

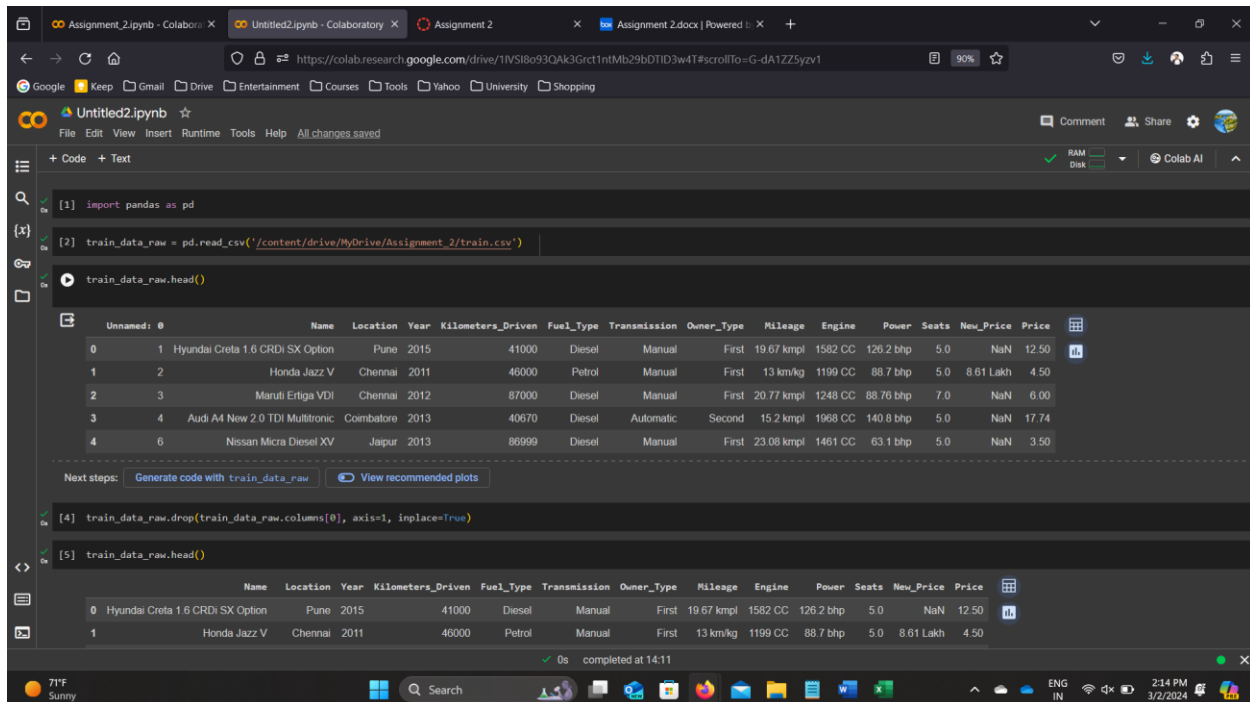
Principles of Data science - Assignment – 2

Report on results.

Name: Prakash Raj Baskar Raj

Student ID: 16353058

I view the provided data set as the unprocessed data. I'm removing the first column, which is the serial number-containing nameless column. The reason I am deleting this column is that the serial numbers do not match.



The screenshot shows a Google Colab notebook with the following code and output:

```
[1] import pandas as pd
```

```
[2] train_data_raw = pd.read_csv('/content/drive/MyDrive/Assignment_2/train.csv')
```

```
train_data_raw.head()
```

	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	88.70 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

Next steps: [Generate code with train_data_raw](#) [View recommended plots](#)

```
[4] train_data_raw.drop(train_data_raw.columns[0], axis=1, inplace=True)
```

```
[5] train_data_raw.head()
```

	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	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	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	13 km/kg	1199 CC	88.7 bhp	5.0	8.61 Lakh	4.50

- I'm trying to find the values that are missing from each column in this query. I am using the mode to impute missing values for the categorical columns since doing so allows you to keep the data that is currently in that category field. This lessens the possibility of bias by ensuring that the imputed values are reflective of the entire set of data.

I am using the mean to impute missing values for the numerical columns because doing so preserves the distribution and central tendency of the available data in the numerical column. By using this method, the possibility of introducing errors is decreased and the imputed values are guaranteed to match the distribution of the whole data set.

