

Data Science Capstone Project

Data Driven Decision for the best place to open a restaurant in New York

TOC

- Introduction
 - Problem Description and background
 - Target Audience
- Data
 - Sources
 - Usage
- Methodology
 - EDA
 - Inferential statistical testing
 - ML Algorithms
- Notebook
- Results
- Discussions
- Conclusions

Introduction

Introduction

- Prepared for IBM Data Science Capstone Project - The Battle of Neighborhoods
- Use techniques learned in the course for
 - Exploring and comparing geographical location
 - Leveraging the Foursquare APIs
- To lead to a data driven suggestion for opening a new restaurant in New York

Target Audience

Target Audience

- Target Audience for this presentation are the potential Restaurateurs who would like to open a new restaurant
- It can also be used by potential investors for gauging the popularity of a particular restaurant
- Sample Template for Data Scientists for Exploratory Data Analysis techniques

Problem Description

Introduction

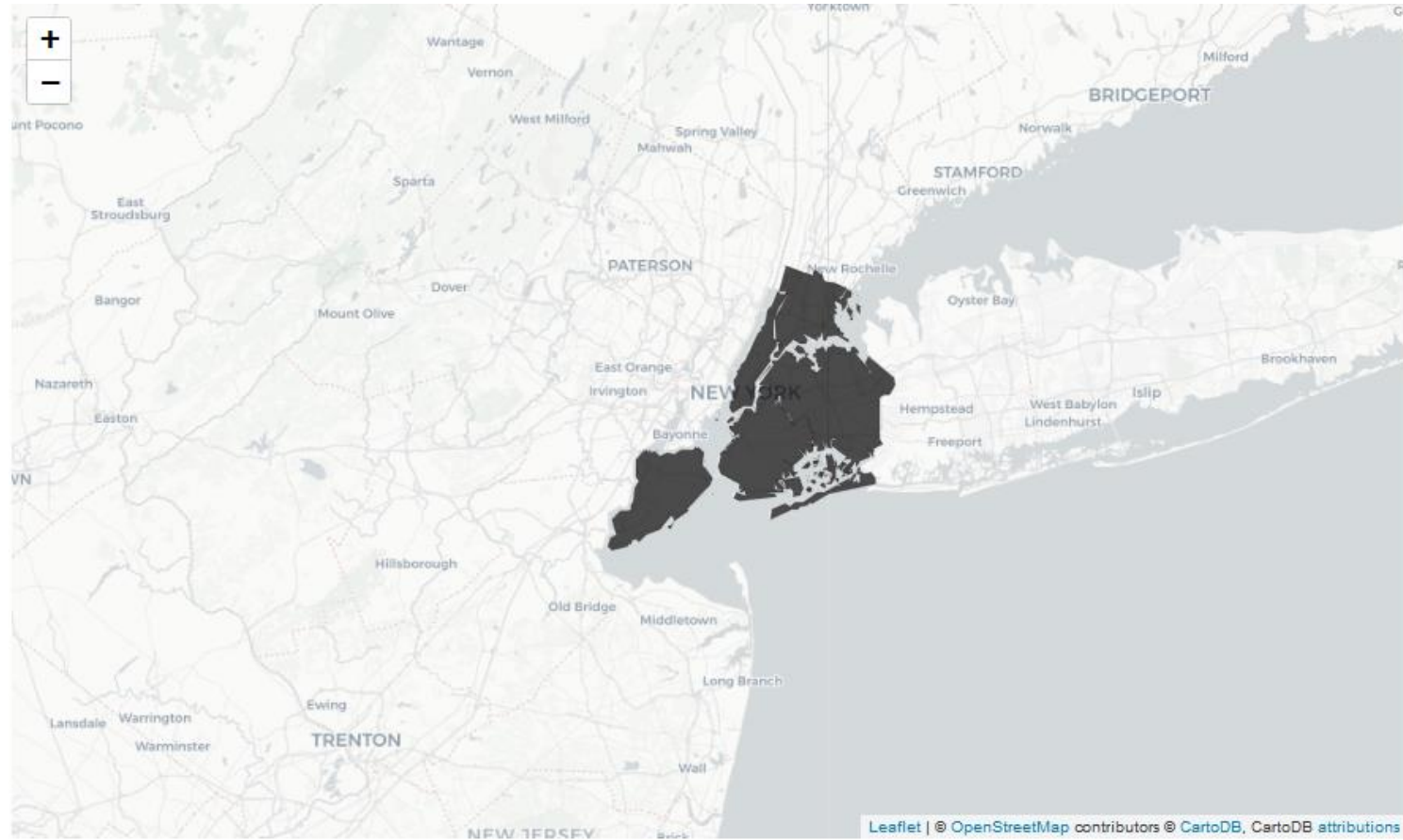
- Opening a new restaurant in New York
 - Concept
 - Target Market - Identify gaps in cuisine, service.
 - Competitors – Who are the competitors?
 - Location
 - Where ? – Depends upon the Target Market, Business Model and competitors
 - Parking? – Without adequate parking, people may not come
 - Visibility? – Is the location at a prominent place?

