Humana-Mays Healthcare Analytics

Case Competition - 2022

Leveraging Analytics to Address Housing Insecurity

Sunday 16th October, 2022

Contents

1 Introduction **4**

1.1 Goal of the study . . . 4

1.2 Key Performance Indicators . . . 4

1.3 Financial Motive . . . 5

2 Problem Background **8**

2.1 Literature Survey . . . 8

2.2 Previous Approaches to the Problem . . . 9

2.3 Why Data could solve this problem? . . . 10

3 Exploratory Data Analysis **11**

3.1 Feature Groups . . . 11

3.1.1 Population & Regional Features (atlas) . . . 11

3.1.2 Behavioral Health Features (bh) . . . 12

3.1.3 Charlson Comorbidity Features (cci) . . . 12

3.1.4 CMS Features (cms) . . . 13

3.1.5 CMS Level 1 & Level 2 (cmsd1, cmsd2) . . . 13

3.1.6 Member Interactions (cnt) . . . 13

3.1.7 Constructed Indices (cons) . . . 14

3.1.8 Loan Accounts (credit) . . . 14

3.1.9 Latest Non-Behavioral Claims (med) . . . 14

3.1.10 Claim Lines by Revenue Code (rev) . . . 14

3.1.11 Robert Wood Johnson Foundation Data (rwjf) . . . 15

3.1.12 Prescription Features (rx) . . . 15

3.1.13 In-Patient Facility Usage Data (total) . . . 16

3.1.14 Categorical Feature Analysis . . . 16

3.2 Data Cleaning . . . 22

3.2.1 Removal of features with cardinality = 1 . . . 22

3.2.2 Convert features with missing values and cardinality = 2 to binary . . . 22

3.2.3 Handling missing information . . . 23

3.2.4 Categorical Features Processing . . . 24

3.2.5 Standard Scaling . . . 24

3.3 Baseline LightGBM . . . 25
3.4 Explanatory Variables . . . 25
3.4.1 Prominent Features Identification . . . 25
3.4.2 Target Variable: Housing Insecurity Indicator . . . 27
3.5 Analytical Modeling . . . 28
3.5.1 Comparative Study of Different Machine Learning Baselines . . . 28
3.5.2 LightGBM Model - Parameter Tuning . . . 29
3.5.3 LightGBM Model Evaluation . . . 30

4 Analysis of Results **34**
4.1 Top Features . . . 34
4.2 Fairness & Equity in Model . . . 36

5 Strategies & Implementations **38**
5.1 Mental Health Care Improvement in Rural Areas . . . 38
5.1.1 Expansion of Humana Neighborhood Centers . . . 39
5.1.2 Temporary locations in collaboration with local bodies . . . 40
5.1.3 Benefit Analysis . . . 40
5.1.4 Solutions for Metro Areas . . . 41
5.2 Generic Drugs Usage . . . 41
5.2.1 Strategies to Pivot to Generic Drugs . . . 41
5.2.2 Improve Availability of Generic Drugs . . . 42
5.2.3 Incentivize Customers to Switch to Generics . . . 43
5.2.4 Benefit Analysis . . . 44
5.3 CT Scans & Radiology Diagnostics Tests . . . 44
5.3.1 Alternatives and Advancements in the field of medical imaging . . . 45
5.4 Individuals on AIDS Medication . . . 45
5.4.1 Long Term Care Solutions . . . 45

6 Conclusion **46**

Executive Summary

Housing insecurity is a critical social problem that is deeply intertwined with the mental and physical well-being of a person. As a result, housing insecurity is costing Humana $153 million annually. Humana provided an anonymized sample of 48,300 people from their MAPD subscriber base of 4.9 million for this study. Out of the 880 features provided by Humana in the dataset, 156 key predictors were identified and a LightGBM model was used to accurately predict housing insecurity. Overall, we obtained a AUC score of 0.7507 and ensured fairness across gender and race demographies. We further investigated mental health and geographic factors, and proposed cost-saving strategy worth $120M by leveraging Humana's Neighbourhood Center and wellness programs. Furthermore, we analyzed the prescription drug spending patterns and proposed cost rationalization strategies to obtain additional savings worth $225M for Humana.

1 Introduction

1.1 Goal of the study

Housing insecurity and the medical well-being of a person are highly interlinked and feed into each other. Research has shown that housing insecurity puts a person at multiple physical and mental health risks and vice-versa [4, 5]. In this study, we investigate the significant interactions between the housing insecurity of a person and their health. Specifically, we focus on the Medicare Advantage Plan D members of Humana, identify the specific medical-related pain points that are leading to and/or effected by housing insecurity, and provide appropriate recommendations based on data analytics. The broad objectives of this study include:

- Create a Fair Prediction Model for each member of MAPD to identify members who are most likely to be Housing Insecure (HI).
- Identify features that Humana may use to classify a member as Housing Insecure.
- Provide recommendations based on insights from the data which translate into actionable outcomes.
- Come up with a strategic plan to implement the suggested solutions.

1.2 Key Performance Indicators

One of the key objectives of this study is to accurately predict if a person is likely to be housing insecure based on various health, economic and geographic factors available with Humana. This prediction helps us in two ways. Firstly, if housing insecurity is seen to cause certain medical needs, then accurate prediction of housing insecurity could help Humana proactively predict the medical needs of the person and assist accordingly. Secondly, if certain medical features (ex. AIDS prescription, CT scans, etc.) are seen to cause housing insecurity, then Humana could focus on making them affordable. Measures could include finding affordable alternatives such as generic drugs, and strengthening the pharmacy and diagnostics networks to lower cost per transaction. In turn, these lowered costs make the medical services more affordable and reduce the costs for Humana as well as the end user.

Quantitatively, these objectives translate to building a model to predict housing insecurity and improving the model performance by tracking metrics such as ROC-AUC, Precision and Recall. We cannot use accuracy as a metric for this study because of the severe imbalance in the dataset - only 4% of the data is housing insecure and the remaining data does not have have housing insecurity. Therefore, if the model is tuned to optimize of accuracy, if could simply predict all the data points as housing insecure and achieve 96% accuracy. Hence we use metrics such as AUC that balance the accuracy for both the classes of data (with and without housing insecurity).

This is reflected in business terms as follows: choosing a model based on ROC-AUC reduces the likelihood that we over-estimate the housing insecurity of the customers and squander the resources. Simultaneously, higher AUC score reduces the likelihood that the medical expenses arising out of housing insecurity over shoot the budgeted estimates for payout.

1.3 Financial Motive

Humana's National housing strategy is focused on three key areas.

- Housing Stability and Homelessness Prevention
- Stabilizing Individuals with Significant Health Risks with Incremental Clinical Support
- Strategic Investments to Increase Community Capacity.

Improving the housing conditions has shown to lead to reduced medical costs in the past. Philadelphia and Arizona has running very successful efforts to reduce medical costs.[11]

Looking into the data we see that Housing Insecure MAPD members on average have 35.7 USD greater medical costs than non Housing Insecure. A split of these cost differences are given below:

Drug Label	Housing Secure	Housing Insecure	Cost Difference
rx_branded_pmpm_cost	242.1877946	309.7823229	67.59452835
rx_overall_pmpm_cost	317.8012836	353.5408499	35.73956624
rx_maint_pmpm_cost	214.2855587	243.1138763	28.82831764
rx_nonmail_pmpm_cost	74.23808627	97.05923513	22.82114886
rx_nonbh_pmpm_cost	231.5040832	251.1074693	19.60338612
rx_nonspecialty_pmpm_cost	187.6233264	206.8642351	19.24090872
rx_nonotc_pmpm_cost	247.4422	262.2355524	14.79335242
rx_specialty_pmpm_cost	35.49414166	49.22642587	13.73228422
rx_hum_32_pmpm_cost	69.21699883	82.81221435	13.59521552
rx_specialty_ntwk_pmpm_cost	39.28604911	51.20706327	11.92101416
rx_mail_pmpm_cost	154.6225854	164.0276015	9.405016088

We see that costs per month of drugs related to certain diseases such as Asthma, Hypertension and Psoriasis apart from other speciality drugs have higher per month costs for Housing Insecure MAPD members. The expenditure on branded drugs is especially high and may actually be contributing to housing insecurity. We would suggest a shift to generic drugs from branded ones for Housing Insecure households to reduce the economic burden on them.

We also see higher costs related to diagnostic tests amongst housing insecure MAPD members. Here, the cost for CT scans is of particular significance.CT scans are usually taken after a patient has been admitted. These costs can be avoided with preventative tests.We also notice that housing secure members usually undertake more preventative tests compared to housing insecure members, For example we see a higher number of blood tests amongst housing secure members.

hi_flag	Average Mental health Claims per year	Average CT scans claims per year
0	1.747951	0.416037
1	2.351785	0.49278

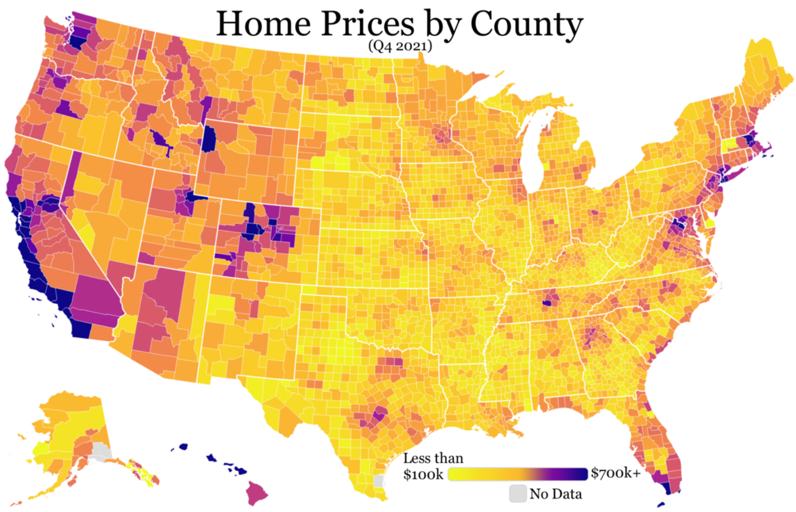
Housing insecure members appear to have a greater frequency for mental health claims (approx 45% more) than Housing secure members. This is also backed by several previous researches[4]. There is significant reason to believe that mental health is strongly correlated with housing stability. There may also be a cycle where poor mental health causes housing insecurity and vice versa.

To sum up there is significant financial motive for Humana to intervene in the state of Housing Insecure MAPD members. The Potential savings of Humana can be calculated per person per month as:

$$\text{No. of MAPD members} \times \text{Percentage of HI} \times (\text{Costs saved from Conversion}) \tag{1}$$

Category	Cost saved per month per person
Prescription Drugs	$35.7
C.T. scans	$21.3
Mental health Claims	$8.1[13]
Total	$65.1

Thus, the potential revenue from solving the housing insecurity problem currently represents a value of approximately **$153,000,000** annually for Humana. This is value is calculated assuming the same rate of Housing Insecurity as represented in our dataset and counting 4.9 million medicare advantage members for Humana. We have also assumed that each mental health claim is also associated with atleast the same number of psych visits. CT scans costs are calculated as product of avg no of scans and avg cost **We have assumed that a person converting from Housing Insecure to Housing secure will follow the medical health patterns of the Housing secure category and the savings that result from this change in behavior will accrue directly to Humana.**



Source: https://en.wikipedia.org/wiki/Housing_insecurity_in_the_United_States

Furthermore, The number of housing insecure population in America is expected to grow given the rising prices of real estate in the country which might lead to greater numbers of Housing Insecurity. This problem thus, requires intervention from organizations committed to improving people's health.

2 Problem Background

2.1 Literature Survey

Housing Insecurity (HI) can be defined as:
"Limited or uncertain availability of, or inability to, acquire stable, safe, adequate, and affordable housing and neighborhoods in socially acceptable ways." [1]

Housing Insecurity may be experienced in several ways, including homelessness, eviction, cost-burdened homes or incidents of violence at home to name a few. These experiences tend to surge in the aftermath of major economic, political or public health events. The housing sector in the United States is experiencing unprecedented levels of unaffordability, with median rent prices and home values increasing by 61% and 112% respectively (inflation adjusted), between 1960 and 2016 [2]. One Example of this can be seen in the aftermath of the COVID-19 pandemic where housing prices increased by up to 26% between March 2020 and November 2021 [3].
However, Housing Insecurity must not be measured simply by whether a person has access to housing, but we must take into other factors which may cause them to be vulnerable. For example,

- Financial factors: Income, Rent, Affordability, etc.
- Domestic factors: Domestic violence, Mental stress, Physical strain, etc.
- Social Factors: Neighborhood safety, Incidents of crime, etc.
- Structural Factors: Age of the building, Materials used, Building maintenance, etc.

Realistically, not all these features may be available to consider while determining whether a person is Housing Insecure or not. Data could be limited due to a variety of reasons including but not limited to privacy norms, member hesitancy to submit this data and simply incomplete database entries.

The effects of Housing Insecurity on a person's health can be observed indirectly through several secondary indicators.

- It has been found that unaffordable housing was associated with increased odds of poor self-rated health and conditions such as hypertension and arthritis [7]
- Housing insecure participants were two times more likely to self-report poor physical health than housing secure participants [6]

- Housing insecurity was also associated with increased risk for adverse kidney outcomes [8]

These results give insights into the process of health deterioration that occurs to individuals who experience housing uncertainty at a physiological level. Thus, it is clear that housing insecurity can have immediate and long-lasting negative implications for various facets of physical health. Additionally, housing insecurity can also contribute to poorer health indirectly through multiple pathways such as environmental, financial and social.

2.2 Previous Approaches to the Problem

The problem of identifying Housing Insecure population has been attempted using several methodologies in the past. Most of these methods have relied upon predetermined factors to come up with a definition of Housing Insecure person. Some common factors that these methodologies have used in the past include:

1. GIS methods[4]
2. Categorizing based on Rental Defaults[1]
3. Categorizing based on Housing Quality[1]
4. Residential Instability[2]
5. Neighborhood Quality[2]

These approaches are mainly concerned with qualitatively identifying Housing Insecurity and provide little to no quantitative parameters. Another problem with categorizing people based on these predetermined factors is that the information needed by these methods might not be readily available or even realistically feasible to get.

One problem with these methods is that they rely mostly on data from surveys and would be greatly biased on the location of the data source, i.e. These models tend to classify large groups of people as housing insecure simply based on their locality. For example, everyone in low income neighborhoods may be at high risk of housing insecurity according to these models or people in higher income neighbourhoods may be excluded from being classified as housing insecure.

In this study we also explore why data is a better solution to this problem and how it addresses the drawbacks of the earlier approaches as well as how it may supplement the results from earlier studies.

