# Features — End-User Reader Guide

The columns given in the dataset are masked according to certain hidden transformations.

This guide is for the participating IITs regarding how to work **with** masked features and interpreting codes **without knowing** proprietary formulas or common indicator names.

---

## 1) What you can (and can't) infer

**You can infer:** - The **family** of a feature (what broad kind of information it encodes): Price-Based, Volatility-Based, Volume-Based, Price+Volume, Band-Based. - The **ordinal index** within that family (e.g., the 7th PB feature). - The **time-scale rank(s)** via `T` indices, where **larger T means longer lookback**.

**You should not infer or request:** - The **formula**, **exact signal name**, or **implementation details** of any feature. - Reverse mappings from codes → proprietary names.

---

## 2) Anatomy of a code

`<Family><Index>[_B<Branch>][_T<t1>[_T<t2> ...]][_<Qualifier>]`

- **Family** ∈ { `PB`, `VB`, `V`, `PV`, `BB` }
- **Index**: integer (1-based) unique within a family (e.g., `PB7`).
- **`_B<Branch>`**: optional sub-variant marker (e.g., `_B3`).
- **`_T...`**: one or more **time-scale ranks**. **Order matters**: leftmost is the **shortest** among the listed T's. Eg: in _TX_TY X is the shorter lookback. Also a larger number after T corresponds to a larger lookback

**Examples (decode):**

- `BB10_T2` → Band-Based, 10th in its family, single horizon T2.

- `PB3_T7` → Price-Based, 3rd, horizon T7 (longer than T2, shorter than T10).

- `V8_T1_T2` → Volume-Based, 8th, cross-horizon (T1 vs T2; T1 is shorter).

- `PV3_B4_T10` → Price+Volume, 3rd, branch 4, long horizon T10.

- `BB21` → Band-Based, 21st, scale-free or implicit horizon.

---

## 3) Families (high-level meaning)

| Prefix | Family | High-level description |
|---|---|---|
| **PB** | Price-Based | Encodes behaviors of the price path and its transformations. |
| **VB** | Volatility-Based | Encodes dispersion/variability/range characteristics of price. |
| **V** | Volume-Based | Encodes participation/activity patterns from traded quantity alone. |
| **PV** | Price + Volume | Encodes joint behavior of movement and participation/flow. |
| **BB** | Band-Based | Encodes relationships to reference bands/envelopes (levels, widths, slopes, positions). |

The table intentionally keeps descriptions **generic** to avoid revealing proprietary constructions.

---

## 4) The T-ladder (time scales)

- `T1 < T2 < T3 < … < T12` is a **strictly increasing** set of lookbacks.
- **Important policy:** a given `Tn` represents the **same absolute lookback across *all* features and families**. (E.g., `T5` is the same horizon everywhere.)
- We have published only **relative ordering**. Treat larger T as **longer context**.

Features with multiple T's (e.g., `_T2_T7`) relate two horizons; the **leftmost** is the **shorter** one.

---

## 5) Branches

- **`_B#` (branch):** a structured sub-variant of the same idea. It does *not* imply a formula; it partitions consistent variants.

---

## 6) FAQ

**Q1. Does T5 mean the same lookback everywhere?**
**A. Yes.** `T5` is the **same absolute lookback** across all families and features.

**Q2. What if a code has no `_T…`?**
The feature is scale-free, session-scoped, or uses an implicit horizon.

**Q3. Can I compare `PB7_T2` with `PB7_T9`?**
Yes—this is a safe cross-horizon comparison **within the same semantic index**.

**Q4. Can I get the exact formula?**
No. You can use family, index, T-rank(s), and qualifiers to work effectively without internals.