Austin Housing

SMU Data Analytics

Greyson Moore

Thi Quang

Jonathan Arnold

Introduction/ Inspiration:

Austin, TX is one of the fastest growing cities in the U.S. Since 2010, the city has gained over half- a- million residents. This research project will look at the effects of this population growth on the housing market in the Texas capital. Using detailed statistics and visualizations, we will identify key areas in Austin where growth is occurring and its influence on pricing, sales, and other trends in the housing market.

Objectives:

Through this project, we want to answer the following questions: What area of Austin is the housing market growing in? How has the frequency of sales been affected by year? Have they increased or decreased? What type of houses are being purchased? And does the market favor investors/buyers or sellers?

Sources of Data:

The three data sources we utilized are Kaggle, U.S. Census and Google Maps API. Our primary data for Austin housing is from Kaggle, which is a CSV that contains information about the houses listed between 2018 to 2021, for the cities within and around Austin. Some of the information this data provided are the latitude and longitude coordinates, the type of homes, latest sale price and year, tax rate, features of the house: living area and lot size, number of bedrooms, bathrooms, stories, and parking spaces, if the house have HOA, cooling system, heating, or spa etc.

U.S. Census data is also a CSV file that provided us population information for 2010 to 2019 for all 50 states. The data breaks it down to individual counties for each state, population estimates, net population change, number of births and deaths, net migration, residual, and many other information for each year.

We also utilized Google Maps API to create our heat map by pulling the longitude and latitude of each listing. This allowed us to visualize the number of sold homes for each month and which area of Austin is the most desirable area.

Data Cleaning:

We began the process by cleaning up the data and identifying outliers. We did not drop any variables from the Austin Housing data as all the information is vital for our research. When looking at the data, we noticed 94% of the homes are single family homes and the remaining 6% are from condo, townhomes, vacant land, and others. We decided to focus on single family homes only as it makes up the majority of our data. Another area decided to focus on is Austin city, which makes up 99% of the data. Even though the data included homes sold from 2018 to 2021, 2021 only make up 7% of the data. Given the little amount of data from the incomplete year, we decided to eliminate homes sold in 2021. After finalizing our focus, we were able to pinpoint our outliers. We eliminated a house with 27 bathrooms and another with 1.5 million square feet in lot size.

With the Census data, we only wanted to focus on the variables that provide the population change from 2010 to 2019 for Austin. We narrowed down the data to only Texas and the main Austin counties. From here, we created a new dataframe which consists of the Net Migration from 2010 to 2019 for each county, and created a dataframe to see the Natural Balance of each county for each year.

Analysis:

We first wanted to look at the Census data to see if population is a factor that drives the sale price in Austin. As shown in the "Population 2010-2019" figure below, the population has had a steady increase between 2.1% and 2.4% annually since 2010. We wanted to look further to see if migration could be the cause of population increase and noticed there has been a positive net migration each year. The positive net migration means there is a higher ratio of immigration than emigration for that year. The data shows that the highest migration year is 2011 with 3.73% increase from the previous year. From 2018 to 2019, there's also an increase of 18% in net migration.
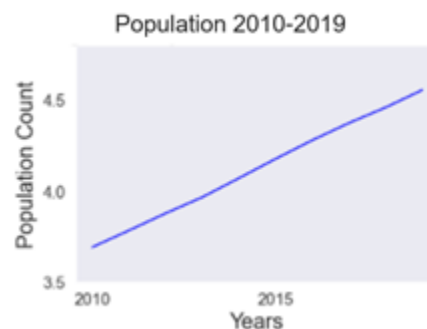

With the increasing number of people moving to Austin, we are also seeing an increase in the price of houses. There is an increase of 2% in the average sale price from 2018 to 2019 and a big jump of 10% from 2019 to 2020.

Factors affecting the market - The housing market in Austin is driven by a number of factors. We aspired to identify which factors had the greatest effect on the potential price of a home. First, we divided home sales regionally. Homes were evenly divided into Northwest, Northeast, Southwest, and Southeast regions.

Northwest: This region of Austin represents the most expensive and quality homes in the city. Median sales price in 2018 was $696,000, $735,000 in 2019, and $798,000 in 2020. This is a 14% increase in 2 years.

