

April 6,
2019

Relationship between Venue Types and House Prices in Perth, WA

REPORT

XUECONG LIU

Table of Contents

1. Introduction	1
1.1 Background	1
1.2 Problem	1
1.3 Interest	1
2. Data acquisition and cleaning	1
2.1 Data acquisition	1
2.2 Data cleaning	2
2.3 Data features	2
3. Methodology and results	3
3.1 Suburb clustering	3
3.2 Relationship between house prices and venue ratios	5
3.3 The effect of distances on house prices	8
3.4 Suburbs on the map	16
4. Discussion	17
4.1 Having 50% venues food related makes the house prices highest	17
4.2 Higher ratios of some venue types means lower house prices	18
4.3 Being 18km away from the airport is best for the house prices	18
4.4 When larger distance from the urban center means a higher house price	18
4.5 Venue composition is an indicator of the local housing market	18
5. Conclusion	19
6. References	20
7. Appendix	20
A.1 Venue category merging keywords	20
A.2 Venue composition of different suburb clusters	20

1. Introduction

1.1 Background

Perth is the capital city of Western Australia, the fourth-most populous city in Australia, and some say the most isolated city in the world. It has over 200 suburbs, stretching from Two Rocks in the north to Singleton in the south, and east inland to The Lakes. ^[1]

When it comes to the house prices of these suburbs, distances to the nearest urban center, airport, and train station are without doubt having a big influence in it. Besides that, considering that different suburbs have different compositions of local venues - some have mainly restaurants and pubs, while some are enjoying more nature related venues - it would be interesting to see if and how the ratios of different venue types in the suburbs also contribute to the house prices.

If there is any relationship, understanding it might help predict house prices, especially when the local venues change significantly, help adjust the housing market by adjusting the policies to different venues, and help find the suburbs that are of the most potential to have a growing housing market this year by looking for the suburbs with a “right” composition of local venues.

1.2 Problem

The aim of this project is to qualitatively explore the relationship between the medium house prices of Perth suburbs, and the ratios of different venue types in these suburbs, with their distances to their nearest urban center, airport and train station also in the consideration.

1.3 Interest

This project might interest and benefit people who are looking to buy houses in a Perth suburb with a potential growing housing market, and government officials in Perth such as city managers who are seeking to adjust the local housing market.

2. Data acquisition and cleaning

2.1 Data acquisition

I scraped the housing prices for 2018 of 351 Perth suburbs in total from [reiwa.com](https://www.reiwa.com.au/),

their geological information from [Postcodes Australia](#) and Wikipedia. The current venue information of 281 unique categories was pulled from [Foursquare API](#) using the suburbs' latitudes and longitudes.

2.2 Data cleaning

There was house price information of 351 suburbs in total scraped, however, there were 49 suburbs with number 0 as their medium house prices, which is obviously due to the lack of data. I dropped these suburbs since they wouldn't be useful in this study.

Among the 302 suburbs left, there were 13 suburbs in total not found on [Postcodes Australia](#), where the latitudes, longitudes, distances to the nearest urban center, nearest airport and train station were acquired for the other suburbs. I scraped 12 out of the 13 suburbs' latitudes and longitudes from Wikipedia, with one of them not found anywhere, which I dropped. These suburbs don't have distances information, but they are still useful for analyzing the relationship between venues and house prices, so I decided to leave them out only when distances are being analyzed.

Foursquare API didn't find venue information for 4 of the remaining suburbs, so they were not studied in this project as well.

The data was not normalized, because this is only a qualitative research, and whether to normalize or not won't affect the pattern.

2.3 Data features

After data cleaning, there was 297 suburbs' data left, with 281 categories of venues in total, and 285 suburbs with information of their distances to the nearest urban center, airport and train station.

The medium house prices of the suburbs range from \$ 222,000 to \$3,575,000, and the suburbs are from 0 to 55 km away from the urban center, from 1.9 to 60 km away from the airports, and from 0 to 24 km away from the train stations.

There were between 1 and 60 venue types for every suburb, with 18.54 being the average number. Because different numbers of venues were pulled for different suburbs, the ratios of venues from every category to all the venues in the suburbs were calculated and used in this study, instead of the number of venues from different categories.

Also, noted that some suburbs only have small numbers of venues pulled, I

decided to merge similar venues categories, cluster the suburbs, and then study the patterns shown in the suburb cluster level, to reduce error caused by inaccuracies of the ratios of every single category in every single suburb.

3. Methodology and results

3.1 Suburb clustering

I used kMeans clustering here to cluster the suburbs. To determine the best number of clusters, F-test score and p-value were used to determine how different the prices are among clusters. The number that results in the biggest F-test score with a reasonably small p-value, i.e., the most significant difference among different clusters' medium house prices with a reasonable degree of confidence, would be the most ideal. I wrote a function to calculate the F-test score and p-value for a given number of clusters, so that I only needed to run it for a range of integers to see which number would be the best.

Before clustering suburbs, I merged similar venue categories on the same purpose of minimizing error. Lists of keywords for new categories were created, so that if one of the keywords was found in the original category, a new category name would be assigned to it. To determine how successful the merging was, I ran the F-test score and p-value calculating function after merging the categories, and then changed some questionable merging keywords and merged the original categories again to see whether the F-test score improved. I chose the set of merging keywords that gave me the highest F-test score in the end, and the original 281 categories were merged into 23 categories: alcohol, animal, arts, automotive, beverage, coast, education, factory, farm, food, healthcare, lodge, nature, office, recreation, rest, service, social, sports, store, structure, studio, transport. The keywords of these categories can be found in appendix A.1.

Using the best merging keywords found, I got the highest F-test score around 21.7 with a small p-value being around 1.6×10^{-9} when the suburbs were clustered into three groups, which is more than enough to guarantee significant differences in the prices among the clusters.

Having clustered the suburbs, I checked the overall ratios of different venue

categories and the 10 most common venues (original venue categories) of every suburb in these three suburb clusters. Here is what I found about them:

The first cluster (Cluster 0) is the closest to the nature, also features stores and alcohol. It has the smallest ratio of food venues, but the largest ratios of alcohol venues, lodges and transport venues (possibly because there are more transient population and less permanent population than the other clusters). It's also big on sports (possibly because the land is cheaper, which is good for building sports venues). Unlike the other two, it has little education venues, offices or social clubs, but with about 10 times more farms and 5 times more studios.

The second cluster (Cluster 1) has more than half venues being food venues, which is the largest among the three, with stores being its second most popular venue category, the ratio of which is nevertheless the lowest compared to the other two clusters. It has the lowest ratios of sports venues and nature venues. Unlike the other two, this cluster has more alcohol venues than nature venues. It's also the only cluster that has a decent amount of social clubs.

The third cluster (Cluster 2) features stores and food, with a decent ratio of sports and nature venues. It has no farms, the smallest ratio of alcohol venues and lodge venues, but the largest ratio of education venues among the three clusters. It also has the most pet related venues, healthcare venues, and automotive venues (including gas stations and auto garages).

Tables of venue category rankings and other related detailed information of these three suburb clusters can be found in appendix A.2.



As shown in the Figure 1, the suburbs in Cluster 1 (the second cluster) have the highest mean of medium house prices, about \$700,000, along with the widest price range. Cluster 2 (the third cluster) has the lowest price mean, just less than \$500,000, but its price range is larger than Cluster 0, with its third quartile value slightly larger than that of Cluster 0. The mean price of Cluster 0 (the first cluster) is very close to \$500,000, and this cluster has the smallest price range.

The only variable I chose here in the boxplot in Figure 1 was the ratios of venue categories, thus this difference in the prices among clusters is a proof that there is a relationship between the medium house price of a suburb and its ratios of different venue categories.

3.2 Relationship between house prices and venue ratios

Having understood that there is a relationship between house prices and venue ratios, I started to check how individual venue categories might contribute to this relationship.

Because the degree of dispersion of the prices is not consistent for different ratios of a venue category, it's not a good idea to use linear models to describe the relationship between the price and the ratio. I decided that it would be better to bin the ratios first, and then explore how the price mean and price standard deviation are influenced by the change of the mean ratio of bins.

For every venue category, I binned the ratios into the maximum number of bins that allows every bin to have at least one sample. I only studied the categories that have more than 3 bins, because if the number of bins is less than that, the range of ratios would be so small that the errors would be too big when trying to find its relationship with prices. Bearing the errors in mind, I also chose not to study the venue categories with less than 3 bins having more than 3 samples, and those that have less than 30 samples in total. After filtering, there were 5 venue categories left for me to explore further: automotive, beverage, food, sports, and store.

The Automotive category (Figure 2) includes mainly gas stations and auto garages. The figure shows that most of the suburbs have less than 7% venues in this category. Within this range, the more Automotive venues there are in the suburbs, the lower their mean medium house price is, and lower the price standard deviation is. From 7% to 12%, this declining trend is slower, with fewer suburbs in the sample. After

12%, the number of suburbs is too small to make any meaningful conclusion out of it.