Assignment_2.ipynb - Colaboratory X Untitled2.ipynb - Colaboratory X Assignment 2 X Assignment 2.docx | Powered by X +

https://colab.research.google.com/drive/1IVSI8o93QAK3Grct1ntMb29bDTID3w4T#scrollTo=OE21SO7Pyg-D 90%

Google Keep Gmail Drive Entertainment Courses Tools Yahoo University Shopping

Untitled2.ipynb File Edit View Insert Runtime Tools Help All changes saved Comment Share Settings

+ Code + Text

Generate randomly select 5 items from a list Close

Generate is available for a limited time for unsubscribed users. Upgrade to Colab Pro

```
missing_values = train_data_raw.isna().sum()
```

```
[7] for column in missing_values.index:
    if missing_values[column] > 0:
        if train_data_raw[column].dtype == 'object':
            mode_value = train_data_raw[column].mode()[0]
            train_data_raw[column].fillna(mode_value, inplace=True)
        else:
            mean_value = train_data_raw[column].mean()
            train_data_raw[column].fillna(mean_value, inplace=True)
```

```
train_data_raw.head(30)
```

	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owmer_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	4.78 Lakh	12.50
1	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	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77 kmpl	1248 CC	88.76 bhp	7.0	4.78 Lakh	6.00
3	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	140.8 bhp	5.0	4.78 Lakh	17.74
4	Nissan Micra Diesel XV	Jaipur	2013	86999	Diesel	Manual	First	23.08 kmpl	1461 CC	63.1 bhp	5.0	4.78 Lakh	3.50
5	Toyota Innova Crysta 2.8 GX AT 8S	Mumbai	2016	36000	Diesel	Automatic	First	11.36 kmpl	2755 CC	171.5 bhp	8.0	21 Lakh	17.50
6	Volkswagen Vento Diesel Comfortline	Pune	2013	64430	Diesel	Manual	First	20.54 kmpl	1598 CC	103.6 bhp	5.0	4.78 Lakh	5.20

