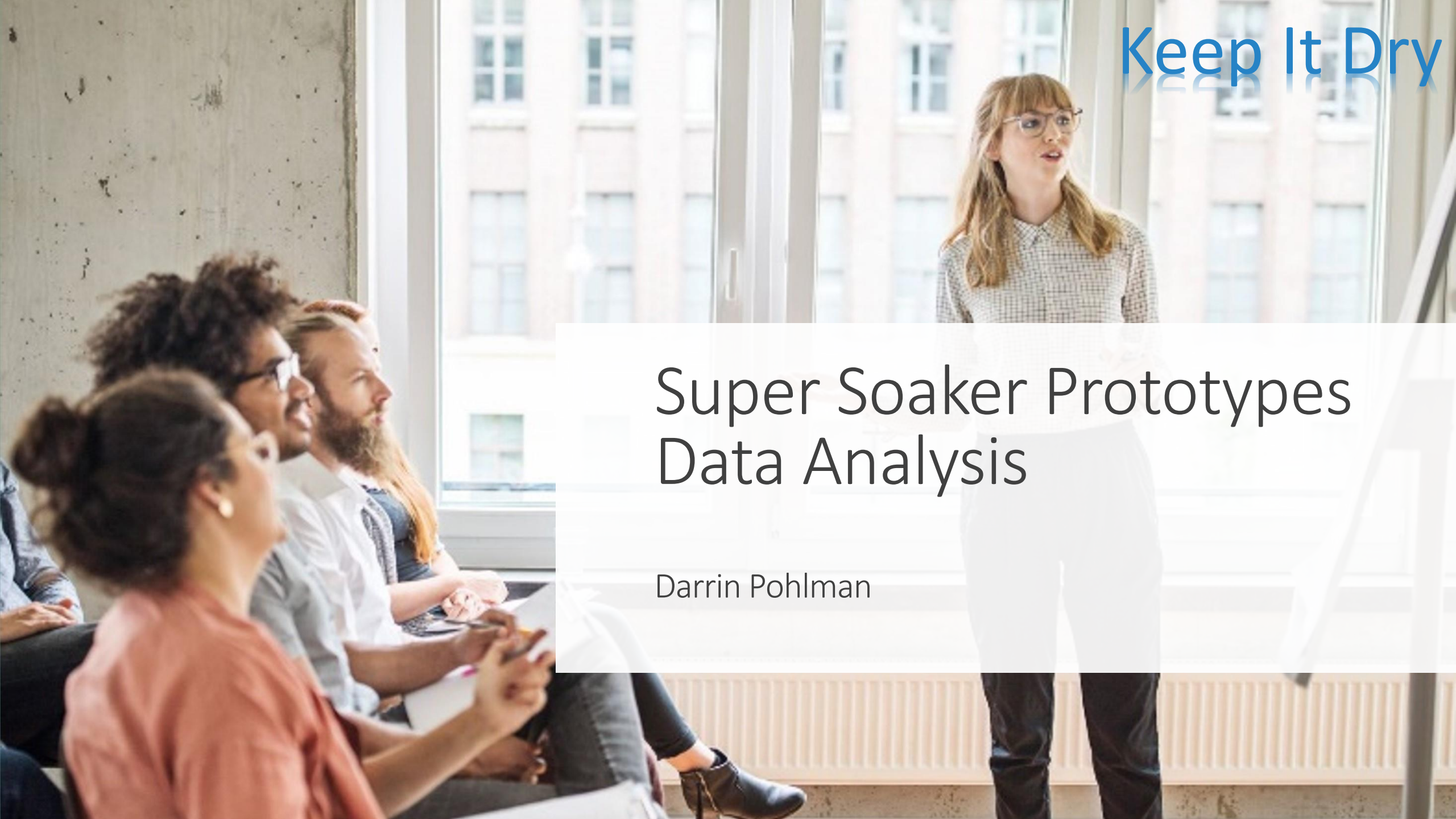


Keep It Dry

Super Soaker Prototypes Data Analysis

Darrin Pohlman

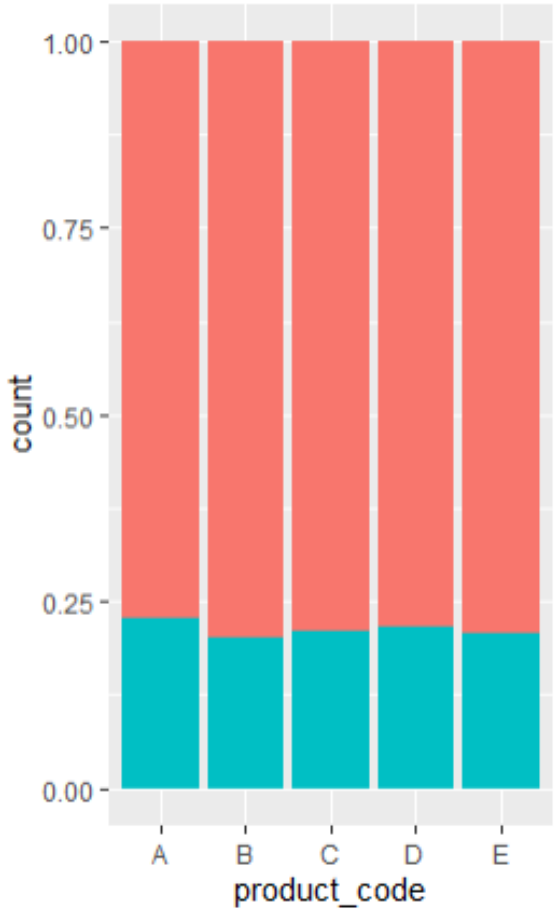
Summary of Train Dataset

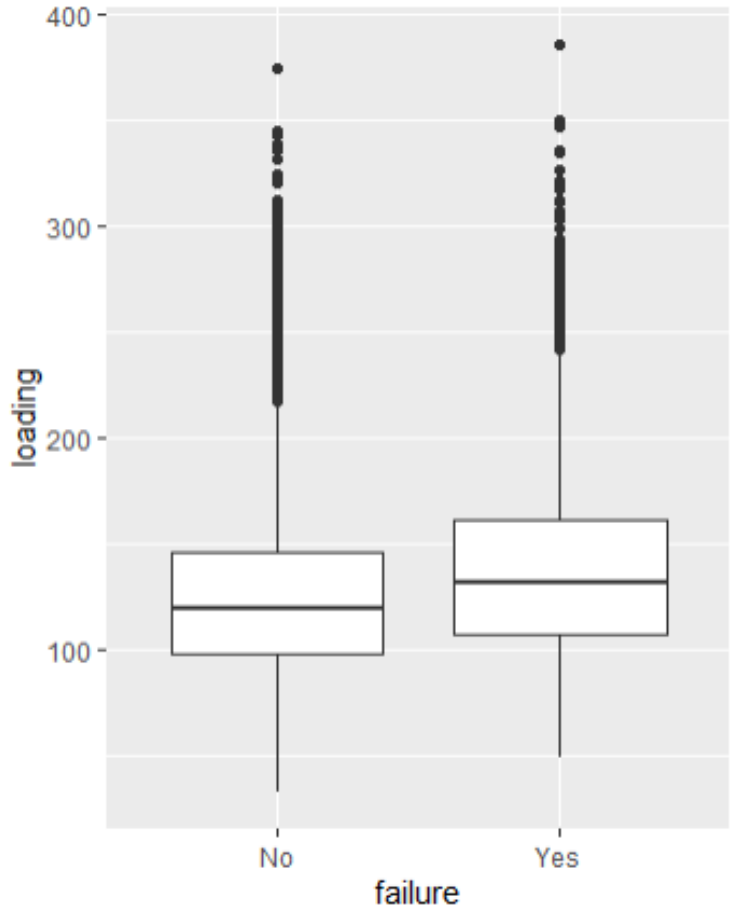
The train dataset contains 26,570 observations with up to 26 variables for each observation.

The failure rate by product code varies slightly but hovers between 23% and 24%.

The loading attribute has a slight impact on the failure rate but there are several other variables that impact failure.

Product_code and Loading may be strong predictors.



