

**UNIVERSITY INSTITUTE OF ENGINEERING**

**Department of Computer Science & Engineering**

# Subject Name: DATA MINING LAB

**Subject Code:** 20CSP-376

**Submitted to: Submitted by:**

Faculty name: Er. Himanshi Name: Sahil Kaundal

UID: 21BCS8197

Section: 616

Group: A

**INDEX**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Ex. No** | **List of Experiments** | **Conduct (MM: 12)** | **Viva**  **(MM: 10)** | **Record (MM: 8)** | **Total**  **(MM: 30)** | **Remarks/Signature** |
| 1 | Demonstration of preprocessing on .arff file using student data. arff |  |  |  |  |  |
| 2 | To perform the statistical analysis of data. |  |  |  |  |  |
| 3 | Demonstration of association rule mining using Apriory algorithm on supermarket data. |  |  |  |  |  |
| 4 | Demonstration of FP Growth algorithm on supermarket data. |  |  |  |  |  |
| 5 | To perform the classification by decision tree induction using WEKA tools. |  |  |  |  |  |
| 6 | To perform classification using Bayesian classification algorithm using R. |  |  |  |  |  |
| 7 | To perform the cluster analysis by k-means method using R. |  |  |  |  |  |
| 8 | To perform the hierarchical clustering using R programming. |  |  |  |  |  |
| 9 | Study of Regression Analysis using R programming. |  |  |  |  |  |
| 10 | Outlier detection using R programming. |  |  |  |  |  |

**Experiment 2**

**Student Name:** Sahil Kaundal **UID:** 21BCS8197

**Branch:** CSE (Lateral Entry)  **Section/Group:** 616/A

**Semester:** 6th **Date of Performance:** 02/03/2023

**Subject Name:** Data Mining Lab **Subject Code:** 20CSP-376

1. **Aim:**

To perform the statistical analysis of data.

1. **Apparatus / Simulation Used:**

* Windows 7 or above
* R Studio

1. **Objective:**

* Represent the reading of file using R Studio.
* Displaying the pattern on Weka Tool.
* Find mean, median and standard deviation of particular columns.

1. **Script and Output:**

library(RWeka)

N = read.arff("super\_sleepers.arff")

print(N)

cat("\n\n\n")

print(head(N,2))

dim(N)

names(N)

N["Animal"]

N["Avg.sleep\_hours"]

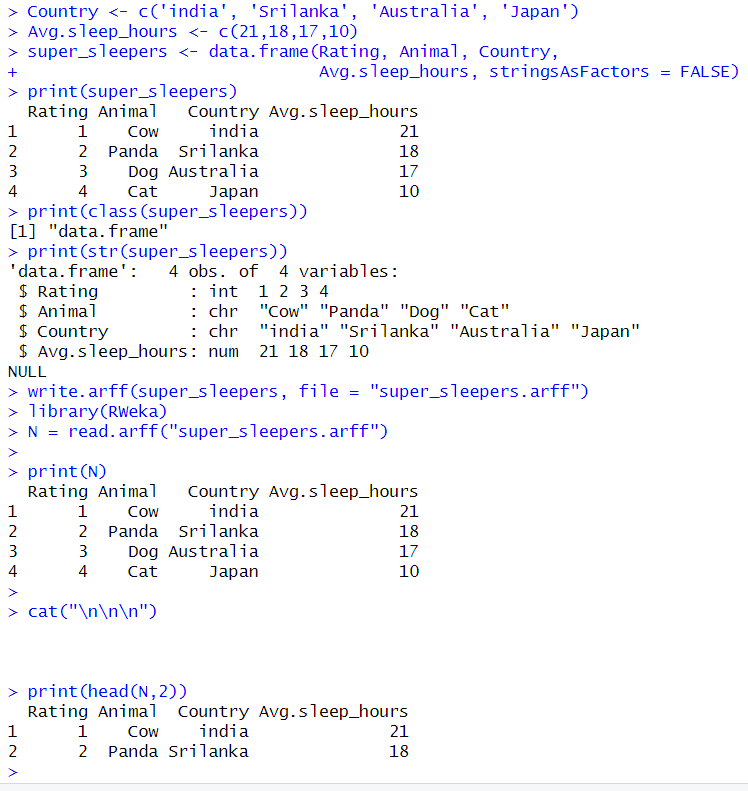
min(Avg.sleep\_hours)

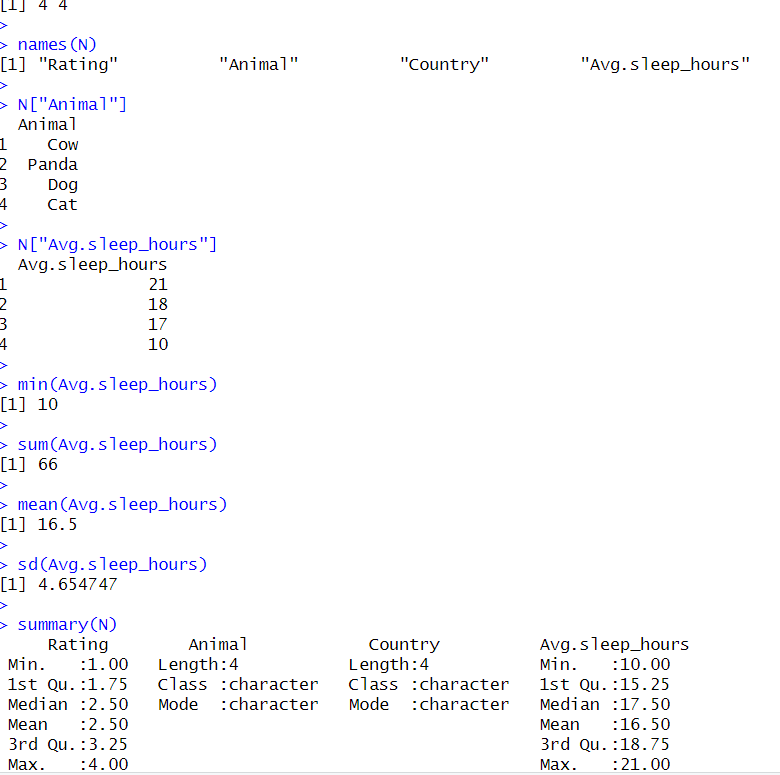
sum(Avg.sleep\_hours)

