

# Choosing a suburb city when you work in Dallas

Juliano Zanardo

January 22, 2021

## 1. Introduction

Dallas is a major city concentrating many businesses and corporate headquarters, therefore hundreds of thousands of jobs.

Housing in the city is either very expensive or is poorly served in terms of services and infrastructure, reason why many people who work in Dallas look for houses in the surrounding suburb cities, where the combination of lower prices and services offer is more attractive.

This Capstone project explores, segments and clusters the suburb cities in the Dallas-Fort Worth Metroplex area, within 35 miles distance from downtown Dallas, to cluster similar cities in terms of offer of education, healthcare, shopping, outdoor options, entertainment, restaurants and hotels, in order to provide an insight of better fit depending on the house buyer interests.

## 2. Data Acquisition and Cleaning

As the project explores, segments and clusters the suburb cities in the Dallas-Fort Worth Metroplex area, the following data components are obtained and steps are executed:

1. Produce a CSV file with the list of all the cities in the DFW area from <https://www.hdavidballinger.com/dfw-metroplex.php> and load it into a dataframe. The list provides the population and the distance from downtown Dallas for each city.
2. Filter Dallas, Fort Worth and cities farer than 35 miles from downtown Dallas out from the list.
3. Obtain, via API, the geolocation (latitude and longitude) of each city and add it to the data frame.
4. Obtain the nearby POIs from FourSquare, via API, for the target cities.
5. Pivot the POIs to count the number of schools, hospitals, groceries / malls, parks, theaters, restaurants and hotels per city.
6. Rank the venue categories for each city, from the 1st to the 7th most common.

This data allows, then, to segment and cluster the suburb cities and perform analysis on the profile of each one in terms of services offered.

### **3. Methodology**

In order to segment the suburb cities surrounding Dallas based on the most common services offered in each one, the KMeans clustering model was applied.

The clustering was based in the 7 most common venue categories of each city and K was set with the value 7.

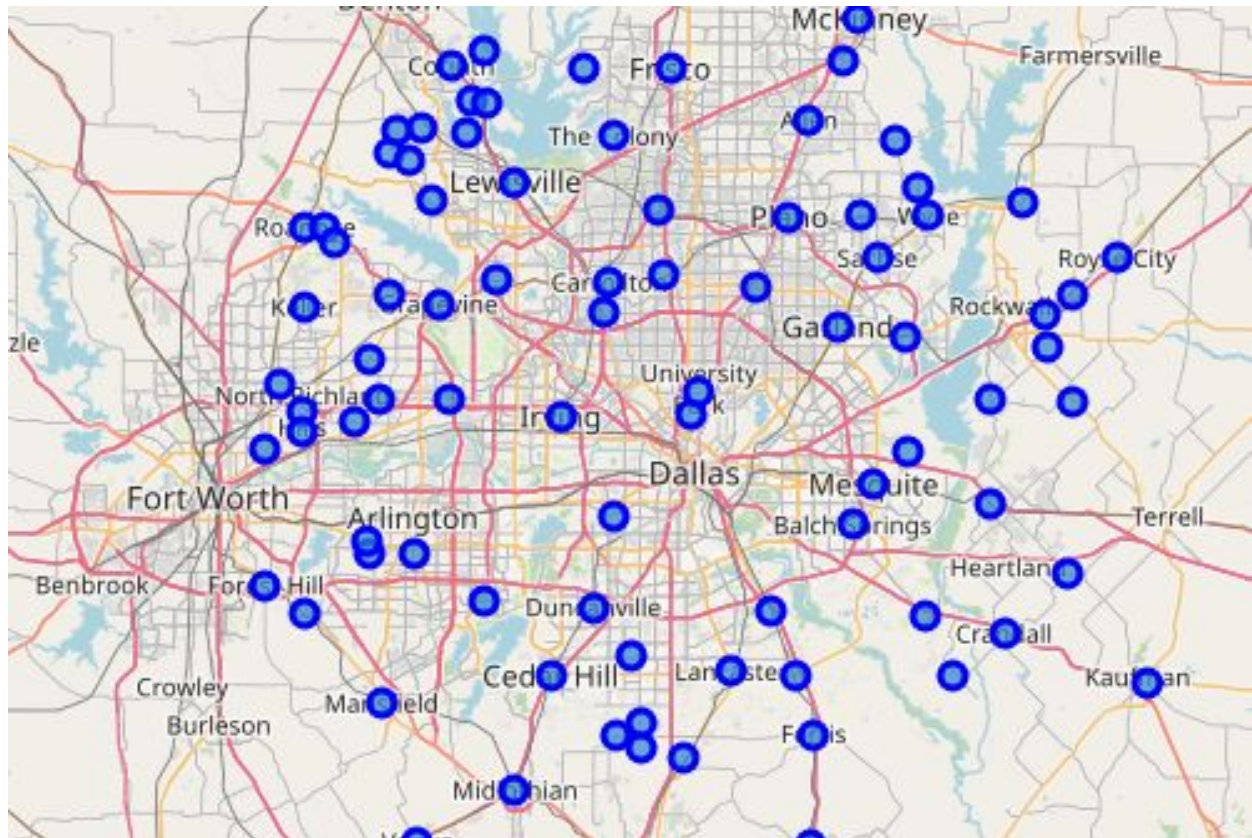
Before being able to determine the most common venue categories of each city, the geolocations (latitude and longitude) were obtained using geopy and then the list of venues within 6 KM were collected using the Foursquare API (providing the geolocation of each city).

