MTL Archives

Making Montréal’s photographic archive easier to explore, and studying what makes people come back.

Product design & engineering 2025–present

Explore the archive ↗ Code ↗ Research notes ↗

Wiel Zouantcha · Read and interact with this case study at https://zouantcha.com/case-studies/mtl-archives

Following an interest in the city

I don't remember exactly what led me to Montréal’s open data portal. I was interested in the city and followed that curiosity. I started browsing the datasets without a particular project in mind.

Among them, I came across the open photographic archive and aerial photothèque. There were photographs of streets, aerial views and records of the city at different points in its history. That was the collection I wanted to spend more time with. As I explored it, I started wondering how someone could find a photograph without already knowing its title or catalogue reference.

I started building MTL Archives in October 2025 as a way to explore that collection. It became a search engine, then a daily location game and a print-order flow. Along the way, it became a research project about the data itself: what a model notices in an old photograph, what makes a search useful, and whether attention on social media leads people back to the archive.

My work spans the ingestion scripts, model experiments, website, visual identity and daily editorial pipeline. The public product now presents 13,000+ records. The larger research datasets include earlier versions and duplicates, so I keep their counts separate.

Aerial photograph | Survey document | Index card

An aerial photograph without the municipal document border. The January analysis placed many of these images together. Open archive record ↗

The first problem was the description

An early audit found that about 97% of the working records had synthetic descriptions. That needs a distinction: my cleaning pipeline had filled missing text with a title, date and archive reference. Those fallback sentences made the rows look complete without adding much meaning. They were not rich descriptions supplied by the city.

That matters for semantic search, which looks for meaning rather than an exact word match. If thousands of records say little more than “photograph, reference, date,” a better search model still has very little to work with. I needed to improve the evidence behind each result before improving its presentation.

Building the collection in layers

The extract, transform and load (ETL) pipeline downloads the city catalogues in batches, normalizes their fields, links records to image files and produces a manifest: a structured inventory of the collection. The source archive reference, or cote, stays attached to the record. Missing files, duplicate records and uncertain locations are data problems to track, not details to hide in the interface.

Some useful words are printed inside the image. I used Tesseract, an optical character recognition (OCR) engine, with French and English support to extract them. A vision-language model, which can interpret an image and produce text, supplies a separate description. Original metadata, extracted text and generated descriptions remain distinct.

From an archive export to a public product

1. Collect | 2. Reconcile | 3. Enrich | 4. Index | 5. Serve

Start with the city’s records.

Download the photographic archive and aerial survey catalogues in batches. Keep source identifiers and file references so each image can be traced back to its record.

> Source records → downloaded files

The preparation jobs run separately from the public website. A visitor’s search does not recaption the collection.

The preparation jobs use Python and TypeScript. Cloudflare R2 stores images; D1 stores records; Vectorize stores the numerical representations used for similarity search. The public Worker answers requests, while the Next.js website presents the results.

The ingestion code streams records in batches and writes checkpoints so an interruption does not require starting over. Failures have their own logs. The serving database is also distinct from the research corpus: a June audit recorded 13,499 deduplicated production records and 14,822 development records. A larger file or vector count does not mean the website has that many distinct photographs.

Captioning was a job to measure

The first recorded large captioning run used LLaVA, the Large Language and Vision Assistant 1.5 7B on a rented graphics processing unit (GPU). In December, the run report recorded 12,304 captioned images, 23 errors, 10.86 hours and $14 in compute. It was a partial run, with roughly 2,500 images still remaining.

By May, I had moved to structured outputs: descriptions, visual categories and fields that later tools could use. The full-run report contains 14,822 output rows, including 14,706 captions, 79 captions that failed the required structure and 116 image or model errors. A valid structure tells me the program can read an answer; it does not prove the description is historically correct.

That run also had to recover from interrupted work. The first 12,100 rows were reconstructed from saved chunks, then the remaining 2,722 were resumed. The report estimates 24.47 GPU hours, but the recovered portion does not have the same complete timing record as the tail. I keep it as an estimate rather than presenting it as a measured invoice.

What the image model was noticing

For visual search I used CLIP, Contrastive Language–Image Pre-training, which represents images and text as numerical vectors. Similar vectors can help match a phrase to a picture. I also built an explorer to inspect the collection rather than judge the system only through a few search queries.

In January, the projection of 14,715 image embeddings showed a striking split: many plain aerial photographs grouped separately from survey documents with municipal headers and borders. Index cards formed their own tight group. The formatting was part of the signal, even when the underlying subject was similar.

