The Battle of Neighborhoods

Finding the best place to build Hotel without Restaurant.

West Java Province and East Java Province

Business Problem

In this project we will try to find the optimal location for **a hotel with lodging facilities only without a restaurant**. Specifically, this report will be addressed to stakeholders who are interested opening **Hotel in West Java Province Hotel and East Java Province, Indonesia.**

Because there are many hotels in West Java Province and East Java province, we will try to provide a busy location with **restaurants, sports bars, basketball courts, food and drink shops and entertainment areas**. We are also very interested because hotel visitors will more freely choose the food menu around the hotel. We also prefer a location that is as close as possible to the city center, assuming that the first conditions are met.

We will use the power of our data science to generate some of the environments that best fit this criterion. The advantages of each region will then manifest the business clearly so that the best outcome can be chosen by the stakeholders.

Purpose of this Project

The dataset combines city coordinates in West Java and East Java. The datasets does not include the venues within these locations. With venue information, it would be easy to find out more information about the neighborhoods. For example, how many sports bars and restaurants there are, and any basketball courts or playgrounds? We could also need to find out about any banks and food and drink shops? It would be better to understand or make good choices about where the hotel will be built

Next, the reason for this project is to, algorithmically, way to use location coordinates and mark each data point into the environment in two Provinces namely West Java and East Java in Indonesia. The algorithm used is k-means clustering. The main idea is to determine neighborhood with venues clustered around each other so that one can make a decision on the right neighborhood to chose based on the proximity of amenities and venues to each other.

K-means Clustering

The k-means clustering algorithm is an unsupervised clustering technique searches for a pre-determined number of clusters within an unlabeled multidimensional dataset. It accomplishes this using a simple conception of what the optimal clustering looks like:

- The "cluster center" is the arithmetic mean of all the points belonging to the cluster.
- Each point is closer to its own cluster center than to other cluster centers in the dataset.

The two assumptions above are presumably the basis of the k-means model.

Data

Based on definition of our problem, factors that will influence our decission are:

- number of **existing restaurants, sports bars, basketball courts, food and drink shops and entertainment areas in the neighborhood**
- number of and distance to restaurants, sports bars, basketball courts, food and drink shops and entertainment areas in the neighborhood, if any
- distance of neighborhood from city center

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

- The dataset for this project consists of information regarding the cities in Indonesia obtained from https://simplemaps.com/data/world-cities. Specifically, the data contain: **City, city_ascii, lat, lng , country, iso2, iso3, admin_name, capital, population and id**.
- I used business intelligence tools for **geocoding** the data to obtain the correct coordinates
- The data was then exported and converted into a .json, read into a pandas dataframe and sliced into West Java and East Java Province data for use in the project.

Data

```
cities_data = 'worldcities.csv'
df = pd.read_csv(cities_data)
print('The dataframe has', df.shape, 'rows and columns respectively.')
df.head()
```

```
The dataframe has (26569, 11) rows and columns respectively.
```

	city	city_ascii	lat	lng	country	iso2	iso3	admin_name	capital	population	id
0	Tokyo	Tokyo	35.6897	139.6922	Japan	JP	JPN	Tōkyō	primary	37977000.0	1392685764
1	Jakarta	Jakarta	-6.2146	106.8451	Indonesia	ID	IDN	Jakarta	primary	34540000.0	1360771077
2	Delhi	Delhi	28.6600	77.2300	India	IN	IND	Delhi	admin	29617000.0	1356872604
3	Mumbai	Mumbai	18.9667	72.8333	India	IN	IND	Mahārāshtra	admin	23355000.0	1356226629
4	Manila	Manila	14.5958	120.9772	Philippines	PH	PHL	Manila	primary	23088000.0	1608618140

Methodology

Exploratory Analysis

Exploratory analysis was performed by examining tables and plots of the download data. This was used to:

1. Segment the data into Cities in West Java and East Java Province

2. identify missing values, verify the quality of the data

3. determine likely approaches to modelling, which might best yield to good clustering.

Segmenting and Slicing, and visualizing the data

An important part of cluster modelling is the careful selection of the variables of available data. A prerequisite of the study is that the foursquare API is used to collect the venue information. Hence it is very important that the dataset for this work includes the coordinates of the cities to be studied. Segmentation and slicing of the data resulted in Table 1. The subjects included in the data for analysis includes: City, city_ascii, lat, lng , country, iso2, iso3, admin_name, capital, population and id

