Relationship between Zillow Rent Index of CA Cities and Nearby Venues

1. Introduction/Business Problem

Tenants are looking for an affordable city or area to live and work.The nearby venues are an important factor to consider.

Investors search a good place to buy houses and use them to collect rentals. The nearby venues can predict the capital return and potential rentals.

Therefore, this project will explore the relationship between Zillow Rent Index of CA cities and nearby venues from Foursquare.

This project helps a group of stakeholders (tenants and landlord) solve a problem.

2. Data Source

2.1. Zillow Rent Index data from Zillow

Zillow Rent Index (ZRI): A smoothed measure of the typical estimated market rate rent across a given region and housing type. ZRI, which is a dollar-denominated alternative to repeat-rent indices, is the mean of rent estimates that fall into the 40th to 60th percentile range for all homes and apartments in a given region, including those not currently listed for rent.

Multifamily, SFR, Condo, Co-op (rental house type)

City level (California)

Columns: City Area Size (SizeRank), Rentals (Zri), Each City Rental Sample Size (ZriRecordCnt)

	City	SizeRank	Zri	ZriRecordCnt
0	Los Angeles	1	2840	824116
1	San Diego	8	2610	406428
2	San Jose	11	3130	257852
3	San Francisco	14	4220	182102
4	Sacramento	29	1700	185705

2.2 Latitude and Longitude of California Cities

Maps of World (California Latitude and Longitude Map, city level)

	City	Latitude	Longitude
0	Acalanes Ridge	37.9	-122.08
1	Acampo	38.17	-121.28
2	Acton	34.5	-118.19
3	Adelanto city	34.59	-117.44
4	Adin	41.2	-120.95

2.3 Foursquare Location Data to Get Nearby Venues

Merge above tables to get nearby venues with city name, latitude and longitude

	City	SizeRank	Zri	ZriRecordCnt	Latitude	Longitude
0	Los Angeles	1	2840	824116	34.11	-118.41
1	San Diego	8	2610	406428	32.82	-117.13
2	San Jose	11	3130	257852	37.3	-121.85
3	San Francisco	14	4220	182102	37.76	-122.69
4	Sacramento	29	1700	185705	38.57	-121.47
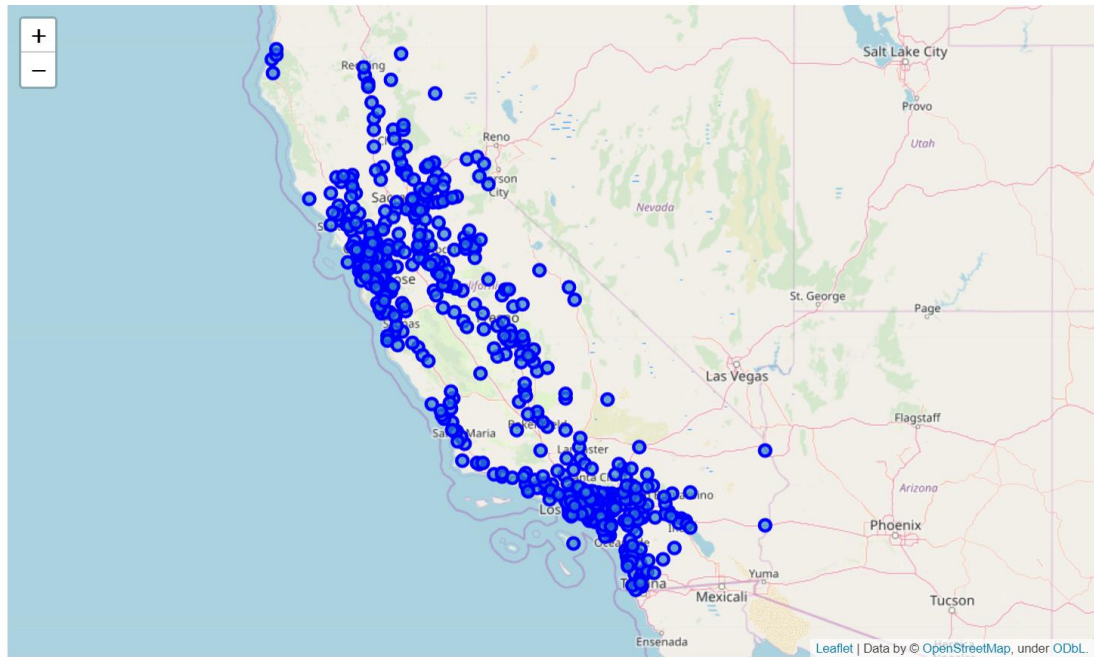
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Los Angeles	34.11	-118.41	Planet Smoothie	34.106260	-118.418041	Juice Bar
1	Los Angeles	34.11	-118.41	Hastain Trail	34.106200	-118.411327	Trail
2	Los Angeles	34.11	-118.41	Franklin Canyon Ranch	34.106318	-118.413352	Park
3	Los Angeles	34.11	-118.41	A Cleaner Pool	34.111485	-118.404465	Home Service
4	Los Angeles	34.11	-118.41	Sai Num Song Fung Khong	34.103432	-118.410871	Vietnamese Restaurant

