Humana/Mays2020
Healthcare Analytics Case Competition

TRANSPORTATION CHALLENGE PREDICTION ANALYSIS

Table of Contents

1. Executive Summary

This analysis focused on providing Humana a better understanding of transportation as a social determinant of health. Our goal was to develop a classification model to identify Medicare members most at the risk for a Transportation Challenge and propose solutions for them to overcome this barrier to accessing care and achieving their best health.

Since our target is a binary feature, we implemented classification models. After thoroughly understanding each feature and its relationship with the target feature - **transportation_issues**, we reduced the number of features from 826 to 138. The shortlisted features belong to different categories like medical claims, pharmacy claims, CMS, demographics, and condition related.

We used Google Cloud Platform as our computation engine, Tableau for data visualization, Python for data manipulation and model building, RStudio (R Programming) for statistical analysis and H20 driverless AI for training multiple models. After testing various models, we found the Light GBM machine learning model to perform best on both train and validation data. **The model resulted in an AUC of 0.77 with 72.5% accuracy on a balanced class data set**.

2. Background

The social and economic environments in which people live are inextricably related to health and well-being. Research has shown that medical care can be linked to just 20 percent of health, while social and economic factors, such as access to nutritious food, housing status, educational achievement and access to transport, account for 40 percent.

Transportation is an economic and social factor that shapes people's daily lives and thus a social determinant of health. It connects people from their origin to their destination, affects land use and shapes our daily lives. It is necessary to access goods, services and activities such as emergency services, health care, adequate food and clothing, education, employment, and social activities. It also can be a vehicle for wellness. Developing affordable and appropriate transportation options, walkable communities, bike lanes, bike-share programs and other healthy transit options can help boost health.

Transportation barriers can affect a person's access to health care services. These barriers may result in missed or delayed health care appointments, increased health expenditures and overall poorer health outcomes. These statistics highlight the scope of the problem:

- 3.6 million people in the U.S. do not obtain medical care due to transportation barriers
- Regardless of insurance status, due to unavailable transportation, 4% of children (approximately 3 million) in the U.S. miss a health care appointment per year; this includes 9% of children in families with incomes of less than $50,000.
- Transportation is the third most commonly cited barrier to accessing health services for older adults.

There is a growing recognition that improving transportation access and support for members can help improve health outcomes and lower health costs.

3. Data Understanding

3.1. Data Summary

To perform our analysis, we were given a training dataset of 69572 rows and 826 columns. The data was collected over a span of 1 year for members enrolled in the Humana Medicare Advantage and Prescription Drug (MAPD) plan. The data included medical claims, pharmacy claims, lab claims, demographics, consumers, credit, clinical conditions, CMS member data, and other member-related features.

3.2 Target Feature

Our goal was to predict if the members are likely to struggle with transportation issues. The target feature was collected as a response to the transportation screening question - ***"In the past 12 months, has a lack of reliable transportation kept you from medical appointments, meetings, work or from getting things needed for daily living?"*** This information is captured in the **transportation_issues** binary feature where 0 indicates the member faced no transportation challenge and 1 indicates the member faced a transportation challenge.

From the distribution of this feature, we can observe that 85% of the members do not face transportation issues which creates a class imbalance problem and should be taken care of while building the predictive model. The document also talks about the descriptive analysis of the dataset.
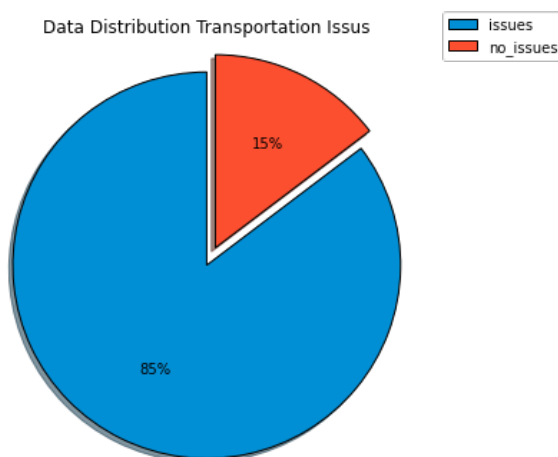
4. Feature Engineering and Exploratory Data Analysis

The data mainly included medical claims, pharmacy claims, lab claims, demographics, consumers, credit, clinical conditions, CMS member data, and other member-related features. This section discusses the inferences, features generated and interesting observations or patterns that are derived in the above-mentioned categories.
Since, the data is augmented, the derived conclusions are limited and may not represent the real-world scenario.

Instead of directly using all the features, we did thorough research and chose the features which we could understand, explain and relate to the final target feature. We believe that this will result in a richer data set that can support the model with fewer parameters.

There are two key types of features across multiple categories:

    a. IND - Binary indicator for different scenarios (Yes - 1/ No – 0)
    b. PMPM – Utilization per member per month calculated using total utilization by the number of months a person stayed with Humana in the past 12 months.

4.1 Medical Claims

Medical claims have 3 sub-categories of features:

4.1.1 CCS (Clinical Classifications Software) Procedure Codes

These features break down diagnosis and procedures into manageable categories that make statistical analysis and reporting easier. The official CCS website helped us to understand that the feature definition of these codes corresponds to the diagnostic code of the CCS and not the processes of the CCS. The data includes 21 such CCS codes. We dropped ccsp_034_ind and ccsp_0120 codes as they had no data. Since these member diagnoses can have an effect on the extent to which members may face transportation problems, we have used few of the features for our model building.

New features:
- A binary indicator - whether the member used any of the CCS service
- Total number of diagnoses services a member has availed

Descriptive statistics:
- 15% of the members had superficial injury and contusion (ccsp_239_ind), and 30% of them faced transportation issues.
- 72% of the members availed at least one of the CCSP services (out of 21) and 15% of them had transportation issues.
- The plot below shows the frequency of top services used in CCSP category

Service usage count for top categories
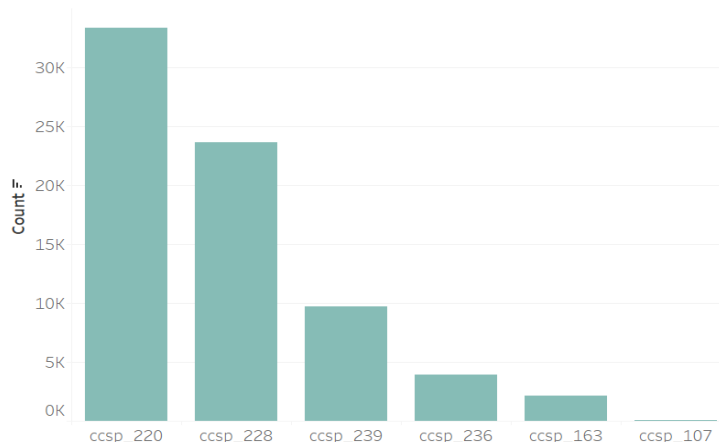
## 4.1.2 BETOS (Berenson-Eggers Type of Service) Procedure codes

BETOS offers information on one of the seven types of services used to track member spending and is used for medical expenditure growth analysis. The type of tests conducted on the member provides insights into the medical conditions that may affect chances of facing transportation issues. There are seven test types (Evaluation and Management, Procedures, Imaging, Tests, Durable Medical Equipment, Other) out of which procedures and imaging does not contain any data. We used some of the default features along with newly curated ones to aggregate data from lower to a higher hierarchy.

New features:

- ind_sum - number of services the member used out of all the services
- pmpm_sum - sum of PMPM for all the services the member used in each category.
- yes_no - binary indicator - did the member use any service in each category
- yes_no_sum - total services a member used across all the categories
- pmpm_total_sum - aggregated PMPM of a member across all the categories

Descriptive statistics:

- Member id d97e73MOSe86T74L0Y5A8Ifb who does not face transportation issues has used 16 services in 5 categories that summed to 15.27 PMPM.
- Evaluation category has maximum PMPM sum of 58935

Bubble chart - sum(pmpm) in each BETOS category
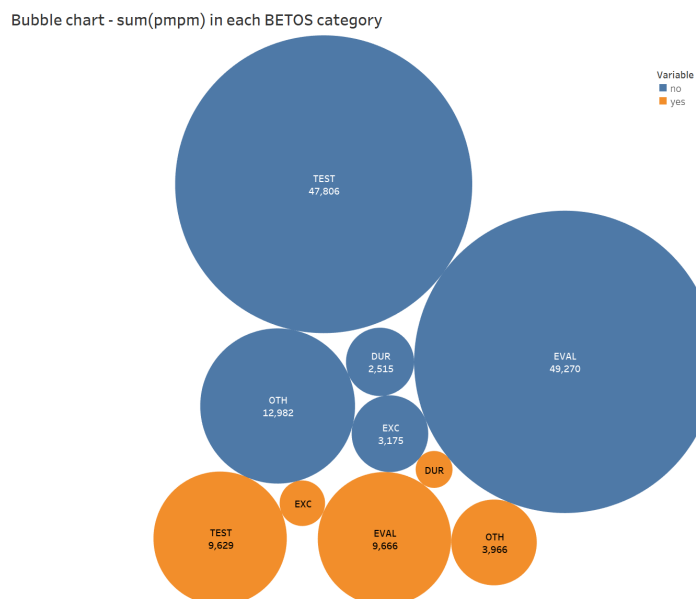
4.1.3 Utilization by category

These features give information about visits (ambulance, urgent care, emergency rooms, physician office, outpatient), hospital admit count and days for chronic, acute and maternity health conditions. The features are spread across two categories - medical and total. We have only focused on the total category as it covers both medical and non-medical categories.

