LloydMartin Redux

Coursera Data Science Capstone Project

July 14th, 2020

Alejandro Erives

Coursera's IBM Data Science Capstone Project

Battle of the Neighborhoods: LloydMartin Redux – Illinois Edition

Introduction

The Problem & Background

LloydMartin was a restaurant in Seattle, Washington.

- It was "a small 34 seat restaurant whom offers a globally-inspired cuisine in which our menu changes daily based on seasonal ingredients. I reach out to the farms that produce the highest quality product based on seasonal availability. Some ingredients I love to serve are foie gras, Wagyu and Prime beef, ramps, truffles, risotto, housemade pastas, dayboat fish, and mushrooms!"

LloydMartin was situated in Seattle's Queen Anne Neighborhood.

- LloydMartin restaurant was located at 1525 Queen Anne Ave N, Seattle, WA 98109 USA. From Wikipedia: "Queen Anne Hill is an affluent neighborhood and geographic feature in Seattle, northwest of downtown."

Chef Sam moved back to Illinois.

- Chef Sam left Seattle and moved back to Illinois in the last year, naturally, his friends and family are enquiring about his next restaurant.

Can Chef Sam replicate the success of LloydMartin by finding a similar location as LloydMartin?

- This project will aim to seek a new restaurant location, based on available restaurant properties near Naperville, Illinois, USA. The goal will be to find which rental properties would be similar to the LloydMartin's Queen Anne location.
- Although the project is specifically undertaken to allow a selection of a similar venue location as Chef Sam's original LloydMartin restaurant, some of the learnings may be more broadly applied to restaurant location selection in general.

Data

Overview:

In this project, I will be comparing different restaurant property locations that are on the market in communities near/surrounding Naperville, Illinois (Chef Sam's new home). I will be comparing the surrounding neighborhoods of each of the available properties to see which may be similar to the LloydMartin location in Seattle, Washington.

Dataset 1: Restaurant Properties

Available properties for evaluation are loaded from a CSV file. The source of the listings was www.loopnet.com. The CSV file contains addresses, price of listing, and the property's square footage (or area).

Dataset 2: Neighborhood information (venues)

The surrounding neighborhoods will be characterized based on the nearby venues to each available property, and the venues surround LloydMartin in Seattle. The data will be sourced from Foursquare, specifically their Venue API.

Dataset – 'Other':

In addition to the venue data and the restaurant properties, I will also use Geocode to retrieve latitude and longitude information for each of the properties to be reviewed. I will use latitude and longitude in our queries to Foursquare API.

Another realization I made while retrieving the Foursquare data, was that the Venue Categories used by Foursquare Ire overly specific. To remedy this, I created an additional file which would be used as a dictionary to map the overly specific venue categories into broader categories (i.e. any type of restaurant would be represented by 'restaurant' instead of "Eastern European Restaurant" or "Diner" or "Steakhouse".

Lastly, I will also use the interactive maps in Folium to display some of our results.

Methodology

The methodology to be used in determining similar properties will be as follows:

1. Loading property addresses and listing information
2. Each property's address will be used to get Latitude and Longitude coordinates
3. Once I have coordinates, I will then use that data to query the Foursquare API to find nearby venues
    a. I restricted the search radius to 150 meters, since I wanted the query to represent the immediate local 'scene' that the restaurant would be located in, rather than the entire neighborhood.
4. Once I have retrieved the venue location set, I will reduce the number of unique venue categories by mapping the Foursquare Venue Categories (there Ire over 200 unique categories in my query).
    a. The resulting data set will only have 11 unique venue categories.
5. To prep the data for our machine learning algorithm, I will perform the following:
    a. Use dummy variables to create a 'one-hot-coding' dataframe for the nearby venues dataframe.
    b. Convert the one-hot codes to a mean-weighted average with a groupby method
6. Once I have the meant-weighted average for each venue category for each properties address, I can use this data to perform an unsupervised machine learning algorithm, namely k-Means clustering.
    a. Why k-Means clustering? The data I are using is unlabeled data. I would like to find properties similar to LloydMartin's location. k-Means clustering can work with multiple dimensions, which is the good since I have 11 venue categories in our dataset (each venue category will serve as a dimension)
7. To visualize our results I can perform the following:
    a. Folium Map: I mapped the restaurant property locations onto a Folium map, with each cluster represented in a different color.
    b. Tabular visualization: Display a dataframe listing each property, the cluster assigned, and the top 5 venue locations.
    c. Graphical visualization: Create scatter plots to visualize the k-Means clustering. I select specific dimensions for the scatter plots using the "top 5" venue locations identified in the Tabular visualization above.

i. The scatter plots really helped to understand some of the impacts of the various locations, such as which were significant and which were not.

