Humana – Mays
2022 Healthcare Analytics Case Competition

Predicting Housing Insecurity

Table of Contents

TABLES AND FIGURES

Executive Summary

Humana, a leading health insurance provider in the United States, is committed to ensuring that social needs do not determine health outcomes. One of the most concerning Health Related Social Needs (HRSN) is housing insecurity. In 2019, more than 30% of homeowners were paying more than 30% of their monthly income on rent or mortgage payments. Housing insecurity disproportionately affects certain social groups. In order to create more equitable access to health care, Humana is attempting to better predict housing insecurity based on available individual customer health and financial attributes.

Housing Insecurity has worsened in recent years, driven by multiple socio-economic factors. Rapidly rising home prices, housing shortages, and inflation are just a few of the issues contributing to the growing prevalence of housing insecurity. Housing insecurity is a broad term that encompasses multiple qualifying factors. They include criterion such as ratio of housing cost and income, substandard housing, presence of sub-families, homelessness and more.

The objective of our analysis is to use data to predict medicare member housing insecurity, illuminate important contributing factors, and provide recommendations to help housing insecure members achieve their best health.

We approached the problem in four phases:

1. Initial Data Review and Exploratory Analysis
2. Domain Research
3. ML Modeling
4. Interpretation and Recommendations

We created an initial base model that took into account all 800+ variables. We improved upon that model by examining prior research on medical conditions with the highest patient cost burden, common health issues among the housing insecure, including mental health diseases, and investigating early life experiences associated with level of housing. Based on findings from this domain research, we grouped the variables into categories, examined correlation between them and built a set of predictors that we felt was most appropriate for further analysis. Feature importance calculation via model output was also used to provide the highest accuracy with minimal model features.

Of all models used, a random forest model of our curated set of predictors produced the highest AUC value on our validation set at 0.6952. Other models used include XG Boost, Ada Boost and Keras DNN. Results of all models are discussed in detail in subsequent sections of this paper.

Our recommendations include focusing on four primary areas:

- Actions to remediate housing insecurity risks for specific social groups as identified in this analysis at a societal level
- Actions to remediate housing insecurity risks for specific social groups as identified in this analysis at a societal level
- Actions to remediate indirect health effects of housing insecurity
- What can Humana do about it?

Dataset Overview

Each row of the dataset corresponds to an individual Medicare Customer in the Humana system in which each column is a specific attribute. The dataset consists of two tables, one for training and one for testing (holdout) our predictive models. The training set consists of 48,300 records of 881 variables. The test set consists of 12,220 records of 880 variables. The training set includes a binary column "hi_flag" which corresponds to whether or not the individual is deemed housing insecure. This column will be used as the dependent variable when training models.

The dataset consists of multiple datatypes, including numeric, binary, and multifactor categorial attributes. Numeric variables account for the vast majority of variables, with a total count of 865. Binary and categorical variables each account for eight variables.

After building the first model with all variables, we created a pivot table of column categories. The table is shown below:

Table 1 - Variable Prefix Pivot Table

Prefix	Description	Feature Count
atlas	Economic and poverty situation	9
bh	behavioral health claims related variables	54
cci	count per month of claims related to Charlson Comorbidity Index	20
cms	Varibles related to Medicare plan	14
cmsd1 & 2	Claims based on CMS diagnosis codes	305
cnt	Number of member interactions with lag	65
cons	Regional census data	7
credit	Customer credit information	10
med	days since last claim for non-behavioral health claims	12
rev	Claim line details based on revenue code for hospital services	74
rwjf	Regional healthcare and economic factors	21
rx	Prescriptions costs and counts	234
total	Cumulative totals of various other categories	44

We segmented columns based on their relation to common risk factors associated with housing insecurity. We then developed a set of final curated features that reflect the most important contributing factors and filtered out highly correlated independent variables.

Dataset Imbalance:

The training data set is highly imbalanced with respect to "hi_flag". The majority class has no housing insecurity and accounts for 96% of all observations. Accounting for this imbalance will be a critical step in the modeling phase.

Exploratory Analysis

We examined distributions of multiple variables families as shown below

Age: Most of members in the dataset are aged from 65 to 80 years. Distribution of members using estimate age feature is show below and follows a near normal distribution.

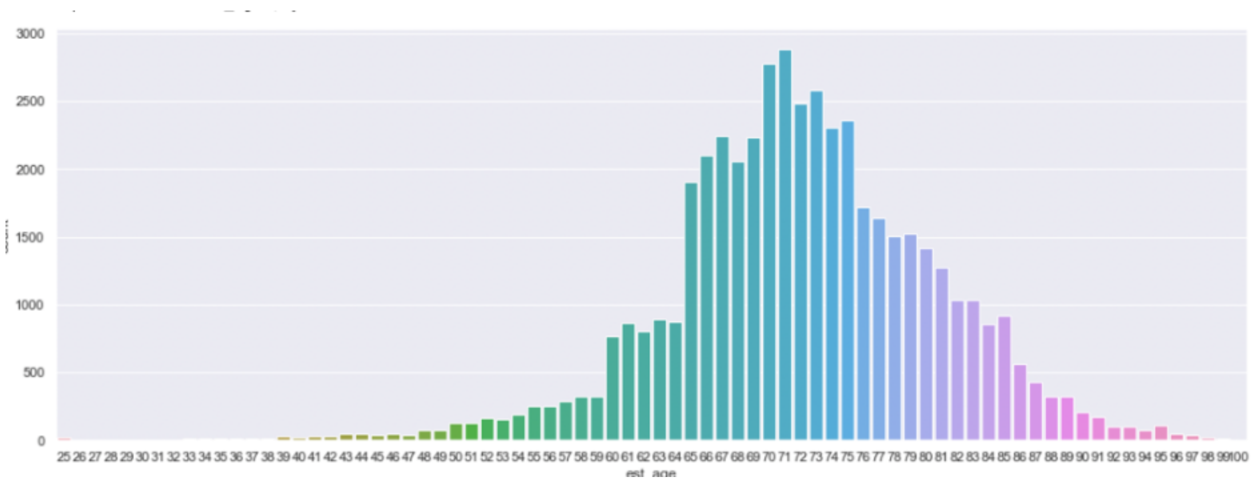
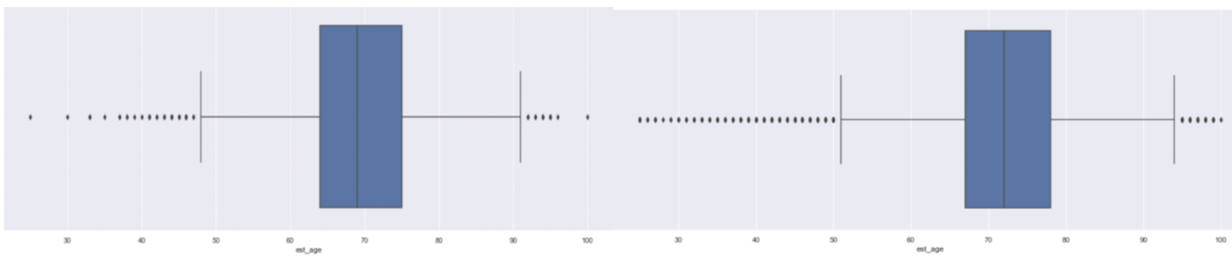


