Feature Engineering (Simple Explanation)

1. Data Preparation
Before creating features, we make sure the dataset is sorted by `store`, `item`, and `date` so that time-based features are computed correctly.

---

2. Lag Features
Lag features represent sales from previous days.

- **lag_1**: Yesterday's sales
- **lag_7**: Sales from the same day last week
- **lag_30**: Sales from about a month ago

■ These help the model remember recent patterns — like short-term memory.

---

3. Rolling Features
Rolling features describe average or variability in sales over recent days.

- **rolling_mean_7**: Average sales over the past 7 days
- **rolling_std_7**: Variation in sales during the last 7 days
- **rolling_mean_30**: Average sales during the last 30 days

■ The `.shift(1)` before `.rolling()` prevents *data leakage*, ensuring only past data is used.

---

4. Change-Based Features
These show how sales change over time.

- **diff_1**: The difference between today’s and yesterday’s sales
- **pct_change_7**: The percentage change compared to 7 days ago

■ They capture direction and strength of sales movement — increasing or decreasing.

---

5. Cyclical Features
Days of week (0–6) and months (1–12) are cyclical; after Sunday comes Monday, not “7.”

We use sine and cosine transforms to capture this cycle:
- **dow_sin**, **dow_cos**: cyclical representation of day_of_week
- **month_sin**, **month_cos**: cyclical representation of month

■ This allows the model to understand that day 0 and day 6 are close in time.

---

6. Handling Missing Values
Because of lag and rolling, the first few rows in each group are `NaN` (no previous data).
We drop or fill them to keep the dataset clean.

---

7. Output
We save the resulting dataset as `features_v1.csv`.
This version includes all the new time-based and cyclical features, ready for modeling.

Summary Table

Feature	Meaning	Purpose
lag_1	Yesterday’s sales	Short-term memory
lag_7	Sales last week	Weekly pattern
lag_30	Sales last month	Long-term trend
rolling_mean_7	7-day average	Short-term trend
rolling_std_7	7-day variability	Sales stability
rolling_mean_30	30-day average	Long-term trend
diff_1	Change from yesterday	Direction of change
pct_change_7	Percent change from last week	Strength of trend
dow_sin / cos	Cyclical day encoding	Weekly cycle
month_sin / cos	Cyclical month encoding	Seasonal cycle