# The Battle of Neighborhoods:

## Exploring 'The Big Durian' for a Coffee Shop Business Opportunity

Coursera Applied Data Science Capstone Project

Week 5 Submission – Full Report

Ginanjar Saputra

16 February 2021

---

**Contents**

1. Introduction ..... 1
   1.1. Background Information ..... 1
   1.2. Problem Statement ..... 2
2. Data ..... 2
3. Methodology ..... 3
   3.1. Jakarta Regions, Postal Codes, and Geographical Coordinates ..... 3
   3.2. Exploring Venues in the Subdistricts of Jakarta ..... 4
   3.3. Most Common Venues Overall ..... 5
   3.4. One-Hot Encoding ..... 5
   3.5. Clustering Subdistricts Based on Venues Similarity ..... 6
   3.6. Examining Each Cluster ..... 8
   3.7. Visualizing Distribution of Coffee Shop Locations in Jakarta ..... 10
4. Results and Discussion ..... 11
5. Conclusion ..... 12

---

## 1. Introduction

### 1.1. Background Information

Jakarta is the special capital region of Indonesia, an archipelago in Southeast Asia. It is located on the northwest coast of Java and is home to a population of 10.5 million. The greater Jakarta metropolitan area, which extends over 6,300 $km^2$, has a staggering population of 35 million and is the second-largest urban agglomeration in the world[1]. A melting pot of many cultures, Jakarta is the center of Indonesia's economic activities

[1] United Nations, Department of Economic and Social Affairs, Population Division (2018). World Urbanization Prospects: The 2018 Revision, Online Edition.

which has attracted people from across the archipelago to move to the city in search of opportunities and a potentially better standard of living.

Business opportunities abound in Jakarta, but the food-and-beverage (F&B) sector has long been an attractive target for investors. It has recorded the largest investment realization among secondary sectors in Indonesia over the last five years, totaling IDR 293 trillion[2]. According to a research by Toffin[3], the coffee shop has been a booming F&B business in Indonesia, reflected on the significant rise in number of outlets and domestic coffee consumptions in the recent years[4]. The market value of coffee shops is also estimated to reach over IDR 4 trillion per year.

### 1.2. Problem Statement

With the aforementioned prospect, various stakeholders (entrepreneurs, investors) may be interested to explore coffee shop business opportunities in Jakarta. This data science project is thus carried out to help them answer the following question:

"*Which of the Jakarta regions are strategic for opening a coffee shop business*?"

Apart from business stakeholders, the project may also be of interest to fellow coffee enthusiasts.

## 2. Data

In order to explore potential answer to the problems, the following data are required:

1. **The names of administrative regions in Jakarta and their corresponding postal codes.** The regions include three levels of subdivision: city, district, subdistrict. The information was scraped from a directory on indonesiapostcode.com. The region names are useful to perform analysis across different sub-regions. The postal codes are needed to obtain coordinates of each subdistricts.
2. **Geographical coordinates** of Jakarta and its subdistricts, which will in turn be needed to utilize Foursquare API in the subsequent step. Coordinates are obtained using Nominatim geocoder from the GeoPy library.
3. Information about venues in Jakarta regions: the names, category, venue latitudes, venue longitudes. These are obtained using Foursquare API. The subdistricts of Jakarta will be clustered based on their surrounding venues to find the best location candidates for opening a coffee shop.

[2] "Food Industry Can Weather Global Economic Shock: BKPM". The Jakarta Post. 27 May 2020.
[3] "The Emerging Business of Coffee Shops in Indonesia". Now! Jakarta. 5 January 2020.
[4] United States Department of Agriculture. Indonesia Coffee Annual Report 2019.

## 3. Methodology

Web scraping was performed to extract data of Jakarta regions and postal codes as well as retrieval of geographical coordinates. Leveraging Foursquare API, these coordinates data were given as inputs to explore venues within the Jakarta subdistricts. Two dataframes were then created for use in the analysis:

1. ***post***: contains postal codes and geographical coordinates of all Jakarta regions (city, district, subdistrict).
2. ***jakvenues***: contains at most 100 venues and venues details (name, category, latitude, longitude) for every subdistrict in Jakarta.

One-hot encoding was performed to analyze and to narrow down the most common venues in each of the subdistricts. Given all the venues surrounding them, subdistricts were clustered using *K*-means algorithm. The number of optimal clusters was decided using the elbow method and silhouette score. Each cluster was separately analyzed in order to examine one discriminating venue that characterizes them. Analysis of the clusters and visualization of coffee shop distribution across Jakarta would provide insights as to where the strategic regions to set up the business are.