Results

The first step in understanding the results was to create a folium map that identified the various clusters (5 clusters in total).

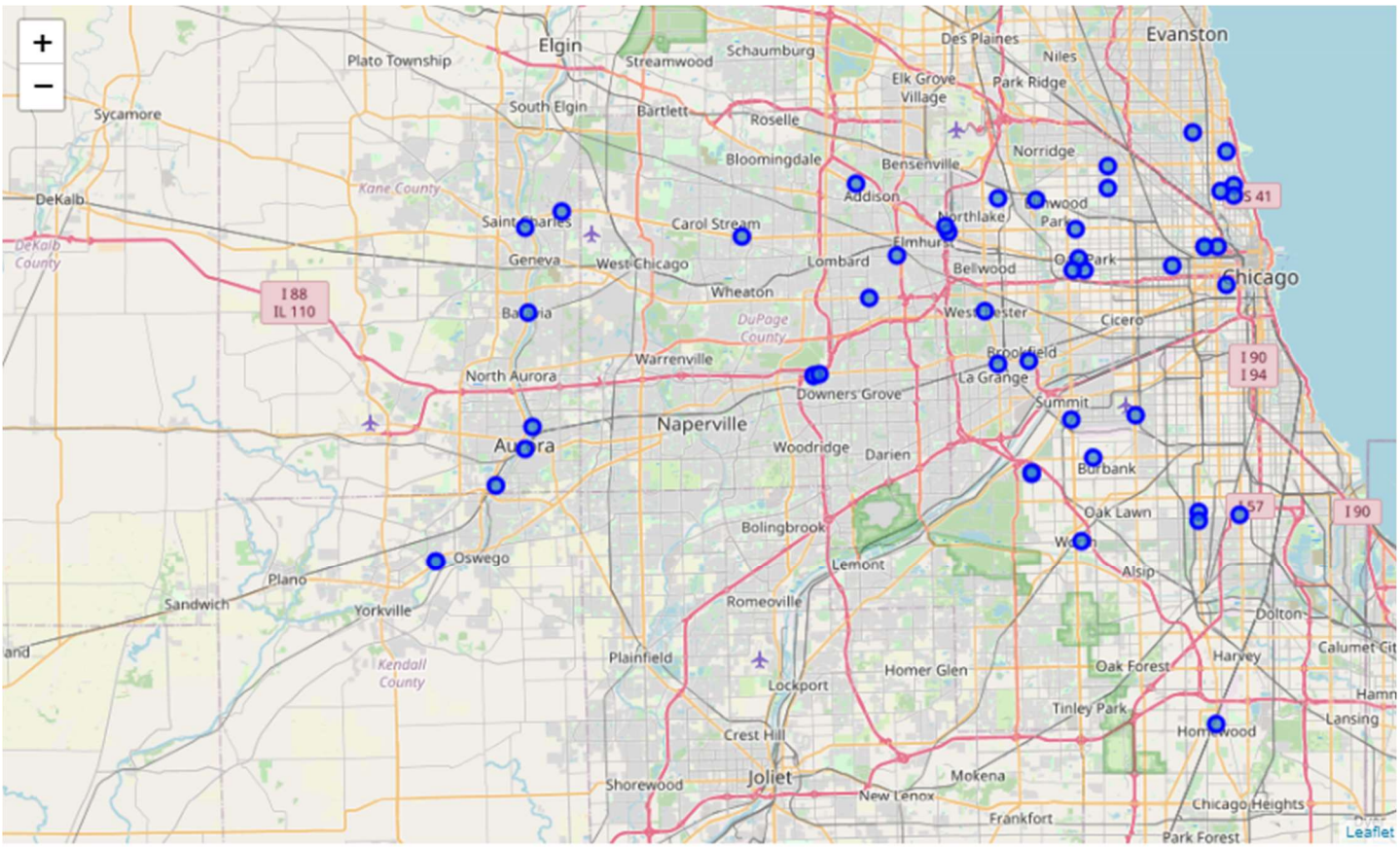


Figure 1. Plot of Restaurant Properties to be clustered. (Before Clustering)

