Coursera Capstone Project

The Battle of Neighborhoods (Week 1 and Week 2) Full Report

Swati

Table of Content

1.Introduction Section

1.1 Background - Discussion of the business problem and relevance of the project:

I am a Software developer and was working in Maryland. But I recently got married and must move to Miami, Florida. I want to I used to live in Owings Mills, Maryland with walking distance to the Grand Junction Metro Station with great connectivity to Baltimore Downtown as well as Washington, D.C., all shopping stores, groceries and pharmacy was nearby to my apartment. But since I am moving to Florida I want to use the skills learned during IBM Data Science Specialization course to help me find a suitable apartment that meets the price range along with good locality which should be connected easily to public transport such as metros and buses and have groceries stores and good restaurants nearby. I can use apartment and renting websites to help me with my situation, but I want to apply my learning and utilize the tools I used to help solve the problem. To help solve the problem, I will evaluate and compare the apartments available for renting in Miami, Florida and they should meet the following measures:

Apartment must be 2 bedrooms.

Desired location is near a metro station in the Manhattan area and within (1.5 km) radius or less than 30 minutes walking distance.

Price of rent not exceed $2,000 per month.

Amenities in the selected neighborhood shall be like my current residence.

Desirable to have venues such as coffee shops, good Restaurants, Gym and Grocery stores nearby.

As a reference, I have included a map of venues near current residence in Maryland.

1.2 Business Problem

The challenge is to find a suitable apartment for rent in Miami, Florida that meets with the demands of appropriate location, price and venues. The data required to resolve this challenge is described in section 2.

1.3 Purpose of Project

This information will be helpful for someone moving to a new city and will help them to find a good apartment, as the basic questions for renting a reasonable apartment in a city are being answered. It will also help individuals interested in exploring coffee shops and restaurants in any city. Lastly, it is serving the purpose of helping me improve my Data Science skills as I can apply in this project.

2. Data Section

2.1 Data description and the sources used to solve the business problem.

Description of the Data - The following data is required to answer the issues of the problem:

List of neighborhoods of Miami, Florida with their geodata (latitude and longitude).

List of apartments for rent in Miami area with their addresses and price.

List of Subway metro stations in Miami with their relative location.

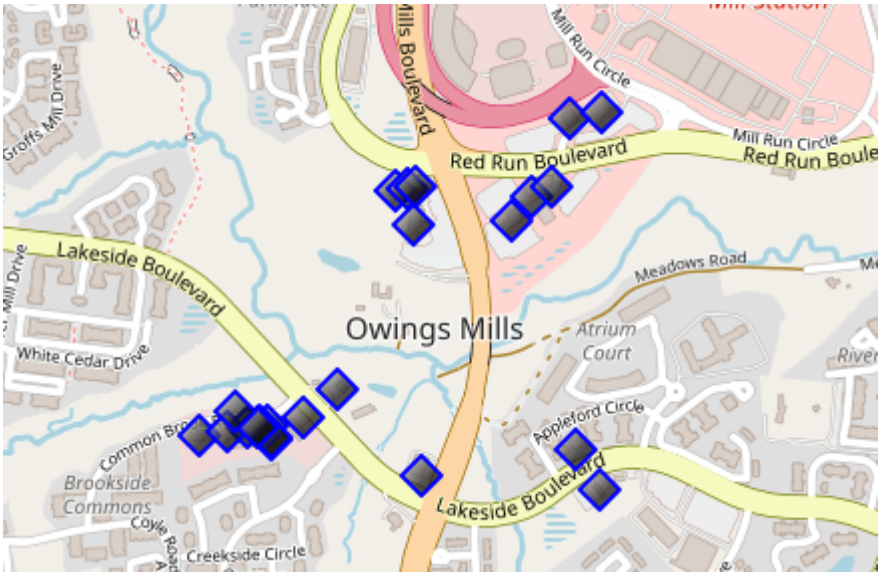
Preferably, a list of apartments for rent with additional information, such as price, address, area, # of beds

Venues for each Miami neighborhoods (that can be clustered).

Venues for subway metro stations.

