IBM Data Science Professional Certificate

Author: Filipe Augusto Sá / March - 2020

Contents

Introduction Section:

1.1 Discussion of the "background situation" leading to the problem at hand:

1.2 Problem to be resolved

1.3 Audience for this project.

Data Section:

2.1 Data of Current Situation (current residence place)

2.2 Data required to resolve the problem

2.3 Data sources and data manipulation

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.

Results Section

Discussion of the results and how they help to take a decision.

Discussion Section

Elaboration and discussion on any observations and/or recommendations for improvement.

Conclusion Section

Decision taken and Report Conclusion.

Author: Filipe Augusto Sá / March - 2020

Powered By

Introduction Section

1.1 Scenario And Background :

I am a data scientist residing in Niterói, Rio de Janeiro, Brazil. I currently live within walking distance to many amenities and venues in the area, such as various international cuisine restaurants, cafes, gyms, food shops and entertainment.

I have been offered a great opportunity to work in Manhattan, NY. Although, I am very excited about it, I am a bit stress toward the process to secure a comparable place to live in Manhattan. Therefore, I decided to apply the learned skills during the Coursera course 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 rewarding doing it myself with learned tools.

1.2 Probleme To Be Resolved:

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

- Top amenities in the selected neighborhood shall be similar to current residence (See item 2.1)
- Desirable to have venues such as Gyms, Restaurants, wine stores, and food shops
- As a reference, I have included a map of venues near current residence in Niterói, Rio de Janeiro.

1.3 Interesed Audience:

I believe this is a relevant project for a person or entity 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 toward the development of Data Science skills.

Author: Filipe Augusto Sá / March - 2020

Powered By

Data Section

2.1 Data Of Current Situation:

I Currently reside in the neighborhood of Niterói in Rio de Janeiro, Brazil. I use Foursquare to identify the venues around the area of residence which are then shown in the Niterói map shown in methodology and execution in section 3.0 . It serves as a reference for comparison with the desired future location in Manhattan NY

2.2 Data Required To Resolve The Problem:

In order to make a good choice of a similar apartment in Manhattan NY, the following data is required:
- List/Information on neighborhoods form Manhattan with their Geodata(latitude and longitude).
- Venues and amenities in the Manhattan neighborhoods (e.g. top 10)

2.3 Sources And manipulation:

The list of Manhattan neighborhoods is worked out during LAB exercise during the course. A csv file was created which will be read in order to create a data frame and its mapping.

2.4 How the data will be used to solve the problem

The data will be used as follows:

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

2.5 Mapping Of Data

The following maps were created to facilitate the analysis and the choice of the palace to live.
- Manhattan map of Neighborhoods
- Manhattan map of clustered venues and neighborhoods

Author: Filipe Augusto Sá / March - 2020

Methodology Section

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 indicates 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 a candidate places for accommodation. The choice is made based on the demands imposed : similar venues to Niterói, Rio de Janeiro, Brazil. This visual approach and maps with popups labels allow quick identification of location, thus making the selection very easy.

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

- What are the venues of the best place to live?
- How venues distribute among Manhattan neighborhoods ?

Author: Filipe Augusto Sá / March - 2020
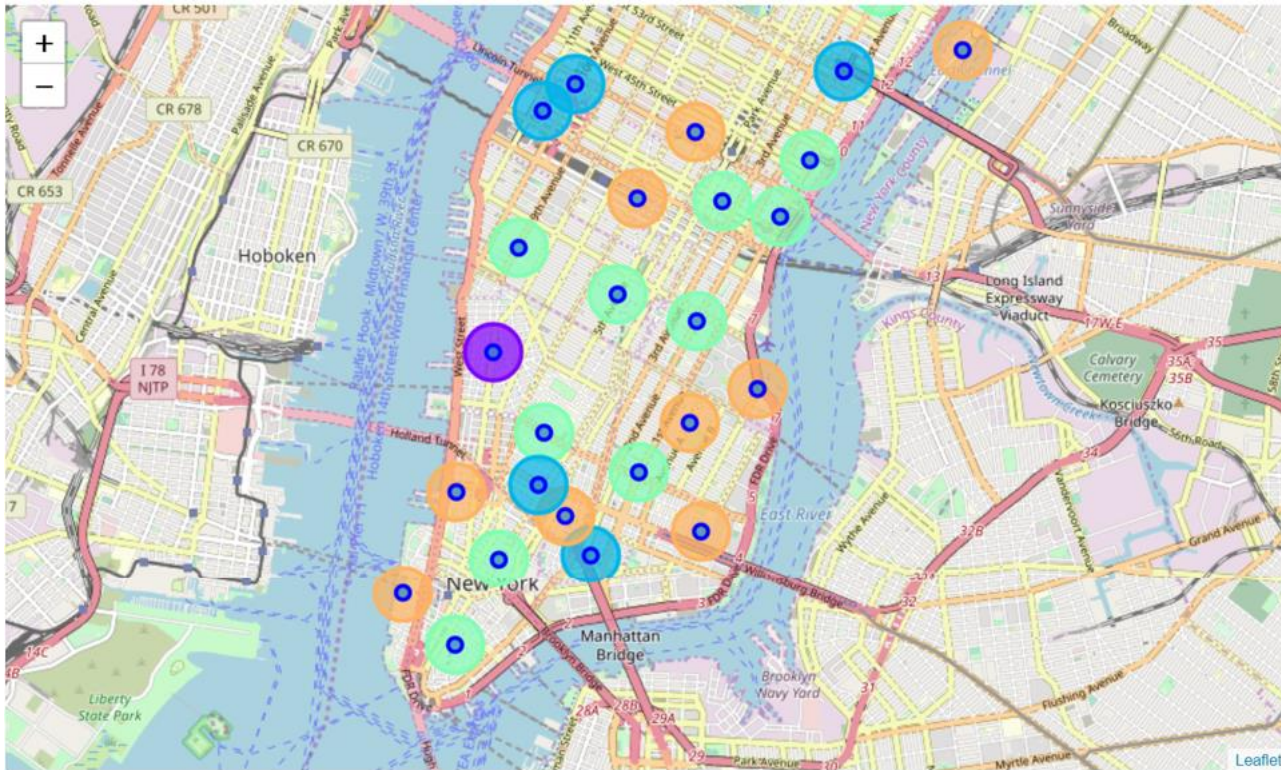
Result Section