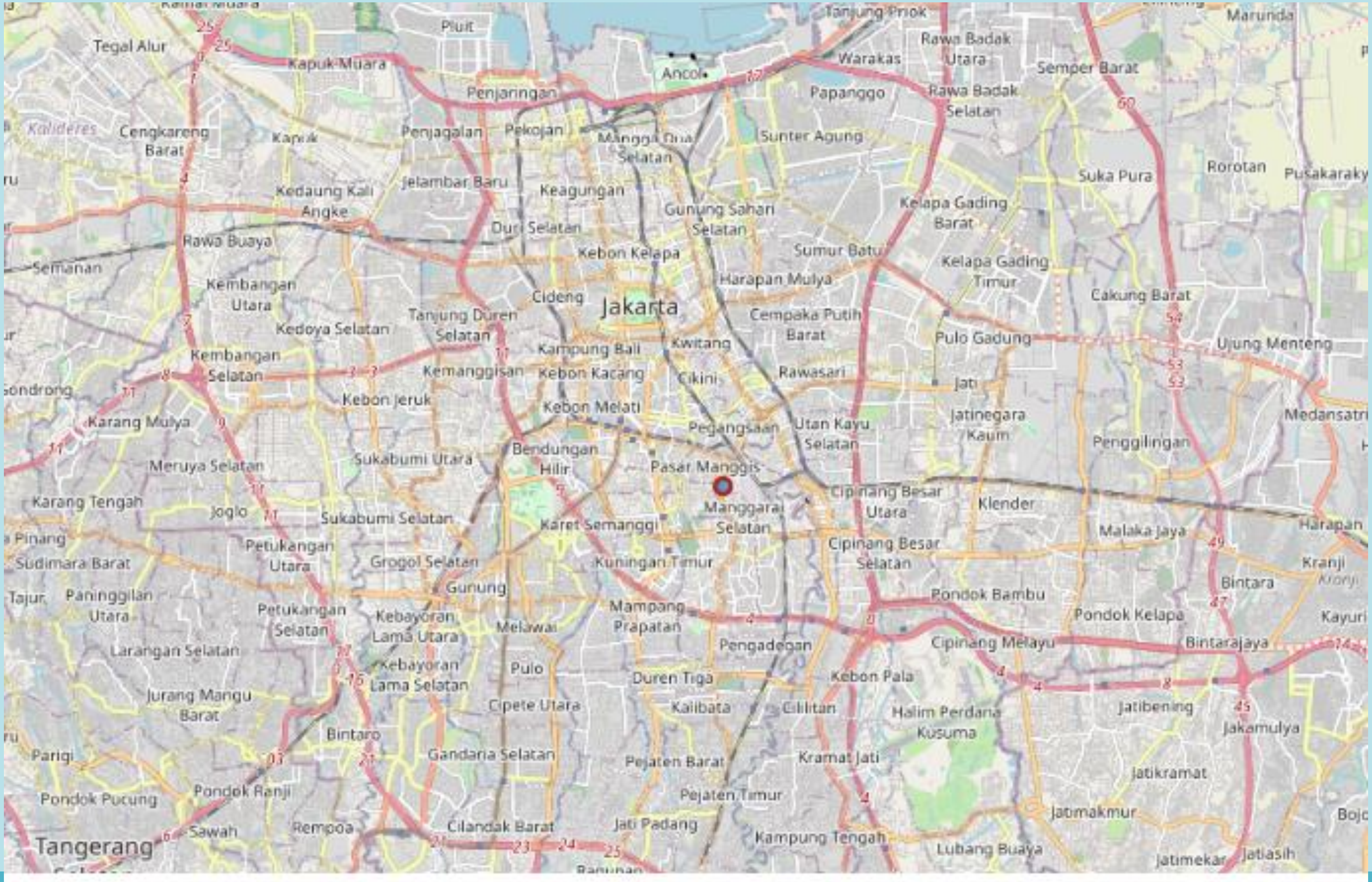
Segmenting Indonesia Data

```
Indonesia_data = df[df.country == 'Indonesia']
Indonesia_data.head(15)
```

	city	city_ascii	lat	lng	country	iso2	iso3	admin_name	capital	population	id
1	Jakarta	Jakarta	-6.2146	106.8451	Indonesia	ID	IDN	Jakarta	primary	34540000.0	1360771077
145	Surabaya	Surabaya	-7.2458	112.7378	Indonesia	ID	IDN	Jawa Timur	admin	4975000.0	1360484663
331	Bandung	Bandung	-6.9500	107.5667	Indonesia	ID	IDN	Jawa Barat	admin	2394873.0	1360313023
336	Bekasi	Bekasi	-6.2333	107.0000	Indonesia	ID	IDN	Jawa Barat	None	2381053.0	1360006015
346	Tangerang	Tangerang	-6.1783	106.6319	Indonesia	ID	IDN	Jawa Barat	None	2237006.0	1360002844
351	Medan	Medan	3.6667	98.6667	Indonesia	ID	IDN	Sumatera Utara	admin	2210625.0	1360543171
447	Makassar	Makassar	-5.1331	119.4136	Indonesia	ID	IDN	Sulawesi Selatan	admin	1651146.0	1360051337
451	Depok	Depok	-6.3940	106.8225	Indonesia	ID	IDN	Jawa Barat	None	1631951.0	1360962899
458	Semarang	Semarang	-6.9667	110.4167	Indonesia	ID	IDN	Jawa Tengah	admin	1621384.0	1360745537
507	Palembang	Palembang	-2.9833	104.7644	Indonesia	ID	IDN	Sumatera Selatan	admin	1452456.0	1360902897
607	Cilacap	Cilacap	-7.7167	109.0170	Indonesia	ID	IDN	Jawa Tengah	minor	1174964.0	1360503809
609	Bandar Lampung	Bandar Lampung	-5.4294	105.2625	Indonesia	ID	IDN	Lampung	admin	1166761.0	1360243491
680	Bogor	Bogor	-6.6000	106.8000	Indonesia	ID	IDN	Jawa Barat	None	1030720.0	1360771925
693	Patam	Patam	1.0678	104.0167	Indonesia	ID	IDN	Kepulauan Riau	None	1029808.0	1360893799
877	Padang	Padang	-0.9556	100.3606	Indonesia	ID	IDN	Sumatera Barat	admin	914970.0	1360900986

Folium

Folium is a powerful python library that builds on the data wrangling strengths of the python ecosystem and the mapping strengths of the Leaflet.js library. Generally, data is manipulating in Python, and then visualize it in on a Leaflet map via Folium.



Example Folium map of Jakarta

Neighborhood Exploration and Cluster – West Java and East Java

For the neighborhood exploration, the Foursquare API was used. The get request was deployed on the Foursquare URL to get the category types of venues limiting the number of venues to 100 within a 500 radius.Because the aim of the project is to determine the cluster of venues in the neighborhoods, one-hot encoding was performed on the venue categories to get dummies for each venue. That is to say, the venues were coded into 0s and 1s. The result was then grouped by neighborhood by taking the mean of the frequency of occurrence of each category.

