The Hunt For Chinese Restaurants In New York City

Brandon Tan

July 30, 2021

1. Introduction

1.1. Business Understanding

New York City, home to 8.4 million people is one of the most famous and affluent cities in the United States of America. The population density is 27,000 people per square mile, with an area of 468.19 square miles. With such density and wealth, it creates tremendous business opportunity for a kind of sectors.

The food industry is one that is particularly competitive, utilizing my skills and knowledge for interpreting data my objective is to find all the locations of Chinese Restaurants in NYC to study the competition.

1.2 Project Objectives

Population density is a key aspect of analyzing data on cities, and it plays an important part in issues such as transport, infrastructure and living standards for residents. Some people prefer a slow pace and less crowded cities with open spaces, while others would rather have the fast-paced city life. In New York, not just Pizzerias or Coffee Shops are famous, but also there is fine Michelin starred restaurants with the most diverse type of food. With this diversity of restaurants with different categories like Chinese, Indian, French, Brazilian, etc. So here, the main goal is to find answers to the following questions:

List and visualize all major parts of New York City that has great restaurants.
Where is the best location in NYC for Chinese restaurants?
Which areas poses a great potential for a new Chinese restaurant business?
Which all areas lack Chinese Restaurants?
Which is the best place to stay if you prefer Chinese Food?

1.3 Interest

A fictitious restaurant chain in China, who wants to start to expand their operation of a new restaurant in New York City, requested this project. The objective is to locate and recommended to the stakeholders which neighborhood of New York City will be best choice to start a restaurant. However is fundamental the stakeholders understand the rationale of the recommendations made.

1.4 Methodology

Every project request, need a good methodology to solve and understand better the business problem. This project will be approach by CRISP-DM (Cross Industry Standard Process for Data Mining) that I believe to be the best lifecycle to develop this project.

The lifecycle of this project will include almost every step defined in the data mining projects that include:

- Business Understanding
- Data Understanding
- Data Preparation & Cleaning
- Modeling
- Evaluation
- Deployment
- Conclusions

2. Data Understanding

2.1 The Data Set

New York City's have a collection of restaurants from across the city that defined by a diverse culture with a lot of diverse type of food, each belonging to different categories like Chinese, Indian, French, Brazilian, etc. For this project, we will use an open data as following below: New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.

- Data set: https://cocl.us/new_york_dataset
- Description: This dataset contains information such as Borough, Neighborhood, Lat & Long
-

Venues of different restaurant cuisines

- Data set: Foursquare API
- Description: Foursquare API allows us to gain access to venues of different restaurants, specifically Chinese restaurants.
-

GeoSpace

- Data set: https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm Description: By using this geo, space data we will get the New York Borough boundaries that will help us visualize choropleth map.

Borough	Neighborhood	Latitude	Longitude
Bronx	Wakefield	40.894705	-73.847201
Bronx	Co-op City	40.874294	-73.829939
Bronx	Eastchester	40.887556	-73.827806
Bronx	Fieldston	40.895437	-73.905643
Bronx	Riverdale	40.890834	-73.912585

3. Data Preparation & Cleaning

I had use the Foursquare API to get the top 100 venues with in a radius of 1000 meters for a given latitude and longitude. And with data um use an Explanatory Data Analysis to understand the data.

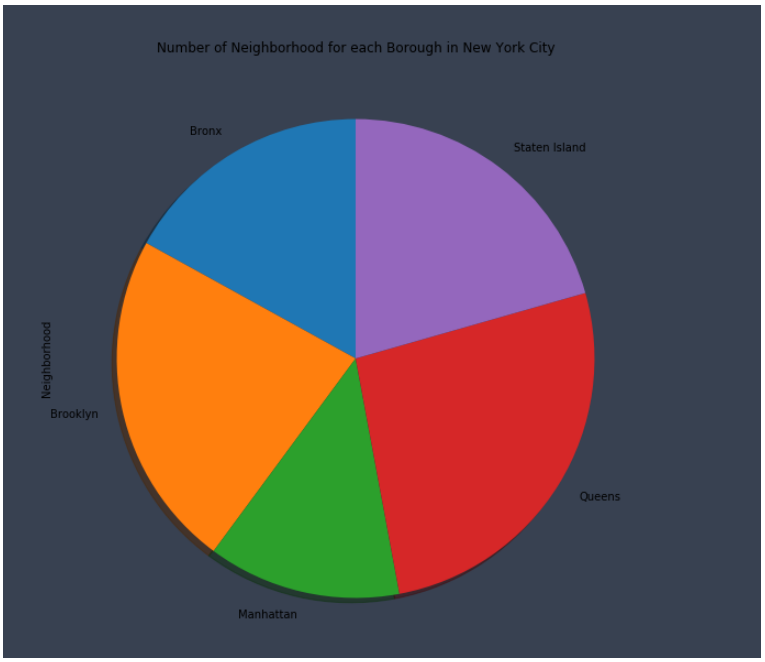

Figure 1: Number of Neighborhood for each Borough in New York City

4. Modeling

This project's main function is data exploration, hence no modeling is required.

5. Evaluation

Based on our findings we now can identify the areas where competiton for chinese restaurants are high, and potentially use this information to our advantage if we want to start our own fried rice shop