



**The xiaolongbao from Wu Wei Din Chinese Cuisine**

# **Explore The Restaurant Venues in DFW Area**

**James (Po-Chih) Chen**

Let's see what is the most popular restaurants in your place!

## **A. Introduction/Business Understanding**

### **A.1 Description of the problem**

The Dallas–Fort Worth metroplex, officially designated Dallas–Fort Worth–Arlington is a conurbation metropolitan statistical area in the U.S. state of Texas encompassing 11 counties. It is the economic and cultural hub of North Texas. Residents of the area also refer to it as DFW (airport code), or the Metroplex. The Dallas–Fort Worth–Arlington metropolitan statistical area's population was 7,573,136 according to the U.S. Census Bureau's 2019 population estimates.

DFW has great people and makes it the most populous metropolitan area in both Texas and the Southern United States, the fourth-largest in the U.S. In 2016, the Dallas–Fort Worth metroplex had the highest annual population growth in the United States. Since the DFW area is a metroplex with a high population and population density, there are a great number of shops, cafeterias, and restaurants in the metroplex where the population is dense. The restaurants will grow quickly as well. This research wants to explore what kind of restaurant is DFW area and

what numbers of restaurants in different districts. Then the investors can choose the potential area to start up their entrepreneur.

The investor can see the type of restaurants and numbers to decide where to build their restaurants and distinct from others.

## **A.2 Data Description**

To consider the problem we can list the data as below:

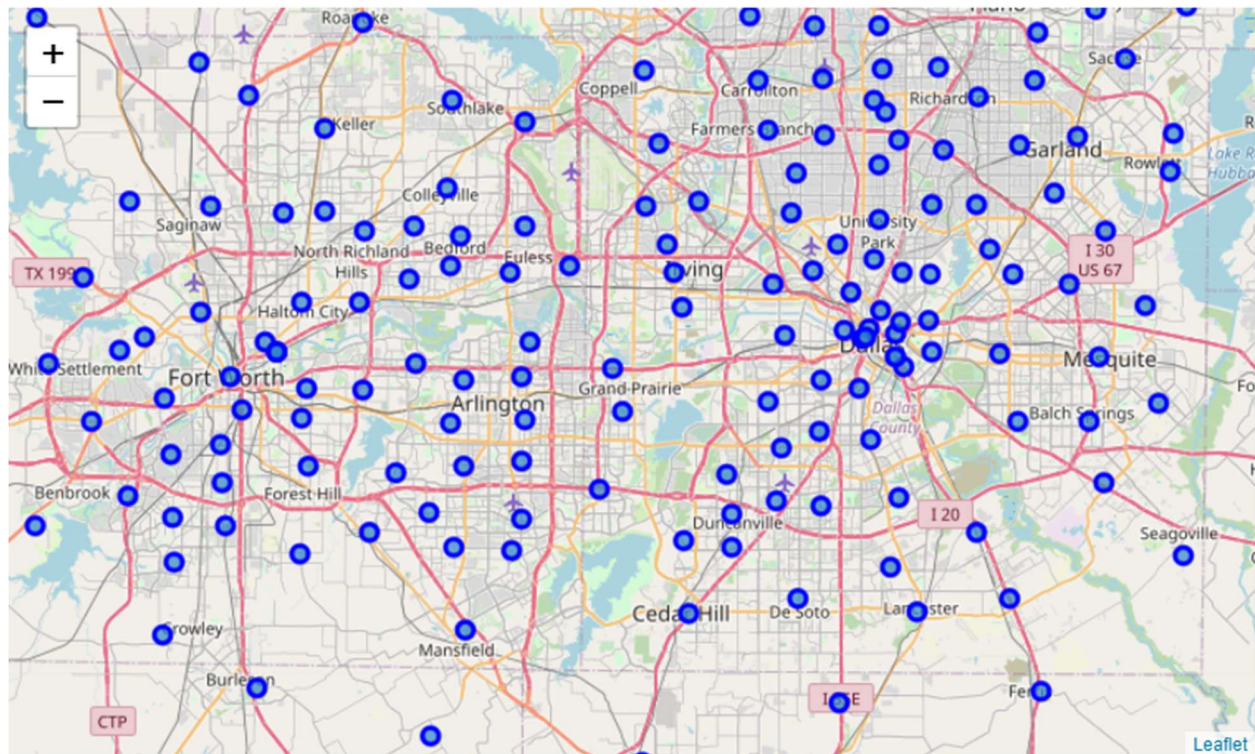
- The sources of zip, city, and population are from the public: Opendatasoft, GitHub, and demographics of Texas gov. The .json file has coordinates of all city of Texas.
- I used Foursquare API to get the most common venues of a given district of the DFW area.
- For convenience, I deleted the city showed in the public data but not showed in the Foursquare search.
- The resulting data is the final data I merged, and I will put it in the file.
- You can open data to find the Postal code, Latitude, Longitude, and Population.
  1. <http://zipatlas.com/us/tx/zip-code-comparison/population-density.htm>
  2. <https://demographics.texas.gov/Data/TPEPP/Estimates/>
  3. <https://github.com/OpenDataDE/State-zip-code-GeoJSON>
  4. <https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/>

