Focusing the Search for Neighborhood to Start a Brewery

Thor Hanson

October 17, 2019

1. Introduction

1.1 Background

The number of craft breweries in the United States has been steadily increasing over the last decade. According to the Brewers Association (brewersassociation.org) website, the count of U.S. based craft breweries has increased by over 482% since 2008. Additionally, the Brewers Association relays that the majority of Americans live within 10 miles of a craft brewery and that one of the defining concepts of craft brewers is their involvement in their communities. The importance of the location of a craft brewery coupled with the support they have from and give to their communities is widely recognized. Therefore, it is necessary, for anyone looking to open a craft brewery, to understand the neighborhood in which they are starting their business.

1.2 Problem

Data that contributes to the understanding of locations and the types of neighborhoods where current craft breweries exist can be useful to those looking for a location to start a craft brewery. This project aims to use data available about the neighborhoods of current craft breweries to focus the location search for neighborhoods that would be the most conducive to starting a new craft brewery.

1.3 Interest

There are many factors that go into the final selection of a site for a new craft brewery so the earlier in the process that one can focus on potential locations is beneficial. The audience for this project is anyone looking to focus the search for a neighborhood in which they intend to start a craft brewery.

2. Data Acquisition and Cleaning

2.1 Data Sources

The data required for this project is current craft breweries and their locations, neighborhood information, and geographic. I obtained data on current craft breweries from the Brewers Association website at https://www.brewersassociation.org/directories/breweries. The

data attributes selected were the brewery name, address (street address, city, state, zip code), and brewery type (taproom, microbrewery, brewpub, regional, large). There were a few additional data attributes available for the breweries such as phone number and web site that were inconsistently available and not likely relevant to this analysis so they were not selected for download. Data for the neighborhood information was obtained through the Foursquare API (https://developer.foursquare.com/). Geographic data was accessed through a geocoding API, ArcGIS for Developers, with more information available on their website (https://developers.arcgis.com/python/).

2.2 Data Filtering and Cleaning

The brewery data that was downloaded from the Brewers Association was very clean. The data was downloaded for Midwest region states of the U.S. and filtered to only look at taproom breweries. The data could be expanded to include more/all regions of the U.S. and for more/other brewery types. The filters applied were selected due to my own preferences and interests in opening a taproom type brewery in the Midwest. There was very little data cleaning required for this data set as the only missing data elements were the street address for only three breweries. Quick web searches provided the information needed and the missing data was added to the set manually.

The neighborhood information data was obtained using the Foursquare API by using the API's explore function to search for venues near a given geographic coordinate. A radius of 2.5km was used with a limit of 1000 venues sampled per location. There was no cleaning of this dataset required beyond translating the returned data format and storing in a Pandas Data Frame for easier use.

The geographic data needed was the geographic coordinates (Latitude, Longitude) that corresponded to the brewery addresses obtained in the brewery dataset from the Brewers Association. This data was stored with the brewery data in the in Pandas Data Frames and was received in a format accepted by the Foursquare API so no further cleaning or translating was required.

2.3 Exploratory Analysis of the Data

Familiarization of areas where this analysis could be applied was the first type of exploratory analysis done. Two cities (Des Moines, IA; Madison, WI) where I had some interest in exploring with this analysis were mapped. They were mapped with nearby cities/suburbs that surrounded the chosen city using zip codes obtained from the website Best Places using the links provided at the end of this paragraph for Des Moines and Madison respectively. The maps were investigated to gain familiarization with surrounding areas to be explored.
(https://www.bestplaces.net/find/zip.aspx?st=ia&msa=19780)
(https://www.bestplaces.net/find/zip.aspx?st=wi&msa=31540)

The second exploratory analysis that was done was to map the 593 breweries to view the distribution of the breweries though the Midwest. As expected, many of the breweries were clustered around the larger cities.

The third exploratory analysis that was done was reviewing the brewery data set including the nearby venues in Foursquare. There are 593 samples with 5 attributes in the brewery data set. After exploring the venues associated with the 593 breweries using the Foursquare API, we discover that only 579 breweries have venues in the Foursquare dataset that are within 2.5km of the breweries. These 579 breweries have a total of 31,287 venues in their neighborhoods. It is decided that this is a large enough sampling to continue with the analysis and this will be reevaluated after initial results are obtained.