Fig2. Relationship between Ratio of Automotive Venues and Medium House Price in a Suburb

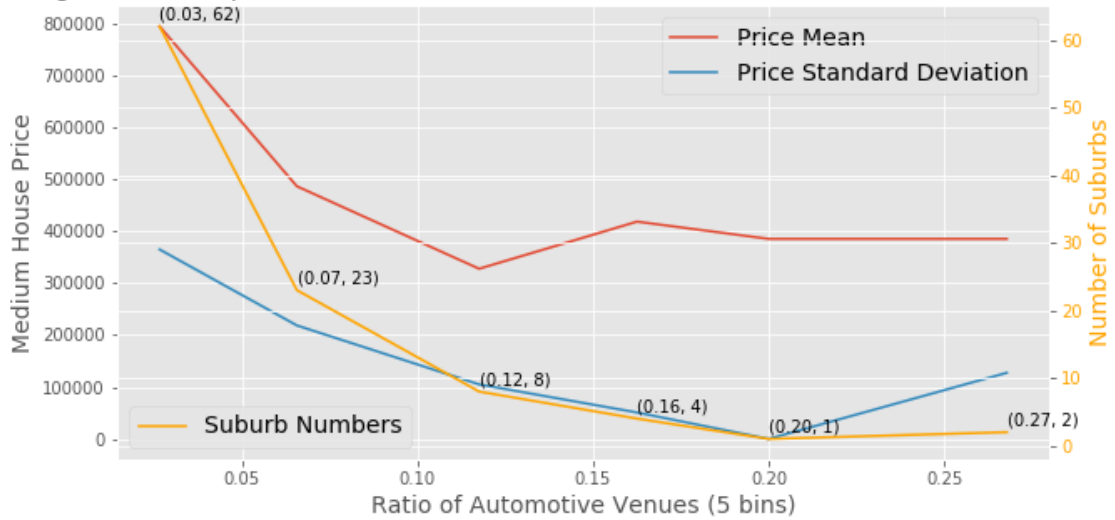
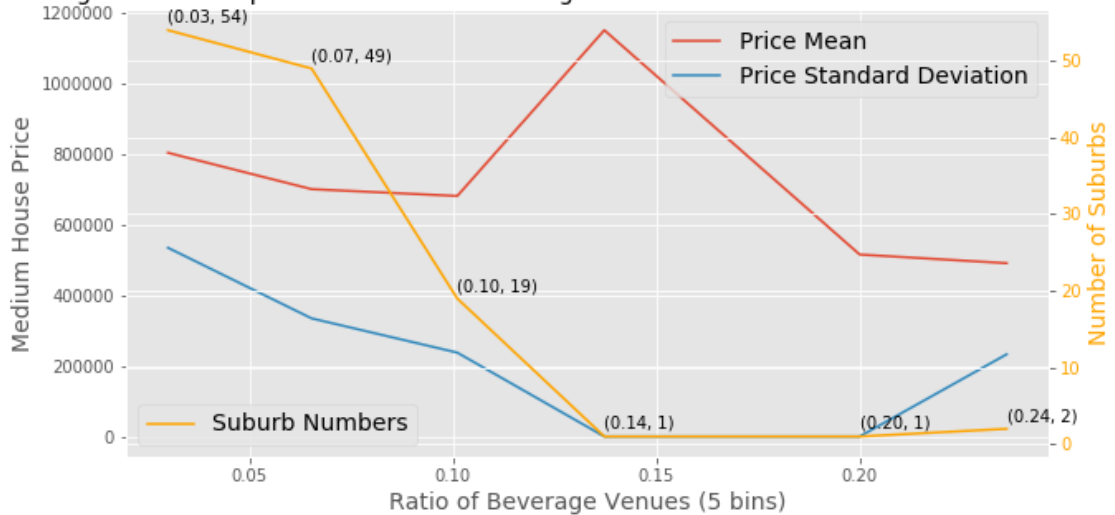


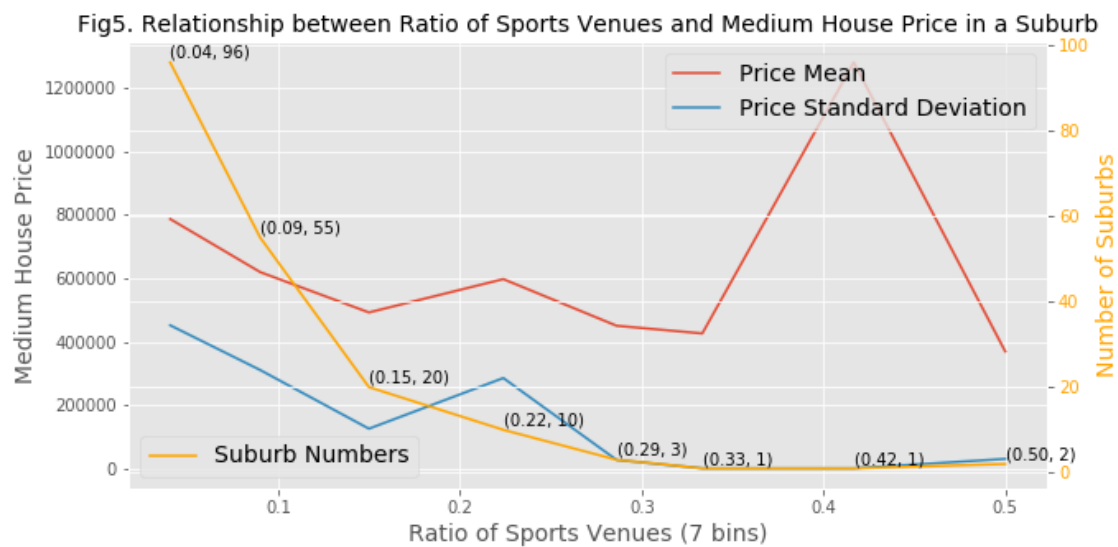
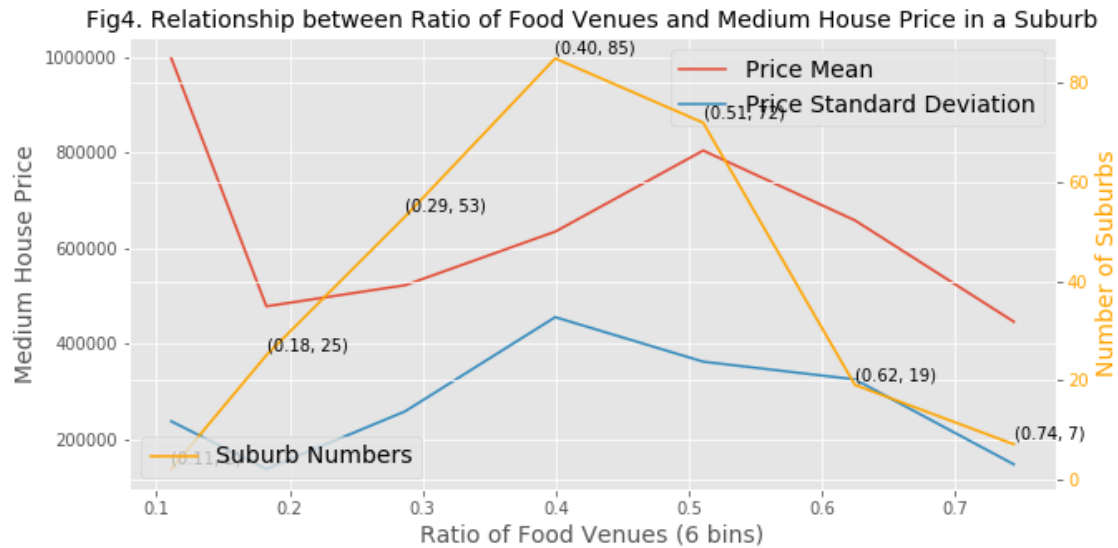
Fig3. Relationship between Ratio of Beverage Venues and Medium House Price in a Suburb



The Beverage category (Figure 3) includes all venue categories that have the keyword "tea" or "coffee". Most of the suburbs have less than 10% of their venues being beverage venues. Between 0% and 10%, the higher the ratio of beverage venues is, the more concentrated the medium house prices are around a lower mean value. After 10%, the sample size is so small that the change of prices is more likely due to the influence of other categories.

The Food category (Figure 4) includes venues that mainly sell food. All suburbs have at least 10% venues being in this category, with most of them having 18% to 62%. There are a very small number of suburbs in the first bin with only 10% Food venues but a very high price mean, and we can't be sure how big a role the Food venues play in this high price, so it's better to leave it out in this analysis. From around 18% on, the

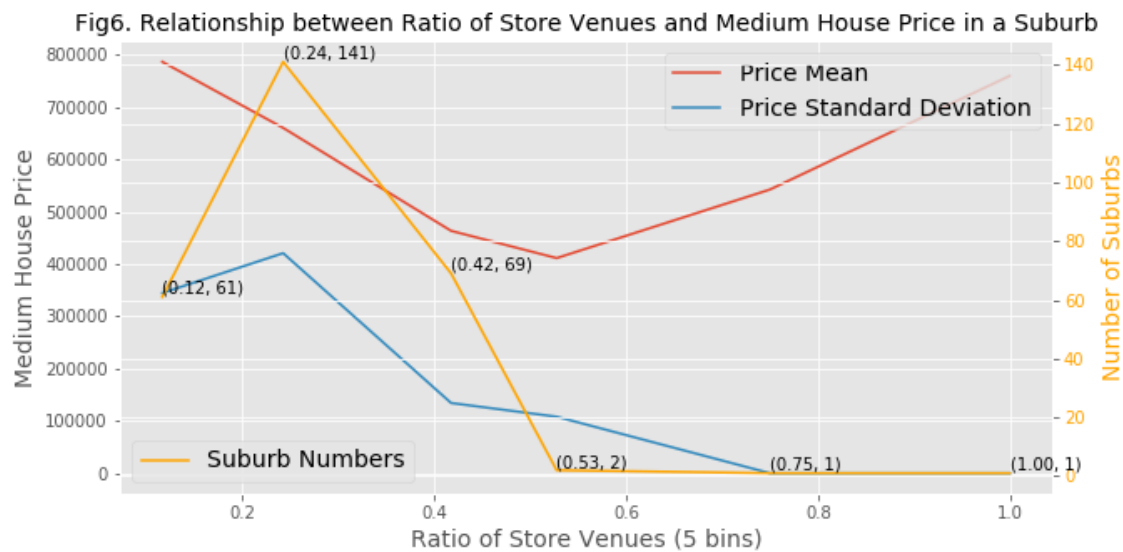
mean price of the medium house prices grows, peaks at 51%, and then declines. The price standard deviation grows as the ratio increases from 18% to 40%, and then declines.