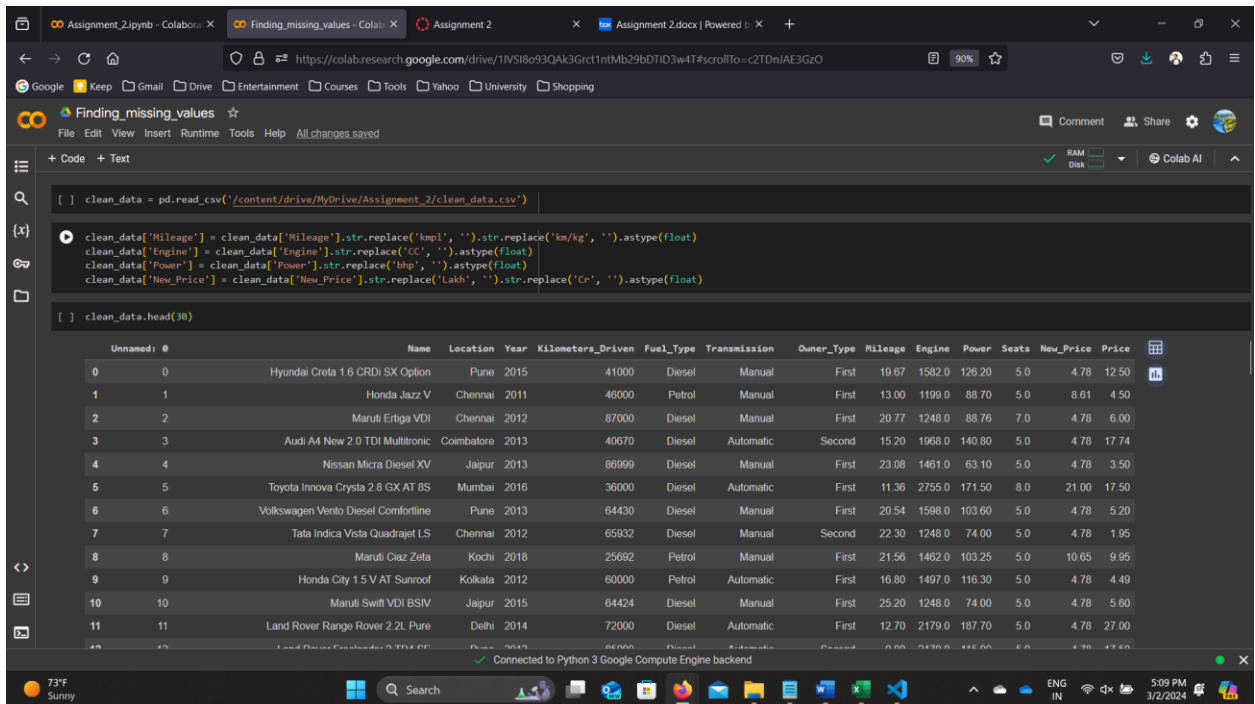
0s completed at 14:17

71°F Sunny Search

2:18 PM 3/2/2024

	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owmer_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	4.78 Lakh	12.50
1	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	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77 kmpl	1248 CC	88.76 bhp	7.0	4.78 Lakh	6.00
3	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	140.8 bhp	5.0	4.78 Lakh	17.74
4	Nissan Micra Diesel XV	Jaipur	2013	86999	Diesel	Manual	First	23.08 kmpl	1461 CC	63.1 bhp	5.0	4.78 Lakh	3.50
5	Toyota Innova Crysta 2.8 GX AT 8S	Mumbai	2016	36000	Diesel	Automatic	First	11.36 kmpl	2755 CC	171.5 bhp	8.0	21 Lakh	17.50
6	Volkswagen Vento Diesel Comfortline	Pune	2013	64430	Diesel	Manual	First	20.54 kmpl	1598 CC	103.6 bhp	5.0	4.78 Lakh	5.20
7	Tata Indica Vista Quadrajet LS	Chennai	2012	65932	Diesel	Manual	Second	22.3 kmpl	1248 CC	74 bhp	5.0	4.78 Lakh	1.95
8	Maruti Ciaz Zeta	Kochi	2018	25692	Petrol	Manual	First	21.56 kmpl	1462 CC	103.25 bhp	5.0	10.65 Lakh	9.95
9	Honda City 1.5 V AT Sunroof	Kolkata	2012	60000	Petrol	Automatic	First	16.8 kmpl	1497 CC	116.3 bhp	5.0	4.78 Lakh	4.49
10	Maruti Swift VDI BSIV	Jaipur	2015	64424	Diesel	Manual	First	25.2 kmpl	1248 CC	74 bhp	5.0	4.78 Lakh	5.60
11	Land Rover Range Rover 2.2 L Pure	Delhi	2014	72000	Diesel	Automatic	First	12.7 kmpl	2179 CC	187.7 bhp	5.0	4.78 Lakh	27.00
12	Land Rover Freelander 2 TD4 SE	Pune	2012	85000	Diesel	Automatic	Second	0.0 kmpl	2179 CC	115 bhp	5.0	4.78 Lakh	17.50
13	Mitsubishi Pajero Sport 4X4	Delhi	2014	110000	Diesel	Manual	First	13.5 kmpl	2477 CC	175.56 bhp	7.0	32.01 Lakh	15.00
14	Honda Amaze S i-Dtech	Kochi	2016	58950	Diesel	Manual	First	25.8 kmpl	1498 CC	98.6 bhp	5.0	4.78 Lakh	5.40
15	Maruti Swift DDIS VDI	Jaipur	2017	25000	Diesel	Manual	First	28.4 kmpl	1248 CC	74 bhp	5.0	4.78 Lakh	5.99
16	Renault Duster 85PS Diesel RxL Plus	Kochi	2014	77469	Diesel	Manual	First	20.45 kmpl	1461 CC	83.8 bhp	5.0	4.78 Lakh	6.34
17	Mercedes-Benz New C-Class C 220 CDI BE Avantgarde	Bangalore	2014	78500	Diesel	Automatic	First	14.84 kmpl	2143 CC	167.62 bhp	5.0	4.78 Lakh	28.00
18	BMW 3 Series 320d	Kochi	2014	32982	Diesel	Automatic	First	22.69 kmpl	1995 CC	190 bhp	5.0	47.87 Lakh	18.55
19	Maruti S Cross DDIS 200 Alpha	Bangalore	2015	55392	Diesel	Manual	Second	23.65 kmpl	1248 CC	88.5 bhp	5.0	4.78 Lakh	8.25
20	Audi A6 2011-2015 35 TFSI Technology	Mumbai	2015	55985	Petrol	Automatic	First	13.53 kmpl	1984 CC	177.01 bhp	5.0	4.78 Lakh	23.50
21	Hyundai i20 1.2 Magna	Kolkata	2010	45807	Petrol	Manual	First	18.5 kmpl	1197 CC	80 bhp	5.0	4.78 Lakh	1.87
22	Volkswagen Vento Petrol Highline AT	Kolkata	2010	33000	Petrol	Automatic	First	14.4 kmpl	1598 CC	103.6 bhp	5.0	4.78 Lakh	2.85
23	Honda City Corporate Edition	Mumbai	2012	51920	Petrol	Manual	First	16.8 kmpl	1497 CC	116.3 bhp	5.0	4.78 Lakh	4.25
24	Nissan Micra Diesel XV	Hyderabad	2012	54000	Diesel	Manual	First	16.48 km/kg	1461 CC	63.1 bhp	5.0	4.78 Lakh	4.25
25	Maruti Alto K10 2010-2014 VXI	Hyderabad	2013	54000	Petrol	Manual	Second	20.92 kmpl	998 CC	67.1 bhp	5.0	4.78 Lakh	2.75
26	Honda WRV i-VTEC VX	Kochi	2018	37430	Petrol	Manual	First	17.5 kmpl	1199 CC	88.7 bhp	5.0	10.57 Lakh	9.90
27	Toyota Innova 2.5 V Diesel 7-seater	Mumbai	2007	262000	Diesel	Manual	Fourth & Above	12.8 kmpl	2494 CC	102 bhp	7.0	4.78 Lakh	4.00
28	Renault Duster 110PS Diesel RxZ Pack	Bangalore	2013	62000	Diesel	Manual	First	19.01 kmpl	1461 CC	108.45 bhp	5.0	4.78 Lakh	7.75

