Open a New Restaurant in Manhattan

Biying Deng

September 25, 2019

1 Introduction

Background

New York City is the most popular and the most densely populated major city in the United States. It has a great impact in multiple field such as entertainment, fashion, financial, media, tourism and especially famous for its commerce. Meanwhile, Manhattan, which is the most representative and well-known boroughs in New York City. We will focus on this borough for this project.

Business Problem:
Location for restaurant is one of the most important key factors in real world, which directly affect sustainability and success in the future. To help our investors in deciding a better location to open a new restaurant, we will utilize different type of data. For this project, our problem will target on Manhattan's neighborhood to explore what are the difference among these neighborhoods. Considering that if we want to pick target neighborhoods for new restaurant, what are possible factors that we need to consider. This project can provide a quick exploration of 40 major neighborhoods in Manhattan by considering different statistics of each neighborhoods.

Target audience:
Investors from all around the world, who want to pick a neighborhood and open a new restaurant in Manhattan.

2. Description of data

- New York Neighborhood data that provide previously in Coursera course, and we narrow down to neighborhoods within Manhattan, which includes 40 major neighborhoods. Source:https://geo.nyu.edu/catalog/nyu_2451_34572

- Obtaining restaurants information, type, location in neighborhood using Foursquare API

- 'zipcode' data was from Wikipedia that I search one by one according to the Neighborhood. Use to locate each specific neighborhood and helps in grabs other demographic data online.

- Income data was searched one by one from website 'https://www.incomebyzipcode.com/' by passing the zipcode from above. It consists Per-capita income, median household and average household income.
- Part of the Population data are from below, and part of them are searched from Wikipedia. It will show approximately how many residents will be in each neighborhood. https://www.worldatlas.com/articles/manhattan-neighborhoods-by-population.html

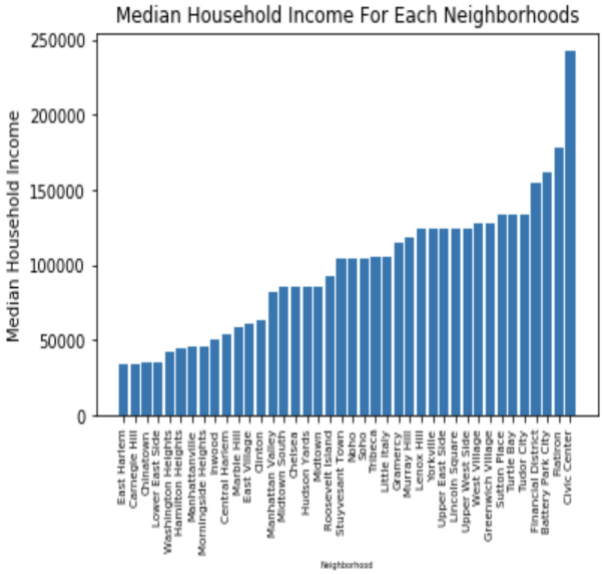
	Borough	Neighborhood	Latitude	Longitude	Zipcode	Median Household Income	Average Household Income	Per-capita Income	Population
0	Manhattan	Marble Hill	40.876551	-73.910660	10463	58881	82103	33630	9481
1	Manhattan	Chinatown	40.715618	-73.994279	10002,10013	35449	65718	30677	100000
2	Manhattan	Washington Heights	40.851903	-73.936900	10032,10033,10040	41890	60656	23217	158318
3	Manhattan	Inwood	40.867684	-73.921210	10034,10040	50909	65924	26216	46746
4	Manhattan	Hamilton Heights	40.823604	-73.949688	10031,10032,10039	44040	64971	26417	48520

