Humana-Mays Healthcare Analytics

2021 Case Competition

Vaccine Hesitancy

Table of Contents

Executive Summary

This study focuses on helping Humana better understand the vaccine hesitancy of individuals and how Humana can best achieve their goal of fully vaccinating their members. Gathered literature showed that vaccination coverage can save millions of lives and still act against Covid variants. Throughout the vaccine rollout process, there was no true national strategy. State level strategies were implemented with a wide variation of reach. Our goal was to generate actionable insights from the data to determine key factors to a member being vaccine hesitant, as well as generate a classification model that predicts the likelihood that a member would be wary of taking the vaccine. The metrics utilized to score our model were ROC and a quantified fairness "disparity" score. The disparity score was calculated as the average of the disparity ratio across each sex and age group. The data was patient-level data collected for Humana patients. The collection for the target variable, vaccinated or unvaccinated, occurred from March of 2020 through March of 2021; whereas the other features like demographics and medical data provided were as of July 2020. After data cleaning, train-test-score using a 70-15-15 split was created for static datasets. To finalize data preparation, data was imputed and scaled using sklearn's StandardScalar. For modeling, we used Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM). Both ensemble tree models tend to perform well on classification models where the data contains non-linear relationships. The average ROC-AUC results for our static test and score datasets was 0.6827. The disparity score for our best model on the score data was 0.994. Disparity with regards to race and gender do not appear to be a pervasive issue in the model. To better understand how the model was making probability predictions, we conducted a post-modeling exploratory analysis with the probability estimates that the model generated on the training dataset. As expected, the probability distribution for the

individuals who were not vaccinated fell in a higher probability range than the individuals who did receive the vaccine. The top 5 most important variables in our model were total partd payment amount, estimated age, child and adult care food program, geographical region, and race. Our group developed a three-dimensional approach to reduce vaccine hesitancy and improve vaccine coverage. The strategies that we developed needed to be able to target not only the population in general, but specific groups based on age, race, and region. We suggested the implementation of Vaccine Mobilization, Healthcare Incentive Plans, and Strategic Public Health Campaigns. Accomplishing these three dimensions will deliver vaccines to local areas, incentivize providers who promote vaccination safety and coverage, as well as provide information to individuals to make educated choices with vaccines. In turn, both vaccine trust and vaccination coverage will be improved.

Case Background

Humana is more than just the third largest health insurance company in the nation, they also offer a wide variety of health and wellness services. Humana serves more than 16.8 million members as of June 2021, and Forbes has ranked them number 41 on the Fortune 500 list. During this unprecedented time for the United States, the healthcare industry has proven just how crucial healthcare workers and services are to the well-being of this nation. As a company, Humana is committed to improving the health of their members, their associates, the communities they serve, and the planet we share.

In early 2020, society watched the Covid-19 pandemic spread across the world. The World Health Organization declared a global pandemic and world leaders began issuing stay at home orders for residents. A year later, in 2021, the focus on ending the pandemic is by way of vaccines. Three vaccines became available, and studies have shown that infection after vaccination is rare, with a vaccinated individual holding only a 0.01% risk of infection (Staff, 2021). Vaccination rollout was strategized at the state level. There was a wide variation of how the phases began, who received the vaccine, and who was being reached within each state. However. the CDC consistently gathers from various sources and releases data over the Covid-19 vaccinations in the United States. As of October 2021, nearly 6 months since the majority of Phase 2 release of the vaccination program rollout, approximately 185 million eligible people have been vaccinated, and approximately 4.74 million having received their booster shot.

Business Problem

The current priority of the pandemic is to reduce vaccine hesitancy. Researchers estimate that 70% to 85% of the country needs immunity against the coronavirus for the Covid variants to stop spreading through our communities (Carlsen, Huang, Levitt, & Wood, 2021). Many studies have been performed to forecast the efficacy and needed measures for vaccination rollout in order to reach this goal. One such study quantified the potential value of decreasing vaccine hesitancy and increasing vaccine coverage, specifically showing the different coverage levels and the impact of the Covid-19 vaccine. Such as, increasing vaccination coverage from 50% to 70% (projected with the vaccination efficacy of 70%, differing from the current 90% efficacy rates), 9.2 million cases could be prevented. Also, the timeline makes a difference, for example, reaching 50% coverage with a 90-day delay will result in an additional 5.8 million cases and $3.5 billion increase in medical costs (Bartsch, et al., 2021). Despite being a successful public health measure, vaccinations are becoming perceived as unwarranted by a growing number of individuals and confidence in vaccines is decreasing. Reducing this hesitancy and reversing this trend will increase vaccine coverage. Increasing this coverage as little as 1% can prevent thousands of cases and save millions in medical costs (Bartsch, et al., 2021).

The purpose of this analysis is to help Humana better understand the determinants of a patient being vaccine hesitant and propose solutions on how to overcome these hurdles to reduce vaccine hesitancy, improve vaccination rates for patients, and have a better understanding on how to prepare and educate for future health crisis.

The ROC (area under the receiver operating characteristic curve) will be used as the key performance indicator. This KPI helps us determine how well the model can predict the probability of no vaccination or vaccine hesitancy. Based on our business problem, our aim is to increase the true positive rate (predicting whether someone is vaccine hesitant) and minimize the false negative rate (we do not want to misclassify someone as non-hesitant when the patient is hesitant). Therefore, the ROC metric is helpful in determining how well the model performs.

ROC is not the only important metric in this analysis. To quantify the model's fairness, a disparity score was calculated for each sex and race group. The disparity score is the average of the disparity ratio (precision of group / precision of privileged group) across each sex and age group. The privileged group in this analysis is white males. The data contained 14 different sex and gender groups.

Data Overview

To perform our analysis, the training data provided consisted of 974,842 records, with 367 features and one target feature. Of these features 101 were externally sourced, 231 were from Humana claims, and 35 are Humana owned. Data collection for the target variable occurred from March of 2020 through March of 2021; whereas the other features provided were as of July 2020. All externally sourced and Humana owned data was not on the individual level, while all the Humana claims data was on the individual level. The patient level data provided included demographic, geographic, medical, and credit information features. The target feature indicated whether a member had received a covid vaccination or not. Data on which vaccine the patients received was not available, simply the binary indicator of vaccinated or not vaccinated within the timeframe of data collection.

Data Exploration and Understanding

An exploratory data analysis was performed to make more sense of the data. We found the target variable to be highly imbalanced with the number of no vaccination being 805,389 (83%), and having a vaccination being 169,453 (17%). Information regarding the geographical regions and other demographic information related to our target variable is outlined below.

The North American regional map (Figure 1) below shows Humana regions located in the Pacific and Central-Midwest areas have the highest percent vaccinated number of patients. Both regions have an average cumulative vaccination percentage of 20%, the west coast regions reach a 19.1% vaccinated, East Coast 17.5% vaccinated, then regions located near the southern U.S (including the gulf states and Florida) have 14.6% vaccinated.

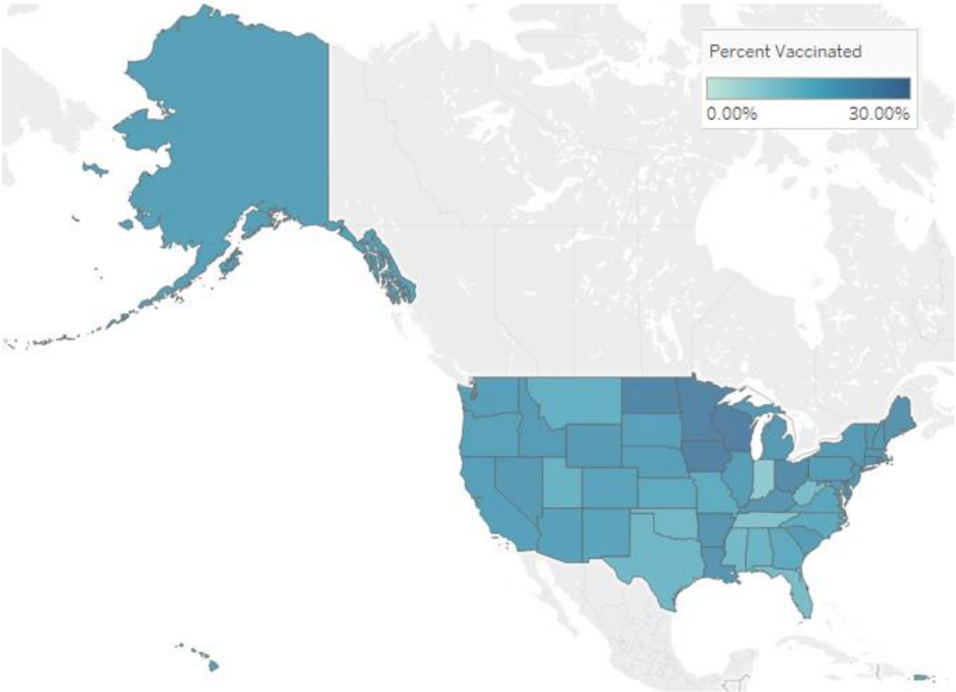


Figure 1: Map of Percent Vaccinated by U.S. State and Territory

There was a visual difference in vaccination rates between gender, where overall females' vaccination percentage is 1.47% higher than that of males (Figure 2). When exploring demographic features further, females of all races have higher vaccination rates than males. Also shown, Black and Hispanic demographics have lower vaccination percentages than other races identified in the dataset (Figure 3, Figure 4).

