Introduction / Business Problem

The San Francisco Bay Area is a popular destination for people who want to live in a diverse cosmopolitan location with significant job opportunities and a comfortable climate. With years of development and growth, many cities have sprung up around San Francisco, contributing to a suburban sprawl that offers vastly distinctive options for potential residents. However, with the sheer number of cities, it is also increasingly difficult to research and find the best city or cities around which to focus a new home search. Data Science can be used collect and analyze data from disparate sources to arrive at a short-list of cities based on potential homeowner preferences.

Data

This analysis uses data from the following two sources.

Wikipedia – This is a good source of reference data compiled by an extensive user base. The data is often displayed in table format from a page written in html. A number of web-scraping techniques can be used to extract the data required for analysis. This analysis uses the html parser BeautifulSoup for this purpose.

1. Cities in the San Francisco Bay Area (sample screenshot below) (https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_the_San_Francisco_Bay_Area)

Name	Type	County	Population (2010)[8][9]	Land area[8]		Incorporated[7]
				sq mi	km^2	
Alameda	City	Alameda	73,812	10.61	27.5	April 19, 1854
Albany	City	Alameda	18,539	1.79	4.6	September 22, 1908
American Canyon	City	Napa	19,454	4.84	12.5	January 1, 1992
Antioch	City	Contra Costa	102,372	28.35	73.4	February 6, 1872
Atherton	Town	San Mateo	6,914	5.02	13.0	September 12, 1923
Belmont	City	San Mateo	25,835	4.62	12.0	October 29, 1926
Belvedere	City	Marin	2,068	0.52	1.3	December 24, 1896
Benicia	City	Solano	26,997	12.93	33.5	March 27, 1850
Berkeley	City	Alameda	112,580	10.47	27.1	April 4, 1878

2. Crime rates for cities in the San Francisco Bay Area (sample screenshot below) (https://en.wikipedia.org/wiki/California_locations_by_crime_rate)

City/Agency	County	Population[5]	Population density[5][3][note 2]	Violent crimes[5]	Violent crime rate per 1,000 persons	Property crimes[5]	Property crime rate per 1,000 persons
Adelanto	San Bernardino	31,213	557.3	189	6.06	790	25.31
Agoura Hills	Los Angeles	20,767	2,664.8	17	0.82	234	11.27
Alameda	Alameda	77,048	7,378.7	145	1.88	1,723	22.36
Albany	Alameda	19,350	10,822.1	31	1.6	478	24.7
Alhambra	Los Angeles	84,931	11,129.7	168	1.98	1,743	20.52
Aliso Viejo	Orange	50,671	7,323.5	35	0.69	273	5.39
Alturas	Modoc	2,615	1,073.9	29	11.09	89	34.03
American Canyon	Napa	20,379	3,351.3	55	2.7	568	27.87
Anaheim	Orange	346,956	6,942.3	1,101	3.17	8,196	23.62
Anderson	Shasta	10,176	1,597.0	96	9.43	617	60.63
Angels	Calaveras	3,716	1,024.3	7	1.88	48	12.92
Antioch	Contra Costa	108,223	3,820.1	849	7.84	4,190	38.72

FourSquare – This is a good source of information about venues of multiple types located around a specified latitude/longitude (sample screenshot below).

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Albany	37.88687	-122.297747	Sam's Log Cabin	37.888589	-122.298258	Breakfast Spot
1	Albany	37.88687	-122.297747	Potala Organic Cafe	37.885131	-122.297013	Vegetarian / Vegan Restaurant
2	Albany	37.88687	-122.297747	Patisserie Rotha	37.884811	-122.296931	Bakery
3	Albany	37.88687	-122.297747	Sprouts Farmers Market	37.885157	-122.297564	Grocery Store
4	Albany	37.88687	-122.297747	Hal's Office	37.890522	-122.295885	Café

Potential residents are likely to be interested in knowing about neighborhood venues when they decide where to relocate, so this information can be used to characterize each Bay Area city and to help focus a search on cities with desirable venues.

Methodology

The list of cities in the Bay Area was retrieved from Wikipedia, and the coordinates (latitude and longitude) for each city was retrieved using the Nominatim call from the geopy.geocoders library.

