COVID-19 Vaccination Intent and Impact

Kate Alden, Bhargav Desai, Prateek Gulati, Ai Vi Truong, Sze Ki (Zalon) Wong

December 2, 2021

Table of Contents

Executive Summary 3

Introduction 4

Business Problem 4

COVID-19 Pandemic 4

Data Overview 5

Descriptive Analysis: COVID-19 Impact 8

Initial Data Analysis 11

Data Preparation and Pre-Processing 12

Cluster Analysis 13

Recommendations 14

Predictive Analysis: Vaccination Intention 16

Initial Data Analysis 16

Data Preparation and Pre-Processing 17

Predictive Modeling 18

Recommendations 19

References 21

Appendix 24

Executive Summary

As consultants for PwC, we used the U.S. Census Bureau's Household Pulse Survey (HPS) to uncover insights about the factors that impacted United States residents during the pandemic, as well as the factors that influence a person's decision to get vaccinated against COVID-19. Our goal for analysis is to assist PwC with accelerating vaccination programs and adopting policies related to COVID-19, by answering the following questions:

- Which characteristics of COVID-19 impact are the most significant and why?
- What characteristics are important in influencing vaccination in the United States and why?

Understanding that the pandemic affected the people of the United States in many ways, we investigated the impacts on mental health, education, employment, income, and age to understand demographics of those most impacted in these ways. Combining our knowledge with the HPS questionnaire's survey questions on topics of employment, physical and mental health, health insurance and access, housing, spending and economic payments, food sufficiency and security, education disruptions, and vaccination intention, we planned to build a descriptive model to understand how these aspects impacted survey respondents, and a predictive model to understand what leads a person to get vaccinated.

We built a clustering analysis model on a randomly sampled dataset from the HPS survey to understand which demographic parameters fell into each group. The insights we gathered from cluster analysis were in line with our research and expectations on how mental health, employment, income, and housing were negatively impacted by COVID-19.

To build our predictive model, we focused on a "get vaccine" variable to predict whether a survey respondent intends to receive the COVID-19 vaccination or not. The Decision Tree classification model produced the best results, with education, age, telework, income and mental health proving to be the most impactful in determining whether an individual would pursue vaccination.

To address COVID-19 impact, we built recommendations to benefit younger people who have been impacted by job loss during the pandemic, benefit females who experienced childcare difficulties, and help those who faced burdens that are housing-related. To address COVID-19 vaccination intent, we built recommendations to build company policies that serve to entice people to get vaccinated by offering compensation, making information on the importance of getting vaccinated available for quick education, and offering mental health services to reduce vaccine hesitancy. If implemented, we believe these programs will have a positive impact on the workforce and increase the number of people vaccinated, to bring our world back to a relative normal post-COVID-19.

Introduction

Business Problem

COVID-19 came as a surprise to the world and changed the course of 2020 and the years to follow. From being declared a pandemic in March 2020 to later creating effective vaccines in fall and winter of 2020, everything about COVID-19 has drastically changed the way United States residents look at life. The challenges that came with COVID-19 continue to impact us today, including the challenge of prompting residents to get vaccinated. Additionally, we are still beginning to understand the impacts that COVID-19 had on the population. While COVID-19 continues to be a global pandemic, the focus of this report will remain on individuals of the United States.

Mental Health, Employment and Education are among the many aspects of life that were significantly affected during the pandemic. People across the globe have been forced to stay home more often as a result and practice quarantine. Unfortunately, the loneliness that quarantine has created has resulted in an increase in the number of depression and anxiety cases among teens and adults (Fernandez). On the topic of employment, the COVID-19 pandemic produced high unemployment rates and furloughs while industries slowed production. Changes in work policies were also enacted, and remote or work from home became the new normal over the course of the pandemic. Education systems also experienced major changes, as strong policies and precautions were placed to cope with existing COVID-19 cases and to prevent new cases.

Although vaccinations are very helpful in combating this pandemic, it will be a while before the world can return to its normal state of the world, we understood pre-pandemic. One important factor that is preventing a return to normalcy is continued vaccine hesitancy. We find it crucial to address the vaccination intent of people to understand what aspects make them hesitant to pursue it and plan to make more effective campaigns. Another importance of analyzing people’s intent to not be vaccinated is to adopt a point of view on COVID-19 and how it currently impacts PwC and its affiliates.

Currently, vaccine hesitancy is disrupting the ongoing vaccination campaigns across the United States. The challenges related to the vaccine include but are not limited to the availability of the vaccine, lack of awareness on the vaccine, and mistrust among members of the public based on preconceived notions and cultures.

The global pandemic is far from over with new variants and outbreaks emerging in many countries. The impact of COVID-19 cannot be taken lightly, as this pandemic has also shed light on the ongoing healthcare crisis across the globe. It is crucial to be aware of the aftereffects and impact of COVID-19, not only to monitor the current state of COVID-19 cases but to prevent a future pandemic. PwC’s determination to address these problems can be related to its medium-term plans, managing the workforce and social responsibility.

COVID-19 Pandemic

First discovered in 2019 in Wuhan Hubei Province, China, the COVID-19 virus was identified as a cause of pneumonia cases, infecting people by inoculation of respiratory samples (Ciotti et al, 2020). The ultra-contagious virus quickly spread from Wuhan Hubei Province to other countries, particularly hitting countries such as China and Italy hard in the early months, with these countries reporting thousands of deaths (Ciotti et al, 2020).

Because of the contagious, deadly nature of the disease, the World Health Organization (WHO) declared it to be a pandemic on March 12, 2020 (Ciotti et al, 2020). COVID-19 hit the United States so hard that health care systems across the nation couldn't handle the influx of critically ill patients with COVID-19 along with patients seeking elective or semi-elective procedures or doctor visits (Ciotti et al, 2020). This made it so that states decided on measures to limit the spread of the disease as much as possible to reduce the number of people needing urgent care for COVID-19. After March 12, each country developed a unique response to the pandemic, with the United States developing point of views on lockdown, quarantine, and mask-wearing to prevent the spread of the disease (Ciotti et al, 2020).

Lockdown affected the people of the United States in many ways, with the shift from "normal" life leading to challenges with mental health, education, and employment, to name a few. Additionally, with 2020 being an election year, people tuned into news more frequently, listening to election updates as well as politician points of view on the development of a vaccine for COVID-19 and the reliance on vaccination to bring society back to "normal" (Ciotti et al, 2020). In our research and analysis, we will discuss the impact of COVID-19 on education, mental health, and employment, and analyze survey responses to predict vaccination intent among Americans.

Data Overview

We obtained datasets of the Household Pulse Survey (HPS), conducted by the US Census Bureau to understand how American's lives have been impacted by the coronavirus pandemic. The survey captured the social and economic effects of COVID-19 in multiple aspects including employment, education, health, childcare, and food sufficiency[1].

The questionnaire explored the social and economic effects by asking questions on the topics of employment, physical and mental health, health insurance and access, housing, spending and economic payments, food sufficiency and security, education disruptions, and vaccination intention.

