Introduction

Dallas-Fort Worth-Arlington metropolitan area has been involving quickly and constantly over the past 10 years, it has been top 10 area in the nation for people to relocate. According to U.S. Census Bureau, it was the fastest-growing metro area in the U.S. in 2018. The population growth has surpassed Las Angeles population growth since 1970s. The major drive for the growth comes from high living quality offered by the area and the local economic stimulation policies and strategies, which provide attractive tax benefits for corporations doing business in the state of Texas. Many fortune 500 companies moved their headquarters to this area, including Exxon Mobile, AT&T, Toyota, American Airlines group and many others. Employment opportunities brought by these corporations continuously attract people to these area from coastal cities and all over the county.

Collin county is one of the popular counties in the Dallas-Fort Worth-Arlington area for the new comers. It resides at the north part of Dallas and encompasses 21 cities. The commute by car takes less than 30 minutes from most of these cities to Dallas.

Collin county is home of many existing and new corporations that moved in in recent years. The low crime rate, excellent schools, growing job opportunity, relatively low housing price, great recreation facilities made it a top choice for many families. As people considering Collin county as their future home, one of the decision to make is which city to settle among the 21 great cities within the county.

Problem Statement

As more than hundreds of thousands of people move into 4th largest U.S metropolitan area Dallas-Fort Worth-Arlington each year, which county and city to settle home becomes a challenge decision, because each county and city has its own charms and the metro-area has so many counties and cities

The objective of this project will analyze cities in Collin county using Foursquare location data and clustering algorithm, explore what Collin county cities can offer along with general overview of Collin county itself. Based on the discovery, cities will be categorized into different groups and comparison analysis will be performed accordingly

The project aims to provide reference about Collin county and its city dynamics for individuals or families who are considering moving into Collin county or to relocate within Collin county cities, they can leverage the analysis conducted in this project in combination with their specific needs to make a suitable decision

Data Overview

<u>Data requirement</u>

To conduct the analysis, following data will be acquired and prepared:

- List of Collin county cities, their geographical coordinators (latitude and longitude), this will be used to plot the cities on the map and acquire their venue information.
- City venue data that will be used to identify the characters of cities being analyzed
- School, crime, population, population, housing, employment data that will facilitate comparison of different cities within Collin county

<u>Data Acquisition</u>

The city geographical coordinator information can be acquired by using Python geocoder package, however through research the internet, Open Data soft offers public free dataset of U.S city geographical coordinator information in various downloadable format. A csv file that contains Collin county city name, zip code, latitude, longitude, time zone, daylight saving time flag, and geo-coordinator is extracted and same as city_zip.csv.

Foursquare has large location data that contains more than 100 million places worldwide, Foursquare API will be used to extract city venues and types to aid the analysis for finding popular top venues of each city

Other data such as school, crime, population, employment, housing will be gathered from various sources that offers free access to public and discussed at certain degree to aid the comparison of each city

All the data source link of origin will be provided in the reference section. The methodology/approach section will discuss how the datasets will be used