Venues Effect on SF Housing Price

Shuhui Jia
Coursera IBM DS Capstone

Overview

- Housing price has increased tremendously in 2020 and 2021
- What is the most attractive thing for buyers?
  - Venues?
  - Location?
- Does buyer concern venues?

Data

Data Source

- SF zip code data
  - Government website
- Venues data
  - Foursquare API
- SF geolocation data
  - Federal website
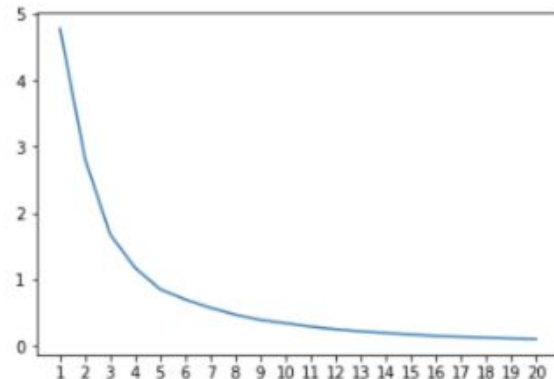- Housing price data
  - Redfin

Data Preprocess

- Data combination
  - 1st: SF zip with geolocation
  - 2nd: Venues with 1st data
  - Final: 2nd with price data
- Clustering data combination

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Cluster Labels
0	94102	37.779583	-122.41934	Louise M. Davies Symphony Hall	37.777976	-122.420157	Concert Hall	0
1	94102	37.779583	-122.41934	War Memorial Opera House	37.778601	-122.420816	Opera House	0
2	94102	37.779583	-122.41934	Herbst Theater	37.779548	-122.420953	Concert Hall	0
3	94102	37.779583	-122.41934	San Francisco Ballet	37.778580	-122.420798	Dance Studio	0
4	94102	37.779583	-122.41934	Asian Art Museum	37.780178	-122.416505	Art Museum	0

Kmeans algorithm

- Clustering lat and lon using kmeans clustering algorithm
- Best k selection
    - Elbow method
- K = 5 is the best one
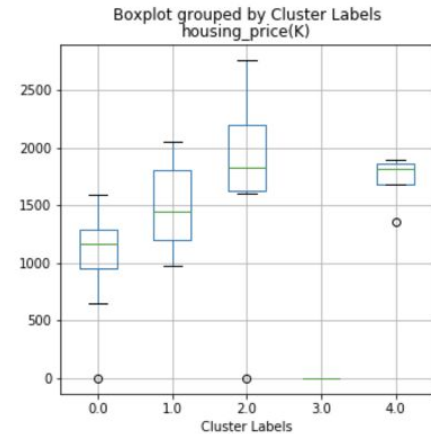- Select the top venue for each zip code

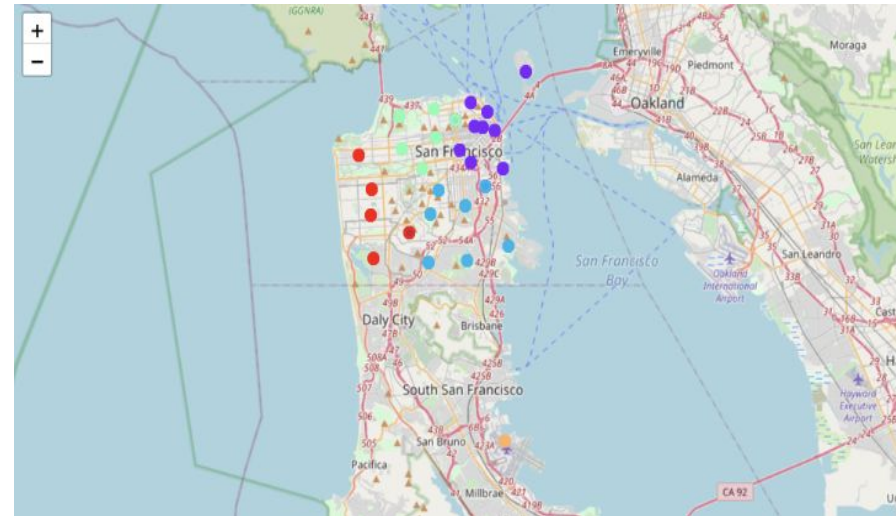| Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	housing_price(K)	Cluster Labels
94102	Beer Bar	Theater	Coffee Shop	Cocktail Bar	Vegetarian / Vegan Restaurant	Marijuana Dispensary	Wine Bar	Vietnamese Restaurant	Sushi Restaurant	Dance Studio	1285.0	0.0
94103	Coffee Shop	Nightclub	Beer Bar	Theater	Gay Bar	Cocktail Bar	Dance Studio	Marijuana Dispensary	Gym / Fitness Center	Gym	905.0	0.0
94104	Men's Store	Coffee Shop	Japanese Restaurant	Gym	Hotel	Salad Place	Mediterranean Restaurant	Sushi Restaurant	Food Truck	New American Restaurant	645.0	0.0
94105	Coffee Shop	Food Truck	Seafood Restaurant	Japanese Restaurant	Gym	Salad Place	Dessert Shop	Bookstore	Art Museum	Mediterranean Restaurant	1000.0	0.0
94107	Café	Coffee Shop	Mexican Restaurant	Sushi Restaurant	Park	Brewery	Sandwich Place	Bubble Tea Shop	Breakfast Spot	Art Gallery	1146.0	0.0

Analysis: housing price with cluster

- Cluster 3 is 0
  - Why? For further analysis
- Significant difference between all other 4 clusters
- Range is from less than 1M to greater than 2.5M

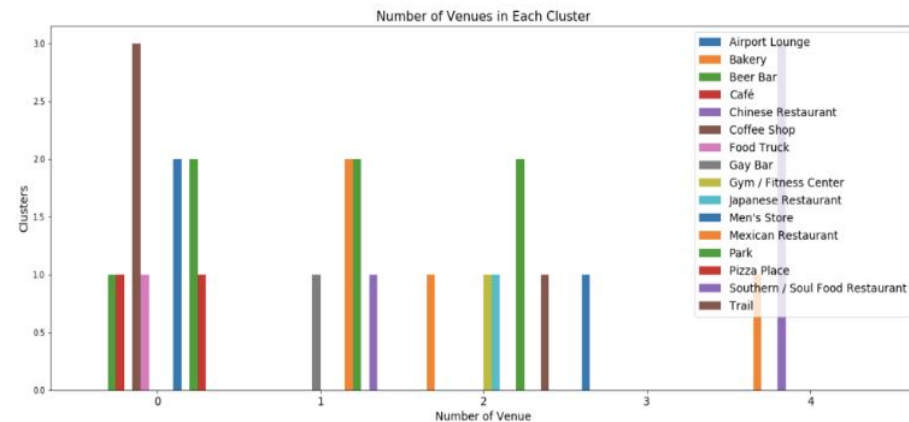Boxplot grouped by Cluster Labels
housing_price(K)

Visualization of clustering

- From cluster 0 to 4
  - purple, blue, green, orange, and red
- Cluster 3 is airport
  - No housing information
- Business center and bridge has lowest price
- Other places are higher than business center

Venues effect on housing price

- From cluster 0 to 4
    - coffee shop, bakery and park, park, nothing, and Chinese restaurant
- Considered with housing price
    - Park is more important for buyer
    - Business center and crowded places are less interests due to covid19

Conclusion

- Location is a very important effect on SF housing price
  - Business center is much cheaper than living places
- More people love to live the place with more park

Thank you.