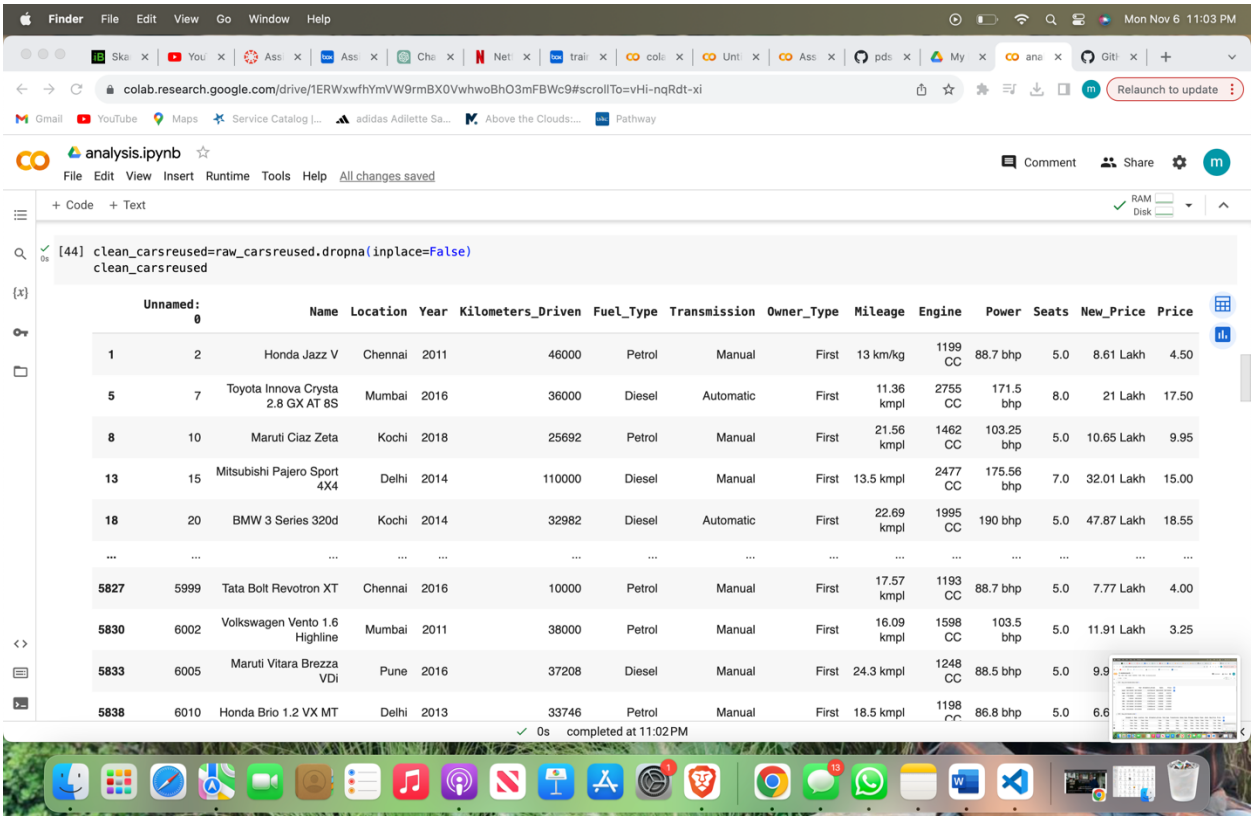
