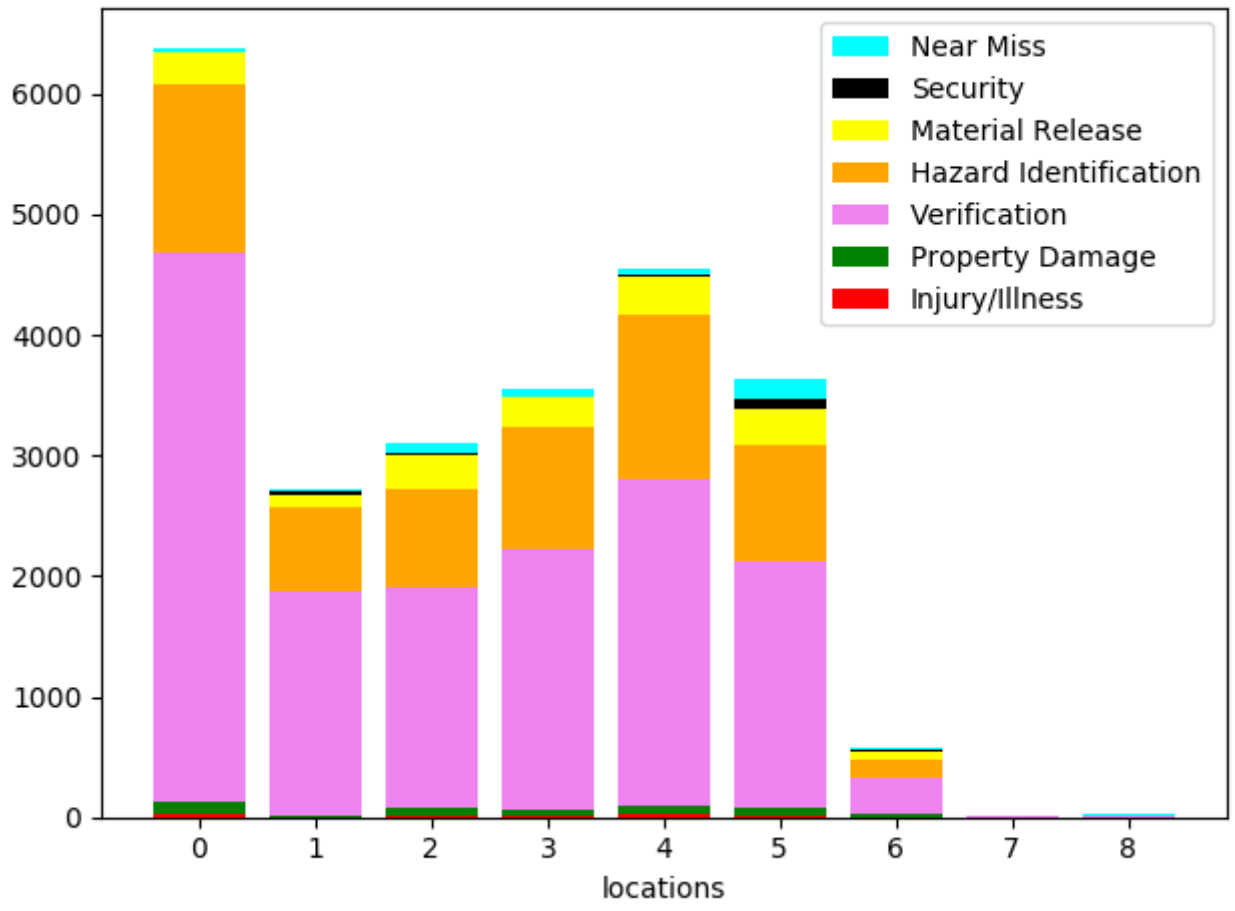
Stay Safe, BP

Anne Morrow

The Data:

The data are SafetyPlus submissions and incident reports from BP well sites. Although there are hundreds of sometimes redundant columns, the features include the date/time of incidents as well as general type (fire/injury/vehicular) and a free-text description. The SafetyPlus data app has changed versions with the older version having more data, so for a quick breakdown of the types of incidents, I made the following chart (with the location names removed for privacy reasons):



Using geolocation data, I can also include information about the weather.

--

The Question:

Most of the event types in the above chart are relatively benign: Near Miss, Hazard Identification, and Verification. Less benign are Material Release, Property Damage, and worst of all Injury/Illness. Were there clues in the benign events that could have helped predict the risk of something truly bad happening? What incident reports in the future should raise red flags?

MVP:

There is a world of EDA in front of me, but one of the first things I'm going to do is classify the incidents as within one week of a bad event or not and see if any supervised learning technique can separate these classes even a little. I also plan on doing natural language processing on the comments to see which benign events had the most in common with (and therefore were perhaps partially responsible for) the bad events. I will also investigate whether I can apply time series machine learning.

And More:

Hopefully, I'll be able to predict the probability of an upcoming bad event and/or make a filter that will look at an incident report and flag it as urgent or not.