Humana-Mays Healthcare Analytics

2021 Case Competition

Determinants of COVID Vaccination Hesitation

TABLE OF CONTENTS

1. SUMMARY 3
2. CASE PREREQUISITES 4
3. DATA PREPARATION 6
   1. DATA ANALYSIS AND CLEANING
   2. DATA PRE-PROCESSING AND FEATURE SELECTION
4. DATA MODELING 9
5. ANALYSIS AND INSIGHTS 10
6. BUSINESS COMPREHENSION AND IMPACT FOR HUMANA #
   1. BUSINESS PROBLEM
   2. BUSINESS SOLUTIONS
   3. BUSINESS IMPACT
7. REFERENCES #
8. Summary

With our study, we aim to determine the groups in the community that are hesitant and/or resistant to partake in the COVID Vaccination. The reasoning behind the hesitancy needs to be analyzed based on certain factors revolving around misinformation, lack of trust in the vaccine, political, socio-economic, socio-demographic, cognitive, and culture.

This multi-factored phenomenon is not a new concept to the United States as it has observed reduced intake of the measles vaccination earlier, even after continuous efforts in successfully combating the spread of a global crisis.

Considering the factors that count for the maximum correlation with the target variable -- Covid Vaccination, we aim to analyze the root cause of hesitation and resistance among the population. The scientists developing the vaccine for an unknown source faced multiple roadblocks from the discovery to the testing and rollout phase. We need to ensure that we utilize the resources we are provided with to mitigate the consequences of this pandemic, as well as any future healthcare breakdowns we may suffer from.

Our motive was to determine a classification model targeting the groups unwilling to take the vaccine to curb the spread and further reduce the impact of the crisis. For our study, we firstly identified our target variable - Covid Vaccination (taken as 0 for not vaccinated, and 1 for vaccinated individuals). Our next step was Data Preprocessing where we cleaned and mined data required for our study to better analyze the variables provided demonstrating the trends. Once we had the data handy, we tested various model architectures until we found a suitable model as per our requirements. We finalized our approach with the Logistic Regression model that helped us achieve a Receiver Operating Characteristic (ROC) Area Under Curve (AUC) of 0.653.

This model helped us analyze our target audience and found Age, Health Insurance, Underlying Health Factors, Net Worth, Mental Health Care, and Poverty to be a major chunk of the significant variables. We have proposed solutions to predict and reduce the hesitancy among the individuals bracketing them into these predominant factors. We focus on Health Insurance and Underlying Health Issues to mitigate the resistance among the population, as well as targeting Healthcare workers to enhance the visibility of our approach.

1. Case Prerequisites

Humana, founded in 1961, has risen to #41 in Fortune 500(2021) leading the Healthcare sector with a variety of insurance practices as well as supreme health/wellness products and services. The mission is to help people achieve lifelong wellbeing. Improving and providing accessible healthcare is the crux of the community. Focusing on social and economic parameters aids in achieving the goals set for our project.

The onslaught of the Global Pandemic highlighted a vast magnitude of social and economic discrepancies all over the world affecting everyone’s day-to-day lives in the most devastating manner.

[1] Approximately half of the world’s population had no access to essential healthcare before the pandemic, which has only worsened since 2019. 50% of the students are still affected by the school closures, -30% investment has been lost in clean energy transition, and +135m are expected to be pushed to below the poverty line by 2030. To top it all, 255m (working hours) full-time jobs were lost in 2020. If this does not unite people to work towards a better future, then we may not have one left to sustain at this point.

A quick look at the above statistics provides an insight into how important it is for the world to tackle the fast-emerging reluctance towards the covid vaccination, and we aim to help Humana analyze the restrictions accompanying the vaccine.

The below statistics help us in segmenting the hesitancy towards a more targeted approach:

Among those ages 12 and older, as of October 4, 2021, 64% of Hispanic people and 61% of White people had received at least one COVID-19 vaccine dose, compared to 55% of Black people. Asian people continued to have the highest vaccination rate at 79%. The narrowing of disparities in vaccination rates among Hispanic people when limiting to the eligible population reflects that a high share of [children under age 12](https://www.kff.org/policy-watch/the-next-stage-of-covid-19-vaccine-roll-out-in-united-states-children-under-12/) who are not yet eligible for the vaccine are Hispanic. The gap in vaccination rates between Black and White people persists among the eligible population but is smaller than the gap among the total population.