## **B. Methodology**

As a database, I used the GitHub repository in my study. My master data has the main components Postal code, District, Population, Latitude, and Longitude information of the DFW area.

	postalcode	District	Latitude	Longitude	Population
0	75001	Addison	32.961290	-96.83751	8259
1	75002	Allen	33.092846	-96.62447	37216
2	75006	Carrollton	32.960374	-96.89163	46784
3	75007	Carrollton	33.005420	-96.89773	49933
4	75009	Celina	33.327927	-96.76129	4951
...	...	...	...	...	...
267	76487	Poolville	33.009411	-97.89968	1670
268	76490	Whitt	32.958020	-98.01942	60
269	76623	Avalon	32.221421	-96.78320	242
270	76651	Italy	32.175783	-96.88018	3242
271	76670	Milford	32.111683	-96.97377	1171

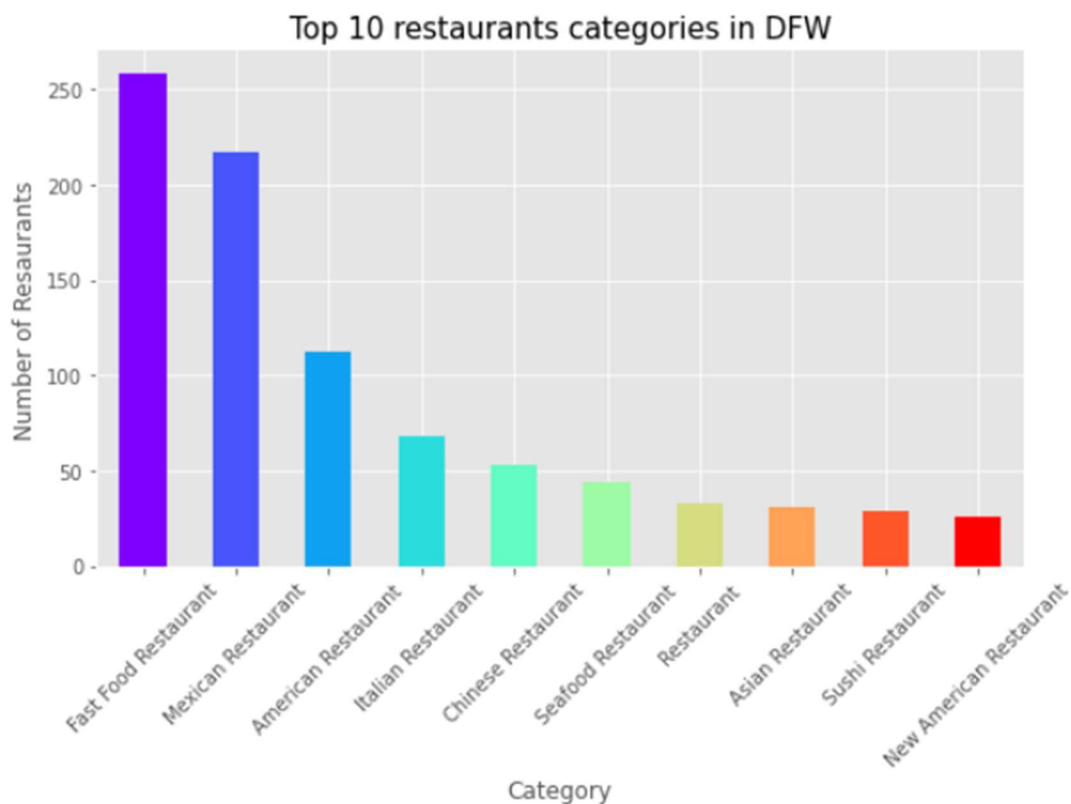
I used Python folium library to visualize geographic details of DFW and its city and created a map of DFW with district superimposed on top. I used latitude and longitude values to get the visual as below:



I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as 100 venues and the radius 1000 meters for each borough from their given latitude and longitude information. Here is ahead of the list Venues name, categories, latitude, and longitude information from Foursquare API. In summary of this data, it is over **5,000** restaurant venues were returned by Foursquare. Here is a merged table of districts and venues.

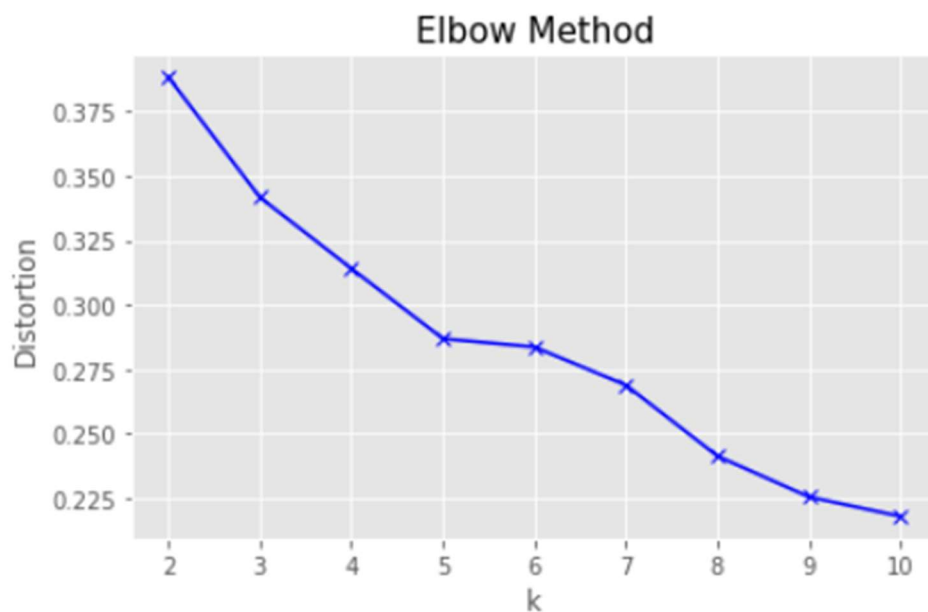
	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	postal code
0	Addison	32.96129	-96.83751	Cindi's N.Y. Delicatessen	32.961864	-96.838873	Diner	75001
1	Addison	32.96129	-96.83751	La Spiga	32.958278	-96.837417	Bakery	75001
2	Addison	32.96129	-96.83751	Elite Cigar Cafe	32.954967	-96.837875	Smoke Shop	75001
3	Addison	32.96129	-96.83751	Enterprise Rent-A-Car	32.960673	-96.838102	Rental Car Location	75001
4	Addison	32.96129	-96.83751	Best Thai Signature	32.954890	-96.841380	Thai Restaurant	75001
5	Addison	32.96129	-96.83751	Fogo de Chao Brazilian Steakhouse	32.953782	-96.837738	Churrascaria	75001
6	Addison	32.96129	-96.83751	Chuy's Tex-Mex	32.953575	-96.835815	Mexican Restaurant	75001

