Setting Up Shop in Rockingham County, VA – Where to Start Your Business

Jonathan Yarber

May 9, 2020

1. Introduction

1.1 Background

Approximately 627,000 new businesses open every year [1]. Each new business' success is often in part by determining and setting up shop in the correct location. It can be a challenging process to find the ideal location for your business, but there are universal pointers that we can use to help consolidate the best possibilities. With so many new businesses and existing businesses in play, many cities may very well already be serving the clientele you may be aiming to serve as well. But cities tend to have primary focuses, whether planned or not, on what kinds of venues have attracted city folk. Each city is also judged on certain city metrics that have been gathered for the purposes of this study.

1.2 Problem

Personal endeavors and details of the business' underlying structure aside, and for the purposes of this study, we will be exploring where the most ideal locations would be to establish a new *restaurant* and a new *building contractor firm* within the Rockingham County area in Virginia. Located in the heart of the beautiful Shenandoah Valley, Rockingham County is home to several cities, major landmarks, historical sites, and smaller unincorporated communities. The county was even named one the of best performing small metropolitan areas within the United States, making it an ideal locale to start a new business.

1.3 Interest

If you are looking to start a new business in the State of Virginia and are able or willing to look to the Rockingham County area in the north-western portion of the state, this analysis and breakdown will show you what you need to know about the cities of the county. As stated before, the recommendations in this study will explicitly target restaurants and building contractors, but the information provided could help determine the right city, of the 10 seen here, for any potential new business.

2. Data Collection and Metrics

2.1 Data Sources

Location data is our primary focus for this study, all of which was loosely unstructured in scope. Some of that data coming from the readily available website Data USA [2]. This allows us to view many statistics for just about any given city or town in the US. In this case study, we used Data USA to collect a couple of our metrics as are further described below, but namely: population and average household income.

Real estate tax information as well as crime rates for our 10 cities was found at Neighborhood Scout [3]. The food tax rates were found by scanning the official government city websites for each city [4][5], as well as a news article that cited the tax rates for two neighboring cities within Rockingham County [6]. To examine the most common venues of each city, we reference Foursquare's API [7]. Lastly, to make visual use of Foursquare's data, we collected the latitudes and longitudes of each city, which was found at a website called Lat-Long [8].

2.2 Data Metrics

After reading numerous articles and other sources on what makes a business successful as it relates to location data, I concluded to use the following city metrics for this study: city population, city average household income, city real estate tax, city food tax, and city crime rate.

These metrics were chosen because each one has potential impact on the success of a new business in the city. With regards to city population size – the higher the population, the more potential customers you may have. The higher an average household income, the more chances that customers will pay for higher-cost items or services. The lower the city real estate tax, the less you will have to pay for your office or business building, which equates to some cost savings. Importantly for the new restaurant, the lower the city food tax rate, the less you will be charged on all prepared food and beverage sales. Lastly, the lower the city crime rate, the safer potential customers will feel to come to your business, generating a healthy continuous stream of customers.

Notably, there is an assumption that must be stated for these metrics: for the purposes of this study, we are assuming an *equal impact relationship of metrics*. This means that each metric holds the same amount of impact as any other. Population of a city will be of the same weighted impactful measure as the crime rate of a city. This is primarily done because there were no sources that could identify which of these metrics is the most impactful to business location performance and success. If impact weights were to be found for these metrics, it could be added to the analysis of this data; but for the time being, these metrics will carry the same impact. Below (Table 1), we can see what our Rockingham County cities' data looks like when the unstructured data was brought together.

One last important note in relation to the common venues section of this report: because Foursquare's API is tracked in real-time, the results can vary slightly from one iteration of the source code behind this study. Within a matter of days of forging the source code, I saw differently formed clusters of common venues because of a shift in which venues were becoming more commonplace throughout time. This is likely due to the unique situation of the nation right now in response to the widespread infection of the COVID-19 virus. So, if this same source code would be run on a future day, the common venues and the grouping of common venues may change some. Knowing this, the most common venues will not hold the most impact on our final decision or recommendation but will serve as guidance to be taken into consideration on the current results of the dataset in this report in our decision-making process.

Table 1. Consolidated city metrics for comparison.

