Segmenting and Clustering Neighborhoods in Dallas

IBA Applied Data Science Capstone Project

Alex P. Blizzard

May 21, 2020

Abstract

Food movements are rising in popularity around the world in younger generations, and the newest southern metroplex to adopt this “foodie” philosophy is Dallas, Texas. With downtown Dallas being so large containing 14 districts with an already massive restaurant population, where would be the ideal location to open a new restaurant? The interpretations of the data are important to a restaurant owner as the restaurant business is one of the riskiest investments, but with data science: one can minimize that risk. The data that will be analyzed in this study is a csv file containing the neighborhoods in downtown Dallas, along with their corresponding coordinates. Venues data will be collected through an API with foursquare, a technology company with a location platform. This study was divided into three segments: data preprocessing (importing data and venues), data processing (assigning a statistical variable to venue category, sorting by frequency), and data modeling (machine learning clustering).

The clustered neighborhoods are primarily associated with their most common venue: cluster 0 has parks, cluster 1 has bars, cluster 2 have stores, cluster 3 has galleries, and cluster 4 has hotels. The practice of data science with data analytics and machine learning clustering of downtown Dallas neighborhood district data with Foursquare’s location platform is an efficient approach to classify best locations to open a restaurant as well as the kind of restaurant to be placed at each location. For a trendy restaurant targeting young professionals, neighborhoods associated with bars would be an ideal location. For a fine dining restaurant targeting upper class travelers, neighborhoods associated with hotels would be an ideal location.

1. Introduction

1.1 Background

Food movements are rising in popularity around the world in younger generations, and the newest southern metroplex to adopt this "foodie" philosophy is Dallas, Texas. This prime destination is known for its wide variety of cuisines from famous Texas barbecue to upscale-hipster ramen bars. For a business owner looking to open a restaurant, downtown Dallas is the perfect location to reach the young "foodie" professionals.

1.2 Problem and Interest

With downtown Dallas being so large containing 14 districts with an already massive restaurant population, where would be the ideal location to open a new restaurant? What kind of restaurant should be open for that district? This study utilizes a variety of data analytics to answer these questions. The interpretations of the data are important to a restaurant owner as the restaurant business is one of the riskiest investments, but with data science: one can minimize that risk.

2. Data

The data that will be analyzed in this study is a excel spreadsheet (csv file) containing the 14 neighborhood districts in downtown Dallas, along with their corresponding coordinates (fig. 1). The coordinates were collected using google maps as some districts were not identified on foursquare. Venues data will be collected through an API with foursquare, a technology company with a location platform. With data science, the venue's category can be assigned a statistical variable and manipulated to identify the most frequent venue in that neighborhood. Using machine learning, the neighborhoods can be clustered together that contain the most similarity in venue frequency.

Neighborhood	Latitude	Longitude
Baylor District	32.797826	-96.778186
The Cedars	32.769897	-96.785852
Civic Center District	32.776255	-96.798500
Dallas Arts District	32.788885	-96.798924
Dallas Farmers Market	32.778543	-96.788246
Deep Ellum	32.784001	-96.778559
Design District	32.789055	-96.821590
Main Street District	32.780931	-96.798114
Reunion District	32.773613	-96.807676
Riverfront District	32.768468	-96.812929
South Side	32.765736	-96.794980
Thanksgiving Commercial Center	32.784866	-96.796783
Uptown	32.802121	-96.800784
Victory Park	32.788206	-96.810155
West End Historic District	32.780877	-96.807525

Figure 1. Dallas neighborhood districts data set with coordinates.

3. Methodology

3.1 Data Preprocessing

This study was divided into three segments: data preprocessing, data processing, and data modeling. Data preprocessing was accomplished with importing necessary libraries, importing Dallas neighborhoods data set, and getting venue data from Foursquare API. The Dallas data set already included coordinates data so therefore did not need to explore neighborhoods on Foursquare's locational platform. A map of downtown Dallas was generated using the folium() function with markers on the neighborhood districts (fig. 2). Venues were collected through the API limiting 100 venues within a 500-meter (1,640 feet) radius. Venues data was explored by the size of data set with the shape() function and the amount of venues per neighborhood (fig. 3).

