Business Opportunities Exploration of Richmond, Virginia

September 22, 2020

James M. Comstock

1. Introduction.

    Richmond, Virginia, is a mid-size American city of approximately 230,000 people. After decades of stagnation, Richmond has enjoyed economic and cultural renewal in recent years. In 2019, *Trailblazer Magazine* showcased the "revolution" in Richmond, highlighting the success of the dining scene, craft breweries, industry, and the arts[1]. Where might investors and entrepreneurs find opportunities in this environment? The research and analysis presented here sought to provide preliminary answers to these questions. Marketing firms, local investors, and entrepreneurs interested in the region could all benefit from a data-driven consideration of the Richmond business scene. To that end, commercial business analysis, using venue information and aggregate personal income data, illustrates a city with business diversity and room for growth.

2. Data.
    a. <u>Sources</u>. This research relied on three primary data sources: the Foursquare online application[2], US Internal Revenue Service (IRS) income data, and City of Richmond neighborhood parcel information. Foursquare provided detailed information on all businesses within the city limits of Richmond. Aggregated IRS income tax data from 2017 provided a zip code-level view of residents' personal income.[3] This information enabled a deeper understanding of each zip code, as it provided some context on how well the residents of each area are doing financially. The City of Richmond data portal includes a Neighborhood Parcels data set.[4] This provided reference data information on which neighborhoods were in each zip code, as well as the geographical planning districts in which each neighborhood resides.

    b. <u>Acquisition and Cleaning</u>. After acquiring the data sets from the City and IRS, the first step was establishing the neighborhood zip code and location information. This required extraction of zip codes and neighborhoods from City data. A parallel effort yielded the 2017 income tax data by zip code for Virginia. Extracting useful income data required aggregation of total gross taxable income in each zip code.

3. Methodology.
    a. <u>Initial Exploration</u>: Examination of City data denoted 148 neighborhoods. Creating a Folium map of neighborhoods provided a first glimpse of the distribution of neighborhoods throughout the city, depicted below in Figure 1. The Folium mapping package enabled the creation of this and all subsequent maps. Visualizing the neighborhoods shows a generally tighter concentration in the city center and wider distribution as you go further south and west. Seeing the number and distribution of neighborhoods in a city of this size provides an initial glimpse of the diversity of the city as well.
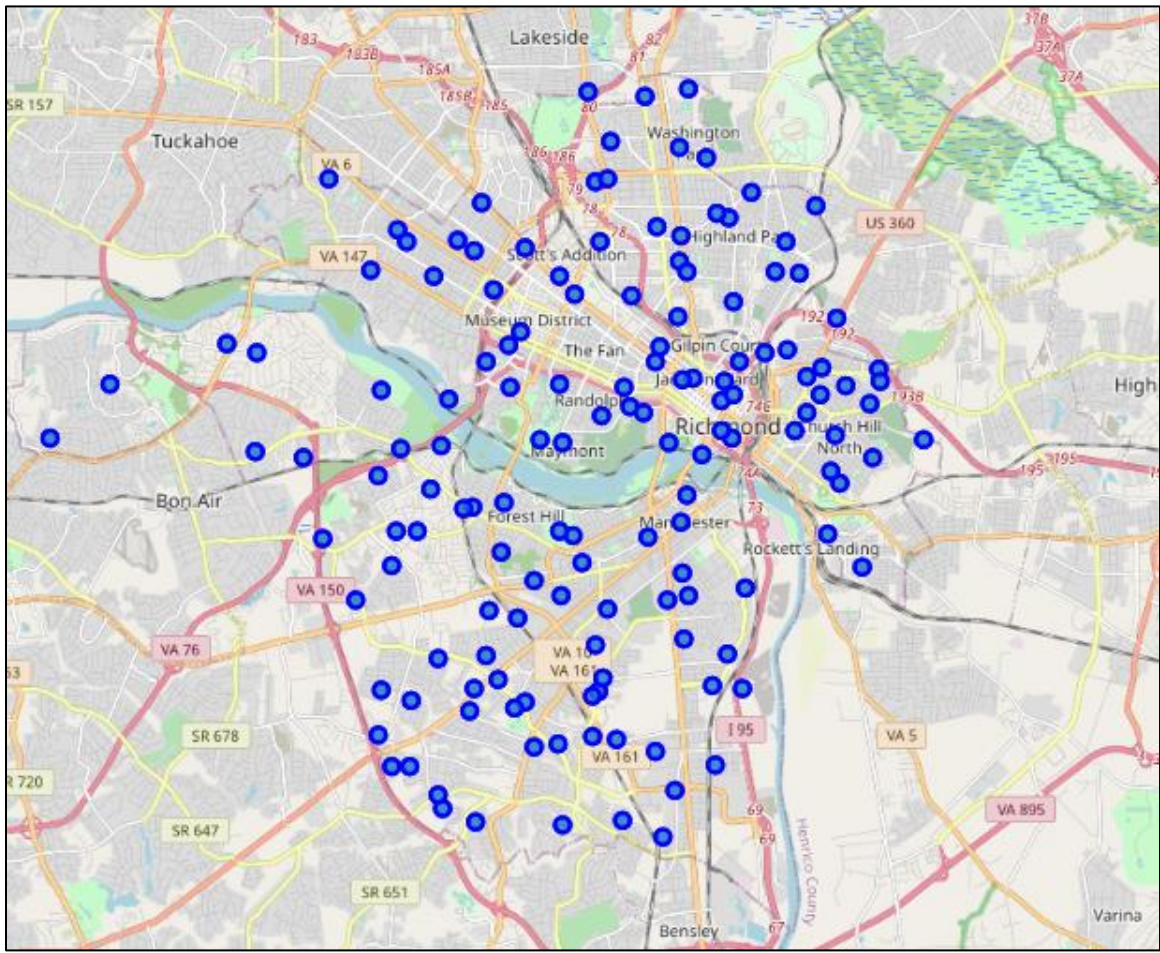
Figure 1. Map of Richmond Neighborhoods.

