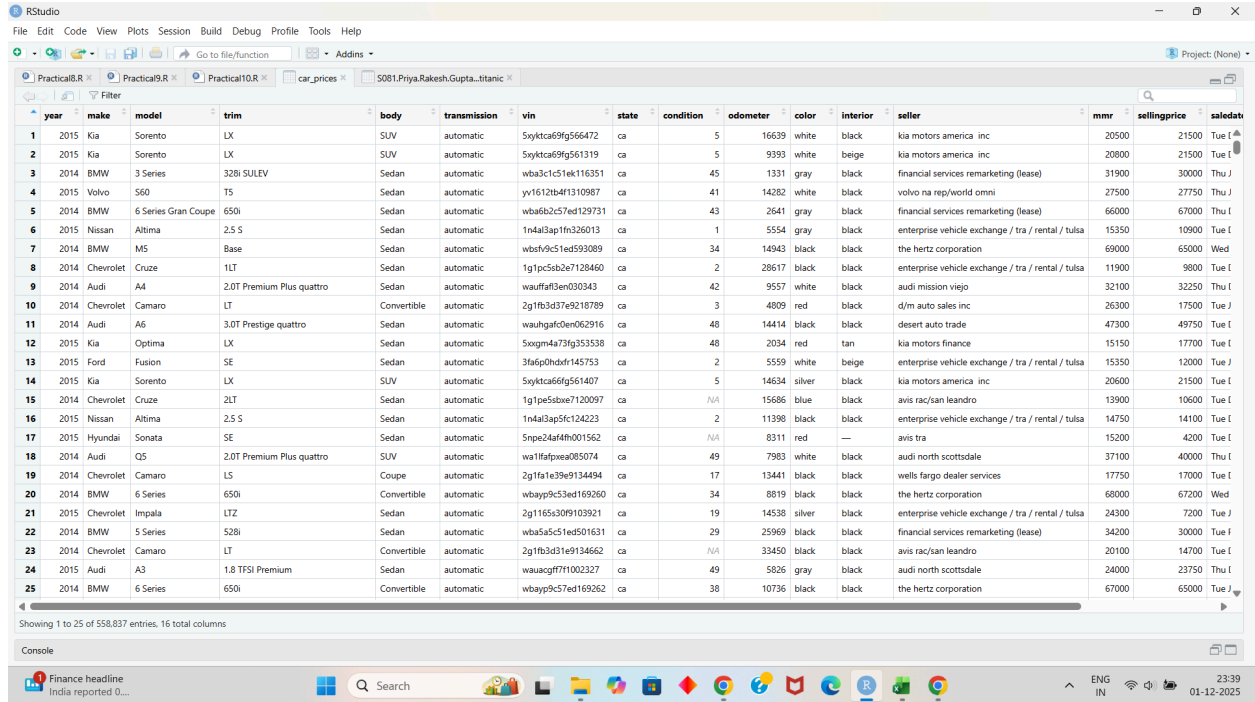


# Sheth L.U.J. & Sir M.V. College

## 10. Creating new variables using transformations and calculations in R. import dataset.



The screenshot displays the RStudio interface with a dataset of car prices loaded. The dataset has 25 rows and 16 columns. The columns are: year, make, model, trim, body, transmission, vin, state, condition, odometer, color, interior, seller, mmp, sellingprice, and saledate. The data shows various car models from different manufacturers, including Kia, BMW, Volvo, Nissan, Chevrolet, Audi, and Ford, with their respective specifications and prices.

