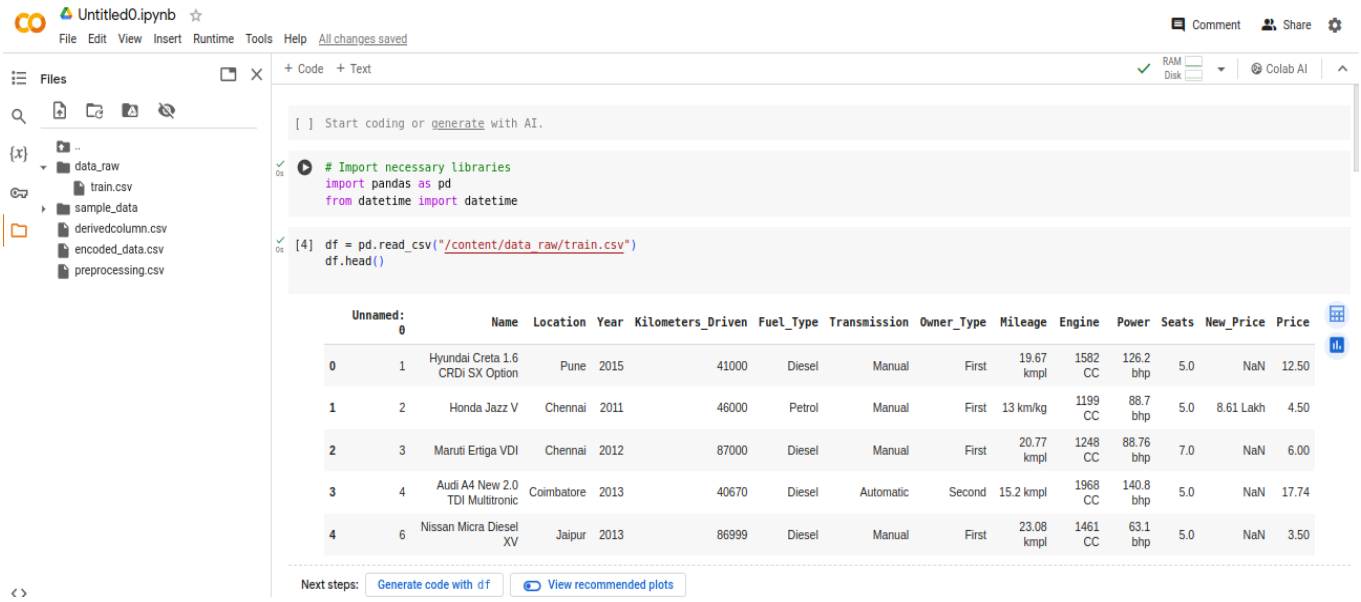
Student Name : Narendra Mamilla

Student ID : 16351892

Assignment - 2

1. Loading the train.csv file into google collab



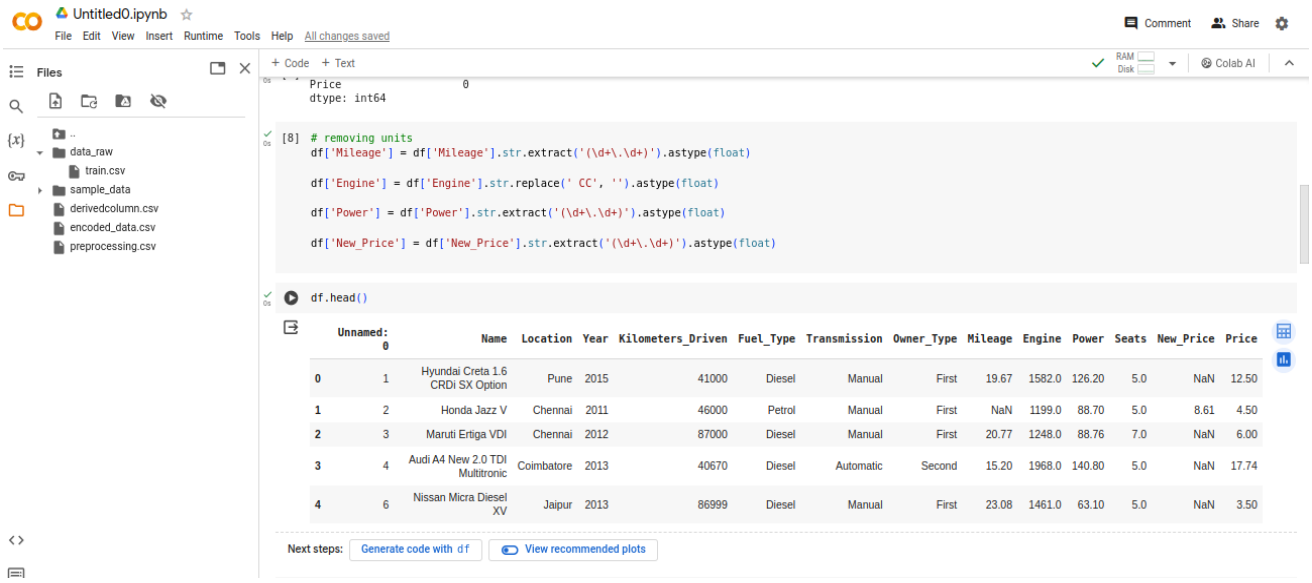
	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	NaN	12.50
1	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	13 km/kg	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

