Exploration of Venues in the State Capitals of the USA

IBM Capstone Project

By Zeynep Akca

April 5, 2020

Introduction

In this final project, I will explore popular venues in 50 state capitals. I target travel agencies who aim to attract tourists to the US Capitals. The aim is to find popular spots in the state capitals. I also aim to both distinguish them in their uniqueness and find similarities among them. This will help agencies to guide their customers better. They can offer destinations diverse enough to satisfy their customers, so that customers would have a satisfying US experience.

I restricted the cities only to the capitals, because I thought that this would be a nice theme to attract tourists who want to do an unconventional US trip, rather than just visiting popular places only in well-known cities. Also, it is limited to downtown areas of the capitals. One of my aims is to capture ordinary life, so that I can explore the choices of locals in their respective cities.

In this project, I will cluster capitals according to the most common venues. And I will explore diverse capitals in detail to help travel agencies to recommend good places to their customers and come up with itineraries unique enough to satisfy customer needs.

Data

I will get the data from two different places. The first one is a github page where I scrape latitudes and longitudes of capitals. The second one is Foursquare data. From there, I will use the explore function to get venues around capitals. I will then get information regarding the venues. Mainly; ratings, number of likes, number of tips, number of photos, price category (ranging between 1 and 4). I will use these to have an estimate of popularity of venues. The reason I chose to use this is because these are the information that I can get from Foursquare with a regular account.

I limit the number of venues for each state to 20. They are located 1500 meters radius from the center of the capital. As you will see below not every city has that many venues. But I decided to keep the radius the same for every city in order to have a consistent comparison.

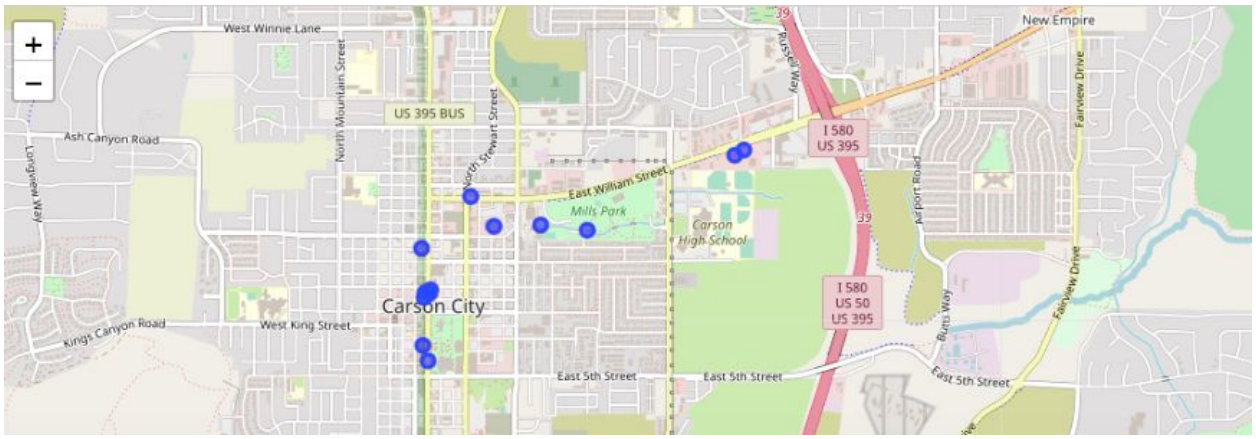
Below I show you can see the capitals and how many venues returned for each of them.

Capital	Capital Latitude	Capital Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Category
Albany	20	20	20	20	20	20	20
Annapolis	20	20	20	20	20	20	20
Atlanta	20	20	20	20	20	20	20
Augusta	20	20	20	20	20	20	20
Austin	20	20	20	20	20	20	20
Baton Rouge	20	20	20	20	20	20	20
Boise	20	20	20	20	20	20	20
Boston	20	20	20	20	20	20	20
Carson City	20	20	20	20	20	20	20
Charleston	20	20	20	20	20	20	20
Cheyenne	20	20	20	20	20	20	20
Columbia	20	20	20	20	20	20	20
Columbus	20	20	20	20	20	20	20
Concord	13	13	13	13	13	13	13
Denver	20	20	20	20	20	20	20
Des Moines	20	20	20	20	20	20	20
Dover	20	20	20	20	20	20	20
Frankfort	20	20	20	20	20	20	20
Harrisburg	20	20	20	20	20	20	20
Hartford	20	20	20	20	20	20	20
Helana	20	20	20	20	20	20	20
Honolulu	20	20	20	20	20	20	20
Indianapolis	20	20	20	20	20	20	20
Jackson	6	6	6	6	6	6	6
Jefferson City	20	20	20	20	20	20	20
Juneau	20	20	20	20	20	20	20