Finally, before clustering the cities, the list of venues was pivoted and grouped in order to count venues by category and by city to then determine the ranking of most common venues by category and by city.

As support for the analysis, folium maps identifying the target cities and displaying the different clusters by colors was applied.

### **4. Results**

92 cities in the DFW Metroplex area are within 35 miles from Downtown Dallas and have less than 400,000 residents. They are shown in the map below.



Picture 1 - Cities in DFW area within 35 miles of Downtown Dallas and with less than 400K habitants

Applying KMeans clustering model, considering the 7 most common venue categories by city and, setting K as 7, the different clusters are calculated and shown in the map below.



15	Dalworthington Gardens	32.702910	-97.155289	0.0	Shopping	Food	Outdoor	Hotels	Health	Entertainment	Education
16	DeSoto	32.606287	-96.865622	0.0	Shopping	Food	Outdoor	Hotels	Health	Entertainment	Education
17	Double Oak	33.065122	-97.110567	0.0	Shopping	Food	Entertainment	Outdoor	Hotels	Health	Education
18	Duncanville	32.651800	-96.908337	0.0	Shopping	Food	Outdoor	Hotels	Health	Entertainment	Education
20	Eules	32.845787	-97.066714	0.0	Shopping	Food	Hotels	Outdoor	Entertainment	Health	Education
21	Fairview	33.157832	-96.631830	0.0	Shopping	Food	Outdoor	Entertainment	Hotels	Health	Education
24	Ferris	32.534026	-96.665549	0.0	Shopping	Food	Outdoor	Hotels	Health	Entertainment	Education
25	Flower Mound	33.028385	-97.086720	0.0	Shopping	Food	Outdoor	Entertainment	Hotels	Health	Education
32	Grand Prairie	32.657368	-97.028466	0.0	Shopping	Food	Outdoor	Hotels	Health	Entertainment	Education
37	Hebron	33.021176	-96.837343	0.0	Shopping	Food	Outdoor	Entertainment	Hotels	Health	Education
38	Hickory Creek	33.122343	-97.043065	0.0	Shopping	Food	Outdoor	Hotels	Health	Entertainment	Education
40	Highland Village	33.091788	-97.046676	0.0	Shopping	Food	Outdoor	Entertainment	Hotels	Health	Education
41	Hurst	32.823462	-97.170568	0.0	Shopping	Food	Outdoor	Entertainment	Hotels	Health	Education
44	Kaufman	32.581021	-96.299271	0.0	Shopping	Food	Outdoor	Hotels	Health	Entertainment	Education
45	Keller	32.929966	-97.227125	0.0	Shopping	Food	Outdoor	Education	Hotels	Health	Entertainment
46	Kennedale	32.646801	-97.225847	0.0	Shopping	Food	Outdoor	Hotels	Health	Entertainment	Education
49	Lantana	33.094028	-97.124497	0.0	Shopping	Food	Entertainment	Outdoor	Hotels	Health	Education
50	Lavon	33.027619	-96.434152	0.0	Shopping	Food	Entertainment	Outdoor	Hotels	Health	Education
51	Lewisville	33.046233	-96.994174	0.0	Shopping	Food	Outdoor	Entertainment	Hotels	Health	Education
52	Lucas	33.084285	-96.576658	0.0	Shopping	Food	Outdoor	Education	Hotels	Health	Entertainment
54	McKinney	33.197650	-96.615447	0.0	Shopping	Food	Outdoor	Entertainment	Hotels	Health	Education
57	Midlothian	32.482361	-96.994449	0.0	Shopping	Food	Hotels	Entertainment	Outdoor	Health	Education