The Sports category (Figure 5) includes gyms and venues used for sports. In the figure, the price mean and price standard deviation decline slowly as the ratio of sports venues increases from 4% to 15%, the range of which nearly all suburbs are in, and then both increase slightly until the ratio reaches 22%. After 22%, the sample size is too small to show any meaningful trends.

The Store category (Figure 6) includes all the supermarkets, markets, stores, malls, etc., the ratio of which is around 24% for most of the suburbs. In all the bins with good amounts of suburbs, the price mean declines as the ratio grows from 12% to

42%, while the price standard deviation rises first before declines, with its peak at 24%.



If only bins with decent amount of samples are considered, in these 5 venue categories, except for food and store venues, the ratio ranges of them are around 10% most of the suburbs, the average medium house price and the price deviation of the suburbs both decline as the ratio increases. Store venues have a bigger range of ratio range being around 30%, and its price deviation increases first before declining when the ratio increases, but its average suburb medium house price is also declining.

The food venue has the largest ratio range for the majority of the suburbs, which is around 44%. The price mean and deviation go up first before they go down as the ratio grows, with the peak of mean around 51%, which happens to be around the average food venue ratio of the second cluster 50.5%.

3.3 The effect of distances on house prices

It's a common sense that distances to the nearest urban center, airport and train station are very influential to house prices, but will they affect house prices differently in different suburb clusters? I made some plots hoping to gain a better understanding in the relationship between these distances and house prices. (Figure 7-9)

From the figure, it's clear that same as when dealing with the venue category ratio data and price data, the degree of dispersion of the prices is not consistent for different distances, making it hard for regression models to make sense. The same method of binning variables and then plotting the means and standard deviations of the prices would be suitable here. Using the same method, I had a closer look at the price –

distance relationships in different clusters (Figure 10 – 18), and also in all suburbs together (Figure 19 – 21).

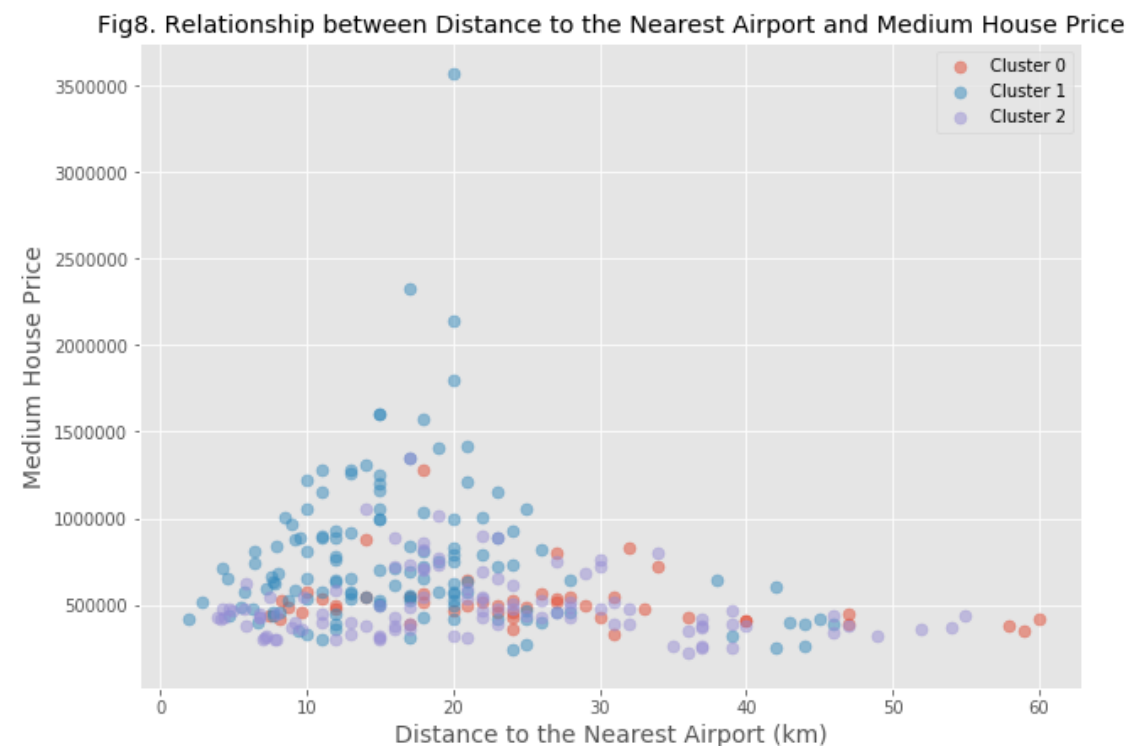
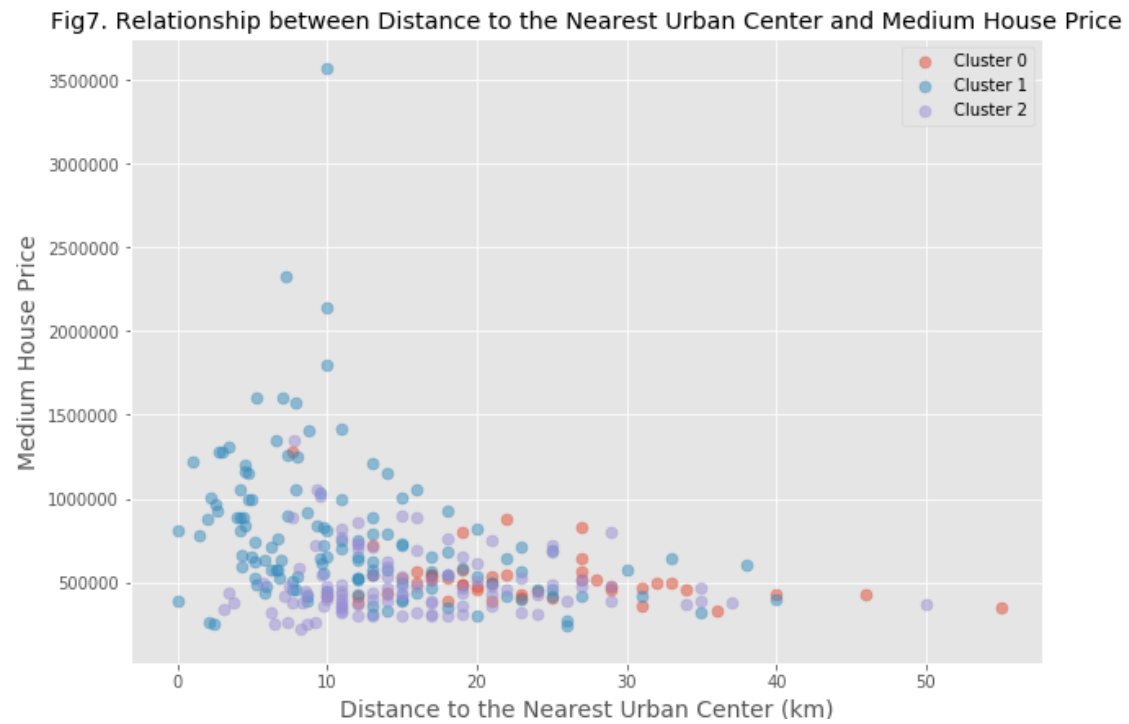
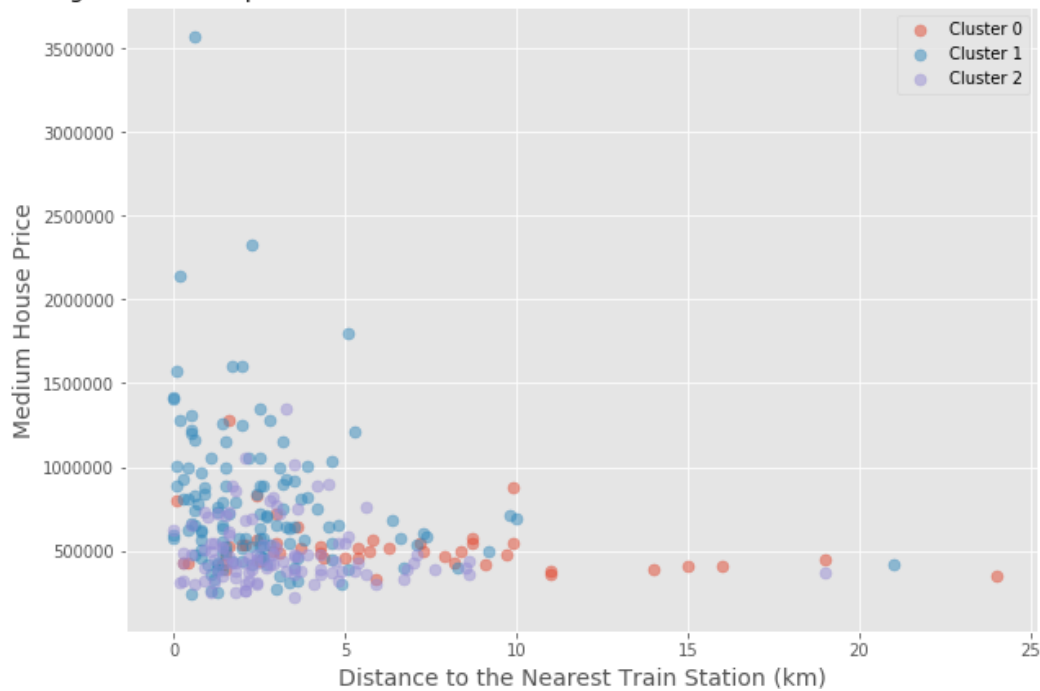
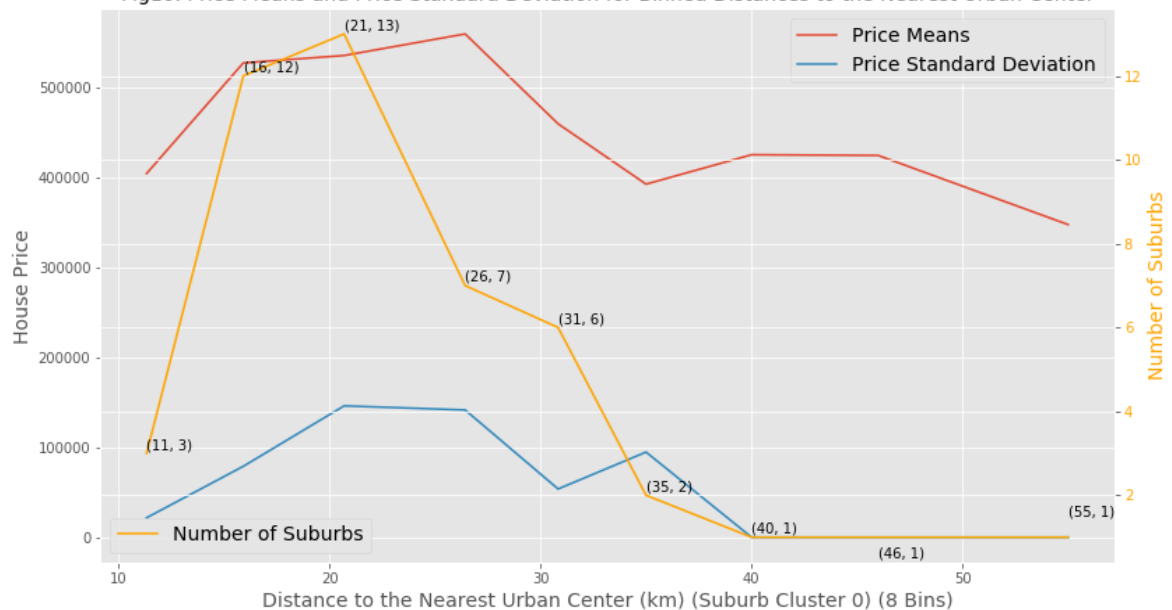


