Hotel Segmentation

Akshat Pattiwar

23 April 2021

1. Introduction

Graduation is the dream of all parents to their children who have completed their studies. Sometimes many parents come all the way to attend the graduation process of their child. **Usually those from out of town will stay at the closest hotel to the university before coming to the graduation ceremony so as not to be late and save time. Interesting places around the hotel can also attract people to stay.**

To determine the recommend hotel, we can use the clustering technique to divide the nearest hotel group into several segments or groups. The data used is the hotel data closest to the Indonesian Computer University (UNIKOM) and the places around the hotel which were taken using the Foursquare API. Furthermore, they will be grouped into several segments using the K-means algorithm to determine the recommend hotel choice.

2. Data

Based on definition of our problem, we need data about:

- Hotels around the Indonesian Computer University with radius 500 meter.
- Places near the hotel with coordinate location.

The data is retrieved using the Foursquare API.

3. Methodology

To access the Foursquare API, we need the following tokens :

- **CLIENT_ID** : Your Foursquare Client ID
- **CLIENT_SECRET** : Your Foursquare Client Secret
- **ACCESS_TOKEN** : Your Four Square Access Token
- **VERSION** : Your Foursquare Version

3.1 Import Necessary Library

The libraries that will be used include:

- **numpy** and **pandas** : data manipulation and analysis
- **request** and **json** : access to url and save into json file
- **geopy** : define coordinate location
- **sklearn** : data modelling

3.2 Data Preparation

Before accessing the Foursquare API we need to know the location of the data we want. Here we will use the UNIKOM location as a reference.

```
In [4]: address = 'Universitas Komputer Indonesia, Jawa Barat, Indonesia'

        geolocator = Nominatim(user_agent="foursquare_agent")
        location = geolocator.geocode(address)
        latitude = location.latitude
        longitude = location.longitude
        print(latitude, longitude)
```

```
-6.8868408 107.6151136
```

Hotels Data

Next, we will take hotel data around UNIKOM with a radius of 500 meters. The steps taken include:

- create url with predefined parameters.

```
search_key = 'hotel'
radius = 500
url_key = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&oauth_token={}&v={}&query={}&radius={}&limit={}'.format(
    CLIENT_ID, CLIENT_SECRET, latitude, longitude,
    ACCESS_TOKEN, VERSION, search_key, radius, LIMIT)
```

- Fetch hotel data in json file format.
- Converts json files to tabular data.

]:

	id	name	categories	referralId	hasPerk	location.address	location.lat	location.lng	location.labeledLatLngs	location.distance	location.po
0	4c72303f9e706dcbf3374a62	Hotel Neo Dipatiukur	[{'id': '4bf58dd8d48988d1fa931735', 'name': 'H...	v-1619096989	False	Jl. Dipatiukur No. 72-74	-6.889919	107.616451	[{'label': 'display', 'lat': -6.88991914858835...	373	
1	515e6be5e4b0c134437ee63a	Patra Jasa Hotel	[{'id': '4bf58dd8d48988d1fa931735', 'name': 'H...	v-1619096989	False	Jl. Ir. H. Juanda No.132	-6.889154	107.614085	[{'label': 'display', 'lat': -6.88915357376354...	281	
2	4bb73d162f70c9b68e668630	Scarlet Dago Hotel	[{'id': '4bf58dd8d48988d1fa931735', 'name': 'H...	v-1619096989	False	Jalan Siliwangi No. 5	-6.885105	107.612117	[{'label': 'display', 'lat': -6.88510543232216...	383	
3	4be81e5dd837c9b6efe4a406	Royal Dago Hotel	[{'id': '4bf58dd8d48988d1fa931735', 'name': 'H...	v-1619096989	False	Jl. Ir. H. Djuanda No. 156-169	-6.886584	107.613328	[{'label': 'display', 'lat': -6.88658352860527...	199	
4	4dfef268483b8f033b425a03	Hotel Puspa Kencana	[{'id': '4bf58dd8d48988d1fa931735', 'name': 'H...	v-1619096989	False	Jl. Siliwangi no. 10	-6.884576	107.612229	[{'label': 'display', 'lat': -6.8845756140164,...	406	

- perform data cleaning such as column filtering and remove similar hotel names.

