Another Restaurant in Greenville, SC ?

Introduction

Deciding the optimum geographical location to open a business depends on many factors, including not only the cost of the real estate involved, but also less easily quantifiable factors such as neighborhood type and demographics, transportation availability, the presence (or absence) of similar establishments (and the associated ratings of those businesses), and even physical security needs. A sole proprietor or silent investor preparing a business plan for a new restaurant ignores these softer factors at his or her peril.

The purpose of this project will be to combine multiple freely available databases to recommend alternatives for a new restaurant in the Greater Greenville, SC area.

Target Audience

Stakeholders would include interested investors- both as the main proprietors or less active backers simply performing due diligence. These could be silent partners, bank loan officers, or small business administration officials. The analysis is not meant to be financial in nature, but rather as an evaluation as to the more subjective "goodness of fit"
for a restaurant in various geographic areas.

Background and Business Problem

The US census Bureau has estimated South Carolina the sixth most rapidly growing state in 2020 (1). Within this state, Greenville and the surrounding counties (The Greenville-Anderson-Mauldin Metropolatin Statisical Area) (2) is the fastest growing, with a population growth rate in Greenville county of greater than 16% from 2010 to 2019 alone (3). Having risen from the ashes of the textile industry collapse (4), the region boasts many large and light industries, including the largest BMW automotive plant in the world (5), the Michelin North American headquarters (6), the largest gas turbine plant in the world (7), an inland port that along with the coast helps the state handle $75B of goods annually (8), a large branch of a fortune 500 engineering and construction services firm (9), a large University (10), and a 300,000 square foot corporate banking campus (11). A massive amount of housing and support industries have grown up to support this economic and population growth. New neighborhoods are built from the ground up annually, along with the retail businesses required to support them.

Opening a restaurant in this geographic area might seem like a slam dunk- 16% more people is roughly 100,000 people based on current county population (there are several counties growing at or near this rate). These people all need to live somewhere, and they all need to eat. Most of them drive to work and home. The labor market is relatively tight, and the weather even allows out of doors dining more months than not during the year. However it is always a good idea to research location when opening any business. It would make little sense to open a five star restaurant in a lower middle class housing development surrounded by "meat and three" diners and fast food establishments (unless perhaps there was some other more appropriate neighborhood nearby. Having lived in this location for several years, the author has no doubt that many great business ideas have failed simply due to the lack of due diligence in location selection. Others have clearly thrived as new clusters have literally sprung up in once deserted local areas.

Data (Extract Transform and Load)

This project combines foursquare data along with other freely available research data in order to examine, compare, and cluster real estate, transportation, and crime statistics for the Greenville county and Anderson County. These data sources and the data available from them are summarized in table 1.

Table 1: Data Sources Used

Data Category	Access Type	Used to Study	Website
Location	html	postal codes	https://www.zip-codes.com/state/sc.asp
Business	API	Business type, location, rating	https://developer.foursquare.com/
Real Estate	web scrape	Median housing value	www.bestplaces.com
Crime	web scrape	Property crime score	www.bestplaces.com
Crime	web scrape	walkability, public transit, bike accessibility	www.bestplaces.com
Crime	web scrape	crime and other statistics	www.walkscore.com

www.bestplaces.com

A listing of all the zip codes from SC was used to create a subset of only those in Greenville and Anderson County (11). This area represents most of what is known as the Greenville-Anderson_Mauldin statistical area (12). The population of this area was just around 725,000 in 2019 (13).

Use of the Zillow and crime score API proved difficult in practice. The business model for both websites is based on web traffic, and a developer must provide a webpage under development. Lacking this item (and worried it would take too long to create one) a simpler web scraping approach was utilized (Appendix A). Mac Automator was used to save webpages from best places for each zip code (where available). These HTML files were combined (again using automator) and then searched using Python (see Jupyter notebook). A list of zip codes was created with the violent crime score, property crime score, and median home price for each zip code. Data was identified for 32 zip codes (of the original 48). This is not entirely surprising as these counties contain some sparsely populated regions. Similarly, a walkscore was scraped for 22 zip codes. It was however unclear how these scores were centered. This score was excluded from further analysis until better data could be obtained.