Data Sources

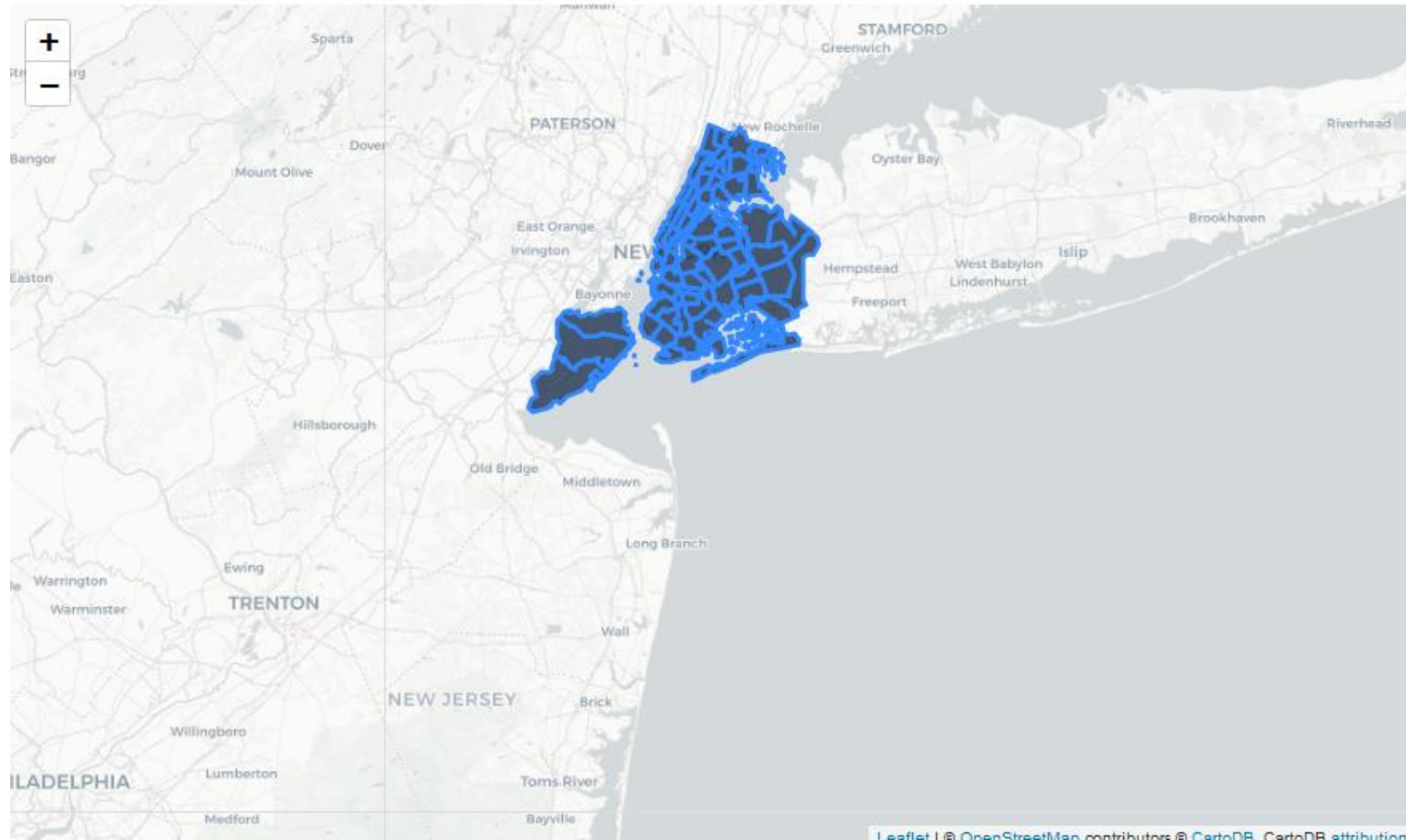
New York

- Divided Into 5 boroughs
- Further each borough is divided into Community Districts
- Each community districts has zoning areas