In summary of this graph **46** unique categories were returned by Foursquare, then I created a table that shows list of the top 10 venue categories for each borough in below bar chart.



We have some common venue categories in the district. For this reason, I used an unsupervised learning **K-means algorithm** to cluster the boroughs. K-Means algorithm is one of the most common cluster methods of unsupervised learning.

I will run K-Means to cluster the boroughs into **7** clusters because when I analyze the K-Means with the elbow method it ensured me the **7** degrees for optimum k of the K-Means.



I also used one-hot encoding to see the frequency of each district. Therefore I can know what the most popular restaurants in each district.

----Plano----

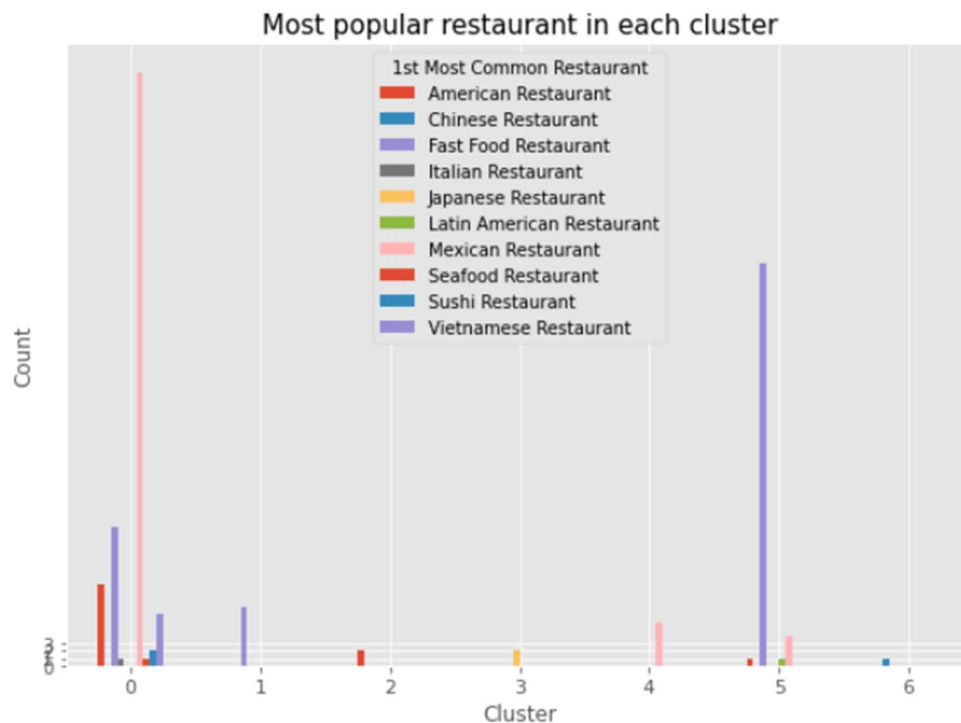
	venue	freq
0	Fast Food Restaurant	0.2
1	Thai Restaurant	0.1
2	Chinese Restaurant	0.1
3	Asian Restaurant	0.1
4	Japanese Restaurant	0.1



Here is my merged table with cluster labels for each district.