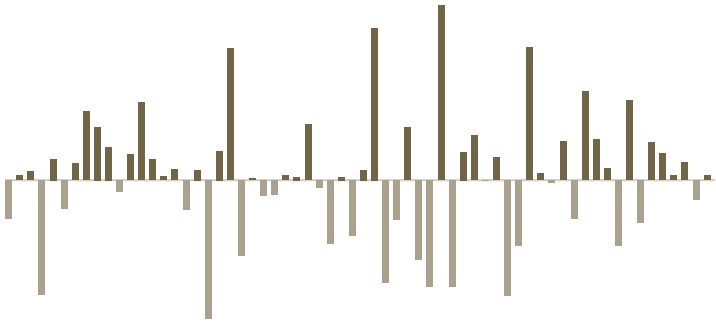
The projection used UMAP, Uniform Manifold Approximation and Projection, to turn high-dimensional vectors into a view I could inspect. It suggested what to investigate; distances on that map were not proof of semantic similarity or a controlled explanation of the model. I wrote about the observation in CLIP Sees Bureaucracy.

One image, two different jobs

Aerial photograph | Survey document | Index card

1. CLIP vector | 2. UMAP position

Original archive image



Dimensions 1–64 of 512 · signed values

CLIP turns the image into 512 numbers. These values describe learned visual features. Search compares vectors for similarity; it does not treat any single bar as a named concept like “street” or “building”. The chart shows the first 64 actual values, using a fixed scale across images.

Actual saved outputs from the research explorer. Changing the image reveals its recorded vector or position; no model runs in your browser. UMAP summarizes a collection of vectors, rather than transforming one photograph in isolation.

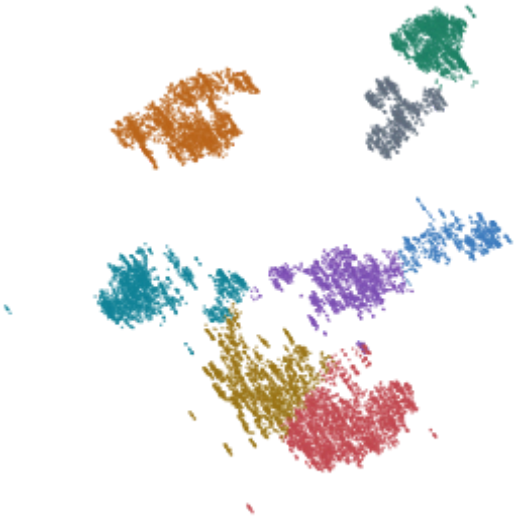
That led to a practical question: should borders and document framing be removed before indexing? A later experiment tried deterministic cropping and tone adjustment on 12 flagged images. Two changed their predicted category. That was enough to justify reviewing individual cases, not enough to justify automatically cropping the archive. Borders can contain evidence worth preserving.

Stepping back to see the whole collection

A search result shows what the system found. I built the MTL Archives Explorer to ask a different question: how had it organized the collection? Each of its 14,715 points represents one photograph. I could move through the projection, open a record and compare a group of images instead of guessing from a handful of thumbnails.

The shape of 14,715 photographs

Flat projection | Date depth | Reset view



All regions | Aerial NW | Street photos | Aerial central | Oblique views | Index cards | Aerial south | Survey documents | Aerial east

Select a region to highlight it. These colours follow the explorer's eight annotated regions; they stay attached to the same photographs as you rotate the view.

Turn | Tilt | Zoom

Drag to rotate, or use the controls. The flat view preserves the saved UMAP layout. Date depth uses the explorer's recorded-date mapping and seeded spacing, not a third UMAP dimension. Colour assignment follows the nearest of eight saved reference centres in the flat projection. The names are research annotations, not labels produced by UMAP or verified categories for every photograph. Directional names do not mark areas of Montréal. Open the full explorer to inspect individual photographs ↗

The flat view uses the saved UMAP coordinates. The 3D view keeps that layout and adds depth from the recorded date, with a little spacing to avoid stacked points. Its shape is a way to navigate the evidence, not the geography of Montréal or a third dimension discovered by UMAP. Colour views offer other ways to inspect it, including dates, photographer labels and annotated visual groups.

That made the split between aerial photographs, framed survey documents and index cards easier to investigate. I could move from an unusual group back to the images that formed it. Labels and anomaly highlights were prompts for review, not new archival facts. The useful outcome was a more specific question about the influence of document formatting, which led to the small cropping

experiment.

The explorer is built with Three.js, a browser graphics library. It loads the saved positions and record identifiers first, then fetches the larger vector file when needed. Similar-image lookup compares the original CLIP vectors, not distances on the flattened map. Selected records can be collected and exported for follow-up. I kept this research workspace separate from the main site's simpler search, game and print flows.

