CAPSTONE PROJECT: Identification of Location to open a new Western Café in Kuala Lumpur, Malaysia

By: Jayaprakash Saththasivam

1.0 Introduction

Little Noshery (LN) is a famous café operating in Marble Hill (MH), Manhattan. The management of the café recently decided to open their very first international outlet in Kuala Lumpur, Malaysia. The business plan is to set up a new western-style café to serve the international community and tourists of Kuala Lumpur (KL). As KL is one of the hotspots for international business and tourism activities in South East Asia, the management believes that operating unique western café at a proper location would be profitable.

As a part of their new plan, the management of LN has hired a data scientist to identify a suitable location in KL. To complete this task, the data scientist adopted the following sequence of approach:-

i. Firstly, the data scientist explored the nearby venues of the existing cafe located at Marble Hill, Manhattan as the management of LN described that one of the reasons behind their successful operation at MH is the location of their café. By doing this, the data scientist will be able to identify the popular venue categories of Marble Hill.

ii. Secondly, the districts in Kuala Lumpur and its nearby venues will be listed. The location for a new café in one of the districts of Kuala Lumpur will be recommended to the LN management by finding similar nearby venue categories as the parent café in Marble Hill.

Using this approach, we can be optimistic as well as certain that the new café will be surrounded by similar venues categories and thus minimizes the business risk that could arise from choosing an improper location. For instance, if the nearby venues in Marble Hill are cafes and restaurants, it will be wise to choose a location in a KL district that is also surrounded by cafes and restaurants rather than choosing a district that is surrounded by a furniture shop or art gallery. One could argue that suggesting a location that is surrounded by café and restaurants could increase the competition and eventually lead to poor sales but the management of LN is confident that their unique menus and reasonable pricing will attract customers. Hence, this approach will be beneficial for LN since they have prior experience operating at a similar location.

2.0 Description of Data

The data scientist used using Foursquare location data to identify the top ten venue categories in Marble Hill. Similarly, he also used Foursquare location data to identify the venues and venue categories in all the districts of Kuala Lumpur. The top ten categories of each district were determined by analyzing the raw data. The recommendation on the final location was made based on the similarities of the venue categories between Marble Hill and Kuala Lumpur.

2.1 Data for Marble Hill

The coordinates for Marble Hill were obtained using geopy library function (Fig 1 and 2). The venues nearby Marble Hill (within 1500 m with a search limit = 100) were explored using Foursquare location data.

```
def LocationName(location1):
    geolocator1 = Nominatim(user_agent='openmoves.net')
    try:
        location1 = geolocator1.geocode(location1)
    except:
        raise Exception("There was a problem with the geolocator function")
    return location1.latitude , location1.longitude
```

Fig.1. Geopy Function to obtain latitude and longitude

```
address = "Marble Hill"
LatMH=LocationName(address)[0]
LongMH=LocationName(address)[1]
print('The geograpical coordinate of {} is {}°N, {}°W'.format(address, LatMH, -1*LongMH))
```

```
The geograpical coordinate of Marble Hill is 40.8762983°N, 73.9104292°W
```

Fig.2. Coordinate of Marble Hill

2.2 Data for districts in Kuala Lumpur

The coordinates for Kuala Lumpur were obtained using geopy library. As there was no information available on the districts of Kuala Lumpur in a tabular format, the data scientist has created a list of

districts (Fig.3) in Kuala Lumpur and obtained their respective coordinates from geopy library using a FOR loop (Fig.4). The location of all the eleven districts in Kuala Lumpur was then plotted using a folium map (Fig 5).

```
KL_district_list = ["Bukit Bintang", "Titiwangsa", "Setiawangsa", "Wangsa Maju","Batu, Kl", "Kepong", "Segambut","Lembah Pantai","Seputeh","Bandar Tun Razak","Cheras"]
df = pd.DataFrame(KL_district_list, columns=["District"])
df
```

