

In the Future

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

I think collection might evolve to include trip paths and conditions and will allow for better urban planning.

I think over time insights will better predict use patterns and needs and operations will become more efficient with prediction accuracy.

I know that citi bike has a bike angels program where users can assist in rebalancing bike flows and hotspots. I think greater analysis of trends will allow for more optimized and dynamic pricing incentives that draw users to redistribute where it is needed most.

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

There is huge value in studying the use of human powered transportation which has a lower environmental impact, takes up less space, and allows people to experience their cities. Knowing how people use bike programs gives much more controlled measures of commuting and patterns over time than user owned bikes can provide. It could help justify new bike lanes, send users to the best stations near them, and quantify conditions that people are most likely to bike in. I think the biggest future potential is within sustainable development and urban planning where bikes are equitably available and supported through transportation systems.

Citi Bike



```
{  
    "num_docks_available": 0,  
    "is_renting": 1,  
    "station_id": "4d24ab76-2f4a-4004-9fc5-09c00c1d9aa4",  
    "is_installed": 1,  
    "num_bikes_disabled": 3,  
    "num_bikes_available": 0,  
    "last_reported": 1741011100,  
    "vehicle_types_available": [  
        {  
            "vehicle_type_id": "1",  
            "count": 26  
        },  
        {  
            "vehicle_type_id": "2",  
            "count": 1  
        }  
    ],  
    "num_scooters_unavailable": 0,  
    "num_bikes_available": 27,  
    "num_docks_disabled": 0,  
    "is_returning": 1,  
    "num_ebikes_available": 1  
}.
```

What is the technology or service?

Citi bike, under parent company Lyft is a bike sharing system that allows users to rent bikes from hundreds of stations and then return them to other stations. They are available at all hours and unlocked through an app. The collection of json docs that I looked at for Citi Bike NYC follow the General Bikeshare Feed Specification that many other mobility systems also use. It displays real-time data, although admittedly the last_reported value is not a time value that I know how to interpret so I am unsure of the update frequency.

What behaviors does it capture?

This data feed captures movement of bikes and scooters among citi bike docks in New York city. It shows how available different modes of transport are at different stations, geographic distribution of stations, and if logged the patterns of use. It doesn't provide historical data or personal data.

How does the raw data look like?

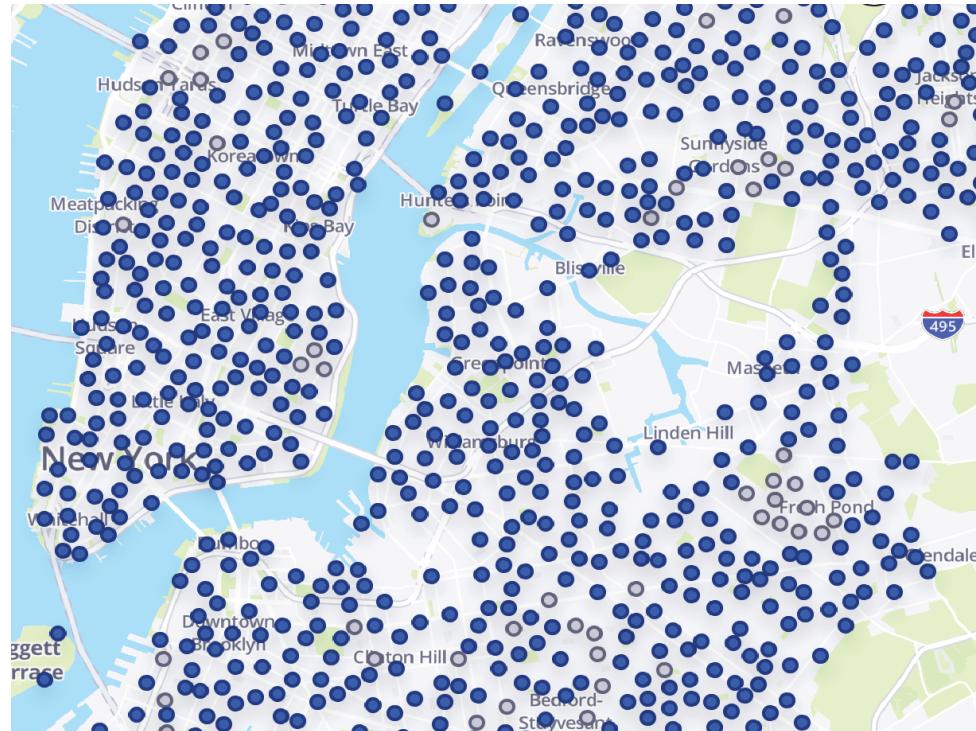
Raw data is split between a few json files. I think I would reference info from station status and station information.