2.2 Comparison with Present Location

I Currently reside in Owings Mills in state of Maryland. I use Foursquare to identify the venues around the area of residence which are then shown in the (Owings Mills, Maryland) map shown in methodology and execution in section 3.0. It serves as a reference for comparison with the desired future location in Miami Florida.



	name	categories	lat	lng
0	Taj Palace Owings MIlls	North Indian Restaurant	39.399413	-76.800307
1	Artful Gourmet Bistro	New American Restaurant	39.399300	-76.801364
2	Asian Kitchen	Asian Restaurant	39.399578	-76.799104
3	Red Robin Gourmet Burgers and Brews	Burger Joint	39.403242	-76.794221
4	Yuki Sushi	Sushi Restaurant	39.399449	-76.799871
5	Olive Garden	Italian Restaurant	39.404531	-76.793403
6	Brookside's Pizzeria, The Original NY Pizza Co.	Pizza Place	39.399343	-76.800756
7	Baskin-Robbins	Ice Cream Shop	39.399234	-76.799813
8	APlus at Sunoco	Coffee Shop	39.400044	-76.798403
9	New York & Company	Women's Store	39.399051	-76.793292

2.3 Data sources and data manipulation techniques

The list of Miami neighborhoods was worked out during Lab exercise during the course assignments. A csv file was created which will be used in order to create a data frame and for mapping. The csv file ‘Miami_neighbourhood.csv' has the following below data structure. The file will be directly used in the Jupyter Notebook for convenience and space savings. The

clustering of neighborhoods and mapping will be shown. An algorithm was used to determine the geodata from Nominatim. The actual algorithm coding is shown in 'markdown' mode because it takes time to run.

	County	Neighborhood	Latitude	Longitude	Cluster Labels
0	Miami	Allapattah	25.815	-80.224	0.0
1	Miami	Arts & Entertainment District	25.799	-80.190	1.0
2	Miami	Brickell	25.758	-80.193	1.0
3	Miami	Buena Vista	25.813	-80.192	1.0
4	Miami	Coconut Grove	25.712	-80.257	2.0

The list of Miami neighborhood data was found on Wikipedia and the changed to csv and then cluster label was evaluated.

The link - https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Miami

A list of Miami metrorail stops was compiled in Numbers (excel) and the information was found from the below link –

http://www.miamidade.gov/transit/WebServices/TrainStations/?StationID=

To find the final consolidated list of metro stops names and their address, the geolocation was obtained via an algorithm using Nominatim. Details will be shown in the execution of methodology in section 3.0. The subway csv file is ''mfl_metrorail.csv'" and the data structure –

	sub_station	sub_address	lat	long
0	Allapattah	3501 NW 12 Avenue, Miami, Fl 33127	25.808705	-80.215403
1	Brickell	1001 SW First Avenue Miami, Fl 33130	25.763735	-80.195469
2	Brownsville	5200 NW 27 Avenuem Miami, Fl 33142	25.822028	-80.240729
3	Civic Center	1501 NW 12 Avenue Miami, Fl 33136	25.789686	-80.215124
4	Coconut Grove	2780 SW 27 Avenue Miami FL 33133	25.739809	-80.238819

A list of places for rent was collected by web-browsing real estate companies in Miami-

https://www.apartments.com/miami-fl/?gclid=CjwKCAjwqJ_1BRBZEiwAv73uwC2KUZFA9EcsibE9JM4UOW49pdDB-jVb63qq2Z3lNVoaXp1QamjbvRoC-IEQAvD_BwE&gclsrc=aw.ds

A csv file was compiled with the rental place that indicated: areas of Miami, address, number of beds, area and monthly rental price. The csv file " Miami_flats_price_LatLong.csv" had the following below structure.

	Address	Area	Price_per_ft2	Rooms	Area-ft2	Rent_Price	Lat	Long
0	West 105th Street	Allapattah	2.94	5	3400	10000	25.815	-80.224
1	East 97th Street	Arts & Entertainment District	3.57	3	2100	7500	25.799	-80.190
2	West 105th Street	Brickell	1.89	4	2800	5300	25.758	-80.193
3	Cypress Bend	Buena Vista	3.03	2	1650	5000	25.813	-80.192
4	owings Run	Coconut Grove	3.45	2	1450	5000	25.712	-80.257

An algorithm was used to create all the geodata using Nominatim, as shown in section 3.0. The actual algorithm coding may be shown in 'markdown' mode because it takes time to run. With the use of geolocator = Nominatim() , it was possible to determine the latitude and longitude for the subway metro locations as well as for the geodata for each rental place listed. 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 the case to calculate average rent price for units around each subway station and at 1.6 km radius. Foursquare is used to find the avenues at Miami neighborhoods in general and a cluster is created to later be able to search for the venues depending of the location shown.

