Humana

HUMANA-MAYS

HEALTHCARE ANALYTICS CASE COMPETITION 2022:
PREDICTION OF HOUSING INSECURITY

MAYS BUSINESS SCHOOL

Table of Contents

1. EXECUTIVE SUMMARY

1.1 Context

Social Determinants of Health (SDoH) are an individual's immediate personal circumstances that may impact their health and well-being. One important SDoH is housing insecurity (HI), which refers to a lack of access to decent, affordable, and secure housing. This SDoH has a large impact on individual's health and poses a large burden to the U.S. healthcare system.

For the affected individual, HI negatively impacts **physical, social, and mental well-being**. The lack of stable housing exposes an individual to high stress levels, increases their risk of exposure to pathogens (e.g. mold, infectious diseases), and negatively impacts their social relationships.

Although housing-insecure individuals are more likely to be exposed to health-damaging conditions, they are also more likely to postpone required *primary medical care* which ultimately results in greater health damage and more *emergency medical care*. Ultimately, HI results in **large, preventable expenses** for healthcare and insurance providers. In fact, a recent study by United Health indicated that moving individuals from unstable to stable housing conditions results in a 20% increase in primary care visits and an 18% decrease in costly ED visits, which resulted in savings of $115 per affected member per month (Anderson, 2018).

In total, HI costs the US healthcare system $11 billion dollars per year (Poblacion et al., 2017). Given this value and information about (A) Medicare and Medicaid market share and (B) Humana's presence in each segment, we estimate that the burden of HI costs Humana $265M each year. This burden could be significantly reduced by early intervention for at-risk members and housing security solutions for Humana's most strongly affected members.

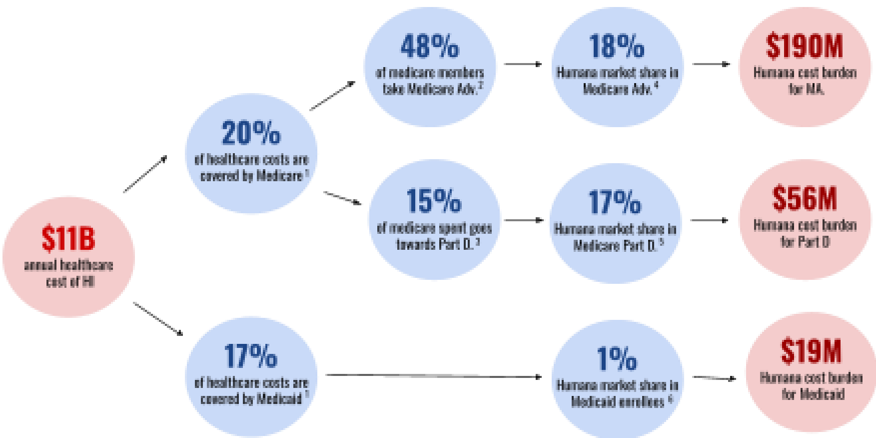


Figure 1: The cost of HI for Humana

This study aids Humana in identifying which Medicare members are housing insecure. The goal of this research is to develop a model to identify Medicare members that are most probable to be struggling with housing insecurity issues and to propose solutions that help people achieve their best health.

---

[1] CMS (2018). *National Health Expenditure 2017 highlights*. Centers for Medicare & Medicaid Services.

Morse, S. (2021). *The Disadvantages of a Medicare Advantage Plan*. Healthcare Finance News.
[3] KFF (2021). *An overview of the Medicare part D prescription drug benefit.* Kaiser Family Foundation.
[4] KFF (2022). *Medicare Advantage in 2022: Enrollment Update and Key Trends.* Kaiser Family Foundation.
[5] KFF (2022). *Key facts about Medicare Part D enrolment and costs in 2022.* Kaiser Family Foundation [6] Humana (2022). *Humana Healthy Horizons in Ohio*. Humana.

1.2 Modeling

Humana provided housing insecurity data for 48,300 individuals, with 881 features collected from primary research and secondary data sources. The features contained demographic, economic, and healthcare data for Medicare members associated with Humana. The objective of the modeling exercise is to predict Medicare members' likelihood of suffering from housing insecurity. To achieve this objective, extensive data cleaning and feature engineering was performed. This was followed by training multiple binary classification models on the clean enriched data to predict housing insecurity.

In the initial data set, 29.5% of columns contained missing values and 13.5% of columns contained no variation in them. The following data cleaning and imputation steps were performed to prepare the data set for model training: (1) Columns with no variation were dropped, (2) columns with more than 80% missing values were dropped, and (3) remaining columns were imputed based on the column type and the information it represented (e.g. county based variables were imputed using KNN imputation with 1,000 neighbors, whereas categorical variables were imputed with the mode or the category "Unknown").

Extensive data augmentation and feature engineering steps were performed to further improve data quality. We were able to identify the county of each individual using the *atlas_totalpopest2016* feature, and subsequently added relevant county-based health features from publicly available data sources (e.g. % uninsured, life expectancy, % food insecure, etc.). Additionally, we created new features from existing features (e.g. health_risk_and_managebility_index) to improve explainability and model performance.

Given the large dimensionality of the data, we performed feature selection using the Boruta method and reduced the number of features from 900+ to 134 to improve scalability and reduce model overfitting. After trying a variety of binary classification models (Gradient Boosted Machines, XGBoost, LightGBM, Random Forest, Neural Networks), we finalized an ensemble of three binary classification models, consisting of a Neural Network model, a Gradient Boosted Machine model, and an XGBoost model. The final model selection was based on its fairness and performance on test data (20% of training data). The ensemble model obtained a fairness score of 0.992 and a holdout AUC score of 0.7598.

To identify important features and explain how the top features influence model predictions, we used feature importance plots and SHAP plots. Despite the fact that Neural Networks are generally considered to be blackbox models, our team calculated shapley values for the Neural Network model and combined its impact with the other two tree-based models in our ensemble. Features such as home ownership, age, and income were negatively correlated with housing insecurity and features related to mental health issues, neurological disorders and CMS subsidies were found to be positively correlated with housing insecurity.

Finally, we applied K-Means clustering for all individuals who have housing insecurity (i.e. *hi_flag=1*) to identify differentiating factors that act as drivers for housing insecurity. We identified three distinct clusters of individuals who experience housing insecurity: (1) The ESRD/Disabled Mental Care cluster,

(2) The Frail OASI cluster, and (3) The HI-Homeowner cluster (*see Section 4.5: Cluster Analysis*). Furthermore, we recognized our model's limitations when it comes to cluster 3 – the HI-Homeowner cluster – where it has a relatively low recall of 37% compared to the other two clusters (i.e. cluster 1 has 96% recall & cluster 2 82% recall). We therefore recommend including additional data points such as mortgage, the individual's income, family size, and recent disaster events by geography to improve model performance.

1.3 Recommendations

The ensemble model combined with the cluster analysis can be used to identify individuals with a high risk of housing insecurity and to generate actionable recommendations to proactively tackle the housing insecurity issues these individuals are facing. There are primarily two aspects to consider when we think about deploying and scaling the solution to the real world. First, we need to understand the model's limitations when scaled, and second, we should carefully consider the actionability of each recommendation and its financial impact.

At a risk score threshold of 0.5, the model can identify 2 out of 3 cases of HI in the data. However, at the same time, the model has a relatively low precision of 10.23%. This means it generates a lot of false positive cases which creates obstacles for Humana when the organization wants to employ intervention techniques to address housing insecurity issues. The high presence of false positive cases increases the operational and financial burden on Humana since the organization has to spend resources to identify true HI cases. Furthermore, the fact that the model produces more false positives also raises the risk of spending resources on individuals who do not really need assistance with housing. Hence, we recommend a "**tiered approach**" in scaling the model in the real world. Humana can take a two-phased approach to weed out false positive cases in a cost-effective fashion.

Phase 1 includes using the model to identify individuals at a high risk of housing insecurity. At a threshold of 0.5, we believe 2 out of 3 true HI cases will be captured. To deal with the false positive cases of HI in phase 2, Humana should employ various communication channels to conduct a short survey with individuals flagged as high risk for HI to detect false positive cases. The channels used to conduct the survey should depend on the risk score to reduce the cost of conducting the survey. We recommend the following:

A. **For individuals with a HI risk score above 0.62:** This group has a high risk of HI and a relatively high precision of 13%, Humana should utilize multiple channels to contact individuals predicted to be HI to screen for true HI cases. The following channels can be used: (1) Survey over email, (2) IVR recorded telephonic survey, (3) survey over mail, and (4) if the person does not respond to multiple communication attempts through the prior channels, then we recommend a telephonic survey with a Humana representative. Humana can also consider (5) in-person surveys limited to areas where there is a high density of HI individuals and Humana was unable to contact them through any of the low-cost channels

B. **For individuals with a risk score between 0.5 and 0.62:** This segment has a low-to-medium risk of HI and a relatively low precision of 7%. Humana should only employ low-cost channels to contact these individuals to screen for true HI cases. Low-cost channels include (1) Survey over email, (2) IVR recorded telephonic survey, and (3) survey over physical mail.

As discussed above, after carefully analyzing and filtering for true cases of housing insecurity, Humana should devote resources towards tracking and tackling housing insecurity. To combat HI, we have provided recommendations in five key areas: (1) **Screening and early identification**, (2) **an emergency response system**, (3) **proactive outreach for select clusters**, (4) **long-term strategic investments**, and (5) **influencing public policy**.

The two-phased approach mentioned above should be used to conduct screening and early identification of HI. Based on the results from this exercise, Humana can take preventive measures for members who are in the early/mild stage of housing insecurity to (A) prevent the situation from worsening and (B) reduce emergency healthcare usage. Humana can **send reminders** through various channels (i.e. emails, fliers, telephone calls) to its members to inform them of the free and low-cost primary health care that they can easily obtain with their Medicare coverage. This way, Humana can motivate members to *not* forgo their primary care visits and reduce emergency healthcare utilization. For eligible members, Humana should also remind them of various housing assistance resources their medicare advantage plan offers around housing quality and/or instability benefits. Specifically, individuals who belong to Cluster 1 (The ESRD/Disabled Mental Care cluster) should be made aware of benefits such as NEMT (non-emergency medical transport) to avail primary care.

Further, for members at high risk of losing their place of residence, Humana should implement an **emergency response system**. Under the emergency response system, Humana should take a case-by-case approach to provide either housing support services, short-term emergency/transitional housing, and/or need-based emergency funding. These measures are aimed at providing immediate relief in the short run and helping its members overcome housing insecurity.

While the above-mentioned measures can help tackle HI issues in the near term, Humana should also focus on **longer-term strategic investments** and expand its efforts to **influence public policy**. Housing insecurity is a complex issue with multiple stakeholders and requires public-private partnerships to resolve it. Under long-term strategic investments, Humana should consider investing resources in building or acquiring low-cost affordable housing in areas with a high density of HI cases. Additionally, partnerships with community benefit organizations on a value-based compensation model can help rope in agencies that are capable of dealing with housing insecurity issues and leverage their expertise in providing relief to Humana's members.

Implementing all of our recommendations would require an estimated initial investment of $37.1M (*see Section 5.3*), however, we expect that this investment can be recouped within a three-year payback period. Along with financial viability, Humana's efforts towards resolving housing insecurity issues will further increase Humana's brand perception among the public and take Humana closer to its vision of improving health outcomes for all its members.

2. INTRODUCTION

2.1 Background

Social Determinants of Health (SDoH) are the immediate health-damaging conditions affecting an individual (Humana Mays, 2022). SDoH includes housing insecurity, food insecurity, transportation barriers, loneliness, and financial strain. Various research has shown that these experiences may negatively affect physical health and make it harder to access healthcare (US Dept. of Health and Human Service, 2022). SDoH screening can inform patients' treatment plans and enable providers to make referrals to any required community services (Center for Medicare & Medicaid Services, 2022).

The SDoH at the core of this paper is **housing insecurity** (HI), which refers to the absence of access to decent, affordable, and secure housing. Situations including chronic or intermittent homelessness, dangerous/overcrowded conditions, frequent evictions, or "couch-surfing" are included in this definition. In the United States, 32.7% of households with older adults experience severe housing issues. Housing insecurity may affect mental health, stress levels, relationships, sleep, the risk of infectious disease, allergies, neurological impairment, heart damage, and more (US Dept. of Health and Human

Service, 2022). There are four pathways connecting housing and health: housing stability, housing quality and safety, housing affordability, and neighborhood (Taylor, 2018).

Stability – An unstable living situation exposes individuals to several health hazards such as chronic stress, trauma, and other mental health issues. This may result in injury, disease, mental illness, or behavioral health issues such as drug abuse. Further, instability tends to reduce the effectiveness of healthcare interventions as it can reduce medication adherence due to improper storage or other instability issues (Maqbool et al., 2015).

Quality and Safety – Low-quality housing poses a multitude of health hazards. Substandard conditions (e.g. mold, poor ventilation, etc.) can cause asthma or allergies. Lead exposure may result in brain and/or nervous system damage, and a lack of temperature control is associated with increased mortality. Besides, an unsafe home environment may increase the risk of fall hazards and associated physical injuries (Humana, 2020b).

Affordability – Housing is typically considered *affordable* when a family considers less than 30% of its income on rent or a mortgage (Commission Health, 2008). An individual or family experiencing HI is likely to be low-income and/or cost-burdened. This implies that housing-insecure groups tend to struggle to afford basic human needs like nutritious food and medical care. In fact, housing-insecure individuals are more likely to delay doctor visits because of costs (Mandy et al., 2015).

Neighborhood – People experiencing HI are more likely to end up in low-provisioned or unsafe neighborhoods, and neighborhood conditions pose a variety of health hazards. A lack of access to public transportation may pose a barrier to obtaining healthcare, and exposure to environmental pollutants may cause disease. Besides, social characteristics of the neighborhood (e.g. segregation and crime) can widen health disparities and influence unhealthy behaviors (Humana, 2020b). Lastly, individuals are also less likely to exercise when they feel unsafe in their neighborhood (Maqbool et al., 2015).

