Group # 8

Charlie Barry (hbe2mx)

Emma Cooper (eyc4xd)

Aishwarya Pradhan (xaz3kw)

Christopher Lee (cl7zn)

INTRODUCTION

Question: What are some key factors leading to hesitancy for the COVID vaccine?

Hypothesis: Demographics (Proximity to medical care), Political Affiliation, Education, Religion, Age, Gender, Race, Socioeconomic Status, Occupation

THE DATA (*Everyone can edit up 4 pm - Group Meeting on 10/3/21 at 4 pm EST)*

https://github.com/owid/covid-19-data/tree/master/public/data/vaccinations/#united-states-vaccination-data

https://ourworldindata.org/us-states-vaccinations

Census Race Data

Precinct level Election Results

County Level Election Data

Interesting query system for medical data based on County

COVID-19 County Hesitancy

Starting dataset: COVID-19 County Hesitancy data (CDC)

Joined dataset #1: Census data

Joined dataset #2: Election data

Joined dataset #3: Healthcare data

All data joined on County (will likely have to use either Census tract or Fips Code)

EXPERIMENTAL DESIGN

- Take data and join into one dataset based on location (either zip or census tracts)
- Pre-processing and data clean-up
- Correlation between variables
- Regression analysis (Model for predicting outcome on new data)
- Machine Learning techniques (Random Forest, Clustering)
- Using ML techniques, identify variables of interest. Then create a model to predict outcomes on new data
- Visualizations

PROJECT MANAGEMENT

Milestones:

1. Identifying data sources
2. Pre-Processing data and combining
3. Initial Analyses to identify variables of interest
4. Model Creation
5. Model Validation
6. Write-up

Team Roles:

- Charlie
    - Submitter
    - Git Leader
- Aishwarya
    - Editor/Compiler
- Emma
    - Model Validation
- Christopher
    - Pre-Processing
- Whole group:
    - Combining data
    - Initial Analyses
    - Model Creation
    - Write-up

RESULTS

- Visualizations
- ML Model Outputs
- ML Demo Video

- Linear Model with use of test data (training/test split)
- PowerPoint Deck

TESTING

- Data will be split into two groups (66% training, 34% test)
- Models and variable selection will be done on training set
- Model validation done on test set
- Will use unit tests for python code

OUTCOME

- Goal is to combine data to effectively analyze populations, groups, and areas that may be predisposed to a higher level of vaccine hesitancy.
- This could be used by governments to focus education on the groups to elaborate on understanding of the vaccine, and alleviate any discomfort/hesitancy towards the vaccine

Summarize your plan and explain how your findings could be used by others (if applicable).