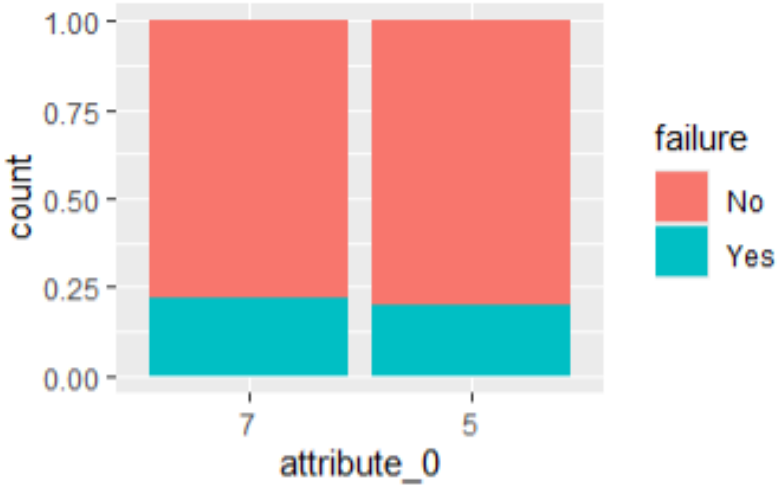


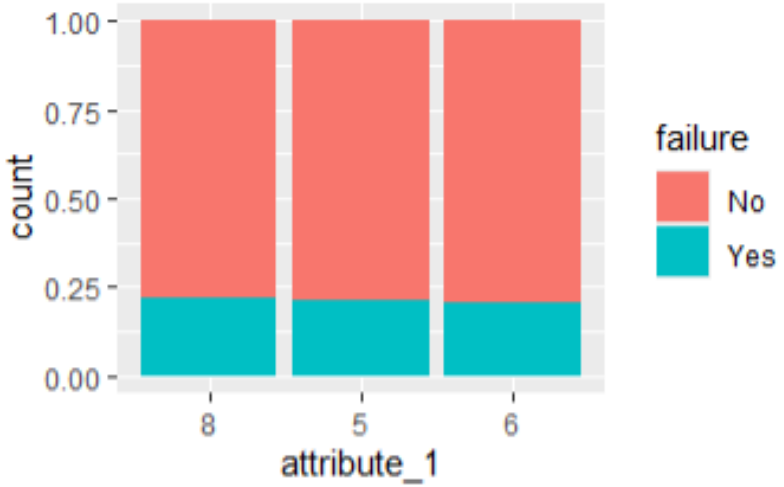
Keep It Dry

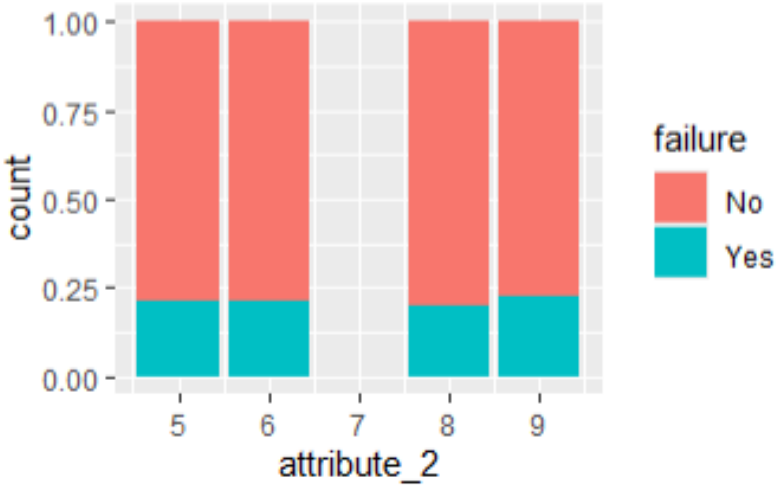
Failure Rate by Attribute

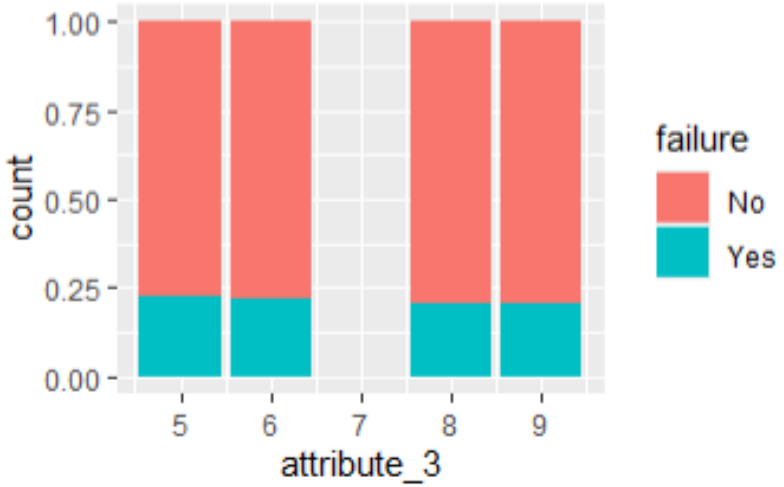
The rate of failure as measured by each attribute combination is less than 25% consistently. No attribute combination appears to perform significantly better than others.

