2022 HUMANA MAYS CASE COMPETITION

[Author name]

[COMPANY NAME] [Company address]

Table of Contents

Executive Summary

Housing is cited as an important social determinant of health, recognizing the range of ways in which a lack of housing, or poor-quality housing, can negatively affect health and wellbeing. The impact of housing on health is now also being widely considered by policy makers. Housing is one of the best-researched social determinants of health, and selected housing interventions for low-income people have been found to improve health outcomes and decrease health care costs. As a result, many health care systems, payers, and government entities are seeking to better understand the totality of the health and housing literature to determine where they might intervene effectively.

As a leading healthcare provider in the US, Humana is dedicated to helping its members improve their overall healthcare. Humana defines housing insecurity as the lack of access to quality and safe housing. Therefore, the objective of this report is to assist Humana in achieving its goal by leveraging the power of Big Data and Machine Learning. The overall objective is broken down into three tasks:

1. Build a predictive model to identify the members with housing instability
2. Based on insights from the model, uncover the underlying reasons and segments among the affected members
3. Design targeted solutions for different segments with comprehensive analysis

Given 48,300 records of Humana's MAPD members with wide ranging information, we started with researching relevant topics, performing exploratory data analysis, and data cleaning to gain a deeper understanding. Feature selection and engineering techniques

were also applied to prepare for modeling. A two-stage process is then carried out to train and compare the performances of 4 different models. We selected the ROC-AUC score as the metric for evaluation and found the Light GBM Classifier (ROC-AUC score of 0.75) to be our final model with the most predictive power. By studying the important features, we extracted insights for potential segmentation.

For the second task: we successfully distinguished three segments from the population: members with financial distress, older adult members, and members in unfavorable neighborhoods. Every segment has a significantly high probability of housing insecurity risks.

To address the three segments, we have devised a two-detail plan to improve the healthcare for members struggling with housing insecurities. The two detailed plans are

1. Housing for America's Older Adults: Making housing affordable, Safe and Quality Oriented
2. Relief to Housing Insecurity Caused due to Neighborhood Pathway

The plan is to implement these schemes which will provide value benefits in terms of improved health and mental benefits, reduction in diseases caused due to stress and environmental triggers. In addition, our plan is to provide recommendations which will intervene before a member ends up being homeless.

Background

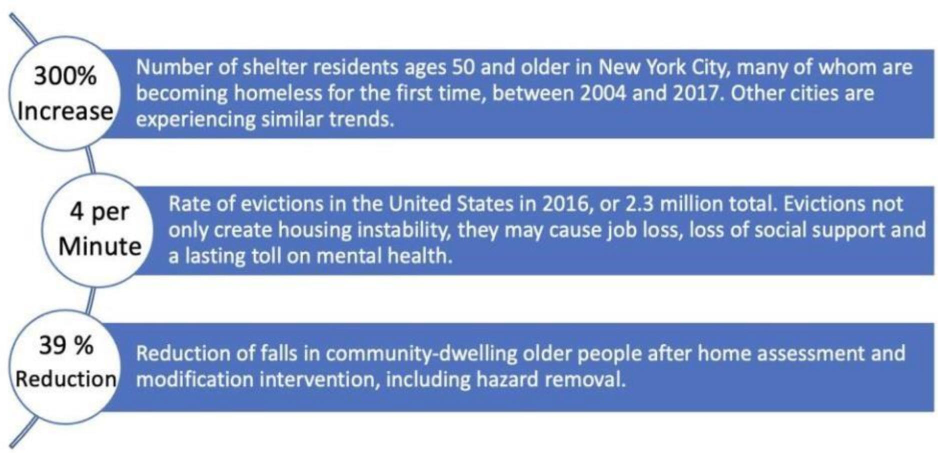
Social Determinants of Health

Social determinants of health (SDOH) are the conditions in the environments where people are born, live, learn, work, play, worship, and age that affect a wide range of health, functioning, and quality-of-life outcomes and risks. SDOH can be grouped into 5 domains[1]:

Social determinants of health (SDOH) have a major impact on people's health, well-being, and quality of life. Examples of SDOH include:

SDOH also contributes to wide health disparities and inequities. For example, people who don't have access to grocery stores with healthy foods are less likely to have good nutrition. That raises their risk of health conditions like heart disease, diabetes, and obesity, and even lowers life expectancy relative to people who do have access to healthy foods.

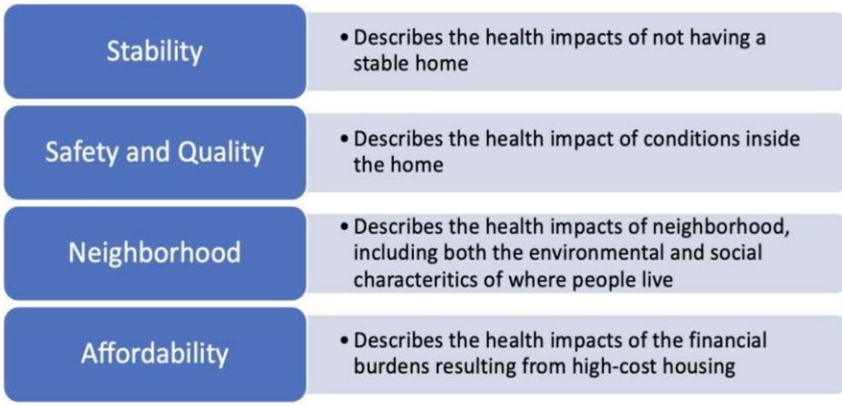
Housing And Health



Housing is cited as an important social determinant of health, recognizing the range of ways in which a lack of housing, or poor-quality housing, can negatively affect health and well-being. The causal relationships between tangible physical housing defects and poor health outcomes are widely accepted with clear evidence of negative physical health effects of toxins within the home, damp and mold, cold indoor temperatures, overcrowding and safety factors, and of negative mental health effects arising from cold indoor temperatures, overcrowding/lack of personal space, and damp and mold. Moreover, analysis of the impact of housing improvement interventions provides evidence for causal direction and pathways.[2]

The impact of housing on health is now also being widely considered by policy makers. Housing is one of the best-researched social determinants of health, and selected housing interventions for low-income people have been found to improve health outcomes and decrease health care costs. As a result, many health care systems, payers, and government entities are seeking to better understand the totality of the health and housing literature to determine where they might intervene effectively.

Existing evidence on housing and health can be understood as supporting the existence of four pathways[3] by which the former affects the latter.



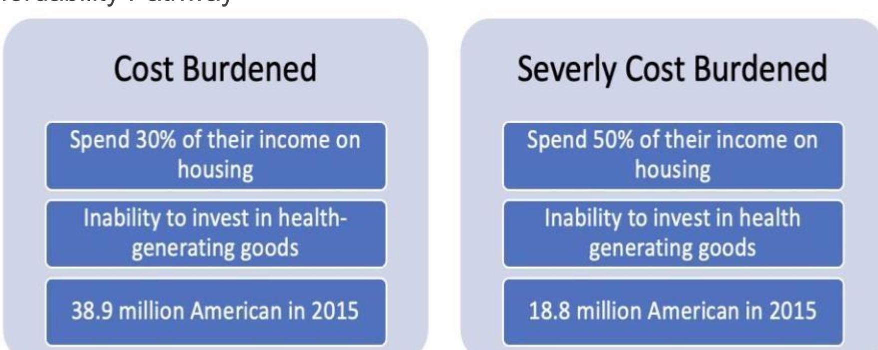
Stability Pathway

People who are chronically homeless or face housing instability in the form of moving frequently, falling behind on rent or couch surfing face higher morbidity in terms of both physical and mental health. These people may face high health care expenditures due to emergency department and inpatient hospital use. The lack of stable housing can also decrease the effectiveness of health care by making proper storage of medications difficult or impossible.

Safety and Quality Pathway

Several environmental factors within homes are correlated with poor health. In-home exposure to lead (Pb) irreversibly damages the brains and nervous systems of children. Substandard housing conditions such as water leaks, poor ventilation, dirty carpets, and pest infestation have been associated with poor health outcomes, most notably those related to asthma. Additionally, exposure to high or low temperatures is correlated with adverse health events including cardiovascular events—particularly among the elderly. Residential crowding has also been linked to both physical illness (for example, infectious disease) and psychological distress.

Affordability Pathway



In some cases, Americans may choose to spend substantially on housing to live in neighborhoods that provide access to health-promoting features such as schools and parks. However, a lack of affordable housing options can affect families' ability to make other essential expenses and can create serious financial strains. Low-income families with difficulty paying their rent or mortgage or their utility bills are less likely to have a usual source of medical care and more likely to postpone needed treatment than those who enjoy more-affordable housing.

Neighborhood Pathway

In the modern era, researchers have found that the availability of resources such as public transportation to one's job, grocery stores with nutritious foods, and safe spaces to exercise are all correlated with improved health outcomes. Living near high-volume roads, in contrast, is a danger to health and can result in increased rates of respiratory diseases such as asthma and bronchitis and increased use of health care.

Business Statement

Housing insecurity as defined by Humana is the lack of access to quality and safe housing. It may include scenarios such as lack of affordability, unsafe/overcrowded conditions or frequent moves. 37.1 million American households are "housing cost burdened," and 32.7% of older adult households have severe housing problems. Resulting Health problems can range from allergies to neurological to heart damage. *Using the provided data we need to identify Medicare members most likely to be struggling with housing insecurity issues and propose solutions that help people achieve their best health.*

Why identifying Medicare members with housing insecurity issues is important?

When we identify Medicare members, we can create member tailored programs to improve their way of life and positively impact their health. These programs have been done previously but were more tailored towards the society or a certain group of people. Below are some past scenarios where interventional studies were performed to see if the housing standards were improved for a member, then their health did see improvement or were better than those who did not have good housing standards.

