

.

## Profile (Account)

- Username (Handle)
- Profile Name (Display Name)
- Internal Profile ID (numeric)
- Profile picture URLs (all resolutions)
- Bio-text (incl. emojis)
- Pronouns (if visible)
- External links (website, linktree, shop, others)
- Verification status
- Account Type Flags (Creator/Business/Personal)
- Category/industry (e.g. „Artist“, „Retail Company“)
- Contact buttons & data, if public (e-mail, telephone, address, „directions“)
- Public/Private Account Status
- Follower number (as shown)
- Following number
- Number of posts (posts/reels)
- Time of last post
- „View Shop“/Shop Presence (if available; Shop URL)

## Posts / Reels / Carousel

- Post-ID (Shortcode + numeric)
- Permalink/URL
- Publication date & time
- Post type (image, video, carousel, reel)
- Caption (complete including hashtags, @mentions, emojis)
- Media URLs (images/video; original/CDN links; cover/thumbnail)

- Video metadata (duration, resolution, if applicable, frame rate/bitrate if visible in JSON)
- Geotag/place (name, place ID, address/coordinates if applicable)
- Tagged accounts in the image/video (handle, ID)
- Collaboration/Co-Author Info for Collab Posts (Handles, IDs)
- Branded-Content-/„Paid Partnership with ...“-Label (Partner-Handle)
- Shopping Tags/Product Tags (Product ID, Product Name, Destination URL)
- Visible like number (if not hidden)
- Visible number of comments
- List of all likers (by like overlay/popup): username, profile name, profile ID, profile picture URL, verification status
- Audio/Sound at Reels: Audio ID, Track, Original/Remix, Audio URL
- Effect/filter (name/ID, if displayed)
- Pinned status (whether post pinned)

#### Comments (per post/reel)

- Comment ID
- Comment text (incl. emojis, hashtags, @mentions)
- Timestamp of the comment
- Commentator: Username, Profile Name, Profile Picture URL, Profile ID, Verification Status
- Likes on comment (if visible)
- Replies/Threads: Parent comment ID, complete answer list with fields above

#### Stories (running stories)

- Story ID
- Time of publication
- Media URL (image/video; CDN link)
- Link Sticker Destination URL (if available)
- Tagged accounts (@mentions in the story)

- Music/Sound (Song Title, Artist)
- Effect/filter (name/ID, if displayed)

#### Story Highlights

- Highlight-Set-ID
- Highlight title
- Highlight Cover Image URL
- Order/position
- All included story items with: story ID, timestamp, media URL, link sticker URL, tagged accounts, music/sound

#### Guides (if available)

- Guide ID
- Guide title and description
- List of referenced posts (each post ID/URL)
- Order of elements

#### Shop/product pages (within Instagram)

- Shop/Catalog URL (if available)
- Product card from posts: product ID, product name, product image URL, destination URL (Insta-Shop or external)

#### Mentions & Relationship Indicators

- @Mentions in

#### Captions/Comments/Stories (Handle, ID)

- Tags/Tags in Images/Videos

#### (Handle, ID)

- „Paid Partnership with ...“-Partner (Handle, ID)

#### External/Deep Links

- Bio-Link target pages

(Linktree/Beacons/etc.): Destination URLs (first level)

- External shop links (e.g. B.

Shopify product/category page URLs)

- Imprint/contact page URLs (first level, if directly linked)

Instagram – complete list of all raw data to crawl from follower and follower lists

Profile (for each follower and each followed profile)

- Username (Handle)
- Profile Name (Display Name)
- Internal Profile ID (numeric)
- Profile picture URLs (all resolutions)
- Biography (full text including emojis)
- Biography entities (extracted @mentions, #hashtags, URLs)
- Pronouns (if visible)
- External links in bio (all target URLs; trace redirects to destination)
- Contact information from profile (public e-mail, telephone, address, contact buttons)
- Verification status (blue check)
- Account type flags (is\_private, is\_business\_account, is\_professional\_account, is\_creator\_account)
- Category / Industry / Profession (Business-Category / Creator Category)
- Follower number (as shown; exact number if present in JSON)
- Following number (as displayed; exact number if present in JSON)
- Number of posts (media\_count / posts)
- Time of last public post (UNIX timestamp, date/time)
- Shop presence flag + shop URL (if available)
- Imprint/company notice (if in bio or linked page)
- Privacy/visibility Flags (restricted\_country, blocked\_by\_viewer)
- Raw HTML of the profile page (Snapshot)
- Raw JSON / GraphQL answer(s) (if available)

