

An aerial photograph of San Francisco, showing the dense urban landscape with numerous buildings, streets, and green spaces. The city's iconic skyline, including the Transamerica Pyramid, is visible in the background against a clear blue sky. The foreground shows a mix of residential and commercial buildings, with some trees and parks interspersed among the structures.

Comparing the Rising Tech City Hubs: Comparing and Investing Similarity to San Francisco

IBM Data Science Specialization
Applied Data Science Course: Capstone Project
Pedro Junior Vicente Valdez

What city is most similar to San Francisco?

- San Francisco is one of the biggest tech hubs in the United States.
- Overpopulation and increasing specialty jobs are causing a issues.
 - Housing Crisis
 - Employment
 - Increasing Cost of Living
- Workforce is looking at other locations in the US with similar environments.
- Some of the rising tech hubs include Seattle, Chicago, Austin and Colorado.

Top Tech Cities in the US (Source: BuiltIn)

AUSTIN AT A GLANCE

950,715

POPULATION

2,578

TECH COMPANIES

90,656

COMMUNITY MEMBERS

#44

MOST EXPENSIVE CITY

BOSTON AT A GLANCE

698,898

POPULATION

2,171

TECH COMPANIES

61,019

COMMUNITY MEMBERS

#8

MOST EXPENSIVE CITY

CHICAGO AT A GLANCE

2,749,360

POPULATION

5,798

TECH COMPANIES

186,729

COMMUNITY MEMBERS

#17

MOST EXPENSIVE CITY

NYC AT A GLANCE

8,558,000

POPULATION

4,918

TECH COMPANIES

184,876

COMMUNITY MEMBERS

#3

MOST EXPENSIVE CITY

COLORADO AT A GLANCE

5,770,000

POPULATION

3,774

TECH COMPANIES

123,659

COMMUNITY MEMBERS

#13

MOST EXPENSIVE CITY

LOS ANGELES AT A GLANCE

3,946,000

POPULATION

5,934

TECH COMPANIES

151,336

COMMUNITY MEMBERS

#10

MOST EXPENSIVE CITY

SEATTLE AT A GLANCE

755,745

POPULATION

2,065

TECH COMPANIES

23,237

COMMUNITY MEMBERS

#12

MOST EXPENSIVE CITY

GEOPY was used to get the latitude and longitude of all 8 cities.

- Geopy makes it “easy for Python developers to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources.”



```
# Specifying the address
address = 'Downtown, San Francisco'

# Gathering the Location coordinates
location_sf = geolocator.geocode(address)

# Placing coordinates to variables
latitude_sf = location_sf.latitude
longitude_sf = location_sf.longitude
print('The geographical coordinate of Downtown, San Francisco are {}, {}'.format(latitude_sf, longitude_sf))
```

The geographical coordinate of Downtown, San Francisco are 37.7875138, -122.407159.

	Cities	Latitude	Longitude
0	San Francisco, California	37.787514	-122.407159
0	Denver, Colorado	39.751770	-105.013873
0	Seattle, Washington	47.604872	-122.333458
0	Los Angeles, California	34.498713	-118.584307
0	New York City, NY	40.599756	-73.946390
0	Chicago, Illinois	41.893648	-87.621960
0	Austin, Texas	30.268054	-97.744764
0	Boston, Massachusetts	42.362918	-71.068737

The Foursquare API was used to obtain venue recommendations and it's ratings.



```
venue_id = "551cfcaf498e23f2c0115449"

url = "https://api.foursquare.com/v2/venues/{}".format(venue_id)
|
params = dict(
    client_id=CLIENT_ID,
    client_secret=CLIENT_SECRET,
    v='20191129')

resp2 = requests.get(url=url, params=params)
data2 = json.loads(resp2.text)
```

```
venue_details = json_normalize(data2)
venue_details[["response.venue.name", "response.venue.rating"]]
```

