Humana-Mays Healthcare Analytics 2020 Case Competition

CONTENTS:

Introduction & Problem Statement

According to the American Hospital Association, each year there are 3.6 million Americans who do not get the medical care they need because of a transportation issue. The results: missed doctor appointments, missed diagnoses, increased healthcare costs and an overall less healthy society.

We ask ourselves, what is the role of health insurance companies to help alleviate transportation barriers that have such detrimental effects? Aiming to answer this question, we have conducted a thorough analysis of a Medicare patients dataset provided by Humana, and have built a model to better predict the presence of a transportation barrier for an individual.

Based on our interpretation of the model output, we have determined several immediate solutions that Humana can implement to alleviate this particular social determinant of health for many of its members.

Executive Summary

Producing a high quality, useful machine learning model depends on significant data preparation. To that end, this report contains extensive information on how we prepared the competition data for modelling. As an overview we:

- Removed features with little or duplicated signal
- better captured by other variables
- Added supplementary external, geospatial data from verified public sources
- Imputed missing values while considering data peculiarities
- Engineered additional features that capture meaningful member characteristics
- Used sophisticated feature selection algorithms to retain only the most relevant features

Ultimately, we built a model that achieved ~0.75 AUC via 5-fold cross-validation on the competition training dataset. But what is the model looking for in the data and what is it telling us?

The indicators with the highest impact on whether a member faces transportation issues are disability status, low income, Medicare prescription drug coverage, age, and homeownership. The model also reveals some lagging indicators i.e. factors which occur *after* or even *because* a user is facing transportation issues. One such lagging indicator is the number of ambulance rides, but there are even more subtle ones. Members with fewer physician visits, fewer lab tests, or fewer specialist visits actually have a *higher* probability of facing transportation issues. These are the members who fly "under the radar", the people who *because* they cannot access the healthcare resources they need are often overlooked.

We believe our model pipeline can help identify at least a portion of these members, namely the ones that cannot access healthcare due to transportation issues. Once identified, how can we provide better service to these most vulnerable members of our society? We envision a few ways. First, Humana could partner with rideshare services such as Lyft to ease the financial and logistical challenges of accessing clinical and preventative healthcare. Similar initiatives such as LyftUp already exist in the context of the Coronavirus pandemic. The business impact for Humana could be significant as ambulance rides easily accrue costs in the thousands of dollars.

Another impactful strategic initiative is working with Community Health Centers (CHCs) to establish their telehealth offerings. As late as 2018 a majority of the CHCs offered no telehealth services. And yet, the potential of technology to improve access to healthcare is evident in the fact that many other health insurance companies have greatly expanded their telemedicine efforts.

Finally, our model made us aware of a factor that has gained more attention in recent years - mental health. It is an established fact that mental wellbeing is a key determinant in physical health. Our analysis shows that members with transportation issues are more likely to be under-prescribed for mental health conditions, particularly when they are low-income. This means that, in the long term, these members would represent a higher cost for Humana.

Establishing or expanding preventative measures targeted towards such members could be a beneficial business strategy.

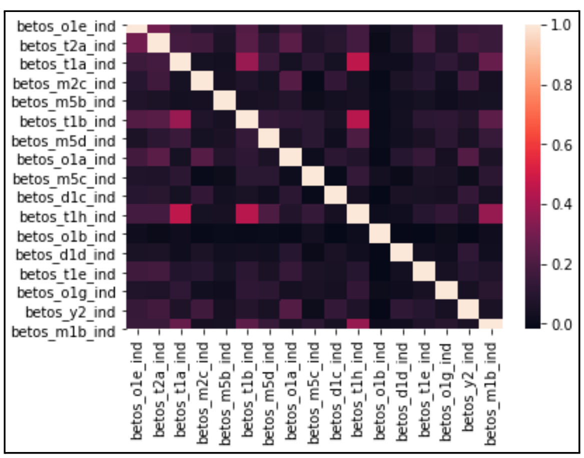
Data Preparation

We spent a significant amount of time getting to know the data, cleaning it, generating features, and even bringing in external data sources. In the following sections we will discuss the main steps taken to prepare the training dataset to fit an interpretable classification model that would predict whether a member will face transportation issues or not.

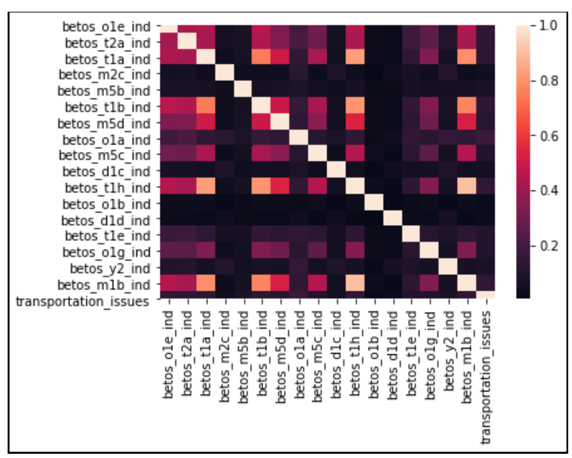
Data Exploration and Cleaning

One of the first things we discovered was that there are 95 columns with no variation. In other words, all values in these columns are the same within a given column, and therefore not useful for predictive purposes. The majority of these are different Medical categories and Rx related. We dropped these 95 columns.

Next, we paid special attention to the many binary features in the data. Our key observation here was that Pearson correlation cannot be used to discover relationships between the many binary features. Instead, we used Jaccard similarity to identify and visualize features that potentially incorporate similar information. To illustrate, let's take the group of BETOS columns. Below, on the left, we have displayed the heatmap for pairwise Pearson correlation between the BETOS binary features. On the right side is a heatmap of the pairwise Jaccard similarity.



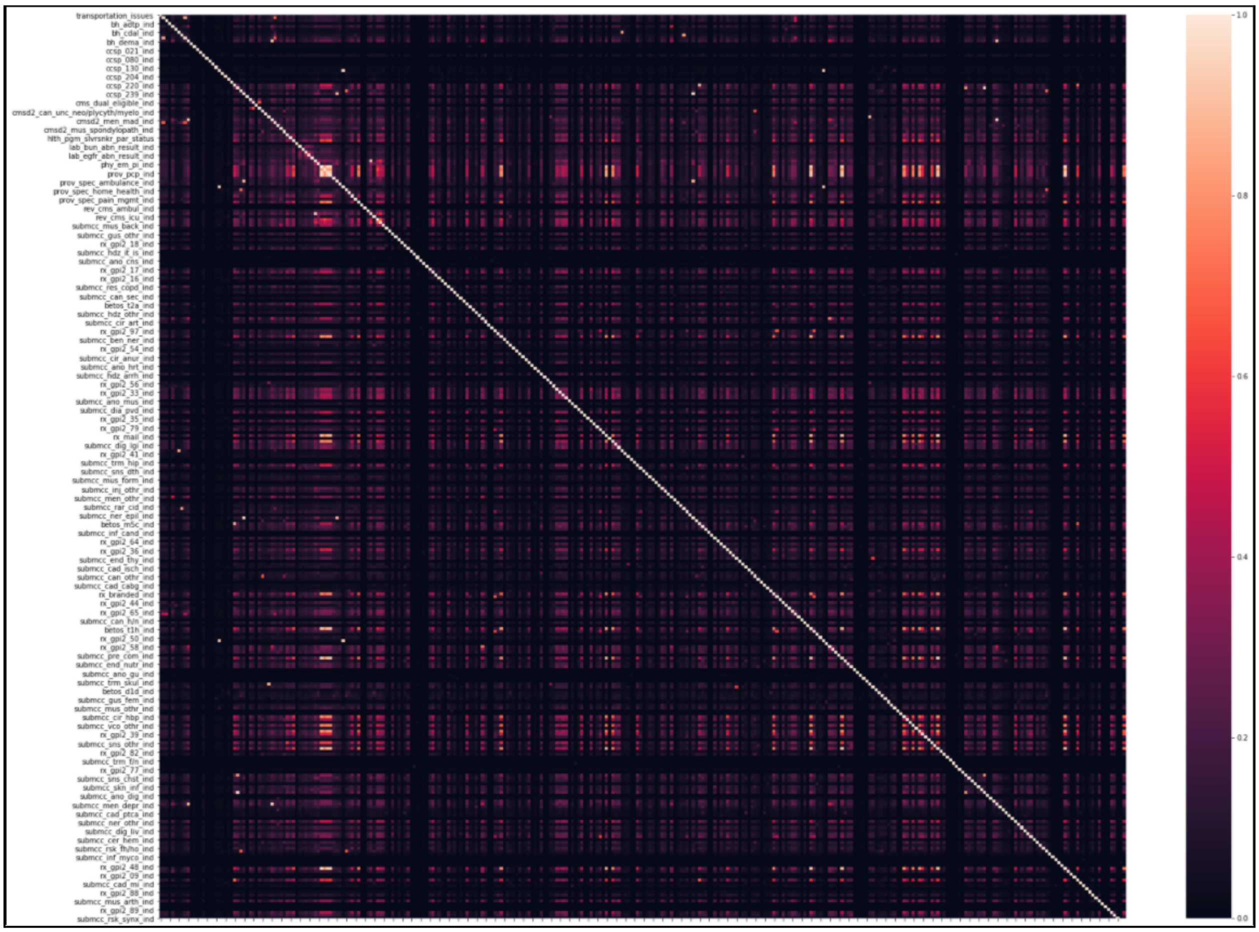
[pairwise Pearson correlation]



[pairwise Jaccard similarity]

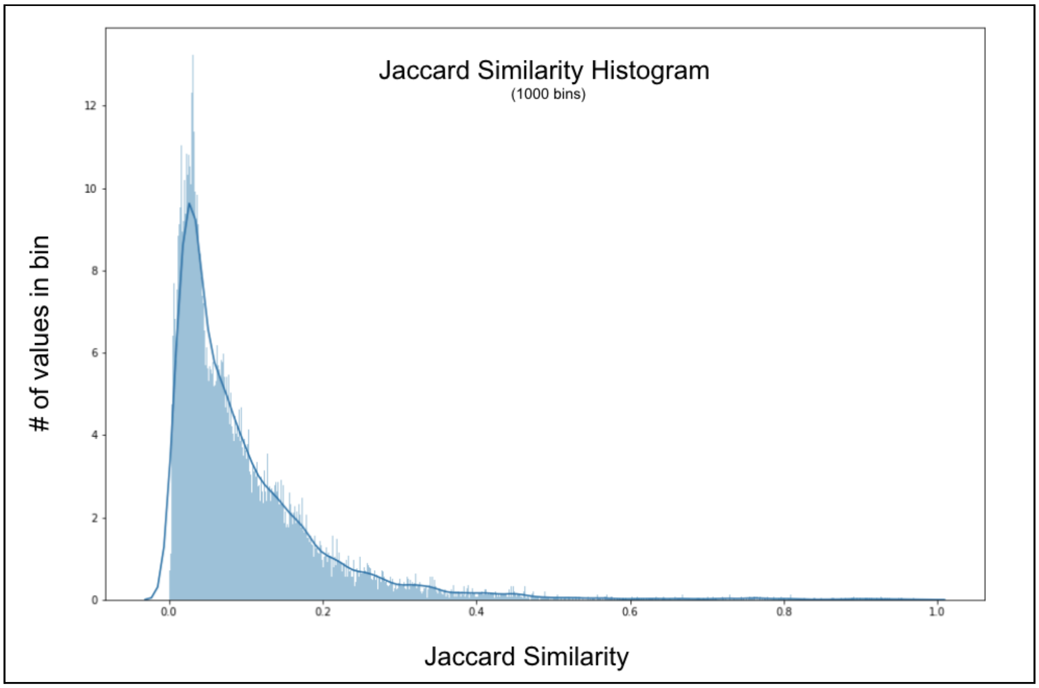
From the two charts it is evident that Jaccard similarity reveals relationships, which are not clear from Pearson correlation. The brighter rectangles signify higher similarity between a pair of features. For example, we discovered that members often get a few types of lab tests done simultaneously. Including all of these features in a model wouldn't be particularly informative and would also incorrectly give certain member aspects undue weight. For example, if all four highly similar binary variables for lab tests are included in a model that wouldn’t tell us anything special and might trick the model into assigning a higher importance than merited to a member getting a lab test.

