# Humana-Mays Healthcare Analytics 2021 Case Competition

COVID-19 Vaccine Access Equity:

Hesitancy Prediction and Segmentation Analysis

October 10th, 2021

# Table of Contents

# 1. Executive Summary

The COVID-19 (Coronavirus Disease 2019) pandemic has dramatically changed the lives of everyone, and countless forces have been joined together to restoring a normal lifestyle. The invention of the COVID-19 vaccination is one of the most effective measures to achieve that objective. However, disparities in various factors have created barriers over equal vaccination opportunities for part of the population. As a leading healthcare provider in the US, Humana is dedicated to helping its underprivileged members gain access to the vaccine. Therefore, the objective of this report is to assist Humana in achieving its goal by leveraging the power of Big Data and Machine Learning.

The overall objective is broken down into three tasks:

1) Build a predictive model to identify the most vaccine-hesitant members

2) Based on insights from the model, uncover the underlying reasons and segments among the hesitant members

3) Design targeted solutions for different segments with comprehensive analysis

Given 974,842 records of Humana's MAPD members with wide-ranging information, we started with researching relevant topics, performing exploratory data analysis, and data cleaning to gain a deeper understanding. Feature selection and engineering techniques were also applied to prepare for modeling. A two-stage process is then carried out to train and compare the performances of 6 different models. We selected the ROC-AUC score as the metric for evaluation and found the XGBoost Classifier (ROC-AUC score of 0.6839) to be our final model with the most predictive power. By studying the important features, we extracted insights for potential segmentation.

For the second task: we successfully distinguished four segments from the population: members with financial distress, racial minority members, members from selected regions, and members with disabilities. Every segment has significantly lower vaccination rates, and the four groups cover 73% of the unvaccinated population. The rest of the population are reckoned to have more personal reasons to cause their hesitancy.

To address the four segments and the rest, we propose the CoVAE (COVID Vaccine Accessibility and Equity) Program as a comprehensive, actionable, measurable solution with targeted strategies. We recommend carrying out CoVAE in three stages, ranked by priority determined by infection risks of segments and ease of implementation:

- **Stage 1** consists of carefully designed financial incentives open to all members in limited times and specifically targets those with financial distress
- **Stage 2** will be provided to selected members only in the opt-in form to avoid extra costs. It includes:
    - Finding the appropriate opinion leaders to influence racial minority groups
    - Providing transportation and special care for people with disabilities
    - Providing pre-vaccination screening and consulting and setting up mobile vaccination sites in regions with high hesitancy

- **Stage 3** targets the unsegmented members in two ways:
    - Building a network model with consented patient information to locate members with higher degrees of connections and prioritize on persuading them first so that their social networks will be driven into vaccination via peer influences
    - Launching a personalized analytical platform

The CoVAE program is well-designed to reduce barriers for each segment and ensure accessibility and equity for the underserved members. By adopting our recommendation, Humana is estimated to convert 14.6% unvaccinated and hesitant members for vaccination and generate valuable assets and tools such as the connection with opinion leaders and the member network model to be used for future purposes.

# 2. Case Context

## 2.1 Case Background

Since the launch of the COVID-19 vaccine, the public has not unanimously welcomed this preventative health measure against the pandemic. According to a study conducted by Facebook, Carnegie Mellon University, and the University of Pittsburgh in the first half-year of 2021, there is still a certain group of people with strong hesitancy in receiving the vaccination, despite a decreasing trend[1]. The relevant factors for the vaccine hesitancy found in the study included younger age, non-Asian race, less than college education qualification, living a rural county, living in a county with a higher Trump-supporting rate, etc. Based on that basic information of vaccine hesitancy, we aimed at having further understanding with the provided information of Humana members.

It is of high priority for Humana to raise the vaccination rates among their members, especially those in the most vulnerable and underserved groups. With an analytical model predicting member-level vaccination status, Humana will be able to apply personalized outreaches to enable these people to receive proper and timely health measures.

Therefore, we are going to leverage the power of the data analytics and the given datasets to (1) build a classification model that can predict whether a Humana member is hesitant to get COVID vaccination and (2) find out the most important features that contribute to the vaccine hesitancy and figure out corresponding solutions.

## 2.2 Business and Data Problem Statement

The current business problem is to identify the underserved groups within the Humana Medicare Advantage Prescription Drug member population (referred as “population” below) which are hesitant to get vaccinated.

Based on the provided data, the data problem is broken down into (1) establishing a precise classification model based on the personalized and regional factors to distinguish whether a Humana member is hesitant to receive vaccination, and (2) identifying the factors that distinguish the underserved subset of the populations from the public.

### 2.3 Statement and Definition of the Metrics

As for the first data problem, our models will return a binary prediction on whether a particular member is hesitant to receive COVID vaccination. To obtain the most accurate prediction, we will select the analytical model with the highest *Area Under the Curve for the Receiver Operator Characteristic curve (AUC-ROC)*, among all model candidates, which is understood as a proxy of model prediction accuracy and performance. Details will be covered in the model evaluation section below.

The second data problem will be evaluated based on a comparison of the *marginal rate of vaccination for underserved groups* against the *marginal rate of vaccination for the complementary groups and overall rate of vaccination* for the entire population. The underserved groups will be defined by some independent factors, which we assume should include race, living regions, and ages based on the past study[2]. For each independent factor, the population is classified into a pair of underserved and complementary groups and the comparison is carried out accordingly.

Assuming the dataset is representative enough for the entire Humana MAPD member populations, all to our dataset, they are defined as follows:

- *The marginal rates of vaccination for underserved groups*: the percentage of a subset of population that has been vaccinated given the underserved personalized or regional attributes.
- *The marginal rate of vaccination for the complementary groups:* the percentage of a subset of the population that has been vaccinated given that they belong to a complementary group to the former underserved groups.
- *The overall rate of vaccination*: the percentage of the entire population that has been vaccinated.

If the marginal rate of vaccination of underserved subgroups is lower than that of the complementary group or the overall rate of vaccination, it implies that the selected factors do contribute to some extent to the vaccine hesitancy.

## 3. Data Preparation

### 3.1 The Dataset

Here is a summary of the dataset provided for our analysis:

- Training data: 974842 records by 368 variable columns, with a response column *covid_vaccination*
- Holdout data: 525158 records by 367 variable columns

### 3.2 Understanding the Data

#### 3.2.1 Understanding Features

Each data row stores data for each a Humana MAPD member, including his or her vaccination status, which will be our targeted response. Other member-level data includes biographical characteristics, financial status, medical expenses, etc.

Based on the information and data dictionary from Humana, the features were initially classified into eight categories:

- Medical Claims Features
- Pharmacy Claims Feature
- Lab Claims Features
- Demographics/Consumer Data
- Credit data
- Condition Related Features
- CMS Features
- Other features

To further understand our features and gain insights for future feature engineering, we investigated each variable according to their prefix.

There were 34 types of prefixes in total. Based on their sources and meanings, we finally classified them into 12 categories as shown in table 1. We have attached hyperlink to each source in the "Group" column.

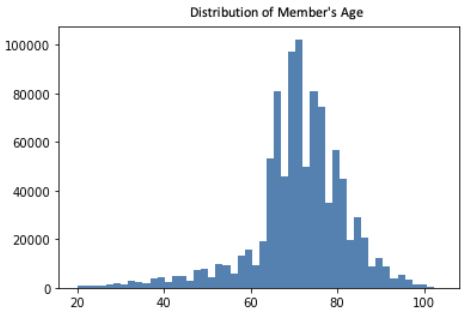
**Table 1 Feature Categories and Other Information**

| `Group | Name/ Prefix | Variable Meaning | Scale | Statistical Level | Null Data Counting |
|---|---|---|---|---|---|
| ATLAS Open Data | Prefix: atlas | Economic and poverty situation | Numerical/ Binary | Region level | Comparatively high null value rate |
| Regional Healthcare Factors | Prefix: rwjf | Regional healthcare and economic factors | Numerical | Region level | Residential Segregation has higher null value rate; others with moderate null value rate |
| **Census Data** | Prefix: cons | Regional census data | Numerical | Region level | Comparatively high null value rate |
| **Acute Admits Record** | Prefix: auth | Authorization for acute admits in the past three months | Numerical | Person level | Comparatively low null value rate with some all-zero columns |
| **Bank Credit** | Prefix: Credit | Personal bank account and statement information | Numerical | Person level | Comparatively low null value rate |
| Medicare Program | Prefix: cms | Claims and Medicare plan payment | Numerical/ Categorical | Person level | Part D related features have higher null value rate; others have comparatively low null value rate |
| Health related claims | Prefix: ccsp, cci | Number of claims related to health service | Numerical | Person level | Comparatively low null value rate |
| **Prescription Cost Trend** | Prefix: rcc, mcc, med, oontnk | Trend of prescription cost for specific diseases | Categorical | Person level | Comparatively low null value rate |
| **Rejected Claims** | Prefix: rej | Trend of rejected non-behavioral health claims | Categorical | Person level | Comparatively low null value rate |

| | | | | | |
|---|---|---|---|---|---|
| **Revenue Code** | Prefix: rev | descriptions and dollar amounts charged for hospital services | Numerical | Person level | Comparatively low null value rate |
| **Total Cost** | Prefix: total | Overall claims/cost | Numerical | Person level | Overall claims have higher null value rate; others have comparatively low null value rate |
| **Other** | ID | Member's unique ID | Categorical | Person level | No missing value |
| | mabh_seg | MAPD behavioral segment | Categorical | Person level | Very high null value rate |
| | cons_mobplus | Mail order buyer | Categorical | Person level | Comparatively high null value rate |
| | hum_region | Geographic information | Categorical | Person level | Comparatively low null value rate |
| | lang_spoken_cd | Language spoken | Categorical | Person level | Very high null value rate |
| | cons_hhcomp | Household composition | Categorical | Person level | Comparatively high null value rate |
| | est_age | Member's age | Numerical | Person level | Comparatively low null value rate |
| | hedis_dia_hba1c_ge9 | Evidence for HBA1C test | Binary | Person level | Comparatively high null value rate |
| | met_obe_diag_pct | Claims related to obesity | Numerical | Person level | Comparatively low null value rate |
| | pdc_lip | Days covered for prescriptions | Numerical | Person level | Comparatively low null value rate |
| | phy_em_px_pct | Claims for physician evaluation | Numerical | Person level | Comparatively low null value rate |
| | src_div_id | Division ID assigned | Categorical | Person level | Comparatively low null value rate |
| | zip_cd | Member's zip code | Categorical | Person level | Low null value rate but contained undetected zip code records |
| | race | Race | Categorical | Person level | Comparatively low null value rate |
| | sex | Sex | Categorical | Person level | Comparatively low null value rate |

### 3.2.2 Understanding Members

- *Age Distribution:* We found that most of the Humana members were elder people. When we grouped the members by their COVID vaccination status, we found that most of those who were vaccinated were around 70 years old.



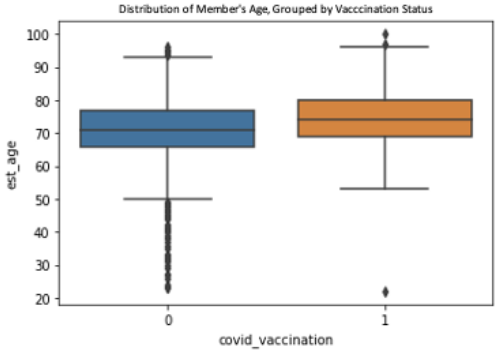



**Figure 1 Member Age Distribution Histogram and Boxplot**

- *Race:* Most of our members in the dataset were whites, who had the highest vaccination rate comparing to other races.

