```
In [45]: import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         import seaborn as sns
```

```
In [46]: psdf = pd.read_csv("co2_emissions_canada.csv")
```

```
In [47]: psdf
```

Out[47]:

	Make	Model	Vehicle Class	Engine Size(L)	Cylinders	Transmission	Fuel Type	Fuel Consumption Comb (L/100 km)	Emission
0	ACURA	ILX	COMPACT	2.0	4	AS5	Z	8.5	
1	ACURA	ILX	COMPACT	2.4	4	M6	Z	9.6	
2	ACURA	ILX HYBRID	COMPACT	1.5	4	AV7	Z	5.9	
3	ACURA	MDX 4WD	SUV - SMALL	3.5	6	AS6	Z	11.1	
4	ACURA	RDX AWD	SUV - SMALL	3.5	6	AS6	Z	10.6	
...	...	...	...	...	...	...	...	...	
7380	VOLVO	XC40 T5 AWD	SUV - SMALL	2.0	4	AS8	Z	9.4	
7381	VOLVO	XC60 T5 AWD	SUV - SMALL	2.0	4	AS8	Z	9.9	
7382	VOLVO	XC60 T6 AWD	SUV - SMALL	2.0	4	AS8	Z	10.3	
7383	VOLVO	XC90 T5 AWD	SUV - STANDARD	2.0	4	AS8	Z	9.9	
7384	VOLVO	XC90 T6 AWD	SUV - STANDARD	2.0	4	AS8	Z	10.7	

7385 rows × 9 columns

```
In [48]: psdf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7385 entries, 0 to 7384
Data columns (total 9 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Make                          7385 non-null   object
 1   Model                         7385 non-null   object
 2   Vehicle Class                 7385 non-null   object
 3   Engine Size(L)                7385 non-null   float64
 4   Cylinders                     7385 non-null   int64
 5   Transmission                  7385 non-null   object
 6   Fuel Type                     7385 non-null   object
 7   Fuel Consumption Comb (L/100 km)  7385 non-null   float64
 8   CO2 Emissions(g/km)           7385 non-null   int64
dtypes: float64(2), int64(2), object(5)
memory usage: 519.4+ KB
```

In [49]: `psdf.describe(include="all").transpose()`

Out[49]:

	count	unique	top	freq	mean	std	min	25%	50%	75%
Make	7385	42	FORD	628	NaN	NaN	NaN	NaN	NaN	NaN
Model	7385	2053	F-150 FFV 4X4	32	NaN	NaN	NaN	NaN	NaN	NaN
Vehicle Class	7385	16	SUV - SMALL	1217	NaN	NaN	NaN	NaN	NaN	NaN
Engine Size(L)	7385.0	NaN	NaN	NaN	3.160068	1.35417	0.9	2.0	3.0	3.7
Cylinders	7385.0	NaN	NaN	NaN	5.61503	1.828307	3.0	4.0	6.0	6.0
Transmission	7385	27	AS6	1324	NaN	NaN	NaN	NaN	NaN	NaN
Fuel Type	7385	5	X	3637	NaN	NaN	NaN	NaN	NaN	NaN
Fuel Consumption Comb (L/100 km)	7385.0	NaN	NaN	NaN	10.975071	2.892506	4.1	8.9	10.6	12.6
CO2 Emissions(g/km)	7385.0	NaN	NaN	NaN	250.584699	58.512679	96.0	208.0	246.0	288.0

In [50]:
```python
a = sorted(list(psdf['Engine Size(L)']))
n = len(a)
print(n/2)
a[(n//2)+1]
```

```
3692.5
```

Out[50]: 3.0

1. What is the median engine size in liters?

The Engine size(L) column of the dataset describes the size of engine in litres that is used in the vehicle.

In the initial analysis, the column has the lowest and largest value of 0.9 and 8.4 respectively, with a mean of about 3.16.

```
In [51]: np.median(psdf['Engine Size(L)'])
```

```
Out[51]: 3.0
```

From the above code, it is clear that the median of Engine Size is **3.0**.

```
In [52]: a = list(psdf['Fuel Type'].unique())
         b = []
         for i in range(len(a)):
             b.append("{} : {}".format(a[i],list(psdf['Fuel Type']).count(a[i])))
         b
```

```
Out[52]: ['Z : 3202', 'D : 175', 'X : 3637', 'E : 370', 'N : 1']
```

Question 2 : What is the average fuel consumption for regular gasoline (Fuel Type = X), premium gasoline (Z), ethanol (E), and diesel (D)?

