1. Intro

As the largest city in California, the City of Los Angeles encompasses a diverse economic and social communities. Millions of people lives in the city, and both domestic and global businesses set their bases here because of the internationalized nature of the city. It's well known that the housing price in LA has been staying at a relatively high level for decades. It has become a major concern, especially to those who are planning to move to LA from other cities, that how to find the ideal neighborhood. This project aims to help resolve this problem by visualizing the data of housing price in all the regions in LA, and also provides information regarding the characteristics of each region.

2. Data acquisition

Data used for this project were gathered from Los Angeles Times Mapping L.A. Boundaries API and Neighborhood Data for Social change, which is a project of USC Price Center for Social Innovation. In later stage, Foursquare API for developers was used to gather data within each area of interest. The first data source provided a geojson-format file, which was used to partition different regions in LA, so as to enable further manipulation of the data and characteristics regarding each region. The second source contains rent of each neighborhood, and for each neighborhood there exist multiple entries that provides abundant information to estimate the average price of each neighborhood.

3. Methodology

The data analysis procedure can be decided into four stages, data processing, preliminary data analysis, data visualization, and data clustering. methods used in this project include manipulation of data frame through Pandas library, data visualization achieved by matplotlib and folium library. Foursquare API for developers was then applied to gather further information of each region. Finally, K-means clustering algorithm was applied to categorize each region.

The first thing after data was imported was to filter out the irrelevant information and to deal with the missing values. Since the rent amounts within each region were not significantly different, and for some regions such as Angeles Crest, there were not sufficient entries to draw conclusion on how representative the data available are, all entries with no rent price were deleted from the data frame.

Considering the purpose of the project is to reflect the approximate rent price throughout LA, region instead of neighborhood was used as the criteria to dive the areas. A bar chart was used to present the rent price of different regions, and the regions were sorted in descending order.

The data visualization stage started with presenting the map of LA using the folium library. A choropleth map was created using the aforementioned geojson file and the rent price data of each region prepared in previous stage. Popup labels were added showing the region names and corresponding average rent price.

To further explore each region, Foursquare API was used to look for the most frequently appeared venues with in each area, so that viewer could have a better understanding the characteristics of each region. The K-means clustering algorithm was used to partition different regions into groups. The machine learning algorithm separates those regions by assigning each region to five clusters randomly, and further optimized based on the most frequently appeared venues within each region. Finally, in the choropleth map each region is labeled with name, average rent price, and cluster number (if applicable) assigned by the K-means algorithm.

4. Results and Discussion
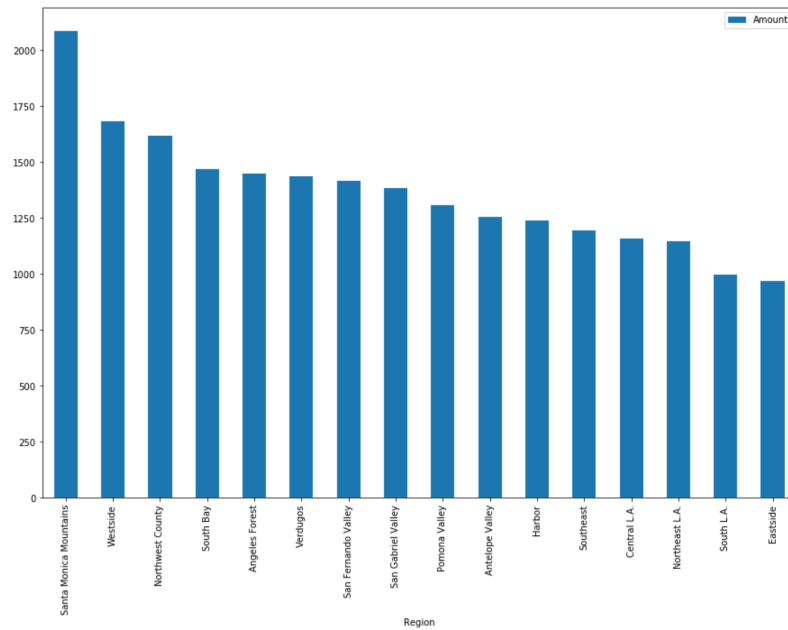
The bar chart above shows the average rent price in each region. The difference between each region is significant, and the average rent price of Santa Monica Mountains is nearly twice that of Eastside.
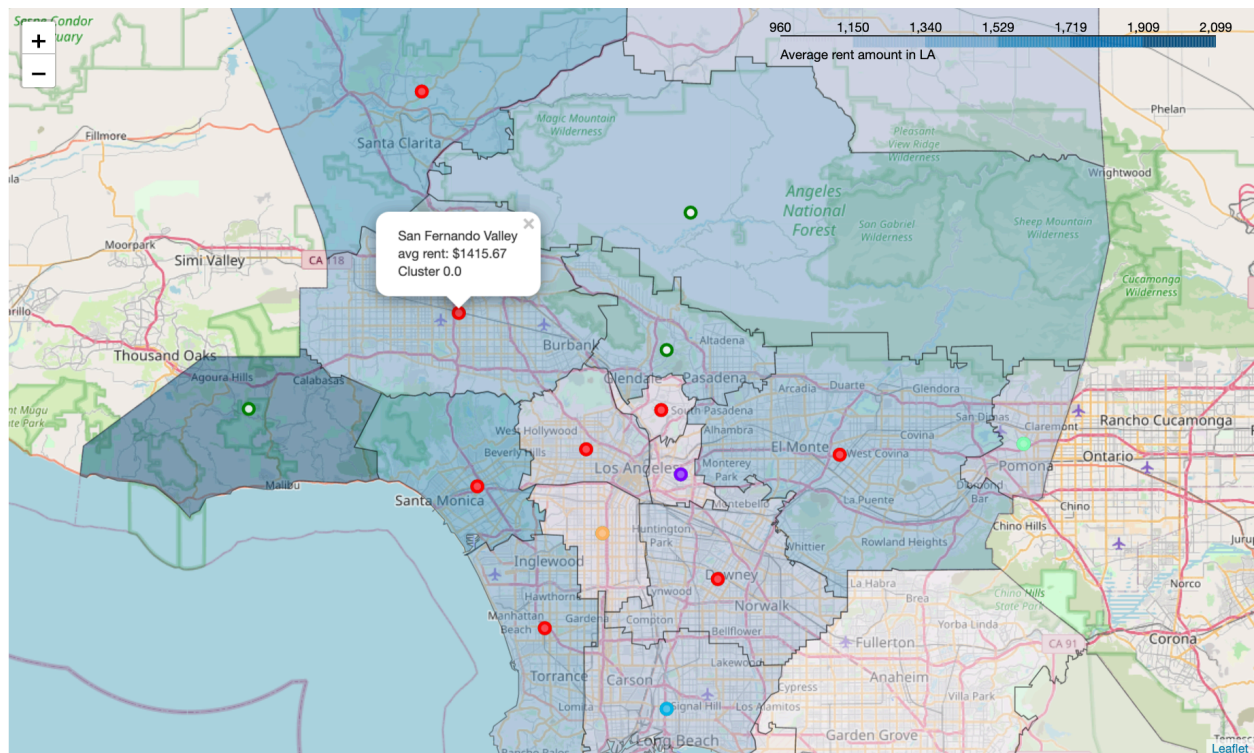
The choropleth map above visualizes the average rent price of each region, and also assigned each region into a cluster based on the similarity between their characteristics. As shown in the

