The anomaly\_list in your script defines a list of anomalies to be introduced into the PMU data before processing it. Each entry in this list is a dictionary specifying the type, column, and range of the anomaly. Here's a breakdown of its purpose and use:

**Purpose of anomaly\_list**

This list is used in the introduce\_anomalies() function to simulate anomalies in the dataset by modifying specific data points. The introduced anomalies serve as test cases to evaluate the performance of the anomaly detection algorithms.

**Structure of anomaly\_list**

Each dictionary in the list contains:

* **type**: Specifies the type of anomaly ('drop' for missing data or 'spike' for sudden magnitude increases).
* **column**: The name of the column where the anomaly should be introduced.
* **start and end**: The row index range within which the anomaly is introduced.
* **magnitude** (only for spikes): The amount by which the selected data points will be increased.

**Usage in introduce\_anomalies()**

The function introduce\_anomalies(data, anomalies) modifies the dataset based on the provided anomaly\_list.

**How it works:**

1. It creates a copy of the input data (anomalous\_data).
2. It initializes a new column, True\_Label, where:
   * 0 indicates normal data.
   * 1 marks anomalous data.
3. It iterates over anomalies and applies:
   * **Drop Anomalies**: Replaces values in the specified range with NaN (simulating missing data).
   * **Spike Anomalies**: Increases values in the specified range by a given magnitude.
4. It counts and returns the number of anomalies introduced.

**Example Execution**

If the dataset contains a Freq column and a VA\_MAG column, after applying introduce\_anomalies(), the data will look like:

| **Time** | **Freq** | **VA\_MAG** | **True\_Label** |
| --- | --- | --- | --- |
| ... | ... | ... | ... |
| 100 | NaN | ... | 1 (Dropped) |
| 101 | NaN | ... | 1 (Dropped) |
| ... | ... | ... | ... |
| 300 | ... | +0.5 | 1 (Spiked) |
| 301 | ... | +0.5 | 1 (Spiked) |
| ... | ... | ... | ... |

**Why is anomaly\_list Useful?**

* **Testing Robustness**: Simulates real-world PMU anomalies for testing detection methods.
* **Ground Truth Creation**: The True\_Label column provides a reference for evaluating model performance.
* **Comparative Analysis**: Helps compare different anomaly detection techniques (e.g., DMD vs ML-based).