SDS 322E

Qusay Ali, Sabreena Delacruz, Matthew Dauber, Ivy Cao, Jack Sweeney, Kim Dang, and Yanbing Lu

December 5th, 2022

Home Ownership Report

Introduction

It has always been the American dream to own a home. However, due to limited inventory and inflated home prices, it has become increasingly difficult in today's economy. The homeownership rate in the US is less than two-thirds of the population, dropping down to only one-third when looking at Americans under the age of 34. Nonetheless, buying and owning a home remains a smarter financial move than renting because owning a home represents stability, independence, and long-term investments. Therefore, if you're financially ready, buying a house is still worth it — even in the current market.

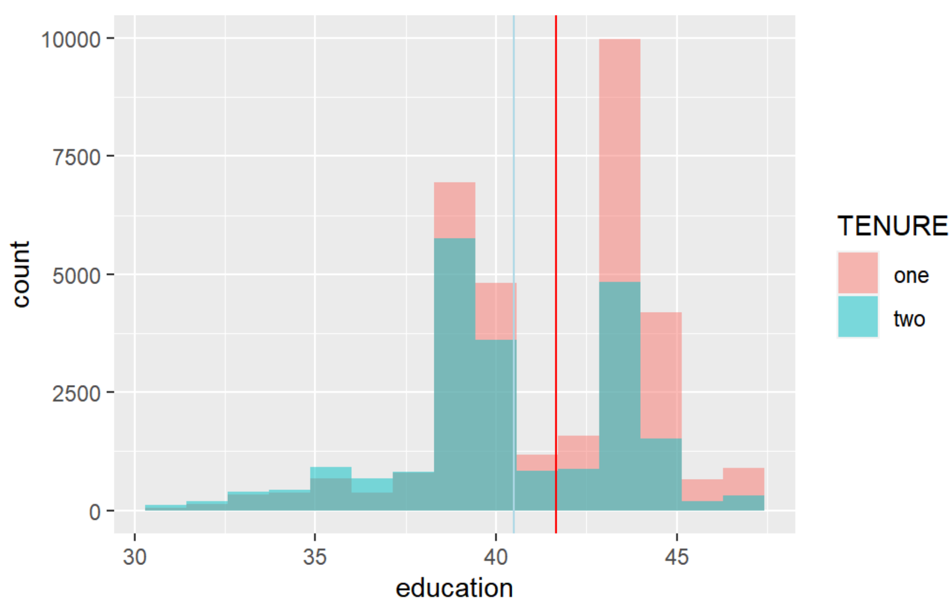
Although, it is a hard decision to make because one needs to weigh many factors when considering what the right time is to become a homeowner. This problem led us to our question of which extrinsic attributes help determine the likelihood of one owning a home. Our goals are to identify the factors that most correlate to homeownership. For our project, we are specifically looking at household demographics like income, sex, and education to predict if they own a home or not.

Data

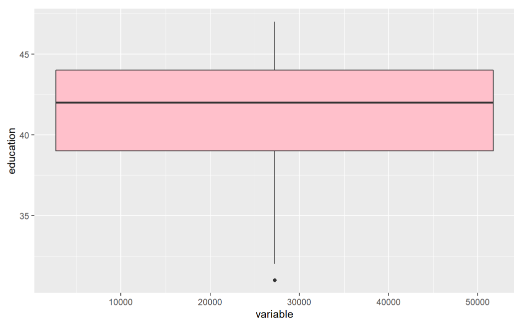
Our Data are from the 2019 American Housing Survey (AHS). AHS is the most comprehensive national housing survey in the U.S., is sponsored by the Department of Housing and Urban Development, and is conducted by the Census Bureau. Therefore, our data is very accurate since it comes from a valid resource. There are 5 different variables used for our data in order to narrow the wide variety of searches in the housing survey. We looked at tenure, which is the type of property they were paying for. Education, which measured the level of education of the people living at the property. Sex, male or female. Household composition, whether it is a single householder, family, or recently married. Income, which measured the household's total income over the past 4 months. The process of cleaning our data began with importing the dataset. We then selected 5 variables from the original 3,360 different variables, removed all cases of "NA"s, and cleaned off all the 's that surrounded each number.

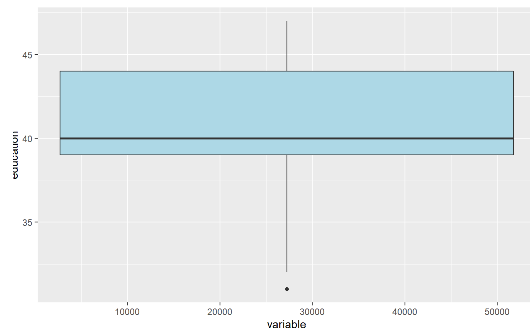
Exploratory Analysis

Hypothesis I: Histogram Showing Education with mean lines



The corresponding boxplot:

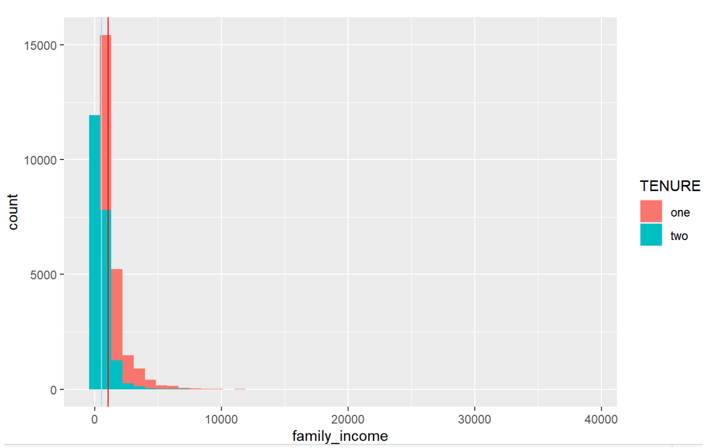




(for tenure = 1) (tenure = 2)

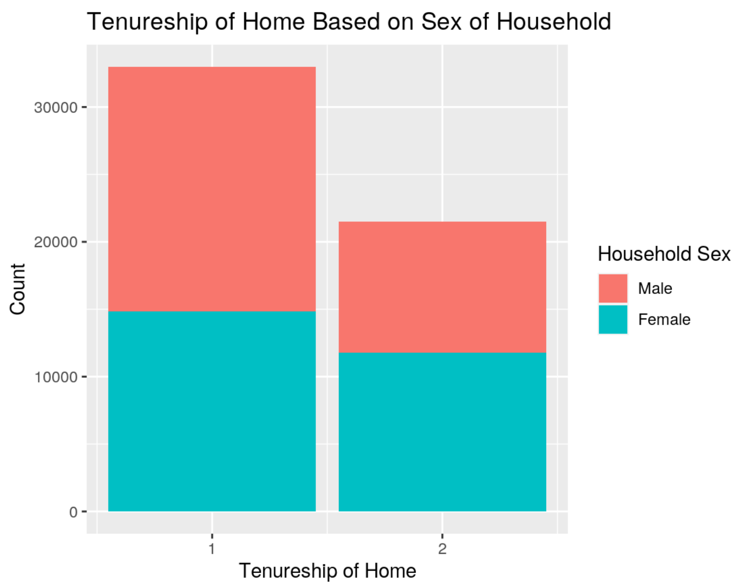
As the level of education increases, so does the likelihood of owning a home. A tenured professor holds a full-time position with job security at the college level. Tenured professors typically enter the academic job market after earning the highest degree in their field, generally a Ph. D., and hold titles like an assistant professor, associate professor, and full professor. A degree is a stepping stone to eventually becoming a homeowner. So while factors such as mortgage fluctuations or housing market volatility may explain the nationwide inability to own a home, this generational shift could answer for the increased ability of those with a degree.

