**HOME TASK**

**EXPERIMENT – 8**

**GROUP NUMBER: 3**

|  |  |  |
| --- | --- | --- |
| **S. No** | **Student Name** | **UID** |
| **1** | Sahul Kumar Parida | **20BCS4919** |

**Aim:**

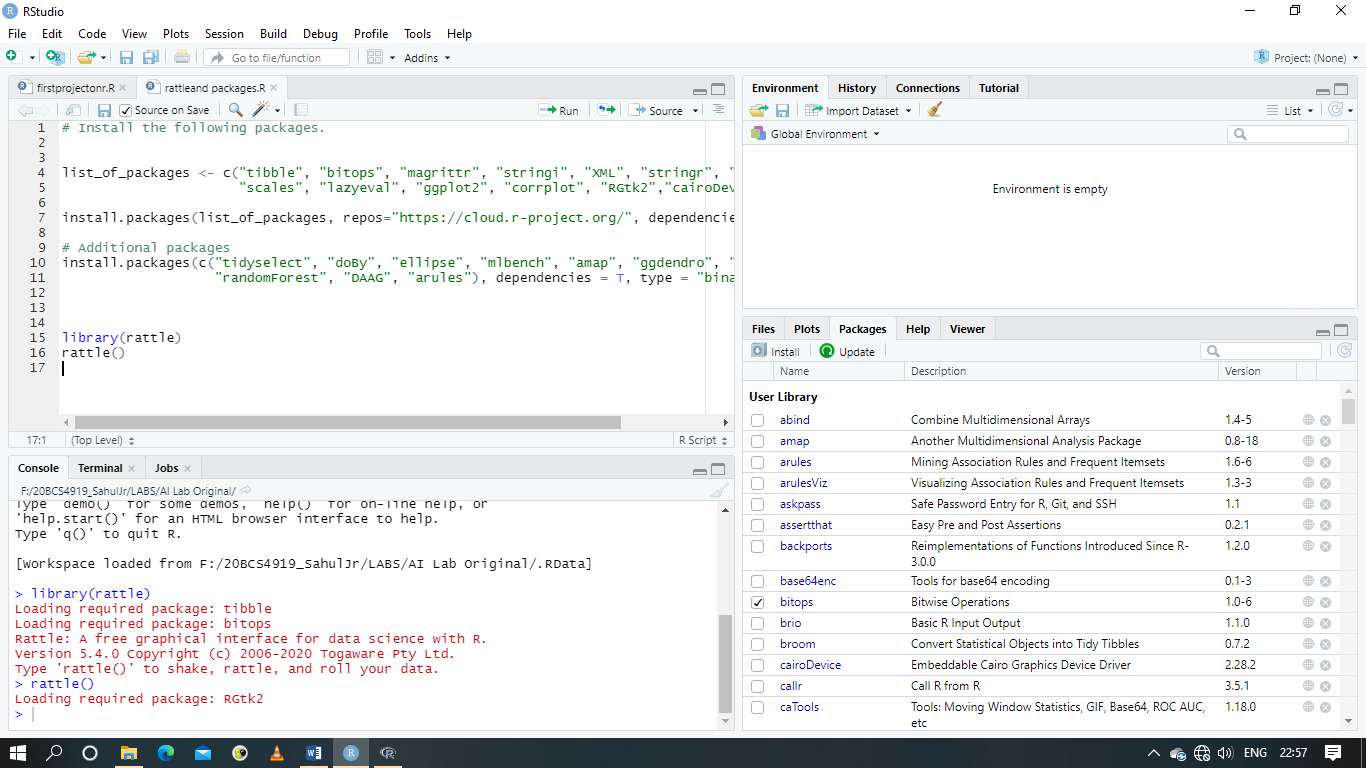
Explore, visualise, transform and summarise input datasets for building classification models.

