



"Help people live healthier lives and help make the health system work better for everyone"

Team 6 Case Analysis on Preventable Hospitalizations



Why?

Roughly 15%-21% of
admitted hospitalizations
per year are a preventable
visit

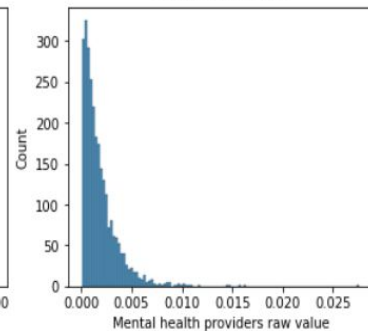
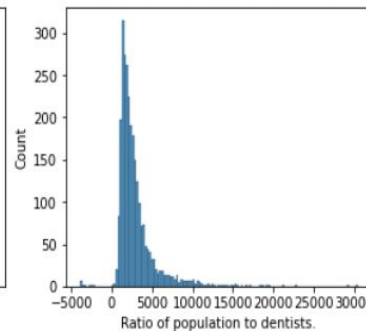
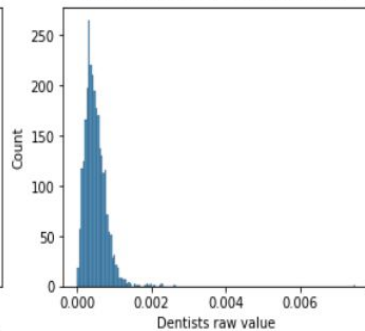
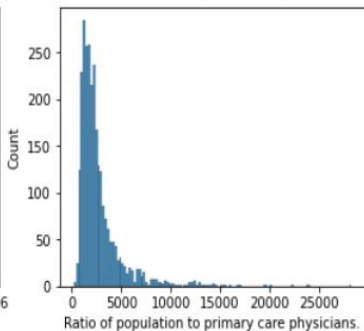
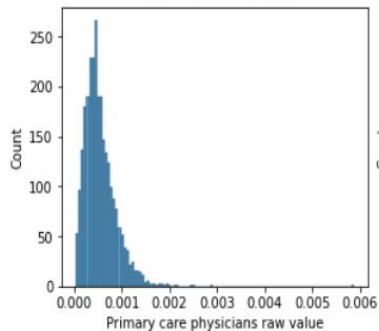
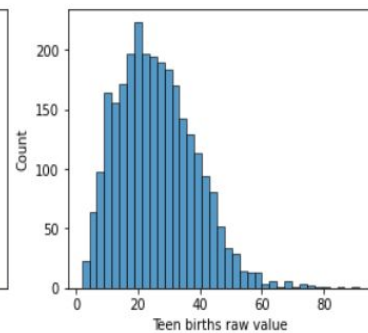
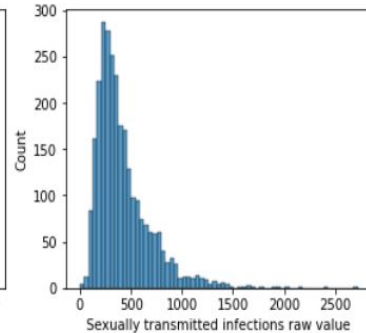
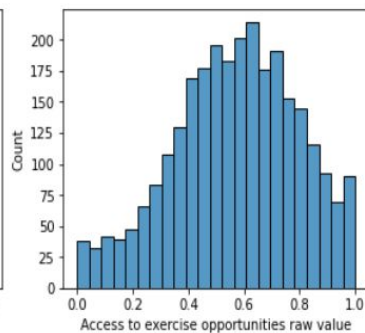
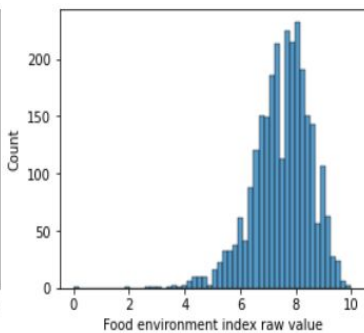
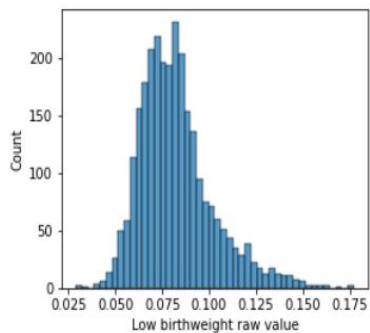
Maximize usage of
outpatient facilities

