

A person is seen from behind, standing on a mountain peak and looking out over a vast, hazy landscape under a warm, golden sky. The person is wearing a dark jacket and has their hands in their pockets. The landscape consists of rolling hills and valleys, with a thin layer of clouds or mist settling in the lower areas. The overall mood is serene and contemplative.

# CLASSIFICATION OF U.S. TOURIST TOWNS

David Sachs

Coursera IBM Data Science Capstone

December 20<sup>th</sup>, 2019



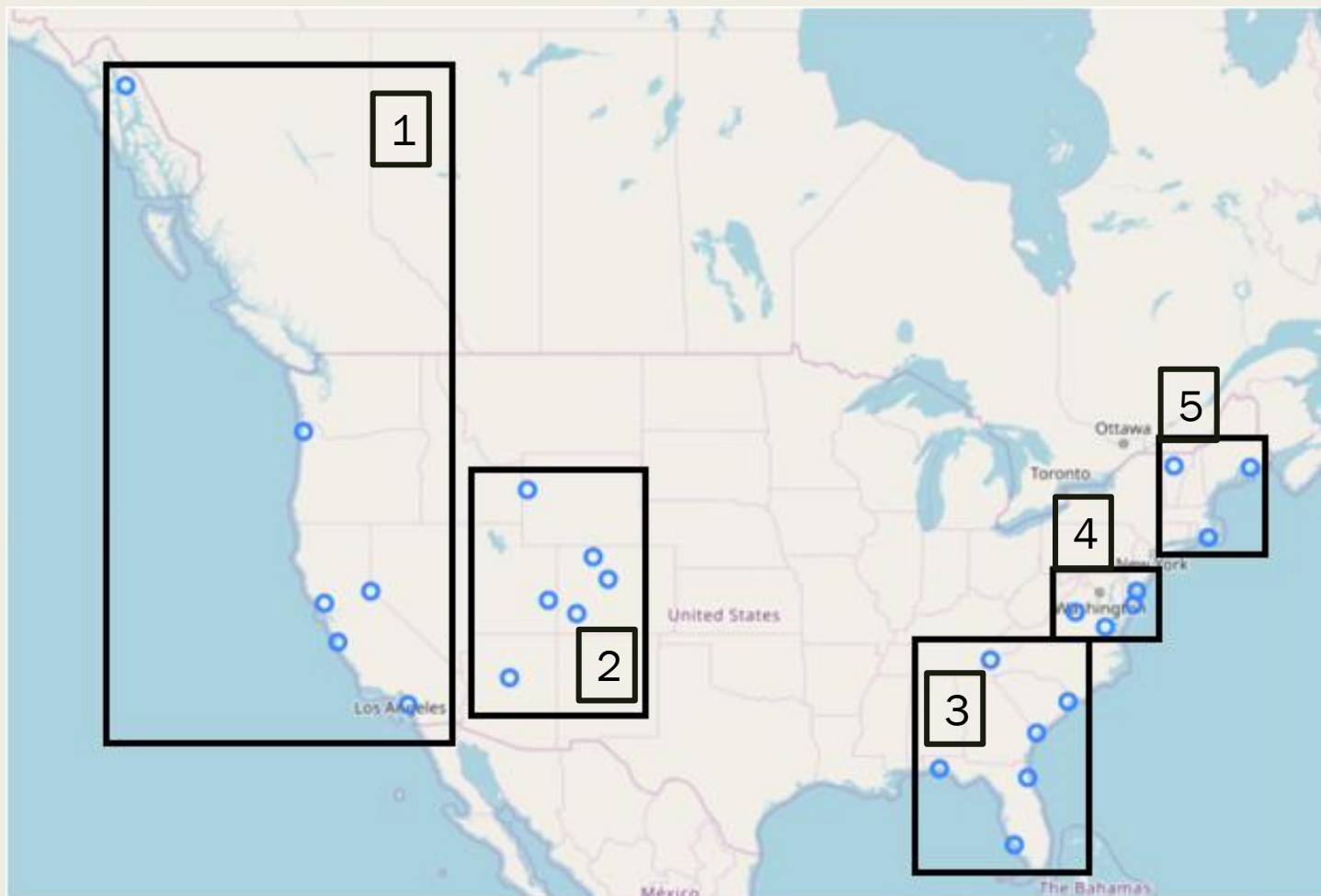
# Why does this matter?