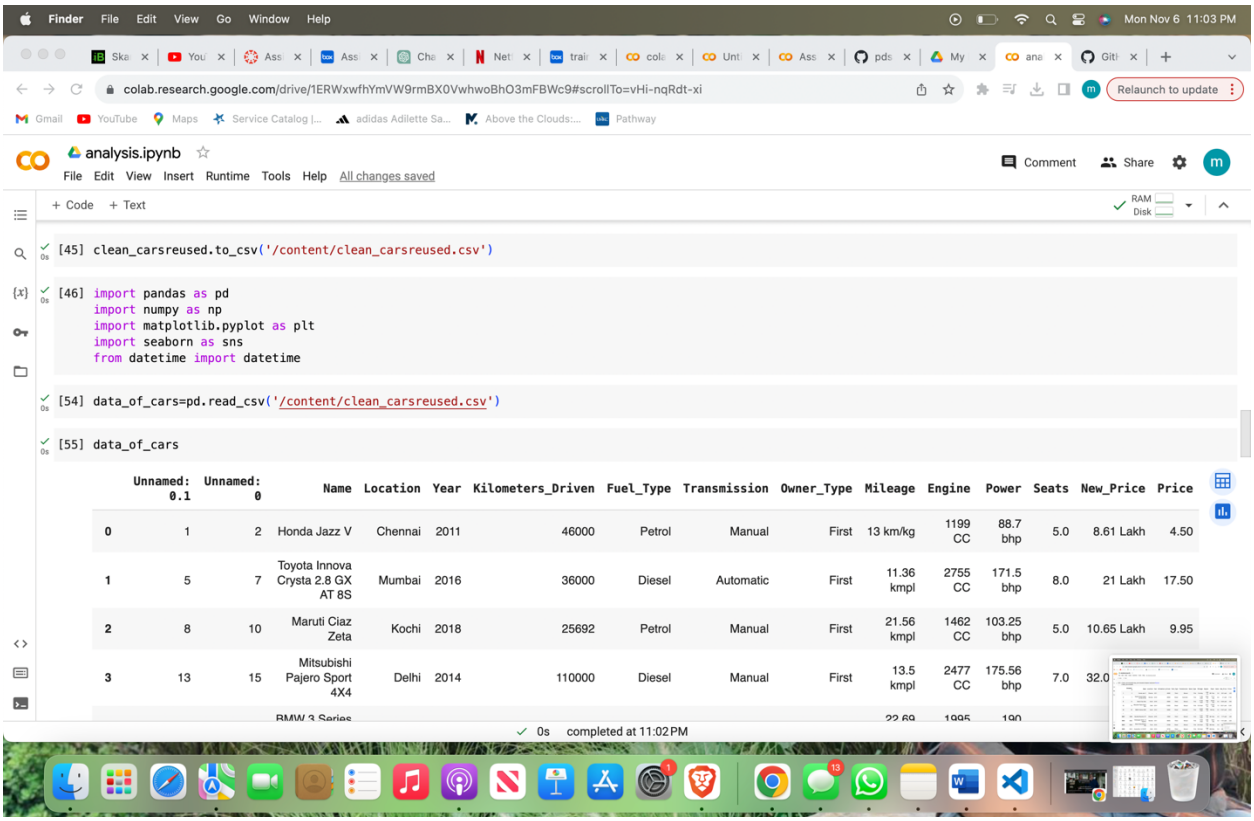
Assignment 2

