Predict House Values for Denton County

Coursera Capstone

Business Case

Background

- **Every Year Denton County appraised values of all the houses within its jurisdiction. A homeowner needs to pay property taxes which are 2.26% of the appraised value of his/her house.**
- **Example: If my house is appraised at $300,000 then I will pay $300,000*2.26% = $6,780 annual taxes.**

Problem Statement

If the appraised value of the house increases by $20,000 from 2018 to 2019 then the homeowner will have to pay an additional $450 in taxes. This is extra money that he may not have planned for in his budget.

Target Customer

- **Home owners whose appraised value has increased in 2019 and they think that their appraised value should be lower than what the county says. They need supporting data on the value of similar homes nearby to argue their case with the county.**
- **Target customers can also be home owners who want a tool to predict what their home value would be in 2019 so they can plan for it ahead.**
- ***Project Phase 2 - Use FourSquare API to build a tool that can give prospective homes to buyers based on criteria such as: area, nearby venues, prices, distance to schools, school ratings etc. For this capstone we will not add school ratings database to the list.***

Data Sources needed

1. Denton County open dataset “County Parcels”

Link: http://data1-dentontxgis.opendata.arcgis.com/datasets/county-parcels/data

Sample: (replaced owner names and address with ****. Only a few fields included here.)

OBJECT ID	LAND SQFT	LIVING AREA	SITUS	OWNER NAME	ADDR_LINE2	CITY	STATE	ZIP	LEGAL_DESC	YR_BLT	APPRAISED	TOTAL LAND	TOTAL PROP
1	15008	1547	****	****	****	KRUGERVILLE	TX	76227	TIMBER MESA BLK C LOT 5	1986	131976	25656	131976
2	21710	0	****	****	****	KRUGERVILLE	TX	76227	COUNTRY ESTATES 2 BLK J LOT 3	0	3039	3039	3039

2. Four Square API data: Currently its location data is the most comprehensive out there, and quite accurate that it powers location data for many popular services like Apple Maps, Uber, Snapchat, Twitter and many others, and is currently being used by over 100,000 developers

Note: We will be using only the 1st 10,000 records from the Denton County database due to the time it takes to parse through the database. 2 minutes per 10k records.

Pre Processing Data

Convert the Denton County json file into a Pandas DataFrame with only the features we need for our study.

	Sub divison	appraised	type	city	land	Subdivison_2	area	year	zip	Latitude	Longitude
9995	MORNINGSIDE 1	168499	Real, Residential, Single Family	CARROLLTON	7548	MORNINGSIDE 1 BLK 2 LOT 15	1684	1980	75006	32.9894	-96.8705
9996	RIVER OAKS ADDN PH 3	182000	Real, Residential, Single Family	CROSSROADS	7607	RIVER OAKS ADDN PH 3 BLK F LOT 11	3063	2001	76227	33.1434	-97.1156
9997	O T JUSTIN	19500	Real, Vacant, Commercial Lot	FORT WORTH	6500	O T JUSTIN BLK 28 LOT 3 & 10' OF ALLEY	0	0	76117	33.0866	-97.2912
9998	WM. LOVE	7	Real, Farm and Ranch, Open Space Qualified Land	DALLAS	5009.4	A0728A WM LOVE, TR 1F, 0.1150 ACRES, (NWISD)(C...	0	0	75287	33.0679	-97.2234
9999	SOUTHRIDGE	237713	Real, Residential, Single Family	DENTON	16950	SOUTHRIDGE BLK 30 LOT 28	3254	1973	76205	33.1874	-97.1218