Consistent with
organizational goals

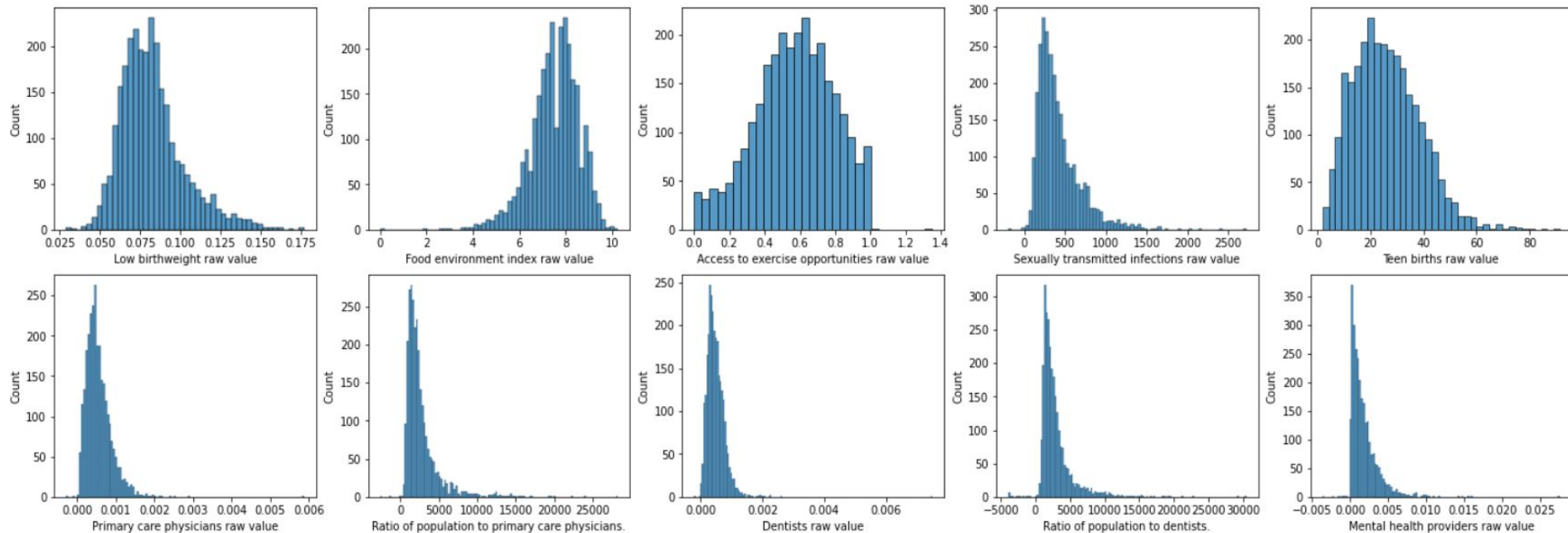
Data Cleaning

- Converting metrics from object type to numeric
 - Removed columns that didn't contain raw values (associated with main variables such as numerator, cihigh, etc.)
 - Got rid of columns that had over 33% data missing
 - Removed counties (rows) that had over 5 missing columns
 - Removed rows where the target class was NaN (Preventable Hospitalization)
-
- Original data set - 3195x725
 - Final data set - 2780x57

