Finding the best place in New York, Manhattan to open an Italian restaurant

COURSERA CAPSTONE PROJECT

A comment to read, please:

- If you cannot open my notebook in github, please, try it here

- https://nbviewer.jupyter.org/github/anasonrisa/couursera_capstone_final_exam/blob/main/Attempt%20final%203%20-%20restaurant%20in%20NY.ipynb

Introduction/Business Problem (1/2)

I decided to look into the business question of opening the restaurant in New York. New York is the most populous city in the United States, with many various generations and nations living in, and thus is a great opportunity for a restaurant opening. As a restaurant opening is quite a wide area, I decided to limit it by:

- Limiting it by focusing on Italian restaurants only (it is as well a personal touch as I am a fun of Italian cuisine). Italian population is around 2,5 mln., so quite a big chunk of total population.
- Limiting the area I would explore by Manhattan, which is a densely populated borough that's among the world's major commercial, financial and cultural centers. So, this is a sign we have a population with a good salaries, in particular, which is good for a restaurant opening.

Introduction/Business Problem (2/2)

Who can be interested in this analysis?

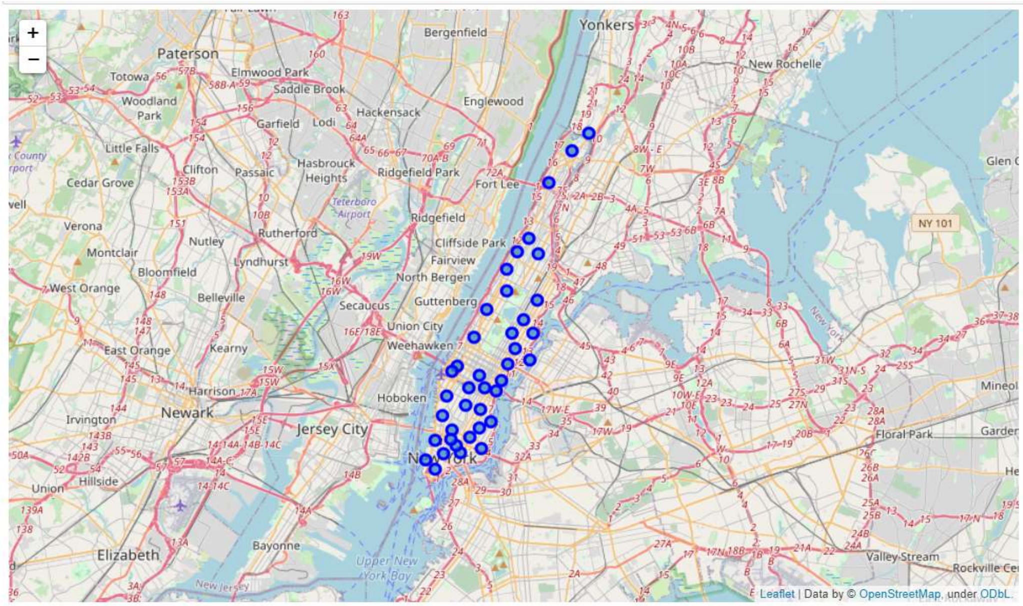
- Entrepreneurs searching for an opportunity in NY
- Government officials of NY city trying to understand the potential of their city better
- Data scientist exploring the ways to analyze the data

Data section: data used and data sources

- For the New York city I leveraged the data shared in one of the labs

Map of Manhattan

Out[17]:



Methodology section

- First of all, I used the conversion of addresses into their equivalent **latitude and longitude** values.
- Also, I used the **Foursquare API** to explore neighborhoods in New York City, in particular Manhattan.
- I used the **explore function** to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. **K-means clustering algorithm** was chosen to cluster the neighborhoods to complete this task.
- To visualize the results I used **Folium library**

Neighborhoods analysis 1/3:

Top-line venue analysis

Out[29]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Battery Park City	78	78	78	78	78	78
Carnegie Hill	96	96	96	96	96	96
Central Harlem	47	47	47	47	47	47
Chelsea	100	100	100	100	100	100
Chinatown	100	100	100	100	100	100
Civic Center	100	100	100	100	100	100
Clinton	100	100	100	100	100	100
East Harlem	38	38	38	38	38	38
East Village	100	100	100	100	100	100
Financial District	100	100	100	100	100	100
Flatiron	100	100	100	100	100	100
Gramercy	99	99	99	99	99	99
Greenwich Village	100	100	100	100	100	100
Hamilton Heights	60	60	60	60	60	60
Hudson Yards	73	73	73	73	73	73
Inwood	54	54	54	54	54	54
Lenox Hill	100	100	100	100	100	100
Lincoln Square	99	99	99	99	99	99
Little Italy	100	100	100	100	100	100
Lower East Side	47	47	47	47	47	47

Neighborhoods analysis 2/3:

Detailed venue analysis

Out[33]:

	Neighborhood	Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auditorium
0	Battery Park City	0.000000	0.000000	0.00000	0.012821	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.012821	0.012821
1	Carnegie Hill	0.000000	0.000000	0.00000	0.010417	0.00	0.010417	0.000000	0.020833	0.000000	0.000000	0.000000	0.000000
2	Central Harlem	0.000000	0.000000	0.06383	0.042553	0.00	0.000000	0.042553	0.000000	0.000000	0.000000	0.000000	0.000000
3	Chelsea	0.000000	0.000000	0.00000	0.040000	0.00	0.000000	0.050000	0.000000	0.000000	0.010000	0.000000	0.000000
4	Chinatown	0.000000	0.000000	0.00000	0.040000	0.00	0.000000	0.000000	0.000000	0.000000	0.020000	0.000000	0.000000
5	Civic Center	0.000000	0.000000	0.00000	0.030000	0.01	0.000000	0.010000	0.000000	0.000000	0.010000	0.000000	0.000000
6	Clinton	0.000000	0.000000	0.00000	0.050000	0.00	0.000000	0.010000	0.000000	0.000000	0.000000	0.000000	0.000000
7	East Harlem	0.000000	0.000000	0.00000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	East Village	0.000000	0.000000	0.00000	0.010000	0.00	0.010000	0.010000	0.000000	0.010000	0.000000	0.000000	0.000000
9	Financial District	0.010000	0.000000	0.00000	0.040000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	Flatiron	0.000000	0.000000	0.00000	0.040000	0.00	0.000000	0.010000	0.000000	0.010000	0.000000	0.000000	0.000000
11	Gramercy	0.000000	0.000000	0.00000	0.030303	0.00	0.000000	0.010101	0.000000	0.000000	0.000000	0.000000	0.000000
12	Greenwich Village	0.010000	0.000000	0.00000	0.020000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
13	Hamilton Heights	0.000000	0.016667	0.00000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
14	Hudson Yards	0.000000	0.000000	0.00000	0.068493	0.00	0.000000	0.013699	0.000000	0.000000	0.000000	0.000000	0.000000

Neighborhoods analysis 3/3:

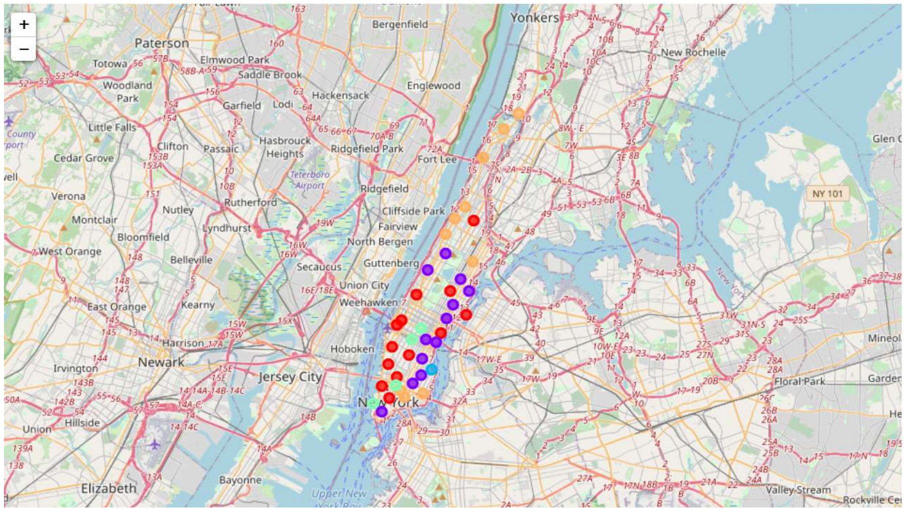
I found the most common places for each:

```
----Battery Park City----
             venue  freq
0             Park  0.10
1      Coffee Shop  0.06
2            Hotel  0.06
3   Clothing Store  0.05
4    Memorial Site  0.04

----Carnegie Hill----
         venue  freq
0  Coffee Shop  0.07
1         Café  0.06
2    Wine Shop  0.04
3  Yoga Studio  0.03
4    Bookstore  0.03

----Central Harlem----
                 venue  freq
0   African Restaurant  0.06
1   Seafood Restaurant  0.06
2   Chinese Restaurant  0.04
3  Gym / Fitness Center  0.04
4           Public Art  0.04
```

Results

I used K-means clustering to create clusters. The result of my work, we can see that there are 5 clusters in Manhattan.

- 3 of them include many boroughs, one has just one and another one is a middle size.