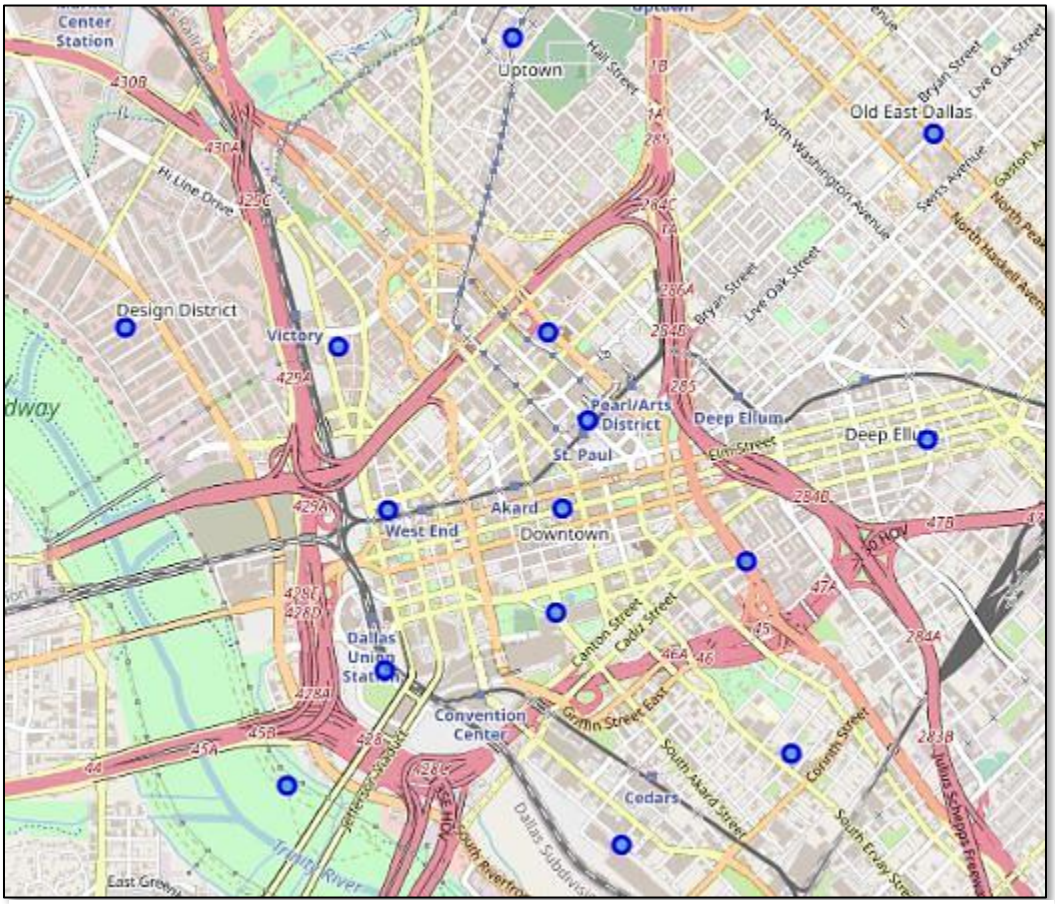

Figure 2. Map of downtown Dallas with markers on neighborhoods.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Baylor District	27	27	27	27	27	27
Civic Center District	28	28	28	28	28	28
Dallas Arts District	37	37	37	37	37	37
Dallas Farmers Market	21	21	21	21	21	21
Deep Ellum	50	50	50	50	50	50
Design District	23	23	23	23	23	23
Main Street District	100	100	100	100	100	100
Reunion District	22	22	22	22	22	22
Riverfront District	6	6	6	6	6	6
South Side	25	25	25	25	25	25
Thanksgiving Commercial Center	49	49	49	49	49	49
The Cedars	12	12	12	12	12	12
Uptown	40	40	40	40	40	40
Victory Park	31	31	31	31	31	31
West End Historic District	57	57	57	57	57	57

Figure 3. Dallas neighborhoods with number of venues from Foursquare.

