

Web Content

CS249i

Modern Websites

Third Party Resources

- Modern websites rely on many different types of *third-party resources* to provide services to keep their websites functional
 - Third party resources are ones served by external parties
 - If you are on cnn.com, any resource served from a domain that is NOT cnn.com (e.g., doubleclick.com, google-analytics.com)
 - These resources could be anything from static images to JavaScript libraries to analytics, advertising, ...

AD

SQUARESPACE

Set up an online store and start selling today.

START YOUR FREE TRIAL



US World Politics Business Opinion Health Entertainment Style Travel Sports Videos

LIVE TV

Edition



PODCAST: Tug of War | TRENDING: World Series | Pope and Biden | Tuesday elections | Halloween train attack | Economic bills | G20 summit | Box office

Trump escalates January 6 cover-up



ANALYSIS

The former President is trying to keep the House select committee probing January 6 from seeing a list of documents as he ramps up his political comeback

KFILE Trump lawyer said 'courage and the spine' would help Pence send election to the House in comments before January 6

Brian Stelter's ominous prediction: Imagine it's 2022 and ...

January 6 committee is losing patience with Trump's former chief of staff Mark Meadows as it seeks his testimony

Washington Post report rebuts the January 6 alt-reality that Tucker Carlson promotes

Biden says US 'continuing to suffer' from Trump's decision to pull out of Iran nuclear deal



LIVE UPDATES

Astros top Braves 9-5 in World Series Game 5

• **Trivia:** Can you name the only player to play in all 3 cities that the Braves have called home?

• **Analysis:** The Braves may win the World Series. But they're striking out with some fans



Students are fed up with raging adults at school board meetings

• A Texas lawmaker is investigating 850 books on race and gender that could cause 'discomfort' to students

• **Opinion:** When parents scream at school board meetings, how can I teach their children?



Southwest launches investigation into pilot reportedly using anti-Biden phrase on flight

Reporter reveals what Lindsey Graham said during January 6 riot

White House press secretary tests positive for Covid, last saw Biden Tuesday

BREAKING Japan's Fumio Kishida defies expectations as ruling party keeps majority

Aurora borealis puts on a gorgeous show

'Step up or step out': Lawmaker calls out attorney general

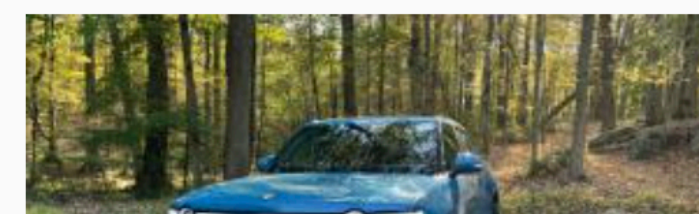
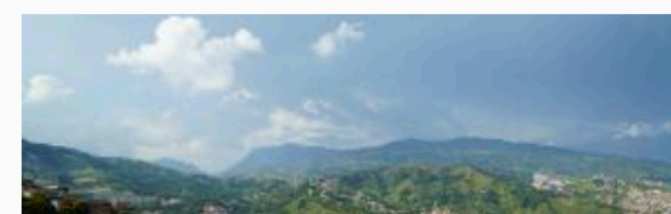
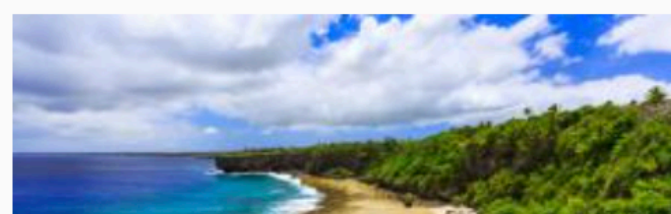
Police investigating desecration of Torah scroll at fraternity

COP26 climate talks off to an ominous start after weak G20 leaders' meeting

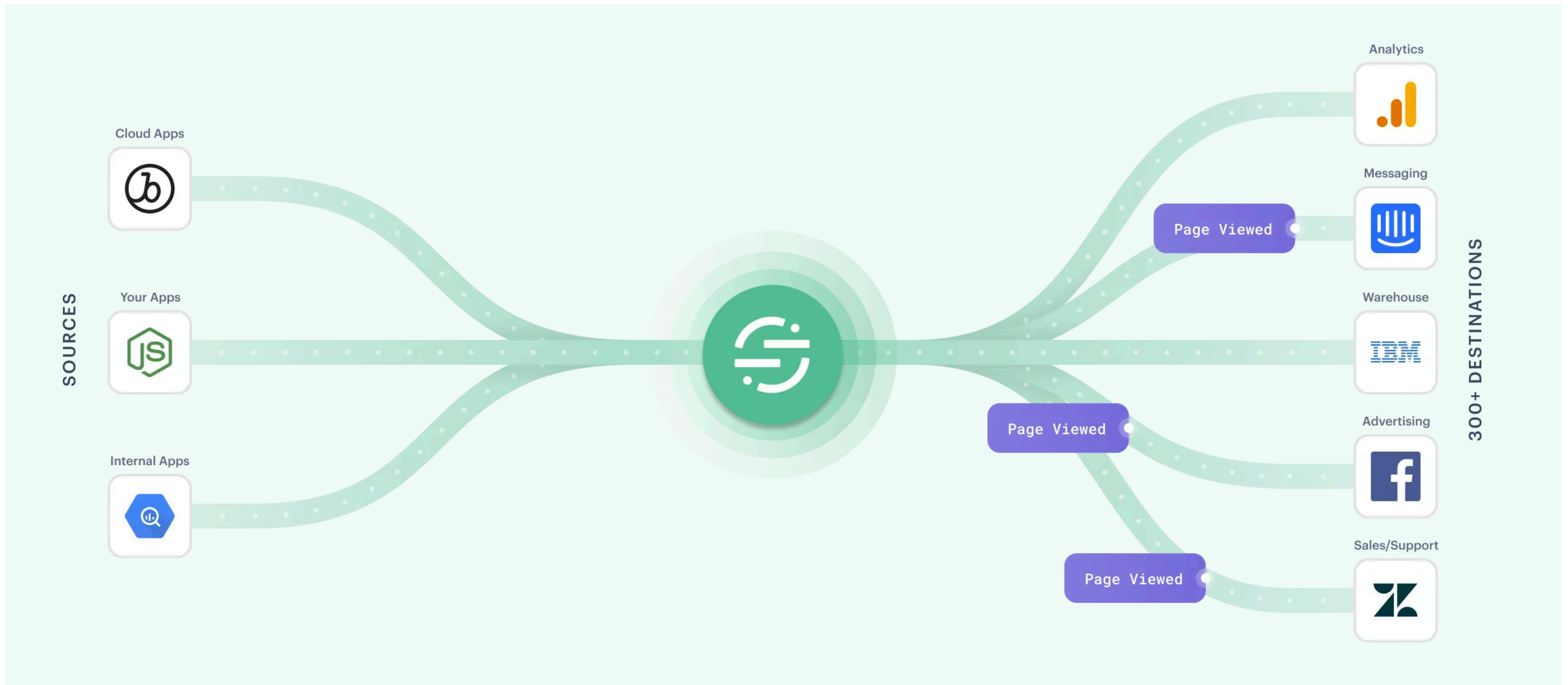
Video shows passengers fleeing knife attack on train

64 Trackers

- AddThis
- Adform
- Adition
- Adobe Audience M...
- Adobe Experience ...
- Aggregate Knowle...
- Amazon Advertising
- AppNexus
- Bidswitch
- Bidtellect
- BlueKai
- Bombora
- Bounce Exchange
- ChartBeat
- Criteo
- Datalogix
- DoubleClick
- Drawbridge
- Eyeota
- Facebook Connect
- FreeWheel
- Google Ads Measu...
- Google Adsense
- Google Dynamic R...
- Google SafeFrame
- Google Tag Manag...
- Index Exchange
- Integral Ad Science
- LiveRamp
- Lotame
- MediaMath
- NetRatings SiteCe...
- OneTag
- OpenX
- Optimizely
- Outbrain
- Outbrain Amplify
- PowerLinks
- PubMatic
- Quantcast
- RTB House
- Rubicon
- Salesforce DMP
- ScoreCard Researc...
- Simpli.fi
- Smaato
- SOASTA mPulse
- SpotX
- Tapad
- TradeDesk

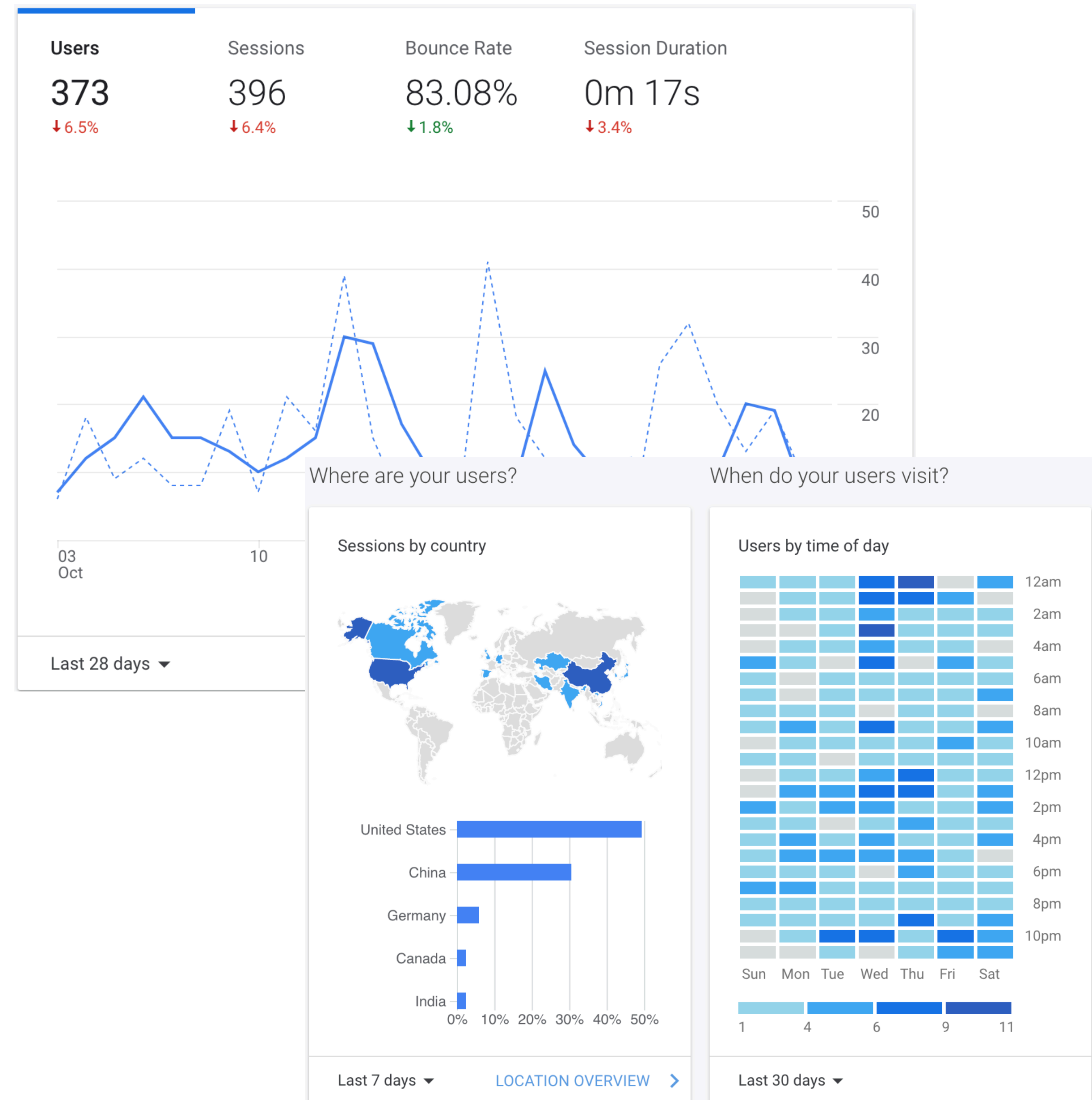


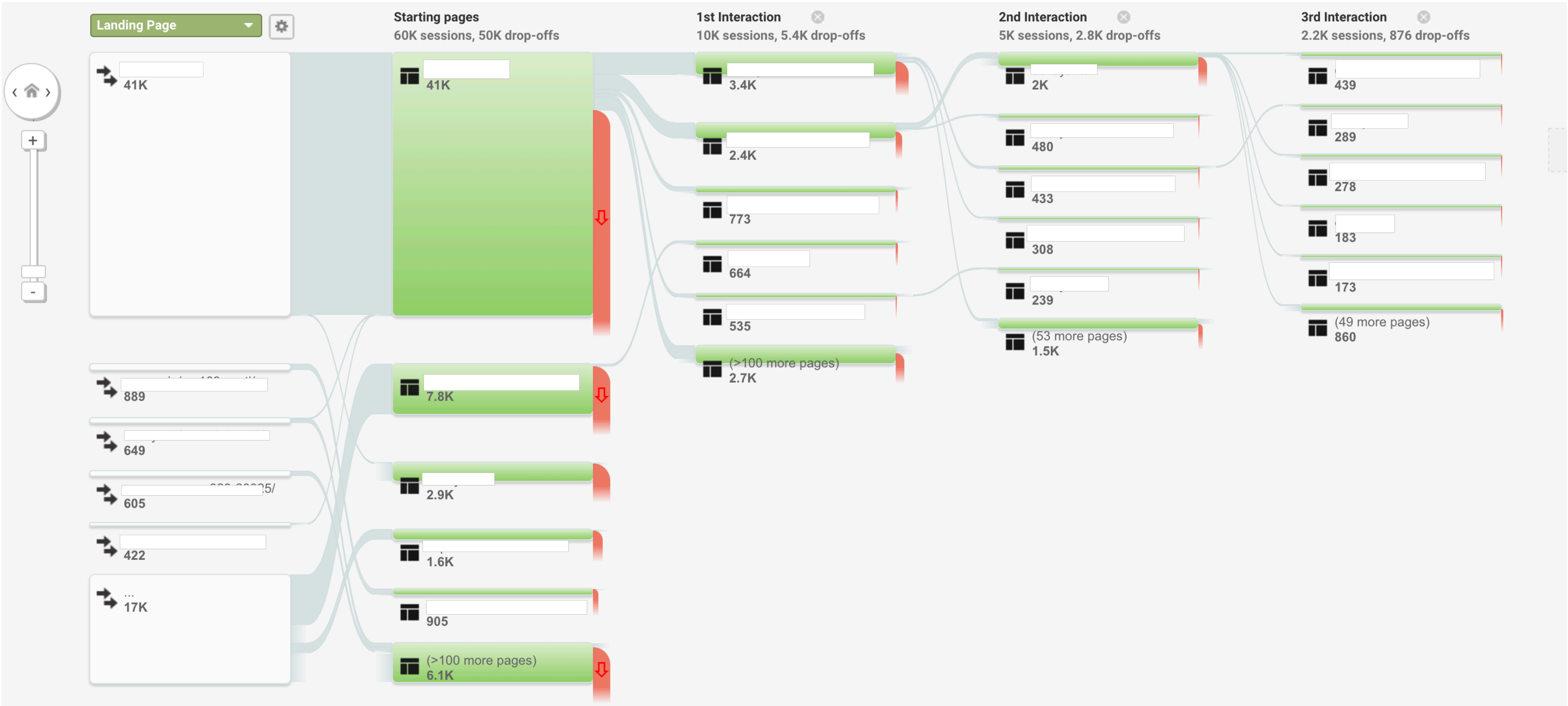
"Twilio's \$3.2B Segment acquisition is about helping developers build data-fueled apps"



Analytics

- Many websites rely on user analytics to improve their services
 - For example, Google Analytics, appears on an estimated 70% of the top websites
- As an analytics user, you can see where your clients are connecting from, you can see how long they spent on the page, what devices they're connecting from, and a ton of other interesting details
 - These are typically scoped to a single request, but in recent years, companies have been expanding the scope of what they know about users...

