PDS ASSIGNMENT - 2

<div align="right">

Jithendra Pavuluri

16343746

</div>

Task A:

Identifying Missing Values:

missing_vals = data.isnull().sum() computes the total number of missing values in each column.

```
Unnamed: 0              0
Name                    0
Location                0
Year                    0
Kilometers_Driven       0
Fuel_Type               0
Transmission            0
Owner_Type              0
Mileage                 2
Engine                 36
Power                  36
Seats                  38
New_Price            5032
Price                   0
dtype: int64
```

missing_vars = list(missing_vals[missing_vals > 0].index) identifies columns with missing values (columns where the sum of missing values is greater than 0).

Output:

['Mileage', 'Engine', 'Power', 'Seats', 'New_Price']

Removing Columns with More than 50% Missing Values:

data = data.dropna(thresh=0.5, axis=1) drops columns where more than 50% of the values are missing. This step helps in eliminating columns with insufficient data, which might not contribute meaningfully to the analysis.

Imputing Missing Values with Median: impute_vals = data.median(axis=0) calculates the median for each column.data = data.fillna(impute_vals) fills in missing values in the remaining columns with their respective column medians. Imputing missing values with the median helps maintain the central tendency of the data and reduces the impact of outliers.

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	NaN	12.50
1	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	13 km/kg	1199 CC	88.7 bhp	5.0	8.61 Lakh	4.50
2	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77 kmpl	1248 CC	88.76 bhp	7.0	NaN	6.00
3	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	140.8 bhp	5.0	NaN	17.74
4	6	Nissan Micra Diesel XV	Jaipur	2013	86999	Diesel	Manual	First	23.08 kmpl	1461 CC	63.1 bhp	5.0	NaN	3.50
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
5842	6014	Maruti Swift VDI	Delhi	2014	27365	Diesel	Manual	First	28.4 kmpl	1248 CC	74 bhp	5.0	7.88 Lakh	4.75

- Task A contributes to data quality and prepares the dataset for further analysis, ensuring more accurate and reliable results.

Task B:

Clean Specific Columns: Remove non-numeric characters from specified columns.

Convert Columns to Numeric: Convert the cleaned columns to numeric types.

Store Numerical Data: Store the cleaned numerical data in a new CSV file.

]		Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	1		Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67	1582.0	126.20	5.0	NaN	12.50
1	2		Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	13.00	1199.0	88.70	5.0	8.61	4.50
2	3		Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77	1248.0	88.76	7.0	NaN	6.00
3	4		Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.20	1968.0	140.80	5.0	NaN	17.74
4	6		Nissan Micra Diesel XV	Jaipur	2013	86999	Diesel	Manual	First	23.08	1461.0	63.10	5.0	NaN	3.50
...	...		...	...	...	...	...	...	...	...	...	...	...	...	...
5842	6014		Maruti Swift VDI	Delhi	2014	27365	Diesel	Manual	First	28.40	1248.0	74.00	5.0	7.88	4.75

Task C:

One-Hot Encoding: Performing one-hot encoding for categorical columns (Fuel_Type and Transmission).

Store Encoded Data: Storing the encoded data into a new CSV file.

Price	Fuel_Type_Diesel	Fuel_Type_Electric	Fuel_Type_Petrol	Transmission_Automatic	Transmission_Manual
12.50	1	0	0	0	1
4.50	0	0	1	0	1
6.00	1	0	0	0	1
17.74	1	0	0	1	0
3.50	1	0	0	0	1
...	...	...	...	...	...

Task D:

Calculate Car Age: Calculates the age of the car using the current year and the year of manufacture.

Store Car Age Data: Stores the data with calculated car age into a new CSV file.
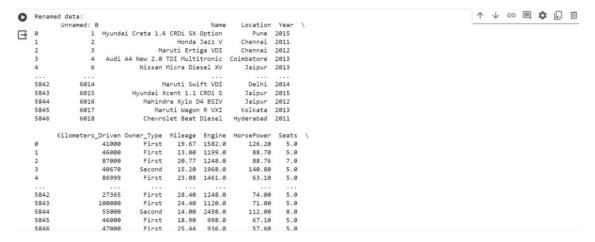
ileage	Engine	Power	Seats	New_Price	Price	Fuel_Type_Diesel	Fuel_Type_Electric	Fuel_Type_Petrol	Transmission_Automatic	Transmission_Manual	Age
19.67	1582.0	126.20	5.0	NaN	12.50	1	0	0	0	1	9
13.00	1199.0	88.70	5.0	8.61	4.50	0	0	1	0	1	13
20.77	1248.0	88.76	7.0	NaN	6.00	1	0	0	0	1	12
15.20	1968.0	140.80	5.0	NaN	17.74	1	0	0	1	0	11
23.08	1461.0	63.10	5.0	NaN	3.50	1	0	0	0	1	11
...	...	...	...	...	...	...	...	...	...	...	...
28.40	1248.0	74.00	5.0	7.88	4.75	1	0	0	0	1	10