Mohana Sai Kumar Reddy Bobba
16337925

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.

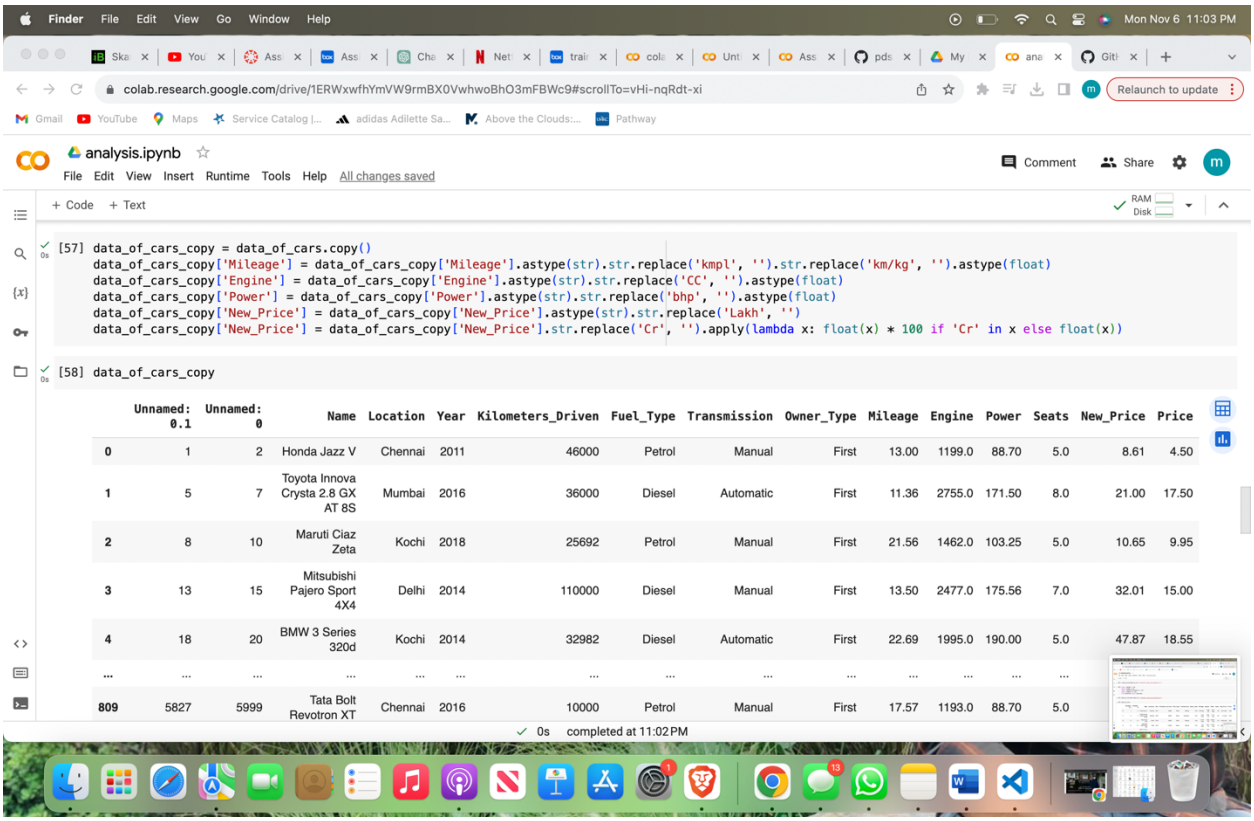


	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
1	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	13 km/kg	1199 CC	88.7 bhp	5.0	8.61 Lakh	4.50
5	7	Toyota Innova Crysta 2.8 GX AT 8S	Mumbai	2016	36000	Diesel	Automatic	First	11.36 kmpl	2755 CC	171.5 bhp	8.0	21 Lakh	17.50
8	10	Maruti Ciaz Zeta	Kochi	2018	25692	Petrol	Manual	First	21.56 kmpl	1462 CC	103.25 bhp	5.0	10.65 Lakh	9.95
13	15	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
18	20	BMW 3 Series 320d	Kochi	2014	32982	Diesel	Automatic	First	22.69 kmpl	1995 CC	190 bhp	5.0	47.87 Lakh	18.55
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
5827	5999	Tata Bolt Revotron XT	Chennai	2016	10000	Petrol	Manual	First	17.57 kmpl	1193 CC	88.7 bhp	5.0	7.77 Lakh	4.00
5830	6002	Volkswagen Vento 1.6 Highline	Mumbai	2011	38000	Petrol	Manual	First	16.09 kmpl	1598 CC	103.5 bhp	5.0	11.91 Lakh	3.25
5833	6005	Maruti Vitara Brezza VDi	Pune	2016	37208	Diesel	Manual	First	24.3 kmpl	1248 CC	88.5 bhp	5.0	9.9	
5838	6010	Honda Brio 1.2 VX MT	Delhi	2013	33746	Petrol	Manual	First	18.5 kmpl	1198 CC	86.8 bhp	5.0	6.6	