New Features:
- Binary indicator - if a person utilized any type of services (derived by pmpm value)

Descriptive statistics:
- pmpm cost is high for outpatient and physician services rather than ambulance service
- 13% of members used ambulance service and 30% of them faced transportation issues
- Total amount of pmpm spent on physician services across all the members is 43983

PMPM sum of top categories in Utilization
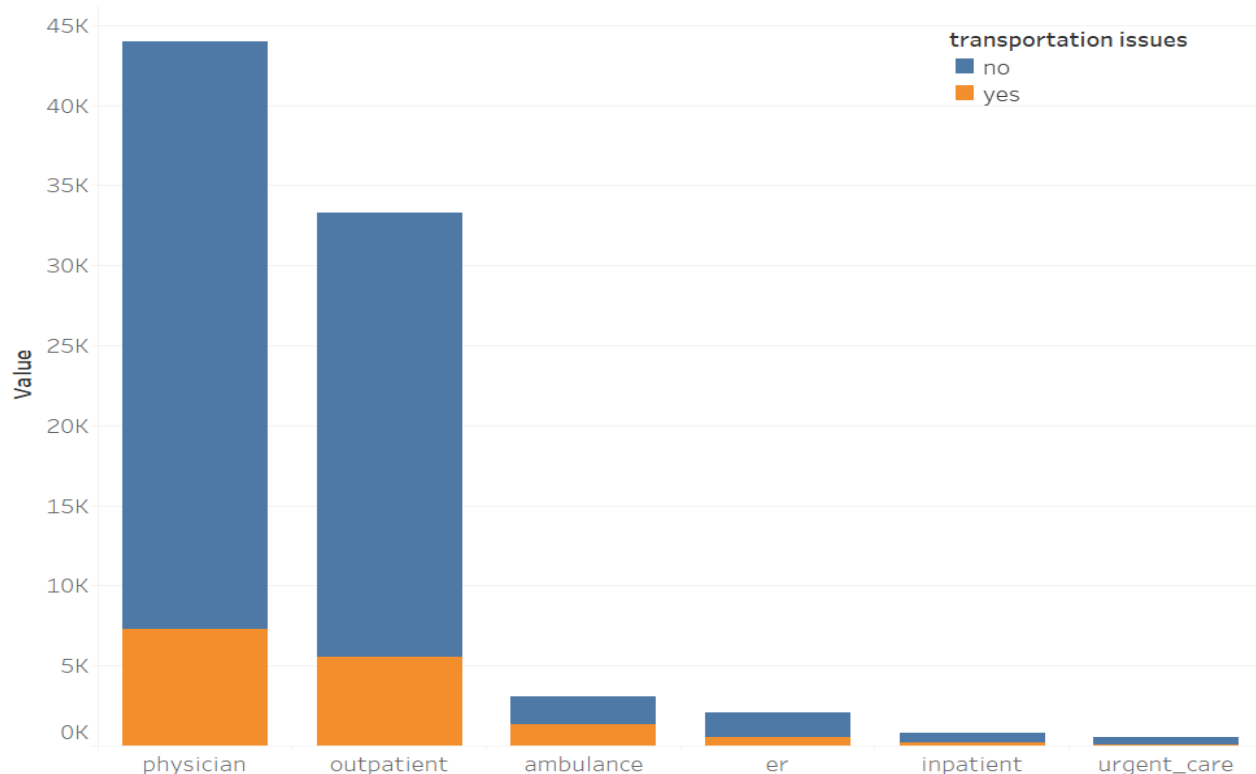
4.2 Pharmacy Claims

These features provide details about 100 different drugs used by the members. It also provides information on prescription and delivery of drugs - whether the prescription was branded or generic or whether the drugs were mailed. After discussing with a pharma expert, we categorized all the drugs based on their health/illness purpose into these 18 categories:

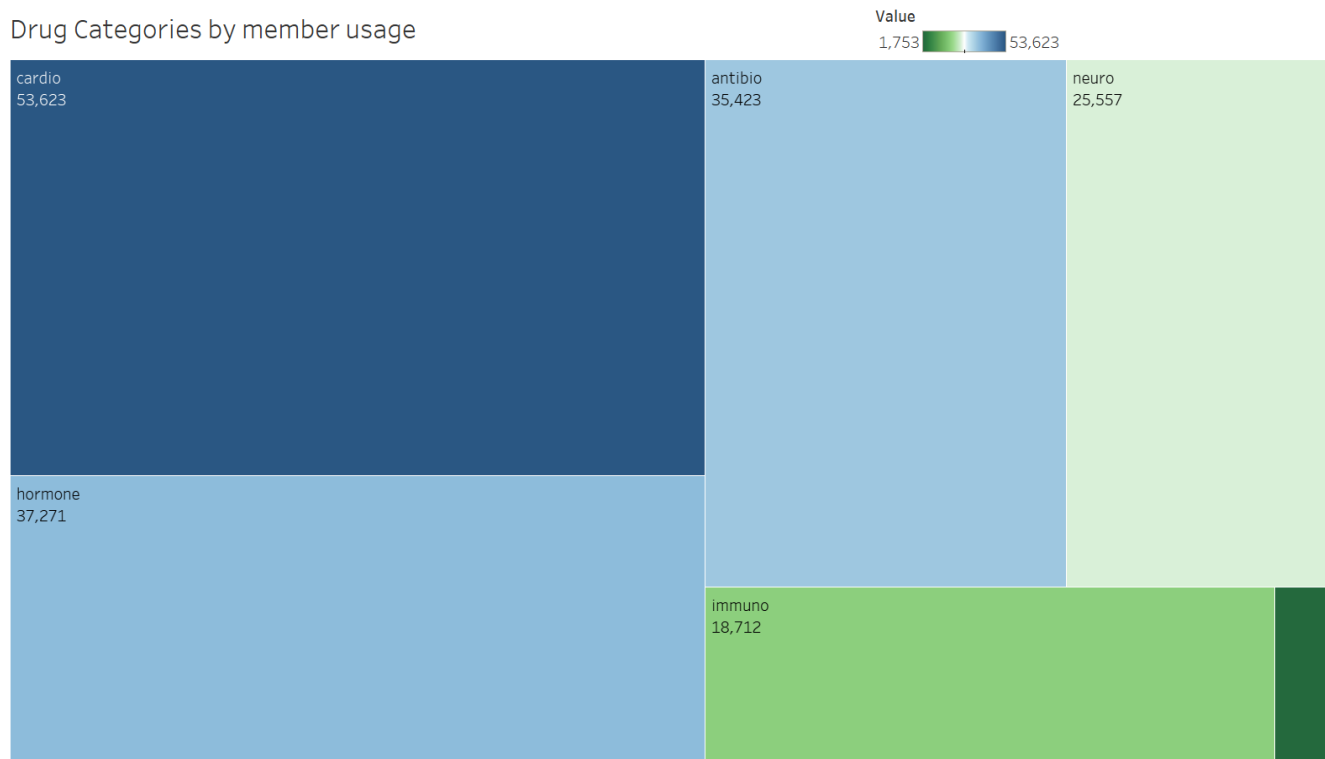|		
Antibiotics	Immunomodulators	Cancer
Hormonal	Cardiovascular	Kidney
Allergic	Respiratory	Intestinal
Neuropharmacological Type 1, 2	NSAIDS	Anesthetics
Hematopoietic agents	Hematopoietic agents	Blood Related
Eye	Ear	Others

New features:
- ind_sum: Aggregated all the indicators in each group
- pmpm_sum: Aggregated pmpm in each group
- yes_no - binary indicator - did the member use drug in any of the 18 categories
- yes_no_sum - Total categories of drugs used

As we had a lot of features, we clubbed the newly generated features with the original data and ran an XGboost algorithm to identify the most relevant features for model building. They are listed in the appendix.

Descriptive statistics:
- The person id 225de3M2OSd6ea8T54LYA1I9 used 15 out of 18 drug categories
- Member id 2Mb6OcdS9fTc711a24LdYAI5 used 33 different types of drugs
- Average pmpm count is higher for cardiovascular and antibiotics drugs than others.
- The plots below display pmpm cost per member across top categories

Drug Categories by member usage

Value
1,753 ■■■ 53,623

| cardio 53,623 | antibio 35,423 | neuro 25,557 |

| hormone 37,271 | |

| | immuno 18,712 | |

4.3 Demographics or Consumer Data

A lot can be derived from the personal information of a member and socio-demographic information that can have a practical effect on the likelihood of facing transport problems. Due to the presence of high NULL values and the inability of imputation models on member level data, we discarded many features that we believe would be useful and selected age, sex, current smoker, former smoker, rural and urban category features from the data.