Data imputation - Looking at distributions for imputing values



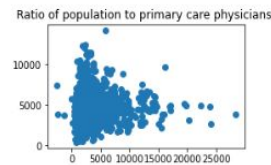
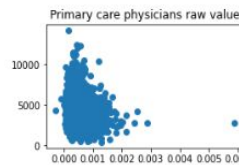
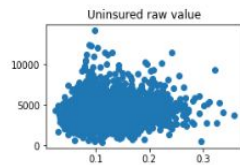
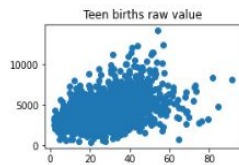
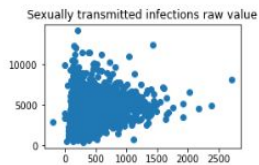
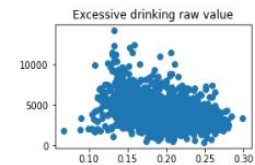
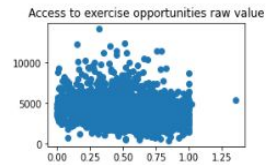
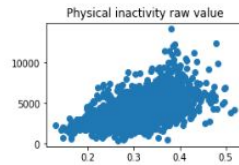
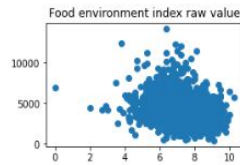
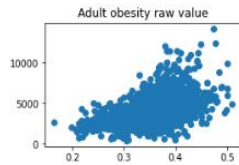
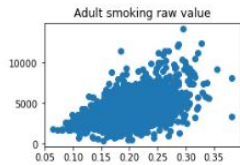
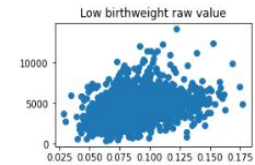
Post Imputation



Dropping columns based on low correlation(0.1)

```
#filtering the columns where absolute value of correlation with the target variable is greater than 0.1
#as that is almost no correlation and would not contribute to preventable hospitalization
data_corr = data.corr()
data_corr_filtered = data_corr[abs(data_corr['v005_rawvalue'])>0.1]
col_names = list(data_corr_filtered['v005_rawvalue'].to_frame().index)

data = data[['statecode', 'countycode', 'fipscode', 'state', 'county', 'year', 'county_ranked'] + col_names]
```



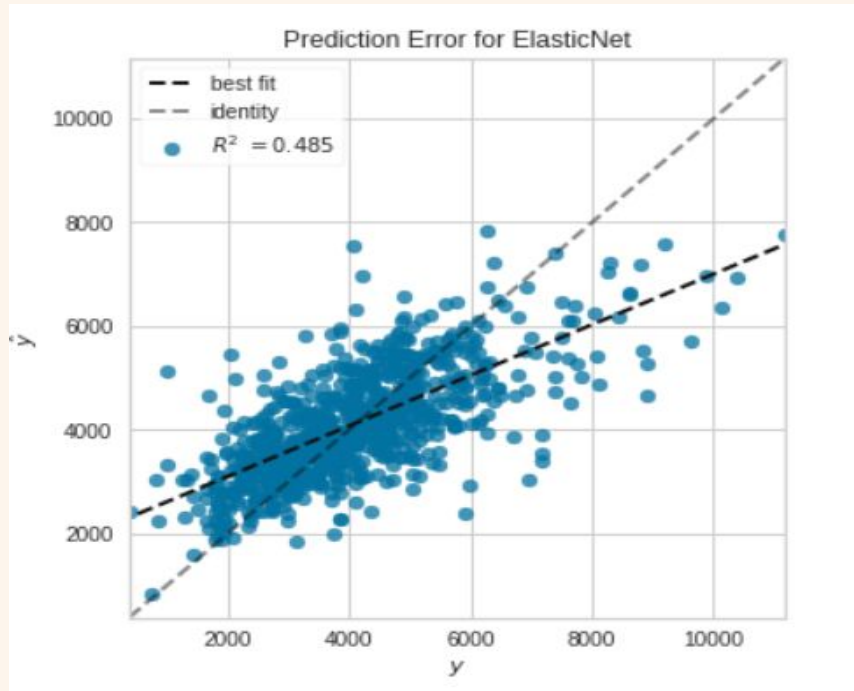
Modeling

OVERVIEW OF MODELING

- Fit Regression models to select the best model
- Utilize the best model to find feature coefficients
- The coefficients determine which features of the model contribute most to the target class
- Utilized Boruta algorithm for feature selection
- Comparing Boruta features with coefficient based feature selection

