Opening Coffee Shop in Jakarta Using Data Analysis Techniques

Budi Irawan Saleh

Abstract

The coffee shop has become an emerging business, especially in a growing business city. Jakarta is one of the growing capital business city in Indonesia, and opening a coffee shop will be a very profitable business. However, the business of coffee shops has been trending, and a new coffee shop has opened in this city. It will require a good analysis of where to locate the coffee shop not just to avoid a highly competitive market but also to reach a higher maximum profit. This research will use the data analysis techniques, which have been thought in the Coursera capstone course. The data of the neighborhood was collected using a web scraping technique to Wikipedia. Then the geo-location of the neighborhood was obtained by using Google API (google maps) and mapped to the Folium map. The venues in the neighborhood collected using Foursquare API and clustered (using K-mean) and analyzed based on the numbers of coffee shops in each cluster. K-mean cluster using the elbow method was found for 2 clusters. Based on the data analysis, the coffee shop can be open in the second cluster because the data analysis showed that numbers of coffee shops in the area are still small and can reduce the competition. However, it also found that the second cluster has the smallest number of area and have a potential smaller profit if the coffee shop is open in the first cluster.

Keywords: *Coffee Shops, Jakarta, Data Analysis, K-mean clustering*

I. INTRODUCTION

Indonesia has become one of the attractive business markets in the world. It has been forecasted by the end of 2024 will be among the highest GDP country in the world. Indonesia also has the most advantages in the demographic (demographic bonuses). Population census data in 2018 shows that populations that are in productive age (15-64 years old) reached 179.13 million people or around 67.6% of Indonesia's total population. Indonesia's most populated city is in Jakarta, where people in the archipelago are gathering in this city for business and government.

Moreover, it holds almost 20% of the country's population. Jakarta is not just a capital city but also a business, education, and travel. It holds almost 80% of the capital movement in the country.

Small enterprise business (SME) is emerging and creates high development in Indonesia, especially in Jakarta. SME's in Indonesia is dominated by a culinary business (i.e : restaurants, coffee shops, bar, etc.). The trends of opening the culinary business are rising because the number of populations is also increased. Jakarta, which has become the most metropolitan city in the country, has also impacted to raised of this SME. The raised of the middle income in Jakarta has increased the trending of the culinary business (i.e.: Coffee shops, bars, etc.)

Coffee shops have become trends, especially in middle-income workers, to hang out after the office and to await a time to go home. It was also used as a place to do a business transaction, working and spending time with family. Coffee shops are business that was the highest attraction by the SME to open due to their high profitably not mention a short time of ROI. Jakarta is estimated to have more than 100 SME's that used coffee shops as their business.

Data analysis has been used significantly to assist business decision and planning for years. Research using these methods has increased significantly. More business has used the methods both supervised or unsupervised learning to help their business.

The objective of This paper is to find the best locations for coffee shops in Jakarta and expecting less competition, among others.

II. METHODOLOGY

Data on the locations of Jakarta, especially its borough area, need to be collected from the trusted place, Wikipedia is the place to gather the data. Web scraping technique is used to gather the borough data and placed it into a panda data frame.

Importing the necessary libraries

In [1]:
```
# import necessary libraries
import pandas as pd
import numpy as np
from bs4 import BeautifulSoup
import requests
import urllib.request
```

extracting table from wikipedia for borough in the City

In [2]:
```
data = requests.get("https://en.wikipedia.org/wiki/List_of_Jakarta_Subdistricts").text
```

In [3]:
```
# parse data from the html into a beautifulsoup object
soup = BeautifulSoup(data, 'html.parser')
```

In [4]:
```
# create a list to store neighborhood data
Borough = []
```

In [5]:
```
# append the data into the list
for row in soup.find_all("div", class_="mw-parser-output")[0].findAll("li"):
    Borough.append(row.text)
```

In [6]:
```
# create a new DataFrame from the list
df = pd.DataFrame({"Borough":Borough})

df
```

Figure 1. Web Scrapping from Wikipedia

Geospatial location was then obtained using the Google API methods. And to check the location accuracy by plotting the data using the Folium map.