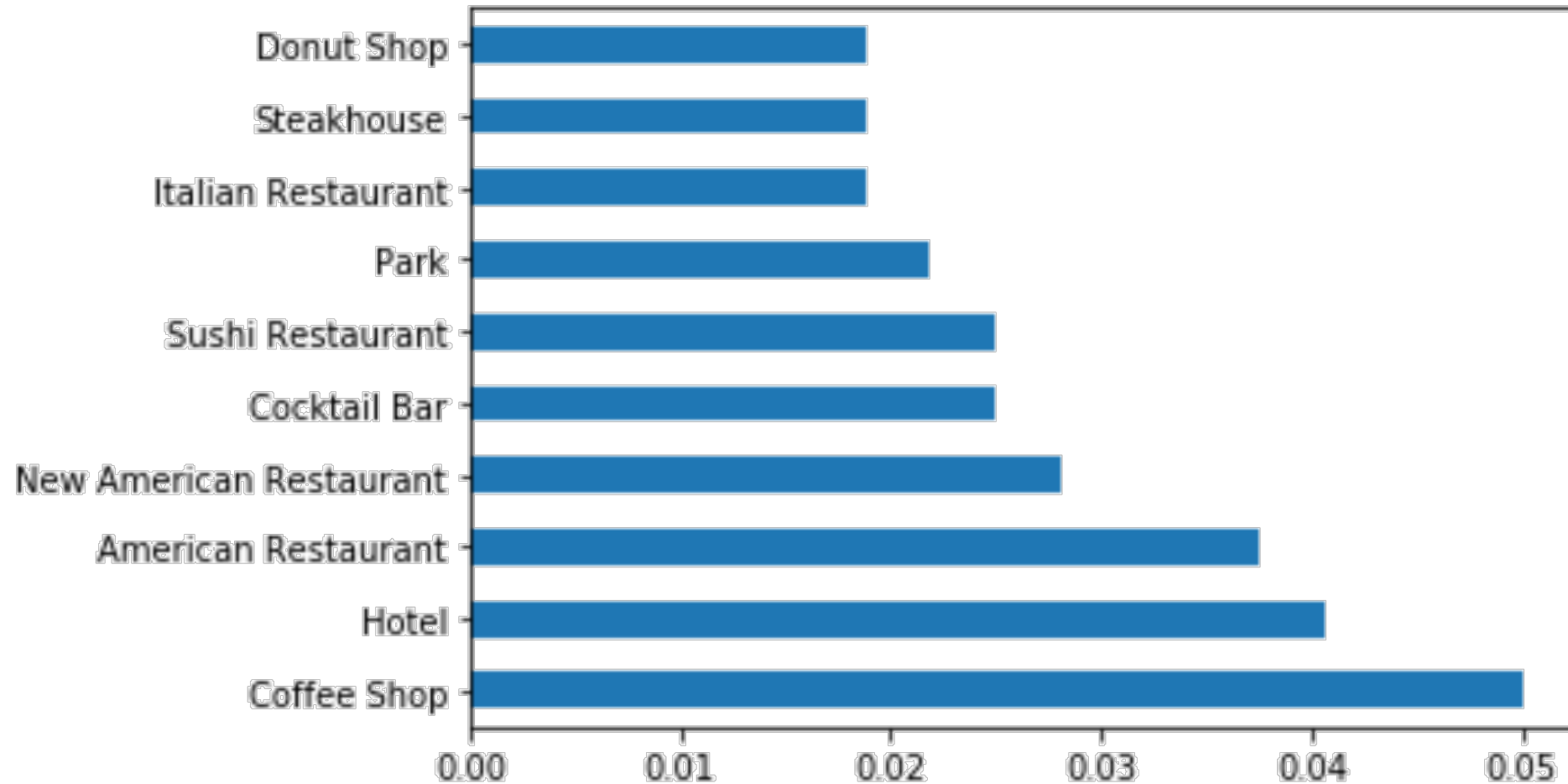
	response.venue.name	response.venue.rating
0	Maison Margiela	9.2

