Multi linear regression

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [53]:

```
#--- importing data
df = pd.read_excel('D:\\1\\co2.xlsx')
df
```

Out[53]:

	MODEL	MAKE	MODEL.1	VEHICLE CLASS	ENGINESIZE	CYLINDERS	TRANSMISSION	FUEL	FUELCONSUMPTION	Unnamed: 9	Unnamed: 10	Unna
0	2002	ACURA	1.7EL	COMPACT	1.7	4	A4	X	9.5	7.3	8.5	
1	2002	ACURA	1.7EL	COMPACT	1.7	4	M5	X	8.8	7.2	8.1	
2	2002	ACURA	3.2CL	COMPACT	3.2	6	AS5	Z	13.6	8.8	11.4	
3	2002	ACURA	3.2TL	MID-SIZE	3.2	6	AS5	Z	13.5	8.8	11.4	
4	2002	ACURA	3.5RL	MID-SIZE	3.5	6	A4	Z	15.0	10.7	13.1	
...	...	...	...	...	...	...	...	...	...	...	...	
735	2002	VOLVO	V70 T5 WAGON TURBO	STATION WAGON - MID-SIZE	2.3	5	AS5	Z	13.4	9.9	11.8	
736	2002	VOLVO	V70 T5 WAGON TURBO	STATION WAGON - MID-SIZE	2.3	5	M5	Z	12.9	9.5	11.3	
737	2002	VOLVO	V70 WAGON	STATION WAGON - MID-SIZE	2.4	5	A5	Z	12.7	9.1	11.1	
738	2002	VOLVO	V70 WAGON	STATION WAGON - MID-SIZE	2.4	5	M5	Z	12.5	9.3	11.0	
739	2002	VOLVO	V70 XC AWD TURBO	SUV	2.4	5	AS5	Z	14.2	10.5	12.5	

740 rows × 13 columns

In [55]:

```
mdf = df[['ENGINESIZE','CYLINDERS','FUELCONSUMPTION','CO2EMISSIONS']]
mdf
```

Out[55]:

	ENGINESIZE	CYLINDERS	FUELCONSUMPTION	CO2EMISSIONS
0	1.7	4	9.5	196
1	1.7	4	8.8	186
2	3.2	6	13.6	262
3	3.2	6	13.5	262
4	3.5	6	15.0	301
...	...	...	...	...
735	2.3	5	13.4	271
736	2.3	5	12.9	260
737	2.4	5	12.7	255
738	2.4	5	12.5	253
739	2.4	5	14.2	288

740 rows × 4 columns

In [56]:

```
print(mdf.describe())
print('======================================================================')
mdf.max()
```

```
       ENGINESIZE   CYLINDERS  FUELCONSUMPTION  CO2EMISSIONS
count  740.000000  740.000000       740.000000    740.000000
mean     3.380000    5.906757        15.065946    300.747297
std      1.262557    1.552233         3.495449     65.574775
min      1.000000    3.000000         4.900000    104.000000
25%      2.300000    4.000000        12.800000    255.000000
50%      3.200000    6.000000        14.500000    288.000000
75%      4.300000    6.000000        17.200000    350.500000
max      8.000000   12.000000        26.500000    481.000000
======================================================================
```

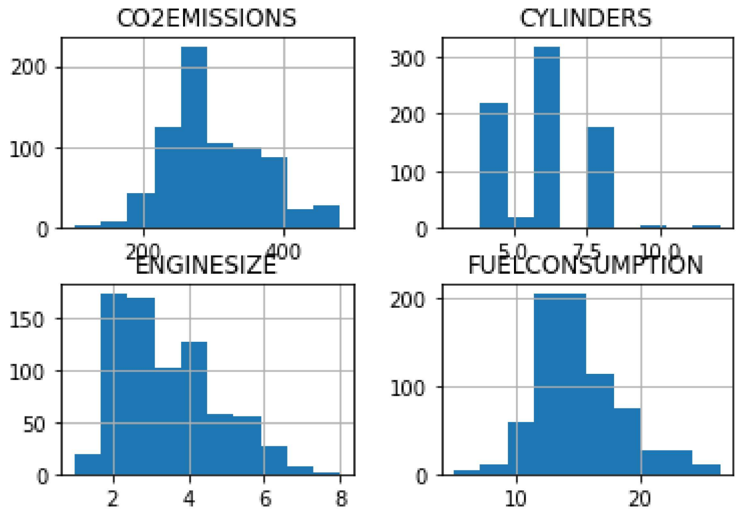
Out[56]:

```
ENGINESIZE           8.0
CYLINDERS           12.0
FUELCONSUMPTION     26.5
CO2EMISSIONS       481.0
dtype: float64
```