- *Moving to Opportunity for Fair Housing Demonstration Program* - As part of this program, people were randomly assigned to groups that either did or did not receive financial and other assistance in moving to lower-poverty areas—a research design that overcame unobservable selection effects inherent in many previous studies. Adults who moved experienced improvements in long-term mental health and some aspects of physical health - for example, reductions in the prevalence of obesity and diabetes in comparison to peers who remained in high-poverty areas.

- Within a population of nearly 10,000 people in Oregon with unstable housing, the provision of affordable housing decreased Medicaid expenditures by 12%. At the same time, use of outpatient primary care increased by 20% and emergency department use declined by 18% for this group.
- The Housing First model, in which chronically homeless people with a diagnosis of a behavioral health condition receive supportive housing, has been shown to be particularly cost-effective, with one study finding that the provision of housing generated cost offsets of up to $29,000 per person per year, after accounting for housing costs.
- Studies in which asthma triggers are removed have repeatedly demonstrated health improvements and cost reductions among both children and adults.
- Research on smoking bans in public and affordable housing has found reductions in the number of smokers, the number of cigarettes smoked per smoker, and secondhand smoke exposure among nonsmokers.
- Children in families participating in the federally funded *Low Income Home Energy Assistance Program (LIHEAP)*, which provides financial assistance for home heating, medically necessary home cooling, and emergencies due to weather-related fuel shortages, were at a healthier weight and at less nutritional risk, compared to their nonparticipant peers. Among community-dwelling older adults, home modifications can reduce falls by 39 percent when delivered by occupational therapists, and a randomized controlled trial of a standardized package of home safety improvements to decrease fall risk is ongoing.
- New York City families with affordable rent payments were found to increase their discretionary income by 77 percent, freeing up funds to spend on health insurance, food, and education or to save for a future down payment on a home.

Commented [1]: Add references for all of these points

Definition of the Metrics

The business problem when converted to data science problem is translated as 'Build a model which will return a binary prediction on whether a particular member is under housing insecurity or not. To obtain the most accurate prediction, we will build multiple models with different parameters and features. The best model will be selected based on the highest Area Under the Curve for the Receiver Operator Characteristic curve (AUC-ROC), among all model candidates, which is understood as a proxy of model prediction accuracy and performance.

Humana is committed to developing and fostering more equitable and inclusive AI for protected classes: race, sex, age, low income status, and disability status. This is measured within our model using a disparity score which is measured using race and sex.

Data Preparation

Dataset Overview

The dataset provided for this analysis consisted of two files containing 880 variables (including "*ID*" column) and one binary target flag variable. The dataset consists of several metrics sourced from external agencies and Humana's internal data collection platforms. It consists of individual level granular information for attributes such as demographic, credit information, medical claims, Clinical condition, etc. The target variable "*hi_flag*" is a binary metric which denotes the housing insecurity mapped to a unique person identifier "*id*".

- **Training data**: *48300* records & 880 variable columns, with a target variable (*hi_flag*)
- **Holdout Data**: *12220* records & 880 variable columns

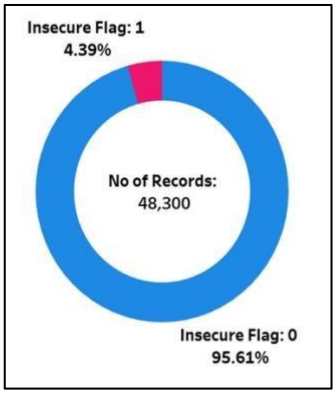
Exploratory Data Analysis

Our initial assessment reveals the dataset to be classified under the following 7 groupings:

- Claims Data
- Credit Data
- CMS Data (Under Department of Health and Human Services)
- Consensus Data
- Interaction\ Communication Data
- Prescription Data
- Other

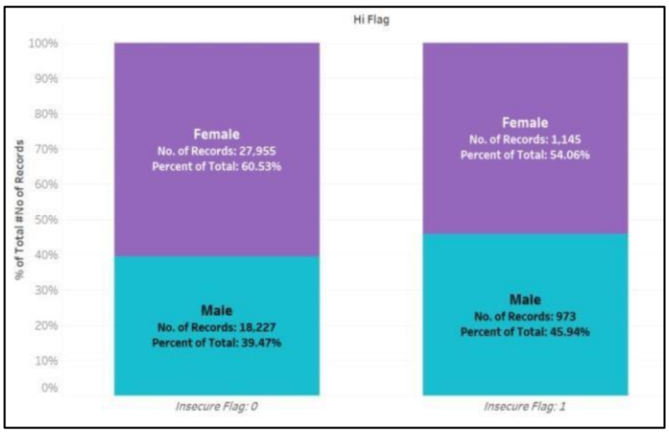
Before we began our modeling approach, we conducted an exploratory data analysis to obtain a high level sense of the data. Through our research we found the following :

1. Target Variable is highly imbalanced



After observing the pie chart on the left we can evidently see that we have an imbalanced dataset where flag variable " *1*" (Insecure) accounts for 4.39% of the overall data. This shows underrepresentation of members that are facing housing insecurity. Most machine learning models don't work well to identify the minority class, hence we must apply data preprocessing steps to handle this class imbalance problem.

Figure 1: Pie Chart for Target Variable "hi_flag"



that we have an

Figure 2: Bar Plot of Sex vs hi_flag indicator

2. Observed Sex Disparity in "hi_flag" = 0

The training dataset is skewed for target variable 0 as there seems to be a lot more records for females than male in people who might not be at the risk of housing insecurity.

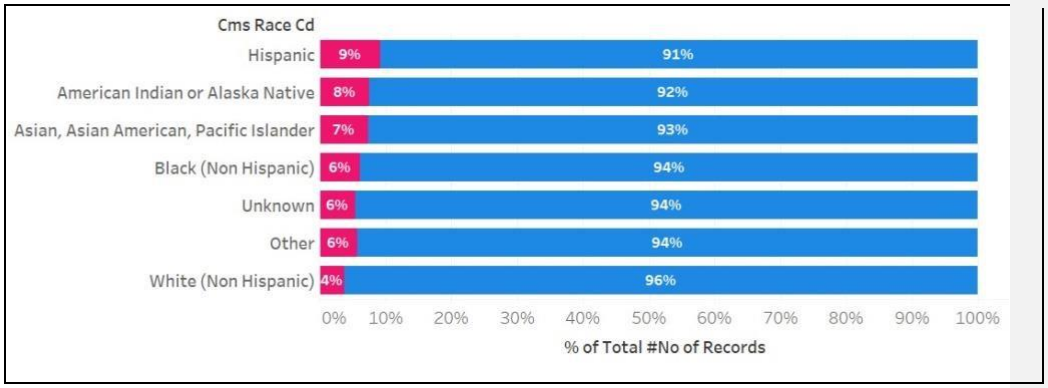


Figure 3: Bar Plot of Race vs hi_flag indicator

3. Observed disparity in Race Variable

The chart above reveals the fact that "Hispanic" individuals are at higher risk of facing housing insecurity than all the other classes. However, the sample for training data is highly skewed where the number of records for classes such as "*White (non hispanic)" & "Black (non hispanic)"* combined dwarfs the representation of all other classes.

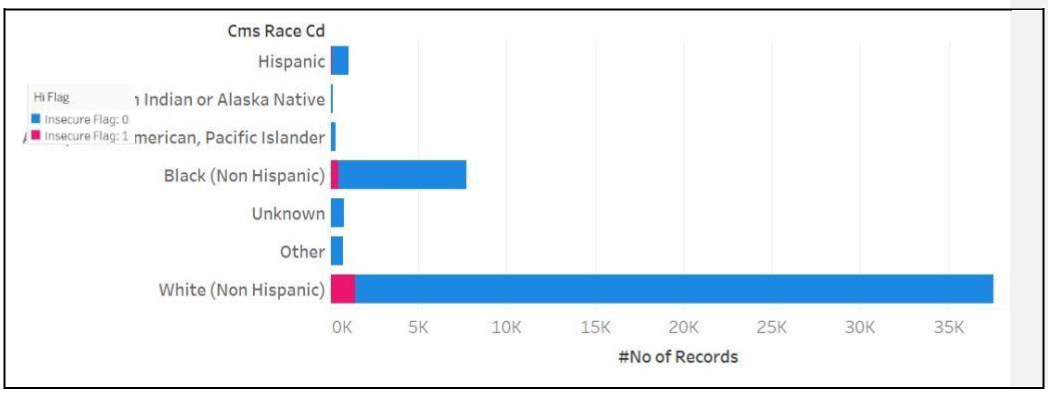


Figure 4: Bar Plot of Number of Records by Race

4. Age does have a marginal relationship with housing insecurity

Individuals that are facing housing insecurity have a median age of 69 years as compared to 72 years for people who are not facing housing insecurity. Though the data might not represent the actual population we can assume that people with higher age tend to be financially stable than the younger population and are thus at less risk of facing housing insecurity.

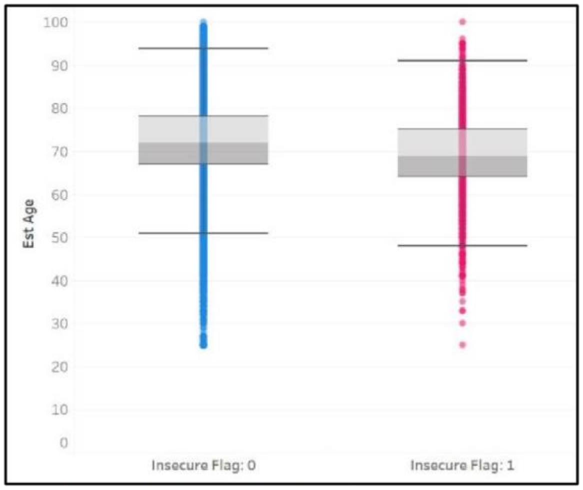


Figure 5: Box plot of Age Vs hi_flag indicator

5. Renter are at higher risk of facing housing insecurity

Plotting the percentage of individuals by the homeowner status reveals the fact that people who are " *Probable Renter" or* identified *"Renters"* face a higher risk of housing insecurity.

