New York Restaurant Data Project

for

ABC Multicuisine Inc

Table of Contents

1. Introduction: Business Problem

ABC Multicuisine Inc (hereafter will be referred to as the Company) is a successfully run food restaurant company that specializes in **Indian**, **Chinese**, **American,** and **Italian** cuisine. The Company is interested in exploring a suitable opportunity to start a new restaurant in the **New York** area by the end of Q3 of 2020.

2. Problem Background

The Company has been successfully running its restaurant business in Asia and Australia region and would like to enter the United States market by setting up its first restaurant in the **New York** region and then expand further in other parts of New York and other cities in the USA. As the company is a new entrant to this part of the world, they have engaged the data science team to research, study and come up with a recommendation on which area in New York would be best suited to open their first restaurant specializing in one among their core strength of **Indian**, **Chinese**, **American** and **Italian** cuisine.

New York City is the financial capital of the USA with a diversified population. It's one of the highest populated cities in the USA with several industries ranging from Finance, Software, Retail, Consumer, Tourism, and so on. The Company would like to make the decision by Q3 of 2020 and looks forward to the data science team to do a thorough analysis and come up with the recommendation in terms of the best location and best cuisine for the new restaurant that can help them gain market share, establish their brand values in New York and help them achieve their best return on investment.

3. Problem Description

The City of New York serves a variety of international cuisine food to its customers. As our company specializes and interested only in **Indian**, **Chinese**, **American,** and **Italian**, we will be focusing only on these four kinds of foods for our data analysis. The New York City is divided into five Boroughs namely:

- Bronx
- Brooklyn
- Manhattan
- Queens
- Staten Island

In order to compete with the existing players and gain market share for our Company and help them grow organically, as part of our data science project, We will be analyzing and taking into account the following areas with respect to each of the above-mentioned Boroughs:

- List of zip codes mapped to Boroughs
- Land Area of Boroughs
- Per Capita Income of People in Each Borough
- Persons Per Square Miles
- Total Population
- Existing Players per cuisine in the market segment of each Borough
- Compare Similarities and Dissimilarities between all five Boroughs

In short, As this will be the first project of the Company in this part of the world, it's very important that we come with the right recommendation in terms of the best location within the five Boroughs in New York and the best restaurant cuisine type within the four categories the Company specializes in that helps them gain market share and get a better return on investment.

4. Target Audience / Stakeholders

ABC Multicuisine Inc has chosen our data science team to understand, study and analyze their problem of finding the right location within New York to start their first restaurant in the USA region. Our objective is to come with the best possible recommendation based on the available data and our research and submit the report to the Board of Directors, Business Head of USA region, and their Executive Leadership team.

5. Success / Exit Criteria

The success criteria for the outcome of this data science project will be decided by the best location and the best category of cuisine recommendation provided by the team that caters the needs of the local population within that selected Borough and meets the demands of the Company's future customer segment.

6. Dataset / Data Provider

The following Data sets will be utilized for this project:

- New York Neighborhood Data
- Land Area / Population Density by Boroughs
- Population By Zip Codes for all Boroughs
- FourSquare Restaurant Categories Data

6.1 Zipcode Definition Data¶

The mapping of available New York zip codes and their corresponding Boroughs can be obtained from: here The New York City has been divided into five Boroughs namely:

- Bronx
- Brooklyn
- Manhattan
- Queens
- Staten Island

We will get all the zip codes that are mapped to their corresponding Boroughs and Neighborhood along with their location coordinates of Latitudes and Longitudes.

ZipCode	Borough	Neighborhood	Population	Density	Latitude	Longitude
10001	Manhattan	Chelsea and Clinton	21102	33959	40.741236	-73.356691
10002	Manhattan	Lower East Side	81410	92573	40.712728	-74.006015
10003	Manhattan	Lower East Side	56024	97188	40.712728	-74.006015
10004	Manhattan	Lower Manhattan	3089	5519	40.712728	-74.006015
10005	Manhattan	Lower Manhattan	7135	97048	40.712728	-74.006015

6.2 Land Area / Population Density by Boroughs

The following key data for each borough can be obtained from here

- Per Capita Income
- Land Area
- People Living Per Square Miles

Per Capita Income data measures the **average income earned per person** in a given area. It is calculated by dividing the area's total income by its total population. Population density is a measurement of population per unit area, or exceptionally unit volume; it is a quantity of type number density.