We use the get_category_type function to get the category types

```
# 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 json is cleaned and the and structured into a pandas dataframe.

```
venues = results['response']['groups'][0]['items']

nearby_venues = json_normalize(venues) # flatten JSON

# filter columns
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
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(10)
```

Using the Foursquare API, the venues within the neighborhoods in both West Java and East Java Province in a vast number of outcomes. The radius defined for the venue returned venues with 677 rows and 7 columns for both West Java and East Java. The one-hot encoding produced a total number of 677 and 128 rows and columns for West Java and East Java Province respectively. Tables below show the results of the top 3 venues in each neighborhood venues for both West Java and East Java Province was grouped by neighborhood.

Cluster of Neighborhoods in West Java and East Java.

For the clustering of venues categories in the neighborhoods, the k-means cluster was employed. to cluster the neighborhood into 4 clusters. The k-means clustering machine learning algorithm is an unsupervised clustering technique searches for a pre-determined number of clusters within an unlabeled multidimensional dataset. It accomplishes this using a simple conception of what the optimal clustering looks like:

- The "cluster center" is the arithmetic mean of all the points belonging to the cluster.
- Each point is closer to its own cluster center than to other cluster centers in the dataset.

The two assumptions above are presumably the basis of the k-means model. To be able to produce the clusters and visualize it on a map, the sliced West Java and East Java data were merged with the grouped venue data. This was done so that the coordinates form the sliced data can aid in visualizing the clusters on a map.

Result

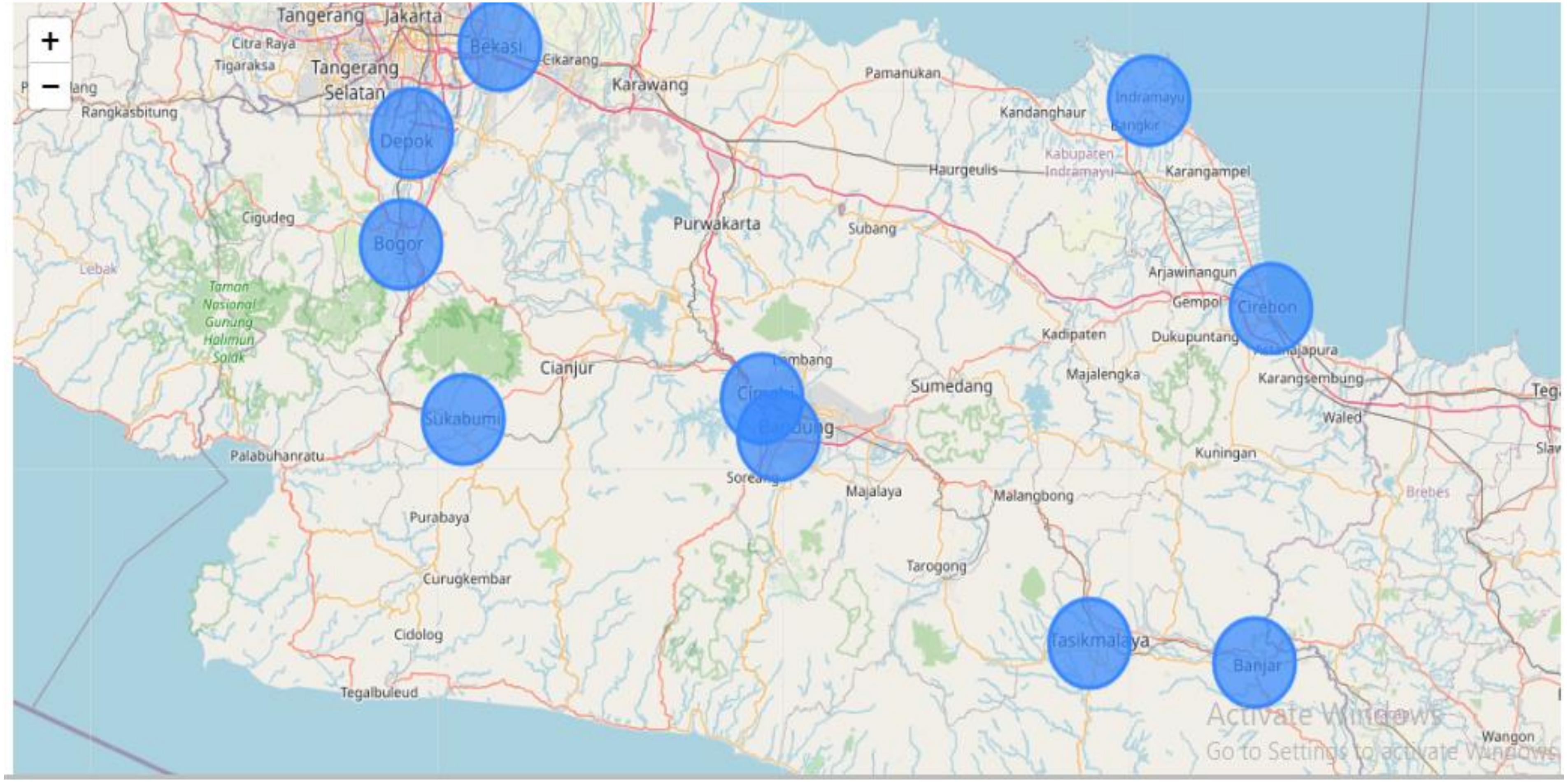
