**Introduction**

Our analysis seeks to develop an investment thesis on what county in the US would be best for investing a hypothetical $100,000 in residential real estate. The primary objective of this investment would be to experience capital gains. A secondary objective is to achieve high rental cash flows. The combination of these will help determine the likelihood of high returns.

In pursuit of our objective, our analysis looked at several factors, demonstrated by the below diagram:

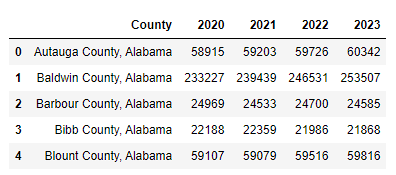
1. Poverty rates by counties (data source: census data)
   1. We wanted to identify counties experiencing much higher population growth vs housing growth, which would create unmet demand for housing, theoretically driving prices up (data source: census data)
   2. We wanted counties that had higher rent prices than mortgage prices; our hypothesis was that these counties would have upwards pressure on home prices, and a high chance for positive cash flows from rents (data source: census data)
   3. To experience higher home price increases, we would like to find communities where the median income to median home price ratio is high, which may attract more qualified buyers for homes, thus pushing prices up (data source: census data for median home price; HUD data for median income by county)

8. We used median increase in home value by county to use in correlation and comparison studies

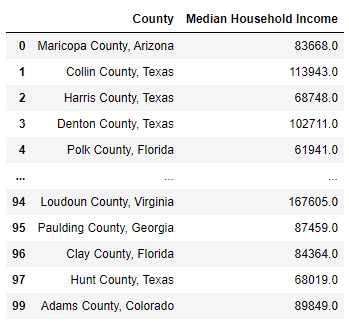
A diagram of a business

Description automatically generated

**How did we do it?**

We began tackling this question by determining the set of counties we wanted to work with. From Census.gov, we obtained a csv file with all 3000+ counties with their population numbers from 2020 to 2023.

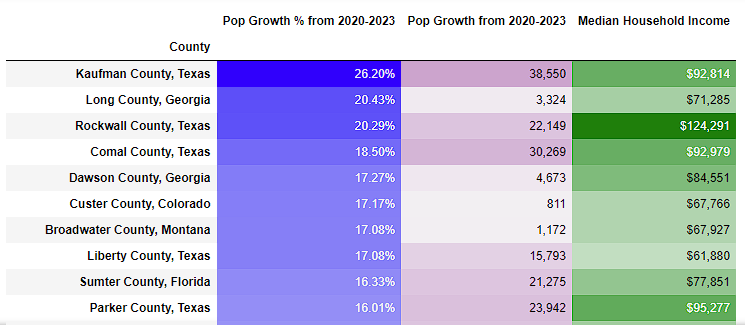
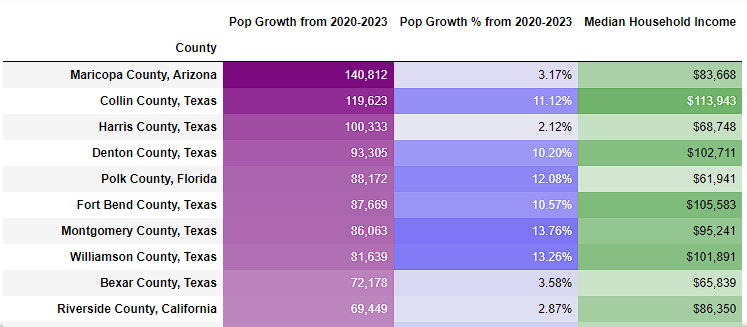
By using the data for years 2023 and 2020, we calculated how much the population had grown in numbers, then took the top 100 counties. We used the same data, and calculated the population growth in percentage, then took those top 100 counties as well. We did this because we wanted a pool of counties that had good population growth, but high population percentage growth did not necessarily mean there was good growth, if the population density in those counties were very low.

From Census.gov, we obtained another csv with the Median Household Income for all of the counties. Then by doing a left merge, we were able to create a dataframe with just the information we were interested in.

