Capstone Project - Restaurant Opening New York

Applied Data Science Capstone by Sujoy Mukherjee

Table of contents

Introduction

For this endeavour, we are trying to find an optimal location for a restaurant in **New York** and also help the person chose the **cuisine**.

We will use the data science algorithms learned to generate promising neighbourhoods based on this criteria.

Data

Based on definition of our problem, factors that will influence our decision are:

- Population profile
- number of existing restaurants in the **Target Market**.
- Location
  - Where
  - Parking lots locations
  - Visibility

Following data sources will be used to extract/generate the required information:

- https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/rs6k-p7g6
- https://geo.nyu.edu/catalog/nyu_2451_34572
- https://data.cityofnewyork.us/City-Government/Zoning-GIS-Data-Shapefile/kdig-pewd
- https://data.cityofnewyork.us/City-Government/Parking-Lots/e2f7-cs7i
- https://data.cityofnewyork.us/Business/parking-garage-list/5bhr-pjxt
- https://en.wikipedia.org/wiki/Economy_of_New_York_City
- https://en.wikipedia.org/wiki/Cuisine_of_New_York_City

- 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 customers between 15 to 44 years

Boroughs

New York City encompasses five county-level administrative divisions called boroughs: The Bronx, Brooklyn, Manhattan, Queens, and Staten Island. Each borough is coterminous with a respective county of New York State. The boroughs of Queens and the Bronx are concurrent with the counties of the same name, while the boroughs of Manhattan, Brooklyn, and Staten Island correspond to New York, Kings, and Richmond counties, respectively.

All five boroughs came into existence with the consolidation of New York City in 1898, when New York County, Kings County, Queens County, and Richmond County consolidated into one municipal government. The former municipalities in Kings, Richmond, and most of Queens counties were abolished, though the counties themselves were retained. However, the five boroughs have not always been coextensive with their five respective counties.

https://en.m.wikipedia.org/wiki/Boroughs_of_New_York_City

Community Districts

The community boards of the New York City government are the appointed advisory groups of the community districts of the five boroughs. There are currently 59 community districts: twelve in Manhattan, twelve in the Bronx, eighteen in Brooklyn, fourteen in Queens, and three in Staten Island:

```
Community boards of Manhattan
Community boards of the Bronx
Community boards of Brooklyn
Community boards of Queens
Community boards of Staten Island
```

They advise on land use and zoning, participate in the city budget process, and address service delivery in their district. Regarding land use they are only advisory and mostly serve as mobilizing institutions for communities opposed to specific projects. The City Charter also allows boards to submit their own plans for the development, growth, and improvement of their communities.

https://en.wikipedia.org/wiki/Community_boards_of_New_York_City

Zoning is the language of the physical city. It aims to promote an orderly pattern of development and to separate incompatible land uses, such as industrial uses and homes, to ensure a pleasant environment. The Zoning Resolution is a legal instrument to regulate and establish limits on the use of land and building size, shape, height, and setback.

The current Zoning Resolution was adopted on December 15, 1961 to apply citywide. When changing conditions warrant them, periodic updates and amendments are made to maintain a rational use of land. Each plot of land within the City's jurisdiction has a zoning designation — residence, commercial, or manufacturing—to establish relevant parameters for building and land use. Numerous zoning districts are mapped in the City's diverse neighborhoods to preserve their varying density and character. These limits help give shape to neighborhoods and predictability to their future. The City continues to adapt the Zoning Resolution as the land use patterns in the City change through private and public actions.

https://www1.nyc.gov/site/planning/zoning/about-zoning.page

Restaurants

For getting a list of Restaurants, there are multiple options and sources available on the net including using Four Square APIs. I decided to infer the Restaurant data from NYC Open Data available for inspections. Since inspections are very thorough and regular, it would cover most of the restaurants.

This list can also be used at a later stage to gain more insights into the targeted locality to figure out additional details.