2.4 How it will help to evaluate the decision of renting

The data will be used as follows:

*Use Foursquare and geopy data to map top 10 venues for all Manhattan neighborhoods and clustered in groups (as per Course LAB)

*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 - or a similar metric. I will be able to quickly point to the popups to know the relative price per subway area.

*Addresses from rental locations will be converted to geodata(lat, long) using Geopy-distance and Nominatim.

*Data will be searched in open data sources if available, from real estate sites if open to reading, libraries or other government agencies.

The processing of these DATA will allow to answer the key questions to decide:

*what is the cost of rent (per square ft) around a mile radius from each subway metro station?

*what is the area of Manhattan with best rental pricing that meets criteria established?

*What are the venues of the two best places to live? How the prices compare?

*How venues distribute among Miami neighborhoods and around metro stations?

*Are there tradeoffs between size and price and location?

*Any other interesting statistical data findings of the real estate and overall data.

3. Methodology Section

3.1 Process steps and strategy to resolve the problem

This section represents the main component of the report where the data is gathered, prepared for analysis. The tools described are used here and the Notebook cells indicate the execution of steps.

The analysis and the strategy:

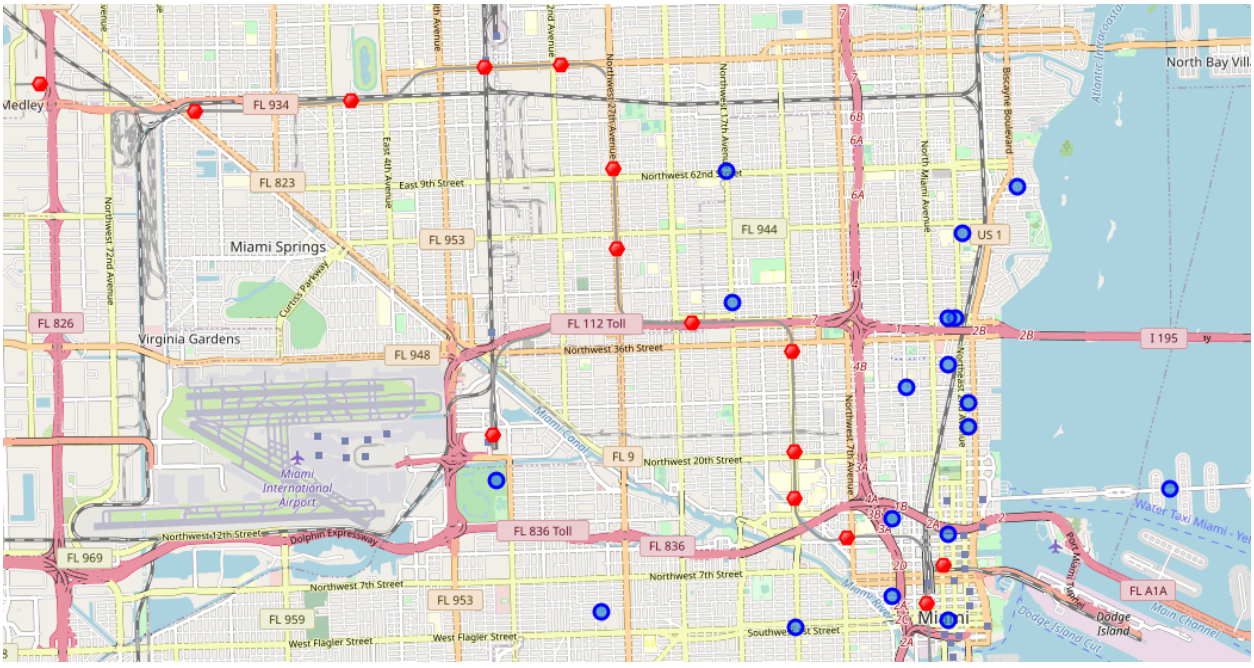
The strategy is based on mapping the above described data in section 2.0, in order to facilitate the choice of at least two candidate places for rent. The choice is made based on the demands imposed: location near a subway, rental price and similar venues to Southbank. This visual approach and maps with pop-up labels allow quick identification of location, price and feature, thus making the selection very easy.

