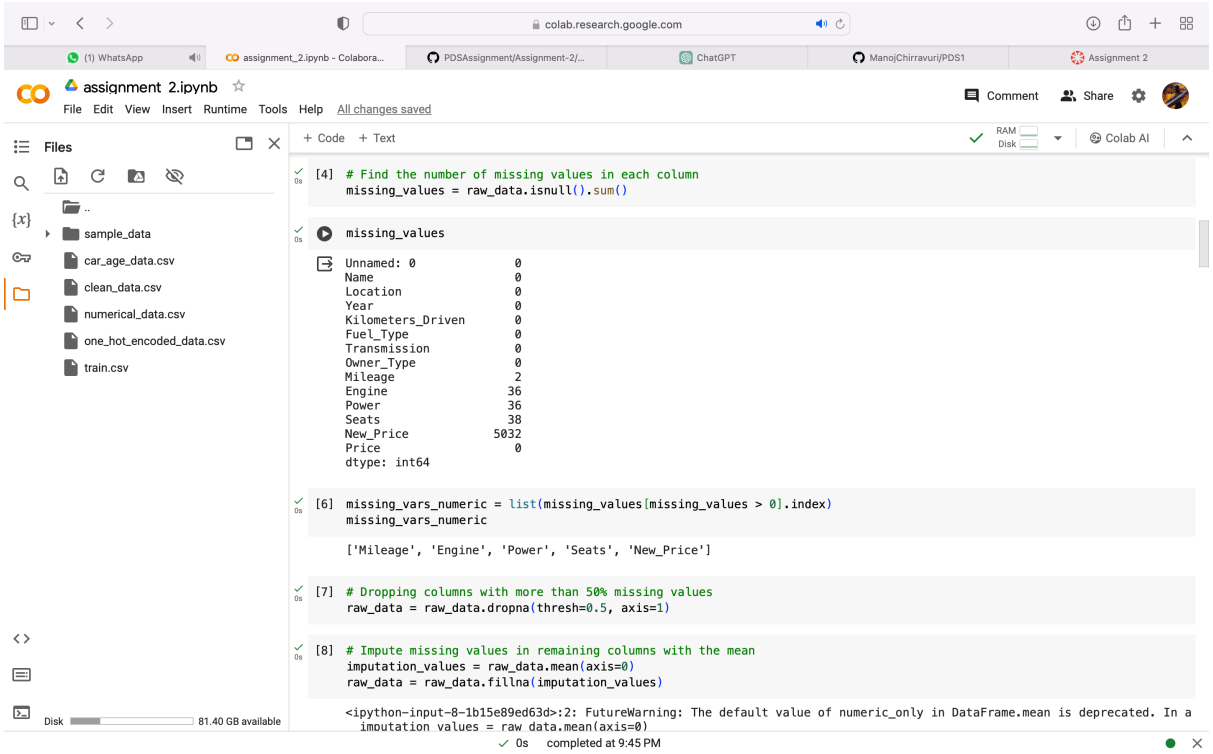
Assignment 2

Naga Venkata Manoj Chirravuri – 16345434

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.



The mean takes into account all the values in the dataset, providing an overall estimate of central tendency. This helps preserve the distribution of the data and may result in less distortion compared to other measures, especially when the data is approximately normally distributed.

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”).

```
clean_data = pd.read_csv('/content/clean_data.csv')

# Define the columns to clean
columns_to_clean = ['Mileage', 'Engine', 'Power', 'New_Price']

# Clean each column
for column in columns_to_clean:
    # Remove non-numerical characters
    clean_data[column] = clean_data[column].str.replace('[^\d.]', '')

# Convert the cleaned columns to numeric
clean_data[columns_to_clean] = clean_data[columns_to_clean].apply(pd.to_numeric)

# Store the updated DataFrame in the file
clean_data.to_csv('/content/numerical_data.csv', index=False)
```

```
<ipython-input-10-54e0ad8e76fd>:9: FutureWarning: The default value of regex will change from True to False in a futu
  clean_data[column] = clean_data[column].str.replace('[^\d.]', '')
```