2.3 Why Data could solve this problem?

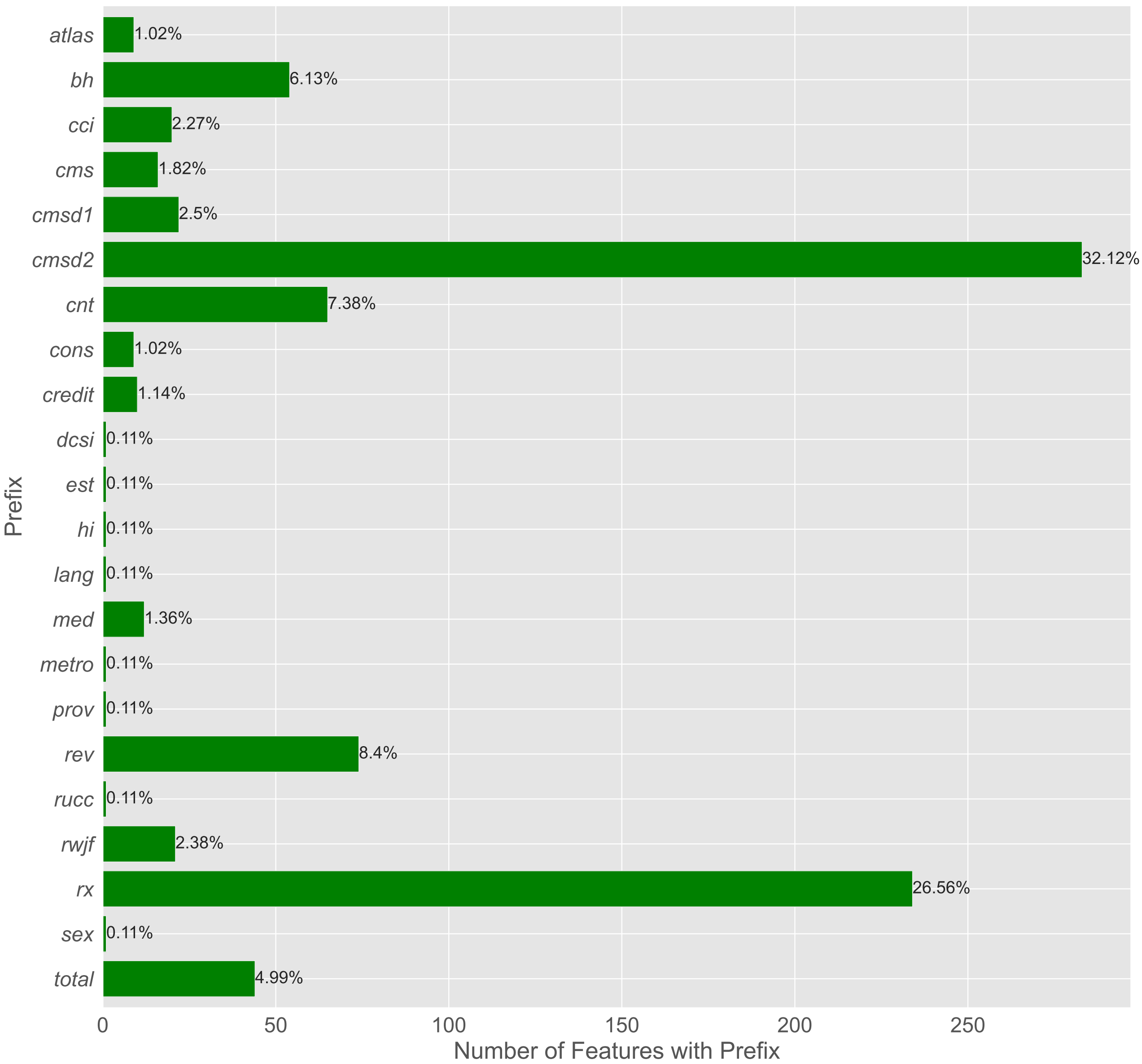
Data is the next oil. In today's information age data is more readily available than ever, owing to advancements in technology and communications. Humana as a healthcare service provider accumulates large amounts of data from its customers over the course of its operations. As we explore the feasibility of solving this classification problem in the sections that follow, we will show how this problem can be solved as a byproduct of the data collected by Humana.

Our problem in this study is quite different from what earlier studies have tried to do. Most of the previous work aims at defining the term "Housing Insecure", while we operate under the assumption that this has already been taken care of. Our job is to judge whether a person may be housing insecure and if so what features or markers could help in identifying them and to what extent. An in-depth analysis of the data given to us reveals several striking correlations between features and also provides insight into the future course of action to take.

3 Exploratory Data Analysis

The training dataset comprises of 48300 data points with 880 features and 1 binary response column stating housing insecurity (hi_flag). While the number of features is extremely high to visualize effectively, they can be grouped into a few categories leveraging the prefix of each column name and evaluated thoroughly.

Figure 1: Counts of features grouped by their prefix



3.1 Feature Groups

3.1.1 Population & Regional Features (atlas)

These features give key indicators of the quality of life of the individual and the general demographics of their surroundings. Indicators related to change in the population and the percentage of the older population are also included in this feature set.

Name	Desc	Sample	NaN Cnt	Uniq Cnt	25%	median	75%
atlas_naturalchangerate1016	Natural population change 2010-2016	NaN	11638	516	1.092	2.876	4.19
atlas_totalpopest2016	Population Size	36913	0	13205	67532.2	237940	748150
atlas_pct_diabetes_adults13	Adult Diabetes Rate	15.9	381	119	9	10.5	12.3
atlas_net_international_migration_rate	Net international migration rate 2010-2016	NaN	11150	517	0.669	1.363	2.052
atlas_totalocchu	Total number of occupied housing units	14122	0	11465	26838.2	89599	278996
atlas_age65andolderpct2010	Percent of population 65 or older	17.13	0	1679	11.08	13.06	15.52
atlas_orchard_farms12	Orchard farms	11	892	145	9	19	53
atlas_snapspth16	SNAP-authorized stores/1,000 pop	NaN	11712	516	0.588	0.726	0.883

3.1.2 Behavioral Health Features (bh)

There are 54 behavioural features that can be classified into three main categories:

- **Allowed Cost per Month (13):** These are allowed costs in the medicare advantage plan across a few psychological and mental health related health care.
  - Residential treatment centers
  - Ambulance place of treatment
  - Rehabilitation inpatient facilities
- **Count of claims per Month (40):** These are counts of claims averaged per month for behavioral health care.
  - Chemical dependence on alcohol or tobacco
  - Bipolar disorder
  - Anxiety
- **Ambulance visits per month(1):** This singular feature shows the number of times individuals have utilized ambulance care for behavioral health care.

Name	Desc	Sample	NaN Cnt	Uniq Cnt	min	25%	median	75%	max
bh_ambulance_allowed_pmpm_cost	Per month allowed cost for Amb treatment	0	0	210	0	0	0	0	303.05
bh_cdto_pmpm_ct	Per month claims (tobacco dependence)	0	0	142	0	0	0	0	4.13
bh_ambulance_visit_ct_pmpm	Ambulance visits per month	0	52	3	0	0	0	0	0.167

3.1.3 Charlson Comorbidity Features (cci)

Charlson Comorbidity Index is aggregated score indicating how sick the patient is based on 19 diseases. This set of 20 features gives information on the overall index as well as the specific claims for the underlying individual diseases.

Name	Desc	Sample	NaN Cnt	Uniq Cnt	min	25%	median	75%	max
cci_score	Charlson Comorbidity Index	4	48	17	0	3	4	6	16
cci_dia_m_pmpm_ct	Per month claims related to diabetes	0.16	0	326	0	0	0	0.16	9.41

3.1.4 CMS Features (cms)

Centre for Medicare Services(CMS) features that indicate Medicare advantage plan details and other information about existing care/health plans for each individual. Categorical features of Race and Risk Adjustment factors are also included in this 16 feature set.

Name	Desc	Sample	NaN Cnt	Uniq Cnt	min	25%	median	75%	max
cms_disabled_ind	Indicator for member age < 65	0	0	2	0	0	0	1	1
cms_institutional_ind	Indicator for institutional member	0	0	1	0	0	0	0	0
cms_tot_partd_payment_amt	Total Part D Payment	NaN	45981	7	0	0	0	0	146.92
cms_risk_adjustment_factor_a_amt	Risk Adj Factor A Amount	NaN	18867	427	0	0.398	0.557	0.818	1.571

3.1.5 CMS Level 1 & Level 2 (cmsd1, cmsd2)

This large set of 22 *level-1* and 283 *level-2* features give information about the count of claims per month across different disease types and sub-types defined by CMS.

CMS Level 1 sample features

Name	Desc	Sample	NaN Cnt	Uniq Cnt	min	25%	median	75%	max
cmsd1_vco_pmpm_ct	contact with health services	0.16	0	560	0	0.25	0.5	0.91	24.58
cmsd1_skn_pmpm_ct	skin and subcutaneous tissue	0.16	0	235	0	0	0	0.08	14.66
cmsd1_bld_pmpm_ct	diseases of the blood and blood-forming organs	0	0	246	0	0	0	0.08	11.75
cmsd1_dig_pmpm_ct	diseases of the digestive system	0	0	319	0	0	0	0.16	9.99

CMS Level 2 sample features

Name	Desc	Sample	NaN Cnt	Uniq Cnt	min	25%	median	75%	max
cmsd2_res_res_up_other_pmpm_ct	diseases of upper respiratory tract	0	0	106	0	0	0	0	3.66
cmsd2_ano_cleft_pmpm_ct	congenital malformations, deformations	0	0	1	0	0	0	0	0
cmsd2_sns_speech_pmpm_ct	symptoms of speech and voice	0	0	56	0	0	0	0	1.75
cmsd2_sns_men_pmpm_ct	symptoms of cognition, perception	0	0	162	0	0	0	0	4.25

3.1.6 Member Interactions (cnt)

The Medicare advantage subscribers interact with Humana through different media like email, calls, physical mail, etc. The features have granularity at a monthly level.

Name	Desc	Sample	NaN Cnt	Uniq Cnt	min	25%	median	75%	max
cnt_cp_emails_pmpm_ct	Per month emails (overall 1 year)	0	0	777	0	0	0.83	3.16	13.66
cnt_cp_emails_2	Per month emails (lag 2)	0	38	16	0	0	0	2	15
cnt_cp_livecall_4	Per month livecall (lag 4)	0	0	1	0	0	0	0	0
cnt_cp_vat_4	Per month vat (lag 4)	0	29	14	0	0	0	1	13

3.1.7 Constructed Indices (cons)

Constructed Indices that give information related to financial stability and health management are covered in this 7 feature set.

Name	Desc	Sample	NaN Cnt	Uniq Cnt	min	25%	median	75%	max
cons_lwcm10	Probability of not exercising	0.518	21417	274	0.056	0.226	0.29	0.345	0.708
cons_hxmioc	Managing Illness or Condition - Index	7	11172	10	0	6	8	9	9
cons_hxmboh	Managing the Business of Health	9	11159	10	0	7	8	9	9
cons_stlnindx	Student Loan Index	8	11168	10	0	7	8	9	9
cons_ccip	Census Income Percentile	27	11190	99	1	28	49	69	99
cons_stlindex	Short Term Loan Index	5	11159	10	0	5	7	8	9
cons_hxmh	Managing Health - Index	9	11187	10	0	7	8	9	9

3.1.8 Loan Accounts (credit)

Data of current credit utilization, balance amounts of loans, and number of due loan accounts give deeper information about the financial situation of the individual. These 10 feature tend to have large number of missing information, which subjectively would mean there are no loan accounts currently for the member.

Name	Desc	Sample	NaN Cnt	Uniq Cnt	min	25%	median	75%	max
credit_prcnt_mtgcredit	% Balance to High Mortgage Credit	NaN	45217	116	61.188	73.755	77.182	79.774	87.125
credit_bal_autobank_new	Auto Bank Loan Accts in 12 months	NaN	45243	116	532.549	1471.31	1985.65	2573.57	4005.18
credit_num_1stmtg_collections	Mortgage Accts - 120 Days Past Due	NaN	43469	117	0	0	0.001	0.001	0.004

3.1.9 Latest Non-Behavioral Claims (med)

This 12 feature set contains information of the number of days since the latest medical claim by the member for particular non-behavioral facilities. This could be a critical feature group in identifying individuals who are shifting into housing insecurity.

The feature seems to max out at 480, which probably is a cut-off for number of days being taken into account.

Name	Desc	Sample	NaN Cnt	Uniq Cnt	min	25%	median	75%	max
med_er_obs_ds_clm	emergency room observations	480	0	10	110	480	480	480	480
med_physician_office_ds_clm	physician office	113	0	398	1	29	58	128	480
med_ip_snf_ds_clm	skilled nursing inpatient facilities	480	0	240	28	480	480	480	480
med_ambulance_ds_clm	ambulance place of treatment	480	0	389	7	480	480	480	480

3.1.10 Claim Lines by Revenue Code (rev)

Features covering information of claim lines per month across 74 different revenue codes.

Name	Desc	Sample	NaN Cnt	Uniq Cnt	min	25%	median	75%	max
rev_pm_pystrp_pmpm_cd_ct	physical therapy	0	0	325	0	0	0	0	8.5
rev_pm_xtrp_pmpm_cd_ct	therapeutic services	0	0	79	0	0	0	0	3
rev_pm_nucl_pmpm_cd_ct	nuclear medicine	0	0	52	0	0	0	0	0.75
rev_pm_mri_pmpm_cd_ct	magnetic resonance imaging (mri)	0	0	46	0	0	0	0	0.66

3.1.11 Robert Wood Johnson Foundation Data (rwjf)

Socioeconomic and health data from Robert Wood Johnson Foundation like income inequality, crime rate, ratio of population to mental health providers, etc at a regional level. These 21 broad features help in identifying at-risk regions where greater percentage of populations are prone to housing insecurity.

Name	Desc	Sample	NaN Cnt	Uniq Cnt	min	25%	median	75%	max
rwjf_premature_mortality	Premature age-adjusted mortality	NaN	11675	516	161.978	301.151	352.572	426.598	670.87
rwjf_child_mortality	Child mortality rate	NaN	11768	514	21.333	44.344	51.442	60.49	105.926
rwjf_median_house_income	Median household income	NaN	11612	514	34583	54242	61429	69455	127898
rwjf_violent_crime_rate	violent crime offenses per 100K	NaN	11847	511	33	261.456	399.615	560.254	1819.51

3.1.12 Prescription Features (rx)

All prescription information for members is contained within this 234 feature set. They can be further divided as,

- **Per Month Count (124):** Averaged count per month of prescription for numerous disease types and drug tiers. Also includes classifications by mode of purchase, type of drug or retailer, for example, delivered through mail or whether the drug is generic or branded or shopped at which pharmacy (public, CVS, etc.)
- **Per Month Cost (106):** Averaged cost per month of prescriptions for fewer categories than the count features.
- **Number of facilities used (3):** Number of pharmacies, physicians associated with pharmacies, and the overall number of prescriptions per month averaged over the past year
- **Latest Prescription (1):** Days since the latest prescription for member showcases if there was any recent drug use.

