

# Reddit

## Table of Contents

### [1 About the Platform](#)

### [2 Available Data](#)

### [3 Restrictions and Limitations](#)

### [4 Available Metadata](#)

### [5 Relevant Application Programming Interface \(API\)](#)

### [6 Searching by Geolocation](#)

### [7 Filtered Content](#)

### [8 References](#)

---

#### About the Platform:



Reddit is a social news aggregation, web content rating, and discussion website. Reddit allows users to post freely without “censorship.” It is a social media platform based on users sharing to subreddits (i.e., sub-threads) that contain different topics of conversation. Moderators, or rulers of those subreddits, watch and guide the conversations. Reddit allows for users to express ideas, opinions, and themes with similar minded people. The platform itself even states, “Whether you're into breaking news, sports, TV fan theories, or a never-ending stream of the Internet's cutest animals,

there's a community on Reddit for you” (Reddit, 2020). Reddit allows for users to find a safe space to share all kinds of unique and interesting ideas.

## **1) What data is available from this platform?**

“Real-time data is available including comments, submissions and links posted to reddit” (r/datasets, 2016). This means that posts, geolocations, comments, usernames, and other resource types are open to researchers. These posts can be collected for analyzing various topics embedded within the Reddit data. Because Reddit has such a cloud of anonymity, users can express ideas free of consequences. This allows for content that is more transparent and honest overall.

## **2) What are the restrictions/limitations of the data?**

Reddit may set and enforce limits on your use of the Reddit APIs (e.g., limiting the number of API requests that can be made or number of users you can serve).

Clients connecting via OAuth2 may make up to 60 requests per minute. The following response headers should be monitored to ensure that limits are not being exceeded:

- X-Ratelimit-Used: the approximate number of requests used in the period
- X-Ratelimit-Remaining: the approximate number of requests left to use
- X-Ratelimit-Reset: the approximate number of seconds until end of period

**3) What variables are available? What meta-data (i.e., the data of the data) is included? For example, what information is available about user profiles (e.g., username, demographic data, number of followers/friends)? What information is linked to the content/post/blog/tweet (e.g., timestamps, re-tweets, location, device/platform that content was shared from, number of likes)?**

```

{
  "data": [
    {
      "author": "MockDeath",
      "author_flair_css_class": null,
      "author_flair_text": null,
      "body": "Knowing more would definitely help. I guess all you can do is find",
      "created_utc": 1270637661,
      "id": "c0nn9iq",
      "link_id": "t3_bne3u",
      "parent_id": "t1_c0nn5ux",
      "score": 2,
      "subreddit": "askscience",
      "subreddit_id": "t5_2qm4e"
    }
  ],
  "metadata": {
    "execution_time_milliseconds": 30.52,
    "results_returned": 1,
    "shards": {
      "failed": 0,
      "successful": 36,
      "total": 36
    },
    "size": 1,
    "sort": "asc",
    "sort_type": "created_utc",
    "timed_out": false,
    "total_results": 134785,
    "version": "v3.0"
  }
}

```

Figure 1. An example response from the Reddit API in the form of JSON objects.

#### 4) Is there an Application Programming Interface (API) available for this platform? If yes, please provide the link for the platform.

Yes, there is both an official and unofficial API for Reddit:

1. **The official Reddit API:**  
<https://www.reddit.com/dev/api/>
2. **The unofficial Reddit API on GitHub (pushshift/api):**  
<https://github.com/pushshift/api>.

Our team opted to use the unofficial API to avoid the frequent “catch-all” HTTP responses that Reddit’s official API was often producing. This response is likely attributed to the heavy traffic that the official API receives, causing the internal server to overload. Fortunately, we could retrieve the same kind of Reddit information from the unofficial source, without having to worry about potential server overload.

**5) Does the API provide a way to limit the search based on geolocation?**

No, Reddit does not provide a way to limit the search based on geolocation.

**6) Does the platform use any kind of filters to remove the content on their sites?**

The company has begun promoting a number of resources containing authoritative information related to COVID-19, stating that unless a subreddit is focused on spreading misleading content, admins will prioritize educating and cooperating with users in the subreddit. If these efforts fail, the platform will then take steps to ban the subreddit, in a process known as “quarantining.” When a community is quarantined, it does not appear in search results. Additionally, if a user tries to visit the quarantined community, they will be notified that the subreddit may contain misleading content and they must explicitly opt-in to viewing the content.

## References

Baumgartner, J. M. (n.d.). *pushshift/api*. GitHub. <https://github.com/pushshift/api>.

Reddit. (2020, January). *Homepage*. <https://www.redditinc.com/>

reddit.com: api documentation. (n.d.). <https://www.reddit.com/dev/api/>.

Stuck\_In\_the\_Matrix. (2015, September 27). *r/datasets - Real-time data is available including comments, submissions and links posted to reddit*. reddit. [https://www.reddit.com/r/datasets/comments/3mk1vg/realtime\\_data\\_is\\_available\\_including\\_comments/](https://www.reddit.com/r/datasets/comments/3mk1vg/realtime_data_is_available_including_comments/).