

The Battle of Neighborhoods

“Which City to Settle in Collin County, Texas”

IBM Applied Science Capstone Project

May 2019



Allen	Celina	dallas	coperville	frisco	plano	lavon	Mckinney	Melissa	Nevada	Princeton	wylie	Anna	prosper
	welington		sam norwood		Blue Ridge		Josephine		farmersville		westminster		weston

Table of contents

Cover Page “The Battle of the Neighborhood” -----	1
Table of Contents -----	2
Introduction -----	3
Problem Statement -----	3
Data Overview -----	4
Methodology/Approach -----	4
Result and Discussion -----	5
Recommendation -----	11
Conclusion -----	11
Appendix -----	12
Reference -----	12

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

Data requirement

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

Data Acquisition

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 to aid the analysis for finding popular top venues of each city county

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

Methodology/Approach

Python is one of the preferred programing language in data science field, it contains rich data science libraries and active development community, for this project work, Python and Jupiter notebook are used for the analysis

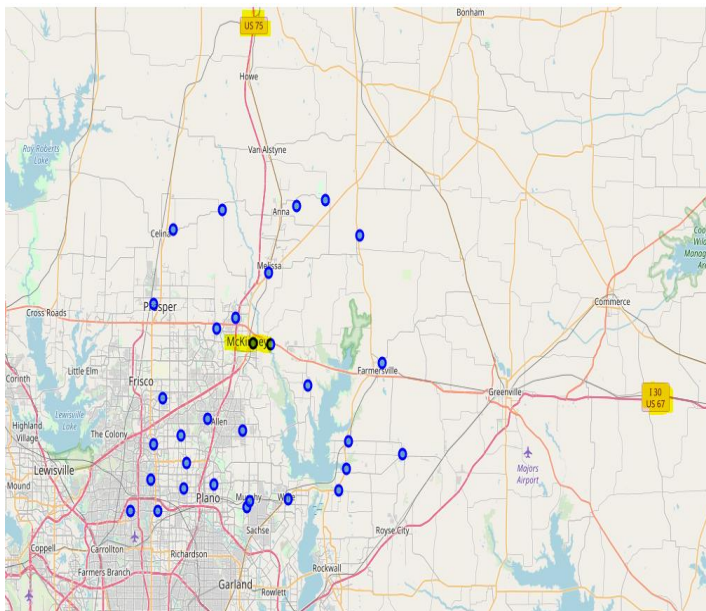
The first step is to gain general impression of cities in Collin county, information like what cities are within the boundary of Collin county, how many cities, where they are in the county, roughly how far away from each other, and what major resources or infrastructures these cities can access and so on. To answer these questions, panda package is used to load the city_zip.csv file into data frame, through counting the unique city name in data frame, 21 cities showed up in Collin county. Then by calling folium package, map is generated with all Collin cities with mark up. (see picture 1 – city location in result section)

Next step is to take a deeper look of each city, explore what it offers, what are the popular venues, this will help those individuals or families who plan to move in understand if those venues or services meet their needs. Foursquare API is called to access its local database where venue name, venue categories are stored. A sorted venue list is retrieved for all the cities being analyzed and stored in a data frame (see picture 2 – ranked venue count in result section). Then retrieved venue data are processed including encode the occurrence of each venue with binary code, sum up the total and normalize the data by mean of the frequency of each category, through ranking the category for each city, the list of top 10 venues in each city is derived and stored in data frame (see picture – 3 in result section). Then k-mean clustering algorithm is used to group the cities. Due to the challenges of k-mean clustering algorithm, which is to find the idea number for K, 10 iterative tests are performed to understand the number of k and its associated cost, the result didn't show a distinct elbow point (k number), this is largely due to the data being analyzed. Combine with the category count of each city, the modeling result showed that when there are more than 2 clusters, most of the new clusters are formed with cities having less number of venue categories and relative far from central area of the county. With the information again about total category of each city, k=4 is used for this analysis. Cluster details are listed in the result section

