



# Capstone Project

## The Battle of Neighborhoods

Applied Data Science Capstone by IBM



Christie Lay





Introduction



Data



Methodology



Results



Discussion



Conclusion



1.

# Introduction

Business Problem



## Business Problem



This project aims to help real estate investors investing in the right location. Los Angeles is a city formed by people with various ethnicity and races. The neighborhood in the county varies from one another. By analyzing different factors, we are able to come up with the location where can give the investors bigger return.





## 2. Data

Data that used to solve the problem  
and the source of the data



# USCities

City name, County, State, Latitude, Longitude. This dataset is used to generate map.



	city	city_ascii	state_id	state_name	county_fips	county_name	lat	lng	population	density	source	military	incorporated	time
0	New York	New York	NY	New York	36061	New York	40.6943	-73.9249	18713220	10715	polygon	False	True	America/New_York
1	Los Angeles	Los Angeles	CA	California	6037	Los Angeles	34.1139	-118.4068	12750807	3276	polygon	False	True	America/Los_Angeles
2	Chicago	Chicago	IL	Illinois	17031	Cook	41.8373	-87.6862	8604203	4574	polygon	False	True	America/Chicago

(Source: <https://simplemaps.com/data/us-cities>)



# Foursquare API

Explore neighborhood with its top venues



	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Los Angeles	34.1139	-118.4068	Poolside BH	34.113805	-118.403169	Pool
1	Long Beach	33.7981	-118.1675	Signal Hill Skyline	33.799343	-118.165590	Park
2	Long Beach	33.7981	-118.1675	Hilltop Park	33.799255	-118.165109	Park
3	Long Beach	33.7981	-118.1675	Mother's Market & Kitchen	33.802214	-118.168039	Organic Grocery
4	Long Beach	33.7981	-118.1675	The Home Depot	33.801868	-118.166359	Hardware Store
...	...	...	...	...	...	...	...
1077	Industry	34.0175	-117.9267	Red Restaurant @ Pacific Palms Hotel & Conferen...	34.020758	-117.926322	Restaurant
1078	Industry	34.0175	-117.9267	St. Andrew's Commons	34.020777	-117.927096	Golf Course
1079	Industry	34.0175	-117.9267	Panaderia La Mexicana	34.014103	-117.923180	Bakery
1080	Vernon	34.0019	-118.2106	Loveseat.com Online Furniture Auction Los Angeles	33.999964	-118.206722	Furniture / Home Store
1081	Vernon	34.0019	-118.2106	Express Mold	34.000675	-118.205650	Business Service

# LA crime data

This dataset contains crime rate for each city in LA county



	City/Agency	County	Population[5]	Populationdensity[5][3] [note 2]	Violent crimes[5]	Violent crime rateper 1,000 persons	Property crimes[5]	Property crime rateper 1,000 persons
0	Adelanto	San Bernardino	31213	557.3	189	6.06	790	25.31
1	Agoura Hills	Los Angeles	20767	2664.8	17	0.82	234	11.27
2	Alameda	Alameda	77048	7378.7	145	1.88	1723	22.36
3	Albany	Alameda	19350	10822.1	31	1.60	478	24.70
4	Alhambra	Los Angeles	84931	11129.7	168	1.98	1743	20.52
...	...	...	...	...	...	...	...	...
454	Yountville	Napa	2978	1945.1	7	2.35	40	13.43
455	Yreka	Siskiyou	7558	756.9	62	8.20	298	39.43
456	Yuba City	Sutter	65525	4473.0	223	3.40	1699	25.93
457	Yucaipa	San Bernardino	52794	1858.9	71	1.34	978	18.52
458	Yucca Valley	San Bernardino	21228	530.5	75	3.53	533	25.11

(Source: [https://en.wikipedia.org/wiki/California\\_locations\\_by\\_crime\\_rate](https://en.wikipedia.org/wiki/California_locations_by_crime_rate))