Task E:

Select Data: Selects specific columns from the dataset.

```
Select data:
      Year  Kilometers_Driven Owner_Type  New_Price
0     2015              41000      First        NaN
1     2011              46000      First       8.61
2     2012              87000      First        NaN
3     2013              40670     Second        NaN
4     2013              86999      First        NaN
...    ...                ...        ...        ...
5842  2014              27365      First       7.88
5843  2015             100000      First        NaN
5844  2012              55000     Second        NaN
5845  2013              46000      First        NaN
5846  2011              47000      First        NaN

[5847 rows x 4 columns]
```

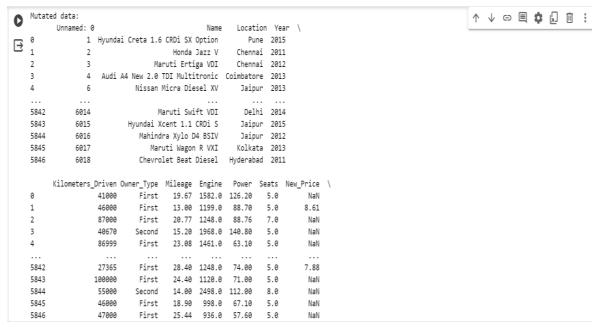**Filter Data**: Filters rows based on a condition.

```
Filtered data:
      Unnamed: 0                           Name    Location  Year  \
0              1  Hyundai Creta 1.6 CRDi SX Option      Pune  2015
1              2                    Honda Jazz V   Chennai  2011
2              3               Maruti Ertiga VDI   Chennai  2012
3              4    Audi A4 New 2.0 TDI Multitronic Coimbatore 2013
4              6             Nissan Micra Diesel XV    Jaipur  2013
...          ...                          ...        ...   ...
5842        6014              Maruti Swift VDI     Delhi  2014
5843        6015          Hyundai Xcent 1.1 CRDi S   Jaipur  2015
5844        6016          Mahindra Xylo D4 BSIV    Jaipur  2012
5845        6017             Maruti Wagon R VXI    Kolkata  2013
5846        6018           Chevrolet Beat Diesel  Hyderabad  2011

      Kilometers_Driven Owner_Type  Mileage  Engine   Power  Seats  New_Price  \
0                 41000      First    19.67  1582.0  126.20    5.0        NaN
1                 46000      First    13.00  1199.0   88.70    5.0       8.61
2                 87000      First    20.77  1248.0   88.76    7.0        NaN
3                 40670     Second    15.20  1968.0  140.80    5.0        NaN
4                 86999      First    23.08  1461.0   63.10    5.0        NaN
...                 ...        ...      ...     ...     ...    ...        ...
5842              27365      First    28.40  1248.0   74.00    5.0       7.88
5843             100000      First    24.40  1120.0   71.00    5.0        NaN
5844              55000     Second    14.00  2498.0  112.00    8.0        NaN
5845              46000      First    18.90   998.0   67.10    5.0        NaN
5846              47000      First    25.44   936.0   57.60    5.0        NaN
```

Rename Columns: Renames specific columns.

```
Renamed data:
      Unnamed: 0                           Name     Location  Year  \
0              1  Hyundai Creta 1.6 CRDi SX Option       Pune  2015
1              2                    Honda Jazz V    Chennai  2011
2              3               Maruti Ertiga VDI    Chennai  2012
3              4   Audi A4 New 2.0 TDI Multitronic  Coimbatore  2013
4              6            Nissan Micra Diesel XV     Jaipur  2013
...          ...                             ...        ...   ...
5842        6014                 Maruti Swift VDI      Delhi  2014
5843        6015           Hyundai Xcent 1.1 CRDi S    Jaipur  2015
5844        6016            Mahindra Xylo D4 BSIV     Jaipur  2012
5845        6017              Maruti Wagon R VXI     Kolkata  2013
5846        6018           Chevrolet Beat Diesel   Hyderabad  2011

      Kilometers_Driven Owner_Type  Mileage  Engine  HorsePower  Seats  \
0                 41000      First    19.67  1582.0      126.20    5.0
1                 46000      First    13.00  1199.0       88.70    5.0
2                 87000      First    20.77  1248.0       88.76    7.0
3                 40670     Second    15.20  1968.0      140.80    5.0
4                 86999      First    23.08  1461.0       63.10    5.0
...                 ...        ...      ...     ...         ...    ...
5842              27365      First    28.40  1248.0       74.00    5.0
5843             100000      First    24.40  1120.0       71.00    5.0
5844              55000     Second    14.00  2498.0      112.00    8.0
5845              46000      First    18.90   998.0       67.10    5.0
5846              47000      First    25.44   936.0       57.60    5.0
```

