A New Chapter in Kuala Lumpur

(Fadil, 2013)

An IBM Applied Data Science Capstone Project

By Tyler Ho

Table of Contents

Contents	Page
Introduction	1
- **Background**	
- **Objective**	
Data and Methodology	
- **Data Collection**	2
- **Data Cleaning**	
- **Data Visualisation**	
- **More on Data Visualisation**	
- **Assumptions on Data**	3
- **Methodology**	4
Results and Discussion	
- **Visualisation on Map**	5
- **Further Analysis**	6
- **Limitations**	10
Conclusion	11
References	12
Appendices	13

Introduction

Background

In the heart of Malaysia lies a bustling and robust city, the capital of the country itself, Kuala Lumpur. With a population of roughly 1.8 million within its 243km^2 area (94 square miles), the city remains the source of great opportunity for the people of Malaysia (DOSM, n.d.). Hence, with this being in mind, it is without a doubt that many will look to migrate to the city as they hunt for new opportunities in life.

Objective

This paper aims itself at helping individuals who are keen on migrating to Kuala Lumpur and are in search of accommodations. As the city is quite small, the public transport and highways make it convenient to travel around. Hence, locations not only in the city itself but also in Greater Kuala Lumpur is taken into account.

The main goal of this paper to determine the best location to live in terms of the availability and variety in the categories of venues as below:

- Food & Beverages Outlets
- Schools
- Healthcare Facilities
- Arts & Entertainment Venues

Data and Methodology

Data Collection

The list of districts in Kuala Lumpur is taken from the website Geographic.org while their locational data is collected through the Geocoder Python Package. The venue data, on the other hand is collected through the Foursquare API.

Data Cleaning

As the Foursquare API may not have very accurate data in Malaysia, data cleaning had to be done on the venue data. The unique categories are listed to check if any venues that does not belong in that category existed. If so, they would be dropped. In the case of the data in this paper, only the F&B outlets had no problem with this.

Data Visualisation

A map whereby all Districts is created to visualise where each District stands. It is worth noting that some Districts are quite close to each other while others aren't. Next, 4 maps are created, each with the different venue to illustrate the compactness and availability of the venues within each district.

More on Data Visualisation

In the F&B category, the venues are categorized further into less groups as initially there had been 76 categories. Bar charts are plotted in each District to picture the count on each types of venues. KMeans clustering on the top 5 venues in each district is also created for more valuable insight.

In the Medical Centres category, a bar chart of the number of venues against each district is created and arranged in descending order.

In the Schools category, a bar chart of the number of venues against each district is created and arranged in descending order. A further plot that depicts the number of Elementary Schools, Middle Schools, High Schools and Private Schools is created.

In the Arts & Entertainment category, a bar chart of the number of venues against each district is created and arranged in descending order.

Assumptions on Data

- Nearby venues within 2.5KM (1.5 Miles) are considered.
- Venues are correctly tagged based on their subcategories.
- The subcategories that are relevant are kept. (*refer to Table 1 in appendices*)
- In the F&B category, subcategories are grouped as shown in *Table 2 in appendices*.
- In the 'Top 5 Highest Number of F&B Venues in each District', venues subcategorised as 'Others' are not considered.
- Elementary School, Nursery School and Preschool are combined as Elementary School.
- Only Elementary Schools, Middle Schools, High Schools and Private Schools are considered as people relocating often bring their younger children along. While their child that are pursuing tertiary education, most likely they do not mind moving away themselves.
- In the Bar Chart 'Type of Schools in each District', venues subcategorised as 'School' are not considered as their level is unknown.

Methodology

All data has been processed using a Jupyter Notebook on a Python kernel. The libraries used to conduct this study are Pandas, Numpy, Requests, GeoPy, MatPlotLib, Folium, IPython and SciKit-Learn. Besides, the FourSquare API is used.

Within the Capstone.ipynb, the code is separated into 3 sections, namely LIBRARIES & FUNCTIONS, DATA RETRIEVAL & CLEANING and DATA ANALYSIS.

In the LIBRARIES & FUNCTIONS section, all the libraries stated above are imported. The credentials for the FourSquare API is loaded as well. Functions that will be used after in the code are defined beforehand in this part too.

In the DATA RETRIEVAL & CLEANING section, a data frame of the list of districts along with their coordinates are created using the GeoPy encoder. The nearby venues in each category are collected through the FourSquare API as well and cleaned.

In the DATA ANALYSIS section, the maps are visualised using Folium. As Folium maps often have a problem showing on GitHub, the IPython library is required to attach screenshots of the results. Next, one hot encoding technique is used on the dataset in order to make the data available for plotting. Both Panda's internal plotting functions and MatPlotLib are used for the Bar Chart visualisations. SciKit-Learn is then used to do KMeans Clustering on the food venues while the elbow point method is used to determine the K used.

Results & Discussion

Visualisation on Map

Figure 1, as seen below acts as an introduction to the location of each district. It can be seen that most districts are pretty spaced out except for Damansara, Petaling Jaya and Setapak, Sentul.

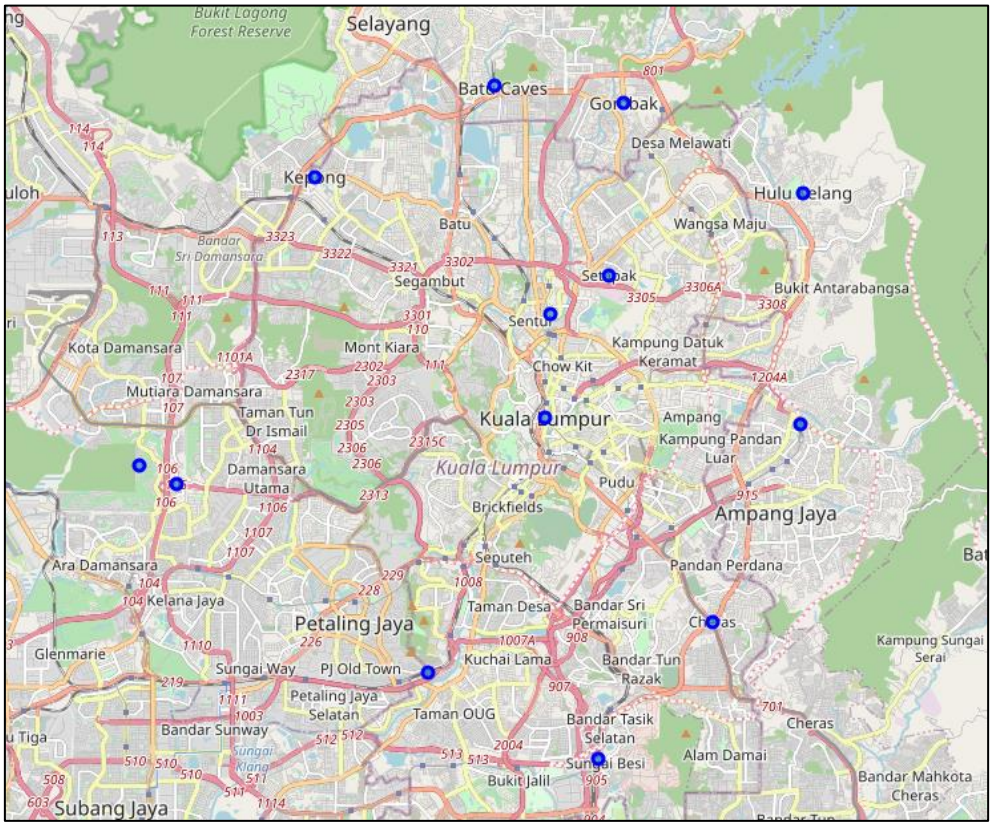


Figure 1: Districts on map of Kuala Lumpur

Maps of each category has also been generated and can be seen if needed in *Figure 2 – 5 in appendices*. From the maps created, it can be seen that food and schools are abundant for all districts. However, when it comes to medical facilities, Kuala Lumpur and Sentul show evidence of significantly higher quantity as compared to the others. The case is the same for Arts & Entertainment venues while Gombak and Petaling Jaya have no such venues.

