GT-Project One
Data Junkies

11/1/2022

Background

- As a team we are trying to aid our first client, Brandyn, in his home search
- We are acting as a consultancy that is establishing a process to analyze housing data
- Our first step is to collect foundational knowledge about the housing market

The Data Junkies

Cathy Chang

Project Manager

Glenn Purcell

Github
Ambassador

Caroline Christy

Data Expert

Brandyn Jackson

Data Expert

Dylan Jackson

Presentation
Coordinator

Client Profile:

Brandyn Jackson

- **Basic wants and needs**
  - ITP (preferably)
  - Things to Do!!!
    - Entertainment
    - Restaurants
    - Nightlife
- **Key areas of interest:**
  - Alpharetta
  - Brookhaven
  - Buckhead
  - North Druid Hills
  - Old 4th Ward
  - Sandy Springs
  - Suwanee
  - Smyrna

How does the housing market relate to data analytics?

- Initial Assumption(s):
  - Low inventory contributed to the recent home sale boom the last two years. We feel there is a shift in the market that can be timed by the buyer to gain an advantage over the seller in negotiating the final price.

- Our Goal → Is inventory related to housing prices?
  - We want to help our client determine any correlation between low inventory and final sales price. Thus...will waiting for periods of high inventory allow for a lower sale price and increased equity??

Hypothesis Testing

Our Hypothesis	The Null	Our Alternative
• Is there a correlation between median sales prices and housing inventory levels?	• Increasing inventories has no effect on median sales prices at all or positive.	• Decreased inventory leads to higher median sales prices.

vs

$$H_0: p = 0$$

$$H_{MSP}: p \geq 0$$

$$H_{MSP}: p < 0$$

Analysis Approach

Identifying Data Source

- Flat file CSV was downloaded from Redfin.com
- 10 years of monthly home sale price and inventory data
- Selected 8 regions within Atlanta to analyze

Plot Everything

- Scatter plots based on inventory and median sales price
- Pairplots via the Seaborn package
- Coefficient correlations
- Sliced data by years and by regions

Test Hypothesis

- Looked for patterns and trends in the plots
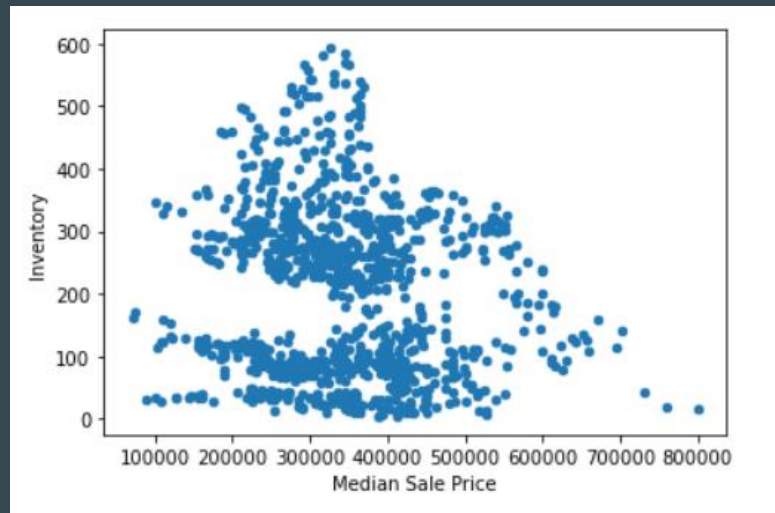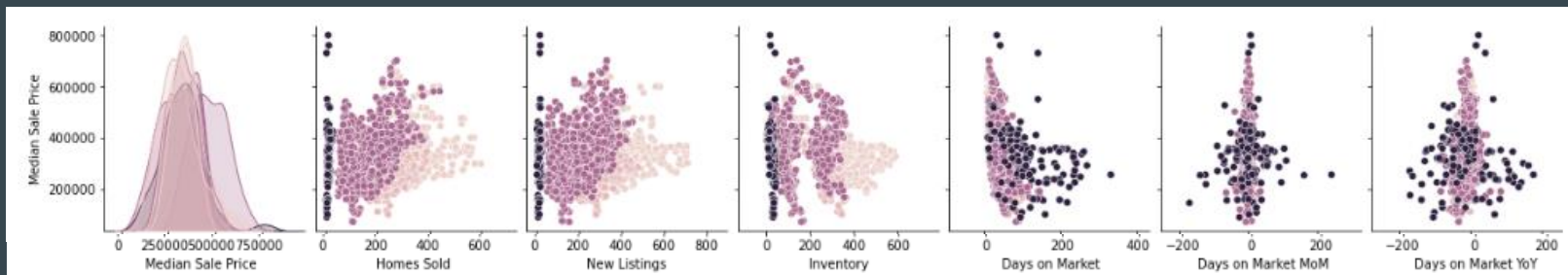- Evaluated null and alternative hypothesis

What the Data Shows

Initial Learnings

- **Data Quality**
  - Outliers were removed
  - Dataset was already manipulated to be average median data
  - This may make it harder to identify patterns/trends
- **Economic Factors**
  - House pricing data isn't adjusted for inflation over time