Lansing	20	20	20	20	20	20	20
Lincoln	20	20	20	20	20	20	20
Little Rock	20	20	20	20	20	20	20
Madison	20	20	20	20	20	20	20
Montgomery	20	20	20	20	20	20	20
Montpelier	20	20	20	20	20	20	20
Nashville	20	20	20	20	20	20	20
Oklahoma City	20	20	20	20	20	20	20
Olympia	20	20	20	20	20	20	20
Phoenix	20	20	20	20	20	20	20
Pierre	20	20	20	20	20	20	20
Providence	20	20	20	20	20	20	20
Raleigh	20	20	20	20	20	20	20
Richmond	20	20	20	20	20	20	20
Sacramento	20	20	20	20	20	20	20
Saint Paul	20	20	20	20	20	20	20
Salem	20	20	20	20	20	20	20
Salt Lake City	20	20	20	20	20	20	20
Santa Fe	20	20	20	20	20	20	20
Springfield	20	20	20	20	20	20	20
Tallahassee	20	20	20	20	20	20	20
Topeka	20	20	20	20	20	20	20
Trenton	20	20	20	20	20	20	20

An example of venues in Carson City

	Capital	Capital Latitude	Capital Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Category
526	Carson City	39.160949	-119.753877	Paul Schat's Bakery	4b92881cf964a520970134e3	39.156239	-119.765484	Bakery
527	Carson City	39.160949	-119.753877	Carson City Aquatic Facility	4bfaf977bbb7c92873830743	39.169057	-119.759542	Gym Pool
528	Carson City	39.160949	-119.753877	Dutch Bros. Coffee	4c6d42d96af58cfacae58817	39.156386	-119.766733	Coffee Shop
529	Carson City	39.160949	-119.753877	Sportsman's Warehouse	5339acbd498e75abcda555c8	39.154992	-119.765967	Outdoor Supply Store
530	Carson City	39.160949	-119.753877	Comma Coffee	4b844baff964a5200a2d31e3	39.162055	-119.767024	Coffee Shop



In total, I ended up with 956 venues from 49 sata capitals (Excluding Bismarck since no venues returned from Foursquare pull). Below an example from the dataset I used for the analysis.

	Capital	Venue ID	Venue Name	Categories	Rating	# of Likes	# of Tips	# of Photos	Price Category	Latitude	Longitude
0	Montgomery	4dcabad652b1c2222a89cc50	Shashy's Bakery & Fine Foods	Bakery	8.4	21	9	7	1.0	32.362289	-86.283226
1	Montgomery	4b9fc51af964a520ff3c37e3	Martin's Restaurant	Fried Chicken Joint	7.9	19	11	26	1.0	32.357262	-86.282862
2	Montgomery	4b50c761f964a520133227e3	Chick-fil-A	Fast Food Restaurant	9.0	57	21	21	1.0	32.368860	-86.270454
3	Montgomery	4bb234ec35f0c9b6b3f4ba83	Zaxby's Chicken Fingers & Buffalo Wings	Fried Chicken Joint	8.2	24	6	15	2.0	32.364406	-86.268742
4	Montgomery	4ba2d789f964a520721d38e3	Subway	Sandwich Place	7.5	4	1	1	1.0	32.357502	-86.283664
5	Montgomery	4bad3674f964a520d3393be3	La Zona Rosa	Mexican Restaurant	8.5	52	29	57	2.0	32.359185	-86.265233
6	Montgomery	4da855c293a021ab13c6d644	Midtown Pizza Kitchen	Pizza Place	8.7	45	29	118	2.0	32.357174	-86.265484
7	Montgomery	576b2c85cd10f350e1b9006b	First Watch - Montgomery	Breakfast Spot	8.7	18	5	10	1.0	32.357405	-86.265341

Here we ended up with 49 state capitals where most of them have 20 venues.

Each venue has details regarding ratings, number of likes, number of tips, number of photos, and price category (1-4).

There are 20 ratings and 349 price categories missing in our data set.

Methodology

In this project, I will limit the analysis to this dataset above (appended_data).

In the initial step, I will try to cluster capitals, this will help to see which capitals are similar and which capitals are different. I will first create a list of most common venue categories in each capital and then cluster them. I will use k means and elbow methods to do clustering.

