A Tale of Two Cities: Capstone Project

San Jose, California vs Austin, Texas

Introduction

In May 2020, Elon Musk announced that Telsa would relocate its headquarters and future programs to Texas. While not the first to announce such a move, Telsa was perhaps the most prominent in the Silicon Valley tech firm exodus. The premier destination for the corporations is Austin, Texas the capitol of the state of Texas. Austin is a city home to 950,807 (2019) residents spread over 319.94 square miles. The city is growing in terms of diversity and boasts a large technology sector. On the departure side of this movement is the heart of Silicon Valley, the city of San Jose. San Jose has been the cultural, financial and political center of Silicon Valley and is the largest city in Northern California by both population and area. San Jose's estimated 2019 population was 1,021,795 (2019) making it the tenth most populous city in the United States. The city of San Jose encompasses 178.24 square miles.

Part of the ongoing narrative regarding Austin, Texas is that the city is not only more accommodating for tech companies but also for employees and people in general. This capstone project will explore the differences in the neighborhoods of both cities and make a comparison based on the common nearby venues. How much truth is there that the two cities are similar? This information would prove to be valuable for any individuals contemplating moving.

Data

The data sources for this project.

https://www.sanjose.org/neighborhoods

https://www.homecity.com/blog/best-neighborhoods-in-austin/

Foursquare API

https://www.google.com/maps

GeoPy

Data Sourcing and Cleaning

```
source = requests.get('https://www.sanjose.org/neighborhoods').text

soup = BeautifulSoup(source, 'lxml')

#body = soup.find('body')
#print(body.prettify()) # run to check that the website has loaded

SJ_list=[]  # create a list
for Districts in soup.find_all('div', class_='hero--event'):
    Name = Districts.h2.a.text
    SJ_list.append(Name)

SJ_list
```

```
url = 'https://www.homecity.com/blog/best-neighborhoods-in-austin/'
data = requests.get(url)
soup = BeautifulSoup(data.text, 'html.parser')
neighborhood = soup.find("div", class_='post-content')
neighborhood_names= neighborhood.find_all("h2")
print(neighborhood_names)
```

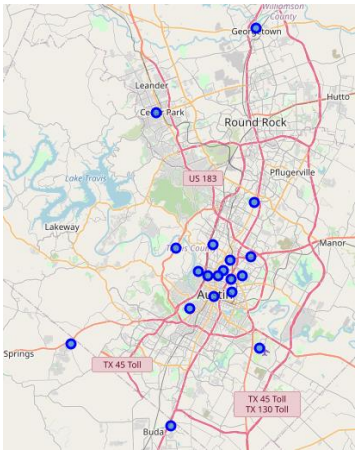
The data set used for San Jose is from the city of San Jose's website that lists the various neighborhoods which loosely follow the city's zip codes. This list was only a list of names and did not include postal codes. The data set used for Austin is from a real estate website which rates neighborhoods within the city limits. Again this was only a list of names and did not include postal codes. Beautiful soup was used to extract the neighborhood names and Geopy Nominatum was used to determine the Geo coordinates. However, some of the Geo coordinates were inaccurate and needed to be manually replaced. I used Google Maps to locate the correct Geospatial coordinates and updated the dataframe. Once the neighborhoods had the correct coordinates the datasets were ready to source the venue data from Foursquare. Many of the San Jose coordinates provided by Geopy were incorrect and the locations of the neighborhoods had to be checked via Google Maps.

data sets used were

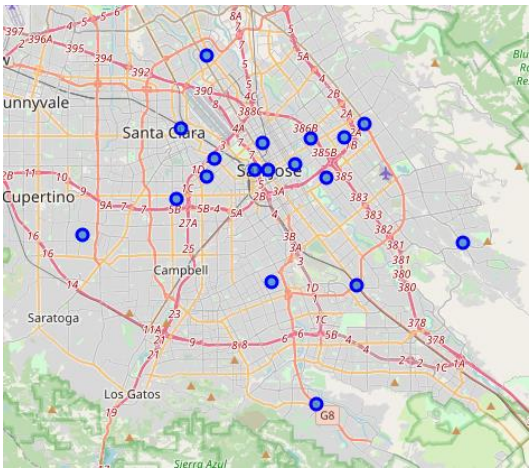
	Neighborhood	SJ_address	query
0	Downtown	San Jose, CA, US	Downtown San Jose, CA, US
1	Japantown	San Jose, CA, US	Japantown San Jose, CA, US
2	Little Italy	San Jose, CA, US	Little Italy San Jose, CA, US
3	Little Portugal	San Jose, CA, US	Little Portugal San Jose, CA, US
4	Little Saigon	San Jose, CA, US	Little Saigon San Jose, CA, US
5	Mayfair	San Jose, CA, US	Mayfair San Jose, CA, US
6	Santana Row	San Jose, CA, US	Santana Row San Jose, CA, US
7	West San Jose	San Jose, CA, US	West San Jose San Jose, CA, US
8	Willow Glen	San Jose, CA, US	Willow Glen San Jose, CA, US
9	The Alameda	San Jose, CA, US	The Alameda San Jose, CA, US
10	Rose Garden	San Jose, CA, US	Rose Garden San Jose, CA, US
11	Naglee Park	San Jose, CA, US	Naglee Park San Jose, CA, US
12	Alviso	San Jose, CA, US	Alviso San Jose, CA, US
13	North San Jose	San Jose, CA, US	North San Jose San Jose, CA, US
14	Alum Rock	San Jose, CA, US	Alum Rock San Jose, CA, US
15	Evergreen	San Jose, CA, US	Evergreen San Jose, CA, US
16	Almaden Valley	San Jose, CA, US	Almaden Valley San Jose, CA, US
17	South San Jose	San Jose, CA, US	South San Jose San Jose, CA, US