The Fuel Type field is a categorical variable with **5** unique values which are **"X", "Z", "E", "D", and "N"** occuring **3637, 3202, 370, 175 and 1** times respectively. The type is not asked in the question. Hence, we will exclude it while calculating the average fuel consumption.

```
In [53]: psdf[psdf['Fuel Type'] == 'N']
```

Out[53]:

	Make	Model	Vehicle Class	Engine Size(L)	Cylinders	Transmission	Fuel Type	Fuel Consumption Comb (L/100 km)	Emissio
2439	CHEVROLET	IMPALA DUAL FUEL	MID-SIZE	3.6	6	AS6	N	12.7	

```
In [54]: fTdf = psdf.groupby('Fuel Type').agg(np.average).reset_index(drop=False)
         fTdf = fTdf[fTdf['Fuel Type'] != 'N'].sort_values(ascending=True, by='Fuel Consumpt
         fTdf
```

```
C:\Users\apal6\AppData\Local\Temp\ipykernel_14236\1442872963.py:1: FutureWarning:
['Make', 'Model', 'Vehicle Class', 'Transmission'] did not aggregate successfully.
If any error is raised this will raise in a future version of pandas. Drop these c
olumns/ops to avoid this warning.
  fTdf = psdf.groupby('Fuel Type').agg(np.average).reset_index(drop=False)
```

	Fuel Type	Fuel Consumption Comb (L/100 km)
0	D	8.835429
3	X	10.084575
4	Z	11.422767
1	E	16.861351

In [55]:
```
colors=['#1fe074','#0069c0','#008ac5','#00a9b5','#00c698','#0045a5']
customPalette = sns.set_palette(sns.color_palette(colors))
sns.countplot(data=psdf,x='Fuel Type',palette=customPalette).set(xlabel='Fuel Type'
plt.show()
```
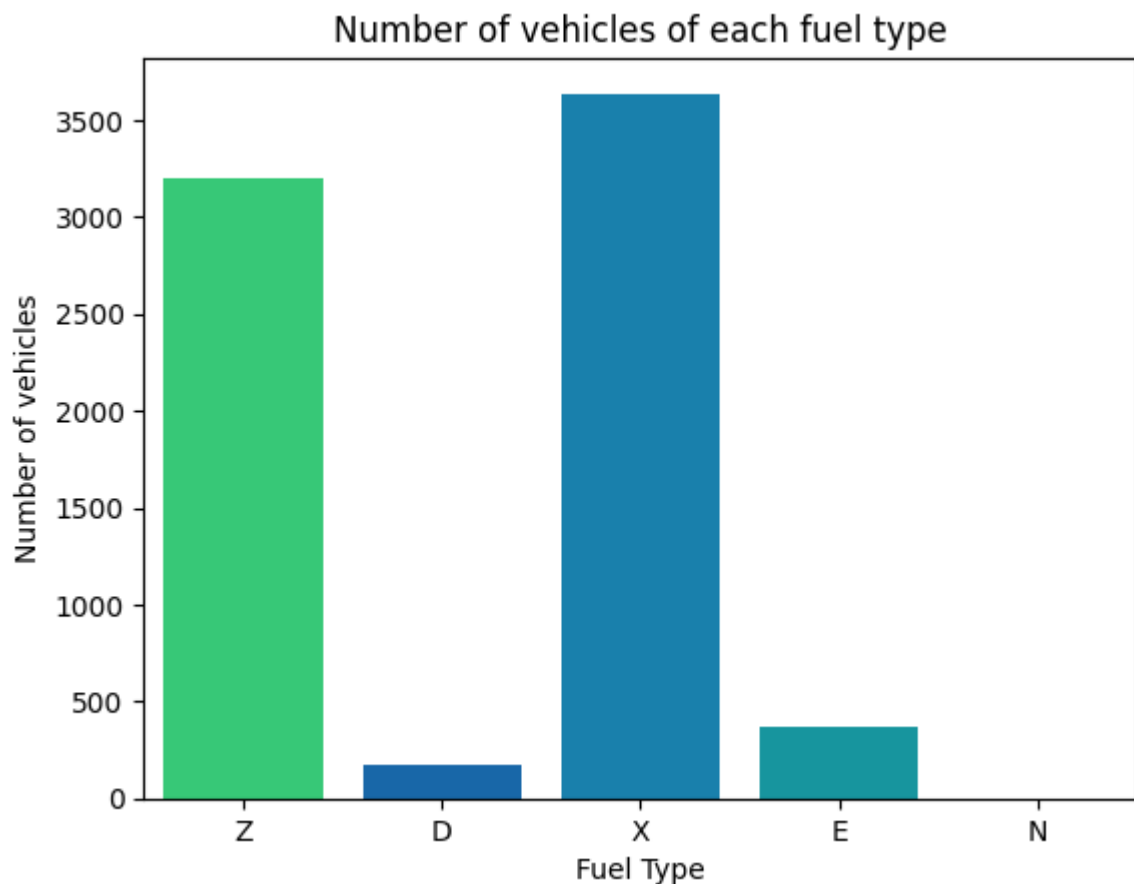

The graph shows the values of each fuel type in the dataset. Now we will calculate the average fuel consumption by grouping Fuel type.