The following Python libraries and dependencies were used: pandas, NumPy, string, Requests, time.sleep, BeautifulSoup, GeoPy (Nominatim geocoder), JSON, Folium, Matplotlib, and scikit-learn.

### 3.1. Jakarta Regions, Postal Codes, and Geographical Coordinates

The data to scrape are the names of all Jakarta regions and their corresponding postal codes. For reference, Jakarta is a province that consists of 5 cities (mainland Jakarta) and 1 regency:

1. Jakarta Pusat (Central Jakarta)
2. Jakarta Utara (North Jakarta)
3. Jakarta Barat (West Jakarta)
4. Jakarta Selatan (South Jakarta)
5. Jakarta Timur (East Jakarta)
6. Kepulauan Seribu (Thousand Islands Regency)

Each of these cities is further subdivided into districts (kecamatan) and then subdistricts (kelurahan). In total, there are 44 districts and 267 subdistricts across Jakarta. Jakarta Selatan and Jakarta Timur are tied as the cities with the highest number of sub-regions, each having 65 subdistricts.

Using the scraped postal codes as inputs, the Nominatim geocoder were used to retrieve latitudes and longitudes of every subdistrict. Figure 1 displays the first 10 rows of the resulting dataframe: ***post***.

| | Postal_Code | City | District | Subdistrict | Latitude | Longitude |
|---|---|---|---|---|---|---|
| 0 | 10210 | Jakarta Pusat | Tanah Abang | Bendungan Hilir | -6.205150 | 106.813743 |
| 1 | 10460 | Jakarta Pusat | Senen | Bungur | -6.175394 | 106.827183 |
| 2 | 10640 | Jakarta Pusat | Kemayoran | Cempaka Baru | -6.168070 | 106.863965 |
| 3 | 10520 | Jakarta Pusat | Cempaka Putih | Cempaka Putih Barat | -6.176444 | 106.860497 |
| 4 | 10510 | Jakarta Pusat | Cempaka Putih | Cempaka Putih Timur | -6.190382 | 106.856887 |
| 5 | 10150 | Jakarta Pusat | Gambir | Cideng | -6.167774 | 106.812574 |
| 6 | 10330 | Jakarta Pusat | Menteng | Cikini | -6.190742 | 106.838564 |
| 7 | 10140 | Jakarta Pusat | Gambir | Duri Pulo | -6.165196 | 106.810215 |
| 8 | 10530 | Jakarta Pusat | Johar Baru | Galur | -6.174859 | 106.856765 |
| 9 | 10110 | Jakarta Pusat | Gambir | Gambir | -6.176077 | 106.826741 |

Figure 1. Dataframe *post*, containing postal codes and geographical coordinates of Jakarta regions.

### 3.2. Exploring Venues in the Subdistricts of Jakarta

A total of 14739 venues were collected through API calls to Foursquare that were made using a user-defined function. The result (Figure 2) is a dataframe containing a maximum of 100 venues within 1 km of a region (i.e., the center of a subdistrict), with the following details: venue name, venue category, venue latitude, venue longitude. On average, there are 55 venues per subdistrict (Figure 3).

| | City | District | Subdistrict | Latitude | Longitude | Venue | Category | Venue_Lat | Venue_Lng |
|---|---|---|---|---|---|---|---|---|---|
| 0 | Jakarta Pusat | Tanah Abang | Bendungan Hilir | -6.20515 | 106.813743 | RM Sederhana | Indonesian Restaurant | -6.205342 | 106.816817 |
| 1 | Jakarta Pusat | Tanah Abang | Bendungan Hilir | -6.20515 | 106.813743 | Harum Manis Indonesian Restaurant | Indonesian Restaurant | -6.207540 | 106.817653 |
| 2 | Jakarta Pusat | Tanah Abang | Bendungan Hilir | -6.20515 | 106.813743 | KLTR | Coffee Shop | -6.207718 | 106.817554 |
| 3 | Jakarta Pusat | Tanah Abang | Bendungan Hilir | -6.20515 | 106.813743 | Shangri-La Hotel, Jakarta | Hotel | -6.203055 | 106.819033 |
| 4 | Jakarta Pusat | Tanah Abang | Bendungan Hilir | -6.20515 | 106.813743 | SATOO | Restaurant | -6.203429 | 106.818783 |
| 5 | Jakarta Pusat | Tanah Abang | Bendungan Hilir | -6.20515 | 106.813743 | Horizon Lounge | Lounge | -6.203015 | 106.819039 |
| 6 | Jakarta Pusat | Tanah Abang | Bendungan Hilir | -6.20515 | 106.813743 | Papaya Fresh Galery | Grocery Store | -6.208744 | 106.818305 |
| 7 | Jakarta Pusat | Tanah Abang | Bendungan Hilir | -6.20515 | 106.813743 | AYANA MidPlaza Jakarta | Hotel | -6.209002 | 106.819677 |
| 8 | Jakarta Pusat | Tanah Abang | Bendungan Hilir | -6.20515 | 106.813743 | Cassis Kitchen | Restaurant | -6.207437 | 106.817549 |
| 9 | Jakarta Pusat | Tanah Abang | Bendungan Hilir | -6.20515 | 106.813743 | Tucano's Churrasco Brazilian Bbq | Brazilian Restaurant | -6.208185 | 106.817631 |