Figure 1 - Age Distribution of all Members



Box plot for housing insecurity flag =1

Box plot for housing insecurity flag =0

Figure 2 - Age and HI Flag, Box Plot

RACE: Majority of the members are white(non Hispanic) followed by black- non Hispanic. Asian, Asian American, or Pacific Islander are the least in count from the entire training dataset

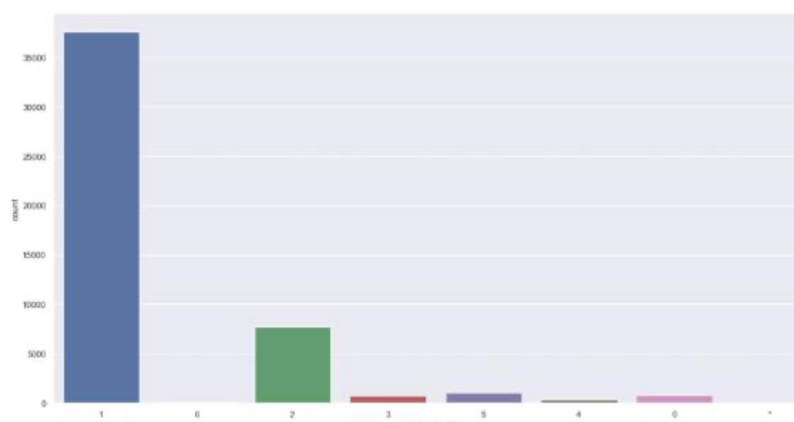


Figure 3 - Race Distribution

RUCC_CATEGORY: It is the member geographic information based on rural urban continuum code. Members in the counties with metro areas of 1 million population or more are much higher than the other metro counties. Whereas in non-metro counties, urban population of 2,500 to 19,999, adjacent to a metro has most members

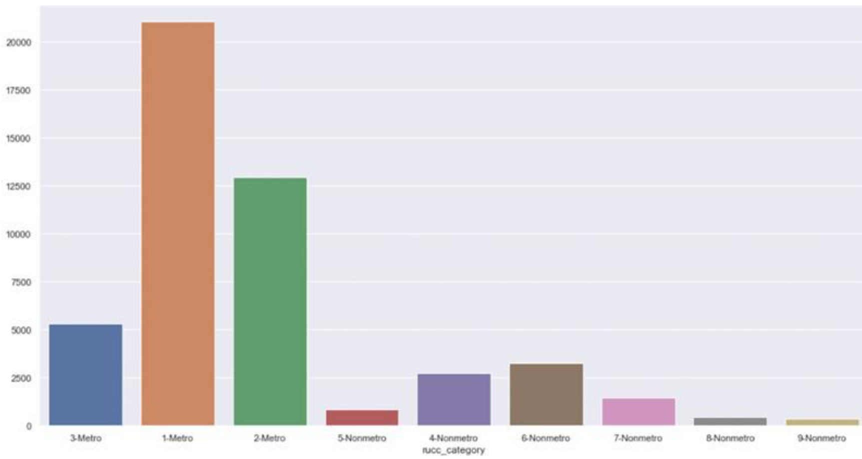
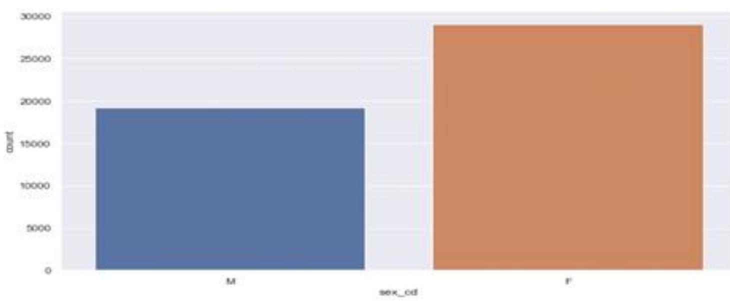


Figure 4 - Rucc Category Distribution

GENDER: Majority of members in the dataset are female. However, percentage of members that are housing insecure is similar between the two gender groups which is around 4 to 5%



	sex_cd	hi_flag	id
0	F	0	27955
1	F	1	1145
2	M	0	18227
3	M	1	973

Figure 5 - Gender Distribution

Prior Research

We examined numerous existing publications and reputable websites with two primary intentions; First, to better understand the context, breadth and significance of the problem. Second, to understand the variables in our dataset and which are important to consider during modeling.

Context, breadth and significance of the problem:

From 2019 to 2020, 30% of households were paying 30% or more of their monthly income on rent or mortgage payments. 14%+ were paying more than 50% house prices soared over the last two years leading to increased cost burden for those earning $30,000 - $45,000/year. Since March 2021, housing prices skyrocketed 20.6% and rent climbed 12.1% (Habitat for Humanity, 2022). As a result, we can expect housing insecurity numbers to climb even higher.

