Where to Explore in Yogyakarta?

Muhammad Faris Herlansyah

August 7, 2021

1. Introduction

1.1 Background

Yogyakarta, a city in Indonesia which also serves as the capital of the Special Region of Yogyakarta, attracts many tourists each year. The city is known for its strong Javanese cultural influence due to the Special Region of Yogyakarta being ruled by a monarchy, its relatively low living cost, and its population which is dominated by young people due to the city hosting some well-known universities. Many places that attract tourists include Keraton Yogyakarta (the seat of the ruling sultan), Jalan Malioboro with its shops selling batik clothes, Kotagede which is well-known for being the centre of silver-based handicrafts, plenty of food stalls, homey hotels, et cetera. Those characteristics make visitors have plenty of choices of what to do while visiting Yogyakarta.

1.2 Problem

However, sometimes first-time visitors to Yogyakarta do not know where or in which districts to focus their vacation on, which may be due to the variety of tourist attractions and distribution of attractions and public facilities (e.g. historical attractions and places to hang out may obviously be concentrated in different parts of the town). Also sometimes when they arrive they may not have yet researched about attractions in Yogyakarta, hence a quick recommendation for first-time visitors may be needed.

1.3 Interests

This analysis is intended to facilitate first-time visitors to Yogyakarta so that they can decide where to visit, what kinds of places they would love, where to stay, where to find cheap foods, etc.

2. Data Preparation

2.1 Data Source

I scraped kodepos.nomor.net to obtain data of all districts and urban villages. The data contains not only all districts in Yogyakarta but also postal codes of each district in Yogyakarta. Besides, I also used Foursquare API to obtain the data containing all venues located within the distance of several hundred metres from each district's centre, and Nominatim from Geopy to acquire geographical coordinates of each district in Yogyakarta.

2.2 Data Cleaning

The webpage from which the districts data were scraped was, in my opinion, aesthetically untidy. Moreover, there were more than two rows representing column names, which was confusing. Also there were doubled place names and a postal code column which contained strings other than postal codes. Hence I performed data cleaning by dropping column name duplicates, resetting indexes, and changing several cells which contain duplicates and unnecessary characters.

Moreover, because we want to cluster districts, it is obvious that we need geographical coordinates data of each district in Yogyakarta. Hence, again I used Nominatim to obtain the geographical coordinates of Yogyakarta, as well as geographical coordinates of each of its districts. Then, I combined the obtained geographical coordinates data with the districts data.

1	Kode POS	Kelurahan	Kecamatan	Kota	Lat_Kecamatan	Long_Kecamatan
0	55253	Pakuncen	Wirobrajan	Yogyakarta	-7.802624	110.350447
1	55251	Patangpuluhan	Wirobrajan	Yogyakarta	-7.802624	110.350447
2	55252	Wirobrajan	Wirobrajan	Yogyakarta	-7.802624	110.350447
3	55163	Giwangan	Umbulharjo	Yogyakarta	-7.814378	110.387374
4	55165	Muja Muju	Umbulharjo	Yogyakarta	-7.814378	110.387374
5	55161	Pandeyan	Umbulharjo	Yogyakarta	-7.814378	110.387374
6	55166	Semaki	Umbulharjo	Yogyakarta	-7.814378	110.387374
7	55162	Sorosutan	Umbulharjo	Yogyakarta	-7.814378	110.387374
8	55167	Tahunan	Umbulharjo	Yogyakarta	-7.814378	110.387374
9	55164	Warungboto	Umbulharjo	Yogyakarta	-7.814378	110.387374
10	55243	Bener	Tegalrejo	Yogyakarta	-7.780455	110.355073
11	55241	Karangwaru	Tegalrejo	Yogyakarta	-7.780455	110.355073
12	55242	Kricak	Tegalrejo	Yogyakarta	-7.780455	110.355073
13	55244	Tegalrejo	Tegalrejo	Yogyakarta	-7.780455	110.355073
14	55111	Gunungketur	Pakualaman	Yogyakarta	-7.800395	110.376249
15	55112	Purwokinanti	Pakualaman	Yogyakarta	-7.800395	110.376249
16	55261	Ngampilan	Ngampilan	Yogyakarta	-7.802183	110.357603
17	55262	Notoprajan	Ngampilan	Yogyakarta	-7.802183	110.357603

