

Behavioral Risk Factor Surveillance System (BRFSS): Prevalence Data (2011 - Present)

an analysis using regression techniques in a
Supervised Learning environment

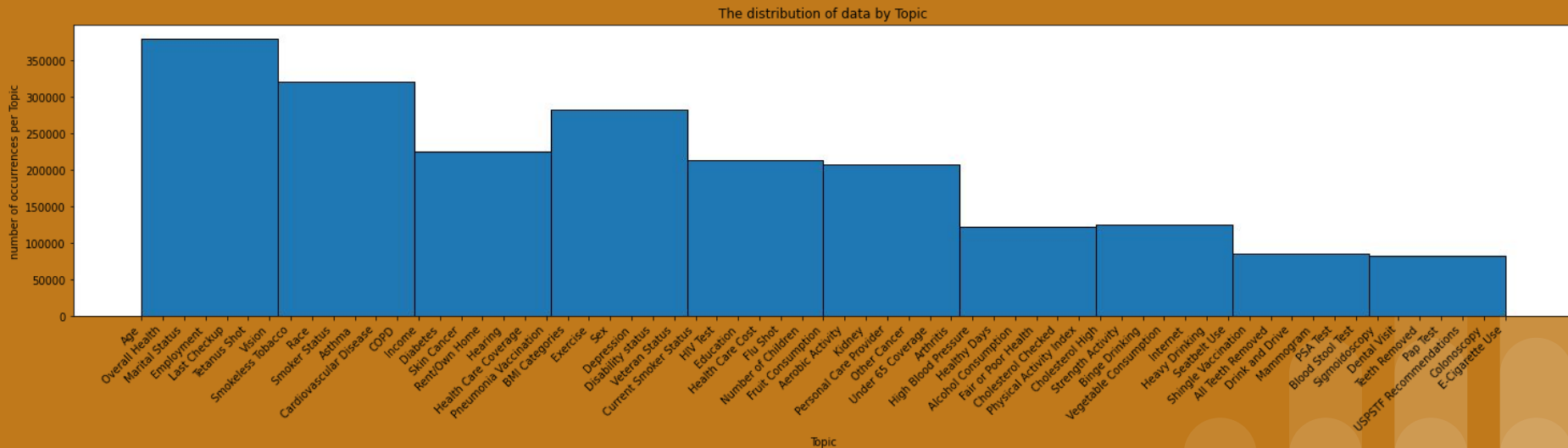




Overview and Problem Statement

- ❑ Chronic disease and health issues = always a problem !
 - ❑ While biological researchers do their work, we can do our own too
- ❑ The Behavioral Risk Factor Surveillance System (BRFSS) plays a major role in collecting data here
 - ❑ Well-established (since 1984), now nationwide, 400,000+ interviews per year
 - ❑ A great asset to have in our data analysis toolbox !
 - ❑ Extensive sponsorship (CDC and federal)
- ❑ Our Question Today: How impactful are less-than-medical factors on this study?
 - ❑ In particular, *Statistical* factors

Visualization of Data





Data Analyzed: [Data_value] Variable

- ❑ Target variable: Data Value
 - ❑ Overall score for each response recorded each time for each question
 - ❑ Lower score → healthy risk factor
 - ❑ Higher score → unhealthy risk factor
- ❑ Features:
 - ❑ Confidence limit high
 - ❑ Confidence limit low
 - ❑ Sample Size
 - ❑ Year
 - ❑ Display Order
 - ❑ Location ID
- ❑ Approximately 2,048,466 entries in the dataset, from 2011 - Present
- ❑ Focus: Confidence Limits
 - ❑ Most highly correlated features
 - ❑ Measure for amount of scope (similar to percent confidence intervals)
- ❑ [Raw data available here](#)
- ❑ [Complete project available here](#)

Cultural Fit: Outbreak → Epidemic → ?

- ❑ Diseases are always beginning, growing, and spreading
 - ❑ COVID pandemic
 - ❑ Monkeypox - current uncertainty
 - ❑ Past outbreaks throughout history
 - ❑ Vaccination to the rescue !
- ❑ The importance of chronic health risk behavioral factors
 - ❑ Ensuring we are as prepared and protected as possible !
 - ❑ Impact on mental wellness !



AIC & BIC

Results: Ordinary Least Squares !

Model	R ²	MAE	MSE	RMSE	MAPE	AIC	BIC
<i>OLS</i>	1.000	1.729e-12	3.094e-24	1.759e-12	inf	-1.051e+08	-1.051e+08
<i>OLS (split)</i>	1.000	8.890e-14	8.977e-27	9.475e-14	inf	-9.364e+07	-9.364e+07
<i>LinReg</i>	1.000	6.027e-14	6.517e-27	8.073e-14	inf	—	—
<i>Lasso</i>	0.999	0.01504	0.00052	0.02275	inf	—	—
<i>Ridge</i>	1.000	7.120e-13	3.955e-24	1.989e-12	inf	—	—
<i>Elastic Net</i>	0.999	0.01504	0.00052	0.02275	inf	—	—

Since the measures of accuracy for the models used appear to contradict themselves, it is not clear which model performed best. Nevertheless, it seems that Ordinary Least Squares (OLS) regression may have.



Thank You!

Questions?

(And any recommendations for discussion!)