**Finding the optimal Zip Code in Richmond, VA for a new/expansion brewery**

**Brandon Kolljeski**

**August 25, 2020**

For this study I am going to try and find the optimal location in Richmond, VA to open a new or expansion brewery. The idea of a business cluster(industry cluster, competitive cluster, Porterian Cluster) was popularized by Michael Porter in his book The Competitive Advantage in 1990. This idea is not new and has been around since the 1890’s with the work of Alfred Marshall and being termed the agglomeration economies. The idea states that similar businesses built close together gain an advantage by generating a few different effects in the localized industry. These effects can include pricing, more customer attraction, innovation in the industry and localizing talent to name a few. This is something I have personally been a part of in the Richmond craft beer community. In late June we went to a district in Richmond to visit a brewery and at the time they were full, so we ended up walking a short distance to the next brewery and spent our money there. If they hadn’t been located within a reasonable distance from one another we would have left the area all together, but because of the convenience we chose to stay in the area and spend our money there. That was what also drew us to the area in the first place. Knowing that if one business did not have seating or a product that we enjoyed we could quickly move on and find something else within a short commute. By looking at the breweries and accommodations in each zip code in Richmond we will be able to see which areas give the best opportunity to attract spill over crowd, people looking for something different or people just enjoying the area in general and stumbling upon your business by simply being there.

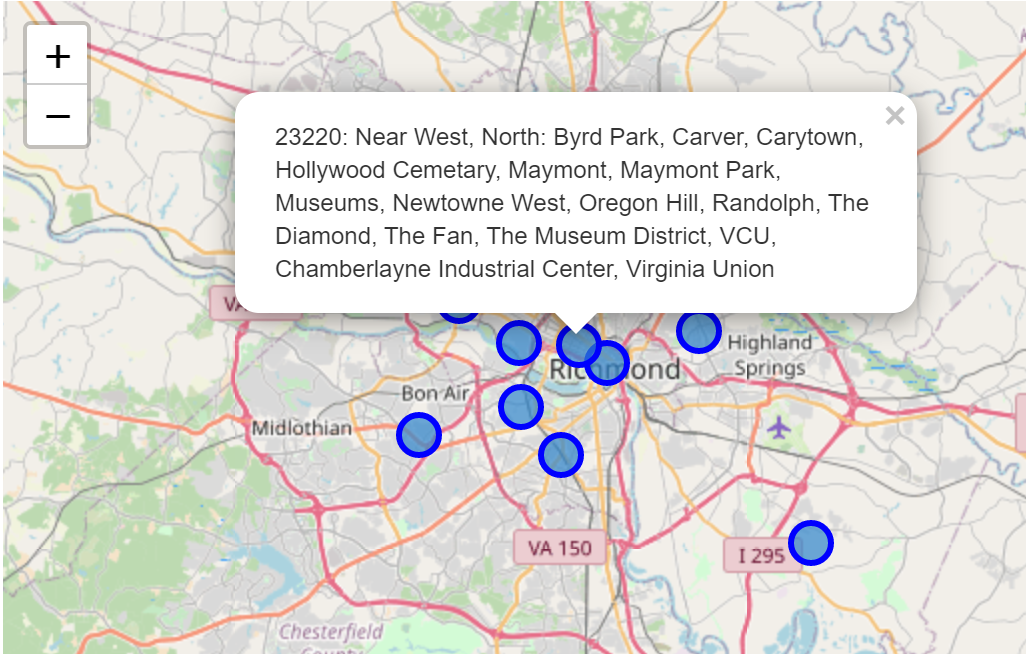
This study would help a new brewery or expansion brewery in making an informed decision on the most probable location to have their business succeed. It would also help the business to attract investors if they could show that there are many current businesses in the area doing well and could benefit from joining the community.