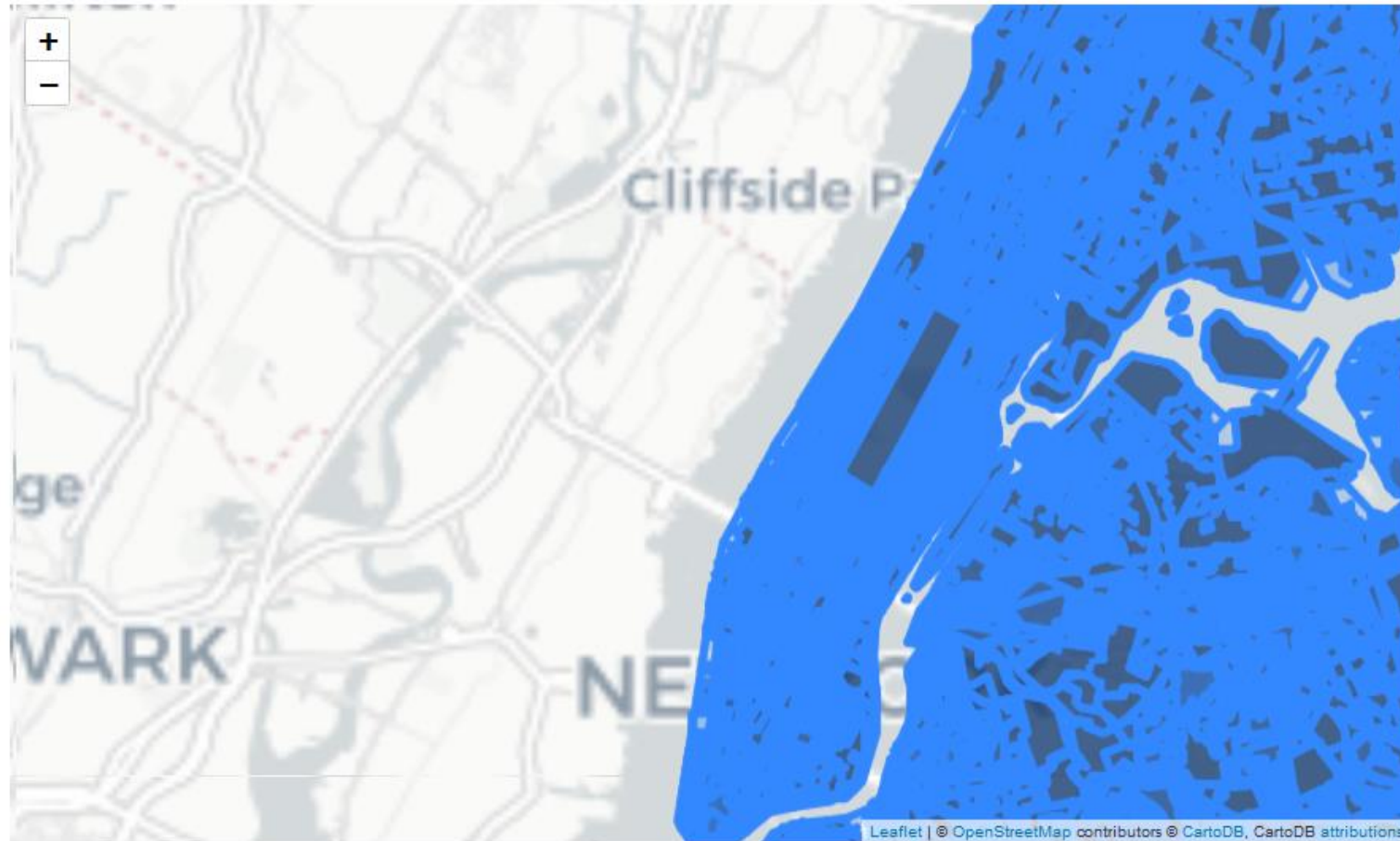
Boroughs



Community Districts



Zoning Areas



Why Finding the Population Base Is Important - Target Market

- You might assume that you know enough about a particular location if you're opening in your own hometown or a local neighbourhood. You might think that doing any further research is unnecessary. This can be a costly mistake.
- Really digging into the numbers in your area, from housing values, average household income to average age, then contrasting this information with the number of competing restaurants in the area, will help you determine the best type of restaurant to open.
- You might tailor your concept to millennial customers rather than a baby boomer population if the average age in the area is 27.
- Data Source
 - https://en.wikipedia.org/wiki/Economy_of_New_York_City

Population Base

Subject	Bronx	Brooklyn	Manhattan	Queens	Staten Island	New York city
SEX AND AGE	NaN	NaN	NaN	NaN	NaN	NaN
Under 5 years	107454.0	193851.0	82024.0	144634.0	27420.0	555383.0
5 to 9 years	100908.0	166770.0	62937.0	127945.0	29083.0	487643.0
10 to 14 years	98242.0	154230.0	59514.0	123991.0	30516.0	466493.0
15 to 19 years	103133.0	150044.0	72486.0	124342.0	29923.0	479928.0
20 to 24 years	121897.0	188262.0	120405.0	158136.0	32042.0	620742.0
25 to 34 years	221024.0	468927.0	370713.0	375623.0	61464.0	1497751.0
35 to 44 years	183782.0	360926.0	239951.0	328408.0	61738.0	1174805.0
45 to 54 years	189692.0	320318.0	206583.0	325624.0	69473.0	1111690.0
55 to 59 years	80591.0	150027.0	93290.0	154210.0	32894.0	511012.0
60 to 64 years	68139.0	135922.0	91914.0	131245.0	29964.0	457184.0