Station status: station_id, count of vehicle types available (bike and ebike), num_bikes_disabled, num_docks_available

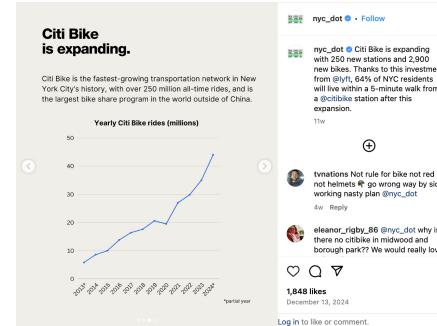
Station information: lat, long, station_id, name, region_id, capacity

Visualizations

Visual Examples of this technology:



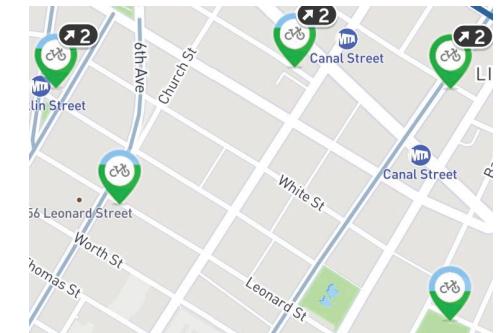
Station map for NYC



Graph on instagram of citi bike expansion



citibike marketing image



Bike angels point map

Insights

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

Station Distribution

Geocoordinates of station locations provide insights to where they are distributed and therefore likely the biking population.

Vehicle type

Electric assist and human powered bikes are both present at many stations. We could learn where electric bikes are most available within the city.

Maintence

Disabled bikes and disabled docks are shown in the dataset so we could see if the most borrowed from station also have the most broken parts.

Total use

The available documents show snapshots of station states not paths of trips, but aggregated, the rise and fall of available bikes could show how many are in use.

Usage

How are the insights currently being used?
By who?

Company

The operations of citi bike depend heavily on display of available vehicles through their app and I'm sure internally as well to track maintenance and needs.

Urban Analysts

Transportation is a huge sector in cities like NYC and knowing real-time usage of different modes helps navigate congestion as well as planning

Sustainability Researchers

Policy proposals and new initiatives for lower environmental impact depend on understanding daily use of current systems and opportunities for growth.

Users

Citi bike users especially those who participate in the bike angels program look at what stations need more or less bikes in order to be compensated most for their redistribution.

In the Future

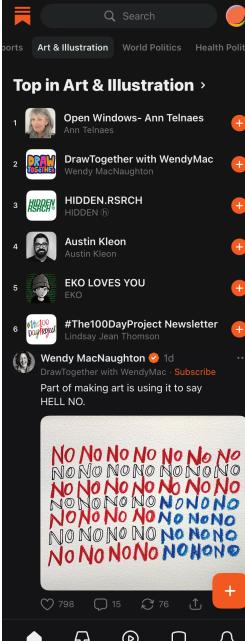
How do you think will the collection and analysis and insights evolve in the next 10 years?

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 opined 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 people flock to niche thinkers it acts as another measure of trends and trust in public opinion.

Substack Api



Newsletters in Art & Illustration (ID: 15417):

- Austin Kleon (austinkleon)
- Open Windows- Ann Telnaes (anntelnaes)
- HIDDEN.RSRCH (hiddenrsrch)
- BoldBrush (sovereignartist)
- PHOTOSNACK (photosnack)
- #The100DayProject Newsletter (the100dayproject)
- BoldBrush Collector (boldbrush)
- DrawTogether with WendyMac (drawtogether)
- Pattern + Process (sarahwattsart)
- EKO LOVES YOU (eko)
- Hello, Wonderful (helloworlderful)
- FlakPhoto Digest (flakphoto)
- Tira do papel (tiradopapel)
- Slowpoke (carsonellis)
- Animation Obsessive (animationobsessive)
- Image Pilgrimage w/ Scott Erickson (scotterickson)
- The Great Women Artists (katyhessel)
- ART DELIVERY from The Jealous Curator (thejealouscurator)
- Ceramics Now Weekly (ceramicssnow)
- Evelyn Henson (evelynhenson)
- Paint Here (jamesgurney)
- Subversive Substack (subversivecrossstitch)
- gamejobs.work's weekly artist jobs (gamejobswork)
- Looking at Picture Books (lookingatpicturebooks)
- Soft Crafts by Ella Emhoff (ellaemhoff)