Latitude and longitude data was obtained for the entire US, and joined with the 48 zip codes identified above (14). Crime and property data was joined with zip code and location data. At this point 33 neighborhoods had been identified to use with Foursquare (Figure 1)

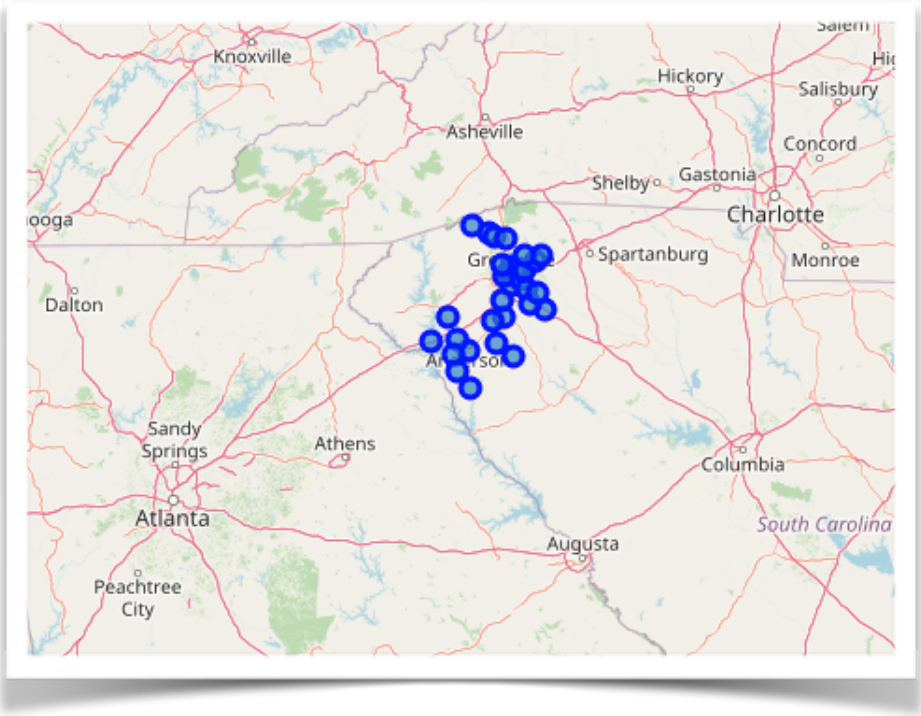


Figure 1: Zip code locations.

Data Analysis

From the resulting 33 postal codes, Foursquare was used to identify top venues in the Greenville and Anderson area. These are presented in Appendix B and Table 2. An examination of the data indicates that Simpsonville has a lower amount of restaurants relative to other areas.

Table 2

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Anderson	Fast Food Restaurant	American Restaurant	Pizza Place	Discount Store	Shoe Store	Bank	Fried Chicken Joint	Sandwich Place	Deli / Bodega	Department Store
1	Belton	Fast Food Restaurant	Video Store	Shopping Mall	Fried Chicken Joint	Mobile Phone Shop	Discount Store	Pharmacy	Bar	Deli / Bodega	Gym / Fitness Center
2	Cleveland	Campground	Scenic Lookout	Video Store	Cosmetics Shop	Deli / Bodega	Department Store	Discount Store	Donut Shop	Ethiopian Restaurant	Fast Food Restaurant
3	Fountain Inn	Fried Chicken Joint	Fast Food Restaurant	Video Store	Food & Drink Shop	Cuban Restaurant	Deli / Bodega	Department Store	Discount Store	Donut Shop	Ethiopian Restaurant
4	Greenville	American Restaurant	Hotel	Steakhouse	Pizza Place	Restaurant	Trail	Art Gallery	Bar	Clothing Store	Performing Arts Venue

Next, k means clustering was applied to the venues. It is immediately obvious that restaurants are centered in Greenville and Anderson (cities with the same name as the containing county). Downtown Greenville is well known within the region for a wide range of dining establishments ranging from fine dining to lighter fare (15). Simpsonville has fewer red clusters (top venue restaurant). The blue clusters are categorized by groceries, retail shopping, and some restaurants.

