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

The data is from a JSON File, the structure is as the following example including neighborhood name, borough name and other information.

{'geometry': {'coordinates': [-73.84720052054902, 40.89470517661],
  'type': 'Point'},
 'geometry_name': 'geom',
 'id': 'nyu_2451_34572.1',
 'properties': {'annoangle': 0.0,
  'annoline1': 'Wakefield',
  'annoline2': None,
  'annoline3': None,
  'bbox': [-73.84720052054902,
   40.89470517661,
   -73.84720052054902,
   40.89470517661],
  'borough': 'Bronx',
  'name': 'Wakefield',
  'stacked': 1},
 'type': 'Feature'}

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.

Combined with neighborhood data and geographical data, the data source should be structured like the following example:

	Borough	Neighborhood	Latitude	Longitude
0	Queens	Astoria	40.768509	-73.915654
1	Queens	Woodside	40.746349	-73.901842
2	Queens	Jackson Heights	40.751981	-73.882821
3	Queens	Elmhurst	40.744049	-73.881656
4	Queens	Howard Beach	40.654225	-73.838138

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.

The example Foursquare API which I will use to collect this venue data is as the following:

```
url =
'https://api.foursquare.com/v2/venues/explore?&client_id={}&cli
ent_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(

    CLIENT_ID, CLIENT_SECRET,VERSION,

    neighborhood_latitude,

    neighborhood_longitude,

    radius,

    LIMIT)
```

After retrieving data from Foursquare and transferred into dataframe, the data source should be structured as the following example:

	name	categories	lat	lng
0	Favela Grill	Brazilian Restaurant	40.767348	-73.917897
1	Orange Blossom	Gourmet Shop	40.769856	-73.917012
2	Titan Foods Inc.	Gourmet Shop	40.769198	-73.919253
3	CrossFit Queens	Gym	40.769404	-73.918977
4	Simply Fit Astoria	Gym	40.769114	-73.912403