**Lab Exercises-3**

**1. Use shopping dataset and perform the following:**

a) Implement Pre-processing functions

- Check for NULL values in the dataset. If any NULL value is present then treat it accordingly.

- extract only feature columns from the dataset.

b) Determine the number of clusters using dendogram. Use single linkage method.

c) Apply agglomerative clustering using single linkage method.

d) Form the clusters.

e) Evaluate the quality of clusters formed by agglomerative clustering.

**2. Use Shopping dataset of Q1 and perform the agglomerative clustering using “complete” linkage. Evaluate the quality of clusters formed by agglomerative clustering and compare the results with the clusters obtained in Q1.**

**3. Use Wholesale Customers dataset and perform the following:**

a) Implement Pre-processing functions

- Check for NULL values in the dataset. If any NULL value is present then treat it accordingly.

- extract only feature columns from the dataset.

b) Determine the number of clusters using dendogram. Use ward linkage method.

c) Apply agglomerative clustering using ward linkage method.

d) Form the clusters.

e) Visualize the clusters formed by agglomerative clustering.

**4. Use Credit Card dataset and perform the following:**

a) Implement Pre-processing functions

- Check for NULL values in the dataset. If any NULL value is present then treat it accordingly.

- extract only feature columns from the dataset.

b) Determine the number of clusters using dendogram. Use ward linkage method.

c) Apply agglomerative clustering using ward linkage method.

d) Form the clusters.

e) Visualize the clusters formed by agglomerative clustering.

d) Evaluate the quality of clusters.

**5. Use wine dataset and perform the following:**

a) Implement Pre-processing functions

- check for NULL values in the dataset. If any NULL value is present then treat it accordingly.

- extract only feature columns from the dataset.

b) Apply Principal component analysis and extract 2 principal components from the dataset.

c) Plot the scatter plot for the extracted components of wine dataset.

d) Now apply *k*-means clustering using the value of *k*=3 (as there are three different wine qualities in the target variable).

e) Form the clusters.

f) Evaluate the quality of clusters formed by *k*-means

6. **Use Breast Cancer Dataset and perform the following:**

a) Implement Pre-processing functions

- check for NULL values in the dataset. If any NULL value is present then treat it accordingly.

- extract only feature columns from the dataset.

b) Apply Principal component analysis and extract 2 principal components from the dataset.

c) Visualize the PCA components of the dataset.

**7. Use digits dataset consisting of 8x8 pixels (i.e. 64 dimensions) handwritten digits images. Your task is to apply PCA to reduce the dimensions of the dataset. Visualize the PCA results.**

**8. Using the dataset of Q7. Impute some noise in the dataset. Visualize the dataset with and without noise. Then apply PCA on noisy dataset to remove the noise.**

**Solution 8:**

**9. Use Labeled faces dataset (having dimensions nearly 3000) available in sklearn. Apply PCA to reduce the dimensions of this dataset.**