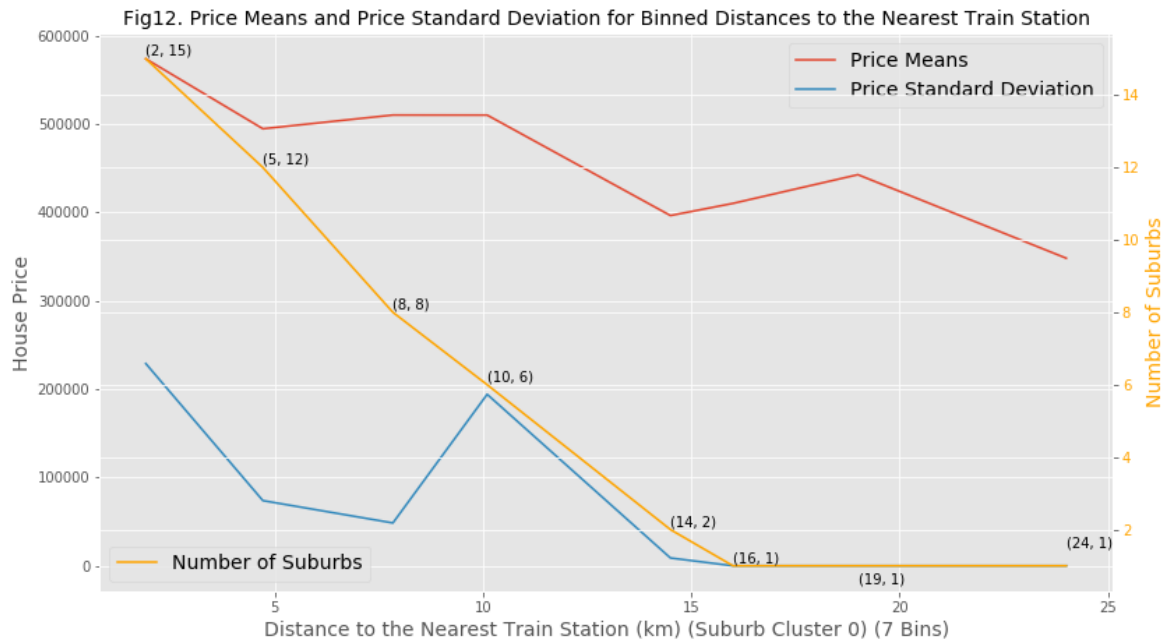
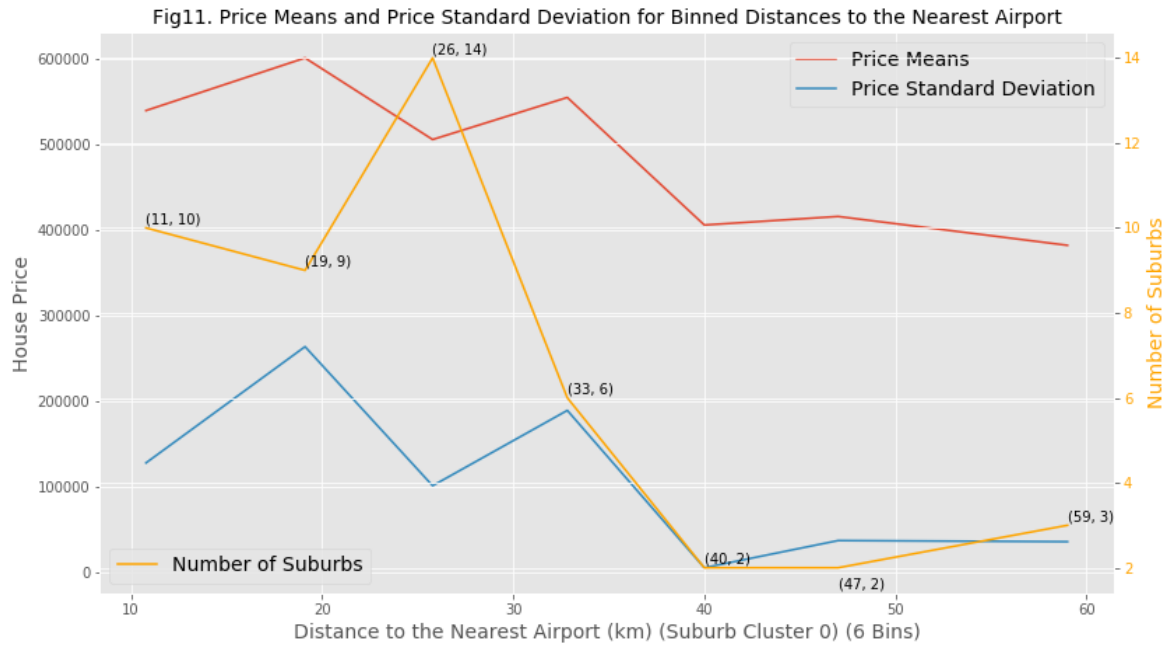
Fig9. Relationship between Distance to the Nearest Train Station and Medium House Price



Most of the suburbs in **the first cluster** are 16 to 31 km away from their nearest urban centers, 11 to 33 km away from their nearest airports, and no more than 10 km away from their nearest train station. Their medium house prices mean is the highest when between 16 to 31 km away from the urban centers, or 11 to 33 km away from airports, along with the largest price ranges. The suburbs with distances to their nearest train station being around 2 km or 10 km have the highest medium house price mean, and also the largest price range. (Figure 10 – 12)