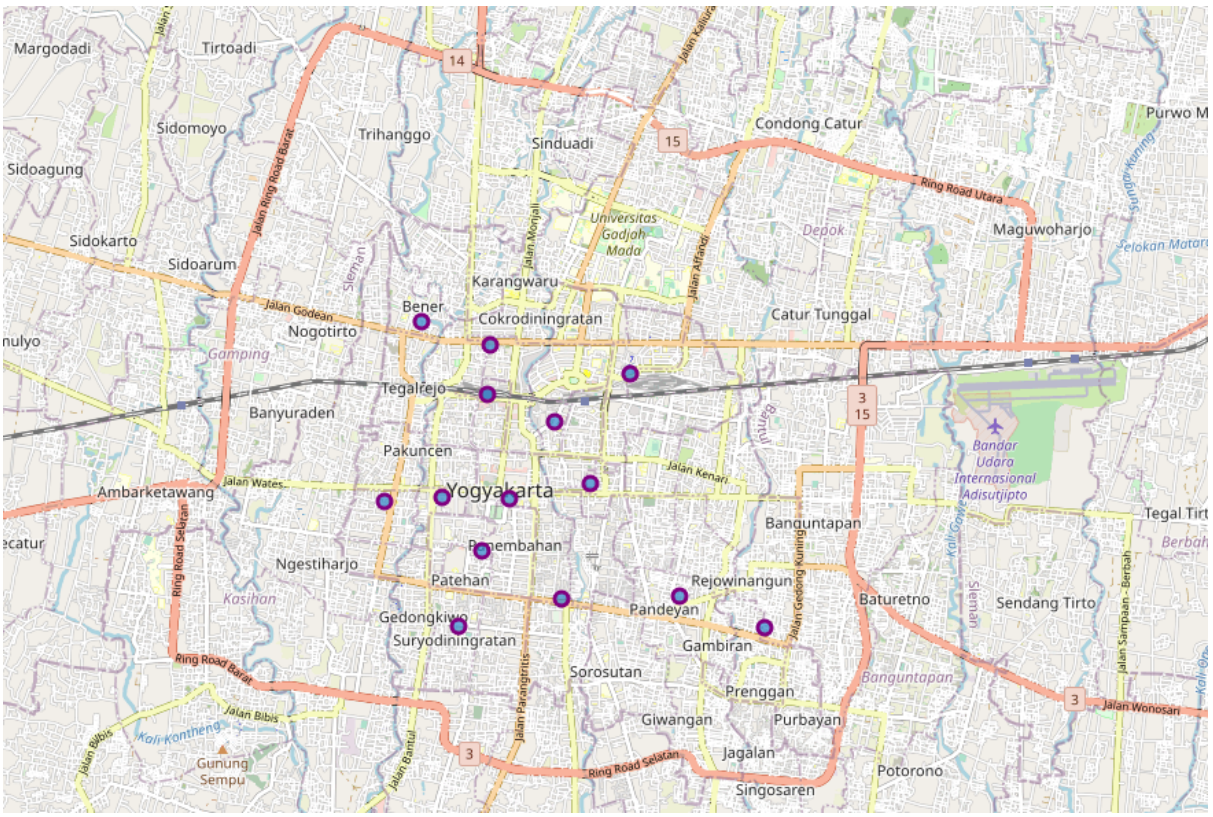
2.3 Feature selection

Given and known that urban villages in Yogyakarta have relatively small areas, which may contain only a small number of venues, I decided not to use the urban villages data; I used only postal code, districts, and city data.

The districts and urban villages data were combined with the latitudes and longitudes data. Then, the data were grouped by districts. Hence, we obtained a data frame containing 14 districts of Yogyakarta and their respective geographical coordinates.