3.2 Data Processing

Data processing was completed by analyzing each neighborhood by assigning a statistical variable (one-hot coding) using the pd.get_dummies() function on venue category. With the new dataframe, neighborhoods were grouped and calculated average frequency of occurrence of each venue category with the mean() function. Neighborhoods can then be analyzed with listing the top 10 most common venues (fig. 4).

	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
0	Baylor District	Pizza Place	Discount Store	Sandwich Place	Mexican Restaurant	Taco Place	Grocery Store	Mobile Phone Shop	Donut Shop	Nail Salon	Fast Food Restaurant
1	Civic Center District	Hotel	Coffee Shop	Bar	Café	Plaza	IT Services	Seafood Restaurant	Salad Place	Cocktail Bar	Department Store
2	Dallas Arts District	Food Truck	Performing Arts Venue	Steakhouse	American Restaurant	Japanese Restaurant	Art Museum	Theater	Sandwich Place	Seafood Restaurant	Dog Run
3	Dallas Farmers Market	American Restaurant	Farmers Market	Food Truck	Vietnamese Restaurant	Pool	Dessert Shop	Sandwich Place	Dog Run	Coffee Shop	Design Studio
4	Deep Ellum	Bar	BBQ Joint	Music Venue	American Restaurant	Mexican Restaurant	Dessert Shop	Dive Bar	Coffee Shop	Art Gallery	Pizza Place

Figure 4. Dallas neighborhoods with 10 most common venues.

3.3 Data Modeling

Data modeling was achieved by utilizing unsupervised machine learning of k means with the Euclidean distance algorithm between each neighborhood with venue category mean frequency data. This process will cluster neighborhoods together that have similarity between the frequencies. The number of clusters that was focused in this study was five, providing enough dissimilarity between the data control. Neighborhoods that were clustered together were assigned a color on the map of downtown Dallas.

4. Results

The result of the machine learning clustering is listed below: cluster 0 (fig. 5), cluster 1 (fig. 6), cluster 2 (fig. 7), cluster 3 (fig. 8), and cluster 4 (fig. 9). The color coordinated clusters are displaced in the map of downtown Dallas (fig. 10). Cluster 0 (red) consist of The Cedars neighborhood with most common venue of a park. Cluster 1 (purple) consist of Dallas Arts District, Dallas Farmers Market, Deep Ellum, South Side, and Uptown with the most common venues with mostly bars. Cluster 2 (blue) consist of Riverfront District with the most common venue of a store. Cluster 3 (green) consist of the Design District with the most common venue of a gallery. Lastly, cluster 4 (orange) consist of the Baylor District, Civic Center District, Main Street District, Reunion District, Thanksgiving Commercial Center, Victory Park, and West End Historic District with the most common venue with mostly hotels.

	Neighborhood	Latitude	Longitude	Cluster Labels2	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	The Cedars	32.769897	-96.785852	0	Skate Park	Food Court	Museum	History Museum	Art Gallery	Coffee Shop	Pharmacy	Dive Bar	BBQ Joint	Dog Run

Figure 5. Cluster 0 neighborhood with most common venues.

	Neighborhood	Latitude	Longitude	Cluster Labels2	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
3	Dallas Arts District	32.788885	-96.798924	1	Food Truck	Performing Arts Venue	Steakhouse	American Restaurant	Japanese Restaurant	Art Museum	Theater	Sandwich Place	Seafood Restaurant	Dog Run
4	Dallas Farmers Market	32.778543	-96.788246	1	American Restaurant	Farmers Market	Food Truck	Vietnamese Restaurant	Pool	Dessert Shop	Sandwich Place	Dog Run	Coffee Shop	Design Studio
5	Deep Ellum	32.784001	-96.778559	1	Bar	BBQ Joint	Music Venue	American Restaurant	Mexican Restaurant	Dessert Shop	Dive Bar	Coffee Shop	Art Gallery	Pizza Place
10	South Side	32.765736	-96.794980	1	Music Venue	Bar	American Restaurant	Coffee Shop	Rock Club	Business Service	Record Shop	Pub	Cocktail Bar	Convenience Store
12	Uptown	32.802121	-96.800784	1	Cocktail Bar	Bar	Seafood Restaurant	Pizza Place	Café	Gym	Italian Restaurant	Burger Joint	Breakfast Spot	Salon / Barbershop

Figure 6. Cluster 1 neighborhood with most common venues.

