Applied Data Science Capstone Project

Optimal Locations for a Italian Restaurant in Manhattan, NY

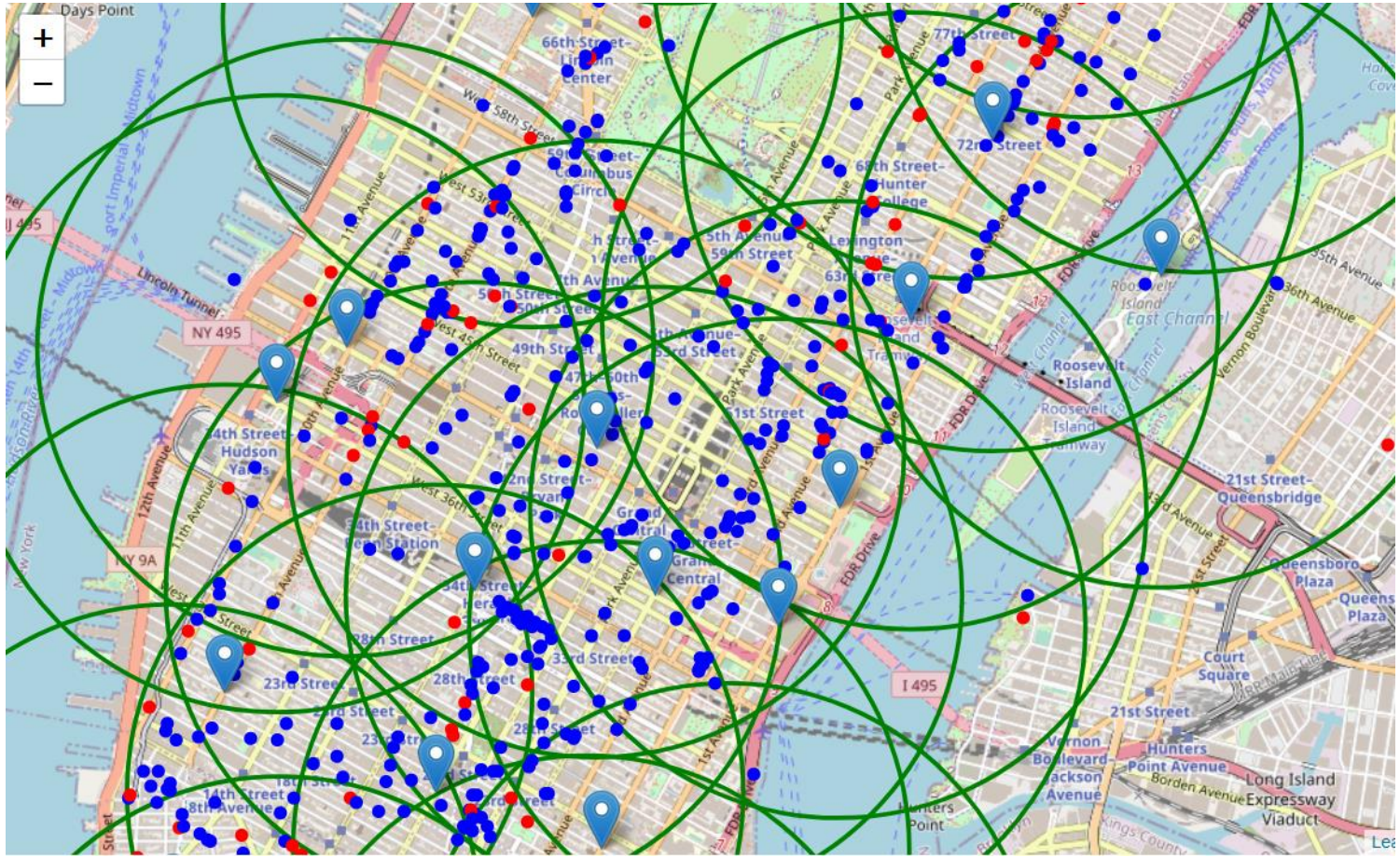
Introduction

- This study focuses on finding good locations where an Italian restaurant can be opened.
- The same technique can also be used for problems where we want to open other businesses like shopping malls, offices etc.

