Data 602
Final Project Proposal
Maxfield Raynolds

Proposal A:

This is a proposal to use a dataset from the company I currently work for. I am awaiting a decision as to whether the dataset can be used or not. If it cannot be used, proposal B will be executed. Due to waiting for a decision no EDA or summary statistics are currently available.

- Research Question:
    - Utilizing historic periodic growth rate data what possible models can be produced to help predict final head weights of lettuce at harvest?
        - Possible models that may be explored:
            - Curvilinear regression
            - Machine learning
- Justification:
    - Being able to predict final head weights at harvest allows for more accurate prediction of product available for sale, allowing for a more secure supply chain and directives for how the sales department should operate.
    - Additionally, prediction of final head weights helps evaluate the performance of the current growing strategy and may help provide indicators of possible adjustments to the strategy
- The data has been collected by myself over the last year as weekly head weight samples as well as data on monitored environmental factors such as:
    - Temperature
    - Humidity
    - Light
    - CO_2
    - Electrical conductivity
    - pH
    - Root length, color, and structure
- Libraries potentially being used
    - Pandas
    - NumPy
    - MatPlotLib
    - SciKit-learn

Data 602
Final Project Proposal
Maxfield Raynolds

- Research question:
    - Can recorded measurements of environmental factors be used to predict plant health?
        - Utilizing a dataset of measured environmental factors combined with a qualitative judgement of the plant's stress level at the time of measurement can the plants stress level be predicted?
- Justification
    - In modern controlled environment agriculture, an extraordinary amount of data is recorded about the environmental factors, from temperature, to humidity to light levels and more. Much of a grower's time is spent interpreting and adjusting influences on the factors to encourage growth and maintain plant health. Analyzing data related to the environment and the plant's health will help establish if there are specific threshold ranges that need to be maintained and possibly provide a framework for predicting plant stress.
- Data Sources
    - A kaggle dataset on plant health will be used for this project. The dataset is available here: https://www.kaggle.com/datasets/ziya07/plant-health-data
- Libraries potentially being used
    - Pandas
    - NumPy
    - MatPlotLib
    - SciKit-learn
- Summary Statistics:
    - For summary statistics please see associated Jupyter Notebook File: data602_final_project_proposal_B