New features:
- Binary indicator - urban or not (combined multiple levels of data)

Descriptive statistics:
- Mean age of members facing transportation issues (66 years) is distinctively different from members who don't (71 years). Ideally, we would expect older people to face transportation issues.
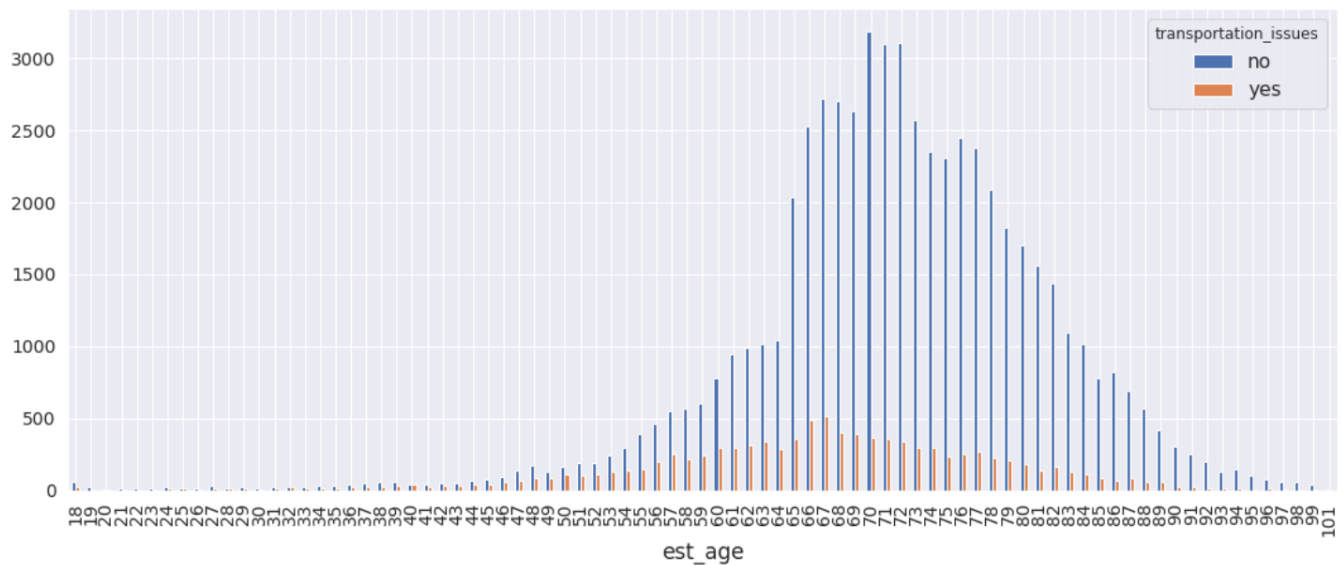
- Percentage of members (in different categories) facing transportation issues:

	% Members having Transportation issue
Male	7.4
Female	6.4
Current Smoker	4.2
Not A Current Smoker	7.5
Former Smoker	7.46
Not A Former Smoker No	6.7
Urban Region	6.7
Rural Region	7.3

4.4 Condition Related Features

4.4.1 Behavioral Health Condition Indicators

Mental health is essential to a person's well-being. These indicators shed light on health conditions like bipolar disorder, alcohol abuse, substance abuse, tobacco use disorder, major

depressive disorder, post-traumatic stress disorder and other anxiety disorders which can be directly linked to the possibility of having transportation issues. The graph below displays the total number of members present in each behavioral category:
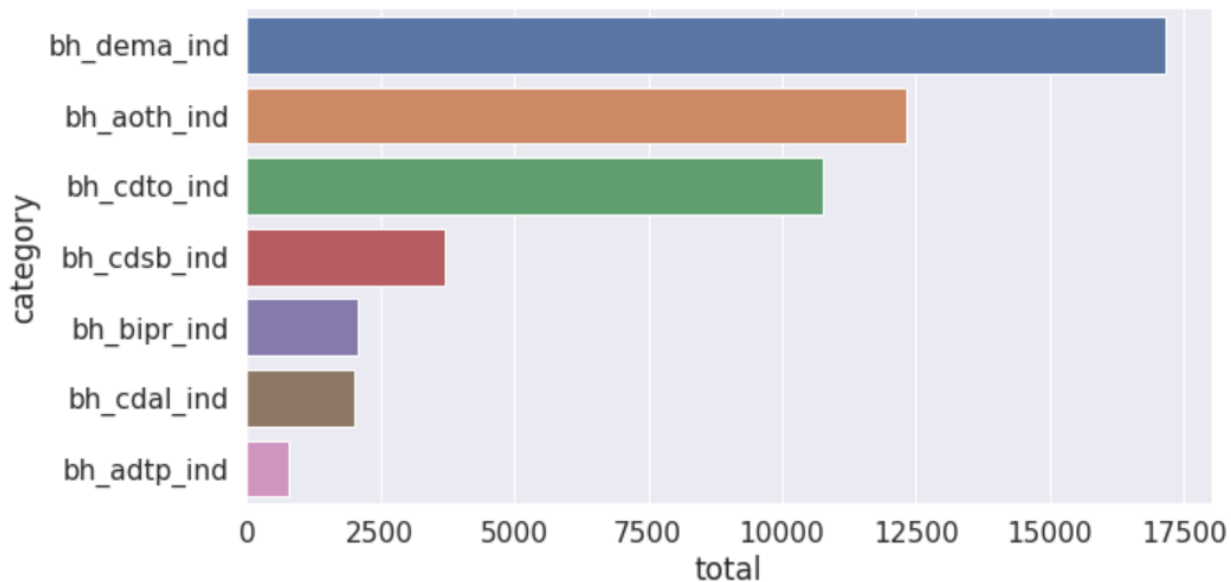
New features:
- Binary indicator - whether the member faced any behavioral issue
- Aggregation - total number of behavioral issues faced by the member

Descriptive Statistics:
- There are more behavioral issues for dementia and other anxiety disorders
- Three members (MO5S536T49e24f3L027Y37AI, bM4OfbS1Te347339LY13AIac and d4MOST0L18c896Y7A9Idc7d2) have all of the behavioral issues

4.4.2 CCI (Charlson Comorbidity Index), FCI (Functional Comorbidity Index) and DCSI (Diabetes Complication Severity Index)

The CCI scores are used to categorize the severity of comorbidity into three grades: mild (CCI scores of 1–2) moderate (CCI scores of 3–4) and severe (CCI scores ≥5). The FCI scores range from 0 to 18 and DCSI scores range from 0 to 13.
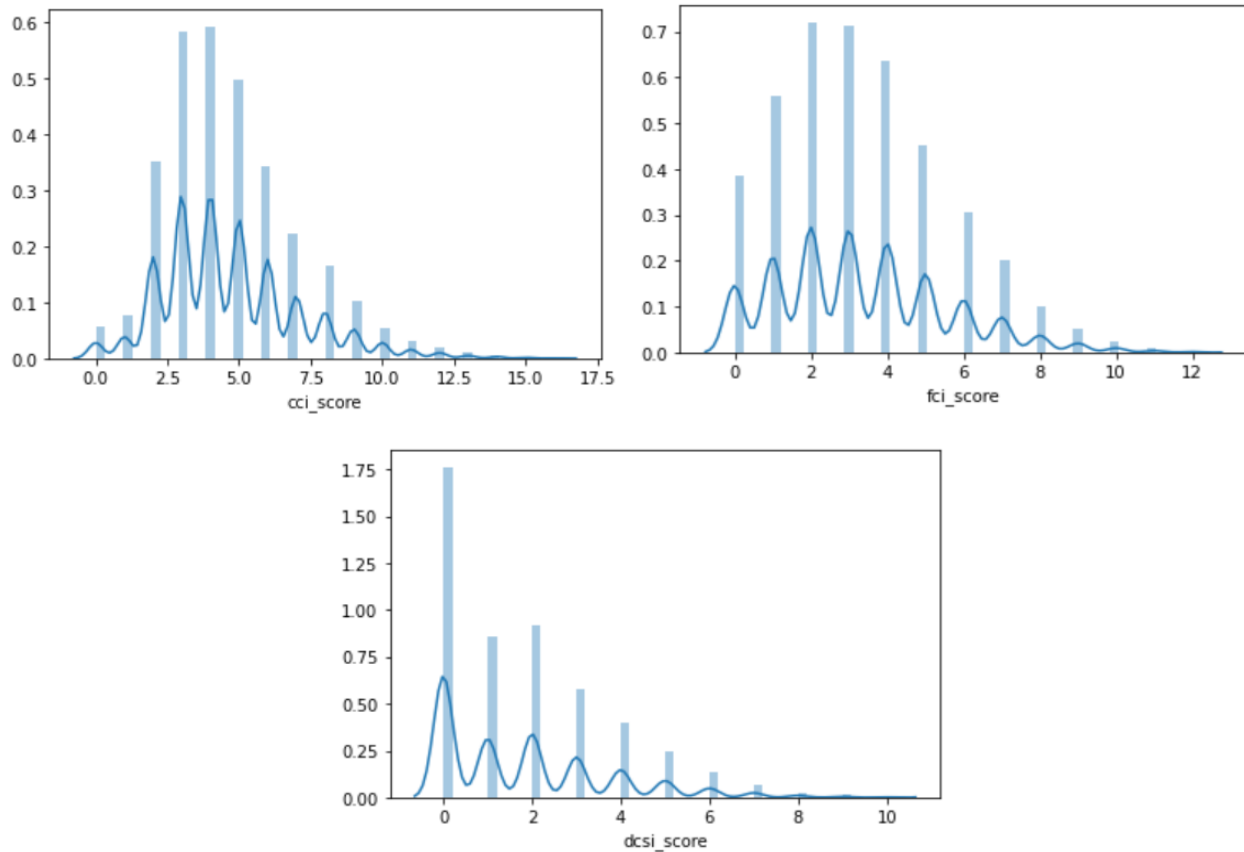
4.4.3 MCC Diagnosis Code Categories