Current Residence in Niterói
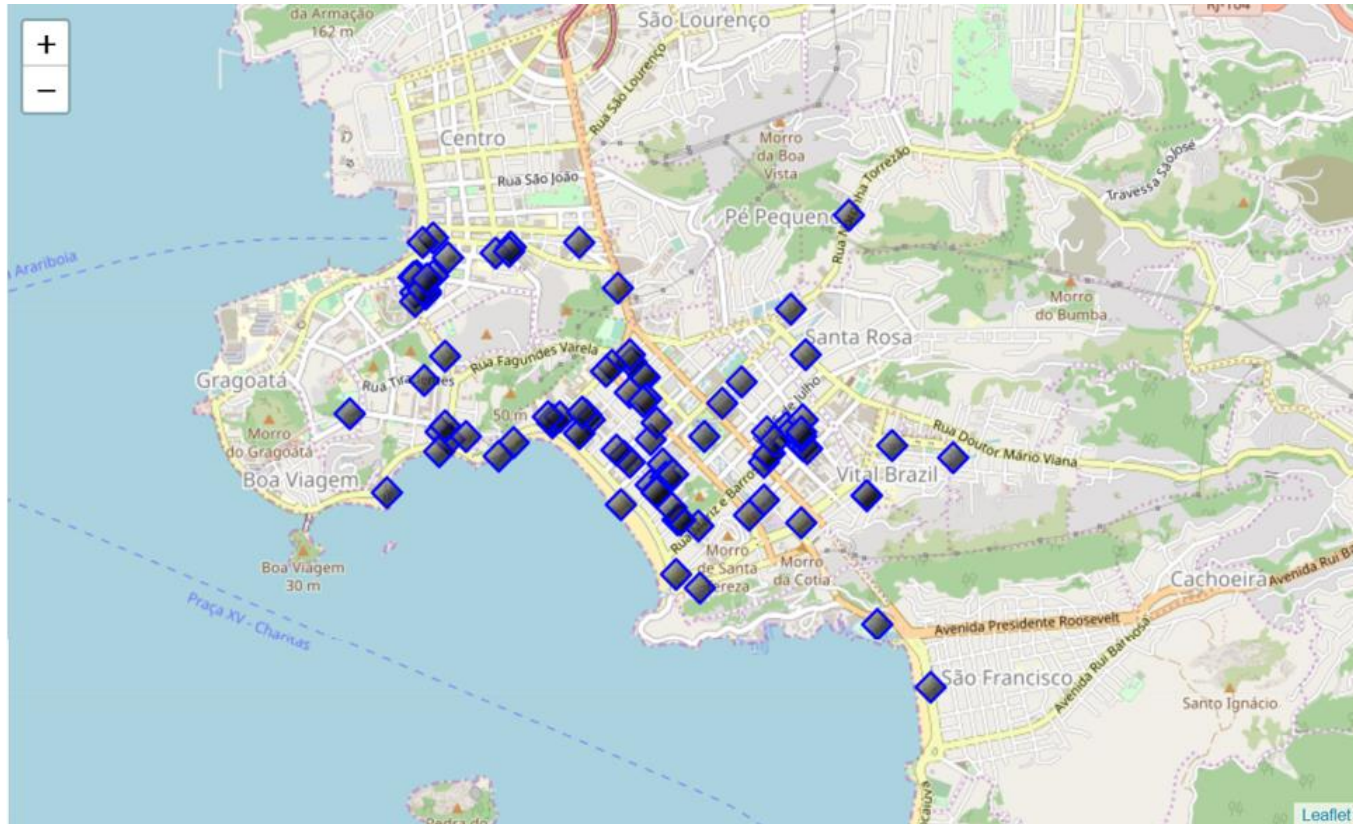

```
[10]:  # Venues near current Brazil residence place
       Nitnearby_venues.head(10)
```

