

# Variable documentation for the somepressa24 corpus

Laaksonen, S.-M., Skänderi, E., Nelimarkka, M., Piitulainen, J., Rikkonen, L., Toroskainen, S., & Pöyry, E. (2026). Suomen presidentinvaalit 2024 sosiaalisessa mediassa, lähdemateriaali [aineisto]. Kielipankki.

<https://urn.fi/urn:nbn:fi:lb-2024030501>

## Instagram posts from candidates *Instagram\_Posts\_candidates.csv*

id	Original id number of the post given by the platform.
published_at	Time of post publication in an ISO 8601 format YYYY-MM-DDTHH:mm:ssZ
published_at_unix	Time of post publication in an unix timestamp format
url	Link to the original post on the platform
body	Full text content of the post
author	Username of the account that posted the post
author_fullname	Visible full name of the account that posted the post
verified	Indicates whether the author account had the verified status on the platform.
author_avatar_url	Link to the profile picture of the author.
coauthors	Potential coauthors of the post, if invited by the original author.
coauthor_fullnames	Visible full names of the coauthors, if any.
coauthor_ids	Id numbers of the coauthors, if any.
media_type	Format of the post (photo, video, mixed)
num_media	Number of media files (images, videos) in the post
hashtags	List of hashtags used in the post text, separated with commas
usertags	List of other users tagged in the post, separated with commas
likes_hidden	Boolean value indicating if post likes are hidden (Yes/No)
num_likes	Number of likes the post received (collected April 4, 2026)
num_comments	Number of comments the post received (collected April 4, 2026)
location_name	Location of posting if inserted by the user
location_id	Instagram internal id for the location
location_latlong	Latitude and longitude coordinates of the chosen location

asset_count	Number of media files related to the post available in the database
asset_path	Path to the first media file related to the post available in the database. If the post had multiple media files, this is the first one (also shown first on the platform)
asset_paths	Paths to media files related to the post available in the database
round	Indicates whether the post was published during the first or second round of the elections

**Instagram stories from candidates** *Instagram\_Stories\_candidates.csv*

published_at	Time of post publication in an ISO 8601 format YYYY-MM-DDTHH:mm:ssZ
published_at_unix	Time of post publication in an unix timestamp format
username	Username of the account that posted the post
round	Indicates whether the post was published during the first or second round of the elections
asset_path	Path to the media file related to the post available in the database

**TikTok from candidates** *TikTok\_candidates.csv*

id	Original id number of the post given by the platform.
thread_id	Original id number of the thread given by the platform.
author	Username of the account that posted the post
author_full	Visible full name of the account that posted the post
body	Full text content of the post
published_at	Time of post publication in an ISO 8601 format YYYY-MM-DDTHH:mm:ssZ
published_at_unix	Time of post publication in an unix timestamp format
is_duet	(True/False)
is_ad	Indicates if the post was a paid advertisement or not (True/False)
music_name	Name of audio attached to the post, if any.
music_id	Id number of audio attached to the post, if any.
tiktok_url	Link to the original post on the platform

likes	Number of likes the post received (collected March 20, 2024)
comments	Number of comments the post received (collected March 20, 2024)
shares	Number of times the post was shared (collected March 20, 2024)
plays	Number of times the post was played (collected March 20, 2024)
hashtags	List of hashtags used in the post text, separated with commas
challenges	List of challenges used in the post text, separated with commas
warning	Shows a potential content warning that was shown with the post on the platform
asset_path_video	Paths to video file related to the post available in the database
asset_path_cover	Paths to cover image file related to the post available in the database
asset_path_image	Paths to image file related to the post available in the database
has_video	Indicates if there is a related video file
has_cover	Indicates if there is a related cover image file
has_image	Indicates if there is a related image file
round	Indicates whether the post was published during the first or second round of the elections (round1, round2)