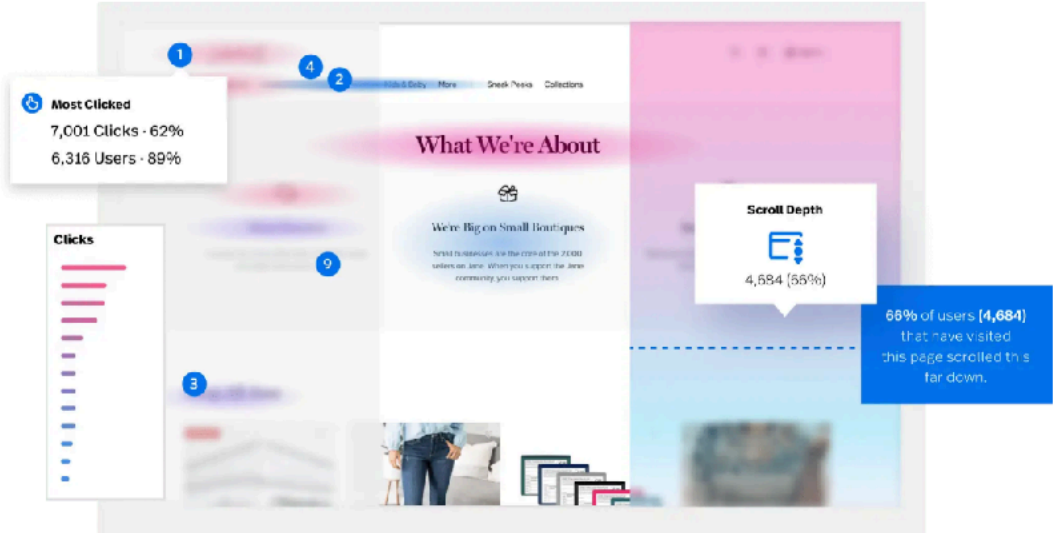




In App Analytics (From FullStory)

Draw insights from data you can trust

Click maps and scroll maps give a crystal-clear picture of user engagement without the confounding factors of screen resolution, dynamic page content, pop-ups, or confusing “splotches.”



Inspect specific user segments

Answer questions like “What do visitors from my ad campaign click on first?”, “How do returning customers interact with our navigation?”, or “Are users clicking on the latest product update?”

Understand the user experience behind every data point and browser tab

Session Replay lets you see exactly what individual users see and do across any number of tabs. Feel their pain so you can cure it, fast.

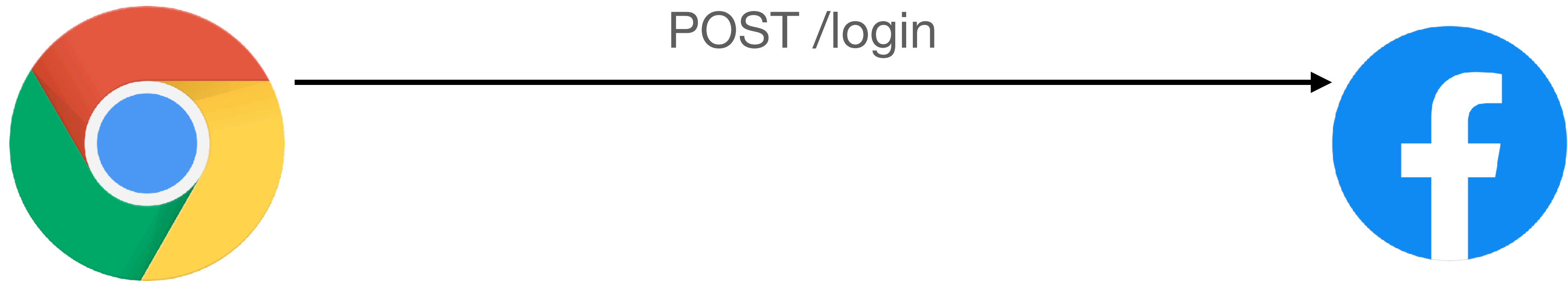
Complete data for a complete understanding

Easy implementation and intuitive UI empowers your whole team to jump in right away. **Tagless autocapture** means you never have to worry about data points slipping through the cracks.

Web Tracking

Cookies and Code

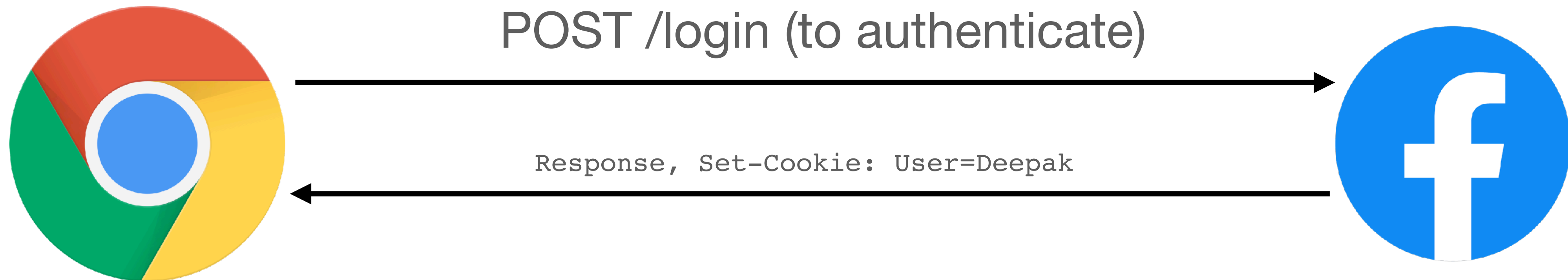
- Major companies typically use *cookies* to offer extended functionality for websites (e.g., keeping you logged in, keeping certain settings stored in your browser, etc.)



Web Tracking

Cookies and Code

- Major companies typically use *cookies* to offer extended functionality for websites (e.g., keeping you logged in, keeping certain settings stored in your browser, etc.)



Web Tracking

Cookies and Code

- Major companies typically use *cookies* to offer extended functionality for websites (e.g., keeping you logged in, keeping certain settings stored in your browser, etc.)
- Once a cookie is set, the browser attaches a cookie to every subsequent request sent out for that particular domain
 - Cookies are by default scoped to the first-party domain that set the cookie
 - No other domains can read the cookie value!
- ...then how does web tracking work?

Web Tracking Cookies and Code

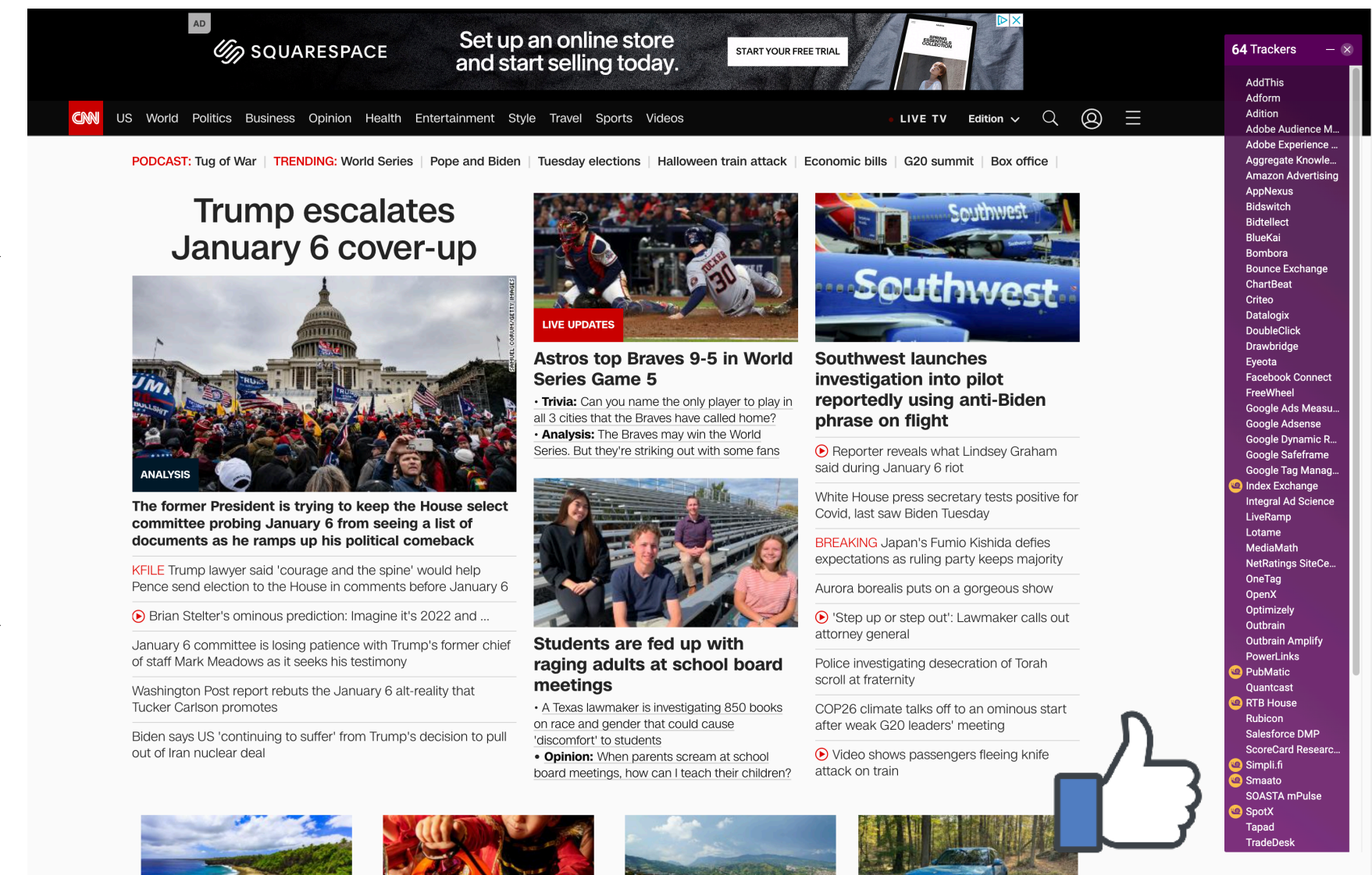
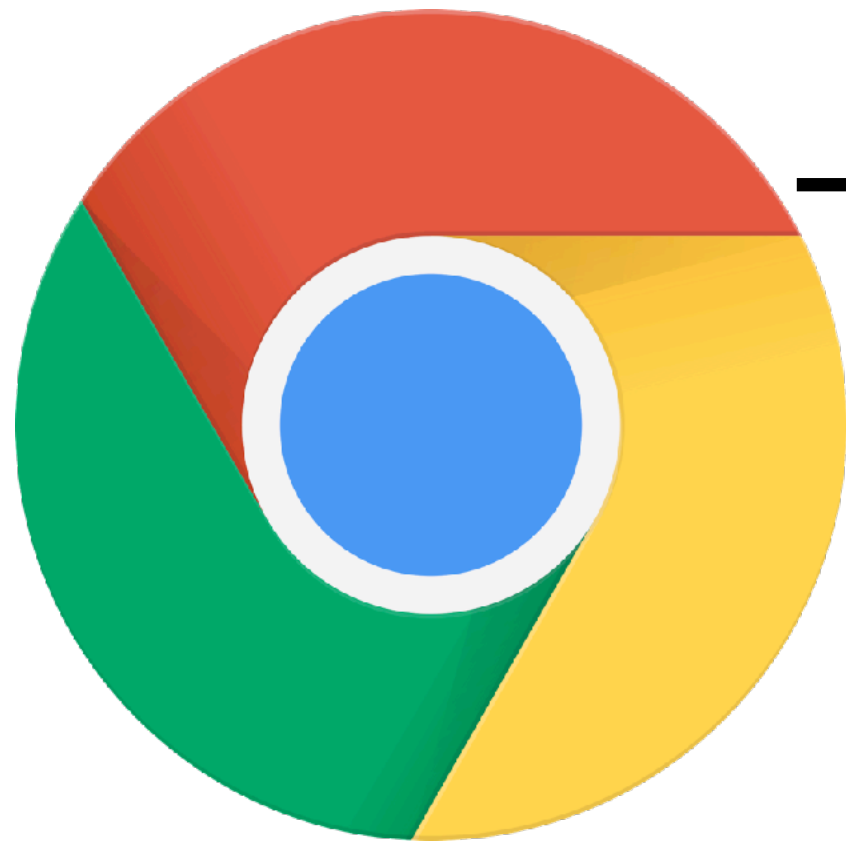
(CNN)

GET / HTTP/3

GET /facebook-like.js HTTP/3

Cookie: User=Deepak, Referer=cnn.com

(Facebook)

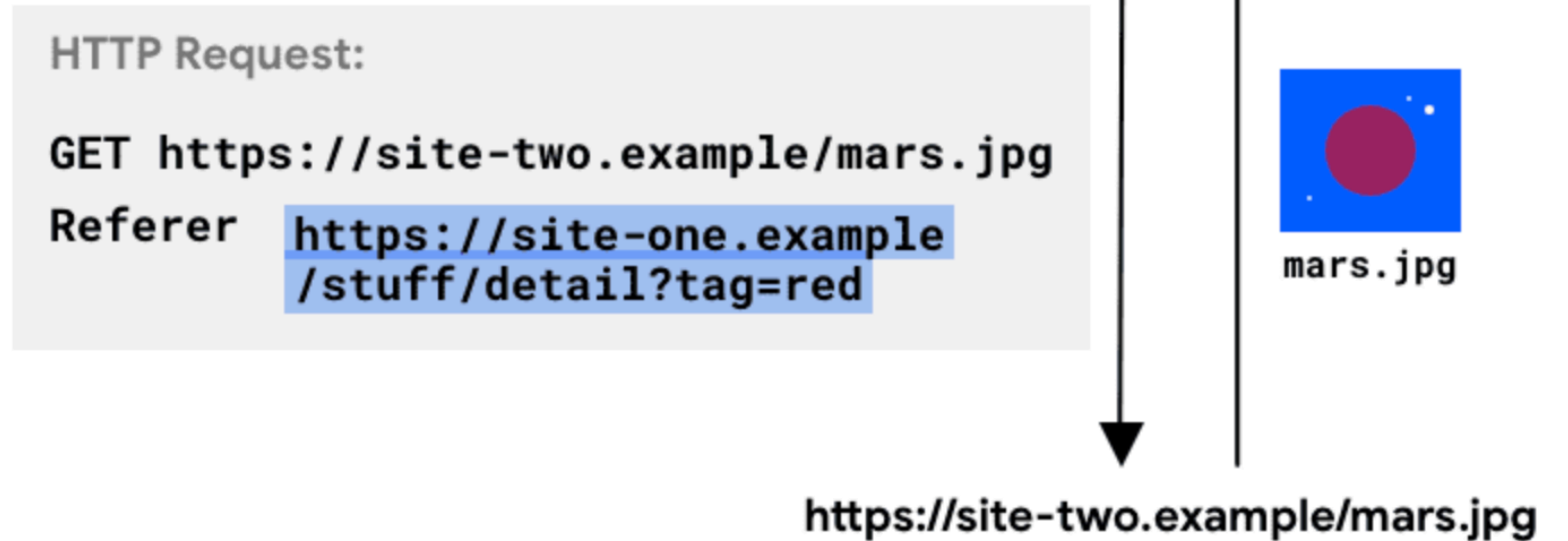


- With this request, companies can link your cookie to your browsing data (e.g., through Referer header, Host headers, Origin, or just JavaScript)

Web Tracking

Cookies and Code

- What exactly is sent in the referer?



