Battle of the Neighborhoods – Dallas Texas Report

A. Introduction

A.1. Background

Dallas is the largest city and seat of Dallas county, with portions extending into Collin, Denton, Kaufman and Rockwall counties.

Although not the fastest growing city in Texas; Dallas, with a population of 1,343,573 Mil (2019) and numerous businesses that is based in and around Dallas is prime ground for restaurants.
Dallas is home to 10 Fortune 500 companies within the city limits. The Dallas–Fort Worth metroplex hosts an additional 23 Fortune 500 companies, including American Airlines (Fort Worth) and ExxonMobil (Irving).
The Dallas-Fort Worth-Arlington core-based statistical area by the U.S. Census Bureau -- had 13,763 restaurants in the fall of 2015. Dallas has 4 times more restaurants per person than New York City.
These business does not always have in-house cafeterias, and if they do, people tend not to always eat at the cafeteria.

A.2. Problem Description:

A client is looking to open a new restaurant in Dallas TX, but she is not sure about the best location for this new venue. The investor is new to the restaurant business and want to start small and is thinking of opening either a pizza restaurant, cafe or wings restaurant and do not want to start a restaurant in a Zip code that have these restaurant categories in the top five.

Dallas is not as expensive as New York or San Francisco, but mortgage or rent can be a substantial chunk of your cost. It is highly preferable to find either a location that is on the outskirts of the city or a type of restaurant that does not have a big footprint.

More importantly the client has children that are Elementary school and Middle school age and has and apartment in Irving, Texas. This means the children will be in a school in the Irving Independent School district. This means she prefers to be with in 5 to 10 kilometers of her home.

The question is what zip code will be the best to open a restaurant in Dallas where competition will not be too much, while within 5 to 10 kilometers from the client's home?

A.3. Objective

The aim of this report is to study and analyze the restaurants in the zip codes of Dallas Texas and group them into similar clusters and, to analyze those clusters to gather meaningful information.

That information can be used to find relevant zip codes for opening a small restaurant that is close to a specific location. In this specific study the location is the client house, but can be modified for any geographical location.

A.4. Target Audience

This information provided by this report would be useful for people who are interested in opening a restaurant in Dallas Texas.

Although this report will focus on café, pizza and wings restaurants, the venue category can be modified to any category that is in Foursquare.

B. Data Description:

To consider the objective stated above, we can list the below data sources used for the analysis.

1. **Dallas Zip code Data**: We need all the zip codes that fall inside the City of Dallas, with their longitude and Latitudes.
   Source: The data is available at Dallas Opendata: https://www.dallasopendata.com/Geography-Boundaries/FY-2017-City-of-Dallas-City-Limits/ad4m-4kje and can be downloaded in geospatial(KML, KMZ, GeoJSON, Shapefile) or non-geospatial formats(CSV, JSON, XML).

   In this case the file was downloaded in non-geospatial format CSV and was transformed into a pandas data frame for further analysis.

2. **Data for each restaurant category in Dallas, TX**: Data for each restaurant category (Pizza, wing and café) will be extracted via Foursquare API.

C. Methodology:

C.1. Import the all Texas zip codes CSV file and convert the data to a pandas data frame

	Zipcodes	City	State	Latitude	Longitude	Timezone	Daylight savings time flag	geopointlat	geopointlon
0	75475	Randolph	TX	33.485315	-96.25525	-6	1	33.485315	-96.25525
1	75757	Bullard	TX	32.136787	-95.36710	-6	1	32.136787	-95.36710
2	78650	McDade	TX	30.283941	-97.23563	-6	1	30.283941	-97.23563
3	75010	Carrollton	TX	33.030556	-96.89328	-6	1	33.030556	-96.89328
4	76054	Hurst	TX	32.858398	-97.17681	-6	1	32.858398	-97.17681

After Importing the data, unnecessary attributes/column needs to be removed. The cleaned-up data frame looks like below.

