

## THE BATTLE OF NEIGHBORHOODS – WHERE TO OPEN A RESTAURANT IN NEW YORK



RODRIGO QUINTELLA – July 2019

## Table of contents

1. Introduction .....	3
2. Business Understanding and Analytic Approach .....	3
2.1. Restaurants .....	3
2.2. New York City .....	3
3. Data Requirements .....	5
3.1. Geographic Data.....	5
3.2. Demographic Data.....	6
4. Methodology.....	7
4.1. Data Collection .....	7
4.2. Data Preparation .....	8
5. Results .....	9
6. Discussion.....	11
7. Conclusion .....	11
8. References.....	12

## 1. Introduction

New York (NY) is one of the richest cities in the world. Businesses easily flourish, as the population and demand for services grow. But if stakeholders plan to open a restaurant in New York City (NYC), where would be the best place? There are, apparently, a lot of options due to the city's huge population and occupied area. A restaurant, though, is a serious investment and implies in heavy periodical costs and aggressive competition, so to define the right place, and additionally, the right category (e.g. Italian Food, Chinese Food) are the first steps to ensure that a dream will come true.

This report intends to explore NY's neighborhoods and to show some of their characteristics to finally suggest better places and categories to open a restaurant.

## 2. Business Understanding and Analytic Approach

### 2.1. Restaurants

"A restaurant or an eatery, is a business which prepares and serves food and drinks to customers in exchange for money. Meals are generally served and eaten on the premises, but many restaurants also offer take-out and food delivery services. Restaurants vary greatly in appearance and offerings, including a wide variety of cuisines and service models ranging from inexpensive fast food restaurants and cafeterias, to mid-priced family restaurants, to high-priced luxury establishments." (Wikipedia, 2019)

If a business like a restaurant offers such variety of alternatives, is highly recommended a thorough study before any investment is made. It's important to understand the population characteristics of the target city, as well as their preferences.

### 2.2. New York City

"The City of New York, usually called either New York City (NYC) or simply New York (NY), is the most populous city in the United States. With an estimated 2018 population of 8,398,748 distributed over a land area of about 302.6 square miles (784 km<sup>2</sup>), New York is also the most densely populated major city in the United States. Located at the southern tip of the state of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass and one of the world's most populous megacities, with an estimated 19,979,477 people in its 2018 Metropolitan Statistical Area and 22,679,948 residents in its Combined Statistical Area. A global power city, New York City has been described as the cultural, financial, and media capital of the world, and exerts a significant impact upon commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. The city's fast pace has inspired the term New York minute. Home to the headquarters of the United Nations, New York is an important center for international diplomacy." (Wikipedia, 2019)

Geography	April 1, 2010		Population Estimate (as of July 1)								
	Census	Estimates Base	2010	2011	2012	2013	2014	2015	2016	2017	2018
New York city, New York	8,175,133	8,174,988	8,190,355	8,272,963	8,348,032	8,398,739	8,437,387	8,468,181	8,475,976	8,438,271	8,398,748

Figure 1 – New York City population by year. Source: (Fact Finder, 2019)

2010 Census Data	New York City compared				
	New York City	Los Angeles	Chicago	New York State	United States
Total population	8,175,133	3,792,820	2,695,598	19,378,102	308,745,538
Population, percent change, 2000 to 2010	+2.1%	+2.6%	-6.9%	+2.1%	+9.7%
Population density	27,012 /sq. mi.	8,092 /sq. mi.	11,864 /sq. mi.	408.7 /sq. mi.	87.4 /sq. mi.
Median household income (1999)	\$38,293	\$36,687	\$38,625	\$43,393	\$41,994
Per capita income (1999)	\$22,402	\$20,671	\$20,175	\$23,389	\$21,587
Bachelor's degree or higher	27%	26%	26%	27%	24%
Foreign born	36%	41%	21.7%	20%	13%
White	44.6%	49.8%	45.0% <sup>[42]</sup>	66.4%	72.4%
Black	25.1%	9.6%	32.9%	15.5%	12.6%
Hispanic (any race)	27.5%	48.5%	28.9%	17.3%	16.3%
Asian	11.8%	11.3%	5.5%	5.9%	4.8%