In [57]:

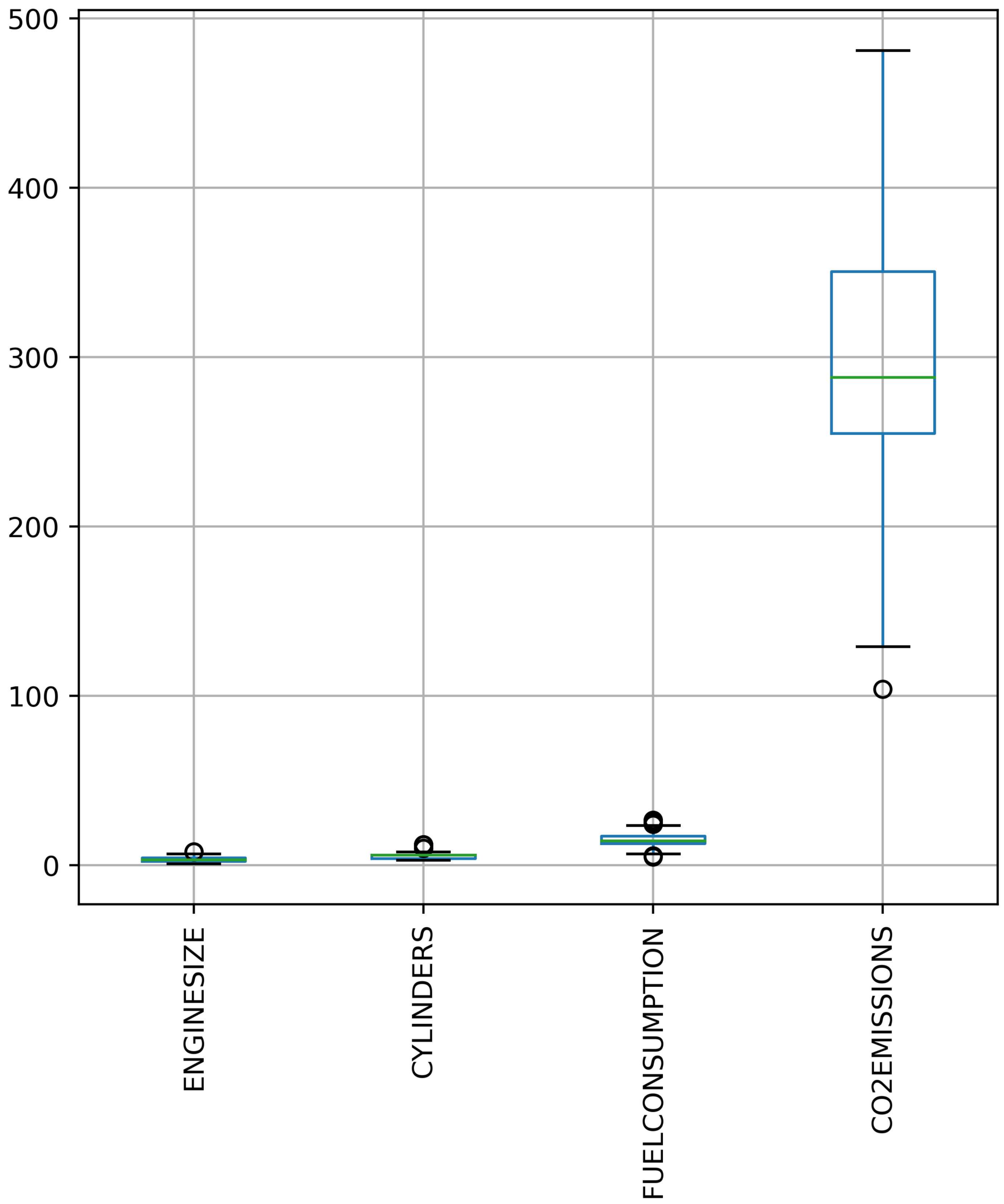
```
mdf.hist()

plt.show()
```



In [58]:

```
plt.figure(figsize=(6,6),dpi=500)
mdf.boxplot()
plt.xticks(rotation='vertical')
plt.show()
```