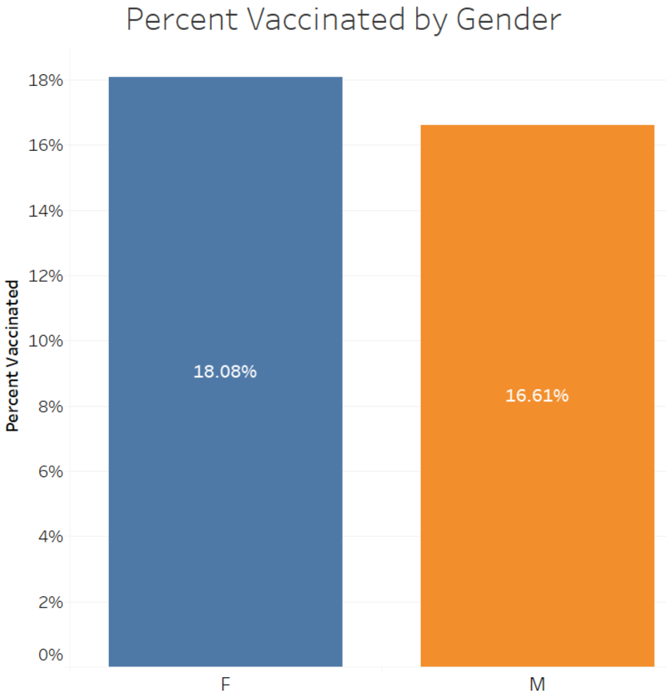


Figure 2: Percent Vaccinated by Gender

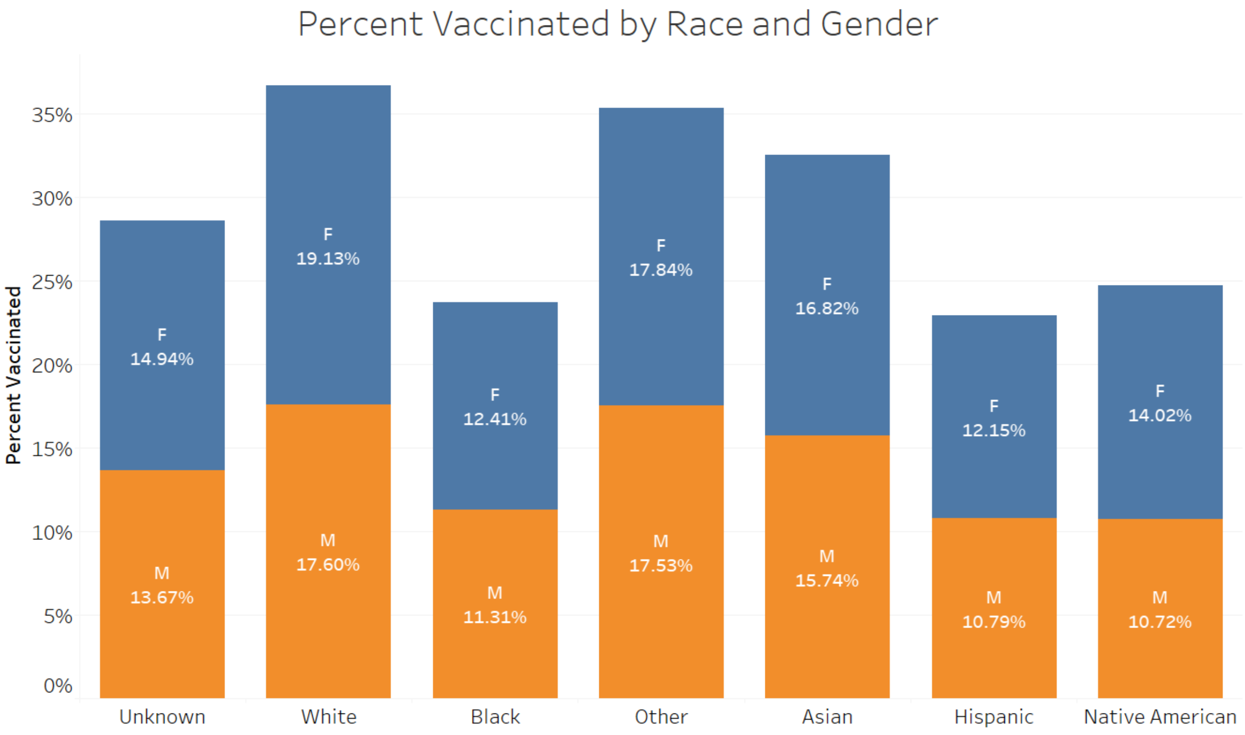


Figure 3: Percent Vaccinated by Race and Gender

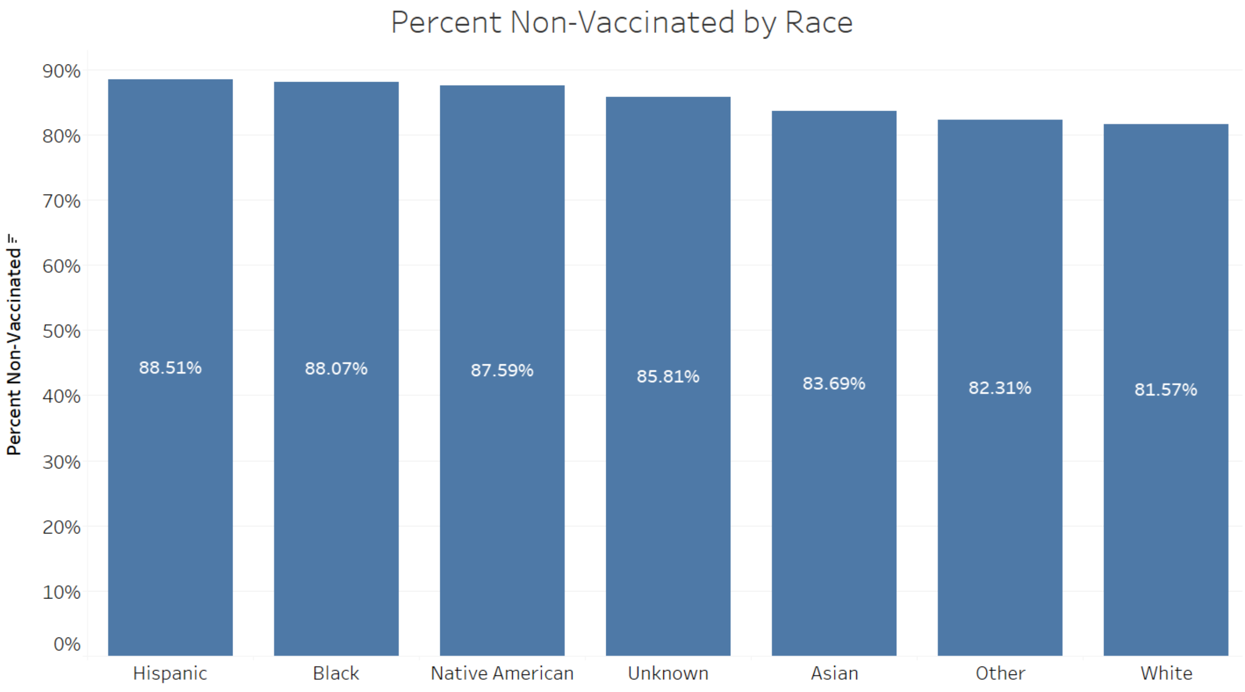


Figure 4: Percent Non-Vaccinated by Race

As mentioned, these demographic groups are accounted for in our disparity score. These figures will be revisited in our analysis section to identify the disparity between subgroups.

The average age of our population is 71.14 years. The age distribution is in the table as follows:

Age Group	Percent of Total
Between 20 and 30	0.63%
Between 31 and 40	1.32%
Between 41 and 50	2.48%
Between 51 and 60	4.69%
Between 61 and 70	25.26%
Greater than 71	65.61%

Table 1: Distribution of Sample Population by Age Group

We expect this distribution of age is due to the time of data collection and the population served by Humana. The individuals would have been part of phase 1 or phase 2 of vaccination rollout in states. These phases included individuals over a specified age and those with high risk or comorbidities. Therefore, due to the data collection timeframe and our population including those patients on Medicare, our data population represents individuals with higher ages or those with disabilities that would have deemed them eligible for vaccination before March of 2021.

Figure 5 below shows the vaccination hesitancy by county poverty percentage. The trend of percent non-vaccinated increases as a county comprises more residents living in poverty stages, with the highest percent non-vaccinated group portrayed being a county that is between 20% and 30% poverty level. In summary, more affluent counties have higher vaccinations while those with 10% and higher poverty composition levels appear to have lower vaccination rates, therefore more likely to be vaccine hesitant.

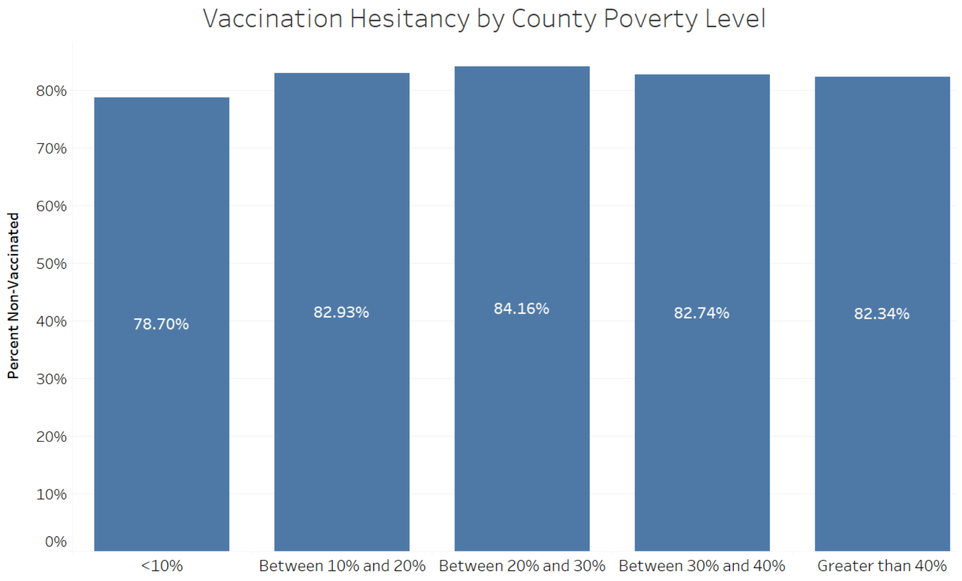


Figure 5: Vaccine Hesitancy by County Poverty Level

The graph below (Figure 6) shows us the percentage of the data population who are unvaccinated by income class. According to *Where Do I Fall in the American Economic Class System* (Snider, 2020), Income classes are distributed as Poor - $32,048 or less, Lower-middle class - $32,048 to $53,413, Middle class - $53,413 to $106,827, Upper-middle class - $106,827 to $373,894, and Rich - $373,894 and up. This shows us that there is not much differentiation between the American Economic classes for those getting vaccinated.

