The Battle of the Cities – Report

This report was prepared to complete the IBM Data Science Capstone Project requirements and offered online by Coursera. The project utilizes a similar approach to that of the Battle of the neighborhoods (NEIs) insofar as to utilize similar information to that collect in the labs and projects completed in the certification. These sources are publicly available GIS data sources, and Foursquare location based points of interest.

1. Introduction and Business Problem

I live and work in California but I "play" in Colorado, Utah, and Arizona, ie , every time I want to have some fun I drive all the way to some remote location in Colorado or Utah and go for hikes in Bryce Canyon, Arches, Zion, Capitol Reef, Grand Canyon, or one of the many awesome National Parks in any of these states.

It is always summer in California and it's great from a weather's point of view, if you like it but there's no seasons. We pay high taxes in almost every aspect of our lives – sales, car, home, state, business, you name it we have it and it's always high taxes.

When both my wife and I retire, I want to move to a state that has all the fun things we like to do, but a state that provides a much better cost of living, much better outdoor activities, low on pollution, has well defined seasons year-round, it is great for foodies, and we don't have to drive for miles just to get a carton of milk.

Telluride is nestled between mountains and has a great yearly film festival attended by famous actors and actresses, who also own their homes outside town (NOTE: the median home price in Telluride is $902,000, likely skewed by famous homes sold at over $10million!!)

Telluride has winter resorts for locals and visitors alike and it's extremely popular as a winter destination.

Telluride still has the old-west feeling, while catering to people who want to retire in a modern town with small businesses, and amenities that support a great community year round.

In the battle of the NEIs we analyzed points of interest between Toronto and NY. These are big cities, with plenty of information available on FourSquare. How will FourSquare handle information not at a neighborhood level but at a city or town level. How will it handle API calls that pull data based on my personal criteria (weather, food, supermarkets, outdoor fun, and trails)?

Once this project is completed, I expect to confirm my assertion about Telluride based on my personal criteria.

1.1. Problem Description

Identifying a town in the US to retire is a daunting task for anyone, and more so if you don't have the right resources available to you. The retiring process is complex and it has many, many pitfalls, related to location, amenities, costs of living, medical facilities, and so on.

Availability of data and research are crucial in identifying a state and a town to retire and provide a smooth transition from active careers to a retirement phase of our lives.

The initial problem to resolve in this phase of the project is to identify the public data sources that have enough information to collate with Foursquare and analyze using one of several machine learning approaches.

Even data from universities is not bullet-proof and during my initial research I found several data points that were not available or had to be highly curated before I could use them (or not at all).

The criteria used to select the best town to retire might not be appropriate for every situation. For my use case I first had to check Foursquare information and verify if data for at least a small group of towns in Colorado was available both in Foursquare and in a GIS JSON file of Colorado towns.

The Weather data, although provided by NOAA and displayed in Wikipedia, is not used at all by some towns. This is primarily because NOAA does not have weather stations to collect data at certain locations in the US. To track the weather in small towns in the US one has to rely on aggregated weather information from larger cities (eg, Boulder, CO). The problem with this approach is that weather in mountain towns is often localized and it expresses microclimates that are not captured by weather stations.

Since weather was one of the major categories in my selection criteria to identify a town for retirement, several interesting cities could not be include in the analysis and recommendations because of a lack of sufficient weather data.

Finally, Foursquare is a site that relies on location data to show interesting points in a map, but it relies heavily on availability of wireless connectivity and precise cell phone signal and that's not always available in remote areas of Colorado. So, if I want to retire to Telluride, CO, known as a small true-american town nestled between mountain peaks, far from civilization, good luck to me finding data to analyze the assertion that Telluride is the best overall town to retire in Colorado.

1.2. Audience

Anyone who wants to retire in Colorado based on similar criteria used to analyze Colorado cities and communities.

Although I'm working on this project for my own purposes, anyone who comes from a state with high income taxes, and high property taxes, will benefit from this research.

Additionally, it will be easy to change the criteria based on other people's preferences and as long as the criteria exists in the data collect from public sources for Colorado cities.

