Ahmed Fakhr Aldin Alfatih

Strategic Decision making in Location Selection.

Introduction

On the 16th of March 2020, and as a result of the covid 19 case numbers - the Malaysian Government, following suit to what many other sister countries instructed - put to place a Movement Restriction order - restricting the movement of its citizens - allowing for only travel within a specific radius. What followed was a year of Standard operating procedures and new normals - a shift in lifestyle that affected both business and family lifestyles.

With the financial, economical and geographic turmoil of 2020 - the need for depending on closer vicinity businesses has become highlighted more than ever. Identifying the correct place to live and start a business has become very relevant and closely tied to the existence of venues and business within the radius of the actual Physical locations identified.

It became very important to scout out and be informed of the facilities available in a certain area - ensuring that both lifestyle priorities are met.

Problem

For our problem statement - we are tasked with identifying a data driven approach to decision making to help us cope with the new normal. Through the use of Geo-Analysis , we aim to Recommend.

1. The Best Area to Live:
   Selecting the best area to live is often a subjective decision as it must align with the personal priorities of the Homeowner/Renter - but to take a data driven approach to the question - we are aiming to create recommendations that can best fit a specific lifestyle - with respect to the availability of the facilities required for that lifestyle within vicinity. For our Case study - we are interested to find an area suitable for an athletic lifestyle interested in quiet night drinks and diversified cuisine.
2. The Best area to start a Restaurant
   Starting a business during a pandemic is a very risky task - as such and in efforts to reduce the assigned risk - we aim to identify the best area to start a business. Specifically we aim to identify the best area to start an arab/Middle Eastern Restaurant with respect to frequency of venues within radius.

For the Purposes of this report - the Locations Referenced will be from Malaysia Specifically the Kuala Lumpur, Johor baru Areas.

Data

For our objectives - we must first identify the data features we need and how we can get them. To do so our target is to identify the areas we will be analyzing.
The Scope of our analysis is based on Kuala Lumpur and Johor Baru - two States of West Malaysia. These States are then further divided into districts and assigned different postal codes to assist with the national referencing and mail proces. For our Analysis - we have chosen to identify the Districts within KL and JB areas as the Larger scale for our analysis. We also require the latitude and Longitude associated with each area / district to identify the venues around it.

○ Sources

For our analysis - we use 3 data sources to complete 4 data migration operations. The initial operation is retracting from a Github Repository to get the Kuala Lumpur and Johor Bahru postal codes and areas.

The Second data file associated is downloaded from the A postal code center to retrieve the latitude and longitude lines associated with each postal code. This is a list of all postal codes in Malaysia, along with additional information such as the state, county, and geographic coordinates. Source from Geonames.org

#https://www.aggdata.com/free/malaysia-postal-codes

○ Exploring and Cleaning the Data Sets

The two sets retrieved from github have the format of {Postal Code , District, Area) . One holding the Districts of Kuala Lumpur and the other the Districts of Johor Bahru.

	Postcode	District	Area
0	80000	Johor Bahru	Johor Bahru
1	81100	Johor Bahru	Bandar Dato' Onn
2	80200	Johor Bahru	Danga Bay
3	81100	Johor Bahru	Johor Jaya
4	81100	Johor Bahru	Desa Jaya

The Kuala Lumpur DataSet holds information about 11 Districts and 66 areas while the johor bahru has 3 districts and 22 areas. The Datasets Contain No duplicates and are Clean of formatting Errors.

Second Data set holds the information of the postal code as created by the malaysian Government - Data set has to be downloaded from an external file and stored into a csv. The features of the data set are {Postal Code, Place Name, State Abbreviation, Latitude, Longitude } - the data set is clean and contains no duplicates. The Data set holds 433 postal codes covering 17 states.

Having identified the first two data sets - we can combine then with the postal code dataset using Join - to ensure only the postal codes that match are merged. We arrive at 2 data sets that hold the information for the KL and JB states and their respective latitude longitude points associated with each postal code. We clean that combined data set to arrive at the features of {Postcode, District, Area, Latitude, Longitude}.

Methodology

To create Analyze our problem and complete our solution we operated Python driven libraries to assist in retrieving, storing, analyzing and visualizing the data associated with this project. The libraries were:

```
]: #identifying the Correlating Libraries

import pandas as pd, numpy as np #Libraries Used to Store, manipulated data
import requests #Used to Pull information from the internte
import json #Used to access and interpret location data
import time
import folium   #for creating maps
from pandas.io.json import json_normalize #converting json to DataFrame
from geopy.geocoders import Nominatim  #converting address to cordinates
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans
```

Using Said libraries we attempted the following methods.