The processing of these DATA and its mapping will allow to answer the key questions to decide:

what is the cost of available rental places that meet the demands? *What* is the cost of rent around a mile radius from each subway metro station? what is the area of Manhattan with best rental pricing that meets criteria established? What are the venues of the two best places to live? How the prices compare? How venues distribute among Manhattan neighborhoods and around metro stations? Are there tradeoffs between size and price and location? Any other interesting statistical data findings of the real estate and overall data.

3.2 Data Science Methods, machine learning, mapping tools and exploratory data analysis

- Creating Maryland Map – Current residence and listing out nearby venues and restaurants.
- Use FourSquare to find venues around current residence.
- Cluster neighborhood data was produced with Foursquare during course lab work. A csv file was produced containing the neighborhoods around the Miami county resulting in 23 neighborhoods. Now, the csv file is just read for convenience and consolidation of report.
- Now the data with initial 10 cluster is worked on to find the venues and other valuable information.
- After examining several cluster data, I found cluster 2 to have resemblance with my current residence and value was assigned to explore the cluster.
- Various web site search resulted in a csv file having various rental apartment listing and relevant information, but it didn't have latitude and longitude. They were found out using algorithm and Nominatim. This can be seen in the lab work.
- Geolocation was obtained for rental properties in Miami area and stored in csv file for easier use helping to save processing time.
- Miami map showing prices and cluster of venues is created o help ease in exploring various possibilities.

