

Julian

Instructor Alex Aklson

Coursera IBM Prof Cert

14. April 2020

Capstone Project: Best Location for my Italian Restaurant

Introduction

Italian food has conquered more nations – or at least their palates – than the Roman Empire ever could have hoped to militarily! Italian restaurants are found all over the world from Tokyo to Toronto.

Opening a restaurant, especially in NYC where we already have many, involves great research and preparation. Therefore I will consider many factors to find the best neighborhood for my new Italian restaurant. For example how many other top rated (Italian) restaurants are around or what other venues are nearby. I am Julian and welcome to my own little culinary empire.

Method

First we import data from a JSON file and display different neighborhoods in Lower Manhattan. The next task is essentially transforming this data of nested Python dictionaries into a *pandas* dataframe and looping through the data and filling the dataframe one row at a time

I used python **folium** library to visualize geographic details of

Figure 1: Lower Manhattan Neighborhoods



As a next step we will use the Foursquare API and import data about top rated venues around our neighborhoods. We then will filter the data and create a data frame that contains the number of venues by type of venue for every neighborhood. I utilized the Foursquare API to explore the neighbourhoods and segment them. I designed the limit as **100 venues** and the radius **500 meters** for each neighborhood from their given latitude and longitude informations.

Through filtering and extracting

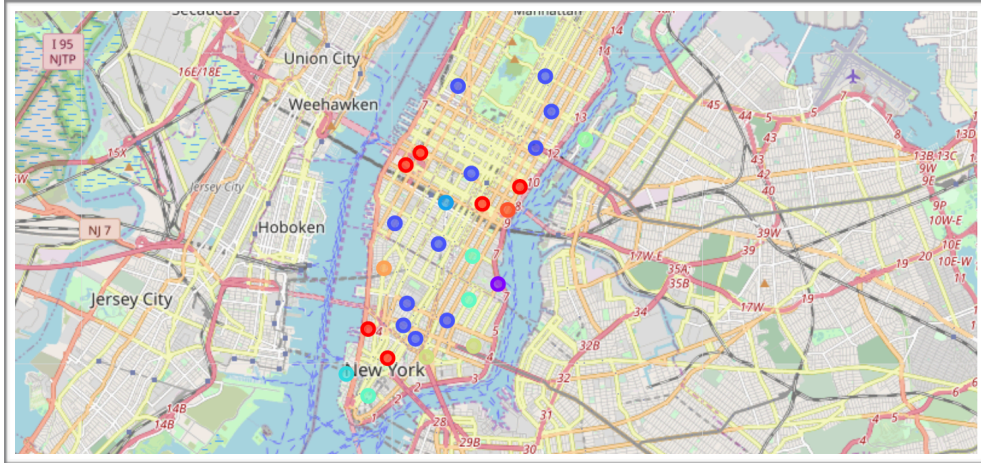
We use this data in combination with a score matrix to get a weighted score based on our requirements for the perfect restaurant environment.

The algorithm works similar to a recommender system. Input it an array with all the information about the importance of a venue for my restaurant. Some criteria was: there should be as little as possible top rated Italian restaurants around, but many other important, helpful and synergetic top rated venues (like theaters, bars or boutiques).