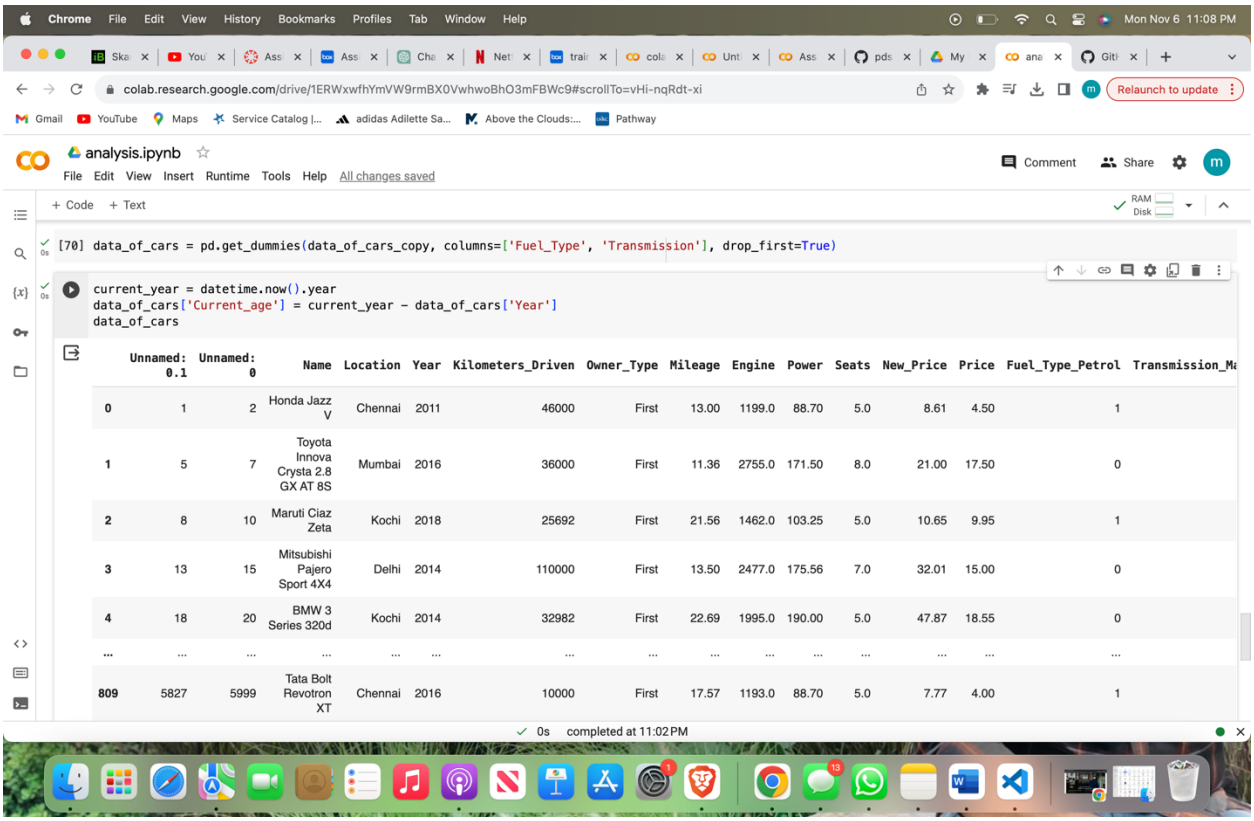
	Unnamed: 0.1	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	1	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	13 km/kg	1199 CC	88.7 bhp	5.0	8.61 Lakh	4.50
1	5	7	Toyota Innova Crysta 2.8 GX AT 8S	Mumbai	2016	36000	Diesel	Automatic	First	11.36 kmpl	2755 CC	171.5 bhp	8.0	21 Lakh	17.50
2	8	10	Maruti Ciaz Zeta	Kochi	2018	25692	Petrol	Manual	First	21.56 kmpl	1462 CC	103.25 bhp	5.0	10.65 Lakh	9.95
3	13	15	Mitsubishi Pajero Sport 4X4	Delhi	2014	110000	Diesel	Manual	First	13.5 kmpl	2477 CC	175.56 bhp	7.0	32.0	

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”). (5 points)



	Unnamed: 0.1	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	1	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	13.00	1199.0	88.70	5.0	8.61	4.50
1	5	7	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
2	8	10	Maruti Ciaz Zeta	Kochi	2018	25692	Petrol	Manual	First	21.56	1462.0	103.25	5.0	10.65	9.95
3	13	15	Mitsubishi Pajero Sport 4X4	Delhi	2014	110000	Diesel	Manual	First	13.50	2477.0	175.56	7.0	32.01	15.00
4	18	20	BMW 3 Series 320d	Kochi	2014	32982	Diesel	Automatic	First	22.69	1995.0	190.00	5.0	47.87	18.55
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
809	5827	5999	Tata Bolt Revotron XT	Chennai	2016	10000	Petrol	Manual	First	17.57	1193.0	88.70	5.0		

C) Change the categorical variables (“Fuel_Type” and “Transmission”) into numerical one hot encoded value. (5 points)

