VIT-AP UNIVERSITY, ANDHRA PRADESH

CSE4027 – Data Analytics - Lab Sheet :4
LAB 4 (Data Cleaning and Imputation)

Academic year: 2022-2023 **Branch/ Class:** B.Tech/M.Tech
Semester: Fall **Date: 07-10-22**
Faculty Name: Dr Syed Khasim **School:** SCOPE
Student name: MAJJIGA JASWANTH **Reg. no.:20BCD7171**

LAB 4 (Data Cleaning and Imputation)

Questions:

1.Do preliminary observations. (head, str…)

```
> data<-read.csv("C:/Users/Sashank K/Desktop/Student_Data_Uncleaned.csv")
> data
```

	sno	regno	name	school	cat1	cat2	da01	fat	lab	quiz1	gt	grade	result
1	1	19BCE7478	NITTOOR VISHNU BHARADWAJ	CSE	43	38.5	19	33.0	87	15	NA	NA	NA
2	2	19BCN7017	INTURI REVANTH	CSE	10	12.0	19	3.0	34	15	NA	NA	NA
3	3	19BCN7045	MUMMANI PURNAVENKTASAIKIRAN	CSE	26	37.0	19	14.0	60	18	NA	NA	NA
4	4	19BCN7050	NISHIT VERMA	CSE	NA	47.5	19	37.0	93	20	NA	NA	NA
5	5	19BCN7064	TADIBOINA ANAND KUMAR	ECE	23	46.0	19	28.0	78	20	NA	NA	NA
6	6	19BCN7079	PATTAPU SAI SRINIVAS	ECE	20	31.5	NA	10.0	61	18	NA	NA	NA
7	7	19BCN7114	BALIVADA PRATYUSH	CSE	28	41.0	18	31.0	75	20	NA	NA	NA
8	8	19BCN7136	AMARA SANTOSH JAYANTH	CSE	39	45.0	19	28.0	85	15	NA	NA	NA
9	9	19BCN7137	JAYAPRAKASH SUGAN PRASAD	CSE	14	17.5	19	9.5	37	NA	NA	NA	NA
10	10	19BCE7475	SHOBHIT KHURANA	CSE	38	40.5	19	30.5	86	18	NA	NA	NA
11	11	19BCE7487	PRIYANSHI SHARMA	ECE	26	31.0	18	16.0	66	15	NA	NA	NA
12	12	19BCN7190	CHEDALAWADA SESHA PRANAV	ECE	32	20.0	18	27.0	69	15	NA	NA	NA
13	13	19BCE7124	PERNI BINDU MADHAVI	CSE	16	27.5	19	14.0	51	20	NA	NA	NA
14	14	19BCN7021	CHADA RADHA AMUKTHA BHARGAVI	CSE	38	35.0	18	34.0	83	20	NA	NA	NA
15	15	19BCN7041	AVVA VENKATA PUNEETH	CSE	36	29.5	18	30.0	73	20	NA	NA	NA
16	16	19BCN7075	RONDA PRIYATHAM		34	26.5	18	22.5	61	13	NA	NA	NA
17	17	19BCN7092	MADATI HRUTHIK YADAV	ECE	17	25.0	NA	18.0	NA	13	NA	NA	NA
18	18	19BES7053	VARIKELA GOPALAKRISHNA	ECE	15	28.5	18	12.5	45	10	NA	NA	NA
19	19	19BCE7131	ANANNYA BARUA	CSE	43	NA	19	24.0	77	20	NA	NA	NA
20	20	19BCE7166	V VIKRAM GANESH	CSE	43	47.0	19	34.5	93	20	NA	NA	NA
21	21	19BCD7074	P.ANUSHRI SOWMYA	CSE	24	23.5	19	7.0	47	10	NA	NA	NA
22	22	19BCE7283	C M ELAKKIA	CSE	28	36.5	19	20.0	66	15	NA	NA	NA
23	23	19BCN7185	M.NAVEEN		37	37.5	19	NA	84	18	NA	NA	NA
24	24	19BCN7186	NISHANTH	ECE	NA	36.0	19	14.0	68	13	NA	NA	NA
25	25	19BCN7255	SURENDHAR A	CSE	13	26.0	19	15.0	55	15	NA	NA	NA
26	26	19BCE7113	THOTA SURYA TEJA	CSE	28	22.0	19	22.5	60	20	NA	NA	NA
27	27	19BCE7653	CHIMAKURTHI B V S N PAVAN KUMAR	CSE	19	34.0	3	29.0	57	5	NA	NA	NA
28	28	19BCN7008	DUTTA GEETHIK	CSE	21	30.0	19	13.5	61	20	NA	NA	NA
29	29	19BCN7225	SEELAM PENCHALA SAI VARA PRASAD	ECE	28	NA	19	20.5	NA	20	NA	NA	NA
30	30	19BEC7051	SIRANJEEVE K	ECE	20	29.0	19	8.0	57	15	NA	NA	NA
31	31	19BCN7022	ANDE SANDEEP	CSE	27	32.5	19	12.0	59	18	NA	NA	NA
32	32	19BCN7242	KATTA LOKESH CHOWDARY	CSE	10	0.0	19	0.0	19	0	NA	NA	NA

```
> summary(data)
      sno          regno              name              school
 Min.   : 1   Length:57          Length:57          Length:57
 1st Qu.:15   Class :character   Class :character   Class :character
 Median :29   Mode  :character   Mode  :character   Mode  :character
 Mean   :29
 3rd Qu.:43
 Max.   :57

     quiz1           gt            grade           result
 Min.   : 0.00   Mode:logical   Mode:logical   Mode:logical
 1st Qu.:15.00   NA's:57        NA's:57        NA's:57
 Median :18.00
 Mean   :16.06
 3rd Qu.:20.00
 Max.   :20.00
 NA's   :4
     cat1            cat2            da01            fat             lab
Min.   : 6.00   Min.   : 0.00   Min.   : 3.00   Min.   : 0.00   Min.   :19.00
1st Qu.:18.00   1st Qu.:23.50   1st Qu.:18.00   1st Qu.:11.00   1st Qu.:49.25
Median :25.00   Median :29.50   Median :19.00   Median :17.50   Median :61.00
Mean   :25.85   Mean   :29.48   Mean   :18.09   Mean   :19.18   Mean   :62.09
3rd Qu.:34.00   3rd Qu.:36.50   3rd Qu.:19.00   3rd Qu.:28.00   3rd Qu.:76.50
Max.   :43.00   Max.   :47.50   Max.   :19.00   Max.   :40.00   Max.   :93.00
NA's   :4       NA's   :4       NA's   :3       NA's   :2       NA's   :3
```

```
> head(data,5)
  sno     regno                        name school cat1 cat2 da01 fat lab quiz1 gt grade result
1   1 19BCE7478    NITTOOR VISHNU BHARADWAJ    CSE   43 38.5   19  33  87    15 NA    NA     NA
2   2 19BCN7017               INTURI REVANTH    CSE   10 12.0   19   3  34    15 NA    NA     NA
3   3 19BCN7045 MUMMANI PURNAVENKTASAIKIRAN    CSE   26 37.0   19  14  60    18 NA    NA     NA
4   4 19BCN7050           NISHIT VERMA         CSE   NA 47.5   19  37  93    20 NA    NA     NA
5   5 19BCN7064       TADIBOINA ANAND KUMAR    ECE   23 46.0   19  28  78    20 NA    NA     NA
> tail(data,5)
   sno     regno                  name school cat1 cat2 da01 fat lab quiz1 gt grade result
53  53 19BCE7491 CYRIL AMBEDKAR KONDRU    ECE   32 29.5   18   5  49     8 NA    NA     NA
54  54 19BCI7098   PENIKALAPATI PRANAY    ECE   26 29.5   19  13  59    18 NA    NA     NA
55  55 19BCI7011     NISCHAL NANDIGAMA          37 37.5   18  NA  84    NA NA    NA     NA
56  56 19BEC7071  ANAND CHOUDARY JASTI    CSE   NA 23.5   19  13  39    13 NA    NA     NA
57  57 19BEC7141          KANDE ANISHA    CSE   18 24.5    3   6  28     5 NA    NA     NA
> str(data)
'data.frame':   57 obs. of  13 variables:
 $ sno   : int  1 2 3 4 5 6 7 8 9 10 ...
 $ regno : chr  "19BCE7478" "19BCN7017" "19BCN7045" "19BCN7050" ...
 $ name  : chr  "NITTOOR VISHNU BHARADWAJ" "INTURI REVANTH" "MUMMANI PURNAVENKTASAIKIRAN" "NISHIT VERMA     " ...
 $ school: chr  "CSE" "CSE" "CSE" "CSE" ...
 $ cat1  : int  43 10 26 NA 23 20 28 39 14 38 ...
 $ cat2  : num  38.5 12 37 47.5 46 31.5 41 45 17.5 40.5 ...
 $ da01  : int  19 19 19 19 19 NA 18 19 19 19 ...
 $ fat   : num  33 3 14 37 28 10 31 28 9.5 30.5 ...
 $ lab   : int  87 34 60 93 78 61 75 85 37 86 ...
 $ quiz1 : int  15 15 18 20 20 18 20 15 NA 18 ...
 $ gt    : logi  NA NA NA NA NA NA ...
 $ grade : logi  NA NA NA NA NA NA ...
 $ result: logi  NA NA NA NA NA NA ...
>
```