- Tourism is a significant industry in the U.S., with small tourist towns often being completely dependent on the industry
- Understanding similar towns and categories of town is useful for travelers, as they can use this to select a trip that will give them the best experience possible
- The towns themselves have interest in this: if a town wishes to be seen a certain way, understanding how they are currently classified can help them create strategy
- Stores and chains may also be interested while looking for locations for expansion

- The sample locations were gathered from a list prepared by U.S. News, available [here](#)
- The location data for the towns, comprised of their latitude and longitude, was gathered using the Geopy Python Library
- FourSquare was used to find venues in each town
- In total, 25 towns and ~2400 venues were considered in this set
- Cleaning was done to ensure the correct towns and corresponding locations were found
- Cleaning was also done to remove ‘filler’ words that provide little analytical value
- Data sets were then created using one-hot encoding and were clustered using K-Means

## Data Acquisition and Cleaning





## Initial Data Exploration - Geography

- Without clustering, some initial groupings based purely on geography appear
- Expected Clusters:
  1. West Coast
  2. The Rocky Mountains
  3. The Southeast
  4. The Mid-Atlantic
  5. The Northeast
- Useful for setting some ideas for potential links

# Initial Data Exploration – Top 10s

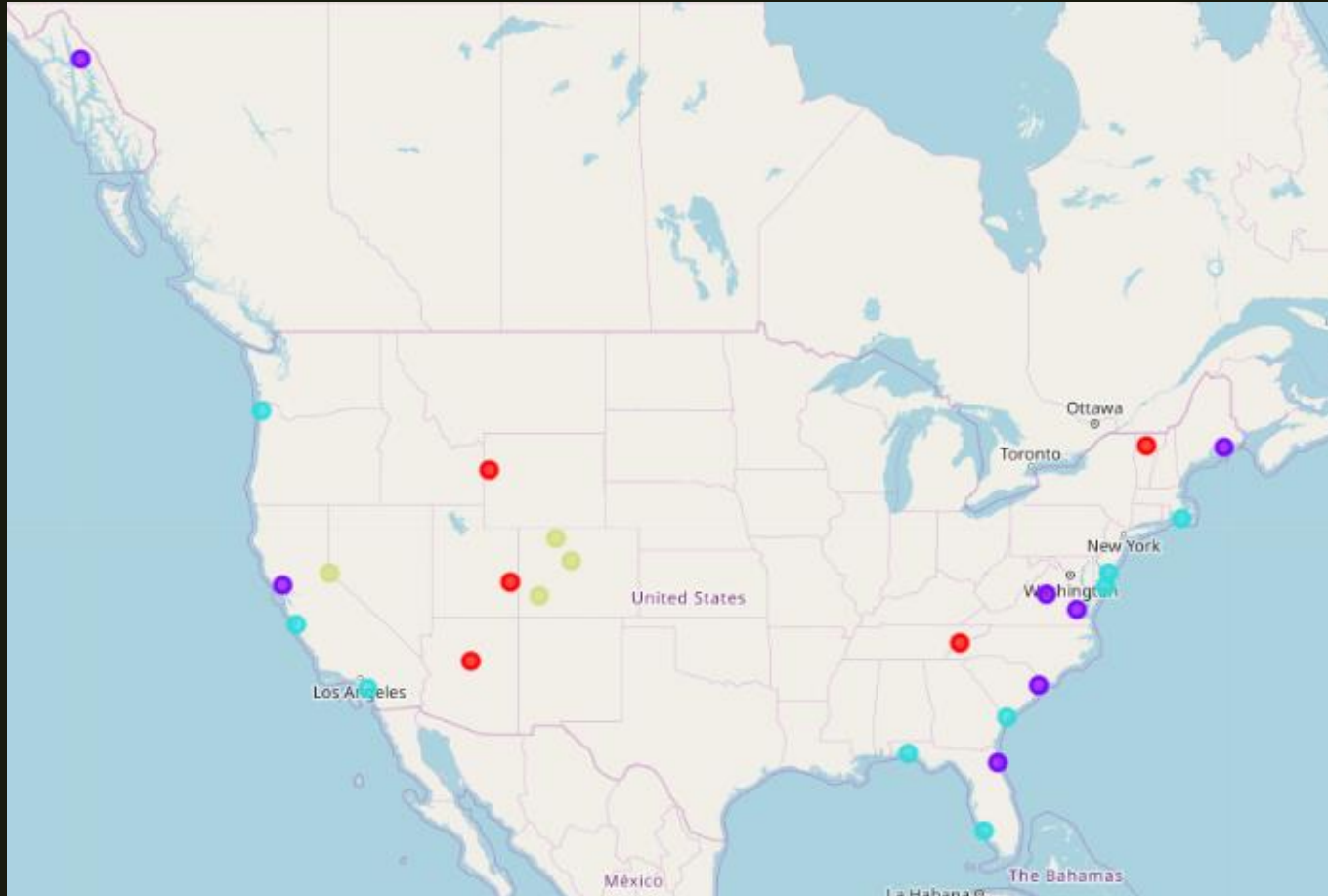
- Some patterns appear while reviewing the top towns
- Almost all towns shares certain words, and certain categories: hotels, restaurants, and bars
- Common words have an issue: many similar words to humans that are distinct to computers
- Tying together varied values may be difficult

