Getting a recommendation on which State Capital to visit based off users palate

Alex Henner

January 20, 2021

1. Introduction

1.1 Background
The world is filled with many different types of food and locations on where to get those types of food. People enjoy exploring new cities based on multiple criteria. One of those criteria happen to be food.

1.2 Problem
Each city has its own ecosystem that may or may not be like others. So trying to figure out if you would like the food that a city has to offer can be a pain. This project aims to simplify those concerns by recommending State Capitals based off the users palate. This project can be easily scaled to more than just the State Capitals.

1.3 Interest
Many users, especially with food concerns, want to explore new cities to broaden their experiences, but have a hard time starting trying to find if a location can handle their food needs.

2. Data acquisition and cleaning

2.1 Data Sources
The data was found using Foursquare API, which gave names of venues and the venue's category. I was limited to 50 venues per search query, which decreased the information gained from more densely populated Capitals. There was also an issue with the lack of premium calls to get the ratings for each venue. We were limited to 500 premium calls per day, so with a fairly large data set of venues that I collected it would have taken a month to collect the ratings. Because of this I focused on the density of the venue categories instead of the average ratings.

2.2 Data Cleaning
After getting the Foursquare information on the venues and their categories, there were cases of duplication and non-food categories that I deleted. I didn't want to have 100 categories that a user would need to rate before getting a recommendation, so I narrowed it down to the following 27 categories.

- African Restaurant
- American Restaurant
- Asian Restaurant
- Bar
- Breakfast
- Buffet
- Chinese Restaurant
- Coffee Shop
- Comfort Food Restaurant
- Dessert Shop
- Eastern European Restaurant
- Fast Food Restaurant
- Food Truck
- French Restaurant
- Indian Restaurant
- Italian Restaurant
- Japanese Restaurant
- Korean Restaurant
- Latin American Restaurant
- Mediterranean Restaurant
- Middle Eastern Restaurant
- Modern European Restaurant
- Sandwich Place
- Seafood Restaurant
- Southern / Soul Food Restaurant
- Spanish Restaurant
- Vegetarian / Vegan Restaurant

To further get the data into those 27 categories, I had to make some assumptions such as 'Bagel Shop' would fall under the breakfast category. Below is a list of all the assumptions that were made.

Main	**Includes**
'Eastern European Restaurant'	
'African Restaurant'	'Moroccan Restaurant', 'Ethiopian Restaurant'
'Seafood Restaurant'	
'Coffee Shop'	'Café', 'Snack Place', 'Bakery', 'Tea Room', 'Bubble Tea Shop', 'Corporate Coffee Shop', 'Juice Bar',
'Food Truck'	'Street Food Gathering',
'Breakfast'	'Breakfast Spot', 'Bed & Breakfast', 'Bagel Shop', 'Donut Shop', 'Creperie',
'Irish Pub'	'Fish & Chips Shop'
'Latin American Restaurant'	'Taco Place', 'Burrito Place', 'Mexican Restaurant', 'Argentinian Restaurant', 'Cuban Restaurant', 'Brazilian Restaurant', 'Portuguese Restaurant', 'Caribbean Restaurant'
'Fast Food Restaurant'	'Food & Drink Shop', 'Food Court', 'Bistro', 'Food Stand', 'Airport Food Court', 'Food', 'Restaurant', 'Theme Restaurant'
'American Restaurant'	'New American Restaurant', 'Diner', 'Steakhouse', 'Burger Joint', 'Hot Dog Joint', 'BBQ Joint', 'Tex-Mex Restaurant', 'Hawaiian Restaurant', 'Wings Joint'
'Buffet'	'Cafeteria', 'College Cafeteria', 'Corporate Cafeteria',
'Indian Restaurant'	'Andhra Restaurant', 'North Indian Restaurant', 'Sri Lankan Restaurant'
'Italian Restaurant'	'Pizza Place'
'Bar'	'Wine Bar', 'Brewery', 'Winery', 'Sake Bar', 'Karaoke Bar', 'Strip Club', 'Dive Bar', 'Sports Bar', 'Speakeasy', 'Pub', 'Hookah Bar', 'Gay Bar', 'Cocktail Bar', 'Whisky Bar', 'Hotel Bar', 'Piano Bar', 'Tiki Bar', 'Beer Bar', 'Gastropub', 'Beer Garden'

