Cluster analysis

Import the dataset *DatasetTelecom,* which contains numerical variables and categorical variables transformed into numerical variables.

To standardize the dataset

```
sTelecom=scale(DatasetTelecom)
```

After installing and loading the package *cluster*, compute the distance matrix

```
distanceTelecom=daisy(sTelecom)
```

To perform hierarchical cluster analysis and represent the dendrogram

```
HCTelecom=hclust(distanceTelecom)
plot(HCTelecom)
```

To analyze the cut of the tree (dendrogram) providing 3 clusters

```
cut3Telecom=cutree(HCTelecom,3)
```

To visualize the membership group of the units

```
n=dim(sTelecom)[1]
groups3Telecom=cbind(seq(1,n),cut3Telecom)
groups3Telecom
```

To compute the mean of the variable AccountWeeks for the groups

```
Account=DatasetTelecom$AccountWeeks
mean(Account[groups3Telecom[groups3Telecom[,2]==1,1]])
mean(Account[groups3Telecom[groups3Telecom[,2]==2,1]])
mean(Account[groups3Telecom[groups3Telecom[,2]==3,1]])
```

To perform hierarchical cluster analysis using the average linkage method

```
HCTelecomAVE=hclust(distanceTelecom, method="single")
plot(HCTelecomAVE)
```

After installing and loading the package *factoextra*, compute the Average Silhouette index con maximum number of clusters equal to 10

```
AverageSilh=fviz_nbclust(sTelecom, hcut,
method='silhouette', k.max=10)
```

To plot the Average Silhouette index vs the number of clusters

```
plot(AverageSilh)
```

	Import the dataset *DatasetStartups3,* which contains numerical variables and categorical variables.
	
	Define the variable *State* as categorical:
	`State=DatasetStartups3$State`
	`State=as.factor(State)`
	`DatasetStartups3$State=State`
	
	The following steps can be implemented
	`distanceStart=daisy(DatasetStartups3)`
	`HCStart=hclust(distanceStart)`
	`plot(HCStart)`