CAO Analysis Report

This report consolidates date coverage, renewal dynamics, salary information, and benefits presence across CAOs. It explains how each metric is constructed and how to read the visualizations.

Total CAOs analyzed: 99

Average number of files per CAO: 15.96

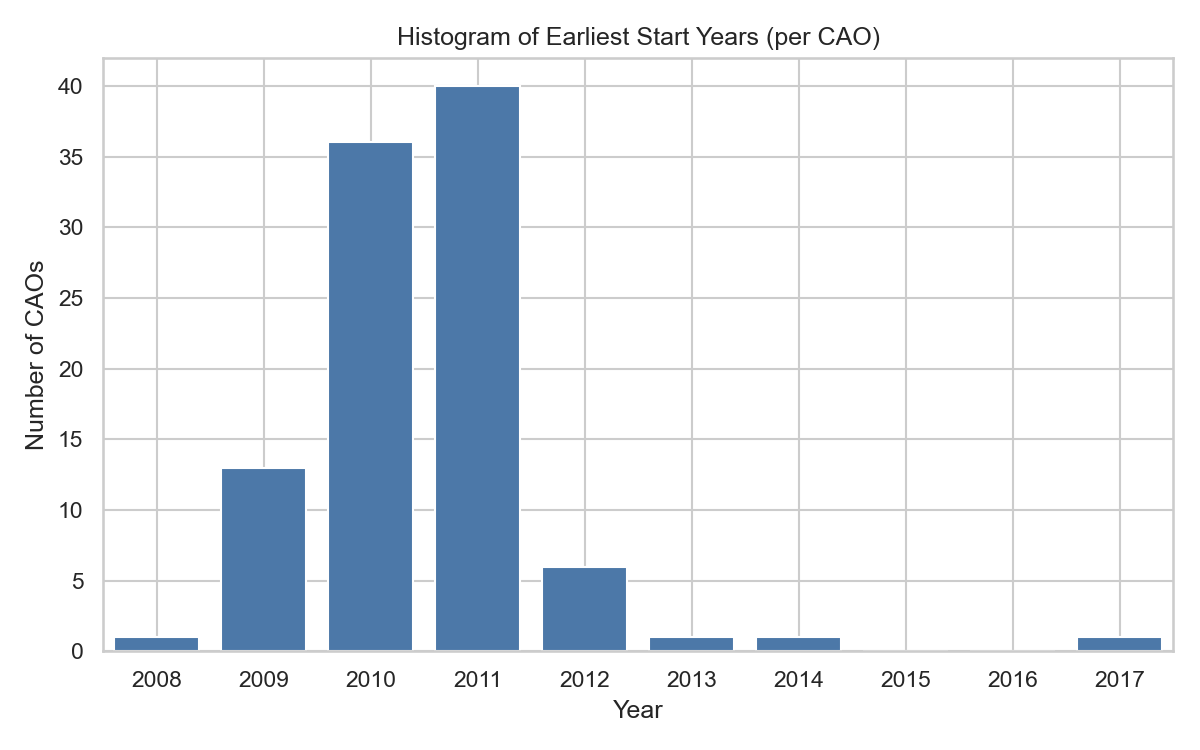
# Coverage and Periods

Coverage is defined per CAO as the union of all validity periods we found in the source files. The earliest start date and latest expiry date establish the outer window. Inside that window there may be internal gaps (days where no period was in force) or overlaps (two or more periods simultaneously covering the same time). This section summarizes those high‑level time bounds and how many files contribute to each CAO.

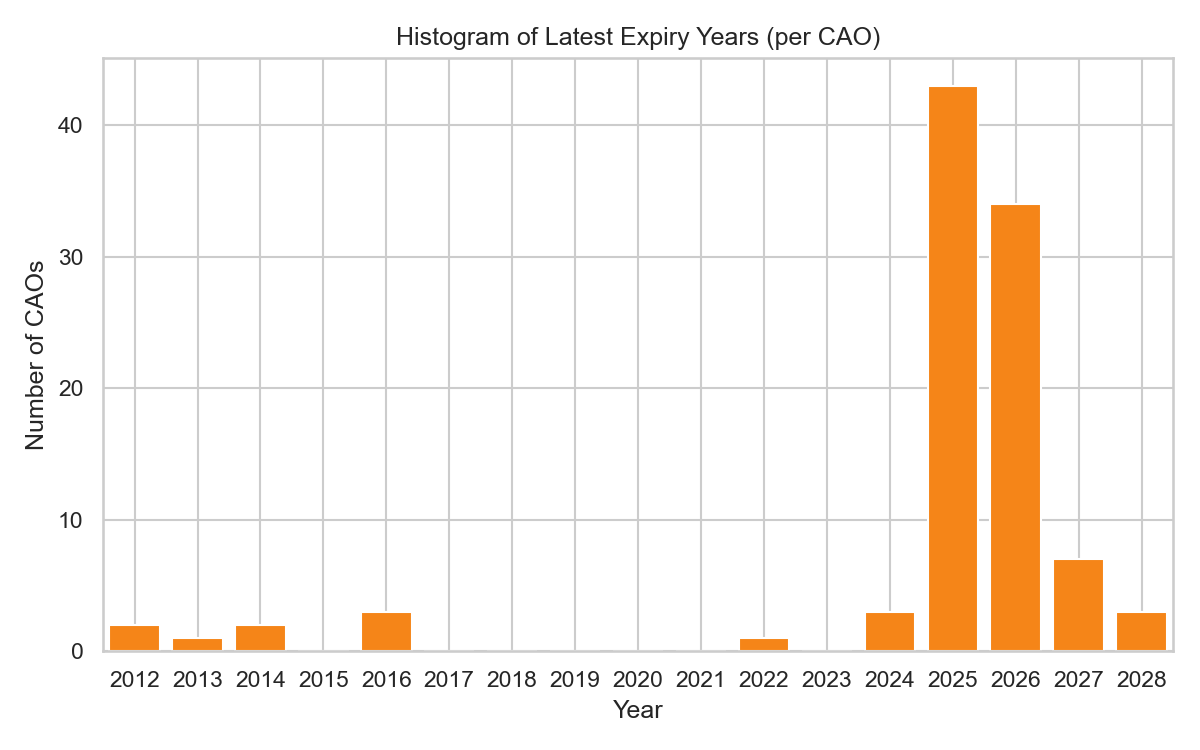
CAOs fully covered without internal gaps: 3