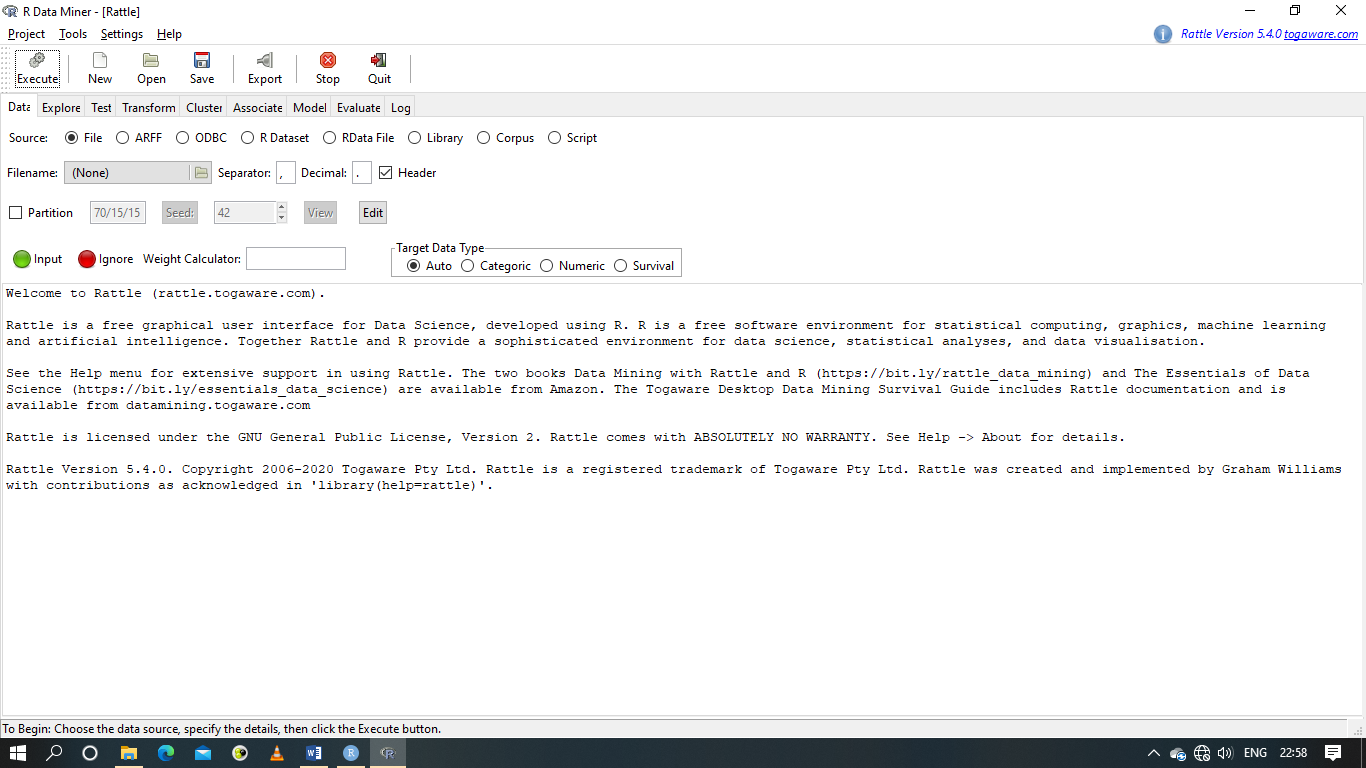
**Requirements:**

PC with internet connectivity, Rattle, R for Windows (Version 4.0.0), RStudio for Windows (Version 1.2.5042) and R Data Miner – [Rattle].