Regression Models trained

Best Model as per results - ElasticNet




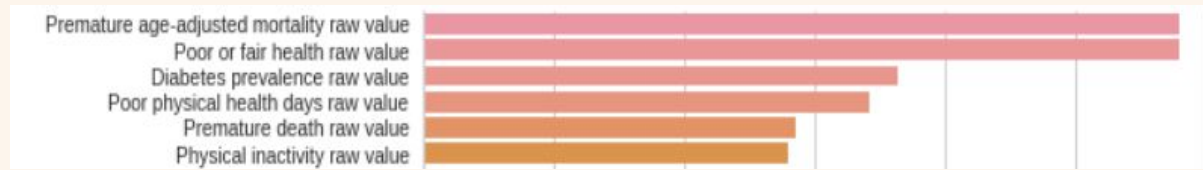
Model performance of Elasticnet on test
Test R2: 0.485
Test RMSE: 1076.801

Model performance of Lasso on test
Test R2: 0.484
Test RMSE: 1078.023

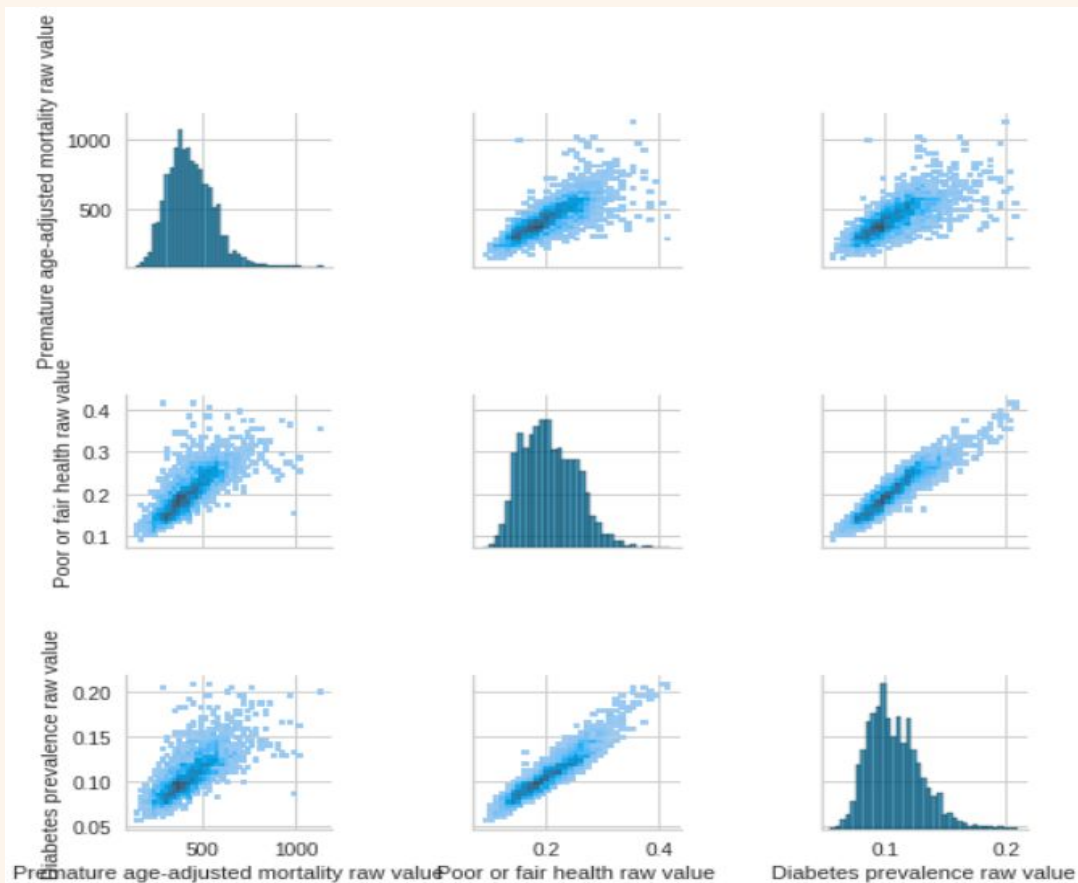
Model performance of Ridge on test
Test R2: 0.485
Test RMSE: 1076.801