	Neighborhood	ATX_address	query
0	Allandale	Austin, TX, US	Allandale Austin, TX, US
1	Barton Hills	Austin, TX, US	Barton Hills Austin, TX, US
2	Buda	Austin, TX, US	Buda Austin, TX, US
3	Cedar Park	Austin, TX, US	Cedar Park Austin, TX, US
4	Cherrywood	Austin, TX, US	Cherrywood Austin, TX, US
5	Downtown	Austin, TX, US	Downtown Austin, TX, US
6	Dripping Springs	Austin, TX, US	Dripping Springs Austin, TX, US
7	East Austin	Austin, TX, US	East Austin Austin, TX, US
8	Georgetown	Austin, TX, US	Georgetown Austin, TX, US
9	Hyde Park	Austin, TX, US	Hyde Park Austin, TX, US
10	Kyle	Austin, TX, US	Kyle Austin, TX, US
11	Mueller	Austin, TX, US	Mueller Austin, TX, US
12	North Loop	Austin, TX, US	North Loop Austin, TX, US
13	North University	Austin, TX, US	North University Austin, TX, US
14	Old West Austin	Austin, TX, US	Old West Austin Austin, TX, US
15	South Austin	Austin, TX, US	South Austin Austin, TX, US
16	Tarrytown	Austin, TX, US	Tarrytown Austin, TX, US
17	Wells Branch	Austin, TX, US	Wells Branch Austin, TX, US
18	Westlake	Austin, TX, US	Westlake Austin, TX, US

Methodology



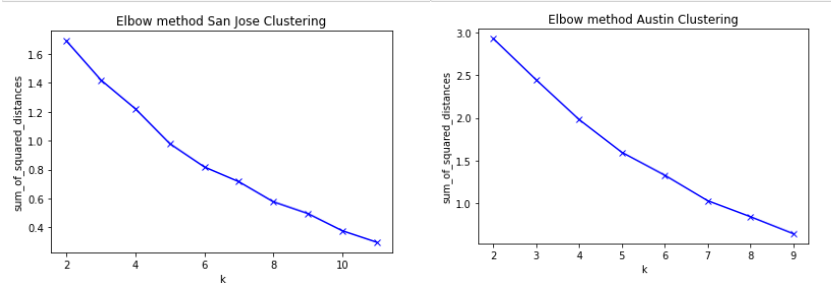
Austin



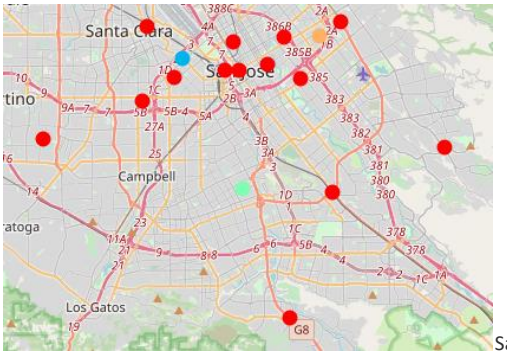
San Jose

Once the data sets were cleaned, the neighborhoods were marked on folium maps of both cities. In doing so provided a visual context of the location of the different neighborhoods as well as confirmation that the location was correct. All neighborhoods were processed through Foursquare API to generate the nearby venues in each neighborhood. The Foursquare parameters were 500 meters and a limit of 100 venues per search (neighborhood). The data was transformed to show the top ten common venues

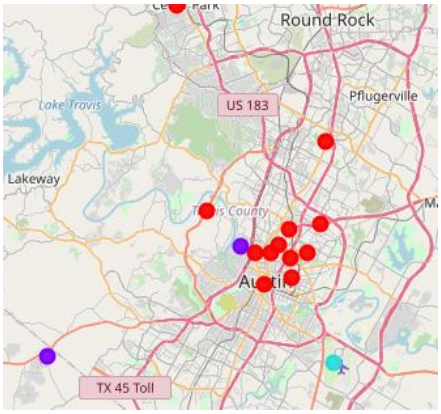
in each neighborhood in terms of frequency. Then neighborhoods were clustered and labeled. The optimal number of clusters was determined by K-means.



