Introduction

New York City is the most populous city in the United States. With a population of 8,804,190 distributed over 300.46 square miles (778.2 km^2), New York City is also the most densely populated major city in the United States. It has been described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports and has sometimes been called the capital of the world.
Situated on one of the world's largest natural harbors, New York City is composed of five boroughs: Brooklyn (Kings County), Queens, Manhattan (New York County), the Bronx, and Staten Island (Richmond County).

The city and its metropolitan area constitute the premier gateway for legal immigration to the United States. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world. New York is home to more than 3.2 million residents born outside the United States, the largest foreign-born population of any city in the world as of 2016.

The city's population in 2010 was 44% White (33.3% non-Hispanic White), 25.5% Black or African American (23% non-Hispanic Black), 0.7% Native American or Alaska Native, and 12.7% Asian. Hispanics or Latinos of any race represented 28.6% of the population, while Asians constituted the fastest-growing segment of the city's population between 2000 and 2010.

Asian Americans in New York City, according to the 2010 census, number more than one million, greater than the combined totals of San Francisco and Los Angeles. New York contains the highest total Asian population of any U.S. city proper.

The Chinese population constitutes the fastest-growing nationality in New York State; multiple satellites of the original Manhattan Chinatown, in Brooklyn, and around Flushing, Queens, are thriving as traditionally urban enclaves.

This has allowed, at least in part, the fusion in some aspects of life, of a multi-ethnic culture, involving the cuisine. The demand for ethnic food has increased exponentially over the years, due to its diffusion, low cost and speed in the availability of food.

Problem Description

The objective of this project is to create a perfect place detection tool which will find the best neighbourhood or place to open a Chinese Restaurant using Foursquare API and location data in New York City.

Using Data science methodology and tools such as data analysis and visualization, this project aims to provide solutions in order to answer to the business question: where, in the city of New York, should the investor open a Chinese Restaurant?

Target Audience

This project is particolary useful to developers and investors looking to open or invest in Chinese restaurant in the city of New York. Overall, New York is a great place to open a restaurant with an ethnical cuisine. Thanks to the integration of different cultures, New York has become the multi-ethnic city par excellence, with more than one million Chinese residents. Chinese food has effectively become part of New York cuisine and its demand is very high.

Data

In order to answer to the previous questions, we will get the services of Foursquare API utilized via the Python Request library to explore the data of New York City, in terms of its neighborhood, including boundaries, latitude and longitude. The data also include the information about the places around each neighborhood like restaurants, hotels, coffee shops, parks, theaters, art galleries, museums and many more, their raitings and tips.

New York City data cointaining the neighboorhoods and boroughs, latitude and longitudes will be obtain from the data source:

https://cocl.us/new_york_dataset

We will use machine learning technique, "clustering" to segment the neighborhoods with similar objects on the basis of each neighborhood data. These objects will be given priority on the basis of foot traffic (activity) in their respective neighborhoods. This will help to locate the tourist's areas and hubs.

