DALLAS NEIGHBORHOODS ANALYSIS—A COMBINATION OF GEOSPATIAL AND DEMOGRAPHICAL APPROACHES

CAPSTONE PROJECT

Introduction of Problem:

My Friend, XY, has been offered a great opportunity to work for a leading firm in Dallas city. He is very excited at this opportunity while a little bit worried about the new environment he is going to move to. Acknowledged that I have learned some geo data analysis skills, he approached me for help.

So the key question is: How can my friend find a convenient and enjoyable place to live in Dallas city?

DATA ACQUISITION AND CLEANING

1. Scraping Dallas Neighborhood Table from website
2. Getting Coordinates of Neighborhoods by using Geocoder package
3. Using Foursquare Location Data

DATA ACQUISITION AND CLEANING

1. special characters such as $, % populated as part of the string values were removed
2. using pandas to-numeric function to transform all object columns into floating numbers.

METHODOLOGY

EXPLORATORY ANALYSIS:

Analysis was done to find out what are the common venues for each neighborhood to get a quick impression first of Dallas neighborhoods. This may give an idea what the final cluster could be.

METHODOLOGY

	Neighborhood	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	
0	Bent Tree	Airport Terminal	Spa	Intersection	History Museum	Italian Restaurant	Shipping Store	Food Truck	Burger Joint	Sandwich Place	
1	Bluffview	Indian Restaurant	Ice Cream Shop	Korean Restaurant	Fried Chicken Joint	Bubble Tea Shop	Burger Joint	Supermarket	Lake	Bookstore	
2	Casa Linda	Park	Sandwich Place	Italian Restaurant	Fast Food Restaurant	Mediterranean Restaurant	Snack Place	Dessert Shop	Grocery Store	Mexican Restaurant	C
3	Casa View	Mexican Restaurant	Chinese Restaurant	Thrift / Vintage Store	Pizza Place	Locksmith	Sandwich Place	Pharmacy	Bank	Dessert Shop	
4	Central Dallas	Hotel	Coffee Shop	Bar	American Restaurant	Cocktail Bar	Mexican Restaurant	Plaza	Park	Italian Restaurant	