2. Identifying the missing values

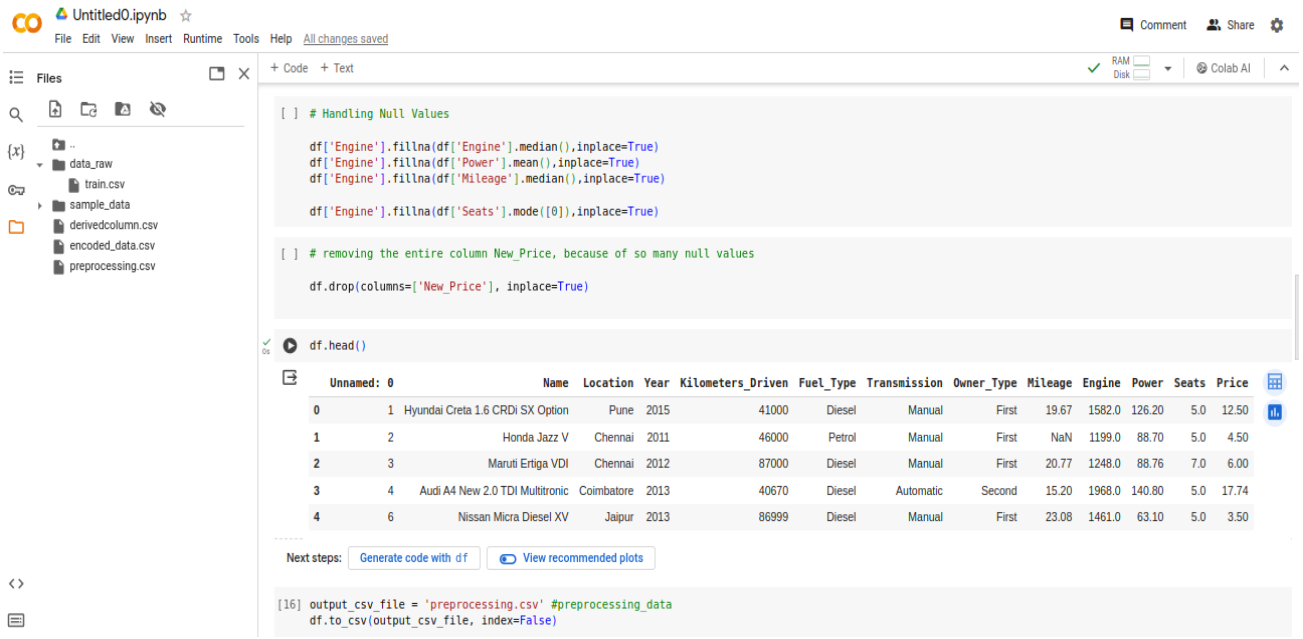
```
# Identify missing values
missing_values = df.isnull().sum()
print("Missing values in each column:")
print(missing_values)
```

```
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
```

