Data Science Restaurant Location in New York Capstone Project

Alexander Friev

Table of Contents

INTRODUCTION

New York City's demographics show that it is a large and ethnically diverse metropolis. It is the largest city in the United States with a long history of international immigration. New York City was home to nearly 8.5 million people in 2014, accounting for over 40% of the population of New York State and a slightly lower percentage of the New York metropolitan area, home to approximately 23.6 million. Over the last decade the city has been growing faster than the region. The New York region continues to be by far the leading metropolitan gateway for legal immigrants admitted into the United States.

Throughout its history, New York City has been a major point of entry for immigrants; the term "melting pot" was coined to describe densely populated immigrant neighborhoods on the Lower East Side. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world. English remains the most widely spoken language, although there are areas in the outer boroughs in which up to 25% of people speak English as an alternate language, and/or have limited or no English language fluency. English is least spoken in neighborhoods such as Flushing, Sunset Park, and Corona.

Opening a new restaurant could prove to be a challenge in any city in the world. The aim of this project is to determine the balance between location, attraction and competition. We want the location to be suitable for the restaurant, while trying to avoid unnecessary competition for this niche cuisine to ensure the profit is maximized.

To find the best spot, we need to analyze the neighborhoods of New York, the location of other restaurants and the most important attractions

METHODOLOGY

Identification of coordinates (latitudes and longitudes) for individual neighborhoods using geocoder package.

Cross referencing with US census Data

Using Foursquare location data to figure out venue details.

Filtering and categorizing restaurants.

Merging and analyze data for calculating rating of place.

Merging and superimposing data sets to create a map representative of neighborhoods with best rating place for new restaurant using Folium.

DATA

In order to address the goals of the project, the data required can be divided into three parts:

1. New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude. Data source: https://cocl.us/new_york_dataset Description : This data set contains the required information. And we will use this data set to explore various neighborhoods of New York City.

2. The coordinates for the city and the neighborhoods (latitudes and longitudes) – to be extracted with the help of the 'geocode' function in python.

3. The data regarding the various 'Restaurants' in the city – **Foursquare** location data will be leveraged to extract this information.

These information points combined will help in making the necessary classifications for the neighborhoods.

WORK FLOW

First download list of Borough, Neighborhood and transform dataset

	Borough	Neighborhood	Latitude	Longitude	restaurants
0	Bronx	Wakefield	40.894705	-73.847201	0
1	Bronx	Co-op City	40.874294	-73.829939	0
2	Bronx	Eastchester	40.887556	-73.827806	0
3	Bronx	Fieldston	40.895437	-73.905643	0
4	Bronx	Riverdale	40.890834	-73.912585	0
5	Bronx	Kingsbridge	40.881687	-73.902818	0
6	Manhattan	Marble Hill	40.876551	-73.910660	0
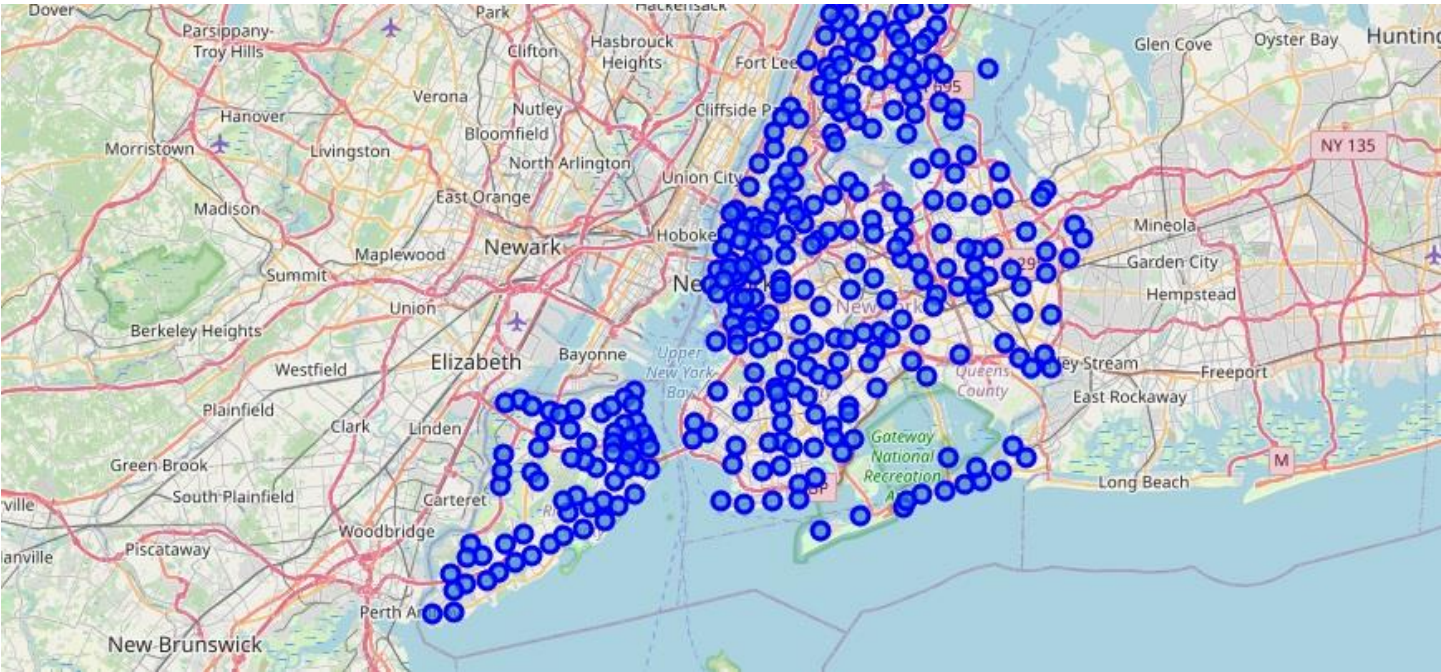
Download Foursquare venues (Italian Restaurant) and delete repeated by id.