Trying a replacement before changing the index

In May I compared the existing CLIP model with SigLIP, Sigmoid Loss for Language Image Pre-Training, on a selected 500-image sample and 13 queries. The sample included aerials, documents and ground photographs, but it was uneven: 174 general aerials and only one ground-transit image. This was a local diagnostic, not a representative search benchmark.

Recorded measure	CLIP ViT-B/32	SigLIP base
Mean reciprocal rank	0.8269	0.4484
Queries with a rule match in top 5	13 / 13	8 / 13

May 26, 2026 · 500 images, 13 queries. Expected matches came from generated category/theme rules, not independent human relevance judgments.

Mean reciprocal rank asks how early the first expected match appears: first place scores one, second place one-half, and so on. The report called its other measure “P@5,” but the code actually checks whether any expected match appears in the first five results. I describe it as a hit rate here. It does not mean all five results were relevant.

Search the archive

tramway

Search

Try a colour, a scene or a Montréal landmark

tramway | red | trees | water | park | snowy street | Marché Bonsecours | pont Jac

6 results for “tramway” · saved starting selection

Rue McGill

1930s

Rue Kelly (devenue boulevard Henri-Bourassa) avant son pavage

1931

Marché Saint-Laurent (aujourd'hui démoli - 1180, boulevard Saint-...

1930s

Pont ferroviaire de la rue de La Montagne

1932

Avenue du Mont-Royal

1928

Pompiers à bord d'un véhicule équipé de la grande échelle devan...

1938

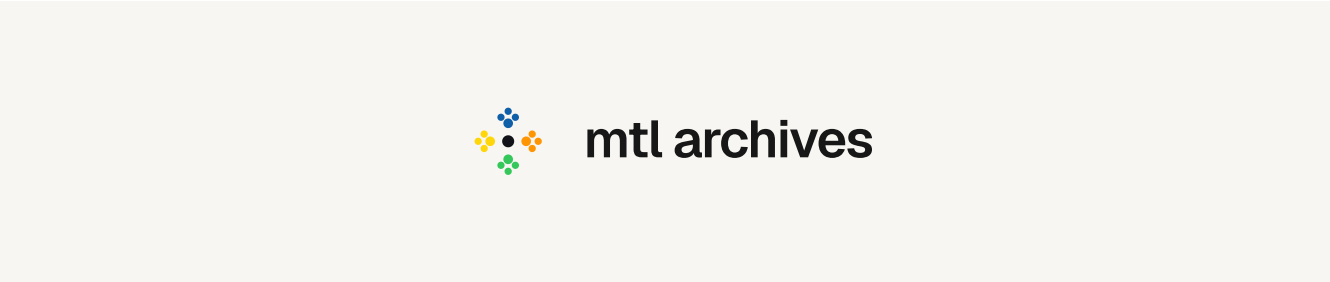
The same smart search used by MTL Archives, combining text and visual retrieval. Landmark names can match catalogue descriptions; colours and scenes invite broader visual associations. Open a photograph for its full record. Images: Archives de la Ville de Montréal. Continue on MTL Archives ↗

The recorded comparison supported keeping CLIP rather than rebuilding the production index around this SigLIP model. They did not show that CLIP was best for every archive question. The expected answers were broad category rules, and some queries were poorly represented in the sample. A stronger comparison needs independent judgments about whether each result answers the actual query.

The ranking experiments brought another useful negative result. Broad boosts from generated categories and quality labels underperformed the existing ranking on the recorded query set. A visually plausible park or waterfront image could move above a more directly relevant result. The documented decision was to keep those signals available for inspection without letting them change scores until better relevance labels support the change.

Designing a way into the archive

I worked through the identity and product states in Paper. The board covers the logo, typography, search, photo detail, game, print ordering, empty states and emails. That let me consider the same photograph as a search result, an archival record and a daily invitation to explore.



24px 48px 80px

The product's dotted rosette, rendered as a vector. The Paper exploration tested dense, medium, minimal and monochrome versions; this is the compact mark used by the site.

The rosette draws on Montréal's civic emblem and turns it into a small arrangement of points. I explored different densities so the idea could survive at icon size. The rest of the system gives the photographs room: paper-toned surfaces, dark text and restrained colour. Spectral carries editorial headings, Figtree handles interface copy, and IBM Plex Mono distinguishes archival details.

The interface is bilingual, with familiar places and subjects as entry points. A person arriving from a phone should be able to browse before learning how the catalogue works. On the record, the source reference and location confidence remain available. The design needs to make discovery easier without making uncertain metadata look authoritative.

