Humana-Mays Healthcare Analytics

2020 Case Competion

Transportation Issues Prediction Analysis

October 11th, 2020

Table of Contents

1. Executive Summary

This case study focuses on helping Humana identify Medicare members suffering from transportation issues. After dividing members with transportation issues into different groups with unique demographic and stress factors, we proposed actionable recommendations for Humana to help members in each group overcome their transportation barrier and achieve their best health.

Since the original training dataset contained over 800 features, we started with exploratory data analysis to give us a basic understanding of each feature. With a general idea of our dataset, we generated new features through feature engineering and selected a final 74 features through a feature selection process. Using these 74 features, we constructed a XGBoost model, which utilizes a gradient boosting algorithm, as our final model with the optimal hyperparameters. Our XGBoost model finished with a ROC (Receiver operating characteristic) AUC (Area under curve) score of 0.7508 and an accuracy of 72%.

With our final model, we explored a number of features that were highly impactful on whether an individual member would struggle with transportation issues. These include their estimated age, stress index level, number of visits to a physician office, and disability. We performed a factor analysis on our Top 20 performing features and used K-means clustering to identify groups based on different factors. The four groups that we identified in need of additional support were: members with high needs for ambulances and low mobility, members with low income and disabilities, members with disabilities and low mobility, and members with low income and credit stress. For each group, we proposed a unique combination of viable solutions for Humana to target their specific needs to solve their transportation issues. By integrating our various solutions, we estimate Humana will be able to save ~ 15.6-million dollars annually from the reduced costs of previously delayed care.

2. Case Background - The Business Problem

Social determinants of health are considered the single biggest factor, contributing as much as 40%, to an individual's health risks and outcomes. Transportation, as one of the social determinants of health, play a massive role in an individual's health outcomes as they affect access to daily necessities, economic opportunities, and health care services. However, with the growing cost of transportation in the United States, there has become an increasing lack of access to reliable transportation. In particular, these transportation issues have had a disproportionately greater impact on elderly people, members of minority groups, and individuals with lower income. In a study of 7,500 Medicare beneficiaries, about 25% trusted sources reported having limited access to transportation. The ongoing coronavirus pandemic has only exacerbated existing transportation issues with lockdowns, social distancing policies, and higher concern for exposure contributing as decisive factors against the use of public transportation and visiting health care locations.

As a result, transportation issues have been identified as a leading cause of patient no-shows, missed appointments that are associated with increased medical care costs for the patient and insurers. Patient no-shows have been linked to disruptions in patient-provider relations, delayed care and increase in emergency room visits. These outcomes have a major effect on the United States healthcare system, contributing annually to an estimated cost of an additional 150 billion dollars. Humana, as one of the largest health insurance companies in the United States, can be at the forefront of mitigating some of these transportation issues to not only help its members reduce their health risks but alleviate the additional financial strain these outcomes have on the United States healthcare system.

Our goal is to build a predictive model that incorporates multiple features from Medicare member data and produces a probability of each member's likelihood of experiencing transportation issues. Using a combination of these probabilities, key performance indicators, and further statistical analysis, we want to construct member group profiles to identify subsections of members as a basis of shaping recommendations. We feel this case competition presents an incredible opportunity for us to share positive transportation recommendations, backed by our data analysis, that will create better health care outcomes for Medicare members, participating hospitals/clinics, and Humana.

3. Data Preparation

The original dataset for this project contained information on 69,572 Medicare members across 826 data fields including categories like Consumer data, Medical claims, Pharmacy Claims, Lab Claims, Demographic/Consumer Data, Credit Data, Condition Related Features, CMS Features, and other Miscellaneous groups. Our target variable, transportation_issues, was a self reported measure from members who indicated whether a lack of reliable transportation has kept them from accessing medical appointments, meetings, work, or getting things needed for daily life in the last 12 months. Members suffering from transportation issues were marked as a 1 in the transportation_issues field, while those that indicated no were coded as a 0.

Medical Claims Features

CCS Procedure Code Categories
BETOS Procedure Code Categories
Utilization by Category (IP admits/ER visits/Outpatient/Ambulance etc.)

Pharmacy Claims Features

Prescription Days Covered
Brand/Generic Prescription
Mailed/Non-mailed Prescription
Maintenance Prescription
GPI2 Level Prescription Utilization

Lab Claims Features

Abnormal Lab Results Indicator
Abnormal Lab Results Indicator by Category (Cholesterol/ EGFR/HbA1c/Hemoglobin etc.)

Demographics/Consumer Data

Age
Geography
Census Education Level
Household Composition
Homeowner Status
Census Percent Motor Vehicle Ownership

Credit data

Balance All Mortgage Accts Past Due
% HH Bank Card Accts - Severe Derogatory Accts
Number All Mortgage Accts - 120 Days Past Due or Collections
% Balance to High Mortgage Credit

Condition Related Features

Behavioral Health Condition Indicator
Charlson Comorbidity Index
Functional Comorbidity Index
Diabetes Complication and Severity Index
CMS Diagnosis Code Categories
MCC Diagnosis Code Categories

CMS Features

Disability
Dual Eligibility
Low Income Subsidy
CMS Risk Score
CMS Total Payment Amount

Other features

Health Program Participation/Status
HEDIS-like Features
Provider Specialty Features
Revenue Code Features
Behavioral Segmentation

| 14

Image 3.1: Description of Primary Dataset Features Humana Kick-Off Slides

3.1 Exploratory Data Analysis

As transportation issues can be attributed to a wide range of factors, we wanted to develop a fundamental understanding of this data as these insights would shape the framework for establishing groupings of members on which we would base our recommendations on.

Age

The distribution of the ages (est_age) shows that this dataset is made up mostly of an elderly population, with a mean age of 70.81 years old, and the middle 50% of values falling between 66 and 77. When the ages are grouped by 10 year increments (Ex 20-29), individuals that are younger reported having more transportation issues than their older counterparts.

Disability

Only 12.19% of members who did not have a reported disability suffered from transportation issues, while 23.18% of members reported with a disability suffered from transportation issues.

Household Composition and Homeowner Status

24.88% of single parent households reported transportation issues, while 9.25% of households with a minimum of two people households and no children reported having transportation issues. Homeowner status plays a similar role with 24.01% of renters indicated they had transportation issues, while only 10.86% of homeowners indicated they had transportation issues.

Health Scores

There appears to be a positive linear relationship between higher weighted health scores (CCI, DSCI, FCI, and HCC), which indicate more health issues, and having transportation issues.

Smoker Status

30.9% of active smokers reported having transportation issues, which is significantly higher than the reported percentage of transportation issues for members who were former smokers (13.4%) and members who have never smoked (13.2%).

Prescription Drug Usage

When generic prescription drug use is grouped into 6 levels (1 as lowest usage , 6 as highest usage), the higher usage of generic prescription drugs had a positive relationship to an increase in transportation issues. Only 11.9% of members in Level 1 reported having transportation issues while 30.4% of members in Level 6 reported issues.

Overall per member per month count for prescriptions (rx_overall_ompm_ct) demonstrated a positive relationship to increased transportation issues. When these values are binned into 7 levels (1 as lowest usage, 7 as highest usage) we see that transportation issues are the highest (36%) at Level 7 than at Level 1 (11.2%).

Hospital Visits:

Members, with existing transportation issues, on average had twice as many hospital claims (0.11 per member per month claims) than those who did not have transportation issues (0.052).

MCC Diagnosis Codes Categories:

More than 50% of the total submitted claims were submitted into these categories: Exam and Screening, Hypertension, Signs and Symptoms(other), Hyperlipidemia, Nervous System(Other), Sense Organs(eye).

These are the categories with highest claim frequencies and at least two claims per member each year: Nervous System(Other), Exam and Screening, Diabetes(Other), Back/Neck Pain, Hyperlipidemia, Signs and Symptoms(Other), Hypertension.

3.2 Feature Engineering

Our next goal was to create new, informative features that would help improve our model performance. We utilized many different feature transformation techniques, outlined in the sections below, along with creating dummy variables for our categorical fields.

2.2.1 Group Binning and Ranking

Our first technique focused on binning categorical data (educational background, household composition, homeowner status, and languages spoken) into narrowly defined groups. Since principal component analysis was going to be completed on many of the numerical variables, percentile ranks were generated to maintain the original ordered structure of the numeric values. These percentile ranks were then summed together to create different scoring metrics based on credit data, SUBMCC claims, CMS data, RX claims, and BETOS claims.

2.2.2 Weighted Metrics

In order to map some relationships between and within categories, we created weighted scores by combining demographic data with specific categories.

Weighted Metric Name	Descirptions
StressIndex	A weighted score based on demographic and credit data aimed at capturing the level of stress a member may experience
MobilityIndex	A weighted score based on demographic and health data intended to capture a member's difficulty to physically move

Table 2.1 Key Weighted Metric Name and Descriptions. Although many more weighted features were created, these two were identified as the most informative to differentiate members with/without transportation issues.

Each of these weighted scores were normalized using a minimum, maximum scaler to represent the values as a percentage making them more intuitive for non-technical shareholders.

2.2.3 K-Means Clustering

K-Means clustering is an unsupervised machine learning algorithm that starts with randomly grouping data based on centroids and performs iterative distance calculations to optimize the position of the centroids, until k centroids have stabilized.

In order to classify members into different groups, K-Means clustering was used on percentile ranking scores and other weighted metrics to identify different clusters relationships to transportation issues. Each value that was placed into the K-Means algorithm was standardized to keep them on the same scale for measuring the distance between different clusters. This process generated 30 new variables, with 3-4 clusters each, grouped based on credit, health, stress, and age Data

2.2.4 Deep Feature Synthesis

One of our challenges during the feature engineering phase was to identify other new, useful features, so we decided to utilize an automated feature engineering technique called Deep Feature Synthesis. Deep Feature Synthesis uses entities and relationships between these entities to create transformations and aggregations of the data. For the purposes of our project, only the transformation aspect was used to create new features from similar categories. Multiplication and addition transformations were created on each combination of

variables among 33 different categories generating over 7000 new features to evaluate for our model.
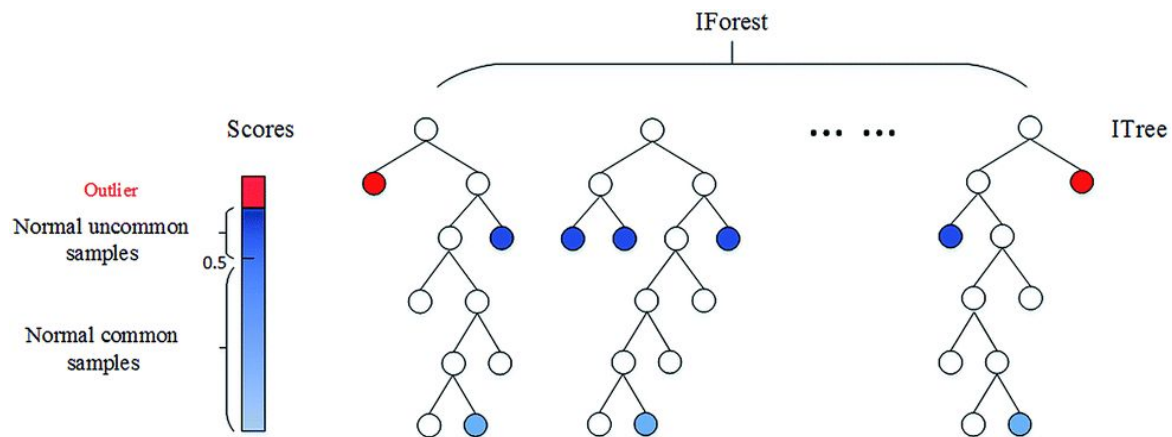
2.2.5 Isolation Forest

Image 2.1 Isolation Forest Diagram. Isolation forest is an unsupervised machine learning algorithm that isolates anomalies based on what is considered normal in the dataset through the use of decision trees.

Another issue that we identified with this dataset was with outlier values or observations that might be considered anomalies. Using an Isolation Forest algorithm, we assigned each individual an anomaly score, based on the likelihood of this member being considered an anomaly, and a binary indicator if these members were an anomaly. Approximately 5600 members were identified as anomalies with 29.54% reporting transportation issues compared to 13.3% of the "normal" members reporting transportation issues.

After creating a large number of features, derived from original fields in our dataset, we noticed many data fields were heavily correlated which might mask their effects when applied in the model. To remedy this issue, we utilized Principal Component Analysis (PCA), a dimensionality-reduction method to reduce variables in a dataset while still accounting for the information of that data. Data categories we applied PCA to included credit data, betos_pmpm_ct data, submcc_pmpm_ct data, and rx_pmpm_ct data. Each of these categories were reduced to the point where their new principal components accounted for at least 85% of the original variance (Appendix B)**.**

The resulting feature engineering and feature extraction with Principal Component Analysis transformed our original dataset to generate over 8000 new features.

3.3 Feature Selection

Given that our new dataset had over 8000 features, we utilized a three-step feature selection process to help reduce it's dimensionality and avoid overfitting our model.

Our first step was to utilize a forward selection process using a base Logistic Regression model to identify the best performing ROC AUC scores. This forward selection process works by running different models in a series of rounds, where each round a single, new feature is placed into the model. The best performing added feature (based on the ROC AUC score) is then placed into the base model and a new round is completed, with the base model holding the best performing features from previous rounds.

The next step followed an intrinsic method with the use of Random Forest to identify which features were most important during the training phase. Approximately 250 features were selected during this stage which were carried over to our final step, another forward selection process, using XGBoost, to return the features that were contributing the most to improving the ROC AUC metric.

Our final feature selection process finished with 74 features. This low number of features helped prevent building an overly complex model that would significantly reduce computational time, generating much quicker predictions.

4. Modeling Approach

Since the original business question centered around predicting individuals that will have transportation issues, we decided that constructing a binary classification model would be the most appropriate way to address this question.

4.1 Model Selection

As a starting point for model evaluation, we started with a base Logistic Regression model and identified a ROC AUC metric of 0.68 as a baseline score to compare with other models.

Another important goal was to reduce the Type II Error (the false negative rate) because we felt it would be better to predict people without transportation issues as having transportation issues rather than predicting people as having no transportation issues when they actually did have transportation issues (Type I Error).

Our dataset was split into a 60% train set, 25% testing set, and 15% validation set. The 25% test set was only used at the end after we chose a final model and our model performance was formed based on this set. With only 14.66% of members reporting transportation issues, we used different sampling strategies: SMOTE which generates synthetic samples for the minority class, downsampling the majority class, and a pipeline combination of both SMOTE and downsampling in order to solve the class imbalance issue. These strategies were tested on different models, however, to meet our Type II Error criteria we decided to utilize class weight hyperparements within our model as it offered the best tradeoff between precision and recall scores.

We constructed base models for the following classification algorithms with their accompanying ROC AUC: Support Vector Machines (0.65), KNN (0.67), Naive Bayes (0.67), Neural Networks (0.67), Logistic Regression(0.71), Random Forest (0.71), CatBoost (0.73) , and XGBoost (0.74).

4.2 Final Model

After evaluating different models selection, we decided to use XGBoost, a gradient boosting algorithm, for building our predictive model. XGBoost implements a gradient boosting decision tree algorithm designed to be incredibly efficient and accurate. Boosting is an ensemble approach where new models are created based on weak predictors, with their residuals or errors being combined together to make a final prediction (Machine Learning Mastery, 2020). In order to minimize loss, it uses a gradient descent algorithm when adding to the new models.
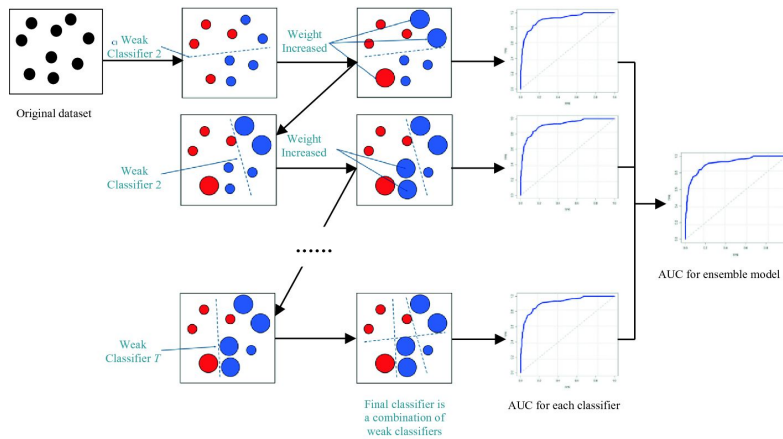
Image 4.1: Gradient Boosting Example

To maximize performance of our XGBoost model, we utilized a 5-fold Grid Search to tune the models hyperparameters. Our final models hyper parameters were as follows: n_estimators: 100, max_depth = 2, scale_pos_weight = 5.5, subsample = 0.6, colsampe_bytree = 0.8, min_child_weight = 5, eta = 0.2, alpha = 0.01, gamma = 2, and reg_lambda = 4.5.

