

Classification of Tourist Towns in the United States

David Sachs

December 20, 2019

1. Introduction

1.1 Background

In 2014, 76.9 million people visited the United States of America. This industry represents roughly 2.8% of the entire US GDP, with a spend of \$251 billion dollars. A key feature of this is that people come to visit different regions and cities within the US for very different reasons. America offers everything from historic locations to world-class skiing to warm beaches to amazing cities. Tourists must research and plan which regions they want to visit based off their interests. The US tourist industry gains value through positive tourism experiences, as they encourage repeat trips. This therefore means that tourists accurately selecting travel locations that match their desired vacation is critical to the overall longevity of US tourism. Being able to create clusters or classifications of tourist towns both enables tourists to determine what cities are interesting, and to find more like them for subsequent trips.

1.2 Problem

Data that might contribute to determining tourist town classifications include the locations and stores that make up that area, as well as the names of those locations. This project aims to classify tourist towns and create clusters to help tourists identify towns of interest and similar locations.

1.3 Interest

People planning trips to the US would be interested in this tool, as it is an easy way to understand what a city has to offer. Cities would be interested as well, as this helps them understand potential competition as well as if there is another cluster they would rather be a part of. Retail companies and restaurants would also benefit, as this may help them find similar markets to open additional locations.

2. Data Acquisition and Cleaning

2.1 Data Sources

The types and names of locations in small towns can be found using the FourSquare API. The list of 25 sample towns was gathered from the US News Best Small Towns list available [here](#). Other potential location lists were considered, but would require removing significant portions of the list such as national parks due to a lack of nearby locations. The latitude and longitude of the locations was gathered using the Geopy library. Examples of the target data would be the locations in Telluride, Colorado, gained from FourSquare. The types of locations and words in the location names would then be used to classify the towns.

2.2 Data Cleaning

Data cleaning was required throughout this project. The list of locations was loaded from the referenced website, had latitudes and longitudes assigned using the Geopy library, and then had locations found using FourSquare. Upon an initial review of the gathered locations, multiple issues were found. Several towns had very few locations and others did not have any locations found.

This was caused by two issues: poorly named locations and inaccurate locations from Geopy. Some locations in the list were county names, or informal names. Examples are Hilton Head, which is technically named Hilton Head Island, and Jackson Hole, which is actually the town of Jackson. In addition, counties like Sonoma had to be focused onto a certain town like Santa Rosa. Other location errors were caused by inaccurate latitudes and longitudes. An example of this issue was Cape May. Geopy returned coordinates nearly a full degree north of the real town.

Finally, cleaning was required to get useful results from the most common words in the venue names. When uncleaned, the most common entries for most towns were ‘the’, ‘&’ and the town name. These values do not provide meaningful insight into the nature of a town. It is also unlikely that a town would ever have locations named after another town. As such, these values were removed from the final data set.

2.3 Feature Selection

A key element to feature selection was determining the radius for the FourSquare call. Unlike cities like New York or Toronto which have hundreds of locations in a few blocks, many of these tourist towns are very rural. Key locations, such as ski slopes, beaches, and trails may be outside of the downtown area. Multiple radii were tested, with 10000 meters selected due to the quality of top categories as well as making sense for a standard vacation. A 6 to 10 mile drive is reasonable to expect to get a location and can still be said to be part of the actual town’s character.

Another key feature for review was the latitude and longitude itself. While useful for determining elements such as weather and proximity, it strongly skews the clusters to be focused on regions. This takes away from the investigation of using venues to determine similar tourist locations.

3. Exploratory Data Analysis

3.1 Initial Mapping

Critical to understanding the data set was to first map the locations with no clustering assigned. This is done to understand what natural groups may be found. I used the corrected latitudes and longitudes to plot the locations into a Folium map. The resulting map is below, along with expected geographic clusters.

