Humana

Humana-Mays Healthcare Analytics

2021 Case Competition

Predicting which members are likely to hesitate towards COVID vaccination

Table of Contents

1. Executive summary 1

2. Case Background 2

3. Data Preparation 3

4. Modeling 6

5. Key Performance Indicator Analysis 8

6. Recommendations and Managerial Implications 10

    6.1 Recommendations based on demographic features 11

    6.2 Recommendations based on health factors 15

    6.3 Recommendations based on wealth factors 18

7. Conclusion 19

8. References 20

1. Executive summary

This study aims to help Humana identify members most likely to be hesitant to the COVID vaccine, draw insights from the data, and more importantly, provide recommendations and potential solutions to drive vaccination among the sub-segments of hesitant members. Plus, we intend to mitigate potential bias inherent in the data to demonstrate fairness and equity.

To begin with, we identified the target variable “covid_vaccination” (“vacc” if vaccinated and “no_vacc” if hesitant to vaccination). The target variable was highly imbalanced. Then, we divided the features into eight categories and examined them using EDAs and statistical tests. Data preprocessing was then done to drop useless columns, transfer data types of variables, and deal with missing values, which helped the model capture underlying patterns. After data preprocessing, we weighed the pros and cons of different models and chose the optimal one - LightGBM to perform our study. After hyperparameter tuning, we got a Receiver Operating Characteristic (ROC) Area Under Curve (AUC) of 0.6848.

We discovered intriguing patterns after performing feature importance analysis. Member age, member’s geographic region, (Risk Adjustment Factor) RAF amount, and percentage of adults under 65 without health insurance, etc. have a significant impact on whether a member will be hesitant to receive the vaccination.

We then divided the hesitant members into four sub-segments and proposed tailored recommendations accordingly. For young members, Humana can leverage social media like Twitter and TikTok to broadcast scientific vaccine information. For members in certain regions where the vaccination rate is low, Humana may partner with local clinics and pharmacies to offer cash incentives to promote vaccines. As for members in doubt of the safety of vaccines, emails/brochures can be sent out to introduce the safety and effectiveness of vaccines to the population. For low-income groups with less access to vaccines, Humana can provide some 24/7 walk-in services to accommodate their needs.

2. Case Background

Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus. It struck the United States in early 2020 and became an ongoing pandemic accumulating 42,966,938 confirmed cases by the end of September in 2021 (Elflein, 2021). The unprecedented pandemic has resulted in severe social and economic disruption. As the pandemic continues, it has become more and more important to increase vaccination in our community. Currently, around 56.5% of the population is fully vaccinated in the US, leaving half of the population unvaccinated and at higher risk to the virus (Ritchie et al., 2020). Successfully vaccinating a large population helps stop the spread of the virus and enhance community immunity. As a leading healthcare company, Humana also pays close attention to the COVID-19 vaccination among its members. Despite that Humana is sparing no effort on providing vaccination opportunities to the most vulnerable and underserved populations, there is still a portion of the members who are hesitant to covid vaccination. Many possible factors may account for the resistance, including lacking trust in science and insufficient education on the advantages and disadvantages of getting vaccinated. Therefore, this project is established to look for the driving forces affecting one’s vaccination attitude. We aim to identify members most likely to hesitate to the COVID vaccine and develop solutions to communicate the benefits of immunization to different populations for Humana.

3. Data Preparation

3.1 Data Understanding

To perform our analysis, we were given a training dataset of 974,842 rows and 368 columns. The data was based on customer level with 8 different subcategories of features including Medical Claims, Pharmacy Claims, Lab Claims, Demographics, Credit Data, Condition Related Features, CMS, and Others. Below are some insights we have gained after doing some preliminary research into the data:

- The data was highly imbalanced with the target variable in the ratio of 805,389 for class 0, i.e., people who are hesitant to get vaccinated; and 169,453 for class 1, i.e., people who are not hesitant to get vaccinated.
- Missing values were displayed in two forms: "*" and NAN. After exploring the distribution for both types of missing values, we found that in columns where "*" and NAN coexisted, "*" only accounted for a very small portion (less than 1%).
- "Zip_cd" contained 83,672 unique values. However, the US has only 41,692 valid zip codes. We can conclude that much noise exists in the feature.
- Categorical features like "cons_hhcomp", "hedis_dia_hba1c_ge9", "lang_spoken_cd", "mabh_seg" had a very high missing rate (over 20%). Specifically, "mabh_seg", one strongly related to one's health consciousness, had about 66% null values.

3.2 Feature Engineering

Many columns came with mixed data types. Some were numeric but came in object forms, while some were categorical but came in numeric forms. We transferred all these columns to the right data types. As for the two forms of missing values "*" and NAN, we replaced all "*" with NAN.

During the feature selection process, we only kept features with more than one unique value. Moreover, if one column had only two unique values and the frequency of one value was over 99.9%, we dropped the column as well. Features with a zero or extremely small variance will not influence our target feature of "covid_vaccination". Thus, we removed the following 40 columns.

auth_3mth_post_acute_rsk	auth_3mth_bh_acute_men
auth_3mth_post_acute_ben	auth_3mth_acute_hdz
auth_3mth_acute_ccs_048	auth_3mth_acute_men
auth_3mth_acute_end	auth_3mth_rehab
auth_3mth_hospice	auth_3mth_acute_ccs_086
auth_3mth_dc_hospice	auth_3mth_acute_cer
auth_3mth_acute_ccs_030	auth_3mth_acute_dia
auth_3mth_acute_skn	auth_3mth_acute_ccs_067
auth_3mth_acute_neo	auth_3mth_acute_ccs_043
auth_3mth_post_acute_vco	auth_3mth_acute_cir
auth_3mth_post_acute_dig	auth_3mth_acute_ccs_094
auth_3mth_post_acute_hdz	auth_3mth_post_acute_cad
auth_3mth_acute_ccs_172	auth_3mth_acute_ccs_044
auth_3mth_acute_ccs_154	auth_3mth_post_acute_ckd
auth_3mth_post_acute_res	auth_3mth_post_acute_ner
auth_3mth_acute_inf	auth_3mth_acute_ccs_042
auth_3mth_acute_cad	auth_3mth_post_acute_inf
auth_3mth_post_acute_cir	auth_3mth_acute_sns
auth_3mth_acute_inj	auth_3mth_post_acute_end
auth_3mth_acute_ccs_153	auth_3mth_acute_gus

Table 1: 40 Dropped Columns

As mentioned above, "zip_cd" had many inaccurate entries. So, we dropped this column as well since "hum_region" already contained a lot of geological information.

