REPORT:
The Battle of Neighborhoods

Coursera Capstone Project Final Assigmnent

REPORT STRUCTURE

1. INTRODUCTION
2. DATA
3. METHODOLOGY
4.RESULTS
5. DISCUSSION
6. CONCLUSION

1. INTRODUCTION: BUSINESS PROBLEM

The Business Problem consists of the following questions:

- Which is the most convenient activity to open in New York?
- Which is the most convenient neighborhood to open this new activity?

This business problem is for investors that wants to differentiate their investments in a solid, not temporary or seasonal activity

2. DATA

Foursquare location data has been used.

Data have been acquired making calls to Foursquare API for exploring the venues of each neighborhoods in New York.

3. METHODOLOGY

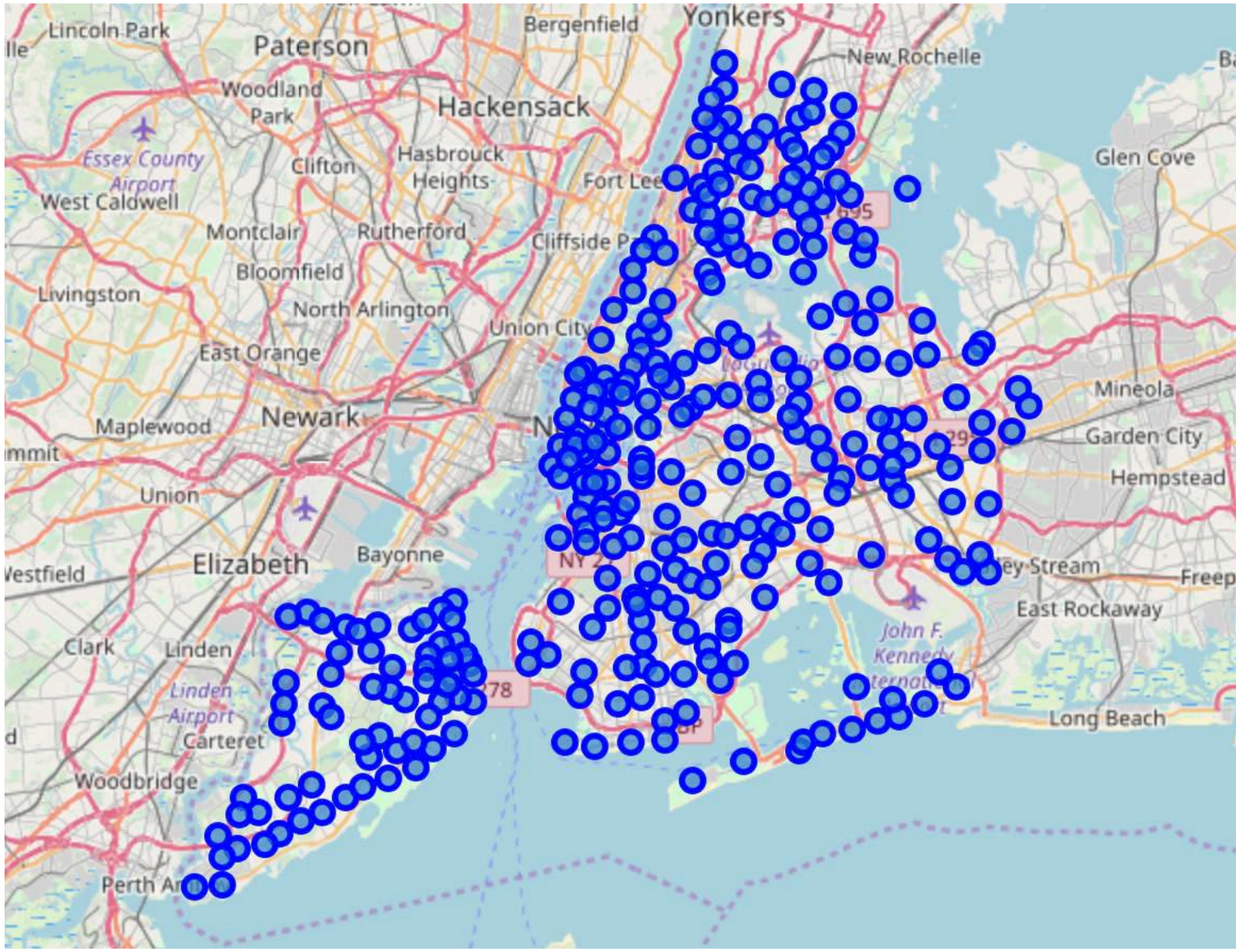
1. Visualize on a map all the neighborhoods in New York as exploratory analysis;