	Neighborhood	Latitude	Longitude	Cluster Labels2	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
9	Riverfront District	32.768468	-96.812929	2	Liquor Store	BBQ Joint	Nightclub	Gay Bar	Business Service	Design Studio	Dive Bar	Discount Store	Diner	Dessert Shop

Figure 7. Cluster 2 neighborhood with most common venues.

	Neighborhood	Latitude	Longitude	Cluster Labels2	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
6	Design District	32.789055	-96.82159	3	Art Gallery	Furniture / Home Store	Antique Shop	Mexican Restaurant	Convenience Store	Shipping Store	Brewery	Pizza Place	Steakhouse	Bridal Shop

Figure 8. Cluster 3 neighborhood with most common venues.

	Neighborhood	Latitude	Longitude	Cluster Labels2	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
0	Baylor District	32.797826	-96.778186	4	Pizza Place	Discount Store	Sandwich Place	Mexican Restaurant	Taco Place	Grocery Store	Mobile Phone Shop	Donut Shop	Nail Salon	Fast Food Restaurant
2	Civic Center District	32.776255	-96.798500	4	Hotel	Coffee Shop	Bar	Café	Plaza	IT Services	Seafood Restaurant	Salad Place	Cocktail Bar	Department Store
7	Main Street District	32.780931	-96.798114	4	Hotel	Coffee Shop	Mexican Restaurant	Sandwich Place	Bar	Cocktail Bar	Salad Place	Park	Café	Sports Bar
8	Reunion District	32.773613	-96.807676	4	Hotel	American Restaurant	Scenic Lookout	Plaza	Bar	Restaurant	Event Space	Coffee Shop	Beer Garden	Food Truck
11	Thanksgiving Commercial Center	32.784866	-96.796783	4	Gym	Coffee Shop	Cocktail Bar	Italian Restaurant	Hotel	Steakhouse	Taco Place	New American Restaurant	Boutique	Sandwich Place
13	Victory Park	32.788206	-96.810155	4	Restaurant	Lounge	Sports Bar	Bar	Yoga Studio	Cocktail Bar	Ramen Restaurant	Furniture / Home Store	Pizza Place	Pharmacy
14	West End Historic District	32.780877	-96.807525	4	History Museum	Sandwich Place	Plaza	Hotel	Music Venue	Liquor Store	Pharmacy	Coffee Shop	Convenience Store	Lounge

Figure 9. Cluster 4 neighborhood with most common venues.

