


```
# Kết nối Google Colab và Google Drive
from google.colab import drive
drive.mount('/content/drive')

↻ Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
import pandas as pd
from sklearn import linear_model
# Bước 1: Thu thập dữ liệu từ file csv
data_frame = pd.read_csv('/content/drive/MyDrive/Machine Learning/car.csv')
data_frame
```

↻

	Car	Model	Volume	Weight	CO2	
0	Toyota	Vios	1000	780	99	
1	Mitsubishi	Xpander	1200	1160	95	
2	Skoda	Citigo	1000	929	95	
3	Fiat	500	900	865	90	
4	Mini	Cooper	1500	1140	105	
5	VW	Up!	1000	929	105	
6	Skoda	Fabia	1400	1109	90	
7	Mercedes	A-Class	1500	1365	92	
8	Ford	Fiesta	1500	1112	98	
9	Audi	A1	1600	1150	99	
10	Hyundai	I20	1100	980	99	
11	Suzuki	Swift	1300	990	101	
12	Ford	Fiesta	1000	1112	99	
13	Honda	Civic	1600	1252	94	
14	Hundai	I30	1600	1326	97	
15	Opel	Astra	1600	1330	97	
16	BMW	1	1600	1365	99	
17	Mazda	3	2200	1280	104	
18	Skoda	Rapid	1600	1119	104	
19	Ford	Focus	2000	1328	105	
20	Ford	Mondeo	1600	1584	94	
21	Opel	Insignia	2000	1428	99	
22	Mercedes	C-Class	2100	1365	99	
23	Skoda	Octavia	1600	1415	99	
24	Volvo	S60	2000	1415	99	
25	Mercedes	CLA	1500	1465	102	
26	Audi	A4	2000	1490	104	
27	Audi	A6	2000	1725	114	
28	Volvo	V70	1600	1523	109	
29	BMW	5	2000	1705	114	
30	Mercedes	E-Class	2100	1605	115	
31	Volvo	XC70	2000	1746	117	
32	Ford	B-Max	1600	1235	104	
33	BMW	2	1600	1390	108	
34	Opel	Zafira	1600	1405	109	
35	Mercedes	SLK	2500	1395	120	

Các bước tiếp theo:

[Tạo mã bằng data_frame](#)[Xem các đồ thị được đề xuất](#)[New interactive sheet](#)

```

#Bước 2.
# Lấy dữ liệu features = input = independent values
X = data_frame[['Weight', 'Volume']]
# Lấy dữ liệu label = output = dependent values
y = data_frame['CO2']
# Bước 3. Sử dụng mô hình Linear Regression để huấn luyện
linear_regression = linear_model.LinearRegression()
linear_regression = linear_regression.fit(X.values, y.values)
# Dự đoán kết quả predict
predict_co2 = linear_regression.predict([[2300, 1300]])
print(predict_co2)

↗ [107.11988043]

#bai1
import pandas as pd
from sklearn.linear_model import LinearRegression

# Đọc dữ liệu từ file CSV
df = pd.read_csv('/content/drive/MyDrive/Machine Learning/ file.csv')

# Chuẩn bị dữ liệu huấn luyện
X = df[["Test score", "IQ test score", "English score", "Interview score", "Years of experiences"]]
y = df["Salary($) per month"]

# Huấn luyện mô hình Linear Regression
model = LinearRegression()
model.fit(X, y)

# Dữ liệu dự đoán cho nhân sự mới
new_employee = pd.DataFrame(
    [[5, 6, 7, 8, 2]],
    columns=["Test score", "IQ test score", "English score", "Interview score", "Years of experiences"]
)

# Dự đoán lương
predicted_salary = model.predict(new_employee)[0]

# In kết quả
print(f"Lương dự đoán cho nhân sự mới là: {predicted_salary:.2f} USD/tháng")

#b
# Tính giá trị trung bình của cột "Test score"
mean_test_score = df["Test score"].mean()

# In kết quả
print(f"Giá trị trung bình của 'Test score' là: {mean_test_score}")

#c
# Tính mode cho cột "Years of experiences"
mode_experience = stats.mode(data["Years of experiences"], keepdims=True)

# In kết quả
print("Giá trị xuất hiện nhiều nhất (mode) trong 'Years of experiences' là:", mode_experience.mode[0])
print("Số lần xuất hiện:", mode_experience.count[0])

#d
# Tính số trung vị (Median) của cột "English score"
median_english = data["English score"].median()

# In kết quả
print("Số trung vị của 'English score' là:", median_english)

↗ Lương dự đoán cho nhân sự mới là: 3207.86 USD/tháng
Giá trị trung bình của 'Test score' là: 5.875
Giá trị xuất hiện nhiều nhất (mode) trong 'Years of experiences' là: 2
Số lần xuất hiện: 11
Số trung vị của 'English score' là: 7.0

#bai2
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression

```

```
from scipy import stats
from google.colab import drive

# Kết nối Google Drive (nếu dùng Colab)
drive.mount('/content/drive')

# Đọc dữ liệu
df = pd.read_csv('/content/drive/My Drive/Machine Learning/Advertising.csv')

# a. Dự đoán doanh số bán hàng (Sales) dựa vào TV, Radio, Newspaper
df_clean = df.dropna()
X = df_clean[["TV", "Radio", "Newspaper"]]
y = df_clean["Sales"]

model = LinearRegression()
model.fit(X, y)

# Dữ liệu dự đoán mới
new_data = pd.DataFrame([{"TV": 200,
                           "Radio": 25,
                           "Newspaper": 50}])

predicted_sales = model.predict(new_data)[0]
print("a. Dự đoán doanh số Sales:", round(predicted_sales, 2))

# b. Mean của cột "TV"
mean_tv = df["TV"].mean()
print("b. Mean TV:", mean_tv)

# c. Mode của cột "Radio"
mode_radio = stats.mode(df["Radio"], keepdims=True).mode[0]
print("c. Mode Radio:", mode_radio)

# d. Median của cột "Newspaper"
median_newspaper = df["Newspaper"].median()
print("d. Median Newspaper:", median_newspaper)
```

 Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

- a. Dự đoán doanh số Sales: 16.75
- b. Mean TV: 147.0425
- c. Mode Radio: 4.1
- d. Median Newspaper: 25.75