Further Analysis

Food & Beverages Outlets

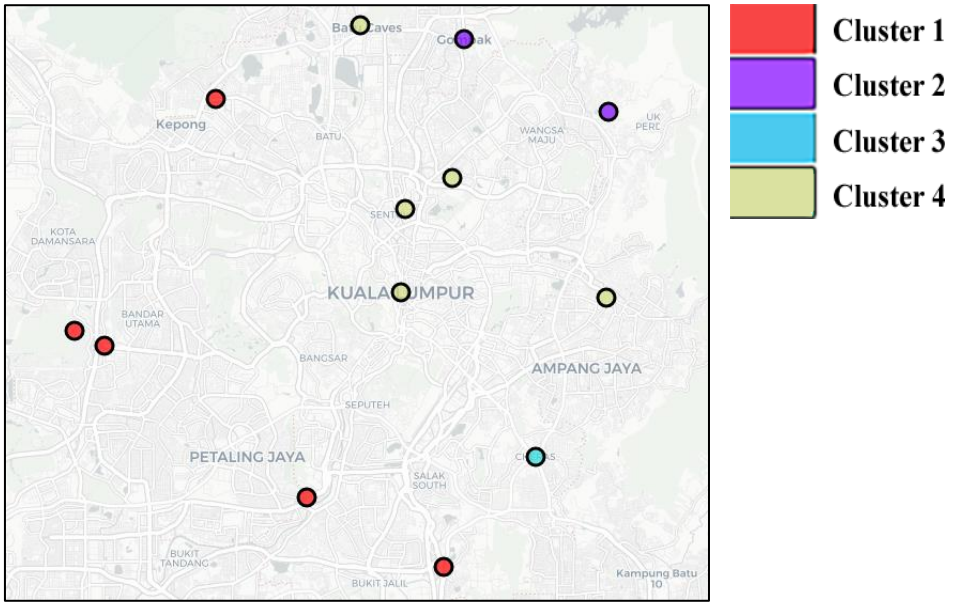


Figure 6: Food Clusters in KL

As seen from Figure 6 after conducting KMeans Clustering, there are a total of 4 clusters which their elements are similar among each other. Districts are distributed into clusters as follow:

Cluster 1: Damansara, Kepong, Petaling, Petaling Jaya and Sungai Besi

Cluster 2: Gombak and Hulu Kelang

Cluster 3: Cheras

Cluster 4: Ampang, Batu Caves, Kuala Lumpur, Sentul and Setapak

Through further examination on the details in the clusters*(refer to tables 3-6 in appendices),* it can be decided that Cluster 1 and 3 are both Café and Chinese Food clusters, difference being Cluster 1 has a hint of bakeries in it; Cluster 2 is a Malay and Fast Food cluster ;Cluster 4 is a Chinese and Malay Food cluster.

Medical Centres

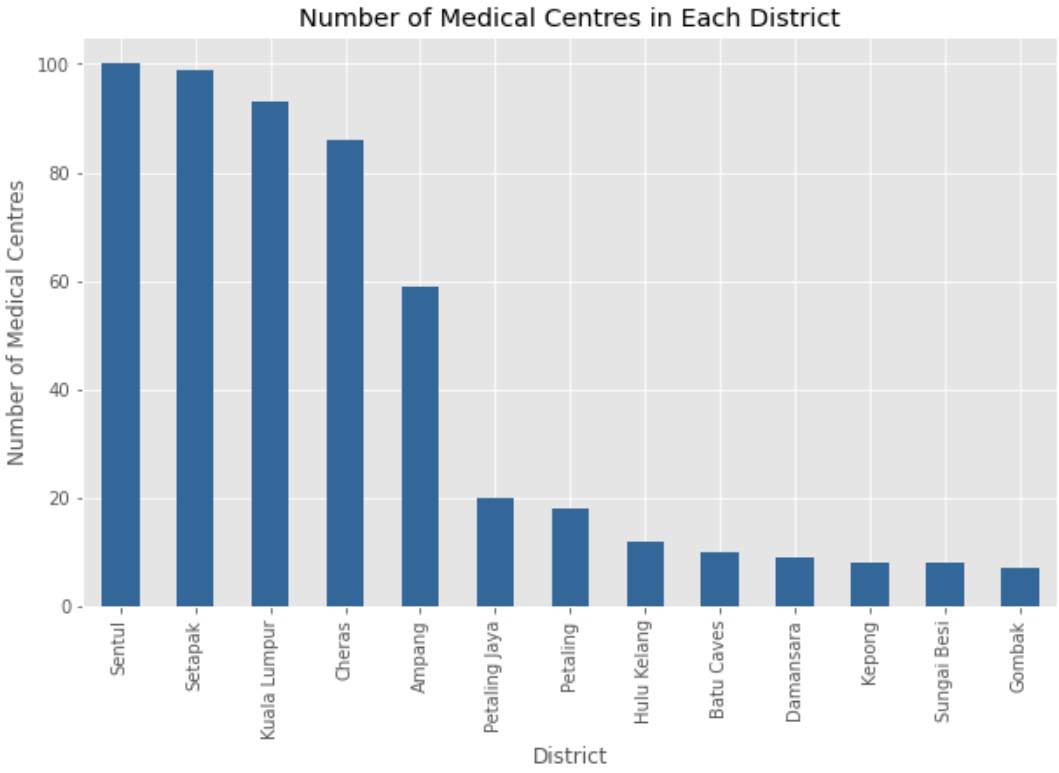


Figure 20: Medical Centres in Each District

As seen in Figure 20, Sentul, Setapak, Kuala Lumpur, Cheras and Ampang has a lot of Medical Centres while the remaining districts have significantly less.

Schools

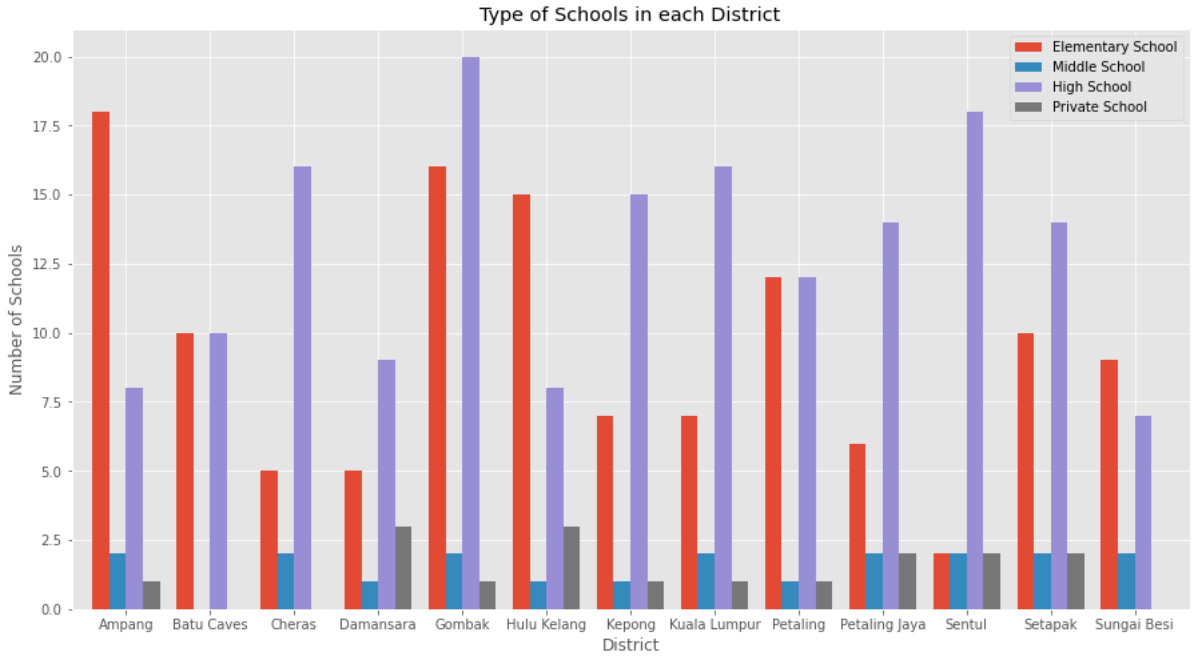


Figure 21: Types of Schools in Each District

Figure 21 illustrates the number of schools by their type in each district. Only Batu Caves has no Middle Schools while Sungai Besi, Batu Caves and Cheras have no Private Schools. Ampang stands as the district with the highest number of Elementary Schools while Sentul the lowest. The number of middle schools in each district is uniformly low. For High School, Gombak has the most while Ampang the least.