2.1 Humana Mays Analytics Challenge

2.1.1 Business Context

Humana believes in a *whole person* healthcare model that supports physical, mental, and social health. The organization believes in a more equitable healthcare where social needs do not determine health outcomes (Humana Mays, 2022). However, as previously established, HI negatively affects physical and mental health, and makes it more difficult to access healthcare. This also bears large financial consequences for Humana, as HI tends to result in the postponement of needed primary care which ultimately results in the utilization of more expensive healthcare resources such as emergency care.

For Humana to achieve its mission and reduce HI-induced preventable expenses, the organization has devoted significant resources to tackling the HI issue. Humana has committed to a **national housing strategy** that aims to address members' housing needs through a three-part approach: (1) housing stability and homelessness prevention, (2) stabilizing individuals with significant health risks with incremental clinical support, and (3) strategic investments to increase community capacity. Regarding the strategic investments: Humana devoted \$25m to building affordable housing both in 2021 and 2022, resulting in a total investment of \$50m.

By building a predictive model, we hope to help Humana in two areas. First, we hope to help the organization better identify which of its members may be experiencing HI. This information can be used to better target Humana's resources at the appropriate individuals and communities. Second, we hope to provide novel information about which features are predictive of housing insecurity. This analysis can help identify various root causes for housing insecurity, and this may help Humana in developing high-impact, long-term solutions and partnerships that tackle the HI problem at its root.

2.1.2 Key Performance Indicators

In order to evaluate our solutions from a business perspective, we will evaluate the following key performance indicators (KPIs):

Table 1: Key Performance Indicators

What	How
Scalability	Build a predictive model that can easily be scaled by Humana.
Financials	Suggest profitable solutions that would require minimal resources for maximum impact.
Reach	Suggest solutions that benefit the maximum number of Humana members.
Fairness	Suggest solutions that benefit all Humana members, regardless of race or sex.

3 DATA PREPARATION

3.1 Data Overview

The training data consists of 48,300 rows and 881 columns, where a single row represents one Medicare member, and a column represents a feature. Similarly, the holdout data consists of 12,220 rows and 880 columns: The only feature available in the training set that is not available in the holdout set is the target variable *hi_flag*.

3.1.1 Target variable

The target variable is *hi_flag* which represents whether or not the individual in question has reported to be housing insecure. This variable takes on the value "1" if the individual experiences HI, and "0" otherwise. This variable is collected through the Accountable Health Communities (AHC) Health-Related Social Needs Screening Tool. This is a 26-question survey across 10 core domains, including food, transportation, utilities, safety, financial strain, employment, family and community support, education, physical activity, substance use, mental health, disabilities, and living situation.

Within the living situation domain of the survey, the HI variable is determined based on the individual's response to the screening question: "What is your living situation today?". There are three possible answers to this question which may result in other a **0** or **1** for our dependent variable:

Table 2: Housing Insecurity variable explained

Response	hi_flag

I have a steady place to live	0
I have a place to live today, but I'm worried about losing it in the future	1
I do not have a steady place to live (I am temporarily staying with others, in a hotel, in a shelter, living outside on the street, on a beach, in a car, abandoned building, bus or train station, or in a park)	1

3.1.2 Features

All features in the data set belong to one of the following four categories: (1) Medical Claims and Condition Features, (2) Pharmacy Claims Features, (3) Demographics-, CMS-, or Consumer Features, and (4) Other Features.

The first category represents **medical claims and condition features**. This includes claims and/or total costs by place of treatment (i.e. inpatient care, outpatient care, emergency room care, etc.), claim count by CMS diagnosis code categories, claim count by behavioral health conditions, and the Charlson Comorbidity Index and utilization, the latter representing a weighted index to predict risk of death based on claims accepted in the past year.

The second category represents **pharmacy claims features**, including claim counts and total cost. This includes a distinction between brand and generic medicine, mailed and non-mailed medicine, maintenance and non-maintenance medicine, behavioral health, and non-behavioral health medicine, over-the-counter (OTC) and non-OTC medicine, and prescription categories as defined by Humana.

The third category represents **consumer demographics**, which includes age, gender, race, disability status, dual eligibility, low-income subsidy receiver, rural category, and CMS risk score and payment amount.

Finally, the fourth category represents **any other features**. This includes rural Atlas SDOH features, Robert Wood Johnson Foundation SDOH features, outreach point features, revenue code-related utilization features, and credit data features.

3.2 Understanding the Data

In this section, we will provide information about the sample composition and important features.

3.2.1 Sample Composition

Our target variable *hi_flag* is a highly unbalanced variable, as only 4.39% of the respondents reported being housing insecure.

Furthermore, we see that 77.74% of the respondents are white (non-Hispanic), 15.95% are black (non-Hispanic), 2.21% are Hispanic, 0.62% are Asian, Asian American or Pacific Islander, 0.27% are American Indian or Alaska Native, 1.57% are in the "other" category, and the race of the remaining 1.63% is unknown. Within the different racial groups, we see different rates of HI, as can be seen in Table 3.

Table 3: Housing Insecurity by Race

Response	Not housing insecure	Housing insecure
Hispanic	90.73%	**9.27%**
American Indian or Alaska Native	92.42%	**7.57%**
Asian, Asian American or Pacific Islander	92.62%	**7.38%**
Black (non-Hispanic)	93.90%	**6.10%**
Other	94.20%	**5.80%**
Unknown	94.42%	**5.58%**
White	96.19%	**3.81%**

Regarding gender imbalance, we see that 60.25% of our respondents are female, and 39.75% are male. We also see that men are more likely to be housing insecure relative to women, as can be seen in Table 4.

Table 4: Housing Insecurity by Sex

Response	Not housing insecure	Housing insecure
Male	94.93%	**5.68%**
Female	96.07%	**3.93%**

The distribution of age is not surprising: 70-75 is the most prevalent with 26.11% of the respondents, closely followed by the 65-70 age group with 23.61%. 75-80, 60-65, and 80-85 are the next three groups with 16.14%, 11.06%, and 10.58% of respondents in these groups, respectively. As can be seen from the distribution in *Figure 2*, there are some younger people in our data set. These younger people would have different reasons to qualify for Medicare other than Old Age Survivors Insurance (OASI).

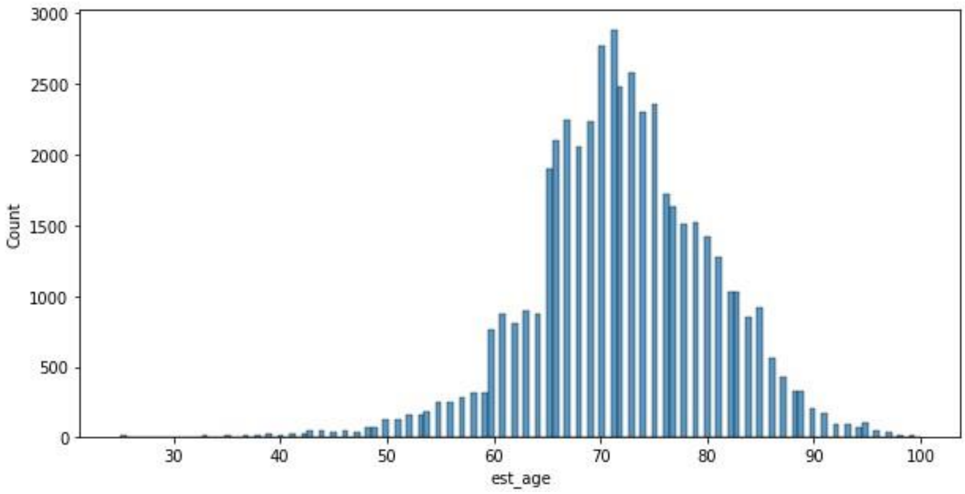


Figure 2: Distribution of estimated age

While 69.07% of our respondents are on Medicare because they are on the OASI, 26.98% is on Medicare because of their disability status, 6.00% are both on OASI and are disabled, and the remaining 3.11% qualified for Medicare due to End Stage Renal Disease (ESRD).

Table 5: Housing Insecurity by Medicare Reason

Reason	Not housing insecure	Housing insecure
End Stage Renal Disease (ESRD) (only)	86.67%	**13.33%**
Both Disabled and End Stage Renal Disease	89.66%	**10.34%**
Disabled (only)	93.02%	**6.98%**
Old Age Survivors Insurance (OASI)	97.01%	**2.99%**

Finally, while analyzing the data, we noted that some data points of younger Medicare members (age < 65 years old) are classified with the reason OASI. This may be due to the introduction of fuzziness by the synthetic data algorithm used to create the dataset. Namely, Humana used a synthetic data algorithm to purposefully introduce noise to the real data, in order to protect the privacy of the individuals from whom the data originated. Obviously, this combination of age and Medicare reason would never occur for true data, and therefore we exclude these data points.

3.2.2 Important Predictors

Based on a review of prior research, we identified various factors that tend to be associated with HI. These factors were taken into consideration when conducting feature engineering and data enrichment (see *section 3.4*). *Table 6* below provides a summary of the variables that may be predictive of HI according to existing literature.

Table 6: Predictors of Housing Insecurity according to the literature

Predictive factor	Exploratory analysis findings
Being from a minority racial class (Shinn et al., 1998)	The predictive factor was **confirmed** by analyzing the *race* variable – the majority group in the US is least likely to experience housing insecurity (3.93%) out of all seven racial groups.
Living in a high-rent area (Early, 2005)	No relevant data provided.
Suffering from a behavioral disorder (Shinn et al., 1998)	The given behavioral health variables (variables starting with *bh*) are **not correlated*** with the *HI_flag* DV.
Having a history of drug and/or alcohol abuse (Shinn et al., 1998)	The three variables related to addiction (tobacco, alcohol, other substances) are **not correlated*** with the *HI_flag* DV.
Having experienced domestic violence or childhood disruptions	No relevant data provided.
A lack of social ties (Shinn et al., 1998)	No relevant data provided.

Home ownership (Shinn et al., 1998)	The predictive factor was **confirmed**: Home ownership (cons_homstat=Y) is negatively correlated with housing insecurity (ρ = -0.11).
Pregnancy or recent birth (Shinn et al., 1998)	No relevant data provided.
Enrolment in welfare benefits (Shinn et al., 1998)	No relevant data was provided. However, external data about SNAP benefits (see *section 3.4.1: Data Enrichment*) **confirmed** the positive correlation between welfare benefits and housing insecurity.
Prior incarceration or arrests are associated HI (Shinn et al., 1998)	No relevant data provided.
Depression and anxiety (Cox et al., 2019)	The given health variables related to depression and anxiety are **not correlated*** with the *HI_flag* DV.

It is notable that *not* all listed medical variables were confirmed to be associated with HI in our dataset. We do not see a relationship between *hi_flag* and the variables related to behavioral disorders, depression, and anxiety. However, we believe that this can be explained by the fact that housing insecure individuals are more likely to postpone and/or avoid needed primary care. The health-related variables in our dataset are solely based on *medical claims* and thus they may not accurately capture the health of those individuals who avoid primary care because of personal circumstances such as housing insecurity. Thus, we believe that **the health data for housing insecure individuals may not accurately represent the true health of the housing insecure individual.** This is one of the primary limitations of our dataset. Because of this short-coming, we may not be able to capture all relevant relationships between provided mental and physical health variables and HI.

3.3 Data Cleaning and Imputation

The original dataset contains 881 columns, where 29.5% of columns contain null values and 13.5% of columns contain no variation – meaning they either consist entirely of null values or contain only a single unique value and no null values. We prepared our data for analysis using column removal, missing value imputation, outlier removal, data enrichment and feature engineering.

3.3.1 Column removal

First, we removed 119 columns that contained no variation. As every observation in the column is equal to the same value, these columns would not add predictive value to our model. After removing no-variation columns, we removed an additional 12 columns that contained over 80% missing values.

3.3.2 Missing data imputation

Second, missing data imputation was carefully tailored for each column that contained missing values. We treated the following categories with a different imputation approach:

1. **Categorical variables → 'Unknown' or most frequent.**
   If the categorical variable contained a pre-existing *Unknown* category, then missing values were imputed with *Unknown*. In cases where there was a large number of missing values and no pre-

existing *Unknown* category, we newly created an *Unknown* category. In cases where there was only a small number of missing values, the missing values were imputed with the most common category value.

2. **Numeric variables with solely zeros and a few null values → impute with 1.**
   As we do not know whether missing data is missing at random or not, we decide to impute missing values with a **1**. This way, subsequent feature selection can determine whether the missing values are related to our outcome variable or not.

3. **Numeric variables with a clear "0" majority class → impute with 0.**
   Numeric variables with a clear "0" majority class include most variables related to medical claims, conditions, and pharmacy expenses. It is most likely that members with missing values did not submit claims or expenses related to the conditions under the relevant feature.

4. **Environmental variables (rwjf, atlas) → KNN imputation**
   County-level social and economic factors tend to be interrelated, and hence KNN imputation may be used to effectively impute missing environmental variables. After parameter tuning, we settled on using KNN with 1,000 neighbors in order to obtain a KNN-localized mean for imputation. Thus, when we encounter a missing value, we look for the 1,000 most environmentally similar data points and impute the missing values with the average value of the 1,000 non-missing values. *Note* that we tried both median imputation and KNN imputation, and KNN imputation led to a better test set AUC.

5. **Numeric variables with a normal or skewed distribution → impute with the median.**
   According to generally accepted best practices, variables with a normal or skewed distribution were imputed with median values.

3.4 Data Processing & Feature Engineering

We augmented our dataset through data enrichment (adding new sources of data), feature combination, feature selection, and outlier removal.