5. Model Evaluation

5.1 Model Performance

Our final model finished with a ROC AUC metric of 0.7508 and a model accuracy of 72%. The weighted average scores for the precision, recall, and F-1 scores were 0.83, 0.72, and 0.75.
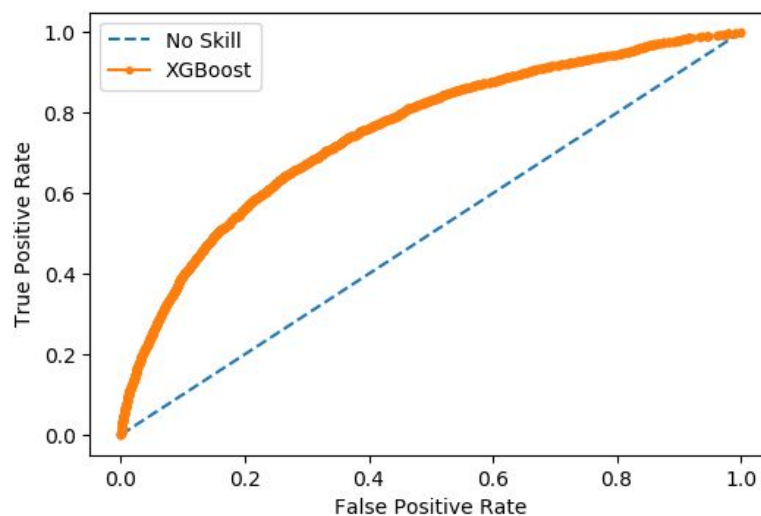
Figure 4.1 ROC AUC Curve for No Skill and XGBoost Model. A No-Skill ROC Curve, predicting only the majority class, has an ROC AUC metric of 0.5. Our XGBoost model returned an ROC of 0.751, representing a 0.251 improvement over the no-skill model.

Confusion Matrix	Predicted 1	Predicted 0
Actual 1	1671 (TP)	914 (FN)
Actual 0	3993 (FP)	10815 (TN)

Table 4.1 Confusion Matrix for our final model The false negative rate (Type II error) was the lowest of the four categories and reflected our original goal for the project. 1 represents transportation issues while 0 represents no transportation issues.

Figure 4.2 displays the breakdown of predicted probabilities for transportation issues, with the colors representing the actual values. From this figure we can see that the model confidently predicts no transportation issues, but when looking at the transportation issues distribution the misclassification percentage is much higher.
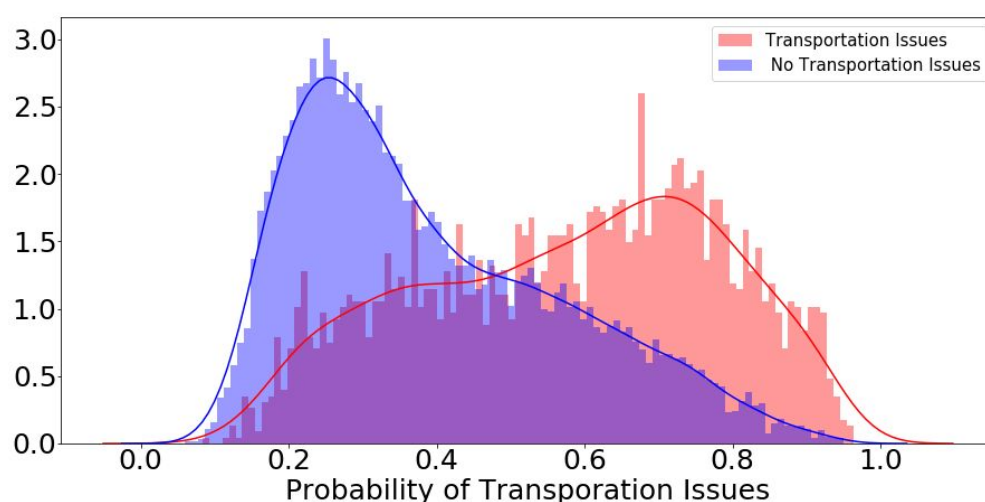

Figure 4.2 Model Predicted Probabilities for Transportation Issues

For building predictions on the holdout dataset, we trained our model on our entire dataset and completed a 10 fold cross validation to evaluate model performance on the entirety of the dataset. The average ROU AUC metric for a 10 fold cross validation was 0.7425.

5.2 Key Performance Indicators/ Variable Importances

Another primary motivation for using an ensemble tree algorithm was that it can give a strong global and individual interpretation when identifying key factors of the model. Our feature importances were measured based on their SHAP value (Shapley Additive Explanations) which examines the impact a certain feature has on the overall model

performance. Based on these SHAP features, Figure 4.3 displays the top 20 most important features with demographic, health, and stress measures contributing the most information into our model. A full list of a description of the top 20 features can be found in Appendix A.
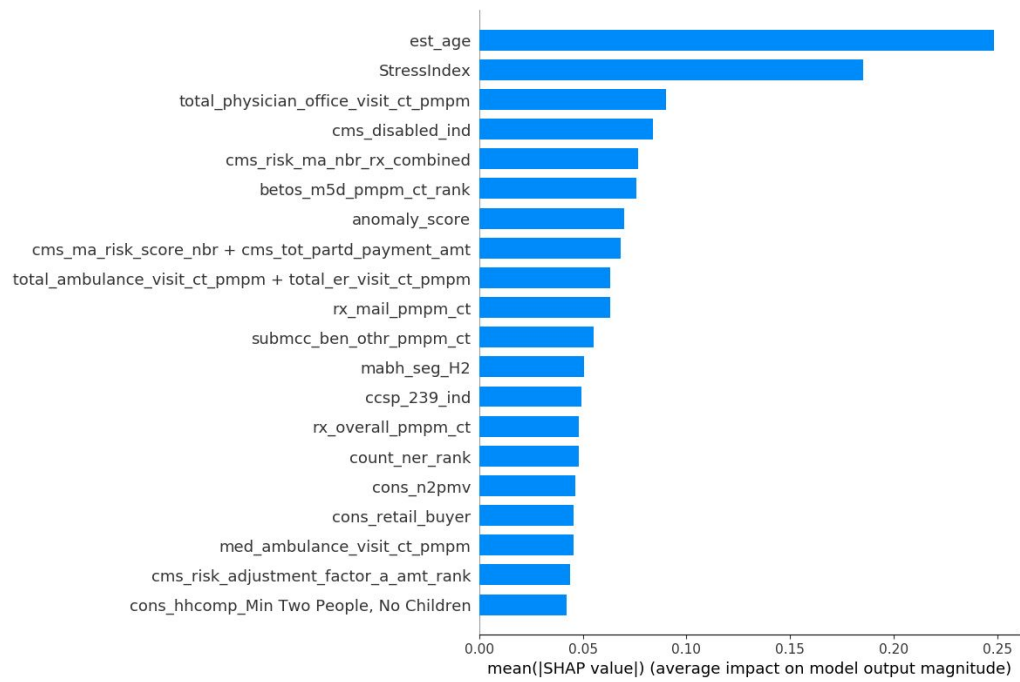

Figure 4.3 Most Important Features. This figure shows the most important features for our model based on their mean shap value, which helps represent their average impact on the model's output prediction.

Figure 4.4 builds on the insights from Figure 4.3, demonstrating how specific values contributed to the predictions of the model. High Stress, Ambulance/ER visits, CMS risk, and overall prescription claim values contribute the most to determining if a member is suffering from transportation issues. Higher Age, Visits to the Physician, and benign neoplasms claims contribute to a lower likelihood of having transportation issues. Binary figures like disability indicators and Contusion injuries play a role in identifying individuals with transportation issues.
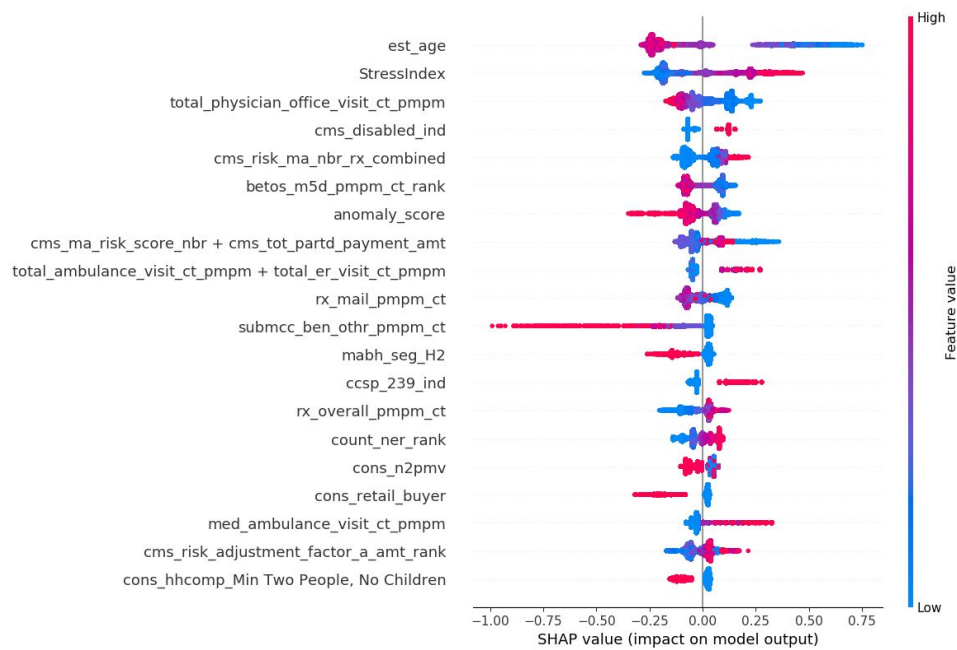
Figure 4.4 Feature Impact and Direction. This figure shows which features correspond to different predictions. For example, low est_age values corresponded to higher strength in predicting transportation issues, while higher ages reflected less transportation issues. Red Values represent observations with high numbers, while blue represents observations with low numbers.