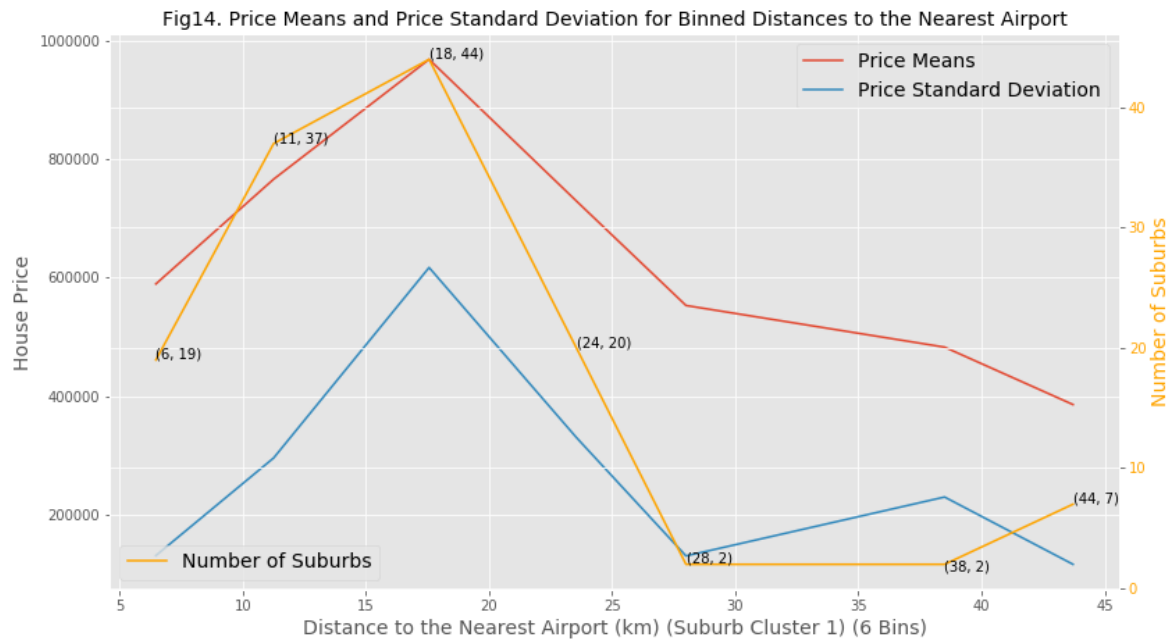
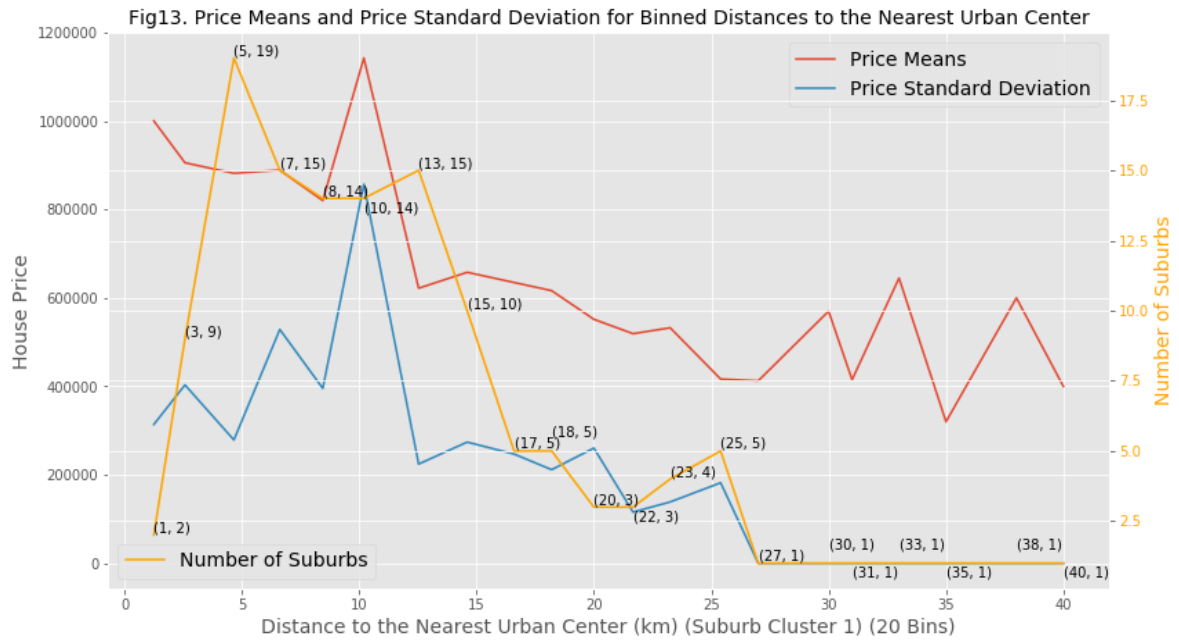
Fig10. Price Means and Price Standard Deviation for Binned Distances to the Nearest Urban Center



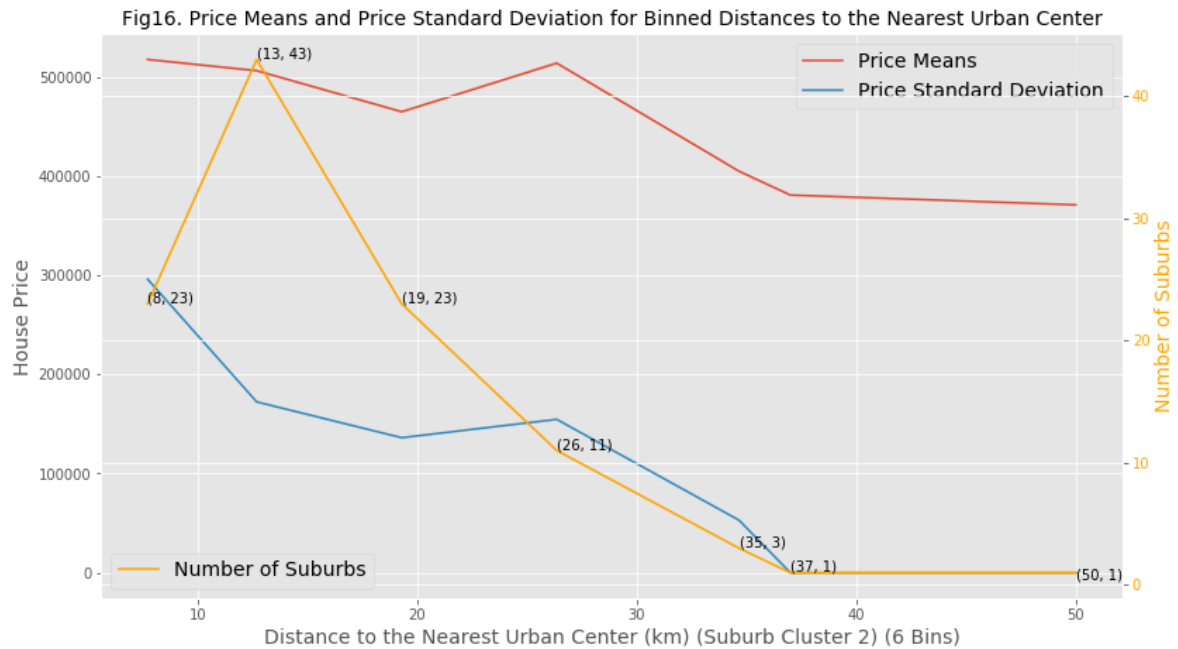
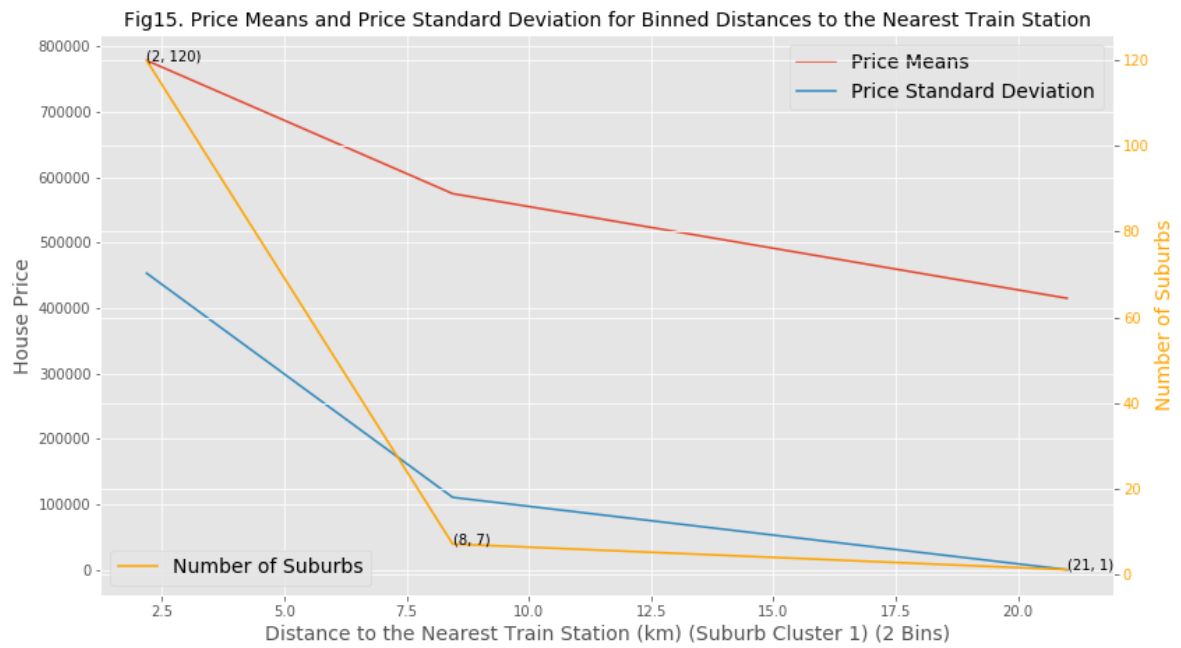


