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.

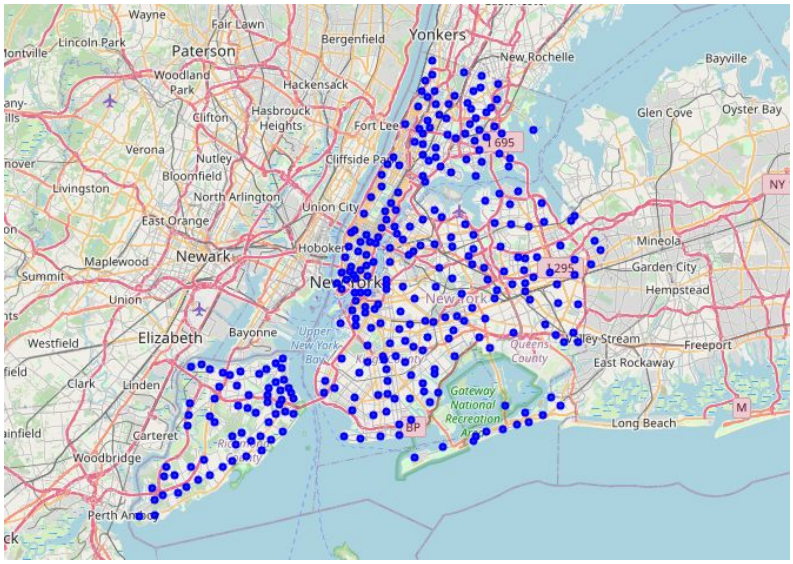
Business Problem

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 make use of 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 of New York City. This will be used to search venues using the Foursquare API and to plot the map. Below is a map of New York City with all the neighborhoods superimposed:



- Population and economic data for the New York City Boroughs. This will be used to narrow down the location of our high-end restaurant within New York City

New York City's five boroughs								
Jurisdiction		Population	Gross Domestic Product		Land area		Density	
Borough	*County*	*Estimate (2019)*[3]	*billions (US$)*[4]	*per capita (US$)*	*square miles*	*square km*	*persons / sq. mi*	*persons / km²*
The Bronx	Bronx	1,418,207	42.695	30,100	42.10	109.04	33,867	13,006
Brooklyn	Kings	2,559,903	91.559	35,800	70.82	183.42	36,147	13,957
Manhattan	New York	1,628,706	600.244	368,500	22.83	59.13	71,341	27,544
Queens	Queens	2,253,858	93.310	41,400	108.53	281.09	20,767	8,018
Staten Island	Richmond	476,143	14.514	30,500	58.37	151.18	8,157	3,150
City of New York		**8,336,817**	**842.343**	**101,000**	**302.64**	**783.83**	**27,547**	**10,636**
State of New York		19,453,561	1,731.910	89,000	47,214	122,284	412	159
Sources:[5] *and see individual borough articles*								

- Venue data from the Foursquare API for New York City. Foursquare classifies each venue with a price qualifier in the form of a Price Tier from 1 (least pricey) to 4 (most pricey). For food venues, in the United States, the Foursquare Price Tiers are defined as follows:
    - 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 stores, bars, nightclubs and any other type of venue returned by the API. This is an example of the output from the Foursquare API with venue names, locations and categories.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Category
2	Chinatown	40.715618	-73.994279	Contra	50234a5ae4b0e5018836a116	40.719889	-73.989250	New American Restaurant
3	Chinatown	40.715618	-73.994279	Balvanera	53e02d67498e78c091589343	40.720547	-73.985399	Argentinian Restaurant
4	Chinatown	40.715618	-73.994279	Atera	4f627061e4b05c1d57815977	40.716752	-74.005712	Molecular Gastronomy Restaurant
7	Hamilton Heights	40.823604	-73.949688	Ponty Bistro Harlem	53e01975498e78c0915599bf	40.817886	-73.941522	African Restaurant
8	Central Harlem	40.815976	-73.943211	Ponty Bistro Harlem	53e01975498e78c0915599bf	40.817886	-73.941522	African Restaurant

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.