Major clinical categories are created to bin the type of diagnosis a member received. According to the description file presented by Humana there are 28 different categories and there are ample members who received more than one type of clinical diagnosis. These provide insights to the type of diagnosis which in turn helps to find the members who are most probable of facing transportation issues. For example, if a member is diagnosed with spine issues, he/she might be facing the transportation issue.
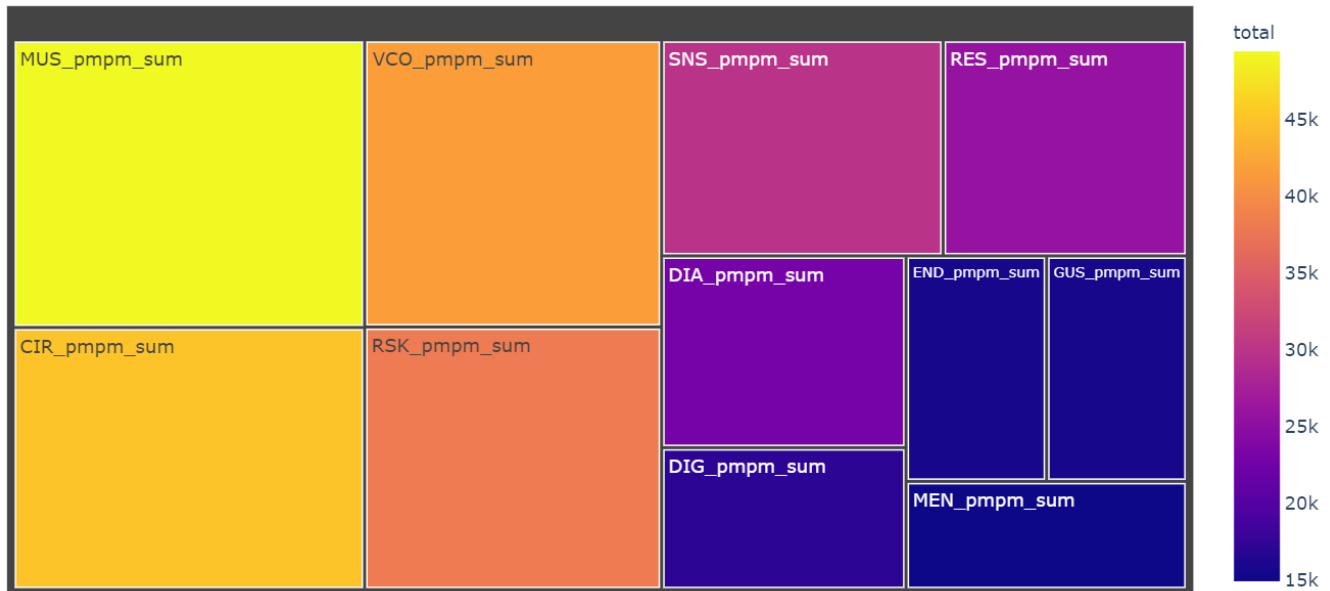
New Features:

As the data is granular and contains the specific type of diagnosis claimed by the member, we have created new features based on each category.
- Aggregating all the features in each category
- Binary indicator - whether the person availed diagnosis in each of the category or not

Descriptive statistics:

- Highest pmpm for a member is 73.03 (Member ID 18M71O8S48eT87dLY81A12I0)
- Highest diagnosis count is 82 for member fM3O00b0S6a12dTLYA032bIe
- The tree map below displays pmpm count greater than 40k across multiple MCC categories

4.5 CMS Features

The Centers for Medicare & Medicaid Services is a federal agency that oversees major healthcare programs. The organization seeks to provide better treatment, access to coverage, and enhanced health CMS features for the healthcare system. It provides interesting details on risk factors and indicators of members' eligibility for various programs.

New features:

The CMS features are divided into two categories CMS diagnosis and CMS risk score. We have created two set of features for each category:

- Aggregation - total number of services availed by a member in each category
- Binary - Whether a member has availed a service in at least one category

We replaced NULL values in some of the numeric features with mean and categorical columns with mode of the column of training data on both training and validation data.

Count of CMS features across issue and no issue
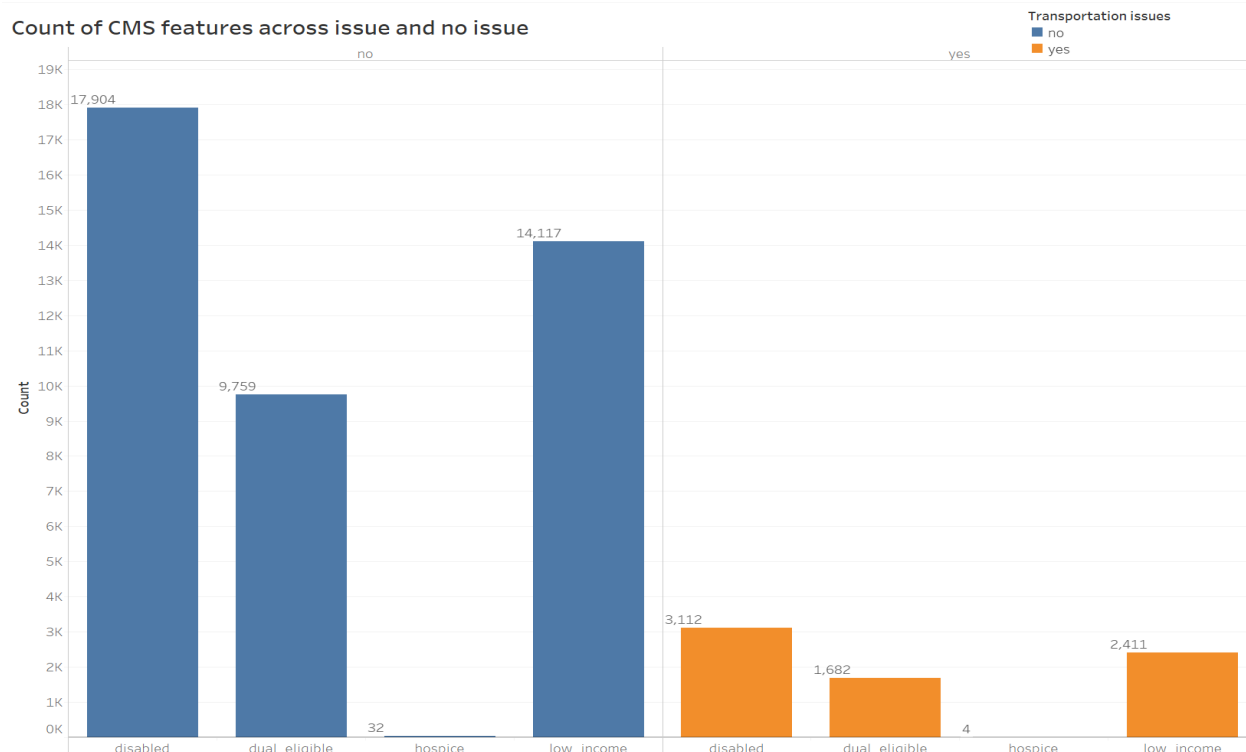
5. Model Building

5.1 Handling class imbalance

The dataset is severely imbalanced with only 14.6% of members struggling with transportation issues. Models built with imbalanced data can lead to biased results. To enhance the performance of the model and correctly classify the minority class (members facing transportation issues), we inspected the following artificial data generation techniques.

1. Under sampling: Removing some observations of the majority class.
2. Oversampling: Adding more observation of the minority class using duplication
3. SMOTE (Synthetic Minority Oversampling Technique): Similar to oversampling, uses nearest neighbors algorithm to generate new and synthetic data.

