

A. Introduction

A.1. Description & Discussion of the Background

Santa Clara County, officially the County of Santa Clara, is California's 6th most populous county, with a population of 1,781,642, as of the 2010 census.[3] The county seat and largest city is San Jose, the 10th most populous city in the United States, California's 3rd most populous city and the most populous city in the San Francisco Bay Area.

Home to Silicon Valley, Santa Clara County is an economic center for high technology and has the third highest GDP per capita in the world (after Zurich, Switzerland and Oslo, Norway), according to the Brookings Institution. The county's concentration of wealth, primarily due to the tech industry, has made it the most affluent county on the West Coast of the United States and one of the most affluent places in America.

Santa Clara County is part of the San Jose-Sunnyvale-Santa Clara, CA Metropolitan Statistical Area as well as the San Jose-San Francisco-Oakland, CA Combined Statistical Area. Located on the southern coast of San Francisco Bay, the urbanized Santa Clara Valley within Santa Clara County is also known as Silicon Valley. Santa Clara is the most populous county in the San Francisco Bay Area and in Northern California.

Santa Clara is a city in Santa Clara County, California. The city's population was 116,468 as of the 2010 United States Census, making it the ninth-most populous city in the San Francisco Bay Area. Located on the southern coast of San Francisco Bay immediately west of San Jose and 45 miles (72 km) southeast of San Francisco, the city was founded in 1777 with the establishment of Mission Santa Clara de Asís, the eighth of 21 California missions. The city was later incorporated in 1852. The mission, the city, and the county are all named for Saint Clare of Assisi.

Santa Clara is located in the center of Silicon Valley and is home to the headquarters of several high-tech companies such as Intel. It is also home to Santa Clara University, the oldest institution of higher learning in the state of California, which was built around Mission Santa Clara de Asís. Levi's Stadium, the home of the National Football League's San Francisco 49ers, is located in the city. Santa Clara is bordered by San Jose on all sides, except for Sunnyvale to the west, and Cupertino to the west.

As you can see from the figures, Santa Clara is a city with high population in a region which is hub to a thriving economy. Being such a popular city with businesses and companies has led to high housing rental prices. The real estate cost is very steep and living near the office comes at a big hit on the paycheck.

As my office is situated in Santa Clara, I chose the city as one of the center and will compare its offerings of restaurants and nightlife with other cities in the San Jose

County area such as San Jose, Sunnyvale, Cupertino, Palo Alto to see which other areas offer similar services and can be alternative to Santa Clara.

When we consider the problem of alternate housing location, we can create a map and information chart where the real estate index is placed on Santa Clara County and each city is clustered according to the venue density.

A.2. Data Description

To consider the problem we can list the datas as below:

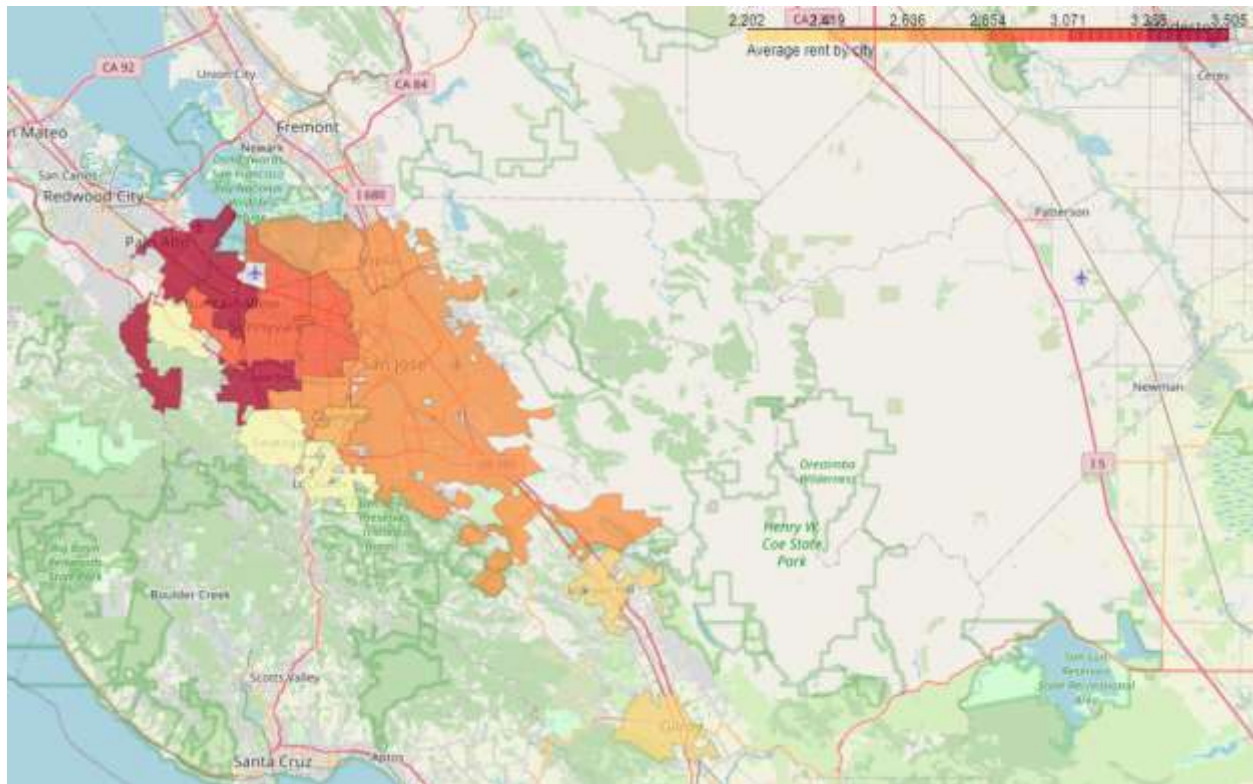
- I found the Spatial data from Open Data Portal of County of Santa Clara.
- The geojson file had coordinates of the cities in the county. I cleaned the data and reproduced it.
- I used **Foursquare API** to get the most common venues of selected Cities.
- I found the Average Rent in Santa Clara, Santa Clara County, CA By Neighborhood from RentCafe.com which has rent trends for US cities.

B. Methodology

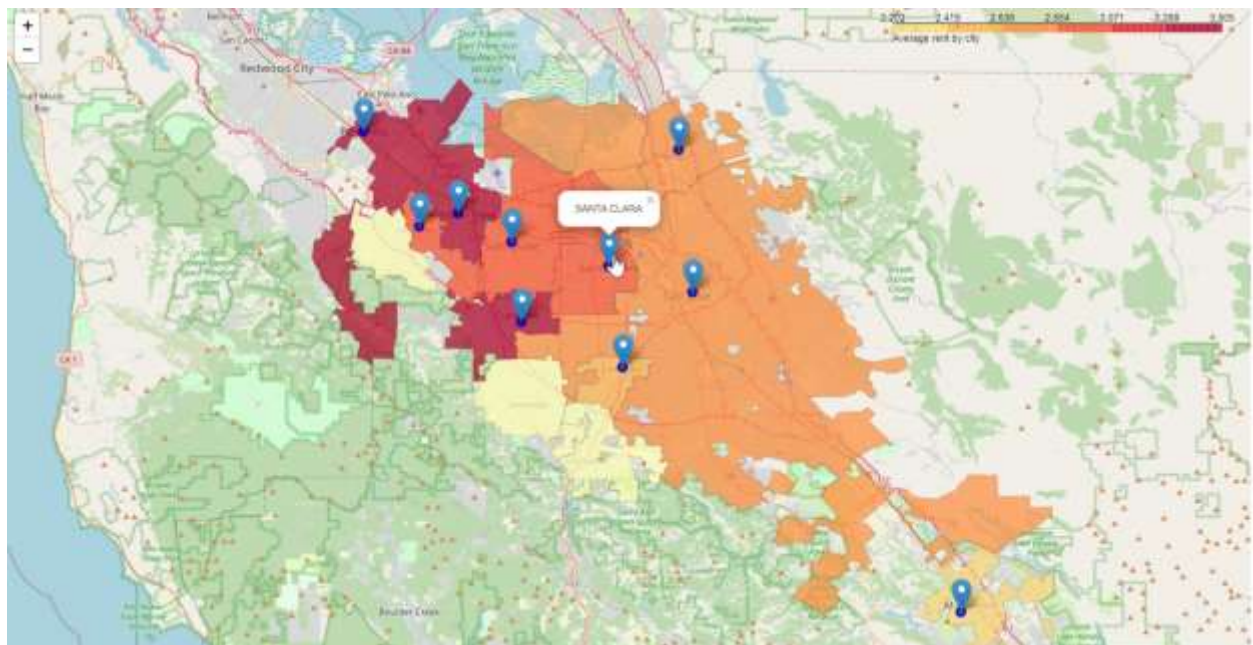
As a database, I used Open Data Portal of County of Santa Clara and Rent Trends from Rentcafe.com in my study. My master data which has the main components *City*, *average rent*, *zip*, *latitude* and *longitude* information of the county.