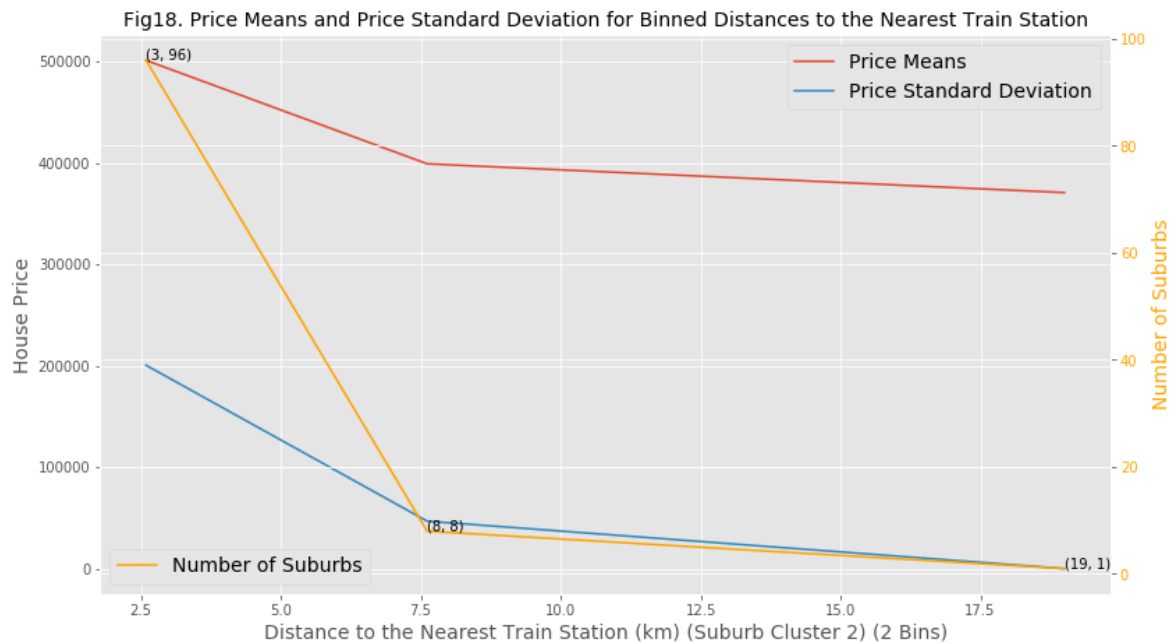
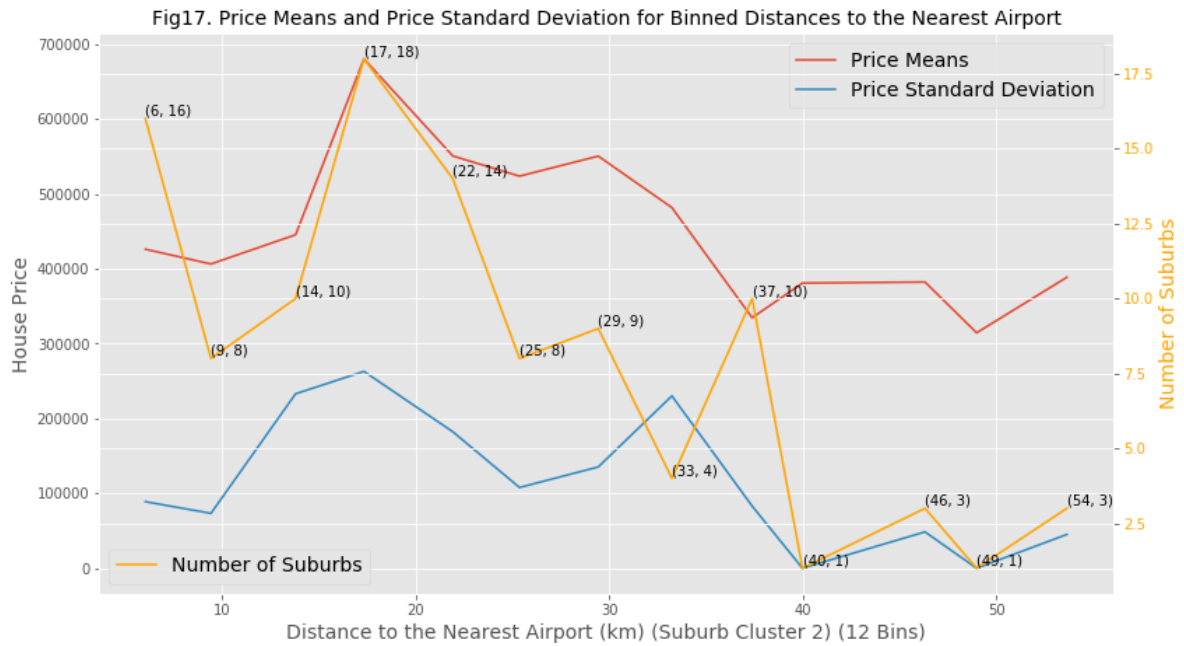
Most of the suburbs in **the second cluster** are 3 to 15 km away from their nearest urban centers and 6 to 24 km away from their nearest airport. Almost all of them are around 2 km away from their nearest train stations. The average medium house price and price standard deviation are both the highest when the suburbs are around 10 km from their nearest urban centers, 18 km from their nearest airports, or 2 km from their nearest train stations. (Figure 13 – 15)

Most of the suburbs in **the third cluster** are 8 to 26 km away from their nearest urban centers, 6 to 37 km away from their nearest airports, and almost all of them are

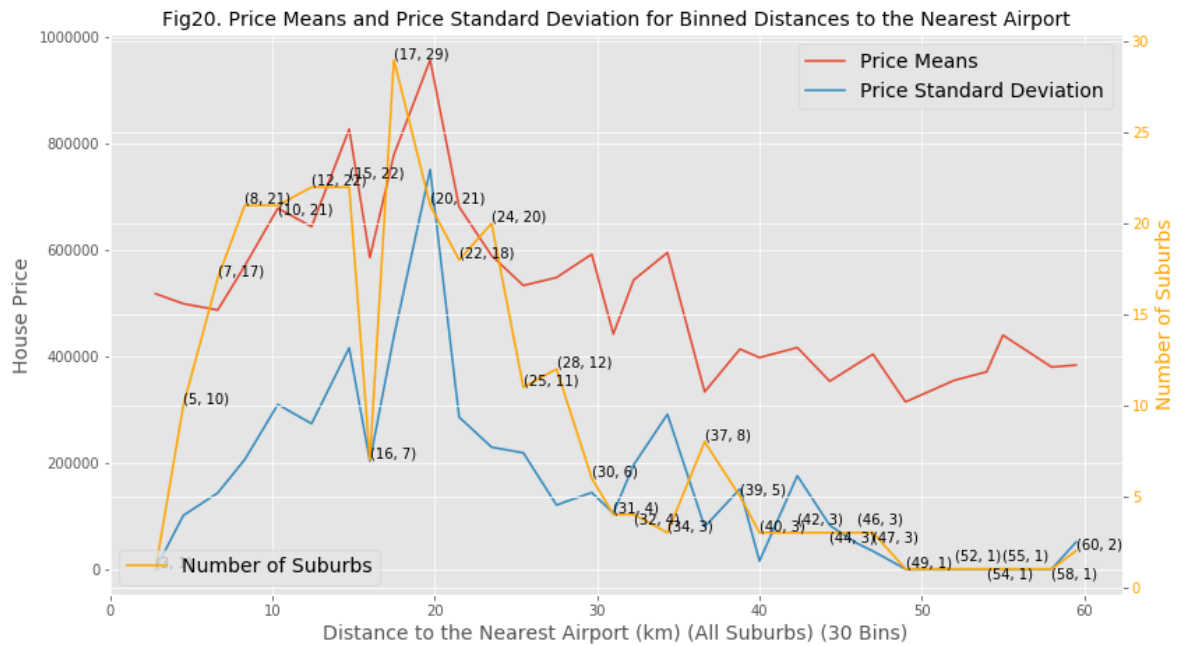
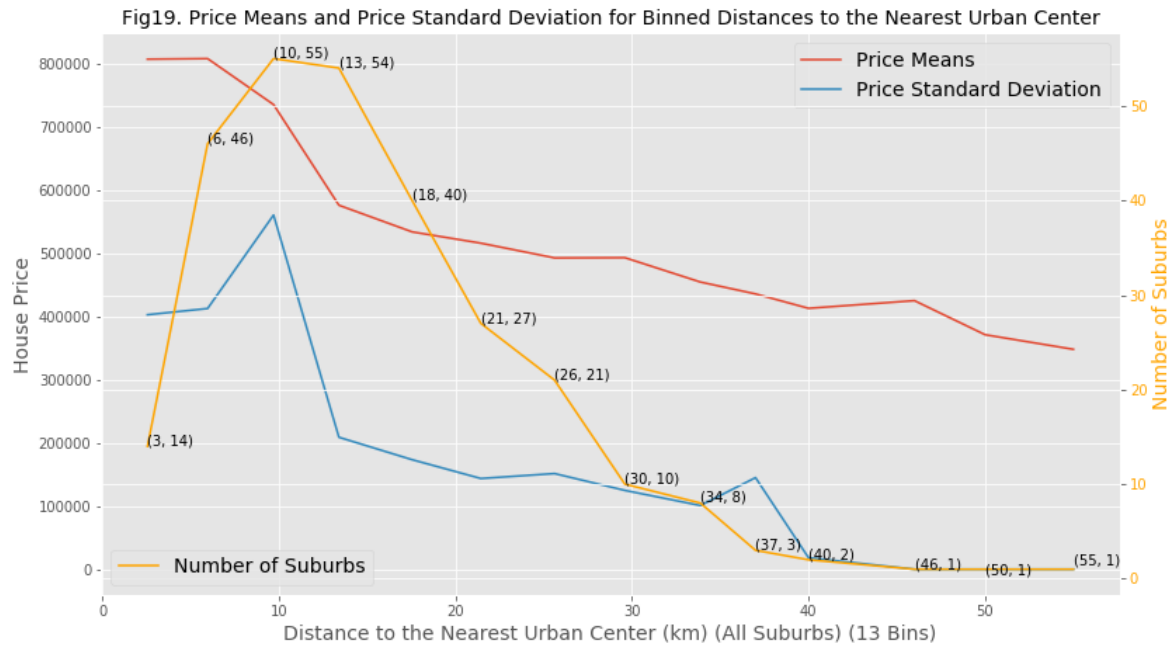


around 3 km away from their nearest train station. The distance to the nearest urban center doesn't affect the medium house price mean much when it's between 8 to 26 km, with only small fluctuation, but the standard deviation does have an overall declining trend. The suburbs that are 14 to 33 km away from their nearest airports have the highest medium house price mean and price deviation, with their peaks both at around 17 km. The average medium house price and the price standard deviation of the suburbs declines when distance to their nearest train station increases. (Figure 16 – 18)