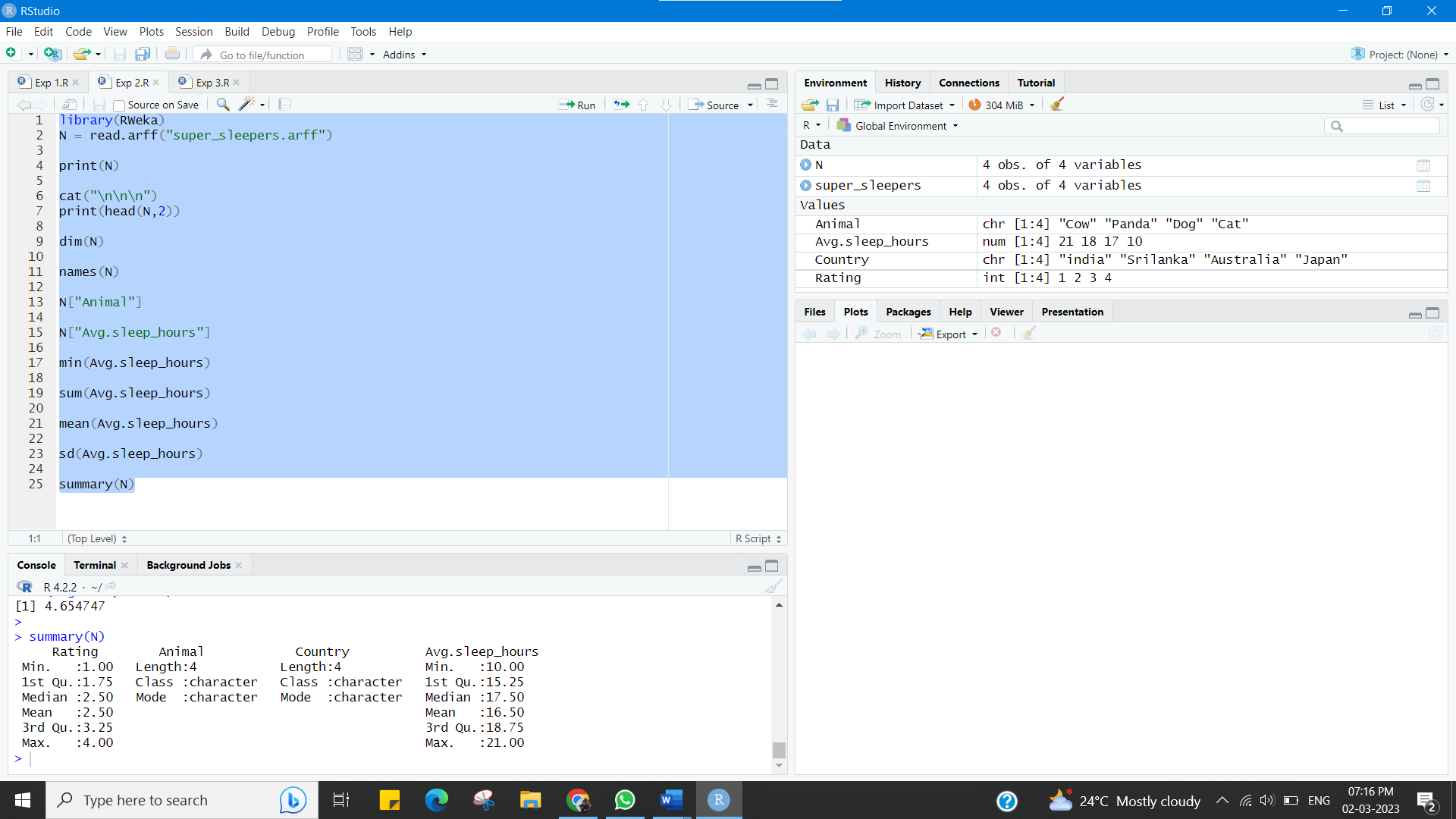
mean(Avg.sleep\_hours)

sd(Avg.sleep\_hours)

summary(N)







**Learning outcomes (What I have learnt):**

* Represent the reading of file using R Studio.
* Displaying the pattern on Weka Tool.
* Find mean, median and standard deviation of particular columns.

**Evaluation Grid (To be created as per the SOP and Assessment guidelines by the faculty):**

|  |  |  |  |
| --- | --- | --- | --- |
| Sr. No. | Parameters | Marks Obtained | Maximum Marks |
| 1. |  |  |  |