MISSING VALUE TREATMENT OF DATA

1. Create a data frame

```
df=data.frame(Name=c("Bhuwanesh","Anil","Jai","Naveen"),
              Physics=c(98,87,91,94),
              Chemistry=c(NA,84,93,87),
              Mathematics=c(91,86,NA,NA))
print(df)
```

	Name	Physics	Chemistry	Mathematics
1	Bhuwanesh	98	NA	91
2	Anil	87	84	86
3	Jai	91	93	NA
4	Naveen	94	87	NA

2. Display the count of missing values in each column

```
df_miss=data.frame(num_missing=colSums(is.na(df)))
print(df_miss)
```

	num_missing
Name	0
Physics	0
Chemistry	1
Mathematics	2

3. Method-1- Replacing NA values with zero

```
df$Chemistry[is.na(df$Chemistry)]=0
df$Mathematics[is.na(df$Mathematics)]=0
print(df)
```

	Name	Physics	Chemistry	Mathematics
1	Bhuwanesh	98	0	91
2	Anil	87	84	86
3	Jai	91	93	0
4	Naveen	94	87	0

4. Method-2 - Imputing the missing values with mean

display mean values of chemistry and mathematics columns

```
print(mean(df$Chemistry,na.rm=T))
print(mean(df$Mathematics,na.rm=T))
```

```
[1] 88
[1] 88.5
```

5. Replacing NA values with mean

df$Chemistry[is.na(df$Chemistry)]=mean(df$Chemistry,na.rm=T)

df$Mathematics[is.na(df$Mathematics)]=mean(df$Mathematics,na.rm=T)

print(df)

```
       Name Physics Chemistry Mathematics
1 Bhuwanesh      98        88        91.0
2      Anil      87        84        86.0
3       Jai      91        93        88.5
4    Naveen      94        87        88.5
```

6. Method 3 - Imputing missing values with median

#Display median values of Chemistry and mathematics columns

print(median(df$Chemistry,na.rm=T))

print(median(df$Mathematics,na.rm=T))

```
> print(median(df$Chemistry,na.rm=T))
[1] 87.5
> print(median(df$Mathematics,na.rm=T))
[1] 88.5
```

7. Replacing NA values with median

df$Chemistry[is.na(df$Chemistry)]=median(df$Chemistry,na.rm=T)

df$Mathematics[is.na(df$Mathematics)]=median(df$Mathematics,na.rm=T)

print(df)

```
       Name Physics Chemistry Mathematics
1 Bhuwanesh      98        87        91.0
2      Anil      87        84        86.0
3       Jai      91        93        88.5
4    Naveen      94        87        88.5
```