2.Using repeat loop add 2 to all CAT1, CAT2 and FAT columns.

```
> data1<-data[,sapply(data,is.numeric)]
> data1
   sno cat1 cat2 da01  fat lab quiz1
1    1   43 38.5   19 33.0  87    15
2    2   10 12.0   19  3.0  34    15
3    3   26 37.0   19 14.0  60    18
4    4   NA 47.5   19 37.0  93    20
5    5   23 46.0   19 28.0  78    20
6    6   20 31.5   NA 10.0  61    18
7    7   28 41.0   18 31.0  75    20
8    8   39 45.0   19 28.0  85    15
9    9   14 17.5   19  9.5  37    NA
10  10   38 40.5   19 30.5  86    18
11  11   26 31.0   18 16.0  66    15
12  12   32 20.0   18 27.0  69    15
13  13   16 27.5   19 14.0  51    20
14  14   38 35.0   18 34.0  83    20
15  15   36 29.5   18 30.0  73    20
16  16   34 26.5   18 22.5  61    13
17  17   17 25.0   NA 18.0  NA    13
18  18   15 28.5   18 12.5  45    10
19  19   43   NA   19 24.0  77    20
20  20   43 47.0   19 34.5  93    20
21  21   24 23.5   19  7.0  47    10
22  22   28 36.5   19 20.0  66    15
23  23   37 37.5   19   NA  84    18
24  24   NA 36.0   19 14.0  68    13
25  25   13 26.0   19 15.0  55    15
26  26   28 22.0   19 22.5  60    20
27  27   19 34.0    3 29.0  57     5
28  28   21 30.0   19 13.5  61    20
29  29   28   NA   19 20.5  NA    20
30  30   20 29.0   19  8.0  57    15
31  31   27 32.5   19 12.0  59    18
32  32   10  0.0   19  0.0  19     0
33  33   22 18.5   19 16.0  50    13
34  34   43 38.0   19 40.0  85    17
```

```
> data2<-data1[,c(2,3,5)]
> data2
   cat1 cat2  fat
1    43 38.5 33.0
2    10 12.0  3.0
3    26 37.0 14.0
4    NA 47.5 37.0
5    23 46.0 28.0
6    20 31.5 10.0
7    28 41.0 31.0
8    39 45.0 28.0
9    14 17.5  9.5
10   38 40.5 30.5
11   26 31.0 16.0
12   32 20.0 27.0
13   16 27.5 14.0
14   38 35.0 34.0
15   36 29.5 30.0
16   34 26.5 22.5
17   17 25.0 18.0
18   15 28.5 12.5
19   43   NA 24.0
20   43 47.0 34.5
21   24 23.5  7.0
22   28 36.5 20.0
23   37 37.5   NA
24   NA 36.0 14.0
25   13 26.0 15.0
26   28 22.0 22.5
27   19 34.0 29.0
28   21 30.0 13.5
29   28   NA 20.5
30   20 29.0  8.0
31   27 32.5 12.0
32   10  0.0  0.0
33   22 18.5 16.0
34   43 38.0 40.0
35   38 38.0 24.5
36   16 18.5 10.0
37   36 42.0 28.0
38   21 25.0  5.5
> length(data2)
[1] 3
> count<-nrow(data2)
> data2[1,2]
[1] 38.5
> i=1
> repeat {
+   for(j in 1:3)
+   {
+     data2[i,j]=data2[i,j]+2
+   }
+   i=i+1
+   if(i>count)
+   {
+     break
+   }
+ }
```

```
> data2
   cat1 cat2  fat
1    45 40.5 35.0
2    12 14.0  5.0
3    28 39.0 16.0
4    NA 49.5 39.0
5    25 48.0 30.0
6    22 33.5 12.0
7    30 43.0 33.0
8    41 47.0 30.0
9    16 19.5 11.5
10   40 42.5 32.5
11   28 33.0 18.0
12   34 22.0 29.0
13   18 29.5 16.0
14   40 37.0 36.0
15   38 31.5 32.0
16   36 28.5 24.5
17   19 27.0 20.0
18   17 30.5 14.5
19   45   NA 26.0
20   45 49.0 36.5
21   26 25.5  9.0
22   30 38.5 22.0
23   39 39.5   NA
24   NA 38.0 16.0
25   15 28.0 17.0
26   30 24.0 24.5
27   21 36.0 31.0
28   23 32.0 15.5
29   30   NA 22.5
30   22 31.0 10.0
31   29 34.5 14.0
32   12  2.0  2.0
33   24 20.5 18.0
34   45 40.0 42.0
35   40 40.0 26.5
36   18 20.5 12.0
37   38 44.0 30.0
```

3.Use for loop to get Not Available data from user for CAT1, CAT2and FAT and update it into CSV

```
> mean(data2[,1],na.rm=TRUE)
[1] 27.84906
> mean(data2[,2],na.rm=TRUE)
[1] 31.48113
> mean(data2[,3],na.rm=TRUE)
[1] 21.18182
> for(i in 1:3)
+ {
+   data2[,i]<-ifelse(is.na(data2[,i]),mean(data2[,i],na.rm=TRUE),data2[,i])
+ }
```

```
> data2
       cat1     cat2      fat
1  45.00000 40.50000 35.00000
2  12.00000 14.00000  5.00000
3  28.00000 39.00000 16.00000
4  27.84906 49.50000 39.00000
5  25.00000 48.00000 30.00000
6  22.00000 33.50000 12.00000
7  30.00000 43.00000 33.00000
8  41.00000 47.00000 30.00000
9  16.00000 19.50000 11.50000
10 40.00000 42.50000 32.50000
11 28.00000 33.00000 18.00000
12 34.00000 22.00000 29.00000
13 18.00000 29.50000 16.00000
14 40.00000 37.00000 36.00000
15 38.00000 31.50000 32.00000
16 36.00000 28.50000 24.50000
17 19.00000 27.00000 20.00000
18 17.00000 30.50000 14.50000
19 45.00000 31.48113 26.00000
20 45.00000 49.00000 36.50000
21 26.00000 25.50000  9.00000
22 30.00000 38.50000 22.00000
23 39.00000 39.50000 21.18182
24 27.84906 38.00000 16.00000
25 15.00000 28.00000 17.00000
26 30.00000 24.00000 24.50000
27 21.00000 36.00000 31.00000
28 23.00000 32.00000 15.50000
29 30.00000 31.48113 22.50000
30 22.00000 31.00000 10.00000
31 29.00000 34.50000 14.00000
32 12.00000  2.00000  2.00000
33 24.00000 20.50000 18.00000
34 45.00000 40.00000 42.00000
35 40.00000 40.00000 26.50000
36 18.00000 20.50000 12.00000
37 38.00000 44.00000 30.00000
38 23.00000 27.00000  7.50000
39 15.00000 24.00000 12.00000
40 24.00000 31.48113 30.50000
41 16.00000 31.00000 11.00000
42 32.00000 31.50000 25.00000
43 27.00000 31.50000 34.50000
44 27.84906 30.00000 19.50000
45 22.00000 25.00000 11.50000
46 26.00000 37.00000 37.00000
47 20.00000 34.00000 19.00000
48 20.00000 19.00000 15.00000
49 22.00000 17.50000 19.50000
50 34.00000 27.50000 28.50000
51 45.00000 31.48113 35.00000
52  8.00000 20.50000  6.00000
53 34.00000 31.50000  7.00000
54 28.00000 31.50000 15.00000
55 39.00000 39.50000 21.18182
56 27.84906 25.50000 15.00000
57 20.00000 26.50000  8.00000
>
```

4.Use mean value to replace Not available data in DA01, QUIZ1 and LAB

```
> data3<-data1[,c(4,6,7)]
> data3
   da01 lab quiz1
1    19  87    15
2    19  34    15
3    19  60    18
4    19  93    20
5    19  78    20
6    NA  61    18
7    18  75    20
8    19  85    15
9    19  37    NA
10   19  86    18
11   18  66    15
12   18  69    15
13   19  51    20
14   18  83    20
15   18  73    20
16   18  61    13
17   NA  NA    13
18   18  45    10
19   19  77    20
20   19  93    20
21   19  47    10
22   19  66    15
23   19  84    18
24   19  68    13
25   19  55    15
26   19  60    20
27    3  57     5
28   19  61    20
29   19  NA    20
30   19  57    15
31   19  59    18
32   19  19     0
33   19  50    13
34   19  85    17
35   19  82    18
36   18  50    15
37   NA  85    NA
38   18  40    13
39   18  42    18
40   18  NA    20
41   18  44    20
42   19  67    20
43   19  71    20
44   19  64    15
45   18  52    18
46   19  75    20
47   19  64    20
48   19  43    10
49   18  43    NA
50   19  66    18
51   19  89    20
52   18  35    18
53   18  49     8
54   19  59    18
55   18  84    NA
56   19  39    13
57    3  28     5
.....
> mean(data2[,1],na.rm=TRUE)
[1] 27.84906
> mean(data2[,2],na.rm=TRUE)
[1] 31.48113
> mean(data2[,3],na.rm=TRUE)
[1] 21.18182
> for(i in 1:3)
+ {
+   data3[,i]<-ifelse(is.na(data3[,i]),mean(data3[,i],na.rm=TRUE),data3[,i])
+ }
```

```
> data3
       da01      lab   quiz1
1  19.00000 87.00000 15.0000
2  19.00000 34.00000 15.0000
3  19.00000 60.00000 18.0000
4  19.00000 93.00000 20.0000
5  19.00000 78.00000 20.0000
6  18.09259 61.00000 18.0000
7  18.00000 75.00000 20.0000
8  19.00000 85.00000 15.0000
9  19.00000 37.00000 16.0566
10 19.00000 86.00000 18.0000
11 18.00000 66.00000 15.0000
12 18.00000 69.00000 15.0000
13 19.00000 51.00000 20.0000
14 18.00000 83.00000 20.0000
15 18.00000 73.00000 20.0000
16 18.00000 61.00000 13.0000
17 18.09259 62.09259 13.0000
18 18.00000 45.00000 10.0000
19 19.00000 77.00000 20.0000
20 19.00000 93.00000 20.0000
21 19.00000 47.00000 10.0000
22 19.00000 66.00000 15.0000
23 19.00000 84.00000 18.0000
24 19.00000 68.00000 13.0000
25 19.00000 55.00000 15.0000
26 19.00000 60.00000 20.0000
27  3.00000 57.00000  5.0000
28 19.00000 61.00000 20.0000
29 19.00000 62.09259 20.0000
30 19.00000 57.00000 15.0000
31 19.00000 59.00000 18.0000
32 19.00000 19.00000  0.0000
33 19.00000 50.00000 13.0000
34 19.00000 85.00000 17.0000
35 19.00000 82.00000 18.0000
36 18.00000 50.00000 15.0000
37 18.09259 85.00000 16.0566 ……
38 18.00000 40.00000 13.0000
39 18.00000 42.00000 18.0000
40 18.00000 62.09259 20.0000
41 18.00000 44.00000 20.0000
42 19.00000 67.00000 20.0000
43 19.00000 71.00000 20.0000
44 19.00000 64.00000 15.0000
45 18.00000 52.00000 18.0000
46 19.00000 75.00000 20.0000
47 19.00000 64.00000 20.0000
48 19.00000 43.00000 10.0000
49 18.00000 43.00000 16.0566
50 19.00000 66.00000 18.0000
51 19.00000 89.00000 20.0000
52 18.00000 35.00000 18.0000
53 18.00000 49.00000  8.0000
54 19.00000 59.00000 18.0000
55 18.00000 84.00000 16.0566
56 19.00000 39.00000 13.0000
57  3.00000 28.00000  5.0000
>
```

5.Find the Grant Total (GT) for all students and update it into the CSV file.
GT = ((CAT1+CAT2+FAT)/150*40 + (DA01/20)*15 + (QUIZ1/20)*15 + (LAB/100) * 30)

```
> data<-transform(data,gt=((data2$cat1+data2$cat2+data2$fat)/150)*40+
+                   (data3$quiz1/20)*15+(data3$lab/100)*30)
> data
```

	sno	regno	name	school	cat1	cat2	da01	fat
1	1	19BCE7478	NITTOOR VISHNU BHARADWAJ	CSE	43	38.5	19	33.0
2	2	19BCN7017	INTURI REVANTH	CSE	10	12.0	19	3.0
3	3	19BCN7045	MUMMANI PURNAVENKTASAIKIRAN	CSE	26	37.0	19	14.0
4	4	19BCN7050	NISHIT VERMA	CSE	NA	47.5	19	37.0
5	5	19BCN7064	TADIBOINA ANAND KUMAR	ECE	23	46.0	19	28.0
6	6	19BCN7079	PATTAPU SAI SRINIVAS	ECE	20	31.5	NA	10.0
7	7	19BCN7114	BALIVADA PRATYUSH	CSE	28	41.0	18	31.0
8	8	19BCN7136	AMARA SANTOSH JAYANTH	CSE	39	45.0	19	28.0
9	9	19BCN7137	JAYAPRAKASH SUGAN PRASAD	CSE	14	17.5	19	9.5
10	10	19BCE7475	SHOBHIT KHURANA	CSE	38	40.5	19	30.5
11	11	19BCE7487	PRIYANSHI SHARMA	ECE	26	31.0	18	16.0
12	12	19BCN7190	CHEDALAWADA SESHA PRANAV	ECE	32	20.0	18	27.0
13	13	19BCE7124	PERNI BINDU MADHAVI	CSE	16	27.5	19	14.0
14	14	19BCN7021	CHADA RADHA AMUKTHA BHARGAVI	CSE	38	35.0	18	34.0
15	15	19BCN7041	AVVA VENKATA PUNEETH	CSE	36	29.5	18	30.0
16	16	19BCN7075	RONDA PRIYATHAM		34	26.5	18	22.5
17	17	19BCN7092	MADATI HRUTHIK YADAV	ECE	17	25.0	NA	18.0
18	18	19BES7053	VARIKELA GOPALAKRISHNA	ECE	15	28.5	18	12.5
19	19	19BCE7131	ANANNYA BARUA	CSE	43	NA	19	24.0
20	20	19BCE7166	V VIKRAM GANESH	CSE	43	47.0	19	34.5
21	21	19BCD7074	P.ANUSHRI SOWMYA	CSE	24	23.5	19	7.0
22	22	19BCE7283	C M ELAKKIA	CSE	28	36.5	19	20.0
23	23	19BCN7185	M.NAVEEN		37	37.5	19	NA
24	24	19BCN7186	NISHANTH	ECE	NA	36.0	19	14.0
25	25	19BCN7255	SURENDHAR A	CSE	13	26.0	19	15.0
26	26	19BCE7113	THOTA SURYA TEJA	CSE	28	22.0	19	22.5
27	27	19BCE7653	CHIMAKURTHI B V S N PAVAN KUMAR	CSE	19	34.0	3	29.0
28	28	19BCN7008	DUTTA GEETHIK	CSE	21	30.0	19	13.5
29	29	19BCN7225	SEELAM PENCHALA SAI VARA PRASAD	ECE	28	NA	19	20.5
30	30	19BEC7051	SIRANJEEVE K	ECE	20	29.0	19	8.0
31	31	19BCN7022	ANDE SANDEEP	CSE	27	32.5	19	12.0
32	32	19BCN7242	KATTA LOKESH CHOWDARY	CSE	10	0.0	19	0.0
33	33	19BEC7079	PAVULURI KHADYOTHAN SAI TRINADH	CSE	22	18.5	19	16.0
34	34	19BCD7243	SIMRAN ANAND	CSE	43	38.0	19	40.0
35	35	19BCE7158	RAYUDU PRANEETH	ECE	38	38.0	19	24.5
36	36	19BCE7168	CHERUKURI SAI RIKSHIT		16	18.5	18	10.0
37	37	19BCN7072	SIMRAN ANAND	CSE	36	42.0	NA	28.0
38	38	19BCD7062	REDROUTHU LAKSHMAN	CSE	21	25.0	18	5.5
39	39	19BCD7113	BORIGORLA BALARAM NAGA SRINIVAS	CSE	13	22.0	18	10.0
40	40	19BCE7225	SEGU SHANMUKA SRINIVAS	CSE	22	NA	18	28.5
41	41	19BCN7230	O.LOKESH CHOWDARY	ECE	14	29.0	18	9.0
42	42	19BCN7111	UTTARILLI SREETEJA	ECE	30	29.5	19	23.0
43	43	19BCN7130	ALUR SAMEER ALI KHAN	CSE	25	29.5	19	32.5
44	44	19BCE7192	MANEESH MADALA	CSE	NA	28.0	19	17.5
45	45	19BCE7469	KASARANENI SRIVINEELA	CSE	20	23.0	18	9.5
46	46	19BCE7483	PASALAPUDI CHANUKYA SAI VARDHAN	CSE	24	35.0	19	35.0
47	47	19BCN7069	CHINTALA LOKESH BABU	ECE	18	32.0	19	17.0
48	48	19BCN7243	KANCHETI DILLIKUMAR		18	17.0	19	13.0
49	49	19BEC7025	SHAIK NAFEESA ANJUM	CSE	20	15.5	18	17.5
50	50	19BES7048	MARNI MAHENDRA CHOWDARY	CSE	32	25.5	19	26.5
51	51	19BCD7025	BUSSA SAIBIPIN	CSE	43	NA	19	33.0
52	52	19BCE7034	MOOLA SAI UDAY KIRAN KUMAR	CSE	6	18.5	18	4.0
53	53	19BCE7491	CYRIL AMBEDKAR KONDRU	ECE	32	29.5	18	5.0
54	54	19BCI7098	PENIKALAPATI PRANAY	ECE	26	29.5	19	13.0
55	55	19BCI7011	NISCHAL NANDIGAMA		37	37.5	18	NA
56	56	19BEC7071	ANAND CHOUDARY JASTI	CSE	NA	23.5	19	13.0
57	57	19BEC7141	KANDE ANISHA	CSE	18	24.5	3	6.0

	lab	quiz1	gt	grade	result
1	87	15	69.483333	NA	NA
2	34	15	29.716667	NA	NA
3	60	18	53.633333	NA	NA
4	93	20	73.926415	NA	NA
5	78	20	65.866667	NA	NA
6	61	18	49.800000	NA	NA
7	75	20	65.766667	NA	NA
8	85	15	68.216667	NA	NA
9	37	NA	35.675786	NA	NA
10	86	18	69.966667	NA	NA
11	66	15	52.116667	NA	NA
12	69	15	54.616667	NA	NA
13	51	20	47.233333	NA	NA
14	83	20	70.033333	NA	NA
15	73	20	63.966667	NA	NA
16	61	13	51.783333	NA	NA
17	NA	13	45.977778	NA	NA
18	45	10	37.533333	NA	NA
19	77	20	65.428302	NA	NA
20	93	20	77.700000	NA	NA
21	47	10	37.733333	NA	NA
22	66	15	55.183333	NA	NA
23	84	18	65.281818	NA	NA
24	68	13	51.976415	NA	NA
25	55	15	43.750000	NA	NA
26	60	20	53.933333	NA	NA
27	57	5	44.316667	NA	NA
28	61	20	52.100000	NA	NA
29	NA	20	56.022746	NA	NA
30	57	15	45.150000	NA	NA
31	59	18	51.866667	NA	NA
32	19	0	9.966667	NA	NA
33	50	13	41.416667	NA	NA
34	85	17	72.116667	NA	NA
35	82	18	66.500000	NA	NA
36	50	15	39.716667	NA	NA
37	85	NA	67.409119	NA	NA
38	40	13	37.083333	NA	NA
39	42	18	39.700000	NA	NA
40	NA	20	56.556080	NA	NA
41	44	20	43.666667	NA	NA
42	67	20	58.700000	NA	NA
43	71	20	61.100000	NA	NA
44	64	15	51.076415	NA	NA
45	52	18	44.700000	NA	NA
46	75	20	64.166667	NA	NA
47	64	20	53.666667	NA	NA
48	43	10	34.800000	NA	NA
49	43	NA	40.675786	NA	NA
50	66	18	57.300000	NA	NA
51	89	20	71.428302	NA	NA
52	35	18	33.200000	NA	NA
53	49	8	40.033333	NA	NA
54	59	18	51.066667	NA	NA
55	84	NA	63.824271	NA	NA
56	39	13	39.676415	NA	NA
57	28	5	26.683333	NA	NA

..

6.Update the grade as per grade policy of our institution.

```
> data<-transform(data1,grade=ifelse(data$gt>90,'S',
+                            ifelse(data$gt<=90 & data$gt>80,'A',
+                            ifelse(data$gt<=80 & data$gt>70,'B',
+                            ifelse(data$gt<=70 & data$gt>=64,'C','F')))))
```

```
> data
   sno cat1 cat2 da01  fat lab quiz1 grade
1    1   43 38.5   19 33.0  87    15     C
2    2   10 12.0   19  3.0  34    15     F
3    3   26 37.0   19 14.0  60    18     F
4    4   NA 47.5   19 37.0  93    20     B
5    5   23 46.0   19 28.0  78    20     C
6    6   20 31.5   NA 10.0  61    18     F
7    7   28 41.0   18 31.0  75    20     C
8    8   39 45.0   19 28.0  85    15     C
9    9   14 17.5   19  9.5  37    NA     F
10  10   38 40.5   19 30.5  86    18     C
11  11   26 31.0   18 16.0  66    15     F
12  12   32 20.0   18 27.0  69    15     F
13  13   16 27.5   19 14.0  51    20     F
14  14   38 35.0   18 34.0  83    20     B
15  15   36 29.5   18 30.0  73    20     F
16  16   34 26.5   18 22.5  61    13     F
17  17   17 25.0   NA 18.0  NA    13     F
18  18   15 28.5   18 12.5  45    10     F
19  19   43   NA   19 24.0  77    20     C
20  20   43 47.0   19 34.5  93    20     B
21  21   24 23.5   19  7.0  47    10     F
22  22   28 36.5   19 20.0  66    15     F
23  23   37 37.5   19   NA  84    18     C
24  24   NA 36.0   19 14.0  68    13     F
25  25   13 26.0   19 15.0  55    15     F
26  26   28 22.0   19 22.5  60    20     F
27  27   19 34.0    3 29.0  57     5     F
28  28   21 30.0   19 13.5  61    20     F
29  29   28   NA   19 20.5  NA    20     F
30  30   20 29.0   19  8.0  57    15     F
31  31   27 32.5   19 12.0  59    18     F
32  32   10  0.0   19  0.0  19     0     F
33  33   22 18.5   19 16.0  50    13     F
34  34   43 38.0   19 40.0  85    17     B
35  35   38 38.0   19 24.5  82    18     C
36  36   16 18.5   18 10.0  50    15     F
37  37   36 42.0   NA 28.0  85    NA     C
38  38   21 25.0   18  5.5  40    13     F
39  39   13 22.0   18 10.0  42    18     F
40  40   22   NA   18 28.5  NA    20     F
41  41   14 29.0   18  9.0  44    20     F
42  42   30 29.5   19 23.0  67    20     F
43  43   25 29.5   19 32.5  71    20     F
44  44   NA 28.0   19 17.5  64    15     F
45  45   20 23.0   18  9.5  52    18     F
46  46   24 35.0   19 35.0  75    20     C
47  47   18 32.0   19 17.0  64    20     F
48  48   18 17.0   19 13.0  43    10     F
49  49   20 15.5   18 17.5  43    NA     F
50  50   32 25.5   19 26.5  66    18     F
51  51   43   NA   19 33.0  89    20     B
52  52    6 18.5   18  4.0  35    18     F
53  53   32 29.5   18  5.0  49     8     F
54  54   26 29.5   19 13.0  59    18     F
55  55   37 37.5   18   NA  84    NA     F
56  56   NA 23.5   19 13.0  39    13     F
57  57   18 24.5    3  6.0  28     5     F
> |
```

.....