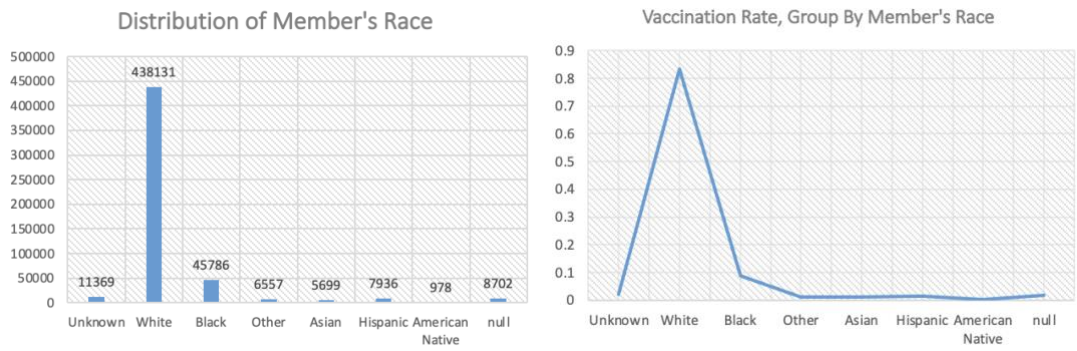



**Figure 2 Member Race Distribution Histogram and Boxplot**

## 3.3 Data Cleaning and Imputation

The original dataset was not readily available for feature engineering due to the problems of improper data types, missing data and presence of categorical variables that cannot be understood by analytical models. Therefore, before diving into the feature engineering and selection, we performed data cleaning and missing data imputation on the original data.

### 3.3.1 Data Type Transformation

The variables in the dataset fell into the following categories: numerical, categorical, binary, ordinal. After we loaded our dataset, we found that some variables' data types were not properly assigned. Thus, we performed several data type regulation steps on the original dataset in advance.

We first replaced all the null value by NaN, to make sure numerical variables were not misclassified as categorical variables. Then we converted all the variables into numerical values that carry the same physical meaning as their original ordinal or categorical data types.

### 3.3.2 Missing Data Imputation

We first checked the null value rate for all the variables and found that about 2 variables had null value rate higher than 60%, in which *spoken language* had the most missing values. 5 variables have null value rates ranging from 20% to 60%, and the remaining variables had null value rate less than 20%.

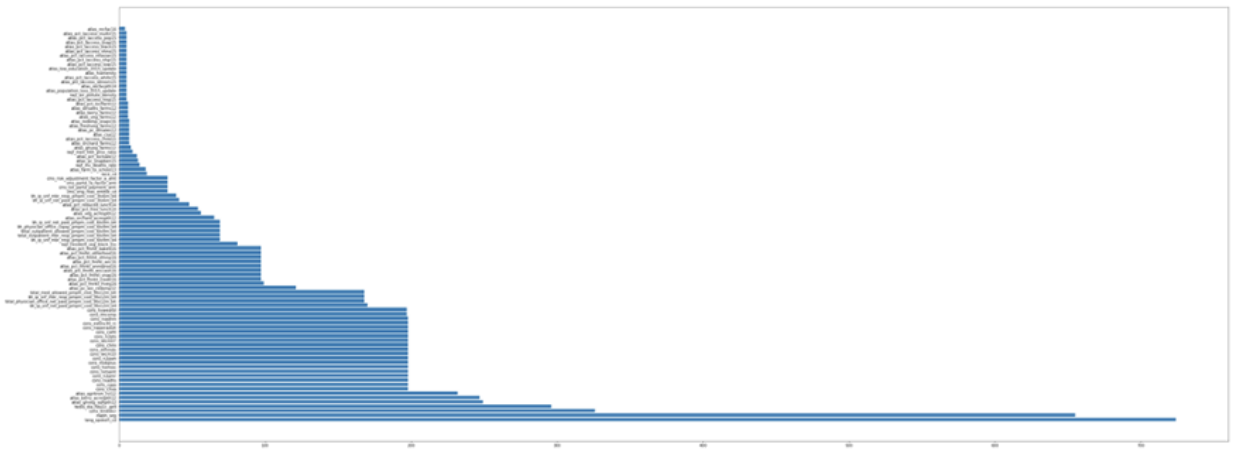


**Figure 3 Features with Highest Null Values**

As we can see from the bar chart (Figure 3), the missing values seemed to have synchronous effects when it came to the variables from the same source. Then we examined the correlation of missing values between variables and discovered that among those features collected from the same source, there existed high correlations among their missing data. The heatmap listed revealed two groups of variables with prefix of *cons* and *atlas*, and we can conclude that if one of the variables were missed for a member, then it's very likely that the other variables in the same group were also missed.

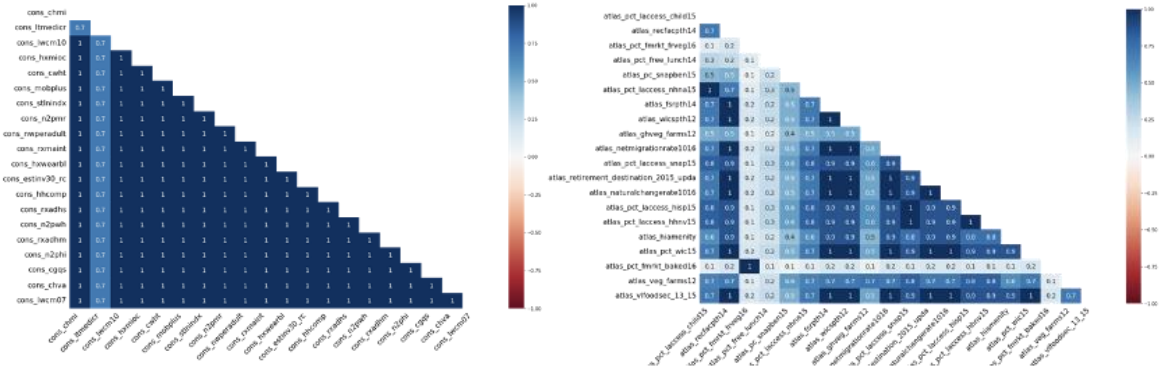


**Figure 4 Feature Null Value Correlation Heatmap**

For each type of variable, we applied different imputation methods according to their measurement level and meaning.

*Numerical variables:*

- *Number of claims, costs and authorizations related to specific disease:*
  Variables of this kind indicated the cost or claims resulting from specific diseases. Most of those variables had skewed distribution with lots of members having zero value. We assume that if the records were lost, it was probably because those members didn't have that disease or didn't have related costs or claims. So, we imputed zero to those null values.
- *Bank credit and medical plan costs:*
  These variables revealed personal credit situations, and we imputed median to those null values.
- *Regional economic and medical situation variables:*

The Atlas, RWJF and census data described the regional statistics for the region where the member was in. We imputed median value for the null values.

*Categorical/Ordinal variables:*

For all the other categorical and ordinal variables, we kept the null value as a new category to keep the original information.

## 3.4 Feature Encoding

In order to build machine learning models, we need to encode our categorical variables into numerical ones [3]. Based on our previous analysis on those categorical variables, we performed following encoding methods [4] on different variables.

- *Label encoding:*
  For binary flag variables, we encoded them into 0 and 1. This method was applied to variables such as *sex_cd*, *hedis_dia_hba1c_ge9* and our target variable *covid_vaccination*.
- *One-hot encoding:*
  For categorical variables without ordinal characteristics, we kept the null value as a new category then applied one-hot encoding method. This method was applied to variables such *as hum_region*, *mabh_seg* etc.
- *Ordinal encoding:*
  For categorical variables related to trend in terms of medical payments or insurance claims, we converted the categorical labels into ordinal data, whose value are assigned based on the order of effects towards COVID vaccination status. This method was applied to all trend-related variables such as *total_allowed_pmpm_cost_t_9-6-3m_b4*, *rx_overall_gpi_pmpm_ct_t_12-9-6m_b4*, etc.

## 3.5 Feature Engineering

### 3.5.1 Additional Data Source: Regional Vaccination Rate and Social Vulnerability Index

When we are exploring the reasons for vaccination rate and hesitancy, we found that there existed regional differences across the country (Figure 5). Meanwhile, the Social Vulnerability Index (SVI) was differently distributed among counties, which indicated the vulnerability of counties when it comes to hazardous events such as natural disaster and disease break. Poverty, lack of access to transportation, and crowded housing were indicators that had an impact on SVI.

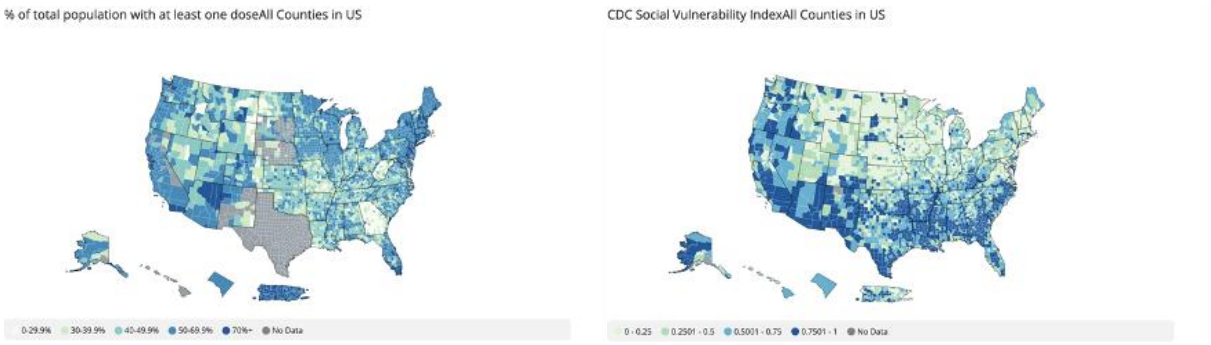


**Figure 5 Vaccination Rate and SVI by County** [5]

Thus, when it came to predicting vaccination hesitancy, we assumed that the members' location might influence their decisions. A region with higher SVI might give the member higher motivation towards vaccination.[6]

We researched for the regional factors that correlated with regional vaccination rates. Since we were given zip code feature in the dataset, we tried to add additional local features to our dataset, which would provide us with extra regional information for the members.

We collected Social Vulnerability Index, Vaccination Rate and Vaccination Hesitancy Survey data by county, and joined them into our dataset based on members' zip code. During the process, we found that about one third of members had invalid zip codes and couldn't be assigned with corresponding values. In this case, we imputed median value to those not matched records.

### 3.5.2 Additional Feature Engineering for *XGBoost*

Besides the feature engineering methods above, more features were created to improve performance for the XGBoost model.

- *Age_bin*: the "est_age" feature was split into bins with a width of 5 years to increase robustness, and each bin stood for a new binary variable.
- Less important but useful features were combined as fit to decrease noise:
    - *"Auth" features*: recent admit and discharge records due to various reasons were added together respectively, considering that each disease might contain too little information.
    - *"Atlas" features*: positive, negative, and farmer's market related features were added together respectively to get more generalized indicators.
    - *"CCSP" features*: claims of different categories were added together.