The above tables shows the Average Fuel Consumption in (L/100 km) for vehicles of Fuel Type **"X", "Z", "E", and "D"** which is **10.08, 11.42, 16.86, and 8.84** after rounding off to the 2 decimal point.
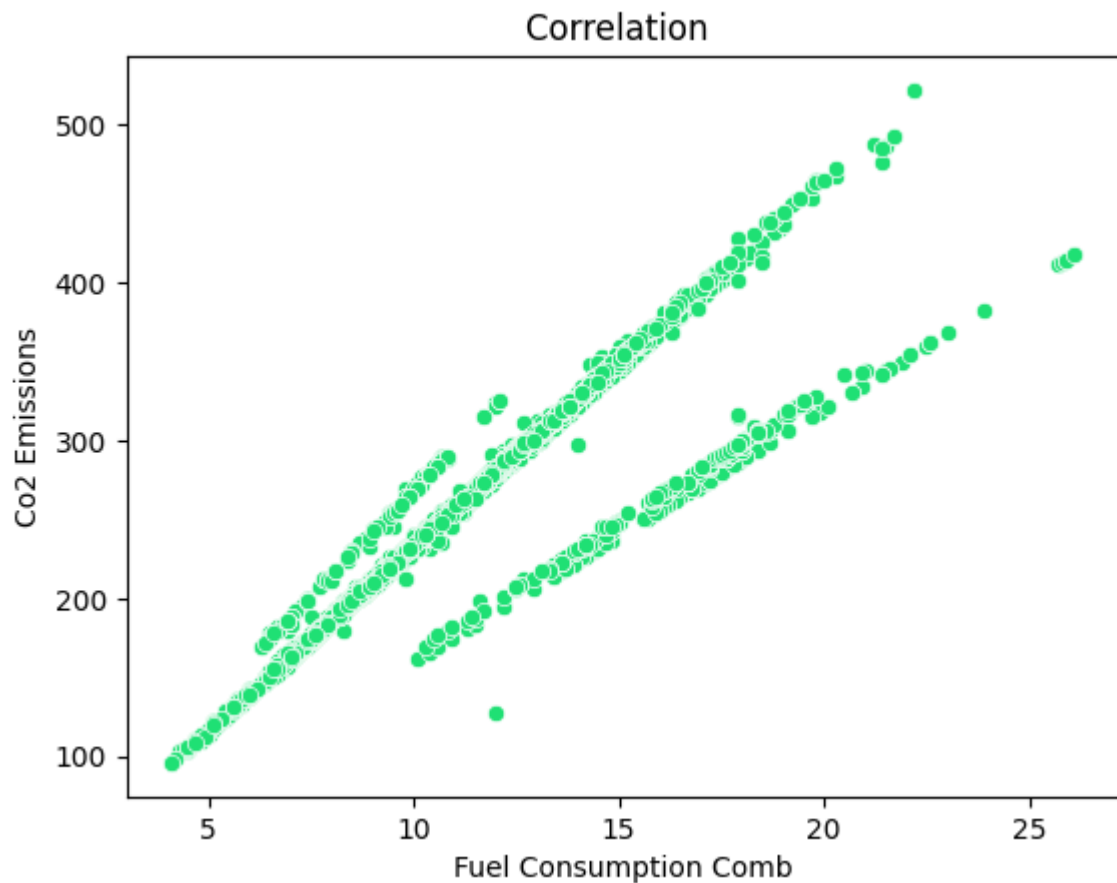
The Average Fuel Consumption is **lowest** for Fuel Type **"D"** and **highest** for Fuel Type **"E"**.

NOTE : The Fuel Type **"N"** only contains one record (row) which has a Fuel Consumption Value of **12.7**

What is the correlation between fuel consumption and CO2 emissions?