Arts & Entertainment

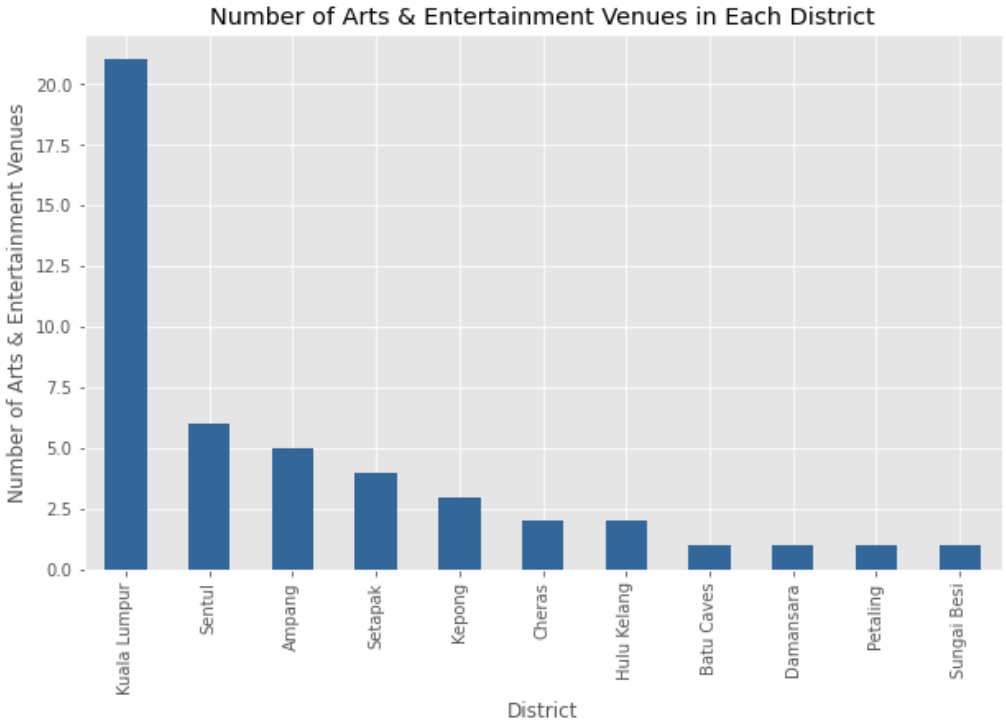


Figure 23: Arts & Entertainment in Each District

As seen from Figure 23, most districts have little to none of such venues except for Kuala Lumpur. It is no surprise that these venues are plentiful in the heart of the city. As Gombak and Petaling Jaya have no venue within their 2.5KM radius, they are not shown in the figure above.

Limitations

- The Sandbox Tier Accounts for the FOURSQUARE API only allows 100 results from each call. Hence, not all data may have been captured during data collection.
- The selection of relevant subcategories is all up to personal judgement and may be subjected to error of judgement.
- The categorisation of F&B Outlet subcategories is all up to personal judgement and may be subjected to error of judgement.

Conclusion

In summary, there are a lot of variables that may impact the choice of lodgings for when one migrates to Kuala Lumpur. However, the different districts allow more flexible options based a person's wants and needs.

When it comes down to F&B Outlets, the districts are more or less similar to one another, providing local delicacies as the most common venues available are that. However, it is worth noting that if Cafes are more sought out, districts in Clusters 1 and 3 are more favourable while Cluster 2 is preferred should you often crave fast foods.

When it comes down to Medical Centres, Sentul, Setapak, Kuala Lumpur, Cheras and Ampang are clearly the winners of this argument. Still, the remaining districts still have such venues within the small investigated radius so it should not cause alert or worry.

When it comes down to Schools, people of foreign origin with children should stray away from Sungai Besi, Batu Caves and Cheras as there are no Private Schools nearby. Elementary schools are not a problem as seen in the sheer amount in each district. Should one be interested in more high school choices then Gombak would be the best bet.

Finally, when it comes to Arts & Entertainment venues, it goes without saying that Kuala Lumpur beats the other districts by a long shot.

References

DOSM. (n.d.). *Federal Territory of Kuala Lumpur*. Retrieved from Department of Statistics Malaysia: https://www.dosm.gov.my/v1/index.php?r=column/cone&menu_id=bjRlZXVGdnBueDJKY1BPWEFPRlhIdz09

Fadil, N. (2013). *MoonRise*. Retrieved from Flickr: https://www.flickr.com/photos/naimfadil/10121316715

Appendices

Category	Subcategory Kept
F&B Outlets	'Thai Restaurant', 'Food Truck', 'Asian Restaurant', 'Fast Food Restaurant', 'Middle Eastern Restaurant', 'Malay Restaurant', 'Chinese Restaurant', 'Burger Joint', 'Pakistani Restaurant', 'Korean Restaurant', 'Halal Restaurant', 'Indian Restaurant', 'Noodle House', 'Restaurant', 'Café', 'Pizza Place', 'Italian Restaurant', 'Indonesian Restaurant', 'Seafood Restaurant', 'Falafel Restaurant', 'Vegetarian / Vegan Restaurant', 'Breakfast Spot', 'Afghan Restaurant', Steakhouse', 'Mexican Restaurant', 'Bakery', 'Fried Chicken Joint', 'Food Court', 'Japanese Restaurant', 'Comfort Food Restaurant', 'Taiwanese Restaurant', 'Sushi Restaurant', 'Soup Place', 'Sandwich Place', 'Diner', 'Food Stand', 'Portuguese Restaurant', 'Bistro', 'Snack Place' ,'BBQ Joint', 'Hainan Restaurant', 'Salad Place', 'Kebab Restaurant', 'Hotpot Restaurant', 'Food', 'Brazilian Restaurant', 'Gastropub', 'American Restaurant' ,'Fish & Chips Shop', 'Donut Shop', 'Eastern European Restaurant', 'Szechuan Restaurant', 'Cantonese Restaurant', 'Hakka Restaurant', 'Pet Café', 'Cafeteria', 'South Indian Restaurant', 'Chettinad Restaurant', 'Latin American Restaurant', 'Tapas Restaurant', 'Spanish Restaurant', 'French Restaurant', 'Dim Sum Restaurant', 'Mamak Restaurant', 'Yakitori Restaurant', 'Udon Restaurant', 'Iraqi Restaurant', 'Hunan Restaurant', 'Hong Kong Restaurant', 'Satay Restaurant', 'Moroccan Restaurant', 'Chinese Breakfast Place', 'Wings Joint', 'Dumpling Restaurant', 'Ramen Restaurant'
Medical Facility	'Hospital', 'Medical Center', 'Emergency Room', 'Office', 'Hospital Ward', 'Pharmacy'
Schools	'School', 'High School', 'Private School', 'Elementary School', 'Preschool', 'Nursery School', 'Middle School'
Arts & Entertainment	'Arts & Crafts Store', 'Art Gallery', 'Dance Studio', 'Music Venue', 'Spa', 'Soccer Field', 'Gym', 'Museum', 'Science Museum', 'History Museum', 'Gym Pool', 'Theater', 'Art Museum', 'Soccer Stadium'

Table 1:Subcategories making up the Categories

New Category	Old subcategory
Chinese Food	Chinese Restaurant, Seafood Restaurant , Noodle House , Taiwanese Restaurant , Soup Place , Hainan Restaurant , Hotpot Restaurant , Szechuan Restaurant , Cantonese Restaurant , Hakka Restaurant , Dim Sum Restaurant , Hunan Restaurant , Hong Kong Restaurant , Chinese Breakfast Place , Dumpling Restaurant
Malay Food	Malay Restaurant , Halal Restaurant , Satay Restaurant
Indian Food	Indian Restaurant , South Indian Restaurant , Chettinad Restaurant , Mamak Restaurant
Food Court	Food Court , Cafeteria
Japanese Food	Japanese Restaurant , Sushi Restaurant , Yakitori Restaurant , Udon Restaurant , Ramen Restaurant
Other Asian Food	Thai Restaurant , Asian Restaurant , Pakistani Restaurant , Korean Restaurant , Afghan Restaurant , Portuguese Restaurant , BBQ Joint
Western Food	Pizza Place , Italian Restaurant , Breakfast Spot , Steakhouse , Mexican Restaurant , Diner , American Restaurant , Fish & Chips Shop , Brazilian Restaurant , Eastern European Restaurant , Latin American Restaurant , Spanish Restaurant , French Restaurant
Fast Food	Food Truck , Fast Food Restaurant , Burger Joint , Fried Chicken Joint , Sandwich Place , Food Stand , Snack Place , Wings Joint
Bakery	Bakery , Donut Shop
Cafés	Café , Pet Café
Vegetarian Food	Vegetarian / Vegan Restaurant , Salad Place
Others	Middle Eastern Restaurant , Restaurant , Indonesian Restaurant , Falafel Restaurant , Comfort Food Restaurant , Kebab Restaurant , Food , Tapas Restaurant , Iraqi Restaurant , Moroccan Restaurant , Bistro , Gastropub