A report shows that COVID-19 vaccinations may have helped prevent hundreds of thousands of new COVID-19 infections and tens of thousands of deaths among seniors. The study, which was conducted by researchers found that vaccinations were linked to a reduction of approximately 265,000 COVID-19 infections, 107,000 hospitalizations, and 39,000 deaths among Medicare beneficiaries between January and May 2021.

1. Data Preparation

The data provided for this project consisted of information about Humana members who are hesitant to the covid vaccine. It contains information on 974,842 members with 83% data of members who are not vaccinated. The dataset consists of 368 data points for each of the members which are broadly categorized into broader features like - COVID19 Vaccination status, medical claims features, Pharmacy claims features, Lab claims features, Demographic / Consumer data, Credit data features, Clinical Condition related features, CMS Member Data elements and other features. The vaccination status was present in a column named ‘covid vaccination’ which was either ‘yes’ or ‘no’ depending on an individual’s vaccination status.

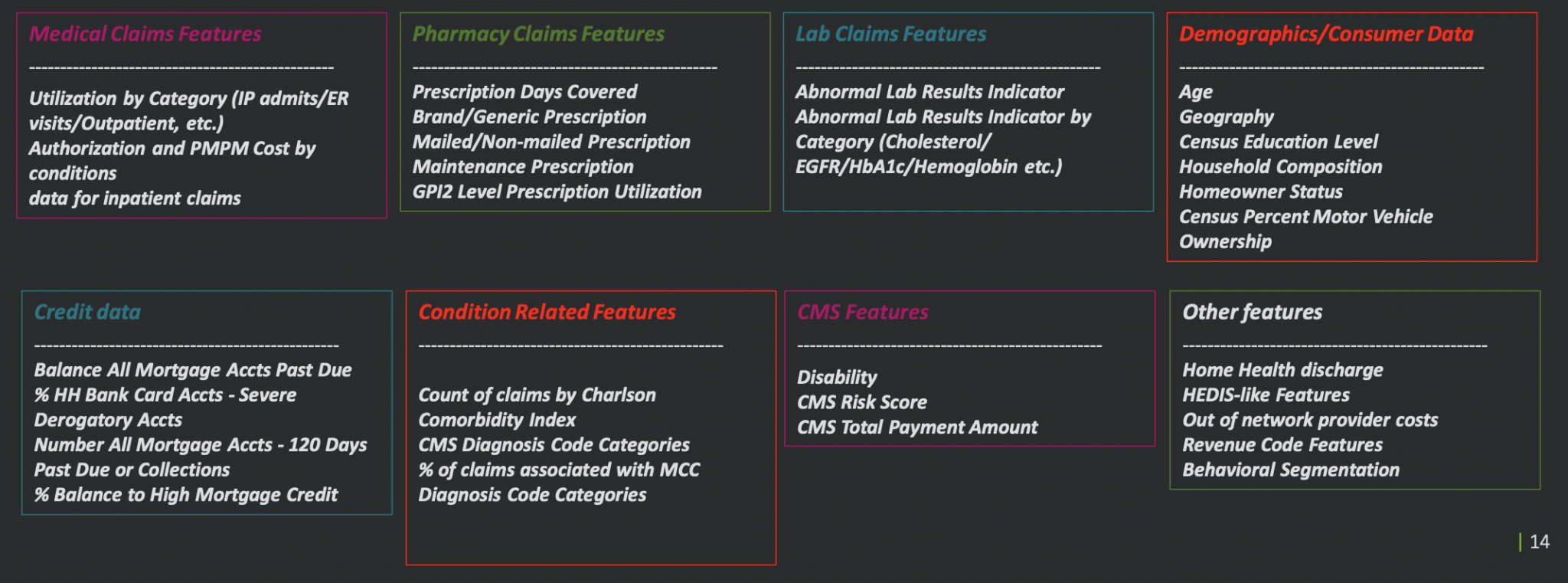


image source: Humana Kick-Off Slides

3.1 Data Analysis and Cleaning

We skimmed through the entire dataset to get a sense of the data types of individual features - then split those into Numeric and Categorical features respectively.

