Opening a New Restaurant in NYC

IBM Applied Data Science Capstone

By: Surya Gurung

Feb 2020

Background

New York City (NYC) is the most populous city in the United States (US) with over 8 million population spread over five boroughs, Brooklyn, Queens, Manhattan, Bronx, and Staten Island. It is well known as global capital of finance, media, and immigrant. NYC is also a global leader in entertainment, fashion, tourism, technology, education, arts, sports, politics, research, and many more industries. Many of the World's most visited tourist attractions are located in NYC. Times Square and Broadway Theater District are few of them. More than 62 million tourists visited in 2017. You can find communities representing whole world, in fact more than 800 languages are spoken in NYC.

Time Square, NYC

NYC is also the home of the World's largest stock exchanges. Many of the World's largest financial and media companies are based in NYC. Recently, new technology and biotechnology companies are growing very fast. In brief, NYC is a power house for every business sector you can imagine. Obviously, restaurant industry has huge opportunity in NYC. After all, the 8+ million New Yorkers and 62+ million tourists need to eat every single day with demands for every possible cuisine on the earth. There is no other better place to open a restaurant than NYC neighborhoods.

Business Problem

Being the most populous city in US and one of most visited cities in the world, is it easy to open a new restaurant successfully in NYC? Even though market for restaurant business is huge in NYC, there are already plenty of restaurants in popular neighborhoods serving every kind of cuisine from every corners of the world. The process of finding the best place for the new business is not as trivial as it may seem. It requires careful and thorough research and

decision to find best possible real estate to give the new eatery best chance to succeed. It requires time, domain expertise, and capital.

So, the problem every new entrepreneur may face is how to find the best location to open a new restaurant in NYC with ease, precision, and quickly so that the new business has best chance to succeed?

There is a better solution for the new restaurant entrepreneurs and current restaurant owners who want to expand their business. The solution is machine learning models such as clustering. The main objective of this capstone is to develop a k-means clustering model and visualize the clusters in the NYC map by using NYC neighborhood datasets along with Foursquare API to help the prospective restaurant owners to select best possible NYC neighborhood for their new business quickly with greater precision. So, this project will greatly simplify the tedious process of opening new restaurant saving the precious time for the restaurant owner.

Data Requirements & Sources

To build the machine learning models, first we need the datasets representative of the business problem. For this project, we are going to use following datasets:

- New York City neighborhood dataset which contain list of neighborhood names and their geographic coordinates in five boroughs, Brooklyn, Bronx, Manhattan, Queens, and Staten Island. This dataset defines the scope of this project which is the neighborhoods in five boroughs of NYC. This dataset also provides the geographic coordinates, latitude and longitude in order to plot the locations of the neighborhoods on the NYC map.
- Census Demographics at NYC Neighborhood Tabulation Area (NTA) dataset. This dataset contains 2010 census population of NYC neighborhoods and it will be used to plot the population density of the neighborhoods on the NYC map. This is obviously very helpful to make decision to start new restaurant. After all, restaurants are for feeding the hungry customers and larger the population more the hungry customers the restaurant is going to get.
- NYC crime dataset based on NYPD complaint dataset. This dataset contains all valid felony, misdemeanor, and violation crimes reported the New York City Police Department in 2019 and the geographic location of the crimes. So, we will plot the crime locations on the NYC map. This will add an additional factor to determine whether the neighborhood is good for the new restaurant.

- Venue dataset returned by the Foursquare API based on the NYC neighborhoods dataset. This data will be used to perform clustering of the NYC neighborhoods based on the most popular venues and the restaurants in the neighborhoods. By clustering the neighborhoods based on the most popular venues and restaurants, it will help make decision if the neighborhood is good for food related venues or if the neighborhood is already too saturated for food related venues.

Sources:

- https://cocl.us/new_york_dataset contains 306 New York City neighborhood names with their geographic coordinates spread over five boroughs. The data will be retrieved using the python package requests as json file and will be converted into pandas dataframe with the borough and neighborhood names along with their latitude and longitude as the columns.
- https://data.ny.gov/resource/rnsn-acs2.json contains 2010 census population of NYC neighborhoods in the five boroughs. This data will also be retrieved using requests package and convert into pandas dataframe.
- https://data.cityofnewyork.us/api/views/5uac-w243/rows.json?accessType=DOWNLOAD contains the crimes reported the New York City Police Department in 2019. This data will also be retrieved and converted similarly as other datasets.
- https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={} is Foursquare API and it will be called for every NYC neighborhoods.