Name	Desc	Sample	NaN Cnt	Uniq Cnt	min	25%	median	75%	max
rx_pharmacies_pmpm_ct	phramacies used per month	0.16	0	199	0	0.16	0.25	0.33	3.32
rx_days_since_last_script	days since last prescription	2	0	372	1	8	17	37	480
rx_tier_1_pmpm_ct	Tier 1 drugs count per month	1.66	0	605	0	0.44	1.03	1.83	10.5
rx_hum_69_pmpm_ct	opthalmology - glaucoma agents count per month	0	0	16	0	0	0	0	0.58
rx_generic_pmpm_cost	generic drugs cost per month	14.96	0	9639	0	1.83	11.03	29.8	3031
rx_hum_73_pmpm_cost	pain mgmt-analgesics drugs cost per month	0	0	1	0	0	0	0	0

3.1.13 In-Patient Facility Usage Data (total)

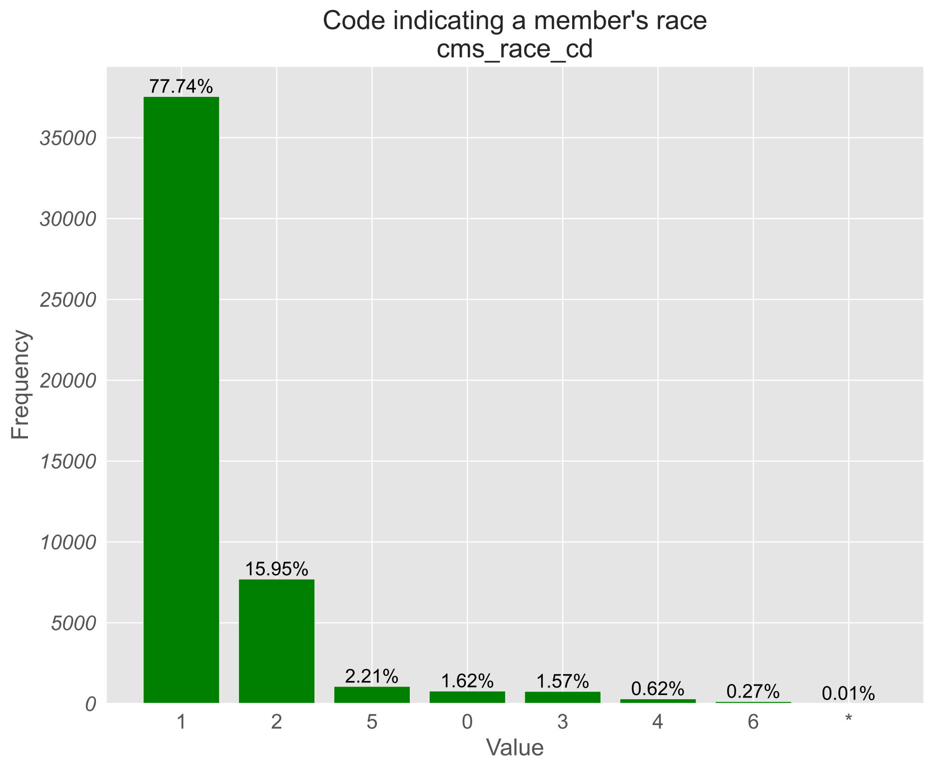
Features related to inpatient admits, facilities usage, and total allowed cost for such use cases. Also captures days since the last claim for any hospitalization or ambulance care.

Name	Desc	Sample	NaN Cnt	Uniq Cnt	min	25%	median	75%	max
total_ip_acute_admit_ct_pmpm	admits per month in acute facilities	0	0	55	0	0	0	0	0.66
total_ip_maternity_admit_days_pmpm	admit days pm in maternity facilities	0	1	1	0	0	0	0	0
total_ip_maternity_allowed_pmpm_cost	allowed cost pm for maternity facilities	0	0	1	0	0	0	0	0
total_er_obs_ds_clm	days since last claim for emergency room	480	0	1	480	480	480	480	480
total_er_visit_ct_pmpm	visits per month for emergency room	0	0	57	0	0	0	0	1.52

3.1.14 Categorical Feature Analysis

Race

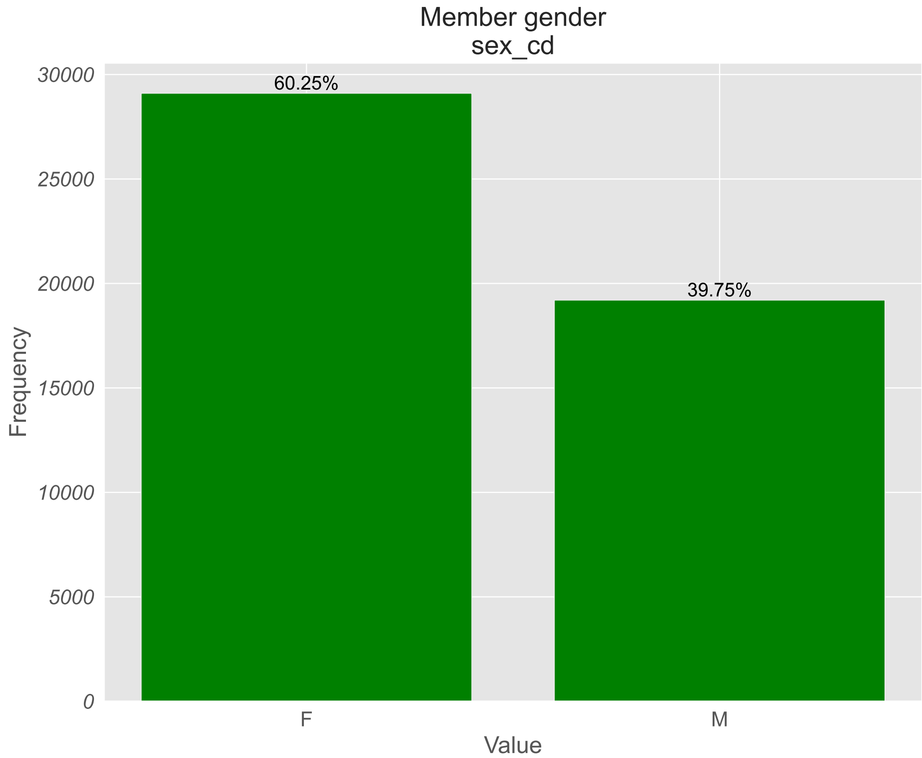
The Race feature is very unbalanced towards non-hispanic white (1) and black (2) members with other members considerably lower in number.



Gender

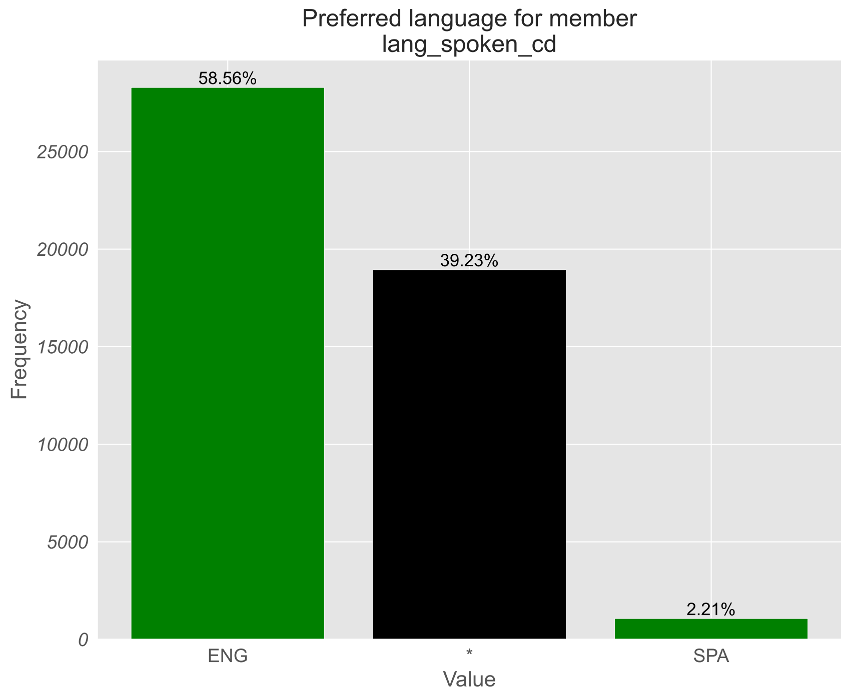
Gender classification is skewed slightly toward female members. This could be interpreted as,

- Female members have greater need of medicare advantage plans due to their social and financial circumstances
- Female members generally live longer and hence form a greater portion of individuals above the age of 65. Consequently, this ratio is reflected in medicare plan memebrs too



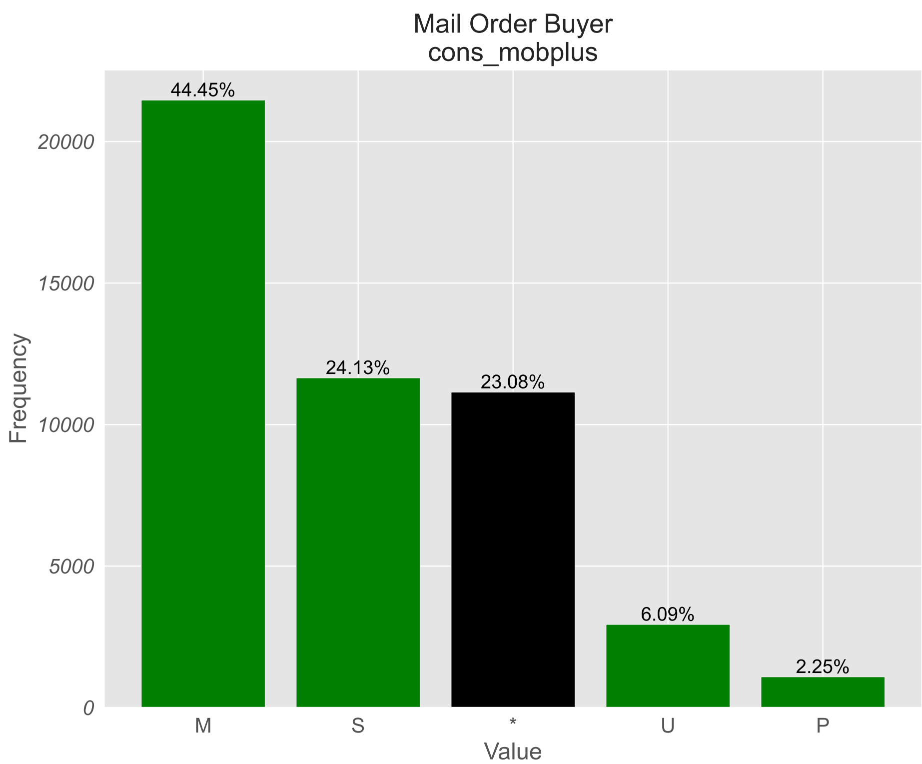
Primary Language

Unfortunately, a lot of the data points for this feature are missing. Subjectively, we could assume these to be any language other than the given languages. But, from the prior information that our data consists of mostly white and black members (93.69%), we expect the missing values for this feature would mostly be "ENG". Nevertheless, we retain the missing information without replacement with "ENG".



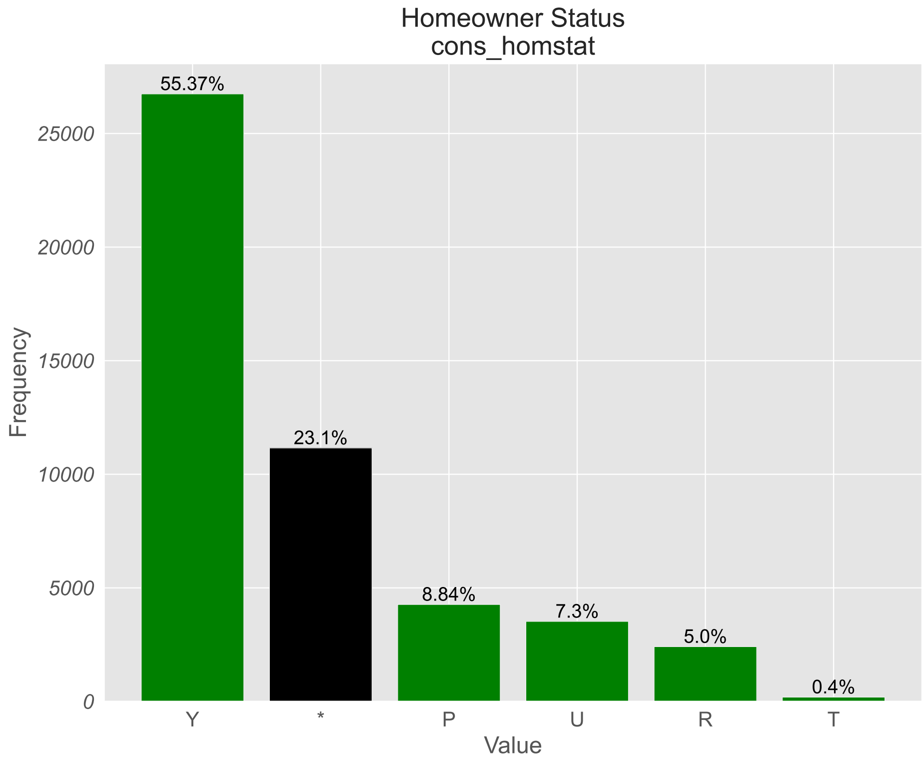
Mail Order Buyer Classification

If members order prescriptions to their homes, it could be an important indicator of housing security and stability. We can see that most members are Multi mail-order buyers (44.45%) or single mail-order buyers (24.13%). Some are also classified as either unknown or probable. A significant 23.08% of members do not have any data available.



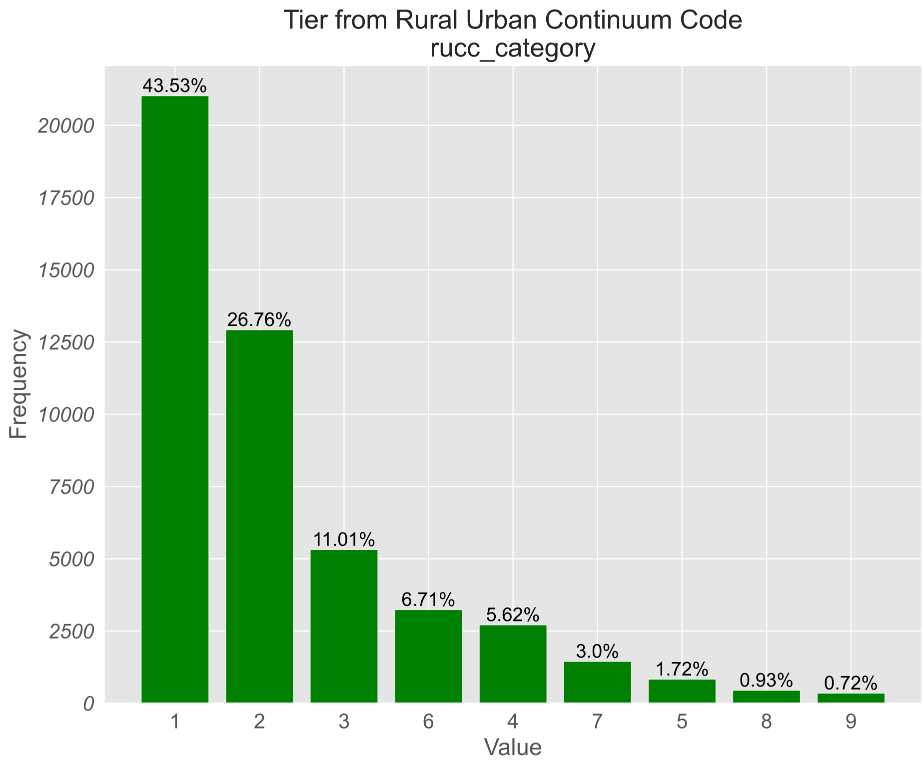
Homeowner Status

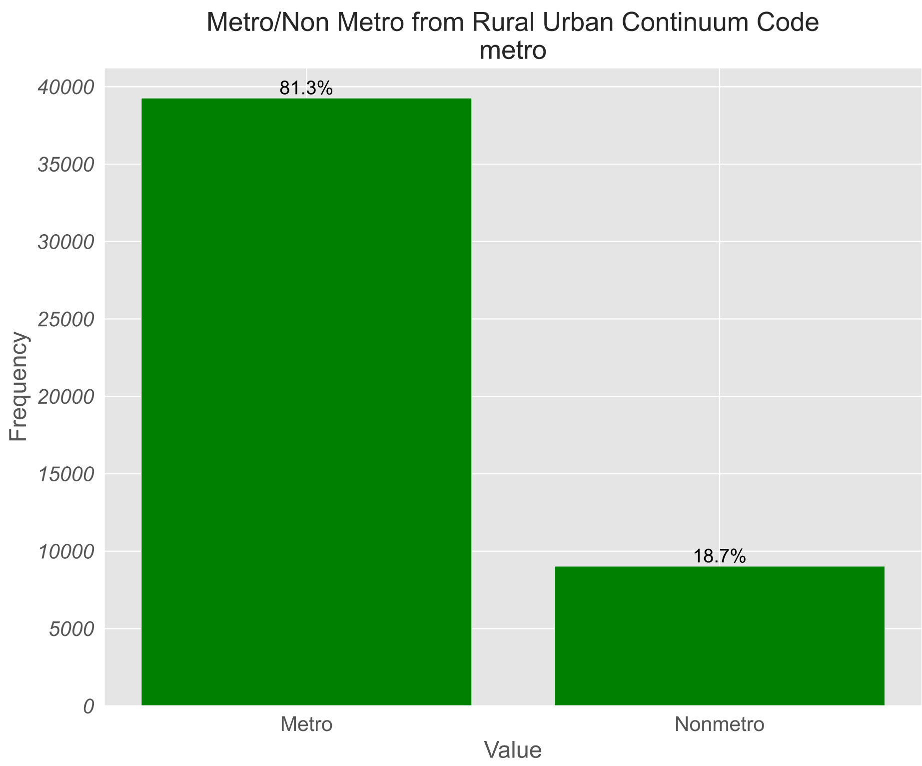
Similar to mail-order buying status, home ownership is a very strong indicator of housing security. It is important to consider that even for memeber that have house ownership, they might not have the best living conditions and could require further maintenance or improvements of their housing.