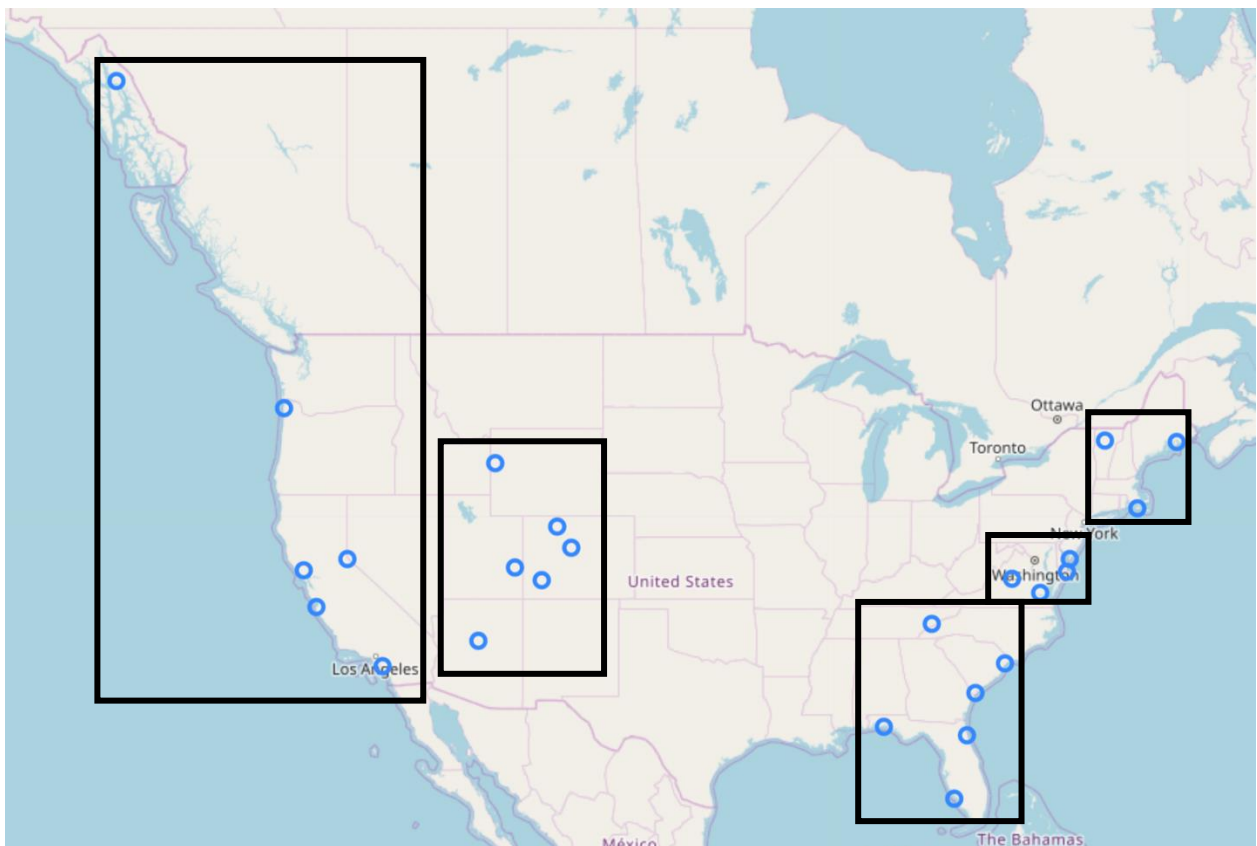


Figure 1. Geographic Clustering of Tourist Towns

Geographic clusters are the West Coast, the Rocky Mountains, the Southeast, the Mid-Atlantic, and the Northeast.

3.2 Most Common Venue Categories

100 venues were selected per tourist town. The 10 most common venue categories were then determined for each town. This was then reviewed to look for common patterns. Some patterns did appear. Certain mainstays of tourist towns, such as hotels, restaurants, and bars were in the top ten of every town.

The most interesting elements of this analysis was the unique values in each town, as well as the greater story. Towns like Breckenridge and Telluride were clearly linked by a plethora of skiing related items. Similar, the many coastal locations shared beach related venues. Cuisine options were also interesting to review, as the tone of the town can be estimated by the naming of places as “joints” versus “restaurants” as well as “pubs” vs. “bars”.