City	Latitude	Longitude	Population	Avg_Household_Income	Real_Estate_Tax	Food_Tax	Crime_Rate
Harrisonburg	38.4496	-78.8689	53064	$43,009.00	$0.86	7.00%	19.12
Elkton	38.4079	-78.6236	2795	$52,808.00	$0.11	7.00%	10.76
Broadway	38.6132	-78.7989	3793	$44,852.00	$0.07	4.00%	2.54
Dayton	38.4138	-78.9389	1622	$52,664.00	$0.08	5.00%	5.54
Massanutten	38.4096	-78.7378	2273	$76,473.00	$0.74	4.00%	7.79
Bridgewater	38.3821	-78.9767	5930	$58,625.00	$0.75	6.00%	4.57
McGaheysville	38.3711	-78.7326	2147	$55,271.00	$0.74	4.00%	7.79
Timberville	38.6390	-78.7739	2592	$47,132.00	$0.11	4.00%	9.02
Mount Crawford	38.3571	-78.9409	403	$68,125.00	$0.07	4.00%	20.30
Linville	38.5204	-78.8375	938	$40,530.00	$0.74	4.00%	14.10

Note: food tax rate for 6 of the 10 cities was not available online, so the average food tax rate of Rockingham County was used for those cities. Similarly, for real estate tax rates, the cities of Massanutten, McGaheysville, and Linville were given the Rockingham County average. These instances of lacking data are likely due to some of these locations being unincorporated communities instead of official cities.

3. Data Analysis

3.1 Exploring Common Venues of Cities

One set of key information that we can use when determining what city to start a new business in is the currently existing venues in each city. This can help us in a couple of theoretical ways: 1) we will see if the most common venues are in alignment with the business you are setting

up, and if so, then 2) we can expect to have greater competition within that city. A new business that does not share itself as a most common venue in a city can equally be both a boon and bane. On the one hand, a different sort of business than what's common means you will not have as much competition. This also means that the city may not need your business' services as much, which may result in fewer customers.

Using Foursquare's API, we can extract the most common venues in each of our ten cities within Rockingham County. I choose to pull the top three most common venues which can be seen below (Table 2). As we can see, food service venues and restaurants are among the most common of venues, comprising 70% of all the most common venues. Each city has a common venue that is food service or a restaurant, but some cities have more than others. We notice here that the city with the highest population, Harrisonburg, and the city with the lowest population, Mount Crawford, have more common venues for food services and restaurants than the other cities; as all of their top three most common venues are all food and beverage focused. Keeping our building contracting firm in mind, the only common venue that we see in the table (Table 2) that compliments building contracts closely would be Linville, where its most common venue is Home Service. If home services are commonplace in Linville, it could be that Linville is seeing a growth in its population or renovating old buildings. Either potential reason could prove useful to have a building contractor in the neighborhood.

Table 2. Three most common venues in each city.

City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Bridgewater	Fast Food Restaurant	Convenience Store	Ice Cream Shop
Broadway	Farmers Market	Café	Wine Shop
Dayton	American Restaurant	Art Gallery	Café
Elkton	Gift Shop	Pharmacy	Food
Harrisonburg	Bar	American Restaurant	Pizza Place
Linville	Home Service	Wine Shop	Café
Massanutten	American Restaurant	Art Gallery	Convenience Store
McGaheysville	Gun Range	Discount Store	Sandwich Place
Mount Crawford	Bakery	Wine Shop	Café
Timberville	American Restaurant	Pizza Place	Gas Station

3.2 Ranking Metrics of Cities

Looking at just the venues of cities is not enough to make a qualified recommendation or conclusion on where to set up your new business, however. The qualities of the city will arguably impact your business success more than the venues of the city. To examine the qualities of the cities, we will use the metrics that were collected as comparison and contrast to see which cities rank the best in these metrics. As mentioned before, a couple of these metrics, namely Population and Average Household Income, we want to be higher value, while we desire lower values for the remaining metrics since they deal with costs and crime rates.