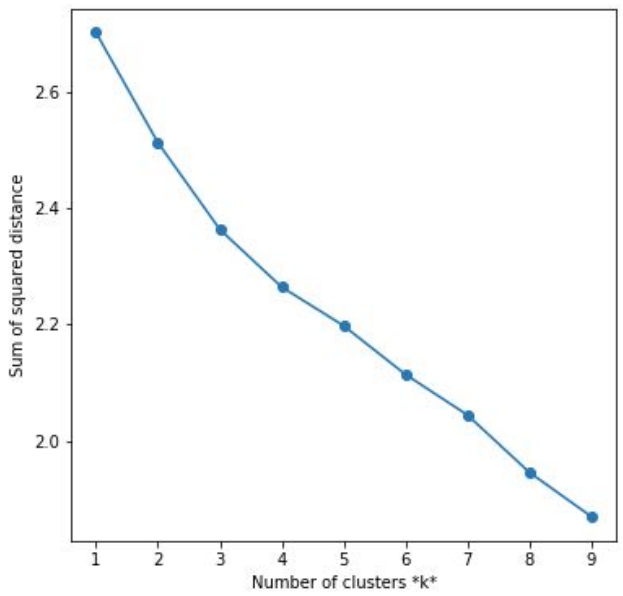
As a second step, I will do in depth exploration of capitals by various factors such as ratings, number of likes etc. I will compare them and try to see if there is a significant difference between the coasts and capitals in general

Analysis

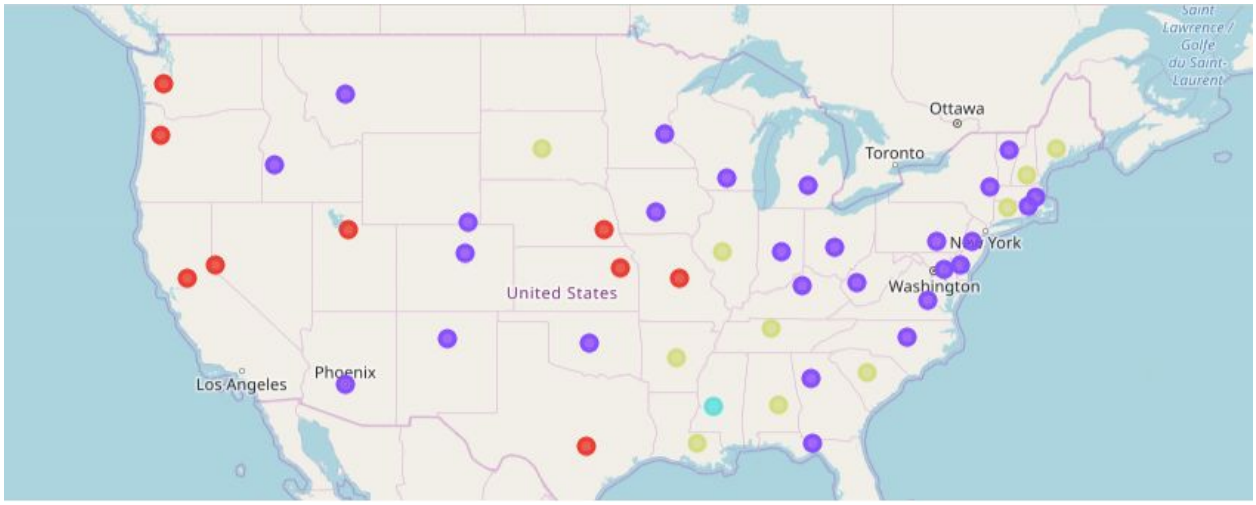
Clustering (Step 1)

First I got the most common venues for each capital.

	Capital	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Albany	Latin American Restaurant	Burger Joint	Sandwich Place	Bar	Sushi Restaurant
1	Annapolis	Bar	Thai Restaurant	Spa	French Restaurant	Gastropub
2	Atlanta	Hotel Bar	Music Venue	Cajun / Creole Restaurant	History Museum	Poke Place
3	Augusta	Pharmacy	American Restaurant	Convenience Store	History Museum	Thai Restaurant
4	Austin	Coffee Shop	Mexican Restaurant	Garden	Dance Studio	Restaurant



As you can see from the left hand side, the elbow method did not give a clear cutoff point. After I tried different k's,
I decided to proceed with k=4, the reason is it is the number of cluster
that shows the clear distinction between the coasts.
I thought that I can reach meaningful conclusions from here to present travel agencies.



