Humana-Mays Healthcare Analytics

2022 Case Competition

Analysis of Housing Insecurity

Table of Contents

1. Executive Summary

Housing insecurity is a serious social issue that includes unaffordable housing prices, limited space to live and dangerous living conditions. It is detrimental to people's health, mental and long-term development in education if people live in a housing insecurity condition for a long time. Therefore, housing insecurity is an important social determinant of health that needs our attention. As a leading provider of healthcare, Humana is dedicated to helping its members to overcome challenges in social determinants of health. This study focuses on analyzing what contributes to housing insecurity and what actions can be taken to eliminate this issue.

Dataset for this competition was very informative so we spent sufficient time on understanding the data. After receiving the dataset, we spent one week doing feature understanding, data distribution exploration and data cleaning. During the feature understanding process, we selected potential features by their meaning as well as by machine learning method Recursive Feature Elimination. As for the data cleaning, we separated our data into two groups based on their data type and cleaned them differently. For the distribution exploration, we made correlation heat maps and feature plots to gain a general impression of the data structure. All those operations helped us build more effective and efficient predictive models in the next step.

To select the best model forecasting that contributes to housing insecurity, we developed multiple machine learning models, applied model comparison and model optimization to improve model forecasting accuracy. In the first step, we built linear regression, logistic regression, random forest, XGBoost, lightGBM, and neural network models with a balanced train dataset. For random forest, XGBoost, lightGBM, and neural network models, we used parameter tuning to determine the best parameters. We then applied AUC to calculate training model accuracy, and find out the most accurate model. Ultimately, we chose to use XGBoost as our final model.

After we obtained the final model, we took a step forward to check the important variables that are best involved in our predictive results. In specific, the homeowner status, risk sense, expenses features and geometric features are four main aspects the model identified that are worth further attention. Based on the identified features, we further consider the impact of housing insecurity from the general economics to specific business impacts, and provide our solutions based on our model results. We also considered the impact on Humana and raised potential solutions from the individual level, community level, and collaboration levels. Finally, we conclude our discoveries and solutions to the potential impact of stockholders.

2. Background

Living in stable, adequate homes that are affordable to families offers a wealth of opportunities and stronger outcomes for children, youth and adults. A stable family provides a platform for improved employment, health and education. However, housing insecurity is still a serious social issue that encompasses a number of problems, including unaffordable housing costs, substandard, overcrowding, homelessness and etc. This is a problem that needs to be addressed urgently, because housing insecurity is detrimental to people's physical health, mental health and long-term development. Based on one report given by the American Hospital Association, the average life expectancy is 27.3 years less for people without stable housing than for people living in a stable home (Baggett et al., 2013). In addition, young people are more prone to experience mental health issues, developmental delays, and cognitive impairments when they lack secure housing.

With the understanding that housing is a core social determinant of personal health and well-being, more and more attention and effort is being devoted to addressing housing insecurity. Humana, a major health insurance company with 17 million medical members within the United States alone, has been working to address social determinants of health and the health-related social needs for the members and communities (*Medical Membership of Humana 2021*, n.d.). For this reason, the company has launched a social health initiative called Bold Goal, of which housing stability is one of the key focuses.

In this analysis, we examined the relationship between housing security and economic and health factors, and used data models to predict the likelihood of members experiencing housing insecurity. By analyzing the characteristics that affect housing security, we hoped to provide the company with multi-faceted business solutions that could be taken to address these issues.

3. Data Preparation

3.1 Data Understanding

To perform the analysis, a training dataset containing 48,300 rows and 881 columns was obtained. These features can be generally classified into 10 different groups, including demographics, medical claims, pharmacy claims, laboratory claims, credit data, CMS features, condition-related features, screening response counts, social determinants of health features and other features. Of these, demographic data were provided at the county level, while the other features were specific to each individual. In addition, each feature was presented in one of three data types: categorical data, numeric data, and binary indicators.

3.2 Data Cleaning

To improve the subsequent machine learning models' training and to lessen the possible negative impact of data abnormalities on model performance, we started our study by doing in-depth data cleaning and imputation. The parts of our process most pertinent to this study are highlighted below.

When first inputting the dataset into R, the missing values are distributed throughout the data table as "null" in string, transforming all data columns into categorical variables. We modified them into NA which can be recognized properly in R and transformed the numerical data into "numeric" data type based on the data dictionary file.

For the data cleaning, we first divided all the variables into numerical variables and categorical variables. For the numerical variables, we used Near-Zero Variance Filter to potentially remove variables that are highly sparse and unbalanced. To deal with the great number of variables in the dataset, we used Recursive feature elimination (RFE) with built-in random forest to remove the weakest features and reduce the potential internal correlation between variables.

In the end, we included 237 variables to fit the models in the next step. Before training models, we also normalized all the variables. The goal of this process was to change the values of numeric columns in the dataset to use a common scale, without distorting differences in the ranges of values or losing information.

For the categorical variables, we first filtered out all string variables into categorical sets. The index variables with integer data type are also included in the categorical variables. Then all variables with only one category, namely "cms_institutional_ind", "cms_hospice_ind", and "cms_ma_plan_ind" are deleted from the dataset. The remaining categorical variables were transformed to “factor” data type in R for model training. As a result, we obtained 18 categorical variables in total (appendix 1). To further synthesize the features into model-preferred styles and identify new potential features, we conduct feature engineering for categorical variables. In specific, rucc_category is merged into “Metro” and “Non-metro” subtypes in the new variable “metro”.

3.3 Data Imputation

The categorical variables and numeric variables were then combined again to prepare for imputation of missing value. To accommodate the high predictive dimensions (large number of columns), “predictive mean match” method (PMM) in the MICE package was used. In specific, linear regression models were applied for each variable to figure out the most similar cases in a proper predictive variables set. The missing values were then imputed with one of the three most similar cases randomly. In practice, we repeated the imputation process five times and calculated the mean value of each imputation for numerical variables, mode value for categorical variables.