A reason to return, and a way to collect

The daily location game asks people to place a photograph on a map. It reuses the archive rather than requiring a separate content library. The code keeps challenges and guesses in the backend, while the map and photo controls belong to the interface. A daily game makes a different invitation from search: you can start with curiosity instead of a query.

Print ordering uses Stripe Checkout. The app validates the shipping details and quote, then a signed payment notification triggers confirmation and fulfilment emails. The print work itself remains manual. A successful browser redirect is not treated as proof of payment.

The newsletter is another return path. Signing up is explicit; playing the game does not subscribe someone automatically. Subscription state and delivery history live in the database. The daily scheduler checks Montréal's local time, including daylight saving changes, rather than assuming the same server hour means morning all year.

These are working product surfaces, but their existence is not evidence of strong conversion. The saved business notes describe revenue as weak relative to attention. That is the next product problem, not a result I can claim to have solved.

Taking the archive to the feed

A searchable website still needs people to find it. I began using Instagram and Facebook as editorial experiments: exact places and dates, street transformations, lost landmarks and questions about what used to be there. Each post had to be worth looking at even if the viewer never ordered a print.

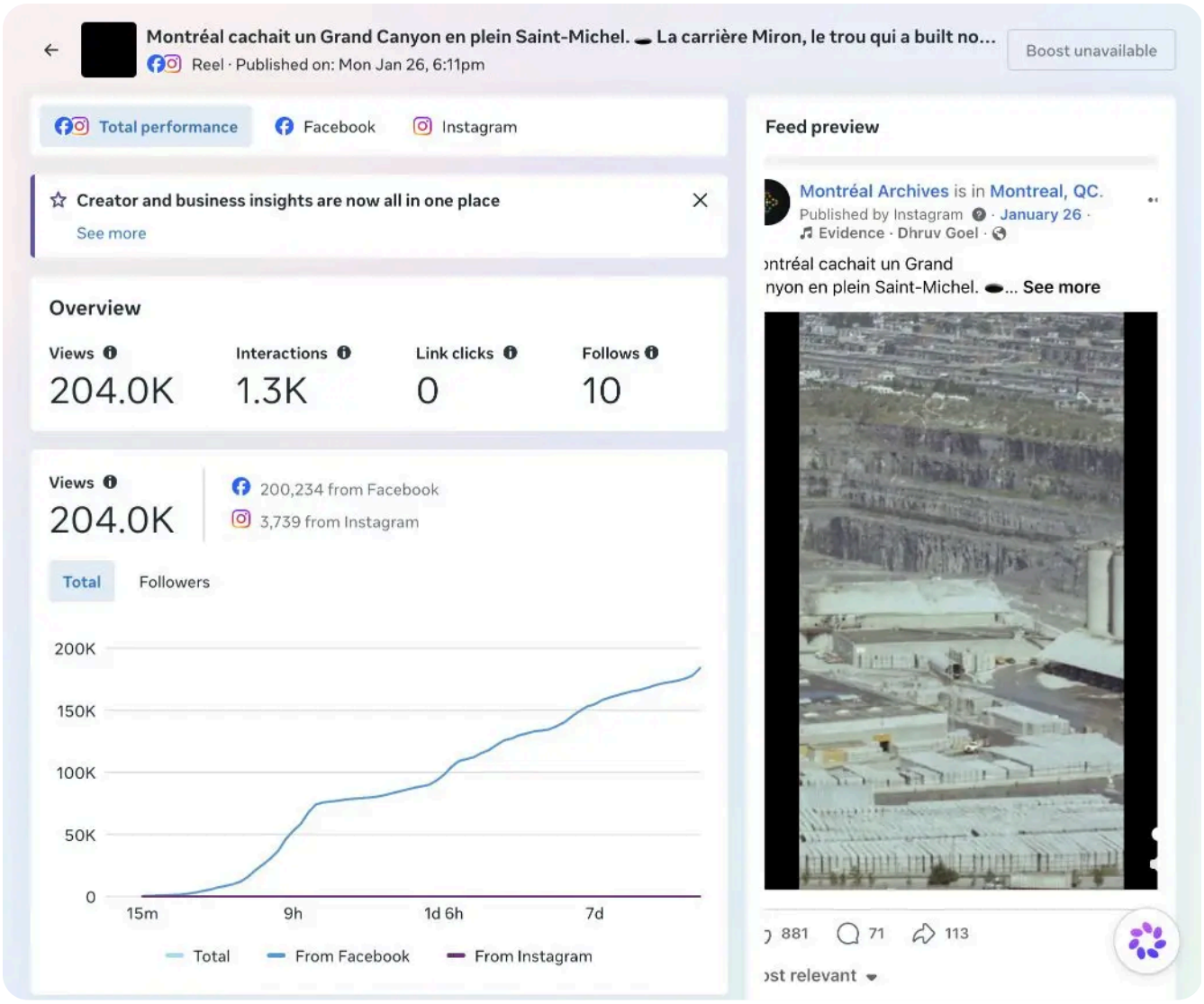
The daily pipeline selects an archive record, assembles its source material, drafts bilingual copy and produces a carousel or reel package. An image interpretation or a web search can suggest context, but unsupported exact locations should not quietly become facts. The code tracks location confidence and can reject copy that reintroduces a place name the evidence does not support.