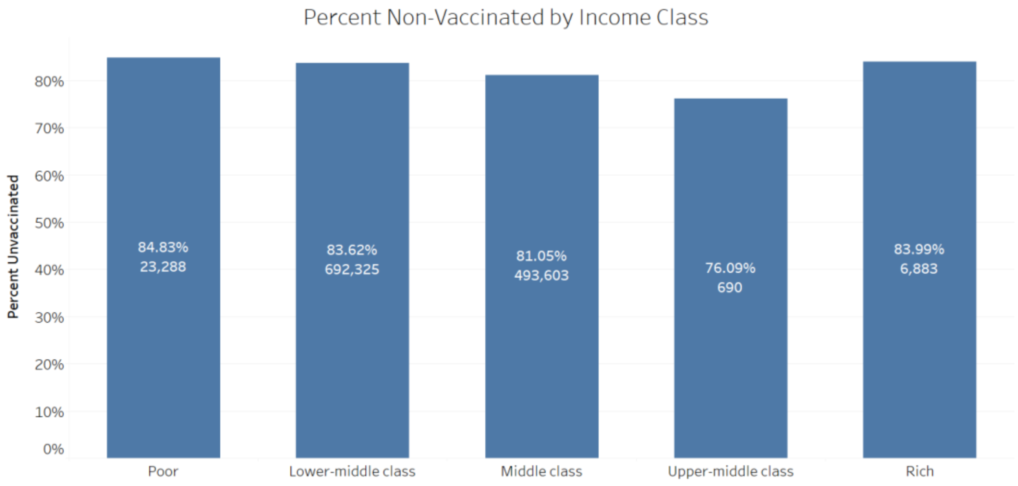


Figure 6: Percent Non-Vaccinated by Income Class

Figure 7 shows us the vaccination percentage by household size. The household sizes are broken into three categories of 1 to 2 people, 2 to 3 people, and more than 3 people. From this, it is shown that the smaller household sizes and the larger household sizes (those made up of 1-2 people and greater than 3 people), are less likely of being vaccinated.

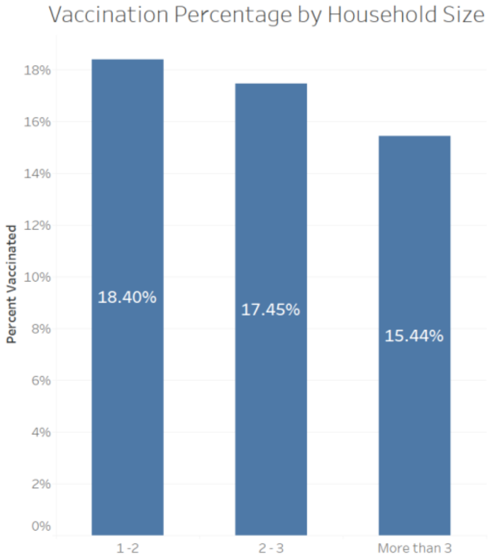


Figure 7: Vaccine Hesitancy by Average Household Size

Figure 8 shown below is the probability that a person is less likely to use a primary care provider as their source information by persons vaccinated and unvaccinated. This gives us information that those who are unvaccinated are less likely to use professionally medical expertise in their medical decisions such as being vaccinated.

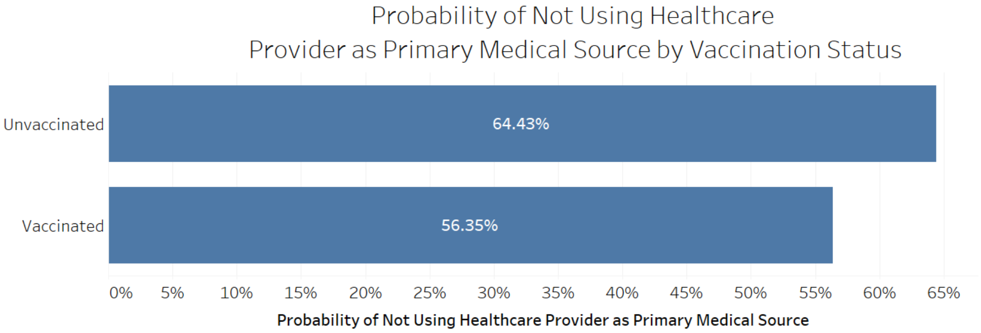


Figure 8: Percentage of those Informed by Healthcare Professionals by Vaccine Status

Data Cleaning and Preprocessing

The training dataset contained 974,842 records and 368 columns while the holdout dataset consisted of 525,158 records and 367 columns. To begin, we dropped the following columns: “Unnamed: 0”, “ID” and “src_div_id.” Before dropping the ID column in the holdout dataset, we saved it to a list so that it could be added back as a column to the data after cleaning.

Next, we replaced “nan”, “*”, “”, “ “, and “null” values with np.nan since they indicate missing data. We identified 15 binary, 9 nominal, 51 ordinal and 289 numeric features in the data. All columns with 0 variance were removed from the data. In total, 54 unary variables were removed.

There were 45 trend variables in the data. These variables contained an ordinal and nominal component. To deal with this, we divided each trend variable into two components. The ordinal represents the magnitude of the change in cost between one 3-month period and another 3-month period. The ordinal component was recoded with the following values:

- Dec_over_8x → -4
- Dec_4x-8x → -3
- Dec_2x-4x → -1
- Dec_1x-2x → -1
- No Change → 0
- No Activity → 0
- Resolved → 0
- New → 0
- Inc_1x-2x → 1
- Inc_2x-4x → 2
- Inc_4x-8x → 3
- Inc_over_8x → 4

Five of the ordinal variables did not contain any variance meaning that no members in the dataset had a change in the cost of claims when comparing a 3-month period to the prior 3-month period. The five ordinal variables were removed from the dataset and are listed below.

- bh_ip_snf_admit_days_pmpm_t_9-6-3m_b4_ord
- bh_urgent_care_copay_pmpm_cost_t_12-9-6m_b4_ord
- rej_med_er_net_paid_pmpm_cost_t_9-6-3m_b4_ord
- rej_med_ip_snf_coins_pmpm_cost_t_9-6-3m_b4_ord
- total_ip_maternity_net_paid_pmpm_cost_t_12-9-6m_b4_ord

The nominal component captures the activity of the client with the following values: Activity, No Activity, Resolved and New. Activity means that there is some cost associated with the claim. Resolved means that there was a cost in the prior period, but no cost in the new period. Finally, No Activity means that there are no claims in either period.

For both race and household composition (cons_hhcomp), we imputed null values with the unknown category label listed in the data dictionary. Missing race values were imputed with 0 and missing cons_hhcomp values were imputed with “U”. For the remaining binary and categorical variables, missing values were imputed with a value of “UNK” for unknown. The binary variables that were imputed with “UNK” for unknown values are now considered categorical values because they have 3 levels. Only three truly binary variables remained: presence of behavioral health condition related to neuro cognition disorder (bhncal_ind), gender (sex_cd) and presence of behavioral health condition related to nc dementia (bh_ncdm_ind). The binary variable “sex_cd” was converted to 1 for Male and 0 for Female.

At this point, a copy of the cleaned training data was exported to later be used in a pipeline with RandomizedGridSearchCV. We also chose to create train-test-score datasets using a 70-15-15 split. We will refer to the train, test, and score datasets as our static datasets because no cross-validation was used with these sets. The static datasets allowed us to quickly try different hyperparameters and analyze the results.

In both the pipeline and static datasets, missing values for numeric and ordinal variables were imputed with the median of the training dataset. Using catboost encoder, all categorical variables were converted to numeric values. The encoder was fit on the training dataset and used to transform the training, test, and score dataset. Finally, each variable was scaled using sklearn's StandardScaler. The scaler was fit to the training data and applied to the train, test, and score datasets. The final datasets contained 350 features. The distribution of the target variable, race and gender can be seen below for the static train, test, and score datasets:

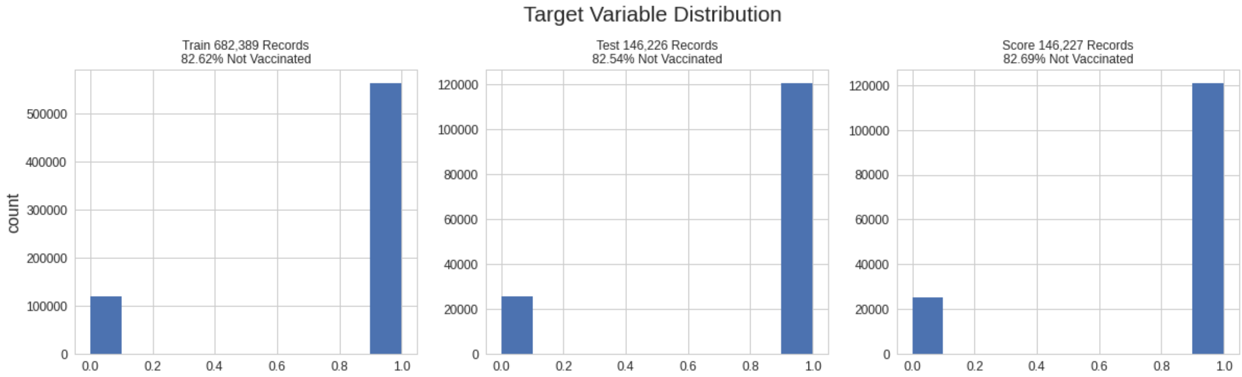


Figure 9: Distribution of Covid Vaccination across Statistic Datasets (1 = Not Vaccinated)