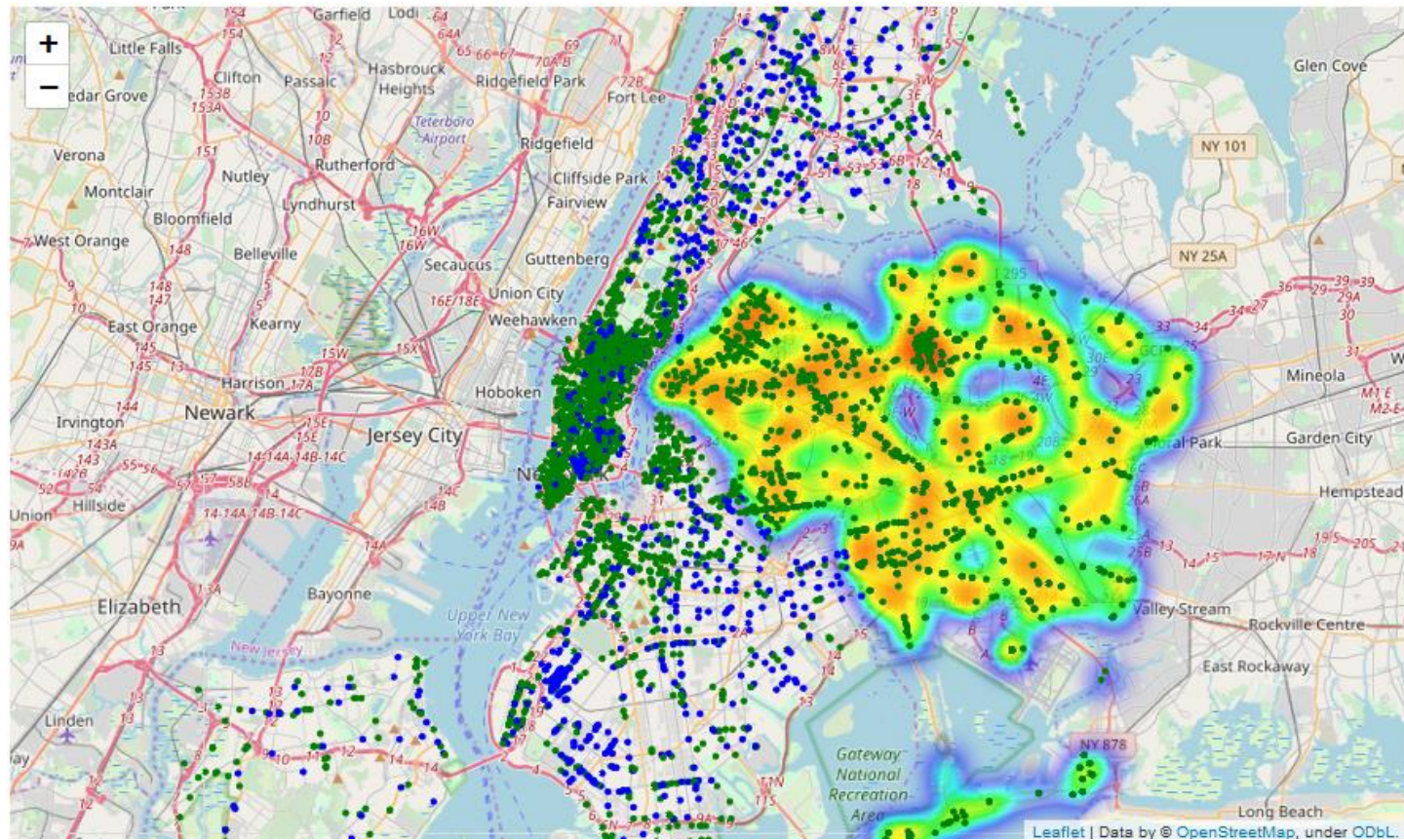
Competition - Know Your Neighbours

- When looking for a restaurant location, consider who else is doing business in the neighbourhood. Are there already half a dozen restaurants with the same concept as yours? Is the area busy or full of empty storefronts? Successful businesses attract other successful businesses.
- Data Source
 - <https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/rs6k-p7g6>
 - https://geo.nyu.edu/catalog/nyu_2451_34572

BORO	BUILDING	STREET	ZIPCODE	PHONE	CUISINE DESCRIPTION	INSPECTION DATE	ACTION	...	RECORD DATE	INSPECTION TYPE	Latitude	Longitude	Community Board
Bronx	1007	MORRIS PARK AVE	10462.0	7188924968	Bakery	06/11/2019	Violations were cited in the following area(s).	...	11/11/2019	Cycle Inspection / Re- inspection	40.848231	-73.855972	211.0
Brooklyn	469	FLATBUSH AVENUE	11225.0	7182875005	Hamburgers	03/13/2018	Violations were cited in the following area(s).	...	11/11/2019	Cycle Inspection / Initial Inspection	40.662652	-73.962081	309.0
Manhattan	351	WEST 57 STREET	10019.0	2122452912	Irish	12/14/2017	Violations were cited in the following area(s).	...	11/11/2019	Cycle Inspection / Re- inspection	40.767326	-73.984310	104.0
Brooklyn	2780	STILLWELL AVENUE	11224.0	7183723031	American	07/20/2017	Violations were cited in the following area(s).	...	11/11/2019	Cycle Inspection / Initial Inspection	40.579920	-73.982090	313.0
Brooklyn	7114	AVENUE U	11234.0	7184443838	Delicatessen	06/03/2017	Violations were cited in the following area(s).	...	11/11/2019	Cycle Inspection / Initial Inspection	40.620112	-73.906989	318.0
...
Queens	17933	HILLSIDE AVE	11432.0	3475619139	American	11/04/2019	Violations were cited in the following area(s).	...	11/11/2019	Pre-permit (Operational) / Initial Inspection	40.712768	-73.783322	408.0