Collecting data for this project was time consuming and tedious. There were no data sets available to the public that included the zip codes, districts and neighborhoods for Richmond, VA, so each one had to be found on its own and then cleaned to fit a larger data set that was used for this project. The first data set I was able to find were the districts in Richmond and their zip codes from a Virginia government website. From there I collected the neighborhoods belonging to each district then cleaned and merged the data sets (Table 1). The last part was collecting the Latitude and Longitude information for each zip code and adding it to the data set to be able to map out the areas using folium (Map 1). Working with the cleaned data sets I was able to start to get a picture of how Richmond was broken up with the zip codes.

Table 1. Zip Code, District, Neighborhood and Lat/Long data table

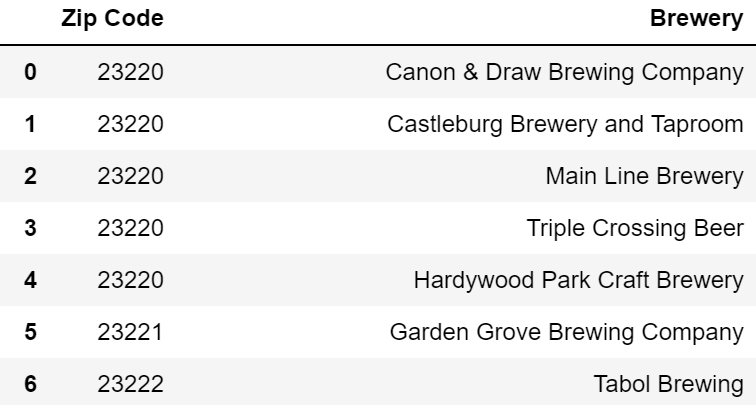
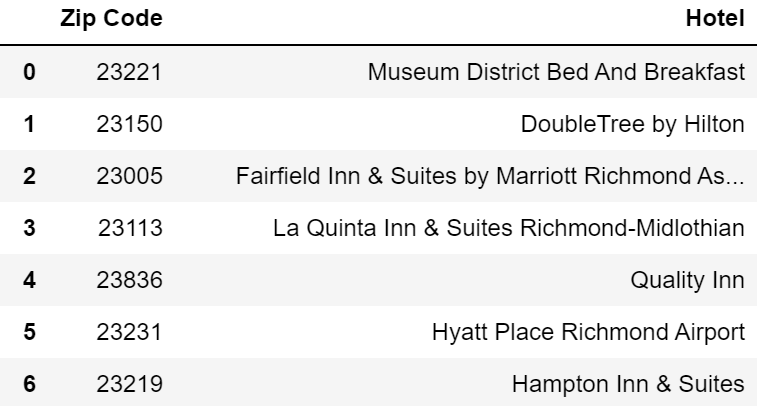


