

Finding Suitable Condominium with Data Science Techniques

1. **Introduction:** This project has been initiated as the final project requirement for Coursera course IBM Data Science Professional Certificate. The aim of this project is to demonstrate how data analytics, machine learning and data visualization techniques can be used to solve real life problem.

1.1 **Business Problem:** Recently one of my old alumni has been offered a job in Kuala Lumpur. With the decision to move in, he called me and asked for my help to suggest him some suitable options for him to stay with his family. I tried to understand what my friend means by “some suitable options”, basically I did a requirement analysis with few more questions. The outcome was a condominium location -

- i) Within 5 KM distance of his new office address, he loves to drive.
- ii) Nearby international school. Within 3 KM distance of the condominium, for his 12 years old kid.
- iii) Nearby supermarket or reputed grocery shop. Within 2 KM distance of the condominium.

I asked for 1-week time and decided to apply my data science and geo coding techniques to conduct a thorough analysis on condominiums available in Kuala Lumpur to find the best options according to my friend's requirements.

1.2 **Audience:** Beside my friend, many people can get benefited from the result of this project who basically need to move to a new location and have similar requirements.

2 Data:

2.1 **Source of Data:** I have decided to use Foursquare data on the locations in Kuala Lumpur as Foursquare is the most trusted, independent location data platform for understanding how people move through the real world.

2.2 **Preparation of Data:** Scrapping venue related data usually comes with lots of attributes which all are not necessary for this project. We need to carefully check the keys and apply filter based on our requirements to save resources and time on processing.

2.3 **Uses of data:** With the location data downloaded from Foursquare, I am targeting to perform below steps of analysis to reach the targeted result.

- a) Generate the list of condominiums in Kuala Lumpur.
- b) Narrow down the number of condominiums within 5000 meters radius of the address provided.
- c) Cluster the condominiums with a school within 3000 meters radius.
- d) Generate a list of schools from 'c' and sort according to distance from condominiums.
- e) Repeat 'c' and 'd' for supermarkets.
- f) Combine the result of 'd' and 'e' to get the list of condominiums that meet the requirements.

3 **Methodology:** We will use Python 3 in our entire journey of data analysis as Python has all the required libraries for this work.

3.1 **Define Algorithm:** At this stage, we plan the step by step procedure to reach our expected result of this project with available data.

- i) Determine the necessary libraries: We need common python libraries such as **pasdas**, **numpy** and **random** as we will be working with data frame, arrays and numeric values. We will use **Nominatim** to convert addresses into latitude and longitude values. **Json_normalize** library will be required to convert json files into data frames. We will also use **Folium** library to plot maps. Lastly, we will use **haversine** library to find actual distance between two geo address.
- ii) With **Nominatim**, we will find the geo coordinates (latitude and longitude) of the given address.
- iii) Communicate with Foursquare API to get list of schools within 8 (5+3) KM range of the geo coordinates for given address.
- iv) Communicate with Foursquare API to get list of condominiums within 5 KM range of the geo coordinates for given address.
- v) Visualize the condominiums on the map with marker and with a circle showing 5 KM range from the center as given address.
- vi) Create a data table of condominiums and international schools from data sets available from steps **iii)** and **iv)** that match first 2 requirements.
- vii) Finally, we will check if the condominiums listed in table of step **vi)** meet the 3rd requirement.

3.1.2 **Execution:** With the strategy ready, we will use Jupiter notebook for coding and testing the result.

