Coursera Capstone

IBM Applied Data Science

Expanding a High-End Irish Pub to New York City

By: Will Pratt

May 2020

Introduction/Business Problem

The city of Boston is well known for many of its citizens having Irish roots. With those roots comes many Irish pubs. The basis for this study is to help a group of investors in Boston expand their successful Pub chain to the New York City area. New York City offers a vast and diverse market that ranks among the most populous cities in the world. Due to the complexity and diversity in the city, the group of investors has asked for assistance in selecting the correct area to expand their restaurant to.

They believe that an area must meet certain criteria for them to consider it:

- It must be considerably wealthy, because of the nature of the high-end restaurant.
- It must have a dense population of people that live near it
- It must not have too many similar restaurants already around their new location.

In the past these criteria have been great indicators on whether the has restaurant fared well or not. The investors believe that leveraging both population data as well as location data, will provide a strong base for a profitable expansion.

Data

To solve the problem the following data sources will be used:

- New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.

  Source: *https://cocl.us/new_york_dataset*

- GeoSpace Data

  Source*: https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm*

  This data will help to get the boundaries for the Boroughs and neighborhoods for visualization

- Venue data

  Source: Foursquare API

  This api will help us to get all the venues from each neighborhood

- Census data

  Source: https://www.kaggle.com/muonneutrino/new-york-city-census-data

  This dataset contains demographic and economic data for NYC

- NYC population data

  Source: https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoodpop.htm

  This data contains regions of New York and their subsequent populations

By leveraging these 5 datasets, we will be able to adequately answer all of the investor's criteria marks for each neighborhood and cluster each based on their desirability for expanding the restaurant.