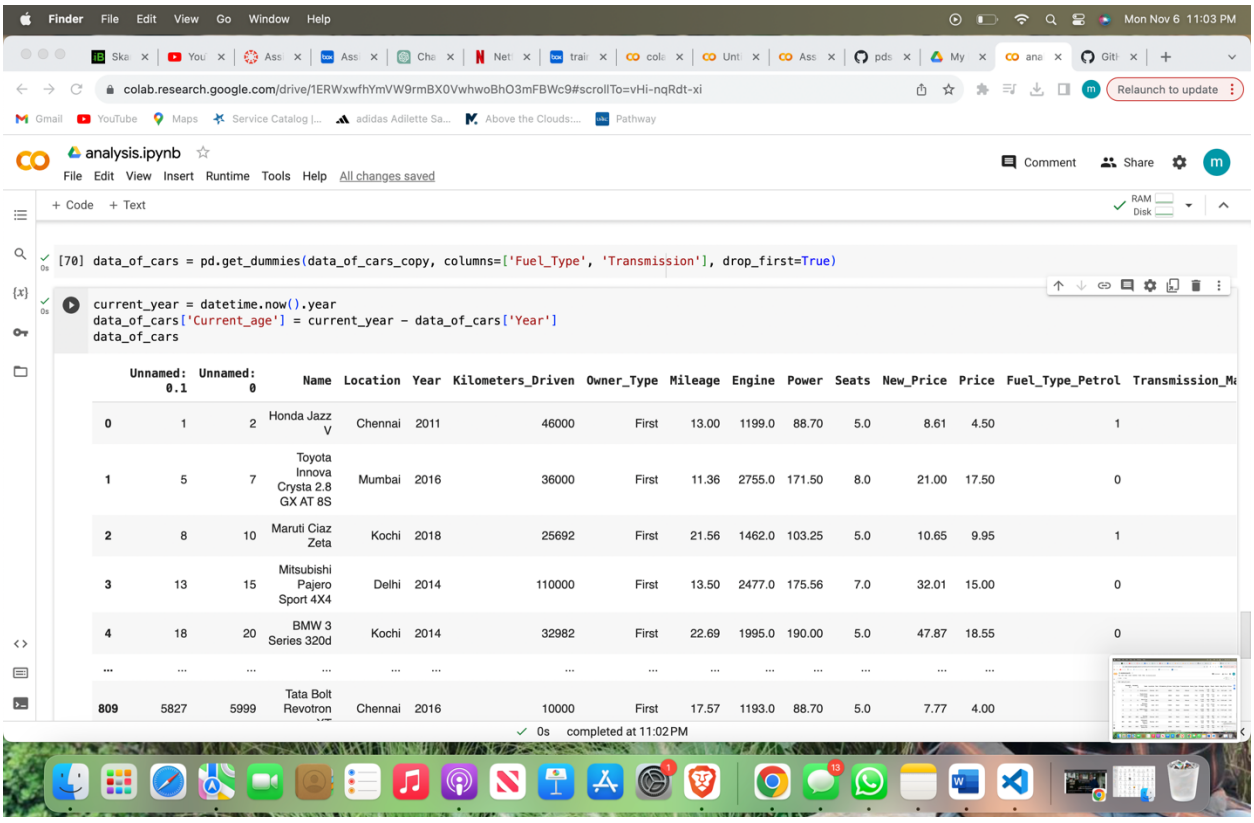


```
[70] data_of_cars = pd.get_dummies(data_of_cars_copy, columns=['Fuel_Type', 'Transmission'], drop_first=True)

current_year = datetime.now().year
data_of_cars['Current_age'] = current_year - data_of_cars['Year']
data_of_cars
```

	Unnamed: 0.1	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price	Fuel_Type_Petrol	Transmission_Ma
0	1	2	Honda Jazz V	Chennai	2011	46000	First	13.00	1199.0	88.70	5.0	8.61	4.50	1	
1	5	7	Toyota Innova Crysta 2.8 GX AT 8S	Mumbai	2016	36000	First	11.36	2755.0	171.50	8.0	21.00	17.50	0	
2	8	10	Maruti Ciaz Zeta	Kochi	2018	25692	First	21.56	1462.0	103.25	5.0	10.65	9.95	1	
3	13	15	Mitsubishi Pajero Sport 4X4	Delhi	2014	110000	First	13.50	2477.0	175.56	7.0	32.01	15.00	0	
4	18	20	BMW 3 Series 320d	Kochi	2014	32982	First	22.69	1995.0	190.00	5.0	47.87	18.55	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
809	5827	5999	Tata Bolt Revotron XT	Chennai	2016	10000	First	17.57	1193.0	88.70	5.0	7.77	4.00	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. (5 points)



	Unnamed: 0.1	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price	Fuel_Type_Petrol	Transmission_Ma
0	1	2	Honda Jazz V	Chennai	2011	46000	First	13.00	1199.0	88.70	5.0	8.61	4.50	1	
1	5	7	Toyota Innova Crysta 2.8 GX AT 8S	Mumbai	2016	36000	First	11.36	2755.0	171.50	8.0	21.00	17.50	0	
2	8	10	Maruti Ciaz Zeta	Kochi	2018	25692	First	21.56	1462.0	103.25	5.0	10.65	9.95	1	
3	13	15	Mitsubishi Pajero Sport 4X4	Delhi	2014	110000	First	13.50	2477.0	175.56	7.0	32.01	15.00	0	
4	18	20	BMW 3 Series 320d	Kochi	2014	32982	First	22.69	1995.0	190.00	5.0	47.87	18.55	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
809	5827	5999	Tata Bolt Revotron	Chennai	2016	10000	First	17.57	1193.0	88.70	5.0	7.77	4.00		