#### Additional fields for Brands / Business Accounts (if detected)

- Business contact fields (business\_email, public\_phone\_number, address\_structured)
- Connected Facebook/Meta Page ID (if visible)
- Shop/Catalog Indicators (catalog\_id, shop\_owner\_id)
- Company name / brand name (from bio or linked page)
- Imprint / Company data (from linked page)
- Shop platform indicator (Shopify, WooCommerce, etc., if recognizable)

#### Additional fields for Creator accounts (if detected)

- Creator Category / Niche Tag (e.g. „Photographer“, „Fitness“)
- Contact CTA (e.g. E-mail button, collaboration notice)
- Total likes sum (if available as field)
- Reels/IGTV-specific flags (has\_reels, has\_igtv)
- Creator-external portfolio/link targets (YouTube, TikTok, Linktree)

#### For followed accounts (following) additionally crawl completely

(For each followed account: all the following data — profiles see above)

- Post list (all posts or complete list; if extremely many → complete metadata for the last N)
- For each post: Post-ID (shortcode + numeric)
- For each post: Permalink/URL
- For each post: publication date & time (UNIX-Timestamp)
- For each post: post type (image, video, carousel, reel, IGTV)
- For each post: Caption (complete including hashtags, @mentions, emojis)
- For each post: Caption-Entities (extracted hashtags, mentions, URLs)
- For each post: media URLs (all images/videos in the carousel; CDN links; cover/thumbnail)
- For videos: Video download URL (mp4, if available)

- For videos: video metadata (duration, resolution, possibly framerate/bitrate if in JSON)
- For images: image metadata (dimensions, file size, mime type, EXIF if available)
- For each post: Geotag/Ort (place name, place id, possibly address/coordinates)
- For each post: Tagged accounts in the image/video (handle + user\_id + possibly bbox)
- Branded-Content-/„Paid Partnership with ...“-Label (Partner-Handle, Partner-ID)
- Shopping tags/product tags (product id, product name, product URL)
- Audio/Music at Reels: Audio ID, Title, Original Creator, Audio URL
- Effect/Filter/AR Effect (Name/ID, if displayed)
- Visible like number (if not hidden)
- Visible number of comments
- Visible share/save indicators (if available)
- List of all likers (like overlay/popup): username, profile name, profile ID, profile picture URL, verification status, private/public flag
- List of all commentators (Comment-ID, Text, Timestamp, Commentator-Profile)
- Pinned status (whether post pinned)
- Raw Post JSON / embedded data

#### Comments (for posts from followed accounts)

- Comment ID
- Comment text (incl. emojis, hashtags, @mentions, URLs)
- Comment Timestamp (UNIX)
- Commentator: username, profile name, profile ID, profile picture URL, is\_verified, is\_private
- Likes on comment (like\_count, if visible)
- Parent/Thread ID (if reply)
- Full thread chain (recursive)
- Raw comment JSON

### Stories / Story-Highlights (for followed accounts)

- Active Story-Items: Story-ID, Veröffentlichungszeit, Medien-URL (CDN), Media-Type (image/video), Duration (video)
- Story Link Stickers Goals (Target URLs)
- Story Mentions / Tags (handles + ids)
- Story Music/Audio Info (title, artist, audio\_id)
- Story Effect/Filter ID
- Story highlights: Highlight set ID, title, cover image URL, order, included story item IDs with item metadata
- Raw Story JSON (if available)

### Follower list (for each account found in the follower list - see profile above)

- Entries of the follower list: the profile fields above apply to each entry
- Position/index in follower list (relative position, if available)
- Mutual-Follow-Flag (follows\_back) if recognizable
- Note: if profile private → only minimally visible header information (username, user\_id, is\_private)

### Following list (for each account followed by the creator - crawl all accounts; see profile above)

- Entries in the following list: the profile fields above apply to each entry
- For each followed account: complete post and story data as described in „For followed accounts“
- For each followed account: follower and following counts (if public) and (if necessary) follower/following lists according to the respective sampling decisions

### Fallback sources with limited follower list visibility

- Accounts that liked the last N posts → each profile fields (see profile)
- Accounts that have commented on the last N posts → each profile fields
- Accounts that were mentioned/tagged in captions, comments or stories → each profile fields
- Accounts that appear multiple times in Reels audio credits or video captions → each profile fields

- Aggregation of all unique accounts from likes/comments/mentions until desired sample size is reached

External links / deep links (for each target found in bio or story/link sticker)

- Destination URL for redirects
- Page Title + Meta Description
- Contact email(s) extracted (Regex)
- Extracted phone numbers
- Imprint / Company data (if available)
- Shop product excerpts (title, image, price, product URL) — first level
- Shop platform indicator (Shopify, WooCommerce, etc.)
- Raw HTML of the landing page