Table 2: Subcategories making up the new F&B Categories

	District	1st Highest Number of F&B Venues	2nd Highest Number of F&B Venues	3rd Highest Number of F&B Venues	4th Highest Number of F&B Venues	5th Highest Number of F&B Venues
3	Damansara	Cafés	Chinese Food	Fast Food	Western Food	Bakery
6	Kepong	Chinese Food	Fast Food	Cafés	Western Food	Vegetarian Food
8	Petaling	Chinese Food	Cafés	Indian Food	Malay Food	Fast Food
9	Petaling Jaya	Chinese Food	Cafés	Fast Food	Bakery	Western Food
12	Sungai Besi	Chinese Food	Fast Food	Malay Food	Cafés	Indian Food

Table 3: Cluster 1 in F&B Categories

	District	1st Highest Number of F&B Venues	2nd Highest Number of F&B Venues	3rd Highest Number of F&B Venues	4th Highest Number of F&B Venues	5th Highest Number of F&B Venues
4	Gombak	Malay Food	Fast Food	Chinese Food	Western Food	Cafés
5	Hulu Kelang	Malay Food	Fast Food	Cafés	Western Food	Chinese Food

Table 4: Cluster 2 in F&B Categories

	District	1st Highest Number of F&B Venues	2nd Highest Number of F&B Venues	3rd Highest Number of F&B Venues	4th Highest Number of F&B Venues	5th Highest Number of F&B Venues
2	Cheras	Chinese Food	Cafés	Malay Food	Fast Food	Western Food

Table 5: Cluster 3 in F&B Categories

	District	1st Highest Number of F&B Venues	2nd Highest Number of F&B Venues	3rd Highest Number of F&B Venues	4th Highest Number of F&B Venues	5th Highest Number of F&B Venues
0	Ampang	Chinese Food	Malay Food	Western Food	Fast Food	Cafés
1	Batu Caves	Malay Food	Chinese Food	Fast Food	Indian Food	Western Food
7	Kuala Lumpur	Cafés	Indian Food	Malay Food	Chinese Food	Western Food
10	Sentul	Chinese Food	Malay Food	Indian Food	Cafés	Fast Food
11	Setapak	Chinese Food	Malay Food	Fast Food	Cafés	Indian Food