b. <u>Mining Foursquare Data</u>: Extraction of business venues in the City of Richmond enabled the creation of a data frame with all of the City's businesses. This involved the use of a

c. <u>Cleaning and Initial Analysis of Venues</u>: Grouping of venues by neighborhood provided the means to see all of the businesses in any given neighborhood within the City. Combining the relevant data into a synthesized whole produced a snapshot of the business and geographical landscape. In the 13 zip codes of Richmond, there are over 1300 commercial venues with over 220 venue category labels. Sorting the top ten venues by neighborhood illustrated the prevalence of business types in each neighborhood. This provided the raw material for machine learning.

d. <u>Clustering</u>: Grouping venues by neighborhood and sorting the top ten venue categories in each neighborhood provided the relevant content to analyze each neighborhood. Subsequent use of one-hot encoding provided the numerical data points for machine learning, in this case K-means clustering. Clustering identifies patterns in the data and determines category labels. In the case at hand, employing a clustering algorithm provides information on what business trends distinguish neighborhoods. K-means clustering analysis yielded 5 clusters. Use of the algorithm with and without income data yielded the exact same results, demonstrating that income data itself was not a determining factor, although it may inform the underlying characteristics.

e. <u>Visualization</u>: Creating a Folium map visualization of clusters showed the layout of clusters within the City.
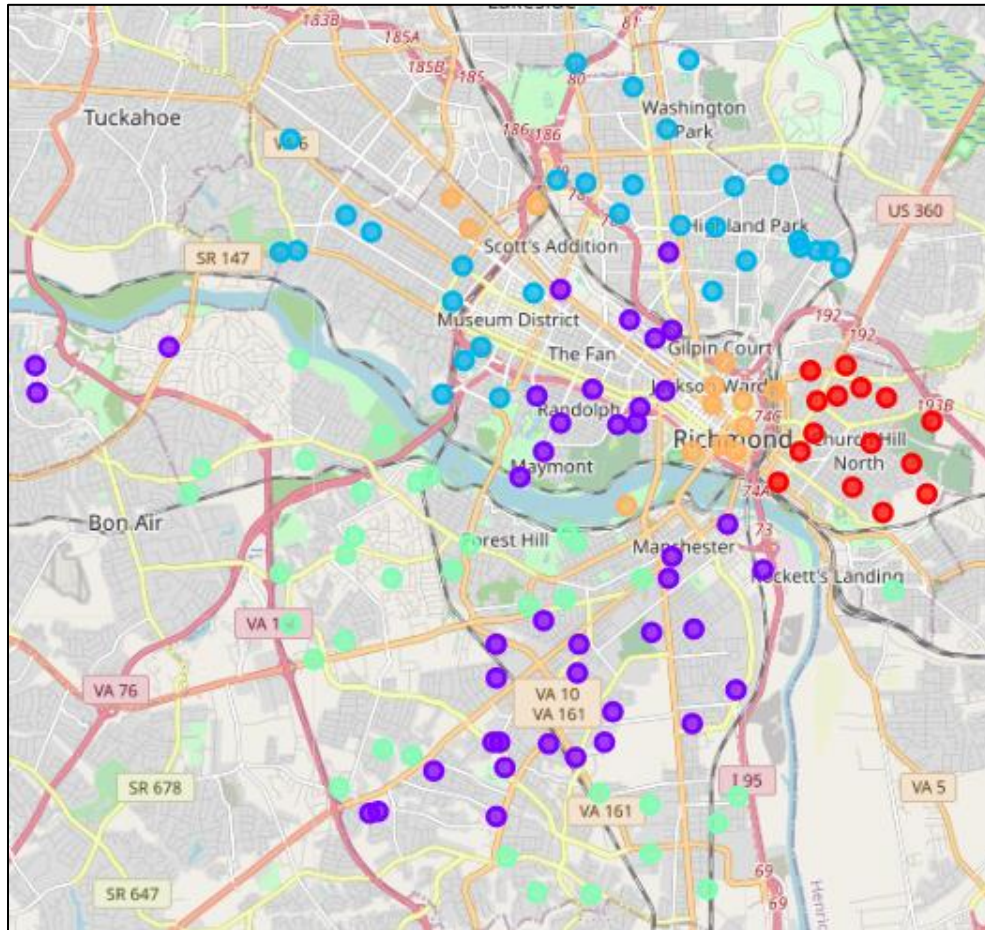

