**BATTLE OF NEIGHBOURHOODS CAPSTONE PROJECT**

## 1. Introduction and Business Problem

What is the best place to open a new italian restaurant? Let's assume the investor is choosing between different boroughs of New York.

The restaurants business is highly competitive, while entry barriers are comparably low. Therefore to solve the problem I am going to analyse how competitive the boroughs are taking into consideration amount of italian restaurans per capita.

Although there is a number of other metrics that can be used, I believe that the described approach would provide a good general conclusion about which location maximizes chances of success for the new business.

## 2. Data

I will use the following data:

1) Population https://en.wikipedia.org/wiki/Boroughs_of_New_York_City

| | Borough | Population |
|---|---|---|
| 0 | Bronx | 1418207 |
| 1 | Brooklyn | 2559903 |
| 2 | Manhattan | 1628706 |
| 3 | Queens | 2253858 |
| 4 | Staten Island | 476143 |

2) NY boroughs and neighborhoods coordinates https://cocl.us/new_york_dataset
3) Foursquare data for existing Italian restaurants and their locations
4) Geospatial data necessary for building a map: world map and NY boroughs json https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm

That would result in density numbers that will allow to (A) compare the boroughs and (B) put restaurants on the map to illustrate the conclusion.

For example, the number of restaurants in staten Island under category “Italian restaurant” will be divided by population in this borough. The obtained metric will allow to compare locations based on the relative market saturation.

## 3. Methodology

i) Import and preprocess population data for each NY borough
ii) Import and preprocess neighborhoods and boroughs data
iii) Use neighborhoods coordinates to find venues nearby using Foursquare

iv) Process venues data: we need (1) a dataframe that contains boroughs and number of italian restaurants in each of them, and (2) full list of italian restaurants in NY with coordinates

v) Use dataframe (1) and population data to calculate density of italian restaurants in each borough.

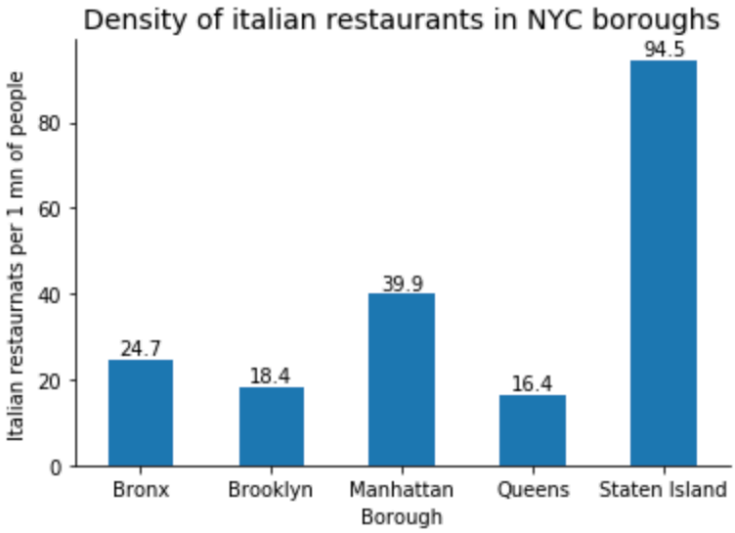
vi) Visualize the solution with a barchart

vii) Visualize the solution with a chronoplet map

## 4. Results

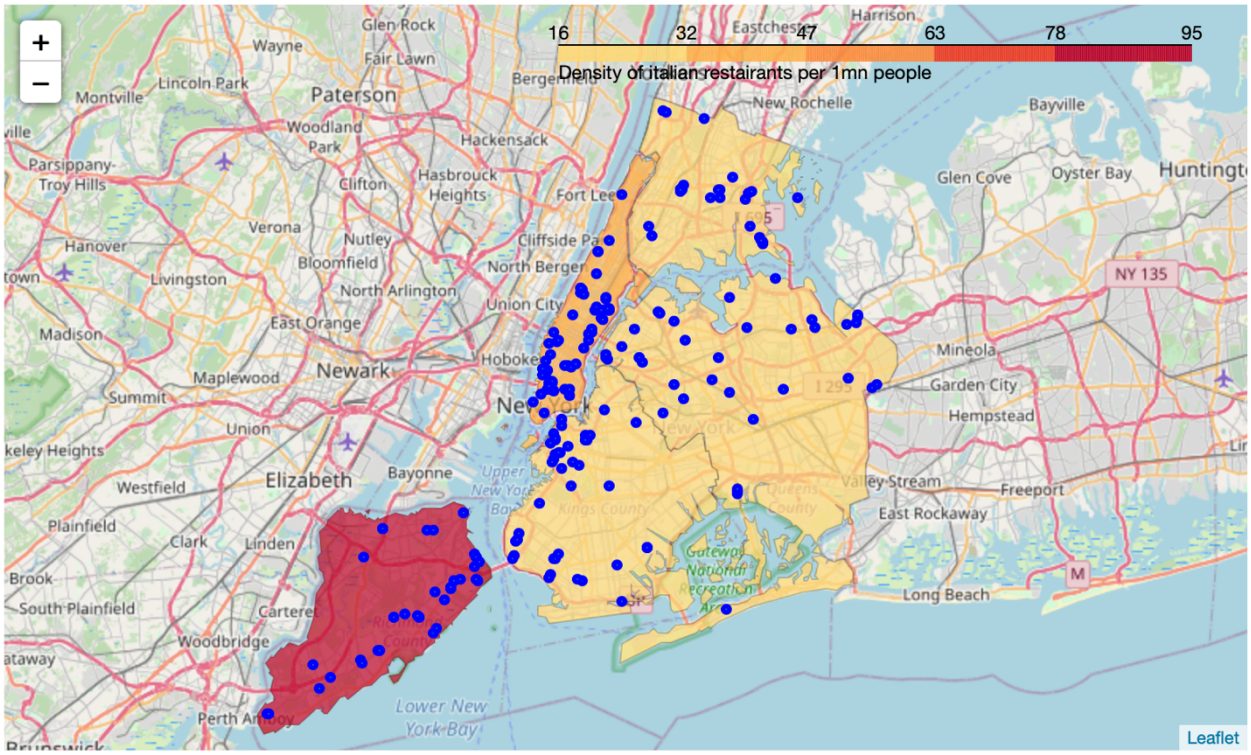
I found out that the italian restaurants are unevenly distributed across NY boroughs:

| | Borough | Number | Population | Per mn of people |
|---|---|---|---|---|
| 0 | Bronx | 35 | 1418207 | 24.679049 |
| 1 | Brooklyn | 47 | 2559903 | 18.360071 |
| 2 | Manhattan | 65 | 1628706 | 39.908983 |
| 3 | Queens | 37 | 2253858 | 16.416296 |
| 4 | Staten Island | 45 | 476143 | 94.509423 |




It's clear the density of venues of the considered category is the least at Queens and Brooklyn. The conclusion is that the one who opens a new italian restaurant in these location will have the highest chances to prosper due to the lowest competetion.

The results of the observation are also evident on a map listed below.




## 5. Discussion

Building the research I faced the limitations the Foursquare API has. Thus, I couldn't download rating for venues as the number of requests exceeeded 50. Such property as rating would enrich the research, as it narrows the range of potential competitors.

Another factor that would be useful for a potential investor is the cost of opening a new venue in different boroughs, as low density in some of them could be a result of high rent.

## 6. Conclusion

Working on this project I practiced the skills in processing data and visualizing the results. As the data could be bulky, all methosed used become very convinient.