Additional Collin county information in terms of school, employment condition, housing and overall safety are researched from various sources and used in combination with the analysis result in the discussion and recommendation section

Result and Discussion

1. City location on map and ranked venue category



Picture 1 – City location

	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
City						
Plano	585	585	585	585	585	585
Mckinney	180	180	180	180	180	180
Dallas	175	175	175	175	175	175
Allen	132	132	132	132	132	132
Wylie	46	46	46	46	46	46
Frisco	35	35	35	35	35	35
Prosper	34	34	34	34	34	34
Melissa	21	21	21	21	21	21
Farmersville	11	11	11	11	11	11
Lavon	11	11	11	11	11	11
Westminister	6	6	6	6	6	6
Princeton	5	5	5	5	5	5
Celina	5	5	5	5	5	5
Copeville	4	4	4	4	4	4
Anna	4	4	4	4	4	4
Nevada	4	4	4	4	4	4
Blue Ridge	3	3	3	3	3	3
Sam Norwood	3	3	3	3	3	3
Wellington	3	3	3	3	3	3
Josephine	2	2	2	2	2	2
Weston	1	1	1	1	1	1

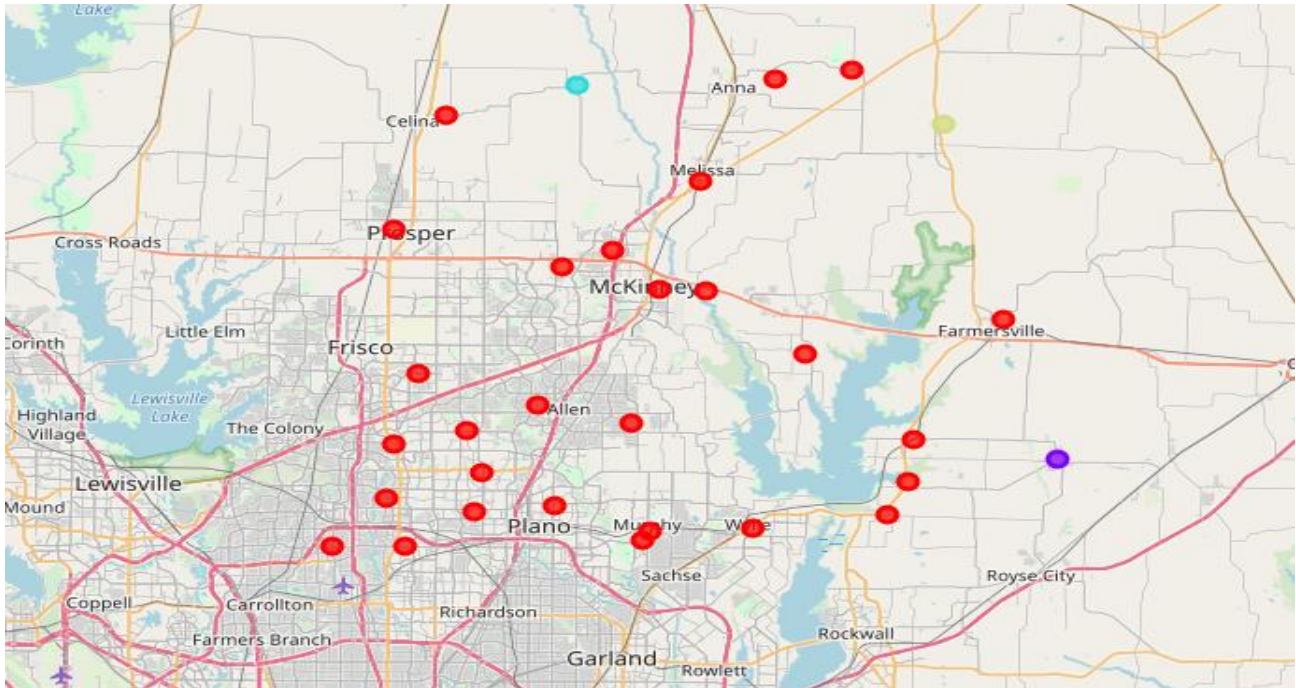
Picture 2 – Ranked Venue Counts

2. City top 10 venues list

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Allen	Fast Food Restaurant	Pizza Place	Pharmacy	Mexican Restaurant	Coffee Shop	Bank	Convenience Store	Burger Joint	Gym / Fitness Center	Gym
1	Anna	Antique Shop	Italian Restaurant	Event Service	Other Repair Shop	Yoga Studio	Fast Food Restaurant	French Restaurant	Football Stadium	Food Truck	Food Court
2	Blue Ridge	Football Stadium	Convenience Store	Clothing Store	Yoga Studio	Farmers Market	French Restaurant	Food Truck	Food Court	Food	Flower Shop
3	Celina	Café	Discount Store	Football Stadium	Pharmacy	Trail	Yoga Studio	French Restaurant	Food Truck	Food Court	Food
4	Copville	Construction & Landscaping	River	Video Store	Farmers Market	French Restaurant	Football Stadium	Food Truck	Food Court	Food	Flower Shop
5	Dallas	Mexican Restaurant	Convenience Store	Fast Food Restaurant	Pizza Place	Sushi Restaurant	American Restaurant	Coffee Shop	Seafood Restaurant	Fried Chicken Joint	Grocery Store
6	Farmersville	Discount Store	Restaurant	Café	American Restaurant	Diner	Mexican Restaurant	Food	Chinese Restaurant	Gift Shop	Fast Food Restaurant
