Recommending a City in the San Francisco Bay Area Based on Venues

Low Saelee

April 2, 2020

Introduction

The San Francisco Bay Area, home to approximately 7.75 million people, is a great place to live in, but there are many choices of cities to choose from. The region ranges from the north bay, up in Cloverdale, all the way to the south bay, down in Gilroy, which are about two hours drive from both endpoints. There are constantly new residents moving into the Bay Area to look for job opportunities, settling down and having a family. As with such, newcomers are always looking for the perfect city in the Bay Area to both buy or rent a property. In the Bay Area many live in a city different from their job location and often commute to work. Some commute up to an hour from their home to their work. Some are lucky enough to live and work in the same city with very little commuting time and distance.

Data

The cities of the San Francisco Bay Area are scraped from Wikipedia, https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_the_San_Francisco_Bay_Area, with a total of 101 cities. After obtaining the list of cities, from Wikipedia, I had to use the Nominatim package to retrieve the latitude and longitude before I can use the Foursquare API. Foursquare API was used to pull the many venues from each city. However, there are a total of seven cities without any venue information pulled from the Foursquare API.

Methodology

I used folium to create an initial visualization of the cities on the map. I centered the map on Alameda as that was the first city on the list, and it turns out that it was somewhat located at the center of all the other cities.
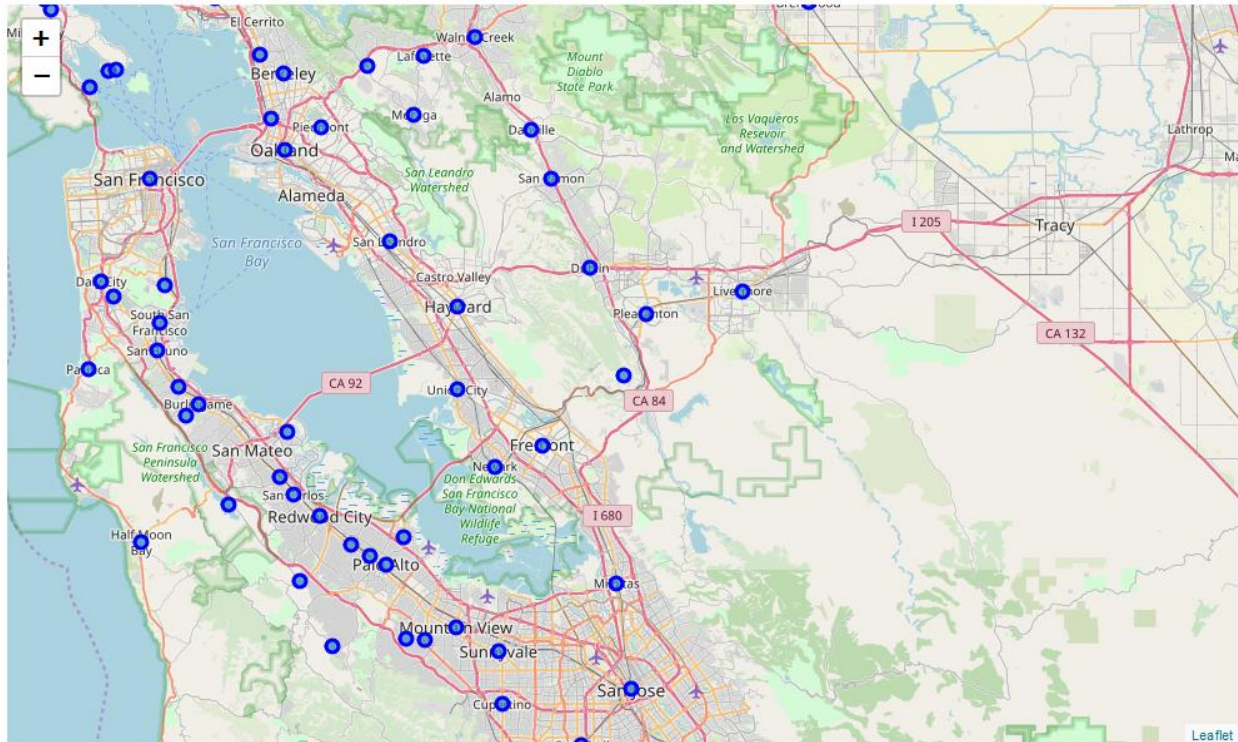
Figure 1. Folium map of some cities in the San Francisco Bay Area