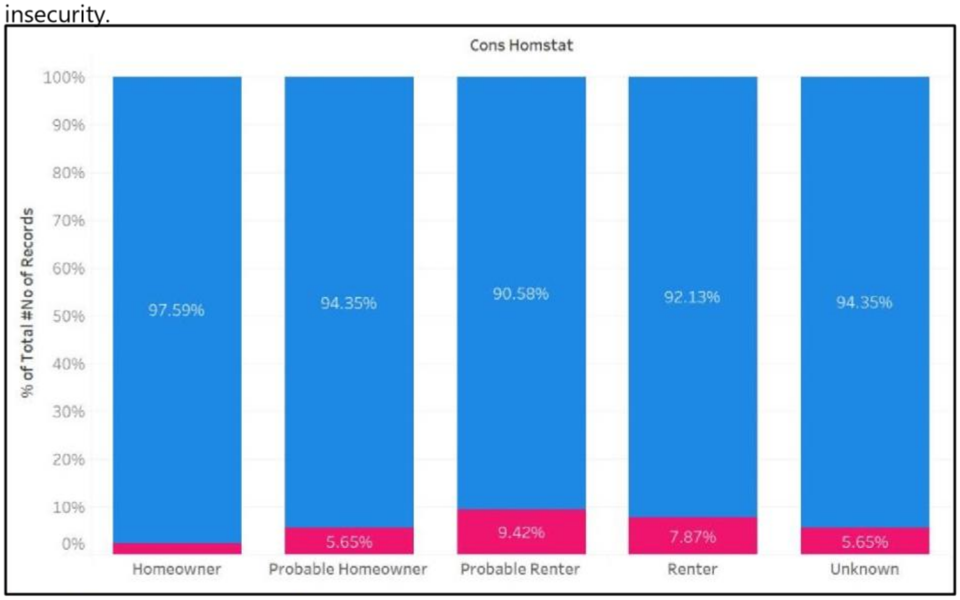


Figure 6: Homeowner status by hi_flag indicator

Data Preprocessing Steps

1. Deleting Zero-Variance Predictors

To reduce the high dimensional data, we needed to employ several techniques to conduct feature elimination. We began our process by removing columns that possessed zero variance and would not help in creating a model that can identify the target variable.
Using this technique, we were able to eliminate 119 columns from the dataset.

2. Dropping Column with missing percent greater than 80%

Our second step included finding the candidate columns for feature elimination based on percentage of missing data. Using this step, we were able to eliminate 18 columns which had missing data greater than 70% of the total dataset.

3. Filling NA value in Categorical Column with "Unknown"

To preserve the information contained in the categorical columns we decided to replace the missing values as "Unknown". We also see an apparent relationship between the missing values of "cons_homstat" and "cons_mobplus" columns. The correlation value of 1 denotes that the presence of null value in one column is directly correlated with the presence of null values in another column.

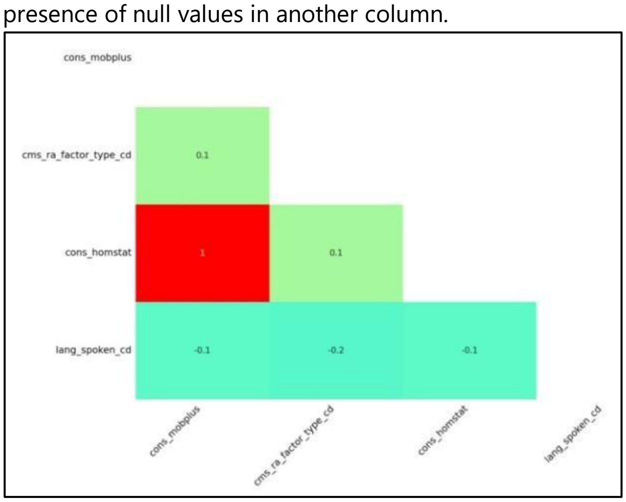


Figure 7: Heat Map Plot for Correlation of Missing Data in Categorical Variables

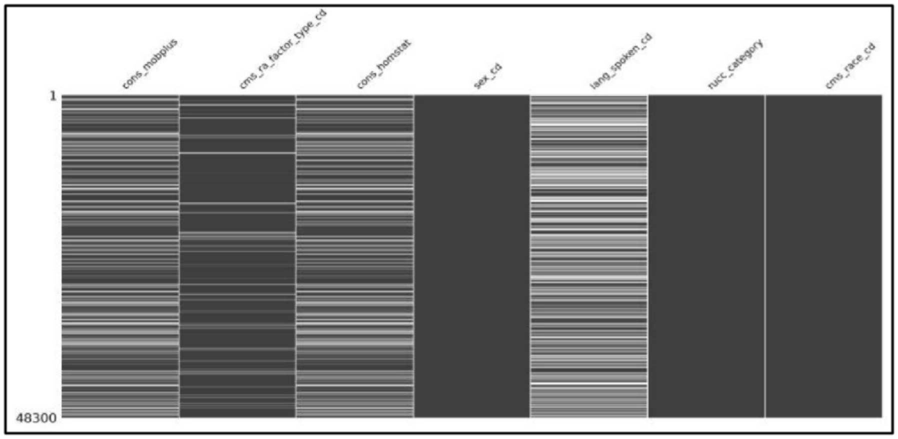


Figure 8: Matrix Plot for Missing Categorical Variables

4. Encoding Categorical variable

After filling the missing “NA” values with “Unknown” we encoded the categorical columns using Pandas Dummy encoder.

5. Running KNN Imputation to fill missing values

Through our analysis we found the number of missing data poi nts to be 314975 after the removal of columns mentioned in the steps above. Given the dataset now contained 775 (excluding flag variable) columns and 48300 rows we had the matrix size of 37,432,500 records out of which 314975 were missing (0.841%). After f actoring this number with a missing check across the rows of the dataset we found the dataset to be a worthy candidate for iterative or KNN imputation. We also imputed the values through mean imputation to perform an alternative check during model evaluation all while making sure we don't cause data leakage through our imputation technique in the test dataset.

6. Column Aggregation for Dimensionality Reduction

Our next step in feature engineering includes column aggregation across fields that hold information on cost and claims data. We ran a sum level scalar aggregation on groups mentioned below to the disease level granularity which eliminated a lot of features in the dataset. Our process approach also included the dataset which did not have column aggregation to serve as a comparison metric for testing the effect of column aggregation in regard to model performance.

a. cmsd2_inj_*, cmsd2_eye_*, cmsd2_ext_*, cmsd2_end_*, cmsd2_ear_*, etc.

b. rx_hum_*_ct, rx_hum_*_cost

c. rx_phar_cat_* , rev_pm_*

7. **Applying SMOTE to balance the underrepresented class variable**

As our target variable is a minority class with an overall representation of 4% we need a robust recognition system or a better preprocessing pipeline to prevent the marginalization of individual members facing housing insecurity. Under the data level approach we performed smote based oversampling on the training dataset to ensure the train model does not overfit and result in poor result on test data. Our second approach to tackle the associated effects of imbalance in training data includes the application of ensemble based models to improve the performance of single classifiers.

Modeling Approach

Given the extensive list of dimensions in the dataset, we simply couldn't rely on traditional statistical tools for modeling to map those complex multiplicative interactions between variables. We understood the need to employ tree based models which implement a system of "if-else" statements that tend to provide better results over high non-linearity & complex relationships between dependent & independent variables.

Model Tuning & Selection

Performing a grid search across a wide range of parameters with five-fold cross validation, we found the best parameters for each model with their respective AUC score. After evaluating each model we selected LightGBM as it performs well on classification problems with its ensemble tree architecture.

Note: We also tried out a slew of stacked ensemble models, but the results did not outperform the single tuned LightGBM model.

Table 1.1 Model Details

Model	Tuned Hyperparameters	Average AUC Score
Logistic Regression	max_iter = 1000 random state = 7 test size = 0.2	0.7165
Light Gradient Boost Classifier	Learning_rate = 0.05, Max_depth = 5, N_estimators = 1800, Min_split_gain = 0.7, Random_state = 12, Subsample = 0.8, Min_sample_leaf = 100	**0.7324**
XGBoost	learning_rate =0.01, n_estimators=5000, max_depth=4, min_child_weight=6, gamma=0, subsample=0.8, colsample_bytree=0.8, reg_alpha=0.005,	0.7003
Random Forest	max_depth = 80, max_features = 3, min_samples_leaf = 5, min_samples_split = 12, n_estimators = 1000	0.6872

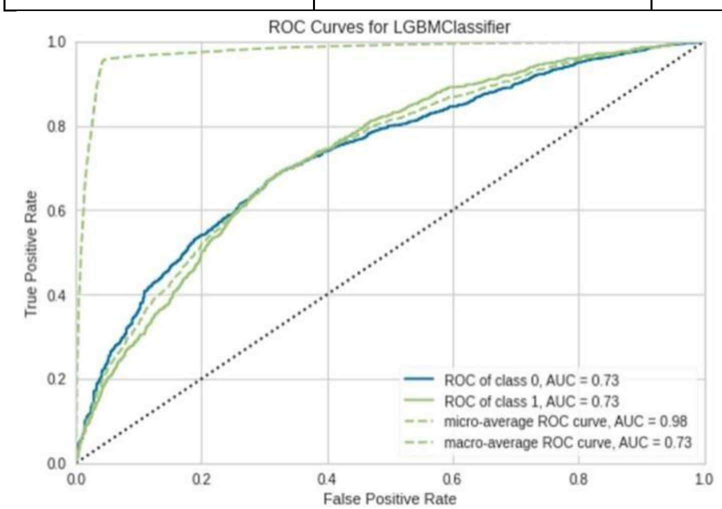


Figure 9: ROC Curve for Light Gradient Boosting Machine

Feature Importance/Key Performance Indicator Analysis

While the models can help us gauge the ability of identifying the individuals that are at the risk of facing housing insecurity, using them alone cannot help us evaluate the ways to prevent the problem. To further our approach and derive some tangible insight we worked to understand the commonalities in the traits of people that might be facing housing risk and created a customized implementation plan to target those segmented groups of individuals.