The datasets acquired from survey results were from Phase 3 (weeks 22 to 27 of 2020) and Phase 3.1 (weeks 28 to 33). Table 1 shows the total number of observations and variables available in each week and the corresponding phase information.

[1] https://www.census.gov/data/experimental-data-products/household-pulse-survey.html

Table 1: Dimensions of datasets collected from phase 3 and 3.1 (HPS)

Phase	Week	Date	Number of Observations	Number of Variables
3	22	6 Jan – 18 Jan	68,348	204
	23	20 Jan - 1 Feb	80,567	204
	24	3 Feb - 15 Feb	77,122	204
	25	17 Feb - 1 Mar	77,788	204
	26	3 Mar - 15 Mar	78,306	204
	27	17 Mar - 29 Mar	77,104	204
3.1	28	14 Apr - 26 Apr	68,913	239
	29	28 Apr - 10 May	78,467	239
	30	12 May - 24 May	72,897	239
	31	26 May - 7 Jun	70,854	239
	32	9 Jun - 21 Jun	68,067	239
	33	23 Jun - 5 Jul	66,262	239

In Phase 3, there are a total of 204 variables and 459,235 observations. In phase 3.1, there are a total of 425,460 observations and 239 variables. In terms of quality, both datasets contain many missing values. As shown, there are also differences in the variables including in the different phases. We recognized that numbers 19 and 44 variables in phases 3 and 3.1, respectively, were distinct variables that appeared in one but not both datasets. Also, we identified that numbers 23 and 30 variables in phases 3 and 3.1, respectively, were modified, but we were able to trace the relevant variables. Finally, we discovered a total of 15 variables with slight variations in variable names, but they have exact matching definitions.

We defined variables as either primary or secondary variables: primary variables are derived from questions that respondents must answer, such as “Have you received a COVID-19 vaccine?”; secondary variables are those generated from dependent questions that respondents may not be offered, such as “Did you receive all required doses?”. Additionally, a variable, Age, was created based on the birth year to facilitate interpretation.

We completed some data exploration and visualization to better understand the breakdown of demographic information in the dataset. In exploring male versus female, we discovered that 59.56% of respondents were female, with the average age 53, with the average age for males as 55. In analyzing the answered questions on the topic of race, we discovered that White made up 82.17% of the survey respondents, with Black as the next prominent with 7.88% of all survey respondents. Regions of the United States were well represented in the survey responses, with West and South making up slightly larger shares over Midwest and Northeast.

Across income, while 25% of respondents did not respond to this question, 38.7% of respondents reported earning an income over $100,000 annually, with the highest share being 100K-149K (18.5%), followed by 50K-74K (17.6%).

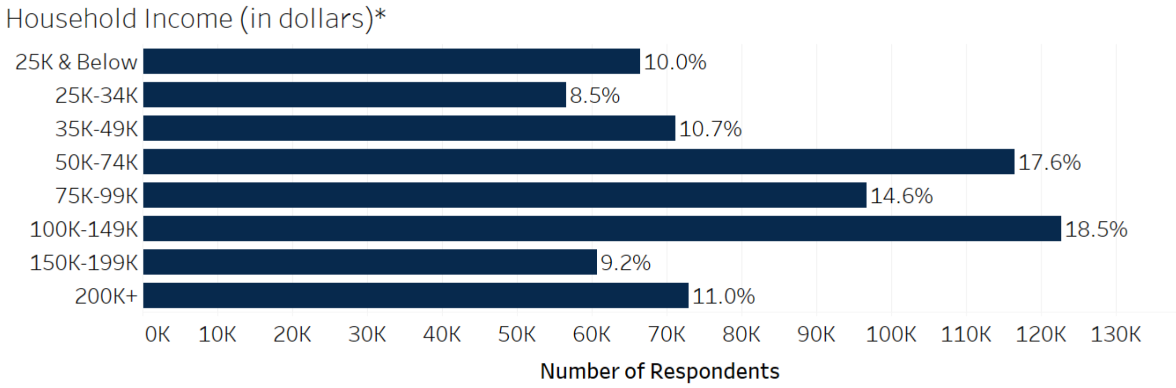

Figure 1: Household Income by Number of Survey Respondents

Across education, 86.4% of survey respondents reported that they had completed at least a high school education, and 54.5% reported they had completed a bachelor's degree or higher. The share of survey respondents who reported having less than a high school education made up the minority, with 2.1%.

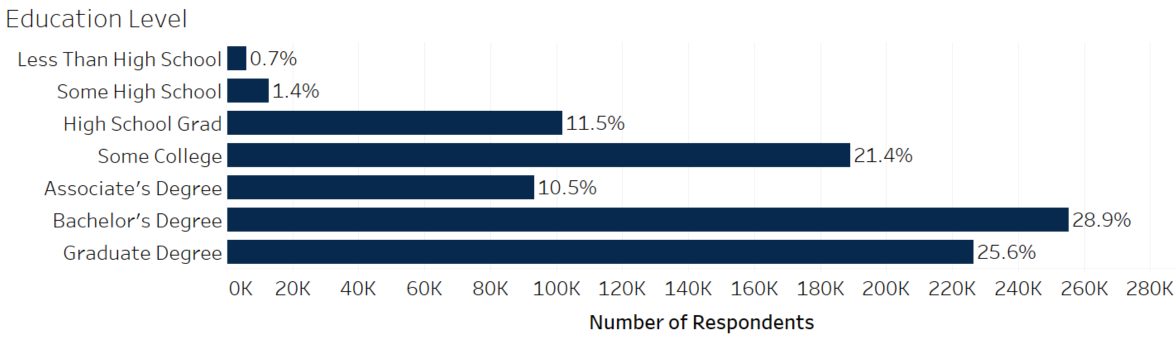

Figure 2: Education Level by Number of Survey Respondents

On the topic of COVID-19 infection, 88.1% of survey respondents confirmed they had not been infected with COVID-19 at the time of taking the survey. While the survey indicated 11.2% of respondents had contracted the disease, at the time of the survey release the US had reported an infection rate of 15%.

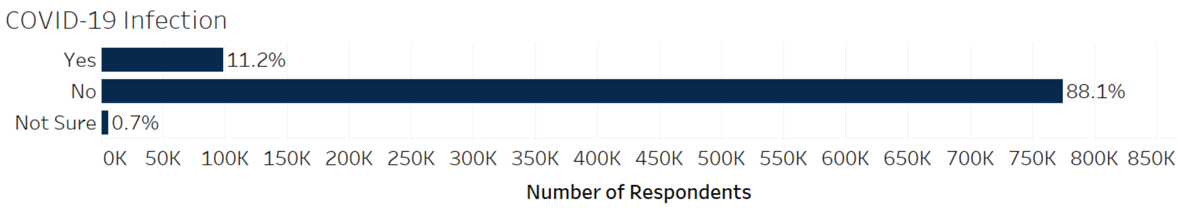

Figure 3: COVID-19 Infection Response by Number of Survey Respondents

Descriptive Analysis: COVID-19 Impact

Mental Health

The COVID-19 pandemic has had a negative effect on our mental health. Since the COVID-19 outbreak, more people have been forced to stay at home in self-isolation and quarantine to prevent the spread of COVID-19. This has caused a negative impact on a person's mental health, since they are experiencing separation from loved one, loss of freedom, boredom, and uncertainty, all of which can cause a deterioration in an individual's mental health status. It has had major effects on multiple demographic groups including children, people with disabilities, health workers, and especially the elderly, as they are most likely to experience mental health issues (Javed et al, 2020).

