



Experiment 1.2

Student Name: Jatin

UID: 20BCS5951

Branch: BE-CSE

Section/Group:605/B

Semester:6th

Date of Performance:21/02/2023

Subject Name: Data Mining Lab

Subject Code:20CSP-376

1. Aim:

⇒ To perform the statistical analysis of data.

2. Objective:

⇒ To represent the creation of file using R Studio and displaying the pattern on Weka Tool for further extraction and analysis of knowledge.

3. Code And Output:

Program

```
#Library library library("RWeka")
```

```
setwd("D:\\Amar Doc\\6th sem\\DM lab") getwd()
```

```
#Creation Of Data Frame
```

```
rating<-1:4
```

```
animal<-c('Dog','Lion','Hen','Panda') country<-
```

```
c('India','Australia','USA','Singapore') avg_sleep_hours<-c(4,5,6,7) name<-
```

```
c("Amarjeet","Manas","Kushagra")
```

```
#make sure to set dtring as factors to false.
```

```
#son that string values are stored as characters and not vectors.
```

```
super_sleepers <- data.frame(rating,animal,country,avg_sleep_hours,stringsAsFactors = FALSE)
```

```
print(super_sleepers) print(class(super_sleepers))
```

```
print(str(super_sleepers))
```

```
write.arff(super_sleepers, file="super_sleepers.arff")
```



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

```
N = read.arff("super_sleepers.arff")
```



Discover. Learn. Empower.

```
#print data print(N)
cat("\n\n\n")
```

```
#printing first two rows from print(head(N,2))
```

```
dim(N)
```

```
names(N) #show all
the      animals
N["animal"]
```

```
#show average sleep hours
N["avg_sleep_hours"] #show max of
average sleep hours
max(N["avg_sleep_hours"])
```

```
#show min of average sleep hours
min(avg_sleep_hours)
```

```
#sum of average sleep hours
sum(avg_sleep_hours)
```

```
#mean of average sleep hours
mean(avg_sleep_hours)
```

```
#median of average sleep hours median(sort(avg_sleep_hours))
```

OUTPUT=



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.



Discover. Learn. Empower.

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

```
R 4.2.2 - D:/Amar Doc/6th sem/DM lab/ <R>
> N = read.arff("super_sleepers.arff")
> #print data
> print(N)
  rating animal  country avg_sleep_hours
1      1   Dog    India          4
2      2  Lion Australia          5
3      3   Hen     USA           6
4      4  Panda Singapore          7
> cat("\n\n")
> #printing first two rows from
> print(head(N,2))
  rating animal  country avg_sleep_hours
1      1   Dog    India          4
2      2  Lion Australia          5
>
> dim(N)
[1] 4 4
>
> names(N)
[1] "rating"      "animal"      "country"     "avg_sleep_hours"
> #show all the animals
> N["animal"]
  animal
1   Dog
2  Lion
3   Hen
4  Panda
>
> #show average sleep hours
> N["avg_sleep_hours"]
  avg_sleep_hours
1           4
2           5
3           6
4           7
>
> #show max of average sleep hours
> max(N["avg_sleep_hours"])
[1] 7
>
> #show min of average sleep hours
> min(avg_sleep_hours)
[1] 4
>
> #sum of average sleep hours
> sum(avg_sleep_hours)
[1] 22
>
> #mean of average sleep hours
> mean(avg_sleep_hours)
[1] 5.5
>
> #median of average sleep hours
> median(sort(avg_sleep_hours))
[1] 5.5
>
```

Environment History Connections

R • Global Environment

Data

N	4 obs. of 4 varia...
super_s...	4 obs. of 4 varia...

Values

animal	chr [1:4]	"Dog" "Li...
avg_sle...	num [1:4]	4 5 6 7
country	chr [1:4]	"India" "..."
name	chr [1:3]	"Amarjeet..."
rating	int [1:4]	1 2 3 4
UID	num [1:4]	9381 5853...