District	Latitude	Longitude	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant	6th Most Common Restaurant	7th Most Common Restaurant	8th Most Common Restaurant	9th Most Common Restaurant	10th Most Common Restaurant
Addison	32.961290	-96.83751	Mexican Restaurant	Seafood Restaurant	Italian Restaurant	Sushi Restaurant	Mediterranean Restaurant	Southern / Soul Food Restaurant	Spanish Restaurant	Asian Restaurant	Tex-Mex Restaurant	Thai Restaurant
Allen	33.092846	-96.62447	Fast Food Restaurant	Tex-Mex Restaurant	Italian Restaurant	Indian Restaurant	Pakistani Restaurant	Mexican Restaurant	Asian Restaurant	Thai Restaurant	Cajun / Creole Restaurant	Chinese Restaurant
Carrollton	32.960374	-96.89163	Mexican Restaurant	Sushi Restaurant	Chinese Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Latin American Restaurant	South American Restaurant	Asian Restaurant	Greek Restaurant	Russian Restaurant
Carrollton	33.005420	-96.89773	Mexican Restaurant	Sushi Restaurant	Chinese Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Latin American Restaurant	South American Restaurant	Asian Restaurant	Greek Restaurant	Russian Restaurant
Carrollton	33.030556	-96.89328	Mexican Restaurant	Sushi Restaurant	Chinese Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Latin American Restaurant	South American Restaurant	Asian Restaurant	Greek Restaurant	Russian Restaurant
Allen	33.106582	-96.69402	Fast Food Restaurant	Tex-Mex Restaurant	Italian Restaurant	Indian Restaurant	Pakistani Restaurant	Mexican Restaurant	Asian Restaurant	Thai Restaurant	Cajun / Creole Restaurant	Chinese Restaurant
Coppell	32.967341	-96.98656	American Restaurant	Fast Food Restaurant	Tex-Mex Restaurant	Mexican Restaurant	Vietnamese Restaurant	Greek Restaurant	Mediterranean Restaurant	Restaurant	Japanese Restaurant	Chinese Restaurant
Flower Mound	33.024778	-97.10206	Sushi Restaurant	Thai Restaurant	Italian Restaurant	Indian Restaurant	Mediterranean Restaurant	Vietnamese Restaurant	French Restaurant	Korean Restaurant	Japanese Restaurant	Indian Restaurant
Plano	33.054671	-96.73506	Fast Food Restaurant	Asian Restaurant	Chinese Restaurant	Thai Restaurant	Italian Restaurant	Japanese Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	Mexican Restaurant
Plano	33.077070	-96.79859	Fast Food Restaurant	Asian Restaurant	Chinese Restaurant	Thai Restaurant	Italian Restaurant	Japanese Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	Mexican Restaurant
Plano	33.086868	-96.74504	Fast Food Restaurant	Asian Restaurant	Chinese Restaurant	Thai Restaurant	Italian Restaurant	Japanese Restaurant	Vietnamese Restaurant	Korean Restaurant	Indian Restaurant	Mexican Restaurant

We can also estimate the number of **1st Most Common Venue** in each cluster. Thus, we can create a bar chart that may help us to find proper labels for each cluster.

