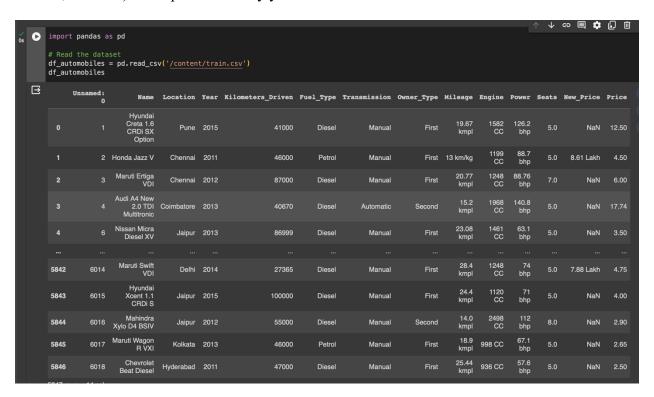
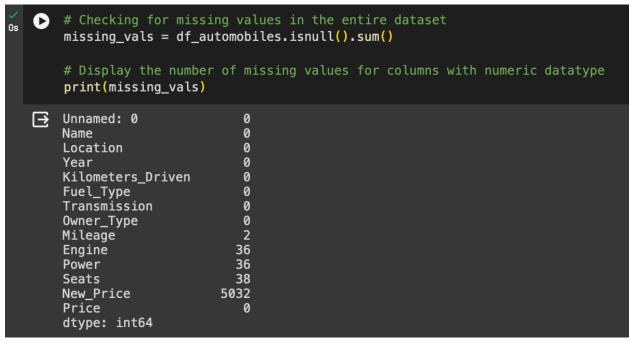
# PDS ASSIGNMENT - 2

**ID**: 16344280

Name: Vinisha Vimal Kumar Shukla

a). Look for the missing values in all the columns and either impute them (replace with mean, median, or mode) or drop them. Justify your action for this task.





```
# Identify and count missing values in columns that have character datatype
    missing_vals_char = df_automobiles.select_dtypes(include=['object']).isnull().sum()
    # Display the number of missing values in columns that have character datatype
    print(missing_vals_char)
Name
    Location
                       0
    Fuel_Type
Transmission
                       0
                       0
    0wner_Type
                       2
    Mileage
    Engine
                      36
    Power
    New_Price
    dtype: int64
```

```
Name 0
Location 0
Year 0
Kilometers_Driven 0
Fuel_Type 0
Transmission 0
Owner_Type 0
Mileage 0
Engine 0
Power 0
Seats 0
Price 0
dtype: int64
<ipython-input-16-0ce23b57a77a>:8: FutureWarning: The default value of numeric_only in DataFrame.mean is deprecated. In a future version, it will defind f_automobiles.fillna(df_automobiles.mean(), inplace=True)
```

```
Name
                                         Location
                                                    Year
                                                           Kilometers_Driven
  Hyundai Creta 1.6 CRDi SX Option
                                              Pune
                                                    2015
                                                                        41000
                        Honda Jazz V
                                           Chennai
                                                    2011
                                                                        46000
                   Maruti Ertiga VDI
2
                                           Chennai
                                                    2012
                                                                        87000
3
    Audi A4 New 2.0 TDI Multitronic Coimbatore
                                                    2013
                                                                        40670
             Nissan Micra Diesel XV
                                                    2013
                                            Jaipur
             Transmission Owner_Type Mileage
1 First 19.67
   Fuel_Type
                                                   Engine
                                                             Power
                                                                    Seats
                                                                            Price
                                                    1582.0
                                                            126.20
                                                                       5.0
                                                                            12.50
                                  First
                                                                       5.0
           1
                                            13.00
                                                   1199.0
                                                             88.70
                                                                             4.50
2
           0
                                  First
                                            20.77
                                                    1248.0
                                                             88.76
                                                                       7.0
                                                                             6.00
                           1
3
           0
                           0
                                 Second
                                            15.20
                                                    1968.0
                                                            140.80
                                                                       5.0
                                                                            17.74
4
                                                             63.10
                                                                       5.0
                                                                             3.50
           0
                                  First
                                            23.08
                                                   1461.0
                                                                     + Code
                                                                                + Text
```