Average coverage window (earliest start to latest end): 174.2 months

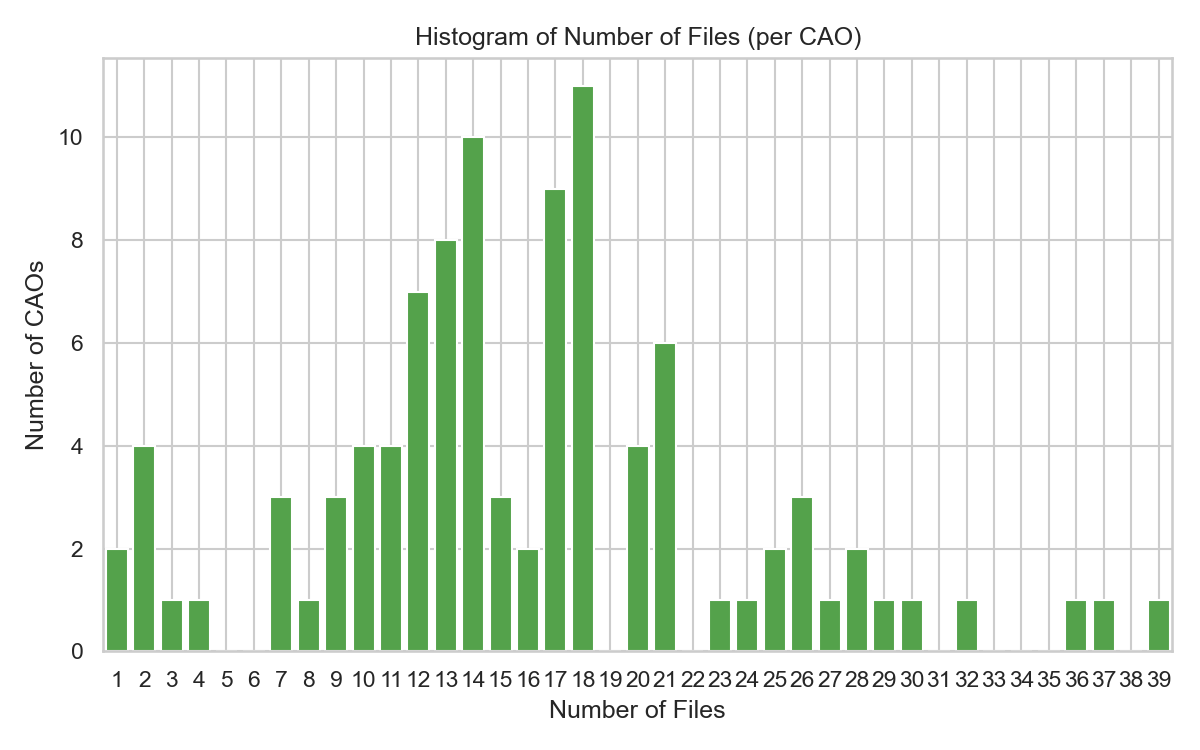
Interpretation guidance: a long coverage window with few files suggests sparse renewals or long multi‑year agreements; a short window with many files suggests frequent renewals, addenda, or multiple document variants.