	year	make	model	trim	body	transmission	vin	state	condition	odometer	color	interior	seller	mmp	sellingprice	saledate
1	2015	Kia	Sorento	LX	SUV	automatic	5y9kta69fg566472	ca	5	16639	white	black	kia motors america inc	20500	21500	Tue 1
2	2015	Kia	Sorento	LX	SUV	automatic	5y9kta69fg561319	ca	5	9393	white	beige	kia motors america inc	20800	21500	Tue 1
3	2014	BMW	3 Series	328i SULEV	Sedan	automatic	wba3c1c51ek116351	ca	45	1331	gray	black	financial services remarketing (lease)	31900	30000	Thu 1
4	2015	Volvo	S60	T5	Sedan	automatic	yv1612b4f1310987	ca	41	14282	white	black	volvo na rep/world omni	27500	27750	Thu 1
5	2014	BMW	6 Series Gran Coupe	650i	Sedan	automatic	wba6b2c57ed129731	ca	43	2641	gray	black	financial services remarketing (lease)	66000	67000	Tue 1
6	2015	Nissan	Altima	2.5 S	Sedan	automatic	1n4a13ap1fn326013	ca	1	5554	gray	black	enterprise vehicle exchange / tra / rental / tulsa	15350	10900	Tue 1
7	2014	BMW	M5	Base	Sedan	automatic	wbstfv9c51ed593089	ca	34	14943	black	black	the hertz corporation	69000	65000	Wed 1
8	2014	Chevrolet	Cruze	1LT	Sedan	automatic	1g1pc5sb3e7128460	ca	2	28617	black	black	enterprise vehicle exchange / tra / rental / tulsa	11900	9800	Tue 1
9	2014	Audi	A4	2.0T Premium Plus quattro	Sedan	automatic	wauffa13en030343	ca	42	9557	white	black	audi mission viego	32100	32250	Thu 1
10	2014	Chevrolet	Camaro	LT	Convertible	automatic	2g1fb3d37e9218789	ca	3	4809	red	black	d/m auto sales inc	26300	17500	Tue 1
11	2014	Audi	A6	3.0T Prestige quattro	Sedan	automatic	wauhgaf0en062916	ca	48	14414	black	black	desert auto trade	47300	49750	Tue 1
12	2015	Kia	Optima	LX	Sedan	automatic	5x9gm4a73fg353538	ca	48	2034	red	tan	kia motors finance	15150	17700	Tue 1
13	2015	Ford	Fusion	SE	Sedan	automatic	3fa6p0hdxr145753	ca	2	5559	white	beige	enterprise vehicle exchange / tra / rental / tulsa	15350	12000	Tue 1
14	2015	Kia	Sorento	LX	SUV	automatic	5y9kta66fg561407	ca	5	14634	silver	black	kia motors america inc	20600	21500	Tue 1
15	2014	Chevrolet	Cruze	2LT	Sedan	automatic	1g1pe5bwe7120097	ca	N/A	15686	blue	black	avis rac/san leandro	13900	10600	Tue 1
16	2015	Nissan	Altima	2.5 S	Sedan	automatic	1n4a13ap5fc124223	ca	2	11398	black	black	enterprise vehicle exchange / tra / rental / tulsa	14750	14100	Tue 1
17	2015	Hyundai	Sonata	SE	Sedan	automatic	5npe24a4fh001562	ca	N/A	8311	red	---	avis tra	15200	4200	Tue 1
18	2014	Audi	Q5	2.0T Premium Plus quattro	SUV	automatic	wa1fafpwea085074	ca	49	7983	white	black	audi north scottsdale	37100	40000	Tue 1
19	2014	Chevrolet	Camaro	LS	Coupe	automatic	2g1fa1e39e9134494	ca	17	13441	black	black	wells fargo dealer services	17750	17000	Tue 1
20	2014	BMW	6 Series	650i	Convertible	automatic	wba9p9c53ed169260	ca	34	8819	black	black	the hertz corporation	68000	67200	Wed 1
21	2015	Chevrolet	Impala	LTZ	Sedan	automatic	2g1165s309f103921	ca	19	14538	silver	black	enterprise vehicle exchange / tra / rental / tulsa	24300	7200	Tue 1
22	2014	BMW	5 Series	528i	Sedan	automatic	wba5a5c51ed501631	ca	29	25969	black	black	financial services remarketing (lease)	34200	30000	Tue 1
23	2014	Chevrolet	Camaro	LT	Convertible	automatic	2g1fb3d31e9134662	ca	N/A	33450	black	black	avis rac/san leandro	20100	14700	Tue 1
24	2015	Audi	A3	1.8 TFSI Premium	Sedan	automatic	wauacgff711002327	ca	49	5826	gray	black	audi north scottsdale	24000	23750	Thu 1
25	2014	BMW	6 Series	650i	Convertible	automatic	wba9p9c57ed169262	ca	38	10736	black	black	the hertz corporation	67000	65000	Tue 1

Name :- Priya Gupta  
Roll no. :- S081

# Sheth L.U.J. & Sir M.V. College

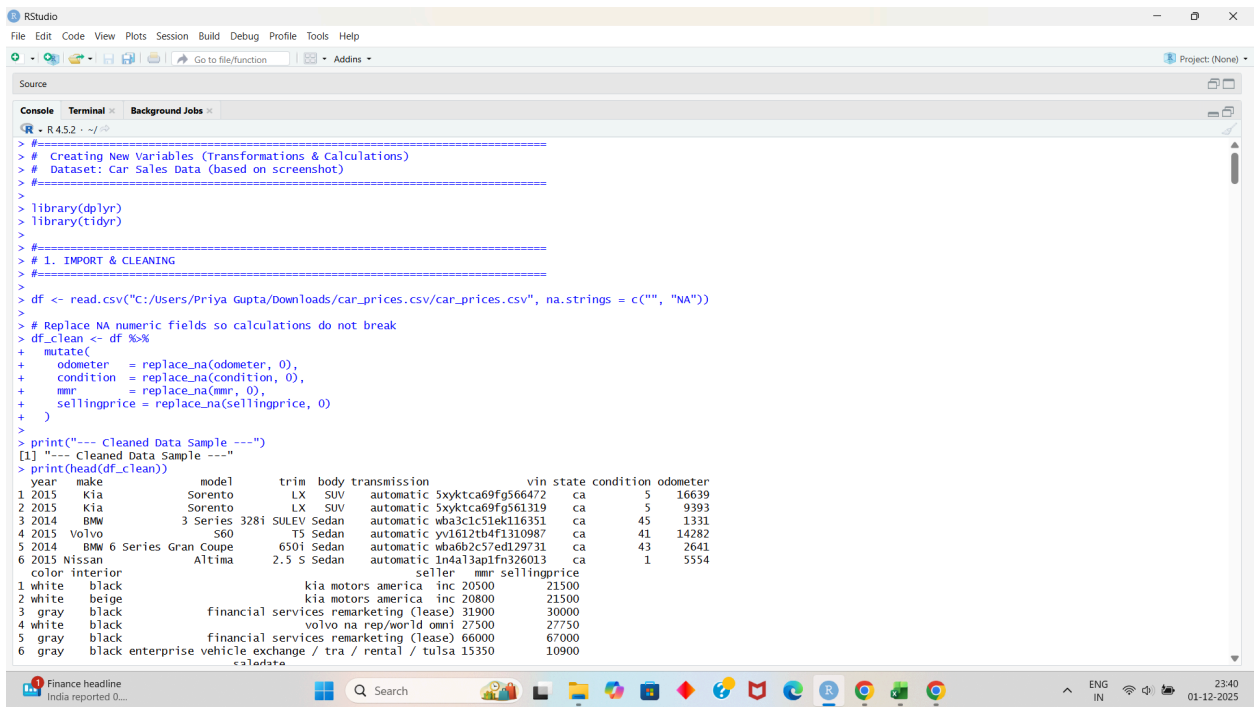
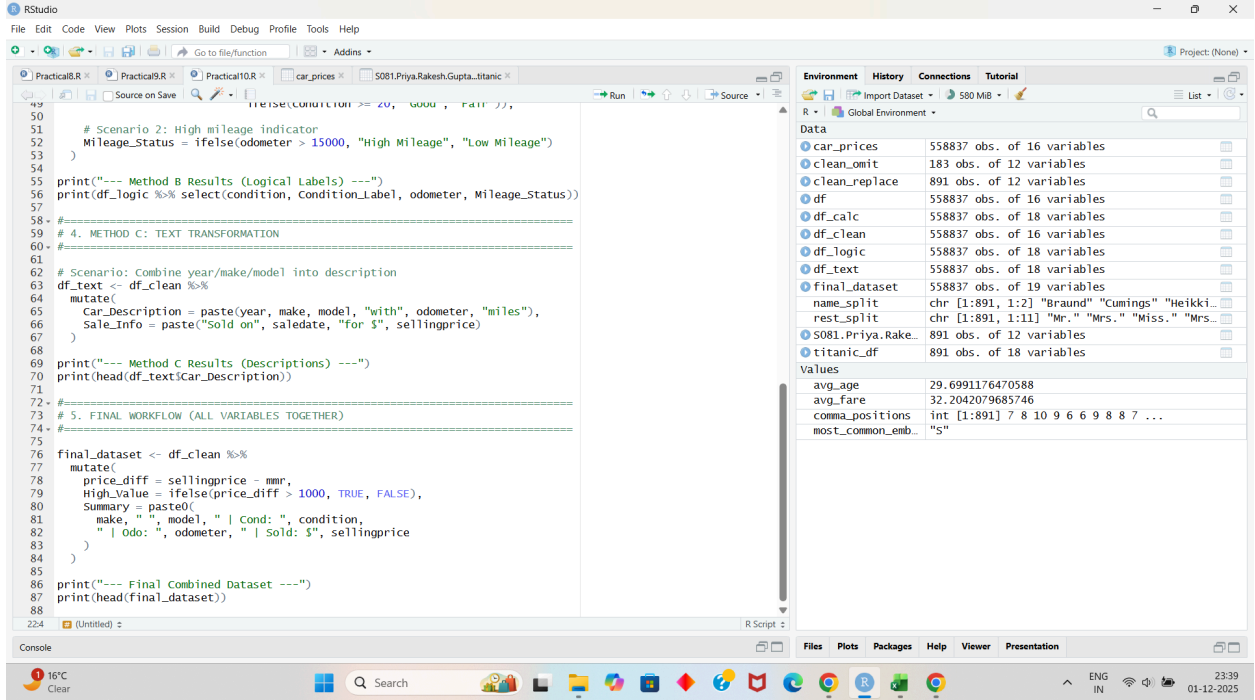
```
1 # =====
2 # Creating New Variables (Transformations & Calculations)
3 # Dataset: Car Sales Data (based on screenshot)
4 # =====
5
6 library(dplyr)
7 library(tidyr)
8
9 # =====
10 # 1. IMPORT & CLEANING
11 # =====
12
13 df <- read.csv("C:/Users/Priya Gupta/Downloads/car_prices.csv/car_prices.csv", na.strings = c("", "NA"))
14
15 # Replace NA numeric fields so calculations do not break
16 df_clean <- df %>%
17   mutate(
18     odometer = replace_na(odometer, 0),
19     condition = replace_na(condition, 0),
20     mmr = replace_na(mmr, 0),
21     sellingprice = replace_na(sellingprice, 0)
22   )
23
24 print("--- Cleaned Data Sample ---")
25 print(head(df_clean))
26
27 # =====
28 # 2. METHOD A: ARITHMETIC CALCULATIONS
29 # =====
30
31 # Scenario: Calculate difference between selling price and MMR value
32 df_calc <- df_clean %>%
33   mutate(
34     price_diff = sellingprice - mmr, # Profit vs MMR
35     price_ratio = sellingprice / (mmr + 1) # Ratio (avoid divide-by-zero)
36   )
37
38 print("--- Method A Results (Price Calculations) ---")
39 print(df_calc %>% select(mmr, sellingprice, price_diff, price_ratio))
40
41 # =====
42 # 3. METHOD B: CONDITIONAL LOGIC
43 # =====
44
45 # Scenario 1: Condition score label
46 df_logic <- df_clean %>%
47   mutate(
48     Condition_Label = ifelse(condition >= 40, "Excellent",
49                             ifelse(condition >= 20, "Good", "Fair")),
50   )
51
52 # Scenario 2: High mileage indicator
53 Mileage_Status = ifelse(odometer > 15000, "High Mileage", "Low Mileage")
54
55 print("--- Method B Results (Logical Labels) ---")
56 print(df_logic %>% select(condition, Condition_Label, odometer, Mileage_Status))
57
58 # =====
59 # 4. METHOD C: TEXT TRANSFORMATION