I tried exploring some venues for the first city on the list, Alameda, but that did not result in any venues. I went to try the second city, Albany. And Albany returned 60 venues by Foursquare. Here is the list of the first 5 venues.

	name	categories	lat	lng
0	Sam's Log Cabin	Breakfast Spot	37.888589	-122.298258
1	Potala Organic Cafe	Vegetarian / Vegan Restaurant	37.885131	-122.297013
2	Patisserie Rotha	Bakery	37.884811	-122.296931
3	Sprouts Farmers Market	Grocery Store	37.885157	-122.297564
4	Hal's Office	Café	37.890522	-122.295885

Figure 2. A preview of 5 venues for the Albany city

```
60 venues were returned by Foursquare.
```

Figure 3. Returns from Foursquare API resulting in 60 venues for Albany

After retrieving all the venues for the 101 cities, I ended up with 3,167 venues.

```
print(bayarea_venues.shape)
bayarea_venues.head()

(3167, 7)
```

Figure 4. The shape of all venues in all cities pulled from Foursquare

Grouping all the venues by cities gave me 92 rows implying that 9 cities did not have any venues associated with them.

San Pablo	22	22	22	22	22	22
San Rafael	52	52	52	52	52	52
San Ramon	20	20	20	20	20	20
Santa Rosa	45	45	45	45	45	45
Sausalito	34	34	34	34	34	34
Sebastopol	1	1	1	1	1	1
South San Francisco	49	49	49	49	49	49
St. Helena	42	42	42	42	42	42
Suisun City	22	22	22	22	22	22
Sunnyvale	51	51	51	51	51	51
Tiburon	37	37	37	37	37	37
Union City	18	18	18	18	18	18
Vacaville	32	32	32	32	32	32
Vallejo	14	14	14	14	14	14
Walnut Creek	95	95	95	95	95	95
Windsor	28	28	28	28	28	28
Woodside	10	10	10	10	10	10
Yountville	35	35	35	35	35	35

92 rows × 6 columns

Figure 5. Venues are grouped by cities, displaying 92 of the 101 cities.

From this list of 3,187 venues, there are 314 unique categories.

```
print('There are {} uniques categories.'.format(len(bayarea_venues['Venue Category'].unique())))

There are 308 uniques categories.
```

