



Vaccine Intent & Impact of the COVID-19 Pandemic

Drexel University LeBow College of Business

November 18, 2021

Group 1: Kaitlyn Bleiweiss, James Butler, Janelle Gillis, Brett Vaccaro & Brian Weiss



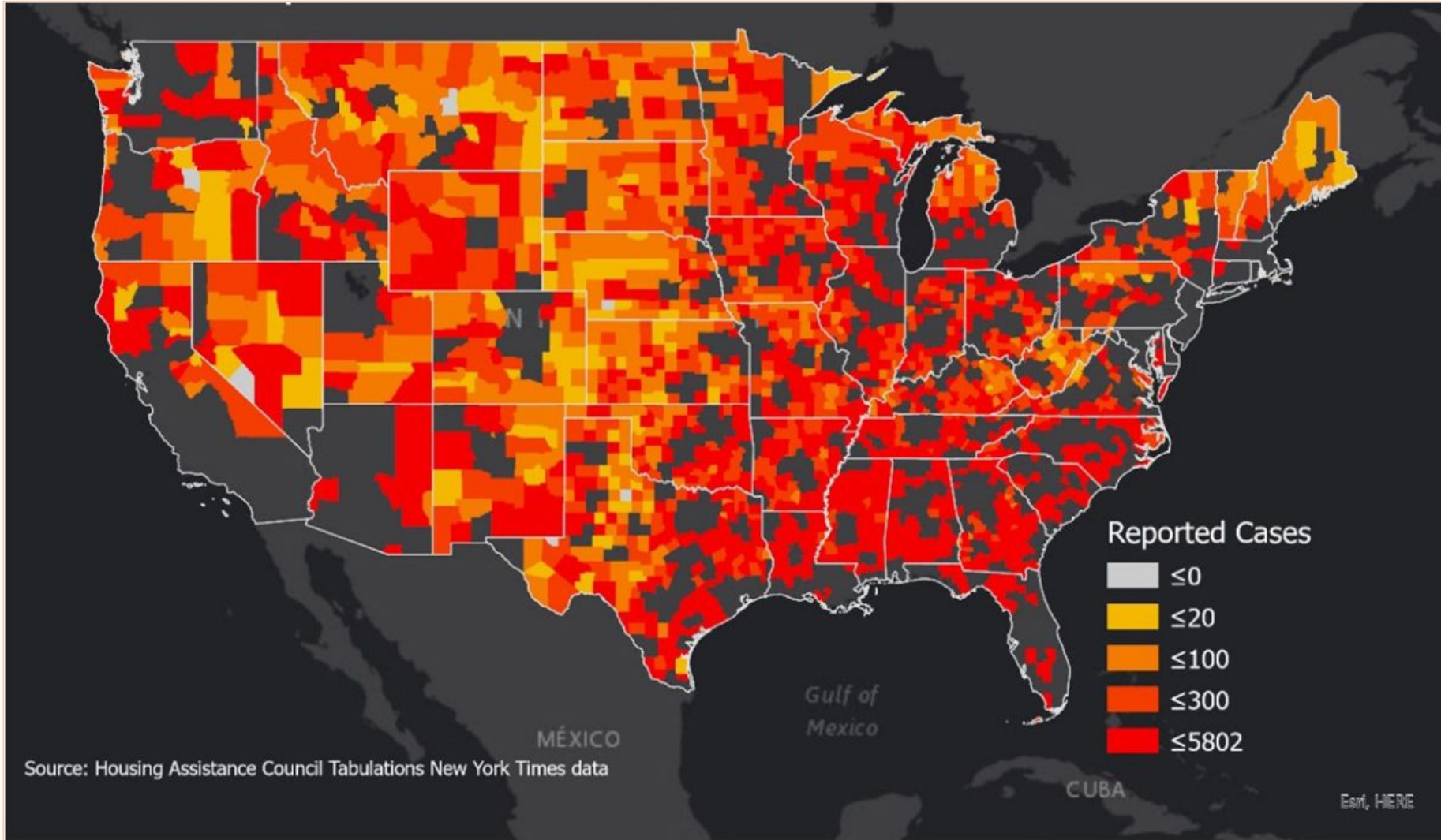
Agenda

1. Business Understanding
2. Data Understanding & Preparation
3. Descriptive Data Analysis
4. Predictive Model
5. Recommendations & Conclusion



BUSINESS UNDERSTANDING





COVID-19

Assessing impact on physical, social, psychological, and economic health indicators.



Descriptive Analysis

What are the effects of COVID-19 on the population?
Can a descriptive explanation be created that defines the different effects on areas such as:

- Childcare
- Education
- Employment
- Energy use
- Food security
- Health
- Housing
- Household spending

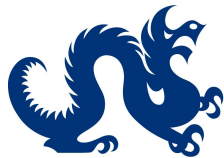
Predictive Model

Select and build model(s) that will accurately predict vaccine intent

- Population level
- Identify factors
- Leading indicators



DATA UNDERSTANDING & PREPARATION



DREXEL UNIVERSITY

LeBow

College of Business





Household Pulse Survey

- Phase 3.0 Q1-2021
- Phase 3.1 Q2-2021

<https://www.census.gov/householdpulsedata>

<u>Dataset</u>	<u>Data File</u>	<u>Date Range</u>	<u># Variables</u>	<u># Observations</u>
3.0	Week 22	1/6 - 1/18	204	68,348
3.0	Week 23	1/20 - 2/1	204	80,567
3.0	Week 24	2/3 - 2/15	204	77,122
3.0	Week 25	2/17 - 3/1	204	77,788
3.0	Week 26	3/3 - 3/15	204	78,306
3.0	Week 27	3/17 - 3/29	204	77,104
3.1	Week 28	4/14 - 4/26	239	68,913
3.1	Week 29	4/28 - 5/10	239	78,467
3.1	Week 30	5/12 - 5/24	239	72,897
3.1	Week 31	5/26 - 6/7	239	70,854
3.1	Week 32	6/9 - 6/21	239	68,067
3.1	Week 33	6/23 - 7/5	239	66,262
Total				884,695



Data Preprocessing

Example of difference between phase 3.0 and 3.1, requiring transformation to merge correctly

QV3 Once a vaccine to prevent COVID-19 is available to you, would you...

- 1 Definitely get a vaccine
- 2 Probably get a vaccine
- 3 Probably NOT get a vaccine
- 4 Definitely NOT get a vaccine

QV3 Once a vaccine to prevent COVID-19 is available to you, would you...