60 # =====
61
62 # Scenario: Combine year/make/model into description
63 df_text <- df_clean %>%
64   mutate(
65     Car_Description = paste(year, make, model, "with", odometer, "miles"),
66     Sale_Info = paste("sold on", saledate, "for $", sellingprice)
67   )
68
69 print("--- Method C Results (Descriptions) ---")
70 print(head(df_text$Car_Description))
71
72 # =====
73 # 5. FINAL WORKFLOW (ALL VARIABLES TOGETHER)
74 # =====
75
76 final_dataset <- df_clean %>%
77   mutate(
78     price_diff = sellingprice - mmr,
79     High_Value = ifelse(price_diff > 1000, TRUE, FALSE),
80     Survived = nacten()
81   )
82
83 # =====
```

```
41 # =====
42 # 3. METHOD B: CONDITIONAL LOGIC
43 # =====
44
45 # Scenario 1: Condition score label
46 df_logic <- df_clean %>%
47   mutate(
48     Condition_Label = ifelse(condition >= 40, "Excellent",
49                             ifelse(condition >= 20, "Good", "Fair")),
50   )
51
52 # Scenario 2: High mileage indicator
53 Mileage_Status = ifelse(odometer > 15000, "High Mileage", "Low Mileage")
54
55 print("--- Method B Results (Logical Labels) ---")
56 print(df_logic %>% select(condition, Condition_Label, odometer, Mileage_Status))
57
58 # =====
59 # 4. METHOD C: TEXT TRANSFORMATION
60 # =====
61
62 # Scenario: Combine year/make/model into description
63 df_text <- df_clean %>%
64   mutate(
65     Car_Description = paste(year, make, model, "with", odometer, "miles"),
66     Sale_Info = paste("sold on", saledate, "for $", sellingprice)
67   )
68
69 print("--- Method C Results (Descriptions) ---")
70 print(head(df_text$Car_Description))
71
72 # =====
73 # 5. FINAL WORKFLOW (ALL VARIABLES TOGETHER)
74 # =====
75
76 final_dataset <- df_clean %>%
77   mutate(
78     price_diff = sellingprice - mmr,
79     High_Value = ifelse(price_diff > 1000, TRUE, FALSE),
80     Survived = nacten()
81   )
82
83 # =====
```