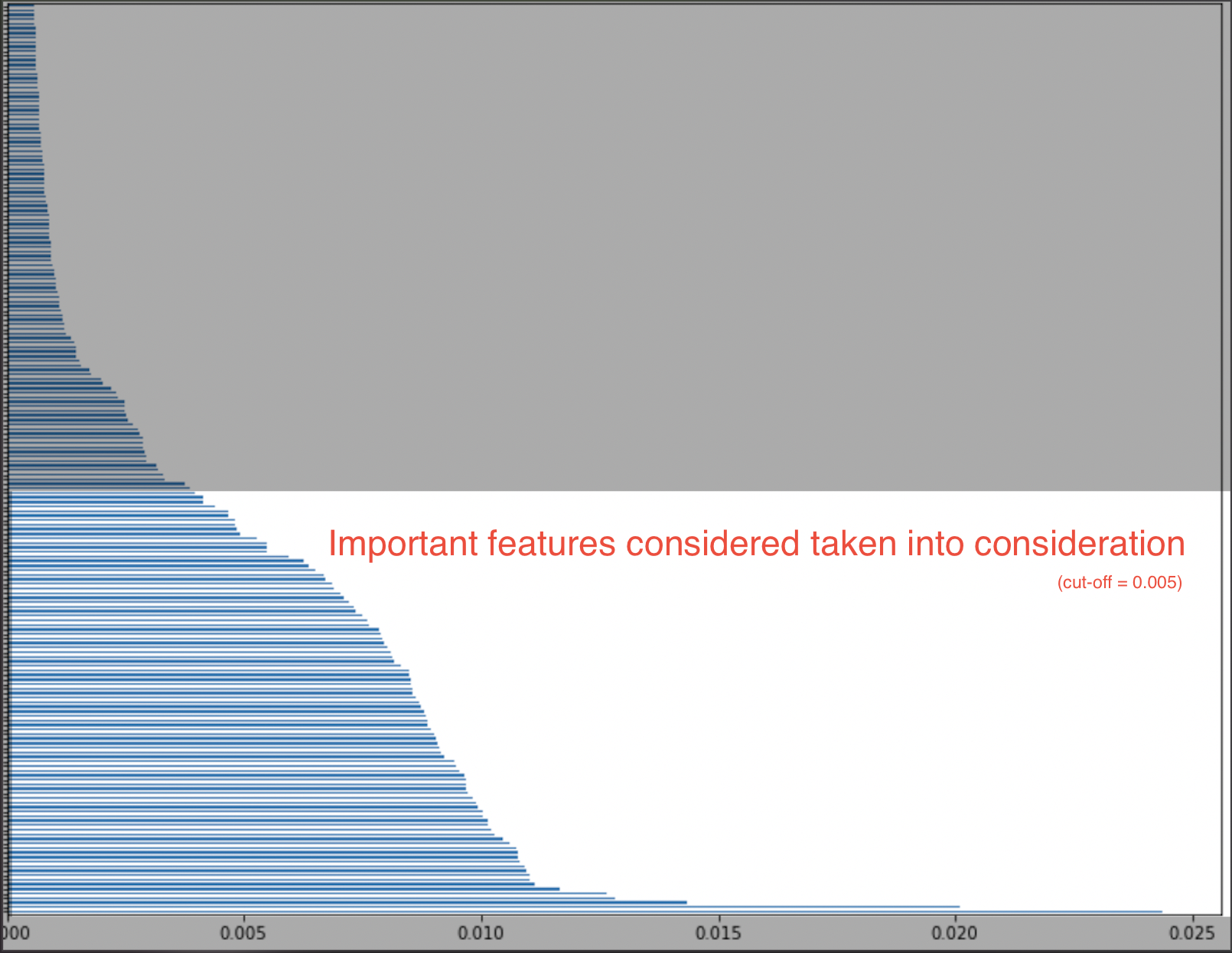
Numeric Data -> Out of 368 columns, 252 contained numeric data straight up and an additional 61 columns that were detected as categorical but were in fact numbers. We had to clean the noise in the data and ended up with 313 numeric features.