Venue Analysis

The city coordinates were subsequently used to retrieve the available venues for each city from FourSquare. The original list from Wikipedia had 101 cities but the resulting list of venues from FourSquare only had 94 cities because there were 7 cities for which FourSquare did not have any venue information. This is not surprising since the list of Bay Area cities includes fairly far-flung rural cities which are not typical destination hotspots.

Since there were 310 categories of venues, the next step involved determining the top venue categories for each city as a way to characterize the profile of the city. The list of venues for all the cities obtained from FourSquare was transformed using one-hot encoding, then grouped by city to produce a list of cities along with the average number of venues in each category for each city. This list was then sorted to identify the top venue categories for each city.

Let's assume potential residents are interested in cities with a number and variety of restaurants. Searching for cities with top venue categories that contained the string "Restaurant" identified the cities for which the top 1 through 5 venue categories are restaurants.

This analysis can easily be replicated for any other venue categories of interest to potential residents.

Crime Analysis

Potential residents are likely to want to know about crime rates in cities, with a natural preference for areas with low crime rates. Sorting the data by either violent crime rates or property crime rates resulted in lists of top 10 cities with low crime rates.

However, cities with low violent crime may not have low property crime, and vice versa. Looking for the intersection of the low violent crime and low property crime lists would be more informative in providing a list of cities with both low violent crime as well as low property crime.

Combined Venue-Crime Analysis

Building on the previous analyses of restaurants and crime, the analysis was extended to identify cities with the attractions of both many restaurants and low crime. A combined dataset was constructed by merging the list of cities with at least one restaurant category in its top 5 venues with the complete list of cities with their associated crime rates. This dataset was then sorted by violent crime to get the top 10 cities with many restaurants and low violent crime, and subsequently sorted by property crime to get the top 10 cities with many restaurants and low property crime. The intersection of these 2 top 10 lists resulted in the list of cities with many restaurants and low crime.

As a variant on this approach, the starting list for the merge was taken as the list of cities with either low violent crime or low property crime (once again, acknowledging that these are different lists), and merging with the list of cities with many restaurants. The intersection of these two lists was used to identify the cities with low crime and many restaurants.

Cluster Analysis

The cluster analysis used the original venue dataset from FourSquare, i.e. not just the subset of cities with many restaurants, and the original crime data set with all cities. In preparation for clustering, the numerical values representing instances of venues and city crime rates were normalized using the StandardScaler object to minimize the potential for skewing effects of large or small numbers.

k-means clustering was used on the combined dataset of venue information and crime rates. Different numbers of clusters were attempted, ranging from 3 to 10, to see if a particular number of clusters would yield more informative results.

Results

The following results were obtained from the analysis of cities with restaurants.

Number of top 5 venue categories that are restaurants	**# Cities**	**Cities**
4 or 5	0	
3	14	Benicia, Brisbane, Fairfield, Hayward, Larkspur, Los Altos, Mill Valley, Millbrae, Milpitas, Morgan Hill, Napa, Newark, Oakland, Orinda, Pleasant Hill, Pleasanton, San Bruno, San Carlos, San Rafael, South San Francisco, Tiburon, Vacaville

In total, the number of cities with at least 1 category of restaurant in its top 5 most common venues is 79.

The following results were obtained from the crime analysis.

Cities with Low Violent Crime	**Cities with Low Property Crime**
Monte Sereno	Ross
Hillsborough	*Monte Sereno*
Tiburon	*Los Altos Hills*
Orinda	Moraga
Los Altos	*Hillsborough*
Los Altos Hills	Saratoga
San Ramon	Windsor
Clayton	St. Helena
Danville	Cotati
Atherton	*Clayton*

The cities that show up on both top 10 lists, i.e. have low violent as well as property crime, are Monte Sereno, Hillsborough, Los Altos Hills, and Clayton.

The following results were obtained from the combined venue-crime analysis.