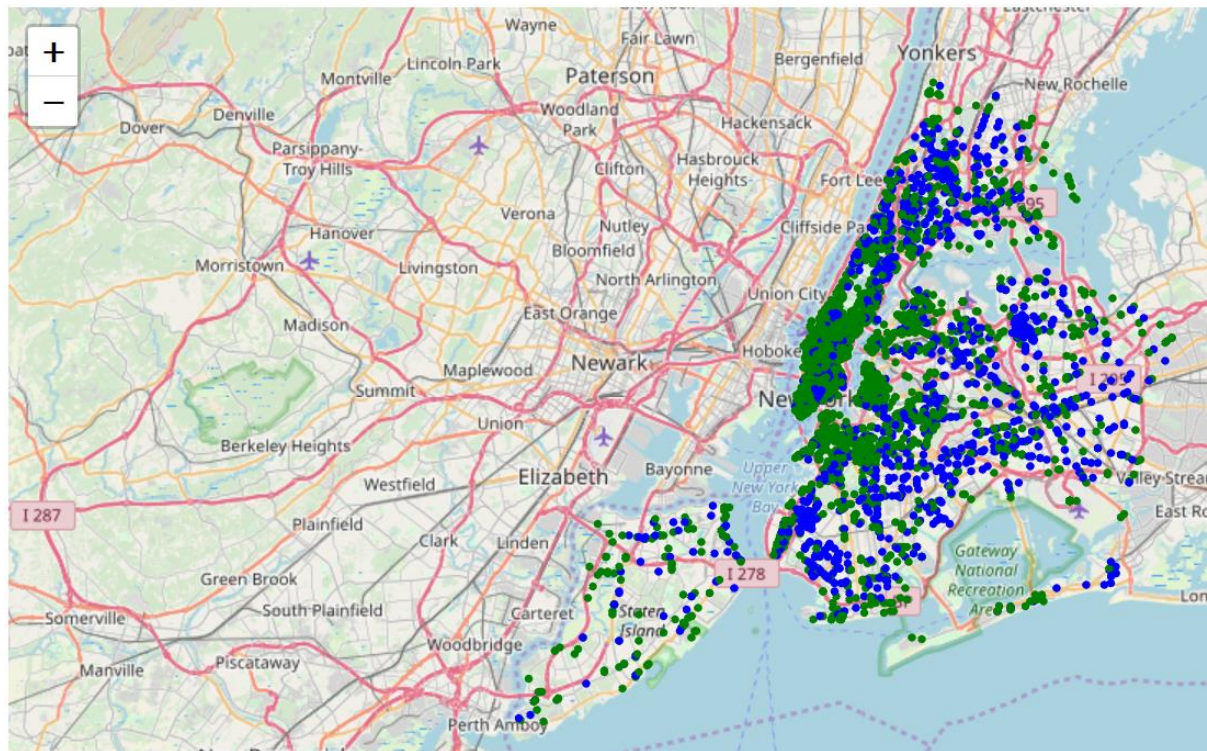|	CAMIS	DBA	BORO	BUILDING	STREET	ZIPCODE	PHONE	CUISINE DESCRIPTION	INSPECTION DATE	ACTION	...	RECORD DATE	INSPECTION TYPE
0	30075445	MORRIS PARK BAKE SHOP	Bronx	1007	MORRIS PARK AVE	10462.0	7188924968	Bakery	06/11/2019	Violations were cited in the following area(s).	...	11/11/2019	Cycle Inspection / Re-inspection
1	30112340	WENDY'S	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
2	30191841	DJ REYNOLDS PUB AND RESTAURANT	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

ick to scroll output; double click to hide

Group the restaurants and count by the Cuisine

| | CAMIS |
CUISINE DESCRIPTION	
Afghan	10
African	32
American	3374
Armenian	16
Asian	192
Australian	14
Bagels/Pretzels	100
Bakery	459
Bangladeshi	19

Now Let's just explore the restaurants serving American and Chinese food as that's the two types of cuisines we are currently interested in.

