



HousePy

Analysis of Houston's Housing Market (2015-2019)

Group 5.

Iretemi Adelaja

Aline Jaimes

Sanjar Rasuly

Elizabeth Shaw

Date: September 11, 2019

Data Analytics Bootcamp, Rice University

Project Summary

Goal: Exploratory analysis of the Houston's housing market from September 1, 2015-August 31 2019

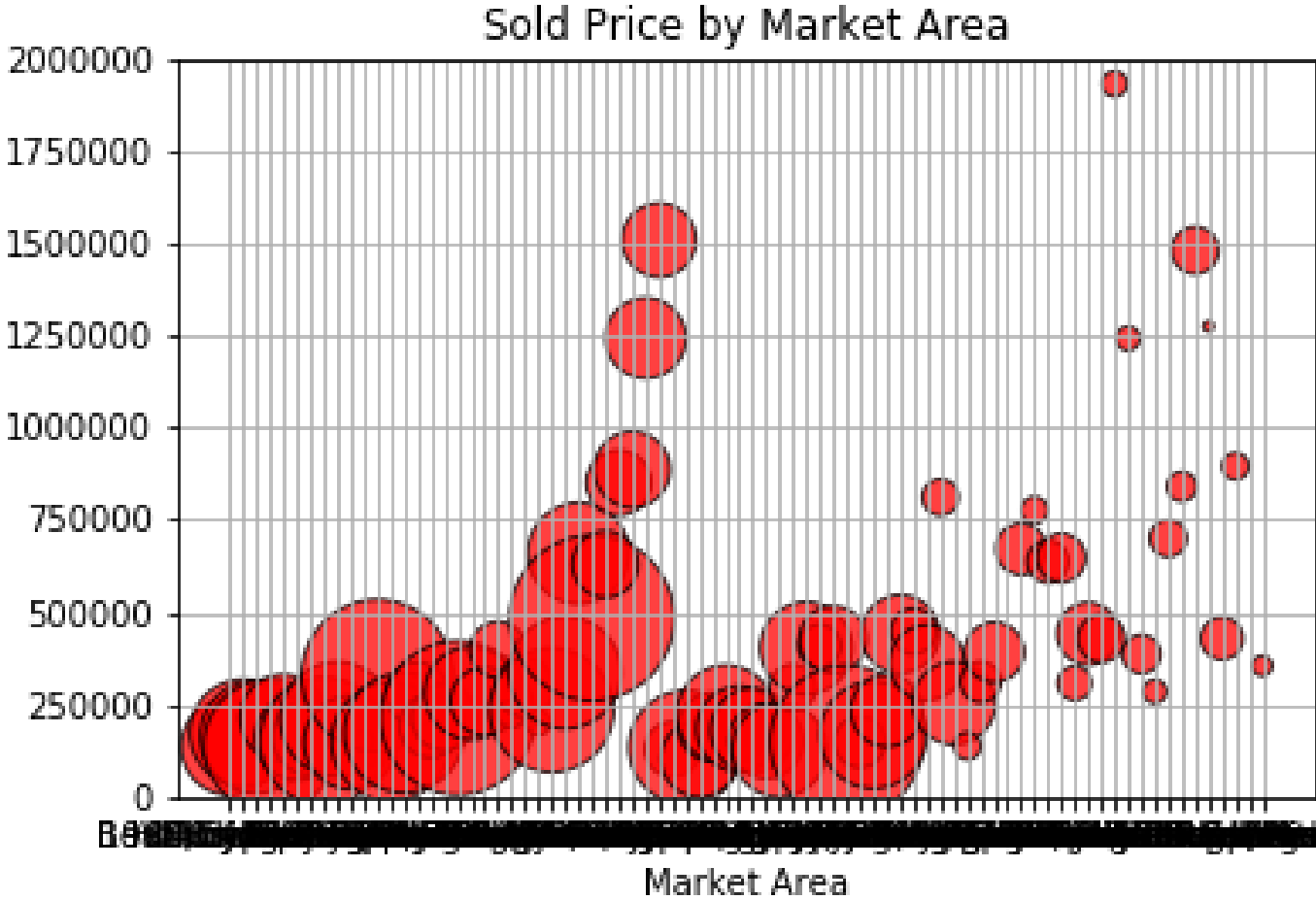
Questions:

1. Which subdivisions are **appreciating/depreciating** and undergoing **gentrification**?
2. What areas are considered **affordable** for average income families?
3. Which market areas have properties with more/less **days in the market?** and what are the **implications** of it?
4. Does **school district** impact market price?