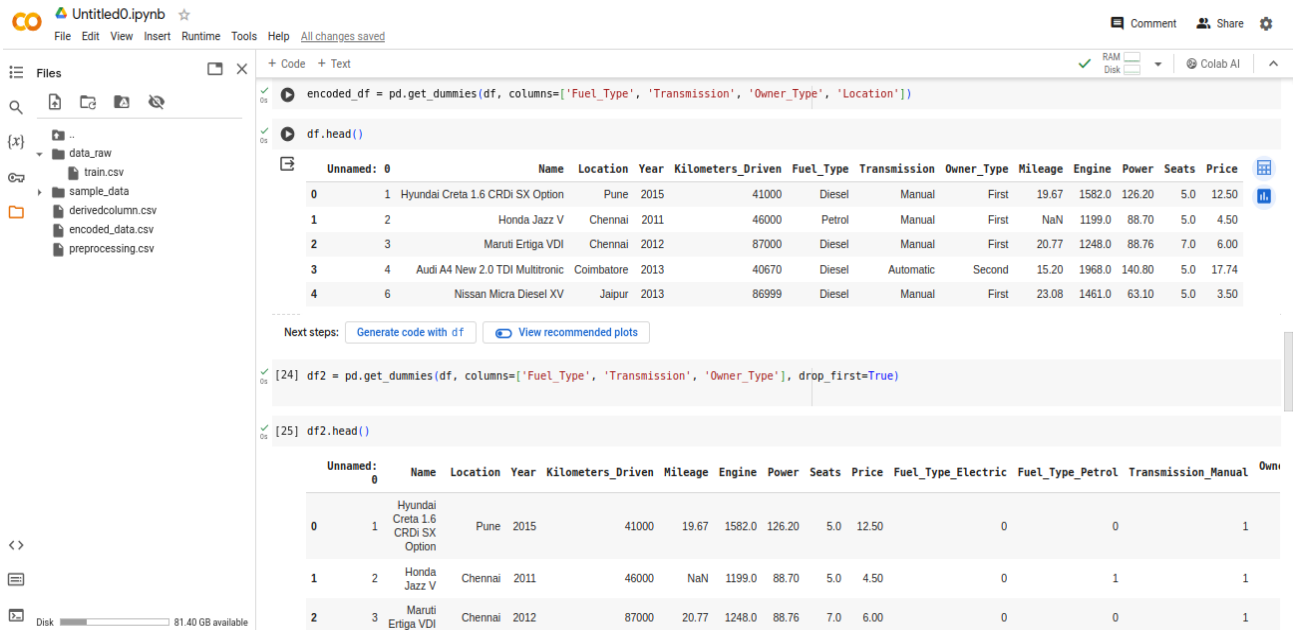
3. Removing the units from Power, Mileage, Engine and New Price



4. Handling the Null values by removing the entire column “New_Price” because it contains so many null values.



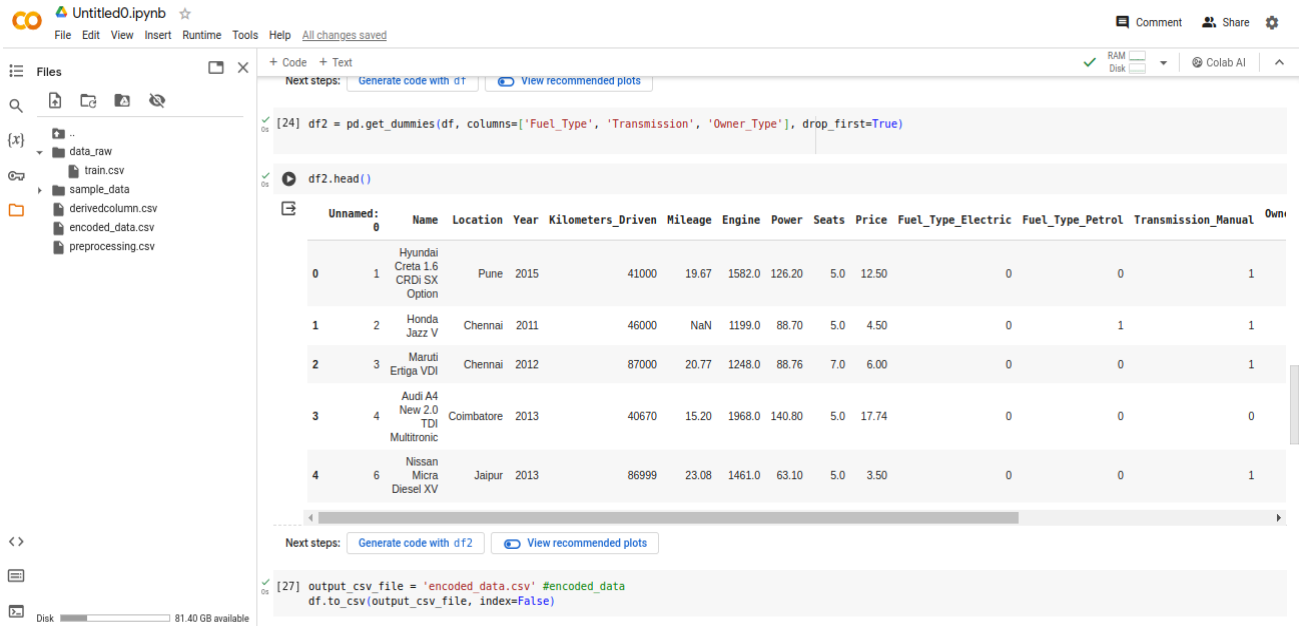
5. Encoding the columns Fuel Types, Transmission, Owner Type and Location to numerical by using 1 hot encoding.