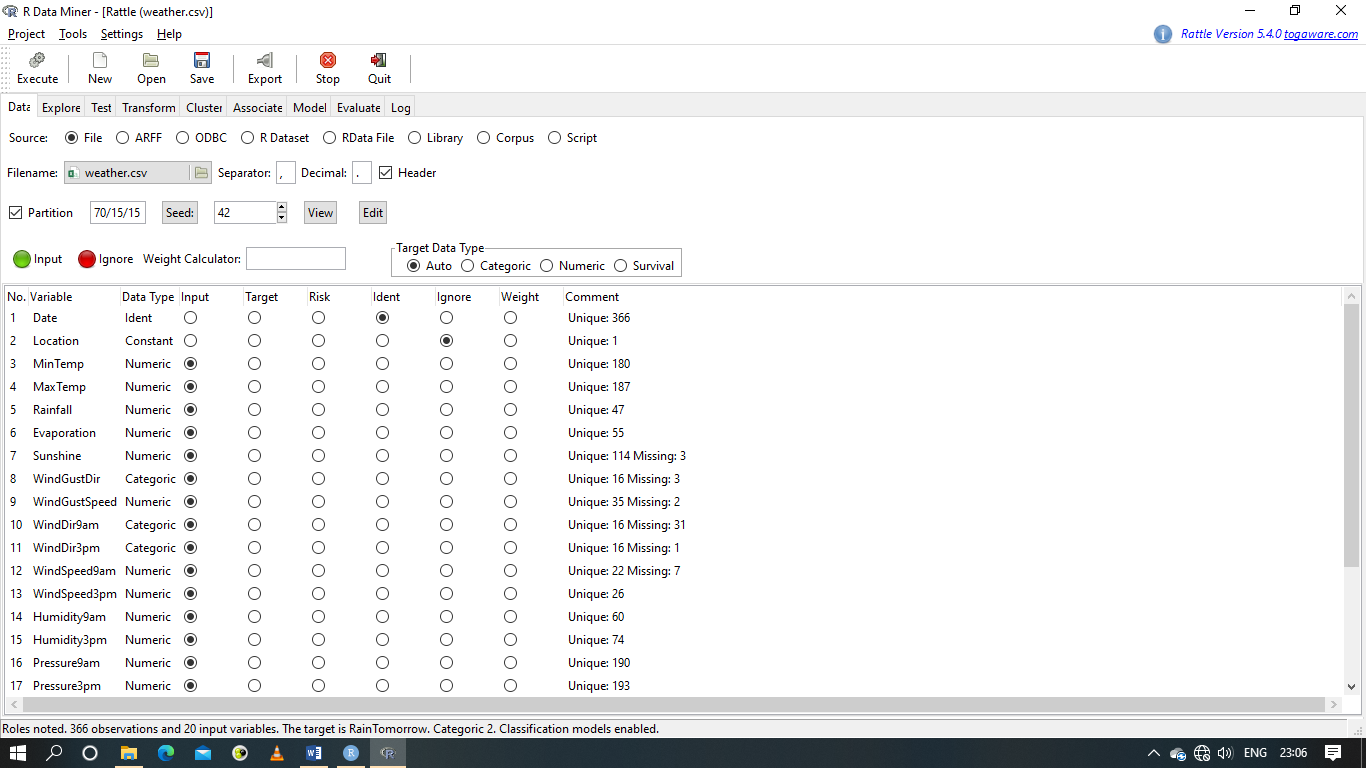
**Command**:

1. First we have to open a file named rattleand packages using RStudio Application.
2. Then we have to install the packages "tibble", "bitops", "magrittr”, “stringi”, “XML", "stringr”, “Hmisc", "R6”, “scales”, “lazyeval”, “ggplot2”, “corrplot", "RGtk2", "cairoDevice", "rattle", tidyselect", "doBy", "ellipse", "mlbench", "amap", "ggdendro", "fpc", "randomForest", "DAAG", "arules" in order to get the R Data Miner – [Rattle] so as to get the desired output.