Data Source: Private database site of [matrix.mls](#) at [Har.com](#). Contains record of every realtor-assisted transaction for single-family homes in Harris County.

Area of interest: Harris county, Cities: {Houston, Spring Valley, Hedwig Village, Hillshire Village, Bunker Hill Village, Piney Point Village, Hunter's Creek Village, Bellaire, Southside Place, West University Place}

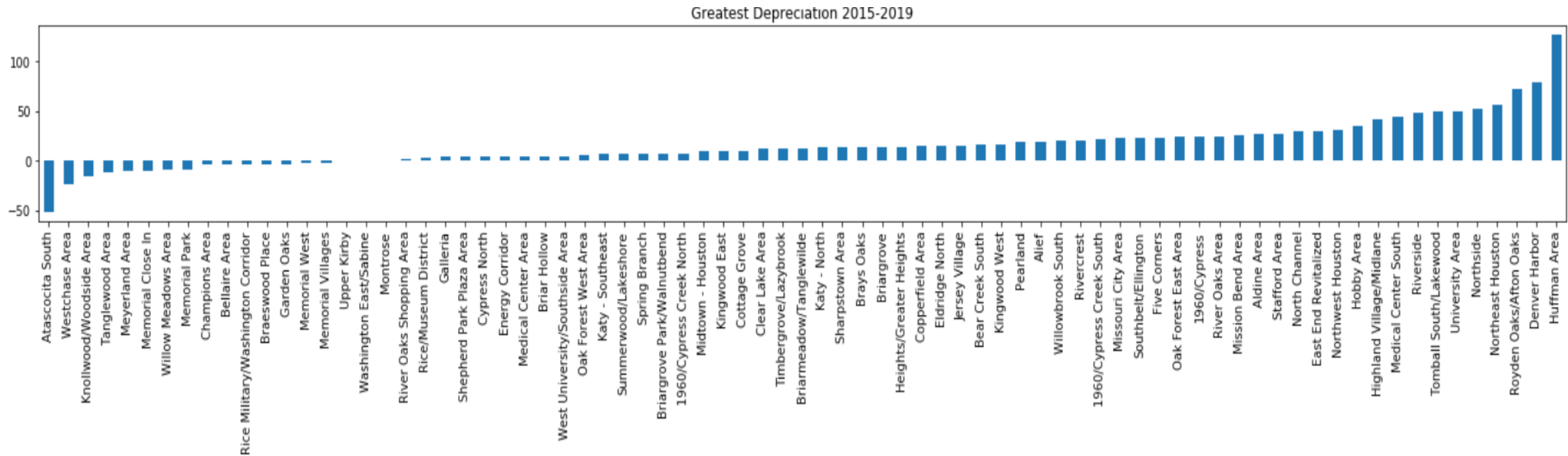
Data Analysis: 1 . Which market areas are appreciating/depreciating?



	Median Sold 2015	Median Sold 2019	Percent Change
Market Area			
Huffman Area	140250.0	317804.5	126.60
Denver Harbor	84500.0	151000.0	78.70
Royden Oaks/Afton Oaks	722500.0	1247500.0	72.66
Northeast Houston	79995.0	125000.0	56.26
Northside	101700.0	155000.0	52.41

	Median Sold 2015	Median Sold 2019	Percent Change
Market Area			
Atascocita South	565000.0	273225.0	-51.64
Westchase Area	731200.0	552000.0	-24.51
Knollwood/Woodside Area	465000.0	389500.0	-16.24
Tanglewood Area	1665000.0	1477500.0	-11.26
Meyerland Area	410000.0	366117.5	-10.70

Data Analysis: Which market areas are appreciating/depreciating?



Above. The percent change in median home sold price from 2015 to 2019.