Figure 6. There 308 unique categories within all the venues

I used clustering to get a sense of how the cities were grouped together with K-Means algorithm. I chose 5 clusters as default.
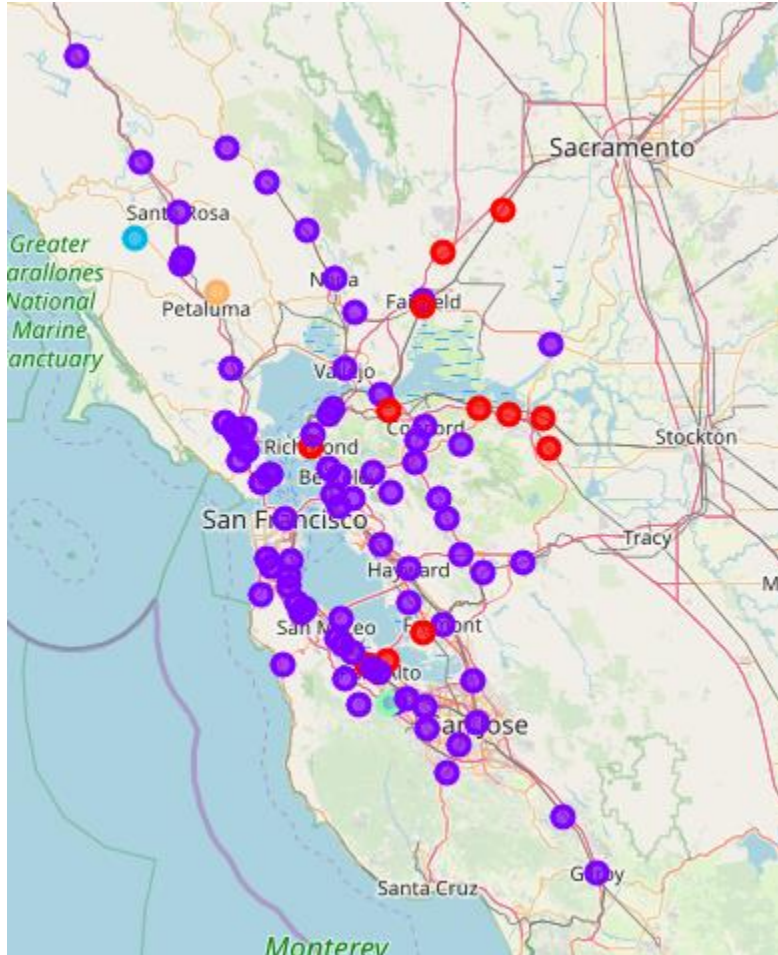

Figure 7. A folium map with 5 different clusters

Cluster 1

From this list of venues that fall under the cluster 1, it looks like most are Mexican restaurants.

	City	1st Most Common Venue
3	Antioch	Fast Food Restaurant
4	Atherton	Home Service
9	Brentwood	Mexican Restaurant
23	Dixon	Mexican Restaurant
25	East Palo Alto	Mexican Restaurant
44	Martinez	Coffee Shop
54	Newark	Mexican Restaurant
57	Oakley	Mexican Restaurant
64	Pittsburg	Mexican Restaurant
69	Richmond	Convenience Store
91	Suisun City	Mexican Restaurant
95	Vacaville	Bar

Figure 8. Several cities from cluster 1

Cluster 2
From this list of venues that fall under the cluster 1, it looks like a mixture of restaurants and coffee shop.

City	1st Most Common Venue
Albany	Coffee Shop
American Canyon	Winery
Belmont	Coffee Shop
Belvedere	Bakery
Benicia	Sushi Restaurant
Berkeley	Sushi Restaurant
Brisbane	Mexican Restaurant
Burlingame	Japanese Restaurant
Calistoga	Hotel
Campbell	Yoga Studio

Figure 9. Several cities from cluster 2

Cluster 3
Although only one venue fall under this cluster, it is a coffee shop.

City	1st Most Common Venue
Sebastopol	Coffee Shop

Figure 10. A city from cluster 3

Cluster 4
Although only one venue fall under this cluster, it is a home service.

City	1st Most Common Venue
Los Altos Hills	Home Service

Figure 11. A city from cluster 4

Cluster 5
Although only one venue fall under this cluster, it is a farm venue.

City	1st Most Common Venue
Petaluma	Farm

Figure 12. A city from cluster 5

Discussion

The San Francisco Bay Area has several big cities where there are a lot of job opportunities and a great place for settling down and starting a family. From the analysis of the data pulled from Wikipedia and Foursquare, all kinds of venues exist within these cities. Many have similar venues except for a few locations where there are wineries and farms. In terms of venues, it looks like there any city in the Bay Area are similar except for one or two cities.

After obtaining the cities, coordinates, and the venues, I used K-Means to cluster the different cities with the different venues. I used 5 different clusters in the machine learning algorithm but found that many with clustered in two main categories.

Conclusion

In conclusion, when people are moving to the San Francisco Bay Area to find jobs and start a family, looking at venues alone, pretty much any city in the Bay Area is a good fit. They all have similar venues that are present in all cities. Perhaps for a better decision on picking a city, it might be best to add in other features such as public transportations, job market, housing price, median rental rate, and school ratings for further analysis.