## 3.6 Feature Selection

After feature engineering, the columns in the dataset increased from 367 to 572. By directly using a vast number of features, the model would cost prolonged run time and be prone to overfitting. Therefore, we employed three different methods for initial feature selection.

Considering the size of the dataset, our first goal was to conservatively remove features with low or zero importance recognized by all methods. Since most of the models in our approach were tree-based, we chose Gini Index, XGBoost, and Random Forest to evaluate the feature importance to ensure reasonable and reliable outcome

We started with Gini Importance, which measures the gain of purity by splits of a given feature. It is the most common approach to measure the feature importance in terms of tree-based models. We selected 166 features with zero Gini importance and defined them as non-important features. Next, we trained a Random Forest (RF) model, which is a tree-based ensemble learning algorithm commonly used in classification tasks. Because the first method already evaluated the features

through Gini importance, which is also the default split criterion for RF, we changed the criterion parameter to "entropy" for different results and found another set of 158 columns with zero importance. Finally, we used XGBoost Classifier with default parameters and got a third set of 234 features with zero importance.

Among the three sets of features, we took the intersection, which sums to 155 features to be removed in total. Further work of feature selection was model-specific and conducted later during model tuning.

# 4. Modeling

## 4.1 Approach

Based on the objective of identifying COVID-19 Vaccine hesitant members, we formulated it as a binary classification problem and utilized a two-stage process to select the best performing model.

Our first step was to randomly split the cleaned "Training Data" into three different sets: *training*, *validation*, and *testing* with the percentages of 70%, 15%, and 15%, respectively.

During the first stage, we selected six different classification models to perform cross-validation with the 70% training data and tested their performance on the 15% validation set. The evaluation metric used was the AUC-score, which measures the area under the Receiver Operating Curve (ROC) and generally reflects how well the model can distinguish the classes. To improve performance, we applied appropriate techniques including parameter tuning, additional feature selection and engineering. The resulting AUC scores summarized in Table 1 suggested three top models with close performances: Gradient Boosting Decision Tree (GBDT), LightGBM, and XGBoost, and these models entered the next stage.

**Table 2 Model Performance on Validation and Test Data**

| **Model** | **Validation AUC Score** | **Test AUC Score** |
|---|---|---|
| Logistic Regression | 0.6706 | -- |
| Random Forest | 0.6443 | -- |
| Neural Networks | 0.6053 | -- |
| **Gradient Boosting Decision Tree** | **0.6814** | 0.6819 |
| **LightGBM** | **0.6828** | 0.6832 |
| **XGBoost** | **0.6825** | **0.6839** |

With the 70% training data and 15% validation set combined, we trained the top three models again with the best parameters determined in the last stage and evaluated their AUC scores on the final 15% testing data. Our approach not only ensured a fair and robust comparison among algorithms by testing on a completely new set of data, but also allowed the most learning opportunities for the algorithms through re-training.

Out of the three models, XGBoost achieved the highest AUC score of 0.6839 and we used all available training data (70%+15%+15%) to train a final XGBoost model to predict the Holdout data for best performance. The details regarding the XGBoost model are explained in the following section.

## 4.2 Final Model

XGBoost, which stands for "eXtreme Gradient Boosting", is a decision tree ensemble learning algorithm with an optimized boosting system [7][8]. Gradient Boosting minimizes loss while adding new models through the gradient descent algorithm to increase the predictive power, and XGBoost builds on this underlying method with systematic and algorithmic performance improvements [9].

With parameters set based on previous experience, we first focused on additional feature selection and engineering. For the first several rounds of training, we removed features with zero importance and examined the less important features for combination to reduce dimensionality. This process was repeated until all non-important features were removed. The exact features generated are written in detail under the feature engineering section. Next, we leveraged the advantages of XGBoost by tuning 7 selected parameters with the "Randomized Search" method. Due to computational capacity limitations, each round of search was limited to 100 estimators, and 5 sets of parameters with similar outcomes were then fully trained for final comparison.

Our final XGBoost Model parameters are:

- Max_depth = 6. This parameter controls the depth of the tree, setting this at a relatively low value saves computational cost since the model prunes backwards and prevents overfitting [8][9].
- Learning_rate = 0.05. This parameter shrinks the feature weights after each boosting step and reduces overfitting. Although a small learning rate will increase train time, the parallelized tree building design of XGBoost has greatly improved run time [8][9].
- Gamma = 0. This parameter sets the minimum decrease in loss for making the next split. The randomized search found the default value to be the best choice [8].
- Subsample = 0.9. This parameter determines the proportion of training data used for each tree [8].
- Colsample_bytree = 0.6. This parameter determines the proportion of features used for each tree, and the smaller helps prevent overfitting [8].
- Scale_pos_weight = 0.21. This parameter deals with an imbalanced dataset, which is true for our case. We calculated the proportion of negative cases (labeled 0, represents those vaccinated) relative to the positive cases (labeled 1, represents those unvaccinated) as recommended [8].
- Min_child_weight = 20. This parameter sets the minimum sum of instance weight in a leaf node before taking the next split. Having this large value makes our model more conservative [8].
- Objective='binary:logistic'. This parameter specifies the type of task for our model is binary classification and the outcome should be probability [8].

- n_estimators=500. This parameter controls the total number of boosting rounds, or the number of gradient boosted trees. Since other parameters were selected to be more conservative, we chose 500 estimators to maximize the learning opportunity for the model [8].
- Alpha = 0.3. This parameter controls the L1 (LASSO) regularization, which is one of the distinguishing attributes of XGBoost to help with overfitting [8].
- Lambda = 1. This parameter controls the L2 (Ridge) regularization, which is another common technique for controlling overfitting [8].

# 5. Modeling analysis

## 5.1 Model Performance

As shown in the figure below, the AUC score of the final XGBoost Classifier is 0.6839 and outperforms the rest. For further evaluation, we examined other metrics as well.

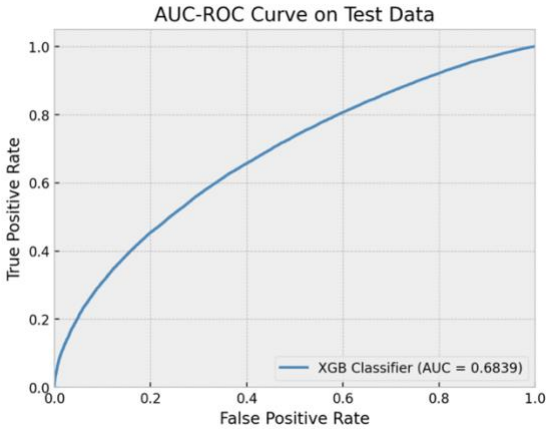



**Figure 6 AUC-ROC Curve on Test Data**

Looking at both the confusion matrix and the classification report, we can observe that our model concentrates on predicting the "Hesitant" members and has a 90% precision rate for this target class. Precision measures the percentage of correct predictions among the "Hesitant" members identified by the model. This indicates that our model has achieved the objective of finding the members most likely to be hesitant. However, our model separates the vaccinated members less accurately. The recall for the "Vaccinated" class is 67%, measuring how many correct predictions we could make among all vaccinated members. But the recall for "Hesitant" class is only 59%, which suggests that a portion of the hesitant members were recognized as vaccinated. At the current default probability threshold, the overall accuracy is 61% and we believe is strong performance to serve Humana's purpose.

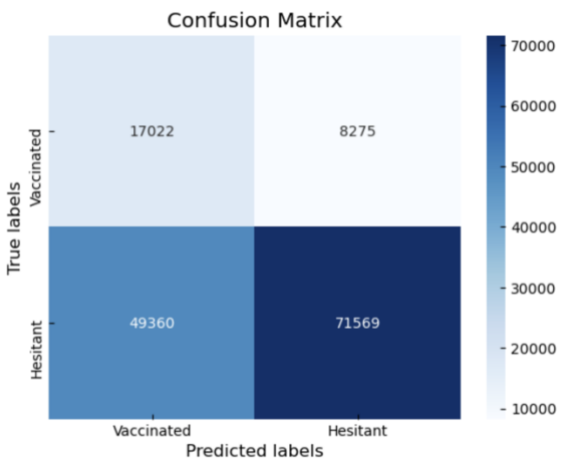



**Figure 7 Confusion Matrix Heatmap**

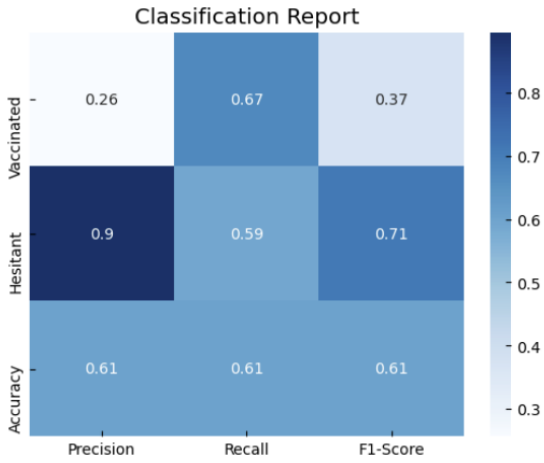



**Figure 8 Classification Report (Partial) Heatmap**

In Figure 9, we can observe the distribution of our predicted probability, where the coloring represents the actual class label, and the yellow line represents the baseline 0.5 probability threshold. The graph aligns with previous findings that the target class prediction is accurate, but the other class is not well distinguished. Therefore, we have considered this conclusion for our recommendations, and we will evaluate the possibility of adjusting the threshold for certain strategies based on cost-benefit analysis.

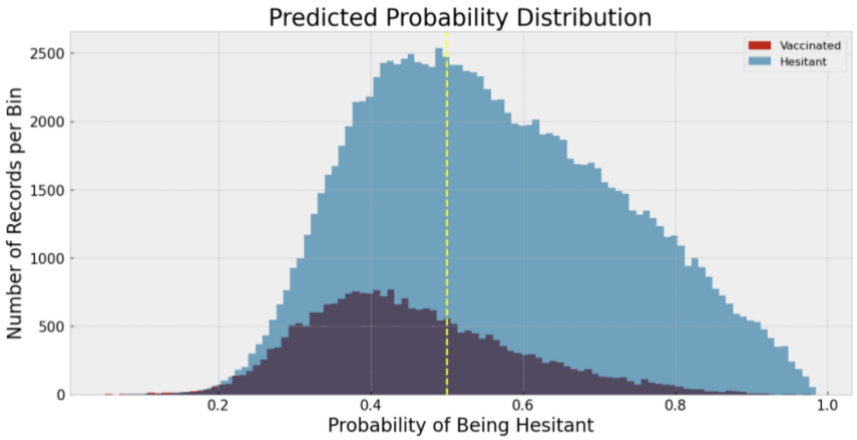



**Figure 9 Predicted Probability Distribution**

## 5.2 Feature importance/Key Performance Indicator Analysis

In order to extract more insights from our predictive model, we looked at the top 20 most important features from both the XGBoost Classifier and SHAP Value.