	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	Price
0	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67	1582.0	126.20	5.0	12.50
1	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	NaN	1199.0	88.70	5.0	4.50
2	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77	1248.0	88.76	7.0	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	17.74
4	6	Nissan Micra Diesel XV	Jaipur	2013	86999	Diesel	Manual	First	23.08	1461.0	63.10	5.0	3.50

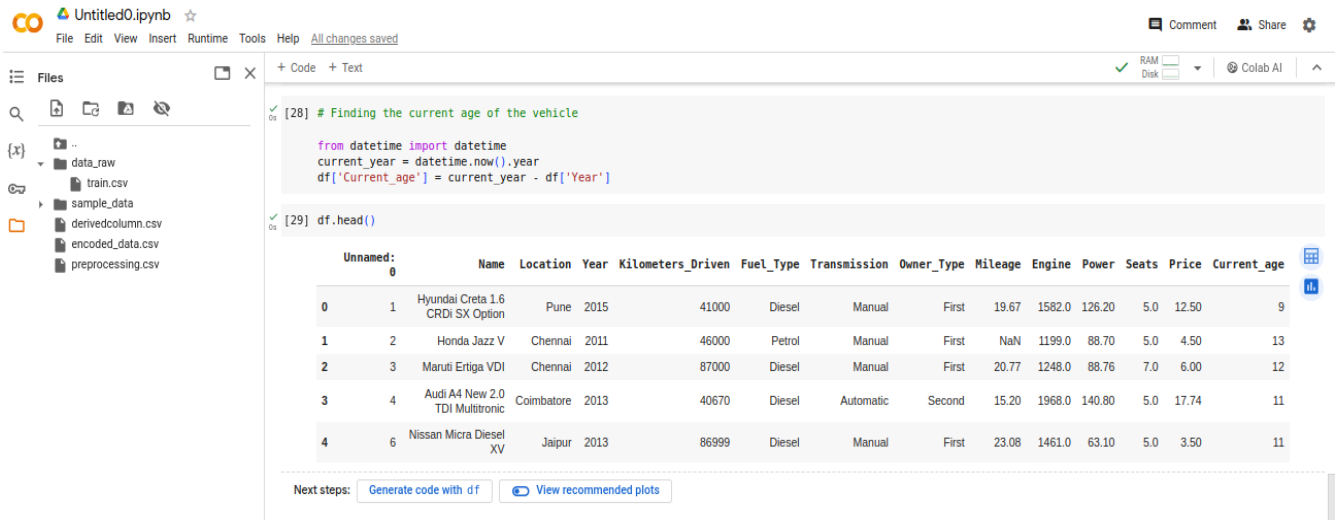
	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Price	Fuel_Type_Electric	Fuel_Type_Petrol	Transmission_Manual	Own
0	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	19.67	1582.0	126.20	5.0	12.50	0	0	1	
1	2	Honda Jazz V	Chennai	2011	46000	NaN	1199.0	88.70	5.0	4.50	0	1	1	
2	3	Maruti Ertiga VDI	Chennai	2012	87000	20.77	1248.0	88.76	7.0	6.00	0	0	1	