	District
0	Bukit Bintang
1	Titiwangsa
2	Setiawangsa
3	Wangsa Maju
4	Batu, Kl
5	Kepong
6	Segambut
7	Lembah Pantai
8	Seputeh
9	Bandar Tun Razak
10	Cheras

Fig.3. Create a list for districts in Kuala Lumpur

```
lat=[]
long=[]
sequence=df["District"]
for i in range(len(sequence)):
    lat1=LocationName((sequence[i]))[0]
    long1=LocationName((sequence[i]))[1]
    lat.append(lat1)
    long.append(long1)
```

```
df["Latitude"] = lat
df["Longitude"]= long
df
```

	District	Latitude	Longitude
0	Bukit Bintang	3.147107	101.708601
1	Titiwangsa	3.173145	101.695933
2	Setiawangsa	3.175725	101.735884
3	Wangsa Maju	3.205667	101.731908
4	Batu, Kl	3.201823	101.671022
5	Kepong	3.205933	101.623711
6	Segambut	3.186437	101.664205
7	Lembah Pantai	3.104444	101.672189
8	Seputeh	3.113687	101.681420
9	Bandar Tun Razak	3.089695	101.712467
10	Cheras	3.107178	101.716490

Fig.4. FOR loop to obtain the coordinates of the districts using geopy function

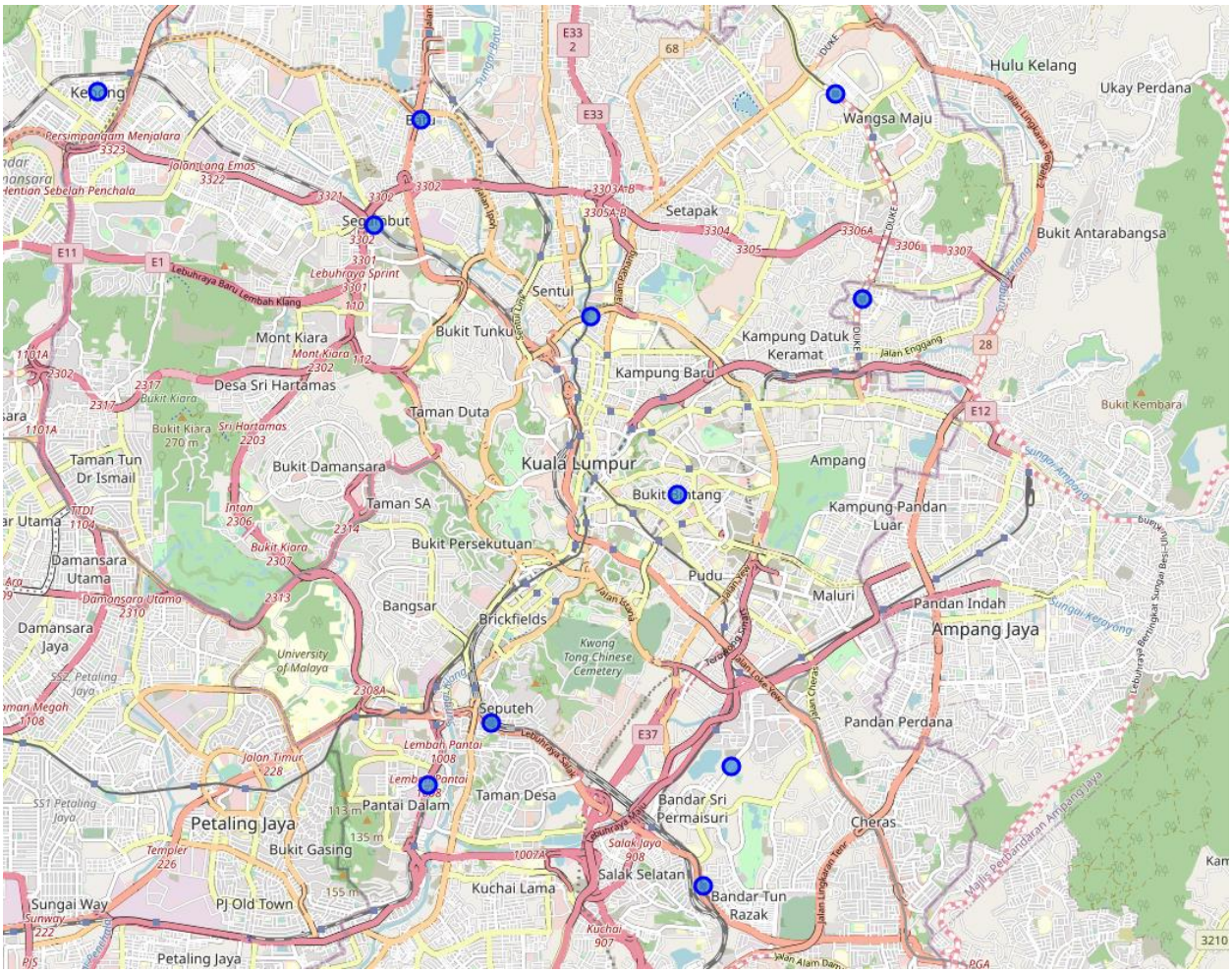

Fig 5 Location of districts in Kuala Lumpur

3.0 Methodology

The data scientist has used several python libraries to achieve his task. The libraries are (i) numpy (ii) pandas (iii) json (iv) geopy (v) matplotlib (vi) folium and (vii) scikit-learn.

After obtaining the location of Marble Hill using geopy function, the venues nearby Marble Hill (within 1500 m with a search limit = 100) were explored using Foursquare location data. The json output data received from Foursquare was then cleaned to obtain the venue name, categories, latitude, and longitude. The results were further analyzed to obtain the frequency of each venue category and finally sorted to identify the top ten venue categories in Marble Hill. This particular result is important as the data scientist would like to identify similar venue categories in the districts of Kuala Lumpur. Using a similar approach, the data scientist again used the Foursquare location to explore all the nearby venues

in each of the eleven districts in Kuala Lumpur. The top ten venues categories of each district were also identified. Finally, KMeans from sklearn cluster was used to cluster the districts in Kuala Lumpur based on the venue categories. The clusters were then compared against Marble Hill.