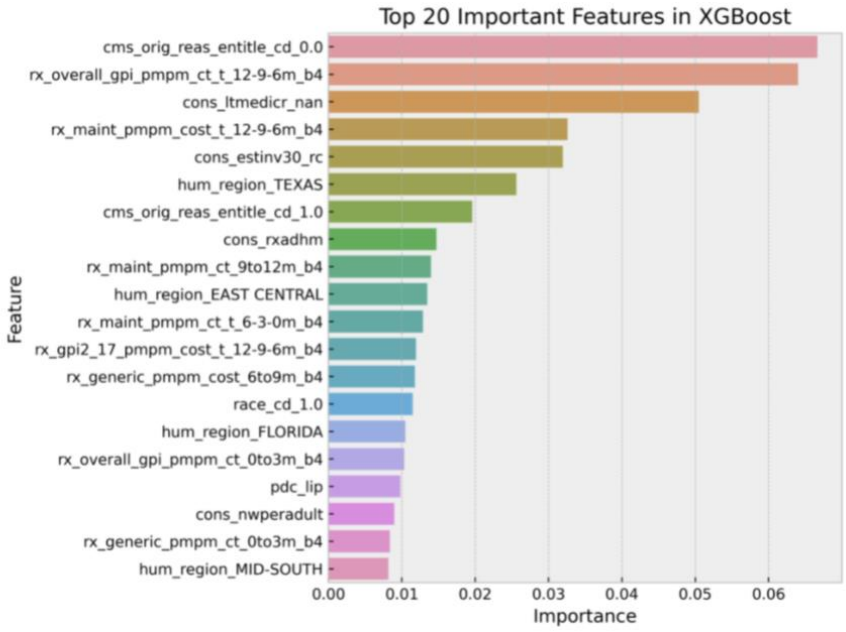



**Figure 10 Top 20 Important Features from XGBoost**

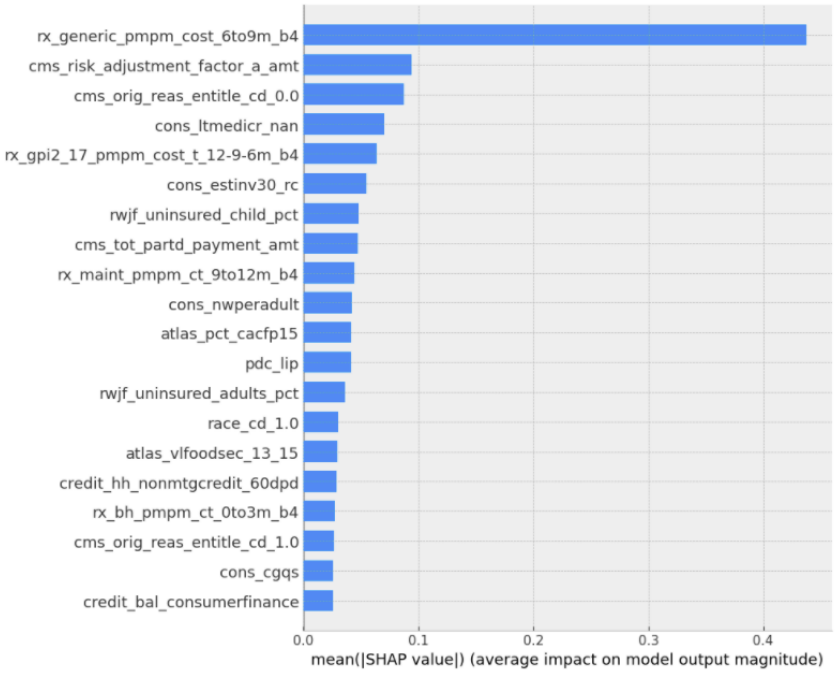



**Figure 11 Top 20 Important Features from SHAP Values**

These features can be categorized as the following:

- "rx" features: different statistics about the member's prescription for various types of diseases. Using selected features in this category, we found that members with no recent prescription (count or spending) are more hesitant to the vaccine.
- Regional features: measures describing characteristics of the member's resident area.

    - Physical locations: Texas, East Central, Florida, and Mid-South are geographical areas with low vaccination rates
    - Socioeconomic: “rwjf” and “atlas” prefixed features reflect an area’s general level of healthcare and income.
- “cms” features: originated from member’s Medicare Advantage information, can suggest reasons related to medical or physical conditions
- Finance-related features: measure the wealth of a member. Low-income or poor members are more hesitant.

Additionally, we used SHAP dependency plots to study the individual effects and interaction effects of key variables.

Cons_estinv_30 is the “Estimated Household Investable Assets Recoded” and the dependency plot shows that lower investable assets correlate with an increase in hesitancy towards the vaccine. The binary variable measuring whether the member entered Medicare based on the age criteria interacted the most. The plot indicates that among the less wealthy members, those who entered Medicare due to other reasons (possibly disability and ESRD) are more hesitant than the elderly. This suggests that lack of access to the vaccine could be the reason behind the low vaccination rate.

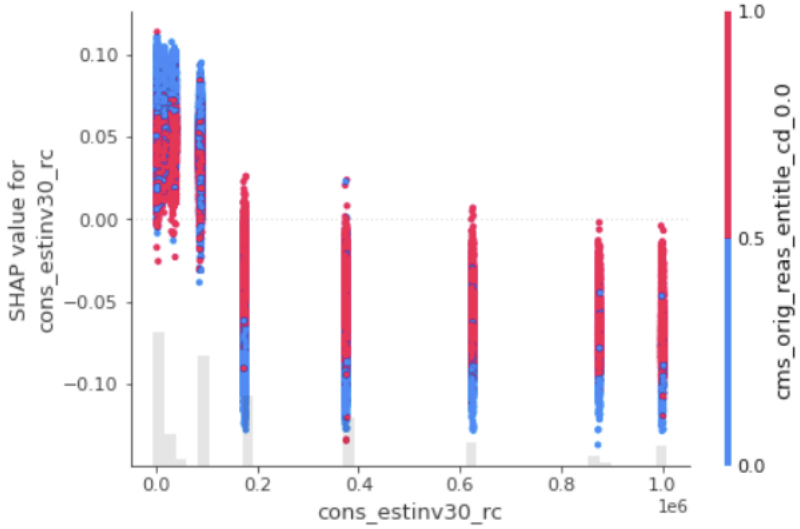



**Figure 12 Cons_estinv_30 Dependency Plot**

Race_cd_1.0 is the binary variable representing the white people and the dependency plot shows that the minority races (non-white) are more prone to be hesitant. The binary variable for Florida interacted with race the most and reveals that the minority races in Florida are especially hesitant, which suggests us to ensure part of the recommendation can reach this group.

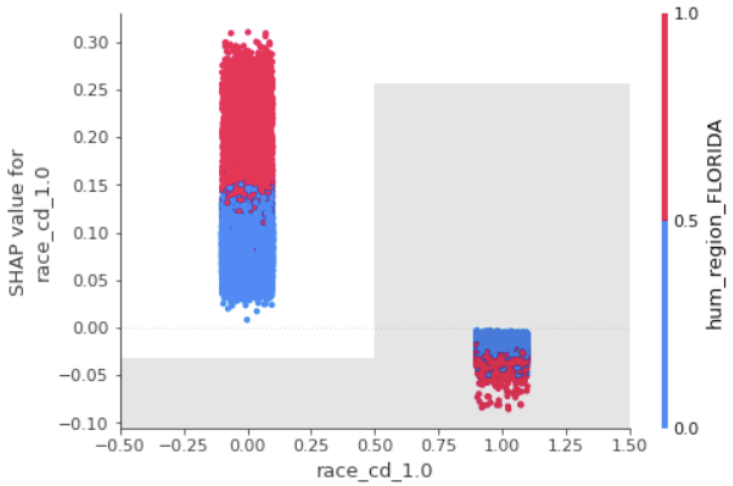



**Figure 13 Race_cd_1.0 Dependency Plot**

# 6. Segmentation

## 6.1 Approach

Based on background research and insights generated from our predictive model, it has become quite clear that the hesitant members are not homogenous. Thus, to successfully design appropriate recommendation for promoting vaccination, we need to dig deeper into the underlying reasons behind their hesitancy and create segments based on different reasons.

The research described earlier in case background suggests that race and region of residence could be relevant factors, as non-Asians and people living a county with higher Trump-supporting rate are more hesitant to vaccination. This finding is consistent with the important features identified by our predictive model, and we selected the following 4 top ranked features for further analysis:

- Cons_estinv30_rc: estimated household investable assets
- hum_region: represents the member's residency region. Texas, Florida, Mid-south and East Central have significantly lower vaccination rates
- cms_orig_reas_entitle_cd: the member's reason for entering Medicare
- race_cd: member's race

Our segmentation approach is to rely on these features and find the appropriate cutoff values to form segments with lower vaccination rates. To better reflect a member's financial status, we added *credit_hh_bankcard_severederog* and *credit_bal_consumerfinance* as additional features and compared the vaccination rates with various cutoff values. The other categorical variables are also examined. Finally, we found the most effective filtering conditions summarized in the table below.

**Table 3 Segment Filtering Conditions**

| Segment | Filtering Conditions |
|---|---|

| | |
|---|---|
| People in Financial Distress | credit_hh_bankcard_severederog>15 or credit_bal_consumerfinance>1500 or cons_estinv30_rc<3000 |
| Racial Minority | race_cd ≠ 1 (Non-white) |
| People in Selected Regions | hum_region=TEXAS or EAST CENTRAL or FLORIDA or MID-SOUTH |
| People with Disabilities | cms_orig_reas_entitle_cd =1 (Disabled) |

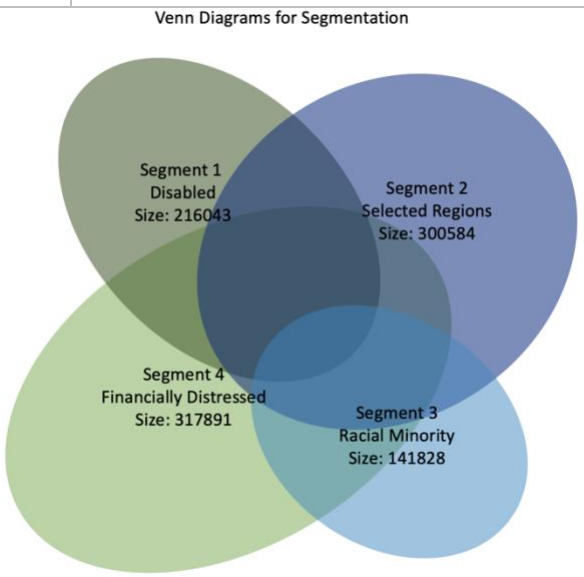



**Figure 14 Segmentation Result Venn Diagram**

The figure above is the Venn diagram describing the result for each segmentation. There is some overlap among different segments but clearly separated overall. The center intersection of all four segments has size of 12,312, equivalent to 1.5% of the whole hesitant population. We also calculated the overlap between any of the two segments and concluded that people with overlap account for no more than 30%. The overlap account for 12.5-22% on average for each of the segments. Therefore, our segmentalized outreach will cover as many people from different segments as possible, while repeated effort on overlapping among segments is relatively small.

We then compared the union set of our four segments with the total number of unvaccinated members and concluded that 73.18% of the unvaccinated members have been covered using our segmentation methods.

## 6.2 Segment 1: People in financial distress

For this segment, we defined a person in financial distress to have more than 15 severe derogatory accounts, have poor credit histories (over $1500 in balance), or whose estimated household investable assets are less than $3000. It is worth noting that 38% of Humana MAPD members are in financial distress to some extent in figure 16. People in financial distress have a lower vaccination rate than the overall population (14.2% vs 17.4%). The reason they are in financial distress may vary. Some of them might have low income or be unemployed, others might be the ones with debt past due and debt in collections. It is likely that those who are in bad financial status are people with less

education. Results from a recent study[10] show that adults with higher education are significantly more likely to get a vaccination and to believe in the vaccine's safety and effectiveness. Or they simply do not want to spend extra money on their way to the clinics, pharmacies, and other locations that offer COVID-19 vaccines. The common characteristic of this segment, however, is that they are extremely sensitive to price. They might be open to getting vaccinated but needs more financial incentive for vaccination.

