

IBM DATA SCIENCE CAPSTONE PROJECT PRESENTATION

Danh Thai-Hoang

Email: danhthaihoang@gmail.com



PROJECT: OPENING A ITALIAN RESTAURANT BUSINESS IN KUALA LUMPUR, MALAYSIA

- Introduction
- Data
- Methodology
- Result and Discussion
- Conclusion



INTRODUCTION – BUSINESS PROBLEM

- More and more restaurants are popping up in the city.
- Location of the restaurant is one of the most important decisions that will determine whether the restaurant will be a success or a failure.
- Project Objectives: To analyze and select the best locations in the city of Kuala Lumpur, Malaysia to open a new Italian restaurant.
- Business Questions:
 - **“Should an entrepreneur start a Italian restaurant business in Kuala Lumpur City, Malaysia?”**
 - **“Which location is suitable and recommended to open it?”**



INTRODUCTION – TARGET AUDIENCES

- Local entrepreneurs
- Foreign investors who are interested in opening restaurant business focus on serving the Italian dishes.



DATA REQUIREMENTS

- List of neighborhoods in Kuala Lumpur
- Latitude and longitude coordinates of the neighborhoods
- Venue data, particularly data related to restaurants



DATA COLLECTION

- **List of neighborhoods in Kuala Lumpur**
 - https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur with more than 70 **neighborhoods** in the list. We will crawl the data from the page by using BeautifulSoup library, Python requests and Pandas.
- **Latitude and longitude coordinates of the neighborhoods**
 - Retrieved using Google Maps Geocoding API. The geometric location values are then stored into the initial dataframe.
- **Venue data, particularly data related to restaurants**
 - From the location data obtained after Web Scraping and Geocoding, the venue data is found out by passing in the required parameters to the **FourSquare API** (<https://developer.foursquare.com/>), and creating another DataFrame to contain all the venue details along with the respective neighborhoods.



METHODOLOGY

- Detect areas of Kuala Lumpur that have low restaurant density, particularly those with low number of Italian restaurants.
 - Collect the required data: location and type (category) of every restaurant within 6km from Kuala Lumpur center (KLCC) from Google APIs
 - Identify Italian restaurants (according to Foursquare categorization)
- Limit our analysis to areas ~6km around the city center.
 - Calculation and exploration of 'restaurant density' across different areas of Kuala Lumpur
 - Use heatmaps to identify a few promising areas close to the center with low number of restaurants in general
- Focus on the most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders



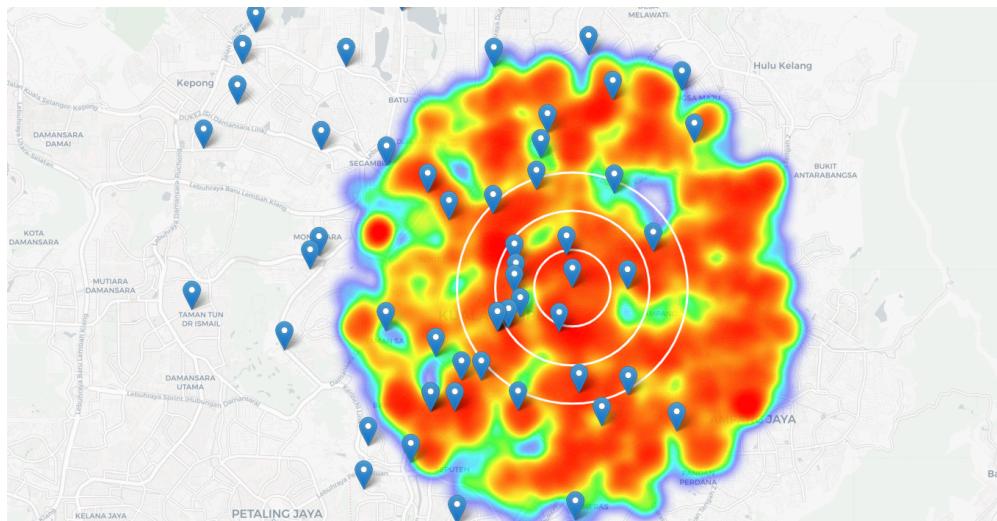
METHODOLOGY

- Take into consideration locations with no more than two restaurants in a radius of 250 meters, and we want locations without Italian restaurants in radius of 400 meters Limit our analysis to areas ~6km around the city center.
- Present a map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses



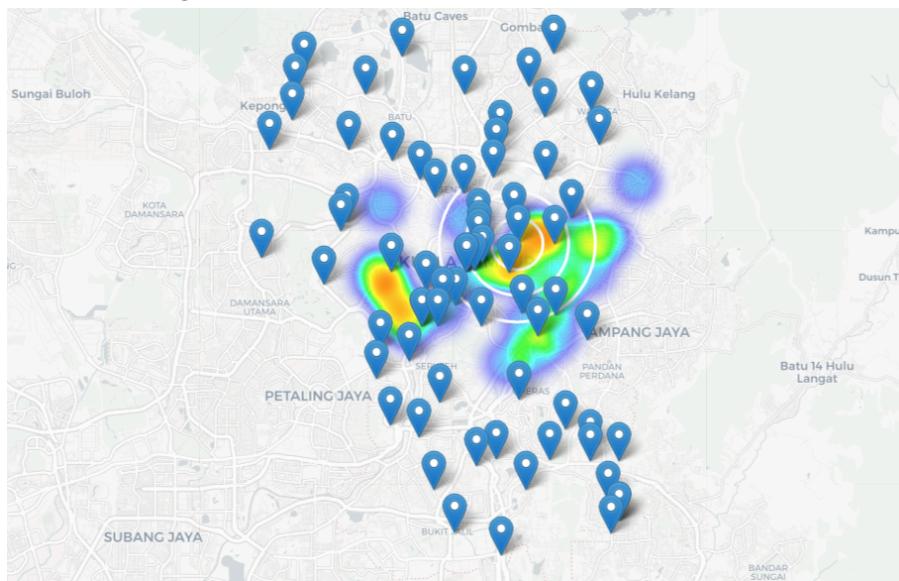
EXPLORATORY ANALYSIS

- A map showing heatmap / density of restaurants and try to extract some meaningful info from that. Also, let's show markers of Kuala Lumpur boroughs on our map and a few circles indicating distance of 1km, 2km and 3km from Suria KLCC (considered as the center of Kuala Lumpur)



