

## Small College Towns – Are they all the same?

### Background

In the spirit of the “The Battle of Neighborhoods” challenge, our neighborhood will be defined by small college towns. Small college towns are communities that are largely defined by the single largest industry in the town, namely higher education. They are the subject of intense interest for both parents and young adults for a relatively short period of time during the search for a college and the subsequent (hopefully four) years spent at that college. Every year, there are analyses performed by different entities that offer their guidance on the “best” colleges. An example of such guidance was offered in December 2019 (pre-pandemic) by WalletHub.com.

<https://wallethub.com/edu/e/best-worst-college-cities-and-towns-in-america/8974/>

The analysis was based on three key dimensions; Wallet Friendliness, Social Environment and Academic & Economic Opportunities. The interested reader can discover the details of their methodology at

<https://wallethub.com/edu/e/best-worst-college-cities-and-towns-in-america/8974/#methodology>

The interest of the author was in looking at other dimensions like climate and the businesses that are within walking distance of the campus. The primary objective of this analysis was to answer the following questions.

Are all small college towns pretty much the same? Does geographical location largely determine climate? Are there distinctive characteristics we can identify based on the walking distance businesses?

Given the limited amount of time available for this analysis, our approach was to take the WalletHub top 25 small college towns (< 150,000 population) out of the available 200 as the set of college towns to analyze. Limiting the set in this way may prove limiting in our understanding and conclusions. We hope not!

If there are distinctive characteristics and college towns are not all the same, here are some possible business uses of the data and analysis

1. Local business development communities at lower ranked college towns could compare themselves to higher ranking college towns and determine what kinds of walking distance businesses they could encourage the development of in their town.
2. National or regional business chains could identify opportunities to expand into otherwise “similar” college towns that are currently under represented by their type of business.
3. College marketing groups could use distinctive characteristics in their outreach efforts for students.

## Data

The list of college towns selected for analysis were:

Town	College Name
Ann Arbor, MI	University of Michigan
Provo, UT	Brigham Young University (UT)
Rexburg, ID	Brigham Young University (ID)
Charleston, IL	Eastern Illinois University
Stevens Point, WI	University of Wisconsin
West Lafayette, IN	Purdue University
Charlottesville, VA	University of Virginia
Orem, UT	Utah Valley University
Cambridge, MA	MIT
Santa Barbara, CA	UC Santa Barbara
San Marcos, TX	Texas State University
Asheville, NC	University of NC Asheville
San Luis Obispo, CA	California Polytechnic State University
College Station, TX	Texas A&M
Berkeley, CA	UC Berkley
Chapel Hill, NC	University of NC Chapel Hill
Oxford, OH	Miami University College of Arts and Science
Boca Raton, FL	Florida Atlantic University
Athens, GA	University of Georgia
State College, PA	Pennsylvania State University
Blacksburg, VA	Virginia Tech
Mount Pleasant, MI	Central Michigan University
Storrs, CT	University of Connecticut
Statesboro, GA	Georgia Southern University
East Lansing, MI	Michigan State University

For each college town in our set, there are two data sets that were gathered.

1. Basic data – student population size, average temperatures, rain and snowfall, sunny days and elevation
2. Foursquare data – venue data within one mile of the college campus up to a limit of 100 venues per college.

The basic data was gathered using web searches and was compiled manually into an Excel spreadsheet which was then converted to a CSV file. The first five rows of the data set are provided below in Table 1.

Table 1. Basic Data

town	college name	latitude	longitude	total	undergraduate	graduate	avg rain	avg snow	sunny days	avg jan temp	avg jul temp	elevation
Ann Arbor, MI	University of Michigan	42.2808	-83.7303	48090	31266	16824	36	43	178	16.5	82.8	879
Provo, UT	Brigham Young University (UT)	40.2486	-111.6493	33511	30745	2766	17.7	42.7	222	21.1	92.8	4550
Rexburg, ID	Brigham Young University (ID)	43.8142	-111.783	34000	34000	0	13.6	41.8	206	12.4	83.8	4865
Charleston, IL	Eastern Illinois University	39.4844	-88.1751	7415	6012	1403	40.3	15.1	202	20.2	86.2	673
Stevens Point, WI	University of Wisconsin	44.5234	-89.5689	7725	7385	340	32.2	45.2	187	6.6	81.6	1089

The Foursquare data was gathered using the free developer API. The venue data from Foursquare includes the name of the venue, a foursquare categorization of the venue and location data for the venue. The walking distance venues characterize the “local neighborhood” for a college town from an on-campus student’s perspective.

