IBM Data Science

Analyzing NYC Neighborhoods For Starting A New Restaurant.

Capstone Project

Shaban Khan

12/19/2019

Introduction

New York City is one of the biggest cities in the world and also one of the most populous cities of United States of America (Cox, 2017). The population of New York City comprises of people from all over the world and belong to Approximately 37% of the city's population is foreign born and more than half of all the children are born to foreign mothers who are immigrants (Semple, 2013). Geographically New York City is divided into five boroughs namely The Bronx, Brooklyn, Manhattan, Staten Island and Queens. The city is full of restaurants serving millions of hungry customers, every day. The diversity in the population of the city has brought in a vast diversity in food habits of people. In this project we will study the food habits of New Yorkers as our business problem belongs in that area.

Business Problem

Our client is an investor who is interested in investing in an Indian restaurant in New York City. They have approached us to study the market and suggest them a location in one of the neighbourhoods which would be in best interest of the business. Our main objectives of this project would be to extract and analyse right data about various neighbourhoods of New York City using various data science techniques and suggest our client a fitting location for their Indian restaurant.

Data

In order to achieve our final goal we will need the following data:

- New York Neighbourhoods data.
- Geographical coordinates of the neighbourhoods.
- Venue data from FourSquare.

Neighbourhoods Data: This data will be extracted from New York Neighbourhoods Wikipedia page (https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City) using web scraping with BeautifulSoup library in Python. This will give us a detailed list of neighbourhoods present in New York City divided in five boroughs.

Geographical Coordinates: Later, the geographical coordinates of various naighbourhoods will be extracted using GeoPy library in Python. Geographical coordinates are necessary for plotting maps during the project for visualizing our data.

Venue Data from FourSquare: In the end, venue data will be extracted using FourSquare API. This venue data will be used to study the restaurants in various neighbourhoods in New York City. This data will provide important details of various restaurants in the area and will help us understand the competition. This data is very important because it will help us draw the main conclusion of the project.

Works Cited

Cox, W. (2017, December 6). *America's Densest Cities*. Retrieved from Huffpost.com: https://www.huffpost.com/entry/americas-densest-cities_b_5888424

Semple, K. (2013, June 8). *Take the A Train to Little Guyana*. Retrieved from nytimes.com: http://archive.nytimes.com/www.nytimes.com/interactive/2013/06/09/nyregion/new-york-citys-newest-immigrant-enclaves.html?pagewanted=all