3.4.1 Data Enrichment

As the provided dataset did not include a *foreign key* to easily join external data (e.g. zip code), data enrichment proved to be a challenging task. However, after closely examining the county-level variables in our dataset, our team was able to identify the **county** that corresponds to each row. We identified the relevant counties by joining the 2018 County Health Rankings National Data (CHRN data) to our original data based on the 2016 county population that was present in both (A) the 2018 CHRN data and (B) the Humana dataset (*atlas_totalpopest2016*).

After we were able to identify the counties of the Humana members, we were able to augment our dataset with external county-level data that we considered relevant based on our literature review:

1. **The 2022 County Health Rankings National Data.**
   % uninsured, COVID-19 death rate, Life Expectancy, % Food Insecure, High School Graduation Rate, Homicide Rate, % Homeowners, % rural, Segregation Index, % Frequent Physical Distress, % Frequent Mental Distress, Gender Pay Gap, % Enrolled in Free or Reduced Lunch, % Severe Housing Cost Burden, % Hispanic, % Black, % Asian, % Non-Hispanic white

2. **US Census Bureau: County-level SNAP benefits and poverty data.**
   Poverty universe, total SNAP benefit enrolment, % SNAP benefit enrolment.

Note that the county join still left us with 26.6% missing data for the *county* column and related county-level variables. We imputed missing county-level variables using KNN imputation, in the same manner that we used KNN imputation for the other environmental variables (*rwjf* and *atlas*).

3.4.2 Feature Combination

We created new features based on (A) our literature review and (B) by hard-coding relevant algorithm-identified feature interactions.

First, based on the literature review, we created two aggregate health indices that are listed below. However, it must be noted that these indices were not deemed to be valuable by our models in feature importance plots, and hence they were ultimately removed from the model.

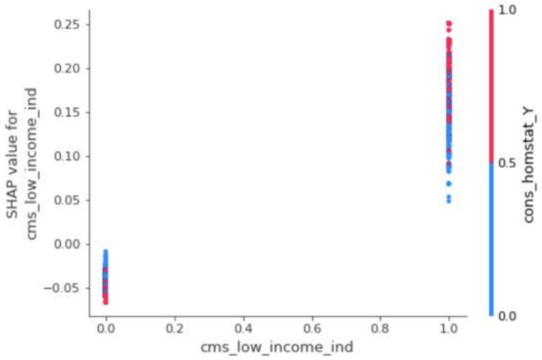
1. **Behavioral health index** – aggregate measure of behavioral health variables related to anxiety, depression, substance abuse, and trauma.
2. **Prescription count features** - aggregate count of Rx for health conditions that can potentially indicate housing insecurity. Features for health conditions include neurological disorders, mental health conditions, cancer (malignant and benign), chemotherapy, respiratory conditions, substance abuse, diabetes, pain, etc.

Second, we hard-coded relevant feature interactions that were identified by a preliminary model so that our subsequent models could more easily identify these relationships and use their training iterations to find other interactions that may be more difficult to find. After running a preliminary model, we used SHAP dependency plots to identify which variables interacted with each other to determine the model's predictions. We identified nine notable interactions. All interactions are listed in *Appendix A*, and two exemplary interactions with corresponding dependency plots are listed below:

1. **Home-ownership (cons_homstat_Y) & low income (cms_low_income_ind)**
   Home-ownership typically makes a member less likely to experience housing insecurity.

   However, if you are a low-income member, then owning a house makes you more likely to experience housing insecurity (*Figure 3*).

2. **Home-ownership (cons_homstat_Y) & days since last claim for physician office (total_physician_office_ds_clm)**
   If a member does not own their house, then the days since their last claim for a physician office visit is an indicator of housing insecurity. The longer it has been since their last claim, the more likely they are to experience housing insecurity (*Figure 4*).



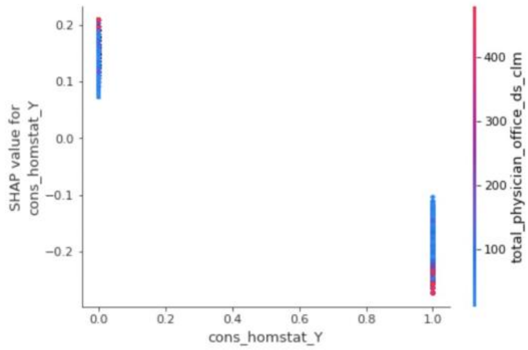


Figure 3: SHAP of low income and home-ownership

Figure 4: SHAP of home-ownership and days since last claim

3.5 Feature Selection

After data cleaning and augmentation, we were left with 810 features. In order to reduce the number of variables, we applied the **Boruta algorithm** for feature selection. This algorithm helps us build a model that is free of highly correlated and/or irrelevant variables. The algorithm works as follows:

Table 7: Feature Selection Algorithm

1. Add randomness	Create shuffled copies of all features: *shadow features*.
2. Apply random forest	Apply an RF classifier on the extended dataset, including both original and shadow features.
3. Evaluate	Evaluate the feature importance of both the original and shadow features.
4. Compare feature importance	For each real feature, check whether its importance is greater than the importance of the shadow features.
5. Decide which features to keep	Remove variables that underperform compared to shadow features.

(Perlato, n.d.)

After applying the Boruta algorithm, we were able to reduce the number of features from 810 to 134. This feature removal only reduced the final test set AUC by a negligible amount and did not notably impact our test set fairness score. Besides, it must be noted that reducing the number of variables largely improves scalability and resource requirements for Humana.

3.6 Removal of outliers

Finally, we removed anomalies from our dataset using an **isolation forest algorithm**. This algorithm is an unsupervised model that isolates anomalies in a dataset through the use of decision trees. *Figure 5* illustrates how this technique works. In simple terms, the isolation forest aims to “isolate” anomalies by creating decision trees over random attributes. Anomalies are typically those observations that are separated *early* and in a *small partition*. If a forest of random trees collectively produces shorter path lengths for a particular data point, then this data point is highly likely to be an anomaly (Arpit, 2020).

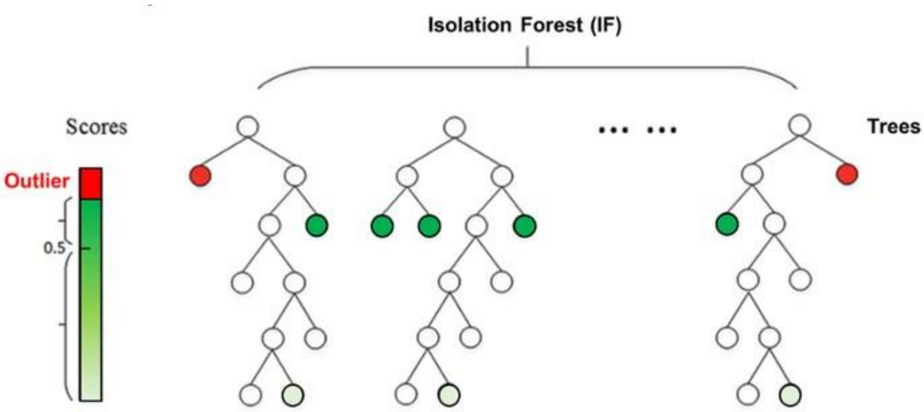


Figure 5: Isolation Forest (Jamalova, 2022)

We ultimately identified 0.3% of data points as anomalous. By removing these anomalous observations, we were able to improve our test AUC from 0.75 to 0.759.

4. MODELING

4.1 Model Objective

Humana has determined two main evaluation criteria for the modeling stage of the competition: (1) to accurately predict whether a Medicare member experiences HI or not, and (2) to minimize bias towards the privileged group in the model used to determine predictions. The final score given to the model consists of the **AUC-ROC** and a so-called **disparity score**.

For each combination of race and sex, the disparity score is defined as follows:

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

where S_n represents the true positive rate for each class, and S_0 represents the true positive rate for the reference group. For this case, the reference group is defined as the white male group, and all other combinations of race and sex are being evaluated against our model's performance on the white male group. be calculated as follows: DR Based on the disparity score of each combination of race and sex, the model's final score can then

$$Disparity\ Score = \frac{\sum_{SV} \sum_{S} min(\frac{S_n}{S_0}, 1)}{N}$$

where SV represents the set of sensitive variables combinations between race and sex, N represents the total number of sensitive variable combinations.

4.2 Modeling Approach

To maximize our AUC and fairness, we tried various different binary classification algorithms, including XGBoost, Random Forest, Gradient Boosted Machines, Neural networks, and LightGBM. The modeling process was iterative where we used **Grid search** and **Cross-Validation** methods to determine optimal hyperparameters to reduce overfitting and maximize ROC-AUC and fairness. *Table 8* lists the test data AUC of a select subset of models that we developed:

Table 8: AUC-ROC for each model

Model	**AUC-ROC** *(20% test set)*
Gradient Boosted Model	**0.7582**
Light GBM	**0.7514**
XGBoost	**0.7471**
Neural Network	**0.7455**
Random Forest	**0.7393**

4.3 Final Model & Performance

After developing individual classification models, we examined whether or not we could improve the performance of our model by creating an *ensemble model* that combines the predictions of the individual models. In order to create an ensemble model that would maximize the holdout data AUC, we created an optimization model that would determine the optimal weights to provide to each individual model in generating combined predictions. Through the optimization model, we can leverage systematic differences in the predictions of each individual model to take advantage of the strengths of each of our five models.

After our first iteration of optimization, we noticed that Random Forest and Light GBM each received a weight of less than 1%. Therefore, we re-ran the optimization with a constraint that the weights for these two models should equal zero. Our final ensemble model combines the predictions of the Neural Network model, Gradient Boosted Model, and XGBoost model with the following weights:

Table 9: Weights allocated to each model by our optimization algorithm

XGBoost	**NN**	**RF**	**Light GBM**	**GBM**
0.234	0.298	0	0	0.468

We can see that most weight is put on the predictions of the Gradient Boosting Model (46.8%), followed by Neural Network (29.8%) and XGBoost (23.4%). By using these weights, we were able to obtain a **final test set AUC of 0.7595** and a **holdout AUC of 0.7598** on the leaderboard. The
ROC-AUC curve of our model is given as follows:

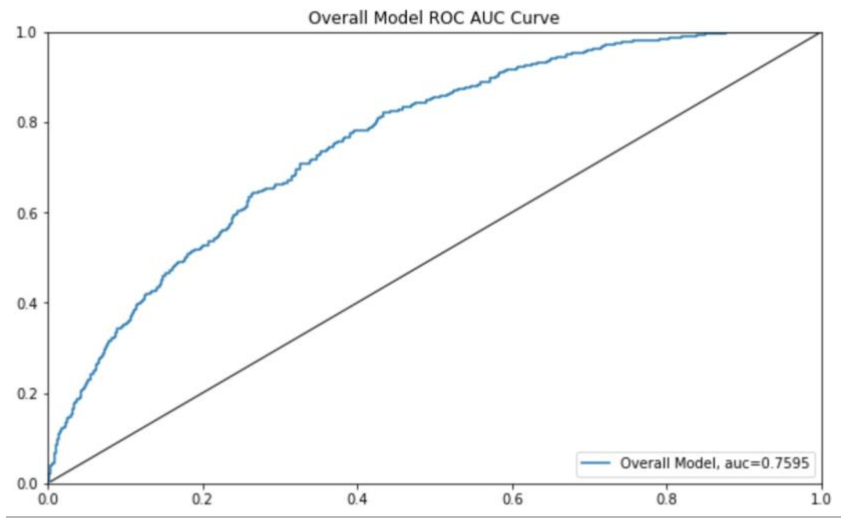


Figure 6: The ROC-AUC curve of the Ensemble model

To further evaluate our model, we will examine the confusion matrix of our predictions:

Table 10: Confusion Matrix of the Ensemble Model (*the threshold for 'hi_flag' was 0.5*)

		Actual Values	
		Housing Insecure	Not Housing Insecure
Predicted Values	Housing Insecure	281	2,464
	Not Housing Insecure	155	6,760

Based on the confusion matrix, we can compute recall, sensitivity, and accuracy:

Sensitivity / Recall: Sensitivity measures to what extent our model is able to detect true positive cases of HI. Our model's sensitivity can be calculated as follows: 281 / (281 + 155) = 64.4%, which means that our model is able to identify about 2 in 3 individuals who actually experience HI.

Specificity: Specificity measures to what extent our model is able to detect true negative cases of HI. Our model's sensitivity can be calculated as follows: 6760 / (6760 + 2464) = 73.3%, which means that our model is able to identify about 3 in 4 individuals who actually are not housing insecure as not being housing insecure.

Accuracy: Accuracy is the overall ability of our model to classify individuals as either housing insecure or not housing insecure accurately. It can be calculated as follows: (281 + 6760) / (281 + 2464 + 155 + 6760) = 72.9%, which means about 72.9% of the individuals in our data set were classified correctly between housing insecure and not housing insecure.

The metrics indicate that our model is effectively able to discriminate between individuals experiencing HI and those who do not. This is further confirmed by (A) the density plot in *Figure 7* and (B) the lift values provided in *Table 11*.

The threshold for classifying ‘hi_flag’ as 1 was set at 0.5. This value was chosen based on the ROC curve (*Figure 6*) and the lift table (*Table 11*), such that we increase Recall while keeping the false positivity rate low.

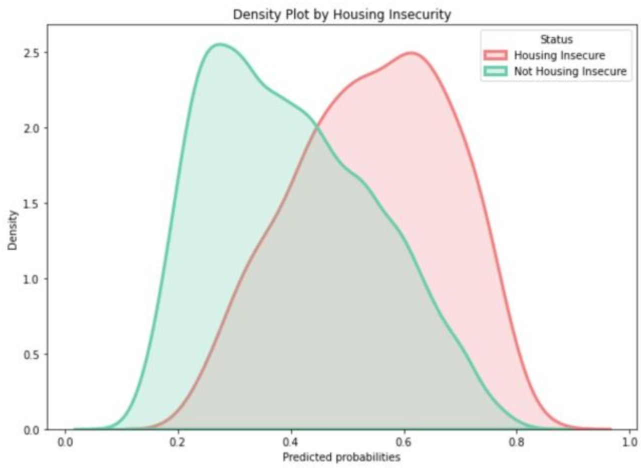


Figure 7: Density Plot of predicted probabilities by actual Housing Insecurity status.

