

The Battle of Neighborhoods Project

1. Introduction

Background:

The Project is the fulfillment of the Coursera IBM Data Science certification course. The project requirements are to leverage the “Foursquare location data to explore or compare neighborhoods or cities of your choice or to come up with a problem that you can use the Foursquare location data and other related data to solve.”

My initial intention is to design this project as a consulting investigation for the catering industry investors by data exploration technologies. I select New York City as the city which we would explore and investigate. Since New York City is one of the most popular travel destinations in the world, and also it has very large kinds of different cuisines in the city.

Problem Definition:

Suppose a Chinese chain restaurant (SWJ) wish to expand its business to North America. SWJ is a very popular brand in China and it has more than 1000 brand restaurants in China. We select Queens, New York City as the first landing location for SWJ in North America. According to its brand awareness in China, so SWJ's target customers are the tourists visiting New York City in China. And the location of new-open restaurant would be near tourism attractions, shopping stores and Chinese-like cuisine restaurants. Because for tourists from China, visiting attraction venues, shopping and dining are the most parts of the whole trip.

So the mission can be the statement:

To find a good location to open a Chinese restaurant in Queens, New York City, which the neighborhood has the most of attractions, shops and Chinese-like cuisine restaurants.

2. Data Sources

Data Source 1 - Neighborhood Data

Queens, which is the target boroughs of this investigation within New York City, has over 100 neighborhoods. We first need to obtain a list of all the locations of the neighborhoods in Queens. This information is available on the following web address: https://geo.nyu.edu/catalog/nyu_2451_34572

Data Source 2 - Geographical Coordinates

Geographical coordinates for each neighborhood will be obtained with the aid of GEOPY Library. Each neighborhood will be assigned a latitude and longitude coordinate.

Data Source 3 - Venue categories

We will use the Foursquare API to retrieve venues, using the coordinates obtained in Data Source 2 above. We shall further obtain a list using Foursquare API for related venues such as attractions, shops and restaurants in Queens.