One of the reasons why the elderly are considered an at-risk population is because aging comes with numerous psychological, social, and environmental vulnerabilities. Social distancing is also a major cause of loneliness, increasing the risk of depression, anxiety, and other disorders. The elderly also struggle with hygiene and outside activities, such as to purchase health care supplies and attend doctor visits. Most elderly people are not comfortable with smart phones, hence they also struggled with attending virtual doctor visits, which resulted in missing their regular doctor checkups (Banerjee, 2020). The British Journal of General Practice conducted a study to identify who are most vulnerable among the elderly (Wong et al, 2020). The result showed that there were significant increases in loneliness, anxiety, and insomnia after the COVID-19 pandemic began. During the start of the pandemic, the number of missed medical appointments for this age group over a 3-month period increased from 16.5% to 22.0%. Additionally, for women who live alone and have four or more chronic health conditions, increased levels of loneliness and anxiety were shown (Wong et al, 2020).

On the other hand, the COVID-19 mental health crisis is also hitting young adults negatively. The Center of Disease Control and Prevention analyzed a survey response and reported that about 75% of respondents from 18-24 that they had one or more adverse mental or behavioral health symptoms (Fernandez, 2021). Young Adults also experienced several pandemic-related consequences, such as school closures and loss of income. During the pandemic, a larger than average share of young adults (ages 18-24) reported symptoms of anxiety and/or depressive disorder (56%). Additionally, compared to all adults, young adults are more likely to report substance use (25% vs. 13%) and suicidal thoughts (26% vs. 11%) (Panchal et al, 2021). Prior to the pandemic, young adults were already at high risk of poor mental health and substance use disorder, though many did not receive treatment (Panchal et al, 2021).

Employment

The COVID -19 pandemic shook the economy, affected the labor force, and disturbed the operation of all industries. Its impacts including unemployment, furlough, changes in mode of work or work hours were asymmetric across demographic groups, occupations, and industries. Much of the research studying the impacts of COVID-19 at the onset of the pandemic shed light on the imbalanced influence.

Uneven Supply Demand Shock Across Industries and Occupations

As the COVID-19 pandemic spread throughout the world, many governments, including the United States, mandated social distancing practices and closure of non-essential businesses as preventive measures to slow down the spread of the virus. These policies created both supply and demand shocks in many industries. The supply-side decline happened mostly because some industries or services were defined as nonessential or work-from-home was not feasible for workers to perform their duties. While demand experienced both upward and downward trends, the reasons for the changes were similar – the behavioral changes of consumers in response to COVID-19 pandemic (del Rio-Chanona et al., 2020).

Based on the supply-demand-shock framework, Del Rio-Chanona et al. (2020) estimated that around 23%of jobs would be jeopardized in the United States while total wage income would drop by 16%. More importantly, the most vulnerable groups in the workforce were hardest hit. For instance, waitresses and dishwashers had low work-from-home flexibility and were unlikely to be part of an essential industry. Occupations like these were most vulnerable to supply shocks and were likely to be low wage positions. Del Rio-Chanona et al. (2020) further concluded that compared with pre-COVID period 41%of jobs in the bottom quartile of the wage distribution were considered vulnerable while only 6%of top quartile jobs was seen as at-risk.

Urban-Rural Divide Among Low-Skilled Workers

Unevenness was not only experienced among different workers but also within low-skills workers located in different geographical locations. In the United States, it was estimated that 37% of jobs, accounting for 46% of all United States wages, can be performed entirely at home and these positions also paid more (Dingel & Neiman, 2020). Althoff et al. (2020) confirmed the urban-rural bias of COVID-19 pandemic's impacts on low-skills workers. Traditionally, large cities with denser populations specialized in high-skill services, whose workers generated job opportunities to low-skill service workers, such as restaurant workers or hairdressers, to make a living in these cities. However, as remote work was enabled by the pandemic, an exodus of skilled scalable services (SSS) [1] workers led to a diminishing need for the non-tradable consumer services provided by low-skill workers in the denser cities. Low-skill workers in urban cities therefore lost more hours of work than their rural counterparts.

Disparities in Employment States, At the Intersection of Age, Gender, Education, & Race/ Ethnicity

Moen, Pedtke & Flood (2020) examined the disparate life-course impacts caused by the COVID-19 pandemic, at the interaction of age, gender, education, and race/ethnicity in the United States. Using monthly Current Population Survey (CPS) data from the US Census Bureau, increases in unemployment and being out of the workforce for experienced by people in all age groups from January through April 2020.

At the intersection of age and gender, while the largest increases in unemployment were seen among emerging adults in their 20's, young women were particularly at risk. Unemployment and NILF-other (not in the labor force for reasons other than retirement and disability) rate of women in their 20's increased by 10.2% and 6.6%, respectively. For comparison, the respective rate changes for women at their 30's were only 6.8% and 3.4% (Moen, Pedtke & Flood, 2020) As age increased, the gap between women and men shrank.

[1] "Skilled Scalable Services" (SSS) refers to skill- and information-intensive services, including Information, Finance and Insurance, Professional Services, and Management of Companies (Althoff et al., 2020.)

Across all age-gender subgroups, college degrees provided protection against COVID-driven job loss, especially for young adults. Individuals in their 20's without a college degree were much more likely to be unemployed or leave the workforce (Table 2). The dynamic within racial/ethnic subgroups were complicated, with some phenomena seeming culturally oriented. Asian women in their 30s, 40s, and 50s without a college degree, for instance, had the highest level of NILF-other, possibly due to increases in family-care responsibilities (Moen, Pedtke & Flood, 2020)

Table 2: Percent in Two Employment States in January 2020, By Gender, Age, and Education

		Women		**Men**	
		20 - 29	30 - 39	20 - 29	30- 39
No college degree	Unemployment	4.7	3.5	5.9	4.1
	NILF-other	27.3	24.8	18.1	6.3
College or above	Unemployment	2.4	1.4	3.5	1.3
	NILF-other	13.3	14.5	11.8	2.9

Most of the published studies regarding the pandemic's impact on employment focused on the onset of the COVID-19 pandemic. Our analysis will examine whether these phenomena are still relevant today and provide an up-to-date overview of the impact of the COVID-19 pandemic.