- Retrieving Venues
  Having Prepared the list of Areas and postal codes with their latitude and longitude assigned, we can Use the Foursquare API to retrieve the venues within vicinity of each postal codes. We do so by using a function getNearbyVenues that retrieves the venues within 500 m of each areas postal code . the venues are returned come in the form of a dataframe holding {Area, Area Latitude, Area Longitude, Venue, Venue Latitude, Venue Longitude, Venue Category.}
- Encoding
  Having retrieved the Venues for both states - we encode the venues from string to numeric to allow us to best compare their distribution, This encoding allows us

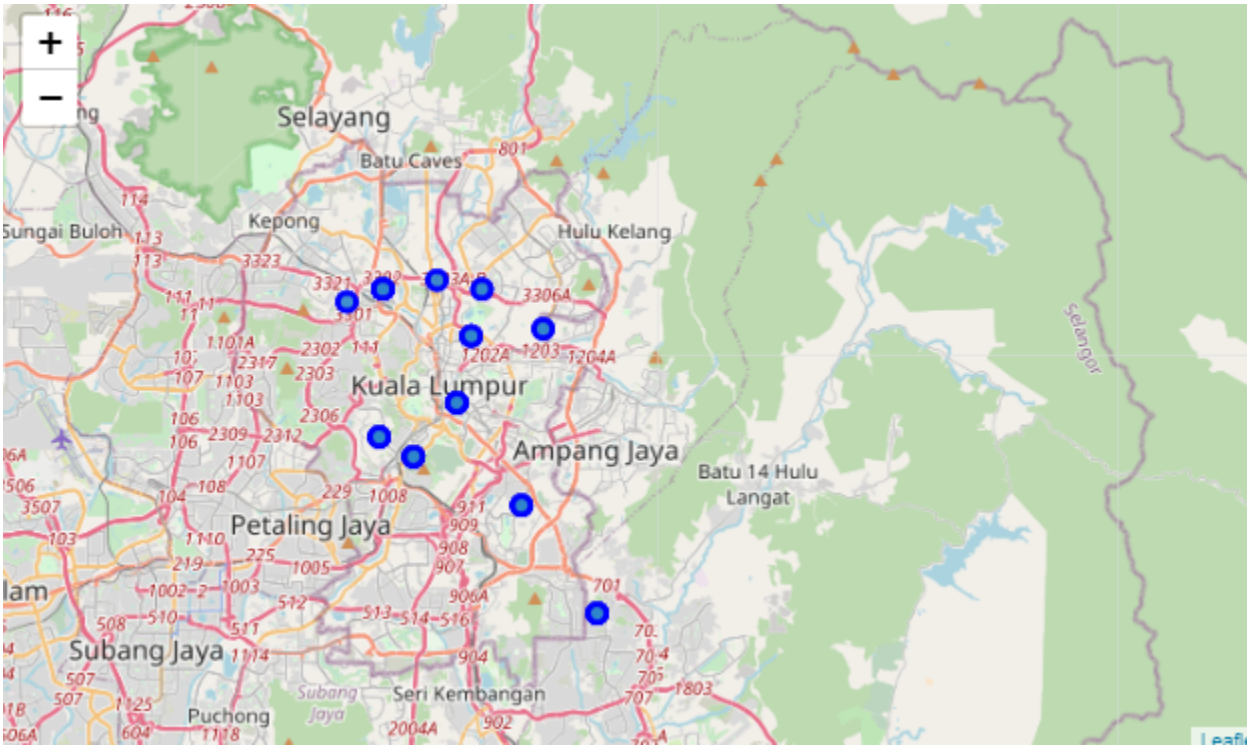
to compare all venues over their categories and numerically then distinquince the frequency of each venu category for each area. This allows us to Tabulate and compare the top 5 categories for each area.

- Visualization
  Having Tabulated the Frequency of each category to each area - we then Create our final intended columns to establish a list of the most repeated venue categories for each area. This list allows us to primarily compare areas with respect to both objectives.
  - Comparing the Best Area to Live
    Having a list tabulating the most popular venus - allows us to compare areas based on how much their venues match the lifestyle of the required user - allowing us to compare and recommend based on the different user priorities.
  - Selecting the Best Place to Start a Restaurant (Arab Middle easter)
    The List of Venues will allow our user to identify which areas are most populated by what categories - clarifying a list of the best locations to establish our Arabic Restaurant.
- Clustering
  After Tabulating our lists - we K means cluster the venues listed in either areas to identify at different categories of venues. These Clusterings allow us to group the venues for easier display to Users - with respect to having common categories or similar degree of Commonality between all the venues - giving the users different layers to study.

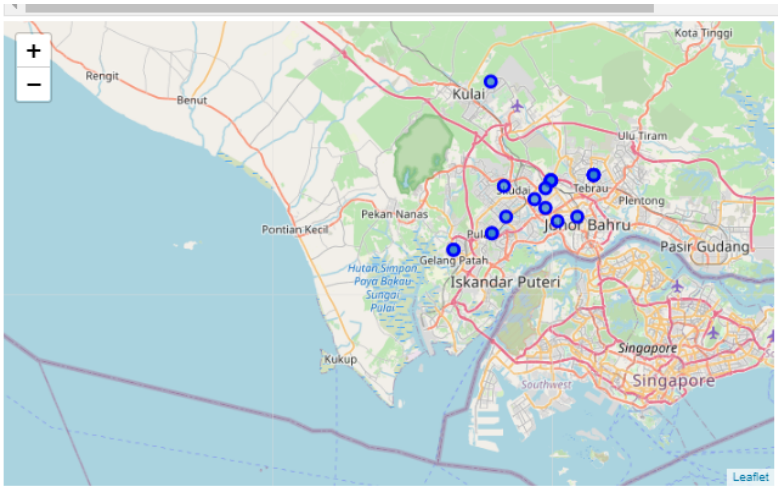
Results

Processing through the data sets we found that the KL state has 11 districts divided and 66 areas - whereas Johor Baru has 3 districts with 11 .