Rural Urban Continuum Code

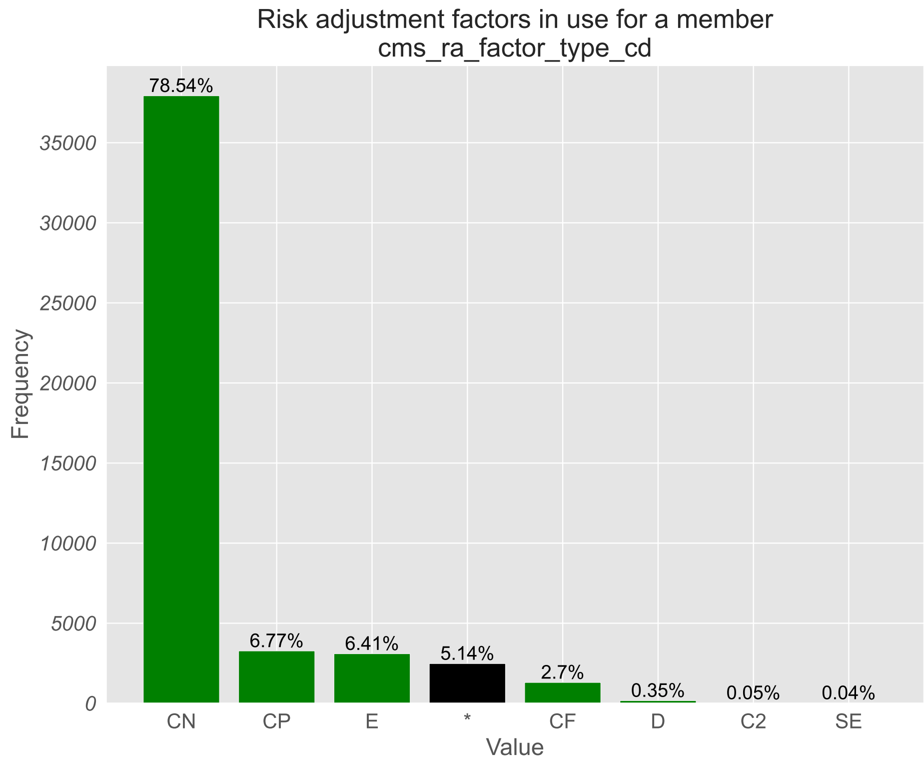
We split the "rucc_category" feature into the Tier and Metro/non-metro classification to further capture any common trends related to metros and non-metros.





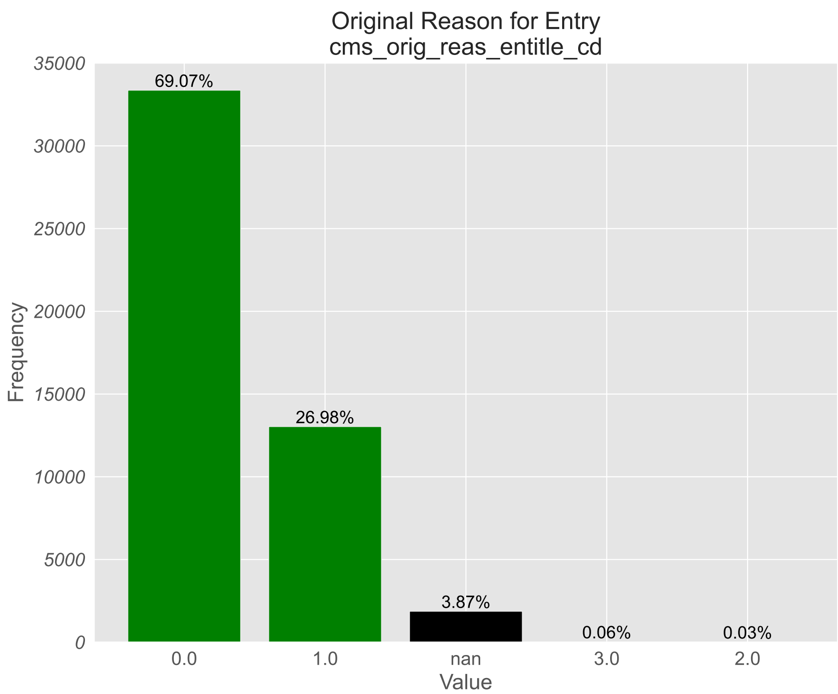
Risk Adjustment Factor Types

This is the code for the risk adjustment in use for the memeber.



Member Reason for Entry

Majority of the member enroll into plans because of Old Age or Disability (96.05%). Some members have missing information.



3.2 Data Cleaning

We performed cleaning operations on our dataset with a few primary objectives:

- Removal or modification of features that do not carry significant information for faster processing
- Imputation of missing features while preserving missing value information
- Conversion of features into numeric data types that can be fed into any model
- Scale features so that any model will give equal weightage to each feature

3.2.1 Removal of features with cardinality = 1

Of the 880 features, there are 119 features whose value is equal for all the data points. Such features do not add any information for modeling tasks and we hence remove them.

Removed Features

Feature name	Feature Desc	Constant Value
total_ip_maternity_ds_clm	days since last claim for overall claims of maternity inpatient facilities	480
total_er_obs_ds_clm	days since last claim for emergency room observations	480
med_ip_maternity_ds_clm	days since last claim for non-behavioral claims of maternity inpatient facilities	480
med_ip_mhsa_ds_clm	days since last claim for non-behavioral claims of mental health and substance abuse	480
rx_phar_cat_express_scripts_pmpm_ct	count per month of prescriptions purchased at express scripts pharmacy	0
	...111 other features	
cnt_cp_livecall_9	count per month of member interactions via livecall (lag 9)	0
cmsd2_ext_occ_3_wheel_pmpm_ct	claims per month of occupant of three-wheeled motor vehicle accidents	0
cmsd2_ext_compl_medical_care_pmpm_ct	claims per month of misadventures to patients during surgical and medical care	0

3.2.2 Convert features with missing values and cardinality = 2 to binary

92 features have only 2 possible values throughout all data points of which 1 value is "NaN". Such features do not carry any scaling information even if they are numerical. We convert such features to binary for faster processing and stability during the modeling process.

Conversion of constant features with missing values

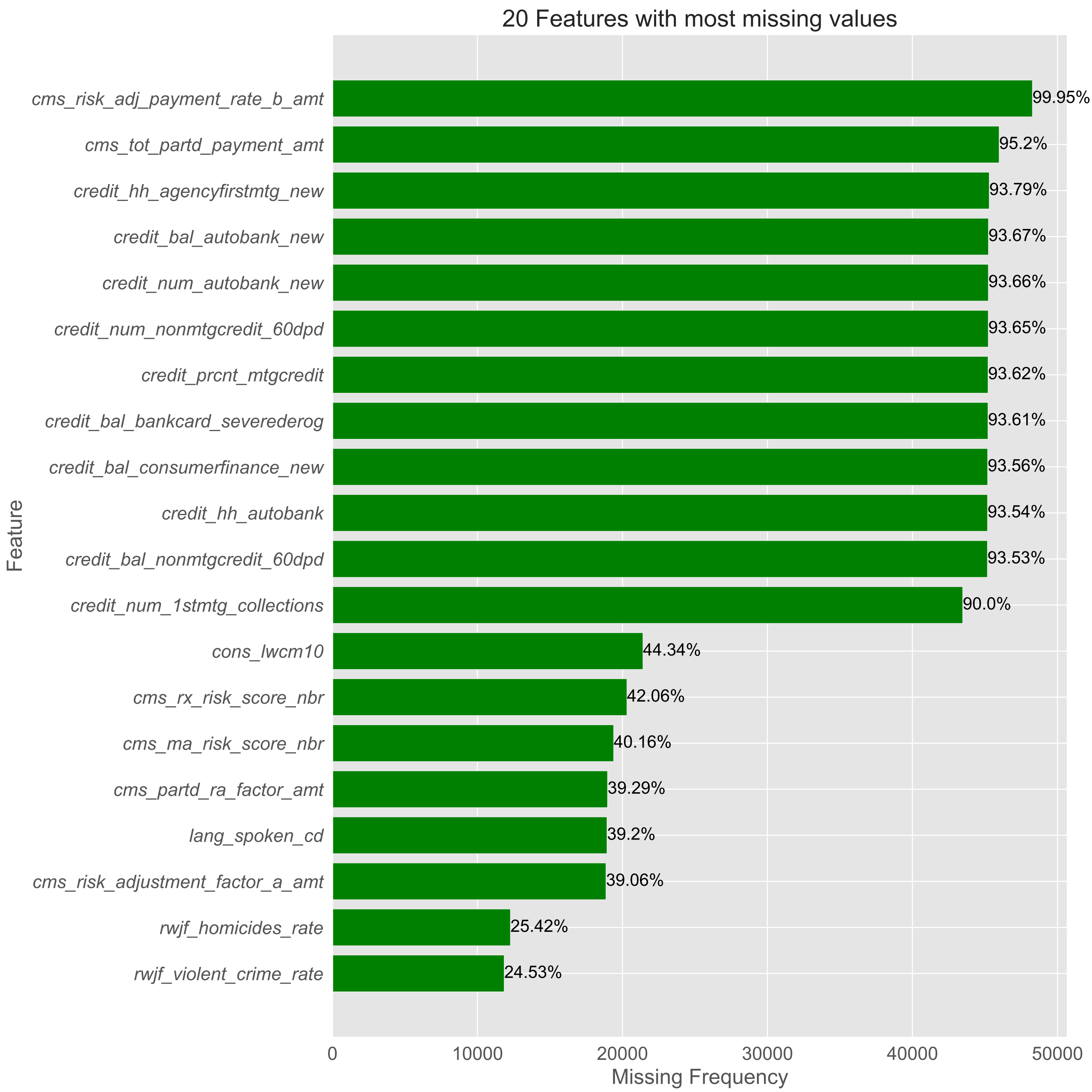
	Constant Sample	NaN Sample
med_ip_ltach_ds_clm	480	NaN
rx_hum_14_pmpm_ct	0	NaN
	...88 other features	
rx_hum_37_pmpm_cost	0	NaN
cmsd1_neo_pmpm_ct	0	NaN

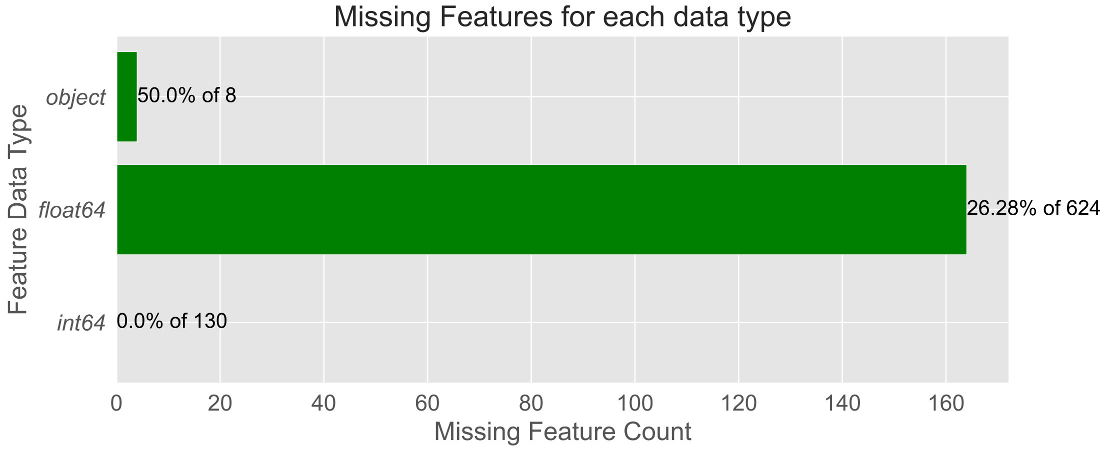
⇒

Converted Constant Sample	Converted NaN Sample
0	1
0	1
...	
0	1
0	1

3.2.3 Handling missing information

Throughout the dataset we have several missing values. Let us take a look at few of columns with the most number of missing values.





There are several methods to deal with missing values:

- Removal of data points with missing values is not feasible as all the rows in our dataset have at least 1 missing value
- Removing individual features also carries the same risks as we lose **168** features or 22.05% of features
- Imputation with mean, median, mode or zero are other possibilities
  - Binary features can be added that preserve missing value information before imputation

We choose to create new features for preserving missing value information and imputed missing values. The numerical features in our data would have a greater chance of being "zero or not applicable" if they are missing compared to having the median value. Hence, we impute missing values with zero everywhere.

For categorical features, we cannot impute with any numerical statistic. We choose to replace missing values with the character (*****), as it already appears in a few columns as a missing value.

3.2.4 Categorical Features Processing

Categorical features cannot be interpreted directly by most models. Tree and boosting tree models like xgboost, lightgbm or catboost have internal routines that can handle categorical features but general models like logistic regression, SVM, KNN can only handle numerical features. Therefore, we employ one-hot encoding to convert categorical features into numerous binary features for use in such models.

3.2.5 Standard Scaling

Similar to categorical features, while tree models are not affect by relative scale of features, models like logistic regression, SVM, KNN weight features depending on their scale. Features with big numbers will be given more weightage or will have greater impact on such model. Hence, we standard scall all features so that any kind of model can be used.

3.3 Baseline LightGBM

As an initial model, we use lightGBM classifier without any hyperparameter tuning. Through this model, we try to understand the relative importance of features.