6. And then Downloading the encoded csv file.



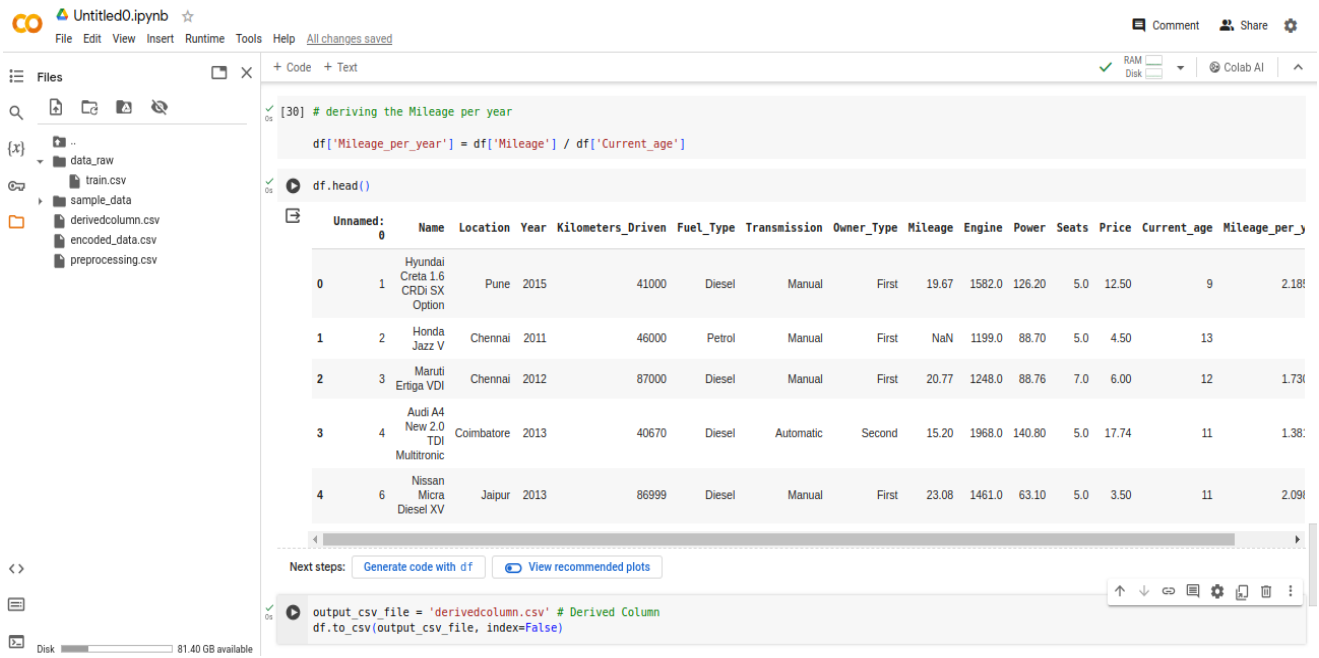
	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Price	Fuel_Type_Electric	Fuel_Type_Petrol	Transmission_Manual	Own
0	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	19.67	1582.0	126.20	5.0	12.50	0	0	1	
1	2	Honda Jazz V	Chennai	2011	46000	NaN	1199.0	88.70	5.0	4.50	0	1	1	
2	3	Maruti Ertiga VDI	Chennai	2012	87000	20.77	1248.0	88.76	7.0	6.00	0	0	1	
3	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	15.20	1968.0	140.80	5.0	17.74	0	0	0	
4	6	Nissan Micra Diesel XV	Jaipur	2013	86999	23.08	1461.0	63.10	5.0	3.50	0	0	1	

