



ANALYZING THE BEST BUSINESS DESTINATION

Kuala Lumpur or Dubai

Coursera IBM Data Science Professional Certification

Final Project Report
Arunjith M
arunjithece@gmail.com

Contents

Introduction	3
Intended Audience	3
Business Problem	4
Data Sources and Requirements	5
Methodology	6
Results	10
Discussion	11
Conclusion	12

Introduction

AmerCave Group is a business entity who has its wings across the United States. It has a chain of shopping malls and restaurants in almost all major cities of different states in America. As part of expanding business, they are strategically stepping outside the US. **AmerCave** have chosen East Asia and Middle East for deploying their first shopping mall and an American restaurant.

The reason for selecting these regions are given below.

A. East Asia and Middle East are two fast growing economies, both having outstanding and competitive regional hubs for starting a business. Middle east, especially Dubai, UAE is an extremely attractive business destination and a great tourism spot. Lot of investments pouring in for the oil sector makes Dubai an advantage for investors. Apart from that, it is the gateway for eastern markets as it acts as a hub for imports and exports. The city attracts a lot of expat investors as well.

B. Kuala Lumpur, Malaysia in the East Asian region have an increasing consumer demand from various nearby countries like India, China etc. It is set to be the center of digital economy and have a multitude of business forms. A diversified workforce is another advantage for opening a business in Malaysia. Kuala Lumpur offers a very competitive market clubbed with greater growth opportunities.

Intended Audience

- a. Any business entity that have a business acumen and ready to start a business either in Middle East or in East Asia
- b. Anybody who is interested in Data Science, who wants to scrap data and analyze those to gain insights & extract meaningful information

Business Problem

As far as **AmerCave** is concerned, they are their own masters and is well established, have the business acumen & glorifies every business opportunity that comes in their way. The difference this time is, that they are stepping out of the United States of America and wanting to have their footprint in Middle East & East Asia.

After much analysis they have shortlisted Dubai & Malaysia as their business destinations.

Strategically armed **AmerCave Group** is now looking for detailed analysis of where to start which business.

The answers to get solved are

1. Where to start the American Restaurant? Dubai or in Kuala Lumpur?
2. Where to start the shopping mall? Dubai or in Kuala Lumpur?

Data Sources and Requirements

To answer the business problems posed by **AmerCave Group**, a couple of inputs are required. It is required to analyze the number of shopping malls and American restaurants in and around the vicinities of both cities - Dubai & Kuala Lumpur

Another requirement brought upfront is that the both business should be opened within 2 Kilometers of the main cities. So, the real requirements can be summarized into four parts below, from which the information on what data to be collected can be finalized.

1. Provide a list of best locations in Kuala Lumpur OR Dubai for starting up the shopping mall
2. Provide a list of best locations in Kuala Lumpur or Dubai for starting the American restaurant
3. Finalize from the list, the best location for starting the shopping mall and the American restaurant
4. All these should be finalized within 2 kilometers from the main city

The Sources

Below is the list of all sources that is used in this project

1. Dubai neighborhoods information which can be web scrapped from Wikipedia page -- https://en.wikipedia.org/wiki/List_of_communities_in_Dubai
2. Kuala Lumpur neighborhood information which can be web scrapped from Wiki page -- https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur
3. Geo-data (Latitude & Longitude information) of Dubai and Kuala Lumpur with top venue information to be collected from Foursquare

Methodology

This part of this report details about the operations that has been performed for arriving at the right results.

Before moving into any type of data analysis, it started with importing the right libraries and package which is required for every single operation within.

```
!pip install geopy
!pip install folium
!pip install geocoder

import pandas as pd #for data analysis
import numpy as np #for handling data
import requests
import folium # map rendering library
import geocoder # to get coordinates
import matplotlib.pyplot as plt # Matplotlib and associated plotting modules
import matplotlib.cm as cm # Matplotlib and associated plotting modules
import matplotlib.colors as colors # Matplotlib and associated plotting modules
import json # Library to handle JSON files

from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe
from bs4 import BeautifulSoup
from geopy.geocoders import Nominatim # convert an address into Latitude and Longitude values
from sklearn.cluster import KMeans # import k-means from clustering stage

def warn(*args, **kwargs):
    pass
import warnings
warnings.warn = warn

print("All required packages installed and imported")
```

Once the necessary packages have been installed, the required data for analysis needs to be collected. As state in the data section, the places information is publicly available in the respective Wikipedia pages.