- b. In this question, I am removing the units from the attributes of Mileage, Engine, Power, New_price columns and keeping only the numerical values and storing this modified data as the cleaned data (data_clean.csv) under data_clean folder.



```
[ ] clean_data = pd.read_csv('/content/drive/MyDrive/Assignment_2/clean_data.csv')

clean_data['Mileage'] = clean_data['Mileage'].str.replace('kmpl', '').str.replace('km/kg', '').astype(float)
clean_data['Engine'] = clean_data['Engine'].str.replace('CC', '').astype(float)
clean_data['Power'] = clean_data['Power'].str.replace('bhp', '').astype(float)
clean_data['New_Price'] = clean_data['New_Price'].str.replace('Lakh', '').str.replace('Cr', '').astype(float)

[ ] clean_data.head(30)
```

Unnamed: 0		Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	0	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67	1582.0	126.20	5.0	4.78	12.50
1	1	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	13.00	1199.0	88.70	5.0	8.61	4.50
2	2	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77	1248.0	88.76	7.0	4.78	6.00
3	3	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.20	1968.0	140.80	5.0	4.78	17.74
4	4	Nissan Micra Diesel XV	Jaipur	2013	86999	Diesel	Manual	First	23.08	1461.0	63.10	5.0	4.78	3.50
5	5	Toyota Innova Crysta 2.8 GX AT 8S	Mumbai	2016	36000	Diesel	Automatic	First	11.36	2755.0	171.50	8.0	21.00	17.50
6	6	Volkswagen Vento Diesel Comfortline	Pune	2013	64430	Diesel	Manual	First	20.54	1598.0	103.60	5.0	4.78	5.20
7	7	Tata Indica Vista Quadrajet LS	Chennai	2012	65932	Diesel	Manual	Second	22.30	1248.0	74.00	5.0	4.78	1.95
8	8	Maruti Ciaz Zeta	Kochi	2018	25692	Petrol	Manual	First	21.56	1462.0	103.25	5.0	10.65	9.95
9	9	Honda City 1.5 V AT Sunroof	Kolkata	2012	60000	Petrol	Automatic	First	16.80	1497.0	116.30	5.0	4.78	4.49
10	10	Maruti Swift VDI BSIV	Jaipur	2015	64424	Diesel	Manual	First	25.20	1248.0	74.00	5.0	4.78	5.60
11	11	Land Rover Range Rover 2.2L Pure	Delhi	2014	72000	Diesel	Automatic	First	12.70	2179.0	187.70	5.0	4.78	27.00