7. Finding the current age of the vehicle to derive the new column “Mileage per year”.



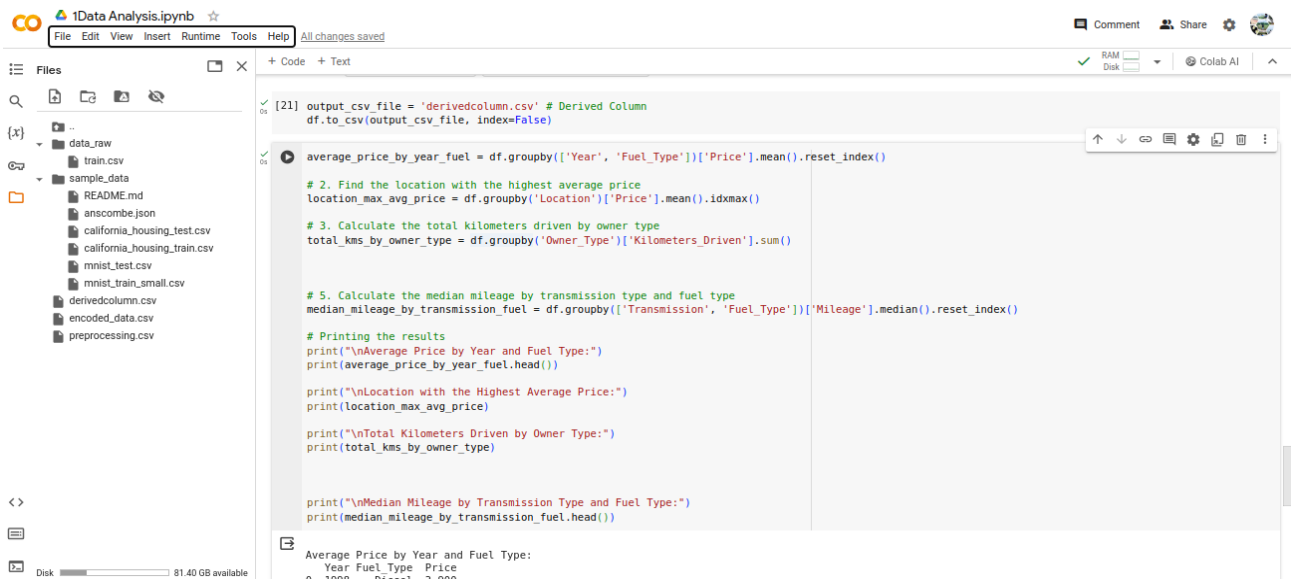
	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	Price	Current_age
0	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67	1582.0	126.20	5.0	12.50	9
1	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	NaN	1199.0	88.70	5.0	4.50	13
2	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77	1248.0	88.76	7.0	6.00	12
3	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.20	1968.0	140.80	5.0	17.74	11
4	6	Nissan Micra Diesel XV	Jaipur	2013	86999	Diesel	Manual	First	23.08	1461.0	63.10	5.0	3.50	11

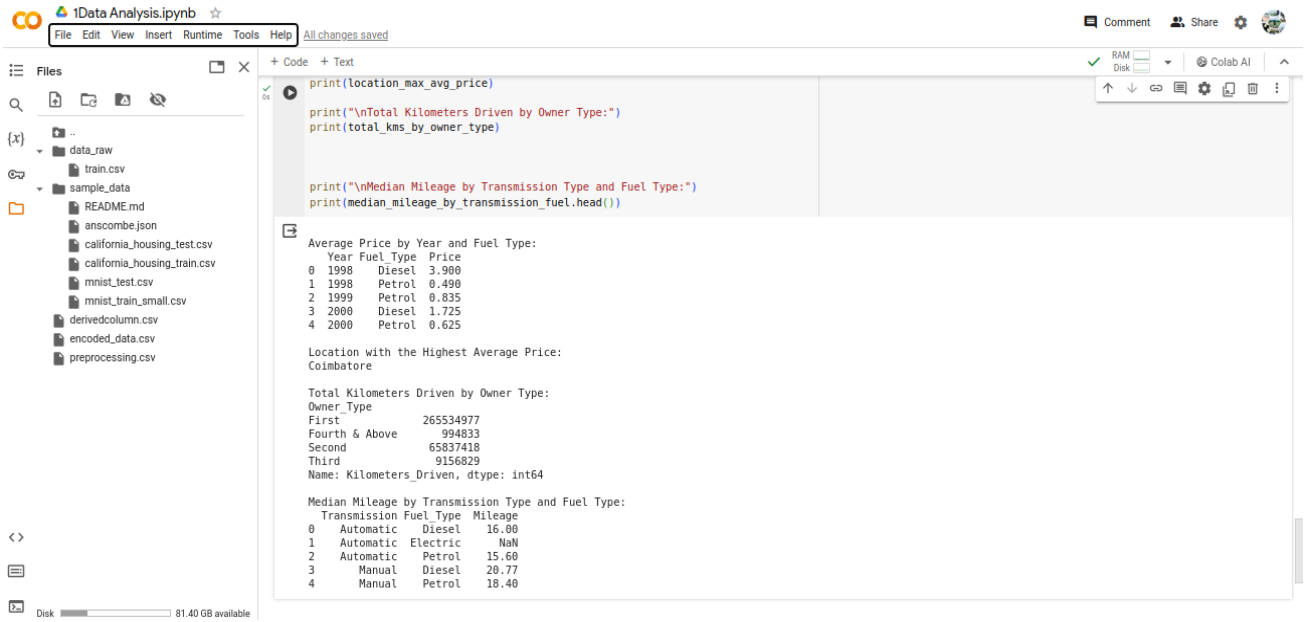
8. Deriving the mileage per year and then downloading the csv file.