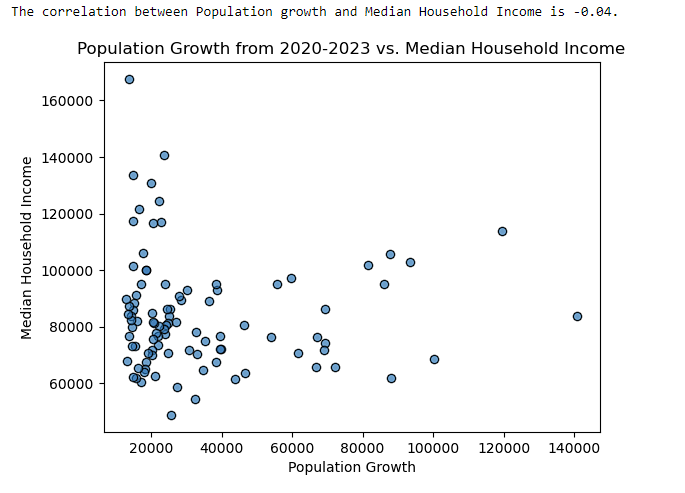
By using the geoapify API, we pulled the Latitude and Longitude information, since this would be helpful in visualization later on. Combining this information, we had a solid start to our data, and adding on to it would be simple:

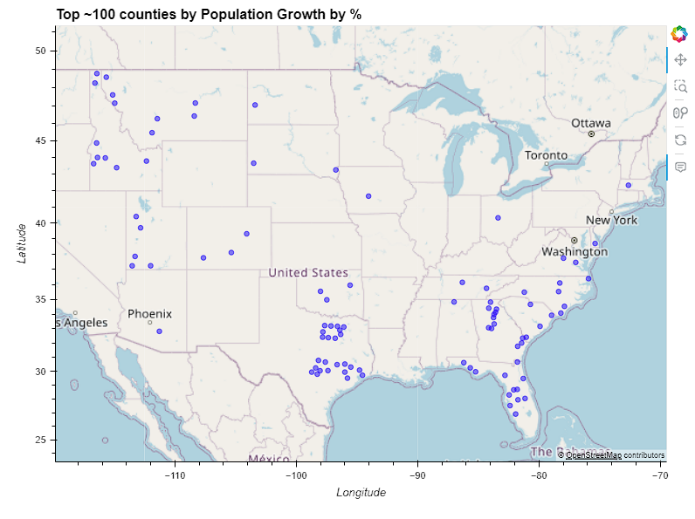


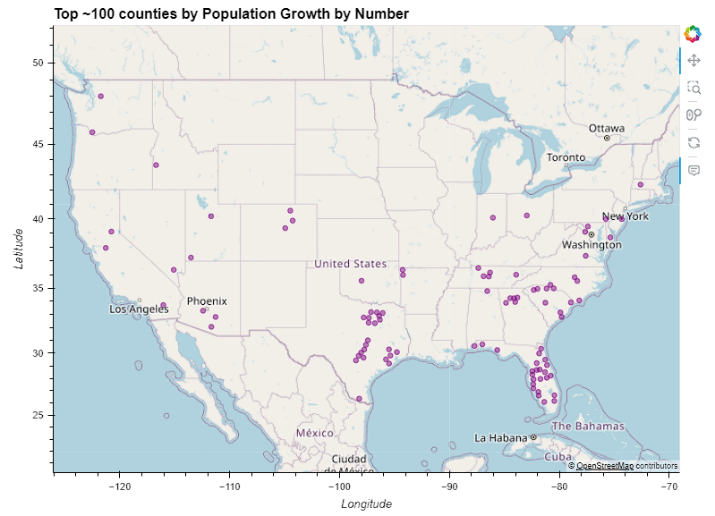
We added a bit of color to our chart in a way so that the gradient dictated which values are better (Dark colors are always better in our analysis). Here we can clearly see that the highest Population Growth had one of the lowest percentages in population growth. However, the Median Household Income does not seem to be affected by the population growth.