Categorical Data -> There were 55 columns that had categorical data including sex, race, language, and features related to medical claims.

We filtered out non-contributing features using filtering and regularization techniques to reduce the dimensionality of the model. We were also careful in this process because the target variable was highly imbalanced (83:17).

3.2. Data Pre-Processing and Feature Selection

With a plethora of features on our plate, we decided to come up with a strategy to select only the top contributing features. For this, we had to first engineer the existing features by filtering rows, imputing missing data, scaling values using Standard Scaler, detecting multicollinearity, and further reducing cardinality, dummifying categorical variables and finally using DecisionTreeClassifier to obtain the most important features -

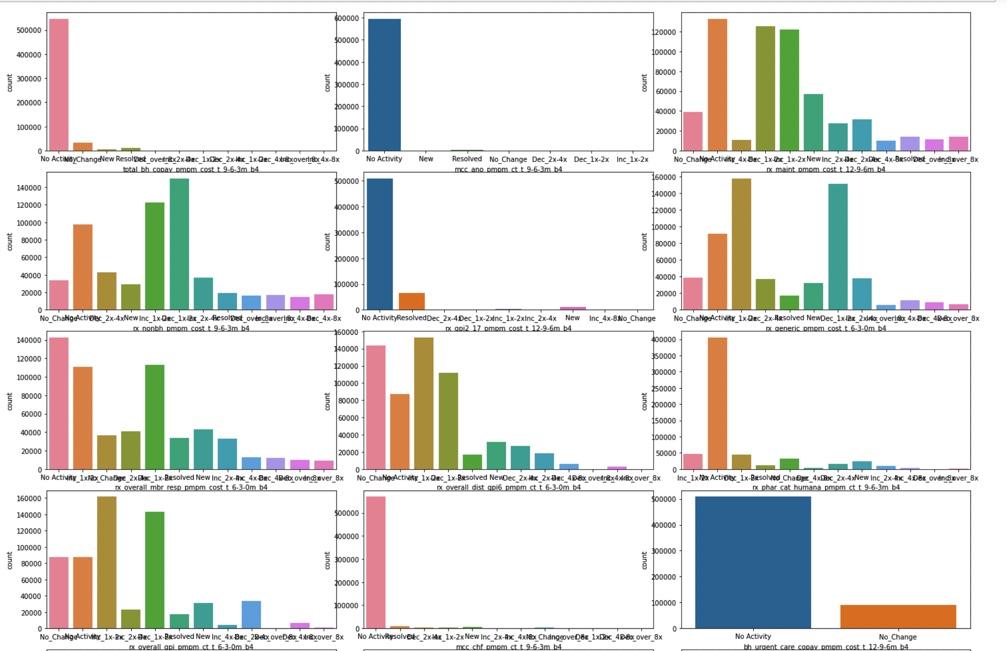


1. Data Modeling
2. Analysis and Insights

We have run a logistic regression model to predict the chances of a member going for a covid jab. And to arrive at the features contributing towards the chances, we have zeroed in on features that cumulatively have a high correlation with the target variable.

For categorical variables we ran a test of variance to determine whether columns had significant spread and we soon realized that a major chunk of those columns had just two values and it can be verified from the countplots below. We dropped columns with just one prominent entry across all rows.

A few categorical variables dropped:



After multiple rounds of wrangling and cleansing, we could establish that these 30 features were the ones contributing most towards a person going for a covid vaccine shot.