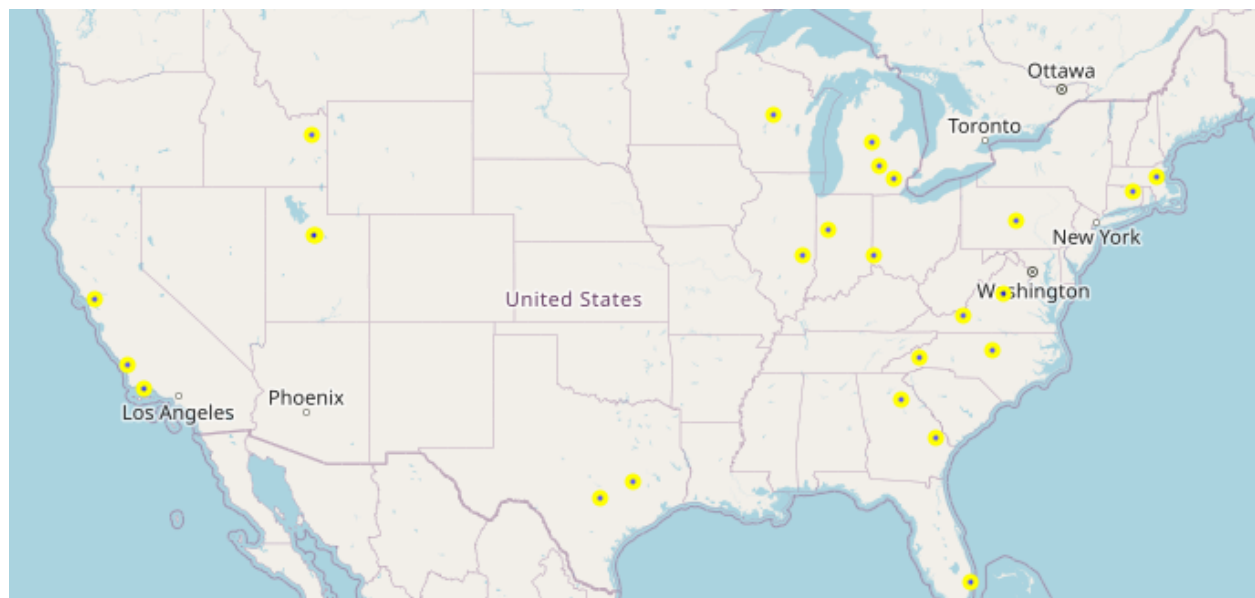
The combination of these two sets of data were explored in an attempt to see how alike or different college towns are. Are there truly distinctive characteristics and can be explained by climate alone.

### Methodology and Results

Data analysis is always performed within a set of constraints. In this case the constraints were time (one week of duration) and a requirement to use Foursquare venue data in the analysis in a “battle of the neighborhoods” themed analysis. The author grew up and was educated in Australia. However, having a child who graduated from a college in a small college town and enjoying that town personally was the motivation to perform this analysis.

This analysis is fundamentally an exploratory data analysis. This was driven by the constraints already described and the small data sets; (25 towns X 100 max venues) being the largest data set in the analysis.

The US is a large country and has very diverse climate conditions. The location of each college town is illustrated in the map below.



Our first step was to examine the student experience from a basic data perspective (student population size and climate). We chose to perform an unsupervised machine learning using the following feature set.

- Total student population
- Undergraduate student population
- Graduate student population
- Average rainfall in inches
- Average snow in inches
- Number of sunny days
- Average January (winter) temperature
- Average July (summer) temperature
- Elevation in feet

Experienced practitioners may notice some issues with this choice. Do you?

