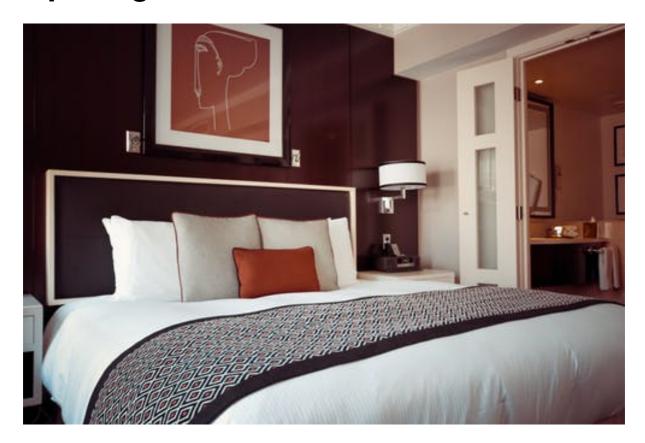Sahil Sharma
Coursera Capstone
IBM Data Science Course
8 April 2020

Opening a New Hotel in New York

Opening a new business can be challenging. Choosing a location for the business can be even more challenging. Hotels are one of the most important thing in any neighborhood. When people visit any place, having a hotel at a convenient location is the best thing they can wish for. In this project, I will be creating a Machine Learning Algorithm which uses Foresquare's data and analyze where can a developer open a new hotel so that there is very little to no competition from existing malls.

Business Problem

The objective of this project is to analyze and select the best locations in New York City where one can open a new Hotel. Using Data Science and Machine learning technology like Clustering, can we answer this simple question: Where to open a new Hotel in New York.

Data

To solve this challenge, I will be using the following data:

- List of neighborhoods in New York - This defines the scope of this project which is confined to New York City.
- Latitude and Longitude of the neighborhoods in New York City - This is required to visualize the places and the the venues.
- Venue data, Particularly related to the hotels in the neighborhoods.

Source of the data

I will be using data from https://geo.nyu.edu/catalog/nyu_2451_34572 where we can find the List of neighborhoods in NY and other features as well. We will be only needing the neighborhoods and the boroughs data from it. The data is in the format of JSON so we will need to modify it ad well. Then we will get the Longitude and the Lattitude of the neighborhoods and use them.

From Foursquare, we can get the common areas nearby and check how many hotels are there in the neighborhood. Foursqaure has one of the largest database of places in the whole world and we can get a lot of information from it. Here we are only interested in the subcategory "Hotel"