From the information seen in Table 1, I converted the values for each metric of each city to an integer value based on highest to lowest rank, which can be seen below (Table 3). The higher the number rank is, the better that metric proved to be in that city. We now have a quick glance view of how each city fairs in its metrics against each other. We quickly notice, for example, that while the city of Harrisonburg ranks the best in population, as it exceeds the next most populated city by 47,000, it ranks the among the worst in every other metric. It holds the highest real estate tax, food tax, and crime rate, and holds the second lowest average household income. By comparison, Mount Crawford has the smallest population and the highest crime rate, but is among the best in average household income, real estate tax, and food tax. Since our food tax rate for several cities was based on the average in Rockingham County, due to the lack of available data, we see those cities tied for best food tax at rank "7."

Table 3. Ranked metrics in each city (the higher the number rank, the better).

City	Latitude	Longitude	Population	Avg_Household_Income	Real_Estate_Tax	Food_Tax	Crime_Rate
Harrisonburg	38.4496	-78.8689	10	2	0	1	1
Elkton	38.4079	-78.6236	7	6	6	1	3
Broadway	38.6132	-78.7989	8	3	9	7	9
Dayton	38.4138	-78.9389	3	5	7	3	7
Massanutten	38.4096	-78.7378	5	10	3	7	6
Bridgewater	38.3821	-78.9767	9	8	1	2	8
McGaheysville	38.3711	-78.7326	4	7	3	7	6
Timberville	38.6390	-78.7739	6	4	6	7	4
Mount Crawford	38.3571	-78.9409	1	9	9	7	0
Linville	38.5204	-78.8375	2	1	3	7	2

These ranked metrics will be summed together and then sorted so that we can easily see from the top to the bottom which cities held the best-ranking metrics. There is a couple of other pieces of information we want to see before we do that, however.

3.3 Correlation of Ranked Metrics

With these ranked city metrics, let us examine if there is a correlation between them. This would allow us to see if there was a relationship between two or more metrics. For example, if a higher population count would equate to a higher or lower average household income. We create a simple correlation matrix of our city metrics after having ranked them. Correlations can only be made for attributes with the same value type, in this case, as the ranked integers. As it stands, there are no positive nor negative correlations between these city metrics that we have for this study. A strong positive correlation would be denoted with a correlation value of 1, while a strong negative correlation would have a value of -1.

Figure 1. Correlation matrix of ranked city metrics.

4. Clustering

4.1 K-Means Clustering of Venues

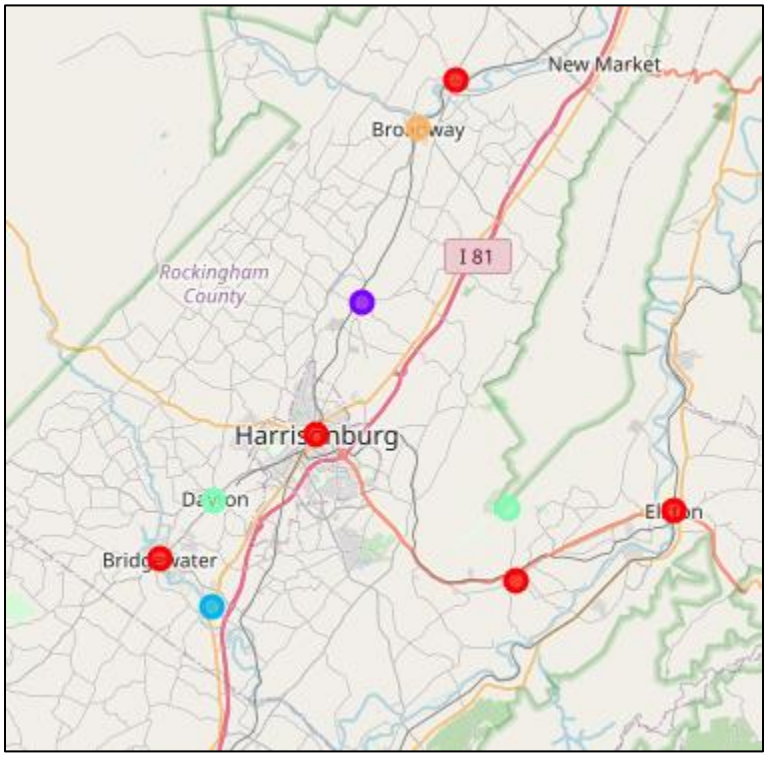
Next, we want to cluster cities together based on the venues and ranked metrics we have looked at. First, let us see what the clusters become when comparing alike common venues.