What’s more, considering fairness among different races and sexes, we want our model to unbiased against certain races or sex. Thus, we removed “race_cd” and “sex_cd” from our model. By doing so, we found it helpful to strengthen the fairness in our model, which resulted in fair and reasonable predictions for all populations.

Next, we segregated the features as numeric and categorical features. Some examples of numeric features were the “atlas_” and the “cons_” features. Some examples of categorical features were the “med_” and the “rx_” features.

a. For numeric features:

We adjusted some data types to “Int64” to reduce memory and speed up our execution. As for missing values, we left them as they were since our model accepted null values. For better performance, we also standardized all the numeric features.

b. For categorical features:

We filled those columns containing 10 percent and more missing values with a new category “blank”, and for those with less than 10 percent missing, we imputed them with the most frequent category. Then, we selected Ordinal Encoding for categorical features for two reasons. One reason is that having more than 50 categorical features with high cardinality, using methods like one-hot encoding or dummy variable encoding will result in a large sparse matrix, which may adversely impact model performance. The other is that our model can take in integer-encoded categorical features. Integer-encoding often performs better than one-hot encoding in LightGBM (Advanced Topic - LightGBM, 2021).

4. Modeling

LightGBM is a gradient boosting framework that makes use of tree-based learning algorithms. It is considered to be a fast-processing algorithm and yields very high accuracy. Moreover, as previously mentioned, LightGBM enables the missing data handled by default, which allows null values in our data. LightGBM also offers good accuracy with integer-encoded categorical features which enables us to use Ordinal Encoding to encode our categorical features (Advanced Topic - LightGBM, 2021). Considering all the above advantages and convenience, we chose the LightGBM model for our predictive analysis.

For our model, the hyperparameters which gave us the highest value of AUC score were:

Parameter	**Value**
boosting_type	gbdt
objective	binary
max_depth	12
num_leaves	42
random_state	1204
learning_rate	0.05
n_estimators	500

Table 2: LightGBM Tuned Hyperparameters

After fitting the model on our train data, we used the model to predict the output on our holdout data.

We got the following validation values:

AUC= 0.6842878745741854

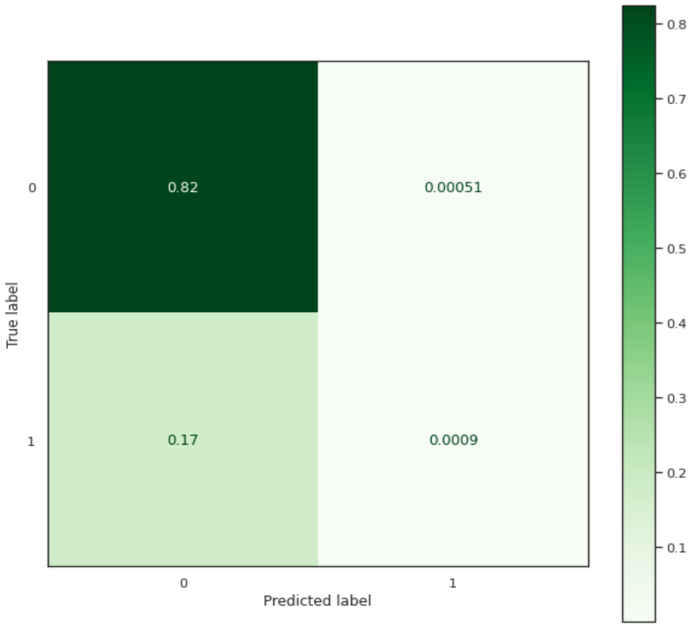

Fig 1: Confusion Matrix for Classification with LightGBM Model

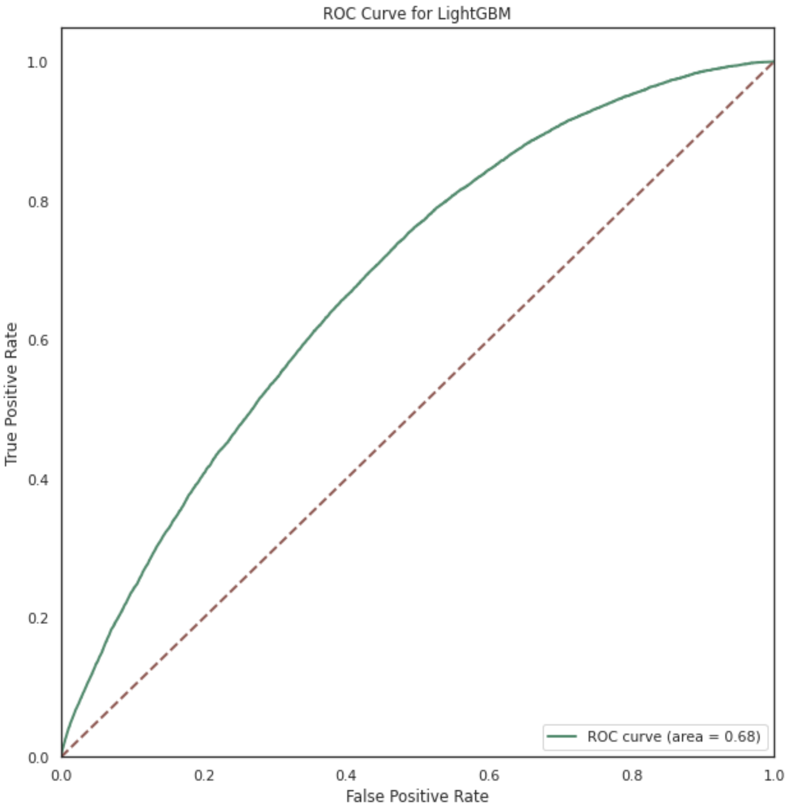

Fig 2: ROC Curve for the tuned LightGBM model

5. Key Performance Indicator Analysis

To further explain the model and provide actionable insights to Humana. It is essential to look at the feature importance. Thus, we extracted the top 50 important features (assessed by the LightGBM model) and plotted them.

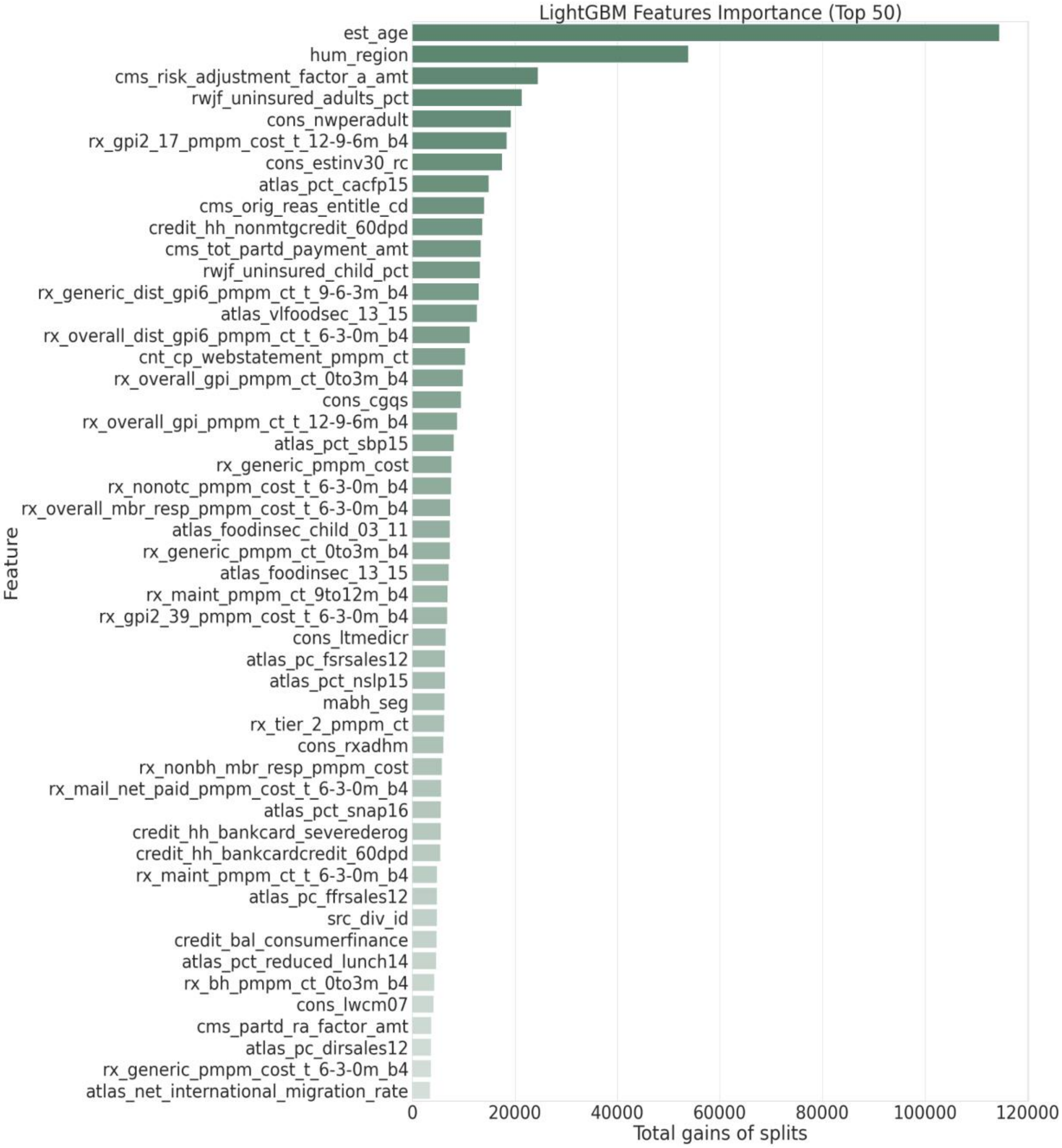

Fig 3: LightGBM Top 50 Important Features

The above features fell under a variety of categories, including demographics, CMS, medical claims, credit data, etc. Since these categories were substantially representative in our top 10 features. We will further analyze these features and make recommendations accordingly. Below we attached the definitions of the10 features for reference, which we use for further recommendations.

Feature	**Meaning**
est_age	Member age
hum_region	Member geographic information
cms_risk_adjustment_factor_a_amt	Risk adjustment Factor A Amount
cons_nwperadult	Net Worth Per Adult
rx_gpi2_17_pmpm_cost_t_12-9-6m_b4	The trend of cost per month of prescription-related to VACCINES drugs in the past sixth to ninth month versus ninth to twelfth month prior to the score date
cons_estinv30_rc	Estimated Household Investable Assets Recoded
credit_hh_nonmtgcredit_60dpd	Percent of Non-Mortgage Loan Accounts 60+ Days Past Due
cons_cgqs	Census Geo-unit Quality Score

Table 3: Top 10 Features Explanation

6. Recommendations and Managerial Implications

The LightGBM model outputted a list of features significantly contributing to the hesitance score. The following section will interpret these features and develop solutions to boost COVID vaccination accordingly. Eight were taken out among the top fifty most important features and put into three categories: demographics, health, and wealth factors.