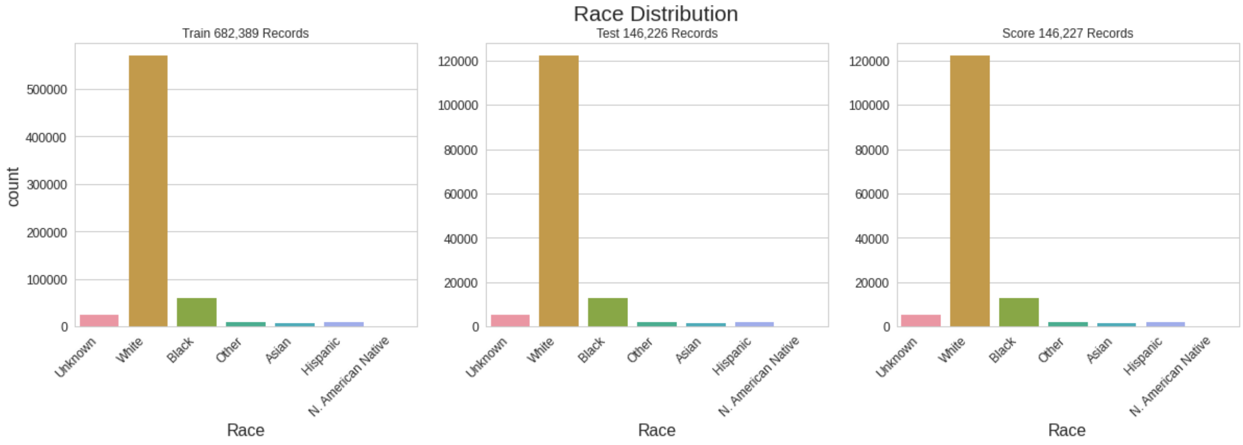


Figure 10: Distribution of Race Categories across Static Datasets

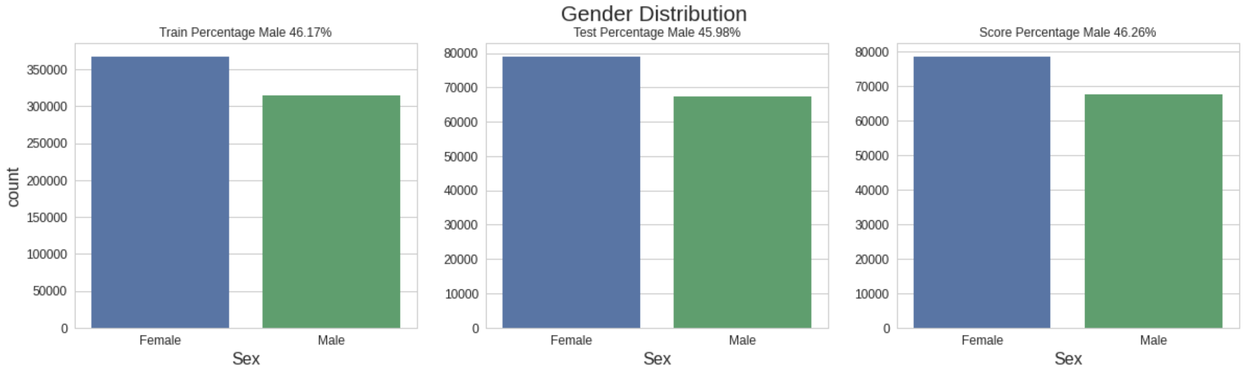


Figure 11: Distribution of Gender across Static Datasets

Analysis

Modeling

For modeling, we used XGBoost and LightGBM. These models were selected because both XGBoost and LightGM are ensemble tree models that tend to perform well on classification problems where the data contains non-linear relationships. LightGBM performs faster and tends to have higher efficiency than XGBoost. This was particularly attractive for our use case since the training dataset provided contained roughly one million records.

After the initial model prototypes, we found LightGBM to produce better results. Given the time constraints on the project, we decided to focus our efforts on optimizing the LightGBM model. To identify optimal hyperparameters, we created a preprocessing pipeline that used RandomizedGridSearchCV to find the best n_estimators, max_depth, num_leaves, learning_rate, feature_fraction and bagging_fraction parameters. The data used for the pipeline did not contain any imputations or transformations to ensure no data leakage occurred during training. The preprocessing steps in the pipeline were imputation of numeric variables with the mean, catboost encoding of categorical variables and scaling the data on a range of 0 to 1. Using 3-fold cross-validation, we ran the data through the preprocessing pipeline and RandomizedGridSearchCV to determine the best parameters for optimizing ROC-AUC.

Using the results from the Randomized Grid Search, we began to train models on our static train, test and score datasets. The Light Gradient Boosting model with the best ROC-AUC used the following hyperparameters:

- boosting_type = 'gbdt'
- objective = 'binary'
- is_unbalance = True
- num_leaves = 70
- n_estimators = 1500
- max_depth = 20
- learning_rate = 0.01
- feature_fraction = 0.6
- bagging_fraction = 0.3
- reg_alpha = 1
- reg_lamba = 1
- seed = 1234

Results

The results for the static datasets can be seen below. Roughly 90% of the observations predicted to be positive were positive. However, the model had a recall score around 60%. The probability threshold used for classifying observations was 0.50. The average ROC-AUC score for between the test and score datasets was 0.6827. The remainder of the model performance metrics can be seen in the Table 2 below.

Dataset	Accuracy	Precision	Recall	F1 Score	ROC-AUC
Train	0.628385	0.917501	0.604552	0.728854	0.737541
Test	0.628623	0.891441	0.626369	0.735759	0.686368
Score	0.594931	0.896199	0.576988	0.702010	0.678979

Table 2: Model Metrics for Final LGBM Model

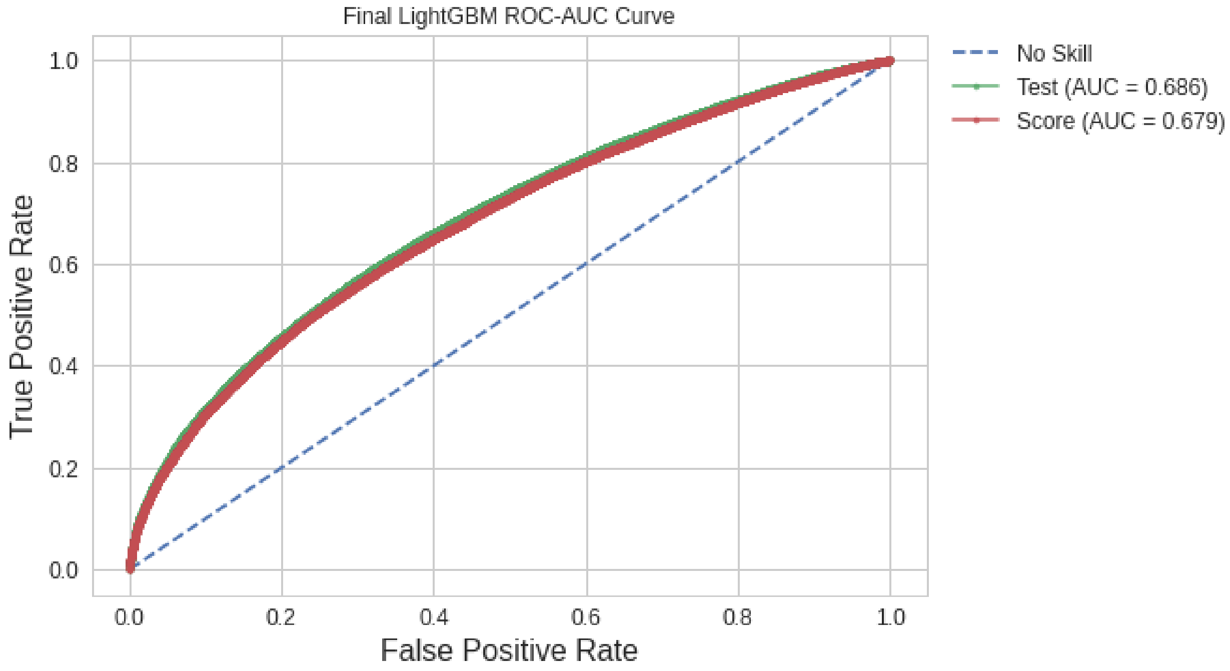


Figure 12: ROC-AUC Curve of Final LGBM Model

To calculate the disparity score, we created a function that found the disparity ratio for each sex, age group and took the average of the ratios to determine the disparity score. The disparity ratio is the precision of each sex, and age group divided by the precision for the privileged group (white, males). The maximum value for each disparity ratio is 1. The disparity score for our best model on the score data was 0.994. Therefore, disparity with regards to race and gender do not appear to be a pervasive issue in the model. The individual disparity ratios for each group can be seen below and the privileged group has been highlighted in yellow.