	index	city	lat	lng	country	admin_name	population
0	331	Bandung	-6.9500	107.5667	Indonesia	Jawa Barat	2394873.0
1	336	Bekasi	-6.2333	107.0000	Indonesia	Jawa Barat	2381053.0
2	451	Depok	-6.3940	106.8225	Indonesia	Jawa Barat	1631951.0
3	680	Bogor	-6.6000	106.8000	Indonesia	Jawa Barat	1030720.0
4	1080	Tasikmalaya	-7.3333	108.2000	Indonesia	Jawa Barat	678027.0
5	1195	Cimahi	-6.8833	107.5333	Indonesia	Jawa Barat	586580.0
6	1878	Sukabumi	-6.9197	106.9272	Indonesia	Jawa Barat	320970.0
7	1926	Cirebon	-6.7167	108.5667	Indonesia	Jawa Barat	316126.0
8	2888	Banjar	-7.3667	108.5333	Indonesia	Jawa Barat	182819.0
9	3908	Indramayu	-6.3356	108.3190	Indonesia	Jawa Barat	123263.0

West Java Data (jabar_data)

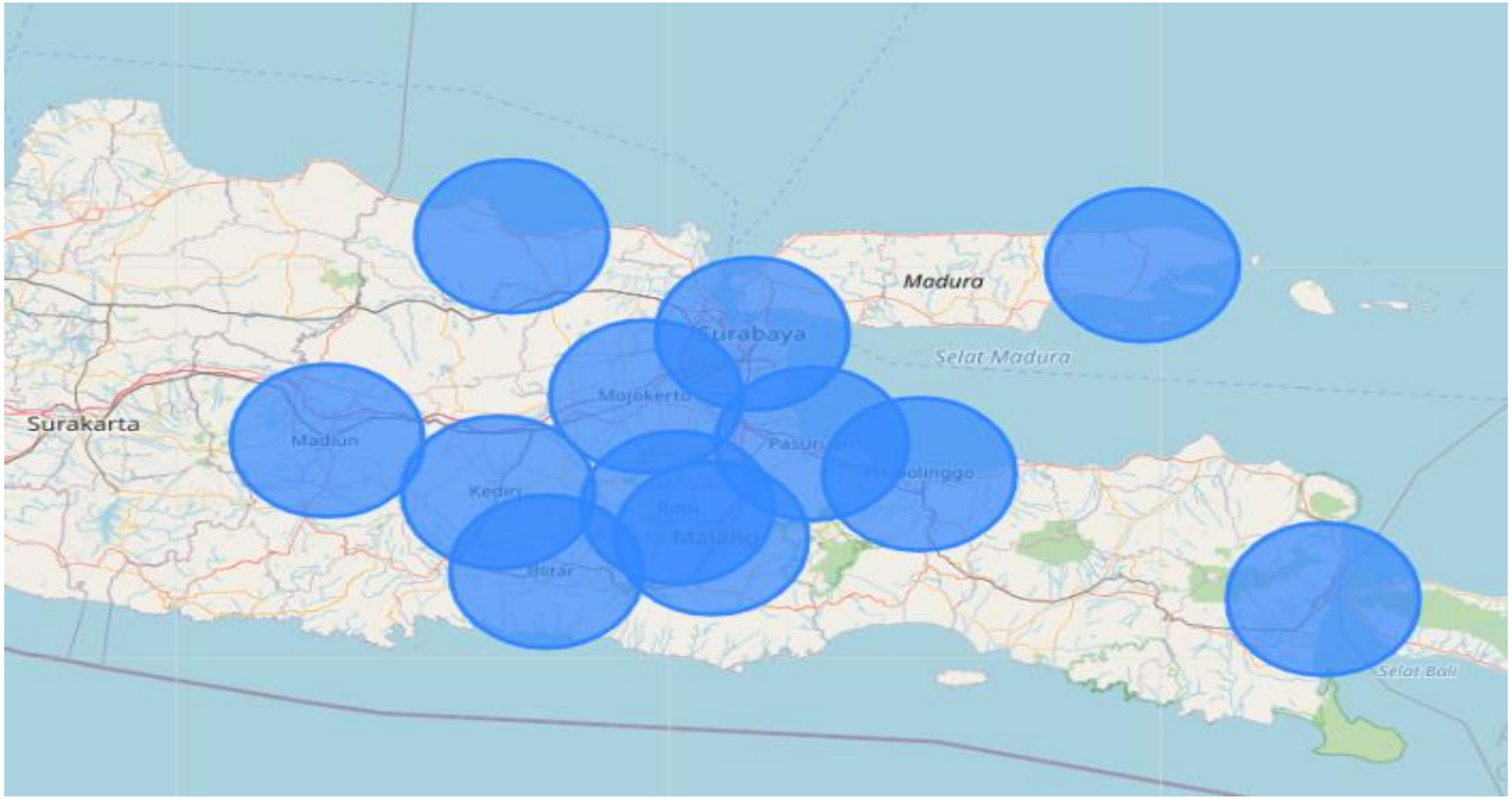
	index	city	lat	lng	country	admin_name	population
0	145	Surabaya	-7.2458	112.7378	Indonesia	Jawa Timur	4975000.0
1	977	Malang	-7.9800	112.6200	Indonesia	Jawa Timur	780000.0
2	2256	Kediri	-7.8166	112.0119	Indonesia	Jawa Timur	252000.0
3	2474	Probolinggo	-7.7500	113.2167	Indonesia	Jawa Timur	223159.0
4	2801	Madiun	-7.6300	111.5231	Indonesia	Jawa Timur	186099.0
5	2807	Pasuruan	-7.6406	112.9065	Indonesia	Jawa Timur	186262.0
6	2813	Batu	-7.8672	112.5239	Indonesia	Jawa Timur	190184.0
7	3040	Banyuwangi	-8.1950	114.3696	Indonesia	Jawa Timur	172424.0
8	3617	Mojokerto	-7.4722	112.4336	Indonesia	Jawa Timur	130196.0
9	3661	Blitar	-8.1000	112.1500	Indonesia	Jawa Timur	132018.0
10	9767	Sumenep	-7.0049	113.8496	Indonesia	Jawa Timur	84656.0
11	9990	Tuban	-6.8995	112.0500	Indonesia	Jawa Timur	76242.0

East Java Data (jatim_data)

Since the aim of the project is to cluster the neighborhoods, the k-means algorithm is applied to the onehot encoded venue dataset, assuming there are 4 different clusters. The tables below show the neighborhood and the cluster labels assigned to it after the k-means algorithm was applied. Cluster label ‘0’ represents the 1st cluster and ‘3’ the 4th cluster. This series of plots shows the data for each pair of