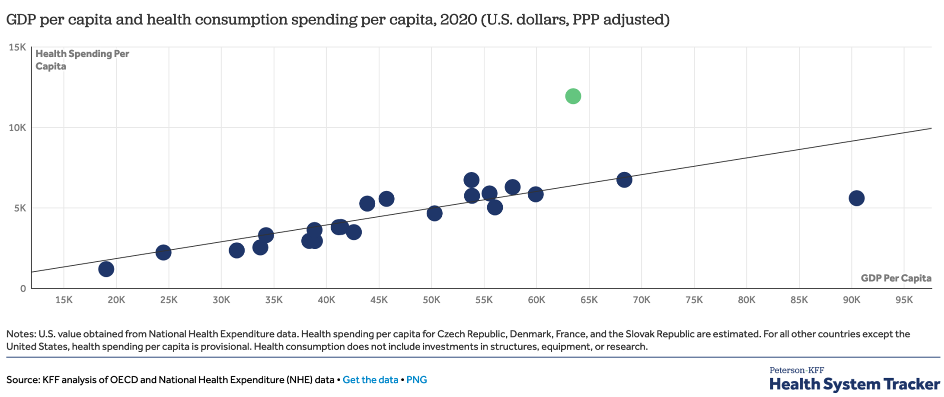


Figure 6 - GDP Per Capita vs Expenditure

Healthcare costs account for 19% of the total U.S. GDP, far higher than any other high-income nation. On average the typical American is spending in excess of $11,000 per year on healthcare costs (Twitter, E. W et al, 2022). Out of pocket health expenditure increased 10% in 2021. (Muoio, D, 2021)

Chronic illnesses and mental health conditions account for 90% of the nation's $3.8 trillion annual healthcare expenditure. The five most expensive conditions are neurological, gastrointestinal, musculoskeletal, heart disease, and cancer. (Murtha, J., 2022)

People struggling with housing insecurity are often suffering from one or more chronic medical conditions. Common conditions include substance abuse, mental health disease, respiratory issues, dental disease and more. Some important lifetime risk factors are shown in the table below.

Frequencies and Percentage of Present-Day Risk Factors for the Marginally Housed and Homeless Groups

Risks Factors	Marginally Housed (*n*=49)	Homeless (*n*=51)
	n (%)	*n* (%)
Sex (Male)	33 (67.3%)	46 (90.2%)
Single	29 (59.2%)	48 (94.1%)
Age		
22–30	14 (28.6%)	16 (31.4%)
31–40	12 (24.5%)	18 (35.3%)
41–50	10 (20.4%)	11 (21.6%)
51–64	13 (26.5%)	6 (11.8%)
Social skills problems	9 (18.4%)	11 (21.6%)
Substance-related problems	14 (28.6%)	15 (29.4%)
Family problems	15 (30.6%)	14 (27.5%)
Level of disability with/without benefits		
No disability	20 (40.8%)	30 (58.8%)
Disability no pension	4 (8.2%)	6 (11.8%)
Physical disability pension	7 (14.3%)	7 (13.7%)
Mental illness pension	8 (16.3%)	2 (3.9%)
Drug-related physical problems pension	4 (8.2%)	1 (2.0%)
Intellectual disability pension	5 (10.2%)	4 (7.8%)
Physical and mental disability pension	1 (2.0%)	1 (2.0%)
Vocational qualifications	35 (71.4%)	27 (52.9%)

Figure 7 – Frequencies and Percentage of Risk Factors (Braver, T., 2012)

Furthermore, housing insecurity disproportionately affects minorities. According to the CEPR, the likelihood that Hispanic and black families were housing insecure rose 11% and 7% in 2019 respectively. (Cai, J., 2021) The figure below provides a visual representation of the trend in housing security as it affects different racial groups.

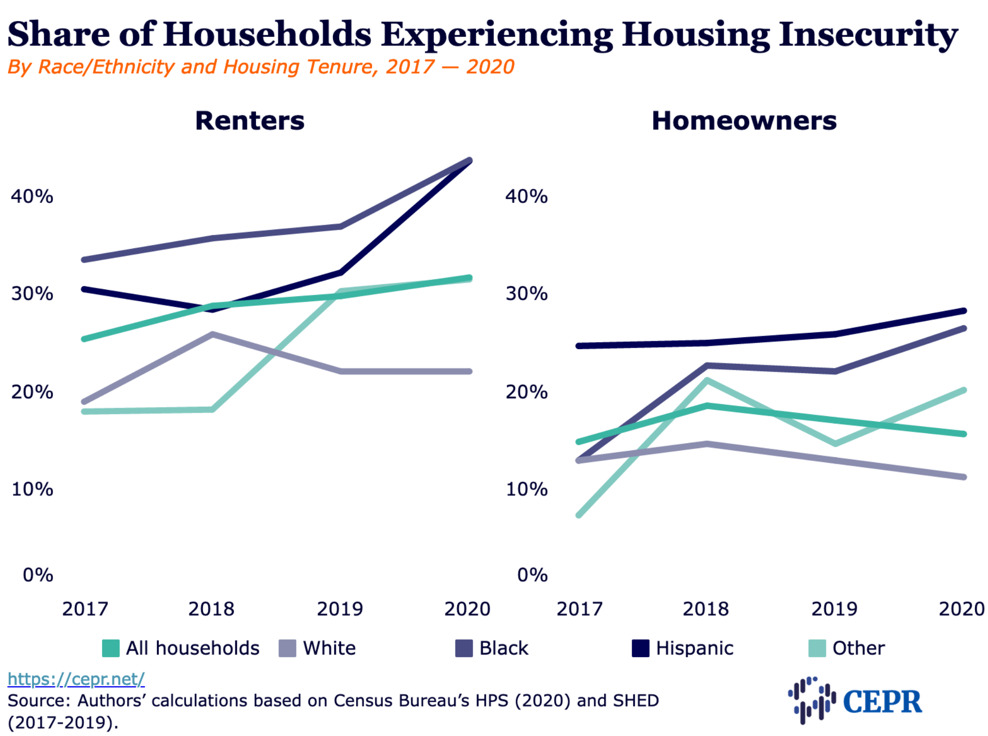


Figure 8 – Share of Households Experiencing HI by Race

Understanding Important Variables and Selection

Based on a deeper understanding of financial variables, high-cost health issues, common health problems associated with housing insecurity, and lifetime risk factors we created attribute "families" of and picked relevant features. We focused on including important financial risk factors like Credit balance amount overdue by 60 days, Substance abuse and mental health related variables, and high-cost chronic health conditions to name a few. Out of the 880 variables, we shortlisted 117 variables.