Referer sent in an request, depending on the Referrer-Policy and the request

	Referer: No data	Referer: Origin only	Referer: Full URL	
Policies that don't take the context of the request into account.	Policy: no-referrer	✓		
	origin		✓	
	unsafe-url			✓
Policies focused on whether the request is made to a destination that is less secure than where the request came from. These help reach security goals.	strict-origin	request from HTTPS to HTTP	request from HTTPS to HTTPS or HTTP to HTTP	
	no-referrer-when-downgrade	request from HTTPS to HTTP		request from HTTPS to HTTPS or HTTP to HTTP
Policies focused on whether the request is cross- or same-origin. These help reach privacy goals.	origin-when-cross-origin		cross-origin request	same-origin request
	same-origin	cross-origin request		same-origin request
Policy focused on the request security AND whether it is cross- or same-origin. It helps reach privacy and security goals.	strict-origin when-cross-origin	request from HTTPS to HTTP	cross-origin request, from HTTPS to HTTPS or HTTP to HTTP	same-origin request

Browser Default

Web Tracking

Browser Fingerprinting

- Websites can also fingerprint you effectively with *browser fingerprinting*, which is a technique that leverages all your settings to identify you, and stores this in a cookie on your browser
 - <https://amiunique.org/>
- So long as JavaScript can run (by third-parties), you run the risk of being “followed” on the web

```
{
  "user-agent": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:93.0) Gecko/20100101 Firefox/93.0",
  "accept": "text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,*/*;q=0.8",
  "accept-encoding": "gzip, deflate, br",
  "accept-language": "en-US,en;q=0.5",
  "upgrade-insecure-requests": "1",
  "referrer": "https://amiunique.org/",
  "userAgent-js": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:93.0) Gecko/20100101 Firefox/93.0",
  "platform": "MacIntel",
  "cookies": "yes",
  "timezone": 420,
  "languages-js": "en-US,en",
  "ad": "no",
  "doNotTrack": "NC",
  "navigator_properties": [
    "vibrate",
    "javaEnabled",
    "getGamepads",
    "getVRDisplays",
    "mozGetUserMedia",
    "sendBeacon",
    "requestMediaKeySystemAccess",
    "registerProtocolHandler",
    "taintEnabled",
    "permissions"
  ]
}
```


MY BROWSER FINGERPRINT

SEE YOUR BROWSER FINGERPRINT PROPERTIES

ARE YOU UNIQUE ?

DOWNLOAD

TIMELINE

TODAY

7 DAYS

15 DAYS

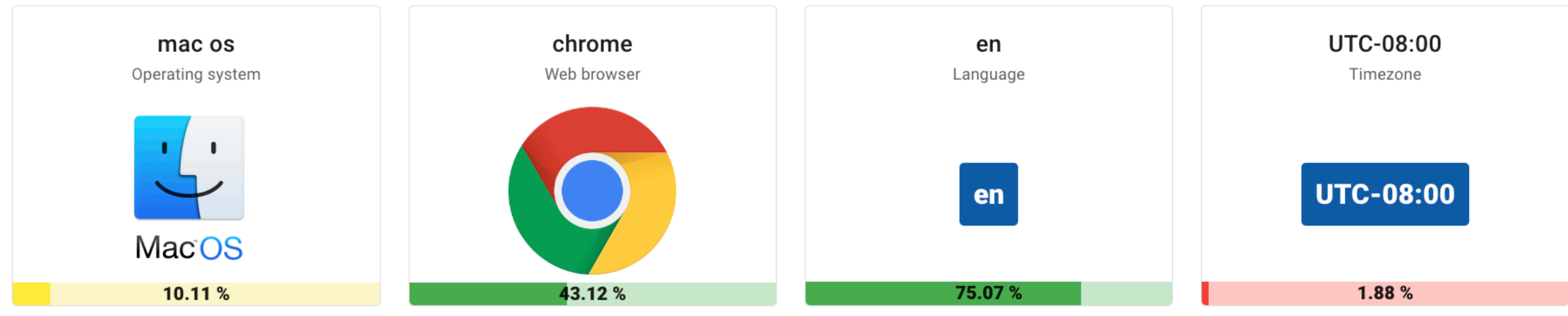
30 DAYS

90 DAYS

ALL TIME

Yes! You are unique among the 2382170 fingerprints in our entire dataset.

The following informations reveal your OS, browser, browser version as well as your timezone and preferred language. Moreover, we show the proportion of users sharing the same elements.



HTTP HEADERS ATTRIBUTES

Search for an attribute

Attribute	Similarity ratio	Value
1 - User agent	0.10 %	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/121.0.0.0 Safari/537.36
2 - Accept	12.02 %	text/html,application/xhtml+xml,application/xml;q=0.9,image/avif,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3;q=0.7
3 - Content encoding	96.52 %	gzip, deflate, br
4 - Content language	19.94 %	en-US,en;q=0.9
5 - Upgrade Insecure Requests	91.00 %	1

JAVASCRIPT ATTRIBUTES

Search for an attribute

Attribute	Similarity ratio	Value
1 - User agent	0.09 %	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/121.0.0.0 Safari/537.36
2 - Platform	10.01 %	MacIntel
3 - Cookies enabled	89.14 %	✓
4 - Timezone	1.88 %	UTC-08:00

20 - Screen width	4.25 %	2560	pdf-viewer.
21 - Screen height	4.58 %	1440	
22 - Screen depth	3.64 %	30	
23 - Screen available top	3.32 %	25	
24 - Screen available Left	83.26 %	0	
25 - Screen available Height	0.01 %	1346	
26 - Screen available width	4.07 %	2560	
27 - Permissions	6.24 %		accelerometer : granted accessibility : Not supported ambient-light-sensor : Not supported camera : prompt clipboard-read : prompt clipboard-write : granted geolocation : prompt background-sync : granted magnetometer : granted microphone : prompt midi : granted notifications : prompt payment-handler : granted persistent-storage : prompt push : Not supported
28 - WebGL Vendor	1.83 %	Google Inc. (Apple)	
29 - WebGL Renderer	0.01 %	ANGLE (Apple, ANGLE Metal Renderer: Apple M2 Max, Unspecified Version)	
30 - WebGL Data	0.13 %		
31 - WebGL Parameters	0.03 %		35 different extensions 25 different general parameters analyzed 36 different shaders precisions analyzed

Web Tracking

Prevalence of Major Companies

- Major companies have large presences on the web, and as a result, can see the majority of websites that you visit
 - Google appears on 82.2% of the Top 1M (by AS), because of analytics and advertising services
 - Facebook appears on 34.1%, to enable social sharing + tracking

Company	Prevalence on Top 1M
Google	82.2%
Facebook	34.1%
Amazon	32.6%
Cloudflare	30.7%
Akamai	20.3%
MaxCDN	19.0%
Edgecast	17.9%
Fastly	15.5%
SoftLayer	11.8%
Twitter	11.2%

Web Tracking

Cookie Syncing

- Even if a company is not available on every website, companies often times *share* cookie information
 - “Cookie Synchronization: Everything You Always Wanted to know but were afraid to ask” – WebConf 2019
- Core idea is simple: If you have a collaboration agreement with another third-party, you *redirect* requests to them upon receiving requests

Web Tracking

Cookie Syncing

The image shows a screenshot of a news website with a purple cookie tracker overlay on the right side. The website content includes a main article titled "Trump escalates January 6 cover-up" with a sub-headline "The former President is trying to keep the House select committee probing January 6 from seeing a list of documents as he ramps up his political comeback". Other articles include "Astros top Braves 9-5 in World Series Game 5", "Southwest launches investigation into pilot reportedly using anti-Biden phrase on flight", and "Students are fed up with raging adults at school board meetings". The purple overlay lists 64 trackers, including AddThis, Adform, Adition, Adobe Audience M..., Adobe Experience..., Aggregate Knowle..., Amazon Advertising, AppNexus, Bidswitch, Bidstreet, BlueKai, Bombora, Bounce Exchange, ChartBeat, Cleo, Datalogix, DoubleClick, Drawbridge, Eyeota, Facebook Connect, FreeWheel, Google Ads Meas..., Google AdSense, Google Dynamic R..., Google Safeframe, Google Tag Manag..., Index Exchange, Integral Ad Science, LiveRamp, Lotame, MediaMath, NetRatings SiteCe..., OneTag, OpenX, Optimizely, Outbrain, Outbrain Amplify, PowerLinks, PubMatic, Quantcast, RTB House, Rubicon, Salesforce DMP, ScoreCard Researc..., Simpli.fi, Smaato, SPASTA mPulse, SpotX, Tapad, and TradeDesk.

GET tracker.com/pixel.jpg

Response, Set-Cookie: User=user123



Web Tracking

Cookie Syncing

GET advertiser.com/pixel.jpg

Response, Set-Cookie: User=userABC

advertiser.com

The screenshot shows a news website interface. At the top, there's a navigation bar with categories like US, World, Politics, Business, Opinion, Health, Entertainment, Style, Travel, Sports, and Videos. Below this, there are several news articles with headlines such as "Trump escalates January 6 cover-up", "Astros top Braves 9-5 in World Series Game 5", and "Southwest launches investigation into pilot reportedly using anti-Biden phrase on flight". A sidebar on the right lists 64 trackers, including AddThis, Adform, Adition, Adobe Audience M..., Adobe Experience..., Aggregate Knowle..., Amazon Advertising, AppNexus, Bidswitch, Bidstreet, BlueKai, Bombora, Bounce Exchange, ChartBeat, Criteo, Datalogix, DoubleClick, Drawbridge, Eyeota, Facebook Connect, FreeWheel, Google Ads Measu..., Google AdSense, Google Dynamic R..., Google Safeframe, Google Tag Manag..., Index Exchange, Integral Ad Science, LiveRamp, Lotame, MediaMath, NetRatings SiteCe..., OneTag, OpenX, Optimizely, Outbrain, Outbrain Amplify, PowerLinks, PubMatic, Quantcast, RTB House, Rubicon, Salesforce DMP, ScoreCard Researc..., Simpli.fi, Smaato, SPASTA mPulse, SpotX, Tapad, and TradeDesk.

Web Tracking Cookie Syncing

Monday, November 1, 2021
Today's Paper

U.S. INTERNATIONAL CANADA ESPAÑOL 中文

PLAY THE CROSSWORD Account

59°F 68° 59"
Nasdaq +0.2% ↑

World U.S. Politics N.Y. Business Opinion Tech Science Health Sports Arts Books Style Food Travel Magazine T Magazine Real Estate Video

LIVE
Climate Change Is 'Ravaging the World,' Biden Tells Summit

António Guterres, the U.N. secretary general, opened the conference with a blistering critique of the world's failure to unite to address global warming.

"We are standing at an inflection point in world history," President Biden said in a speech, calling the need for action a moral imperative. Here's the latest.

China

12 GtCO₂

Historical emissions

1990 2030

Current

Pledged

1.5°C compatible

How much are countries pledging to reduce emissions?

President Biden will try to assure skeptics that the U.S. is serious about climate action.

Europe is worried that the costs of climate action could set off a populist backlash.

Once a leading polluter, the U.K. is now trying to lead on climate change.

LIVE
Supreme Court Is Hearing Oral Arguments on Texas Abortion Law

The question for the justices is whether abortion providers and the Biden administration are entitled to challenge the law. Listen and follow our analysis.

The Supreme Court has revised its procedures in a bid to tame interruptions during oral arguments.

Opinion
KATHRYN KOLBERT AND JULIE F. KAY
Roe Is as Good as Gone. It's Time for a New Strategy.

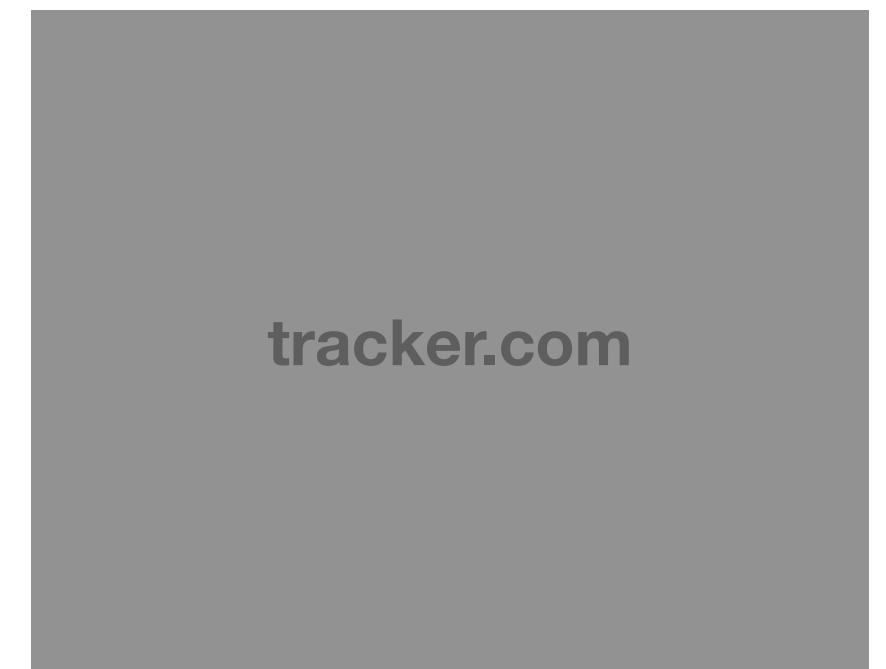
MARGARET RENKL
I Just Turned 60, but I Still Feel 22

Global Virus Death Toll Passes 5 Million

Experts say that the official toll is an undercount, as many

Jen Psaki, the White House press secretary, tested positive for the coronavirus.

GET tracker.com/pixel.jpg, [cookie=user123](https://www.cookie=userid)



Web Tracking

Cookie Syncing

The screenshot shows the New York Times homepage as of Monday, November 1, 2021. The main headline is "Climate Change Is 'Ravaging the World,' Biden Tells Summit". A line chart titled "How much are countries pledging to reduce emissions?" shows historical emissions from 1990 to 2030, with a specific callout for China showing a 1.5°C compatible path. Other articles include "The TV Hit That Wasn't", "Supreme Court Is Hearing Oral Arguments on Texas Abortion Law", and "Global Virus Death Toll Passes 5 Million".

Climate Change Is 'Ravaging the World,' Biden Tells Summit

António Guterres, the U.N. secretary general, opened the conference with a blistering critique of the world's failure to unite to address global warming.

"We are standing at an inflection point in world history," President Biden said in a speech, calling the need for action a moral imperative. Here's the latest.

China

Historical emissions

12 GtCO₂

1.5°C compatible

Current

Pledged

1990

2030

How much are countries pledging to reduce emissions?

The TV Hit That Wasn't

There was hype for the FX series "Impeachment: American Crime Story." But it won't be available on any major streaming platform for another 10 months.

Supreme Court Is Hearing Oral Arguments on Texas Abortion Law

The question for the justices is whether abortion providers and the Biden administration are entitled to challenge the law. Listen and follow our analysis.

Global Virus Death Toll Passes 5 Million

Experts say that the official toll is an undercount, as many

Jen Psaki, the White House press secretary, tested positive for the coronavirus.