Cluster 1 (Red circles on the map)

	Capital	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Sacramento	Coffee Shop	Vietnamese Restaurant	Pet Store	Café	Marijuana Dispensary
10	Honolulu	Japanese Restaurant	Coffee Shop	Scenic Lookout	Mexican Restaurant	State / Provincial Park
15	Topeka	Mexican Restaurant	Coffee Shop	Indian Restaurant	Gym	Breakfast Spot
24	Jefferson City	Pizza Place	Coffee Shop	Mexican Restaurant	Pub	Sandwich Place
26	Lincoln	Mexican Restaurant	Park	Coffee Shop	Zoo	Garden
27	Carson City	Coffee Shop	Thai Restaurant	History Museum	Park	Mexican Restaurant
36	Salem	Coffee Shop	American Restaurant	Theater	State / Provincial Park	Farmers Market
42	Austin	Coffee Shop	Mexican Restaurant	Garden	Dance Studio	Restaurant
43	Salt Lake City	Coffee Shop	Food Stand	Hotel	Food Truck	Burger Joint
46	Olympia	Coffee Shop	Sandwich Place	Park	Brewery	Diner

Cluster 2 (Purple circles on the map)

	Capital	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Juneau	Bar	Coffee Shop	Food Truck	Taco Place	Russian Restaurant
2	Phoenix	Salon / Barbershop	Pub	Theater	Opera House	Breakfast Spot
5	Denver	Yoga Studio	Breakfast Spot	History Museum	Jewelry Store	Park
7	Dover	Pizza Place	BBQ Joint	Flea Market	Monument / Landmark	Seafood Restaurant
8	Tallahassee	Cosmetics Shop	New American Restaurant	Japanese Restaurant	Beer Bar	Cocktail Bar
9	Atlanta	Hotel Bar	Music Venue	Cajun / Creole Restaurant	History Museum	Poke Place
11	Boise	Pub	Chinese Restaurant	Seafood Restaurant	Park	Dive Bar
13	Indianapolis	Brewery	Yoga Studio	Gay Bar	New American Restaurant	Advertising Agency
14	Des Moines	Hotel	Music Venue	Tapas Restaurant	Café	Skating Rink
16	Frankfort	Pizza Place	History Museum	Café	Food Court	Bookstore
19	Annapolis	Bar	Thai Restaurant	Spa	French Restaurant	Gastropub
20	Boston	Breakfast Spot	Clothing Store	Bakery	Donut Shop	Coffee Shop
21	Lansing	Bar	Bakery	Snack Place	Soup Place	Burger Joint
22	Saint Paul	Pizza Place	Brewery	Coffee Shop	Southern / Soul Food Restaurant	Farmers Market
25	Helana	Coffee Shop	Café	Dessert Shop	Taco Place	Pharmacy
29	Trenton	Art Gallery	Bar	Pizza Place	Italian Restaurant	Caribbean Restaurant
30	Santa Fe	Pizza Place	Mexican Restaurant	New American Restaurant	Brewery	Automotive Shop
31	Albany	Latin American Restaurant	Burger Joint	Sandwich Place	Bar	Sushi Restaurant
32	Raleigh	Performing Arts Venue	Sushi Restaurant	Southern / Soul Food Restaurant	BBQ Joint	Cocktail Bar
34	Columbus	Park	Coffee Shop	Theater	Café	Capitol Building
35	Oklahoma City	Bar	Ice Cream Shop	Ramen Restaurant	Sushi Restaurant	Pizza Place
37	Harrisburg	American Restaurant	Brewery	Italian Restaurant	French Restaurant	Speakeasy
38	Providence	Italian Restaurant	Gourmet Shop	Hockey Arena	Theater	Dessert Shop
44	Montpelier	Thai Restaurant	Bar	Italian Restaurant	Steakhouse	Taco Place
45	Richmond	American Restaurant	Dessert Shop	Italian Restaurant	Pizza Place	Tea Room
47	Charleston	Pizza Place	Farmers Market	Bakery	Park	Sandwich Place
48	Madison	Pizza Place	Café	Gastropub	American Restaurant	Italian Restaurant
49	Cheyenne	Pizza Place	Ice Cream Shop	Diner	Mexican Restaurant	Italian Restaurant

Cluster 3 (Blue circle on the map)

	Capital	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
23	Jackson	Zoo	Sandwich Place	Moving Target	Fast Food Restaurant	BBQ Joint

Cluster 4 (Yellow circles on the map)