Cities with Many Restaurants and Low Violent Crime	Cities with Many Restaurants and Low Property Crime
Tiburon	*Orinda*
Orinda	*Los Altos*
Los Altos	*Tiburon*
Mill Valley	*Mill Valley*
Pleasanton	*Morgan Hill*
Benicia	Napa
Morgan Hill	*Pleasanton*
Milpitas	*Benicia*
Pleasant Hill	*South San Francisco*
South San Francisco	Newark

8 out of the top 10 cities are the same in both lists – Tiburon, Orinda, Los Altos, Mill Valley, Pleasanton, Benicia, Morgan Hill and South San Francisco.

The alternative approach to this analysis, constructing the list of cities with low violent and property crime as well as many restaurants by starting with the low crime lists, yielded three cities – Tiburon, Orinda, and Los Altos – all low in violent crime. In fact, there are no cities from the low property crime list that have many restaurants; only cities with low violent crime can have many restaurants.

The following results were obtained from the cluster analysis, using k-means clustering.

Cluster Description	# Cities	Violent Crime[1]	Property Crime[1]
1. Moderate violent crime, high property crime	5	7.77 – 8.65	38.72 – 53.03
2. Low violent crime, low-moderate property crime	34	0.0 – 1.62	8.48 – 20.77
3. Extremely high property crime, high violent crime	2	7.98 – 10.66	146.1 – 180.31
4. Extremely high violent crime, very high property crime	1	16.85	59.43
5. Low-moderate crime	30	1.56 – 4.86	9.29 – 30.2
6. Low-moderate violent crime, high property crime	14	1.1 – 4.71	28.06 – 50.58

[1] incidents per thousand residents

The clusters can be described in terms of crime rate levels, either for violent or property crime or both. They are listed and highlighted in the map below.

Cluster 1

	City	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Violent Crime per thousand	Property Crime per thousand
3	Antioch	0.0	Fast Food Restaurant	Gym	Coffee Shop	Flower Shop	Mexican Restaurant	7.84	38.72
69	Richmond	0.0	Convenience Store	Food Truck	Art Gallery	Grocery Store	Food	7.77	39.47
77	San Francisco	0.0	Coffee Shop	Hotel	Café	Cocktail Bar	Wine Bar	7.95	53.03
81	San Pablo	0.0	Pizza Place	Chinese Restaurant	Supermarket	Mexican Restaurant	Pharmacy	8.08	38.95
96	Vallejo	0.0	Chinese Restaurant	Yoga Studio	Park	Breakfast Spot	Music Venue	8.65	40.81

