

## Step-by-Step XGBoost Regression Model - Canada Per Capita Income

### Step 1: Prepare the Data

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

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

### Step 2: Train/Test Split

```
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(  
    X, y, test_size=0.2, random_state=42  
)
```

### Step 3: Fit XGBoost Regressor

```
from xgboost import XGBRegressor
```

```
xgb_model = XGBRegressor(n_estimators=100, random_state=42)  
xgb_model.fit(X_train, y_train)
```

### Step 4: Make Predictions

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

### Step 5: Evaluate

```
from sklearn.metrics import mean_squared_error, r2_score
```

```
mse = mean_squared_error(y_test, y_pred)  
r2 = r2_score(y_test, y_pred)
```

```
print("MSE:", mse)  
print("R2 Score:", r2)
```

### Step 6: Visualize

```
import matplotlib.pyplot as plt
```

```
plt.scatter(X, y, color='blue', label='Actual Data')  
plt.scatter(X_test, y_pred, color='darkred', marker='x', label='Predicted (XGB)')
```

```
plt.xlabel('Year')  
plt.ylabel('Income')  
plt.title('XGBoost Regression - Per Capita Income')  
plt.legend()  
plt.grid(True)  
plt.tight_layout()  
plt.show()
```

## Evaluation Output

Mean Squared Error: 3,034,850.75

R<sup>2</sup> Score: 0.975