We will use K-means clustering, which is a type of unsupervised machine learning that is capable of clustering data into a multitude of clusters even with a multitude of comparisons, such as our top three most common venues and ranked city metrics. K-means will separate

the cities into several clusters equal to *k*, where we initialize *k*. The clusters are formed by calculating the sum of squared distance of the samples (in this case, values respective of the venues) to the nearest center of a cluster, which is called the cluster centroid. For clustering the cities based on their most common venues, we initialized *k* as 5, so we will have five separate clusters.

In the below map (Map 1), we see our clusters represented by the colored circles over-top each city. The same color circle indicates that the cities are in the same cluster. We see that one of our clusters contains five of the ten cities, with these cities tending to have a variety of different types of venues across the board, but repeating themes between the cities. Whereas other clusters limit themselves to having one or two cities in a single cluster. The higher our number of clusters *k*, the more clusters that would likely contain a fewer number of cities. This helps us segment our cities' most common venues into groups, which could help in our decision-making process of where to set up a new business. In tandem with the ranked metric data, we could, for example, establish a new business in a cluster of complimenting venues and then choose the city with the best ranked metrics from that cluster.

Map 1. Clustering of Rockingham County's cities by common venues.



To show an example of our clustered data by common venues, the below table (Table 4) shows that Dayton and Massanutten were placed in the same cluster because they had the same 1st and 2nd most common venues – American Restaurant and Art Gallery, respectively.

Table 4. Clustering: Dayton and Massanutten with similar common venues.

City	Real_Estate_Tax	Food_Tax	Crime_Rate	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Dayton	$0.08	5.00%	5.54	3	American Restaurant	Art Gallery	Café
Massanutten	$0.74	4.00%	7.79	3	American Restaurant	Art Gallery	Convenience Store

4.2 Elbow Method

Identifying the appropriate number of clusters is a unique challenge presented in K-means clustering. For clustering the cities by most common venues, we initialized *k* to be 5. Before examining the clusters of cities by the ranked metrics, let us use a methodology of determining how many clusters we should have with this dataset. One means of determining the appropriate number of *k* clusters is through what is called the elbow method. After scaling our data so that all ranked metrics are of equal importance, we initialize k-means and use the *inertia* attribute to identify the sum of squared distances of samples to the nearest cluster centroid. As *k* increases, the sum of squared distances grows closer to zero. If we initialized *k* as its maximum possible value, which would be the number of cities in this case, then each city would be in its own cluster and would therefore have a sum of squared distance of zero.

This elbow method will plot a graph (Figure 2) that is vaguely in the form of an elbow. The sharpest point of the elbow is how many *k* clusters is the most optimal for the dataset. While a bit harder to discern from this dataset, our optimal number of *k* clusters is seven. So, we will have seven clusters of cities by their similarly ranked metrics.