In *Figure 7*, we see that the individuals who are actually housing insecure have a left-skewed probability distribution (i.e. the density of the probability distribution peaks at a high probability value, namely around 0.65), whereas the individuals who are not housing insecure have a right-skewed probability distribution (i.e. the density of the probability distribution peaks at a low probability value, namely around 0.25). This means our model generally allocates higher probabilities to actual housing-insecure individuals, while allocating lower probabilities to individuals who are not housing insecure.

Finally, in *Table 11*, we see a steady increase in lift values from the bottom deciles of predicted probabilities to the top deciles of predicted probabilities. This indicates that the actual rate of HI steadily increases as our predicted probabilities increase, which evidences good discrimination of the data.

Table 11: Lift Table on the Test Data Set (20% of total training data, N = 9,660)

HI Risk Score Decile	# People	# People with HI	Cumulative count of HI people	Lift
(0.62, 1]	966	145	145	3.33
(0.55, 0.62]	966	78	223	1.79
(0.49, 0.55]	966	61	284	1.40
(0.44, 0.49]	966	49	333	1.12
(0.39, 0.44]	966	35	368	0.80
(0.35, 0.39]	966	29	397	0.67
(0.30, 0.35]	966	19	416	0.44
(0.26, 0.30]	966	13	429	0.30
(0.22, 0.26]	966	7	436	0.16
(0, 0.22]	966	0	436	0.00

4.4 Feature Importance

To identify important features and their relationship with our target variable "*hi_flag*", we calculated shapley values for a sample of 1,500 observations for each model in our final ensemble. We were limited to a sample of 1,500 observations due to the computational complexity of obtaining shapley values for the Neural Network component of our ensemble model.

Shapley values help explain the prediction of each observation in the data by computing the contribution of each feature's value to the final prediction of the observation. The SHAP explanation method is based on coalitional game theory, where each feature's contribution to the final prediction is fairly assessed and represented. The important features are those features that strongly contribute towards the prediction.

To arrive at the final overall SHAP plot, we performed a weighted aggregation of SHAP values for each model in our ensemble. The weighting was done based on the weights assigned to the model's predictions, as explained in *Section 4.3: Final Model*.

In *Figure 8*, the features with **blue dots** on the RHS are features that are negatively associated with HI, while the features with **red dots** on the RHS are features that are positively associated with HI.

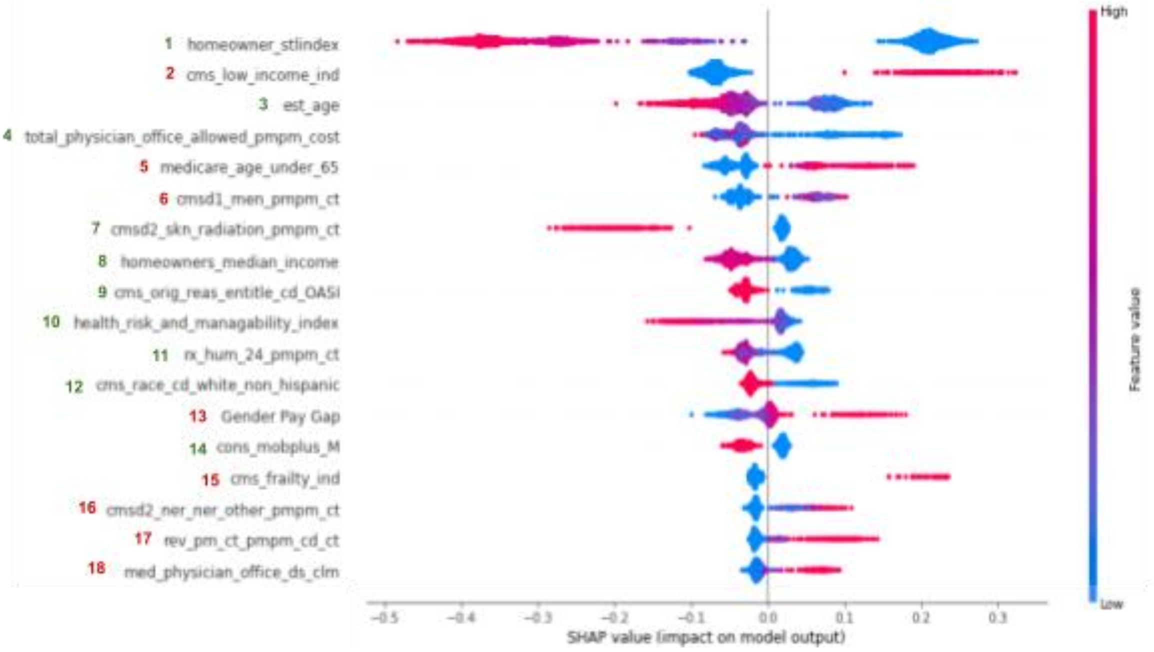


Figure 8: Shap Plot of the Ensemble Model

4.4.1 Features that reduce the risk of housing insecurity

The features that reduce the risk of HI are listed and explained below:

(1) Being a homeowner with a high value for the short-term loan index (STLI)* – Homeowners with higher STLI are less likely to be at risk of HI. The short-term loan index is a demographic-based analytical model which predicts the likelihood that someone in the household has applied for a short term loan (a higher index indicates a higher likelihood). Seemingly, the demographic variables used to predict the STLI are related to HI as well. Unfortunately, we do not have information about these demographic variables. In the future, Humana should analyze between the STLI predictors and HI.

(3) Older age – Older individuals are less likely to face housing insecurity. This might be due to the fact that the likelihood of homeownership monotonically increases with age: whereas only 37.8% of Americans under 35 years owned their home in 2021, about 80% of Americans over 65 did (Statista, 2021). Besides, relatively older Medicare members may be more likely to obtain housing assistance than younger Medicare members who are – on average – healthier.

(4) Higher total physician office allowed cost per month (total_physician_office_visit_ct_pm) – Higher allowed cost per month for overall claims is associated with lower risk of housing insecurity. As previously explained, housing-insecure individuals are more likely to postpone and/or avoid needed primary care. Hence, the costs associated for physician visits would be lower for this group.

(8) Homeowners with higher income* – Homeowners living in a higher-income area are less likely to be housing insecure than (A) homeowners living in a lower-income area and (B) non-homeowners. The fact that these individuals live in a high-income area may indicate that they are financially secure.

(10) Higher health risk manageability index (cons_hxmioc/cms_ma_risk_score_nbr)* – Individuals who are (A) not managing an illness or condition or (B) better able to manage their health conditions are less likely to be experiencing HI. As indicated by our previous literature review, mental and/or physical illnesses cause stress, complicate daily activities, and may hinder economic well-being (eg. by making it more difficult to obtain and sustain a job). Hence, individuals in this group would be more likely to experience financial difficulties and housing insecurity.

(12) Race – Medicare individuals who belong to minority race (i.e. non-White) are at higher risk of housing insecurity. This might due to the life-long marginalization and discrimination these groups have faced and continue to face.

(14) Multiple Mail Order Buyer – Individuals who have made multiple purchases through mail are less likely to be housing insecure. Mail-order buying evidently requires a home-address, and hence repeated mail-order buying indicates that the member has a somewhat stable, long-term address.

4.4.2 Features that increase the risk of housing insecurity

The features that increase the risk of HI are listed and explained below:

(2) Receiving subsidy from CMS – Individuals who receive a subsidy to pay for prescription drugs are more likely to suffer from HI. People who receive CMS subsidies have been classified as low-income individuals, and hence may be less likely to afford necessities such as housing.

(5|9) Reason for Medicare: Disabled/ESRD – Medicare members who are disabled and/or suffer from ESRD are more likely to experience HI in comparison to OASI Medicare members. When an individual with ESRD chooses to receive in-center hemodialysis, Medicare will only start covering the expenses from the fourth month onwards, meaning that individuals are forced to burn through their savings to be able to pay for the first three months of treatment, thereby having less money to pay the mortgage or rent and thus are more likely to become HI (Options for Dialysis, 2022; Collins, 2016). Furthermore, individuals with a disability also have to face a tough financial situation: After being determined to be disabled, an individual will only start receiving Social Security Disability benefits after a five-month waiting period, and will only qualify for Medicare when they have received Social Security Disability benefits for 24 months. Besides the waiting period, the caveat about Social Security Disability is that there is a cap: Individuals who qualify receive somewhere up to $3,345 per month (SSA, 2022), which may not be sufficient for many disabled people.

(6) Claims for mental, behavioral, and neurodevelopmental disorders (cmsd1_men_pmpm_ct) – A higher number of claims per month related to mental, behavioral and neurodevelopmental disorders is associated with an increased risk of HI. These disorders could be both the *cause* or a *consequence* of HI. Individuals with these disorders may be less likely to find and/or maintain their job, rendering them more likely to be in a tough financial situation. However, these disorders may also be caused by stress, anxiety, and inferior physical circumstances that may result from HI.

(12) High county-level gender pay gap* – Living in a county with a higher gender pay gap increases the risk of HI. In these areas, women are at higher risk of HI relative to men as the higher gender pay gap makes it more difficult for women to afford housing in the county.

(15) Frail individuals – Individuals who are deemed frail based on specific diagnoses, multiple serious chronic conditions, functional impairments, or other factors are at a higher risk of HI. These individuals need more accessibility support and might not be able to work, thereby suffering from a lower household income and therefore more likely to be housing insecure.

(16) Neurological disorders – Individuals who have received treatment for miscellaneous neurological disorders are at higher risk of HI. Similarly, as to neurodevelopmental disorders, neurological disorders might be the cause of housing insecurity through the fact that these individuals might face higher barriers to finding a job and therefore are more likely to have lower household income and/or default on their mortgage or rent.

(18) Days since last non-behavioral physician claim – Individuals who have not visited their physician for a non-behavioral issue for a relatively long time are more likely to experience HI. Again, we see this relationship because housing-insecure individuals are more likely to avoid and/or postpone primary care visits. Our data seems to indicate that they are more likely to postpone non-behavioral primary care visits than behavioral primary care visits: This may be because housing-insecure individuals experience higher levels of stress, anxiety, and depression relative to non-HI individuals.

4.5 Cluster Analysis

There are 2118 Medicare members in the dataset who were labeled as HI. However, these individuals vary widely in their demographic characteristics and they experience HI for different reasons. To help them overcome housing challenges, our solution needs to address the unique challenges faced by each individual. This can be partially addressed by identifying different clusters of HI individuals. K-Means clustering was performed using 15 important features that define various characteristics of an individual such as age, race, the reason for medicare, ability to manage their wellness, etc. These 15 variables were selected based on our literature review and our model's feature importance. The features were decomposed into principal components before conducting K-means clustering. *Figure 9* depicts the identified clusters based on the first two principal components.

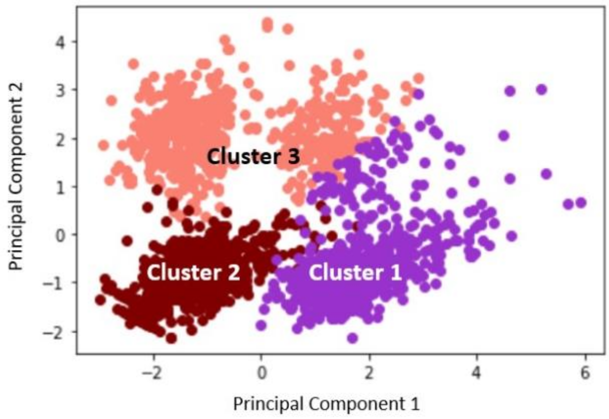


Figure 9: Cluster Analysis

After analyzing the characteristics of each cluster, we identified demographic patterns and potential reasons for HI within each cluster. A brief description of each cluster can be found in *Table 12*.

Table 12: Cluster Details

	Cluster	**Description**	**# People**
ESRD/Disabled mental care cluster (1)	Non-OASI individuals with mental health conditions and/or neurological disorders	This cluster encompassess ESRD/Disabled individuals. They are not frail but predominantly suffer from mental health issues and neurological disorders. This cohort has a low income, which could possibly be due to difficulty in maintaining a steady income stream due to their physical and mental health conditions. They are also less capable of managing their health conditions.	740
Frail OASI cluster (2)	OASI individuals who are frail and do not own their own homes	This cluster contains Medicare individuals who are eligible for Medicare because of their age (65+ years). They have relatively higher income and are better able to manage their health conditions. However, they are most frail and do not own their place of residence which could be direct factors for their housing insecurity.	832
HI-Homeowner cluster (3)	HI Medicare members who are actually homeowners	This cluster contains a mix of OASI and ESRD/ Disabled Medicare individuals. They are ALL house owners. Compared to the other two clusters, this cohort has decent income and ability to manage their health conditions. They are not frail.	546

The HI-homeowner cluster is an interesting cluster as it contains Medicare members whose economic, health, and demographic indicators do not indicate any direct cause for HI. However, there could be multiple reasons for their HI which the current database does not capture. This is further supported by the fact that our model has a very low recall of 37% for this cohort. In contrast, the cluster 1 and cluster 2 have 96% recall and 82% recall.

One hypothesis for the HI status of the individuals in the HI-homeowner cluster may be that this cohort could contain homeowners who are **still paying their mortgage** and who experienced an unexpected adverse event. For instance, job insecurity due to Covid or a sudden loss of an income stream could directly affect the individuals' ability to pay their mortgage.