Attribute_2 and Attribute_3 include material 7 but no direct connection to failure.







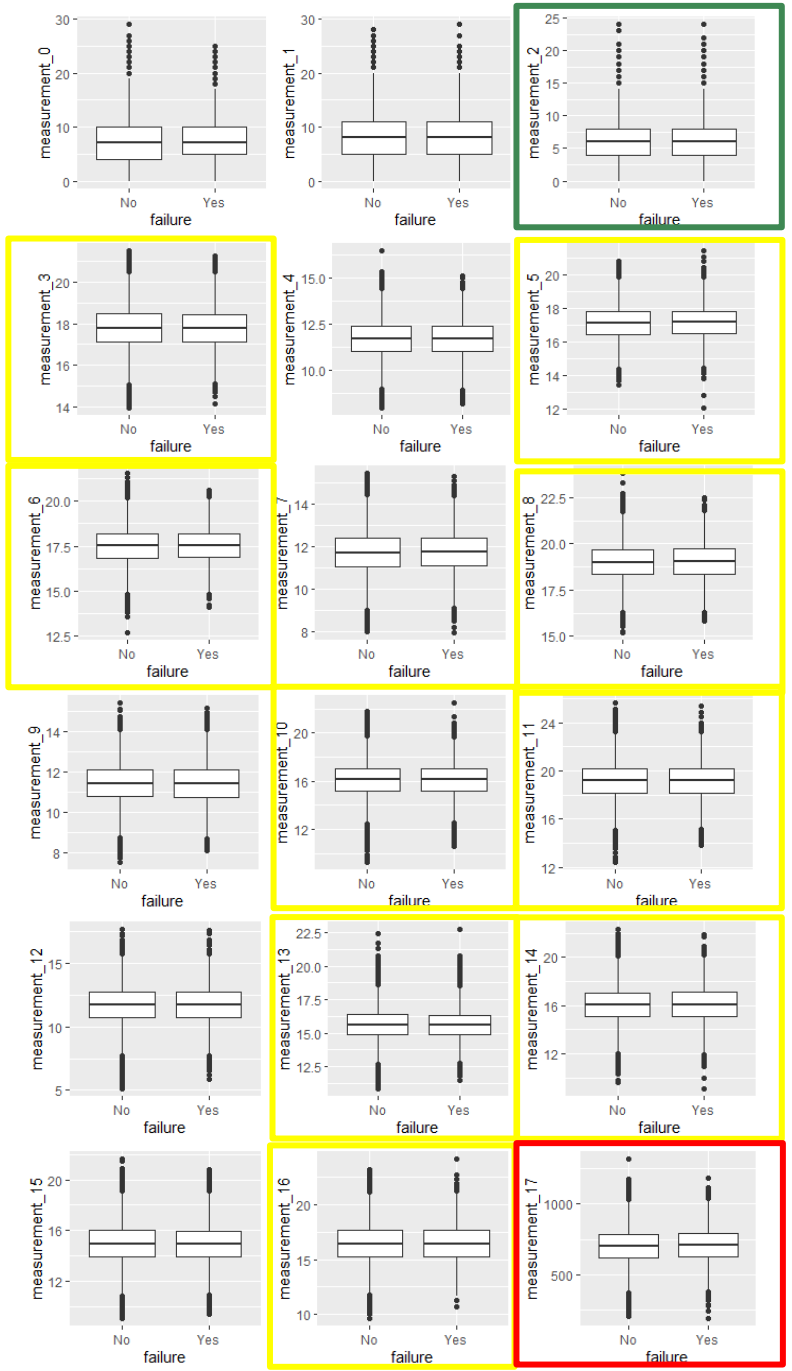


Keep It Dry

Failure Rate by Measurement

Each measurement includes a similar median range for both success and failure. The scale for each measurement varies by color with the most common mean between 16 and 17.

Measurement_17 may be a strong predictor.