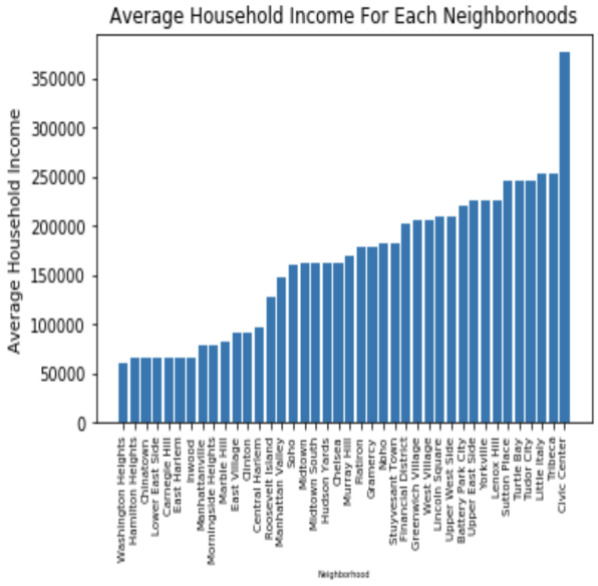
3. Methodology

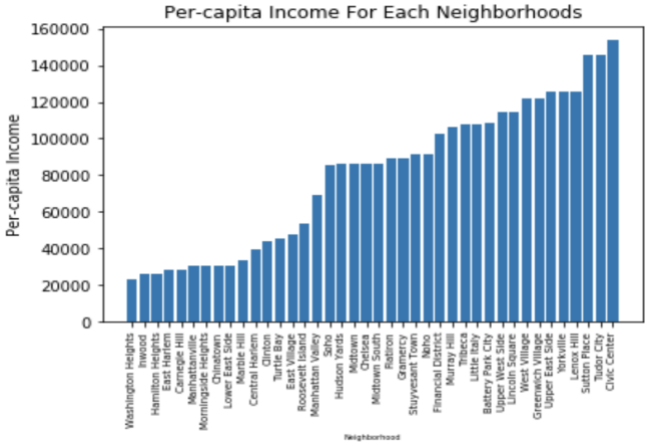
For Data cleaning, Since I scraped most of the demographic data one by one according to each neighborhood, there is no data cleaning needed. Then we plot histogram of all income statistics (Median household income, Average household income, Per-capita income) we have in data and see what's the difference for each neighborhood, which are the top 5 neighborhoods. For crime complaints data, I created an interactive crime maps with folium, which enhance the visualization that shows which neighborhoods have more number of complaints. After EDA of demographic data, by combing our new_york dataset, we scrape venues by using Foursquare API for each neighborhood. Based on location and information of venues such as type of venues, we find Top 10 most common venue type in each neighborhood. Lastly, by applying k-means clustering method to cluster venue, we can find out what type of venues are in each clusters and what neighborhoods are being clustered together.

4. Result

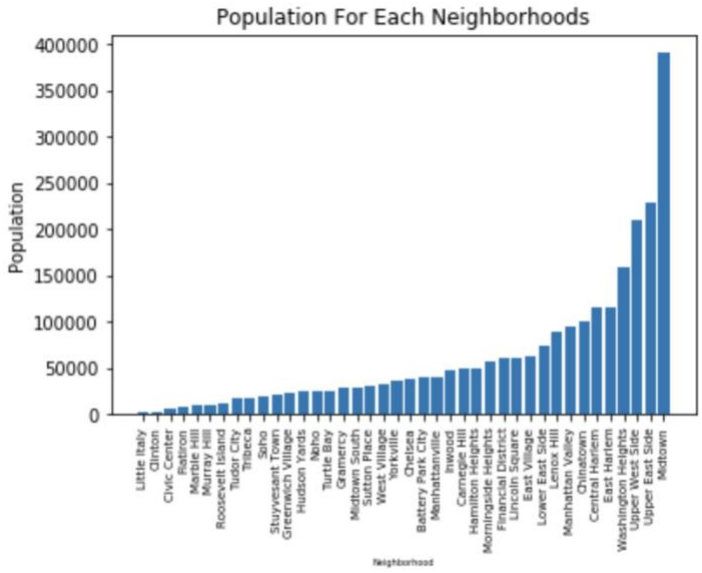
The purpose I include household income data is because It is possible that people with higher income may have more money or more likely to try a new restaurant. With higher surrounding household income, it will be better for business sustainability of this restaurant as well. From the below plot, it shows how each neighborhoods' median household income different from each other. Top five neighborhoods are Turtle Bay, Tudor City, Financial District, Battery Park City, Radio Row and Civic Center.



