Venues Data Analysis For Opening A Restaurant In Dallas

A. Introduction

A.1. Description & Discussion of the Background

Dallas, located in North Texas, is the 9th largest city in the United States and the 3rd largest in Texas. The current population of Dallas in 2019 is estimated at 1.3 million. The Dallas here we are talking about is Dallas County, which included in the Dallas–Fort Worth–Arlington. It is Texas' second-most populous county and the ninth-most populous in the United States.
As you can see from the figures, Dallas is a city with a high population and will become more and more attractive. When we think of it by the investor, we plan to choose a popular place to open a restaurant in Dallas.
Here are some questions lead us and need to solve in this topic：

1. Where is the best place to open?
    - Which city?
    - Which zip code area?
2. What type of restaurant we want to open?
    - Fast food/ Asian/ Mexican/ etc..

When we consider all these problems, we can create a map for checking clusters according to the venue density and data chart for showing the most popular type for restaurant.

A.2. Data Description

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

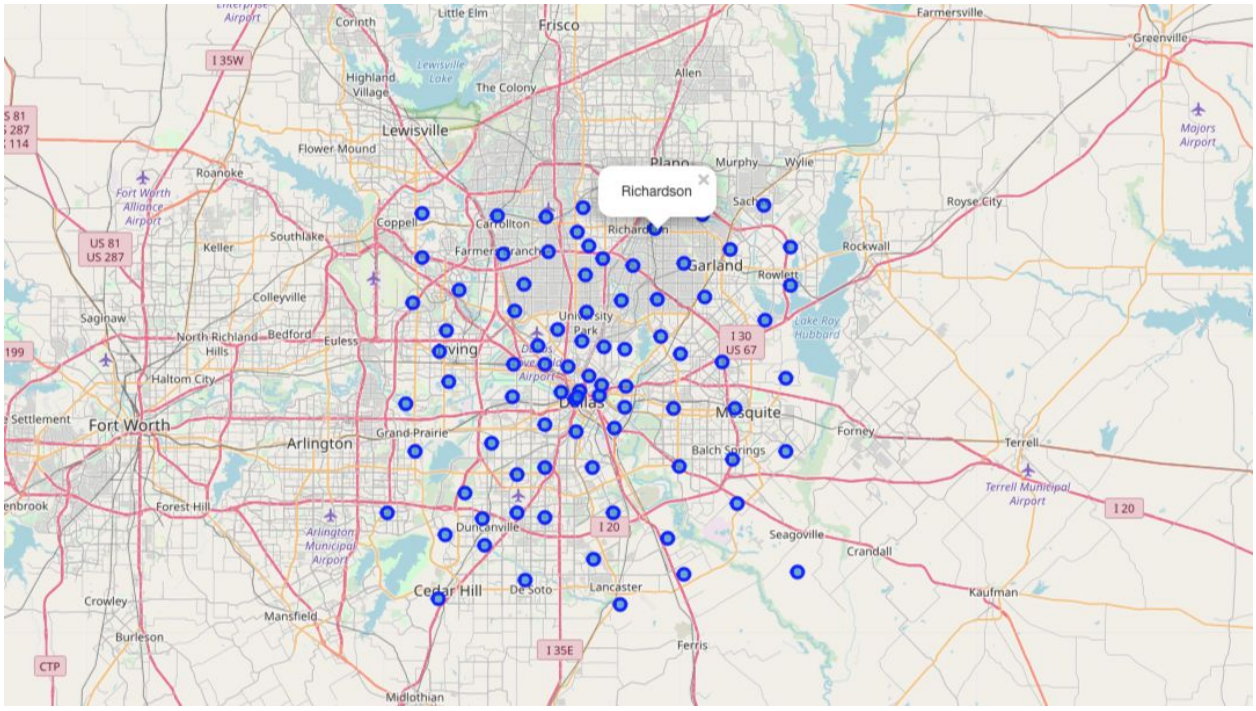
- I found the list of United States within all city, county and zip info. I cleaned the data and reduced it to state of Texas only to create map.
- I used Foursquare API to get the most common venues of given County of Dallas.
- I used Tableau to generate charts to check the most popular type of restaurants as a reference to recommend me which type of restaurant to choose to open.

