ENHANCING PREVENTIVE CARE IN HEALTHCARE SYSTEMS USING DATA-DRIVEN APPROACHES

Humana – Mays Healthcare Analytics Competition 2024

Humana

TEXAS A&M UNIVERSITY
Mays Business School

Table Of Contents

Table of Figures

1. Executive Summary

The *Humana-Mays Healthcare Analytics Case Competition* aims to address the challenge of low engagement in preventive care among Humana's LPPO plan members. LPPO plans, while offering members flexibility in choosing healthcare providers, face higher rates of disengagement compared to other plans like HMOs. This disengagement poses significant risks to both member health outcomes and Humana's performance in the CMS Stars and Risk Adjustment programs, which are crucial for delivering enhanced member benefits. Health insurance companies operate by pooling risks among their members, collecting premiums to fund healthcare services, and ensuring members have access to necessary medical care. They play a critical role in managing costs and risks associated with health services, while also improving the quality of care through initiatives such as preventive care programs.

Preventive care is a key focus, as it helps reduce long-term healthcare costs by identifying and managing health issues early. Insurers, like Humana, benefit from strong member engagement in preventive care, as it improves overall health outcomes and reduces the cost burden of untreated or late-stage conditions. This competition seeks innovative solutions to identify members at risk of disengagement from preventive care and to recommend strategies for increasing engagement. The focus is on leveraging data to proactively pinpoint unengaged members, understanding key behavioral factors that contribute to disengagement, and proposing actionable interventions that drive higher engagement in preventive primary care visits. The data provided includes over 1.5 million LPPO members with approximately 300 features. These features capture a comprehensive view of member demographics, claims, plan details, tenure, medical/pharmacy claims, and engagement metrics, such as call center interactions and web activity. The data-driven recommendations provided in this report focus on feasible and proven strategies to increase engagement in preventive care visits. By implementing targeted outreach, leveraging pharmacy partnerships, offering financial incentives, and using digital tools, healthcare organizations can improve preventive care engagement, reduce long-term healthcare costs, and ensure better patient outcomes.

Action Plan and Recommendations:

1. **Automated Outreach Campaigns**: Initiate campaigns targeting members with lower healthcare expenditure with SMS and email reminders, along with incentives.
2. **Pharmacy Partnerships**: Collaborate with pharmacies to offer preventive services during prescription pickups.
3. **Veteran-Specific Programs**: Launch veteran-specific outreach through partnerships with veteran organizations.
4. **Digital Engagement Tools**: Implement SMS/email reminders and promote online scheduling to increase participation.
5. **Incentives for High-Risk Members**: Offer waived copays and other financial incentives to reduce barriers for high-risk individuals.

2. Case Context: Understanding Preventive Care Engagement in Humana's LPPO Plans

2.1 Case Objective

The primary objective of this case is to identify and address the key factors that lead to low engagement in preventive care among Humana's Local Preferred Provider Organization (LPPO) members. Specifically, the goal is to:

1. Proactively identifying members most likely to be disengaged and not completely preventive primary care visits.
2. Analyze key characteristics, behaviors, and external factors that influence member engagement.
3. Develop strategies to increase engagement in preventive services, improve health outcomes, and positively impact Humana's performance in key metrics such as CMS Stars ratings and Medicare Risk Adjustment (MRA).

Preventive care plays a crucial role in maintaining overall health, managing chronic conditions, and preventing more serious health issues. By ensuring members engage in preventive care, Humana can help reduce long-term healthcare costs, improve member satisfaction, and achieve better performance in government rating programs, ultimately delivering greater value to its members and stakeholders.

2.2 Problem Statement

The problem facing Humana's LPPO plans is the significantly higher rate of unengaged members in preventive care compared to other health plans, such as Health Maintenance Organizations (HMOs). This disengagement is concerning because:

- **Preventive care**—such as annual wellness visits, vaccinations, and screenings—helps detect health issues early, reducing the need for more expensive interventions later.
- LPPO members have the flexibility to seek care outside a defined network, which may lead to inconsistent touchpoints with healthcare providers, particularly primary care physicians (PCPs). Without regular interactions with a PCP, members may miss out on important health screenings and preventive services.

Key problem areas include:

- **Lower PCP Engagement**: Members may not establish a relationship with a PCP, which is crucial for ongoing preventive care and management of health risks.
- **Higher Reliance on Specialists**: Members who visit specialists may not receive comprehensive preventive care, as specialists typically focus on specific health issues rather than overall wellness.
- **Lack of Awareness**: Many members are unaware of the importance and availability of preventive care services, leading to disengagement.
- **Behavioral Gaps**: Members with lower health literacy, younger members, and those with socio-economic challenges tend to be less engaged in preventive care.

2.3 Risks of Disengagement for Humana

If Humana's LPPO members continue to disengage from preventive care, several key challenges and risks could emerge, both in terms of member health outcomes and business performance:

2.3.1 Impact on Member Health Outcomes

a) *Delayed Detection of Health Issues*: Without preventive care, conditions like diabetes, hypertension, and cancers may go undetected or untreated until they progress to more serious, costly stages. This puts members at greater risk for hospitalizations, emergency room visits, and long-term complications. Early detection through preventive care can significantly reduce these risks, but disengaged members miss these opportunities for timely intervention.

b) Increased Burden of Chronic Conditions: Preventive care helps in the ongoing management of chronic conditions, ensuring that treatments are effective, and members stay on track with medications and lifestyle changes. Without this support, chronic conditions can worsen, leading to higher healthcare utilization and poorer quality of life for members.

c) Lower Member Satisfaction: Disengaged members may feel disconnected from their healthcare providers, leading to frustration and dissatisfaction with their health plan. This could result in higher rates of member turnover, which can negatively affect Humana's customer retention rates.

2.3.2 Business and Financial Impact on Humana

a) Decreased CMS Stars Ratings: The Centers for Medicare and Medicaid Services (CMS) Stars rating system rewards health plans that achieve high-quality care metrics, including preventive care. A decline in preventive care can lead to lower Star ratings, resulting in reduced bonuses and government funding. Lower Stars ratings also make it harder for Humana to market its plans as high-quality options, potentially impacting member acquisition and retention.

b) Incomplete Risk Adjustment Documentation: Medicare Risk Adjustment (MRA) requires accurate documentation of member health risks. If members are not visiting their PCPs for preventive services, important health conditions may go undiagnosed and undocumented, leading to lower risk scores and reduced reimbursement from CMS. This could result in financial losses for Humana, as the plan will receive less funding to cover the true health risks of its members.

c) Higher Healthcare Costs: Without preventive care, members are more likely to require acute interventions, such as hospitalizations or emergency room visits, which are more costly than routine preventive visits. This could increase Humana's overall healthcare expenditures, eroding profitability and the ability to offer enhanced plan benefits

d) Member Retention and Plan Competitiveness: Engaged members who experience positive health outcomes and strong provider relationships are more likely to remain loyal to their health plan. Conversely, disengaged members may seek out other options, leading to higher member turnover, reduced market share, and lower plan competitiveness.

2.4 Strategic Importance of Solving the Problem

Solving the issue of low preventive care engagement is crucial for Humana, as it aligns with both its mission to improve member health outcomes and its financial performance objectives. Increasing engagement will:

- Improve member health through early detection and better management of conditions.
- Enhance Humana's CMS Stars ratings and Medicare Risk Adjustment scores, leading to higher reimbursement and better market positioning.
- Allow Humana to reinvest in additional member benefits, creating a positive feedback loop that attracts and retains more members.

By addressing these challenges, Humana can lead the way in preventive care engagement and solidify its position as a top healthcare provider in the LPPO market.

3. Data Understanding and Data Overview

3.1 Introduction to the Dataset

The dataset provided for this competition consists of multiple tables containing various features related to member demographics, healthcare utilization, social determinants of health, pharmacy use, and more. These tables collectively offer a comprehensive view of over 1.5 million unique members, allowing for detailed analysis of their engagement in preventive care.

Based on the data provided by Humana, two primary datasets have been included: the training dataset covering the people registered in Humana LPPO plan in 2023 and their past two-year engagement and patterns.

- *Training Dataset:* This dataset consists of 1,527,904 unique Humana members.
- *Holdout Dataset:* This dataset includes 381,976 unique members and is intended for making predictions based on historical patterns observed in the training data

Target Variable:

The target variable for this analysis is **'***preventive_visit_gap_ind***'**, which indicates whether a member has a gap in preventive care. This binary variable is central to the model's goal of predicting which members are less likely to complete a preventive care visit. The objective is to identify members with this gap and intervene proactively to improve their engagement with healthcare services.

3.2 Data Aggregation and Structure

The Quality dataset, member visit claims, and member condition datasets contained a higher number of records due to multiple entries per member:

1. *Multiple claims:* Members may have more than one claim throughout the year, necessitating aggregation of cost and utilization data.
2. *Repeated health conditions*: A member might have several diagnoses or interactions with different healthcare providers, requiring the consolidation of condition-related information.

To perform analysis at the member level, the data needs to be aggregated so that each row represents a unique member. Key considerations in this aggregation process include:

- Summing or counting numeric features like member condition, number of visits, or prescription fills.
- Flagging binary variables such as whether a member has certain health conditions or has used services.
- Categorical data such as member demographics (e.g., age group, gender) can remain unaggregated but serve as key variables for segmentation.

3.3 Data Exploration

In the data exploration phase, member demographics, geographic location, healthcare access, and plan tenure are examined to uncover patterns and factors related to missed preventive visits. This analysis provides an understanding of the key trends and relationships within the data, helping to highlight potential

barriers and contributing factors to missed preventive care. Through this exploration, we gain a clearer picture of how different variables may influence adherence to preventive visit schedules.

Target Variable Distribution: Nearly 45% of members miss preventive visits (indicated as 1), indicating significant engagement challenges across the entire member population. Behavioral patterns such as low adherence to scheduled visits or lack of knowledge about the importance of preventive care may contribute to this gap.

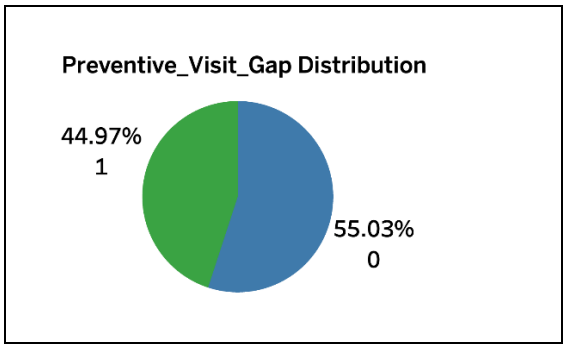


Fig. 3.1 *Target Variable Distribution*

1. **Demographic and Geographic Factors**

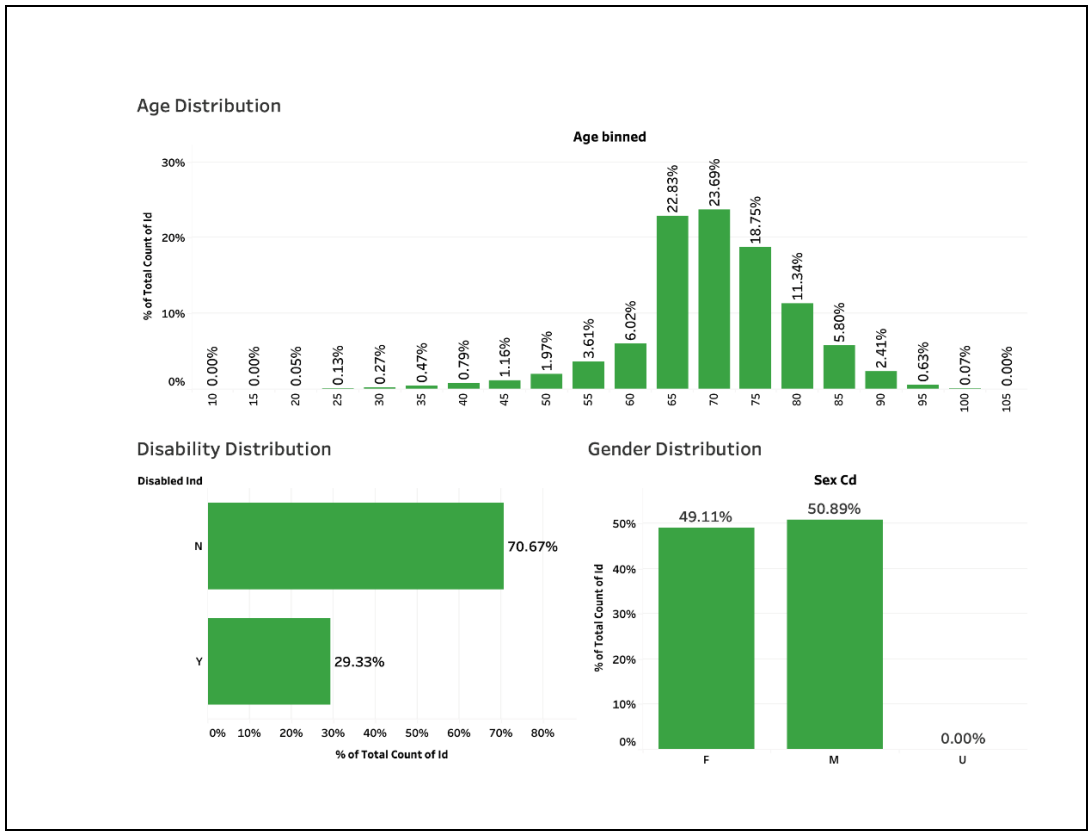


Fig. 3.2 *Distribution of 'preventive visit gap ind = 1' in age, gender and disability status*

- *Age and Gender Distribution*: Close to 83% of the members missing preventive care visits are aged between 65-80. Younger and Middle-aged members contribute to lower preventive visit gaps. There is no significant gender disparity, as preventive visit gaps are nearly evenly split between males (50.89%) and females (49.11%).
- *Disability Status*: While 70.67% of members with preventive visit gaps are non-disabled, a significant 29.33% are disabled.