	Capital	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Montgomery	Fried Chicken Joint	Fast Food Restaurant	Sandwich Place	Mexican Restaurant	Breakfast Spot
3	Little Rock	Hotel	Mexican Restaurant	Pharmacy	Zoo	Bank
6	Hartford	Theater	American Restaurant	Boutique	Mexican Restaurant	Sandwich Place
12	Springfield	Pharmacy	Restaurant	Diner	Italian Restaurant	Donut Shop
17	Baton Rouge	Gas Station	Discount Store	Convenience Store	Grocery Store	Fast Food Restaurant
18	Augusta	Pharmacy	American Restaurant	Convenience Store	History Museum	Thai Restaurant
28	Concord	Science Museum	Sandwich Place	Pizza Place	Sporting Goods Shop	Storage Facility
39	Columbia	American Restaurant	Seafood Restaurant	Pizza Place	Bakery	Burger Joint
40	Pierre	Mobile Phone Shop	Pizza Place	Liquor Store	Restaurant	Movie Theater
41	Nashville	Hotel	Music Venue	Sandwich Place	Park	Steakhouse

Even if we did not get a clear cutoff point from elbow methow, clustering by 4 shows that there is a clear distinction between the coasts. Before diving into that, we should be aware of the fact that there is only one capital in the third cluster. The reason might be that there are only 6 venues returned for Jackson and all of them belong to different venue categories, therefore the order does not reflect the most common venues in the city. And the k-means could not put it into any other cluster. The fourth cluster looks different than the first two but it does not have a distinct feature as the other two have.

From the map it is clear that there is a difference between the West and East Coasts of the country. The West Coast capitals have concentrated more on coffee shops and the East Coast capitals more pizza places and bars.

Comparison of the Two Coasts (Step 2)

In order to see if there is a significant difference between how people engage in places on the coasts, I will compare them by various factors.

I will compare coffee shops, pizza places and bars on two coasts. Since this analysis is based on coasts, I decided to exclude certain capitals. The capital which is included to the Cluster 1 but location-wise counted as the east (based on the middle meridian); Jefferson, and the capitals which are included to the Cluster 2 but location-wise counted as the west; Boise, Helane, Cheyenne, Phoenix, Santa Fe, Oklahoma City, and Denver are excluded from the analysis.

Coffee Shops

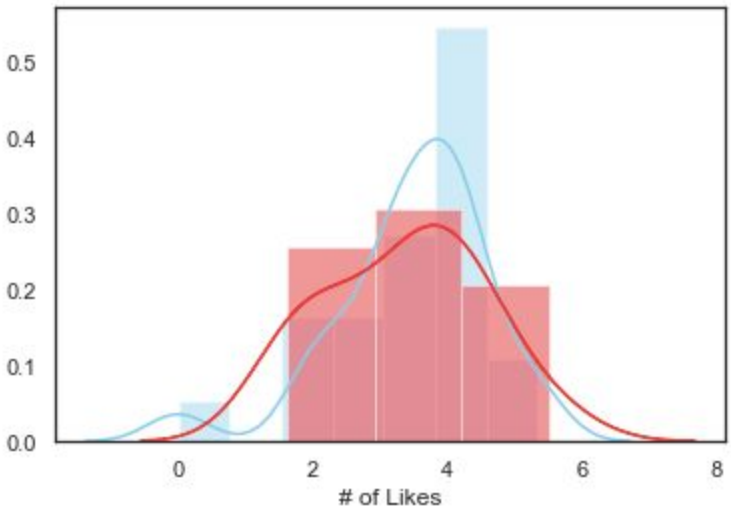
Firstly, I got descriptive stats for coffee shops in each coast. There are 24 coffee shops in the west, and 15 in the east. At the first glance, means and stds for each factor are pretty similar to each other except the number of photos.

