UNDERVALUED ZIP CODES OF SILICON VALLEY

A CAPSTONE PROJECT FOR THE

IBM PROFESSIONAL CERTIFICATE IN DATA SCIENCE

-

ALISSA MCFARLIN

JUNE 2021

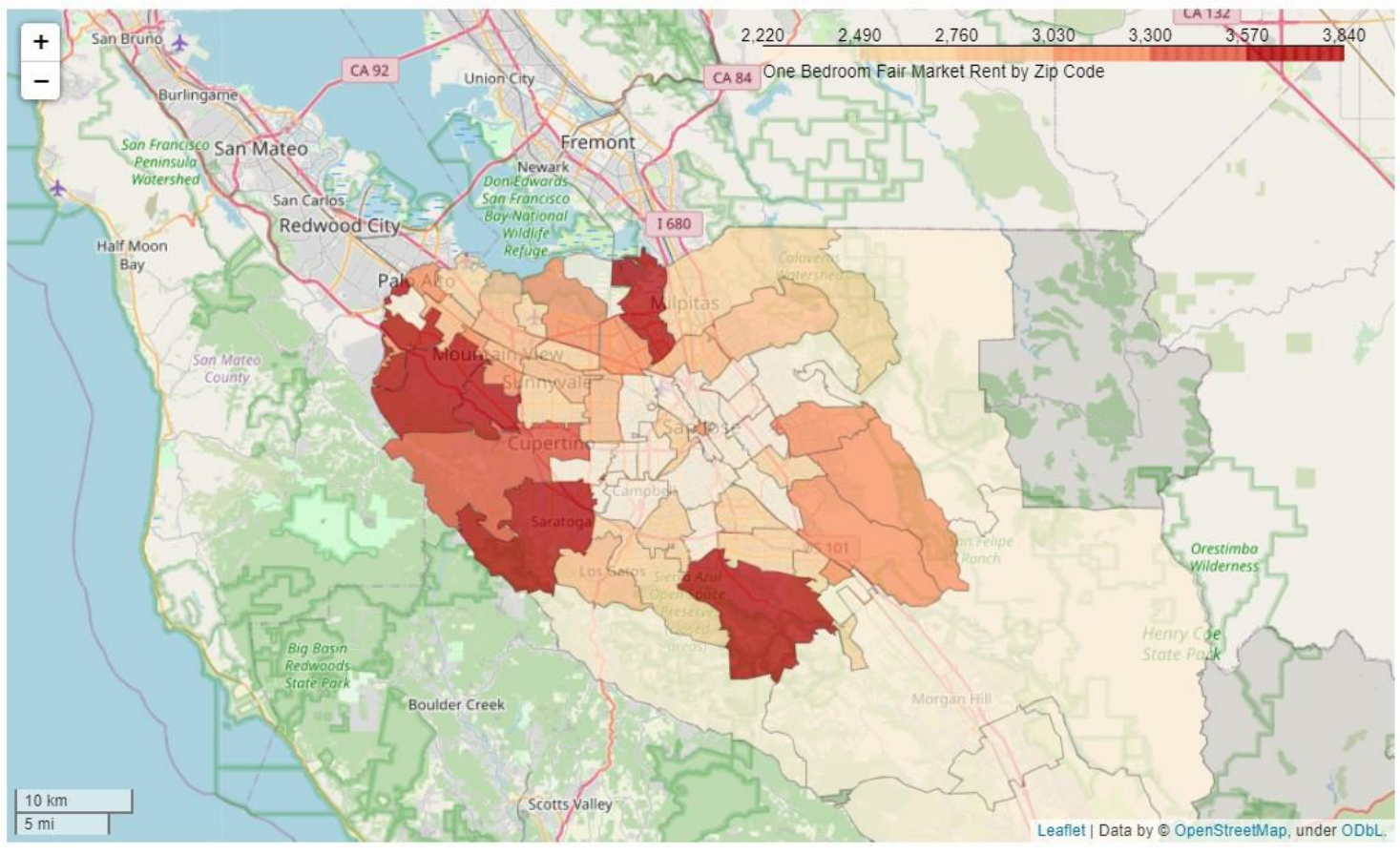
WHERE IN SILICON VALLEY ARE THE BEST PLACES FOR INVESTMENT IN APARTMENT BUILDINGS?

- Silicon Valley rents: high and getting higher
- Which ZIP Codes offer the most potential for gain?
- Group ZIP Codes by mix of restaurants, stores, and amenities
- ZIP Codes with venue mix of expensive areas but with lower rents = POTENTIAL!