B. Methodology

As a database, I used GitHub repository in my study. My master data which has the main components City , Zip, Latitude and Longitude informations of the County of Dallas.

	zip	lat	lng	city	state_id	state_name	county_name
0	75001	32.9600	-96.8385	Addison	TX	Texas	Dallas
1	75006	32.9619	-96.8970	Carrollton	TX	Texas	Dallas
2	75019	32.9633	-96.9855	Coppell	TX	Texas	Dallas
3	75038	32.8746	-96.9976	Irving	TX	Texas	Dallas
4	75039	32.8875	-96.9422	Irving	TX	Texas	Dallas

I used python folium library to visualize geographic details of Dallas and all cities with marks and 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 venue and the radius 500 meter for each borough from their given latitude and longitude informations. Here is a head of the list Venues name, category, latitude and longitude informations from Foursquare

API.

	city	city latitude	city longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Addison	32.96	-96.8385	La Spiga	32.958278	-96.837417	Bakery
1	Addison	32.96	-96.8385	Cindi's N.Y. Delicatessen	32.961864	-96.838873	Diner
2	Addison	32.96	-96.8385	Enterprise Rent-A-Car	32.960673	-96.838102	Rental Car Location
3	Addison	32.96	-96.8385	Five Star Fabric	32.959985	-96.838228	Arts & Crafts Store
4	Addison	32.96	-96.8385	Tone Shop Guitars	32.959020	-96.839050	Music Store

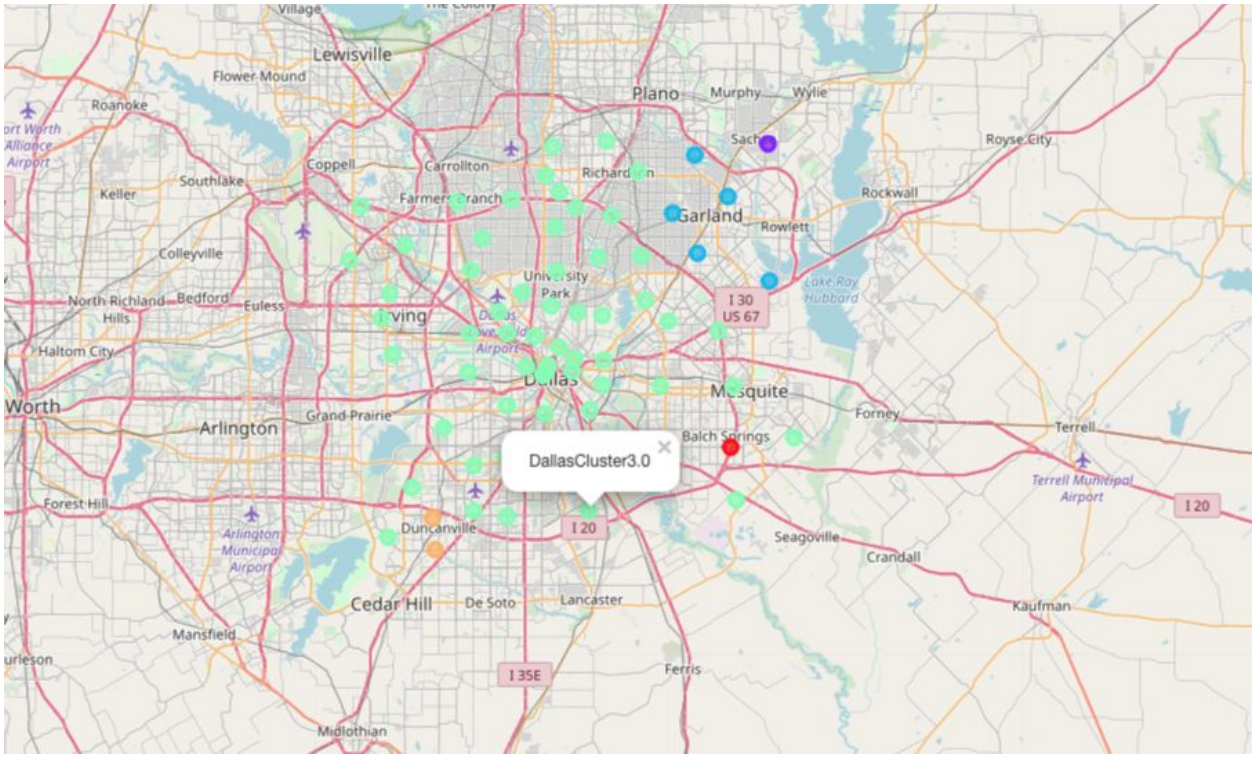
I filtered venue category only for Restaurant and get 5 most common venues:

	city	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Balch Springs	Fast Food Restaurant	Asian Restaurant	Chinese Restaurant	Seafood Restaurant	Mexican Restaurant
1	Dallas	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
2	Duncanville	Fast Food Restaurant	Vietnamese Restaurant	Thai Restaurant	Asian Restaurant	Cajun / Creole Restaurant
3	Garland	American Restaurant	Mexican Restaurant	Japanese Restaurant	Asian Restaurant	Cajun / Creole Restaurant
4	Irving	Indian Restaurant	Fast Food Restaurant	American Restaurant	Southern / Soul Food Restaurant	South Indian Restaurant

I run k-means and merged two data tables with 5 cluster labels:

	zip	lat	lng	city	state_id	state_name	county_name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
3	75038	32.8746	-96.9976	Irving	TX	Texas	Dallas	3.0	Indian Restaurant	Fast Food Restaurant	American Restaurant	Southern / Soul Food Restaurant	South Indian Restaurant
4	75039	32.8875	-96.9422	Irving	TX	Texas	Dallas	3.0	Indian Restaurant	Fast Food Restaurant	American Restaurant	Southern / Soul Food Restaurant	South Indian Restaurant
5	75040	32.9277	-96.6201	Garland	TX	Texas	Dallas	2.0	American Restaurant	Mexican Restaurant	Japanese Restaurant	Asian Restaurant	Cajun / Creole Restaurant
6	75041	32.8809	-96.6515	Garland	TX	Texas	Dallas	2.0	American Restaurant	Mexican Restaurant	Japanese Restaurant	Asian Restaurant	Cajun / Creole Restaurant
7	75042	32.9139	-96.6749	Garland	TX	Texas	Dallas	2.0	American Restaurant	Mexican Restaurant	Japanese Restaurant	Asian Restaurant	Cajun / Creole Restaurant

Also, use rainbow colors for theses 5 Clusters in the visual map to get a quick view, and it is clear to see Cluster 3 got a main area for the restaurants:



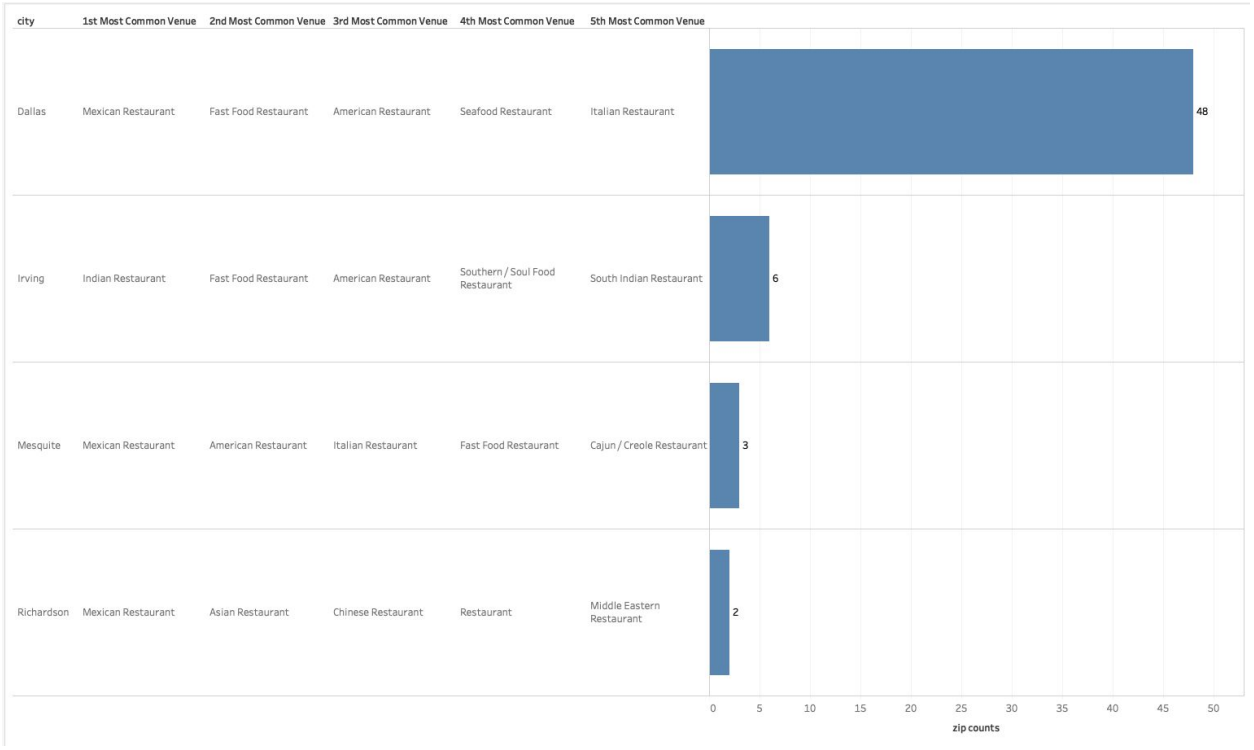
So I continued to pull out the Cluster 3 with total 59 lines detail:

	zip	city	county_name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
3	75038	Irving	Dallas	3.0	Indian Restaurant	Fast Food Restaurant	American Restaurant	Southern / Soul Food Restaurant	South Indian Restaurant
4	75039	Irving	Dallas	3.0	Indian Restaurant	Fast Food Restaurant	American Restaurant	Southern / Soul Food Restaurant	South Indian Restaurant
14	75060	Irving	Dallas	3.0	Indian Restaurant	Fast Food Restaurant	American Restaurant	Southern / Soul Food Restaurant	South Indian Restaurant
15	75061	Irving	Dallas	3.0	Indian Restaurant	Fast Food Restaurant	American Restaurant	Southern / Soul Food Restaurant	South Indian Restaurant
16	75062	Irving	Dallas	3.0	Indian Restaurant	Fast Food Restaurant	American Restaurant	Southern / Soul Food Restaurant	South Indian Restaurant
17	75063	Irving	Dallas	3.0	Indian Restaurant	Fast Food Restaurant	American Restaurant	Southern / Soul Food Restaurant	South Indian Restaurant
18	75080	Richardson	Dallas	3.0	Mexican Restaurant	Asian Restaurant	Chinese Restaurant	Restaurant	Middle Eastern Restaurant
19	75081	Richardson	Dallas	3.0	Mexican Restaurant	Asian Restaurant	Chinese Restaurant	Restaurant	Middle Eastern Restaurant
29	75149	Mesquite	Dallas	3.0	Mexican Restaurant	American Restaurant	Italian Restaurant	Fast Food Restaurant	Cajun / Creole Restaurant
30	75150	Mesquite	Dallas	3.0	Mexican Restaurant	American Restaurant	Italian Restaurant	Fast Food Restaurant	Cajun / Creole Restaurant
34	75181	Mesquite	Dallas	3.0	Mexican Restaurant	American Restaurant	Italian Restaurant	Fast Food Restaurant	Cajun / Creole Restaurant
36	75201	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
37	75202	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
38	75203	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
39	75204	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
40	75205	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
41	75206	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
42	75207	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
43	75208	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
44	75209	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
45	75210	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
46	75211	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
47	75212	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
48	75214	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
49	75215	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
50	75216	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
51	75217	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
52	75218	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant
53	75219	Dallas	Dallas	3.0	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant

C. Results

In final section, I used Tableau to get bar charts for popular city and show the below results:

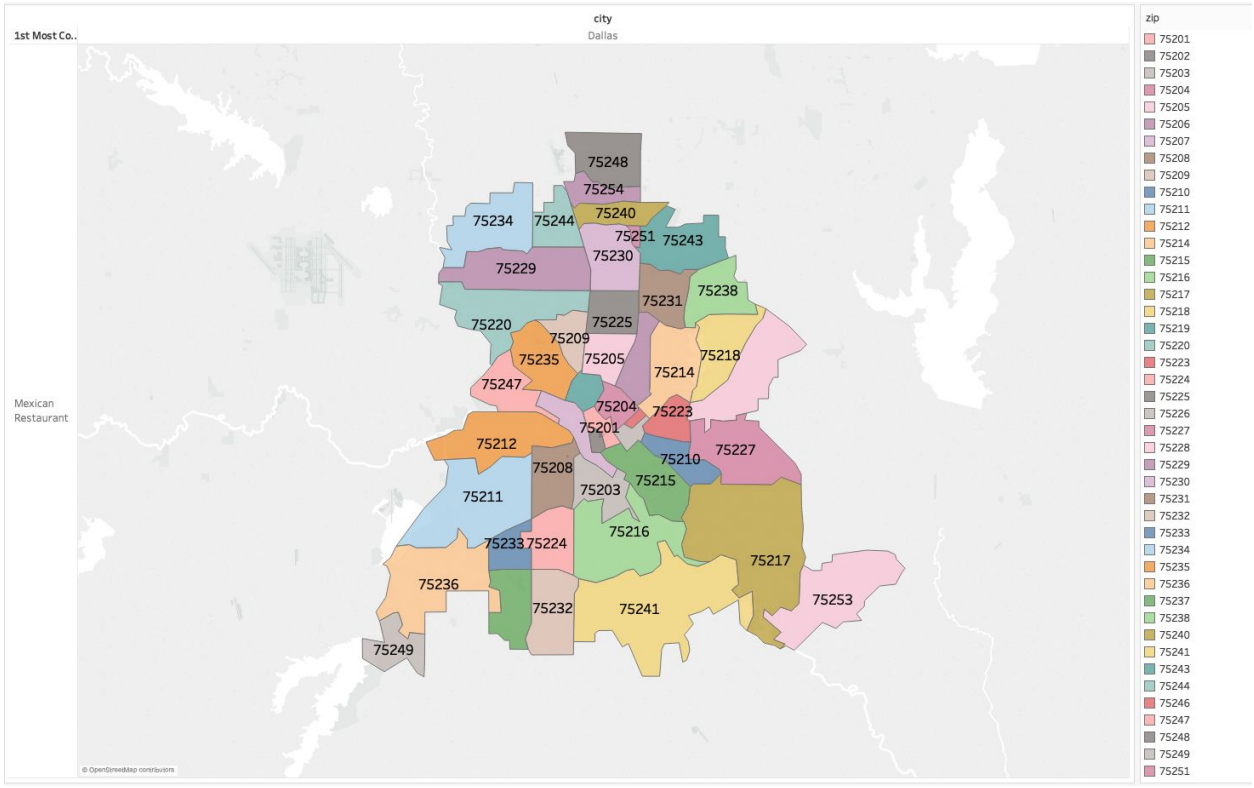
- Dallas is the most popular city for opening restaurants
- Mexican restaurant is the most popular restaurants in Dallas city



D. Discussion

However, there are too much zip codes in Dallas city than I expected, so I still need to narrow my conditions, so below are some discussions need to consider further:

1. Check the financial budget, if the budget is low, I probably need exclude the zip areas in uptown and downtown dallas with high pay store location.
2. Check the population data for Mexican people in these zip area, may narrow down some zip area.



E. Conclusion

I used the K-mean algorithm as part of this clustering study. From the Foursquare data venues, I dropped some NaN data because they cannot match with some US zip codes. Data accuracy will always need to be concerned in this study. I also performed data visualization analysis through bar chart and geographic mapping by using Tableau. In further details for codes you can also find on my GitHub.

F. Reference

[1] Foursquare API
[2] Tableau
[3] GitHub

To the future,

Elva Jin