Score Aggregation Methodology

Configuration Parameters

Parameter	Description
WINDOW_DAYS	Number of most recent daily partitions to scan (default: 7)
MIN_SCORE_GAIN	Minimum percent_gain threshold for rising terms (default: 0)
DMA_COUNT	Total number of DMAs considered (constant: 210)

Rising Terms (trend_rising_.csv)

1. Window Selection: Select rows from top_rising_terms whose refresh_date is within the latest WINDOW_DAYS partitions.

2. Raw Data Extraction: Retrieve term, dma_id, percent_gain as metric. Filter by percent_gain ≥ MIN_SCORE_GAIN.

3. Statistics Aggregation (stats CTE): • dma_hits (count distinct dma_id per term) • coverage_ratio (dma_hits / DMA_COUNT) • median_gain (via APPROX_QUANTILES(metric, 2)[OFFSET(1)]) • spread_intensity_score (coverage_ratio × median_gain).

4. Final Selection: Output top 200 terms ordered by spread_intensity_score descending.

Top Terms (trend_top_.csv)

1. Window Selection: Select rows from top_terms whose refresh_date is within the latest WINDOW_DAYS partitions.

2. Raw Data Extraction: Retrieve term, dma_id, score as metric, rank.

3. Best DMA per Term (best_dmas CTE): Keep only the row with the highest metric per term/DMA across the window.

4. Statistics Aggregation (stats CTE): • dma_hits (count distinct dma_id) • avg_rank (average rank across DMAs) • total_score (maximum metric across DMAs) • coverage_ratio (dma_hits / DMA_COUNT).

5. Final Selection: Output top 200 terms ordered by total_score descending.

Output Files

• trend_rising_.csv: term, dma_hits, coverage_ratio, median_gain, spread_intensity_score

• trend_top_.csv: term, dma_hits, avg_rank, total_score, coverage_ratio