	venue.name	venue.id	venue.location.formattedAddress	venue.categories	venue.location.lat	venue.location.lng
0	Maison Margiela	551cfcaf498e23f2c0115449	[134 Maiden Ln, San Francisco, CA 94108, Unite...	[{'id': '4bf58dd8d48988d104951735', 'name': 'B...	37.788261	-122.405765
1	Saint Laurent	528d4fe211d2543b7663f4fd	[108 Geary St, San Francisco, CA 94108, United...	[{'id': '4bf58dd8d48988d104951735', 'name': 'B...	37.787774	-122.405412
2	Williams-Sonoma	4aa45625f964a5207b4620e3	[340 Post St (btwn Powell & Stockton), San Fra...	[{'id': '58daa1558bbb0b01f18ec1b4', 'name': 'K...	37.788377	-122.407446
3	Tiffany & Co.	4a791992f964a520efe61fe3	[350 Post St (btwn Powell & Stockton), San Fra...	[{'id': '4bf58dd8d48988d111951735', 'name': 'J...	37.788598	-122.407708
4	UNIQLO	50043438e4b0f448ea4f447f	[111 Powell St, San Francisco, CA 94102, Unite...	[{'id': '4bf58dd8d48988d103951735', 'name': 'C...	37.785850	-122.408041

A Pandas dataframe was created and filtered to contain only the following features.

- City
- City Latitude and Longitude:
- Venue Name
- Venue ID
- Venue Latitude and Longitude
- Venue Category
- Venue Rating

Coffee Shops and Restaurants were among the most recurring recommended venues



The category “Park” is missing from the most common venues in San Francisco and Boston

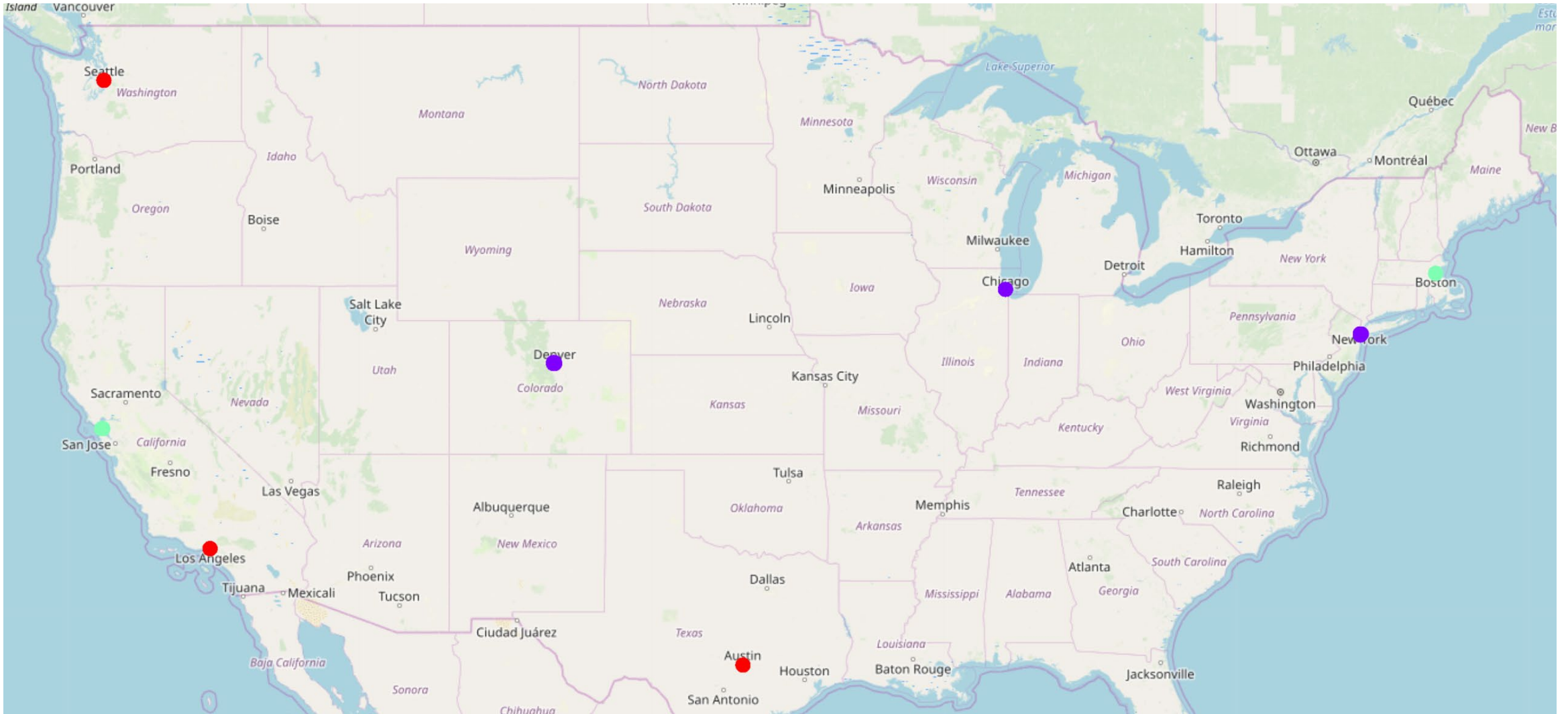
- These type of observations are expected to be learned by the model. If this is true, we expect Boston and San Francisco to belong to the same cluster.

