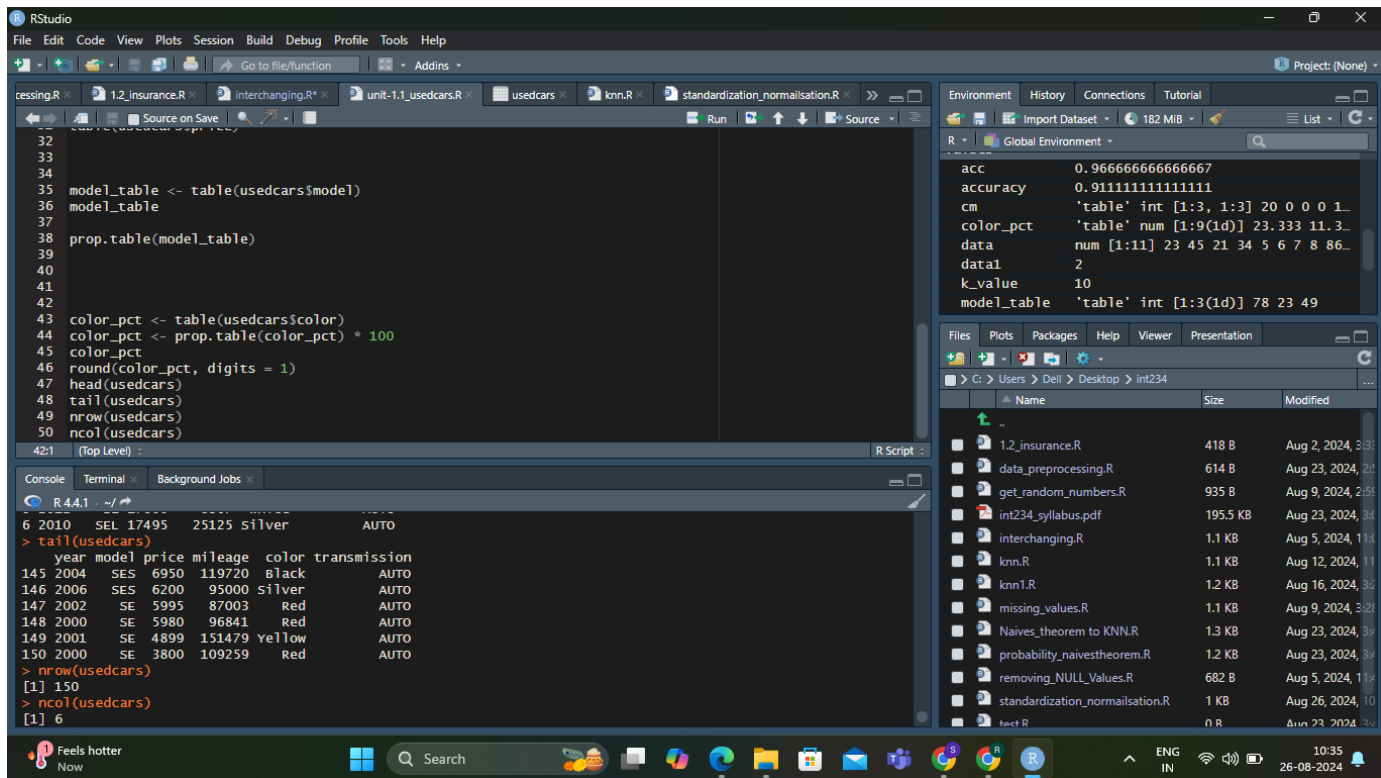


Name: Rohith Reddy Somuguttu



RStudio interface showing the analysis of the 'usedcars' dataset. The script in the editor defines 'model_table' and 'color_pct' from the 'usedcars' dataset. The console shows the tail of the dataset and the number of rows and columns. The Environment pane shows the objects created.