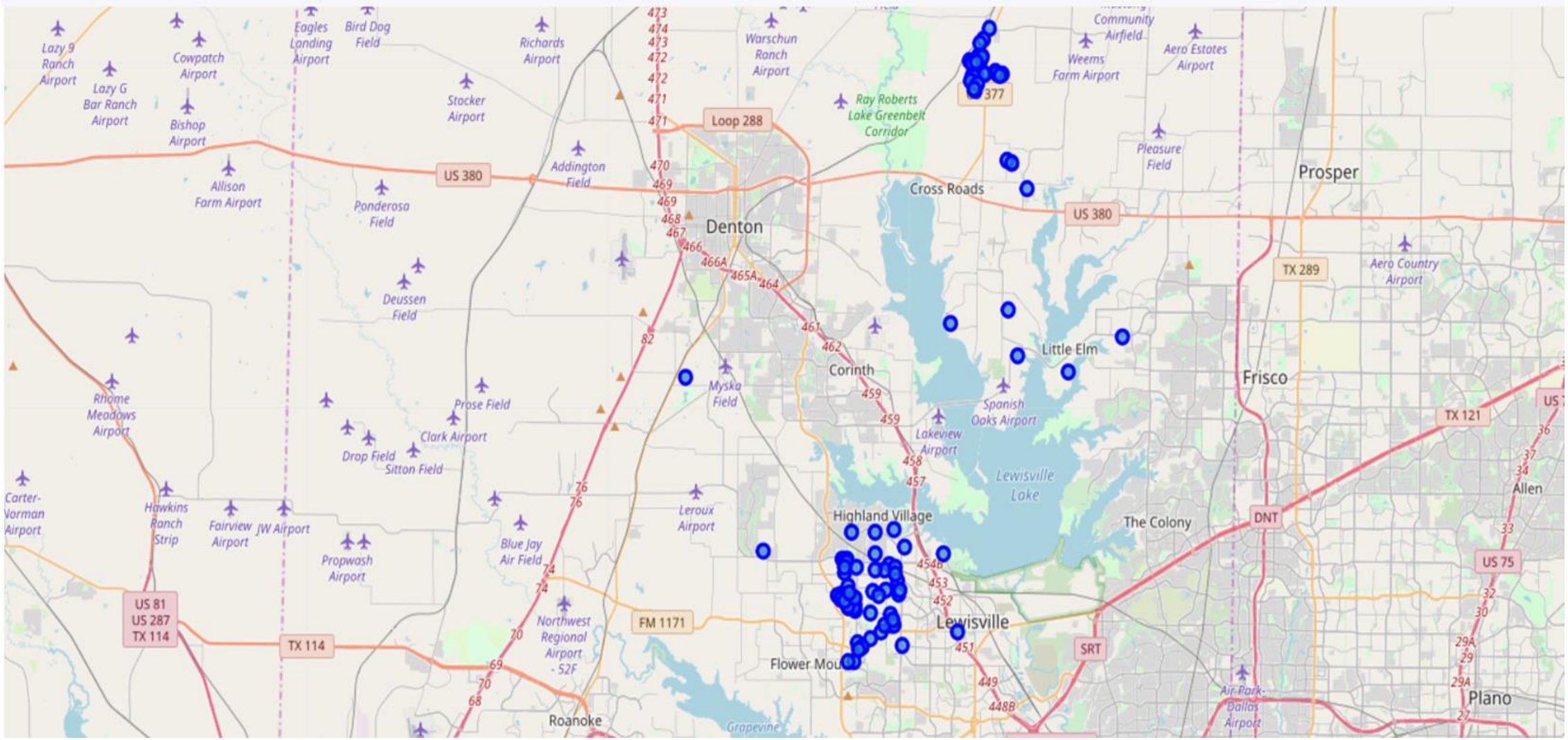
to numeric:

Remove records not needed:

Records with only land & no living area constructed (flag: living area = 0)
Records which are rural, commercial, multi family etc. (flag: type != Single family)

Feature Engineering: Add a feature age to the data frame.
Formatting: Convert columns such as appraised, area, land, latitude, longitude to numeric.