Figure 2. Dataframe *jakvenues*, containing venues information across Jakarta subdistricts.

| | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| **Venue Count** | 267.0 | 55.202247 | 34.397558 | 1.0 | 24.5 | 50.0 | 92.0 | 100.0 |

Figure 3. Statistical distribution of venue counts in every Jakarta subdistricts.

### 3.3. Most Common Venues Overall

As shown in Figure 4, various kinds of restaurant top the list of most common venues in Jakarta. Coffee shop, which is our venue of interest, comes in second. With almost 1000 coffee shops in Jakarta, it sure is a quite competitive business.

| | Category | Count |
|---|---|---|
| 0 | Indonesian Restaurant | 999 |
| 1 | Coffee Shop | 997 |
| 2 | Fast Food Restaurant | 705 |
| 3 | Asian Restaurant | 681 |
| 4 | Noodle House | 542 |
| 5 | Chinese Restaurant | 537 |
| 6 | Hotel | 517 |
| 7 | Convenience Store | 515 |
| 8 | Café | 384 |
| 9 | Food Truck | 359 |

Figure 4. Top 10 most common venues in all of Jakarta.

| | Subdistrict | Accessories Store | Acehnese Restaurant | African Restaurant | Airport | Airport Lounge | Airport Terminal | American Restaurant | Antique Shop | Aquarium | ... |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Ancol | 0.00000 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.014706 | ... |
| 1 | Angke | 0.01087 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.000000 | ... |
| 2 | Balekambang | 0.00000 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.000000 | ... |
| 3 | Bali Mester | 0.00000 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.000000 | ... |
| 4 | Bambu Apus | 0.00000 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.000000 | ... |
| 5 | Bangka | 0.00000 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.033333 | 0.0 | 0.000000 | ... |
| 6 | Baru | 0.01087 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.000000 | ... |
| 7 | Batu Ampar | 0.00000 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.000000 | ... |
| 8 | Bendungan Hilir | 0.00000 | 0.01 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | 0.000000 | ... |
| 9 | Bidara Cina | 0.00000 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.023256 | 0.0 | 0.000000 | ... |

Figure 5. One-hot encoding, showing the mean frequencey of venue occurrence in each subdistrict.

### 3.4. One-Hot Encoding

One-hot encoding converts categorical variables (i.e., venues) into numeric variables. In this case, a dummy of all the venues was made and the mean of the frequency of venue occurrence were calculated. The dataframe is then grouped by subdistrict, as

shown in Figure 5. The data were then filtered with a user-defined function to obtain 5 most common venues in each subdistrict, as displayed in Figure 6.

| | Subdistrict | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|
| 0 | Ancol | Theme Park Ride / Attraction | Coffee Shop | Theme Park | Hotel | Seafood Restaurant |
| 1 | Angke | Indonesian Restaurant | Coffee Shop | Fast Food Restaurant | Hotel | Asian Restaurant |
| 2 | Balekambang | Fast Food Restaurant | Mediterranean Restaurant | Restaurant | Bakery | Gym |
| 3 | Bali Mester | Convenience Store | Fast Food Restaurant | Asian Restaurant | Jewelry Store | Donut Shop |
| 4 | Bambu Apus | Convenience Store | Indonesian Restaurant | Garden | Soup Place | High School |
| 5 | Bangka | Restaurant | Coffee Shop | Bar | Bistro | American Restaurant |
| 6 | Baru | Indonesian Restaurant | Coffee Shop | Fast Food Restaurant | Hotel | Asian Restaurant |
| 7 | Batu Ampar | Fast Food Restaurant | Donut Shop | Shopping Mall | Cupcake Shop | Movie Theater |
| 8 | Bendungan Hilir | Coffee Shop | Japanese Restaurant | Indonesian Restaurant | Pizza Place | Lounge |
| 9 | Bidara Cina | Convenience Store | Hotel | Fast Food Restaurant | Café | Food Truck |

Figure 6. Five most common venues in each of the Jakarta subdistricts.