SHAP dependency plots based on individual variables provide additional context on how these individual variables are related to performance of the model and other variables. Appendix C contains four different SHAP dependence plots for the features: Stress Index, est_age, rx_overall_pmpm_ct, and total_physician_office_visit_ct_pmpm. Rx_overall_pmpm_ct and Stress Index values have positive relationships with transportation issues, while total_physician_office_visit_ct_pmpm and est_age have negative relationships with transportation issues.

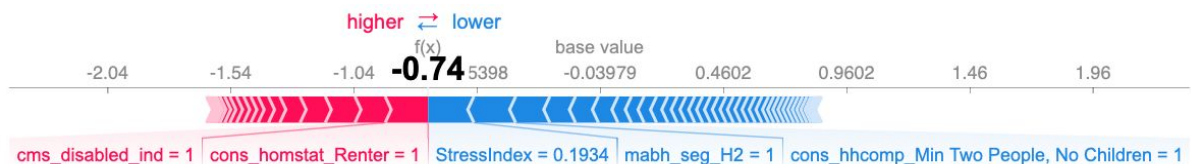**Figure 4.5 Correctly Classified Observation (Transportation Issues)**

Figure 4.6 Correctly Classified Observation (No Transportation Issues)

15

higher ⇄ lower
base value
0.07

-2.04   -1.54   -1.04   -0.5398   -0.0   0.4602   0.9602   1.46   1.96

cms_disabled_ind = 1 | cons_homstat_Renter = 1 | StressIndex = 0.4834 | submcc_ben_othr_pmpm_ct = 0.6005 | betos_m5d_pmpm_ct_rank =

Figure 4.7 Misclassified Observation (Predicted 1, Actual 0)

These individual estimator SHAP plots help us understand how each variable is interacting with a specific observation. In Figure 4.5, the values that correspond to stronger transportation issues (Stress, Disability, Contusion injuries) are pushing this model to predict this individual as someone with a transportation issue. Figure 4.6 demonstrates the general findings that a low Stress Index and household with no children decreases the likelihood of having transportation issues. Although Figure 4.7 was an incorrectly predicted observation, we can view some of the key drivers as to why this was misclassified and further enquiry into that individual's situation.

5.3 Model Conclusions

From these model insights, we discovered that disability, age, prescription claims, Stress, and motor vehicle percentage are all major factors for transportation issues and we can use these variables as foundations for forming user groups. While we believe that our model performs well enough to give a general estimate, we recognize that our model has potential flaws, in particular the poor precision score for positive cases. Our recommendation would be to use this model's predictions as a guide and dedicate more time on its specific components to identify customized solutions for each member.

Using an explainer model (like in Figures 4.5 - 4.7), we can build intuitive visualizations that can inform medical staff and insurance providers of the probabilities of transportation issues and which factors are affecting individuals the most. This information can be combined with existing clusters to build recommendations that will more accurately address concerns of multiple members. The fear of simply predicting individuals with transportation issues, is that we would fail to fully understand their individual situation and thus be unable to accommodate their individual needs.

If the eventual end goal is to generate a more accurate model for predicting transportation issues, we believe that there is a massive potential in utilizing specific locational data. Since our existing dataset was missing PPI (Personally Identifiable Information) regarding residence location, future models could leverage this data with the locations of Humana's existing coverage network of hospitals and clinics. This hypothetical model could take into account the number of insured locations within a certain range of an individual's address along with calculating the distances between these locations. This specific locational data could also inform the model about other modes of transportation and determine if these other modes are feasible for the member to utilize and thus build a more accurate prediction metric.

6. Recommendations

6.1 Grouped Transportation Needs

Medicare generally covers people who are above 65 or have a disability. To identify different groups in the Medicare plan that need additional support on transportation issues, we completed factor analysis on our top 20 performing features (See Appendix A) and used K-means clustering to identify groupings according to different key factors.

We identified the following 5 most important factors:

"Ambulance_Usage": the necessity and frequency of ambulance usage

"Mobility": the chance of a member having motor vehicle or mobile home

"Disabled_Stress": a member's level of disability and related mental health conditions

"Lowincome_Stress": a member's level of low income status and related financial stress

"Credit_Status": a member' credit status including the number of days past due and total credit balance

We used K-means clustering to group all samples into different numbers of clusters based on the factors chosen and presented cluster visualizations in 2D. Larger values correspond with stronger correlations with positive values representing a positive correlation and negative values representing a negative correlation. Below are the four groups we identified with the most transportation issues:

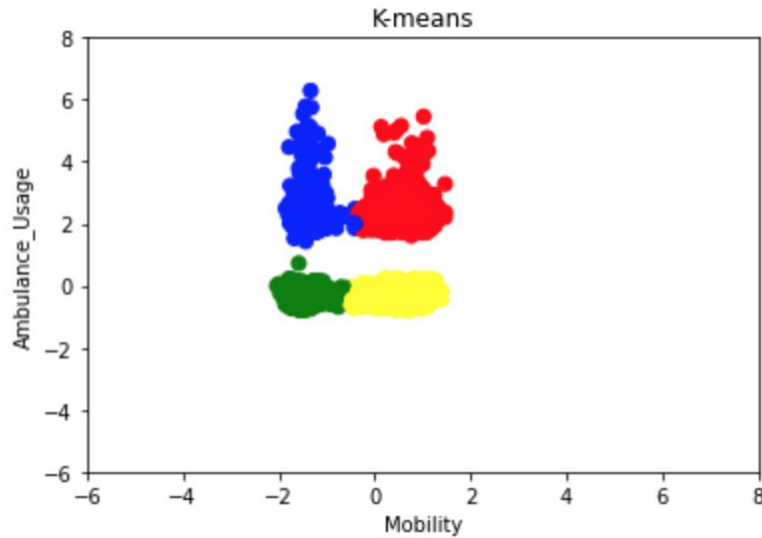## Group 1: High demand for ambulances with low mobility

Figure 6.1 High Demand for Ambulances and Low Mobility. The group in color blue is the group of members who have high demand for ambulances but don't have motor vehicles for transportation.
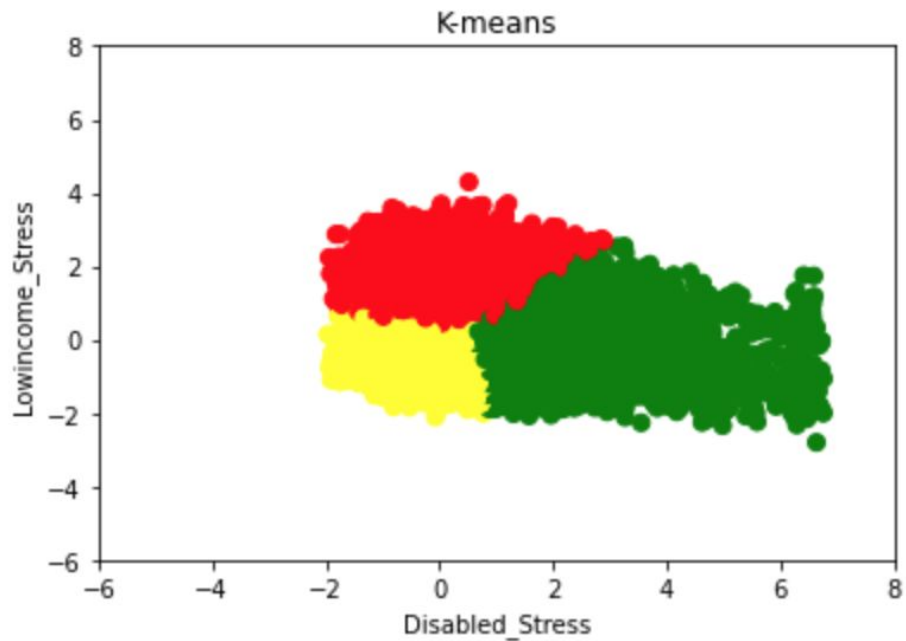
Group 2: Low income with disability

Figure 6.2 Low Income with Disability. The group in color red is the group of people who have low income and disability conditions.

