

Experiment No. 2

Code:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score

df = pd.read_csv(r'FuelConsumptionCo2.csv')

df = df[['ENGINE SIZE','CYLINDERS','FUELCONSUMPTION_CITY','FUELCONSUMPTION_HWY','FUELCONSUMPTION_COMB','FUELCONSUMPTION_COMB_MPG','CO2EMISSIONS']]

plt.scatter(df.ENGINE SIZE, df.CO2EMISSIONS, color='blue')
plt.xlabel("Engine size")
plt.ylabel("Emission")
plt.show()

X = df.iloc[:, :-1]
y = df.iloc[:, -1]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20)

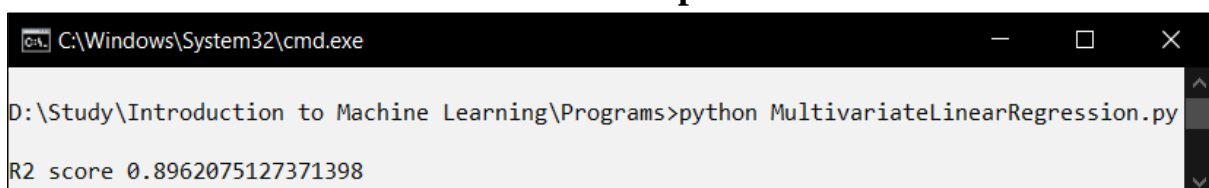
reg = LinearRegression()
reg.fit(X_train, y_train)
y_pred = reg.predict(X_test)

plt.scatter(X_test.ENGINE SIZE, y_pred, color='blue')
plt.xlabel("Engine size")
plt.ylabel("Emission")
plt.show()

print('R2 score', r2_score(y_test, y_pred))
```

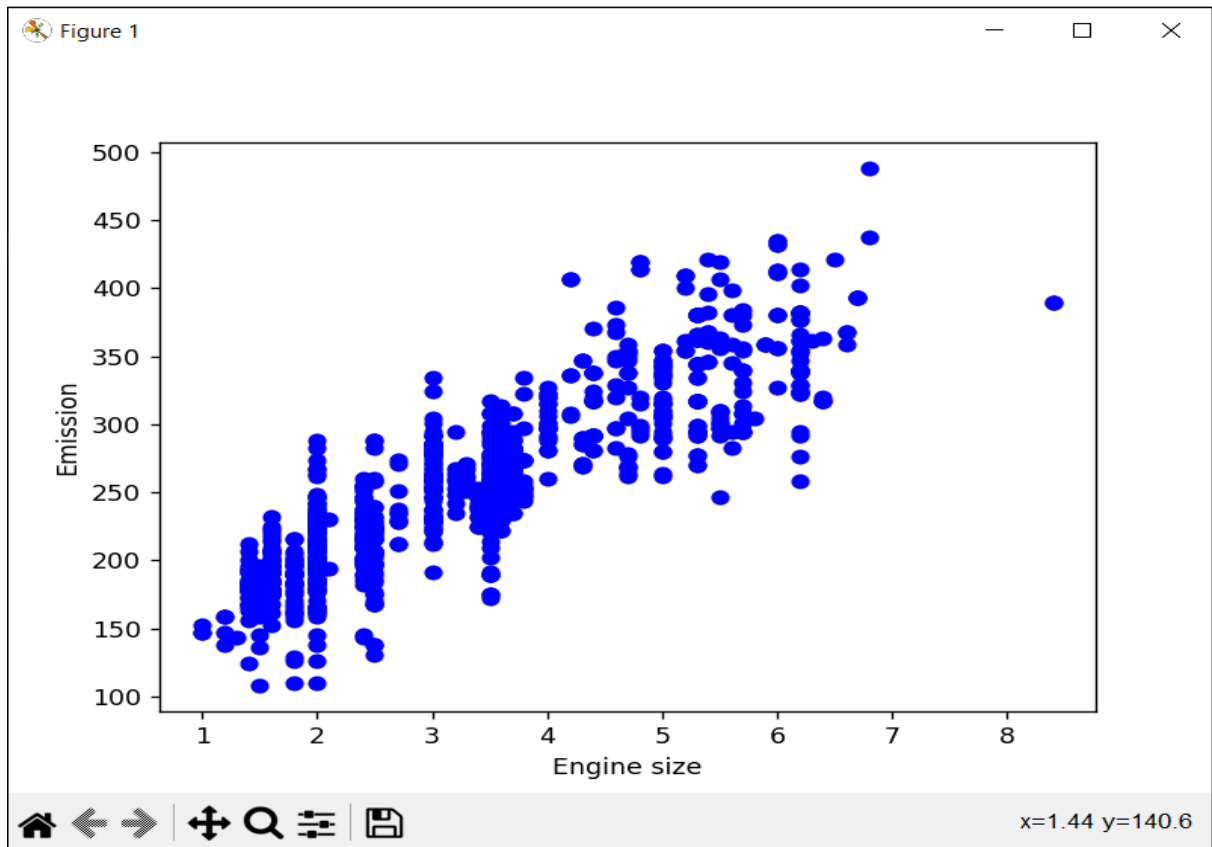
Output:

R2 Score for the test and predicted values



```
C:\Windows\System32\cmd.exe
D:\Study\Introduction to Machine Learning\Programs>python MultivariateLinearRegression.py
R2 score 0.8962075127371398
```

Scatter-plot of all observations of Emission and Engine size.



Scatter-plot of all observations of Emission and Engine size.