Relationship and collaboration indicators (to be captured from profile, posts, captions, comments, stories)

- @Mentions in Captions/Comments/Stories (Handle, ID)
- Tags/Tags in Images/Videos (Handle, ID)
- „Paid Partnership with ...“ Partner-Handle/Partner-ID
- Co-Author/Collab-Info (if post marked as Collab/Co-Author)
- Common mentions between creator and account (evidence of interaction)
- Liker/Commenter-Edges (Post-ID → user\_id of liker/commenter)
- Mutual Follows / Reciprocal-Follow-Edges (if recognizable)

Heuristics / Flags for classification (derivable from profile & external information)

- Is\_brand\_candidate (bio contains shop/brand keywords or link leads to shop)
- Is\_creator\_candidate (category flagged, follower\_count threshold, active Reels/IGTV)
- Is\_private\_flag
- Is\_verified\_flag
- Contactable\_flag (public email or contact-button available)

Other potentially relevant raw data

- Pinned posts / Marked posts in profile (IDs + metadata)
- Guide references (guide ID, title, referenced post IDs)
- Hashtag-Use-History in profile (list of all hashtags used in bios/captions)
- Mentioned third-party platforms (YouTube/TikTok/Twitter handles in Bio/Captions)
- Possible affiliate/tracking links (parameters in bio links)
- Visible promotions/ads in profile or posts

Note on restricted access extraction

- If direct follower lists blocked/invisible: replace by systematic collection of likes/comments/mentions/tags; verify identity overlap (username + id) and eliminate duplicates.

From followers and followed and one himself additionally:

Profile information (optional, supplementary)

- Language/Locale of the profile (if recognized from posts/stories)
- Registration date / Account-Age (if visible from GraphQL)
- Number of archived posts / story highlights (if any)
- Category of external links (shop, social media, portfolio)
- Flags for linked meta/Facebook page
- Organic keywords for brand/product identification (e.g. „Official“, „Shop“, „Collab“)

Posts / Reels / Carousel

- Visible shares/saves (if derivable)
- Polls, quizzes, questions in carousel or reel posts (if interactive)
- Analytics notes from captions (e.g. „Link in Bio“ click prompts)
- Linked products outside the Instagram store (external links in post captions)

Comments

- Comment Author Account Type (Brand, Creator, Private)
- History of interaction between commentator and creator (if previous posts exist)

Stories / Story Highlights

- Story viewer list (if accessible or derivable from reactions)
- Interactive sticker answers (e.g. Answers to surveys/questions)
- Story duration per viewer (if derivable; optional)

#### Follower/Following

- Follower/Following-Engagement-Level (e.g. Share of active interactions in the last N posts)
- Relative Position in Follower List (Top, Mid, Tail) for Sampling
- Mutual Follows / Overlap with other relevant accounts (Creator, Brands)

#### External Links / Deep Links

- Domain Type (Shop, Social Media, Blog, Affiliate, Portfolio)
- Affiliate/tracking parameters in links
- Presence of imprint or legally relevant information

#### Network and relationship data

- Collaboration patterns: who is frequently tagged or mentioned
- Relevant follower interactions with brands (likes/comments on brand posts)
- Frequent interaction between followers (e.g. Comments on posts from the same brands/creators)

#### Sampling rules / crawl quantities

##### 1. Creator-Followers

- < 5,000 followers → crawl all followers
- 5,000–50,000 → 30% of followers, stratified by position in list (Top/Mid/Tail) and activity
- 50,000–1,000,000 → 20% of followers, stratified, priority on active creator/brand accounts
- 1,000,000 → 10% of followers, maximum 100,000 accounts, priority Top + Brands + Creator

##### 2. Own Followed

- Usually < 1,000 → crawl all

##### 3. Brand Accounts

- Always crawl completely, including followers, posts, stories, shop/product pages

#### 4. Followers of Followers / Following of Following

- For micro-creators (<50k) → sampling 20–30% with relevance (e.g. Brand or creator)
- For mid-accounts (50k–1M) → 10–20 % sampling, active priority/highly interacting accounts
- For large accounts (>1M) → 5–10% sampling, max 50,000–100,000 accounts, priority on brands/creators

#### 5. Additional rules

- Private profiles → only publicly visible data (if follower list blocked → use likes/comments/mentions)
- Top interaction accounts (regularly likes/comments) → always crawl, even if otherwise lower sampling
- At Stories → always crawl all current stories (as time-limited)