7.Update the result as “PASS” if their mark is greater than or equal to 50. Else result is “FAIL”

```
> data<-transform(data,result=ifelse(data$gt>=50,'PASS','FAIL'))
> data
   sno     regno                          name school cat1 cat2 da01  fat lab quiz1       gt grade result
1    1 19BCE7478        NITTOOR VISHNU BHARADWAJ    CSE   43 38.5   19 33.0  87    15 83.73333     A   PASS
2    2 19BCN7017                 INTURI REVANTH    CSE   10 12.0   19  3.0  34    15 43.96667     F   FAIL
3    3 19BCN7045     MUMMANI PURNAVENKTASAIKIRAN    CSE   26 37.0   19 14.0  60    18 67.88333     C   PASS
4    4 19BCN7050             NISHIT VERMA       CSE   NA 47.5   19 37.0  93    20 88.17642     A   PASS
5    5 19BCN7064          TADIBOINA ANAND KUMAR    ECE   23 46.0   19 28.0  78    20 80.11667     A   PASS
6    6 19BCN7079           PATTAPU SAI SRINIVAS    ECE   20 31.5   NA 10.0  61    18 63.36944     F   PASS
7    7 19BCN7114          BALIVADA PRATYUSH        CSE   28 41.0   18 31.0  75    20 79.26667     B   PASS
8    8 19BCN7136          AMARA SANTOSH JAYANTH    CSE   39 45.0   19 28.0  85    15 82.46667     A   PASS
9    9 19BCN7137       JAYAPRAKASH SUGAN PRASAD    CSE   14 17.5   19  9.5  37    NA 49.92579     F   FAIL
10  10 19BCE7475               SHOBHIT KHURANA    CSE   38 40.5   19 30.5  86    18 84.21667     A   PASS
11  11 19BCE7487              PRIYANSHI SHARMA    ECE   26 31.0   18 16.0  66    15 65.61667     C   PASS
12  12 19BCN7190       CHEDALAWADA SESHA PRANAV    ECE   32 20.0   18 27.0  69    15 68.11667     C   PASS
13  13 19BCE7124           PERNI BINDU MADHAVI    CSE   16 27.5   19 14.0  51    20 61.48333     F   PASS
14  14 19BCN7021    CHADA RADHA AMUKTHA BHARGAVI    CSE   38 35.0   18 34.0  83    20 83.53333     A   PASS
15  15 19BCN7041          AVVA VENKATA PUNEETH      CSE   36 29.5   18 30.0  73    20 77.46667     B   PASS
16  16 19BCN7075               RONDA PRIYATHAM            34 26.5   18 22.5  61    13 65.28333     C   PASS
17  17 19BCN7092           MADATI HRUTHIK YADAV    ECE   17 25.0   NA 18.0  NA    13 59.54722     F   PASS
18  18 19BES7053         VARIKELA GOPALAKRISHNA    ECE   15 28.5   18 12.5  45    10 51.03333     F   PASS
19  19 19BCE7131                ANANNYA BARUA      CSE   43   NA   19 24.0  77    20 79.67830     B   PASS
20  20 19BCE7166             V VIKRAM GANESH       CSE   43 47.0   19 34.5  93    20 91.95000     S   PASS
21  21 19BCD7074             P.ANUSHRI SOWMYA      CSE   24 23.5   19  7.0  47    10 51.98333     F   PASS
22  22 19BCE7283                 C M ELAKKIA       CSE   28 36.5   19 20.0  66    15 69.43333     C   PASS
23  23 19BCN7185                   M.NAVEEN              37 37.5   19   NA  84    18 79.53182     B   PASS
24  24 19BCN7186                  NISHANTH       ECE   NA 36.0   19 14.0  68    13 66.22642     C   PASS
25  25 19BCN7255                SURENDHAR A      CSE   13 26.0   19 15.0  55    15 58.00000     F   PASS
26  26 19BCE7113              THOTA SURYA TEJA    CSE   28 22.0   19 22.5  60    20 68.18333     C   PASS
27  27 19BCE7653 CHIMAKURTHI B V S N PAVAN KUMAR    CSE   19 34.0    3 29.0  57     5 46.56667     F   FAIL
28  28 19BCN7008                 DUTTA GEETHIK     CSE   21 30.0   19 13.5  61    20 66.35000     C   PASS
29  29 19BCN7225 SEELAM PENCHALA SAI VARA PRASAD    ECE   28   NA   19 20.5  NA    20 70.27275     B   PASS
30  30 19BEC7051                 SIRANJEEVE K      ECE   20 29.0   19  8.0  57    15 59.40000     F   PASS
31  31 19BCN7022               ANDE SANDEEP        CSE   27 32.5   19 12.0  59    18 66.11667     C   PASS
32  32 19BCN7242          KATTA LOKESH CHOWDARY     CSE   10  0.0   19  0.0  19     0 24.21667     F   FAIL
33  33 19BEC7079 PAVULURI KHADYOTHAN SAI TRINADH    CSE   22 18.5   19 16.0  50    13 55.66667     F   PASS
```

23	23	19BCN7185	M.NAVEEN		37	37.5	19	NA	84	18	79.53182	B	PASS
24	24	19BCN7186	NISHANTH	ECE	NA	36.0	19	14.0	68	13	66.22642	C	PASS
25	25	19BCN7255	SURENDHAR A	CSE	13	26.0	19	15.0	55	15	58.00000	F	PASS
26	26	19BCE7113	THOTA SURYA TEJA	CSE	28	22.0	19	22.5	60	20	68.18333	C	PASS
27	27	19BCE7653	CHIMAKURTHI B V S N PAVAN KUMAR	CSE	19	34.0	3	29.0	57	5	46.56667	F	FAIL
28	28	19BCN7008	DUTTA GEETHIK	CSE	21	30.0	19	13.5	61	20	66.35000	C	PASS
29	29	19BCN7225	SEELAM PENCHALA SAI VARA PRASAD	ECE	28	NA	19	20.5	NA	20	70.27275	B	PASS
30	30	19BEC7051	SIRANJEEVE K	ECE	20	29.0	19	8.0	57	15	59.40000	F	PASS
31	31	19BCN7022	ANDE SANDEEP	CSE	27	32.5	19	12.0	59	18	66.11667	C	PASS
32	32	19BCN7242	KATTA LOKESH CHOWDARY	CSE	10	0.0	19	0.0	19	0	24.21667	F	FAIL
33	33	19BEC7079	PAVULURI KHADYOTHAN SAI TRINADH	CSE	22	18.5	19	16.0	50	13	55.66667	F	PASS
34	34	19BCD7243	SIMRAN ANAND	CSE	43	38.0	19	40.0	85	17	86.36667	A	PASS
35	35	19BCE7158	RAYUDU PRANEETH	ECE	38	38.0	19	24.5	82	18	80.75000	A	PASS
36	36	19BCE7168	CHERUKURI SAI RIKSHIT		16	18.5	18	10.0	50	15	53.21667	F	PASS
37	37	19BCN7072	SIMRAN ANAND	CSE	36	42.0	NA	28.0	85	NA	80.97856	A	PASS
38	38	19BCD7062	REDROUTHU LAKSHMAN	CSE	21	25.0	18	5.5	40	13	50.58333	F	PASS
39	39	19BCD7113	BORIGORLA BALARAM NAGA SRINIVAS	CSE	13	22.0	18	10.0	42	18	53.20000	F	PASS
40	40	19BCE7225	SEGU SHANMUKA SRINIVAS	CSE	22	NA	18	28.5	NA	20	70.05608	B	PASS
41	41	19BCN7230	O.LOKESH CHOWDARY	ECE	14	29.0	18	9.0	44	20	57.16667	F	PASS
42	42	19BCN7111	UTTARILLI SREETEJA	ECE	30	29.5	19	23.0	67	20	72.95000	B	PASS
43	43	19BCN7130	ALUR SAMEER ALI KHAN	CSE	25	29.5	19	32.5	71	20	75.35000	B	PASS
44	44	19BCE7192	MANEESH MADALA	CSE	NA	28.0	19	17.5	64	15	65.32642	C	PASS
45	45	19BCE7469	KASARANENI SRIVINEELA	CSE	20	23.0	18	9.5	52	18	58.20000	F	PASS
46	46	19BCE7483	PASALAPUDI CHANUKYA SAI VARDHAN	CSE	24	35.0	19	35.0	75	20	78.41667	B	PASS
47	47	19BCN7069	CHINTALA LOKESH BABU	ECE	18	32.0	19	17.0	64	20	67.91667	C	PASS
48	48	19BCN7243	KANCHETI DILLIKUMAR		18	17.0	19	13.0	43	10	49.05000	F	FAIL
49	49	19BEC7025	SHAIK NAFEESA ANJUM	CSE	20	15.5	18	17.5	43	NA	54.17579	F	PASS
50	50	19BES7048	MARNI MAHENDRA CHOWDARY	CSE	32	25.5	19	26.5	66	18	71.55000	B	PASS
51	51	19BCD7025	BUSSA SAIBIPIN	CSE	43	NA	19	33.0	89	20	85.67830	A	PASS
52	52	19BCE7034	MOOLA SAI UDAY KIRAN KUMAR	CSE	6	18.5	18	4.0	35	18	46.70000	F	FAIL
53	53	19BCE7491	CYRIL AMBEDKAR KONDRU	ECE	32	29.5	18	5.0	49	8	53.53333	F	PASS
54	54	19BCI7098	PENIKALAPATI PRANAY	ECE	26	29.5	19	13.0	59	18	65.31667	C	PASS
55	55	19BCI7011	NISCHAL NANDIGAMA		37	37.5	18	NA	84	NA	77.32427	B	PASS
56	56	19BEC7071	ANAND CHOUDARY JASTI	CSE	NA	23.5	19	13.0	39	13	53.92642	F	PASS
57	57	19BEC7141	KANDE ANISHA	CSE	18	24.5	3	6.0	28	5	28.93333	F	FAIL

>

(Use Regression-Analysis-Data.csv)

1. Perform Exploratory Analysis

```
> data<-read.csv("C:/Users/Sashank K/Downloads/Regression-Analysis-Data.csv")
> data
```

	Observation	Dist_Taxi	Dist_Market	Dist_Hospital	Carpet	Builtup
1	1	9796	5250	10703	1659	1961
2	2	8294	8186	12694	1461	1752
3	3	11001	14399	16991	1340	1609
4	4	8301	11188	12289	1451	1748
5	5	10510	12629	13921	1770	2111
6	6	6665	5142	9972	1442	1733
7	7	13153	11869	17811	1542	1858
8	8	5882	9948	13315	1261	1507
9	9	7495	11589	13370	1090	1321
10	10	8233	7067	11400	1030	1235
11	11	4278	10646	8243	1187	1439
12	12	8066	11149	12936	1751	2098
13	13	7693	9130	14684	1746	2064
14	14	5236	10853	13054	1615	1931
15	15	6027	6707	10176	1469	1756
16	16	9648	14789	12812	1644	1950
17	17	11079	13102	13076	1578	1907
18	18	6698	11519	13441	1703	2045
19	19	9609	9066	13304	1438	1731
20	20	6209	7839	10660	1837	NA
21	21	8155	8085	9837	1940	2340
22	22	9669	12385	13589	1421	1700
23	23	11613	11675	15476	1458	1746
24	24	10664	14161	15076	1719	2065
25	25	7309	7468	11460	1449	1752
26	26	7103	9832	10943	1234	1488
27	27	6615	8178	12804	1732	2073
28	28	10870	12040	15221	1475	1777
29	29	7186	8400	11151	1226	NA
30	30	11735	10687	16954	1390	1648
31	31	13441	13852	17989	1642	1943
32	32	8876	13372	15845	1715	2071
33	33	4066	9318	10159	1439	1746
34	34	8248	8464	11292	1250	1508
35	35	9904	11866	14566	1331	1608
36	36	6526	7876	11215	1784	2163
37	37	6166	9680	12079	1375	1648