- 1 Definitely get a vaccine
- 2 Probably get a vaccine
- 3 Be unsure about getting a vaccine
- 4 Probably NOT get a vaccine
- 5 Definitely NOT get a vaccine

Descriptive Analysis

- 136 variables and 884,695 observations
- 114 categorical variables
(survey data = multiple choice)
- 14 ordinal, only 7 numeric

Predictive Model

- Target variable = intent
- Originally ordinal, range 1-5
- Eventually binarized: intent = yes or no



DESCRIPTIVE DATA ANALYSIS



Initial Analysis of Respondents

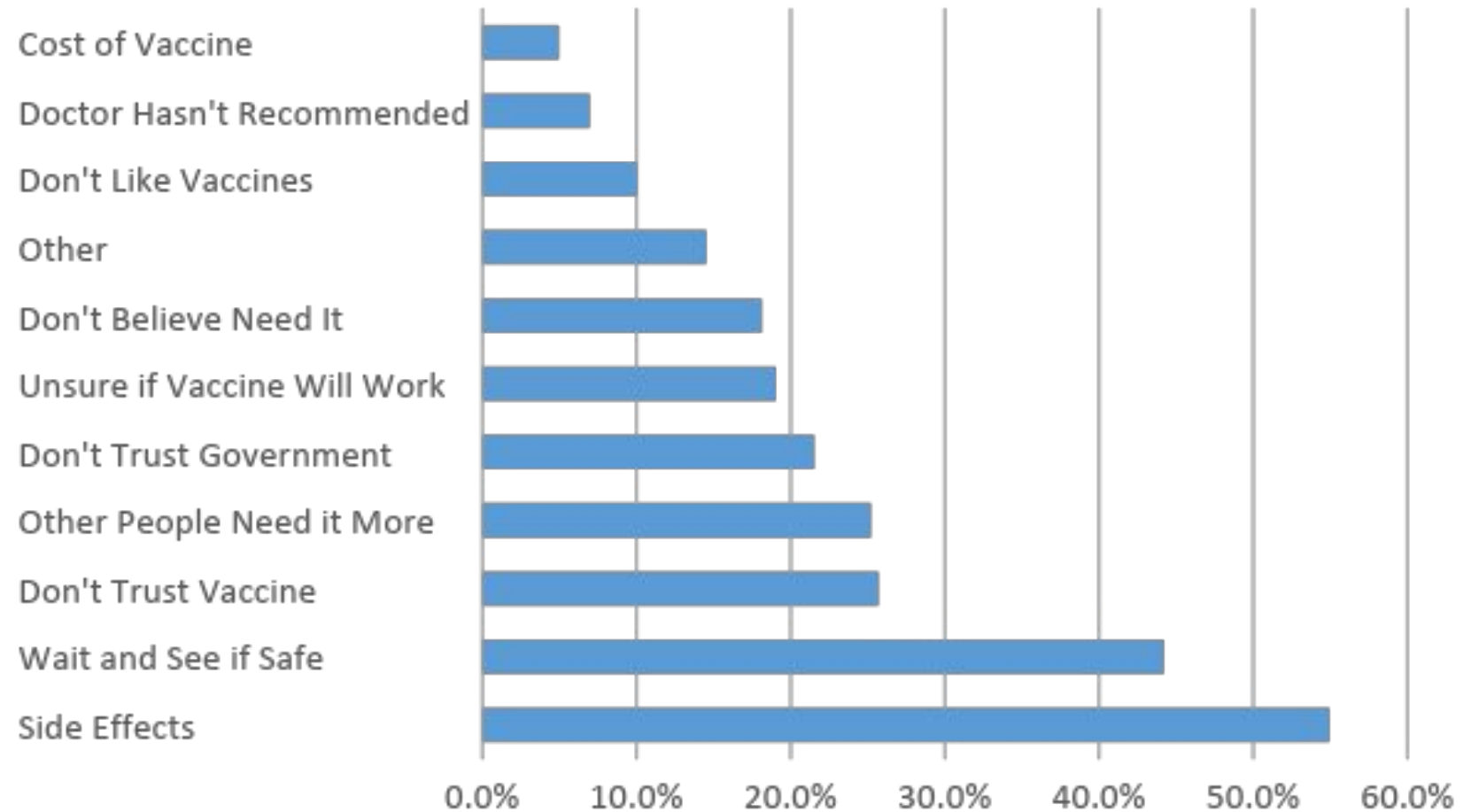
- 59.5% female and 84.0% Caucasian
- Education: 13.6% high school or lower, 60.8% at least some college, 18.4% graduate education
- Marital Status: 18.4% never married, 58.3% married, 22.5% widowed, separated, or divorced
- Median birth year of 1966
- Households contain average of 2.72 people



Vaccination and COVID-19

57.6% of respondents received the COVID-19 vaccination

Reason for Vaccine Hesitancy



Mental Health Impacts

- 11.1% respondents received mental health care services, lower than 19.2% reported by CDC in 2019 (Terlizzi & Zablotsky, 2020)
- 10.2% indicated not receiving needed mental health services
- Those that reported higher levels of anxiety and nervousness were disproportionately reporting delaying needed medical care
- Anxiety, lack of interest, and worry were related to anticipating employment related income loss and experiencing difficulty paying for normal expenses