Map 1. Zip Codes, Districts and Neighborhoods mapped out



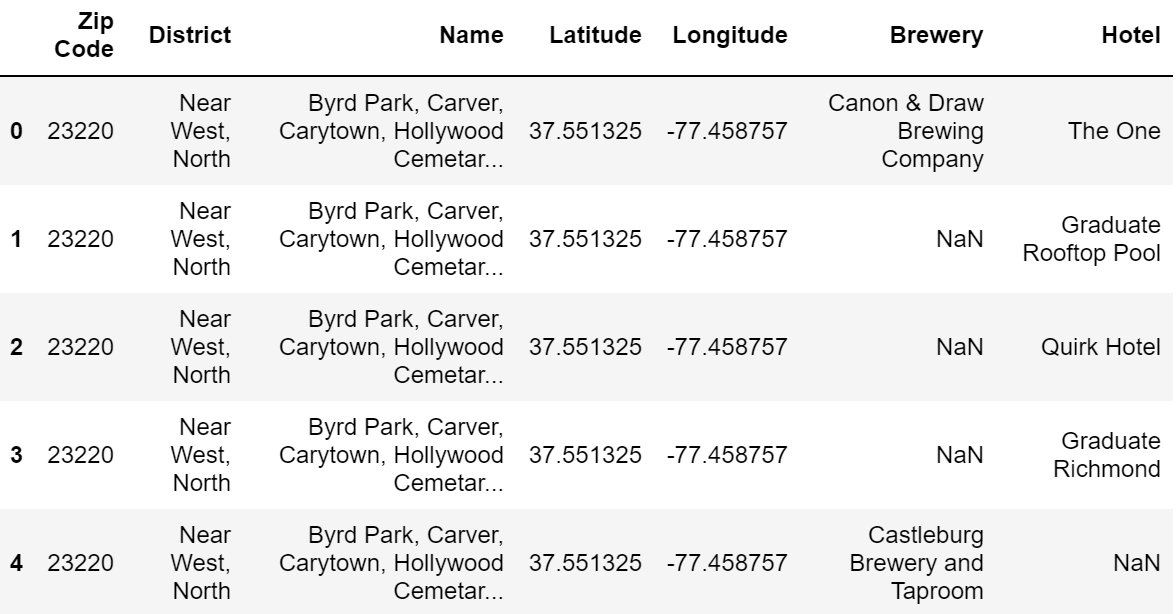
After collecting the location data I moved on to start collecting brewery and hotel data from FourSquare using the API, I made the get requests to find breweries (Table 2) and hotels (Table 3) in Richmond. These data sets were taken from the JSON files and then moved into Excel for cleaning and processing using Power Query. Cleaning the data was fairly straight forward and I only needed to remove a few rows because of missing information or the Get request pulling in places outside of the immediate Richmond area.

Table 2. Breweries and Zip Codes Table 3. Hotels and Zip Codes

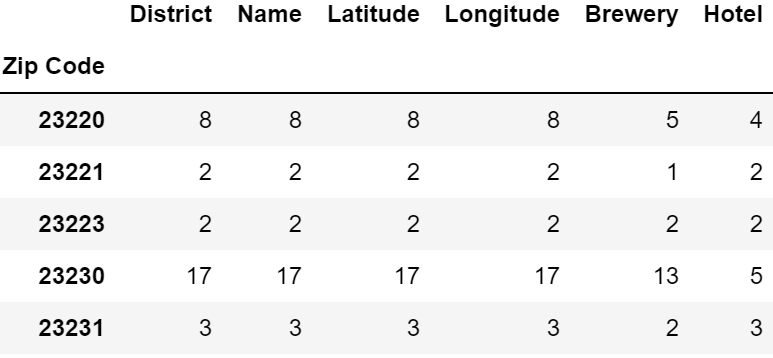
Once the needed information was cleaned and ready these tables were merged with the bigger Richmond data set by their zip codes (Table 4). At this time I chose to only use zip codes that had both breweries and hotels available to give the best showing of possible areas for the new business.

Table 4. Full data set showing zip codes, with breweries and hotels



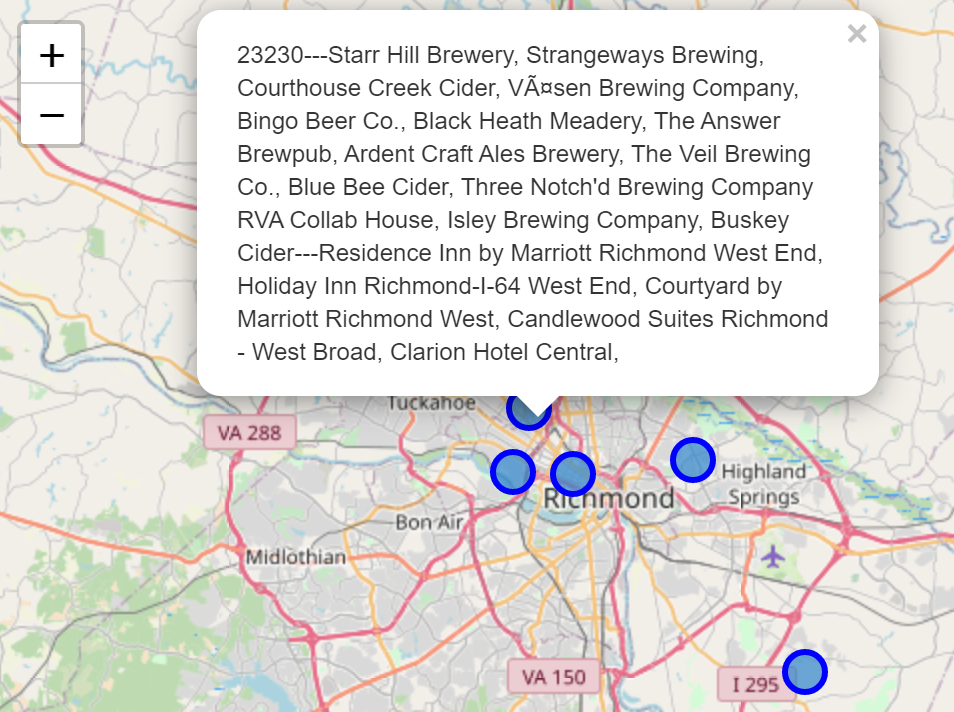
Now with all the data merged into a single table I was able to start counts and to look at areas with the highest number of breweries and hotels. Only 5 of the zip codes ended up containing both of the businesses we were looking for (Table 5).