Borough	PerCapitaIncome	LandArea	PersonsPerSqM
Bronx	30100	42.10	33867
Brooklyn	35800	70.82	36147
Manhattan	368500	22.83	71341
Queens	41400	108.53	20767
Staten Island	30500	58.37	8157

6.3 Population data By Zip codes for All Boroughs

We will collect the following category of data from here

- Population (Number of people living in a given zip code area)
- Density (Number of people living per square mile in a given zip code area)

Population and density will provide us a clear picture of how densely each zip code area are populated.

ZipCode	Borough	Neighborhood	Population	Density	Latitude	Longitude
10001	Manhattan	Chelsea and Clinton	21102	33959	40.741236	-73.356691
10002	Manhattan	Lower East Side	81410	92573	40.712728	-74.006015
10003	Manhattan	Lower East Side	56024	97188	40.712728	-74.006015
10004	Manhattan	Lower Manhattan	3089	5519	40.712728	-74.006015
10005	Manhattan	Lower Manhattan	7135	97048	40.712728	-74.006015

6.4 Foursquare Venues Data By Restaurant Category

Forsquare.com provides access to firmographic data and rich community-sourced content for more than 60 million commercial places around the world—via flat file or API. We will be using their Places API that provides location data with the list of restaurant venues for a given restaurant category and Borough in JSON format. Since ABC Multi Cuisine Inc specializes only in a certain kind of cuisine, We are collecting the restaurant data for the following four categories of restaurants:

- **American**
- **Italian**
- **Chinese**
- **Indian**

ZipCode	Borough	Neighborhood	Latitude	Longitude	Name	Category
10009	Manhattan	Chelsea and Clinton	40.727510	-73.979324	Khiladi NYC	Indian
10024	Manhattan	Chelsea and Clinton	40.786166	-73.976414	Alachi Masala	Indian
10022	Manhattan	Chelsea and Clinton	40.755620	-73.968666	Amma	Indian
10009	Manhattan	Chelsea and Clinton	40.727285	-73.979602	Desi Galli - Avenue B	Indian
10016	Manhattan	Chelsea and Clinton	40.741393	-73.983367	Saravanaa Bhavan	Indian

ZipCode	Borough	Neighborhood	Latitude	Longitude	Name	Category
11238	Brooklyn	Northwest Brooklyn	40.681505	-73.955770	Golda	American
11211	Brooklyn	Northwest Brooklyn	40.710783	-73.953704	Lighthouse	American
11238	Brooklyn	Northwest Brooklyn	40.682846	-73.963835	Otway	American
11222	Brooklyn	Northwest Brooklyn	40.733427	-73.958201	Alameda	American
11238	Brooklyn	Northwest Brooklyn	40.681470	-73.955800	Hart's	American

Known assumptions

This project is done with the known API rate limiting imposed by foursquare.

7. Methodology

7.1 Business Understanding

Our primary aim of this project is to come with the right recommendation in terms of the best-desired location within the five Boroughs in New York and the best restaurant cuisine type within the four categories the Company specializes in that helps them gain market share and get a better return on investment.