a. The KL Map



B. The JB Map



When we matches the Area with the District and Called the Near By venues - we retrieved 5002 venues for KL spanning 164 different categories - where as when observing JB we only retrieved 709 venues within 103 unique categories.
After encoding the Data - we arrive at the frequency distribution points of the areas - allowing us to identify the most frequent category type for each area :

```
----Alam Damai----
                  venue  freq
0     Chinese Restaurant  0.21
1      Convenience Store  0.11
2       Asian Restaurant  0.11
3  Vietnamese Restaurant  0.11
4           Soccer Field  0.05


----Ampang----
              venue  freq
0  Malay Restaurant  0.20
1  Asian Restaurant  0.10
2             Café  0.08
3      Soccer Field  0.08
4        Restaurant  0.05
```

(Example)

Having Identified the frequency of categories in each area - we are able to showcase the most repeated venues for each category on each area - allowing us to make recommendations based on the area to which we are discussing.
For the KL area we identified :

Area	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
Alam Damai	Chinese Restaurant	Vietnamese Restaurant	Asian Restaurant	Convenience Store	Steakhouse	Playground	Thai Restaurant	Athletics & Sports	BBQ Joint	Food Truck
Ampang	Malay Restaurant	Asian Restaurant	Café	Soccer Field	Thai Restaurant	Restaurant	Food Court	Coffee Shop	Pizza Place	Chinese Restaurant
Bandar Baru Sentul	Malay Restaurant	Asian Restaurant	Breakfast Spot	Thai Restaurant	Food Truck	Gym	Residential Building (Apartment / Condo)	Noodle House	Night Market	Indian Restaurant
Bandar Malaysia	Indian Restaurant	Hotel	Cosmetics Shop	Gym	Park	Scenic Lookout	Seafood Restaurant	Café	Food Court	Food Truck
Bandar Menjalara	Chinese Restaurant	Asian Restaurant	Indian Restaurant	Gym	Convenience Store	Business Service	Burger Joint	Gas Station	Halal Restaurant	Gym Pool

For the JB District we Identified:

	Area	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	Tanjung Langsat	Fried Chicken Joint	Chinese Restaurant	BBQ Joint	Motorcycle Shop	Food Court	Flea Market	Fireworks Store	Fast Food Restaurant	Electronics Store	Dessert Shop
1	Bandar Dato' Onn	Video Game Store	Furniture / Home Store	Fireworks Store	Gym	Malay Restaurant	Warehouse Store	Food Court	Fast Food Restaurant	Electronics Store	Dessert Shop
2	Bandar Seri Alam	Fried Chicken Joint	Chinese Restaurant	BBQ Joint	Motorcycle Shop	Food Court	Flea Market	Fireworks Store	Fast Food Restaurant	Electronics Store	Dessert Shop
3	Bukit Indah	Golf Course	Fast Food Restaurant	Brewery	Warehouse Store	Chinese Restaurant	Food Court	Flea Market	Fireworks Store	Electronics Store	Dessert Shop
4	Danga Bay	Malay Restaurant	Food Truck	Café	Food Court	Rest Area	Gaming Cafe	Grocery Store	Burger Joint	Pharmacy	Flea Market

This tables showcase the head of Files stored Containing a list of the most common venues for each area - dedicated for Both Kuala Lumpur and Johor Bahru areas.

Conclusion

- As per our Case Study Selected -
  A. Place to live :
  Under the lifestyle Recommended in our Problem statement - we start by initially selecting KL as it contains more venues per area - allowing for larger number of restaurants to explore. Along that route - within KL we recommend :

  1. Segambut 51200 Postal Code with its areas of : The Existence of a gym and diversified cuisine options—and a variety of cafes for the nights.

  B . Best Place to Start an Arab/Middle Eastern Restaurant
  Evaluating where to start the business follows a different set of priorities in that the research is done to avoid competition. As such - our initial recommendation would start with focusing the restaurant plan towards the Johor Bahru State as it has less Venues listed overall. Furthermore - the category of middle eastern restaurant is not listed in the JB area - so we don't have to worry about competition within the category. Alternatively - to select the right area - we choose one with less international cuisine competition. As such we recommend:

10	Johor Jaya	Malay Restaurant	Seafood Restaurant	Café	Convenience Store	Fast Food Restaurant	Hotel	Hookah Bar	Halal Restaurant	Hakka Restaurant

- Assumptions

Our research does not account for updated foursquare data .
As part of our assumptions for this project - we noticed that all areas with in the same district are noted to have the same latitude longitude values - although the venues and area names listed in foursquare show Greater details . Further more - within the categories - there exists a lot of single use categories skewing the biased - as such - any category containing less than the frequency distribution of 0.05 is omitted.