We have found 30 international schools within 8 KM radius of the given office address (*348 Jalan Tun Razak, Kuala Lumpur*).

	id	name	categories	referralId	hasPerk	location.address	location.lat	location.lng	location.labeledLatLngs	location.distance	to
0	4bf0c88c6f8aa5938249c23c	Sayfol International School	[[{"id": "4bf58dd8d48988d13d941735", "name": "H..."}]]	1624284617	V-	False	No. 261, Jalan Ampang	3.159513	101.730375	[[{"label": "display", "lat": 3.159513368061476...}]]	1127
1	53507b6e498ecfbc2055ee10	Children DC International School	[[{"id": "4bf58dd8d48988d13d941735", "name": "H..."}]]	1624284617	V-	False	NaN	3.159316	101.719661	[[{"label": "display", "lat": 3.159316, "lng": ...}]]	200
2	52ddfe5811d2f77b311cdc2	Children Discovery International School	[[{"id": "4bf58dd8d48988d13d941735", "name": "H..."}]]	1624284617	V-	False	NaN	3.163110	101.720704	[[{"label": "display", "lat": 3.16311, "lng": 1...}]]	239
3	4d40ab3090f9224ba387081e	Fairview International School , Ampang	[[{"id": "4bf58dd8d48988d13d941735", "name": "H..."}]]	1624284617	V-	False	Jl.Ampang	3.157351	101.723390	[[{"label": "display", "lat": 3.157351, "lng": ...}]]	527
4	4dddaee6c65bb2aa738411be	SUB International School Of Beauty & Makeup	[[{"id": "4bf58dd8d48988d13d941735", "name": "T..."}]]	1624284617	V-	False	129-131,1st Floor ,Jalan Imbi	3.142694	101.710518	[[{"label": "display", "lat": 3.142694208237098...}]]	2310
5	56fe243e498eca2ef93b55ec	The International School of Kuala Lumpur (Ampa...	[[{"id": "52e81612bc57f1066b7a46", "name": "P..."}]]	1624284617	V-	False	Jalan Ampang Hillr	3.155995	101.740848	[[{"label": "display", "lat": 3.155994624198063...}]]	2346

Figure 1: top 6 of the 30 international schools

In the next phase We have found 50 international schools within 5 KM radius of the given office address.

	id	name	categories	referralId	hasPerk	location.address	location.crossStreet	location.lat	location.lng	location.labeledLatLngs	
0	4c6b75ca20fe3704a54154f8	Condo Chinta	'{id': '4d954b06a243a5684965b473', 'name': 'R...	1624284901	v-	False	CHINTA	JLN. UTHANT	3.156203	101.726743	[{'label': 'display', 'lat': 3.156203179674576...
1	4ddcdf8322713339cd1eec23	Titivangsa Sentral Condominium	'{id': '4d954b06a243a5684965b473', 'name': 'R...	1624284901	v-	False	Jalan Cemur, Kompleks Damai	NaN	3.171410	101.696432	[{'label': 'display', 'lat': 3.171409957975216...
2	4bda84e0c79cc928cc147fe9	Avare Condo	'{id': '4d954b06a243a5684965b473', 'name': 'R...	1624284901	v-	False	2 Lorong Kuda	Jalan Tun Razak	3.155632	101.720271	[{'label': 'display', 'lat': 3.155631893287337...
3	4b83e51cf964a520971531e3	Seri Maya Condominium	'{id': '4d954b06a243a5684965b473', 'name': 'R...	1624284901	v-	False	Jalan Jelatek, off Jalan Ampang	NaN	3.175808	101.734131	[{'label': 'display', 'lat': 3.175808165630073...
4	4b5bb942f964a520551229e3	The Palladium Condominium	'{id': '4bf58dd8d48988d130941735', 'name': 'B...	1624284901	v-	False	Jalan Gurney 2,	Jalan Gurney 2, Kampung Datok Keramat	3.167132	101.717500	[{'label': 'display', 'lat': 3.167132097169949...
5	4d637bb4a5be236a5232c2db	Panorama Condominium	'{id': '4bf58dd8d48988d130941735', 'name': 'B...	1624284901	v-	False	No 2, Persiaran Hampshire	NaN	3.160965	101.714794	[{'label': 'display', 'lat': 3.160964620643164...
6	4cfe2564ee9cb60ca7169ead	Kedai Makan Bistari Condo	'{id': '4bf58dd8d48988d156941735', 'name': 'M...	1624284901	v-	False	Bistari Condominium, North Bld, 2nd Flr, Jalan...	NaN	3.167995	101.693892	[{'label': 'display', 'lat': 3.167994851777907...