CUISINE DESCRIPTION

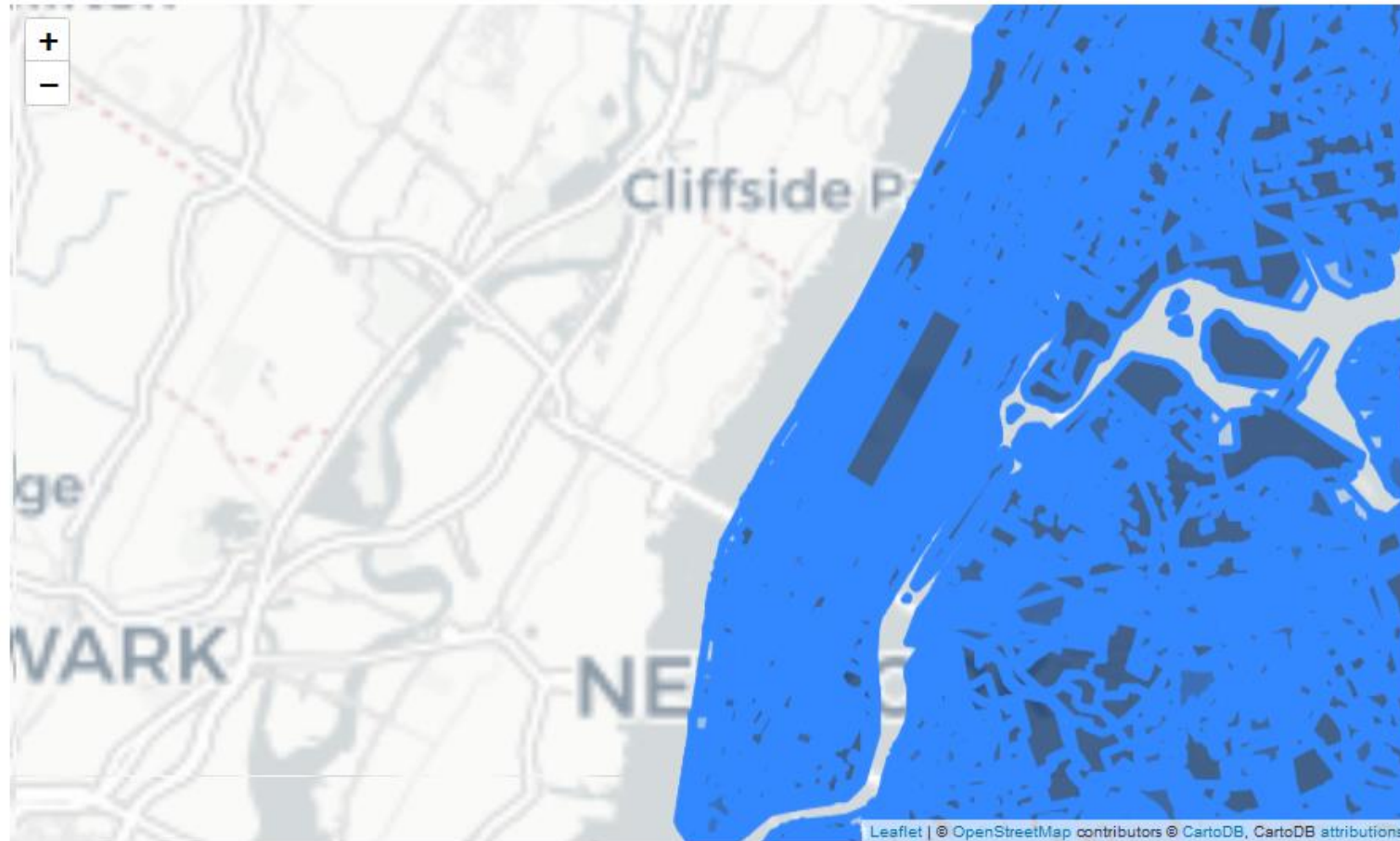
Afghan	10
African	32
American	3374
Armenian	16
Asian	192
Australian	14
Bagels/Pretzels	100
Bakery	459
Bangladeshi	19
Barbecue	25



Location - Zoning Ordinances – Where?

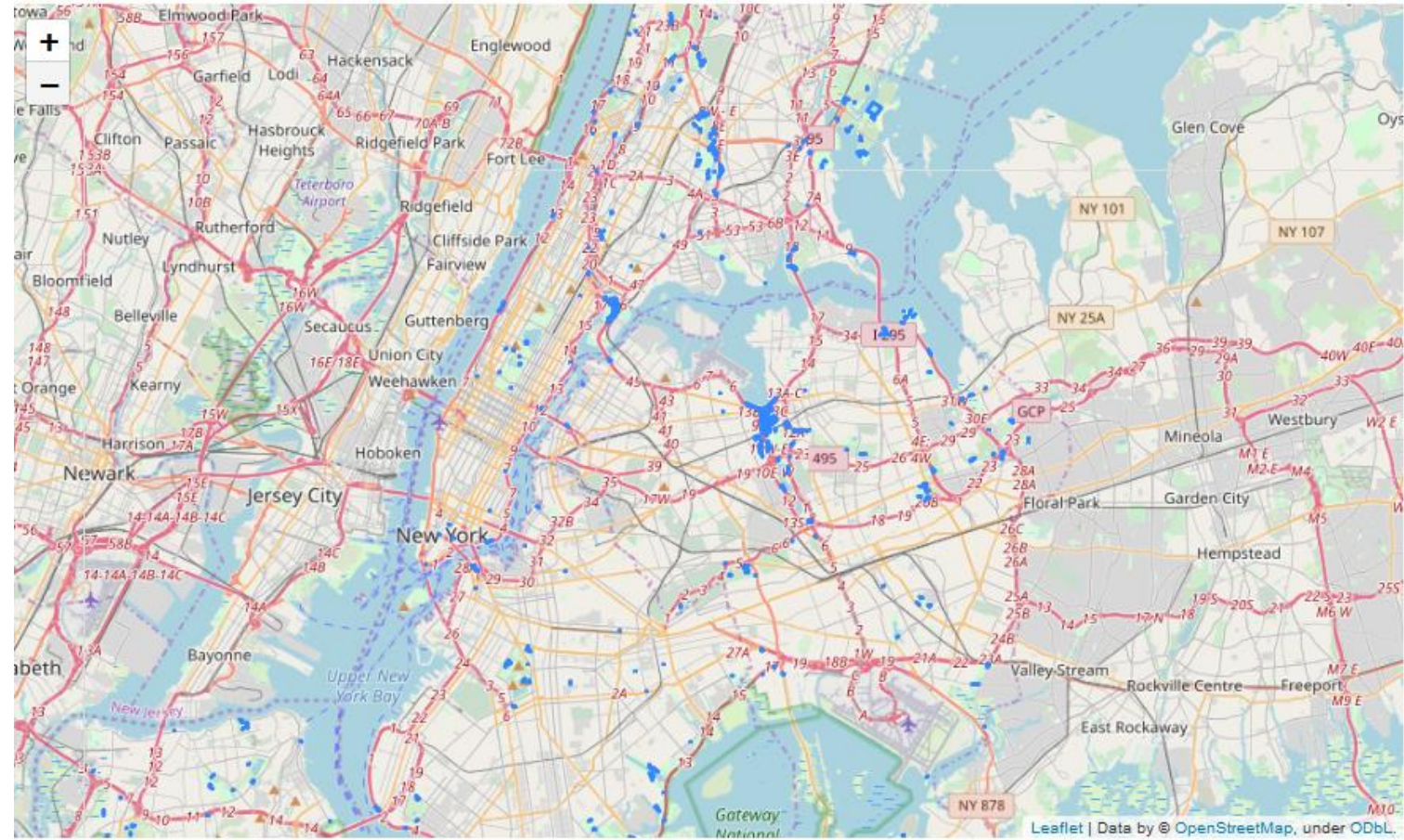
- Some locations are obviously in commercial zones, such as busy downtowns and developed strips of highway. Other locations may be on the fringes of a commercial and or residential zone. Find out if the building is properly zoned for a restaurant. Some towns [prohibit the sale of alcohol](#) within so many feet of a church or house of worship, while other towns still maintain “dry” ordinances left over from prohibition.
- Data Source
- <https://data.cityofnewyork.us/City-Government/Zoning-GIS-Data-Shapefile/kdig-pewd>