The following Scatterplot indicates a high correlation between Fuel Consumption and CO2 Emission. If you carefully see the scatterplot, you would notice three straight line diagonally in the plot.

```
In [56]: sns.scatterplot(data=psdf,x='Fuel Consumption Comb (L/100 km)',y='CO2 Emissions(g/k
         plt.show()
```



```
In [57]: sns.regplot(data=psdf,x='Fuel Consumption Comb (L/100 km)',y='CO2 Emissions(g/km)')
         plt.show()
```
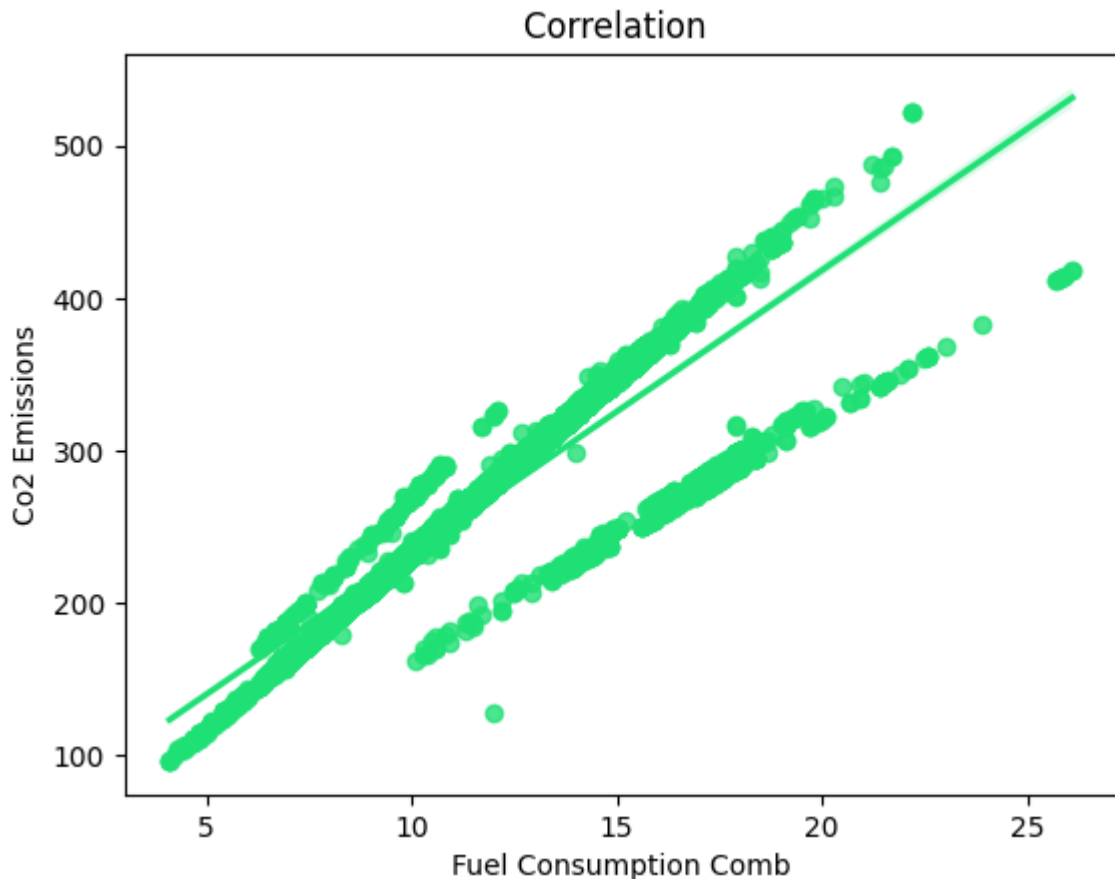
Correlation

```
In [58]:  round(psdf['Fuel Consumption Comb (L/100 km)'].corr(psdf['CO2 Emissions(g/km)']),2)
```

```
Out[58]:  0.92
```

From above, it is clear that **Fuel Consumption and Co2 Emissions** have highly positive correlation.

Which vehicle class has lower average CO2 emissions, 'SUV - SMALL' or 'MID-SIZE'?

```
In [59]:  lCo2Emi = psdf.sort_values(ascending=False,by="Vehicle Class")
          lCo2Emi = lCo2Emi.groupby('Vehicle Class').agg(np.mean).reset_index(drop=False)
          lCo2Emi = lCo2Emi[['Vehicle Class','CO2 Emissions(g/km)']]
          lCo2Emi2 = lCo2Emi[lCo2Emi['Vehicle Class'].isin(["MID-SIZE","SUV - SMALL"])]
          lCo2Emi2
```

```
C:\Users\apal6\AppData\Local\Temp\ipykernel_14236\3485578738.py:2: FutureWarning:
The operation <function mean at 0x000001FE7F488DC0> failed on a column. If any err
or is raised, this will raise an exception in a future version of pandas. Drop the
se columns to avoid this warning.
  lCo2Emi = lCo2Emi.groupby('Vehicle Class').agg(np.mean).reset_index(drop=False)
```

	Vehicle Class	CO2 Emissions(g/km)
2	MID-SIZE	222.455428
11	SUV - SMALL	236.292523

```python
MSco2 = lCo2Emi[lCo2Emi['Vehicle Class'].isin(["MID-SIZE"])]
SSco2 = lCo2Emi[lCo2Emi['Vehicle Class'].isin(["SUV - SMALL"])]
```

```python
MSco2
```

	Vehicle Class	CO2 Emissions(g/km)
2	MID-SIZE	222.455428

```python
SSco2
```

	Vehicle Class	CO2 Emissions(g/km)
11	SUV - SMALL	236.292523

```python
np.mean(MSco2['CO2 Emissions(g/km)'])
```

222.45542806707854

```python
np.mean(SSco2['CO2 Emissions(g/km)'])
```

236.29252259654888

```python
sns.barplot(x='Vehicle Class',y='CO2 Emissions(g/km)',data=lCo2Emi2).set(xlabel="Ve
plt.show()
```