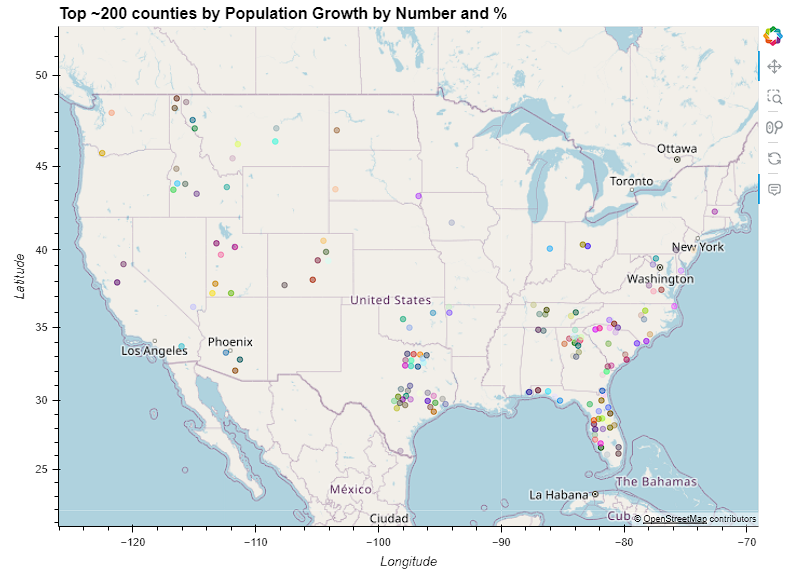
We take this a step further and ran a correlation between population growth and Median Household Income, and we can see that the correlation coefficient is -0.04, little to no correlation:



So, we cannot use one factor as a predictor for another factor, but both of these factors are important to figuring out which county would be the best place to buy or build. By using the Latitude and Longitude information, we created a visualization of the top counties based on growth by numbers, as well as by percentage:

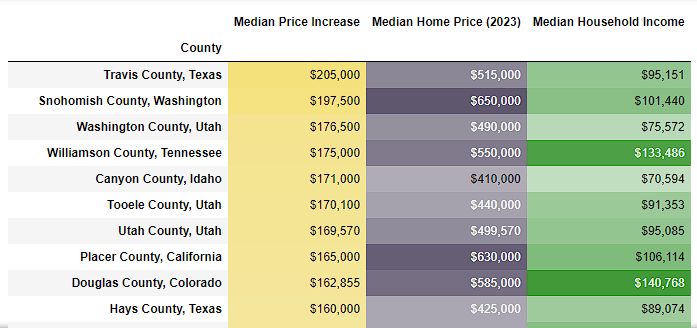


Also, by merging all the data into one dataframe, we are able to visualize all the counties put together:



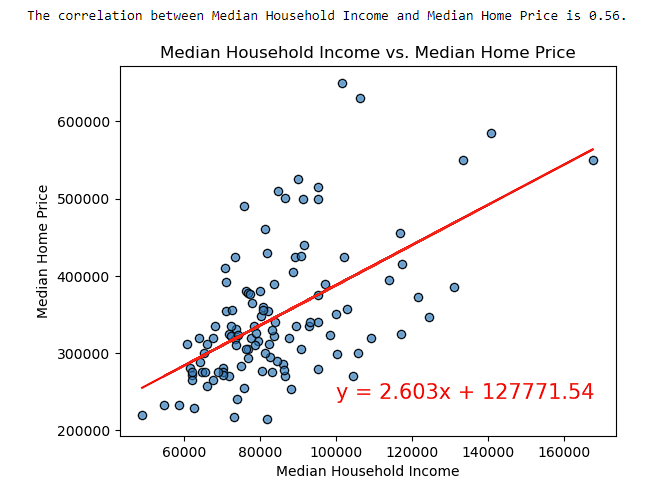
**Introducing additional factors: Median Home Prices and Home Price Increase**

Now that we have a list of counties we want to work with, we begin introducing various factors that we would use to narrow down our choices. From the HUD (Housing and Urban Development) we obtained a csv with Median Home Prices. From there, we were able to calculate Median Price Increases and put them side by side with the Median Home Price of 2023, as well as the Median Household Income.