Although non-linear regression-based machine learning models are generally able to handle multicollinearity, some of the benefits of Random Forest and other model types are diminished with the presence of multicollinearity. Specifically, their ability to detect interactions between features can be masked. As a result, we used a combination of domain knowledge and collinearity matrices to include/exclude variables within a given family. Selected Variables are shown in the table below

Table 2 - Variable Selection

Prefix	Feature Columns Selected
atlas	atlas_age65andolderpct2010,atlas_naturalchangerate1016,atlas_net_international_migration_rate,atlas_orchard_farms12,atlas_pct_diabetes_adults13,atlas_snapspth16,atlas_totalocchu,atlas_totalpopacs,atlas_totalpopest2016
bh	bh_anxiety,bh_chemical_dp,bh_depression,bh_depressionbh_nc_disorder,bh_nc_disorder,bh_neuro_dev_disorder,bh_otsl_pmpm_ct,bh_suicide_ct
cci	cci_chronic,cci_liver,cci_mlg_pmpm_ct,cci_renal,cci_score
cms	cms_disabled_ind,cms_dual_eligible_ind,cms_low_income_ind,cms_orig_reas_entitle_cd,cms_ra_factor_type_cd_CF,cms_ra_factor_type_cd_CN,cms_ra_factor_type_cd_CP,cms_ra_factor_type_cd_E,cms_race_cd
cmsd1	cmsd1_bld_pmpm_ct,cmsd1_can_pmpm_ct,cmsd1_cir_pmpm_ct,cmsd1_dig_pmpm_ct,cmsd1_ear_pmpm_ct,cmsd1_end_pmpm_ct,cmsd1_ext_pmpm_ct,cmsd1_eye_pmpm_ct,cmsd1_gus_pmpm_ct,cmsd1_inf_pmpm_ct,cmsd1_inj_pmpm_ct,cmsd1_men_pmpm_ct,cmsd1_mus_pmpm_ct,cmsd1_ner_pmpm_ct,cmsd1_res_pmpm_ct,cmsd1_skn_pmpm_ct,cmsd1_sns_pmpm_ct,cmsd1_unc_pmpm_ct,cmsd1_vco_pmpm_ct
cmsd2	cmsd2_eye_lacrimal_pmpm_ct,cmsd2_eye_vitreous_pmpm_ct,cmsd2_gus_urinary_other_pmpm_ct,cmsd2_men_mad_ind,cmsd2_skn_radiation_pmpm_ct,cmsd2_sns_digest_abdomen_pmpm_ct,cmsd2_sns_skn_pmpm_ct
cnt	cnt_cp_emails_pmpm_ct,cnt_cp_print_pmpm_ct,cnt_cp_vat_pmpm_ct,cnt_cp_webstatement_0,cnt_cp_webstatement_10,cnt_cp_webstatement_2,cnt_cp_webstatement_pmpm_ct
cons	cons_ccip,cons_homstat,cons_hxmh,cons_stlindex,cons_stlnindx
med	med_ambulance_ds_clm,med_er_ds_clm,med_ip_acute_ds_clm,med_ip_snf_ds_clm,med_outpatient_ds_clm,med_physician_office_ds_clm,med_urgent_care_ds_clm
rev	rev_pm_asc_pmpm_cd_ct,rev_pm_lab_pmpm_cd_ct
rwjf	rwjf_air_pollute_density,rwjf_child_mortality,rwjf_dentists_ratio,rwjf_drinkwater_violate_ind,rwjf_food_env_inx,rwjf_hiv_rate,rwjf_homicides_rate,rwjf_income_inequ_ratio,rwjf_median_house_income,rwjf_men_hlth_prov_ratio,rwjf_mv_deaths_rate,rwjf_pcp_rate,rwjf_poor_men_hlth_days,rwjf_

	poor_phy_hlth_days,rwjf_population,rwjf_premature_death_rate,rwjf_premature_mortality,rwjf_preventable_ip_rate,rwjf_std_infect_rate,rwjf_teen_births_rate,rwjf_violent_crime_rate
rx	rx_bh_pmpm_ct,rx_branded_pmpm_cost,rx_days_since_last_script,rx_generic_pmpm_cost,rx_hum_56_pmpm_ct,rx_hum_91_pmpm_cost,rx_maint_pmpm_ct,rx_nonmail_pmpm_cost,rx_nonmaint_pmpm_cost,rx_overall_pmpm_cost,rx_pharmacies_pmpm_ct
other	sex_cd,total_physician_office_allowed_pmpm_cost,dcsi_score,est_age,prov_line_pmpm_cnt

Feature engineering:

We created 10 new derived features from the existing set of available columns. They are summations of multiple critical factors that correlate to housing insecurity. The derived features and the features from which they were created are shown in the table below.

Table 3 - Derived Features

Derived Feature	Sub Features
bh_anxiety	bh_agad_pmpm_ct,bh_aoth_pmpm_ct,bh_apan_pmpm_ct,bh_apho_pmpm_ct,bh_atad_pmpm_ct,bh_adtp_pmpm_ct,bh_atot_pmpm_ct
bh_chemical_dp	bh_cdal_pmpm_ct, bh_cdsb_pmpm_ct,bh_cdto_pmpm_ct
bh_depression	bh_dema_pmpm_ct,bh_deot_pmpm_ct
bh_nc_disorder	bh_ncml_pmpm_ct,bh_ncot_pmpm_ct
bh_neuro_dev_disorder	bh_ndad_pmpm_ct,bh_ednd_pmpm_ct,bh_ndid_pmpm_ct, bh_ndlr_pmpm_ct,bh_ndot_pmpm_ct
bh_suicide_ct	bh_suat_pmpm_ct,bh_suid_pmpm_ct
cci_chronic	cci_dia_c_pmpm_ct,cci_dia_m_pmpm_ct
cci_liver	cci_lvr_m_pmpm_ct,cci_lvr_s_pmpm_ct
cci_renal	cci_ren_m_pmpm_ct,cci_ren_s_pmpm_ct
credit_bal	credit_bal_nonmtgcredit_60dpd,credit_bal_consumerfinance_new,credit_bal_bankcard_severederog,credit_bal_autobank_new

Modeling and Outputs

The main objective of our machine learning model is to identify the members that are most likely to be housing insecure. The training data set provided is highly imbalanced with the target variable “hi_flag” that determines the class of housing insecurity. Regular classification models would not perform well because we are interested in predicting minority class, a positive class that constitutes only 4% of the overall dataset.