		Train	Test
LightGBM (is_unbalance = True)	AUC	0.922	0.602
	Precision	0.250	0.100
	Recall	0.978	0.353
	Accuracy	0.871	0.828

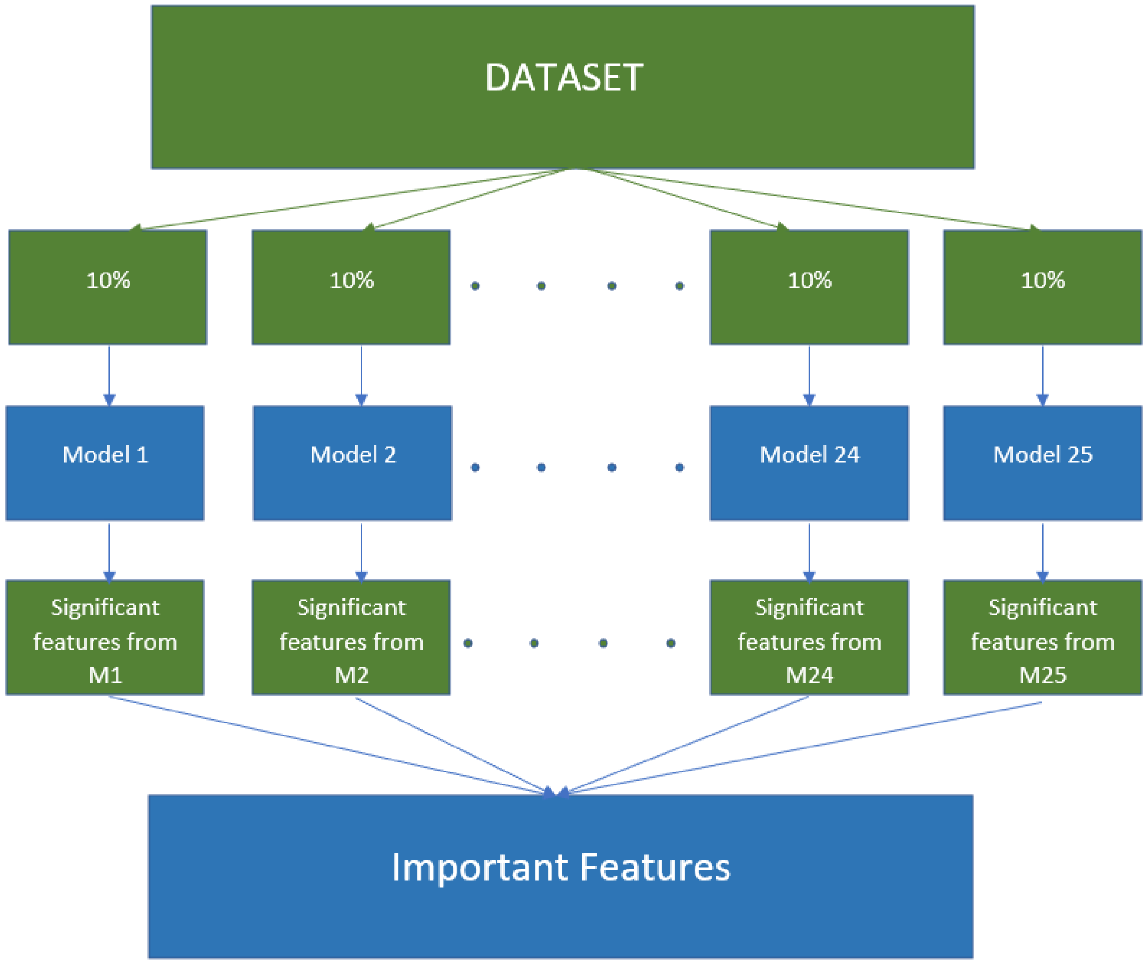
From the train and test performance metric, we suspect that the model is overfitting. We plan to correct this during the hyperparameter tuning setup.

3.4 Explanatory Variables

3.4.1 Prominent Features Identification

After finalizing our baseline and completing data imputation, we moved on to selecting prominent features which can later be used to build better fitted Machine Learning Models. To achieve this, we implemented our own bootstrapping procedure to identify the most important features by leveraging the feature importance metrics available through the lightGBM model. The **feature importance** value is the number of times each variable is used to create a split/branch in a tree-based model. The procedure is outlined below,

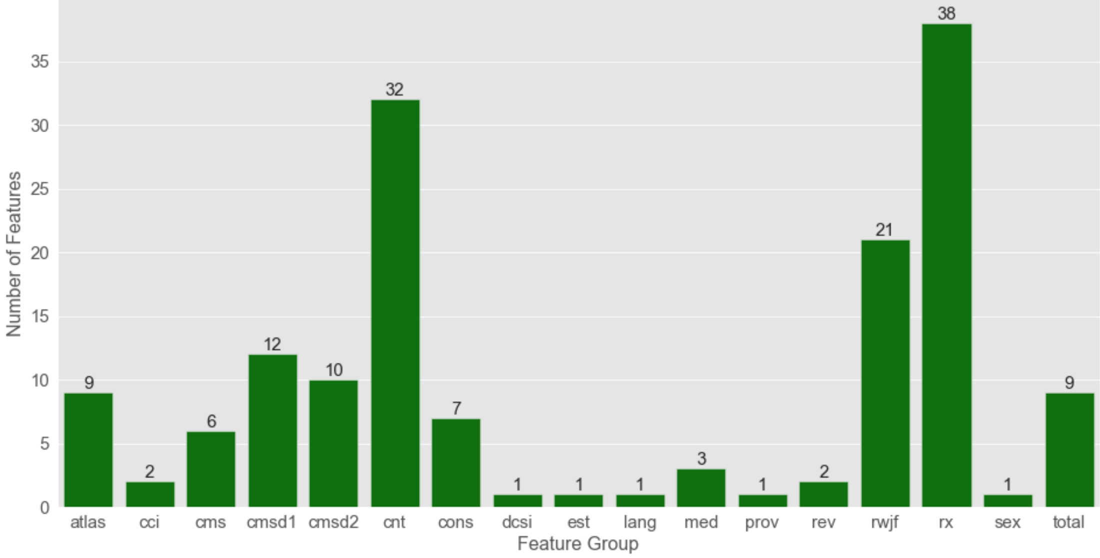
- Create 25 sample data sets that cover 10% of the total data each, with replacement
- Build the baseline lightGBM model over each sample and extract feature importance metric
- Select features that have an importance value of at least 1



Creating different models for different partitions of data allows us to find the various features that can be significant depending on the underlying data distribution. We enforce the condition that the feature importance should have a minimum value of 1 because it ensures that the feature has contributed to the branching of the boosting tree at least once.

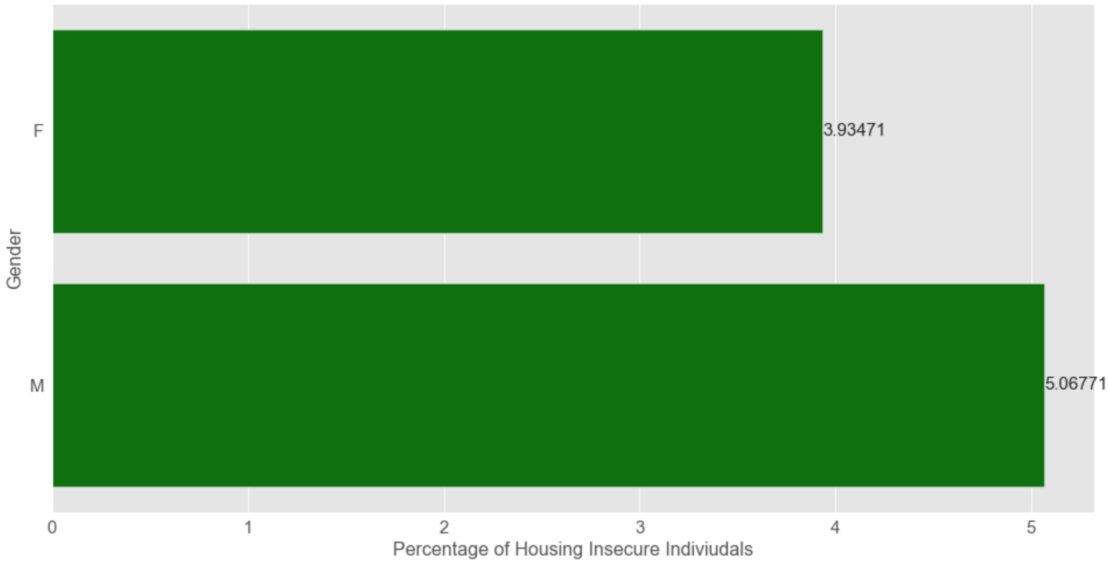
After performing this exercise, we have a resulting data set of 156 predictors. We were successful in decreasing the feature space from 880 features to 156 features (approximately 82% decrease) while maintaining the predictive power of the variables. The final features that we obtained can be classified into the following groups.

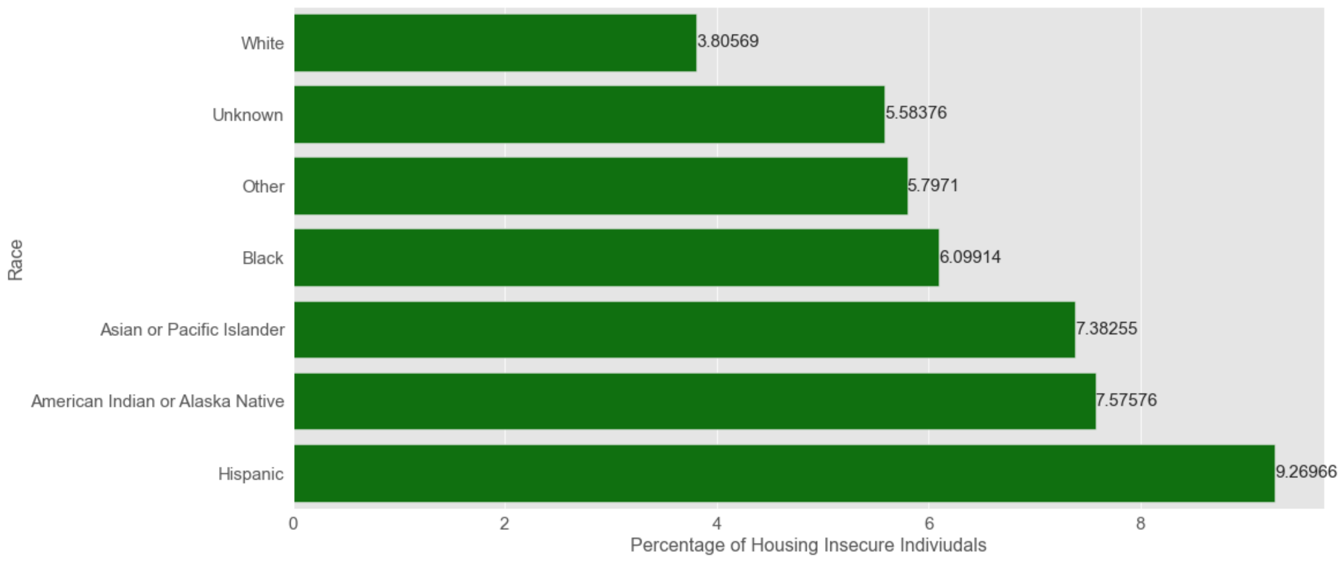
Figure 2: Final selected 156 feature set



3.4.2 Target Variable: Housing Insecurity Indicator

The target variable is the key indicator which is used to identify patterns and explore correlations with important features. In our case, the target variable is a self-reported Housing Insecurity indicator that was collected in the June/July timeframe of 2022. Over the entire population of 48,300 individuals, we observed that 4.385% of them considered themselves to be housing insecure. We further deep dive into different demographics to better understand the problem.





We observed that different demographics have different percentages of individuals who self-identify as housing insecure. For example, we can clearly see that minorities such as Asians or Hispanics have a disproportionately higher rate of housing insecurity. Hence, we recognise that demographic factors such as gender and race are important to analyze the fairness of our model and bring them back while evaluating our final model.

3.5 Analytical Modeling

Once we finalized and reduced the feature space of the dataset from 880 features to 156 features, we started developing machine learning models which would help us identify the likelihood of a Humana member having housing insecurity based on the extracted features.

3.5.1 Comparative Study of Different Machine Learning Baselines

We performed a comparative study of different machine learning baselines to find the one best suited to the problem in hand. A 90-10 split of our data was done to generate train and test data sets.

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
catboost	CatBoost Classifier	0.9547	0.7231	0.0000	0.0000	0.0000	-0.0004	-0.0031	13.9100
lda	Linear Discriminant Analysis	0.9536	0.7201	0.0000	0.0000	0.0000	-0.0024	-0.0077	0.8500
gbc	Gradient Boosting Classifier	0.9536	0.7154	0.0000	0.0000	0.0000	-0.0024	-0.0077	50.7100
ada	Ada Boost Classifier	0.9549	0.7070	0.0000	0.0000	0.0000	0.0000	0.0000	9.3300
lightgbm	Light Gradient Boosting Machine	0.9547	0.7039	0.0046	0.3333	0.0090	0.0078	0.0346	0.5700
rf	Random Forest Classifier	0.9549	0.6598	0.0000	0.0000	0.0000	0.0000	0.0000	2.1400
xgboost	Extreme Gradient Boosting	0.9542	0.6570	0.0138	0.3333	0.0264	0.0229	0.0600	1.5600
et	Extra Trees Classifier	0.9547	0.6464	0.0000	0.0000	0.0000	-0.0004	-0.0031	1.2400
nb	Naive Bayes	0.8656	0.5862	0.1468	0.0646	0.0898	0.0289	0.0318	0.0700
knn	K Neighbors Classifier	0.9536	0.5515	0.0000	0.0000	0.0000	-0.0024	-0.0077	0.0300
lr	Logistic Regression	0.9545	0.5215	0.0000	0.0000	0.0000	-0.0008	-0.0044	5.7600
dt	Decision Tree Classifier	0.9101	0.5181	0.0872	0.0748	0.0805	0.0336	0.0337	6.3000
ridge	Ridge Classifier	0.9549	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0800
qda	Quadratic Discriminant Analysis	0.9549	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2400
svm	SVM - Linear Kernel	0.7915	0.4931	0.1651	0.0418	0.0667	-0.0057	-0.0075	2.9600

Out of all the models we tested out, we found out that boosted tree-based classifiers were performing better with relatively high AUCs. Among these model, we chose the LightGBM model to further tune because of its computational efficiency and better baseline Precision and Recall scores.

3.5.2 LightGBM Model - Parameter Tuning

LightGBM is a gradient boosting framework that uses tree based learning algorithms. It uses two novel techniques: Gradient-based One Side Sampling and Exclusive Feature Bundling (EFB) which makes it work efficiently and provide it a cutting edge over other Gradient Boosting Decision Tree frameworks. [9] Before starting hyper-parameter tuning we did a 90-10 split to get training and test data. The training data was used for GridSearch-based parameter tuning through a process of 10-fold cross validation. The test data was used to evaluate the performance of the tuned model. We aimed to maximize both AUC and Recall for our model. Recall is considered important here as it measures the model's ability to detect Housing Insecure individuals.

We performed hyper-parameter tuning to find the best fitted LightGBM model by performing a GridSearch on a specific set of hyper parameters.

- Is Unbalanced (is_unbalance) : True

  This parameter ensures that the data imbalance between Housing Secure and Housing Insecure members is addressed. The algorithm will try to automatically balance the weight of the dominated label.

- Number of Leaves (num_of_leaves) : [20,50,100]

  This parameter sets the maximum number of leaves each learner has. It controls the complexity of the model. Large number of leaves increases accuracy on the training set but also increases the chance of overfitting.

- Max Depth (max_depth): [2, 5,10]

  This parameter controls the max depth of each trained tree. Large values of this parameter will likely lead to overfitting on the training data set.

- Learning Rate (learning_rate) : [0.05,0.1,0.2]

  This parameter controls the Boosting learning rate.