COLLECT AND EXTRACT DATA

- US Dept. of Housing and Urban Development Fair Market Rent (FMR)
  - ZIP Code-level data
  - One-bedroom FMR = consistent, accurate rent metric
- Foursquare API
  - Rich venue information, including venue categories
  - ZIP Code coordinates used to extract nearby venue data
- Calculate proportion of each ZIP Code's venues belonging to each category

SILICON VALLEY ZIP CODES BY FAIR MARKET RENT



FOURSQUARE VENUE CATEGORY DATA

- Rural ZIP Codes not in Silicon Valley excluded
- 4057 venues for 57 ZIP Codes
- 331 unique venue categories
- Categories one hot encoded, grouped by ZIP Code, and proportions calculated

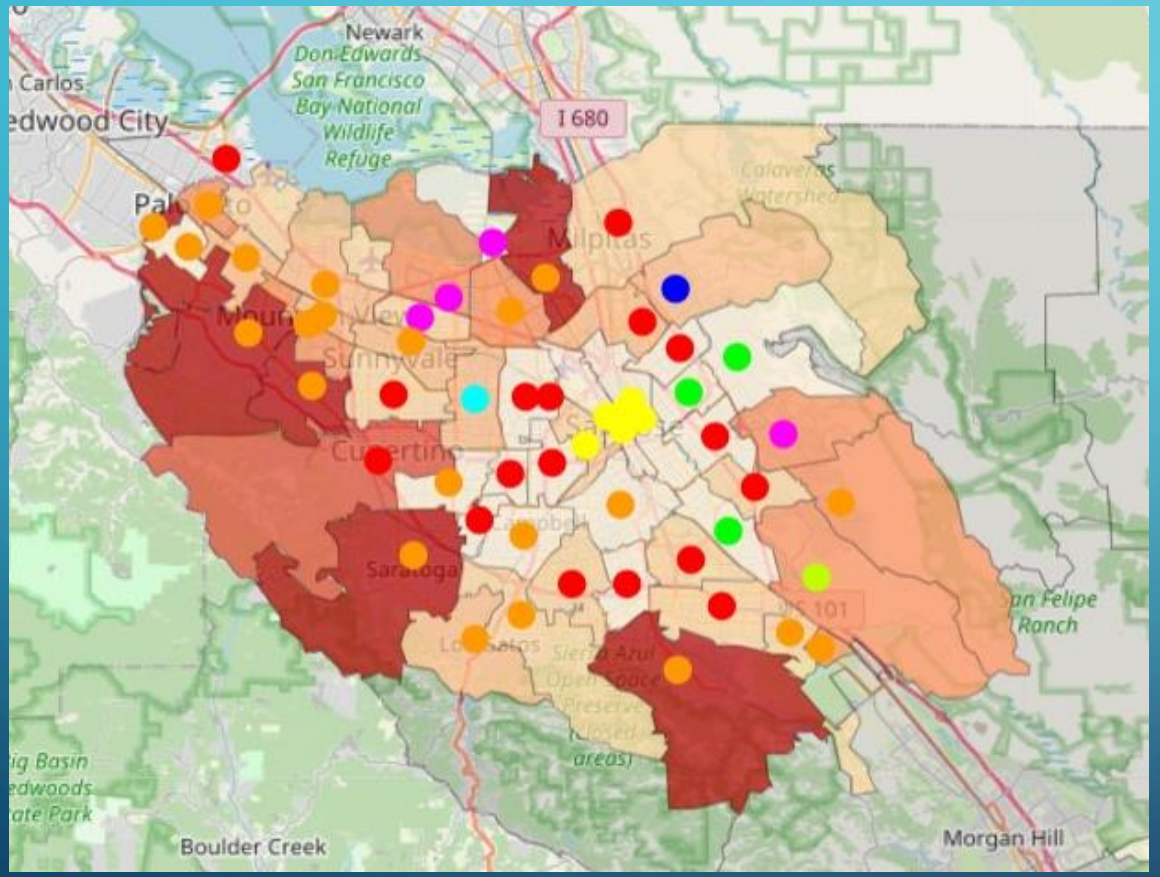
	zipcode	zip_lat	zip_long	venue	venue_lat	venue_long	venue_category
0	94022	37.3814	-122.1258	Tin Pot Creamery	37.378643	-122.118322	Ice Cream Shop
1	94022	37.3814	-122.1258	Linden Tree	37.379216	-122.116963	Bookstore
2	94022	37.3814	-122.1258	Manresa Bread	37.379234	-122.117080	Bakery
3	94022	37.3814	-122.1258	Asa	37.379490	-122.116821	Spanish Restaurant
4	94022	37.3814	-122.1258	Satura Cakes	37.378835	-122.115860	Bakery

