

Capstone Project - The Battle of the Neighborhoods

Lei Guo
April 29, 2020

1. Introduction

1.1 Background

New York City is experiencing a Chinese food renaissance. Never before have the city's offerings been so diverse; not only are multiple regions represented, but price points range, too. There are fast-casual spots like Xi'an Famous Foods, and there's ambitious pan-regional fine dining like Atlas Kitchen. And where restaurateurs once needed to cater to Western tastes, many of today's New York Chinese restaurants don't have to so in order to survive.

1.2 Problem

In this project we will try to find an optimal location for a restaurant. Since there are lots of restaurants in New York we will try to detect the best location that are not already crowded with Chinese restaurants. Moreover, we are also particularly interested in areas with less competitor Chinese restaurants with good market reputation.

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

1.3 Interest

Specifically, this report will be targeted to stakeholders interested in opening a Chinese restaurant in New York city, USA.

2. Data acquisition and cleaning

2.1 Data sources

Based on definition of our problem, factors that will influence our decision are:

- number of existing Chinese restaurants in the neighborhood.
- number of existing Chinese restaurants in the neighborhood with good market reputation (likes, tips etc.).

Following data sources will be needed to extract the required information:

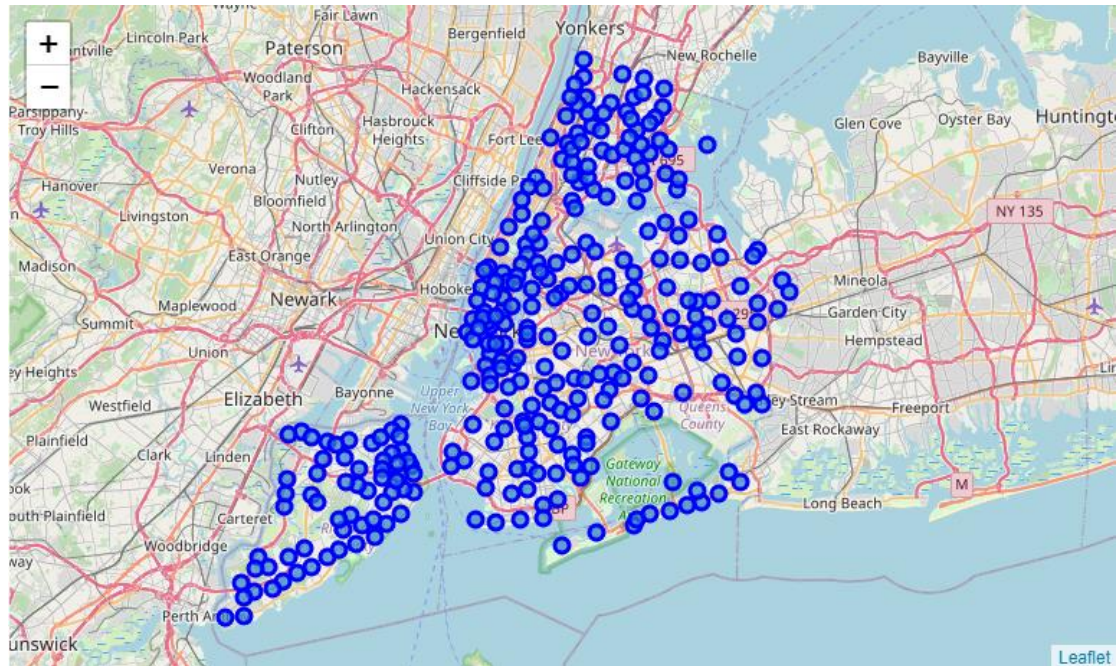
- New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.
Data source: https://cocl.us/new_york_dataset
- Chinese restaurants in each neighborhood of New York city.

Data source : Fousquare API

2.2 Data cleaning

New York has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

Create a map of New York with neighborhoods superimposed on top.



Next, we are going to start utilizing the Foursquare API to explore the neighborhoods and segment them. The details are in the below part.