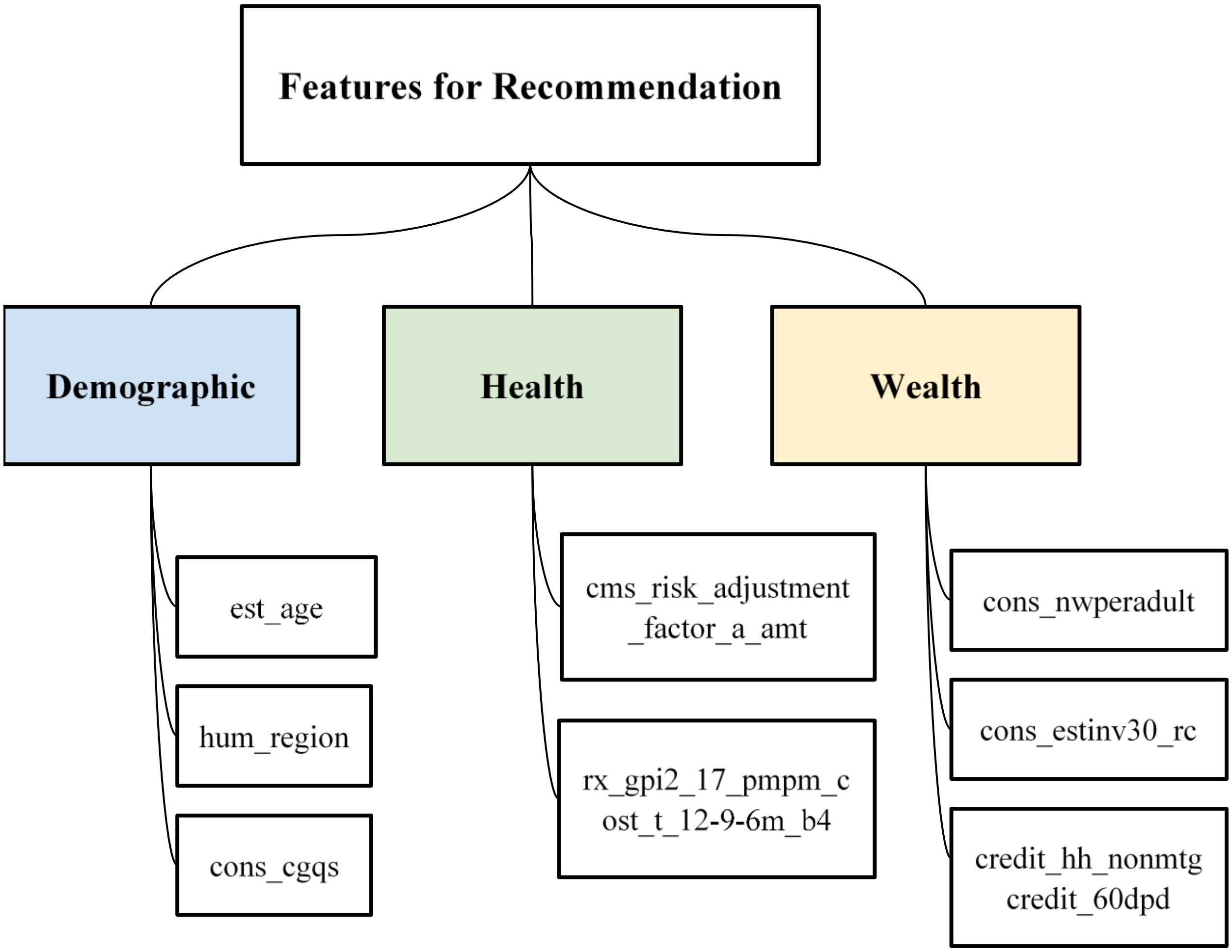
Features for Recommendation
Demographic
Health
Wealth
est_age
hum_region
cons_cgqs
cms_risk_adjustment_factor_a_amt
rx_gpi2_17_pmpm_cost_t_12-9-6m_b4
cons_nwperadult
cons_estinv30_rc
credit_hh_nonmtg credit_60dpd

Fig 4: Features for Recommendation

6.1 Recommendations based on demographic features

6.1.1 Age affects one's attitude toward COVID vaccines: young people are more hesitant to COVID vaccination

Our model suggests the highest contribution comes from the age factor. Comparing the vaccination rate among different age groups, the younger groups stand out with a relatively low vaccination rate. Specifically, the vaccination rate of people aged between twenty and forty years old was only 6.36%, while that of people aged over eighty was around 22.3%. There was a huge difference between people in different age groups, which required Humana to adopt tailored strategies to communicate with them.

