

2. Data

2.1 Influencing Factors on Data

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

- Majority ethnic group in the neighborhood
- Number of existing restaurants in the neighborhood (any type of restaurant)

2.2 Data Acquisition

Data sources are listed below:

- Demographic data of ethnic groups in Kuala Lumpur are found on TindakMalaysia Website
- Coordinates of Kuala Lumpur center are obtained using Geocoders geocoding of well known Kuala Lumpur location
- Neighborhood areas of Kuala Lumpur are found on TindakMalaysia Github Repository
- Coordinates of centers of candidate areas are generated algorithmically from the neighborhood areas of Kuala Lumpur
- Number of restaurants and their type and location in every neighborhood are obtained using Foursquare API

2.3 Data Cleaning

Due to the lack of Malaysia's demographic data online, I have chosen to scrape the Kuala Lumpur voters' demographic information from TindakMalaysia website, which originally served for election purpose. This information is based on census data of Q4, Year 2017. Kuala Lumpur has totally 11 administrative districts. Ethnic groups are listed according to their percentage for each district (Figure 1). After that, the data is being cleaned up to show only the majority ethnic group, with other redundant information removed (Figure 2).

	District	Chinese	Malay	India	Bumiputera Sabah	Bumiputera Sarawak	Orang Asli	Others
0	BANDAR TUN RAZAK	29.10	61.10	8.27	0.42	0.53	0.01	0.57
1	BATU	38.27	42.52	17.47	0.32	0.44	0.02	0.96
2	BUKIT BINTANG	75.13	12.53	10.95	0.37	0.28	0.02	0.73
3	CHERAS	82.93	9.66	6.74	0.22	0.06	0.01	0.38
4	KEPONG	88.74	4.56	6.03	0.30	0.05	NaN	0.32
5	LEMBAH PANTAI	18.91	62.11	16.26	0.67	0.48	0.02	1.55
6	SEGAMBUT	58.91	27.53	11.78	0.14	0.12	0.01	1.50
7	SEPUTEH	82.02	9.04	7.88	0.22	0.08	NaN	0.75
8	SETIAWANGSA	25.87	61.08	10.39	0.75	1.08	0.04	0.78
9	TITIWANGSA	17.31	70.48	9.42	0.49	0.37	0.01	1.91
10	WANGSA MAJU	29.57	59.81	8.19	0.45	0.59	0.04	1.36

Figure 1: Table of Proportion of Ethnic Groups in Kuala Lumpur Districts

	District	Majority Race	Percentage
0	BANDAR TUN RAZAK	Malay	61.10
1	BATU	Malay	42.52
2	BUKIT BINTANG	Chinese	75.13
3	CHERAS	Chinese	82.93
4	KEPONG	Chinese	88.74
5	LEMBAH PANTAI	Malay	62.11
6	SEGAMBUT	Chinese	58.91
7	SEPUTEH	Chinese	82.02
8	SETIAWANGSA	Malay	61.08
9	TITIWANGSA	Malay	70.48
10	WANGSA MAJU	Malay	59.81

Figure 2: Table of The Majority Race in Kuala Lumpur Districts with Percentages

As we need to visualize the data using map, the majority race in each district area is represented as a number, so the choropleth map can display the areas with different colors based on their respective main ethnic group in stay. As from the previous table we know that there are only 2 major groups presented, therefore the Malays are labeled as 0, and Chinese are labeled as 1 (Figure 3).

	District	Majority Race	Percentage	Majority Race No
0	BANDAR TUN RAZAK	Malay	61.10	0
1	BATU	Malay	42.52	0
2	BUKIT BINTANG	Chinese	75.13	1
3	CHERAS	Chinese	82.93	1
4	KEPONG	Chinese	88.74	1
5	LEMBAH PANTAI	Malay	62.11	0
6	SEGAMBUT	Chinese	58.91	1
7	SEPUTEH	Chinese	82.02	1
8	SETIAWANGSA	Malay	61.08	0
9	TITIWANGSA	Malay	70.48	0
10	WANGSA MAJU	Malay	59.81	0

Figure 3: Extended Table with The Majority Race Shown In Numbers

In the meanwhile, the coordinates of neighborhoods in Kuala Lumpur is scraped from the TindakMalaysia Github Repository. The centers of each neighborhood are generated by simple calculation of mean method based on boundary points of each neighborhood. However, since some areas are in irregular shapes and therefore the calculated centers are placed far from the middle of their areas, therefore adjustments on some candidate centers have to be made accordingly (Figure 4).