Map of Denton County



Scatter Plots

- Strong positive relationship between Living Area and Price/value of house
- Weak positive relationship between Land Area and Price. This could be due to the fact that most houses in the Sub division have similar land area.
- Age of House does not seem to have a correlation with Price

Predict House Values: Multiple Regression

Tools/Algorithms

- **Multiple Linear Regression**

Sample

- **Sample dataset**
- Lewisville Valley Subdivision
- **Sample size**
- 559 records

Attributes

- **Features (X):**
- Age of House, Living Area, Land Area
- **Response (Y):**
- Price

R-sq. = 61%

Next Steps: Try KNN Regression & compare the results

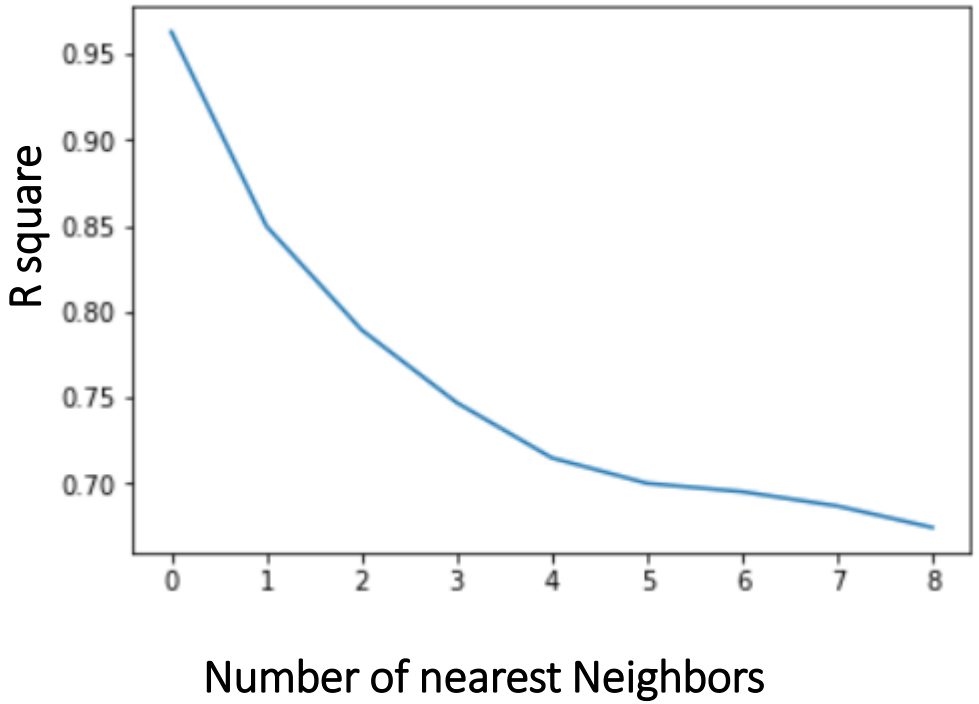
Predict House Values: KNN Regression

Tools/Algorithms	Sample	Attributes
• **KNN Regression**	• **Sample dataset** • Lewisville Valley Subdivision • **Sample size** • 559 records	• **Features (X):** • Age of House, Living Area, Land Area • **Response (Y):** • Price

R-sq. = 79%

KNN results look better at k =3

Selecting the right k for KNN



Elbow at k =3. Let us use k = 3 for our model.

