The Battle of Neighborhoods: Coursera Capstone Project

Opening a new Japanese Restaurant in Manhattan and Queens, New York City.

Tongxi Zhao

April 30th, 2021

Introduction

New York city is the most populous city in the United States, the core of the New York metropolitan area, and one of the largest cities in the world. It is an international metropolis with great influence on the global economy, commerce, finance, media, politics, education, and entertainment. New York City consists of five boroughs: Manhattan, Queens, Brooklyn, Bronx, and Staten Island, each of which is a county of the state of New York.

New York is also the most densely populated major city in the United States. In 2019, there were 8.419 million people living in the New York city according to United States Census Bureau, Eurostat[1].

New York city is the headquarters of the United Nations, so it is also considered the center of world diplomacy. New York city is also known as the "cultural capital of the United States". People describe the New York City as being diversity, including diverse people coming from the everywhere across the world, diverse languages, diverse food culture etc. According to 2010 Census[2], the New York City has become the home to more than one million Asian Americans, it contains the highest total Asian population.

In addition, according to the 2000 census, over half of the 37,279 people of Japanese ancestry in the U.S. state of New York lived in New York City. As of 2012, the New York City metropolitan area was home to the largest Japanese community on the East Coast of the United States[3]. After Italian, Chinese and Mexican, Japanese food is probably the most popular ethnic cuisine in the United States. Based on these real facts, to open a new Japanese restaurant is becoming the planned business plan in New York City, and this final project explores the possible neighborhoods to start new Japanese restaurants throughout the Queens and Manhattan, because these two boroughs are having taste food from around the globe, people are more likely to explore delicacy here.

Business Problem

I will be leveraging the Foursquare location data to find numbers of popular or people favored Japanese restaurants for each neighborhood in Queens and Manhattan, I will also look at the Japanese restaurants' frequency in venue category for each neighborhood in

[1] https://datacommons.org/place/geoId/3651000?utm_medium=explore&mprop=count&popt=Person&hl=en
[2] http://www.ameredia.com/resources/demographics/asian_american.html
[3] https://books.google.com/books?id=O_9B1qXjohsC&pg=PA105#v=onepage&q&f=false

Queens and Manhattan. The project assumes that the neighborhood with larger number of favored Japanese restaurants are not optimal choice for opening a new Japanese restaurant become the competition in these neighborhoods could be fierce. The assumption is very naïve without considering other dimensions, such as demographics, geographic preferences, traffic routes, and so on due to the limitation of data availability and time. To improve the results, I will be implementing K-means clusters model to have a cluster on types of venue category with string 'Restaurant', the cluster results will indicate the similar neighborhoods and ranking of types of food in these neighborhoods. By combing the numbers of favored Japanese Restaurants and clustered restaurants, I will suggest the possible neighborhoods to open a new Japanese restaurant.

Data:

1. Data Downloaded from:

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json

It's a New York city data with city features, the features in *json* file as Python dictionary key provides a list of neighborhoods. To extract other desired dictionary keys, I can create a *Pandas Dataframe* with columns including boroughs in New York Cities, neighborhoods in each boroughs and geographic coordinates for each neighborhood.

2. Given the geographic coordinates of each neighborhoods, the project explores the popular venues in each neighborhood using Foursquares API. The venues information dataset will be leveraged to analyze the business problem.