# Rent data

This dataset contains rent data in different area



ZIP Code	HUD Area Code	HUD Metro Fair Market Rent Area Name	SAFMR0BR	SAFMR0BR - 90% Payment Standard	SAFMR0BR - 110% Payment Standard	SAFMR 1BR	SAFMR 1BR - 90% Payment Standard	SAFMR 1BR - 110% PaymentStandard	SAFMR 2BR	SAFMR 2BR - 90% Payment Standard	SAFMR 2BR - 110% PaymentStandard	SAFMR 3BR
0 76437	METRO10180M10180	Abilene, TX MSA	\$570	\$513	\$627	\$590	\$531	\$649	\$780	\$702	\$858	\$1,034
1 76443	METRO10180M10180	Abilene, TX MSA	\$580	\$522	\$638	\$600	\$540	\$660	\$780	\$702	\$858	\$1,034
2 76464	METRO10180M10180	Abilene, TX MSA	\$570	\$513	\$627	\$590	\$531	\$649	\$780	\$702	\$858	\$1,034
3 76469	METRO10180M10180	Abilene, TX MSA	\$590	\$531	\$649	\$610	\$549	\$671	\$800	\$720	\$880	\$1,034
4 79501	METRO10180M10180	Abilene, TX MSA	\$700	\$630	\$770	\$720	\$648	\$792	\$940	\$846	\$1,034	\$1,210

(Source: [https://www.huduser.gov/portal/datasets/fmr.html#2021\\_data](https://www.huduser.gov/portal/datasets/fmr.html#2021_data))

## Zipcode data

This dataset contains zipcode, city name, Latitude and Longitude in CA



	Zip	City	State	Latitude	Longitude	Timezone	Daylight savings time flag	geopoint	Unnamed: 8
0	91131	Pasadena	CA	33.786594	-118.298662	-8	1	33.786594	-118.298662
1	92399	Yucaipa	CA	34.033564	-117.037760	-8	1	34.033564	-117.037760
2	95957	Meridian	CA	39.070323	-121.830650	-8	1	39.070323	-121.830650
3	93225	Frazier Park	CA	34.815266	-119.004040	-8	1	34.815266	-119.004040
4	94244	Sacramento	CA	38.377411	-121.444429	-8	1	38.377411	-121.444429

(Source: <https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/>)



The background is a detailed historical map of the North Atlantic and Europe. The left side shows North America, including Hudson Bay, James Bay, and the Great Lakes region, with cities like Montreal and New York. The right side shows Europe, including the British Isles, France, Germany, and the Mediterranean Sea. Red lines are drawn across the map, representing shipping routes or trade paths. A white rectangular box with a thin blue border is centered over the map, containing the title and a compass rose.

### 3. Methodology



# Methodology



## Why Use K-means Clustering?

“The K-means clustering algorithm is used to find groups which have not been explicitly labeled in the data. This can be used to confirm business assumptions about what types of groups exist or to identify unknown groups in complex data sets.”