To lay out those understanding we began by evaluating the best features that helped the model's ability to identify the individual's risk of facing housing insecurity. The added benefit of using ensemble-based models is that they automatically provide estimates for feature importance from a trained predictive model. The feature importance chart below showcases the important variables that are a worthy candidate for analyst gathering information as they directly affect the model's ability to predict those individuals.

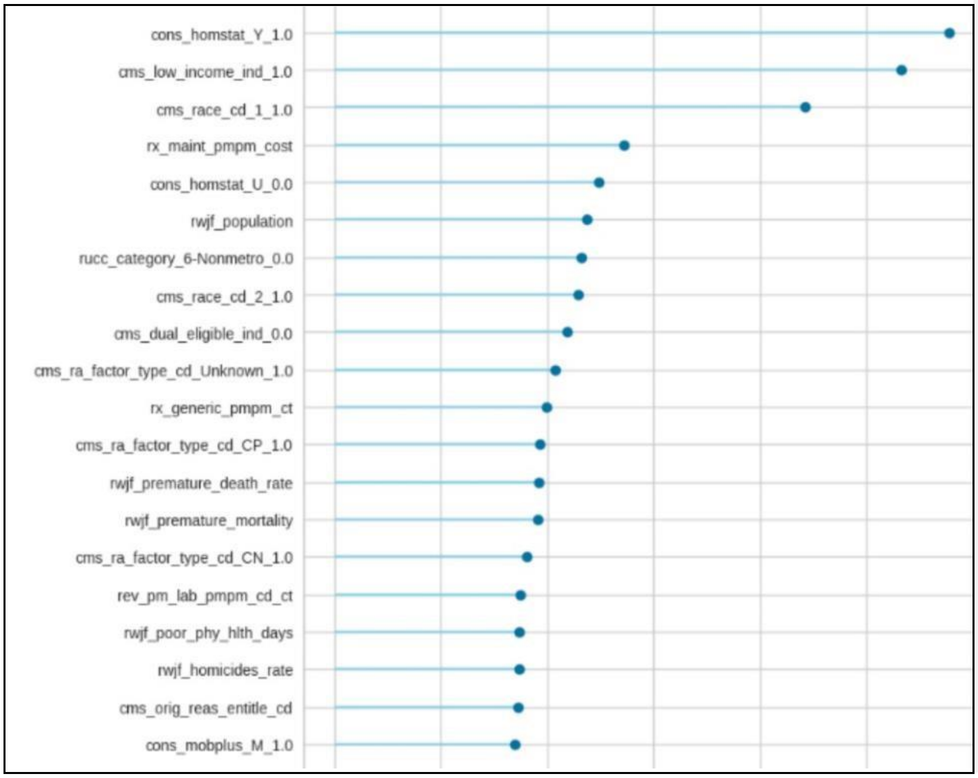

cons_homstat_Y_1.0
cms_low_income_ind_1.0
cms_race_cd_1_1.0
rx_maint_pmpm_cost
cons_homstat_U_0.0
rwjf_population
rucc_category_6-Nonmetro_0.0
cms_race_cd_2_1.0
cms_dual_eligible_ind_0.0
cms_ra_factor_type_cd_Unknown_1.0
rx_generic_pmpm_ct
cms_ra_factor_type_cd_CP_1.0
rwjf_premature_death_rate
rwjf_premature_mortality
cms_ra_factor_type_cd_CN_1.0
rev_pm_lab_pmpm_cd_ct
rwjf_poor_phy_hlth_days
rwjf_homicides_rate
cms_orig_reas_entitle_cd
cons_mobplus_M_1.0

Figure 10: Feature Importance of Top 20 Variables for LightGB model

Cons_homstat_y: Homeowner status is the most important predictor in the model as it contains information on the accommodation status of the member. Overall people or individuals who have a status of "Probable Renter" or "Renter" have a 114% higher risk of facing housing insecurity than "Homeowner" or "Probable Homeowner".

Cms_low_income_ind_1.0: The rising cost of housing relative to income makes this variable extremely important in predicting the risk of facing housing insecurity. Lower income often makes it diffi cult for individuals to pay the rent which entails them to cut down expenses on necessary medication that might keep them healthy and safe. A study presented by Joint Center for Housing Studies of Harvard University [JCHS] 2020, showcases that one in every four renters is severely cost burdened, spending more than half of their household income on housing cost. This combined with the severe shortage of affordable housing gives us the by-product that is housing instability, often represented by frequent and involuntary residential moves.

Cms_race_cd_1: The model corroborates the fact that race does dictate the likelihood of an individual facing housing insecurity as people of color are more likely than white households to be extremely low-income renters and disproportionately struggle to pay rent compared to white households. Lower wages along with discrimination prevents these classes from ensuring a good and safe housing environment.

Census Bureau's Household Pulse Survey: For example, in April 2021, nearly two-thirds of Black adults and seven in ten Hispanic adults (64% and 70%, respectively) reported difficulty paying household expenditures compared to 42% of White adults; 7% of Black adults and 12% of Hispanic adults reported no confidence in their ability to make next month's housing payment compared to 4% of White adults, and 14% of Black adults and 16% of Hispanic adults reported food insufficiency in the household compared to 5% of White adults.

This growing disparity in an individual's ability to pay rent exposes them to the risk of instability, eviction and even homelessness which researchers link to an array of negative consequences such as food insecurity, poor health, lower academic achievement and low economic mobility.

Rx_maint_pmpm_cost: The cost per month for prescriptions related to maintenance drugs does find its place among the important variables present in the model. The data suggest a relationship between an individual risk of facing housing and the cost for maintenance drugs where the risk factor increases as the cost of prescription drugs increase.

Rwjf: The model also finds considerable information in variables presented by Robert Wood Johnson Foundation as they do track structural barriers to health.

Rucc_category_6-nonmetro: A lack of decent affordable housing underlies both rural and urban areas. Though the housing costs tend to be lower in rural areas, so are rural incomes, leading to similarly high rent burdens. This shows that member geographic information does dictate the likelihood of facing housing insecurity.

In addition to the feature ranking system based on estimators we also used Shapely Additive Explanations (SHAP) plots to interpret the effect of individual features on the probability of "hi_flag" variable.

FUTURE STEPS - ANALYSIS AND ACTIONABLE INSIGHTS

SHAP (Shapely Additive graphs) help us understand how a model makes its decisions. It quantifies how important each input variable is to a model for making predictions. This can be a useful sanity check that the model is behaving in a reasonable way, and that the features used by the model are reasonable in explaining our output.

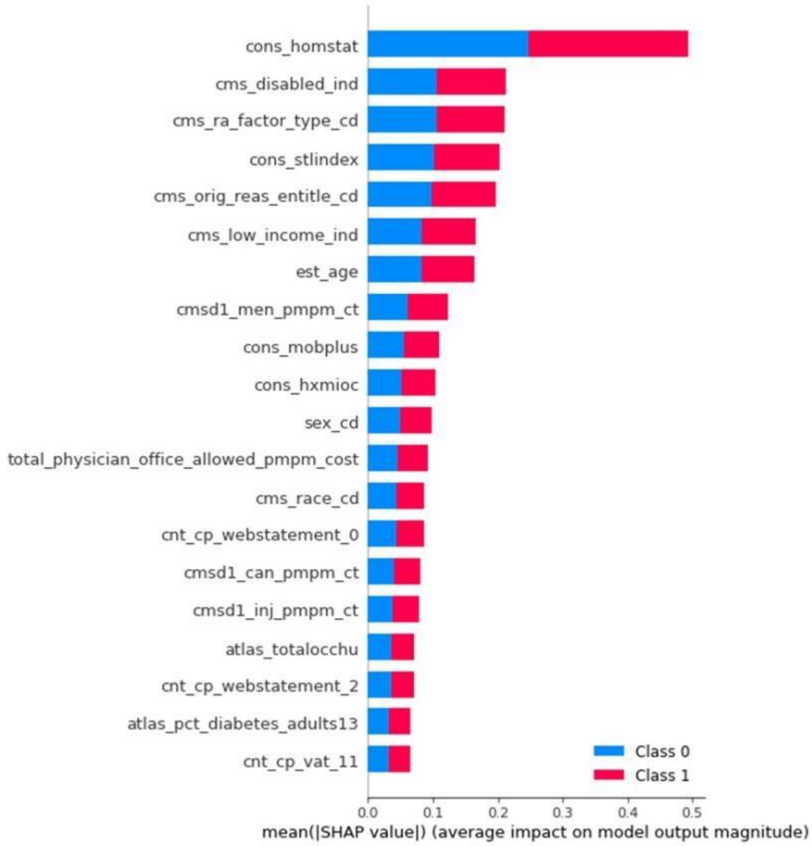


Figure 11: SHAP plot for features used by model for decisionmaking.

The top 22 Variables that were selected as important by our LightGBM Model were further fed into SHAP for model interpretability. This helped us understand the magnitude of influence of features in determining the presence of housing insecurity. So me of these plots are as follows:

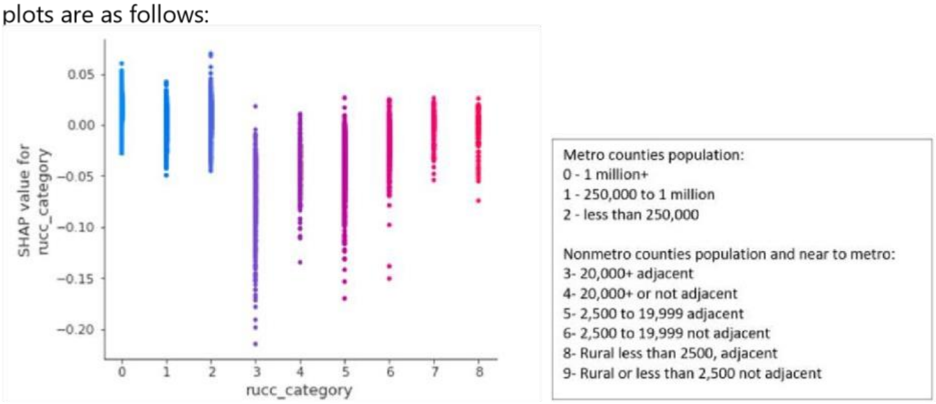


Figure 12: Individual analysis for RUCC Category

RUCC categories show that the likelihood of undergoing housing insecurity is higher for populations residing in Rural areas.