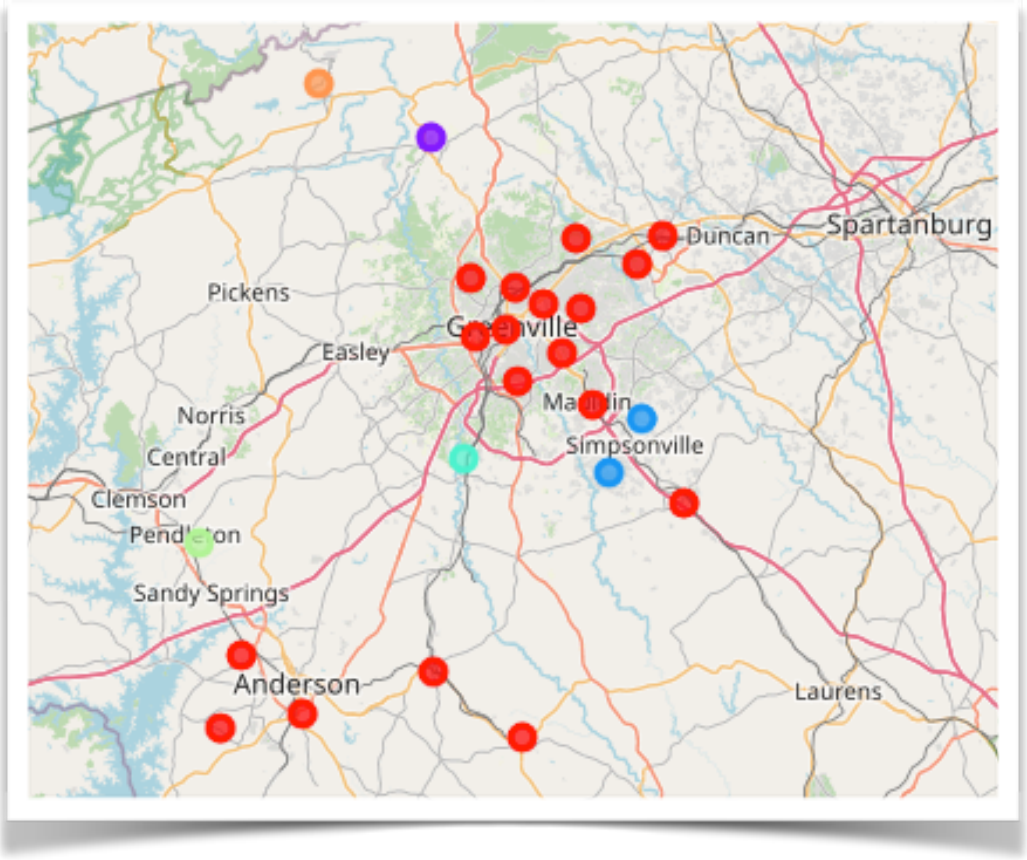


Figure 2: K Means Cluster Analysis of Venue type Greenville Anderson Counties

The next step was to overlay the crime data with the cluster data. Best places conveniently provides the US average scores for property (35.4) and violent (22.7) crime. Clusters with a score higher than average were colored red (with the outline kept the color of the cluster) and scores below average green (Figure 3 and Figure 4). Unfortunately, crime appears higher than the national average for most of these areas (in particular the metropolitan areas. However the Simpsonville area scores well on both metrics.

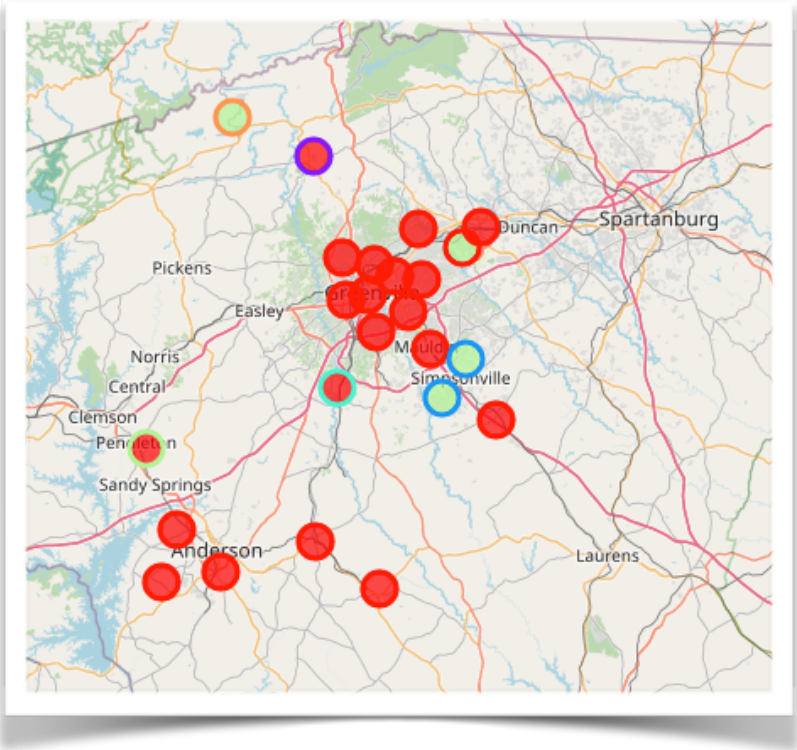


Figure 2: Violent Crime Overlay (green is below average)