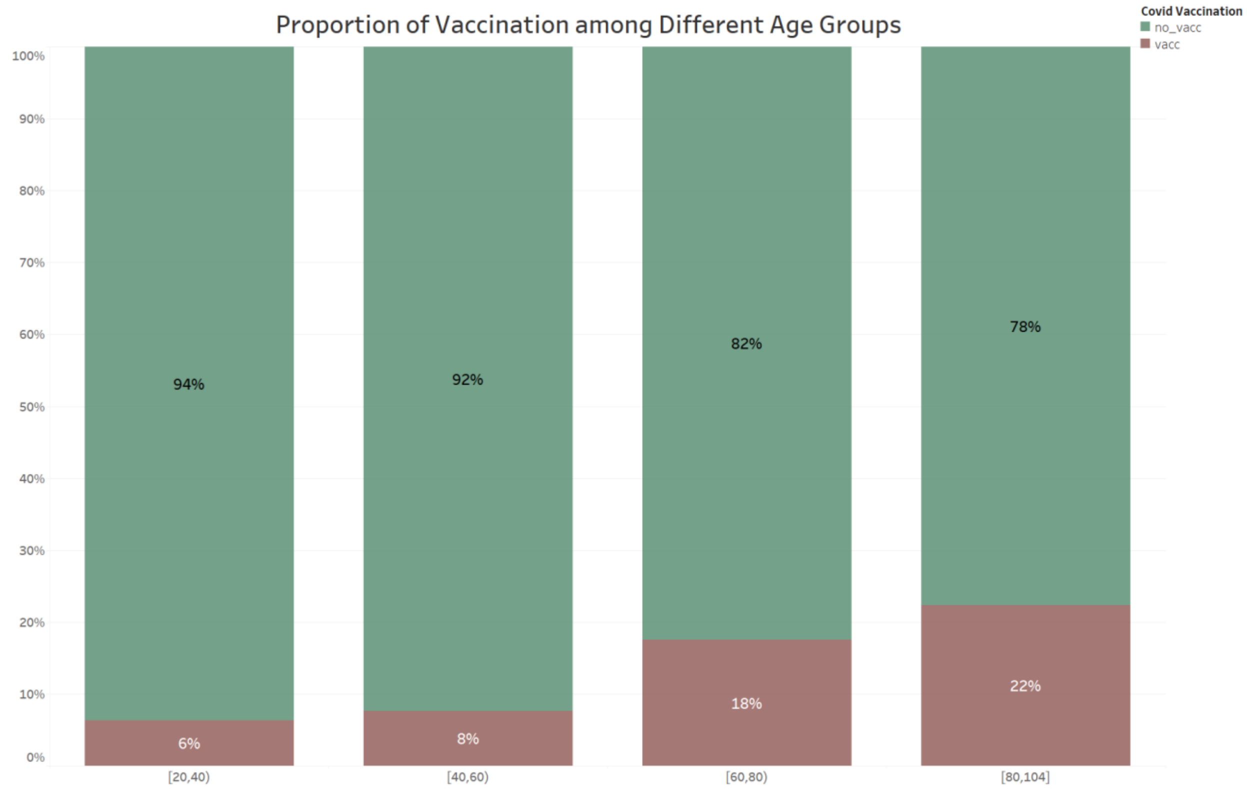

Fig 5: Proportion of Vaccination among Different Age Groups

Several reasons explain why young people are more likely to resist vaccination. Young people tend to overlook the harmful effects of COVID and not take COVID seriously. It is widely acknowledged that young people are generally less vulnerable to COVID with a lower hospitalization, severity, and death rate. However, there are still a large number of young folks infected and spreading the virus around. Lacking sufficient understanding of the pandemic makes young people unaware of the harm to themselves and others. What's worse, young people are fed with much inaccurate information on the Internet, further distorting one's understanding of COVID and discouraging them from getting vaccinated.

How to encourage young people to vaccinate?

- *Launch a marketing campaign on mainstream social media to broadcast scientific vaccine information*
- *Partner with influencers to enhance the propaganda effects*

To target the younger generation, Humana must aim at where they are. Most young people are deep users of social media, so we recommend Humana to advertise on mainstream social media like Twitter, Tiktok, and Instagram, emphasizing the benefit of COVID vaccination. Also, Humana can employ an influencer marketing strategy, inviting influencers on social media to endorse COVID vaccination campaigns. By partnering with influencers, Humana can help broadcast scientific vaccine information verified by CDC or other authoritative medical agencies, push credible vaccine news, and clarify the misinformation about the covid vaccine. It is essential to let Humana's young members realize the importance of vaccination as well as their individual responsibility to help stop the pandemic.

6.1.2 Regional differences also affect COVID vaccination: The South tends to have a higher hesitation to COVID vaccination