Small Range

Common Range

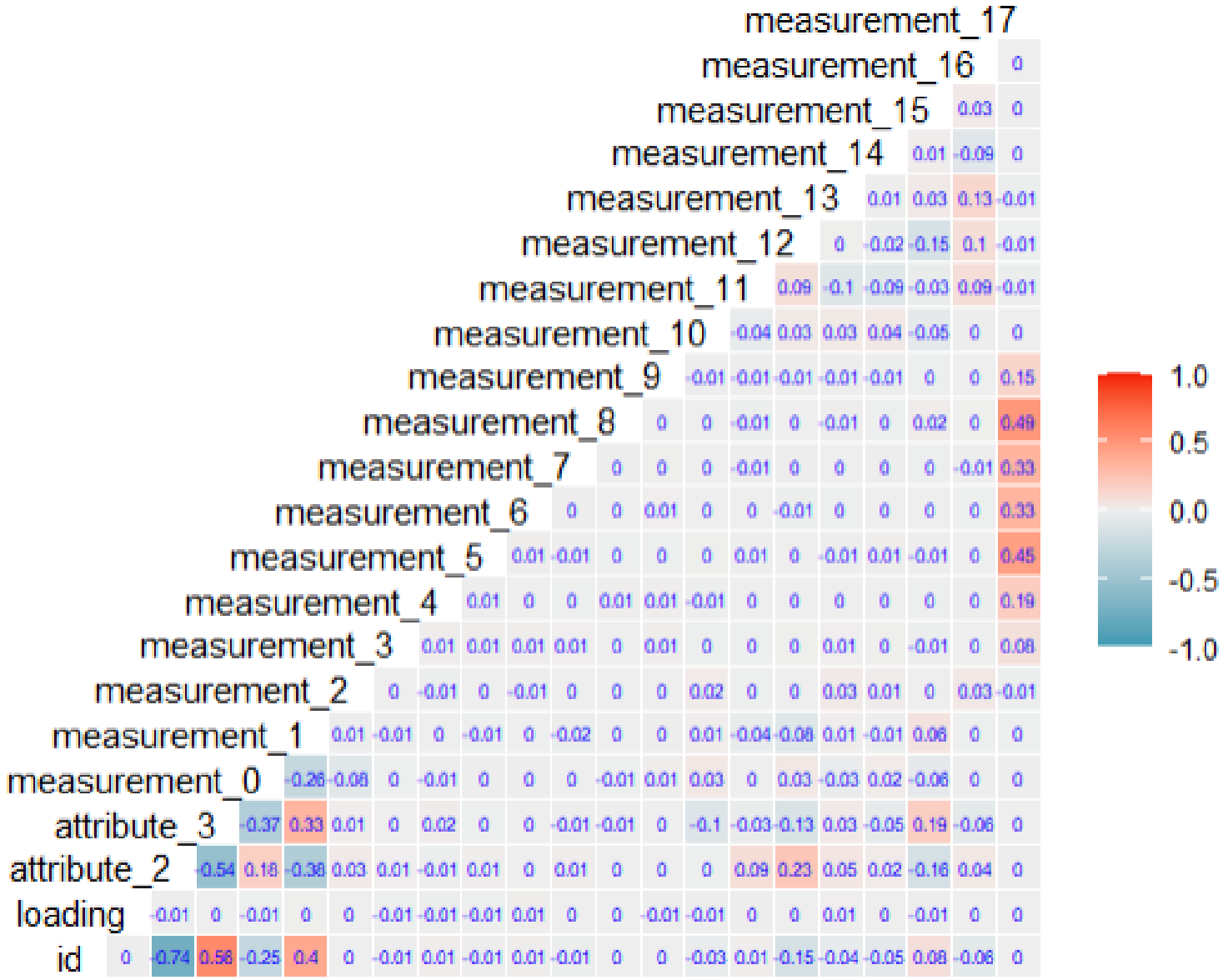
Large Range

Keep It Dry

Correlation Matrix View

This diagram shows a pairwise correlation between data attributes. Pairs that are darker colored have a higher correlation than faint pairs.