Table 6: Cluster 4 in F&B Categories

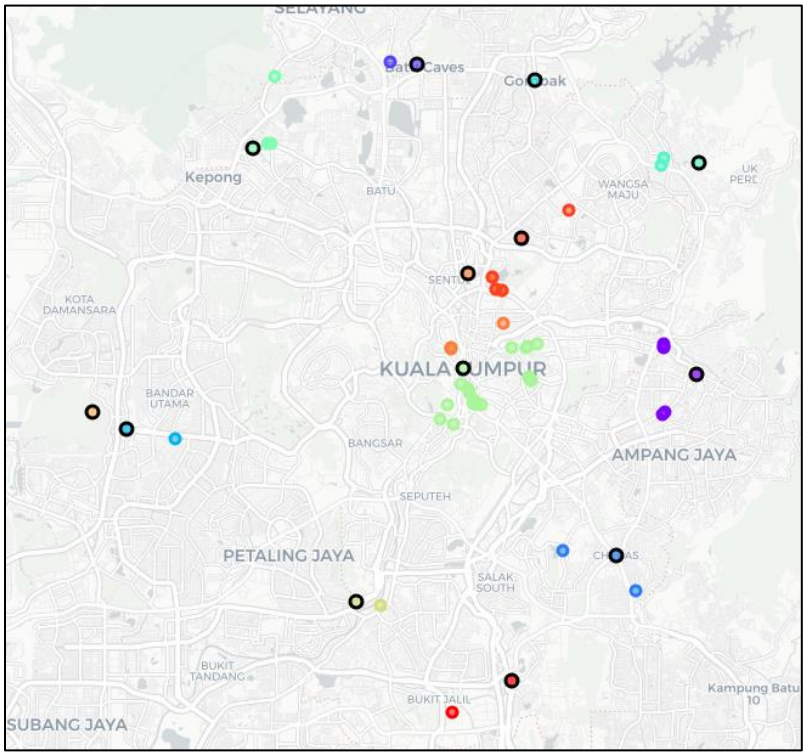


Figure 2: Arts & Entertainment Venues in Kuala Lumpur

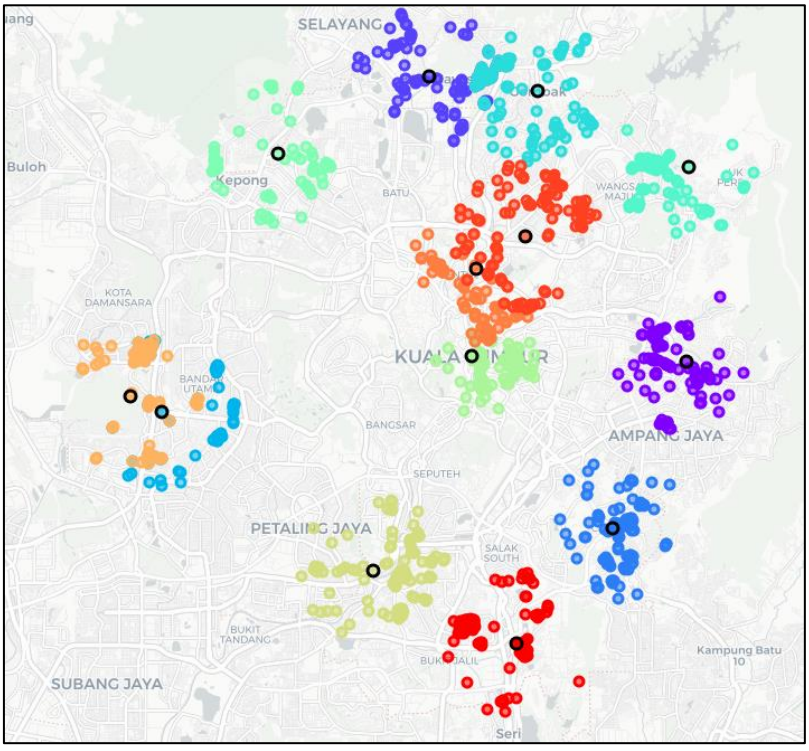


Figure 3: F&B Outlets in Kuala Lumpur

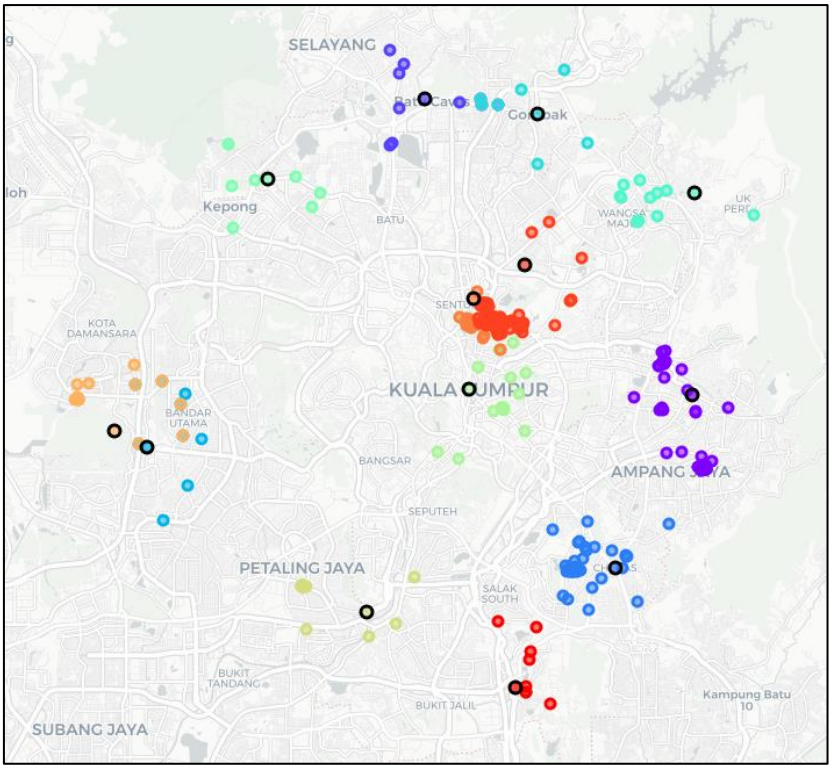


Figure 4: Medical Facilities in Kuala Lumpur

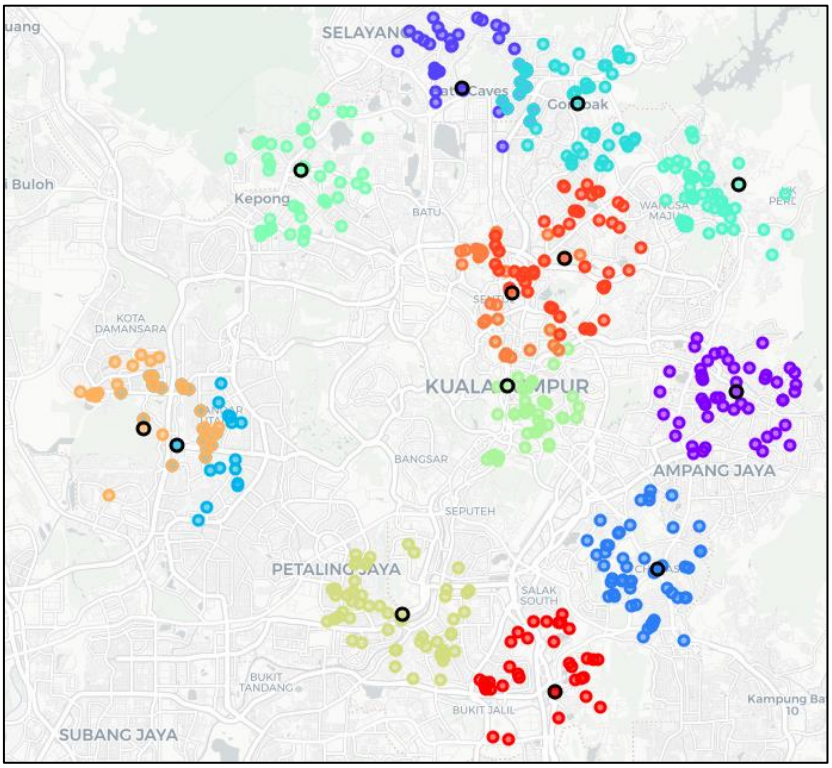
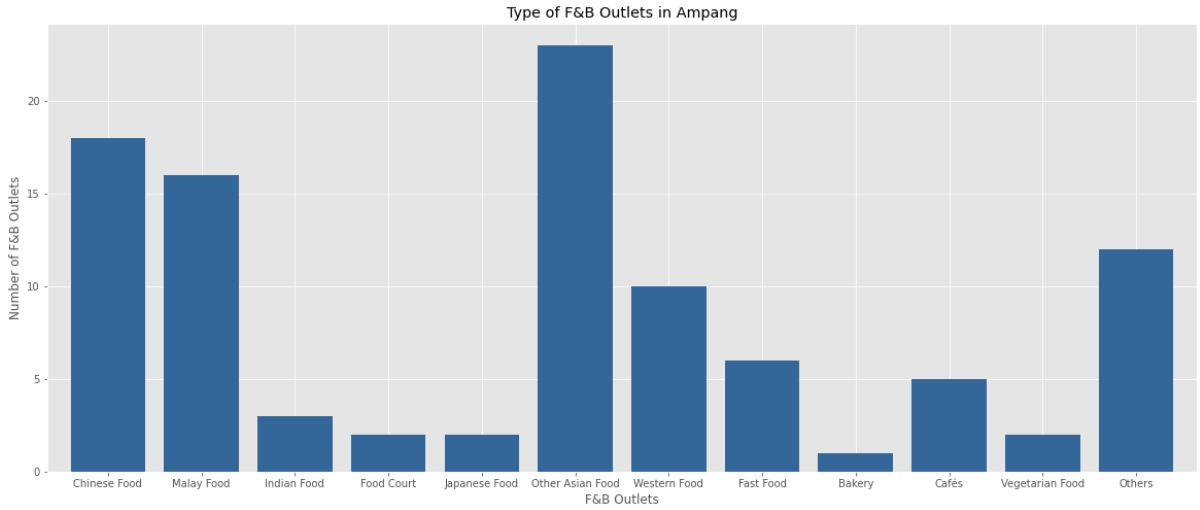
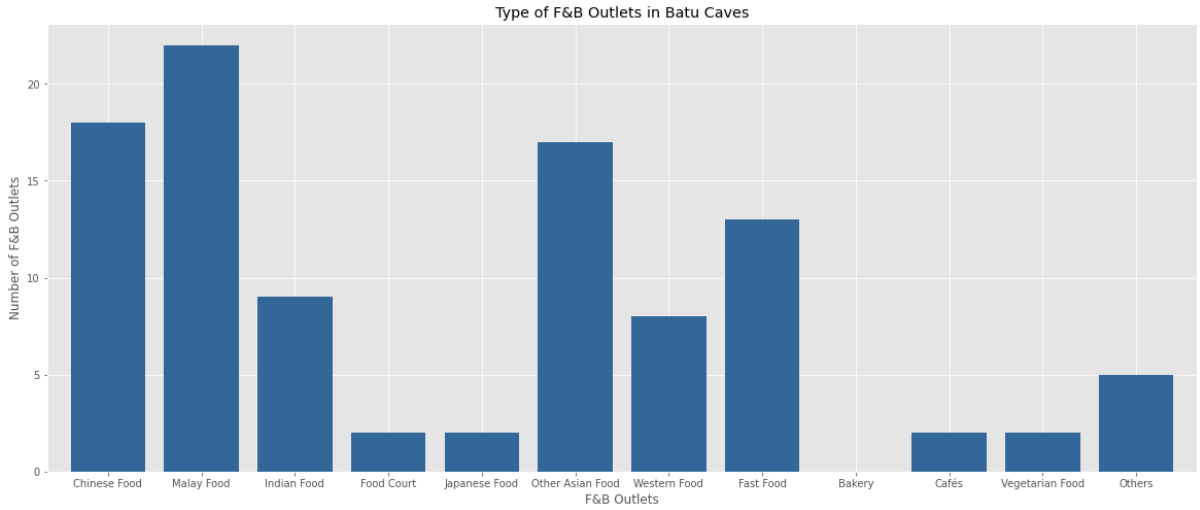
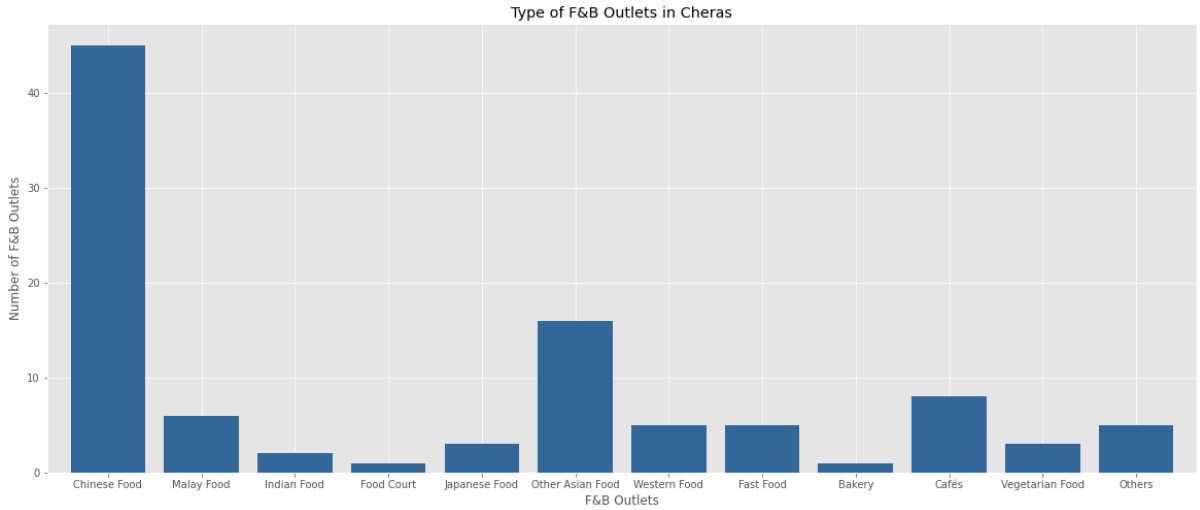