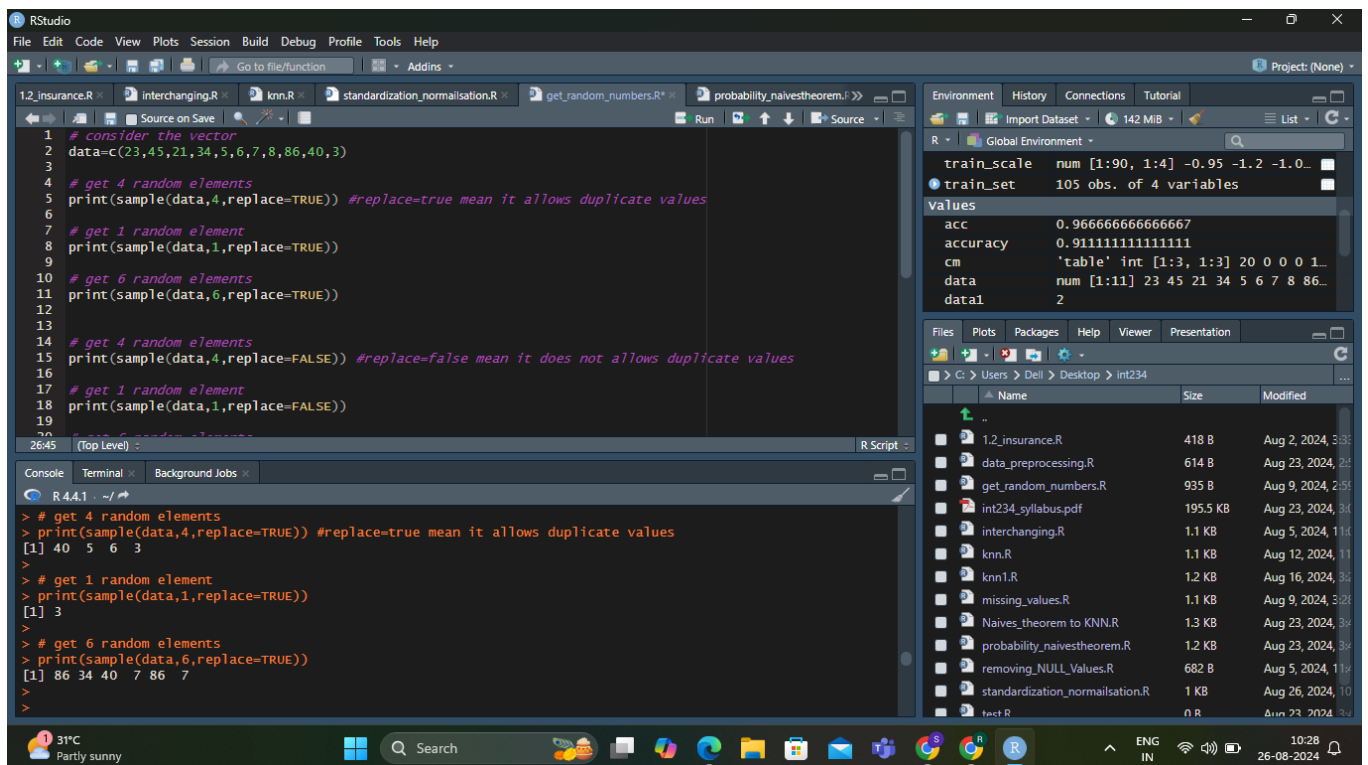
```
32  
33  
34  
35 model_table <- table(usedcars$model)  
36 model_table  
37  
38 prop.table(model_table)  
39  
40  
41  
42  
43 color_pct <- table(usedcars$color)  
44 color_pct <- prop.table(color_pct) * 100  
45 color_pct  
46 round(color_pct, digits = 1)  
47 head(usedcars)  
48 tail(usedcars)  
49 nrow(usedcars)  
50 ncol(usedcars)
```

Console output:

```
6 2010 SEL 17495 25125 Silver AUTO  
> tail(usedcars)  
  year model price mileage color transmission  
145 2004 SES 6950 119720 Black AUTO  
146 2006 SES 6200 95000 Silver AUTO  
147 2002 SE 5995 87003 Red AUTO  
148 2000 SE 5980 96841 Red AUTO  
149 2001 SE 4899 151479 Yellow AUTO  
150 2000 SE 3800 109259 Red AUTO  
> nrow(usedcars)  
[1] 150  
> ncol(usedcars)  
[1] 6
```

Environment pane:

Object	Class	Value
acc	num	0.966666666666667
accuracy	num	0.911111111111111
cm	'table' int	[1:3, 1:3] 20 0 0 1...
color_pct	'table' num	[1:9(1d)] 23.333 11.3...
data	num	[1:11] 23 45 21 34 5 6 7 8 86...
data1	num	2
k_value	num	10
model_table	'table' int	[1:3(1d)] 78 23 49