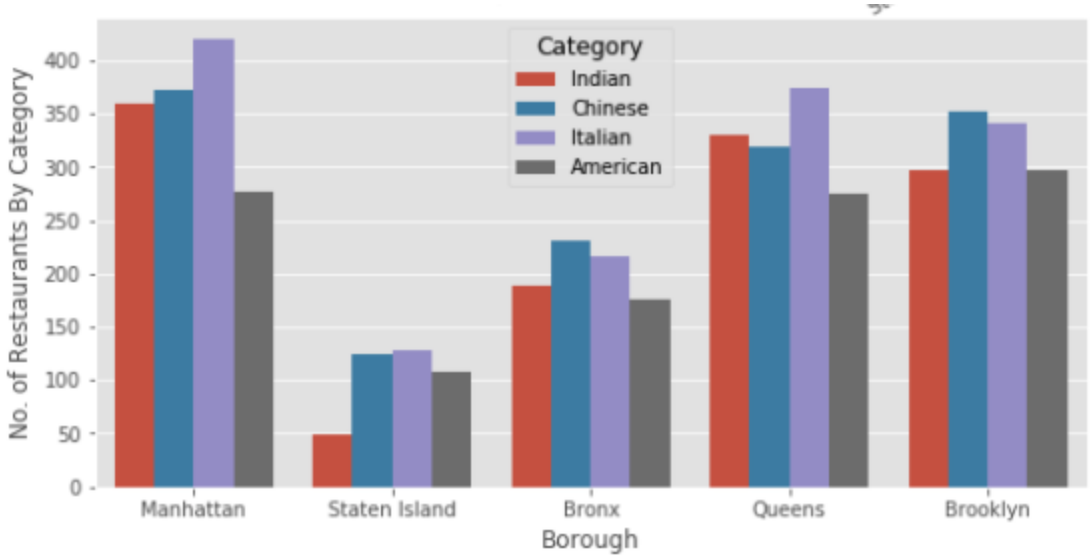
7.2 Exploratory Data Analysis

The City of New York is divided into five Boroughs and the company **ABC Multicuisine Inc** specializes only in **Indian**, **Chinese**, **American,** and **Italian** cuisine. Based on this, We will be collecting data with respect to the five Boroughs and the neighborhoods within those Boroughs. Restaurants data filtering out the above mentioned four categories will be collected and analyzed.

7.2 Data Visualization

Restaurants By Category

Restaurants data for the four major categories that we are interested in are collected using the Forsquare.com APIs. We used their Places API that provides location data with the list of restaurant venues for a given restaurant category and neighborhood. We then group the collected data by each Boroughs.



Restaurants Frequency By Neighborhood

For each of the neighborhoods within the five boroughs, we create the frequency data that provides us a clear indication in terms of which type of category has the higher and lower frequency for a given neighborhood.

	Neighborhood	American	Chinese	Indian	Italian
0	Borough Park	0.314286	0.157143	0.228571	0.3
1	Bronx Park and Fordham	0.000000	0.000000	1.000000	0.0
2	Bushwick and Williamsburg	0.314286	0.157143	0.228571	0.3
3	Canarsie and Flatlands	0.314286	0.157143	0.228571	0.3
4	Central Bronx	0.000000	0.000000	1.000000	0.0

Restaurants Ranking By Frequency

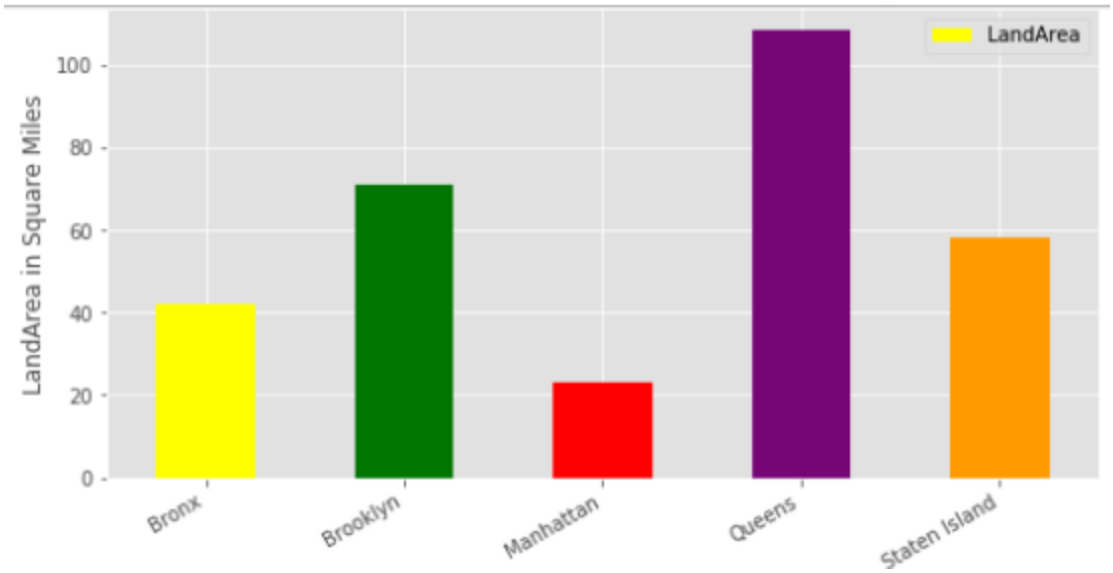
Based on the frequency of the type of category, we rank each neighborhood. For Example, in the Central Bronx neighborhood, Indian restaurants are most common and the Italian restaurant types are second most common among the chosen four categories.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
Borough Park	American	Italian	Indian	Chinese
Bronx Park and Fordham	Indian	Italian	Chinese	American
Bushwick and Williamsburg	American	Italian	Indian	Chinese
Canarsie and Flatlands	American	Italian	Indian	Chinese
Central Bronx	Indian	Italian	Chinese	American