Cluster 2

	City	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Violent Crime per thousand	Property Crime per thousand
4	Atherton	1.0	Business Service	Spa	Mexican Restaurant	Food & Drink Shop	Baseball Field	0.42	10.39
5	Belmont	1.0	Pet Store	Sushi Restaurant	Coffee Shop	Grocery Store	Sandwich Place	1.34	13.8
6	Belvedere	1.0	Bakery	Deli / Bodega	Bay	Harbor / Marina	Chinese Restaurant	0.47	16.86
7	Benicia	1.0	Mexican Restaurant	Café	Wine Bar	American Restaurant	Italian Restaurant	0.94	17.43
12	Calistoga	1.0	Hotel	Bed & Breakfast	Wine Bar	American Restaurant	Bakery	0.57	13.83
14	Clayton	1.0	Sandwich Place	Gym	Steakhouse	Liquor Store	Bar	0.34	9.53
20	Cupertino	1.0	Chinese Restaurant	Coffee Shop	Hotel	Furniture / Home Store	Bank	0.66	16.94
22	Danville	1.0	Pizza Place	Sandwich Place	American Restaurant	Coffee Shop	Juice Bar	0.39	10.05
24	Dublin	1.0	Furniture / Home Store	Korean Restaurant	Thrift / Vintage Store	Men's Store	American Restaurant	1.3	14.83
30	Foster City	1.0	Fast Food Restaurant	Food Truck	Lake	Coffee Shop	Asian Restaurant	0.43	11.15
31	Fremont	1.0	Pizza Place	Bagel Shop	Grocery Store	Bakery	Falafel Restaurant	1.25	17.18
36	Hercules	1.0	Pub	Bay	Playground	American Restaurant	Asian Restaurant	1.08	11.43
37	Hillsborough	1.0	Farm	Business Service	Yoga Studio	Financial or Legal Service	Eye Doctor	0.09	9.05
38	Lafayette	1.0	Construction & Landscaping	Yoga Studio	Fish & Chips Shop	Eye Doctor	Falafel Restaurant	0.67	17.31
41	Los Altos	1.0	Pizza Place	Italian Restaurant	Mexican Restaurant	American Restaurant	Bakery	0.23	10.64
42	Los Altos Hills	1.0	Music Venue	Home Service	Yoga Studio	Fish & Chips Shop	Eye Doctor	0.24	8.68
43	Los Gatos	1.0	Hotel	Food	Pool	Baseball Field	Moving Target	0.75	20.12
45	Menlo Park	1.0	Coffee Shop	Café	Food Truck	Japanese Restaurant	Park	1.56	16.96
46	Mill Valley	1.0	Pizza Place	Italian Restaurant	American Restaurant	Coffee Shop	Indian Restaurant	0.76	13.05
49	Monte Sereno	1.0	Home Service	Yoga Studio	Fish & Chips Shop	Eye Doctor	Falafel Restaurant	0	8.54
50	Moraga	1.0	Coffee Shop	Sandwich Place	Italian Restaurant	Shipping Store	Farmers Market	0.47	8.85
51	Morgan Hill	1.0	Italian Restaurant	Brewery	Mexican Restaurant	American Restaurant	Burger Joint	1.56	14.72
55	Novato	1.0	Mexican Restaurant	Coffee Shop	Bakery	Bar	Breakfast Spot	1.46	14.06
57	Oakley	1.0	Grocery Store	Ice Cream Shop	Hawaiian Restaurant	Mexican Restaurant	Home Service	1.16	12.07
58	Orinda	1.0	Coffee Shop	American Restaurant	Burger Joint	Sushi Restaurant	Mexican Restaurant	0.21	9.89
60	Palo Alto	1.0	Café	Ice Cream Shop	Coffee Shop	French Restaurant	Japanese Restaurant	0.88	19.34
62	Piedmont	1.0	Pool	Art Gallery	Soccer Field	Lawyer	Theater	0.72	20.77
66	Pleasanton	1.0	Italian Restaurant	Ice Cream Shop	Sushi Restaurant	Mexican Restaurant	Coffee Shop	0.81	16.57
72	Ross	1.0	Park	Restaurant	Theater	Café	Deli / Bodega	1.62	8.48
73	St. Helena	1.0	American Restaurant	Pharmacy	Bank	Grocery Store	Spa	1	9.37
83	San Ramon	1.0	Sandwich Place	Grocery Store	Sushi Restaurant	Furniture / Home Store	American Restaurant	0.31	9.97
86	Saratoga	1.0	Food	Yoga Studio	Fish & Chips Shop	Eye Doctor	Falafel Restaurant	0.61	9.25
92	Sunnyvale	1.0	Coffee Shop	Grocery Store	Chinese Restaurant	Bank	Burger Joint	1.12	15.77
93	Tiburon	1.0	Chinese Restaurant	Italian Restaurant	Clothing Store	Hotel	American Restaurant	0.11	11.59

Cluster 3

	City	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Violent Crime per thousand	Property Crime per thousand
16	Colma	2.0	Flower Shop	Electronics Store	Hardware Store	Automotive Shop	Rental Car Location	7.98	180.31
27	Emeryville	2.0	Pet Store	Mobile Phone Shop	Bakery	Furniture / Home Store	Cupcake Shop	10.66	146.1