58	Mobile City	32.9228 23	-96.4108 49	0.0	Shoppi ng	Food	Outdoor	Hotels	Health	Entertainm ent	Education
60	North Richland Hills	32.8342 95	-97.2289 03	0.0	Shoppi ng	Food	Outdoor	Entertainm ent	Hotels	Health	Education
61	Oak Leaf	32.5204 15	-96.8547 22	0.0	Shoppi ng	Food	Outdoor	Hotels	Health	Entertainm ent	Education
62	Ovilla	32.5334 15	-96.8828 75	0.0	Shoppi ng	Food	Outdoor	Hotels	Health	Entertainm ent	Education
65	Parker	32.7599 48	-97.7935 77	0.0	Shoppi ng	Food	Entertainm ent	Hotels	Outdoor	Health	Education
69	Richland Hills	32.8159 62	-97.2280 70	0.0	Shoppi ng	Food	Outdoor	Entertainm ent	Hotels	Health	Education
70	Roanoke	33.0040 13	-97.2258 48	0.0	Shoppi ng	Food	Outdoor	Hotels	Entertainm ent	Education	Health
72	Rowlett	32.9029 02	-96.5638 80	0.0	Shoppi ng	Food	Outdoor	Entertainm ent	Hotels	Health	Education
74	Sachse	32.9762 33	-96.5952 70	0.0	Shoppi ng	Food	Outdoor	Entertainm ent	Hotels	Health	Education
76	Seagoville	32.6447 92	-96.5418 61	0.0	Shoppi ng	Food	Outdoor	Hotels	Health	Entertainm ent	Education
78	Southlake	32.9412 36	-97.1341 78	0.0	Shoppi ng	Food	Outdoor	Hotels	Entertainm ent	Health	Education
79	St. Paul	33.0412 31	-96.5502 68	0.0	Shoppi ng	Food	Entertainm ent	Outdoor	Hotels	Health	Education
80	Sunnyvale	32.7965 17	-96.5608 24	0.0	Shoppi ng	Food	Outdoor	Entertainm ent	Hotels	Health	Education
82	Terrell	30.1990 74	-102.097 488	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
87	Watauga	32.8579 06	-97.2547 37	0.0	Shoppi ng	Food	Outdoor	Hotels	Health	Entertainm ent	Education
88	Waxahachi e	32.3944 91	-96.8439 36	0.0	Shoppi ng	Food	Outdoor	Entertainm ent	Education	Hotels	Health
91	Wylie	33.0151 20	-96.5388 79	0.0	Shoppi ng	Food	Entertainm ent	Outdoor	Hotels	Health	Education

## Cluster 1

	city	latitud e	longitu de	Clust er Label s	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
3	Balch Springs	32.728 741	-96.622 771	1.0	Shopping	Food	Outdoor	Entertainm ent	Hotels	Health	Education
7	Cedar Hill	32.588 807	-96.955 367	1.0	Shopping	Food	Outdoor	Entertainm ent	Hotels	Health	Education

