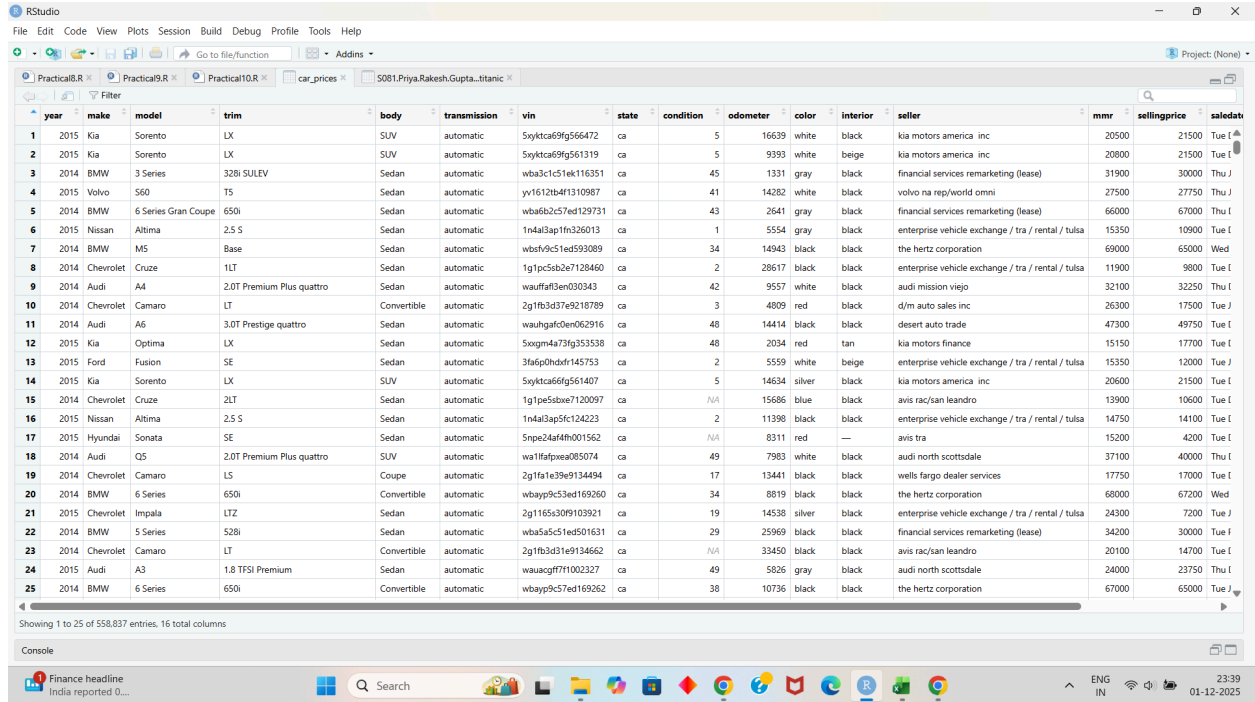# Sheth L.U.J. & Sir M.V. College

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

Name :- Priya Gupta

Roll no. :- SO81

```r
#=======================================================
# Creating New Variables (Transformations & Calculations)
# Dataset: Car Sales Data (based on screenshot)
#=======================================================

library(dplyr)
library(tidyr)

#=======================================================
# 1. IMPORT & CLEANING
#=======================================================

df <- read.csv("C:/Users/Priya Gupta/Downloads/car_prices.csv/car_prices.csv", na.strings = c("", "NA"))

# Replace NA numeric fields so calculations do not break
df_clean <- df %>%
  mutate(
    odometer    = replace_na(odometer, 0),
    condition   = replace_na(condition, 0),
    mmr         = replace_na(mmr, 0),
    sellingprice = replace_na(sellingprice, 0)
  )

print("--- Cleaned Data Sample ---")
print(head(df_clean))

#=======================================================
# 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) ---")
print(df_calc %>% select(mmr, sellingprice, price_diff, price_ratio))
```

```r
#=======================================================
# 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) ---")
print(df_logic %>% select(condition, Condition_Label, odometer, Mileage_Status))

#=======================================================
# 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) ---")
print(head(df_text$Car_Description))

#=======================================================
# 5. FINAL WORKFLOW (ALL VARIABLES TOGETHER)
#=======================================================

final_dataset <- df_clean %>%
  mutate(
    price_diff = sellingprice - mmr,
    High_Value = ifelse(price_diff > 1000, TRUE, FALSE),
    Summary = paste0(
```