Cluster 4

click to scroll output; double click to hide

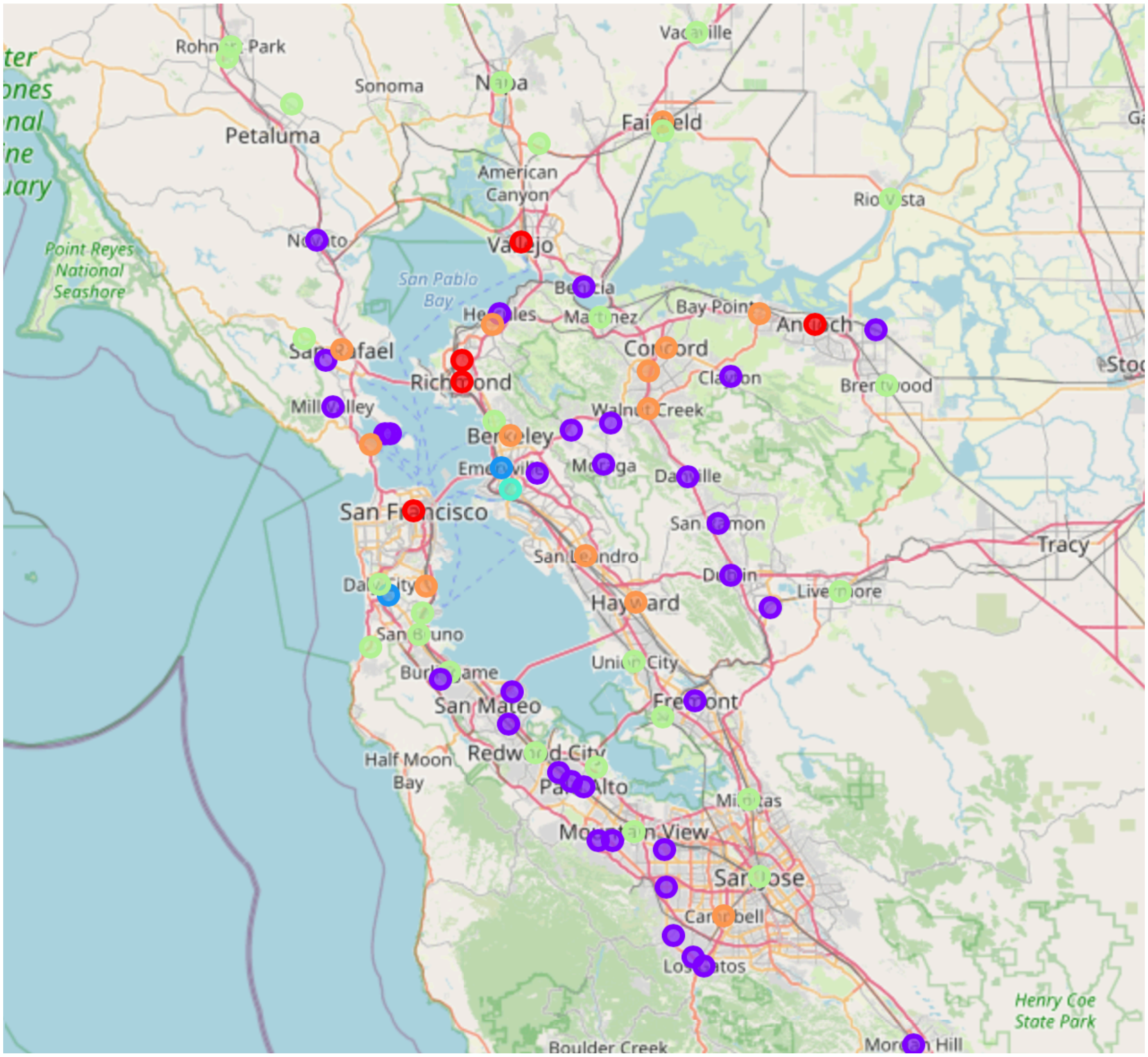
	City	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Violent Crime per thousand	Property Crime per thousand
56	Oakland	3.0	Bar	Chinese Restaurant	Japanese Restaurant	Sandwich Place	Vietnamese Restaurant	16.85	59.43