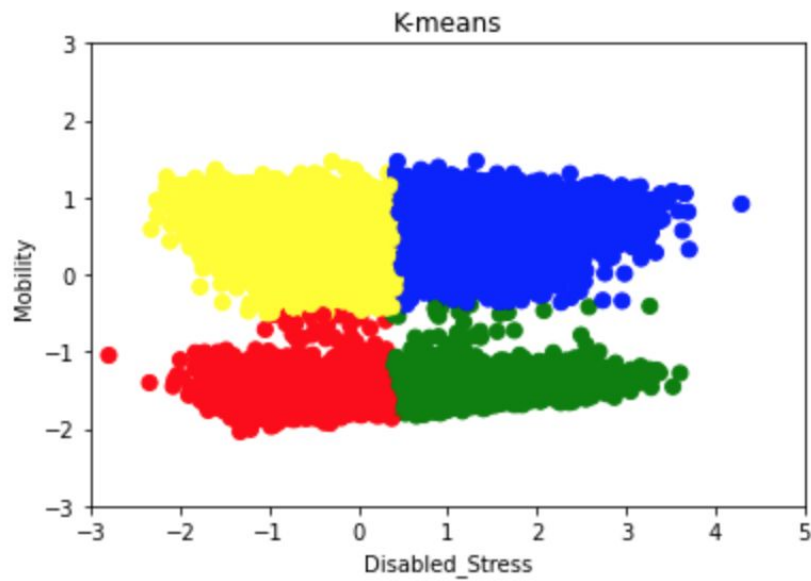## Group 3: Disability with low mobility

Figure 6.3 Disability with Low Mobility. The group in color green is the group of people who have disabilities and don't have motor vehicles for transportation.
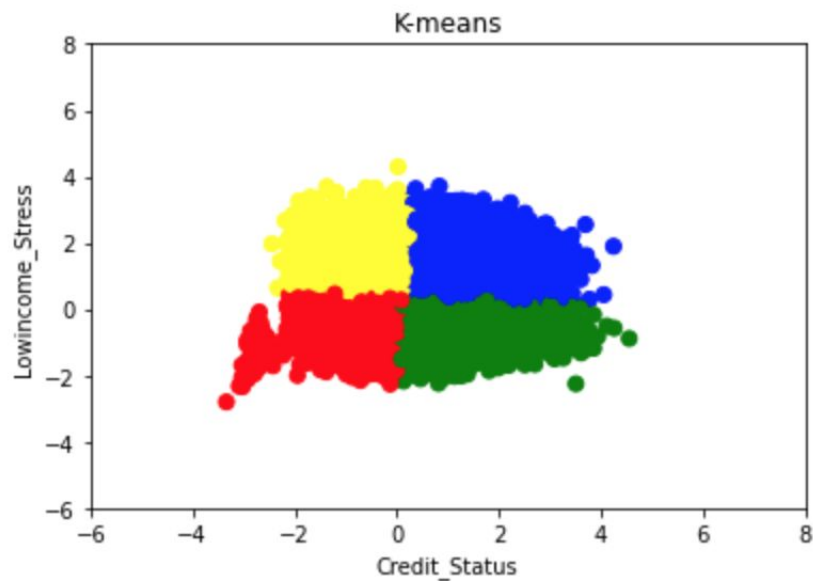
Group 4: Low income with credit stress

Figure 6.4 Low Income with Credit Stress. The group in color blue is the group of people who have low income and credit balances that are overdue

6.2 Recommendations for Different Groups

Based on the four Medicare member personas in the previous section, we came up with four recommendations with specific support for each group:

Recommendation 1: Use Marketing Campaign to Increase Telehealth Adoption

Telehealth, which allows remote treatment through telecommunications systems, is our primary solution to remedy transportation barriers for healthcare members. Current Medicare telehealth services include virtual office visits, psychotherapy, consultations, and other medical or health services depending on an eligible provider. While telehealth is becoming a more common practice during the COVID-19 public health emergency, patient usage rate is often underutilized. According to ASPE, only 43.5% of Medicare primary care visits were provided via telehealth in April 2020, compared with less than one percent before COVID-19 in February (0.1%). Moreover, providers in rural counties had smaller increases in Medicare primary care telehealth visits compared with providers in urban areas (ASPE, 2020). This suggests while current telehealth usage is largely driven by concerns with COVID-19, patients and providers are ready for its more widespread adoption.

Therefore, we recommend using marketing campaigns to familiarize and educate people with telehealth and its benefits, especially in rural areas, to encourage increasing and continuing the usage of telehealth even past pandemic. Possible social marketing efforts include community tabling, flyers and instruction manuals through direct mails, and social media advertising.

We establish a marketing budget of 5% out of the expected 18 million lost from missed appointments, leading to a campaign cost of $900,000. Our participation increase estimates a 20% increase of utilization by transportation issues individuals, and resulting in a 50% decrease of missed appointments from this new participating group. While studies show that telehealth meetings are generally cheaper, an increased utilization keeps the total costs constant. However, this new telehealth system would mean that up 3.6 million dollars would no longer be lost from the missed appointments. Additionally, studies have supported as much as a 65% decrease in ER visits due to consistent telehealth checkups, using a conservative 10% decrease in multiple visits to ER resulting in an annual savings of 2.4 million dollars.

Marketing Cost: -$900,000

Net Reduced ER visit Savings: + $2,400,000

Recommendation 2: Enhance Prescription Medicine Delivery Experience

Prescription medicine delivery service, which offers home delivery per prescription refill requests for patients, is an effective way to accomodate people with transportation issues and saves them the trouble of visiting pharmacies each month. As prescription medicine home delivery is increasingly utilized during the current COVID-19, several problems and concerns have been discovered. For example, delivery delays on critical prescription medicine are experienced nationwide due to the USPS disruptions, which could cause serious problems for customers who run out of their medication. However, their pick-up options are sometimes limited due to specific pharmacy locations or additional transportation expenses needed.

To prevent such problems in the future and resolve the general concerns about prescription mail delivery, we recommend Huamana to partner with ride-sharing or delivery pick-up Apps as alternative delivery options to mail delivery for customers who need faster delivery in urban areas. For customers in rural areas who don't have access to Apps delivery but need immediate delivery for their medicine, we believe offering secure, next-day delivery, such as USPS Priority Mail Express or FedEx Standard Overnight, for free would be a good alternative to normal USPS ground delivery. To reduce the costs from offering faster deliveries, we recommend Humana only offering such service to Medicare members on critical prescriptions who cannot afford any prescription delay.

Establishing partnerships to provide express shipping for medication can reduce these costs to an estimated $20 per package. By identifying individuals with transportation issues that have poor health or disabilities, we can offer free shipping on prescription refills in an attempt to reduce future ER and inpatient stays. We anticipate the need to send an additional 150,000 prescriptions packages for these individuals annually. Our expected reduction in Impatient and ER Visits for these individuals is 30%.

Maximum Shipping Costs: - $3,000,000

Net Impatient and ER Savings: + $3,150,000

Recommendation 3: Set up a Community Volunteer Network and Mobile Health Station

According to the ASPE, the general cost of an in-home telehealth unit is approximately $3,500, including a TeleStation, measurement device and web-based clinical review software. For most families, this amount of money is enough to stop them from considering telehealth as a healthcare option. Therefore, we recommend Humana provide a subsidy for a certain amount of in-home telehealth units for each community based on the number of patients and their usage. Humana could distribute telehealth units to different patients at different times depending on their needs.

To set up a telehealth unit at patients' homes, we recommend Humana partnering with local volunteer organizations, focusing on assistance from high school and college students. Considering most healthcare members are relatively elderly aged and might not be familiar with the latest technology, they will need some young volunteers to guide them through the telehealth process.

Besides doctor visits through telehealth, members will benefit from easily accessing and tracking their own biometrics, so they will have an idea of their overall health and when to consult with doctors. For members to do so, we recommend two methods to Humana based on the average income of each community. For communities with an average or high income, Humana could partner with Higi, a consumer health engagement company, to set up smart health stations at local drugstores to automatically assess members' general health. Members in this area with access to smartphones, could keep track of their biometrics and activity online. For communities with a relatively low income, Humana could set up mobile health stations to provide quick check-ups to its members.

Given the high capital costs, these in-home stations should be aimed towards communities with high percentages of low income individuals and rural communities. Without having a full understanding of Humana's healthcare network and locations, it is difficult to determine the best locations for these mobile stations, but we have allocated a total of 1,000 units to be

subsidised to be used in communities across the country. Partnering with healthcare organizations with existing home health solutions could further reduce future ER and Inpatient visits in rural/low income communities by 15%.

Volunteer Training and Health Care Costs: - $1,500,000 ($4,000,000 as initial capital investment)
Net Preventative Care Savings: $ + 8,700,000

Recommendation 4: Provide Enhanced Ride-share Services

To make transportation available to every healthcare member, Humana could partner with ride-sharing giants such as Uber or Lyft to provide Medicare members with a free, safe, and private drive. This option is especially helpful for individuals who are in low-income categories and provides them with more better health outcomes, which can reduce future health costs and decrease the amount of missed appointments.

This final solution focuses on covering ride-sharing for low-income individuals. By covering to and from the hospital for these low-income individuals with transportation issues in metro areas, we estimate that the preventative savings could reach up to 55% reduced visits to the ER and inpatient stays. In addition, the 1.65 million dollars would no longer be wasted due to these missed appointments.

Total Cost of Ride Share Program: - $3,000,000
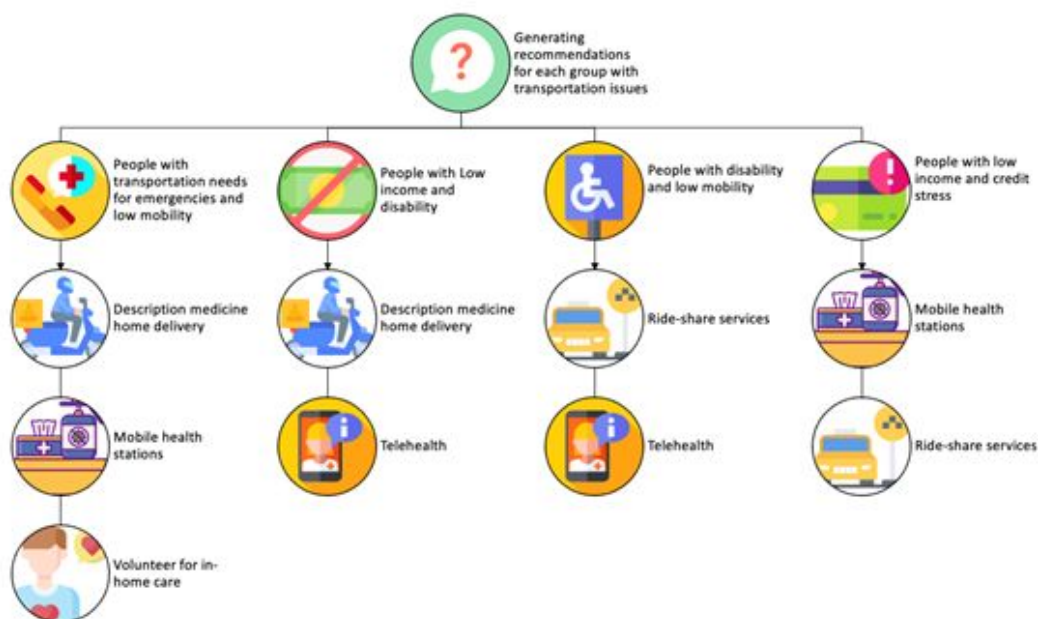Net Preventative Care Savings:+ 9,750,000

Figure 6.1 Recommendation break-down diagram

Below are our four detailed recommendations to each group:

Group 1: Members with transportation needs for emergencies and low mobility

For Medicare members with need for urgent care, we need to provide regular check-ups for them to give them a clear tracking of their overall health.

Group 2: Members with low income and disability

Among all the Medicare members, members with low income and disability might be facing the biggest challenge of physically accessing hospitals. Low-income status may also affect their access to necessary technology, such as smartphone or laptop, required for face-to-face telehealth. Therefore, we want to make sure they can receive the proper care and treatment in the comfort of their own home.

Group 3: Members with disability and low mobility

Medicare members with disability and low mobility also face transportation barriers. We need to assist them to set up in-home telehealth units so they can talk to doctors at home. Mail prescription might be necessary for them to stick to the prescription plan. When they need to go to hospitals for surgeries or other medical operations, they will need special

ride-share services because it is very possible that they are not able to drive even if they have a car.

Group 4: Members with low income and credit stress

Medicare members in this group do not experience as many physical mobility issues, but may suffer from their access to transportation. For this group, ride-sharing services can be provided for when they need to go to the hospital.

6.3 Financial Analysis

Transportation issues can contribute to three major costs that would financially directly affect Humana: the cost of missing appointments, the cost of failing to refill/continue taking prescriptions, and the potential future cost of missing/delaying healthcare.

According to a 2006 study, estimates range as high as a $700 cost to the healthcare system when an individual misses an appointment, leading to an annual cost to the total system at over 150 billion dollars (Harvard Business Reviews, 2020). A 2016 study estimates that hospitals experience up to 62 no-shows a year, with as many as 20% of appointments being attributed as a no show. Based on these estimates approximately 500,000 appointments are missed each year from Humana members. Doctor rates for missed appointments can run "as high as 20% to 30%" meaning a total of 18 million dollars could be wasted annually from these missed appointments.

Another key financial figure is the cost of failing to refill prescriptions as it is estimated that as many as 20% - 30% of individuals don't fill prescriptions, 50% don't take prescriptions as prescribed, accounting for a total cost to the US of over 100 billion to 289 billion per year in the form of future hospitalizations and estimated deaths (Nytimes.com, 2020).

The final financial factor, and perhaps the most costly, is related to missing initial health appointments due to transportation barriers can contribute to "2.6 times more likely to report multiple ER and 2.2 times more likely to report IP visit over a 12-monh period" (McKinsey, 2020). Based on a second ER visit rate of 6.5%, we can generalize that transportation issues account for an additional 400,000 ER visits each year. If we define one twentieth of these estimates as "true emergencies" (20,000 additional ER visits) which is covered under

Humana's plans, the total cost for these additional ER visits would be roughly 30 million per year and another additional 20,000 inpatient stays can range to another 20 million dollars annually.
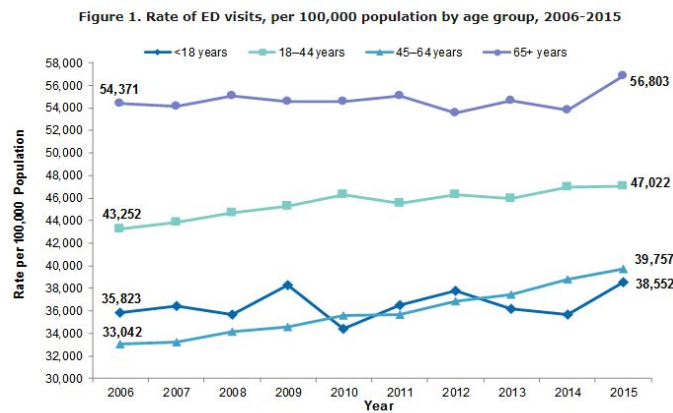

Figure 6.2 Amount of ER visits Per 100,000 People

We recognize that there is potential for overlap for each of these solutions' savings, so to account for this we reduce our total savings estimate by 35%.

Total Savings: 15.6 million dollars annually
Net Savings: 7.2 million dollars annually

Based on our generalized analysis we estimate that of the 68 million dollars lost for health outcomes due to transportation issues, 4.65 million would now be utilized for appointments. An initial investment of 4 million dollars along with another 8.4 million contributed annually can equate to a future savings of ~7.2 million dollars with the potential of further reducing future health costs.

7. Conclusion

Given the Humana competition training dataset of 69,572 observations and 826 columns, we constructed an XGBoost model, a gradient boosting algorithm, that predicted transportation issues with an AUC of 0.7508. Comparing this model to the baseline model, it is clear that our model provides meaningful insights into the likelihood of a Humana Medicare member experiencing transportation issues. The key features from our model corresponded to actionable recommendations towards alleviating transportation challenges that are faced by Humana Medicare members.

The key features and indicators of our model offered us interpretable insights that allowed us to formulate four persona groups. To tackle transportation issues of each group specifically, we proposed a unique combination of our four recommendations: 1. Use Social Marketing Campaign to Increase Telehealth Adoption, 2. Enhance Prescription Medicine Delivery Experience, 3. Set up a Community Volunteer Network and Mobile Health Station, 4. Provide Enhanced Ride-share Services. Our estimates indicate that utilizing these four recommendations can lead to a total of 15.6 million in savings from preventive care and with an additional 4.6 million no longer wasted due to missed appointments. These viable solutions will not only help Humana solve its members' transportation issues to a large extent, but also improve Medicare members' overall health in new ways, which exactly aligns with Humana's core business value.

8. References

Aspe.Hhs.Gov, 2020
https://aspe.hhs.gov/system/files/pdf/263866/HP_IssueBrief_MedicareTelehealth_final7.29.20.pdf
Accessed 11 Oct 2020.

"A Startup That Operates Health Kiosks in Grocery Stores Landed a $30M Funding Round Led by Babylon." *FierceHealthcare*
www.fiercehealthcare.com/tech/a-startup-operates-health-kiosks-grocery-stores-landed-30m-backed-by-babylon
Accessed 11 Oct. 2020.

Brownlee, Jason. "A Gentle Introduction To Xgboost For Applied Machine Learning".
Machine Learning Mastery, 2020
https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/
Accessed 11 Oct 2020.

"Gradient Boosting – What You Need To Know — Machine Learning — DATA SCIENCE".
DATA SCIENCE, 2020
https://datascience.eu/machine-learning/gradient-boosting-what-you-need-to-know/
Accessed 11 Oct 2020.

"Social determinants of health series: Transportation and the role of hospitals." *Chicago, IL: Health Research & Educational Trust,* 2017
http://www.aha.org/transportation
Accessed 11 Oct 2020.

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

"Home Health Aide: Educational Requirements." *Study.Com*
study.com/articles/Home_Health_Aide_Educational_Requirements.html

"How Behavioral Economics Can Help Cure The HealthCare Crisis". *Harvard Business Review*, 2020
https://hbr.org/2010/03/how-behavioral-economics-can-h
Accessed 11 Oct 2020.

"Insights From The Mckinsey 2019 Consumer Social Determinants Of Health Survey - Humana". *Humana*, 2020

https://populationhealth.humana.com/stories/insights-from-the-mckinsey-2019-consumer-social-determinants-of-health-survey/
Accessed 11 Oct 2020.

KIM, Donghwa. "Isolation Forest". *Donghwa-Kim.Github.Io*, 2020
https://donghwa-kim.github.io/iforest.html
Accessed 11 Oct 2020.

Providers, Humana, and Provider Covid-19. "COVID-19 Telemedicine - Humana".
Humana.Com, 2020
https://www.humana.com/provider/coronavirus/telemedicine
Accessed 11 Oct 2020.

Reddy, Madhuri. "Professional Caregiver for Older Adults - Part I: Responsibilities."
Blog.Careacademy.Com
blog.careacademy.com/what-does-it-mean-to-be-a-professional-caregiver-of-elders-part-i-responsibilities
Accessed 11 Oct. 2020.

Ryvicker M, Bollens-Lund E, Ornstein KA, 2020, "Driving Status and Transportation
Disadvantage Among Medicare Beneficiaries. Journal of Applied Gerontology",
2020;39(9):935-943, doi:10.1177/0733464818806834.
Accessed 11 Oct 2020.

Srinivasan, Venkat. "Some Of The Most Effective Health Clinics Are On Wheels". *Slate
Magazine*, 2020
https://slate.com/technology/2015/05/mobile-health-clinics-outcomes-community-relationships-privacy-concerns-success-rates.html
Accessed 11 Oct 2020.

Wallace, Richard & Mull, Hillary, et all, 2005 "Access to Health Care and Nonemergency
Medical Transportation", *Transportation Research Record: Journal of the Transportation
Research Board*
https://journals.sagepub.com/doi/abs/10.1177/0361198105192400110
Accessed 11 Oct 2020.

"The Cost Of Not Taking Your Medicine (Published 2017)". *Nytimes.Com*, 2020,
https://www.nytimes.com/2017/04/17/well/the-cost-of-not-taking-your-medicine.html
Accessed 11 Oct 2020.

How Do Mobile Health Clinics Impact Patient Access to Care?
PatientEngagementHIT
https://patientengagementhit.com/news/how-do-mobile-health-clinics-impact-patient-access-to-care Accessed 11 Oct 2020.

9. Appendix

Appendix A (Top 20 Most Important Feature Descriptions)

Est_age	Log Transformation of a Members Age
StressIndex	Weighted Score combining demographic, health, and credit information to identify a members level of stress
total_physician_office_visit_ct_pmpm	Per Member Per Month total amount of physician claims
cms_disabled_ind	Binary Disability Indicator
cms_risk_ma_nbr_rx_combined	Combined CMS risk and prescription risk score
betos_m5_pmpm_ct_rank	Percentile Rank of members Per Member Per Month Usage of Speciality Physician
anomaly_score	Likelihood that a member is an anomaly
cms_ma_risk_score_nbr + cms_tot_partd_payment_amt	Combined Risk Score and Total Part D payment from CMS categories
total_ambulance_visit_ct_pmpm + total_er_visit_ct_pmpm	Combined total Per Member Per Month ambulance and ER visit claims
rx_mail_pmpm_ct	Total mail prescription claims per member per month
submcc_ben_othr_pmpm_ct	Neoplasms Other claims per member per month
mabh_seg_H2	Binary Indicator if member belongs to segmentation of Medicare Segmentation Group H2
ccsp_239_ind	Binary Indicator for CCS code Superficial Injury, Contusion
rx_overall_pmpm_ct	Per Member Per Month Overall prescription claims
count_ner_rank	Percentile Ranking of Total Nervous System Claims
cons_n2pmv	Census Data of Percent of Motor Vehicle Ownership

cons_retail_buyer	Binary Indicator if member is a retail buyer
med_ambulance_visit_ct_pmpm	Medical Ambulance Per Member Per Month Claim
cms_risk_adjustmnet_factor_a_amt_rank	Percentile ranking of members Risk Adjustment Factor Type A amount
cons_hhcomp_Min Two People, No Children	Binary Indicator if member belongs in a household with two people, no children

Appendix B (Explained Variance From PCA

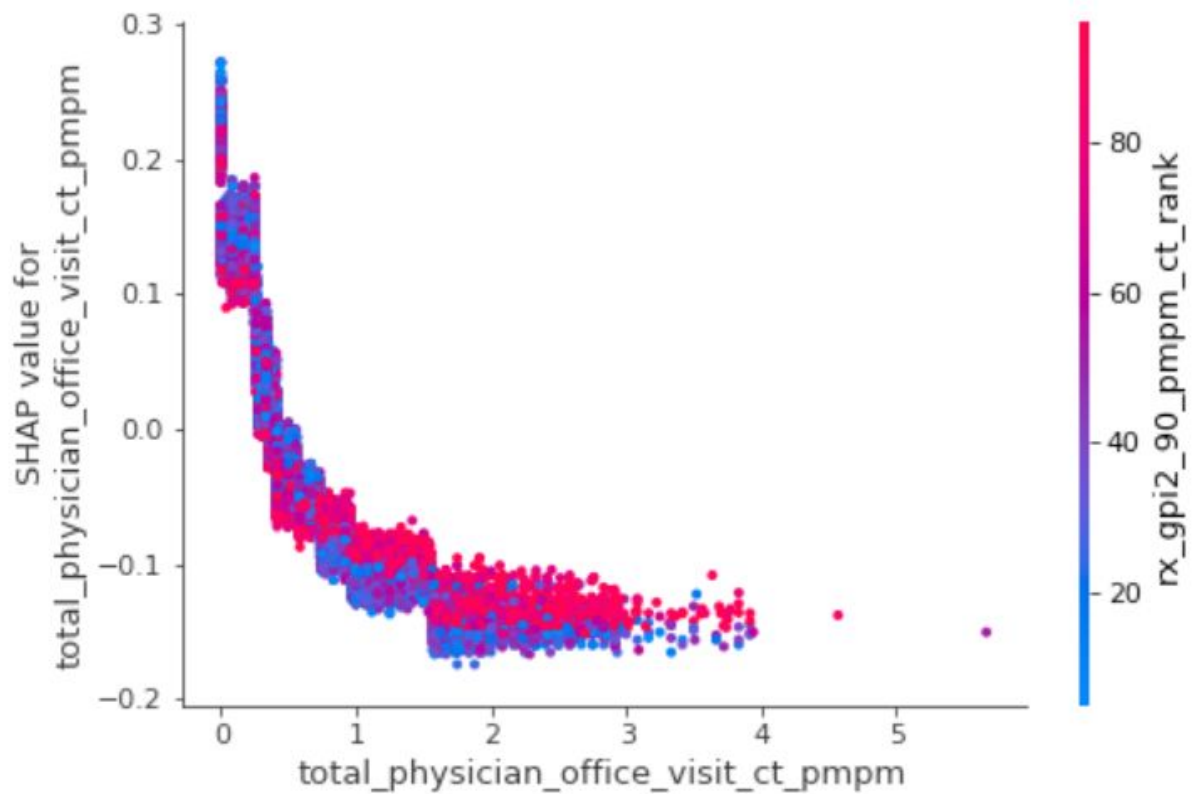| Category	Explained Variance
Credit Balance	0.861
Credit Accounts	0.947
Credit Percent of Household	0.911
SUBMCC	0.909
RX	0.935
Betos	0.865
CMS	0.979

Appendix C (SHAP Dependency Plots)