Data Analysis: Which market areas are gentrifying?

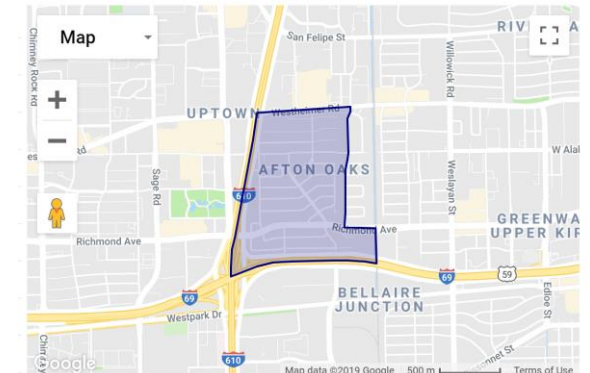
According to Zillow gentrifying is:

- Homes are built over 50 years ago.
- Bordered by two neighborhoods that are at least double median price.
- Transportation access.
- Not too many vacant lots.

Denver Harbor Boundary [View Homes](#)



Royden Oaks/Afton Oaks Boundary [View Homes](#)



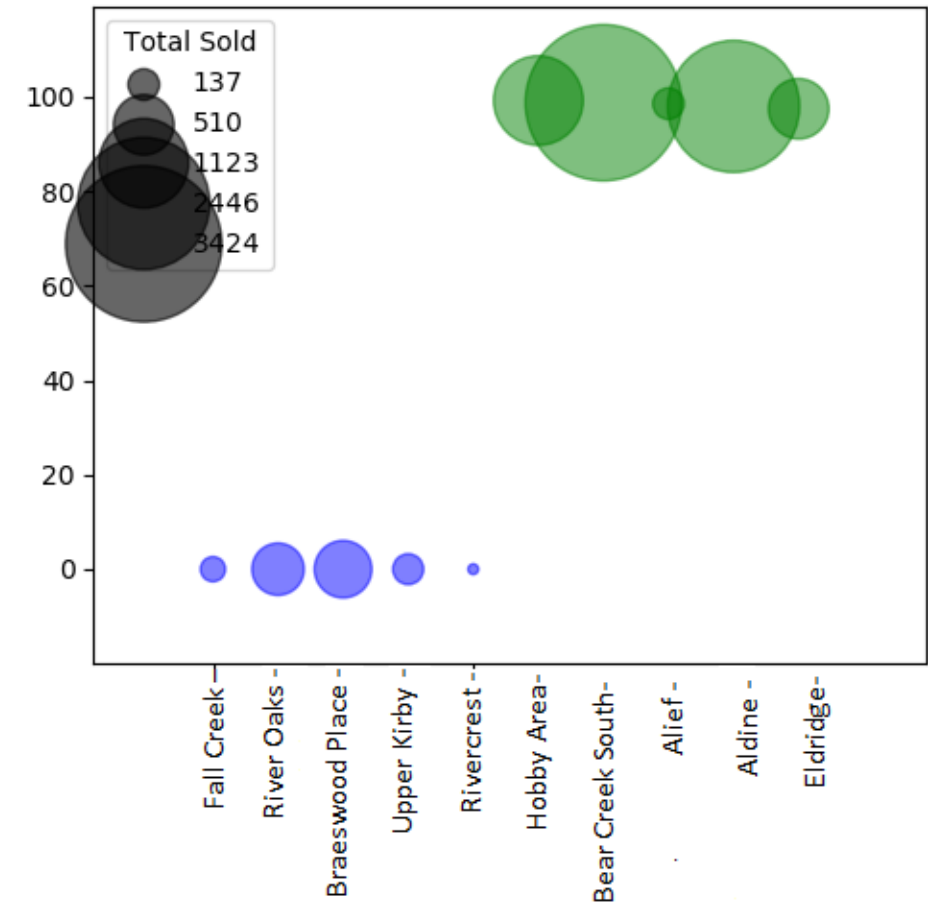
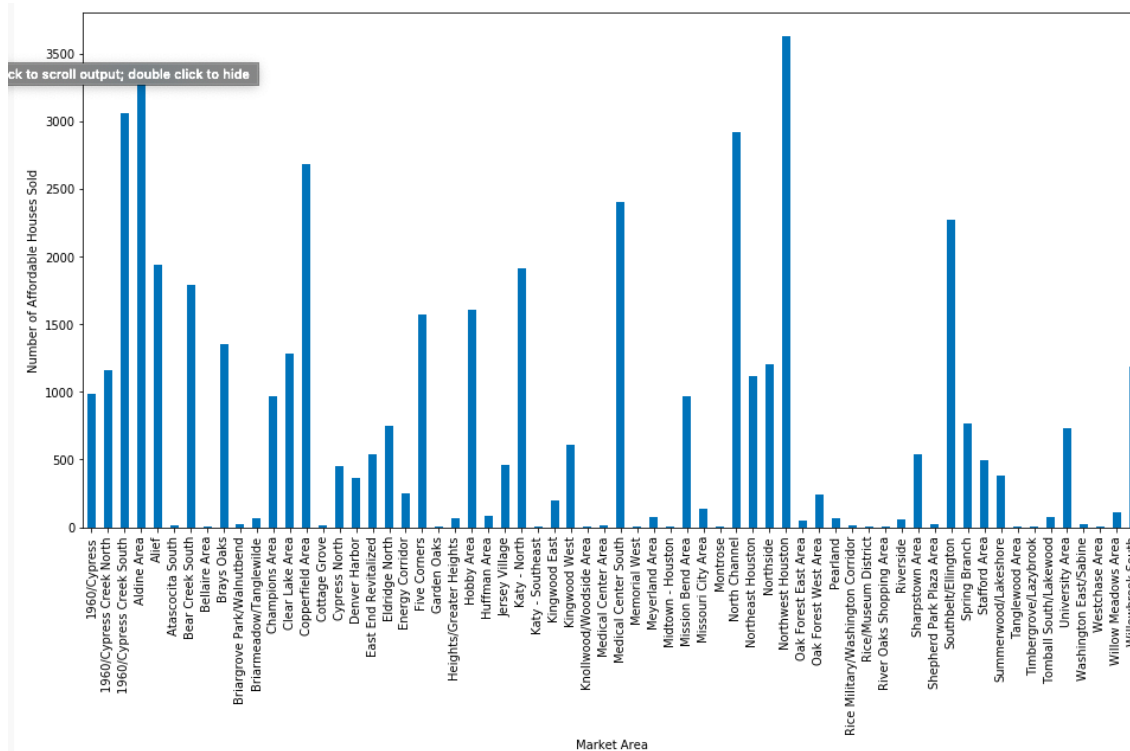
Northside Boundary [View Homes](#)