Town	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
Bar Harbor	Seafood Restaurant	Trail	Hotel	American Restaurant	Ice Cream Shop	Mountain	Park	Beach	Breakfast Spot	Brewery
Breckenridge	Ski Area	Coffee Shop	Hotel	American Restaurant	Ski Chairlift	American Restaurant	Ski Lodge	Beer Garden	Restaurant	Pizza Place
Cannon Beach	Beach	Hotel	American Restaurant	Coffee Shop	Seafood Restaurant	Resort	State / Provincial Park	Brewery	Restaurant	Trail
Cape May	American Restaurant	Beach	Seafood Restaurant	Hotel	Coffee Shop	Ice Cream Shop	Campground	Breakfast Spot	Italian Restaurant	Deli / Bodega
Charlottesville	Brewery	American Restaurant	Music Venue	Coffee Shop	Burger Joint	Pizza Place	Mexican Restaurant	Bagel Shop	Park	Hotel
Destin	Seafood Restaurant	Beach	Resort	Grocery Store	Pizza Place	Coffee Shop	Mexican Restaurant	Bar	Furniture / Home Store	Bookstore
Gatlinburg	Resort	Hotel	Scenic Lookout	Breakfast Spot	Candy Store	Trail	BBQ Joint	Ski Chairlift	American Restaurant	Donut Shop
Hilton Head Island	Beach	Resort	American Restaurant	Pizza Place	Hotel	Restaurant	Seafood Restaurant	Park	Harbor / Marina	Grocery Store
Jackson	Hotel	American Restaurant	Coffee Shop	Grocery Store	Resort	Sandwich Place	Bar	BBQ Joint	Breakfast Spot	Pizza Place
Juneau	Seafood Restaurant	Coffee Shop	Bar	Pizza Place	American Restaurant	Café	Dive Bar	Mexican Restaurant	Mountain	Trail
Laguna Beach	Beach	Mexican Restaurant	Grocery Store	American Restaurant	Seafood Restaurant	New American Restaurant	Resort	Hotel	Park	Scenic Lookout
Martha's Vineyard	American Restaurant	Beach	Seafood Restaurant	Bar	Bakery	Coffee Shop	Farm	Diner	Ice Cream Shop	Hotel
Moab	Hotel	Trail	Campground	American Restaurant	National Park	Café	Motel	Food Truck	Sporting Goods Shop	Mexican Restaurant
Monterey	Beach	Seafood Restaurant	Aquarium	Coffee Shop	Breakfast Spot	Grocery Store	Sushi Restaurant	Golf Course	American Restaurant	Hotel
Myrtle Beach	Hotel	Burger Joint	Italian Restaurant	American Restaurant	Seafood Restaurant	Bar	Department Store	Steakhouse	Gift Shop	Pub
Ocean City	Beach	Seafood Restaurant	Bar	Hotel	American Restaurant	Clothing Store	Restaurant	Café	Deli / Bodega	Italian Restaurant
Sanibel Island	Resort	Beach	Seafood Restaurant	American Restaurant	Trail	Hotel	Harbor / Marina	New American Restaurant	Bar	Pizza Place
Santa Rosa	Coffee Shop	Grocery Store	Brewery	Thai Restaurant	American Restaurant	Mexican Restaurant	Pizza Place	Café	Italian Restaurant	Bakery
Sedona	Trail	Hotel	American Restaurant	Mexican Restaurant	Scenic Lookout	Resort	Italian Restaurant	Brewery	New American Restaurant	Pizza Place
South Lake Tahoe	Coffee Shop	Hotel	American Restaurant	Beach	Italian Restaurant	Ski Area	Mexican Restaurant	Pizza Place	Brewery	Resort
St. Augustine	American Restaurant	Seafood Restaurant	Pizza Place	Cocktail Bar	Pub	Music Venue	History Museum	Historic Site	Café	Brewery
Steamboat Springs	American Restaurant	Ski Area	Brewery	Mexican Restaurant	Trail	Pub	Coffee Shop	Italian Restaurant	Steakhouse	Pizza Place
Stowe	Trail	Hotel	American Restaurant	Brewery	Resort	Bakery	New American Restaurant	Breakfast Spot	Bed & Breakfast	Gift Shop
Telluride	Hotel	Bar	Pizza Place	Performing Arts Venue	Sandwich Place	Sports Bar	Ski Chairlift	Trail	French Restaurant	Brewery
Williamsburg	American Restaurant	History Museum	Seafood Restaurant	Hotel	Gift Shop	Historic Site	Clothing Store	Coffee Shop	Restaurant	Brewery

Table 1. Towns with Top 10 Venue Categories

Other interesting standouts include Moab, with its clear association to the outdoors seen in national parks, trails, and campgrounds.

3.3 Most Common Words in Venue Names

Following the venue selection, the names of the locations were parsed into individual words. The most common words were then determined for each town. As mentioned in the previous section, extensive cleaning was required here. Without cleaning, the most common words in every town were the town name, the word ‘the’, and various forms of the word ‘and’.