In [59]:

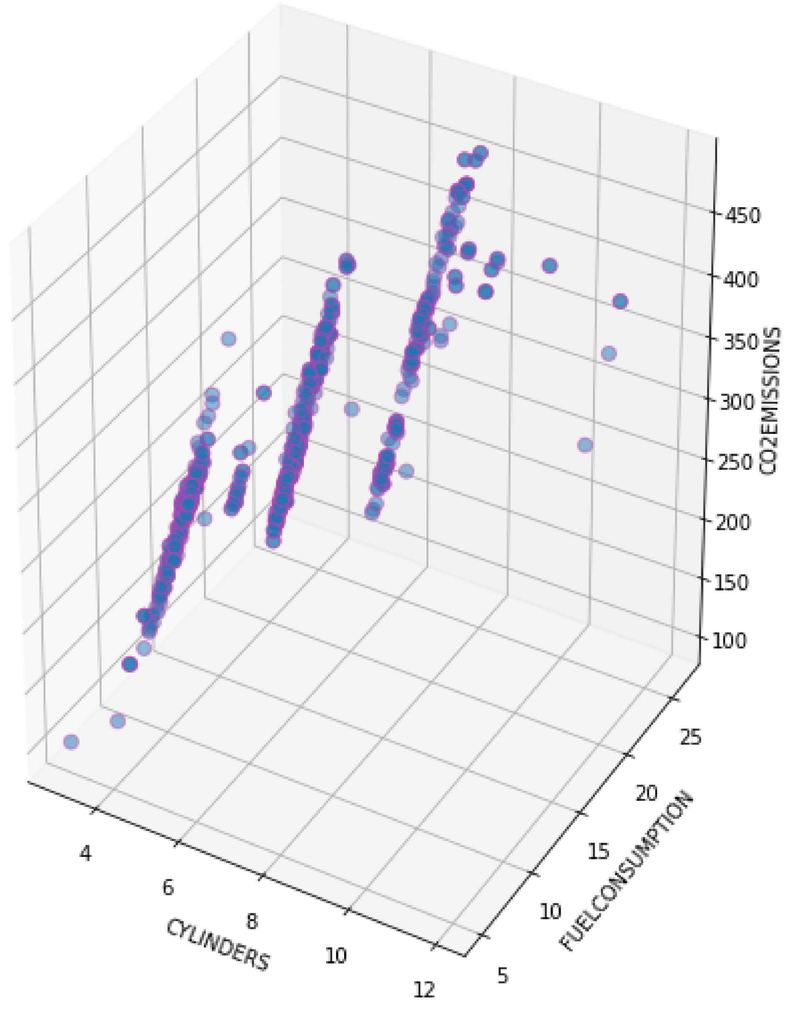
```
Rmdf = mdf[['CYLINDERS','FUELCONSUMPTION','CO2EMISSIONS']]

fig = plt.figure(figsize=(8,10))
ax = fig.add_subplot(111, projection='3d')

x_axis =  mdf['CYLINDERS']
y_axis = mdf['FUELCONSUMPTION']
z_axis = mdf['CO2EMISSIONS']
ax.scatter(x_axis, y_axis, z_axis, s=50, alpha=0.5, edgecolors='#c443b1')

ax.set_xlabel('CYLINDERS')
ax.set_ylabel('FUELCONSUMPTION')
ax.set_zlabel('CO2EMISSIONS')
plt.savefig('D:\\1\\3Dplot.png', dpi=500)

plt.show()
```



spliting out data

In [65]:

```
msk = np.random.rand(len(df)) < 0.8
train = mdf[msk]
test = mdf[~msk]
```

In [66]:

```
from sklearn import linear_model
M_regression = linear_model.LinearRegression()
x = np.asanyarray(train[['CYLINDERS','FUELCONSUMPTION']])
y = np.asanyarray(train[['CO2EMISSIONS']])
M_regression.fit(x,y)
print ('Coefficients: ', M_regression.coef_)
```

```
Coefficients:  [[ 3.63780295 16.90007071]]
```