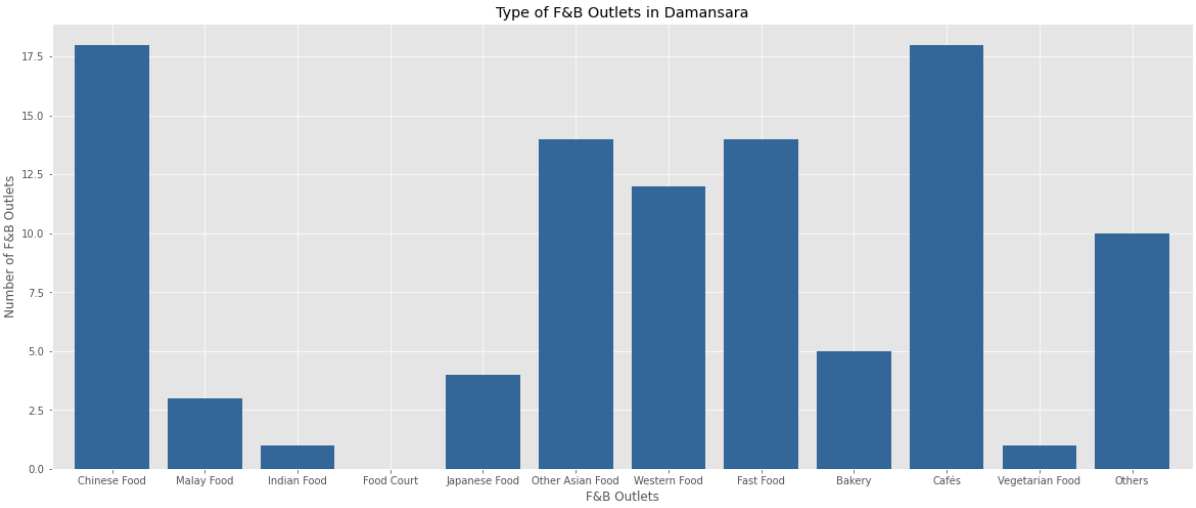
Figure 5: Schools in Kuala Lumpur

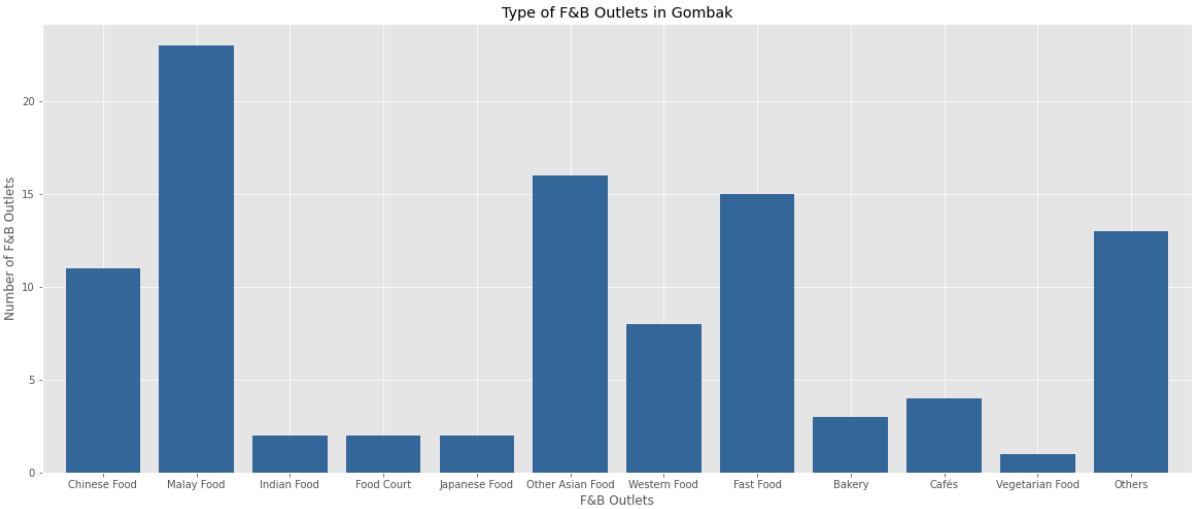
Figures 7-19: Types of F&B Outlets in {district}