'Sandwich Place'	'Deli / Bodega',
'Chinese Restaurant'	'Dumpling Restaurant', 'Szechuan Restaurant', 'Malay Restaurant', 'Shanghai Restaurant', 'Dim Sum Restaurant'
'Asian Restaurant'	'Mongolian Restaurant', 'Thai Restaurant', 'Filipino Restaurant', 'Vietnamese Restaurant', 'Noodle House'
'Vegetarian / Vegan Restaurant'	'Salad Place'
'Spanish Restaurant'	'Tapas Restaurant'
'Mediterranean Restaurant'	'Greek Restaurant'
'Korean Restaurant'	'Korean BBQ Restaurant'
'Japanese Restaurant'	'Sushi Restaurant', 'Shabu-Shabu Restaurant', 'Hotpot Restaurant', 'Ramen Restaurant',
'Middle Eastern Restaurant'	'Halal Restaurant', 'Israeli Restaurant', 'Turkish Restaurant', 'Turkish Home Cooking Restaurant', 'Doner Restaurant', 'Falafel Restaurant', 'Persian Restaurant', 'Afghan Restaurant'
'Dessert Shop'	'Ice Cream Shop', 'Cupcake Shop', 'Frozen Yogurt Shop', 'Chocolate Shop', 'Gelato Shop',
'Southern / Soul Food Restaurant'	'Mac & Cheese Joint', 'Fried Chicken Joint'
'Comfort Food Restaurant'	'Soup Place'
'French Restaurant'	'Cajun / Creole Restaurant'
'Modern European Restaurant'	
'Swiss Restaurant'	

2.3 Feature Selection

Once the 27 categories were cleaned, I proceeded to take each of the Capitals and find the density of each of the categories. This would provide me with the data needed to explore in more detail.

3. Exploratory Data Analysis

3.1 Top 10

I started by looking at the top 10 categories from each state. As you can see in *Figure 1* we have a high level view at a glance.

	Capital	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
0	Albany	Bar	Fast Food Restaurant	American Restaurant	Chinese Restaurant	Coffee Shop	Sandwich Place	Food Truck	Latin American Restaurant	Asian Restaurant	Breakfast
1	Annapolis	Coffee Shop	Bar	Fast Food Restaurant	Breakfast	Seafood Restaurant	Italian Restaurant	American Restaurant	Latin American Restaurant	Japanese Restaurant	Food Truck
2	Atlanta	Food Truck	Coffee Shop	Fast Food Restaurant	Breakfast	Bar	American Restaurant	Chinese Restaurant	Southern / Soul Food Restaurant	Italian Restaurant	Seafood Restaurant
3	Augusta	American Restaurant	Fast Food Restaurant	Coffee Shop	Italian Restaurant	Bar	Breakfast	Chinese Restaurant	Southern / Soul Food Restaurant	Latin American Restaurant	Seafood Restaurant
4	Austin	Food Truck	Coffee Shop	Bar	Fast Food Restaurant	Latin American Restaurant	American Restaurant	Chinese Restaurant	Breakfast	Asian Restaurant	Italian Restaurant