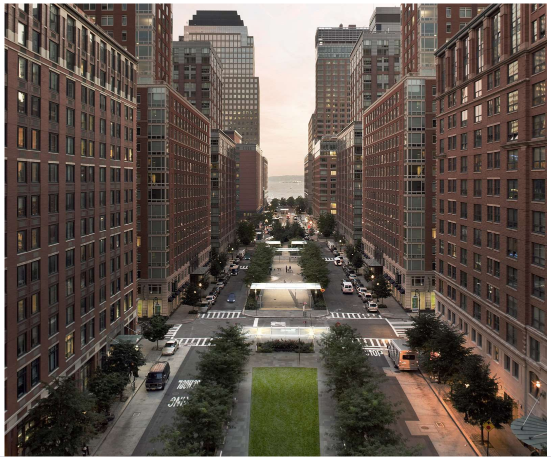


Discussion section

- Let us explore the clusters a bit. 1 and 2 has many restaurants inside and moreover quite a lot of Italian restaurants. So I do not see them as a great opportunity to start our business.
- Cluster 3 is too small to consider to get a good revenue.
- **While cluster 4 and 5 seem to be the most interesting for me.**
  - Cluster 4 does not have many restaurants / café in as for now – so we can be a “new category opener” and give people there an opportunity to try the Italian cuisine.
  - Cluster 5 is already developed as a restaurant / café center, and does not have many Italian restaurants , which gives us a great opportunity to take this place.

Cluster 4 details: "**grow it up!**"

Cluster 4 does not have many restaurants / café in as for now – so we can be a "new category opener" and give people there an opportunity to try the Italian cuisine.



Cluster 4

```
In [44]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 3, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1
```

Out[44]:

	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
15	Midtown	Hotel	Coffee Shop	Theater	Sporting Goods Shop	Clothing Store	American Restaurant	Bookstore	Steakhouse	Indian Restaurant	Gym
23	Soho	Clothing Store	Italian Restaurant	Boutique	Bakery	Mediterranean Restaurant	Coffee Shop	Women's Store	Shoe Store	Art Gallery	Dessert Shop
28	Battery Park City	Park	Coffee Shop	Hotel	Clothing Store	Memorial Site	Gym	Playground	Sandwich Place	Food Court	Pizza Place
33	Midtown South	Korean Restaurant	Hotel	Cosmetics Shop	Japanese Restaurant	Gym / Fitness Center	Dessert Shop	American Restaurant	Clothing Store	Hotel Bar	Coffee Shop

Cluster 5: “**compete and win!**”

Cluster 5 is already developed as a restaurant / café center, and does not have many Italian restaurants , which gives us a great opportunity to take this place.

Cluster 5

In [45]:
```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 4, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1
```

Out[45]:

	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
0	Marble Hill	Gym	Coffee Shop	Sandwich Place	Yoga Studio	Department Store	Supplement Shop	Steakhouse	Seafood Restaurant	Deli / Bodega	Diner
1	Chinatown	Chinese Restaurant	Bakery	American Restaurant	Cocktail Bar	Hotpot Restaurant	Dessert Shop	Salon / Barbershop	Optical Shop	Spa	Ice Cream Shop
2	Washington Heights	Café	Bakery	Mobile Phone Shop	Grocery Store	New American Restaurant	Gym	Coffee Shop	Park	Bank	Sandwich Place
3	Inwood	Mexican Restaurant	Café	Restaurant	Bakery	Chinese Restaurant	Deli / Bodega	Caribbean Restaurant	Pizza Place	Lounge	Wine Bar
4	Hamilton Heights	Pizza Place	Coffee Shop	Café	Mexican Restaurant	Deli / Bodega	Yoga Studio	Bakery	Liquor Store	Indian Restaurant	Park
5	Manhattanville	Coffee Shop	Deli / Bodega	Chinese Restaurant	Italian Restaurant	Bar	Mexican Restaurant	Lounge	Park	Sushi Restaurant	Bus Station
7	East Harlem	Thai Restaurant	Mexican Restaurant	Bakery	Latin American Restaurant	Sandwich Place	Deli / Bodega	Park	Historic Site	Seafood Restaurant	Taco Place
20	Lower East Side	Chinese Restaurant	Pizza Place	Art Gallery	Coffee Shop	Bakery	Ramen Restaurant	Café	Grocery Store	Park	Speakeasy
22	Little Italy	Bakery	Italian Restaurant	Ice Cream Shop	Café	Mediterranean Restaurant	Coffee Shop	Tea Room	Sandwich Place	Salon / Barbershop	Chinese Restaurant
26	Morningside Heights	Coffee Shop	American Restaurant	Park	Bookstore	Café	Burger Joint	Deli / Bodega	Grocery Store	Pharmacy	Seafood Restaurant

Conclusion

- This reports gave an overview of the project of finding the best place to open an Italian restaurant in New York, Manhattan. We looked at the data used, analysis types, as well as the results and observations.

Observations:

- I chosen to have 5 clusters in this exercise. The result was good enough to make a statement, but I believe it might be good to try to create less (3-4) or more (7-8) clusters to see how it can help resolve the problem.

Thank you!