7	Frisco	Video Store	Pizza Place	Coffee Shop	Athletics & Sports	Convenience Store	Dessert Shop	Home Service	Donut Shop	South Indian Restaurant	Mobile Phone Shop
8	Josephine	Gas Station	Park	Yoga Studio	Fried Chicken Joint	Football Stadium	Food Truck	Food Court	Food	Flower Shop	Flea Market
9	Lawn	Mexican Restaurant	Discount Store	Convenience Store	BBQ Joint	Gas Station	Auto Garage	Fast Food Restaurant	Sandwich Place	Intersection	Donut Shop
10	McKinney	Pizza Place	Mexican Restaurant	American Restaurant	Sandwich Place	Café	Convenience Store	Burger Joint	Fried Chicken Joint	Discount Store	Pharmacy
11	Melissa	Gas Station	Convenience Store	Donut Shop	Home Service	Thai Restaurant	Pizza Place	Sandwich Place	Construction & Landscaping	Pharmacy	Fast Food Restaurant
12	Nevada	Dive Bar	Taco Place	Italian Restaurant	Gas Station	Yoga Studio	Fast Food Restaurant	French Restaurant	Football Stadium	Food Truck	Food Court
13	Plano	Pizza Place	Mexican Restaurant	Fast Food Restaurant	Sandwich Place	American Restaurant	Coffee Shop	Park	Pharmacy	Convenience Store	Burger Joint
14	Princeton	Farm	Stables	Insurance Office	Auto Workshop	Fast Food Restaurant	French Restaurant	Football Stadium	Food Truck	Food Court	Food
15	Prosper	Clothing Store	Café	American Restaurant	Pizza Place	Burger Joint	Fast Food Restaurant	Health & Beauty Service	Supermarket	Steakhouse	Spa
16	Sam Norwood	Hotel	Fast Food Restaurant	Motel	Discount Store	Fish & Chips Shop	Fried Chicken Joint	French Restaurant	Football Stadium	Food Truck	Food Court
17	Wellington	Diner	Fast Food Restaurant	Construction & Landscaping	Fried Chicken Joint	French Restaurant	Football Stadium	Food Truck	Food Court	Food	Flower Shop
18	Westminster	Campground	Golf Course	Burger Joint	Restaurant	Dive Bar	Discount Store	Football Stadium	Food Truck	Food Court	Food
19	Weston	American Restaurant	Yoga Studio	Fast Food Restaurant	Fried Chicken Joint	French Restaurant	Football Stadium	Food Truck	Food Court	Food	Flower Shop
20	Wylie	Fast Food Restaurant	Pizza Place	American Restaurant	Mexican Restaurant	Pharmacy	Grocery Store	Coffee Shop	Sandwich Place	Gas Station	Italian Restaurant