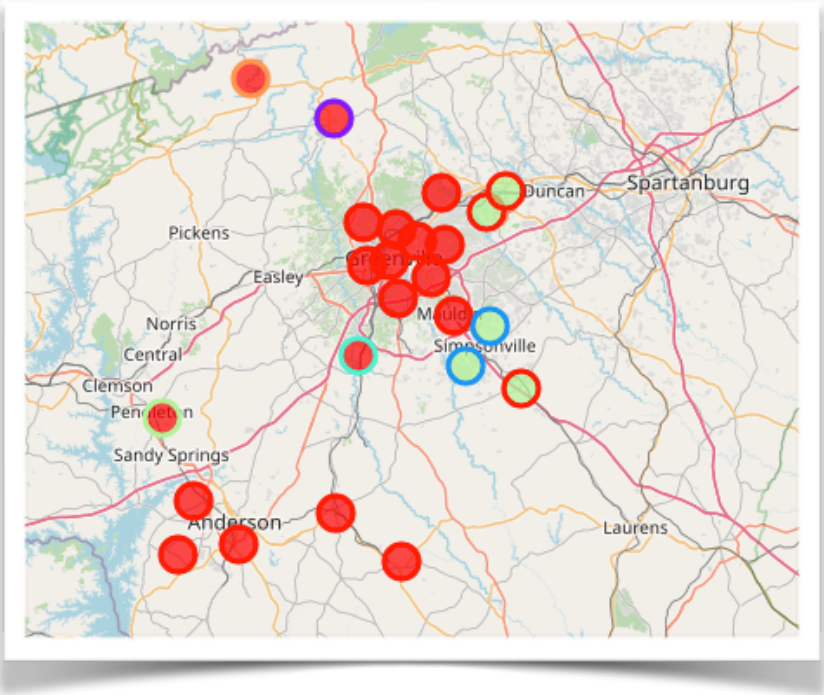


Figure 3: Property Crime Overlay (green is below average)

Lastly the property values were overlaid on the clusters (Figure 4). This time the color was made red for below the local average (all data) and green for above. Higher property values are found on the eastern part of the Greenville area, and in the Simpsonville/Mauldin area.

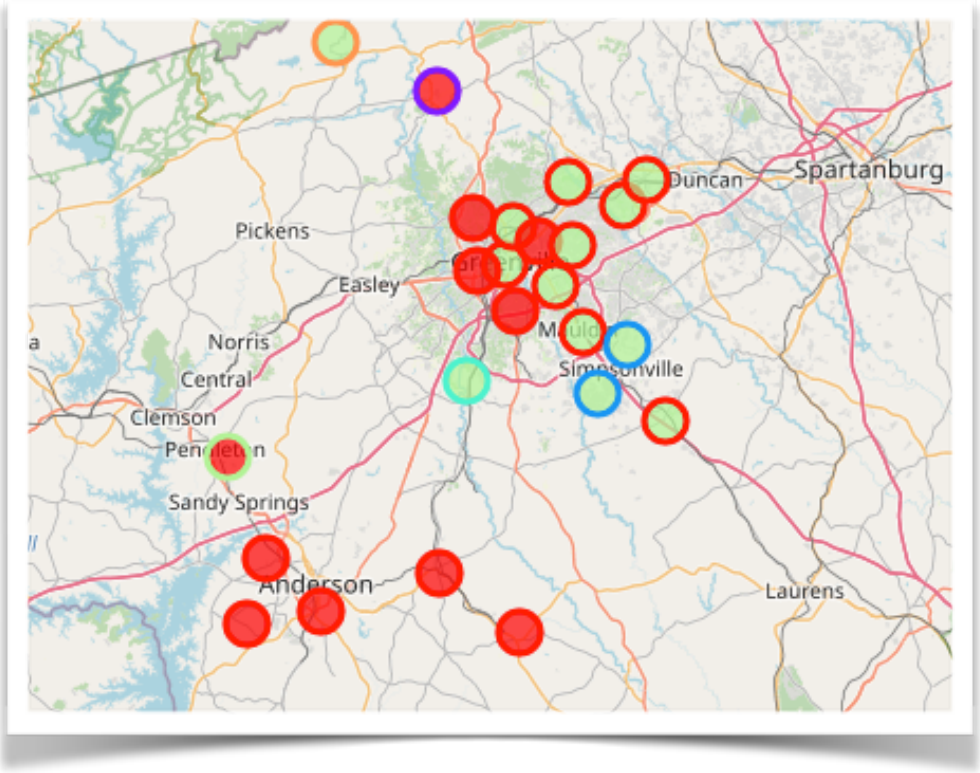


Figure 4: Property value (green is above average)

The crime statistics were examined again using scatterplots. Violent crime appears to decrease with increasing property values (Figure 5). High and low crime areas are identified below. Simpsonville is the lowest on the chart (both zip codes). The highest values are in Anderson and Greenville zip codes.