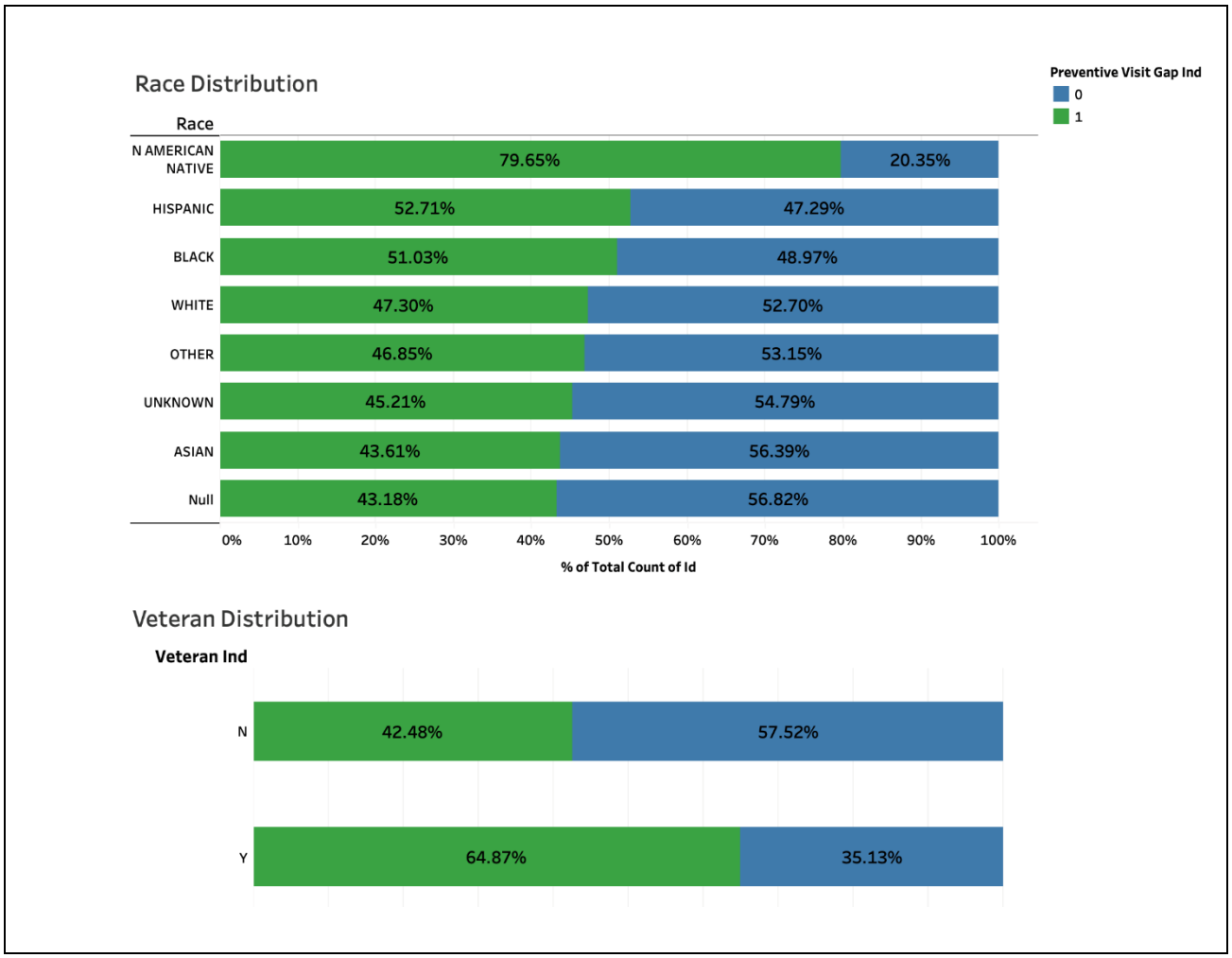


Fig. 3.3 *Distribution of 'preventive visit gap ind = 1' in Race and Veteran Ind*

- *Race Distribution*: Native American members have the largest gap in preventive care, with 79.65% missing their preventive visits, making them the least engaged group. Other racial groups show higher engagement, with 52.70% of White members, 48.97% of Black members, and 47.29% of Hispanic members missing preventive care visits. Asian members and those with unknown race data have slightly lower engagement, with 56.39% and 54.79% respectively missing preventive visits. This emphasizes the need for targeted interventions, especially among Native American members, to close the preventive care gap.

- *Veteran Status*: Most members with gaps are veterans (64.87%), although non- veterans (42.48%) also experience notable engagement issues.

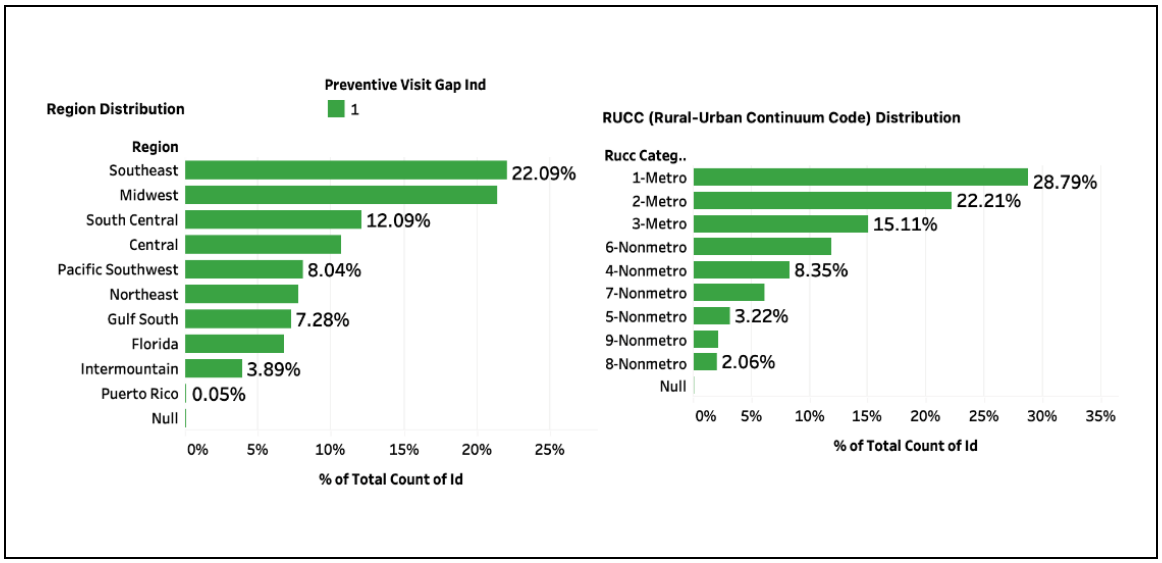


Fig. 3.4 *Distribution of 'preventive visit gap ind = 1' in Region and RUCC*

- *Region Distribution*: Members in the Southeast (22.09%) and Midwest (21.37%) contribute to a higher percentage of missing preventive visits, with the South Central and Central regions also showing significant representation. Florida, Intermountain, and Puerto Rico show fewer issues.
- *RUCC (Rural-Urban Continuum Code) Distribution*: Preventive visit gaps are more prevalent in metro areas (e.g., 1-Metro at 28.79%, 2-Metro at 22.21%), though non-metro areas still face notable gaps (e.g., 4-Nonmetro at 11.89%).

2. **Healthcare Access and Plan Factors**

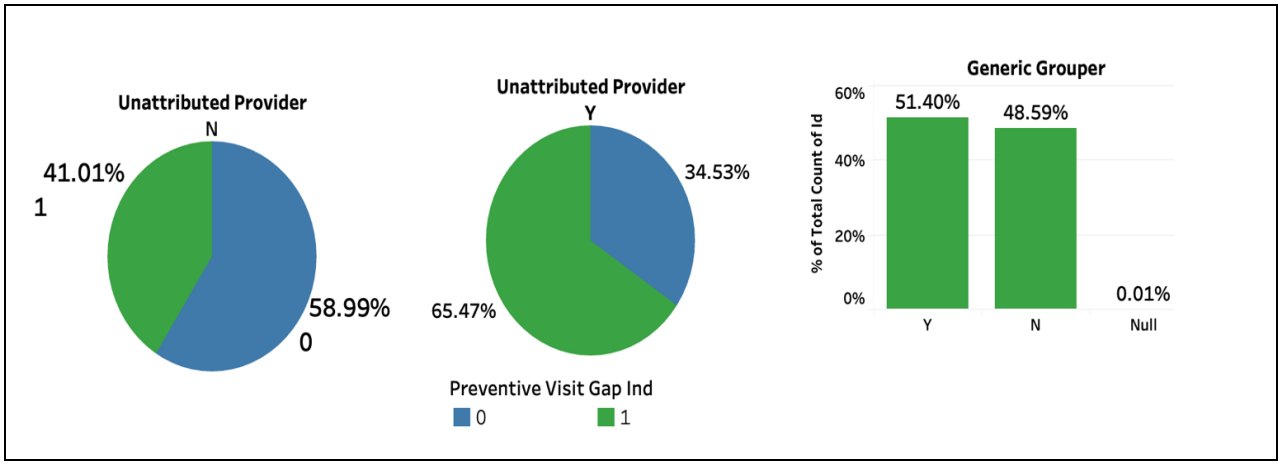


Fig. 3.5 *Distribution of 'preventive visit gap ind' = 1' in Unattributed provider and generic grouper*

- *Unattributed Providers*: Among members who are not visiting preventive care, Members without an attributed provider (Y) have a preventive visit gap of 65.47%, though 40.01% of attributed members (N) also miss preventive visits.
- *Generic Grouper Classification*: This chart shows that members classified as Y in the Generic Grouper have the highest percentage of preventive visit gaps (56.32%), while those classified as N are less likely to have gaps (37.06%).

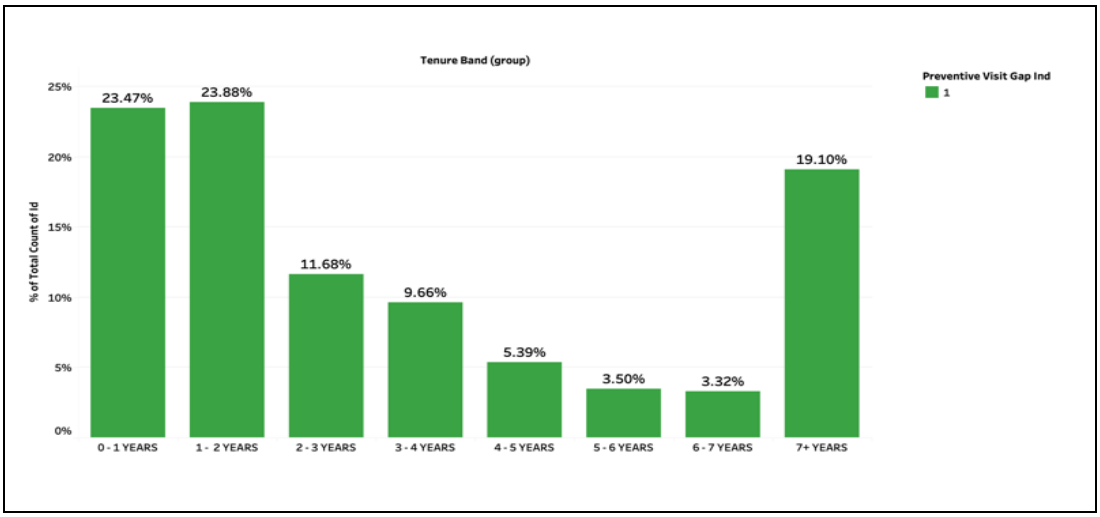


Fig. 3.6 *Tenure Band*

- *Tenure Distribution*: Among members who are not visiting preventive care, newer members (with 0-1 years of tenure at 23.47% and 1-2 years at 23.88%) have more preventive visits gap. Members with 7+ years also have a notable gap rate (19.10%), while members with 5-6 years of tenure show the lowest gap rates (3.50%).

