CAPSTONE PROJECT THE BATTLE OF THE NEIGHBORHOODS

Date : 03 January 2021

Name : Miyoung Shin

INTRODUCTION

In this project I will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an **Italian restaurant** in **New York City**, United States. Since there are lots of restaurants in Manhattan, we will try to detect **locations that are not already crowded with restaurants**. We are also particularly interested in **areas with no Italian restaurants in vicinity**. I will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

DATA

Based on definition of our problem, factors that will influence our decision are:

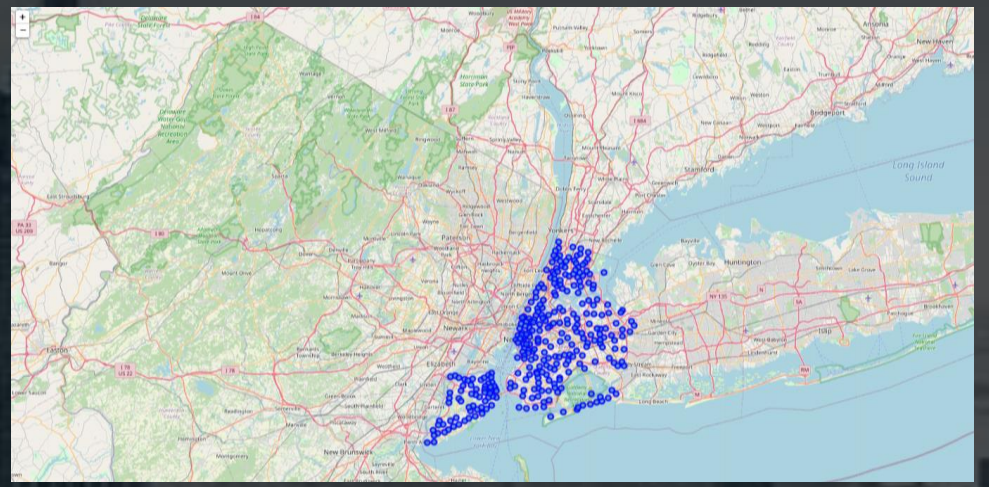
- number of existing restaurants in the neighborhood (any type of restaurant)
- number of and distance to Italian restaurants in the neighborhood, if any
- distance of neighborhood from Manhattan
- Following data sources will be needed to extract/generate the required information:
- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Google Maps API reverse geocoding**
- number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**
- coordinate of Manhattan will be obtained using **Google Maps API geocoding** of NYC location.

METHODOLOGY

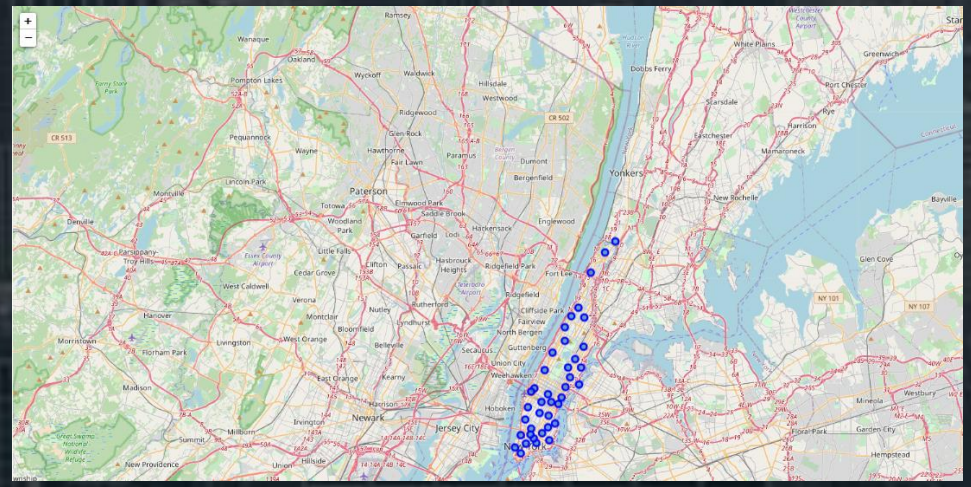
- In this project we will direct our efforts on detecting areas of New York City that have low restaurant density, particularly those with low number of Italian restaurants. We will limit our analysis to Manhattan area.
- In first step we have collected the required **data: location and type (category) of every restaurant within 1 km from Washington Heights**. We have also **identified Italian restaurants** (according to Foursquare categorization).
- Second step in our analysis will be calculation and exploration of '**restaurant density**' across different areas of Manhattan - we will use **folium** to identify a few promising areas close to each neighborhood with low number of restaurants in general (*and* no Italian restaurants in vicinity) and focus our attention on those areas.
- In third and final step we will focus on most promising areas and within those create **clusters of locations that meet some basic requirements** established in discussion with stakeholders: we will take into consideration locations with **no more than three restaurants in radius of 1000 meters**. We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final the categorized neighborhood in Manhattan area and search for optimal venue location by stakeholders.

RESULTS (1)

- Map of New York with Neighborhood :



- Segment and cluster only the neighborhood in Manhattan.



RESULTS (2)

- Find out how many unique categories can be curated from all the returned venues.
- Out of 315 unique venues, grouping rows by neighborhood and by taking the mean of the frequency of occurrence of each category.
- Pick the interest neighborhoods along with the top 5 most common venues.
- Create a new data that includes the cluster as well as the top 10 venues for each neighborhood.

CONCLUSION

- Purpose of this project was to identify the low number of restaurants (particularly Italian restaurants) close to ***Washington Heights*** area in order to aid stakeholders in narrowing down the search for optimal location for a new Italian restaurant. By calculating restaurant density distribution from Foursquare data and considering the frequency of each categorized venue :

----Washington Heights----

	venue	freq
0	Pizza Place	0.06
1	Café	0.05
2	Latin American Restaurant	0.05
3	Bakery	0.05
4	Grocery Store	0.03

- Clustering of those locations was then performed in order to create major frequency of restaurant nearby and addresses of those neighborhoods were created to be used as starting points for final exploration by stakeholders :