C) Change the categorical variables (“Fuel_Type” and “Transmission”) into numerical one hot encoded value.

```
[11] data = pd.read_csv('/content/numerical_data.csv')
```

```
[12] one_hot_encoded_data = pd.get_dummies(data, columns=['Fuel_Type', 'Transmission'])
     one_hot_encoded_data.to_csv('/content/one_hot_encoded_data.csv', index=False)
```

```
one_hot_encoded_data
```

wer	Seats	New_Price	Price	Fuel_Type_Diesel	Fuel_Type_Electric	Fuel_Type_Petrol	Transmission_Automatic	Transmission_Manual
3.20	5.0	NaN	12.50	1	0	0	0	1
3.70	5.0	8.61	4.50	0	0	1	0	1
3.76	7.0	NaN	6.00	1	0	0	0	1
0.80	5.0	NaN	17.74	1	0	0	1	0
3.10	5.0	NaN	3.50	1	0	0	0	1
...	...	...	...	...	...	...	...	...
4.00	5.0	7.88	4.75	1	0	0	0	1
1.00	5.0	NaN	4.00	1	0	0	0	1

d) Create one more feature and add this column to the dataset (you can use mutate function in R for this). For example, you can calculate the current age of the car by subtracting “Year” value from the current year.

```
#current year
current_year = pd.to_datetime('today').year

#age of the car
data['Age'] = current_year - data['Year']

data.to_csv('/content/car_age_data.csv', index=False)
```

```
data
```

Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price	Age
undai a 1.6 Di SX ption	Pune	2015	41000	Diesel	Manual	First	19.67	1582.0	126.20	5.0	NaN	12.50	9
onda azz V	Chennai	2011	46000	Petrol	Manual	First	13.00	1199.0	88.70	5.0	8.61	4.50	13
laruti a VDI	Chennai	2012	87000	Diesel	Manual	First	20.77	1248.0	88.76	7.0	NaN	6.00	12
di A4 w 2.0 TDI ronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.20	1968.0	140.80	5.0	NaN	17.74	11
issan Micra el XV	Jaipur	2013	86999	Diesel	Manual	First	23.08	1461.0	63.10	5.0	NaN	3.50	11
...	...	...	...	...	...	...	...	...	...	...	...	...	...
laruti t VDI	Delhi	2014	27365	Diesel	Manual	First	28.40	1248.0	74.00	5.0	7.88	4.75	10
undai nt 1.1 RDi S	Jaipur	2015	100000	Diesel	Manual	First	24.40	1120.0	71.00	5.0	NaN	4.00	9
indra lo D4 BSIV	Jaipur	2012	55000	Diesel	Manual	Second	14.00	2498.0	112.00	8.0	NaN	2.90	12

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

Select

```
data = pd.read_csv('/content/car_age_data.csv')
selected_data = data[['Name', 'Engine', 'Power', 'Age']]

print("Selected data:\n", selected_data)
```

```
Selected data:
                                  Name  Engine   Power  Age
0     Hyundai Creta 1.6 CRDi SX Option  1582.0  126.20    9
1                         Honda Jazz V  1199.0   88.70   13
2                    Maruti Ertiga VDI  1248.0   88.76   12
3      Audi A4 New 2.0 TDI Multitronic  1968.0  140.80   11
4                Nissan Micra Diesel XV  1461.0   63.10   11
...                                 ...     ...     ...  ...
5842                 Maruti Swift VDI  1248.0   74.00   10
5843          Hyundai Xcent 1.1 CRDi S  1120.0   71.00    9
5844             Mahindra Xylo D4 BSIV  2498.0  112.00   12
5845                Maruti Wagon R VXI   998.0   67.10   11
5846             Chevrolet Beat Diesel   936.0   57.60   13

[5847 rows x 4 columns]
```

Filter:

```
filtered_data = data[data['Year'] > 2010]

print("\nFiltered data:\n", filtered_data)
```

```
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 Fuel_Type Transmission Owner_Type  Mileage  Engine  \
0                41000    Diesel       Manual      First    19.67  1582.0
1                46000    Petrol       Manual      First    13.00  1199.0
2                87000    Diesel       Manual      First    20.77  1248.0
3                40670    Diesel    Automatic     Second    15.20  1968.0
4                86999    Diesel       Manual      First    23.08  1461.0
...                ...       ...          ...        ...      ...     ...
5842             27365    Diesel       Manual      First    28.40  1248.0
5843            100000    Diesel       Manual      First    24.40  1120.0
5844             55000    Diesel       Manual     Second    14.00  2498.0
5845             46000    Petrol       Manual      First    18.90   998.0
5846             47000    Diesel       Manual      First    25.44   936.0

       Power  Seats  New_Price  Price  Age
0     126.20    5.0        NaN  12.50    9
1      88.70    5.0       8.61   4.50   13
2      88.76    7.0        NaN   6.00   12
3     140.80    5.0        NaN  17.74   11
4      63.10    5.0        NaN   3.50   11
```