3. Methodology

Use the nearby venues table to analyze each city

	Neighborhood	ATM	Accessories Store	Adult Boutique	Airport	Airport Food Court	Airport Gate	Airport Service	Airport Terminal	Alternative Healer	...	Warehouse Store	Water Park	Waterfall	Weight Loss Center
0	Acampo	0.000000	0.00	0.0	0.000	0.0	0.0	0.0	0.0	0.0	...	0.00	0.0	0.0	0.000000
1	Adelanto	0.000000	0.00	0.0	0.000	0.0	0.0	0.0	0.0	0.0	...	0.00	0.0	0.0	0.000000
2	Agoura Hills	0.014286	0.00	0.0	0.000	0.0	0.0	0.0	0.0	0.0	...	0.00	0.0	0.0	0.000000
3	Agua Dulce	0.000000	0.00	0.0	0.125	0.0	0.0	0.0	0.0	0.0	...	0.00	0.0	0.0	0.000000
4	Alameda	0.000000	0.00	0.0	0.000	0.0	0.0	0.0	0.0	0.0	...	0.00	0.0	0.0	0.000000
5	Alamo	0.000000	0.00	0.0	0.000	0.0	0.0	0.0	0.0	0.0	...	0.00	0.0	0.0	0.000000
6	Albany	0.000000	0.00	0.0	0.000	0.0	0.0	0.0	0.0	0.0	...	0.00	0.0	0.0	0.000000
7	Alhambra	0.000000	0.00	0.0	0.000	0.0	0.0	0.0	0.0	0.0	...	0.00	0.0	0.0	0.000000
8	Aliso Viejo	0.000000	0.00	0.0	0.000	0.0	0.0	0.0	0.0	0.0	...	0.00	0.0	0.0	0.000000
9	Alpine	0.000000	0.00	0.0	0.000	0.0	0.0	0.0	0.0	0.0	...	0.00	0.0	0.0	0.000000
10	Altadena	0.000000	0.00	0.0	0.000	0.0	0.0	0.0	0.0	0.0	...	0.00	0.0	0.0	0.000000

Here, we can see each city with different nearby venues. The total venues is 455.

Create the new table and display the top 10 venues for each city

|	Neighborhood	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	Acampo	Winery	Business Service	Yoga Studio	Field	Ethiopian Restaurant	Event Service	Event Space	Eye Doctor	Fabric Shop	Factory
1	Adelanto	Cosmetics Shop	Department Store	Yoga Studio	Filipino Restaurant	Event Service	Event Space	Eye Doctor	Fabric Shop	Factory	Fair
2	Agoura Hills	Fast Food Restaurant	Deli / Bodega	Pharmacy	Mexican Restaurant	Bakery	Gym / Fitness Center	Gym	Chinese Restaurant	Shipping Store	Athletics & Sports
3	Agua Dulce	Bakery	Convenience Store	Café	Mexican Restaurant	Pizza Place	Gift Shop	Grocery Store	Airport	Eye Doctor	Factory
4	Alameda	Beach	Bus Station	Liquor Store	Home Service	Trail	Tennis Court	Dog Run	Other Great Outdoors	Gym	Park

For instance, Alameda is a city in Alameda County, California, United States. It is located on Alameda Island and Bay Farm Island. It is adjacent to and south of Oakland and east of San Francisco across the San Francisco Bay. Bay Farm Island, a portion of which is also known as "Harbor Bay Isle", is part of the mainland adjacent to the Oakland International Airport. Therefore, the 1st most common venue is beach.

The k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. k-Means minimizes intragroup distance, and maximizes intergroup distance.

This method can group different cities according to different nearby venues. After that, we could compare cluster labels with SizeRank, Zri, and ZriRecordCnt to find the relationship between Zillow Rent Index of CA cities and nearby venues.