38	38	5064	10573	9796	1871	2230
39	39	6712	10341	15629	1442	1744
40	40	8317	13286	12785	1174	1411
41	41	7796	17101	16192	1839	2204
42	42	9171	8368	12566	1270	1516
43	43	8525	11543	13589	1209	NA
44	44	9357	11584	14219	1435	1725
45	45	8101	11176	13502	965	1152
46	46	8794	12199	14275	1665	2001
47	47	6017	14147	12642	1780	2117
48	48	11265	12999	14621	1009	1194
49	49	6407	10134	13433	1227	1471
50	50	NA	NA	10643	1565	1885
51	51	7193	10163	10715	1769	2087
52	52	9723	11620	13693	1660	1982
53	53	11363	14001	17453	1472	1776
54	54	13157	11247	16385	1408	1688
55	55	11657	12184	16450	1514	1820
56	56	10269	7505	13552	1565	1880
57	57	9480	15910	16887	1074	1288
58	58	10451	11465	16488	1864	2240
59	59	7331	11247	10787	1570	1898
60	60	8458	13941	15721	1417	1701
61	61	9620	15033	14775	1734	2060
62	62	10566	10728	16951	1470	1763
63	63	5568	11538	14767	1761	2104
64	64	8210	9740	12485	1756	2070
65	65	4533	6617	10052	1704	2045
66	66	7471	9029	14080	NA	1614
67	67	6239	7326	9085	2011	2391
68	68	10504	9938	13747	1472	1748
69	69	12715	10229	15329	1310	1561
70	70	9472	9686	13464	1544	1821
71	71	4195	9050	7517	1707	2052
72	72	5858	8114	11026	1881	2262
73	73	9849	8773	15371	1416	1681
74	74	7983	6338	11916	1631	1941
75	75	5174	9293	10704	1318	1576
76	76	7629	8786	12052	1692	2019
77	77	8841	11909	11625	1152	1380
78	78	9327	9867	14244	891	1073
79	79	6637	10365	9537	1468	1749

80	80	4589	12404	12558	1539	1833
81	81	9604	13622	15501	1635	1956
82	82	8868	12356	13209	1267	1520
83	83	11676	14735	15825	1250	1475
84	84	11703	12778	16910	1598	NA
85	85	5909	10455	13125	1720	2044
86	86	9674	12676	14459	1462	1761
87	87	4997	9411	11468	1431	1711
88	88	9371	12701	16615	1539	1858
89	89	14306	12591	18029	1441	1723
90	90	6591	14567	13985	1572	1884
91	91	7566	9599	16454	1287	1525
92	92	8463	8825	7307	1468	1760
93	93	1200	7579	7376	1931	2342
94	94	6871	10557	12097	1252	1506
95	95	6956	9048	11991	1238	1468
96	96	6541	7537	11474	1694	NA
97	97	2222	10106	8858	1479	1758
98	98	6194	8473	11978	1590	1912
99	99	7025	10344	11682	2169	2617
100	100	9233	12290	14032	1838	2205

	Parking	City_Category	Rainfall	House_Price
1	Open	CAT B	530	6649000
2	Not Provided	CAT B	210	3982000
3	Not Provided	CAT A	720	5401000
4	Covered	CAT B	620	5373000
5	Not Provided	CAT B	450	4662000
6	Open	CAT B	760	4526000
7	No Parking	CAT A	1030	7224000
8	Open	CAT C	1020	3772000
9	Not Provided	CAT B	680	4631000
10	Open	CAT C	1130	4415000
11	Covered	CAT A	1090	7128000
12	No Parking	CAT B	720	5762000
13	Open	CAT B	1050	6047000
14	Covered	CAT B	1160	5913000
15	Open	CAT B	770	6636000
16	Covered	CAT A	790	7887000
17	Open	CAT A	1440	7725000
18	Covered	CAT B	670	3817000
19	Open	CAT A	1030	6354000
20	Open	CAT B	790	4922000

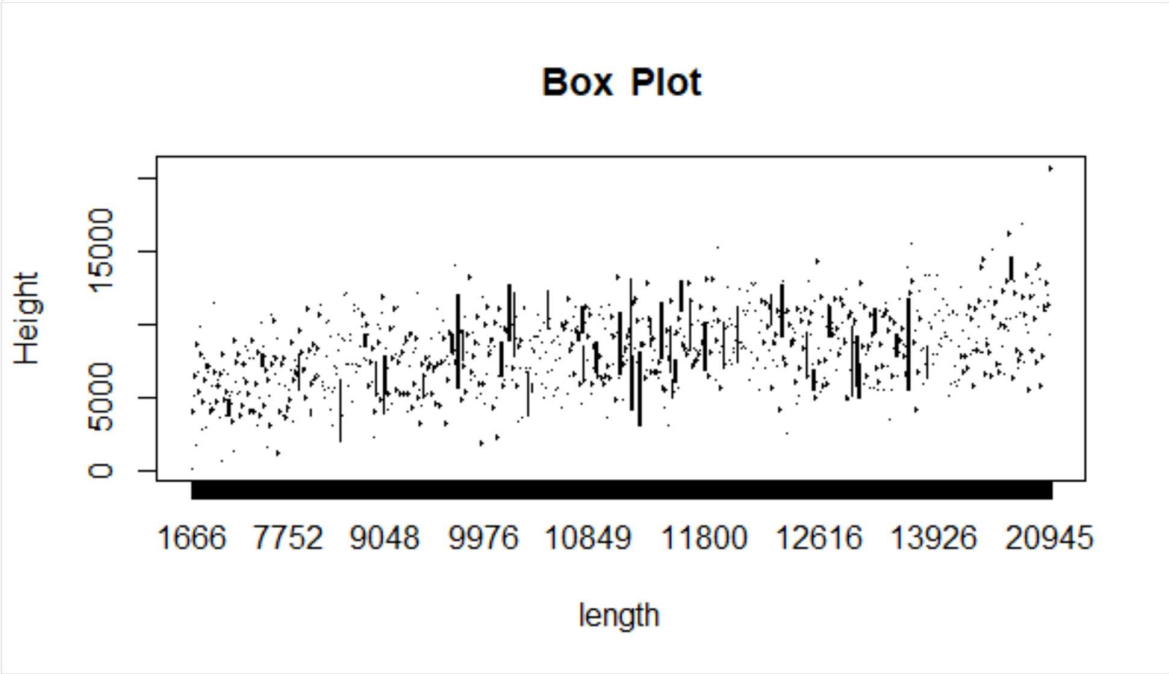
21	Covered	CAT C	980	4019000
22	No Parking	CAT C	370	4346000
23	Open	CAT A	690	6889000
24	Covered	CAT B	950	5278000
25	Not Provided	CAT B	620	5068000
26	No Parking	CAT B	840	4188000
27	Not Provided	CAT C	820	7005000
28	Open	CAT C	1100	2582000
29	Open	CAT C	760	4766000
30	No Parking	CAT B	980	6053000
31	No Parking	CAT B	710	7810000
32	Open	CAT B	650	8486000
33	Open	CAT A	990	8014000
34	Open	CAT C	990	1492000
35	Open	CAT B	880	6046000
36	Covered	CAT B	620	7456000
37	Not Provided	CAT A	1020	7243000
38	Covered	CAT B	700	4546000
39	Not Provided	CAT C	610	4157000
40	Covered	CAT A	1080	6287000
41	No Parking	CAT B	1010	5517000
42	Not Provided	CAT B	0	4548000
43	Open	CAT A	1200	6041000
44	Open	CAT C	1250	4955000
45	Not Provided	CAT C	600	4814000
46	No Parking	CAT A	730	6764000
47	Open	CAT C	780	3973000
48	Not Provided	CAT C	520	3580000
49	No Parking	CAT A	870	4927000
50	Not Provided	CAT A	660	6707000
51	No Parking	CAT B	690	6651000
52	Open	CAT B	910	6253000
53	Open	CAT B	1260	5998000
54	Not Provided	CAT A	1040	7635000
55	Covered	CAT B	910	8182000
56	Covered	CAT B	1300	5783000
57	Not Provided	CAT B	320	7045000
58	No Parking	CAT B	530	6700000
59	No Parking	CAT B	980	6409000
60	Open	CAT B	740	4867000
61	Not Provided	CAT B	1240	6621000
62	Covered	CAT B	1080	8366000

```
63  Not Provided         CAT A     1080     7077000
64          Open         CAT C      460     5387000
65    No Parking         CAT A      300     7130000
66       Covered         CAT A      870     7669000
67  Not Provided         CAT C      530     4610000
68          Open         CAT B      600     4910000
69  Not Provided         CAT B      860     6592000
70       Covered         CAT B      590     5866000
71          Open         CAT B      920     6913000
72    No Parking         CAT C      570     4908000
73  Not Provided         CAT A      290     7271000
74          Open         CAT B      650     3343000
75       Covered         CAT C      710     4243000
76          Open         CAT B      850     5665000
77  Not Provided         CAT B      530     3358000
78  Not Provided         CAT A      630     6737000
79  Not Provided         CAT A      700     7178000
80  Not Provided         CAT A      650     8484000
81       Covered         CAT A      720     8207000
82  Not Provided         CAT B      450     2677000
83          Open         CAT B     1390     6556000
84       Covered         CAT A      530     8116000
85          Open         CAT B      960     3866000
86  Not Provided         CAT A      600     6566000
87    No Parking         CAT C      620     4070000
88  Not Provided         CAT A     1020     8891000
89          Open         CAT B      330     5786000
90    No Parking         CAT C     1410     4267000
91  Not Provided         CAT A     1200     9018000
92  Not Provided         CAT C      280     5128000
93       Covered         CAT A      940     9726000
94          Open         CAT B      850     3965000
95          Open         CAT C      960     4572000
96       Covered         CAT B      560     4830000
97       Covered         CAT C      420     4206000
98          Open         CAT B      830     6535000
99          Open         CAT B      600     6708000
100      Covered         CAT B      400     5403000
 [ reached 'max' / getOption("max.print") -- omitted 832 rows ]
> |
```