Name :- Priya Gupta

Roll no. :- S081

**Sheth L.U.J. & Sir M.V. College**



**Name :- Priya Gupta**

**Roll no. :- S081**

# Sheth L.U.J. & Sir M.V. College

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)

Source
Console Terminal Background Jobs
R - R4.5.2 - ~/...
1 Tue Dec 16 2014 12:30:00 GMT-0800 (PST)
2 Tue Dec 16 2014 12:30:00 GMT-0800 (PST)
3 Thu Jan 15 2015 04:30:00 GMT-0800 (PST)
4 Thu Jan 29 2015 04:30:00 GMT-0800 (PST)
5 Thu Dec 18 2014 12:30:00 GMT-0800 (PST)
6 Tue Dec 30 2014 12:00:00 GMT-0800 (PST)
>
> #=====
> # 2. METHOD A: ARITHMETIC CALCULATIONS
> #=====
>
> # Scenario: calculate difference between selling price and MMR value
> df_calc <- df_clean %>%
+   mutate(
+     price_diff = sellingprice - mmr,      # Profit vs MMR
+     price_ratio = sellingprice / (mmr + 1) # Ratio (avoid divide-by-zero)
+   )
>
> print("--- Method A Results (Price Calculations) ---")
[1] "--- Method A Results (Price Calculations) ---"
> print(df_calc %>% select(mmr, sellingprice, price_diff, price_ratio))
  mmr sellingprice price_diff price_ratio
1  20500      21500        1000    1.0487293
2  20800      21500         700    1.0336042
3  31900      30000       -1900    0.9404094
4  27500      27750         250    1.0090542
5  66000      67000         1000    1.0151361
6  15350      10900       -4450    0.7100515
7  69000      65000       -4000    0.9420153
8  11900      9800       -2100    0.8234602
9  32100      32250         150    1.0046416
10 26300      17500      -8800    0.6653739
11 47300      49750         2450    1.0517748
12 15150      17700         2550    1.1682397
13 15350      12000       -3350    0.7817080
14 20600      21500         900    1.0436387
15 13900      10600       -3300    0.7625351
16 14750      14100        -650    0.9558674
17 15200         4200      -11000    0.2762976
18 37100      40000         2900    1.0781381

