Deciding the Best County in Virginia to Open a New Supermarket

Shahrouz Sadeghi

3/17/2020

1. Introduction

1.1. Background

With many new supermarkets joining the market, the competition is very cutthroat. Many factors determine whether a new business is successful in this industry or not, like the brands they offer, the quality of the products, the demographic, and the location of the store. There are different tier supermarkets and with different offerings that can each be successful in different locations. Whole Foods will not have much success in a rural area with a low average household income. On the other hand, Since Shoppers competitive advantage is their pricing, they might benefit more in areas with low to mid average household income. Furthermore, determining market in which area is not saturated yet and there is room for more supermarkets in the area is essential.

1.2. Problem

Once it is determined which county is ideal to open a supermarket in, the company needs to delve deeper to pinpoint the exact location. Some factors that this script does not look into are cities within a county, different ethnicities in each county, and consumer loyalty to the current competitors within each county. All these factors have a huge effect on determining the best location. But the most important ones are being investigated by this program: population, income, and number of current competitors.

1.3. Interest

Any supermarket company looking to expand operation into a new territory will surely be interested in this information since this data will pull accurate and up-to-date information from multiple sources and merge them together to segment the market into different clusters.

2. Data Acquisition and Cleaning

2.1. Data Sources

The name of the counties and their population was pulled from the website
https://demographics.coopercenter.org, which had updated information for all the counties. In order to
obtain the average household income for each county, a different website needed to be used.
https://datausa.io/profile/geo/virginia had the most recent information that could be used with minimal
cleaning. The geographical information of each county was obtained using the Geocoder function of
Python. Finally, the number of supermarkets available in each county was obtained from Foursquare
using the category ID for supermarkets in order to limit the results to supermarkets only.

2.2. Data Cleaning and Feature Selection

The original table included county names with population from multiple years and the changes in
population by year. Since we are only interested in the latest population, population from previous years
and changes in population were dropped from the table. Since there are 133 counties in Virginia, I tried to
focus on the top 15 most populated counties. Therefore, after soring the table according to the population
in descending order, only the top 15 were kept.

Second table which included the income needed a lot more preprocessing. Firstly, for each county there
were multiple rows for average household income for different years and different ethnicities. All rows
that pertained to years other than most recent one were dropped. Then, only the total income for each
county was kept instead of income based on race. Since our two tables needed to be merged, the keys had
to match. The format of key in the first table was similar to "Fairfax County" whereas in the second one it
was "Fairfax county, VA". So, not only ", VA" had to be dropped, but "county" had to be capitalized as
well. All this was done using sting slicing and .title() function.

Latitude and Longitude coordinates were easily obtained and added to each row using geocoder. Then,
using these coordinates, number of supermarkets were obtained from Foursquare. By using .groupby()
and .count() functions, number of supermarkets for each county were added to the table as well.

3. Exploratory Data Analysis

By plotting different features for counties, a better picture was obtained of the standing of each county.
Looking at the data, we can see Fairfax county has the highest population followed by Prince William
County and Virginia Beach County. Based Solely on population, we can see that Fairfax county has the
highest potential for opening a new business there.

