Coursera Capstone Athens Project

The Battle of Neighborhoods - Final Report (Week 1)

1. Introduction Section :

Discussion of the business problem

1.1 Scenario and Background

I just started working as an Application Support Engineer although my real passion is to find a job and work as a data scientist. I currently live within walking distance from "Megaro Mousikis metro station" near the center of Athens therefore I have access to good public transportation to work. Likewise, I enjoy many amenities in the neighborhood , such as international cousine restaurants, cafes, food shops and entertainment. Although you don' t really care about my personal life I have just recently started looking for a bigger house in order to live with my girlfriend and start a family. Although, I am very excited about it, I am a bit stressed towards the process to secure a comparable place to live in Athens, since there is a significant raise in rental prices within the last few years. Therefore, I decided to apply the skills I have learned during studying for the Coursera courses to explore ways to make sure my decision is factual and rewarding. Of course, there are alternatives to achieve the answer using available Google and Social media tools, but it will be rather rewarding doing it myself with the tools used so far.

1.2 Problem to be resolved:

The challenge to resolve is being able to find a rental apartment unit in Attica GR that offers similar characteristics and benefits to my current situation. Therefore, in order to set a basis for comparison, I want to find a rental unit subject to the following conditions:

- Apartment with min 2 bedrooms with monthly rent not to exceed 700 euro/month
- Unit located within walking distance (<= 1.5 km) from a subway metro station in Attica
- Area with amenities and venues similar to the ones described for current location (See item 2.1)

1.3 Interested Audience:

I believe this is a relevant project for everyone considering moving to a major city in Europe, US or Asia, since the approach and methodologies used here are applicable in all cases. The use of FourSquare data and mapping techniques combined with data analysis will help resolve the key questions arisen. Lastly, this project is a good practical case towards the development of Data Science skills.

2. Data Section:

Description of the data and its sources that will be used to solve the problem

2.1 Data of Current Situation

I Currently reside in the neighborhood of 'Kolonaki' near Athens city center. Foursquare will be used to identify the venues around the area of residence which will be shown in Athens map displayed in methodology and execution in section 3.0 . It serves as a reference for comparison with the desired future location.

2.2 Data Required to resolve the problem

In order to make a good choice of a similar apartment, the following data is required: List/Information of Attica neighborhoods with their geodata (latitude and longitude). List/Information of the subway metro stations with Geodata. Listed apartments for rent in Athens area with descriptions (number of bedrooms, apartment size, price, location). Venues and amenities in Athens neighborhoods (e.g. top 10).

2.3 Sources and manipulation

The list of Athens neighborhoods is scraped from Wikipedia link https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Athens and is passed to a list and then to the pandas DataFrame "df_neighborhoods" along with latitude and longitude retrieved from Nominatim.

```
                            Neighborhood   Latitude  Longitude
0     Agios Eleftherios, Athens, Greece  38.020044  23.731724
1  Agios Panteleimonas, Athens, Greece  37.996564  23.726957
2             Akadimia Platonos, Greece  37.989357  23.711217
3              Akadimia, Athens, Greece  37.980285  23.734528
4                     Anafiotika, Greece  37.972351  23.728043
```

A list of Athens subway metro stations was scraped once again from Wikipedia (https://en.wikipedia.org/wiki/List_of_Athens_Metro_stations). The geolocation was obtained again using Nominatim and passed to the pandas DataFrame "df_stations".

```
                          Station   Latitude  Longitude
0         Nerantziotissa station  38.045158  23.792984
1               Piraeus station  37.943159  23.647059
2         Elliniko metro station  37.907554  23.737044
3   Agia Paraskevi metro station  38.020815  23.816783
4  Agios Dimitrios metro station  37.946833  23.737825
```

A list of places for rent was collected by web-browsing nepstick site:
https://www.nestpick.com/athens/ working as search engine for rental apartments, retrieving data from different real etate sites. Afterwards data are passed to a DataFrame with the following columns ['name', 'category', 'normalized_price', 'number_of_bedrooms', 'apartment_size', 'latitude', 'longitude']. The loop algorithms used are shown in the execution of data in section 3.0. "Great_circle" function from geolocator was used to calculate distances between two points, as in this case to calculate average rent price for units around each subway station and at 1.5 km radius. Foursquare is used to find the avenues at Athens neighborhoods in general and a cluster is created to later be able to search for the venues depending on the location shown.

2.4 How the data will be used to solve the problem

The data will be used as follows: Use Foursquare and geopy data to map top 10 venues for all Athens neighborhoods and clustered in groups. Use foursquare and geopy data to map the location of subway metro stations, separately and on top of the above clustered map in order to be able to identify the venues and amenities near each metro station, or explore each subway location separately. Use Foursquare and geopy data to map the location of rental places, in some form, linked to the subway locations. Create a map that depicts, for instance, the average rental price per square ft, around a radius of 1.5 km around each subway station. I will be able to quickly point to the popups to know the relative price per subway area.