1.3. Success Criteria

The success criteria of the project will confirm or invalidate the assertion that Telluride, CO is one of the best towns in Colorado to retire based on the preferences highlighted in the problem description. Although Telluride, CO is assumed to be the best town to retire, we need to be open to the fact that there are other more viable options based on the same criteria defined in the problem description.

2. Data

Data will be loaded from publicly available GIS data sets maintained by Stanford University, Harvard University, and Colorado State GIS Repository.

Note that the use of this research and the use of each site mentioned in this report is not a sponsorship of the website, the company or services offered there, and I did not receive any sponsorship from anyone to do this analysis.

Foursquare items of interest

We will look at cities in Colorado and compare key elements to help decide where to retire:

- Weather (NOAA / Wikipedia)
- Food (4Square)
- Trails (4Square)
- Supermarkets (4Square)
- Outdoor fun (4Square)

Websites used for initial research

My initial assumption was that Telluride would be a great place to retire and as such I started researching Telluride, CO, using the **Best Places**

This site offers simple analysis on each place population, unemployment rate, median income, median age, median home price, and comfort index. It also provides comparisons between two places on costs of living, real estate, crime, climate, schools, economy, health, religion, politics, commute, and so on.

- ASSERTION: Preferred City to Retire
    - Telluride: https://www.bestplaces.net/cost_of_living/city/colorado/telluride

I also used the site SmartAsset to find out additional information about Colorado and specifically reasons to move to Colorado (https://smartasset.com/mortgage/15-things-to-know-before-moving-to-colorado). Note that I did not use this site directly in my data analysis but rather to ground the research on specific cities and aspects that I wanted to analyze for each town in Colorado.

The normalization and scope of the project was defined by a bit more research using the site SmartAsset to identify the top 10 cities to retire in Colorado (https://smartasset.com/retirement/best-places-to-retire-in-colorado)

Although I started with 10 cities I ended up with only 7 cities since some cities did not have enough data on weather, or population demographics or failed to report on one of the categories listed in my criteria.

List of cities used to compare NEIs and that fit retirement criteria

Sorted by ascending order of overall quality (as defined in the SmartAsset site):

- 1 - Littleton: https://www.bestplaces.net/city/colorado/littleton
- 2 - Englewood: https://www.bestplaces.net/city/colorado/englewood
- 3 - Evergreen: https://www.bestplaces.net/city/colorado/evergreen
- 5 - Estes Park: https://www.bestplaces.net/city/colorado/estes_park
- 6 - Monument: https://www.bestplaces.net/city/colorado/monument
- 8 - Salida: https://www.bestplaces.net/city/colorado/salida
- 10 - Montrose: https://www.bestplaces.net/city/colorado/montrose

Available Weather data in Wikipedia:

- 0 - Telluride: https://en.wikipedia.org/wiki/Telluride,_Colorado
- 1 - Littleton: https://en.wikipedia.org/wiki/Littleton,_Colorado
- 2 - Englewood: https://en.wikipedia.org/wiki/Englewood,_Colorado
- 3 - Evergreen: https://en.wikipedia.org/wiki/Evergreen,_Colorado
- 5 - Estes Park: https://en.wikipedia.org/wiki/Estes_Park,_Colorado
- 6 - Monument: https://en.wikipedia.org/wiki/Monument,_Colorado
- 8 - Salida: https://en.wikipedia.org/wiki/Salida,_Colorado
- 10 - Montrose: https://en.wikipedia.org/wiki/Montrose,_Colorado

Each location or city listed above contains weather data from NOAA and listed on Wikipedia as below:

Climate data for Telluride 4WNW 1981-2010, extremes 1900 to present [hide]