Food Security

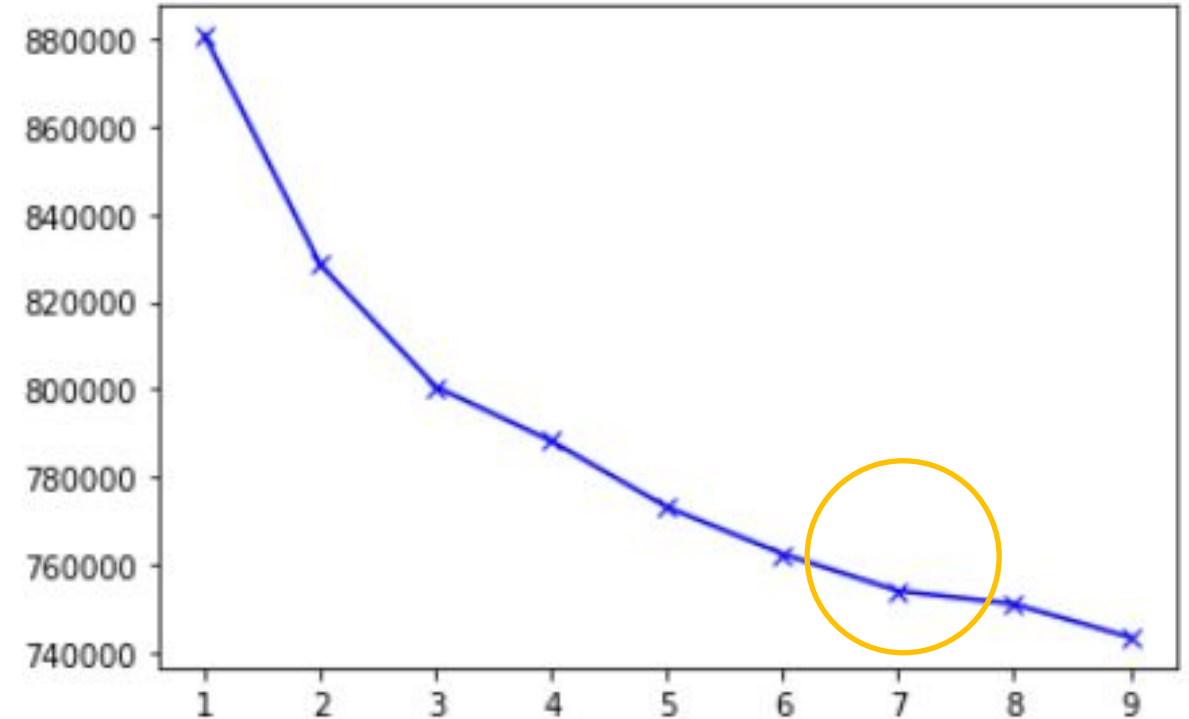
- Food insecurity has increased during the pandemic (Niles et al., 2021)
- 5.5% of data reported not having enough food
- 18.5% reporting having enough to eat, but not the types of food they want
- USDA reported that in 2020 10.5% experienced some form of food insecurity (Coleman et al., 2021)
- Supplemental Nutrition Assistance Program (SNAP) recipients reported higher levels of experiencing difficulty paying for normal expenses



K-Mode Clustering

Sample Size Dimensionality: 51,121 observations X 46 variables

- Optimum Clusters: 7
 - Cluster 1 Size: 7,045
 - Cluster 2 Size: 11,769
 - Cluster 3 Size: 8,591
 - Cluster 4 Size: 3,834
 - Cluster 5 Size: 8,058
 - Cluster 6 Size: 3,769
 - Cluster 7 Size: 8,055



K-Mode Clustering: What Did We Learn?

Cluster 4



- Female
- Born in the 1970s
- From the West
- Some College but No Degree
- Not Receiving Regular Income Like Before Pandemic
- Income: \$25k-\$35k
- Renting
- Delayed Getting Medical Care Due to COVID
- More Than Half of the Days Little Interest in Activities
- Worried and Anxious Nearly Every Day
- Enough Food but Not Always the Kind Liked
- Very Difficult Paying Bills
- Never Married
- No Work in 7 Days and Using Savings for Income or Funds
- Non-Vaccinated (COVID)



K-Mode Clustering: What Did We Learn?

Cluster 5



- Male
- Born in the 1960s
- From the South
- Graduate Degree
- Income: \$150K - \$200k
- Single Family Home: Mortgage
- Married
- Private Insurance
- Worked in the Past 7 days
- Not Experiencing Worry, Anxiety, or Lack of Interest
- No Difficulty Paying Bills
- No Food Insecurities
- Vaccinated (COVID)