Figure 2: top 7 of the 50 condominiums

The figure below shows the office location with 5 KM radius and the position of the condominium with blue center with orange border circle.

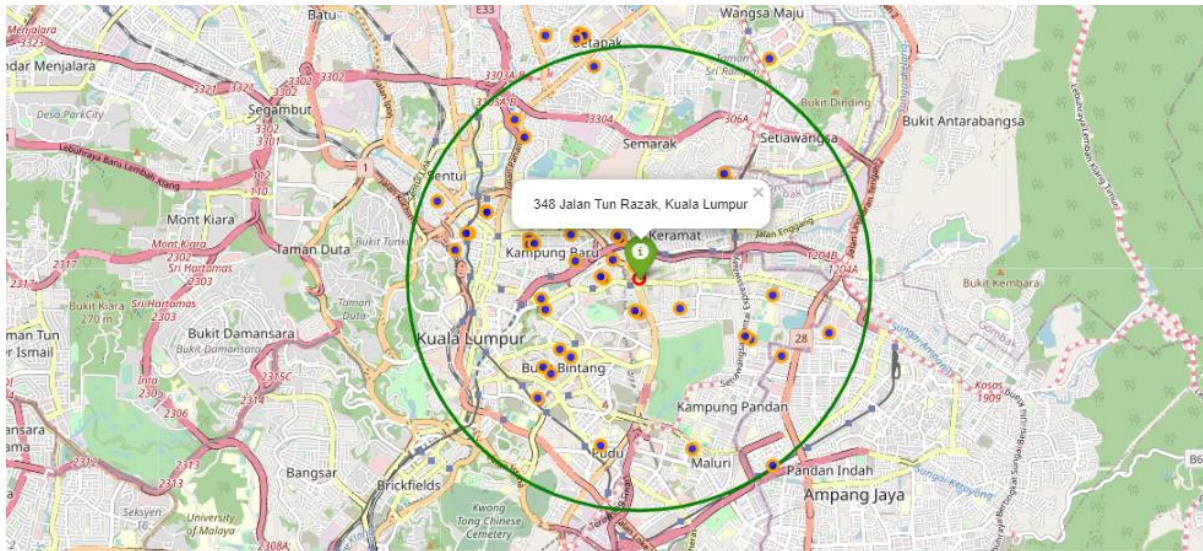


Figure 3: Map with office address and condominium within 5 KM radius

Data analysis has given us 718 results of condominium and school combinations that meet the first 2 requirements.

	Condo_ID	Condo_Name	Condo_lat	Condo_lon	School_lat	School_lon	Distance(Meter)	School_Name
469	5579bdfa498e427237000f6d	Vista Damai Condominium	3.162958	101.719687	3.163110	101.720704	114.137813	Children Discovery International School
197	4cfe2564ee9cb60ca7169ead	Kedai Makan Bistari Condo	3.167995	101.693892	3.167098	101.692283	204.612703	UKL International School of Entrepreneurship
666	4d7e1455f635236af2fc6016	Villa Putra Condominium	3.165287	101.691711	3.167098	101.692283	211.097194	UKL International School of Entrepreneurship
617	578b4cca498edc44cdd1928f	Condo Tingkat 7 IPB	3.134568	101.728676	3.136862	101.731109	371.482976	Library Taylor's International School KL (Perp...
119	4d85b59b48ee6ea8ccf75d68	Desa Kudalari Condo	3.155916	101.719723	3.159316	101.719661	378.107373	Children DC International School
...
282	5072889be4b09145cdcd80f2	library Bestari Condo	3.120597	101.735141	3.155995	101.740848	3986.721479	The International School of Kuala Lumpur (Ampa...
22	4ddcdf8322713339cd1eec23	Titivangsa Sentral Condominium	3.171410	101.696432	3.159513	101.730375	3994.033534	Sayfol International School
693	545993a0498e6ea3efce8d54	K155 Serimaya Condo	3.174470	101.734007	3.159338	101.701373	3994.842235	UnikL International School of Entrepreneurship...
466	4d0e20fc3905b1f7651c9e4b	PV12 Platinum Lake Condominium	3.206494	101.717895	3.195683	101.752228	3996.763816	Fairview International School
279	4d035fe0c2e53704cd3fb167	Rez Cafe D5 Gurney Height Condo	3.170009	101.728191	3.167098	101.692283	3999.760868	UKL International School of Entrepreneurship

