

The Battle of Neighborhoods

New York Asian Restaurant



Introduction

- City will be analyzed in this project: *New York*.
- Different kinds of venues located within this 302.6 mi² area, in order to fulfill all kinds of requirements. Based on this fact, I'd like to conduct area classification and segmentation of this big city, combining with my own experience, to explore new, uncovered secrets await under the map.



Initial Ideas

- Assume I'm a business owner, what is the best place for me to open my own business? (including restaurant, cafe etc.)
- Assume I'm an XXX critic (including restaurant, cafe etc.), which area should I visit most in order to work out my post?

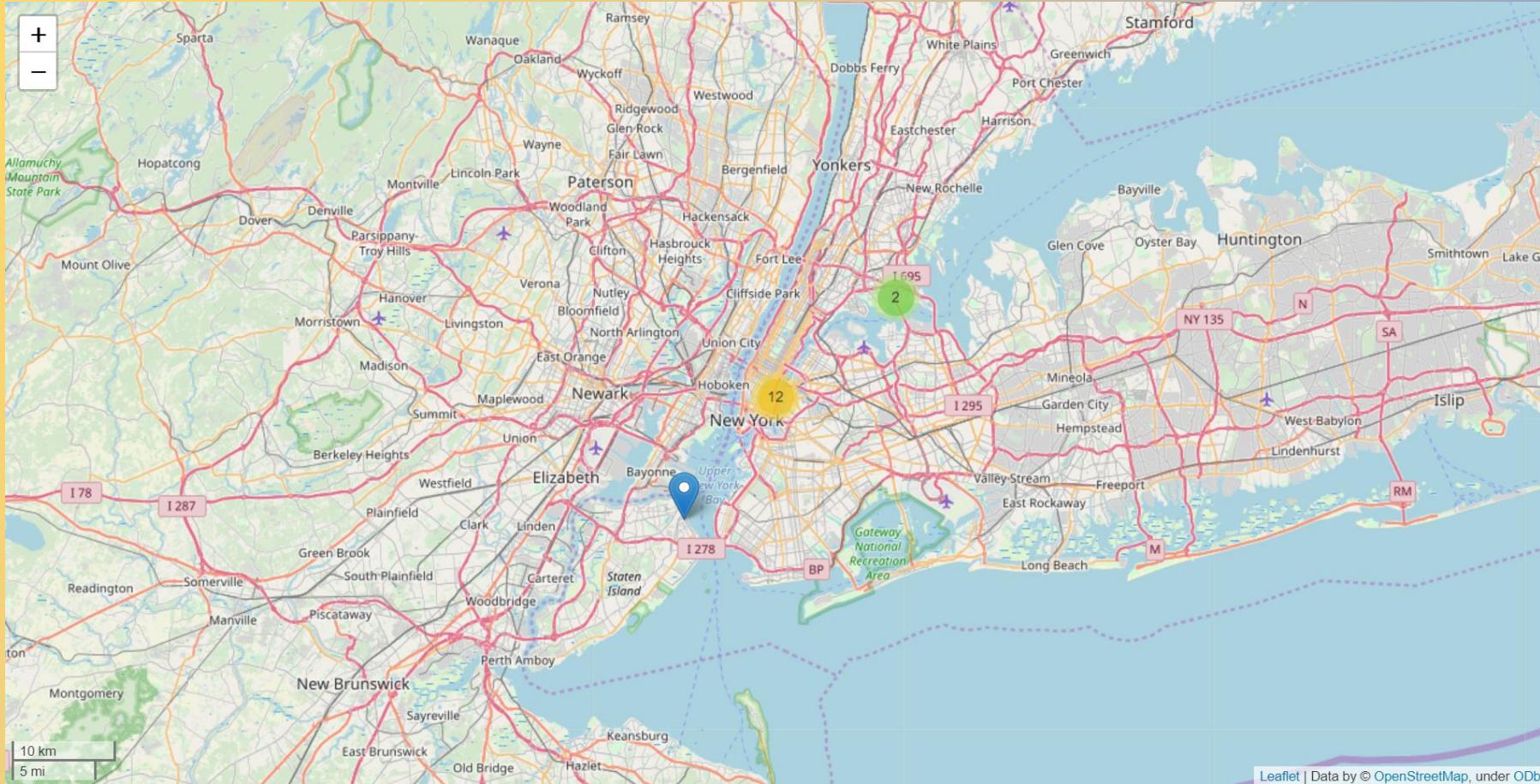


More thoughts during the process

- During the data exploring phase, I find out another dataset which contains rolling sales data of New York's real estates. This inspires me to further conduct my project (after analyzing the data get from Foursquare API).
- Based on these rolling sales data, I'm aiming at provide some more general suggestions to potential business owners about the commercial real estate's prices. I just do clustering & mapping here, but regression and other machine learning methodologies can be used if data source is more informative.



