Multiple Linear Regression Assignment – 3 problems

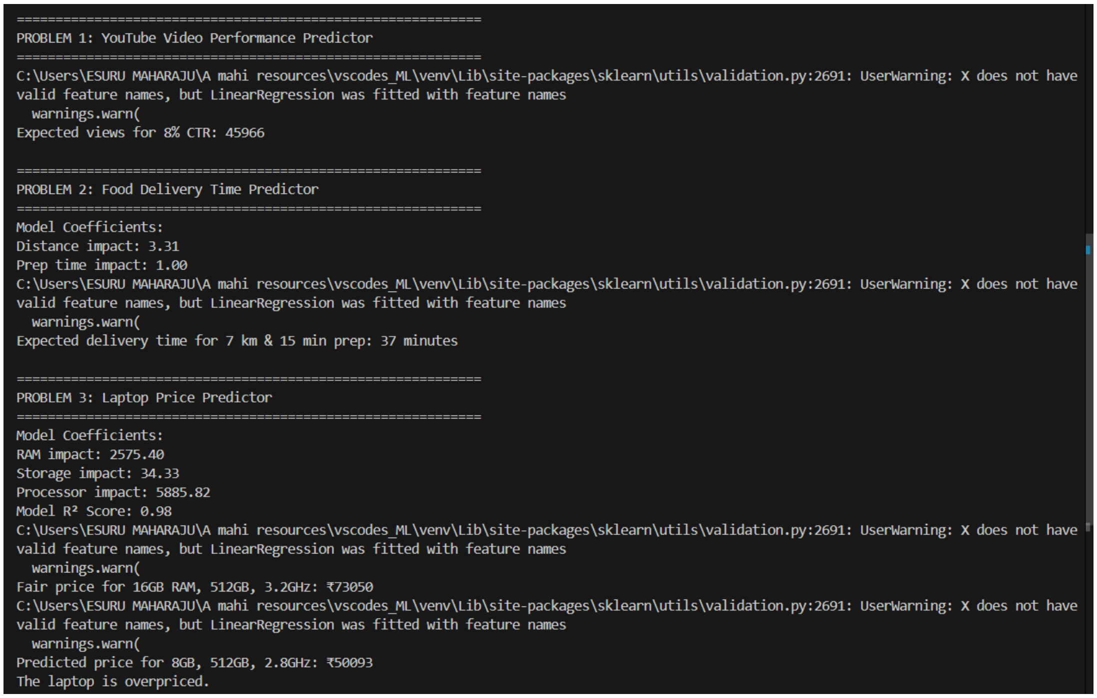
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

# ============================================================
# PROBLEM 1: YOUTUBE VIDEO PERFORMANCE PREDICTOR
# ============================================================

data1 = {
    'ctr': [3.2, 5.8, 2.1, 7.4, 4.5, 6.2, 1.8, 5.1, 3.9, 8.5,
            4.2, 2.8, 6.8, 3.5, 7.1, 2.4, 5.5, 4.8, 6.5, 3.1,
            7.8, 2.6, 5.3, 4.1, 6.9, 3.7, 8.1, 2.2, 5.9, 4.6],

    'total_views': [12000, 28000, 8500, 42000, 19000, 33000, 7000, 24000, 16000, 51000,
                    18000, 11000, 38000, 14500, 44000, 9500, 26000, 21000, 35000, 13000,
                    47000, 10000, 25000, 17500, 40000, 15000, 49000, 8000, 31000, 20000]
}

df1 = pd.DataFrame(data1)

plt.figure()
plt.scatter(df1['ctr'], df1['total_views'])
plt.xlabel("CTR (%)")
plt.ylabel("Total Views")
plt.title("CTR vs Total Views")
plt.show()

X1 = df1[['ctr']]
y1 = df1['total_views']

X1_train, X1_test, y1_train, y1_test = train_test_split(X1, y1, test_size=0.2, random_state=42)

model1 = LinearRegression()
model1.fit(X1_train, y1_train)
```

```
pred_views = model1.predict([[8]])[0]
print("Expected views for 8% CTR:", int(pred_views))


# ============================================================
# PROBLEM 2: FOOD DELIVERY TIME PREDICTOR
# ============================================================

data2 = {
    'distance_km': [2.5, 6.0, 1.2, 8.5, 3.8, 5.2, 1.8, 7.0, 4.5, 9.2,
                    2.0, 6.5, 3.2, 7.8, 4.0, 5.8, 1.5, 8.0, 3.5, 6.8,
                    2.2, 5.5, 4.2, 9.0, 2.8, 7.2, 3.0, 6.2, 4.8, 8.2],

    'prep_time_min': [10, 20, 8, 25, 12, 18, 7, 22, 15, 28,
                      9, 19, 11, 24, 14, 17, 6, 26, 13, 21,
                      10, 16, 14, 27, 11, 23, 12, 18, 15, 25],

