

The Detroit Project

I. Introduction – The Business Problem

Like many cities around the United States and the world, Detroit was profoundly impacted economically by the financial crisis of 2008. In addition to this global shock, Detroit was already in the process of an economic decline over the preceding years and decades. The Detroit of world renown was an automotive manufacturing superpower, but its peak may be traced back as far as the early 1950's when the city was the fourth largest city in the United States and home to 1.8 million people.¹ Since this peak, people and businesses have left Detroit for a variety of reasons, dropping the total population drastically down to a current level of 670,000.² The city even experienced a major bankruptcy in 2013.³

In the intervening years since the 2008 financial crisis, the city of Detroit has been the subject of several programs to revitalize the city and grow the business community. These programs were initiated by a variety of prominent organizations and individuals, including the University of Michigan Business School, Quicken Loans founder Dan Gilbert and Democratic presidential candidate Andrew Yang. Despite these tremendous efforts, attracting businesses and people to a city with the economic struggles of Detroit is quite difficult. Much work still needs to be done to return Detroit to its former vibrancy. The goal of this particular project is to leverage Foursquare API data for Detroit and Manhattan to make some recommendations for businesses that can be started in the city of Detroit and the neighborhoods where those businesses may flourish. If successful, the results of this project may be useful to those who are trying attract businesses to Detroit, those who may be considering starting a business in Detroit and the local communities who would benefit from the arrival of new and successful businesses.

a. Methodology

This project and its analysis will entail (i) collecting latitude and longitude data on neighborhoods in both Detroit and Manhattan, (ii) using that geographical data to collect information on business types in those neighborhoods and (iii) using a K-means clustering algorithm to cluster the neighborhoods based on their respective business types. The goal is to get a cluster or two that contain both Manhattan and Detroit neighborhoods and find recommendations for a business or business type from a Manhattan neighborhood that can be started in a Detroit neighborhood.

¹ https://en.wikipedia.org/wiki/Demographic_history_of_Detroit

² <https://en.wikipedia.org/wiki/Detroit>

³ Ibid

II. The Data

Geolocation Data

Neighborhood names and latitude/longitude pairs for Detroit can be scraped from https://en.wikipedia.org/wiki/Template:Neighborhoods_Detroit_Index, and the corresponding information for New York can be obtained from https://geo.nyu.edu/catalog/nyu_2451_34572.

Once this data is scraped and cleaned into a data frame, it can be used to access the Foursquare API.

Foursquare API Data

The Foursquare API provides a variety of data about venues visited by Foursquare app users. Using the geolocation data for the neighborhoods, the Foursquare API will be queried to provide information on venues near or within the neighborhoods. The data that will be primarily used for clustering analysis is the category of each venue. As an example, the following are a few of the venues returned for the Marble Hill area in New York:

	name	categories	lat	lng
0	Arturo's	Pizza Place	40.874412	-73.910271
1	Bikram Yoga	Yoga Studio	40.876844	-73.906204
2	Tibbett Diner	Diner	40.880404	-73.908937
3	Starbucks	Coffee Shop	40.877531	-73.905582
4	Astral Fitness & Wellness Center	Gym	40.876705	-73.906372

Between the Detroit and New York neighborhoods, there should be a few hundred venue categories which can be used to cluster the neighborhoods. The K-means clustering will be based on the frequency of venue types in each neighborhood on the belief that a successful business in a New York neighborhood can also be successful in a Detroit neighborhood of the same cluster.

If a particular business seems like a good recommendation after the cluster analysis, the Foursquare API can provide further detail on that business. For example, if Bikram Yoga seems like a good business to bring to Detroit, the Foursquare API can provide other important details, such as hours, price level and user ratings.

Other Data

Other data publicly available on the internet may be used in the analysis for this project, such as basic demographic or business information. However, it is not currently anticipated that other large datasets or APIs will be used.