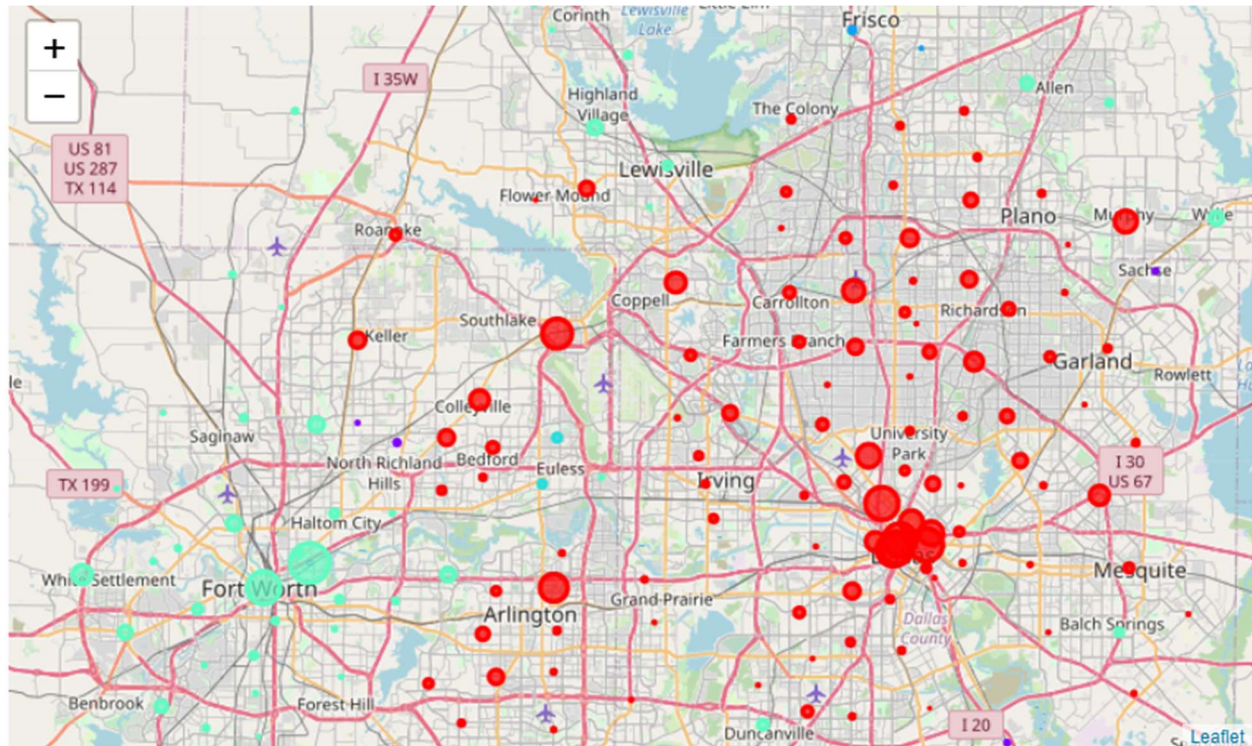


We can label this 7 cluster as follows:

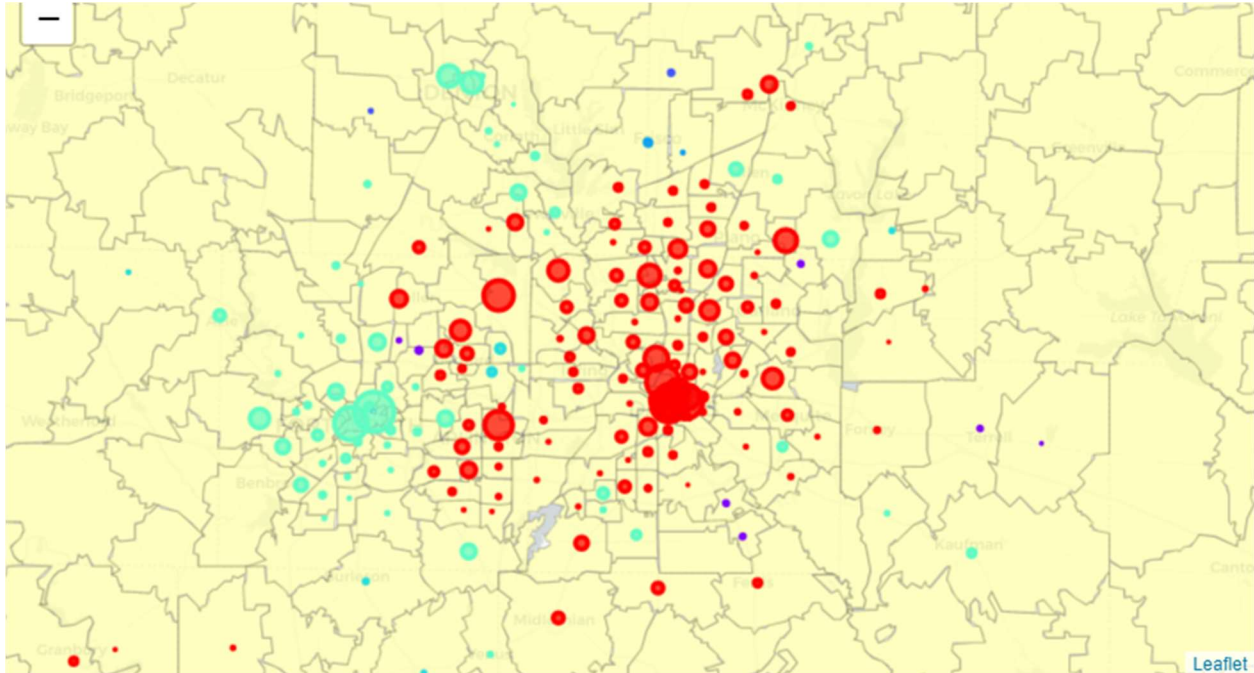
1. Great Dallas area metro.
2. Fast Food lover: Between cities and border area: Sachse, Hutchins, and Terrel.
3. The border area: Prosper and Ponder.
4. Frisco area.
5. Mexican Food lover: Lavon, Alvarado, Euless.
6. Great Fort Worth area.
7. Sanger: Chinese Food Lover.