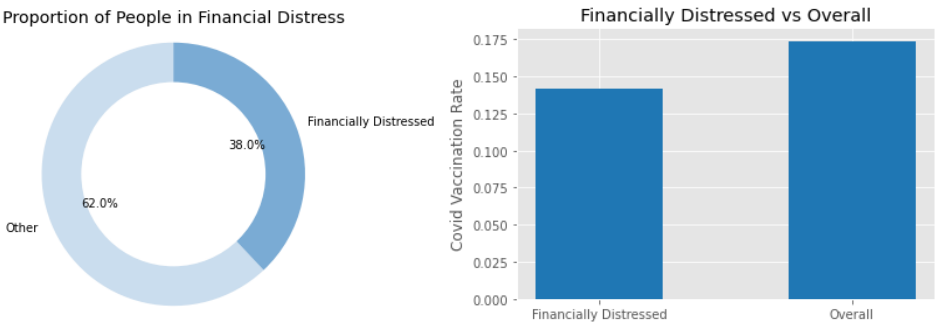



**Figure 15 Proportion of people in financial distress and their vaccination rate compared with the overall population**

## 6.3 Segment 2: Racial minority

Vaccine hesitancy among ethnic minority groups needs to be highlighted. We separated the non-white population as a segment from the total population. According to our dataset, 90% of the non-them are Black and Hispanic, with Black people accounting for 68.8% of the total non-white population in figure 17. A survey conducted by Kricorian and Turner[11] shows that Black and Hispanic individuals were less willing than Whites to receive the vaccine. Moreover, mistrust of the vaccine among Black respondents was significantly higher than other racial groups. In order to promote vaccination equity among racial and ethnic minority groups, we need to address their concerns and provide solutions targeting specifically for this segment.

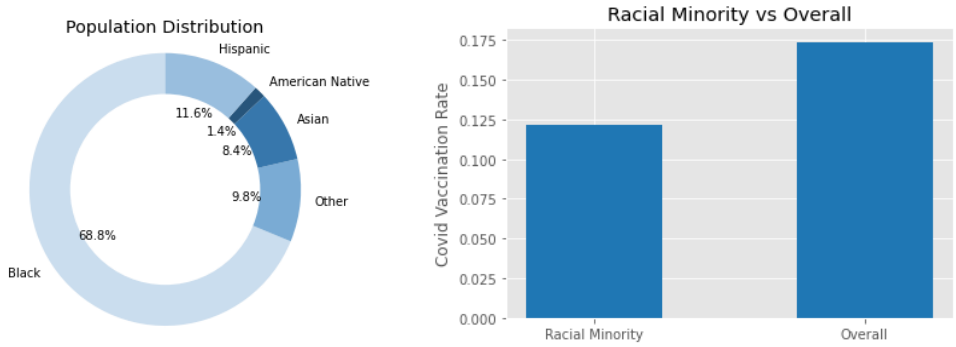



**Figure 16 Race distribution and the racial minority vaccination rate compared with the overall population**

## 6.4 Segment 3: People from selected regions

Vaccination rates also vary widely by county across the United States (figure 18). Especially, the average vaccination rate in counties that voted for Trump in the 2020 election is 28.5% compared to 35.0% in counties that voted for Biden [12].

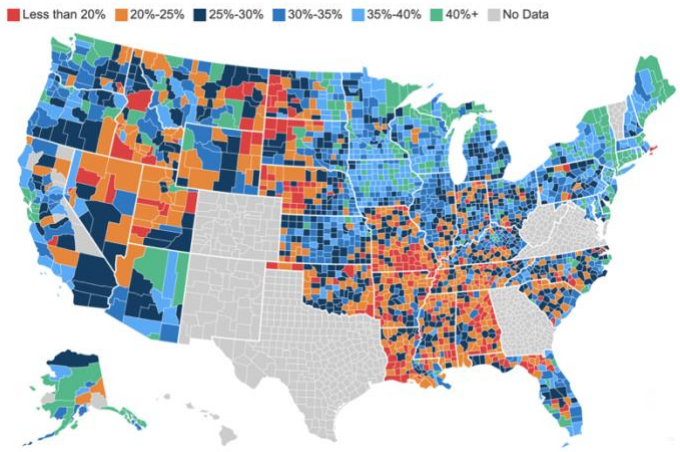



**Figure 17 Vaccination Rates by County [12]**

We selected 4 regions with low vaccination rates including East Central, Mid-south, Texas and Florida and segmented the residents from these regions, which account for 36% of the total population shown in figure 19. Because an individual's behavior is often influenced by people around them, grouping potential members by region helps us target the unvaccinated members more easily.

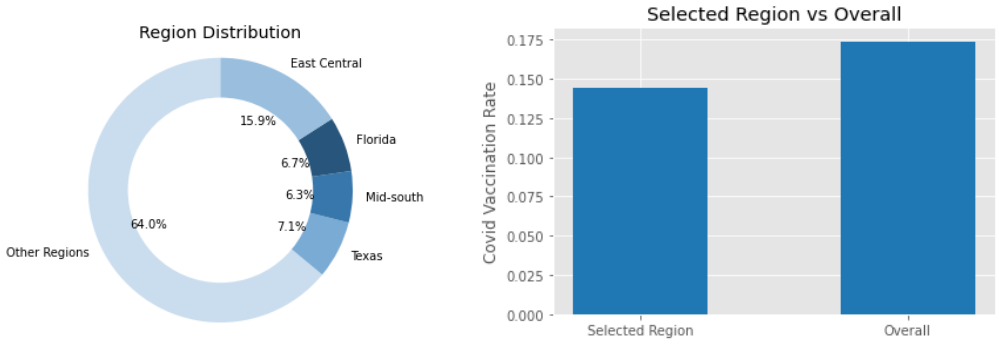



**Figure 18 Region distribution and the selected region vaccination rate compared with the overall population**

## 6.5 Segment 4: People with Disabilities

Among Humana MAPD members, 25.1% have a certain type of disability (figure 20). Though vaccines are now available nationally for free, many people with disabilities have difficulty accessing them. Study shows that COVID-19 vaccination coverage was lower among adults with a disability than among those without a disability, even though adults with a disability reported less hesitancy to getting vaccinated [13]. Therefore, it is essential to ensure equal access to vaccines and make vaccination process accommodate the needs of people with disabilities.

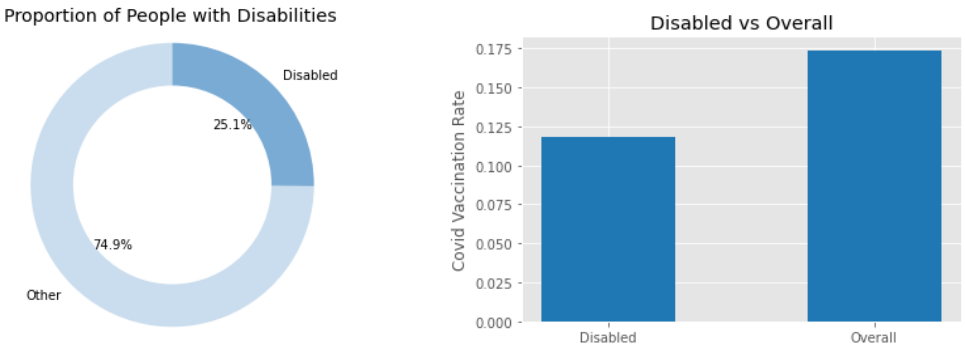



**Figure 19 Proportion of people with disabilities and the people with disabilities vaccination rate compared with the overall population**

The remaining population excluded from the four segments accounts for 26.82% of the total population. Our preliminary assumption is that those are the people who do not have specific reasons for not getting vaccinated, or the reasons are too complicated to explain. We have already ruled out the possibility of lacking access to vaccination. Further, they are either from regions with low vaccination rate or from minority groups. Therefore, there might be some personal reasons such as hesitant family members, religious-based objection to vaccine, fearing for side-effects or they simply do not take COVID-19 as a serious disease.

## 7. Recommendation

Based on the segmentation analysis, we have proposed a CoVAE Program to enhance Covid Vaccine Accessibility and Equity.

For the sake of feasibility, we have determined the priority among segments based on the risks of COVID-19 infections and ease of execution.

Our priority is to encourage people in critical financial distress to get vaccinated. This is because financial distress might indicate low living standards and medical resources inaccessibility, which results in high risks of COVID-19 infections. Using financial incentives is considered as the easiest and most direct measure to boost vaccination rate in this group.

*Second*, the other three identified segments, known as racial minority, people with disabilities and from selected regions, are vulnerable because their close-knit social circles tend to have lower vaccination motivations. But at the same time, the reasons behind the low vaccination rate of each segment are quite different. We designed separate strategies to overcome physical and mental barriers to improve vaccination rate.

*Third*, the remaining segment is least prioritized because the difficulty of motivating the segment is the highest. Since this segment is mostly comprised of white people without financial distress and minority, they are less likely to be infected.

**Table 4 Priority for the Segment**

| | **High Risk of Infection** | **Low Risk of Infection** |
|---|---|---|

| | | |
|---|---|---|
| **Easy to motivate** | 1st Priority: People in Financial Distress | |
| **Difficult to motivate** | 2nd Priority: Racial Minority, People with Disabilities, People from Selected Region | 3rd Priority: remaining Segment |

For each strategy below, we will start by elucidating the rationale behind the strategy and how it addresses the program. Then, we will describe the detailed plan of implementation. It will be followed by the analysis of effectiveness, i.e., conversion rate into vaccination, and costs estimation.

**CoVAE Program Overview:**

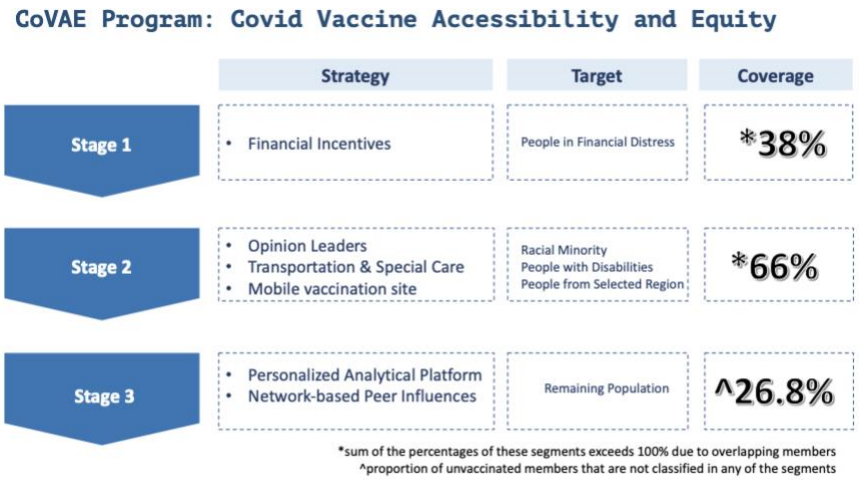



**Figure 20 Timeline for the 3-Staged CoVAE Program**

## 7.1 Stage 1: Financial Incentives

### 7.1.1 People in Financial Distress: Nudging with Financial Incentives

Strategy Rationale

Humana should provide monetary rewards to incentivize people to receive COVID-19 vaccination. This solution is population-wide and available to all Huaman members, but it is expected to primarily affect the people in financial distress, who should be price-sensitive. Among all the underserved groups identified in the segmentation analysis, the members in financial distress are likely to be debt-ridden. They might be hesitant to receive the vaccination for various reasons, but financial incentives could overcome their hesitancy to some extent and encourage them to receive the vaccination. By providing a monetary incentive, a portion of them will be induced to administer the vaccination.

