

Restaurant Expansion from Boston to Manhattan and Chicago

IBM Applied Data Science Capstone Project

By Tiffany Brand

August 2019

Introduction

Successful businesses looking to expand may seek similar cities and neighborhoods to open new locations. The type of neighborhood including types and numbers of venues present can contribute to the success of the business' current location, so looking for similar neighborhoods in other cities can help focus the search for business expansion to new areas. Entrepreneur.com in their article "Expanding with a Second Location" suggests that business owners should look at the strengths of their existing location to inform an expansion to a new location. They recommend considering other shops located nearby that may help generate and sustain traffic and performing market research on proposed new locations. The research conducted in this project will focus on looking at the venue makeup of the current neighborhood and finding a similar composition in potential expansion neighborhoods using clustering techniques.

Business Problem

The purpose of this project is to assist a potential client with a successful restaurant business located in Boston to choose possible neighborhoods to open other restaurant locations in Manhattan and Chicago. This project will look at the numbers and types of venues in Boston, Manhattan, and Chicago and use data science clustering methods to find clusters of similar neighborhoods.

Target Audience

A business owner who wishes to expand a successful business in one city to other cities.

Data

To solve this problem, the following data will be used:

Neighborhood boundaries for Boston, Chicago, and Manhattan:

Boston: The Boston Neighborhoods dataset can be found here:

<https://data.boston.gov/dataset/boston-neighborhoods> - This dataset is found on the Analyze Boston website created by Boston's GIS department. The data is a combination of zoning neighborhood boundaries, zip code boundaries, and 2010 Census tract boundaries. They state that the neighborhood boundaries are not official but can be used in a broad sense for visualization and planning. The GeoJSON version of the data will be downloaded. The following features will be extracted: Neighborhood name and the latitude and longitude data marking the boundaries of the neighborhood. From the boundaries, centroid coordinates of the neighborhoods will be calculated.

Chicago: The Chicago Boundaries – Neighborhoods dataset can be found here: <https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Neighborhoods/bbvz-uum9> - This dataset is found on the Chicago Data Portal and was developed by the Office of Tourism to describe the approximate neighborhood names of Chicago. The GeoJSON version of the data will be downloaded. The following features will be extracted: Neighborhood name and the latitude and longitude data marking the boundaries of the neighborhood. From the boundaries, centroid coordinates of the neighborhoods will be calculated.

Manhattan: The New York City neighborhoods data can be found here in a JSON file: https://cocl.us/new_york_dataset (this was the dataset used for the week 3 lab in this capstone course). The dataset is found on the NYU Spatial Dataset Repository https://geo.nyu.edu/catalog/nyu_2451_34572. It was created as a guide to New York City's neighborhoods. The following features will be extracted: Neighborhood name, Borough, Latitude and Longitude. Using the borough feature, the neighborhoods from Manhattan will be extracted.

Foursquare Venue Data:

Using the Foursquare API (<https://api.foursquare.com>), data on nearby venues to each neighborhood will be extracted. Foursquare houses a global database of 105 million places with more than 70 venue attributes and more than 900 categories. Features extracted will be Venue Name, Venue Category, Venue Latitude and Longitude.

References

“Expanding With a Second Location.” *Entrepreneur*, 23 Feb. 2003, www.entrepreneur.com/article/47552.