Oversampling worked best for our dataset without causing bias-variance problems.

5.2 Model Building

While choosing the machine learning model, we considered the accuracy vs. interpretability trade-off. Since we are dealing with healthcare and finance data, we wanted to build a model which emphasizes on introspection, causal effects, and processes.

We followed a sophisticated well-designed process to build our model. After understanding each feature, we conducted exploratory data analysis, identified some interesting patterns and shortlisted useful features for our classification model. To prepare the data for model building, we encoded the categorical features and standardized the numeric features.

Instead of simply splitting the dataset into train and test, we used stratified k-fold cross validation to keep the same percentage of samples of target class for both training and validation. This significantly reduces bias and variance as we use most of the data for both fitting and validation.

We pipelined the features into machine learning and statistical models to check the performance and choose the best model for prediction. We had a tradeoff between choosing different models, either to choose Whitebox models for clear understanding of impact of each feature on target feature or to choose Blackbox models which have high predictability and less understandability of features impact- to this end we did both. We have implemented LightGBM as the machine learning model and used the generalized linear model with binary distribution (logistic regression) to validate ML interpretations. We believe this has struck a balance of complexity and interpretability.

Light GBM is almost 7 times faster than XGBOOST and is a much better approach when dealing with large datasets. Light GBM is a fast, distributed, high-performance gradient boosting framework based on a decision tree algorithm, used for ranking, classification and many other machine learning tasks. The parameters used for the final model are colsample_bytree=0.8(Subsample ratio of columns when constructing each tree). learning_rate=0.02(The amount that the weights are updated during training), max_depth=8(Maximum tree depth for base learners), n_estimators=74(Number of boosted trees to fit), reg_lambda=1.0(L2 regularization term on weights), subsample=0.7(subsample

ratio of the training instance). The Logistic model summarizes the coefficients and standard errors of how a dependent feature varies when an independent feature is changed by a unit. The interpretation is in terms of log-odds and can be easily converted to probabilities.

Receiver Operating Characteristic

AUC = 0.77
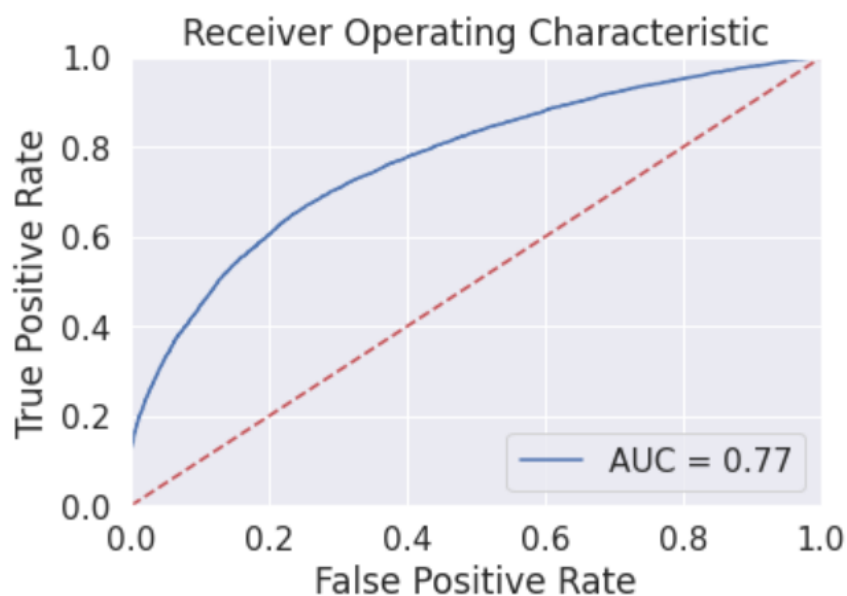
Accuracy=0.725
Precision=0.310
Recall=0.661
F1 Score=0.422
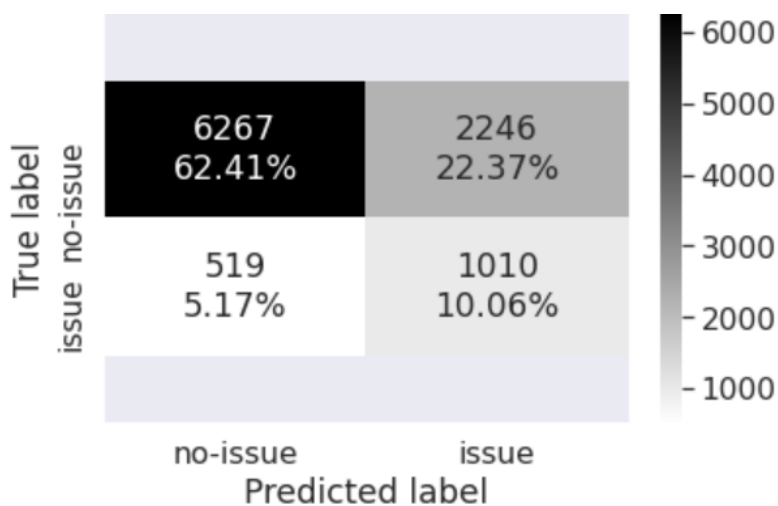
Implications from logistic model:

Feature	Estimate	Std. Error
(Intercept)	1.5060	0.0638
cms_tot_partd_payment_amt	0.0025	0.0001
est_age	-0.0319	0.0008
cms_disabled_ind1	0.0383	0.0179
ccsp_239_ind1	0.6879	0.0232
cms_rx_risk_score_nbr	0.0565	0.0121
betos_o1a_pmpm_ct_betos	0.1895	0.0345
cms_tot_ma_payment_amt	0.0001	0.0000
cms_low_income_ind1	0.0379	0.0273
bh_sum_ind	0.1516	0.0094
rx_mail_pmpm_ct_RX	-0.0410	0.0040
MEN_pmpm_sum_MCC	0.0912	0.0181
cms_partd_ra_factor_amt	0.0890	0.0110
total_ambulance_visit_ct_pmpm	0.6042	0.0601
total_physician_office_visit_ct_pmpm	-0.2018	0.0166
ccsp_220_ind1	-0.1058	0.0179
fci_score_MCC	0.0132	0.0044
EVAL_ind_sum_betos	-0.1469	0.0105
cms_risk_adj_payment_rate_b_amt	0.0000	0.0000
pmpm_sum_neuro_2_RX	0.2109	0.0190
cms_ma_risk_score_nbr	0.0988	0.0105

* Refer to Appendix for snapshot on summary details of logistic regression.

* All effects in the above model might not hold true in the real world.
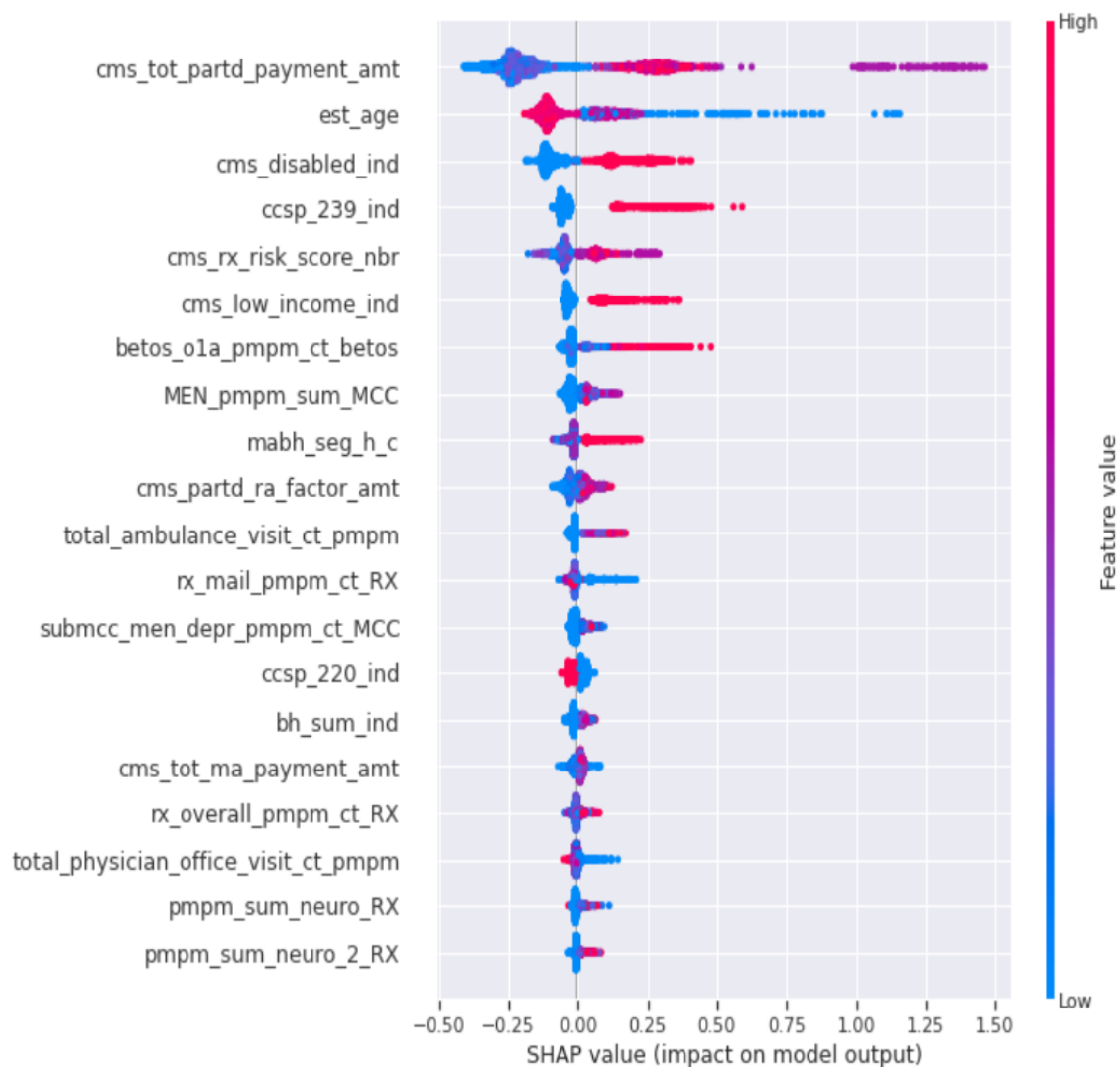
The Shapley plot gives clear output on the effect of each feature and the correlation. We can clearly see the cms_total_partd_payment_amt, cms_disabled_ind, ccsp_239_ind and betos_01a_ct_betos have high and positive impact on transportation issue (The high comes from the red color and positive impact shows from the X-axis). With increase in the value present in these columns the chances of a member facing the transportation issue goes up. On the other hand, est_age, ccsp_220_ind have negative correlation.