	Male		Female	
Race	**Precision**	**Disparity Ratio**	**Precision**	**Disparity Ratio**
Unknown	0.938699	1.0000	0.929665	1.0000
White	0.896206	1.0000	0.887529	0.9903
Other	0.890351	0.9935	0.879237	0.9811
Black	0.912296	1.0000	0.902204	1.0000
Hispanic	0.902968	1.0000	0.893271	0.9967
Asian	0.887805	0.9906	0.870098	0.9709
North American Native	0.884956	0.9874	0.942308	1.0000

Table 3: Disparity Ratios for the Score Data using the Final LGBM Model

After the final model was selected, we trained the model on the entire training dataset which was preprocessed according to the methodology described previously. To better understand how the model was making probability predictions, we conducted a post-modeling exploratory analysis with the probability estimates the model generated on the training dataset. As expected, the probability distribution for the individuals who were not vaccinated fell in a higher probability range than the individuals who did receive the vaccine as seen in the Figure 13 below. The average probability of hesitancy among those who received the vaccine was 43.7% compared to 56.3% probability of hesitancy among those who did not receive the vaccine.

	Vaccinated	**Not Vaccinated**
count	169,453	805,389
mean	0.4370	0.5630
std	0.1253	0.1665
min	0.0877	0.0971
25%	0.3459	0.4308
50%	0.4169	0.5437
75%	0.5096	0.6848
max	0.9674	0.9873

Table 4: Summary Statistics of Vaccine Hesitancy Predicted Probabilities

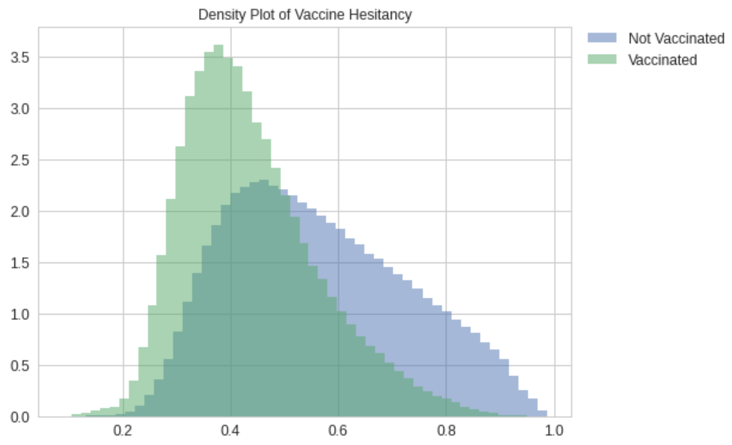


Figure 13: Density Plot of Predicted Probabilities on Entire Training Dataset

Post Modeling EDA

To better understand the model, we conducted a post model exploratory analysis on the most important variables in the final LightGBM. It should be noted that the findings in this section are based on the predicted probabilities from the model for the entire training dataset. As seen in the results, the model is not perfect. The model's recall (or sensitivity) is only performing slightly better than random chance. Therefore, we should be aware of the model's error when interpreting the predicted probabilities. However, it is still important to interpret the results of the predicted probabilities to better understand how the variables are influencing the model's predictions.

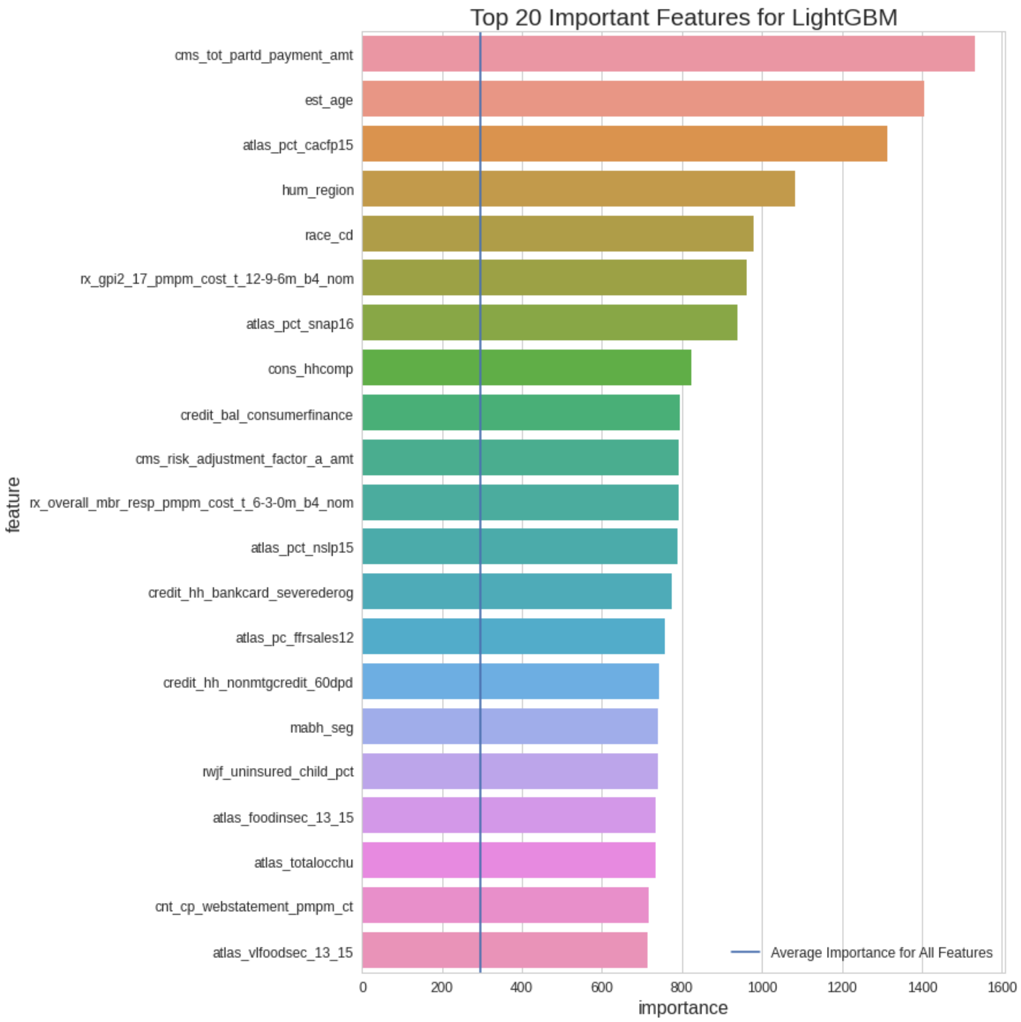


Figure 14: Top 20 Features by Importance for Final LightGBM

The top 5 most important variables in our model were total partd payment amount, estimated age, child/adult care, geographical region, and race. The top 20 variables by feature importance can be seen in Figure 14. The most important variable is Total Part D Payment Amount which we understood to be the monthly fee that the patient must pay for a premium drug coverage plan. The distribution of the variable is similar for both the vaccinated and non-vaccinated groups. However, the non-vaccinated group has a higher average part d payment amount than the vaccinated group. Where the two groups begin to differ drastically is the difference in the 75th percentile. Half of the non-vaccinated group has a cost between $49.35 and $226.60 compared to $45.44 and $150.55 for the vaccinated individuals Table 5. This means there is a greater cost variance for the non-vaccinated group than the vaccinated group. To help illustrate this, the violin plot in Figure 15 shows the transformed distributions between the two groups. The orange line is the non-vaccinated group, and the red line is the vaccinated group. Both the mean and 75th percentile for the non-vaccinated group is clearly greater than the vaccinated group.

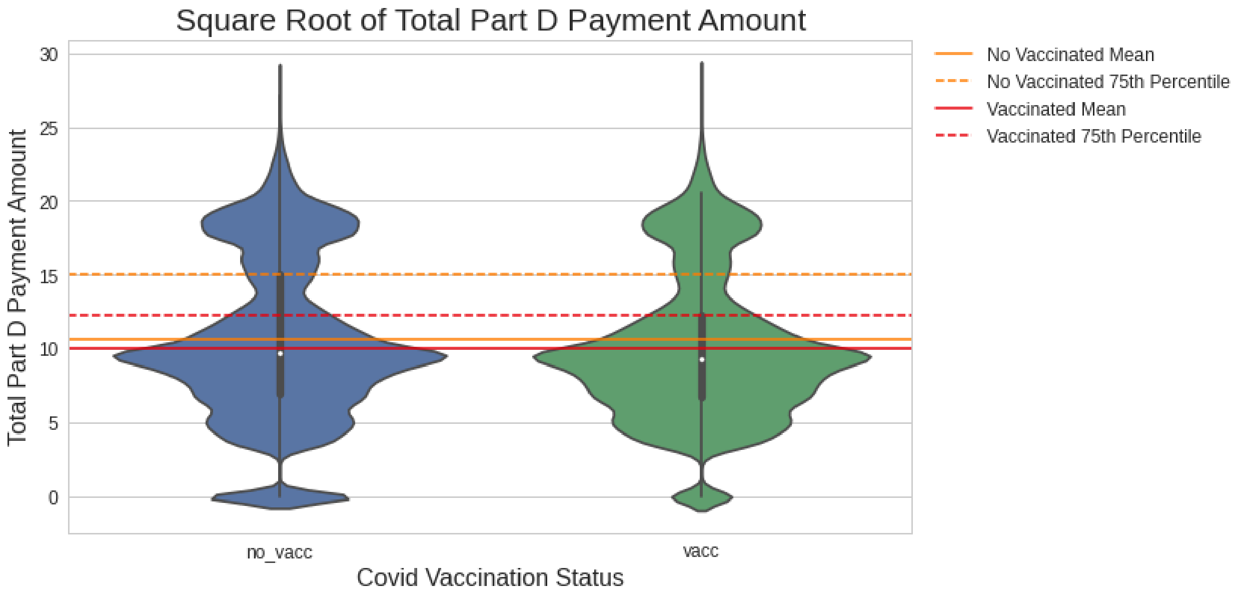


Figure 15: Square Root of Total Part D Payment Amount by Vaccination Status

	Not Vaccinated	Vaccinated
count	774,162	165,665
mean	143.69	125.63
std	128.65	116.47
min	0	0
25%	49.35	45.44
50%	94.34	86.60
75%	226.60	150.55
max	813.06	813.06

Table 5: Summary Statistics of Total Part D Payment Amount

The second most important variable in the model is estimated age of the member. Figure 16 illustrates the average predicted probability for each year of age. The size of the point represents the number of individuals at a specific year of age. As the age of the patient increases, the probability of vaccine hesitancy in our model predictions declines. The greatest decline happens between 50 years of age and 70 years of age. We can assume from this plot that the younger members are more vaccine hesitant than those of more advanced age. To test this assumption, we compared the average predicted probability of age with the actual probability within each age group. We found a similar trend in the actual probabilities even though the probability range of vaccine hesitancy was higher in the actual data. Given that this data was collected in March of 2021, we would assume that more people in the population have received the vaccine since collection time.

