1. Missing value replenishment

2.Merging data files

3.Appending the data files

4.Transformation or normalization

5.Random Sampling

**Missing Value Handling**

Suppose a telecom company wants to analyse the performance of its circles based on the following parameters

1. Current Month’s Usage
2. Last 3 Month’s Usage
3. Average Recharge
4. Projected Growth

Example: Read data and variables to R from **Missing\_Value\_Telecom**

> mydata = Missing\_Values\_Telecom

>cmusage = mydata[,2]

> l3musage = mydata[,3]

> avrecharge = mydata[,4]

Option 1: Discard all records with missing values

>newdata = na.omit(mydata)

>write.csv(newdata,"newdata.csv")

Option 2: Replace the missing values with variable mean, median, etc

Replacing the missing values with mean

Compute the mean excluding the missing values

>cmusage\_mean = mean(cmusage, na.rm = TRUE)

>l3musage\_mean = mean(l3musage\_mean, na.rm = TRUE)

> l3musage\_mean = mean(l3musage, na.rm = TRUE)

> avrecharge\_mean = mean(avrecharge, na.rm = TRUE)

Replace the missing values with mean

> cmusage[is.na(cmusage)]=cmusage\_mean

> l3musage[is.na(l3musage)]= l3musage\_mean

> avrecharge[is.na(avrecharge)]=avrecharge\_mean

**Making the new file**

> mynewdata = cbind(cmusage, l3musage, avrecharge, mydata[,5],mydata[,6])

> write.csv(mynewdata, "mynewdata.csv")

**DATA MERGING**

Exercise: The data of 30 customers on credit card usage in INR1000 is given in CC\_Usage.txt. Similarly the user profile namely gender (1: male, 2: female) and whether they have done shopping or banking (1: yes , 2: no) with credit card are given in cc\_Profile.csv. Can you merge the two files into a single data set?

Read the files

>myprofile = CC\_Profile

>myusage = CC\_Usage

**Merge the files by “ID” field**

>mydata = merge(myprofile, myusage, by = “ID”)

**DATA APPEND**

Exercise: The data on user profile of customers whom are included in the previous mailing campaign is compiled into two files namely classification1.csv and classification2.txt. Can you append the second data set with the first one and store the new data set in a new file?

**DATA APPEND**

Read the files

>class1 = Classification1

> class2 = Classification2

Append class1 with class2

>mydata = rbind(class1, class2)