	name	categories	lat	lng
0	Campo de São Bento	Park	-22.904364	-43.107562
1	Temakeria e Cia	Japanese Restaurant	-22.904560	-43.110627
2	Confeitaria Beira Mar	Bakery	-22.905444	-43.112172
3	Big Boss Bistrô & Grill	Steakhouse	-22.901284	-43.110953
4	Dice'n'Roll Coffee Tales	Coffee Shop	-22.902703	-43.106526
5	Squasso Centro de beleza e Spa	Salon / Barbershop	-22.905827	-43.109923
6	Porção Mágica	Dessert Shop	-22.903673	-43.110272
7	Kopenhagen	Chocolate Shop	-22.902128	-43.111771
8	Calçadão de Icaraí	Pedestrian Plaza	-22.904246	-43.114734
9	Crema & Cioccolato Café	Ice Cream Shop	-22.909079	-43.107834

Author: Filipe Augusto Sá / March - 2020

Powered By

Result Section

Clusters of Neighbourhoods in Manhattan

Author: Filipe Augusto Sá / March - 2020

Result Section

On Careful Examination, Financial District in Cluster 3 resembles in amenities to our Current Residence

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

[23]:

	Neighborhood	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
3	Inwood	Mexican Restaurant	Lounge	Pizza Place	Cafe	Wine Bar	Bakery	American Restaurant	Park	Frozen Yogurt Shop	Spanish Restaurant
5	Manhattanville	Deli / Bodega	Italian Restaurant	Seafood Restaurant	Mexican Restaurant	Sushi Restaurant	Beer Garden	Coffee Shop	Falafel Restaurant	Bike Trail	Other Nightlife
10	Lenox Hill	Sushi Restaurant	Italian Restaurant	Coffee Shop	Gym / Fitness Center	Pizza Place	Burger Joint	Deli / Bodega	Gym	Sporting Goods Shop	Thai Restaurant
12	Upper West Side	Italian Restaurant	Bar	Bakery	Vegetarian / Vegan Restaurant	Indian Restaurant	Coffee Shop	Cosmetics Shop	Wine Bar	Mexican Restaurant	Sushi Restaurant
16	Murray Hill	Sandwich Place	Hotel	Japanese Restaurant	Gym / Fitness Center	Coffee Shop	Salon / Barbershop	Burger Joint	French Restaurant	Bar	Italian Restaurant
17	Chelsea	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bakery	Nightclub	Theater	Art Gallery	Seafood Restaurant	American Restaurant	Hotel
18	Greenwich Village	Italian Restaurant	Sushi Restaurant	French Restaurant	Clothing Store	Chinese Restaurant	Cafe	Indian Restaurant	Bakery	Seafood Restaurant	Electronics Store
27	Gramercy	Italian Restaurant	Restaurant	Thrift / Vintage Store	Cocktail Bar	Bagel Shop	Coffee Shop	Pizza Place	Mexican Restaurant	Grocery Store	Wine Shop

Author: Filipe Augusto Sá / March - 2020

Venue Selection

- Using the "one map" above, I was able to explore all possibilities since the popups provide the information needed for a good decision.

- Financial District having Gyms, Hotels and Restaurants like Niterói residence is my preferable choice for a future residence.

- Based on current Niterói venue, I feel that Cluster 3 type of venues is a closer resemblance to my current place. That means that Financial District is a better choice since it has Gyms, Hotels and Restaurant and is worth the conveniences it provides.

Author: Filipe Augusto Sá / March - 2020

Powered By

Discussion

- In general, I am positively impressed with the overall organization, content and lab works presented during the Coursera IBM Certification Course

- I feel this Capstone project presented me a great opportunity to practice and apply the Data Science tools and methodologies learned.

- I have created a good project that I can present as an example to show my potential.

- I feel I have acquired a good starting point to become a professional Data Scientist and I will continue exploring to creating examples of practical cases

Author: Filipe Augusto Sá / March - 2020

Powered By

Conclusion

- I feel rewarded with the efforts, time and money spent. I believe this course with all the topics covered is well worthy of appreciation.

- This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools.

- The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision thoroughly and with confidence. I would recommend for use in similar situations.

- One must keep abreast of new tools for Data Science that continue to appear for application in several business fields.

Author: Filipe Augusto Sá / March - 2020