Market Area	Median Sold 2015	Median Sold 2019	Percent Change
Huffman Area	140250.0	317804.5	126.60
Denver Harbor	84500.0	151000.0	78.70
Royden Oaks/Afton Oaks	722500.0	1247500.0	72.66
Northeast Houston	79995.0	125000.0	56.26
Northside	101700.0	155000.0	52.41

Data Analysis. What market areas are considered affordable?

- Affordable Housing = median income x4.12 (Source: Zillow)
- The Median income of Harris County is \$57,791.00 (Source: According to census.gov)



Above. Top 5 areas with highest (green) and lowest (blue) percentage of affordability. Percentage of affordability calculated as number of sold homes with sold prices below the 4.12 * median income.

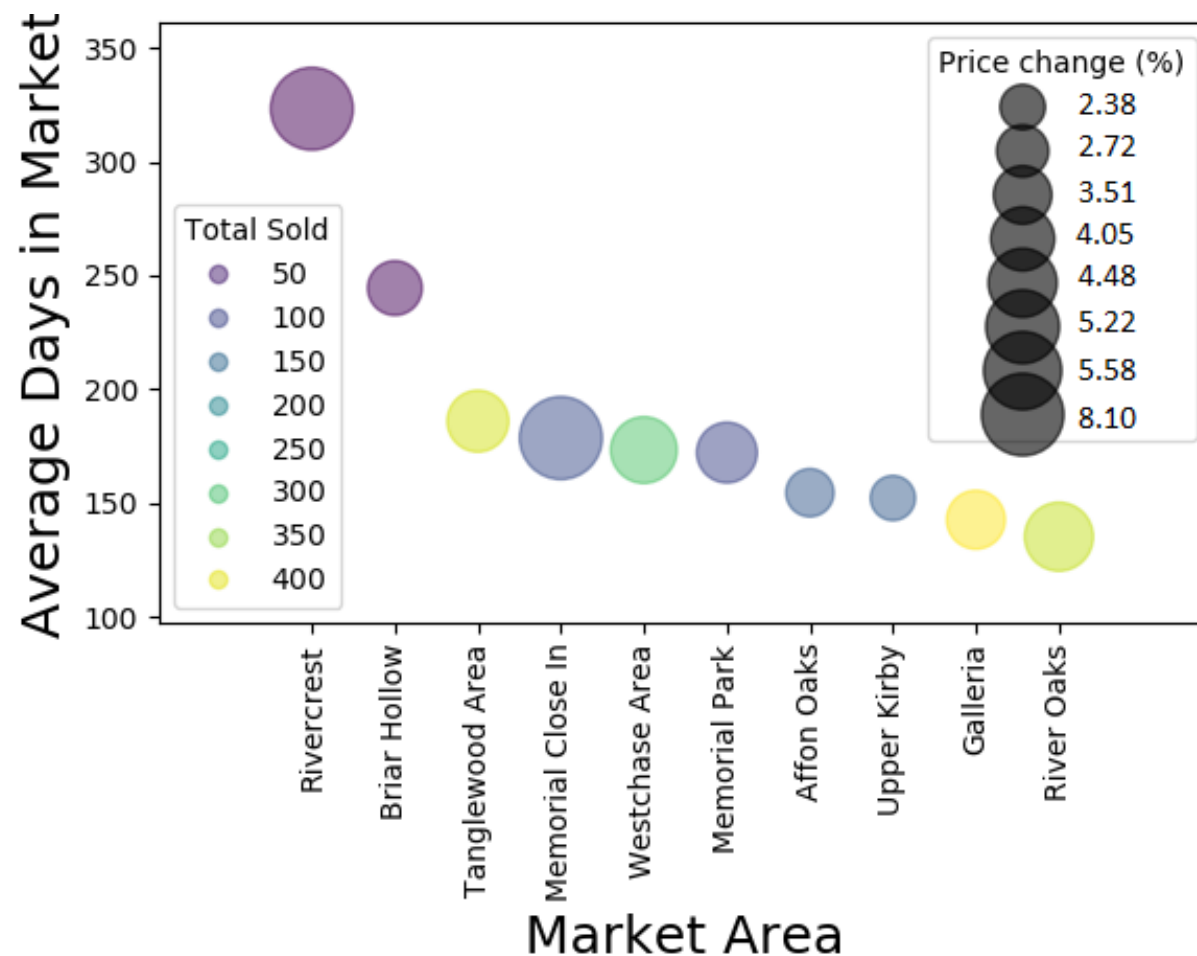
Data Analysis: Which market areas have properties with more/less days in the market?

Highlights

- Market areas with properties that remain in the market longer, reduce the total number of homes sold.
- The longer a property remains in the market, the largest percentage of price change from listed to sold.

Table. Summary of the top ten areas with lowest average number of days in the market.

Market Area	Avg Days in Market	Listed Price	Total Sold	% of Price Changed
Rivercrest	323.266667	3100000.0	15	8.06
Briar Hollow	244.428571	1425000.0	21	3.51
Tanglewood Area	185.917949	1562500.0	390	4.48
Memorial Close In	178.495495	2100000.0	111	8.10
Westchase Area	173.203279	670000.0	305	5.22
Memorial Park	172.079208	1295000.0	101	4.32
Royden Oaks/Afton Oaks	154.437500	920000.0	128	2.72
Upper Kirby	152.106061	799000.0	132	2.38
Galleria	142.562500	674800.0	416	4.05
River Oaks Area	135.099738	2200000.0	381	5.58



Above. Top ten areas with highest average number of days in the market.