2. **Assumption 1**: The best neighborhood has to be found in the neighborhoods that have the higher number of venues because, normally, they are the most crowded locations and because they have been found as convenient to the bigger part of investors/entrepreneurs;

3. The neighborhoods with the bigger number of venues have been found and shown on a map. Taking in account that the neighborhood with the higher number of venues is Chinatown (238 venues), only the neighborhoods with a number of venues bigger that the 80% of Chinatown's venues have been considered;

4. The venues have been clustered both in 3 and 5 clusters with Kmeans algorithm in order to understand their similarity and, finally, if there is a particular type of activity that is still missing in one of the considered neighborhoods and extract the final solution.

3.1 Visualize on a map all the neighborhoods in New York as exploratory analysis



Neighborhoods in New York

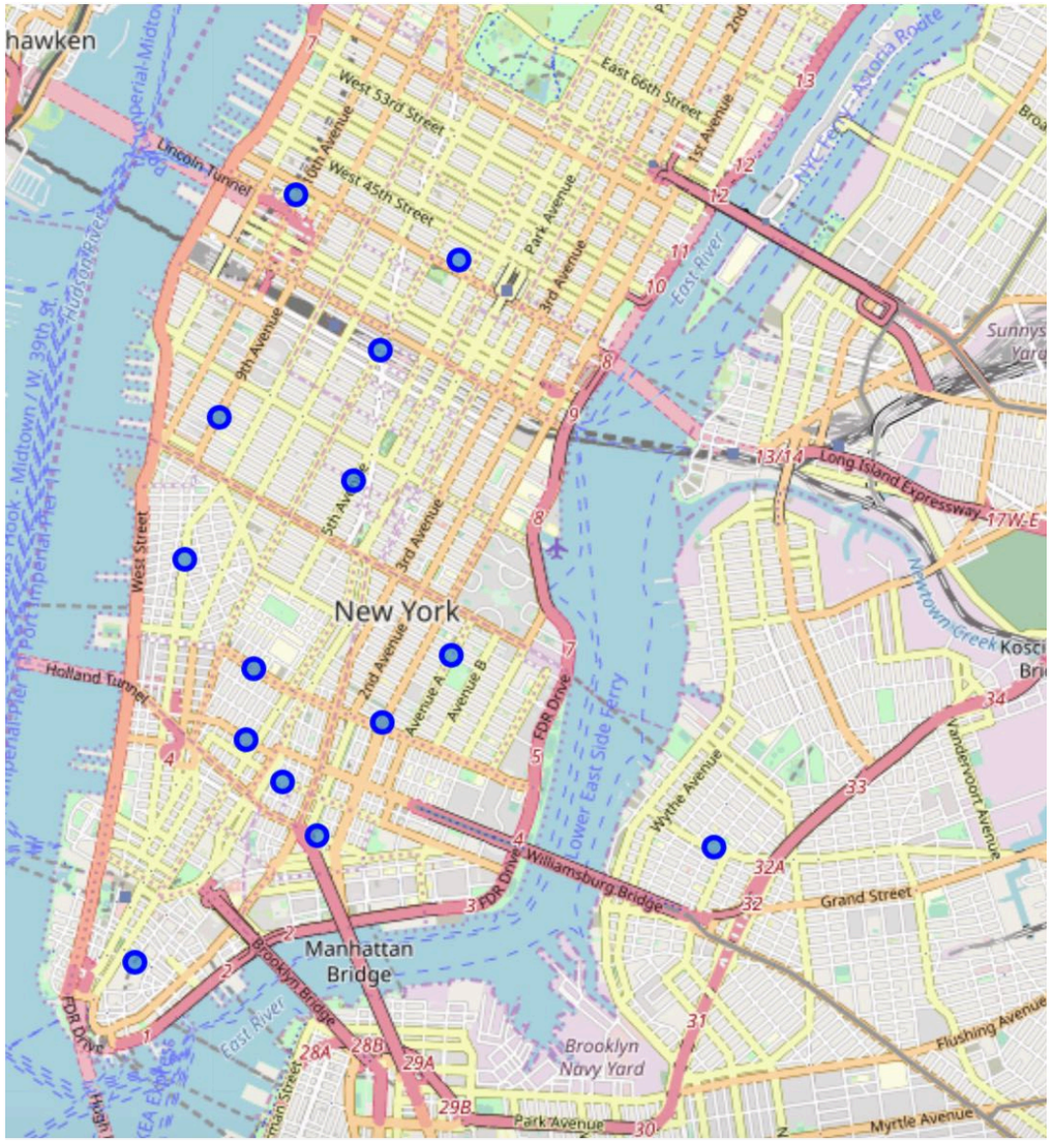
3.2 Assumption N.1