Unnamed: 0		Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	0	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67	1582.0	126.20	5.0	4.78	12.50
1	1	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	13.00	1199.0	88.70	5.0	8.61	4.50
2	2	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77	1248.0	88.76	7.0	4.78	6.00
3	3	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.20	1968.0	140.80	5.0	4.78	17.74
4	4	Nissan Micra Diesel XV	Jaipur	2013	86999	Diesel	Manual	First	23.08	1461.0	63.10	5.0	4.78	3.50
5	5	Toyota Innova Crysta 2.8 GX AT 8S	Mumbai	2016	36000	Diesel	Automatic	First	11.36	2755.0	171.50	8.0	21.00	17.50
6	6	Volkswagen Vento Diesel Comfortline	Pune	2013	64430	Diesel	Manual	First	20.54	1598.0	103.60	5.0	4.78	5.20
7	7	Tata Indica Vista Quadrajet LS	Chennai	2012	65932	Diesel	Manual	Second	22.30	1248.0	74.00	5.0	4.78	1.95
8	8	Maruti Ciaz Zeta	Kochi	2018	25692	Petrol	Manual	First	21.56	1462.0	103.25	5.0	10.65	9.95
9	9	Honda City 1.5 V AT Sunroof	Kolkata	2012	60000	Petrol	Automatic	First	16.80	1497.0	116.30	5.0	4.78	4.49
10	10	Maruti Swift VDI BSIV	Jaipur	2015	64424	Diesel	Manual	First	25.20	1248.0	74.00	5.0	4.78	5.60
11	11	Land Rover Range Rover 2.2L Pure	Delhi	2014	72000	Diesel	Automatic	First	12.70	2179.0	187.70	5.0	4.78	27.00
12	12	Land Rover Freelander 2 TD4 SE	Pune	2012	85000	Diesel	Automatic	Second	0.00	2179.0	115.00	5.0	4.78	17.50
13	13	Mitsubishi Pajero Sport 4X4	Delhi	2014	110000	Diesel	Manual	First	13.50	2477.0	175.56	7.0	32.01	15.00
14	14	Honda Amaze S i Dtech	Kochi	2016	58950	Diesel	Manual	First	25.80	1498.0	98.60	5.0	4.78	5.40
15	15	Maruti Swift DDIS VDI	Jaipur	2017	25000	Diesel	Manual	First	28.40	1248.0	74.00	5.0	4.78	5.99
16	16	Renault Duster 85PS Diesel RxL Plus	Kochi	2014	77469	Diesel	Manual	First	20.45	1461.0	83.80	5.0	4.78	6.34
17	17	Mercedes-Benz New C-Class C 220 CDI BE Avantgarde	Bangalore	2014	78500	Diesel	Automatic	First	14.84	2143.0	167.62	5.0	4.78	28.00
18	18	BMW 3 Series 320d	Kochi	2014	32982	Diesel	Automatic	First	22.69	1995.0	190.00	5.0	47.87	18.55
19	19	Maruti S Cross DDIS 200 Alpha	Bangalore	2015	55392	Diesel	Manual	Second	23.65	1248.0	88.50	5.0	4.78	8.25
20	20	Audi A6 2011-2015 35 TFSI Technology	Mumbai	2015	55985	Petrol	Automatic	First	13.53	1984.0	177.01	5.0	4.78	23.50
21	21	Hyundai i20 1.2 Magna	Kolkata	2010	45807	Petrol	Manual	First	18.50	1197.0	80.00	5.0	4.78	1.87
22	22	Volkswagen Vento Petrol Highline AT	Kolkata	2010	33000	Petrol	Automatic	First	14.40	1598.0	103.60	5.0	4.78	2.85
23	23	Honda City Corporate Edition	Mumbai	2012	51920	Petrol	Manual	First	16.80	1497.0	116.30	5.0	4.78	4.25
24	24	Nissan Micra Diesel XV	Hyderabad	2012	54000	Diesel	Manual	First	16.48	1461.0	63.10	5.0	4.78	4.25
25	25	Maruti Alto K10 2010-2014 VXI	Hyderabad	2013	54000	Petrol	Manual	Second	20.92	998.0	67.10	5.0	4.78	2.75
26	26	Honda WRV i-VTEC VX	Kochi	2018	37430	Petrol	Manual	First	17.50	1199.0	88.70	5.0	10.57	9.90
27	27	Toyota Innova 2.5 V Diesel 7-seater	Mumbai	2007	262000	Diesel	Manual	Fourth & Above	12.80	2494.0	102.00	7.0	4.78	4.00
28	28	Renault Duster 110PS Diesel RxZ Pack	Bangalore	2013	62000	Diesel	Manual	First	19.01	1461.0	108.45	5.0	4.78	7.75

- c. We are loading the clean data and encoding the categorical variables “Fuel_Type” and “Transmission” into numerical values and storing this modified data in data_encoded.csv under the results folder.

The screenshot shows a Google Colab notebook titled "Finding_missing_values". The code in the cell is as follows:

```
processed_data = pd.get_dummies(clean_data, columns=['Fuel_Type', 'Transmission'])  
  
processed_data.to_csv('/content/drive/MyDrive/Assignment_2/clean_data.csv', index=False)  
  
processed_data.head(30)
```

The output of the code is a preview of the DataFrame. The columns are: Unnamed: 0, Name, Location, Year, Kilometers_Driven, Owner_Type, Mileage, Engine, Power, Seats, New_Price, Price, Fuel_Type_Diesel, Fuel_Type_Electric, Fuel_Type_Petrol, Transmission_Automatic, and Transmission_Manual. The first 6 rows of data are shown:

Unnamed: 0	Name	Location	Year	Kilometers_Driven	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price	Fuel_Type_Diesel	Fuel_Type_Electric	Fuel_Type_Petrol	Transmission_Automatic	Transmission_Manual
0	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	First	19.67	1582.0	126.20	5.0	4.78	12.50	1	0	0	0	
1	Honda Jazz V	Chennai	2011	46000	First	13.00	1199.0	88.70	5.0	8.61	4.50	0	0	1	0	
2	Maruti Ertiga VDI	Chennai	2012	87000	First	20.77	1248.0	88.76	7.0	4.78	6.00	1	0	0	0	
3	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Second	15.20	1968.0	140.80	5.0	4.78	17.74	1	0	0		1
4	Nissan Micra Diesel XV	Jaipur	2013	86999	First	23.08	1461.0	63.10	5.0	4.78	3.50	1	0	0	0	
5	Toyota Innova Crysta 2.8	Mumbai	2016	36000	First	11.36	2755.0	171.50	8.0	21.00	17.50	1	0	0		1