We also need to focus on

- an area where there are enough restaurants - Shows us that there is a demand
- Paying Population (around 18 - 40 years) would be our target demographics

Demographics

Selected demographic and housing estimates data citywide and by borough. Five year estimates of population data from the Census Bureau's American Community Survey. https://catalog.data.gov/dataset/demographic-social-economic-and-housing-profiles-by-borough

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

From the data, also figure out the density of population for ethnicity of our target audience.

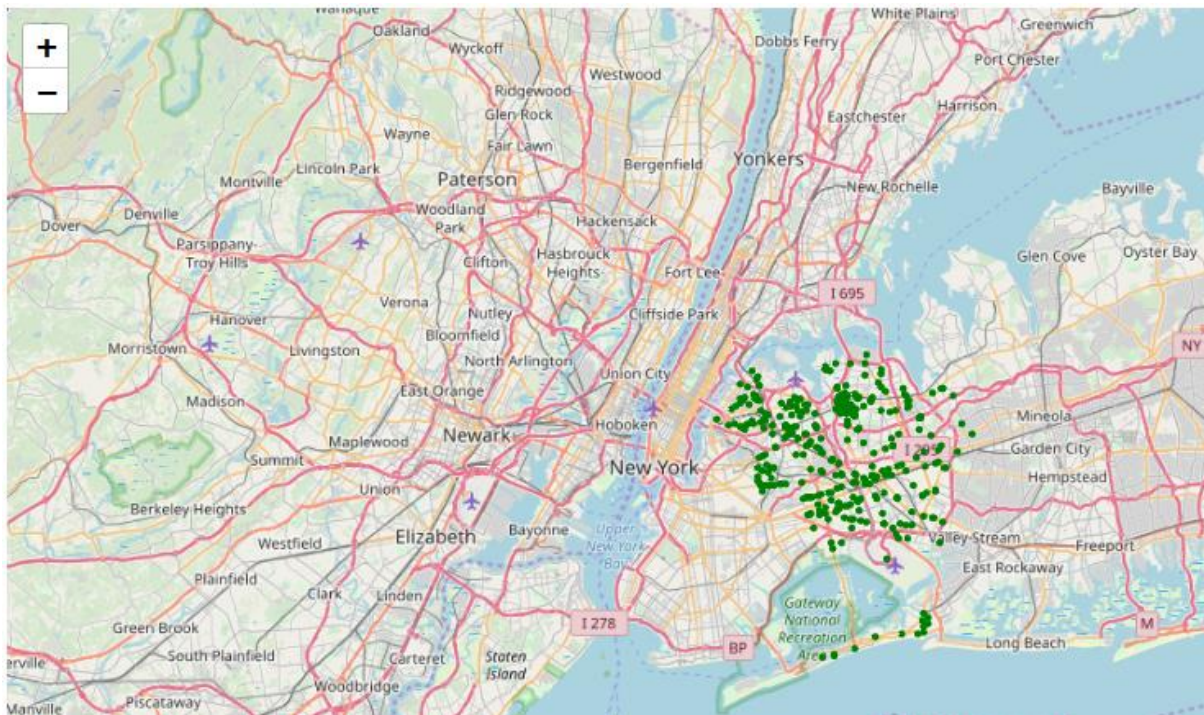| Subject	Bronx	Brooklyn	Manhattan	Queens	Staten Island	New York city
Asian	52462.0	304037.0	192573.0	572759.0	38052.0	1159883.0

From the above, we have done initial EDA and now we have our target customer base, cuisine and borough where we would concentrate on for opening our restaurant