Hypothesis II: Histogram for Family Income(/$1000)



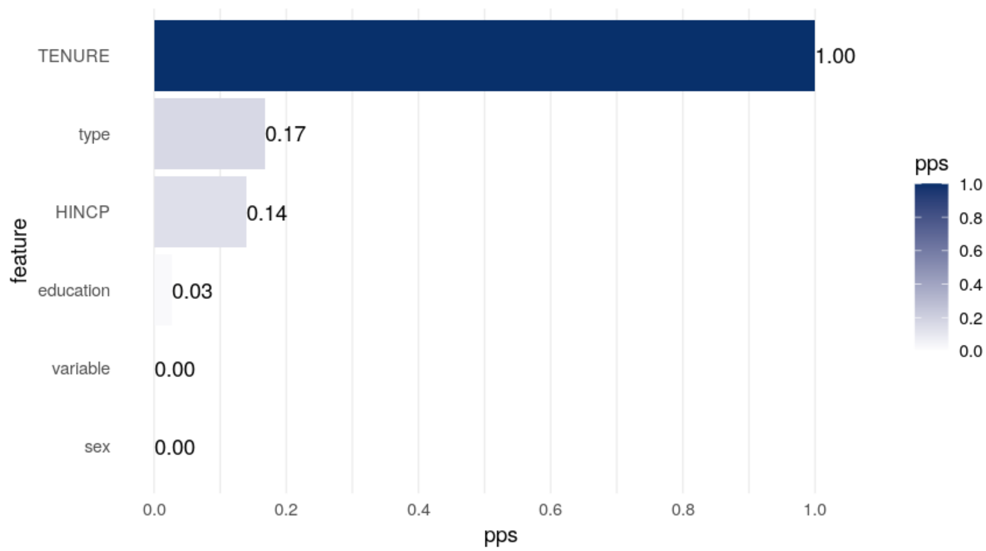
Higher family income increases the chances of owning a home. Homebuyers have a median household income of about $90,000, compared with the national median of about $66,000. Given the challenges of the pandemic real-estate market, that means that first-time homebuyers are a shrinking part of the market. Those with a higher family income and especially those who have generational wealth are more likely to own a home.

Hypothesis III: Tenure vs Sex

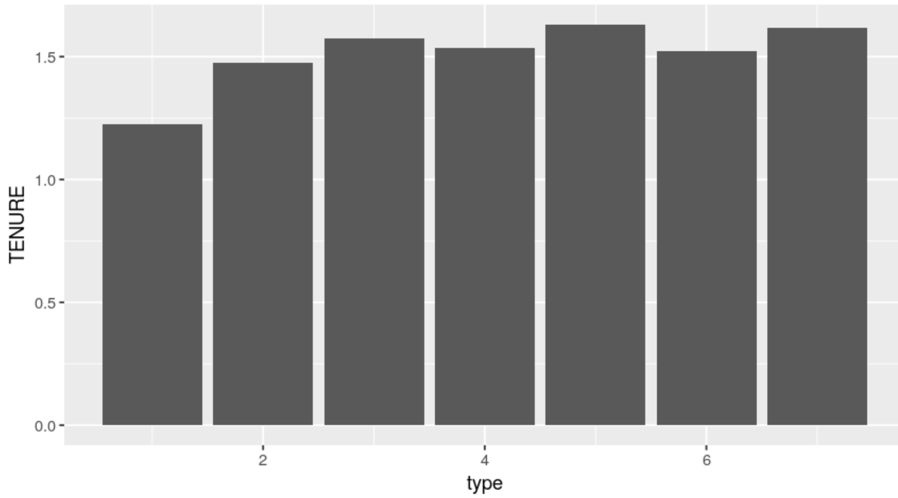


The hypothesis tested was that single women tend to have higher homeownership rates than single men. The hypothesis was developed as past data from the National Association of Realtors suggested women are more likely to own homes. A stacked histogram was created in order to visually view the ratios between males and females for when they own the home versus when they do not own the home. When tenureship has a value of 1, the residence is owned and when tenureship has a value of 2, the residence is not owned. Viewing the histogram, women do not tend to have higher home ownership rates versus that of men. This can be seen as an equal ratio is present for homeownership between males and females. Furthermore, neither men nor women are more likely to not own a home as seen in the visualization above.