4.0 Results

Results related to Marble Hill are in shown in Fig 6. It can be seen that Foursquare Location returned 100 venues and 51 unique venue categories in Marble Hill. Frequency calculation shows that Pizza place and Mexican Restaurant are the top two venue categories in Marble Hill with a combined frequency of 14%. These were then followed by parks, bakeries, cafés, coffee shops, bars, and diners at 4% each. These distributions revealed that nearby venues in Marble Hill are dominated by food outlets of western style.

	Venue	Venue Category	Venue Latitude	Venue Longitude
0	Bikram Yoga	Yoga Studio	40.876844	-73.906204
1	Tibbett Diner	Diner	40.880404	-73.908937
2	Arturo's	Pizza Place	40.874412	-73.910271
3	Sam's Pizza	Pizza Place	40.879435	-73.905859
4	El Malecon	Caribbean Restaurant	40.879338	-73.904457

143]

```
print('There are {} venues in Marble Hill.'.format(nearby_venues.count()[1]))
print('There are {} uniques categories in Marble Hill.'.format(len(nearby_venues['Venue Category'].unique())))
```

```
There are 100 venues in Marble Hill.
There are 51 uniques categories in Marble Hill.
```

Find the top ten venue categories in Marble Hill

144]

```
top10marblehill = nearby_venues['Venue Category'].value_counts()/ len(nearby_venues['Venue Category'])
top10marblehill.head(10)
```

```
Pizza Place           0.07
Mexican Restaurant    0.07
Park                  0.04
Bakery                0.04
Café                  0.04
Coffee Shop           0.04
Bar                   0.04
Diner                 0.04
Wine Shop             0.03
Deli / Bodega         0.03
Name: Venue Category, dtype: float64
```

-]

Fig.6. Top ten venue categories in Marble Hill obtained from Foursquare

Analyses using Foursquare data for the districts in Kuala Lumpur are shown in Figure 7. It can be seen that there are a total of 1040 venues with 188 different unique venues categories across the eleven districts. It can be also summarized from the data that Batu, KL has the lowest number of venues. The top 10 venue categories of each district in Kuala Lumpur are shown in Figure 8.