4. Data Preparation and Feature Engineering

4.1 Data Aggregation

The first step in preparing the dataset was to aggregate the data across the various tables, particularly those that contained multiple entries per member. There were three datasets that required specific aggregation techniques to ensure that each unique member was represented by a single row, combining information across multiple datasets, interactions, claims, and records.

4.1.1 Member Conditions Dataset

The Member Conditions dataset provides detailed information on chronic health conditions affecting members. Given that members may have multiple conditions recorded over the observation period, the dataset needed to be aggregated to create a concise member-level view that summarizes the health risks each member faces.

Approach to Aggregation

To prepare the Member Conditions data for analysis, we applied the following aggregation steps:

1. *Categorizing Health Conditions*: The first step was to categorize the individual conditions into broader health categories based on common medical themes. Instead of analyzing each individual condition in isolation, similar conditions were grouped into high-level categories such as:
    - **Cardiovascular Diseases**: This category includes conditions like heart failure, cardiomyopathy, and arrhythmias.
    - **Diabetes-Related**: Includes both chronic complications of diabetes and acute conditions such as severe diabetic eye disease.
    - **Kidney and Liver Diseases**: Captures conditions like chronic kidney disease, liver failure, and hepatitis.
    - **Respiratory Diseases**: Includes chronic obstructive pulmonary disease (COPD), asthma, and lung fibrosis.
    - **Cancers**: Covers various cancer types, including metastatic cancers and acute leukemia.
    - **Mental Health Disorders**: Includes conditions such as major depressive disorder, schizophrenia, and substance use disorders.
    - **Neurological Disorders**: Includes Parkinson's disease, cerebral palsy, and seizure disorders.
    - **Immunological Disorders**: Consists of immune system disorders such as lupus and rheumatoid arthritis.
    - **Miscellaneous**: Includes other significant conditions like morbid obesity and chronic ulcers.

This categorization simplified the dataset, reducing hundreds of potential conditions into a manageable set of categories. By focusing on these broader categories, we were able to highlight the major chronic disease burdens faced by each member.

2. *Summarizing Conditions at the Member Level:* Once the conditions were grouped into these high-level categories, the next step was to create a summary of each member's health conditions. For each member, we counted how many conditions they had in each category.
3. *Handling Multiple Records per Member*: Since members could have multiple conditions or multiple records for the same condition, it was important to ensure that the aggregation captured the number of unique conditions rather than just the number of records. This involved summarizing each member's conditions by counting the occurrences of each condition category across the observation period.

4.1.2 Member Claims Visit Dataset

The Member Visit Claims dataset contains detailed records of healthcare services accessed by members, including preventive services, specialist visits, emergency room visits, and telehealth consultations. Given that members may have multiple claims throughout the year, it was essential to aggregate these data to create a comprehensive, member-level view of healthcare utilization.

Approach to Aggregation

To aggregate this data at the member level, the following techniques were applied:

- For fields such as number of visits, ER visits, and preventive services received, numeric values were summed per member. This helped quantify the total healthcare utilization for each member over the observed period.

To gain deeper insights into utilization trends, the data was aggregated separately for the years 2021 and 2022 also. This allows for a better understanding of how healthcare engagement, particularly preventive care, changed over time.

4.1.3 Quality Dataset

The Quality Dataset captures data for multiple quality measures related to HEDIS, Patient Safety, and Patient Experience from 2020 to 2022. Each measure evaluates different aspects of member healthcare, such as preventive screenings, chronic condition management, and patient satisfaction. Since the dataset contains detailed information for each member across multiple quality measures and years, we aggregated it at the member level to summarize their overall compliance with key quality measures.

Approach to Aggregation

1. **Aggregating by Measure Type (HEDIS, Patient Safety, and Patient Experience)**:
   - We aggregated the data by measure type to summarize member compliance with each type of quality measure. This provided an overall view of how members performed across the three main quality domains:

- **HEDIS**: Health Effectiveness Data and Information Set, measuring preventive care and chronic disease management.
- **Patient Safety**: Measures related to safe healthcare delivery, such as avoiding adverse drug reactions or readmissions.
- **Patient Experience**: Measures of patient satisfaction and engagement with their healthcare services.

The aggregation summed up the compliant count for each member across all measures in each type.

2. **Aggregating by Measure Name**:
   - We also aggregated the data by the specific measure name to capture compliance with individual measures. For each member, we summed the eligible count and compliant count for each measure.
   - This aggregation resulted in fields that represent a member's compliance with each specific measure, such as the number of times a member was eligible for and compliant with diabetes screenings or breast cancer screenings.
3. **Creating Compliance and Eligibility Features**:
   - After aggregating the data, we created two sets of features for each measure:
     - **Eligible Count**: The total number of times the member was eligible for the specific measure.
     - **Compliant Count**: The total number of times the member was compliant with the measure.

4.2 Feature Engineering

Feature engineering is a crucial step in the data preparation process, where new features are derived from existing data to improve the predictive power of the model. In this project, we introduced several ratio-based features to enhance the model's ability to detect patterns in member behavior, particularly in the Quality Dataset and Member Visit Dataset. These new features provide insights into member compliance with healthcare measures and trends in healthcare utilization over time.

4.2.1 Compliance Ratios

In the Quality Dataset, we introduced several ratio-based features that measure the proportion of compliance for various quality measures. These ratios represent how well members adhered to key quality indicators, such as preventive care, chronic condition management, and patient safety. The compliance-to-eligibility ratios provide a normalized measure of compliance, making it easier to compare members with varying levels of eligibility for specific measures.

Key Features Created

The following are the key compliance to eligibility ratios introduced in the Quality Dataset, representing the ratio of compliant members to those eligible for each measure:

- *MRP_ratio*: Medication Reconciliation Post-Discharge compliance.
- *OMW_ratio*: Osteoporosis Management in Women compliance.

- *Patient_Safety_ratio*: Compliance with overall patient safety measures.
- *HEDIS_ratio*: Compliance with HEDIS preventive care measures.
- *BCS_ratio*: Breast Cancer Screening compliance.
- *ADH_(ACE)_ratio*: Medication Adherence for ACE Inhibitors compliance.
- *COL_ratio*: Colon Cancer Screening compliance.
- *SUPD_ratio*: Statin Use in Persons with Diabetes compliance.

These ratios are designed to capture member adherence to preventive measures, chronic disease management, and patient safety protocols, providing a clear measure of engagement and compliance with healthcare guidelines.

Importance of Compliance Ratios

These features offer valuable insights into:

- Preventive Care Engagement: Higher ratios indicate greater compliance with preventive care measures, which correlate with better health outcomes and fewer gaps in care.
- Chronic Disease Management: Ratios such as *ADH_(ACE)_ratio* and *SUPD_ratio* capture compliance with chronic condition management protocols, highlighting members who are successfully managing conditions like hypertension and diabetes.
- Overall Healthcare Adherence: Ratios like *Patient_Safety_ratio* and *HEDIS_ratio* summarize compliance with broader healthcare quality measures, helping to identify members who are consistently engaged in maintaining their health.

4.2.2 Member Visit Dataset: Year-over-Year Visit Ratios

In the Member Visit Dataset, we introduced several ratio-based features that compare a member's healthcare utilization between 2021 and 2022. These ratios help track changes in healthcare engagement, highlighting members who may be increasing or decreasing their reliance on healthcare services.

Key Features Created

The following are the visit ratios created, comparing healthcare usage between 2021 and 2022:

- *pcp_visit_ratio*: Ratio of primary care physician visits in 2022 to 2021.
- *preventative_visit_ratio*: Ratio of preventive visits in 2022 to 2021.
- *er_visit_ratio*: Ratio of emergency room visits in 2022 to 2021.
- *urgent_care_visit_ratio*: Ratio of urgent care visits in 2022 to 2021.
- *telehealth_ratio*: Ratio of telehealth visits in 2022 to 2021.
- *specialist_visit_ratios*: Ratios for various specialist visits, such as *endocrinologist_visit_ratio, cardiologist_visit_ratio, oncologist_visit_ratio*, and more.

These year-over-year ratios provide a dynamic view of how members' healthcare utilization patterns are changing. Members with a declining ratio may be disengaging from care, while those with increasing ratios may be more proactive about their health.

Importance of Year-over-Year Visit Ratios

- *Detecting Changes in Healthcare Engagement:* These ratios allow us to monitor changes in member behavior, identifying members who are either increasing or decreasing their healthcare engagement. For example, a low or declining *preventative_visit_ratio* may signal that a member is at risk of missing critical preventive care services.
- *Assessing Reliance on Acute Care:* Ratios like *er_visit_ratio* and *urgent_care_visit_ratio* help identify members who are increasingly relying on emergency or urgent care, indicating possible gaps in preventive care or regular physician visits.
- *Tracking Telehealth Utilization:* The telehealth_ratio helps capture the shift toward digital healthcare solutions. An increasing ratio may indicate a preference for virtual care, while a declining ratio may reflect members reverting to in-person visits.

4.3 Data Imputation

Imputation is a vital part of the data preparation process, especially when dealing with large datasets with missing or incomplete data. In this project, the imputation strategy was tailored to the specific characteristics of the different datasets and feature types. We aimed to ensure that missing values were handled appropriately without distorting the overall dataset, enabling accurate analysis and model development. The chosen imputation techniques ensured that missing values were handled in a way that preserved the overall structure of the data without introducing bias or skew.

4.3.1 Identifying Missing Data

Missing values were present across several datasets, particularly in fields related to healthcare claims, social determinants of health, and quality measures. Proper handling of these missing values was critical to maintain the integrity of the data, particularly for aggregated columns, demographic data, and ratio fields.

4.3.2 Imputation Strategy Overview

The imputation strategy was applied differently depending on the nature of the data:

1. *Imputation for Aggregated Columns*

For aggregated columns (such as total visits and condition counts), missing values were imputed with zero. This approach assumes that if a record is missing, it indicates no healthcare event occurred, or the member did not use the service.

- Example: If a member had missing data for preventive care visits, the value was set to zero, indicating that no preventive visit took place for that member during the observation period.

2. Mean Imputation for Social Demographic Columns

For columns related to social demographics, such as income, education, and other socioeconomic factors, imputing with zero did not make sense, as it would distort the meaning of these fields. Instead, mean imputation was used for these fields to ensure that missing values were replaced with representative values from the dataset.

- Example: If a member's income data was missing, it was imputed with the mean income for all members in the dataset. This maintains a reasonable estimation for the missing data without skewing the results.

3. Imputation for Ratio Columns

For ratio columns (where one value is divided by another, such as a 2022_visits/2021_visits ratio) missing values were imputed based on the behavior of the numerator and denominator:

- If the numerator was zero, the ratio was set to zero.
- If the denominator was zero, the ratio was set to -1 to represent a default reasonable value without overinflating the result.

This strategy was applied to ensure that missing values in ratio columns were handled logically and didn't create division errors or misrepresent the underlying data.

5. Statistical Analysis and Modeling

The statistical analysis and modeling process forms the core of our predictive approach. In this section, we detail the steps taken to build a robust predictive model for identifying members at risk of not completing preventive care visits. While various classification models were tested, the XGBoost classifier was selected as the final model based on its superior performance.

To evaluate the performance of the model and ensure it generalizes well to unseen data, we split the dataset into training and testing sets using a 70/30 split. This means that 70% of the data was used to train the model, while the remaining 30% was reserved for testing and validating the model's performance on new, unseen data.