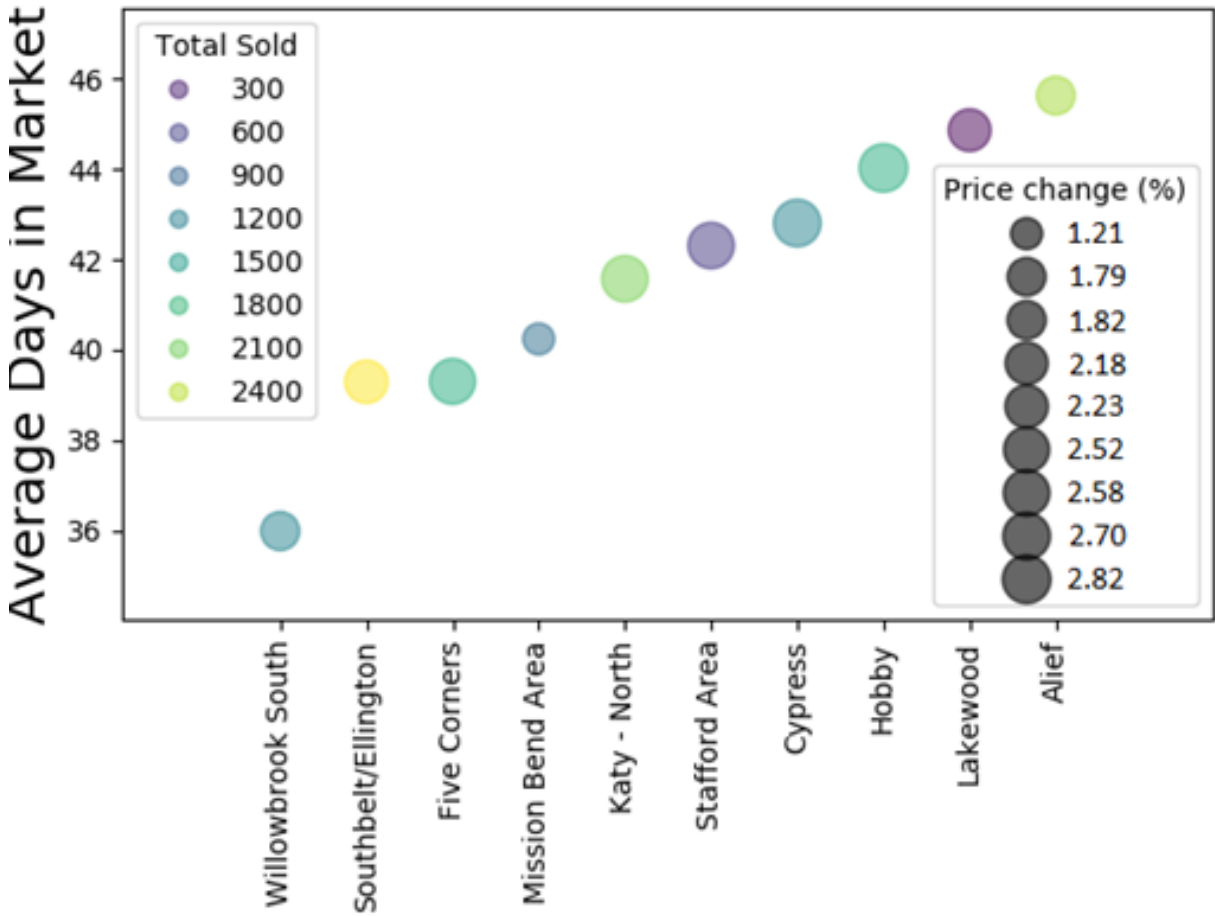
Data Analysis: Which market areas have properties with more/less days in the market?

Highlights

- Market areas with properties that sell quickly significantly increase the total number of homes sold.
- The percentage of price change does not change largely.

Table. Summary of the top ten areas with lowest average number of days in the market.