Northeast: The northeast area of Austin median house price increased from $371,000 to $443,000 from 2018 to 2020. This represents a 19.4% growth. This extra growth is likely from people who are outpriced in the Northwest but still want to live in north Austin, shown by this histogram which displays a right skew in housing prices.
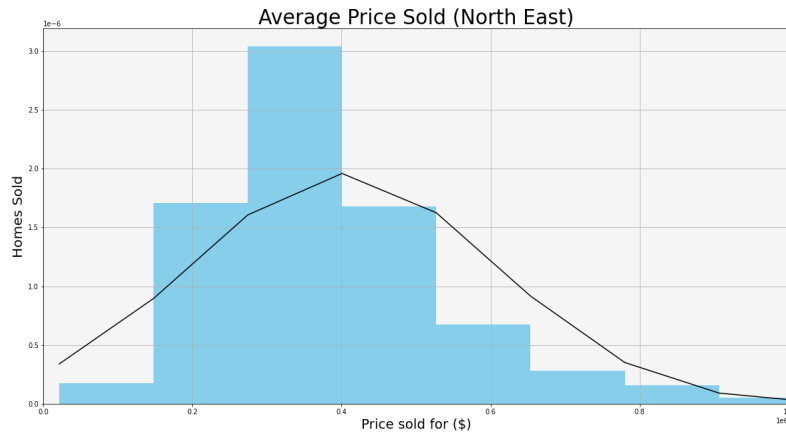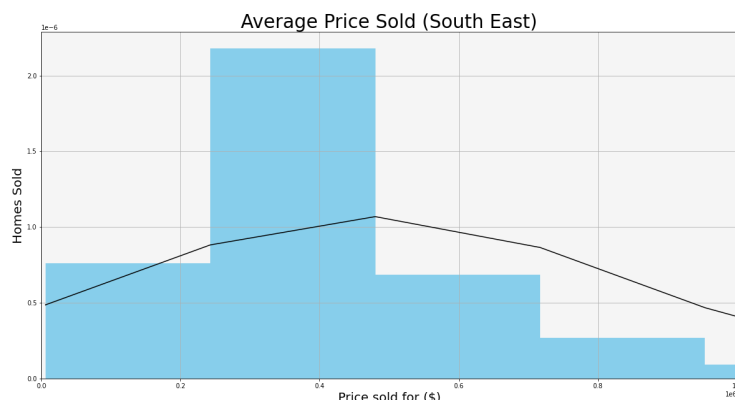

Southwest: This region experienced the least amount of growth of all four areas. From 2018 to 2020 prices only increased from $456,000 to $488,000, representing a 7% growth. This graph shows the increase in price over the four years.
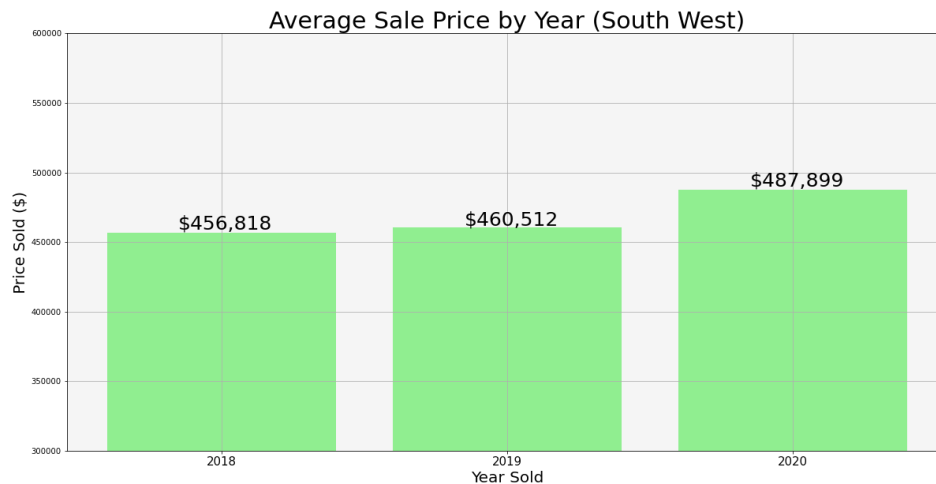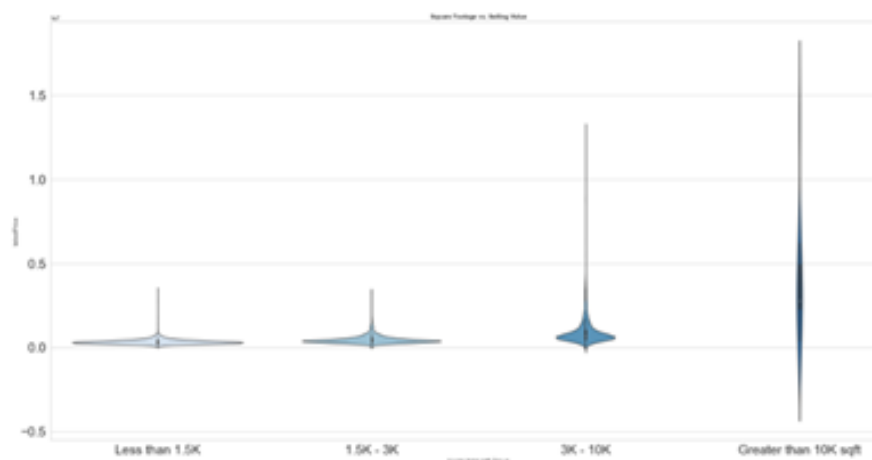
Southeast: Finally, the southeast region median house price swelled from $438,000 to $517,000.

