

# In the Future

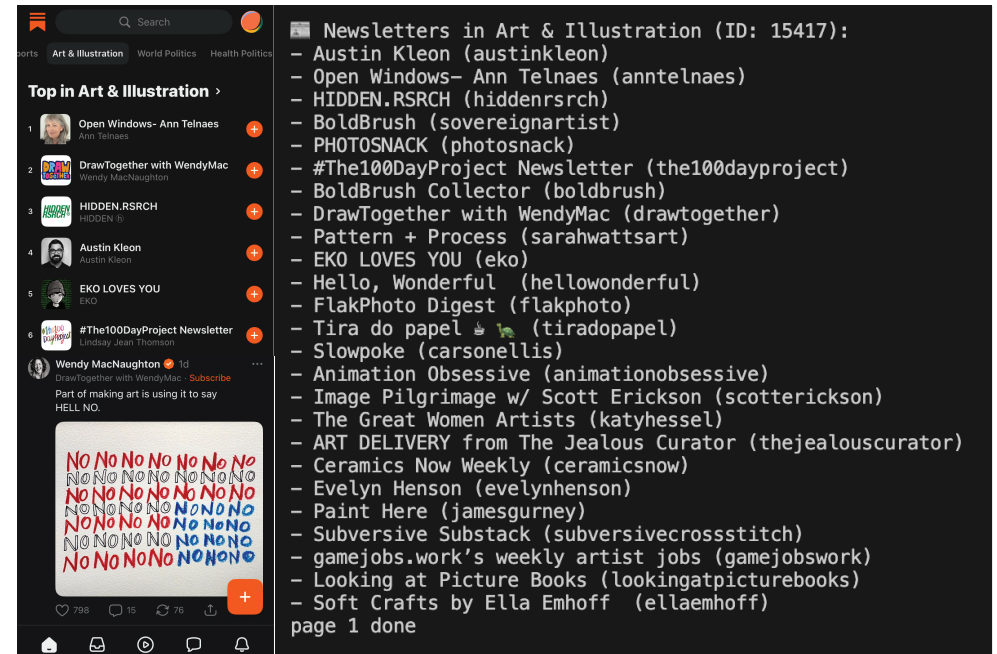
How do you think will the collection and analysis and insights evolve in the next 10 years?

I think that there will be lots of evolution in the next 10 years as Substack has grown and the role and structure of long form content continues to shift. As a collection I wouldn't be surprised if it became harder to access much of the data. It may be shifting towards more of a social media hybrid rather than primarily newsletter distribution which might mean, like other large platforms, it restricts how data can be collected. I think before that happens, there will be greater analysis of Substack data as an example of opinionated long form content and an alternative from mainstream media.

In which areas do you see the biggest future potentials or value of the data?

I think the biggest potential is in contrast to other available platforms which together make up the digital landscape of how audiences get information, generate content, and build community. I assume others see potential in AI with text generating and suggestions with this data, though to me that is not of much value as all, as I feel it dilutes the expression and is not necessary to optimize and personalize. I think in an attention scarce society, data about how willing people are to spend time on new ideas is important. I also think as an people flock to niche thinkers it acts as another measure of trends and trust in public opinion.

# Substack Api



What is the technology or service?

Substack is an American online platform that allows journalists, writers, and other content creators to publish newsletters directly to their audience. It has a subscription model where readers can sign up for free or paid digital newsletters, enabling writers to monetize their content without relying on ads. The company retains 10% of subscription revenue. While primarily text-based, it also offers podcasts, discussion threads, and videos. Substack has less content moderation than many other platforms which has been the main reason for criticism against it.

How does the raw data look like?

I'm using a Python client library so I cannot see the full data, just formatted requests which can be accessed through retrieving newsletter or user.

category\_id, category\_name, newsletter\_reccomendations, slug, newsletter\_slug, title, url

user\_id, username, user\_likes, post, user\_notes, user\_reads

What behaviors does it capture?

The behaviors that it captures would be publications of newsletter posts, and user likes, user reads, user notes (when user id is known). It can capture likes on a post but I'm not sure if that is possible over time or just at the moment the user liked it what the count was.

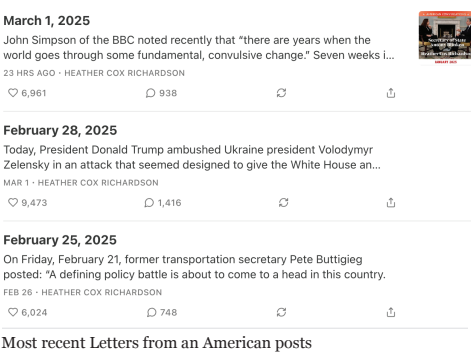
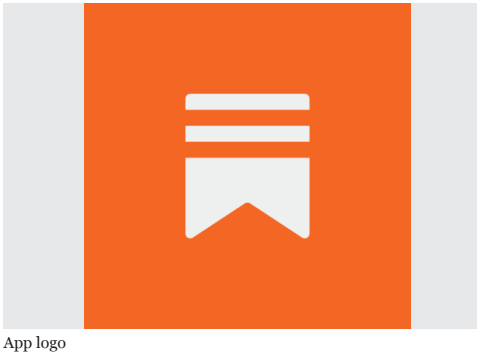
I think it would also be possible to track changes to username and bios.

# Visualizations

Visual Examples of this technology:



Graph of highest earning Substack Newsletters (2024) Backlinko



# Insights

What are the main types of insights that can be derived from the data?

- Headlines  
We can get titles for posts using a specific newsletter or for a category.
- User interests  
By finding trends in the user reads or user likes posts we can see what an individual interests with.

- Connections  
How much in common you have with another reader or writer.
- Sentiment  
Text analysis on collections of posts with use of additional tools to determine how positive/negative they are or what emotions are present.

# Usage

How are the insights currently being used?  
By who?

- Github users  
I'm working with an unofficial Substack API that a Github user made and so I can see that 41 people have starred it.
- Bloggers  
I think mostly data about Substack is used by media blogs and marketers who want to follow trends in technology.

- Researchers  
People who study public opinion: academics and market researchers might look at Substack data.
- Content Creators  
People who distribute content on Substack use the data it provides them to track their engagement and optimize content.