Results / Conclusion - Part One

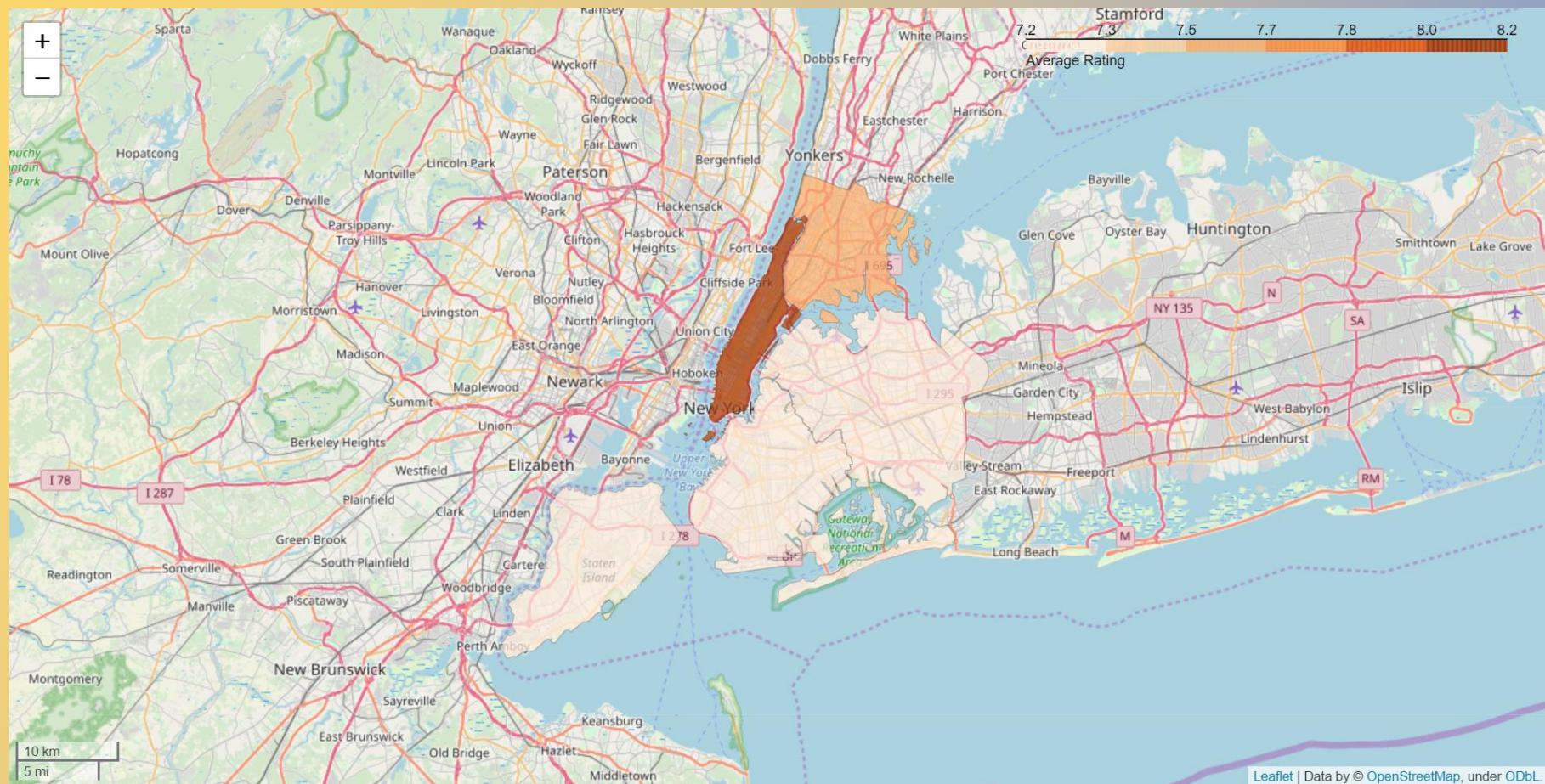


If you take a look at the simple map above, you will find it's pretty strange that there are just 15 neighborhoods within New York City show that they have some Asian restaurants. This situation is not what I have expected, and