We are targetting a low cost Chinese Restaurant in Queens catering to a group of people in the age group of 5 to 44 years, who may prefer takeaways and delivery
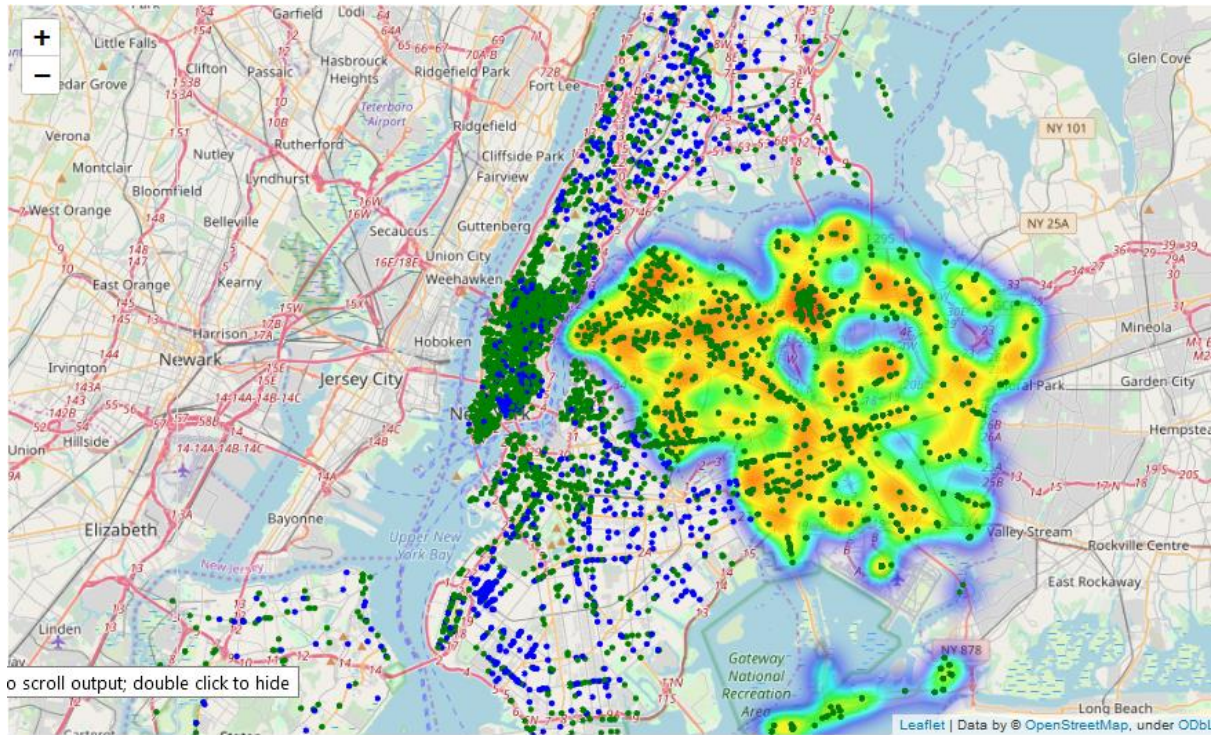
Methodology

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

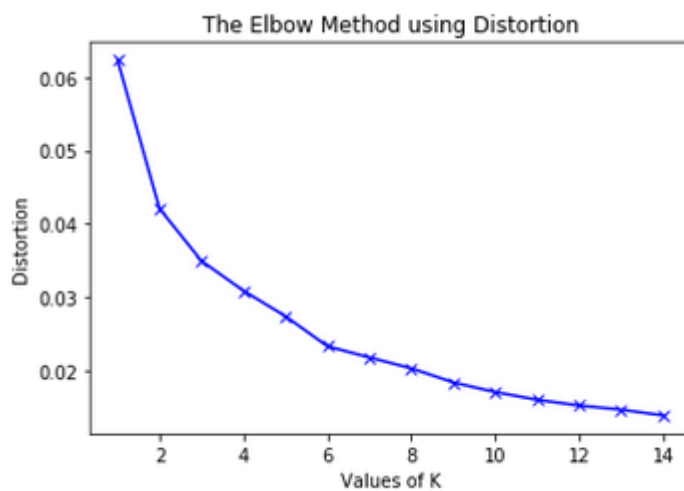Plot the data of Chinese Restaurants in Queens

