

Applied Data Science Capstone

Exploring Richmond, Virginia – Case Study for a Supermarket

Randy Michaud, June 22, 2020

Introduction:

Richmond, Virginia was incorporated in 1742 and established as a city in 1871. Richmond bears a rich history as one of the largest cities in the commonwealth of Virginia and is the state's capital. The Richmond metro area is one of diversity; a hub of employment opportunities in manufacturing, healthcare, education, and financial sector. The city of Richmond has a vibrant downtown culture with a balance of history, culture, and modernization. This case study will seek to use publicly available data and data science modeling methods to explore:

1. The neighborhood demographics of Richmond, Virginia metro area.
2. What types of businesses populate Richmond, Virginia metro area.
3. Seeking a business opportunity to open a new supermarket;
 - a. Is there a neighborhood lacking a supermarket as a grocery option?
 - b. What neighborhoods/areas of Richmond would most benefit from investment into a supermarket as a grocery option?

Data Sources:

Richmond, VA City portal: <https://data.richmondgov.com/>

Richmond neighborhoods: https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Richmond,_Virginia

Richmond Foursquare data: <https://developer.foursquare.com/places>

Analysis:

The ultimate target variable is supermarket locations within Richmond, Virginia neighborhoods. I started the analysis by importing all necessary libraries to support data analysis and creating a model. Data for neighborhoods was downloaded from Wikipedia for Richmond, Virginia and cross referenced with the Richmond, Virginia data portal. A list of neighborhoods were created in a file, with the first couple lines removed that corresponded to titles in the table.

```
In [8]: rva_df.drop([0,1], axis=0).head(10)
```

Out [8]:

	Neighborhood
2	Arts District, Richmond, Virginia
3	Barton Heights
4	Biotech and MCV District
5	Brandermill, Virginia
6	Carillon, Richmond, Virginia
7	Carver, Richmond, Virginia
8	Carytown, Richmond, Virginia
9	Central Office District
10	Chamberlayne Industrial Center
11	Chestnut Hill-Plateau Historic District

```
In [21]: # convert the venues list into a new DataFrame
venues_df = pd.DataFrame(venues)

# define the column names
venues_df.columns = ['Neighborhood', 'Latitude', 'Longitude', 'VenueName', 'VenueLatitude', 'VenueLongitude', 'VenueCategory']

print(venues_df.shape)
venues_df.head()
```

```
(4772, 7)
```

```
Out[21]:
```

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	List of neighborhoods in Richmond, Virginia	37.5658	-77.44864	SCOR: Sports Center of Richmond	37.563313	-77.455388	Gym
1	List of neighborhoods in Richmond, Virginia	37.5658	-77.44864	Hardywood Park Craft Brewery	37.564810	-77.457135	Brewery
2	List of neighborhoods in Richmond, Virginia	37.5658	-77.44864	Sugar Shack Donuts	37.557084	-77.453144	Donut Shop
3	List of neighborhoods in Richmond, Virginia	37.5658	-77.44864	Food Truck Court At Hardywood	37.564559	-77.457302	Food Court
4	List of neighborhoods in Richmond, Virginia	37.5658	-77.44864	CrossFit RVA	37.567474	-77.463065	Gym