The 70/30 train-test split ensures that the model has enough data to learn complex patterns during training, while the test set allows us to evaluate how well the model generalizes. The train-test split was used consistently throughout the feature selection, hyperparameter tuning, and model evaluation process to ensure unbiased performance assessment.

5.1 Overview of Modeling Process

The primary goal of this analysis was to build a classification model to predict whether a member would experience a preventive care gap (the target variable: *preventive_visit_gap_ind*). The dataset provided a rich set of features, including aggregated data on member visits, quality measures, and claims.

5.1.1 Initial Model Selection

We began by testing multiple classification algorithms to evaluate their performance on this binary classification task. These models included:

- Decision Tree Classifier: For capturing non-linear relationships between the features.
- Random Forest Classifier: An ensemble method that combines multiple decision trees.
- XGBoost Classifier: A gradient boosting algorithm that iteratively improves prediction accuracy by combining weak learners.

Each model was evaluated using standard performance metrics such as accuracy, precision, recall, and the AUC-ROC curve. After careful evaluation, XGBoost was chosen as the final model due to its superior performance, flexibility, and ability to handle complex interactions between features.

5.2 Understanding XGBoost

XGBoost (Extreme Gradient Boosting) is a powerful machine learning algorithm that falls under the category of ensemble learning methods, specifically using gradient boosting techniques. The key idea behind XGBoost is to build an ensemble of weak learners (typically decision trees) and combine their

predictions to create a strong learner. The algorithm improves iteratively, correcting the errors of previous trees in each subsequent step.

5.2.1 Mathematical Foundation of XGBoost

XGBoost works by optimizing an objective function that combines the prediction accuracy (loss function) with model complexity (regularization). The objective function is typically defined as:

$$Obj(\theta) = \sum_{i=1}^{n} L\left(y_i, \hat{y}_i\right) + \sum_{k=1}^{T} \Omega(f_k)$$

Where:

- $L(y_i, \hat{y}_i)$ is the loss function, representing the difference between the true label y_i and the predicted label $\hat{y}_i$.
- $\Omega(f_k)$ is the regularization term that penalizes the complexity of the trees to prevent overfitting.
- T represents the total number of trees used in the ensemble.

In each iteration, XGBoost adds a new tree $f_t(x)$ to minimize the objective function:

$$\hat{y}^{(t)} = \hat{y}^{(t-1)} + \eta f_t(x)$$

Where:

- $\hat{y}^{(t)}$ is the prediction at iteration t,
- $\hat{y}^{(t-1)}$ is the prediction from the previous iteration,
- η is the learning rate that controls the contribution of each tree.

The gradient boosting technique improves the model by minimizing the residuals (errors) from the previous iteration, hence boosting the performance step-by-step.

5.2.2 Advantages of XGBoost in our model:

1. Regularization: The algorithm includes both L1 (lasso) and L2 (ridge) regularization, which helps to prevent overfitting by penalizing overly complex models.
2. Feature Importance: XGBoost can handle large feature sets and provides insights into the relative importance of features, aiding in feature selection.

5.3 Early Stopping

To prevent the model from overfitting, we applied early stopping during training. Early stopping halts the training process if the model's performance on the test dataset does not improve after a certain number of

rounds. In this case, we used early stopping rounds = 50, meaning that if there was no improvement in the model's performance on the test set for 50 consecutive boosting rounds, training would stop.

5.3.1 Advantages of Early Stopping:

1. Prevents Overfitting: By monitoring the model's performance on the test data, early stopping ensures that the model does not overfit the training data, leading to better generalization.
2. Efficiency: It reduces the computation time by halting the training once improvements plateau, rather than continuing for all specified boosting rounds.

Implementation of Early Stopping in XGBoost: During training, after each iteration, the model's performance is evaluated on the test set. If no improvement in performance is observed for 50 consecutive rounds, the model stops training, retaining the parameters from the round with the best performance.

5.4 Hyperparameter Tuning with Random Search

To optimize the XGBoost model, we employed Random Search for hyperparameter tuning. Random Search allows for a more efficient exploration of hyperparameters by randomly sampling combinations, offering an advantage over Grid Search in terms of computational efficiency.

5.4.1 Why Random Search?

Random Search was chosen because:

- It explores a random subset of hyperparameter combinations, leading to faster optimization compared to the exhaustive search performed by Grid Search.
- Random Search is often equally effective in finding optimal hyperparameters, especially when certain hyperparameters have minimal impact on the final model performance.

5.4.2 Random Search Process

Random Search was paired with 3-fold cross-validation to evaluate different combinations of hyperparameters and ensure the model generalizes well to unseen data. AUC-ROC was used as the primary scoring metric during the random search process, as it reflects the model's ability to distinguish between members with and without preventive care gaps.

Hyperparameter	Parameter Grid
n_estimators	[1000, 1100, 1200, 1250, 1300]
max_depth:	[8, 9, 10, 11, 12]
reg_lambda:	[100, 200, 300, 400, 500, 600]
learning_rate	[0.125, 0.1,0.075, 0.05, 0.025]
reg_alpha	[0.5, 0.6, 0.7, 0.8]
colsample_bylevel	[0.6, 0.7, 0.8, 0.9]
min_child_weight	[10, 30, 50, 70, 100]
eval_metric	′auc′
early_stopping_rounds	50
subsample	0.8

Note: The **'learning_rate**' controls the contribution of each tree to the final model, influencing the speed of learning. '**Max depth'** determines the maximum depth of each decision tree, affecting model complexity, while the '**n_estimators'** refers to the total number of boosting rounds (trees). The **'subsample'** ratio specifies the fraction of training data used for each tree to help prevent overfitting, and gamma sets the minimum loss reduction required for further partitioning a leaf node. '**Min_child_weight'** ensures that nodes have sufficient data to avoid overfitting by setting the minimum sum of instance weights in a child node. The '**colsample_bylevel'** controls the fraction of features sampled at each tree level, introducing randomness to improve generalization. '**reg_lambda'** (L2 regularization) penalizes large coefficients to prevent overfitting, and '**reg_alpha**' (L1 regularization) encourages sparsity by penalizing large feature weights, aiding in feature selection. Finally, '**early_stopping_rounds'** halts training if no improvement is observed over a specified number of rounds, further mitigating overfitting.

Table 5.1 *parameter grid used*

These values were used to fine-tune the model and achieve optimal performance.

5.4.3 Best Hyperparameters from first round of random search

The optimal hyperparameters identified after 100 iterations with 3-fold cross validation were:

- *reg_lambda*: 500,
- *'reg_alpha'*: 0.8,
- *'n_estimators'*: 1200,
- *'min_child_weight'*: 30,
- *'max_depth'*: 8,
- *'colsample_bytree'*: 0.9,
- *'learning_rate'*: 0.1,
- *subsample*: 0.8
- *early_stopping_rounds*: 50

***ROC-AUC on test dataset*=0.7661**

These hyperparameters provided the best balance between model accuracy and generalization.

5.5 Iterative Feature Selection and Importance-Based Refinement

To further improve the model's performance, we performed iterative feature selection based on feature importance scores provided by the XGBoost model. This process aimed to remove less important features thereby removing the noise ensuring that only the most impactful variables remained in the final model. XGBoost provides a ranking of features based on their importance in the prediction process. We used these feature importance scores to guide feature selection: Features with low importance (5%) were systematically removed to reduce noise and potential overfitting.

5.5.1 Iterative Process for Feature Reduction

After identifying less important or highly correlated features, we removed them one at a time and retrained the model, monitoring the performance after each iteration. This process was repeated until the model's performance metrics (such as AUC-ROC) were maximized.

Feature Reduction Process:

1. Evaluate feature importance scores from the current model and sort it in ascending order.
2. Retain features with cumulative contribution > 95 % to the prediction.
3. Remove the low-important features.
4. Retrain the model and assess its performance using the validation set.
5. Repeat the process until no further improvement in performance on the test dataset is observed.

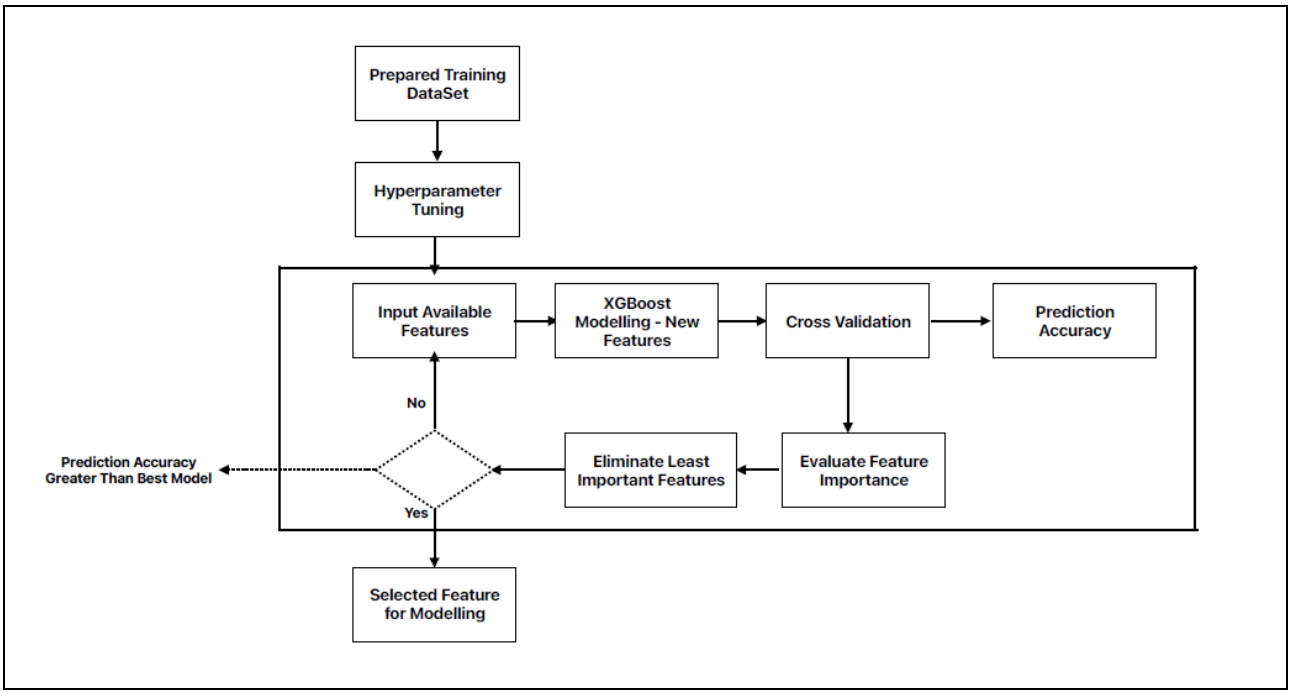


Fig 5.1 *Flow of Model Tuning Process*

5.5.2 Final set of Hyperparameters

After the feature reduction process, the XGBoost model was optimized using the following final set of hyperparameters:

- *n_estimators*: 5000
- *max_depth*: 14
- *reg_lambda*: 100
- *reg_alpha*: 0.3
- *colsample_bytree*: 0.9
- *min_child_weight*: 10
- *learning_rate*: 0.1
- *subsample*: 0.8
- *early_stopping_rounds*: 50

ROC-AUC on test dataset: 0.7789

These final parameters were selected based on their performance in maximizing the model's predictive power while minimizing overfitting, resulting in the most effective configuration for predicting preventive care gaps.

Key Parameters	Original XGBoost Model	After Random Search	Final hyperparameters after iterations
n_estimators	500	1200	5000
max_depth	6	8	14
reg_lambda	None	500	100
reg_alpha	None	0.8	0.3
colsample_bytree	1	0.9	0.9
min_child_weight	150	30	10
learning_rate	0.2	0.1	0.1
subsample	None	0.8	0.8
early_stopping_rounds	50	50	50

Table 5.2 *Hyperparameters used for optimization*

5.6 Final Feature Set

The iterative feature selection process, combined with hyperparameter tuning, resulted in a more refined feature set that significantly improved the model's performance by reducing noise and multicollinearity. The final set of features was selected based on their contribution to the model's predictive power, as evaluated through the feature importance scores generated by XGBoost.

The top 30 most impactful features that were retained after the feature reduction process are listed below. These features were critical to the model's success in identifying members at risk of missing preventive care:

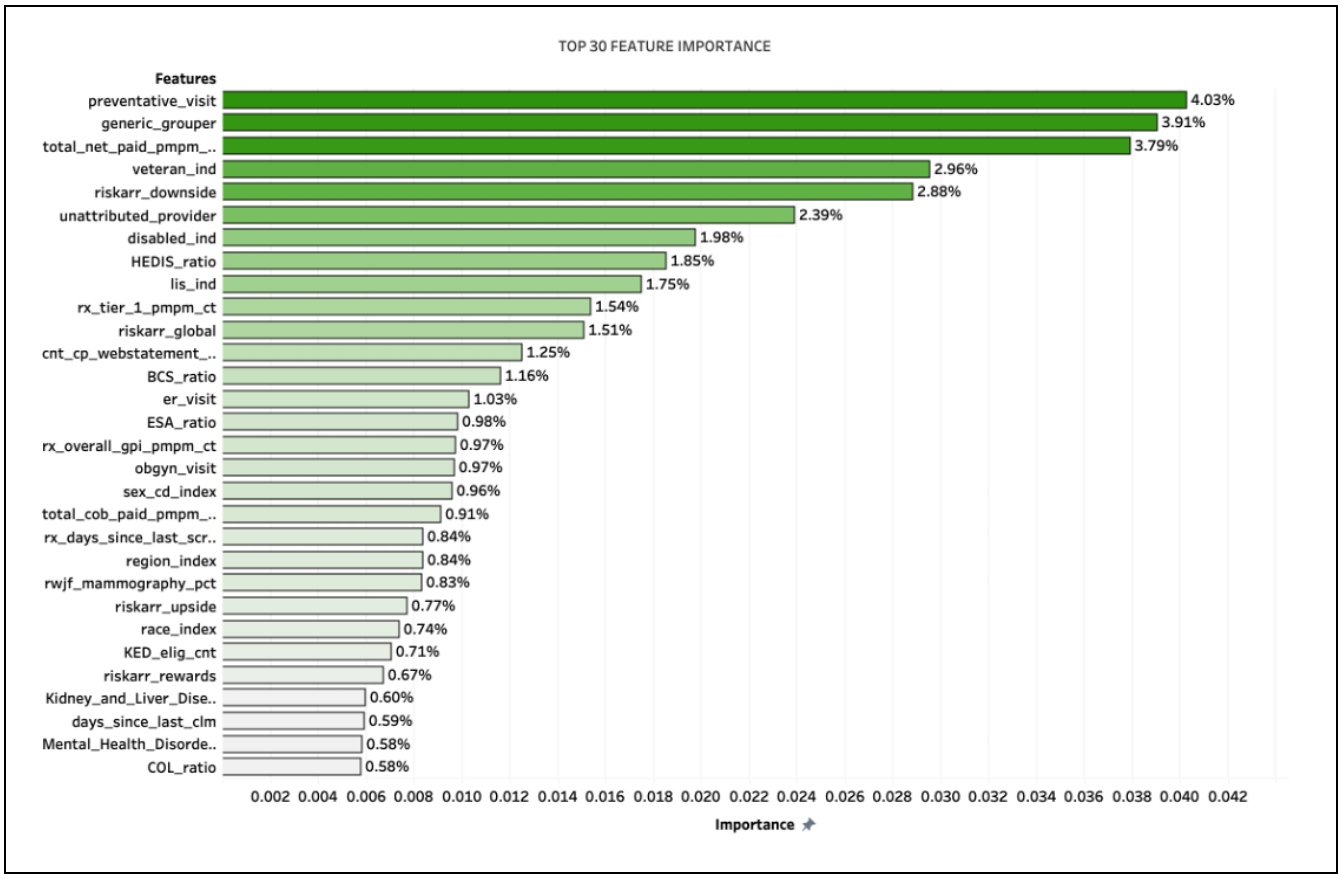


Fig 5.2 *Top 30 features*

The plot shows features such as ‘preventative_visit,’ ‘generic_grouper’, and ‘total_net_paid_pmpm_cost’ as the most influential. Features are ranked based on their contribution to the model's predictions, with higher bars indicating greater impact on the outcome. This chart highlights the key variables that drive model performance and influence the likelihood of missed preventive visits.

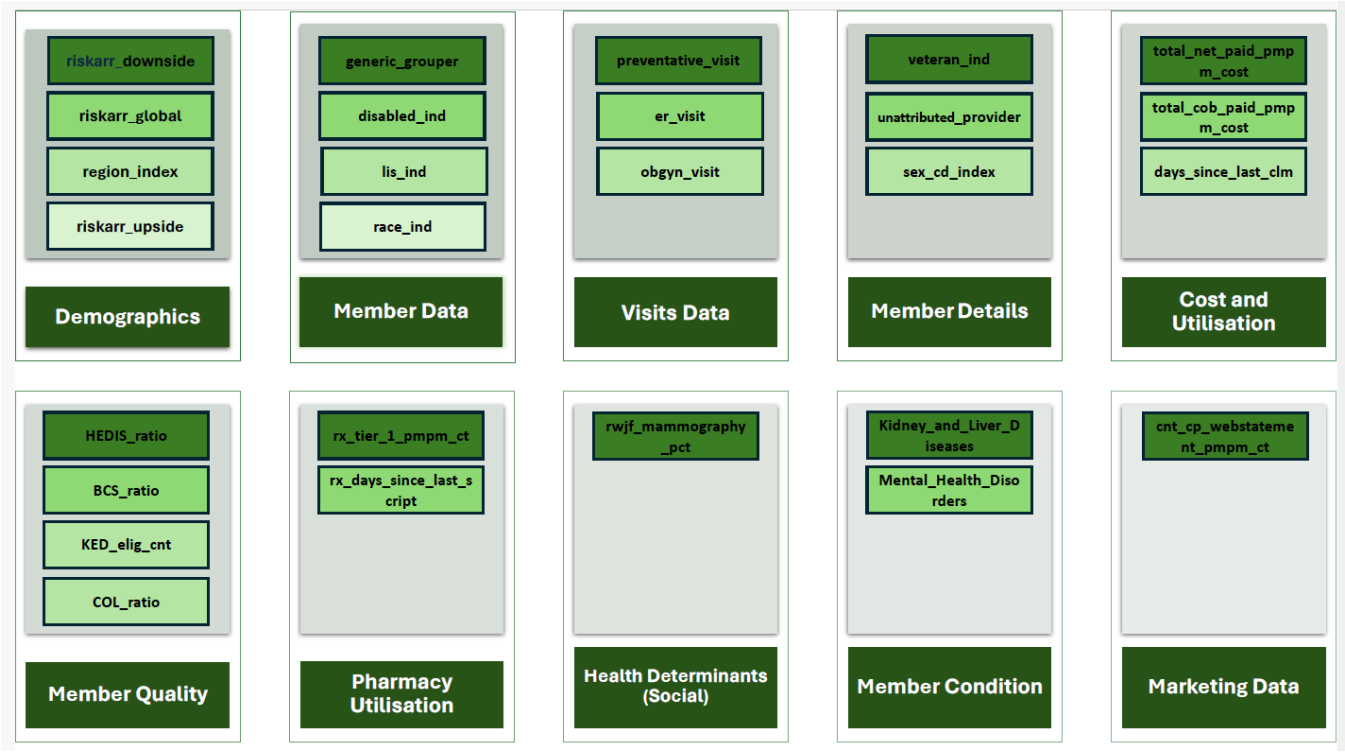


Fig 5.3 *Grouped categories of top 30 features*

This figure illustrates the key feature categories analyzed in the study, including Demographics, Member Data, Visits Data, Member Details, Cost and Utilization, Member Quality, Pharmacy Utilization, Health Determinants (Social), Member Condition, and Marketing Data. Each category contains specific features that were used to explore patterns and trends associated with missed preventive visits. By grouping the features in this way, the analysis provides a clearer understanding of how different factors contribute to healthcare engagement.

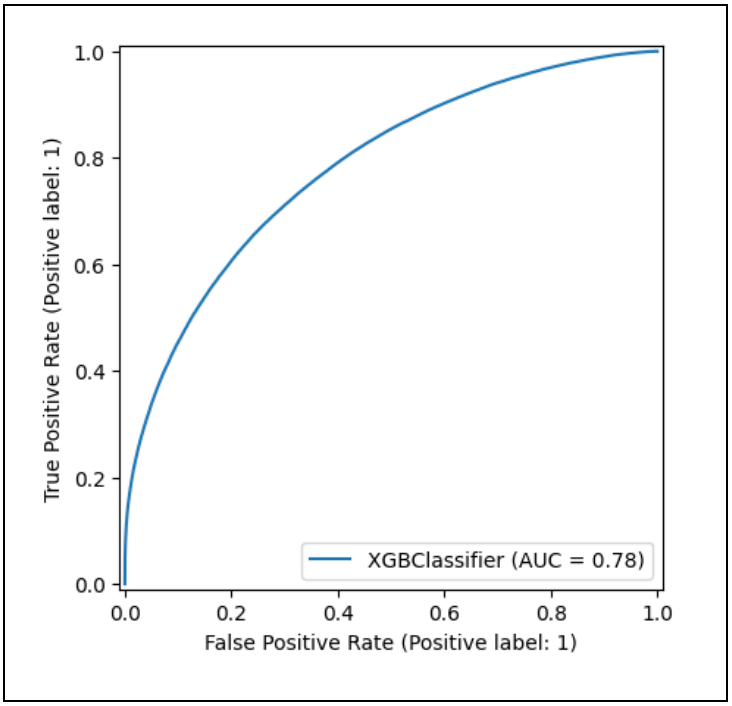
5.7 Model Evaluation and Metrics

After applying early stopping, hyperparameter tuning, and iterative feature selection, the final XGBoost model was evaluated using various performance metrics to ensure it met the predictive goals.

5.7.1 Evaluation Metrics

Final AUC-ROC Score: 0.782

This metric confirms that the XGBoost model is highly effective in predicting preventive care gaps.

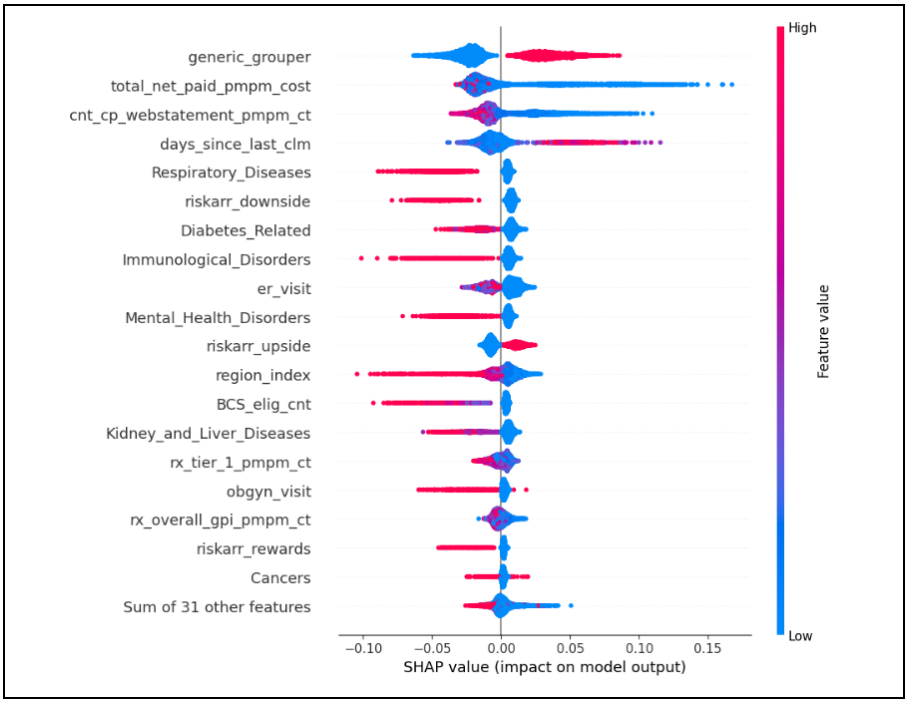


Note: The **AUC-ROC** curve (Area Under the Receiver Operating Characteristic Curve) is a performance measurement for classification models. The ROC curve plots the **True Positive Rate** (Sensitivity) against the **False Positive Rate** (1 - Specificity) at various threshold settings. The AUC represents the degree of separability between the classes, with a value of 1 indicating perfect classification and 0.5 representing a model no better than random guessing. A higher AUC indicates better model performance, as it shows that the model distinguishes well between positive and negative classes.

Fig 5.4 *AUC-ROC curve*

5.7.2 SHAP Beeswarm Plot Analysis

The SHAP beeswarm plot provides an in-depth analysis of the contribution of each feature to the model's predictions, ranking them by their importance. Each point in the plot represents a single observation's SHAP value for a specific feature.