Unnamed: 0	Name	Location	Year	Kilometers_Driven	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price	Fuel_Type_Diesel	Fuel_Type_Electric	Fuel_Type_Petrol	Transmission_Automatic	Transmission_Manual
0	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	First	19.67	1582.0	126.20	5.0	4.78	12.50	1	0	0	0	
1	Honda Jazz V	Chennai	2011	46000	First	13.00	1199.0	88.70	5.0	8.61	4.50	0	0	1	0	
2	Maruti Ertiga VDI	Chennai	2012	87000	First	20.77	1248.0	88.76	7.0	4.78	6.00	1	0	0	0	
3	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Second	15.20	1968.0	140.80	5.0	4.78	17.74	1	0	0		1
4	Nissan Micra Diesel XV	Jaipur	2013	86999	First	23.08	1461.0	63.10	5.0	4.78	3.50	1	0	0	0	
5	Toyota Innova Crysta 2.8 GX AT BS	Mumbai	2016	36000	First	11.36	2755.0	171.50	8.0	21.00	17.50	1	0	0		1
6	Volkswagen Vento Diesel Comfortline	Pune	2013	64430	First	20.54	1598.0	103.60	5.0	4.78	5.20	1	0	0	0	
7	Tata Indica Vista Quadrajet LS	Chennai	2012	65932	Second	22.30	1248.0	74.00	5.0	4.78	1.95	1	0	0	0	
8	Maruti Ciaz Zeta	Kochi	2018	25692	First	21.56	1462.0	103.25	5.0	10.65	9.95	0	0	1	0	
9	Honda City 1.5 V AT Sunroof	Kolkata	2012	60000	First	16.80	1497.0	116.30	5.0	4.78	4.49	0	0	1		1
10	Maruti Swift VDI BSIV	Jaipur	2015	64424	First	25.20	1248.0	74.00	5.0	4.78	5.60	1	0	0	0	
11	Land Rover Range Rover 2.2L Pure	Delhi	2014	72000	First	12.70	2179.0	187.70	5.0	4.78	27.00	1	0	0		1
12	Land Rover Freelander 2 TD4 SE	Pune	2012	85000	Second	0.00	2179.0	115.00	5.0	4.78	17.50	1	0	0		1

- d. We will first load the data_encoded.csv data obtained from the previous step. Then I am adding a new column Current_Age and storing this final data as the data_final_results.csv under the results folder.

```
[ ] import datetime

[ ] current_year = datetime.datetime.now().year

[ ] data = pd.read_csv('/content/drive/MyDrive/Assignment 2/clean_data.csv')

[ ] data['Current_Age'] = current_year - data['Year']

[ ] data.head(30)
```

Unnamed: 0	Name	Location	Year	Kilometers_Driven	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price	Fuel_Type_Diesel	Fuel_Type_Electric	Fuel_Type_Petrol	Transmission_Automatic
0	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	First	19.67	1582.0	126.20	5.0	4.78	12.50	1	0	0	0
1	Honda Jazz V	Chennai	2011	46000	First	13.00	1199.0	88.70	5.0	8.61	4.50	0	0	1	0
2	Maruti Ertiga VDI	Chennai	2012	87000	First	20.77	1248.0	88.76	7.0	4.78	6.00	1	0	0	0
3	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Second	15.20	1968.0	140.80	5.0	4.78	17.74	1	0	0	1

Unnamed: 0	Name	Location	Year	Kilometers_Driven	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price	Fuel_Type_Diesel	Fuel_Type_Electric	Fuel_Type_Petrol	Transmission_Automatic	Transmission_Manual
0	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	First	19.67	1582.0	126.20	5.0	4.78	12.50	1	0	0	0	
1	Honda Jazz V	Chennai	2011	46000	First	13.00	1199.0	88.70	5.0	8.61	4.50	0	0	1	0	
2	Maruti Ertiga VDI	Chennai	2012	87000	First	20.77	1248.0	88.76	7.0	4.78	6.00	1	0	0	0	
3	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Second	15.20	1968.0	140.80	5.0	4.78	17.74	1	0	0	1	
4	Nissan Micra Diesel XV	Jaipur	2013	86999	First	23.08	1461.0	63.10	5.0	4.78	3.50	1	0	0	0	
5	Toyota Innova Crysta 2.8 GX AT BS	Mumbai	2016	36000	First	11.36	2755.0	171.50	8.0	21.00	17.50	1	0	0	1	
6	Volkswagen Vento Diesel Comfortline	Pune	2013	64430	First	20.54	1598.0	103.60	5.0	4.78	5.20	1	0	0	0	
7	Tata Indica Vista Quadrigel LS	Chennai	2012	65932	Second	22.30	1248.0	74.00	5.0	4.78	1.95	1	0	0	0	
8	Maruti Ciaz Zeta	Kochi	2018	25692	First	21.56	1462.0	103.25	5.0	10.65	9.95	0	0	1	0	
9	Honda City 1.5 V AT Sunroof	Kolkata	2012	60000	First	16.80	1497.0	116.30	5.0	4.78	4.49	0	0	1	1	
10	Maruti Swift VDI BSIV	Jaipur	2015	64424	First	25.20	1248.0	74.00	5.0	4.78	5.60	1	0	0	0	
11	Land Rover Range Rover 2.2L Pure	Delhi	2014	72000	First	12.70	2179.0	187.70	5.0	4.78	27.00	1	0	0	1	
12	Land Rover Freelander 2 TD4 SE	Pune	2012	85000	Second	0.00	2179.0	115.00	5.0	4.78	17.50	1	0	0	1	