Hierarchical clustering was chosen for the initial approach. This choice was made primarily because hierarchical clustering does not require a cluster size and the output Dendrogram provides a good intuition as to any potential clusters.

The clustering activity was refined in an iterative fashion. Multiple passes were made with different feature sets. The final feature set chosen was

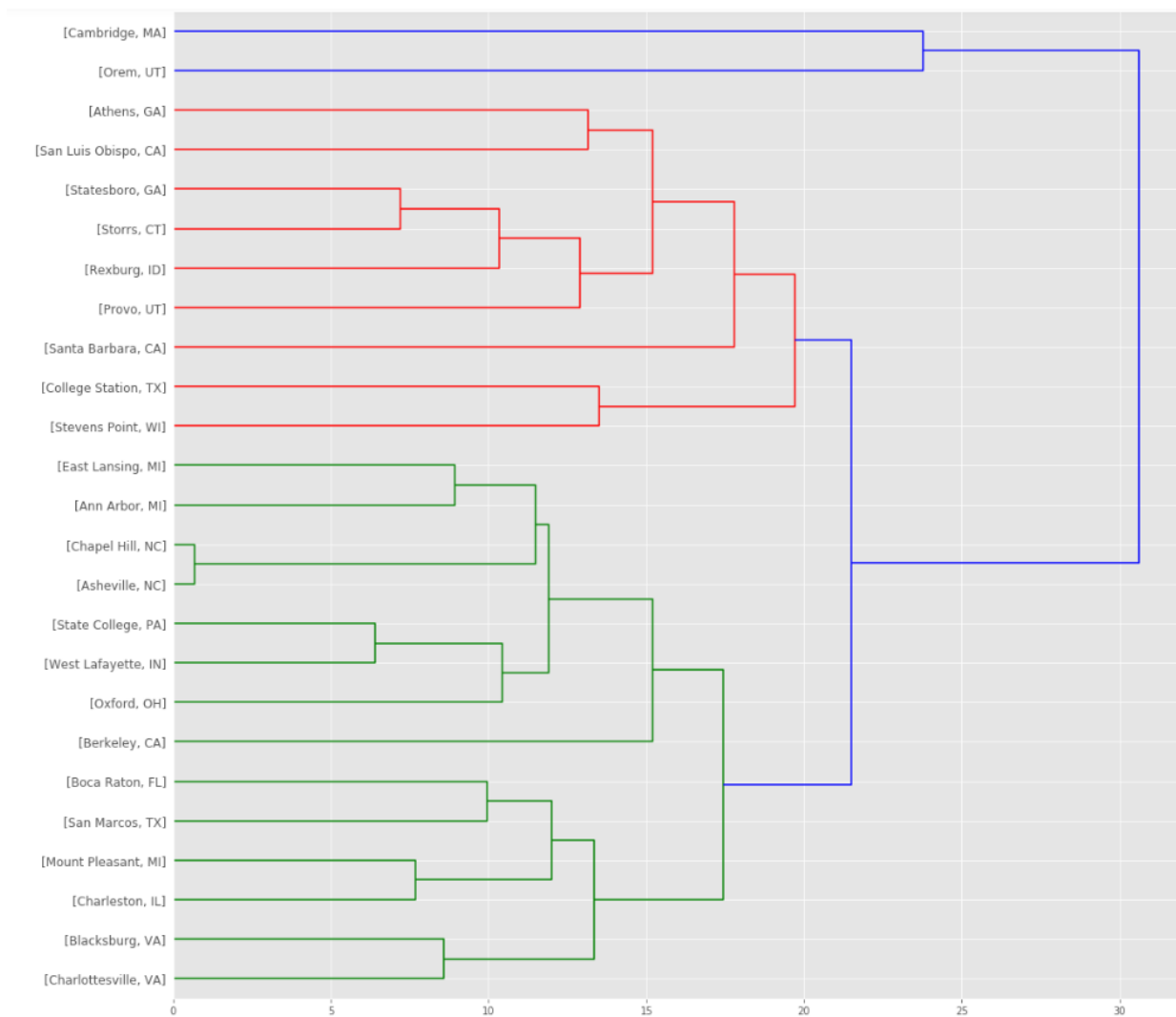
- Total student population
- Average precipitation
- Number of sunny days
- Average January (winter) temperature