In [14]:
```
import googlemaps
```

In [15]:
```
gmaps_key = googlemaps.Client(key = "AIzaSyAxPJUAU4xQsSBF9UJtJrCDCSRNPG57W3k")
```

In [16]:
```
dataset["Latitude"] = None
dataset["Longitude"] = None
for i in range (0, len(dataset),1 ):
    geocode_result =gmaps_key.geocode(df.iat[i,0])
    try:
        lat = geocode_result[0]["geometry"]["location"]["lat"]
        long = geocode_result[0]["geometry"]["location"]["lng"]
        dataset.iat[i, dataset.columns.get_loc("Latitude")] = lat
        dataset.iat[i, dataset.columns.get_loc("Longitude")] = long
    except:
        lat = None
        long = None
```

Figure 2. Getting the latitude and longitude using Google API

In [43]:
```
#  center map on mean of Latitude/Longitude
map_world = folium.Map(location=[-6.22818 , 106.934385], zoom_start = 11)

#  add locations to map
for lat, lng, label in zip(dataloc.Latitude, dataloc.Longitude, dataloc.Borough):
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        fill=True,
        color='Blue',
        fill_color='Yellow',
        fill_opacity=0.6
        ).add_to(map_world)

#  display interactive map
map_world
```

Figure 3. Mapping the location using the Folium Map

The venue then collected using Foursquare API and placed into the panda data frame.

Use the Foursquare API to explore the neighborhoods

In [25]:
```
import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe
```

In [26]:
```
# define Foursquare Credentials and Version
CLIENT_ID = 'JB3NHQWKPKI14RWWLY3R0GMKF3X04NTN4KNB04R3J10G2V0X' # your Foursquare ID
CLIENT_SECRET = 'W5JL4J5T0JS2WEZOMWHK1KSKUMYURD44VCCXIXAQBHOTZJEB' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version

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

```
Your credentials:
CLIENT_ID: JB3NHQWKPKI14RWWLY3R0GMKF3X04NTN4KNB04R3J10G2V0X
CLIENT_SECRET:W5JL4J5T0JS2WEZOMWHK1KSKUMYURD44VCCXIXAQBHOTZJEB
```

In [27]:
```
radius = 2000
LIMIT = 100

venues = []

for lat, long, neighborhood in zip(dataloc['Latitude'], dataloc['Longitude'], dataloc['Borough']):