**YouTube from candidates** *YouTube\_candidates.csv*

**YouTube search results** *YouTube\_general.csv*

video_id	Original id number of the video given by the platform.
channel_id	Original id number of the posting channel given by the platform.
channel_title	Visible full name of the channel that posted the post
published_at	Time of post publication in an ISO 8601 format YYYY-MM-DDTHH:mm:ssZ
published_at_unix	Time of post publication in an unix timestamp format
video_title	Title of the video as shown on the platform
video_description	Textual description of the video as shown on the platform
tags	List of tags used with the video, separated with commas
video_category_id	Original id number of the video category, defined by the platform.

video_category_label	Label of the video category, defined by the platform.
topic_categories	A list of topic keywords that provide a high-level description of the video's content. Most likely automatically generated by YouTube.
duration_iso	Duration of the video in
duration_sec	Duration of the video in seconds.
dimension	Video dimensions.
definition	Video definition quality
caption	Boolean value (true/false) which indicates if there are captions in the video.
default_language	The language of the text in the video resource's title and description properties.
default_audio_language	Specifies the language spoken in the video's default audio track.
thumbnail_maxres	Maximum resolution version of the thumbnail image, url to file.
licensed_content	Boolean value (1/0) which indicates whether the video represents licensed content, which means that the content was uploaded to a channel linked to a YouTube content partner and then claimed by that partner.
location_description	The text description of the location where the video was recorded, if available.
latitude	Latitude coordinates of the chosen location
longitude	Longitude coordinates of the chosen location
view_count	Number of times the video was played (collected Jan 31, 2024)
like_count	Number of likes the video received (collected Jan 31, 2024)
dislike_count	Number of dislikes the video received (collected Jan 31, 2024)
favorite_count	Number of times the video was favorited (collected Jan 31, 2024)
comment_count	Number of comments the video received (collected Jan 31, 2024)
youtube_url	Link to the original video on YouTube.
original_query	The query used on YouTube when collecting the data
source_kind	Indicates whether the data was collected using the channel name as a starting point or keyword search

**Facebook from candidates** *Facebook\_candidates.csv*

page_name	Visible full name of the Facebook page that posted the post
user_name	Username of the account that posted the post
facebook_id	Id of the Facebook page that published the post.

published_at	Time of post publication in an ISO 8601 format YYYY-MM-DDTHH:mm:ssZ
published_at_unix	Time of post publication in an unix timestamp format
page_category	A topical category of the page, selected from a list given by the platform.
page_admin_top_country	page's top administrative countries
page_description	The description of the page as documented in Page Transparency information.
page_created	Time when the page was created.
likes_at_posting	Number of page likes at the time of post publication
followers_at_posting	Number of page followers at the time of post publication
type	Post content type (album, igtv, link, live_video, live_video_complete, live_video_scheduled, native_video, photo, status, video, vine, youtube)
total_interactions	Total number of interactions received for the post (collected March 6, 2024)
likes	Number of likes the post received (collected March 6, 2024)
comments	Number of comments the post received (collected March 6, 2024)
shares	Number of times the post was shared (collected March 6, 2024)
love	Number of times the post received a love reaction (collected March 6, 2024)
wow	Number of times the post received a wow reaction (collected March 6, 2024)
haha	Number of times the post received a haha reaction (collected March 6, 2024)
sad	Number of times the post received a sad reaction (collected March 6, 2024)
angry	Number of times the post received an angry reaction (collected March 6, 2024)
care	Number of times the post received a care reaction (collected March 6, 2024)
video_share_status	Indicates if the video is recognized as original content shared by the page.
is_video_owner	Boolean value to indicate is the publisher the owner of the video.
post_views	Number of times the post was viewed (collected March 6, 2024)
total_views	Number of times the post was viewed (collected March 6, 2024)
total_views_for_all_crossposts	The total number of views, across all crossposts, for a Facebook native video that has been cross-posted on Facebook on several pages.
video_length	Video length hh:mm:ss
url	Link to the original post on the platform
message	Text content of the post.

link	Link shared in the post by the publisher.
final_link	The unshortened link, if a URL has been shortened.
image_text	Text extracted from the image using OCR, processed by CrowdTangle.
link_text	The headline of a link URL or the title of a native video (often the html page title).
description	For posts that include a link, the sub-header of a link URL, i.e., the text that shows up under a link, extracted from the HTML of the linked page.
round	Indicates whether the post was published during the first or second round of the elections