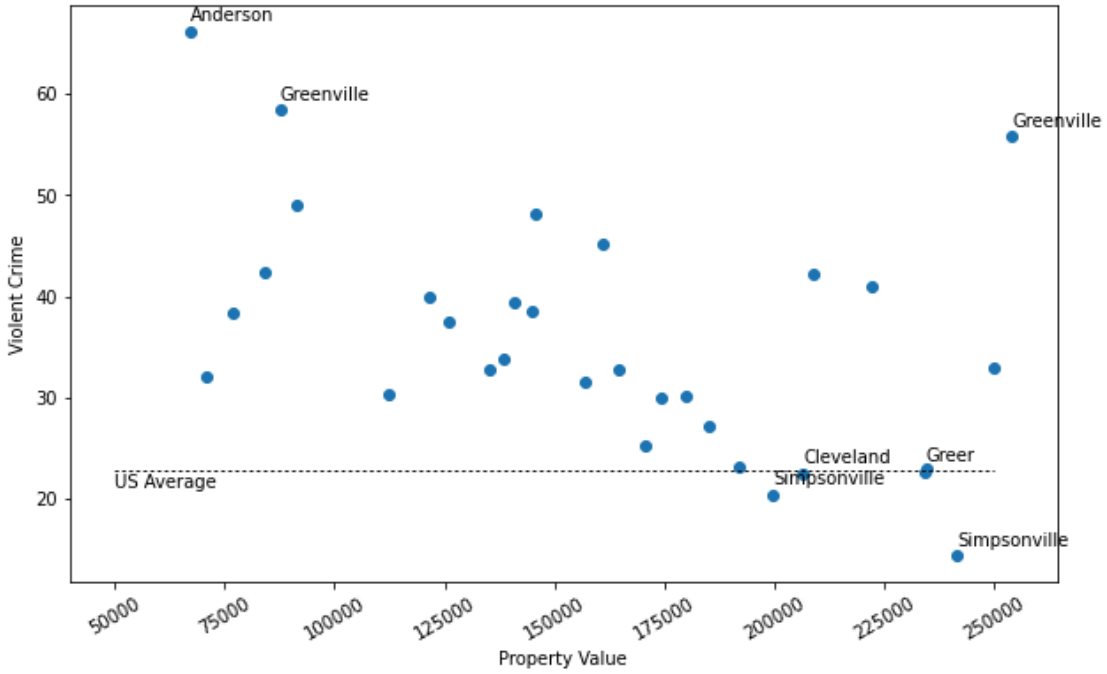


Figure 5: Violent Crime Versus Property Value

Examining property crime, the chart looks similar (Figure 6). Simpsonville again scores at the bottom (top) of this metric. It should be noted that it is also at the top of the property value scale. For full disclosure the author was relieved to learn he lives in one of the locations at or below the average crime lines.

These two types of crime are highly correlated. Again Simpsonville is identified as a lower crime alternative. Anderson and Greenville (both cities) contain some very high crime areas. Not exactly groundbreaking, but important information to consider when placing a business and attempting to attract customers.