Note: The SHAP Beeswarm plot visualizes the impact of each feature on the model's predictions. Each dot represents a SHAP value for a specific feature and instance. The color gradient from blue to red shows the feature value (low to high), and the position on the x-axis indicates whether the feature pushes the prediction higher or lower. Features are ranked by importance, with the most impactful features at the top. This plot provides an intuitive overview of feature influence across the entire dataset.

***Fig 5.5** SHAP value of features*

Key Observations from the Beeswarm Plot:

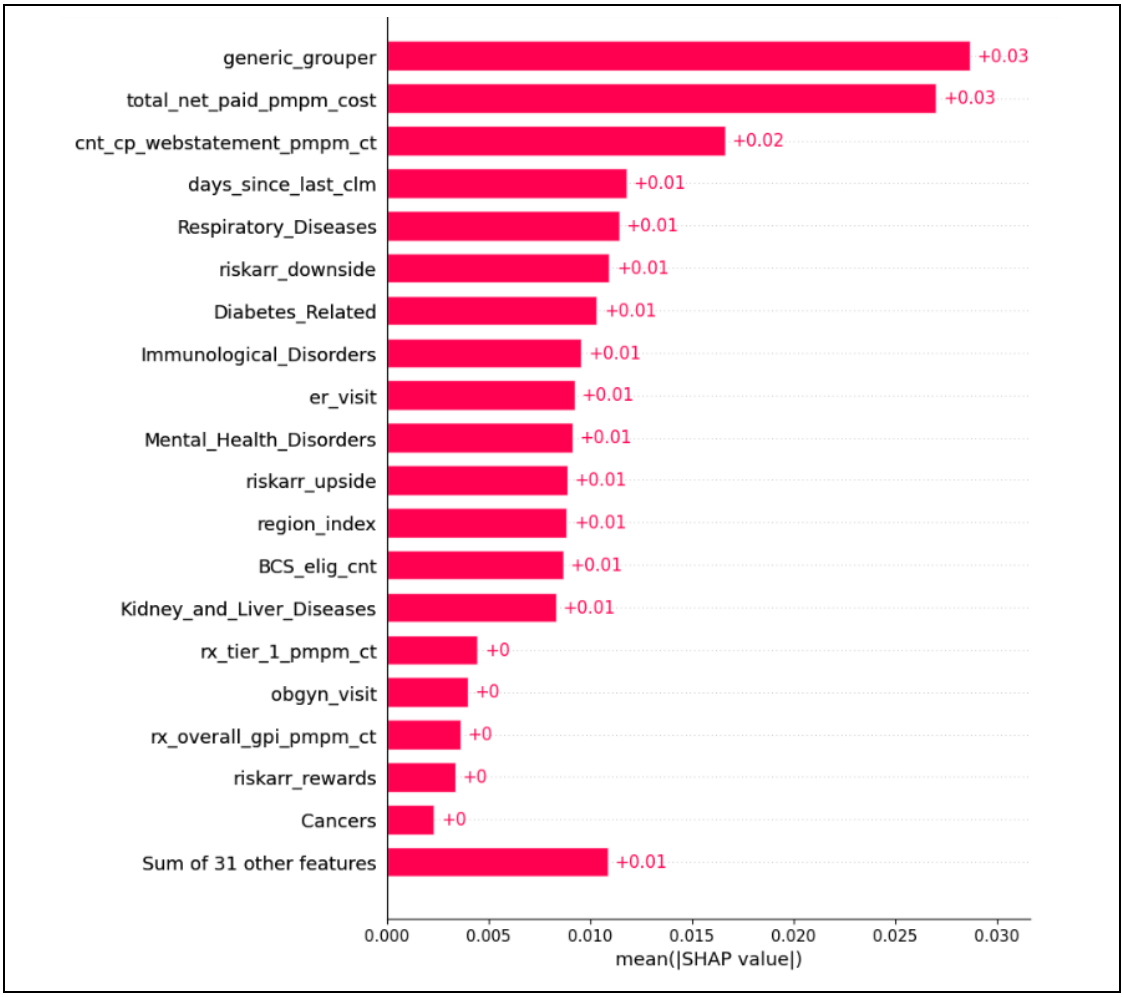
1. *Top Features:*
   - *generic_grouper, total_net_paid_pmpm_cost,* and *cnt_cp_webstatement_pmpm_ct* have the largest spread and strongest impact on the model's predictions. High values of *generic_grouper* (shown in red) and *total_net_paid_pmpm_cost* tend to have a significant positive effect on the predicted outcome, pushing it towards higher values.
   - *total_net_paid_pmpm_cost*: Smaller values (blue dots) are associated with positive SHAP axis, suggests that patients with smaller costs lead to higher chance of missing preventive care.
   - *cnt_cp_webstatement_pmpm_ct*: Smaller values (blue dots) are associated with **positive SHAP values**, suggesting that fewer web statement views or accesses are linked to a higher likelihood of missing preventive care.
2. *Moderate Features*:
   - Features like *days_since_last_clm, Respiratory_Diseases, riskarr_downside*, and *Diabetes_Related* contribute to the model, but their impact is less pronounced.
   - For instance, *Respiratory_Diseases* has mostly blue points with negative SHAP values, indicating that the presence of this condition generally has higher likelihood of going to preventive care.
3. *Lower Impact Features:*

- Features like *Cancers, riskarr_rewards*, and *rx_overall_gpi_pmpm_ct* have minimal impact on predictions as their SHAP values hover around zero. This suggests that these features do not significantly affect the model's outputs across most data points.

4. *Color Distribution:*

- The color gradient (from blue to red) provides further insight into how different feature values influence the prediction. For instance, red dots (high values) for *generic_grouper* and *total_net_paid_pmpm_cost* are mostly on the positive side of the SHAP value axis, indicating that higher values of these features push the predictions up.

5.7.3 Overview of mean absolute SHAP Value Graph:



Note: The SHAP Bar Plot displays the mean absolute SHAP values for each feature, ranking them by their overall importance to the model's predictions. The longer the bar, the more impact that feature has on the model. This plot provides a clear summary of which features are most influential in driving predictions across the dataset, helping to identify the key variables affecting model performance.

***Fig 5.6** SHAP bar plot*

1. *Feature Importance Ranking:*
    - The top three features (*generic_grouper, total_net_paid_pmpm_cost, cnt_cp_webstatement_pmpm_ct*) have the highest average SHAP values (around 0.02 to 0.03). This means that these features have the most significant influence on the model's predictions.
        - *generic_grouper* has the largest mean SHAP value (0.03), indicating that, on average, this feature has the strongest impact on predictions.
        - *total_net_paid_pmpm_cost* also ranks highly, suggesting that higher healthcare costs are consistently associated with higher predicted outcomes.
2. *Mid-Tier Features*:
    - Several features, such as *days_since_last_clm, Respiratory_Diseases,* and *riskarr_downsi*de, have mean SHAP values around 0.01. While they contribute to the model's predictions, their influence is less pronounced than the top features.
3. *Lower Impact Features:*
    - Features like *rx_tier_1_pmpm_ct, obgyn_visit, Cancers,* and *riskarr_rewards* have mean SHAP values close to zero, indicating that they have little to no influence on the model's predictions.
    - The "Sum of 31 other features" suggests that a combination of many smaller, less impactful features contributes minimally to the model's output..

6. Model Interpretation

The XGBoost model used in this analysis provides a robust prediction of patients who are likely to miss preventive care services. By leveraging SHAP values and feature importance rankings, we can group features into key categories to better understand the drivers behind preventive care engagement. Below is a detailed interpretation of the model, focusing on the different feature categories and how they contribute to the model's predictions.

1. Member Data

Member data, including features like *generic_grouper*, *disabled_ind*, and *lis_ind*, play a significant role in predicting preventive care behavior. Among these, *generic_grouper* stands out as the most influential feature in both the SHAP analysis and XGBoost feature importance rankings. This feature likely captures essential demographic or behavioral groupings that are critical in determining preventive care utilization. *generic_grouper* ='Y' indicate an increased likelihood of missing preventive care, possibly representing specific subpopulations that are less engaged in healthcare services. Additionally, the feature *disabled_ind*, which indicates whether a patient is disabled, also contributes moderately to the predictions. Patients marked as disabled may interact with the healthcare system differently, either increasing their engagement due to the need for regular medical attention or reducing it due to barriers in accessing care. The *lis_ind* feature, which flags patients eligible for a low-income subsidy (LIS), is another moderate predictor, reflecting the economic factors that might influence a patient's ability to engage with preventive care. Patients eligible for LIS may face access issues or have different healthcare utilization patterns, which affects their likelihood of receiving preventive services.

2. Demographics

Demographic factors, such as *riskarr_downside*, *riskarr_global*, *region_index*, and *riskarr_upside*, are important indicators of patient behavior. The feature *riskarr_downside*, which captures the predicted cost and utilization downside risk, is particularly impactful. Patients with higher downside risk are likely under closer monitoring and may receive more proactive interventions from healthcare providers, thus reducing their likelihood of missing preventive care. On the other hand, *riskarr_upside* captures the opposite risk, where patients with lower costs and utilization patterns might not receive as much attention, leading to a higher risk of missing care. *Region_index* also plays a role in understanding healthcare access. This feature represents geographic differences in healthcare services, and areas with lower access or fewer healthcare resources may have patients who are less likely to engage in preventive care. These demographic variables provide important context for understanding how external factors, such as geography and projected healthcare costs, influence patient behavior in preventive healthcare.

3. Member Details

The details of a member's healthcare journey, including features like *veteran_ind, unattributed_provider*, and *sex_cd_index*, also contribute to predicting preventive care behavior. *Veteran_ind* is particularly significant, indicating that veterans may have different healthcare-seeking behaviors compared to non-

veterans. Veterans often have access to specialized healthcare services, which might influence their engagement with preventive care. Meanwhile, the feature *unattributed_provider*, which flags patients who do not have an attributed healthcare provider, is a strong predictor of missed preventive care. Patients without a regular healthcare provider are less likely to receive the reminders and continuity of care that promote preventive services. *Sex_cd_index*, representing gender, while less important than other features, still provides valuable insight into how gender differences impact healthcare engagement. These members highlight how personal characteristics and healthcare continuity can affect preventive care utilization, with veterans and those with assigned providers being more likely to participate in preventive care.

4. Visits Data

Features capturing healthcare visits, such as *preventative_visit*, *er_visit*, and *obgyn_visit*, are highly predictive of future preventive care behavior. The feature *preventative_visit* emerges as the top factor in the XGBoost model's feature importance rankings. Patients who have already participated in preventive visits are much more likely to continue engaging with preventive care services. This finding aligns with established healthcare behavior models, where past behavior is a strong indicator of future engagement. Similarly, *er_visit*, which captures emergency room visits, indicates a moderate impact on the model. Patients with frequent ER visits may rely on urgent care rather than preventive services, which could influence their future healthcare behavior. *Obgyn_visit* is particularly relevant for female patients, as regular OB/GYN visits may correlate with higher preventive care utilization in certain demographic groups. These visit-related features show that engagement with the healthcare system, whether through preventive or emergency services—provides valuable insight into a patient's likelihood of participating in preventive care.

5. Cost and Utilization

Cost and utilization metrics, including *total_net_paid_pmpm_cost*, *total_cob_paid_pmpm_cost*, and *days_since_last_clm*, are among the most significant predictors of preventive care behavior. The feature *total_net_paid_pmpm_cost*, which represents the total healthcare costs paid per member per month, is consistently ranked as one of the top predictors. Higher healthcare costs typically indicate greater engagement with the healthcare system, resulting in a lower likelihood of missed preventive care. Patients who spend less on healthcare services, on the other hand, may be less engaged and more likely to miss preventive services. *Total_cob_paid_pmpm_cost*, which refers to coordination of benefits payments, is another important feature. Higher values of this feature suggest that the patient is receiving care from multiple sources, which may either increase their access to preventive care or complicate coordination, depending on the situation. *Days_since_last_clm* is another key indicator, as patients who have recently filed a claim are more likely to engage in preventive care. Recent claims indicate a higher level of healthcare interaction, reducing the chances of missing preventive services.