2. Perform visual Exploratory Analysis

```
#Perform Exploratory Analysis
 str(data)
 head(data,5)
 colSums(is.na(data))
 boxplot(Dist_Taxi~Dist_Market,data=file1,xlab="length",ylab="Height",main="Box Plot",col=c("green","orange"))
```

```
> #Perform Exploratory Analysis
>             str(data)
'data.frame':   932 obs. of  10 variables:
 $ Observation  : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Dist_Taxi    : int  9796 8294 11001 8301 10510 6665 13153 5882 7495 8233 ...
 $ Dist_Market  : int  5250 8186 14399 11188 12629 5142 11869 9948 11589 7067 ...
 $ Dist_Hospital: int  10703 12694 16991 12289 13921 9972 17811 13315 13370 11400 ...
 $ Carpet       : int  1659 1461 1340 1451 1770 1442 1542 1261 1090 1030 ...
 $ Builtup      : int  1961 1752 1609 1748 2111 1733 1858 1507 1321 1235 ...
 $ Parking      : chr  "Open" "Not Provided" "Not Provided" "Covered" ...
 $ City_Category: chr  "CAT B" "CAT B" "CAT A" "CAT B" ...
 $ Rainfall     : int  530 210 720 620 450 760 1030 1020 680 1130 ...
 $ House_Price  : int  6649000 3982000 5401000 5373000 4662000 4526000 7224000 3772000 4631000 4415000 ...
>             head(data,5)
  Observation Dist_Taxi Dist_Market Dist_Hospital Carpet Builtup      Parking City_Category Rainfall House_Price
1           1      9796        5250         10703   1659    1961         Open          CAT B      530     6649000
2           2      8294        8186         12694   1461    1752 Not Provided          CAT B      210     3982000
3           3     11001       14399         16991   1340    1609 Not Provided          CAT A      720     5401000
4           4      8301       11188         12289   1451    1748      Covered          CAT B      620     5373000
5           5     10510       12629         13921   1770    2111 Not Provided          CAT B      450     4662000
>             colSums(is.na(data))
  Observation     Dist_Taxi   Dist_Market Dist_Hospital        Carpet       Builtup       Parking City_Category      Rainfall   House_Pr
ice
            0            13            13             1             8            15             0             0             0
  0
>             boxplot(Dist_Taxi~Dist_Market,data=file1,xlab="length",ylab="Height",main="Box□Plot",col=c("green","orange"))
> |
```



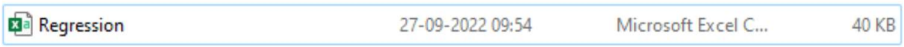
3 Perform Data cleaning operation and upload the corrected csv file.

```
write.csv(file1,"Regression.csv",row.names=FALSE)


#Deleting/Selecting a particular column and are stored in different.
 data_2<-data[,7]
 data_2

 #Perform Data Imputation using Mean/ Mode/ Median Imputation
```

Output:



(Use iris.csv)

1. Perform Data Imputation using **Deletion**

```
#Deleting/Selecting a particular column and are stored in different.
 data_2<-data[,7]
 data_2

 data_3 <- data[,-2:-4]
 data_3
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Weight.in.gm	Species	Season
1	1	5.1	3.5	1.4	0.2	20	Iris-setosa	spring
2	2	4.9	3.0	1.4	0.2	35	Iris-setosa	summer
3	3	4.7	3.2	1.3	0.2	33	Iris-setosa	fall
4	4	4.6	3.1	1.5	0.2	27	Iris-setosa	winter
5	5	5.0	3.6	1.4	0.2	41	Iris-setosa	spring
6	6	5.4	3.9	1.7	0.4	17	Iris-setosa	summer
7	7	4.6	3.4	1.4	0.3	5	Iris-setosa	fall
8	8	5.0	3.4	1.5	0.2	33	Iris-setosa	winter
9	9	4.4	2.9	1.4	0.2	5	Iris-setosa	spring
10	10	4.9	3.1	1.5	0.1	31	Iris-setosa	summer
11	11	5.4	3.7	1.5	0.2	33	Iris-setosa	fall
12	12	4.8	3.4	1.6	0.2	7	Iris-setosa	winter
13	13	4.8	3.0	1.4	0.1	NA	Iris-setosa	spring
14	14	4.3	3.0	1.1	0.1	19	Iris-setosa	summer
15	15	5.8	4.0	1.2	0.2	8	Iris-setosa	fall
16	16	5.7	4.4	1.5	0.4	19	Iris-setosa	winter
17	17	5.4	3.9	1.3	0.4	38	Iris-setosa	spring
18	18	5.1	3.5	1.4	0.3	5	Iris-setosa	summer
19	19	5.7	3.8	1.7	0.3	35	Iris-setosa	fall
20	20	5.1	3.8	1.5	0.3	24	Iris-setosa	winter
21	21	5.4	3.4	1.7	0.2	14	Iris-setosa	spring
22	22	5.1	3.7	1.5	0.4	8	Iris-setosa	summer
23	23	4.6	3.6	1.0	0.2	29	Iris-setosa	fall
24	24	5.1	3.3	1.7	0.5	41	Iris-setosa	winter
25	25	4.8	3.4	1.9	0.2	24	Iris-setosa	spring
26	26	5.0	3.0	1.6	0.2	NA	Iris-setosa	summer
27	27	5.0	3.4	1.6	0.4	41	Iris-setosa	fall
28	28	5.2	3.5	1.5	0.2	7	Iris-setosa	winter
29	29	5.2	3.4	1.4	0.2	20	Iris-setosa	spring
30	30	4.7	3.2	1.6	0.2	28	Iris-setosa	summer
31	31	4.8	3.1	1.6	0.2	3	Iris-setosa	fall
32	32	5.4	3.4	1.5	0.4	3	Iris-setosa	winter
33	33	5.2	4.1	1.5	0.1	15	Iris-setosa	spring
34	34	5.5	4.2	1.4	0.2	25	Iris-setosa	summer

```
36   36          5.0         3.2          1.2         0.2          8     Iris-setosa winter
37   37          5.5         3.5          1.3         0.2         39     Iris-setosa spring
38   38          4.9         3.1          1.5         0.1         37     Iris-setosa summer
39   39          4.4         3.0          1.3         0.2         35     Iris-setosa   fall
40   40          5.1         3.4          1.5         0.2         NA     Iris-setosa winter
41   41          5.0         3.5          1.3         0.3         40     Iris-setosa spring
42   42          4.5         2.3          1.3         0.3         37     Iris-setosa summer
43   43          4.4         3.2          1.3         0.2         32     Iris-setosa   fall
44   44          5.0         3.5          1.6         0.6         25     Iris-setosa winter
45   45          5.1         3.8          1.9         0.4         19     Iris-setosa spring
46   46          4.8         3.0          1.4         0.3          6     Iris-setosa summer
47   47          5.1         3.8          1.6         0.2         NA     Iris-setosa   fall
48   48          4.6         3.2          1.4         0.2         40     Iris-setosa winter
49   49          5.3         3.7          1.5         0.2         24     Iris-setosa spring
50   50          5.0         3.3          1.4         0.2         37     Iris-setosa summer
51   51          7.0         3.2          4.7         1.4         42 Iris-versicolor   fall
52   52          6.4         3.2          4.5         1.5         19 Iris-versicolor winter
53   53          6.9         3.1          4.9         1.5         10 Iris-versicolor spring
54   54          5.5         2.3          4.0         1.3         29 Iris-versicolor summer
55   55          6.5         2.8          4.6         1.5          6 Iris-versicolor   fall
56   56          5.7         2.8          4.5         1.3         NA Iris-versicolor winter
57   57          6.3         3.3          4.7         1.6         19 Iris-versicolor spring
58   58          4.9         2.4          3.3         1.0         36 Iris-versicolor summer
59   59          6.6         2.9          4.6         1.3         16 Iris-versicolor   fall
60   60          5.2         2.7          3.9         1.4         32 Iris-versicolor winter
61   61          5.0         2.0          3.5         1.0         12 Iris-versicolor spring
62   62          5.9         3.0          4.2         1.5         25 Iris-versicolor summer
63   63          6.0         2.2          4.0         1.0         NA Iris-versicolor   fall
64   64          6.1         2.9          4.7         1.4         16 Iris-versicolor winter
65   65          5.6         2.9          3.6         1.3         29 Iris-versicolor spring
66   66          6.7         3.1          4.4         1.4         29 Iris-versicolor summer
67   67          5.6         3.0          4.5         1.5         36 Iris-versicolor   fall
68   68          5.8         2.7          4.1         1.0         31 Iris-versicolor winter
69   69          6.2         2.2          4.5         1.5         32 Iris-versicolor spring
70   70          5.6         2.5          3.9         1.1         23 Iris-versicolor summer
71   71          5.9         3.2          4.8         1.8         12 Iris-versicolor   fall
72   72          6.1         2.8          4.0         1.3         14 Iris-versicolor winter
91   91          5.5         2.6          4.4         1.2         20 Iris-versicolor   fall
92   92          6.1         3.0          4.6         1.4         NA Iris-versicolor winter
93   93          5.8         2.6          4.0         1.2         29 Iris-versicolor spring
94   94          5.0         2.3          3.3         1.0         33 Iris-versicolor summer
95   95          5.6         2.7          4.2         1.3         39 Iris-versicolor   fall
96   96          5.7         3.0          4.2         1.2         18 Iris-versicolor winter
97   97          5.7         2.9          4.2         1.3         14 Iris-versicolor spring
98   98          6.2         2.9          4.3         1.3         23 Iris-versicolor summer
99   99          5.1         2.5          3.0         1.1         27 Iris-versicolor   fall
100 100          5.7         2.8          4.1         1.3         24 Iris-versicolor winter
101 101          6.3         3.3          6.0         2.5         18  Iris-virginica spring
102 102          5.8         2.7          5.1         1.9         NA  Iris-virginica summer
103 103          7.1         3.0          5.9         2.1         10  Iris-virginica   fall
104 104          6.3         2.9          5.6         1.8         19  Iris-virginica winter
105 105          6.5         3.0          5.8         2.2         37  Iris-virginica spring
106 106          7.6         3.0          6.6         2.1          8  Iris-virginica summer
107 107          4.9         2.5          4.5         1.7         29  Iris-virginica   fall
108 108          7.3         2.9          6.3         1.8         10  Iris-virginica winter
109 109          6.7         2.5          5.8         1.8         42  Iris-virginica spring
110 110          7.2         3.6          6.1         2.5         42  Iris-virginica summer
111 111          6.5         3.2          5.1         2.0         40  Iris-virginica   fall
112 112          6.4         2.7          5.3         1.9         17  Iris-virginica winter
113 113          6.8         3.0          5.5         2.1         17  Iris-virginica spring
114 114          5.7         2.5          5.0         2.0         NA  Iris-virginica summer
115 115          5.8         2.8          5.1         2.4         40  Iris-virginica   fall
116 116          6.4         3.2          5.3         2.3         NA  Iris-virginica winter
117 117          6.5         3.0          5.5         1.8         24  Iris-virginica spring
118 118          7.7         3.8          6.7         2.2         18  Iris-virginica summer
119 119          7.7         2.6          6.9         2.3         32  Iris-virginica   fall
120 120          6.0         2.2          5.0         1.5         11  Iris-virginica winter
121 121          6.9         3.2          5.7         2.3         14  Iris-virginica spring
122 122          5.6         2.8          4.9         2.0         40  Iris-virginica summer
123 123          7.7         2.8          6.7         2.0         37  Iris-virginica   fall
124 124          6.3         2.7          4.9         1.8         21  Iris-virginica winter
125 125          6.7         3.3          5.7         2.1         22  Iris-virginica spring
 [ reached 'max' / getOption("max.print") -- omitted 25 rows ]
> |
```