Finance headline
India reported 0...
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)

Source
Console Terminal Background Jobs
R - R4.5.2 - ~/...
212 14400      21000       -400    0.9816020
213 16450      17500       1050    1.0637651
214 22500      23750       1250    1.0555086
215 23000      21000       -2000    0.9130038
216 12250      12700         450    1.0366501
217 15350      16500       1150    1.0748485
218 21200      21400         200    1.0093863
219 23600      25750       2150    1.0910555
220 23300      25000       1700    1.0729153
221 23900      22000       -1900    0.9204636
222 17750      18200         450    1.0252943
223 22600      18000       -4600    0.7964249
224 21000      22200       1200    1.0570925
225 25300      26250         950    1.0375084
226 11500      11900         400    1.0346926
227 23300      23000       -300    0.9870821
228 27900      27500       -400    0.9856278
229 36800      25000      -11800    0.6793294
230 22600      26500       3900    1.1725145
231 12000      12200         200    1.0165820
232 11900         8700       -3200    0.7310310
233 13150      10800       -2350    0.8212303
234 22400      21750       -650    0.9709388
235 23900      23500       -400    0.9832225
236 23400      22000       -1400    0.9401308
237 11400         4500       -6900    0.3947022
238 23200      23500         300    1.0128874
239 15050      16300       1250    1.0829845
240 23700      22750       -950    0.9598751
241 23800      24500         700    1.0293685
242 15450      14900       -550    0.9643389
243 23100      21000       -2100    0.9090516
244 23000      24750       1750    1.0760402
245 11650         8400       -3250    0.7209682
246 23900      23500       -400    0.9832225
247 22100      21000       -1100    0.9501832
248 11400         9400       -2000    0.8244891
249 26100      26000        -100    0.9961304
250 21800      21400       -400    0.9816063
[ reached 'max' / getOption("max.print") -- omitted 558587 rows ]
>
```

**Name :- Priya Gupta**

**Roll no. :- S081**

**Sheth L.U.J. & Sir M.V. College**

The screenshot shows the RStudio IDE interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The toolbar below the menu contains icons for file operations and running code. The 'Source' pane on the right displays R code for data manipulation using dplyr. The 'Console' pane on the left shows the execution of the code, resulting in a data frame with 22 rows and 4 columns: condition, condition\_Label, odometer, and Mileage\_Status. The data frame is printed to the console, showing the first 22 rows of the dataset.

**R Code in Source Pane:**

```
#=====
> # 3. METHOD B: CONDITIONAL LOGIC
#=====
>
> # Scenario 1: Condition score label
> df_logic <- df_clean %>%
+   mutate(
+     Condition_Label = ifelse(condition >= 40, "Excellent",
+                             ifelse(condition >= 20, "Good", "Fair")),
+
+     # Scenario 2: High mileage indicator
+     Mileage_Status = ifelse(odometer > 15000, "High Mileage", "Low Mileage")
+   )
>
> print("--- Method B Results (Logical Labels) ---")
[1] --- Method B Results (Logical Labels) ---
> print(df_logic %>% select(condition, condition_Label, odometer, Mileage_Status))
```

**Console Output:**

