Texas City Segmentation for Urban Agriculture Projects

Coursera Capstone Project

IBM Data Science Specialization

David Campos

https://github.com/vidcampos.

Introduction

Background

Increasing population over time, less usable land and hard-to-predict climate conditions.

changes on the traditional agricultural practices, leading in a continuously changing new paradigm

Problem

Transportation contamination, transport logistic problems, products quality loss and added charges which the customers pay at the end of the supply chain.

Objective

To segment Texas counties and cities by its house income, commercial water costs and potential clients search for intersting cities to establish urban agriculture projects.

Data adquisition

Texas Water Cost data

Dataset by the Texas Municipal League (https://www.tml.org/) 'Water and Wastewater Rate Survey' for https://www.tml.org/229/Water-Wastewater-Survey-Results

Texas Household Incomes by County data

. The information was compiled by the The County Information Program, Texas Association of Counties from the U.S. Census Bureau. Small Area Income & Poverty Estimates of 2019. The data can be access here: https://www.txcip.org/tac/census/morecountyinfo.php?MORE=1013

Data adquisition

Texas cities ZIP Codes and coordenates data

Dataset was taken from OpenDataSoft and you can get the original dataset in different formats here: https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/?refine.state=TX

Texas cities by county data.

Information of Texas cities and its belonging County(s) from Wikipedia, You can access data here: https://en.wikipedia.org/wiki/List_of_cities_in_Texas

Variables in the merged database

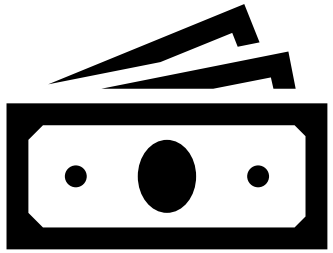
Numerical Information

- County Income.
- Total water customers by city.
- Average water usage by city.
- Water cost for residential use (5k and 10k gallons per month).
- City population
- Water cost for commercial user (50k and 200k gallons per month).

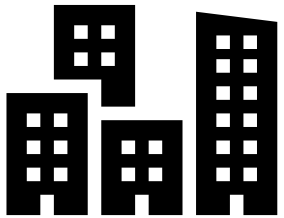
Location Information

- County.
- City.
- Zipcode.
- City latitude.
- City longitude

Parameters to consider cities



Average Income over $56.5000



Over 20.000 citizens per city

Water Cost under average in all cities ($280)

Table 1. Resulted table after data merging and filter parameters applied.

	County	Income	City	Zip	State	Latitude	Longitude	CityPopulation	TotalCustomers	AverageUsage	5k Gal	10k Gal	50k Gal	200k Gal
0	Brazoria	74225.0	Lake Jackson	77566	TX	29.036879	-95.441030	27604	9061	5800	26.40	47.90	248.40	968.40
1	Collin	96936.0	Allen	75013	TX	33.106582	-96.694020	100685	31488	10000	34.20	54.20	245.65	760.15
2	Comal	76523.0	New Braunfels	78135	TX	29.738502	-98.087157	79152	39285	6500	20.55	33.53	153.41	537.56
3	Dallas	59838.0	Coppell	75099	TX	32.771030	-96.799630	41941	13398	10000	33.60	49.35	213.94	686.44
4	Dallas	59838.0	Dallas	75294	TX	32.767268	-96.777626	1341075	334211	8300	18.84	38.84	231.84	839.50
5	Dallas	59838.0	Duncanville	75138	TX	32.767268	-96.777626	39826	12626	0	29.04	51.11	260.31	1077.81
6	Dallas	59838.0	Grand Prairie	75050	TX	32.759922	-97.012160	193837	48338	9390	35.02	55.22	265.78	940.78
7	Denton	88384.0	Carrollton	75010	TX	33.030556	-96.893280	135710	39577	8250	23.36	40.06	158.92	505.42
8	Denton	88384.0	Denton	76204	TX	33.207430	-97.116282	136268	35947	7117	36.75	57.50	274.00	941.50
9	Galveston	71959.0	Friendswood	77546	TX	29.516873	-95.194720	39839	12978	10000	21.30	35.80	189.95	624.95
10	Guadalupe	73864.0	Seguin	78155	TX	29.564780	-97.962830	30006	7507	4500	42.55	62.35	273.59	947.09
11	Harris	60241.0	Deer Park	77536	TX	29.687657	-95.120100	33891	10738	6241	26.92	51.02	243.82	966.82
12	Harris	60241.0	Houston	77046	TX	29.733181	-95.431310	2312717	486293	6000	30.39	57.31	239.71	920.71
13	Harris	60241.0	Pasadena	77505	TX	29.650492	-95.146320	153887	35665	7500	20.61	37.30	266.24	963.74
14	Montgomery	77598.0	Conroe	77304	TX	30.331460	-95.507030	84378	20358	10000	17.80	31.10	282.70	1228.20
15	Smith	56848.0	Tyler	75712	TX	32.411237	-95.289903	109000	35172	9396	25.70	43.68	160.97	523.94
16	Tarrant	66059.0	Colleyville	76034	TX	32.885062	-97.149230	26674	10370	17942	34.26	54.88	254.14	875.47
17	Tarrant	66059.0	Fort Worth	76107	TX	32.738481	-97.384240	874168	253704	7179	26.85	46.33	244.77	754.10
18	Tarrant	66059.0	Grapevine	76099	TX	32.771419	-97.291484	53982	14788	12072	25.92	45.67	203.40	795.90
19	Williamson	87817.0	Round Rock	78681	TX	30.518975	-97.714390	123678	35190	7950	35.80	48.60	249.51	669.51

20 cities fullfill the filtering parameters

K means to seek similar cities.