    'delivery_time_min': [18, 38, 12, 52, 24, 34, 14, 45, 29, 58,
                          15, 40, 21, 50, 27, 35, 11, 54, 23, 43,
                          17, 32, 26, 56, 19, 47, 20, 37, 30, 53]
}

df2 = pd.DataFrame(data2)

plt.figure()
plt.scatter(df2['distance_km'], df2['delivery_time_min'])
plt.xlabel("Distance (km)")
plt.ylabel("Delivery Time (min)")
plt.title("Distance vs Delivery Time")
plt.show()

plt.figure()
plt.scatter(df2['prep_time_min'], df2['delivery_time_min'])
plt.xlabel("Prep Time (min)")
plt.ylabel("Delivery Time (min)")
plt.title("Prep Time vs Delivery Time")
plt.show()

X2 = df2[['distance_km', 'prep_time_min']]
y2 = df2['delivery_time_min']

X2_train, X2_test, y2_train, y2_test = train_test_split(X2, y2, test_size=0.2, random_state=42)
```

```
model2 = LinearRegression()
model2.fit(X2_train, y2_train)

print("Distance coefficient:", model2.coef_[0])
print("Prep time coefficient:", model2.coef_[1])

pred_time = model2.predict([[7, 15]])[0]
print("Expected delivery time:", int(pred_time), "minutes")


# ============================================================
# PROBLEM 3: LAPTOP PRICE PREDICTOR
# ============================================================

data3 = {
    'ram_gb': [4, 8, 4, 16, 8, 8, 4, 16, 8, 16,
               4, 8, 4, 16, 8, 8, 4, 16, 8, 16,
               4, 8, 4, 16, 8, 8, 4, 16, 8, 16],

    'storage_gb': [256, 512, 128, 512, 256, 512, 256, 1024, 256, 512,
                   128, 512, 256, 1024, 256, 512, 128, 512, 256, 1024,
                   256, 512, 128, 512, 256, 512, 256, 1024, 256, 512],

    'processor_ghz': [2.1, 2.8, 1.8, 3.2, 2.4, 3.0, 2.0, 3.5, 2.6, 3.0,
                      1.6, 2.8, 2.2, 3.4, 2.5, 2.9, 1.9, 3.1, 2.3, 3.6,
                      2.0, 2.7, 1.7, 3.3, 2.4, 3.0, 2.1, 3.5, 2.6, 3.2],

    'price_inr': [28000, 45000, 22000, 72000, 38000, 52000, 26000, 95000, 42000, 68000,
                  20000, 48000, 29000, 88000, 40000, 50000, 23000, 70000, 36000, 98000,
                  25000, 46000, 21000, 75000, 39000, 53000, 27000, 92000, 43000, 73000]
}

df3 = pd.DataFrame(data3)

for feature in ['ram_gb', 'storage_gb', 'processor_ghz']:
    plt.figure()
    plt.scatter(df3[feature], df3['price_inr'])
    plt.xlabel(feature)
    plt.ylabel("Price (INR)")
    plt.title(feature + " vs Price")
    plt.show()

X3 = df3[['ram_gb', 'storage_gb', 'processor_ghz']]
```

```
y3 = df3['price_inr']

X3_train, X3_test, y3_train, y3_test = train_test_split(X3, y3, test_size=0.2, random_state=42)

model3 = LinearRegression()
model3.fit(X3_train, y3_train)

print("Model coefficients:", model3.coef_)

y3_pred = model3.predict(X3_test)
print("R2 Score:", r2_score(y3_test, y3_pred))

pred_price = model3.predict([[16, 512, 3.2]])[0]
print("Fair price:", int(pred_price))

bonus_price = model3.predict([[8, 512, 2.8]])[0]
print("Predicted price for bonus laptop:", int(bonus_price))

if 55000 > bonus_price:
   print("Laptop is overpriced")
else:
   print("Laptop price is fair or underpriced")
```

Output:



```
============================================================
PROBLEM 1: YouTube Video Performance Predictor
============================================================
C:\Users\ESURU MAHARAJU\A mahi resources\vscodes_ML\venv\Lib\site-packages\sklearn\utils\validation.py:2691: UserWarning: X does not have
valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
Expected views for 8% CTR: 45966

============================================================
PROBLEM 2: Food Delivery Time Predictor
============================================================
Model Coefficients:
Distance impact: 3.31
Prep time impact: 1.00
C:\Users\ESURU MAHARAJU\A mahi resources\vscodes_ML\venv\Lib\site-packages\sklearn\utils\validation.py:2691: UserWarning: X does not have
valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
Expected delivery time for 7 km & 15 min prep: 37 minutes

============================================================
PROBLEM 3: Laptop Price Predictor
============================================================
Model Coefficients:
RAM impact: 2575.40
Storage impact: 34.33
Processor impact: 5885.82
Model R² Score: 0.98
C:\Users\ESURU MAHARAJU\A mahi resources\vscodes_ML\venv\Lib\site-packages\sklearn\utils\validation.py:2691: UserWarning: X does not have
valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
Fair price for 16GB RAM, 512GB, 3.2GHz: ₹73050
C:\Users\ESURU MAHARAJU\A mahi resources\vscodes_ML\venv\Lib\site-packages\sklearn\utils\validation.py:2691: UserWarning: X does not have
valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
Predicted price for 8GB, 512GB, 2.8GHz: ₹50093
The laptop is overpriced.
```