The best neighborhood to start the new activity has to be found in the neighborhoods that have the higher number of venues because, normally, they are the most crowded locations and because they have been found as convenient to the bigger part of investors/entrapeneurs

3.3 The neighborhoods with the bigger number of venues have been found

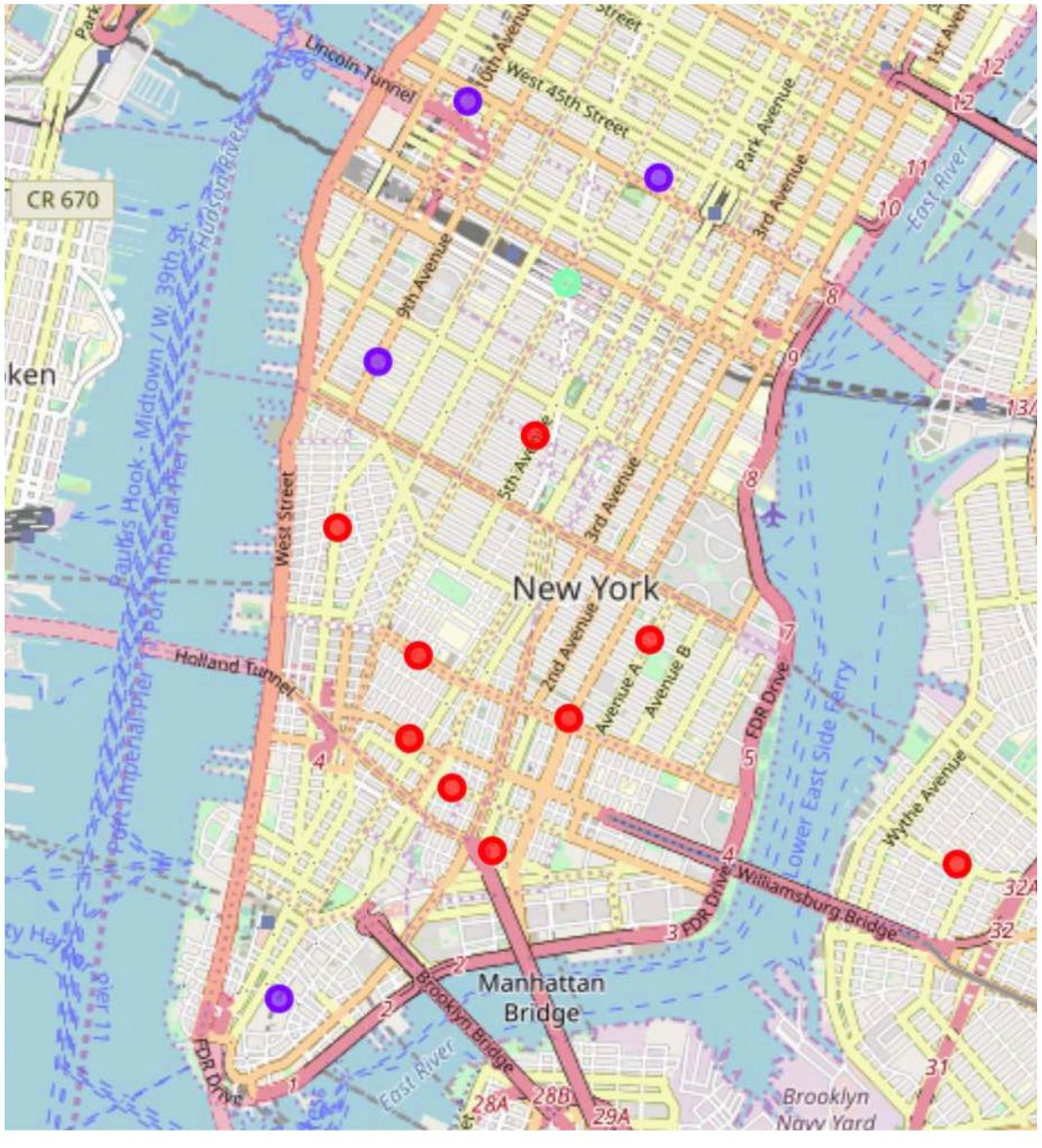
	Borough	Latitude	Longitude	N. of venues	Neighborhood
100	Manhattan	40.715618	-73.994279	238	Chinatown
117	Manhattan	40.726933	-73.999914	237	Greenwich Village
121	Manhattan	40.719324	-73.997305	234	Little Italy
248	Manhattan	40.723259	-73.988434	227	Noho
118	Manhattan	40.727847	-73.982226	226	East Village
128	Manhattan	40.707107	-74.010665	225	Financial District
250	Manhattan	40.748510	-73.988713	225	Midtown South
96	Brooklyn	40.714823	-73.958809	221	North Side
114	Manhattan	40.754691	-73.981669	221	Midtown
276	Manhattan	40.739673	-73.990947	219	Flatiron
113	Manhattan	40.759101	-73.996119	205	Clinton
116	Manhattan	40.744035	-74.003116	195	Chelsea
123	Manhattan	40.734434	-74.006180	193	West Village
122	Manhattan	40.722184	-74.000657	192	Soho