```
condition condition_Label odometer Mileage_Status
1         5         Fair   16639   High Mileage
2         5         Fair   9393    Low Mileage
3        45      Excellent   1331    Low Mileage
4        41      Excellent  14282    Low Mileage
5        43      Excellent   2641    Low Mileage
6         1         Fair   5554    Low Mileage
7        34         Good  14943    Low Mileage
8         2         Fair  28617    High Mileage
9        42      Excellent   9557    Low Mileage
10        3         Fair   4809    Low Mileage
11       48      Excellent  14414    Low Mileage
12       48      Excellent   2034    Low Mileage
13        2         Fair   5559    Low Mileage
14        5         Fair  14634    Low Mileage
15        0         Fair  15686    High Mileage
16        2         Fair  11398    Low Mileage
17        0         Fair   8311    Low Mileage
18       49      Excellent   7983    Low Mileage
19       17         Fair  13441    Low Mileage
20       34         Good   8819    Low Mileage
21       19         Fair  14538    Low Mileage
22       29         Good  25969    High Mileage
```

The image shows a screenshot of the RStudio IDE. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The top toolbar contains icons for file operations and a search bar. The top right corner shows the project name 'Project: (None)'. The main workspace is divided into three panes: Console, Terminal, and Background Jobs. The Console pane is active and displays the output of an R command: 'R 4.5.2 - v7.0'. The output shows a table of car data with columns for car model, condition, mileage, and price. The table is displayed in the Source pane. The Console pane shows the output of a command to print the first 50 rows of the data. The output is as follows:

car	condition	mileage	price
213	42	Excellent	19109
214	43	Excellent	26396
215	21	Good	23507
216	38	Good	34745
217	48	Excellent	10564
218	47	Excellent	16132
219	47	Excellent	16854
220	44	Excellent	19551
221	44	Excellent	14486
222	42	Excellent	17831
223	27	Good	26792
224	43	Excellent	19152
225	44	Excellent	19492
226	34	Good	50160
227	45	Excellent	19368
228	39	Good	12154
229	43	Excellent	11566
230	47	Excellent	25635
231	37	Good	39062
232	2	Fair	29006
233	48	Excellent	8559
234	39	Good	26907
235	34	Good	14765
236	43	Excellent	18673
237	0	Fair	37279
238	37	Good	20707
239	42	Excellent	8467
240	44	Excellent	16009
241	4	Fair	15260
242	44	Excellent	9539
243	36	Good	20976
244	46	Excellent	22294
245	2	Fair	44135
246	4	Fair	30556
247	41	Excellent	29731
248	1	Fair	36957
249	44	Excellent	13321
250	36	Good	6552

**Name :- Priya Gupta**

**Roll no. :- S081**

# Sheth L.U.J. & Sir M.V. College

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)

Source
Console Terminal Background Jobs
R - R 4.5.2 - ~/ /
> #=====
> # 4. METHOD C: TEXT TRANSFORMATION
> #=====
>
> # Scenario: Combine year/make/model into description
> df_text <- df_clean %>%
+   mutate(
+     Car_Description = paste(year, make, model, "with", odometer, "miles"),
+     Sale_Info = paste("Sold on", saledate, "For $", sellingprice)
+   )
>
> print("--- Method C Results (Descriptions) ---")
[1] "--- Method C Results (Descriptions) ---"
> print(head(df_text$Car_Description))
[1] "2015 Kia Sorento with 16639 miles"      "2015 Kia Sorento with 9393 miles"
[3] "2014 BMW 3 Series with 1331 miles"      "2015 Volvo S60 with 14282 miles"
[5] "2014 BMW 6 Series Gran Coupe with 2641 miles" "2015 Nissan Altima with 5554 miles"
>
> #=====