GET tracker.com/pixel.jpg, cookie=user123

REDIRECT, advertiser.com?syncID=user123&publisher=nytimes.com

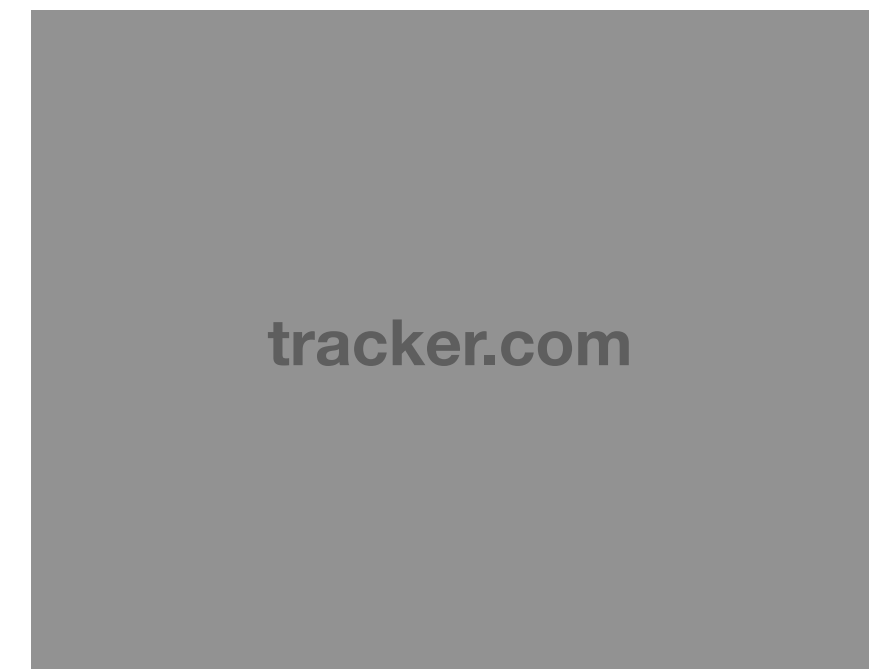
tracker.com

advertiser.com

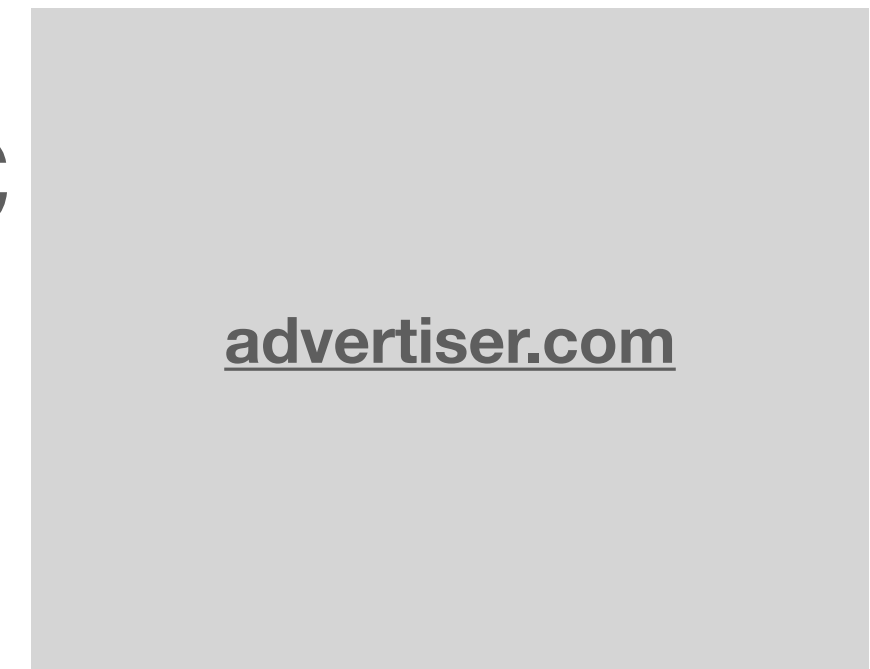
Web Tracking Cookie Syncing



GET tracker.com/pixel.jpg, [cookie=user123](https://www.cookie=userid)



REDIRECT, advertiser.com?syncID=user123&publisher=nytimes.com



GET [syncID=user123](https://www.syncID=userid), [cookie=userABC](https://www.cookie=userid)

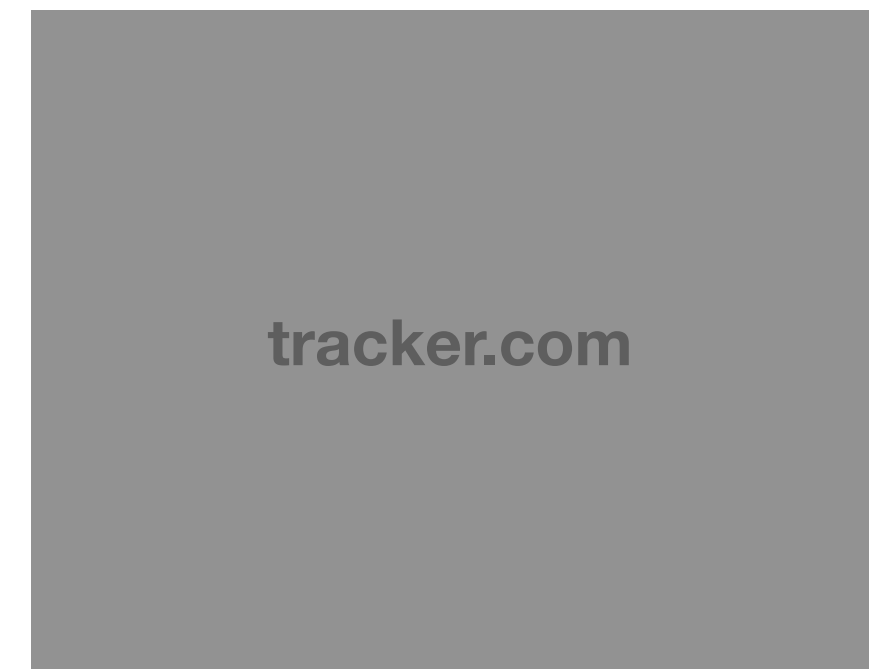


Web Tracking

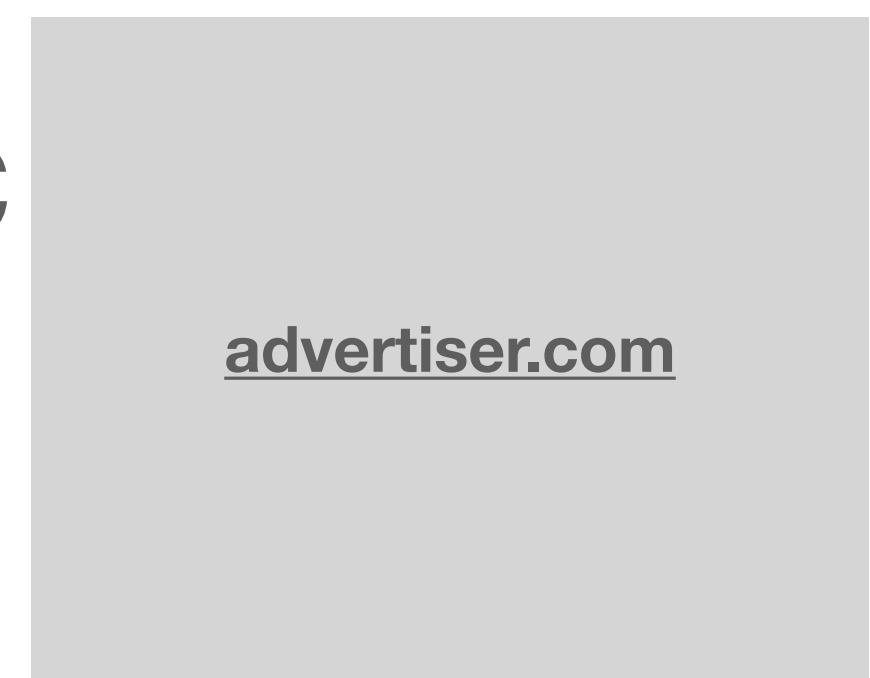
Cookie Syncing



GET tracker.com/pixel.jpg, cookie=user123



REDIRECT, advertiser.com?syncID=user123&publisher=nytimes.com



GET syncID=user123, cookie=userABC



- Third-parties with cookie syncing are enabled on 78% of modern websites

Web Tracking

Cookie Ghostwriting

- Not all first-party cookies *should* be treated the same!

Web Tracking Cookie Ghostwriting

- Not all first-party cookies *should* be treated the same!

Monday, November 1, 2021
Today's Paper

U.S. INTERNATIONAL CANADA ESPAÑOL 中文

PLAY THE CROSSWORD Account

59°F 68° 59°
Nasdaq +0.2% ↑

World U.S. Politics N.Y. Business Opinion Tech Science Health Sports Arts Books Style Food Travel Magazine T Magazine Real Estate Video

LIVE
Climate Change Is 'Ravaging the World,' Biden Tells Summit

António Guterres, the U.N. secretary general, opened the conference with a blistering critique of the world's failure to unite to address global warming.

"We are standing at an inflection point in world history," President Biden said in a speech, calling the need for action a moral imperative. Here's the latest.

China

12 GtCO₂

Historical emissions

1990 2030

Current

Pledged

1.5°C compatible

How much are countries pledging to reduce emissions?

President Biden will try to assure skeptics that the U.S. is serious about climate action.

Europe is worried that the costs of climate action could set off a populist backlash.

Once a leading polluter, the U.K. is now trying to lead on climate change.

The TV Hit That Wasn't

There was hype for the FX series "Impeachment: American Crime Story." But it won't be available on any major streaming platform for another 10 months.

He ran in the first New York City Marathon. Next week, he'll run in the 50th.

A pair of N.F.L. teams made big impressions in Week 8. Here's what we learned from Sunday's games.

Opinion

KATHRYN KOLBERT AND JULIE F. KAY

Roe Is as Good as Gone. It's Time for a New Strategy.

MARGARET RENKL

I Just Turned 60, but I Still Feel 22

LIVE
Supreme Court Is Hearing Oral Arguments on Texas Abortion Law

The question for the justices is whether abortion providers and the Biden administration are entitled to challenge the law. Listen and follow our analysis.

LIVE
Global Virus Death Toll Passes 5 Million

Experts say that the official toll is an undercount, as many

[imate/paris-pledges-tracker-cop-26.html](#) Covid-19 deaths accurately.

Jen Psaki, the White House press secretary, tested positive for the coronavirus.

GET tracker.com/script.js



Web Tracking Cookie Ghostwriting

- Not all first-party cookies *should* be treated the same!

The screenshot shows the New York Times homepage with several articles and a chart. The chart, titled "How much are countries pledging to reduce emissions?", shows historical emissions from 1990 to 2030. It highlights China's current emissions and a 1.5°C compatible path. Other articles include "Climate Change Is 'Ravaging the World,' Biden Tells Summit", "The TV Hit That Wasn't", "Supreme Court Is Hearing Oral Arguments on Texas Abortion Law", "Roe Is as Good as Gone. It's Time for a New Strategy.", and "Global Virus Death Toll Passes 5 Million".

GET tracker.com/script.js



script.js

```
document.cookie = "user=userABC"
```

Web Tracking Cookie Ghostwriting

- 42% of identifier cookies are *ghostwritten* in modern websites

The screenshot shows the New York Times homepage with several articles and a chart. The chart, titled "How much are countries pledging to reduce emissions?", shows historical emissions from 1990 to 2030. It highlights China's current emissions and a pledged path that is 1.5°C compatible. Other articles include "Climate Change Is 'Ravaging the World,' Biden Tells Summit", "The TV Hit That Wasn't", "Supreme Court Is Hearing Oral Arguments on Texas Abortion Law", "Roe Is as Good as Gone. It's Time for a New Strategy.", and "Global Virus Death Toll Passes 5 Million".

GET `tracker.com/script.js`

tracker.com

script.js

`document.cookie = "user=userABC"`

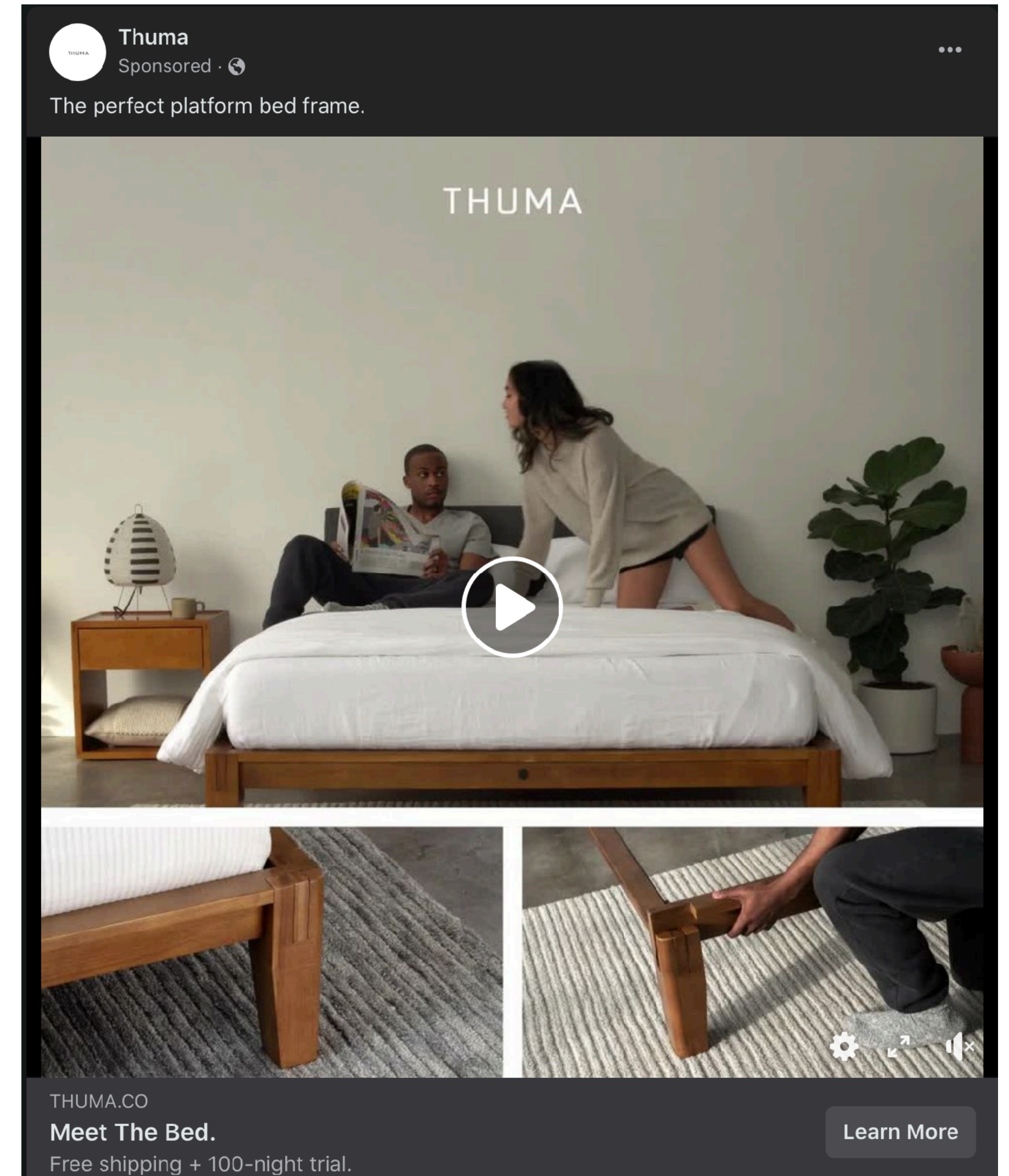
advertiser.com

Online Advertising Ecosystem

Online Advertising

**The Best Thing Since Sliced Bread!
Available for \$4.99 at your local Costco.**

- Companies typically track you around the web to build profiles for *targeted advertising*
- The more targeted your advertising, the more revenue you can make from advertisers who are potentially willing to give you more money to sell the ad spot
- Useful for advertisers to know if people with your browsing habits, your properties, your whatever are browsing on the web



Thuma
Sponsored · 🌐

The perfect platform bed frame.