```
>  #Deleting/Selecting a particular column and are stored in different.
>                data_2<-data[,7]
>                data_2
  [1] "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"
  [8] "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"
 [15] "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"
 [22] "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"
 [29] "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"
 [36] "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"
 [43] "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"     "Iris-setosa"
 [50] "Iris-setosa"     "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor"
 [57] "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor"
 [64] "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor"
 [71] "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor"
 [78] "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor"
 [85] "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor"
 [92] "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor" "Iris-versicolor"
 [99] "Iris-versicolor" "Iris-versicolor" "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"
[106] "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"
[113] "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"
[120] "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"
[127] "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"
[134] "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"
[141] "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"  "Iris-virginica"
[148] "Iris-virginica"  "Iris-virginica"  "Iris-virginica"
>
```

```
>                data_3 <- data[,-2:-4]
>                data_3
     Id PetalWidthCm Weight.in.gm         Species Season
1     1          0.2           20     Iris-setosa spring
2     2          0.2           35     Iris-setosa summer
3     3          0.2           33     Iris-setosa   fall
4     4          0.2           27     Iris-setosa winter
5     5          0.2           41     Iris-setosa spring
6     6          0.4           17     Iris-setosa summer
7     7          0.3            5     Iris-setosa   fall
8     8          0.2           33     Iris-setosa winter
9     9          0.2            5     Iris-setosa spring
10   10          0.1           31     Iris-setosa summer
11   11          0.2           33     Iris-setosa   fall
12   12          0.2            7     Iris-setosa winter
13   13          0.1           NA     Iris-setosa spring
14   14          0.1           19     Iris-setosa summer
15   15          0.2            8     Iris-setosa   fall
16   16          0.4           19     Iris-setosa winter
17   17          0.4           38     Iris-setosa spring
18   18          0.3            5     Iris-setosa summer
19   19          0.3           35     Iris-setosa   fall
20   20          0.3           24     Iris-setosa winter
21   21          0.2           14     Iris-setosa spring
22   22          0.4            8     Iris-setosa summer
23   23          0.2           29     Iris-setosa   fall
24   24          0.5           41     Iris-setosa winter
25   25          0.2           24     Iris-setosa spring
26   26          0.2           NA     Iris-setosa summer
27   27          0.4           41     Iris-setosa   fall
28   28          0.2            7     Iris-setosa winter
29   29          0.2           20     Iris-setosa spring
30   30          0.2           28     Iris-setosa summer
31   31          0.2            3     Iris-setosa   fall
32   32          0.4            3     Iris-setosa winter
33   33          0.1           15     Iris-setosa spring
34   34          0.2           25     Iris-setosa summer
```

43	43	0.2	32	Iris-setosa	fall
44	44	0.6	25	Iris-setosa	winter
45	45	0.4	19	Iris-setosa	spring
46	46	0.3	6	Iris-setosa	summer
47	47	0.2	NA	Iris-setosa	fall
48	48	0.2	40	Iris-setosa	winter
49	49	0.2	24	Iris-setosa	spring
50	50	0.2	37	Iris-setosa	summer
51	51	1.4	42	Iris-versicolor	fall
52	52	1.5	19	Iris-versicolor	winter
53	53	1.5	10	Iris-versicolor	spring
54	54	1.3	29	Iris-versicolor	summer
55	55	1.5	6	Iris-versicolor	fall
56	56	1.3	NA	Iris-versicolor	winter
57	57	1.6	19	Iris-versicolor	spring
58	58	1.0	36	Iris-versicolor	summer
59	59	1.3	16	Iris-versicolor	fall
60	60	1.4	32	Iris-versicolor	winter
61	61	1.0	12	Iris-versicolor	spring
62	62	1.5	25	Iris-versicolor	summer
63	63	1.0	NA	Iris-versicolor	fall
64	64	1.4	16	Iris-versicolor	winter
65	65	1.3	29	Iris-versicolor	spring
66	66	1.4	29	Iris-versicolor	summer
67	67	1.5	36	Iris-versicolor	fall
68	68	1.0	31	Iris-versicolor	winter
69	69	1.5	32	Iris-versicolor	spring
70	70	1.1	23	Iris-versicolor	summer
71	71	1.8	12	Iris-versicolor	fall
72	72	1.3	14	Iris-versicolor	winter
73	73	1.5	16	Iris-versicolor	spring
74	74	1.2	14	Iris-versicolor	summer
75	75	1.3	NA	Iris-versicolor	fall
76	76	1.4	15	Iris-versicolor	winter
77	77	1.4	38	Iris-versicolor	spring
78	78	1.7	24	Iris-versicolor	summer
79	79	1.5	7	Iris-versicolor	fall
115	115	2.4	40	Iris-virginica	fall
116	116	2.3	NA	Iris-virginica	winter
117	117	1.8	24	Iris-virginica	spring
118	118	2.2	18	Iris-virginica	summer
119	119	2.3	32	Iris-virginica	fall
120	120	1.5	11	Iris-virginica	winter
121	121	2.3	14	Iris-virginica	spring
122	122	2.0	40	Iris-virginica	summer
123	123	2.0	37	Iris-virginica	fall
124	124	1.8	21	Iris-virginica	winter
125	125	2.1	22	Iris-virginica	spring
126	126	1.8	12	Iris-virginica	summer
127	127	1.8	38	Iris-virginica	fall
128	128	1.8	16	Iris-virginica	winter
129	129	2.1	32	Iris-virginica	spring
130	130	1.6	36	Iris-virginica	summer
131	131	1.9	20	Iris-virginica	fall
132	132	2.0	14	Iris-virginica	winter
133	133	2.2	33	Iris-virginica	spring
134	134	1.5	NA	Iris-virginica	summer
135	135	1.4	38	Iris-virginica	fall
136	136	2.3	31	Iris-virginica	winter
137	137	2.4	19	Iris-virginica	spring
138	138	1.8	40	Iris-virginica	summer
139	139	1.8	11	Iris-virginica	fall
140	140	2.1	15	Iris-virginica	winter
141	141	2.4	28	Iris-virginica	spring
142	142	2.3	12	Iris-virginica	summer
143	143	1.9	24	Iris-virginica	fall
144	144	2.3	13	Iris-virginica	winter
145	145	2.5	9	Iris-virginica	spring
146	146	2.3	11	Iris-virginica	summer
147	147	1.9	20	Iris-virginica	fall
148	148	2.0	18	Iris-virginica	winter
149	149	2.3	12	Iris-virginica	spring
150	150	1.8	11	Iris-virginica	summer

>

2.Perform Data Imputation using **Mean/ Mode/ Median Imputation**

```
>                  #Perform Data Imputation using Mean/ Mode/ Median Imputation
>                   mean(data$SepalLengthCm)
[1] 5.843333
>                   median(data$SepalLengthCm)
[1] 5.8
>                     mean(data$SepalWidthCm)
[1] 3.054
>                   median(data$SepalWidthCm)
[1] 3
>                     mean(data$PetalLengthCm)
[1] 3.758667
>                   median(data$PetalLengthCm)
[1] 4.35
>                     mean(data$PetalWidthCm)
[1] 1.198667
>                   median(data$PetalWidthCm)
[1] 1.3
>                     getmode <- function(v) {
+                       uniqv <- unique(v)
+                       uniqv[which.max(tabulate(match(v, uniqv)))]
+                       }
>                   getmode(data$Species)
[1] "Iris-setosa"
>
```

3. Perform Data Imputation using **Prediction Model**

```
>                       iris<-data
>                       split<-sample.split(iris,SplitRatio = 0.8)
>                        split
  [1]  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE
```

```
> train<-subset(iris,split==TRUE)
>                  train
    Observation Dist_Taxi Dist_Market Dist_Hospital Carpet Builtup      Parking City_Category Rainfall House_Price
1             1      9796        5250         10703   1659    1961         Open         CAT B      530     6649000
2             2      8294        8186         12694   1461    1752 Not Provided         CAT B      210     3982000
3             3     11001       14399         16991   1340    1609 Not Provided         CAT A      720     5401000
4             4      8301       11188         12289   1451    1748      Covered         CAT B      620     5373000
7             7     13153       11869         17811   1542    1858   No Parking         CAT A     1030     7224000
8             8      5882        9948         13315   1261    1507         Open         CAT C     1020     3772000
9             9      7495       11589         13370   1090    1321 Not Provided         CAT B      680     4631000
10           10      8233        7067         11400   1030    1235         Open         CAT C     1130     4415000
11           11      4278       10646          8243   1187    1439      Covered         CAT A     1090     7128000
12           12      8066       11149         12936   1751    2098   No Parking         CAT B      720     5762000
13           13      7693        9130         14684   1746    2064         Open         CAT B     1050     6047000
14           14      5236       10853         13054   1615    1931      Covered         CAT B     1160     5913000
17           17     11079       13102         13076   1578    1907         Open         CAT A     1440     7725000
18           18      6698       11519         13441   1703    2045      Covered         CAT B      670     3817000
19           19      9609        9066         13304   1438    1731         Open         CAT A     1030     6354000
20           20      6209        7839         10660   1837      NA         Open         CAT B      790     4922000
21           21      8155        8085          9837   1940    2340      Covered         CAT C      980     4019000
22           22      9669       12385         13589   1421    1700   No Parking         CAT C      370     4346000
23           23     11613       11675         15476   1458    1746         Open         CAT A      690     6889000
24           24     10664       14161         15076   1719    2065      Covered         CAT B      950     5278000
27           27      6615        8178         12804   1732    2073 Not Provided         CAT C      820     7005000
28           28     10870       12040         15221   1475    1777         Open         CAT C     1100     2582000
29           29      7186        8400         11151   1226      NA         Open         CAT C      760     4766000
30           30     11735       10687         16954   1390    1648   No Parking         CAT B      980     6053000
31           31     13441       13852         17989   1642    1943   No Parking         CAT B      710     7810000
32           32      8876       13372         15845   1715    2071         Open         CAT B      650     8486000
33           33      4066        9318         10159   1439    1746         Open         CAT A      990     8014000
34           34      8248        8464         11292   1250    1508         Open         CAT C      990     1492000
37           37      6166        9680         12079   1375    1648 Not Provided         CAT A     1020     7243000
38           38      5064       10573          9796   1871    2230      Covered         CAT B      700     4546000
39           39      6712       10341         15629   1442    1744 Not Provided         CAT C      610     4157000
40           40      8317       13286         12785   1174    1411      Covered         CAT A     1080     6287000
41           41      7796       17101         16192   1839    2204   No Parking         CAT B     1010     5517000
```

50	50	NA	NA	10643	1565	1885	Not Provided	CAT A	660	6707000	
51	51	7193	10163	10715	1769	2087	No Parking	CAT B	690	6651000	
52	52	9723	11620	13693	1660	1982	Open	CAT B	910	6253000	
53	53	11363	14001	17453	1472	1776	Open	CAT B	1260	5998000	
54	54	13157	11247	16385	1408	1688	Not Provided	CAT A	1040	7635000	
57	57	9480	15910	16887	1074	1288	Not Provided	CAT B	320	7045000	
58	58	10451	11465	16488	1864	2240	No Parking	CAT B	530	6700000	
59	59	7331	11247	10787	1570	1898	No Parking	CAT B	980	6409000	
60	60	8458	13941	15721	1417	1701	Open	CAT B	740	4867000	
61	61	9620	15033	14775	1734	2060	Not Provided	CAT B	1240	6621000	
62	62	10566	10728	16951	1470	1763	Covered	CAT B	1080	8366000	
63	63	5568	11538	14767	1761	2104	Not Provided	CAT A	1080	7077000	
64	64	8210	9740	12485	1756	2070	Open	CAT C	460	5387000	
67	67	6239	7326	9085	2011	2391	Not Provided	CAT C	530	4610000	
68	68	10504	9938	13747	1472	1748	Open	CAT B	600	4910000	
69	69	12715	10229	15329	1310	1561	Not Provided	CAT B	860	6592000	
70	70	9472	9686	13464	1544	1821	Covered	CAT B	590	5866000	
71	71	4195	9050	7517	1707	2052	Open	CAT B	920	6913000	
72	72	5858	8114	11026	1881	2262	No Parking	CAT C	570	4908000	
73	73	9849	8773	15371	1416	1681	Not Provided	CAT A	290	7271000	
74	74	7983	6338	11916	1631	1941	Open	CAT B	650	3343000	
77	77	8841	11909	11625	1152	1380	Not Provided	CAT B	530	3358000	
78	78	9327	9867	14244	891	1073	Not Provided	CAT A	630	6737000	
79	79	6637	10365	9537	1468	1749	Not Provided	CAT A	700	7178000	
80	80	4589	12404	12558	1539	1833	Not Provided	CAT A	650	8484000	
81	81	9604	13622	15501	1635	1956	Covered	CAT A	720	8207000	
82	82	8868	12356	13209	1267	1520	Not Provided	CAT B	450	2677000	
83	83	11676	14735	15825	1250	1475	Open	CAT B	1390	6556000	
84	84	11703	12778	16910	1598	NA	Covered	CAT A	530	8116000	
87	87	4997	9411	11468	1431	1711	No Parking	CAT C	620	4070000	
88	88	9371	12701	16615	1539	1858	Not Provided	CAT A	1020	8891000	
89	89	14306	12591	18029	1441	1723	Open	CAT B	330	5786000	
90	90	6591	14567	13985	1572	1884	No Parking	CAT C	1410	4267000	
91	91	7566	9599	16454	1287	1525	Not Provided	CAT A	1200	9018000	
92	92	8463	8825	7307	1468	1760	Not Provided	CAT C	280	5128000	
93	93	1200	7579	7376	1931	2342	Covered	CAT A	940	9726000	
94	94	6871	10557	12097	1252	1506	Open	CAT B	850	3965000	
119	119	9588	10711	13963	1485	1787	Open	CAT B	800	6295000	
120	120	8276	11728	11932	1897	2248	Not Provided	CAT A	1330	5601000	
121	121	7739	12930	12703	1607	1927	No Parking	CAT B	940	4508000	
122	122	10106	13805	16769	1909	2287	Not Provided	CAT B	1210	4683000	
123	123	8510	10891	13380	1274	1503	Open	CAT A	660	6475000	
124	124	5553	13174	11586	1157	1379	Not Provided	CAT A	770	8078000	

[reached 'max' / getOption("max.print") -- omitted 646 rows]