West Coast

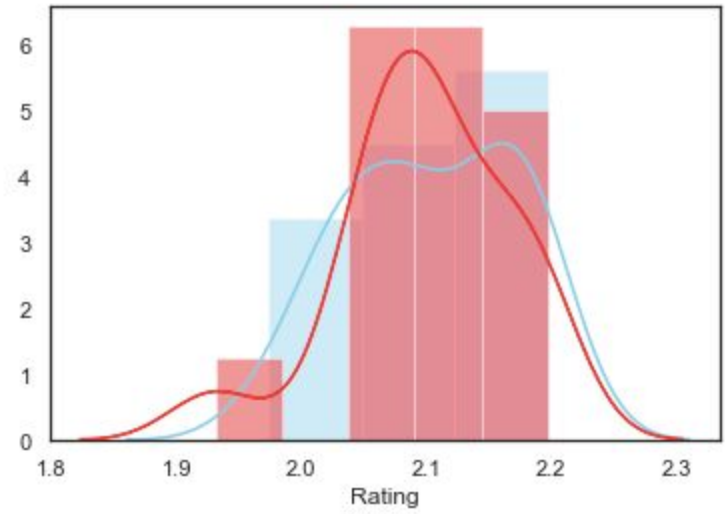
	Rating	# of Likes	# of Tips	# of Photos	Price Category	Latitude	Longitude
count	24.000000	24.000000	24.000000	24.000000	24.000000	24.000000	24.000000
mean	8.225000	51.958333	20.125000	89.416667	1.333333	39.500645	-118.588800
std	0.550296	50.341857	21.297709	99.488656	0.481543	7.268677	16.364630
min	7.200000	1.000000	0.000000	0.000000	1.000000	21.298552	-157.838272
25%	7.775000	20.750000	4.750000	15.750000	1.000000	38.561583	-123.021490
50%	8.150000	39.500000	14.500000	49.500000	1.000000	39.171711	-121.471916
75%	8.725000	62.500000	25.750000	113.750000	2.000000	44.937042	-114.247557
max	9.000000	208.000000	88.000000	373.000000	2.000000	47.045435	-95.695906

East Coast

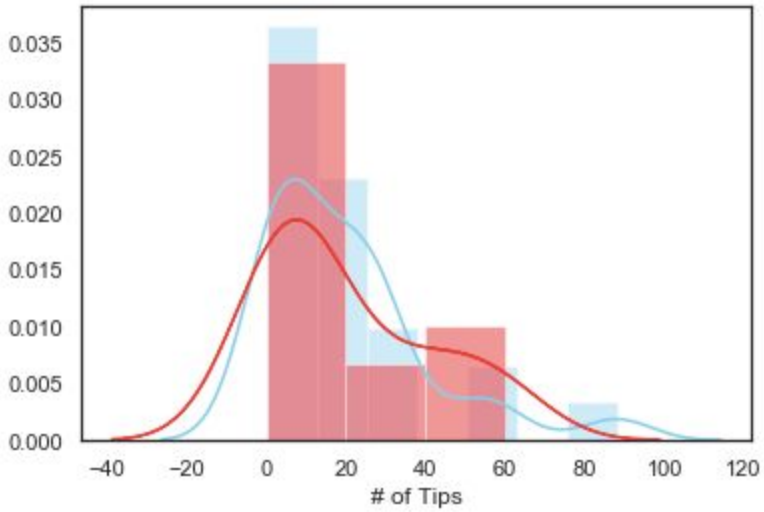
	Rating	# of Likes	# of Tips	# of Photos	Price Category	Latitude	Longitude
count	15.000000	15.000000	15.000000	15.000000	15.000000	15.000000	15.000000
mean	8.180000	51.533333	20.200000	57.066667	1.200000	42.096926	-88.412531
std	0.544059	61.901842	20.918891	68.371325	0.414039	7.402665	19.571899
min	6.900000	5.000000	0.000000	4.000000	1.000000	30.454443	-134.421394
25%	7.900000	11.000000	5.500000	17.500000	1.000000	39.068161	-87.760729
50%	8.100000	37.000000	13.000000	32.000000	1.000000	39.963640	-83.000568
75%	8.550000	63.500000	35.500000	78.000000	1.000000	42.904774	-76.978404
max	9.000000	246.000000	60.000000	271.000000	2.000000	58.301657	-71.031759



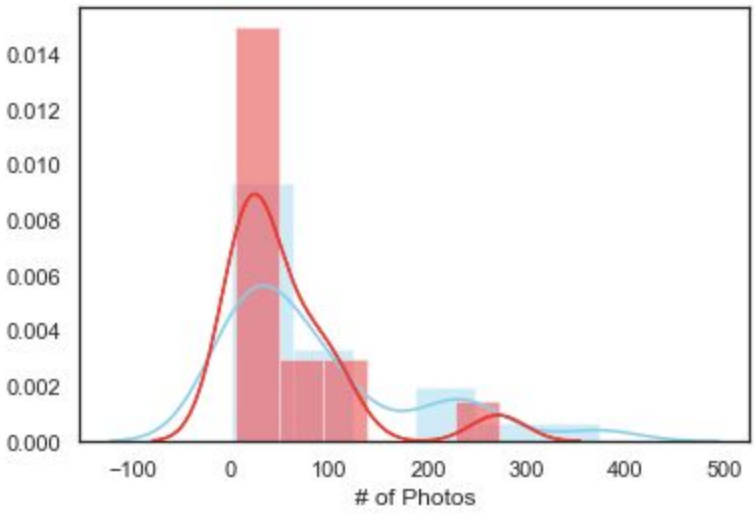
T-test did not give a significant result for the number of likes (p-value=0.75). This means that there is no significant difference between the coasts in terms of how many times people liked the coffee shops.