BeautifulSoup package has been used for parsing the html files for both places. The information pertaining to Kuala Lumpur cities is available in the mw-category tag field in the html output and that of Dubai is available in the mw-parser-output tag field.

The first five rows of the resulting dataframe is shown in the next page. The dataframes listed for storing Kuala Lumpur & Dubai cities information is kl_df and dxb_df respectively. Also it should be noted that the first 9 tags can be removed from Dubai city information, as it has nothing to do with the actual name of places.

	KL_Neighborhood		DXB_Neighborhood
0	Alam Damai	0	Abu Hail
1	Ampang, Kuala Lumpur	1	Al Baraha
2	Bandar Menjalara	2	Al Buteen
3	Bandar Sri Permaisuri	3	Al Dhagaya
4	Bandar Tasik Selatan	4	Al Garhoud

Once the places name have been successfully derived, the next step would be to get the latitude and longitude information pertaining to each and every places. This is extracted using the Geocoder package.

The following piece of code is just for Kuala Lumpur. A similar one is done for Dubai as well.

```
#define function for getting KL coordinates
def get_latlng_kl(neighborhood):
    # initialize your variable to None
    lat_lng_coors = None
    # loop until you get the coordinates
    while(lat_lng_coors is None):
        g = geocoder.arcgis('{}', Kuala Lumpur, Malaysia'.format(neighborhood))
        lat_lng_coors = g.latlng
    return lat_lng_coors

#Creating a list of co-ordinates using the list comprehension method
kl_coors = [ get_latlng_kl(neighborhood) for neighborhood in kl_df["KL_Neighborhood"].tolist() ]

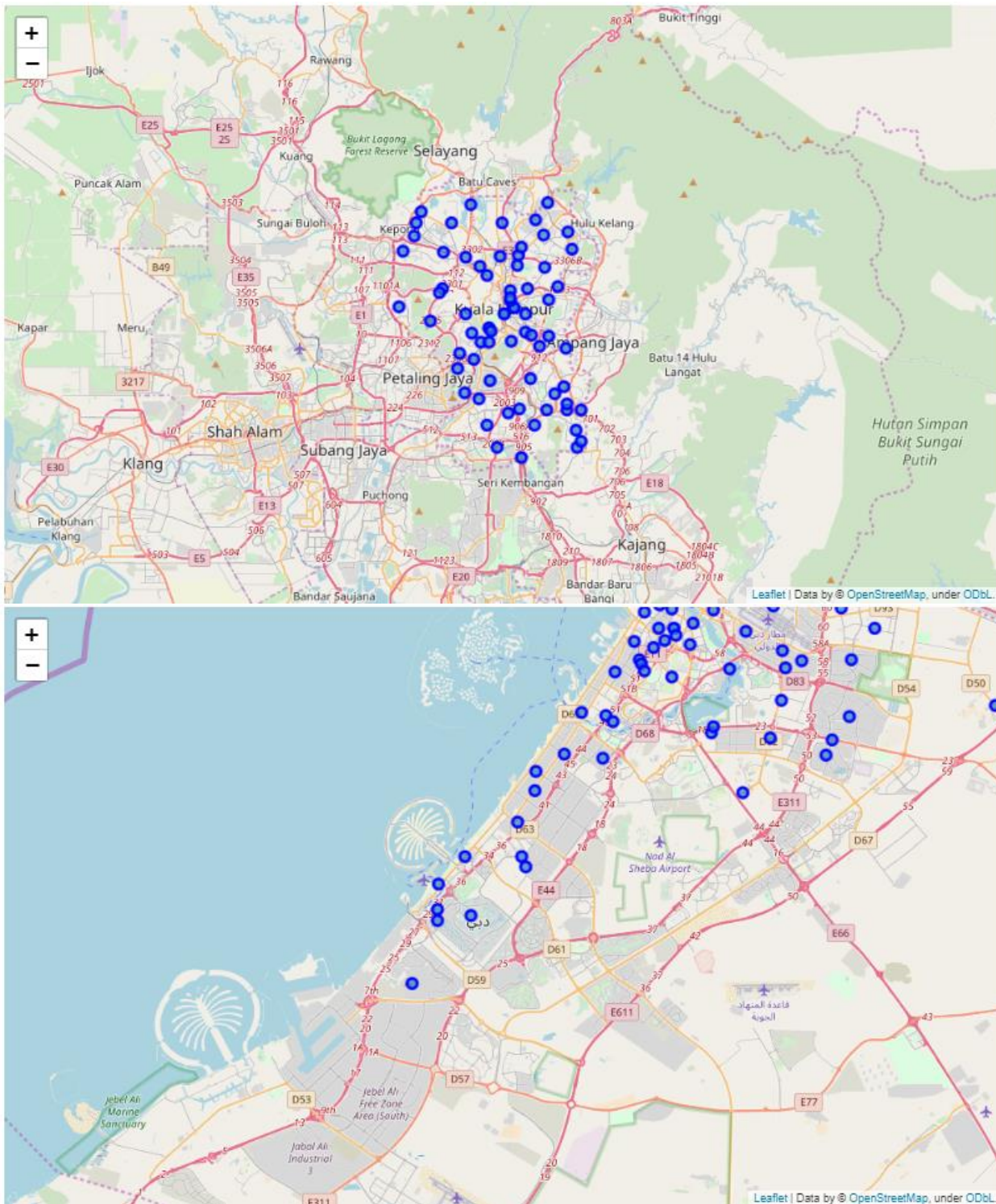
# define a function to get Dubai coordinates
def get_latlng_dxb(neighborhood):
    # initialize your variable to None
    lat_lng_coors = None
    # loop until you get the coordinates
    while(lat_lng_coors is None):
        g = geocoder.arcgis('{}', Dubai, UAE'.format(neighborhood))
        lat_lng_coors = g.latlng
    return lat_lng_coors

#Creating a list of co-ordinates using the list comprehension method
dxb_coors = [ get_latlng_dxb(neighborhood) for neighborhood in dxb_df["DXB_Neighborhood"].tolist() ]
```