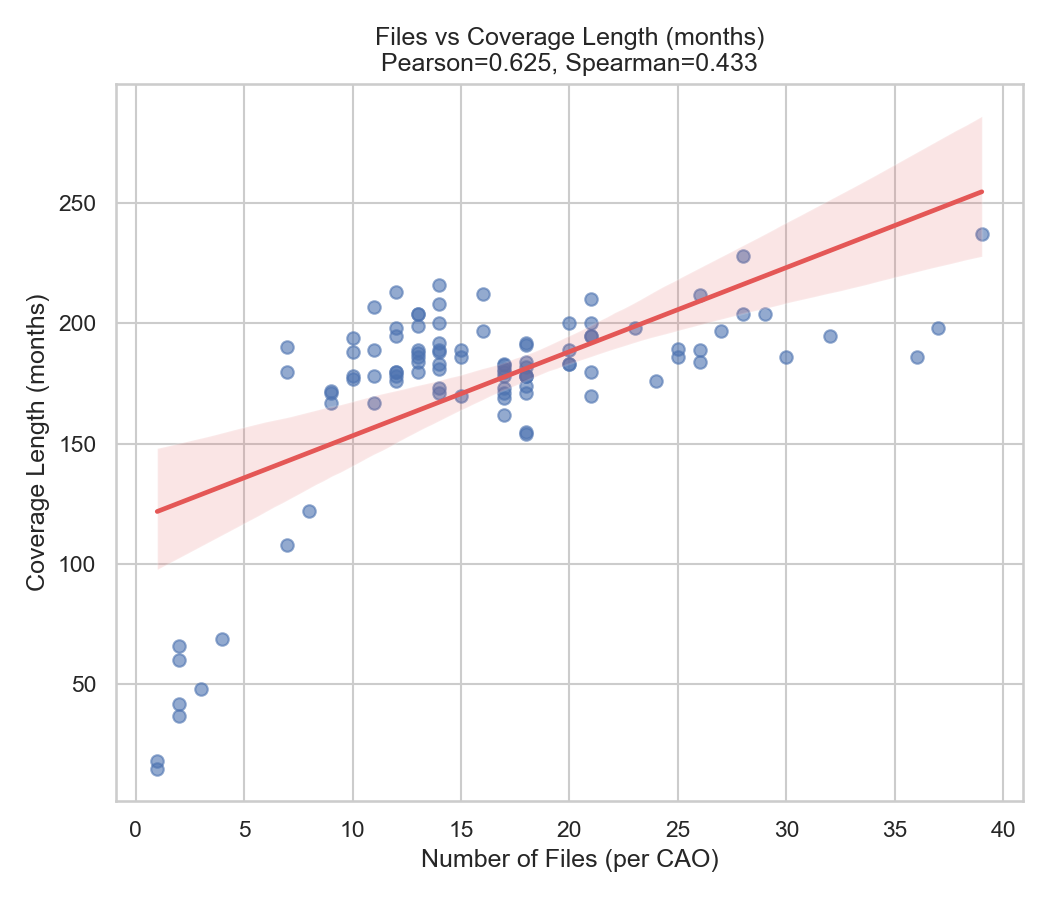
*Earliest start years distribution*



*Latest expiry years distribution*



*Files per CAO distribution*



*Files vs coverage length (months)*

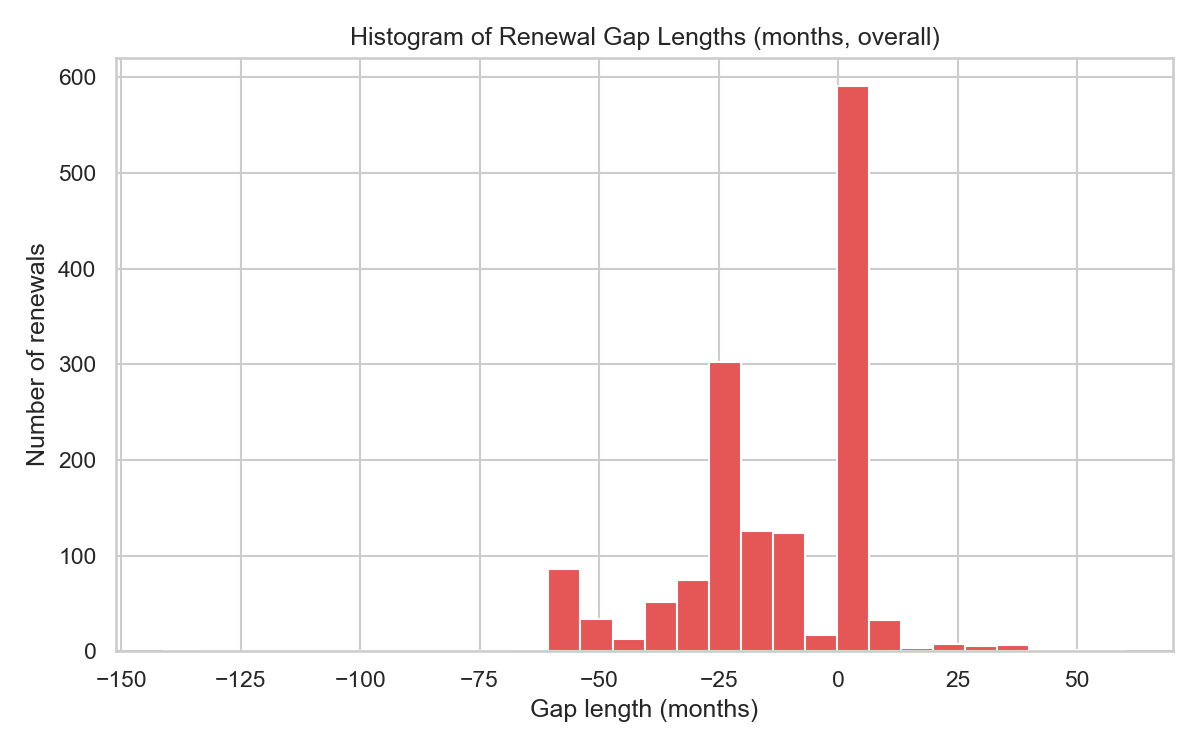
# Gaps and Renewals

Renewal gap = next period's start date minus previous period's end date after sorting by start date within a CAO. We distinguish: (a) negative gaps = overlaps (the next starts before the previous ends), (b) zero gaps = contiguous renewals (back‑to‑back), and (c) positive gaps = uncovered time. Only positive gaps represent true holes in coverage.

Large negative gaps typically arise when published periods use placeholder end dates (e.g., 2028‑12‑31) that were superseded early. Use the positive‑only statistics to understand actual breaks; use the 'all transitions' statistics to understand administrative cadence including overlaps.

Renewal gaps (all transitions): mean -14.38 months, median -11.96 months.

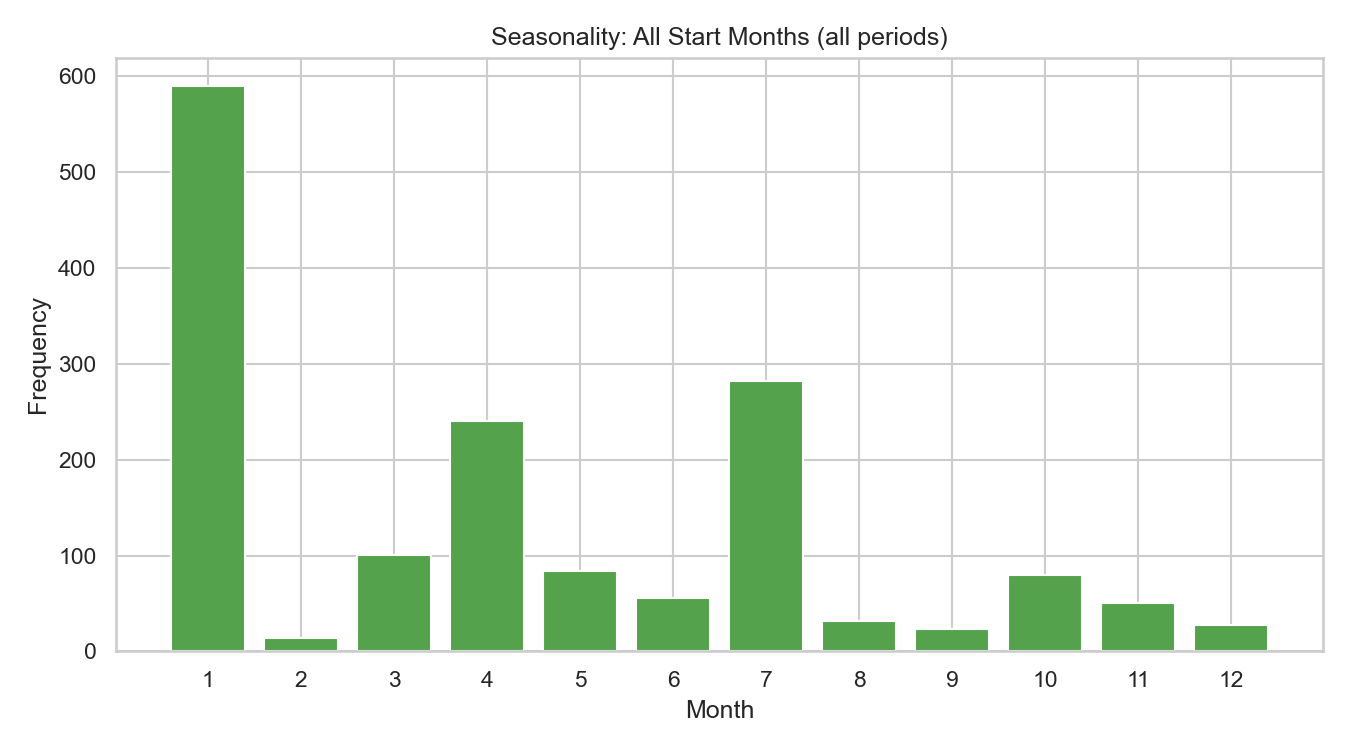
Renewal gaps (positive-only): mean 1.96 months, median 0.03 months.



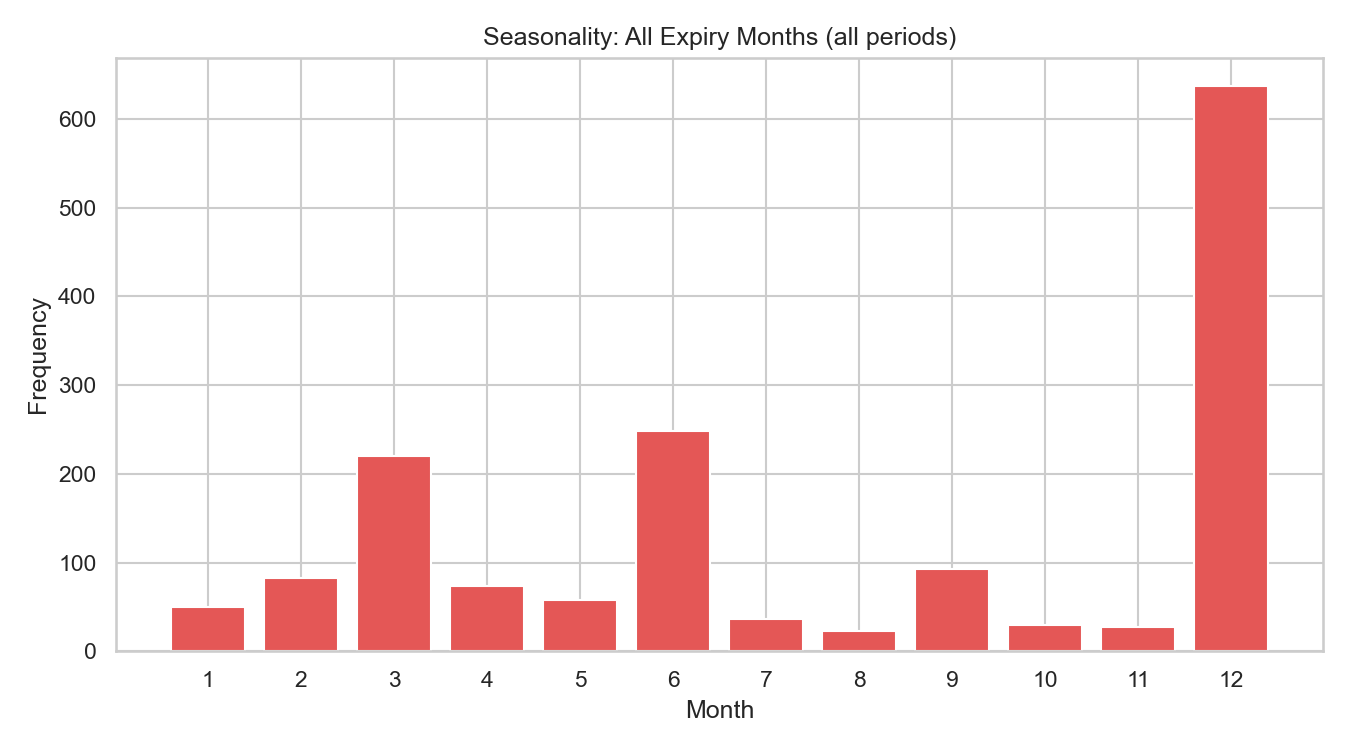
*Renewal gap lengths histogram (months)*

# Seasonality of Dates

We show the month of all period starts and expiries across the corpus. This can reveal preferred renewal months (e.g., January or July) or administrative clustering (e.g., expiries at year‑end). These charts are descriptive; they do not account for period length or document weight.