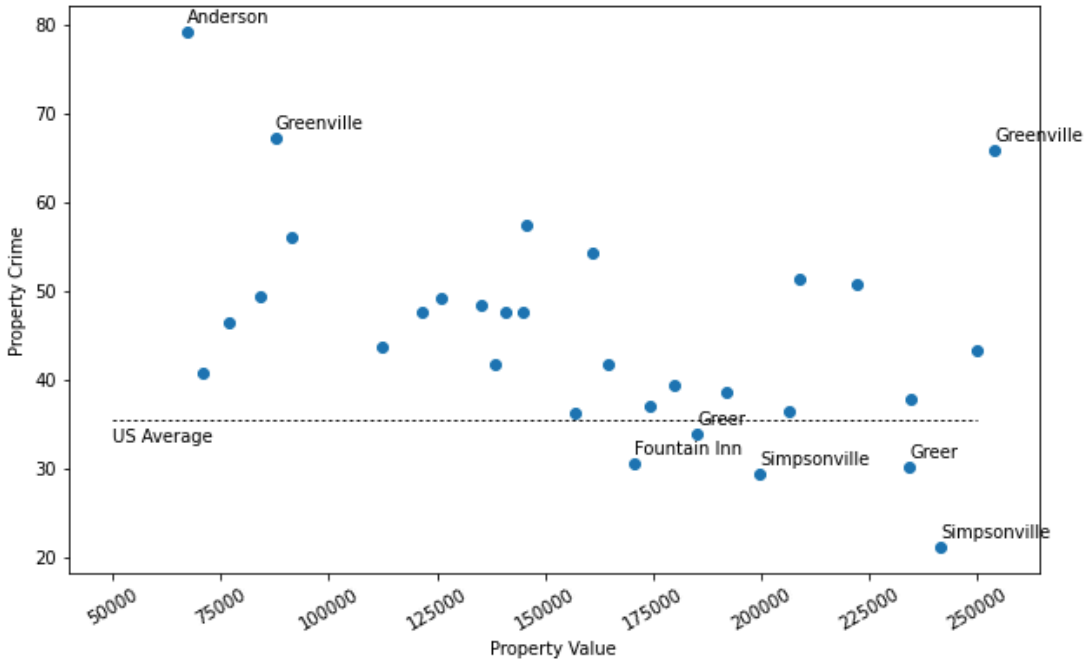


Figure 6: Property Crime Versus Property Value

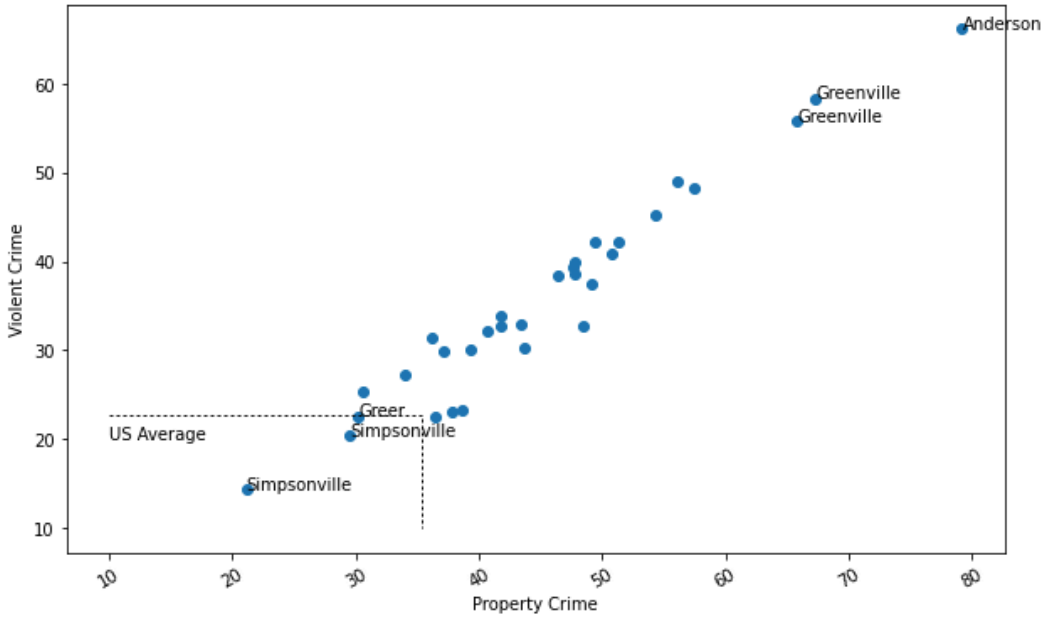


Figure 7: Crime Correlation

Lastly walk scores were accessed directly for Simpsonville, Greenville, and Anderson, and the results presented in Table 3. Simpsonville is the least walkable or bike able of the three locations. Greenville scores the best on these metrics.

Table 3: Transit Scores

	Walk Score	Bike Score
Simpsonville	22	26
Greenville	43	41
Anderson	33	34

Data(Summary and Conclusions)
A vast array of data is aggregated on the best places website (and on census and local government websites). This data can be easily scraped for use in analysis. Large scale scraping in the real would require checking the website for a robots.txt file or other restrictions and user agreements.