Data of the neighborhoods with the bigger number of venues



Maps of the neighborhoods with the bigger number of venues

3.4 Neighborhoods' clustering



Neighborhoods clustering with k=3. The three closest red points in the bottom part of the image above are Chinatown, Little Italy and Soho neighborhoods. They have been considered a unique cluster in terms of searching for the less diffused activities in this area because of their nearness.

Neighborhoods clustering with k=5. The green point in both images is the Midtown South neighborhood: it resulted as a unique cluster and its most common venues have been used to assess the possible vacancy in terms of types of activity in the Chinatown, Little Italy and Soho area.

4. RESULTS

1. The 14 venues have been clustered both in 3 and 5 clusters with Kmeans algorithm in order to understand their similarity and, finally, if there is a particular type of activity that is still missing in one of the considered neighborhoods. As result, the Midtown South neighborhood has been resulted in both analysis as a different (1-element) cluster from all the other clusters;
2. **ASSUMPTION N. 2**: Because Little Italy neighborhood is in between Chinatown and Soho and all the three neighborhoods are very close to each others (Chinatown and Soho are distant 1km), Little Italy can be considered the right place for starting the new activity as soon as there is a missing/low-diffused type of activity that is successful in the other main neighborhoods;
3. Searching for the diffusion of the most common activities for the Midtown South neighborhood in the Chinatown, Little Italy and Soho neighborhoods, the Korean Restaurant has been found as best result.

5. Discussion

- Analyzing the top 10 most common venues in Midtown South venues, resulted that Korean Restaurant, Japanese Restaurant, Hotel Bar, Gym/Fitness Center were not in the top 10 venues in Little Italy, Chinatown and Soho.
- In particular, only 2 Korean Restaurant resulted in three-neigborhoods-area (Little Italy analyzed with a radius of 500 m) and only 14 Japanese Restaurant. With a bigger radius of 1 km, only 26 Japanese restaurants have been found in the discussed area.
- About the position, although Little Italy neighborhood has been selected mainly for its number of total venues together with other neighborhoods, the final choice is based on the assumption that the nearness with other two main neighborhoods is an important and unique value. By the way, this fact has been considered in terms of the choice of the type of activity.
- The “pure” Korean restaurant results in almost an absence of concurrency in the whole area but the similarity between the Korean and the Japanese kitchens is a factor that has been taken into account. In facts, as mentioned, also the Japanese Restaurant results in a low concurrency business because there are 26 activities of this type in a radius of 1km from Little Italy. To clarify this number, it’s important to note that in the three neighborhoods area there are 664 venues (Chinatown:238, Little Italy:234, Soho: 192).
- Because the Soho neighborhood is plenty of cocktail bar and it’s famous for its after dinner attitude, it’s a good factor to have a restaurant close to a neighborhood in which the customers can continue to enjoy themselves.

6. Conclusion

The analysis gives a data driven and reasonable solution of the business problem. For sure the achieved result can be refined and improved taking into account other Foursquare location data such as the users' tips and trending venues and also other external sources of data.