

Abie Safdie

CS 670 Project Proposal

- 1) **Data:** I plan to use the CalCOFI Oceanographic dataset (<https://www.kaggle.com/datasets/sohier/calcofi/data>). This dataset has 60 years of ocean data off the coast of California. It has information stored such as temperature, salinity, oxygen, phosphate, nitrate and nitrite, chlorophyll, plankton biomass and biodiversity.
- 2) **Questions:** I plan to answer questions such as: how does climate change and rising ocean temperatures affect plankton biomass and the health of marine ecosystems. I believe that plankton biomass will be a good indicator of the health of these ecosystems as they are a keystone species. Further, I plan to ask more questions regarding how salinity and other ocean health factors affect ecosystems. I plan for these questions to develop and become more robust as I play around with the dataset.
- 3) **Models:** I plan to fit some multiple linear regression models to the data. I plan to use KNN as well and try to receive insight from that. Overall though, I plan to see what fits best as I learn more techniques in lecture and as I play with the dataset.

Overall, I think this data should bring about some insights and I'm looking forward to working with it! :)