REPORT

Description of Database:

Predicting the age of abalone from physical measurements. The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope -- a boring and time-consuming task. Other measurements, which are easier to obtain, are used to predict the age. Further information, such as weather pattern and location (hence food availability) may be required to solve the problem. Description of all the attributes is given below.

Name	Data Type	Measurement	Description
Sex	nominal	M, F, and I (infant)	Gender
Length	continuous	mm	longest shell measurement
Diameter	continuous	mm	perpendicular to length
Height	continuous	mm	with meat in shell
Whole weight	continuous	grams	whole abalone
Shucked weight	continuous	grams	weight of meat
Viscera weight	continuous	grams	gut weight (after bleeding)
Shell weight	continuous	grams	after being dried
Ring's	integer		+1.5 gives the age in years

K-Means Algorithm:

1. Specify number of clusters K.
2. Initialize K data point randomly without replacement as centroids from the dataset.
3. Compute the sum of the squared distance between data points and all centroids.
4. Assign each data point to the closest cluster (centroid).
5. Compute the centroids for the clusters by taking the average of the all-data points that belong to each cluster.
6. Repeat steps 3, 4, 5 until there's no change in the clusters.
7. Then I calculated Davies Boulding Score.
8. Then plotted the clusters by reducing the dimension into 2 using PCA.

Screenshot of K-Means Output: