ASSIGNMENT-2

Name : Siddartha Rao Damarla

I'D:
16337125

	...1	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kml
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	13 km/kg
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77 kml
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl
5	6	Nissan Micra Diesel XV	Jaipur	2013	86999	Diesel	Manual	First	Second mp
6	7	Toyota Innova Crysta 2.8 GX AT 8S	Mumbai	2016	36000	Diesel	Automatic	First	11.36 kml
7	8	Volkswagen Vento Diesel Comfortline	Pune	2013	64430	Diesel	Manual	First	20.54 kml
8	9	Tata Indica Vista Quadrajet LS	Chennai	2012	65932	Diesel	Manual	Second	22.3 kmpl
9	10	Maruti Ciaz Zeta	Kochi	2018	25692	Petrol	Manual	First	21.56 kml
10	11	Honda City 1.5 V AT Sunroof	Kolkata	2012	60000	Petrol	Automatic	First	16.8 kmpl
11	12	Maruti Swift VDI BSIV	Jaipur	2015	64424	Diesel	Manual	First	25.2 kmpl
12	13	Land Rover Range Rover 2.2L Pure	Delhi	2014	72000	Diesel	Automatic	First	12.7 kmpl
13	14	Land Rover Freelander 2 TD4 SE	Pune	2012	85000	Diesel	Automatic	Second	0.0 kmpl
14	15	Mitsubishi Pajero Sport 4X4	Delhi	2014	110000	Diesel	Manual	First	13.5 kmpl
15	16	Honda Amaze S i-Dtech	Kochi	2016	58950	Diesel	Manual	First	25.8 kmpl
16	17	Maruti Swift DDiS VDI	Jaipur	2017	25000	Diesel	Manual	First	28.4 kmpl
17	18	Renault Duster 85PS Diesel RxL Plus	Kochi	2014	77469	Diesel	Manual	First	20.45 kml

Showing 1 to 18 of 5,847 entries, 14 total columns

Console **Terminal** **Background Jobs**

R 4.3.2 · ~/

```
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> library(readr)
> train <- read_csv("Downloads/train.csv")
New names:
• `` -> `...1`
Rows: 5847 Columns: 14
── Column specification ─────────────────────────────────────
Delimiter: ","
chr (9): Name, Location, Fuel_Type, Transmission, Owner_Type, Mileage, Engine...
dbl (5): ...1, Year, Kilometers_Driven, Seats, Price

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
> View(train)
> |
```

> is.na(train)

	...1	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
[1,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[2,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[3,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[4,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[5,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[6,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[7,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[8,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[9,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[10,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[11,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[12,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[13,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[14,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[15,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[16,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[17,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[18,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[19,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[20,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[21,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[22,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[23,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[24,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[25,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[26,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[27,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
[28,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[29,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
[30,]	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE

```
> is.na(train$New_Price)
  [1]  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE
 [21]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
 [41]  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE
 [61]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
 [81]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
[101]  TRUE FALSE  TRUE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE
[121]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[141] FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[161]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE FALSE
[181]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[201]  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE FALSE
[221]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[241]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
[261]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
[281]  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
[301]  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[321]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE  TRUE
[341]  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE
[361]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
[381]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE
[401]  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[421]  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE
[441]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[461]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
[481]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE
[501]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
> 
> train$New_Price=ifelse(is.na(train$New_Price),median(train$Mileage,na.rm=TRUE),
train$Mileage)
> 
```

Seats	New_Price	Price
5	4.59 Lakh	12.50
5	8.61 Lakh	4.50
7	4.59 Lakh	6.00
5	4.59 Lakh	17.74
5	4.59 Lakh	3.50
8	21 Lakh	17.50
5	4.59 Lakh	5.20
5	4.59 Lakh	1.95
5	10.65 Lakh	9.95
5	4.59 Lakh	4.49
5	4.59 Lakh	5.60
5	4.59 Lakh	27.00
5	4.59 Lakh	17.50
7	32.01 Lakh	15.00
5	4.59 Lakh	5.40
5	4.59 Lakh	5.99
5	4.59 Lakh	6.34

Remove the units from some of the attributes and only keep the numerical for Mileage by removing kmpl, CC from Engine,bhp from power and lakh from New_price.

Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
Diesel	Manual	First	19.67	1582	126.20	5	4.59 Lakh	12.50
Petrol	Manual	First	18.19	1199	88.70	5	8.61 Lakh	4.50
Diesel	Manual	First	20.77	1248	88.76	7	4.59 Lakh	6.00
Diesel	Automatic	Second	15.20	1968	140.80	5	4.59 Lakh	17.74
Diesel	Manual	First	23.08	1461	63.10	5	4.59 Lakh	3.50
Diesel	Automatic	First	11.36	2755	171.50	8	21 Lakh	17.50
Diesel	Manual	First	20.54	1598	103.60	5	4.59 Lakh	5.20
Diesel	Manual	Second	22.30	1248	74.00	5	4.59 Lakh	1.95
Petrol	Manual	First	21.56	1462	103.25	5	10.65 Lakh	9.95
Petrol	Automatic	First	16.80	1497	116.30	5	4.59 Lakh	4.49
Diesel	Manual	First	25.20	1248	74.00	5	4.59 Lakh	5.60
Diesel	Automatic	First	12.70	2179	187.70	5	4.59 Lakh	27.00
Diesel	Automatic	Second	0.00	2179	115.00	5	4.59 Lakh	17.50

Mileage	Engine	Power	Seats	New_Price
19.67	1582	126.20	5	4.59
18.19	1199	88.70	5	8.61
20.77	1248	88.76	7	4.59
15.20	1968	140.80	5	4.59
23.08	1461	63.10	5	4.59
11.36	2755	171.50	8	21.00
20.54	1598	103.60	5	4.59
22.30	1248	74.00	5	4.59
21.56	1462	103.25	5	10.65
16.80	1497	116.30	5	4.59
25.20	1248	74.00	5	4.59
12.70	2179	187.70	5	4.59
0.00	2179	115.00	5	4.59
13.50	2477	175.56	7	32.01
25.80	1498	98.60	5	4.59

Change the categorical variables the Fuel_Type and Transmission into numeric values using the hot value encoded.

```
> train_encoded <- cbind(train, model.matrix(~ Fuel_Type + Transmission - 1,data = train))
> train_encoded <- train_encoded[, !(names(train) %in% c("Fuel_Type", "Transmission "))]
> head(train_encoded)
  ...1                          Name    Location Year Kilometers_Driven Transmission
1    1  Hyundai Creta 1.6 CRDi SX Option      Pune 2015            41000       Manual
2    2                    Honda Jazz V   Chennai 2011            46000       Manual
3    3               Maruti Ertiga VDI   Chennai 2012            87000       Manual
4    4  Audi A4 New 2.0 TDI Multitronic Coimbatore 2013            40670    Automatic
5    6           Nissan Micra Diesel XV    Jaipur 2013            86999       Manual
6    7 Toyota Innova Crysta 2.8 GX AT 8S    Mumbai 2016            36000    Automatic
  Owner_Type Mileage Engine Power Seats New_Price Price Fuel_TypeDiesel Fuel_TypeElectric
1      First   19.67     NA    NA     5     19.67 12.50               1                 0
2      First      NA     NA    NA     5     18.19  4.50               0                 0
3      First   20.77     NA    NA     7     20.77  6.00               1                 0
4     Second   15.20     NA    NA     5     15.20 17.74               1                 0
5      First   23.08     NA    NA     5     23.08  3.50               1                 0
6      First   11.36     NA    NA     8     11.36 17.50               1                 0
  Fuel_TypePetrol TransmissionManual
1               0                  1
2               1                  1
3               0                  1
4               0                  0
5               0                  1
6               0                  0
>
```

Calculating the current age by subtracting the year value from current year.

```
> Current_year <- as.numeric(format(Sys.Date(),"%Y"))
> train$Current_Age <- Current_year - train$Year
> 
```

New_Price	Price	Current_Age
4.59 Lakh	12.50	8
8.61 Lakh	4.50	12
4.59 Lakh	6.00	11
4.59 Lakh	17.74	10
4.59 Lakh	3.50	10
21 Lakh	17.50	7
4.59 Lakh	5.20	10
4.59 Lakh	1.95	11
10.65 Lakh	9.95	5
4.59 Lakh	4.49	11
4.59 Lakh	5.60	8
4.59 Lakh	27.00	9
4.59 Lakh	17.50	11
32.01 Lakh	15.00	9