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

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

    # return only relevant information for each nearby venue
    for venue in results:
        venues.append((
            neighborhood,
            lat,
            long,
            venue['venue']['name'],
            venue['venue']['location']['lat'],
            venue['venue']['location']['lng'],
            venue['venue']['categories'][0]['name']))
```

Figure 4. Using Foursquare API to get the venue list

The K-mean then used to cluster the location based on geospatial location. The cluster number was obtained using the elbow methods.
Data then analyzed to obtain conclusions for the location of the coffee shops.

III. DISCUSSIONS

From the web scrapping technique from Wikipedia, the borough data is stored in the data frame as follow :

Out[12]:

	Borough
0	Angke
1	Cakung
2	Cempaka Putih
3	Cengkareng
4	Cilandak
5	Cipayung
6	Ciracas
7	Duren Sawit
9	Duri Kepa
10	East Cengkareng
11	Gambir, Jakarta
12	Grogol

Figure 5. Borough Data from Wikipedia

The data of geospatial location then obtained using Google Map API as follow :

Out[21]:

	Borough	Latitude	Longitude
0	Angke	41.493662	-81.704854
1	Cakung	-6.182629	106.947666
2	Cempaka Putih	-6.182671	106.867990
3	Cengkareng	-6.148665	106.735258
4	Cilandak	-6.284528	106.800140
5	Cipayung	-6.327251	106.900447
6	Ciracas	-6.323116	106.870940
7	Duren Sawit	-6.232191	106.915202
8	Duri Kepa	-6.232191	106.915202
9	East Cengkareng	-6.169185	106.774860
10	Gambir, Jakarta	-6.142332	106.733784
11	Grogol	-6.170340	106.814805

Figure 6. Geospatial Location

Plotted using folium map to check the locations, and it was found that no data error and all area is covered

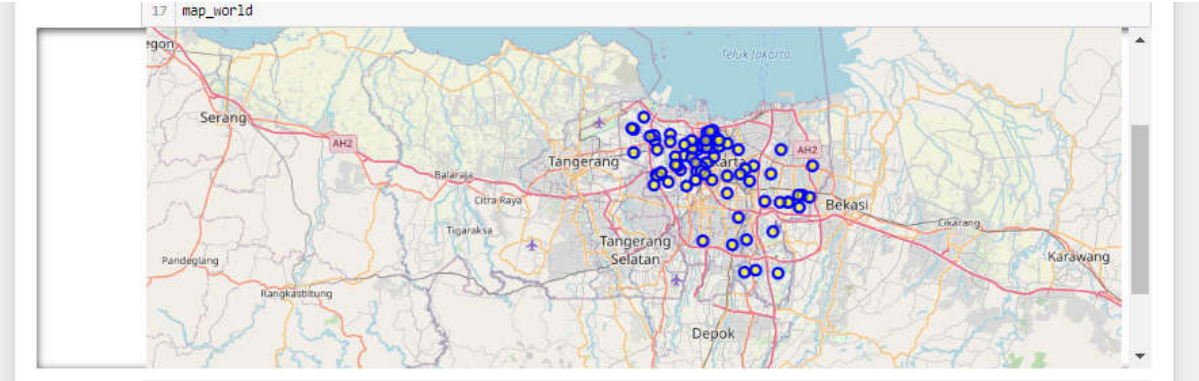
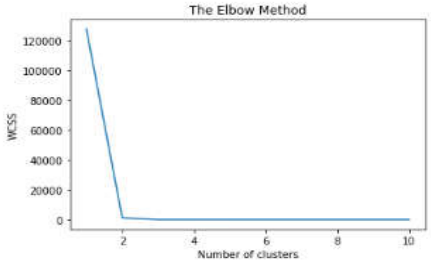


Figure 6. Map to folium map

The foursquare methods then used to obtain the venues in the location and put into the data frame, and it was found that a total of 272 uniques venues is at Jakarta. The K-mean clustering methods are used to cluster based on the location of the borough. The elbow method then used to obtain a sufficient clustering number. Elbow methods found that 2 clusters need to be used for clustering.

Using the elbow method to get the K mean clustering value

In [56]:
```
from sklearn.cluster import KMeans
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state = 42)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
plt.plot(range(1, 11), wcss)
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
```



From the elbow methods it was found that the Kcluster value is - 2

Figure 7. The elbow method

The data then plotted based on the cluster to check the logic of the cluster number. And have found no error in the clustering number.

Analyzing the cluster value

In [59]:
```
plt.scatter(X[y_kmeans1 == 0, 0], X[y_kmeans1 == 0, 1], s = 100, c = 'red', label = 'Cluster 1')
plt.scatter(X[y_kmeans1 == 1, 0], X[y_kmeans1 == 1, 1], s = 100, c = 'blue', label = 'Cluster 2')
plt.title('Clusters Area')
plt.xlabel('Latitude')
plt.ylabel('Longitude')
plt.legend()
plt.show()
```

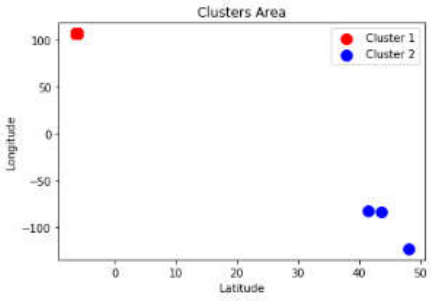


Figure 8. Clustering Plot

The clustered data then analyzed from the five most venues in the area based on the Foursquare data. And from the data it was found that the second cluster has fewer coffee shops in the area as can be seen in the bellow table :