Table 5. Zip Codes showing count of breweries and hotels



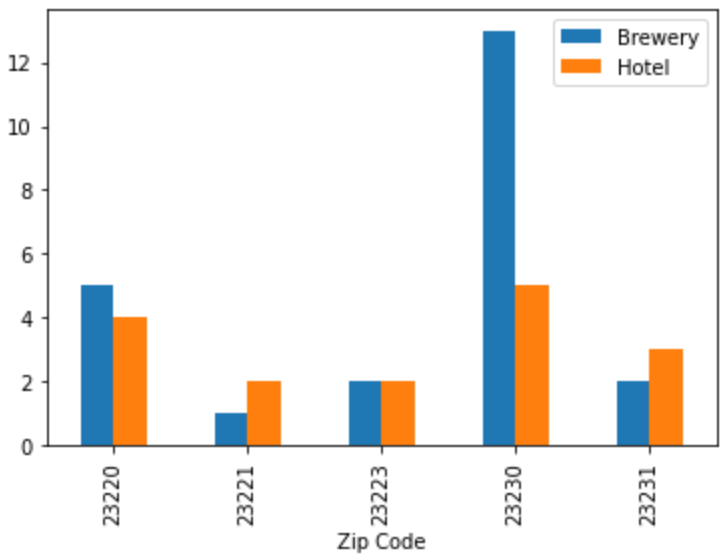
Using the combined data sets I was able to create another map using folium that used markers to show each zip code and the breweries and hotels contained in that area (Map 2).

Map 2. Showing zip codes with their breweries and hotels



Also using that table I created a bar graph to show each zip code and the number of breweries and hotels for easier consumption in a visual format (Figure 1).

Figure 1. Bar graph showing number of breweries and hotels for zip codes



Using the data sets combined we are able to use kMeans clustering to see which zip codes group together (Map 3). With such a small data set from the Richmond area it was already clear where the best zip code would be, but to be sure the algorithm was run and the groups ended up with Cluster 0 being 23231 having only 40% breweries and 60% hotels. Cluster 1 was 23230 with 72.2% breweries and 27.8% hotels. 23220 and 23223 in Cluster 2 because of sharing almost a 50/50 split between breweries and hotels. Leaving Cluster 3 in 23221 having 33% breweries and 67% hotels (Tables 6&7).