6. Insights and Recommendations

The summary of logistic regression and plots from SHAP were used to find insights and share relevant recommendations.

- cms_total_partd_payment is the payment made to private insurers for delivering prescription drug benefits to Medicare beneficiaries. This feature has a strong positive influence on transport problems. If the sum is higher than $200, the chances of having transportation difficulties would increase. This can be further supported by the statistical model, for every $100 increase in cms_total_partd_payment_amt, the log odds of transportation issue increase by 0.254.

- **Recommendation:** Humana, with the aid of pharmacy partners, can assist patients spending more than $200, by delivering the prescription medications to their home / work along with FFS (fee for service) options by partnering with Uber/Lyft. This will alleviate the transportation efforts made by patients and improve the level of service offered to individual customers.

- From the plot on the right, we can observe that the probability of predicting transportation_issues dependent feature increases with decrease in age and its more evident when age is less than 60. The statistical model output also aligns with this pattern i.e. log-odds reduces by 0.03 when age increases by a year. The Medicare services are provided to members whose age is greater than 65, certain young people with disabilities and people with end-stage renal disease.

- **Recommendation:** Younger members with disabilities and members with serious health conditions are at larger risk than few senior members among Medicare Beneficiaries. Providing disability services like in-home treatment will be beneficial.
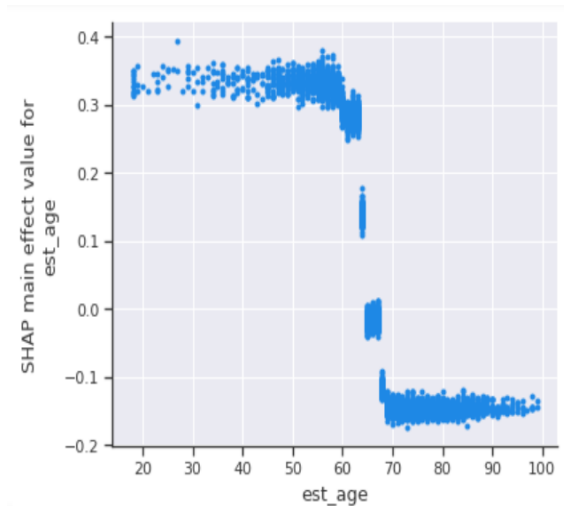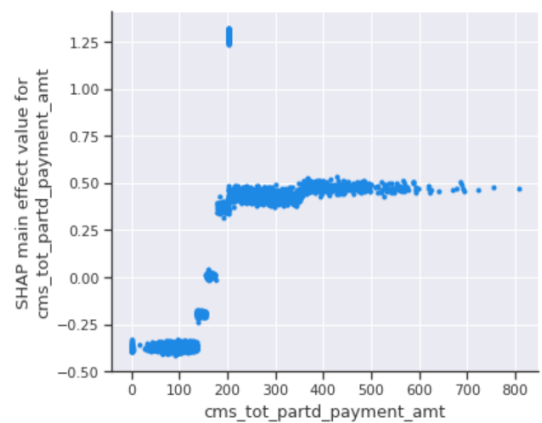
- Betos_01a_pmpm_ct_betos is the ambulance service code in the barenson-eggers type of service codes. The plot on the right clearly shows that the members who have pmpm count greater than 1, have higher chance of facing transportation issues and the log-odds increases by 0.189 with increase in betos_01a_pmpm_ct_betos by 1. The possible reason behind this can be that there is no reliable mode of transportation in their locality or the severity of their issue is high which needs frequent ambulance service.
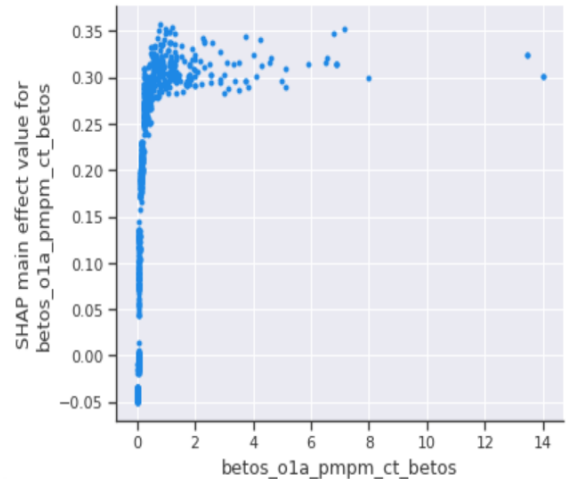
- **Recommendation:** Contact members to know the issue for which the ambulance service is used and embrace telehealth services, where the severity of the problem is low.
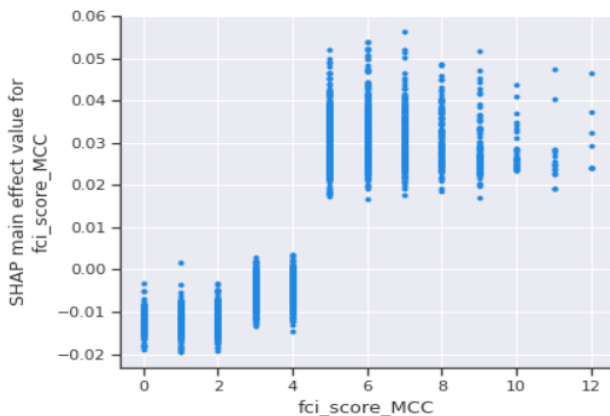
- Functional Comorbidity Issues (FCI), is calculated on 18 aspects which reveal the overall health status of a person. From the plot, it is evident that members having score greater than 5 have higher chance of facing transportation issues and with every increase in score, the log odds of having transportation issues increases by 0.013.

- **Recommendation:** The clear distinction of threshold helps to target members and introduce fitness plans and activities to indulge which can improve the health. These Health literacy initiatives can be achieved by partnering with fitness services.

- There is an interesting pattern with interaction of cms_risk_adj_payment_rate_b_amt and cms_diasbled_ind. If a member is disabled then the increase in predictability of transportation is almost zero irrespective of the increase in payment_rate_b (as most of the red points are on the zero line), but if the member is not disabled and the cms_risk_adj_payment_rate_b_amt is greater than 500 then the predictability increases.
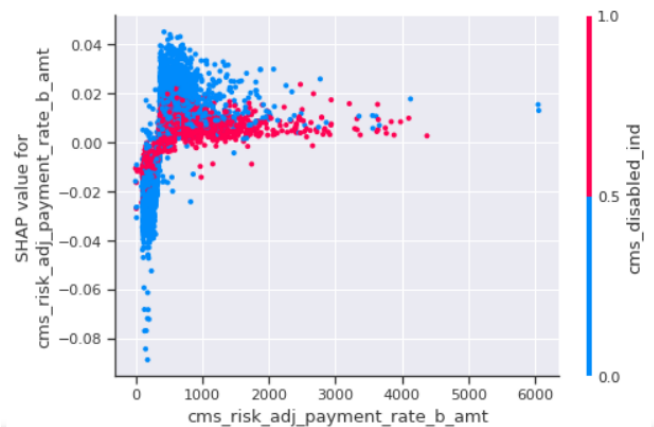
- Increase in member awareness about the effects of various problems like the superficial and contusion injury (ccsp_239) which was one of the reasons for transportation issue. By studying more about this issue, we can educate members through pop-up notification or share it on the main website.
- The need for virtual connection is more needed than ever due to Covid. Creating a platform for members to meet the Primary care Physician (PCP) virtually might help reduce the emergency ambulance rides.
- Calculate in-house treatments for disabled members, to find optimal way of services to members.