Figure 2 – New York City compared. Source: (Wikipedia, 2010)



Figure 3 – Pictures from New York City. Source: (Wikipedia, 2019)



The demand for all kinds of services is huge but the competition for customer's attention is fierce and tough. New York is a city of neighborhoods, which offers lots of business options and different demographic features. Therefore, it's of great importance to explore all the alternatives and be sure of what are the goals to run any kind of businesses there and a restaurant is not an exception.

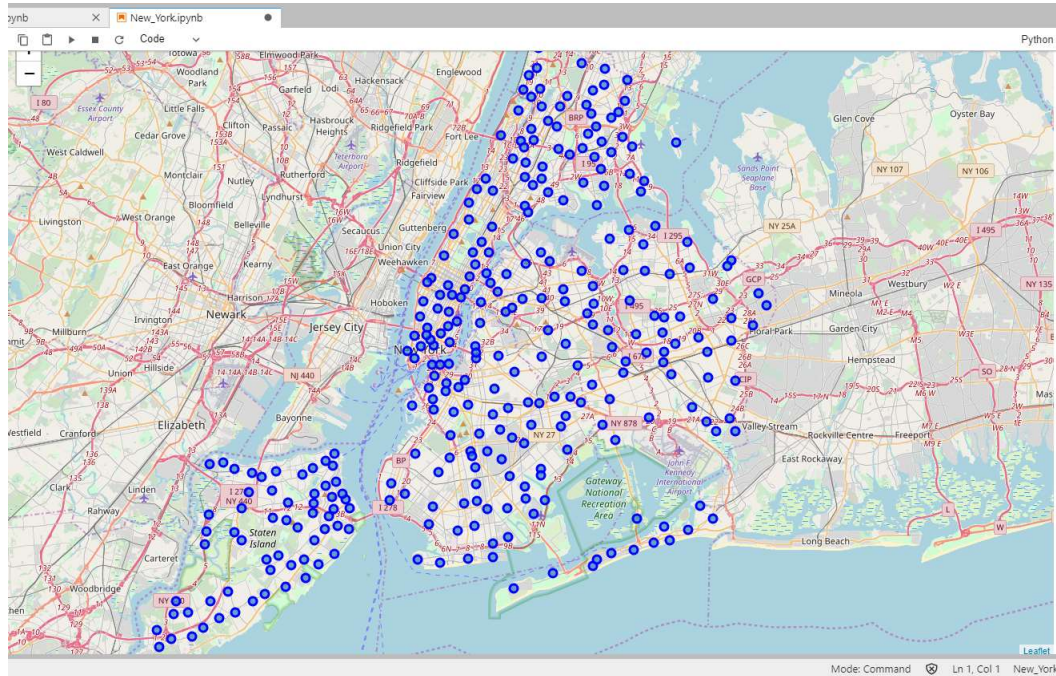


Figure 4 – Neighborhoods on New York Map. Source: Final assignment

### 3. Data Requirements

#### 3.1. Geographic Data

Since the work involves to choose a place for a restaurant, it's essential to have geographic data in hands. This means to combine location coordinates with other kinds of information like population, income and so on.

Firstly, the basic information involves the New York City neighborhoods and their centroid latitude and longitude. Using the dataframe analyzed during the Capstone Project, it's possible to retrieve the necessary information.

```
[6]: neighborhoods.head()
```

```
[6]:
```

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Figure 5 – Initial geographic dataframe. Source: Final assignment

Then, the Foursquare tool will be necessary to explore the venues in each neighborhood. By using API queries, it's possible to aggregate venues location data and their categories, which means the kind of business of each venue (e.g. Hotel, Restaurant, School).