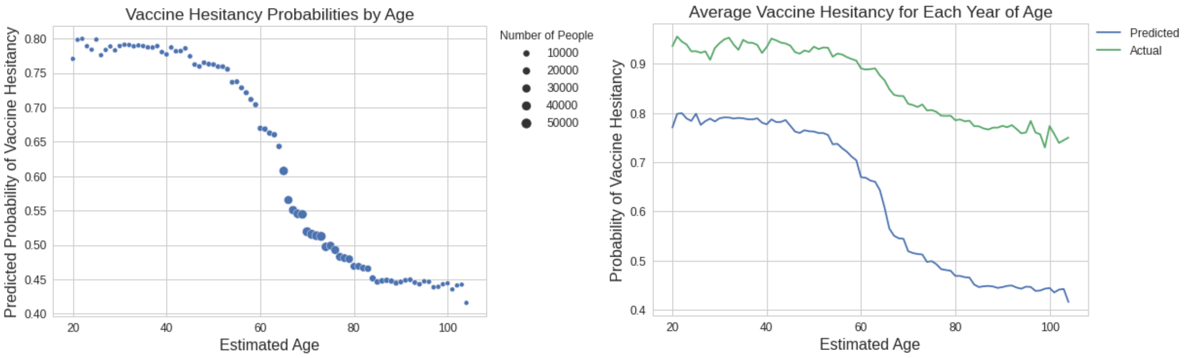


Figure 16: Average Predicted Vaccine Hesitancy for each Year of Age

Race is another variable that we found to be important for predicting vaccine hesitancy. Individuals with an unknown race had a higher average vaccine hesitancy at each year of age compared to other racial groups as seen in Figure 17. The white population is the least vaccine hesitant group. Foreseeably, minority groups are more vaccine hesitant than the white population. For all racial groups, vaccine hesitancy appears to decrease with older age groups.

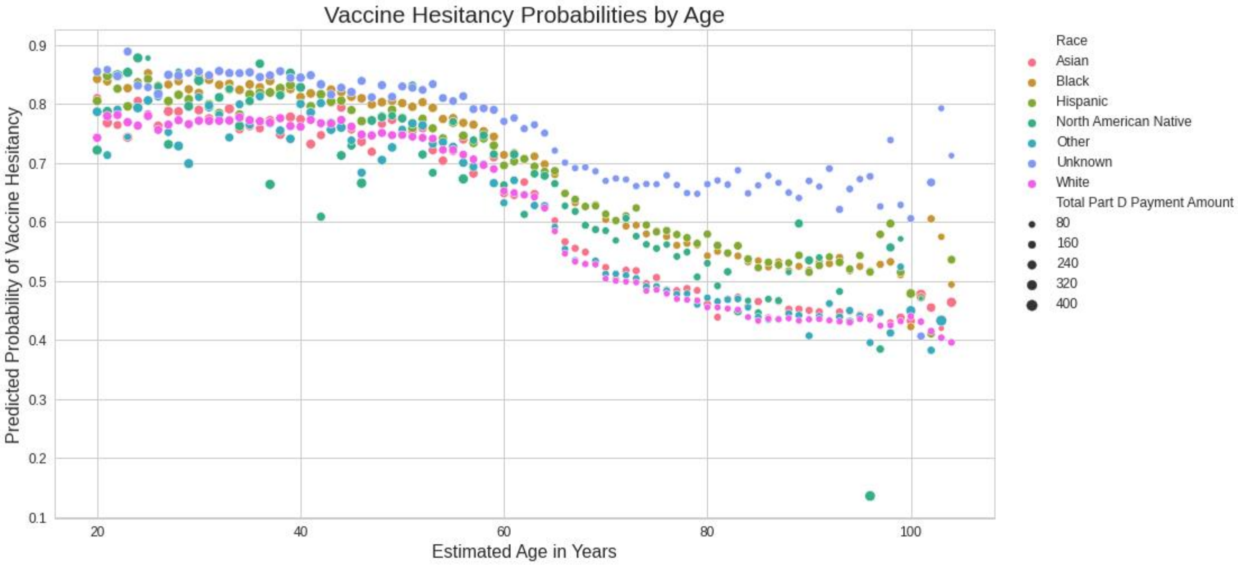


Figure 17: Average Predicted Vaccine Hesitancy for Each Year of Age Across Different Races

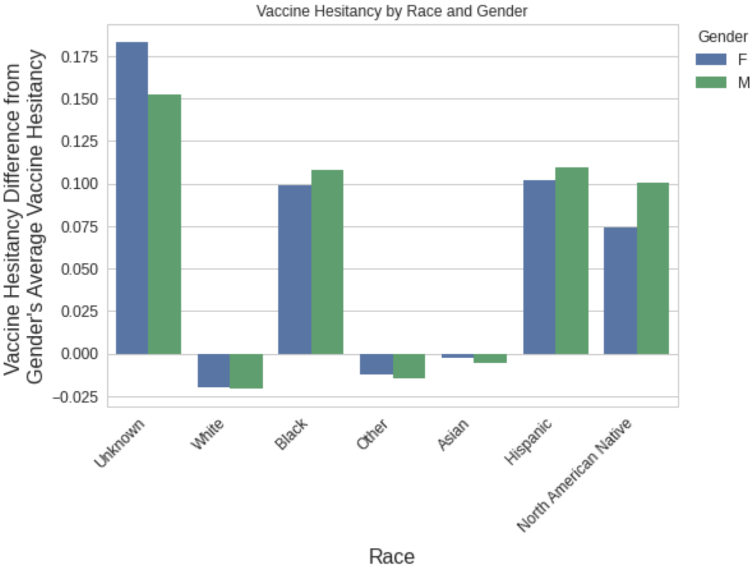


Figure 18: Vaccine Hesitance for each Race and Gender group minus Overall Vaccine Hesitancy for each Sex

Florida is shown to be the most vaccine hesitant group and the Great Lakes/Central North region appears to be the least vaccine hesitant group.

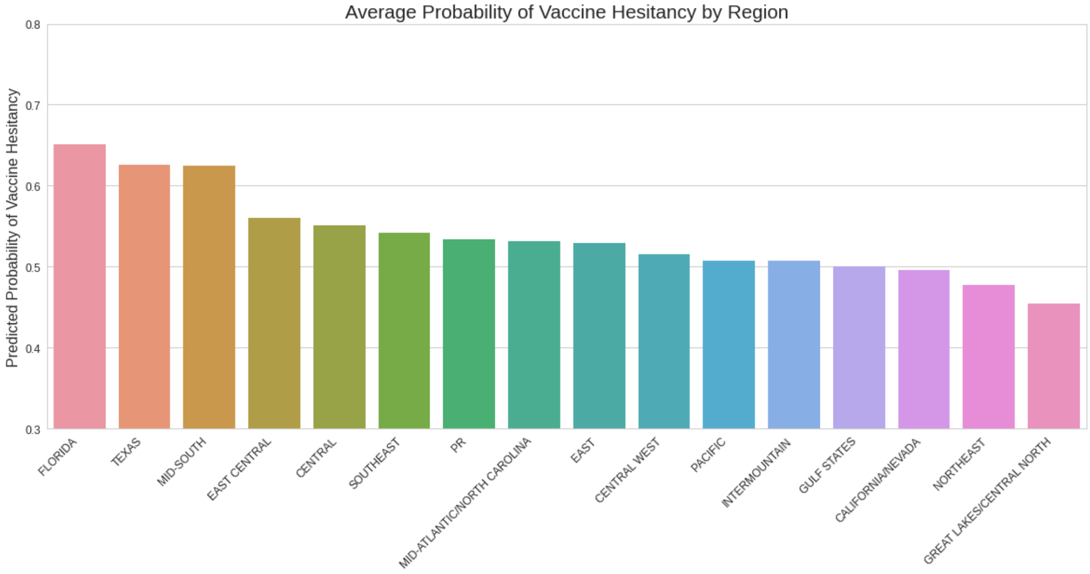


Figure 19: Average Predicted Vaccine Hesitancy for each Region

Region	Vaccine Hesitancy	CACFP (% population)	Number of Records	Percentage of Dataset
FLORIDA	65.14%	1.329896	65608	6.73%
TEXAS	62.57%	1.727749	68980	7.08%
MID-SOUTH	62.44%	1.089524	61079	6.27%
EAST CENTRAL	55.98%	1.083277	155468	15.95%
CENTRAL	55.15%	1.474547	61266	6.28%
SOUTHEAST	54.16%	1.23256	67117	6.88%
PR	53.36%	1.371695	4631	0.48%
MID-ATLANTIC/NORTH CAROLINA	53.18%	1.1052	92716	9.51%
EAST	52.96%	1.2438	69441	7.12%
CENTRAL WEST	51.55%	0.777099	29656	3.04%
PACIFIC	50.79%	1.265542	1123	0.12%
INTERMOUNTAIN	50.72%	1.08108	27280	2.80%
GULF STATES	50.02%	2.057363	47137	4.84%
CALIFORNIA/NEVADA	49.58%	1.272387	43909	4.50%
NORTHEAST	47.78%	1.459561	48535	4.98%
GREAT LAKES/CENTRAL NORTH	45.42%	1.358488	130896	13.43%

Table 6: Summary Statistics of Regional Vaccine Hesitancy and CACFP Percent

The percentage of the population receiving Child and Adult Care Food Program (CACFP) benefits is also in the top five most important variables for our model. CACFP provides reimbursement to child and adult care institutions for nutritious meal and snacks served to children and older adults or chronically impaired persons with disabilities in their care (Child and Adult Care Food Program (CACFP)). To understand how this variable is correlated to vaccine hesitancy, we created 8 bins for vaccine hesitancy and looked at the average of CACFP percentage of the population for each bin as seen in Figure 20. Due to the low number of records in bins 0-9 and 10-19, the bins were combined with 30-29 to create the 0-29 bin. For the most vaccine hesitant groups (80-89 and 90-100) there is a higher average percentage of the population with CACFP benefits. This means that in areas that have higher utilization of CACFP benefits, the individuals in the region are more likely to be vaccine hesitant. This is a generalization based on the information provided in the dataset. We would recommend further investigation before making any decisions based on the information.