Undergraduate and graduate student population were eliminated as they are redundant with total student population. Students are largely absent from college in the summer so Average July temperature was also dropped. Average rainfall and snow were combined (snow converted to rain based on a common 10inches of snow = 1 inch of rain ration) into Average precipitation.

A set of clusters were identified based on the four features. The clusters are

- Medium size, cloudy, cold and wet
- Medium size, sunny, cold and dry
- Medium size, sunny, cold and wet
- Medium size, super sunny, temperate and dry
- Large, sunny, temperate and wet
- Medium size, sunny, hot and wet

The choice of six clusters was based on a review of the dendrogram presented below and the summary statistics of each cluster generated with six clusters.



total   sunny days   avg jan temp   avg precip

hierarchy cluster

0	30891.111111	182.666667	15.666667	43.074444
1	27614.571429	213.428571	27.085714	45.325714
2	53826.000000	220.500000	38.850000	37.655000
3	35814.000000	217.000000	18.166667	20.563333
4	30209.000000	275.333333	43.166667	21.333333
5	29772.000000	231.000000	57.800000	60.000000

Each of our college towns was associated with the cluster. The results are presented below.

Town	College Name	Hierarchy Cluster
Ann Arbor, MI	University of Michigan	Medium size, cloudy, cold and wet
Provo, UT	Brigham Young University (UT)	Medium size, sunny, cold and dry
Rexburg, ID	Brigham Young University (ID)	Medium size, sunny, cold and dry
Charleston, IL	Eastern Illinois University	Medium size, sunny, cold and wet
Stevens Point, WI	University of Wisconsin	Medium size, cloudy, cold and wet
West Lafayette, IN	Purdue University	Medium size, cloudy, cold and wet
Charlottesville, VA	University of Virginia	Medium size, sunny, cold and wet
Orem, UT	Utah Valley University	Medium size, sunny, cold and dry
Cambridge, MA	MIT	Medium size, cloudy, cold and wet
Santa Barbara, CA	UC Santa Barbara	Medium size, super sunny, temperate and dry
San Marcos, TX	Texas State University	Large, sunny, temperate and wet
Asheville, NC	University of NC Asheville	Medium size, sunny, cold and wet
San Luis Obispo, CA	California Polytechnic State University	Medium size, super sunny, temperate and dry
College Station, TX	Texas A&M	Large, sunny, temperate and wet
Berkeley, CA	UC Berkley	Medium size, super sunny, temperate and dry
Chapel Hill, NC	University of NC Chapel Hill	Medium size, sunny, cold and wet
Oxford, OH	Miami University College of Arts and Science	Medium size, cloudy, cold and wet
Boca Raton, FL	Florida Atlantic University	Medium size, sunny, hot and wet
Athens, GA	University of Georgia	Medium size, sunny, cold and wet
State College, PA	Pennsylvania State University	Medium size, cloudy, cold and wet
Blacksburg, VA	Virginia Tech	Medium size, sunny, cold and wet
Mount Pleasant, MI	Central Michigan University	Medium size, cloudy, cold and wet
Storrs, CT	University of Connecticut	Medium size, cloudy, cold and wet
Statesboro, GA	Georgia Southern University	Medium size, sunny, cold and wet
East Lansing, MI	Michigan State University	Medium size, cloudy, cold and wet

Anyone familiar with the US will recognize that geographic location is a key factor in these clusters. The clusters are strongly associated with California, Texas, Florida, the Northwest, Northeast and Southeast. So we have really just confirmed our intuition that geographic location strongly influences climate.

Our second step was to gather and explore the venue data for each college town.

