Step-by-Step Linear Regression Model - Canada Per Capita Income

Step 1: Import Necessary Libraries

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt


from sklearn.model_selection import train_test_split

from sklearn.linear_model import LinearRegression

from sklearn.metrics import mean_squared_error, r2_score
```

Step 2: Load the Dataset

```
df = pd.read_excel("/mnt/data/canada_per_capita_income.xlsx")

print(df.head())
```

Step 3: Check Data Types & Structure

```
print(df.info())
```

Step 4: Handle Missing Values (if any)

```
print(df.isnull().sum())
```

Step 5: Split Data into X (features) and y (target)

```
X = df[['year']]  # feature

y = df['income']  # target
```

Step 6: Split Data into Training & Testing Sets

```
X_train, X_test, y_train, y_test = train_test_split(

    X, y, test_size=0.2, random_state=42

)
```

Step 7: Train the Linear Regression Model

```
lr_model = LinearRegression()

lr_model.fit(X_train, y_train)
```

Step 8: Make Predictions

```
y_pred = lr_model.predict(X_test)
```

Step 9: Visualize Results

```
plt.scatter(X, y, color='blue', label='Actual Data')

plt.plot(X_test, y_pred, color='red', label='Regression Line')

plt.xlabel('Year')

plt.ylabel('Income')

plt.title('Linear Regression - Canada Per Capita Income')

plt.legend()

plt.grid(True)

plt.tight_layout()

plt.show()
```

Step 10: Evaluate the Model

```
mse = mean_squared_error(y_test, y_pred)

r2 = r2_score(y_test, y_pred)


print("Mean Squared Error:", mse)

print("R² Score:", r2)
```

Evaluation Output

```
Mean Squared Error: 15147815.5

R² Score: 0.875
```