Zoning Areas



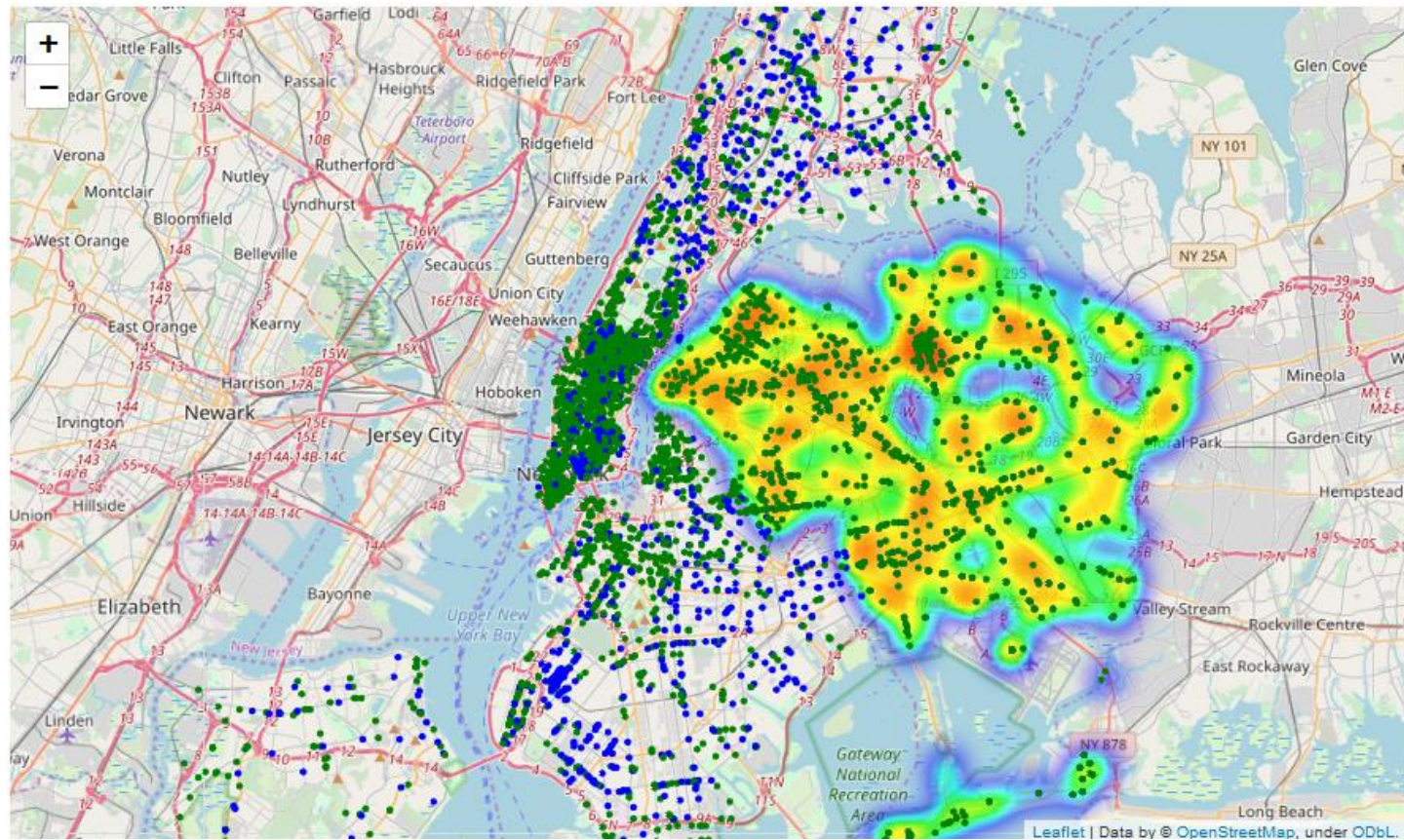
Location - Parking

- People are lazy. There is just no way around it. If they have to walk a fair distance to get to your restaurant, they may opt to go somewhere else "more convenient." If you live in an urban area where everyone walks and there is public transportation, this is less of a factor. If you are thinking of a restaurant location out of town, in a place that requires you to drive to get there, you'd better have parking available. If your [restaurant location](#) doesn't come with a large parking lot, is it near a municipal parking lot for patrons to use?
- Data Sources
 - <https://data.cityofnewyork.us/City-Government/Parking-Lots/e2f7-cs7i>
 - <https://data.cityofnewyork.us/Business/parking-garage-list/5bhr-pjxt>