it finally leads to an average or even poor analysis.



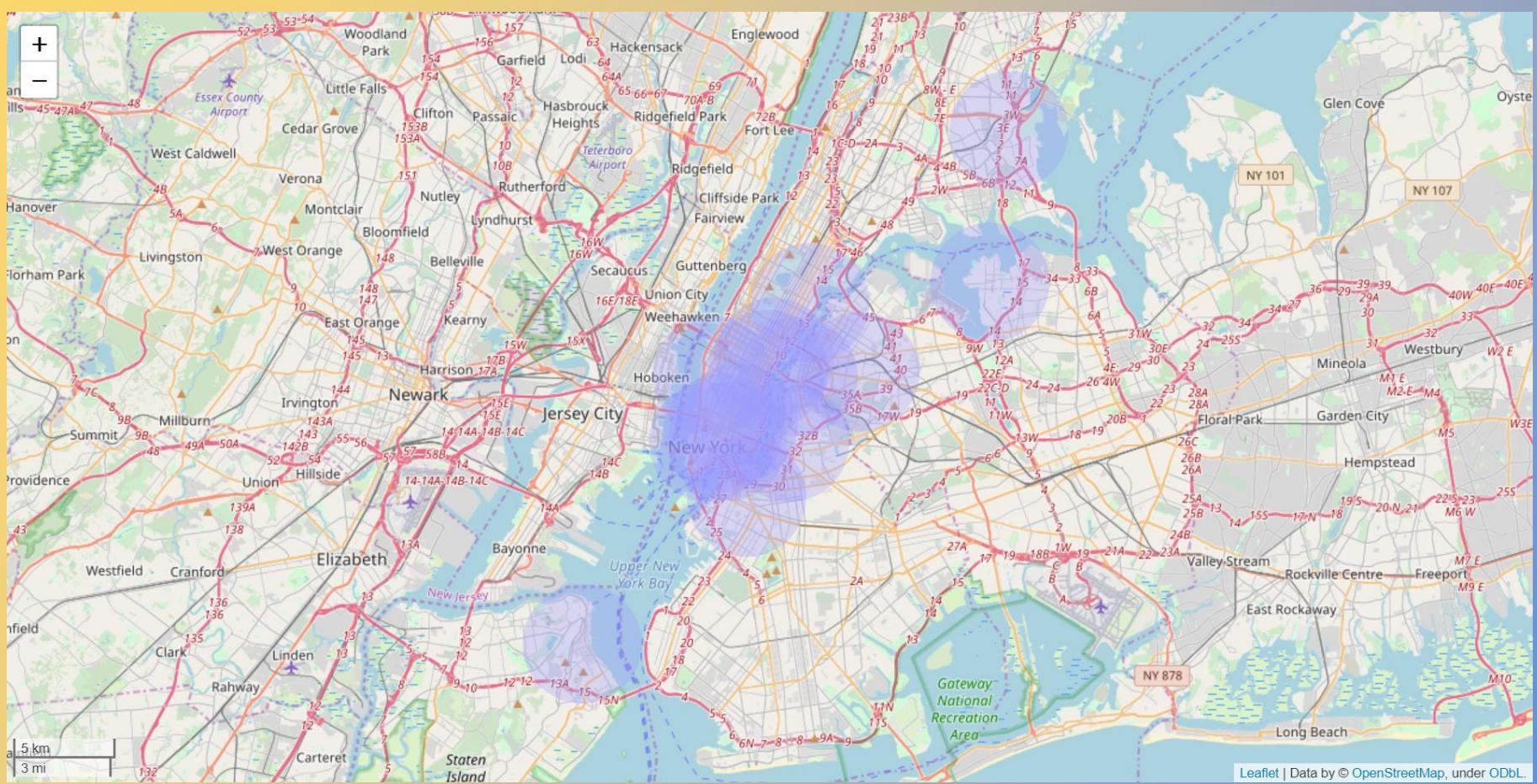
Even worse, like what has shown above, if we consider the numbers of Asian restaurants within these five boroughs, we can see there are only two restaurants show up in "Bronx", while "Manhattan" has 12.