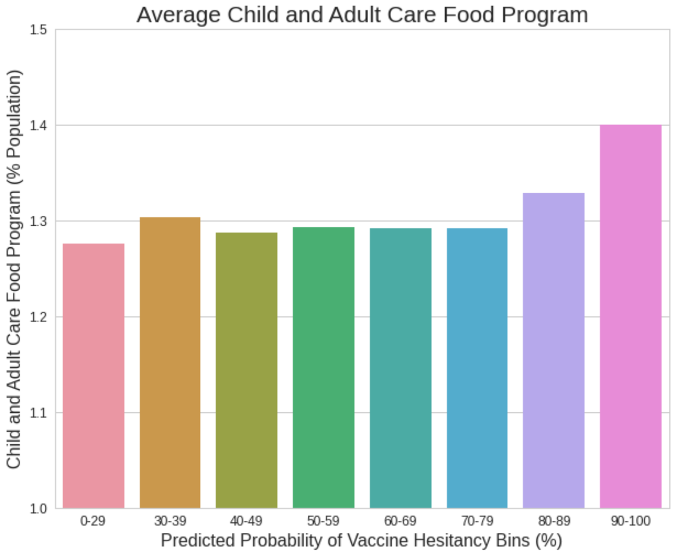


Figure 20: Predicted Vaccine Hesitancy and Child/Adult Care Food Program Percentage

Finally, we wanted to investigate household composition as the final variable for post modeling EDA. Although this variable was not in the top 5, it appeared in the top twenty most important variables. The heatmap in Figure 21 shows the average vaccine hesitancy across different household compositions and race. North American Natives were combined with the other category due to the low number of records. The groups that appear to the most vaccine hesitant on average across all races are the household compositions with a single householder and children present. Those who are married without children in the household are the least hesitant on average across all races.

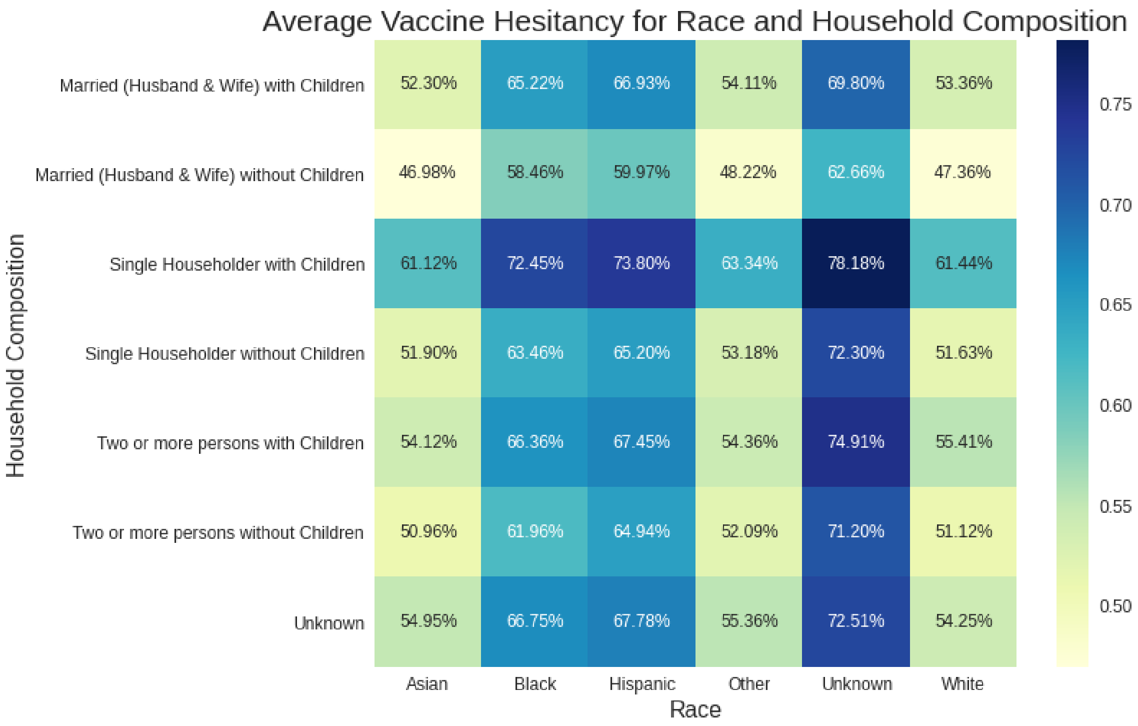


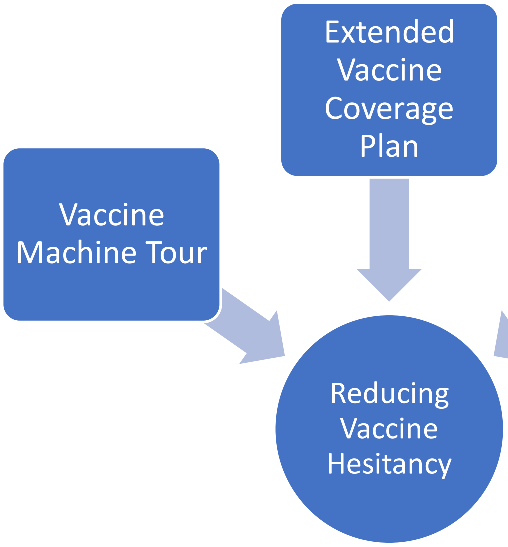
Figure 21: Average Vaccine Hesitancy for Race and Household Composition

Actionable Insights and Recommendations

The aim of this project was to predict those Humana members that are likely to be vaccine hesitant and to design an outreach plan to provide the health solutions that increase vaccine coverage. In this analysis, we developed a predictive model that identifies the probability of a member not being vaccinated also called being vaccine hesitant. The model's probability predictions resulted in an average probability of hesitancy of 43.7% among those who received the vaccine compared to 56.3% probability of hesitancy among those who did not receive the vaccine. The individuals in the non-vaccinated group had a higher partd cost variance than the vaccinated group. Other significant features in predicting non-vaccinated members were age, CACFP subsidized care program, race, and region. The predictions show as the age of the patient increases, the probability of vaccine hesitancy in our model predictions declines. Unknown, Black, and Hispanic races have a higher probability of being vaccine hesitant than the other racial groups.

Healthcare providers have experienced this business problem first-hand. Any solutions geared towards solving the issue with the insights above must be practical for those in the system, patients, and healthcare providers both. Our group interviewed several providers including nurses and physicians working in various medical areas in our state of residence. Insights gained from these discussions revealed that majority of their patients get information from the internet or social media. Overall, the providers felt that vaccines were easily accessible in their areas. Concerns or reasons that an individual is not vaccinated include the reduced time for initial vaccine release (patients are concerned that the vaccine was developed too quickly), misinformation about the vaccines, and a patient does not want to miss work or have time to go get the vaccine. These healthcare providers bared that most of their patients have a primary healthcare provider, but the regularity of visits between the provider and patients were unknown. Through this knowledge, our group determined it is important for proper communication to members should occur and healthcare providers are key players in vaccine coverage. The interaction between providers and patients has been named as a foundation of keeping confidence in vaccination rates, if a provider has deep knowledge and a positive attitude towards vaccines then the provider vaccination coverage rate is higher. There is a positive association with this provider coverage and promotion of the vaccine to patients (Laberg, Guay, Bramadat, Roy, & Bettinger, 2013).

Our group developed a three-dimensional approach to reducing vaccine hesitancy and improving vaccine coverage. Based on our model findings, our goal is to have strategies that inform Humana patients as well as make the vaccine available to those with the barriers preventing vaccine accessibility. The strategies outlined will be able to target not only the population in general, but specific groups based on age, race, socio-economic status and region. The three-dimensional strategies have a wide variety of practical marketing strategies for these targeted segments. For Humana to have a marketing strategy that can effectively target the segments above, we suggest the adoption and implementation of Vaccine Mobilization, Healthcare Extended-Incentive Plans, and Strategic Public Health Campaigns. These strategies will be able to target the specific segments, as well as distribute covid vaccines, information, and incentives that will benefit Humana patients and partners.



Vaccine Machine Tour

Vaccine mobilization should remain a top priority to ensure maximum vaccine coverage. This first dimension directs local health departments to increase mobilization efforts, delivering vaccines to areas in the community. The ability to distribute vaccinations to areas that have lower levels of transportation services could help induce a person's aptitude to become vaccinated. This also would be beneficial to rurally populated areas that have much greater distances to and from locations that provide vaccinations. Finally, this program also enables families with non-traditional makeup, minimum availability, and time constraints to easily receive a vaccine. Creating a "Vaccine Machine Tour" would help identify these types of areas within communities and regions that are healthcare – desolate and provide quick and easy accessibility for those who are unable to get vaccinated due to location and distance constraints.

The program includes county health departments sending out staff to local areas to distribute vaccines to individuals in the community. If a vaccine mobilization effort isn't already in place for a county within a Humana region, there are a few easy steps to implement this program. This program would also use existing resources available to a county health department.