Out[7]:

	id	name	address	lat	lng
0	4c72303f9e706dcbf3374a62	Hotel Neo Dipatiukur	Jl. Dipatiukur No. 72-74	-6.889919	107.616451
1	515e6be5e4b0c134437ee63a	Patra Jasa Hotel	Jl. Ir. H. Juanda No.132	-6.889154	107.614085
3	4be81e5dd837c9b6efe4a406	Royal Dago Hotel	Jl. Ir. H. Djuanda No. 156-169	-6.886584	107.613328
4	4dfef268483b8f033b425a03	Hotel Puspa Kencana	Jl. Siliwangi no. 10	-6.884576	107.612229
6	4bef544ff2712d7f21a9fbd8	Geulis Boutique Hotel	Jl. Ir. H Juanda	-6.891400	107.612839

Nearby Places Around The Hotels

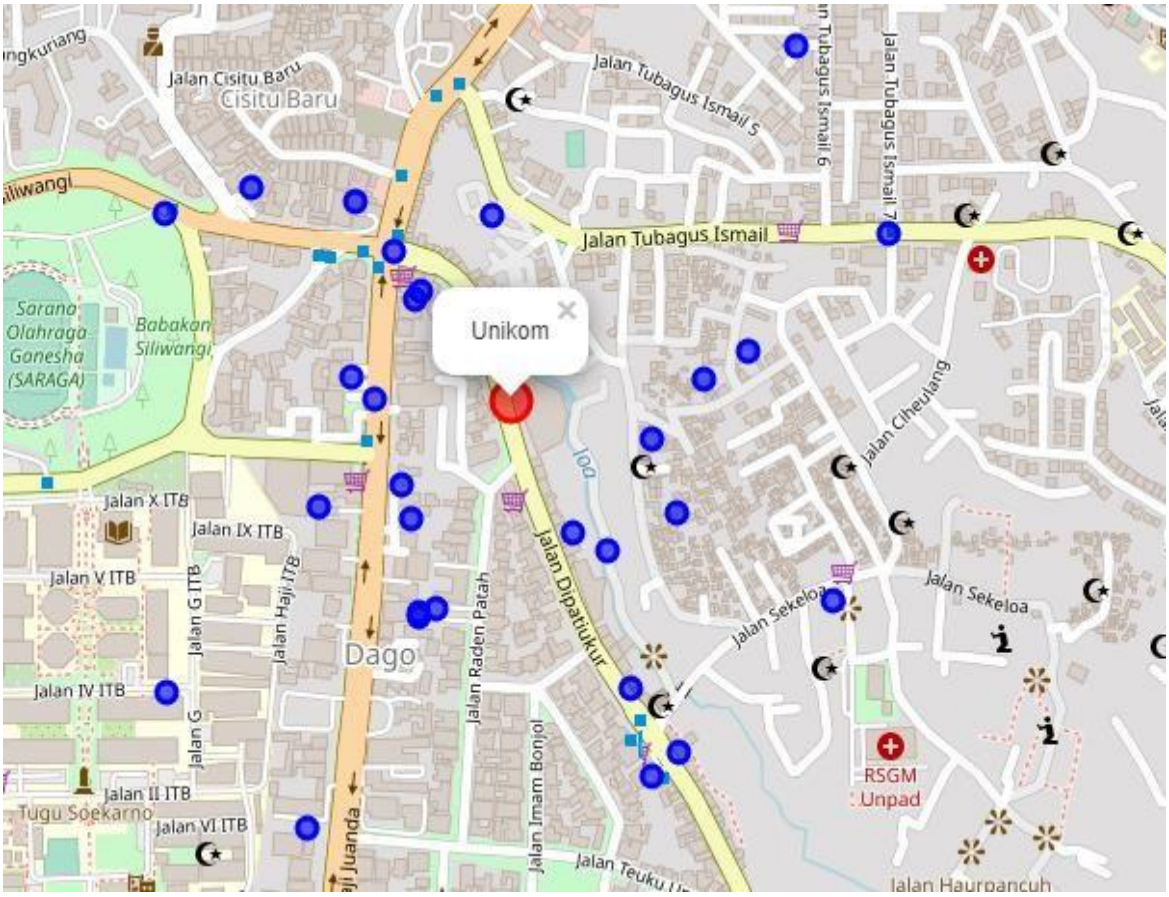
Next, retrieve data on the closest places around each hotel using the Foursquare API.