Further, other hypotheses include the following: (1) The individual may be facing housing challenges due to environmental factors such as wildfires, floods, hurricanes, etc. (2) The individual may have large health expenses that are not covered by Medicare and not captured by our dataset. (3) The individual may have health issues that he/she did not seek care for, and hence those health issues would not appear in our dataset. (4) Finally, it may also be possible that some individuals incorrectly reported housing insecurity.

To identify the root cause of HI for individuals in this cohort, Humana should try to include additional data points around mortgage, family size, income, recent natural disasters across US, etc.

4.6 Model Limitations

There are three primary data limitations that may affect the performance of our model: (1) the inclusion of synthetic data in Humana's dataset, (2) sample bias in the *HI_flag* column, and (3) the unrepresentativeness of claims columns for members experiencing housing insecurity.

As previously noted, our group identified some issues with the member classifications (e.g. some OASI members were younger than 65 years old). After communicating with Humana, it became apparent that the organization has used a **synthetic data algorithm** and this algorithm may have caused some fuzziness and inconsistencies in the data. This likely reduced the predictive performance of our model.

Second, we believe that it is very likely that the *HI_flag* column contains **sample bias** as it would be difficult to reach individuals experiencing housing instability or homelessness. As this group of people is difficult to reach, they would be under-reported and under-represented in our dataset.

Finally, as previously noted, individuals experiencing housing insecurity are more likely to avoid and/or postpone required medical care. Because of this, **our health data could be providing an inaccurately representation of the true health of the housing insecure members in our sample**. For instance, even though a particular member may be experiencing depression and/or anxiety because of their stressful housing circumstances, we would see zero values for the medical claims related to mental health issues if the member is postponing or avoiding their required mental healthcare. Previous studies have found a relationship between mental health issues and housing insecurity, however, this relationship is not fully captured by our model. This may be because the provided data could inaccurately represent the actual health of housing-insecure Humana members. In the future, we should consider asking individuals about their mental health through another medium (e.g. surveys) rather than solely relying on claims data.

5. BUSINESS ANALYSIS AND RECOMMENDATIONS

5.1 Existing solutions for housing insecurity

Before providing recommendations to Humana, we should consider existing solutions to HI. Below, we provide an overview of solutions provided by CMS, federal and state institutes, and Humana.

5.1.1 Medicare Advantage

Medicare Advantage (MA) plans offer housing support through their **Special Supplemental Benefits for the Chronically Ill (SSBCI)** for MA members with chronic health conditions**.** Eligible members can obtain these benefits as long as there is a reasonable expectation of improving or maintaining their health or overall ability to function (ATI Advisory, 2021). Examples of permitted housing-related SSBCI include pest control products, particulate air filters, carpet cleaning, structural home modifications that may assist the member's movement and health (e.g. mobility ramps), and even subsidies for rent or assisted living communities (CMS, 2019).

5.1.2 Medicaid

While Medicaid funds cannot be used to directly pay for housing development or rental assistance, the funds can be used to pay for certain housing-related services and community projects.

First of all, Medicaid funds can be used to finance **services related to obtaining and maintaining housing**. These include housing transition services (e.g. assistance with housing applications) and housing sustaining services (e.g. education on the rights/responsibilities of tenants) (MACPAC, 2018).

Second, states can use Medicaid funds for **housing-related collaborative agreements** (American Institutes for Research, 2017). For instance, Medicaid funds can be used for (A) identifying housing opportunities and (B) organizing collaborations between state housing agencies and community development agencies. This way, Medicaid helps facilitate the development of new housing resources.

Third, states can use Medicaid funds to pay for certain public health activities such as the **lead abatement of buildings**. This includes the removal, enclosure, or encapsulation of lead-based paint and dust hazards (MACPAC, 2018).

Fourth, Medicaid funds can be used to support the ***Olmstead* Implementation**. The *Olmstead* ruling guarantees equal opportunity to access all public programs to people with abilities. As such, Medicaid funds can be used to ensure equal access to supportive housing for people with disabilities (MACPAC, 2018).

5.1.3 Federal and state solutions

Within the U.S., it is estimated that about 1.2 million households currently live in some form of public housing. However, there is a much larger demand, leading to increasing waitlist lengths and an increasing need for additional federal funding to combat this crisis. For this reason, there are a number of federal housing programs organized by the US Department of Housing and Urban Development (HUD) and the US Department of Health and Human Services (HHS).

First, the HUD invests in the development of affordable housing and runs various programs that provide direct assistance to families or individuals experiencing housing instability. For instance, the HUD runs the **Housing Choice Voucher program** which provides rent subsidies to eligible low-income families, as well as a Public Housing program that provides affordable apartments for low-income families, the elderly, and people with disabilities (American Institutes for Research, 2017). Moreover, the HUD has a couple of projects, including the "Continuum of Care Program (CoC)" and "Emergency Solutions Grant (ESG)" to get the financial needs for permanent-, transitional-, and emergency housing, and to prevent homelessness.

Second, the **National Housing Trust Fund (HTF)** is a program that gives block grants to states to construct, maintain, or renovate housing that is affordable to households with incomes at or below 30% of the area median income and/or incomes at or below the federal poverty line. Not only can HFT grants be used for securing housing, it also can be used to help first-time homebuyers with their homeownership activities, including down payment assistance or rehabilitation of the owner-occupied house. The fund specifically targets vulnerable citizens including but not limited to people who are homeless, disabled, elderly, veterans, or victims of domestic violence. Each state maintains the liberty to allocate HFT funds based on their consideration of the most severe and/or addressable housing needs. For instance, Texas was able to build 132 affordable single-adult apartments in Austin partially based on HTF funds (Mayors & CEOs for US Housing Investment, 2022).

Third, the HHS runs programs that provide **in-home and community-based, long-term services**. For instance, the HHS has a program "Projects for Assistance in Transition from Homelessness (PATH)" that helps people facing homelessness in terms of mental health, substance abuse, and housing services. Under HHS, the "Services in Supportive Housing (SSH)" program provides monetary grants to provide individuals and/or families with mental illness and/or substance abuse disorders with a permanent place to stay. Also, the HHS runs programs related to the Older Americans Act and Medicaid home and community-based waivers (Humana, 2020b).

5.1.4 Humana's current solutions

Humana's national housing strategy is focused on three key areas: (1) housing stability and homelessness prevention, (2) stabilizing individuals with significant health risks through incremental clinical support, and (3) strategic investments to increase community capacity (Humana, 2022b).

Housing Stability and Homelessness Prevention

First, in order to assist members with eviction prevention and possible housing diversions, Humana has developed **strategic partnerships** with various organizations that target housing instability. For instance, Humana has a strategic partnership with Volunteers of America (VOA) – a national non-profit organization that provides various support services (incl. housing support) for vulnerable groups. All in all, Humana's partnerships aid the organization to:

- Understand risks that may lead to potential eviction;
- Develop a viable housing plan for members;
- Negotiate with landlords or other housing authorities on behalf of members; ● Provide necessary legal aid to support housing security; and ● Coordinate medical respite care (Humana, 2020b).

Second, Humana launched an innovative **value-based model to address SDoH** in 2020. This model offers healthcare providers resources and tools to identify and address SDoH (incl. Housing insecurity), and provides compensation for coordinating patient care based on three components: (1) patient screenings, (2) documentation of assessment findings and (3) connecting the patient to relevant resources (Humana, 2020a). Research has indicated that *value-based care* results in higher rates of preventative care and screenings, fewer ER visits and hospital admissions, and higher scores on indices related to healthcare effectiveness (Humana, 2020a).

Stabilizing individuals with large health risks with incremental clinical support
Humana has committed to embracing the aforementioned **SSBCI benefits** for Medicare Advantage plan holders, in order to support members in building a safe home environment (Humana, 2020a). For instance, Humana partnered with Home Advantage to provide fall prevention assessments and minor home repairs for eligible members (Humana, 2020b).

Strategic Community Investment
The Humana Foundation has made multiple large investments to advance equity in healthcare by addressing SDoH in Bold Goal Communities. Accordingly, Humana has invested significantly in programs that provide support to families experiencing HI (Humana, 2020a). For instance, during the 2021-2022 period, Humana invested a total of $50m to create low-cost rental units in many of its key communities by rehabilitating or newly constructing affordable housing units (Humana, 2021).

5.2 Next Steps & Recommendations

After examining feature importances, the clustering of housing-insecure individuals and existing solutions to housing insecurity, we came up with five classes of recommendations for Humana. The recommendations are summarized in *Figure 10*, and they are elaborated upon below.

Figure 10: Recommendations Overview

5.2.1 Recommendation 1: Screening and Early Identification

Our first recommendation to Humana is to set up screening processes for the early identification of HI. They should do so through three avenues: (1) deploying our predictive model for regular, automated screening of the data, (2) collaborating with primary care providers to integrate SDoH screening in general primary care visits, and (3) initiating an information-sharing arrangement with housing assistance providers.

Part 1: Predictive model deployment
First, we believe Humana should conduct regular, automatic screening for housing insecurity based on our predictive model. Screening should occur daily, and relevant employees should receive an automatic notification whenever a new housing-insecure individual is flagged by our model. This way, Humana employees can (1) identify at-risk individuals more rapidly, (2) open up an investigation to validate the model's flagging, and (3) if deemed necessary, connect the individual to the relevant resources that can keep them in their home and prevent a cascade of harmful events (see *Recommendation 2: emergency response system*). However, before deploying our machine learning model in the real world, we must determine an appropriate probability cut-off above which we would classify an individual as housing insecure.

Predictive model deployment: Determining the appropriate probability cut-off

In determining the appropriate probability cut-off, we must take into consideration that there exists a tradeoff between recall (i.e. ability to identify true HI cases) vs. precision (i.e. ability to reduce the number of individuals the model falsely predicts as HI), and there are costs associated with both factors. (1) When the model falsely classifies an individual to be HI, then Humana has to bear the cost of verifying their actual housing situation and also bear the risk of spending housing assistance resources to help individuals who did not really need it. (2) When the model falsely classified an individual as not HI, Humana may encounter high preventable healthcare claims. In order to minimize overall costs, we recommend a two-step screening process to effectively identify Medicare members who are truly suffering from housing insecurity:

<u>Phase 1 Screening: Use the ML model for all Medicare members part of Humana's network</u>

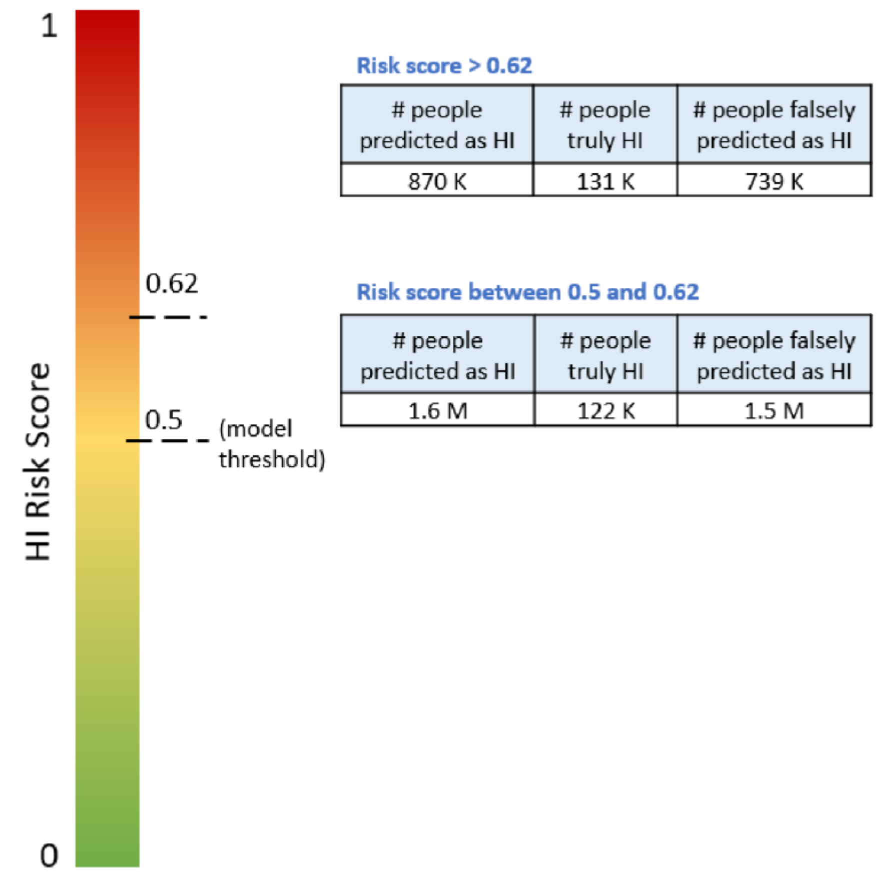
For the first screening phase, we recommend a model threshold of 0.5 based on the provided lift table (*see Table 11*). At this threshold, we expect to capture 2 in 3 true HI cases. However, the model may still falsely classify ~2M members as HI when applied to Humana's 8.7M Medicare members (Humana, 2022a). To reduce this error, we recommend an additional screening step as explained below.

<u>Phase 2 Screening: Use different communication channels to weed out falsely identified HI cases</u>

In phase 2, we recommend a screening approach based on risk scores (i.e. predicted probabilities). On the next page, *Figure 11* illustrates the tiered precision of our model across different probability thresholds. This tiered approach can be used to identify high-risk and low-to-medium risk members, and we recommend the use of different communication channels for each group based on their assigned risk to minimize cost:

- **For individuals who have a risk score of 0.62 or higher:**
  Since this group has a high risk of HI and a relatively high precision of 13%, Humana should utilize multiple channels to contact these individuals. The following sequence of channels can be used to reach the high-risk Humana members:
  1. Survey through email (low cost & scalable)
  2. Interactive Voice Response (IVR) recorded survey (low cost & scalable)
  3. Survey through mail (low cost & scalable)
  4. Survey over phone call with a Humana representative when the member does not respond to multiple calls from IVR or to emails (high cost & not scalable)
  5. In-person visits are limited to areas with a high density of high-risk members when the above channels did not yield a response (very high cost and least scalable).