Picture 3 – top 10 venue list

3. City cluster details



Cluster 1

```
collin_merged.loc[collin_merged['Cluster_Labels'] == 0, collin_merged.columns[[2] + list(range(5, collin_merged.shape[1]))]]
```

	City	State	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Allen	TX	0	Fast Food Restaurant	Pizza Place	Bank	Mexican Restaurant	Coffee Shop	Pharmacy	Convenience Store	Gym	Baseball Field	Park
1	Celina	TX	0	Café	Discount Store	Mexican Restaurant	Pharmacy	Yoga Studio	French Restaurant	Football Stadium	Food Truck	Food Court	Food
2	Allen	TX	0	Fast Food Restaurant	Pizza Place	Bank	Mexican Restaurant	Coffee Shop	Pharmacy	Convenience Store	Gym	Baseball Field	Park
3	Plano	TX	0	Pizza Place	Fast Food Restaurant	Mexican Restaurant	Sandwich Place	American Restaurant	Coffee Shop	Park	Pharmacy	Convenience Store	Burger Joint
4	Plano	TX	0	Pizza Place	Fast Food Restaurant	Mexican Restaurant	Sandwich Place	American Restaurant	Coffee Shop	Park	Pharmacy	Convenience Store	Burger Joint
5	Plano	TX	0	Pizza Place	Fast Food Restaurant	Mexican Restaurant	Sandwich Place	American Restaurant	Coffee Shop	Park	Pharmacy	Convenience Store	Burger Joint
6	Plano	TX	0	Pizza Place	Fast Food Restaurant	Mexican Restaurant	Sandwich Place	American Restaurant	Coffee Shop	Park	Pharmacy	Convenience Store	Burger Joint
7	Frisco	TX	0	Video Store	Pizza Place	Dessert Shop	Home Service	Flower Shop	Athletics & Sports	Convenience Store	Donut Shop	Business Service	Lake
8	McKinney	TX	0	Pizza Place	American Restaurant	Sandwich Place	Mexican Restaurant	Burger Joint	Café	Convenience Store	Pharmacy	Fast Food Restaurant	Discount Store
9	McKinney	TX	0	Pizza Place	American Restaurant	Sandwich Place	Mexican Restaurant	Burger Joint	Café	Convenience Store	Pharmacy	Fast Food Restaurant	Discount Store
10	McKinney	TX	0	Pizza Place	American Restaurant	Sandwich Place	Mexican Restaurant	Burger Joint	Café	Convenience Store	Pharmacy	Fast Food Restaurant	Discount Store
11	Plano	TX	0	Pizza Place	Fast Food Restaurant	Mexican Restaurant	Sandwich Place	American Restaurant	Coffee Shop	Park	Pharmacy	Convenience Store	Burger Joint
12	Plano	TX	0	Pizza Place	Fast Food Restaurant	Mexican Restaurant	Sandwich Place	American Restaurant	Coffee Shop	Park	Pharmacy	Convenience Store	Burger Joint
13	Prosper	TX	0	Café	Clothing Store	Gym / Fitness Center	Burger Joint	Fast Food Restaurant	Pizza Place	American Restaurant	Dessert Shop	Coffee Shop	Food Truck
14	Plano	TX	0	Pizza Place	Fast Food Restaurant	Mexican Restaurant	Sandwich Place	American Restaurant	Coffee Shop	Park	Pharmacy	Convenience Store	Burger Joint
15	Plano	TX	0	Pizza Place	Fast Food Restaurant	Mexican Restaurant	Sandwich Place	American Restaurant	Coffee Shop	Park	Pharmacy	Convenience Store	Burger Joint
16	Plano	TX	0	Pizza Place	Fast Food Restaurant	Mexican Restaurant	Sandwich Place	American Restaurant	Coffee Shop	Park	Pharmacy	Convenience Store	Burger Joint
18	Wylie	TX	0	Fast Food Restaurant	Pizza Place	American Restaurant	Mexican Restaurant	Gas Station	Italian Restaurant	Grocery Store	Coffee Shop	Sandwich Place	Pharmacy
19	Copville	TX	0	Pet Store	Racetrack	Construction & Landscaping	Video Store	Farm	Football Stadium	Food Truck	Food Court	Food	Flower Shop
21	Lavon	TX	0	Mexican Restaurant	Auto Garage	BBQ Joint	Discount Store	Fast Food Restaurant	Donut Shop	Intersection	Sandwich Place	Convenience Store	Gas Station
22	Nevada	TX	0	Grocery Store	Taco Place	Italian Restaurant	Gas Station	Farmers Market	Football Stadium	Food Truck	Food Court	Food	Flower Shop
23	Dallas	TX	0	Mexican Restaurant	Fast Food Restaurant	Pizza Place	Convenience Store	Sushi Restaurant	American Restaurant	Coffee Shop	Fried Chicken Joint	Sandwich Place	Gym
24	Dallas	TX	0	Mexican Restaurant	Fast Food Restaurant	Pizza Place	Convenience Store	Sushi Restaurant	American Restaurant	Coffee Shop	Fried Chicken Joint	Sandwich Place	Gym
25	Princeton	TX	0	Farm	Stables	Construction & Landscaping	Auto Workshop	Fast Food Restaurant	French Restaurant	Football Stadium	Food Truck	Food Court	Food