*All start months*

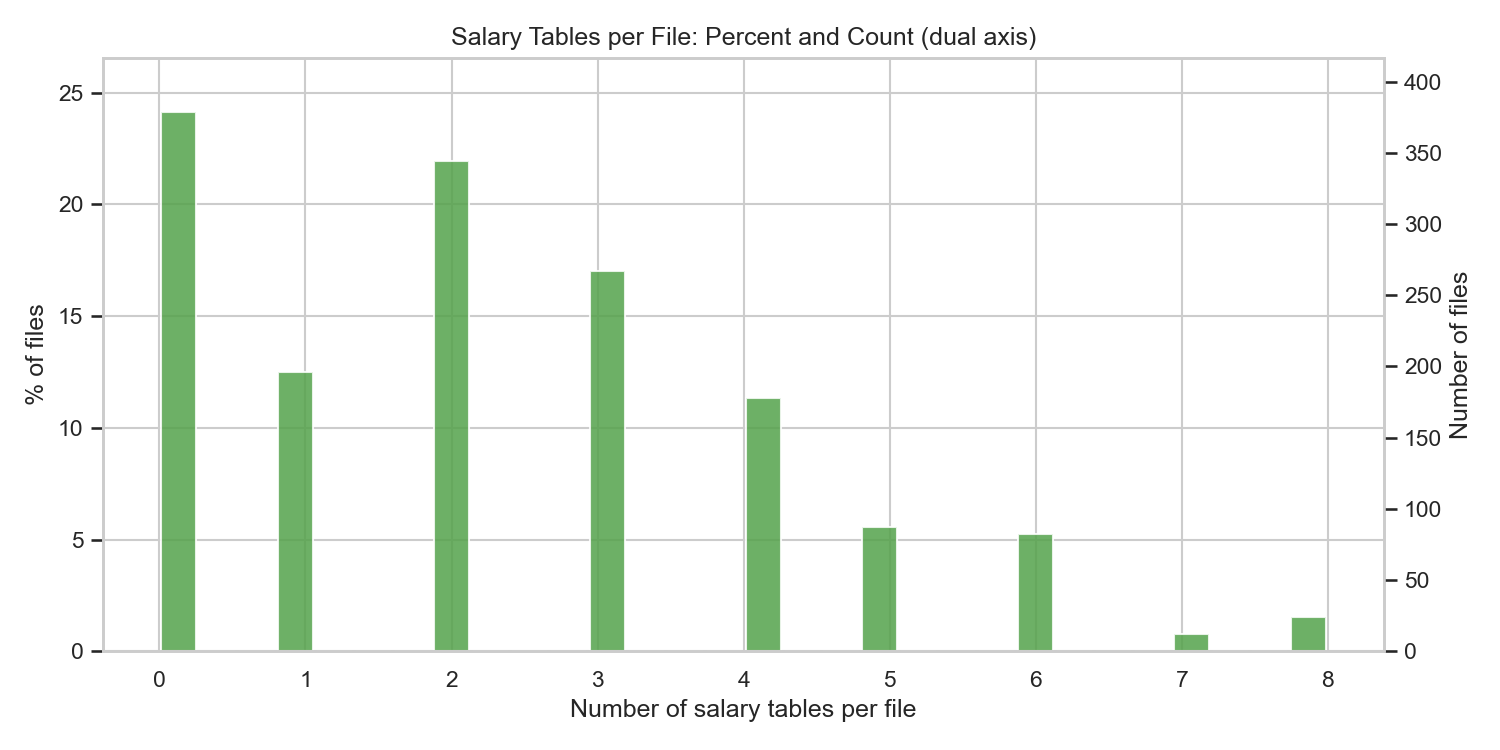


*All expiry months*

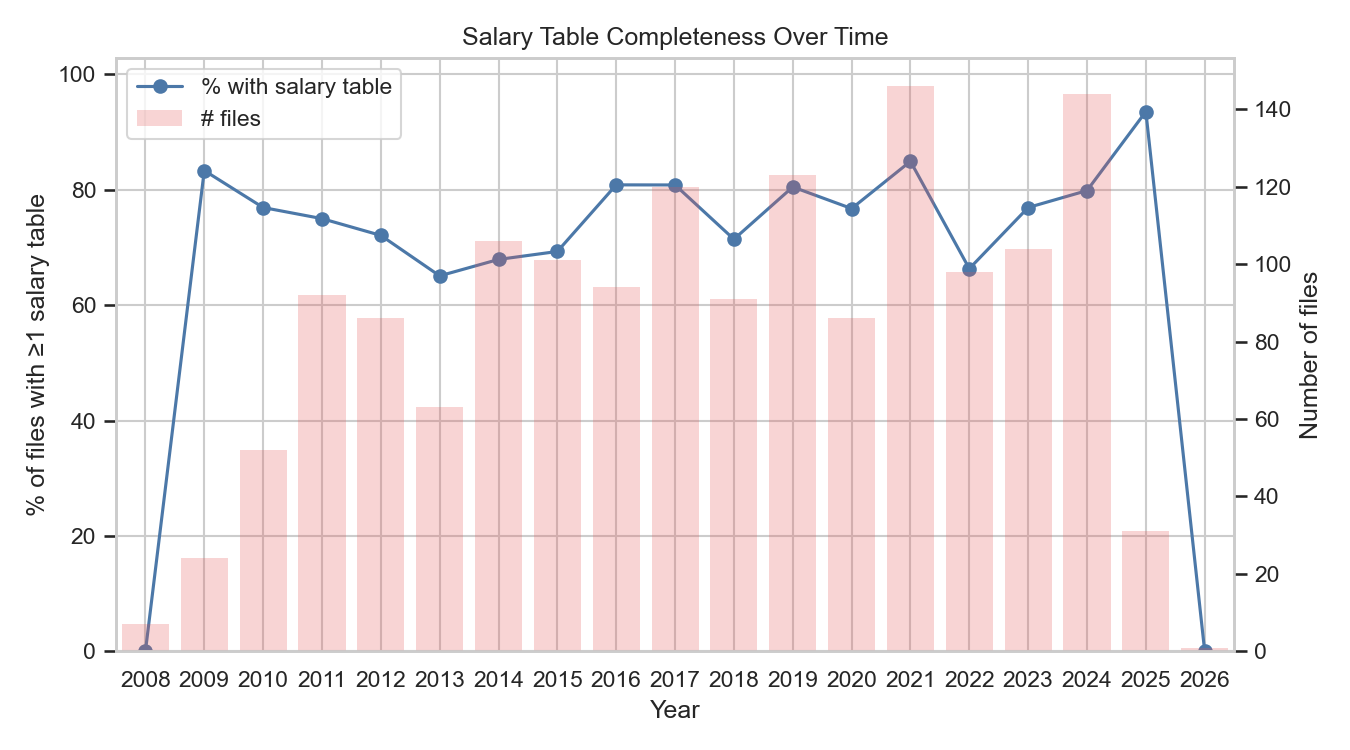
# Salary Tables

We count salary tables per file by checking filled columns salary\_1…salary\_7. If all seven are filled and more\_salaries is 'yes', we mark '8+' (internally 8). Completeness over time is the share of files per start year that contain ≥1 salary table. This reflects the presence of structured pay information, not its depth or correctness.