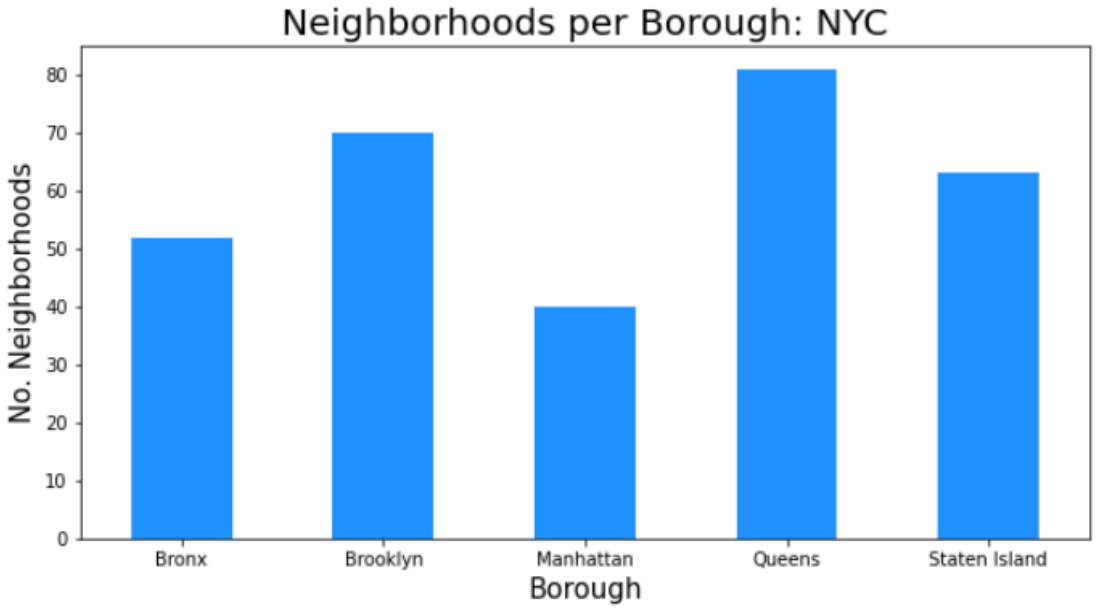
Methodology

After importing the Python libraries necessary for our analysis, we can start by getting the data from the public database https://cocl.us/new_york_dataset, perform cleaning (eliminating "Not

assigned" or missing values, clean not labeled values) and processing into a dataframe. We will save them in .csv file, this is because the API calls we can make are limited.

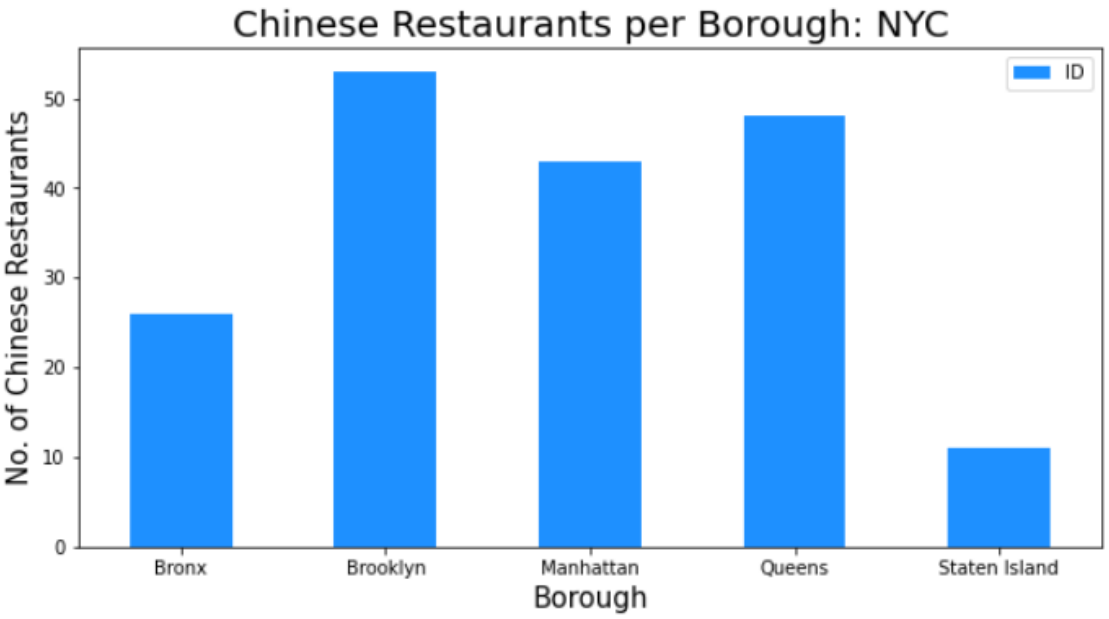
We are finally ready for data analysis.

