**FINOLEX ACADEMY OF MANAGEMENT AND TECHNOLOGY, RATNAGIRI**

**DEPARTMENT OF MCA**

**PRACTICAL NO .06**

**INTRODUCTION TO R GRAPHICS AND DATA**

**PREPROCESSING**

QUE.1.Import employee.csv file and perform following -

1. Extract only following columns "Name", "Age", "Salary", "isLocal" into dataframe "employee\_subset"
2. Rename the following columns ""Name", "Age", "Designation", "Salary", "isLocal" from employee\_subset dataframe
3. Check if a value is missing in employee\_subset
4. Calculate the mean of Age and Salary column in employee\_subset
5. Replace missing values by mean of that variable/column

ANS.

1.Display the content.

* + employee<-read.csv("employee.csv")
  + employee

id Name Age Designation Salary isLocal

1. 1 Michelle 44 Manager 72000 NA
2. 2 Ryan 27 Clerk 48000 NA
3. 3 Gary 30 Clerk 54000 NA
4. 4 Guru 38 Engineer 61000 NA
5. 5 Harsh 40 Clerk NA NA
6. 6 Brad 35 Engineer 58000 NA
7. 7 James NA Clerk 52000 NA
8. 8 Tina 48 Senior\_manager 79000 NA
9. 9 Mina 50 CEO 83000 NA
10. 10 Tara 37 Engineer 67000 NA
11. Extract only following columns "Name", "Age", "Salary", "isLocal" into dataframe "employee\_subset"

> employee\_subset=select(employee,'Name','Age','Salary','isLocal')

> employee\_subset

Name Age Salary isLocal

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | Michelle | 44 | 72000 | NA |
| 2 | Ryan | 27 | 48000 | NA |
| 3 | Gary | 30 | 54000 | NA |
| 4 | Guru | 38 | 61000 | NA |
| 5 | Harsh | 40 | NA | NA |
| 6 | Brad | 35 | 58000 | NA |
| 7 | James | NA | 52000 | NA |
| 8 | Tina | 48 | 79000 | NA |
| 9 | Mina | 50 | 83000 | NA |
| 10 | Tara | 37 | 67000 | NA |

1. Rename the following columns ""Name", "Age", "Designation", "Salary", "isLocal" from employee\_subset dataframe

>employee\_subset=select(employee,'name'='Name','age'='Age','design ation'='Designation','salary'='Salary','islocal'='isLocal')

>employee\_subset

name age designation salary islocal

1. Michelle 44 Manager 72000 NA
2. Ryan 27 Clerk 48000 NA
3. Gary 30 Clerk 54000 NA
4. Guru 38 Engineer 61000 NA
5. Harsh 40 Clerk NA NA
6. Brad 35 Engineer 58000 NA
7. James NA Clerk 52000 NA
8. Tina 48 Senior\_manager 79000 NA
9. Mina 50 CEO 83000 NA
10. Tara 37 Engineer 67000 NA
11. Check if a value is missing in employee\_subset

>is.na(employee\_subset)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | name | age | designation | salary | islocal |
| [1,] | FALSE | FALSE | FALSE | FALSE | TRUE |
| [2,] | FALSE | FALSE | FALSE | FALSE | TRUE |
| [3,] | FALSE | FALSE | FALSE | FALSE | TRUE |
| [4,] | FALSE | FALSE | FALSE | FALSE | TRUE |
| [5,] | FALSE | FALSE | FALSE | TRUE | TRUE |
| [6,] | FALSE | FALSE | FALSE | FALSE | TRUE |
| [7,] | FALSE | TRUE | FALSE | FALSE | TRUE |
| [8,] | FALSE | FALSE | FALSE | FALSE | TRUE |
| [9,] | FALSE | FALSE | FALSE | FALSE | TRUE |
| [10,] | FALSE | FALSE | FALSE | FALSE | TRUE |

1. Calculate the mean of Age and Salary column in employee\_subset
   * + mean\_age<-mean(employee\_subset$age,na.rm=TRUE)
     + mean\_age [1] 38.77778

mean\_salary<-mean(employee\_subset$salary,na.rm=TRUE)

* + - mean\_salary [1] 63777.78

1. Replace missing values by mean of that variable/column

>employee\_subset$age<-ifelse(is.na(employee\_subset$age),ave(empl oyee\_subset$age, FUN = function(x) mean(x, na.rm = TRUE)), employe e\_subset$age)

>employee\_subset

name age designation salary islocal

1 Michelle 44.00000 Manager 72000 NA

2 Ryan 27.00000 Clerk 48000 NA

3 Gary 30.00000 Clerk 54000 NA

1. Guru 38.00000 Engineer 61000 NA
2. Harsh 40.00000 Clerk NA NA
3. Brad 35.00000 Engineer 58000 NA
4. James 38.77778 Clerk 52000 NA

8 Tina 48.00000 Senior\_manager 79000 NA

9 Mina 50.00000 CEO 83000 NA

10 Tara 37.00000 Engineer 67000 NA

>employee\_subset$salary<-ifelse(is.na(employee\_subset$salary),ave(employee\_subset$salary, FUN = function(x) mean(x, na.rm = TRUE)), employee\_subset$salary)

>employee\_subset

name age designation salary islocal

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 | Michelle | 44.00000 | Manager | 72000.00 | NA |
| 2 | Ryan | 27.00000 | Clerk | 48000.00 | NA |
| 3 | Gary | 30.00000 | Clerk | 54000.00 | NA |
| 4 | Guru | 38.00000 | Engineer | 61000.00 | NA |
| 5 | Harsh | 40.00000 | Clerk | 63777.78 | NA |
| 6 | Brad | 35.00000 | Engineer | 58000.00 | NA |
| 7 | James | 38.77778 | Clerk | 52000.00 | NA |
| 8 | Tina | 48.00000 | Senior\_manager | 79000.00 | NA |
| 9 | Mina | 50.00000 | CEO | 83000.00 | NA |
| 10 | Tara | 37.00000 | Engineer | 67000.00 | NA |