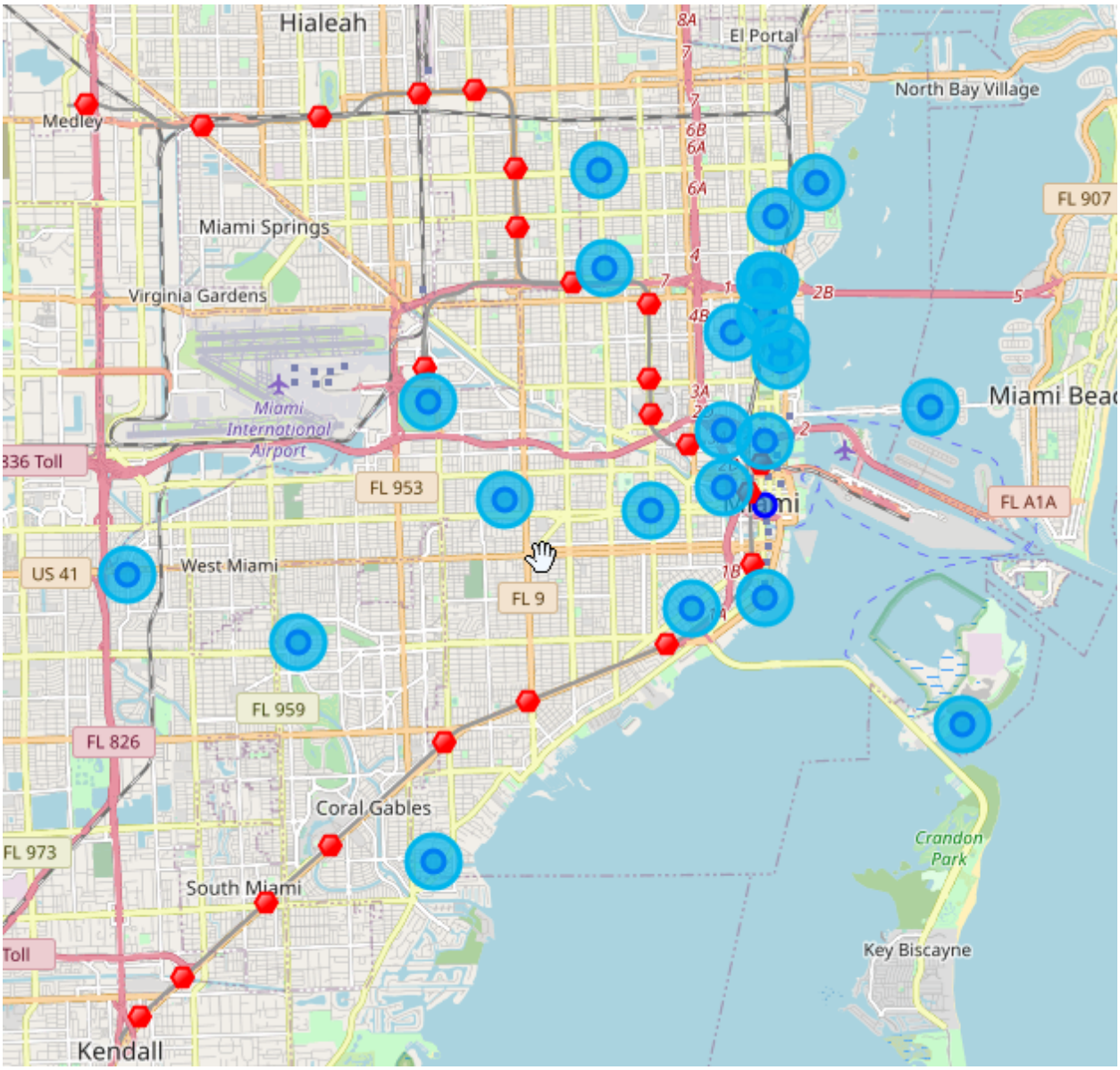


4.0 Results

4.1. Result and Discussion

Map is consolidated with all required information to help with apartment selection. Map of Miami with apartment rentals, metro location and venue cluster.

Red dots are Subway stations, Blue dot are apartment available for rent, Bubbles are the clusters of venues.



```
kk = 2
miami_merged.loc[miami_merged['Cluster Labels'] == kk, miami_merged.columns[[1] + list(range(5, miami_merged.shape[1]))]]
```

	County	Cluster Labels	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
4	Miami	2.0	Park	American Restaurant	Boat or Ferry	Event Space	Food Truck	Food	Flea Market	Fish Market	Fast Food Restaurant	Wings Joint
20	Miami	2.0	Park	Lounge	Boat or Ferry	Empanada Restaurant	Food	Flea Market	Fish Market	Fast Food Restaurant	Event Space	Wings Joint

Apartment 1 Locations

```
## kk is the cluster number to explore
kk = 3
miami_merged.loc[miami_merged['Cluster Labels'] == kk, miami_merged.columns[[1] + list(range(5, miami_merged.shape[1]))]]
```

	County	Cluster Labels	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
21	Miami	3.0	Beach	Wings Joint	French Restaurant	Dessert Shop	Diner	Discount Store	Dive Bar	Dog Run	Donut Shop	Empanada Restaurant

Apartment 2 Locations

Apartment Selection-Comparison analysis Using the "one consolidated map" above, I was able to explore all possibilities since the popups provided the information needed for a good data-driven decision.

The apartment in blue dot 113 Greenwich way is the best location in terms of rent, transportation and venue.

Based on current Maryland venues, I feel that Cluster 3 type of venues is a closer resemblance to my current place. That means that APARTMENT is a better choice and cheaper which means I can use it for other expenses.

5.0 DISCUSSION

5.1. Elaborate discussion

I believe that transportation and location both matter a lot. Having to spend 2000 USD a month in Maryland and enjoying life. This exercise was way helpful to figure out big life decisions as well as providing ease to figure out the option.

The complete Specialization Course is well structured with enough on hand lab work and assignments and helps to give insights to concepts learned during classes. It helped me to learn various tools and provided me with great knowledge.

This project provided an opportunity to apply and practice Data Science tools and skills acquired during learning process.

6.0 Conclusion

6.1. Conclusion

I am very happy to be able to complete the 9-course specialization on time and it was worth the time spent. It has provided with various skills and tools that will help to grow and build a career in Data Science.

Thank you for reviewing my work and thanks to the IBM/Coursera community for this course!