'est\_age --> Member age {calculated using est\_bday, relative to score/index date} --> ',

'rwjf\_uninsured\_adults\_pct --> Clinical Care - Percentage of adults under age 65 without health insurance --> ',

'cons\_lwcm10 --> The probability of the individual not exercising at all --> ',

'credit\_num\_autobank\_new --> Number Auto Bank Loan Accts New w/in 12 months --> ',

'cons\_lwcm07 --> The probability of the individual being less likely to use doctor/physician as a primary source for medical information --> ',

'credit\_num\_agencyfirstmtg --> Number Agency 1st Mortgage Accts --> ',

'cons\_cwht --> Percent White Collar Employed --> ',

'credit\_bal\_bankcard\_severederog --> Balance Bank Card Accts - Severe Derogatory Accts --> ',

'credit\_num\_consumerfinance\_new --> Number Consumer Finance Accts New w/in 12 months --> ',

'credit\_bal\_nonmtgcredit\_60dpd --> Balance Non-Mortgage Loan Accts 60+ Days Past Due --> ',

'cons\_nwperadult --> Net Worth Per Adult --> ',

'atlas\_netmigrationrate1016 --> Net Migration 2010-2016 --> ',

'cms\_tot\_partd\_payment\_amt --> Total Part D Payment Amount --> ',

'cons\_cgqs --> Census Geo-unit Quality Score --> ',

'atlas\_percapitainc --> Per capita Income in the past 12 months 2014-2018 --> ',

'atlas\_fsrpth14 --> Full-service restaurants/1,000 pop --> ',

'atlas\_pc\_dirsales12 --> Direct farm sales per capita --> ',

'atlas\_veg\_acrespth12 --> Vegetable acres harvested/1,000 pop --> ',

'credit\_num\_autofinance --> Number Auto Finance Loan Accts --> ',

'rx\_overall\_mbr\_resp\_pmpm\_cost --> member responsibilty cost per month for prescriptions in the past one year --> ',

'rwjf\_men\_hlth\_prov\_ratio --> Clinical Care - Ratio of population to mental health providers --> ',

'atlas\_pct\_laccess\_nhna15 --> American Indian or Alaska Native, low access to store --> ',

'atlas\_pc\_wic\_redemp12 --> WIC redemptions per capita --> ',

'atlas\_pct\_laccess\_white15 --> White, low access to store --> ',

'atlas\_grocpth14 --> Grocery stores/1,000 pop --> ',

'atlas\_recfacpth14 --> Recreation & fitness facilities/1,000 pop --> ',

'rwjf\_resident\_seg\_black\_inx --> Social and Economic Factors - Residential segregation - black/white --> ',

'atlas\_pct\_laccess\_hisp15 --> Hispanic ethnicity, low access to store --> ',

'cms\_partd\_ra\_factor\_amt --> Medicare Part D Risk Adjustment Factor Amount --> ',

'cons\_n2pmr --> Census Percent Married

Doing a deep dive into the most relevant items from the dataset, we could see that member age is a factor that hugely influences a person’s decision to go for vaccine shots. We could see a positive correlation between the member age and likelihood to go for covid jabs, so the older the person the more likely he/she takes the jab.

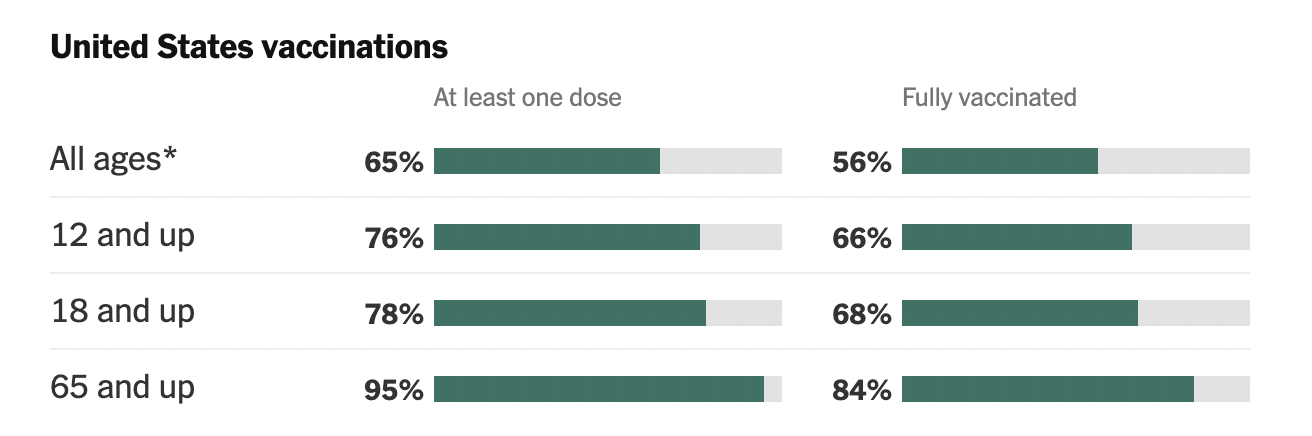
Another contributing factor has been the population of folks under 65 and without a health insurance plan, such folks are more likely to go for vaccines as compared to the ones with health insurance plans.