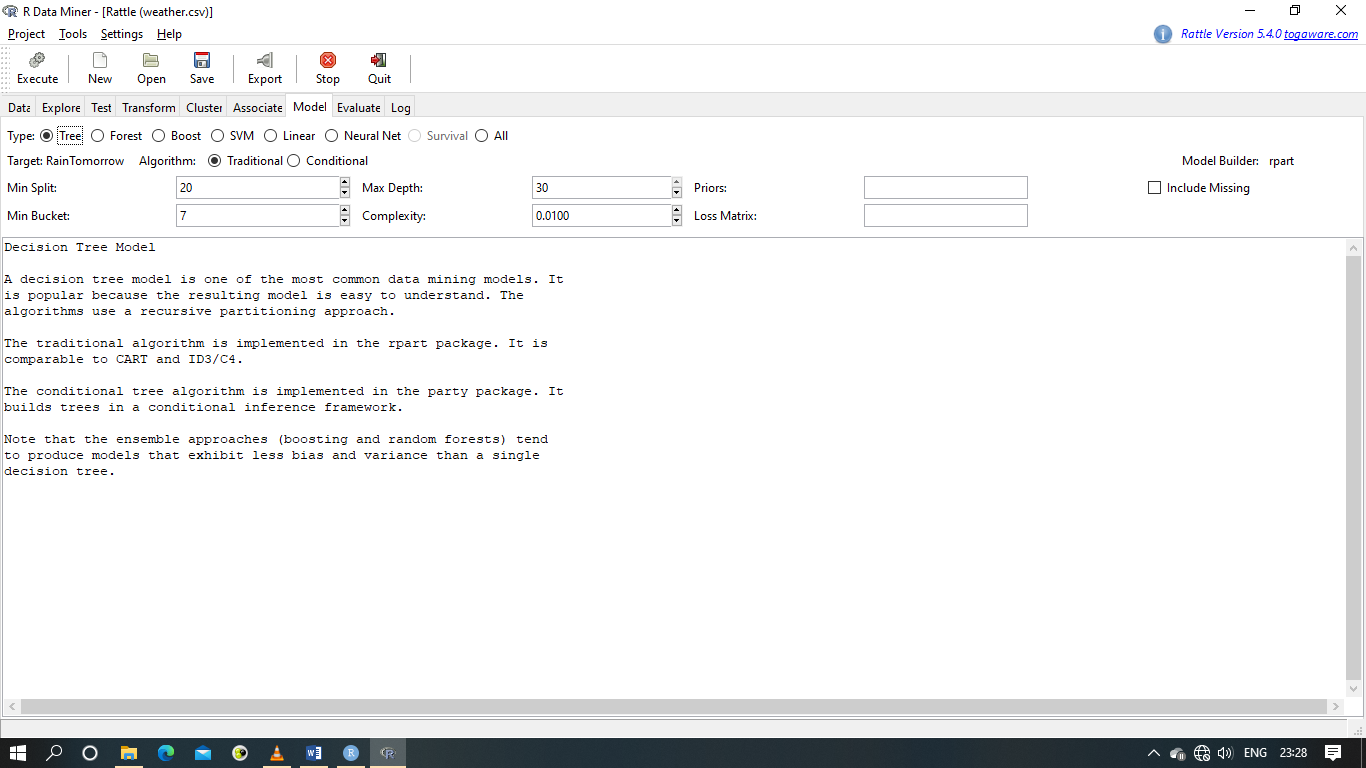




1. We have to open the file named weatherAUS.csv from R Data Miner – [Rattle].

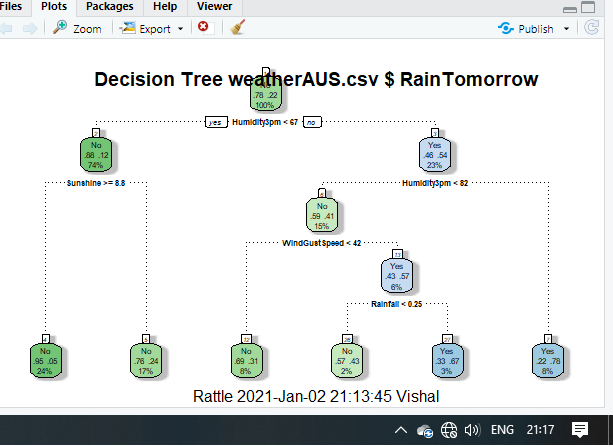


1. In the Data tab, we have to Ignore the four values namely Date, Location, Rain Today and Risk\_MM and also we have to set the Target to Rain Tomorrow. Next we have to execute the file.
2. Next we have to go the Model tab, set the Type to Tree and execute the file.