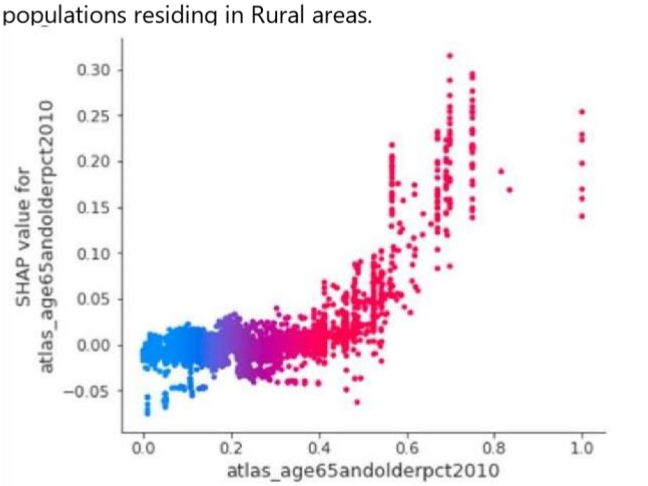


Figure 13: Individual analysis for Rate of change of people over 65 years of age

With the increase in the rate of change of people over 65 years of age we see increase in the SHAP values indicating the effect of increase in Housing insecurity.

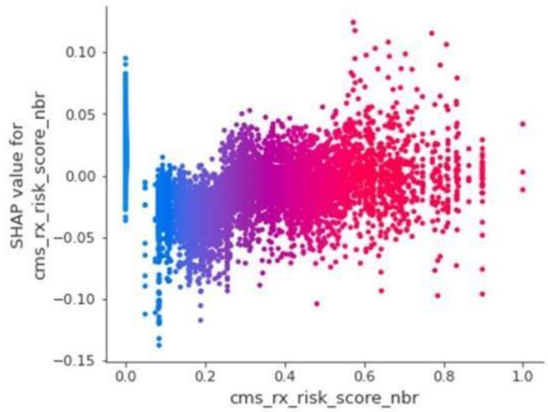


Figure 14: Individual analysis for cms_risk_score_nbr

The SHAP values for cms_risk_score_nbr indicate that housing insecurity increases with increase in the risk score.

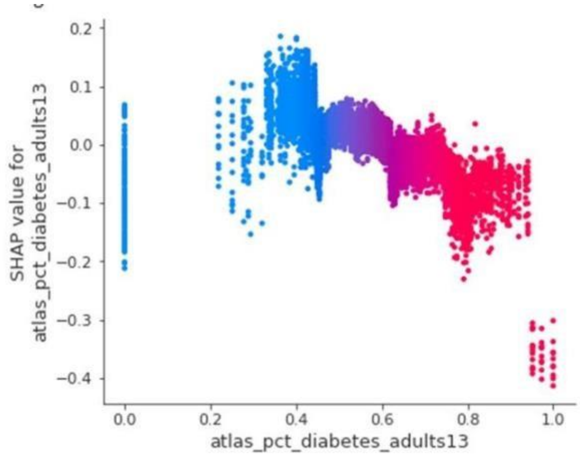
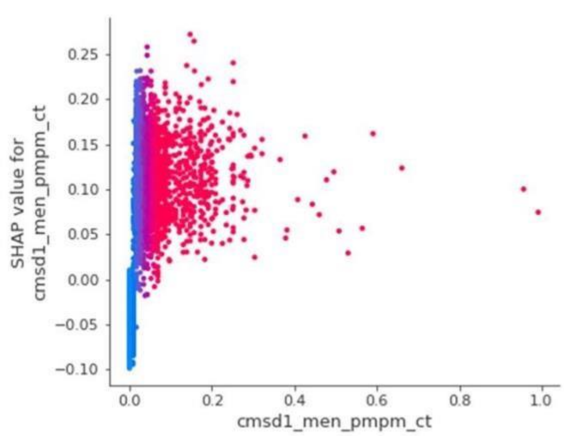


Figure 15: Individual analysis for atlas_pct_diabetes_adults13

As seen the SHAP values for percentage of diabetes in adults indicate linkage to increased risk of housing insecurity.



Figure

16: Individual analysis for cmsd1_men_pmpm_ct

The spread of SHAP values for the count of claims related to mental health and neurological shows that with a higher number of claims per month we see increased risk of Housing insecurity.

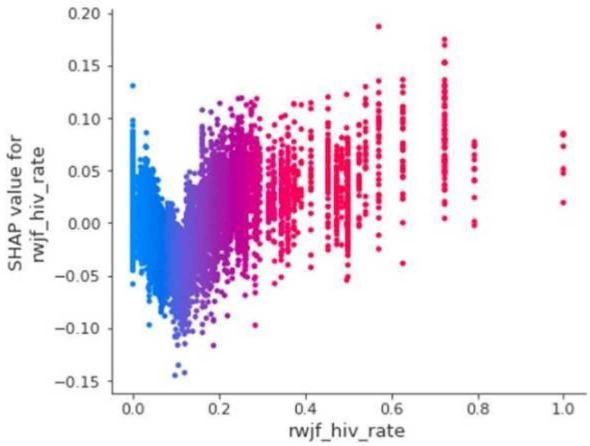


Figure 17: Individual analysis for rwjf_hiv_rate

The SHAP values for HIV Rate in a county indicate that housing insecurity is influenced, and an individual is likely to be housing insecure in a neighborhood characterized by this feature.

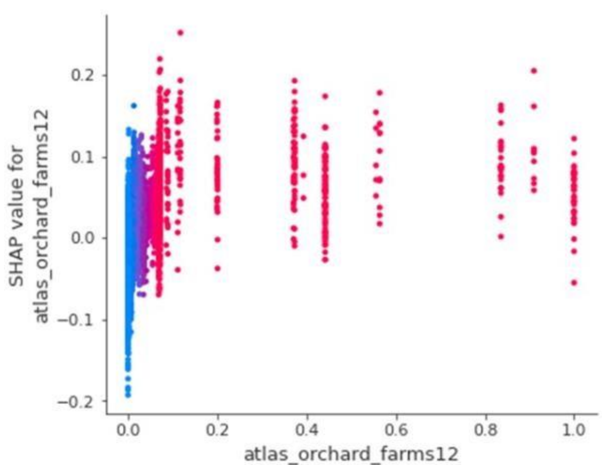


Figure 18: Individual analysis for atlas_orchard_farms

The variable atlas_orchard_farm indicates the number of orchard farms in a given county. A greater number of orchard farms might be an indicator of a healthy and engaged community as often these orchard farms are maintained and taken care of by the communities around it. The SHAPly plot here seems to be counterintuitive as it shows that with a greater number of orchard farms the presence of housing insecurity is also high.

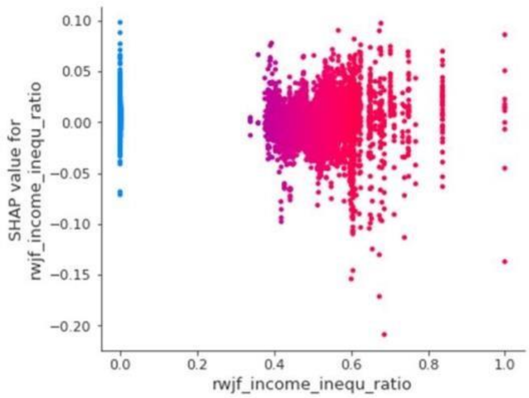


Figure 19: Individual analysis for rwjf_income_inequ_ratio

We see that with the increase in the inequality ratio in a country the chances of being housing vulnerable increases.

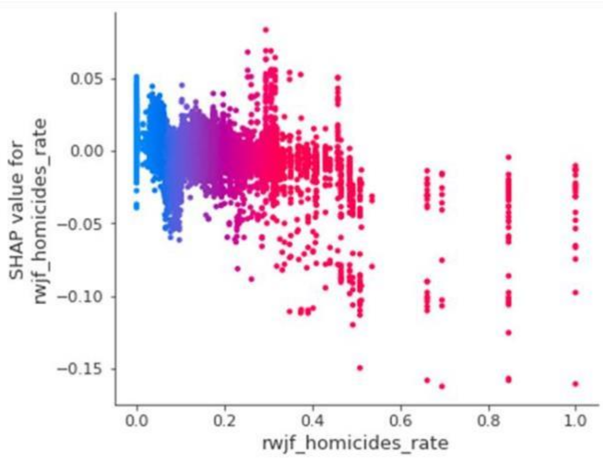


Figure 20: Individual analysis for rwjf_homicides_rate

SHAP values for Homicide rates in a county indicate that housing insecurity is affected, and an individual is likely to be housing insecure in a neighborhood characterized by this feature.