|	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Central Deli	40.896728	-73.844387	Deli / Bodega
1	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
2	Wakefield	40.894705	-73.847201	Subway	40.897792	-73.855219	Sandwich Place
3	Wakefield	40.894705	-73.847201	Golden Krust Caribbean Restaurant	40.903773	-73.850051	Caribbean Restaurant
4	Wakefield	40.894705	-73.847201	Burger King	40.895540	-73.856460	Fast Food Restaurant

Count restaurants by Borough and Neighborhood

	Borough	Neighborhood	Latitude	Longitude	restaurants	Gross Domestic Product	Population
0	Bronx	Wakefield	40.894705	-73.847201	17	0.0	0
1	Bronx	Co-op City	40.874294	-73.829939	3	0.0	0
2	Bronx	Eastchester	40.887556	-73.827806	29	0.0	0
3	Bronx	Fieldston	40.895437	-73.905643	8	0.0	0
4	Bronx	Riverdale	40.890834	-73.912585	13	0.0	0

Number of Neighborhood for each Borough in New York City
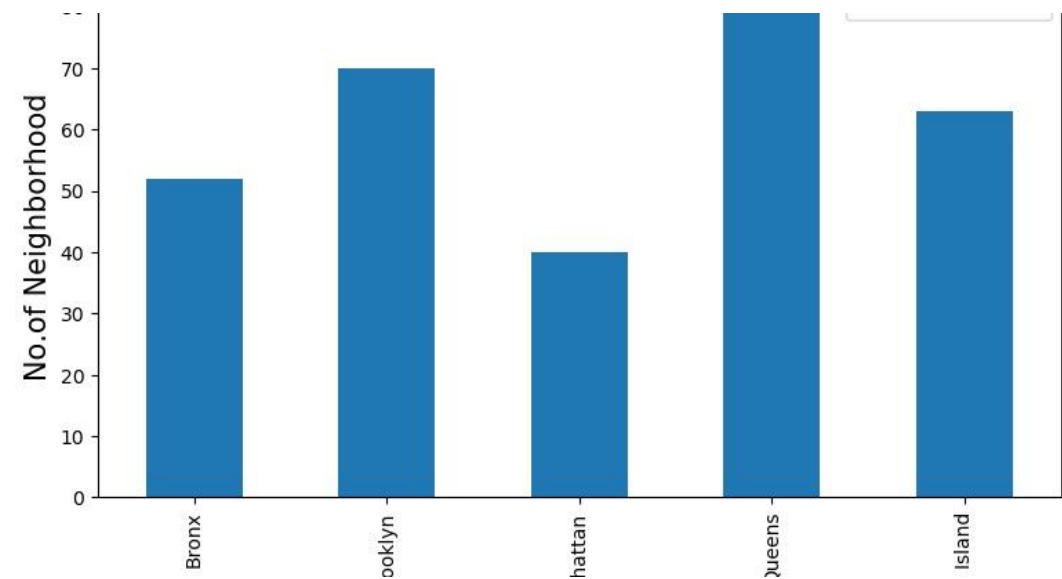
Italian Restaurant for each Borough in New York City
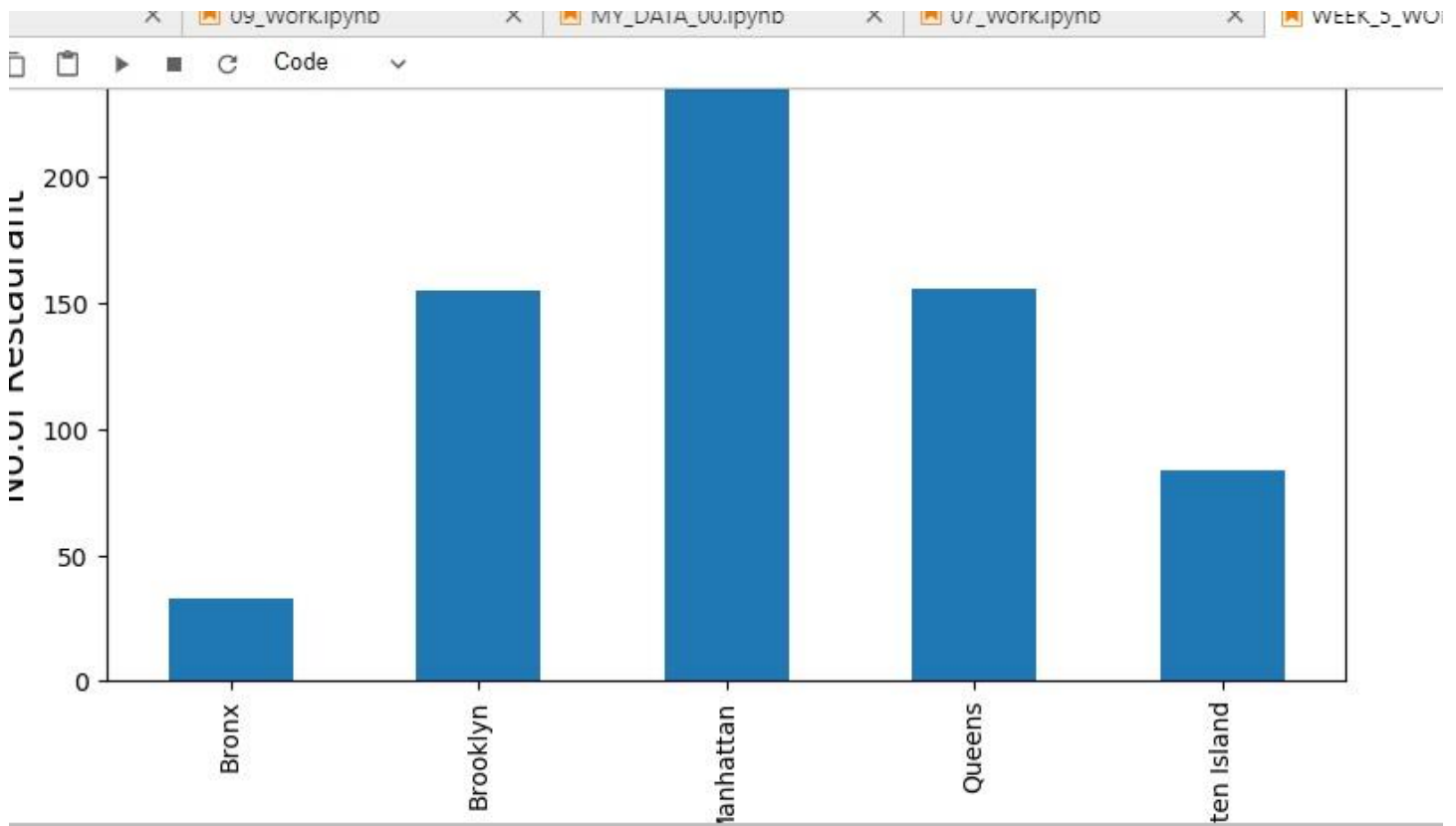
We can see Bronx has a Min Italian restaurants, but we will try to calculate rating