However, if you search "Asian restaurant" in Yelp or other kinds of map app like Google Map, then you will see the numbers of Asian restaurants are way more than 36 in New York. In that case, seems like Foursquare API need to further expand & update its database in order to provide better user experience and fulfill all kinds of user requirements



From the graphs above we can see that there isn't any borough has an "Average Rating" lower than 7.2, which is good news to Asian food lovers, but it indicates a negative news to business owners who are seeking new opportunities.

Within these five boroughs, "Manhattan" & "Bronx" have the highest and second highest "Average Rating" 8.2 & 7.6, respectively. In that case, these two boroughs should be "must visit" if people are looking for some fantastic Asian food, while other boroughs also have some good Asian restaurants but needs more careful searching & preparation in order to avoid lower rating venues.



I also use shadow markers to generate a guide map which tells business owners where are the places they need to avoid -- those deep blue areas (the radius of each circle marker is 3km, which is commonly considered as the average radioactive radius for every restaurant).

Again, this map doesn't really help since I have to use the geo info of each neighborhood instead of the geo info of each restaurant, to locate the center of these circle markers, which results in lots of information loss.

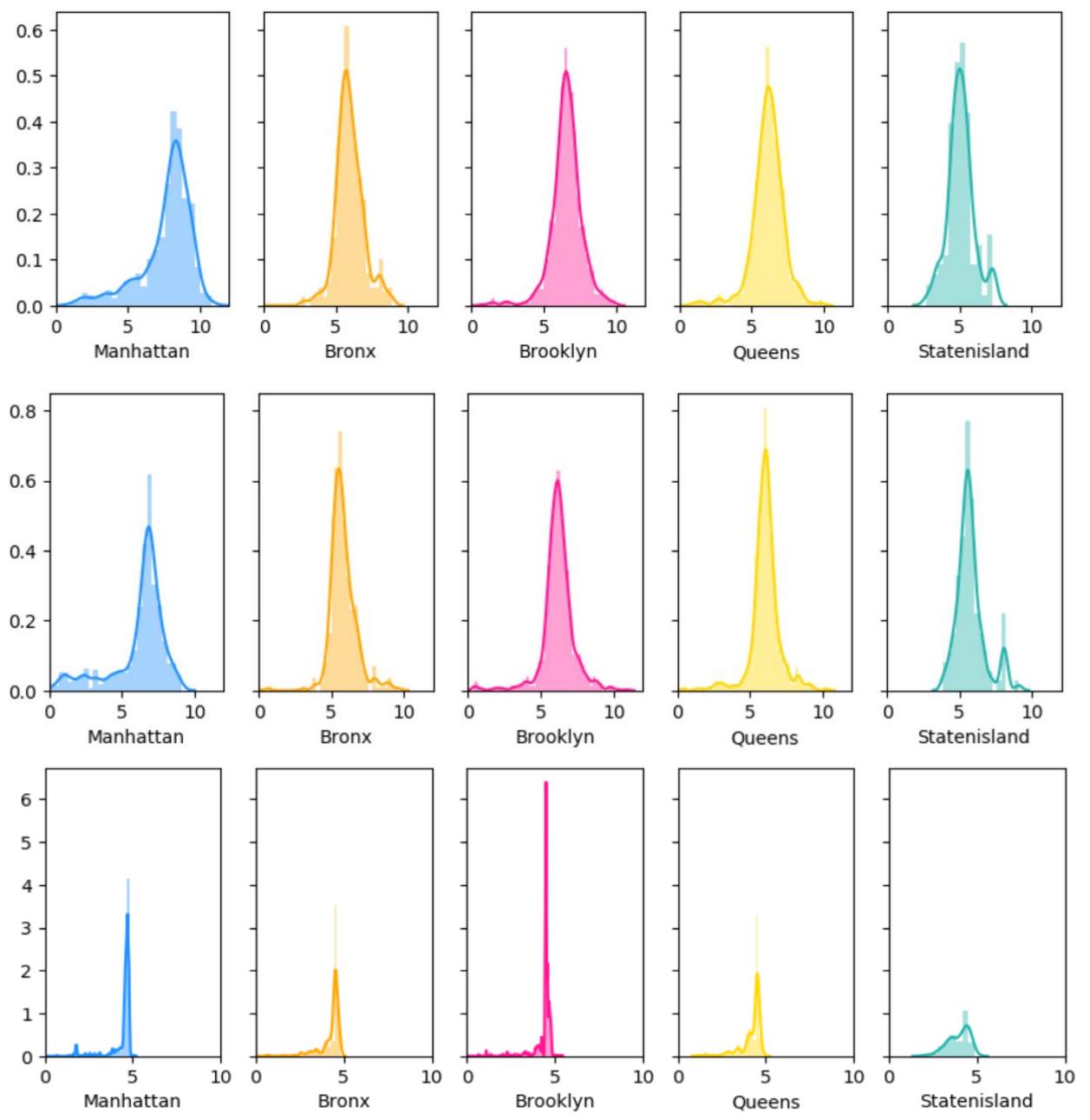


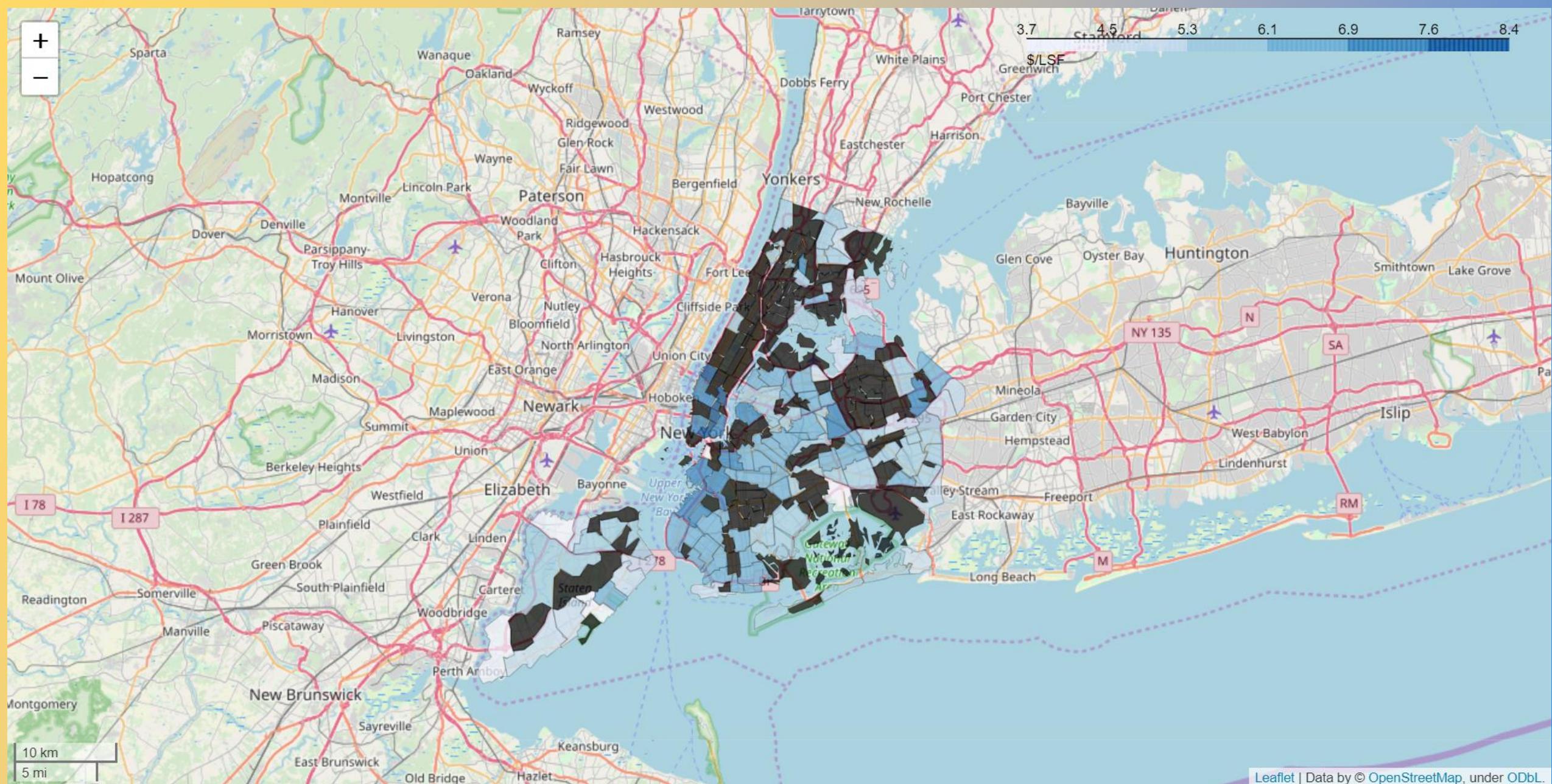
As a staged conclusion, besides what I have mentioned above to food lovers & business owners, I'd like to recommend using other APIs such as Yelp API to conduct a more detailed report. That may indeed help people to find places they are looking for.