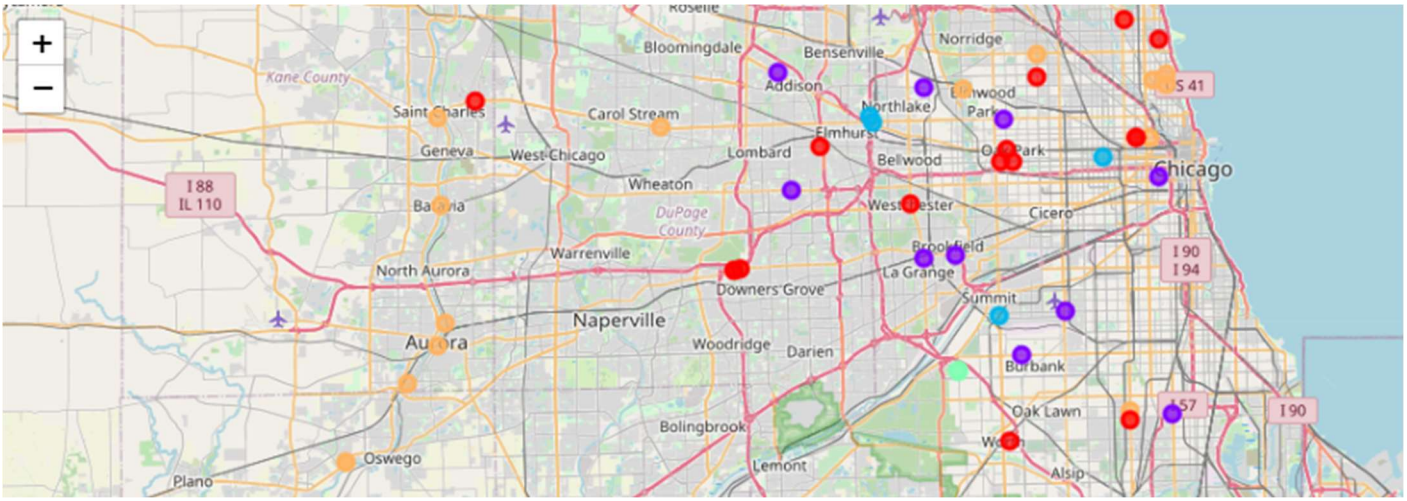


Figure 2. Plot of Restaurant Properties after clustering (Orange is the cluster with LloydMartin Restaurant)

The orange markers in Figure 2 correspond to Cluster 3, which is the cluster that includes our restaurant of interest (LloydMiller Restaurant). Right away, I found it interesting that the rental properties up and down the Fox River (on the left half of the plot) were in the LloydMiller cluster, as well as a few select properties in Chicago.

Visually, it seems k-Means clustering performed a good job of dividing the property list into smaller sets of data.

We will further explore & discuss the results in the Discussion section below:

Discussion

Cluster "0"

This cluster appears to have "Restaurant" as the most common venue, but interestingly, LloydMiller did not fall into this cluster.

	Address	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
3	4715 W 63rd St, Chicago, IL	41.778222	-87.740447	Restaurant	Shop	Entertainment	Transportation	Other
4	10200 S Halsted St, Chicago, IL	41.708391	-87.643175	Restaurant	Shop	Food Store	Transportation	Other
10	8819 Ogden Ave, Brookfield, IL	41.816379	-87.841179	Restaurant	Food Store	Entertainment	Transportation	Shop
18	30 S La Grange Rd, La Grange, IL	41.814616	-87.869891	Restaurant	Entertainment	Shop	Food Store	Gastropub
19	6 W Roosevelt Rd, Lombard, IL	41.860545	-87.990435	Restaurant	Shop	Transportation	Food Store	Other
22	6831 W North Ave, Oak Park, IL	41.908909	-87.797416	Restaurant	Shop	Entertainment	Food Store	Transportation
24	7141 79th Street, Burbank, IL	41.748429	-87.781433	Restaurant	Food Store	Entertainment	Transportation	Shop
26	1152 W Taylor St, Chicago, IL	41.869629	-87.656305	Restaurant	Shop	Entertainment	Food Store	Bank
41	9769 W Grand Ave, Franklin Park, IL	41.929738	-87.869557	Restaurant	Shop	Entertainment	Transportation	Food Store
47	672 N Mill Rd, Addison, IL	41.939613	-88.003237	Restaurant	Entertainment	Shop	Food Store	Bank

Cluster "1"

This cluster is notable for what types of venues are lacking... Notice that none of the top 5 most common venues for these properties include "Entertainment".