Feature importance from the best model

	feature	coefficient	
35	Premature age-adjusted mortality raw value	579.868537	
1	Poor or fair health raw value	579.561185	
38	Diabetes prevalence raw value	363.262998	
2	Poor physical health days raw value	342.585874	
0	Premature death raw value	285.480733	
8	Physical inactivity raw value	280.160172	



Features correlation



**Boruta variables for features selection vs
coefficient analysis from regression model**

0



0 Poor or fair health raw value

1 Poor physical health days raw value

2 Poor mental health days raw value

3 Adult smoking raw value

4 Adult obesity raw value

5 Physical inactivity raw value

6 Sexually transmitted infections raw value

7 Teen births raw value

8 Ratio of population to mental health providers.

9 Unemployment raw value

10 Air pollution - particulate matter raw value

11 Driving alone to work raw value

12 Life expectancy raw value

13 Premature age-adjusted mortality raw value

14 Frequent physical distress raw value

feature coefficient



35 Premature age-adjusted mortality raw value 579.868537

1 Poor or fair health raw value 579.561185

38 Diabetes prevalence raw value 363.262998

2 Poor physical health days raw value 342.585874

0 Premature death raw value 285.480733

8 Physical inactivity raw value 280.160172

Top 3 Categories

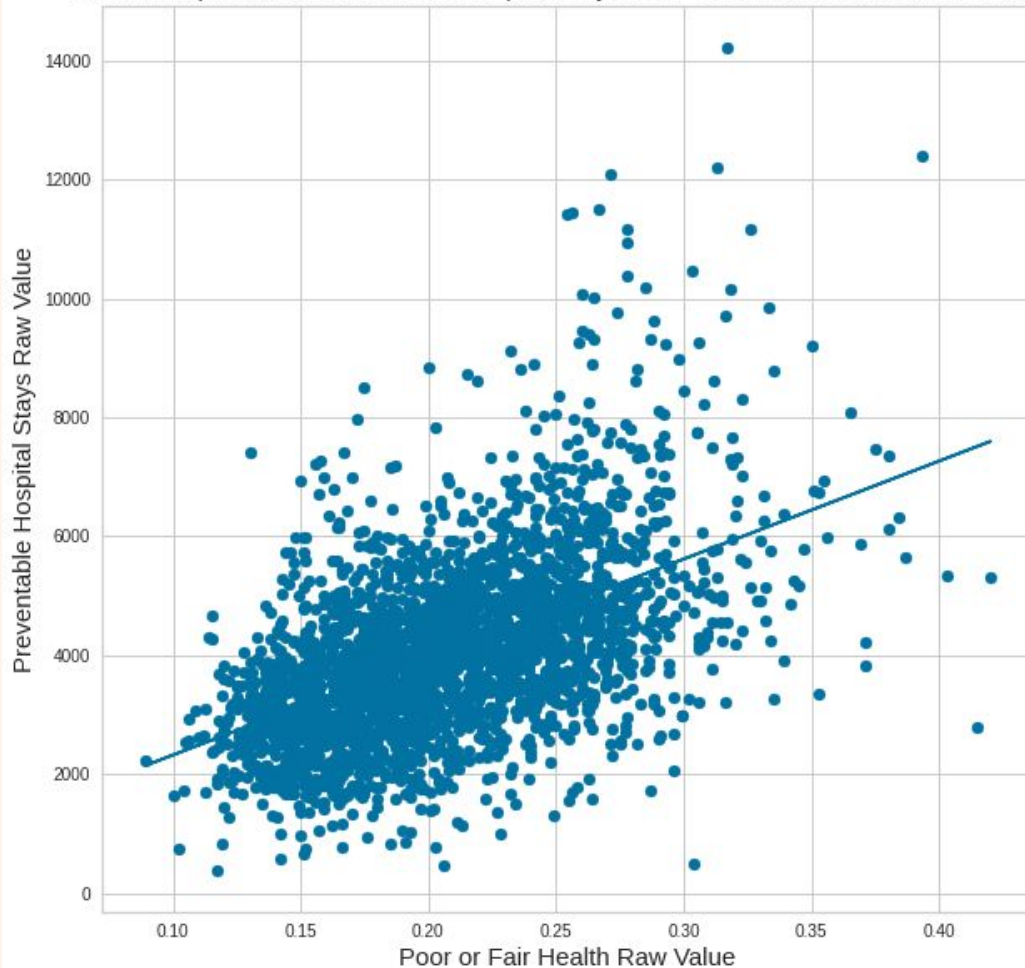
Premature Age Adjusted Mortality- Death that occurs before the average age of death for a certain population. In the United States its widely known at 75 years.