Neighborhood cluster for West Java



Neighborhood cluster for East Java

West Java most common venues

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bandung	Chinese Restaurant	BBQ Joint	Asian Restaurant	Coffee Shop	Indonesian Restaurant	Noodle House	Hotel	Convenience Store	Bakery	Pizza Place
1	Banjar	Department Store	Indonesian Restaurant	Rest Area	Asian Restaurant	Restaurant	Train Station	Park	Market	Donut Shop	Dog Run
2	Bekasi	Indonesian Restaurant	Fast Food Restaurant	Asian Restaurant	Coffee Shop	Pizza Place	Pool	Multiplex	Hotel	Chinese Restaurant	Food Court
3	Bogor	Indonesian Restaurant	Hotel	Café	Coffee Shop	Bakery	Restaurant	Noodle House	Asian Restaurant	Dessert Shop	Bistro
4	Cimahi	Convenience Store	Café	Indonesian Restaurant	Coffee Shop	Noodle House	Japanese Restaurant	Pool	Pizza Place	Supermarket	Department Store
5	Cirebon	Hotel	Indonesian Restaurant	Coffee Shop	Steakhouse	Food Truck	Café	Bakery	Diner	Soccer Stadium	Seafood Restaurant
6	Depok	Indonesian Restaurant	Fast Food Restaurant	Café	Coffee Shop	Japanese Restaurant	Asian Restaurant	Bakery	Chinese Restaurant	Seafood Restaurant	Snack Place
7	Indramayu	Hotel	Stadium	Food Truck	Indonesian Restaurant	Motel	Plaza	Eastern European Restaurant	Department Store	Chinese Restaurant	Food
8	Sukabumi	Indonesian Restaurant	Coffee Shop	Hotel	Snack Place	Noodle House	Seafood Restaurant	Karaoke Bar	Café	Food Truck	Steakhouse
9	Tasikmalaya	Sundanese Restaurant	Indonesian Meatball Place	Hotel	Indonesian Restaurant	Café	Coffee Shop	Supermarket	Juice Bar	Diner	Soup Place

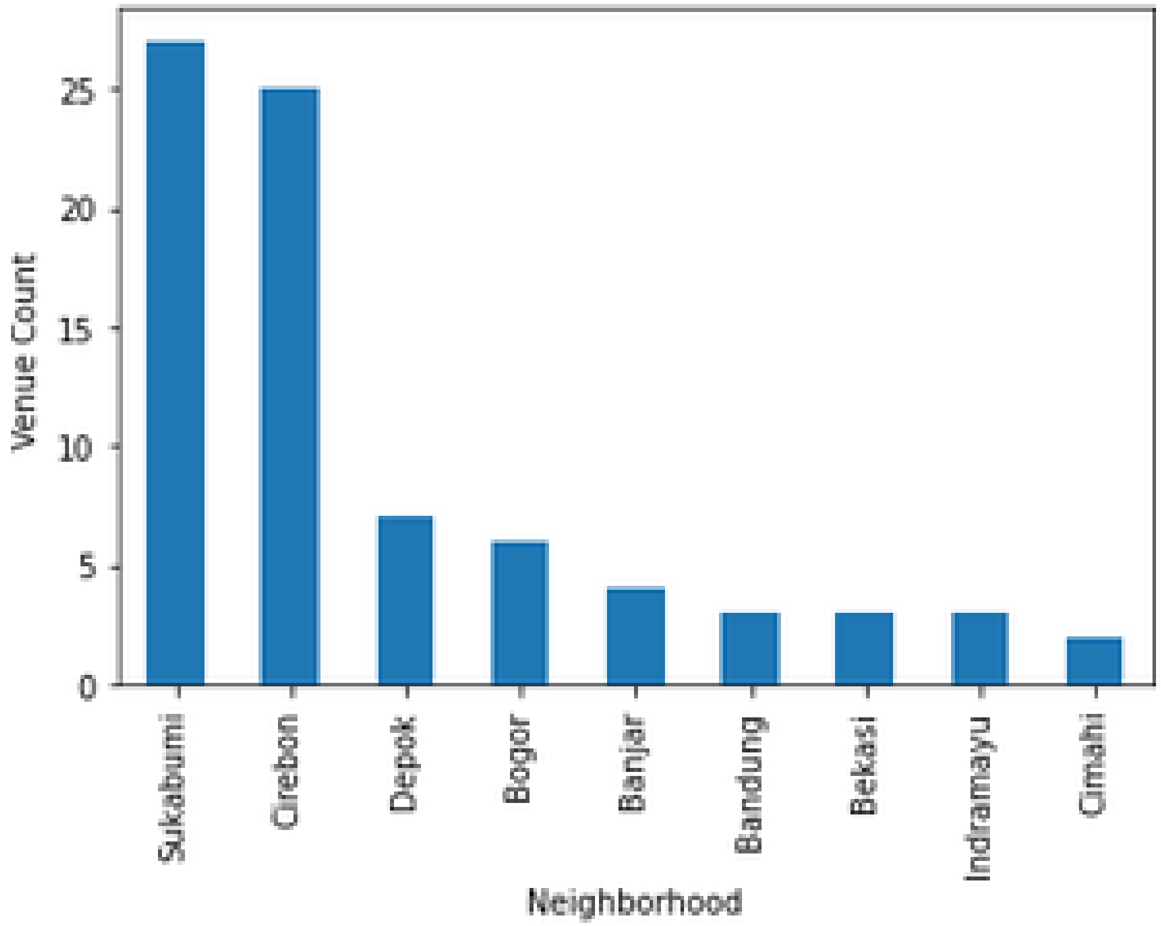
East Java most common venues

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	1	Banyuwangi	Hotel	Indonesian Restaurant	Coffee Shop	Steakhouse	Food Truck	Café	Bakery	Diner	Soccer Stadium	Seafood Restaurant
1	2	Batu	Indonesian Restaurant	Coffee Shop	Hotel	Snack Place	Noodle House	Seafood Restaurant	Karaoke Bar	Café	Food Truck	Steakhouse