The Foursquare API was used to get a maximum of 100 venues within a one mile radius of the college itself. The key features of this data set are

- Venue Name
- Venue Category
- Venue Latitude
- Venue Longitude

The average count of venues per college town was 76. So not all college towns have 100 venues within a mile radius. There were a total of 271 different Foursquare categories in the venue data set. Each Foursquare venue category was itself categorized into Category1 and then Category1 categorized into Category2. These categorizations are provided below in Table 2.

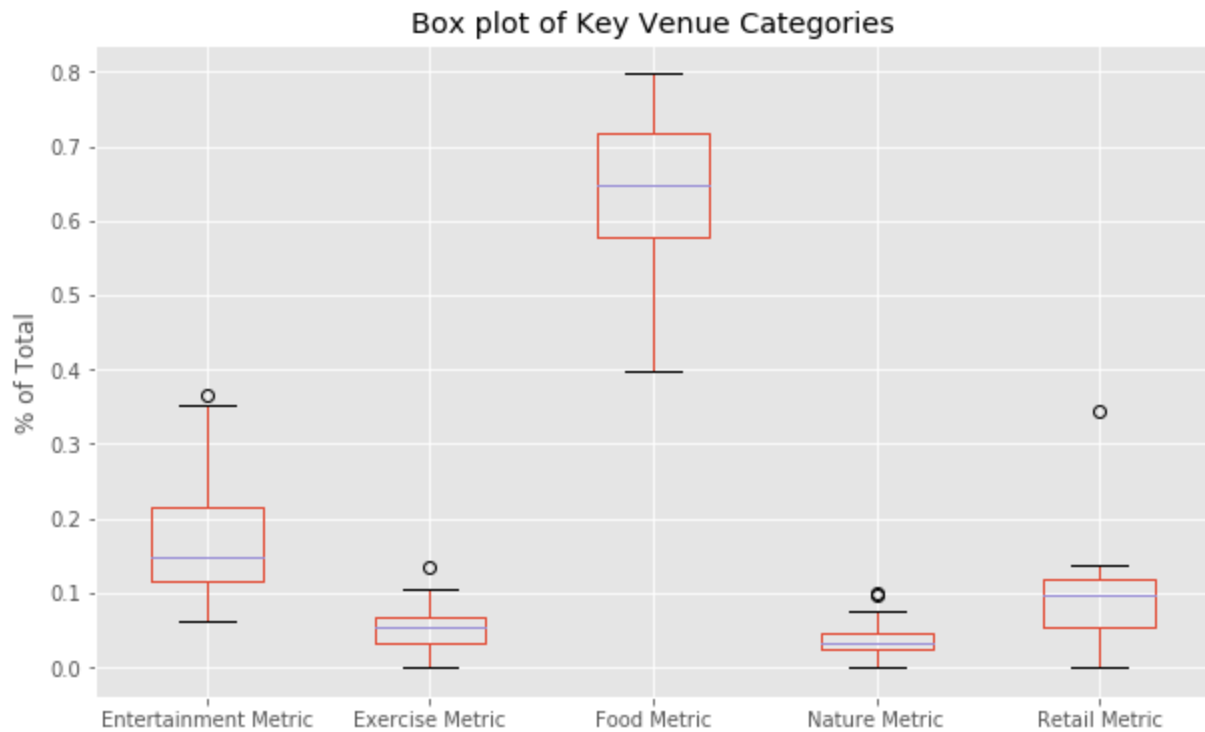
Table 2. Custom categories for Foursquare venue categories

Category1	Category2
Arts Venue	Entertainment
Drinking Venue	Entertainment
Other Venue	Entertainment
Museum	Entertainment
Exclude	Exclude
Gym	Exercise
Pool	Exercise
Sporting-Leisure	Exercise
Yoga Studio	Exercise
Grocery	Food
Casual Dining	Food
Coffee	Food
Dessert	Food
Dining	Food
Fast Food	Food
Ice Cream	Food
International Dining	Food
Vegetarian Dining	Food
Nature	Nature
Retail	Retail

Note: There were some Foursquare categories that were considered insignificant in determining the character of the college town. An example is ATM or Bank. Such Foursquare categories were included in the Exclude Category1 and were eliminated from the analysis.