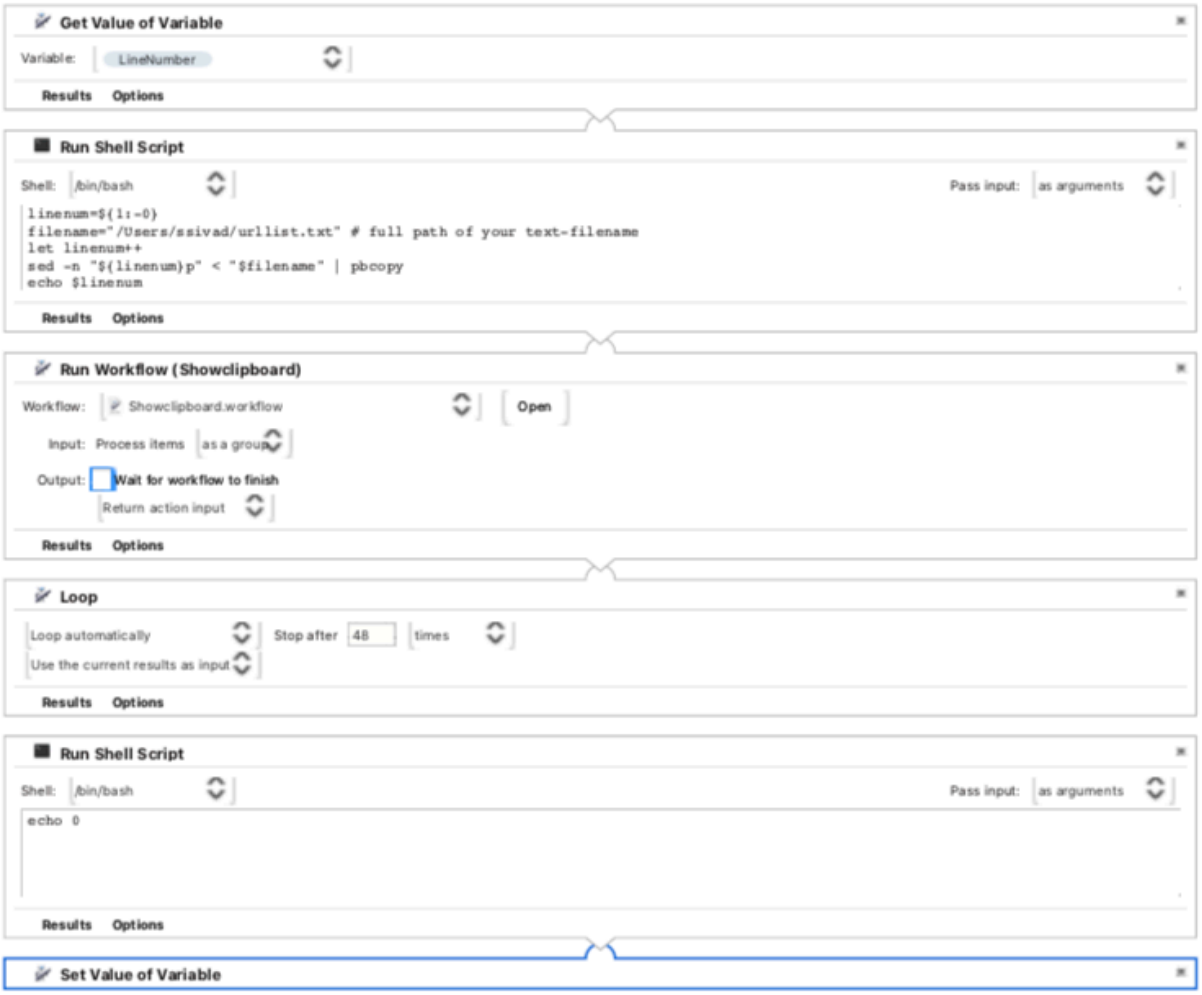
Discussion
Based on the above analysis, a location within Simpsonville is recommended for a new restaurant. The analysis indicates that this area is not saturated with restaurants, but does have other shopping available. The area is low on crime and high on property value. In addition, zip codes in Greenville and Anderson can probably be ruled out due to extremely high crime scores (property and violence).

Conclusions and Recommendations
Foursquare and publicly available real estate, transportation, and crime data was used together to reccomend a location (Simpsonville) for a new restaurant. Data was obtained from the internet (scraped automatically using Automator), and analyzed via Python (Jupyter notebook). It is recommended to do do further business analysis more local to the recommended location. In the end there is no substitute for actually visiting a geographic location. This approach however greatly narrowed down the options.