Figure 2. Neighborhoods by Cluster.

f. <u>Case study example</u>: Pizza restaurants, or parlors, are a common feature of the American food service industry and provide a narrow focus for further analysis. Extracting pizza venues from the assembled data yielded some interesting observations. The distribution of parlors throughout the city varies greatly and does not neatly correspond with income. Linearly, there is a large concentration along the Monument Avenue corridor. In terms of proximity, there are distinct clusters around the Museum District and southeast end of The Fan.
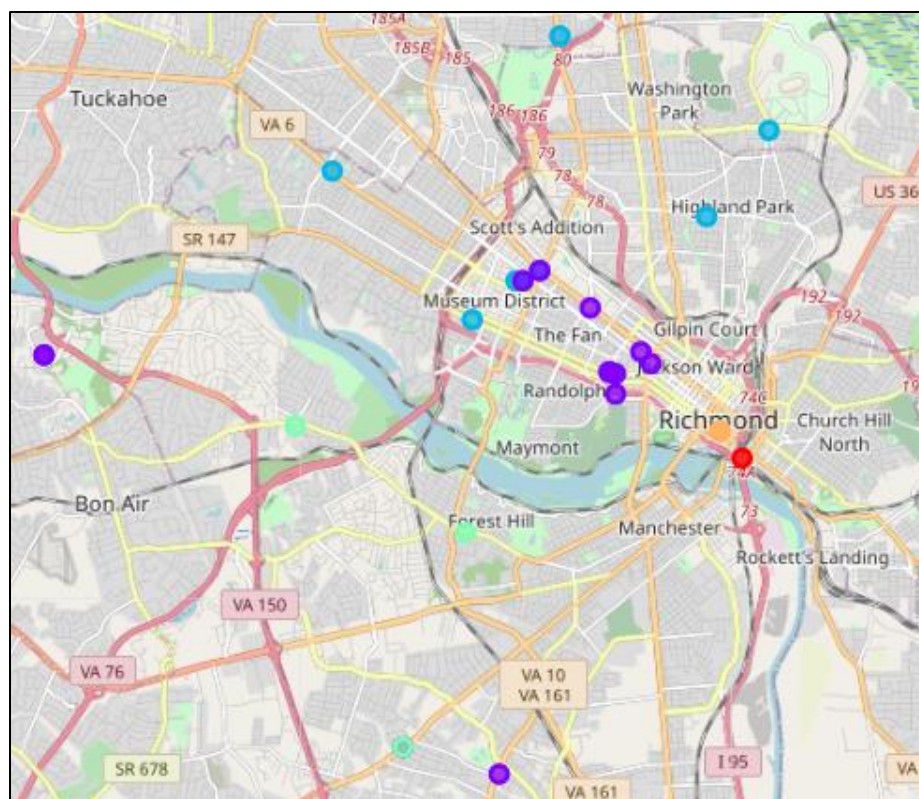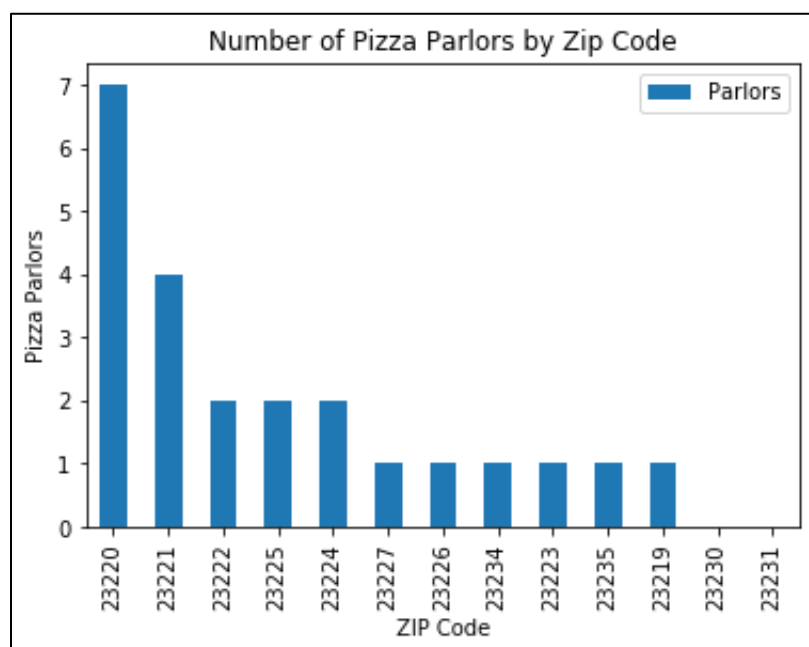
Figure 3. Richmond Pizza Parlors (color by cluster)