- **For individuals who have a risk score between 0.5 & 0.62:**
  Since this group has a low-to-medium risk of HI and a relatively low precision of 7%, Humana should only employ low-cost channels to contact individuals predicted to be HI to screen for true HI cases.
  1. Survey through email (low cost & scalable)
  2. IVR recorded survey (low cost & scalable)
  3. Survey through mail (low cost & scalable)



Risk score > 0.62

# people predicted as HI	# people truly HI	# people falsely predicted as HI
870 K	131 K	739 K

Risk score between 0.5 and 0.62

# people predicted as HI	# people truly HI	# people falsely predicted as HI
1.6 M	122 K	1.5 M

Figure 11: Two-step screening for housing insecurity

Note that the values in *Figure 11* are scaled to represent the 8.7M Humana Medicare members (Humana, 2022a).

Predictive model deployment: Improving data availability

Next to determining the appropriate probability cutoff, Humana should also undertake the following actions to improve the input data on which the model relies:

1. **Improve the availability of financial data.**
   As previously explained, some of the most predictive variables in our model are the short-term loan index (STLI), the gender pay gap, median household income, and CMS subsidies. We believe that other financial variables (e.g. loan repayments, credit scores, recent credit card usage, etc.) could be used to paint a better picture of the member's financial state. HI is strongly associated with financial difficulties, hence this additional data would help us better identify Humana members experiencing housing insecurity.

2. **Obtain other sources of information w.r.t health data (other than medical claims).** As previously noted, individuals experiencing housing insecurity are more likely to postpone or avoid primary care. Hence, the absence of claims for medical variables (e.g. depression, anxiety, substance abuse) among housing-insecure individuals may be an inaccurate representation of the actual health of the housing-insecure individual. In our cluster analysis (*Section 4.5*), we identified a **cluster of housing-insecure homeowners** where our model could not pin down effective predictors for housing insecurity. In order to obtain a better

picture of the health of the housing insecure members in this cluster, Humana should examine other sources of health information.

3. **Obtain additional demographic and social information from our members.**
Finally, we recommend that Humana collects more relevant information about the demographic and social characteristics of the member. This will help us further identify potential reasons behind the housing insecurity of the **HI homeowners cluster**. Based on our literature review, we suggest that Humana collects additional information about the following factors: mortgage pay-off data, homeownership type (shared homeownership vs. individual homeownership), abuse history, relationship status (married, going through diverse, divorced, etc.), LGBTQ+ status and social ties (e.g. the number of small children for adults).

Part 2: Collaborating with primary care providers

Second, next to deploying the predictive model, we recommend that Humana collaborates with primary care providers (PCPs) to include SDoH screening in regular health check-ups. We previously identified that two out of our three housing insecurity clusters are effectively able to manage their health conditions (see *section 4.5: clustering analysis*). Given that two out of our three clusters visit their PCPs, we recommend that Humana leverages this interaction to obtain information about housing insecurity and other SDoH. Humana should organize a short SDoH survey and share this with PCPs, so that PCPs can administer this survey to their patients during regular health check-ups. Based on the answers to this survey, Humana can link the individual to key players in their housing network (see *Recommendation 2: Emergency response system*).

Part 3: Collaborating with housing assistance providers

Finally, we recommend that Humana initiates an information-sharing arrangement between Humana and housing assistance providers. In doing so, Humana should focus on key operational areas where Humana sees the largest opportunities for improving housing insecurity: Florida, Georgia, Illinois, Indiana, Kentucky, Louisiana, Ohio, Oklahoma, South Carolina, Virginia, and Wisconsin (Humana, 2022b). In general, when individuals are experiencing housing insecurity, they are more likely to reach out to housing assistance providers than they are to reach out to their healthcare provider(s). As such, housing assistance providers may be aware of housing-insecure Humana members before our predictive model has identified these members as housing insecure.

5.2.2 Recommendation 2: Emergency Response System

We recommend that Humana organizes an emergency response system for at-risk individuals, to ensure that these individuals do not fully become *homeless* and to prevent a cascade of harmful events. We recommend setting up this system through two core modules: (1) Humana-arranged housing support services with personalized case management and (2) an external social care network for patients referred by the healthcare system

Housing Support Systems: Eviction Prevention and Relocation

First, we recommend that Humana grows the responsibilities of its service department that aids members who are at risk of losing their place of residence. Each individual who was identified to be *at-risk* should be connected with a **personal case manager**. This case manager will be responsible for the supporting the member in the following areas:

1. **Housing support services**.
   First, the case manager will be responsible for support services in the areas of eviction prevention and relocation. This means the case manager should help the member in their housing search, communications with property managers, and link the member to any required external services such as low-cost legal representation. Note that Humana already provides housing support services, however, we recommend upgrading this service through a personal case manager.
2. **Short-term emergency/transitional housing.**
   Second, the case manager should leverage the external care network (see *External Social Care Network*) to connect the member to short-term emergency or transitional housing.
3. **Need-based emergency funding.**
   Finally, Humana should set an emergency fund to provide limited financial resources deemed critical for the member's stabilization. We recommend that Humana should sponsor the following elements when deemed critical for relocation or eviction prevention: housing deposits, move-in fees, and moving costs (trucks and temporary storage). The case manager should determine which resources are critical for stabilization on a case-by-case basis.

External Social Care Network

In a recent press release, one of Humana's key competitors – Highmark Health – announced the launch of a multi-year initiative to create a social care network that will compensate nonprofits that address SDoH for patients referred by the healthcare system. In the first year of the pilot, 20 nonprofits will participate in the program, where they can earn value-based reimbursement. Through this system, Highmark Health aims to create the much-needed link between at-risk patients and social services, so that patients' health outcomes are improved and avoidable healthcare costs are reduced (Highmark Health, 2022). We recommend that Humana closely monitors the effectiveness of Highmark Health's pilot, and if deemed effective, Humana should arrange a similar social care network in its key communities: Florida, Georgia, Illinois, Indiana, Kentucky, Louisiana, Ohio, Oklahoma, South Carolina, Virginia and Wisconsin (Humana, 2022b).

5.2.3 Recommendation 3: Proactive Outreach and Coverage Reminders for Cluster 1

In our cluster analysis, we saw that one of our three housing-insecure clusters tends to be poor at managing their health conditions: the ESRD/Disabled mental care cluster. However, the postponement or avoidance of required health check-ups tends to result in large expenses in *avoidable* emergency care. In order to prevent these large emergency care expenses, we recommend that Humana undertakes the following two actions:

1. **Send Medicare coverage reminders to members of Cluster 1**
   Humana should proactively reach out to the individuals in this cluster to remind them of the free and low-cost healthcare that they can easily obtain with their Medicare coverage. Humana should send **weekly emails** to these members that highlight their free and low-cost healthcare resources in the email's title, to remind members of their coverage and to motivate them to *not* forgo their primary care visits. If emails do not prove to be effective for certain members, Humana employees should **politely call the member on a bi-monthly basis** to check in about

their health and remind the member of any healthcare resources that may be relevant for the member based on the conversation.

2. **Highlighting non-emergency medical transport (NEMT) availability for Cluster 1.** As previously noted, a large share of the members of the **ESRD/Disabled mental care cluster** are evidently disabled. A large body of literature indicates that disabled individuals often forego primary care because of transportation issues (Bruns, 2020). Humana already offers NEMT benefits through Logisticare for all Medicaid members and 42% of Medicare Advantage members (Humana, 2019). To reduce the transportation hurdle for members of Cluster 1, we recommend that Humana highlights the availability of NEMT coverage in the aforementioned regular weekly emails and bi-monthly calls. Individuals may be unaware of NEMT benefits and forego treatment simply because of a lack of knowledge.

5.2.4 Recommendation 4: Long-term strategic investments

As a fourth recommendation, we suggest that Humana should engage in long-term strategic investments in two areas: (1) low-rent affordable housing and (2) permanent supportive housing through external parties and a pay-for-success model.

1. **Invest in low-rent affordable housing.**
   As previously noted, Humana has invested $25M in 2021 and another $25M in 2022 to create low-cost rental units in many of its key communities (Humana, 2021). We recommend that Humana repeats this investment annually: the organization should invest another $25M in 2023 and the years after that. Another insurance provider who has engaged in similar investments – United Health – has openly stated that its investments in low-cost rental units have proven to be a *profitable* investment (Khemlani, 2022). We believe the same would be true for Humana, hence we recommend that Humana continues its annual investments in affordable housing. Humana should ideally build **multi-unit and multi-family rental properties**, to strive for the maximum possible impact. In doing so, Humana should take advantage of (A) the Section 8 Housing Choice Voucher Program and (B) the low-income housing tax credits provided by the US government. Under Section 8, the US government pays for 70% of housing and utilities for eligible low-income households, and the renters are solely responsible for the remaining 30% of the rent. The tax credits go up to 9% and can be used for the rehabilitation and/or acquisition of affordable housing (Schreiber, 2021).

2. **Permanent Supportive Housing through a pay-for-success compensation model.**
   Permanent supportive housing (PSH) refers to long-term rental assistance paired with support services. PSH is typically designed for individuals/families who have experienced chronic homelessness or who face large barriers to stability because of chronic health conditions, disabilities, mental illness or a history of substance abuse (Mayors & CEOs for U.S. Housing Investment, n.d.). We recommend that Humana organizes PSH primarily for **cluster 1 (ESRD/Disabled mental care cluster)**. Humana should not focus its PSH resources on cluster 2 (Frail OASI cluster), as this cluster is largely served by government and non-profit organizations. Regarding cluster 3 (Homeowners), eligibility for PSH should be determined on a case-by-case basis. The current dataset does not allow for the identification of the

reasons behind this cluster's housing insecurity status. Humana should investigate the reasons behind HI for these cluster members before determining eligibility for PSH.

All in all, cluster 1 (ESRD/Disabled mental care cluster) is relatively more underserved and represents a larger financial burden for Humana as an insurance provider, so investing in this cluster will (A) have a larger positive health impact and (B) result in more long-term savings as this cluster is less effective at managing their health conditions (which tends to result in expensive emergency department visits).

We recommend that PSH is implemented through the increasingly popular **pay-for-success model,** where the investor receives payment for the achievement of measurable progress and outcomes (Nonprofit Finance Fund, 2019). The mission-driven investor typically covers the initial investment cost, but they will be gradually repaid once they have achieved certain predetermined goals. In the context of PSH, Humana should align these payment-triggering milestones to its priority objectives. We believe that the type of PSH as well as the payment-triggering milestones should differ across the two selected clusters:

Table 13: PSH structure and associated goals

Suggested PSH structure	**Pay-for-success: Goal**
Members of cluster 1 have high rates of neurological and mental disorders, and they are relatively ineffective at managing their health. Hence, we recommend structuring PSH in the form of **multi-unit group homes for disabled individuals with in-house care**, ideally built in the proximity of mental- and neurological healthcare providers.	(1) Emergency care utilization p.p. (2) % change in yearly medical costs p.p.

5.2.5 Recommendation 5: Influence public policy

Finally, it must be highlighted that housing insecurity does not only impact healthcare. The issue is also interconnected with criminal justice, racial inequality, education and employment. Given HI's impact on multiple national systems, addressing housing insecurity should be a multi-stakeholder effort with significant governmental support. After all, the government (or the taxpayer) is the primary beneficiary of housing programs: a reduction in housing insecurity results in lower costs for social welfare programs over time, which subsequently boosts economic productivity and development while reducing tax fund consumption (The Health Equity Project, 2022).

However, it is widely known that the US does not spend a comparable amount on social support relative to other OECD countries (incl. housing, food, education, cash assistance, and care for children and the elderly). At the same time, the US spends a very large share of its GDP on healthcare (18%) as compared to the other OECD countries (8.6%).

In order to influence public policy and government spending, we recommend that Humana becomes an active member of major healthcare and housing associations. For instance:

1. **Example Healthcare Association**
   Humana should join the large and impactful *Healthcare Anchor Network* – an association that aims to promote industry collaboration for proactively addressing economic and racial inequities that create poor health (SDoH) (Healthcare Anchor Network, n.d.). By joining this association, Humana can actively promote the importance of *housing* as a SDoH on the agenda of the association's 400 members.

2. **Example Housing Association**
   Humana should consider becoming a private sector partner for the *Mays and CEOs for US Housing Investment* like its competitor, Kaiser Permanente (Institute for Health Policy, 2022). Through this initiative, Humana can directly communicate with mayors and CEOs who are major decision-makers for housing-related investments. Hence, Humana could use this opportunity to (A) promote affordable housing reforms and (B) oppose cuts to affordable housing.

5.3 Expected Value for Humana

In total, housing insecurity costs the US healthcare system $11 billion dollars per year (Poblacion et al., 2017). Given this value and information about (A) Medicare and Medicaid market share and (B) Humana's presence in each segment, we estimate that the burden of HI costs Humana $265M per year. This burden could be significantly reduced by implementing the recommendations from *Section 5.2*.