Each venue was then categorized and counted by college town. A new measure was created in the data set for each of the Category1 and Category2 entries. For example, a Food Metric that counted each venue with a Category2 = Food was now available for each college town.

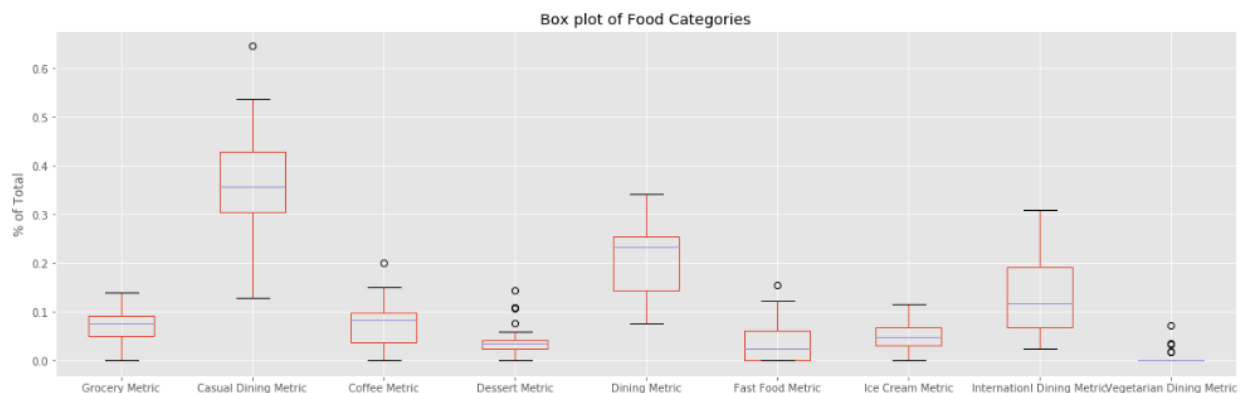
The Box plot below illustrates the distribution of venues into Category2 as a percent of total for the college town.



The obvious takeaway is that Food and Entertainment are the most common venues in college towns; not surprising, students have to eat and like to play after a hard day learning (or not!)

It is also interesting to see that there are some college towns that have no Exercise, Nature or Retail venues within walking distance. That was a surprise to the author.

The boxplot below shows the same percent of total distribution within Food venues.