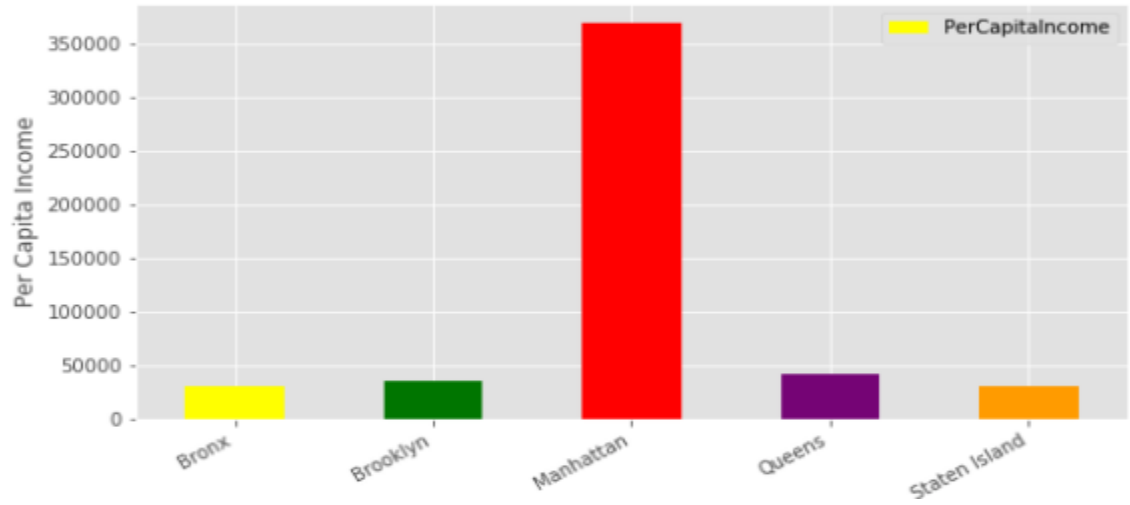
Borough specific metrics

We then collected, analyzed and visualized the borough specific metrics like total land area of boroughs, per capita income, total population and population density from here