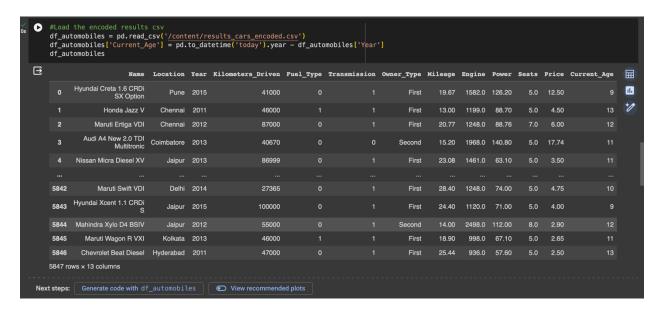
# **Justification:**

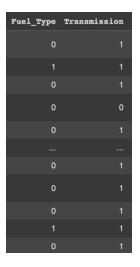
- It ensures data integrity and completeness by retaining all available information in the dataset.
- By keeping observations with missing values, it maintains the sample size, preventing data loss and preserving statistical power.
- Imputation methods like mean and mode utilize existing data distribution, providing representative values for missing entries.
- This approach mitigates the risk of introducing bias into the analysis, as imputed values are derived from the observed data.
- Utilizing available information maximizes the dataset's potential for accurate analysis and modeling.
- It enhances the robustness of the analysis by leveraging the entire dataset, even in the presence of missing values.
- Imputing missing values facilitates downstream tasks like modeling and visualization, as the dataset remains intact and ready for analysis without additional preprocessing steps.

**b).** Remove the units from some of the attributes and only keep the numerical values (for example remove kmpl from "Mileage", CC from "Engine", bhp from "Power", and lakh from "New price").

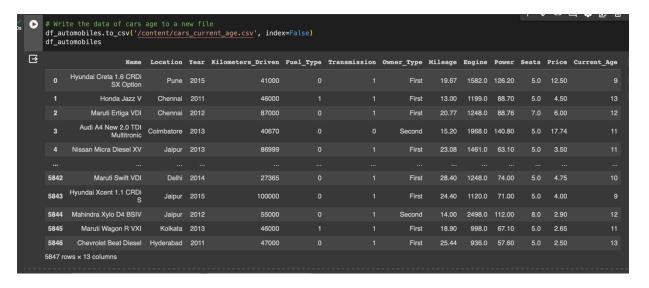
```
df_automobiles['Mileage'] = df_automobiles['Mileage'].str.replace('[^0-9.]', '', regex=True).astype(float)
df_automobiles['Engine'] = df_automobiles['Engine'].str.replace('[^0-9.]', '', regex=True).astype(float)
df_automobiles['Power'] = df_automobiles['Power'].str.replace('[^0-9.]', '', regex=True).astype(float)
df_automobiles['New_Price'] = df_automobiles['New_Price'].str.replace('[^0-9.]', '', regex=True).astype(float)
                                                                                                                           , regex=True).astype(float)
                                                                                                                              '', regex=True).astype(float)
     print(df_automobiles.describe())
\blacksquare
                                                      Kilometers_Driven
5.847000e+03
                                                                                   Mileage
5845.000000
                Unnamed: 0 5847.000000
                                                                                                      Engine 5811.000000
                                   5847.000000
      count
                3013.181461
                                   2013.448435
                                                              5.841013e+04
                                                                                      18.158496
                                                                                                       1631.552573
                1736.398890
                                       3.194949
                                                              9.237971e+04
                                                                                       4.358246
                                                                                                        601.972587
                                   1998.000000
                                                              1.710000e+02
3.346750e+04
     min
25%
                                                                                                      72.000000
1198.000000
                    1.000000
                                                                                       0.000000
                1509.500000
                                   2012.000000
                                                                                      15.260000
                3015.000000
                                   2014.000000
      75%
                4517.500000
                                   2016.000000
                                                               7.249050e+04
                                                                                      21.100000
                                                                                                       1991.000000
                                   2019.000000
     max
                6018.000000
                                                              6.500000e+06
                                                                                      28.400000
                                                                                                      5998,000000
                                                        New_Price
                5811.000000
                                   5809.000000
                                                       815.000000
                                                                        5847.000000
      mean
                 113.803144
                                        5.286452
                                                        20.484564
                                                                             9.653742
                                                        20.248764
      std
                   53.896719
                                        0.806668
                                                                            11.275966
                   34.200000
                                        2.000000
                                                         1.000000
7.880000
                                                                             0.440000
     min
25%
                   78.000000
                                        5.000000
                                                                             3.550000
     50%
                   98.600000
                                        5.000000
                                                        11.480000
      75%
                  139.010000
                                        5.000000
                                                        24.015000
                                                                            10.250000
                                                        99.920000
                  560.000000
                                      10.000000
                                                                          160.000000
     max
```