Location - Visibility

- Setting up shop in a location with either high foot or car traffic is ideal. Making your restaurant (or restaurant sign) visible to the public is like free advertising. It reminds them that your restaurant exists and they should stop by for dinner sometime.
- Data Sources
 - https://en.wikipedia.org/wiki/Economy_of_New_York_City
 - https://en.wikipedia.org/wiki/Cuisine_of_New_York_City



Open Data – NYC

- Open Data is a veritable source of openly available data for New Yorkers that is produced and used by City government. The relevant data sets have been referenced in the above slides, with the specific problems which they will be used to tackle.
- Wikipedia is another major source of data, which can be used to cross reference the various aspects of New York.

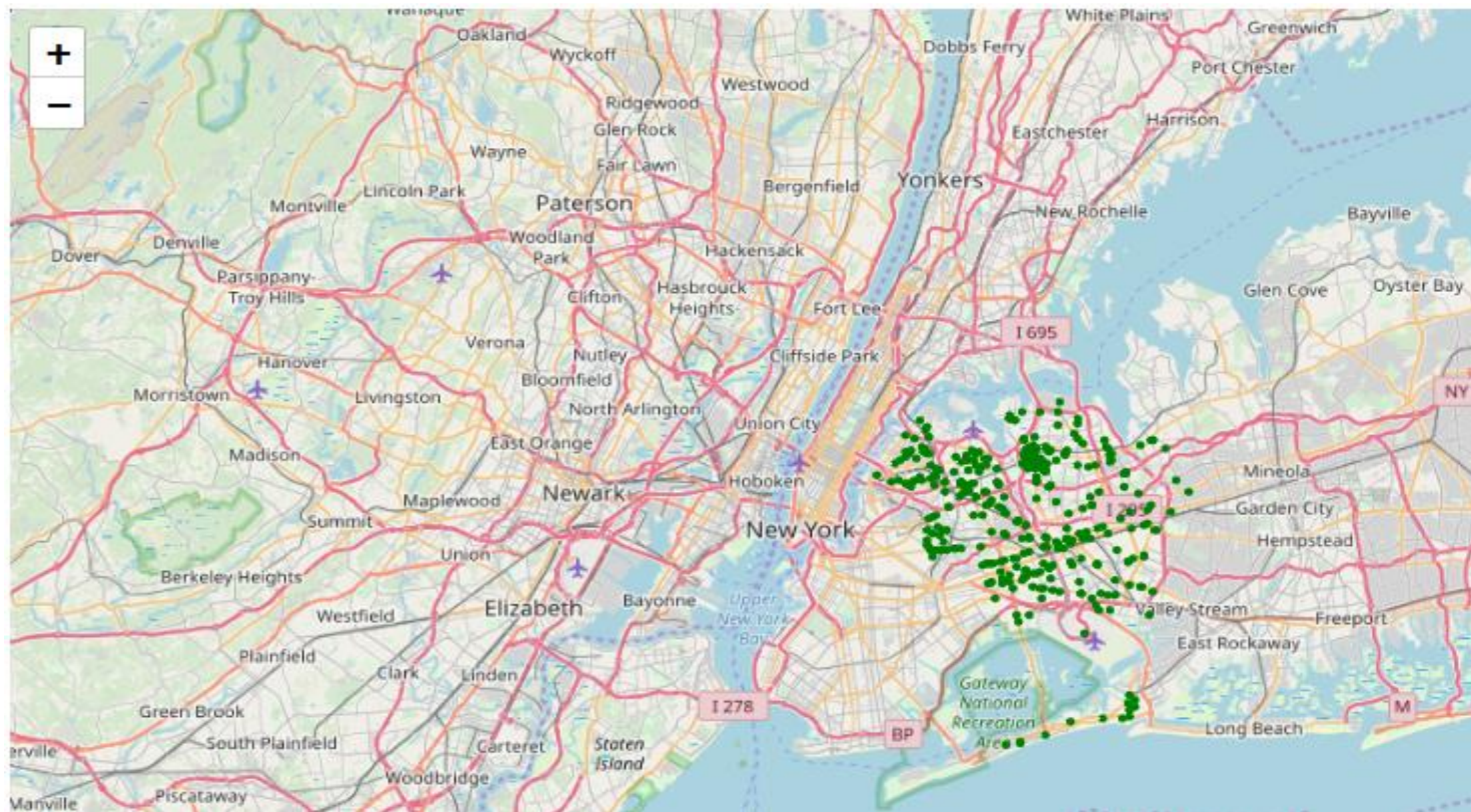
References

- These links actually were very helpful in collating and collecting information for understanding the business domain
 - <https://www.thebalancesmb.com/how-to-find-out-the-population-base-of-an-area-2888660>
 - <https://www.fundera.com/blog/starting-a-restaurant>
 - <https://www.thebalancesmb.com/before-you-choose-a-restaurant-location-2888548>
 - <https://www.thebalancesmb.com/choosing-restaurant-location-2888543>