Figure 4. Pizza Parlors by Zip Code

Zip Code	Zip Code Total Income	No. of Parlors	Cluster Label
23223	24430	1	0
23234	20630	1	3
23225	20110	2	3
23231	18370	0	3
23235	16860	1	1
23224	16330	2	1
23220	15050	7	1
23227	12540	1	2
23222	11970	2	2
23226	8480	1	2
23221	8400	4	2
23230	4170	0	4
23219	2380	1	4

Figure 5. Parlors by Zip Code and Income Levels. Income shown in $1000s.

4. Discussion.

Exploratory analysis indicates a diverse business environment. Examining the results of venue analysis, there are characteristics that uniquely define the 5 clusters of city neighborhoods. For an investor or entrepreneur, there are at least two general methods to select a location for a new venture. With a business category already in mind, the investor could select a neighborhood based on the absence or near absence of comparable businesses. For example, in the pizza parlor case, there were two zip codes with no parlors. Another option is to select a location based on a combination of income level and relative absence of desired business type. Applying this method to the pizza case, two high-income zip codes (23223 and 23234) have only one parlor.

The pizza parlor case study provides multiple avenues for investment. While some areas are oversaturated, the fact that there are zip codes with few or zero parlors suggests an opportunity-rich environment. High income areas are ripe for development of a boutique parlor venue. For a supply vendor, targeting the parlors by location and/or zip code income could enable either high volume or specialized partnership.

5. Conclusion.

Richmond has a diverse business community and is ripe for continued investment. Higher income neighborhoods are ripe for targeted investment opportunities, depending on zoning. A case study of pizza parlors alone highlighted multiple avenues for entrepreneurism. Some readers may be aware that, as of September 2020, Richmond has suffered some business losses and elevated crime following the novel coronavirus pandemic of 2020 and civil unrest. The research presented here began before such decline occurred. In drawing conclusions about growth opportunities, the author assumes that public safety and business conditions will improve over the long term.

Overall, this analysis demonstrates the potential for a local investor or entrepreneur to utilize data analysis of the market to make well-informed decisions. Future work could include the use of zoning codes and crime data to add richness and additional meaning to the clustering analysis and recommendations.

---

[1] http://trailblazermagazine.net/2019/06/richmond-revolution/

[2] Foursquare.com.

[3] 2017 Virginia Income data. Internal Revenue Service. Accessed May 20, 2020. Available at https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-2017-zip-code-data-soi

[4] Parcel Geographic Summary: City of Richmond Data Portal. Accessed May 18, 2020. Available at https://data.richmondgov.com/browse?category=Unique+and+Inclusive+Neighborhoods