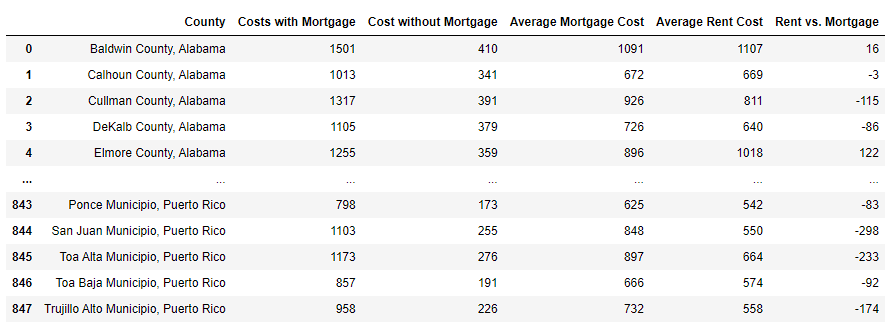
It also looked from the table colors that the homes that had a higher Median Home Price also experienced the highest Median Price Increase. We did not have the time to run a correlation analysis on these two factors, but it was interesting to see how the smallest Median Home Prices were at the bottom of this chart that was sorted by Median Price Increase.

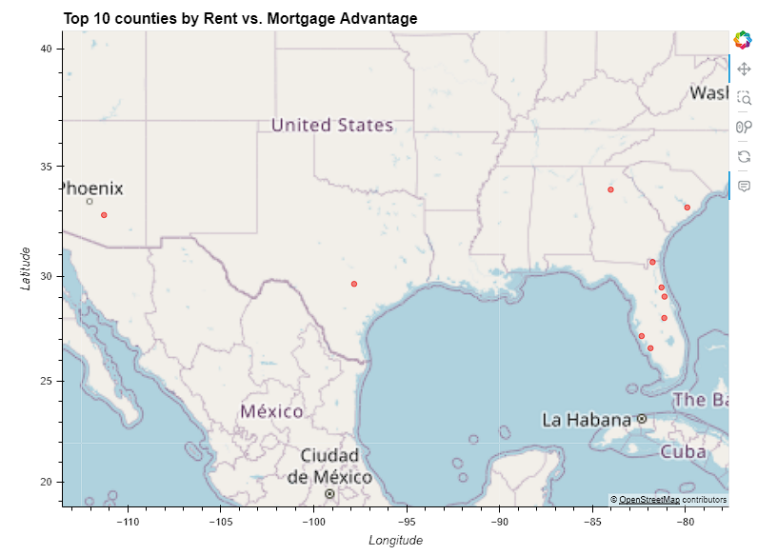
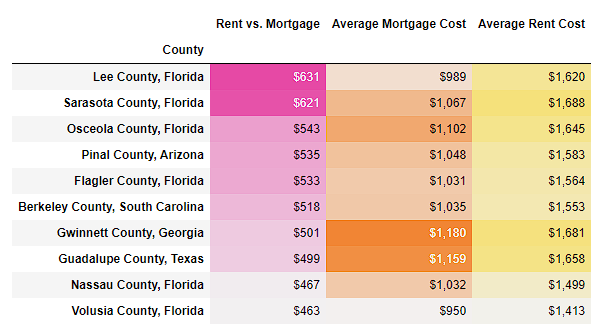
We did run an analysis comparing Median Household Income and Median Home Price, and it was interesting to see that there was a correlation between the two factors. The correlation coefficient was 0.56, and with the regression equation we would be able to estimate the Home Price of a family if given the household Income:

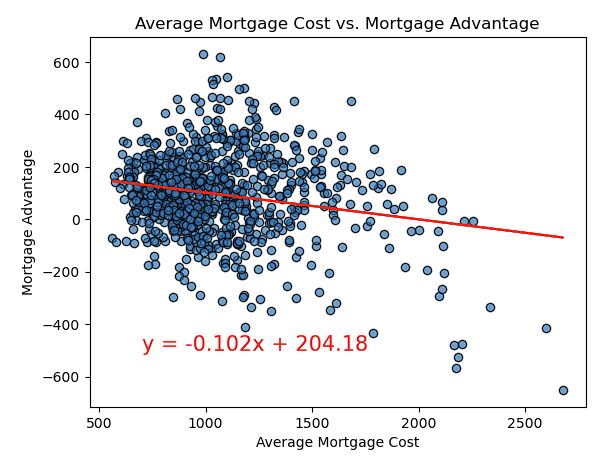


**Introducing additional factors: Rent vs. Mortgage, Average Mortgage Cost**

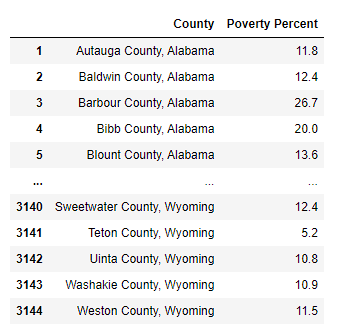
After speaking to an actual real estate agent, one of the factors that we wanted to look at was a comparison between costs of renting and costs of owning a home. This became one of the factors we used to score our counties. We obtained information from the HUD, calculated the Average Mortgage Cost, then compared each county with the Average Rental Cost. By subtracting the two factors, we were able to see how much cheaper it was to own a home than to rent:



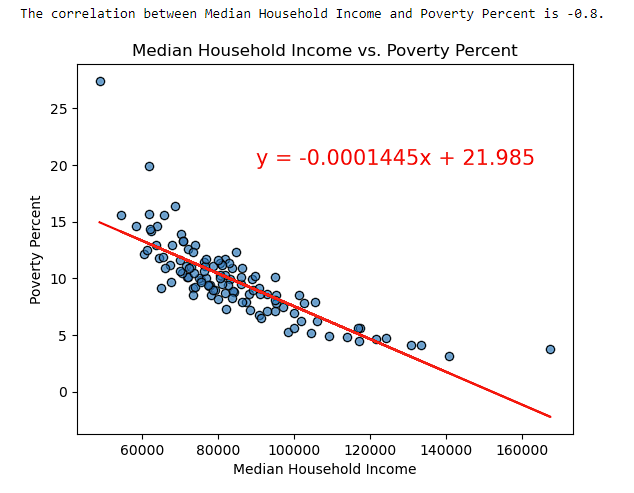
By merging this data with our list of counties, we were able to filter our top 10 counties. It was interesting to see that the majority of counties with the best Rent vs. Mortgage statistic was in Florida.

We did further comparison of the Average Mortgage cost, and the Rent vs. Mortgage statistics (which we labeled as Mortgage Advantage) and found that there was a weak correlation between the two factors. Homes with a higher mortgage cost had less of an advantage to own:

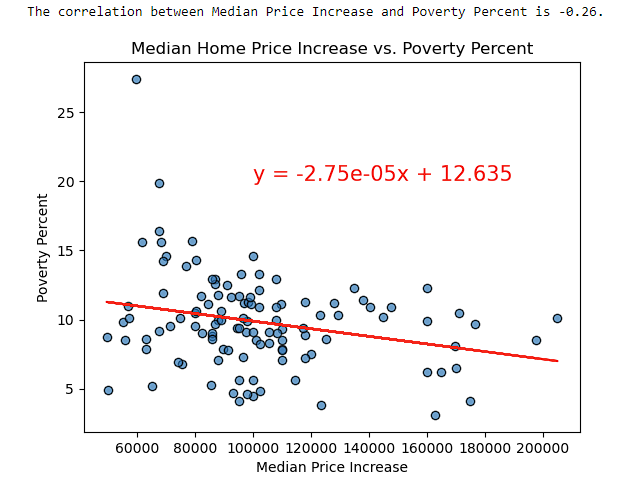
**Introducing additional factors : Poverty**

Our inquiry with a real estate agent also revealed that one of the factors to look at was the poverty statistic. Locations with lower poverty would be a better place to make a good investment. We could have pulled a csv from the Census.gov website, but we were determined to get the Census API to work. We were not able to figure out an easy way to run a for-Loop for the counties we were working with, but we did manage to pull poverty information for all 3000+ counties in the US. Like before, we were able to simply merge this data with our previous data frame to keep the information we wanted. One challenge we had was that the API did not provide the county and state information in one column like in the CSVs. In fact, the state information was abbreviated, so we wrote code to convert all the abbreviations and combine the county and state information so that it matched the rest of the data. 

