Coursera Capstone project Coursera

IBM Data Science Certification

Arif Zainurrohman

Nov 2nd, 2020

Report Content

1. **Introduction Section** : The “business problem” to be solved by this project and who may be interested
2. **Data Section:** Describe Data requirements and Sources needed to solve the problem
3. **Methodology section:** Main component of the report - Execute data processing, describe/discuss any exploratory data analysis and/or inferential statistical testing performed, and/or machine learnings used.
4. **Results section:** Discussion of the results and finding of answer
5. **Discussion section:** Discussion of observations noted and any recommendations
6. **Conclusion section:** Answer chosen and conclusions.

1.0 Introduction

1.1 Scenario and Background

I am a business analyst currently residing in Jakarta Barat Indonesia. I currently live within walking distance to buy food therefore I have access to good public transportation to geet it. Likewise, I enjoy many ammenities in the neighborhood , such as international cousine restaurants, cafes, food shops and entertainment. I have been offered a great opportunity to work in Jakarta. Although, I am very excited about it, I am a bit stress toward the process to secure a comparable place to live in here.

1.2 Problem to be resolved

The major purpose of this project, is to suggest a better neighborhood in a new city for the person who are shiffting there. Social presence in society in terms of like minded people. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

1.Show list of venue in Jakarta Barat

2.Show list of Hospital in Jakarta Barat

1.3 Interested Audience

I believe the methodology, tools and strategy used in this project is relevant for a person or entity considering moving to a major city in Indonesia. Likewise, it can be helpful approach to explore the opening of a new business. The use of FourSquare data and mapping techniques combined with data analysis will help resolve the key questions. Lastly, this project is a good practical case for a person developing Data Science skills.

2.0 Data Section

2.1 Data Requirements

- Geodata for current residence in Jakarta Barat, Indonesia with venues established using Foursquare.
- List of Jakarta Barat postcode

  https://www.indonesiapostcode.com/location/jakarta/jakarta-barat/
- List of Hospital in Jakarta

  https://raw.githubusercontent.com/cahyati/Coursera_Capstone/master/Hospital%20for%20treatment%20covid-19.csv

2.2 Data Sources, Data Processing and Tools used

- Jakarta Barat data and map is to be created with use of Nominatim , Foursquare and Folium mapping
- Jakarta Barat Postcode were obtained from Indonesia postcode and organized by Neighborhoods with geodata via Nominatim for mapping with Folium.
- List of Subway stations was obtained via Wikipedia, NY Transit web site and Google map,
- List of hospital for rent was consolidated from web-scraping. The geolocation (lat,long) data was found with algorithm coding and using Nominatim.
- Folium map was the basis of mapping with various features to consolidate all data in ONE map where one can visualize all details needed.

3.0 Methodology

The Strategy to find the answer:

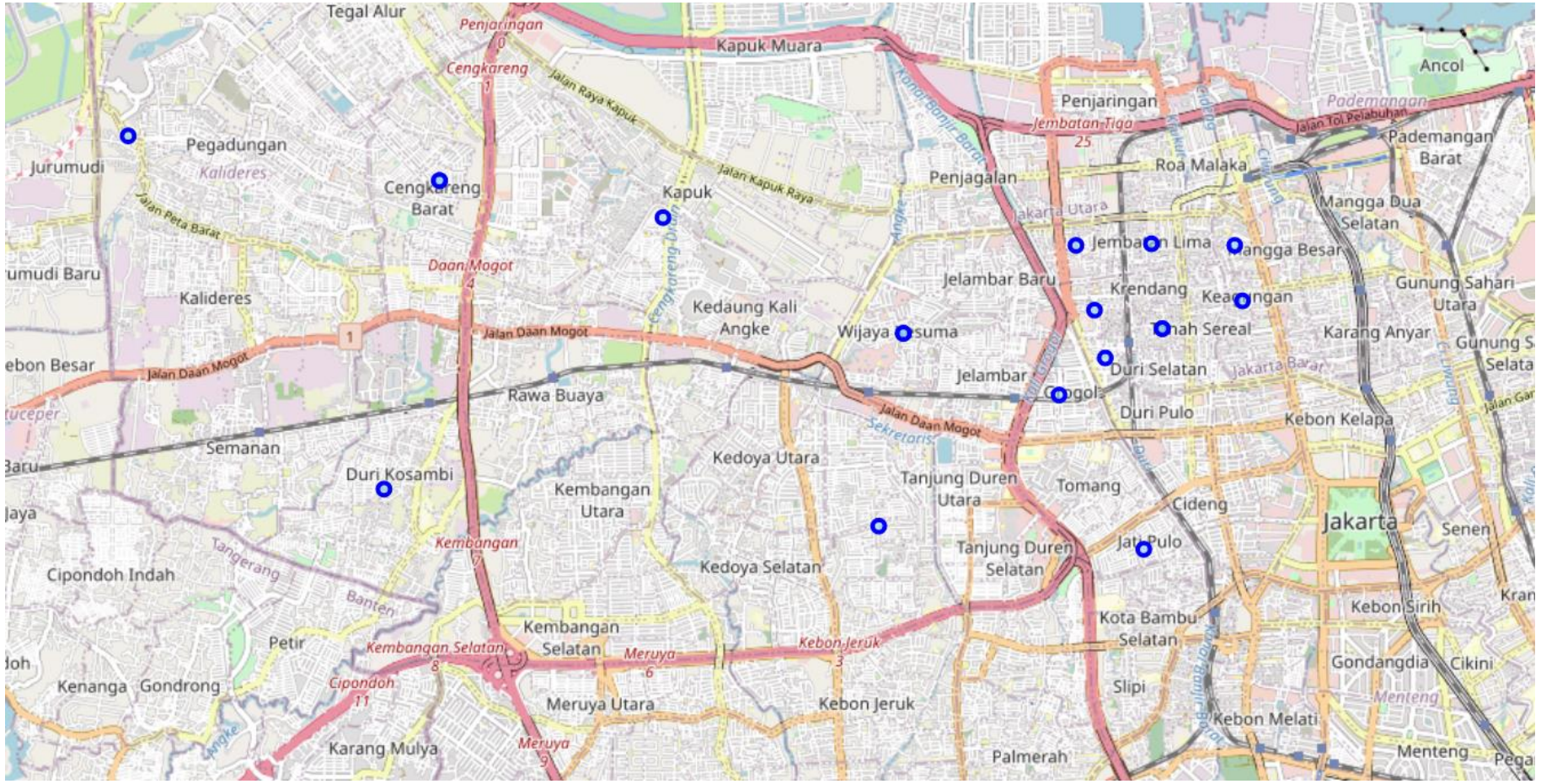
The strategy is based on mapping the described data in section 2.0. The information will be consolidated in ONE MAP where one can see the details of the venue

The Tools:

Web-scraping of sites is used to consolidate data-frame information which was saved as csv files for convenience and to simply the report. Geodata was obtained by coding a program to use Nominatim to get latitude and longitude of Kecamatan and also for each venue listed. Geopy_distance and Nominatim were used to establish relative distances.