	District	Latitude	Longitude
0	LEMBAH PANTAI	3.109179	101.664066
1	TITIWANGSA	3.153691	101.736048
2	CHERAS	3.098256	101.734837
3	SETIAWANGSA	3.185183	101.733304
4	BUKIT BINTANG	3.139090	101.700725
5	BATU	3.202750	101.681007
6	KEPONG	3.207030	101.652003
7	SEPUTEH	3.088583	101.690155
8	SEGAMBUT	3.166146	101.657389
9	WANGSA MAJU	3.210742	101.720206
10	BANDAR TUN RAZAK	3.076840	101.713364

Figure 4: Table of Kuala Lumpur Districts with Coordinates

For the easy readability, both previous tables are merged together to provide a master table with complete information of coordinates and demographic information in Kuala Lumpur's districts (Figure 5).

	District	Latitude	Longitude	Majority Race	Percentage	Majority Race No
0	LEMBAH PANTAI	3.109179	101.664066	Malay	62.11	0
1	TITIWANGSA	3.153691	101.736048	Malay	70.48	0
2	CHERAS	3.098256	101.734837	Chinese	82.93	1
3	SETIAWANGSA	3.185183	101.733304	Malay	61.08	0
4	BUKIT BINTANG	3.139090	101.700725	Chinese	75.13	1
5	BATU	3.202750	101.681007	Malay	42.52	0
6	KEPONG	3.207030	101.652003	Chinese	88.74	1
7	SEPUTEH	3.088583	101.690155	Chinese	82.02	1
8	SEGAMBUT	3.166146	101.657389	Chinese	58.91	1
9	WANGSA MAJU	3.210742	101.720206	Malay	59.81	0
10	BANDAR TUN RAZAK	3.076840	101.713364	Malay	61.10	0

Figure 5: Master Table of Complete Information

On the other hand, since we will explore venues in each district area using Foursquare API, the magnitude of radius from the city center is determined by plotting it on the map and inspect it visually, such that the coverage is sufficient enough to cover most of the areas of districts in Kuala Lumpur (Figure 6).

After deciding the magnitude of radius, we would also request the result to be containing 'food' category only from Foursquare, as we are interested in restaurant only. Furthermore, we would want to search for those that are proper restaurants. So the list obtained from the result is filtered out to only include venues that have 'restaurant' in their category name (Figure 7).

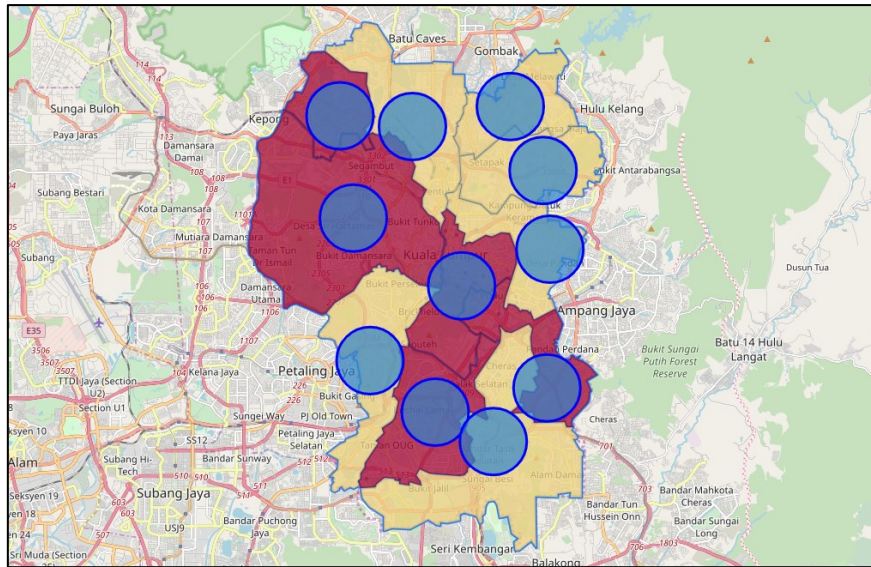


Figure 6: Map of Kuala Lumpur and Its Districts together with 1500m Radius Coverage from Centers of Districts

	District	Venue Name	Venue Category	Venue Coordinates	Venue Distance
0	LEMBAH PANTAI	明記家鄉小食	Chinese Restaurant	(3.1104848185266993, 101.678110759793)	271
1	LEMBAH PANTAI	Sanuki Udon	Udon Restaurant	(3.110877105565396, 101.68498225960003)	566
2	LEMBAH PANTAI	Nihon Kai	Japanese Restaurant	(3.106273820844563, 101.67897077220967)	349
3	LEMBAH PANTAI	Gyutaro Yakiniku	Japanese Restaurant	(3.1041169632637438, 101.6777439142475)	624
4	LEMBAH PANTAI	Sukaphat Thai Vegetarian restaurant	Vegetarian / Vegan Restaurant	(3.113072985090071, 101.67842572798531)	475

Figure 7: Filtered Table of Restaurants Only Requested from Foursquare