Cluster 2

```
In [121]: collin_merged.loc[collin_merged['Cluster_Labels'] == 1, collin_merged.columns[[2] + list(range(5, collin_merged.shape[1]))]]
```

```
Out[121]:
```

	City	State	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
20	Josephine	TX	1	Gas Station	Park	Yoga Studio	Frozen Yogurt Shop	French Restaurant	Football Stadium	Food Truck	Food Court	Food	Flower Shop

Cluster 3

```
In [122]: collin_merged.loc[collin_merged['Cluster_Labels'] == 2, collin_merged.columns[[2] + list(range(5, collin_merged.shape[1]))]]
```

```
Out[122]:
```

	City	State	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
17	Weslton	TX	2	American Restaurant	Yoga Studio	Fast Food Restaurant	Fried Chicken Joint	French Restaurant	Football Stadium	Food Truck	Food Court	Food	Flower Shop

Cluster 4

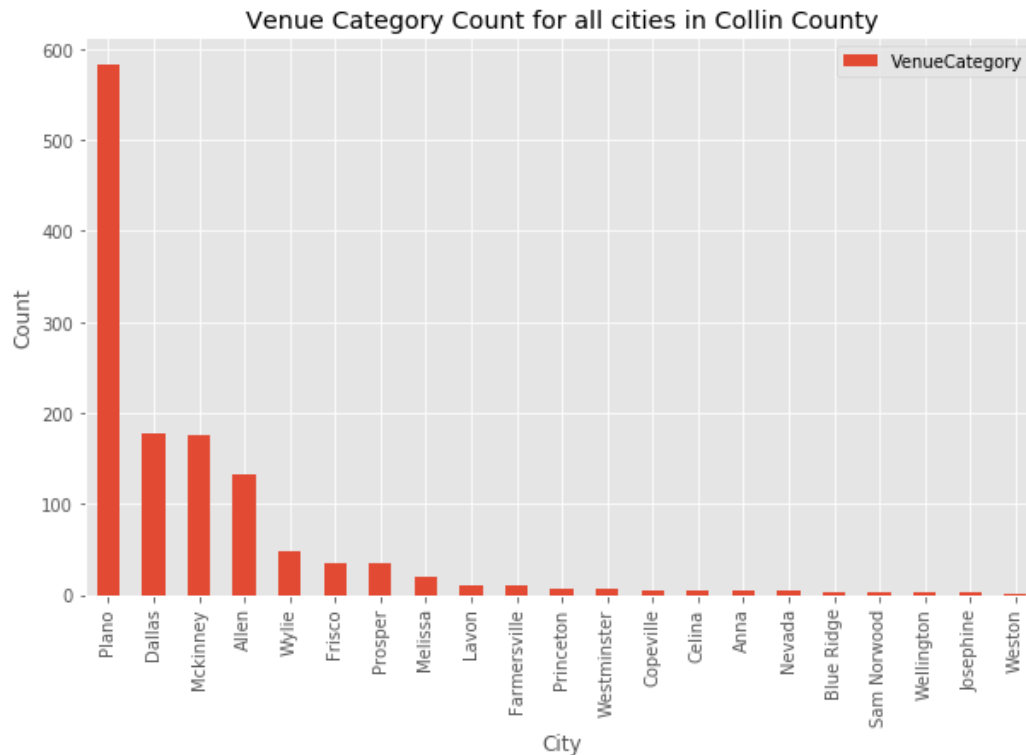
```
In [123]: collin_merged.loc[collin_merged['Cluster_Labels'] == 3, collin_merged.columns[[2] + list(range(5, collin_merged.shape[1]))]]
```

```
Out[123]:
```