### 3.5. Clustering Subdistricts Based on Venues Similarity

The subdistricts were clustered based on a set of similar characteristics or features, i.e., their surrounding venues. *K*-Means clustering, which was used in this part of the analysis, is a machine learning algorithm that creates homogeneous subgroups/clusters from unlabeled data such that data points in each cluster are as similar as possible to each other according to a similarity measure (e.g., Euclidian distance).

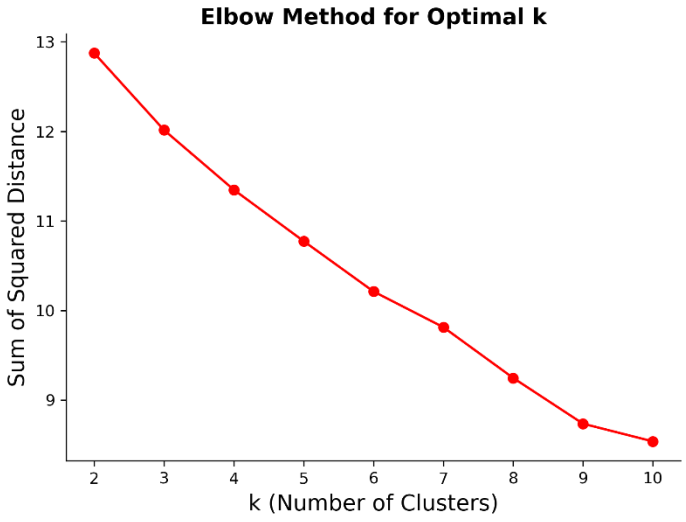


Figure 7. Elbow method: sum of squared distances for different *k* values.

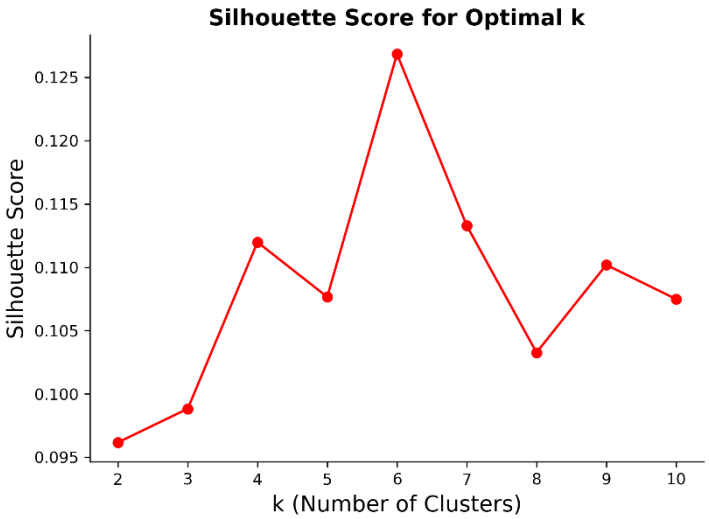


Figure 8. Silhouette scores across different values of *k*.

| | Postal_Code | City | District | Subdistrict | Latitude | Longitude | Cluster | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 10210 | Jakarta Pusat | Tanah Abang | Bendungan Hilir | -6.205150 | 106.813743 | 5 | Coffee Shop | Japanese Restaurant | Indonesian Restaurant | Pizza Place | Lounge |
| 1 | 10460 | Jakarta Pusat | Senen | Bungur | -6.175394 | 106.827183 | 5 | Indonesian Restaurant | Coffee Shop | Fast Food Restaurant | Hotel | Asian Restaurant |
| 2 | 10640 | Jakarta Pusat | Kemayoran | Cempaka Baru | -6.168070 | 106.863965 | 5 | Indonesian Restaurant | Noodle House | Coffee Shop | Restaurant | Hotel |
| 3 | 10520 | Jakarta Pusat | Cempaka Putih | Cempaka Putih Barat | -6.176444 | 106.860497 | 5 | Indonesian Restaurant | Indonesian Meatball Place | Hotel | Café | Convenience Store |
| 4 | 10510 | Jakarta Pusat | Cempaka Putih | Cempaka Putih Timur | -6.190382 | 106.856887 | 1 | Convenience Store | Fast Food Restaurant | Seafood Restaurant | Pharmacy | Indonesian Restaurant |
| 5 | 10150 | Jakarta Pusat | Gambir | Cideng | -6.167774 | 106.812574 | 5 | Indonesian Restaurant | Coffee Shop | Chinese Restaurant | Hotel | Asian Restaurant |
| 6 | 10330 | Jakarta Pusat | Menteng | Cikini | -6.190742 | 106.838564 | 5 | Indonesian Restaurant | Hotel | Coffee Shop | Fast Food Restaurant | Café |
| 7 | 10140 | Jakarta Pusat | Gambir | Duri Pulo | -6.165196 | 106.810215 | 2 | Indonesian Restaurant | Fast Food Restaurant | Chinese Restaurant | Hotel | Convenience Store |
| 8 | 10530 | Jakarta Pusat | Johar Baru | Galur | -6.174859 | 106.856765 | 5 | Indonesian Restaurant | Hotel | Indonesian Meatball Place | Café | Chinese Restaurant |
| 9 | 10110 | Jakarta Pusat | Gambir | Gambir | -6.176077 | 106.826741 | 5 | Indonesian Restaurant | Coffee Shop | Asian Restaurant | Hotel | Café |