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Record high °F (°C)	58 (14)	65 (18)	73 (23)	78 (26)	83 (28)	91 (33)	96 (36)	91 (33)	88 (31)	83 (28)	73 (23)	60 (16)	96 (36)
Average high °F (°C)	37.5 (3.1)	39.5 (4.2)	44.5 (6.9)	52.9 (11.6)	62.6 (17.0)	73.2 (22.9)	77.5 (25.3)	74.7 (23.7)	68.1 (20.1)	58.5 (14.7)	45.5 (7.5)	37.2 (2.9)	56.0 (13.3)
Average low °F (°C)	4.5 (−15.3)	8.9 (−12.8)	15.7 (−9.1)	23.7 (−4.6)	30.8 (−0.7)	36.1 (2.3)	42.2 (5.7)	42.3 (5.7)	35.2 (1.8)	25.6 (−3.6)	15.5 (−9.2)	5.8 (−14.6)	23.9 (−4.5)
Record low °F (°C)	−32 (−36)	−36 (−38)	−24 (−31)	−10 (−23)	3 (−16)	15 (−9)	26 (−3)	22 (−6)	9 (−13)	0 (−18)	−22 (−30)	−31 (−35)	−36 (−38)
Average precipitation inches (mm)	1.49 (38)	1.58 (40)	2.02 (51)	1.94 (49)	1.81 (46)	1.12 (28)	2.50 (64)	3.02 (77)	2.68 (68)	1.92 (49)	1.87 (47)	1.46 (37)	23.41 (594)
Average snowfall inches (cm)	27.7 (70)	26.3 (67)	34.7 (88)	20.0 (51)	8.3 (21)	0.4 (1.0)	0 (0)	0 (0)	1.0 (2.5)	8.6 (22)	24.4 (62)	25.6 (65)	177 (449.5)
Average precipitation days (≥ 0.01 in)	10.0	10.1	11.8	10.5	8.9	6.9	13.5	16.6	12.2	9.3	8.9	10.0	128.7
Average snowy days (≥ 0.1 in)	9.8	9.4	10.4	6.8	2.3	0.1	0	0	0.4	2.9	7.2	9.7	59

Source: Nowdata, Telluride 4WNW from Grand Junction area

Publicly available GIS Data Sources

There are several sources of GIS information for each state in the US but I prefer to use the following data sources:

- Cities and Towns of the United States, 2014
  - https://geo.nyu.edu/catalog/stanford-bx729wr3020
  - Includes a downloadable JSON data source with coordinates for each city in the united states as of 2014

- Colorado Department of Local Affairs
  - https://demography.dola.colorado.gov/gis/gis-data/#gis-data
  - Includes different aspects of each county, borders, districts, and locations for each city. Updated often.

- US Department of Commerce – Colorado View
  - https://www.coloradoview.org/united-states-gis/
  - Lists several census data sources organized in different types of data files, GIS, ShapeFiles, Raster files, and others

- New York University, Spatial Data Repository
  - https://geo.nyu.edu/?f%5Bdc_rights_s%5D%5B%5D=Public&f%5Bdct_spatial_sm%5D%5B%5D=Colorado
  - NYU maintains a repo of spatial data aggregated from other universities and state departments. The data can be pulled in different formats, including JSON, shape, raster, or TIFF.

3. Methodology

The main goal is to confirm the assertion about Telluride, CO being one of the best cities to retire in Colorado.

Exploratory Data Analysis Approach

The following steps were taken to explore the publicly available datasets and identify characteristics (or features) that need further analysis.

- Get a list of all cities in the US
- Create a new dataframe with just the cities in colorado
- Find Colorado on the map
- Position Telluride and all cities in Colorado
- Add markers for all other cities listed as best cities to retire to in Colorado
- Sort the cities by population in ascending order
- Add all the points of interest to the map
- Weather Analysis
- Start analyzing temperature data for Telluride
- Start analyzing Precipitation data for Telluride
- Play with Seaborn library
- Get top 100 points of interest in Telluride within a radius of 15 miles
- Finding points of interest based on my criteria
- Create a function to analyze all cities of interest in Colorado
- Analyze unique categories that can be found from all the returned venues
- Analyze each City - ONE HOT Analysis
- Group rows by city and take the mean of the frequency of occurrences of each category
- Begin the CLUSTERING Process using k-means
- Create a new dataframe that includes the cluster as well as the top 10 venues for each city
- Visualize clusters
- Examine Clusters