7. Scaling and Deployment

- We used python to process data and develop ML models. We have written reusable code functions and pipelines which can easily adapt to new data.
- The code is highly scalable and can be support ML frameworks with few tweaks.
- We used Tableau to create data visualizations and dashboard which can be liked to live database for real time insights.

- Since all the features in the model change their value over time, it is important to retrain the model with updated features and use the updated model for better realistic predictions.

8. Future Improvements

- The model is built on LightGBM (an ensemble method), and the predictability can be increased further by employing complex models and creating new features through interactions and auto-encoders.
- We had to discard most of the demographics features which are vital in model building as they can provide more information about the area where the member is living, hospital accessibility, the neighborhood quality, etc. This data can be clubbed with other online sources such as number of hospitals in a zipcode to come up with more personalized solutions.
- Along with AUC we can employ the precision-recall threshold trade-off after building a cost matrix on how much of misclassification in each category is acceptable.
- A similar approach can be applied to other insurance plans to gain insights on the SDOH factors.
- Scheduling meetings with relevant personals will help to learn more about the business problem, desired outcome, and relevant recommendations.

9. Conclusion

SDOH is now being incorporated in insurance and public health domain. With new rules from CMC, various programs can be initiated based on the insights generated from the above analysis. By mitigating the transportation issue faced by its members, Humana will be able to solve one of the Social Determinants of health which in turn would not only improve health status and affordability of its members, but also cut-costs, reduce spending and support communities. By Identifying a small number of the high-risk members responsible for majority of healthcare costs, reaching, and engaging with those members is critical for improved outcomes. Since, transportation is a broader/holistic

issue dependent on many economic factors. The increase in personal data combined with machine learning has potential to identify the vulnerable members.

10. Appendix

10.1 Statistical Model Output

```
Call:
glm(formula = transportation_issues ~ ., family = binomial(link = "logit"),
    data = df_train)

Deviance Residuals:
    Min       1Q    Median       3Q       Max
-3.8872   -0.9687   -0.2381    1.0401    1.9890

Coefficients:
                                     Estimate Std. Error z value Pr(>|z|)
(Intercept)                         1.506e+00  6.379e-02   23.605  < 2e-16 ***
cms_tot_partd_payment_amt           2.546e-03  1.258e-04   20.240  < 2e-16 ***
est_age                            -3.188e-02  8.322e-04  -38.303  < 2e-16 ***
cms_disabled_ind1                   3.832e-02  1.792e-02    2.138  0.03253 *
ccsp_239_ind1                       6.879e-01  2.320e-02   29.653  < 2e-16 ***
cms_rx_risk_score_nbr               5.652e-02  1.206e-02    4.688 2.76e-06 ***
betos_o1a_pmpm_ct_betos             1.895e-01  3.447e-02    5.497 3.86e-08 ***
cms_tot_ma_payment_amt              8.535e-05  1.843e-05    4.631 3.64e-06 ***
cms_low_income_ind1                 3.786e-02  2.733e-02    1.385  0.16598
bh_sum_ind                          1.516e-01  9.427e-03   16.078  < 2e-16 ***
rx_mail_pmpm_ct_RX                 -4.098e-02  3.988e-03  -10.276  < 2e-16 ***
MEN_pmpm_sum_MCC                    9.120e-02  1.811e-02    5.035 4.77e-07 ***
cms_partd_ra_factor_amt             8.895e-02  1.099e-02    8.092 5.85e-16 ***
total_ambulance_visit_ct_pmpm       6.042e-01  6.011e-02   10.051  < 2e-16 ***
total_physician_office_visit_ct_pmpm -2.018e-01  1.655e-02  -12.193  < 2e-16 ***
ccsp_220_ind1                      -1.058e-01  1.792e-02   -5.905 3.52e-09 ***
fci_score_MCC                       1.316e-02  4.367e-03    3.012  0.00259 **
EVAL_ind_sum_betos                 -1.469e-01  1.046e-02  -14.047  < 2e-16 ***
cms_risk_adj_payment_rate_b_amt     1.377e-05  3.175e-05    0.434  0.66444
pmpm_sum_neuro_2_RX                 2.109e-01  1.904e-02   11.077  < 2e-16 ***
cms_ma_risk_score_nbr               9.878e-02  1.049e-02    9.415  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 133733  on 96467  degrees of freedom
Residual deviance: 116360  on 96447  degrees of freedom
AIC: 116402

Number of Fisher Scoring iterations: 5
```

10.2 Model Features

10.2.1 Medical Claims Category

New Features	
DUR_ind_sum_betos	Sum of all the indicators present under durable medical equipment category
EVAL_ind_sum_betos	Sum of all the indicators present under evaluation and management category
EXC_ind_sum_betos	Sum of all the indicators present under exception/unclassified category
yes_no_sum_betos	Total categories from which a service was used by the member
CCSP_sum	Total number of ccsp services availed by the patient
TOTAL_ind_sum	Number of services used in combined med & bh category
TOTAL_pmpm_sum	Sum of number of services used
Default features	
betos_d1d_ind_betos	Binary indicator for durable medical equipment - wheelchairs
betos_m1b_ind_betos	Binary indicator for evaluation and management- office visits - established
betos_m5b_ind_betos	Binary indicator for evaluation and management - specialist - psychiatry
betos_m5b_pmpm_ct_betos	Per member per month count of logical claims for specialist - psychiatry
betos_m5d_ind_betos	Binary indicator for evaluation and management - specialist - other
betos_o1a_ind_betos	Binary indicator for ambulance
betos_o1a_pmpm_ct_betos	Per member per month count of logical claims for ambulance
betos_o1b_ind_betos	Binary indicator for chiropractic

betos_t1a_ind_betos	Binary indicator for lab tests - routine venipuncture (non-Medicare fee schedule)
ccsp_107_ind	Binary indicator for ccs code - cardiac arrest and ventricular fibrillation
ccsp_163_ind	Binary indicator for ccs code - genitourinary symptoms and ill-defined conditions
ccsp_205_ind	Binary indicator for ccs code - spondylosis, intervertebral disc disorders, back problems
ccsp_220_ind	Binary indicator for ccs code - intrauterine hypoxia and birth asphyxia
ccsp_228_ind	Binary indicator for ccs code - skull and face fractures
ccsp_236_ind	Binary indicator for ccs code - skull and face fractures
ccsp_239_ind	Binary indicator for ccs code - superficial injury, contusion
total_ambulance_visit_ct_ind	Combined med & bh values - ambulance visits
total_ambulance_visit_ct_pmpm	Combined med & bh values - ambulance visits
total_er_visit_ct_ind	Combined med & bh values - emergency room visits
total_er_visit_ct_pmpm	Combined med & bh values - emergency room visits
total_ip_acute_admit_days_ind	Combined med & bh values - inpatient acute
total_ip_mhsa_admit_ct_pmpm	Combined med & bh values - inpatient mhsa
total_outpatient_visit_ct_pmpm	Combined med & bh values - outpatient
total_physician_office_visit_ct_ind	Combined med & bh values - physician office
total_physician_office_visit_ct_pmpm	Combined med & bh values - physician office
total_urgent_care_visit_ct_ind	Combined med & bh values - urgent care

10.2.2 Pharmacy Claims Category

New features	
immuno_yes_no_RX	Binary indicator - whether the member used drugs in immuno category

pmpm_sum_immuno_RX	Sum of pmpm used by member in immuno drug category
pmpm_sum_neuro_2_RX	Sum of pmpm used by member in neuro 2 drug category
pmpm_sum_neuro_RX	Sum of pmpm used by member in neuro 1 drug category
Default features	
rx_bh_ind_RX	Binary indicator for prescriptions broken out by category
rx_generic_pmpm_ct_RX	Per member per month count for prescriptions broken out by category
rx_gpi2_43_pmpm_ct_RX	Per member per month count for prescriptions broken out by *cough/cold/allergy* category
rx_gpi2_44_ind_RX	Binary indicator for prescriptions broken out by *antiasthmatic and bronchodilator agents* category
rx_gpi2_50_ind_RX	Binary indicator for prescriptions broken out by *antiemetics* category
rx_gpi2_58_ind_RX	Binary indicator for prescriptions broken out by *antidepressants* category
rx_gpi2_59_ind_RX	Binary indicator for prescriptions broken out by *antipsychotics/antimanic agents* category
rx_mail_ind_RX	Binary indicator for prescriptions broken out by category
rx_mail_pmpm_ct_RX	Per member per month count for prescriptions broken out by category
rx_overall_pmpm_ct_RX	Per member per month count for prescriptions

