**House Prices with Ethnicity**

**Austin Lien, Sneh Patel, Reid Erickson**

**1. Introduction**

According to U.S. Department of Treasury1, the average white household whose head is younger than 35 has $30,000 and $20,000 more in housing equity wealth than the average Black or Hispanic household in that same age group. By the time household heads are 55 or older, these gaps will have widened to $175,000 and $145,000, respectively. Housing affordability presents a significant challenge across the United States. By understanding how factors like demographics influence housing prices, we can gain valuable insights to guide policymakers, real estate professionals, and individuals in their decision-making. This project focuses on investigating the intricate relationship between demographics and housing prices within the state of Iowa. We will analyze county-level data to uncover potential correlations, regional disparities, and trends that could illuminate affordability issues and market dynamics within the state.

In this project, we plan to use data about home prices and demographics about counties in Iowa to determine if there are correlations between the prices of homes and the demographics of counties.

**2. Data**

This project uses two sources of data which are statistics of home prices from Realtor.com2 and data about demographics of each country from the U.S. Census Bureau3.

*2.1 Housing Prices*

For the housing data we used the csv RDC\_Inventory\_Core\_Metrics\_County\_History that was renamed to realtorData from the website <https://www.realtor.com/research/data/> that included various housing metrics for each county in the United States for the last 5 years. The realtorData csv contained 288386 observations with 40 variables which is loaded into the file *project\_data\_gather.R* as the data frame *‘realtorData’*. After loading this in, we isolated the data just to 2022 which is noted as the data frame *‘now’* and the state of Iowa noted in the data frame *‘iowa’* to match the counties and year for the data. After this, cleaned the data through matching the county names, correcting data types, and averaging the values for each month in the year to get single averages for each category across the whole year which can be seen in the data frame *‘yearly\_averages’*. The data frames are all contained in the *project\_data.RData* file*.* After this, we removed unnecessary columns not relevant to our analysis.

*2.2 Ethnicity Data*

To scrape and integrate this data, we first created an API key for the United States Government’s API for the annual census data Americans must fill out. The census has thousands of data points, but we chose to focus on county demographics including how many White, Black, Asian, Native American, Pacific Islander, and other races were in each county. To do this, in the file *project\_data\_gather.R* we used the package tidycensus in R and the function get\_acs to get the data from the 2022 annual census data, focused on counties from Iowa. We also had to specify which variable we were asking for, as demographic data was labeled from B02001\_001E to B02001\_007E. Using this we created several data frames (df1, df2, df3, etc. in ) containing the county, variable name (B02001\_001E), and estimate for each race. With these data frames we integrated them horizontally using a for loop that created a final combined data frame that included all estimates of race population for each county in Iowa in 2022. The final step to get the census data ready for merger with the housing data was to remove all unnecessary columns.

*2.3 Combining Housing Prices and Ethnicity*

To combine the housing data and the census data we did a horizontal merger by *County* combining the data frames *combined\_df* for census data and *yearly\_averages* for housing price data into one data frame names *final\_df* which contains 98 observations with 22 variables such as *population, black, median\_listing\_price, etc*. This is the data frame that will be used to conduct all of our analysis and built visualizations. All this work was done in the *project\_data\_gather.R* file. A description of each variable in the data frame *final\_df* is contained below in Table 1.