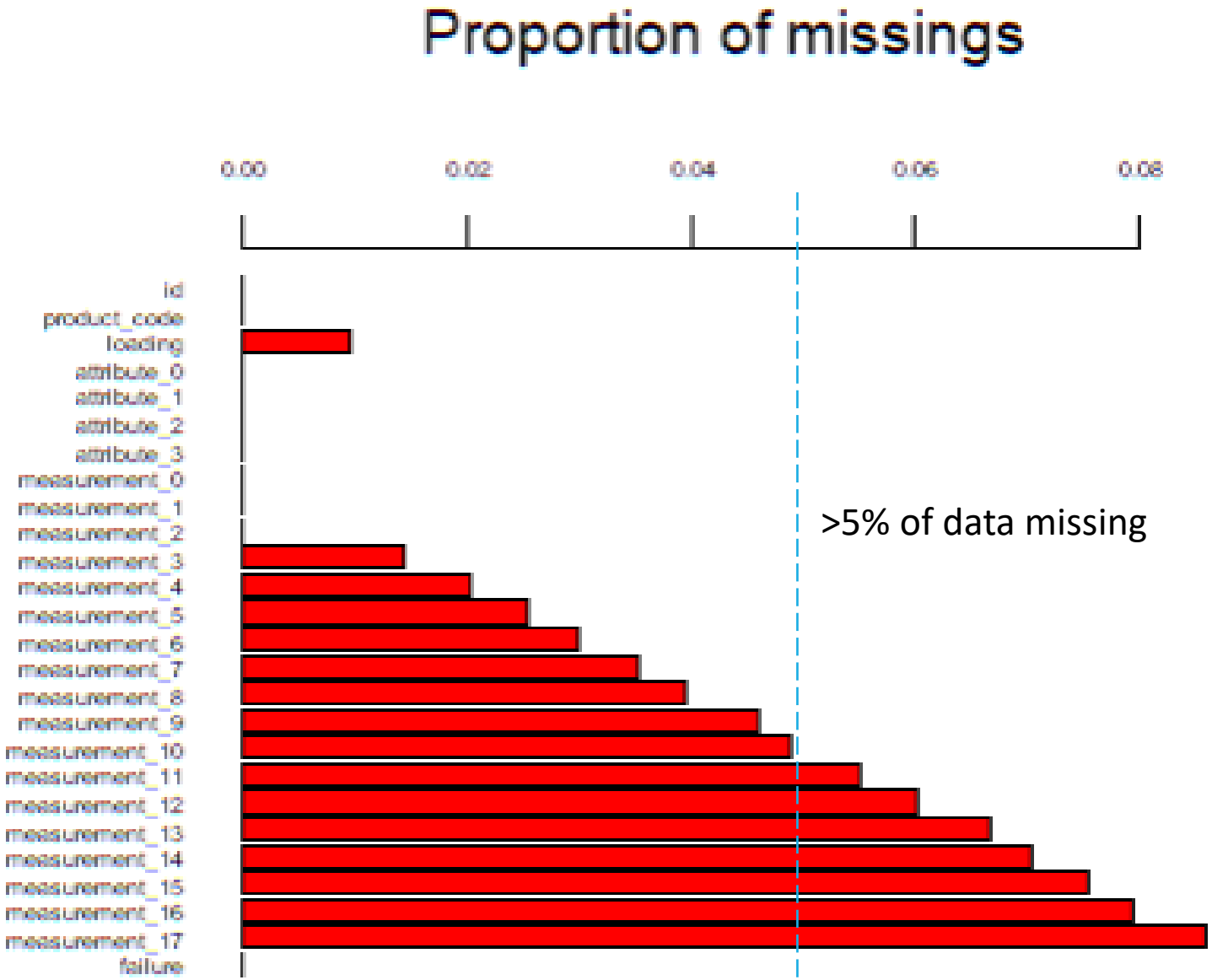
Attribute_2 and Attribute_3 may be strong predictors.



Keep It Dry

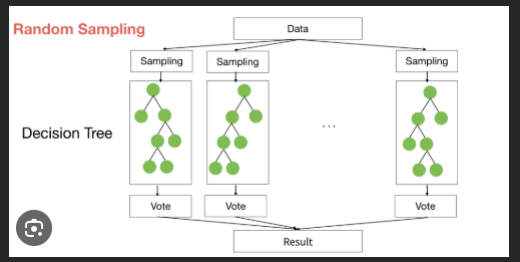
Managing Missing Data

The prototype testing results lack a large volume of measurement data. We recommend removing and ignoring any measurement from the predictive model that is missing more than 5% of the results from the training dataset. For data attributes missing less than 5%, we will create reasonable data to complete the model.



Keep It Dry

Steps to create a viable predictive model



1. Handle missing data
2. Determine variable importance
3. Create Multiple Models
4. Test / Tune Accuracy
5. Predict using Test Data
6. Iterate and adapt