Problem: Forecasting Harmful Algae Blooms

Harmful algae blooms occur when algae grows out of control and produce harmful toxins. These toxins contaminate the air we breath, the water we drink, and the local ecosystems. Ultimately, harmful algae blooms can impact the economy as people become sick and local marine life die. While we don't completely understand the causes, we know that warmer bodies of water are a major contributor to the increased frequency and severity of the blooms. As part of a constant pursuit for public safety, it is important to predict these harmful algae blooms so the public can prepare and avoid incoming danger.

We have data from the National Oceanic and Atmospheric Administration (NOAA) that records the state and local region of the bloom, the cell count, the salinity and temperature of the water, the wind direction and speed, and the date the measurements were taken. The wind direction and speed are the least consistent features with many empty entries. The water salinity and temperature, which we believe is vital for the forecast, is missing some entries but is a much more reliable feature than the wind speed/direction. The cell count and the date of measurement, which are the bare minimum needed for the forecast, are consistent throughout the dataset.

Stakeholders:

- NOAA
- People that consume from a water source susceptible to algae blooms
- People in proximity to the coast/lakes
- State Tourism Board
- Fish and Wildlife Department

Key Performance Indicators:

- Predicting/Forecasting an algae bloom accurately
- Build the model on more reliable measurements
- Predicting the severity of the bloom (duration of bloom)
- Predicting on location as well?,