6. Member Quality

Quality-related features, such as *HEDIS_ratio*, *BCS_ratio*, *ESA_ratio*, *KED_elig_cnt*, and *COL_ratio*, also contribute to the model's predictions. *HEDIS_ratio* is a key indicator of healthcare quality, showing

how well a patient adheres to healthcare measures that are linked to positive health outcomes. Patients with higher *HEDIS_ratio* values are more likely to engage in preventive care, as they are more closely aligned with healthcare quality standards. Other quality metrics, such as *BCS_ratio, ESA_ratio,* and *KED_elig_cnt,* while less significant, still provide valuable insights. These features reflect patient adherence to specific healthcare protocols and services, which can predict their likelihood of receiving preventive care. Overall, patients who score well on these quality measures are more engaged in the healthcare system and, consequently, less likely to miss preventive services.

7. Pharmacy Utilization

Pharmacy utilization features, including *rx_tier_1_pmpm_ct, rx_overall_gpi_pmpm_ct*, and *rx_days_since_last_script*, provide insights into a patient's medication usage and its correlation with preventive care behavior. Rx_tier_1_pmpm_ct captures the use of lower-cost, tier 1 medications, and its moderate importance in the model suggests that patients who frequently use these medications may have different engagement patterns with the healthcare system. *Rx_overall_gpi_pmpm_ct,* which represents overall pharmacy utilization, also contributes to the model's predictions. Patients with higher pharmacy usage may have greater healthcare engagement overall, leading to more frequent preventive care participation. Finally, *rx_days_since_last_script* indicates how recently a patient has filled in a prescription. Patients who have recently obtained medications are more likely to engage in preventive care, reflecting ongoing interactions with healthcare providers.

8. Marketing Data

Engagement with healthcare portals, represented by *cnt_cp_webstatement_pmpm_ct*, is a key feature in predicting preventive care behavior. Patients who access their healthcare information online tend to be more engaged and informed, which increases their likelihood of participating in preventive services. The use of online statements and portals indicates higher engagement with the healthcare system, as these patients are likely taking a proactive approach to managing their health. The SHAP analysis shows that patients with higher *cnt_cp_webstatement_pmpm_ct* values are less likely to miss preventive care, making this feature an important predictor in the model.

9. Member Condition

Finally, patient conditions, such as *Kidney_and_Liver_Diseases* and *Mental_Health_Disorders*, provide valuable context for understanding preventive care behavior. Patients with chronic conditions, like kidney and liver diseases, are often more engaged with the healthcare system due to their need for regular monitoring and treatment. This engagement reduces their likelihood of missing preventive care. *Mental_Health_Disorders*, while not as highly ranked, still plays a role in predicting preventive care participation. Patients with mental health disorders may face additional challenges in accessing preventive services but may also have more frequent interactions with healthcare providers, depending on their condition.

7 Business Implications and Recommendations

Preventive care is a critical component of maintaining long-term health and reducing healthcare costs, yet many populations show low engagement with preventive services, leading to missed opportunities for early disease detection and improved overall well-being. The following recommendations address key challenges in preventive care engagement by targeting specific groups, such as low-cost members, individuals with long gaps since their last claim, veterans with disabilities, and the crucial role that primary care physicians (PCPs) and specialists play in promoting these services. Through the implementation of tailored outreach programs, partnerships with pharmacies and veteran organizations, and enhanced communication between healthcare providers, these strategies aim to increase participation in preventive care. Offering incentives, leveraging pharmacy visits, and improving care coordination can significantly boost engagement and ensure more individuals receive the preventive services they need to maintain optimal health.

7.1 Low-Cost Members Show Declining Engagement Over Time

People who incur lower healthcare costs for insurance companies and have lower prescription usage are the least likely to attend preventive visits, with 81.50% (~100,000 members) missing these important appointments, highlighting the need for targeted interventions. To close this gap, automated outreach campaigns can be an effective strategy to engage this group, especially those with an average prescription count of less than 0.5 per month. These campaigns should focus on the long-term benefits of preventive care, such as early disease detection and improved overall health, to encourage participation. Offering small incentives, such as a free annual wellness checkup and credits for completing preventive care visits, can further motivate members to act and attend health screenings. Research shows that financial incentives can greatly enhance preventive care engagement, making this approach a promising solution[1].

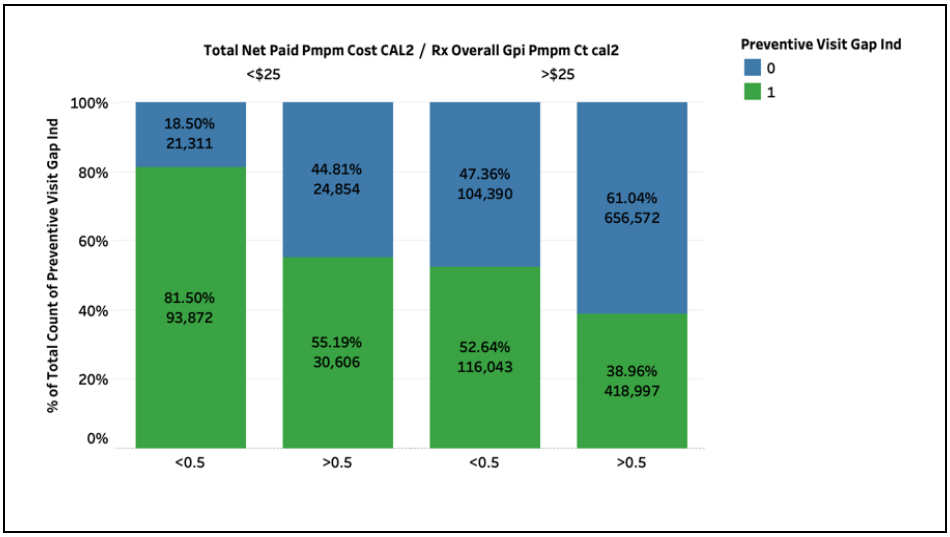


***Fig 7.**1 Preventive Care Gaps by Prescription Usage and Healthcare Costs*

1Cohen, D., et al. (2014). "Effect of Preventive Counseling During Acute Care Visits." Journal of General Internal Medicine.

Strategy for Implementation:

• Automated Reminders: Leverage Humana's CRM system to send regular SMS, postcards, and phone call reminders to this targeted population.

• Incentive Programs: Partner with local urgent care or pharmacy to offer low-cost incentives (e.g., health-related gift cards). These should be redeemable upon completing preventive care screenings.

• Pilot Program: Test these incentives in specific regions and evaluate effectiveness after 6 months, adjusting incentives as necessary based on participation rates.

7.2 Long gaps since last claim led to lower preventive care engagement

As the number of days since the last claim increases, there is a clear trend showing that the percentage of the preventive visit gap rises, with members who have gone longer without making a claim being progressively less likely to engage in preventive care. To address this, a two-pronged approach can boost engagement. First, implement pharmacy-based services for members who haven't visited in over 160 days, offering convenient preventive care options during prescription pickups and over the counter (OTC) medicine purchases. OTC medicine visits offer an opportunity to engage members who may not be visiting for prescriptions but are still using pharmacy services. For members with more than 280 days since their last claim—who rarely visit pharmacies or hospitals—partner with major chains like Walgreens or CVS to advertise Humana's preventive care programs. This multi-channel approach ensures consistent outreach, closes the preventive care gap, and improves overall health outcomes[2].

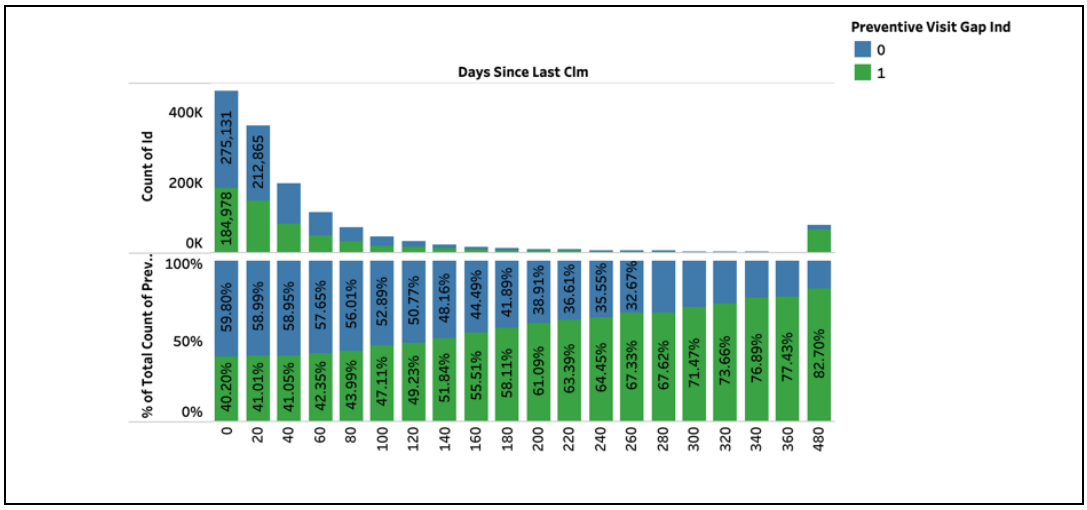


***Fig 7.2** Preventive Care Gaps by Days Since Last Claim*

2Meeker, D., et al. (2018). "Impact of Personalized PCP Reminders on Preventive Care Engagement." American Journal of Managed Care.

Strategy for Implementation:

- **Partner with Pharmacies**: Collaborate with in-house and external pharmacy chains like Walgreens, CVS, or local pharmacies to provide preventive care services such as flu shots, blood pressure checks, and wellness screenings.
- **Leverage OTC Purchases**: Use over the counter (OTC) medicine purchases as touchpoints to engage members who may not be filling prescriptions but are visiting the pharmacy for other needs. Display reminders for preventive care services near OTC aisles.
- **In-Pharmacy Signage and Outreach**: Set up educational displays or kiosks in pharmacies to inform customers about available preventive care services. Offer printed materials at checkout points and during OTC purchases.

7.3 Veterans, especially Disabled Veterans, have lower Preventive Care Engagement

Non-veterans without disabilities are the most engaged in preventive care, while veterans with disabilities show the largest preventive care gap, with 73.24% not engaging. This highlights the need for tailored outreach and support to boost preventive care participation among veterans, particularly those with disabilities. To close this gap, developing veteran- and disability-specific programs is crucial. Partnering with organizations like Veterans of Foreign Wars to offer wellness screenings and preventive health days can make preventive services more accessible to these groups. For individuals with disabilities, introducing in-home preventive screening services can help address accessibility challenges. Additionally, targeted campaigns directed at the caregivers of these individuals can further drive engagement. By collaborating with veteran organizations, healthcare providers can gain access to valuable data, allowing them to create personalized outreach efforts and prioritize high-risk members for preventive care. Offering discounts on assistive technology for people with disabilities can further encourage engagement, making it easier for them to access and participate in preventive care. Research confirms that veteran-focused programs significantly improve engagement, making this strategy both effective and essential[3].

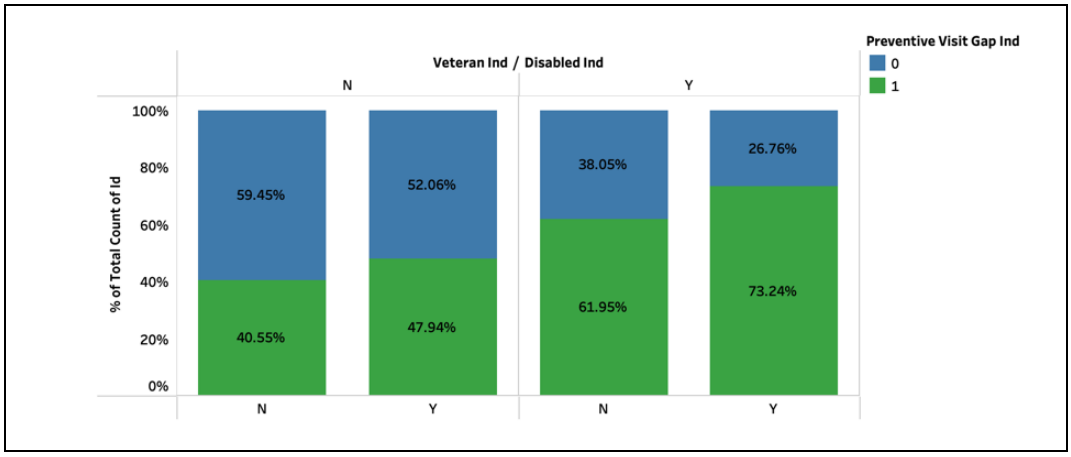


Fig 7.3 *Preventive Care Gaps Among Veterans and Disabled Individuals*

3Zinzow, H. M., Brooks, J. J., & Sternberg, J. E. (2020). "Veterans and Preventive Care: Analyzing the Impact of Tailored Health Interventions." Veterans Health Journal.