Out[11]:

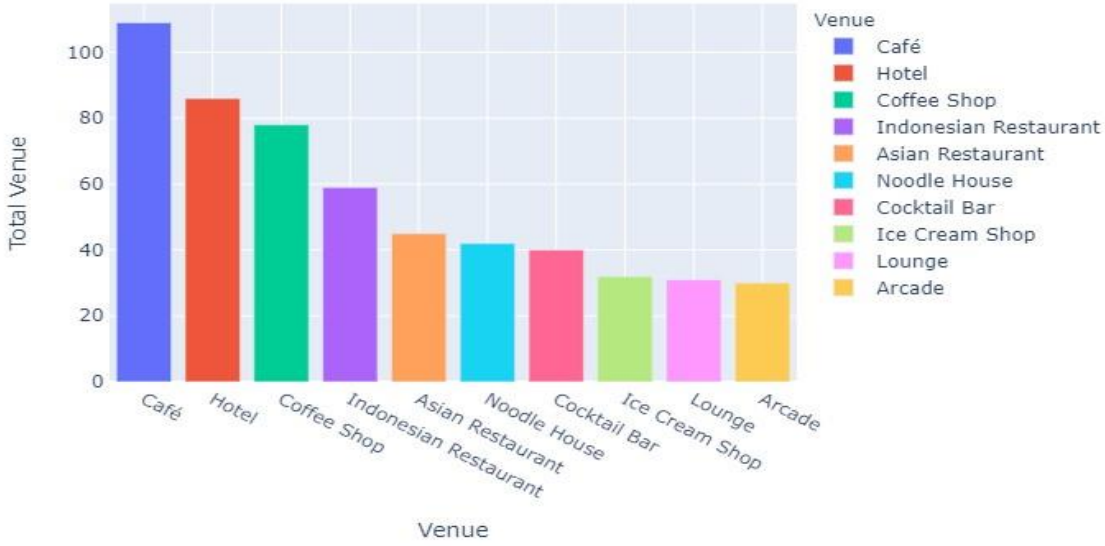
	Venue Name	Venue Latitude	Venue Longitude	Venue Category	Hotel Name	Hotel Latitude	Hotel Longitude
0	Hotel Neo Dipatiukur	-6.889919	107.616451	Hotel	Hotel Neo Dipatiukur	-6.889919	107.616451
1	Koffie Opa	-6.891955	107.616803	Coffee Shop	Hotel Neo Dipatiukur	-6.889919	107.616451
2	Mie Merapi	-6.891459	107.616998	Noodle House	Hotel Neo Dipatiukur	-6.889919	107.616451
3	Hotel Patra Jasa Bandung	-6.889111	107.614090	Hotel	Hotel Neo Dipatiukur	-6.889919	107.616451
4	Bebeke Om Aris	-6.892662	107.617820	Asian Restaurant	Hotel Neo Dipatiukur	-6.889919	107.616451

3.3 Data Visualization

Hotels Map



Top 10 Venue Categories



3.4 Data Modeling

We will do hotel clustering based on the surrounding places. Because several hotels next to each other, then we will eliminate the category of hotel as the place around. After that, **one hot encoding** will be carried out on the **Venue Category** data, and then grouping each hotel.

]:

	Hotel Name	Acehnese Restaurant	American Restaurant	Arcade	Asian Restaurant	Bakery	Bar	Bistro	Breakfast Spot	Bridge	Burger Joint	Café	Chinese Restaurant	Cocktail Bar	Coffee Shop	College Auditorium	College Gym	College Stadium
0	Crowne hotel	0.000000	0.0	0.035714	0.035714	0.000000	0.0	0.035714	0.000000	0.0	0.035714	0.178571	0.0	0.071429	0.142857	0.0	0.0	0.0
1	D'best Sofia Hotel Bandung	0.000000	0.0	0.086957	0.086957	0.043478	0.0	0.000000	0.043478	0.0	0.000000	0.130435	0.0	0.000000	0.173913	0.0	0.0	0.0
2	Geulis Boutique Hotel	0.000000	0.0	0.120000	0.040000	0.040000	0.0	0.000000	0.000000	0.0	0.000000	0.160000	0.0	0.000000	0.120000	0.0	0.0	0.0
3	Grand Ball Room Panghegar Hotel	0.000000	0.0	0.000000	0.074074	0.000000	0.0	0.000000	0.000000	0.0	0.037037	0.185185	0.0	0.074074	0.111111	0.0	0.0	0.0
4	Grand Kopo Hotel	0.035714	0.0	0.000000	0.071429	0.000000	0.0	0.000000	0.000000	0.0	0.035714	0.178571	0.0	0.035714	0.178571	0.0	0.0	0.0

Next, do hotel clustering with **Venue Category** as feature data using the **K-means** algorithm with the desired number of clusters, namely 5. Finally, combine the data clusters with hotel data along with the top 10 venue categories in each hotel.