Folks who are less likely to exercise are more likely to take the covid jab; there exists a negative correlation between the target variable and the probability of a person not exercising at all.

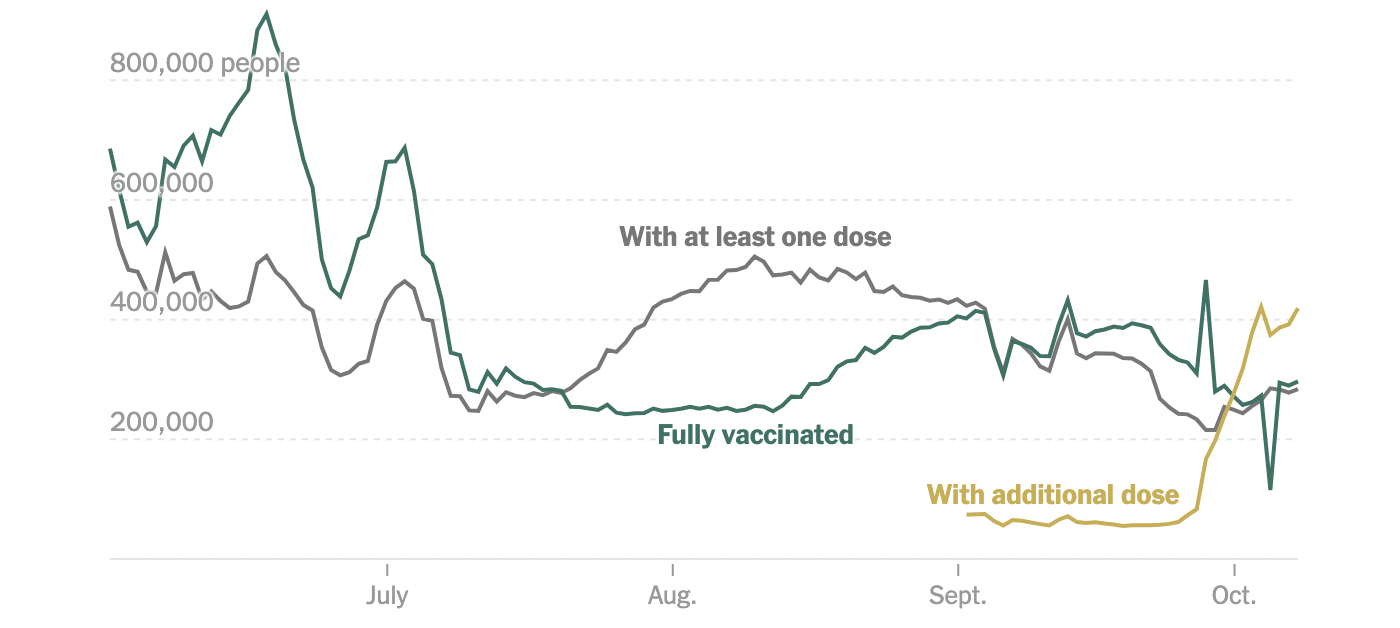
1. Business Comprehension and Impact for Humana