The elbow method was used once the points were graphed. The number of clusters seemed to be 5 clusters for San Jose and 4 clusters for Austin. The neighborhoods were mapped again this time with the color coded clusters and the maps were able present a visual of the dominant types of venues in each neighborhood.



San Jose



Austin

Results

San Jose

	Cluster Labels	1st Most Common Venue
0	0	Sandwich Place
1	0	Japanese Restaurant
2	0	Coffee Shop
3	0	Mexican Restaurant
4	0	Vietnamese Restaurant
5	4	Mexican Restaurant
6	0	Accessories Store
7	0	Pool
8	3	Pool
9	2	Mexican Restaurant
10	0	Theater
11	0	Pizza Place
12	0	Park
13	1	Food Truck
14	0	Light Rail Station
15	0	Dessert Shop
16	0	Coffee Shop
17	0	Mexican Restaurant

Austin

	Cluster Labels	1st Most Common Venue
1	3	Pizza Place
2	0	Food Truck
3	0	Theater
4	0	Accessories Store
5	0	Mexican Restaurant
6	1	Mexican Restaurant
7	0	Pool
8	0	Pool
9	0	Coffee Shop
10	0	Light Rail Station
11	0	Park
12	0	Sandwich Place
13	0	Dessert Shop
14	0	Japanese Restaurant
15	2	Coffee Shop
16	1	Mexican Restaurant
17	0	Vietnamese Restaurant
18	0	Mexican Restaurant

Five clusters were set for San Jose, 14 of 17 neighborhoods were clustered in label 1, the remaining minority clusters have one neighborhood each. The top venue for the San Jose minority clusters is Café, Park, Financial or Legal Service, and Cocktail Bar. Label 1 was centered on food and entertainment. Four clusters were set for Austin, 14 of 18 neighborhoods were clustered on label 0, the remaining minority clusters have 1 neighborhood with one cluster having 2 neighborhoods. The top venue for the Austin minority clusters is Pizza Place, Financial or Legal Service, Food stand, Vietnamese Restaurant.

Discussion

The largest clusters in both cities include similar venues types whose categories are centered on food and entertainment. In both cities Mexican food has a strong presence; in San Jose five neighborhoods have Mexican Food as the 2nd and 3rd most common venue. Austin has six neighborhoods with Mexican Food as the 2nd , 3rd , and 5th most common venue. In both cities, Asian food also has a nearly identical presence; two neighborhoods have Japanese and Vietnamese cuisine as 1st most common venue, then Thai, Hawaii, Chinese make appearances in the subsequent most common venues. Both Cities have a (1st most) café/coffee centric neighborhoods and a light rail system (2nd and 3rd). That Austin and San Jose would share similar venues is not unexpected. The demographics for both cities have a high percentage of

Latino residents (35.1% Austin 33.2% San Jose). Interestingly is how the Japanese cuisine and Vietnamese cuisine are the most common venues in two Austin neighborhoods. While Austin's Asian population has doubled from 3% to 6% since 1990 to 2010 is still far below San Jose's 32%. Both cities have similar infrastructure in having an airport, light rail station, and university.

Conclusion

This comparison certainly provides some insight on the nature of the neighborhoods but it definitely fails to capture all aspects of each city that one might want to factor in prior to moving. One example is that Austin has world renowned barbeque venues. Yet this is not captured in the foursquare data. Perhaps these venues are spread out too much to constitute a most common venue. In contrast for San Jose, one would expect a higher number most common venues of Asian cuisines in the neighborhoods. However, there are only two San Jose neighborhoods with Asian cuisines as their 1st most common Venue and predictably enough those neighborhoods are Little Saigon and Japantown. One factor might be that most Asian venues tend to congregate next to one another.

Lastly it is also important to point out that Silicon Valley is not just San Jose but rather the entire peninsula from San Francisco to San Jose. The entire route is populated with small suburban to mid-sized cities. Many of these cities have their own demographics that may vary from San Jose's and result in their own unique venues. Austin maybe roughly one hour from San Antonio but the route is not as populated with cities. If one were to consider leaving the Bay Area for Austin they may want to include more cities along with San Jose in their comparison analysis.