Introduction

* Intent of the application
  + To explore the Iris Dataset
* Dataset to be used, including source
  + Iris Dataset
  + Source: [Iris Dataset | Machine Learning, Deep Learning, and Computer Vision (ritchieng.com)](https://www.ritchieng.com/machine-learning-iris-dataset/)
* Use case
  + A biology student stumbles over a ginormous Iris and wants to seek revenge. Since the student doesn’t fully recall what species of Iris they stumbled over, they go out into the field to collects 50 samples from 3 different species of Iris (150 samples total) and analyze the data to narrow down the culprit.

Dataset Analysis

* Define variables
  + sepal length, sepal width, petal length, petal width
  + 'sepal length (cm)', 'sepal width (cm)', 'petal length (cm)', 'petal width (cm)'
* Define labels
  + "0": setosa
  + "1": versicolor
  + “2": virginica

Inputs

* Data import
  + Iris dataset
    - https://archive.ics.uci.edu/ml/datasets/iris

Proposed Libraries

* Libraries
  + Pandas
  + Pandas-Profiling
  + Matplotlib
  + Scikit-learn
  + NumPy
  + Seaborn
* Library source
  + https://pandas.pydata.org
  + <https://github.com/ydataai/pandas-profiling>
  + <https://matplotlib.org>
  + https://scikit-learn.org/stable/
  + <https://numpy.org>
  + https://seaborn.pydata.org

Proposed Solution

* Functional description
  + Import the dataset and explore and display the features of the dataset.
  + Create a scatter plot of the sepal length verses the sepal width.
  + Add the petal length as a third dimension to the plot.
  + Explore the dataset using tools and libraries available in Python.
  + Calculate the dot product between the sepal length vector vs the sepal width vector.
  + Calculate the norm of the sepal length vectors.
  + Calculate the distance between the sepal length vector and the sepal width vector

Proposed Outputs

* Define application outputs
  + Dot product between the sepal length vector vs the sepal width vector
  + Norm of the sepal length vectors
  + Distance between the sepal length vector and the sepal width vector

Proposed Visualization

* Define visualization of outputs
  + Features of the dataset
  + Scatter plot of the sepal length verses the sepal width
  + EDA of the dataset

Conclusions

* Analysis of results
  + We expect to find that each species of Iris will have a correlated length and width of the sepal and pedals and that the culprit of the species of Iris with the most extreme length and width of the sepal and pedals