3.4 Data Balancing

Original dataset was heavily imbalanced. 2,118 objects had housing insecurity, while 46,182 objects did not have housing insecurity (Figure 3.6.1). In this project, we used several machine learning algorithms for our model building. Machine learning algorithms tend to tremble when faced with an imbalanced classification dataset. Additionally, imbalanced datasets contribute to inaccurate forecasts and skewed accuracies (avcontentteam, 2016). Therefore, we aimed to transfer the imbalanced dataset to a balanced dataset to better evaluate our models. Three data balancing methods were applied during the process: oversampling, undersampling, and ROSE.

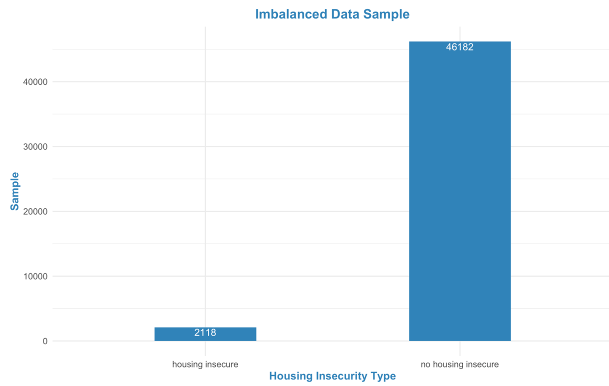

Figure 3.4.1 Original Imbalanced Dataset

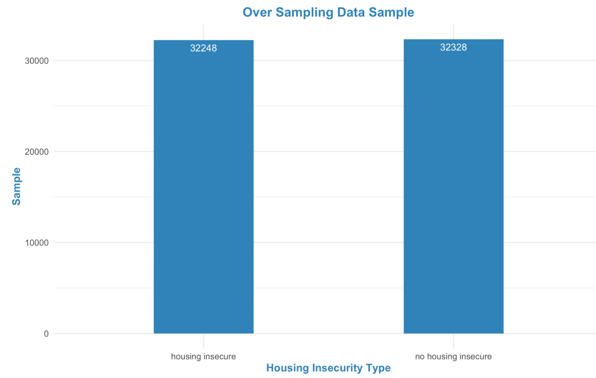

Figure 3.4.2 Oversampled Dataset

Oversampling method randomly selected data from housing insecure groups and repeatedly duplicated until there are an equal number of data samples in housing insecure and housing secure groups. After applying this method, 32,248 objects had housing insecurity, while 32,328 objects did not have housing insecurity (Figure 3.6.2). The Undersampling method counted the number of housing insecure groups, then randomly selected a similar number of data points from housing secure groups. After applying this method, 1,483 objects had housing insecurity, while 1,536 objects did not have housing insecurity (Figure 3.6.3). The ROSE (Random Over Sampling Examples) method used smoothed bootstrapping to draw artificial samples from the feature space neighborhood around the housing insecure group. After applying this method, 16,811 objects had housing insecurity, while 17,000 objects did not have housing insecurity (Figure 3.6.4).

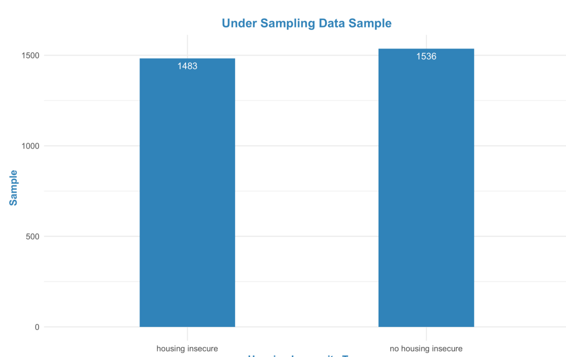

Figure 3.4.3 Undersampled Dataset

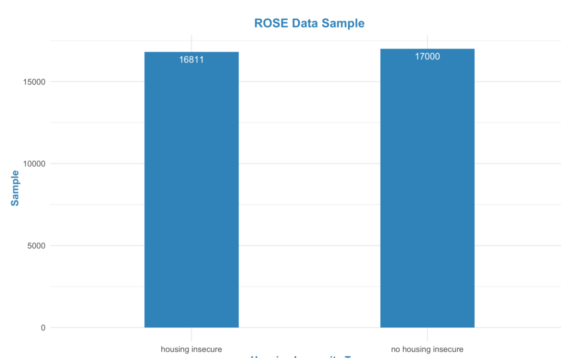

Figure 3.4.4 Balanced Dataset with ROSE

To select the best data balancing method, we used AUC (Area Under ROC Curve) values to evaluate and compare these methods. AUC of oversampling, undersampling, and ROSE methods are 0.641, 0.656, 0.648, respectively (Figure 3.6.5). Even though the undersampling method has the highest AUC score, it only selected 3,019 data points, or 6.25% data from original training datasets. Unsampled dataset is likely to lose potentially important data, as well as cause biased housing secure groups. Therefore, we chose to use ROSE to balance our dataset. The ROSE method had a good performance eliminating data bias and overall provided a better dataset for model development.