Overall Poor or Fair Health- Poor state of conditional health ailments.

Diabetes Prevalence- Related cases of hospitalizations that have to do with diabetes.

Exploratory Data Analysis

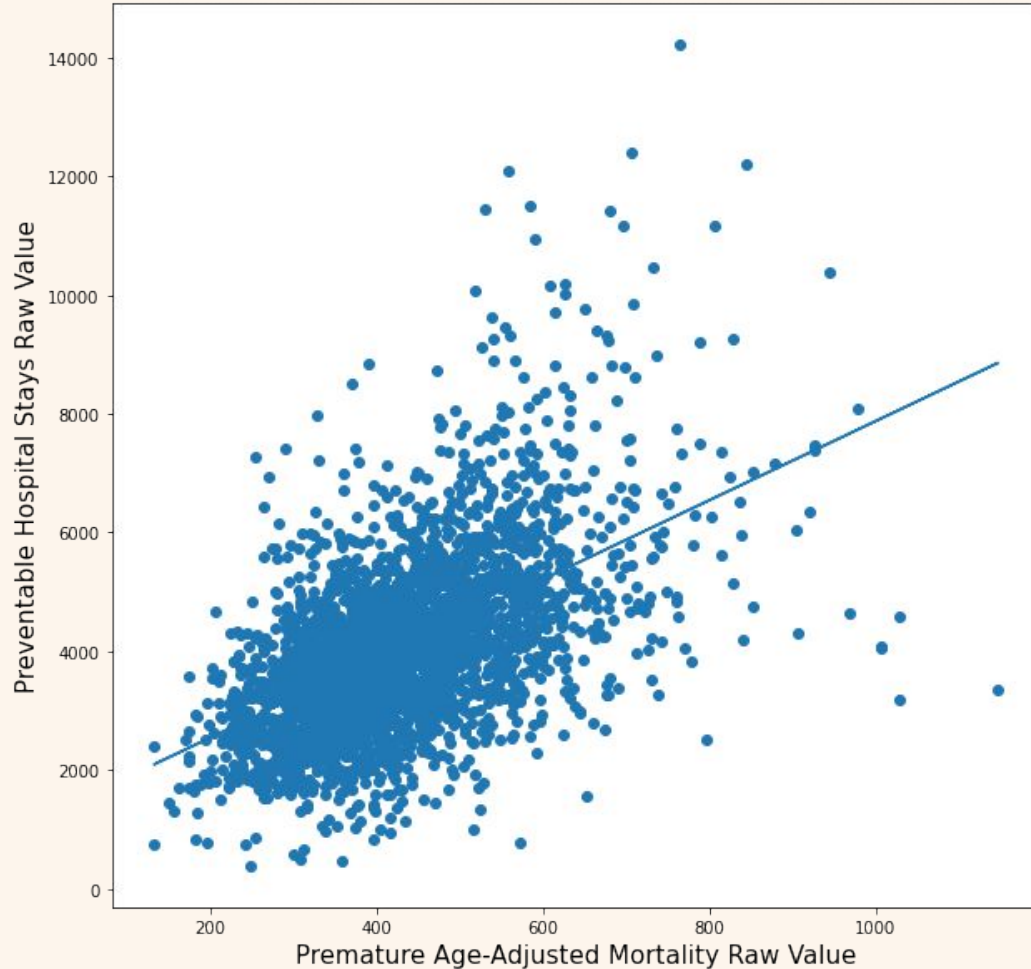
Relationship between Preventable Hospital Stays and Poor or Fair Health in US Counties