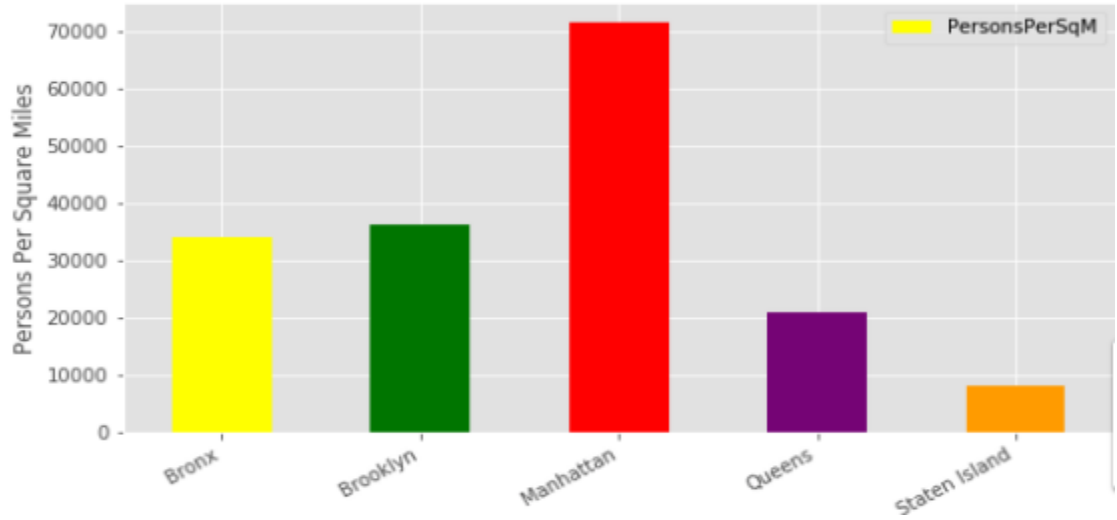
Land Area of Boroughs



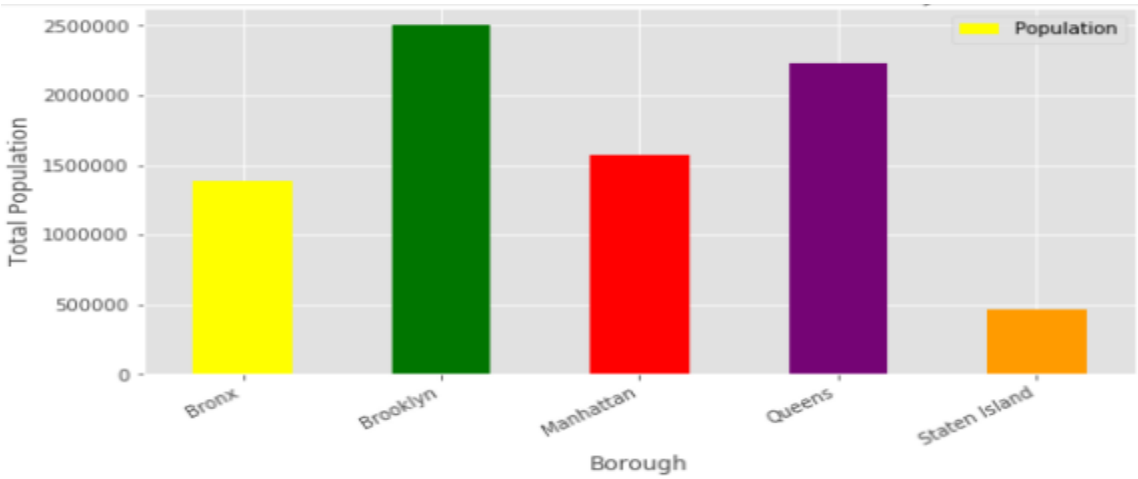
Per Capita Income by Boroughs



Persons Per Square Mile By Boroughs



Population By Boroughs

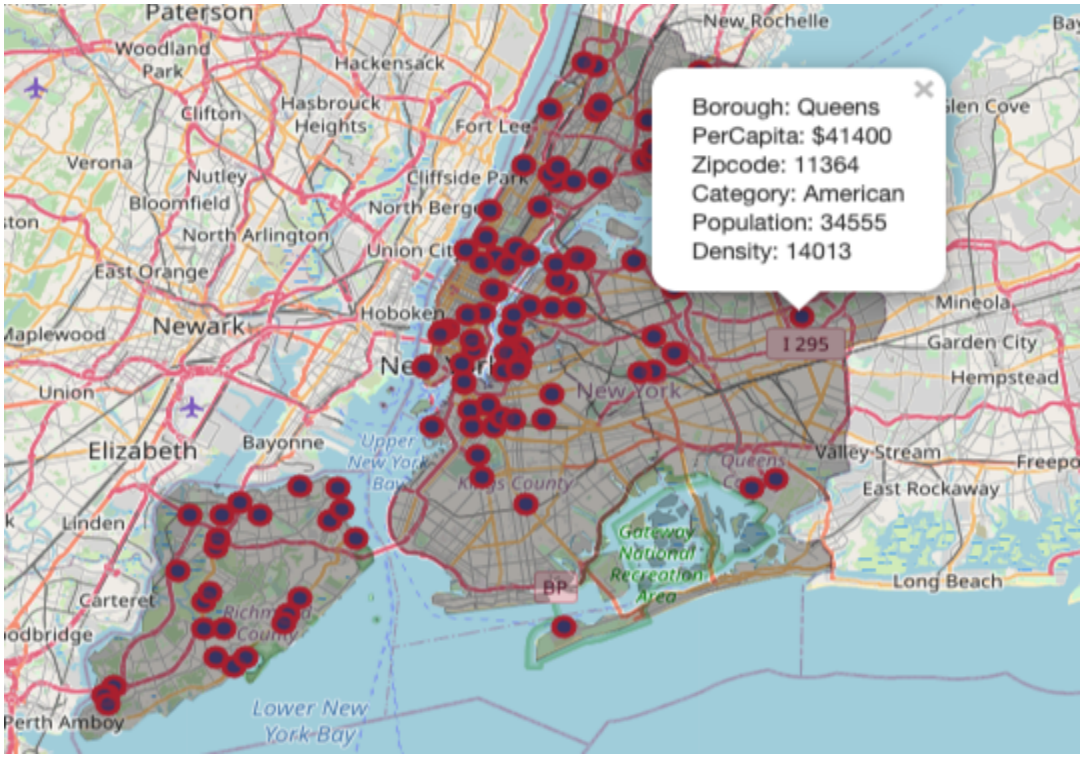


Restaurants data visualization By Category

- Use New York's borough boundaries geo data and choropleth to mark the boundaries of boroughs
- Use folium map and mark all the restaurant venue location
- Each category of restaurants are color coded with their respective country flag color
- Pop up added for each location with important data related to that location