[375]
```
KL_venues.groupby('District').count()
VenueCatperDis=KL_venues.groupby('District').count()
VenueCatperDis[["Venue"]]
```

District	Venue
Bandar Tun Razak	100
Batu, KL	60
Bukit Bintang	100
Cheras	92
Kepong	88
Lembah Pantai	100
Segambut	100
Seputeh	100
Setiawangsa	100
Titiwangsa	100
Wangsa Maju	100

Let's find out how many unique categories can be curated from all the returned venues

[376]
```
print('There are a total of {} venues in Kuala Lumpur.'.format(VenueCatperDis['Venue Category'].sum()))
print('There are {} uniques categories.'.format(len(KL_venues['Venue Category'].unique())))
```

```
There are a total of 1040 venues in Kuala Lumpur.
There are 188 uniques categories.
```

Fig.7 Number of venues and venue categories obtained from Foursquare for all the districts in Kuala Lumpur.

	District	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	10th Most Common Venue
0	Bandar Tun Razak	Chinese Restaurant	Asian Restaurant	Malay Restaurant	Indian Restaurant	Convenience Store	Gas Station	Café	Thai Restaurant	Fast Food Restaurant	Indonesian Restaurant
1	Batu, KL	Chinese Restaurant	Asian Restaurant	Malay Restaurant	Seafood Restaurant	Food Truck	Restaurant	Burger Joint	Motorcycle Shop	Sandwich Place	Café
2	Bukit Bintang	Hotel	Café	Clothing Store	Bar	Spa	Japanese Restaurant	Boutique	Coffee Shop	Shopping Mall	Malay Restaurant
3	Cheras	Chinese Restaurant	Malay Restaurant	Asian Restaurant	Indian Restaurant	Convenience Store	Indonesian Restaurant	Pizza Place	Italian Restaurant	Food Truck	Seafood Restaurant
4	Kepong	Chinese Restaurant	Café	Noodle House	Asian Restaurant	Restaurant	Food Truck	Coffee Shop	Fast Food Restaurant	Dessert Shop	Park
5	Lembah Pantai	Japanese Restaurant	Café	Malay Restaurant	Chinese Restaurant	Coffee Shop	Asian Restaurant	Restaurant	Pet Store	Fast Food Restaurant	Convenience Store
6	Segambut	Chinese Restaurant	Korean Restaurant	Malay Restaurant	Convenience Store	Coffee Shop	Asian Restaurant	Japanese Restaurant	Thai Restaurant	Noodle House	Indian Restaurant
7	Seputeh	Japanese Restaurant	Ice Cream Shop	Café	Bakery	Clothing Store	Cosmetics Shop	Chinese Restaurant	Food Court	Department Store	Boutique
8	Setiawangsa	Malay Restaurant	Asian Restaurant	Thai Restaurant	Café	Bakery	Middle Eastern Restaurant	Food Truck	Park	Seafood Restaurant	Fast Food Restaurant
9	Titiwangsa	Malay Restaurant	Chinese Restaurant	Asian Restaurant	Café	Motorcycle Shop	Hotel	Thai Restaurant	Indian Restaurant	Coffee Shop	Clothing Store
10	Wangsa Maju	Chinese Restaurant	Asian Restaurant	Malay Restaurant	Thai Restaurant	Ice Cream Shop	Café	Burger Joint	Coffee Shop	Boutique	Convenience Store

Fig.8 Top 10 common category venues in all the districts in Kuala Lumpur

The results obtained based one-hot encoding for the venue categories were used as input for the K-MEANS clustering model. The data scientist chose to group the districts into four different clusters. The results of the clusters are shown in a Folium map (Fig 9) and the analyses of each cluster are shown in Figure 10-13.

