TAMIZHAN SKILLS

PROJECT:

Sales Forecasting With Linear Regression

PROBLEM STATEMENT:

Many businesses face challenges in predicting future sales. Without accurate forecasts, they struggle with inventory management, staffing, budgeting, and growth planning.

OBJECTIVE:

To *build a machine learning model using Linear Regression that can predict future sales* based on historical data. This model will help businesses make data-driven decisions.

REQUIREMENTS:

The dataset should be in **CSV format** and include the following columns:

- date – the date of the sale
- product – the item sold (can be filtered or encoded if multiple products)
- quantity – units sold
- revenue – money earned from the sales

STEP 1: Data Preprocessing

GOAL: Prepare clean and usable data for the model.

WHY: Machine learning models perform poorly with missing or noisy data.

TASK:

- Load the Data
- View first few rows
- Check for missing values
- Convert date column to datetime

STEP 2: Feature Engineering

GOAL: Create a time-series suitable format.

WHY: We want the model to learn trends based on time progression.

TASK:

- Aggregate the data

STEP 3: Linear Regression

GOAL: Train a simple model to learn the trend and forecast future sales.

WHY: Linear regression is a good baseline model. It's fast, interpretable, and works well with linear trends.

TASK:

- Import necessary libraries
- Train-test split
- Initialize and train model
- Predict

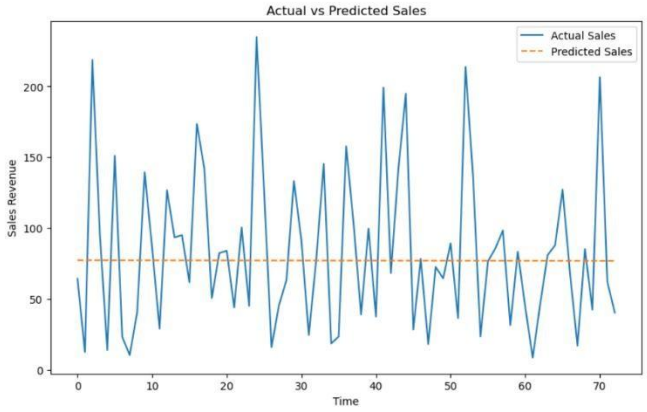
STEP 4: Evaluate Model Performance

GOAL: Check how well the model is performing.

STEP 5: Visualization

GOAL: Compare actual vs predicted sales visually.

WHY: A graph gives stakeholders a quick idea of accuracy and trend matching.



What This Graph Explains:

1. Comparison Over Time

- The **X-axis** (Time) represents the timeline — days or periods (e.g., last few months).
- The **Y-axis** shows the **sales revenue**.

You are plotting two lines:

- y_test = **Actual sales** from the real data
- y_pred = **Predicted sales** from your linear regression model

2. Model Accuracy

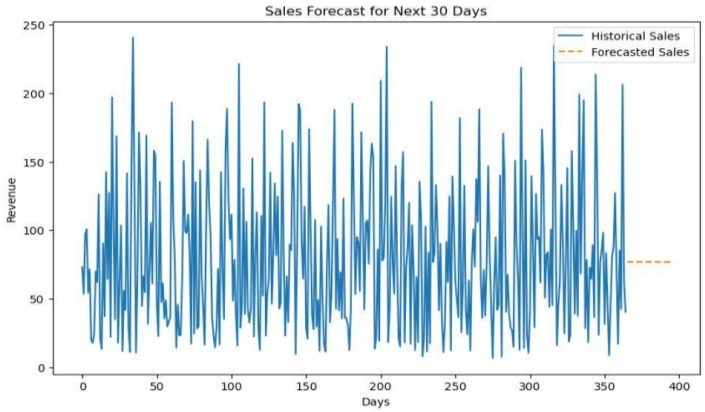
- If the **two lines are close together**, it means your model has predicted the sales very accurately.
- If the **predicted line (dashed)** deviates from the actual line, it indicates where your model made errors.

3. Trend Learning

- This graph visually shows whether your model has **learned the correct trend** — for example:
    - Is it capturing increases and decreases properly?
    - Is it able to follow seasonal patterns or sudden spikes?

STEP 6: Forecast Future Sales

GOAL: Use the model to predict future sales.



What It Explains:

1. **Trend Understanding**:
    - Helps you see whether your sales have been **increasing**, **decreasing**, or **stable** over time.
    - Linear Regression assumes a trend and extends it into the future.
2. **Model Forecast**:
    - Shows how the model **projects the next 30 days** of revenue. ○ If your business is growing steadily, the forecast line will slope upward.
    - If it's declining, the line will slope downward.
3. **Business Planning Tool**:
    - Useful for **inventory, staffing, and marketing** decisions. ○ Helps you answer: "How much are we likely to sell next month if the trend continues?"

EXPECTED OUTCOME:

- **A trained regression model** that can predict sales based on past trends.
- **A visual comparison** of actual vs predicted values.
- **Forecasts of future sales** in both graphical and tabular form.
- **Evaluation metrics** (like MSE, R^2 score) to judge accuracy.