Town	1st Most Common Word	2nd Most Common Word	3rd Most Common Word	4th Most Common Word	5th Most Common Word	6th Most Common Word	7th Most Common Word	8th Most Common Word	9th Most Common Word	10th Most Common Word
Telluride	gondola	station	mountain	sheridan	lodge	village	brown	pizza	cafe	new
Jackson	restaurant	inn	teton	pearl	street	hotel	center	grill	pizza	bakery
South Lake Tahoe	cafe	heavenly	coffee	grill	dog	pizza	edgewood	hut	south	lodge
Sedona	park	resort	trail	creek	spa	park	oak	red	cafe	trailhead
Bar Harbor	park	acadia	cafe	hotel	lobster	trail	street	inn	cove	desert
Gatlinburg	smoky	resort	falls	mountain	ole	center	bear	great	pancake	cabin
Monterey	bay	del	monte	grill	park	market	house	state	co.	old
Steamboat Springs	restaurant	grill	resort	creek	express	chairlift	brewery	cafe	mountain	storm
Williamsburg	shop	tavern	house	grill	market	square	cafe	dog	palace	colonial
Moab	park	center	inn	suites	arches	trail	cafe	arch	campground	national
Stowe	trapo	family	mountain	lodge	inn	market	cafe	resort	topnotch	green
Juneau	alaska	alaskan	company	cafe	brewing	co.	heritage	salmon	treadwell	adventure
Laguna Beach	park	grill	cove	market	crystal	pelican	bay	cafe	montage	also
Santa Rosa	coffee	market	sonoma	brewing	company	cafe	street	center	restaurant	county
Breckenridge	peak	lodge	center	superchair	restaurant	blue	ski	summit	mountain	coffee
Sanibel Island	lighthouse	club	grill	restaurant	cafe	store	inn	resort	bay	marina
Martha's Vineyard	market	restaurant	farm	edgartown	west	view	state	house	cafe	inn
Hilton Head Island	resort	cafe	sea	palmetto	dunes	cove	harbour	pinet	shelter	park
St. Augustine	restaurant	company	house	cafe	seafood	old	museum	historic	wine	coffee
Destin	seafood	oyster	house	park	henderson	mexican	grill	resort	home	starbucks
Myrtle Beach	resort	grill	broadway	restaurant	house	inn	oceanfront	grand	red	new
Ocean City	street	grill	restaurant	cafe	on	shop	fries	italian	center	pool
Cannon Beach	cafe	park	inn	restaurant	state	arch	coffee	company	lounge	ecola
Charlottesville	brewery	downtown	kitchen	park	bodo's	bagels	coffee	monticello	center	beer
Cape May	restaurant	hotel	company	congress	house	sunset	wildwood	inn	washington	hall

Table 2. Towns with Top 10 Venue Name Words

The intention of the exercise was to understand what things each town used as names. For example, Jackson frequently mentions Teton, the famous mountain range that surrounds the town. Telluride continues to exemplify

mountain living, with the most common words including gondola, mountain, and lodge, all words associated with skiing. Some groupings could be viewed here, as well as some towns that were very distinct. Juneau is dominated by words referring to Alaska. Breckenridge and Telluride are clearly linked via skiing.

Unfortunately, the unique elements are somewhat drowned out in this data set. The words grill, café, and restaurant are omnipresent while offering minimal information on the true similarity between the towns.

3.4 Merged Data Set

In the merged data set, the top 10 venue categories and the top 10 words in venue names were combined. The intention in this data set is to strengthen the category data by providing additional information into the venues. Having many locations with the word “mountain” in the name can help strengthen a mountain cluster where sharing “restaurants” may not.

An immediate observation in this data set is that the most common words do more to reinforce the category values than to provide additional insight. Many of the most common words are very similar or identical to location categories. While a human may be able to view this data set and see clear similarities, a computer will struggle. Example include how Hilton Head has common words of sea, dunes, cove, and harbor. While a human immediately knows that these all to the ocean and would expect a sea town, an algorithm will not know to associate these with other nautical terms unless they are exact matches.