Initial Data Analysis

Employment

Similar to the existing literature, age and gender disparities were observed in the Household Pulse Survey (HPS) data, especially in Phase 3. Individuals in their 20's or younger reported the highest percentage of employment loss (34%) among all age groups, while women experienced job loss more often than men in all age groups, except for those aged 60 and above (Figure 4)

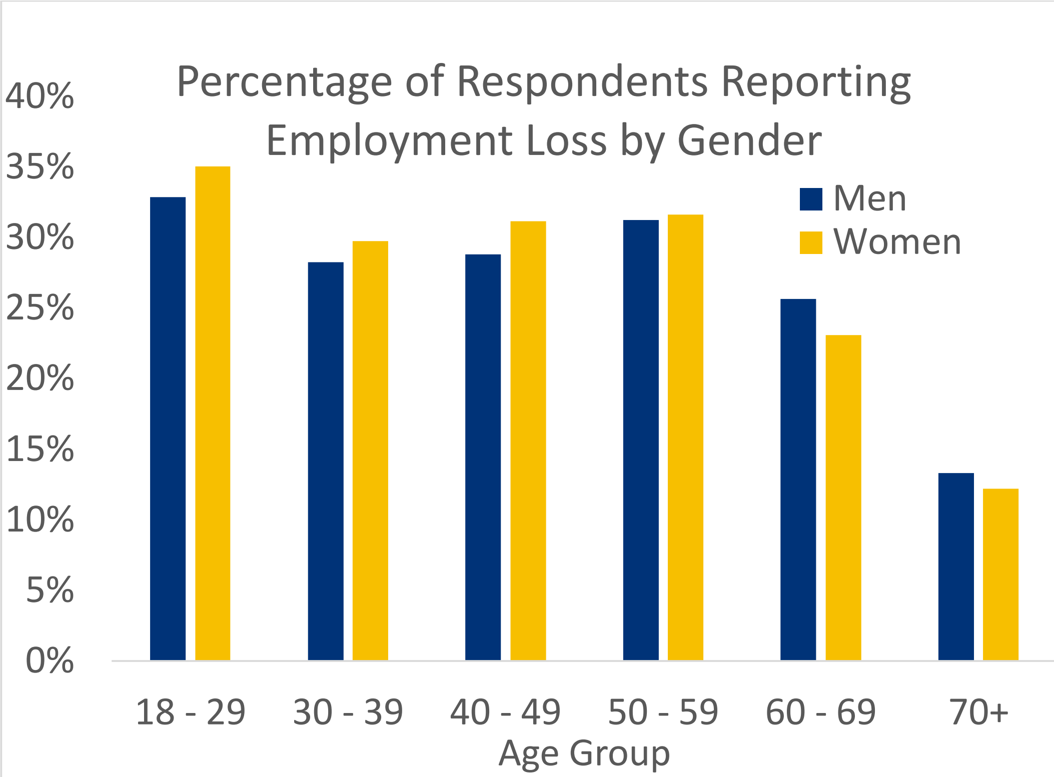

Figure 4: Percentage of Employment Loss Experienced by Men and Women by Age

Focusing on the state of recent employment, around 42% of respondents stated they did not work for pay or profit in the past seven days before answering the survey. This represents a larger percentage of respondents reporting employment income loss, which was 26%. Among those sharing the reason why they did not work, there were different patterns among men and women in the different age groups. Table 3 shows the two most common reasons for not working in the past seven days for men and women, grouped by age (excluding the category "Others, please specify" as details were not provided). Young adults in their 20's or younger indicated that it was their personal choice to not work while no other group ranked this reason in the top two. Men in their 30's to 40's did not work, mostly because of external conditions of where they worked due to COVID's impacts (furlough or laid off). Unsurprisingly, a large proportion of women younger than 50 years old did not work, as they largely assumed childcare duty. Also, as expected, retirement played an important role in why individuals 50 years or older did not work.

Table 3: Top Two Reasons of Men and Women Aged 18 – 59 for Not Working Recently (Excluding "Others")

	Reason	Men	Women
20's & Below	Primary	Personal Choice (23%)	Childcare duty (17%)
	Secondary	Being laid off / furloughed due to coronavirus pandemic (16%)	Personal Choice (15%)
30's	Primary	Being laid off/ furloughed due to coronavirus pandemic (26%)	Childcare duty (36%)
	Secondary	Childcare duty (9%)	Being laid off/ furloughed due to coronavirus pandemic (13%)
40's	Primary	Being laid off/ furloughed due to coronavirus pandemic (24%)	Childcare duty (21%)
	Secondary	Sickness/ disability (not COVID) (9%)	Being laid off/ furloughed due to coronavirus pandemic (13%)
50's	Primary	Being laid off/ furloughed due to coronavirus pandemic (19%)	Sickness/ disability (not COVID) (15%) Retirement (15%)
	Secondary	Retirement (18%)	

Data Preparation and Pre-Processing

The data pre-processing task for descriptive and predictive modeling were done independently to create a dataset appropriate for the respective analyses. For the Cluster Analysis, all missing data was imputed, and sample size was reduced to match with the limited computing capacity.

61 consistent primary variables were initially retained from Phase 3 and Phase 3.1 data sets. 6 variables were renamed, of which 3 variables were recoded for consistency. Given the focus of the COVID-19 pandemic's impact on employment and mental health, variables unrelated to these two aspects or demographic data were removed from the combined data set. Redundant variables including "THHLD_NUMPER", "WEEK", "SCRAM", "EST_MSA" (high missingness) and "HLTHINS1" to "HLTHINS8" (recorded into "PRIVHLTH" and "PUBHLTH") were also dropped. Of the 61 primary original primary variables, 28 were retained (shown in Table 4). Out of the complete data sets totaling 884,713 observations, approximately 70% (617,321 observations) had complete data for the retained variables. To enable cluster analysis, the incomplete observations were removed. Finally, random sampling was performed to create a sample of 10,000 observations for Cluster Analysis.

Table 4: Variables Retained for Cluster Analysis

Demographic	COVID-19 Related	Employment	Mental Health
AGE	HADCOVID	WRKLOSSRV	ANXIOUS
EEDUC	RECVDVACC	EXPCTLOSS	DOWN
EGENDER		ANYWORK	INTEREST
EST_ST		UI_APPLYRV	WORRY
INCOME		SSA_APPLYRV	PRESCRIPT
MS		SSA_RECV	MH_NOTGET
REGION		TW_YN	MH_SVCS
RHISPANIC			
RRACE			
TNUM_PS			
THHLD_NUMADLT			
THHLD_NUMKID			

Model Building

The Cluster Analysis model was built on a randomly sampled dataset of Phase 3 and Phase 3.1, containing 28 variables and 10,000 observations. The distribution of demographics parameters such as RACE, GENDER, and AGE, were similar to the original combined dataset. Hierarchical Cluster Analysis was performed using Ward's Minimum Variance Method and the distance used was Gower distance, to account for the mixed data types. The number of clusters was chosen to be 3 based on relative cluster validation using average silhouette coefficient and a scree plot showing the within-cluster sum of square error (SSE). After clustering, Cluster 1 had 34% of observations, Cluster 2 had 42% of observations and the smallest one cluster 3 had 24% observations as shown below in Figures 5 and 6.