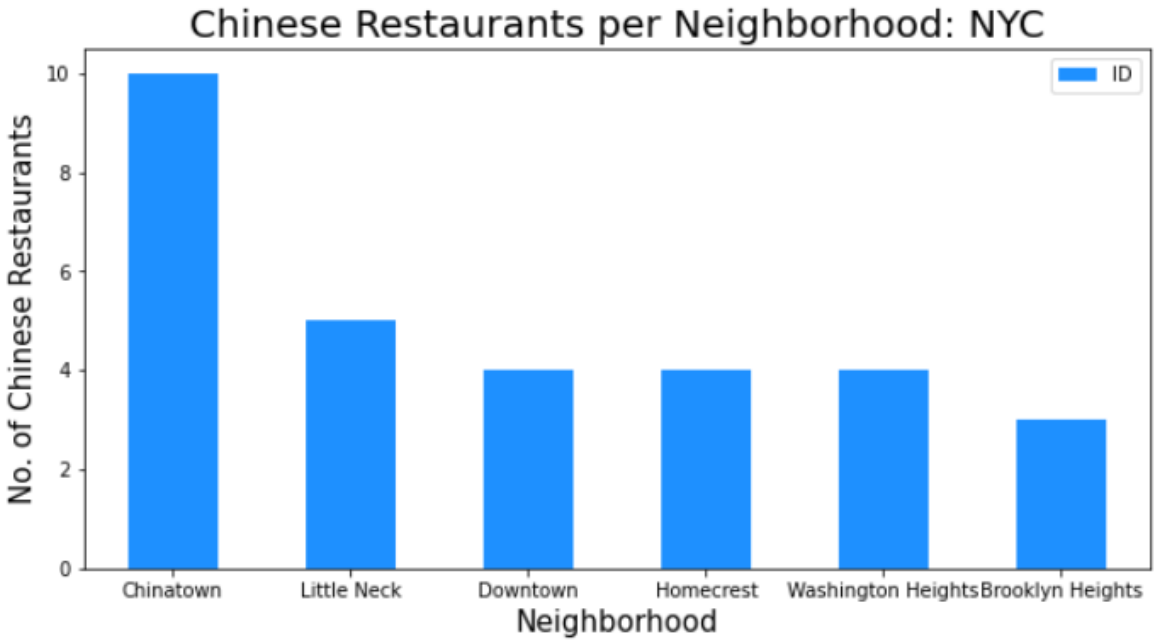
As a preliminary analysis, we ran .head and .shape to explore our dataframe and find that the total number of boroughs in New York is 306.



Now, we explore our data through data visualization.
The borough that has more neighboorhoods is Queens.

As a result of further analysis, there are 183 Chinese restaurants in New York City.