	Address	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	441 W Lake Street, Elmhurst, IL	41.905975	-87.917068	Shop	Restaurant	Food Store	Transportation	Other
7	6312 S Harlem Ave, Summit, IL	41.775356	-87.800936	Restaurant	Food Store	Transportation	Shop	Other
32	3206 W Washington Blvd, Chicago, IL	41.882873	-87.706439	Shop	Restaurant	Food Store	Transportation	Other
46	144 E North Ave, Northlake, IL	41.910254	-87.919077	Transportation	Other	Food Store	Shop	Restaurant

Cluster "2"

This small cluster also lacks any connection to "Entertainment" venues, and has "Restaurant", "Transportation", and "Shop"s as top venues.

	Address	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
17	8420 S 88th Ave, Justice, IL	41.737605	-87.838356	Restaurant	Transportation	Shop	Other	Office
33	8350 S 88th Ave, Justice, IL	41.738789	-87.838392	Restaurant	Transportation	Shop	Other	Office

Cluster "3"

A total of 16 available properties' neighborhoods were clustered with the LloydMiller restaurant's Queen Anne location. We will explore this cluster in more detail in the next section (after we look at Cluster 4 properties) to see what makes these properties 'special'.

	Address	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
8	1422-1424 W Chicago Ave, Chicago, IL	41.896376	-87.663448	Restaurant	Entertainment	Shop	Transportation	Other
9	600 E Veterans Parkway, Yorkville, IL	41.676452	-88.396870	Restaurant	Entertainment	Transportation	Shop	Other
11	155 E Veterans Parkway, Yorkville, IL	41.676452	-88.396870	Restaurant	Entertainment	Transportation	Shop	Other
15	8347 W Grand Ave, River Grove, IL	41.928426	-87.834947	Entertainment	Shop	Restaurant	Transportation	Other
16	3155 N Halsted Street, Chicago, IL	41.939462	-87.648950	Restaurant	Entertainment	Shop	Food Store	Transportation
20	2723 N Halsted St, Chicago, IL	41.932290	-87.648669	Restaurant	Entertainment	Shop	Food Store	Office
21	10036 S Western Ave, Chicago, IL	41.710766	-87.682041	Restaurant	Entertainment	Transportation	Shop	Food Store
23	460 S 1st St, Saint Charles, IL	41.909459	-88.313139	Entertainment	Restaurant	Food Store	Transportation	Shop
25	227 W Wilson St, Batavia, IL	41.850034	-88.310957	Restaurant	Entertainment	Shop	Food Store	Bank
27	211 N River St, Montgomery, IL	41.729291	-88.341051	Restaurant	Food Store	Entertainment	Shop	Transportation
28	720 E North Ave, Carol Stream, IL	41.903060	-88.109868	Entertainment	Office	Transportation	Shop	Restaurant
31	2913 N Lincoln Ave, Chicago, IL	41.934850	-87.661586	Entertainment	Restaurant	Shop	Transportation	Food Store
37	75-77 S Lasalle St, Aurora, IL	41.754934	-88.313622	Entertainment	Shop	Restaurant	Food Store	Transportation
38	3938 N Central Ave, Chicago, IL	41.952488	-87.767228	Entertainment	Restaurant	Shop	Transportation	Food Store
40	18027 Dixie Highway, Homewood, IL	41.561843	-87.665303	Restaurant	Entertainment	Shop	Transportation	Other
42	47 E Illinois Ave, Aurora, IL	41.770698	-88.307201	Restaurant	Entertainment	Transportation	Shop	Other
48	1525 Queen Anne Ave N, Seattle, WA	47.633140	-122.357260	Entertainment	Restaurant	Shop	Food Store	Bank

Cluster "4"

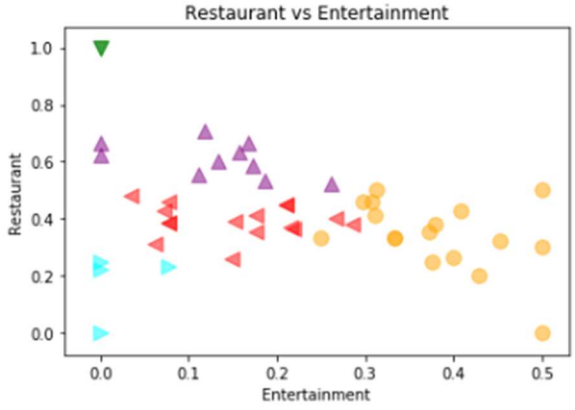
The 4th and largest cluster has "Restaurant" and "Shops" predominantly as the most common Venues.