- Minimum number of data needed in a child (min_child_samples): [5,10,15] This parameter controls the to minimum number of instances needed to be in each node. Larger values of this parameter will lead to more conservative branching of the trees.

- Number of Estimators (n_estimators) : [150, 250]

  This parameter controls the total number of boosting rounds, or the number of gradient boosted trees

- L1 regularization term on weights (reg_alpha) : [0, 0.01, 0.03]

  This is a regularization parameter that combats overfitting

3.5.3 LightGBM Model Evaluation

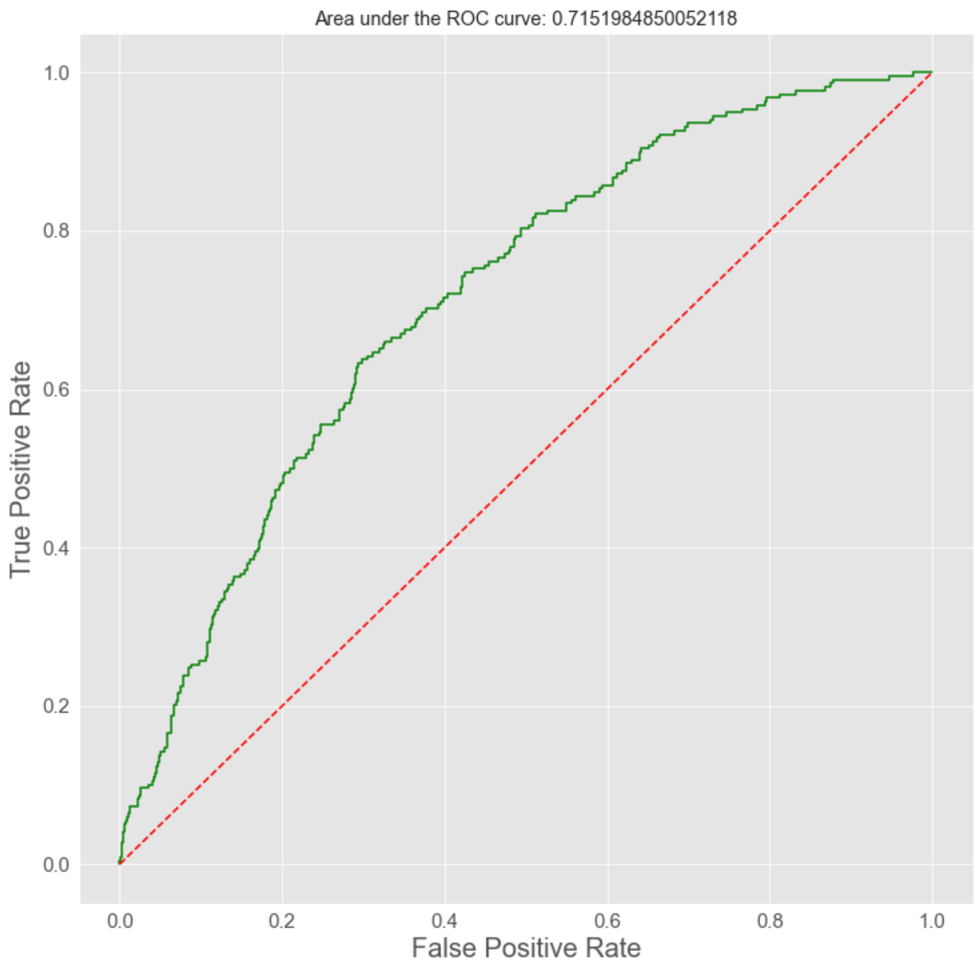
After tuning the parameter, we obtained the following model.

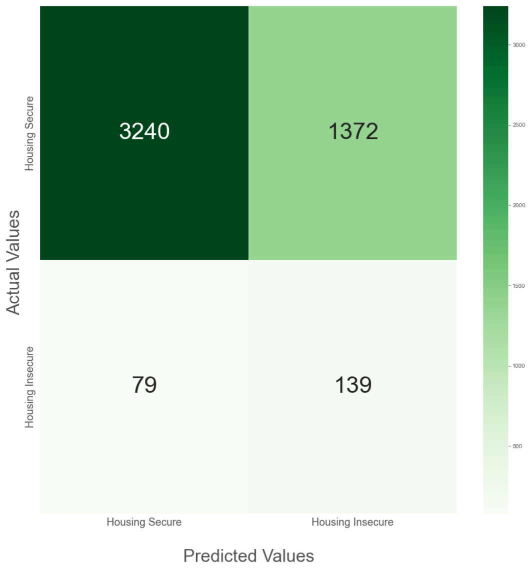
```
LGBMClassifier

LGBMClassifier(is_unbalance=True, learning_rate=0.05, max_depth=2,
               min_child_samples=5, n_estimators=250, num_leaves=20,
               reg_alpha=0.03)
```

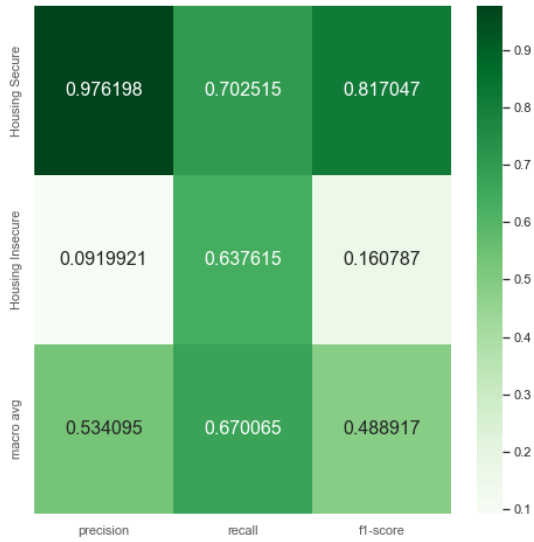
Our tuned model has the following evaluation metrics on unseen test data.

- AUC = 0.7152
- Recall : 0.6376
- Precision : 0.09199
- Accuracy : 0.6996

Due to our targeted tuning, we were able to increase our model's AUC to 0.715 with a relatively high recall of 0.642. For further evaluation, we examined other metrics as well.

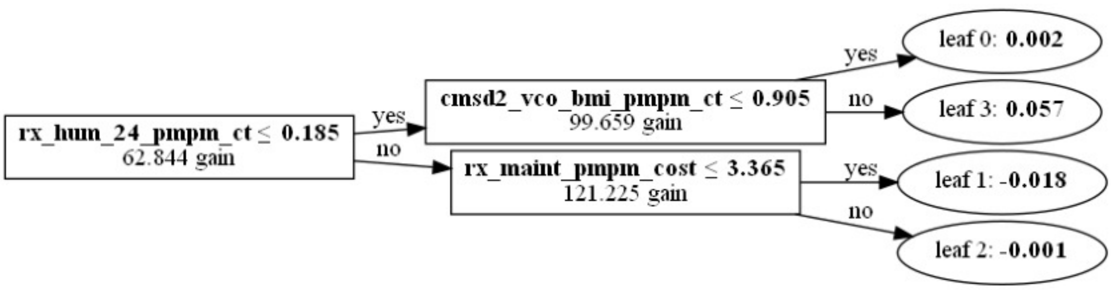


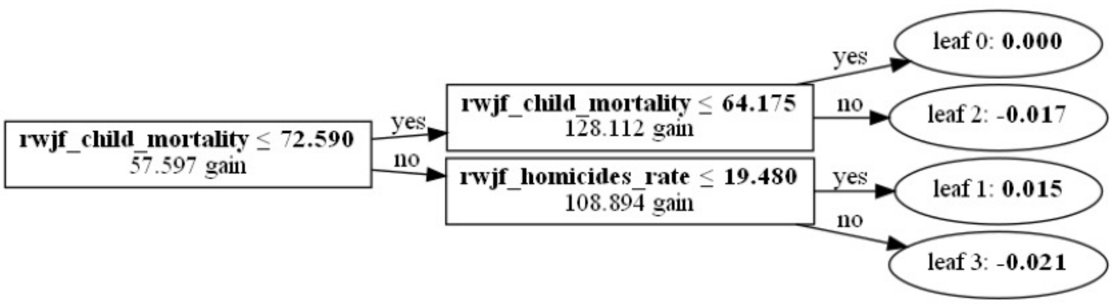


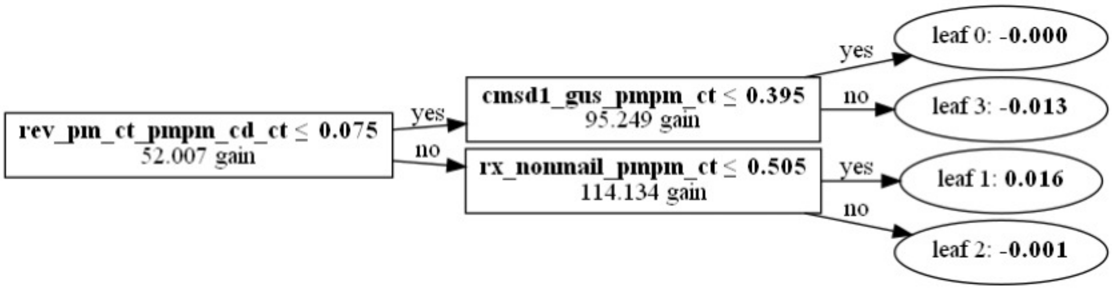


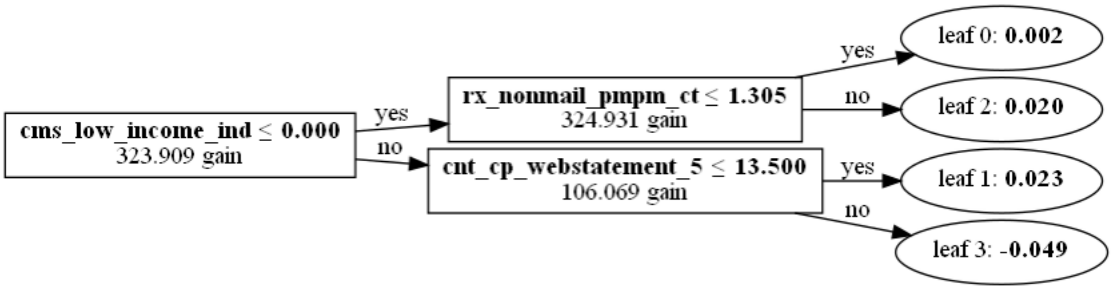
We can see from the classification report that our model is extremely precise in finding if an individual is Housing Secure. However, the main takeaway from the model is in its ability to find out all Housing Insecure individuals in the given population with high recall. This allows the model to learn patterns and feature dependencies which allows it to find out if the an individual is housing insecure.

Additionally, we can observe some of the decision trees used in our LightGBM model and understand different decision pathways.





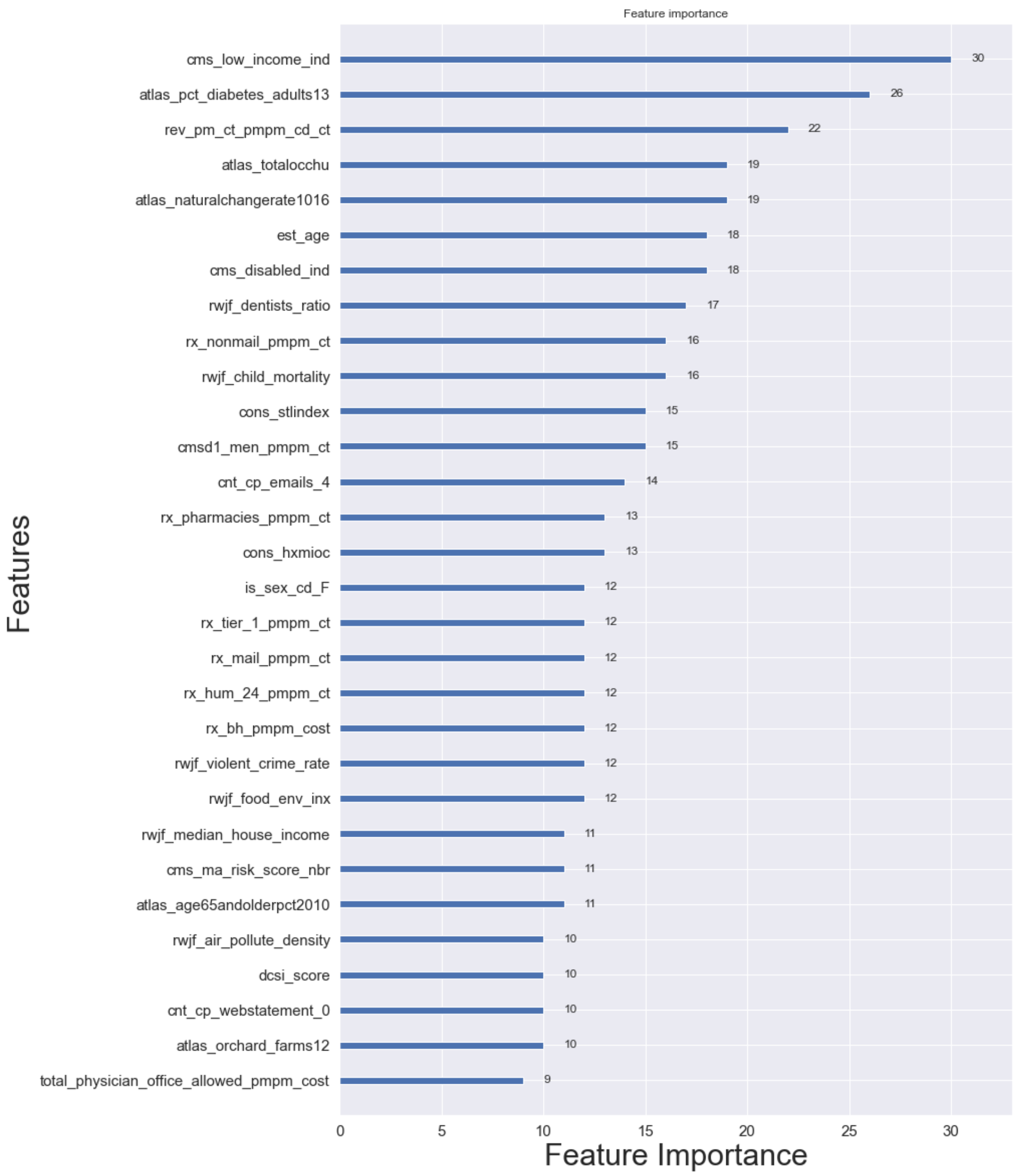

rev_pm_ct_pmpm_cd_ct ≤ 0.075
52.007 gain
yes
no
cmsd1_gus_pmpm_ct ≤ 0.395
95.249 gain
rx_nonmail_pmpm_ct ≤ 0.505
114.134 gain
yes
no
yes
no
leaf 0: -0.000
leaf 3: -0.013
leaf 1: 0.016
leaf 2: -0.001