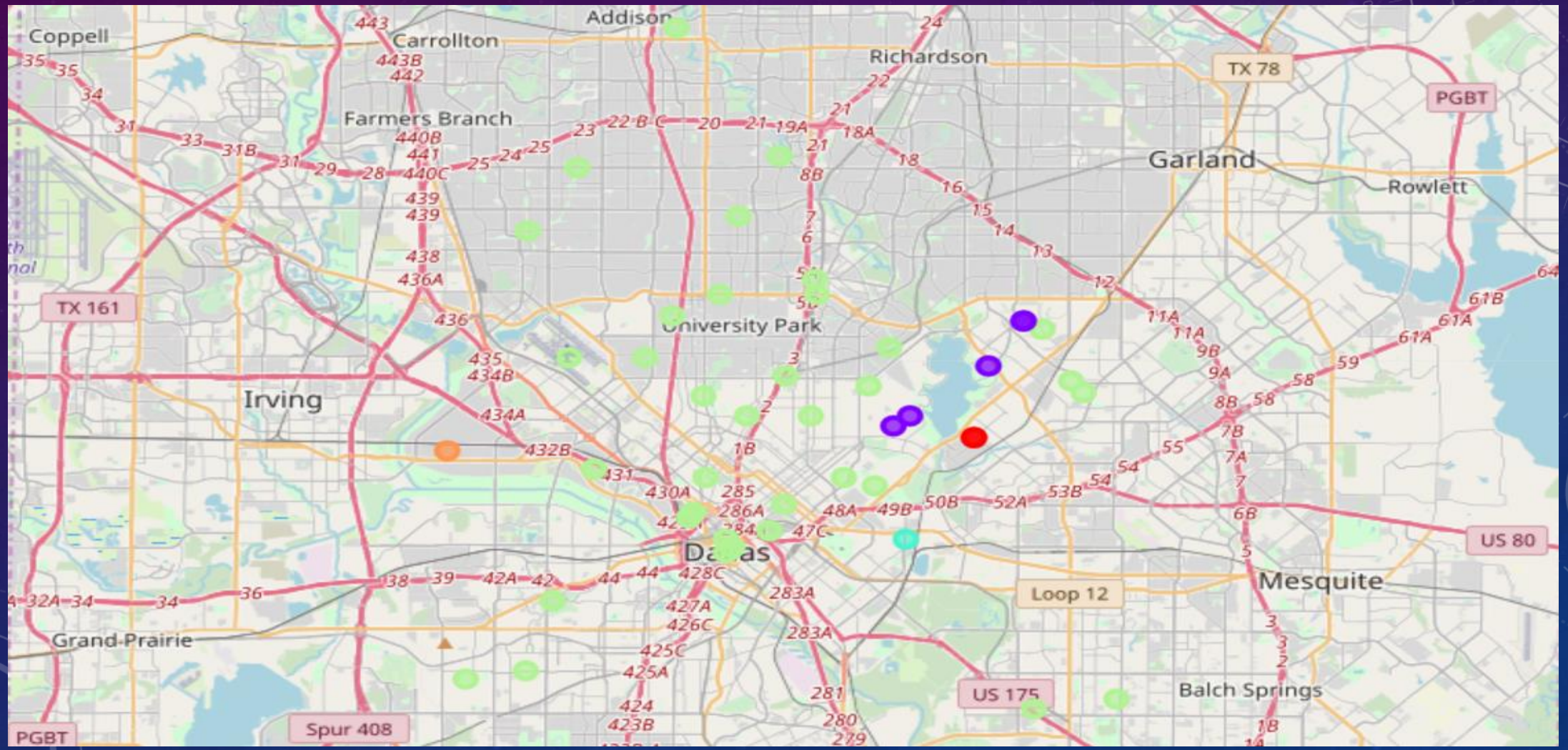
METHODOLOGY

Clustering technique:

In this project, the 1st part is to clustering of 57 neighborhood using venues in 1000 meters around the longitude and latitude of each neighborhood. The goal of this cluster analysis to find the areas that is more accessible to parks.

The second part is to recluster of the neighborhoods with demographic variables, which is particular important when considering a home purchase.

METHODOLOGY



RESULTS

Neighborhoods in cluster 1 are the best communities to access parks nearby.

e	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	1	Park	Sandwich Place	Italian Restaurant	Fast Food Restaurant	Mediterranean Restaurant	Snack Place	Dessert Shop	Grocery Store	Mexican Restaurant	Coffee Shop
0	1	Park	IT Services	Pharmacy	Food	Video Store	Chinese Restaurant	Fast Food Restaurant	Pizza Place	Cosmetics Shop	Dive Bar
0	1	Park	Fishing Spot	Lake	Boat or Ferry	Trail	Golf Course	Harbor / Marina	Scenic Lookout	Dive Bar	Dog Run
0	1	Locksmith	Golf Course	Park	Zoo	Dog Run	Donut Shop	Dry Cleaner	Electronics Store	Ethiopian Restaurant	Discount Store
0	1	Park	IT Services	Pharmacy	Food	Video Store	Chinese Restaurant	Fast Food Restaurant	Pizza Place	Cosmetics Shop	Dive Bar
0	1	Trail	Convenience Store	Shopping Mall	Gym	Park	Light Rail Station	Fast Food Restaurant	Pizza Place	Fried Chicken Joint	Donut Shop
0	1	Park	Bakery	Harbor / Marina	Gas Station	Used Bookstore	Italian Restaurant	Tex-Mex Restaurant	Beer Bar	Public Art	Wine Bar

RESULTS

Characteristics of each cluster in terms of demographics

dcluster	Pop_density	Median_age	age18+	Household_income	Median_home_value	Owner percent	Commute_time
0	5283.071429	32.457143	0.723143	42252.714286	180224.000000	0.446071	29.000000
1	4122.937500	41.595833	0.777208	85448.541667	328356.458333	0.754583	25.791667
2	6218.800000	34.166667	0.900667	66325.833333	281634.333333	0.143833	22.166667
3	4522.600000	37.350000	0.738500	159716.250000	964580.000000	0.787000	19.000000
4	3268.350000	30.700000	0.685000	35234.500000	89608.000000	0.559000	32.500000
5	2337.216667	49.741667	0.842283	101305.833333	646573.000000	0.613850	22.333333

RESULTS

Park friendly neighborhoods:

'Casa Linda', 'Lake Highlands', 'Lakewood', 'Lakewood Heights', 'Old Lake Highlands', 'Southwest Dallas', 'The Peninsula'.

Demo Friendly Neighbourhoods:

'Bluffview', 'Casa Linda', 'Eastwood', 'Forest Hills', 'Hillside', 'Hollywood Heights - Santa Monica', 'Kiestwood', 'Lake Park Estates', 'Lakewood', 'Lakewood Heights', 'Little Forest Hills', 'Lochwood', 'Love Field', 'Lower Greenville', 'M Streets - Vickery Place', 'Midway Hollow', 'Northaven Park', 'Northwood Hills - Valley View', 'Old Lake Highlands', 'Preston Highlands', 'Prestonwood', 'Ridgewood Park', 'The Peninsula', 'Wilshire Heights'.

Recommended:

'Old Lake Highlands', 'The Peninsula', 'Lakewood', 'Lakewood Heights', 'Casa Linda'.

DISCUSSION

This project gives us some very important preliminary information on possibilities of developing a tool to help new movers in the relocation.

With the booming of Dallas economy and its population, this analysis is also applicable for anyone interested in exploring new opportunities in any city.