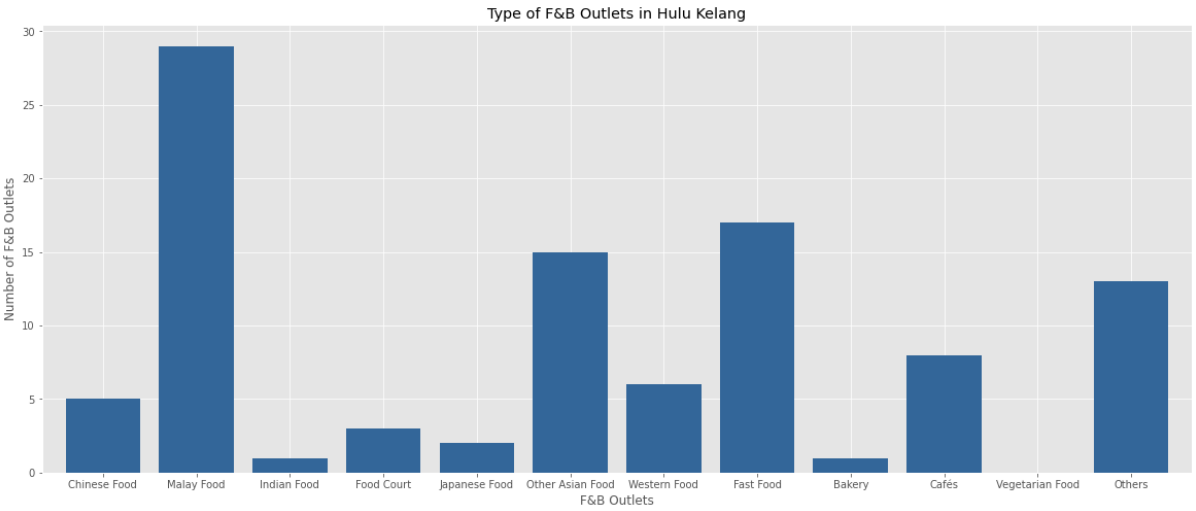


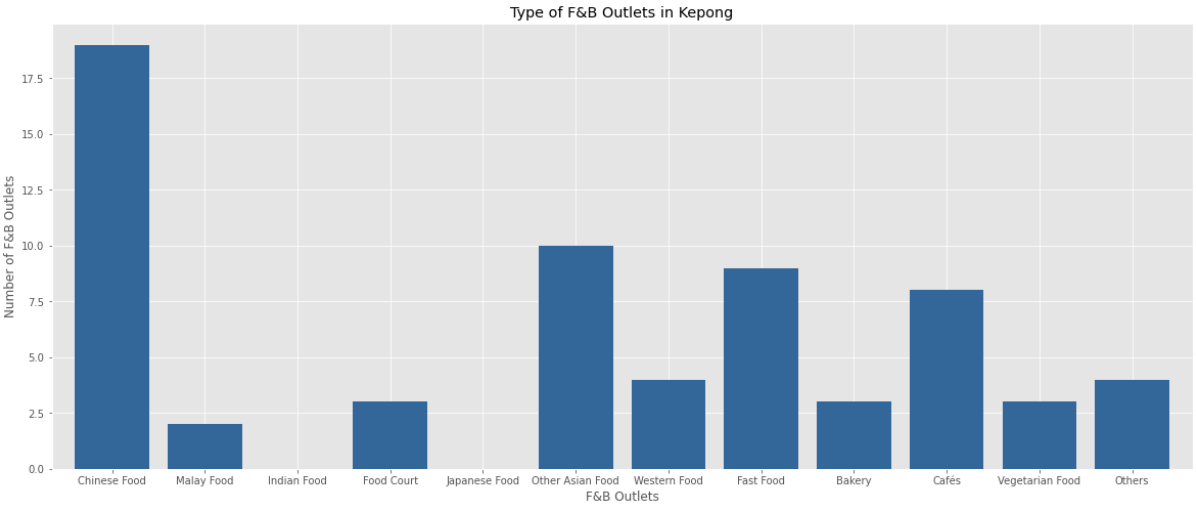


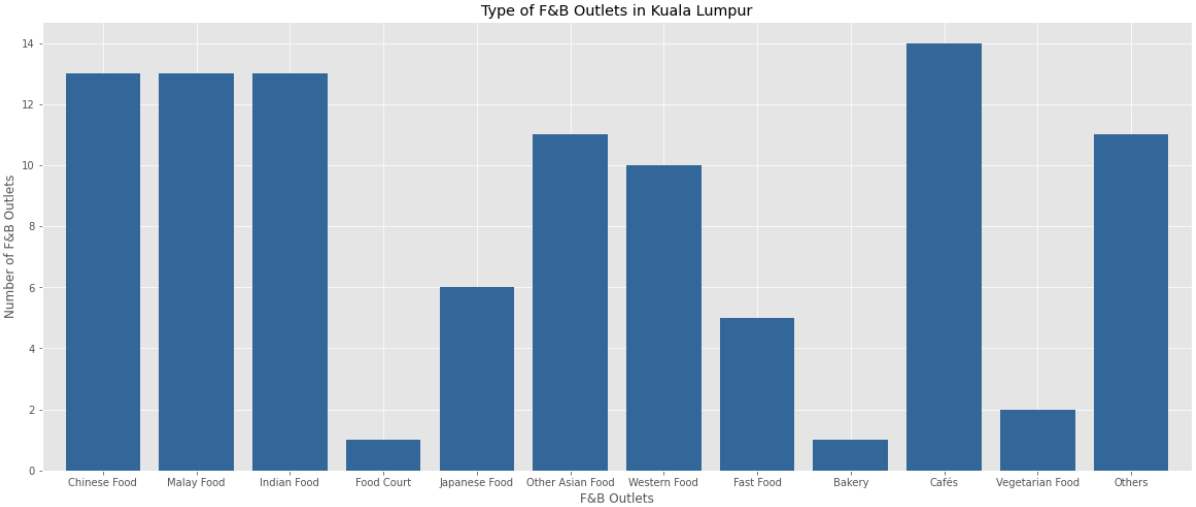


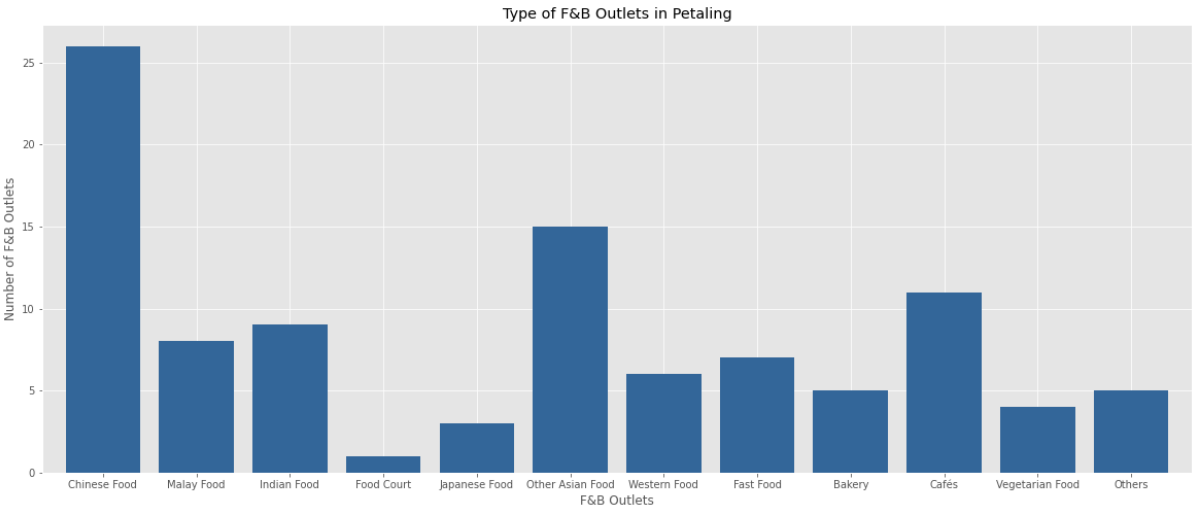

Type of F&B Outlets in Damansara
Number of F&B Outlets
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Chinese Food
Malay Food
Indian Food
Food Court
Japanese Food
Other Asian Food
Western Food
Fast Food
Bakery
Cafés
Vegetarian Food
Others
F&B Outlets

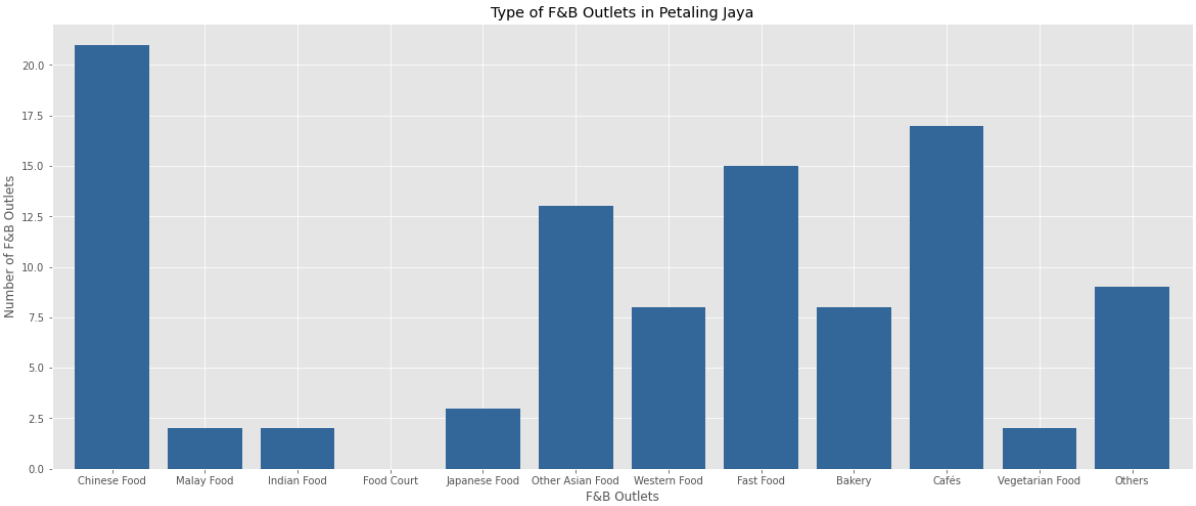

Type of F&B Outlets in Gombak
Number of F&B Outlets
0
5
10
15
20
Chinese Food
Malay Food
Indian Food
Food Court
Japanese Food
Other Asian Food
Western Food
Fast Food
Bakery
Cafés
Vegetarian Food
Others
F&B Outlets

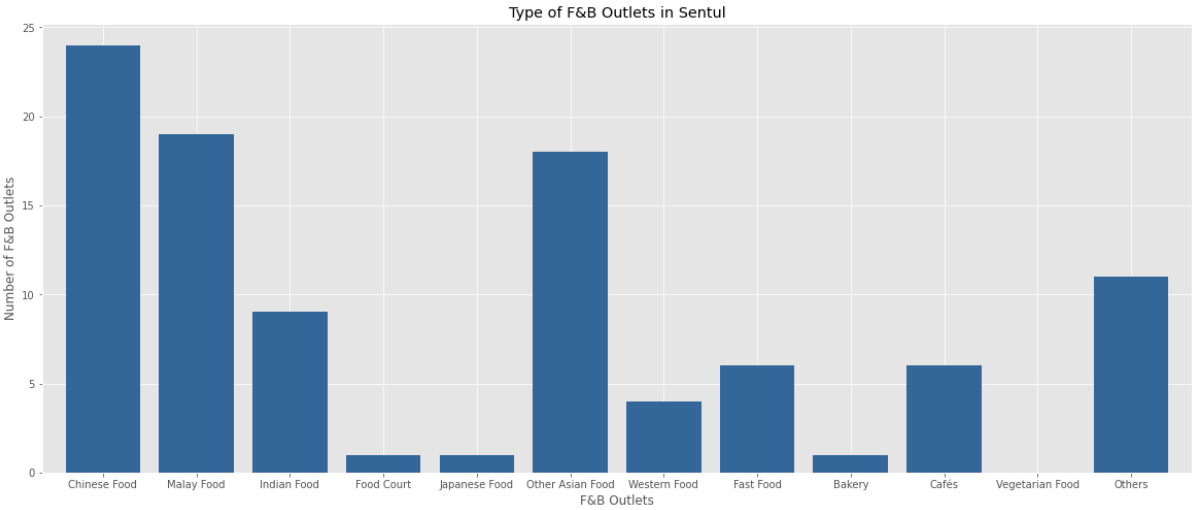

Type of F&B Outlets in Hulu Kelang
Number of F&B Outlets
0
5
10
15
20
25
30
Chinese Food
Malay Food
Indian Food
Food Court
Japanese Food
Other Asian Food
Western Food
Fast Food
Bakery
Cafés
Vegetarian Food
Others
F&B Outlets