For this case, we explored various ensemble models like XGBoost and AdaBoost to better account for the class imbalance. We also implemented a Deep Neural Network using the Keras package. The performance comparison of these models is discussed in the later part of this section. Here we focus on implementation methodology of two of our top performing models: XGBoost and Balanced Random Forest Classifier.

For model testing we employed the use of a “validation” set. We split the training data using a 70%-30% split. This allows us to use our validation set as a mock test set to evaluate model effectiveness.

We focus primarily on optimizing AUC value as it is the primary evaluation metric being used by the judging panel and it accounts for both precision and recall.

XGBoost

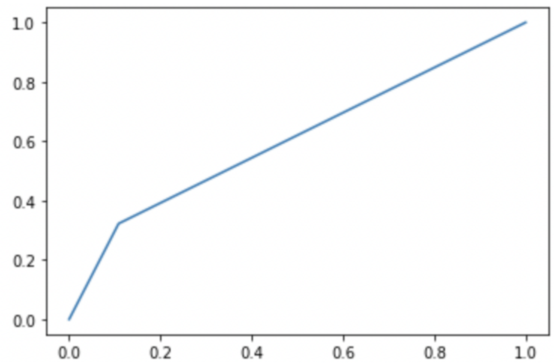
It is an optimized distributed gradient boosting technique using stochastic gradient boosting algorithm. Using an ensemble technique, more trees are added to existing trees to correct error gradients in this model. The model can be further tuned with the help of hyper parameter tuning that provides a complete control of training methodology. scale pos weight , a key parameter is used to adjust the model to give minority class a greater consideration.

Important XGBoost Model parameters:

learning_rate (0.01)
It controls the speed of learning. A low learning rate helps ensure the model doesn’t miss local minima.
max_depth (4)
It is the maximum depth of a tree. Higher values will tend to produce strong results on the traning set, they make the model susceptible to overfitting. The objective is to find the “sweet spot” – the lowest value that produces the highest test set accuracy.
n_estimators (1100)
The number of trees in the forest
scale_pos_weight(25)
Used to balance the positive and negative weights of the unbalanced class which can increase the overall performance (AUC)

AUC ,Confusion matrix and other metrics

Tuned Balanced Random Forest Classifier produced the better performance over the other models with **0.6069 AUC value**.



	True Positive	True Negative
Predicted Positive	6341	771
Predicted Negative	164	78

Measure	Value
Sensitivity	0.9748
Specificity	0.0919
Precision	0.8916
Negative Predictive Value	0.3223
False Positive Rate	0.9081
False Discovery Rate	0.1084
False Negative Rate	0.0252
Accuracy	0.8729
F1 Score	0.9313
Matthews Correlation Coefficient	0.1194

Balanced Random Forest Classifier:

Regular Random Forest model is an ensemble model that uses unpruned trees using random feature selection for each induced tree. The final prediction is the aggregation of all prediction based on majority voting. To avoid the absence of one class during bootstrap sampling, this ensemble method uses a sampling with replacement form each class during model training. This adaption helped the random forest model to be more robust to imbalance datasets

Key parameters used to build the BalancedRandomForestClassifier

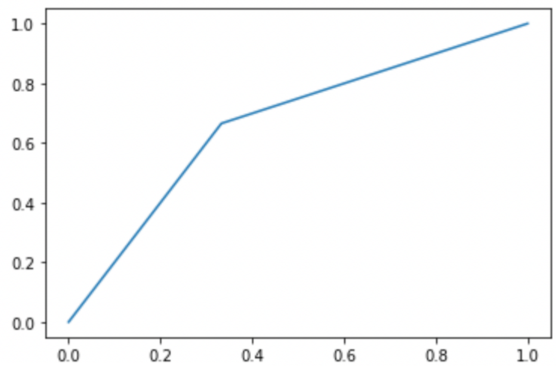
n_estimators (800)
Number of trees in the model
class_weight (balanced_subsample)
Weights are calculated using the sample drawn for every tree in the ensemble and adjust inversely proportional to the frequencies of the classes.
min_samples_split (2)
Minimum number of samples to split the node
min_samples_leaf (2)
Minimum samples to select for each leaf
max_features (sqrt)
Maximum features to consider during split. Sqrt uses square root of the total features available
max_depth (5)
Maximum depth of the trees
criterion (gini)
Impurity function to measure the quality of split
n_jobs (-1)

To use all CPU cores to perform cross validation
Bootstrap (True)
Use bootstrap samples when building trees

AUC Curve, Confusion matrix and other metrics

Tuned Balanced Random Forest Classifier produced the better performance over the other models with **0.6958 AUC value**.