County Population
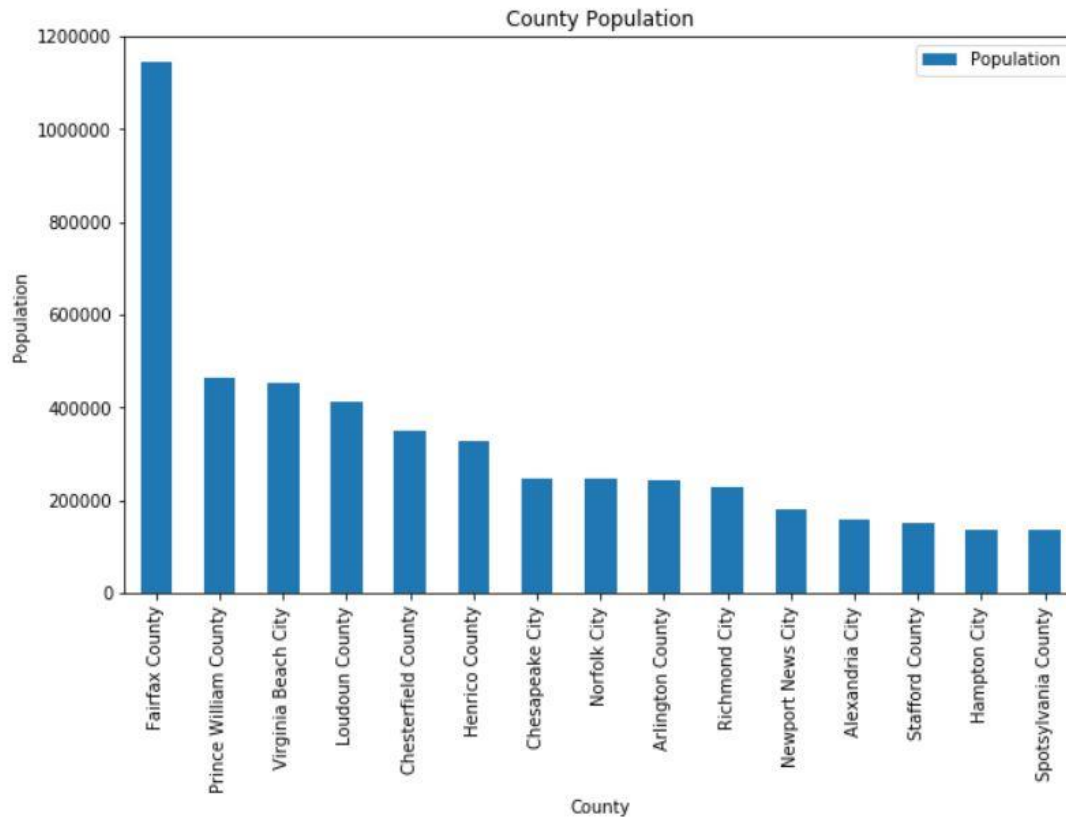
Next, we take a look at the average household income for each county. Although Fairfax County is still one of the highest, Loudon County and Arlington County both have the highest average income. Therefore, it is not logical to base our decision solely on one factor.