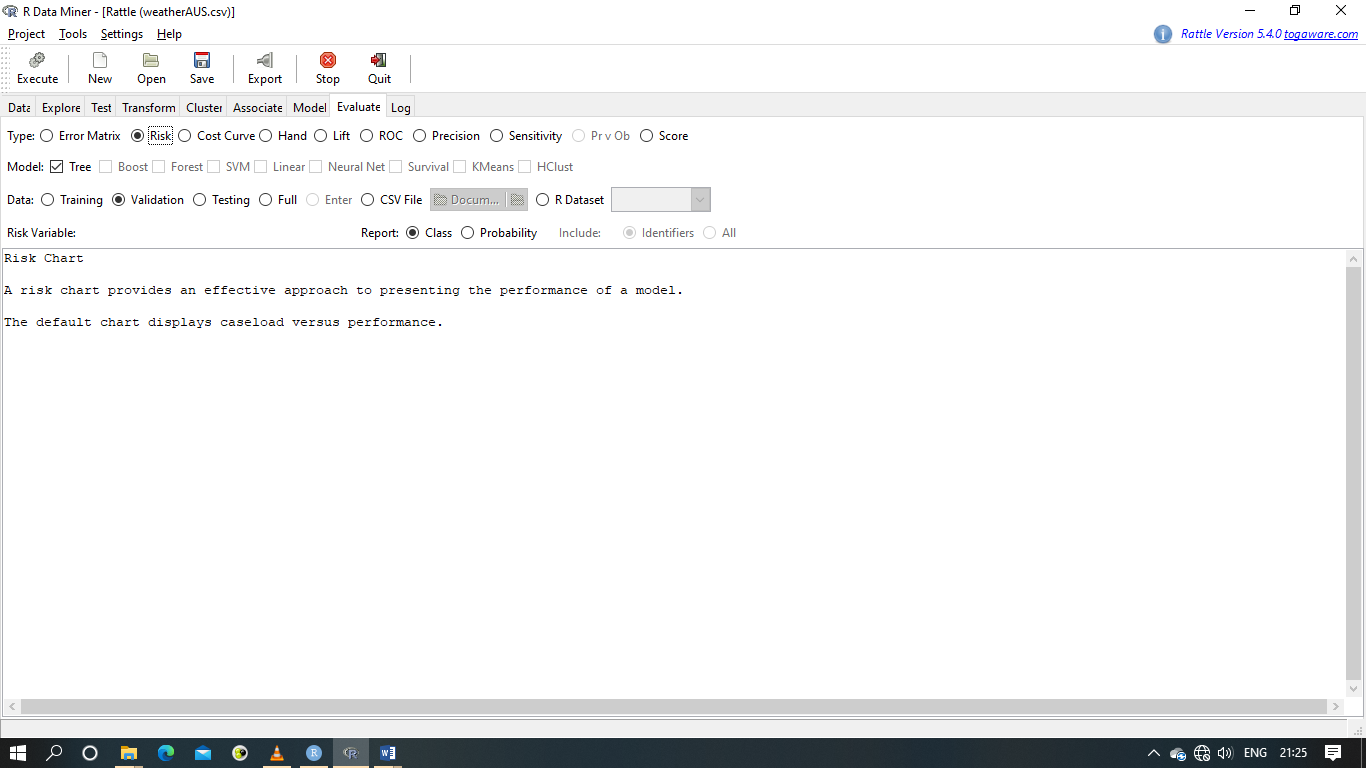


**Expected output:**

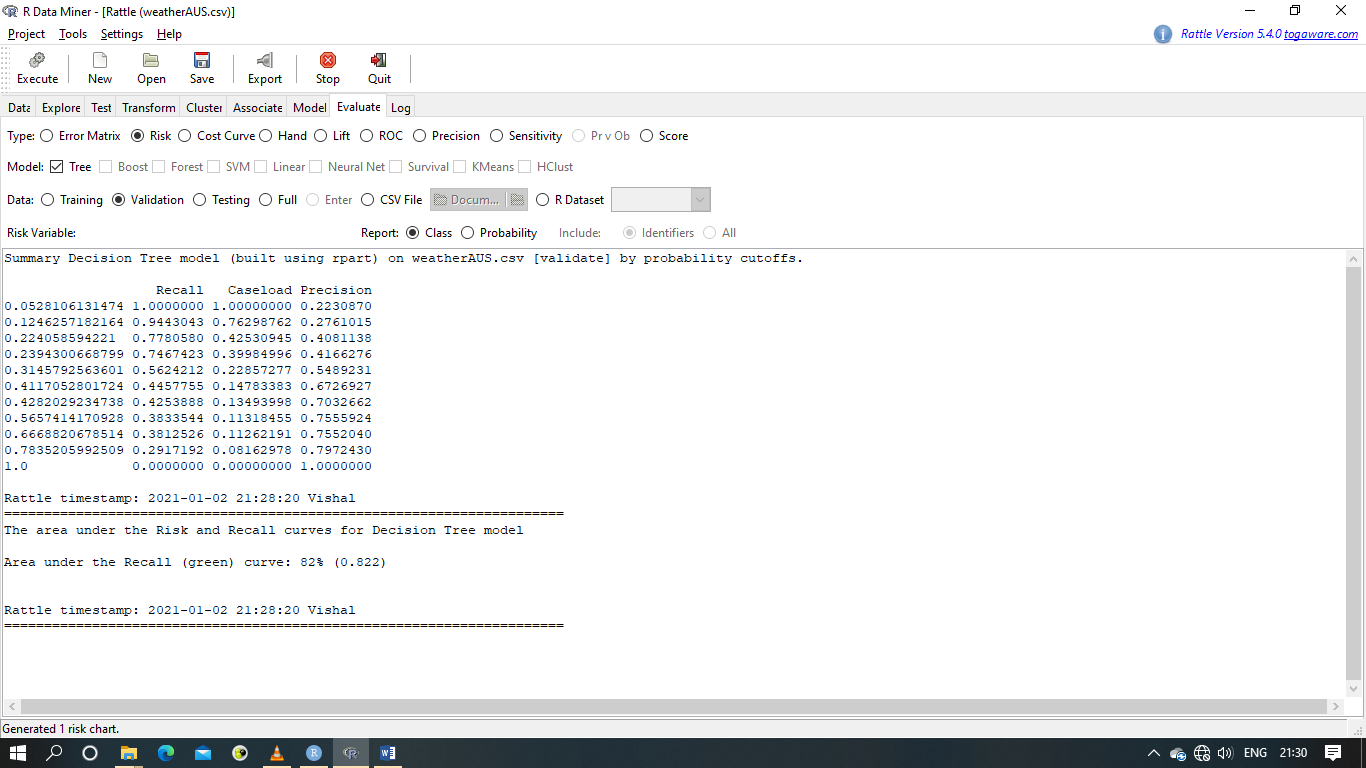
There are two things on the right side of the Model tab namely Rules and Draw. We have to click on Draw on the right side under the Model Tab in order to get the flowchart in the Plotting Region of RStudio Application.



1. We have to go to the Evaluate tab. Set the Type to Risk and Data to Validation.



Then, we have to click on execute in order to get the final output.



Summary Decision Tree model (built using rpart) on weatherAUS.csv [validate] by probability cutoffs.