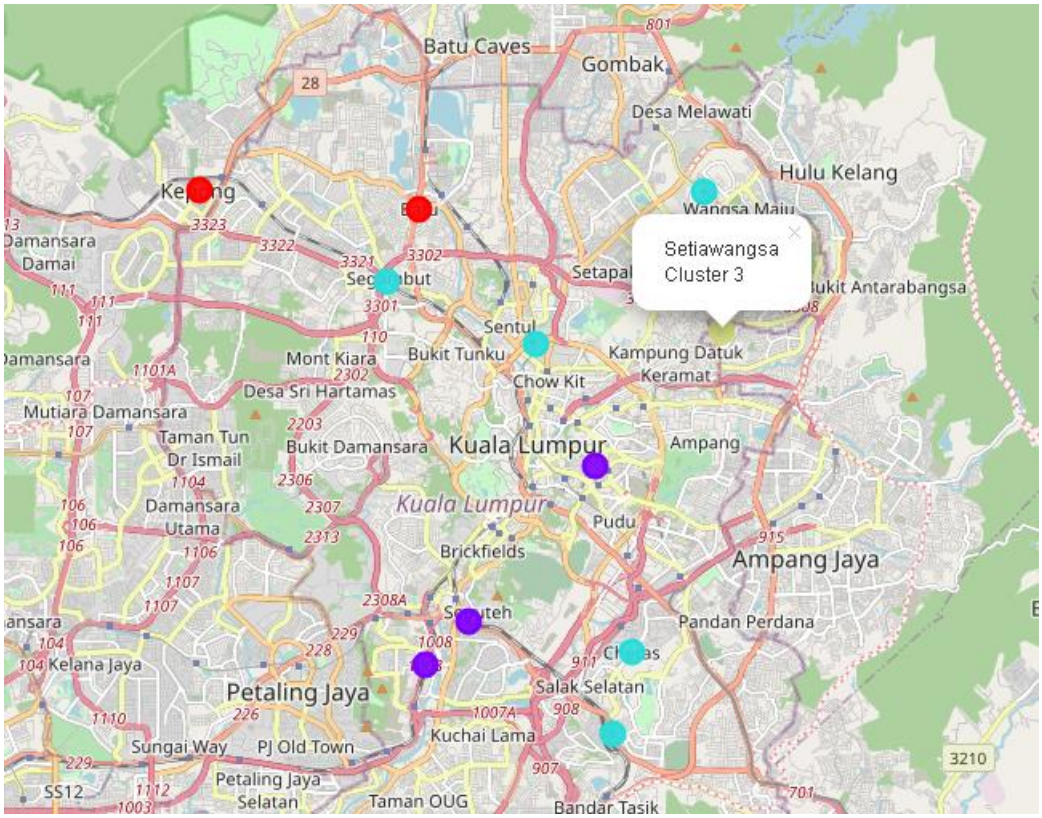

Fig.9. The cluster of districts in Kuala Lumpur (Cluster 0: Red, Cluster 1: Purple and Cluster 2: cyan and Cluster 3: light green)

Cluster 0

```
[388]
KL_merged.loc[KL_merged['Cluster Labels'] == 0, KL_merged.columns[[0] + list(range(4, KL_merged.shape[1]))]]
```

	District	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	10th Most Common Venue
4	Batu, Kl	Chinese Restaurant	Asian Restaurant	Malay Restaurant	Seafood Restaurant	Food Truck	Restaurant	Burger Joint	Motorcycle Shop	Sandwich Place	Café
5	Kepong	Chinese Restaurant	Café	Noodle House	Asian Restaurant	Restaurant	Food Truck	Coffee Shop	Fast Food Restaurant	Dessert Shop	Park

Fig.10. Popular venues categories of districts labeled under Cluster 0

Cluster 1

```
[389] KL_merged.loc[KL_merged['Cluster Labels'] == 1, KL_merged.columns[[0] + list(range(4, KL_merged.shape[1]))]]
```