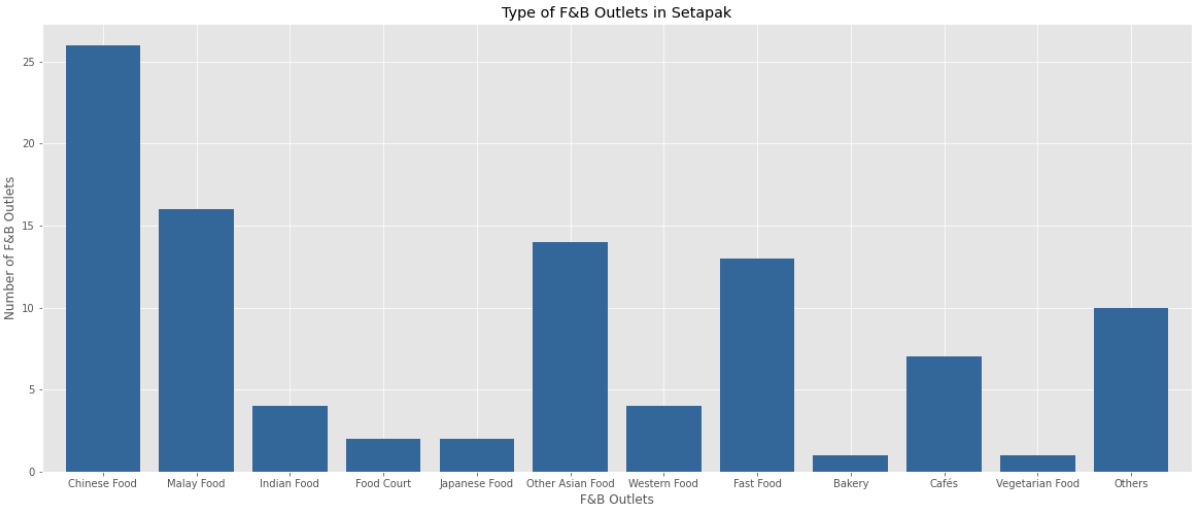

Type of F&B Outlets in Kepong
Number of F&B Outlets
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Chinese Food
Malay Food
Indian Food
Food Court
Japanese Food
Other Asian Food
Western Food
Fast Food
Bakery
Cafés
Vegetarian Food
Others
F&B Outlets

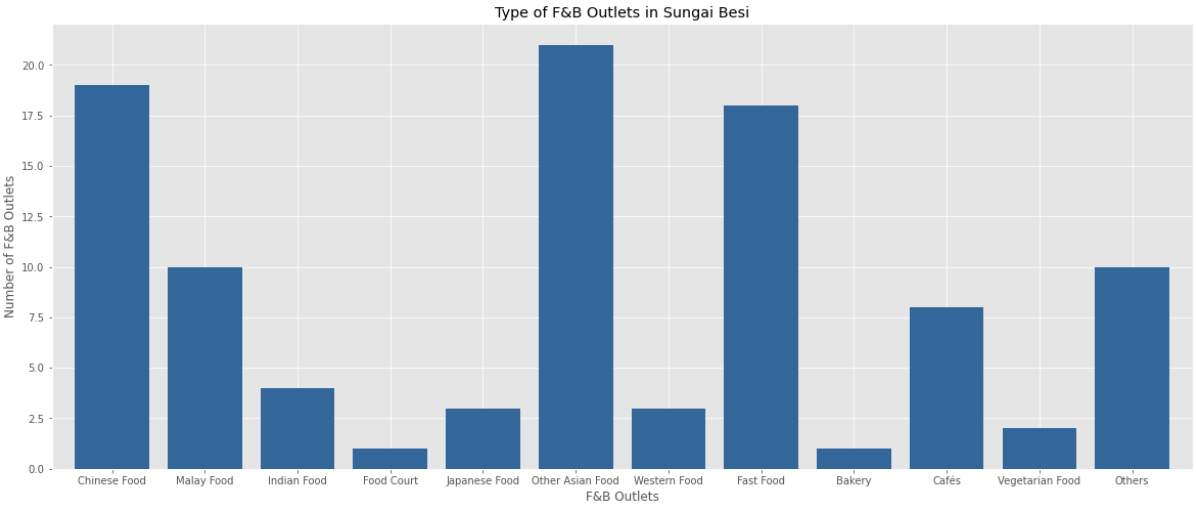

Type of F&B Outlets in Kuala Lumpur
Number of F&B Outlets
0
2
4
6
8
10
12
14
Chinese Food
Malay Food
Indian Food
Food Court
Japanese Food
Other Asian Food
Western Food
Fast Food
Bakery
Cafés
Vegetarian Food
Others
F&B Outlets


Type of F&B Outlets in Petaling
Number of F&B Outlets
0
5
10
15
20
25
Chinese Food
Malay Food
Indian Food
Food Court
Japanese Food
Other Asian Food
Western Food
Fast Food
Bakery
Cafés
Vegetarian Food
Others
F&B Outlets


Type of F&B Outlets in Petaling Jaya
Number of F&B Outlets
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Chinese Food
Malay Food
Indian Food
Food Court
Japanese Food
Other Asian Food
Western Food
Fast Food
Bakery
Cafés
Vegetarian Food
Others
F&B Outlets


Type of F&B Outlets in Sentul
Number of F&B Outlets
0
5
10
15
20
25
Chinese Food
Malay Food
Indian Food
Food Court
Japanese Food
Other Asian Food
Western Food
Fast Food
Bakery
Cafés
Vegetarian Food
Others
F&B Outlets


Type of F&B Outlets in Setapak
Number of F&B Outlets
0
5
10
15
20
25
Chinese Food
Malay Food
Indian Food
Food Court
Japanese Food
Other Asian Food
Western Food
Fast Food
Bakery
Cafés
Vegetarian Food
Others
F&B Outlets


Type of F&B Outlets in Sungai Besi
Number of F&B Outlets
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Chinese Food
Malay Food
Indian Food
Food Court
Japanese Food
Other Asian Food
Western Food
Fast Food
Bakery
Cafés
Vegetarian Food
Others
F&B Outlets

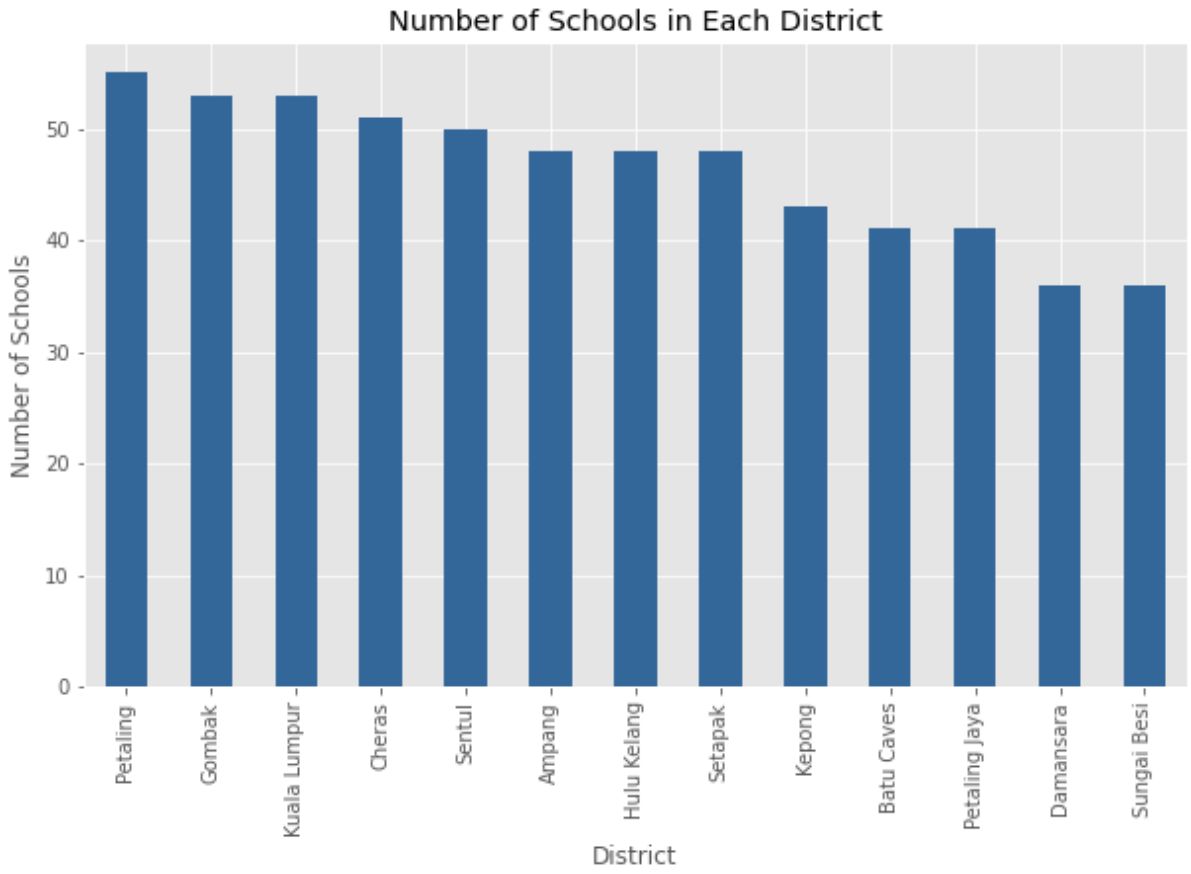


Figure 22: Number of Schools in Each District