## C. Result

Join the numbers of restaurants and DFW map.



Join the number of restaurants and DFW in each district.



## D. Discussion

### 1. Interpret the Maps

1. Maps 1: The cities in the DFW area. One city stands for one zip code.
2. Maps 2: The restaurants in the DFW area. The bigger the bubble stands for the more restaurants and vice versa. The different color shows the different groups.
3. Maps 3: The restaurants in the DFW area, which are shown in the district. Therefore, we can see which district has the most restaurants.

### 2. Name for Clustering

**According to the clustering, we got 7 groups. And the explanation for the clustering:**

1. Great Dallas area metro: Since the Great Dallas area is the largest city in the DFW area, they got a lot of choices. Most of them choose Mexican food because Mexican restaurants are the most in DFW area because these two types of restaurants are more than other restaurants.



2. Fast Food lover: The area between cities and border area: Sachse, Hutchins and Terrel. Basically, the restaurants are below 20, so they had few choices.
3. The border area: Prosper and Ponder. The top three restaurants are American food, French food, and Italian food.
4. Frisco area: Frisco had a lot of Japanese food and people like to eat.
5. Mexican Food lover: Lavon, Alvarado, Euless area.
6. Great Fort Worth area: People got top 2 choices: Fast food and Mexican food.
7. Sanger: Chinese Food Lover. Sanger has more Chinese restaurants than others.
6. Great Fort Worth area: People got top 2 choices: Fast food and Mexican food.

### **3. Investors View**

1. Investors can see the size of the bubble and decide where to start up their new restaurants.
2. Investors can use the most popular restaurants in different district table and find the potential chance. For example, although most restaurants in Plano and Allen are Mexican restaurants, Chinese restaurants are the third place in Plano. However, Chinese restaurants stand only 10th in Allen.

## **E. Conclusion**

1. Investors can open their restaurants according to the final file--result. In this file, we will know the most popular restaurants, the population, and the number of restaurants in a specific area.
2. The bubble in the map shows the first and second metroplex in the DFW area, which are the Great Dallas and Fort Worth areas. Investors can choose their ideal location to start up.

## **F. References:**

- [1] DFW Wiki, <https://en.wikipedia.org/wiki/DFW>
- [2] The demographics Texas of government, <https://demographics.texas.gov/>
- [3] Foursquare API, <https://foursquare.com/>
- [4] Opendatasoft, <https://public.opendatasoft.com/>

=====