

How *The Local* did this story

All inspection-related data used in this story came from [Ontario's public long-term care database](#), which publishes [general information](#) and inspection reports for every licensed long-term care facility in the province.

The database contains tens of thousands of [inspection reports](#)—forms filled out by provincial inspectors documenting violations and punitive enforcement actions tied to how individual homes are run. Individual reports typically include the inspection date and type (proactive, or related to a complaint or critical incident report), the inspectors involved, and any findings of non-compliance related to resident care, staffing, infection control, medication management, and more. It's a window into the day-to-day issues inside Ontario's long-term care homes.

Using Python and the web-scraping library [Beautiful Soup](#), *The Local* built a custom [web scraper](#) to collect all of the inspection reports from the more than [600 long-term care home pages](#) listed on the province's public reporting website.

In total, excluding duplicate French-language versions of reports, the database we've created contains more than 17,500 inspection reports dating from Jan. 1, 2018—the year Doug Ford's Progressive Conservatives took office in Ontario—through to December 31, 2025.

Collecting the reports was the relatively easy part. Making sense of them was something else entirely.

Over the past eight years, Ontario's inspection forms have changed repeatedly, particularly following the passage of the 2021 [Fixing Long-Term Care Act](#). And while most reports in the database were clean, searchable, digitally-generated PDFs, others were little more than scanned images. In rare cases, inspectors appeared to have filled forms out by hand before scanning and uploading them, creating major inconsistencies in formatting and readability from one report to the next.

To deal with those inconsistencies, *The Local* built a multi-stage parsing system to classify each report by form type and extract structured data from the hundreds of thousands of pages collected. Whenever possible, the system pulled text directly from PDFs. When reports existed only as scanned images or handwriting, it switched to [optical character recognition](#) to read the text.

From there, each report was scanned for recurring phrases using [regular expression \(REGEX\) pattern matching](#). Terms like "Inspection Summary," "Written Notification," and "Compliance Order" helped us extract the individual non-compliances cited in each inspection report,

alongside accompanying details like the type of enforcement action issued, and the text of the violation.

No AI tools were used to extract or analyze information from any of the documents we scraped from the ministry's website; only custom Python scripts alongside manual reporting and verification were employed.

The final product was a large structured database where each row represented a single cited violation tied to a specific inspection report and long-term care home. That line-by-line dataset was then aggregated and analyzed to identify broader trends, homes with recurring issues, the ministry's enforcement patterns, and other noteworthy factors across Ontario's long-term care inspection system.