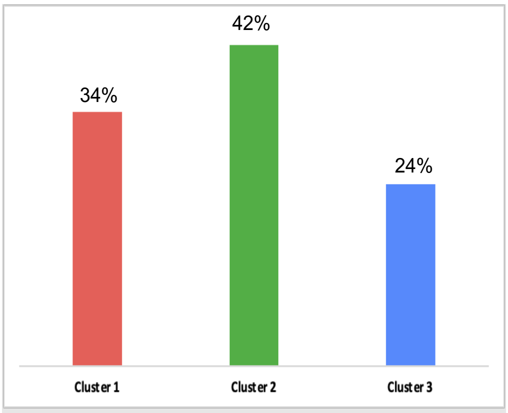

Figure 5: Proportion of observations per cluster

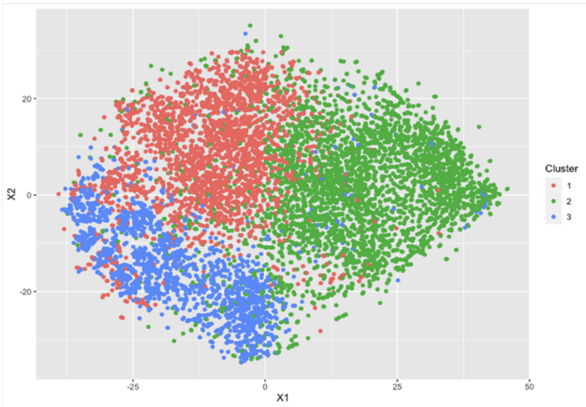

Figure 6: Hierarchical cluster analysis

Cluster Analysis

With respect to demographic parameters, Cluster 1 had an average age of 52 years while Cluster 2 had an average age of 47 years. For living alone, Cluster 1 had a higher percentage of people who had not completed a college education; this cluster also was represented with mostly females. Cluster 3 was comprised of relatively older people.

With respect to employment, Cluster 1 had the highest income levels and experienced some work loss, and few were expecting work loss in near future. Where job loss was reported, they didn't specify a reason for job loss. Although they were primarily employed in private companies with an option to work from home, they also did some volunteer work related to healthcare and other essential work during COVID times. On the other hand, Cluster 2 had mid-level income. Roughly 40% of people lost their job and more than 20% are expecting to lose their job in the upcoming days on the pretext of either their employer being temporarily closed, or it completely went out of business. Cluster 2 had a higher proportion of people who applied and received unemployment incentive and stimulus checks. This economic support was used by most people in this group to pay of their debts. Cluster 3 had the lowest income level and did not see much job loss, as most of the people were retired and most of them received retirement benefits. Cluster analysis insights were consistent with the existing literature, where non-college degree holders and younger people having some option of work from home were the most impacted due to the COVID-19 pandemic.

With mental health impacted due to COVID-19, Cluster 2 displayed increased anxiousness, depression, and worry. Additionally, the people in this cluster received mental health counseling and were given a prescription related to mental health issues. In contrast to the existing literature, which suggested that females who were living alone suffered from mental health problems, based on our data females living with more than 1 household members also suffered from mental health issues.

With respect to housing, people in Cluster 2 were least confident in paying their mortgages and were anticipating a foreclosure of their homes in the next 2 months. This clearly indicated loss of employment probably caused mortgage and foreclosure problems which could have contributed to the increased mental health concerns present in Cluster 2. Based on our analysis, Cluster 2 was the most impacted in terms of employment and mental health, followed by Cluster 1. Cluster 3, which represented older populations, was the least impacted of the 3 groups.

Recommendations

As employment of young individuals and women appeared to be impacted more severely during the COVID-19 pandemic, we recommend three short-term initiatives to help these groups to re-enter or remain in the workforce and improve PwC's talent acquisition and retention efficiency. Simultaneously, these initiatives can help PwC demonstrate its commitment to fulfilling social responsibility.

First, PwC may focus its effort on recruiting young talents (individuals in their 20's) for entry-level positions and provide designated on-board training. Since the great resignation, companies generally experienced loss of talents especially mid-career employees. Cook (2021) explained that because of the challenge of provide in-person training, companies were hesitant to hire entry-level employees leading a shortage of mid-career talents. The same phenomenon may also explain why young individuals reported higher employment loss. By targeting recent university graduates for entry-level positions, PwC could recruit the top-tiered young talent while competitor companies fight for mid-career talents. This initiative will be better paired with on-board training designed for young new entrants to ensure a smooth transition.

Secondly, PwC is recommended to further enhance its childcare support to employees with one or more young children to release the childcare burden from the workforce. Currently, PwC offers employees various childcare assistance programs including childcare discounts and back-up care and nanny assistance (PricewaterhouseCoopers, n.d.). Yet, finding a spot in childcare centers may not be as easy as expected. It will be very beneficial if PwC can reserve some childcare spots for internal application within the company. By giving extra support to employees with children, it does not only release more working hours but can also improve staff morale and loyalty.

Finally, we recommend that direct mortgage support should be provided to eligible employees so that PwC can help reduce housing- related burdens. As the current Employee Mortgage Program mainly provides connection to vendors who help with refinancing existing mortgages or applying for new mortgages (PricewaterhouseCoopers, n.d.), it may not alleviate the pressures facing homeowner employees. Addressing root causes of mental health issues like financial burden may be more effective than spending company budget on hiring counselors or providing mental health services to treat mental illnesses.

Overall, these short-term programs can help PwC establish its leadership in Environmental, Social & Governance (ESG) and bring long-term value to the company as it builds a stronger workforce to serve the company.

Predictive Analysis: Vaccination Intention

Since the start of the pandemic, many studies have been done on the intent and non-intent to be vaccinated. One such study conducted by the CDC focuses on the reasons for prioritized groups to not be vaccinated between September and December 2020. Both surveys showed that the intent to be vaccinated was highest in people 65 years of age and older. While the overall vaccination non-intent declined from 38.1 % to 32.1 % in all adults, some of the subpopulations such as women, non-Hispanic Black persons, younger adults, adults living in nonmetropolitan areas with a lower educational attainment and socio-economic circumstances showed a strong aversion to getting vaccinated based on the surveys conducted (CDC,2020). These subpopulations also showed limited access to healthcare and insurance which could contribute to their hesitancy.

The effectiveness of COVID-19 vaccines, along with any possible side-effects of the vaccine are the main concerns of people with a strong non-intent. To ensure that more people are vaccinated, the efforts of vaccination campaigns must be encouraged and public confidence in the vaccine should also be addressed and restored.

Another study which was performed on faculty, staff, postdoctoral research associates and medical students, found that 64.6% reported that they would get the COVID-19 vaccine. Additionally, respondents who had tested positive for COVID-19 reported that they are more interested in receiving the vaccine (Dowdle et al, 2021). Based on this research, we understand that individuals in the health sciences or education fields are more likely to be interested in receiving the COVID-19 vaccine.

Initial Data Analysis

Based on the data, we found that as levels of education increased among survey respondents, people were more likely to consider getting the vaccine. As shown in Table 5, while greater than 50% of survey respondents in our sample reported that they were intending to be vaccinated, the lowest percentages of willingness were among those who had graduated High School. Similar levels of willingness were shown when education levels are lower than High School. The largest sample of survey respondents held a bachelor’s degree (98,961) and 79% were willing to get vaccinated.