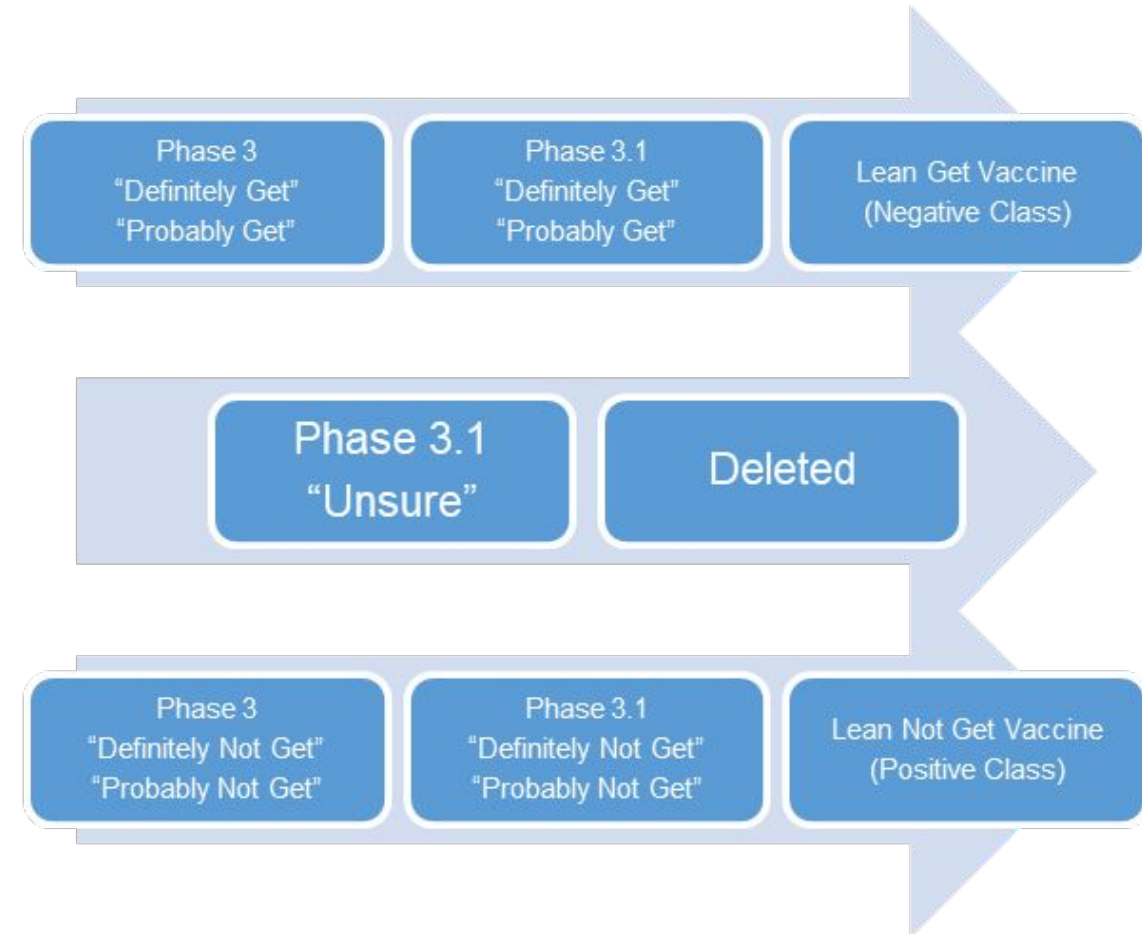


PREDICTIVE MODEL



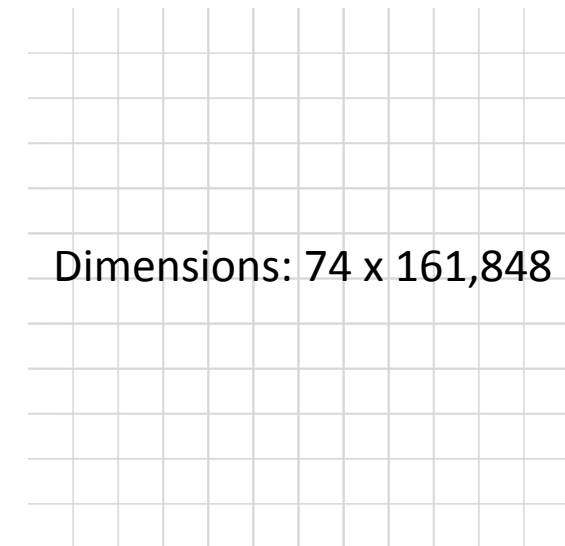
Target Variable: Vaccine Intent

- Filtered to exclude vaccinated participants
- Filtered to exclude missing values
- Follow-up questions to target variable removed to avoid potential bias



Dimensionality Reduction:

- Removed survey questions unique to Phase 3 or 3.1
- Removed conditional survey questions
- Removed irrelevant variables related to vaccination
- Removed rows with more than 50% missing values
- Applied down sampling to correct for class imbalance



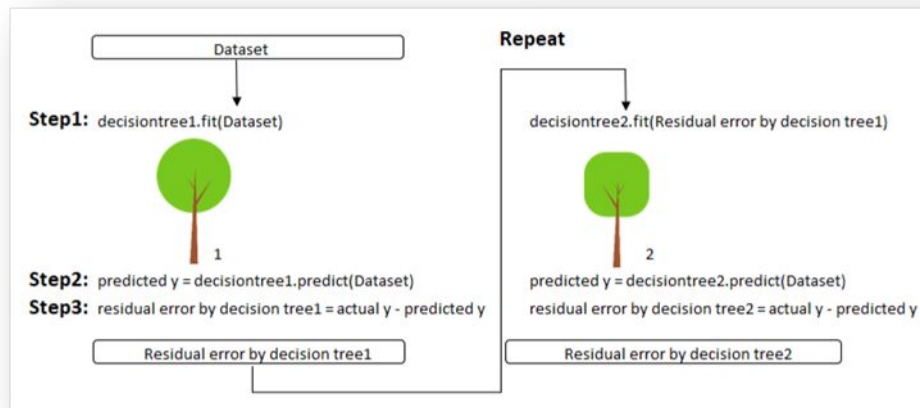
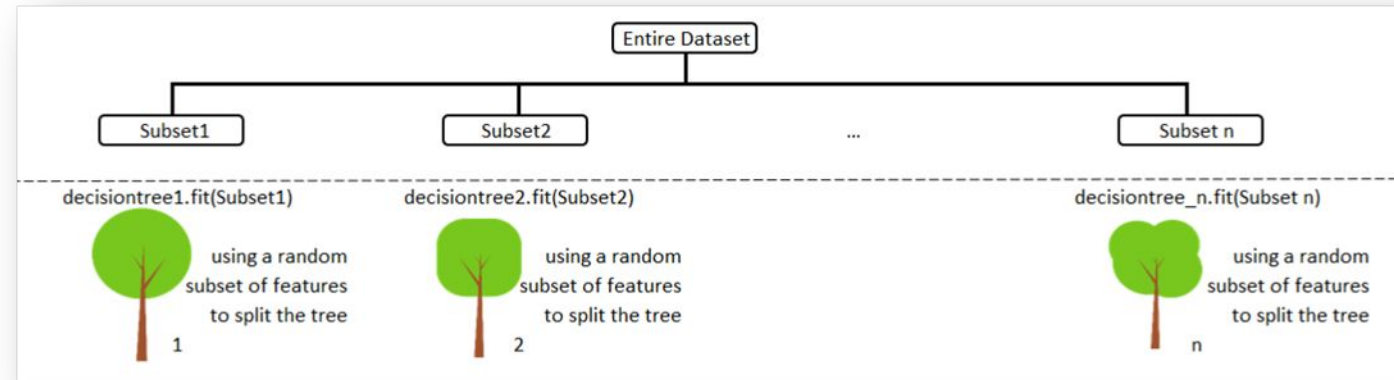
Multivariate Imputation of Missing Values:

- 10 iterations
- Logistic regression for 2-factor variables
- Predictive mean matching for all other variables



Random Forest:

- Multiple decision trees
- Random subset of variables to split the tree



(Chen, 2019)

Gradient Boosting:

- Multiple decision trees in series
- Prediction errors of previous tree trains the next tree



Extreme Gradient Boosting

- Training/Testing partition
 - 80:20
- Cross validation
 - 5-fold with 3 repeats
- Optimal tuning parameters
 - Number of boosting rounds = 130
 - Maximum nodes allowed from root to farthest leaf = 8
 - Fraction of variables to use = 0.9
 - Fraction of observations to sample at each step = 1
 - Learning rate = 0.09
 - Minimum weight required to create a new node = 6
 - Minimum error reduction to make a new code = 0

	Training	Testing
Accuracy	0.7158	0.7150
Kappa	0.4315	0.4299
Sensitivity	-	0.6961
Specificity	-	0.7339
F-score	-	0.7095