After retrieving the venues information and combining them with the neighborhoods data, it will be possible to filter the eatery venues and have interesting visualizations, like the different restaurant categories placed on their respective neighborhoods.

### 3.2. Demographic Data

To make more accurate analysis, it will be necessary to use demographic data, including population number, ethnical and income data. This information will be combined with the geographic data using the NTA (Neighborhood Tabulation Areas) codes. Since the NTA codes are not unique for each neighborhood in the basic dataframe, it will be necessary to perform some neighborhood combination by working with NTA codes geographic boundaries.

The goal of this project is to make statistical inferences in order to find correlations between population number, income, ethnical features and eatery categories in order to determine which is the best place to open a restaurant in New York. The demographic data will be downloaded from NYC Open Data website.

This study is of great interest for any stakeholder interested in investing on a restaurant in New York City.

## 4. Methodology

### 4.1. Data Collection

To make a rich collection of venues in New York City, the Foursquare queries were made by borough separately:

	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bronx	Wakefield	40.894705	-73.847201	Lollipop Gelato	40.894123	-73.845892	Dessert Shop
1	Bronx	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
2	Bronx	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
3	Bronx	Wakefield	40.894705	-73.847201	Cooler Runnings Jamaican Restaurant Inc	40.898276	-73.850381	Caribbean Restaurant
4	Bronx	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop

Figure 6 – Venues collection for Bronx Borough. Source: Final Assignment

After collecting the venues for each of the 5 boroughs in New York, there were 10,408 venues to be analysed.

Then, accessing the NYC Open Data site, geographic data concerning the NTA codes for the neighborhoods were downloaded. This NTA codes were downloaded as polygon coordinates, which delimited the area of each neighborhood.

	A	B	C	D	E
1	GeoType	GeogName	GeoID	Borough	Tot_Pop
2	NTA2010	Bath Beach	BK27	Brooklyn	33.163
3	NTA2010	Bay Ridge	BK31	Brooklyn	79.134
4	NTA2010	Bedford	BK75	Brooklyn	75.318
5	NTA2010	Bensonhurst East	BK29	Brooklyn	64.267
6	NTA2010	Bensonhurst West	BK28	Brooklyn	91.646
7	NTA2010	Borough Park	BK88	Brooklyn	105.913

Figure 7 – Population data. Source: (NYC Open Data, 2018)

	A	B	C	D	E	F	G	H	I	J	K
1	GeoID	HHIU10E	HHI10t14E	HHI15t24E	HHI25t34E	HHI35t49E	HHI50t74E	HHI75t99E	HI100t149E	HI150t199E	HHI200pIE
2	BK27	895	618	1.223	927	1.290	2.074	1.352	1.758	706	585
3	BK31	2.313	1.573	3.235	3.047	3.542	5.570	4.108	4.827	2.497	2.775
4	BK75	3.673	2.317	3.385	2.509	3.124	3.706	2.269	2.847	1.195	1.034
5	BK29	2.356	1.413	2.718	2.185	2.655	3.898	2.558	2.416	1.144	844
6	BK28	2.330	1.746	3.383	3.246	4.036	5.117	3.428	3.743	1.649	1.231
7	BK88	2.208	2.630	4.619	3.829	3.442	3.888	2.221	2.751	978	856
8	BK19	2.781	1.404	1.522	1.273	1.391	2.147	995	1.228	689	667

Figure 8 – Economic data. Source: (NYC Open Data, 2018)

	A	B	C	D	E	F	G	H	I
1	GeolD	Pop_Hsp	Pop_Wht	Pop_BlK	Pop_Ind	Pop_Asn	Pop_Pac	Pop_Oth	Pop_2Rc
2	BK27	5.861	14.490	404	11	11.975	29	71	322
3	BK31	14.134	50.197	1.511	181	11.515	19	151	1.426
4	BK75	14.682	26.407	30.566	66	2.290	0	149	1.158
5	BK29	7.923	29.894	531	0	24.858	0	348	713
6	BK28	13.378	39.112	801	111	37.132	0	319	793
7	BK88	10.244	80.190	946	140	13.614	0	162	617
8	BK19	3.903	23.746	261	149	5.105	0	142	851