Files Plots Packages Help View

Home

Name	Size
.RData	2.6
.Rhistory	1 K
Adobe	
Amar.jpg	29..
Custom Office Templates	
desktop.ini	41B
Sound Recordings	
vikas scholar.pdf	192

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

```
R 4.2.2 - D:/Amar Doc/6th sem/DM lab/ <R>
> dim(N)
[1] 4 4
>
> names(N)
[1] "rating"      "animal"      "country"     "avg_sleep_hours"
> #show all the animals
> N["animal"]
  animal
1   Dog
2  Lion
3   Hen
4  Panda
>
> #show average sleep hours
> N["avg_sleep_hours"]
  avg_sleep_hours
1           4
2           5
3           6
4           7
>
> #show max of average sleep hours
> max(N["avg_sleep_hours"])
[1] 7
>
> #show min of average sleep hours
> min(avg_sleep_hours)
[1] 4
>
> #sum of average sleep hours
> sum(avg_sleep_hours)
[1] 22
>
> #mean of average sleep hours
> mean(avg_sleep_hours)
[1] 5.5
>
> #median of average sleep hours
> median(sort(avg_sleep_hours))
[1] 5.5
>
```

Environment History Connections

R • Global Environment

Data

N	4 obs. of 4 varia...
super_s...	4 obs. of 4 varia...

Values

animal	chr [1:4]	"Dog" "Li...
avg_sle...	num [1:4]	4 5 6 7
country	chr [1:4]	"India" "..."
name	chr [1:3]	"Amarjeet..."
rating	int [1:4]	1 2 3 4
UID	num [1:4]	9381 5853...

Files Plots Packages Help View

Home

Name	Size
.RData	2.6
.Rhistory	1 K
Adobe	
Amar.jpg	29..
Custom Office Templates	
desktop.ini	41B
Sound Recordings	
vikas scholar.pdf	192

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

```
R 4.2.2 - D:/Amar Doc/6th sem/DM lab/ <R>
> dim(N)
[1] 4 4
>
> names(N)
[1] "rating"      "animal"      "country"     "avg_sleep_hours"
> #show all the animals
> N["animal"]
  animal
1   Dog
2  Lion
3   Hen
4  Panda
>
> #show average sleep hours
> N["avg_sleep_hours"]
  avg_sleep_hours
1           4
2           5
3           6
4           7
>
> #show max of average sleep hours
> max(N["avg_sleep_hours"])
[1] 7
>
> #show min of average sleep hours
> min(avg_sleep_hours)
[1] 4
>
> #sum of average sleep hours
> sum(avg_sleep_hours)
[1] 22
>
> #mean of average sleep hours
> mean(avg_sleep_hours)
[1] 5.5
>
> #median of average sleep hours
> median(sort(avg_sleep_hours))
[1] 5.5
>
```

Environment History Connections

R • Global Environment

Data

N	4 obs. of 4 varia...
super_s...	4 obs. of 4 varia...

Values

animal	chr [1:4]	"Dog" "Li...
avg_sle...	num [1:4]	4 5 6 7
country	chr [1:4]	"India" "..."
name	chr [1:3]	"Amarjeet..."
rating	int [1:4]	1 2 3 4
UID	num [1:4]	9381 5853...

Files Plots Packages Help View

Home

Name	Size
.RData	2.6
.Rhistory	1 K
Adobe	
Amar.jpg	29..
Custom Office Templates	
desktop.ini	41B
Sound Recordings	
vikas scholar.pdf	192



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

Discover. Learn. Empower.

Weka Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Generate...

Undo

Edit...

Save...

Filter

Choose

None

Apply

Stop

Current relation

Relation: R_data_frame

Attributes: 4

Instances: 4

Sum of weights: 4

Attributes

All

None

Invert

Pattern

No.	Name
1	<input checked="" type="checkbox"/> rating
2	<input type="checkbox"/> animal
3	<input type="checkbox"/> country
4	<input type="checkbox"/> avg_sleep_hours

Remove

Selected attribute

Name: rating

Missing: 0 (0%)

Distinct: 4

Type: Numeric

Unique: 4 (100%)

Statistic	Value
Minimum	1
Maximum	4
Mean	2.5
StdDev	1.291

Class: avg_sleep_hours (Num)

Visualize All

4

1

2.5

4

Status

OK

Log

x 0