Cluster 5

	City	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Violent Crime per thousand	Property Crime per thousand
1	Albany	4.0	Pizza Place	Thai Restaurant	Coffee Shop	Burger Joint	Sushi Restaurant	1.6	24.7
2	American Canyon	4.0	Winery	Yoga Studio	Fish & Chips Shop	Eye Doctor	Falafel Restaurant	2.7	27.87
9	Brentwood	4.0	Pizza Place	Mexican Restaurant	American Restaurant	Bar	Sandwich Place	1.83	22.39
11	Burlingame	4.0	Japanese Restaurant	Italian Restaurant	Sandwich Place	Coffee Shop	Breakfast Spot	1.56	24.87
15	Cloverdale	4.0	Airport	Recreation Center	Skydiving Drop Zone	Yoga Studio	Financial or Legal Service	1.71	19.51
19	Cotati	4.0	Pizza Place	Music Store	Park	Bar	Karaoke Bar	4.45	9.43
21	Daly City	4.0	Sandwich Place	Fast Food Restaurant	Mexican Restaurant	Pizza Place	Gym / Fitness Center	1.84	15.95
23	Dixon	4.0	Mexican Restaurant	Sushi Restaurant	Bistro	Bakery	Auto Workshop	2.77	22.4
25	East Palo Alto	4.0	Mexican Restaurant	Bagel Shop	Gym / Fitness Center	Grocery Store	Market	4.22	19.54
28	Fairfax	4.0	Coffee Shop	Indian Restaurant	Italian Restaurant	Bar	Park	2.09	13.6
40	Livermore	4.0	Mexican Restaurant	Bar	Dive Bar	Coffee Shop	Ice Cream Shop	2.74	17.42
44	Martinez	4.0	Coffee Shop	Mexican Restaurant	American Restaurant	Plaza	Sandwich Place	1.95	26.03
48	Milpitas	4.0	Indian Restaurant	Vietnamese Restaurant	Korean Restaurant	Sandwich Place	Café	1.59	30.2
52	Mountain View	4.0	Coffee Shop	Sushi Restaurant	Bakery	Park	Yoga Studio	1.98	20.42
53	Napa	4.0	Wine Bar	American Restaurant	Italian Restaurant	Sushi Restaurant	Lounge	3.13	16.53
54	Newark	4.0	Mexican Restaurant	Asian Restaurant	Sandwich Place	Bubble Tea Shop	Chinese Restaurant	2.45	21.84
59	Pacifica	4.0	Garden	Grocery Store	BBQ Joint	Steakhouse	Trail	2.34	16.39
61	Petaluma	4.0	Farm	Dog Run	Yoga Studio	Fish & Chips Shop	Falafel Restaurant	3.34	17.96
68	Redwood City	4.0	Sandwich Place	Mexican Restaurant	Coffee Shop	Burger Joint	Grocery Store	2.37	21.11
70	Rio Vista	4.0	Gym	Post Office	Chinese Restaurant	BBQ Joint	American Restaurant	4.86	14.82
71	Rohnert Park	4.0	Pub	Disc Golf	Salon / Barbershop	Park	Athletics & Sports	3.85	17.7
75	San Bruno	4.0	Japanese Restaurant	Korean Restaurant	Grocery Store	Mexican Restaurant	Rental Car Location	2.57	24.63
78	San Jose	4.0	Mexican Restaurant	Cocktail Bar	Sandwich Place	Pub	Sushi Restaurant	3.21	24.34
85	Santa Rosa	4.0	Clothing Store	Brewery	Lingerie Store	Cosmetics Shop	Coffee Shop	3.68	22.26
90	South San Francisco	4.0	Mexican Restaurant	Coffee Shop	Italian Restaurant	Chinese Restaurant	Diner	2.34	19.07
91	Suisun City	4.0	Mexican Restaurant	American Restaurant	Trail	Deli / Bodega	Convenience Store	2.35	20.91
94	Union City	4.0	Yoga Studio	Filipino Restaurant	Park	Coffee Shop	Fried Chicken Joint	2.83	21.86
95	Vacaville	4.0	Mexican Restaurant	Bar	Sandwich Place	Sushi Restaurant	Italian Restaurant	2.83	27.32
98	Windsor	4.0	Coffee Shop	Mexican Restaurant	Market	BBQ Joint	Indian Restaurant	3.14	9.29
100	Yountville	4.0	Wine Bar	Hotel	Bakery	Deli / Bodega	French Restaurant	2.35	13.43