	City	State	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
27	Blue Ridge	TX	3	Football Stadium	Convenience Store	Clothing Store	Yoga Studio	Farmers Market	French Restaurant	Food Truck	Food Court	Food	Flower Shop

Looking at the cities marked up within Collin county boundary, if take McKinney as county center, use I30 US 67 as horizontal axis and US75 as vertical axis, it appears that southwest is most condensed area, next is southeast, both northwest and northeast area are relatively less populated in terms of number of cities.

Among all the cities, Plano is the most vibrant and contains more than 585 venues, then followed by McKinney, Dallas and Allen. Next group consists with Frisco, Prosper, Mellissa and Wylie, these 4 cities had tremendous developments in recent years and added quite number of services as population increased. The last group contains cities scattered relatively far from county center, each of them has less number of venues comparatively, however some of them are next to great lakes, green land and natural resources.



Glance over top 10 venue categories of all the cities in cluster 1, Collin county in general is a great place to live, plenty of dining options, both high end and fast food, sports and fitness facilities like baseball fields, football stadiums, gym and yoga studios, pharmacies, parks and lakes. If people like country style, there are several cities have farms. Around county central area, Plano, McKinney, Dallas, prosper are top choices for dinning, more than 70% of top venues are restaurants and fast food places; Frisco and Allen have great sports facilities in addition to parks and lakes. Little away from center area, country style life is accessible including farms, ranches and big lakes. There are also plenty of convenience stores and pharmacies, a good place for aspiring pharmacist to find employment opportunities. Another interesting observation is that people in Collin county love flowers, flower shops appeared in top 10 venue categories in most of the cities

Cluster 2, 3, 4, contain total 3 cities, with each in its own cluster. These three cities sit at far north, northeast and southeast border area of Collin county, away from most of the cities in cluster 1, which are around center area. Same as cities in cluster 1, good number of restaurants and convenient food stores are offered by these cities, the uniqueness about these cities is that they are favorite place for football players. Football stadium is ranked as No. 1 top venues in blue ridge (city in cluster 3) and No. 6 in Weston (city in cluster 4)