Generating a package and publishing it are separate events. The pipeline records attempts, successful post identifiers and permalinks, which helps prevent duplicate delivery and makes later analysis possible. The home server holds operational state; an optional Obsidian mirror holds the editorial notes. A local fallback can prepare a package when the server is unavailable.

One archive, two editorial experiments

I was also experimenting with the writing around each photograph. The January caption formatter assembled a location and era, researched context, a surprising detail, a follow prompt and hashtags. The published reels tested a more dramatic opening: something lost, hidden or changed beyond recognition. The photograph supplied the evidence; the first line gave someone a reason to stop scrolling.

Facebook responded strongly to that approach. In the March 31 analysis of 135 first-quarter posts across both platforms, Facebook reels using loss or erasure language averaged 100,260 views, compared with 35,974 for reels without it. Openings such as “Une rue fantôme” made urban change the story. These were observed post-level averages at the time of the export, not views earned only in February or a controlled test of the recommendation algorithm.



The Miron quarry reel, published January 26. Captured September 9: 200,234 Facebook views and 3,739 Instagram views. These are cumulative post results, not February-only views. The same reel travelled very differently on the two platforms.

February concentrated that experiment: 18 Facebook reels averaged 62,519 views in the saved cohort. But the same format did not travel equally well to Instagram. There, February had 17 reels and only five carousels. The carousels averaged 9,468 views; the reels averaged 1,533. Optimizing both accounts around the largest Facebook number would have missed the difference.

Instagram's stronger pattern was more documentary: lead with a place and date, point out a detail, explain what changed or survived, and provide context in French and English. Across the first-quarter snapshot, place-and-date openings averaged 4,866 views versus 2,315 without them. Bilingual contextual captions also performed better in that comparison. Format, subject and posting date varied together, so I treated those findings as directions to test rather than isolated effects of the caption.

I shifted the writing toward that local-archivist voice. The later pipeline separates carousel and reel captions, checks whether generated text fits the story, and falls back to a structured template when it does not. Its reel instructions ask for a location and date, something visible to look at, concrete historical context and what survived. The March revision also filters generic mystery language. The engineering work was making that editorial choice repeatable, not just asking a model to write something engaging.

The change in mix is visible in March: Instagram moved to 12 carousels and eight reels, while its monthly account views recovered from 48,996 to 68,263. Facebook still published 17 reels, but their average in the saved cohort fell to 17,800. That helps explain why the February peak should not be read as a steady growth rate. It does not establish that a caption change alone caused the later decline.

