Feature Guide

- State
  - Categorical: Each state is assigned a numerical value from 1-50 in alphabetical order, additional values are provided for US territories.
- Race
  - Categorical: White only, Black or African American only, American Indian or Alaskan Native only, Asian only, Native Hawaiin or Pacific Islander only, and Multi racial. This variable was one hot encoded.
- Sex
  - Categorical: Male and Female. This variable was one hot encoded.
- Age
  - Ordinal: Ages were binned and assigned numerical values based on the following categories the following are listed in key:value pairs: 1:18-24 yrs old, 2:25-29 yrs old, 3:30-34 yrs old, 4:35-39 yrs old, 5:40-44 yrs old, 6:45-49 yrs old, 7:50-54 yrs old, 8:55-59 yrs old, 9:60-64 yrs old, 10:65-69 yrs old, 11:70-74 yrs old, 12:75-79 yrs old, 13:80+ yrs old.
- Level of Education
  - Ordinal: Level of education were binned with the following categories 1:Did not graduate high school, 2: Graduate High School, 3: Attended College or Technical School, 4: Graduated from College or Technical School.
- Income
  - Ordinal: Participants are binned into income brackets based on their household income 1: less than 15k, 2: 15k-25k, 3: 25k-35k, 4: 35-50k, 5: 50k-100k, 6: 100k-200k, 7: 200k+.

- Smoker Status
  - Ordinal: Participants are asked whether they currently smoke everyday (1), they sometimes smoke (2), they used to smoke (3), or if they have never smoked (4).
- Drug Use Variables (E-Cigarette Use, Alcohol Use)
  - Ordinal: Particpants are asked whether they have (1) or have not (2) used an e-cigarette in the past 12 months. For alcohol use participants are asked whether they have drank more than 14 drinks for men or 7 drinks for women per week. 1 if they do, 2 if they do not.
- Health Assessment (General Health, Physical Health, Mental Health)
  - Ordinal: Participants self evaluate their general health by rating their health on a scale from 1-5, with 1 being excellent and 5 being poor.
  - Continuous: Participants are asked to provide the number of days that they are physically healthy for the physical health feature or mentally healthy for the mental health feature.
- Health Access
  - Ordinal: Participants are asked whether they have (1) or have not (2) avoided medical care in the past 12 months due to the cost of treatment.
- Exercise Level

- ○ Ordinal: Participants ask whether they do (1) or do not (2) exercise outside of work regularly.
- Sleep
    - ○ Continuous: Participants are asked to provide an integer value for the average number of hours they sleep per night.
- Health History Variables(Heart Attack, Heart Disease , Stroke, Asthma, Cancer, COPD, Depression, Kidney Disease, Diabetes)
    - ○ Ordinal: Participants provide a binary yes (1) or no (2) based on if they have one of the listed health conditions. Each health conditions is its own feature.
- Urban Level
    - ○ Ordinal: Participants are asked whether they do (1) or do not (2) live in a metropolitan county.
- BMI
    - ○ Continuous: BMI is listed is calculated using the respondents weight in kilograms and height in meters to return the appropriate BMI value.
- Use of Food Stamps
    - ○ Ordinal: Participants are asked whether they have (1) or have not (2) received food stamps in the past year. This was dropped due to a large number of missing values.
- COVID Vaccination Status.
    - ○ Ordinal: This variable is coded as (0) has never received a COVID vaccine, and (1) has received at least 1 COVID vaccine in the past. This was dropped due to a large number of missing values.
- COVID
    - ○ Ordinal: Participants responded with yes (1) or no (2) to whether or not they have had COVID in the past. This was removed as all participants had COVID
- Pregnant
    - ○ Ordinal: Participants responded with yes (1) or no (2) to whether or not they are currently pregnant. No participant that responded to the Long-COVID question was pregnant at the time and this variable was removed.
- **Target Variable:** Long COVID
    - ○ A binary 0,1 variable representing if a respondent has (1) or has not (0) had Long COVID. This is the variable are algorithms are attempting to predict.