San Francisco Bay Area Analysis

Data Science Capstone Project

Ruth Werner

Agenda

- Introduction / Business Problem
- Data
- Methodology / Results
- Results Summary
- Conclusion

Introduction / Business Problem

- The San Francisco Bay Area is a popular destination for homeowners
  - Diverse cosmopolitan location
  - Significant job opportunities
  - Comfortable climate
- Suburban sprawl with many cities offers vastly distinctive options
- Difficult to leverage massive amounts of data to narrow down cities
- Data Science can help focus a new home search

Data: Wikipedia

Cities

Name	Type	County	Population (2010)[8][9]	Land area[8]		Incorporated[7]
				sq mi	km²	
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

https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_the_San_Francisco_Bay_Area)

Crime Rates

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

https://en.wikipedia.org/wiki/California_locations_by_crime_rate)

Data: FourSquare

Venues of multiple types located around a specified latitude/longitude

	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

Methodology / Results: Venue Analysis

- Focus on restaurant categories, as example of attribute of interest to potential residents
- Cities with restaurants as at least 3 out of top 5 most common venues:

1. Benicia
2. Brisbane
3. Fairfield
4. Hayward
5. Larkspur
6. Los Altos
7. Mill Valley
8. Millbrae
9. Milpitas
10. Morgan Hill
11. Napa
12. Newark
13. Oakland
14. Orinda
15. Pleasant Hill
16. Pleasanton
17. San Bruno
18. San Carlos
19. San Rafael
20. South San Francisco,
21. Tiburon
22. Vacaville

Methodology / Results: Crime Analysis

Cities Lowest in Violent Crime

1. Monte Sereno
2. Hillsborough
3. Tiburon
4. Orinda
5. Los Altos
6. Los Altos Hills
7. San Ramon
8. Clayton
9. Danville
10. Atherton

Cities Lowest in Property Crime

1. Ross
2. Monte Sereno
3. Los Altos Hills
4. Moraga
5. Hillsborough
6. Saratoga
7. Windsor
8. St. Helena
9. Cotati
10. Clayton

- Cities low in both violent and property crimes:
  - Monte Sereno, Hillsborough, Los Altos Hills, Clayton

Methodology / Results: Venue-Crime Analysis

- Low crime rates alone may not be sufficient to make cities attractive, so combine venue and crime analyses

Cities with Many Restaurants that are Low in Violent Crime

1. Tiburon
2. Orinda
3. Los Altos
4. Mill Valley
5. Pleasanton
6. Benicia
7. Morgan Hill
8. Milpitas
9. Pleasant Hill
10. South San Francisco

Cities with Many Restaurants that are Low in Property Crime

1. Orinda
2. Los Altos
3. Tiburon
4. Mill Valley
5. Morgan Hill
6. Napa
7. Pleasanton
8. Benicia
9. South San Francisco
10. Newark

- Low crime cities with many restaurants – Tiburon, Orinda, Los Altos

Methodology / Results: Cluster Analysis

- Results of k-means cluster analysis heavily weighted crime rates relative to venues despite normalization, probably because of the large number of venue categories (310)