14	Crandall	32.627911	-96.455820	1.0	Shopping	Food	Outdoor	Hotels	Health	Entertainment	Education
26	Forest Hill	32.672078	-97.269181	1.0	Shopping	Food	Outdoor	Hotels	Entertainment	Health	Education
48	Lancaster	32.592080	-96.756108	1.0	Shopping	Food	Outdoor	Hotels	Health	Entertainment	Education
55	McLendon Chisholm	32.842348	-96.380539	1.0	Shopping	Outdoor	Food	Hotels	Health	Entertainment	Education
56	Mesquite	32.766610	-96.599472	1.0	Shopping	Food	Outdoor	Entertainment	Hotels	Health	Education
63	Palmer	32.431252	-96.667770	1.0	Shopping	Food	Entertainment	Outdoor	Hotels	Health	Education
71	Rockwall	32.892346	-96.406699	1.0	Shopping	Food	Outdoor	Hotels	Health	Entertainment	Education

## Cluster 2

	city	latitude	longitude	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
10	Combine	32.588303	-96.511797	2.0	Food	Entertainment	Shopping	Outdoor	Hotels	Health	Education

## Cluster 3

	city	latitude	longitude	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
6	Carrollton	32.953735	-96.890282	3.0	Food	Shopping	Outdoor	Entertainment	Hotels	Health	Education
22	Farmers Branch	32.926514	-96.896115	3.0	Food	Shopping	Outdoor	Entertainment	Hotels	Health	Education
43	Irving	32.829518	-96.944218	3.0	Food	Shopping	Outdoor	Hotels	Health	Entertainment	Education
47	Lake Dallas	33.119287	-97.025564	3.0	Shopping	Food	Outdoor	Hotels	Health	Entertainment	Education
81	Talty	32.683187	-96.385539	3.0	Food	Shopping	Outdoor	Hotels	Health	Entertainment	Education
84	Trophy Club	33.004677	-97.205599	3.0	Food	Shopping	Hotels	Outdoor	Entertainment	Education	Health
86	Venus	32.433474	-97.102508	3.0	Food	Shopping	Outdoor	Hotels	Health	Entertainment	Education

## Cluster 4

	city	latitude	longitude	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
2	Arlington	32.701939	-97.105624	4.0	Food	Shopping	Outdoor	Hotels	Health	Entertainment	Education
5	Bedford	32.844017	-97.143067	4.0	Food	Shopping	Outdoor	Entertainment	Hotels	Health	Education
9	Colleyville	32.880960	-97.155012	4.0	Food	Shopping	Outdoor	Entertainment	Hotels	Health	Education
19	Ennis	32.329311	-96.625268	4.0	Food	Shopping	Outdoor	Hotels	Health	Entertainment	Education
23	Fate	32.941511	-96.381372	4.0	Food	Shopping	Education	Outdoor	Hotels	Health	Entertainment
27	Forney	32.747893	-96.471929	4.0	Food	Shopping	Entertainment	Outdoor	Hotels	Health	Education
28	Frisco	33.150674	-96.823612	4.0	Food	Shopping	Entertainment	Outdoor	Education	Hotels	Health
29	Garland	32.912624	-96.638883	4.0	Food	Shopping	Outdoor	Entertainment	Hotels	Health	Education
30	Garrett	32.363476	-96.654713	4.0	Food	Shopping	Outdoor	Hotels	Entertainment	Health	Education
31	Glenn Heights	32.543873	-96.855183	4.0	Food	Shopping	Outdoor	Hotels	Health	Entertainment	Education
33	Grapevine	32.933738	-97.078875	4.0	Food	Shopping	Outdoor	Hotels	Health	Entertainment	Education
34	Hackberry	33.152342	-96.917782	4.0	Food	Shopping	Outdoor	Hotels	Health	Entertainment	Education
35	Haltom City	32.799574	-97.269182	4.0	Food	Shopping	Outdoor	Hotels	Health	Entertainment	Education
36	Heath	32.844530	-96.471846	4.0	Food	Shopping	Hotels	Outdoor	Entertainment	Health	Education
39	Highland Park	32.831563	-96.800773	4.0	Food	Shopping	Outdoor	Hotels	Entertainment	Health	Education
42	Hutchins	32.649300	-96.713052	4.0	Food	Shopping	Hotels	Entertainment	Outdoor	Health	Education
53	Mansfield	32.563192	-97.141677	4.0	Food	Shopping	Outdoor	Entertainment	Hotels	Health	Education
59	Murphy	33.015121	-96.613048	4.0	Food	Shopping	Hotels	Entertainment	Outdoor	Education	Health
64	Pantego	32.714299	-97.156400	4.0	Food	Shopping	Outdoor	Hotels	Health	Entertainment	Education