0.6958



	True Positive	True Negative
Predicted Positive	**4944**	**2168**
Predicted Negative	**60**	**182**

Measure	Value
Sensitivity	0.9880
Specificity	0.0774
Precision	0.6952
Negative Predictive Value	0.7521
False Positive Rate	0.9226
False Discovery Rate	0.3048
False Negative Rate	0.0120
Accuracy	0.6970
F1 Score	0.8161

Hyper Parameter Tuning

With the help of RandomizedSearchCV, optimized model parameters are identified by cross validated search. Unlike grid search, all parameters are not tried but a fixed number of random parameters are sampled.
Below is the list of parameters used to tune **BalancedRandomForestClassifier**

```
random_grid = {'n_estimators': [int(x) for x in np.linspace(start = 200, stop = 2000, num = 10)],
           'max_features': ['log2','sqrt'],
           'max_depth': [int(x) for x in np.linspace(5, 50, num = 11)],
           'min_samples_split': [2, 5, 10,15],
           'min_samples_leaf': [1,2, 4],
           'class_weight':['balanced_subsample'],
           'bootstrap': [True, False]}
```

Cross fold Validation:

A RepeatedStratifiedKFold library is used to perform K fold validation n times on the model in each repetition where Kfold divides the sample into k group of samples.
Parameters:
n_splits (5) - Number of folds
n_repeats (4) - Number of times to repeat cross validation
random_state (1) - generates the random state for each repetition for reproducible output

After tuning the models with hyperparameters from randomized search library, balanced random forest classifier produces the better performance values than any of our models. Hence, we used this model output to analyze our case study.

A comparison table of models is shown below:

Table 4 - Model Comparison

Model	AUC
XGBoost	0.606
AdaBoost	0.591
Keras	0.6424
Balanced Random Forest	0.6958

Explaining machine learning models

We examined the important features as shown on the SHAP explainer plot and model feature importance output. Top features from both are shown below

SHAP Explainer Output from XG Boost Model

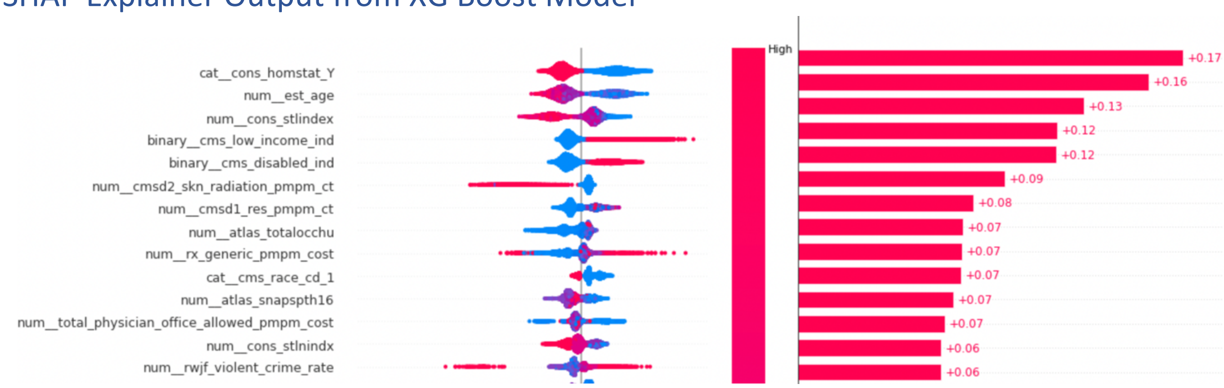


Figure 9 - SHAP Explanatory Plot (XG Boost)

Table 5 - Feature Importance Table (Random Forest)

Important Features from Random Forest	Feature Description	Importance Value
cons_stlindex	Short Term Loan Index	0.04507
cons_homstat	Homeowner Status	0.04035
est_age	Member age {calculated using est_bday, relative to score/index date}	0.03464
cms_disabled_ind	Binary indicator that a Medicare Supplement member is under age 65	0.03153
cms_orig_reas_entitle_cd	Code indicating the original reason for entry into Medicare	0.02458
cons_stlnindx	Student Loan Index	0.02384
cms_low_income_ind	Binary indicator that a member is receiving a subsidy from CMS	0.02068
rwjf_population	Demographics - Population	0.01858
atlas_totalocchu	Total number of occupied housing units	0.01788
med_physician_office_ds_clm	days since last claim for non-behavioral health claims related to physician office in the past one year	0.01522
total_physician_office_allowed_pmpm_cost	allowed cost per month for overall claims related to physician office in the past one year	0.01495
med_outpatient_ds_clm	days since last claim for non-behavioral health claims related to outpatient facilities in the past one year	0.01464
atlas_net_international_migration_rate	Net international migration rate 2010-2016	0.01443
rx_nonmaint_pmpm_cost	cost per month of prescriptions related to non maintenance drugs in the past one year	0.01440

The eight variables highlighted in green above are common between the two feature importance outputs. We looked further into each to understand how they might be affecting housing insecurity and found some very interesting mutual effects. Findings about the top 5 are discussed in further detail below.

Short Term Loan Index

Short term loan index refers to the likelihood that an individual has a short term loan. Short term loans are generally taken out to pay for basic necessities that correlate strongly with housing insecurity. Housing payments are the second most common necessity for which short term loans are taken out. The top eight necessities are shown in the table below

Table 6 - Use of Short Term Loans

Uses of short-term loans.

Use Category	*% (Frequency)*
Food	54% (33)
Housing	49% (30)
Utilities	41% (25)
Personal goods	38% (23)
Education	21% (13)
Vacation	21% (13)
Medical expenses	15% (9)
Child or dependent expenses	13% (8)

Furthermore, those with a history of short-term loans had significantly worse health across a range of measures. (Sweet, E. et al., 2018)

Homeowner Status

Homeowner status was a key feature in predicting housing insecurity. The highest count of housing insecure flagged medicare customers in the dataset were homeowners. This makes sense as mortgage payments have risen in recent years as noted earlier in the domain research portion of this paper.

Estimated Age

Age is a significant factor in determining housing insecurity. People aged 65+ are eligible for medicare. Once Medicare coverage kicks in, it can help shield seniors vulnerable to housing insecurity. (Bhat, AC., et al. 2022). This is validated by the fact that, the highest percentage of housing insecure people are between the ages of 32-65.