Hypothesis IV: Household type and Homeownership



In this graph, we charted each variable's predictive power score in relation to tenure. Obviously, tenure itself is at the top at 1, it makes sense that if you know someone's tenure you would be able to predict their tenure status perfectly. The second and third variables are the only ones with any predictive power, and those are household type and household income. This proves our hypothesis that household type is the most useful variable in predicting homeownership.



This graph shows us Tenure on the y axis with 1 being owned and 2 being not owned, and household type as the x axis. The variables for household type are as follows:

1: Married-couple family household

2: Other family household: male householder, no wife present

3: Other family household: female householder, no husband present

4: Nonfamily household: male householder, living alone

5: Nonfamily household: male householder, not living alone

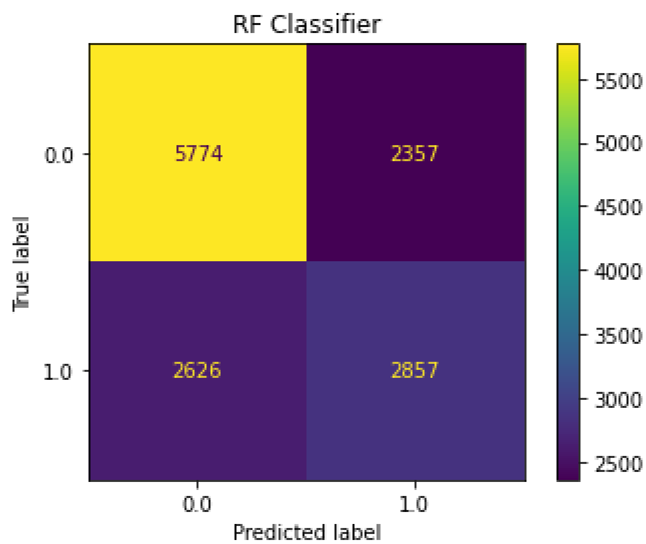
6: Nonfamily household: female householder, living alone

7: Nonfamily household: female householder, not living alone

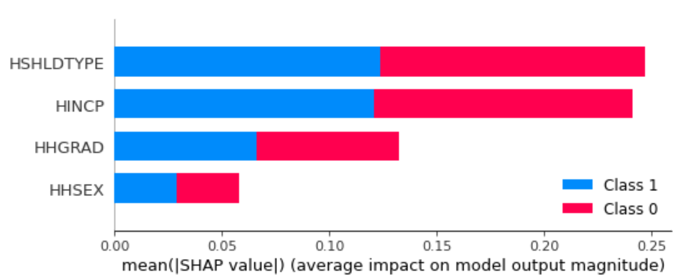
From this graph, we can see that a married-couple family household is by far the most likely to own their own house, while nonfamily households are generally less likely to own with nonfamily households that don't live alone being the least likely. This makes intuitive sense because married households are likely pooling income and have enough to support a family while nonfamily households have people living together usually temporarily only paying for themselves.

Modeling

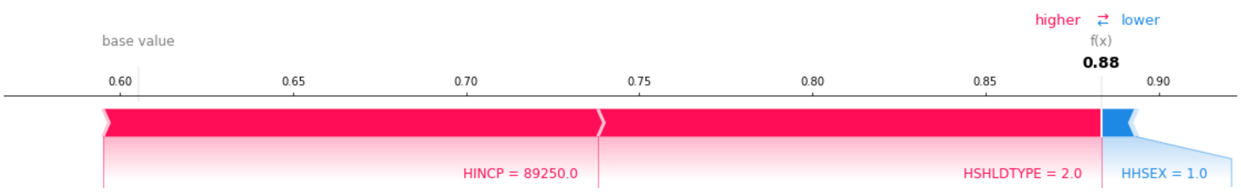
For our project, we looked at homeownership using US census data, which we measured using the variable TENURE. We wanted to look at predicting homeownership using education - HHGRAD, income - HINCP, sex - HHSEX, and household type - HSHLDTYPE. We performed a binary classification task, with TENURE as our dependent variable. We coded 0 to represent owning a home, and 1 as not owning a home. Then we created a RandomForest classifier and plotted the confusion matrix. We also used SHAP to evaluate the importance of features.