c). Change the categorical variables ("Fuel\_Type" and "Transmission") into numerical one hotencoded value.





**d).** Create one more feature and add this column to the dataset. you can calculate the current age of the car by subtracting "Year" value from the current year.



Price	Current_Age
12.50	9
4.50	13
6.00	12
17.74	11
3.50	11
4.75	10
4.00	9
2.90	12
2.65	11
2.50	13

e). Perform select, filter, rename, mutate, arrange and summarize with group by operations (or their equivalent operations in python) on this dataset.

## **Select:**

```
# Display the results
print("Selected Columns:")
print(selected_columns.head())
Selected Columns:
                                                   Year
                                                          Kilometers_Driven
                                Name
                                         Location
  Hyundai Creta 1.6 CRDi SX Option
                                             Pune
                                                    2015
                                                                       41000
                                          Chennai
                                                    2011
                                                                       46000
                        Honda Jazz V
                   Maruti Ertiga VDI
                                          Chennai
                                                    2012
                                                                       87000
3
    Audi A4 New 2.0 TDI Multitronic
                                       Coimbatore
                                                    2013
                                                                       40670
             Nissan Micra Diesel XV
                                           Jaipur
                                                    2013
                                                                       86999
              Transmission Owner_Type Mileage
1 First 19.67
                                                            Power
                                                                           Price
   Fuel_Type
                                                   Engine
                                                                    Seats
0
                                                   1582.0
                                                           126.20
                                                                      5.0
                                                                           12.50
                                  First
                                                                           4.50
1
                                           13.00
                                                   1199.0
                                                            88.70
                                                                      5.0
           1
           0
                          1
                                  First
                                           20.77
                                                   1248.0
                                                            88.76
                                                                      7.0
                                                                            6.00
3
                                           15.20
                                                   1968.0
                                                           140.80
                                                                      5.0
                                                                           17.74
           0
                          0
                                 Second
           0
                                                                      5.0
                                                                            3.50
                                                            63.10
                                  First
                                           23.08
                                                  1461.0
```

#### Filter:

```
[33] print("\nFiltered Data (Petrol only):")
print(filtered_data.head())

Filtered Data (Petrol only):
Empty DataFrame
Columns: [Name, Location, Year, Kilometers_Driven, Fuel_Type, Transmission, Owner_Type, Mileage, Engine, Power, Seats, Price, Current_Age]
Index: []
```