This is an 18% increase. Similar to the Northeast, this histogram shows a right skew in house price.

By dividing the home sales by region, we can see huge increases in east Austin. Clearly, this is where people are moving to as the large growth in price reflects a demand for houses in these regions. Despite relatively small growth in the Southwest, all four regions increased year to year in median house price.

House features vs price – When comparing each feature of the house to the sale price, we can see that the square footage of the house and the number of features are significant to the sale price. The violin plot on the left is showing that there are many houses that are less than 3,000 sq ft in living area. As the square footage increases, so does the price. And the same applies to the lot size, number of bedrooms, number of bathrooms, etc.


Regression - Regression modeling allows us to compare a dependent variable and its relationship with multiple independent variables. The process returns an r-squared value, which

predicts the probability that any variance of the dependent variable is produced by the independent variables. We selected home price as the dependent variable. The independent variables included living area square feet, lot area square feet, property tax rate, as well as house features such as number of stories, bedrooms and bathrooms. Further, we divided Austin into four regions (mentioned previously in "Factors affecting the market"). Then we graphed each home's actual price versus its predicted price.

Northwest: Looking at the Northwest produced a r-squared value of 0.357. This means that 36% of price variance can be attributed to the independent variables.


Predicted vs Actual Price (North West)
R squared is: 0.3574394932463125
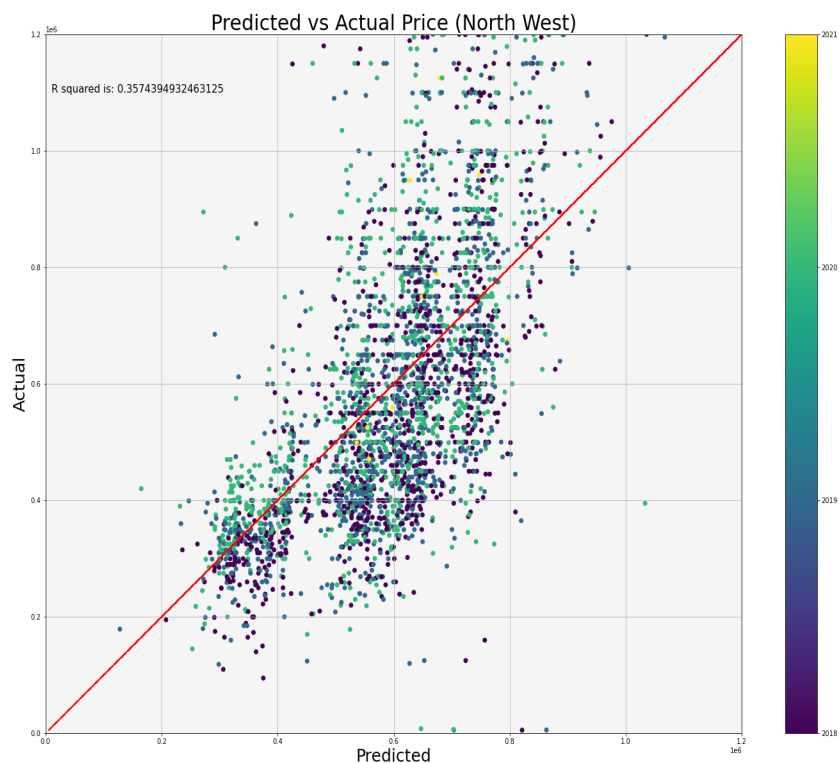
Northeast: This region shows a r-squared value of 0.249, so 25% of price variance is attributed to the independent variables.