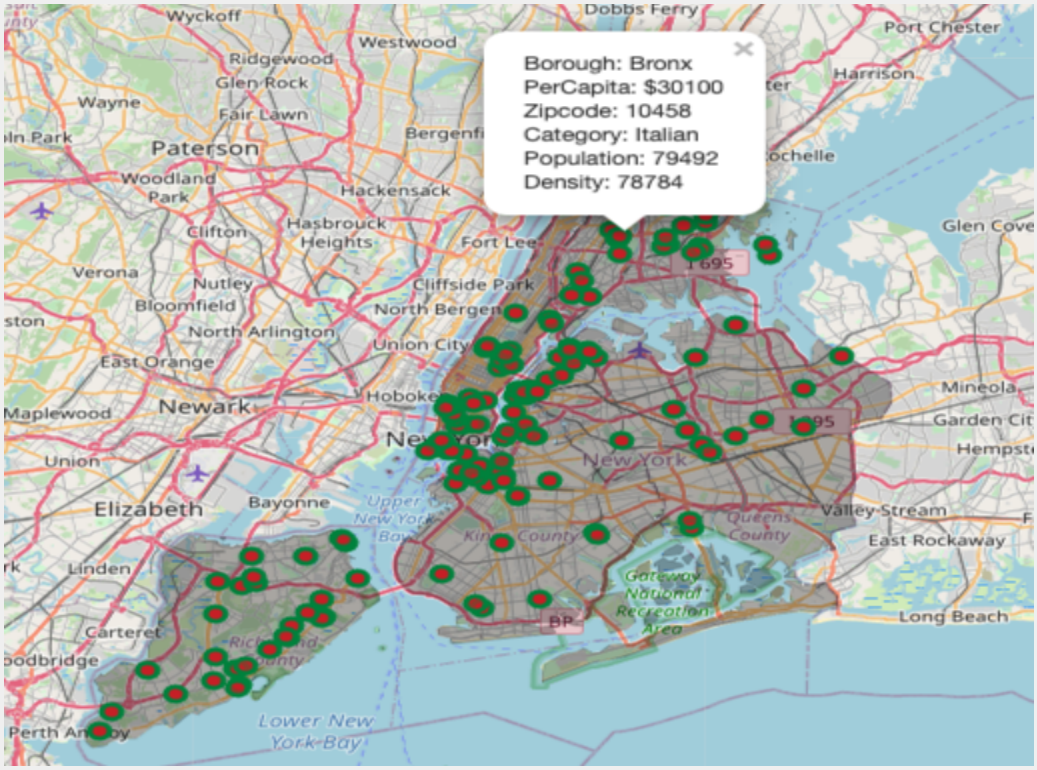
American Category Restaurants in New York Boroughs

In the neighborhood of zip code 11364, the per capita income for that borough is $41,400 and the population in that zip code is 34,555 and the population density is 14,013.



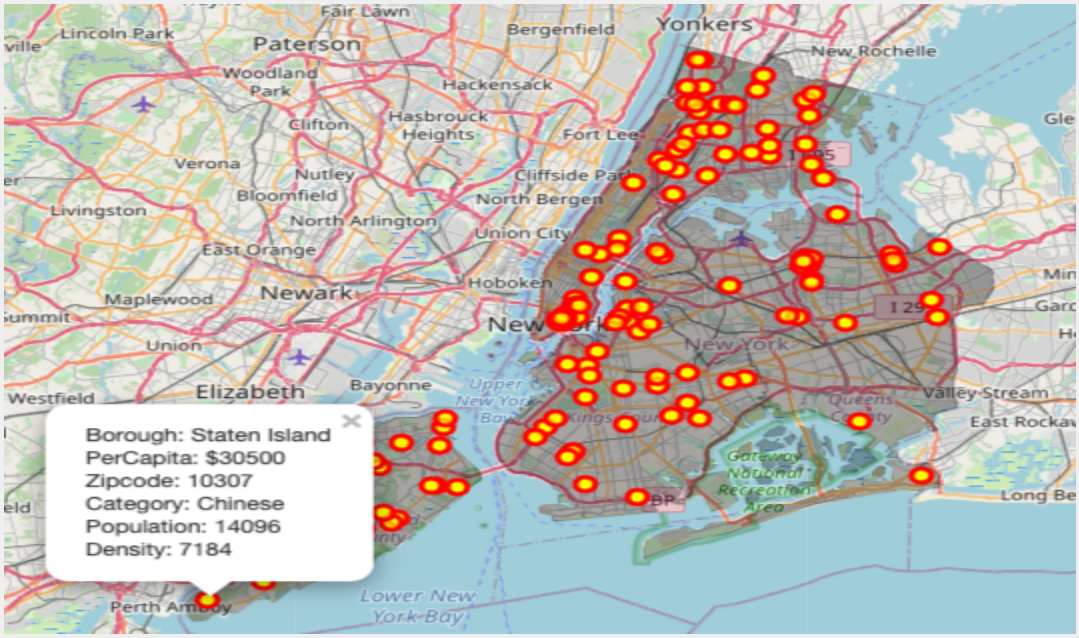
Italian Category Restaurants in New York Boroughs

In the neighborhood of zip code 10458, the per capita income for that borough is $30,100 and the population in that zip code is 79492 and the population density is 78784.