We extended this analysis to all 312 binary (0/1) features. Below we include the Jaccard similarity heatmap. Note the distinctive “checker-marked” pattern. The many bright spots signify widespread similarity between features in the data.

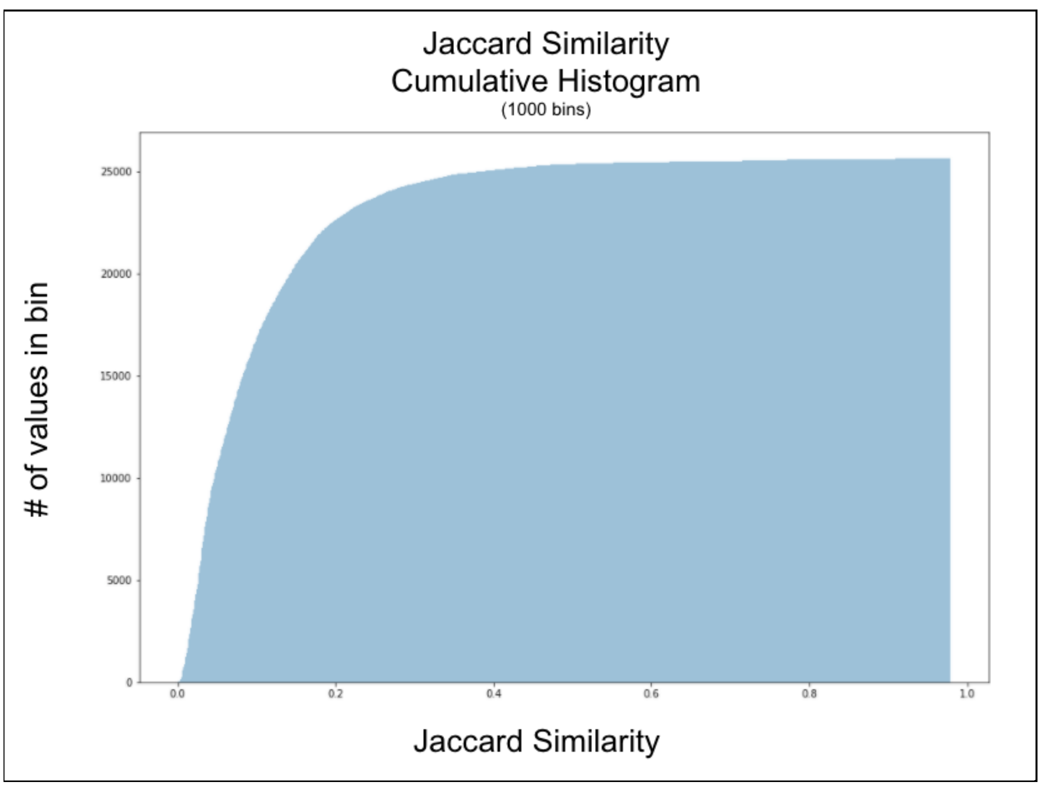


[pairwise Jaccard similarity]
(Not all features displayed on y-axis)

To better understand and represent the above phenomena, we made a histogram of all pairwise Jaccard similarities.



The cumulative version of the above histogram is below:

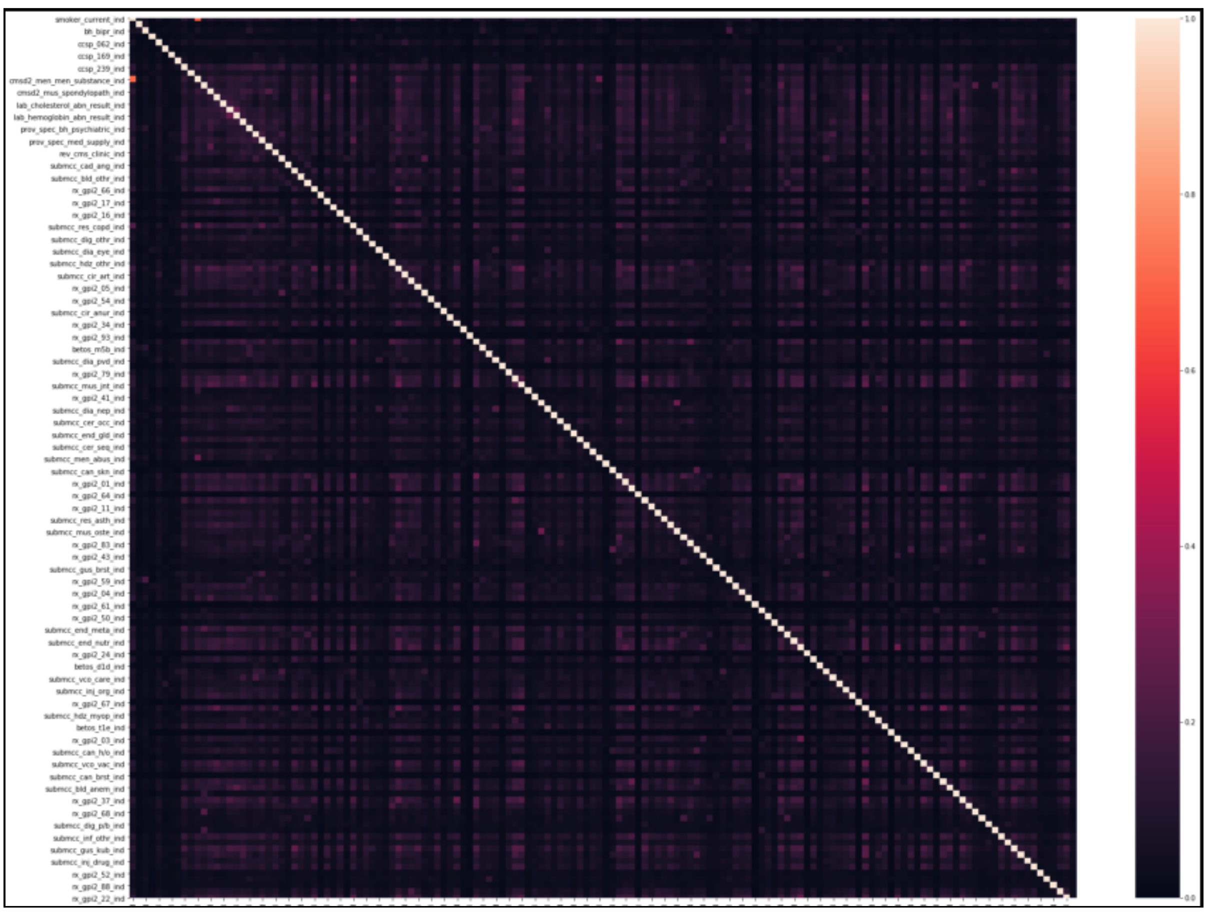


The conclusion we drew from the above charts is that we can safely use a threshold of **0.3** Jaccard similarity as an initial cutoff to limit the useful features we use to train the model. In other words, the majority of the features have lower similarity than this threshold and eliminating features with higher values would address the issues of improper weighing of member characteristics we described previously while still preserving most features.

We did not, however, simply delete all the features in a pair with pairwise similarity above 0.3. That would remove a lot of signal from the data. Instead, we implemented the following heuristic:

- Sort all the feature pairs by Jaccard similarity in descending order.
- Start with the first feature pair.
- Check if the Jaccard similarity is above the threshold of 0.3 and neither feature in the pair has been marked for removal.
    - If yes, then mark for removal the feature that has ***lower*** Jaccard similarity with the target “transportation issues”.
    - If no, then continue with the next feature pair.
- Stop when you’ve gone through all the pairwise similarities.

In the end, we had 92 features marked for deletion. Below we show how the pairwise similarity heatmap looks after we dropped them:



[pairwise Jaccard similarity]

The above heatmap clearly shows that the “checker-marked” pattern we pointed out earlier, while still visible, is much less pronounced. The result of the above procedure helped remove features which bring the least amount of signal not already captured by other features.

We also decided to clean several CMS binary features: cms_disabled_ind, cms_dual_eligible_ind, and cms_low_income_ind. Based on our initial analyses of these three features, we discovered that the sample of disabled individuals had a statistically significant ($p<0.0001$) higher proportion of those with transportation issues than the sample of non-disabled individuals. We saw analogous statistically significant results for dual-eligible or low-income patients.

With the additional knowledge that NaN values in these columns were coded as 0s, we were concerned that our models could lose pertinent patient information if disabled, dual-eligible, or low income individuals were improperly encoded. Hence, we created a criterion for identifying such data points that may be improperly encoded and reverted them back to NaN. This criterion involves the continuous CMS variables in the dataset; we’ll use cms_partd_ra_factor_amt as an example, though all continuous CMS features showed the same trend we will now describe. When we compare the sample of individuals having a NaN value for cms_partd_ra_factor_amt with the sample of individuals having a numerical value for cms_partd_ra_factor_amt, the NaN sample has a statistically lower proportion of disabled, dual-eligible, or low income individuals (i.e. value of 0 for the binary CMS variables). This has no conceivable explanation other than a significant amount of the 0’s being incorrectly encoded. Hence, we decided to revert any 0 value for the three binary CMS features to NaN if any of the continuous CMS features in that row contained NaN. We relied on our imputation technique, multiple imputation by chained equations (MICE), to assign a more accurate value to these data points.