	CITY	Average_Rent	ZIP	LATITUDE	LONGITUDE
0	SANTA CLARA	2974	95050	37.355531	-121.955573
1	GILROY	2215	95020	37.004815	-121.572442
2	MORGAN HILL	2390	95037	37.125569	-121.662109
3	CAMPBELL	2633	95008	37.288126	-121.944362
4	SAN JOSE	2790	95113	37.337845	-121.886027
5	MILPITAS	2853	95035	37.432342	-121.897441
6	SUNNYVALE	3016	94086	37.371408	-122.036903
7	LOS ALTOS	3060	94022	37.381406	-122.113581
8	MOUNTAIN VIEW	3291	94039	37.390176	-122.081790
9	PALO ALTO	3347	94301	37.444255	-122.160528
10	CUPERTINO	3492	95014	37.318313	-122.028747

I used python **folium** library to visualize the county and its cities and I created a heat map based on average rent by city. I used latitude and longitude values to get the visual as below:



After that I superimposed the cities as markers with pop-up labels to better identify the city locations in the county.



I utilized the Foursquare API to explore the boroughs and segment them. I set the limit as **100 venue** and the radius **1000 meter** radius for each city from the city hall. Here is

a brief reflection of the list of venue, category, its latitude and longitude information from Foursquare API.

(389, 7)

	CITY	CITY Latitude	CITY Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	SANTA CLARA	37.355531	-121.955573	Kettle'e	37.352102	-121.954609	Indian Restaurant
1	SANTA CLARA	37.355531	-121.955573	BEST WESTERN University Inn Santa Clara	37.352545	-121.955239	Hotel
2	SANTA CLARA	37.355531	-121.955573	Maple Leaf Donuts	37.355301	-121.959146	Donut Shop
3	SANTA CLARA	37.355531	-121.955573	Triton Museum of Art	37.355971	-121.955246	Art Museum
4	SANTA CLARA	37.355531	-121.955573	Chick-fil-A	37.352039	-121.958524	Fast Food Restaurant

In summary of this data **389** venues were returned by Foursquare. Here is a summary table of venues by city.

	CITY	CITY Latitude	CITY Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
	CAMPBELL	45	45	45	45	45	45
	CUPERTINO	19	19	19	19	19	19
	GILROY	16	16	16	16	16	16
	LOS ALTOS	48	48	48	48	48	48
	MILPITAS	46	46	46	46	46	46
	MORGAN HILL	7	7	7	7	7	7
	MOUNTAIN VIEW	38	38	38	38	38	38
	PALO ALTO	73	73	73	73	73	73
	SAN JOSE	37	37	37	37	37	37
	SANTA CLARA	16	16	16	16	16	16
	SUNNYVALE	44	44	44	44	44	44

We can see that Palo Alto had 73 Venues (under given limitations), which was the highest among the cities. Morgan Hill had the least number of Venues (7). From the table we can see that the distribution is clustered more around the range of 30-50.

