Aditya Koushik, Keane Haesle, Julie Young, Nathan Zhou

Methods

We collected data on factors that potentially relate to depression **by county in the US**. We normalized the data using the MinMax Scalar and dividing by total population.

We then analyzed the data using **Random Forest** and **Linear Regression.** We can then gain insight into which factors of 60 are the most correlated with depression. This information can then be used to inform policy and individuals' lifestyle choices.

### Correlation Heatmap (of all 60+ variables)
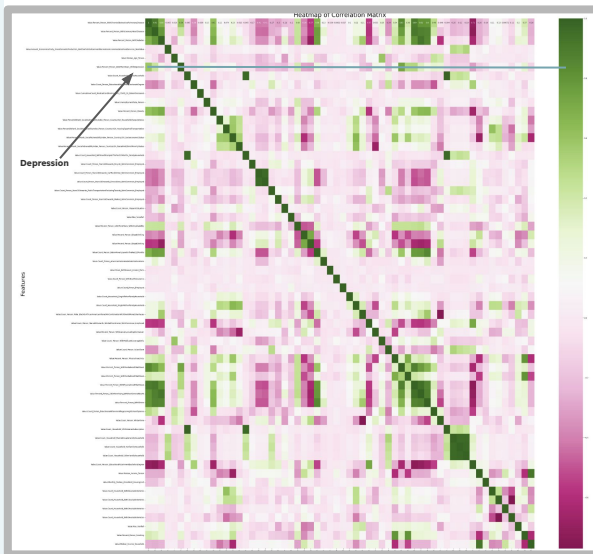
We trained a machine learning regression model to identify depression rates across counties based on the 60 variables we obtained from google common dataset.
Test Train Split **70%** | KNN (n_neighbors = 2): **60.82%**
Random Forest Algorithm Accuracy Score (n=500 estimators): **83.87%**
Accuracy of Linear Regression: **78.23%**

% of people with depression over 18 (per county)
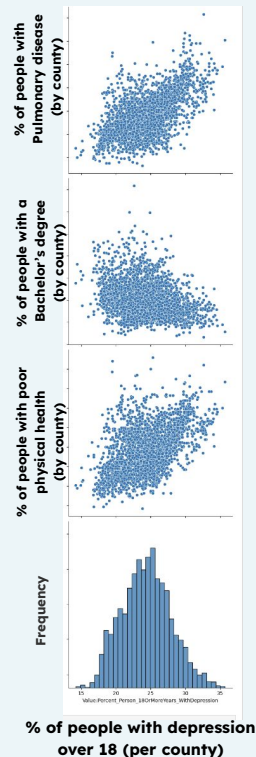
Implications/Action Items

- Develop program that integrates mental health support into routine care for patients with chronic conditions such as Chronic Obstructive Pulmonary disease
- Companies should offer more remote/hybrid options to support mental health.
- Educating citizens has implications beyond degrees and employment, but with mental health as it has a negative correlation.
- The heatmap as a whole can be useful to counties to understand not only factors related to depression, but how these variables relate to each other

Conclusions: Based on our data analysis, we found a strong connection between depression and Chronic Obstructive Pulmonary Disease (0.59), Poor Physical Health (0.51), Coronary Heart Disease (0.43) Disabilities (0.37) Percent With a Bachelor's Degree (-0.31), and interestingly, People Working From Home (-0.24), all with p value below 0.001. Machine learning models had strong performance.