Strategy for Implementation:

- Veteran Outreach Campaign: Create targeted communication campaigns with veteran-specific messaging. Utilize veteran organizations to amplify these campaigns through their channels.
- **Veteran Health Clinics:** Establish in-home health clinics specifically for veterans, where nurses visit the veterans' homes to conduct preventive screenings and provide wellness education. These clinics can also be offered at veteran centers or through mobile health units to increase accessibility and convenience for veterans.
- **Peer Support Programs:** Create peer support systems where veterans motivate fellow veterans to engage in preventive health services. Leveraging behavioral psychology principles, such as Authority Bias (*Influence: The Psychology of Persuasion*[4]), target respected community leaders, like a general in the veteran community or a pastor in church groups, to lead and promote these programs. Their influence can significantly boost participation and trust within the veteran community.

7.4 Strategies for PCPs and specialists to improve engagement in preventive care visits

Both PCPs and specialists play crucial roles in improving patient engagement in preventive care by proactively discussing its importance during every interaction. PCPs should incorporate preventive care conversations into every visit, regardless of the primary reason, as this approach has been shown to increase screening rates by 30%[5]. Similarly, specialists can reinforce these messages during their consultations, with studies showing that when specialists emphasize the need for preventive screenings, patients are 35% more likely to follow through[6]. In addition to direct discussions, care coordination between PCPs, specialists, and care coordinators is vital for ensuring consistent follow-up and support, especially for high-risk patients. This collaborative approach has been found to increase preventive care use by 40% among at-risk populations[7]. Moreover, multidisciplinary collaboration between providers further boosts engagement, with coordinated efforts leading to a 22% increase in preventive care participation[8]. By working together, sharing information, and maintaining a focus on preventive care, both PCPs and specialists can significantly improve patient outcomes.

[4] Cialdini, R. B. (2009). *Influence: The psychology of persuasion* (Revised ed.). Harper Business.
[5] Fisher, E., et al. (2017). "Impact of Care Coordination on Preventive Service Use." Commonwealth Fund.
[6] Chen, L., et al. (2016). "Specialist Referral and Preventive Screening Engagement." JAMA.
[7] Gottlieb, L., et al. (2019). "Addressing Social Determinants of Health to Increase Preventive Care." Health Affairs.
[8] McCarthy, D., et al. (2017). "The Benefits of Coordinated Care in Preventive Service Engagement." Kaiser Permanente

7.5 Targeting Preventive Care Engagement for Unattributed Providers with Low Web Engagement and Low Prescription Count Populations

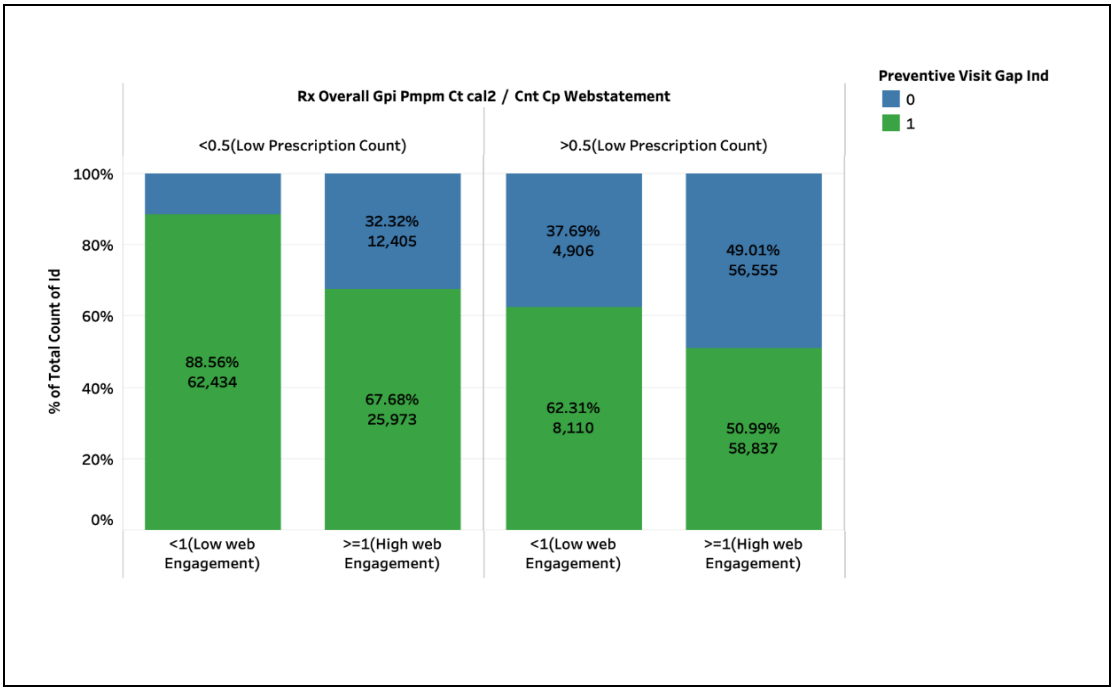


Fig 7.4 *Preventive Care Gaps by Prescription Count and Web Engagement*

The graph highlights *that individual* with low prescription counts, low web engagement *and* who *have* no primary care physician assigned to them have the highest preventive care gaps, with 88.56% not participating in preventive visits. In contrast, those with high web engagement and higher prescription use show improved engagement, though gaps *remain*. To address this, a multi-channel approach is recommended. For unattributed members with low web engagement, direct mail campaigns and phone outreach should be utilized, offering personalized education on the importance of preventive care and providing easy scheduling options. Additionally, Community outreach programs in senior centers and local organizations can further reach elderly populations, with local influencers and community leaders advocating for preventive care. Studies show that such strategies can reduce preventive care gaps by 18% to 30%, making them effective in increasing engagement among this underserved group.

8. Competitor Strategies to Engage Members in Preventive Care

Competitors of Humana, such as UnitedHealthcare, Aetna, Cigna, and Blue Shield, have implemented several innovative strategies to engage members and close gaps in preventive care. For example, UnitedHealthcare uses predictive analytics and its HouseCalls program, where nurse practitioners visit members' homes to provide preventive services for those who may not engage with traditional healthcare settings. A study in *Health Affairs* reported that UnitedHealthcare's personalized outreach through predictive analytics increased preventive service utilization by 25%[1]. Aetna, in partnership with CVS Health, has leveraged retail clinics like MinuteClinics to offer easy access to preventive care at convenient locations, resulting in increased flu vaccinations and health screenings[2]. Cigna has integrated behavioral health services into its preventive care strategies, recognizing that addressing mental health improves engagement in physical preventive care, leading to a 27% increase in preventive service use, according to a study published in *JAMA*[3]. Lastly, Blue Shield has focused on community partnerships and telehealth expansion to improve preventive care access, especially in underserved areas. Blue Shield has partnered with local community health organizations to host preventive care events and deployed telehealth options for routine checkups, which has led to a 20% increase in preventive care participation, as reported by the *Blue Shield Annual Report*[4]. These competitors' strategies, driven by data analytics, personalized outreach, and convenient care options, have proven effective in closing preventive care gaps and can serve as benchmarks for Humana's efforts.

[1] *Health Affairs* (2021): "Predictive Analytics in Healthcare."
[2] *American Journal of Managed Care* (2021): "Retail Clinics and Preventive Care Uptake."
[3] *JAMA* (2022): "Behavioral Health Integration and Preventive Care."
[4] *Blue Shield Annual Report* (2022): "Community Partnerships and Telehealth for Preventive Care."

9. Future Scope

Incorporating additional data sources can provide deeper insights into member engagement in preventive care and reveal underlying barriers that may currently be overlooked. Geospatial Information can identify areas where members have limited access to healthcare facilities, enabling a more targeted approach to improving engagement in regions with fewer health resources. Additionally, Social Determinants of Health (SDOH)—including factors like socioeconomic status, education, employment, and transportation access—have a significant impact on healthcare engagement. Integrating SDOH data would allow for a more holistic understanding of the challenges faced by certain populations, helping to create more effective interventions for those most at risk of disengagement.

Moreover, leveraging Natural Language Processing (NLP) on unstructured data sources, such as call center transcripts, survey responses, or member feedback, could uncover specific barriers to preventive care that may not be evident from structured data alone. NLP could provide valuable insights into member concerns, frustrations, and preferences, which can inform outreach strategies. Additionally, understanding the psychographic data of members, such as their attitudes toward healthcare services, willingness to use technology, or trust in healthcare providers, would allow for a more personalized and targeted approach to improving preventive care engagement. These combined data points would provide a more nuanced and comprehensive framework for addressing preventive care gaps across various member groups.

10. Conclusion

In summary, preventive care engagement is a crucial determinant of both health outcomes for members and the financial sustainability of health plans like Humana's LPPO. Through data analysis and advanced modeling techniques such as XGBoost, this report has explored several key factors that influence preventive care engagement. Members with lower prescription usage, veterans—particularly those with disabilities—and individuals with long gaps since their last claims represent the most significant areas of disengagement. Each of these groups faces unique barriers to preventive care, which if addressed properly, could lead to significant improvements in member health and reductions in long-term healthcare costs for Humana.

The data-driven approach detailed in this report emphasizes the importance of tailored interventions for these specific populations. By utilizing feature engineering, statistical modeling, and SHAP analysis, we have identified critical predictors of preventive care gaps and provided actionable insights for engaging these members. The analysis shows that strategic outreach efforts, such as offering incentives, creating veteran-specific programs, and leveraging pharmacy services, are effective in closing these gaps. Furthermore, predictive models allow for better identification of high-risk members, enabling Humana to implement targeted solutions efficiently.

Moving forward, the application of machine learning techniques and a focus on continuous data refinement will allow Humana to proactively address preventive care gaps before they widen. By implementing these insights and fostering stronger member engagement in preventive services, Humana can improve health outcomes while also mitigating financial risks associated with unengaged populations.

This report underscores the strategic importance of enhancing preventive care engagement and provides a comprehensive framework for improving the health and well-being of LPPO members. With tailored solutions and ongoing efforts to refine predictive modeling, Humana is well-positioned to lead the way in preventive care excellence.

11 Additional References

For Tableau visualizations – Click here!

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. https://doi.org/10.1145/2939672.2939785

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS '17)*. https://doi.org/10.5555/3295222.3295230

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—Big data, machine learning, and clinical medicine. *The New England Journal of Medicine*, 375(13), 1216–1219. https://doi.org/10.1056/NEJMp1606181

Hainmueller, J., & Hazlett, C. (2014). Random forests for causal inference: Statistical adjustments for selection bias. *Political Analysis*, 22(4), 492–502. https://doi.org/10.1093/pan/mpu017

Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542, 115–118. https://doi.org/10.1038/nature21056

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *The New England Journal of Medicine*, 380(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259

Jiang, F., Jiang, Y., Zhi, H., et al. (2017). Artificial intelligence in healthcare: Past, present, and future. *Stroke and Vascular Neurology*, 2(4), 230–243. https://doi.org/10.1136/svn-2017-000101

Reddy, S., Fox, J., & Purohit, M. P. (2019). Artificial intelligence-enabled healthcare delivery. *Journal of the Royal Society of Medicine*, 112(1), 22–28. https://doi.org/10.1177/0141076818815510

Shortreed, S. M., Cook, A. J., Coley, R. Y., et al. (2019). Challenges and opportunities for using big data to improve patient care. *Annals of Internal Medicine*, 171(9), 685–690. https://doi.org/10.7326/M19-1511

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning with applications in R. *Springer Texts in Statistics*. https://doi.org/10.1007/978-1-4614-7138-7

Vashistha, S., Burman, D., & Gottlieb, L. M. (2021). Evaluating social determinants of health in preventive care outreach. *Journal of General Internal Medicine*, 36(8), 2485–2492. https://doi.org/10.1007/s11606-020-06313-4

https://spark.apache.org/docs/latest/api/python/

Reference: Apache Spark™ - Unified Analytics Engine for Big Data
Apache Software Foundation. (2023). Retrieved from Website: https://scikit-learn.org/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830. Retrieved from https://jmlr.org/papers/v12/pedregosa11a.html

SHAP (SHapley Additive exPlanations) Documentation:
SHAP is a popular framework for explaining machine learning models.

- Website: https://shap.readthedocs.io/