Comparsions between MID-SIZE and SUV-SMALL Co2 Emissions
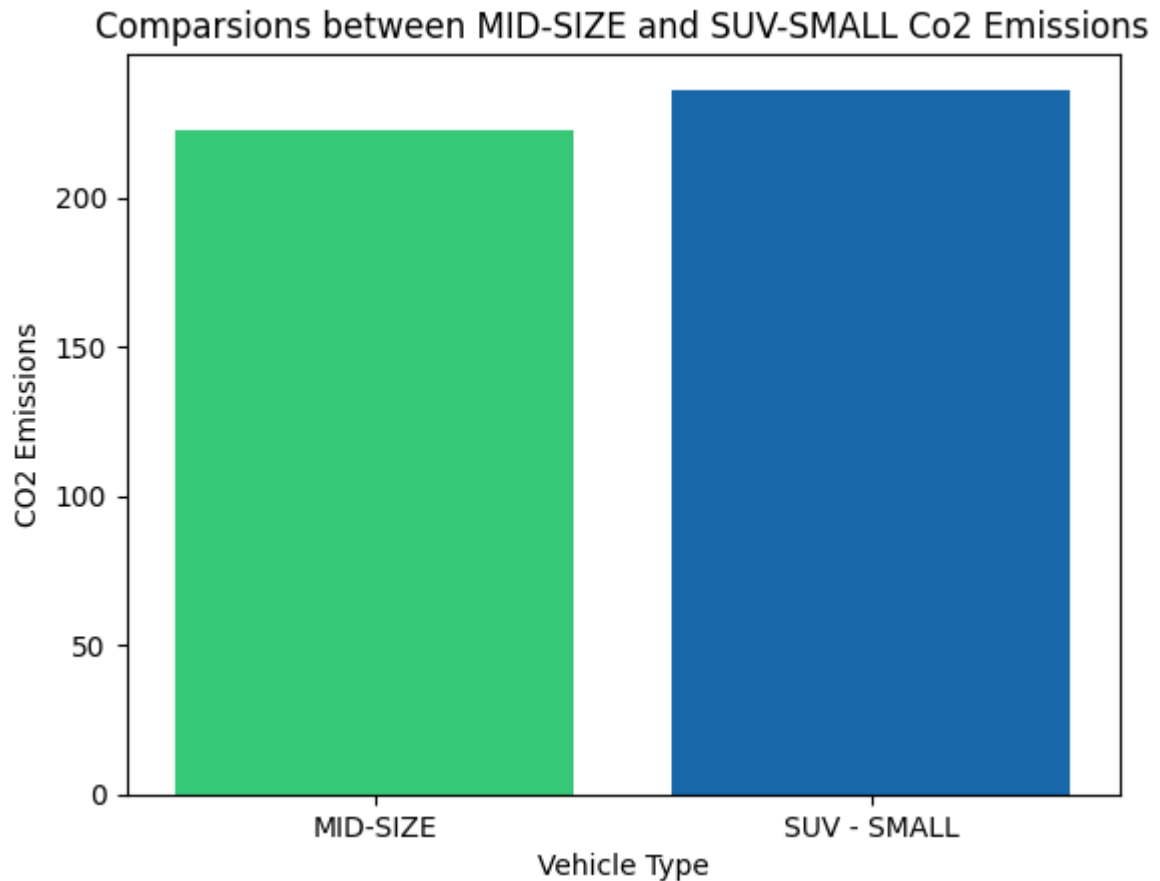
From above barplot and dataframe, it is clear that vehicle of **MID-SIZE** class have lower **CO2 Emissions** than vehicle of **SUV-SMALL** class.

The Average CO2 Emissions for MID-SIZE vehicle class is **222.45**.

The Average CO2 Emissions for SUV-SMALL vehicle class is **236.30**.

What are the average CO2 emissions for all vehicles?

```
In [66]:   np.mean(psdf['CO2 Emissions(g/km)'])
```

Out[66]:   250.58469871360867

The average CO2 Emissions for all the vehicles is **250.58**.

For vehicles with an engine size of 2.0 liters or smaller?

```
In [67]:   np.mean((psdf[psdf['Engine Size(L)'] <= 2])['CO2 Emissions(g/km)'])
```

Out[67]:   198.26783530370975

The average CO2 Emissions for all the vehicles with an engine size of 2.0 Litres or smaller is **198.27**.

Other Insights Calculations

The other insights that i have found during the analysis are :

1 - The CO2 Emissions are highest and lowest for **VAN-PASSENGER** and **STATION WAGON - SMALL** respectively.

