AP4 Report

This dataset was created using the measured geometric properties of three types of wheat, gathered from the Catholic University of Lublin. The dataset contained 203 points of data that cover a number of categories as well as what type of wheat it falls under. The three cluster k-mean was created using the KMean function of sklearn.

The k-mean cluster centroids are the center-most cluster with the smallest error. The data point errors are compared with the nearest cluster, the one with the smallest error is selected, and the process is then repeated for however many times is specified. The k-means accuracy turned out to be around 81.4%.

The centroids are as follows:

**Variety 1**

[('Area', 14.653088235294117), ('Perimeter', 14.471617647058823), ('Compactness', 0.8781205882352942), ('Length of Kernel', 5.573558823529412), ('Width of Kernel', 3.2738529411764707), ('Asymmetry Coefficient', 2.6449), ('Length of Kernel Groove', 5.194411764705882), ('Variety of Wheat', 1.0)]

**Variety 2**

[('Area', 18.721803278688522), ('Perimeter', 16.297377049180326), ('Compactness', 0.8850868852459016), ('Length of Kernel', 6.208934426229508), ('Width of Kernel', 3.722672131147541), ('Asymmetry Coefficient', 3.603590163934426), ('Length of Kernel Groove', 6.066098360655738), ('Variety of Wheat', 2.0)]

**Variety 3**

[('Area', 11.994324324324324), ('Perimeter', 13.287837837837838), ('Compactness', 0.8526527027027027), ('Length of Kernel', 5.232864864864864), ('Width of Kernel', 2.877689189189189), ('Asymmetry Coefficient', 4.777621621621622), ('Length of Kernel Groove', 5.094027027027027), ('Variety of Wheat', 3.0)]