Figure C.1 Total_Physician_Office_Visit_CT_PMPM Dependency Plot. The graph demonstrates that as physician office visits increase, there is a decrease in the expected transportation issues probability. Physician office visits interacted the most with RX 90, dermatology prescriptions.
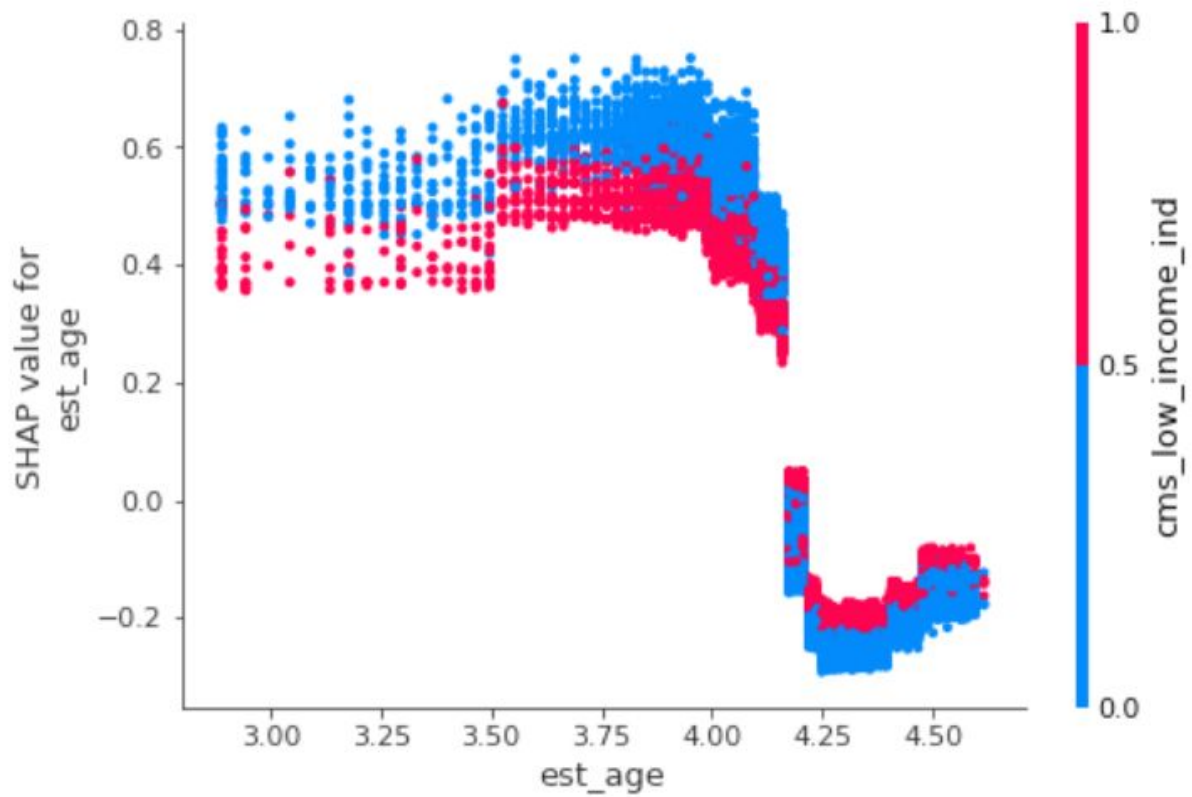
Figure C.2 Est_Age Dependency Plot. The graph demonstrates that as estimated age increases, there is a decrease in the expected transportation issues probability. Estimated Age interacted the most with low income indicators.
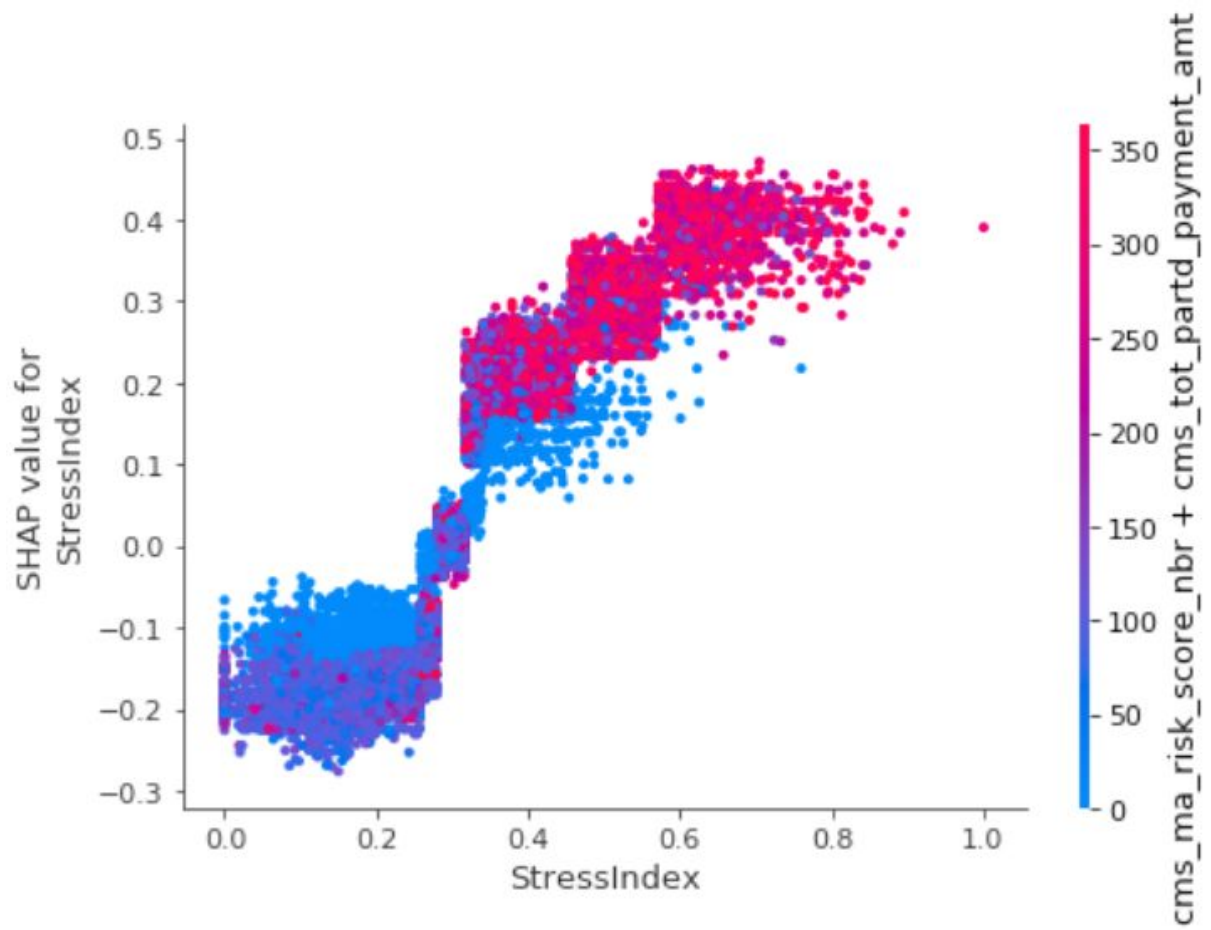
Figure C.3 Stress Index Dependency Plot. The graph demonstrates that as estimated Stress increases, there is an increase in the expected transportation issues probability. Stress Index interacted the most with the combination of CMS risk factors and payment amount.
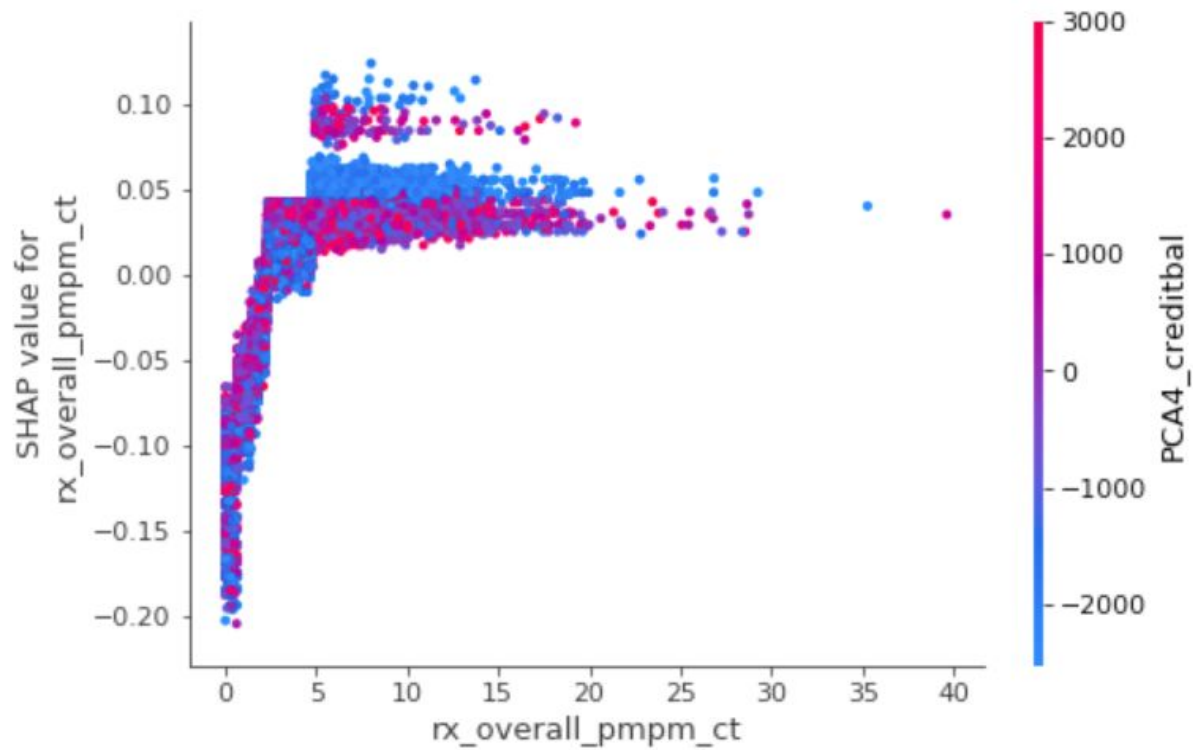
Figure C.4 RX Overall PMPM Ct Dependency Plot. The graph demonstrates that as prescription claims increase, there is an increase in the expected transportation issues probability. Total prescription claims correlated the most with a PCA breakdown of credit balance.