Data Sets: Scatter Plot

Data Sets: PairPlot

Data Sets: Correlation Coefficients - Overall

	Region	Median Sale Price	Homes Sold	New Listings	Inventory	Days on Market	Days on Market MoM	Days on Market YoY
Region	1.000000	0.028586	-0.801742	-0.769079	-0.671665	0.469361	-0.008603	-0.011169
Median Sale Price	0.028586	1.000000	0.177396	0.134965	-0.114685	-0.346852	0.033298	0.092914
Homes Sold	-0.801742	0.177396	1.000000	0.937816	0.781690	-0.435069	0.002485	0.017996
New Listings	-0.769079	0.134965	0.937816	1.000000	0.849334	-0.398315	-0.047753	0.024785
Inventory	-0.671665	-0.114685	0.781690	0.849334	1.000000	-0.157612	-0.006999	0.033668
Days on Market	0.469361	-0.346852	-0.435069	-0.398315	-0.157612	1.000000	0.228128	0.286618
Days on Market MoM	-0.008603	0.033298	0.002485	-0.047753	-0.006999	0.228128	1.000000	0.302177
Days on Market YoY	-0.011169	0.092914	0.017996	0.024785	0.033668	0.286618	0.302177	1.000000

Data Sets: Correlation Coefficients - Alpharetta vs. Brookhaven

Alpharetta

	Region	Median Sale Price	Homes Sold	New Listings	Inventory	Days on Market	Days on Market MoM	Days on Market YoY
Region	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Median Sale Price	NaN	1.000000	0.068856	-0.059854	-0.756464	-0.755697	0.052495	0.252741
Homes Sold	NaN	0.068856	1.000000	0.660295	0.258507	-0.437519	0.049503	0.110173
New Listings	NaN	-0.059854	0.660295	1.000000	0.534306	-0.286747	-0.390677	0.085014
Inventory	NaN	-0.756464	0.258507	0.534306	1.000000	0.500273	-0.130299	0.031766
Days on Market	NaN	-0.755697	-0.437519	-0.286747	0.500273	1.000000	0.102705	-0.130780
Days on Market MoM	NaN	0.052495	0.049503	-0.390677	-0.130299	0.102705	1.000000	0.229401
Days on Market YoY	NaN	0.252741	0.110173	0.085014	0.031766	-0.130780	0.229401	1.000000

Brookhaven

	Region	Median Sale Price	Homes Sold	New Listings	Inventory	Days on Market	Days on Market MoM	Days on Market YoY
Region	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Median Sale Price	NaN	1.000000	0.294800	0.134524	-0.534855	-0.702080	0.051660	0.343769
Homes Sold	NaN	0.294800	1.000000	0.596960	-0.051858	-0.589161	-0.054648	0.196823
New Listings	NaN	0.134524	0.596960	1.000000	0.381186	-0.437181	-0.383347	0.204817
Inventory	NaN	-0.534855	-0.051858	0.381186	1.000000	0.418074	-0.042750	0.143149
Days on Market	NaN	-0.702080	-0.589161	-0.437181	0.418074	1.000000	0.182242	-0.190266
Days on Market MoM	NaN	0.051660	-0.054648	-0.383347	-0.042750	0.182242	1.000000	0.230842
Days on Market YoY	NaN	0.343769	0.196823	0.204817	0.143149	-0.190266	0.230842	1.000000

Data Sets: Correlation Coefficients - Alpharetta vs. Old 4th Ward

Alpharetta

	Region	Median Sale Price	Homes Sold	New Listings	Inventory	Days on Market	Days on Market MoM	Days on Market YoY
Region	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Median Sale Price	NaN	1.000000	0.068856	-0.059854	-0.756464	-0.755697	0.052495	0.252741
Homes Sold	NaN	0.068856	1.000000	0.660295	0.258507	-0.437519	0.049503	0.110173
New Listings	NaN	-0.059854	0.660295	1.000000	0.534306	-0.286747	-0.390677	0.085014
Inventory	NaN	-0.756464	0.258507	0.534306	1.000000	0.500273	-0.130299	0.031766
Days on Market	NaN	-0.755697	-0.437519	-0.286747	0.500273	1.000000	0.102705	-0.130780
Days on Market MoM	NaN	0.052495	0.049503	-0.390677	-0.130299	0.102705	1.000000	0.229401
Days on Market YoY	NaN	0.252741	0.110173	0.085014	0.031766	-0.130780	0.229401	1.000000

Old 4th Ward

	Region	Median Sale Price	Homes Sold	New Listings	Inventory	Days on Market	Days on Market MoM	Days on Market YoY
Region	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Median Sale Price	NaN	1.000000	0.291453	0.198791	-0.526394	-0.638600	0.087167	0.480364
Homes Sold	NaN	0.291453	1.000000	0.687846	-0.016290	-0.457862	-0.209067	-0.075675
New Listings	NaN	0.198791	0.687846	1.000000	0.287785	-0.393808	-0.292715	0.024197
Inventory	NaN	-0.526394	-0.016290	0.287785	1.000000	0.448275	0.012736	-0.189959
Days on Market	NaN	-0.638600	-0.457862	-0.393808	0.448275	1.000000	0.158043	-0.230031
Days on Market MoM	NaN	0.087167	-0.209067	-0.292715	0.012736	0.158043	1.000000	0.302926
Days on Market YoY	NaN	0.480364	-0.075675	0.024197	-0.189959	-0.230031	0.302926	1.000000

Conclusions

$$H_0: p = 0$$

$$H_{MSP}: \geq 0$$

$$H_{MSP}: < 0$$

- We are rejecting the Null Hypothesis of:
  - Increasing inventories and it has no effect on median sales prices.
- The absolute value of our Pearson coefficients are between 0.4 and 0.79, indicating Moderate to High Correlation

- Inverse correlation between inventory and median sales price
- Further research is needed
- Old 4th Ward and Brookhaven may be ideal places to begin the house hunt