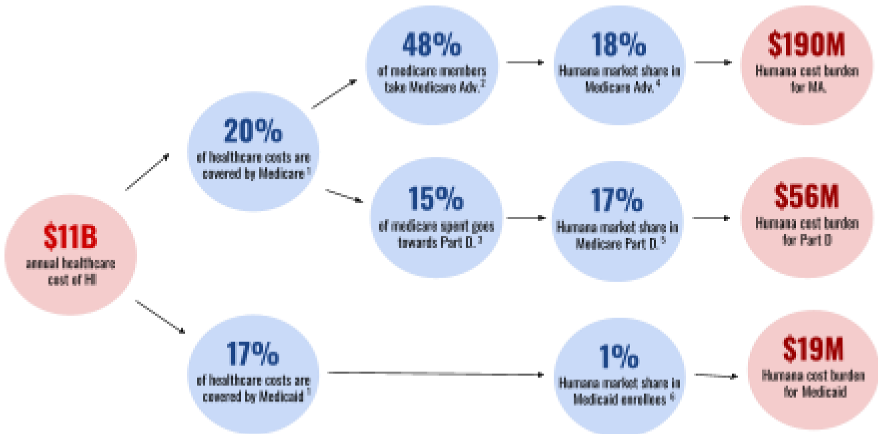


Figure 12: The cost of HI for Humana

In *Table 14* on the next page, we provide an overview of the estimated costs and revenues associated with the recommendations provided. The sources and calculations behind these financials can be found in *Appendix B*.

To implement all of our recommendations, Humana would need to commit ~$37.1M as an initial investment. To maintain the implemented solutions over time, we estimate annual costs of ~$7.4M. This resource commitment would return an estimated annual revenue of ~$20.2M. Based on these numbers, we expect a **payback period** of roughly three years:

$$payback\ period = \underline{initial} annual\ \underline{investment}\ profit = 20.237M. - 1M7.4M \approx$$ 2.9 years

[1] CMS (2018). *National Health Expenditure 2017 highlights*. Centers for Medicare & Medicaid Services.

[2] Morse, S. (2021). *The Disadvantages of a Medicare Advantage Plan*. Healthcare Finance News.

[3] KFF (2021). *An overview of the Medicare part D prescription drug benefit.* Kaiser Family Foundation.

[4] KFF (2022). *Medicare Advantage in 2022: Enrollment Update and Key Trends.* Kaiser Family Foundation.

[5] KFF (2022). *Key facts about Medicare Part D enrolment and costs in 2022.* Kaiser Family Foundation [6] Humana (2022). *Humana Healthy Horizons in Ohio*. Humana.

Table 14: Financial Analysis

		Est. initial investment	Est. annual revenue	Est. annual costs
Screening and early identification	Predictive model deployment	**-$8K**	N/A – *This recommendation drives*	**-$12K**
	Collaboration with primary care providers	**-$70K**		**-$811K**
	Collaboration with housing assistance firms	**-$70K**	*solutions*	**-$14K**
Emergency response system	Housing support services	N/A – *Limited services already offered*	**+$2.48M**	**-$800K**
	Short-term emergency & transitional housing	N/A – *Limited services already offered*	**+$1.24M**	**-$400K**
	Need-based emergency stabilization fund	N/A	**+$414K**	**-$400K**
	External social care network	**-$2M**	**+$552K**	**-$160K**
Proactive Outreach for Cluster 1	Coverage reminders	N/A – *Infrastructure in place.*	**+$2.10M**	**-$200K**
	Non-emergency medical transportation	N/A – *Infrastructure in place.*	**+$6.18M**	**-$1.46M**
Long-term strategic investments	Invest in low-rent affordable housing for members with proven housing insecurity	**-$25M**	**+$5.76M**	**-$2.10M**
	Permanent Supportive Housing *for members with chronic homelessness or members at high risk of homelessness*	**-$10M**	**+1.44M**	**-$1.01M**

revenue by supporting the other

6. CONCLUSION

To help Humana address the issue of housing insecurity, we built a predictive ensemble model based on a Neural Network, XGBoost, and Gradient Boosted Decision Trees. We ultimately achieved a holdout AUC of 0.7598 and an excellent fairness score of 0.992. By considering the feature importances of our final model, we were able to identify key drivers and protective factors with respect to housing insecurity. Features such as home ownership, age, and income were found to be negatively correlated with housing insecurity and features related to mental health and/or neurological disorders and CMS subsidies were found to be positively correlated with housing insecurity.

Further, by conducting K-means clustering based on key demographic and predictive features, we were able to identify three distinct segments of Humana members experiencing housing insecurity:

1. **ESRD/Disabled mental care cluster** – Non-OASI individuals with mental health conditions and/or neurological disorders.
2. **Frail OASI cluster –** OASI individuals who are frail and do not own their place of residence.
3. **HI-homeowners cluster –** HI Medicare members who actually are homeowners.

Our model has the highest error rate within the HI-homeowners cluster. Based on the given dataset, we are unable to identify the exact reasons for housing insecurity within this cluster. To better predict HI within this cluster in the future, Humana should include additional data points such as mortgage, the individual's income, family size, and recent disaster events by geography to improve model performance.

Finally, based on the cluster analysis and feature importances, we generated actionable and cluster-tailored recommendations. We carefully analyzed and evaluated our predicted HI probabilities to determine the optimal probability cutoff above which Humana should classify a member as potentially housing insecure, and devote resources towards tracking and tackling housing insecurity. To combat HI, we provided recommendations in five key areas: (1) Screening and early identification, (2) an emergency response system, (3) proactive outreach for select clusters, (4) long-term strategic investments, and (5) influencing public policy.

Implementing all of our recommendations would require an initial investment of $37.1M, however, we expect that this investment can be recouped within a three-year payback period.

7. REFERENCES

Anderson, C. (2018, July 13). *Finding the Right Fit: Imagining Health & Housing Partnerships that Work* [PowerPoint slides]. Housing Is. https://housingis.org/sites/default/files/CLPHA%20Housing%20Healthcare%20_C.%20Anderson%207.14.17%20Final.pdf

ATI Advisory. (2021). *Data Insight: Special Supplemental Benefits for the Chronically Ill in Plan Year 2021.* Retrieved 11 October 2022, from https://atiadvisory.com/wp-content/uploads/2021/01/ATI-Advisory-Data-Insight_Special-Supplemental-Benefits-for-the-Chronically-Ill-in-Plan-Year-2021.pdf

Bagi, N. (2022). Anomaly Detection in Financial Data Using Isolation Forest. Retrieved 11 October 2022, from https://codime.medium.com/anomaly-detection-in-financial-data-using-isolation-forest-2515b 4572bc8

Bhayani, A. (2020). Isolation Forest algorithm for anomaly detection. Retrieved 11 October 2022, from https://medium.com/@arpitbhayani/isolation-forest-algorithm-for-anomaly-detection-f88af2d 5518d

Bruns, K. (2020). Kansans with Disabilities: Transportation Barriers = Health Barriers. Retrieved 16 October 2022 from https://ihdps.ku.edu/sites/ihdps/files/documents/Transportation%204.28.2020--fixed.pdf

Coleman, K. (2019, April 24). *Implementing Supplemental Benefits for Chronically Ill Enrollees.* Department of Health & Human Services, Centers for Medicare & Medicaid Services. Retrieved 11 October 2022, from https://www.hhs.gov/guidance/sites/default/files/hhs-guidance-documents/Supplemental_Ben efits_Chronically_Ill_HPMS_042419.pdf

Collins, K. (2016). Kidney Failure and Medicare: What you should know. Retrieved 11 October 2022, from https://www.medicarerights.org/guest-posts/2016/09/27/blog-kidney-failure-medicare-know

Commonwealth of Virginia. (2015, December). *Performance and Pricing of Medicaid Non-Emergency Transportation.* Retrieved 15 October 2022 from http://jlarc.virginia.gov/pdfs/reports/Rpt477.pdf

Cox, R., Henwood, B., Rodnyansky, S., Rice, E., & Wenzel, S. (2019). Road map to a unified measure of housing insecurity. *Cityscape*, *21*(2), 93-128.

Early, D. W. (2005). An empirical investigation of the determinants of street homelessness. *Journal of Housing Economics*, *14*(1), 27-47.

Gramlich, E. (2022, September). *The National Housing Trust Fund: An Overview of 2017 State Projects.* The National Low Income Housing Coalition. Retrieved 11 October 2022, from https://www.nlihc.org/sites/default/files/2022-09/nhtf-overview-2017-state-projects.pdf

Healthcare Anchor Network. (n.d.). *About the Healthcare Anchor Network.* Retrieved 16 October 2022 from https://healthcareanchor.network/about-the-healthcare-anchor-network/

Highmark Health. (2022, June 2). Highmark Health, Allegheny Health Network announce 20 community benefit organizations to participate in $1M value-based arrangement [Press release]. https://www.highmarkhealth.org/hmk/newsroom/pr/2022/2022-06-02-Value-Based-Arrangement.shtml

American Institutes for Research (2017). *H^2 Final Report.* U.S. Department of Housing and Urban Development. Retrieved 11 October 2022, from https://files.hudexchange.info/resources/documents/H2-Final-Report.pdf

Housing Instability - Healthy People 2030 | health.gov. (2022). Retrieved 10 October 2022, from https://health.gov/healthypeople/priority-areas/social-determinants-health/literature-summaries/housing-instability

Humana. (2019, June). *Transportation Issue Brief.* Retrieved 11 October 2022, from https://populationhealth.humana.com/wp-content/uploads/2019/08/BG_Transportation-IssueBrief_1Q2019_EXTERNAL.pdf

Humana. (2020a, April 3). *Humana Launches Innovative Value-Based Program to Address Social Determinants of Health.* Retrieved 11 October 2022, from https://press.humana.com/news/news-details/2020/program-to-address-social-determinants-of-health/default.aspx#gsc.tab=0

Humana. (2020b, June). *Housing Issue Brief.* Retrieved 15 October 2022 from https://populationhealth.humana.com/wp-content/uploads/2020/06/Humana_HousingBrief_Final_External_version_2020.pdf

Humana. (2021, September 9). *Humana Expands Upon Its Bold Goal Initative with $25 Million Affordable Housing Investment.* Retrieved 11 October 2022, from https://press.humana.com/news/news-details/2021/Humana-Expands-Upon-Its-Bold-Goal-Initiative-with-25-Million-Affordable-Housing-Investment/default.aspx#gsc.tab=0

Humana. (2022a, January 10). *Humana significantly expands Medicare Advantage health plan offerings in 2023 with focus on greater value for members.* Retrieved October 15 2022 from https://press.humana.com/news/news-details/2022/Humana-significantly-expands-MedicareAdvantage-health-plan-offerings-in-2023-with-focus-on-greater-value-for-members/default.as px#gsc.tab=0

Humana. (2022b, April 13). *Humana Expands National Commitment to Affordable Housing With Additional $25 Million Investment.* Retrieved 11 October 2022, from https://press.humana.com/news/news-details/2022/Humana-Expands-National-Commitment-to-Affordable-Housing-With-Additional-25-Million-Investment/#gsc.tab=0

Institute for Health Policy. (2022, April 2022). *Addressing Housing Affordability.* Kaiser Permanente. Retrieved 16 October 2022 from https://www.kpihp.org/wp-content/uploads/2021/05/IHP_Housing_flyr_final.pdf

Khemlani, A. (2022, July 12). *Why United Health, CVS, Humana and Others Are Investing in Affordable Housing.* Yahoo Finance. Retrieved 11 October 2022, from https://finance.yahoo.com/news/why-united-health-cvs-humana-and-others-are-investing-in-affordable-housing-194319528.html

Krishnan, S., Cozier, Y. C., Rosenberg, L., & Palmer, J. R. (2010). Socioeconomic status and incidence of type 2 diabetes: results from the Black Women's Health Study. *American journal of epidemiology*, *171*(5), 564-570.

Lagasse, J. (n.d.). *Kaiser Permanente is spending $400 million on affordable housing initiatives.* Retrieved 15 October 2022 from https://www.healthcarefinancenews.com/news/kaiser-permanente-spending-400-million-affor dable-housing-initiatives

Maqbool, N., Viveiros, J., & Ault, M. (2015). The Impacts of Affordable Housing on Health: A Research Summary. Retrieved 10 October 2022, from https://nhc.org/wp-content/uploads/2017/03/The-Impacts-of-Affordable-Housing-on-Health-A -Research-Summary.pdf

Mayors & CEOs for U.S. Housing Investment (2022).. *Questions and Answers About Homelessness.* Retrieved 11 October 2022, from https://housinginvestment.org/q-and-a/

Medicaid and CHIP Payment and Access Commission. (2018, October). *Medicaid's Role in Housing.* MACPAC. Retrieved 11 October 2022, from https://www.macpac.gov/wp-content/uploads/2018/10/Medicaid%E2%80%99s-Role-in-Housi ng.pdf

National Ambulatory Medical Care. (2018). *National Ambulatory Medical Care Survey: 2018 National Summary Tables.* Retrieved 15 October from https://www.cdc.gov/nchs/data/ahcd/namcs_summary/2018-namcs-web-tables-508.pdf

Nonprofit Finance Fund. (2019, May). *Pay for Success: The First 25.* Retrieved 11 October 2022, from https://nff.org/sites/default/files/paragraphs/file/download/pay-for-success-first-25.pdf

Options for Dialysis. (2022). Retrieved 11 October 2022, from https://lifeoptions.org/living-with-kidney-failure/options-for-dialysis/

Perlato, A. (2022). Feature Selection using Boruta Algorithm. Retrieved 11 October 2022, from https://www.andreaperlato.com/mlpost/feature-selection-using-boruta-algorithm/

Poblacion, A., Bovell-Ammon, A., Sheward, R., Sandel, M., Ettinger de Cuba, S., Cutts, D. and Cook, J. (2017, July). *Stable Homes Make Healthy Families.* Children's HealthWatch. http://childrenshealthwatch.org/wp-content/uploads/CHW-Stable-Homes-2-pager-web.pdf

Robbins, J. M., Vaccarino, V., Zhang, H., & Kasl, S. V. (2005). Socioeconomic status and diagnosed diabetes incidence. *Diabetes research and clinical practice*, *68*(3), 230-236.

