########## Problem Statement ############

Introduction:

Wholesale Customer dataset refers to clients of a wholesale distributor. It includes the annual spending in monetary units (m.u.) on diverse product categories.

Number of Instances: 440 Number of Attributes: 8 Data Set Characteristics: Multivariate Attribute Characteristics: Integer

Objective:

The major aim is to perform clustering analysis using algorithms like hClust,PAM(partition around medoids), kMeans and provide inferences accordingly.

Attribute Information:

FRESH: annual spending (m.u.) on fresh products (Continuous)

MILK: annual spending (m.u.) on milk products (Continuous)

GROCERY: annual spending (m.u.)on grocery products (Continuous)

FROZEN: annual spending (m.u.)on frozen products (Continuous)

DETERGENTS\_PAPER: annual spending (m.u.) on detergents and paper products (Continuous)

DELICATESSEN: annual spending (m.u.)on and delicatessen products (Continuous)

CHANNEL: customersale Channel - Horeca (Hotel/Restaurant/Cafe) or Retail channel (Nominal)

REGION: customersale Region - Lisnon, Oporto or Other (Nominal)

We will use KMeans algorithm to derive customer insights.

Task:

1. Download and visualize data (draw plots)

2. Get rid of nominal attributes (such as channel and region)

3. Apply K-means and hClust with and without outlier removal

4. Plot Elbow curve for different values of K (for K-means with outliers removed)

5. Describe your inferences