2	1	Blitar	Hotel	Stadium	Food Truck	Indonesian Restaurant	Motel	Plaza	Eastern European Restaurant	Department Store	Chinese Restaurant	Food
3	1	Kediri	Indonesian Restaurant	Fast Food Restaurant	Café	Coffee Shop	Japanese Restaurant	Asian Restaurant	Bakery	Chinese Restaurant	Seafood Restaurant	Snack Place
4	1	Madiun	Sundanese Restaurant	Indonesian Meatball Place	Hotel	Indonesian Restaurant	Café	Coffee Shop	Supermarket	Juice Bar	Diner	Soup Place
5	1	Malang	Indonesian Restaurant	Fast Food Restaurant	Asian Restaurant	Coffee Shop	Pizza Place	Pool	Multiplex	Hotel	Chinese Restaurant	Food Court
6	1	Mojokerto	Department Store	Indonesian Restaurant	Rest Area	Asian Restaurant	Restaurant	Train Station	Park	Market	Donut Shop	Dog Run
7	3	Pasuruan	Convenience Store	Café	Indonesian Restaurant	Coffee Shop	Noodle House	Japanese Restaurant	Pool	Pizza Place	Supermarket	Department Store
8	1	Probolinggo	Indonesian Restaurant	Hotel	Café	Coffee Shop	Bakery	Restaurant	Noodle House	Asian Restaurant	Dessert Shop	Bistro
9	0	Surabaya	Chinese Restaurant	BBQ Joint	Asian Restaurant	Coffee Shop	Indonesian Restaurant	Noodle House	Hotel	Convenience Store	Bakery	Pizza Place

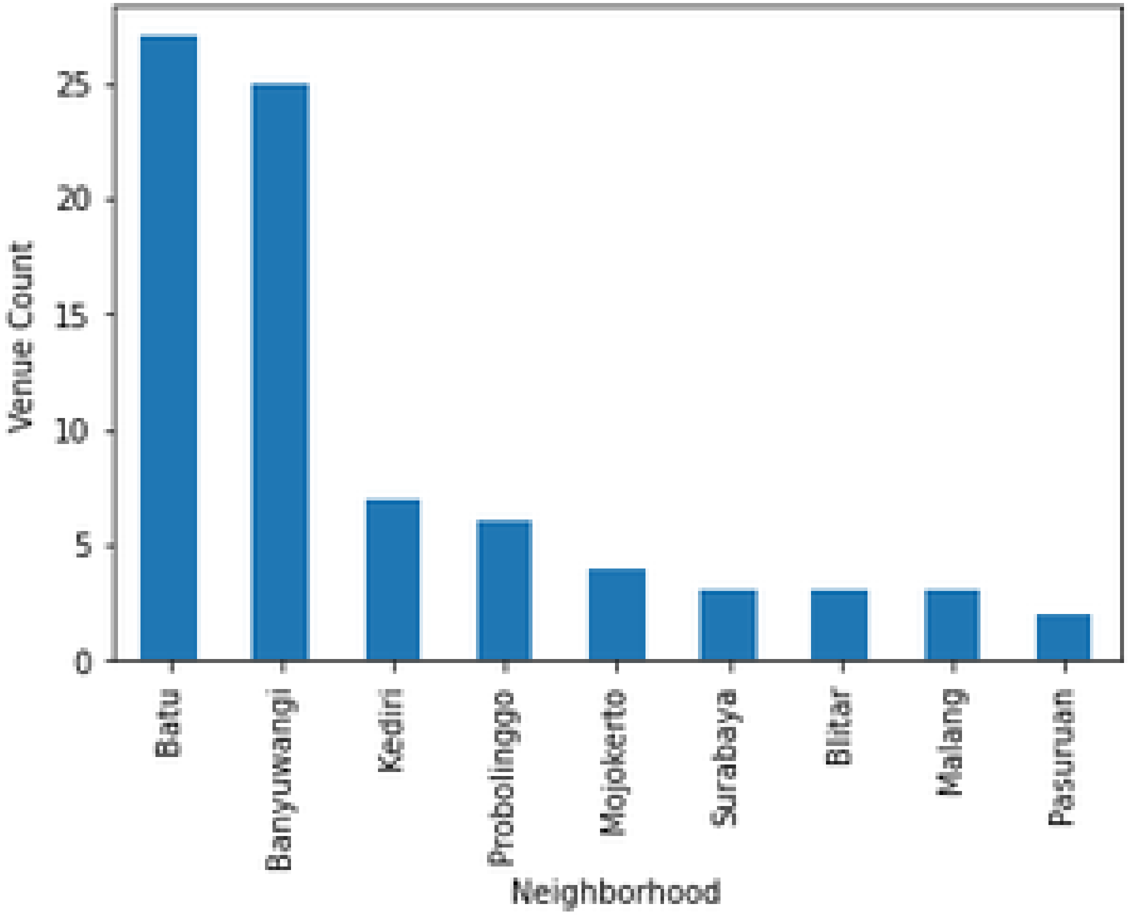
From the exploratory analysis, it seems a lot of the neighborhoods are in the cluster for both West Java and East Java Province. When we look at the clusters for West Java, it becomes clear that the first two most common venues in the neighborhoods contain a lot of mixed amenities: Chinese Restaurant, Indonesian Restaurant, Fast Food Restaurant, Sundanese Restaurant, Snack Place, Steakhouse, Japanese Restaurant, Fast Food Restaurant, Pool, Pizza Place, Karaoke Bar, Eastern European Restaurant.

Again, looking at that of East Java Province, we have: Convenience Store, Coffee Shop, Department Store, Juice Bar, Noodle House, Japanese Restaurant, Soup Place, Park, and Donut Shop.

West Java Venue Count Bar graph



East Java Venue Count Bar graph



Cluster in West Java

Cluster 1 - Authentic Restaurant