Figure 9. A merged dataframe with cluster labels for every subdistrict.

A value of *k* (number of clusters) needs to be defined before proceeding with the clustering. The “Elbow Method” was used, which calculates the sum of squared distances of data points to their closest centroid (cluster center) for different values of *k*. The optimal value of *k* is the one after which there is a plateau (no significant decrease in sum of squared distances). However, because there is no discernible "elbow" from the plot (Figure 7), another measure was used: “Silhouette Score”. Silhouette score varies from -1 to 1. A score value of 1 means the cluster is dense and well-separated from other clusters. A value nearing 0 represents overlapping clusters, data points are close to the decision boundary of neighboring clusters. A negative score indicates that the samples might have been assigned into the wrong clusters. Given that there is a peak at *k* = 6 (Figure 8), the *K*-Means clustering was proceeded with that value.

After each subdistrict had been assigned a cluster label (Figure 9), the clusters were color-coded and visualized on a map of Jakarta (Figure 10) to understand how they are distributed across the regions.

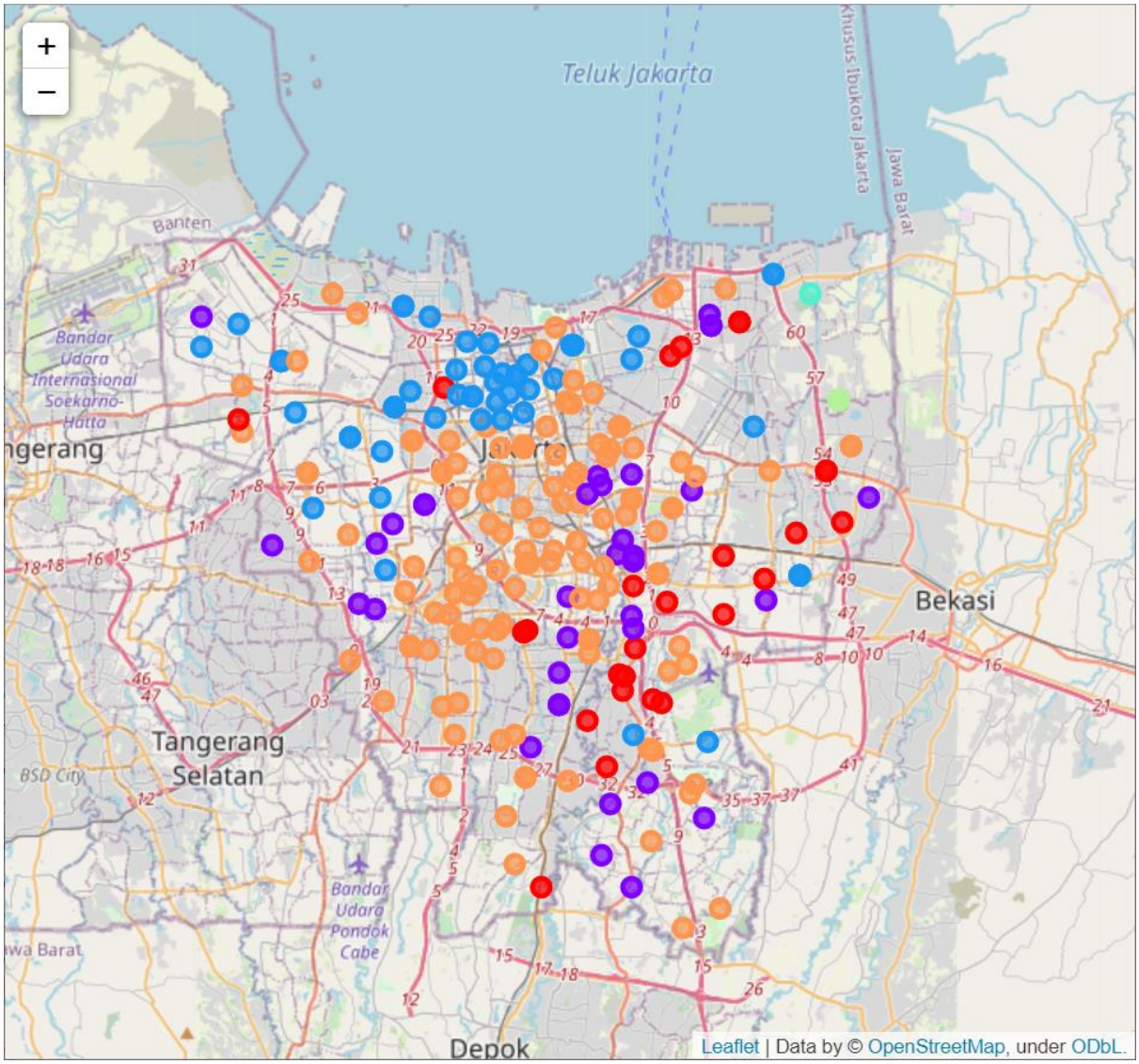


Figure 10. Clusters of Jakarta subdistricts based on similarity of venues.

### 3.6. Examining Each Cluster

Each cluster was filtered from the dataframe previously created in the clustering stage. The clusters were separately analyzed in order to gain an understanding of a discriminating venue that characterize each of them. The number one most common venue categories from each cluster, as well as the regions (cities) in which a particular cluster is highly concentrated were singled out.

| | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|
| count | 27 | 27 | 27 | 27 | 27 |
| unique | 6 | 13 | 14 | 19 | 23 |