Mutate Data: Creates a new column by performing a computation on existing columns.

```
Mutated data:
      Unnamed: 0                           Name     Location  Year  \
0              1  Hyundai Creta 1.6 CRDi SX Option       Pune  2015
1              2                    Honda Jazz V    Chennai  2011
2              3               Maruti Ertiga VDI    Chennai  2012
3              4   Audi A4 New 2.0 TDI Multitronic  Coimbatore  2013
4              6            Nissan Micra Diesel XV     Jaipur  2013
...          ...                             ...        ...   ...
5842        6014                 Maruti Swift VDI      Delhi  2014
5843        6015           Hyundai Xcent 1.1 CRDi S    Jaipur  2015
5844        6016            Mahindra Xylo D4 BSIV     Jaipur  2012
5845        6017              Maruti Wagon R VXI     Kolkata  2013
5846        6018           Chevrolet Beat Diesel   Hyderabad  2011

      Kilometers_Driven Owner_Type  Mileage  Engine   Power  Seats  New_Price  \
0                 41000      First    19.67  1582.0  126.20    5.0        NaN
1                 46000      First    13.00  1199.0   88.70    5.0       8.61
2                 87000      First    20.77  1248.0   88.76    7.0        NaN
3                 40670     Second    15.20  1968.0  140.80    5.0        NaN
4                 86999      First    23.08  1461.0   63.10    5.0        NaN
...                 ...        ...      ...     ...     ...    ...        ...
5842              27365      First    28.40  1248.0   74.00    5.0       7.88
5843             100000      First    24.40  1120.0   71.00    5.0        NaN
5844              55000     Second    14.00  2498.0  112.00    8.0        NaN
5845              46000      First    18.90   998.0   67.10    5.0        NaN
5846              47000      First    25.44   936.0   57.60    5.0        NaN
```

Arrange Data: Sort the data based on a specific column.

```
Arranged data:
       Unnamed: 0                              Name    Location  Year  \
1315         1361          Maruti Alto 800 2016-2019 VXI    Mumbai  2019
5451         5606             Maruti Wagon R ZXI AMT 1.2    Mumbai  2019
1203         1242          Jaguar XE 2.0L Diesel Prestige     Delhi  2018
165           173       Hyundai Grand i10 1.2 Kappa Asta   Kolkata  2019
1123         1161           Tata Tigor 1.2 Revotron XTA  Ahmedabad  2018
...           ...                                   ...       ...   ...
2997         3092                  Honda City i VTEC SV   Kolkata  2015
344           358                  Hyundai i10 Magna 1.2   Chennai  2009
1805         1860        Volkswagen Vento Diesel Highline   Chennai  2013
326           340  Skoda Octavia Ambition Plus 2.0 TDI AT   Kolkata  2013
2259         2328              BMW X5 xDrive 30d M Sport   Chennai  2017

      Kilometers_Driven Owner_Type  Mileage  Engine   Power  Seats  New_Price  \
1315                171      First    24.70   796.0   47.30    5.0        NaN
5451                600      First    21.50  1197.0   81.80    5.0       6.80
1203               1000      First    13.60  1999.0  177.00    5.0      52.77
165                1000      First    18.90  1197.0   81.86    5.0       7.39
1123               1000      First    20.30  1199.0   84.00    5.0        NaN
...                 ...        ...      ...     ...     ...    ...        ...
2997             480000      First    17.40  1497.0  117.30    5.0        NaN
344              620000      First    20.36  1197.0   78.90    5.0        NaN
1805             720000      First    20.54  1598.0  103.60    5.0        NaN
326              775000      First    19.30  1968.0  141.00    5.0        NaN
2259            6500000      First    15.97  2993.0  258.00    5.0        NaN
```

Summarize Data: Summarize data using the group by function.

```
Summarized data:
               Mileage  Kilometers_Driven
Owner_Type
First            18.6        55193.302224
Fourth & Above   13.3       124354.125000
Second           17.0        71175.587027
Third            16.1        88901.252427
```