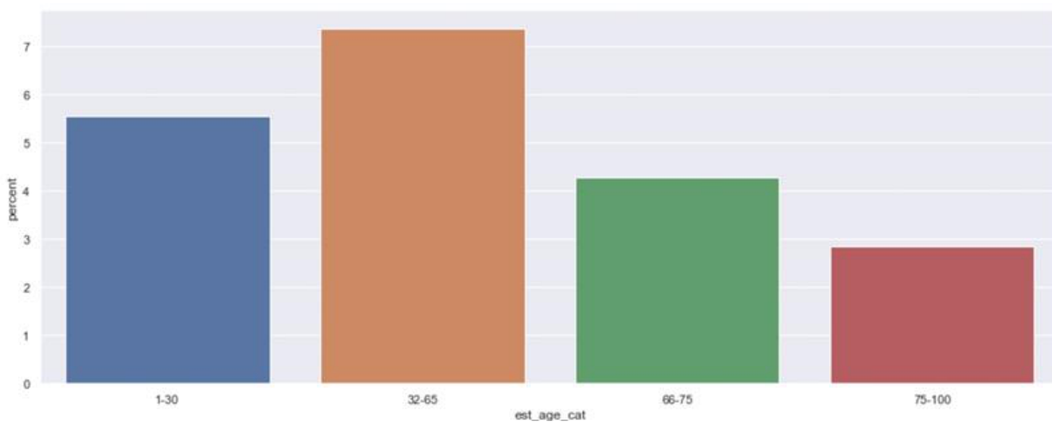


Figure 10 - Percentage of HI by Age Category

Student Loan Index

This variable corresponds to likelihood that an individual has a student loans. The higher the likelihood, the higher the index value. Our data indicates a clear trend that as the student loan index rises, so does the count of housing insecure people. This trend is depicted in the plot below.

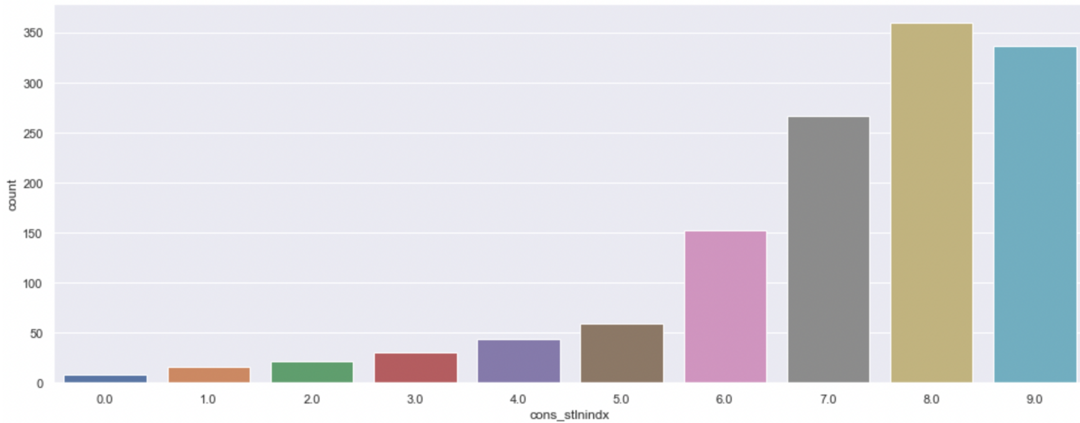


Figure 11 - HI Trend for Student Loan Index

Further investigation revealed two important things. Firstly, students are high risk individuals for housing insecurity. As noted by the Student Borrower Protection Center, "Forty-five million Americans now owe a combined $1.7 Trillion in student debt…impact of student debt ripples across borrower's financial lives forcing the to push off home ownership and other milestones" (Kaufman, B., 2021)

Furthermore, people with higher amounts of student debt reported higher levels of depressive symptoms, indicating that there is an interaction between student loan index and mental health diseases that needs to be further explored. Consumers older than 60 years with existing student loans are more likely to skip out on necessary health care needs such as prescription medicines, doctors' care and dental care due to unaffordability. (Kaufman, B., 2021)

This one variable seems to affect several other key variables including Age, homeowner status short term loan index

Medicare Supplement Member is under 65

Those under 65 who are eligible for Supplementary Medicare insurance are typically suffering from disability. Disabilities are among the most important factors that contribute to housing insecurity as noted in multiple longitudinal studies. 25% of housing insecure people in the United States have a disability. (Novack, V. et al, 2022) According to the Center for American Progress, "*This trend affects disabled women and disabled people of color at disproportionate rates. Disabled people also hold higher shares of medical debt, experience higher rates of food insecurity, and receive lower pay.*"

Recommendations

What we found most interesting over the course of this analysis was how certain seemingly unrelated variables actually affect one another. This is best exemplified by the Student Loan Index variable. As noted in the previous section, SLI has corollary effects on age, homeowner status and short term loan index. Other variables like medicare supplement member under age 65 also seem to have interactions with other important variables. This leads us to believe that recommendations should be focused on taking preventive measures in two broad areas; Remediating housing insecurity risk for high risk demographics, and preventive measures for the health outcomes that housing insecure people often face.

1. **Actions to remediate housing insecurity risks for specific social groups as identified in this analysis at a societal level**

    Housing insecurity affects certain demographics more than others. Here we examine three population demographics with disproportionately high risk for housing insecurity as identified via our analysis.

    Members with Student and Short Term loans:

    The index variable for these types of loans were very significant features in predicting housing insecurity. It is common knowledge that higher education in this country is becoming more unaffordable every year. Even wealthy families struggle to pay for their children's college tuition. Students without family financial backing are at very high risk to delay major homeownership and other life milestones. These issues also correlate to increased depression and mental health issues for those with persistent student loan payments.

    As a society we have to reevaluate how much college tuition costs and find ways to help struggling students that go beyond local, state and federal scholarships. President Biden's student loan forgiveness plan is a good start but does not address the root cause of the problem, which is inability to repay loans due to lack of sufficient income after graduating. We need to put into place more effective advisories that help funnel students into higher paying career paths. More than 40% of graduates wish they had chosen a different career path because of their student loan burden. (Hornsby, T. 2022)

    Members between ages 32-65

    As noted earlier, this demographic seemed to have the highest rates of housing insecurity. This segment of the population constitutes the primary workforce that drives the overall economy. We obviously cannot alter peoples' age, so it may seem that this variable is difficult to take action on. However, we can help reduce HI risk in this segment of the population by finding ways to reward those who are working hard, but not able to pay their mortgages and rental obligations. We can work on providing financial incentives for people in this category to make better financial and health related life choices. In the same way that so many fintech startups such as Greenlight and Bloom are focused on helping kids learn the ropes of financial decision making, so too can we look to provide federal funding for startups looking to do the same for this population demographic.