The percent days covered features were also cleaned. A value of 1.1 in these columns implied NaN, so rows were re-encoded as such. For the depression, asthma, heart failure, hyperlipidemia, and diabetes PDC variables, some of the NaN rows were imputed as 0 if there was another feature in the data that indicated that the individual had the disease. For instance, if an individual had a NaN value for pdc_dep (depression) but the BH indicator for major depressive disorder was equal to 1, then that NaN pdc_dep value was imputed as 0. We assumed that these patients were undercovered and wanted to preserve that information in the dataset.

Finally, we dropped a few more columns, which are irrelevant for predictive purposes, like src_platform_cd and person_id_syn. There are several other minor data wrangling actions we implemented, but for the sake of brevity we won’t list all of them. After completing data cleaning on the original dataset, we engineered several new features, which we detail in the next section.

Feature Engineering

The team set out to find lagging indicators - that is, patient characteristics that may have been *caused by* a transportation issue. For instance, someone with a transportation issue may be undercovered. In contrast, an obvious leading indicator would be someone with a disability having a transportation issue.

Using the GPI2 Level Prescription Utilization features (these will be subsequently referred to as Rx features) in conjunction with the SubMCC medical claims features, we set out to determine which patients may be underprescribed. First, we utilized the binary indicators for each SubMCC medical claim and Rx claim to discover useful SubMCC-Rx feature pairs. A useful pair is essentially a particular drug used for a particular disease. In order to prevent our limited medical knowledge from introducing bias into our model, we iterated through each SubMCC-Rx pair and used the following criteria to designated it as a useful combination:

- At least 50% of individuals with the disease (value of 1 for SubMCC binary indicator) were taking the drug (value of 1 for Rx binary indicator)
- The proportion of patients taking the drug was at least 1.5x higher in the sample of patients with the disease than in the sample of patients without the disease

Once these combinations were identified, we created new features by dividing the Rx PMPM column by the SubMCC PMPM column for each combination (but only when both binary indicators had a value of 1, otherwise we would assign NaN), to determine Rx claims per SubMCC claim. An example of one of the new features is rx_gpi2_90/submcc_brn_othr. Rx_gpi2_90 refers to dermatological prescriptions and submcc_brn_othr refers to medical claims for burns (other). Hence, this feature signifies how many Rx claims for dermatologicals an individual made per medical claim for burns (other).

Additionally, we used the BETOS feature, Specialist - Psychiatry (code M5B), in a similar manner to the Rx codes above, to create features describing psychiatry claims per SubMCC claims for the following SubMCC categories under the Mental Health Conditions MCC Category: Substance abuse, alcoholism, depression, schizophrenia, and other.

As a result, we had over 200 new features, so we decided to reduce the dimensionality of the prescriptions per medical claim signal. Thus, for each MCC (not SubMCC) category, we combined each Rx/SubMCC feature by taking the maximum value. This is best illustrated by example. For the MCC category CIR (Other Circulatory), three Rx/SubMCC features were created previously:

- rx_gpi12_33/submcc_cir_anur
- rx_gpi12_36/submcc_cir_hbp
- rx_gpi12_39/submcc_cir_hbp

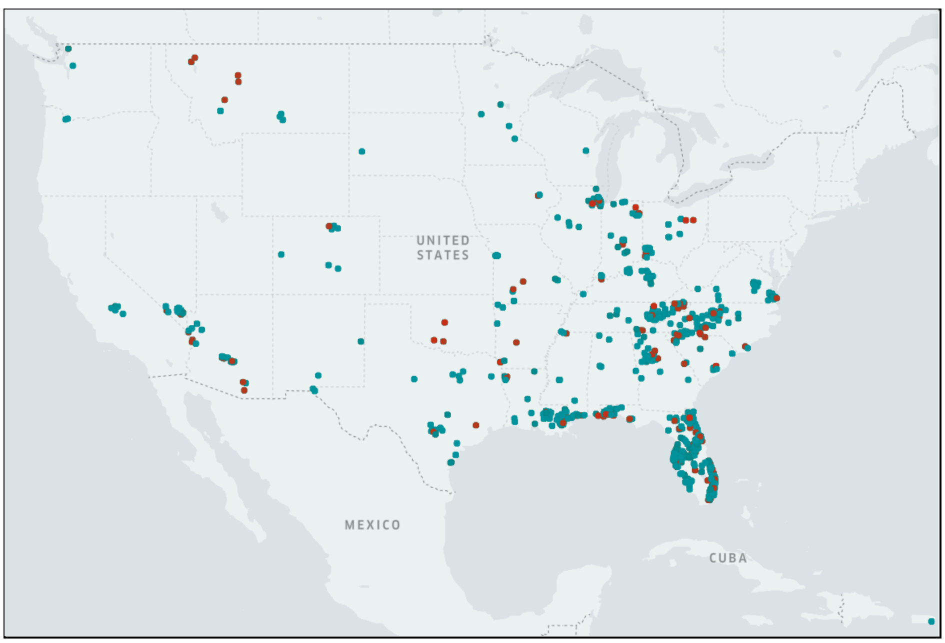
A new column, rx/submcc_cir_max replaced all three columns by taking the maximum value in each row.

Furthermore, the team looked at the HEDIS variables to see if any new features could be engineered. We were interested in the patients who received abnormal lab results in the three following tests: LDL Cholesterol, Diabetic HbA1c, and Diabetic LDL Cholesterol. We made three new columns for each test, respectively. If a patient received a screening for a test but had an "N" value in the control column for that test, that meant the patient saw an abnormal lab result. If a patient had a "Y" in both the screening and control columns, that patient did not see an abnormal lab result. If a patient never got a screening, a NaN value was assigned to that individual.

Finally, we generated 3 features based on the definitions of the RUCC (Rural-Urban Continuum Codes). We parsed these codes to generate binary features for metro/non-metro area, adjacent/non-adjacent to metro area, and rural/non-rural.

External Data

One of the first data exploratory actions we did with the training data was map the locations of subscribers for the subset of subscribers that have location data (33% of the total training data set). We wanted to see if there were any clear geographical indicators of subscribers that face transportation challenges versus those that do not. Below is our geospatial visualization using kepler.gl:



The red dots signify subscribers with transportation issues and the green represents the majority class with no transportation issues.

We noticed pockets in rural Montana and Oklahoma with patients indicating transportation issues. Looking at major cities, it was difficult to see any clear patterns, but we hypothesized that there may be a relationship between subscribers in food deserts, areas with a lack of public transportation, and/or minimal healthcare facilities nearby that may indicate subscribers facing the target social determinant of health, transportation issues.

In order to test this hypothesis further, we used two primary external data sources:

- USDA Food Access Research Atlas: incorporates food environment factors such as store/restaurant proximity, food prices, food and nutrition assistance programs, etc. at the census tract level. USDA provides data at the census tract level regarding whether certain areas are food deserts. This can indicate areas where access to food and even healthcare is problematic and thus contribute to a patient having transportation issues.

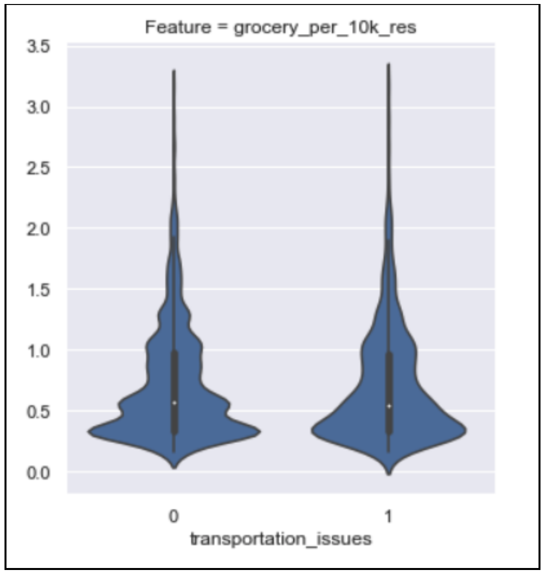
- Measure of America by the Social Science Research Council: contains information on the number of primary care physicians per 100,000 and other health indicators at the county level. Using the data collected by the Social Science Research Council at the county level, we can capture some key health stats, such as the amount Primary Care Physicians Per 100,000 in order to judge access to healthcare for a given location.

We mapped these to the competition dataset by using a dictionary of the FIPS county codes from the Housing and Urban Development website. With these external data sources, we were able to generate the following features:

- HUNVFlag- proportion of tracts in a given zip code that indicate issues with vehicle access
- lahunv1- number of housing units in a given zip code with low vehicle access
- lapop1- population count beyond 1 mile from supermarket
- PCP_100k - # primary care physicians per 100,000 in a given county
- FoodInsecure_per - % of pop indicating food insecurity in a given county
- SNAP_Benefits_per - % of pop enrolled for SNAP benefits in a given county
- grocery_per_10k_res - # of grocery stores per 10,000 residents in a given county

It's important to note that many of these data sources include data at the census tract level or county level, whereas the finest granularity in the competition dataset is at the zip code level. Therefore, we've aggregated via sums or averages depending on what makes sense for a given variable.

Ultimately, after running feature selection algorithms among the hundreds of features initially included, we ended up including only one external geospatial feature in our final model which was the grocery_per_10k_res. This indicated food deserts may be an indicator of areas where there may be low access to food and possibly healthcare. The violin plot below shows that there are indeed two slightly different distributions for subscribers that indicate transportation issues versus those that do not.



The distribution plot on the right has a lower mean (based on the small white dot along the middle axis). This reflects areas with fewer grocery stores (indicative of food deserts or rural areas) and may help predict subscribers with transportation issues.

The other external geospatial features were not as important in improving model performance. It is important to note that the training data contains zip codes for ~33% of the data. Therefore, we were able to supplement only that portion of the data with the additional data described above.

Data Imputation

Some models in the sklearn library do not support NaN values in the training data. To address this issue, we needed to handle missing and null data in the competition datasets. We implemented a version of Multiple Imputations using Chained Equations (MICE) in sklearn. The multivariate imputer estimates each missing feature from all the other features in the dataset. More specifically, we used a Bayesian Ridge estimator to impute the missing values and made sure the model imputed the features with the least number of missing values first before proceeding to impute the features with more missing values.

Feature Selection

After cleaning the data, generating features, and joining with external data sources, we ended up with 623 features. In order to improve model performance, interpretability, and training speed, we resorted to Recursive Feature Elimination (RFE). In this algorithm, a Gradient Boosted Classifier is initially trained on all features. Then, up to 10% of features with the lowest feature importance are dropped from the data. This process is repeated multiple times until a pre-selected number of features are left. For our purposes, we kept the top 100 features.

The most important features were cms_low_income_ind, est_age, and cms_disabled_ind. Interestingly enough, we also noticed a few highly ranked features related to ambulance usage: total_ambulance_visit_ct_pmpm, med_ambulance_visit_ct_pmpm, and betos_o1a_pmpm_ct. At first, we noticed that none of the external data features made the cut, and it was because only 33% of the training data had location data. We tried running RFE on the subset of data that had location data to see if the geospatial external features would then be important. Finally, we saw that the number of grocery stores per 10k residents became a more important feature. We thus decided to include this feature in addition to the top 100 features generated by RFE before in our modelling efforts.