```
>
>                   test<-subset(iris,split==FALSE)
>                    test
```

	Observation	Dist_Taxi	Dist_Market	Dist_Hospital	Carpet	Builtup	Parking	City_Category	Rainfall	House_Price
5	5	10510	12629	13921	1770	2111	Not Provided	CAT B	450	4662000
6	6	6665	5142	9972	1442	1733	Open	CAT B	760	4526000
15	15	6027	6707	10176	1469	1756	Open	CAT B	770	6636000
16	16	9648	14789	12812	1644	1950	Covered	CAT A	790	7887000
25	25	7309	7468	11460	1449	1752	Not Provided	CAT B	620	5068000
26	26	7103	9832	10943	1234	1488	No Parking	CAT B	840	4188000
35	35	9904	11866	14566	1331	1608	Open	CAT B	880	6046000
36	36	6526	7876	11215	1784	2163	Covered	CAT B	620	7456000
45	45	8101	11176	13502	965	1152	Not Provided	CAT C	600	4814000
46	46	8794	12199	14275	1665	2001	No Parking	CAT A	730	6764000
55	55	11657	12184	16450	1514	1820	Covered	CAT B	910	8182000
56	56	10269	7505	13552	1565	1880	Covered	CAT B	1300	5783000
65	65	4533	6617	10052	1704	2045	No Parking	CAT A	300	7130000
66	66	7471	9029	14080	NA	1614	Covered	CAT A	870	7669000
75	75	5174	9293	10704	1318	1576	Covered	CAT C	710	4243000
76	76	7629	8786	12052	1692	2019	Open	CAT B	850	5665000
85	85	5909	10455	13125	1720	2044	Open	CAT B	960	3866000
86	86	9674	12676	14459	1462	1761	Not Provided	CAT A	600	6566000
95	95	6956	9048	11991	1238	1468	Open	CAT C	960	4572000
96	96	6541	7537	11474	1694	NA	Covered	CAT B	560	4830000
105	105	5725	16967	11221	1816	2171	Open	CAT A	1160	6118000
106	106	13866	13472	15427	1785	2147	Open	CAT C	820	6366000
115	115	11475	14768	15998	1929	2311	Open	CAT B	630	7921000
116	116	10148	10679	11618	1573	1878	Not Provided	CAT C	650	2374000
125	125	7522	11974	11433	1712	2046	Not Provided	CAT C	460	3118000
126	126	8702	13088	14238	1500	1798	No Parking	CAT C	860	5671000

335	335	7492	12257	11623	1513	1825	Covered	CAT A	980	7595000
336	336	9464	10762	13998	1208	1459	Open	CAT C	930	4149000
345	345	7259	11982	13297	1640	1948	Covered	CAT C	960	5127000
346	346	7197	7872	11113	1060	1273	Open	CAT B	860	5750000
355	355	7364	7633	12153	966	1172	Not Provided	CAT A	900	8547000
356	356	8692	9271	12352	1922	2290	Open	CAT A	290	8041000
365	365	9666	10668	16002	1766	2138	Open	CAT B	580	4998000
366	366	7318	9441	12192	1362	1634	Not Provided	CAT A	770	8294000
375	375	9415	12467	12841	1514	1822	Covered	CAT C	670	4809000
376	376	6792	11510	14652	1015	1216	No Parking	CAT C	460	2731000
385	385	10027	11992	15660	1207	1476	Open	CAT A	720	8737000
386	386	5558	8275	10846	1484	1774	No Parking	CAT A	940	5980000
395	395	7508	10841	11927	1496	1774	Not Provided	CAT A	810	7549000
396	396	3772	6427	10119	1412	1699	Open	CAT B	680	3920000
405	405	6860	12712	12638	1491	1800	Not Provided	CAT A	980	7562000
406	406	7270	9499	11654	1415	1697	Open	CAT B	1200	3942000
415	415	7935	7911	10691	1546	1829	Not Provided	CAT B	380	2772000
416	416	5787	14151	14705	1689	2044	Open	CAT C	1140	3226000
425	425	5440	12550	11734	1613	1893	Not Provided	CAT C	1180	4603000
426	426	11037	9154	12853	1151	NA	Open	CAT B	1300	6701000
435	435	5900	10980	12617	1378	1670	Covered	CAT C	950	5861000
436	436	11537	13214	17191	1462	1732	Not Provided	CAT A	840	6826000
445	445	6266	13767	12066	1443	1755	Not Provided	CAT C	880	4933000
446	446	5359	8372	11239	1247	1523	Open	CAT C	1350	5350000
455	455	6669	11500	11948	1249	1503	Open	CAT A	950	6197000
456	456	4846	9003	10135	1373	1635	No Parking	CAT A	840	6104000
465	465	9089	14320	13391	1307	1570	Not Provided	CAT B	1080	4171000
466	466	7753	8307	13508	1628	1959	Not Provided	CAT C	890	7148000
475	475	4783	11103	9602	1808	2154	Open	CAT C	990	3522000
476	476	5381	5177	8992	1501	1790	Not Provided	CAT B	450	5394000
485	485	7971	8165	12202	1845	2233	No Parking	CAT B	370	4319000
486	486	NA	NA	14676	1529	1820	No Parking	CAT B	800	6093000
495	495	9590	10173	13574	1611	1919	Covered	CAT A	960	7124000
496	496	11139	14537	17452	1187	1417	Open	CAT C	1230	5894000

[reached 'max' / getOption("max.print") -- omitted 86 rows]

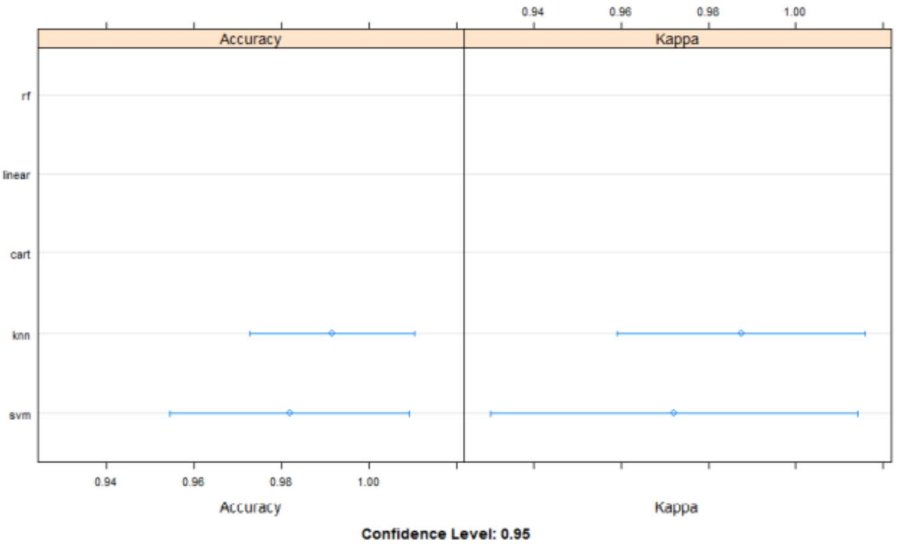
```
> |
```

```
Call:
summary.resamples(object = results)

Models: linear, svm, rf, cart, knn
Number of resamples: 10

Accuracy
             Min. 1st Qu. Median      Mean 3rd Qu. Max. NA's
linear 1.0000000       1      1 1.0000000       1    1    0
svm    0.9090909       1      1 0.9818182       1    1    0
rf     1.0000000       1      1 1.0000000       1    1    0
cart   1.0000000       1      1 1.0000000       1    1    0
knn    0.9166667       1      1 0.9916667       1    1    0

Kappa
             Min. 1st Qu. Median      Mean 3rd Qu. Max. NA's
linear 1.0000000       1      1 1.0000000       1    1    0
svm    0.8607595       1      1 0.9721519       1    1    0
rf     1.0000000       1      1 1.0000000       1    1    0
cart   1.0000000       1      1 1.0000000       1    1    0
knn    0.8750000       1      1 0.9875000       1    1    0
```



Confidence Level: 0.95

4. Perform Data Imputation using **MICE Package**

```
#Perform Data Imputation using MICE Package
  install.packages("mice")
  library(mice)
md.pattern(data)
```

```
 /\     /\
{  `---'  }
{  O   O  }
==>  V <==  No need for mice. This data set is completely observed.
 \  \|/  /
  `-----'
```

```
    Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm Weightgm Species Season
150  1             1            1             1            1        1       1      1 0
     0             0            0             0            0        0       0      0 0
```