- e. In the final step, we summarized the data by grouping it based on the unique 'Location' values. Using the groupby() function, we aggregated the 'Power' column by summing its values for each group. This produced a new DataFrame summarized_df, showing the total 'Power' for each location. These operations, along with others like selecting specific columns and filtering rows, provide valuable insights into the dataset's distribution and characteristics.

```

import pandas as pd
df = pd.read_csv('/content/drive/MyDrive/Assignment_2/clean_data.csv')
selected_df = df[['Location', 'Mileage']]
filtered_df = df[df['Mileage'] > 20]
renamed_df = df.rename(columns={'Mileage': 'new_mileage', 'Power': 'new_power'})
df['mileage_power_ratio'] = df['Mileage'] / df['Power']
arranged_df = df.sort_values(by='Mileage', ascending=False)
summarized_df = df.groupby('Location').agg({'Power': 'sum'})
print("Selected Columns:")
print(selected_df)
print("\nFiltered Rows:")
print(filtered_df)
print("\nRenamed Columns:")
print(renamed_df)
print("\nMutated DataFrame:")
print(df)
print("\nArranged DataFrame:")
print(arranged_df)
print("\nSummarized DataFrame:")
print(summarized_df)

```

Selected Columns:

	Location	Mileage
0	Pune	19.67
1	Chennai	13.00
2	Chennai	20.77
3	Coimbatore	15.20
4	Jaipur	23.08
...
5842	Delhi	28.40
5843	Jaipur	24.40
5844	Jaipur	14.00
5845	Kolkata	18.90
5846	Hyderabad	25.44

[5847 rows x 2 columns]

```

Selected Columns:
0      Location  Mileage
1      Chennai    13.00
2      Chennai    20.77
3  Coimbatore    15.20
4      Jaipur     23.08
...
5842    Delhi     28.40
5843    Jaipur     24.40
5844    Jaipur     14.00
5845    Kolkata    18.90
5846  Hyderabad    25.44

[5847 rows x 2 columns]

Filtered Rows:
Unnamed: 0      Name      Location  Year
2      2      Maruti Ertiga VDI  Chennai  2012
4      4      Nissan Micra Diesel XV  Jaipur  2013
6      6  Volkswagen Vento Diesel Comfortline  Pune  2013
7      7    Tata Indica Vista Quadrajet LS  Chennai  2012
8      8      Maruti Ciaz Zeta  Kochi  2018
...
5833    5833      Maruti Vitara Brezza VDI  Pune  2016
5841    5841    Honda Amaze VX i-DTEC  Coimbatore  2015
5842    5842      Maruti Swift VDI  Delhi  2014
5843    5843  Hyundai Xcent 1.1 CRDI S  Jaipur  2015
5846    5846    Chevrolet Beat Diesel  Hyderabad  2011

Kilometers_Driven  Owner_Type  Mileage  Engine  Power  Seats  New_Price
2      87000      First      20.77  1248.0  88.76  7.0  4.78
4      86999      First      23.08  1461.0  63.10  5.0  4.78
6      64430      First      20.54  1598.0  103.60  5.0  4.78
7      65932      Second     22.30  1248.0  74.00  5.0  4.78
8      25692      First      21.56  1462.0  103.25  5.0  10.65
...
5833    37200      First      24.30  1248.0  88.50  5.0  9.93
5841    70602      First      25.80  1498.0  98.60  5.0  4.78
5842    27365      First      28.40  1248.0  74.00  5.0  7.88
5843    100000      First      24.40  1120.0  71.00  5.0  4.78
5846    47000      First      25.44  936.0  57.60  5.0  4.78

Price  Fuel_Type_Diesel  Fuel_Type_Electric  Fuel_Type_Petrol
2      6.00      1      0      0
4      3.50      1      0      0
6      5.20      1      0      0
7      1.95      1      0      0
8      9.95      0      0      1
...
5833    7.43      1      0      0
5841    4.83      1      0      0
5842    4.75      1      0      0
5843    4.00      1      0      0
5846    4.40      1      0      0