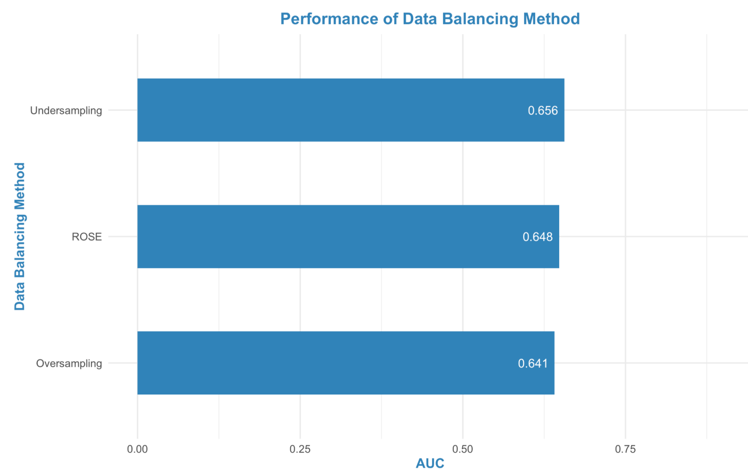

Figure 3.4.5 Performance of Three Data Balancing Methods

4. Modeling

4.1 Model Comparison

In this report, we were interested in housing insecurity (hi_flag), which was a binary variable with class "0" and "1". Therefore, we applied binary classification machine learning models to predict the response and set AUC value as primary performance metrics.

To compare the performance of different machine learning models and select the final model with the best predictive performance, we first split the preprocessed dataset into a training set

(70%) and a testing set (30%). We further applied five-fold cross-validation (CV) to the training set and compared the average AUC score of each training fold to obtain the final evaluation score. In specific, the training set was equally split into five subsets. In each training round, we use one subset as the validation set and the remaining as the training set. The AUC score was then calculated in each validation set and averaged to obtain the final AUC score. The same five-fold CV method is applied for hyperparameter-tuning when the model requires the selection of hyper-parameters (GLMNET, Random Forest, Neural Networks, XGBoost, LighGBM models).

Finally, we selected the parameter-tuned model with the highest AUC score.

Table 4.1.1: Model Performance with AUC metrics in validation set and test set

Model	Validation AUC	Test AUC
Logistic Regression	0.7353	0.7246
GLMNET	0.7537	0.7315
Random Forest	0.7278	0.7059
Neural Networks	0.9632	0.6770
XGBoost	0.8002	0.7352
LightGBM	0.7556	0.7343

4.2 Final Model

After comparing the performance of six models, we selected XGBoost (eXtreme Gradient Boosting) as our final model. XGBoost is an optimized gradient boosting algorithm that avoids overfitting and bias through parallel processing, tree-pruning, and handling missing values (Sharma, 2021). In addition to its good performance regarding high AUC values, we chose XGBoost also based on the following reasons:

- The XGBoost algorithm provides a large range of hyperparameters. Hyperparameter tuning provides optimized values for hyperparameters, which maximize model's predictive accuracy. A wide range of hyperparameters can give us more parameter options and ultimately allow us to build more accurate models.

- XGBoost provides various intuitive features, such as parallelization, distributed computing, cache optimization, etc. Parallelizing the whole boosting process can greatly eliminate the model training time. Distributed computing helps XGBoost improve model performance, scalability, resilience, and overall improves the model effectiveness and efficiency. In addition, XGBoost stores its intermediate calculations and statistics in optimized cache, which makes the prediction process much faster than some other machine learning models (Jorly, 2021).

4.3 Parameter Tuning

XGBoost (Extreme Gradient Boosting) is an optimized distributed gradient boosting library. Here we list some key hyperparameters we turned in the cross validation part to parallel comput, regularize and prune on our medel, in order to increase the accuracy of the model.

- **Nrounds:** nround means the maximum number of iterations. For classification, it is similar to the number of trees to grow. In our turning process, we set the lower bound 50 and upper bound 300 for the automatic selection.
- **Max_depth:** This parameter controls the depth of the tree. With the increase of this parameter, the model will become more complex, with higher chances of overfitting. Thus, we set a reasonable range here between 1 and 10.
- **Eta:** Eta means learning rate, which is the rate at which our model learns patterns in data. After every round, it shrinks the feature weights to reach the best optimum. Lower eta leads to slower computation and smaller change per time, while higher eta leads to faster computation but has a hard time shrinking to a small area. Here, we set the lower bound 0.01 and upper bound 0.4 for the automatic selection.
- **Lambda:** Lamda controls L2 regularization on weights. It is used to avoid overfitting. we set the lower bound 0.1 and upper bound 1 for the automatic selection.

5. Performance Analysis

5.1 Accuracy

The first thing we evaluate is the model accuracy. A ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. After the model selection, for our final model XGBoost, we evaluate its performance on the testing dataset by test ROC curve plot.

