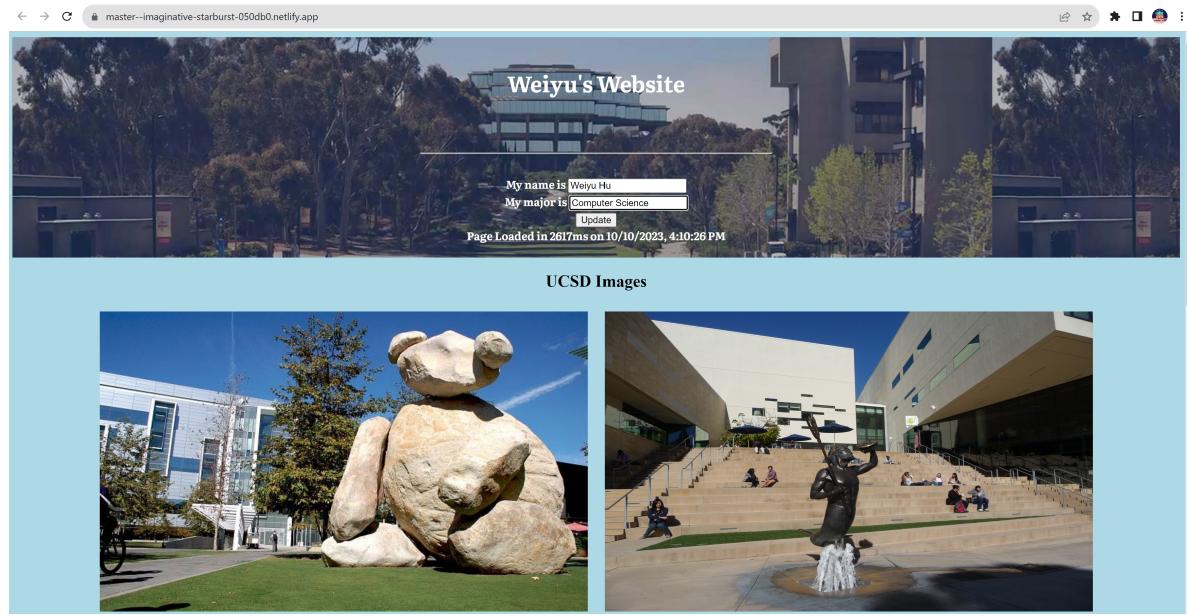


# Part 1

---

## Q1 Introduction to Web Technologies



Link: <https://master--imaginative-starburst-050db0.netlify.app>

---

## Q2 Chrome DevTools - Network

### 1. # of Requests by Content Type

- HTML: 1
- CSS: 1
- JS: 1
- Font: 1
- GIF: 1
- JPG: 2
- PNG: 1
- MP4: 1
- SVG: 17
- ICO: 1

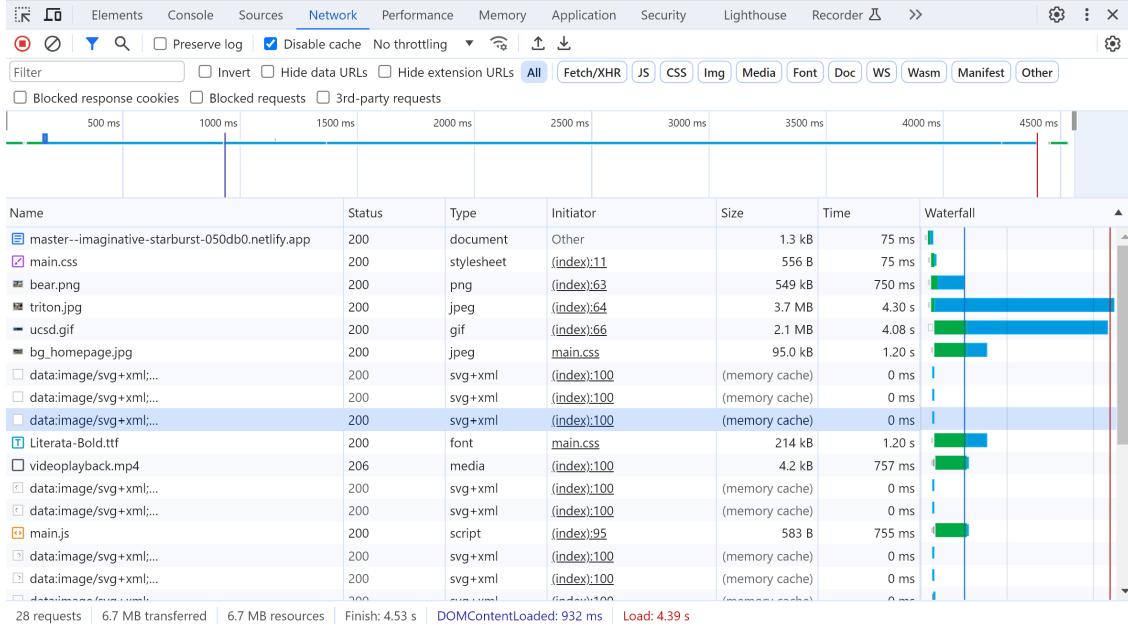
### 2. Total # of Requests

28 requests

### 3. Total Bytes Sent

6.7 megabytes

#### 4. Waterfall of Requests Screen Capture



## Q3 Client-Side Inherently Insecure Demo

The screenshot shows a web application with a header 'ALL THE POINTS' and a sub-header 'Author: Give me the points .... Lorem ipsum dolor sit amet, consectetur adipiscing elit. I want more points'. Below this is a form with fields for 'Major: Computer Science' and 'My name is' (containing 'Give me the points ...'). There's also an 'Update' button and a timestamp 'Page Loaded in 2765ms on 10/10/2023, 4:35:03 PM'. The background features a photograph of a modern building with a large rock sculpture in front. On the right side, the Chrome DevTools Elements tab is open, showing the DOM structure and some inline JavaScript comments. The Network tab shows a request for 'UCSD Images'.

## Part 2

1. Were any parts of navigating ESPN site easy? Were any parts difficult?