```
jabar_merged.loc[jabar_merged['Cluster Labels'] == 0.0, jabar_merged.columns[[0]+[1]+[2] + list(range(5, jabar_merged.shape[1]))]]
```

	city	lat	lng	population	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1080	Tasikmalaya	-7.3333	108.2	678027.0	0	Sundanese Restaurant	Indonesian Meatball Place	Hotel	Indonesian Restaurant	Café	Coffee Shop	Supermarket	Juice Bar	Diner	Soup Place

Cluster 2 - Indonesia Restoran Nearby Jakarta

```
jabar_merged.loc[jabar_merged['Cluster Labels'] == 1, jabar_merged.columns[[0]+[1]+[2] + list(range(5, jabar_merged.shape[1]))]]
```

	city	lat	lng	population	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
331	Bandung	-6.9500	107.5667	2394873.0	1	Chinese Restaurant	BBQ Joint	Asian Restaurant	Coffee Shop	Indonesian Restaurant	Noodle House	Hotel	Convenience Store	Bakery	Pizza Place
336	Bekasi	-6.2333	107.0000	2381053.0	1	Indonesian Restaurant	Fast Food Restaurant	Asian Restaurant	Coffee Shop	Pizza Place	Pool	Multiplex	Hotel	Chinese Restaurant	Food Court
451	Depok	-6.3940	106.8225	1631951.0	1	Indonesian Restaurant	Fast Food Restaurant	Café	Coffee Shop	Japanese Restaurant	Asian Restaurant	Bakery	Chinese Restaurant	Seafood Restaurant	Snack Place
680	Bogor	-6.6000	106.8000	1030720.0	1	Indonesian Restaurant	Hotel	Café	Coffee Shop	Bakery	Restaurant	Noodle House	Asian Restaurant	Dessert Shop	Bistro
1195	Cimahi	-6.8833	107.5333	586580.0	1	Convenience Store	Café	Indonesian Restaurant	Coffee Shop	Noodle House	Japanese Restaurant	Pool	Pizza Place	Supermarket	Department Store
1878	Sukabumi	-6.9197	106.9272	320970.0	1	Indonesian Restaurant	Coffee Shop	Hotel	Snack Place	Noodle House	Seafood Restaurant	Karaoke Bar	Café	Food Truck	Steakhouse
1926	Cirebon	-6.7167	108.5667	316126.0	1	Hotel	Indonesian Restaurant	Coffee Shop	Steakhouse	Food Truck	Café	Bakery	Diner	Soccer Stadium	Seafood Restaurant

Cluster in West Java

Cluster 3 - Shoping and Nearby Central Java

```
jabar_merged.loc[jabar_merged['Cluster Labels'] == 2, jabar_merged.columns[[0]+[1]+[2] + list(range(5, jabar_merged.shape[1]))]]
```

	city	lat	lng	population	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2888	Banjar	-7.3667	108.5333	182819.0	2	Department Store	Indonesian Restaurant	Rest Area	Asian Restaurant	Restaurant	Train Station	Park	Market	Donut Shop	Dog Run

Cluster 4 - Beach view

```
jabar_merged.loc[jabar_merged['Cluster Labels'] == 3, jabar_merged.columns[[0]+[1]+[2] + list(range(5, jabar_merged.shape[1]))]]
```

	city	lat	lng	population	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3908	Indramayu	-6.3356	108.319	123263.0	3	Hotel	Stadium	Food Truck	Indonesian Restaurant	Motel	Plaza	Eastern European Restaurant	Department Store	Chinese Restaurant	Food

Cluster in East Java

Cluster 1 - Restaurant and City

```
jatim_merged.loc[jatim_merged['Cluster Labels'] == 0.0, jatim_merged.columns[[0]+[1]+[2] + list(range(5, jatim_merged.shape[1]))]]
```

	city	lat	lng	population	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
145	Surabaya	-7.2458	112.7378	4975000.0	0.0	Chinese Restaurant	BBQ Joint	Asian Restaurant	Coffee Shop	Indonesian Restaurant	Noodle House	Hotel	Convenience Store	Bakery	Pizza Place

Cluster 2 - Restaurant in Highland

