done specifically. This can be used by investor to invest and build business for each of neighborhood in Jakarta, it can be from Restaurant or Hotel, Resort, or even Workshop :)

END

In [ ]: