

What one release contains.

Every food post on social, video included, decomposed into facts an agent can cite. This is the shape of the bundle we deliver: the tables, one post taken apart, and two queries that run against it.

THE TABLES

posts 5,776 rows	Every retrieved post, in the cohort or not: caption, transcript, a timestamped account of what happens in the video, relevance, and why it was excluded.
post_tags 176,825	Every finding: the node, the quoted span, and the passage it came from. Direct and rolled, so you can work at any level of the tree.
nodes 3,846	The taxonomy itself, induced from the corpus: labels, paths, tiers, and the editorial flag.
sentiment · sentiment_events 5,550 · 5,712	Stance per post with strength; every opinion with who said it, what kind of statement it was, and the quote.
topic_metrics 181	Share of conversation early vs late, with confidence intervals and a false-discovery correction.
post_metrics_history 20,643	Hourly engagement over a post's first days. How fast it took off, not how it did over the quarter.
creators 2,894	Perceived age band and presentation, each with its evidence sentence and the posts it was made from.
manifest.json SHA-256	Every input, hashed. Rebuild the export and get the same answer.

METHOD, IN BRIEF

Source	Instagram, collected through official platform APIs.
What we analyze	The caption, the transcript, the on-screen text, and a timestamped account of what happens in the video. Every finding names which one it came from.
Evidence	Every finding carries the post permalink, the quoted passage, and its passage id.
Provenance	Frozen inputs, each hashed in manifest.json, pinned to a model version and taxonomy build. Rebuild the export and get the same answer.
Personal data	Captions, transcripts, handles and permalinks ship. Perceived age band and presentation ship with their evidence. Base region, perceived race or ethnicity, accent labels and embeddings are not exported, by decision.
Freshness	Median 28 minutes from post to analyzed record in the last release. Six in ten posts within the hour; the rest were picked up by later retrieval runs.

ONE POST, TAKEN APART

[instagram.com/reel/DbLpibiRKt6/](https://www.instagram.com/reel/DbLpibiRKt6/) · @chloecleroux · 72.5K likes · Jul 24, 2026

"Just a simple summer table setting to gather your friends and family this season 🍅"

video 0:01 **preparations** · layering · "stacking large red tomatoes into a pyramid shape"

video 0:06 **preparations** · threading · "threading cherry tomatoes onto a string using a needle"

video 0:12 **paired_ingredients** · basil · in_dish · "sliced tomatoes and fresh basil leaves"

caption **occasions** · summer · "summer table setting"

video 0:16 **sentiment** · ▲ positive · tomato salad · creator · "takes a bite of the tomato salad, and smiles at the camera"

relevance central · in cohort · 11 direct findings exported, five shown. A caption-only tool sees "tomato", "summer", and three brand handles.

TWO QUERIES

```
-- Distinct posts per family. Families overlap; never sum them.
SELECT n.family_id, COUNT(DISTINCT t.post_id) AS posts
FROM read_parquet('post_tags.parquet') t
JOIN read_parquet('nodes.parquet') n USING (node_id)
WHERE t.kind = 'direct'
GROUP BY 1 ORDER BY 2 DESC;
-- paired_ingredients 4905
-- occasions          3028
-- forms              2856

-- What tomato is eaten with, and how.
SELECT n.label, t.relation, COUNT(DISTINCT t.post_id) AS posts
FROM read_parquet('post_tags.parquet') t
JOIN read_parquet('nodes.parquet') n USING (node_id)
WHERE t.kind = 'direct' AND n.family_id = 'paired_ingredients'
GROUP BY 1, 2 ORDER BY 3 DESC LIMIT 3;
-- onion  in_dish 1081
-- garlic in_dish  948
-- salt   in_dish  936
```

Count distinct post ids, never rows: a post can carry several findings in one family.