Assignment: Clustering problem(Whole sellers data)

You are given with excel file named "Wholesale.csv". The file contains data of annual spending of 430 clients of a wholesale distributor. Eight columns containing information on various attributes of interest of these clients are as follows

- Channel - Horeca (Hotel/Restaurant/Cafe or Retail channel (Nominal)
- Regions - Lisnon, Oporto or Other (Nominal)
- 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)
- Delicassen: annual spending (m.u.)on and delicatessen products (Continuous)

1 Fit K means clustering model to the above data, using following steps

a) Using the elbow method, find the optimal number of clusters
b) Using the optimal number of clusters determined above, fit K-Means clustering model to the dataset
c) Using the model, predict the cluster to which each customer belongs to. Add the cluster values for each customer in the dataset.

2 Fit hierarchical clustering model to the above data, using following steps

a) Using Dendrogram method, determine the optimal number of clusters.
b) Applying Hierarchial Clustering, divide the data in the number of clusters determined using Dendrogram method.
c) Add the cluster values for each customer in the dataset determined using Hierarchial Clustering algorithm.

3 Compare the similarity of prediction arrived at using the above two approaches