- Schedule time at local community centers, stores, neighborhoods, and other businesses to create pop-up vaccine distribution sites
- Use health department resources such as vehicles and allotted vaccines to transport to site
- Schedule healthcare workers to provide the vaccinations to individuals in the community

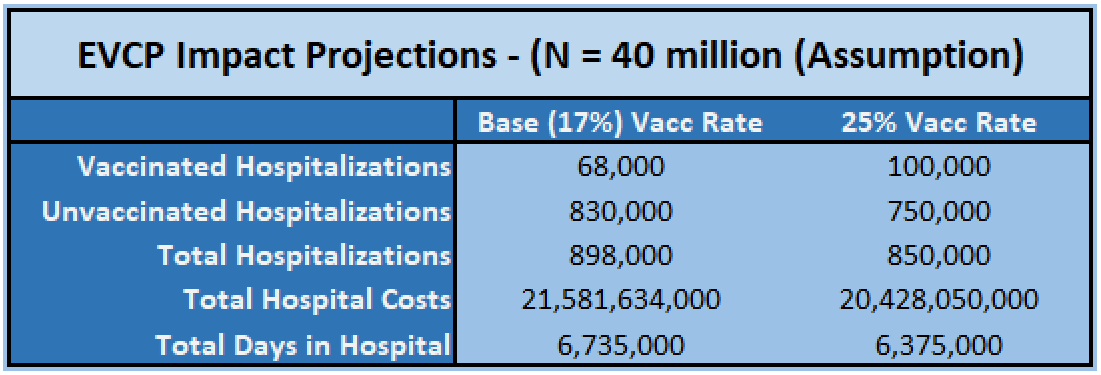
Extended Vaccine Coverage Plan

The second dimension, the Extended Vaccine Coverage Plan (EVCP), attempts to remove the fear of hospitalization costs due to the vaccine and other infection factors. We propose the implementation of out-of-pocket cost deduction incentives for Humana patients to reduce direct medical costs to those individuals.

When looking at assumptive-based financial models, as the vaccination rate of Humana members increase, the savings retained from fewer cases of hospitalizations greatly outweigh the amount of out-of-pocket cost paid by each member. Our group developed the following cost savings model to justify the implementation of this program. This model includes the following assumptions:

- Number of Records in Dataset: 974,842
- Average Cost of COVID-19 Hospitalization: $24,033 (Covid19 Data SnapShot Public Release)
- Average Length of Hospital Stay (Days): 7.5 (Covid19 Data SnapShot Public Release)
- Vaccinated COVID Cases Leading to Hospitalization: 1% (Science Brief: COVID-19 Vaccines and Vaccination, 2021)
- Unvaccinated COVID Cases Leading to Hospitalization: 2.5% (Science Brief: COVID-19 Vaccines and Vaccination, 2021)

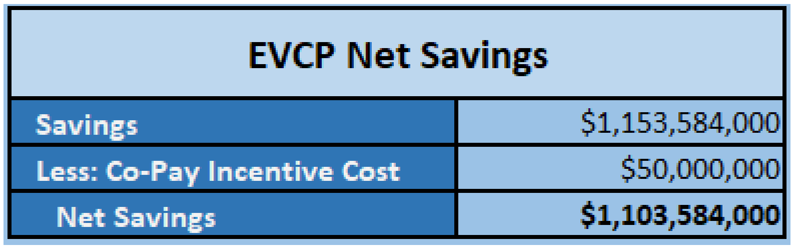
We use these assumptions to calculate the figures shown in Table 7 below. Using Humana's annual report, we concluded that Medicare Advantage plan has around 40 million members (Humana 2020 Annual Report). Therefore, projections using the same assumptions are based on a sample size of 40 million. The projected figures include vaccinated hospitalizations, the number of unvaccinated hospitalizations, total hospitalizations, total hospital costs and total days in hospital.



EVCP Impact Projections - (N = 40 million (Assumption)		
	Base (17%) Vacc Rate	25% Vacc Rate
Vaccinated Hospitalizations	68,000	100,000
Unvaccinated Hospitalizations	830,000	750,000
Total Hospitalizations	898,000	850,000
Total Hospital Costs	21,581,634,000	20,428,050,000
Total Days in Hospital	6,735,000	6,375,000

Table 7: Calculated Figures for EVCP Projections

The training dataset showed 17% of Humana members being vaccinated. When we successfully promote vaccine coverage, we would anticipate the number of vaccination hospitalizations in the population to increase and the unvaccinated hospitalizations to decrease. Above we calculated cost savings if 25% of Humana members were vaccinated. In the population sample of 40 million, approximately billion dollars is potentially saved. When broken down into how many vaccinated hospitalizations are projected to happen, this equates to about $11,535.84 of savings per projected hospitalization.



EVCP Net Savings	
Savings	$1,153,584,000
Less: Co-Pay Incentive Cost	$50,000,000
Net Savings	**$1,103,584,000**

Table 8: Net Savings Calculation for EVCP

We found that the typical out-of-pocket cost per hospital stay of Humana Medicare Advantage hospitalizations is around $500 (Medicare costs at a glance). Through the EVCP, Humana would waive the out-of-pocket cost of hospital visits during this pandemic for those who have received the vaccine. Household income had above average importance in our model. Additionally, economic indicators such as CACFP and SNAP were in the top 20 important variables. Therefore, we have reason to believe that concerns for healthcare costs could be contributing to vaccine hesitancy due to the risk of vaccine symptoms healthcare cost incurred. Table 8 above shows the net savings Humana would incur through the implementation of this program. Costs savings from reduced hospitalizations comes to $1,153,584,000 less the out-of-pocket cost coverage for patients of $1,103,584,000 amounting to a net savings of $1,103,584,000.

Vacc Facts Campaign

In this third dimension, Humana should focus public health communication efforts to reach those groups that are most vaccine hesitant. Our model predicted that younger individuals are more likely to be non-vaccinated, and through discussions with healthcare providers we found that a common platform used to gain information is social media. Humana could gain more information delivery coverage and improve trust in vaccination information by allocating marketing money to a social media campaign. In this social media marketing campaign, information should be visualized and posted as infographics and relatable stories. Along this, referencing and directing to a FAQ online page. This would allow individuals to see the most common asked questions and email questions and concerns about the vaccine. This FAQ will allow those people who are 50-70% probable to be hesitant to submit questions and gain information to allow them to make those educational decisions for themselves.

Future Studies and Suggestions

Given the time constraints and time-period of this study, we would like to consider the following in future studies.

- Vaccine clinic location such as number of vaccines sites per zip code
- An outline of vaccination rollout plans for each state
- Rerun this analysis with information on same individuals today
- Reason field of why vaccinations were received (do a survey of members)

Expanding the information of covid vaccination clinic locations; this will help provide in depth information on patient's distance to these clinics which our group believes could help indicate trends for vaccinations based on rurality. Gaining an outline of vaccination rollout plans would enable researchers to align timeframes with demographic variables like age and disability. As each state had a different schedule, our group found it difficult to truly understand the individuals that would have been eligible as a whole during and at the end of the data collection period to receive the vaccination doses.

Following this, for future studies include rerunning this analysis with the same individuals' current vaccination status. Sentiments and vaccination statuses may have change over time. Reperforming this analysis on the new information could give insightful clues to whether the same drivers of vaccine hesitancy persist.

Finally, including a reason for vaccination field in a survey to integrate as a variable would be useful in analysis. In discussion with healthcare providers, it was noted that motivation to vaccinate would be to protect themselves, family, friends or viewed it as a social norm. It would be interesting to detect whether this altruistic sentiment is the major driving motivation that the data detects as well and incorporate that information into the marketing strategy performance in our solutions.

References

Bartsch, S., Wedlock , P., O'Shea, K., Cox, S., Strych, U., Buzzo, J., . . . Lee, B. (2021, May 06). Lives and Costs Saved by Expanding and Expediting Coronavirus Disease 2019 Vaccination. *Journal of Infectious Diseases*. Retrieved from https://academic.oup.com/jid/article/224/6/938/6267841?login=true

Carlsen, A., Huang, P., Levitt, Z., & Wood, D. (2021, October 1). How is the COVID-19 vaccination campaign going in your state? Retrieved from https://www.npr.org/sections/health-shots/2021/01/28/960901166/how-is-the-covid-19-vaccination-campaign-going-in-your-state

Child and Adult Care Food Program (CACFP). (n.d.). Retrieved from https://www.benefits.gov/benefit/5871

Covid19 Data SnapShot Public Release. (n.d.). Retrieved from https://www.cms.gov/files/document/medicare-covid-19-data-snapshot-services-through-2021-03-20.pdf

Humana 2020 Annual Report. (n.d.). Retrieved from https://humana.gcs-web.com/static-files/78c99040-2eed-4231-89f9-e12b3e9ec333

Laberg, C., Guay, M., Bramadat, P., Roy, R., & Bettinger, J. A. (2013). Vaccine hesitancy. *taylor & Francis Online*. Retrieved from https://www.tandfonline.com/doi/full/10.4161/hv.24657

Medicare costs at a glance. (n.d.). Retrieved from https://www.medicare.gov/your-medicare-costs/medicare-costs-at-a-glance

Science Brief: COVID-19 Vaccines and Vaccination. (2021, September 15). Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/fully-vaccinated-people.html

Snider, S. (2020, Dec 8). *Where Do I Fall in the American Economic Class System?* Retrieved from U.S. News: https://money.usnews.com/money/personal-finance/family-finance/articles/where-do-i-fall-in-the-american-economic-class-system

Staff, A. (2021, June 3). A Timeline of COVID-19 Vaccine Developments in 2021. Retrieved from https://www.ajmc.com/view/liquid-biopsy-radiological-response-predict-posttreatment-outcomes-in-braf-mutated-melanoma