page 1 done

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
```

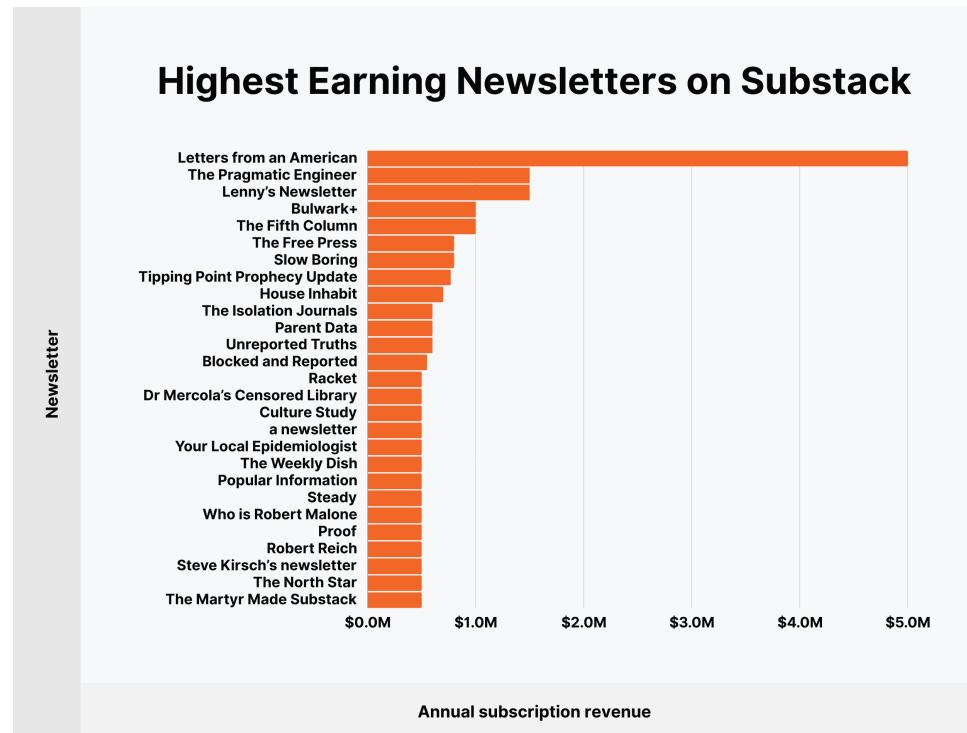
What behaviors does it capture?

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.

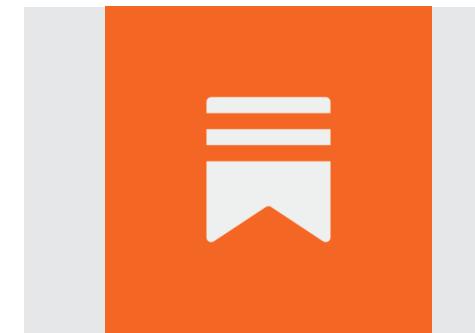
I think it would also be possible to track changes to username and bios.

Visualizations

Visual Examples of this technology:



Graph of highest earning Substack Newsletters (2024) Backlinks



App logo

Recent Posts from Newsletter:
- DHS Now Allows for Surveillance based on Sexual Orientation or Gender Identity (<https://www.erininthemorning.com/p/dhs-now-allows-for-surveillance-based>)
- "The Crime Is Being Trans": Montana House Passes Indecent Exposure Bill Applying To Trans People (<https://www.erininthemorning.com/p/the-crime-is-being-trans-montana>)
- University of Cincinnati Protests Lead to "Biological" Bathroom Signs Reversal (<https://www.erininthemorning.com/p/university-of-cincinnati-protests>)
- Iowa Legislature Votes To Strip Transgender From Civil Rights Protections (<https://www.erininthemorning.com/p/iowa-legislature-votes-to-strip-transgender-from-civil-rights-protections>)
- No, Trump's Trans Military Ban Does Not Have A "waiver" For Openly Trans People Serving (<https://www.erininthemorning.com/p/no-trumps-trans-military-ban-does-not-have-a-waiver-for-openly-trans-people-serving>)
- The New McCarthyism: LGBTQ+ Purges In Government Begin (<https://www.erininthemorning.com/p/the-new-mccarthyism-lgbtq-purges>)
- Marco Rubio May Have Just Banned Trans Foreigners Seeking Visas From US Entry (<https://www.erininthemorning.com/p/marco-rubio-may-have-just-banned>)
- Several Nonprofits Overcomply With Trump EOs, Erase Trans People From Websites (<https://www.erininthemorning.com/p/several-nonprofits-overcomply-with>)
- "We Will Continue To Stonewall": Transphobia Flares (<https://www.erininthemorning.com/p/we-will-continue-to-stonewall>)
- Nazis Burned Trans Books To Usher In Fascism: Now Trump Does The Same (<https://www.erininthemorning.com/p/nazis-burned-trans-books-to-usher>)

API most recent posts from Erin in the Morning

March 1, 2025

John Simpson of the BBC noted recently that "there are years when the world goes through some fundamental, convulsive change." Seven weeks i...

23 HRS AGO · HEATHER COX RICHARDSON

♡ 6,961

○ 938



February 28, 2025

Today, President Donald Trump ambushed Ukraine president Volodymyr Zelensky in an attack that seemed designed to give the White House an...

MAR 1 · HEATHER COX RICHARDSON

♡ 9,473

○ 1,416



February 25, 2025

On Friday, February 21, former transportation secretary Pete Buttigieg posted: "A defining policy battle is about to come to a head in this country."

FEB 26 · HEATHER COX RICHARDSON

♡ 6,024

○ 748



Most recent Letters from an American posts

You Are Doing A Good Enough Job

A newsletter for people who beat themselves up.

7,000+ subscribers

Subscribed

About page in app for You Are Doing A Good Enough Job

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.