Town	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	1st Most Common Word	2nd Most Common Word	3rd Most Common Word	4th Most Common Word	5th Most Common Word	6th Most Common Word	7th Most Common Word	8th Most Common Word	9th Most Common Word	10th Most Common Word
0 Bar Harbor	Seafood Restaurant	Trail	Hotel	American Restaurant	Ice Cream Shop	Mountain	Park	Beach	Breakfast Spot	...	park	acadia	cafe	hotel	lobster	trail	street	inn	cove	desert
1 Breckenridge	Ski Area	Coffee Shop	Hotel	Ski Chairlift	American Restaurant	Bar	Ski Lodge	Beer Garden	Restaurant	...	peak	lodge	center	superchair	restaurant	blue	ski	summit	mountain	coffee
2 Cannon Beach	Beach	Hotel	American Restaurant	Coffee Shop	Seafood Restaurant	Resort	State / Provincial Park	Brewery	Restaurant	...	cafe	park	inn	restaurant	state	arch	coffee	company	lounge	ecole
3 Cape May	American Restaurant	Beach	Seafood Restaurant	Hotel	Coffee Shop	Ice Cream Shop	Campground	Breakfast Spot	Italian Restaurant	...	restaurant	hotel	company	congress	house	sunset	wildwood	inn	washington	hall
4 Charlottesville	Brewery	American Restaurant	Music Venue	Coffee Shop	Burger Joint	Pizza Place	Mexican Restaurant	Bagel Shop	Park	...	brewery	downtown	kitchen	park	bodo's	bagels	coffee	monticello	center	beer
5 Destin	Seafood Restaurant	Beach	Resort	Grocery Store	Pizza Place	Coffee Shop	Mexican Restaurant	Bar	Furniture / Home Store	...	seafood	oyster	house	park	henderson	mexican	grill	resort	home	starbucks
6 Gatlinburg	Resort	Hotel	Scenic Lookout	Breakfast Spot	Candy Store	Trail	BBQ Joint	Ski Chairlift	American Restaurant	...	smoky	resort	falls	mountain	ole	center	bear	great	pancake	cabin
7 Hilton Head Island	Beach	Resort	American Restaurant	Pizza Place	Hotel	Restaurant	Seafood Restaurant	Park	Harbor / Marina	...	resort	cafe	sea	palmetto	dunes	cove	harbour	pinos	shelter	park
8 Jackson	Hotel	American Restaurant	Coffee Shop	Grocery Store	Resort	Sandwich Place	Bar	BBQ Joint	Breakfast Spot	...	restaurant	inn	teton	pearl	street	hotel	center	grill	pizza	bakery
9 Juneau	Seafood Restaurant	Coffee Shop	Bar	Pizza Place	American Restaurant	Café	Dive Bar	Mexican Restaurant	Mountain	...	alaska	alaskan	company	cafe	brewing	co.	heritage	salmon	treadwell	adventure
10 Laguna Beach	Beach	Mexican Restaurant	Grocery Store	American Restaurant	Seafood Restaurant	New American Restaurant	Resort	Hotel	Park	...	park	grill	cove	market	crystal	pelican	bay	cafe	montage	aliso
11 Martha's Vineyard	American Restaurant	Beach	Seafood Restaurant	Bar	Bakery	Coffee Shop	Farm	Diner	Ice Cream Shop	...	market	restaurant	farm	edgartown	west	view	state	house	cafe	inn
12 Moab	Hotel	Trail	Campground	American Restaurant	National Park	Café	Motel	Food Truck	Sporting Goods Shop	...	park	center	inn	suites	arches	trail	cafe	arch	campground	national
13 Monterey	Beach	Seafood Restaurant	Aquarium	Coffee Shop	Breakfast Spot	Grocery Store	Sushi Restaurant	Qof Course	American Restaurant	...	bay	del	monte	grill	park	market	house	state	co.	old
14 Myrtle Beach	Hotel	Burger Joint	Italian Restaurant	American Restaurant	Seafood Restaurant	Bar	Department Store	Steakhouse	Gift Shop	...	resort	grill	broadway	restaurant	house	inn	oceanfront	grand	red	new
15 Ocean City	Beach	Seafood Restaurant	Bar	Hotel	American Restaurant	Clothing Store	Restaurant	Café	Del / Bodega	...	street	grill	restaurant	cafe	on	shop	fries	italian	center	pool
16 Sanibel Island	Resort	Beach	Seafood Restaurant	American Restaurant	Trail	Hotel	Harbor / Marina	New American Restaurant	Bar	...	lighthouse	club	grill	restaurant	café	store	inn	resort	bay	marina
17 Santa Rosa	Coffee Shop	Grocery Store	Brewery	Thai Restaurant	American Restaurant	Mexican Restaurant	Pizza Place	Café	Italian Restaurant	...	coffee	market	sonoma	brewing	company	cafe	street	center	restaurant	county
18 Sedona	Trail	Hotel	American Restaurant	Mexican Restaurant	Scenic Lookout	Resort	Italian Restaurant	Brewery	New American Restaurant	...	rock	resort	trail	creek	spa	park	oak	red	cafe	trailhead
19 South Lake Tahoe	Coffee Shop	Hotel	American Restaurant	Beach	Italian Restaurant	Ski Area	Mexican Restaurant	Pizza Place	Brewery	...	cafe	heavenly	coffee	grill	dog	pizza	edgewood	hut	south	lodge
20 St Augustine	American Restaurant	Seafood Restaurant	Pizza Place	Cooktail Bar	Pub	Music Venue	History Museum	Historic Site	Café	...	restaurant	company	house	cafe	seafood	old	museum	historic	wine	coffee
21 Steamboat Springs	American Restaurant	Ski Area	Brewery	Mexican Restaurant	Trail	Pub	Coffee Shop	Italian Restaurant	Steakhouse	...	restaurant	grill	resort	creek	express	chairlift	brewery	cafe	mountain	storm
22 Stowe	Trail	Hotel	American Restaurant	Brewery	Resort	Bakery	New American Restaurant	Breakfast Spot	Bed & Breakfast	...	trap	family	mountain	lodge	inn	market	cafe	resort	topnotch	green
23 Telluride	Ski Area	Hotel	Bar	Pizza Place	Performing Arts Venue	Sandwich Place	Sports Bar	Ski Chairlift	Trail	...	gondola	station	mountain	sherman	lodge	village	brown	pizza	cafe	new
24 Williamsburg	American Restaurant	History Museum	Seafood Restaurant	Hotel	Gift Shop	Historic Site	Clothing Store	Coffee Shop	Restaurant	...	shop	tavern	house	grill	market	square	cafe	dog	palace	colonial

Table 3. Towns with 10 Top 10 Venue Categories and Top 10 Venue Name Words

4. Clustering

4.1 Methodology

The towns were clustered using the K-means method. The values were prepared for K-means using one-hot encoding to convert string values into float values between 0 and 1. This method was selected because the problem pertains to categorizing items and lacks the varied binary metrics to make something such as a decision tree.