The 389 venues were split into **141** unique categories. From this, I created a table which shows list of top 10 category for each City in below table by taking into account the frequency for each category as follows:

```

----CAMPBELL----
      venue  freq
0    Sandwich Place 0.07
1  Italian Restaurant 0.07
2  Mexican Restaurant 0.07
3    Ice Cream Shop 0.04
4      Pizza Place 0.04

```

	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	CAMPBELL	Sandwich Place	Italian Restaurant	Mexican Restaurant	Pizza Place	Ice Cream Shop	Boutique	Frozen Yogurt Shop	Record Shop	Arcade	Park
1	CUPERTINO	Bank	Mexican Restaurant	Coffee Shop	Pharmacy	Chinese Restaurant	Electronics Store	Sandwich Place	Mattress Store	Massage Studio	Spa
2	GILROY	American Restaurant	Mexican Restaurant	Cafe	Breakfast Spot	Bar	Wine Bar	Bookstore	Indian Restaurant	Coffee Shop	Bus Station
3	LOS ALTOS	Bakery	Coffee Shop	Sushi Restaurant	Mexican Restaurant	Pizza Place	American Restaurant	Bank	Bookstore	Breakfast Spot	Greek Restaurant
4	MILPITAS	Shopping Plaza	Grocery Store	Mexican Restaurant	Optical Shop	Sandwich Place	Chinese Restaurant	Bank	Coffee Shop	Seafood Restaurant	Salon / Barbershop
5	MORGAN HILL	Ice Cream Shop	Grocery Store	Convenience Store	Chinese Restaurant	Sushi Restaurant	Pizza Place	Vietnamese Restaurant	Fast Food Restaurant	Concert Hall	Hobby Shop
6	MOUNTAIN VIEW	Coffee Shop	Park	Yoga Studio	Pool	Optical Shop	Sushi Restaurant	Health & Beauty Service	Salon / Barbershop	Salad Place	Pastery Shop
7	PALO ALTO	Cafe	Ice Cream Shop	Coffee Shop	Hotel	American Restaurant	Juice Bar	Gym / Fitness Center	Spa	Greek Restaurant	Pizza Place
8	SAN JOSE	Sandwich Place	Ice Cream Shop	Vietnamese Restaurant	Mexican Restaurant	Coffee Shop	Convenience Store	Pub	Pharmacy	Music Store	Mediterranean Restaurant
9	SANTA CLARA	Fast Food Restaurant	Fried Chicken Joint	Indian Restaurant	Bank	Hotel	Breakfast Spot	Intersection	Art Museum	Italian Restaurant	Food Truck
10	SUNNYVALE	Grocery Store	Coffee Shop	Chinese Restaurant	Indian Restaurant	Spa	Bank	Hotel Restaurant	Rental Car Location	Shoe Repair	Dumping Restaurant

We have some common venue categories amongst the cities. Due to this reason I used **K-means algorithm** to cluster the cities. K-Means is one of the most common clustering method and provides good results as well.

I divided the cities into **3** clusters. Here is the merged table with cluster labels for each borough.

	CITY	ZIP	LATITUDE	LONGITUDE	Cluster Labels	Average Rent	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	SANTA CLARA	95050	37.355531	-121.955573	2	2974	Fast Food Restaurant	Fried Chicken Joint	Indian Restaurant	Bank	Hotel	Breakfast Spot	Intersection	Art Museum	Italian Restaurant	Food Truck
1	GILROY	95020	37.004815	-121.572442	2	2215	American Restaurant	Mexican Restaurant	Cafe	Breakfast Spot	Bar	Wine Bar	Bookstore	Indian Restaurant	Coffee Shop	Bus Station
2	MORGAN HILL	95037	37.125569	-121.862109	2	2390	Ice Cream Shop	Grocery Store	Convenience Store	Chinese Restaurant	Sushi Restaurant	Pizza Place	Vietnamese Restaurant	Fast Food Restaurant	Concert Hall	Hobby Shop
3	CAMPBELL	95008	37.280126	-121.944362	2	2633	Sandwich Place	Italian Restaurant	Mexican Restaurant	Pizza Place	Ice Cream Shop	Boutique	Frozen Yogurt Shop	Record Shop	Arcade	Park
4	SAN JOSE	95113	37.337845	-121.886027	2	2790	Sandwich Place	Ice Cream Shop	Vietnamese Restaurant	Mexican Restaurant	Coffee Shop	Convenience Store	Pub	Pharmacy	Music Store	Mediterranean Restaurant