Borough	Longitude	Cluster_Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Angke	-81.7049	0	Brewery	Coffee Shop	Bar	Hotel	Gym / Fitness Center
Cakung	106.9477	0	Fast Food Restaurant	Department Store	Coffee Shop	Theme Park	Asian Restaurant
Cempaka Putih	106.868	0	Indonesian Restaurant	Coffee Shop	CafÃ©	Convenience Store	Restaurant
Cengkareng	106.7353	0	Asian Restaurant	Noodle House	Indonesian Restaurant	Chinese Restaurant	Pizza Place

Borough	Longitude	Cluster_Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Cilandak	106.8001	0	Coffee Shop	Indonesian Restaurant	CafÃ©	Padangnese Restaurant	Gym
Cipayung	106.9004	0	Indonesian Restaurant	Convenience Store	Food & Drink Shop	Fast Food Restaurant	Asian Restaurant
Ciracas	106.8709	0	Convenience Store	Noodle House	Pizza Place	Fast Food Restaurant	Arcade
Duren Sawit	106.9152	0	Fast Food Restaurant	Indonesian Restaurant	Indonesian Meatball Place	Asian Restaurant	Bakery
Duri Kepa	106.9152	0	Restaurant	Indonesian Restaurant	Indonesian Meatball Place	Asian Restaurant	Bakery
East Cengkareng	106.7749	0	Noodle House	Chinese Restaurant	Restaurant	Indonesian Restaurant	Convenience Store
Grogol	106.8148	0	Noodle House	Indonesian Restaurant	Coffee Shop	Chinese Hotel	Coffee Shop
Jalembar	106.7928	0	Coffee Shop	Noodle House	Chinese Restaurant	Indonesian Restaurant	Soup Place
Jembatan Besi	106.8827	0	Asian Restaurant	Fast Food Restaurant	Indonesian Restaurant	Coffee Shop	CafÃ©
Jembatan Lima	106.8046	0	Coffee Shop	Indonesian Restaurant	Dessert Shop	Soup Place	Shopping Mall
Kali Anyar	-122.756	0	Hotel	Thai Restaurant	Seafood Restaurant	Coffee Shop	American Restaurant
Kalideres	106.8577	0	Indonesian Restaurant	Pizza Place	Coffee Shop	Hotel	Asian Restaurant
Kebon Jeruk, Jakarta Barat	106.7691	0	Noodle House	Indonesian Restaurant	Seafood Restaurant	Coffee Shop	Asian Restaurant
Kedaung Kali Angke	106.7559	0	Coffee Shop	Flea Market	Convenience Store	Bank	Seafood Restaurant
Kelapa Gading	106.7647	0	Noodle House	Indonesian Restaurant	Coffee Shop	Seafood Restaurant	Chinese Restaurant
Kemanggisan	106.7647	0	Noodle House	Indonesian Restaurant	Coffee Shop	Seafood Restaurant	Chinese Restaurant
Kemayoran	106.7559	0	Coffee Shop	Flea Market	Fast Food Restaurant	Restaurant Pizza Place	Convenience Store
Kembangan, Jakarta	106.7364	0	Coffee Shop	Massage Studio	Indonesian Restaurant	Indonesian Multiplex	Clothing Store
Klender	106.9055	0	Noodle House	Coffee Shop	Restaurant	Asian Restaurant	Dessert Shop
Kota Bambu	106.7913	0	Coffee Shop	Dessert Shop	Hotel	Chinese Clothing Store	Chinese Restaurant
Krendang	106.7426	0	Coffee Shop	Indonesian Restaurant	Noodle House	Asian Restaurant	CafÃ©
Krukut	-83.9019	0	Fast Food Restaurant	Convenience Store	Department Store	Sandwich Place	Mobile Phone Shop
Makasar, Jakarta	106.7987	0	Coffee Shop	Dessert Shop	Clothing Store	Chinese Restaurant	Hotel
Malaka Jaya	106.8591	0	Restaurant	Food Truck	Chinese Restaurant	Pizza Place	Middle Eastern Restaurant
Maphar	106.8945	0	Coffee Shop	Indonesian Restaurant	Fast Food Restaurant	Golf Course	Seafood Restaurant
Matraman	106.9344	0	Indonesian Restaurant	Indonesian Meatball Place	Noodle House	Fast Food Restaurant	Food Truck
Menteng	106.9285	0	Indonesian Restaurant	Indonesian Meatball Place	Noodle House	Fast Food Restaurant	Pool
North Kembangan	106.8621	0	Coffee Shop	Indonesian Restaurant	Convenience Store	Pizza Place	Asian Restaurant
North Meruya	106.8331	0	Hotel	Coffee Shop	Indonesian Restaurant	Chinese Restaurant	Japanese Restaurant