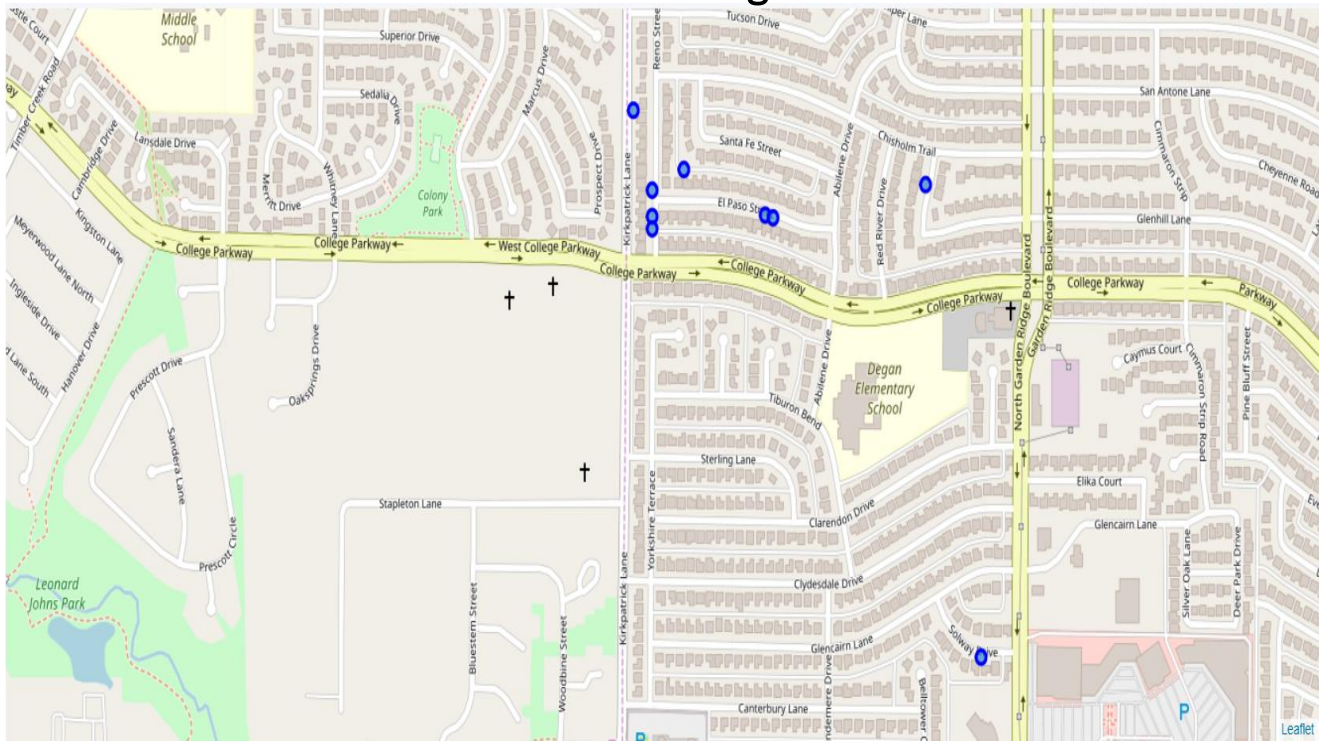
Predictions

If a homeowner in the Lewisville Valley area has a 40 year old house with 1800 sqft. of living space and 7000 sqft. of total land area, they can use the tool to predict the value of their house.

If the county appraises their house within the 147,000-157,000 range then they should be good. But if the house gets appraised more then they can file a dispute.

Land Area (sqft.)	7000
Living Area (sqft.)	1800
Age	40
Multiple Regression	147,600
KNN	156,800

Nearest Neighbors Plot



Explore Venues around the neighbors

Cosmetics Shop	336
Park	204
Gym / Fitness Center	168
American Restaurant	152
Disc Golf	150
Ice Cream Shop	142
Donut Shop	130
Health & Beauty Service	127
Restaurant	115
Coffee Shop	110

10 most common venues around the Lewisville area

Discuss Findings

The model can be used to assist home owners with

1 – Review the value of their houses. If they feel that the appraised value of their house by the county is too high/low we can use the model to not only assess it but also pull data on nearest neighbors.

2 – Help the homeowners predict the value of their houses for the coming year so they can plan for the change in house taxes

3 – Help home buyers decide the neighborhood they want to purchase a house in based on their criteria: Living area, age of the house, total land area and nearby venues.

Conclusion

We can make this analysis more robust by adding additional information such as:

1 – Assigned Elementary/Middle and High schools & their ratings
2- Year on Year change in home prices
3 – Listings of houses on sale
4 – Sale prices of all houses sold in the year

This will become an effective tool for home buyers, home sellers, home owners who want to dispute the appraised value of their houses & a tool for planning property taxes for the next year.