

Dallas Restaurant Analysis

Data Science Capstone Project

Khan Ismail

Table of Contents

- I. Introduction..... 1
- II. Data Collection & Usage 2
 - A. Data Sources: 2
 - B. Data Usage..... 2
 - C. Dallas Neighborhoods 2
 - D. K-Means & Clustering 3
- 3. Conclusion 4

I. Introduction

This report contains an analysis for new restaurant location in Dallas Metro Area in Texas, USA. The client is interested in understanding different types of restaurants available in the Dallas metro area and looking for recommendations identifying the best place to open a new restaurant.

With an estimated 2019 population of 1,343,573, it is the ninth most-populous city in the U.S. and the third largest in Texas. Over 41 colleges and universities are located within its metropolitan area, which is the most of any metropolitan area in Texas. The city has a population from a myriad of ethnic and religious backgrounds and one of the largest LGBT communities in the U.S.

The city of Dallas is home to many areas, neighborhoods, and communities. Dallas can be divided into several geographical areas which include larger geographical sections of territory including many subdivisions or neighborhoods, forming macro-neighborhoods namely;

Bishop Arts District
Casa Linda
Casa View
Cedars, The
Deep Ellum
Design District
Downtown
Exposition Park
Fair Park
Kessler Park
Knox-Henderson
Lakewood
Bishop Arts District
Casa Linda

Lake Highlands
Lower Greenville
"M" Streets
Oak Cliff
Oak Lawn
Park Cities
Pleasant Grove
Preston Hollow
Southwestern Medical District
Trinity Groves
Turtle Creek
Uptown
Victory Park
West End

II. Data Collection & Usage

A. Data Sources:

Demographic data can be obtained via "<https://neighborhoods.dmagazine.com/>" whereas location data (coordinates) can be obtained using geocode.

B. Data Usage

Dataframes will be created using pandas library in python to merge the information found in the above to data sets. Folium library will be used to visualize these on a geographical map followed by K_means clustering to identify similarity of neighborhoods. FOURSQUARE API will be used to determine local venues based on the clusters. Once the data is available, this can allow us to visualize whether a certain type of food is over or undersaturated. Meaning, does a market exist for the type of food we would like to introduce in Dallas area. (see jupyter notebook for libraries imported).

C. Dallas Neighborhoods

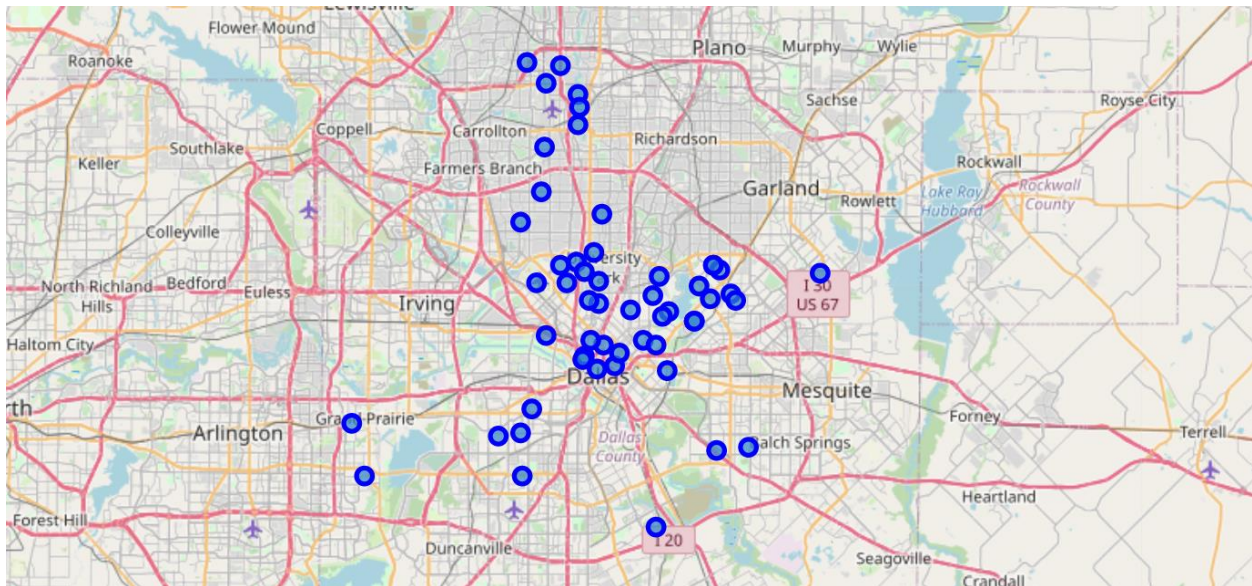


Fig 1: Map of Dallas Neighborhoods

D. K-Means & Clustering

Using Elbow Method, optimum K value is identified for this dataset:

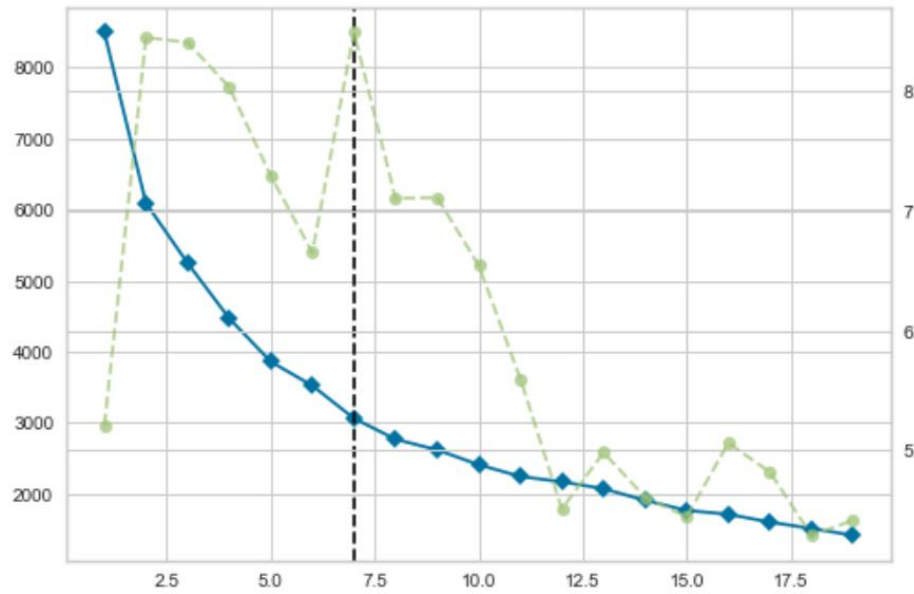


Fig 2: K means versus distortion

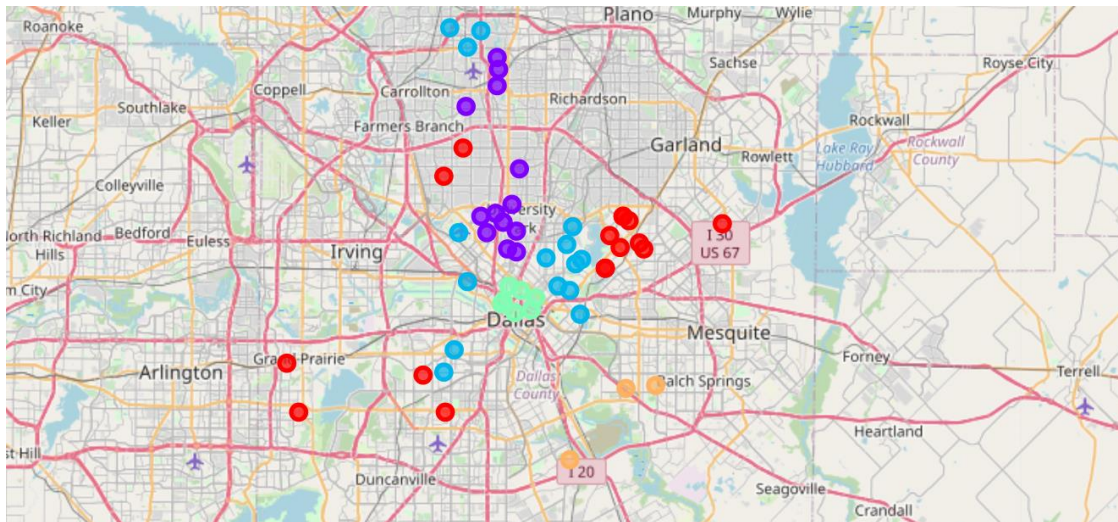


Fig 3: 5 Clusters of restaurants around Dallas Metro Area

3. Conclusion

As represented in the above maps, best place to open a new restaurant based on location data available would be the SOUTH side of Dallas city. This is because there isn't a high concentration of restaurants available there and would make a suitable place to inaugurate a new restaurant.

As with any study, there are limitations to this analysis. One of the biggest missing components is the Socio-Economic-Status (SES) of the neighborhoods. That would determine whether it would make business sense to operate a high-end restaurant in areas where average income is less than \$50,000.