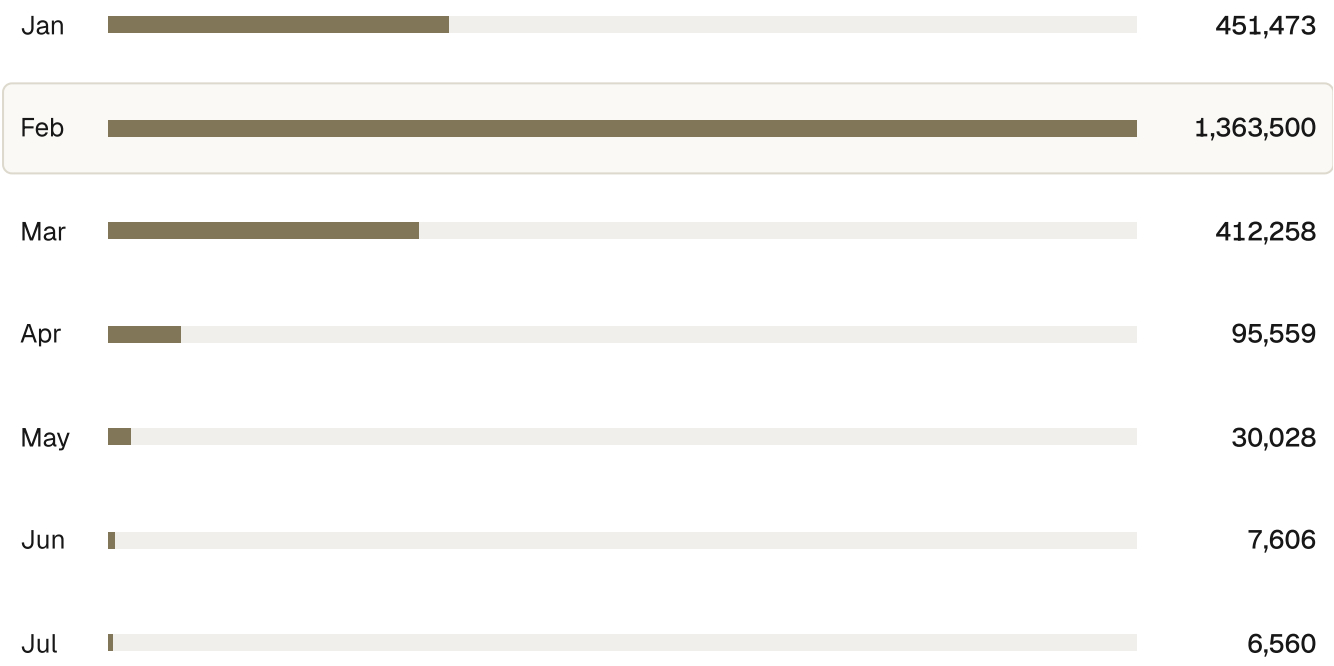
I could have kept testing the dramatic Facebook pattern. Instead, I chose to put more weight on context, specificity and the kind of archive I wanted people to return to. Repeating the same framing would not have guaranteed the same distribution. The result I can stand behind is narrower and more useful: I found two different editorial patterns, measured their tradeoffs, and changed the pipeline to reflect that choice.

The reach was real. It did not stay there.

The saved January–July reports total about 2.66 million Facebook and Instagram account-level views. February was the peak: 1,363,500 Facebook views and 48,996 Instagram views. By July, Facebook was down to 6,560 while Instagram recorded 25,787. Showing only the peak would miss most of what the experiment taught me.