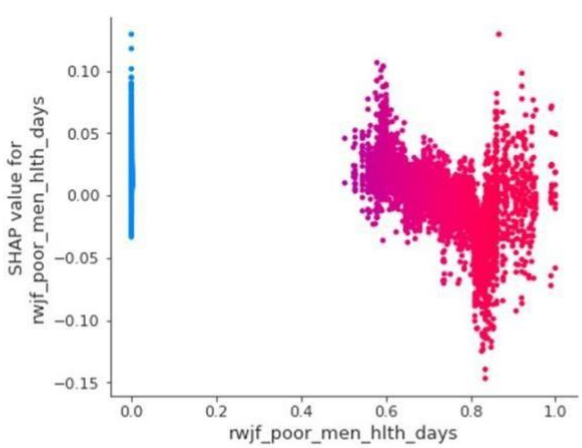
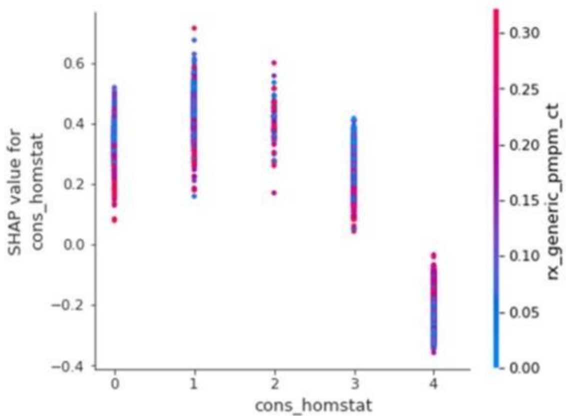


Figure 21: Individual analysis for rwjf_poor_men_hlth_days

to



Counties that have more number of mentally unhealthy days reported in past 30 days show that residents in these counties are liklier be housing insecure.

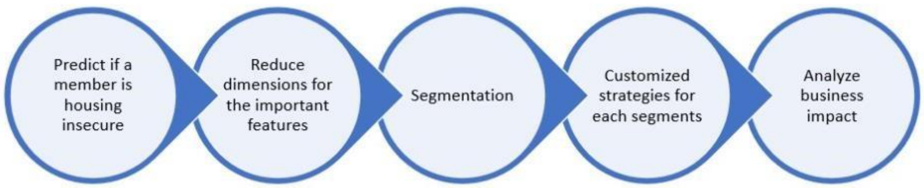
0 - Probable Homeowner
1- Renter
2- Probable Renter
3- Unknown
4 - Homeowner

Figure 22: Interaction analysis for cons_homstat vs rx_generic_pmpm_ct

The higher SHAP values for the homeowner status Renter and probable Renters validates our intuitive understanding that being a homeowner rather than a renter indicates housing stability. Being a probable renter acts as an indicator for housing insecurity and shows an increased count of prescriptions of generic prescription drugs which are cheaper than their branded counterparts.

Proposed Solutions

We propose differentiated strategies for member sub-segments to prevent and mitigate housing instability amongst the members. Our post modeling analysis process has been described in the following flow chart.



Key factors contributing to Housing Insecurity

To identify the key factors contributing to housing insecurity, we picked the top 22 important variables from our model and business understanding, and further conducted Principal Component Analysis (PCA) to categorize them into 4 factors (Refer Appendix B) with 80.2% variance explained.

Factor 1: Demographic factors

The features contributing to this factor includes: Percent of population 65 or older and Natural population change(2010-2016). According to the previous SHAP analysis, the increase in the rate of change of people over 65 years of age would indicate an increase in Housing insecurity. This indicates that a member with a higher score on this component is likely at higher risk of housing instability.

Factor 2: Health Related Factors

The clinic visits and claims per month for different kinds of diseases (neoplasms, mental and behavioral disorders, injury and poisoning) and the risk score based on pharmacy claims mainly contributes to the second component. The SHAP values for the risk score are positively related to housing insecurity which indicates that members with higher risk scores are at higher risk of housing insecurity. A member with a higher score on this component is generally in a better health condition.

Factor 3: Financial Stability Factors

This factor consists of homeowner status and income inequality ratio (ratio of household income at the 80th percentile to income at the 20th percentile). According to the SHAP analysis, members owning a house are at a lower risk of housing instability compared to renting a house. A higher score on this component indicates a member has better financial stability.

Factor 4: Neighborhood Factors

Rural Urban geographic information (RUCC Category) and the neighborhood factors (HIV rate, mortality rate, violent crime rate, homicides rate etc) mostly contribute to this factor. According to the SHAP analysis, the presence of orchard firms don't clearly indicate a higher risk of housing instability but counties with a higher HIV rate are at a higher risk of housing instability

Segmentation and Reasoning

Values	Group 1	Group 2	Group 3
Financial Stability Factors	-4.73	1.71	0.39
Demographic Factors	-0.26	-0.68	1.78
Health Related Factors	0.13	0.1	-0.35
Neighborhood Factors	-0.56	-0.05	0.61
Members	459 (22%)	1150 (54%)	509(24%)
Risk Score based on pharmacy claims *	0.26	0.31	0.28

Reason for entry into medicare **	Old Age Survivors Insurance (11%), Disabled (10%)	Old Age Survivors Insurance (33%), Disabled (21%)	Old Age Survivors Insurance (13%), Disabled (11%)

After identifying key factors contributing to housing insecurity, we are identifying different groups in humana's members that are at the risk of housing insecurity. This approach would help us provide customized and efficient solutions for the different groups at risk of housing insecurity. We found members falling into primarily 3 main clusters with the highest Calinski-Harabasz score at 2,068. This is an indicator that the segments are well separated into non-overlapping sub-groups.

The key features of every group are provided in the following table. The top 4 factors are components extracted from the Principal Component Analysis while the others are general profiles of clusters.

Table 1.2 Cluster Groups

* Identified by the “cms_rx_risk_score_nbr” feature

**Identified by the "cms_orig_reas_entitle_cd" feature. Rest 2 categories had negligible amount of members.

First Segment: Financially unstable
The first group mainly shows financial instability as the major contributing factor. This severe sign of financial constraints is depicted by their low ratio of 80 percentile household income to a 20 percentile household income. The percentage of Humana Members from our data that are impacted is 22%.

Second Segment: Seniors with higher risk score
While financially better than the other groups, this group has a higher spread of older members with health factors possibly linked to diabetes which is known to affect one's' quality of living. This was further validated when we profiled our clusters and found that Segment 2 had 33% members that had OASI (Old Age Survivor Insurance) as the reason of entry when compared to Segments 1 and 3 at 11% and 13% each.

Third Segment: Unhealthy Neighborhood
The second biggest contributor to this group are the neighborhood factors and also health factors that are indicated by higher number of physician visits and greater number of diagnoses based on external injuries.

Business Insights: Value Benefit Proposition

When Healthcare Insurance providers like Humana look to identify their members who are most likely to be going through Housing Insecurity it saves up on resources and time of Homelessness prevention organizations across the region that process several thousands Housing assistance program applications out of which only a fraction of the applicants can be helped due to the limited resources these programs have.
Early identification of Housing Instability can give time to implement different steps of intervention as proposed later.
Consequences of Housing Insecurity on Health:

- Mental and Behavioral Health:
  One of the important factors in determining or influencing Housing Insecurity as identified by our model was the count of claims made related to mental, behavioral and neurodevelopmental disorders. Average pay-out for Humana as an Insurance provider would be around

- Increase in Diabetes, asthma, cardio-vascular diseases due to stress and environment triggers.[9] People with diagnosed diabetes incur average medical expenditures of $16,752 per year. People with diagnosed asthma incur average medical cost of $3,266 per year. People with diagnosed cardio-vascular disease incur average medical cost of $18,953 per year.

- In acute cases of Housing insecurity, or when failing to intervene before this issue it can result in homelessness. While this has dire health and safety consequences for people, the lack of a physical address also makes it difficult for health providers and community workers to reach, communicate and/or help them. They are less likely to be able to access medicines that they may need which can result in cycles of sickness leading upto expensive Emergency rooms visits with severe health conditions. [3] Also, people who are homeless have higher rates of illness and die on average 12 years sooner than the general U.S. population.[9]

Average cost of Emergency visits owed to Medicare as stated by HEALTHCARE COST & UTILIZATION PROJECT[5] was around $23 billion in 2017.
Market share of Humana is 7 percent as of 2022. Extrapolating from this we can say that percent of shared expenses for medicare by Humana may amount to 7% of 23 billion dollars = 1.61 billion dollars.
Out of these the ones that may be a result of exacerbated health issues due to homelessness is estimated using out percentage of Humana members that are Housing insecure. The percentage of Housing insecure vulnerable is 4.3% from our Training data. The group that is most vulnerable and can have homelessness as a dire consequence is Group 3 at 22% of members.
Cost of Emergency Department visits as a result of Homelessness to Humana = 22% * 4.4% * 1.61 billion = $15.23 million.
Another source of estimated costs shows that the average cost of expenses incurred as a result of Emergency Visits when homeless is $3,700. On average people struggling with Homelessness visit ERs 5 times a year. This amounts to $18,500. [6]
The number of people in our most vulnerable group is 509 members. This amounts to 509* 18500 = $9.4165 million.
Above two estimates give us a range for potential preventable costs ranging from $9 to $15 million.

Cost Benefit Calculation:

- Number of records with "hi_flag" = 1 in the dataset: 2118
- Humana Medicare Advantage Plan (Part B Premium) : $170 (Approx Monthly)
- Total Opportunity: 170 x 12 x 2118 = $4,320,720 (Per Year)

Overarching solutions:

1. Teaming up with safety engineers and hygiene inspection workers to detect mold, leakages, air conditioning and heater issues as well as water filters.
2. Assigning care managers or mental health providers to certain households based on health or functional need criteria. Peer support can help coordinate care and social support services, facilitating linkages to housing, transportation, employment, nutrition services, and other community-based supports.
3. Providers can aid the "Health Home State Plan" established by ACA to coordinate care for people with chronic diseases. These health home care services include

comprehensive care management, care coordination, health promotion, comprehensive transitional care, patient and family support, as well as referrals to community and social support services (such as housing, transportation, employment, or nutritional services).