RStudio interface showing the 'get_random_numbers.R' script. The script demonstrates sampling from a vector 'data' with and without replacement. The console shows the results of these sampling operations. The Environment pane shows the objects created.

```
1 # consider the vector  
2 data=c(23,45,21,34,5,6,7,8,86,40,3)  
3  
4 # get 4 random elements  
5 print(sample(data,4,replace=TRUE)) #replace=true mean it allows duplicate values  
6  
7 # get 1 random element  
8 print(sample(data,1,replace=TRUE))  
9  
10 # get 6 random elements  
11 print(sample(data,6,replace=TRUE))  
12  
13  
14 # get 4 random elements  
15 print(sample(data,4,replace=FALSE)) #replace=false mean it does not allows duplicate values  
16  
17 # get 1 random element  
18 print(sample(data,1,replace=FALSE))  
19
```

Console output:

```
> # get 4 random elements  
> print(sample(data,4,replace=TRUE)) #replace=true mean it allows duplicate values  
[1] 40 5 6 3  
>  
> # get 1 random element  
> print(sample(data,1,replace=TRUE))  
[1] 3  
>  
> # get 6 random elements  
> print(sample(data,6,replace=TRUE))  
[1] 86 34 40 7 86 7  
>  
>
```

Environment pane:

Object	Class	Value
train_scale	num	[1:90, 1:4] -0.95 -1.2 -1.0...
train_set	num	105 obs. of 4 variables
acc	num	0.966666666666667
accuracy	num	0.911111111111111
cm	'table' int	[1:3, 1:3] 20 0 0 1...
data	num	[1:11] 23 45 21 34 5 6 7 8 86...
data1	num	2

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Source on Save Run Source

```
18 dataset$Sales[is.na(dataset$Sales)] <-
19   mean(dataset$Sales, na.rm = TRUE)
20
21 nrow(dataset[!complete.cases(dataset),])
22
23 #there are lot of missing values in columns such as order.Priority, Sales, Ship.Mode,
24 #Profit, Unit.Price and Customer.Name
25 dataset1 <- na.omit(dataset) #removing rows containing missing values and giving it a new
26
27
28 #replace categorical missing values by random value from each variable
29 dataset$order.Priority[is.na(dataset$order.Priority)] <-
30   sample(levels(dataset$order.Priority),
31     size = sum(is.na(dataset$order.Priority)),
32     replace = TRUE)
33 nrow(dataset[!complete.cases(dataset),])
34
35 sum(is.na(dataset))
36
```

30:42 (Top Level) R Script

Console Terminal Background Jobs

```
> dataset1 <- na.omit(dataset) #removing rows containing missing values and giving it a new
>
> #replace categorical missing values by random value from each variable
> dataset$order.Priority[is.na(dataset$order.Priority)] <-
+   sample(levels(dataset$order.Priority),
+     size = sum(is.na(dataset$order.Priority)),
+     replace = TRUE)
> nrow(dataset[!complete.cases(dataset),])
[1] 5
>
> sum(is.na(dataset))
[1] 5
```

Environment History Connections Tutorial

R Global Environment

- dataset1 665 obs. of 12 variables
- iris 150 obs. of 5 variables
- iris_n 150 obs. of 4 variables
- test_cl 60 obs. of 5 variables
- test_scale num [1:60, 1:4] -1.278 -1.392 ...
- test_set 45 obs. of 4 variables
- train_cl 90 obs. of 5 variables
- train_scale num [1:90, 1:4] -0.95 -1.2 -1.0

Files Plots Packages Help Viewer Presentation

C:\Users\> Dell\> Desktop\> int234

Name	Size	Modified
1_2_insurance.R	418 B	Aug 2, 2024, 3:32
data_preprocessing.R	614 B	Aug 23, 2024, 2:44
get_random_numbers.R	935 B	Aug 9, 2024, 2:55
int234_syllabus.pdf	195.5 KB	Aug 23, 2024, 3:44
interchanging.R	1.1 KB	Aug 5, 2024, 11:44
knn.R	1.1 KB	Aug 12, 2024, 11:44
knn1.R	1.2 KB	Aug 16, 2024, 3:44
missing_values.R	1.1 KB	Aug 26, 2024, 10:44
Naives_theorem to KNN.R	1.3 KB	Aug 23, 2024, 3:44
probability_naivestheorem.R	1.2 KB	Aug 23, 2024, 3:44
removing_NULL_Values.R	682 B	Aug 5, 2024, 11:44
standardization_normalisation.R	1 KB	Aug 26, 2024, 10:44
test.R	0 B	Aug 23, 2024, 3:44
unit-1.1_usedcars.R	952 B	Aug 2, 2024, 3:32

31°C Partly sunny

Search

ENG IN 10:37 26-08-2024