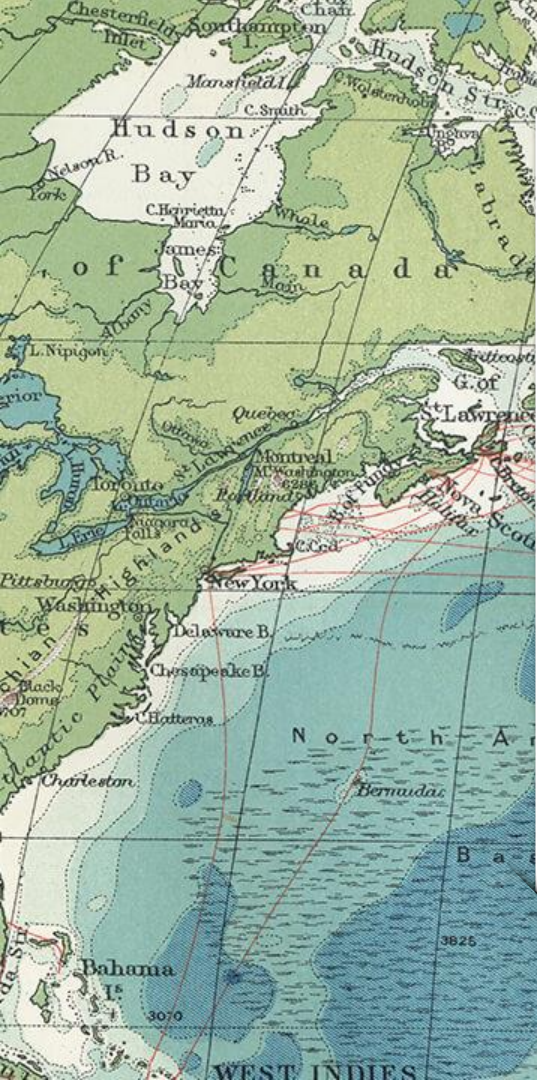
(Source: <https://blogs.oracle.com/ai-and-datascience/post/introduction-to-k-means-clustering>)

## Why Use Data Visualization?

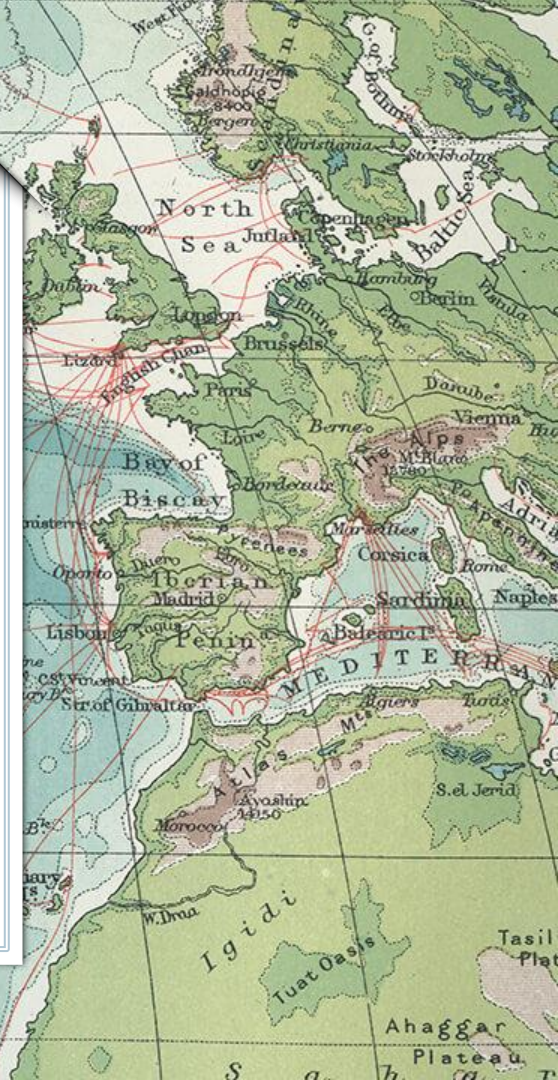
“Data visualization gives us a clear idea of what the information means by giving it visual context through maps or graphs. This makes the data more natural for the human mind to comprehend and therefore makes it easier to identify trends, patterns, and outliers within large data sets”

(Source: <https://analytiks.co/importance-of-data-visualization/>)