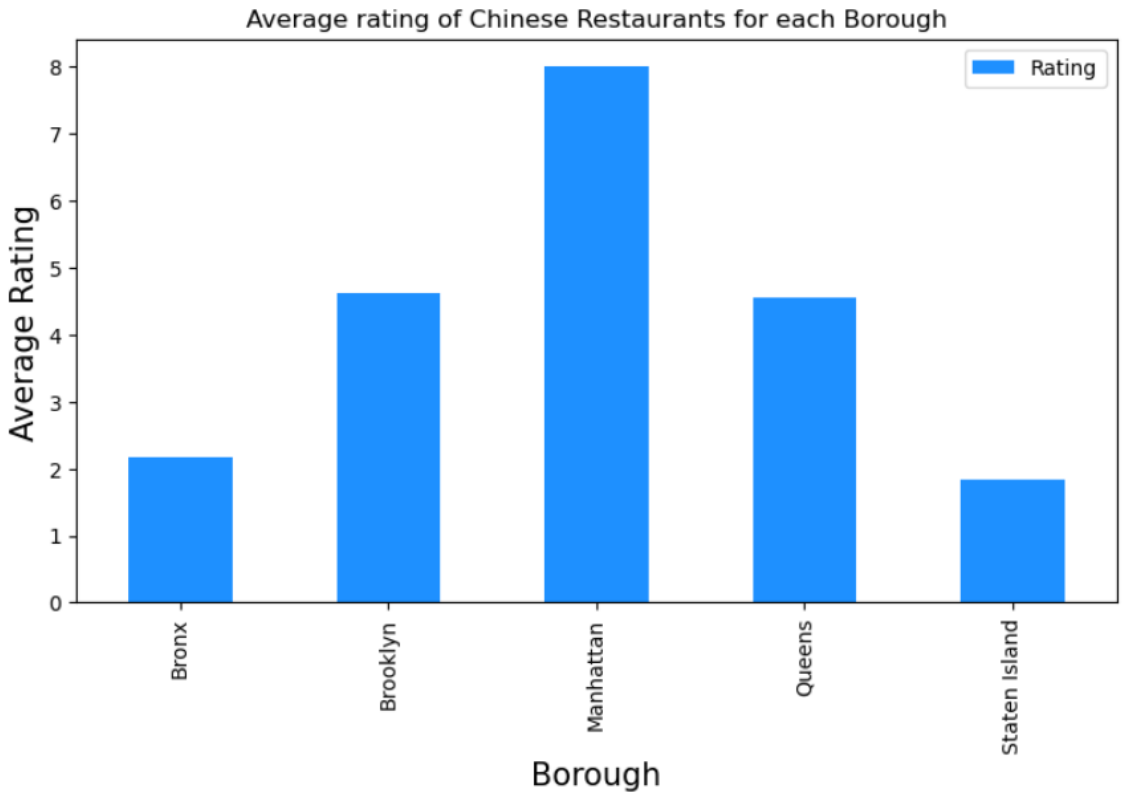




Continuing our analysis, we see below that although Manhattan had the least number of neighborhoods, the largest number of Chinese restaurants is located in Chinatown.

	Borough	Neighborhood	ID	Name
62	Manhattan	Chinatown	4db3374590a0843f295fb69b	Spicy Village
63	Manhattan	Chinatown	4a96bf8ff964a520ce2620e3	Wah Fung Number 1 Fast Food 華豐快餐店
64	Manhattan	Chinatown	5894c9a15e56b417cf79e553	Xi'an Famous Foods
65	Manhattan	Chinatown	3fd66200f964a520b1ea1ee3	Great N.Y. Noodletown
66	Manhattan	Chinatown	59d828f0916bc1155fde2c04	Hwa Yuan
67	Manhattan	Chinatown	5c965dad5455b2002c058659	Yi Ji Shi Mo Noodle Corp
68	Manhattan	Chinatown	4afff2b4f964a520583a22e3	Sun Hing Lung Ho Fun Tofu
69	Manhattan	Chinatown	4bc34a294cdfc9b6a01a9721	Happy Cafe (Happy Express Cafe)
70	Manhattan	Chinatown	3fd66200f964a520ceea1ee3	Deluxe Green Bo Restaurant
71	Manhattan	Chinatown	3fd66200f964a520ede41ee3	Big Wong King 大旺

At this point, we analyze the scores of individual restaurants with their likes, ratings and tips in order to create a statistic of average rating of Chinese restaurants for each borough and view the results obtained.