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	spoon	station	mountain	gherkin	ledge	slight	amen	pizza	cafe	nest
Jackson	restaurant	inn	lston	pead	street	hotel	center	gill	pizza	bakery
South Lake Tahoe	cafe	heavenly	coffee	grill	dog	pizza	edgewood	hut	south	ledge
Sedona	rock	resort	trail	oreak	spa	park	oak	red	cafe	trainhead
Bar Harbor	park	acadia	cafe	hotel	lobster	trail	street	inn	cove	desert
Gatlinburg	smoky	resort	falls	mountain	ole	center	bear	great	pancake	cabin
Monterey	city	north	park	grill	park	market	house	flat	ch	old
Steamboat Springs	restaurant	grill	resort	oreak	express	charlft	brewery	cafe	mountain	stom
Williamsburg	shop	tavern	house	grill	market	squire	cafe	dog	palace	colonial
Moab	park	center	inn	suiles	arches	trail	cafe	arch	campground	national
Stowe	trapp	family	mountain	ledge	inn	market	cafe	resort	toproth	green
Juneau	alaska	alaskan	company	cafe	brewing	co	heritage	salmon	treadwell	adventure
Laguna Beach	park	grill	cove	market	crystal	pelican	bay	cafe	mortgage	also
Santa Rosa	coffee	market	company	brewing	company	cafe	street	center	restaurant	sturdy
Blackridge	peak	ledge	center	superhair	restaurant	blue	ski	summit	mountain	coffee
Bonibel Island	lighthouse	club	grill	restaurant	cafe	store	inn	resort	bay	marina
Martha's Vineyard	market	restaurant	farm	edgartson	west	view	state	house	cafe	inn
Hilton Head Island	resort	cafe	sea	palmetto	dunes	harbour	pine	shelter	park	cafe
St. Augustine	restaurant	company	house	cafe	seafood	old	museum	historic	wine	coffee
Destin	seafood	house	cafe	handerson	house	mexican	grill	report	home	starbucks
Myrtle Beach	resort	grill	broadsley	restaurant	house	inn	oceanfront	grand	red	nest
Ocean City	street	grill	restaurant	cafe	on	shop	fris	italian	center	pool
Cannon Beach	cafe	park	inn	restaurant	state	arch	coffee	company	lounge	skole
Charlottesville	brewery	downtown	kitchen	park	body's	bagels	coffee	monticello	center	bear
Cape May	restaurant	hotel	company	congress	house	sunset	wildwood	inn	washington	hall

## Towns with the Most Common Venue Words

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
Blackridge	St. Area	Coffee Shop	Hotel	St. Chant	American Restaurant	Bar	St. Lodge	Bear Garden	Restaurant	Pizza Place
Cannon Beach	Beach	Hotel	American Restaurant	Coffee Shop	Seafood Restaurant	Brewery	State / Provincial Park	Brewery	Restaurant	Trail
Cape May	American Restaurant	Beach	Seafood Restaurant	Hotel	Coffee Shop	Ice Cream Shop	Campground	Breakfast Spot	Italian Restaurant	Del / Bologna
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	BBO Joint	St. Chant	American Restaurant	Diner 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	BBO Joint	Breakfast Spot	Pizza Place
Juneau	Seafood Restaurant	Coffee Shop	Bar	Pizza Place	American Restaurant	Cafe	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	Cow	Ice Cream Shop	Hotel
Moab	Hotel	Trail	Campground	American Restaurant	National Park	Cafe	Hotel	Food Truck	Sporting Goods Shop	Mexican Restaurant
Monterey	Beach	Seafood Restaurant	Aquarium	Coffee Shop	Breakfast Spot	Grocery Store	Such 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	Cafe	Del / Bologna	Italian Restaurant
Sandibel Island	Beach	Seafood Restaurant	Hotel	American Restaurant	Trail	Harbor / Marina	Restaurant	Cafe	Bar	Pizza Place
Santa Rosa	Coffee Shop	Grocery Store	Brewery	The Restaurant	American Restaurant	Pizza Place	New American Restaurant	Cafe	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	St. Area	Mexican Restaurant	Pizza Place	Brewery	Resort
St. Augustine	American Restaurant	Seafood Restaurant	Pizza Place	Coolaid Bar	Pub	Music Venue	History Museum	Historic Site	Cafe	Brewery
Steamboat Springs	American Restaurant	St. Area	Brewery	Mexican Restaurant	Trail	Pub	Coffee Shop	Italian Restaurant	Steakhouse	Pizza Place
Stowe	Trail	American Restaurant	Brewery	Resort	Bar	New American Restaurant	Breakfast Spot	Bed & Breakfast	Gift Shop	Trail
Telluride	St. Area	Hotel	Bar	Pizza Place	Performing Arts Venue	Sandwich Place	Sports Bar	St. Chant	Trail	French Restaurant
Williamsburg	American Restaurant	History Museum	Seafood Restaurant	Hotel	Gift Shop	Historic Site	Clothing Store	Coffee Shop	Restaurant	Brewery