Rudden, J. (2022, March 22). *Medical membership of Humana in the United States from 2008 to 2021* [Infographic]. Statista. Retrieved 15 October 2022 from https://www.statista.com/statistics/210651/total-medical-membership-of-humana/

Schreiber, L. (2021, November 9). *Investing in Affordable Housing: Everything You Need to Know.* NorthEast Private Client Group. Retrieved 11 October 2022, from https://northeastpcg.com/blog/investing-in-affordable-housing/

Shinn, M., Weitzman, B. C., Stojanovic, D., Knickman, J. R., Jimenez, L., Duchon, L., ... & Krantz, D. H. (1998). Predictors of homelessness among families in New York City: from shelter request to housing stability. *American Journal of Public Health*, *88*(11), 1651-1657.

Social Security Administration. (2022). *Fact Sheet Social Security.* Retrieved 16 October 2022 from https://www.ssa.gov/news/press/factsheets/colafacts2022.pdf

Stahre, M., VanEenwyk, J., Siegel, P., & Njai, R. (2015). Peer reviewed: Housing insecurity and the association with health outcomes and unhealthy behaviors, Washington State, 2011. *Preventing chronic disease*, *12*.

Strawser, J., Monsees, G., Dimitrion, G., Franklin, S., Haugh, G. (2022, September 14). *Sixth Annual Humana/Mays 2022 Healthcare Analytics Case Competition.* Humana / Texas A&M University Mays Business School. http://mays.tamu.edu/humana-tamu-analytics/wp-

content/uploads/sites/113/2022/09/Humana_
Mays_Case_Competition_Kickoff_FINAL_08182022.pdf

Taylor, L. (2018). Housing And Health: An Overview Of The Literature. Retrieved 10 October 2022, from https://www.healthaffairs.org/do/10.1377/hpb20180313.396577/

The Accountable Health Communities Health-Related Social Needs Screening Tool. (2017). Retrieved 10 October 2022, from https://innovation.cms.gov/files/worksheets/ahcm-screeningtool.pdf

The Health Equity Project. (2022). *Building the value case: Medicaid's role in addressing social determinants of health. Housing Insecurity.* Algorex, Pareto Intelligence, Third Horizon Strategies. Retrieved October 15 2022 from https://projecthealthequity.org/wp-content/uploads/Housing-instability-brief.pdf

Tompkins, L. (2022). *If Housing Is a Health Care Issue, Should Medicaid Pay the Rent?.* New York Times. Retrieved 11 October 2022, from https://www.nytimes.com/2022/06/14/headway/medicaid-housing-rent-health.html

U.S. Department of Housing and Urban Development. (2022). *Rental Assistance.* Retrieved 11 October 2022, from https://www.hud.gov/topics/rental_assistance

U.S. Department of Housing and Urban Development. (n.d.). *Public Housing.* Retrieved 15 October from https://www.hud.gov/program_offices/public_indian_housing/programs/ph

U.S. homeownership rate by age 2021 | Statista. (2022). Retrieved 11 October 2022, from https://www.statista.com/statistics/1036066/homeownership-rate-by-age-usa/

Where We Live Matters for Our Health: The Links Between Housing and Health. (2008, October 16). *Housing and Health Issue Brief.* Robert Wood Johnson Foundation. http://www.commissiononhealth.org/PDF/e6244e9e-f630-4285-9ad7-16016dd7e493/Issue%20Brief%202%20Sept%2008%20-%20Housing%20and%20Health.pdf

ZipRecruiter. (n.d.). *Housing Authority Salary in California.* Retrieved 15 October from https://www.ziprecruiter.com/Salaries/Housing-Authority-Salary--in-California#:~:text=As%20of%20Oct%201%2C%202022,be%20approximately%20%2426.08%20an%20hour

Zumper. (2020, March 5). *A Guide to Section 8 Housing in Los Angeles.* Retrieved 15 October 2022 from https://www.zumper.com/blog/a-guide-to-section-8-housing-in-los-angeles/

8. APPENDIX

Appendix A: Feature Engineering

Variable A	Variable B.	Interaction
Home-ownership (cons_homstat_Y)	low income (cms_low_income_ind)	Multiplication

Home-ownership (cons_homstat_Y)	days since last claim for physician office (total_physician_office_ds_clm)	Multiplication
Home-ownership (cons_homstat_Y)	Short term loan (cons_stlindex)	Multiplication
Risk score assigned based on medical claims (cms_ma_risk_score_nbr)	Managing an illness/condition index (cons_hxmioc)	Division
Age (est_age)	The binary variable indicating that the member is under age 65 (cms_disabled_ind)	Multiplication
allowed cost per month for overall claims related to physician office in the past one year (total_physician_office_allowed_pmpm_cost)	Indicator that the member is disabled (cms_orig_reas_entitle_cd_Disable)	Multiplication
Total occupied housing units (atlas_totalocchu)	Male gender (sex_cd_M)	Multiplication
Adult Diabetes Rate (atlas_pct_diabetes_adults13)	Homeowner Status (cons_homstat_Y)	Multiplication
Binary indicator that a member is receiving a subsidy from CMS (cms_low_income_ind)	Homeowner Status (cons_homstat_Y)	Multiplication
Binary indicator that a member is receiving a subsidy from CMS (cms_low_income_ind)	days since last claim for overall claims related to physician office in the past one year (total_physician_office_ds_clm)	Multiplication

Appendix B: Financial Analysis

		Est. initial investment	Est. annual revenue	Est. annual costs
Screening and early identification	Predictive model deployment	**-$8K** 2 headcount at $100K for 2 weeks: $100K * 2 * 2 / 52 = $8K		**-$12K** -$60K[1] / 5 years = -$12K
	Collaboration with primary care providers	**-$70K** 2 headcount at $70K for 6 months: $70K * 2 * 0.5 = $70K		**-$811K** 8.7M[2] Humana Medicare members 326.69[3]million US citizens in 2018 860.4M[4] doctor visits per year in 2018 85%[5] of US citizens have a smartphone → automatic data collection, other 15% needs to be a paper-based and inserted into the data system manually (2 min) avg. medical receptionist salary: $29,462[6] 860.4M / 326.69M * 8.7M * (1 min/ 60 min) * 15% * $29,462 / (40 hrs/week * 52 weeks) = $811K

	Collaboration with housing assistance providers	**-$70K** 2 headcount at $70K for 6 months: $70K * 2 * 0.5 = $70K		**-$14K** 1.2M[7] families living in public housing 326.69[3] million US citizens in 2018 8.7M[2] Humana Medicare members Avg. salary of housing authority: $54,248[8] 1.2M / 326.69M * 8.7M * (1 min / 60 min) * $54,248 / (40 hrs/week * 52 weeks) = -$14K
Emergency response system	Housing support services		**+$2.48M** *A case manager can deal with ~15 housing insecurity cases on average in a month. Healthcare savings would be ↓$115[9] p.p. per month.* *(10*15*12*115)*12*	**-$800K** 10 additional headcount at $80K annually to handle the additional responsibility: $80K * 10 = $800K
	Short-term emergency and transitional housing		**+$1.24M** *A case manager can deal with ~15 housing insecurity cases on average in a month. Healthcare savings would be ↓$115[9] p.p. per month.* *(5*15*12*115)*12*	**-$400K** 5 additional headcount at $80K annually to handle the additional responsibility : $80K * 5 = $400K

	Need-based emergency stabilization fund		**+$414K** *Across 200 households with an average of 1.5 medicare individuals, healthcare savings would be ↓$115[9] p.p. per month. (200 * 1.5 * 115)* 12 months*	**-$400K** Annual fund: $400,000 *up to $2k per household, support up to 200 high-need households/year*
	External social care network	**-$2M** Initial suggested investment of $2M	**+$552K** *Average value of benefits each individual receives ~$5,000. Total number of individuals = $2M/ $5 K = 400 individuals. Healthcare savings would be ↓$115[9] p.p. per month.* *(400 * 115 * 12mo)*	**-$160K** 2 additional headcount at $80K annually to handle the additional responsibility : $80K * 2 = $160K
Proactive Outreach for Cluster 1	Coverage reminders	**N/A** *Infrastructure is already in place.*	**+$2.10M** *5% of the 8.7M[2] Humana members experience HI. 35% are in cluster 1. Healthcare cost savings would be $115[9] p.p. per month. An incremental increase of 1% in response rate is anticipated* *(0.05*8.7M*0.35*115*0.01)*12mo*	**-$200K** *Two customer service employees for weekly emails and bi-monthly calls for the HI-insure members in cluster 1.*

	Non-emergency medical transportation	**N/A** *Infrastructure is already in place.*	**+$6.18M** *5% of the 8.7M[2] Humana members experience HI. 35% are in cluster Y. 42%[10] have NEMT coverage. Healthcare savings would be ↓$115[9] p.p. per month. 7% incremental utilization in this service is anticipated* *12*(0.05*8.7M*0.35*0.42*0.07)*115*	**-$1.46M** *Estimated 12 rides p.p. per year. 5% of the 8.7M[2] Humana members experience HI. 35% are in cluster Y. 42%[10] have NEMT coverage. Cost is $19[11] per trip. 10% incremental utilization in this service is anticipated* *12*(0.05*8.7M*0.35*0.42*0.1)*19*
Long-term strategic investments	Invest in low-rent affordable housing for members with proven housing insecurity	**-$25M** $25m[12] investment for 250 units[13] *50% one-bedroom units* *50% two-bedroom units* *→ 375 individuals affected*	**+$5.76M** Healthcare savings: ● ↓$115[9] p.p. per month. Monthly rent (affordable apt)[14]: ● $1700 per 1-bedroom apt. ● $2100 per 2-bedroom apt. *(($1700*115+$2100*115)+($115*375))*12*	**-$2.10M** For an apartment building, the operating expenses typically fall between 35-45% . *[0.40*($1700*115+$2100*115)]*12*
	Permanent Supportive Housing for members with chronic homelessness or at high risk of homelessness	**-$10M** *Funding for 100 two-bedroom units using a pay-for-success model.* → *200 individuals affected*	**+1.44M** Healthcare savings: ↓39%[15] in total cost of services p.p. for high-cost individuals. Homeless individuals cost the healthcare system $18.5k per year. *18,500*0.39*200*	**-$1.01M** For an apartment building, the operating expenses typically fall between 35-45% of the rent. *[0.40*($2100*100)]*12*

Sources to the calculations:

[1]

Coop, R. (2021, May 20). *What is the Cost to Deploy and Maintain a Machine Learning Model?* phData. Retrieved October 16 2022 from https://www.phdata.io/blog/what-is-the-cost-to-deploy-and-maintain-a-machine-learningmodel/ [2]

Humana. (2022a, January 10). *Humana significantly expands Medicare Advantage health plan offerings in 2023 with focus on greater value for members.* Retrieved October 15 2022 from https://press.humana.com/news/news-details/2022/Humana-significantly-expands-MedicareAdvantage-health-plan-offerings-in-2023-with-focus-on-greater-value-for-members/default.as px#gsc.tab=0 [3]

US Census Bureau (2022). *US Population by Year.* Multpl. Retrieved October 16 2022 from https://www.multpl.com/united-states-population/table/by-year [4]

National Ambulatory Medical Care. (2018). *National Ambulatory Medical Care Survey: 2018 National Summary Tables.* Retrieved 15 October from https://www.cdc.gov/nchs/data/ahcd/namcs_summary/2018-namcs-web-tables-508.pdf [5]

Pew Research Center. (2021, April 7). *Mobile Fact Sheet.* Retrieved October 16 2022 from https://www.pewresearch.org/internet/fact-sheet/mobile/ [6]

ZipRecruiter. (2022). *Medical Receptionist Salary.* Retrieved October 16 2022 from https://www.ziprecruiter.com/Salaries/Medical-Receptionist-Salary--in-California [7]

U.S. Department of Housing and Urban Development. (n.d.). *Public Housing.* Retrieved 15 October from https://www.hud.gov/program_offices/public_indian_housing/programs/ph

[8]

ZipRecruiter. (2022). *Housing Authority Salary in California.* Retrieved October 16 2022 from https://www.ziprecruiter.com/Salaries/Housing-Authority-Salary--in-California#:~:text=As%20of%20Oct%201%2C%202022,be%20approximately%20%2426.08%20an%20hour [9]

Anderson, C. (2018, July 13). *Finding the Right Fit: Imagining Health & Housing Partnerships that Work* [PowerPoint slides]. Housing Is. https://housingis.org/sites/default/files/CLPHA%20Housing%20Healthcare%20_C.%20Ander son%207.14.17%20Final.pdf [10]

Humana. (2019, June). *Transportation Issue Brief.* Retrieved 11 October 2022, from https://populationhealth.humana.com/wp-content/uploads/2019/08/BG_Transportation-IssueBrief_1Q2019_EXTERNAL.pdf [11]

Commonwealth of Virginia. (2015, December). *Performance and Pricing of Medicaid Non-Emergency Transportation.* Retrieved 15 October 2022 from http://jlarc.virginia.gov/pdfs/reports/Rpt477.pdf [12]

Humana. (2021, September 9). *Humana Expands Upon Its Bold Goal Initative with $25 Million Affordable Housing Investment.* Retrieved 11 October 2022, from https://press.humana.com/news/news-details/2021/Humana-Expands-Upon-Its-Bold-Goal-Initiative-with-25-Million-Affordable-Housing-Investment/default.aspx#gsc.tab=0 13

Lagasse, J. (n.d.). *Kaiser Permanente is spending $400 million on affordable housing initiatives.* Retrieved 15 October 2022 from

https://www.healthcarefinancenews.com/news/kaiser-permanente-spending-400-million-affordable-housing-initiatives 14

Zumper. (2020, March 5). *A Guide to Section 8 Housing in Los Angeles.* Retrieved 15 October 2022 from https://www.zumper.com/blog/a-guide-to-section-8-housing-in-los-angeles/

15 2

American Institutes for Research (2017). *H Final Report.* U.S. Department of Housing and Urban Development. Retrieved 11 October 2022, from https://files.hudexchange.info/resources/documents/H2-Final-Report