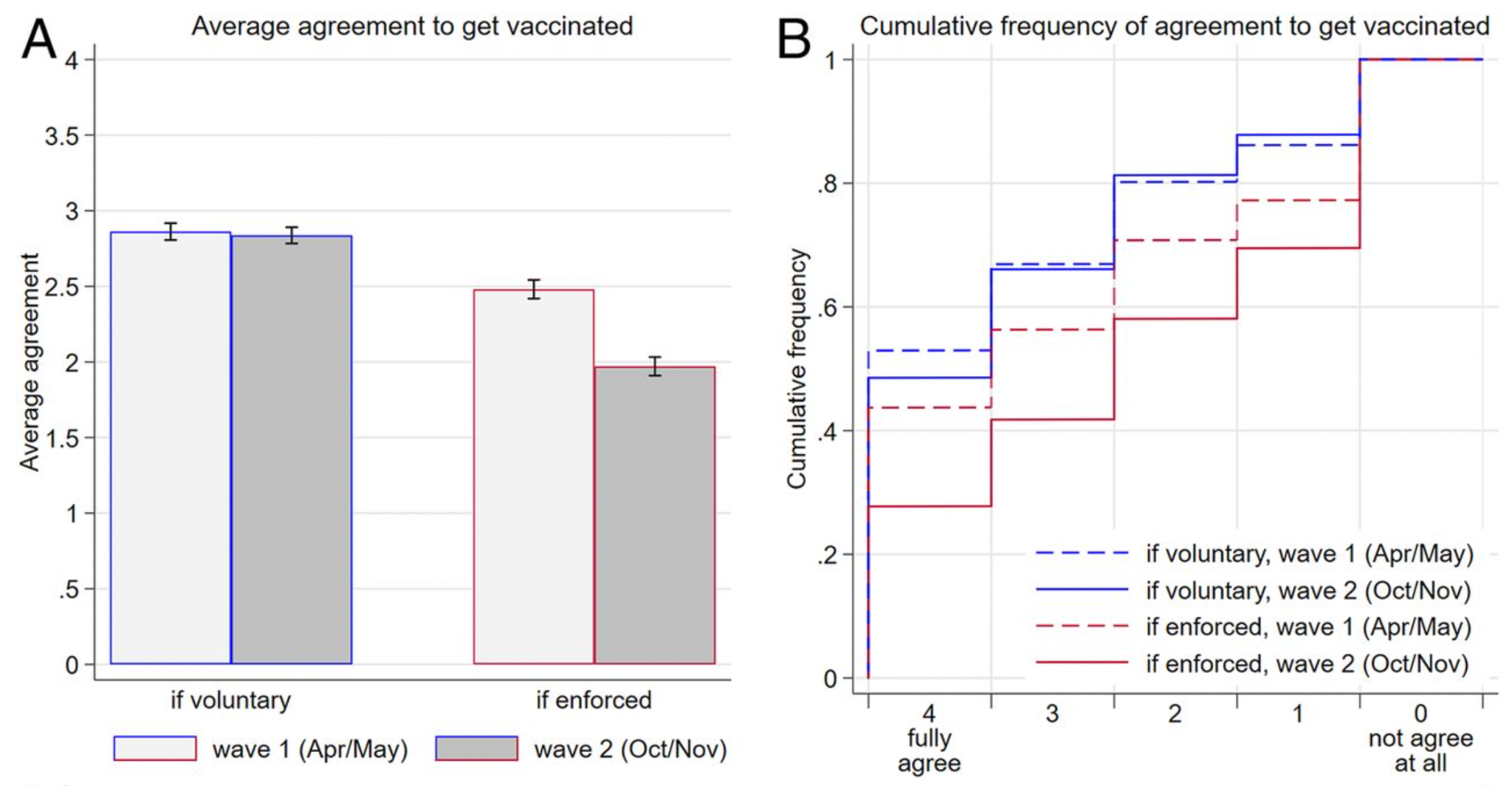
6.1 Business Problem

[3] We have noticed multiple Anti-Vaccine protestors in the United States shutting down major COVID vaccine sites citing COVID as a Scam fighting to end the lockdown claiming it as a loss of freedom. In the United States, 41% of those surveyed in March said they were “very concerned” that they “might be required to get the COVID-19 vaccine even if [they] don’t want to,” and another 21% were “somewhat concerned”. This requires a boost in the vaccine confidence structure, which enables herd immunization along with healthcare worker motivation to achieve total immunity from the virus.