Strategy Implementation Plan

The design of the rewards schemes could be broken into 3 points. *First*, a clear validity period for the rewards has been added to all monetary rewards to limit costs and maximize the immediate

urge for the rewards. *Second*, on top of the expiry period of the offering, we could limit the quota under each reward scheme to control the costs. *Third*, rewards can be provided in various forms and scales to nudge people into vaccinations. Here is a suggestion of a list of reward schemes:

1) Modest Rewards linked to Humana Insurance:

    A one-off discount for an annual insurance premium of $35 could be provided, within a promotion period of one to two months. The renewal offering can continually tie the members to Humana insurance service.

2) Modest Immediate Rewards:

    A monetary refund of around 25 can be offered in the form of gift card to each Humana member if they are fully vaccinated, within a promotion period of one to two months. It is the most direct monetary inducement to vaccination.

3) Modest Lucky Draw on Daily Scale

    For vaccination sites located in a shopping mall or food courts, Humana members who have administered vaccination on the same day are eligible to enter a lucky draw to win a gift card of around $80 to $100. Given the short span of lucky draws, price-sensitive members are more likely to be driven into action immediately. Humana should locate the states where most Humana members are concentrated and select the busiest 500 shopping malls *next to vaccination sites* as test run places for the daily lucky draw.

4) Sumptuous Lucky Draw

    In contrast to the daily modest lucky draw, members can be invited to a lucky draw for an attractive prize, such as a limousine, if they receive vaccination within two months. The sumptuous lucky draw is expected to attract members who favor rattle.

<u>Expected Effectiveness</u>

Nudging with financial incentive into vaccination is a scientifically proven method. This idea of financial incentives has been commonly practiced by governments and private sectors[14], but Humana could still implement the aforementioned reward schemes specifically targeting Humana members to raise the member's vaccination rate. According to a study in Sweden conducted by *Science*, the modest monetary incentive of 24 US dollars can push up the vaccination rate by 4.2% from the 71.6% baseline rate as compared to the control group[15]. It means that 0.042/(1-0.716) = 14.789% of unvaccinated are moved by the monetary rewards. One important observation from the above study is that monetary incentives increased vaccination *irrespective of people's background*.

The Swedish study has referential value in the United States because Sweden and the US have similar levels of human development (HDI: 0.945 for Sweden vs 0.926 for the US)[16]. From the Training Data, there are 82.62% of unvaccinated Humana MAPD members. Assuming the monetary nudges have a similar conversion rate in the overall Humana population as the study, it implies that the vaccination uptake can be 0.83*0.15 = **12.22%** for the Humana population. Given that Humana population size of 4 million (from 2019 data), it is estimated that 4* 12.22% =0.48 million members will be vaccinated due to the monetary rewards. Below shows the regression estimated impact for financial incentives and other nudges, i.e. the quantified influence of nudges on the vaccination uptake.

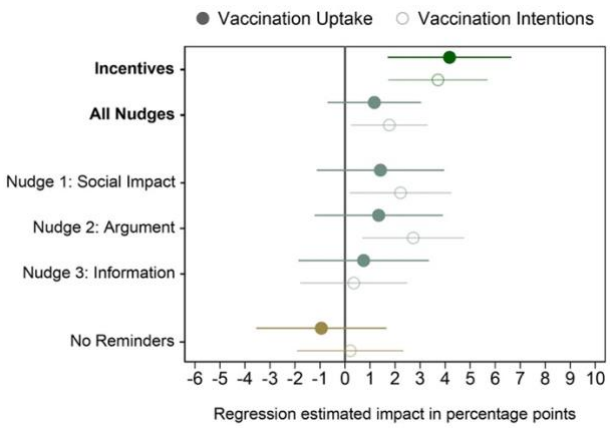



**Figure 21 Regression-estimated effects of experimental conditions on vaccination uptake and vaccination intentions against the control condition[17]**

Cost Analysis

The costs of the financial incentives can be easily controlled and estimated by imposing expiry periods and quotas. The amount of reward takes reference with $24 rewards used in the Swedish study, which is used as a baseline per person reward in the above rewards schemes, as in scheme B. The reward in scheme A has priced in the additional benefits of member retention on top of baseline rewards. The lucky draw prize in Scheme C is nudging a group of members thus choosing $100, 4 times the baseline rewards as value of rewards. To limit the costs, a quota of 50,000 is set for individual reward scheme A and B each, which sum up to account for 20% of potential converted members, the luck draw scheme C and D will be expected to account for the remaining 80% of potential converted members.

The variable costs of scheme A and B can be estimated by the following formula:

$$Cost\ for\ A\ and\ B = \ Quota\ of\ Rewarded\ Members * Reward\ Amount.$$

The variable cost for scheme A is $50{,}000 * 35 = \$1{,}750{,}000$, which will be in the form of *reduced earnings due to a discount in premium.* The potential benefits from the effects of member retention have not been considered. The variable cost for scheme (b) is $50{,}000 * 25 = \$1{,}250{,}000$. For scheme C and D, their costs are estimated based on the amounts of prizes of lucky draws. The variable cost for scheme C is estimated by:

$$Cost\ for\ C = \ Number\ of\ Daily\ Lucky\ draw * Number\ of\ Days * Prize\ Value$$

Thus, the estimated costs for Scheme C are $500 * 24 * 90 = \$1{,}080{,}000$. The cost for scheme D is estimated based on the reference price of a Tesla limousine Model S to be $129,990.

The total costs will account for the sum of the variable costs and fixed costs such as the administrative costs in organizing the rewards schemes and marketing fees. While the costs of the rewards schemes may be seemingly significant, the potential cost-saving for Humana could be up to $rate\ of\ hopstilization * Humana\ population * newly\ vaccinateed\ popluation *$ $per\ person\ hopstialization\ cost = \frac{14.7}{100{,}000} * 4{,}000{,}000 * \$73{,}000 * 12.22\% = \$5{,}245{,}313,$ (assuming the vaccinated members will unlikely be infected with COVID-19, based on the rate of

hospitalization for people aged 65 or above as of Oct 2, 2021[18]). The potential benefits of monetary nudge still outweigh its costs by $\$5{,}245{,}313 - \$4{,}209{,}990 = \$1{,}035{,}323$.

## 7.2 Stage 2: Targeted Groups

### 7.2.1 Racial Minority: Reducing Language Barriers by Cooperating with Opinion Leaders

Strategy Rationale

For people belonging to a minority race, language barrier can be a real obstacle for them to receive vaccine-related information. When it came to the detailed medical explanations on the vaccine and its possible side effects, it would be difficult for non-English speakers to fully understand English information and build trust to vaccination, thus causing hesitancy towards vaccination. These barriers might result in conformity effects and had negative peer impact on those who didn't get vaccination.

Strategy Implementation Plan

Our recommended approach is to work with appropriate opinion leaders among the underserved races to promote equal informational accessibility [19].

When finding opinion leaders, they should have the following characteristics:

1) Equipped with basic medical knowledge that can help eliminate peers' concerns.

2) Having close relationships with communities.

To find appropriate opinion leaders, we propose two possible sources:

1) Minority Serving Institutions Program [20], which has institutions aimed at minority people education, had wide network and close relationship with minority communities [21].

2) Minority influencers [22] on social media platform.

Through building connections with people from these two sources and providing them with related materials, they would become opinion leaders in their communities to help others.

There are several possible ways of influencing their peers:

1) Sharing their stories and experiences of getting vaccination (face to face/social media platform).

2) Discussing reasons to get vaccination and possible side effects.

3) Sharing vaccination resources accessible to community members.

Cost & Effectiveness Analysis

The cost mainly comes from the process of cooperating with opinion leaders. We can calculate the total cost with this formula:

$$Nol_1 = Number\ of\ opinion\ leaders\ on\ social\ media\ platforms$$

$$Nol_2 = \ Number\ of\ opinion\ leaders\ in\ communities$$

$$NC_1 = \ Cost\ per\ leader\ on\ social\ media\ platforms$$

$$NC_2 = \ Cost\ per\ leader\ in\ communities$$

$$Total\ Cost = Nol_1 * NC_1 + Nol_2 * NC_2$$

According to our segmentation, 15% of the unvaccinated members might need opinion leaders to reach out. As most of them are elder people, we estimate that 70% of them need to be reached out by community members, and the rest can be reached via social media platforms. Thus, we get 4.4m*83%*15%*70% = 380k people requiring community opinion leader to connect, and 4.4m*83%*15%*30% = 160k people requiring influencers to reach out. Assuming a community leader can reach 100 members in a month, then we will need 3800 community opinion leaders. For an influencer, we assume he can reach 1000 members in a month, then we would need 160 influencers.

The cost per opinion leader can be estimated based on our research, the cost for influencers could be $500 per month, while the cost for opinion leaders could be $1000 per month considering their extra efforts of meeting community members in-person. So, the total cost would be 3800*1000+160*500 = 3.88 million.

The effectiveness of this strategy could be evaluated by its conversion rate. Among the 540k members, we estimate the percentage of reached members as 70%. We assume the conversion rate as 8% when they are reached by a community member, 4% when they are reached by an influencer. Thus, we can get the estimated number of members that can be convinced to get vaccine: 380k*70%*8%+160k*70%*4% = 25.8k members.

### 7.2.2 People with Disabilities: Vaccination Special Care and Transportation Assistance

Strategy Rationale:

For people with disabilities, they will be faced with more difficulties on the way to vaccination. For example[23], people with mobility issues need careful transportation assistance, while people with hearing loss need special communication.

Strategy Implementation Plan

To promote equitable vaccination access, we suggest providing special care and assistance to people with disabilities. Special care includes professional help for people with communication or understanding disabilities, door-to-door services, suitable transportation methods, etc.

After we detected our members in need, we could reach out to them and provide them with more convenient solutions to get vaccination. We could collect their responses through a survey and arrange appointments, special care staff and transportation solutions.

Our research revealed that, there are plenty of local and national institutions having been providing transportation solutions for people who need to go to vaccination sites. Cooperating with such

institutions, such as Lyft, 211, and providing transportation solutions for Humana's members in need would be an efficient way.

As for the special care service, Humana could build connections with local medical providers and seek corresponding help according to the needs of members.