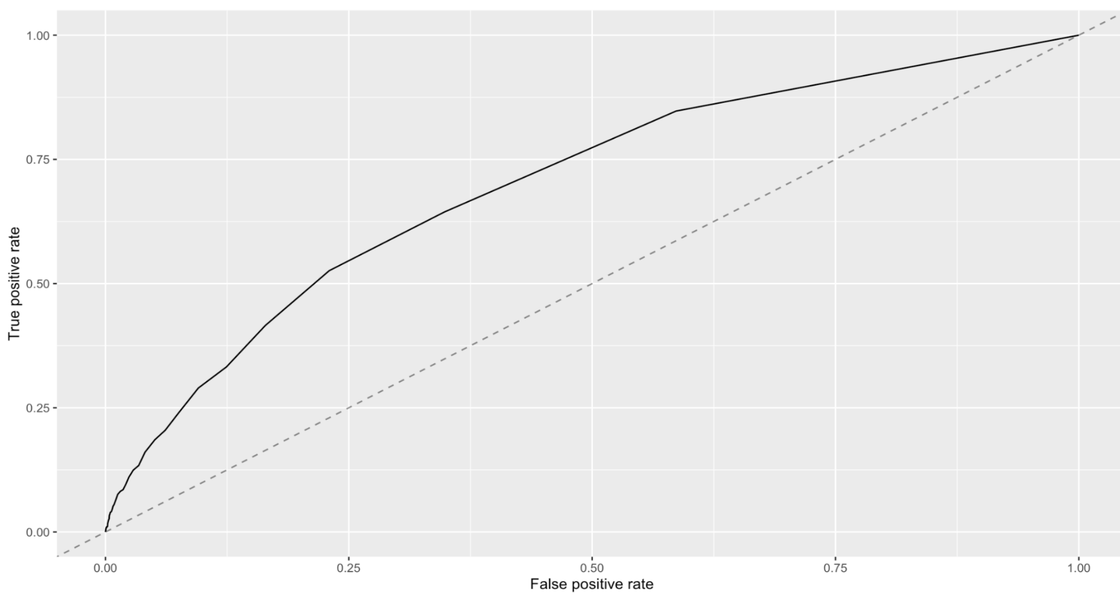

Figure 5.1.1 Test ROC Curve Plot by XGBoost

The testing ROC tells us the model performs well on the testing data, which means that it has a fair ability to predict new data. Also, since the training AUC and testing AUC are not divergent, it means that our model did not overfit and is qualified to do future prediction.

The second thing for the accuracy evaluation is about Threshold-Moving performance in False positive rate, True positive rate and Mean misclassification error. A binary classification model primarily returns a like-probability score for each class in the target variable, which gives a measure of how likely it is that the prediction obtained for that observation is the positive class. The plots below show the performance across the threshold grid and whether the performance estimates were aggregated.

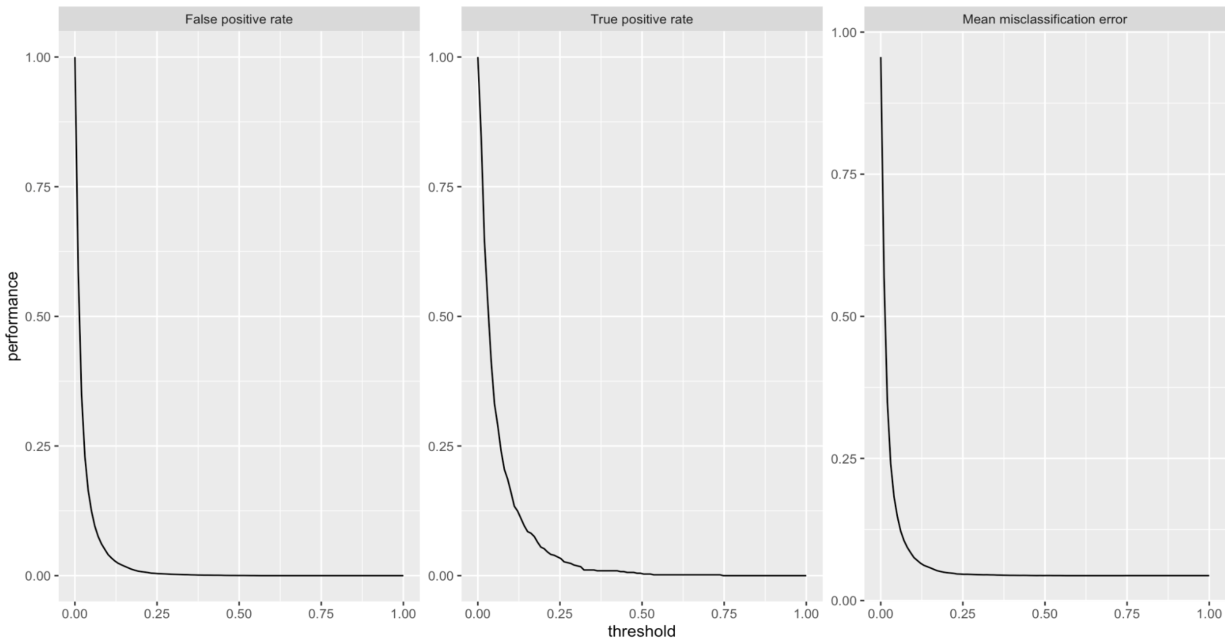

Figure 5.1.2 Threshold-moving Plot for two class classifications

5.2 Variables importance

Variable importance refers to how much a given model "uses" that variable to make accurate predictions. The more a model relies on a variable to make predictions, the more important it is for the model. We try to evaluate the variable importance based on the XGBoost model. And here are the top 20 important variables we selected from XGBoost:

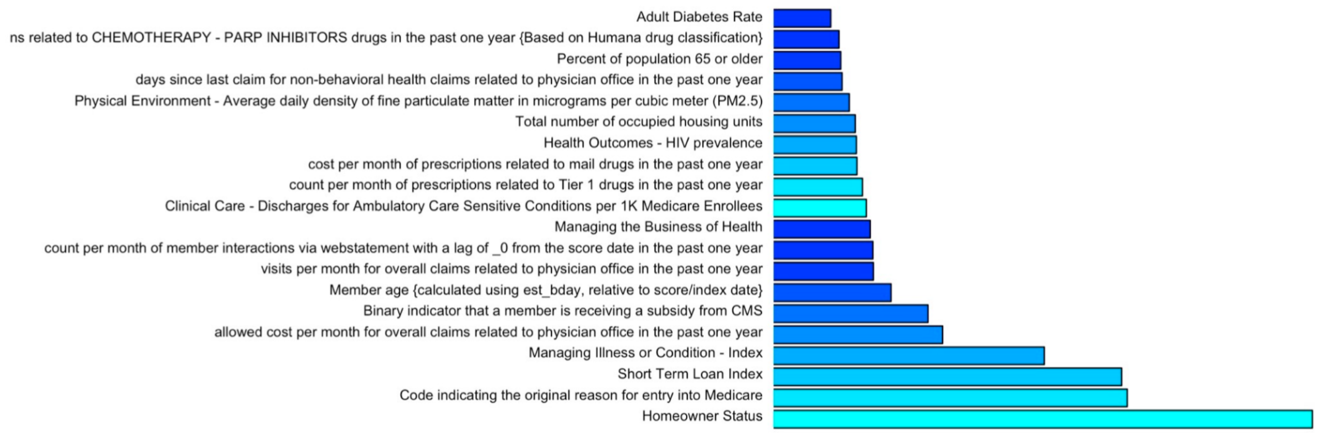

Figure 5.2.1 The Top 20 Variable Importance Barplot by XGBoost

Accordingly, “Homeowner Status” is the most important variable for the prediction of housing insecurity. Risk sense of the subjects is also an important dimension, features including “the original reason for entry into Medicare”, “Short Term Loan”, and “Managing Illness or Condition”.

In the category of risk sense, medical risk sense is a subcategory that we can not ignore. In the top 20 important variables, we have 5 variables that can imply the medical awareness of a person. In addition, there are also several features related to health status or healthcare frequency, for example, “visits per month for overall claims related to physician office in the past year”.

Also, some expenses features, including “allowed cost per month for overall claims related to physician office in the past one year”, “Binary indicator that a member is receiving a subsidy from CMS”.

Demographic features also have a great influence on the result. The most significant one is age. “Member age” and “Percent of population 65 or older” are the eighth and nineteenth most important variables in the list respectively.

5.3 Partial Dependence analysis

For the partial dependence analysis, we mainly focus on the marginal effect one or two features have on the predicted outcome of a machine learning model. We analyze some significant variables that might influence the outcome in the real world.

For the cost related features including: “cost per month for prescriptions in the past one year”, “allowed cost per month for overall claims related to physician office in the past one year” and “allowed cost per month for overall claims in the past one year”, we find out that lower cost input is related to higher risk of housing insecure.

As for the age indicators, we find out that in the group of people who are older than 65, with the increasing age, the risk of getting housing insecure is also increasing.

Also, for some health indicators like “diabetes rate”, we find that adults with worse health status would be more likely to suffer from housing insecurity.

A pretty unexpected result we find here is that the physical environment will also influence housing insecurity. People who live in a worse air-conditioned environment are slightly more likely to be housing insecure than people who live in a less polluted area.

5.4 Fairness Analysis

To ensure fairness of prediction in sensitive feature groups, we calculate the disparity ratio (DR) in each feature group subsetted by Race and Sex. The Disparity Ratio is defined as:

$$DR = \frac{S_n}{S_0}$$

Where S_n is the True Positive Rate (TPR) of each class, and S_0 is the TPR for reference group (Race: White, Sex: Male). The result showed that in our model, the DR in different groups were comparable (figure xxx) and no severe fairness problem was identified. Most of the groups could even predict better than the reference group. The disparity level could be further indicated with an overall Disparity Score (DS) across groups, calculated by averaging across all sensitive variables (SV with N total groups) the sum of each disparity ratio capped at 100%:

$$DS = \frac{\sum_{SV}\sum_{S} min(DR, 1)}{N}$$

Here we get a final score of 0.9976 (> 90%), suggesting an equal performance across sensitive feature groups.