**Twitter from candidates** *Twitter\_candidates.csv*

**Twitter general discussion** *Twitter\_general.csv*

published_at	Time of post publication in an ISO 8601 format YYYY-MM-DDTHH:mm:ssZ
published_at_unix	Time of post publication in an unix timestamp format
tweet_id	Original id number of the tweet given by the platform.
thread_id	Original id number of the thread given by the platform, corresponds to the id of the first tweet in a thread.
author_id	Original id number of the author account, given by the platform.
author_screen_name	Username of the account that tweeted the post
author_name	Visible full name of the account that tweeted the post
author_followers	Number of followers the account had at the time of posting
author_friends	Number of accounts followed by the account (i.e. friends) at the time of posting
author_statuses	Total number of posts the account had posted at the time of posting
lang	Original language the tweet was posted in
full_text	Full text of the tweet
hashtags	List of hashtags used in the tweet
mentions	List of users tagged in the tweet
urls	List of urls mentioned in the tweet
media_types	Type of media content shared in the post, if any (photo, video)
views	The total number of times a post has been viewed.

likes	Number of likes the tweet received (collected daily during campaigning)
replies	Number of replies the tweet received (collected daily during campaigning)
retweets	Number of times the tweet was retweeted (collected daily during campaigning)
quotes	Number of times the tweet was quoted (in a retweet) (collected daily during campaigning)
bookmarks	Number of times the tweet was bookmarked (collected daily during campaigning)
is_quote_status	Boolean value that indicates if the the tweet references another Tweet (True/False)
possibly_sensitive	Boolean value that indicates if the tweet potentially has sensitive content. Can originate, for example, from user settings or flagged URLs.

**News comments** *News\_Comments\_general.csv*

source_media	The news media outlet that published the news the comment relates to
article_id	The id of the article on the original hosting news media outlet (given by the media)
topic_id	The topic id of the article on the original hosting news media outlet (given by the media)
comment_id	Original id number of the comment given by the hosting news site.
parent_id	Id number of a parent comment if the discussion was in threads
comment_text	Text content of the comment
published_at	Time of comment publication in an ISO 8601 format YYYY-MM-DDTHH:mm:ssZ
published_at_unix	Time of comment publication in an unix timestamp format
positive_votes_hs	Positive votes received by the comment (only on HS)
likes_yle	Likes received by the comment (only on YLE)
author_display_name	Visible full name of the account that posted the post
author_realname_id	Id number of the account that posted the post
parent_comment_text_hs	Full text of a parent comment if the discussion was in threads (HS only)

**Combined** *somepressa24\_candidates\_allposts.csv*

platform	The platform on which the post was published on Facebook 774 Instagram 1570 TikTok 1286 Twitter 73 Youtube 709
post_id	Original id number of the post given by the platform.
username	Username of the account that posted the post
name	Visible full name of the account that posted the post
published_at	Time of post publication in an ISO 8601 format YYYY-MM-DDTHH:mm:ssZ
published_at_unix	Time of post publication in an unix timestamp format
posttype	Type of the post, including platform and the post type if available.  Facebook post types: Link_FB 50 Live_Video_Complete_FB 3 Live_Video_Scheduled_FB 25 Native_Video_FB 79 Photo_FB 587 Status_FB 26 Video_FB 1 YouTube_FB 3 (= Youtube video shared on Facebook)  Instagram post types: mixed_IG 35 (= both photo and video on Instagram) photo_IG 1083 video_IG 452  Tiktok post types: video_TT 1286  Twitter post types: Tweet_TW 73  YouTube post types: Video_YT 709
posttext	Full text content of the post
media_type	TO BE REMOVED
commentsreplies	Number of comments or replies the post received (collection date varies)
likes	Number of likes the post received (collection date varies)
shares	Number of times the post was shared (collection date varies)
original_url	Link to the original post on the platform

tags	List of (hash)tags used in the post text, separated with commas
shared_urls_TW	List of urls mentioned in the post (Twitter only)
video_title_YT	Title of the video as shown on the platform (YouTube only)