T-test did not give a significant result for ratings (p-value=0.80). This means that there is no significant difference between the coasts in terms of how much people rated the coffee shops.



T-test did not give a significant result for the number of tips (p-value=0.99). This means that there is no significant difference between the coasts in terms of how much they engaged with the coffee shops and left tips for other people.



Lastly, the number of photos uploaded is not significantly different either. Here I used Mann-Whitney U Test (the non parametric version of student-t test), because log transformation was inapplicable. And I got p-value=0.24.

Bars

Since there is no bar returned from our sample in the west coast, there is no further analysis done for bars.

Pizza Places

There are 3 pizza places in our sample for the West Coast, compared to 17 returned for the East Coast. The sample is not enough to do a comparison.

Comparing the Coasts

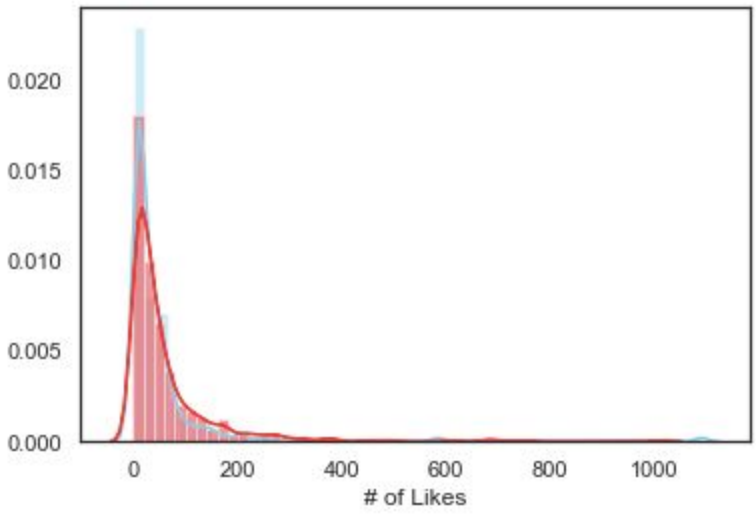
I decided to do a comparison between coasts in terms of the number of likes in general, rather than doing a venue category-wise comparison. This would be helpful to see if there is a popularity difference between capitals in each coast.

West Coast Venues sorted by the number of likes

	Capital	Venue ID	Venue Name	Categories	Rating	# of Likes	# of Tips	# of Photos	Price Category	Latitude	Longitude
6	Austin	4c77cbe5947ca1cd90694837	Austin City Limits Live	Performing Arts Venue	9.4	1092	98	3212	NaN	30.265288	-97.747260
15	Austin	49e32bbaf964a52068621fe3	La Condesa	Mexican Restaurant	8.6	586	269	808	3.0	30.265466	-97.747734
3	Austin	4db8a87c6a2334682d9809a9	Violet Crown Cinema	Indie Movie Theater	8.9	330	75	331	NaN	30.265524	-97.748189
2	Sacramento	4b0586b4f964a520816a22e3	Gunther's Quality Ice Cream	Ice Cream Shop	9.2	251	79	288	1.0	38.553600	-121.475792
4	Austin	554377bc498e6cb88b23f4bf	Trader Joe's	Grocery Store	9.2	208	8	195	NaN	30.267585	-97.752687
7	Sacramento	4b1c1d5ff964a520680224e3	Temple Coffee & Tea	Coffee Shop	9.0	208	88	373	2.0	38.563899	-121.472408
2	Austin	4a61f288f964a520b1c21fe3	Juan Pelota Café	Coffee Shop	8.9	177	58	255	1.0	30.267953	-97.749365
8	Sacramento	4b0586b9f964a520386b22e3	Sacramento Natural Foods Co-op	Grocery Store	9.1	163	32	156	NaN	38.564475	-121.472676