	City	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	Austin	Hotel	Cocktail Bar	Lounge	Steakhouse	Coffee Shop	Burger Joint	Park	Cajun / Creole Restaurant	Chinese Restaurant	Salad Place
1	Boston	Coffee Shop	New American Restaurant	Gym / Fitness Center	Steakhouse	Sandwich Place	Salad Place	Gastropub	Hotel	Falafel Restaurant	Restaurant
2	Chicago	American Restaurant	New American Restaurant	Donut Shop	Grocery Store	Restaurant	Cosmetics Shop	Yoga Studio	Café	Salon / Barbershop	Resort
3	Colorado	Theme Park Ride / Attraction	Coffee Shop	Park	Yoga Studio	Ice Cream Shop	Café	Sushi Restaurant	Brewery	Pizza Place	Seafood Restaurant
4	Los Angeles	Plaza	Ice Cream Shop	Speakeasy	Coffee Shop	Theater	Jazz Club	Park	School	Candy Store	Historic Site
5	New York	Pizza Place	Sushi Restaurant	Bakery	Italian Restaurant	Food & Drink Shop	Bagel Shop	Deli / Bodega	Farmers Market	Mexican Restaurant	Bubble Tea Shop
6	San Francisco	Boutique	Hotel	Men's Store	Bubble Tea Shop	Clothing Store	Plaza	Gym / Fitness Center	Shoe Store	Music Venue	Food Truck
7	Seattle	Hotel	Coffee Shop	Cocktail Bar	Concert Hall	Café	Donut Shop	Deli / Bodega	Seafood Restaurant	Scenic Lookout	Sandwich Place