Figure 1

3.2 Clusters

This allowed me to cluster the data using Kmeans. I used 5 clusters for the analysis. The first cluster seemed to be grouped by high density of food trucks and Chinese food as seen in Figure 2. Most of the clusters seemed to be grouped in a similar fashion with similar densities, but a few were grouped by lack of having certain groups all together. In Figure 3 you can see that the clusters are somewhat regional as well, which is interesting since that data was not used for clustering.
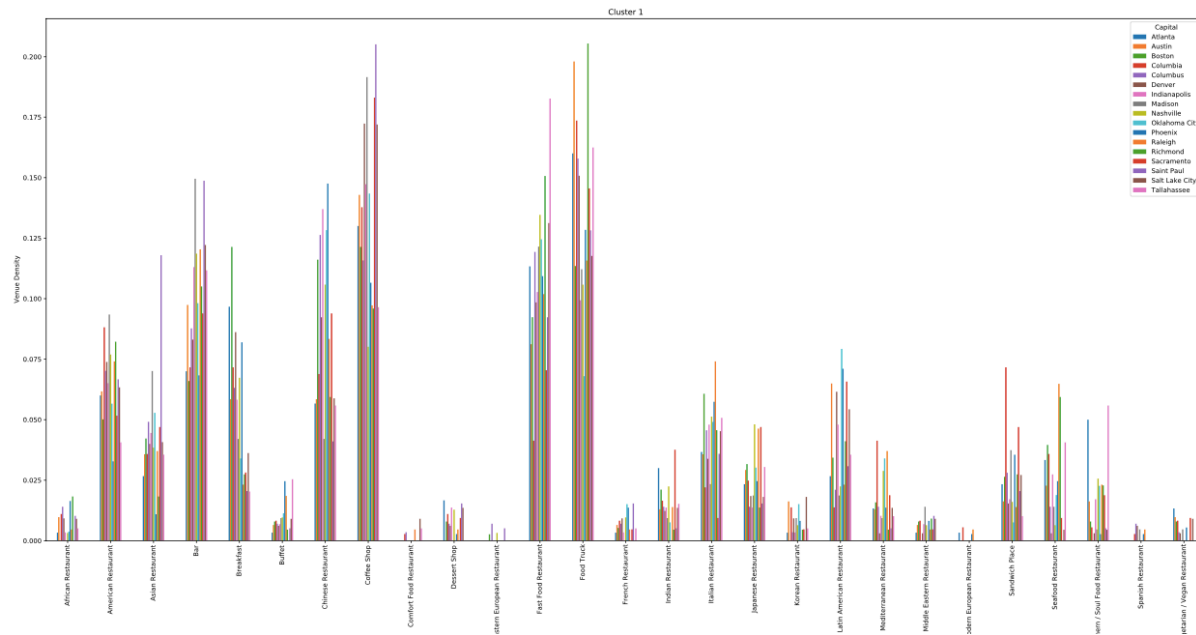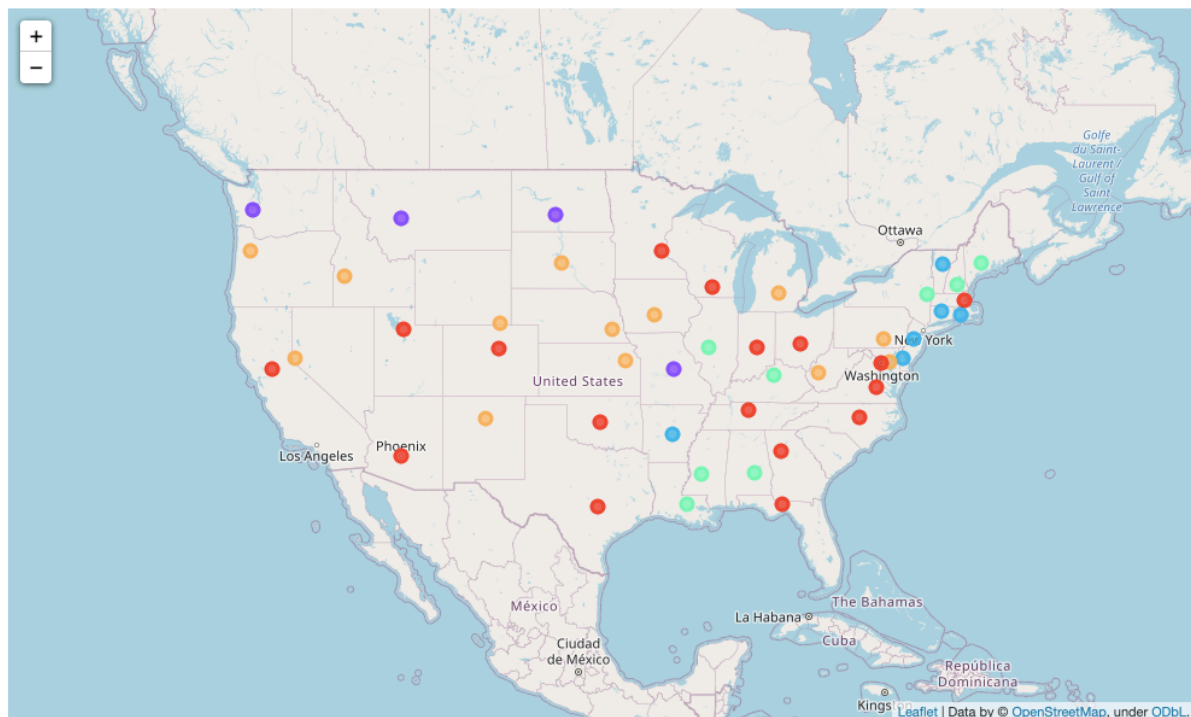

Figure 2

Figure 3

4. Results

4.1 Model

Users were asked to rate each of the 27 categories using a 0-5 scale. I collected this data using Google Forms, which made it easier to get bulk recommendations. The program would then list the top 5 State Capitals for each user as seen in Figure 4.

```
Lindsay  will enjoy:
          Capital         State
1          Juneau        Alaska
18        Augusta         Maine
27    Carson City        Nevada
30       Santa Fe    New Mexico
40         Pierre  South Dakota


Chris will enjoy:
         Capital          State
9        Atlanta        Georgia
18       Augusta          Maine
20        Boston  Massachusetts
39      Columbia  South Carolina
42        Austin          Texas
```

Figure 4

5. Discussion

5.1 Observations

It is interesting to see that the results don't always fall into the clusters. What I did find fascinating is that of the 20 user profiles that I did using friends and family, 25% of them received Atlanta, Ga in their top 5. This is interesting because the 20 users are all from Atlanta, so their tastes were shaped from growing up in Atlanta. I want to think of this as a sign that the system works fairly well. It would be interesting to scale this into the top 5 populous cities from each state, since a good portion of the State Capitals aren't in the top 5 for their State.

6. Conclusion

In conclusion, while I believe that the recommendation system works, I believe that it could be better. I would really like to see recommendations based off the ratings of the venues instead of just the density. Just because a Capital may have a lot of a certain category, it may not be very good.