Correlation Coefficient = 0.53

Changes directly with
Preventable Hospital stays

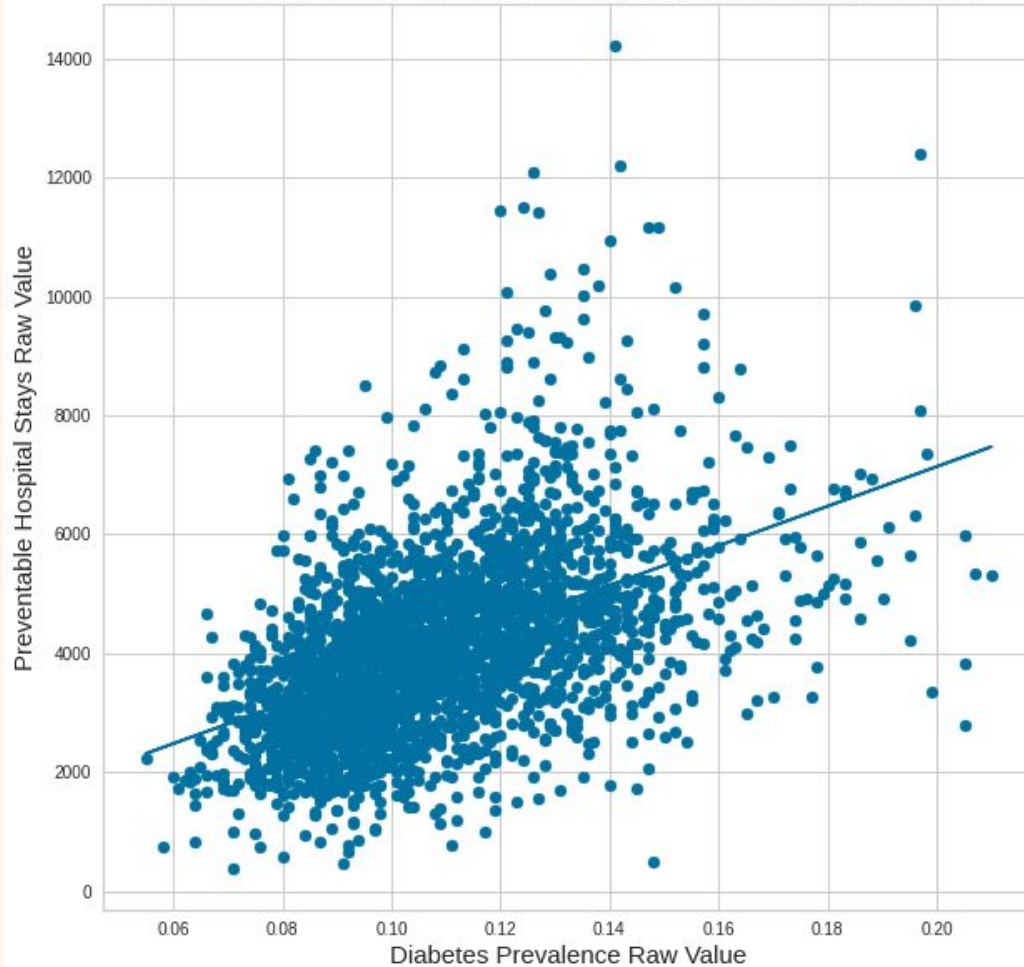
Relationship between Preventable Hospital Stays and Premature Age-Adjusted Mortality



Correlation Coefficient = 0.53

Changes directly with
Preventable Hospital stays

Relationship Between Preventable Hospital Stays and Diabetes Prevalence



Correlation Coefficient = 0.5

Changes directly with
Preventable Hospital stays

Solutions

Most Effective = Blanket Solution

Integrative Prevention Plans:

Implement preventative consultation between cross-sectional groups:

- Come together to establish high-risk patients and priority base them for treatment.
- Alleviate the pressure on hospitals because patients at most risk for hospitalizations are already being taken care of.

Create incentive plans for outpatient facilities and general practitioners:

- Financial or commending them, have outpatient facilities and general practitioners develop a more intimate relationship among patients.
- Address their concerns and problems in a preventive manner before something happens.

Thank you