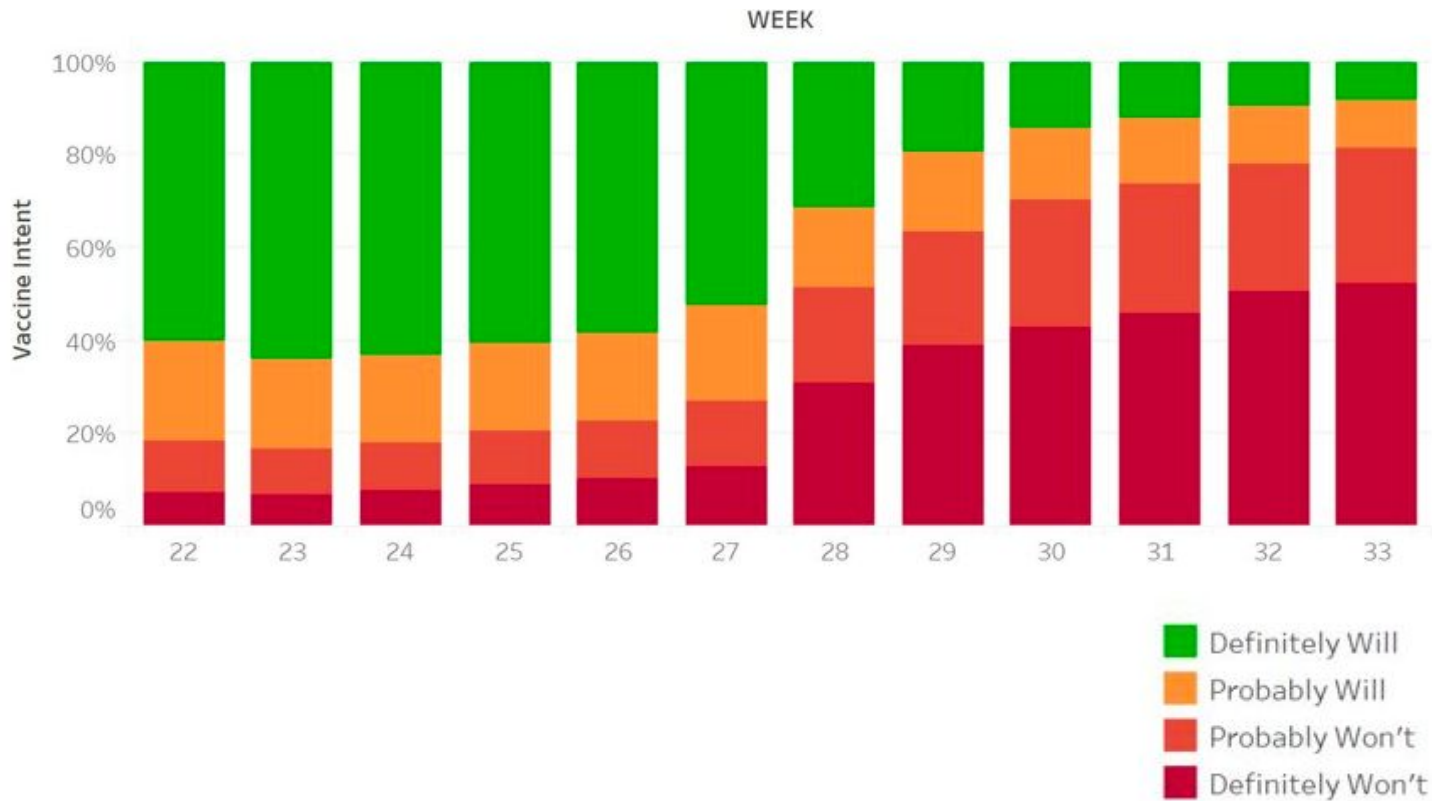


Which variables are most important in predicting survey participants' intent to vaccinate?

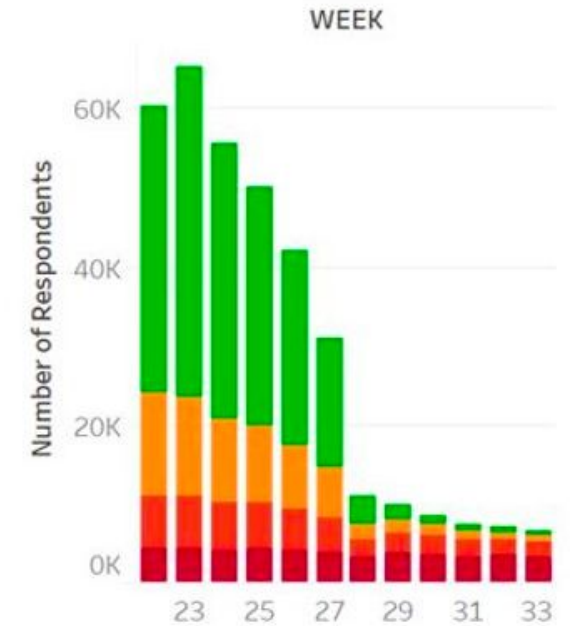
<u>Rank</u>	<u>Variable Name</u>	<u>Variable Description</u>	<u>Importance</u>
1	WEEK	Week of the survey	125.4
2	FEWTRANS	Fewer public transit or rideshare trips due to pandemic	71.4
3	TBIRTH_YEAR	Survey participant birth year	63.5
4	EEDUC	Survey participant education level	61.6
5	INCOME	Household income level	47.6
6	EXPNS_DIF	Difficulty with expenses	46.0
7	ANXIOUS	Frequency of anxiety over the past 7 days	42.7
8	THHLD_NUMKID	Total number of people under 18-years-old in household	35.0
9	CURFOODSUF	Household food sufficiency for last 7 days	33.5
10	INTEREST	Frequency of having little interest in things over the past 7 days	32.7



Vaccination Intent Over Time

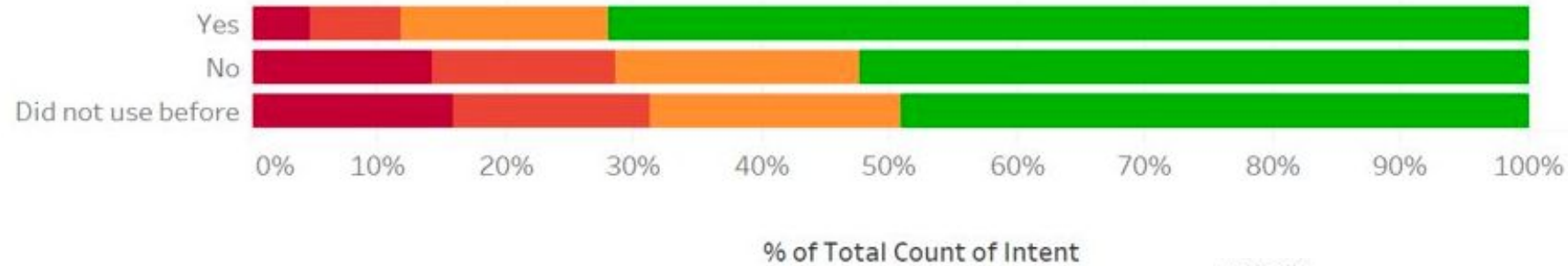


Results filter out those who have received the vaccine.

