ML & DS Capstone Project:

	Model_Year	Engine_Size	Cylinders	Fuel_Consumption_in_City(L/100 km)	Fuel_Consumption_in_City_Hwy(L/100 km)	Fuel_Consumption_comb(L/100km)	CO2_Emissions
Mean	2021.0	3.212727	5.712299	12.458289	9.290588	11.033048	257.898128
Median	2021.0	3.0	6.0	12.0	9.0	10.7	255.0
Mode	2021	2.0	4	10.1	8.7	9.1	214.0
Min	2021	1.0	3	4.0	3.9	4.0	94.0
Q1:25%	2021.0	2.0	4.0	10.1	7.7	9.1	213.0
Q2:50%	2021.0	3.0	6.0	12.0	9.0	10.7	255.0
Q3:75%	2021.0	4.0	8.0	14.8	10.8	13.1	303.5
99%	2021.0	6.6	12.0	21.765	15.45	19.1	408.0
Q4:100%	2021	7.0	14	21.85	15.45	19.1	439.25
IQR	0.0	2.0	4.0	4.7	3.1	4.0	90.5
Lesser_IQR	2021.0	-1.0	-2.0	3.05	3.05	3.1	77.25
Greater_IQR	2021.0	7.0	14.0	21.85	15.45	19.1	439.25
Skewness	0.0	0.849973	1.022354	0.377239	0.469603	0.395946	0.20424
Kurtosis	0.0	-0.222371	1.049699	0.121686	-0.098956	0.059601	-0.207735
Variance	0.0	1.909613	3.830416	11.159259	4.676448	7.673585	3859.275411
Std_deviation	0.0	1.381887	1.957145	3.340548	2.16251	2.770124	62.123067

UNIVARIATE ANALYSIS:

1. **Mean**: From the above table, we can infer that the average fuel consumption in city side is higher than the average fuel consumption in highways and the combination of city and highway. The average engine size and cylinder size of different car models are 3.2 and 5.7 respectively. The average CO_2 Emissions is 256.

2. **Median**: The mean and median are similar for all the columns. Hence, there is no potential outliers present in these columns.

3. **Mode**: The repeated values for different columns of Model year, Engine size, cylinders, fuel consumption in city, fuel consumption in highway, combo fuel consumption, CO_2 Emissions and smog level are:
   2021, 2, 4, 10.1, 8.7, 9.1, 214 and 5 respectively.

4. **Percentile**:
   i. Engine Size: There is 1% increment in engine size for every percentile from 25% to 75% and in the last 99% and 100%, there is 2% increment in each percentile.
   ii. Cylinders: There is 2% increment in cylinder size for every percentile from 25% to 75% and in the last 99% and 100%, there is 4% increment in each percentile.
   iii. Fuel Consumption in City: There is 2% increment in cylinder size for every percentile from 25% to 75% and in the last 99% and 100%, there is 7% and 9% increment respectively.

iv. Fuel Consumption in City Hwy: There is 2% increment in cylinder size for every percentile from 25% to 75% and in the last 99% and 100%, there is 5% increment in each percentile.

v. Fuel Consumption comb: In the first 25% increment, there is only 1.6% increment. From the 50th percentile to 75th percentile, there is 3% increment in fuel consumption and in the last 99% and 100%, there is 6% increment in each percentile.

vi. CO_2 Emissions: There is high increment only in the CO_2 Emissions column in the dataset. In the first 25% increment, the emission is 42 numbers high. In the 2nd increment, there is 48 numbers high emission. From Q3 to 99%, there is 105 numbers increment in CO_2 Emission. In the 100th percentile, there is 200 numbers increment.

vii. Smog level: There is 2% increment in smog level from 25% to 50% and only 1% increment in the next percentile ranges. There is no increment in smog level from 99% to 100%.

5. **Inter Quartile Range**: There is no lesser range of outliers present in the dataset. However greater range of outliers are present in the following columns:
Engine Size, Cylinders, Fuel Consumption in City (L/100 km), Fuel Consumption in City Hwy (L/100 km), Fuel Consumption comb (L/100km), CO2 Emissions.
Even though greater range of outliers are present, they are not very potential as there is only small differences between the greater IQR and maximum values of the columns.

6. **Skewness**: The skewness for smog level column is negative; hence the central tendency is in the order of Mean>Median>Mode.
For all the other columns, we have positive skewness; hence the central tendency is in the order of Mean<Median<Mode.

7. **Kurtosis**: It can be seen that all the values are <3 for kurtosis, hence all the columns follow Platykurtic type.