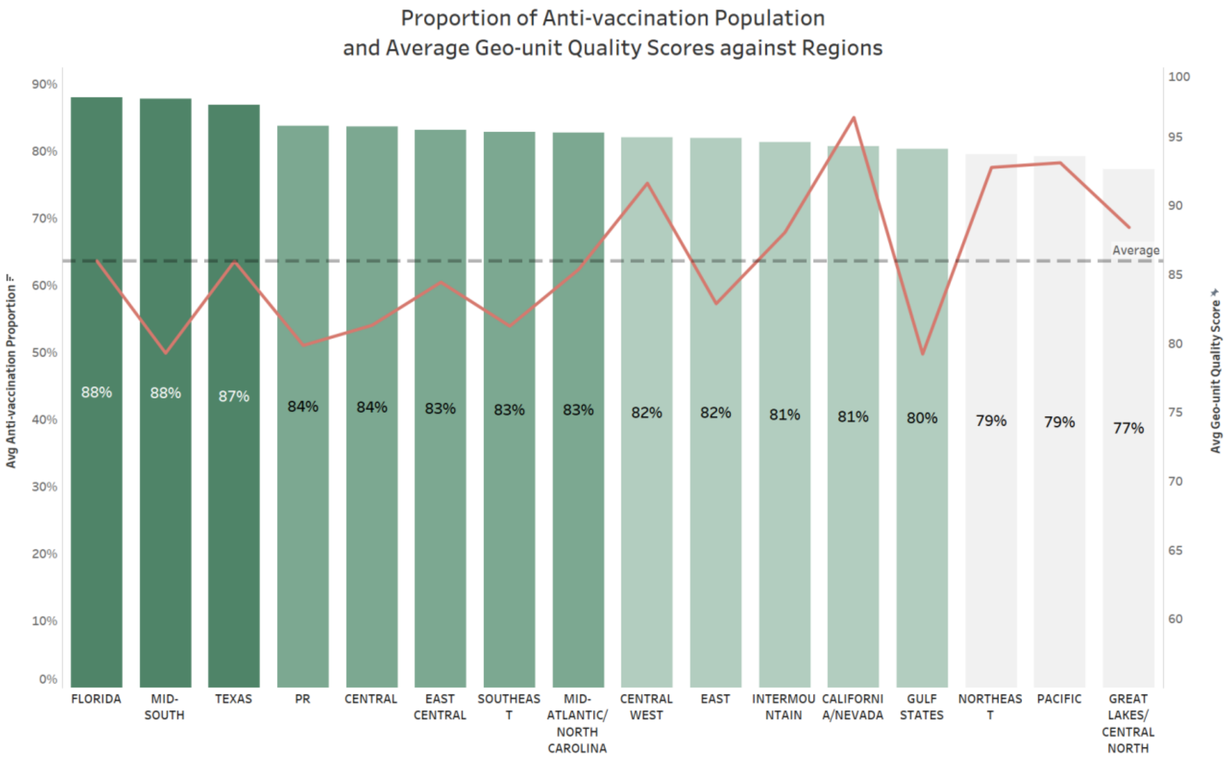

Fig 6: Proportion of Anti-vaccination Population and Average Geo-unit Quality Scores against Regions

The regional difference is a complex issue caused by many factors. We found that people residing in the South tend to be more reluctant to get vaccinated, particularly in Florida, Texas, and some Mid-South states where the non-vaccination rate was larger than 85%. The trend in the above graph also conforms with the CDC data. By Oct 10, 2021, the fully vaccinated rate of nine states in the mid-south region was lower than 50%, not only lower than the national rate of 56% but also much lower than that of the west and northeast states (Mayo Foundation, 2021). Conservatism and partisan divide may explain the gap. Most of the states with a lower vaccine rate support the Republican Party. Several Republican lawmakers and governors still don’t actively support vaccination and anti-vaccine voices are frequently heard on conservative media, which results in vaccination resistance among a large portion of Republican supporters. Also, different education levels can account for the regional discrepancy. People from mid-south states have lower average education attainment (Least educated states, 2021) and tend to be more resistant to vaccination. Education levels affect one's information source and perception of science, less educated people are more likely to be deluded by conspiracy and disinformation.

How to boost vaccination in the South?

- *Collaborate with local traditional media to broadcast the importance of vaccination*
- *Partner with local clinics and pharmacies to offer cash rewards to vaccinated members*

Most residents in these regions obtain information mainly from local traditional media like newspapers or radios. We suggest Humana advertise on these channels to broadcast the importance of vaccination. Given the conservatism in the South, vaccine incentives can be a good approach to boost vaccination in a short period of time. We recommend Humana offer discounts on insurance fees or a small number of cash rewards as incentives to its eligible members who chose to get vaccinated.

6.2 Recommendations based on health factors

6.2.1 Members with higher RAF amount are more hesitant towards COVID vaccines

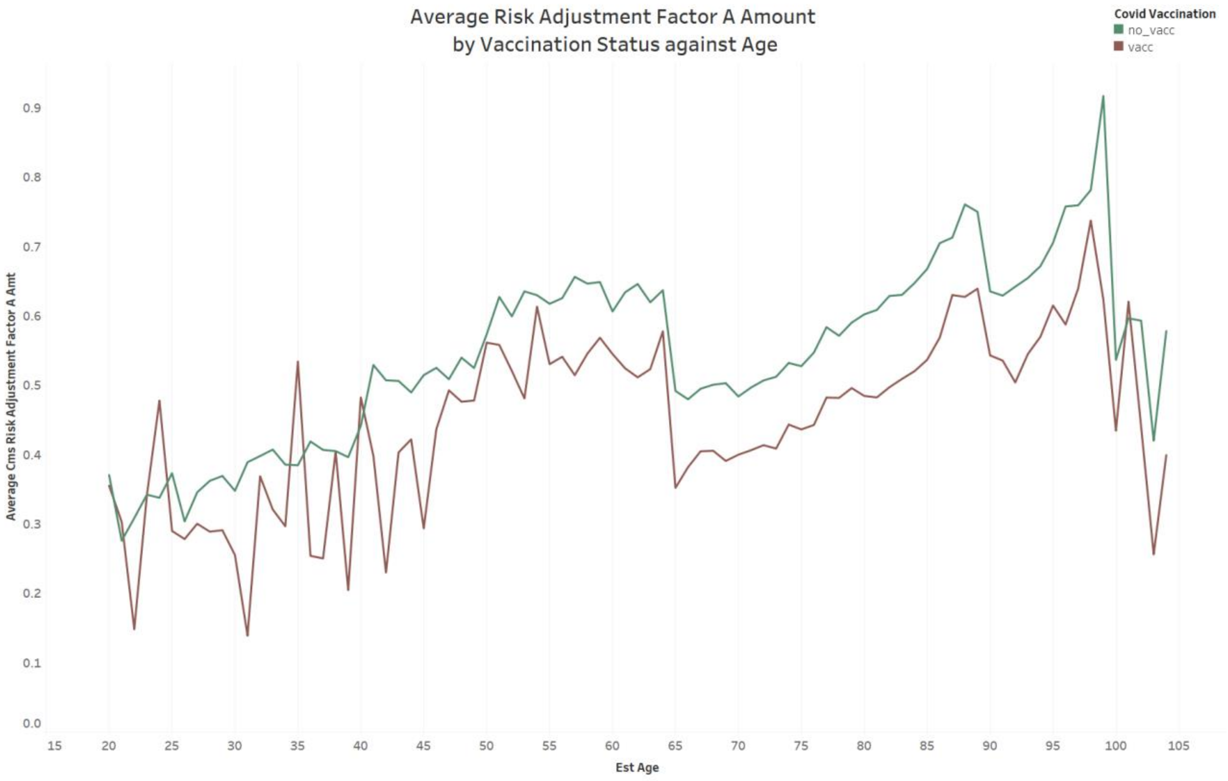

Fig 7: Average Risk Adjustment Factor A Amount by Vaccination Status Against Age

RAF score, or risk adjustment factor score, is a medical risk adjustment model used by the Centers for Medicare & Medicaid Services (CMS) and insurance companies to represent a patient’s health status. A higher RAF score indicates a higher expenditure on health.