The following barplot shows the difference.

```
In [68]: sns.barplot(x='Vehicle Class',y='CO2 Emissions(g/km)',data=lCo2Emi).set(xlabel="Veh
         plt.xticks(rotation = 90)
         plt.show()
```
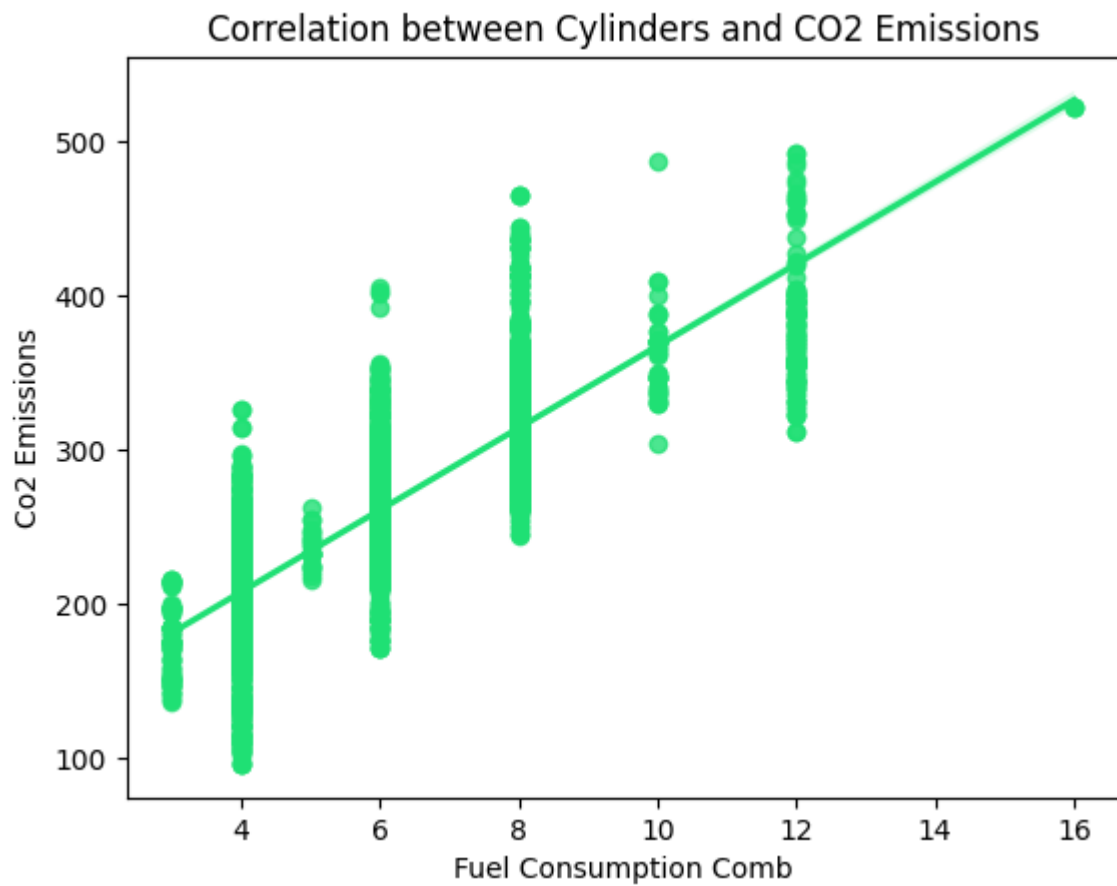

Comparsions of Co2 Emissions of different vehicle class
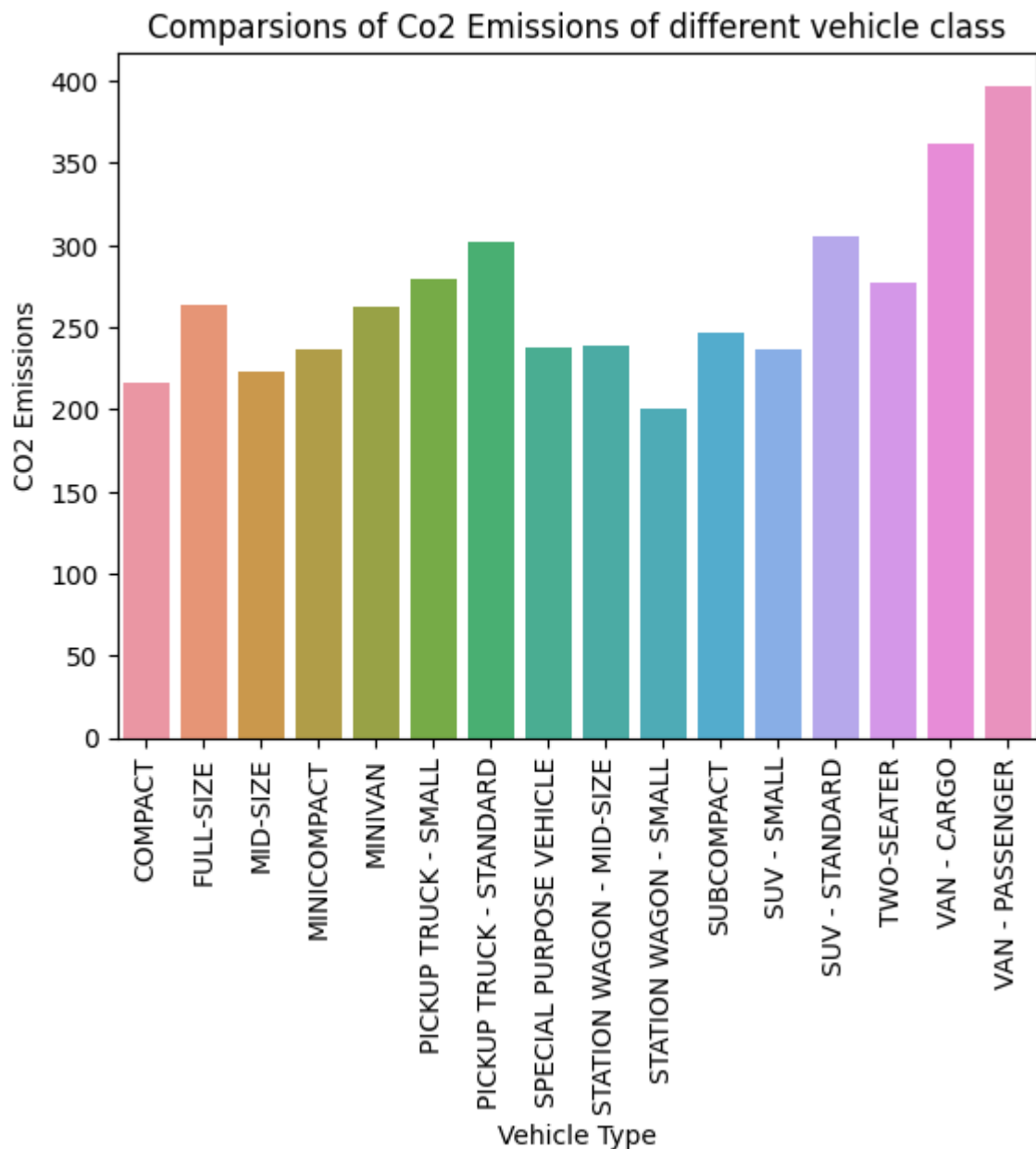
```
In [69]: round(psdf['Cylinders'].corr(psdf['CO2 Emissions(g/km)']),2)
```

```
Out[69]: 0.83
```