Figure 9 – Ethnical data. Source: (NYC Open Data, 2018)

## 4.2. Data Preparation

As the number of neighborhoods listed in Foursquare were different from the NTA codes, it was necessary to group the initial neighborhoods by NTA codes. This was made by applying the centroid coordinates of the NTA codes for the neighborhoods and, for the ones which it wasn't possible, applying the closest neighborhood within the same borough.

```
[43]: # Now let's group the names of Neighborhoods by NTA Code
nta_neigh_grouped = neighborhoods.groupby(['NTA Code', 'Borough', 'Latitude', 'Longitude'])['
nta_neigh_grouped.head()
```

[43]:	NTA Code	Borough	Latitude	Longitude	Neighborhood
0	BK09	Brooklyn	-73.994871	40.695469	Brooklyn Heights
1	BK17	Brooklyn	-73.941511	40.588300	Sheepshead Bay, Canarsie, Mill Island, Manhatt...
2	BK19	Brooklyn	-73.961216	40.580923	Brighton Beach
3	BK21	Brooklyn	-73.991236	40.576483	Coney Island, Sea Gate
4	BK25	Brooklyn	-73.964334	40.599954	Homecrest

Figure 10 – Grouped Neighborhoods by NTA Code. Source: Final Assignment.

There were also in NYC Open Data site the demographic data broke by NTA Codes. The data were downloaded and cleaned up to be combined with the venues dataframe. This data included total population, segmented income data and also ethnical data.

The main goal of the project were to find any correlations involving the number of venues, venues categories, population, income profile and ethnical features of each neighborhood.

In addition, the most common restaurant categories for each NTA code were listed in order to analyse the competition to be confronted or avoided.



## 5. Results

Although there wasn't identified any important correlation involving population, income or ethnical features, it was identified a strong correlation between the number of restaurants in each neighborhood and the number of other venues.

```
[ ]: from scipy import stats
# To avoid errors calculating the pearson coefficient, let's drop the 0 rows
nta_econ_other_venues = nta_econ_other_venues[nta_econ_other_venues['Restaurant'] != 0]
# Calculating pearson coefficient and p_value
pearson_coef, p_value = stats.pearsonr(nta_econ_other_venues['Restaurant'], nta_econ_other_venues['Venue'])
print("The Pearson Correlation Coefficient is", pearson_coef, " with a P-value of P =", p_value)

The Pearson Correlation Coefficient is 0.9553395236294961 with a P-value of P = 1.5735208178602133e-76
```

**So, the restaurants tend to be near other venues !!!**

Figure 11 – Correlation between number of restaurants and number of other venues. Source: Final Assignment.

So, by counting the number of venues for each NTA code, it was possible to identify the “hottest” neighborhoods for opening a restaurant.

```
[72]: # Now, let's see the top 10 neighborhoods Locations
top_10_neigh = nta_econ_other_venues_sorted.head(10)

top_10_neigh = pd.merge(top_10_neigh, nta_neigh_grouped, how='left', on='NTA Code')
top_10_neigh
```