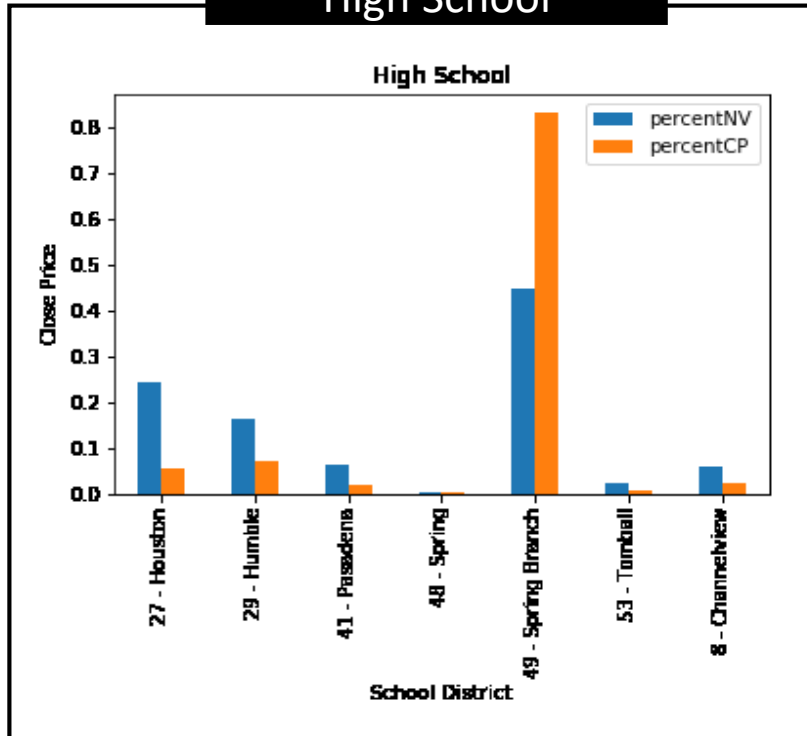
	Avg Days in Market	Listed Price	Total Sold	% of Price Changed
Market Area				
Willowbrook South	35.991235	164999.0	1255	1.82
Southbelt/Ellington	39.293478	179000.0	2668	2.23
Five Corners	39.302133	154900.0	1688	2.52
Mission Bend Area	40.241951	165000.0	1025	1.21
Katy - North	41.572965	174500.0	2138	2.58
Stafford Area	42.307843	155000.0	510	2.58
1960/Cypress	42.806041	185000.0	1258	2.70
Hobby Area	44.025354	139900.0	1696	2.82
Tomball South/Lakewood	44.861111	206000.0	108	2.18
Alief	45.627214	168000.0	2315	1.79



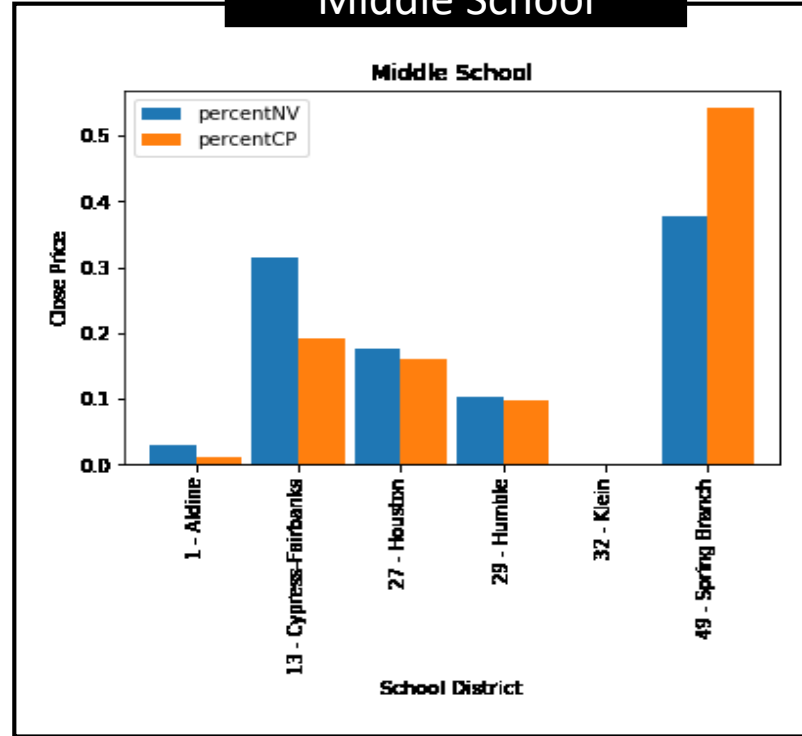
Above. Top ten areas with lowest average number of days in the market.