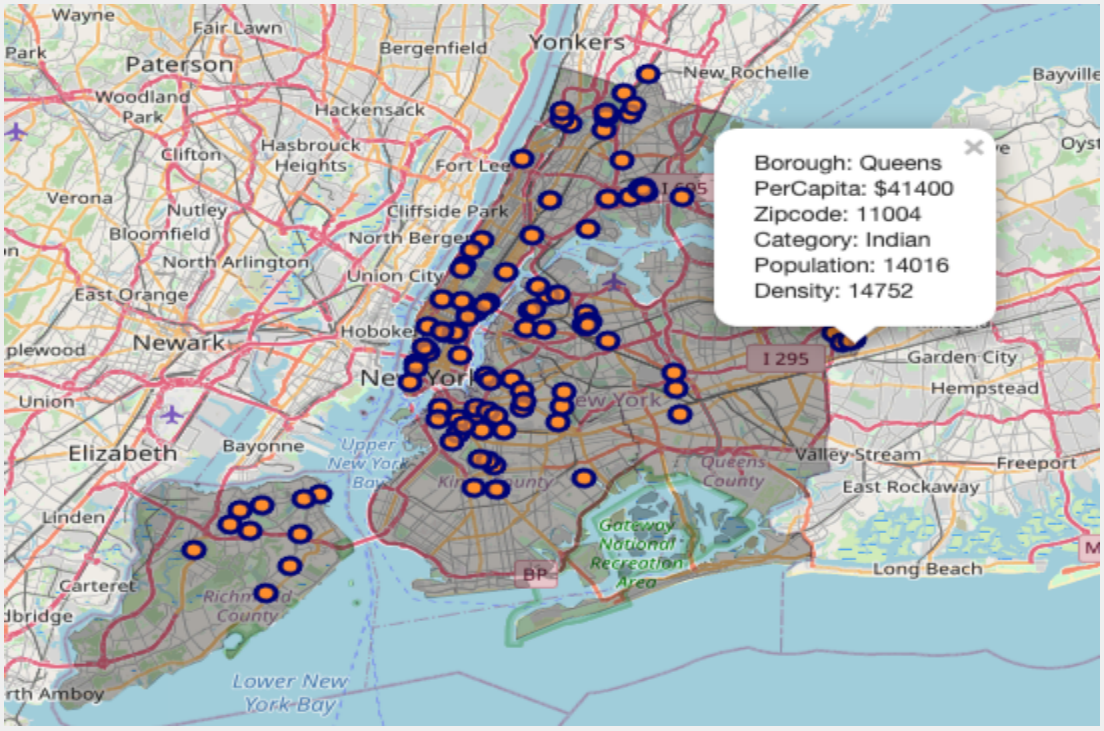
Chinese Category Restaurants in New York Boroughs

In the neighborhood of zip code 10307, the per capita income for that borough is $30,500 and the population in that zip code is 14,096 and the population density is 7184.