Modelling & Model Evaluation

We started our modelling efforts by using H2O AutoML to get an idea of what types of models would perform well and where we should focus on hyperparameter tuning. Here are the top performing models:

model_id	auc	logloss	aucpr	mean_per_class_error	rmse	mse
StackedEnsemble_AllModels_AutoML_20201007_181110	0.744283	0.367857	0.368639	0.338582	0.332644	0.110652
StackedEnsemble_BestOfFamily_AutoML_20201007_181110	0.743879	0.36816	0.367181	0.328887	0.332825	0.110772
GBM_1_AutoML_20201007_181110	0.742205	0.366604	0.35948	0.331029	0.332618	0.110635
GBM_2_AutoML_20201007_181110	0.74077	0.367514	0.356436	0.328444	0.333149	0.110988
GBM_3_AutoML_20201007_181110	0.738785	0.368216	0.356343	0.332019	0.333419	0.111168
XGBoost_3_AutoML_20201007_181110	0.73698	0.36901	0.354796	0.333153	0.333713	0.111364
GLM_1_AutoML_20201007_181110	0.735334	0.369838	0.34779	0.328348	0.334099	0.111622
GBM_5_AutoML_20201007_181110	0.735197	0.368531	0.358372	0.331614	0.333176	0.111006
GBM_4_AutoML_20201007_181110	0.732787	0.37177	0.344853	0.335047	0.335141	0.112319
DRF_1_AutoML_20201007_181110	0.728261	0.372565	0.346933	0.344013	0.335063	0.112267

The two best performing models are stacked ensembles, which would be difficult to interpret. Given the low AUC difference between them and the next models, we decided to focus on tree based models and ignore the stacked ensemble models.

Concerned that our model may be biased towards individuals without transportation issues given the imbalanced nature of the dataset, we decided to replicate each row of the minority class (presence of a transportation issue) six times. Due to this data manipulation, we use the following formula to correct probabilities produced by a model trained on the upsampled data:

$$p(\mathbf{x}) = \frac{p_s(\mathbf{x})\,O}{O_s - p_s(\mathbf{x})\,(O_s - O)}$$

- $p_s(x)$: probability of transp. issue from upsampled data
- O_s: Population odds, sampled data
- O : Population odds, original data

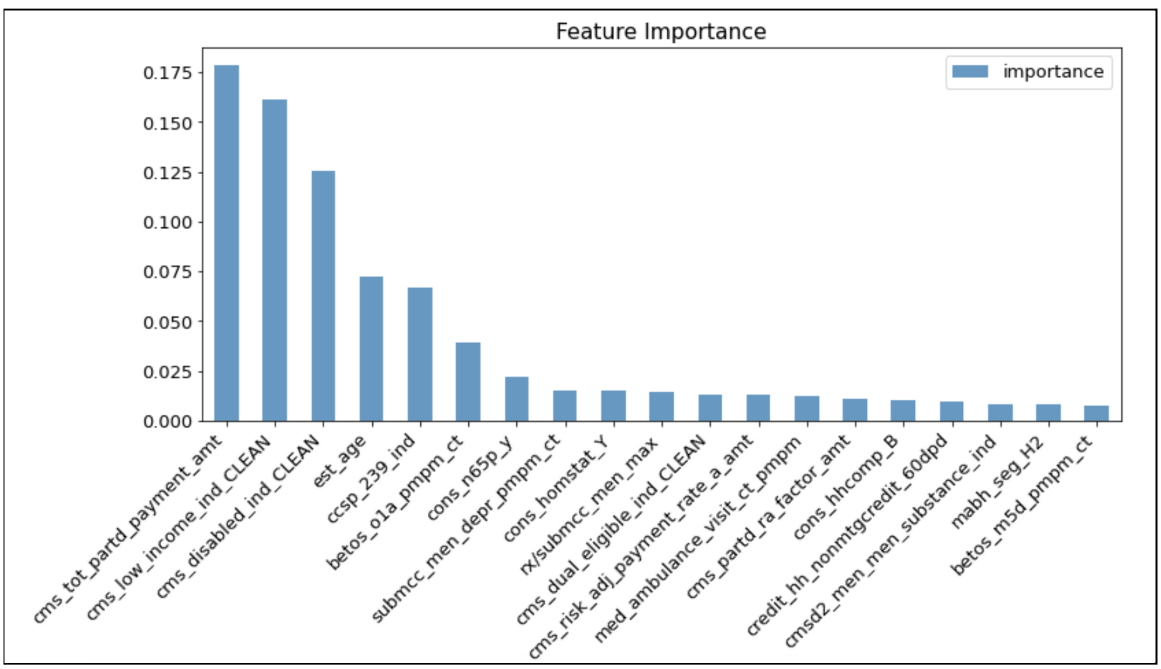
Our approach to hyperparameter tuning and model evaluation was to use 5-Fold Cross Validation and evaluate our models based on the average AUC across the five folds. To ensure that no data points were present in multiple folds (i.e. data leakage), we conducted the aforementioned upsampling within each CV partition. Our modelling results are summarized:

Model	**5-fold Cross-Validation AUC**
Gradient Boosted Trees	0.75
Random Forests	0.73
CatBoost	0.65
XGBoost	0.54

Clearly, the best-performing model is the Gradient Boosted Trees. The best hyperparameters were 125 iterations (trees) with a maximum depth of 3. We continue the next sections with analysis of feature importance and interpreting the model results.

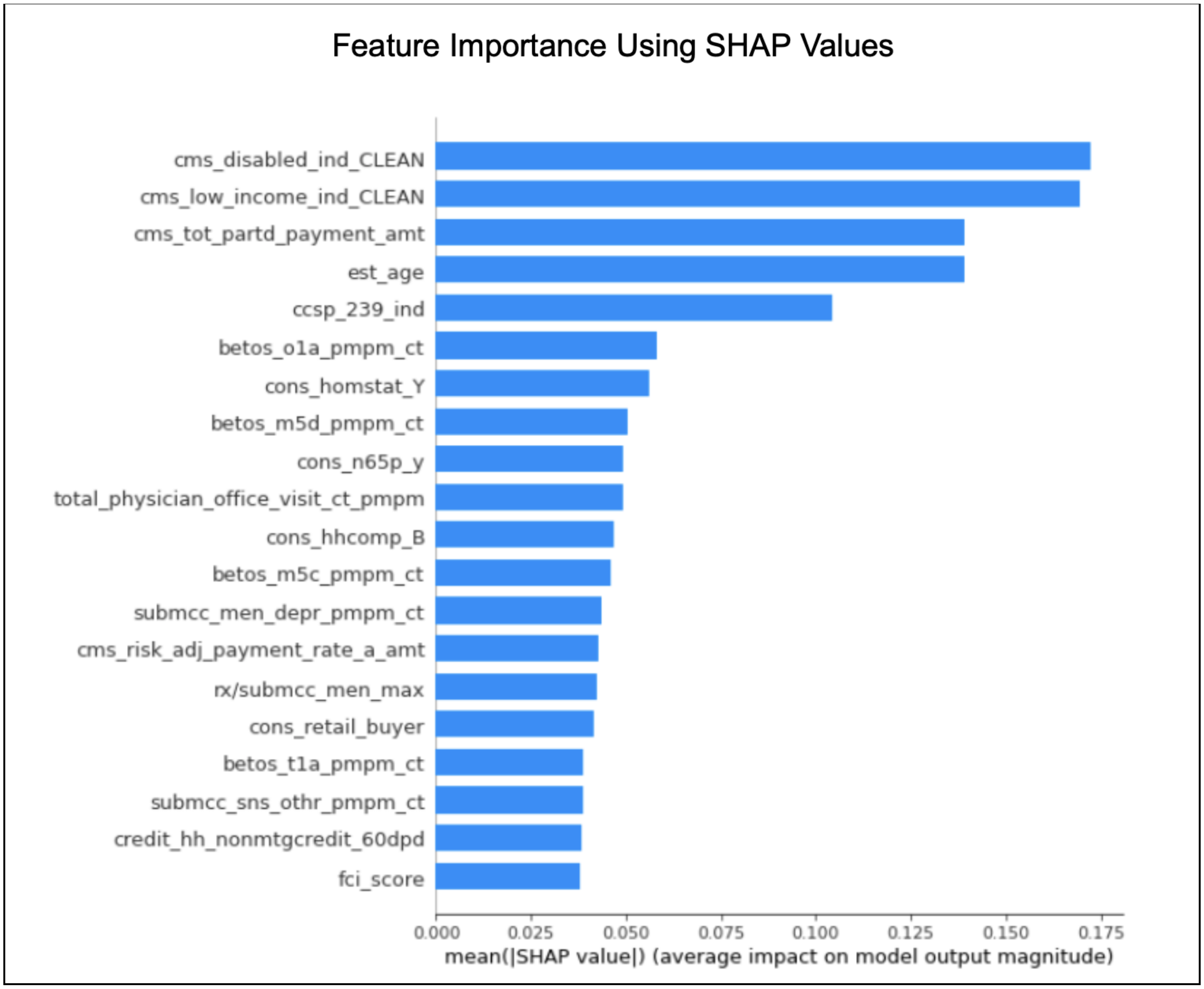
Feature Importance

Out of the box, most trained model objects from the sklearn library provide feature importance values. These indicate the magnitude of influence that a feature has on the model predictions. The values for the top features from our final model are in the chart below:

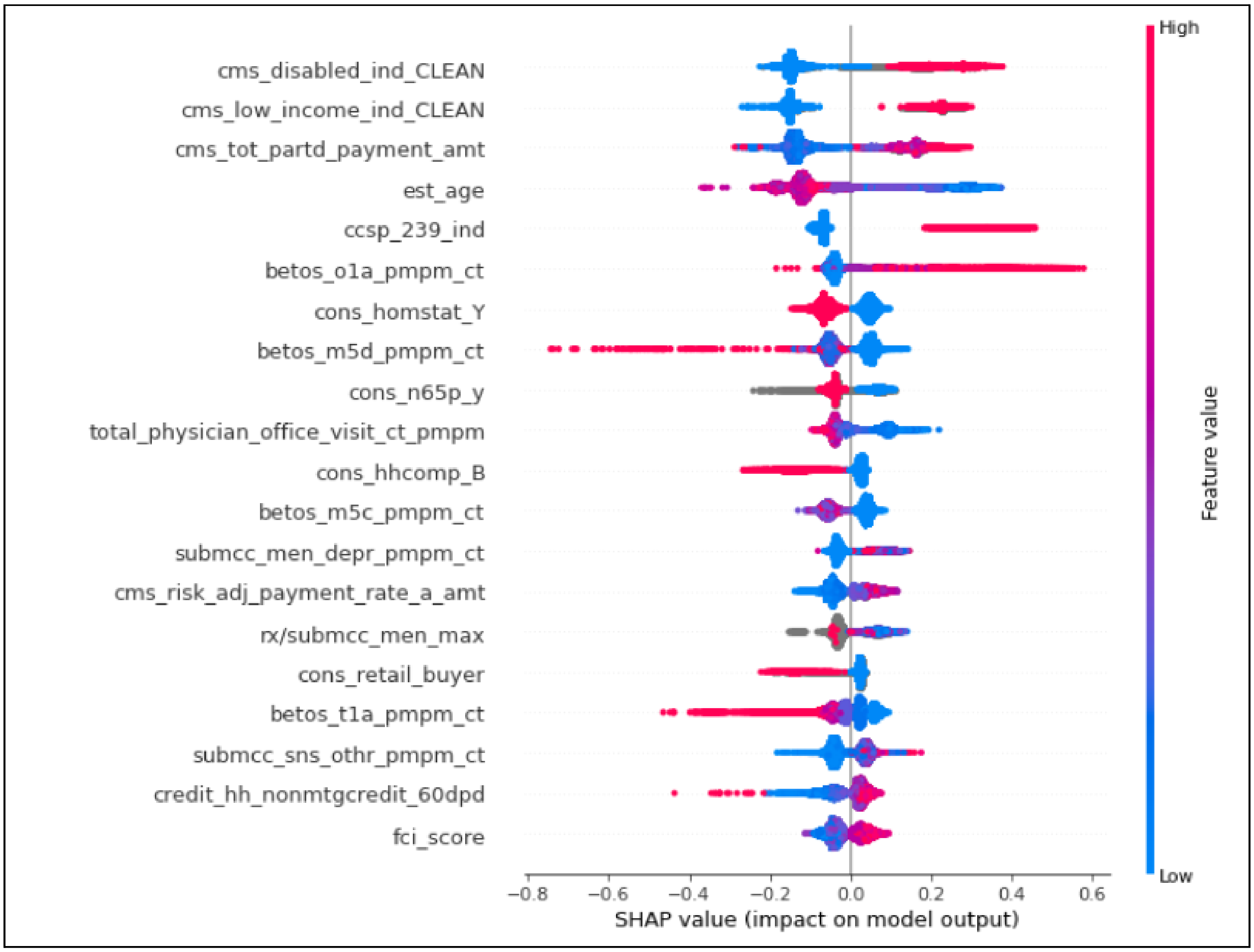


[Feature Importance from GBT trained model object]

However, there are known issues with the above mentioned feature importance estimates. More specifically, these tend to place higher weight on numerical features over categorical ones. Since our final training data contains many binary features, we explored alternative feature importance estimates and ended up using Shapley values. The results are shown below:



SHAP (SHapley Additive exPlanations) is a game theoretic approach to explain the output of any machine learning model. Essentially, features are treated as “players" competing with each other in trying to impact the output variable (i.e. likelihood an individual faces transportation issues). SHAP values measure the positive and/or negative impact of each feature on the prediction. The plot above shows the top 20 predictors that impact the outcome solely based on the magnitude. Evidently, the SHAP feature importance places appropriate weight to binary indicators such as the cms_disabled_ind_CLEAN and cms_low_income_ind_CLEAN. The cms_tot_partd_payment_amt which is a continuous variable dropped from the most important using traditional feature importance methods to third most important, so it is evident that using the SHAP method reduces the bias towards favoring continuous variables. Below is a slightly different visualization of the feature importance that highlights the impact via SHAP values but additionally indicates the direction each feature impacts the final outcome.

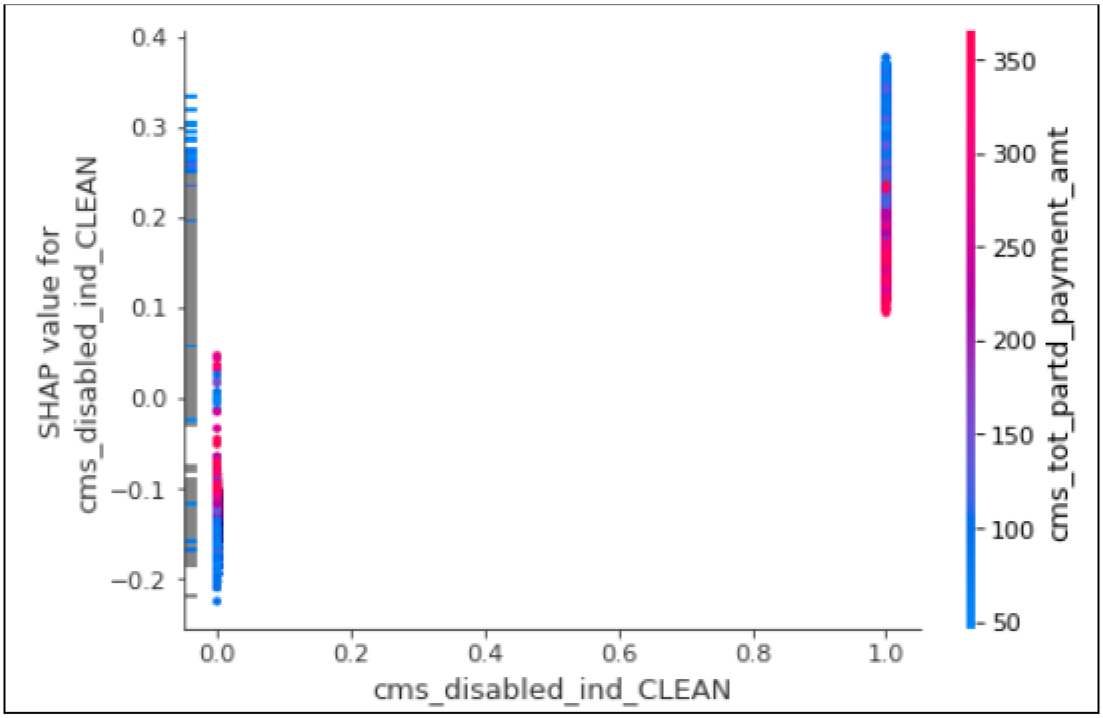


In the figure above, the top 20 features in the model are shown. The red color indicates that particular instance has a high feature value and blue indicates that low feature value. For binary variables such as cms_disabled_ind_CLEAN, red would indicate a value of 1 and blue would indicate a value of 0. There is a range of SHAP values for each variable because of the interactions of that particular feature with all the other features in the model. In the next section, some of the features are explored further by looking how each feature impacts the output variable via SHAP values.

Impact on Response - SHAP Dependence Plots

In order to understand the impact of each individual predictor on the target - the probability a given subscriber would face transportation issues - we decided to look at the SHAP dependence plots to visualize the relationship between the predictor and the response variable. One added benefit with SHAP dependence plots is that they also incorporate the next best predictor that interacts with the predictor in question the most. This way it shows the relationship between the main predictor as well as interaction with other variable(s) on the outcome to draw valuable insights. Positive SHAP values as indicated on the y-axis indicates a “positive” impact on the prediction. In other words, larger positive SHAP values represents an increase in odds for a given subscriber of facing transportation issues. On the other hand, larger (in absolute terms) negative SHAP values represent an decrease in odds for a given subscriber of having a transportation barrier. In order to focus on the values of variables that are given, we analyzed the feature dependency plots based on the un-imputed data.

The first plot looks at the top feature which is cms_disabled_ind_CLEAN:

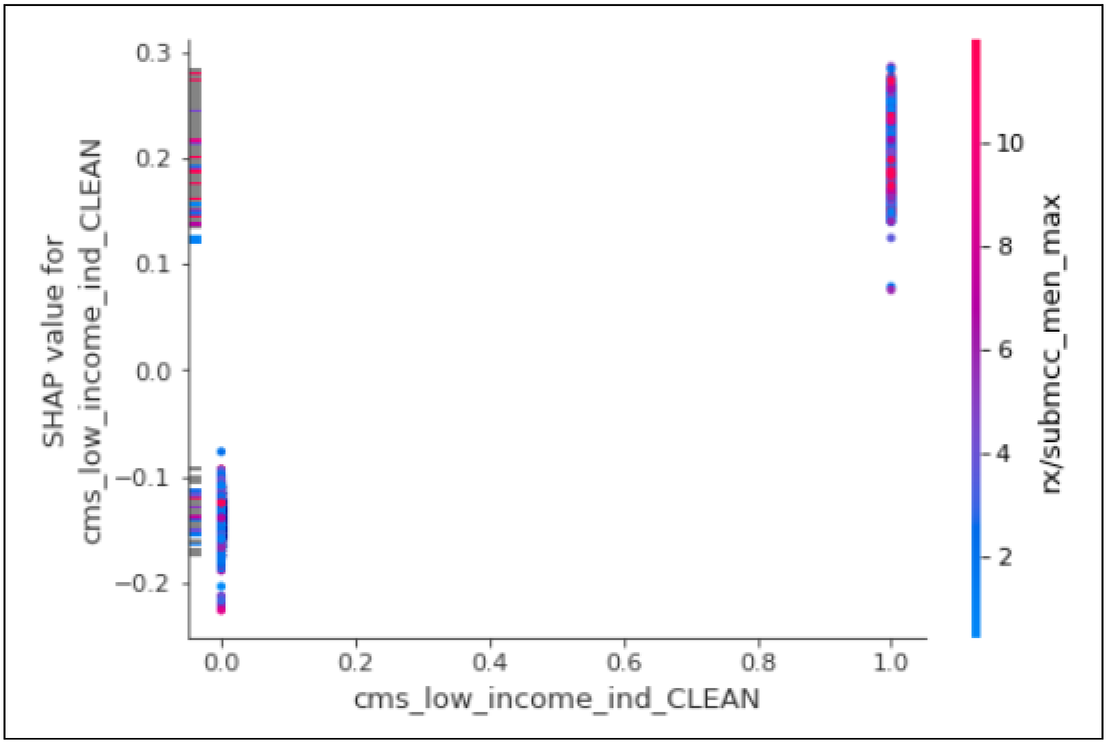


This variable signifies if a given subscriber is on disability. As you can see above, subscribers with a disability have a larger SHAP value and thus a higher probability of facing transportation challenges. The tick marks on the y-axis represents the subscribers in the training dataset for which this value was blank. The color represents the value of the most interacted with variable which in this case is cms_tot_partd_papyment_amt. The subscribers that are on disability and have a lower part-d prescription payment amount have a larger probability of facing transportation issues. This may indicate that patients that are on disability but are spending less on prescriptions might not have access to healthcare facilities, and thus are not receiving the appropriate prescriptions. If this is the case, the disability indicator is a leading indicator and the amount spent on prescriptions is a lagging indicator.

Knowing that individuals on disability are more likely to face transportation hurdles, so much so that they may not be receiving the proper prescriptions due to lack of access can lead to some useful actions. For these types of individuals, it may be necessary to offer prescription delivery services or even other modes of transportations (i.e. rideshares, taxi vouchers) to access their healthcare sites for prescription fulfillments.