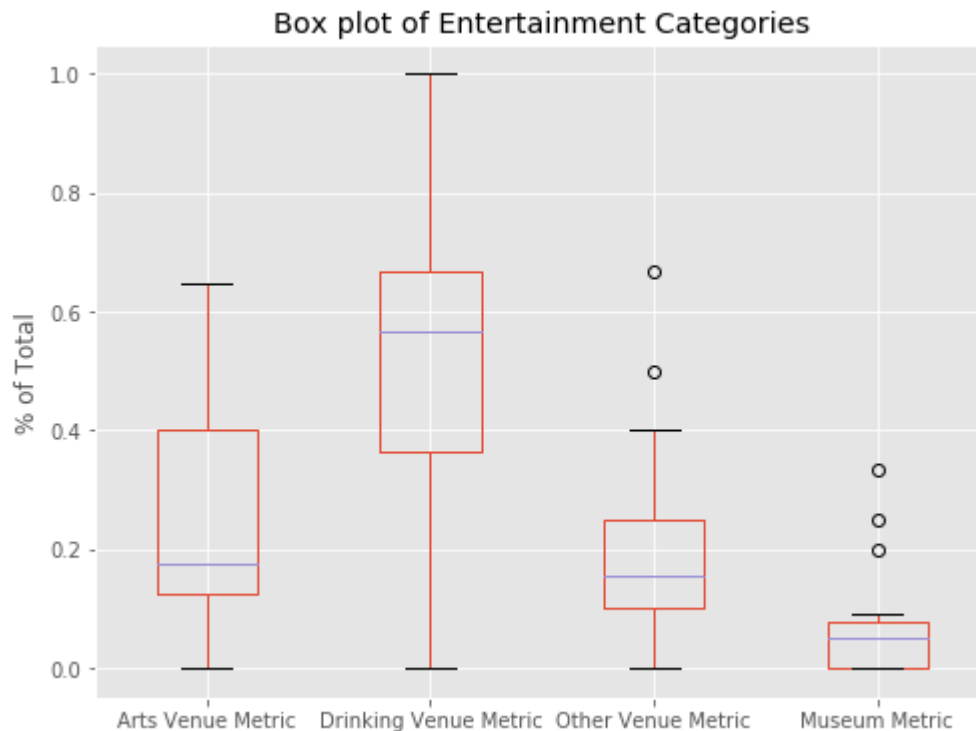


There are some obvious takeaways here. Casual dining is the most frequently occurring category. "Don't forget you are a poor college student" is a common parental mantra. The Dining category is the next step up in price and second in frequency on average. International Dining was the third most frequent category albeit with a very broad range of values. Fast food was surprisingly low but there is an outlier



that may be worth exploring further. There are some college towns without a Coffee shop! How can that be you ask? Stay tuned. Finally, Vegetarian Dining is non-existent with the exception of three outliers.

A similar analysis of the Entertainment category is presented in the boxplot below.



Con

The most surprising category is the Drinking Venue. It ranges from zero to all venues within a single college town. Read that again! The Arts Venue also has a very significant range. The takeaway here is that within the entertainment venue category, college towns can differ significantly.

### Discussion

The exploratory data analysis was successful in showing that while college towns have the same basic types of venues albeit in different relative quantities, there are definitely some significant differences in the specific types of venues one might find.

There are four college towns that have either zero or a single drinking establishment and coffee shop. How can that be you ask? The first three of the four college towns listed below are all in states that have significant Mormon populations. Mormons do not drink coffee or alcohol so the venue data lines up well.

- Provo, UT
- Rexburg, ID
- Orem, UT
- Statesboro, GA

At the other end of the spectrum is Santa Barbara, CA where every single entertainment venue is a drinking venue. All other college towns fall somewhere in between. The mean is about 60% but skewed towards the lower end.

Frequency of Coffee drinking venues is not correlated with a colder climate. This was a little surprising as the author expected to see more coffee shops in colder climates. The top five college towns for coffee shop venues were

- Ann Arbor, MI
- College Station, TX
- Oxford, OH
- Athens, GA
- East Lansing, MI

The top five college towns for International Dining were

- Cambridge, MA
- Berkeley, CA
- Asheville, NC
- Chapel Hill, NC
- Ann Arbor, MI

While Cambridge, Berkeley and Ann Arbor seem like they might be obvious choices for a more diverse population, the two North Carolina college towns are a surprise just considering that North Carolina is generally considered a conservative and less diverse state. The implication is that small college towns might not necessarily follow the state “average”. It would be interesting to examine other similar sized towns in NC that do not contain colleges and compare them on this metric.

The ONLY college towns that have a Vegetarian restaurant were

- Cambridge, MA
- Berkeley, CA
- Ann Arbor, MI
- Blacksburg, VA
- Santa Barbara, CA

There is a significant overlap between International Dining and Vegetarian Dining. Both are markers for a more diverse and potentially progressive student population.

### Conclusions

College towns are not all the same. There is a wide range of venues within walking distance and the character of the student population can be discerned to some degree from the walking distance venues.

Given the constraints described previously in this document, the exploratory data analysis indicates that venue data is likely a fertile area for investigation.

The small sample size of college towns was definitely a limiting factor on more sophisticated analyses that would seek to identify relationships between climate and walking distance venues. Given more time the author would have gathered the basic and venue data sets for all 200 college towns in the original WalletHub.com report and attempted to find dependencies in the feature set.

This project required the execution of all steps in a typical data analysis and required multiple iterations including some trial and error and dead-ends. In the business world, it has been my experience that there is typically more direction given on a specific problem to be solved. Understanding that this was an academic activity was important to the author in the frustrating early stages of problem formulation.