graph, the central regions: Central L.A., South L.A., Northeast L.A., and Eastside has lower average rent price than the outer regions. Properties of each region are listed below:

Cluster 1

	Region	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
1	Westside	0.0	Mexican Restaurant	Japanese Restaurant	Yoga Studio	Pharmacy	Movie Theater	Liquor Store	Indian Restaurant	Health & Beauty Service	Gym	Government Building
2	Northwest County	0.0	Mexican Restaurant	Coffee Shop	Breakfast Spot	Middle Eastern Restaurant	Pharmacy	Massage Studio	Salon / Barbershop	Sandwich Place	Shopping Mall	Video Store
3	South Bay	0.0	Mexican Restaurant	Cuban Restaurant	Smoke Shop	Noodle House	Mediterranean Restaurant	Donut Shop	Martial Arts Dojo	Sandwich Place	Market	Liquor Store
6	San Fernando Valley	0.0	Peruvian Restaurant	Intersection	Bagel Shop	Cosmetics Shop	Marijuana Dispensary	Fast Food Restaurant	Doctor's Office	Dog Run	Donut Shop	Dumpling Restaurant
7	San Gabriel Valley	0.0	Paper / Office Supplies Store	Optical Shop	Convenience Store	Coffee Shop	Food	Yoga Studio	Electronics Store	Dog Run	Donut Shop	Dumpling Restaurant
11	Southeast	0.0	Convenience Store	Discount Store	Doctor's Office	Mexican Restaurant	Sandwich Place	Seafood Restaurant	Fast Food Restaurant	Burger Joint	Sushi Restaurant	Gym
12	Central L.A.	0.0	Korean Restaurant	Coffee Shop	Cocktail Bar	Chinese Restaurant	Seafood Restaurant	Mexican Restaurant	Dumpling Restaurant	Juice Bar	Pizza Place	Ramen Restaurant
13	Northeast L.A.	0.0	BBQ Joint	Dog Run	Convenience Store	Playground	Liquor Store	Mexican Restaurant	Scenic Lookout	Yoga Studio	Electronics Store	Donut Shop

Cluster 2

	Region	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
15	Eastside	1.0	Tree	Yoga Studio	Food	Dive Bar	Doctor's Office	Dog Run	Donut Shop	Dumpling Restaurant	Electronics Store	Fast Food Restaurant

Cluster 3

	Region	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
10	Harbor	2.0	Health & Beauty Service	Convenience Store	Chinese Restaurant	Yoga Studio	Food	Dog Run	Donut Shop	Dumpling Restaurant	Electronics Store	Fast Food Restaurant

Cluster 4

	Region	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
8	Pomona Valley	3.0	Pharmacy	Pizza Place	Grocery Store	Gas Station	Chinese Restaurant	Electronics Store	Dive Bar	Doctor's Office	Dog Run	Donut Shop

Cluster 5

	Region	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
14	South L.A.	4.0	American Restaurant	Check Cashing Service	Wine Bar	Health & Beauty Service	Gym / Fitness Center	Gym	Grocery Store	Government Building	Gas Station	Dive Bar

5. Conclusion

This project analyzed the average rent price in different areas in L.A. As shown in the result section, central area has lower average rent place than the regions on the surrounding regions. The outcome of the clustering algorithm is perhaps not so informative as the other data analysis approaches, and that is possibly due to the small number size of the regions and the shear amount of properties used as the criteria to categorize each region. The rest part, however, shed some light on the features of the city of Los Angeles. For those who are planning to move to L.A. and are looking for accommodation with lower price, the central regions seems to be ideal, since those regions also have comprehensive community infrastructures according to the cluster information. For business owners, the central regions also seems economic, and they may use the cluster information as reference before they make the decision of where to start their businesses.