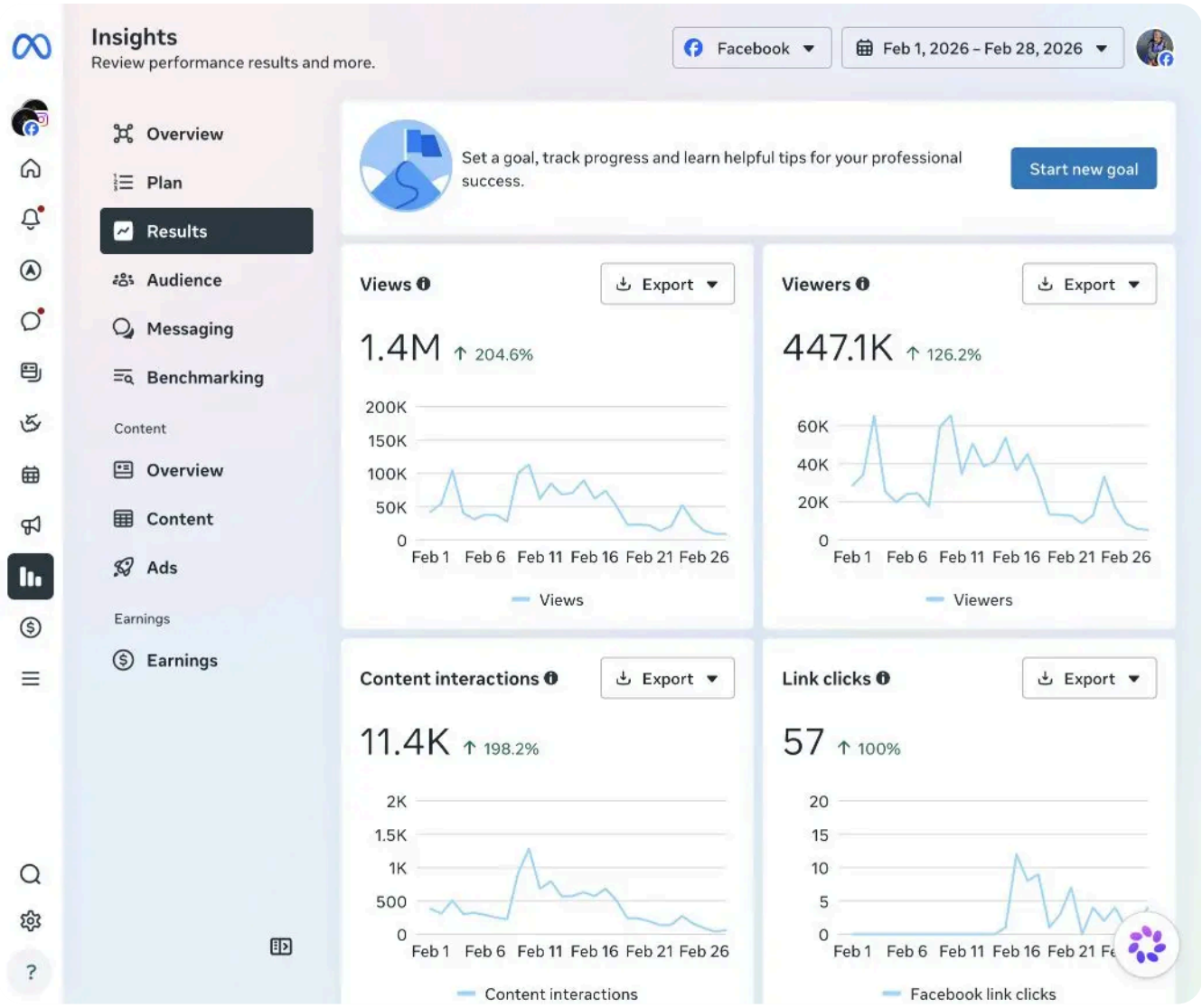
The spike, and what came after

Facebook views | Instagram views | Website visitors | Website page views

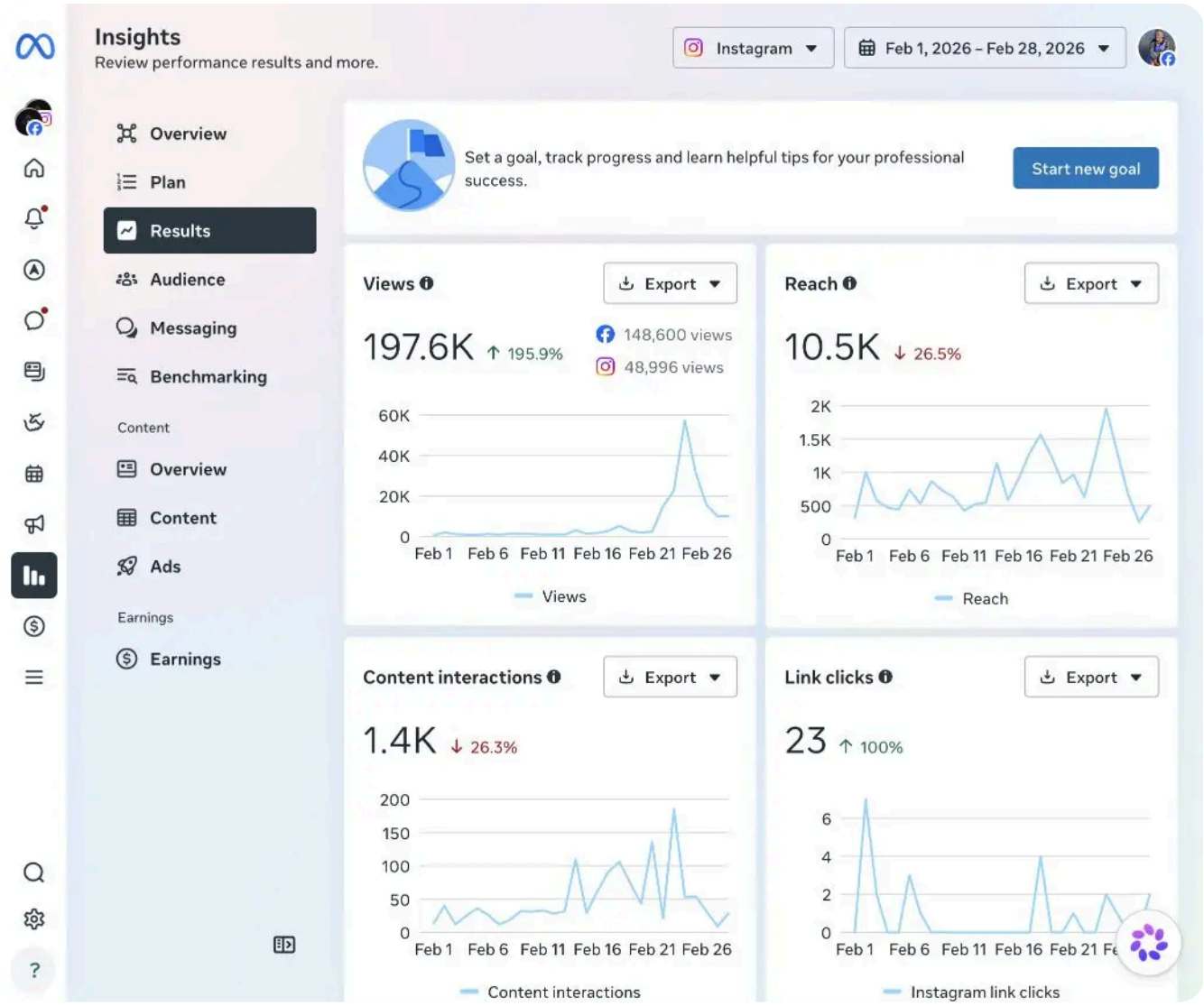


February: 1,363,500 Facebook views, 48,996 Instagram views and 4,747 website page views from 875 reported visitors.