	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
1	Upper East Side	Coffee Shop	Italian Restaurant	Bakery	Exhibit	Yoga Studio	Spa	Juice Bar	French Restaurant
2	Lenox Hill	Italian Restaurant	Coffee Shop	Pizza Place	Sushi Restaurant	Cocktail Bar	Gym / Fitness Center	Gym	Burger Joint
4	Lincoln Square	Gym / Fitness Center	Café	Plaza	Concert Hall	Performing Arts Venue	Theater	Wine Shop	American Restaurant
6	Midtown	Coffee Shop	Hotel	Steakhouse	Sporting Goods Shop	Clothing Store	Sandwich Place	Bakery	Theater
8	Chelsea	Coffee Shop	Art Gallery	American Restaurant	Italian Restaurant	Bakery	Cycle Studio	Cupcake Shop	Seafood Restaurant
9	Greenwich Village	Italian Restaurant	Café	Coffee Shop	Gym	Sushi Restaurant	French Restaurant	Dessert Shop	Gourmet Shop
13	Little Italy	Bakery	Café	Chinese Restaurant	Bubble Tea Shop	Hotel	Hotpot Restaurant	Italian Restaurant	Mediterranean Restaurant
14	Soho	Coffee Shop	Italian Restaurant	Café	Bakery	Sandwich Place	Clothing Store	Mediterranean Restaurant	French Restaurant
19	Noho	Italian Restaurant	Coffee Shop	Pizza Place	Grocery Store	French Restaurant	Cocktail Bar	Japanese Restaurant	Hotel
22	Sutton Place	Coffee Shop	Italian Restaurant	Gym / Fitness Center	Gym	Park	Furniture / Home Store	Pizza Place	Yoga Studio
26	Flatiron	Gym / Fitness Center	New American Restaurant	Café	Yoga Studio	Cosmetics Shop	Mediterranean Restaurant	Italian Restaurant	Gym

Figure 2: Neighborhoods and venues

By clustering the neighborhoods with k-means we can find similar ones. The matches we found an can further compare, evaluate and discuss the used model.

**Figure 3: Clustered Neighborhoods**

Results

Based on my algorithm the result for the best neighbourhood is “Stuyvesant Town” with a score of 0.470588 . Followed by “East Village” and “Sutton Place”.

Neighborhood	
0	Stuyvesant Town
1	East Village
2	Sutton Place
3	Battery Park City
4	Gramercy

Discussion

By plotting the factors of the Neighbourhood Score we can see that the number of values is not always optimally balanced. Also in case of “Stuyvesant Town” (0). We can argue that not all factors came into account. Therefore the result diversity can be taken into account. If that’s the case we could select Sutton Place (3) as our location. By en-

larging the volume of data, improving the score algorithm or by adding different data a more precise picture could be made.

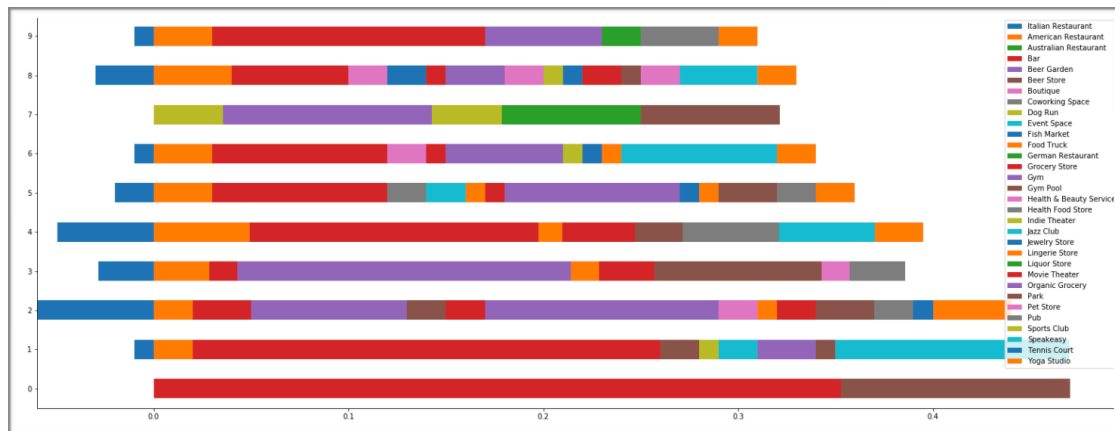


Figure 4: "Neighbourhood Score" stacked by venue type

Conclusion

The results show which neighborhood matches my needs best. This information will give me an orientation about where to locate my new business for the best profits. We can see that even for small businesses we can make decisions based on data. Data Visualisation and plotting maps is a great way to understand the Data given here.

I hope you enjoyed this little research. Thank you for your reading!