Analysis NYC Neighborhoods for New Chinese Restaurant Business

Yizheng Li   03/22/2020

1. Introduction

1.1 Background

United Sates of America is an immigration country, which create ethnic diversity. New York as the economic center of U.S., because of this unsurpassed ethnic diversity, the restaurant goer will feel that he has died and gone to heaven while in New York City. The infinite numbers of Chinese dishes, flavors, and methods of cooking which make eating Chinese food joyful and exciting. So, when someone want to open a Chinese restaurant business in NYC, which location and neighborhood is the best for business? Location is very important for business, a bustling block can help restaurant attract more people, meanwhile, the restaurant trend are also very important for a new restaurant.

1.2 Problem

For a restaurant owner, the tradition way is field investigation by ask information from people and check out each location personally.   However, this is very inefficient way, people waste money and time during this process. In addition, their information may not be accurate or complete, this will cause huge economic losses when the new restaurant performance does not meet owner's expectation. The worst thing is the owner doesn't even realize the problem of location and waste more money on propaganda and other expense. Using data science to analysis and cluster Chinese food to find the best location is best way for owner, this will save both money and time. We will use data to make the decision on the place that is the most convenient for owner's new business.

2.Data acquisition and cleaning
2.1 Data source

For data and location analysis, we need geographic data, borough and neighborhoods Data, Chinese restaurant data from data sources.

- NYC borough and neighborhoods from use Wikipedia information and NYC data Json file from previous Lab: https://cocl.us/new_york_dataset
- NYC geophysics data from Foursquare location data
- Chinese Restaurants data in NYC from NYC Open Data
- NYC population from NYC Department of Health

2.2 Data Cleaning

The initial Chinese restaurant data frame contained redundant information such as street, zip codes, phone, etc. So, a new data frame was created with just the columns containing boroughs and Chinese Restaurant Name.

The borough and neighborhoods data frame were cleaned and leave Borough, Neighborhoods, coordinate to give us better observation.