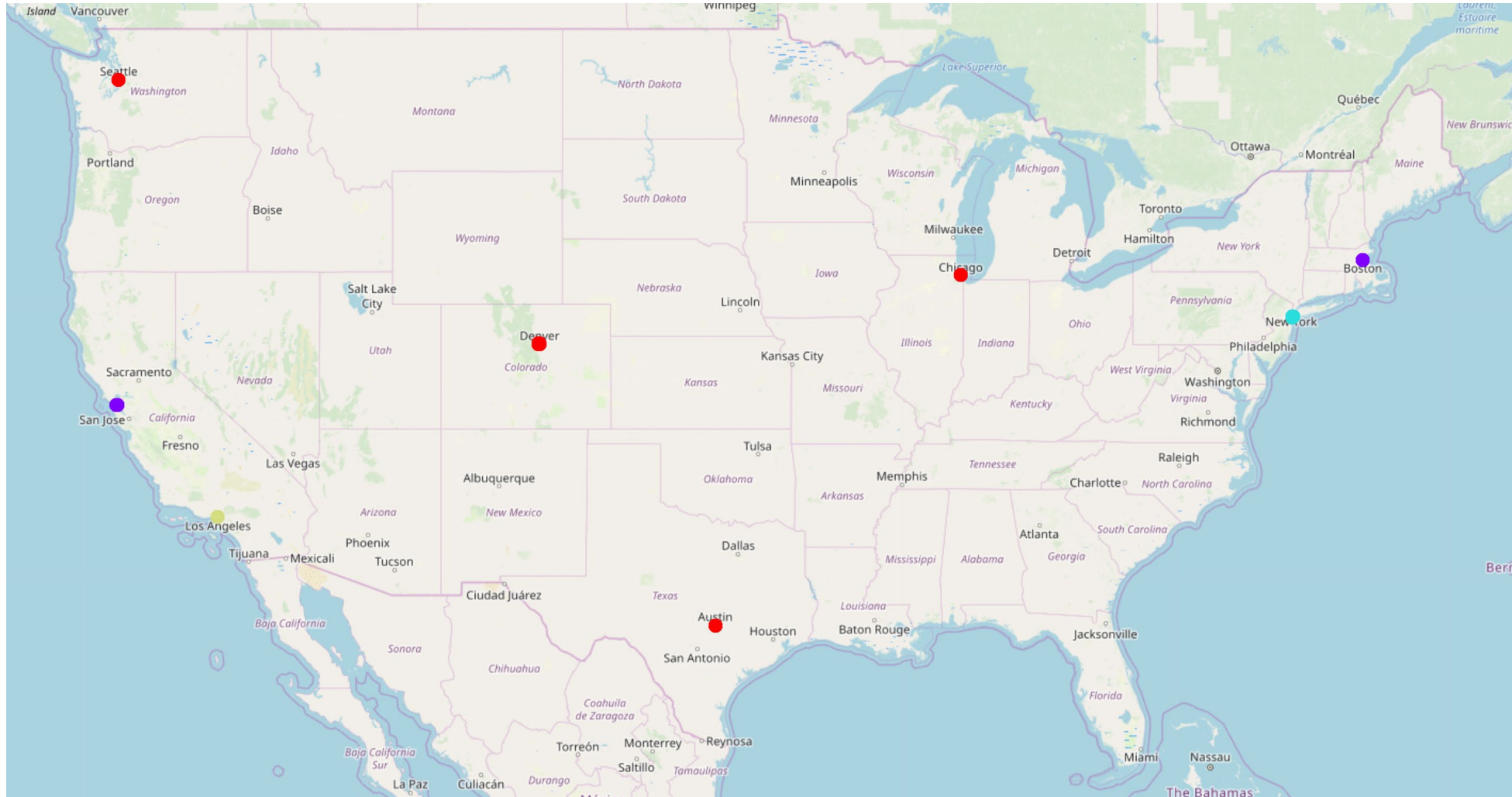
Hypothesis

- Based purely on distance, either Denver, Los Angeles, or Seattle are expected to be similar to San Francisco.
- Based on the observation that “Parks” are missing in San Francisco in terms of recommended venues we can expect Boston to be similar to San Francisco

Boston is the only city belonging to the same cluster as San Francisco with $K=3$



Results were corroborated by increasing the number of clusters to four.



Conclusion: Boston is the most similar city.

- Our hypothesis based on distance was wrong (expectedly). This problem is more complex than just using Euclidean distances.
- It is important to know the limitations of the current results. This is the result of a clustering algorithm fitted to a limited number of recommended venues (limited by the number of premium calls that allowed us to get the venue's ratings). This is not representative of an entire city but for the purposes of this project it will suffice.
- From the qualitative analysis we can see that both, Boston and San Francisco, have a high venue density of Boutiques, Stores and Hotels followed by Tea Places and Gyms.
- Other clusters include Parks as one of the most common categories. Parks are not even in the top 10 most common venues of San Francisco cluster #1. This is in contrast to other cities like Denver, Chicago, and New York which has Parks as the 3rd most common venue.

Future Work

- Architecture has a big impact on similarity and therefore the venues picture could be used by a convolutional neural network to classify visual similarity amongst cities based on their venues (capturing bigger vs smaller cities).
- Other parameters that could be included is Census data including population, car density, population density, etc.