6.2.2 Members without vaccine drugs activities are more hesitant towards COVID vaccines

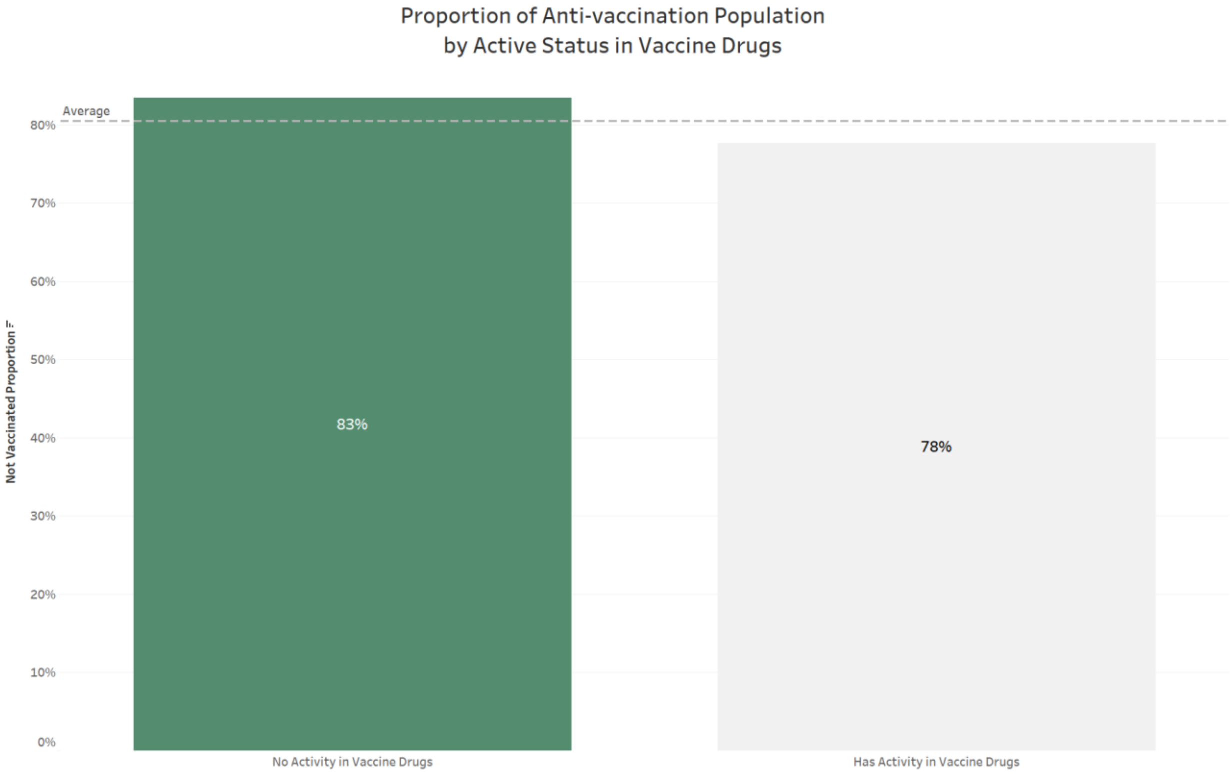

Fig 8: Proportion of Anti-vaccination Population by Active Status in Vaccine Drugs

Based on feature descriptions, trends including “Inc”, “Dec”, “New”, “Resolved”, and “No Change” can all be categorized as “Has Activity” while the rest excluding missing values is “No activity”.

6.2.3 Both health conditions and prior vaccine activities relate to one’s vaccination attitude

The line chart above shows that people with a higher RAF score are more hesitant to COVID vaccines. The RAF score relates to one’s health status, and a higher RAF score indicates higher expenditure on health. We found less healthy members are more resistant to the covid vaccine.

Given the unsatisfactory health conditions, these members could be more concerned about the side effects and unexpected harm caused by the vaccines. Moreover, the second bar chart tells that member with prior activities related to vaccination tend to accept the COVID vaccines more.

How can Humana encourage vaccination considering these health factors?

- *Send out emails/brochures introducing the safety and effectiveness of vaccines*

We recommend Humana to convince its members by showing accurate data statistics including the clinical trials results and billions of real-world data about COVID vaccines, illustrating that the rate of serious side effects is significantly low and the rare death cases are verified to be unrelated to the covid vaccine. What Humana wants to convey to the MAPD members is that, although there could be potential risks of getting vaccinated, the gains still greatly outweigh the losses.