We did a correlation between Median Household Income vs. Poverty Percentage, and as expected there was a very strong correlation.



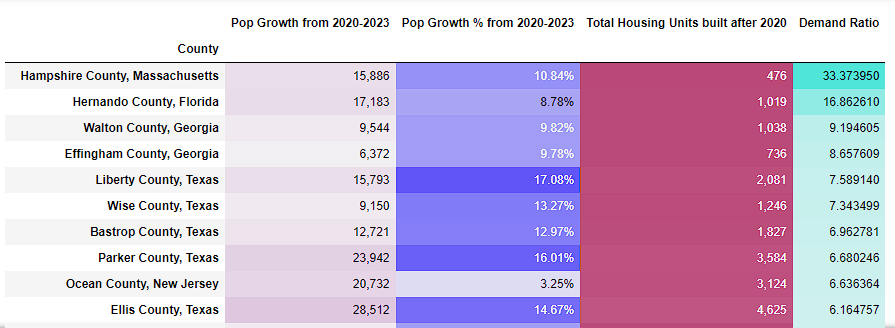
We also did a correlation between the Median Home Price Increase and the Poverty statistic, and found that there was a weak negative correlation, which was unexpected. The Median Price of homes in the area were only very slightly correlated with the percentage of poverty in the county. We expected that in areas with high Median Home prices, the poverty percent would be much lower.



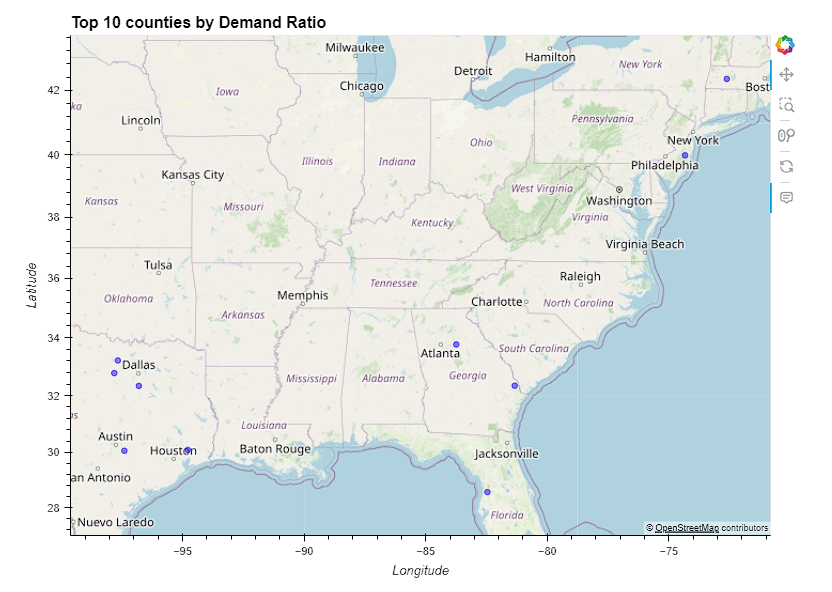
**Putting it all together: Demand Ratio, Affordability Ratio**

To filter down our list of counties into the top few counties of choice, we needed a way to make a scoring system. We already had information regarding poverty, as well as mortgage advantage for each county, but we also wanted to compare more factors.

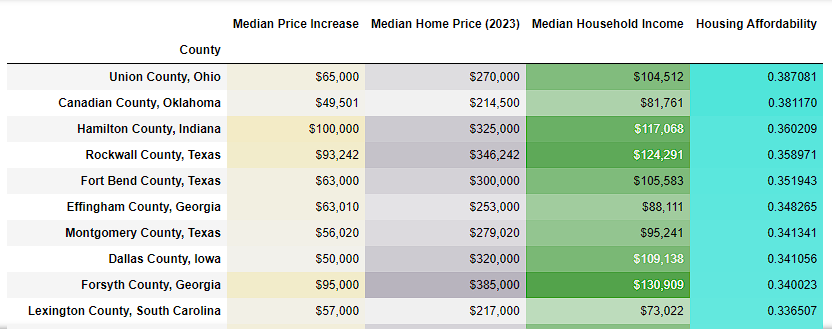
We created a Demand ratio, which would tell us which counties would have the highest demand for homes to buy. We calculated this by taking the Population Growth by Numbers in each city and dividing it by the Total Housing Units built after 2020. The higher the demand ratio, it means less homes would be available for purchase:

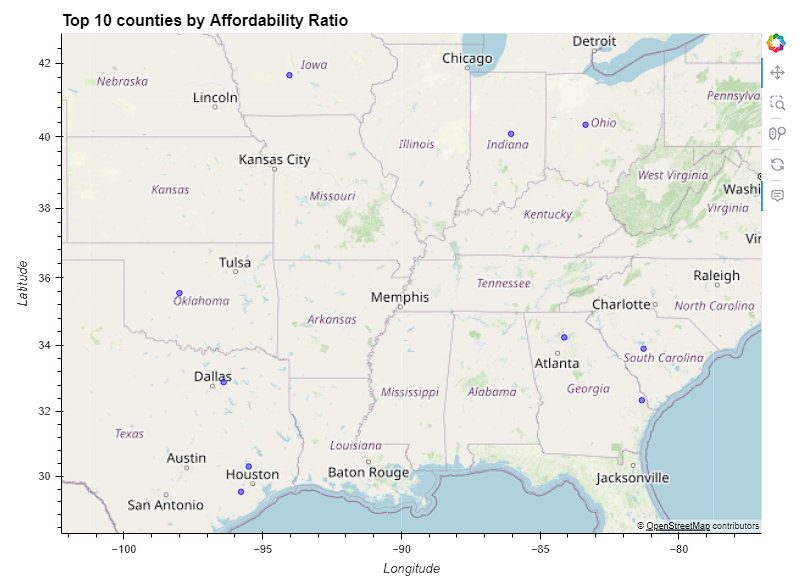


Mapping these out we can see the top 10 counties with the highest demand for homes.