Predicted vs Actual Price (North East)
R squared is: 0.24948302649699206
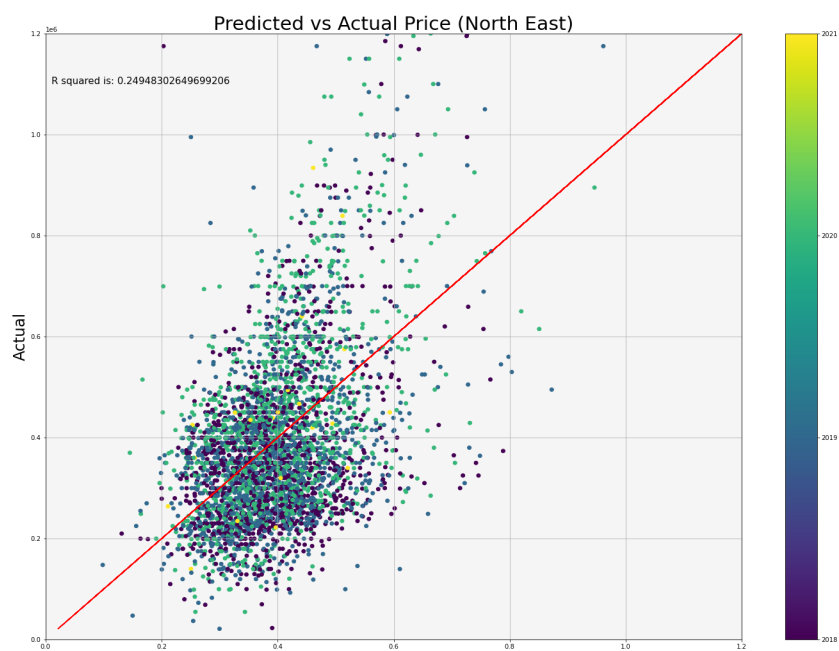
Southwest: The Southwest regression resulted in a r-squared value of 0.600, meaning that 60% of variance is a result of the independent variables.

Southeast: Finally, the Southeast regression model's r-squared value was .335, so price variance is attributed to 33% of the independent variables.
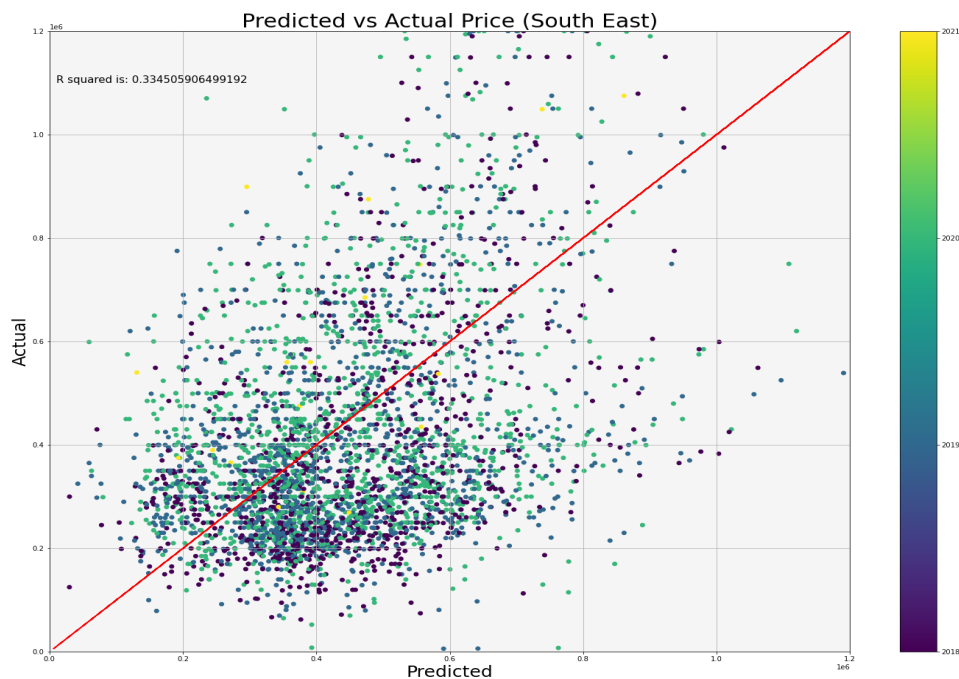
These regression models allowed us to draw several conclusions. First, none of the r-squared values are particularly high. We would have wished to see values closer to 1 which would allow us to more accurately predict a house's price. Values were especially low in eastern Austin. This is likely because of the substantial growth we have seen in this area. House prices are more difficult to predict as the market in that region has high demand. The high r-squared value of the southwest region follows a symmetrical pattern. House prices have increased the least in the area so prices are more steady and therefore easier to predict.