66	Plano	33.013676	-96.692510	4.0	Food	Shopping	Outdoor	Entertainment	Hotels	Health	Education
67	Red Oak	32.511710	-96.808363	4.0	Food	Shopping	Outdoor	Hotels	Health	Entertainment	Education
73	Royse City	32.976104	-96.331334	4.0	Food	Shopping	Entertainment	Outdoor	Hotels	Health	Education
77	Shady Shores	33.167856	-97.029868	4.0	Food	Shopping	Hotels	Entertainment	Outdoor	Health	Education
83	The Colony	33.089009	-96.886392	4.0	Food	Shopping	Outdoor	Entertainment	Hotels	Education	Health
85	University Park	32.851034	-96.792580	4.0	Food	Shopping	Outdoor	Entertainment	Hotels	Health	Education
89	Westlake	32.991226	-97.194370	4.0	Food	Shopping	Outdoor	Hotels	Entertainment	Education	Health
90	Wilmer	32.589024	-96.685272	4.0	Food	Shopping	Outdoor	Hotels	Health	Entertainment	Education

### Cluster 5

	city	latitude	longitude	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
75	Scurry	32.722517	-100.905202	5.0	Shopping	Hotels	Food	Entertainment	Outdoor	Health	Education

### Cluster 6

	city	latitude	longitude	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
0	Addison	32.960431	-96.830260	6.0	Food	Shopping	Outdoor	Entertainment	Hotels	Health	Education
68	Richardson	32.948179	-96.729721	6.0	Food	Shopping	Hotels	Outdoor	Entertainment	Health	Education

## 5. Discussion

The criteria adopted to cluster the suburb cities was to rank the most common venues on those cities based on the following 7 categories: Shopping, Food, Outdoor, Entertainment, Health, Hotels and Education.

Venues of other categories were not considered.

Other aspects like distance from Downtown Dallas and population were not considered for clustering as well. The first one because the resulting commute time doesn't vary in a meaningful way to turn it into a variable of interest; the second because a higher population in a city in DFW area doesn't mean more concentration, usually these cities spread around equally over the land.

With the KMeans clustering with K=7, a similar result than when applying K=4 was obtained because the frequency of the categories selected for clustering are very similar across the suburb cities.

Food and Shopping services are almost always the most frequent, with outdoor activities usually in third place and schools (education) predominantly in the last place.

If we try to force a clear difference among the clusters, the most notable one is that in 3 clusters the shopping activity prevails, while in 4 clusters the restaurants (food) are more frequent.

Tagging a name to the clusters to classify them in terms of services more offered ends up to:

- Cluster 0 : **Predominance of Shopping services, with Theaters more numerous than Hotels** - 45 cities
- Cluster 1 : **Predominance of Shopping services, with Hotels more numerous than Theaters** - 9 cities
- Cluster 2 : **Restaurants first, Theaters second** - 1 city
- Cluster 3 : **Restaurants first, Shopping second** - 7 cities
- Cluster 4 : **Restaurants first, Shopping second - similar to Cluster 3** - 27 cities
- Cluster 5 : **Higher proportion of Hotels** - 1 city
- Cluster 6 : **Restaurants first, Shopping second - similar to Cluster 3** - 2 cities

## 6. Conclusion

When it comes to services offered in the suburb cities in the DFW Metroplex area, the frequency by service category is quite similar across them all, with one only notable division: the group where restaurants are the most common venue and the group where shopping venues are more frequent; these 2 categories are the most frequent for the vast majority of cities, though.

Therefore, differentiating the cities as per the frequency of categories of services offered in each one is not very helpful, as they are very similar from that perspective. Quantitative economic attributes like house prices and household income or, qualitative ones like school rates likely produce a better differentiation than venue categories.