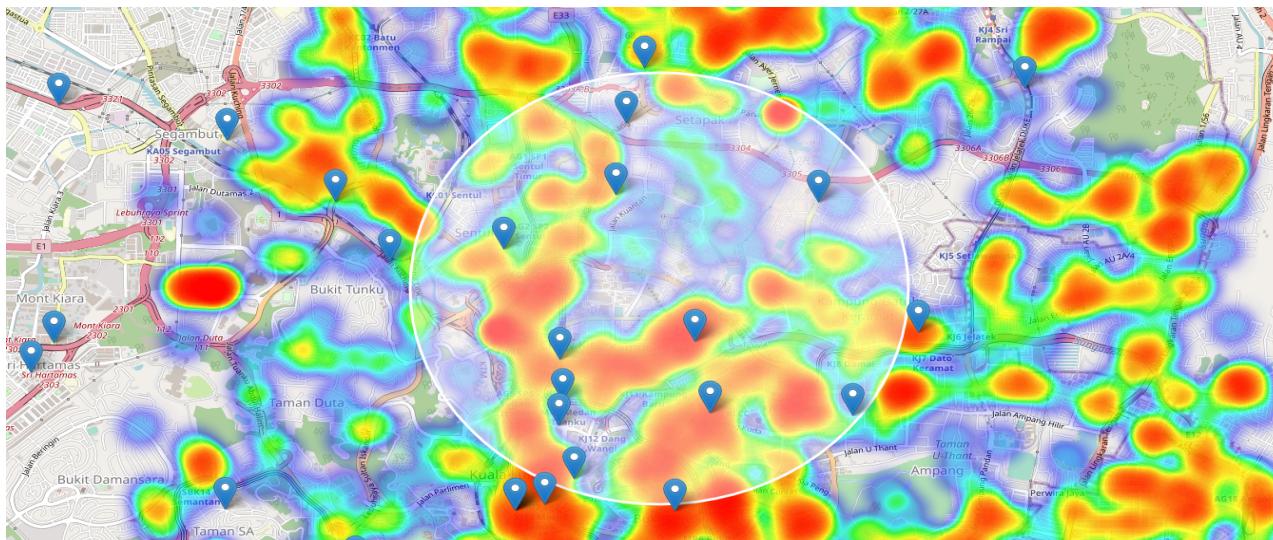
EXPLORATORY ANALYSIS

- Heatmap map showing heatmap/density of Italian restaurants only



EXPLORATORY ANALYSIS

- Low-restaurant-count parts closest to Suria KLCC.



EXPLORATORY ANALYSIS

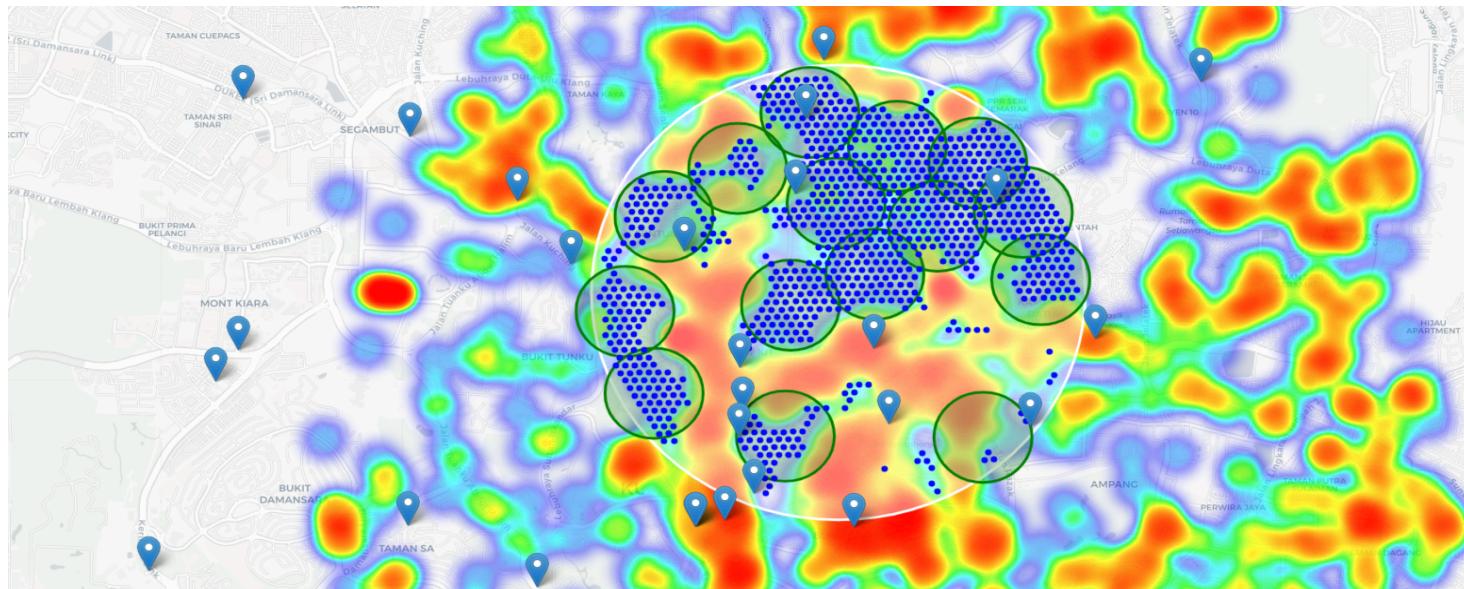
- Calculate two most important things for each location candidate: number of restaurants in vicinity (we'll use a radius of 250 meters) and distance to closest Italian restaurant.