Colorado Data

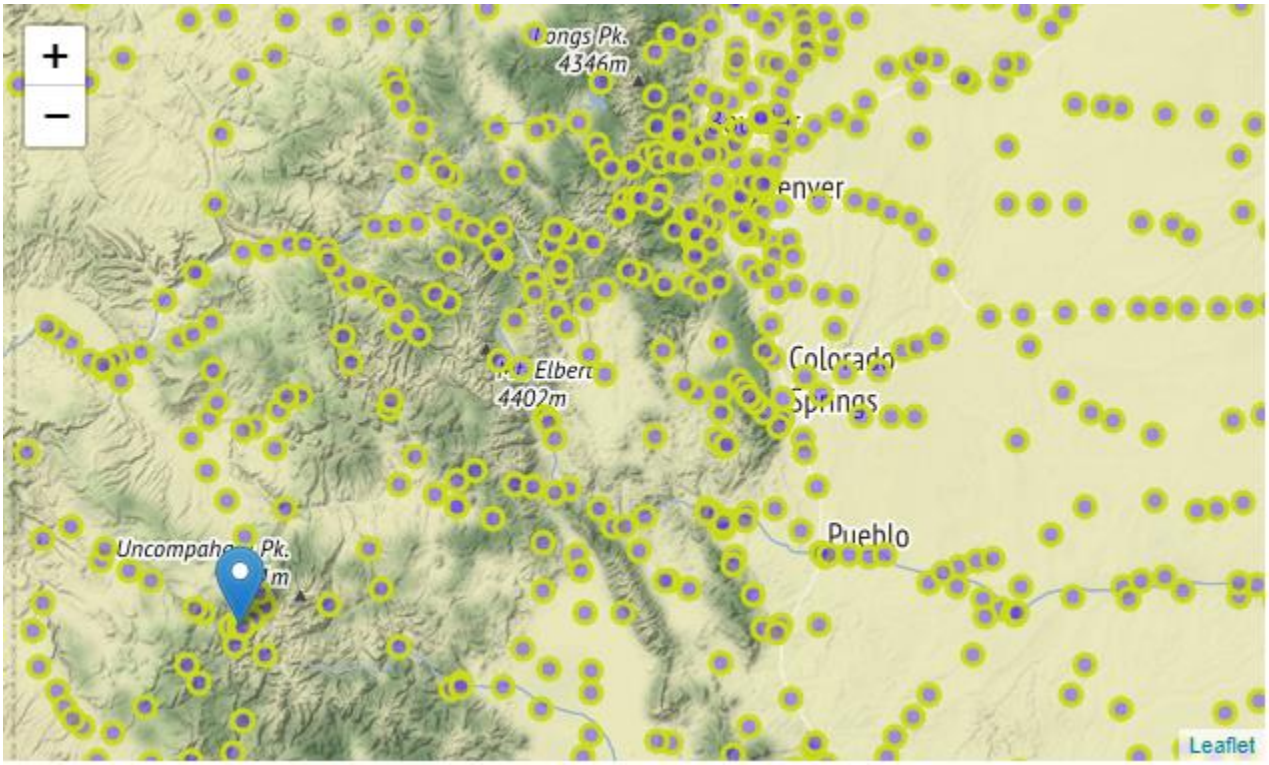
The original dataset contains all cities in the US, and thus had to be trimmed to contain only the cities in Colorado.

[14]:

	City	State	Population	County	Fips	Latitude	Longitude
17496	Atwood	CO	133	Logan	075	40.547762	-103.269657
17498	Laporte	CO	2450	Larimer	069	40.626371	-105.137758
17526	Allenspark	CO	528	Boulder	013	40.194429	-105.525555
17527	Eldora	CO	142	Boulder	013	39.948598	-105.563889
17528	Eldorado Springs	CO	585	Boulder	013	39.932486	-105.276935
...	...	...	...	...	...	...	...
38161	Mineral Hot Springs	CO	-999	Saguache	109	38.173661	-105.925413
38162	Lawson	CO	-999	Clear Creek	019	39.764162	-105.628844
38163	Dumont	CO	-999	Clear Creek	019	39.764337	-105.598796
38164	El Rancho	CO	-999	Jefferson	059	39.699880	-105.331404
38177	Doyleville	CO	-999	Gunnison	051	38.451721	-106.609497

591 rows × 7 columns

Once the cities in Colorado were identified and we created a dataframe with the results it was easy to reduce the scope of the project to only those cities that I was interested in, based on the “best cities to live in Colorado” research done in the first week of this project.