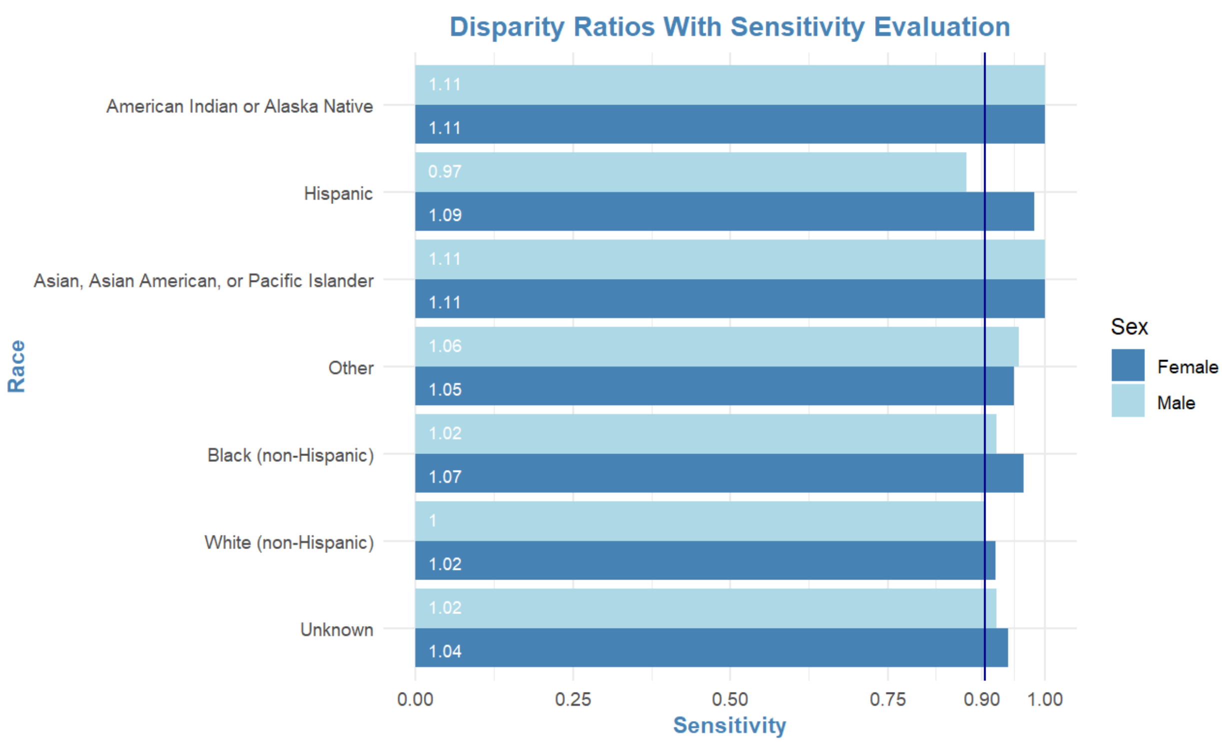

Figure 5.4.1: Disparity Ratio across Race and Sex Groups

6. Business Implication

6.1 Business impact in Important Variables

From the model analysis we have discussed in the previous chapter, we can tell that many key points drive the Housing insecurity, including social economics influencers, healthcare resource and environment conditions etc. Most of those important variables that cause high housing insecurity are more about the society rather than an individual. The lack of social support in the U.S. particularly in housing is one of the key effects. "In theory, if a household income is less than 50% of the area median income (AMI), they can get housing assistance. But in reality, only one in four receive it. The housing dimension of the social safety net is broken," said Dr. Khadduri. Thus, companies need to focus more on the macro environment, the key organizations and government, rather than specific cases.

In addition, housing insecurity is more than a social problem. We can tell from the analysis that there are many health status indicators that are highly related to housing insecurity. For example, the rate of adult diabetes is a crucial predictor of housing insecurity. Based on the result of partial independent analysis, people who have a higher rate of diabetes will lead to high risk of housing insecurity. We can conclude that people who are housing insecure are often dealing with serious medical and behavioral health conditions. Even though we can not identify the causal relationship between insecurity and poor health, we can explain this relationship in both directions: Housing insecurity could lead to poor health, and some illnesses might be a cause of housing insecurity. There are some people that might be bankrupted because of unaffordable medical care and become homeless. There are also people who might get a high risk of chronic or severe illness because of the bad housing conditions. Thus, when it comes to a business decision, we need to make plans in a combination of both health status and housing insecurity. Solving one without the other one won't work.

Another insight for business decisions is that we need to do stratification on the people. Different groups of people may have different obstacles and different needs. In our model analysis, we observed that age is a significant feature of housing insecurity. Older people might face a higher risk of housing insecurity. It is quite understandable since older people always have less ability to earn money and have feeble bodies as well as numerous potential diseases. Also, according to a report from the U.S. Department of Housing and Urban Development, many college students are also getting housing insecurity (U.S. Department of Housing and Urban Development, 2015). Some of those college students are forced to be homeless because they can not find a place to live in the city of their college. The two phenomena above show that every group of people are having its own problems. Thus, stratification for our future products is a really vital choice.

6.2 Potential Impacts on the Economy

Housing insecurity is an economic issue. This issue can be contributed to multiple factors: high housing costs relative to income, poor housing quality, unstable neighborhoods, overcrowding, homelessness, and more (Amy Johnson & Meckstroth, 1998). We were trying to determine what housing insecurity may result in the U.S. economy. In order to ascertain this topic, we analyzed the impacts of housing insecurity on three dimensions.

6.2.1 Income Inequity

Housing insecurity leads to unstable neighborhoods and education gaps. As housing costs go up, people who cannot afford costly areas choose to move further out. Moving away from hot markets may be stopping people from working in locations with high wages. In the meanwhile, the low cost of housing brings less-educated and lower-income households to the community. Less tax generation leads to less funding for public schools. Low salary forces teachers to leave the community. Children who grow up in communities with unstable neighborhoods and poor education quality are more likely to quit school at an early age. Under-educated workers find it hard to find high-income jobs and harder to move to good communities. As time goes by, income inequity is greatly exacerbated.

6.2.2 Unemployment rate

Housing insecurity also triggers an increased unemployment rate. Communities with low housing costs obtain relatively less-educated and low-income workers. Without paid or unpaid leave, low-income workers may find it challenging to take time off to cope with the eviction process and locate a new home without compromising job performance. They are more likely to lose their jobs, and if they do, they frequently struggle to obtain new ones (Benjamin Keys and Danzinger, 2008). During the past few years of pandemic, many jobs were transmitted from in-person work environments to work from home patterns. These jobs tended to require high-educated and high-experienced employees, most likely in technical areas. Traditional industries especially retails and restaurants had severe financial crises during the pandemic. Most jobs located around low-cost housing communities are retail based. People who live in these communities ultimately face a higher risk of unemployment.

6.3 Solution

The main causes of housing insecurity are the lack of adequate safe housing and the high cost of available housing. Therefore, to address housing insecurity, it is natural to think about building more safe housing. The government should reduce or remove regulatory barriers to building new homes and apartments. For example, previously planned single-family homes should be rezoned. Rezoning single-family homes to duplexes and triplexes could help to provide more living space in limited land and reduce the cost of housing.

In addition, some new technologies can be utilized in construction. In recent years, the price of construction materials, such as steel, has increased dramatically. The import and export of many materials has also been significantly delayed due to the impact of the pandemic. As a result, the cost of building traditional housing has skyrocketed, which can make housing more unaffordable. However, with the development of technology, more and more new materials are being used. For example, 3D printing housing is a new type of housing construction being developed. The advantage of 3D printing compared to traditional house building techniques is that it is very material efficient. Because the printing technology itself offers a wide variety of unique shapes, 3D printing reduces the amount of material that needs to be cut and pieced together in the traditional house building process. The use of different printing materials can also be resistant to earthquakes, warmth, water release and other different effects. Therefore, safety can also be guaranteed. The government can provide more safe housing at a lower cost in a shorter period of time by building more innovative housing, such as 3D printing housing.