Interestingly, those not on disability and with a lower part-d prescription payment amount have a much smaller probability of facing transportation issues. It may be the case that non-disabled individuals are not as sick as those on disability, and thus may not need to spend as much on prescriptions.

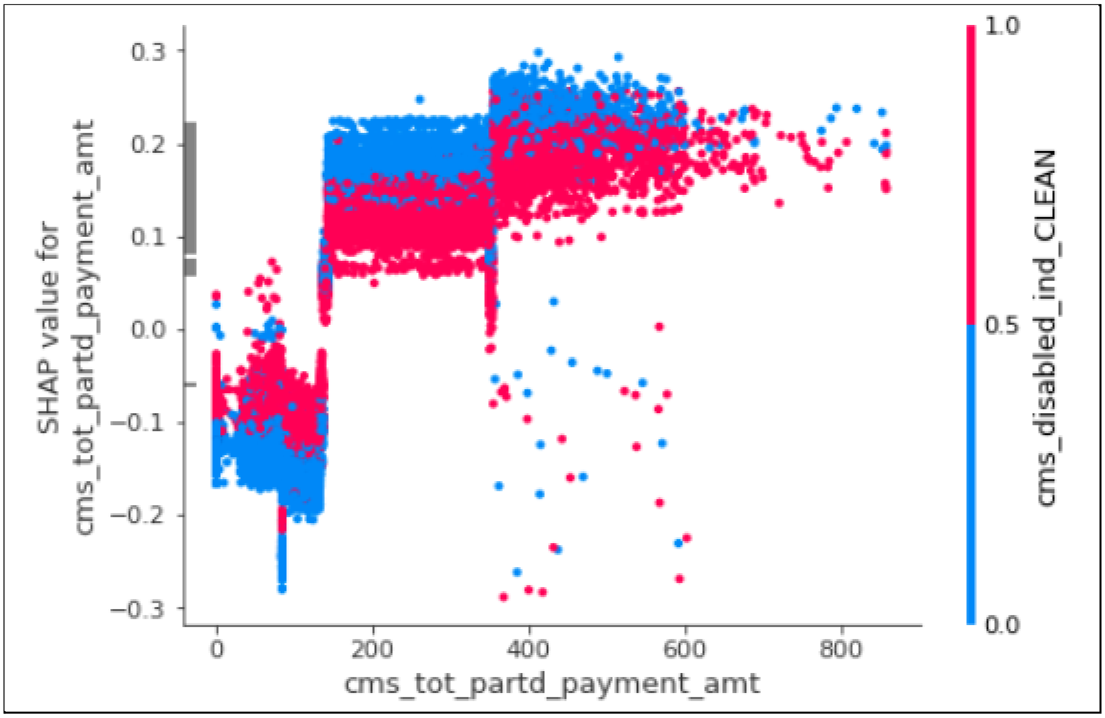
The next top feature in the model is cms_low_income_ind_CLEAN:



This feature shows whether an individual has a low income. As per expectations, individuals with a low income are more likely to face transportation issues. The other feature that seemed to interact the most with this low income feature in predicting transportation issues is an engineered feature (rx/submcc_men_max) which represents prescriptions per mental health disease claims. A low rx/submcc_men_max would indicate that this individual suffering from a mental health disorder takes fewer prescriptions for this disorder as compared to their peers. This can possibly be a lagging indicator of facing transportation issues, as these individuals are not gaining access to their appropriate prescriptions. It is more likely that this inability to fulfill their prescription needs is due to low income and thus access to healthcare.

In order to relieve these subscribers of their transportation issues that they may be facing, it'll be important to possibly subsidize transportation with rideshare or taxi credits. This might be another added reason why those subscribers with low income tend to have difficulty accessing healthcare facilities. In order to alleviate this, it may be important to look where healthcare facilities are located in order to ensure equitable access for all income classes.

The following dependency plot looks at the SHAP values of cms_tot_partd_payment_amt:

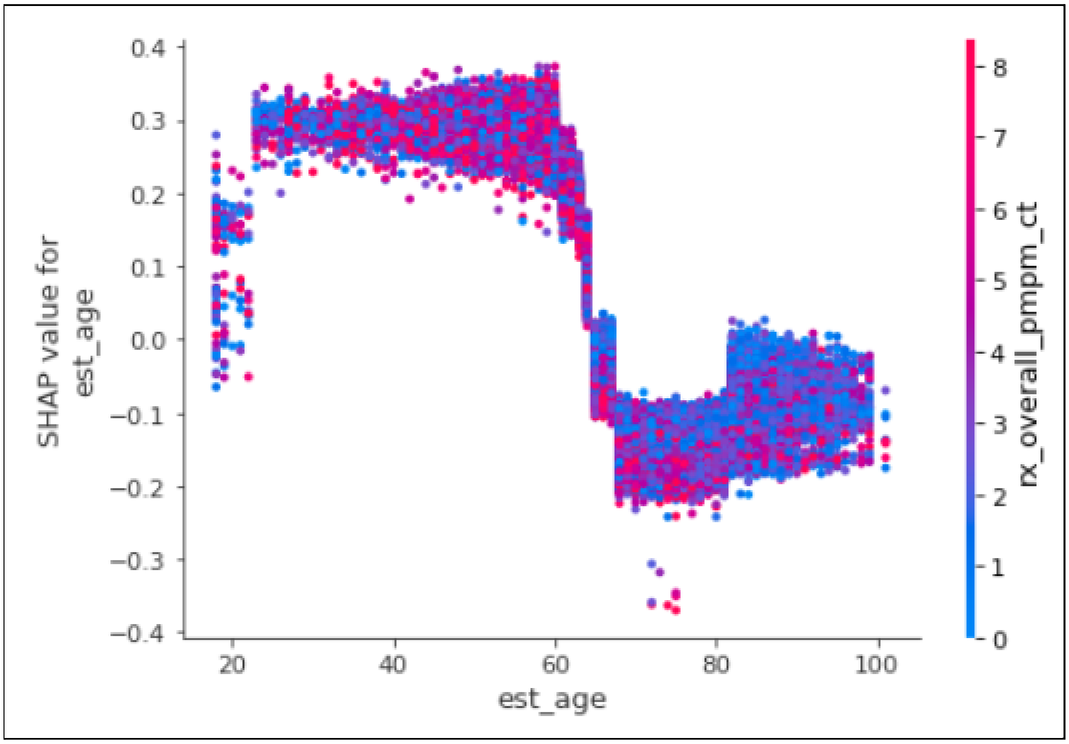


o

This variable essentially reflects the total cost of medications for those enrolled in the Medicare Part D program, which is insurance for medication needs. Based on the SHAP dependency plot, the more spent on prescriptions via Part D Medicare, the higher the likelihood of suffering from transportation issues. Assuming that the amount spent on medications are reflective of how frequent or how expensive pharmaceutical treatment is for a given condition, this can be used to determine how sick an individual subscriber is. Thus, this implies that subscribers that are sicker are more likely to face transportation issues, as these individuals may be incapacitated from accessing public transportation or their own vehicle. The variable that is most closely linked, cms_disabled_ind_CLEAN, was discussed earlier, but this reassures that relationship.

Humana can possibly use the Part D medicare total payments as a proxy indicator to determine how sick and possibly incapacitated a particular subscriber is, and thus can help identify individuals facing transportation issues. Further analysis has to be done to confirm this relationship.

This dependency plot looks at the relationship between est_age and the outcome:



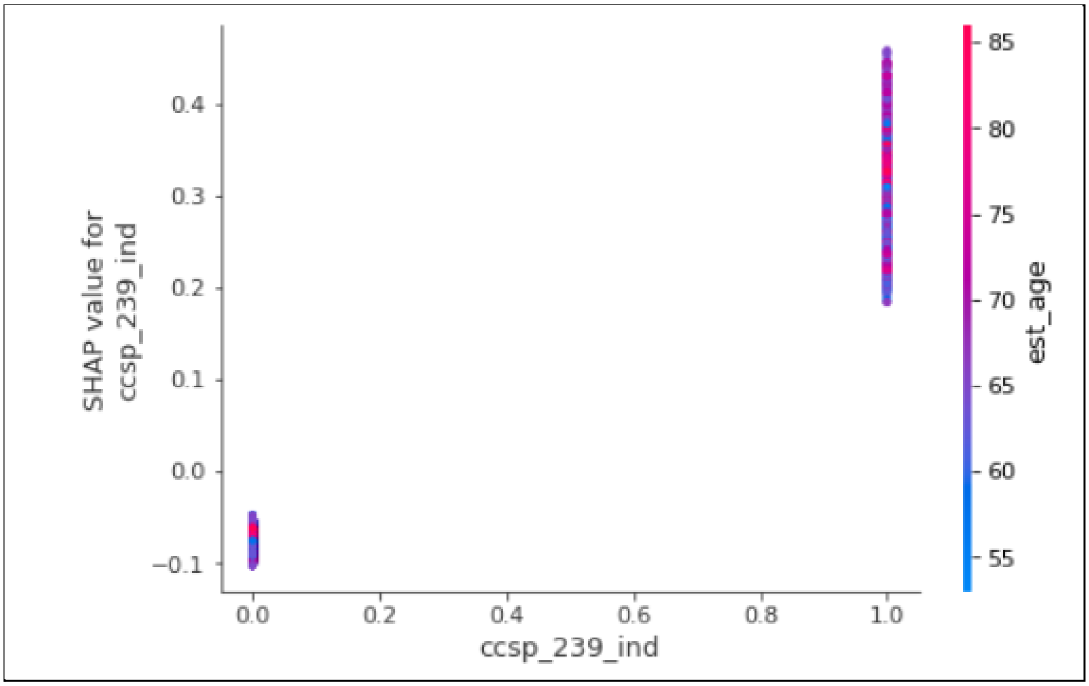
The est_age variable really demonstrates the age requirements to qualify for Medicare. Medicare is available to those older than 65 and individuals that qualify for disability according to Social Security standards. The est_age SHAP dependency plot shows a stark decline in SHAP value at the 65 year mark. Those younger than 65 have a much higher SHAP value indicating increased odds of facing transportation issues. After 65, there is a sharp decline in the SHAP value, and then there is a slight step increase after the age of 80.

The individuals below 65 years are on disability, and thus the increase in SHAP values for this group is because these individuals may be incapacitated and thus unable to get themselves to the appropriate healthcare facility as they face transportation issues. There is a negative SHAP value associated with those above 65. This is probably focused on individuals who have just started enrollment in Medicare. These individuals are not necessarily incapacitated based on Medicare requirements. There is a slight increase when past 80, but it tends to stay negative.

The other feature that interacts the most with this particular variable is the rx_overall_pmpm_ct variable which is the average prescription per month for each subscriber. The amount of prescriptions tends to be much higher for individuals below 65 and thus on disability which coincides with the previous analysis on Part D Medicare expenditure. In order to significantly reduce the impact of transportation as a social determinant of health, at least for Medicare subscribers, the focus should be on individuals who are on disability.