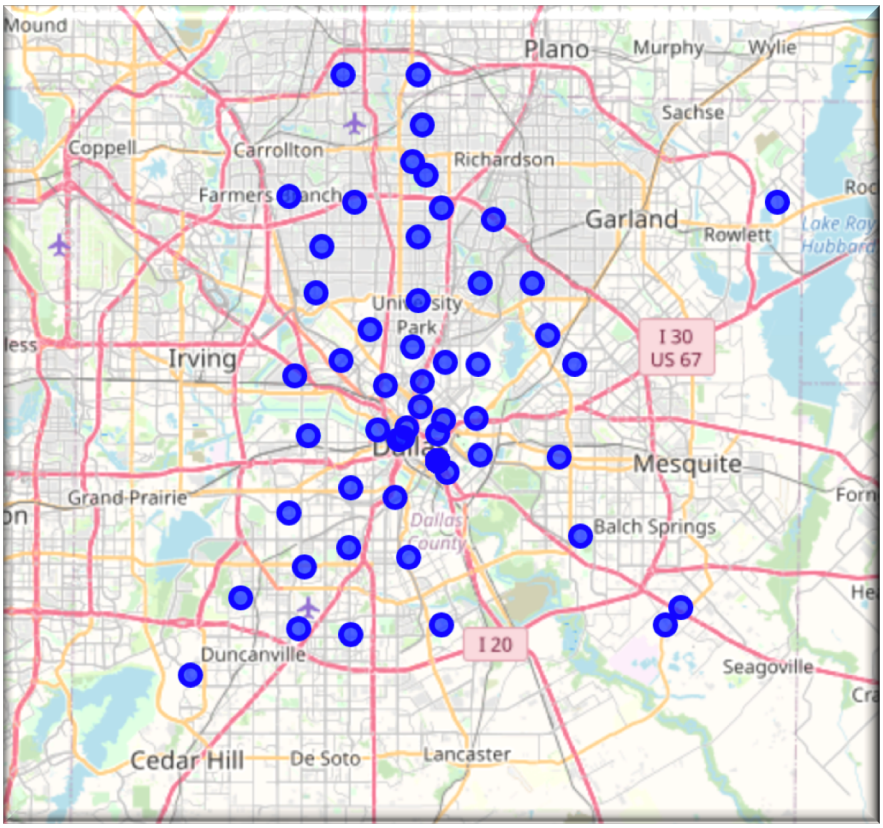
	Zipcodes	City	State	Latitude	Longitude
0	75475	Randolph	TX	33.485315	-96.25525
1	75757	Bullard	TX	32.136787	-95.36710
2	78650	McDade	TX	30.283941	-97.23563
3	75010	Carrollton	TX	33.030556	-96.89328
4	76054	Hurst	TX	32.858398	-97.17681

The dataframe still contains all Texas zip codes, thus we need to find only zip codes in Dallas, and remove any zip codes that is in cities that have Dallas in their name (example: Lake Dallas).

	Zipcodes	City	State	Latitude	Longitude
0	75294	Dallas	TX	32.767268	-96.777626
1	75255	Dallas	TX	32.669783	-96.614921
2	75374	Dallas	TX	32.767268	-96.777626
3	75252	Dallas	TX	32.998132	-96.790880
4	75275	Dallas	TX	32.767268	-96.777626

C.2. Generating a map of Dallas and plotting the zip codes data on it

We then use the python **folium** library to visualize geographic details of Dallas and its zip codes. I created a map of Dallas with the zip code centers superimposed on top using the latitude and longitude values to get the visual as below:



C.3. Utilizing Foursquare API to explore the zip codes

Next, we utilized the Foursquare API to explore all venues with category “food” in the zip codes and segment them. We set the LIMIT parameter to **100**, which would limit the number of food venues returned by the Foursquare API and the radius of 600 meters.

Below is the Dallas venues dataframe for all food venues in Dallas as retrieved from Foursquare.

	Zipcode	Zipcode Latitude	Zipcode Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	75255	32.669783	-96.614921	Sid's Food Mart	32.669854	-96.614021	Deli / Bodega
1	75202	32.779880	-96.805020	Y. O. Ranch Steakhouse	32.781296	-96.806402	Steakhouse
2	75202	32.779880	-96.805020	Tutta's	32.781305	-96.807423	Pizza Place
3	75202	32.779880	-96.805020	Record Grill	32.779976	-96.806781	American Restaurant
4	75202	32.779880	-96.805020	Latin Deli	32.778747	-96.805873	Latin American Restaurant

C.4. Analyze each zip code

We use One Hot Encoding, use the zip codes to group data, and find out the top five venue categories present in each zip code, and then use the unsupervised machine

learning algorithms, k-Means clustering (kclusters = 5), with to determine the top 5 venues per zip code.

