Texas Cities Segmentation for Urban Agriculture Projects

I. Introduction

Background

Today's worldwide agriculture have a great dilemma. An increasing population over time, less usable land and hard-to-predict climate conditions. Due to this researches, public and private organizations, companies and farmers have to made changes on the traditional agricultural practices, leading in a continuously changing new paradigm of how agriculture is made. For example, in the new paradigm control environments and implementing techniques such as hydroponics/Aeroponics have been greatly accepted and has let farmers cultivated in less land and maintaining, or even, increasing their yields.

Problem

Most of products tend to come from countryside areas and is transported to cities, which leads to contamination, transport logistic problems, products quality loss and added charges which the customers pay at the end of the supply chain.

Interest

Segmenting cities could be a great start-point to embrace new profitable, efficient and ecofriendly urban agriculture projects that adapts to its particular food security strengths, problems, necessities.

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.

II. Methodology

Texas Water Cost data

To analyze the water costs in Texas cities we will use a dataset by the Texas Municipal League (https://www.tml.org/), this dataset are the results of the 'Water and Wastewater Rate Survey' for 2019 and we will use the 'Residential and Commercial Water Costs Details'. You can access the original database and more of their results here: https://www.tml.org/229/Water-Wastewater-Survey-Results

Texas Household Incomes by County data

We will use each county Median Household Income of each Texas county. 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

Texas cities ZIP Codes and coordenates data

We will import the Texas Zip Codes for location segmentation and to perform the client search, the data set 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.

The zipcode data does not have County information and will be needed to link the income data, water cost per city and each cities coordenates . It was used Wikipedia´s data wich contains some Texas cities and its belonging County(s). You can access data here: https://en.wikipedia.org/wiki/List_of_cities_in_Texas

Data Cleaning and merging

Data was imported to jupyter notebooks using the Pandas library, afterwards each dataset was cleaned: changing datatypes, column names, dropping NaN values.

Exploratory Analysis

After cleaning a exploratory analysis was made using the df.describe method and visualization using seaborn barplots.

Data Filtering

Each dataframe was filtered that meet the following criteria after the exploratory analysis:

- Water Cost: Selected all cities below the average water cost for commercial use of 50k gallons.
- County income: Selected all counties that surpassed the 3rd quartile with the highest incomes.
- Wikipedia cities by county: Drop all the cities wich were between more than one county (Duplicate cities) keeping just the first city listed.

Resulted Dataframe

After merging the NaN values was drop. The resulted dataframe had a shape contained:

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

Clustering cities using Kmeans.

A clustering of cities where performed using Sklearn library Kmeans. A Total of 3 clusters where made taking into consideration the following data:

- County.
- City.
- County Income.
- City population.
- Water cost for residential use (per 5k and 10k gallons per month).

After the segmentation a cluster that fit the best the initial data filtering criteria was used to perform a client search and mapping.

Potential Customer search

Cities venues belonging to the clustered that met the filtering criteria were searched using the Foursquare API. Only venues catalog as Food and beverages store and Food places where search in a 5 km radius from the city coordinates. The city venues with more potential customers near was map using folium library. Finally, a dataset filtered by cluster selected cities was made and contained the following summed parameters:

- County.
- City.
- County Income.
- City population.
- Water cost for residential use (per 5k and 10k gallons per month).
- 50 k gallons of water cost for commercial user.
- Potential clients in 5km radius.

III. Results

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

Table 1 shows the 20 cities around Texas that has a water cost under the average ($280), a city population over 20.000 citizens and a county income over the 3rd quartile of all Texas counties ($56. 500).

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

According with table 2, Cluster 0 is a more heterogenous cluster, incomes, population intervals are really far from each other, also have the cities with the most expensive water prices. On the other hand, cluster 1 incomes are between USD 60k, it has average water prices and the highest populations. Finally, cluster 2 have medium-high incomes, heterogenous intervals regarding population and the cheapest water prices. Based on these information Cluster 2 looks like the best choice because have a balance between all our parameters.

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

According to table 3 Coppell City in Dallas County could be a great city to establish the Urban Farm beacuse it have a >20.000 population, Water Cost under the average in Texas a County Average Income over 75% of other Texas Counties and also have 87 potential bussiness clients categorized as Groceries Stores and Restaurantes nearby in a 5 km radius. In spite Carrollton in Denton Conunty and Graprevine in Tarrant County also met all previous requirements but did not have far less ¿ potential clients nearby as Coppell who almost doubled both cities. Due to this, Coppell City potential clients was mapped, as observed in figure 1, to established

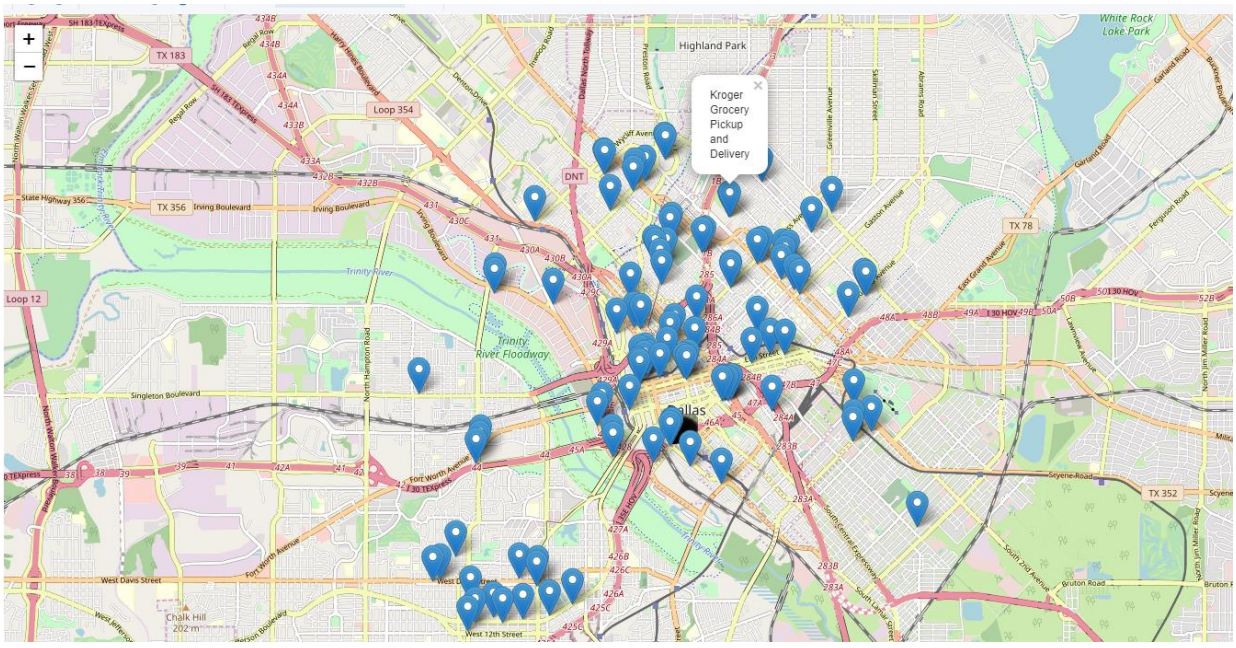

Figure 1. Potential Client Distribution in Coppell City, Dallas, TX in a 5km radius.

IV. Conclusions

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

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

Acknowledgement

This project was made by David Campos as part of the IBM Data Science Specialization Final Capstone Project in Coursera, you can find this project step-by-step Jupyternotebook in my Github.