Cost & Effectiveness Analysis

For this strategy, we mainly have two parts of cost: one is from transportation, the other is from professional help. The calculation would be:

$$p_s = \%\ of\ Members\ who\ respond\ and\ need\ special\ care$$

$$p_t = \%\ of\ Members\ who\ respond\ and\ need\ transportation$$

$$C_s = Cost\ per\ member\ who\ need\ special\ care$$

$$C_t = Cost\ per\ member\ who\ need\ transportation$$

$$Total\ Cost = (N * p_s * C_s) + (N * p_t * C_t)$$

Based on our segmentation, 20% of the unvaccinated members might need extra help. The US government census[24] showed us types of disability and the distribution:

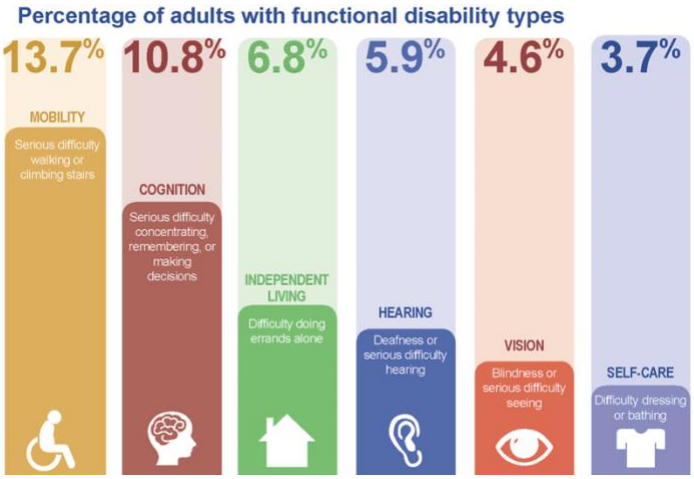



**Figure 22 Percentage of Disability Types**

Thus, we get 4.4m*83%*20%*13.7% = 100k people requiring transportation assistance, and 4.4m*83%*20%*(10.8%+5.9%+4.6%) = 150k people need special care.

We assume that the percentage of responding to Humana's survey would be 30%, so we finally get 30k people for transportation assistance and 45k people for special care.

According to research[25] conducted in North Carolina, we estimate the average distance of these members to the nearest hospital, community health center, rural health clinic, would be 5 miles, which will cost about 36 dollars for round trip. For special care service, we estimate each member would need about 2 hours' help from a medical specialist, which costs about 50 dollars.

So, the total cost would be 30k*36+45k*50= 3.33 million.

### 7.2.3 People from Selected Region: Resolving concerns with medical consultation and providing accessible vaccination site

Strategy rationale

According to the segmentation analysis, Humana should provide special care, such as provision of mobile vaccination sites, home visits and free or discounted pre-vaccination screening, to the members living in the designated 4 regions, known as Texas, East Central, Florida and Mid-South, which are particularly hesitant to receive vaccinations. The possible reasons for the hesitancy for the residents in these regions are health concerns over vaccine's side effects and development seed and inconvenience due to distance from vaccination sites, which are the opinions collected from the residents in Texas [26] and Florida [27]. By providing pre-vaccination, medical home visits and mobile vaccination sites, it is expected to alleviate their health concerns and overcome the spatial barrier, thus instilling the idea that the benefit of vaccine outweigh the transportation and health costs and encouraging the members in these selected areas to receive vaccination.

Strategy Implementation Plan

1) Pre-vaccination Screening

Since nearly half of the hesitant people possibly have concerns about side effects[28], Humana can offer free pre-vaccination screening the members in these selected regions, such that they can understand better their body conditions and alleviate their concerns over potential side effects.

2) Home Visits

Home visits accompanied by certified nurses and medical practitioners can also provide an informal medical consultation for the Humana members living in these regions and clear their doubts about the vaccines. The above measures are on a reservation basis that requires a member's active opt-in, to avoid resource wastage.

3) Mobile Vaccination Sites

For those members who are hesitant to take vaccination due to inconvenience and remote location of vaccination centers, Humana can locate the neighborhoods and counties with the highest density of Humana members and set up mobile vaccination clinics using trailers.

Effectiveness Analysis

As for the estimate of conversion rate, a conversative estimate for the conversion rate of the above policy is estimated to be 12%, on par with monetary incentive for several reasons. Although, the above services are provided *before* the member receive vaccination, with less guarantee for conversion, the measures are directly addressing the core reason of hesitancy, i.e., health concerns, as compared to indirect monetary nudge. Some members only hesitate, but not completely deny, to take vaccination. If they are willing to opt-in for the free screening or home-visit, it indicate their willingness to understand vaccine-related information and more likely to receive vaccination once their concerns can be properly answered. Therefore, a safe estimate of conversion is 12% for the

regional outreaches. According to Training Data, there are 36% of entire Humana population living in these 4 regions ($4{,}000{,}000 * 0.36 = 1{,}440{,}000$), so, it is estimated that 182,400 members will be converted into vaccination.

Cost Analysis

As for the cost estimation, the cost estimates assume that 15% of the members in selected regions are interested and opt in for the above services, slightly higher than the estimated conversion rate. The estimated costs for pre-vaccination screening are equal to the *cost of basic body screening* $*$ *estimated number of opt-in members* $\$39 * 0.15 * 1{,}520{,}000 = 8{,}892{,}000$, based on the cost of COVID-19 basic assessment in MinuteClinic[29]. The cost estimates for home visits and mobile clinics varies, depending on the availability of medical personnel and set-up costs for a mobile vaccination trailer.

## 7.3 Stage 3: Personalized Solutions

### 7.3.1 Remaining Population: Exerting node-based Peer Influences through Social Network

Strategy Rationale

According to our segmentation method, there is a remaining group of unvaccinated members, who could be hesitant to receive vaccination by the various reasons, ranging from health concerns to political distrust in the government. Promoting vaccination via peer influence is one of the possible driving forces to overcome these barriers.

Humana could motivate the remaining unsegmented members into receiving vaccination by spreading vaccination tendency via the peer influence on the social networks of the members. Viewing the social network of members in a connected graph, Humana could identify the most influential members (being the central node in a connected graph), and provide personalized persuasion towards the targeted members, such that they could possibly spread their personal choice of vaccination to other members in their social network.

Individual persuasion is more effective than mass education despite being expensive, so we suggest building a social network model to analyze connections between Humana members. Over a half of the hesitant respondents in a survey have concerns over side effects and lack trusts in the vaccine itself, while one-third of them expressed distrust towards the government[30]. Educating the benefits and risks of vaccines and instilling a scientific perspective towards vaccination over politicalized views is difficult and costly, but it could be one of the most effective measure to clear the non-scientific biases towards vaccinations from Humana members. If we can identify the members that can spread the most peer influence over their social networks (“central members”), we can focus on persuasions in these central members and limit the costs.

Strategy Implement Plan

As for the building of the models, connection (edges) between the members can be established by linking the different features of member-level data into a network. For instance, zip code enables us to relate members living in the same county as one neighborhood and estimate the primary

connections. Also, it is also reasonable to infer connections from the members that visited for same doctors, clinics, hospital, or pharmacies or that are inflicted with the same chronic diseases as it is commonplace for patients to exchange opinion about medical treatments and reviews about medical institutions. Finally, in compliance with the data privacy law, consent must have been obtained from Humana members before using these data for this specific purpose of model and network building. In order to attract Humana members to enroll into this consent-based model building scheme, Humana could, in return for their consent of data usage, provide benefits of providing digital communication platforms (e.g., a channel in messaging app) and meetings among members who are interested in to interact with other members that share medical conditions and live nearby. Even if they do not have strong connections, the connections can be established and bolstered with the Humana-host networking scheme.

**Figure 23 Demonstration of social network with a high degree of social connectedness.[31]**

Cost & Effectiveness Analysis

As soon as the model is completed, we need to quantify the influence of the peer influence and estimate the conversion rate for this targeted persuasion policy. A study about social network connected graph showed the opinion converges to consensus under the presence of increasing peer pressure[32]. Another study about peer influence in vaccination dynamics indicates that vaccination can be strongly promoted by peer influence the relative cost of vaccination is below a critical value because the vaccinated individuals will reinforce each other in using it[33]. Also, both models in the above studies acknowledge the common understanding that the stronger the bonding between peers, the faster and easier it is to have a convergence in the opinion dynamics. In response to the finding of the study, to maximize the peer influence from central members after persuading them into receiving vaccination, Humana should strengthen the inter-member network by coordinating the meetings and networking among the members, such that the bonding and peer influence can be increased within the member's social network graph. Only after the social network graph specific to Humana member population is built can we simulate the dynamics of peer influence and estimate the time for the opinion convergence as well as conversion rate into vaccination. Below is the vaccination dynamics in the square lattice population depending on the costs of vaccination (C) and the strength of peer pressure ($\alpha$) from the reference study, showing that the stronger the peer influence and the lower cost of vaccination, the higher fraction of vaccination for the social network.

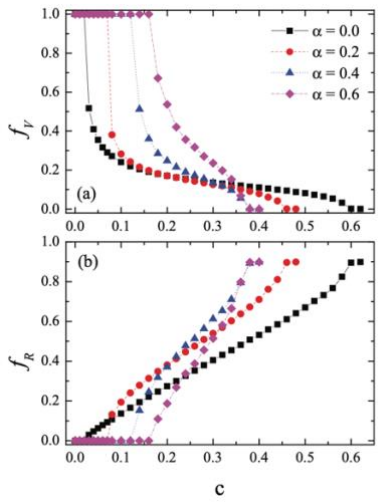



**Figure 24 Vaccination dynamics in the square lattice population.** (The fractions of vaccinated fV (a) and infected fR (b) individuals are shown as a function of the relative cost of vaccination, c, for different values of α, the relative strength of peer influence. The selected parameters can be referred to the original paper)[34]

As for the cost estimation, the cost will depend on the model suggested above. As soon as we simulate the opinion dynamics, we can estimate the minimum number of central members (central nodes) required to lead to a convergence into a positive opinion towards vaccination over the entire or a subset of social network graph. The cost will be gauged based on $per\ person\ cost\ of\ persuasion * minimum\ number\ of\ central\ member\ needed$.

Although there are not definite estimates as to this solution in this report, it is a long-term effective measure to induce influence over the behaviors of the Humana member populations against other factors (such as political and mental hesitancy against vaccination).

### 7.3.2 Personalized Analytical Platform for Vaccination: Based on Health Records and Statistics

Strategy Rationale

If we investigate the reasons for not getting vaccination, we may find the top reasons are lack of trust and uncertainty in its side effects[35], though vaccinations have been more widely used and tested. From a survey[36] on vaccination trust conducted this year, we can find that lack of trust is usually resulted from the opaque and unmatched information. People that are hesitant to get vaccination have questions such as "How do I know about the side effects?" "How do I know if it is effective in face with virus's quick mutations?" "What if I got sick because of the vaccine?" All in all, they need more information to make their decision.

Our goal for building a personalized information platform is to provide more specified information on vaccination with the help of statistics and data analytics. We believe that if one could get information that is more applicable to his situation, it is helpful for building his trust on vaccine.

*Strategy Implementation Plan:*

With the clear consent from members, we suggest adding an analytical module in Humana's user online profile, that can take in members' non-sensitive health information and output personalized assessment and instructions on vaccination as well as corresponding vaccination statistics collected from open sources.

At the same time, this module will contain information on the nearest vaccination sites and appointment access. After the member is vaccinated, they could also reserve follow-up check for the antibody level, and track on their health status through the new module.

This module will have three parts:

1) Analytical part: Based on basic personal information (such as age, sex), health records (such as disease, chronic health problems, medical claims, medicine costs, immunization records, etc.) and voluntarily provided information (such as exercise habits, diet habits, etc.), comparing with public health statistics, then provide information on personal health and vaccination suggestions to the member.
2) Appointment part: Provide information on nearest vaccination sites.
3) Health record tracking part: After vaccination, provide vaccine information related to antibody test.