1	Kode POS	Kecamatan	Kota	Lat_Kecamatan	Long_Kecamatan
0	55211	Danurejan	Yogyakarta	-7.792842	110.371795
1	55271	Gedongtengen	Yogyakarta	-7.789338	110.363376
2	55221	Gondokusuman	Yogyakarta	-7.786791	110.381157
3	55121	Gondomanan	Yogyakarta	-7.802395	110.366112
4	55231	Jetis	Yogyakarta	-7.783297	110.363649
5	55171	Kotagede	Yogyakarta	-7.818311	110.397941
6	55131	Kraton	Yogyakarta	-7.808799	110.362726
7	55141	Mantrijeron	Yogyakarta	-7.818067	110.359731
8	55151	Mergangsan	Yogyakarta	-7.814734	110.372558
9	55261	Ngampilan	Yogyakarta	-7.802183	110.357603
10	55111	Pakualaman	Yogyakarta	-7.800395	110.376249
11	55241	Tegalrejo	Yogyakarta	-7.780455	110.355073
12	55161	Umbulharjo	Yogyakarta	-7.814378	110.387374
13	55251	Wirobrajan	Yogyakarta	-7.802624	110.350447

After that, I created a visualization of Yogyakarta's districts using Folium. The resulting map is as follows, with the dots representing all districts of Yogyakarta:



2.4 Using Foursquare API to Obtain Venues Data

Now we start using the Foursquare API to obtain data of venues in Yogyakarta.

First, we define our Foursquare credentials, which comprises client ID, client secret, API version, and limit value to determine the number of venues obtained. Then we try obtaining venue data at one district, to ensure that the API works well in obtaining data. If the API works well, it will return data in JSON format.

After obtaining the data in JSON format, we can reformat the data to be in the form of a data frame. Below is an example of a data frame containing venues data in the district of Gedongtengen:

	name	categories	lat	lng
0	Stasiun Yogyakarta Tugu	Train Station	-7.789425	110.363460
1	Loko Cafe	Café	-7.789171	110.363488
2	Jogja Scrummy Stasiun Tugu	Food & Drink Shop	-7.789426	110.363686
3	Roti Cane Maryam Sta. Tugu	Bakery	-7.789235	110.363395
4	Top Roof Inna Garuda Hotel	Roof Deck	-7.789274	110.363896

Because the above process has worked well, we apply the similar process for all districts in Yogyakarta. By creating a function using Foursquare APIs and method to convert JSON data into data frame, the following data frame is generated:

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Danurejan	-7.792842	110.371795	Laura's Backpackers	-7.790985	110.370988	Hostel
1	Danurejan	-7.792842	110.371795	Jagung Bakar Menthel	-7.792026	110.373640	Snack Place
2	Danurejan	-7.792842	110.371795	BAKPIAKU Gajah Mada (pusat)	-7.797168	110.372700	Bakery
3	Danurejan	-7.792842	110.371795	Bakpia Citra Premium	-7.793618	110.368021	Snack Place
4	Danurejan	-7.792842	110.371795	Rumah Makan Murni 83	-7.791199	110.372408	Asian Restaurant
5	Danurejan	-7.792842	110.371795	Sate Ayam Podomoro	-7.795822	110.369001	Asian Restaurant
6	Danurejan	-7.792842	110.371795	Indomaret Jl.Hayam Wuruk	-7.795520	110.372715	Department Store
7	Danurejan	-7.792842	110.371795	Spesial Sambal Mentah "Bu Saring"	-7.792306	110.373451	Asian Restaurant
8	Danurejan	-7.792842	110.371795	warung makan pak parno sate kambing dan tongseng	-7.792491	110.373380	Diner
9	Danurejan	-7.792842	110.371795	Kedai Kopi Mataram	-7.792888	110.367511	Coffee Shop

We can also check the shape of the above data frame:

```
[24] print(yogya_venues.shape)

     (373, 7)
```

We can check how many venues returned for each district:

District	Venue
Danurejan	16
Gedongtengen	39
Gondokusuman	33
Gondomanan	53
Jetis	66
Kotagede	13
Kraton	30
Mantrijeron	25
Mergangsan	24
Ngampilan	9
Pakualaman	21
Tegalrejo	12
Umbulharjo	15
Wirobrajan	17