

Experiment 1.2

Student Name: Anshuman Singh

UID: 20BCS2665

Branch: CSE

Section/Group: 20BCS_DM-902/A

Semester: 6th

Date of Performance: 20/02/2023

Subject Name: Data Mining Lab

Subject Code: 20CSP-376

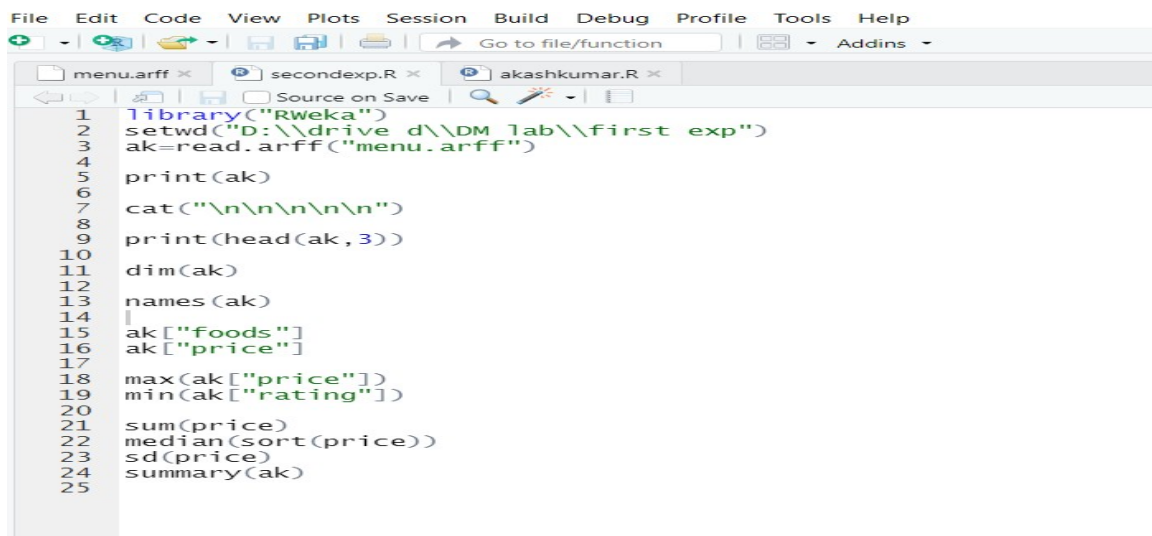
Aim: To perform the statistical analysis of data.

Objective: To analyse the .arff file we created in first experiment.

Steps:

- Install and import the RWeka library.
- Set the working directory.
- Read the arff file and store it.
- Now we can perform different operation on that.

Code:



```
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function | Addins
menu.arff x secondexp.R x akashkumar.R x
Source on Save
1 library("Rweka")
2 setwd("D:\\drive d\\DM lab\\first exp")
3 ak=read.arff("menu.arff")
4
5 print(ak)
6
7 cat("\n\n\n\n\n")
8
9 print(head(ak,3))
10
11 dim(ak)
12
13 names(ak)
14 |
15 ak["foods"]
16 ak["price"]
17
18 max(ak["price"])
19 min(ak["rating"])
20
21 sum(price)
22 median(sort(price))
23 sd(price)
24 summary(ak)
25
```

Code and its output:

`ak=read.arff("menu.arff")`-Reads the arff file.

`print(ak)`- Prints the table's data

`cat("\n\n\n\n\n")`- here are printing 5 blank lines

```
D:/drive d/DM lab/first exp/ ↵
> library("Rweka")
> setwd("D:\\drive d\\DM lab\\first exp")
> ak=read.arff("menu.arff")
>
> print(ak)
  srno   foods price rating
1     1 Frech Fries   50     4
2     2   Burgers   40     5
3     3    Pizza  120     5
4     4   Drinks   40     3
>
> cat("\n\n\n\n\n")
>
```

`print(head(ak,3))`- prints first three rows. `dim(ak)`-it prints the dimension of the table. `names(ak)`- it prints the column names.

```
> print(head(ak,3))
  srno   foods price rating
1     1 Frech Fries   50     4
2     2   Burgers   40     5
3     3    Pizza  120     5
>
> dim(ak)
[1] 4 4
>
> names(ak)
[1] "srno" "foods" "price" "rating"
```

`ak["foods"]` – It prints the details of column foods. `ak["price"]` - It prints the details of column price.

`max(ak["price"])`- It prints the maximum of all price.

`min(ak["rating"])` – It prints the minimum of all the

rating `sum(price)` – This prints the sum of all prices.

`median(sort(price))` – Prints the median of prices data.

`sd(price)` – Standard deviation. `summary(ak)`- Summary

of the data of the table.

```
> ak["foods"]
      foods
1 Frech Fries
2   Burgers
3   Pizza
4   Drinks
> ak["price"]
      price
1      50
2      40
3     120
4      40
>
> max(ak["price"])
[1] 120
> min(ak["rating"])
[1] 3
>
> sum(price)
[1] 250
> median(sort(price))
[1] 45
> sd(price)
[1] 38.6221
> summary(ak)
      srno      foods      price      rating
Min.   :1.00 Length:4   Min.   : 40.0   Min.   :3.00
1st Qu.:1.75 Class :character 1st Qu.: 40.0   1st Qu.:3.75
Median :2.50 Mode  :character Median : 45.0   Median :4.50
Mean   :2.50          Mean   : 62.5   Mean   :4.25
3rd Qu.:3.25          3rd Qu.: 67.5   3rd Qu.:5.00
Max.   :4.00          Max.   :120.0   Max.   :5.00
> |
```



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.