Recall Caseload Precision

0.0528106131474 1.0000000 1.00000000 0.2230870

0.1246257182164 0.9443043 0.76298762 0.2761015

0.224058594221 0.7780580 0.42530945 0.4081138

0.2394300668799 0.7467423 0.39984996 0.4166276

0.3145792563601 0.5624212 0.22857277 0.5489231

0.4117052801724 0.4457755 0.14783383 0.6726927

0.4282029234738 0.4253888 0.13493998 0.7032662

0.5657414170928 0.3833544 0.11318455 0.7555924

0.6668820678514 0.3812526 0.11262191 0.7552040

0.7835205992509 0.2917192 0.08162978 0.7972430

1.0 0.0000000 0.00000000 1.0000000

Rattle timestamp: 2021-01-02 21:28:20 Vishal

======================================================================

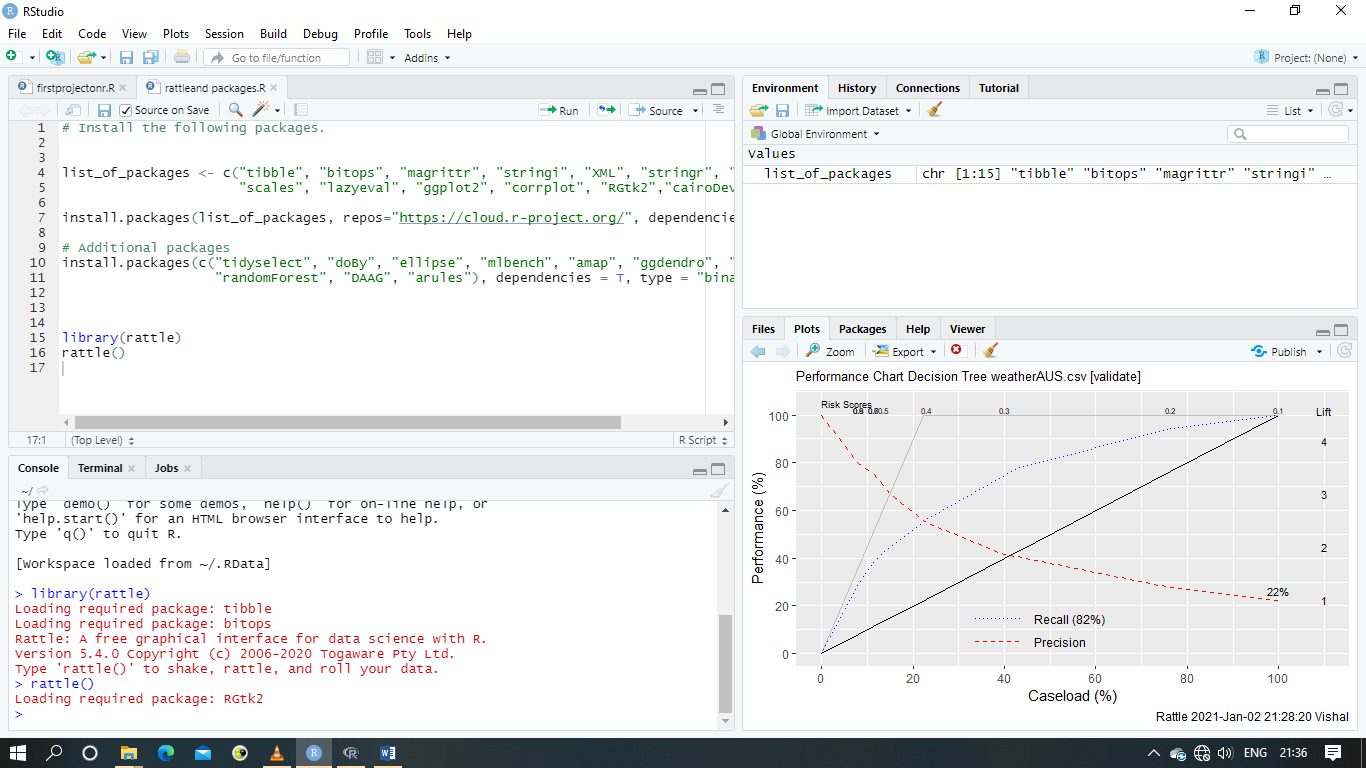
The area under the Risk and Recall curves for Decision Tree model

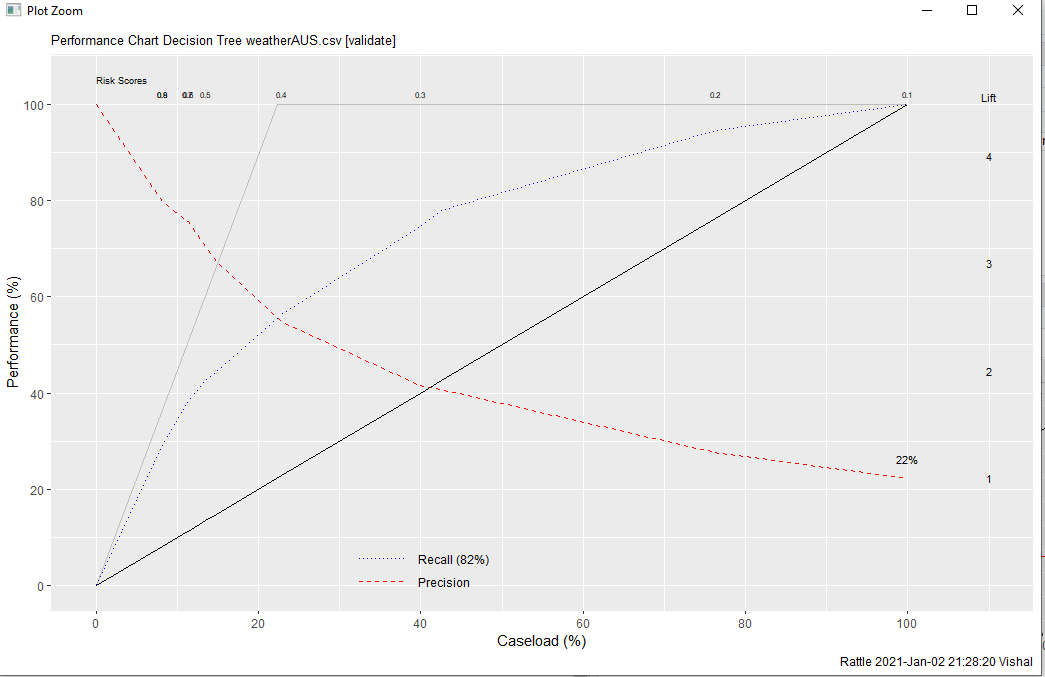
Area under the Recall (green) curve: 82% (0.822)