Cluster 0 incomes, population intervals are really far from each other, also have the cities with the most expensive water prices.

Cluster 1 incomes are between USD 60k, it has average water prices and the highest populations.

Cluster 2 have medium-high incomes, heterogenous intervals regarding population and the cheapest water prices.

Chosen Cluster 2 to search potential clients

Table 2. Clustering using k-means based in cluster label after fitting, County household income, city population and commercial water cost for 50k gallons.

Labels	Income	CityPopulation	50k Gal
0	73185.545455	8.643036e+04	260.376364
1	62046.000000	1.509320e+06	238.773333
2	69935.166667	7.660400e+04	180.098333

Potential Customers using FourSquare API

Potential customers

Filtered as restaurants and food and beverage stores

Radius: 5km

Coppell City had the most clients

Table 3. Texas cities resumed information in the chosen cluster after k-means segmentation.

	County	City	County Income	Population	Water Cost	Clients Nearby
0	Dallas	Coppell	59838.0	41941	213.94	87
1	Denton	Carrollton	88384.0	135710	158.92	37
2	Tarrant	Grapevine	66059.0	53982	203.40	37
3	Galveston	Friendswood	71959.0	39839	189.95	18
4	Comal	New Braunfels	76523.0	79152	153.41	16

Potential Customers in Coppell City using Folium Library.

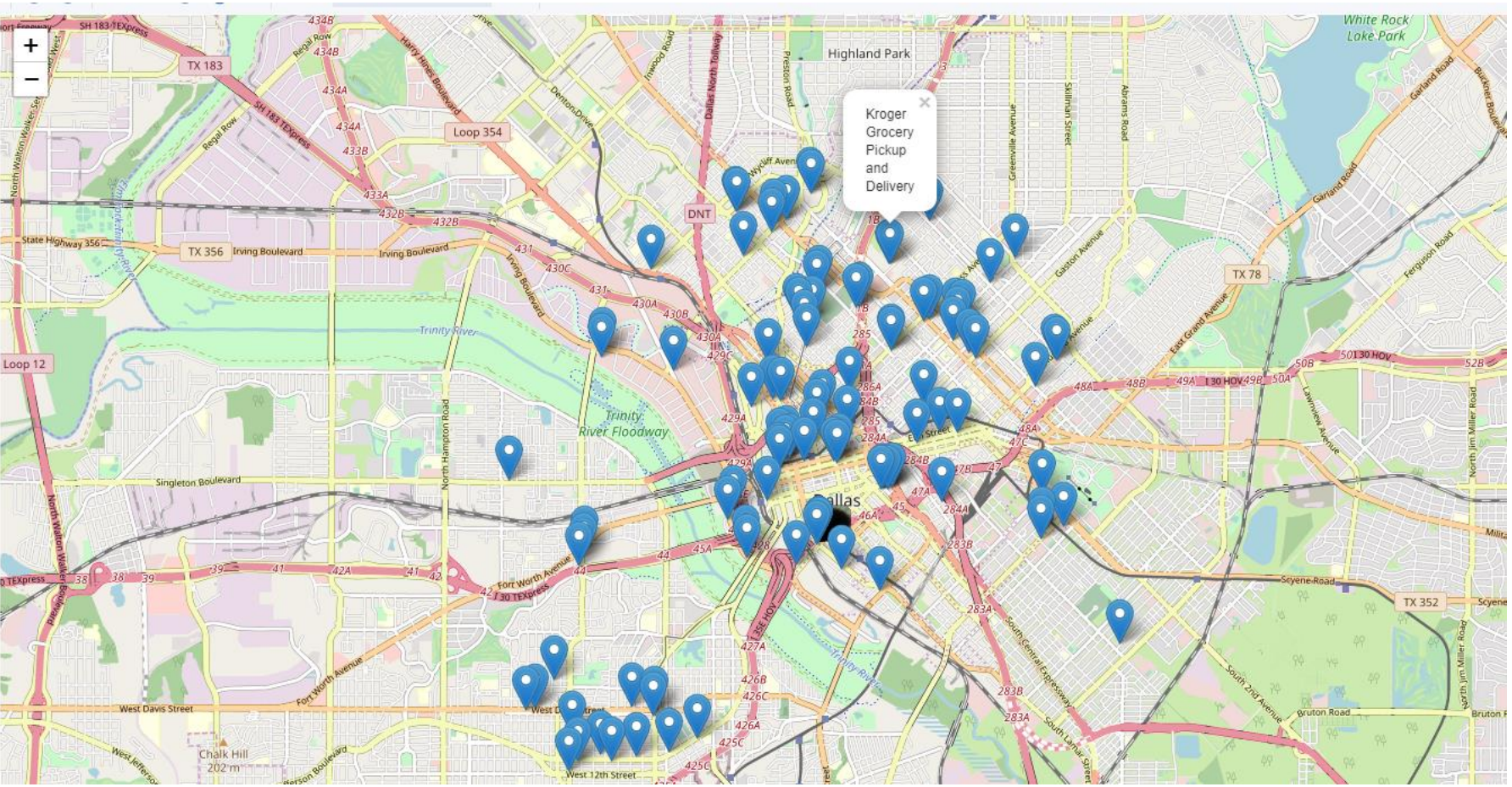


Figure 1. Potential Client Distribution in Coppell City, Dallas, TX in a 5km radius.

Conclusions

Coppell City was the city that meet all the parameters applied regarding water cost, average county income, city population and potential clients proximity.

Carrellton, Grapevine, Friendswood and New Braunfels could be other potential cities to consider establishing agricultural urban projects