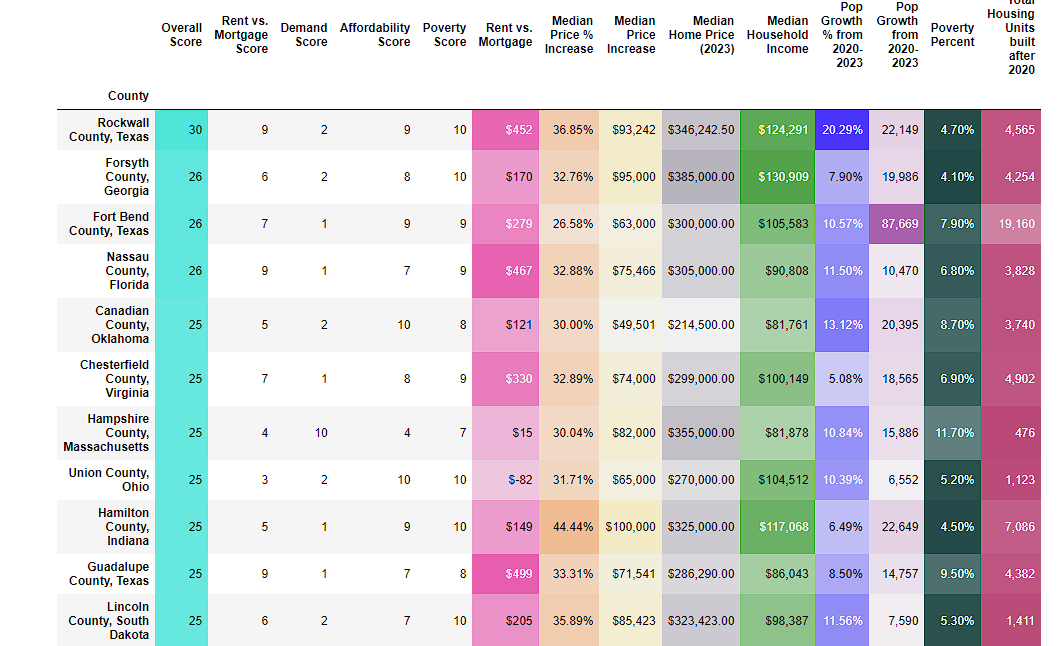


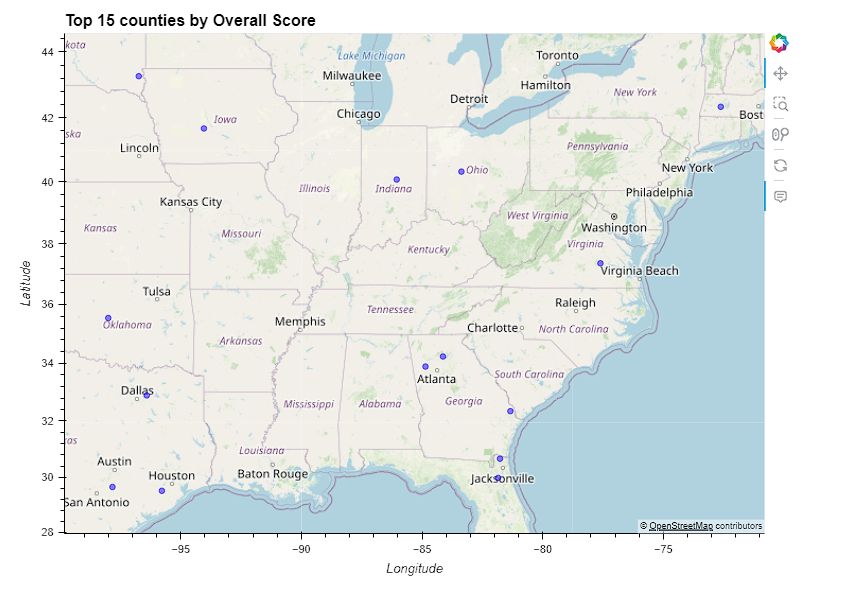
Another factor we wanted to figure out was affordability. We calculated the Affordability ratio in each county by taking the Median Household Income, and dividing by the Median Home Prices in the county. The higher the Affordability Ratio, the more affordable it was for households in each county to afford a home.



The top 10 counties by Housing Affordability mapped out as follows: 

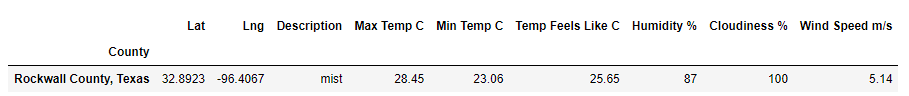
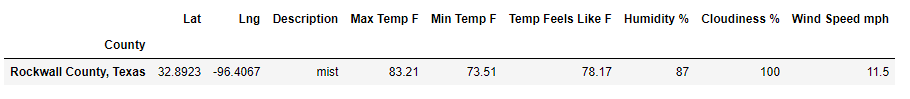
**Applying Scores to each County**

To create scores for each factor, we used the binning method which applied scores between 1-10. Since Poverty Percentage was better with a lower percentage, we used a set of reversed labels to make sure we applied the correct score. Then, by adding the scores together, we were able to filter our counties down to the best Overall counties.

Mapping our top counties, we notice that none of them were in the west coast:

Rockwall County, Texas is our top choice for location to invest in a home! Based on our data, this county had one of the highest Population Growth by Percentage, and strong Mortgage Advantage, and a low Poverty Percent. Median Household Incomes in the area were high as well.

By using the OpenWeather API, we pulled the weather for Rockwall County, Texas in both Metric and Imperial units:

**Conclusion**

Out top three counties were Rockwell County, Texas, Forsynth County, Georgia, and Fort Bend County, Texas. If we had more time, we would conduct additional analysis to improve our data. One weakness of our data is it is skewed by major sociological trends caused by the COVID crisis. This caused major population shifts across the country, and triggered spending that caused inflation and the highest mortgage rates seen in many years. These two have cast a freeze across the real estate market that makes further investment much trickier. In order to overcome this, we would need to reduce dependence on debt as an investment vehicle, and look for opportunities with lower operating costs and higher rents to offset the higher costs of debt. Adding sensitivity analysis and experimenting with weighting factors could improve the usefulness of our analysis.