2 - The correlation between Cylinders and CO2 Emissions is **0.83**.

In [70]:
```python
sns.regplot(data=psdf,x='Cylinders',y='CO2 Emissions(g/km)').set(xlabel ="Fuel Cons
plt.show()
```


Correlation between Cylinders and CO2 Emissions

In [76]:
```python
!pip install nbconvert[webpdf]
```

```
Requirement already satisfied: nbconvert[webpdf] in c:\users\apal6\appdata\local\p
rograms\python\python310\lib\site-packages (7.2.8)
Requirement already satisfied: bleach in c:\users\apal6\appdata\local\programs\pyt
hon\python310\lib\site-packages (from nbconvert[webpdf]) (5.0.1)
Requirement already satisfied: defusedxml in c:\users\apal6\appdata\local\programs
\python\python310\lib\site-packages (from nbconvert[webpdf]) (0.7.1)
Requirement already satisfied: markupsafe>=2.0 in c:\users\apal6\appdata\local\pro
grams\python\python310\lib\site-packages (from nbconvert[webpdf]) (2.1.2)
Requirement already satisfied: traitlets>=5.0 in c:\users\apal6\appdata\local\prog
rams\python\python310\lib\site-packages (from nbconvert[webpdf]) (5.8.1)
Requirement already satisfied: jupyter-core>=4.7 in c:\users\apal6\appdata\local\p
rograms\python\python310\lib\site-packages (from nbconvert[webpdf]) (5.1.3)
Requirement already satisfied: nbformat>=5.1 in c:\users\apal6\appdata\local\progr
ams\python\python310\lib\site-packages (from nbconvert[webpdf]) (5.7.3)
Requirement already satisfied: pandocfilters>=1.4.1 in c:\users\apal6\appdata\loca
l\programs\python\python310\lib\site-packages (from nbconvert[webpdf]) (1.5.0)
Requirement already satisfied: jinja2>=3.0 in c:\users\apal6\appdata\local\program
s\python\python310\lib\site-packages (from nbconvert[webpdf]) (3.1.2)
Requirement already satisfied: packaging in c:\users\apal6\appdata\local\programs
\python\python310\lib\site-packages (from nbconvert[webpdf]) (23.0)
Requirement already satisfied: pygments>=2.4.1 in c:\users\apal6\appdata\local\pro
grams\python\python310\lib\site-packages (from nbconvert[webpdf]) (2.14.0)
Requirement already satisfied: tinycss2 in c:\users\apal6\appdata\local\programs\p
ython\python310\lib\site-packages (from nbconvert[webpdf]) (1.2.1)
Requirement already satisfied: nbclient>=0.5.0 in c:\users\apal6\appdata\local\pro
grams\python\python310\lib\site-packages (from nbconvert[webpdf]) (0.7.2)
Requirement already satisfied: beautifulsoup4 in c:\users\apal6\appdata\local\prog
rams\python\python310\lib\site-packages (from nbconvert[webpdf]) (4.11.1)
Requirement already satisfied: mistune<3,>=2.0.3 in c:\users\apal6\appdata\local\p
rograms\python\python310\lib\site-packages (from nbconvert[webpdf]) (2.0.4)
Requirement already satisfied: jupyterlab-pygments in c:\users\apal6\appdata\local
\programs\python\python310\lib\site-packages (from nbconvert[webpdf]) (0.2.2)
Collecting pyppeteer<1.1,>=1
  Downloading pyppeteer-1.0.2-py3-none-any.whl (83 kB)
     ------------------------------------- 83.4/83.4 kB 4.6 MB/s eta 0:00:00
Requirement already satisfied: pywin32>=1.0 in c:\users\apal6\appdata\local\progra
ms\python\python310\lib\site-packages (from jupyter-core>=4.7->nbconvert[webpdf])
(305)
Requirement already satisfied: platformdirs>=2.5 in c:\users\apal6\appdata\local\p
rograms\python\python310\lib\site-packages (from jupyter-core>=4.7->nbconvert[webp
df]) (2.6.2)
Requirement already satisfied: jupyter-client>=6.1.12 in c:\users\apal6\appdata\lo
cal\programs\python\python310\lib\site-packages (from nbclient>=0.5.0->nbconvert[w
ebpdf]) (7.4.9)
Requirement already satisfied: jsonschema>=2.6 in c:\users\apal6\appdata\local\pro
grams\python\python310\lib\site-packages (from nbformat>=5.1->nbconvert[webpdf])
(4.17.3)