## Towns with the Most Common Venue Categories

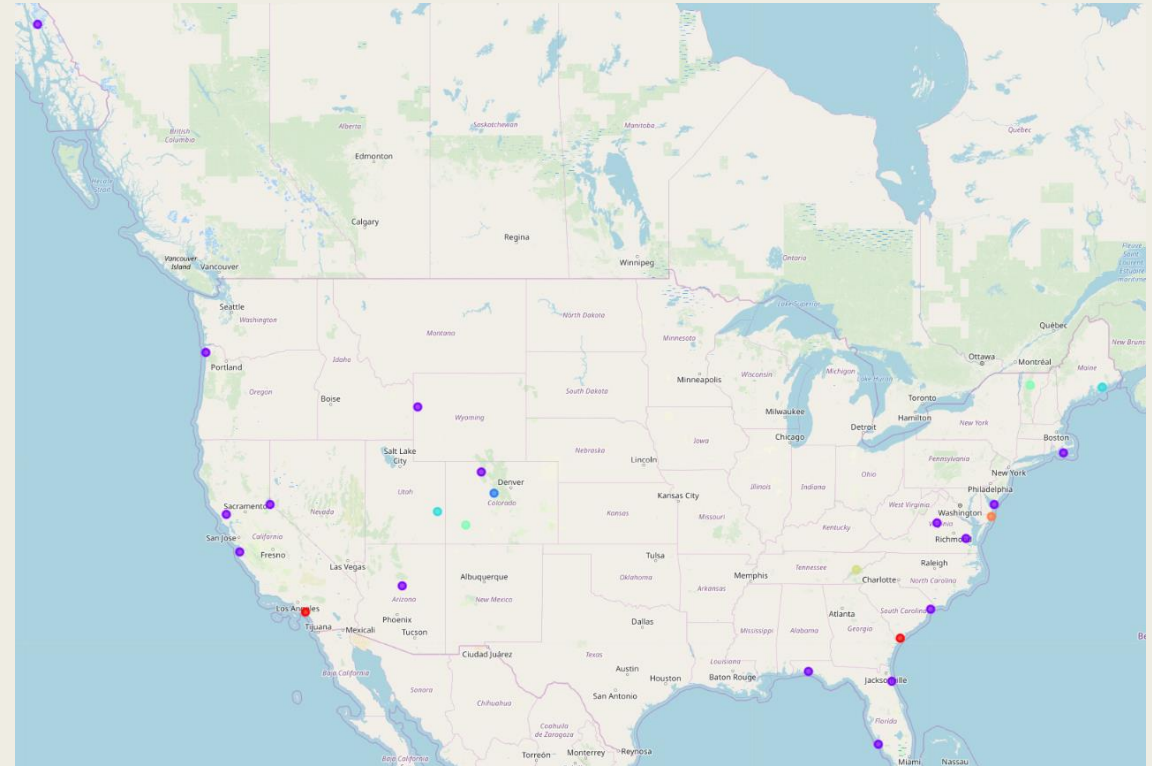


# Clustering via Most Common Venues

- Clustering was completed with levels of K from 3 to 8
- K = 4 proved to be the best result, with clear clusters as follows:
  - *Cluster 1: Ski Towns [Gold color]*
  - *Cluster 2: Mountain/Outdoor Towns [Red color]*
  - *Cluster 3: Beach Towns [Teal color]*
  - *Cluster 4: Come and Relax Towns [Purple color]*

# Clustering via Other Methods

- Clustering with other methods proved generally ineffective
- Clustering was done from  $K=3$  through  $K=8$
- No clear clusters formed
- Cluster accuracy was likely limited by the amount of noise: the most common words one-hot data set had nearly 600 features
- The merged data set, which comprised both the venues and most common words, was heavily skewed by the poor performance of the week common words dataset



- $K=4$  for Top Venue Categories ended up producing a model that met the original goal: clearly classify tourist towns
- This result proves that it is possible to simplify the tourism experience by helping people understand the town they are visiting, as well as the next town to visit
- There is still room for improvement: further classifying refinement could be done
- The above will require either new features from different sources or more data to help form clusters

## Results





# Conclusion and Future Direction

- The goal of the overall project was a success
- K-Means classification was successfully used
- Additional towns and features can be used to improve the model
- Word count can be refined using classification in itself: turn 20 low word count terms into a broader category
- Additional work would be to include other regions (EU, Asia)