Data Analysis: 4. Does school district impact market price?

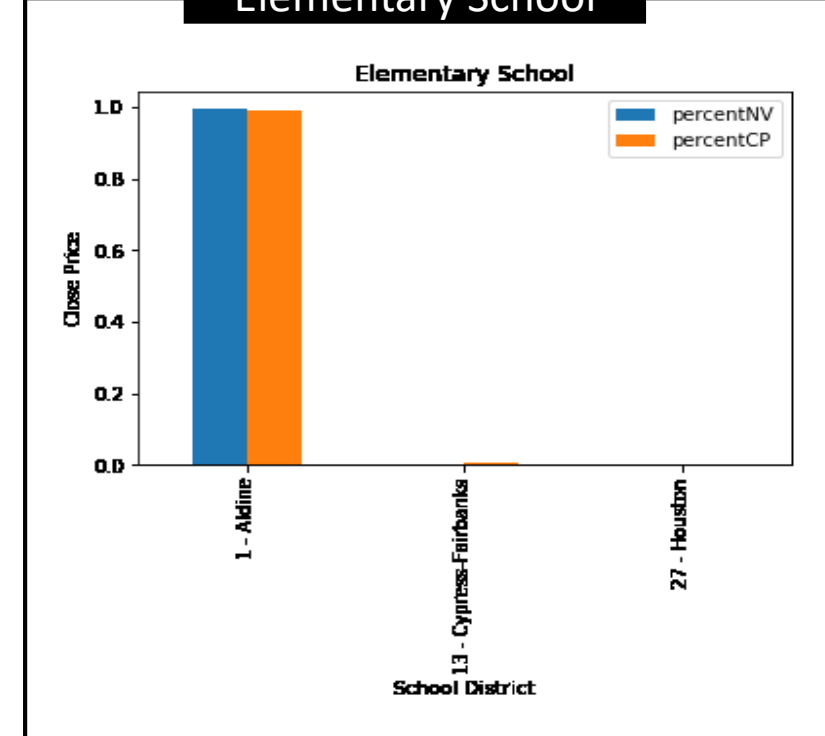
High School



Middle School



Elementary School



Above. Percent of homes sold in Harris County from 2015 to 2019. Based on percent of “Numerical Value” and percent of “Close Price”

Conclusions



Summary of Market Areas that are:

Appreciating

- Huffman Area
- Denver Harbor
- Royden Oaks/Afton Oaks

Fastest Sell, Less lost of value

- Willow brook South
- Ellington
- Harrison Bend Area
- Katy-North

Affordable

- Hobby Area
- Bear Creek South
- Alief
- Aldine
- Eldrige

Data Cleanup & Exploration

1. Dropped columns and datatypes were changed from object to integer by removing punctuation characters.
2. New columns were created to calculate sales over time.
3. When applying functions it was determined about 25% of pre-defined Market Areas by HAR had less than 10 sales for the four year data period. These Market Areas were dropped from the data as outliers.

1. Subset data frame
2. Identify/Convert data types
3. Exploratory plots
4. group the data by "School District" to identify top performing school
5. Created three random sample from the school districts
6. `#f_rand1 = df_schoolbyprice.sample`
7. Using the merge method, combined "Top school" with "Housepy"
8. `#df_Elementary_School_D_join = pd.merge(df_Elementary_rating, housepy_df, on="School Elementary")`
9. Performed the analysis using matplotlib

```
df_rand1 = df_schoolbyprice.sample(8)

print(df_rand1)
```

School District	
41 - Pasadena	490608073
29 - Humble	1019082411
23 - Goose Creek	239999
99 - Other	2179600
53 - Tomball	28033525
13 - Cypress-Fairbanks	3210190620
1 - Aldine	613422258
26 - Hitchcock	140000

Name: List Price, dtype: int64