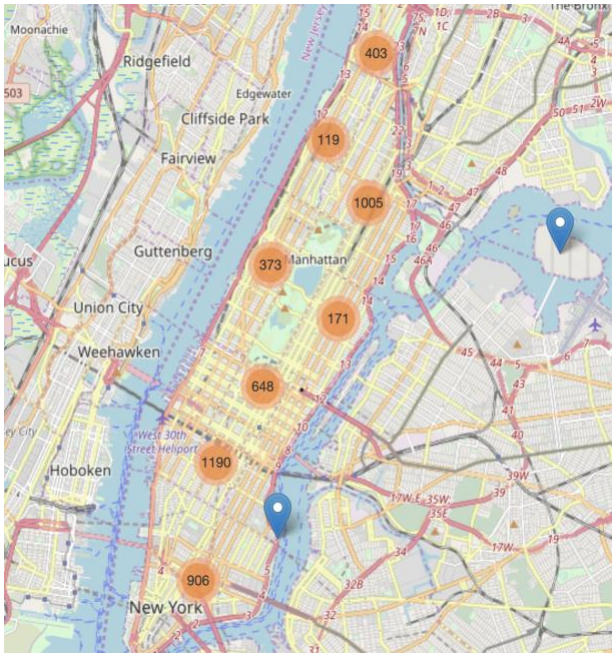




As the plot of average household income for each neighborhood in Manhattan shown, the top 5 neighborhoods with highest average household income are Turtle Bay, Tudor City, Little Italy, Tribeca and Civic Center. As the plot of per-capita income for each neighborhood shown, Top 5 neighborhoods with highest per-capita income are Yorkville, Lenox Hill, Sutton Place, Tudor City, and Civic Center. In general, there are still slightly different ranking base on the above statistics.



Another factor that may affect local business of restaurant will be the local population density in each neighborhood. More residents around may bring more new customers for this restaurant. From the above plot, top 5 neighborhoods with the most population are East Harlem (Same as Central Harlem), Washington Heights, Upper West Side, Upper East Side and Midtown.



By creating this interactive crime map (Right) and neighborhoods map(Left), we can see how many crime (felony, misdemeanor, violation) in specific area reported to NYPD so far till this year 2019. We can observe that Midtown, Midtown South and Central Harlem area tend to have more crime complaints by looking at both map, while Upper West side, Lincoln Square

and Yorkville area have less crime complaints. In fact, since Midtown is where Time Square located, and Midtown South has Korean town and Penn Satiation, both are attractive to visitors from all around the world and residents to shop around. It is possible that increase rate of crime happened due to high density of population and tourism.

Top 10 most common venue for higher income neighborhoods

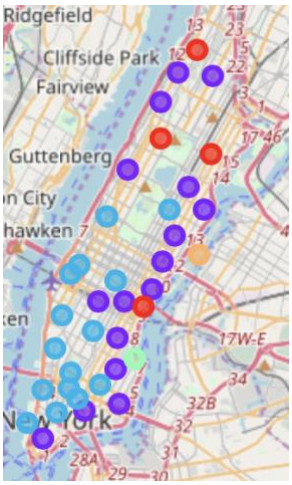
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	Civic Center	Gym / Fitness Center	Italian Restaurant	Coffee Shop	French Restaurant	Hotel	Sandwich Place	Sporting Goods Shop	American Restaurant	Cocktail Bar	Park
9	Financial District	Coffee Shop	Pizza Place	Wine Shop	Hotel	Gym	Event Space	American Restaurant	Gym / Fitness Center	Cocktail Bar	Steakhouse
33	Tudor City	Park	Café	Mexican Restaurant	Sushi Restaurant	Greek Restaurant	Deli / Bodega	Pizza Place	Dog Run	Diner	Gym
34	Turtle Bay	Italian Restaurant	Sushi Restaurant	Steakhouse	Coffee Shop	Wine Bar	Café	Japanese Restaurant	Park	French Restaurant	Ramen Restaurant

From the above, it shows top 10 most common venues in each neighborhood. If we look at those common neighborhoods with higher household income such as Civic Center, Turtle Bay, Financial District and Tudor City. The common venue they have in common are Italian Restaurant, Café, Coffee Shop, Steakhouse, Wine Bar, etc.