*Table 1 Data Dictionary*

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Data Type** |
| County | Name of the county | Character |
| Total | Total estimated population of the county | Numeric |
| White | Estimated White population in the county | Numeric |
| Black | Estimated Black population in the county | Numeric |
| Native | Estimated Native American population in the county | Numeric |
| Asian | Estimated Asian population in the county | Numeric |
| Islander | Estimated Islander population in the county | Numeric |
| Other | Estimated population of other races/ethnicities in the county | Numeric |
| median\_listing\_price | Median listing price of properties in the county | Numeric (currency) |
| median\_days\_on\_market | Median number of days properties stay on the market | Numeric (days) |
| active\_listing\_count | Amount of active listings | Numeric |
| median\_listing\_price\_per\_square\_foot | Median listing price per square foot of property | Numeric (currency per unit area) |
| average\_listing\_price | Average listing price of properties in the county | Numeric (currency) |
| median\_square\_feet | Median square footage of properties in the county | Numeric (area) |
| total\_listing\_count | Total number of property listings in the county | Numeric |
| pending\_ratio | Ratio of pending listings to active listings | Numeric |
| White\_Percent | Percent of total population that is White in decimal form | Numeric |
| Black\_Percent | Percent of total population that is Black in decimal form | Numeric |
| Native\_Percent | Percent of total population that is Native in decimal form | Numeric |
| Asian\_Percent | Percent of total population that is Asian in decimal form | Numeric |
| Islander\_Percent | Percent of total population that is Islander in decimal form | Numeric |
| Other\_Percent | Percent of total population that is Other in decimal form indicating any race not already described | Numeric |

**3. Analysis**

*3.1* Average Home Price by Ethnicity

In this project we wanted to find out whether there are certain ethnicities that on average indicate higher/lower home prices within Iowa. To do this analysis, we decided to use the top 10 counties with the highest population for each different ethnicity. This yielded 6 new data frames named *top\_10(Ethnicity).* To figure out the average price for homes by ethnicity we then took the newly created data frames and took the mean of the median listing price for each data frame. This yielded a range of home prices from 168,553 to 260,584 rounded to the nearest dollar.

At first, due to the introductory research we conducted, we thought that *Black* would have the lowest home price of all the ethnicities. However, after creating the bar chart using ggplot of Average Home Price shown in Figure 1, we were able to determine that the lowest price for homes was actually *Native.* The highest price for homes based on averages went to *Asians* which is not surprising based on what the average Asian makes in the U.S. compared to other ethnicities. Asians are constantly among the top average earners every year.

A graph of blue rectangular bars

Description automatically generated

Figure 1 Average Home Price by Ethnicity

Based on the results of this analysis we can conclude that in Iowa, *Asians* have the highest average home price, while on the other hand, *Natives* have the lowest average home price. However, since the data we have used is from 2022 it may be possible that the data has changed in the last 2 years and further analysis would not need to be conducted with the newest available data to determine any changes.

*3.2 Correlation Analysis*

Analyzing the correlation between variables, we unearth interesting characteristics that one would not automatically assume for this data. One point we see is the correlation between Other\_Percent / Asian\_Percent and median housing price seen in figure 2, with both having a similar correlation of around 0.35. This isn’t a particularly strong correlation, but it indicates that there is a small correlation indicating that the higher percentage of Asian or Other (likely being Hispanic or another race not mentioned in the census) in the total county population, there is typically a higher median housing price for that county.

Figure 2 Other Percentage vs. Median Listing Price

A graph showing the difference between median and median prices

Description automatically generated

Another interesting trend we see is the negative correlation between median\_listing\_price\_per\_square\_foot and White\_Percent, seen in figure 3, a negative correlation with a large clustering in the bottom right of the graph indicating a common high percentage of Whites in counties with a low median price per square foot. This clustering can suggest that many Iowa counties have an average median price per square foot between 75 and 150, with only a few counties, likely higher population counties, having higher values.