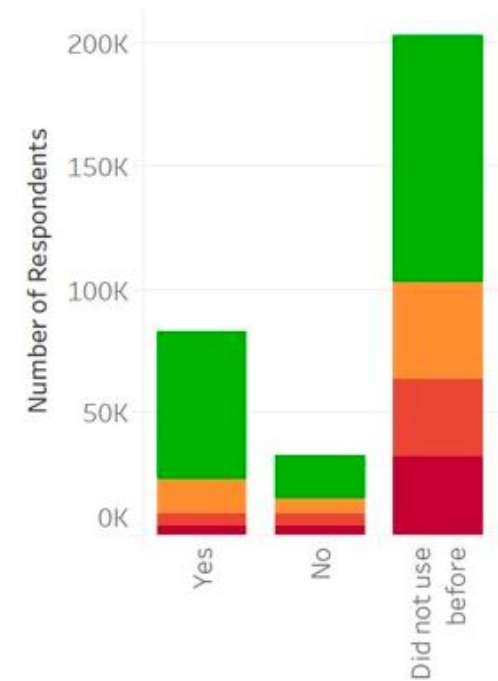


Reduced Public Transit vs Vaccination Intent

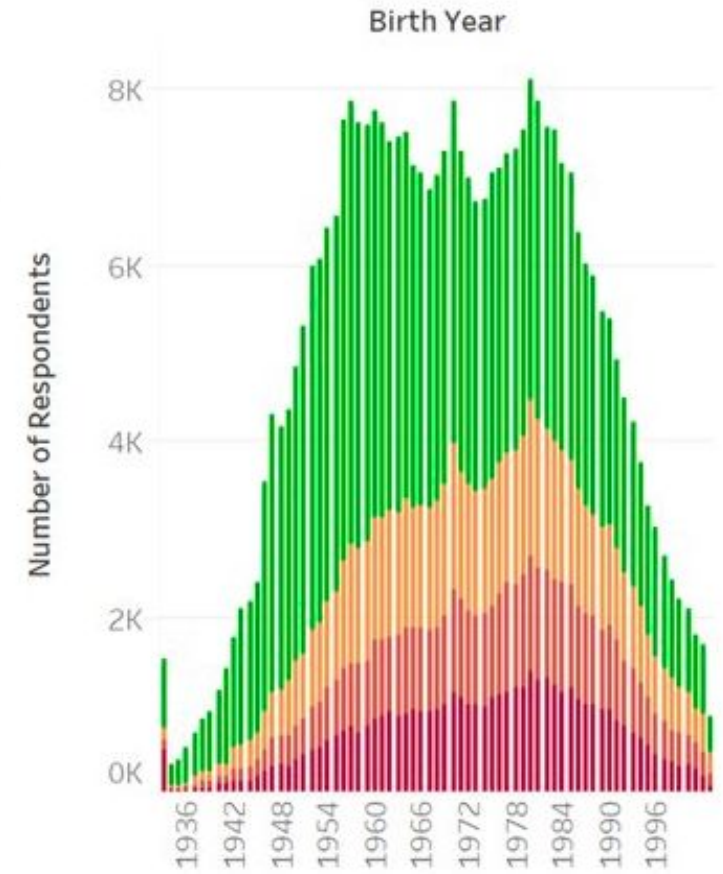
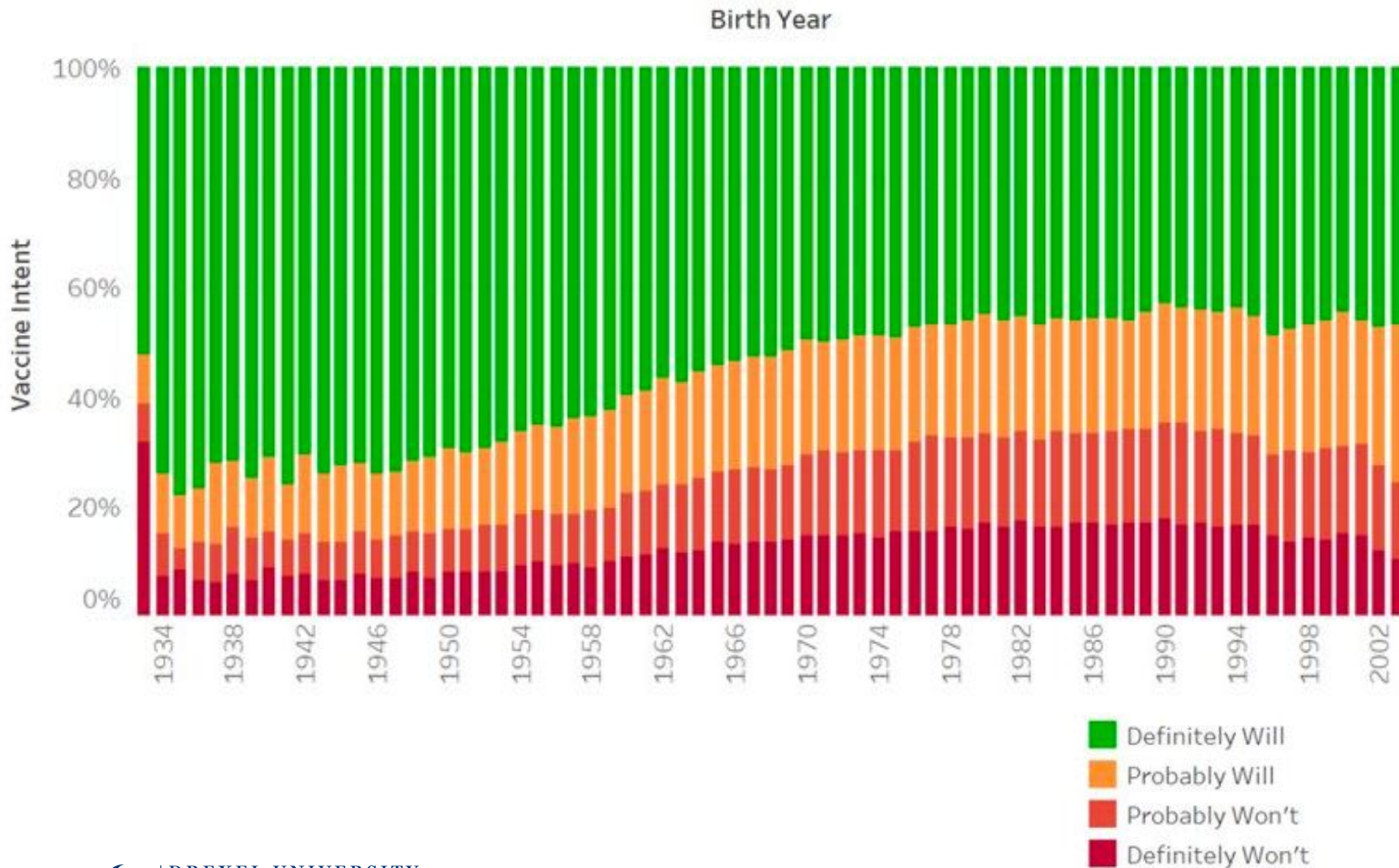
fewrtrans %