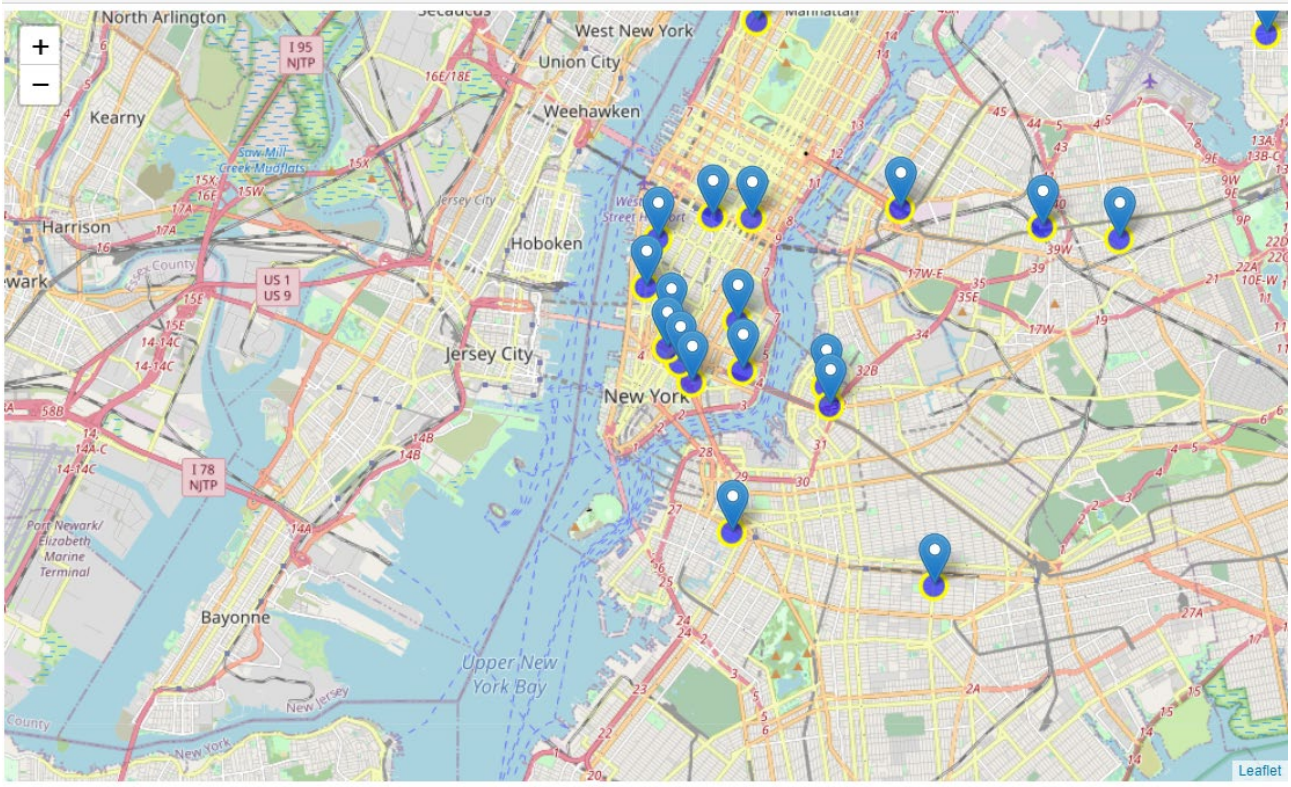
Results

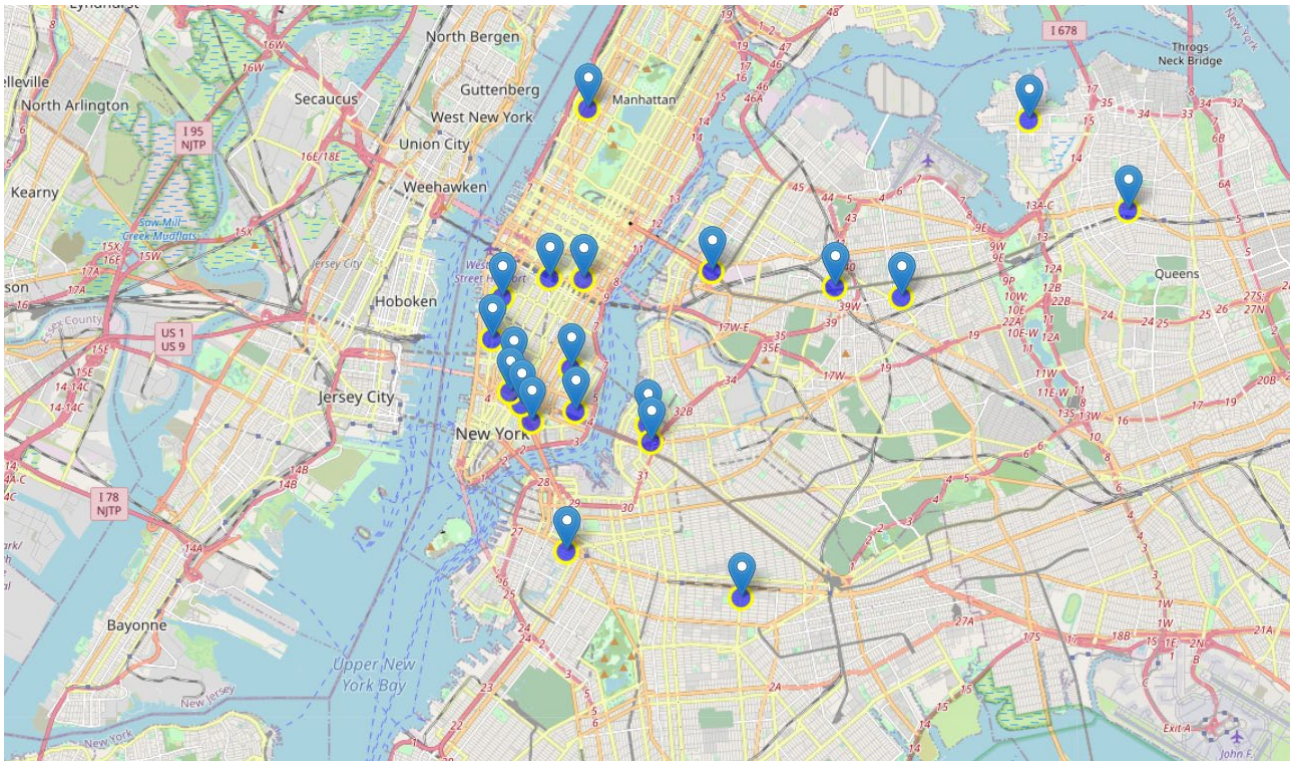
As a final result, we can say that the borough that contains more neighboorhood in New York City is Queens, followed by Brooklyin and Staten Island. Based on the database in our possession, we can say that there are 183 Chinese restaurants throughout the city, of which:

- most of the Chinese restaurants divided by borough are located in Brooklyn, followed by Queens and Manhattan.
- most of the restaurants divided by neighborhood are located in Chinatown, where the number is really large. The other neighborhoods follow roughly evenly.

On the other hand, analyzing the data concerning the rankings concerning a restaurant (likes, ratings, tips) divided by borough, the most popular is Manhattan. Queens and Brooklyn follow a tie much later.

Let's now consider all the neighborhoods with average rating greater or equal 8.0 to visualize on a map and merge them with the dataset to original New York data to get longitude and latitude. We finally display this data on a map via Python's Folium library and add a new field to the dataframe for labeling purposes.





From the conclusive analysis of these data, we can therefore say that for an investor who wants to open a Chinese restaurant in New York, the two ideal areas are Manhattan and Queens. Manhattan is the most touristic area in the entire city, which includes Times Square, the Empire State Building, Central Park and is close to the Statue of Liberty.

At the same time, Manhattan is the area where Chinese restaurants, especially in Chinatown, are more numerous and have the highest reviews in all of New York, so I suggest opting for

Manhattan, excluding the Chinatown neighborhood. The second option could be Queens, which remains in a great position for less competition and lower rental and costs than Manhattan.

The area that I strongly advise against is Brooklyn where the competition is really high. The number of Chinese restaurants divided by borough is the highest, and to face the competition it would be necessary to focus on an exclusive and top quality restaurant.

As a final note, the entire analysis and design performed is based on the accuracy of the Four Square data. More thorough analysis and future work should incorporate data from other external databases.