Table 5. Education and Vaccination Intent

	Not Intending to Get Vaccine	Intending to Get Vaccine	Percentage of Willing to Be Vaccinated
Less than High School	1,202	2154	64%
Some High School	2,445	4848	66%
High School Graduate or Equivalent	18,417	31555	63%
Some College	28,037	58387	68%
Associate’s Degree	13,822	25894	65%
Bachelor’s Degree	20,482	78479	79%
Graduate Degree	10,757	60256	85%

Age also proved to be an important factor in willingness to get vaccinated. As seen in Figure 7, survey respondents who are middle aged are more likely to get vaccinated, with the age group 55-64 showing the largest share of willingness, followed by 45-54. While the age group 35-44 has a high intent to be vaccinated, this age group also had a high percentage of people who did not intend to get vaccinated. Shares of younger and older survey respondents showed less significant results because most respondents were in the age range of 25-75.

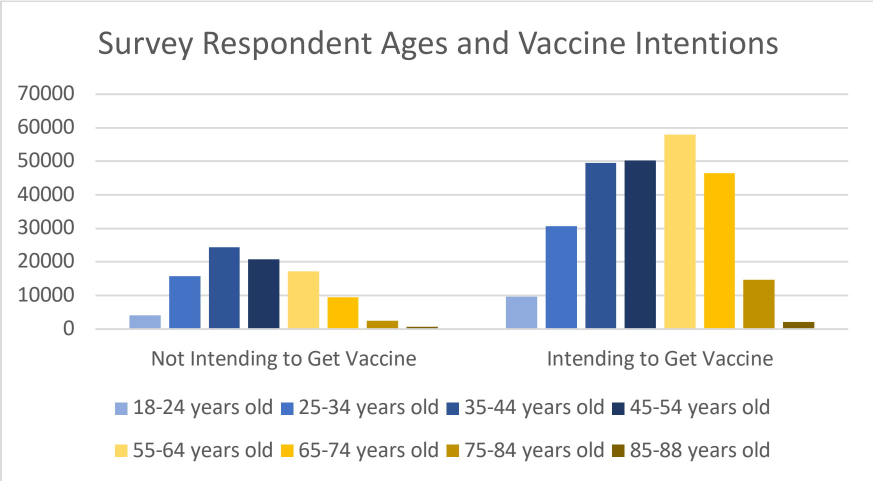

Figure 7: Survey Respondent Ages and Vaccine Intentions

Data Preparation and Pre-Processing

Data preprocessing was carried out before combining Phases 3 and 3.1. The “GETVACC” variable of Phase 3 and the “GETVACRV” variable of Phase 3.1 were determined as the target variable to predict whether an individual intends to receive the COVID-19 vaccine or not. Observations with missing data in the target variable were removed and responses indicating “Be unsure about getting a vaccine” (category 3) in Phase 3.1 were also excluded from the modeling process. Category 1 (“Definitely get a vaccine”) and 2 (“Probably get a vaccine”) were then grouped together as 1 for “Intend to get a vaccine” while category 3 (“Probably NOT get a vaccine”, equivalent to category 4 in Phase 3.1) and 4 (“Definitely NOT get a vaccine”, equivalent to category 5 in Phase 3.1) were grouped together as 0 for “Not intend to get a vaccine” (Table 6).

Table 6. Definition of Target Variable in Predictive Model

	Class 0: “Intend to get a vaccine”	**Class 1: “Not intend to get a vaccine”**
Phase 3	“Definitely get a vaccine” (category 1) “Probably get a vaccine” (category 2)	“Probably NOT get a vaccine” (category 3) “Definitely NOT get a vaccine” (category 4)
Phase 3.1	“Definitely get a vaccine” (category 1) “Probably get a vaccine” (category 2)	“Probably NOT get a vaccine” (category 4) “Definitely NOT get a vaccine” (category 5)

Remark: Categories not mentioned were removed from data sets

Out of the 204 and 239 variables in Phases 3 and 3.1 respectively, 131 variables were consistent (regardless of primary and secondary level), and 25 variables were mostly consistent with some slight modifications in the questions, response categories, or variable names. Including the target variable, 25 variables were renamed, and 9 variables were recoded for consistency. The Phase 3 and Phase 3.1 data sets were then combined.

Variable missingness was studied for those retained from the renaming and recoding step. While we initially planned to impute missing values to replace values lost, we believed it would be better to remove observations with missing values. The imputation method would have served to guess survey respondent answers to add to the analysis, whereas our approach would instead consider those who provided valuable answers to assist our analysis.

Although Decision Trees can handle missing data well, we decided to remove 9 primary variables (“EST_MSA”, “SPNDSRC1” to “SPNDSRC9”) with high missingness because it indicated a low response rate. As for secondary variables, since not every respondent was eligible for the secondary questions since they hadn’t answered a primary previously, these variables had a high magnitude of missingness. Because of these reasons, these secondary variables were removed. Various redundant variables were also identified and removed; these variables included “THHLD_NUMPER”, “WEEK”, “SCRAM”, “RECVDVACC”, “EST_ST” (reason being there were too many categories) and “HLTHINS1” to “HLTHINS8” (we opted to recode these into “PRIVHLTH” and “PUBHLTH”).

40 variables and 356,735 observations were eventually retained for predictive modeling. Appendix A shows details of variable consistency and the renaming and recoding steps for the retained predictor variables.

Predictive Modeling

The objective of the analysis is to build a predictive model for vaccination intention. We explored different classification analysis models and decision tree was chosen, since it produced the most robust results.

The split for train and test is 70% and 30%, respectively, based on the target variable vaccination intention. We observed that the observations in class 1 are higher than the observations in class 0. Random under sampling was applied to the training dataset to treat this class imbalance. We then performed hyperparameter tuning and 10-fold cross validation, repeated 3 times. The hyperparameter tunes is the complexity parameter (cp); we find an optimal value for this to reduce the impact of overfitting on the Decision Tree model. As a result of hyperparameter tuning, we determined a complexity parameter of 0.005 would provide the highest overall accuracy. Figure 8 shows the training results with respect to cp and accuracy.

