New York City Neighborhood Venue Analysis

Manthan Shah

August 3rd, 2019

1. Introduction

1.1 Background

New York city was called many names, Big Apple, Financial Capital of the World, the Garment Capitol, Wall Street, and many more. Jay Z even wrote a song about it where he refers to the city as "a concrete jungle where dreams are made of". New York was called many things, but never has it been called slow, boring, or uneventful. One can find interesting places to visit, eat, drink, enjoy, in one of the five boroughs of the city, namely Bronx, Brooklyn, Manhattan, Queens, and Staten Island. The city, knowing to be one of the top tourist destinations of the world, attracts a large number of travelers everyday who throng the city streets morning to night. This provides a good opportunity for a business-person to open a business in areas that are frequently visited by large groups of people, tourwasts. This helps the business-person personally and the nation economically.

1.2 Problem and Target Audience

Due to the vastness of the area and the ample locations available to visit, it can sometimes become quite intimidating for a new timer to know one's way around the city. Many a times it can become quite confusing given with so many options to visit. It would be quite helpful to know which and where are interesting places to visit. Having knowledge about different locations are nearby might help one plan their day around the city.

Having knowledge about different types of locations around a neighborhood can help a business-person make an informed decision about the location where a new shop should be set up to incur profit.

This paper attempts to give a general idea of different types of venues situated across 300 neighborhoods spanning 5 boroughs of the city. The target audience was classified under two cohorts of population: firstly, people with little to no knowledge about the locations of different venues around the 5 boroughs of the city, and secondly, prospective business-people looking for locations to set up new businesses.

2. Data acquisition and cleaning

2.1 Data Sources
Firstly, to know the neighborhoods of data, spatial information (latitude and longitude coordinates) was needed. A JSON (JavaScript Object Notation) was obtained containing spatial coordinates of 306 neighborhoods of New York City.

Secondly, and most important, source of data was FOURSQUARE. Its API (Application Programming Interface) was used extensively to obtain venue data using latitude and longitude coordinates obtained from the JSON file.

2.2 Data Collection Methodology
Firstly, the JSON file was downloaded and parsed. From it, 306 neighborhood's name, spatial coordinates, and the borough in which its situated was collected. Using this spatial data, and python's Folium mapping library

Then, using the spatial coordinates of the neighborhood and the developer account of FOURSQUARE, requests were made to FOURSQUARE's API by providing the neighborhood's spatial coordinates and user developer account credentials to obtain 100 venues situated within 500 meters of the spatial coordinates provided. These requests were answered with JSON file containing information about 100 different venues (if present) situated in the spatial coordinates provided for a neighborhood and metadata associated with each venue, like **v**enue id, venue name, venue category, venue spatial coordinates, and more. Using the venue id, requests were further made to the API to obtain detailed metadata on each individual venue, which included, but not limited to, venue rating, venue tips, venue distance, venue menu, venue photos, venue likes, venue comments by users, user photos, etc.

A collective table of 6201 venues and 13 features was generated where each venue had information on its 13 features which provided information on venue's borough, venue's neighborhood, neighborhood latitude, neighborhood longitude, venue id, venue name, venue latitude, venue longitude, venue category, venue shortname, venue costliness, venue likes, and venue rating, respectively.

3. Exploratory Data Analysis

3.1 Venues by Neighborhood