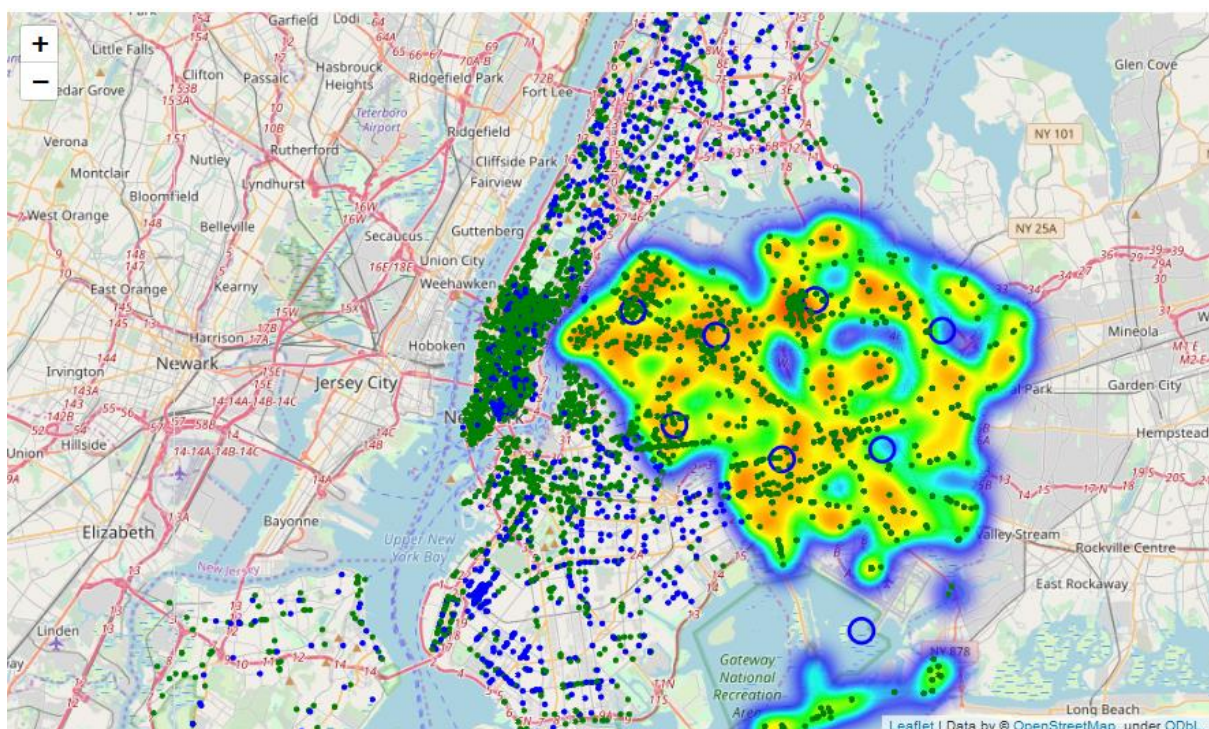Explore the hotspots and overlay it with Chinese Restaurants

Analysis

From the heatmap there are a couple of areas where restaurant density is high, but Chinese restaurants are less. We would get the centroid of those areas using KMeans and then suggest the shortlisted prospective locations

Use KMeans algorithm and the elbow method to find out the centroids and ideal number of clusters

Use the above data to find out the centroid of all the clusters and Plot the centroid found using KMeans along with the Chinese restaurants in Queens to visually look at the data

Now we need to give the location of the centroids as the shortlisted places, with a high enough demand (as suggested by the Restaurant concentration), but also lesser availability of Chinese restaurants.

```
===============================================================
Addresses of centres of areas recommended for further analysis
===============================================================

['40-71 Denman St, Flushing, NY 11373, USA',
 '108-31 173rd St, Jamaica, NY 11433, USA',
 '144-60 Roosevelt Ave, Flushing, NY 11354, USA',
 '87-40 109th St, Jamaica, NY 11418, USA',
 '64-03 229th St, Flushing, NY 11364, USA',
 '32-77 37th St, Long Island City, NY 11103, USA',
 'Perimeter Rd, Jamaica, NY 11430, USA',
 '65-6 Fresh Pond Rd, Ridgewood, NY 11385, USA']
```

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

Conclusion

This notebook could be used as a starting point for a potential restaurant owner to shortlist areas for opening a restaurant. Due to the 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.