Rename

```
renamed_data = data.rename(columns={'Engine': 'CC', 'Price': 'Old_Price'})

print("\nRenamed data:\n", renamed_data)
```

```
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 Fuel_Type Transmission Owner_Type  Mileage      CC  \
0                41000    Diesel       Manual      First    19.67  1582.0
1                46000    Petrol       Manual      First    13.00  1199.0
2                87000    Diesel       Manual      First    20.77  1248.0
3                40670    Diesel    Automatic     Second    15.20  1968.0
4                86999    Diesel       Manual      First    23.08  1461.0
...                ...       ...          ...        ...      ...     ...
5842             27365    Diesel       Manual      First    28.40  1248.0
5843            100000    Diesel       Manual      First    24.40  1120.0
5844             55000    Diesel       Manual     Second    14.00  2498.0
5845             46000    Petrol       Manual      First    18.90   998.0
5846             47000    Diesel       Manual      First    25.44   936.0

       Power  Seats  New_Price  Old_Price  Age
0     126.20    5.0        NaN      12.50    9
1      88.70    5.0       8.61       4.50   13
```

Mutate:

```
mutated_data = data.assign(Increase_in_price=data['New_Price'] - data['Price'])

print("\nMutated data:\n", mutated_data)
```

```
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 Fuel_Type Transmission Owner_Type  Mileage  Engine  \
0                 41000    Diesel       Manual      First    19.67  1582.0
1                 46000    Petrol       Manual      First    13.00  1199.0
2                 87000    Diesel       Manual      First    20.77  1248.0
3                 40670    Diesel    Automatic     Second    15.20  1968.0
4                 86999    Diesel       Manual      First    23.08  1461.0
...                 ...       ...          ...        ...      ...     ...
5842              27365    Diesel       Manual      First    28.40  1248.0
5843             100000    Diesel       Manual      First    24.40  1120.0
5844              55000    Diesel       Manual     Second    14.00  2498.0
5845              46000    Petrol       Manual      First    18.90   998.0
5846              47000    Diesel       Manual      First    25.44   936.0

       Power  Seats  New_Price  Price  Age  Increase_in_price
0     126.20    5.0        NaN  12.50    9                NaN
1      88.70    5.0       8.61   4.50   13               4.11
2      88.76    7.0        NaN   6.00   12                NaN
3     140.80    5.0        NaN  17.74   11                NaN
4      63.10    5.0        NaN   3.50   11                NaN
...      ...    ...        ...    ...  ...                ...
```

Arrange:

```
arranged_data = data.sort_values(by='Year', ascending=False)

print("\nArranged data:\n", arranged_data)
```

```
Arranged data:
       Unnamed: 0                                 Name    Location  Year  \
5405        5560          Renault KWID RXT Optional       Kochi  2019
942          975    Ford Endeavour 2.2 Trend AT 4X2       Kochi  2019
5533        5690          Maruti Omni 5 Seater BSIV  Coimbatore  2019
770          796              Mahindra XUV500 W9 AT  Coimbatore  2019
4267        4399         Maruti Swift Dzire AMT ZDI     Chennai  2019
...          ...                                ...         ...   ...
1185        1224                    Maruti Zen VX      Jaipur  1999
1791        1845                Honda City 1.3 EXI        Pune  1999
3630        3749  Mercedes-Benz E-Class 250 D W 210      Mumbai  1998
5558        5716                    Maruti Zen LX      Jaipur  1998
3039        3138                   Maruti Zen LXI      Jaipur  1998

      Kilometers_Driven Fuel_Type Transmission Owner_Type  Mileage  Engine  \
5405               6568    Petrol       Manual      First    25.17   799.0
942               11209    Diesel    Automatic      First    12.62  2198.0
5533               4721    Petrol       Manual      First    14.00   796.0
770               19654    Diesel    Automatic      First    14.00  2179.0
4267              65000    Diesel    Automatic      First    26.59  1248.0
...                 ...       ...          ...        ...      ...     ...
1185              70000    Petrol       Manual     Second    17.30   993.0
1791             140000    Petrol       Manual      First    13.00  1343.0
3630              55300    Diesel    Automatic      First    10.00  1796.0
5558              95150    Petrol       Manual      Third    17.30   993.0
3039              95150    Petrol       Manual      Third    17.30   993.0

      Power  Seats  New_Price  Price  Age
5405   53.3    5.0       4.78   5.09    5
942   158.0    7.0        NaN  31.15    5
5533   35.0    5.0        NaN   4.11    5
770   155.0    7.0      21.33  17.63    5
4267   74.0    5.0        NaN   6.75    5
```

Group by:

```
grouped_data = data.groupby('Fuel_Type').agg({'Mileage': 'mean'})

print("\nGrouped data:\n", grouped_data)
```

```
Grouped data:
              Mileage
Fuel_Type
Diesel     18.652661
Electric         NaN
Petrol     17.576509
```