We all have an interest in public safety and have created governmental agencies to act on our behalf – airline safety (FAA) --- food and drug safety (FDA) – highway safety (DOT/FHSA). Bridge safety (BTS/FHA). And yet most of us never think about bridge safety until there is a disaster such as the Florida bridge collapse in 2018 that killed six (<https://www.usatoday.com/story/news/2018/11/15/ntsb-miami-bridge-collapse-design-errors/2012020002/>) or the Silver Bridge collapse in Ohio (1967) that killed 46 (<https://en.wikipedia.org/wiki/Silver_Bridge>). The National Bridge Inventory maintained by the FHA lists 616,087 bridges. 7.63% are rated “structurally deficient.” (47052/616087) and need urgent repairs. Estimated cost to complete all needed bridge work is $171 Billion. <https://artbabridgereport.org/>

Americans cross these deficient bridges 178 million times a day. (<https://artbabridgereport.org/reports/2019-ARTBA-Bridge-Report.pdf>).

The FHA (Federal Highway Administration) National Bridge Inventory is huge:

Over 600,000 rows

Over 100 features

Most features have several sub-features.

As you can see, there are a lot of bridges and a lot of data. Most reports are like the ones mentioned above where they tell you how many or what percentage or the cost -- and may give a simple prioritization to the bridges. I will survey this data and apply several ML models and statistical tests to better understand the true state of our bridges.

I will do this by creating a model to predict which bridges are “structurally deficient.” I will run my model using FHA database. I will score it against the bridges already identified as deficient.

My audience will be the general public but, in particular, individuals and entities responsible for bridge maintenance and safety.

I will, hopefully, present a method to prioritize bridge evaluations that is more granular, and cost effective, than the current practice.

Math:

After EDA – data cleaning, outlier detection, feature engineering -- I will subject the data to several ML models including Linear Regression, Random Forest and XGBoost -- ultimately recommending the one with the highest accuracy on the test data.