Name :- Priya Gupta

Roll no. :- SO81

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


```r
        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) ---")
print(df_logic %>% select(condition, Condition_Label, odometer, Mileage_Status))

#==============================================================
# 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) ---")
print(head(df_text$Car_Description))

#==============================================================
# 5. FINAL WORKFLOW (ALL VARIABLES TOGETHER)
#==============================================================

final_dataset <- df_clean %>%
  mutate(
    price_diff = sellingprice - mmr,
    High_Value = ifelse(price_diff > 1000, TRUE, FALSE),
    Summary = paste0(
      make, " ", model, " | Cond: ", condition,
      " | Odo: ", odometer, " | Sold: $", sellingprice
    )
  )

print("--- Final Combined Dataset ---")
print(head(final_dataset))
```
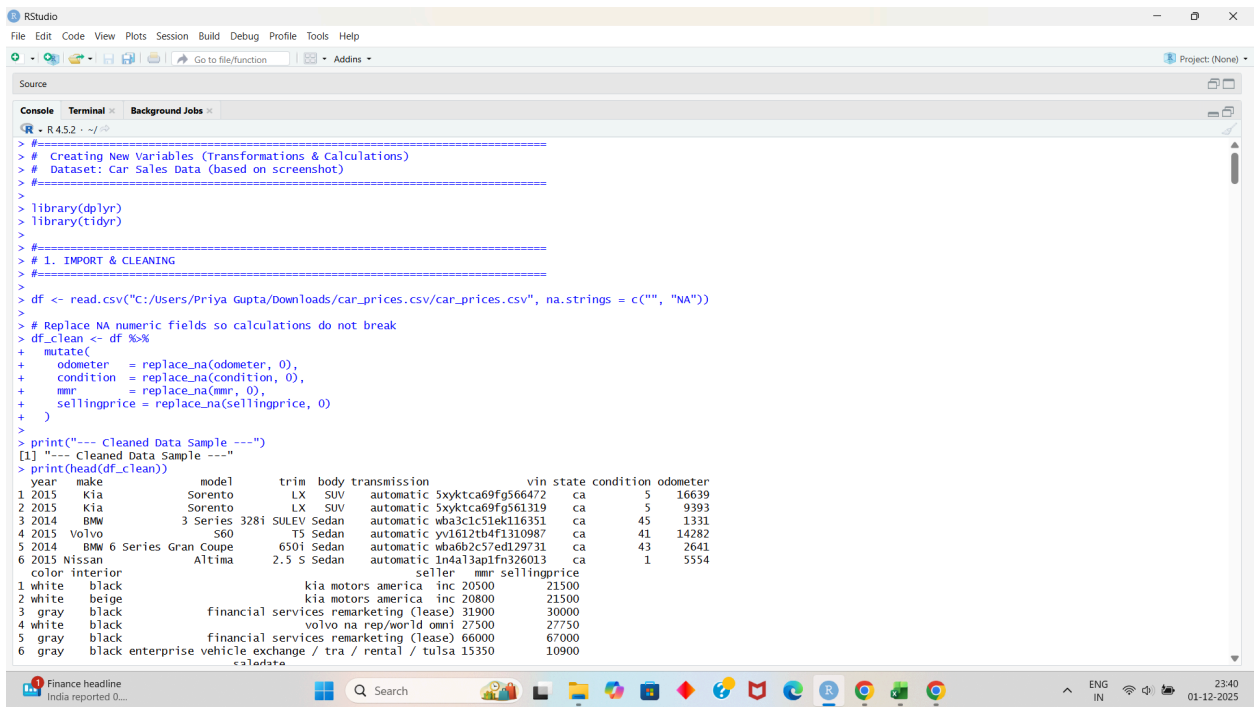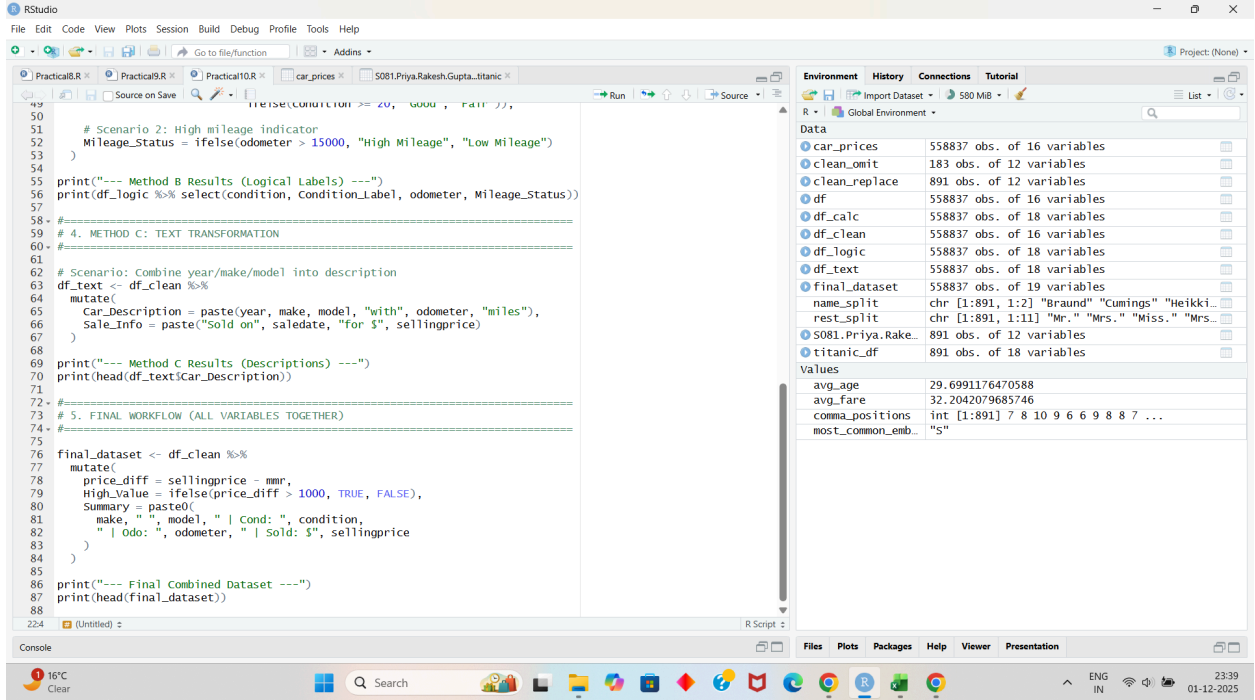


```r
> #==============================================================
> #  Creating New Variables (Transformations & Calculations)
> #  Dataset: Car Sales Data (based on screenshot)
> #==============================================================
>
> library(dplyr)
> library(tidyr)
>
> #==============================================================
> # 1. IMPORT & CLEANING
> #==============================================================
>
> df <- read.csv("C:/Users/Priya Gupta/Downloads/car_prices.csv/car_prices.csv", na.strings = c("", "NA"))
>
> # Replace NA numeric fields so calculations do not break
> df_clean <- df %>%
+   mutate(
+     odometer    = replace_na(odometer, 0),
+     condition   = replace_na(condition, 0),
+     mmr         = replace_na(mmr, 0),
+     sellingprice = replace_na(sellingprice, 0)
+   )
>
> print("--- Cleaned Data Sample ---")
[1] "--- Cleaned Data Sample ---"
> print(head(df_clean))
  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   mmr 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
```