	District	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	10th Most Common Venue
0	Bukit Bintang	Hotel	Café	Clothing Store	Bar	Spa	Japanese Restaurant	Boutique	Coffee Shop	Shopping Mall	Malay Restaurant
7	Lembah Pantai	Japanese Restaurant	Café	Malay Restaurant	Chinese Restaurant	Coffee Shop	Asian Restaurant	Restaurant	Pet Store	Fast Food Restaurant	Convenience Store
8	Seputeh	Japanese Restaurant	Ice Cream Shop	Café	Bakery	Clothing Store	Cosmetics Shop	Chinese Restaurant	Food Court	Department Store	Boutique

Fig.11. Popular venues categories of districts labeled under Cluster 1

Cluster 2

```
[390] KL_merged.loc[KL_merged['Cluster Labels'] == 2, KL_merged.columns[[0] + list(range(4, KL_merged.shape[1]))]]
```

	District	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	10th Most Common Venue
1	Titiwangsa	Malay Restaurant	Chinese Restaurant	Asian Restaurant	Café	Motorcycle Shop	Hotel	Thai Restaurant	Indian Restaurant	Coffee Shop	Clothing Store
3	Wangsa Maju	Chinese Restaurant	Asian Restaurant	Malay Restaurant	Thai Restaurant	Ice Cream Shop	Café	Burger Joint	Coffee Shop	Boutique	Convenience Store
6	Segambut	Chinese Restaurant	Korean Restaurant	Malay Restaurant	Convenience Store	Coffee Shop	Asian Restaurant	Japanese Restaurant	Thai Restaurant	Noodle House	Indian Restaurant
9	Bandar Tun Razak	Chinese Restaurant	Asian Restaurant	Malay Restaurant	Indian Restaurant	Convenience Store	Gas Station	Café	Thai Restaurant	Fast Food Restaurant	Indonesian Restaurant
10	Cheras	Chinese Restaurant	Malay Restaurant	Asian Restaurant	Indian Restaurant	Convenience Store	Indonesian Restaurant	Pizza Place	Italian Restaurant	Food Truck	Seafood Restaurant

Fig.12. Popular venues categories of districts labeled under Cluster 2

```
KL_merged.loc[KL_merged['Cluster Labels'] == 3, KL_merged.columns[[0] + list(range(4, KL_merged.shape[1]))]]
```

	District	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	10th Most Common Venue
2	Setiawangsa	Malay Restaurant	Asian Restaurant	Thai Restaurant	Café	Bakery	Middle Eastern Restaurant	Food Truck	Park	Seafood Restaurant	Fast Food Restaurant

Fig.13. Popular venues categories of districts labeled under Cluster 3

5.0 Discussions

It can be seen from Figure 10-13, that there are distinct differences between these clusters. It can be summarized as follows:-

Cluster 0: It is dominated by Chinese and Asian style food outlets

Cluster 1: It has less influence from Malaysian style cuisine and it seems to be more associated with the international community and tourismas there are many hotels, spas, and bars located in this cluster.

Cluster 2: Heavy influence by Malaysian style food outlet that caters the Malaysian population that mainly consists of Malay, Chinese and Indian ethnics.

Cluster 3: This cluster can be associated with the Asian and Muslim population as there are Malay, Asian, and Middle Eastern style restaurants.

As Marble Hill is operating at a venue that that has a western influence, it would be logical to set up their new branch in a district that has exposure to the international environment. Cluster 1 is most likely the best option for their new operation as it is likely to have the presence of international communities and tourists.

6.0 Conclusion

Based on the observation made from these clusters and the top ten venue categories in Marble Hill, the data scientist is inclined to recommend the management of Little Noshery to open their new branch in one of the districts listed in Cluster 1 as it has more exposure to international settings. It will suit them as their current location in Marble Hill is dominated by western cuisine. The presence of hotels, cafés, coffee shops, shopping malls, and spas suggesting that the districts in this cluster are more likely a tourist hot spot. More analyses are required to determine the most suitable location among the three districts listed in Cluster 1. For instance, details analyses related to the rental cost, distance from raw material supplier, demographic of the international community in the district and other factors need to be further investigated.