In addition to the housing itself, a safe and well-developed community is essential to housing safety. According to our research, people in poor health are more likely to fall into housing insecurity. Therefore, providing them with more medical help in the community can help improve their health and reduce the financial burden of their health problems, thus reducing their housing stress.

In addition to the housing itself, a safe and well-developed community is essential to housing safety. According to our research, people in poor health are more likely to fall into housing insecurity. Therefore, providing them with more medical help in the community can help improve their health and reduce the financial burden of their health problems, thus reducing their housing stress. At the same time, a well-established community brings more jobs and boosts the regional economy. With sufficient income, people are less likely to be impacted by housing insecurity.

6.4 Humana Insight

Humana's Bold Goal social health initiative seeks to provide people with more Healthy Days and better lives by focusing on critical social determinants of health, including housing stability. In the past two years, Humana has investigated $50 Million in funding to increase the supply of affordable housing across the country. To maximize the housing security impact of funding and ensure equal opportunity and proper benefits to all vulnerable members and communities, we hereby adopted our model results into Humana's Housing strategies to provide more detailed insights into the particular insecurity problems.

6.4.1 Individual Housing Supportive Plan (IHSP) Based on Housing Insecurity Model

To identify the potential members who suffer from challenging times in housing security problems, we could apply our predictive model to estimate an overall security score for every individual member and find the members that are most likely suffering from the housing insecurity issues. An Individual Housing Supportive Plan (IHSP) could be provided to those with high probability of housing problems to obtain additional personal funding and supportive services, including but not limited to, providing house renting subsidy, keeping active connection with housing service expertise and advisors. The priority and enrollment eligibility of the service list of millions of members could also be suggested based on the model evaluation probability. In specific, we also identified several groups that need particular concern on the housing security issue:

- **People with high Short-term Loan**
  According to our model, we identified that with the “cons_stlindex” column, people with higher short-term loans (as indicated with smaller index) have higher probability to be in the situation of housing insecurity, which makes sense since housing cost is always the large burden in people’s daily expenses. Therefore, short-term loans could be settled as a candidate marker that is involved in the eligibility evaluation of IHSP.
- **People with less ability in managing illness or Condition**
  Similarly, the “cons_hxmioc” column also showed great importance in determining people’s housing insecurity status. In specific, people with higher demands in managing their illness conditions would have higher pressure in ensuring their housing status, since they might probably have difficulties with earning money due to their worse health condition. For those people, we should pay additional attention to help them go through the difficult time, especially providing related services and accommodations for them to get the best care and achieve the best health conditions.
- **People already received a subsidy from CMS**
  The people who have a record in receiving a subsidy from CMS are shown with higher probability to be in the housing insecure status. Therefore, Humana could provide additional housing subsidies to those with low incomes and limit the use of the subsidy in renting a house or providing access to built community accommodations.
- **People with younger age**
  It was surprising that a higher percentage of young people are suffering from the housing insecurity condition than the elder people. To provide additional support to these young people, we could co-considering the above situations and provide additional help to them, especially providing potential job opportunities to healthy young people or providing lower rental fees for them.

To ensure this additional housing subsidy could be fully used to secure member’s housing status, an overall enrollment eligibility should be applied and a clear indication and funding limitations should be declared in the plan. For example, people enrolled in the IHSP could have renting discounts or tax waiver abilities in specific accommodations. Or the subsidies could be provided within the member account and only when they want to pay for housing related services, the subsidies could be used.

By providing the IHSP plan, members could have more freedom on accommodating the funds to their realistic situation and optimize the use of the money. Also, they could have priority in resources on related housing services to relieve their burden. However, the drawback for Humana is that more human power and a developed system should be established, and this could take time.

The SWOT analysis of the plan thus lies in:

- **Strength:** With the support of large claim data of Humana members, we are able to provide a robust evaluation of housing insecurity status and rank the priority level based on the member's situation in all aspects. Besides, Humana has investigated $50 million in improving housing status, which allows the program to cover as many people as possible with an estimated average subsidy per member and a dynamic evaluation system to provide the proper amount of funds.
- **Weakness:** The model does not achieve 100% accuracy and additional criteria are needed for proper evaluation. Besides, adding a new nationwide individual plan could take great time and human effort to test if it works properly, extending the result delivery over a long time scale.
- **Opportunities:** Government has made great efforts to solve the housing security problem and we could establish active collaboration with government or non-profit associations to accelerate the plan promotion process and ensure the plan could cover those people in urgent need.
- **Threats:** Improper communication with members may lead to misuse of the funds and people with the most urgent needs might have less access to the individual plan due to their health, education, or willingness considerations.

By carefully dealing with the SWOT aspects, Humana could benefit from making a more individualized and precise plan to best accommodate their member's status in housing.

6.4.2 Community involvement

Community involvement positively impacts the community, supporting people in needs, and aspiring employees participate in volunteering opportunities. Community involvement is also a quick and efficient approach to build brand recognition, strengthen a solid reputation, and expand a company. Local businesses have closer access to customers than big chain companies. Comparing companies with similar products and services, community engagement is a critical way for residents to choose which company to trust. In addition, community engagement helps businesses establish long standing, effective partnerships with government organizations and public residents, benefiting in reputation development in not only communities but also nationwide. Housing insecurity plays a vital role in the country's economy, and society. Supporting communities to overcome housing insecurity secures communities' safety, eliminates income inequity, importantly boosts economic development, and most benefits Humana's long-term growth.

- **Volunteering**

  Volunteering is the most straight-forward way to get business involved in the community. In order to executive volunteering plans, Humana can support employees to participate in volunteering events by offering incentives and time off. Setting routine volunteer days for employees to involve in local shelters, food banks, church services encourages employees to pay more attention to people who are experiencing housing insecurity. It also allows employees to stay away from their daily work and meet people holding the same beliefs. Volunteering events increases employees' social network ability, and promotes involved people a sense of belonging. Volunteering is a fantastic opportunity to meet potential partners and future clients with the same beliefs and values. Sharing the same values helps businesses build collaboration in a faster way.