	zipcode	ATM	Accessories Store	Acupuncturist	Afghan Restaurant	Airport Terminal	American Restaurant	Andhra Restaurant	Arcade	Art Gallery	...	Vietnamese Restaurant	Volleyball Court	Warehouse Store
0	94022	0.020833	0.0	0.0	0.0	0.0	0.062500	0.0	0.020833	0.0	...	0.000000	0.0	0.0
1	94024	0.030303	0.0	0.0	0.0	0.0	0.030303	0.0	0.000000	0.0	...	0.030303	0.0	0.0
2	94035	0.010000	0.0	0.0	0.0	0.0	0.010000	0.0	0.000000	0.0	...	0.000000	0.0	0.0
3	94040	0.000000	0.0	0.0	0.0	0.0	0.010000	0.0	0.000000	0.0	...	0.000000	0.0	0.0
4	94041	0.010000	0.0	0.0	0.0	0.0	0.010000	0.0	0.000000	0.0	...	0.010000	0.0	0.0

K-MEANS CLUSTERING

- Machine learning algorithm groups ZIP Codes with similar proportions of venue categories together into clusters.
- Number of clusters (k) determined by experiment. Find point where adding more clusters results in diminishing gains in similarity within clusters.
- Optimal: k = 8 clusters

CLUSTERS OF ZIP CODES WITH SIMILAR VENUES



Clusters with more than one ZIP Code:

- Downtown San Jose (yellow)
- Inner suburbs (red)
- Western suburbs (orange)
- Eastern suburbs (green)
- Northern suburbs (pink)

Oddball ZIP Codes in their own clusters:

- Koreatown (light blue)
- Eastern ZIP Codes with small venue sample sizes (dark blue and lime green)