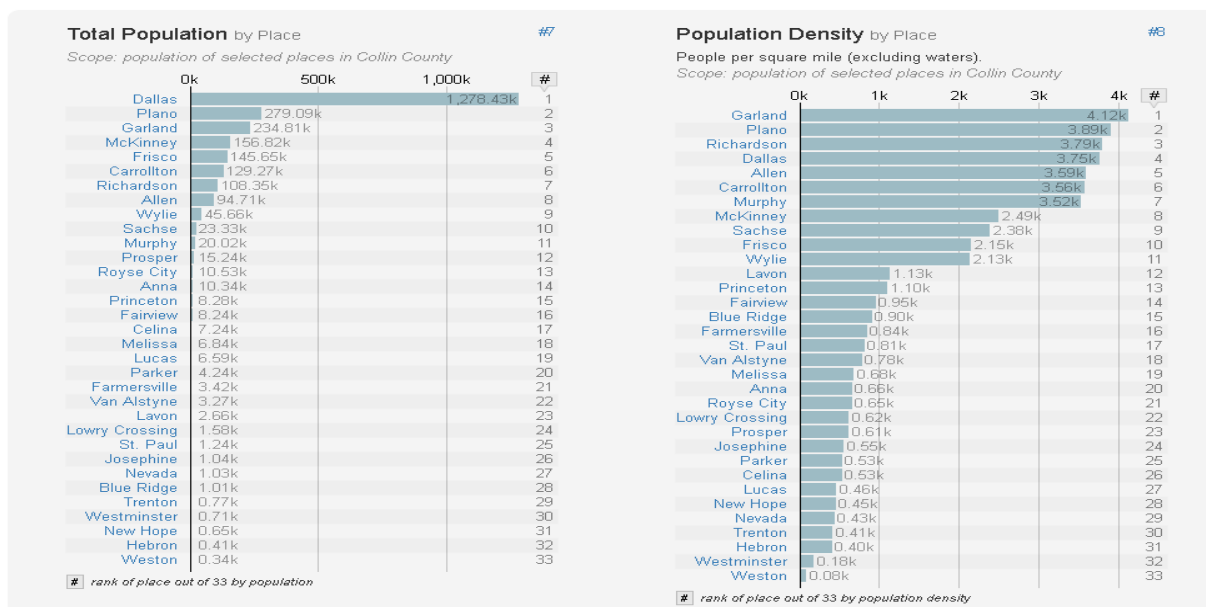
Collin county has great public schools with high diversity. According to public school review, Plano independent school district is most diversified school district. Student to teacher ratio is 15:1. Among all the school district, Plano school is also known as the most competitive school district, especially in

west Plano areas. Frisco, Allen, McKinney also have many top-ranked schools, lots of them are rated 9 out of 10 or 10

As mentioned in the beginning of this report, people from all over the country continuous move into the state of Texas, population has grown tremendously, city and area development is expanding quickly. Among all the cities of Collin county, Plano is most populated followed by McKinney, Frisco, Allen, WY line. Following chart (source of origin in reference section) shows the population distribution

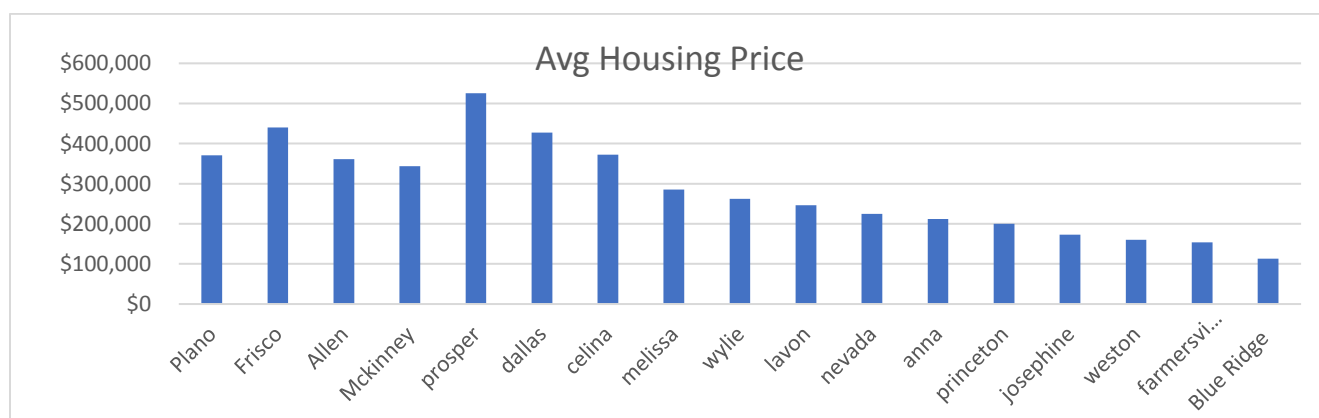
Population by Place in Collin County

There are 33 places that are fully or partially contained within Collin County (18 fully and 15 partially). This section compares all 33 of those to each other.