10.2.3 Demographics Category

Default features	
est_age	Member age calculated using est_bday, relative to score/index date

rucc_category	Member geographic information - rural urban continuum code
sex_cd	Member gender
smoker_current_ind	Current smoker based on the presence of smoking indication from membership or medical claims data
smoker_former_ind	Former smoker based on the presence of smoking indication from membership or medical claims data

10.2.4 Condition Related Category

New features	
CAN_ind_sum_MCC	Number of services used by member in cancer category
CAN_pmpm_sum_MCC	Sum of pmpm used by member in cancer category
CER_pmpm_sum_MCC	Sum of pmpm used by member in cerebrovascular category
DIA_ind_sum_MCC	Number of services used by member in diabetes category
GUS_pmpm_sum_MCC	Sum of pmpm used by member in genito urinary category
INF_ind_sum_MCC	Number of services used by member in infection category
INJ_pmpm_sum_MCC	Sum of pmpm used by member in injury and poisoning category
MEN_ind_sum_MCC	Number of services used by member in mental health condition category
MEN_pmpm_sum_MCC	Sum of pmpm used by member in mental health conditions category
NER_ind_sum_MCC	Number of services used by member in nervous system category
PRE_ind_sum_MCC	Number of services used by member in pregnancy and complications category
SKN_pmpm_sum_MCC	Sum of pmpm used by member in skin category
SNS_pmpm_sum_MCC	Sum of pmpm used by member in signs and symptoms category

SOR_ind_sum_MCC	Number of services used by member in sense organs category
VCO_ind_sum_MCC	Number of services used by member in vcodes category
Default features	
cci_score_MCC	Charlson comorbidity index score
fci_score_MCC	Functional comorbidity index
submcc_ben_othr_ind_MCC	Binary indicator - neoplasms (benign) - other
submcc_ben_othr_pmpm_ct_MCC	Per member per month count of logical claims for neoplasms (benign) - other
submcc_bld_othr_pmpm_ct_MCC	Per member per month count of logical claims for diseases of blood and blood-forming organs - other
submcc_can_h/o_ind_MCC	Binary indicator -malignant neoplasms - v codes
submcc_can_h/o_pmpm_ct_MCC	Per member per month count of logical claims for malignant neoplasms - v codes
submcc_can_othr_ind_MCC	Binary indicator -malignant neoplasms - other
submcc_can_othr_pmpm_ct_MCC	Per member per month count of logical claims for malignant neoplasms - other
submcc_cir_art_ind_MCC	Per member per month count of logical claims for binary indicator -other circulatory - arteriosclerosis
submcc_dia_neu_pmpm_ct_MCC	Per member per month count of logical claims for diabetes - neuropathy
submcc_gus_kub_ind_MCC	Binary indicator – genito - urinary system - kidney/ureter/bladder
submcc_inj_drug_pmpm_ct_MCC	Per member per month count of logical claims for injury and poisoning - drug complications
submcc_inj_othr_pmpm_ct_MCC	Per member per month count of logical claims for infections - other
submcc_men_abus_ind_MCC	Binary indicator -mental health conditions - substance abuse
submcc_men_alco_pmpm_ct_MCC	Per member per month count of logical claims for mental health conditions - alcoholism

submcc_men_depr_ind_MCC	Binary indicator -mental health conditions -depression
submcc_men_depr_pmpm_ct_MCC	Per member per month count of logical claims for mental health conditions -depression
submcc_mus_arth_ind_MCC	Binary indicator -musculoskeletal and connective tissue - arthritis (other than ra)
submcc_mus_form_ind_MCC	Binary indicator -musculoskeletal and connective tissue - malformations
submcc_mus_oste_ind_MCC	Binary indicator -musculoskeletal and connective tissue - osteoporosis
submcc_mus_othr_pmpm_ct_MCC	Per member per month count of logical claims for musculoskeletal and connective tissue - other
submcc_ner_epil_pmpm_ct_MCC	Per member per month count of logical claims for nervous system - epilepsy
submcc_ner_othr_pmpm_ct_MCC	Per member per month count of logical claims for nervous system - other
submcc_res_asth_pmpm_ct_MCC	Per member per month count of logical claims for respiratory - asthma
submcc_res_copd_ind_MCC	Binary indicator -respiratory - chronic obstructive pulmonary disease
submcc_res_copd_pmpm_ct_MCC	Per member per month count of logical claims for respiratory - chronic obstructive pulmonary disease
submcc_res_fail_pmpm_ct_MCC	Per member per month count of logical claims for respiratory - respiratory failure
submcc_res_inf_ind_MCC	Binary indicator -respiratory - infections
submcc_rsk_chol_ind_MCC	Binary indicator -risk behaviors - hyperlipidemia
submcc_rsk_chol_pmpm_ct_MCC	Per member per month count of logical claims for risk behaviors - hyperlipidemia
submcc_rsk_obe_pmpm_ct_MCC	Per member per month count of logical claims for risk behaviors - obesity
submcc_rsk_smok_pmpm_ct_MCC	Per member per month count of logical claims for risk behaviors - smoking
submcc_sns_othr_ind_MCC	Binary indicator -signs and symptoms - other

submcc_sns_othr_pmpm_ct_MCC	Per member per month count of logical claims for signs and symptoms - other
submcc_sor_ear_pmpm_ct_MCC	Per member per month count of logical claims for sense organs - ear
submcc_sor_eye_ind_MCC	Binary indicator - sense organs - eye
submcc_vco_end_ind_MCC	Binary indicator -v-codes - endocrine
submcc_vco_exam_ind_MCC	Binary indicator -v-codes - exam and screening
submcc_vco_vac_ind_MCC	Binary indicator - v-codes - vaccination

10.2.5 CMS Category

Default features	
cms_disabled_ind	Disability indicator
cms_dual_eligible_ind	Dual eligibility indicator - eligible for both Medicare and Medicaid
cms_low_income_ind	Low income subsidy indicator from CMS
cms_ma_risk_score_nbr	CMS Medicare advantage risk score
cms_partd_ra_factor_amt	CMS Medicare part d risk adjustment factor amount
cms_risk_adj_payment_rate_b_amt	CMS risk adjustment payment rate b
cms_rx_risk_score_nbr	CMS Medicare rx risk score
cms_tot_ma_payment_amt	Total Medicare advantage payment amount
cms_tot_partd_payment_amt	Total partd payment amount
cmsd2_can_unc_neo	Binary indicator for CMS level 2 diagnosis category - neo
cmsd2_men_mad_ind	Binary indicator for CMS level 2 diagnosis category - mental health
cmsd2_men_men_substance_ind	Binary indicator for CMS level 2 diagnosis categories present in the reference table
cmsd2_skn_radiation_ind	Binary indicator for CMS level 2 diagnosis category - skin radiation

cmsd2_sns_general_ind	Binary indicator for CMS level 2 diagnosis category - sign and symptoms category

10.2.6 Other Category

New features	
bh_sum_ind	Number of symptoms showed by member in behavioral category
bh_yes_no	Binary indicator - member showed any behavioral symptom or not
phy_sum_ind	Total number of physician services availed by the patient
prov_yes_no	Binary indicator - member used any of the prov category service or not
rev_sum_ind	Sum of number of revenue services used
Default features	
bh_bipr_ind	Binary indicator for bipolar disorder
bh_cdto_ind	Binary indicator for tobacco use disorder
bh_dema_ind	Binary indicator for major depressive disorder
mabh_seg_h_c	Multi-level indicator - medical segmentation
phy_em_pe_ind	Binary indicator for each physican e & m category
prov_spec_ambulance_ind	Binary indicator for a select group of categories using std_hipaa_prov_spec_cd
prov_spec_bh_psychiatric_ind	Binary indicator for a select group of categories using std_hipaa_prov_spec_cd
prov_spec_pain_mgmt_ind	Binary indicator for a select group of categories using std_hipaa_prov_spec_cd
rev_cms_clinic_ind	Binary indicator for revenue code clinal cms category
rev_cms_er_ind	Binary indicator for revenue code emergency room category
rev_cms_lab_ind	Binary indicator for revenue code lab category
rev_cms_phar_ind	Binary indicator for revenue code pharmacy category

11. References

- https://www.aha.org/ahahret-guides/2017-11-15-social-determinants-health-series-transportation-and-role-hospitals#:~:text=Transportation%20issues%20can%20affect%20a,be%20a%20vehicle%20for%20wellness.

- https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MedicareFeeforSvcPartsAB/Downloads/BETOSDescCodes.pdf

- http://health.utah.gov/opha/IBIShelp/codes/CCS.htm

- https://www.cms.gov/newsroom/fact-sheets/medicare-part-d-direct-and-indirect-remuneration-dir

- http://www.hpoe.org/resources/ahahret-guides/3078

- https://www.ensocare.com/knowledge-center/how-demographics-impact-health-care-delivery

- https://www.healthedge.com/blog-social-determinants-health-what-are-payers-doing

- http://chirblog.org/insurance-company-investments-social-determinants-of-health/

- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4265215/https://christophm.github.io/interpretable-ml-book/shap.html