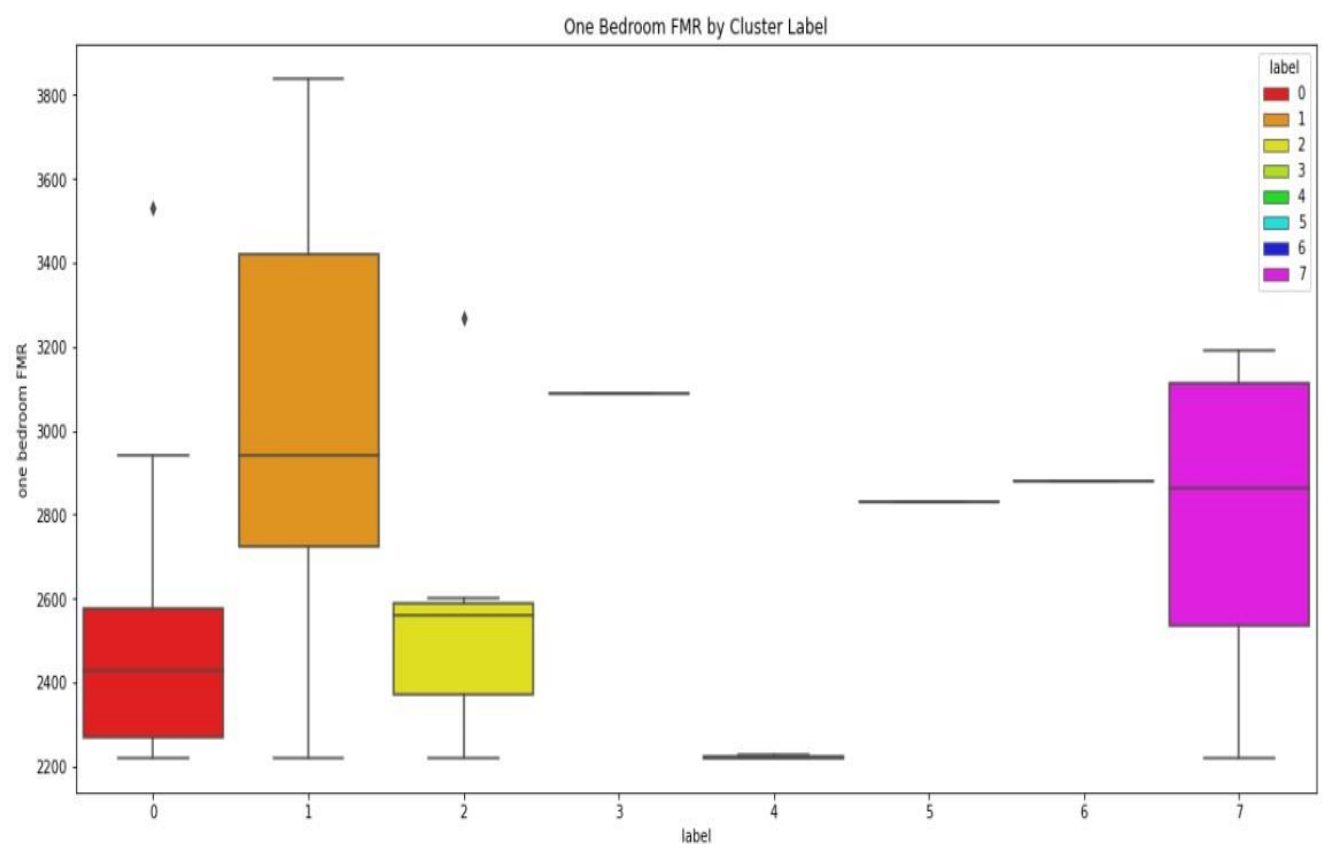
TOP VENUE CATEGORIES BY CLUSTER

	label	color	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	0	red	Coffee Shop	Sandwich Place	Vietnamese Restaurant	Pizza Place	Grocery Store	Fast Food Restaurant	Mexican Restaurant	Convenience Store	Bubble Tea Shop	Chinese Restaurant
1	1	orange	Coffee Shop	Park	Pizza Place	Mexican Restaurant	Grocery Store	Sandwich Place	Chinese Restaurant	Fast Food Restaurant	Gym / Fitness Center	Italian Restaurant
2	2	yellow	Mexican Restaurant	Coffee Shop	Bar	Cocktail Bar	Sandwich Place	Sushi Restaurant	Pub	Pizza Place	Ice Cream Shop	Grocery Store
3	3	lime	Yoga Studio	Hotel	Playground	Seafood Restaurant	Music Venue	Sushi Restaurant	Motel	Gym / Fitness Center	Mexican Restaurant	Baseball Field
4	4	green	Mexican Restaurant	Fast Food Restaurant	Convenience Store	Intersection	Pizza Place	Sandwich Place	Vietnamese Restaurant	Bakery	Video Store	Chinese Restaurant
5	5	light blue	Korean Restaurant	Indian Restaurant	Chinese Restaurant	Coffee Shop	Sushi Restaurant	BBQ Joint	Sandwich Place	Park	Spa	Rental Car Location
6	6	blue	Park	Light Rail Station	Tennis Court	Fried Chicken Joint	Playground	Food & Drink Shop	Bubble Tea Shop	Volleyball Court	Burrito Place	Donut Shop
7	7	pink	Park	Hotel	Food Truck	Mexican Restaurant	Convenience Store	Gym	Indian Restaurant	Vietnamese Restaurant	Video Store	Gym / Fitness Center

PATTERNS IN VENUE CATEGORIES

- Downtown: more bars, cocktail bars, and pubs
- Inner suburbs: not as many parks as outer suburbs
- Western suburbs: similar to inner suburbs, but more parks, gyms, and Italian restaurants
- Eastern suburbs: compared to western suburbs, more convenience stores, video stores, and bakeries but fewer coffee shops and parks

HOW DO CLUSTERS BASED ON VENUE CATEGORIES RELATE TO RENT?