Reading tips: spikes in the distribution at 1–2 tables often reflect single‑table salary structures or one update table alongside an existing scale. '8+' indicates rich multi‑table structures. Over‑time completeness can shift as extraction quality or document formats evolve.



*Salary tables per file: % and count (aligned axes)*



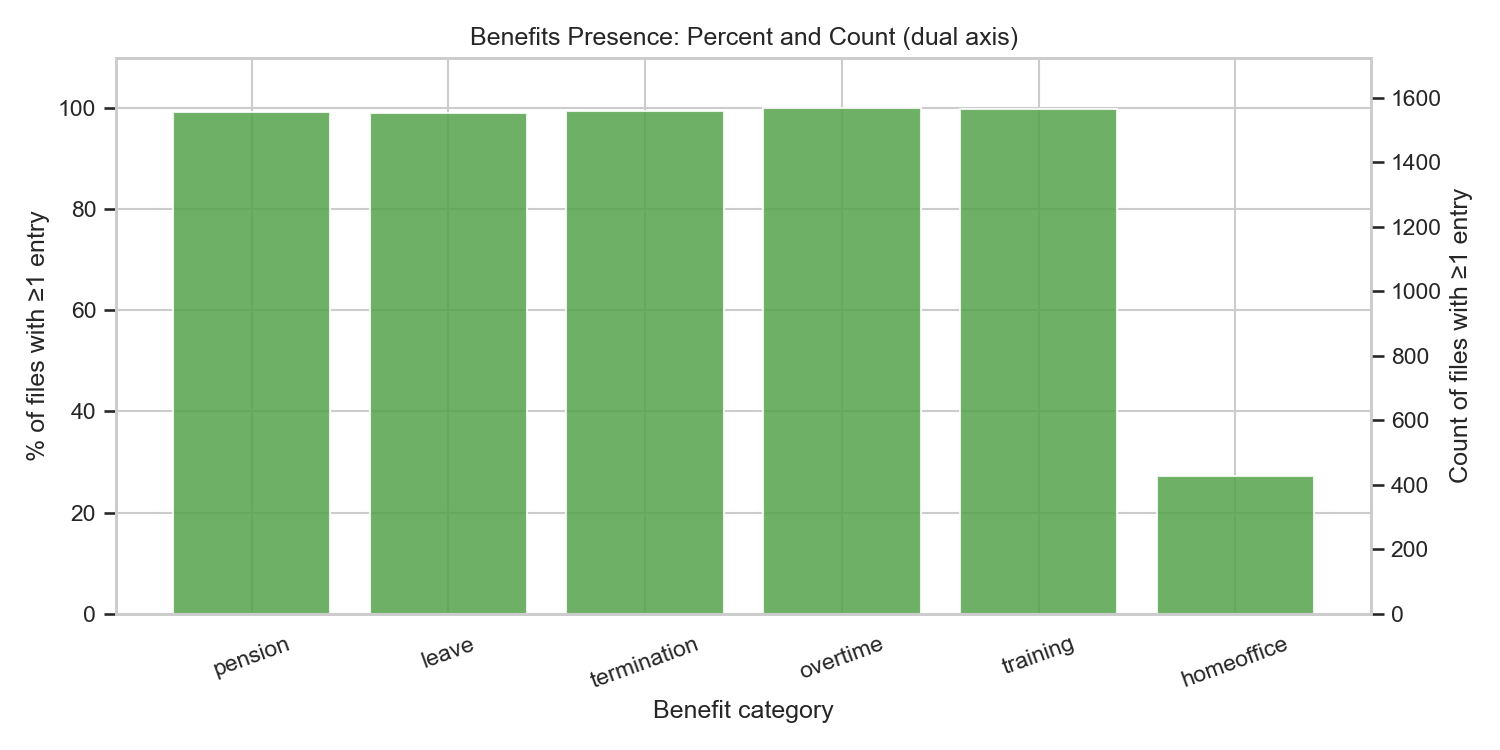
*% of files with ≥1 salary table by start year, with file counts*

Latest year completeness: 2026.0: 0.0% of files with ≥1 salary table (n=1).