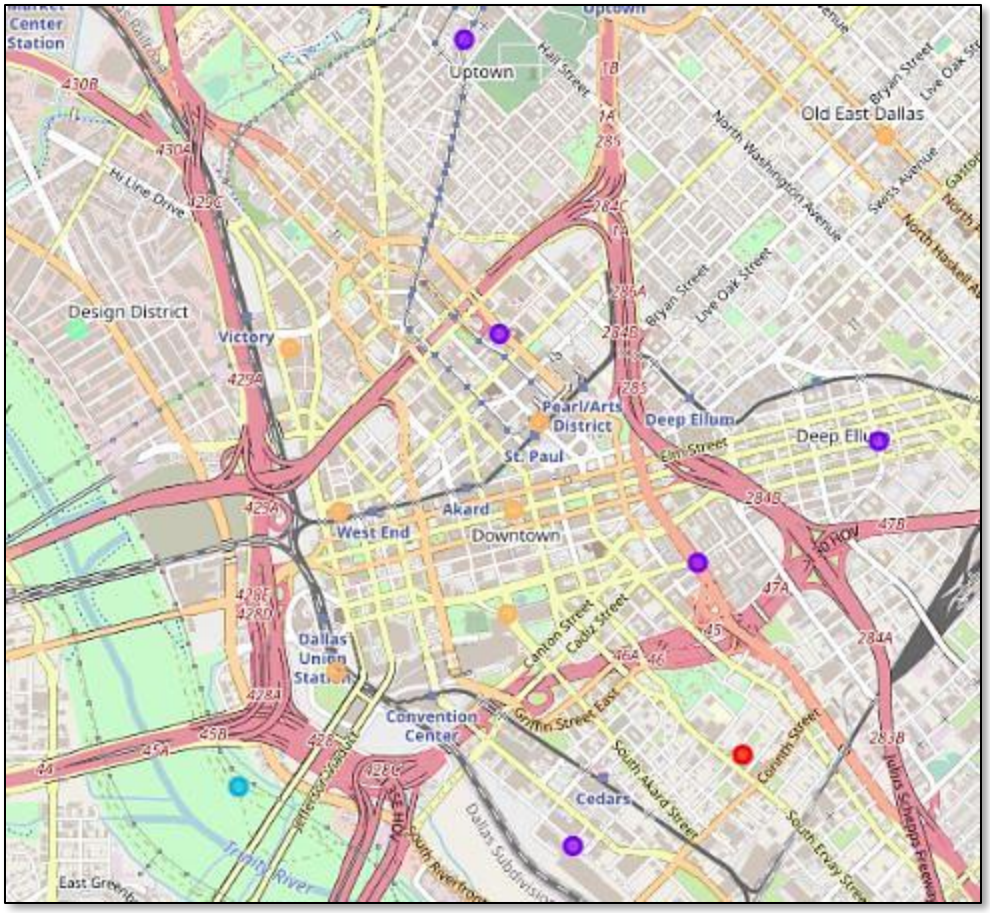

Figure 10. Map of downtown Dallas with clustered neighborhood markers.

5. Discussion

The clustered neighborhoods are primarily associated with their most common venue: cluster 0 has parks, cluster 1 has bars, cluster 2 have stores, cluster 3 has galleries, and cluster 4 has hotels. In order to decide the best location to open a restaurant, other variables must be kept in mind. A large population of young professional's travel to downtown Dallas looking for a fun evening which might include dinner and drinks at nearby bars. If your restaurant is targeting young professionals in a trendy environment, then cluster 1 (Dallas Arts District, Dallas Farmers Market, Deep Ellum, South Side, and Uptown) would be an ideal choice. Another large population in downtown Dallas is the upper-class tourist staying in hotels. These groups of individuals are usually travelers that lack personal transportation and seek out restaurants within walking distance of hotels. If your restaurant is targeting upper-class professionals in a fine dining environment, then cluster 4 (Baylor District, Civic Center District, Main Street District, Reunion District, Thanksgiving Commercial Center, Victory Park, and West End Historic District) would be an ideal location.

6. Conclusion

The practice of data science with data analytics and machine learning clustering of downtown Dallas neighborhood district data with Foursquare's location platform is an efficient approach to classify best locations to open a restaurant as well as the kind of restaurant to be placed at each location. The following are concluding remarks that were interpreted using the data explored and manipulated in this study:

1) Dallas neighborhoods were characterized into 5 clusters each with a unique primarily most common venue.

2) The Cedars is primarily associated with parks.

3) Dallas Arts District, Dallas Farmers Market, Deep Ellum, South Side, and Uptown is primarily associated with bars.

4) Riverfront District is primarily associated with stores.

5) Design District is primarily associated with galleries.

6) Baylor District, Civic Center District, Main Street District, Reunion District, Thanksgiving Commercial Center, Victory Park, and West End Historic District are primarily associated with hotels.

7) For a trendy restaurant targeting young professionals, neighborhoods associated with bars would be an ideal location.

8) For a fine dining restaurant targeting upper class travelers, neighborhoods associated with hotels would be an ideal location.