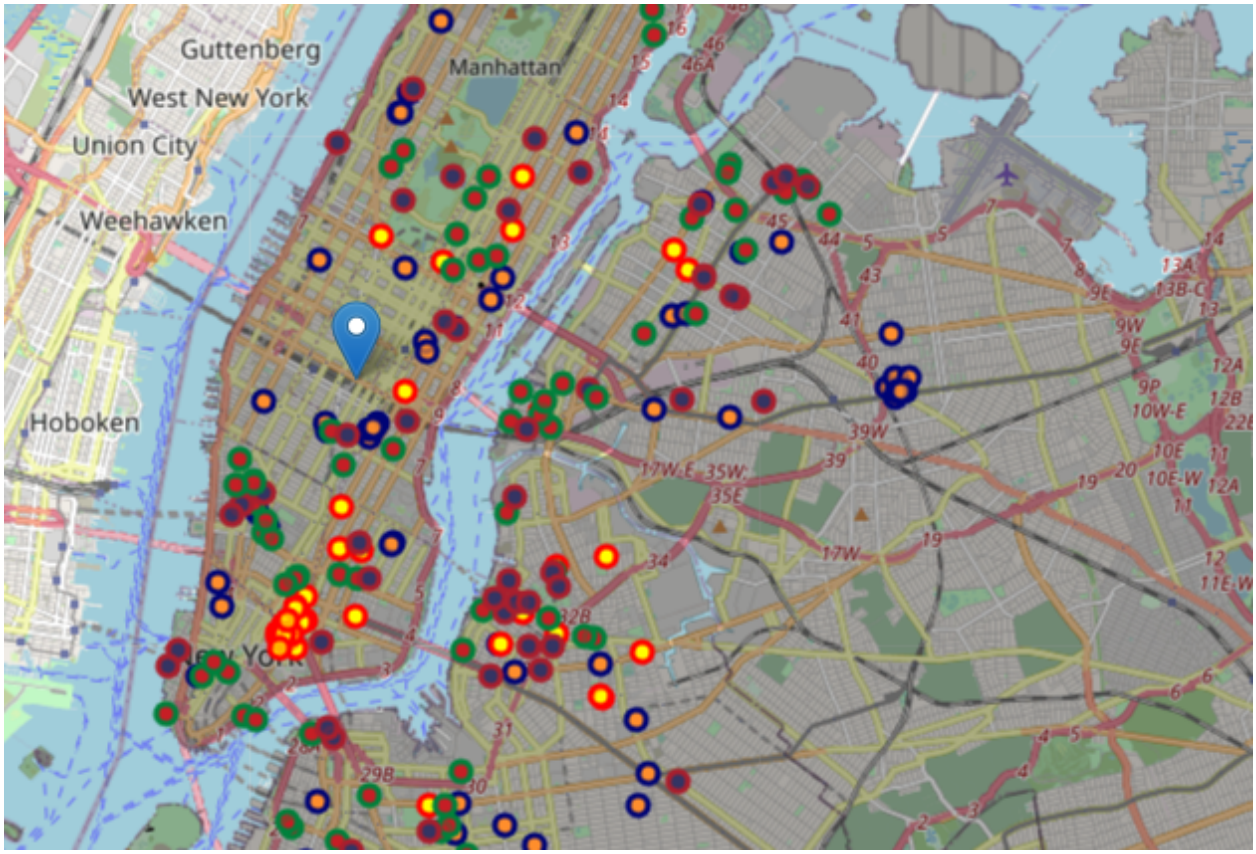
Indian Category Restaurants in New York Boroughs

In the neighborhood of zip code 11004, the per capita income for that borough is $41,400 and the population in that zip code is 14,016 and the population density is 14,752.



Neighborhood Clustering Model

For a given target location, we collect all the restaurant venues for the four categories of restaurant and apply the clustering algorithm to cluster the data into American, Italian, Chinese and Indian restaurants. We then superimposed this cluster data on the New York map. This cluster data provides a clear indication of the existing players within the 5 miles radius from the targeted region.



8. Results

From the restaurant's data, we filtered out the restaurants that are within the 5 miles radius. For a given target location, we collected the 5 miles radius coordinates and then calculated the distance from the target location to all the restaurants within the 5 miles radius. We then used the clustering algorithm to cluster the data and color coded the different restaurant categories.

For a given neighborhood, We also identified the top most common restaurant within the four categories of restaurants and ranked them. We now have a complete picture of each neighborhood and the most common and least common restaurant types within that neighborhood.

ABC Multi Cuisine Inc can utilize this data and decide if it would like to venture into the most common or least common category of restaurants within the targeted region. The population data, per capita income and population density also assist in selecting the right kind of neighborhood that optimize their return on investment.

9. Discussion

We can also consider the land area data. Keep in mind the future growth of some of the low population density with higher land area. In the future, depending on the business needs of **ABC Multi Cuisine Inc,** We can also consider other category of restaurants as the New York city is well diversified

10. Conclusion

This data analysis was done with limited data as we were restricted by the rate limiting imposed by four square, the location data provider. Getting an enterprise license can provide an endless number of possibilities with vast amounts of data.

Based on the current analysis, Manhattan and Brooklyn, the high density areas have the most number of restaurants. These areas have a variety of existing players in the market segment. If the cost is not a constraint, then opening a restaurant in these two boroughs can have a bigger impact as the per capita income in these regions are high compared to other boroughs.

Queens with higher land area could be a good candidate for future expansion. Bronx and Staten Island boroughs can be considered for small size restaurants as the per capita income in these regions are lower than others.