Source from statisticalatas.com

Housing price is steadily rising in recent years, following chart shows some of the Collin county average housing price based on 2018 data from Zillow.com



Employment outlook is great, there are plenty of employment opportunities in Collin county that have kept county's unemployment rate at the lower edge of 3%. Most of the corporate employees work in

the central area cities like Plano, Frisco, McKinney, Allen, and Dallas, and a large portion of them live in Plano, Frisco, Allen and McKinney (source from Collin County economic overview).

In terms of overall safety, according to sheriff's office and county policy report, the overall crime rate of Collin County is very low compare to other counties in Dallas-Fort Worth-Arlington metropolitan area and nationwide.

Recommendation

Collin county in general is a very safe, fast-growing and relative low-cost area to live. For individuals who are seeking great employment opportunity and vibrant amenities, or families hold good income, have school kids and looking for the best schools in the area, Plano, Frisco, McKinney, Allen are great options, they all located around center area of the county and access to popular city venues are very convenient. The downside is average housing price in these four cities are rising and relatively higher than other cities far from central area. Among these four cities, Frisco and Allen also offer attractive sports facilities, Plano and Allen have many parks and natural trails for people to relax.

For people or families like quiet rural community feeling but are within the proximity to bigger-city amenities, if top schools are not the primary concern and willing for longer commute, Weston, Blue Ridge, Josephine, and some of the cities around the county border are good choices. As an incentive, average housing prices are much lower compare to those in the center area.

If families need relatively good school (school ratings are 6 or higher out of 10) and mixed city-country living, doesn't want to sacrifice convenience to good dining, shopping and recreation, but willing to commute at certain degree in exchange of bigger and better housing with same or lower prices compare to those in Plano, Frisco, Allen and McKinney, a group of cities like Melissa, Princeton, Wyllie, Anna identified in cluster 1 are good alternatives. Prosper and Celina are two exceptional in terms of housing price in this group, this is due to explorative demand of housing in recent years along with people seeking some level of rural living environment.

Conclusion

The project used data and information collected from various sources and provided high level information about Collin county, then took a deeper look about cities within the county by leveraging Foursquare local data through applying K-Mean clustering machine learning algorithm. Combining the clustering results with collected school, housing, employment, commute information of each city group, three distinct options provided for individuals or families who are considering move into Collin county or planning to relocate within Collin county cities. Hopefully this will help them to have a better understanding about city dynamics and make a suitable decision that meet their specific needs.

Regardless which option people finally decide, Collin county in general is a fast growing, relatively low-cost living area that offers different style of high quality living

Appendix

1. Link to GitHub Notebook that contains the code work for this report <https://github.com/copperAn/DS-Projects-/blob/master/wk5%20capstone%20code%20book.ipynb>
2. Link to presentation in GitHub that related to this report <https://github.com/copperAn/DS-Projects-/blob/master/Applied%20Data%20Science%20Capstone%20Presentation.pdf>

References

1. Dallas vs Las Angeles population growth research <https://www.star-telegram.com/news/business/growth/article212129739.html>
2. Collin county public school information <https://www.publicschoolreview.com/texas/collin-county>
3. Collin county population information <https://statisticalatlas.com/county/Texas/Collin-County/Population>
4. School Ranking <https://www.niche.com/k12/search/best-school-districts/c/collin-county-tx/>
5. Employer Source <http://www.discovercollincounty.com/top-employers-collin-county/> and <https://www.niche.com/k12/search/best-school-districts/c/collin-county-tx/>