Figure 3 White Percentage vs. Median Listing Price per Square foot

A graph of black dots

Description automatically generated

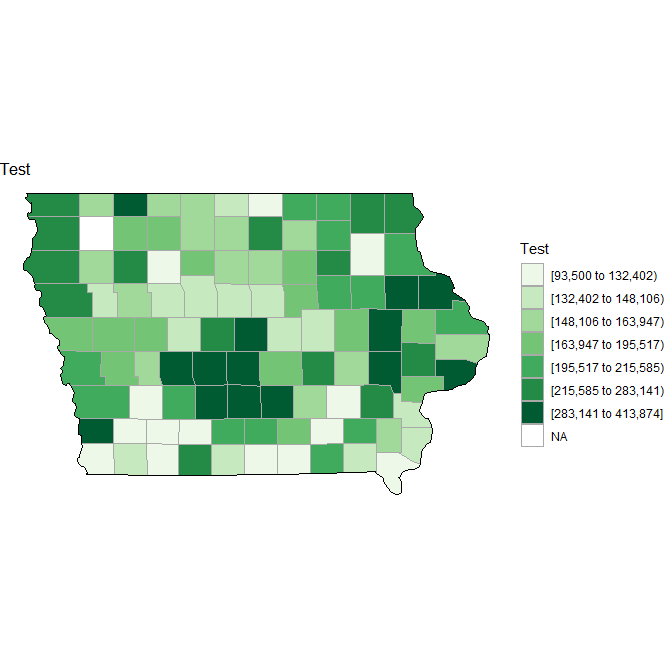
*3.3 Median Listing Price by Region*

In our project, we aimed to uncover not only which counties boasted the highest median listing prices but also to discern any discernible trends that might shed light on why certain areas commanded higher prices than others.

Initially, we assumed the Des Moines area and the Iowa City area would top the charts in median listing prices. We reasoned that the high demand to live there, driven by job opportunities and nearby attractions, would naturally inflate prices. Conversely, we expected smaller towns to exhibit lower prices due to limited job prospects and fewer attractions, making them less desirable places to live.

However, it's important to acknowledge that such assumptions might oversimplify the real estate landscape. Factors beyond size and economic activity can significantly influence housing prices. Some smaller towns might surprise us with unique charms that elevate their appeal and, consequently, their prices. Similarly, within urban centers, localized factors like neighborhood dynamics or recent development trends can lead to variations in housing prices.

To gain a deeper understanding, we need to consider a range of factors, including population growth rates, employment opportunities, access to amenities, and recent real estate developments. This comprehensive approach can help us uncover patterns and drivers behind median listing prices across different counties.

Figure 4 Median Listing Price by County

The choropleth map corroborates our initial hypotheses, revealing that the Des Moines and Iowa City areas boast the highest median listing prices per house. This outcome aligns with expectations, considering the abundant job opportunities and attractions in these regions.

On the flip side, the map highlights lower median listing prices in smaller areas across the north side of the state, consistent with our anticipated findings. These counties, lacking the economic vibrancy and amenities of larger urban centers, naturally exhibit more modest housing prices.

**4. Conclusion**

In this project, we analyzed 3 different aspects of the housing market in the state of Iowa. We looked at the median listing price by county, if there were any ethnicities that would have a higher or lower average listing price, and if there was a strong correlation between different counties of Iowa and their demographics and prices of their homes. In summary, from the analysis questions asked and the analysis that was done, here are our results.

1. Are there certain ethnicities that indicate higher/lower prices than the average? (i.e. if a county has a high majority of whites, are the home prices considerably higher)

*Yes, when looking at the bar graph that is presented, the highest average home price by ethnicity we can see that Asian has the highest average home price while Native has the lowest.*

1. Is there a strong correlation between counties of Iowa demographics and pricing of homes?

*When looking at both figure 2 and figure 3 we can see that there is a negative correlation between median price per sqft and the percentage of White people in these counties and there was a negative correlation with some clustering. This can bring us to the conclusion that many Iowa Counties have an average median price per sqft between 75 and 150 with potentially some of the more heavily populated areas having a higher value.*

1. Is there a certain region in Iowa in which pricing of homes is considerably higher than the rest/average of Iowa?

*Yes, the regions that are more populated such as the Des Moines and Iowa City areas have the higher prices whereas the regions with less people have lower prices.*

There were several limitations when completing this project such as this data is not the most current data it is from 2022 along with when looking through the ethnicities in our data, there is an other category which is not overly specific and as we assumed throughout this project we assumed Hispanic was not mentioned which could be a large percentage of the other category. For the future work being done on a project like this we recommend finding a dataset that is more current such as this year's data if available and try and find data that is more specific when it comes to ethnicity and make sure that no large ethnicities are being left out in your data.