Targeting these age groups to complete their dosage requirements is the priority, especially for children and teenagers.





Reduced support for enforced vaccination. (*A*) Average agreement to get vaccinated if it is voluntary or enforced in the two waves of the survey (in Likert scale units). Error bars represent 95% CI. (*B*) Cumulative distributions of agreement in case of enforced versus voluntary vaccination for the two waves of the survey. For example, the dashed and solid red lines show that 44% and 28% of respondents fully agreed to get vaccinated in case of enforcement in the first and second waves of the survey, respectively. The sum of those expressing either agreement level 3 or 4 under enforcement amounts to 56% in wave 1 and 42% in wave 2. Opposition to enforcement (levels 0 and 1) was expressed by 29% in wave 1 and 42% in wave 2 (1–0.71 and 1–0.58, respectively, that is, the final two steps in the graph).

6.2 Business Solutions

6.2.1 Building Trusted healthcare for Bracketed Age Groups

* As per our analysis, most of the unvaccinated sections lie in the Age Groups below 18 and above 65.
* Providing digital solutions backed with AI to support senior citizens would assist in timely and trusted healthcare all round the clock. By building a web and mobile application to support our strategies enforced with assistance by healthcare workers by appointment service or customer service, we can ensure reduced misinformation and guided support.
* School-located vaccination sites can aid in removing the logistics and transportation barrier for children and teenagers.
* Home-delivered vaccination can reduce the logistic barrier for the elderly.

6.2.2 Building Trusted healthcare for the Underprivileged

* A vast section of our analysis provides us with insights to combat the barriers of privileges.
* Homeless-shelter vaccination sites can mitigate the logistic barrier for the homeless individuals and families situated in the shelters.
* Home-delivered vaccination can combat the issue of transportation for individuals and families with no vehicles to receive the vaccination.
* Providing financial incentives to the underprivileged can motivate them to reduce their resistance towards the vaccine intake. For example, providing a free meal or a coupon after getting the vaccination completed. Partnering with a successful food chain or a utility provision company would be useful in tackling this issue at hand.

6.2.3 Building Trusted healthcare for the Individuals with no HealthCare Plan

* Increasing the trusted healthcare network for individuals with no health care plan can be marked as a necessity since our analysis holds a major chunk of the population with no access to the healthcare plans.
* Humana can revamp its digital platform to build a trusted network of healthcare workers aiming at this goal by assigning vaccine ambassadors and pulling in brand ambassadors with known public faces or movie stars that have a huge fan following.
* Major establishments can pair up with Humana to provide healthcare insurance to their employees and workers (including their families) to mitigate the population with no healthcare plans.

6.2.4 Building Trusted healthcare by combating misinformation

* Partnering with brand ambassadors as mentioned above can combat this issue of mistrust and the spread of misinformation amongst the population.
* Sending out trusted fact checks and polls/graphs on the proposed digital platform can aid in reducing the spread of misinformation among the public.

6.3 Business Impact

Utilizing the recommendations of our key insights and model predictions, we can assist Humana in saving ~ $30 million annually in Reimbursements Costs. Our solutions aim to provide Humana with a Revamping of their Digital platform (Web and Mobile) to accelerate their service offerings for the aged and have better Data Security; thus gaining a competitive advantage.

1. References

[1] https://www.news-medical.net/health/How-has-the-COVID-19-Pandemic-Impacted-Global-Health.aspx

[2] https://www.kff.org/coronavirus-covid-19/issue-brief/latest-data-on-covid-19-vaccinations-by-race-ethnicity/

[3] <https://www.washingtonpost.com/nation/2021/01/30/anti-vaccine-protest-dodger-stadium/>