Average Household Income of Counties in Virginia
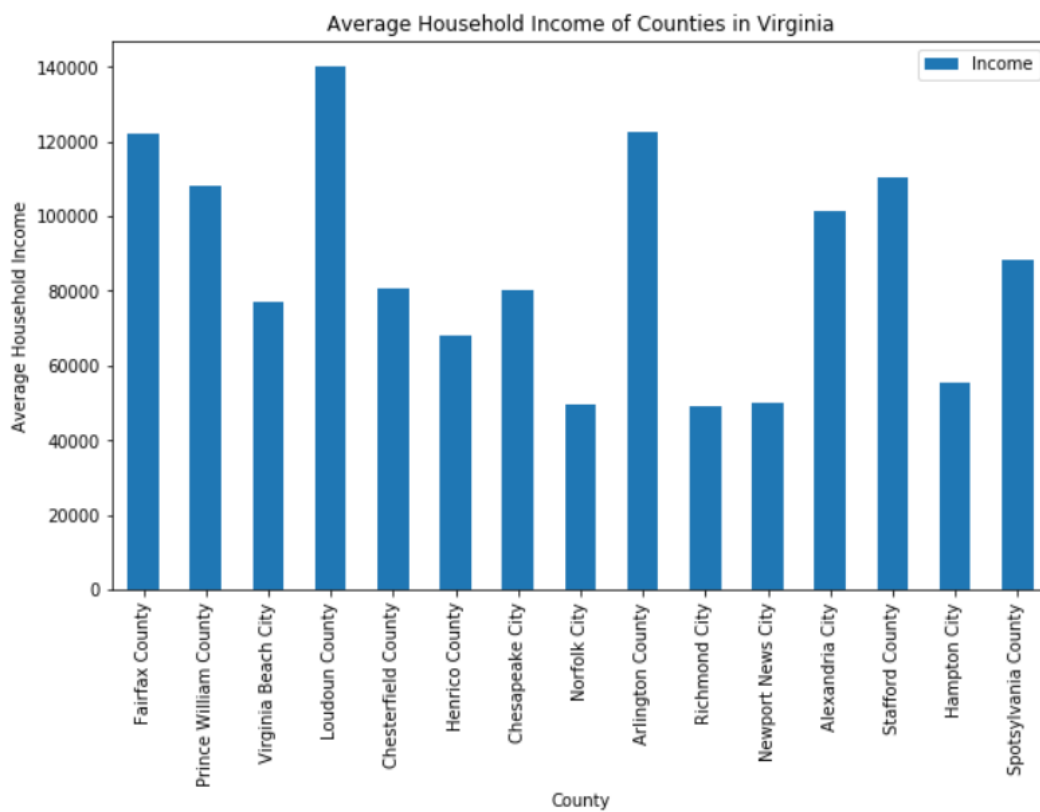
Another feature we can consider is the number of supermarkets. Interestingly, Fairfax county has the highest number of supermarkets, but even though the population is much larger than other counties, the number of supermarkets is in fact only slight larger than other counties. A better way to determine whether there is still a demand in a particular market for a new supermarket is to see how many supermarkets there are for every 10,000 residents.
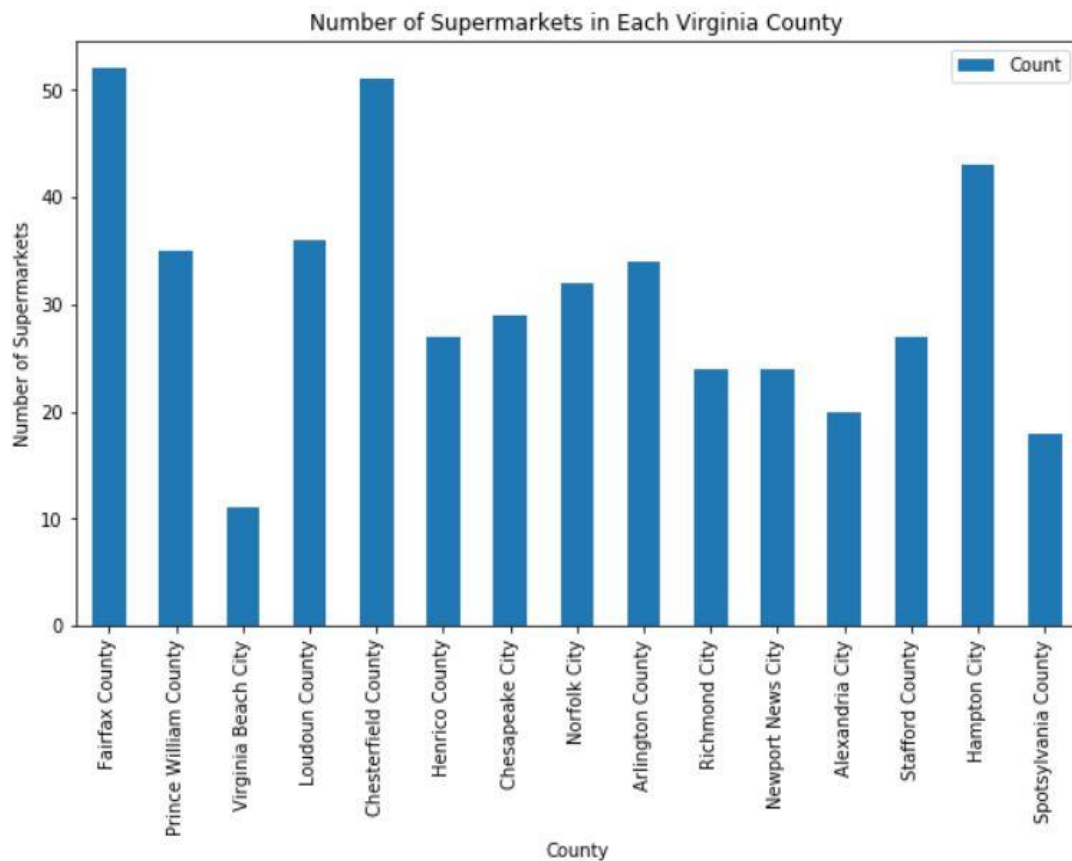
We can obtain this information by dividing the number of existing supermarkets in each county by the population divided by 10,000. By plotting the chart for our new variable, we can see that residents of Hampton City have the highest access to a supermarket with 3.17 supermarkets for every 10,000 residents, whereas Virginia Beach City only has 0.23 supermarkets for every 10,000 residents. By just looking at all the available data, it seems that Virginia Beach City might be the best location for a supermarket that can provide high end and pricy products and a place like Richmond City is ideal for economy supermarkets with average and below average prices.