6.3 Recommendations based on wealth factors

6.3.1 Low-income populations hesitate more toward the vaccines: Make vaccination more accessible to them

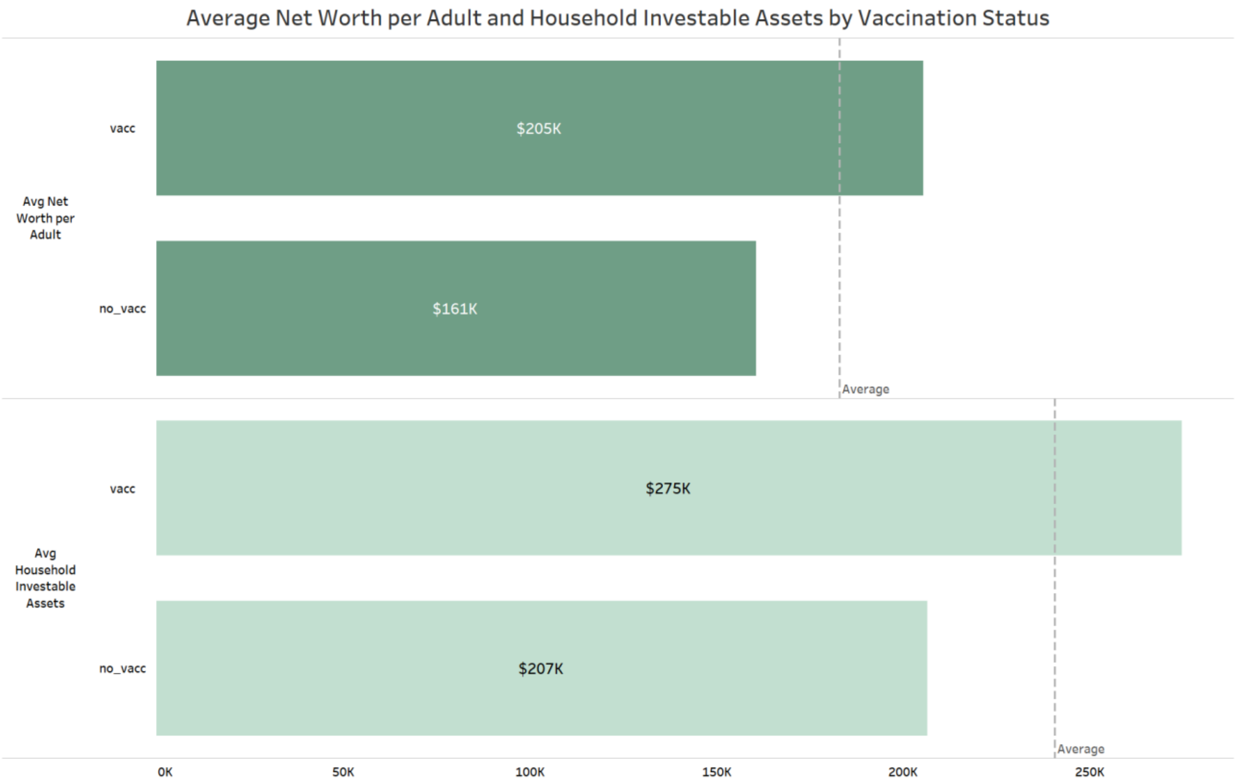

Fig 9: Average Net Worth per Adult and Household Investable Assets by Vaccination Status

How can Humana encourage vaccination considering these wealth factors?

- *Provide 24/7 vaccination services to low-income members*
- *Strengthen on-site marketing to enhance the awareness of COVID vaccination through community volunteers*

Wealth status affects people’s vaccination status. People with lower net worth and household investable assets are more hesitant to vaccinate. Wealth is correlated with many other factors like education, race, health and living areas, etc. Here we focus on the low-income group who are not aware of vaccination or who are not able to vaccinate conveniently. Many low-income people work in the service industry, where they have irregular work hours contradicting most vaccine locations' operation hours. We recommend Humana cooperate with the government or local medical agency to set up a few 24/7 vaccination sites nearby the low-income communities with walk-in services. Also, Humana can consider inviting community volunteers to promote the accessible vaccine opportunities through face-to-face communication, which will be an effective way to increase the awareness and willingness of vaccination in poor communities.

7. Conclusion

The goal of this study was to predict one’s possibility of being resistant to the covid vaccine. We filtered over 367 features down to about 320 features for use in the LightGBM model. The model achieves an AUC of 0.684. After performing the feature importance analysis, we were able to divide hesitant members into several sub-segments based on age, region, health, and wealth. Then we provided actionable insights and potential solutions to help Humana promote vaccines among these groups. The covid vaccine situation in the US has been evolving at an unprecedented pace in human history. For future considerations, Humana may advance with the times, adjust their marketing and publicity accordingly, and iterate on this process to refine their practice.

8. References

Adeline, S., Jin, C. H., Hurt, A., Wilburn, T., Wood, D., & Talbot, R. (2021, October 4). *Tracking coronavirus around the U.S.: See how your state is doing*. NPR. Retrieved October 10, 2021, from https://www.npr.org/sections/health-shots/2020/09/01/816707182/map-tracking-the-spread-of-the-coronavirus-in-the-u-s.

Advanced topics. Advanced Topics - LightGBM 3.3.0.99 documentation. (n.d.). Retrieved October 10, 2021, from https://lightgbm.readthedocs.io/en/latest/Advanced-Topics.html#categorical-feature-support.

Elflein, J. (2021, October 6). *U.S. covid-19 cases by day*. Statista. Retrieved October 7, 2021, from https://www.statista.com/statistics/1103185/cumulative-coronavirus-covid19-cases-number-us-by-day/.

Least educated states 2021. (n.d.). Retrieved October 10, 2021, from https://worldpopulationreview.com/state-rankings/least-educated-states.

Mayo Foundation for Medical Education and Research. (n.d.). *U.S. COVID-19 vaccine tracker: See your state's progress*. Mayo Clinic. Retrieved October 10, 2021, from https://www.mayoclinic.org/coronavirus-covid-19/vaccine-tracker.

Ritchie, H., & Mathieu, E. (2020, March 5). *Coronavirus (COVID-19) vaccinations - statistics and research*. Our World in Data. Retrieved October 10, 2021, from https://ourworldindata.org/covid-vaccinations?country=USA.

Smith, M., Heyward, G., & Kasakove, S. (2021, June 28). *Why young adults are among the biggest barriers to mass immunity*. The New York Times. Retrieved October 8, 2021, from https://www.nytimes.com/2021/06/28/us/covid-vaccine-immunity.html.

World Health Organization. (n.d.). *Coronavirus disease (COVID-19)*. World Health Organization. Retrieved October 7, 2021, from https://www.who.int/health-topics/coronavirus#tab=tab_1.