DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
Discover. Learn. Empower.

CHANDIGARH UNIVERSITY
Discover. Learn. Empower.

# UNIVERSITY INSTITUTE OF ENGINEERING

## Department of Computer Science & Engineering

**Subject Name:** Data Mining LAB

**Subject Code:** 20CSP-376

**Submitted to:**

Er. Navneet Kaur

**Submitted by:**

Name: Aayush Gurung

UID: 20BCS5323

Section: DM_607

Group: A

## INDEX

| Ex. No | List of Experiments | Conduct (MM: 12) | Viva (MM: 10) | Record (MM: 8) | Total (MM: 30) | Remarks/Signature |
|---|---|---|---|---|---|---|
| 1.1 | | | | | | |
| 1.2 | | | | | | |
| 1.3 | | | | | | |
| 2.1 | | | | | | |
| 2.2 | | | | | | |
| 2.3 | | | | | | |
| 2.4 | | | | | | |
| 3.1 | | | | | | |
| 3.2 | | | | | | |
| 3.3 | | | | | | |

# Experiment1.1

**Student Name: Aayush Gurung** **UID:20BCS5323**
**Branch: CSE** **Section/Group:DM_607(A)**
**Semester: 6** **Date of Performance:13-02-2023**
**Subject Name: DM LAB** **Subject Code: 20CST-376**

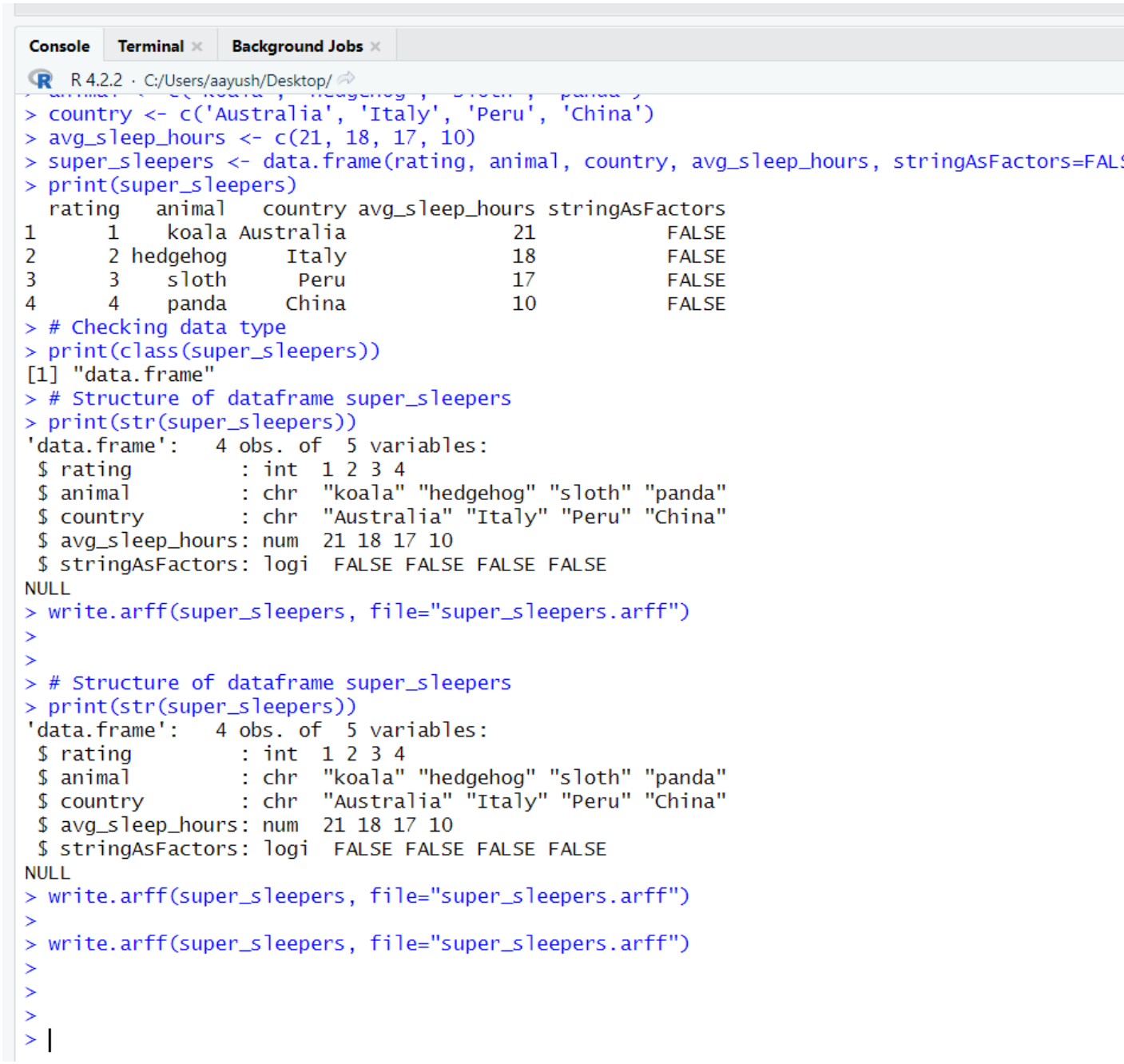
## 1. Aim:

Demonstration of preprocessing on .arff file using student data .arff.

## 2. Code:

```
# Loading library
library(RWeka)

# Setting Working Directory
setwd("C\Users\aayush\Desktop\Rstudio")

# Checking Working Directory
getwd()

# Creating Data frame
rating <- 1:4
animal <- c('koala', 'hedgehog', 'sloth', 'panda')
country <- c('Australia', 'Italy', 'Peru', 'China')
avg_sleep_hours <- c(21, 18, 17, 10)

# Make sure to set stringAsFactors to FALSE
#so that string values are stored as characters and not vectors

super_sleepers <- data.frame(rating, animal, country, avg_sleep_hours, stringAsFactors=FALSE)
print(super_sleepers)

# Checking data type
print(class(super_sleepers))

# Structure of dataframe super_sleepers
print(str(super_sleepers))

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

## 3. Output:


Console
Terminal
Background Jobs
R 4.2.2 · C:/Users/aayush/Desktop/
> country <- c('Australia', 'Italy', 'Peru', 'China')
> avg_sleep_hours <- c(21, 18, 17, 10)
> super_sleepers <- data.frame(rating, animal, country, avg_sleep_hours, stringAsFactors=FAL
> print(super_sleepers)
rating animal country avg_sleep_hours stringAsFactors
1 1 koala Australia 21 FALSE
2 2 hedgehog Italy 18 FALSE
3 3 sloth Peru 17 FALSE
4 4 panda China 10 FALSE
> # Checking data type
> print(class(super_sleepers))
[1] "data.frame"
> # Structure of dataframe super_sleepers
> print(str(super_sleepers))
'data.frame': 4 obs. of 5 variables:
$ rating : int 1 2 3 4
$ animal : chr "koala" "hedgehog" "sloth" "panda"
$ country : chr "Australia" "Italy" "Peru" "China"
$ avg_sleep_hours: num 21 18 17 10
$ stringAsFactors: logi FALSE FALSE FALSE FALSE
NULL
> write.arff(super_sleepers, file="super_sleepers.arff")
>
>
> # Structure of dataframe super_sleepers
> print(str(super_sleepers))
'data.frame': 4 obs. of 5 variables:
$ rating : int 1 2 3 4
$ animal : chr "koala" "hedgehog" "sloth" "panda"
$ country : chr "Australia" "Italy" "Peru" "China"
$ avg_sleep_hours: num 21 18 17 10
$ stringAsFactors: logi FALSE FALSE FALSE FALSE
NULL
> write.arff(super_sleepers, file="super_sleepers.arff")
>
> write.arff(super_sleepers, file="super_sleepers.arff")
>
>
>
>