By Gross Domestic Product and Population.

RESULTS

index		Borough	Population	Gross Domestic Product	Land area	latitude	longitude	restaurants	Rating
0	1	Manhattan	1628701	600.244	22.83	40.789624	-73.959894	254	3848889
1	4	Bronx	1432132	42.695	42.10	40.846651	-73.878594	33	1852875
2	0	Brooklyn	2582830	91.559	70.82	40.650104	-73.949582	155	1525686
3	2	Queens	2278906	93.310	108.53	40.749824	-73.797634	156	1363107
4	3	Staten Island	476179	14.514	58.37	40.583456	-74.149605	84	82276

It is certain that - Manhattan more attractive even already have more restaurants than Bronx, Brooklyn, Queens, Staten island. By high family income in area. Bronx on second place, with less restaurants, if we will add rent, transport, food cost, result may change.

DISCUSSION

The map with the cluster of neighborhoods and restaurants can be very revealing. It clearly depicts the neighborhoods with a higher concentration of restaurants. At the same time, it neatly indicates the localities that the stakeholder population can choose from to build up on their food experiences.

In addition to these, the neighborhood clusters are also indicative of the high concentration of restaurants in a particular neighborhood. This can also help the potential investors identify spots which are not so crowded with restaurants and can help them narrow down an investment area.

Another potential of the project and the exercise is that it can help entrepreneurs further filter the data of the types of restaurants and the cuisines they offer, and help them decide which restaurant and what cuisine would they like to offer in case they want to venture in a neighborhood with already high number of restaurants.

CONCLUSION

The project uses the Foursquare data on an as it is basis. The study is good as a pilot, but more rigor needs to be added to further refine the results for more robust conclusions. The neighborhoods do indicate the presence of several restaurant categories, however, some eateries or cafes, or even restaurants could have remained excluded due to a different description of venue. Such categories need to be refined and added for a better representation of the ground reality.

The project model can lay a basis for further analysis with addition of more data points, and with integration of a richer location data. The analysis went well, but not taking into account factors like prices or accessibility. These are both factors that could alter the conclusions entirely for those interested in open new restaurants in New York.

REFERENCES

United States Census
Foursquare API