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

By: Jayaprakash Saththasivam

Problem Statement

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.

Description of Data

The data scientist will be using Foursquare location data to identify the top ten venue categories in Marble Hill. He will be also using Foursquare location data to identify the venues and venue categories in all the districts of Kuala Lumpur. The top ten categories of each district will be also determined by analyzing the raw data. Recommendation on the final location will be made based on the similarities of the venue categories between Marble Hill and Kuala Lumpur.

Data for Marble Hill

The coordinates for Marble Hill are 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 Fousquare location data. The data is then analyzed to obtain the top ten venue categories in Marble Hill (Fig.2)

```
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

```
# one hot encoding
marblehill_onehot = pd.get_dummies(nearby_venues[['Venue Category']], prefix="", prefix_sep="")

marblehill_onehot.head()
marblehill_grouped = marblehill_onehot.mean().reset_index()

temp = marblehill_grouped
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(10))
print('\n')
```

```
venue  freq
0  Mexican Restaurant  0.07
1         Pizza Place  0.07
2         Coffee Shop  0.04
3              Bakery  0.04
4                 Bar  0.04
5                Park  0.04
6                Café  0.04
7               Diner  0.04
8       Deli / Bodega  0.03
9          Restaurant  0.03
```

Fig.3. Top ten venue categories in Marble Hill obtained from Foursquare

Data for districts in Kuala Lumpur

The coordinates for Kuala Lumpur are obtained using geopy library. As for the districts in Kuala Lumpur, there is no available online table that lists all the districts. Hence, I have created a list of districts (Fig.4) in Kuala Lumpur and obtained their respective coordinates from geopy library using a FOR loop (Fig.5). Next, the nearby venues in each district (within 1000 m with a search limit = 100) was obtained using Fousquare location data. The number of venues and venue categories obtained from Foursquare for all the districts in Kuala Lumpur is shown in Fig.6. It can be seen that there are 172 uniques categories across the eleven districts. Further analyses will be presented in the upcoming reports.

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

District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bandar Tun Razak	44	44	44	44	44	44
Batu, Kl	84	84	84	84	84	84
Bukit Bintang	100	100	100	100	100	100
Cheras	62	62	62	62	62	62
Kepong	64	64	64	64	64	64
Lembah Pantai	75	75	75	75	75	75
Segambut	39	39	39	39	39	39
Seputeh	100	100	100	100	100	100
Setiawangsa	71	71	71	71	71	71
Titiwangsa	100	100	100	100	100	100
Wangsa Maju	85	85	85	85	85	85

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

```
print('There are {} uniques categories.'.format(len(KL_venues['Venue Category'].unique())))
```

```
There are 172 uniques categories.
```

Fig.6 Number of venues and venue categories obtained from Foursquare for all the districts in Kuala Lumpur.