Opening a High-end Restaurant in New York

By: Juan Burgos
May 17, 2020

Introduction

There are countless options for entrepreneurs, developers and investors as to where they could potentially invest their money, time and resources. One of these possible investment opportunities are businesses such as restaurants, bars, night clubs, etc. The risk of opening such a business can be mitigated by choosing an appropriate location and type. Then as investors look to protect their money from potentially risky enterprises, being able to leverage existing business and population data to determine what kind of venues and in which locations to invest would be of great value.

In this report, we aim to build a model to determine what kind of venue and in which location in New York City would a prospective entrepreneur chose for their new business. We will restrict our report to restaurants, as restaurants and eateries are one of the most common venues and one of the riskiest. In particular, our focus will be on high-end or expensive restaurants in New York City.

Business Problem

The objective of this project is to inform what kind of restaurant and where should it be located for a person or group looking to open a new high-end restaurant. Using Data Science methodologies and Foursquare location data this project aims to answer the question; what type of eatery should somebody chose for their new high-end restaurant and where in New York City would it be located?

Data

To address this problem we will need the following data:

- List of boroughs in New York City with their respective coordinates. This will define the scope of this project to New York City.
- Coordinates of each of the neighborhoods in each borough in New York City. This will be used to search venues using the Foursquare API and to plot the map.
- Population and economic data for the New York City Boroughs. This will be used to narrow down the location of our high-end restaurant.
- Venue data from the Foursquare API for New York City. Foursquare classifies each venue with a price qualifier containing the Price Tier from 1 (least pricey) - 4 (most pricey). For food venues, in the United States, the Foursquare Price Tiers are defined as:
    - 1 is < $10 an entree
    - 2 is $10-$20 an entree
    - 3 is $20-$30 an entree
    - 4 is > $30 an entree

  In particular, we will restrict our search to venues in the most pricey tier, as we define these as high-end venues. Additionally, we will use the Venue Category to further filter our data to eateries and restaurants, thus excluding bars, clubs and any other type of venue returned by the API.

Data Sources

Demographic information for New York City consisting of population and economic data for each of the boroughs is scraped from the following source on Wikipedia: https://en.wikipedia.org/wiki/Demographics_of_New_York_City

Then we will get the geographical coordinates of the New York City neighborhoods from a JSON file obtained using the Python Geocoder package. After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. The Foursquare API will provide many categories for each of the queried venues. We are particularly interested in the Restaurant Category (eg. French Restaurant), filtered by the priciest Price Tier in order to help us to solve the business problem at hand.