Simple Linear Regression

```
import matplotlib.pyplot as plt
import pandas as pg
import pylab as pl
import numpy as np
```

upload csv file

```
df = pg.read_csv("FuelConsumptionCo2.csv")
df.head()


df.describe()

cdf=df[['ENGINESIZE','CYLINDERS','FUELCONSUMPTION_COMB','CO2EMISSIONS']]
cdf.head(9)

viz=cdf[['CYLINDERS','ENGINESIZE','CO2EMISSIONS','FUELCONSUMPTION_COMB']]
viz.hist()
plt.show()

plt.scatter(cdf.ENGINESIZE,cdf.CO2EMISSIONS,color="blue")
plt.xlabel("ENGINESIZE")
plt.ylabel("EMISSION")
plt.show()

plt.scatter(cdf.FUELCONSUMPTION_COMB,cdf.CO2EMISSIONS,color="blue")
plt.xlabel("FUELCONSUMPTION_COMB")
plt.ylabel("EMISSION")
plt.show

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


from sklearn import linear_model
regr=linear_model.LinearRegression()
train_x=np.asanyarray(train[['ENGINESIZE']])
train_y=np.asanyarray(train[['CO2EMISSIONS']])
regr.fit(train_x,train_y)
#The coefficients
print('Coefficient:',regr.coef_)
print('Intercept:',regr.intercept_)

from sklearn.metrics import r2_score
test_x=np.asanyarray(train[['ENGINESIZE']])
test_y=np.asanyarray(train[['CO2EMISSIONS']])
test_y_=regr.predict(test_x)
print("Mean absolute error:%.2f"%np.mean(np.absolute(test_y_-test_y)))
print("Residual sum of squares(MSE):%2f"%np.mean((test_y_-test_y)**2))
print("R22-score:%.2f"%r2_score(test_y,test_y))
```