	Address	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	2801 E. Main Street, St Charles, IL	41.921177	-88.278703	Restaurant	Shop	Transportation	Food Store	Entertainment
1	515 Madison Street, Oak Park, IL	41.879764	-87.789129	Restaurant	Transportation	Shop	Food Store	Entertainment
5	2410 Ogden Ave, Downers Grove, IL	41.805908	-88.043546	Restaurant	Shop	Transportation	Entertainment	Other
6	40W089 IL-64, Saint Charles, IL	41.805908	-88.043546	Restaurant	Shop	Transportation	Entertainment	Other
12	3108 N Central Ave, Chicago, IL	41.936984	-87.766869	Restaurant	Shop	Entertainment	Bank	Food Store
13	6900 W 111th Street, Worth, IL	41.690221	-87.791149	Restaurant	Shop	Food Store	Entertainment	Bank
14	1420 Ogden Ave, Downers Grove, IL	41.806905	-88.037795	Shop	Restaurant	Entertainment	Transportation	Food Store
29	1948 W Chicago Ave, Chicago, IL	41.896109	-87.676522	Restaurant	Shop	Entertainment	Food Store	Transportation
30	3011 Ogden Ave, Lisle, IL	41.896109	-87.676522	Restaurant	Shop	Entertainment	Food Store	Transportation
34	107 N Oak Park Ave, Oak Park, IL	41.887525	-87.794593	Restaurant	Shop	Entertainment	Food Store	Transportation
35	2306 W Foster Ave, Chicago, IL	41.976117	-87.687126	Restaurant	Shop	Entertainment	Transportation	Food Store
36	10350 S Western Ave, Chicago, IL	41.704946	-87.681962	Restaurant	Shop	Entertainment	Food Store	Transportation
39	4437 N Broadway, Chicago, IL	41.962816	-87.655454	Restaurant	Entertainment	Shop	Transportation	Food Store
43	1938 S Mannheim Rd, Westchester, IL	41.851464	-87.881940	Restaurant	Food Store	Bank	Transportation	Entertainment
44	1114 Madison St, Oak Park, IL	41.879695	-87.799518	Restaurant	Entertainment	Shop	Food Store	Bank
45	621 E Saint Charles Rd, Villa Park, IL	41.890118	-87.964809	Shop	Restaurant	Entertainment	Food Store	Bank

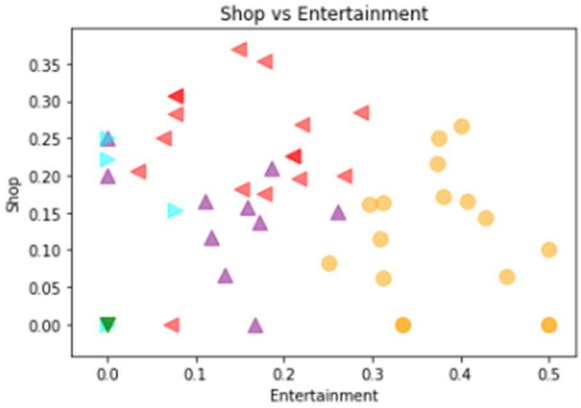
The tabular data above allowed us to understand the broad relevance that each of the venue categories had on determining which cluster a particular property was grouped. However, we also can further understand the impacts of each venue category by graphically plotting the mean-weighted data that we used in the k-Means machine learning algorithm.

The following plots show the results of plotting the mean-weighted variables: (Note that in each plot, Category 3 is shown in **ORANGE**.)

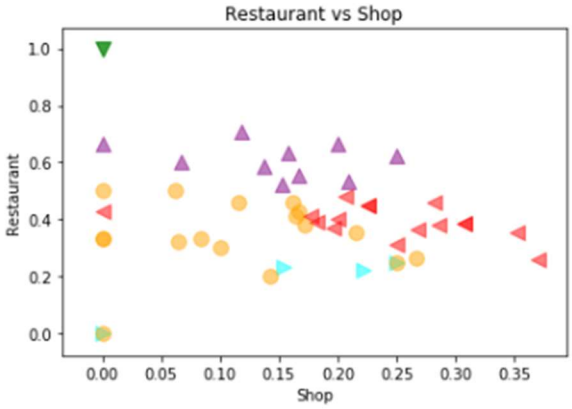
Restaurant vs Entertainment: From this scatterplot, we can see that Cluster 3 includes properties where the 'density' of Entertainment venues was above a threshold of ~0.25. LloydMiller restaurant was in a desirable location, with lots to do in the area. It is possible that having good entertainment options is a good thing to look for when selecting a new restaurant location.