Data Acquisition

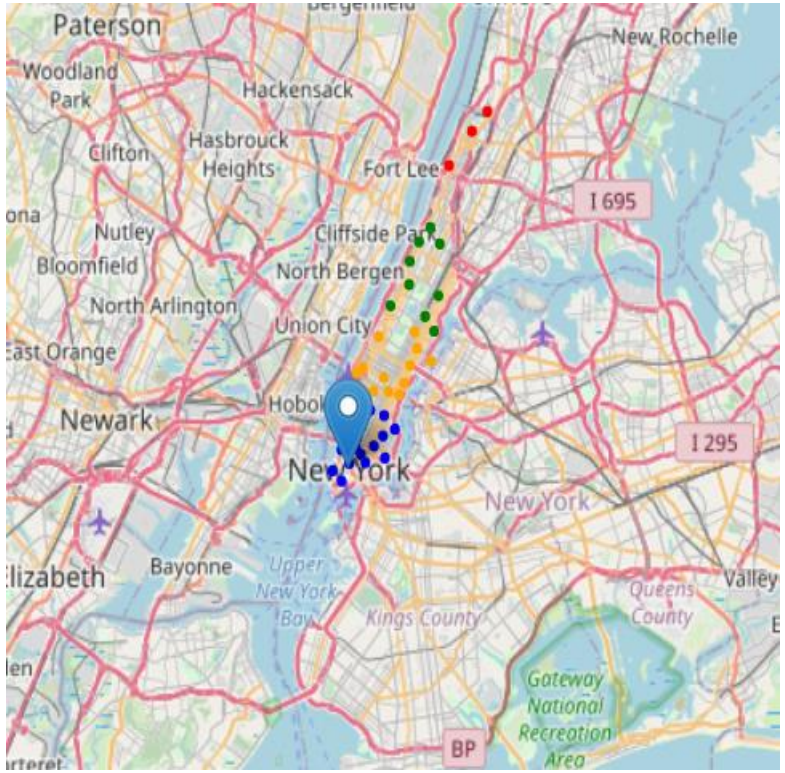
- Using the Manhattan Latitude and Longitude data for all 40 neighbourhoods from the New York data used in the Lab of Week 3.
- Using the Foursquare API for data about the restaurants around the neighbourhood centers.
- Using reverse geocoding from Google APIs to get the optimal address from their latitude, longitude value.