	NTA Code	Restaurant	Tot_Pop	Venue	Borough	Latitude	Longitude	Neighborhood
0	MN27	264.0	44898.0	652.0	Manhattan	-73.994630	40.715607	Marble Hill, Chinatown, Inwood, Yorkville, Len...
1	MN24	132.0	42979.0	364.0	Manhattan	-74.004747	40.720766	Greenwich Village, Tribeca, Little Italy, Soho...
2	MN13	98.0	77205.0	275.0	Manhattan	-73.999021	40.747585	Clinton, Chelsea, Flatiron, Hudson Yards
3	BK38	54.0	42245.0	242.0	Brooklyn	-73.985987	40.692927	Downtown, Boerum Hill, Vinegar Hill, Dumbo
4	QN15	49.0	54264.0	183.0	Queens	-73.755317	40.602455	Howard Beach, College Point, Far Rockaway, Bro...
5	BK33	59.0	42463.0	182.0	Brooklyn	-74.003899	40.678772	Cobble Hill, Carroll Gardens, Red Hook
6	MN17	48.0	27811.0	150.0	Manhattan	-73.983503	40.755734	Midtown, Midtown South
7	BX59	61.0	28437.0	146.0	Bronx	-73.848643	40.832100	City Island, Mott Haven, Port Morris, Soundvie...
8	MN40	48.0	58593.0	144.0	Manhattan	-73.961174	40.774730	Upper East Side, Carnegie Hill
9	BK73	59.0	54330.0	133.0	Brooklyn	-73.958625	40.714916	North Side, South Side

Figure 12 – The “hottest” neighborhoods for opening a restaurant. Source: Final Assignment.

```
[84]: ny_rest_rank = ny_rest['Venue Category'].value_counts().reset_index()
ny_rest_rank.columns = ['Venue Category', 'Count']
ny_rest_rank.head(10)
```

[84]:

	Venue Category	Count
0	Italian Restaurant	310
1	Chinese Restaurant	230
2	Mexican Restaurant	174
3	American Restaurant	170
4	Sushi Restaurant	116
5	Japanese Restaurant	98
6	Fast Food Restaurant	98
7	Thai Restaurant	89
8	Restaurant	85
9	Seafood Restaurant	82

Figure 13 – Most common restaurant categories in NYC. Source: Final assignment.