Actionable insights:

Buyer strategies - Our data shows Winter months are the ideal months for investors to make a purchase. As you can see from the "House Sales by Month" figure below, the number of sales is significantly lower than Summer months, particularly January. The best time to sell would be during summer. The number of sales peaked around June and July for all three years.

Seller strategies – Based on the data and analysis compiled in this study, we can propose several helpful strategies to home sellers in Austin, TX. One of the most important factors is the time of the home sale. Sellers should list their homes in the summer rather than the winter months. The following chart shows the home sales by month each year:
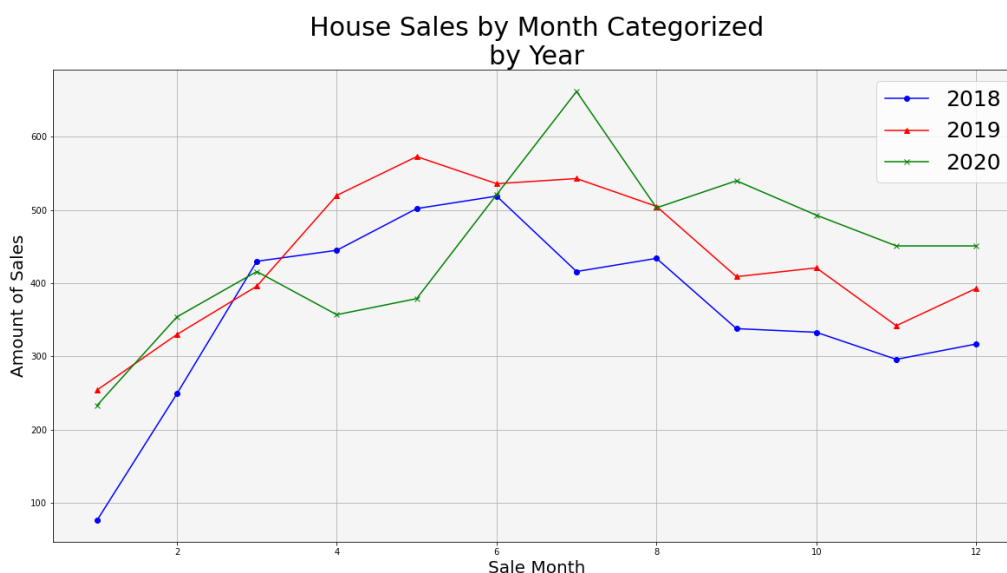
Home sales peaked in the summer months of May, June and July, while it dipped during the winter months, particularly in January.

Sellers should also be aware that the population is increasing at a lower rate than the increase in house price. The housing market may not see as high of prices as population growth and housing prices come to equilibrium. The best advice is to sell now as the market favors sellers exponentially.

Conclusion:

The Austin housing market has seen huge growth in the past few years. Housing prices have increased in every area of the city as Austin's population has grown. By compiling and analyzing this data, we are able to determine where home sales are occurring, what time of the year homes are sold and how the market affects both buyers and sellers. Using regression models, we surmised the predictability of house prices and the variables that affect it. We expect the same trends found in our analysis to hold true as long as Austin continues to grow.

Limitations:

When analyzing the Austin Housing data, we recognized there is some missing information that could also influence the sale price. One of them would be the listed date of the house. With the listed date, we can see how long the houses have been on the market. Parallel to that would be the initial listed price. This would allow us to see how the price fluctuated during the timeframe they were on the market. Additional information such as the condition of the house or the neighborhood quality could also allow us to verify their correlation to the sale price.

We also encountered limitations when it comes to the Census data. Information for 2020 was not provided, so we were not able to determine how big of a change in population compared to the increase in sale price for that year. The data was also divided into states and counties only, instead of zip code. This prevented us from merging the two data together to directly compare population to the sale price for each listing.

Future Work:

For future work, we would want to look at 2021 and see how Covid-19 has impacted the housing market in Austin. We also would like to collect additional data to see what other factors are driving the population growth and sales price, such as the number of companies moving to Austin or the increase of investors.