THUMA

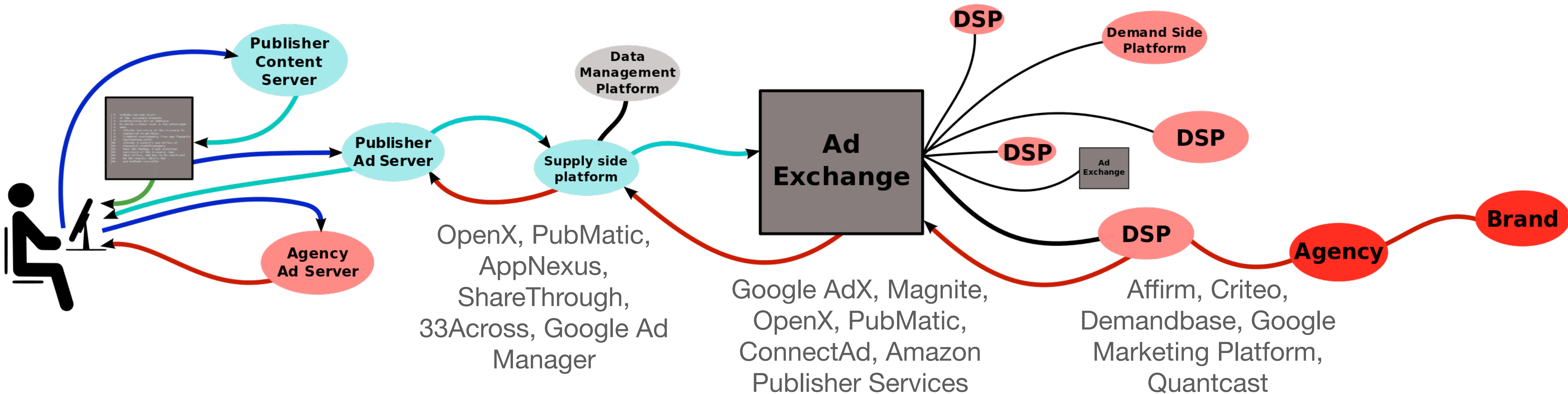
THUMA.CO
Meet The Bed.
Free shipping + 100-night trial.

Learn More

The advertisement is a video player interface. At the top left, it shows the Thuma logo and 'Sponsored' status. Below that is the headline 'The perfect platform bed frame.' The main video area shows a man sitting on a bed reading a magazine while a woman leans over the side of the bed. The word 'THUMA' is overlaid in the top right of the video. Below the main video are two smaller inset images: one showing a close-up of the bed's wooden frame and another showing a person's hand touching the frame. At the bottom, there is a dark grey bar with the website 'THUMA.CO', the product name 'Meet The Bed.', a promotional offer 'Free shipping + 100-night trial.', and a 'Learn More' button.

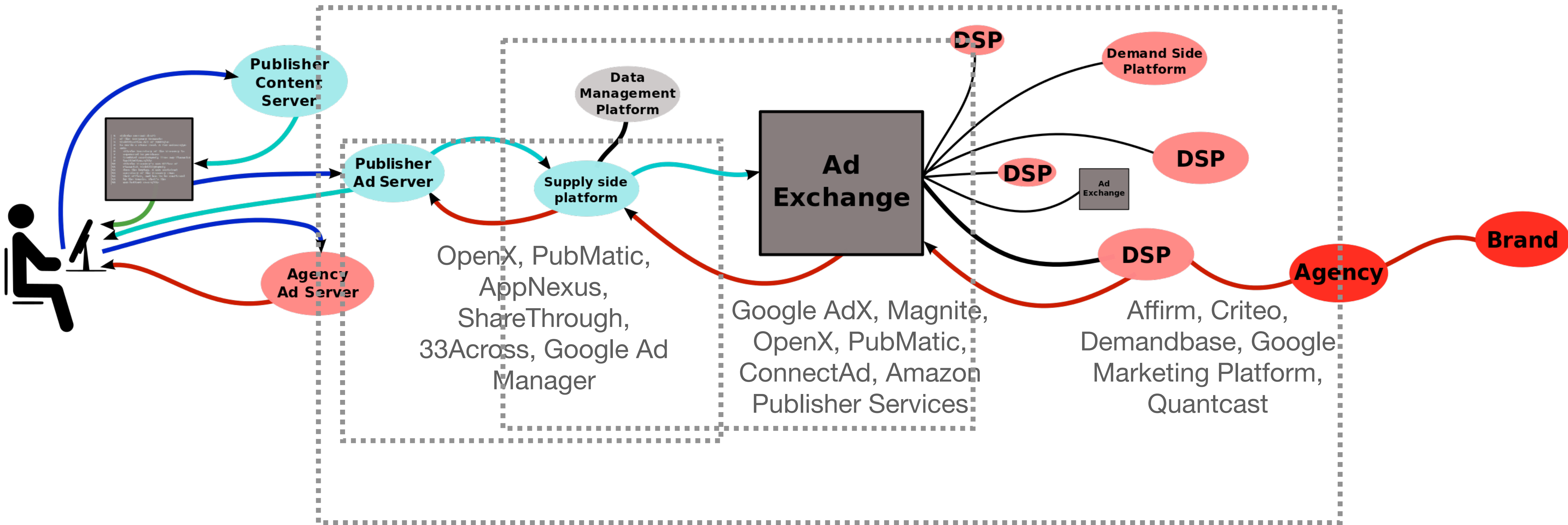
Online Advertising

The Many Internet Players in Advertising



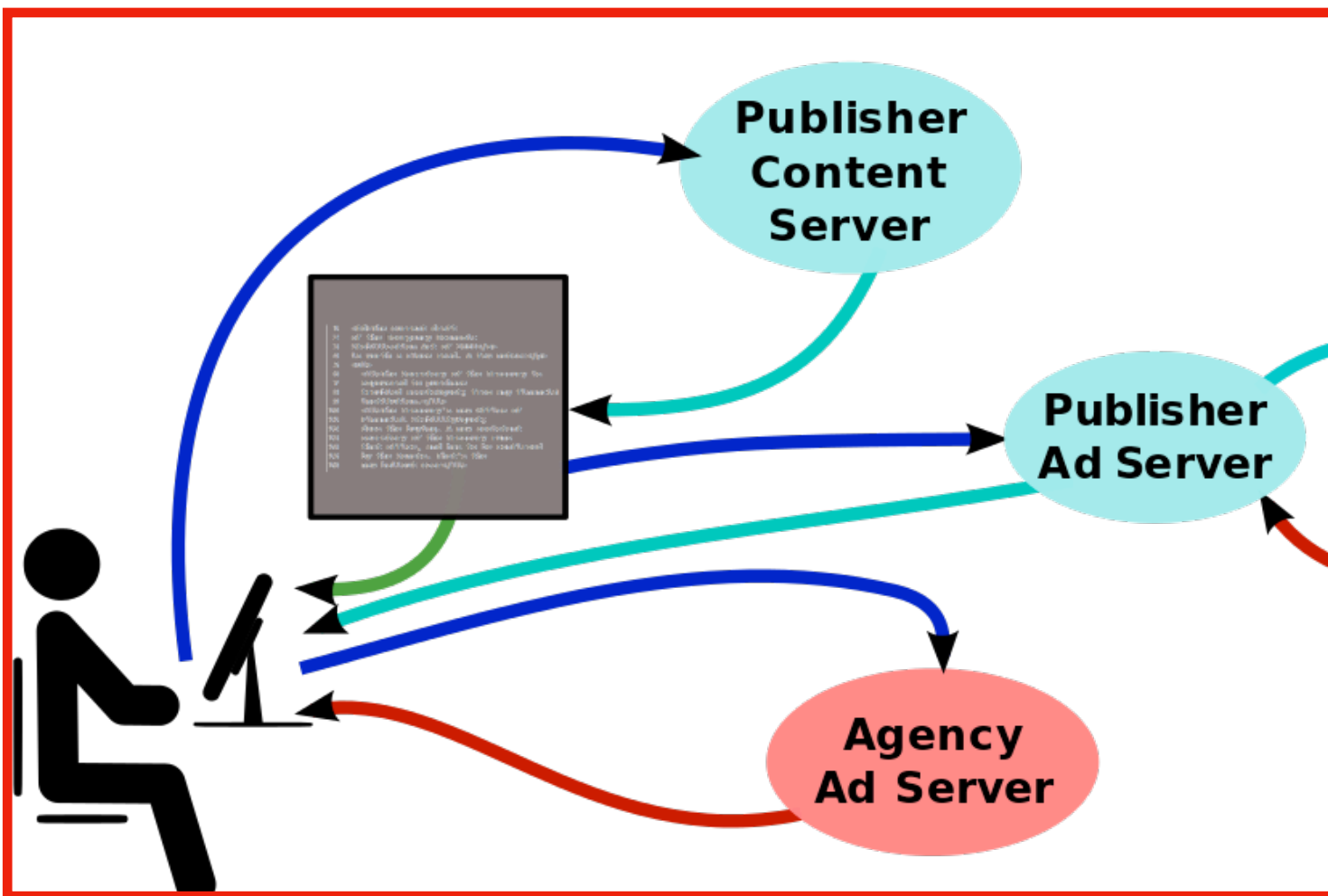
Online Advertising

The Many Internet Players in Advertising



Online Advertising

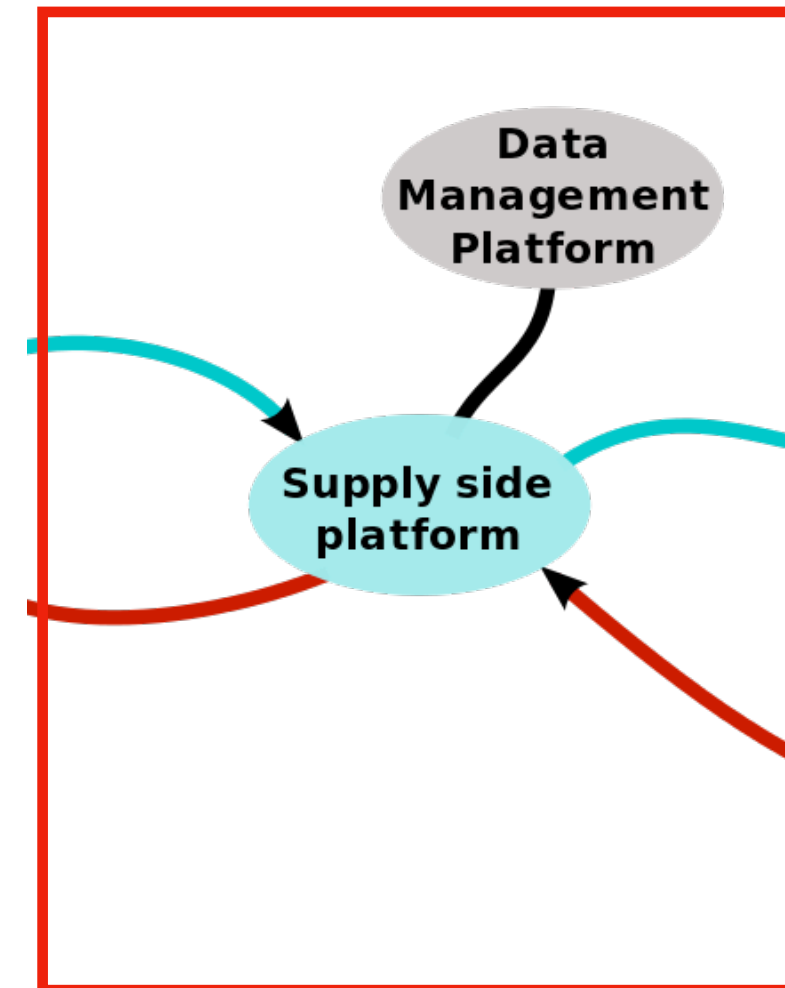
Publishers



- Publishers (e.g., nytimes.com, cnn.com, other websites) often have advertising space that they are hoping to make revenue off of
- In some cases, publishers have explicit agreements with specific companies and can sell their space directly that way too

Online Advertising

Supply Side Platforms



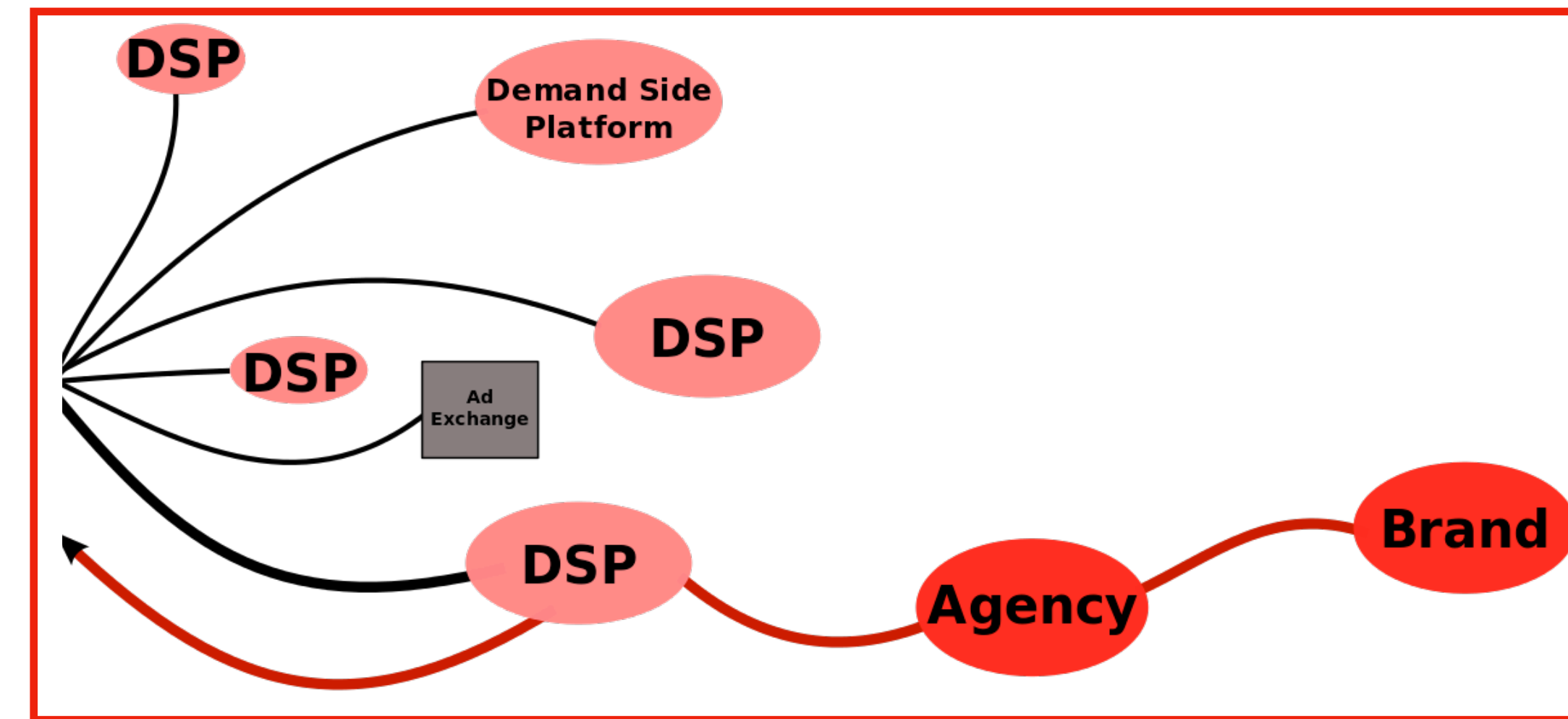
- **Multiple Types of Data in DMP:**

- "First Party", e.g., CRM Data. This can include data from customer behaviors, actions, purchases or interest
- "Second party" — statistics related to cookie pools on external publications and platforms. The
- "Third party" — sourced from external providers and aggregated across websites. Businesses sell third-party data