- **Fundraising**

  Providing financial support is the most effective way to support the community to solve housing insecurity. It is also a great way to build connections between businesses and governments. Humana can collaborate with local non-profit organizations to contribute a small portion of profit to support while non-profit organizations help businesses to build positive brand awareness in the local community. Building collaboration with the government is also an effective way to achieve a win-win situation. Governments can help Humana promote the IHSP plan while Humana supports particular local events to help eliminate housing insecurity and instability.

Solving housing insecurity issues does not only bring benefits to people who have housing insecurity, it also strengthens communities' stability, improving the living environment of employees, customers, and clients.

6.4.3 Partnership & collaboration

Social Impact Bonds (SIB) are a unique type of public-private partnership that invests in effective social services through performance-based contracts. If the project the bonds invest in achieves results that generate public value, the government repays those investors. Even though SIBs have only been issued by the public sector, Humana could still find some way to cooperate and expand its social responsibility. For example, because the return on investment for SIB is tied to the outcome of the research project, Humana can support the success of the research project by paying for the funding. In this way, Humana not only supports scientific research but through the SIB, enables the return on investment to be utilized for housing security-related projects such as housing construction. In the long run, this investment approach will also help Humana improve its reputation.

6.4.4 Stakeholder profit analysis

Eliminating housing insecurity provides a more stable community. A healthy community supplies a good living and work environment to both employees and potential customers. Solving housing insecurity also benefits the economy, avoiding people spending the extra money to migrate. Spare money in hand furnishes people with the ability to consume what they want. Partnering with local businesses, involving in communities, and promoting IHSP plans help Humana broaden brand awareness, increase social impact, attract more customers, and enlarge business services. These increase the demand of Humana's services and products. Based on supply and demand balance, in order to fill increased demand, supply is then increased, and ultimately product price can be raised up. These provide more profit to the company executives. With more profits, executives are capable of executing more business plans. In the meanwhile, directors and managers are able to collaborate with more clients and hand on more diverse projects. Employees will be able to manage more cases and gain comprehensive experience. This trend ultimately surges stock price, positively impacts stockholders and overall the economy.

7. References

Amy Johnson, & Meckstroth, A. (1998, June 21). *Ancillary Services to Support Welfare to Work*. ASPE. https://aspe.hhs.gov/reports/ancillary-services-support-welfare-work

avcontentteam. (2016). *Imbalanced Classification Problems in R*. https://www.analyticsvidhya.com/blog/2016/03/practical-guide-deal-imbalanced-classification-problems/

Baggett, T. P., Hwang, S. W., O'Connell, J. J., Porneala, B. C., Stringfellow, E. J., Orav, E. J., Singer, D. E., & Rigotti, N. A. (2013). Mortality Among Homeless Adults in Boston: Shifts in Causes of Death Over a 15-year Period. *JAMA Internal Medicine*, *173*(3), 189–195. https://doi.org/10.1001/jamainternmed.2013.1604

Department of Housing and Urban Development. 2015. Barriers to Success: Housing Insecurity for U.S. College Students U.S. https://www.huduser.gov/portal/periodicals/insight/insight_2.pdf

Jorly, J. (2021, June 5). *XGBOOST — IN A NUTSHELL*. Medium. https://ai.plainenglish.io/xgboost-in-a-nutshell-211e170e8b48

Keys, Benjamin and Sheldon Danzinger. 2008. "Hurt the Worst: The Rise of Unemployment among Disadvantaged and Advantaged Male Workers, 1968-2003." Pp. 56–73 in Laid Off, Laid Low: Political and Economic Consequences of Employment Insecurity, edited by Katherine Newman. New York: Columbia University Press.

Medical membership of Humana 2021. (n.d.). Statista. Retrieved October 15, 2022, from https://www.statista.com/statistics/210651/total-medical-membership-of-humana/

Sharma, H. (2021, June 6). XGBoost- The Miracle Worker. *AlmaBetter*. https://medium.com/almabetter/xgboost-the-miracle-worker-7518dd55abf3

8. Appendix: List of Categorical Variables

"cons_mobplus"	Mail Order Buyer
"cms_ra_factor_type_cd"	Code indicating the type of risk adjustment factors in use for a member
"cons_homstat"	Homeowner Status
"cms_orig_reas_entitle_cd"	Code indicating the original reason for entry into Medicare
"sex_cd"	Member gender
"lang_spoken_cd"	Preferred language for member
"rucc_category"	Member geographic information - Rural Urban Continuum Code
"cms_race_cd"	Code indicating a member's race
"cms_disabled_ind"	Binary indicator that a Medicare Supplement member is under age 65
"cons_hxmioc"	Managing Illness or Condition - Index
"cons_hxmboh"	Managing the Business of Health
"cons_stlnindx"	Student Loan Index
"cmsd2_men_mad_ind"	a binary value indicating if a service related to mental, behavioral and neurodevelopmental disorders : mood [affective] disorders was performed in the past one year {based on CMS diagnosis code level2}

"cms_dual_eligible_ind"	Binary indicator that a member is eligible for Medicaid and Medicare with PartD Low Income Multiplier
"cons_stlindex"	Short Term Loan Index
"cms_low_income_ind"	Binary indicator that a member is receiving a subsidy from CMS
"cons_hxmh"	Managing Health - Index
"cms_frailty_ind"	Binary indicator that a member is deemed frail {specific diagnoses, multiple serious chronic conditions, functional impairments or other factors}