

Finding the best place to open a beer bar in Nashville (TN)

Capstone Project - The Battle of Neighborhoods (Nashville)

Applied Data Science Capstone by IBM/Coursera

2. Data

Based on definition of the problem, factors that will influence the decision are:

- number of existing bars in the neighborhood (any type of bar)
- number of and distance to beer bars in the neighborhood, if any
- distance of neighborhood from city center (Broadway)

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding,
- number of bars and their type and location in every neighborhood will be obtained using Foursquare API,
- coordinate of Nashville center will be obtained using **Google Maps API geocoding** of famous location (Broadway).

To solve the problem the following data is needed:

- List of issued beer permits and locations
- latitude and longitude coordinates of locations. This is required in order to plot the map and get the venue data.
- data related to bars. Data will be used to perform clustering on the neighbourhoods.

Sources of data and methods to extract them:

The Nashville Open Data Portal <https://data.nashville.gov> will be used as a source of data of all active beer permits in Davidson County. The data is available from <https://data.nashville.gov/Licenses-Permits/Beer-Permit-Locations-Map/a93h-xhg6> and can be exported in CVS format.

There are 2,065 active beer permits in Davidson County. We will use read method to obtain data. After that, we will use Foursquare API to get the venue data for those places. Foursquare has one of the largest database of 105+ million places and it will provide many categories of the venue data. We are particularly interested in the bar category in order to help us to solve the business problem put forward. In this project we will make use of many data science skills, from reading data, working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).