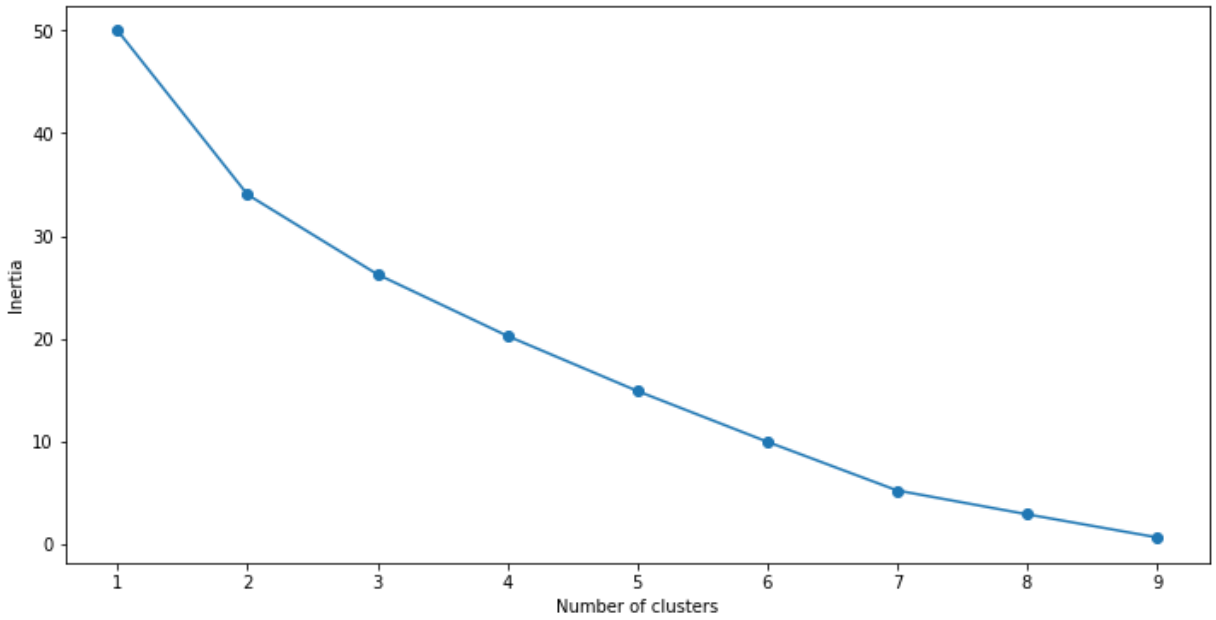


Figure 2. Elbow graph to show ideal number of clusters in our K-means clustering (7).

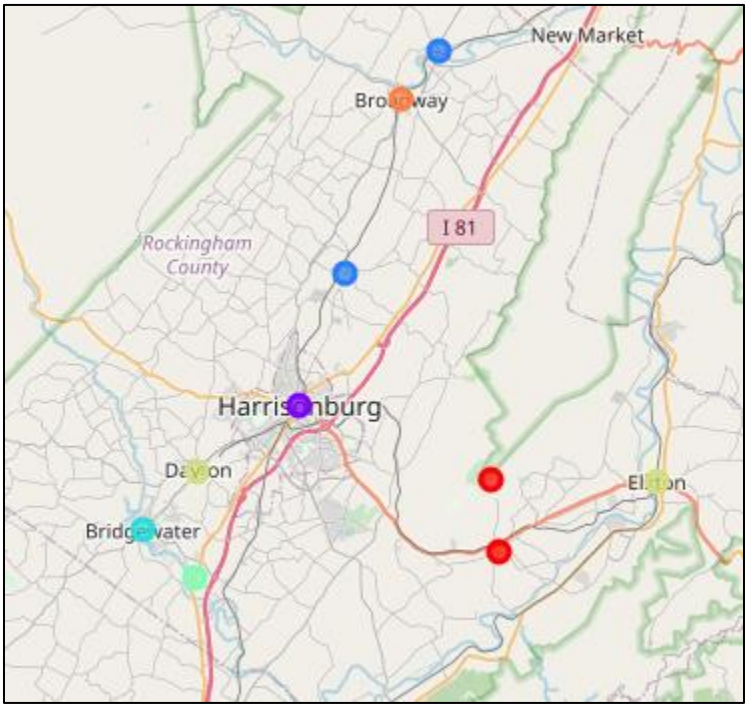
4.3 K-Means Clustering of Ranked Metrics

Now that we know the optimal number of *k* clusters for clustering cities by their similarly ranked metrics, we can create a map (Map 2) showing the same cities within their new clusters.

Interestingly, none of the cities share the same clusters for both common venues and ranked metrics, and unsurprisingly, we have more independent cities in individual clusters when by clustered by ranked metrics. While it will not be an exact correlation, we will still be able to relay a city with a higher rank amongst the clusters of common venues.

There are three clusters that contain two cities each. Massanutten and McGaheysville are in the same cluster due to having the same tied ranks in real estate tax, food tax, and crime rate. Timberville and Linville likewise share a cluster with equal ranks in food tax, and close ranks in crime rate. Dayton and Elkton, in the same cluster, do not have any exact matching ranked metrics, but are similar in average household income, real estate tax, and food tax.

Map 2. Clustering of Rockingham County's cities by similarly ranked metrics.