	Zipcode	1th Most Common Venue	2th Most Common Venue	3th Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	75201	Food Truck	American Restaurant	Sandwich Place	New American Restaurant	BBQ Joint
1	75202	Sandwich Place	Asian Restaurant	Fried Chicken Joint	Fast Food Restaurant	American Restaurant
2	75204	Fast Food Restaurant	Wings Joint	Chinese Restaurant	Food Court	Food
3	75205	Food	Wings Joint	Chinese Restaurant	Food Court	Fast Food Restaurant
4	75206	Mexican Restaurant	French Restaurant	Restaurant	Donut Shop	Café

C.5. Adding the Cluster Labels to the Venue Data

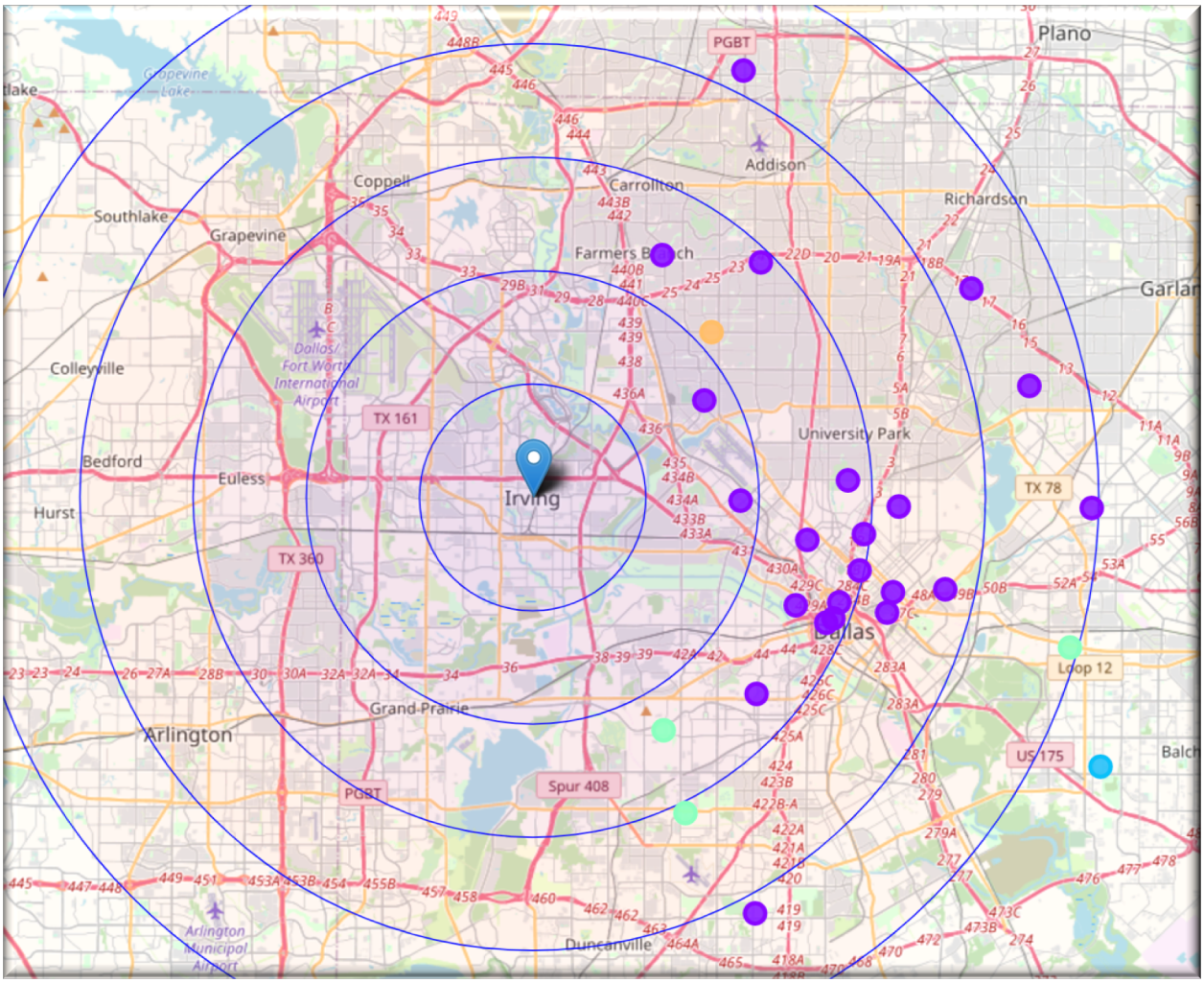
The below table depicts the clustered data along with the top 5 most common venues.

Zipcodes	City	State	Latitude	Longitude	Cluster Labels	1th Most Common Venue	2th Most Common Venue	3th Most Common Venue	4th Most Common Venue	5th Most Common Venue
75255	Dallas	TX	32.669783	-96.614921	4.0	Deli / Bodega	Wings Joint	Chinese Restaurant	Food Court	Food
75202	Dallas	TX	32.779880	-96.805020	0.0	Sandwich Place	Asian Restaurant	Fried Chicken Joint	Fast Food Restaurant	American Restaurant
75270	Dallas	TX	32.781330	-96.801980	0.0	Sandwich Place	Mexican Restaurant	Salad Place	Fast Food Restaurant	Fried Chicken Joint
75220	Dallas	TX	32.867977	-96.863060	0.0	Pizza Place	Food	Bakery	Mexican Restaurant	Chinese Restaurant
75234	Dallas	TX	32.925975	-96.883220	0.0	Breakfast Spot	Mexican Restaurant	American Restaurant	Fast Food Restaurant	Donut Shop

D. Results:

D.1. Visualizing the resulting Clusters

We use the matplotlib and folium packages to visualize the clusters on a map of Dallas, then add kilometer circles around the home location (Irving, Texas) of the customer to determine which zip codes fall inside an acceptable radius from her home.