Observations

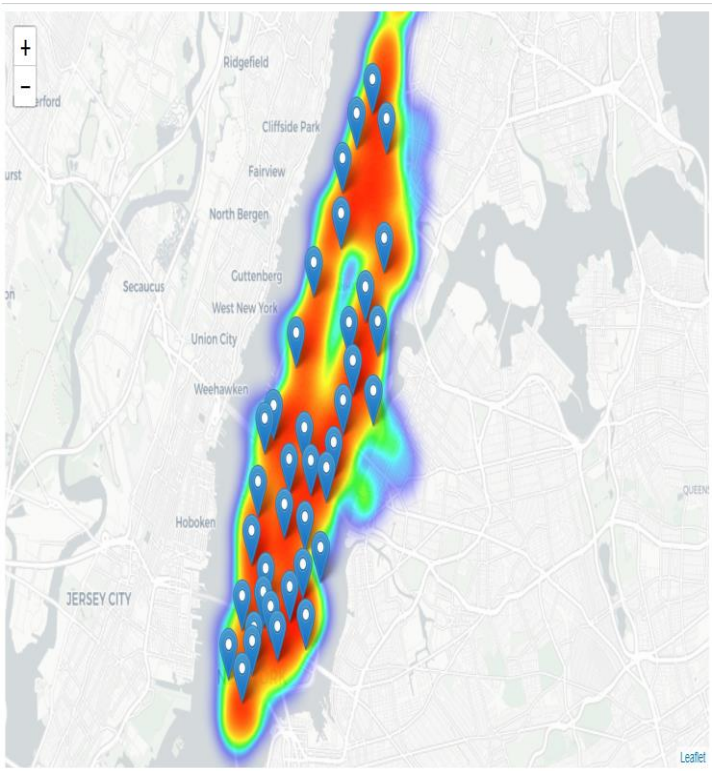


Restaurants are in blue and Italian Restaurants are in red

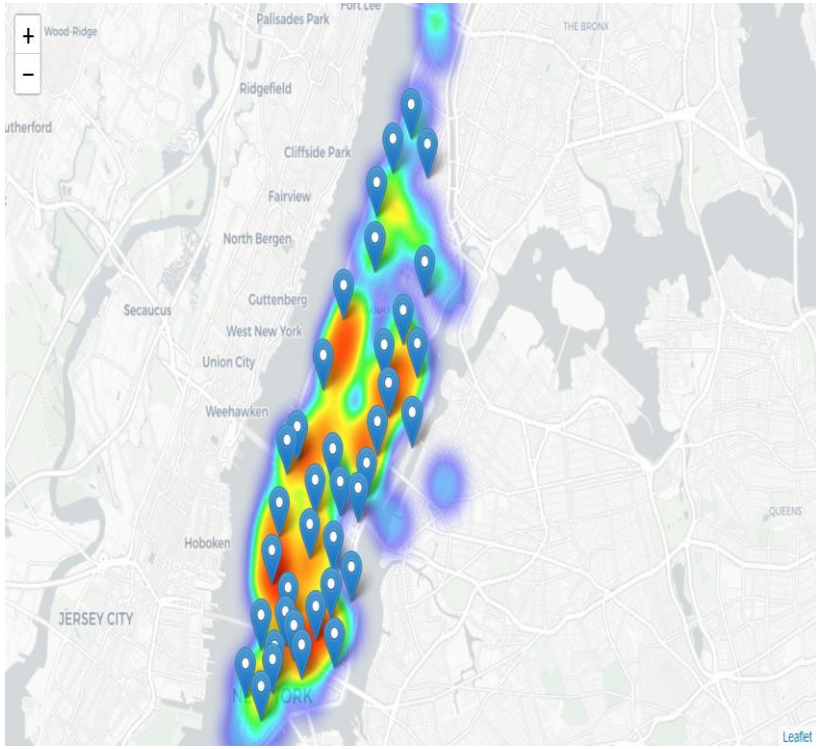
Now we can use this data to find density of the restaurants and the Italian restaurants which will help us in finding regions where there are few Italian restaurants.



My attempt to group the nearby neighbourhoods of Manhattan



Heatmap of Restaurants in Manhattan



Heatmap of Italian restaurants in Manhattan

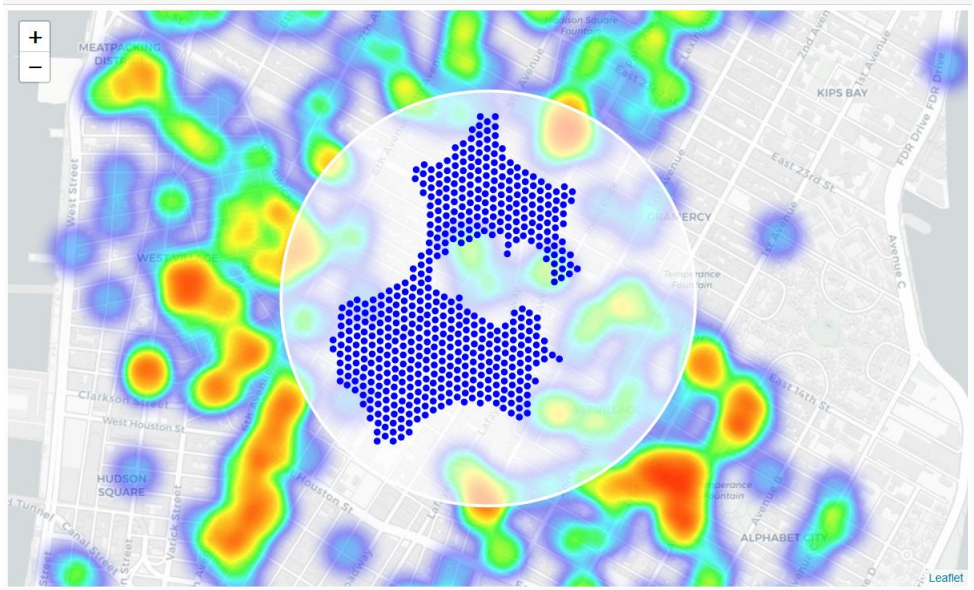
From the above plots, one can see that concentrating on the blue and orange shaded neighbourhoods would be ideal.

Heatmap of Restaurants around the first cluster of neighbourhoods with the circles of radius .5,1,1.5 KM.

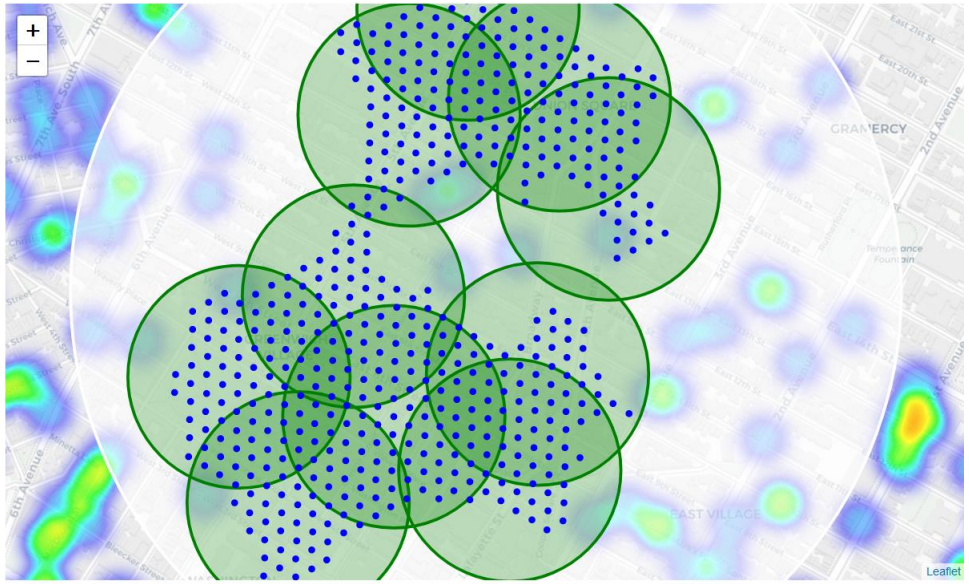
Heatmap of Italian restaurants around the first cluster of neighbourhoods with the circles of radius .5,1,1.5 KM.