718 rows x 8 columns

Figure 4: Condominiums, Schools, and the distance between them

Next, we have checked for first 5 condominiums in our list if they have at least 1 supermarket within 2 KM radius. All 5 was successful. Below figure is the result of first query.

	id	name	categories	referralId	hasPerk	location.address	location.lat	location.lng	location.labeledLatLngs	location.distance	lo
0	4ba0902cf964a520687237e3	Hock Choon Supermarket	'52f2ab2ebcbc57f1066b8b46', 'name': 'S...	1624279298	V-	False	241-1, Jalan Ampang	3.160428	101.728618	[[{"label": "display", "lat": 3.160428163170332...}	514
1	4c28a1f7ed0ac9b6627760aa	Giant Supermarket	'52f2ab2ebcbc57f1066b8b46', 'name': 'S...	1624279298	V-	False	Sungei Wang Plaza	3.143862	101.711091	[[{"label": "display", "lat": 3.143861950401681...}	2216
2	4caec256562d224b04c50888	Pink Evil's Fashion Supermarket	'4bf58dd8d48988d104951735', 'name': 'B...	1624279298	V-	False	Pavilion Kuala Lumpur	3.149258	101.713271	[[{"label": "display", "lat": 3.14925764955174...}	1685
3	5c04fde0a08ab002cd0de7d	Big Supermarket	'52f2ab2ebcbc57f1066b8b46', 'name': 'S...	1624279298	V-	False	Jalan Tun Razak	3.159670	101.720675	[[{"label": "display", "lat": 3.15967, "lng": 1...	777
4	5813150738fa3d570ebc6198	Supermarket @The Forum	'4bf58dd8d48988d118951735', 'name': 'G...	1624279298	V-	False	NaN	3.144235	101.721295	[[{"label": "display", "lat": 3.144235, "lng": ...}	1463
5	4d0c32675c46a0936d4007b4	T.C Permata Supermarket	'4bf58dd8d48988d118951735', 'name': 'G...	1624279298	V-	False	Ampang Park	3.162959	101.727228	[[{"label": "display", "lat": 3.162958713643700...}	753
6	4d26a9708292236a18be08bb	ori supermarket ampang park	'52f2ab2ebcbc57f1066b8b46', 'name': 'S...	1624279298	V-	False	NaN	3.160492	101.720049	[[{"label": "display", "lat": 3.160492392114407...}	884

Figure 5: Supermarkets within 2 KM radius of the first condominium.

4 **Result:** As the result of the above analysis we have got 50 condominium names and addresses that meets the requirements of my friend. Also, he has 25 options as international schools within the required 3 KM distance of the 50 condominiums.

5 **Discussion:** First thing comes out of the analysis is my friend does not need to worry about the school for his kid and grocery shopping as there are plenty of options. Next he can look for other amenities such as gym, swimming pool, playground, park, medical facilities and the living cost to narrow down the options to find the best one for him.

One observation is that rating information for international schools are not available in Foursquare dataset. I would like to suggest them to include this one as many parents may look for ratings in the selecting suitable one for their children

6 **Conclusion:** The goal has been achieved from this analysis. However, this has also opened opportunities for further analysis specially in real-state sector. This is just one small piece of example of data science uses cases that one can realize applying technology which is available for free today.

- - - - -