The random forest classifier above demonstrates the accuracy of the model by testing it against true values. Analyzing the visualization, the classifier was able to predict the correct label when the home was owned 42% of the time, and was able to correctly predict when the home was not owned 21% of the time. The other percentage of the predicted values were errors and predicted the wrong label.



The SHAP analysis above contains class 0 if the home is owned, and class 1 if the home is not owned. Viewing the feature importance using the visualization above, the most important features that go into homeownership was the type of household and the household income. This was determined because the magnitude of the two features were the largest out of all the features in the SHAP model.



The SHAP visualization above further proves the feature importance of the dataset. The visualization further illustrates that the household income and the household type were the biggest features in predicting the homeownership of the household.

Discussion

Our model was accurate 63.4% of the time, but only had a precision of 54.8%. This means that since precision measures how many of the predicted positives are actually positive, our model might lean towards false positives over false negatives. Additionally, it may not be a very good predictor of homeownership considering it was not correct 75% of the time. Having more variables would have potentially made these predictions more accurate. We were limited by the time frame we had to work on this project, as processing and cleaning the data for just these 5 variables took a while, so it ended up being untenable to do so for the entire dataset, as many were filled with lots of NAs. With regard to the SHAP analysis, the variables with the most influence on homeownership were household type, and household income, with household type being slightly more impactful. The 2nd SHAP plot(plot 3) shows the probability of an individual data point belonging to class 0 or 1, or rather, of owning a home or not, and specifying what amount of weight came from each variable. The f-score of the binary classifier was 0.53 which is quite low, meaning that the classification model used was not very accurate at predicting homeownership.

Accuracy: 0.6339797267518731

Precision: 0.5479478327579593

recall score: 0.5210651103410542

f-score: 0.5341684584462935

Conclusion

In conclusion, based on our data we have, household type and income have the most influence on homeownership. As the level of education increases, the likelihood of owning a home does not increase as seen in the first visualization. Higher family income increases the chances of owning a home. Viewing the stacked histogram, we rejected the hypotheses that women tend to be homeowners more than men. Using the SHAP models, we were able to determine that household type and household income were the most important features when predicting using our random forest classifier.

The next steps we could take is to broaden this study and see how we can increase homeownership in the United States. With the rapidly expanding global population, an alarming amount of new homes will be needed by the end of the century. There is an urgency to somehow make homeownership more feasible. For example, Romania introduced the First Home program in 2009 to help first-time buyers struggling after the 2008 financial crisis. As a result, the vast majority of Romanians are now homeowners. Romania has the highest homeownership rate in the world at 96.4%. The national rate of homeownership in the US is 64.8%. First-time buyers make up only 33% of all home buyers. It would be beneficial for the United States to adopt a better homeownership program.

Acknowledgment

Name	Percentage Contribution
Sabreena Delacruz	100
Matthew Dauber	100
Qusay Ali	100
Kim Dang	100
Ivy Cao	100
Yanbing Lu	100
John Sweeney	100

Bibliography

"Codebook." *United States Census Bureau*, https://www.census.gov/data-tools/demo/codebook/ahs/ahsdict.html?sortby=name&s_variablelist=.

Stasha, S. (2022, September 29). *Home ownership statistics for 2021: Policy advice*. Home Ownership Statistics for 2021 | Policy Advice | Policy Advice. Retrieved November 27, 2022, from https://policyadvice.net/insurance/insights/home-ownership-statistics/

Lautz, Jessica. "Single Women Buyers Outpace Men, but Not without Sacrifices." www.nar.realtor, December 8, 2021. https://www.nar.realtor/blogs/economists-outlook/single-women-buyers-outpace-men-but-not-without-sacrifices.

Staff, C. B. J. N. (2022, April 6). *Study: College graduates more likely to own their own home*. Corridor Business Journal. Retrieved December 5, 2022, from https://corridorbusiness.com/study-college-graduates-more-likely-to-own-their-own-home/