Those under 65 with Disabilities

Those under 65 with disabilities are also at a disproportionately high risk to suffer from HI. The primary driving factors of this phenomenon are inability to find gainful employment, and low wages when they do. The median sub-minimum wage for disabled people is $2.15/hr, which is more than 3x lower than the federally mandated level of $7.25/hr.

We need to create more employment opportunities with competent wages that help the disabled people to pay their expenses. Investing in disability-forward housing policies and initiatives will provide housing security. Try to eliminate housing discrimination. provide the assistance programs to create safe working places and improve accessibility.

Finally we need to help promote the disabled to develop skills that harness digital technologies, and do not require specific physical abilities. We can do this by investing in training and apprenticeship programs for various IT sectors, and discounts for online training courses for those who have qualified for medicare supplementary assistance under the age of 65

2. Actions to remediate indirect health effects of housing insecurity

We know that numerous chronic health conditions are associated with, and made worse by, housing insecurity. They include cardiac, respiratory, and mental health diseases to name a few. We can use the predictive modeling approach to flag those that may not yet be but are likely to become housing insecure. Based on HI prediction values, and cross referencing common health conditions we can implement preventive care measures to help avoid the dire consequences of chronic health conditions.

For example if we are able to predict that a certain member is likely to be housing insecure, and has a family history of certain chronic conditions like heart disease or substance abuse, we can proactively encourage them to participate in preventive care workshops. Abbott Laboratories, a major US pharmaceutical company has implemented a similar health program for its employees, where in employees who participate in regular physical screenings, daily exercise and more can redeem points earned for things like Amazon gift cards.

3. What can Humana do about it?

The digital generation has made gamification of social systems a very effective tool to incentivize people to make choices. Humana can leverage this trend to help its most vulnerable Medicare members to make healthy life decisions to proactively address their risk. Focusing on proactive, preventive care is a cost benefit for providers like Humana. We do not have the figures on how much health insurance providers save on healthy vs unhealthy members, so a cost benefit analysis is difficult to provide here. However, we do know that healthy people pay health insurers ~$2,500 less per year than their health compromised peers, indicating that insurance providers clearly save substantially when members make healthy choices. Preventive care measures save the health system ~$7 Billion/year (Yong et.al, 2010)

For those members with student and/or short term loans, and for those who are already deemed housing insecure and fall in the 32-65 age range, Humana can look into gamifying health and lifestyle choices. Similar to how companies like Abbott Labs have incentivized their employees to make healthy life choices, Humana can incentivize exercise, healthy eating, and cessation of high risk activities like smoking. This will help not only mitigate the cost burden of health diseases, but also promote prevention of mental health issues..

For members with disabilities who have qualified for Medicare supplementary plans under the age of 65, Humana can focus on investing in companies and organizations that help provide financial opportunities for this demographic, especially those who are part of the underrepresented minority segment who have significantly less access to positive financial outcomes.

References

2022 state of the nation's Housing Report: 4 key takeaways for 2022. Cost of Home. (n.d.). Retrieved October 15, 2022, from https://www.habitat.org/costofhome/2022-state-nations-housing-report-lack-affordable-housing

Braver, T., & Jenvey, V. B. (2012). Lifetime risk factors associated with level of housing among Australian poor. *Poverty & Public Policy*, *4*(1), 1–30. https://doi.org/10.1515/1944-2858.1204

Cai, J. Y., Fremstad, S., & Kalkat, S. (2021, April 5). *Housing insecurity by race and place during the pandemic*. Center for Economic and Policy Research. Retrieved October 16, 2022, from https://cepr.net/report/housing-insecurity-by-race-and-place-during-the-pandemic/

Kaufman, B. (2021, January 8). *Student debt: Bad for your wallet, bad for your health*. Student Borrower Protection Center. Retrieved October 16, 2022, from https://protectborrowers.org/student-debt-bad-for-health/

Muoio, D. (2021, August 3). *Nationwide out-of-pocket spending jumped 10% in 2021. expect that growth to continue through 2026*. Fierce Healthcare. Retrieved October 16, 2022, from https://www.fiercehealthcare.com/payer/nationwide-out-pocket-spending-grew-10-to-1-650-per-person-2021-expect-to-continue-through

Murtha, J. (2022, February 15). *The top five most expensive employee health conditions*. Canopy Health. Retrieved October 16, 2022, from https://www.canopyhealth.com/the-top-five-most-expensive-employee-health-conditions/

Novack, V., Lake, J., & Ives-Rublee, M. (2022, June 8). *Recognizing and addressing housing insecurity for disabled renters*. Center for American Progress. Retrieved October 16, 2022, from https://www.americanprogress.org/article/recognizing-addressing-housing-insecurity-disabled-renters/

Travis Hornsby, C. F. A. (2022, June 18). *Top 10 jobs with most student debt regret*. Student Loan Planner. Retrieved October 16, 2022, from https://www.studentloanplanner.com/top-jobs-career-regret-student-debt/

Twitter, E. W., Twitter, E. W., Ortaliza, J., & Twitter, C. C. (2022, February 14). *How does health spending in the U.S. compare to other countries?* Peterson-KFF Health System Tracker. Retrieved October 15, 2022, from https://www.healthsystemtracker.org/chart-collection/health-spending-u-s-compare-countries-2/

Yong, P. L., Saunders, R. S., & Olsen, L. A. (2010). *The Healthcare Imperative: Lowering costs and improving outcomes: Workshop Series Summary*. National Academies Press.

Young adults disproportionately affected by food and housing insecurity and mental health symptoms during pandemic. Chapin Hall. (2021, June 15). Retrieved October 16, 2022, from https://www.chapinhall.org/news/young-adults-disproportionately-affected-by-food-and-housing-insecurity-and-mental-health-symptoms-during-pandemic/