Shop vs Entertainment: From the next scatterplot, we can see that Shop density does not really distinguish the cluster when compared to Entertainment.

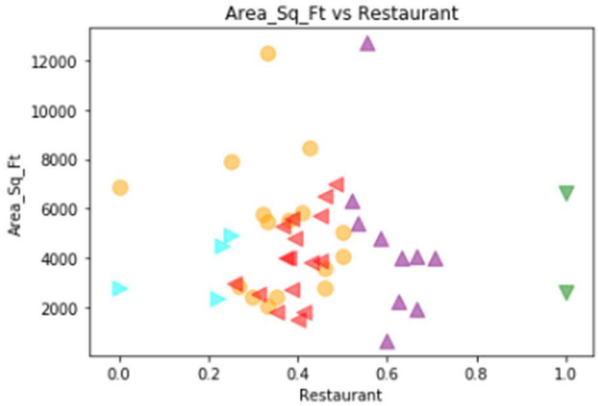


Restaurant vs Shop: From this next scatterplot, we can see that cluster 3 includes properties where the 'density' of Restaurants and Shop venues were both below certain thresholds (imagine a rectangular region in the lower right quadrant of this scatter plot, with Restaurants < ~0.5 & Shop < ~0.25). If Entertainment is important to have nearby for LloydMiller's new location, then it gathers that we would not have an excess of the other venue categories nearby. For restaurants, this is intuitive, as we would not want too large of a density of nearby restaurants (as this also means competition). For shops, this is less intuitive, but it could be that higher shop density may take away from Entertainment options.

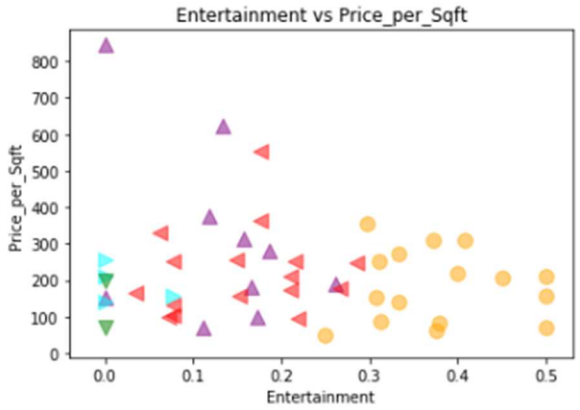


I wanted to explore the variables that were not used in our clustering algorithm to see if there was any incidental correlation.

Area_Sq_Ft vs Restaurant: From this scatterplot, we can see that Area_Sq_Ft density does not really distinguish the cluster when compared to Restaurant density.



Price_Sq_Ft vs Entertainment: From this scatterplot, we can see that Price_Sq_Ft density does not really distinguish the cluster when compared to Entertainment density.



Conclusion

The following list of properties is the reduced set of rental properties available near Naperville, IL that may be similar neighborhoods as the LloydMiller restaurant in the Queen Anne neighborhood in Seattle.

These properties are all characterized by having higher "Entertainment" venue density, moderate "Restaurant" density and lower "Shop" density as compared to the other property clusters.

Address	Price	Price_per_Sqft
211 N River St, Montgomery, IL	599000.0	48.699187
75-77 S Lasalle St, Aurora, IL	475000.0	59.974747
47 E Illinois Ave, Aurora, IL	289000.0	71.534653
18027 Dixie Highway, Homewood, IL	449000.0	81.134803
10036 S Western Ave, Chicago, IL	425000.0	84.830339
155 E Veterans Parkway, Yorkville, IL	286000.0	141.865079
227 W Wilson St, Batavia, IL	419000.0	150.882247
720 E North Ave, Carol Stream, IL	1080000.0	157.022390
460 S 1st St, Saint Charles, IL	1199000.0	208.449235
3938 N Central Ave, Chicago, IL	500000.0	209.555742
8347 W Grand Ave, River Grove, IL	625000.0	219.298246
2723 N Halsted St, Chicago, IL	1460000.0	251.031637
600 E Veterans Parkway, Yorkville, IL	1500000.0	273.922571
2913 N Lincoln Ave, Chicago, IL	750000.0	308.896211
1422-1424 W Chicago Ave, Chicago, IL	2625000.0	311.018957
3155 N Halsted Street, Chicago, IL	1275000.0	356.145251

This is a much more manageable list of properties to review. Additionally, we have learned that having nearby Entertainment options was an important variable in the k-means clustering, and it seems to make sense as well.