Palmerah	106.7426	0	Coffee Shop	Indonesian Restaurant	Noodle House	Asian Restaurant	CafÃ©
Palmerah, Palmerah	106.7382	0	Coffee Shop	Food Truck	Indonesian Restaurant	Chinese Restaurant	Bookstore
Pancoran, South Jakarta	106.778	0	Indonesian Restaurant	Coffee Shop	Pizza Place	Steakhouse	Asian Restaurant
Pasar Minggu	106.7839	0	Coffee Shop	Noodle House	Chinese Restaurant	Indonesian Restaurant	CafÃ©
Pasar Rebo	106.7972	0	Coffee Shop	Dessert Shop	Hotel	Restaurant Pizza Place	Chinese Restaurant
Pegadungan	106.7898	0	Coffee Shop	Indonesian Restaurant	Asian Restaurant	Pizza Place	Food Truck
Pekojan	106.8473	0	Coffee Shop	Indonesian Restaurant	Noodle House	Food Truck	Asian Restaurant
Pinangsia	106.84	0	Coffee Shop	Restaurant	Asian Restaurant	Pizza Place	Food Truck
Pondok Bambu	106.8562	0	Noodle House	Pizza Place	Indonesian Restaurant	Convenience Store	Fast Food Restaurant
Rawa Buaya	106.9034	0	Fast Food Restaurant	Indonesian Restaurant	Grocery Store	Asian Restaurant	CafÃ©

Borough	Longitude	Cluster_Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Roa Malata	106.93	0	Indonesian Restaurant	Food Truck	Asian Restaurant	Restaurant	Salon / Barbershop
Sawah Besar	106.9432	0	Noodle House	Indonesian Restaurant	Food Truck	Pool	Department Store
Semanan	106.8912	0	Indonesian Restaurant	Coffee Shop	Restaurant	Pizza Place	Chinese Restaurant
Senen	106.7382	0	Coffee Shop	Fast Food Restaurant	Pizza Place	Noodle House	CafÃ©
South Duri	106.7058	0	Indonesian Restaurant	Asian Restaurant	Fast Food Restaurant	Noodle House	Convenience Store
South Kedoya	106.8503	0	Indonesian Restaurant	Coffee Shop	Hotel	CafÃ©	Restaurant
South Kembangan	106.8326	0	Coffee Shop	Hotel	Clothing Store	Multiplex	Bakery
South Meruya	106.8016	0	Coffee Shop	Dessert Shop	Indonesian Restaurant	Pizza Place	Hotel
South Tanjung Duren	106.7618	0	Noodle House	Coffee Shop	Indonesian Restaurant	Seafood Restaurant	Chinese Restaurant
Srengseng	106.7426	0	Coffee Shop	Indonesian Restaurant	Noodle House	Asian Restaurant	CafÃ©
Taman Sari (Jakarta)	106.7338	0	Indonesian Restaurant	Noodle House	Convenience Store	Pool	Fast Food Restaurant
Tambora, Jakarta Barat, Tambora, Jakarta Barat	106.7898	0	Coffee Shop	Noodle House	Dessert Shop	Indonesian Restaurant	Chinese Restaurant
Tanah Abang	106.753	0	Coffee Shop	Noodle House	Indonesian Restaurant	Seafood Restaurant	Restaurant
Tomang	106.8119	0	Coffee Shop	Hotel	Indonesian Restaurant	Restaurant	Asian Restaurant
West Cengakareng, Jakarta Barat	106.7285	0	Asian Restaurant	Noodle House	Chinese Restaurant	Pizza Place	Steakhouse Chinese Restaurant

Table 1 Cluster Analysis for First Cluster

Latitude	Longitude	Cluster_Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
-6.14233	106.7338	1	Asian Restaurant	Chinese Restaurant	Pizza Place	Indonesian Restaurant	Noodle House
-6.1617	106.7846	1	Chinese Restaurant	Noodle House	Coffee Shop	Korean Restaurant	Restaurant
-6.15982	106.7854	1	Chinese Restaurant	Noodle House	Coffee Shop	Restaurant	Convenience Store
-6.14882	106.7847	1	Restaurant	Noodle House	Asian Restaurant	Indonesian Restaurant	Coffee Shop
-6.15301	106.7987	1	Noodle House	Chinese Restaurant	Restaurant	Indonesian Restaurant	Asian Restaurant
-6.14517	106.8046	1	Chinese Restaurant	Noodle House	Asian Restaurant	Coffee Shop	CafÃ©
-6.15745	106.7994	1	Restaurant	Noodle House	Asian Restaurant	Indonesian Restaurant	Pharmacy
-6.13429	106.7058	1	Noodle House	Asian Restaurant	Chinese Restaurant	Pizza Place	Food Truck