4. Results

	City	SizeRank	Zri	ZriRecordCnt	Latitude	Longitude	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Los Angeles	1	2840	824116	34.11	-118.41	1	Los Angeles	Home Service	Park	Trail	Juice Bar	Vietnamese Restaurant
1	San Diego	8	2610	406428	32.82	-117.13	0	San Diego	Café	Sandwich Place	Real Estate Office	Hotel	Diner
2	San Jose	11	3130	257852	37.3	-121.85	0	San Jose	Vietnamese Restaurant	Fast Food Restaurant	Mexican Restaurant	Marijuana Dispensary	Asian Restaurant
3	Sacramento	29	1700	185705	38.57	-121.47	0	Sacramento	Coffee Shop	Mexican Restaurant	Bar	Pizza Place	Dive Bar
4	Fresno	32	1330	152740	36.78	-119.79	0	Fresno	Mexican Restaurant	Hotel	Chinese Restaurant	Video Store	Liquor Store
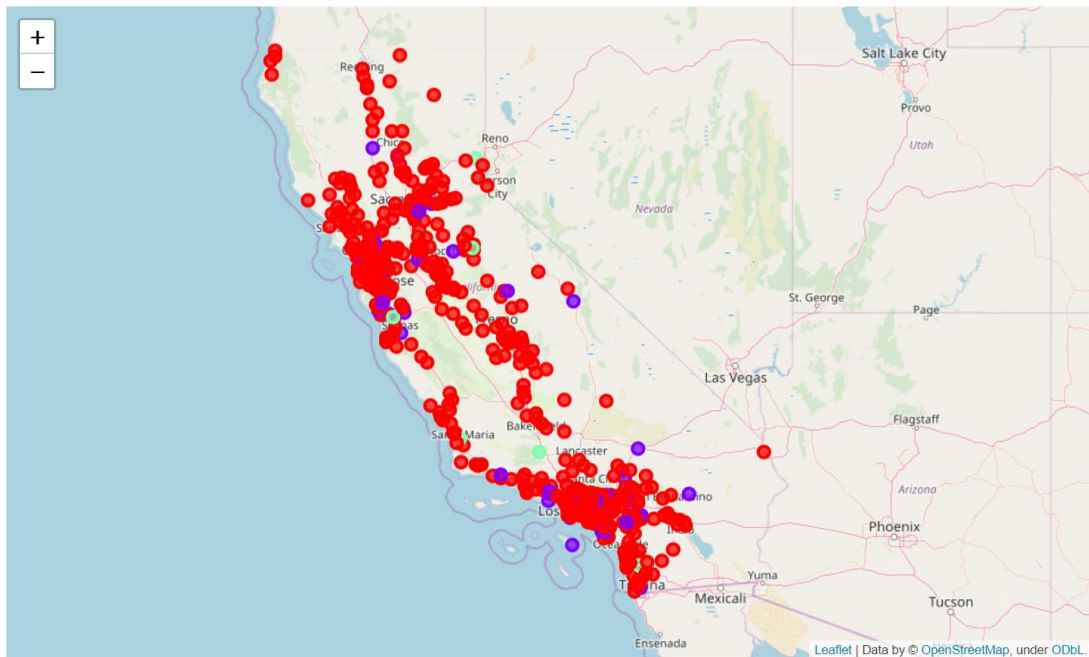
From map, we can see there are 3 groups. Now we combine the cluster labels, SizeRank, Zri, and ZriRecordCnt to analyze the question.

5. Discussion

Cluster Labels	SizeRank	Zri	ZriRecordCnt
0	2536.021484	2329.003906	16427.994141
1	2643.822222	2583.777778	33324.555556
2	4334.285714	2094.285714	8312.000000

SizeRank measures the region area (lower value means a big size city group). ZriRecordCnt is the sample size (number of rental houses) for each city. Therefore, Label 1 city group has high rental fees with large demand (large city). Labels 2 city group has low rental fees with low demand (small city). Label 0 city group is between Label 1 and Label 2 (rental price and demand), but the city has a big size.

Now, let's explore more details from nearby venues.
Top 3 venues of 1st Most Common Venue for each city group

```
Park            27
Trail            4
Home Service     2
Name: 1st Most Common Venue, dtype: int64
```

The metropolitan has the relaxed common venue.

```
Mexican Restaurant       74
Fast Food Restaurant     55
Pizza Place              30
Name: 1st Most Common Venue, dtype: int64
```

The common living area provide restaurant.

```
Business Service    6
Winery              1
Name: 1st Most Common Venue, dtype: int64
```

The suburb and countryside are with business service.

6. Conclusion

Tenants are looking for an affordable city or area to live and work.The nearby venues are an important factor to consider. Investors search a good place to buy houses and use them to collect rentals. The nearby venues can predict the capital return and potential rentals. According to results, the metropolitan, such as downtown Los Angeles, is an expensive rental area. The tenants prefer some relaxed venues near rental houses, such as parks. For most tenants, they prefer live in the convenient place with restaurants, supermarkets, and stores. Investors should invest money in large and medium cities. The good environment attracts the high-income tenants. However, convenient living service provides more opportunities to tenants and landlords.