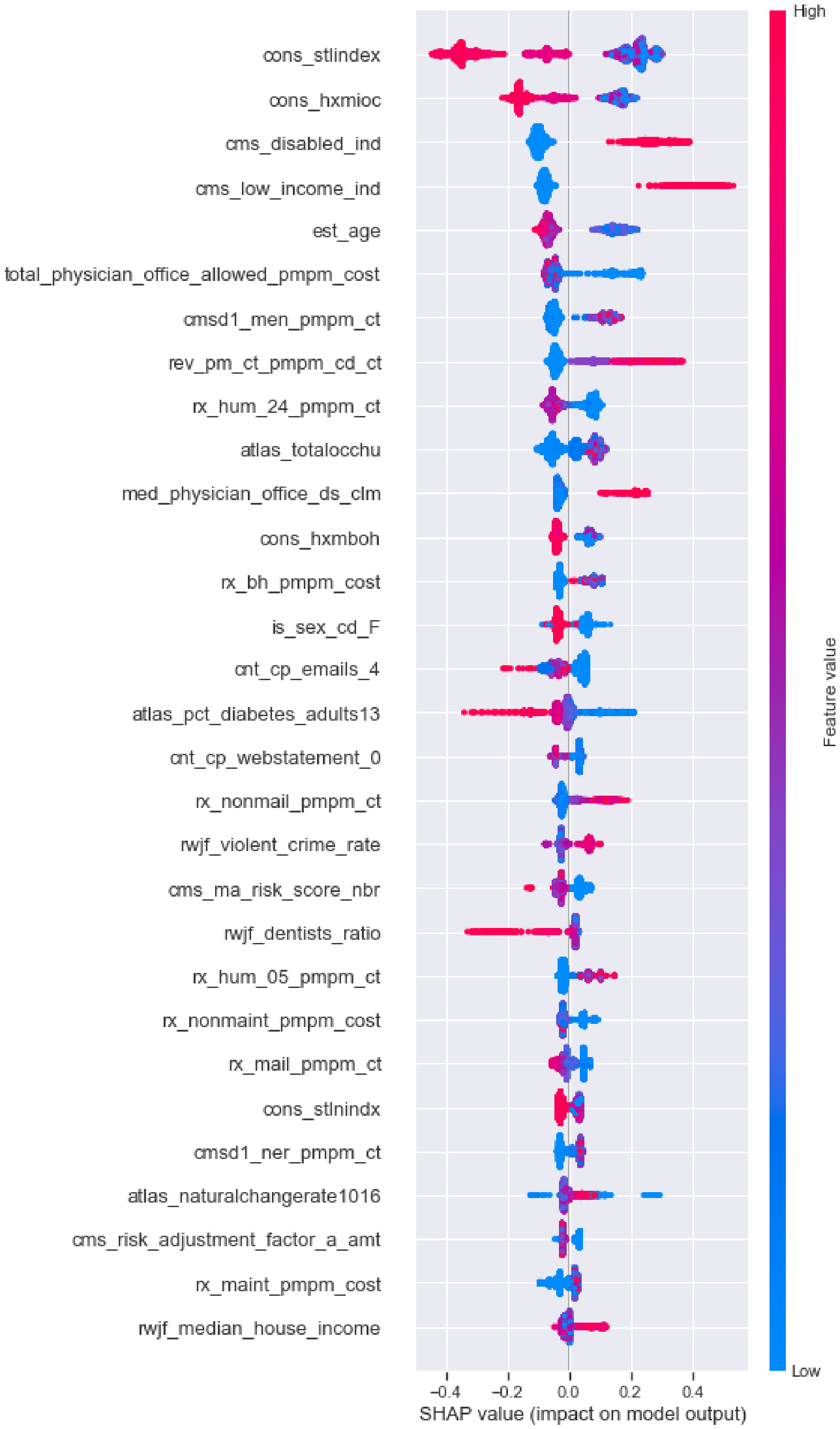

cms_low_income_ind ≤ 0.000
323.909 gain
yes
no
rx_nonmail_pmpm_ct ≤ 1.305
324.931 gain
cnt_cp_webstatement_5 ≤ 13.500
106.069 gain
yes
no
yes
no
leaf 0: 0.002
leaf 2: 0.020
leaf 1: 0.023
leaf 3: -0.049

4 Analysis of Results

4.1 Top Features

Since our model does a good job of identifying housing insecure individuals, we use SHAP and Feature Importance Plot to gain further insights into feature dependencies and model interpretability.





Feature Importance Plot tells us how vital the feature was while the tree-based model was trained and branching occurred. On the other hand, SHAP quantifies the impact each feature had on the model output. This also removes the need to manually inspect all individual trees to have a deeper understanding of the model.

The relationship between housing insecurity and predictors is multi-faceted in nature and SHAP helps us quantify the impact of each feature's value has on the likelihood of an individual having housing insecurity. From SHAP analysis, we get a deeper understanding of many important features.

- Humana constructed indices like 'cons_stlindex' and 'cons_hxmioc' which capture the likelihood of a person having Short Term Loan and person self-monitoring their Health conditions are very impactful in our model. Our model suggests that Housing Insecurity is inversely proportional to these indices.
- Prescription ('RX') features which captures a person's prescriptions and associated costs are important too. For example, we see that 'rx_hum_05_pmpm_ct' is directly proportional to a increased likelihood of feeling housing insecurity. This suggests that individuals who prescribe more number of anticonvulsants(which are used in treatment of seizures and mental disorders) are more likely to feel housing insecurity.
- Environmental Features such as 'rwjf_violent_crime_rate' and 'rwjf_hiv_rate' also have an effect. The model suggests that higher rates of Violent Crime and HIV is closely related to higher chances of feeling housing insecurity.
- 'Atlas' have an impact on the model output as they act as proxies for an individual's socio-economic conditions.
- Claims features ('cmsd1/d2') which indicate the kind of claims made by a person are impactful too. For example, 'cmsd1_men_pmpm_ct' which captures the number of claims per month related to mental disorders is directly related to increased chances of housing insecurity.

4.2 Fairness & Equity in Model

As we are dealing with critical data, we need to ensure that our model is fair and bias-free. To understand if our model is picking up any unwarranted biases, we analyse the model performance in each demographic.

Demographic	AUC
White	0.734143
Asian or Pacific Islander	0.670274
Hispanic	0.697898
Unknown	0.75568
American Indian or Alaska Native	0.685714
Other	0.666822
Black	0.720545

Demographic	AUC
Female	0.7165861
Male	0.7071216

We can see that our model consistently gives comparable performance (AUC ranges from 0.67 to 0.72) across all racial demographics. Additionally, The AUC does not change much between Males and Females of the dataset. This further proves that our model does not pick up any gender biases.

After evaluating the performance and verifying fairness in the model, We evaluated the model on the HOLDOUT dataset and achieved an AUC of 0.7507 with a Disparity score of 0.9868.

5 Strategies & Implementations

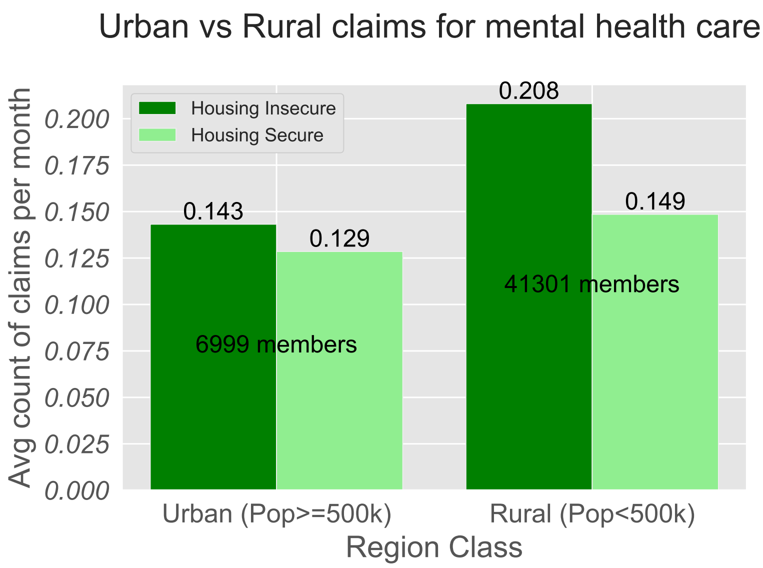
Based on our analysis, we identified key drivers of housing insecurity,

- Regionally, in Rural or non-metro regions, behavioral health related diseases have a great impact on housing security
- Comparatively, in urban or metro locations, housing insecurity is driven by financial issues affecting the member.
- Members classified as housing insecure have disproportionate, increased claims related to branded drugs
- Sophisticated screening tests like CT scan, MRI could drive individuals towards housing insecurity

5.1 Mental Health Care Improvement in Rural Areas

The drivers of housing uncertainty vary with geographic location. Specifically, in metros and large cities, the cause of housing uncertainty seems to be economic in nature, whereas behavioral health factors are strongly correlated with housing insecurity in smaller cities, towns, and other rural regions.

In order to understand this phenomenon deeper, we divide members depending on the population number in their region.



We can observe considerable gap in the claims made by rural members based on whether they are housing secure or insecure. Comparatively, in urban areas the gap in mental health claims is not too distinct. To get a deeper understanding of what the trend looks like we can further split the lower population regions into multiple bins and observe the trend.

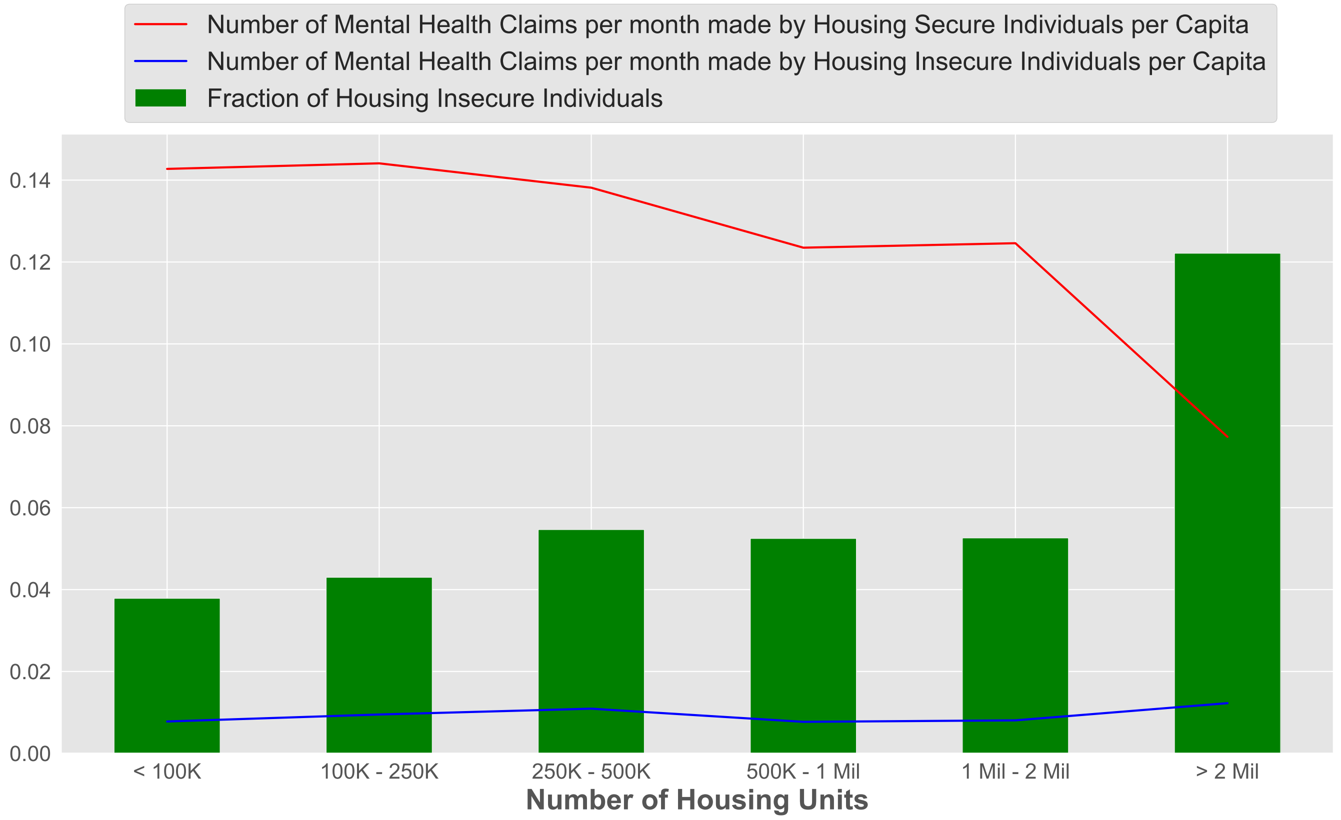


Figure 3: Bar chart showing the fraction of housing insecure individuals for various locations as a function of occupied housing units (a proxy for population).

For the Housing Insecure members we can see that the lower population areas have more or less similar claim rates per capita which are almost twice that of urban centers with (>2 million occupied housing units). This provides a clear direction for action towards lower population areas. In contrast the Housing secure members have similar mental health claim rates across all population centers with a slight increase towards the higher density population centers.

5.1.1 Expansion of Humana Neighborhood Centers

Humana Neighborhood Centers provide several resources to improve community health. Each location conducts special events, programs and activities to help improve physical and mental health, many of them at no cost to attendees. Currently there are 47 centers across the country with 14 centers concentrated in Florida.

We suggest expansion of these center to further locations in order to improve mental health of rural populations. This could be done by with collaboration with local bodies to setup temporary locations.

5.1.2 Temporary locations in collaboration with local bodies

As MAPD members are primary the elderly, we believe that most of the intervention must be in-person. Hence, we prioritize this venture to provide in-person care.

In order to setup temporary locations, the primary requirement is infrastructure. We plan to utilize existing infrastructure of local bodies like churches, town councils or educational institutions in rural regions for mental health care assistance. Further, we can collaborate with these entities to make

Specialized doctors, therapists and medical personnel could conduct wellness drives and create awareness through these temporary centers. This can be expanded across the country to reach as many MAPD members as possible.

5.1.3 Benefit Analysis

The solutions will cover nearly 85% of the Housing Insecure MAPD members. We can estimate the benefit per person in a similar way to how we determined the financial motive. The reduction in medical costs per member can be broadly divided into three parts

- Early Diagnosis
- Consultation
- Treatment and Therapy

Of these we want to avoid the latter two to maximize benefits. The Treatment and Therapy would include prescription costs. The estimates for medical costs and savings are given below.

hi_flag	Rural/Urban	rev_pm_ct_pmpm_cd_ct	rx_overall_pmpm_cost	cmsd1_men_pmpm_ct
0	Rural	0.035801	323.433	0.148522
0	Urban	0.0278907	284.05	0.128527
1	Rural	0.0444051	377.979	0.208085
1	Urban	0.0271646	246.943	0.14319

Category	Cost saved per month per person
Prescription Drugs	$35.7
C.T. scans	$ 28.17
Psychiatrist Visits	$5.9
Total	$69.77

The value of the low population density market is estimated to be at approximately **$120 million** assuming that the percentage of housing insecure members is the same as represented in our dataset. The costs are calculated assuming the costs for average CT scans and Psych visits from external sources.[12][13]

5.1.4 Solutions for Metro Areas

For metro areas the driving factors are usually economic circumstances.

- The long term solution would be to provide affordable housing facilities for the housing insecure population.
- Specialized Medical Plans with reduced co-pay for diseases caused by environmental factors.
- Awareness drives to have simple diagnostic testing such as blood tests at a greater frequency.
- Collaborate with Medicaid and other governmental agencies and invest in affordable housing solutions.

5.2 Generic Drugs Usage

Based on the drug usage data provided, we see that housing insecure people tend to spend 25% more on branded drugs.

Drug Label	Housing Secure	Housing Insecure	Cost Difference
Branded drugs	242.19	309.78	67.59
Non-branded drugs	75.61	43.76	-31.85
Total	317.80	353.54	35.74
Maintenance drugs	214.29	243.11	28.82

Sated differently, housing secure clients spend about 76% of their total prescription drug costs towards branded drugs, while housing insecure clients spend about 87% of their total prescription drug costs towards branded drugs.