The results have been embedded to actual data frame and looks like the one shown below here

	KL_Neighborhood	Latitude	Longitude		DXB_Neighborhood	Latitude	Longitude
0	Alam Damai	3.057690	101.743880	0	Abu Hail	25.28308	55.33435
1	Ampang, Kuala Lumpur	3.153153	101.700413	1	Al Baraha	25.28280	55.31678
2	Bandar Menjalara	3.190350	101.625450	2	Al Buteen	25.26925	55.29944
3	Bandar Sri Permaisuri	3.103910	101.712260	3	Al Dhagaya	25.27217	55.30157
4	Bandar Tasik Selatan	3.072620	101.714710	4	Al Garhoud	25.24337	55.35267

As there is enough input for Folium package for plotting maps, both have been done for visualizing purpose. Screenshots of both are given below.



Now comes the role of Foursquare API for extracting the venue details of both cities. And it resulted in the following.

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Alam Damai	3.05769	101.74388	Pengedar Shaklee Kuala Lumpur	3.061235	101.740696	Supplement Shop
1	Alam Damai	3.05769	101.74388	Machi Noodle 妈子面	3.057695	101.746635	Noodle House
2	Alam Damai	3.05769	101.74388	Restoran Ikbal	3.061134	101.750220	Restaurant
3	Alam Damai	3.05769	101.74388	閒茶素食店 Leisure Tea Vegetarian	3.057673	101.747258	Vegetarian / Vegan Restaurant
4	Alam Damai	3.05769	101.74388	Ivy Sekinchan Seafood Noodle House 道耕莊特刺魚丸海鮮面	3.065749	101.748718	Noodle House