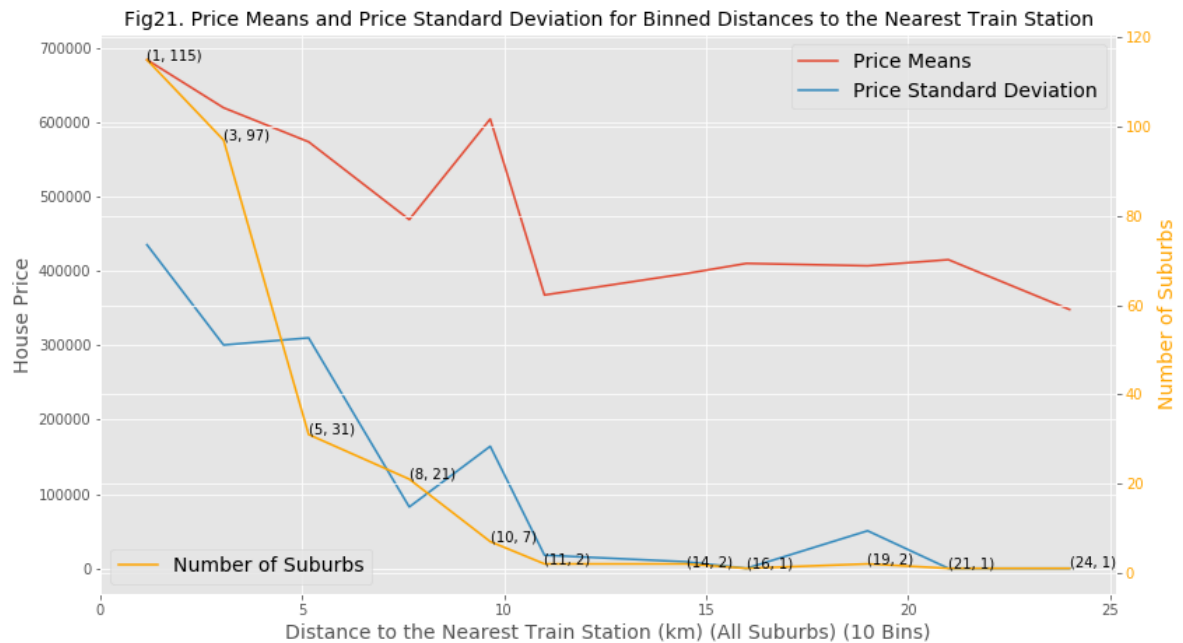


Most of the suburbs in Perth are within 30 km from the urban center, 5 to 28 km away from the airports, and within 8 km from the train stations. Overall, the mean of the medium house prices of the suburbs declines as the distance to the urban center increases, while the price deviation increases first before it starts to decline after 10 km. The price mean is the highest when the suburbs are 8 to 22 km away from the airports, along with the largest price deviation. Less than 5 km or around 10 km away from the train stations is where you'll find the suburbs with the highest average medium house price, and the price range of the suburbs within 5 km away from the train stations is the largest. (Figure 19 – 21)



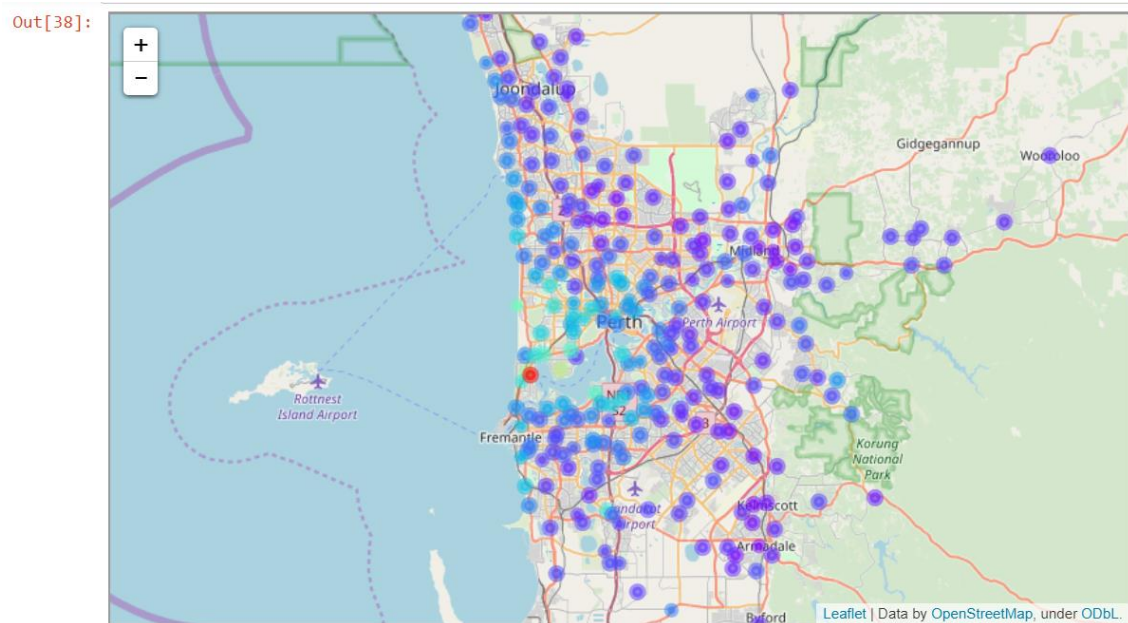
These figures shows that distances do have slightly different impacts on house prices for suburbs from different clusters. For example, the suburbs from the first cluster experience a rise in average house price first as the distance to the urban center increases, and only decrease when the distance is as big as around 30 km away, while those from the other two clusters generally have the highest average house prices when they are the closest to the urban center. The suburbs from the second and third clusters are mostly within 3 km from the train stations, while the first cluster suburbs are generally farther away but with their house prices less affected by this distance. Nevertheless, they all have the highest average house prices when they are

around 18 km away from the airport.



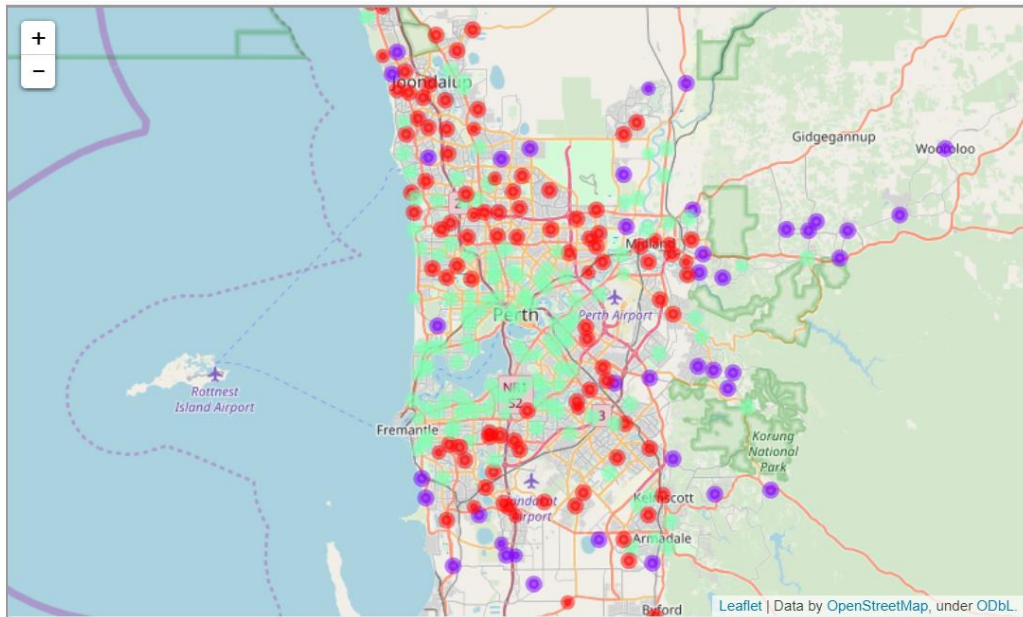
3.4 Suburbs on the map

I finally marked these suburbs on Perth map with their marker colors representing their medium house price (Map1) / cluster label (Map2).



Map1. Suburbs Marked with Colors representing their house price on Perth Map, the darker the marker, the cheaper the medium house price for the suburb

Out[39]:



Map2. Suburbs Marked with Colors representing their house price on Perth Map (Cluster 0 purple, Cluster 1 green, Cluster 2 red)

As shown in the first map, the suburbs close to the rivers or the coast are generally the suburbs with the highest medium house prices. Interestingly, in the second map, we can see those suburbs are mostly in the second cluster (Cluster 1). The suburbs in the first cluster (Cluster 0) are the farthest from the urban center, with many of them being the farthest from the coast. The third cluster (Cluster 2) suburbs are mostly between the Cluster 1 area and the Cluster 0 area.

Please feel free to check out the interactive version of these two maps in my notebook EDA.ipynb.

4. Discussion

4.1 Having 50% venues food related makes the house prices highest

Within a range, the more food venues there are in the local venue composition, the higher the possibility that the food venues are not only serving the local residents, but also people from other areas, which is likely suggesting the more capitals people bring to this suburb, along with the rise in local house prices.

When food venue ratio reaches a threshold, it's likely that the higher the ratio, the fewer other venue categories there are in this suburb, which eventually makes this suburb a big busy food market instead of a livable area, thus goes the decrease of the

house prices.

As analyzed in the methodology section, the best ratio is around 50% for the suburbs in Perth, which is nearly the same as the average food venue ratio of the suburbs from the second cluster that has the highest house price mean and deviation.

4.2 Higher ratios of some venue types means lower house prices

For the 5 venue type ratios studies in the 3.2 section, the higher the ratio are, the lower average house price is, with food venue being the only exception. It's possibly because, unlike food venue having a large ratio range and reflecting the non-locals and capitals the suburb attracts, venues like automotive venues, beverage venues, sports venues and store venues usually only have a small ratio and a small range to serve the locals. When the ratio deviates from this range, the venues would be overserving and consuming the capitals that could have been used somewhere more in need. Or it might also mean the local population and capitals are so small that there are only a limited number of venues there, and the ratios are easy to become a bit extreme in this case.

4.3 Being 18km away from the airport is best for the house prices

Considering the distance between Perth airport and the beach is approximately 19 km and how much Australians love the sandy beaches, it's not hard to understand why the average house prices are the highest when the suburbs are around 18 km away from the airport.

4.4 When larger distance from the urban center means a higher house price

As for the suburbs from the first cluster, it's shown in Figure 10 that their average house price rises when the distance between the urban center and the suburbs increases from 11 to 26 km. It's possible that within this range, the number of suburbs from the other two clusters are decreasing rapidly, making first cluster suburbs the most popular choice, hence, its average house price rises.

