Lab Work 2: Linear Regression for Predicting Profit and Transactions

Objective:

To apply linear regression for predicting Profit and Transactions based on Sales data.

Steps:

1. Set Up Python Environment:

- Ensure that Python and necessary libraries like pandas, sklearn, and matplotlib are installed.
- Use Python's package manager (pip) to install required libraries if not already done:

```
PS C:\Aarchi_022bim003> pip install pandas scikit-learn m
Defaulting to user installation because normal site-packa
Collecting pandas
  Using cached pandas-2.3.0-cp313-cp313-win_amd64.whl.met
Collecting scikit-learn
  Using cached scikit_learn-1.7.0-cp313-cp313-win_amd64.w
Collecting matplotlib
  Using cached matplotlib-3.10.3-cp313-cp313-win_amd64.wh
Requirement already satisfied: numpy>=1.26.0 in c:\users\
ythonsoftwarefoundation.python.3.13_qbz5n2kfra8p0\localca
te-packages (from pandas) (2.3.1)
Collecting python-dateutil>=2.8.2 (from pandas)
```

2. Load and Prepare data:

- Load the provided CSV data into a DataFrame using pandas.
- Preprocess the data by ensuring all necessary columns (Sales, Profit, & Transactions) are correctly formatted.

```python
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

# Step 1: Load CSV file
df = pd.read_csv('Lab_2_Data.csv')

# Step 2: Preprocessing
# Drop rows with missing Profit and Transactions separately
profit_train = df.dropna(subset=['Profit'])
trans_train = df.dropna(subset=['Transactions'])
```

3. Implement Linear Regression Models:

- Use LinearRegression from sklearn to create two separate models:
    - One to predict Profit based on Sales.
    - Another to predict Transactions based on Sales.

```
# Step 3: Train Linear Regression Models
profit_model = LinearRegression()
trans_model = LinearRegression()
```

4. Make Predictions:

- Predict the missing Profit and Transactions for the given Sales data using the trained models.

```
PS H:\SXC BIM 6TH SEM\Business Information Systems\Aarchi_022bim003> python
_022bim003\Lab2.py"
Profit model R² score: 0.9948128102748437
Transactions model R² score: 0.992944108499159
```

5. Visualize the output.

- The diagram consists of two scatter plots, both sharing the same x-axis labeled "Sales" with values ranging from 10,500 to 26,221. The two plots are:

Output:

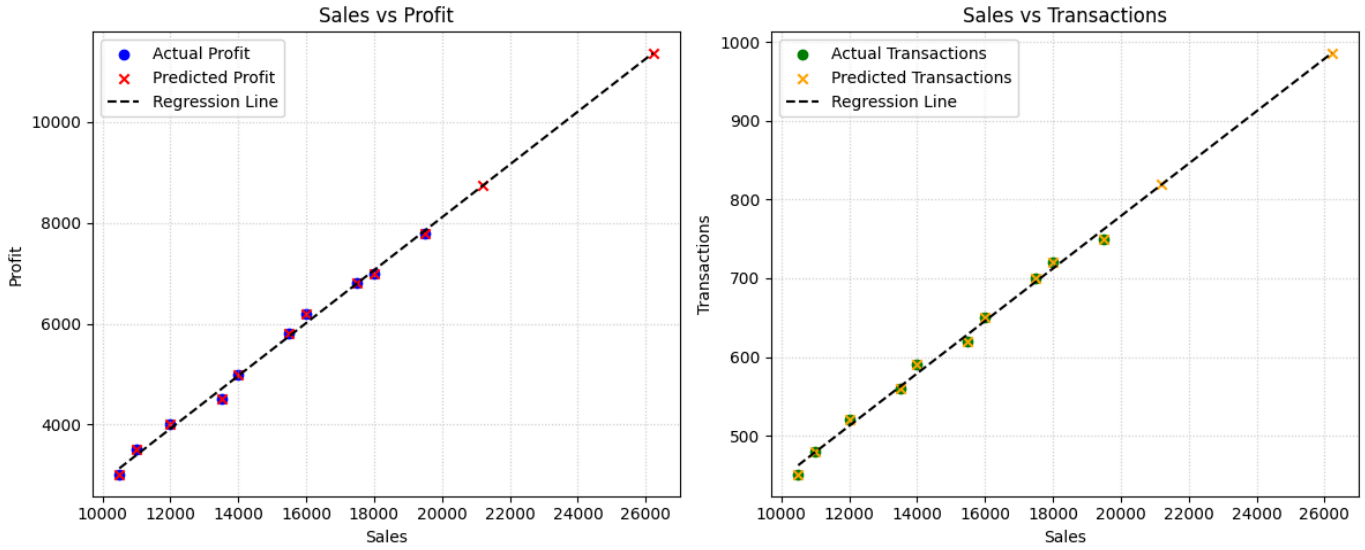


Fig: Output plots for Original Dataset

A. Sales vs Profit

- **Description:**

    a. Y-axis: Profit, ranging from 3,000 to 7,800.

    b. Data points:
        i. "Actual Profit" (scattered points).
        ii. "Predicted Profit" (likely overlapping or close to the actual points).

    c. Regression Line: A line representing the trend between Sales and Profit.

- **Interpretation:**

    a. The regression line suggests a positive linear relationship between Sales and Profit. As Sales increase, Profit tends to increase as well.

    b. The closeness of "Actual Profit" and "Predicted Profit" points to the regression line indicates that the model fits the data well, meaning Sales is a good predictor of Profit.

B. Sales vs Transactions:

- **Description:**

    a. Y-axis: Transactions, ranging from 450 to 750.

    b. Data points:
        i. "Actual Transactions" (scattered points).
        ii. "Predicted Transactions" (likely overlapping or close to the actual points).

    c. Regression Line: A line representing the trend between Sales and Transactions.

- **Interpretation:**

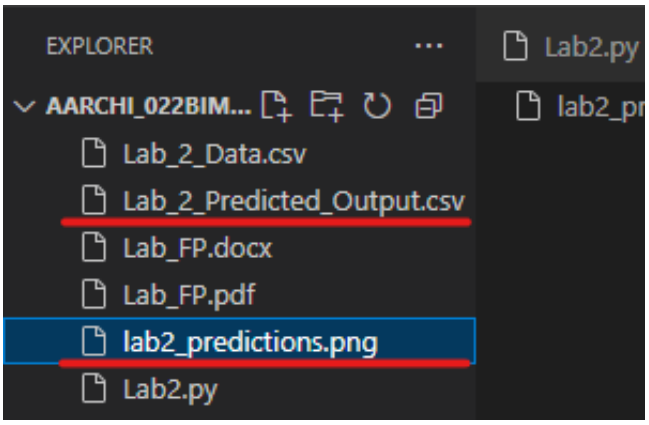
    a. The regression line here also shows a positive linear relationship, indicating that higher Sales are associated with a higher number of Transactions.

    b. Similar to the first plot, the alignment of "Actual Transactions" and "Predicted Transactions" with the regression line suggests a strong predictive relationship between Sales and Transactions.

6. Save Results:

- Save the updated data (including predictions) to a new CSV file.

- Save the generated plots as images for review.



7. Modification of original data

- The diagram consists of two scatter plots, both sharing the same x-axis ("Sales"), which ranges from 7,000 to 18,000. Each plot examines the relationship between Sales and a different business metric:

Output:

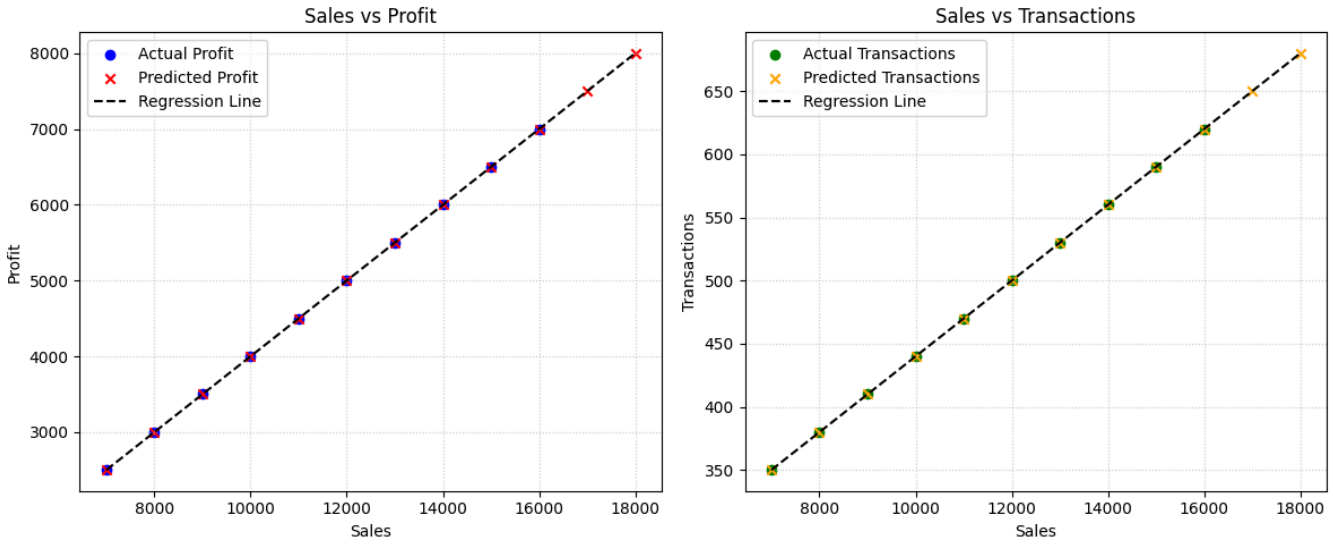


Fig: Output plots for Modified Dataset

A. Sales vs. Profit

- Description:

a. Y-axis: Profit, ranging from 2,500 to 7,000.

b. Data points:
   i. Actual Profit (observed values, scattered points).
   ii. Predicted Profit (model-estimated values, likely close to actual points).

c. Regression Line: A straight trend line showing the linear relationship between Sales and Profit.

- **Interpretation:**

a. Positive Correlation: The upward-sloping regression line indicates that higher sales lead to higher profits.

b. Model Fit: The close alignment between Actual Profit and Predicted Profit suggests the regression model is reliable for forecasting.

c. Profit Sensitivity:
   i. At Sales = 7,000, Profit = 2500.
   ii. At Sales = 18,000, Profit = 8,000.
   iii. Observation: Profit grows at a slightly diminishing rate (not perfectly proportional), possibly due to rising costs at higher sales volumes.

B. Sales vs. Transactions

- **Description:**

a. Y-axis: Transactions, ranging from 350 to 620.

b. Data points:
   i. Actual Transactions (observed values, scattered points).
   ii. Predicted Transactions (model-estimated values, likely close to actual points).

c. Regression Line: A straight trend line showing the linear relationship between Sales and Transactions.

- **Interpretation:**

a. Positive Correlation: The regression line confirms that higher sales are associated with more transactions.

b. Model Fit: The tight clustering of Actual vs. Predicted Transactions implies a strong predictive relationship.

c. Transaction Efficiency:
   i. At Sales = 7,000, Transactions = 350.
   ii. At Sales = 18,000, Transactions = 680.
   iii. Observation: The increase in transactions is less steep than sales growth.

Conclusion:

This lab demonstrated how linear regression can predict Profit and Transactions from Sales data, providing valuable insights for business decisions. Key takeaways:

1. Forecasting – Predict profit and transaction trends to set realistic targets.
2. Optimization – Identify inefficiencies (e.g., slowing profit growth at higher sales) to adjust pricing or costs.
3. Strategy – Focus on increasing average transaction value rather than just sales volume.

The model enables data-driven decisions, but should be refined with additional variables for greater accuracy. By integrating regression analysis into business planning, companies can enhance profitability and operational efficiency.

Hence this lab work confirms that linear regression is a powerful tool for transforming sales data into actionable business intelligence.