Name :- Priya Gupta
Roll no. :- SO81

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


```
R RStudio
File  Edit  Code  View  Plots  Session  Build  Debug  Profile  Tools  Help
Go to file/function                              Addins
Source
Console  Terminal  Background Jobs
R 4.5.2 · ~/
                              JaceDate
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
```



```
R RStudio
File  Edit  Code  View  Plots  Session  Build  Debug  Profile  Tools  Help
Go to file/function                              Addins
Source
Console  Terminal  Background Jobs
R 4.5.2 · ~/
212 21400        21000       -400   0.9812626
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. :- SO81

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

Console Terminal × Background Jobs ×

R 4.5.2 · ~/

```
> #=============================================================
> # 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))
   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
```

Finance headline
India reported 0....

Q Search

ENG
IN

23:41
01-12-2025

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source

Console Terminal × Background Jobs ×

R 4.5.2 · ~/

```
213        42       Excellent    19109   High Mileage
214        43       Excellent    26396   High Mileage
215        21            Good    23507   High Mileage
216        38            Good    34745   High Mileage
217        48       Excellent    10564    Low Mileage
218        47       Excellent    16132   High Mileage
219        47       Excellent    16854   High Mileage
220        44       Excellent    19551   High Mileage
221        44       Excellent    14486    Low Mileage
222        42       Excellent    17831   High Mileage
223        27            Good    26792   High Mileage
224        43       Excellent    19152   High Mileage
225        44       Excellent    19492   High Mileage
226        34            Good    50160   High Mileage
227        45       Excellent    19368   High Mileage
228        39            Good    12154    Low Mileage
229        43       Excellent    11566    Low Mileage
230        47       Excellent    25635   High Mileage
231        37            Good    39062   High Mileage
232         2            Fair    29006   High Mileage
233        48       Excellent     8559    Low Mileage
234        39            Good    26907   High Mileage
235        34            Good    14765    Low Mileage
236        43       Excellent    18673   High Mileage
237         0            Fair    37279   High Mileage
238        37            Good    20707   High Mileage
239        42       Excellent     8467    Low Mileage
240        44       Excellent    16009   High Mileage
241         4            Fair    15260   High Mileage
242        44       Excellent     9539    Low Mileage
243        36            Good    20976   High Mileage
244        46       Excellent    22294   High Mileage
245         2            Fair    44135   High Mileage
246         4            Fair    30556   High Mileage
247        41       Excellent    29731   High Mileage
248         1            Fair    36957   High Mileage
249        44       Excellent    13321    Low Mileage
250        36            Good     6552    Low Mileage
[ reached 'max' / getOption("max.print") -- omitted 558587 rows ]
>
```

Finance headline
India reported 0....

Q Search

ENG
IN

23:42
01-12-2025

Name :- Priya Gupta
Roll no. :- SO81

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

Name :- Priya Gupta

Roll no. :- SO81

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

Name :- Priya Gupta

Roll no. :- S081