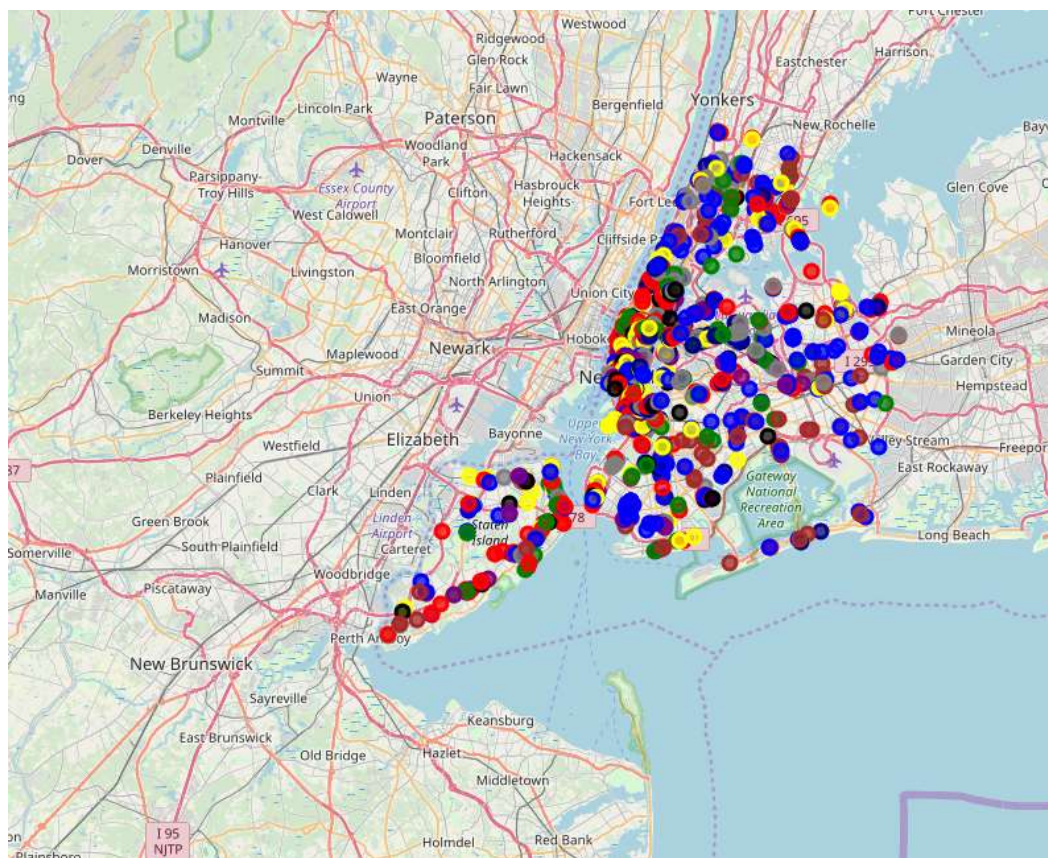


Figure 14 – Restaurant categories on NY map. Source: Final Assignment.

The most common restaurant categories were also listed for the “hottest” Neighborhoods.

NTA Code	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
BK33	Italian Restaurant	Thai Restaurant	American Restaurant	Sushi Restaurant	Restaurant
BK38	American Restaurant	Thai Restaurant	Mexican Restaurant	Japanese Restaurant	Italian Restaurant
BK73	American Restaurant	Chinese Restaurant	Japanese Restaurant	Sushi Restaurant	Mexican Restaurant
BX59	Italian Restaurant	Chinese Restaurant	Fast Food Restaurant	American Restaurant	Mexican Restaurant
MN13	Italian Restaurant	American Restaurant	Japanese Restaurant	Thai Restaurant	Restaurant
MN17	Japanese Restaurant	American Restaurant	Italian Restaurant	Sushi Restaurant	Restaurant
MN24	Italian Restaurant	American Restaurant	Sushi Restaurant	Chinese Restaurant	Japanese Restaurant
MN27	Italian Restaurant	Mexican Restaurant	Sushi Restaurant	Chinese Restaurant	American Restaurant
MN40	Italian Restaurant	Japanese Restaurant	Sushi Restaurant	American Restaurant	Mexican Restaurant
QN15	Chinese Restaurant	Italian Restaurant	Fast Food Restaurant	Mexican Restaurant	Japanese Restaurant

Figure 15 – Most common restaurant categories by neighborhoods. Source: Final Assignment.

## 6. Discussion

Based on the results presented, the recommendation to the stakeholders who plan to open a restaurant is to find a place in one of the “hottest” neighborhoods and choose one of the categories listed above, considering the most common type of restaurant for each neighborhood, which will be useful to avoid massive competition. For example, it’s recommended for one who wanted to open a Chinese restaurant (the second most common restaurant in New York) to try it in the area of the ‘MN13’ NTA code, which is Clinton, Chelsea, Flatiron and Hudson Yards in Manhattan. In that area, Chinese restaurants are not so common. Another recommendation could be open a Chinese restaurant in the area of ‘BK38’ NTA code area, including Downtown, Boerum Hill, Vinegar Hill, Dumbo in Brooklyn.

## 7. Conclusion

This case study will be very useful to restaurant stakeholders and could be extended to other venue categories, as other correlations could be identified and explored using statistical inferences and Data Science techniques.

## 8. References

(2010). Retrieved July 20, 2019, from Wikipedia:

[https://en.wikipedia.org/wiki/Demographics\\_of\\_New\\_York\\_City](https://en.wikipedia.org/wiki/Demographics_of_New_York_City)

(2018, September 19). Retrieved July 26, 2019, from NYC OPen Data:

<https://data.cityofnewyork.us/City-Government/Demographics-and-profiles-at-the-Neighborhood-Tabu/hyuz-tij8>

(2019). Retrieved July 20, 2019, from Fact Finder:

<https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=CF>

(2019). Retrieved July 20, 2019, from Wikipedia: <https://en.wikipedia.org/wiki/Restaurant>