## 4. Results



# K-Means Clustering

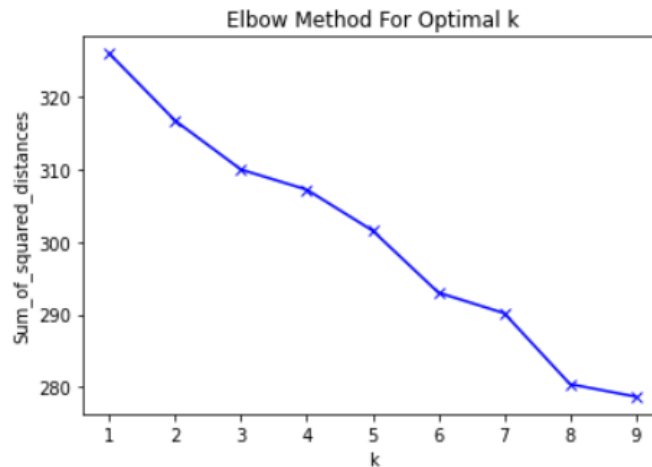
## Elbow Method For Optimal k



We implement elbow method to find the best k for clustering. According to the result, we set the k to 8

```
Sum_of_squared_distances = []
K = range(1,10)
for k in K:
    km = KMeans(n_clusters=k)
    km = km.fit(data_transformed)
    Sum_of_squared_distances.append(km.inertia_)
```

```
plt.plot(K, Sum_of_squared_distances, 'bx-')
plt.xlabel('k')
plt.ylabel('Sum_of_squared_distances')
plt.title('Elbow Method For Optimal k')
plt.show()
```





# K-Means Clustering

## Cluster 0 Result



	state_name	lng	Cluster Labels	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
93	California	-118.1675	0	Park	Hardware Store	Food Truck	Trail	Scenic Lookout	Organic Grocery	Event Service	Donburi Restaurant	Donut Shop	Electronics Store
694	California	-118.2017	0	Pharmacy	Gym	Park	Yoga Studio	Donburi Restaurant	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant
998	California	-118.0686	0	Food Truck	Park	Yoga Studio	Donburi Restaurant	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market
1352	California	-117.7153	0	Home Service	Park	Yoga Studio	Donburi Restaurant	Food	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant
1520	California	-117.7702	0	Park	Yoga Studio	Dog Run	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Farm
2044	California	-117.9300	0	Food Court	Park	Dog Run	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Farm
2111	California	-118.2362	0	Convenience Store	Park	Dog Run	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Farm
2399	California	-117.9041	0	Food	Park	Dog Run	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Farm
2437	California	-118.2353	0	Food	Park	Dog Run	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Farm

# K-Means Clustering

## Cluster 1 Result



	state_name	lng	Cluster Labels	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
125	California	-118.1753	1	Discount Store	Department Store	Fast Food Restaurant	Sandwich Place	Chinese Restaurant	Pet Store	Cosmetics Shop	Paper / Office Supplies Store	Optical Shop	ATM
143	California	-118.4964	1	Gym	Mexican Restaurant	Chinese Restaurant	Burger Joint	Donut Shop	Liquor Store	Tattoo Parlor	Gastropub	Greek Restaurant	Hardware Store
309	California	-118.3417	1	Cosmetics Shop	Pool	Steakhouse	Recreation Center	Garden	Yoga Studio	Dog Run	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant
365	California	-118.1691	1	Food	Chinese Restaurant	Intersection	Taco Place	Liquor Store	Bus Station	Donut Shop	Flower Shop	Flea Market	Fish & Chips Shop
389	California	-118.0291	1	Pharmacy	Hotel	Juice Bar	Bank	Sandwich Place	Discount Store	Burger Joint	Donut Shop	Veterans' Organization	Video Store
...	...	...	...	...	...	...	...	...	...	...	...	...	...
11518	California	-118.3783	1	Construction & Landscaping	Baseball Field	Vegetarian / Vegan Restaurant	Golf Course	Donburi Restaurant	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant
12096	California	-118.2990	1	Vineyard	Restaurant	Yoga Studio	Dog Run	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market
14264	California	-117.9684	1	Home Service	Donburi Restaurant	Food	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Farm
18633	California	-118.4588	1	Campground	Yoga Studio	Donburi Restaurant	Food	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market