Number of Supermarkets in Each County Per 10,000 Residents
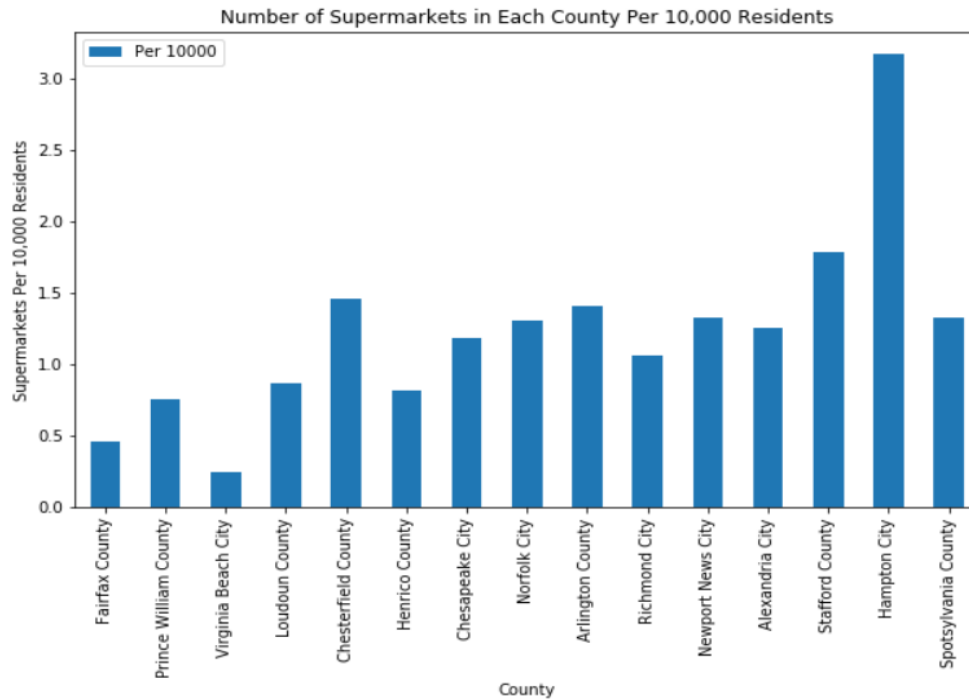
4. Segmentation

The best method to determine which areas are the best location to open a supermarket is to the segment the counties with the same characteristics into different groups and by analyzing the characteristics of each group, we can determine the best location for our new supermarket.

4.1. Clustering

We used K-Means clustering to segment the counties into different groups. After trying different number of clusters, the best segmentation was achieved using 5 clusters. Although the clusters do not have equal members, but they seem to have very unique characteristics.

4.2. Results

By looking at the cluster characteristics, we see the cluster 2 has the highest population with the highest income and the lowest number of supermarkets per 10,000 residents. Therefore, it is the best option to open a new supermarket in. When we look at our table with their cluster number assigned to it, we see that cluster 2 has only one member and that is Fairfax County, one of the most affluential counties in the country considering its proximity to the nation's capital. Given how ideal all of Fairfax County's characteristics are, no other county was even close to Fairfax and that is why it had a cluster of its own. The next cluster that would be a great place to open a supermarket after opening the first branch in Fairfax

County is cluster 0 with approximately 450,000 people residents per county and an average income of just over $108,000 and only 0.62 supermarket per 10,000 residents. This cluster has three members: Prince William County, Loudon County, and Virginia Beach City. Interestingly, geographically, Prince William and Loudon Counties are next to Fairfax County in Northern Virginia and Virginia Beach City is in South of Virginia but it is a tourist attraction.

Cluster Labels	Population	Income	Count	Per 10000
0	4.438957e+05	108299.666667	27.333333	0.621807
1	1.526618e+05	81119.800000	26.400000	1.771282
2	1.143528e+06	122227.000000	52.000000	0.454733
3	2.399480e+05	75308.250000	29.750000	1.237002
4	3.398795e+05	74362.000000	39.000000	1.137328

	Cluster Labels	Population	Income	Count	Per 10000	County
0	2	1143528	122227	52	0.454733	Fairfax County
1	0	465498	107925	35	0.751883	Prince William County
2	0	452643	77059	11	0.243017	Virginia Beach City
3	0	413546	139915	36	0.870520	Loudoun County
4	4	350760	80655	51	1.453986	Chesterfield County
5	4	328999	68069	27	0.820671	Henrico County
6	3	245745	80265	29	1.180085	Chesapeake City
7	3	245054	49587	32	1.305835	Norfolk City
8	3	242152	122394	34	1.404077	Arlington County
9	3	226841	48987	24	1.058010	Richmond City
10	1	181000	50180	24	1.325967	Newport News City
11	1	159152	101215	20	1.256660	Alexandria City
12	1	151689	110307	27	1.779958	Stafford County
13	1	135753	55466	43	3.167517	Hampton City
14	1	135715	88431	18	1.326309	Spotsylvania County

5. Conclusion

By looking at different aspects and features of each county, we were able to determine the ideal location for a new supermarket. The features that were helpful were population, income, and number of supermarkets. I also calculated a new variable that gave us the number of supermarket per 10,000 residents to determine which market has the most need for a new supermarket. Any corporation that is looking to open a new branch of their supermarket in a new area, can use this algorithm to determine the best location. Although this script is only for Virginia, with a little modification to this code, it can be used for any location.