| top | Fast Food Restaurant | Fast Food Restaurant | Food Truck | Indonesian Restaurant | Indonesian Restaurant |
| freq | 19 | 7 | 6 | 4 | 2 |

Figure 11. Most common venue, Cluster 0 (Red): Fast food restaurant.

| | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|
| count | 35 | 35 | 35 | 35 | 35 |
| unique | 9 | 14 | 19 | 24 | 20 |
| top | Convenience Store | Fast Food Restaurant | Fast Food Restaurant | Jewelry Store | Asian Restaurant |
| freq | 21 | 10 | 5 | 3 | 4 |

Figure 12. Most common venue, Cluster 1 (Purple): Convenience store.

| | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|
| count | 51 | 51 | 51 | 51 | 51 |
| unique | 6 | 12 | 15 | 18 | 20 |
| top | Chinese Restaurant | Noodle House | Asian Restaurant | Convenience Store | Asian Restaurant |
| freq | 25 | 17 | 14 | 8 | 5 |

Figure 13. Most common venue, Cluster 2 (Blue): Chinese restaurant .

| | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|
| count | 1 | 1 | 1 | 1 | 1 |
| unique | 1 | 1 | 1 | 1 | 1 |
| top | Noodle House | Accessories Store | Outdoors & Recreation | Pet Store | Pet Service |
| freq | 1 | 1 | 1 | 1 | 1 |

Figure 14. Most common venue, Cluster 3 (Cyan): Noodle house

| | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|
| count | 2 | 2 | 2 | 2 | 2 |
| unique | 1 | 1 | 1 | 1 | 1 |
| top | Paper / Office Supplies Store | Donut Shop | Arcade | Asian Restaurant | Accessories Store |
| freq | 2 | 2 | 2 | 2 | 2 |

Figure 15. Most common venue, Cluster 4 (Light Green): Paper / office supplies store.

| | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|
| count | 151 | 151 | 151 | 151 | 151 |
| unique | 21 | 32 | 40 | 45 | 44 |
| top | Indonesian Restaurant | Coffee Shop | Fast Food Restaurant | Hotel | Asian Restaurant |
| freq | 64 | 41 | 33 | 39 | 36 |

Figure 16. Most common venue, Cluster 5 (Orange): Indonesian restaurant

Table 1. Concentration of cluster members in the cities of Jakarta.