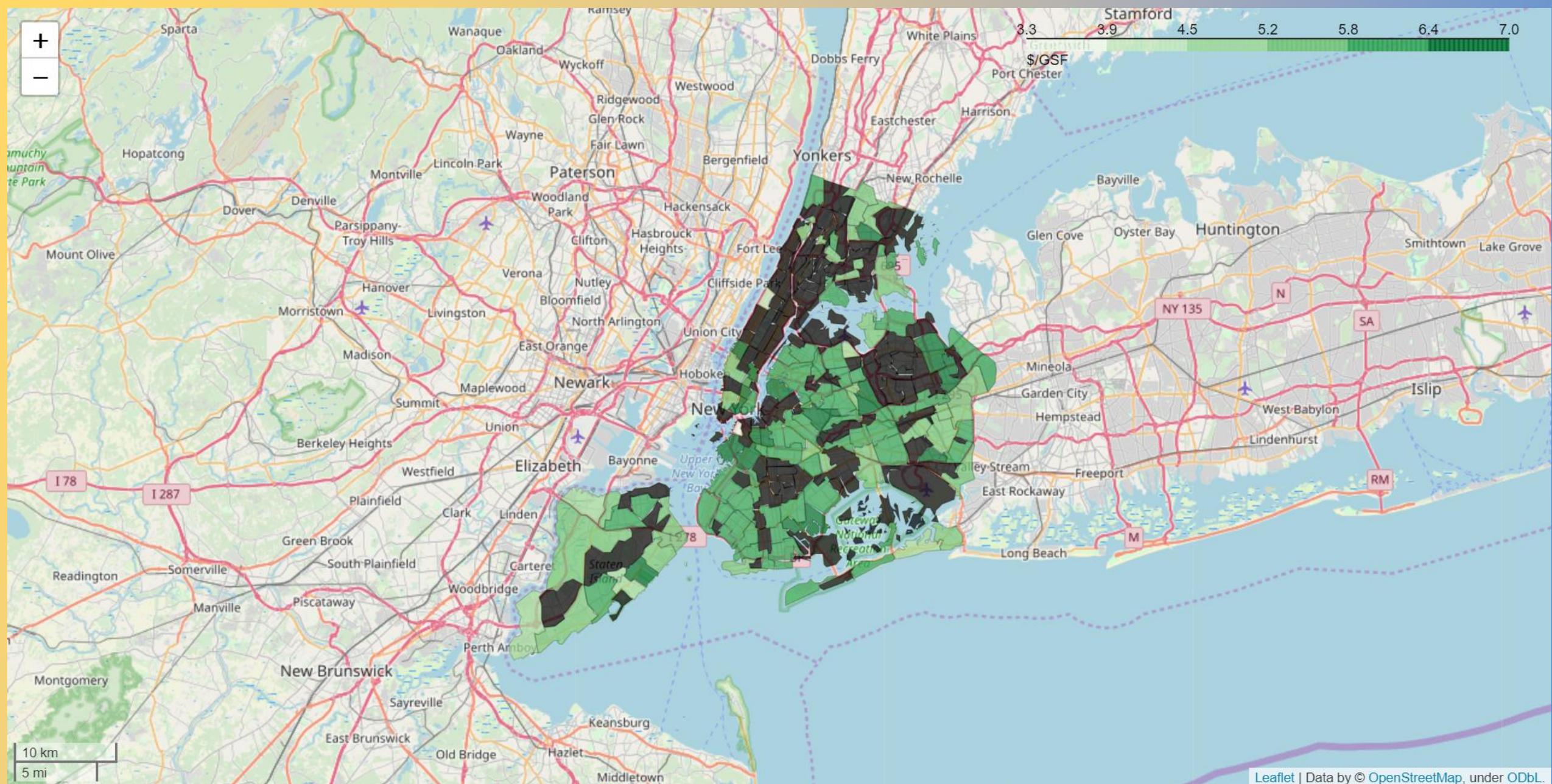
Results / Conclusion - Part Two

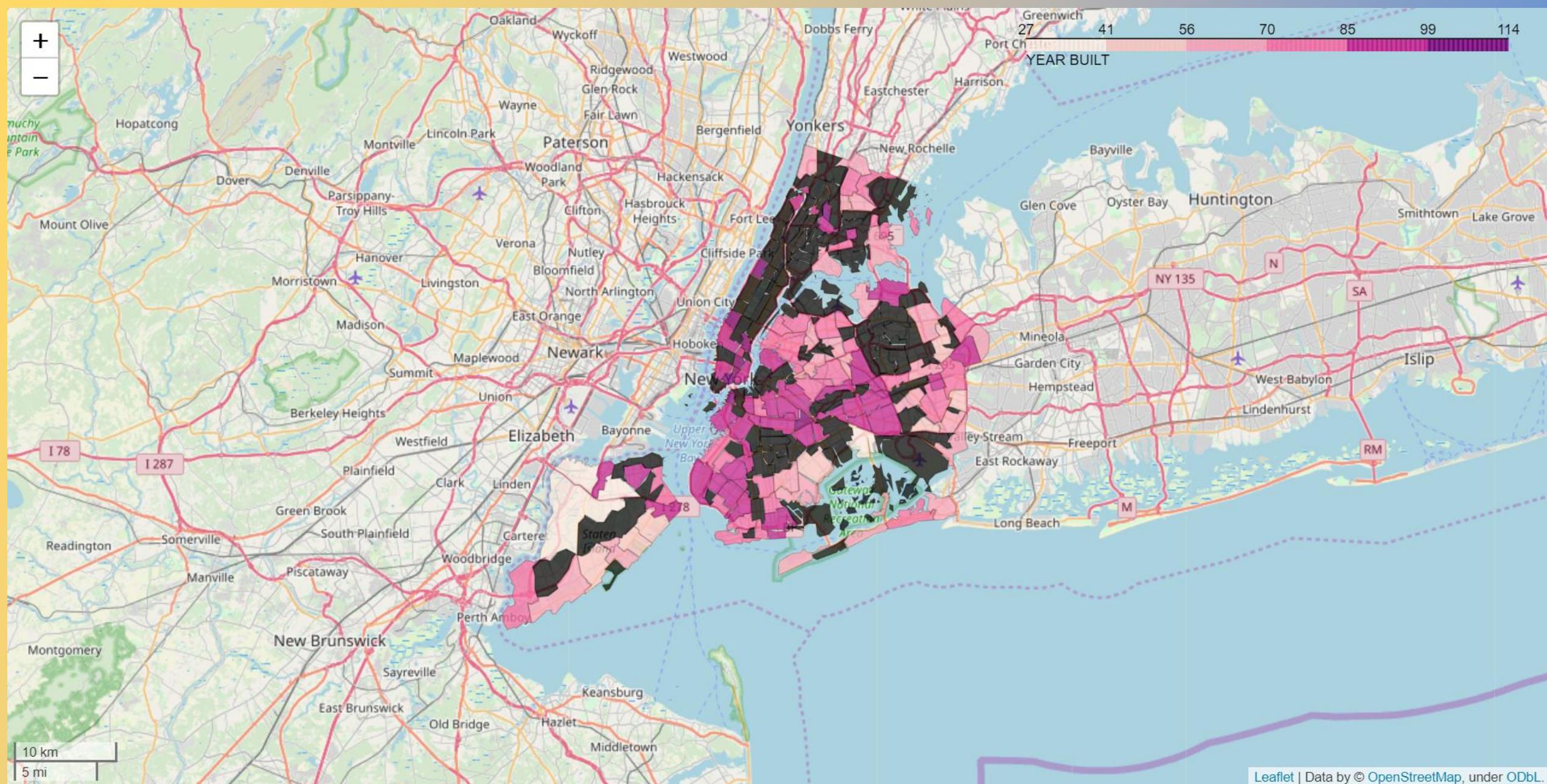
Besides using the data provided by Foursquare API, I try to build up some maps with New York real estate's pricing data, which might generally help business owners to make their decisions easier, no matter of what kind of business they want to pursue.

There are several parameters within rolling sales dataset, but I pick four as the most important: "LAND SQUARE FEET", "GROSS SQUARE FEET", "YEAR BUILT", and "SALE PRICE". In order to generating more useful analysis, certain methods are used to normalize all these variables.









Due to the differences between my rolling sales data source and geojson I download from officially government website, there are some missing areas on these maps. However, we can still tell some patterns & features based on them

- Avoid deep color area might be your initial ideas, but this is not always the case. Deep color area means higher "dollars per LSF", higher "dollars per GSF", and older "building age". However, if you filter the best place for your business in this way, then you might put your restaurant in Staten island, which makes no sense since a restaurant there will have very low potential customer numbers. This is not what we want to see.
- There are still some "good locations" in Queens & Brooklyn & Bronx if we put these 3 graphs with "average ratings" map together and check very carefully (those areas with lower prices, younger average building age, medium average rating). However, since the "average ratings" map is not a unbiased one and we lack other kinds of data (like security situation, customer average spending etc.), this kind of conclusion might be misleading to some degree. Detailed preparation is needed but these areas worth watching.
- Manhattan is expensive, full of sense of history, but it definitely contains the largest potential customer group. To those business owners who are rich enough, Manhattan is always their first & best choice.



New York City is place full of "treasure",
everyone can find their benefits in here.