Use the clustering methodology to segment customers into groups:  
 Use the following clustering algorithms:

1. K means
2. Hierarchical

• Identify the right number of customer segments.  
• Provide the number of customers who are highly valued.  
• Identify the clustering algorithm that gives maximum accuracy and explains robust clusters.  
• If the number of observations is loaded in one of the clusters, break down that cluster further using the clustering algorithm. [ hint: Here loaded means if any cluster has more number of data points as compared to other clusters then split that clusters by increasing the number of clusters and observe, compare the results with previous results.]

Here's the dataset: [https://github.com/Simplilearn-Edu/Data-Science-with-R.git](https://github.com/Simplilearn-Edu/Data-Science-with-R)

**Solution:**

Since we want customers with high price purchases we calculate sum(Unit Price\* Quantity) for each customer

Graphical user interface, text, application, Word

Description automatically generated

Graphical user interface, text, application, Teams

Description automatically generated

**K-Means Clustering**

Now we perform K-means clustering on sum(Price\*Quantity)

Let’s find suitable number of clusters

Graphical user interface, text, application, email

Description automatically generated

Chart, line chart, histogram

Description automatically generated

Elbow bend is observed at K=3. We choose K=3.

Text

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Cluster 3 has very high value customers, Cluster 2 has reasonably high value customers and cluster 1 has low value customers. We are interested in Cluster 3 and Cluster 2

Cluster 3: Three premium level customers

Text

Description automatically generated

Cluster 2: Reasonably high value customers (30 in number)

Table

Description automatically generated

**Hierarchical Clustering**

Text

Description automatically generated

Chart

Description automatically generated

Here cluster 3 are valued customers while cluster 1 and cluster 2 are low valued

Table

Description automatically generated