4.0 Execution and Results

Current Kecamatan Neighborhood in Jakarta Barat



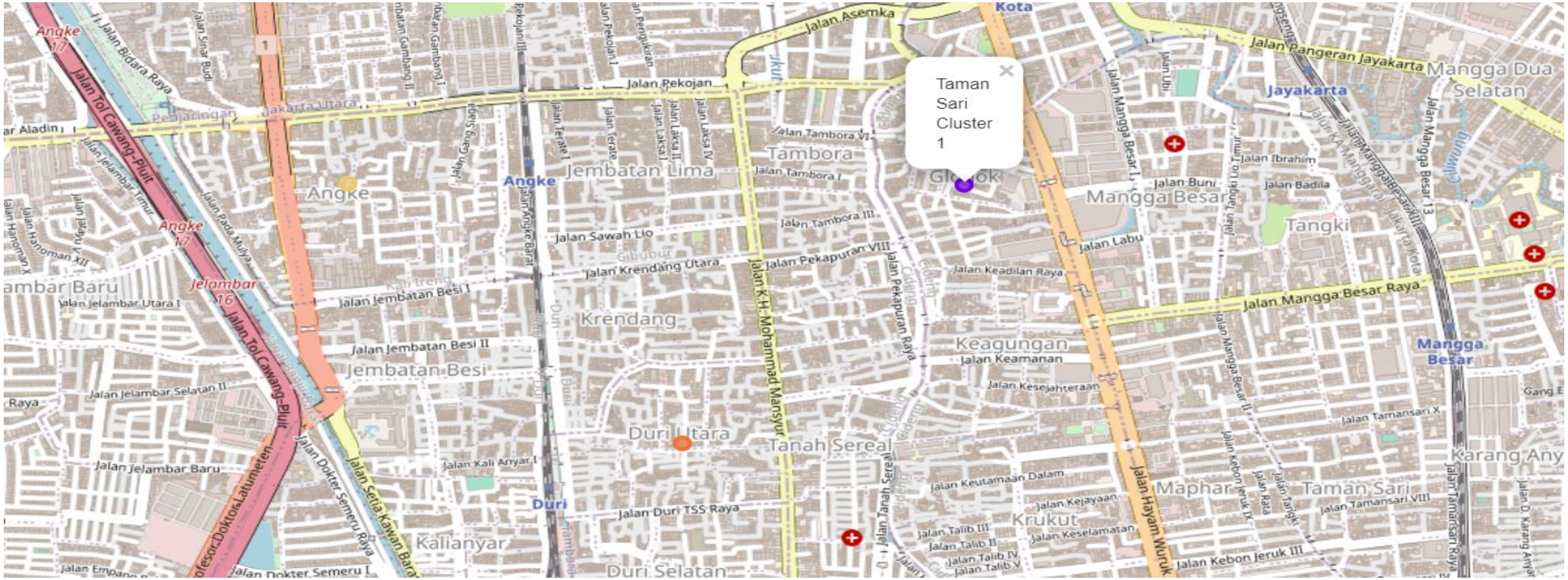
Venues around Kecamatan in

```
jakarta_venues.head()
```

```
(269, 7)
```

	Kecamatan	Kecamatan Latitude	Kecamatan Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Tambora	-6.145099	106.795553	Bakmi Garing 21	-6.144851	106.793806	Noodle House
1	Tambora	-6.145099	106.795553	The Harvest - Patissier & Chocolate	-6.142911	106.793866	Café
2	Tambora	-6.145099	106.795553	Jus 'Kembung' Moro Seneng	-6.142308	106.796334	Juice Bar
3	Tambora	-6.145099	106.795553	Jus Kembung	-6.142329	106.796294	Lounge
4	Tambora	-6.145099	106.795553	Trattoria Cucina Italiana	-6.146542	106.796548	Italian Restaurant

Jakarta Barat Map - Kecamatan and Cluster of Venues



Hospital in Jakarta

	Hospital	City	District
0	RSPI Sulianti Saroso	Jakarta Utara	Tanjung Priok
1	RSUP Persahabatan	Jakarta Timur	Pulo Gadung
2	RSPAD Gatot Soebroto	Jakarta Pusat	Senen
3	RSUP Fatmawati	Jakarta Selatan	Cilandak
4	RSU Bhayangkara Said Sukanto	Jakarta Timur	Kramat Jati
5	RSAL Mintohardjo	Jakarta Pusat	Tanah Abang
6	RSUD Cengkareng	Jakarta Barat	Cengkareng
7	RSUD Pasar Minggu	Jakarta Selatan	Pasar Minggu
8	RSKD Duren Sawit	Jakarta Timur	Duren Sawit
9	RS Pelni	Jakarta Barat	Palmerah
10	RSUD Tarakan	Jakarta Pusat	Cideng
11	RSUD Koja	Jakarta Utara	Koja
12	RSU Pertamina Jaya	Jakarta Pusat	Cempaka Putih Timur

Map of Hospital in Jakarta

Top 5 venue in each Kecamatan

```
num_top_venues = 5
for neigh in jakarta_grouped['Kecamatan']:
    print("----"+neigh+"----")
    temp = jakarta_grouped[jakarta_grouped['Kecamatan'] == neigh].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')
```

```
----Cengkareng----
                   venue  freq
0       Asian Restaurant  0.20
1           Noodle House  0.14
2     Chinese Restaurant  0.13
3     Seafood Restaurant  0.07
4  Indonesian Restaurant  0.06


----Grogol Petamburan----
               venue  freq
0      Noodle House  0.23
1  Basketball Court  0.10
2        Restaurant  0.06
3        Food Court  0.06
4         Bookstore  0.06
```

Most Common venues near Kecamatan

```
Kecamatan_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
Jakarta_merged = df
Jakarta_merged = Jakarta_merged.join(Kecamatan_venues_sorted.set_index('Kecamatan'), on='Kecamatan')

Jakarta_merged.head()
```

	Desa	Kecamatan	KodePos	Latitude	Longitude	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
0	Angke	Tambora	11330	-6.145099	106.795553	1	Chinese Restaurant	Asian Restaurant	Seafood Restaurant	Noodle House	Train Station	Indonesian Restaurant	Food Court	Fast Food Restaurant	Indonesian Meatball Place
1	Cengkareng Barat	Cengkareng	11730	-6.138010	106.724683	6	Asian Restaurant	Noodle House	Chinese Restaurant	Seafood Restaurant	Fried Chicken Joint	Dessert Shop	Indonesian Restaurant	Café	Indonesian Meatball Place
2	Cengkareng Timur	Cengkareng	11730	-6.138010	106.724683	6	Asian Restaurant	Noodle House	Chinese Restaurant	Seafood Restaurant	Fried Chicken Joint	Dessert Shop	Indonesian Restaurant	Café	Indonesian Meatball Place
3	Duri Kepa	Kebon Jeruk	11510	-6.176304	106.773504	2	Food Truck	Park	Coffee Shop	Soccer Field	Vegetarian / Vegan Restaurant	Fast Food Restaurant	Dog Run	Dumpling Restaurant	Eastern European Restaurant
4	Duri Kosambi	Cengkareng	11750	-6.172171	106.718389	6	Asian Restaurant	Noodle House	Chinese Restaurant	Seafood Restaurant	Fried Chicken Joint	Dessert Shop	Indonesian Restaurant	Café	Indonesian Meatball Place

5.0 Discussion

- In general, I am positively impressed with the overall organization, content and lab works presented during the Coursera IBM Certification Course
- My analysis shows that although there is a great number of restaurants in Jakarta , there are pockets of low restaurant density fairly close to city center. Another borough was identified as potentially interesting, but our attention was focused on Restaurant and Café which offer a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics and a number of pockets of low restaurant density.
- Those location Restaurants were then clustered to create zones of interest which contain greatest number of location. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.
- Result of all this is containing largest number of potential new restaurant locations based on number of and distance to existing venues - Chinese restaurants in general and Asian restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide info on areas close to Jakarta center but not crowded with existing restaurants (particularly Chinese) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

6.0 Conclusions

- I feel rewarded with the efforts, time and money spent. I believe this course with all the topics covered is well worthy of appreciation.
- Purpose of this project was to identify Jakarta Barat areas close to center with low number of restaurants (particularly Chinese restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Asian restaurant or Indonesian restaurant. By calculating restaurant density distribution from Foursquare data we have first identified 8 Kecamatan that justify further analysis, and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.
- Final decission this project help me to explore Jakarta Barat about food.
- End of Project and Course/ Thanks to Coursera Team and Students!.