- If a publisher wants to place the ad spot on the open advertising market, they typically go through an intermediary called a Supply Side Platform (SSP)
 - Examples: Pubmatic, Rubicon Project, Verizon Media, etc.
- This aggregates information about the client (through a DMP) and participates in ad exchange

Online Advertising

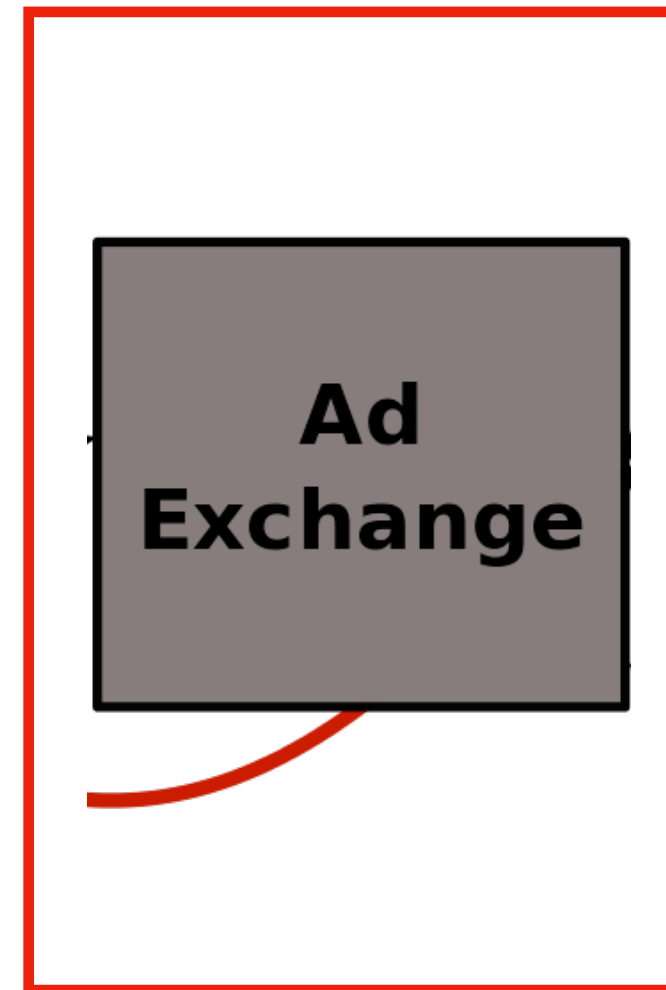
Demand Side Platforms



- On the other end of the pipeline, you have advertisers
- There are analogous entities called demand side platforms, which participate in Real-Time Bidding, which is a real-time auction for ad space (examples: Google DoubleClick, QuantCast, Criteo, Adform)
 - Typically happens in < 100ms

Online Advertising

Ad Exchanges



- Advertising exchanges receive spots from supply side, and facilitate real time bidding from the demand side based on properties of the ad spot
- Examples: Google DoubleClick, Facebook Exchange, PubMatic, Microsoft Advertising

Online Advertising

Bid Requests

```
"site": {
  "id": "1234",
  "name": "Example Site",
  "domain": "examplesitedomain.com",
  "mobile": 1,
  "amp": 0,
  "pub": {
    "id": "9876",
    "name": "Example Publisher, Inc.",
    "domain": "examplepubdomain.com"
  }
},
"user": {
  "id": "a0af45c77890045deec100acb8443baff57c",
  "consent": "ihdknkhkq8y",
  "buyeruid": "fcd4282456238256034abcdef220d9aa5892",
  "yob": 1990,
  "gender": "F",
  "ext": {
    "consented_providers_settings": {
      "consented_providers": [
        1,
        52,
        45,
        23
      ]
    }
  }
},
"device": {
  "type": 4,
  "ifa": "8846d6fa10008bceaaf322908dfcb221",
  "ip": "1.2.3.4",
  "ua": "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10.6; en-US; rv:1.9.2.16) Gecko/20110319
Firefox/3.6.16",
  "make": "Apple",
  "model": "iPhone",
  "hvw": "6s",
  "os": 13,
  "osv": "11.4.1",
  "mccmnc": "310-005",
  "geo": {
```

[https://
protocol.bidswitch.com/
rtb/request-
examples.html](https://protocol.bidswitch.com/rtb/request-examples.html)

Online Advertising

Bid Response

```
{
  "id": "d7d1e107-987h",
  "cur": "usd",
  "ext": {
    "protocol": "6.0"
  },
  "seatbid": [
    {
      "seat": "4",
      "bid": [
        {
          "id": "qwerty-098765",
          "item": "asdf-7890",
          "price": 1.45,
          "cid": "app-mraid-campaign-3442",
          "burl": "https://adserver.com/winnotice?impid=102&winprice=${AUCTION_PRICE}",
          "macro": [
            {
              "key": "TIMESTAMP",
              "value": "1127987134"
            }
          ],
          "ext": {
            "agency_id": "agency_123",
            "advertiser_name": "example advertiser"
          },
          "media": {
            "ad": {
              "id": "creative_id_1234",
              "adomain": [
                "example.com",
                "example.io"
              ],
              "cat": [
                "cat_1",
                "cat_2"
              ],
              "lang": "en",
              "attr": [
                3,
                7
              ]
            }
          }
        }
      ]
    }
  ]
}
```

[https://
protocol.bidswitch.com/
rtb/response-
examples.html](https://protocol.bidswitch.com/rtb/response-examples.html)

Online Advertising

Bidding for Ad Spots

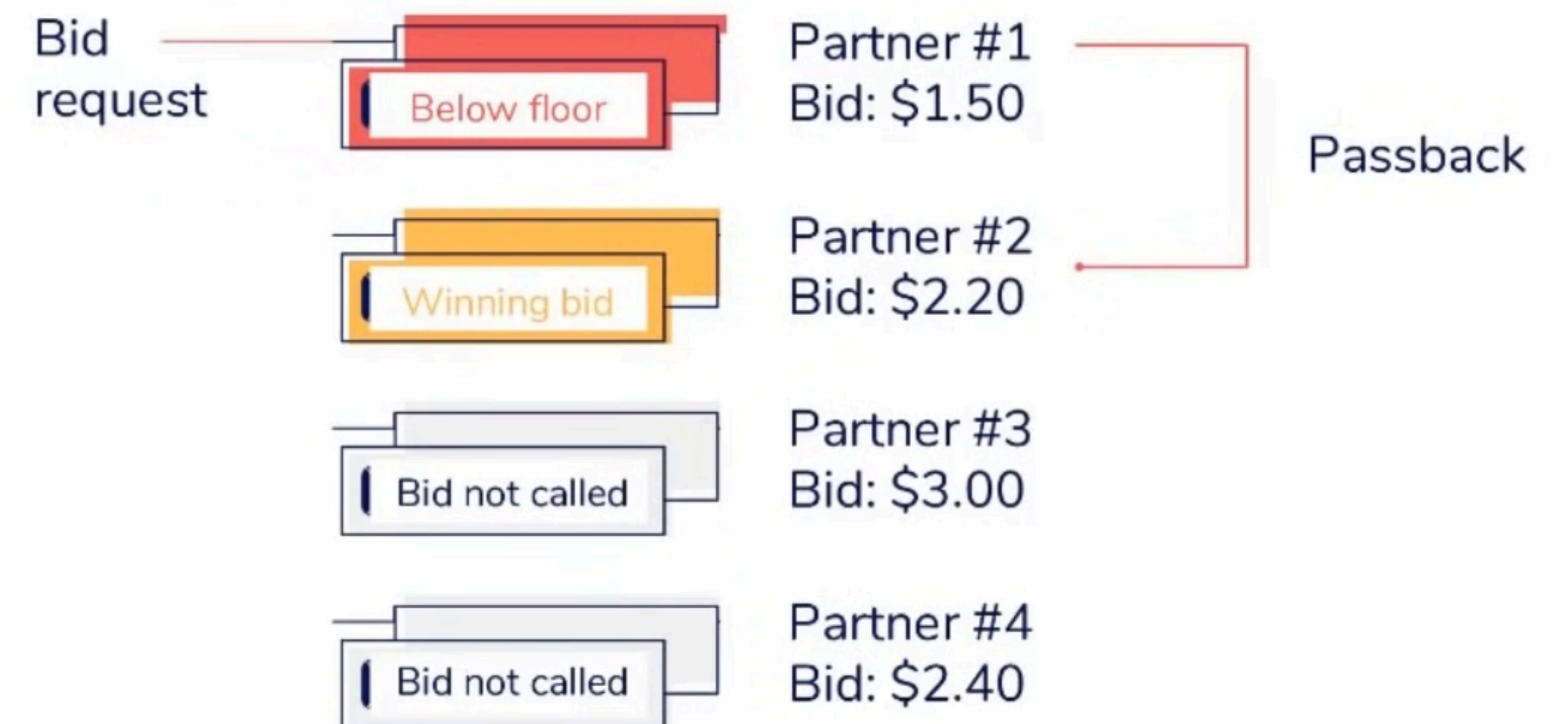
- Real-time bidding is an auction process that is kicked off when a publisher tells an advertising network that they have an open ad-spot with certain properties
- Two most widely used methods of auctioning
 - Waterfall bidding
 - Header bidding

Online Advertising

Waterfall Bidding

- Publishers would pre-define a hierarchy of advertising networks that they wanted to ask in order (e.g., in a waterfall) about any given advertising spot
- Publishers would then set a floor bid rate that they needed for the ad spot
 - The first network to fulfill the floor would win the spot, but floor price goes down with lower priority
- Problems:
 - Slow (serial computation)
 - Anti-competitive!
 - Google had both an SSP and a DSP, which often meant they got first pick at ad spots

\$2.00 FLOOR PRICE



Online Advertising

Header Bidding

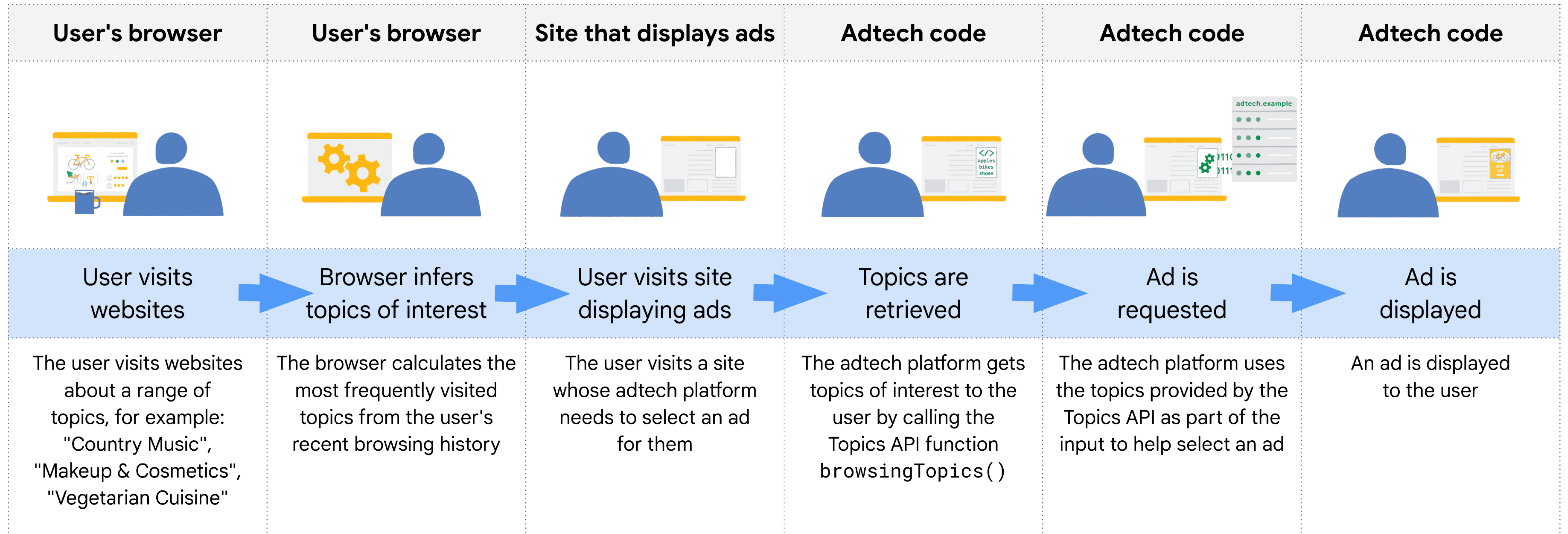
- Every DSP is offered the auction at the *same time*, and DSPs are incentivized to provide their true value for the advertising spot (theoretically)
 - This typically happens in under 100 ms
- Two options:
 - Client-side header bidding (happens in JavaScript), potentially makes the page slower, but have finer grained access to cookies
 - Server-side header bidding (happens in the SSP), can be faster, but requires cookie syncing, could make things slower

~~2024 — The Year of Linux on the Desktop~~

2024 — The Year of the End of Third Party Cookies?

- Firefox:
 - Third-Party Cookies from known trackers are dropped
 - Third-party cookies use separate cookie jar per site, so they can't be used to track users across sites
- Safari: Blocks third-party cookies
- IE: blocks some third-party cookies based on baked-in blacklist
- Edge does not block third-party cookies by default
- Chrome announced that they will drop support for third party cookies by the end of 2024

Google Topics



Topics are selected from a taxonomy consisting of hierarchical categories such as /Arts & Entertainment/Music & Audio/Soul & R&B and /Business & Industrial/Agriculture & Forestry.

The (maximum) three topics returned for a user are chosen at random from the top five for the past three epochs (with a 5% chance of getting a random topic).

Google Topics

Site	Topics	API callers on site
yoga.example	Fitness	adtech1.example adtech2.example
knitting.example	Crafts	adtech1.example
hiking-holiday.example	Fitness, Travel & Transportation	adtech2.example
diy-clothing.example	Crafts, Fashion & Style	[none]

At the end of the epoch (currently one week) the Topics API generates the browser's top topics for the week.

- adtech1.example is now eligible to receive the "Fitness" and "Crafts" topics, since it observed them on yoga.example and also on knitting.example.
- adtech1.example is not eligible to receive the "Travel & Transportation" topic for this user as it is not present on any sites the user visited recently that are associated with that topic.
- adtech2.example has seen the "Fitness" and "Travel & Transportation" topics, but has not seen the "Crafts" topic.

The user visited diy-clothing.example, which has the "Fashion & Style" topic, but there were no calls to the Topics API on that site. At this point, this means the "Fashion & Style" topic would not be returned by the API for any caller.

In week two, the user visits another site:

Site	Topics	API callers on site
sewing.example	Crafts	adtech2.example

In addition, code from adtech2.example is added to diy-clothing.example:

Site	Topics	API callers on site
diy-clothing.example	Crafts, Fashion & Style	adtech2.example

As well as "Fitness" and "Travel & Transportation" from week 1, this means that adtech2.example will now be able to receive the "Crafts" and "Fashion & Style" topic — but not until the following epoch, week 3. This ensures that third parties can't learn more about a user's past (in this case, an interest in fashion) than they could with cookies.