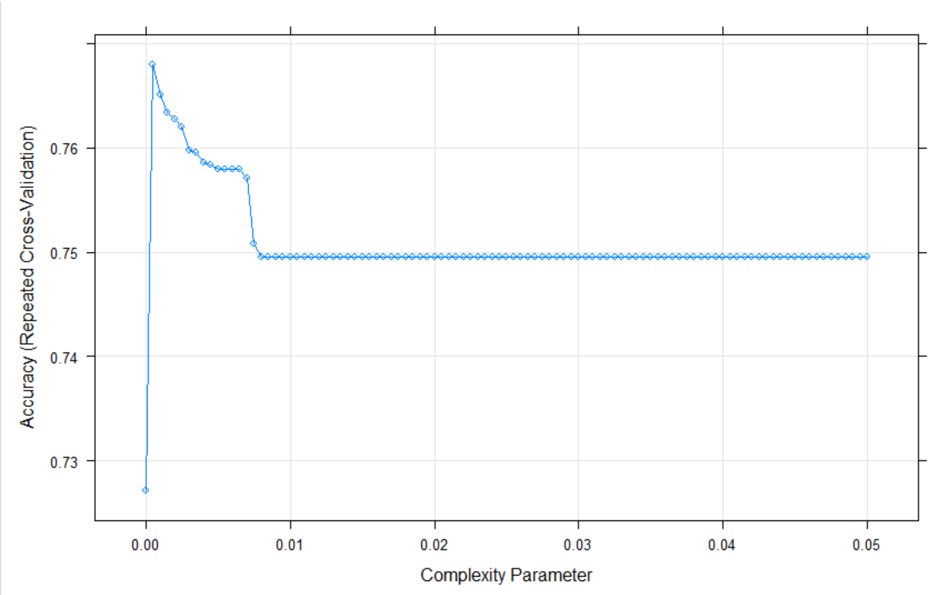

Figure 8: Complexity Parameter vs Accuracy

We then applied the classification Decision Tree method to the train dataset, and it produced 68% accuracy. The train dataset result was used to predict the test dataset, and it had an accuracy of 69%. Based on the low accuracy of the model, it is currently underfitting, and additional analysis methods could be considered to reduce the underfitting. Further, future analysis can include more data or variables to possibly reduce underfitting.

For running the model, we recommend defining the target variable vaccination intention, in two classes: Class 0 being definitely not get the vaccine and probably not get the vaccine, and Class 1 being definitely get the vaccine and probably get the vaccine. Based on our findings, the Decision Tree is the best predictive model to use, but with an accuracy of 68%, we recommend further modifications of variables used in the model, more variables be considered as predictors, and additional analysis methods be considered to improve the accuracy of the predictive model.

Recommendations

As a result of our analysis, we found the most important variables to be education, age, telework, income and mental health. For younger age groups and those with some high school education, vaccine hesitancy was widespread, in fact the lowest percentage of those who are willing to be vaccinated comes from those with some high school education (Nguyen, 2021). On the other side, telework proved to be an important indicator in the level of education a survey respondent had, as 62% of workers using Telework to work remotely reported having at least a bachelor's degree (Parker, 2021), and were less likely to show vaccine hesitancy because of their level of education.

Mental health also proved to be significant in predicting vaccine intention, as those who reported having mental health issues during the pandemic were among the demographics less likely to pursue vaccination: Young adults and the elderly (Fernandez, 2021). Finally, income was an important factor in vaccination, as those with lower reported incomes were less likely to get vaccinated (Baack, 2021).

We recommend implementing several initiatives to be used to build a stronger workforce as everyone adapts to new normal post-pandemic life.

As income proved to be a major factor in a person's willingness to get vaccinated, we recommend emulating other company's compensation perks to entice employees to get vaccinated, such as offering time off for vaccination and recovery time. Aldi is a company that pledged to pay hourly wages for time taken to get vaccinated (Terrell, 2021) and they have seen success with that program. Keeping monetary incentives in mind may be the most successful way to launch a vaccination program.

To bolster education programs on vaccines, we recommend distributing helpful materials to employees that serve to inform on the importance of getting vaccinated. The CDC recommends several PDFs to ease conversations; we found one of the most compelling to be "Diseases and the vaccines that prevent them" (CDC, 2019) as it starts by describing the diseases that can be better fought with vaccines.

Finally, we recommend helping employees access mental health benefits, to boost mental health strength in younger employees as well as old. This effort should help counteract vaccine hesitancy, prompt people to discuss their anxieties in therapy and eventually lead them to want to get vaccinated. Vaccine counseling services were tested in Palermo, Italy and showed interestingly positive results from encouraging counseling to overcome hesitations (Constantino, 2020).

References

Althoff, L., Eckert, F., Ganapati, S., & Walsh, C. (2021). *The Geography of Remote Work* (No. w29181). National Bureau of Economic Research.

Baack, B. N., Abad, N., Yankey, D., Kahn, K. E., Razzaghi, H., Brookmeyer, K., Kolis, J., Wilhelm, E., Nguyen, K. H., & Singleton, J. A. (2021). COVID-19 vaccination coverage and intent among adults aged 18–39 years — United States, March–May 2021. *Centers for Disease Control and Prevention*, *70*.

Banerjee, D. (2020). The impact of COVID-19 pandemic on Elderly Mental Health. *International Journal of Geriatric Psychiatry*, *35*(12), 1466–1467. https://doi.org/10.1002/gps.5320

Centers for Disease Control and Prevention. (2019, May 7). *Education info about vaccines and diseases they prevent*. Centers for Disease Control and Prevention. Retrieved December 1, 2021, from https://www.cdc.gov/vaccines/ed/patient-ed.html.

Centers for Disease Control and Prevention. (2021, February 11). *COVID-19 vaccination intent, perceptions, and reasons for not vaccinating among groups prioritized for early vaccination - United States, September and December 2020*. Centers for Disease Control and Prevention. Retrieved October 21, 2021, from https://www.cdc.gov/mmwr/volumes/70/wr/mm7006e3.htm.

Chin, E. T., Huynh, B. Q., Lo, N. C., Hastie, T., & Basu, S. (2020). Projected geographic disparities in healthcare worker absenteeism from COVID-19 school closures and the economic feasibility of childcare subsidies: A simulation study. BMC Medicine, 18(1). https://doi.org/10.1186/s12916-020-01692-w

Costantino, C., Caracci, F., Brandi, M., Bono, S. E., Ferro, A., Sannasardo, C. E., Scarpitta, F., Siddu, A., Vella, C., Ventura, G., Vitale, F., Casuccio, A., & Restivo, V. (2020). Determinants of vaccine hesitancy and effectiveness of vaccination counseling interventions among a sample of the general population in Palermo, Italy. *Human Vaccines & Immunotherapeutics*, *16*(10), 2415–2421. https://doi.org/10.1080/21645515.2020.1728157

Cook, I. (2021, November 10). *Who is driving the great resignation?* Harvard Business Review. Retrieved November 12, 2021, from https://hbr.org/2021/09/who-is-driving-the-great-resignation.

del Rio-Chanona, R. M., Mealy, P., Pichler, A., Lafond, F., & Farmer, J. D. (2020). Supply and demand shocks in the COVID-19 pandemic: An industry and occupation perspective. *Oxford Review of Economic Policy*, *36*(Supplement_1), S94-S137.

Department of Education. (2021). Education in a Pandemic: The Disparate Impacts of COVID-19 on America's Students. *Office for Civil Rights Blog*. Retrieved October 20, 2021, from https://www2.ed.gov/about/offices/list/ocr/docs/20210608-impacts-of-covid19.pdf.

Dingel, J. I., & Neiman, B. (2020). How many jobs can be done at home? *Journal of Public Economics*, *189*, 104235. https://doi.org/10.1016/j.jpubeco.2020.104235

Dowdle, Travis, and Jeff Dennis. “Intention to Receive COVID-19 Vaccine by U.S. Health Sciences University Employees - Travis S. Dowdle, Jeff Dennis, Kenneth M. Nugent, Theresa Byrd, 2021.” SAGE Journals, https://journals.sagepub.com/doi/full/10.1177/21501327211036611.