# K-Means Clustering

## Cluster 2 Result



	state_name	lng	Cluster Labels	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
293	California	-117.7626	2	Seafood Restaurant	Mexican Restaurant	Fried Chicken Joint	Gas Station	Donut Shop	Yoga Studio	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant
318	California	-118.1390	2	Mexican Restaurant	Thai Restaurant	Currency Exchange	Grocery Store	Donut Shop	Food	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service
437	California	-117.9112	2	Mexican Restaurant	Juice Bar	Trail	Butcher	Sports Bar	Department Store	Electronics Store	Food & Drink Shop	Food	Flower Shop
872	California	-118.2167	2	Mexican Restaurant	Fried Chicken Joint	Fast Food Restaurant	Bakery	Café	Yoga Studio	Donut Shop	Flower Shop	Flea Market	Fish & Chips Shop
1170	California	-118.1550	2	Mexican Restaurant	Grocery Store	Donut Shop	Food	Fast Food Restaurant	Fried Chicken Joint	Liquor Store	Burger Joint	Thai Restaurant	Sporting Goods Shop
1370	California	-118.0579	2	Mexican Restaurant	Restaurant	Donburi Restaurant	Food	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market
1384	California	-118.1798	2	Mexican Restaurant	Market	Burger Joint	Yoga Studio	Electronics Store	Food	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service
1781	California	-118.1877	2	Mexican Restaurant	Pizza Place	Convenience Store	Taco Place	Grocery Store	Gas Station	Latin American Restaurant	Fast Food Restaurant	Donut Shop	Dive Bar

# K-Means Clustering

## Cluster 3 Result



	state_name	lng	Cluster Labels	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
226	California	-118.2468	3	Trail	Yoga Studio	Donburi Restaurant	Food	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market
1184	California	-118.3637	3	Trail	Nature Preserve	Yoga Studio	Donburi Restaurant	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market
1348	California	-117.9921	3	Trail	Yoga Studio	Donburi Restaurant	Food	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market



# K-Means Clustering

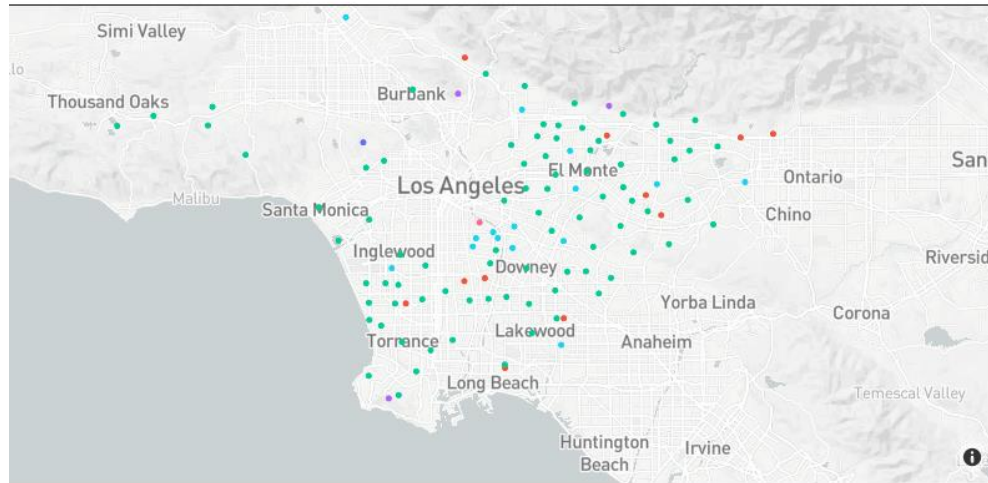
## Cluster 4 Result



	state_name	lng	Cluster Labels	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
2469	California	-118.6317	4	Furniture / Home Store	Yoga Studio	Food Court	Food	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market
25948	California	-118.2106	4	Furniture / Home Store	Business Service	Yoga Studio	Donut Shop	Food	Flower Shop	Flea Market	Fish & Chips Shop	Financial or Legal Service	Fast Food Restaurant