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Abu Hail	25.28308	55.33435	Habib Bakery	25.281124	55.332774	Bakery
1	Abu Hail	25.28308	55.33435	Gold's Gym	25.282698	55.341019	Gym
2	Abu Hail	25.28308	55.33435	Al Douri Roastery	25.277057	55.328223	Bakery
3	Abu Hail	25.28308	55.33435	Union Co-Operative Society	25.282769	55.340896	Department Store
4	Abu Hail	25.28308	55.33435	Fitness Time (وقت اللياقة)	25.289077	55.347913	Gym

Now, a more detailed and closer look into both city venues are required. Using one-hot encoding and KMeans clustering, further analyses were performed. The total number of clusters used was 3 and it resulted in 12 different values. The value inside each cluster was summed up in order get the density information of each business entities in each city. More details of how this is performed are very much detailed in the code section. Visit this link for more into coding part and to understand how it is performed.

https://github.com/arunjithm/Coursera_Capstone/blob/master/Coursera_Capstone_Project_Final.ipynb

Results and inferences drawn out of the results is detailed in the Results & Discussion that follows in next page.

Results

As mentioned, the three cluster analysis performed contributed to 12 different values which indicates the densities of the two business entities (shopping malls and American restaurants) in both cities. This is given in the table below.

	Cluster 0	Cluster 1	Cluster 2
KL_ShoppingMall	0.38	0.51	0
KL_AmericanRestaurant	0	0.04	0
DXB_ShoppingMall	0.27	0.42	0.14
DXB_AmericanRestaurant	0	0.37	0.14

Note that this is obtained by summing up the corresponding values in each clusters. This is a harmless approach as it directly indicates the presence of such business in every clusters. The greater the value, the greater the presence of corresponding business figure there. The information drawn out of this analysis is presented in the discussion part in next page.

Discussion

We need to compare the number of neighborhoods in every cluster of Dubai and Kuala Lumpur to draw insights from the above data.

From the resultant data frame it is pretty much clear that there is a high density of shopping malls in the locations in Cluster 0 and Cluster 1 of Kuala Lumpur. And there is very less density of shopping malls in locations of Cluster 2 in Kuala Lumpur.

Also analysing the values of shopping malls in Dubai, all the three clusters have sufficiently good amount of shopping malls.

A similar analysis can be done for American Restaurant. From the data frame it is pretty much clear that there is very less density of American restaurants in locations of various clusters in Kuala Lumpur. Also, from the values it is clear that Kuala Lumpur is not a potential market for American restaurants.

While on the other hand, there is sufficiently greater number of American Restaurants in locations of Cluster 1 and Cluster 2 of Dubai. And there is negligible amount of American restaurants in Dubai. The values of those in Cluster 1 & 2 reveal that Dubai is a potential market for American restaurants.

Scope Extension

The project can be extended by analyzing the number of people residing in each location, the number of nationalities residing in each location, the economic conditions of both cities can be taken into comparison & analysis, the development index can be considered etc. This analysis will help bring out a better figure and pinpoint the location where the business can be launched off.

Conclusion

It is not advised to start a new shopping mall in the vicinity of locations in either of the locations in Dubai clusters, as from the resultant table it is pretty obvious that there is a huge concentration of shopping malls in Dubai. It won't be profitable to start afresh in a competitive and an established market.

Comparatively, there is much diminished presence of shopping malls in Cluster 2 locations of Kuala Lumpur. Therefore, this project is advising **AmerCave Group** to start a new shopping mall in any of the locations of Cluster 2 in Kuala Lumpur, as there will be very less competition and have a positive growth index.

The American restaurants have very low presence in either of the clusters in Kuala Lumpur. This means a afresh American restaurant business would not be profitable for **AmerCave**. There is sufficiently good amount of such restaurants in the Cluster 1 and Cluster 2 locations in Dubai.

Therefore, this project is advising **AmerCave Group** to start a new American restaurant in any of the locations of Cluster 0 in Dubai, as there will be very less competition and there is a potential for growth when comparing the values with other two clusters.

Thanks for going through this report.

Arunjith M