Rattle timestamp: 2021-01-02 21:28:20 Vishal

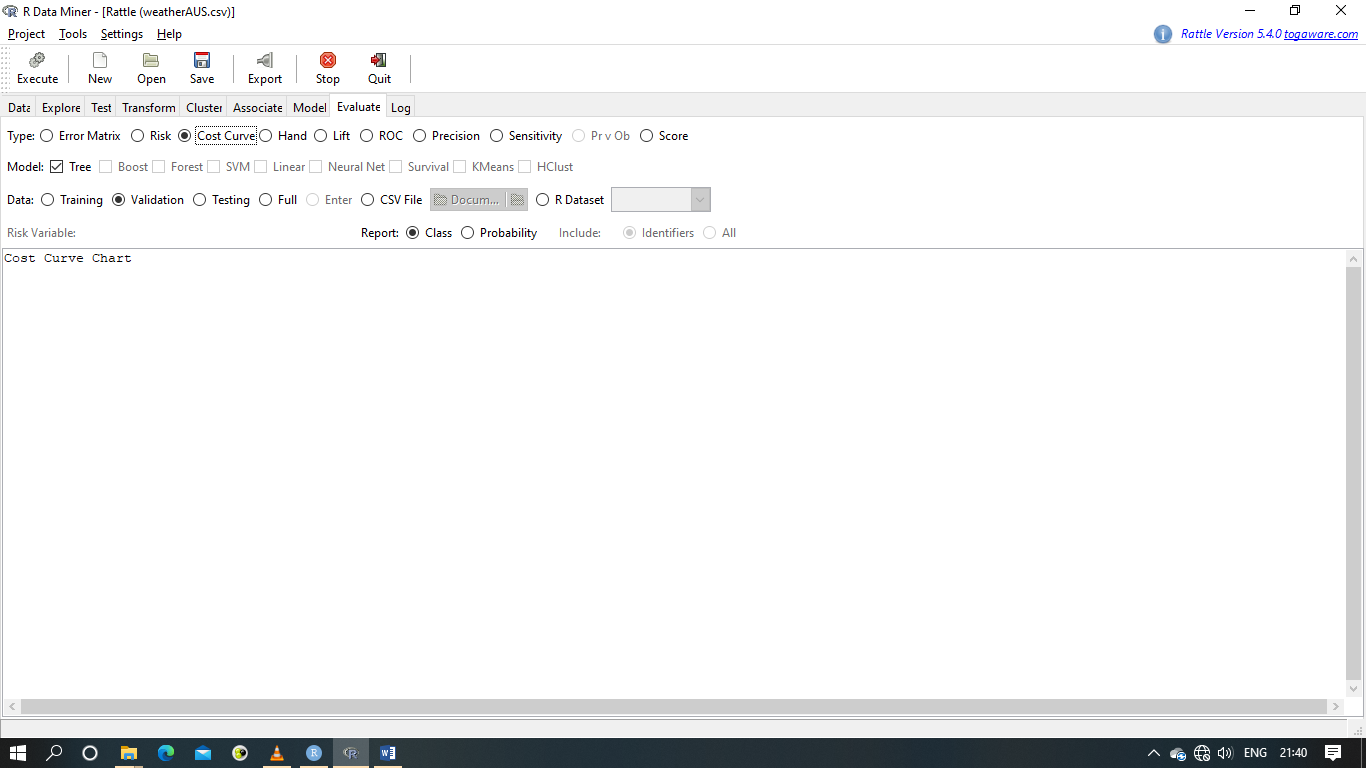
======================================================================

We also get a Performance Chart in the Plotting Region of RStudio Application.





2. Set the Type to Cost Curve and Data to Validation.



Then, we have to click on execute in order to get the final output.

We also get a Probability Function Graph in the Plotting Region of RStudio Application.