```

```

Transmission_Automatic Transmission_Manual
2 0 1
4 0 1
6 0 1
7 0 1
8 0 1
...
5833 0 1
5841 0 1
5842 0 1
5843 0 1
5846 0 1

[1868 rows x 17 columns]

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

Kilometers_Driven Owner_Type new_mileage Engine new_power Seats \
0 41000 First 19.67 1582.0 126.20 5.0
1 46000 First 13.00 1199.0 88.76 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

New_Price Price Fuel_Type_Diesel Fuel_Type_Electric \
0 4.78 12.50 1 0
1 8.61 4.50 0 0
2 4.78 6.00 1 0
3 4.78 17.74 1 0
4 4.78 3.50 1 0
...
5842 7.88 4.75 1 0
5843 4.78 4.00 1 0
5844 4.78 2.90 1 0
5845 4.78 2.65 0 0
5846 4.78 2.50 1 0

```

```

Fuel_Type_Petrol Transmission_Automatic Transmission_Manual
0 0 0 1
1 1 0 1
2 0 0 1
3 0 1 0
4 0 0 1
...
5842 0 0 1
5843 0 0 1
5844 0 0 1
5845 1 0 1
5846 0 0 1

[5847 rows x 17 columns]

Mutated DataFrame:
Unnamed: 0 Name Location Year \
0 0 Hyundai Creta 1.6 CRDi SX Option Pune 2015
1 1 Honda Jazz V Chennai 2011
2 2 Maruti Ertiga VOI Chennai 2012
3 3 Audi A4 New 2.0 TDI Multitronic Coimbatore 2013
4 4 Nissan Micra Diesel XV Jaipur 2013
...
5842 5842 Maruti Swift VOI Delhi 2014
5843 5843 Hyundai Xcent 1.1 CRDi S Jaipur 2015
5844 5844 Mahindra Xylo D4 BSIV Jaipur 2012
5845 5845 Maruti Wagon R VXI Kolkata 2013
5846 5846 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 4.78
1 46000 First 13.00 1199.0 88.76 5.0 8.61
2 87000 First 20.77 1248.0 88.76 7.0 4.78
3 40670 Second 15.20 1968.0 140.80 5.0 4.78
4 86999 First 23.08 1461.0 63.10 5.0 4.78
...
5842 27365 First 28.40 1248.0 74.00 5.0 7.88
5843 100000 First 24.40 1120.0 71.00 5.0 4.78
5844 55000 Second 14.00 2498.0 112.00 8.0 4.78
5845 46000 First 18.90 998.0 67.10 5.0 4.78
5846 47000 First 25.44 936.0 57.60 5.0 4.78

Price Fuel_Type_Diesel Fuel_Type_Electric Fuel_Type_Petrol \
0 12.50 1 0 0
1 4.50 0 0 1
2 6.00 1 0 0
3 17.74 1 0 0
4 3.50 1 0 0
...
5842 4.75 1 0 0
5843 4.00 1 0 0
5844 2.90 1 0 0
5845 2.65 0 0 1
5846 2.50 1 0 0

```

```

864      4.78  3.55      1      0
2361      9.14  4.00      1      0
2670      4.78  7.50      1      0
1288      7.52  6.24      1      0
4198      4.78  8.40      1      0
...      ...      ...      ...
4877      4.78 26.00      1      0
2998      4.78 28.00      1      0
524      49.49 38.99      0      0
2950      4.78  1.15      0      0
186      4.78  2.95      0      0

      Fuel_Type_Petrol  Transmission_Automatic  Transmission_Manual  \
864              0              0              1
2361              0              0              1
2670              0              1              0
1288              0              0              1
4198              0              0              1
...              ...              ...              ...
4877              0              1              0
2998              0              1              0
524              1              1              0
2950              1              0              1
186              1              0              1

      mileage_power_ratio
864              0.383784
2361              0.385085
2670              0.383784
1288              0.383784
4198              0.383784
...              ...
4877              0.000000
2998              0.000000
524              0.000000
2950              0.000000
186              0.000000

[5847 rows x 18 columns]

Summarized DataFrame:
      Location      Power
Ahmedabad    24002.910
Bangalore    46404.660
Chennai      51568.510
Coimbatore    79088.775
Delhi        65192.080
Hyderabad    72923.730
Jaipur       37797.135
Kochi        73728.960
Kolkata      53308.160
Mumbai       95231.600
Pune         61647.350

```