#### Rename:

```
print("\nRenamed Columns:")
    print(renamed_columns.head())
\square
    Renamed Columns:
                                             Location Year
                                                             Kilometers_Driven
       Hyundai Creta 1.6 CRDi SX Option
                                                 Pune
                                                       2015
                                                                          41000
                            Honda Jazz V
                                             Chennai
                                                       2011
                                                                          46000
                      Maruti Ertiga VDI
                                             Chennai 2012
                                                                          87000
    3
        Audi A4 New 2.0 TDI Multitronic
                                                       2013
                                           Coimbatore
                                                                          40670
                 Nissan Micra Diesel XV
                                                       2013
                                                                          86999
                                               Jaipur
                                                  Engine
       Fuel_Type
                  TransType OwnerType Mileage
                                                           Power
                                                                   Seats
                                                                          Price
                                 First
                                           19.67
                                                  1582.0
    0
                                                          126.20
                                                                     5.0
                                                                          12.50
                                 First
                                                  1199.0
                                                                           4.50
                                           13.00
                                                           88.70
                                                                     5.0
    2
3
4
                                                                     7.0
               0
                                 First
                                           20.77
                                                  1248.0
                                                           88.76
                                                                           6.00
                                           15.20
                                                  1968.0
                                                                     5.0
                                                                          17.74
               0
                           0
                                Second
                                                          140.80
                                                                     5.0
                                                                           3.50
               0
                                 First
                                           23.08
                                                  1461.0
                                                           63.10
       Current_Age
9
    0
                13
    1
    2
                12
                11
```

#### Mutate:

```
[35] print("\nMutated Data (Added Power_per_Cylinder column):")
     print(df_cars[['Power', 'Seats', 'Power_per_Cylinder']].head())
    Mutated Data (Added Power_per_Cylinder column):
               Seats Power_per_Cylinder
        Power
       126.20
                  5.0
                                    25.24
                  5.0
                                    17.74
        88.70
        88.76
                  7.0
                                    12.68
                                    28.16
       140.80
                  5.0
                  5.0
        63.10
                                    12.62
```

# **Sort/Arrange:**

```
print("\nSorted Data (Descending order of Year):")
print(sorted_data.head())
Sorted Data (Descending order of Year):
                                                         Year
                                                                Kilometers_Driven \
                                              Location
                                     Name
             Renault KWID RXT Optional
                                                         2019
2019
5405
                                                  Kochi
                                                                               6568
      Ford Endeavour 2.2 Trend AT 4X2 Kochi
Maruti Omni 5 Seater BSIV Coimbatore
942
                                                 Kochi
                                                                              11209
                                                         2019
2019
5533
                                                                              4721
                 Mahindra XUV500 W9 AT Coimbatore
                                                                              19654
770
4267
            Maruti Swift Dzire AMT ZDI
                                               Chennai
                                                          2019
                                                                              65000
      Fuel_Type Transmission Owner_Type
1 1 First
                                               Mileage
                                                                           Seats
                                                 25.17
12.62
5405
                                                           799.0
                                                                   53.3
                                                                             5.0
                                                                   158.0
942
                                       First
                                                          2198.0
                                                                             7.0
5533
                                                          796.0
                                                                   35.0
                                                                             5.0
                                       First
                                                  14.00
                               0
                                                  14.00
                                                          2179.0
770
                0
                                       First
                                                                  155.0
                                                                             7.0
4267
                                                                             5.0
                                       First
                                                  26.59
                                                          1248.0
              Current_Age
                             Power_per_Cylinder
10.660000
      Price
       5.09
5405
                                       22.571429
942
      31.15
5533
                                        7.000000
       4.11
                                       22.142857
770
      17.63
                          5
4267
       6.75
                                       14.800000
```

## Group By:

```
print("\nGrouped Data (Mean Price and Total Kilometers_Driven per Location):")
print(grouped_data.head())
Grouped Data (Mean Price and Total Kilometers_Driven per Location):
                Price Kilometers_Driven
Location
Ahmedabad
             8.567248
                                11984485
Bangalore
            13.482670
                                 20496207
Chennai
             7.958340
                                43075555
            15.160206
                                 29625311
Coimbatore
Delhi
             9.881944
                                 30880339
```

## **Conclusion:**

The dataset consists of attributes of used cars, with the target variable being the price measured in lakhs. Missing values were addressed by imputing them or dropping them based on data integrity. Units were removed from specific attributes, and categorical variables were transformed into numerical one-hot encoded values. Additionally, a new feature indicating the current age of the car was created. Various data manipulation operations were performed for analysis and summary.