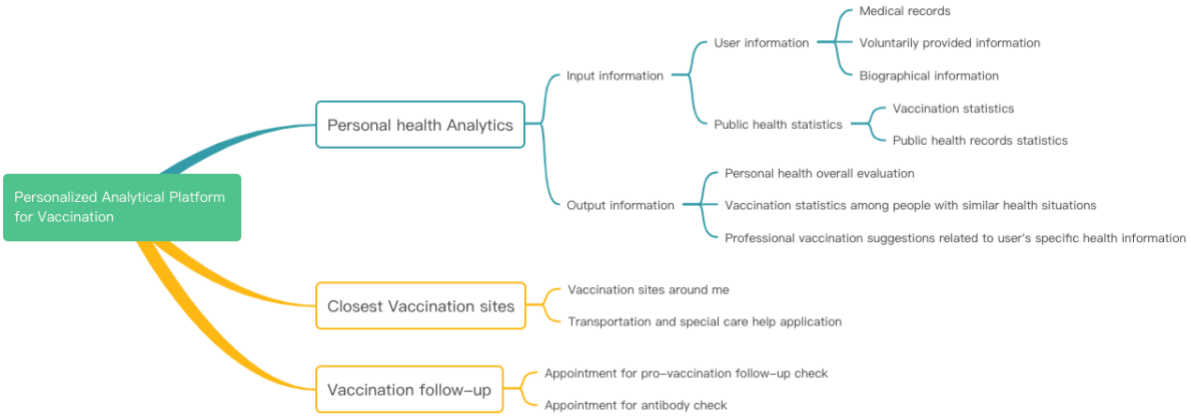



**Figure 25 Personalized Analytics Module Structure**

The unique value for a personalized vaccination information module is that, comparing to the general suggestions for all, these suggestions and statistics would be more focused on personal situations and provide more specific information.

## 8. Conclusion

Humana is interested in identifying Medicare members who are most likely to be hesitant about COVID-19 vaccine. In this case study, we built a model to predict vaccine hesitancy based on XGBoost. Our model shows excellent performance with an AUC score of 0.6839. We then analyzed

the features of paramount importance in our model and utilized them to further perform segmentation among the entire Medicare member population.

In the segmentation analysis, we divided the overall members into four segments based on their biographical, medical, and financial records. We finally obtained 4 segments (people in financial distress, racial minority, people from selected regions and people with disabilities), enabling us to devise corresponding strategies. Our segmentation method covers majority of the unvaccinated group with low overlap among each other.

Finally, we proposed CoVAE (COVID Vaccination Accessibility and Equity) Program, a three-stage program to address vaccine hesitancy of each segment based on data analytical insights. Stage 1 is to provide financial incentives, which particularly targets members who are in financial distress. At Stage 2, a target group program covering 66% of the unvaccinated members, focuses on racial minority segments, disabled segment, and selected region segments. Solutions involves cooperating with opinion leaders, providing special care and transportation assistance for the disabled and setting up mobile vaccination site. At stage 3, we introduced personalized analytical module and network-based peer influence on the group members (account for about 13% of the total population) who are not identified in our segmentation. By taking the additive effects of our strategies into consideration, our weighted success rate for converting unvaccinated members into vaccinated ones is 14.6% of all unvaccinated Humana members.

With our proposed CoVAE program, Humana will effectively help the most vulnerable and underserved groups gain equitable access to COVID-19 vaccination. Part of our solution targets intrinsic motivation by building trust and disseminating accurate information about vaccination. Meanwhile, Humana can help overcome external barriers like distance and disability for proper segments of the unvaccinated members, ensuring accessibility and equity for every individual. By deploying our CoVAE program, Humana can promote the vaccination and proper allocation of health solutions among its members.

# 9. Reference


[1] Wendy C King et al., "Time Trends, Factors Associated with, and Reasons for Covid-19 Vaccine Hesitancy in US Adults: January-May 2021," 2021, https://doi.org/10.1101/2021.07.20.21260795.

[2] same as [1]

[3] *Feature encoding: Machine learning in the Elastic Stack [7.15]*. Elastic. (n.d.). Retrieved October 11, 2021, from https://www.elastic.co/guide/en/machine-learning/current/ml-feature-encoding.html#:~:text=Feature%20encodingedit,process%20is%20called%20feature%20encoding.&text=The%20vector%20represent%20whether%20the,)%20or%20not%20(0).

[4] *8 categorical data encoding techniques to boost your model in python!* Analytics Vidhya. (2020, August 16). Retrieved October 11, 2021, from https://www.analyticsvidhya.com/blog/2020/08/types-of-categorical-data-encoding/.

[5] Centers for Disease Control and Prevention. (n.d.). *CDC Covid Data tracker*. Centers for Disease Control and Prevention. Retrieved October 11, 2021, from https://covid.cdc.gov/covid-data-tracker/#vaccinations-county-view|SVI|RPL_THEMES|all.

[6] *CDC's Social Vulnerability index (SVI)*. (n.d.). Retrieved October 11, 2021, from https://svi.cdc.gov/Documents/FactSheet/SVIFactSheet.pdf.

("Here's All you Need to Know About Encoding Categorical Data (with Python code)," 2020) ("Encoding Categorical Data," 2020)

[7] Chen, T., & Guestrin, C. (2016). XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 785–794. https://doi.org/10.1145/2939672.2939785

[8] Vishal Morde. (2019, April 8). *XGBoost Algorithm: Long May She Reign!* Medium; Towards Data Science. https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d

[9] Xgboost Developers. (2021). *XGBoost Parameters — xgboost 1.2.0-SNAPSHOT documentation*. Xgboost.readthedocs.io. https://xgboost.readthedocs.io/en/latest/parameter.html

[10] Miller, Jenesse. "Education Is Now a Bigger Factor than Race in Desire for COVID-19 Vaccine." USC News, 25 Feb. 2021, news.usc.edu/182848/education-covid-19-vaccine-safety-risks-usc-study/. Accessed 10 Aug. 2021.

[11] Kricorian, K., & Turner, K. (2021). COVID-19 Vaccine Acceptance and Beliefs among Black and Hispanic Americans. PloS one, 16(8), e0256122. https://doi.org/10.1371/journal.pone.0256122

[12] Tolbert, J., & 2021. (2021, May 12). Vaccination is Local: COVID-19 Vaccination Rates Vary by County and Key Characteristics. Retrieved October 10, 2021, from KFF website: https://www.kff.org/coronavirus-covid-19/issue-brief/vaccination-is-local-covid-19-vaccination-rates-vary-by-county-and-key-characteristics/

[13] Ryerson AB, Rice CE, Hung M, et al. Disparities in COVID-19 Vaccination Status, Intent, and Perceived Access for Noninstitutionalized Adults, by Disability Status — National Immunization Survey Adult COVID Module, United States, May 30–June 26, 2021. MMWR Morb Mortal Wkly Rep 2021;70:1365–1371. DOI: http://dx.doi.org/10.15585/mmwr.mm7039a2

[14] Covid-19 vaccine incentives. National Governors Association. (2021, August 9). Retrieved October 9, 2021, from https://www.nga.org/center/publications/covid-19-vaccine-incentives/.

[15] P. Campos-Mercade et al., Science 10.1126/science.abm0475 (2021).

[16] Human development reports. Latest Human Development Index Ranking | Human Development Reports. (n.d.). Retrieved October 10, 2021, from http://hdr.undp.org/en/content/latest-human-development-index-ranking.

[17] same as [15]

[18] Centers for Disease Control and Prevention. (n.d.). Covid-19 hospitalizations. Centers for Disease Control and Prevention. Retrieved October 10, 2021, from https://gis.cdc.gov/grasp/COVIDNet/COVID19_3.html.

[19] *COVID-19 Is Crushing Black Communities. Some States Are Paying Attention.* (n.d.-b). Pew.org. https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2020/05/27/covid-19-is-crushing-black-communities-some-states-are-paying-attention

[20] Minority Serving Institutions Program. (2015, July 1). Www.doi.gov.https://www.doi.gov/pmb/eeo/doi-minority-serving-institutions-program

[21] *Minority-serving institution*. (2021, September 29). Wikipedia. https://en.wikipedia.org/wiki/Minority-serving_institution

[22] *US turns to social media influencers to boost vaccine rates*. (2021, August 10). AP NEWS. https://apnews.com/article/lifestyle-technology-joe-biden-social-media-business-a2992b2881fcef68e1144efa7b869844

[23] *Strategies for Helping Older Adults and People with Disabilities Access COVID-19 Vaccines*. (n.d.). https://acl.gov/sites/default/files/2021-04/ACLStrategiesVaccineAccess_Final.pdf

[24] CDC. (2019, March 8). *Disability Impacts All of Us Infographic*. Centers for Disease Control and Prevention. https://www.cdc.gov/ncbddd/disabilityandhealth/infographic-disability-impacts-all.html

[25] Cochran, A. L., Wang, J., Prunkl, L., Oluyede, L., Wolfe, M., & McDonald, N. (2021). Access to the COVID-19 Vaccine in Centralized and Dispersed Distribution Scenarios. *Findings*, 23555. https://doi.org/10.32866/001c.23555

[26] Cai, M., & DeGuzman, C. (2021, August 3). Covid-19 is spreading fast among Texas' unvaccinated. here's who they are and where they live. The Texas Tribune. Retrieved October 10, 2021, from https://www.texastribune.org/2021/08/03/unvaccinated-texas-demographics/.

[27] Harding, A. (2021, June 17). USF survey reveals biggest driver of vaccine hesitancy in Florida. WJXT. Retrieved October 10, 2021, from https://www.news4jax.com/health/2021/06/17/usf-survey-reveals-biggest-driver-of-vaccine-hesitancy-in-florida/.

[28] Wendy C King et al., "Time Trends, Factors Associated with, and Reasons for Covid-19 Vaccine Hesitancy in US Adults: January-May 2021," 2021, https://doi.org/10.1101/2021.07.20.21260795.

[29] Price List of Screening. Minuteclinic. (n.d.). Retrieved October 10, 2021, from https://www.cvs.com/minuteclinic/services/price-lists#screen.

[30] same as [1]

[31] Nanawati, S. (2019, August 2). Social Network Analytics. Medium. Retrieved October 10, 2021, from https://medium.com/analytics-vidhya/social-network-analytics-f082f4e21b16.

[32] Semonsen, J., Griffin, C., Squicciarini, A., & Rajtmajer, S. (2018, November). Increasing peer pressure on any connected graph leads to consensus. NASA/ADS. Retrieved October 10, 2021, from https://ui.adsabs.harvard.edu/abs/2017arXiv170207912S/abstract.

[33] Wu, Z.-X., & Zhang, H.-F. (2013). Peer pressure is a double-edged sword in vaccination dynamics. EPL (Europhysics Letters), 104(1), 10002. https://doi.org/10.1209/0295-5075/104/10002

[34] same as [33]

[35] Inc, G. (2021, July 30). *Vaccine Hesitancy and U.S. Public Opinion*. Gallup.com. https://news.gallup.com/opinion/polling-matters/352976/vaccine-hesitancy-public-opinion.aspx

[36] Latkin, C. A., Dayton, L., Yi, G., Konstantopoulos, A., & Boodram, B. (2021, February). Trust in a COVID-19 vaccine in the U.S.: A Social-Ecological Perspective. Social science & medicine (1982). Retrieved October 11, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7834519/.