Map 3. Showing clusters created with kMeans

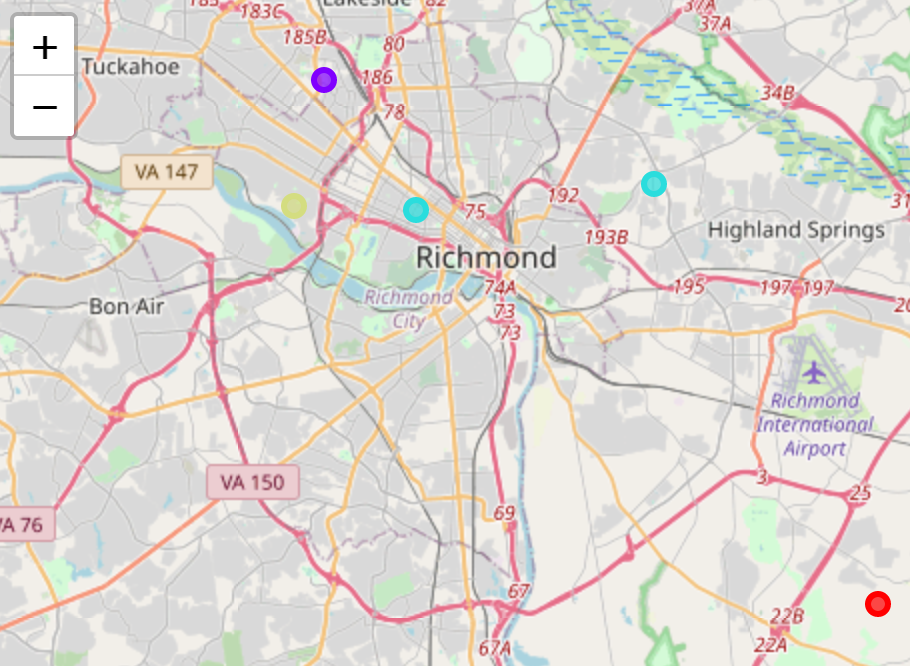
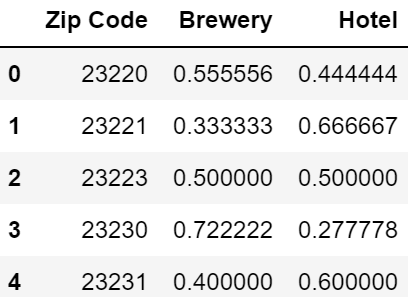
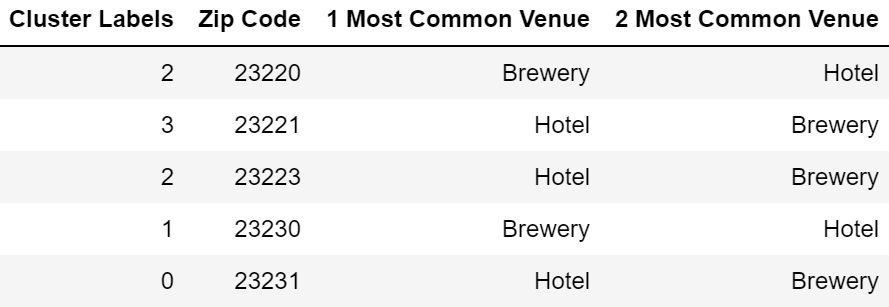


Table 6. Showing brewery and hotel % Table 7. Showing the cluster labels and most common venues

Using this data to create the maps and graph shown, it was clear that the 23230 zip code had the most breweries by a large margin and most hotels. From this simple look into the data we can say that the 23230 zip code would have the largest chance of putting you next to similar businesses and having accommodations close to your business to help bring in more business and the possibility to generate larger profits.

For a future endeavor there are many extra data sets I would like to explore to dive further into this problem. The profitability of breweries, count of check-ins at each brewery, untappd ratings, which beers sell the best or are rated the best from each brewery and many more data points would give a better picture of the breweries themselves. Looking at the hotels I would like to see what their occupancy is at certain times of the year and if that trends with brewery profitability, and if the higher occupancy hotels are nearer to the breweries with the highest profits. Other aspects of the area would be also be of interest such as concert venues, museums, art galleries, and many other businesses in these areas that could possibly attract a crowd that would bring in larger profits for the business. Looking into when the hotels were built and if that coincided with any other business types in the area, or if the breweries were all started around the same time and if that helped to grow the area would also be an interesting data set to be able to pull into the overall analysis.