Below are 15 potential buckets for classifying employees, along with the features to use and example thresholds. (Note: Adjust thresholds based on your actual data distribution.)

Bucket 1: High vs. Low Communication

- **Features:** `Teams_Messages_Sent_sum`, `Emails_Sent_sum`
- **Thresholds:**
  - **High:** Values above the 75th percentile.
  - **Low:** Values below the 25th percentile.

Bucket 2: Consistent vs. Inconsistent Communication

- **Features:** `Teams_Messages_Sent_std`, `Emails_Sent_std`
- **Thresholds:**
  - **Consistent:** Standard deviation below the median (or a z-score < -0.5).
  - **Inconsistent:** Standard deviation above the median (or a z-score $> +0.5$).

Bucket 3: High vs. Low Meeting Attendance

- **Features:** `Meetings_Attended_sum`
- **Thresholds:**
  - **High:** Above the 75th percentile.
  - **Low:** Below the 25th percentile.

Bucket 4: Overworked vs. Underworked

- **Features:** `Work_Hours_sum`, `Work_Hours_mean`
- **Thresholds:**
  - **Overworked:** Sum above the 90th percentile or mean $>$ 50 hours/week.
  - **Underworked:** Sum below the 10th percentile or mean $<$ 30 hours/week.

Bucket 5: New Hires vs. Experienced

- **Features:** `Days_Since_Joining` (or derived tenure)
- **Thresholds:**
  - **New Hires:** Lower 25th percentile (e.g., less than 1 year).
  - **Experienced:** Upper 75th percentile (e.g., more than 5 years) with a mid-tenure group in between.

Bucket 6: High vs. Low Performance

- **Features:** `Performance_Rating`, `Manager_Feedback`, `Total_Decayed_Reward_Points` (inverse relation)
- **Thresholds:**
  - **High Performers:** Top 25% in rating and below 25th percentile in decayed reward points.
  - **Low Performers:** Bottom 25% in rating and above 75th percentile in decayed reward points.

Bucket 7: Highly Recognized vs. Under-Recognized

- **Features:** `Best Team Player_Count`, `Innovation Award_Count`, `Leadership Excellence_Count`, `Star Performer_Count`
- **Thresholds:**
  - **Highly Recognized:** Composite score above the 75th percentile.
  - **Under-Recognized:** Composite score below the 25th percentile.

Bucket 8: Balanced vs. Imbalanced Leave Usage

- **Features:** `Annual Leave_Factor`, `Casual Leave_Factor`, `Sick Leave_Factor`, `Unpaid Leave_Factor`
- **Thresholds:**
  - **Balanced:** Each factor within the interquartile range.
  - **Imbalanced:** One or more factors outside (below 25th or above 75th percentile).

Bucket 9: High Engagement in Activities

- **Features:** `Total_activity_entry`, `Last_activity_entry`
- **Thresholds:**
  - **Highly Engaged:** Total above the 75th percentile and recent activity (e.g., within the last week).
  - **Low Engagement:** Below the 25th percentile and older activity timestamps.

Bucket 10: Successful Onboarding

- **Features:** `Onboarding_Feedback`, `Initial_Training_Completed` (Boolean), `Onboarding_Factor`
- **Thresholds:**

  - **Successful:** Positive feedback (e.g., "Excellent" or score above a set threshold), training completed, and onboarding factor above the median.
  - **Struggling:** Otherwise.

Bucket 11: Mentor Potential

- **Features:** `Mentor_Assigned` (Boolean), `Performance_Rating`, `Days_Since_Joining`
- **Thresholds:**
  - **High Potential:** `Mentor_Assigned` is True or top 25% in performance, with tenure in the upper 50%.
  - **Not Mentor Candidates:** The opposite.

Bucket 12: Promotion Candidates

- **Features:** `Promotion_Consideration` (flag), `Performance_Rating`, `Manager_Feedback`
- **Thresholds:**
  - **Promotion-Ready:** Flagged for promotion and with ratings above the 75th percentile.
  - **Not Promotion-Ready:** Otherwise.

Bucket 13: Emotional/Morale State

- **Features:** `Decayed_Emotion_Zone`, `Decayed_Vibe`
- **Thresholds:**
  - **Positive Morale:** Both scores above a set threshold (e.g., above the median).
  - **Negative Morale:** One or both scores in the bottom 25%.

Bucket 14: High Variability in Work Hours

- **Features:** `Work_Hours_std`
- **Thresholds:**
  - **High Variability:** Above the 75th percentile.
  - **Consistent:** Below the 25th percentile.

Bucket 15: Reward Efficiency

- **Features:** Ratio of `Total_Decayed_Reward_Points` to `Total_activity_entry` or `Work_Hours_sum`

- **Thresholds:**
  - **High Efficiency:** Ratio below the 25th percentile.
  - **Low Efficiency:** Ratio above the 75th percentile.

Implementation Notes:

- Normalize features with different scales before computing percentiles or z-scores.
- Adjust thresholds (e.g., 25th/75th percentiles, fixed cutoffs) based on the data distribution.
- Consider creating composite scores (weighted sums) for buckets that involve multiple features.
- Revisit and recalibrate thresholds periodically as the data evolves.