Here the white circle represents the area of our interest since the density of Italian restaurants is very low.

I decided to partition our area of interest into locations spaced 50 metres apart and choose locations where the number of restaurants closer than 250 metres is less than 2 and the closest Italian restaurant is farther than 400 metres.



The blue dots represent the optimal locations according to the above criteria.



Since there were over 500 such locations all situated close to each other, I decided to use k-means clustering algorithm to group them into 10 locations

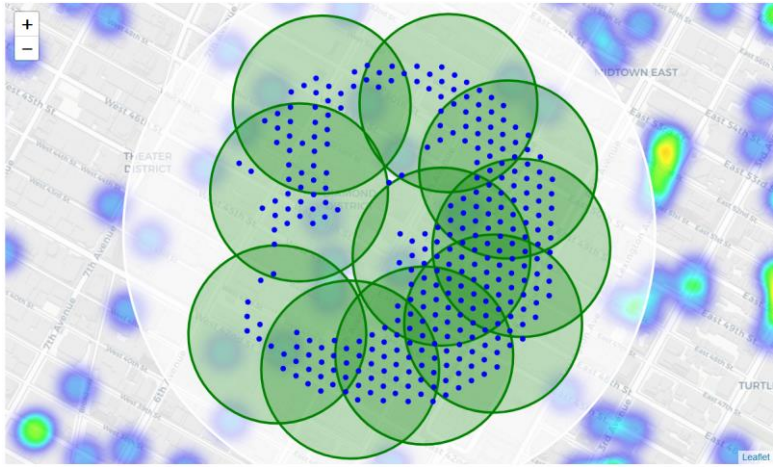
The same procedure for the second cluster

Heatmap of
restaurants around second cluster

Heatmap of Italian
restaurants around second cluster

The white circle represents
the region of interest

The optimal locations



The optimal location clusters

Results

- The 10 location address for each of the two neighbourhood clusters was computed and made available to the stake holders which they can use to do final street-level exploration to find their optimal location to open a new Italian restaurant.

These are the 10 candidate addresses from our cluster 0 neighborhoods

```
for lon, lat in cluster_centers:
    addr = get_address(google_api_key, lat, lon)
    print(addr)
```

```
7 E 9th St, New York, NY 10003, USA
89 5th Ave, New York, NY 10003, USA
260 Greene St, New York, NY 10003, USA
780 Broadway, New York, NY 10003, USA
2 Union Square E, New York, NY 10003, USA
76 5th Ave, New York, NY 10011, USA
44 West 4th Street, New York, NY 10012, USA
18 Washington Square N, New York, NY 10011, USA
Union Square Park, 201 Park Ave S, New York, NY 10003, USA
13 Astor Pl, New York, NY 10003, USA
```

```
for lon, lat in cluster_centers:
    addr = get_address(google_api_key, lat, lon)
    print(addr)
```

```
341 Madison Ave, New York, NY 10017, USA
22 E 50th St, New York, NY 10022, USA
1180 6th Ave, New York, NY 10036, USA
66 E 46th St, New York, NY 10017, USA
33 W 42nd St, New York, NY 10036, USA
6 Av/W 48 St, 6th Ave, New York, NY 10020, USA
14 E 47th St, New York, NY 10017, USA
16 W 51st St, New York, NY 10111, USA
280 Park Ave # 27e, New York, NY 10017, USA
511 5th Ave, New York, NY 10017, USA
```

Conclusion and future directions

- Built a model to find good locations in Manhattan to open a new Italian restaurant.
- Building a model that takes input(for example we may want to open a Greek restaurant instead or a Movie Theatre) and gives the desired locations would be an improvement . (Note : It is feasible since the problems are such that, optimal locations are those where there are no competitors nearby.)