4.5 Venue composition is an indicator of the local housing market

It's true that venues can bring changes to the local economy to an extent. However, in the same city and under the same policies, geological locations of the suburbs are probably an important, if not the most important factor of their economy. For example, some coastal suburbs are more connected to the world than inland areas, thus they are usually more developed.

The difference in the geological distribution of the suburbs from different cluster in Map2 is probably suggesting that, instead of affecting house prices and local economy, venue types do a better job reflecting the local economy and hence are an indicator of local house prices.

If these assumptions are true, the suburbs in the second cluster that have low house prices at the moment might be of the most potential to have a growing housing market in 2019, especially the coastal ones.

5. Conclusion

In this project, I clustered the Perth suburbs by their venue compositions, explored the differences in the house prices among different suburb clusters, including how distances to the nearest urban center, the nearest airport, the nearest train station affect the suburbs in different clusters differently, and investigated the relationships of venue category ratio and house prices for some venue categories.

The differences in house prices among different cluster clearly showed that local venue composition and affect / indicates the local house prices to a degree. This relationship between venue compositions and house prices is complicated and non-linear, with the food venues having the clearest and most obvious relationship with house prices. It's also found that the distances to the nearest urban center, airport, and train station affect suburbs with different venue compositions slightly differently.

The limit of this study is, there were only small numbers of venues pulled for some suburbs. Thus, the relationships of the ratios of some venue categories and the house prices are not studied thoroughly. If there were more data, there might have been different conclusions drawn for these venue categories. Besides that, this project only studied the suburbs in Perth, WA, so the conclusions might not be applicable to other areas.

6. References

[1] (2019) *Perth* - *Wikipedia*, Available at: <https://en.wikipedia.org/wiki/Perth#Isolation> (Accessed: 6th April 2019).

7. Appendix

A.1 Venue category merging keywords

Alcohol: 'bar', 'pub', 'nightclub', 'speakeasy', 'wine', 'beer'

Animal: 'pet', 'dog', 'cat', 'animal'

Arts: 'art', 'museum', 'opera', 'music', 'concert', 'exhibit', 'historic', 'library', 'cultur', 'planetarium', 'sculpture', 'memorial'

Automotive: 'auto', 'gas station'

Beverage: 'tea', 'coffee'

Coast: 'waterfront', 'beach', 'pier', 'harbor', 'boat', 'lighthouse'

Education: 'school', 'college'

Factory: 'factory', 'distillery', 'brewery'

Farm: 'farm', 'vineyard', 'stable'

Food: 'buffet', 'food', 'restaurant', 'bakery', 'chicken', 'wings joint', 'café', 'cafe', 'soup', 'salad', 'sandwich', 'pizza', 'bbq', 'diner', 'steak', 'noodle', 'deli', 'bistro', 'burrito', 'burger', 'breakfast', 'taco', 'snack', 'butcher', 'bagel', 'chocolate', 'donut', 'cupcake', 'dessert', 'chips', 'yogurt', 'pastry', 'pie', 'sausage', 'creperie'

Healthcare: 'pharmacy', 'doctor', 'medical', 'chiropractor', 'hospital'

Lodge: 'hotel', 'motel', 'hostel', 'bed', 'resort'

Nature: 'camp', 'tree', 'nature', 'zoo', 'aquarium', 'garden', 'river', 'park', 'lookout', 'nightlife', 'mountain', 'forest', 'lake', 'fountain', 'landscap'

Office: 'office', 'agency', 'department'

Recreation: 'casino', 'entertainment', 'vacation', 'theater', 'circus', 'comedy', 'play', 'recreation', 'arcade', 'newsstand', 'game', 'multiplex', 'radio'

Rest: 'rest', 'lounge'

Service: 'information', 'service', 'cleaner', 'salon', 'tattoo', 'spa', 'massage', 'repair', 'currency', 'locksmith', 'laundromat', 'tailor'

Social: 'social', 'club', 'community center'

Sports: 'rink', 'ski', 'rugby', 'surf', 'gym', 'sports', 'tennis', 'soccer', 'weight loss', 'yoga', 'pool', 'skating', 'golf', 'ball', 'outdoors', 'bowling', 'bike', 'hockey'

Store: 'store', 'shop', 'boutique', 'market', 'grocery', 'plaza', 'retailer'

Storage: 'storage', 'warehouse'

Structure: 'building', 'neighborhood', 'canal', 'bridge', 'stadium', 'housing'

Studio: 'studio'

Transport: 'airport', 'bus', 'train', 'transportation', 'rail', 'track', 'metro', 'intersection', 'heliport', 'platform', 'moving target', 'tram'

A.2 Venue composition of different suburb clusters

	count	mean	std	min	25%	50%	75%	max
Nature	49	0.220555	0.197753	0	0.066667	0.2	0.333333	1
Store	49	0.206576	0.180044	0	0	0.2	0.307692	0.5
Alcohol	49	0.109378	0.217102	0	0	0	0.142857	1
Sports	49	0.083277	0.133286	0	0	0	0.142857	0.5
Transport	49	0.074811	0.135414	0	0	0	0.125	0.5
Food	49	0.070818	0.094216	0	0	0	0.166667	0.25
Lodge	49	0.044898	0.165594	0	0	0	0	1
Farm	49	0.042371	0.109364	0	0	0	0	0.5
Coast	49	0.031388	0.0948	0	0	0	0	0.5
Service	49	0.030661	0.090335	0	0	0	0	0.5
Recreation	49	0.023922	0.07458	0	0	0	0	0.428571
Arts	49	0.020563	0.082642	0	0	0	0	0.5
Healthcare	49	0.009256	0.032532	0	0	0	0	0.142857
Beverage	49	0.00737	0.038779	0	0	0	0	0.25
Factory	49	0.006122	0.031677	0	0	0	0	0.2
Studio	49	0.005183	0.025592	0	0	0	0	0.142857
Rest	49	0.005102	0.035714	0	0	0	0	0.25
Automotive	49	0.002704	0.013421	0	0	0	0	0.076923
Structure	49	0.002551	0.017857	0	0	0	0	0.125
Animal	49	0.002494	0.01227	0	0	0	0	0.066667
Office	49	0	0	0	0	0	0	0
Social	49	0	0	0	0	0	0	0
Education	49	0	0	0	0	0	0	0

Table1. Venue Composition of Suburb Cluster 0

	count	mean	std	min	25%	50%	75%	max
Food	134	0.505087	0.099848	0.25	0.444444	0.5	0.559286	0.8
Store	134	0.16493	0.071331	0	0.118848	0.175439	0.214286	0.3125
Alcohol	134	0.063815	0.053993	0	0.020944	0.057418	0.092495	0.25
Nature	134	0.046687	0.047558	0	0	0.037386	0.071429	0.25
Sports	134	0.044543	0.045344	0	0	0.035714	0.066667	0.25
Beverage	134	0.034466	0.037585	0	0	0.027402	0.060624	0.222222
Transport	134	0.022895	0.044263	0	0	0	0.022222	0.25
Arts	134	0.020224	0.031805	0	0	0	0.031622	0.25
Recreation	134	0.017296	0.026141	0	0	0	0.029202	0.125
Coast	134	0.016086	0.040327	0	0	0	0.01	0.272727
Automotive	134	0.012112	0.022047	0	0	0	0.016949	0.111111
Factory	134	0.008482	0.047883	0	0	0	0	0.5
Animal	134	0.008065	0.023471	0	0	0	0	0.2
Lodge	134	0.007397	0.022696	0	0	0	0	0.192308
Healthcare	134	0.005272	0.012456	0	0	0	0	0.090909
Structure	134	0.005204	0.020941	0	0	0	0	0.2
Service	134	0.005016	0.025625	0	0	0	0	0.25
Farm	134	0.004472	0.026401	0	0	0	0	0.222222
Education	134	0.002745	0.009695	0	0	0	0	0.052632
Office	134	0.001851	0.007854	0	0	0	0	0.071429
Rest	134	0.001549	0.005093	0	0	0	0	0.035714
Studio	134	0.000913	0.004906	0	0	0	0	0.035714
Social	134	0.000894	0.003998	0	0	0	0	0.03125

Table2. Venue Composition of Suburb Cluster 1

	count	mean	std	min	25%	50%	75%	max
Store	114	0.357438	0.11019	0.142857	0.285714	0.34699	0.4	1
Food	114	0.321603	0.101583	0	0.272727	0.337719	0.3907	0.571429
Sports	114	0.058277	0.067314	0	0	0.044949	0.08961	0.3
Nature	114	0.052965	0.055414	0	0	0.045455	0.083333	0.25
Automotive	114	0.032633	0.054958	0	0	0	0.043247	0.285714
Arts	114	0.027742	0.044571	0	0	0	0.044729	0.25
Alcohol	114	0.024013	0.035747	0	0	0	0.03828	0.142857
Beverage	114	0.022574	0.03717	0	0	0	0.04	0.2
Transport	114	0.021414	0.034908	0	0	0	0.035406	0.166667
Coast	114	0.021233	0.07298	0	0	0	0	0.461538
Recreation	114	0.018816	0.036002	0	0	0	0.02987	0.25
Healthcare	114	0.017606	0.040535	0	0	0	0.019616	0.25
Animal	114	0.010763	0.032023	0	0	0	0	0.25
Factory	114	0.003456	0.014067	0	0	0	0	0.117647
Education	114	0.003166	0.024505	0	0	0	0	0.25
Service	114	0.00206	0.008218	0	0	0	0	0.043478
Lodge	114	0.001887	0.007879	0	0	0	0	0.047619
Office	114	0.001357	0.007299	0	0	0	0	0.05
Structure	114	0.000649	0.005077	0	0	0	0	0.047619
Social	114	0.000351	0.003746	0	0	0	0	0.04
Rest	114	0	0	0	0	0	0	0
Farm	114	0	0	0	0	0	0	0
Studio	114	0	0	0	0	0	0	0

Table3. Venue Composition of Suburb Cluster 2