One-hot encoded data frames were created for both the word and venue category data sets. These were then combined into the merged data frame representing both data sets. Following clustering, the clusters were plotted onto a map for review. Clustering was generated with different numbers of clusters ranging from 3 to 8. Less than three was clearly underfitting the data, while more than 8 clearly demonstrated overfitting. Due to the number of clusters, I will focus primarily on the strongest cluster and explain the issues with other clusters.

4.2 Clustering with Venue Category

While clustering with venue category, 4 clusters proved to be the number of clusters with the clearest groupings. 3 was able to group clusters into high level categories of outdoors, beach, and general tourist. However, at K=4, the outdoors group was split into two clearly different sub-categories. 5 was like 4 but provided no real improvement in information. Beyond 5, additional clusters were generally reserved for individual towns, adding little information to the model.

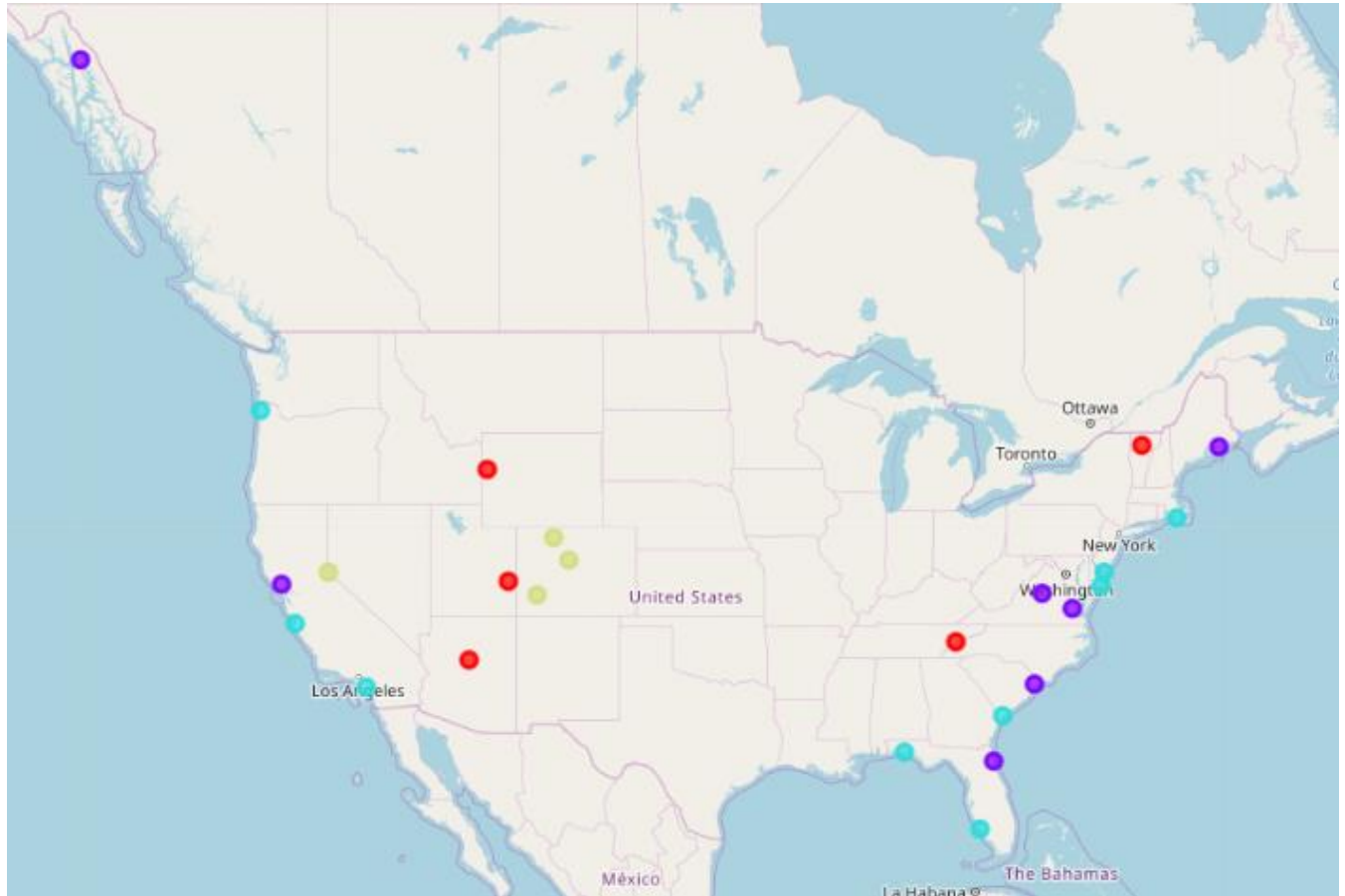


Figure 2. Towns Clustered by Venue Category, K = 4

Venue Cluster Labels	Town	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	Lat	Long
0	1	Bar Harbor	Seafood Restaurant	Trail	Hotel	American Restaurant	Ice Cream Shop	Mountain	Park	Beach	Breakfast Spot	Brewery	44.3885 -88.2119
1	3	Breckenridge	Ski Area	Coffee Shop	Hotel	Ski Chairlift	American Restaurant	Bar	Ski Lodge	Beer Garden	Restaurant	Pizza Place	39.4817 -106.038
2	2	Cannon Beach	Beach	Hotel	American Restaurant	Coffee Shop	Seafood Restaurant	Resort	State / Provincial Park	Brewery	Restaurant	Trail	45.8918 -123.952
3	2	Cape May	American Restaurant	Beach	Seafood Restaurant	Hotel	Coffee Shop	Ice Cream Shop	Campground	Breakfast Spot	Italian Restaurant	Del / Bodega	38.0255 -74.9137
4	1	Charlottesville	Brewery	American Restaurant	Music Venue	Coffee Shop	Burger Joint	Pizza Place	Mexican Restaurant	Bagel Shop	Park	Hotel	38.0293 -78.4787
5	2	Destin	Seafood Restaurant	Beach	Resort	Grocery Store	Pizza Place	Coffee Shop	Mexican Restaurant	Bar	Furniture / Home Store	Bookstore	30.3935 -86.4958
6	0	Gatlinburg	Resort	Hotel	Scenic Lookout	Breakfast Spot	Candy Store	Trail	BBQ Joint	Ski Chairlift	American Restaurant	Donut Shop	35.7143 -83.5102
7	2	Hilton Head Island	Beach	Resort	American Restaurant	Pizza Place	Hotel	Restaurant	Seafood Restaurant	Park	Harbor / Marina	Grocery Store	32.1618 -80.7513
8	0	Jackson	Hotel	American Restaurant	Coffee Shop	Grocery Store	Resort	Sandwich Place	Bar	BBQ Joint	Breakfast Spot	Pizza Place	43.4743 -110.765