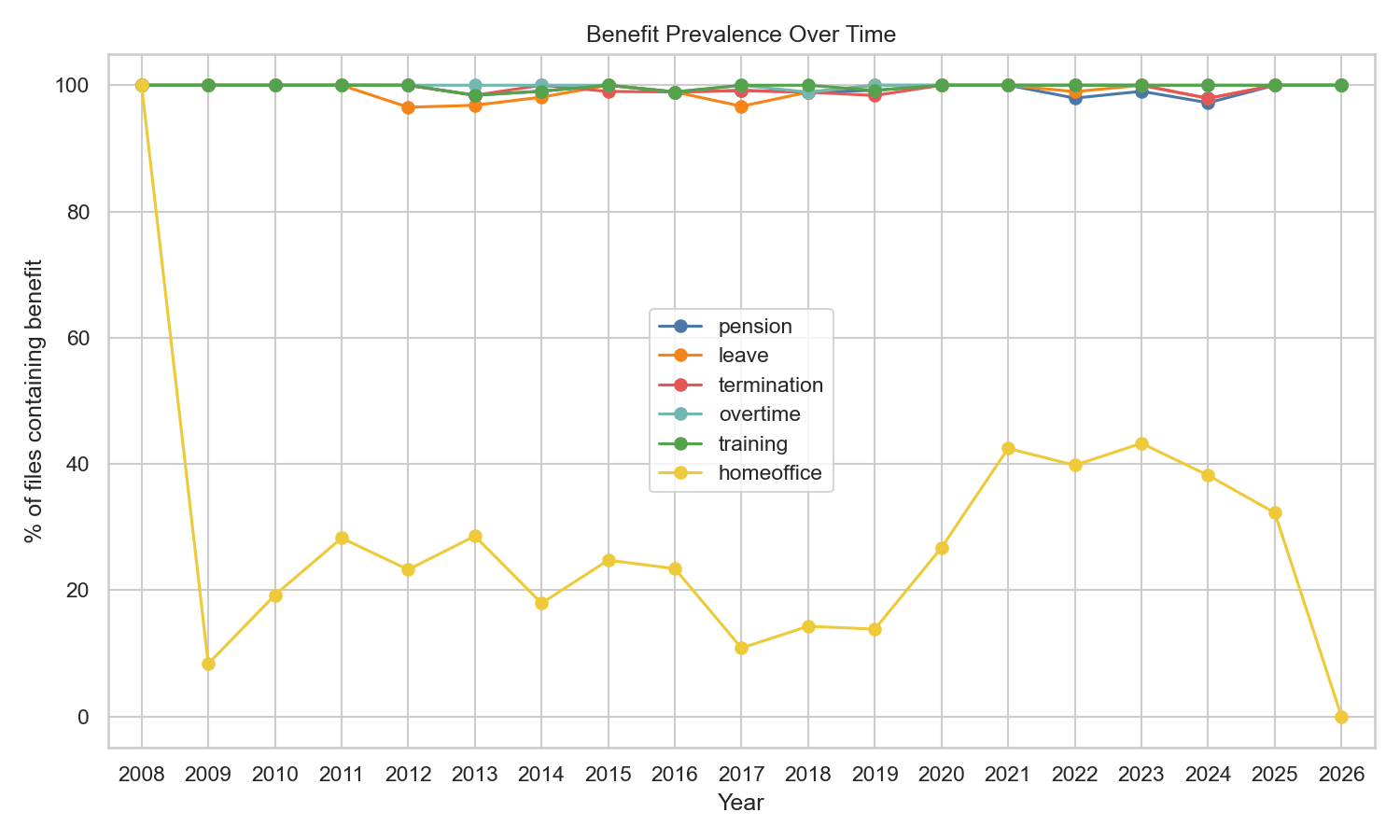
# Benefits

Benefit presence is detected when any mapped column for a topic is non‑empty in a file. Mapping includes: pension (pension, retire), leave (vacation, maternity, vakantie, verlof), termination (term\_\*, ontslag, beëindiging, probation), overtime (overtime, shift compensation, max/min hours), training, and homeoffice. Presence indicates the topic is mentioned with data, not necessarily that it is exhaustive or standardized.

Use the prevalence‑over‑time chart to see adoption trends (e.g., rise of homeoffice). Presence can be conservative for sparse entries and liberal for verbose narrative; interpret comparatively across topics and years rather than as absolute compliance.