> # 5. FINAL WORKFLOW (ALL VARIABLES TOGETHER)
> #=====
>
> final_dataset <- df_clean %>%
+   mutate(
+     price_diff = sellingprice - mmp,
+     High_Value = ifelse(price_diff > 1000, TRUE, FALSE),
+     Summary = paste0(
+       make, " ", model, " | Cond: ", condition,
+       " | Odo: ", odometer, " | Sold: $", sellingprice
+     )
+   )
>
> print("--- Final Combined Dataset ---")
[1] "--- Final Combined Dataset ---"
> print(head(final_dataset))
  year make      model      trim  body transmission      vin state condition odometer
1 2015  Kia      Sorento      LX   SUV      automatic 5xyktca69fg566472 ca    5      16639
2 2015  Kia      Sorento      LX   SUV      automatic 5xyktca69fg561319 ca    5      9393
3 2014  BMW      3 Series 328i SULEV Sedan      automatic wba3c1c51ek116351 ca    45     1331
4 2015  Volvo      S60        T5 Sedan      automatic yv1612tb4f1310987 ca    41    14282
5 2014  BMW      6 Series Gran Coupe 650i Sedan      automatic wba6b2c57ed129731 ca    43     2641
6 2014  BMW      6 Series Gran Coupe 650i Sedan      automatic wba6b2c57ed129731 ca    43     2641
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)

Source
Console Terminal Background Jobs
R - R 4.5.2 - ~/ /
+   price_diff = sellingprice - mmp,
+   High_Value = ifelse(price_diff > 1000, TRUE, FALSE),
+   Summary = paste0(
+     make, " ", model, " | Cond: ", condition,
+     " | Odo: ", odometer, " | Sold: $", sellingprice
+   )
+ )
>
> print("--- Final Combined Dataset ---")
[1] "--- Final Combined Dataset ---"
> print(head(final_dataset))
  year make      model      trim  body transmission      vin state condition odometer
1 2015  Kia      Sorento      LX   SUV      automatic 5xyktca69fg566472 ca    5      16639
2 2015  Kia      Sorento      LX   SUV      automatic 5xyktca69fg561319 ca    5      9393
3 2014  BMW      3 Series 328i SULEV Sedan      automatic wba3c1c51ek116351 ca    45     1331
4 2015  Volvo      S60        T5 Sedan      automatic yv1612tb4f1310987 ca    41    14282
5 2014  BMW      6 Series Gran Coupe 650i Sedan      automatic wba6b2c57ed129731 ca    43     2641
6 2015  Nissan    Altima      2.5 S Sedan      automatic 1n4al3ap1fn326013 ca    1     5554
  color interior      seller      mmp sellingprice
1 white  black      kia motors america  inc 20500      21500
2 white  beige      kia motors america  inc 20800      21500
3 gray  black      financial services remarketing (lease) 31900      30000
4 white  black      volvo na rep/world omni 27500      27750
5 gray  black      financial services remarketing (lease) 66000      67000
6 gray  black enterprise vehicle exchange / tra / rental / tulsa 15350      10900
  saledate price_diff High_Value
1 Tue Dec 16 2014 12:30:00 GMT-0800 (PST) 1000 FALSE
2 Tue Dec 16 2014 12:30:00 GMT-0800 (PST) 700 FALSE
3 Thu Jan 15 2015 04:30:00 GMT-0800 (PST) -1900 FALSE
4 Thu Jan 29 2015 04:30:00 GMT-0800 (PST) 250 FALSE
5 Thu Dec 18 2014 12:30:00 GMT-0800 (PST) 1000 FALSE
6 Tue Dec 30 2014 12:00:00 GMT-0800 (PST) -4450 FALSE
  Summary
1 Kia Sorento | Cond: 5 | Odo: 16639 | Sold: $21500
2 Kia Sorento | Cond: 5 | Odo: 9393 | Sold: $21500
3 BMW 3 Series | Cond: 45 | Odo: 1331 | Sold: $30000
4 Volvo S60 | Cond: 41 | Odo: 14282 | Sold: $27750
5 BMW 6 Series Gran Coupe | Cond: 43 | Odo: 2641 | Sold: $67000
6 Nissan Altima | Cond: 1 | Odo: 5554 | Sold: $10900
> |
```

Name :- Priya Gupta

Roll no. :- S081

**Sheth L.U.J. & Sir M.V. College**

**Name :- Priya Gupta**

**Roll no. :- S081**