Requirement already satisfied: fastjsonschema in c:\users\apal6\appdata\local\prog
rams\python\python310\lib\site-packages (from nbformat>=5.1->nbconvert[webpdf])
(2.16.2)
Collecting importlib-metadata>=1.4
  Downloading importlib_metadata-6.0.0-py3-none-any.whl (21 kB)
Collecting appdirs<2.0.0,>=1.4.3
  Downloading appdirs-1.4.4-py2.py3-none-any.whl (9.6 kB)
Collecting urllib3<2.0.0,>=1.25.8
  Downloading urllib3-1.26.14-py2.py3-none-any.whl (140 kB)
```

```
                   ------------------------------------ 140.6/140.6 kB 8.7 MB/s eta 0:00:00
Collecting websockets<11.0,>=10.0
  Downloading websockets-10.4-cp310-cp310-win_amd64.whl (101 kB)
                   ------------------------------------ 101.4/101.4 kB 5.7 MB/s eta 0:00:00
Collecting certifi>=2021
  Downloading certifi-2022.12.7-py3-none-any.whl (155 kB)
                   ------------------------------------ 155.3/155.3 kB ? eta 0:00:00
Collecting tqdm<5.0.0,>=4.42.1
  Downloading tqdm-4.64.1-py2.py3-none-any.whl (78 kB)
                   ------------------------------------ 78.5/78.5 kB ? eta 0:00:00
Collecting pyee<9.0.0,>=8.1.0
  Downloading pyee-8.2.2-py2.py3-none-any.whl (12 kB)
Requirement already satisfied: soupsieve>1.2 in c:\users\apal6\appdata\local\progr
ams\python\python310\lib\site-packages (from beautifulsoup4->nbconvert[webpdf])
(2.3.2.post1)
Requirement already satisfied: webencodings in c:\users\apal6\appdata\local\progra
ms\python\python310\lib\site-packages (from bleach->nbconvert[webpdf]) (0.5.1)
Requirement already satisfied: six>=1.9.0 in c:\users\apal6\appdata\local\programs
\python\python310\lib\site-packages (from bleach->nbconvert[webpdf]) (1.16.0)
Collecting zipp>=0.5
  Downloading zipp-3.11.0-py3-none-any.whl (6.6 kB)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in
c:\users\apal6\appdata\local\programs\python\python310\lib\site-packages (from jso
nschema>=2.6->nbformat>=5.1->nbconvert[webpdf]) (0.19.3)
Requirement already satisfied: attrs>=17.4.0 in c:\users\apal6\appdata\local\progr
ams\python\python310\lib\site-packages (from jsonschema>=2.6->nbformat>=5.1->nbcon
vert[webpdf]) (22.2.0)
Requirement already satisfied: entrypoints in c:\users\apal6\appdata\local\program
s\python\python310\lib\site-packages (from jupyter-client>=6.1.12->nbclient>=0.5.0
->nbconvert[webpdf]) (0.4)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\apal6\appdata\lo
cal\programs\python\python310\lib\site-packages (from jupyter-client>=6.1.12->nbcl
ient>=0.5.0->nbconvert[webpdf]) (2.8.2)
Requirement already satisfied: tornado>=6.2 in c:\users\apal6\appdata\local\progra
ms\python\python310\lib\site-packages (from jupyter-client>=6.1.12->nbclient>=0.5.
0->nbconvert[webpdf]) (6.2)
Requirement already satisfied: nest-asyncio>=1.5.4 in c:\users\apal6\appdata\local
\programs\python\python310\lib\site-packages (from jupyter-client>=6.1.12->nbclien
t>=0.5.0->nbconvert[webpdf]) (1.5.6)
Requirement already satisfied: pyzmq>=23.0 in c:\users\apal6\appdata\local\program
s\python\python310\lib\site-packages (from jupyter-client>=6.1.12->nbclient>=0.5.0
->nbconvert[webpdf]) (25.0.0)
Requirement already satisfied: colorama in c:\users\apal6\appdata\local\programs\p
ython\python310\lib\site-packages (from tqdm<5.0.0,>=4.42.1->pyppeteer<1.1,>=1->nb
convert[webpdf]) (0.4.6)
Installing collected packages: pyee, appdirs, zipp, websockets, urllib3, tqdm, cer
tifi, importlib-metadata, pyppeteer
Successfully installed appdirs-1.4.4 certifi-2022.12.7 importlib-metadata-6.0.0 py
ee-8.2.2 pyppeteer-1.0.2 tqdm-4.64.1 urllib3-1.26.14 websockets-10.4 zipp-3.11.0
```

```
In [ ]:  jupyter nbconvert --to webpdf --allow-chromium-download .ipynb
```