*Benefits presence: % (left) and counts (right scale)*

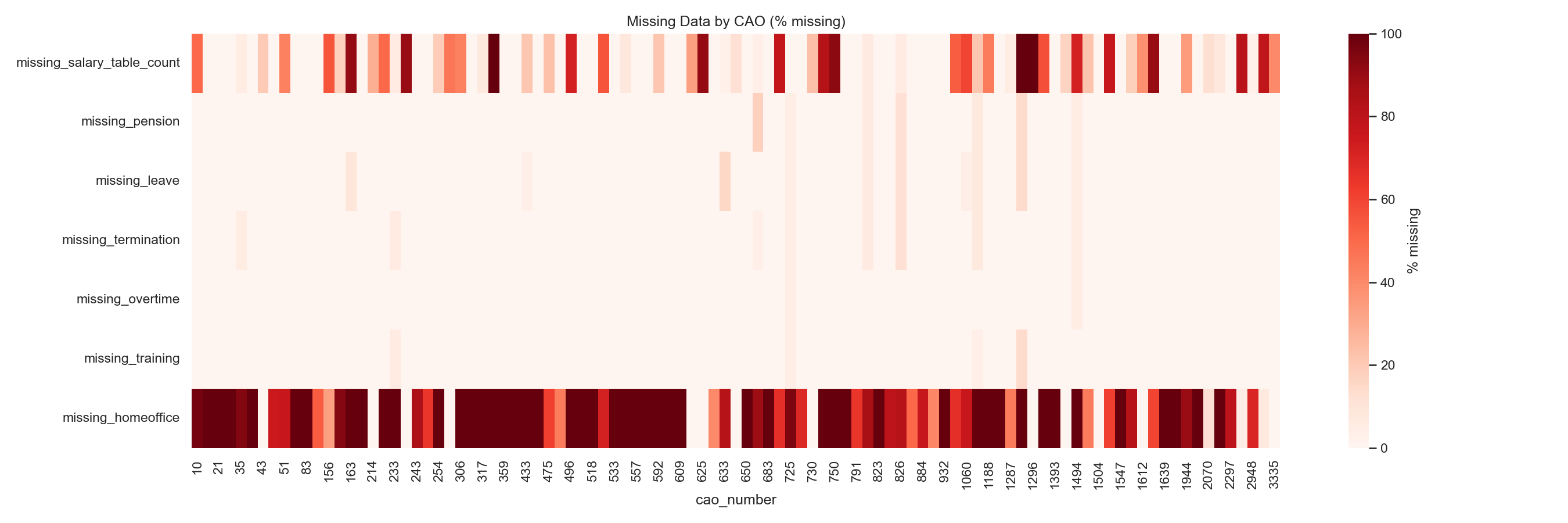


*Benefit prevalence over time (% of files)*

Top benefits by presence: overtime: 99.9%, training: 99.7%, termination: 99.4%.

# Missingness by CAO

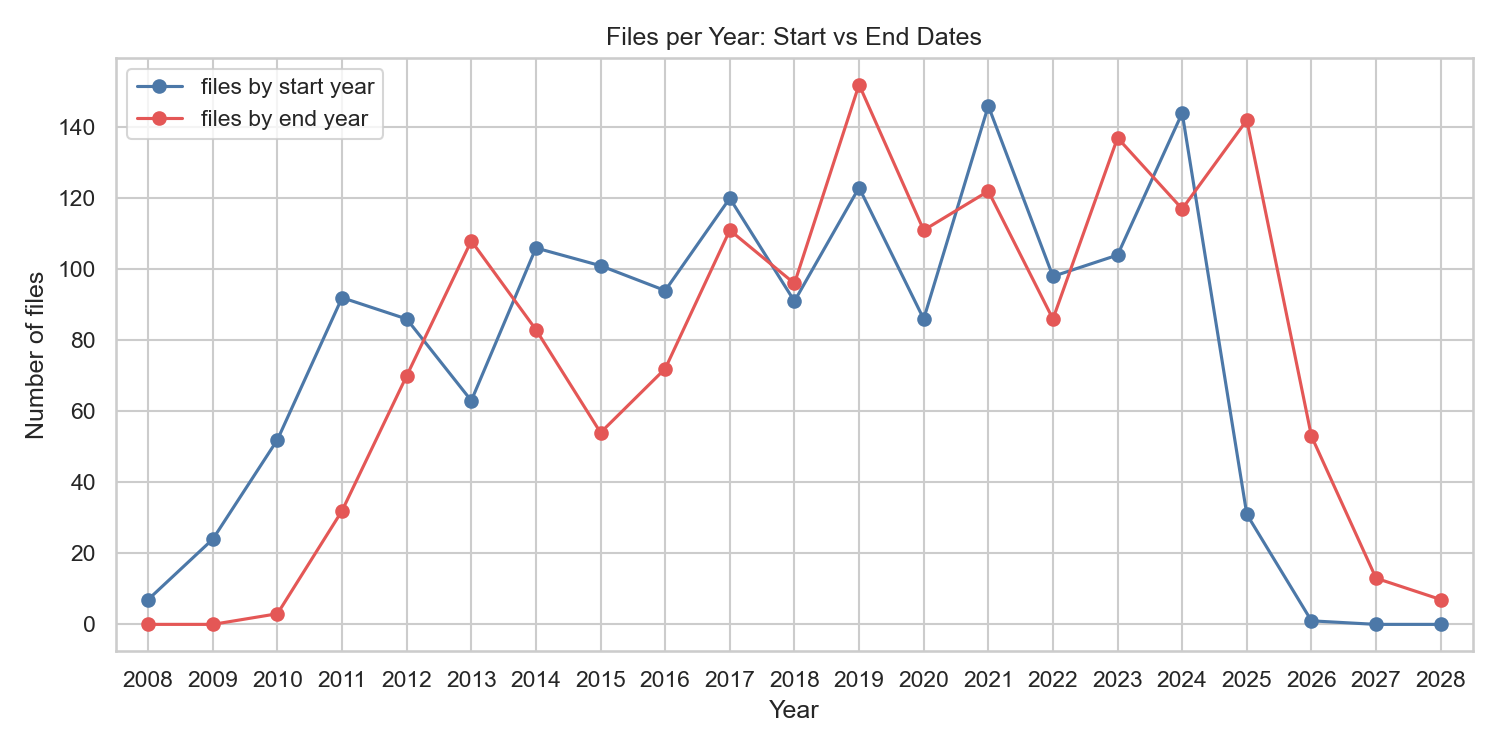
Missingness is computed at file level per CAO: salary is missing when the file has zero salary tables; a benefit is missing when none of its mapped columns are filled. We then average within each CAO. High missingness can reflect either true absence in the documents or extraction gaps; use alongside presence and completeness metrics.



*Missingness heatmap (% missing by CAO)*

# Files Over Time

We compare counts of files by earliest start year and by end year. Divergence between the lines can indicate long periods (many end in later years) or back‑dated agreements (starts cluster earlier). This provides context for interpreting completeness/prevalence trends.



*Files per year: start vs end*