References

1. https://www.census.gov/programs-surveys/popest/technical-documentation/research/evaluation-estimates.html
2. https://www.bls.gov/oes/current/oes_24860.htm
3. https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-total.html
4. https://independenttribune.com/news/decline-of-an-industry-what-happened-to-american-textiles/article_65325fc0-b20e-552c-a1bf-80b52e94ba06.html
5. https://www.bmwgroup-werke.com/spartanburg/en.html/
6. https://michelinmedia.com/about/
7. https://www.ge.com/news/reports/turbines-born-inside-look-ges-big-iron-maternity-ward
8. http://scspa.com/about/statistics/cargo-value/
9. https://www.fluor.com/about-fluor/locations
10. https://www.clemson.edu
11. https://www.downloadexcelfiles.com/us_en/download-list-zip-codes-south-carolina-state
12. https://datausa.io/profile/geo/greenville-anderson-mauldin-sc
13. https://www.census.gov/quickfacts/fact/table/andersoncountysouthcarolina,greenvillecountysouthcarolina/PST045219
14. https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/?q=SC
15. https://www.fallforgreenville.net/

APPENDIX A


Get Value of Variable
Variable: LineNumber
Results Options
Run Shell Script
Shell: /bin/bash
Pass input: as arguments
linenum=$(($1 - 0))
filename="/Users/ssivad/urllist.txt" # full path of your text-filename
let linenum++
sed -n "${linenum}p" < "$filename" | pbcopy
echo $linenum
Results Options
Run Workflow (Showclipboard)
Workflow: Showclipboard.workflow
Open
Input: Process items as a group
Output: Wait for workflow to finish
Return action input
Results Options
Loop
Loop automatically
Stop after 48 times
Use the current results as input
Results Options
Run Shell Script
Shell: /bin/bash
Pass input: as arguments
echo 0
Results Options
Set Value of Variable


Get Contents of Clipboard
Results Options
Set Value of Variable
Variable: clipval
Results Options
Extract URLs from Text
Extract URLs from text
Results Options
Set Value of Variable
Variable: clipval
Results Options
Download URLs
Where: CRIMEARCHIVES
Results Options
Ask for Confirmation


Get Contents of Clipboard
Results Options
Set Value of Variable
Variable: clipval
Results Options
Extract URLs from Text
Extract URLs from text
Results Options
Set Value of Variable
Variable: clipval
Results Options
Download URLs
Where: CRIMEARCHIVES
Results Options
Ask for Confirmation

APPENDIX B

```
Venue Frequency by Location

----Anderson----
                  venue  freq
0           Pizza Place  0.17
1   American Restaurant  0.17
2  Fast Food Restaurant  0.17
3        Discount Store  0.17
4            Shoe Store  0.08


----Belton----
                  venue  freq
0  Fast Food Restaurant  0.33
1           Video Store  0.11
2         Shopping Mall  0.11
3        Discount Store  0.11
4     Mobile Phone Shop  0.11


----Cleveland----
            venue  freq
0  Scenic Lookout   0.5
1      Campground   0.5
2             ATM   0.0
3     Pizza Place   0.0
4       Pet Store   0.0


----Fountain Inn----
                   venue  freq
0    Fried Chicken Joint   0.5
1   Fast Food Restaurant   0.5
2                    ATM   0.0
3             Pet Store   0.0
4  Performing Arts Venue   0.0


----Greenville----
```

```
                 venue  freq
0  American Restaurant  0.05
1                Hotel  0.04
2           Steakhouse  0.04
3          Pizza Place  0.04
4           Restaurant  0.04


----Greer----
                  venue  freq
0           Coffee Shop  0.18
1         Grocery Store  0.09
2                Lounge  0.09
3                   Pub  0.09
4  Ethiopian Restaurant  0.09


----Honea Path----
            venue  freq
0  Discount Store  0.25
1  Sandwich Place  0.25
2             Bar  0.25
3   Moving Target  0.25
4             ATM  0.00


----Mauldin----
                  venue  freq
0              Pharmacy   0.2
1          Home Service   0.2
2                  Bank   0.2
3        Hardware Store   0.2
4  Fast Food Restaurant   0.2


----Pendleton----
                   venue  freq
0                   Pool   1.0
1            Music Store   0.0
2               Pharmacy   0.0
```

```
3              Pet Store   0.0
4  Performing Arts Venue   0.0


----Piedmont----
                   venue  freq
0            Supermarket   1.0
1                    ATM   0.0
2                 Museum   0.0
3              Pet Store   0.0
4  Performing Arts Venue   0.0


----Simpsonville----
                   venue  freq
0            Pizza Place   1.0
1            Music Store   0.0
2               Pharmacy   0.0
3              Pet Store   0.0
4  Performing Arts Venue   0.0


----Slater----
         venue  freq
0         Park   1.0
1          ATM   0.0
2  Music Store   0.0
3     Pharmacy   0.0
4    Pet Store   0.0


----Taylors----
            venue  freq
0  Lingerie Store   0.5
1          Lounge   0.5
2             ATM   0.0
3     Music Venue   0.0
4        Pharmacy   0.0
```