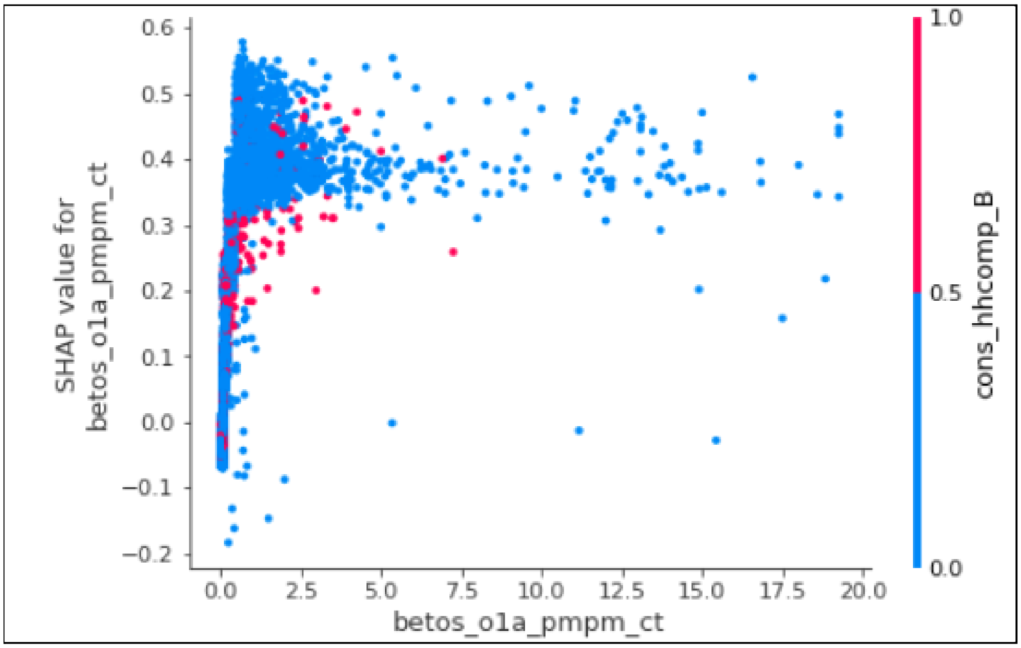
There also seems to be a decrease of prescription counts when the subscribers are older than 80 which could be indicative of these individuals' ability to access their pharmacies to refill their prescriptions. One potential solution is to implement telemedicine visits and prescription mail delivery services to assist these individuals.

Another top feature in the model is ccsp_239_ind:



Ccsp_239 is a binary indicator for CCS code - superficial injury, contusion. Looking at the other CCS codes in the dataset, this one seemed to be indicative of subscribers that face any type of accident or physical trauma. The plot looks like there are a lot more people that have an indicator of 1, but that is not the case as the values are spread out in SHAP values based on the interaction with the est_age variable. Looking closely at the data, there are only 9737 subscribers out of the 69122 individuals in the training data set that have been marked as having a superficial injury and/or contusion. Nevertheless, the model is indicating that those individuals with this indicator are more likely to face transportation issues.

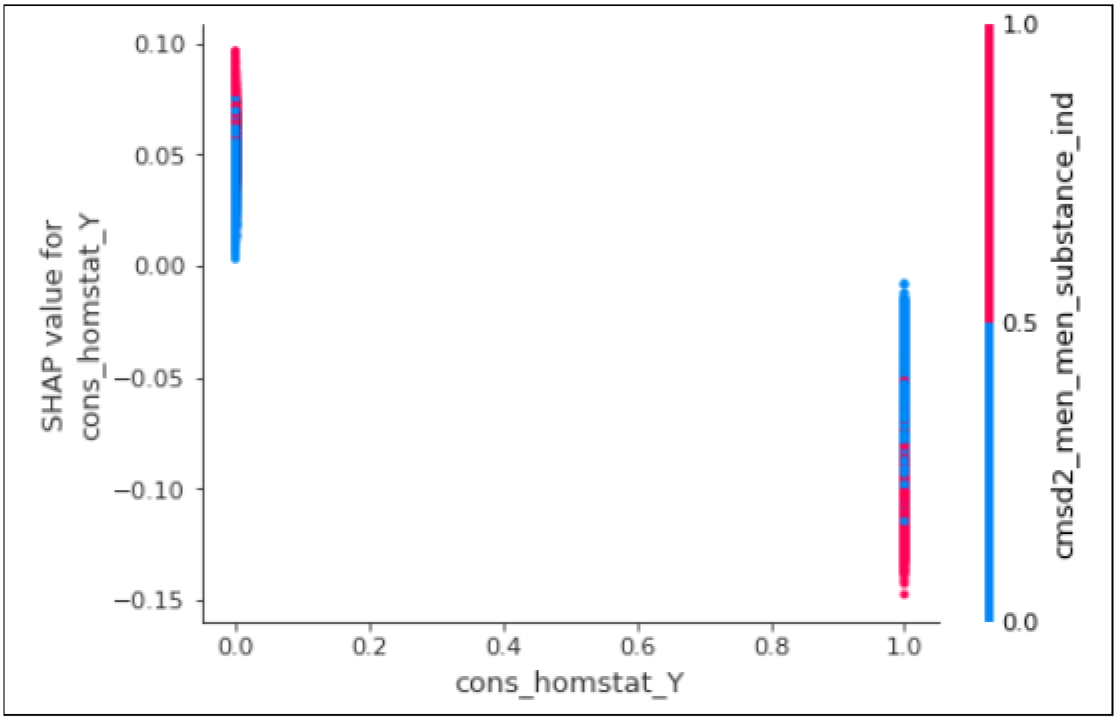
This dependency plot shows the impact of the betos_o1a_pmpm_ct variable in the model:



The scale of the SHAP values for the betos_o1a_pmpm_ct variable is much larger than other SHAP dependency plots. This suggests that this variable may have the potential largest contribution range to increasing the odds of a subscriber facing a transportation barrier. This BETOS variable indicates the per member per month count of ambulance claims. Essentially, it indicates the average frequency of ambulance frequency per month for a particular subscriber. This is most likely a lagging indicator of transportation issues, as individuals who need medical attention but are unable to transport themselves to the nearest facility may have to resort to using ambulance services to get medical help. The majority of subscribers (i.e. ~60,000) in the training data have a value of 0 for this particular feature. The dependency plot shows a stark increase in the SHAP value, as individuals who use ambulances frequently are the most likely to face transportation issues.

This dependency plot can be the justified case for action to come up with a solution for individuals facing transportation issues. Medicare providers like Humana would rather have subscribers use any other means of transport versus using ambulance transport because ambulance services are expensive and it may tie up ambulance resources unnecessarily when other solutions may work. It may be in the Medicare insurance provider’s best interest to promote the use of rideshare or taxi credits for these individuals who frequently use ambulance services to gain access to healthcare. Providing rideshare credits will most likely result in significant savings from avoiding unnecessary ambulance rides when alternatives are made available. One thing to note is that it is most likely that these individuals are suffering from a disability of some sort, so ensuring viable transportation options (i.e. wheelchair accessible when needed) will ensure increased use of these alternative services.

The cons_homstat_Y variable was another important factor in the model:

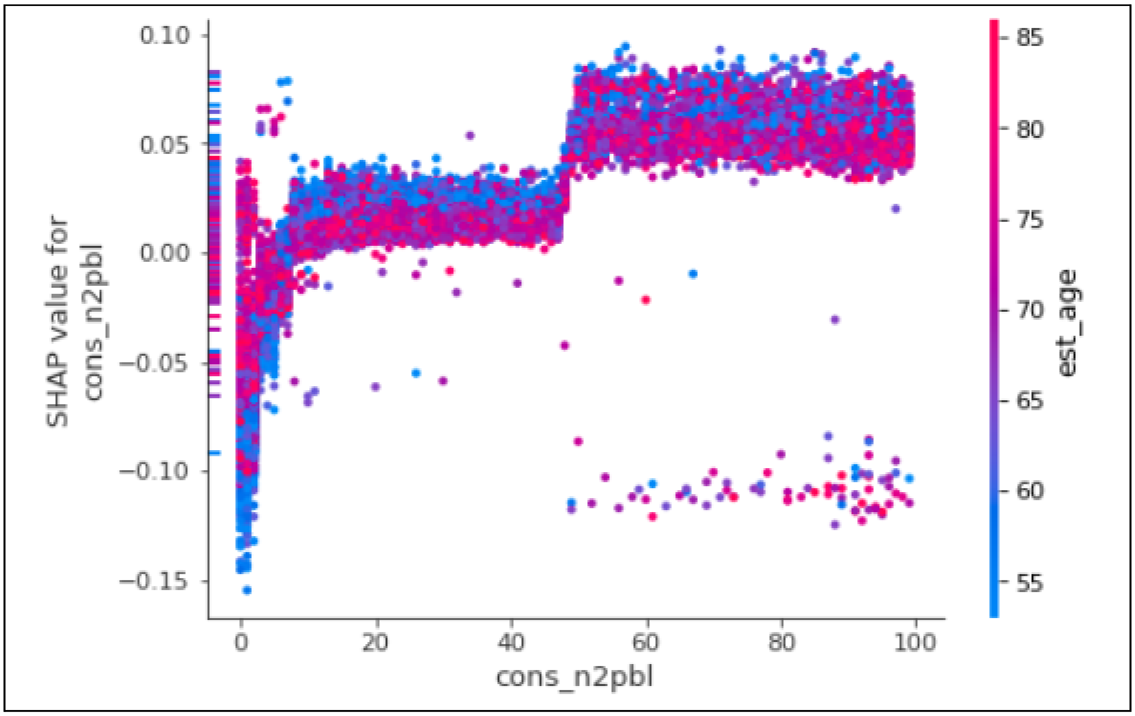


The cons_homstat_Y variable indicates whether the subscriber is a homeowner. As expected, individuals who are homeowners tend to have a negative SHAP value indicating decreased likelihood of facing transportation issues. This is essentially a proxy variable for people that are financially sound and thus would not have issues accessing healthcare.

The SHAP package chose the cmsd2_men_men_substance_ind variable as the one other variable that interacted with cons_homstat_y the most. The cmsd2_men_men_substance_ind indicates individuals who have mental and behavioral disorders due to psychoactive substance abuse. Individuals that have such mental disorders and are not homeowners have an increased risk of facing transportation issues. It is interesting to note that individuals that are indicating this mental health disorder and are homeowners tend to have an even lower probability of facing transportation issues. This may be because the homeowners have other family members in their home who can transport them, allowing them to access their medical needs.

Ultimately, this dependency plot shows that financial instability and substance abuse are indicative of a subscriber facing transportation issues. These individuals need to be treated for their mental health issues and should be offered resources to achieve financial stability to avoid being stuck in this vicious circle.

