```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error


from google.colab import files
uploaded = files.upload()
```

```
Choose Files petrol_consumption.csv
- petrol_consumption.csv(text/csv) - 1211 bytes, last modified: 3/24/2025 - 100% done
Saving petrol_consumption.csv to petrol_consumption.csv
```

```
df = pd.read_csv('petrol_consumption.csv')


df
```

	Petrol_tax	Average_income	Paved_Highways	Population_Driver_licence(%)	Petrol_Consumption
0	9.00	3571	1976	0.525	541
1	9.00	4092	1250	0.572	524
2	9.00	3865	1586	0.580	561
3	7.50	4870	2351	0.529	414
4	8.00	4399	431	0.544	410
5	10.00	5342	1333	0.571	457
6	8.00	5319	11868	0.451	344
7	8.00	5126	2138	0.553	467
8	8.00	4447	8577	0.529	464
9	7.00	4512	8507	0.552	498
10	8.00	4391	5939	0.530	580
11	7.50	5126	14186	0.525	471
12	7.00	4817	6930	0.574	525
13	7.00	4207	6580	0.545	508
14	7.00	4332	8159	0.608	566
15	7.00	4318	10340	0.586	635
16	7.00	4206	8508	0.572	603
17	7.00	3718	4725	0.540	714
18	7.00	4716	5915	0.724	865
19	8.50	4341	6010	0.677	640
20	7.00	4593	7834	0.663	649
21	8.00	4983	602	0.602	540
22	9.00	4897	2449	0.511	464
23	9.00	4258	4686	0.517	547
24	8.50	4574	2619	0.551	460
25	9.00	3721	4746	0.544	566
26	8.00	3448	5399	0.548	577
27	7.50	3846	9061	0.579	631
28	8.00	4188	5975	0.563	574
29	9.00	3601	4650	0.493	534
30	7.00	3640	6905	0.518	571
31	7.00	3333	6594	0.513	554
32	8.00	3063	6524	0.578	577
33	7.50	3357	4121	0.547	628
34	8.00	3528	3495	0.487	487
35	6.58	3802	7834	0.629	644
36	5.00	4045	17782	0.566	640
37	7.00	3897	6385	0.586	704
38	8.50	3635	3274	0.663	648
39	7.00	4345	3905	0.672	968
40	7.00	4449	4639	0.626	587
41	7.00	3656	3985	0.563	699
42	7.00	4300	3635	0.603	632
43	7.00	3745	2611	0.508	591
44	6.00	5215	2302	0.672	782
45	9.00	4476	3942	0.571	510

46	7.00	4296	4083	0.623	610
47	7.00	5002	9794	0.593	524

Next steps: Generate code with df | View recommended plots | New interactive sheet

```
X = df.drop(columns=['Petrol_Consumption'])  # Assuming 'Petrol_Consumption' is the target column
y = df['Petrol_Consumption']

# Split dataset (80% training, 20% testing)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)


model = DecisionTreeRegressor()
model.fit(X_train, y_train)

# Predict on test data
y_pred = model.predict(X_test)

# Evaluate the model


mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)

print(f"Mean Absolute Error (MAE): {mae}")
print(f"Mean Squared Error (MSE): {mse}")
print(f"Root Mean Squared Error (RMSE): {rmse}")
```

```
Mean Absolute Error (MAE): 91.2
Mean Squared Error (MSE): 17170.8
Root Mean Squared Error (RMSE): 131.03739924159058
```

```
import matplotlib.pyplot as plt
from sklearn.tree import plot_tree
feature_importance = model.feature_importances_
features = X.columns

# Print feature importance
print("\nFeature Importance:")
for feature, importance in zip(features, feature_importance):
    print(f"{feature}: {importance:.4f}")

# Plot feature importance
plt.figure(figsize=(8, 5))
plt.barh(features, feature_importance, color='skyblue')
plt.xlabel("Feature Importance Score")
plt.ylabel("Features")
plt.title("Feature Importance for Petrol Consumption Prediction")
plt.show()

# Visualize the Regression Tree Structure
plt.figure(figsize=(12, 6))
plot_tree(model, feature_names=features, filled=True, rounded=True)
plt.show()
```