| Cities | Cluster 0 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 |
|---|---|---|---|---|---|---|
| Jakarta Pusat | | 4 | 8 | | | 32 |
| Jakarta Utara | 4 | 3 | 9 | 1 | 2 | 12 |
| Jakarta Barat | 2 | 6 | 29 | | | 19 |
| Jakarta Selatan | 4 | 8 | | | | 53 |
| Jakarta Timur | 17 | 14 | 5 | | | 29 |
| Kepulauan Seribu | | | | | | 6 |
| **Total Data Points** | 27 | 35 | 51 | 1 | 2 | 151 |

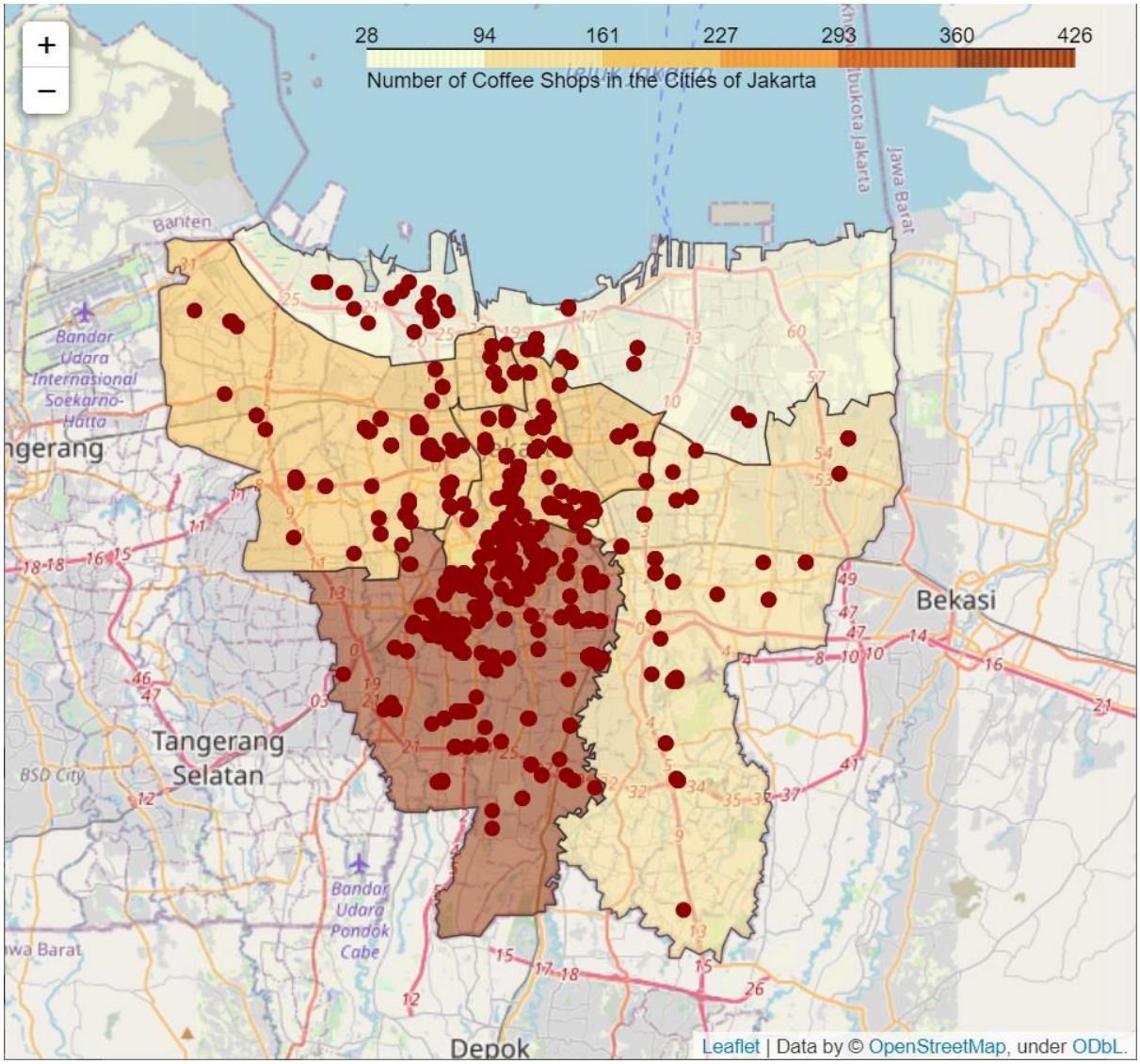


Figure 17. Concentrations of coffee shops across Jakarta.

