

UCLA Statistics and Data Science Winter Datathon

Column Description

Here is a short description of each column in the dataset.

Columns:

Timestamp: Timestamp when the post was scraped, seconds from January 1, 1970

subreddit: The name of the subreddit where the post was made

selftext: The text content of the post

is_gallery: Whether the post contains multiple images in a gallery format

title: The title of the post

thumbnail_height: Height of the thumbnail image (if applicable)

upvote_ratio: Ratio of upvotes to total votes

ups: Number of upvotes the post received

thumbnail_width: Width of the thumbnail image

score: The overall score (upvotes minus downvotes)

author_premium: Whether the author is a Reddit Premium user

edited: Whether the post was edited

content_categories: Categories assigned to the post (if any)

allow_live_comments: Whether live comments are enabled (similar to live chat)

media_only: Whether the post is media-only (no text)

spoiler: Whether the post is marked as a spoiler

id: Unique post ID

author: The username of the post's author

num_comments: number of comments

stickied: Whether the post is stickied (stays at the top of the subreddit)

url: The URL of the post

subreddit_subscribers: Number of subscribers in the subreddit

created_utc: Timestamp when the post was created in UTC, seconds from January 1, 1970

num_crossposts: The number of times this post was posted on another subreddit

is_video: Whether the post contains a video

post_hint: Type of post (image, link, etc.)