Fernandes, N.; Costa, D.; Costa, D.; Keating, J.; Arantes, J. (2021). Predicting COVID-19 Vaccination Intention: The Determinants of Vaccine Hesitancy. Vaccines, 1161. https://doi.org/10.3390/vaccines9101161

Fernandez, M. E. (2021, March 5). *COVID-19 mental health crisis is hitting Young Adults*. www.heart.org. Retrieved October 19, 2021, from https://www.heart.org/en/news/2021/03/05/COVID-19-mental-health-crisis-is-hitting-young-adults.

Javed, B., Sarwer, A., Soto, E. B., & Mashwani, Z. U. (2020). The coronavirus (COVID-19) pandemic's impact on mental health. *The International journal of health planning and management*, *35*(5), 993–996. https://doi.org/10.1002/hpm.3008

Loomba, S., de Figueiredo, A., Piatek, S.J. et al. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 5, 337–348 (2021). https://doi.org/10.1038/s41562-021-01056-1

Moen, P., Pedtke, J. H., & Flood, S. (2020). Disparate disruptions: Intersectional COVID-19 employment effects by age, gender, education, and Race/Ethnicity. *Work, Aging and Retirement*, *6*(4), 207–228. https://doi.org/10.1093/workar/waaa013

Nirmita Panchal, Nirmita, and Rabah Kama. “The Implications of COVID-19 for Mental Health and Substance Use.” KFF, 20 July 2021, https://www.kff.org/coronavirus-COVID-19/issue-brief/the-implications-of-COVID-19-for-mental-health-and-substance-use/.

Nguyen, Kimberly. “Changes in Covid-19 Vaccination Receipt and Intention to Vaccinate by Socioeconomic Characteristics and Geographic Area, United States, January 6 – March 29, 2021.” https://www.tandfonline.com/doi/full/10.1080/07853890.2021.1957998.

Parker, Kim, et al. “How Coronavirus Has Changed the Way Americans Work.” *Pew Research Center's Social & Demographic Trends Project*, Pew Research Center, 25 May 2021, https://www.pewresearch.org/social-trends/2020/12/09/how-the-coronavirus-outbreak-has-and-hasnt-changed-the-way-americans-work/.

Petts, R. J., Pepin, J. R., & Carlson, D. L. (2020). A gendered pandemic: Childcare, homeschooling, and parents' employment during COVID-19. Gender, Work & Organization, Supplement: Feminist Frontiers, 28(S2), 515–534. https://doi.org/10.1111/gwao.12614

Prados, M., & Zamarro, G. (2020). Gender Differences in Couples’ Division of Childcare, Work and Mental Health During COVID-19. CESR-Schaeffer Working Paper No. 003. Retrieved October 10, 2021, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3667803.

PricewaterhouseCoopers. (n.d.a). *Employee benefits*. PwC. Retrieved November 15, 2021, from https://www.pwc.com/us/en/careers/why-pwc/employee-benefits.html.

Terrell, K. (2021, September 10). *Companies paying employees to get COVID-19 vaccine*. AARP. Retrieved November 30, 2021, from https://www.aarp.org/work/working-at-50-plus/info-2021/companies-paying-employees-covid-vaccine.html.

United Nations. (2020). *Policy Brief: Education During COVID-19 and Beyond*. Retrieved October 19, 2021, from https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/08/sg_policy_brief_COVID-19_and_education_august_2020.pdf.

Wong, S. Y., Zhang, D., Sit, R. W., Yip, B. H., Chung, R. Y.-nork, Wong, C. K., Chan, D. C., Sun, W., Kwok, K. O., & Mercer, S. W. (2020). Impact of COVID-19 on loneliness, mental health, and Health Service Utilisation: A prospective cohort study of older adults with multimorbidity in primary care. *British Journal of General Practice*, *70*(700). https://doi.org/10.3399/bjgp20x713021

Appendix

Appendix A: Summary of Predictor Variables for Decision Tree Model

<u>Consistent Variables</u>

Phase 3	Phase 3.1	Predictive Model
EGENDER	EGENDER	EGENDER
RHISPANIC	RHISPANIC	RHISPANIC
RRACE	RRACE	RRACE
EEDUC	EEDUC	EEDUC
MS	MS	MS
THHLD_NUMKID	THHLD_NUMKID	THHLD_NUMKID
THHLD_NUMADLT	THHLD_NUMADLT	THHLD_NUMADLT
HADCOVID	HADCOVID	HADCOVID
EXPCTLOSS	EXPCTLOSS	EXPCTLOSS
ANYWORK	ANYWORK	ANYWORK
SSA_RECV	SSA_RECV	SSA_RECV
EXPNS_DIF	EXPNS_DIF	EXPNS_DIF
CURFOODSUF	CURFOODSUF	CURFOODSUF
FREEFOOD	FREEFOOD	FREEFOOD
SNAP_YN	SNAP_YN	SNAP_YN
TSPNDFOOD	TSPNDFOOD	TSPNDFOOD
TSPNDPRPD	TSPNDPRPD	TSPNDPRPD
ANXIOUS	ANXIOUS	ANXIOUS
WORRY	WORRY	WORRY
INTEREST	INTEREST	INTEREST
DOWN	DOWN	DOWN
DELAY	DELAY	DELAY
NOTGET	NOTGET	NOTGET
PRESCRIPT	PRESCRIPT	PRESCRIPT
MH_SVCS	MH_SVCS	MH_SVCS
MH_NOTGET	MH_NOTGET	MH_NOTGET
TENURE	TENURE	TENURE
TNUM_PS	TNUM_PS	TNUM_PS
INCOME	INCOME	INCOME
PRIVHLTH	PRIVHLTH	PRIVHLTH
PUBHLTH	PUBHLTH	PUBHLTH
REGION	REGION	REGION
AGE	AGE	AGE

Renamed Variables

Phase 3	Phase 3.1	Predictive Model
WRKLOSS	WRKLOSSRV	WRKLOSSRV
UI_APPLY	UI_APPLYRV	UI_APPLYRV
SSA_APPLY	SSA_APPLYRV	SSA_APPLYRV

Recoded and Renamed Variables

Phase 3	Phase 3.1	Predictive Model
TW_START (Teleworking start due to COVID)	TW_YN (Teleworking start due to COVID)	TW_YN
1) Yes, at least one adult substituted some or all their typical in-person work for telework 2) No, no adults substituted their typical in-person work for telework 3) No, there has been no change in telework	1) Yes 2) No	Follow Phase 3.1 definition. For Phase 3, category 1 recoded as category 1 (Yes); category 2 and 3 recoded as 2 (No).
EIP (Receipt and use of Economic Impact Payment (Stimulus))	EIP_YN (Receipt and use of Economic Impact Payment (Stimulus))	EIP_YN
1) Mostly spend it 2) Mostly save it 3) Mostly use it to pay off debt 4) Not applicable, I did not receive the stimulus payment	1) Yes 2) No	Follow Phase 3.1 definition. For Phase 3, category 1, 2, and 3 recoded as category 1 (Yes); category 4 recoded as category 2 (No)
LIVQTR	LIVQTRRV	LIVQTRRV