C. Results

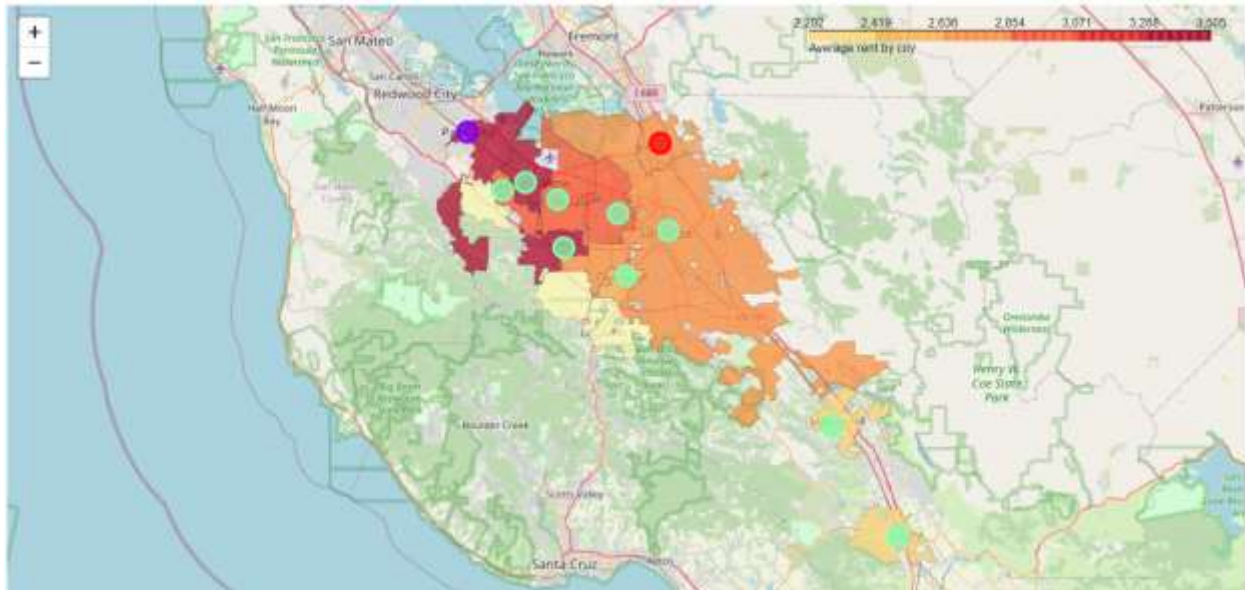
In clustering, I also took average rent into consideration with the venues as rent is a crucial factor for selecting an accommodation and city for my study.

The clustering showed that areas surrounding Santa Clara City have a similar lifestyle offering and an individual or business can choose an alternative from the cities surrounding Santa Clara City if lifestyle is of high preference. Palo Alto offers a different lifestyle and hence it was not part of the cluster of Santa Clara City which is reflected by the highest rental value for accommodation and distance from Santa Clara City.

I plotted the clustered cities on a Choropleth map.

In the final section, the Choropleth map contained the necessary information for the study for each city in the county of Santa Clara:

- County Borders
- City Borders
- Average Rental Price (USD)
- Optimum alternative locations to Santa Clara City



D. Discussion

Santa Clara County is a pretty big and thriving country with a high population density in a narrow area. The total number of measurements and population densities of the 11 cities in total can vary. As there is a lot of complexity, multiple different approaches can be tried in clustering and classification studies.

I used the K-means algorithm as part of this clustering study. For more detailed and accurate guidance, the data set can be expanded and the details of the neighborhood or street can also be drilled.

I ended the study by visualizing the data and clustering information on the Santa Clara County map. In future studies, web or mobile applications can be carried out to direct people who are finding accommodation in the county.

Using these tools, it was observed that surrounding cities have similar features while distant cities offer different characteristics.

F. Conclusion

People are turning to clusters where companies are starting business. This is driving the realty prices through the roof and people find it difficult to make an informed choice for their accommodation. People can make an informed choice if they have access to a lifestyle comparison platform which suggests them areas based on their lifestyle.

Business owners can also leverage this to make smarter investments and ventures. They can identify the gaps in an area and try to fulfill them in that region or join a saturated market for a safer ROI.

G. References:

- [Foursquare API](#)
- [Rentcafe – Rental prices Santa Clara County](#)
- [County of Santa Clara – Open Data Portal](#)