Cluster 6

	City	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Violent Crime per thousand	Property Crime per thousand
8	Berkeley	5.0	Sushi Restaurant	Theater	Brewery	Music Venue	Electronics Store	3.66	43.33
10	Brisbane	5.0	Mexican Restaurant	Deli / Bodega	Vietnamese Restaurant	Indian Restaurant	Paper / Office Supplies Store	2.45	29.67
13	Campbell	5.0	Yoga Studio	Mexican Restaurant	Sandwich Place	Italian Restaurant	Cosmetics Shop	1.98	33.96
17	Concord	5.0	Mexican Restaurant	Indian Restaurant	Café	Coffee Shop	Food Truck	3.67	41
29	Fairfield	5.0	Chinese Restaurant	Indian Restaurant	Bank	Thai Restaurant	Beer Bar	4.71	35.11
32	Gilroy	5.0	American Restaurant	Tapas Restaurant	Theater	Shipping Store	Train Station	3.76	28.35
34	Hayward	5.0	Fast Food Restaurant	Bar	Vietnamese Restaurant	Pizza Place	Mexican Restaurant	3.95	31.78
63	Pinole	5.0	Liquor Store	Spa	Sporting Goods Shop	Comic Shop	Chinese Restaurant	3.63	33.2
64	Pittsburg	5.0	Mexican Restaurant	Fast Food Restaurant	Park	Supermarket	Fried Chicken Joint	2.59	34.99
65	Pleasant Hill	5.0	Sushi Restaurant	Burger Joint	American Restaurant	Pizza Place	Chinese Restaurant	1.75	50.58
79	San Leandro	5.0	Pharmacy	Sushi Restaurant	Coffee Shop	Burger Joint	Mexican Restaurant	4.16	42.36
82	San Rafael	5.0	Thai Restaurant	Indian Restaurant	Coffee Shop	Mexican Restaurant	Bar	3.26	28.06
87	Sausalito	5.0	Café	Thai Restaurant	Seafood Restaurant	Coffee Shop	Pizza Place	2.94	32.51
97	Walnut Creek	5.0	Coffee Shop	Pizza Place	Ice Cream Shop	American Restaurant	Italian Restaurant	1.1	36.1



Discussion

There are two limitations in using FourSquare data that were encountered in this project.

1. The quality of the venue analysis is the degree of dependence on the labels used for each venue category. For instance, in searching for cities with restaurants, the analysis excluded categories like “Café”, “Deli”, and “Pizza Place” because these labels did not contain the term “Restaurant” although arguably they should have been included as dining establishments.
2. FourSquare data is live and updated frequently so conclusions drawn from one set of results may change over a short period of time.

The analyses of venues and crime rates illustrate typical trade-offs faced by potential residents as they are forced to prioritize multiple desirable attributes of cities. If low crime is of paramount importance while having many restaurants in the neighborhood is a strong preference, recommended cities to consider are (in descending order):

1. Tiburon, Orinda, and Los Altos – cities with low crime and many restaurants
2. Mill Valley, Pleasanton, Benicia, and Morgan Hill– cities with many restaurants that are relatively low on crime
3. The 27 remaining cities in cluster 2 ("low crime") of the cluster analysis since this is the cluster shared by the cities listed above in 1. and 2.

If having many restaurants in the vicinity is the most important attribute and crime rates matter less, then also consider the cities that came up on top in the restaurant venue analysis – Brisbane, Fairfield, Hayward, Larkspur, Millbrae, Milpitas, Napa, Newark, Oakland, Pleasant Hill, San Bruno, San Carlos, San Rafael, South San Francisco, and Vacaville.

Despite the normalization of all the numeric values used in clustering, the crime data still seems to have carried more weight than the venue data in the clustering process. This may be because there were so many venue categories (310) that the instance data was spread too thin across many possible values (categories) relative to the crime data which had only two values (violent or property).

Conclusion

Data Science can be applied to readily-available locational information, e.g. cities, venues within cities, and crime rates by city, in order to arrive at short lists of cities based on different criteria. In this analysis, the focus was on potential residents looking to settle in cities with desirable characteristics. It was demonstrated that multiple different analyses could be used to enable alternative perspectives on the information available, with the ultimate objective of providing insights on which to base decisions.

Further work in this area could include the development of a front-end user interface that would allow the selection of criteria, in this case venue categories of interest, so that alternative custom analyses can be conducted. Another potential area for expansion could be the inclusion of more data sources to complement the venue and crime data and that may be relevant for different potential residents, e.g. information on the housing market by city (house prices, % buyers vs. renters, etc.), information on schools within each city, and so on.

The tools of Data Science today enable effective data manipulation and analysis by applying the power of computing to massive quantities of information.