	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	Price	Current_age	Mileage_per_y
0	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67	1582.0	126.20	5.0	12.50	9	2.18
1	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	NaN	1199.0	88.70	5.0	4.50	13	
2	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77	1248.0	88.76	7.0	6.00	12	1.73
3	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.20	1968.0	140.80	5.0	17.74	11	1.38
4	6	Nissan Micra Diesel XV	Jaipur	2013	86999	Diesel	Manual	First	23.08	1461.0	63.10	5.0	3.50	11	2.09

9. Manipulations of given data set like group by...


1Data Analysis.ipynb
File Edit View Insert Runtime Tools Help
All changes saved
Comment
Share
Files
+ Code + Text
Colab AI
[21] output_csv_file = 'derivedcolumn.csv' # Derived Column
df.to_csv(output_csv_file, index=False)
average_price_by_year_fuel = df.groupby(['Year', 'Fuel_Type'])['Price'].mean().reset_index()
# 2. Find the location with the highest average price
location_max_avg_price = df.groupby('Location')['Price'].mean().idxmax()
# 3. Calculate the total kilometers driven by owner type
total_kms_by_owner_type = df.groupby('Owner_Type')['Kilometers_Driven'].sum()
# 5. Calculate the median mileage by transmission type and fuel type
median_mileage_by_transmission_fuel = df.groupby(['Transmission', 'Fuel_Type'])['Mileage'].median().reset_index()
# Printing the results
print("\nAverage Price by Year and Fuel Type:")
print(average_price_by_year_fuel.head())
print("\nLocation with the Highest Average Price:")
print(location_max_avg_price)
print("\nTotal Kilometers Driven by Owner Type:")
print(total_kms_by_owner_type)
print("\nMedian Mileage by Transmission Type and Fuel Type:")
print(median_mileage_by_transmission_fuel.head())
Average Price by Year and Fuel Type:
Year Fuel_Type Price
data_raw
train.csv
sample_data
README.md
anscombe.json
california_housing_test.csv
california_housing_train.csv
mnist_test.csv
mnist_train_small.csv
derivedcolumn.csv
encoded_data.csv
preprocessing.csv
Disk
81.40 GB available


1Data Analysis.ipynb
File Edit View Insert Runtime Tools Help
All changes saved
Comment
Share
Files
+ Code + Text
Colab AI
print(location_max_avg_price)
print("\nTotal Kilometers Driven by Owner Type:")
print(total_kms_by_owner_type)
print("\nMedian Mileage by Transmission Type and Fuel Type:")
print(median_mileage_by_transmission_fuel.head())
Average Price by Year and Fuel Type:
Year Fuel_Type Price
0 1998 Diesel 3.900
1 1998 Petrol 0.490
2 1999 Petrol 0.835
3 2000 Diesel 1.725
4 2000 Petrol 0.625
Location with the Highest Average Price:
Coimbatore
Total Kilometers Driven by Owner Type:
Owner_Type
First 265534977
Fourth & Above 994833
Second 65837418
Third 9156829
Name: Kilometers_Driven, dtype: int64
Median Mileage by Transmission Type and Fuel Type:
Transmission Fuel_Type Mileage
0 Automatic Diesel 16.00
1 Automatic Electric NaN
2 Automatic Petrol 15.60
3 Manual Diesel 20.77
4 Manual Petrol 18.40
data_raw
train.csv
sample_data
README.md
anscombe.json
california_housing_test.csv
california_housing_train.csv
mnist_test.csv
mnist_train_small.csv
derivedcolumn.csv
encoded_data.csv
preprocessing.csv
Disk
81.40 GB available