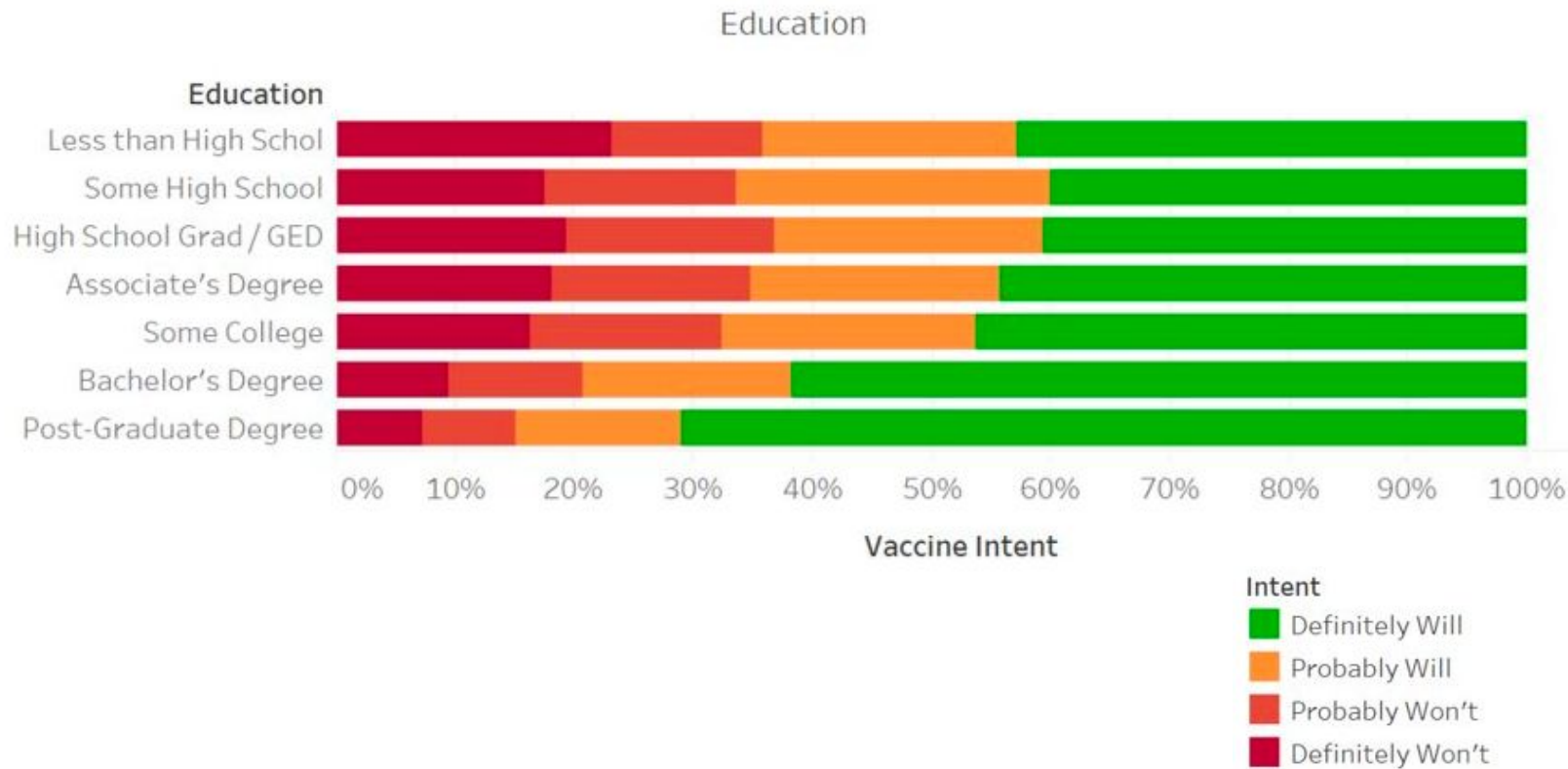
fewrtrans #



Vaccination Intent by Birth Year

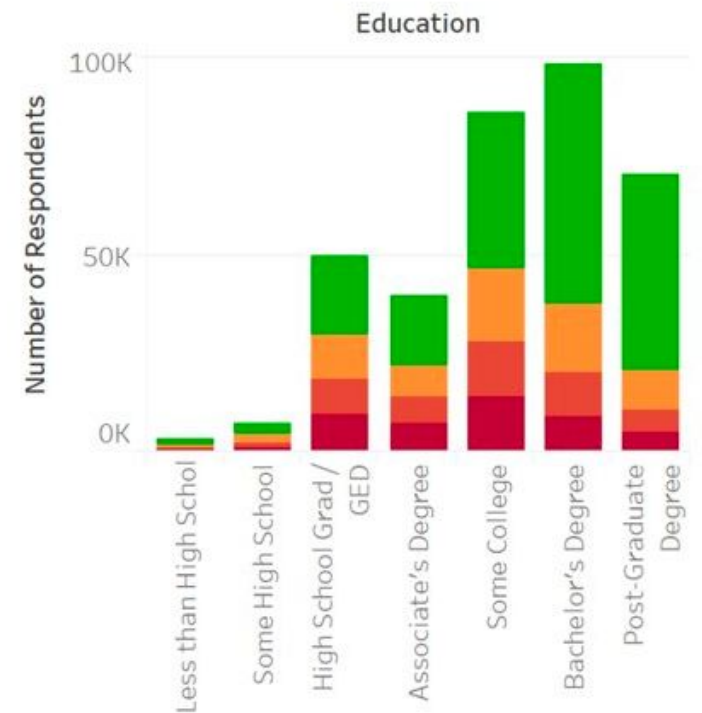


Education vs Vaccination Intent

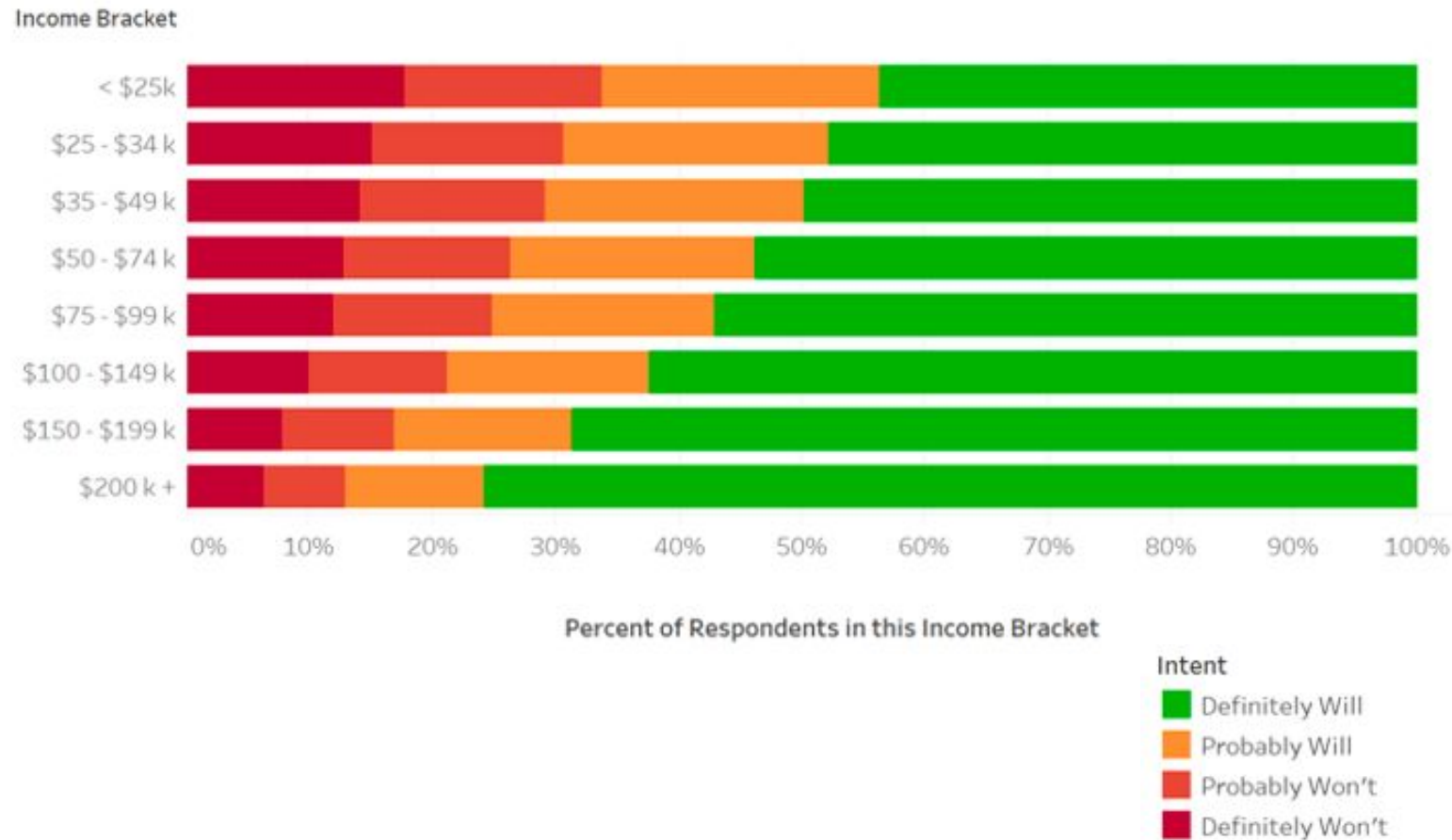


Distribution of Education Level

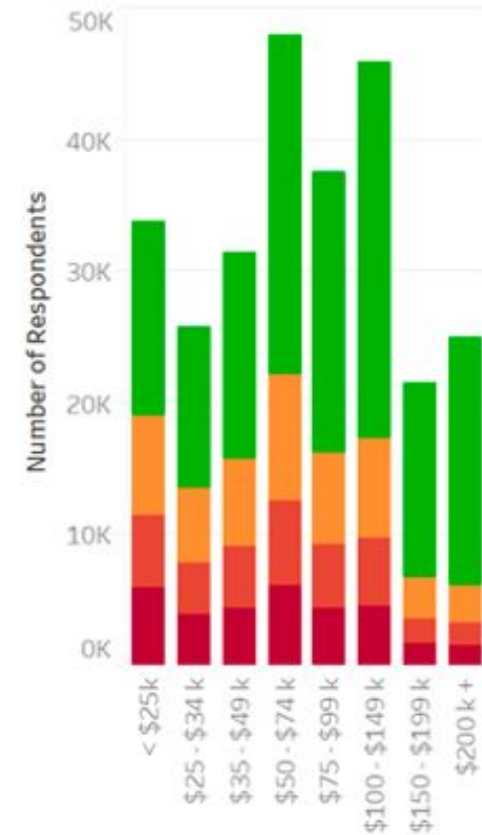
Results filter out those who have received the vaccine.



Household Income & Vaccination Intent



Distribution of Household Income



RECOMMENDATIONS & CONCLUSION





[This Photo](#) by Unknown Author is licensed under [CC BY-SA-NC](#)

- Extreme gradient boosting for predicting vaccination intent
- Cloud based computing



- A vaccination education project
- Employee retention and recruitment strategies
- Mental health public service campaigns
- Government contractor projects that assist with public health initiatives, such as food security



Thank you!



References

- Chen, L. (2019, January 2). *Basic ensemble learning (random forest, AdaBoost, gradient boosting)- step by step explained*. Towards Data Science. Retrieved November 17, 2021, from <https://towardsdatascience.com/basic-ensemble-learning-random-forest-adaboost-gradient-boosting-step-by-step-explained-95d49d1e2725>