5. Conclusions and Recommendations

5.1 Findings

Taking all this information together, we can see a sorted summation of the cities by the highest-ranking metrics, as well their cluster labels (Table 5). Before the data was assembled and

before any assumptions about the metrics were made, I was expecting to see Harrisonburg dominating the list. Since our metrics all have the same weighted impact, the only metric Harrisonburg ranked high on was its over-exceeding population when compared to the other cities. In the end, with these metric assumptions, Harrisonburg falls to the bottom of the list of cities. Broadway comes in at the top with a summed metric ranked score of 36 out of the possible 50.

Venues is not shown in this table, but we can use the tables before and our knowledge of the clusters of similar common venues to make some more finite conclusions and recommendations on where is the most ideal city to establish a new restaurant and building contractor firm.

Table 5. Sorted summation of highest ranked city's metrics.

City	Latitude	Longitude	Population	Avg_Household_Income	Real_Estate_Tax	Food_Tax	Crime_Rate	Cluster Labels	Sum
Broadway	38.6132	-78.7989	8	3	9	7	9	6	36
Massanutten	38.4096	-78.7378	5	10	3	7	6	0	31
Bridgewater	38.3821	-78.9767	9	8	1	2	8	3	28
McGaheysville	38.3711	-78.7326	4	7	3	7	6	0	27
Timberville	38.6390	-78.7739	6	4	6	7	4	2	27
Mount Crawford	38.3571	-78.9409	1	9	9	7	0	4	26
Dayton	38.4138	-78.9389	3	5	7	3	7	5	25
Elkton	38.4079	-78.6236	7	6	6	1	3	5	23
Linville	38.5204	-78.8375	2	1	3	7	2	2	15
Harrisonburg	38.4496	-78.8689	10	2	0	1	1	1	14

5.2 New Restaurant

For the most ideal city to start a new restaurant, let us look at some of the highest-ranking city metric clusters that have complimentary venues that a restaurant could comfortably coincide with without there being too much competition.

Massanutten and McGaheysville were in the same cluster for similarly ranked metrics, with Massanutten having the higher overall ranks. Massanutten also shared a cluster with Dayton on similar common venues, which had American Restaurant as the most common venue. The second and third most common venues in Massanutten, Art Gallery and Convenience Store, do not have as much of an impact to restaurant success. If your new restaurant were an American-style restaurant, it would certainly fit in well in Massanutten, but your level of competition will be higher.

In the end, I would recommend Broadway as the most ideal city to set up a new restaurant. This for several reasons. Firstly, it has the lowest real estate tax rate, tied for the lowest food tax rate, and the lowest crime rate, so you will be paying less to maintain your office and services while the citizens of the city feel safe to come to your restaurant. Secondly, the most

common venues of Broadway are all complimentary for restaurants. Competition? Not necessarily. Especially if your restaurant caters to the cafes and wine shops that are in the city. And with delicious product at the farmer's market, which is the most common venue in Broadway, customers of yours may even try to recreate your recipes and keep coming back to taste more as a result. Lastly, Broadway does boast the highest overall ranked city metrics out of all ten cities.

5.3 New Building Contractor

Similarly, we can evaluate the ideal location for a new building contractor as we did for the restaurant.

Originally, I was thinking that Linville would be the most ideal location for a building contractor due to the most common venue there being Home Service, which a building contractor could pair well with. Unfortunately, Linville ranks second-to-last in overall ranked city metrics, so the conditions are not as ideal.

Looking at the top ranked cities for metrics, McGaheysville has the most diverse sort of most common venues, which could yield to a higher need for multiple building contractors to help construct all these different buildings that provide completely different services. I would recommend McGaheysville for this reason, as the other cities equal to or higher in overall city metrics have common venues that are remarkably similar, so there isn't as much diversity to use as a foundation for building projects.

It would be interesting to see what the growth rates were for these different types of venues in the cities, as that could potentially change where the most ideal city for building contractor would be, but that is out of scope of this study.

6. References

[1] https://smallbusiness.chron.com/information-small-business-startups-2491.html

[2] https://datausa.io/

[3] https://www.neighborhoodscout.com/

[4] https://www.harrisonburgva.gov/business-taxes

[5] http://www.elktonva.gov/treasurer.html

[6] https://www.dnronline.com/news/harrisonburg/the-food-tax-is-high-enough/article_3fd1758e-289d-11e7-8164-db5f0349f9d5.html

[7] https://foursquare.com/

[8] http://www.lat-long.com/