4. Funding or providing partial aid to eviction prevention programs.
   a. Individuals under “Renter” & “Probable Renter” Status (Train Data): 208
   b. Percentage of people at risk of facing eviction[7]: 6%
   c. No of individuals under eviction risk: 6% x 208 = 12 (Approx)
   d. Dollar Contribution of those members: 12 (member) x 12 (months) x 170 (ppm) = $24480
   e. Average cost to clear back rent and avert an eviction for a household in the Eviction Diversion Program per individual : $1,067 (per person)
   f. Total cost to prevent eviction: 12 (members) x 1067 = $12804
   g. Total Cost Savings for members who might be at risk of facing Housing Insecurity: d - f = 24480 - 12804 = **$11676**

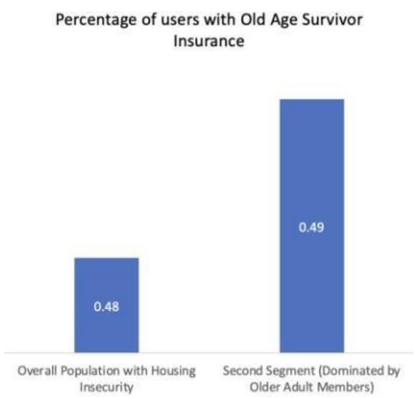
Long-Term Solutions:

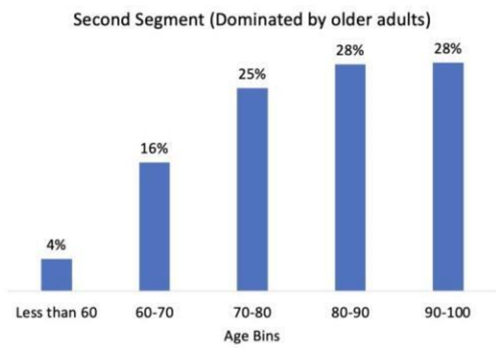
1. Influx of volunteering and Community health activities.
2. Investing in home developers that provide affordable housing and drafting a working relation with the selected developer by incorporating health insurances for its employees and their family members at lower premiums.

Conclusion

Housing for America’s Older Adults : Making housing affordable, Safe and Quality Oriented

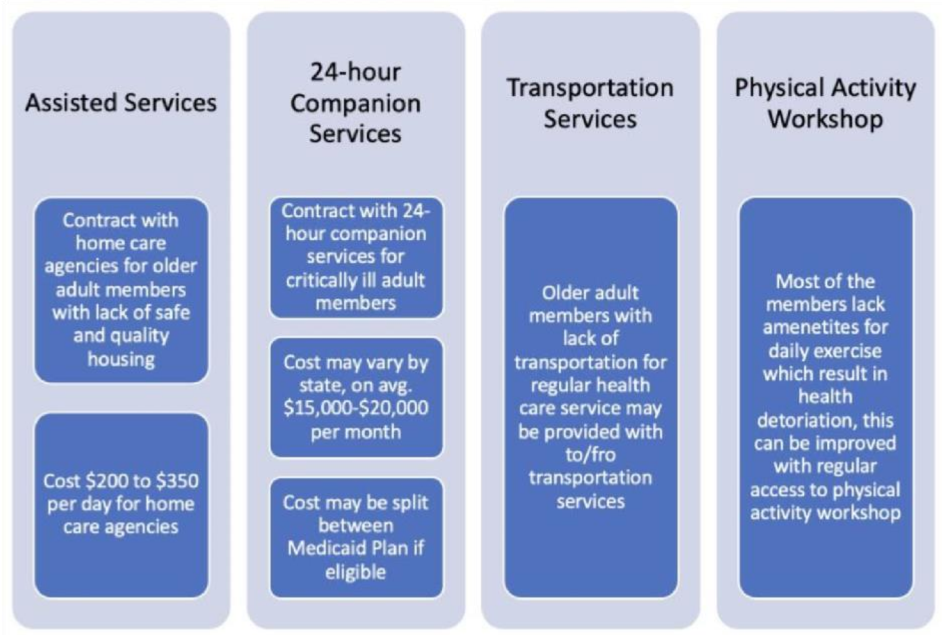
Literature review suggests older adults in future may need to live alone and on limited incomes, and many will have mobility and other health challenges. By 2035, the Census Bureau projects that the **population 80 and over will grow to nearly 24 million people, fully doubling from 2016.** Looking at the statistics, there is a huge need in the future for affordable, accessible housing, in-home services, and neighborhood support and amenities. In addition, other statistics say that there are over **10 million households headed by someone 65 and over who are cost burdened** (paying more than a third of their income on housing); half of these pay more than 50 percent. To compensate, households often cut back on food and medical care, which can be detrimental for those with chronic health conditions.





In our training data, **for the Second Segment** we can see in our housing insecurity population, the average age is higher compared to people without the non-housing insecurity. These older adults may need homes which offer a <u>no-step entry, single-floor living,elevators in multi-storey buildings and wide enough doors and hallways to accommodate a wheelchair</u>. Currently, in the US these types of homes are less than 4% percent. So, there is a huge need to make modifications and maintain housing in safe condition. <u>This is something that is plausible only with the help of policymakers</u>. And they need to investigate it because research suggests service-enriched affordable housing has been shown to support independence—and reduce healthcare costs—but need outstrips supply. Demand will grow for supports and services delivered to middle-income older adults who typically cannot afford assisted living settings.

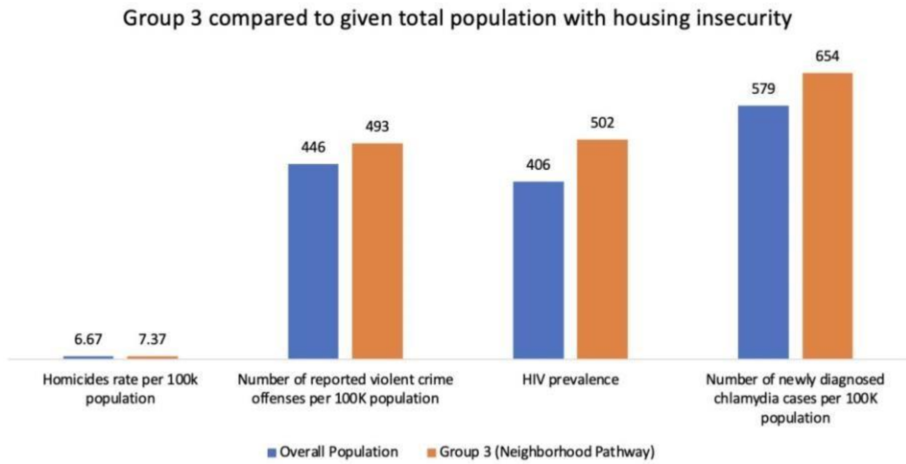
We also have seen literature which suggest older adults who live in places that lack livability features, such as neighborhood services, transportation alternatives, safe streets, and opportunities for engagement are at higher risk of health issues. These all contribute to wellbeing, and can even combat isolation and loneliness, both serious health issues.



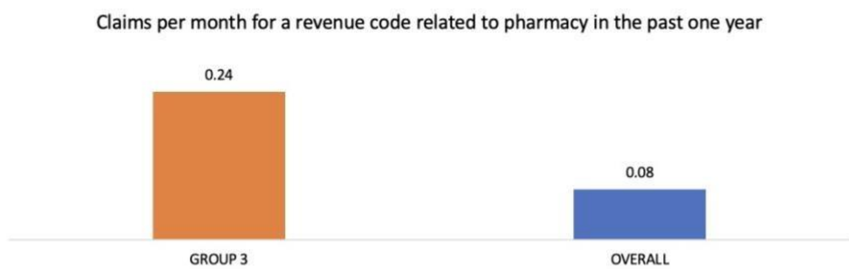
Relief to Housing Insecurity Caused due to Neighborhood Pathway

In high -poverty, urban communities with a predominantly racial and ethnic minority population, there is evidence to showcase consistent relationship between neighborhood crime and poor access to health-enabling resources. People reporting poor neighborhood safety were less likely to have large grocers, pharmacies, and fit ness resources within 1 mile from home. Even among those who did have resources within 1 mile from home, a prior experience of neighborhood crime was associated with bypassing pharmacies. These findings provide evidence to suggest that neighborhood crime m ay impede access to health-enabling resources in two important ways. First, neighborhood crime may reduce the total number of available resources (i.e., potential access), perhaps because health - enabling businesses avoid locating to high -crime neighborhoods. Second, neighborhood crime may impede utilization of available resources (i.e., realized access). Notably, we found that experiences of neighborhood crime were consistently associated with poor access to pharmacies. In addition to having fewer pharmacies, participants were also more likely to bypass nearby pharmacies if they had been a victim of theft or property crime.

Considering all this, we speculate that health prevention efforts, which have heavily emphasized the quantity of resources in each community, will also need to address the social barriers that impede access to and utilization of resources that support and sustain a healthy life.[43] **For the Third Segment,** the above scenario holds true. We notice the following

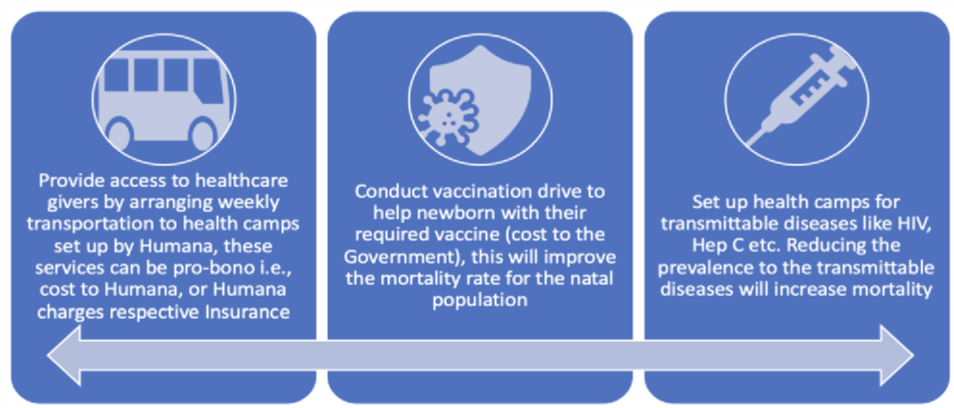


Due to these social and economic conditions in these neighborhoods, access to healthcare is severely compromised. This in turn impacts the whole healthcare system and the underlying financial institution supporting this. Some of the health issues arising here can be stopped or investigated at the initial phase thus reducing the future burden to healthcare. This can be seen in the below graph.