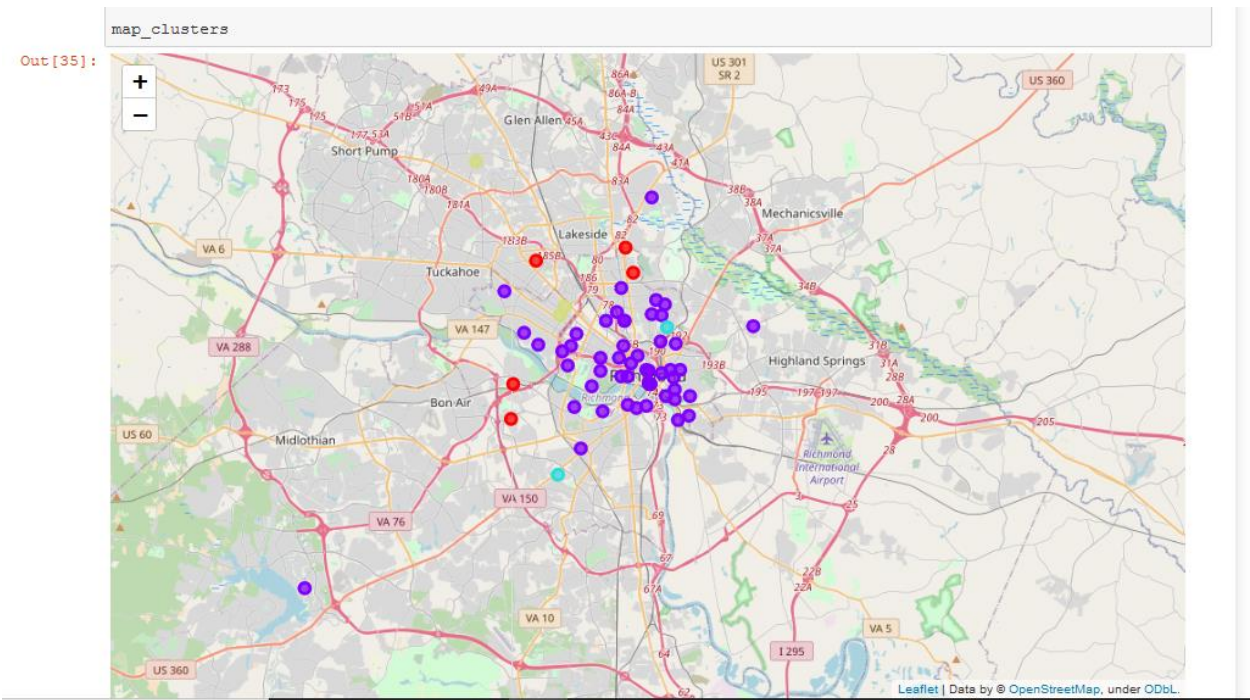
Clusters 0-3 Analysis

```
In [37]: # Cluster 0 Analysis
rva_merged.loc[rva_merged['Cluster Labels'] == 0]
```

Out[37]:

	Neighborhood	Supermarket	Cluster Labels	Latitude	Longitude
60	Washington Park, Richmond	0.023256	0	37.589940	-77.443010
25	Jahnke, Richmond, Virginia	0.021277	0	37.517070	-77.520420
7	Central Office District	0.017544	0	37.534194	-77.519045
40	Pine Camp, Richmond, Virginia	0.021739	0	37.602040	-77.448160
44	Shed Town	0.020000	0	37.595884	-77.504261

Chart above is an example of Cluster 0 Analysis



Richmond, Virginia Cluster Visualization

The Applied Data Science Capstone Project jupyter notebook hosted on GitHub goes into much further detail of the analysis of each cluster and the conclusions drawn in the next section.

Conclusions:

Cluster 2 shows the highest area of supermarkets within the Richmond metro area, specifically the South Garden and Chestnut Hill neighborhoods. Clusters 0 and 3 show moderate population of supermarkets in those areas. Cluster 1 shows the greatest opportunity for opening a supermarket within that cluster to service those neighborhoods – it appears as those neighborhoods are not served by a large chain supermarket. It is interesting to note that the inner-city neighborhoods could be serviced by smaller grocery stores and small markets, leading to the lack of a large supermarket chain in those areas. This does present an opportunity for a supermarket chain to enter Richmond in the Cluster 1 segment, however an assessment on available real estate and proximity to those inner-city neighborhoods would need to be studied for a complete business case. Based on the Richmond neighborhood data and Foursquare data, an opportunity for a supermarket to service Cluster 1 exists.