Based on the 7 cities under review we created a table that lists out all the retirement-ready cities in Colorado, sorted by population

[24]:

	index	City	State	Population	County	Fips	Latitude	Longitude
87	20214	Littleton	CO	41737	Arapahoe	005	39.613321	-105.016650
85	20212	Englewood	CO	30255	Arapahoe	005	39.647765	-104.987760
98	20252	Montrose	CO	19132	Montrose	085	38.478320	-107.876174
11	17539	Evergreen	CO	9038	Jefferson	059	39.633321	-105.317215
130	21383	Estes Park	CO	5858	Larimer	069	40.377206	-105.521665
267	22475	Monument	CO	5530	El Paso	041	39.091659	-104.872758
179	21474	Salida	CO	5236	Chaffee	015	38.534719	-105.998902
292	22537	Telluride	CO	2325	San Miguel	113	37.937494	-107.812285

Now the data looks much better and we can do further analysis on each city:

Weather Data Analysis

This is was the most painful, time consuming, and rewarding aspects of this project. Although pulling data from Wiki seemed easy to begin with using the BeautifulSoup4 package, data in Wikipedia is entered and edited in many different ways, and none of it is standard. In the future this portion of the analysis should be done using data from NOAA.

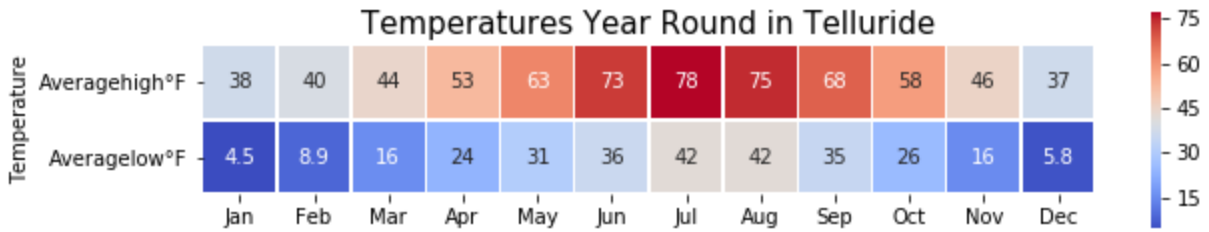
Parsing the HTML data include using Chrome Developer tools to identify where the table with the weather was, and pull the data into a JSON tree and eventually a CSV file for easy analysis.

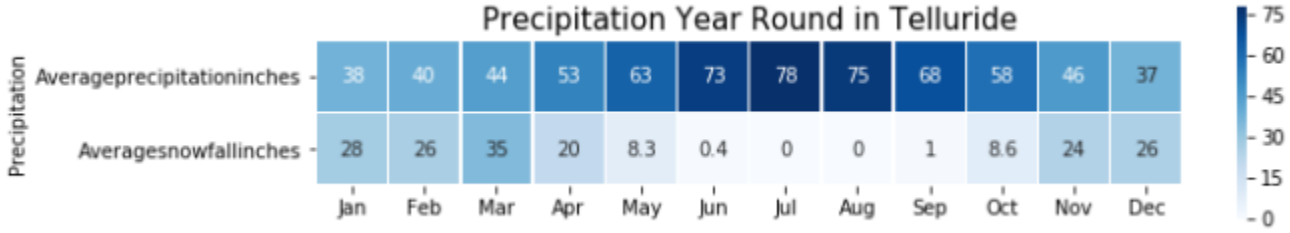
Before the data could be analyzed, we had to remove several extraneous characters (eg, \r\n and (**)) so that only the actual values of interest were used. I decided on just using Averages for each value, instead of the highs and lows for each feature.

But after so many data manipulations and cleaning the results are incredible.