```
jatim_merged.loc[jatim_merged['Cluster Labels'] == 1, jatim_merged.columns[[0]+[1]+[2] + list(range(5, jatim_merged.shape[1]))]]
```

	city	lat	lng	population	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
977	Malang	-7.9800	112.6200	780000.0	1.0	Indonesian Restaurant	Fast Food Restaurant	Asian Restaurant	Coffee Shop	Pizza Place	Pool	Multiplex	Hotel	Chinese Restaurant	Food Court
2256	Kediri	-7.8166	112.0119	252000.0	1.0	Indonesian Restaurant	Fast Food Restaurant	Café	Coffee Shop	Japanese Restaurant	Asian Restaurant	Bakery	Chinese Restaurant	Seafood Restaurant	Snack Place
2474	Probolinggo	-7.7500	113.2167	223159.0	1.0	Indonesian Restaurant	Hotel	Café	Coffee Shop	Bakery	Restaurant	Noodle House	Asian Restaurant	Dessert Shop	Bistro
2801	Madiun	-7.6300	111.5231	186099.0	1.0	Sundanese Restaurant	Indonesian Meatball Place	Hotel	Indonesian Restaurant	Café	Coffee Shop	Supermarket	Juice Bar	Diner	Soup Place
3040	Banyuwangi	-8.1950	114.3696	172424.0	1.0	Hotel	Indonesian Restaurant	Coffee Shop	Steakhouse	Food Truck	Café	Bakery	Diner	Soccer Stadium	Seafood Restaurant
3617	Mojokerto	-7.4722	112.4336	130196.0	1.0	Department Store	Indonesian Restaurant	Rest Area	Asian Restaurant	Restaurant	Train Station	Park	Market	Donut Shop	Dog Run
3661	Blitar	-8.1000	112.1500	132018.0	1.0	Hotel	Stadium	Food Truck	Indonesian Restaurant	Motel	Plaza	Eastern European Restaurant	Department Store	Chinese Restaurant	Food

Cluster in East Java

Cluster 3 - Restaurant and Coffe Shop

```
jatim_merged.loc[jatim_merged['Cluster Labels'] == 2, jatim_merged.columns[[0]+[1]+[2] + list(range(5, jatim_merged.shape[1]))]]
```

	city	lat	lng	population	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2813	Batu	-7.8672	112.5239	190184.0	2.0	Indonesian Restaurant	Coffee Shop	Hotel	Snack Place	Noodle House	Seafood Restaurant	Karaoke Bar	Café	Food Truck	Steakhouse

Cluster 4 - Convenience Store and Cafe

```
jatim_merged.loc[jatim_merged['Cluster Labels'] == 3, jatim_merged.columns[[0]+[1]+[2] + list(range(5, jatim_merged.shape[1]))]]
```

	city	lat	lng	population	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2807	Pasuruan	-7.6406	112.9065	186262.0	3.0	Convenience Store	Café	Indonesian Restaurant	Coffee Shop	Noodle House	Japanese Restaurant	Pool	Pizza Place	Supermarket	Department Store

Discussion

So, the question is where one should consider establishing a hotel in West Java and East Java Provinces?

<u>For West Java</u>, Cluster 1 covering the city of Tasikmalaya will be selected by a stakeholder to build a hotel with authentic Sundanese nuances, because there are several restaurants around it including a Sundanese restaurant. Cluster 2 covering several cities will be chosen because the area is close to Jakarta and surrounded by many Indonesian restaurants, here it is more suitable as an alternative for people to go to Jakarta. Cluster 3 includes banjar which is somewhat less suitable for building hotels because there are fewer restaurants. Cluster 4 covers the city of Indramayu which is close to the beach, here there is a potential to build a hotel with a beach style.

<u>For East Java</u>, Cluster 1 covering the city of Surabaya will be chosen by a stakeholder to build a hotel with a sparkling city feel, because Surabaya is the second largest city in Indonesia and the capital of the province of East Java. Cluster 2 includes several cities in the highlands, will be chosen because of the cold area and there are a variety of Indonesian restaurants, so as to satisfy the appetite of visitors. Cluster 3 covers Batu City, can be an alternative to the highlands, there are many restaurants and karaoke bars, perfect for people who want to cool off. Cluster 4 covers the city of Pasuruan, here it is suitable for building a hotel with a beach style because it is not as busy as Surabaya, suitable for people who want to avoid noisy atmosphere.

Discussion

The purpose of this job is to provide the necessary facilities to help people decide where to build a hotel if they think about it. Using public data sets obtained from the web, I was able to overcome several factors by analyzing the environment in the two main Provinces of Indonesia, West Java and East Java based on the spatial distribution of places in the selected neighborhoods.

My analysis shows that using the folium-python library which helps build fast interactive data visualizations and the Foursquare API for environmental data collection, it is feasible to segment data for neighboring cities based on a known and accepted machine learning technique - the K-Means Algorithm. These results should be considered bound within the scope of the data set used, as no information is available about their origin. The results will be of interest to both business people and people aiming to compare different environments when thinking about staycation in different environments, given the ease of accessing multiple places in a clustered setting.

Of course there is still a lot of room for improvement. For example, obtaining more than the current neighborhood location to analyze and classify wide expanses of geographic settings. We were also able to use and analyze crime data - publicly available in the two provinces - to help provide sufficient room for decision-making regarding the selection of sites to relocate. This information may be very useful because we certainly don't want to do it. living in a crime-ridden environment. While the approach used here may not be robust, it does demonstrate the usefulness of environmental data analysis.

Thank You