Top 10 most common venue for less crimes neighborhood

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
7	Lincoln Square	Theater	Gym / Fitness Center	Café	Plaza	Concert Hall	Italian Restaurant	Park	Performing Arts Venue	Indie Movie Theater	French Restaurant
6	Upper West Side	Italian Restaurant	Wine Bar	Coffee Shop	Indian Restaurant	Bakery	Bar	Vegetarian / Vegan Restaurant	Thai Restaurant	Pub	Bookstore
9	Yorkville	Italian Restaurant	Gym	Bar	Coffee Shop	Pizza Place	Deli / Bodega	Sushi Restaurant	Diner	Mexican Restaurant	Wine Shop

From the above it shows top 10 most common venues within these neighborhood with less crime reported or safer for people to eat around are Upper West side, Lincoln Square and Yorkville area. The common venue they have in common are Italian Restaurant and Coffee Shop.



K-means clustering neighborhoods

By applying K-means clustering, we can cluster similar neighborhoods with similar type of venue into different clusters. Then we could utilize this information to help our investors to decide which neighborhood is more profitable for their restaurant concept, or even could help them to known what types of restaurant are more popular in each neighborhood. From the above picture, we can observe that the majority neighborhoods were clustered into Cluster 1 (Purple), Cluster 2 (Blue) and Cluster 3(Red), with only 1 neighborhood in each cluster 4 and 5. In general, cluster1 includes neighborhoods with majority venue such as Café, Mexican Restaurant, Pizza Place, Indian Restaurant and Deli / Bodega. Cluster 2 includes neighborhoods that are characterized by venue such as Coffee Shop, Bar, Wine Bar, Gym/Fitness Center. Cluster 3 includes neighborhoods with majority venues in these neighborhoods are Italian Restaurant, Clothing Store, Gym / Fitness Center, Coffee Shop/Café.

5.Discussion

Observation:

- 1.Neighborhoods characteristics (less crime, higher population density, higher household income) can be associate with surrounding restaurant concept.
- 2.Base on observation, neighborhoods with higher household income and less crime complaints tend to have more Italian restaurant and Coffee shop. Investors could make use of this information when deciding their restaurant concept or restaurant location.
- 3.Midtown (Time Square) and Midtown South (Korean town and Penn Station) are the most representative popular place that includes a lot of different type of restaurants. They are both favorite place for visitors and residents to shop around, even though they may high in crime complaints according to our crime map.

Recommendation for future investigation:

- We could use Elbow Method to find the optimal K to optimize our clustering results.
- Include demographic data when do our clustering.
- Include more data (transportation, rent space price, etc.)

- Build a basic recommendation system to help investors in decision-making.

6. conclusion

Base on the exploration so far, even though we already have some useful information about what kind of restaurant are most common and how the surrounding residents and environments are in each neighborhood. It is still need more investigation and exploration on different kind of data to better help investors in decision-making. This report can only provide a general direction in helping investors to make decision on where to open their new restaurant, or what kind of restaurant concept they can pick if it is opened in specific neighborhoods base on our clustering results. However, it is a broad topic that we can investigate deeper in different way. Results in this report may not be perfect and may need further improvement in general.

7. Reference

New York Neighborhood data: https://geo.nyu.edu/catalog/nyu_2451_34572
Income data : https://www.incomebyzipcode.com/
Part of Population data: https://www.worldatlas.com/articles/manhattan-neighborhoods-by-population.html