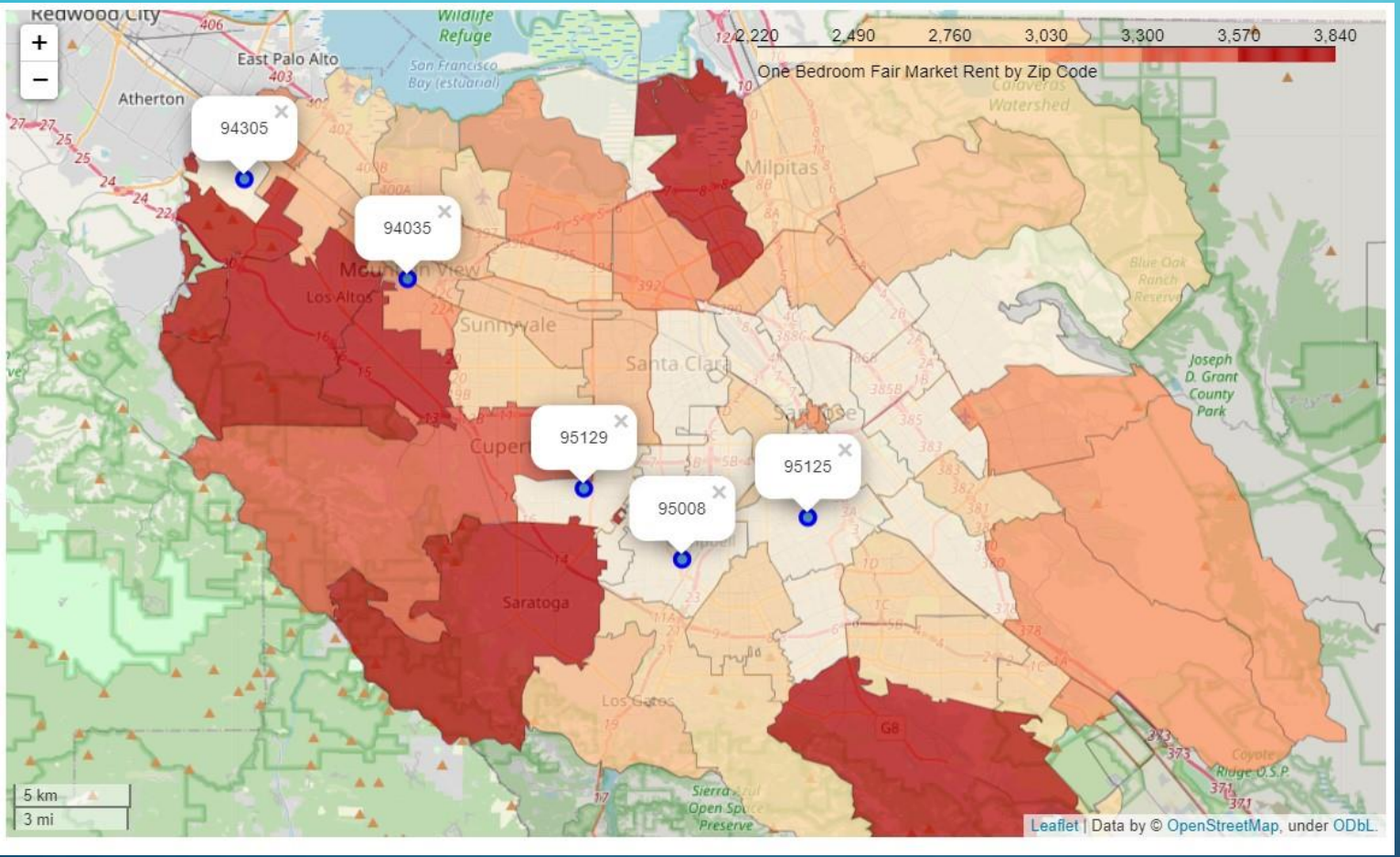


- Western suburb cluster (orange) has higher mean Fair Market Rent (FMR) than other clusters
- Kruskal-Wallis H test for the two clusters with the most ZIP Codes, the western suburbs and the adjacent inner suburbs (red) shows difference in mean FMR is statistically significant with a p value of 0.00027.

POTENTIALLY UNDERVALUED ZIP CODES

- Western cluster, with its parks, gyms, and Italian restaurants = overall higher rents
- ZIP Codes in this cluster with lower rents may be areas on the rise
  - Have already attracted the venue mix of expensive ZIP Codes
  - Rents potentially positioned to go up
- ZIP Codes in this cluster with the lowest FMRs:
  - 94305, 95125, 95129, 95008, and 94035

A CLOSER LOOK…



- 94305 is Stanford University. Can't invest here.
- 95125 is Willow Glen. Cute shopping and dining area. Worth a look.
- 95129 is West San Jose. Also worth a look.
- 95008 is Campbell. Revitalized core shopping and dining area. Again worth a look.
- 94035 is Moffett Field, a joint civil-military airport. Can't invest here.

CONCLUSION

- Three ZIP Codes with potential:
  - 95125 (Willow Glen)
  - 95129 (West San Jose)
  - 95008 (Campbell)
- Venue mix of expensive areas with currently lower rents
- Factors not considered: school ratings, crime rates, traffic noise, etc.
- Short candidate list decreases cost of further research

DATA SOURCES

- GeoJSON (ZIP Code boundaries) from the City of San Jose https://data.sanjoseca.gov/dataset/zip-code-boundary1/resource/cde0aa77-a995-4457-b6d1-6449c0274bfa
- ZIP Code Tabulation Zone center coordinates from GeoNames https://download.geonames.org/export/zip
- US Dept. of Housing and Urban Development small area Fair Market Rents https://www.huduser.gov/portal/datasets/fmr/smallarea/index.html
- Foursquare API endpoints https://developer.foursquare.com/docs/places-api/endpoints/