January–July 2026 saved reports. Each measure has its own scale. March/April use saved monthly summaries; May has partial social coverage. Website January covers January 1–30; later months include dashboard captures. These are separate measures, not an attributed conversion funnel.



Facebook, February 1–28, 2026, captured September 9. Meta rounds the views total to 1.4M and reports 447.1K viewers. The saved report supplies the more precise 1,363,500 views used above.



Instagram, February 1–28, 2026, captured September 9. The headline 197.6K includes 148,600 Facebook views and 48,996 Instagram views. I use the Instagram breakdown, not the combined headline, in the monthly chart.

By September 9, Meta Business Suite showed approximately 8.7K Facebook followers and 3.7K Instagram followers. Those are account totals at the time of review, rather than followers gained during the February experiment.

The saved March 19 account export recorded 3,340 followers and 425 posts on Instagram (@mtlarchives). That is a dated account snapshot, separate from the view totals above. You can also explore the published work on Facebook.

A small number of reels accounted for much of the observed Facebook attention. In the saved August post snapshot, the top five unique January reels represented 82.4% of that month's reel cohort's cumulative views. These are lifetime post counts, not views accrued during January, so I do not add them to the monthly account totals.

Concrete places and a reason to be curious appeared repeatedly among the stronger posts. On Instagram, examples included Parc Marquette, 1969 and Avenue du Mont-Royal at Saint-Denis, 1928. The saved analysis found different patterns for documentary carousels and curiosity-led reels. These were observational comparisons; timing, format and platform distribution changed together, so they do not establish a causal recipe.

The data needed its own cleanup. Facebook's published-post export contained 196 reel rows for 98 unique reels. Counting the rows as separate pieces of content would double the denominator. Some monthly reports also lacked preserved daily exports. Those limitations remain in the supporting notes instead of being smoothed into an uninterrupted growth story.

What happened on the website

February brought 875 reported website visitors and 4,747 page views. March had fewer visitors, 714, but more page views, 5,296. The seven-month reports contain 13,783 page views in total. I do not sum monthly visitor counts and call that a unique audience: the same person can return in several months.

The gap between feed attention and website activity is the important result. The available exports do not reliably connect an individual post to a visit, game session or order. I can describe when activity rose and fell, but I cannot turn the social totals into an attributed conversion rate.

The later website captures show the smaller scale clearly: July recorded 130 visitors and 442 page views; August 1–10 recorded 33 and 122. The partial August period is kept out of the full-month chart. Better campaign links, preserved month-end exports and product events are needed to tell which forms of discovery lead to repeat archive use.

What changed how I work

I started by wanting to look through old photographs. I ended up learning how much the choices before the interface shape what someone can find. My own fallback descriptions made the records look more complete than they were. Keeping the city's metadata, extracted text, and generated captions separate became part of the product, not just housekeeping in the pipeline.

The model experiments changed how I read a benchmark. A promising replacement did not beat the simpler search model on my test set. Then I had to look harder at the test itself: a category rule is not a person deciding whether a photograph answers their question. I would now build a small, carefully reviewed set of real search tasks earlier, before spending more time tuning scores.

The explorer gave me another way to question the data. The islands in the map were compelling, but some reflected document formats rather than subjects. A visualization can reveal a pattern and still leave its meaning unresolved. I learned to keep the original photographs close enough that someone can check the interpretation for themselves.

Publishing taught me something different. A caption could travel widely on Facebook and do little to bring people into the archive. The February experiments made that distinction hard to ignore. I would set up campaign links and product events from the beginning, and review what actually gets published. A successful pipeline run does not tell me whether the caption is specific, useful, or even finished.

What I value most is that I could follow the work all the way through: from public records to search results, from a photograph to a post, and from attention to what people did next. Each part challenged an assumption I had made in another. That is the kind of research I want to keep doing: building something people can use, then letting the evidence change it.

The code and source notes for this case study show the implementation and limits behind the story. If you are opening up a difficult collection, evaluating search, or building a product around specialist data, I'd be happy to talk.