- Coleman-Jensen, A., Rabbitt, M.P., Gregory, C.A., & Singh, A. (2021). *Household food security in the United States in 2020 (ERR-298)*. U.S. Department of Agriculture, Economic Research Service. <https://www.ers.usda.gov/webdocs/publications/102076/err-298.pdf?v=8240.1>
- Housing Assistance Council. (2020, October 5). *Covid-19 cases surpass 900,000 in rural America – 20,000 deaths*. Rural Home. Retrieved November 17, 2021, from <https://ruralhome.org/covid-19-cases-surpass-900000-in-rural-america-20000-deaths/>.
- Niles, M.T., Beavers, A.W., Clay, L.A., Dougan, M.M., Pignotti, G.A., Rogus, S., Savoie-Roskos, M.R., Schattman, R.E., Zack, R.M., Acciai, F., Allegro, D., Belarmino, E.H., Bertmann, F., Biehl, E., Birk, N., Bishop-Royse, J., Bozlak, C., Bradley, B., Brenton, B.P., ...Yerxa, K. (2021). *A multi-site analysis of the prevalence of food security in the United States, before and during the COVID-19 pandemic*. MedRxiv. <https://www.medrxiv.org/content/10.1101/2021.07.23.21260280v1>
- Terlizzi, E.P. & Zablotsky, B. (2020, September). *Mental health treatment among adults: United States, 2019*. (NCHS Data Brief, No. 380). National Center for Health Statistics. <https://www.cdc.gov/nchs/data/databriefs/db380-H.pdf>