### 3.7. Visualizing Distribution of Coffee Shop Locations in Jakarta

Based on data in the *jakvenues* dataframe, the total number of coffee shops within each of the Jakarta cities and districts were calculated to examine the distribution of coffee

shop businesses and to help figure out strategic locations. This distribution was visualized in the choropleth map above (Figure 17).

| | City | Coffee Shop |
|---|---|---|
| 0 | Jakarta Selatan | 426.0 |
| 1 | Jakarta Pusat | 189.0 |
| 2 | Jakarta Barat | 172.0 |
| 3 | Jakarta Timur | 112.0 |
| 4 | Jakarta Utara | 70.0 |
| 5 | Kepulauan Seribu | 28.0 |

Figure 18. Number of coffee shops in Jakarta cities.

| | District | Coffee Shop | City |
|---|---|---|---|
| 0 | Setiabudi | 108 | Jakarta Selatan |
| 1 | Kebayoran Baru | 101 | Jakarta Selatan |
| 2 | Tanah Abang | 60 | Jakarta Pusat |
| 3 | Pancoran | 41 | Jakarta Selatan |
| 4 | Tambora | 41 | Jakarta Barat |
| 5 | Tebet | 36 | Jakarta Selatan |
| 6 | Cilandak | 35 | Jakarta Selatan |
| 7 | Grogol Petamburan | 34 | Jakarta Barat |
| 8 | Menteng | 33 | Jakarta Pusat |
| 9 | Taman Sari | 33 | Jakarta Barat |

| | District | Coffee Shop | City |
|---|---|---|---|
| 34 | Duren Sawit | 7 | Jakarta Timur |
| 35 | Kepulauan Seribu Selatan | 7 | Kepulauan Seribu |
| 36 | Kalideres | 6 | Jakarta Barat |
| 37 | Kelapa Gading | 6 | Jakarta Utara |
| 38 | Johar Baru | 6 | Jakarta Pusat |
| 39 | Cempaka Putih | 4 | Jakarta Pusat |
| 40 | Cakung | 2 | Jakarta Timur |
| 41 | Kramat Jati | 2 | Jakarta Timur |
| 42 | Cengkareng | 2 | Jakarta Barat |
| 43 | Koja | 0 | Jakarta Utara |

Figure 19. Districts with the highest (left) and lowest (right) count of coffee shops.

## 4. Results and Discussion

Exploratory data analysis as well as machine learning and visualization techniques have provided us with some insights into the problem at hand.

A total of 14739 venues from all Jakarta regions (267 subdistricts) were returned at the time the API call was made. There are on average 55 venues within a kilometer of a subdistrict center, where two of the most common categories overall are Indonesian Restaurants and Coffee Shops.

After deciding on an optimal *k* value of 6, *K*-Means algorithm was run to cluster the subdistricts based on their most common surrounding venues. Each of the six clusters, labeled 0-5, is characterized by a dominant venue as follows:

Table 2. Results of K-Means clustering.

| Cluster | Member | Common Venues | Concentrated Region |
|---|---|---|---|
| 0 | 27 | Fast Food Restaurant | Jakarta Timur |
| 1 | 35 | Convenience Store | Jakarta Timur |
| 2 | 51 | Chinese Restaurant | Jakarta Barat |
| 3 | 1 | Noodle House | Jakarta Utara |
| 4 | 2 | Office Supplies Store | Jakarta Utara |
| 5 | 151 | Indonesian Restaurant, Coffee Shop | Jakarta Selatan, Jakarta Pusat |

A considerable number of coffee shops can be found within Cluster 5 (41 shops out of 151 venues). In fact, it is the second-most common venues in that cluster. Choropleth map of coffee shop locations across mainland Jakarta shows that Jakarta Selatan has a very high concentration of the business, i.e., 426 shops while the rest are below 200. The districts in Jakarta Selatan, therefore, are not viable options for opening up a coffee shop business because they are already way too saturated.

It is recommended that stakeholders look into opportunities in Jakarta Timur (e.g., Cakung, Kramat Jati) and Jakarta Utara (e.g., Kelapa Gading), as these two cities have the least concentration of coffee shops and would significantly minimize competition. If, however, a moderate competition is not a concern then districts in Jakarta Pusat (e.g., Cempaka Putih, Johar Baru) and Jakarta Barat (e.g., Kalideres, Cengkareng) are also recommended.

## 5. Conclusion

Stakeholders searching for opportunities to open a coffee shop in Jakarta may want to consider setting up their business someplace where competitions are not severe. All of Jakarta sub-regions were explored and then clustered based on the similarity of their surrounding venues using *K*-Means clustering algorithm. Analysis results show that districts in Jakarta Utara and Jakarta Timur are among the best candidates for a new coffee shop location.