This last dependency report included looks closely at the cons_n2pbl variable:



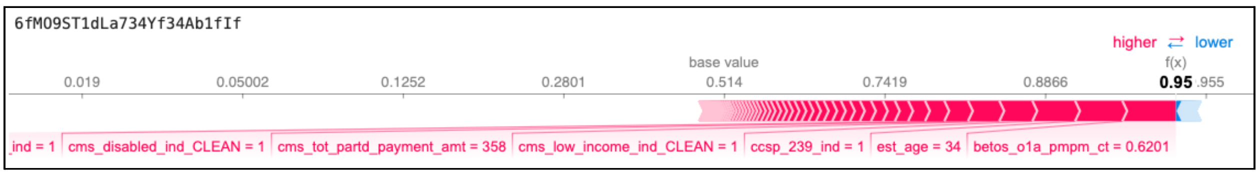
The cons_n2pbl variable indicates the % Black in census tracts that individual subscribers belong to. This data is only available for approximately 33% of the dataset, as most of the subscribers do not have location data available. This ended up not being a top 20 predictor in the model, but it may be because of the limited availability of location data. Even with the fewer data points, the SHAP dependency plot shows a problematic issue on racial equity. Based on this plot, it can be seen that areas with increasing Black population tend to have an adverse effect on access to transport for healthcare needs. There is a significant step increase in the SHAP value for areas where the Black population is greater than 50% of the area. Essentially, individuals living in predominantly Black areas have higher odds of facing transportation issues.

This lack of equity contributing to lack of transportation access can be attributed to the fact that Black neighborhoods tend to be more impoverished. It may also be the case that healthcare facilities are located further away from Black neighborhoods and thus more difficult to access. Further analysis should be done to understand this possible relationship between Black neighborhoods and the lack of access to transportation for healthcare needs. One possible solution to this problem in the short term could be to implement telemedicine in these areas or even have healthcare resources visit these “healthcare deserts”. Longer term solutions will require leveraging municipal and state legislation to promote building healthcare facilities in these areas to increase equity.

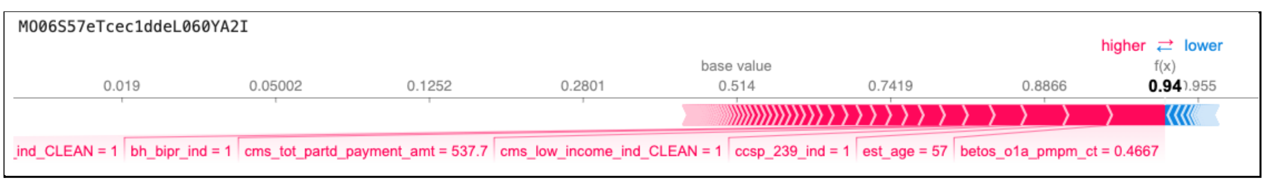
Individual Prediction Explainability - SHAP Force Plots

SHAP Force plots look at the SHAP values of each of the predictors in the model and highlight their contribution to the prediction. The red indicates a positive “force”, pushing the predicted value (i.e. probability of subscriber facing transportation issues) higher. The blue bars represent a negative “force” pushing the predicted value in the other direction. Taking a look at individual SHAP plots for each prediction can possibly show what kind of factors are pushing a particular subscriber to face a transportation problem based on the model. It is also important to note that the final probability is based on the upsampled dataset used to train the model. This model output probability has to be adjusted per the formula in the Modelling & Model Evaluation section above.

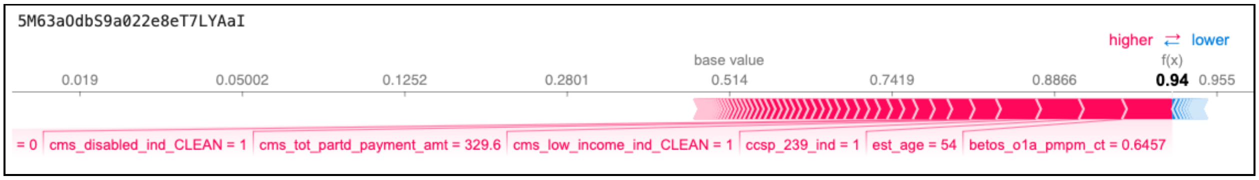
Here are the SHAP Force for the top 3 ranked members that may face a transportation problem in the holdout data:



For this individual, the top contributing feature is betos_o1a_pmpm_ct which is the number of times a subscriber uses an ambulance per month on average. This individual has an average count of 0.6201 which is rather high. This particular feature is most likely a lagging indicator of this individual facing transportation issues. This subscriber uses an ambulance often because he/she does not have access to another mode of transportation. This individual is also disabled, low income, and has faced a contusion/superficial injury (ccsp_239). The cms_tot_partd_payment_amt indicates the amount spent on Medicare Part D prescriptions which is also quite high.

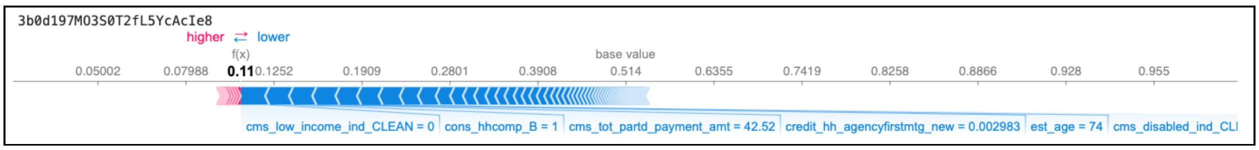


The subscriber above has similar indicators as mentioned before. In addition, this individual has a bh_bipr_ind value of 1 meaning that this individual also faces a bipolar mental disorder. It is interesting to note that one of the negative factors pushing this subscriber’s likelihood to face transportation issues is cons_n2pbl which is the percentage of Black people in this individual’s census tract which was valued at 2%, indicating that this person is located in a predominantly non-Black area. This is a negative force because the model has identified that the likelihood of facing a transportation issue tends to be higher for individuals located in predominantly Black census tracts.



This subscriber above is in line with the previous 2 top subscribers most likely facing transportation issues. This individual also has frequent ambulance use, superficial injury/ contusions, low income, and is disabled.

SHAP force plots also help in the other direction to highlight why an individual may be likely not facing a transportation problem. The next force plot shows the plot for the subscriber ranked last in terms of probability of facing a transportation issue:



This particular individual is ranked this low mainly because they do not have low income, are married with no kids (i.e. partner can possibly drive or at least accompany this individual to a healthcare site if needed), are not disabled, and spend relatively low on Medicare Part D prescriptions.

Going forward, using these individual SHAP Force plots can highlight the likely reasons why a particular subscriber may have transportation problems and allow for Humana to address these concerns individually. For instance, the top ranked patient (ID: 6fMO9ST1dLa734Yf34Ab1flf) has frequent ambulance use, disability, and low income as top indicators of facing transportation issues. Humana can help alleviate this individual's need to frequently use an ambulance, which is also costly to Humana, by possibly providing other alternatives like disability-accommodating taxi vouchers or even rideshare credits for this particular subscriber.

Recommendations

Partnerships with Rideshare Companies

Ambulance rides are expensive. Humana should consider offering rideshare opportunities to its subscribers instead. Lyft, for instance, has begun an initiative called "LyftUp", which aims to provide affordable and reliable transportation to those in need. Humana could partner with a company such as Lyft in order to provide rides to medical appointments for individuals.

LyftUp is not limited to offering transportation. Part of this initiative also includes what Lyft calls the Grocery Access Program, which provides easier access to healthy food options in underserved communities. Healthy eating habits are critical to an individual's overall health. Moreover, considering that many individuals who have a transportation barrier may be under-prescribed, Humana could work with Lyft to extend this program to include prescription delivery services as well.

Uplifting Community Health Centers in the Age of Telemedicine

According to McKinsey and Co., health providers are seeing up to 175 times as many patients via telehealth as they were prior to the COVID-19 pandemic. Surely, telemedicine will play a vital role in healthcare for decades to come, and it has the potential to alleviate transportation barriers for many. However, at this point, telehealth is simply not an option for some low-income individuals.

According to the HealthAffairs journal, community health centers (CHCs) serve over 28 million low-income patients in the U.S. While the telemedicine industry boomed as the pandemic spread, CHCs experienced unprecedented layoffs and site closures. As late as 2018, a majority of CHCs offered no telehealth services - of these, only a few were equipped to even begin actively implementing it.

Humana should consider partnering with CHCs to help them improve their telemedicine capabilities, thereby giving low-income communities access to this incredibly beneficial service and alleviating transportation barriers for many.

Lastly, for optimal telemedicine service, a member needs access to the internet and a phone or laptop with a working camera and microphone. Some subscribers may lack access to this technology. Humana can solve this problem by extending the CHC partnerships to include a technology or communication services company.

Emphasis on Mental Health

In recent years, conversations on mental health have become more prevalent in media and politics. It is becoming increasingly clear that mental health is essential to physical health. For the longest time, this topic did not get the attention it deserved. Perhaps that is still so; our model output suggests that a member with a transportation barrier is more likely to be under-prescribed for mental health conditions, particularly when they are a low-income individual.

Humana should place a higher emphasis on mental health treatment, perhaps offering increased subsidies for certain prescriptions. This would likely pay for itself, as improved mental health tends to lead to similar outcomes in physical health, according to the Mental Health Foundation.

References

Health Research & Educational Trust. (2017, November). Social determinants of health series: Transportation and the role of hospitals. Chicago, IL: Health Research & Educational Trust. Accessed at www.aha.org/transportation

USDA ERS. Food Access Research Atlas Data 2015, 18 May 2017 https://www.ers.usda.gov/data-products/food-access-research-atlas/download-the-data/

Measure of America. "Mapping America: HDI Indicators." New York: Social Science Research Council, 2018. https://measureofamerica.org/maps/?state^doctors^all_all^Health^health

"Physical Health and Mental Health." *Mental Health Foundation*, 10 Feb. 2020, www.mentalhealth.org.uk/a-to-z/p/physical-health-and-mental-health.

Cole, June-Ho Kim, Eesha Desai, Megan B. "How The Rapid Shift To Telehealth Leaves Many Community Health Centers Behind During The COVID-19 Pandemic." *How The Rapid Shift To Telehealth Leaves Many Community Health Centers Behind During The COVID-19 Pandemic | Health Affairs*, 2 June 2020, www.healthaffairs.org/do/10.1377/hblog20200529.449762/full/.

Bestsennyy, Oleg, et al. "Telehealth: A Quarter-Trillion-Dollar Post-COVID-19 Reality?" *McKinsey & Company*, McKinsey & Company, 1 June 2020, www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/telehealth-a-quarter-trillion-dollar-post-covid-19-reality.

Shapley sampling values: Strumbelj, Erik, and Igor Kononenko. "Explaining prediction models and individual predictions with feature contributions." Knowledge and information systems 41.3 (2014): 647-665.