Introduction

1. Data Exploration: Initially this application will import the iris dataset and visually display features of it for analysis. Visualizations will include a scatter plot of sepal length versus the sepal width.
   1. Dataset to be used: The Iris dataset imported from the Seaborn Python Library from [seaborn: statistical data visualization — seaborn 0.12.2 documentation (pydata.org)](https://seaborn.pydata.org/)
   2. Use Case: Students can use this application to explore the iris data set and get an idea of what data is contained in it. Also students will understand what challenges may present themselves when performing more complex operations on it such as clustering and regression.

Dataset Analysis

* Variables: Sepal length, sepal width, petal length, and petal width
* Labels: Setosa, versicolor, and virginica

Inputs

* Data import from the Iris Data Set contained in the Seaborn Python Library.

Proposed Libraries and Their Sources

* Seaborn: <https://seaborn.pydata.org/>
* Pandas: [pandas - Python Data Analysis Library (pydata.org)](https://pandas.pydata.org/)
* Numpy: [numpy · PyPI](https://pypi.org/project/numpy/)
* Matplotlib: [Matplotlib — Visualization with Python](https://matplotlib.org/)
* Pandas Profiling: [ydata-profiling · PyPI](https://pypi.org/project/ydata-profiling/)
* Sklearn: [scikit-learn: machine learning in Python — scikit-learn 1.2.2 documentation](https://scikit-learn.org/stable/index.html)

Proposed Solution

* This application will explore the Iris dataset using visualizations and by performing mathematical operations on the data such as calculating the dot product and norm of sepal width.
  + Import the dataset
  + Display description of data
  + Display first 10 lines of data
  + Create scatter plot of sepal length against sepal width and add petal length as a third dimension.
  + Calculate distance between sepal length vector and sepal width vector
  + Using Numpy mathematical calculations will be performed on the data.

Proposed Outputs

* Outputs will include:
  + Scatter plot visualizations
  + Dot product between sepal length vector vs sepal width vector
  + Norm of the sepal length vectors
  + Distance between the sepal length vector and the sepal width vector

Proposed Visualization

* Scatter plots and what you would like.

Conclusions

* Your observations.