-6.13429	106.7058	1	Noodle House	Asian Restaurant	Chinese Restaurant	Pizza Place	Food Truck
-6.1506	106.8142	1	Restaurant	Noodle House	Asian Restaurant	Coffee Shop	Hotel
-6.16037	106.8473	1	Hotel	Chinese Restaurant	Noodle House	Coffee Shop	Indonesian Restaurant
-6.14965	106.8038	1	Chinese Restaurant	Noodle House	Asian Restaurant	Coffee Shop	Bakery
-6.15764	106.8134	1	Chinese Restaurant	Noodle House	Coffee Shop	Asian Restaurant	Hotel
-6.14535	106.8186	1	Chinese Restaurant	Noodle House	Coffee Shop	Asian Restaurant	Hotel
-6.15674	106.8215	1	Restaurant	Noodle House	Asian Restaurant	Hotel	Coffee Shop
-6.15471	106.8053	1	Noodle House	Chinese Restaurant	Asian Restaurant	Coffee Shop	Soup Place
-6.17079	106.7618	1	Noodle House	Indonesian Restaurant	Chinese Restaurant	Coffee Shop	Seafood Restaurant
-6.13184	106.7028	1	Noodle House	Chinese Restaurant	Food Truck	Pizza Place	Asian Restaurant
-6.13751	106.8046	1	Chinese Restaurant	Noodle House	Asian Restaurant	Coffee Shop	CafÃ©
-6.13468	106.8134	1	Chinese	Noodle House	Asian Restaurant	Coffee Shop	Historic Site

Latitude	Longitude	Cluster_Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
-6.13671	106.8097	1	Restaurant Chinese Restaurant	Noodle House	Asian Restaurant	Coffee Shop	Historic Site
-6.15319	106.8326	1	Noodle House	Chinese Restaurant	Seafood Restaurant	Hotel	Asian Restaurant
-6.15854	106.8053	1	Noodle House	Chinese Restaurant	Coffee Shop	Asian Restaurant	Hotel
-6.14847	106.8208	1	Chinese Restaurant	Noodle House	Asian Restaurant	Coffee Shop	Hotel
-6.14847	106.8208	1	Chinese Restaurant	Noodle House	Asian Restaurant	Coffee Shop	Hotel
-6.14904	106.8031	1	Noodle House	Chinese Restaurant	Asian Restaurant	Coffee Shop	Restaurant
-6.11786	106.7193	1	Noodle House	Pizza Place	Coffee Shop	Chinese Restaurant	Asian Restaurant
-6.15446	106.7749	1	Chinese Restaurant	Asian Restaurant	Noodle House	Indonesian Restaurant	Coffee Shop

Table 2 Cluster Analysis for Second Cluster

From the data on the table, the best place to start the business for a coffee shop is the second cluster, where the number of coffee shops is less than the first cluster and can avoid the highly competitive market. However, because the number of the area in the second cluster is also less than the first cluster, it was expected that in the second cluster will have a potential of less profit.

IV. CONCLUSION

The data analysis technique was used in this research to find the best location for opening the coffee shop in the Jakarta area. It was found by using the K-mean clustering that a total of clusters area in Jakarta divided into two clusters. The clusters were analyzed to seek the best location based on the numbers of coffee shops venue, and it was found that the second cluster that has fewer coffee shops venues is the most appropriate location to avoid high competition. However, because of the number of locations of the second cluster far less then the first clusters, it can potentially impact to the profit of the business.
This research limitation is being conducted in the Jakarta area using K-mean clustering for analyzing. Furter clustering methods can be used to analyze the data, and also regression can be used to predict the future growth of the business.
Detail data and code are available in the following link:

https://github.com/budisaleh/github-courseracapst/blob/master/Week%204/coursera_caps.ipynb

V. REFERENCES

Wikipedia.org
Google.com
Stackoverflow.com