Studying the Web

Headless Chrome

- Headless Chromium allows running Chromium in a headless/server environment.
- "Expected use cases include loading web pages, extracting metadata (e.g., the DOM) and generating bitmaps from page contents — using all the modern web platform features provided by Chromium and Blink."
- Let's you load and manipulate websites DOM's after Chromium has loaded them using Node.js

Top Lists

- Most studies about the web use “Top Lists”
 - one million domain long lists of who thinks the most popular sites on the web are
- Most popular list was “Alexa Top Million”, which went away in 2022!
- Where does this data come from??! 🤔

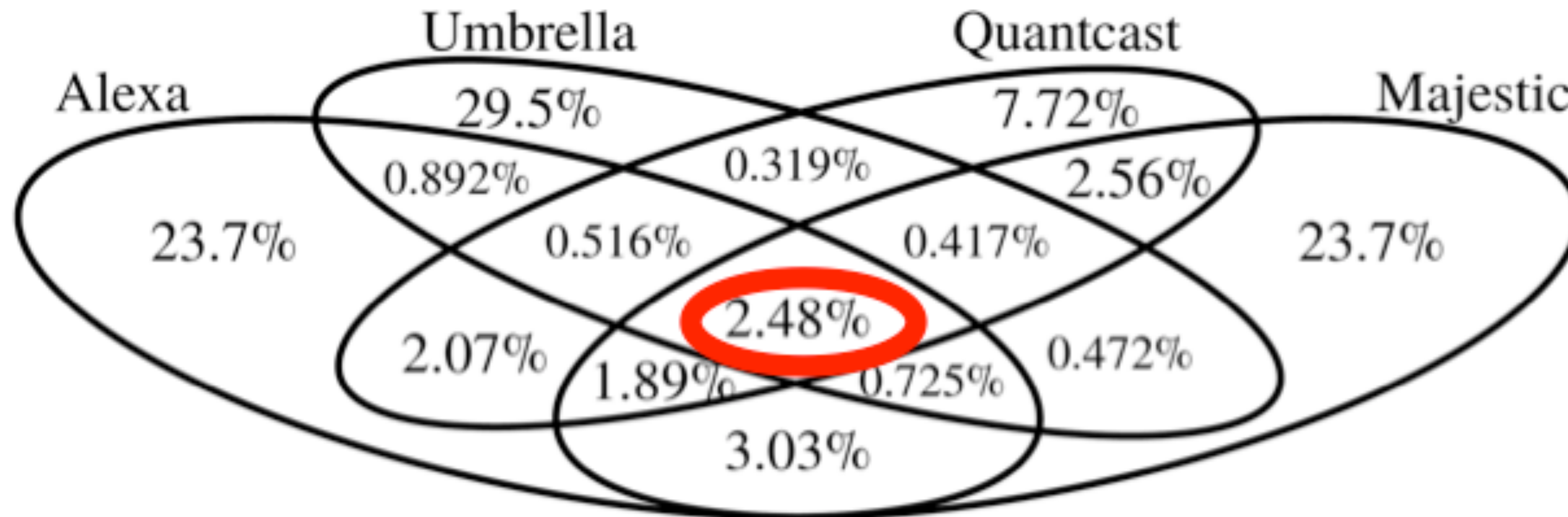
Example Top Sites

Rank	Site
1	google.com
2	amazonaws.com
3	facebook.com
4	a-msedge.net
5	microsoft.com
6	apple.com
7	googleapis.com
8	youtube.com
9	akamaiedge.net
10	akamai.net
11	azure.com
12	twitter.com
13	instagram.com
14	googlevideo.com

Top Lists – Up Until Recently

- Alexa: approximates site popularity by tracking the browsing behavior of several million users through partnerships with a reported 25K browser extensions as well as through websites that install Alexa Certify code
- Cisco Umbrella 1 Million is a list of the most popular names (e.g., .com is ranked #1) looked up using Cisco Umbrella's DNS service
- Majestic Million [20] is a list of popular websites maintained by Majestic SEO, which is calculated based on the number of backlinks that each site has
- Secrank list is a researcher-built list that aggregates several features of DNS data from a major resolver in China.
- Tranco aggregates data from the Alexa, Umbrella, and Majestic lists over a 30 day window to form a ranking that is more temporally stable and resistant to adversarial manipulation

Top Lists

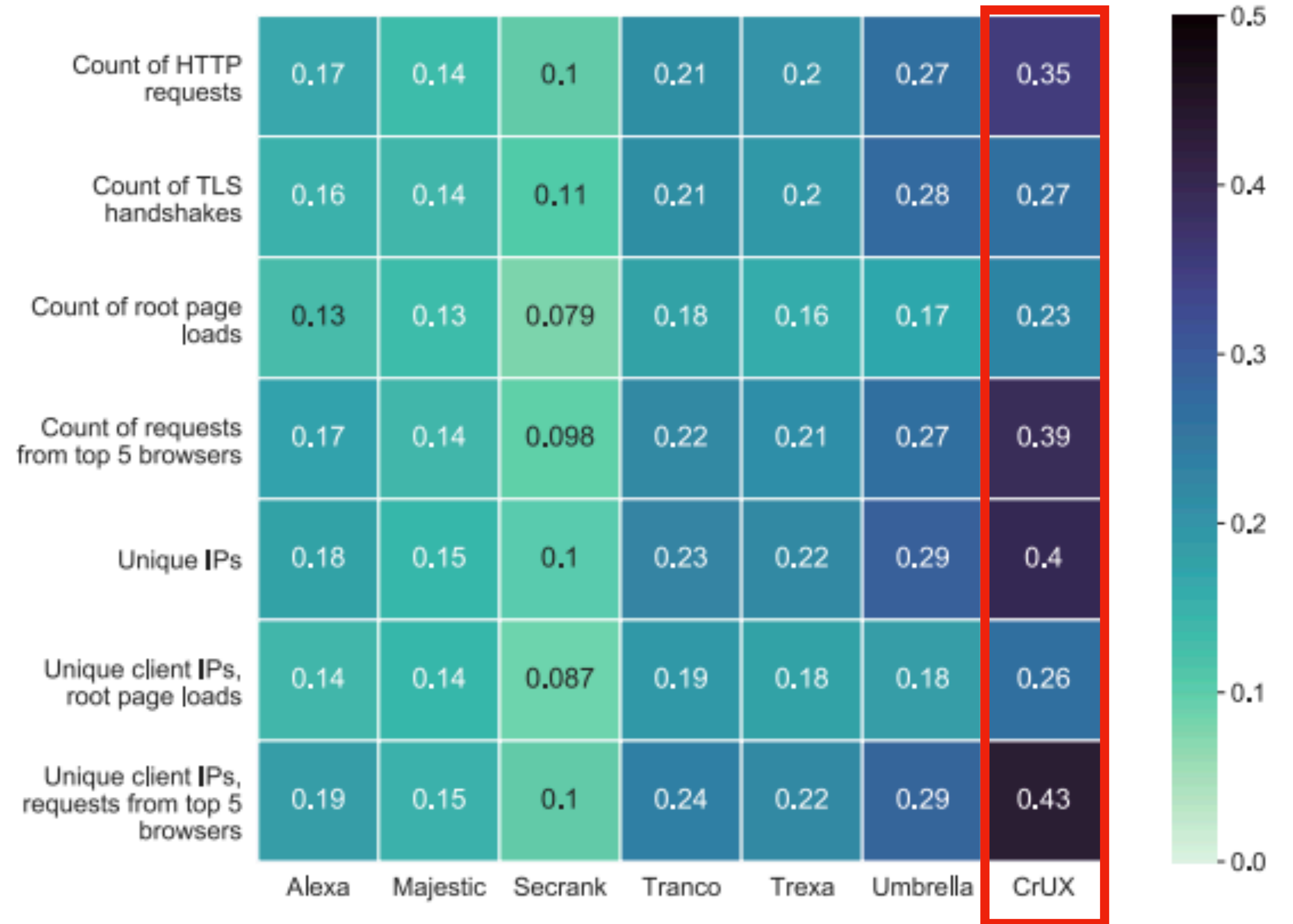


The Tale of Website Popularity Rankings: An Extensive Analysis (2019)

Tajalizadehkhoob et al.

Top Lists

- In 2022, Google Chrome agreed to start publishing a list of top sites as seen by anonymous telemetry
- Gathered from Chrome Users who have explicitly enabled sharing URLs with Google and who have usage statistic reporting enabled.
- Part of Chrome User Experience Report (CrUX) dataset
- Can be downloaded from <https://github.com/zakird/crux-top-lists>



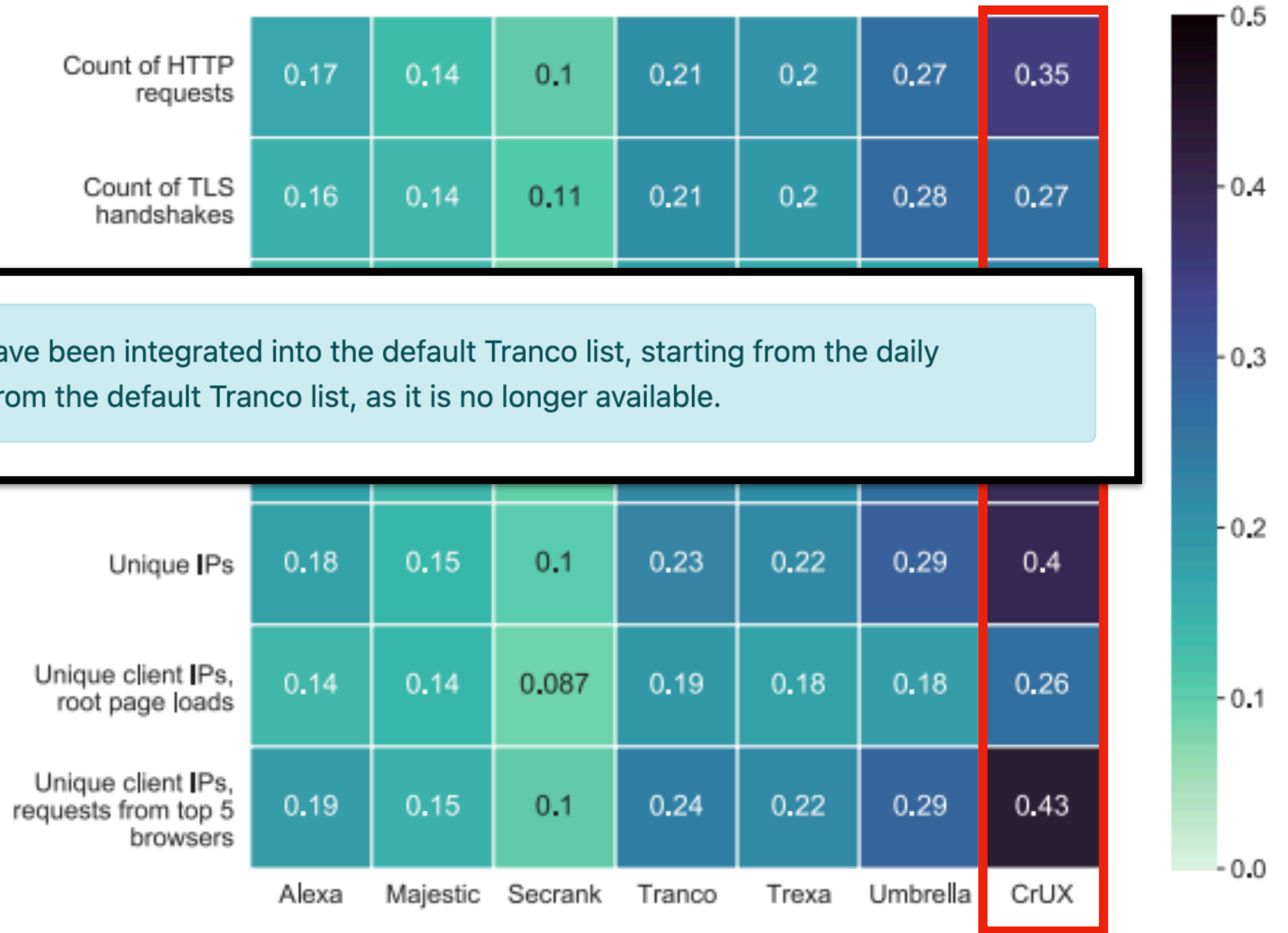
(a) Jaccard Index

Top Lists

- In 2022, Google Chrome agreed to start publishing a list of top sites

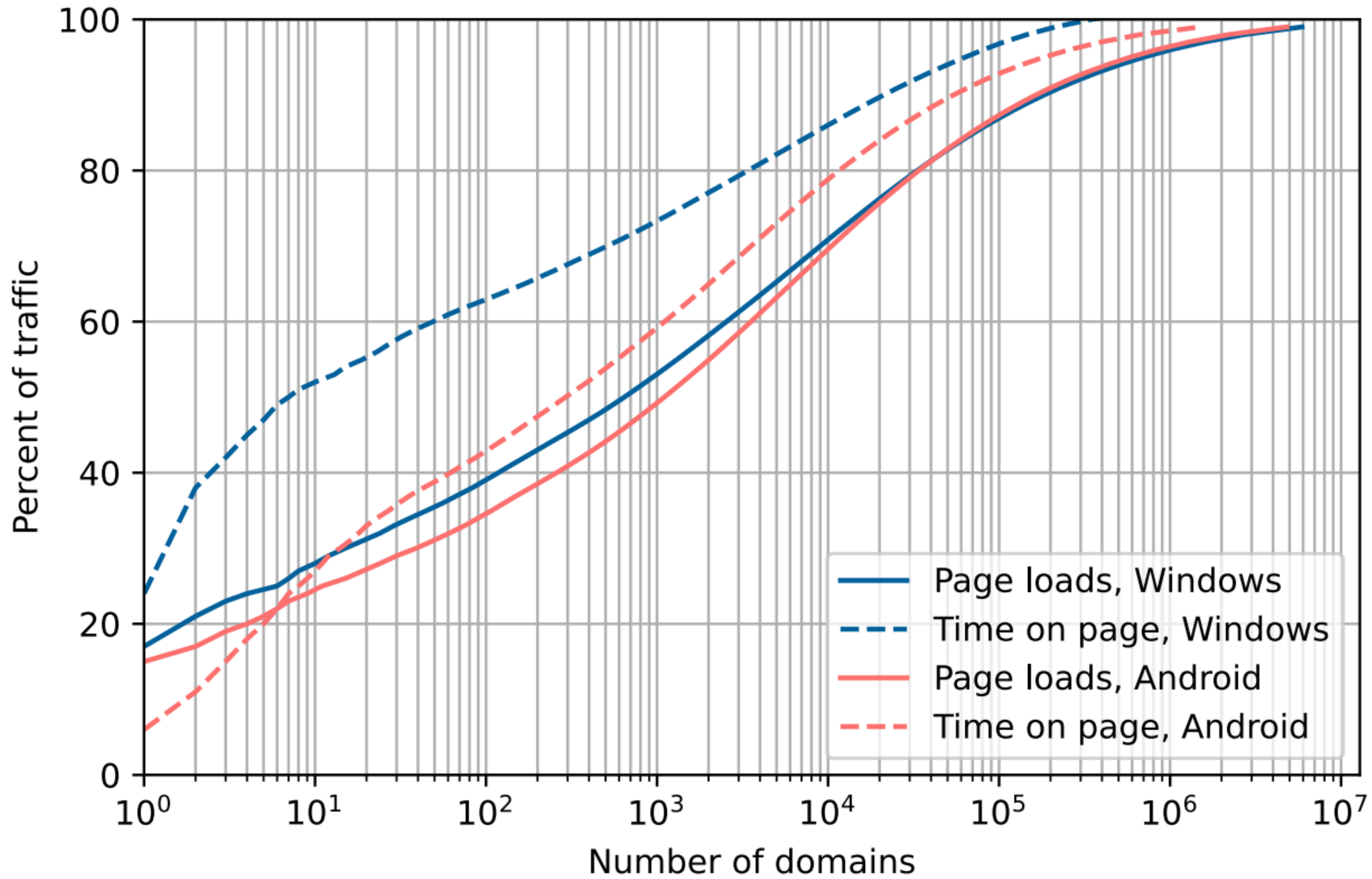
The Chrome User Experience Report and Cloudflare Radar rankings have been integrated into the default Tranco list, starting from the daily updated list of August 1, 2023. The Alexa ranking has been removed from the default Tranco list, as it is no longer available.