9	1	Juneau	Seafood Restaurant	Coffee Shop	Bar	Pizza Place	American Restaurant	Café	Dive Bar	Mexican Restaurant	Mountain	Trail	58.3019 -134.42
10	2	Laguna Beach	Beach	Mexican Restaurant	Grocery Store	American Restaurant	Seafood Restaurant	New American Restaurant	Resort	Hotel	Park	Scenic Lookout	33.5427 -117.785
11	2	Martha's Vineyard	American Restaurant	Beach	Seafood Restaurant	Bar	Bakery	Coffee Shop	Farm	Diner	Ice Cream Shop	Hotel	41.3919 -70.6209
12	0	Moab	Hotel	Trail	Campground	American Restaurant	National Park	Café	Motel	Food Truck	Sporting Goods Shop	Mexican Restaurant	38.5738 -109.546
13	2	Monterey	Beach	Seafood Restaurant	Aquarium	Coffee Shop	Breakfast Spot	Grocery Store	Sushi Restaurant	Golf Course	American Restaurant	Hotel	36.6003 -121.88
14	1	Myrtle Beach	Hotel	Burger Joint	Italian Restaurant	American Restaurant	Seafood Restaurant	Bar	Department Store	Steakhouse	Gift Shop	Pub	33.6955 -78.89
15	2	Ocean City	Beach	Seafood Restaurant	Bar	Hotel	American Restaurant	Clothing Store	Restaurant	Café	Del / Bodega	Italian Restaurant	38.3349 -75.0848
16	2	Sanibel Island	Resort	Beach	Seafood Restaurant	American Restaurant	Trail	Hotel	Harbor / Marina	New American Restaurant	Bar	Pizza Place	26.4404 -82.1137
17	1	Santa Rosa	Coffee Shop	Grocery Store	Brewery	Thai Restaurant	American Restaurant	Mexican Restaurant	Pizza Place	Café	Italian Restaurant	Bakery	38.4405 -122.714
18	0	Sedona	Trail	Hotel	American Restaurant	Mexican Restaurant	Scenic Lookout	Resort	Italian Restaurant	Brewery	New American Restaurant	Pizza Place	34.8897 -111.781
19	3	South Lake Tahoe	Coffee Shop	Hotel	American Restaurant	Beach	Italian Restaurant	Ski Area	Mexican Restaurant	Pizza Place	Brewery	Resort	39.0332 -119.984
20	1	St. Augustine	American Restaurant	Seafood Restaurant	Pizza Place	Cocktail Bar	Pub	Music Venue	History Museum	Historic Site	Café	Brewery	29.8947 -81.3145
21	3	Steamboat Springs	American Restaurant	Ski Area	Brewery	Mexican Restaurant	Trail	Pub	Coffee Shop	Italian Restaurant	Steakhouse	Pizza Place	40.4848 -106.832
22	0	Stowe	Trail	Hotel	American Restaurant	Brewery	Resort	Bakery	New American Restaurant	Breakfast Spot	Bed & Breakfast	Gift Shop	44.4054 -72.886
23	3	Telluride	Ski Area	Hotel	Bar	Pizza Place	Performing Arts Venue	Sandwich Place	Sports Bar	Ski Chairlift	Trail	French Restaurant	37.9375 -107.812
24	1	Williamsburg	American Restaurant	History Museum	Seafood Restaurant	Hotel	Gift Shop	Historic Site	Clothing Store	Coffee Shop	Restaurant	Brewery	37.2789 -76.0845