East Coast Venues sorted by the number of likes

	Capital	Venue ID	Venue Name	Categories	Rating	# of Likes	# of Tips	# of Photos	Price Category	Latitude	Longitude
3	Atlanta	4a05d34ef964a52083721fe3	Centennial Olympic Park	Park	9.2	1012	209	3505	NaN	33.760356	-84.393507
13	Des Moines	4df4d43522718759f8245edd	Zombie Burger + Drink Lab	Burger Joint	8.7	749	317	1310	2.0	41.590380	-93.613471
12	Atlanta	4aa08dedf964a520094020e3	Atlanta Marriott Marquis	Hotel	8.6	695	128	2267	NaN	33.761600	-84.385929
0	Atlanta	40e0b100f964a5209b071fe3	The Tabernacle	Music Venue	9.1	678	93	1746	NaN	33.758719	-84.391455
3	Madison	4afcc582f964a520bc2522e3	The Old Fashioned Tavern & Restaurant	Gastropub	9.0	614	259	906	2.0	43.076153	-89.383526
10	Boston	4a5c1457f964a5202fbc1fe3	South Shore Plaza	Shopping Mall	8.0	493	82	256	NaN	42.221995	-71.023768
13	Raleigh	4e091acc1f6e21103396e069	Beasley's Chicken + Honey	Southern / Soul Food Restaurant	9.2	453	173	590	2.0	35.776968	-78.638175
8	Des Moines	4b43e42ef964a52058ed25e3	Fong's Pizza	Pizza Place	9.0	388	198	532	2.0	41.585925	-93.621893



There is indeed a significant difference between the coasts (p-value = 0.004) in terms of the number of likes. We can also observe that the East Coast capitals are richer in terms of venue categories, such as parks, stadiums, museums, arenas, markets etc. However, when we check the sorted list for the West, we can see that capital centers do not provide that many variety to people.

Results

The analysis showed that there is a clear distinction between the West and East Coast capitals. Coffee Shops are more common in the West Coast whereas bars and pizza places are common in the East. The comparison of coffee shops in the two coasts in terms of the number of likes, ratings, the number of tips, the number of photos did not give significant results. Therefore we can conclude that how people engage with these coffee shops is not significantly different in the two coasts. This also means that coffee shops in the West are not unique in itself to attract people's special attention and effort to rate, like or upload photos that are substantially different than people do in the East Coast coffee shops. The same comparison was not possible for bars and pizza places because there are not enough samples from the West Coast to reach meaningful conclusions. But this in general is a sign that bars and pizza places are not common in the West. Therefore travel agencies can set their customers' expectations accordingly. Tourists may have a hard time to find bars and pizza places in the West.

Secondly, we see that capital centers in the East Coast are richer in content. There are more variety of venues on this coast. People engage with these venues more actively compared to the West. The states are smaller in the East, so there are more capitals and they can be more reachable in a limited time period. So travel agencies may consider recommending their customers to allocate more time in the East Coast.

Discussion

There are limits to this project where it can be overcomed if it is turned to be a more extensive project. The biggest obstacle was not being able to pull data from Foursquare because of certain restrictions. This could be overcomed by doing extra payment and switching to a private account. I was able to pull 20 venues for each capital, increasing this to 50 or 100 would provide a dataset to do more confident analysis and to reach better conclusions.

Also, I limited the center to 1500 m radius. This might be increased or decreased for some capitals. There could be more extensive research on how each capital defines its limits of the downtown area. Besides that, I was able to work with 49 capitals and had to omit Bismarck because no venues returned from the Foursquare pull for this capital.

Another thing is I only got records of the venues which are registered in Foursquare, there might be places which are missed because they are not in the database. Furthermore, I formed my analysis based on factors that are shared in the website. There might be a huge number of people who did not share their opinions for the venues that they visited. So the information I had about the venues is limited to the people who are active users.

In this project, I used ratings, number of likes, number of tips, number of photos as indicators of a place's popularity. There are factors that I wanted to include but could not do that because it is not possible to get it if you are not the owner of the place. Such as the total count of check-in of a place so far and the number of different people who visited a place. These two factors could have been useful to know in terms of figuring out the crowdedness of a place and how many new people a place could attract.

Conclusion

My aim in this project is to see the differences and similarities in the capitals in terms of venue categories and have meaningful insights for the travel agencies to attract their customers with appealing recommendations. I believe this project is helpful to travel agencies for creating an itinerary on each coast. The results showed that it is better if they have unique travel recommendations for each coast. They can look more into the substitutes of bar and pizza places in the West Coast closely. Also they can investigate more closely the places in the East, and calculate how much time is needed to have an extensive trip compared to the West.