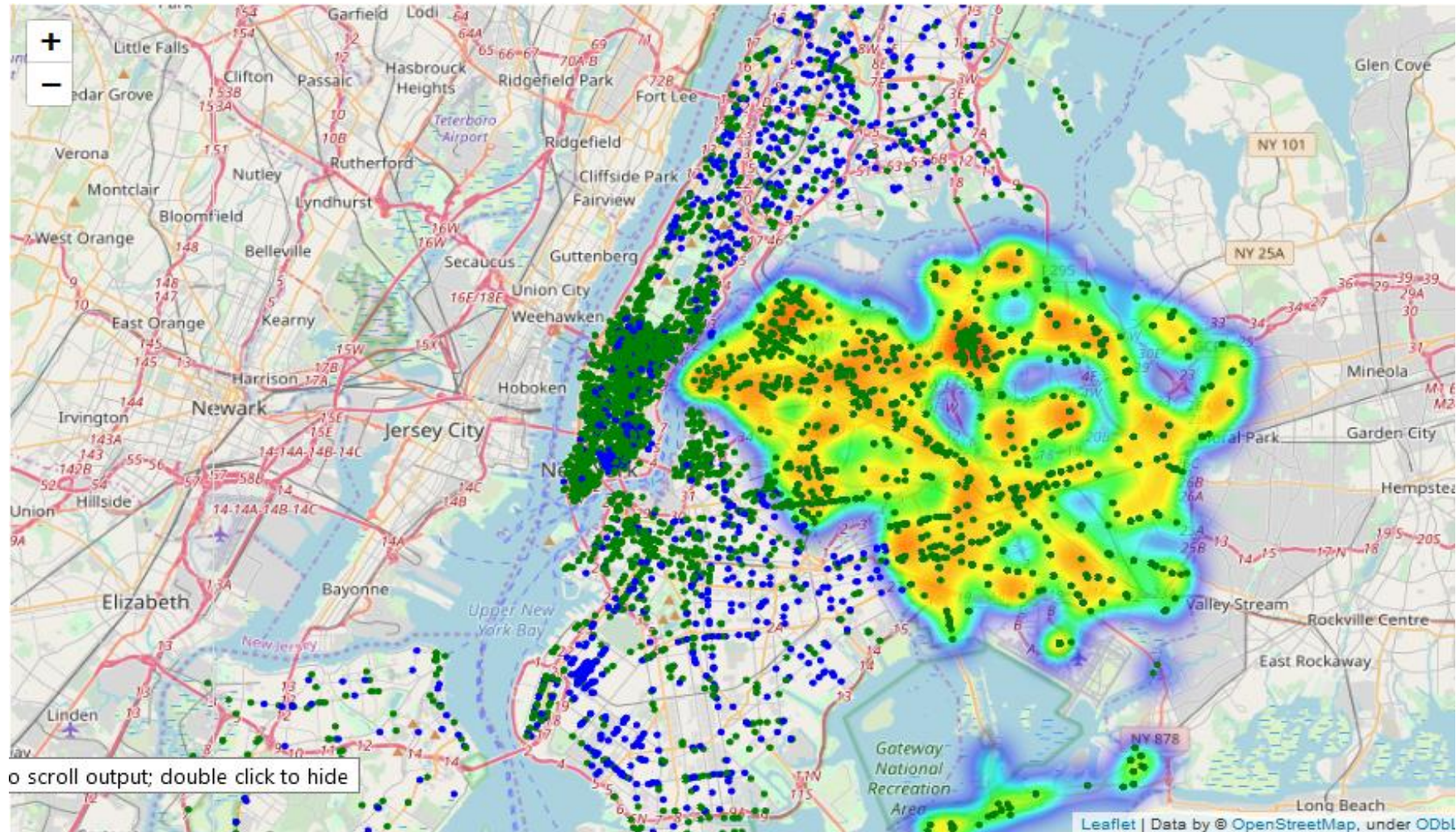
Methodology

Locations

- Find out the location of the restaurant now..
- find out how many Chinese restaurants are there in the Queens and plot
- find a neighborhoods where there are not many chinese restaurants, but has a high concentration of other restaurants
- Plot the data of Chinese Restaurants in Queens



Explore the hotspots and overlay it with Chinese Restaurants



Notebook Link

- Running all the cells in the notebook and trying to upload it to GitHub was giving a file too big error, as the size was becoming more than 25mb. Hence I have run only relevant cells in the notebook. All the maps/data outputs are there in the report and the presentation.

Please consider that before final grading. Thanks

https://github.com/smukherjee/Coursera_Capstone/blob/master/CapstoneRestaurantOpening.ipynb

Results and Discussions

- From our Analysis,
- We have been able to identify the Restaurant hotspots in Queens
- We have been able to identify and shortlist the areas where there are enough Restaurants, but there are not too many Chinese Restaurants
- As a result of our analysis there are 8 areas which a potential restaurant owner can further explore on opening up of a Chinese Restaurants. The people in these areas have an habit of eating out, and the market is not saturated with Chinese restaurants.
- Further Analysis could be done on
 - Parking availability
 - Zoning Licenses and Liquor licenses
 - Existing properties available

Conclusions

- This project could be used as a starting point for a potential restaurant owner to shortlist areas for opening a restaurant.
- Extensive analysis and availability of the data, it will also become easier in case, the owner is not interested in opening a Chinese Restaurant but interested in looking at an alternate cuisine.
- Since all the dataset used are publicly available and updated constantly, I hope that the work done remains relevant for some time to come.

References

- Report: https://cocl.us/coursera_capstone_report
- Notebook: https://cocl.us/coursera_capstone_notebook
- Presentation: https://cocl.us/coursera_capstone_presentation