Table 4. Towns Clustered by Venue Category, K = 4

The clusters found are best described as below:

Cluster 1: Ski Towns [Gold]

This cluster, composed of Lake Tahoe, Telluride, Breckenridge, and Steamboat Springs, is dominated by skiing related venues.

Cluster 2: Mountain/Outdoor Towns [Red]

This cluster is dominated by mountain and adventure terms, but more broadly so than the very focused skiing cluster. It is made up of Jackson, Sedona, Moab, Gatlinburg, and Stowe.

Cluster 3: Beach Towns [Teal]

This cluster is made up of towns that are very beach oriented. Venue highlights include the titular beach along with seafood restaurants. It is made up of Cannon Beach, Cape May, Destin, Hilton Head, Laguna Beach, Martha's Vineyard, Ocean City, and Sanibel Island.

Cluster 4: Come and Relax Towns [Purple]

This cluster is made up of towns that are more defined by a plethora of restaurants, hotels, and museums than any other activity. It is made up of Williamsburg, St. Augustine, Santa Rosa, Myrtle Beach, Juneau, Charlottesville, and Bar Harbor.

4.3 Clustering with Most Common Words

Clustering with common words generally proved to be less effective than clustering via venue categorization. Clusters changed dramatically between different numbers of clusters, and generally tended to single out towns more than group them. This is likely driven by the commonality of words like restaurant, hotel, and bar across all towns. However, unlike venues, other values were less likely to have similar values in other towns, leading them to have less impact. For example, no level of clustering grouped the towns of Breckenridge, Telluride, and Steamboat Springs despite being commonly grouped with venues with a clear explanation of skiing dominated towns.

None of the different clustering methods ended up producing understandable groupings. Low levels of K formed one or two mega clusters with singular towns in their own clusters. High levels of K would have three major clusters with standalones, though with no clear groupings.

A likely cause for this is the sheer number of unique words. With nearly 600 features, it is likely that the noise significantly outweighs the signal in the naming.

4.4 Clustering with the Merged Data Set

Clustering with the merged data sets performed similarly to clustering with the most common words. Clusters shifted dramatically with K and were largely unintelligible. The groupings also had internal conflict, where a cluster would feature radically different venue categories. The assumption again is that the sheer number of features in the word count created too much noise in the data set to find a valid group.

4.5 Comparison of Methods

As discussed above, grouping via venue categories was clearly the strongest method. Multiple levels of K had meaningful clusters produced, and clusters were easy to comprehend. Both methods that included the most common words generally were drowned in the noise of many features. In addition, the discussed issue of words that are similar to a human but not a computer would clearly take away from the algorithms ability to form coherent groups.

5. Conclusion

This modeling exercise proved that it is possible to group tourist towns into meaningful similar categories. The venue categories in tourist towns can be used to find similar towns for additional adventures. However, the direct use of counts of most common words in venue names introduced so much noise that the signal ended up being lost. This can be very helpful for visitors to these towns to find similar towns for additional trips as well as for the

towns to understand the competition and neighbors. For a town such as Jackson, they may wish to strengthen their skiing credentials to join the exclusive Skiing cluster. This was all done using K-means clustering to determine classifications for a sample of U.S. towns.

6. Future Direction

Additional information is clearly needed to strengthen the model. Further information such as the size of the towns, average cost of a night's stay, average cost of a meal, and the top activities in the town could be used to further enhance the model. I still firmly believe there is value somewhere in the naming of venues in a town. Additional steps in the algorithm such as weighting for the most common terms could help filter out the noise and substantially improve performance. It would also be worth investigating automated grouping of the words themselves. This would help turn 20 low count words relating to hiking into a clear category that can be used to help clearly identify a town. Finally, the model would be strengthened by adding more towns to the data set.