## Data Visualization - Plotly

### Scatterplot of clusters

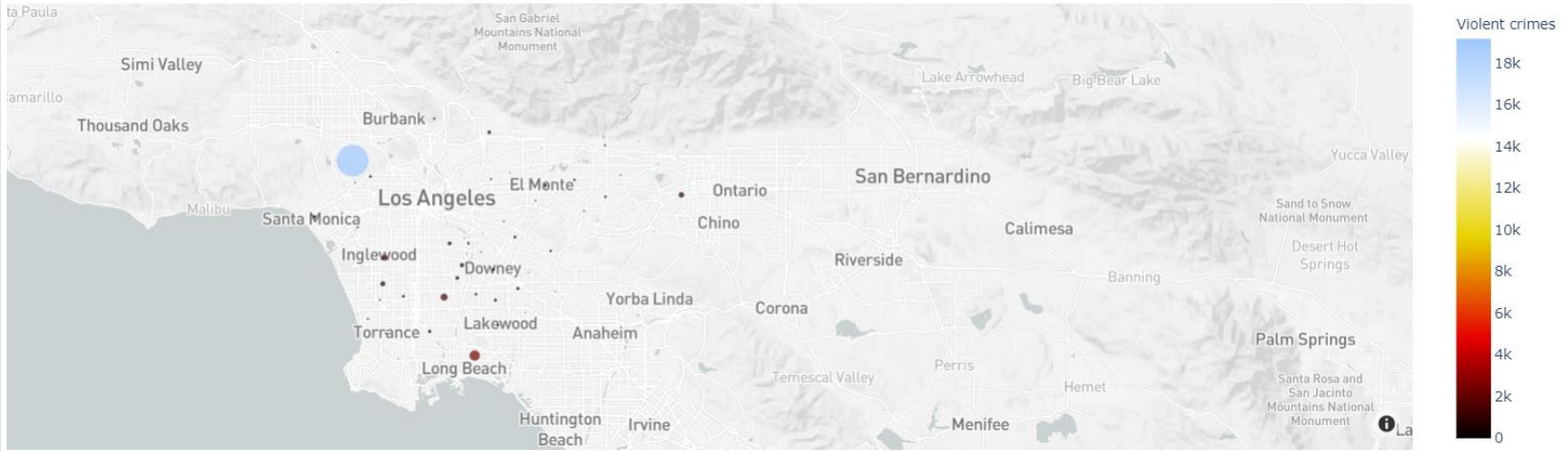


- Cluster Labels=5
- Cluster Labels=0
- Cluster Labels=1
- Cluster Labels=3
- Cluster Labels=6
- Cluster Labels=2
- Cluster Labels=4
- Cluster Labels=7