In [67]:

```
## another way without sklearn
import matplotlib.pyplot as plt
import statsmodels.formula.api as smf
model = smf.ols(formula='CO2EMISSIONS ~ CYLINDERS + FUELCONSUMPTION', data=df2)
results_formula = model.fit()
results_formula.params
```

Out[67]:

```
Intercept          29.465722
CYLINDERS           3.975922
FUELCONSUMPTION    16.447475
dtype: float64
```

In [75]:

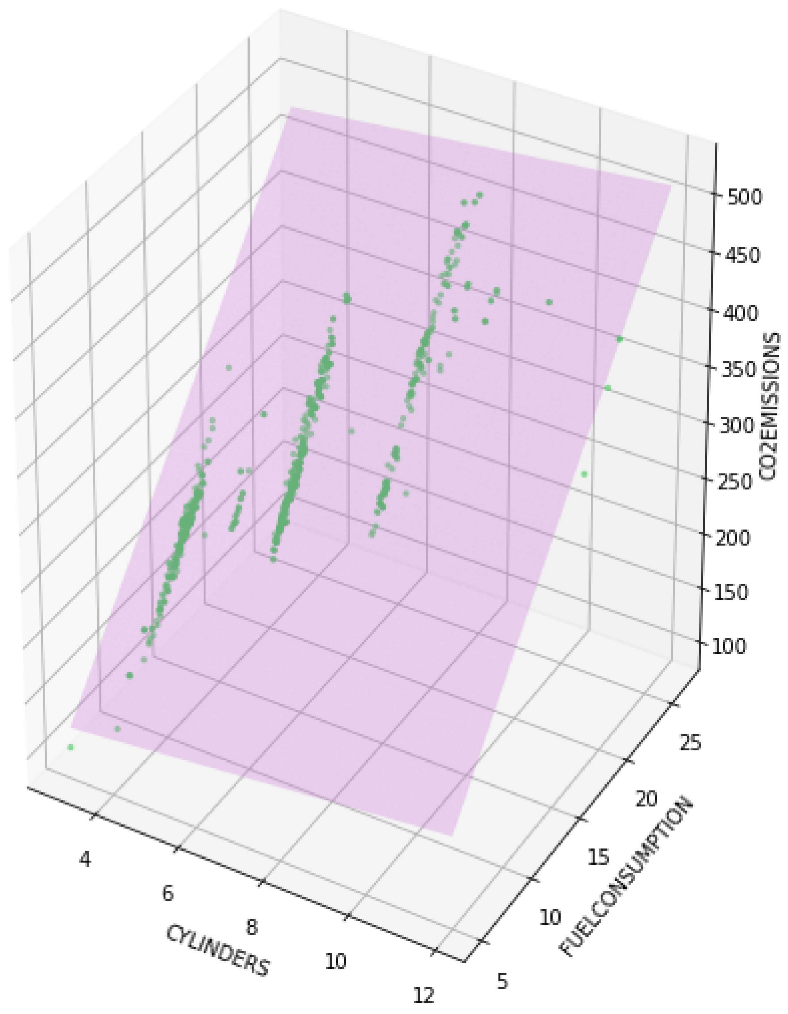
```
x_surf, y_surf = np.meshgrid(np.linspace(mdf.CYLINDERS.min(), mdf.CYLINDERS.max(), 100),np.linspace(mdf.FUELCONSUMPTIO
N.min(), mdf.FUELCONSUMPTION.max(), 100))
onlyX = pd.DataFrame({'CYLINDERS': x_surf.ravel(), 'FUELCONSUMPTION': y_surf.ravel()})
fittedY=results_formula.predict(exog=onlyX)

fittedY=np.array(fittedY)


fig = plt.figure(figsize=(8,10))

ax = fig.add_subplot(111, projection='3d')
ax.scatter(mdf['CYLINDERS'],mdf['FUELCONSUMPTION'],mdf['CO2EMISSIONS'],c='#3fcc54', marker='.', alpha=0.5)
ax.plot_surface(x_surf,y_surf,fittedY.reshape(x_surf.shape), color='#d97ce6', alpha=0.3)
ax.set_xlabel('CYLINDERS')
ax.set_ylabel('FUELCONSUMPTION')
ax.set_zlabel('CO2EMISSIONS')
plt.savefig('D:\\1\\3Dplot2.png', dpi=500)

plt.show()
```


CO2EMISSIONS
FUELCONSUMPTION
CYLINDERS