[49]:

	Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	Averagehigh°F	37.5	39.5	44.5	52.9	62.6	73.2	77.5	74.7	68.1	58.5	45.5	37.2
2	Averagelow°F	4.5	8.9	15.7	23.7	30.8	36.1	42.2	42.3	35.2	25.6	15.5	5.8
4	Averageprecipitationinches	1.49	1.58	2.02	1.94	1.81	1.12	2.50	3.02	2.68	1.92	1.87	1.46
5	Averagesnowfallinches	27.7	26.3	34.7	20.0	8.3	0.4	0	0	1.0	8.6	24.4	25.6





Further analysis of Temperature and Precipitation for each of the contender cities should be done in the future and it is outside of the scope for this project.

Points of Interest in Telluride, CO

Using Foursquare to identify points of interest within a 15miles range from the city center (based on the Telluride latitude and longitude) resulted in 94 venues.

Further filtering the data based on my criteria resulted in 21 venues that represent Restaurants, Hiking Trails, Supermarkets and Outdoor fun, with corresponding name, latitudes and longitudes.

```
My criteria includes: Food/Restaurants, Hiking Trails, Supermarkets, and Outdoor Fun
```

	name	categories	lat	lng
0	Brown Dog Pizza	Pizza Place	37.937042	-107.810671
3	siam	Thai Restaurant	37.937935	-107.817829
5	Telluride Ski Resort	Ski Area	37.936505	-107.846016
18	Bridal Veil Falls	Outdoors & Recreation	37.928554	-107.776266
21	The Phoenix Bean	Café	37.937812	-107.812414
22	Station St. Sophia 10,540ft	Ski Area	37.931300	-107.833022
28	See Forever Trail	Trail	37.912764	-107.823630
35	Tracks Cafe & Bar	Pizza Place	37.936769	-107.846014
37	Telluride Main Street	Ski Area	37.937239	-107.811622
38	The Market at Mountain Village	Grocery Store	37.932439	-107.853891
46	Mountian Villiage Telluride	Ski Area	37.925297	-107.846884
47	Telluride Gondola Station (Station 1)	Ski Area	37.935985	-107.813518
48	Crazy Elk Pizza	Pizza Place	37.936541	-107.846594
49	High Pie Pizzeria & Taproom	Pizza Place	37.937119	-107.810927
53	Box Canyon Park	Trail	38.018060	-107.677608
55	Golden Block Brewery Pizza	Pizza Place	37.812175	-107.664309
62	Cascade Falls Park	Trail	38.025135	-107.666768
68	Silverton Mountain	Ski Area	37.811759	-107.664699
76	Brown Bear Cafe	Café	37.811210	-107.664912
82	Ice Falls	Trail	37.999940	-107.660825
83	Hurricane Pass	Trail	37.919340	-107.627940

Telluride is great but we need to compare it to other cities in the CO.

And so, based on the same criteria we created a table that lists similar venues in each city, creating a new dataframe with all the results merged.

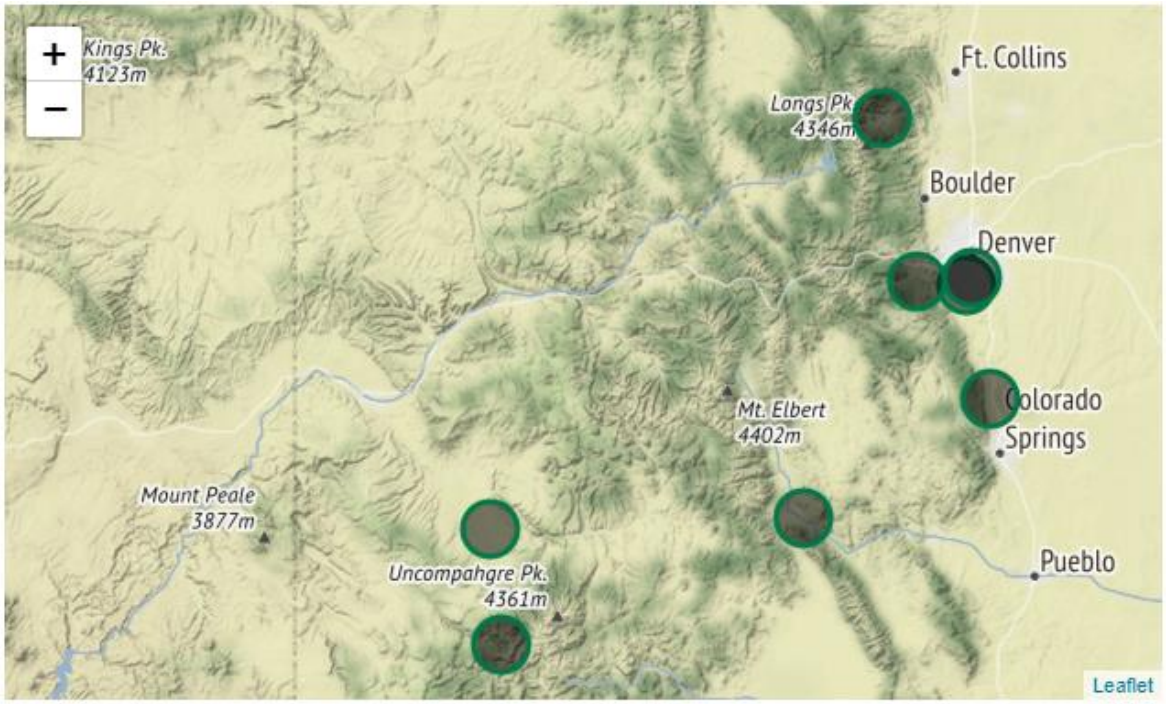
[64]:

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Telluride	37.937494	-107.812285	Brown Dog Pizza	37.937042	-107.810671	Pizza Place
1	Telluride	37.937494	-107.812285	Taco del Gnar - Telluride	37.937216	-107.813271	Taco Place
2	Telluride	37.937494	-107.812285	The Butcher & Baker Cafe	37.936921	-107.809715	Sandwich Place
3	Telluride	37.937494	-107.812285	La Marmotte	37.935731	-107.812275	French Restaurant
4	Telluride	37.937494	-107.812285	New Sheridan Hotel & Chop House	37.937689	-107.812670	Hotel
...	...	...	...	...	...	...	...
223	Montrose	38.478320	-107.876174	Sweet Bites Bakery	38.480149	-107.875576	Bakery
224	Montrose	38.478320	-107.876174	Sams Tavern	38.479927	-107.877550	Dive Bar
225	Montrose	38.478320	-107.876174	Elevate Salon	38.480019	-107.878447	Health & Beauty Service
226	Montrose	38.478320	-107.876174	Backstreet Bagel & Deli	38.479873	-107.879339	Bagel Shop
227	Montrose	38.478320	-107.876174	Crash Burger	38.481299	-107.873507	Burger Joint

228 rows × 7 columns

Because the data by this point was already highly massaged, filtered, and aggregated by interests, it was a straight forward process to use 1-hot analysis, followed by clustering using k-means with 5 clusters (in hindsight I probably should've used 3 or even 2 clusters to aggregate potential cities similar to Telluride in a "tighter" configuration)

The clustering process resulted in similar map as described above.



At this point it was clear that from all cities in Colorado there were a very few that had the right mix of characteristics and that were candidates for retiring.

City	Fips	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Telluride	113	37.937494	-107.812285	0	American Restaurant	Hotel	Bar	Pizza Place	Gift Shop
Littleton	005	39.613321	-105.016650	0	Restaurant	Gourmet Shop	Coffee Shop	Mexican Restaurant	Bar
Evergreen	059	39.633321	-105.317215	0	Bar	Restaurant	Candy Store	Breakfast Spot	Pizza Place
Estes Park	069	40.377206	-105.521665	0	Gift Shop	American Restaurant	Ice Cream Shop	Mexican Restaurant	Arts & Crafts Store

4. Results and Conclusion

Based on the results of different criteria (weather, population, restaurants, hiking trails, supermarkets, and outdoor activities) it is clear to see that Telluride is still a great town to retire too.

Of course, this analysis is biased towards Telluride and in the real world we would need to remove bias and include a much larger dataset (possibly ALL cities in Colorado) and a much larger set of features to analyze.

It is possible that using different clusters would also yield different results and point to a different conclusion.

The weather data and demographics for each city was extremely time consuming and in future analysis it should be included as a relevant characteristic in the cluster formation (eg, if it snows in Telluride on average x inches, how does that compare to other cities further east or further south).