

## **Table of Content**

#### 1.Introduction Section:

- 1.1 Background
- 1.2 Business Problem
- 1.3 Purpose of Project

#### 2.Data Section:

- 2.1 Data description and the sources used to solve the business problem.
- 2.2 Comparison with Present Location
- 2.3 Data sources and data manipulation techniques
- 2.4 How it will help to evaluate the decision of renting

#### 3. Methodology section:

- 3.1 Process steps and strategy to resolve the problem
- 3.2 Data Science Methods, machine learning, mapping tools and exploratory data analysis
- 4. Results section
- 5. Discussion section
- 6.Conclusion section

## Introduction

#### 1.1 Background

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.

#### 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.

### **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.

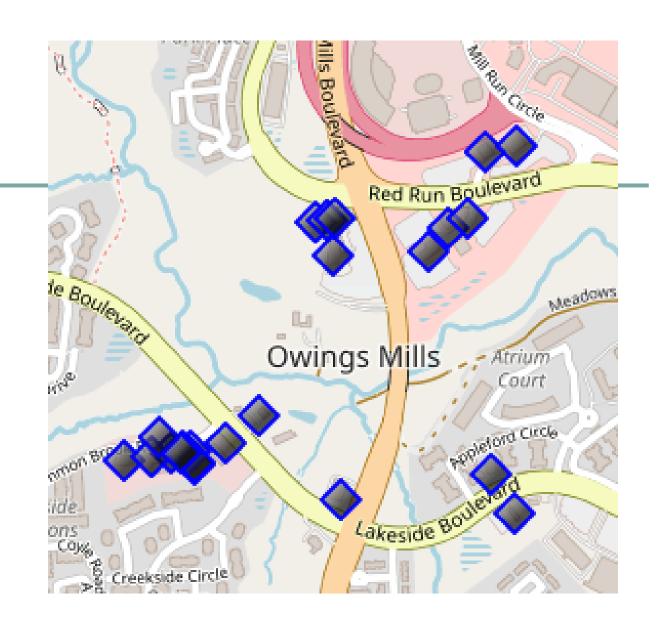
Preferably, a list of apartments for rent with additional information, such as price, address, area, # of beds

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	ιаτ	ıng
0	Taj Palace Owings MIIIs	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.



# 2.4 How it will help to evaluate the decision of renting.

- It will help to answer the below questions:
- what is the cost of rent (per square ft) around a mile radius from each Metrorail station?
- what is the area of Miami 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.



# 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.

# **Methodology Section**

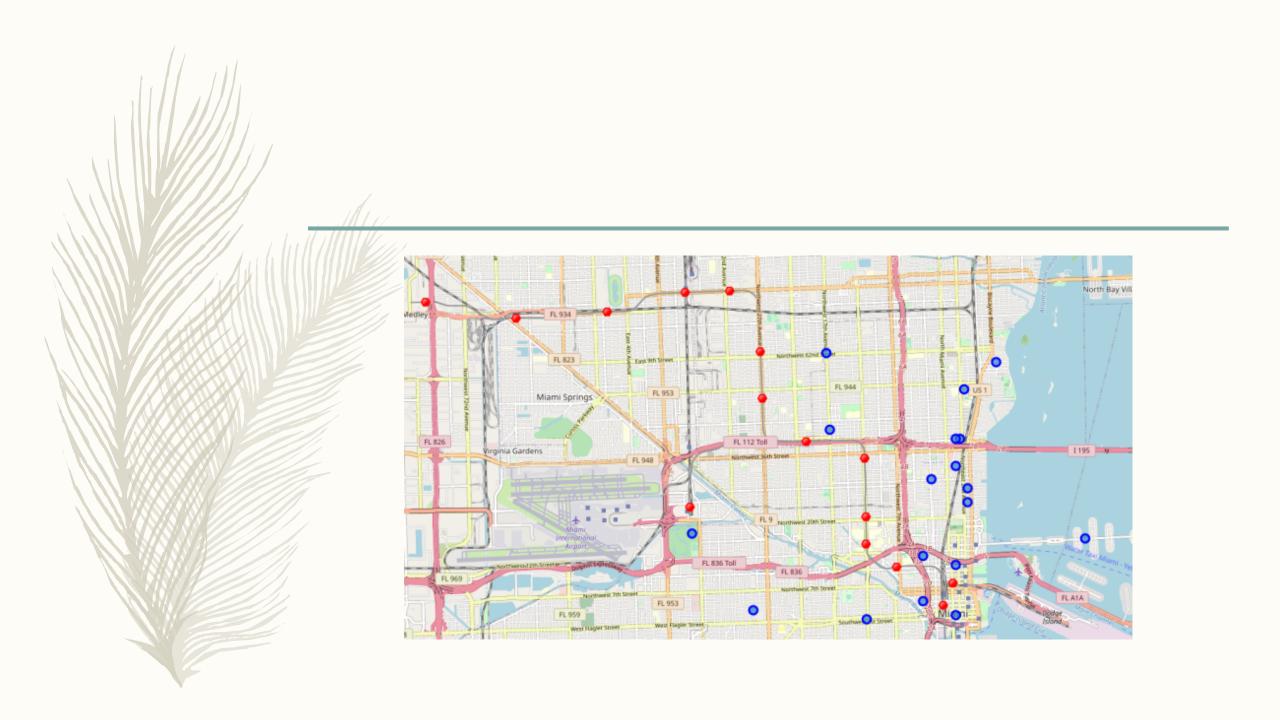
#### 3.1 Process steps and strategy to resolve the problem