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

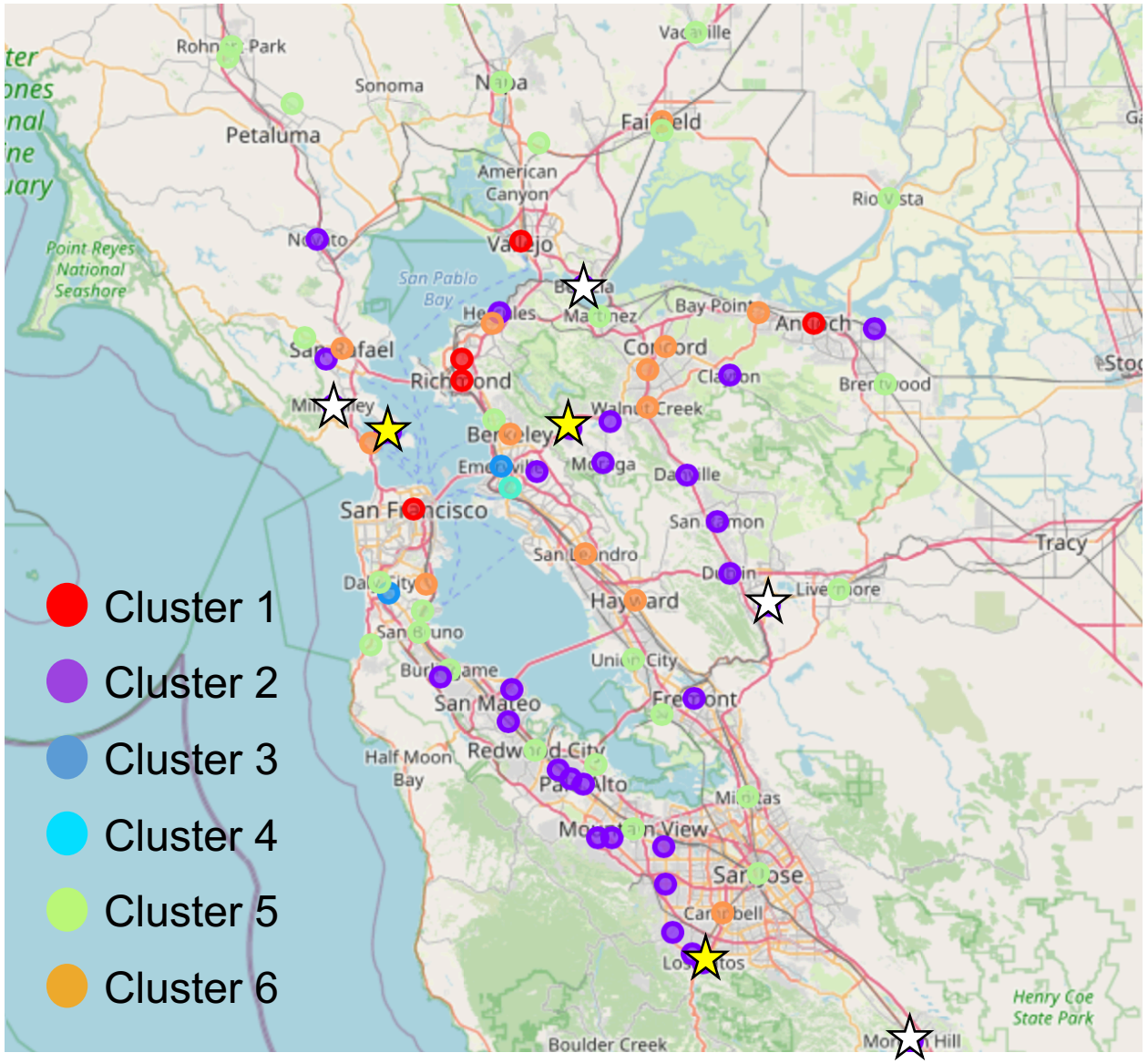
Methodology / Results: Cluster Analysis (cont'd)

Cluster	Cities
1. Moderate violent crime, high property crime	Antioch, Richmond, San Francisco, San Pablo, Vallejo
2. Low violent crime, low-moderate property crime	Atherton, Belmont, Belvedere, Benicia, Calistoga, Clayton, Cupertino, Danville, Dublin, Foster City, Fremont, Hercules, Hillsborough, Lafayette, Los Altos, Los Altos Hills, Los Gatos, Menlo Park, Mill Valley, Monte Sereno, Moraga, Morgan Hill, Novato, Oakley, Orinda, Palo Alto, Piedmont, Pleasanton, Ross, St. Helena, San Ramon, Saratoga, Sunnyvale, Tiburon
3. Extremely high property crime, high violent crime	Colma, Emeryville
4. Extremely high violent crime, very high property crime	Oakland
5. Low-moderate crime	Albany, American Canyon, Brentwood, Burlingame, Cloverdale, Cotati, Daly City, Dixon, East Palo Alto, Fairfax, Livermore, Martinez, Milpitas, Mountain View, Napa, Newark, Pacifica, Petaluma, Redwood City, Rio Vista, Rohnert Park, San Bruno, San Jose, Santa Rosa, South San Francisco, Suisun City, Union City, Vacaville, Windsor, Yountville
6. Low-moderate violent crime, high property crime	Berkeley, Brisbane, Campbell, Concord, Fairfield, Gilroy, Hayward, Pinole, Pittsburg, Pleasant Hill, San Leandro, San Rafael, Sausalito, Walnut Creek

Results Summary

Recommended Cities

1. Tiburon, Orinda, and Los Altos – low crime, many restaurants
2. Mill Valley, Pleasanton, Benicia, and Morgan Hill – many restaurants, relatively low crime
3. Remaining cities in Cluster 2 – similar crime profile, mixed frequency of restaurants



Conclusion

- This project shows how Data Science can be applied to readily-available locational information, e.g. cities, venues within cities, and crime rates by city, to arrive at short lists of cities based on different criteria.
  - These analyses enable alternative perspectives on the information to provide insights on which to base decisions.
- Further work can include:
  - Development of a front-end interface to allow user selection of criteria, e.g. venue categories, for custom analyses
  - Inclusion of other data sources relevant for potential residents, e.g. housing market by city (house prices, % buyers vs. renters, etc.), 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.