E. Discussion:

The intent with which analysis was carried out was to find the best zip code in Dallas to open a restaurant.

One of the criteria was that the location must be, preferably with in 5 to 10 kilometers of Irving where the clients home is as she has children in Irving ISD.

As we analyze the results section, we can see that there are only two zip code that fall in the less that 10 kms range as well as inside the City of Dallas.

They are 75220 and 75235.

Investigating these two zip codes further to see if café's, pizza restaurant or wing restaurant are in the top 5 categories listed in the top 5 venues, we notice that (see figure below):

- Zip code 75220's most common venue is Pizza place, while not café and Wings restaurants are not in the top 5 venues.
- Zip code 75235 has none of the three preferred restaurants types in the top 5.

	Zipcodes	Cluster Labels	1th Most Common Venue	2th Most Common Venue	3th Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	75202	0	Sandwich Place	Asian Restaurant	Fried Chicken Joint	Fast Food Restaurant	American Restaurant
10	75270	0	Sandwich Place	Mexican Restaurant	Salad Place	Fast Food Restaurant	Fried Chicken Joint
11	75220	0	Pizza Place	Food	Bakery	Mexican Restaurant	Chinese Restaurant
12	75234	0	Breakfast Spot	Mexican Restaurant	American Restaurant	Fast Food Restaurant	Donut Shop
22	75246	0	American Restaurant	Food	New American Restaurant	Chinese Restaurant	Fast Food Restaurant
27	75238	0	Breakfast Spot	Wings Joint	Chinese Restaurant	Food Court	Food
31	75207	0	American Restaurant	Café	Restaurant	BBQ Joint	Bakery
34	75223	0	Mexican Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Taco Place	Fried Chicken Joint
35	75287	0	Sushi Restaurant	Fried Chicken Joint	Food	Wings Joint	Chinese Restaurant
38	75232	0	Pizza Place	Fried Chicken Joint	Donut Shop	Sandwich Place	Cajun / Creole Restaurant
46	75204	0	Fast Food Restaurant	Wings Joint	Chinese Restaurant	Food Court	Food
47	75226	0	Food Truck	Taco Place	BBQ Joint	Soup Place	Sandwich Place
50	75205	0	Food	Wings Joint	Chinese Restaurant	Food Court	Fast Food Restaurant
55	75221	0	Fast Food Restaurant	Noodle House	Diner	Mexican Restaurant	Restaurant
59	75219	0	Café	Pizza Place	Asian Restaurant	Taco Place	Sushi Restaurant
61	75228	0	Food	Sandwich Place	Mexican Restaurant	Wings Joint	Chinese Restaurant
72	75206	0	Mexican Restaurant	French Restaurant	Restaurant	Donut Shop	Café
73	75235	0	Fried Chicken Joint	Fast Food Restaurant	Mexican Restaurant	Comfort Food Restaurant	Restaurant
76	75201	0	Food Truck	American Restaurant	Sandwich Place	New American Restaurant	BBQ Joint
90	75208	0	Mexican Restaurant	Fast Food Restaurant	Italian Restaurant	Bakery	Donut Shop
104	75243	0	Fast Food Restaurant	Burger Joint	Mexican Restaurant	Wings Joint	Chinese Restaurant
112	75244	0	Sandwich Place	Pizza Place	Fast Food Restaurant	Donut Shop	Restaurant

F. Recommendation:

Assuming the client want to strictly open a restaurant in the City of Dallas with in 10 kilometers from her home, the best solution will be investigate further opening either a Wings restaurant or a café in Zip code 75235.

IF the above zip code is not satisfactory, it is suggested to rerun the analysis around other city, including Irving, TX, as this will minimize travel the most.

G. Conclusion:

Like seen in the example above, data was used to cluster neighborhoods in Dallas based on the most common venues in those zip codes and then narrowing down the zip codes based on distance.

The solution can also be used to solve other problems, and it is quite easy to change the Foursquare Category and cluster zip codes, or change the center of base (home location).

G. References:

1. *"Fortune 500"* Fortune. Retrieved June 29, 2020.
2. The data is available at Dallas *Opendata:*
3. Foursquare API