	Distance to Italian restaurant	Latitude	Longitude	Restaurants nearby	x	y
0	3424.963359	3.192860	101.707235	6	2.219629e+06	1.963000e+07
1	3389.552037	3.192846	101.706367	4	2.219729e+06	1.963000e+07
2	3261.375329	3.192177	101.712021	7	2.219079e+06	1.963009e+07
3	3269.031296	3.192164	101.711153	11	2.219179e+06	1.963009e+07
4	3279.719834	3.192151	101.710285	11	2.219279e+06	1.963009e+07
5	3293.411418	3.192137	101.709417	12	2.219379e+06	1.963009e+07
6	3310.068783	3.192124	101.708549	12	2.219479e+06	1.963009e+07
7	3329.647417	3.192111	101.707681	12	2.219579e+06	1.963009e+07
8	3338.242594	3.192097	101.706813	5	2.219679e+06	1.963009e+07
9	3289.592130	3.192084	101.705945	0	2.219779e+06	1.963009e+07



EXPLORATORY ANALYSIS

- Cluster those locations to create centers of zones containing good locations



FINAL ANALYSIS

▪ Good locations to open the Italian Restaurants in Kuala Lumpur

2218187.45758
3, Lorong Kuda, Kuala Lumpur, 50450 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia => 1.
0km from Suria KLCC
2219103.26718
B-32-3A Bennington Residensi, Taman Ayer Panas, 53200 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia => 3.1km from Suria KLCC
2220182.9704
1111, Jalan Dr Latif, 50586 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia => 1.7km from Suria KLCC
2217355.0079
Markaz Latihan TD, Kampung Datuk Keramat, 54000 Kuala Lumpur, Federal Territory of Kuala Lumpur, Malaysia => 2.3km from Suria KLCC
2218629.12425
Jalan Johor, Pusat Latihan Polis (PULAPOL), 54100 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia => 2.3km from Suria KLCC
2221495.79092
139, Jalan Union, Sentul, 51000 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia => 3.3km from Suria KLCC
2219687.7264
Taman Tasik Titiwangsa, Tasik, Titiwangsa, 53200 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia => 2.6km from Suria KLCC
2221919.50886
Jalan 11, Jalan Ipoh, 51200 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia => 3.0km from Suria KLCC
2220205.50968
Dang Wangi, Kuala Lumpur, 50300 Kuala Lumpur, Federal Territory of Kuala Lumpur, Malaysia => 1.1km from Suria KLCC
2219947.87425
416, Jalan Pahang, Setapak, 53000 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia => 3.6km from Suria KLCC
2219294.7869
Gardens, Titiwangsa, 53200 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia => 1.7km from Suria KLCC
2217726.70489
Unnamed Road, Kampung Padang Tembak, 54100 Kuala Lumpur, Federal Territory of Kuala Lumpur, Malaysia => 2.8km from Suria KLCC
2218196.98139
2, Lorong Ayer Leleh, Taman Ayer Panas, 53200 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia => 1.1km from Suria KLCC
2220730.11235
43-0-8, Jalan 1/48A, Sentul Perdana, Bandar Baru Sentul, WP Kuala Lumpur, Bandar Baru Sentul, 51000 Kuala Lumpur, Federal Territory of Kuala Lumpur, Malaysia => 3.3km from Suria KLCC
2221629.12425
Club House, Jalan Tun Ismail, Kuala Lumpur, 50480 Kuala Lumpur, Federal Territory of Kuala Lumpur, Malaysia => 2.5km from Suria KLCC



RESULTS AND DISCUSSION

- Our analysis shows that although there is a great number of restaurants in Kuala Lumpur (~ 2800 in our initial area of interest which was **12x12km** around Suria), there are pockets of low restaurant density fairly close to the city center
- Highest concentration of restaurants was detected north and west from West, so we focused our attention to areas south, south-east and east.
- After directing our attention to this more narrow area of interest (covering approx. **5x5km** south-east from Suria KLCC) we first created a dense grid of location candidates (spaced **100m** apart); those locations were then filtered so that those with more than two restaurants in a radius of **250m** and those with an Italian restaurant closer than **400m** were removed.
- Result of all this is **15 zones** containing the largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Italian restaurants in particular



CONCLUSION

- Purpose of this project was to identify Kuala Lumpur areas close to the center with a low number of restaurants (particularly Italian restaurants) in order to aid stakeholders in narrowing down the search for the optimal location for a new Italian restaurant
- 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 decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations.



REFERENCE

- Wiki page
- Google APIs
- Foursquare APIs