]:

	Hotel Name	Hotel Latitude	Hotel Longitude	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	Cluster
0	Hotel Neo Dipatiukur	-6.889919	107.616451	Café	Coffee Shop	Dessert Shop	Arcade	Noodle House	Breakfast Spot	Indonesian Restaurant	Lounge	Toy / Game Store	Bakery	1
1	Royal Hotel	-6.886584	107.613328	Café	Indonesian Restaurant	Coffee Shop	Arcade	Asian Restaurant	Lounge	Cocktail Bar	Convenience Store	Bridge	Noodle House	3
2	Hotel Puspa Kencana	-6.884576	107.612229	Café	Coffee Shop	Cocktail Bar	Ice Cream Shop	Indonesian Restaurant	Lounge	Bridge	Noodle House	Fast Food Restaurant	College Gym	3
3	Geulis Boutique Hotel	-6.891400	107.612839	Café	Indonesian Restaurant	Arcade	Coffee Shop	Shoe Store	Noodle House	Toy / Game Store	Kids Store	Lounge	Ice Cream Shop	1
4	Hotel Sheraton	-6.885765	107.614036	Café	Fast Food Restaurant	Coffee Shop	Cocktail Bar	Asian Restaurant	Ice Cream Shop	Indonesian Restaurant	Bridge	Noodle House	Lounge	3

Clusters Map

4. Result and Discussion

The following is the hotel data with the top 10 places nearby which are quite interesting and their location after clustering.

Cluster 0

```
In [85]: # cluster 0
         hotels_merged[hotels_merged['Cluster'] == 0]
```

Out[85]:

	Hotel Name	Hotel Latitude	Hotel Longitude	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	Cluster
6	Hotel Patra Jasa	-6.889111	107.614090	Coffee Shop	Café	Noodle House	Cocktail Bar	Indonesian Restaurant	Bakery	Lounge	Fast Food Restaurant	Sporting Goods Shop	Arcade	0
7	Hotel Akasia	-6.886620	107.617260	Café	Coffee Shop	Indonesian Restaurant	Noodle House	Fast Food Restaurant	Ice Cream Shop	Dessert Shop	Convenience Store	Asian Restaurant	Lounge	0
14	Crowne Hotel	-6.888255	107.615798	Café	Coffee Shop	Noodle House	Convenience Store	Indonesian Restaurant	Cocktail Bar	Dessert Shop	Burger Joint	Ice Cream Shop	Lounge	0
15	Grand Ball Room Panghegar Hotel	-6.887248	107.616684	Café	Coffee Shop	Noodle House	Fast Food Restaurant	Convenience Store	Asian Restaurant	Indonesian Restaurant	Cocktail Bar	Dessert Shop	Sundanese Restaurant	0
16	Fitness Center Galeri Ciumbuleuit Hotel & Apar...	-6.888450	107.616193	Café	Coffee Shop	Convenience Store	Indonesian Restaurant	Cocktail Bar	Dessert Shop	Noodle House	Burger Joint	Food Court	Lounge	0
22	Grand Kopo Hotel	-6.888047	107.616955	Café	Coffee Shop	Indonesian Restaurant	Asian Restaurant	Convenience Store	Dessert Shop	Noodle House	Food Truck	Lounge	Burger Joint	0

Cluster 1

```
In [86]: # cluster 1
         hotels_merged[hotels_merged['Cluster'] == 1]
```

Out[86]:

	Hotel Name	Hotel Latitude	Hotel Longitude	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	Cluster
0	Hotel Neo Dipatiukur	-6.889919	107.616451	Café	Coffee Shop	Dessert Shop	Arcade	Noodle House	Breakfast Spot	Indonesian Restaurant	Lounge	Toy / Game Store	Bakery	1
3	Geulis Boutique Hotel	-6.891400	107.612839	Café	Indonesian Restaurant	Arcade	Coffee Shop	Shoe Store	Noodle House	Toy / Game Store	Kids Store	Lounge	Ice Cream Shop	1
18	Hotel Arimbi Stasion	-6.890589	107.616992	Coffee Shop	Indonesian Restaurant	Café	Arcade	Asian Restaurant	Noodle House	Breakfast Spot	Kids Store	Dessert Shop	Toy / Game Store	1
23	Hotel Puri Gardenia	-6.890862	107.616674	Coffee Shop	Arcade	Café	Noodle House	Breakfast Spot	Ice Cream Shop	Indonesian Restaurant	Kids Store	Dessert Shop	Toy / Game Store	1

Cluster 2

```
In [87]: # cluster 2
         hotels_merged[hotels_merged['Cluster'] == 2]
```

Out[87]:

	Hotel Name	Hotel Latitude	Hotel Longitude	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	Cluster
9	Hotel Efita	-6.886307	107.617765	Asian Restaurant	Café	Indonesian Restaurant	Coffee Shop	Convenience Store	Dessert Shop	Fast Food Restaurant	Noodle House	Ice Cream Shop	Food Truck	2
21	Hotel Lotus	-6.885058	107.619329	Convenience Store	Asian Restaurant	Coffee Shop	Dessert Shop	Indonesian Restaurant	Food Truck	Juice Bar	Bistro	Burger Joint	Café	2
24	Hotel Alila Villas Soori, Tabanan ,Bali	-6.883052	107.618287	Convenience Store	Ice Cream Shop	Asian Restaurant	Dessert Shop	Sundanese Restaurant	Coffee Shop	Indonesian Restaurant	Chinese Restaurant	Café	Burger Joint	2

Cluster 3

```
In [88]: # cluster 3
         hotels_merged[hotels_merged['Cluster'] == 3]
```

Out[88]:

	Hotel Name	Hotel Latitude	Hotel Longitude	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	Cluster
1	Royal Hotel	-6.886584	107.613328	Café	Indonesian Restaurant	Coffee Shop	Arcade	Asian Restaurant	Lounge	Cocktail Bar	Convenience Store	Bridge	Noodle House	3
2	Hotel Puspa Kencana	-6.884576	107.612229	Café	Coffee Shop	Cocktail Bar	Ice Cream Shop	Indonesian Restaurant	Lounge	Bridge	Noodle House	Fast Food Restaurant	College Gym	3
4	Hotel Sheraton	-6.885765	107.614036	Café	Fast Food Restaurant	Coffee Shop	Cocktail Bar	Asian Restaurant	Ice Cream Shop	Indonesian Restaurant	Bridge	Noodle House	Lounge	3
5	Hotel Royale	-6.884864	107.614899	Café	Cocktail Bar	Coffee Shop	Ice Cream Shop	Indonesian Restaurant	Asian Restaurant	Burger Joint	Lounge	Convenience Store	Noodle House	3
8	Hotel Safari Garden	-6.887974	107.612964	Café	Fast Food Restaurant	Coffee Shop	Indonesian Restaurant	Cocktail Bar	Bridge	College Stadium	Gas Station	Ice Cream Shop	Lounge	3
10	Hotel Shangri-La	-6.884710	107.613377	Café	Coffee Shop	Cocktail Bar	Asian Restaurant	Ice Cream Shop	Indonesian Restaurant	Lounge	Bridge	Noodle House	Fast Food Restaurant	3
11	Prime Park Hotel	-6.887748	107.613901	Café	Fast Food Restaurant	Coffee Shop	Indonesian Restaurant	Cocktail Bar	Bridge	College Stadium	Gas Station	Ice Cream Shop	Lounge	3
12	Proyek Hotel Lor In Lembang	-6.885684	107.614102	Café	Fast Food Restaurant	Coffee Shop	Cocktail Bar	Asian Restaurant	Ice Cream Shop	Indonesian Restaurant	Bridge	Noodle House	Lounge	3

Cluster 4

```
In [89]: # cluster 4
         hotels_merged[hotels_merged['Cluster'] == 4]
```

Out[89]:

	Hotel Name	Hotel Latitude	Hotel Longitude	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	Cluster
17	Hotel Cemerlang	-6.888986	107.618706	Asian Restaurant	Coffee Shop	Noodle House	Convenience Store	Indonesian Restaurant	Café	Burger Joint	Dessert Shop	Food Truck	Breakfast Spot	4

When viewed in the map, it can be seen that:

- **Cluster 0 (red)** is closer to UNIKOM but nearby places are less popular and some hotels require detour access.

- **Cluster 1 (purple)** is located before UNIKOM with a fairly strategic location but it appears that there are other universities that are possible if the road is quite busy. Apart from that there are bus stops too.
- **Cluster 2 (blue)** is located quite far from UNIKOM and less strategic places around it.
- **Cluster 3 (sky blue)** is located in the most strategic location, although not too close to UNIKOM, road access is quite easy because it is the main road.
- **Cluster 4 (orange)** does not look strategic, other than that it is not on the main road.

5. Conclusion

Hotels in Cluster 3 are highly recommended because there are more hotel choices depending on your taste and road access is easier. In addition, the places around the hotel are more or less the same because they are included in the top 10 nearby places. Maybe it can be developed further because the distance of the hotel to UNIKOM is also affected by the traffic flow as well.