

REPORT

COURSERA CAPSTONE PROJECT FINAL

ASSIGNMENT:

THE BATTLE OF NEIGHBORHOODS

INTRODUCTION:

The business problem consists of following questions:

1. Which is the most convenient activity to open in New York?
2. Which is the most convenient neighborhood to open this new activity?

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

DATA:

In order to solve the problem, Foursquare location data have been used.

Data have been acquired making calls to the Foursquare API for different purposes: for example, for searching for a specific type of venues, for exploring a particular venue and for exploring a geographical location also producing maps.

In particular, the Foursquare API have been queried for exploring the venues of each neighborhood of the city.

METHODOLOGY

The methodology can be divided in two main parts: the former is the research to find the best neighborhood in which to open a new activity is most convenient; the latter is the research to find the type of activity is most convenient to open given the neighborhood extracted in the previous part.

Accomplishing with these tasks, the adopted steps and assumptions are the following:

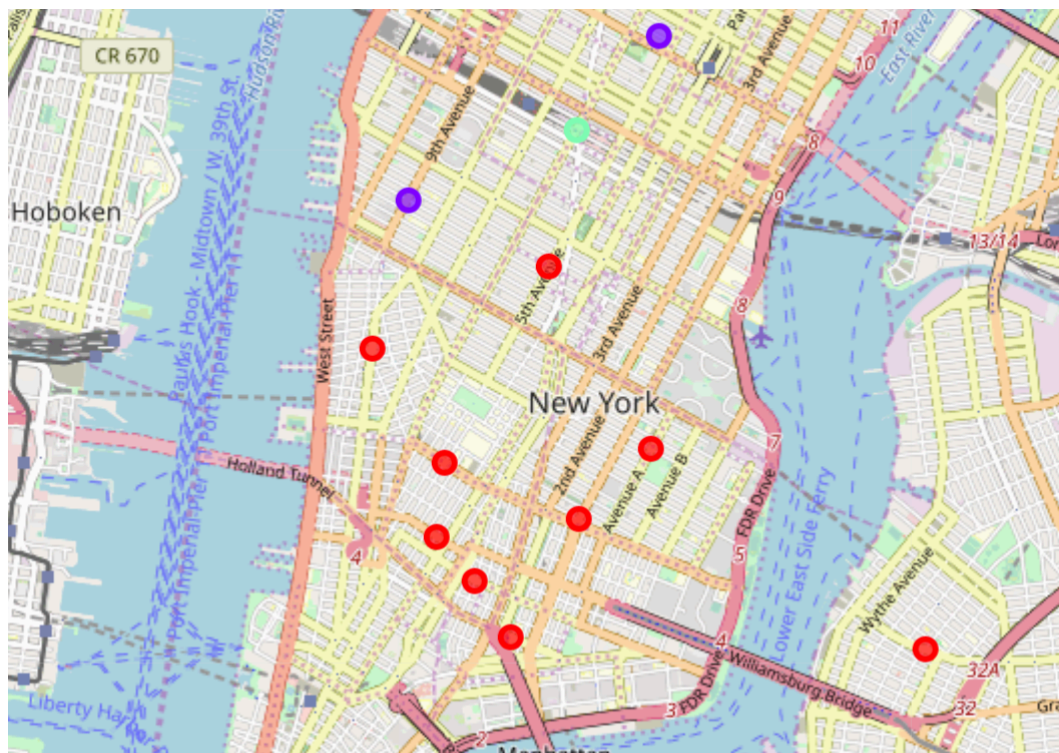
1. Visualize on a map all the neighborhoods of New York as exploratory analysis;
2. **Assumption 1:** The best neighborhood to start the new activity has to be found in the neighborhoods that have the higher number of venues because, normally, they are the most crowded locations and because they have been found as convenient to the bigger part of investors/entrapeneurs;
3. 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. As results, 14 venues (13 in Manhattan) area have been found;
4. 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;

5. **Assumption 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;
6. As consequence of steps 4 and 5, the most diffused activities in the Midtown South neighborhood have been searched in the Chinatown, Little Italy and Soho neighborhoods. At the end, the Korean Restaurant has been found as best result.

RESULTS

In the following images the neighborhoods with the higher number of venues have been shown:

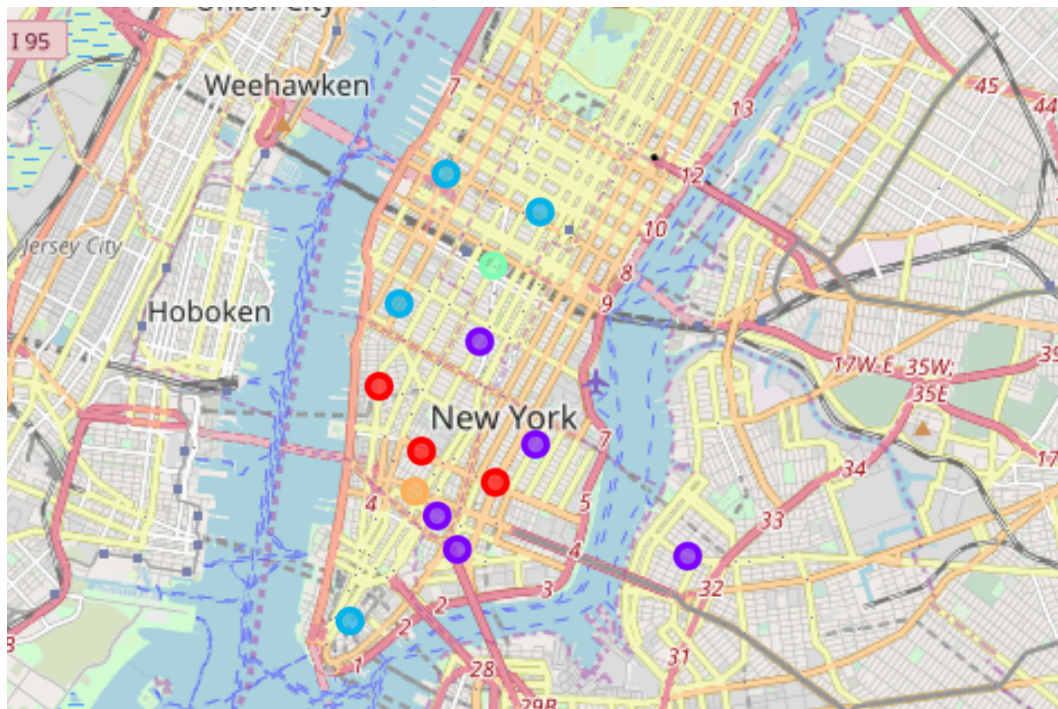


NEIGHBORHOODS WITH THE HIGHER NUMBER OF VENUES CLUSTERED WITH K=3.

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.

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.



NEIGHBORHOODS WITH THE HIGHER NUMBER OF VENUES CLUSTERED WITH K=5.

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-neighborhoods-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.

As final result, a Korean/Japanese Restaurant in Little Italy is suggested in case of opening a new activity in New York.

DISCUSSION

As mentioned in the previous parts, the Korean Restaurant in Little Italy should be the best choice in case of a decision about investing on a new activity in New York.

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). In addition, 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.

A high quality korean/japanese restaurant is also the last piece of a puzzle for a famous area in which different types of kitchen are already embedded.

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.