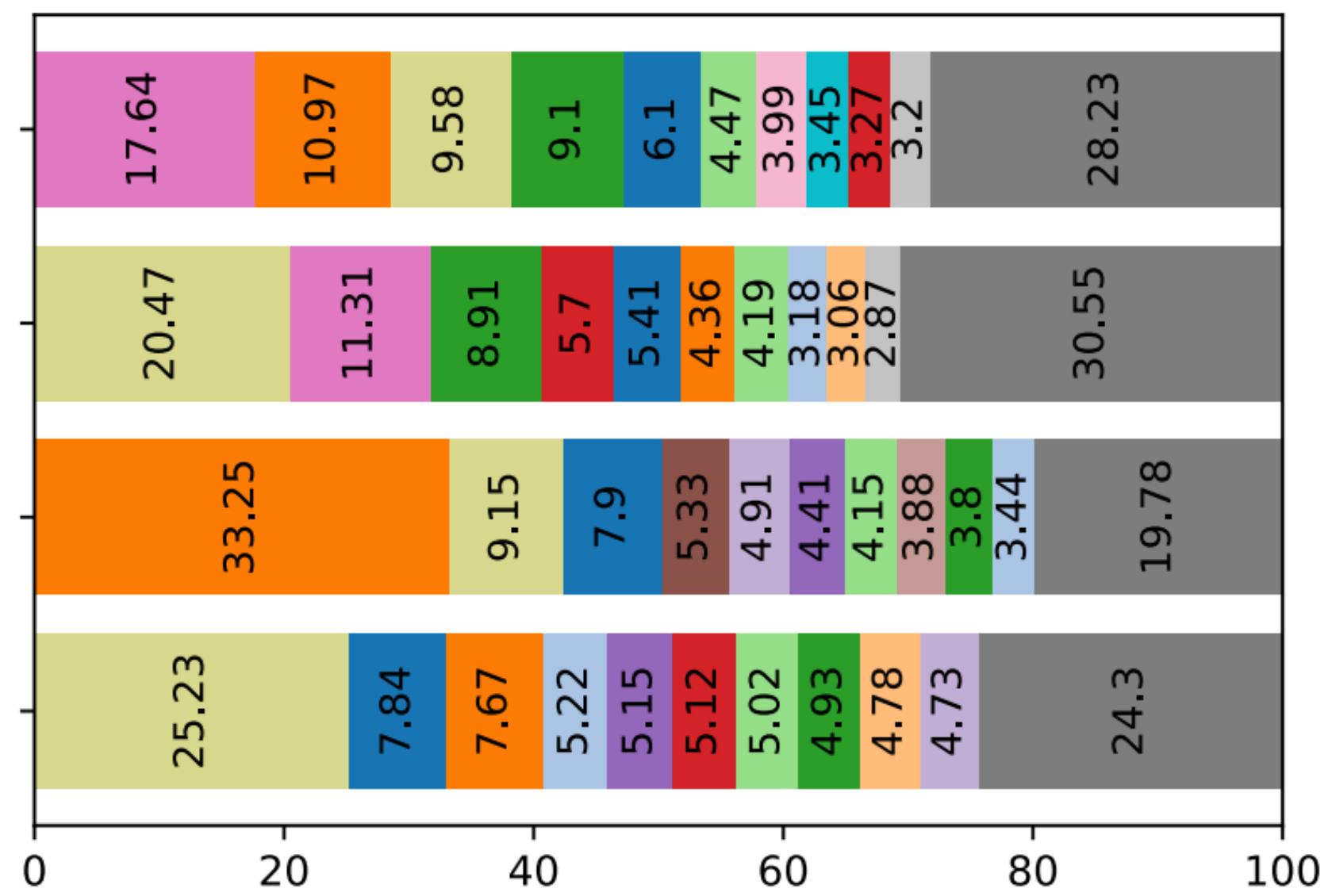
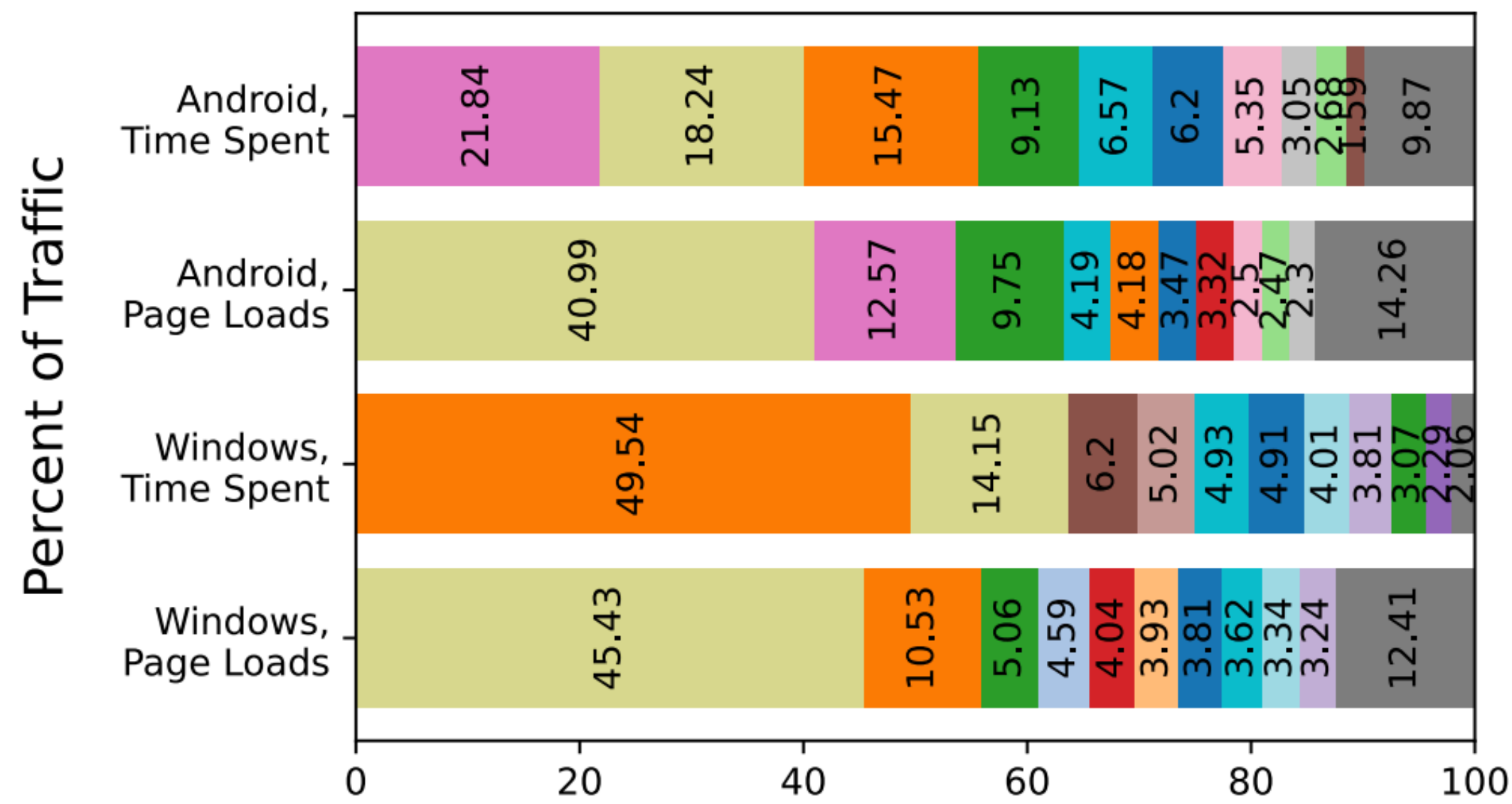
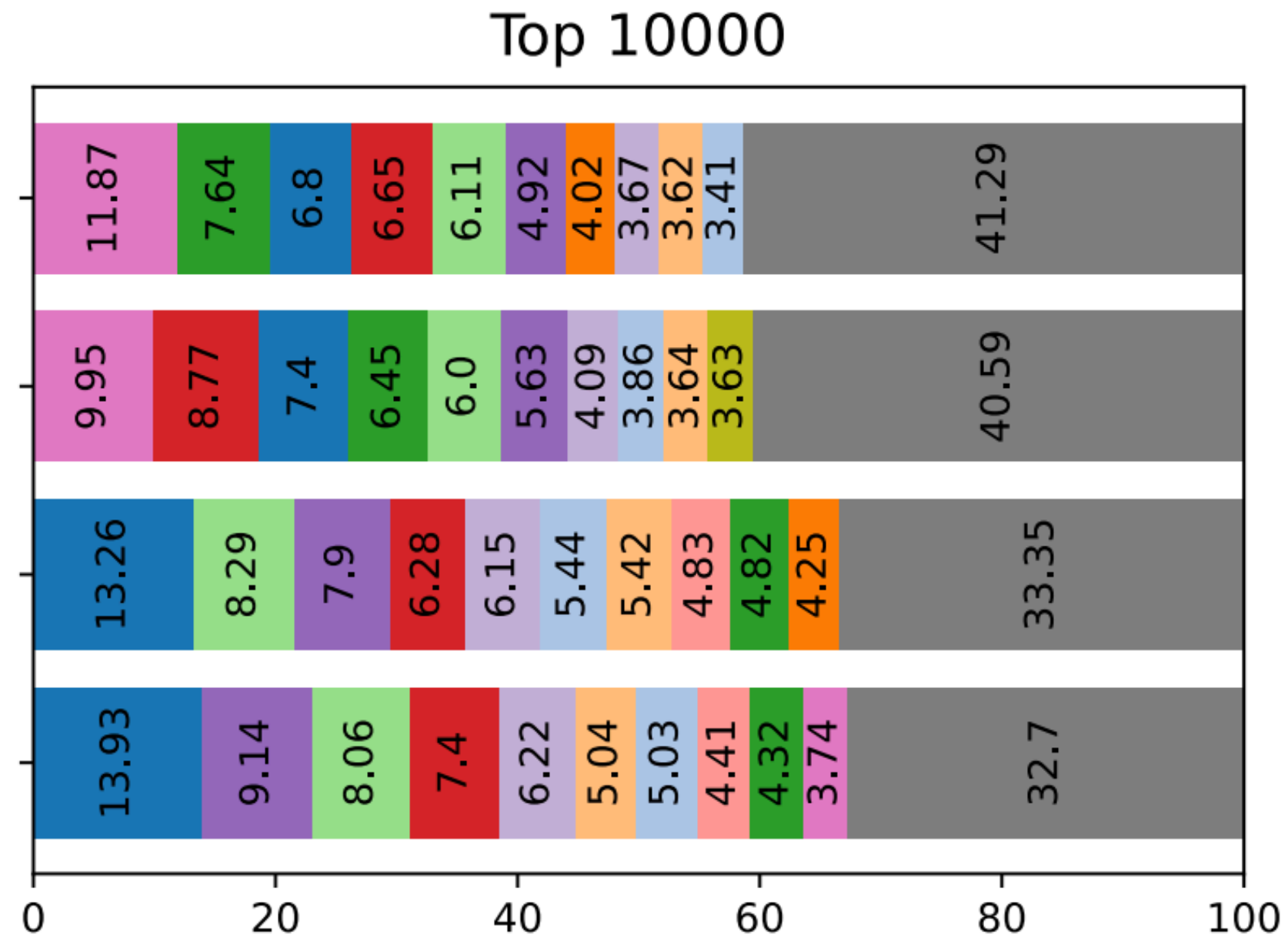
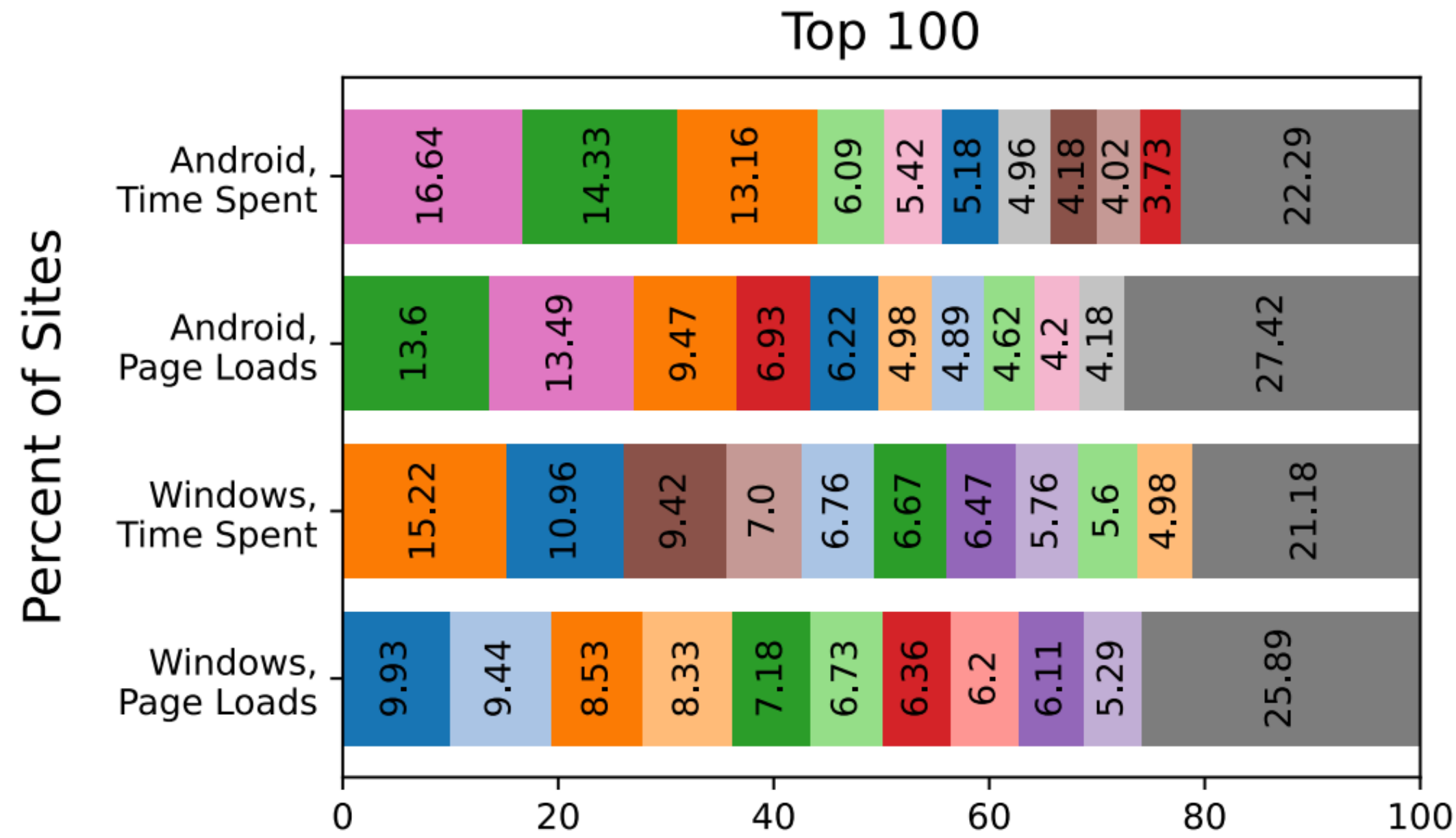
- URLs with Google and who have usage statistic reporting enabled.
- Part of Chrome User Experience Report (CrUX) dataset
- Can be downloaded from <https://github.com/zakird/crux-top-lists>



(a) Jaccard Index

Shape of the Web





- Technology
- Economy & Finance
- Video Streaming
- Politics and Advocacy
- News & Media
- Education
- Ecommerce
- Educational Institutions
- Business
- Gaming
- Movies and Home Video
- Television
- Pornography
- Gambling
- Sports
- Clothing and Fashion
- Search Engines
- Social Networks
- Chat and Messaging
- Other