As you can see this group has higher claims per month compared to other members with housing insecurity. Thus, this group not only adds up to the future cost, but also reduces the overall healthcare index of the locality. The difference in the characteristics of these members to other members with housing insecurity lies with the impact of the neighborhood pathway.

This can be solved by various measures:



In conclusion to the neighborhood pathway intervention, we now have seen how having to live in unsafe or economically distressed neighborhoods is another potential form of housing insecurity. Living in high crime neighborhoods can have a negative impact on one's overall health or to the healthcare system. Conversely, moving to "opportunity neighborhoods" that have high-quality schools and access to transportation and employment has been associated with better educational, employment, and health outcomes for families.

Based on ACS measures for "distressed neighborhoods" where at least 40% of households in a census tract are below the poverty line, researchers have found the total impact in this category can be 11.2 million people who live in distressed neighborhoods. [44]

APPENDIX

A: Data Description and Variable Aggregation

Sr. No	Description/ Group	Prefix	No of Col	Type
1	Admit Per Month for Overall Claims	total_ip_*_admit_ct_pmpm	6	double
2	Admitted Days Per Month for Overall Claims	total_ip_*_admit_days_pmpm	6	double
3	Allowed Cost Per Month for Behavioral Health Claims	bh_*_*_allowed_pmpm_cost	12	double

4	Allowed Cost Per Month for Overall Claims	total_*_*_allowed_pmpm_cost	11	double
5	Credit Balance	credit_bal_*_*	4	double
6	Cms Indicator	cms_*_ind	7	int
7	Claim Lines Per Month for Revenue Code	rev_pm_*_pmpm_cd_ct	73	float
8	Claims Per Month Related to Certain Conditions Originating In The Perinatal Period (Immeditely Before Or After Birth)	cmsd2_neo_*_pmpm_ct	12	float
9	Claims Per Month Related to Certain Infectious And Parasitic Diseases	cmsd2_inf_*_pmpm_ct	21	float
10	Claims Per Month Related to Congenital Malformations, Deformations And Chromosomal Abnormalities	cmsd2_ano_*_pmpm_ct	11	float
11	Claims Per Month Related to Diseases of The Blood And Blood-forming Organs And Certain Disorders Involving The Immune Mechanism	cmsd2_bld_*_pmpm_ct	7	float
12	Claims Per Month Related to Diseases of The Circulatory System	cmsd2_cir_*_pmpm_ct	10	float
13	Claims Per Month Related to Diseases of The Digestive System	cmsd2_dig_*_pmpm_ct	10	float
14	Claims Per Month Related to Diseases of The Ear And Mastoid Process	cmsd2_ear_*_pmpm_ct	5	float
15	Claims Per Month Related to Diseases of The Eye And Adnexa	cmsd2_eye_*_pmpm_ct	12	float

16	Claims Per Month Related to Diseases of The Genitourinary System	cmsd2_gus_*_pmpm_ct	11	float
17	Claims Per Month Related to Diseases of The Musculoskeletal System And Connective Tissue	cmsd2_mus_*_pmpm_ct	20	float
18	Claims Per Month Related to Diseases of The Nervous System	cmsd2_ner_*_pmpm_ct	11	float
19	Claims Per Month Related to Diseases of The Respiratory System	cmsd2_res_*_pmpm_ct	11	float
20	Claims Per Month Related to Diseases of The Skin And Subcutaneous Tissue	cmsd2_skn_*_pmpm_ct	9	float
21	Claims Per Month Related to Endocrine, Nutritional And Metabolic Diseases	cmsd2_end_*_pmpm_ct	10	float
22	Claims Per Month Related to External Causes of Morbidity	cmsd2_ext_*_pmpm_ct	30	float
23	Claims Per Month Related to Factors Influencing Health Status And Contact With Health Services	cmsd2_vco_*_pmpm_ct	14	float

24	Claims Per Month Related to Injury, Poisoning And Certain Other Consequences of External Causes	cmsd2_inj_*_pmpm_ct	22	float
25	Claims Per Month Related to Mental, Behavioral And Neurodevelopmental Disorders	cmsd2_men_*_pmpm_ct	11	float
26	Claims Per Month Related to Neoplasms	cmsd2_can_*_pmpm_ct	21	float

27	Claims Per Month Related to Pregnancy, Childbirth And The Puerperium	cmsd2_pre_*_pmpm_ct	9	float

28	Claims Per Month Related to Symptoms, Signs And Abnormal Clinical And Laboratory Findings, Not Elsewhere Classified	cmsd2_sns_*_pmpm_ct	14	float
29	Clinical Care	rwjf_*	4	float
30	Cost Per Month of Prescriptions {Based On Humana Drug Classification}	rx_hum_*_pmpm_cost	92	float
31	Count Per Month of Behavioral Health Claims	bh_*_pmpm_ct	40	float
32	Count Per Month of Claims {Based On Charlson Comorbidity Index Categories}	cci_*_pmpm_ct	19	float
33	Member Count Per Month of Interactions Via Emails	cnt_cp_emails_*	12	int
34	Member Count Per Month of Interactions Via Livecall	cnt_cp_livecall_*	12	int
35	Member Count Per Month of Interactions Via Print	cnt_cp_print_*	12	int

36	Count Per Month of Member Interactions Via Vat	cnt_cp_vat_*	12	int
37	Count Per Month of Member Interactions Via Webstate ment	cnt_cp_webstatement_*	12	int
38	Count Per Month of Prescriptions Purchased	rx_phar_cat_*_pmpm_ct	15	float
39	Count Per Month of Prescriptions	rx_hum_*_pmpm_ct	92	float

40	Count Per Month of Prescriptions Related to Tier	rx_tier_*_pmpm_ct	4	float

41	Days Since Last Claim for Non-Behavioral Health Claims	med_*_ds_clm	12	int
42	Days Since Last Claim for Overall Claims	total_*_ds_clm	12	int
43	Visits Per Month for Overall Claims	total_*_ct_pmpm	5	double
44	Population Variables	atlas_*	9	double
45	Claims Per Month Related to {Based On Cms Diagnosis Code Level1}	cmsd1_*_pmpm_ct	22	float
46	Count Per Month of Member Interactions Via {Platform} In The Past One Year	cnt_cp_*_pmpm_ct	5	float
47	Robert Wood Johnson Foundation	rwjf_*	17	float
48	Prescription Variables	rx_*	31	float
49	total Allowed Cost And Claims Per Month	total_*	5	double
50	Credit Variables	credit_*	6	double
51	Consensus Variables	cons_*	9	int
52	Other		22	int/cat

B: PCA Factors

Variables	Category
atlas_age65andolderpct2010	Demographic Factors
atlas_naturalchangerate1016	

cons_homstat	Financial Stability Factors
rwjf_income_inequ_ratio	
atlas_pct_diabetes_adults13	Health Related Factors
cms_rx_risk_score_nbr	
cmsd1_can_pmpm_ct	
cmsd1_men_pmpm_ct	
rx_tier_1_pmpm_ct	
total physician_visit_ct	
total outpatient visit	
cmsd1_inj_pmpm_ct	
atlas_orchard_farms12	Neighborhood
rucc_category	
rwjf_child_mortality	
rwjf_hiv_rate	
rwjf_homicides_rate	
rwjf_poor_men_hlth_days	
rwjf_poor_phy_hlth_days	
rwjf_std_infect_rate	

rwjf_teen_births_rate	
rwjf_violent_crime_rate	

7. REFERENCES

[1]Healthy People 2030, U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. From https://health.gov/healthypeople/objectives-and-data/social-determinants-health

[2]"Housing And Health: An Overview of The Literature," Health Affairs Health Policy Brief, June 7, 2018. DOI: 10.1377/hpb20180313.396577

[3]Adapted by the author from *Gibson et al. 2011*, *Sandel et al. 2018*, *Maqbool et al. 2015*, and *Braveman et al. 2011*.

[43] Tung EL, Boyd K, Lindau ST, Peek ME. Neighborhood crime and access to healthenabling resources in Chicago. Prev Med Rep. 2018 Jan 31;9:153-156. doi: 10.1016/j.pmedr.2018.01.017. PMID: 29527469; PMCID: PMC5840856.

[44] Kneebone, Elizabeth. 2014. "The Growth and Spread of Concentrated Poverty, 2000 to 2008–2012." Washington, DC: Brookings Institution.

1. 2. Johnson A, Meckstroth A. Ancillary services to support welfare to work. Washington, DC: US Dept of Health and Human Services; June 22, 1998:20–23. Available at: http://aspe.hhs.gov/hsp/isp/ancillary/front.htm. Accessed April 28, 2010 [Google Scholar]
2. Swope, C. B., & Hernández, D. (2019). Housing as a determinant of health equity: A conceptual model. *Social Science & Medicine, 243*, 112571. https://doi.org/10.1016/j.socscimed.2019.112571
3. Humana Press News: Humana-Expanding-Medicare-Advantage-Health-Plans-in-2022-to-Address-Beneficiaries-Most-Important-Needs-DeliveringPredictableAffordable-and-Understandable-Health-Care
4. Millions of Americans Are Housing Insecure: Rent Relief and Eviction Assistance Continue to Be Critical
5. https://www.hcup-us.ahrq.gov/reports/statbriefs/sb268-ED-Costs-2017.jsp

6. Greendoors: The cost of Homelessness Facts
7. https://www.census.gov/data/tables/2021/demo/hhp/hhp35.html
9. https://nhchc.org/wp-content/uploads/2019/08/homelessness-and-health.pdf