- 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.



# 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.

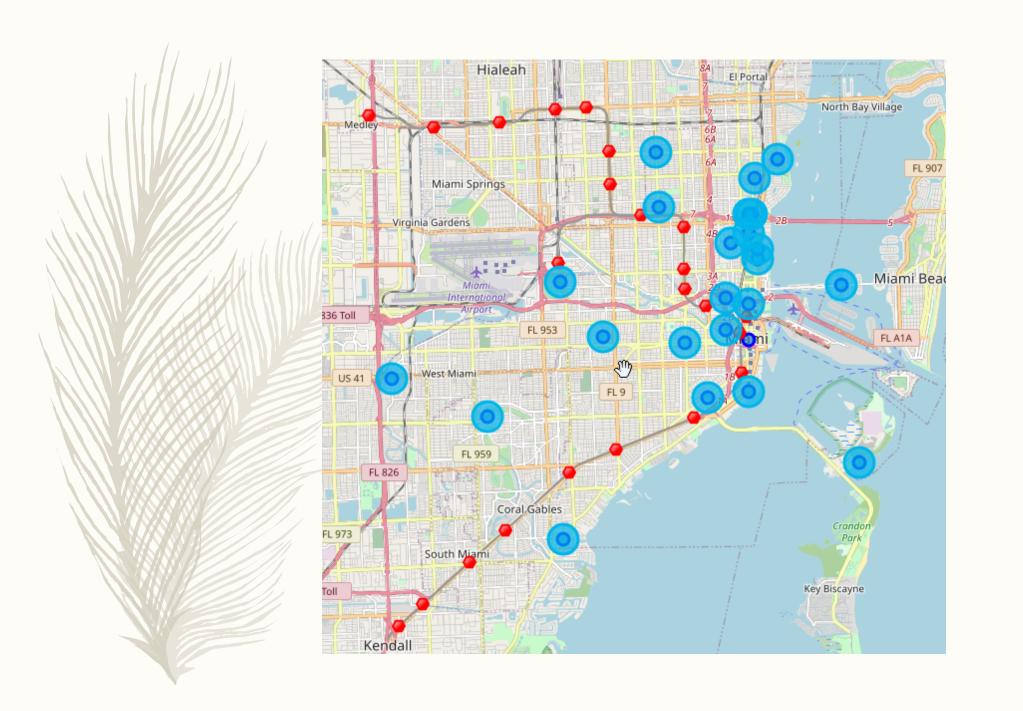




# Results

#### 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 = :

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**



# **DISCUSSION**

#### 5.1. Elaborate discussion

- 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.
- 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.



# 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!