5.2.1 Strategies to Pivot to Generic Drugs

United States Food and Drug Administration (FDA) 2021 annual report [14] shows that 90% of all prescriptions in United State are generic drugs and yet cost only 26% of the total drug costs. The housing secure population follows this branded vs non-brand split in prescription drug expenses. However, the branded drug costs for housing insecure population is a significantly higher fraction of their total drug costs. This adds to the economic burden and pushes them further into housing insecurity. Furthermore, vast majority of the drug cost is going towards maintenance drugs. This presents an opportunity to reduce branded drugs usage and in turn reduce the financial burden on the housing insecure population.

The higher usage of branded drugs can be reduced by a two-pronged strategy.

- The first prong of the strategy involves efforts to increase the availability of generic drugs
- The second prong of the strategy involves getting the customers to switch from their branded drug to a generic alternative.

5.2.2 Improve Availability of Generic Drugs

- Humana should actively work towards finding generic alternatives - add generic drugs frequently to the formulary. Many studies have found that generic drugs to be as effective as branded drugs and are substantially cheaper - branded drugs typically cost 4-6 times the cost of generic alternatives [15]
- Since a significant fraction of monthly prescription expenses go towards maintenance drugs, we can assume that majority of the branded drugs fall under the maintenance drugs category. As these branded maintenance drugs have a regular usage, Humana can accurately predict the demand for these drugs, aggregate the demand and place forward contracts with their pharmacy partners. This would reduce the overall costs of the branded drugs.
- Shortage of generic drugs has been shown to increase the usage of branded drugs [16]. Therefore, accurate demand prediction of various drugs will help reduce the usage of branded drugs. In particular, since more than 85% of Humana's drug prescriptions are handled by WalGreen and CVS pharmacies (see Figure 4) for both housing insecure and housing secure categories, the negotiations on forward contracting and pricing in general could be streamlined.

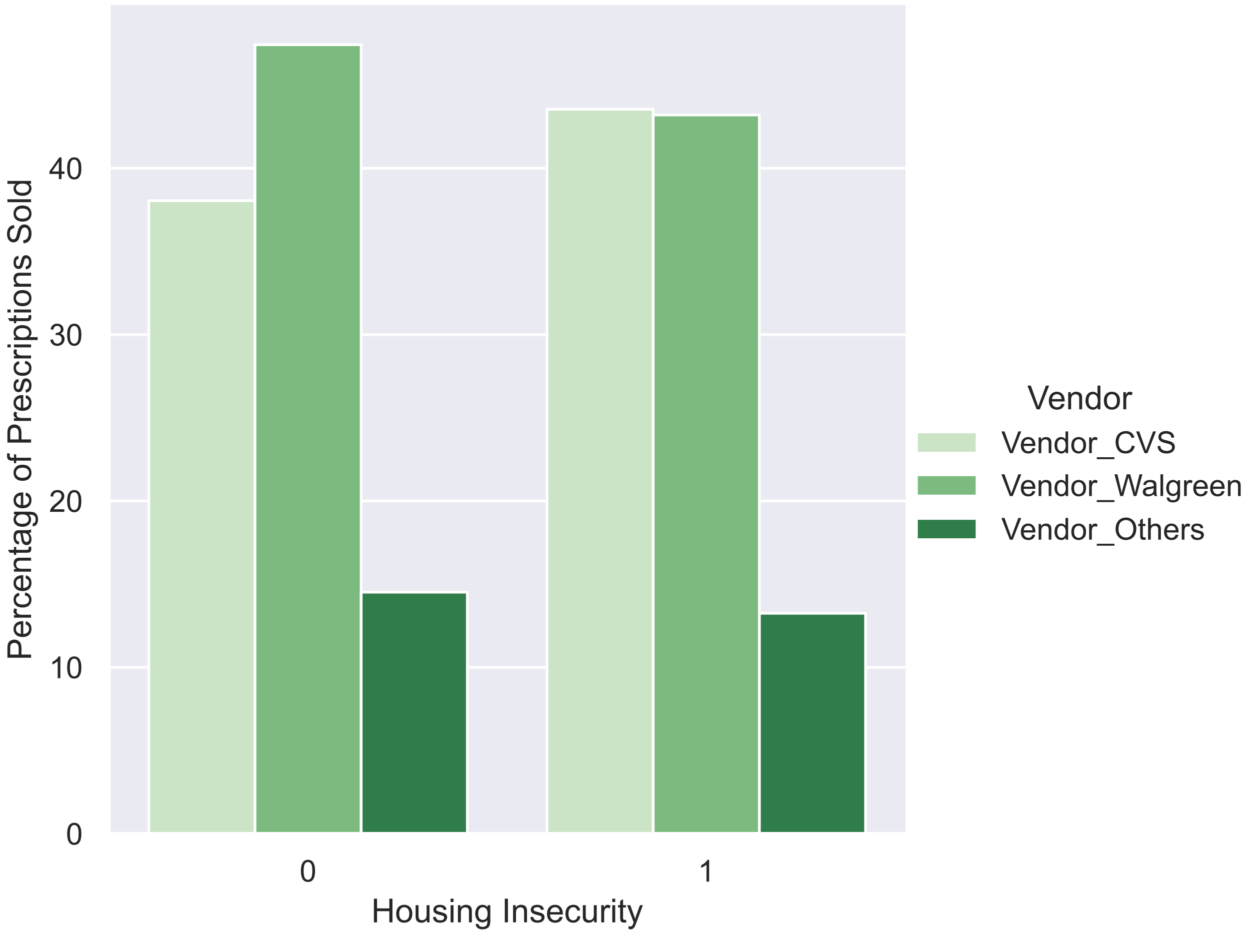


Figure 4: Bar chart showing the percentage of prescriptions fulfilled by various pharmacies for both housing secure and housing insecure groups.

5.2.3 Incentivize Customers to Switch to Generics

Research has shown that several strategies such as reverse payment, product hopping are being used to delay the adoption of generic drugs [10]. The solution to this problem is cosumer awareness. Enhancing the customer trust in generic alternatives and creating purchase preference for generics would go a long way in combating the aforementioned challenges. This can be achieved int he following ways.

- Collaborate with generic drug manufacturers and launch a massive awareness campaign among the Humana and educate them on the efficacy of generic drugs.
- All MAPD subscribers need to pay a monthly premium for Medicare Plan B. So, if Humana could incentivize the customer who switch from their branded drug to generic alternatives could be rewarded with coupons (say 10 to 20 dollars) towards monthly Plan B premium.

5.2.4 Benefit Analysis

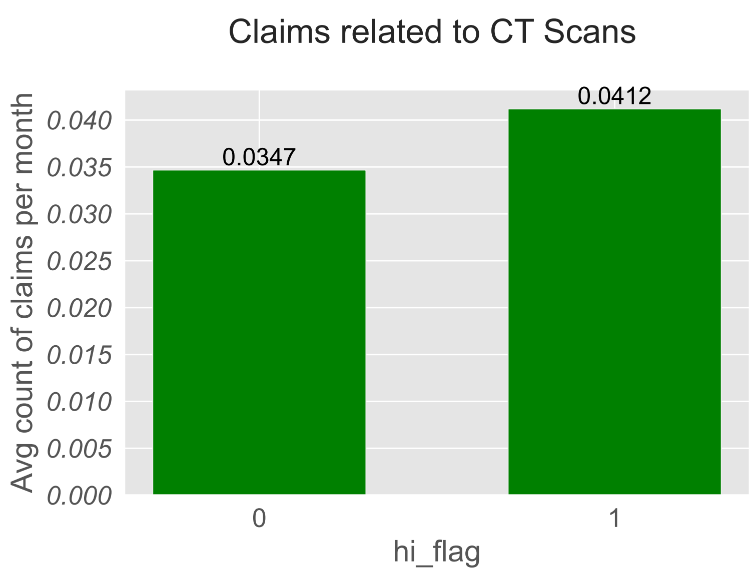
Any reduction in branded drug usage is likely to impact both groups with and without housing insecurity. Therefore, to evaluate the benefit of our strategy, we assume a modest reduction of 25% in generic drug usage across all groups. Furthermore, we conservatively assume that the cost of the replacement generic drug is only one-fourth of the branded cost (recall that the generic drugs tend to be up to 6 times cheaper than branded drugs). Based on these assumptions, we obtain a potential savings of 212.75M$ for the group without housing insecurity and the 12.48M$ on the group with housing insecurity.

The calculations are shown in the table below:

	Housing Secure	Housing Insecure
Current branded drugs cost	242.19	309.78
Reduction in branded drug cost	60.55	77.45
Generic drug replacement cost	15.14	19.36
Total savings per person (in $)	45.41	58.08
MAPD Subcribers (in million)	4.69	0.21
Total savings (in M$)	212.75	12.48

5.3 CT Scans & Radiology Diagnostics Tests

People with housing insecurity are seen to have 18% more claims for CT scans than those without housing insecurity.



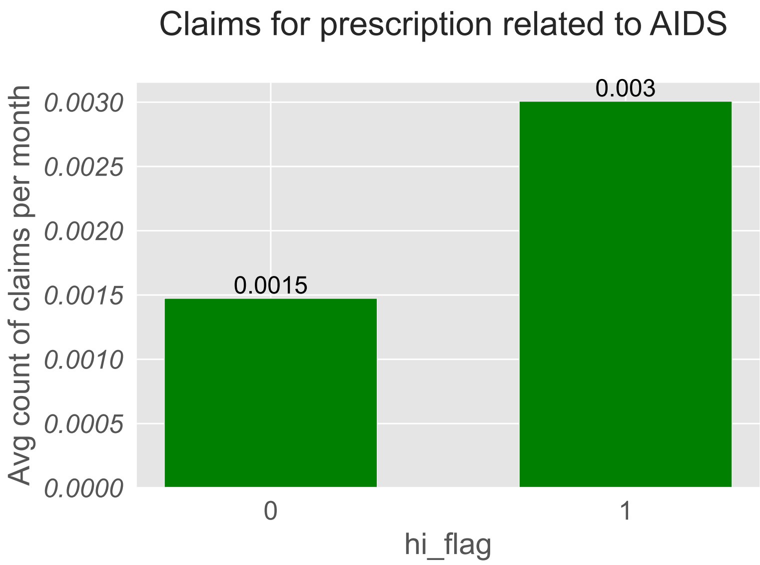
The national average for CT scan cost is $3275[12]. Given the high cost of this common diagnostic scan, a 18% increase in CT scan rates puts the housing insecure group at further stress and pushes them deeper into housing insecurity.

5.3.1 Alternatives and Advancements in the field of medical imaging

Humana can leverage IoT and Machine Learning to automate the preliminary diagnosis. This reduces the workload for per scan for radiologist and overall cost for CT scan. Various groups and investigators have successfully tested these automations in the real world. Hence, by leveraging these these technological developments, Humana could significantly reduce the radiology diagnostic costs.

5.4 Individuals on AIDS Medication

People with housing insecurity are seen to have about 100% higher rate of using AIDS medication. (*rx_hum_05_pmpm_ct* claims in the dataset).



The data shows that Housing Insecure MAPD members have 100% higher claim counts for HIV related prescriptions, the causality indicates that people suffering from AIDS (or using HIV medications) and belonging to a low income category are much more likely to also suffer from housing insecurity.

5.4.1 Long Term Care Solutions

Given that the AIDS medications need long-term usage, Humana can accurately predict the demand for these drugs. Aggregating this information over its insurees, Humana can negotiate future contract deals with pharmacy networks and procure these medications at lower price. This reduces the expense for both insuree and the insurer (in this case Humana). This would ensure adequate treatment/care for AIDS patients, improve their ability to work and reduce their financial vulnerability/housing insecurity.

6 Conclusion

In this study we have examined the factors that are correlated with housing insecurity to identify members of Humana's Medicare Advantage part D plan who may be at risk of Housing Insecurity. We have come up with possible reasons for why each of the factors may be either a consequence of housing insecurity or why it may be the causative reason for Housing Insecurity. At the same time we have explored the trends in the occurrence of Housing Insecurity amongst different groups as well.

During the modelling process we have undertaken an in-depth analysis into each feature and considered categorical features with their imbalances. Further, we identified the top features through non-linear feature selection methods and built a refined, hyperparameter optimized final model. Our final model is a LGBMClassifier which provides an excellent performance on the holdout dataset with an AUC score of 0.7507. We have also evaluated the most important features used by the model to ensure fairness and equity in identifying housing insecure individuals.

On a final note we have recommended solutions that Humana could implement based on the results of the dataset. We came up with customized solutions for each of the problems that the significant features provided insights into. Further, we identified specific target groups for the initial stages of the implementation. We provided benefit analysis of investing into some of the main solutions and supplemented this with an emphasis on in-person contact and care over virtual care considering the target population demographics. The solutions listed in our study focuses on benefits for both Humana and its Medicare members.

We hope that Humana will be able to meet its goal of improving the health and wellness of its members using the results from this study.

References

[1] Road Map to a Unified Measure of Housing Insecurity https://www.jstor.org/stable/10.2307/26696378.

[2] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9149198/

[3] https://www.jpmorgan.com/insights/research/post-pandemic-housing-affordability

[4] Carrere, J., Vásquez-Vera, H., Pérez-Luna, A. et al. Housing Insecurity and Mental Health: the Effect of Housing Tenure and the Coexistence of Life Insecurities. J Urban Health 99, 268–276 (2022). https://doi.org/10.1007/s11524-022-00619-5

[5] Bhat, A.C. et al. (2022) A longitudinal analysis of the relationship between housing insecurity and physical health among midlife and aging adults in the United States, SSM - Population Health. Elsevier. Available at: https://www.sciencedirect.com/science/article/pii/S2352827322001070 (Accessed: October 14, 2022).

[6] Stahre M, VanEenwyk J, Siegel P, Njai R. Housing Insecurity and the Association With Health Outcomes and Unhealthy Behaviors, Washington State, 2011. Prev Chronic Dis. 2015 Jul 9;12:E109. doi: 10.5888/pcd12.140511. PMID: 26160295; PMCID: PMC4509099.

[7] Leifheit KM, Pollack CE, Raifman J, et al. Variation in State-Level Eviction Moratorium Protections and Mental Health Among US Adults During the COVID-19 Pandemic. JAMA Netw Open. 2021;4(12):e2139585. doi:10.1001/jamanetworkopen.2021.39585

[8] Housing Insecurity and Risk of Adverse Kidney Outcomes Tessa K. Novick, Chiazam Omenyi, Dingfen Han, Alan B. Zonderman, Michele K. Evans, Deidra C. Crews

[9] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., . . . Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146–3154.

[10] Jones, G. H., Carrier, M. A., Silver, R. T., Kantarjian, H. (2016). Strategies that delay or prevent the timely availability of affordable generic drugs in the United States. Blood, 127(11), 1398-1402.

[11] https://www.nytimes.com/2022/06/14/headway/medicaid-housing-rent-health.html

[12] https://www.newchoicehealth.com/ct-scan/cost

[13] https://www.khealth.com/learn/healthcare/how-much-does-a-psychiatrist-cost-without-insurance/: :text=The

[14] https://www.fda.gov/drugs/generic-drugs/office-generic-drugs-2021-annual-report

[15] https://www.ehealthinsurance.com/medicare/coverage/medicare-part-d-brand-name-vs-generic-drugs/

[16] https://www.uspharmacist.com/article/factors-involved-in-us-generic-drug-shortages