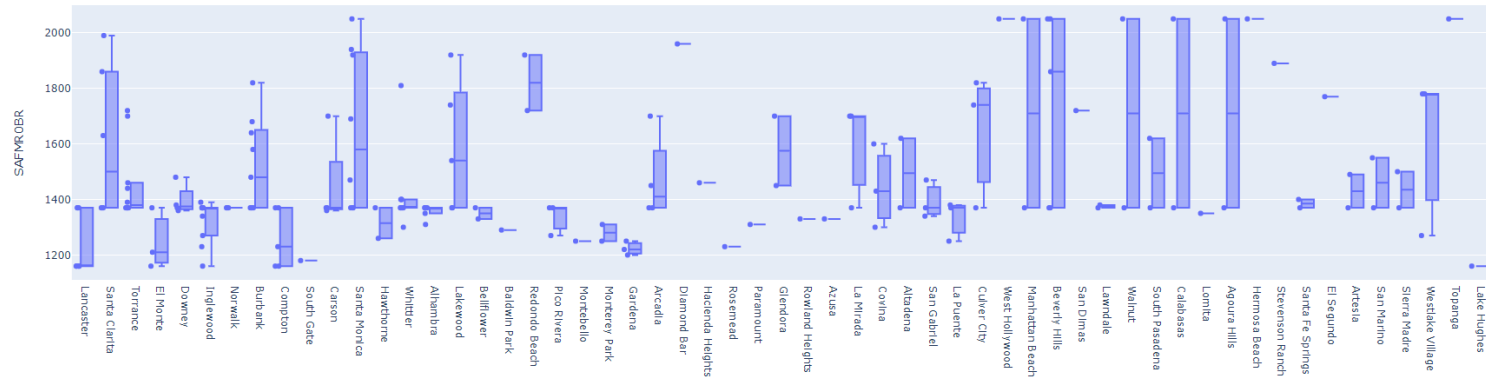
## Data Visualization - Plotly

### LA crime rate

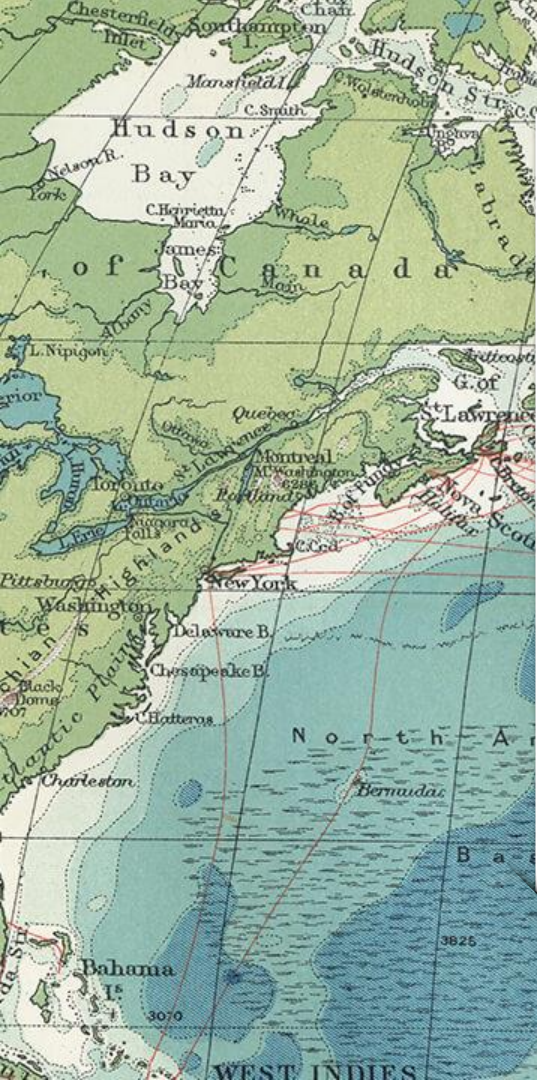


## Data Visualization - Plotly

### Cluster 1 rental price boxplot







5.

# Discussion



## Discussion



- Among all the clusters, the location in cluster 2 tend to have higher crime rate, so we are not taking cluster 2 into consideration.
- According to the result, the location in cluster 1 might be good choices. There's various kinds of venues in the neighborhood. However, There's a wide a range of price differences in the rental market. We can see that the rent on the west side of LA is quite pricey. Therefore, the investors should consider investing properties located on the east side of LA.



## 6. Conclusion

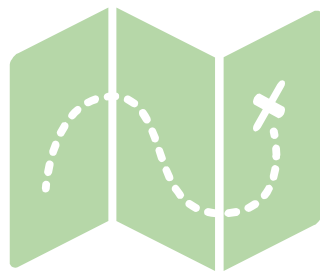


## Conclusion



- This project aims to help real estate investors to target the right property to invest.
- We've look at the crime rate and rental price in the area and came out with some idea of the desired investing properties.
- There are lots of different factors that could possibly effect the value of the property. For example, household income or school district can also be the reason why people choose to live in certain area. In future research, as more factors are adding in, the more accurate the result will be.





THANKS!