Easy: It was easy to go through every item. Therefore, I just have to wait until I hear **Top Headlines**.

Difficult: It took me a long time to hear **Top Headlines** because it is located at a far right corner.

2. Were any parts of navigating webaim.org's site easy? Were any parts difficult?

Easy: It was quick to find the answer about archive once the page is correct.

Difficult: There are many similar terms that sound like *Web Accessibility Virtual Training*, and there are many links included in these clickable links. It was easy to enter other pages that do not have the information about archive classes.

## Part 3

### Q1 HTTP Response headers

#### 1. UCSD:

##### ▼ Response Headers

Accept-Ranges:	bytes
Content-Length:	46316
Content-Type:	text/html; charset=UTF-8
Date:	Wed, 11 Oct 2023 01:57:37 GMT
Etag:	"b4ec-607652303a260"
Last-Modified:	Tue, 10 Oct 2023 23:33:40 GMT
Server:	Apache/2.4.6 (Red Hat Enterprise Linux)
Strict-Transport-Security:	max-age=0; includeSubDomains

#### 2. UCI:

##### ▼ Response Headers

Raw

Connection:	Keep-Alive
Content-Type:	text/html; charset=UTF-8
Date:	Wed, 11 Oct 2023 02:00:39 GMT
Keep-Alive:	timeout=2
Server:	Apache/2.4.6 (CentOS) OpenSSL/1.0.2k-fips mod_auth_kerb/5.4 mod_perl/2.0.11 Perl/v5.16.3
Strict-Transport-Security:	max-age=31536000; includeSubDomains
Transfer-Encoding:	chunked
X-Frame-Options:	SAMEORIGIN
X-Xss-Protection:	1; mode=block

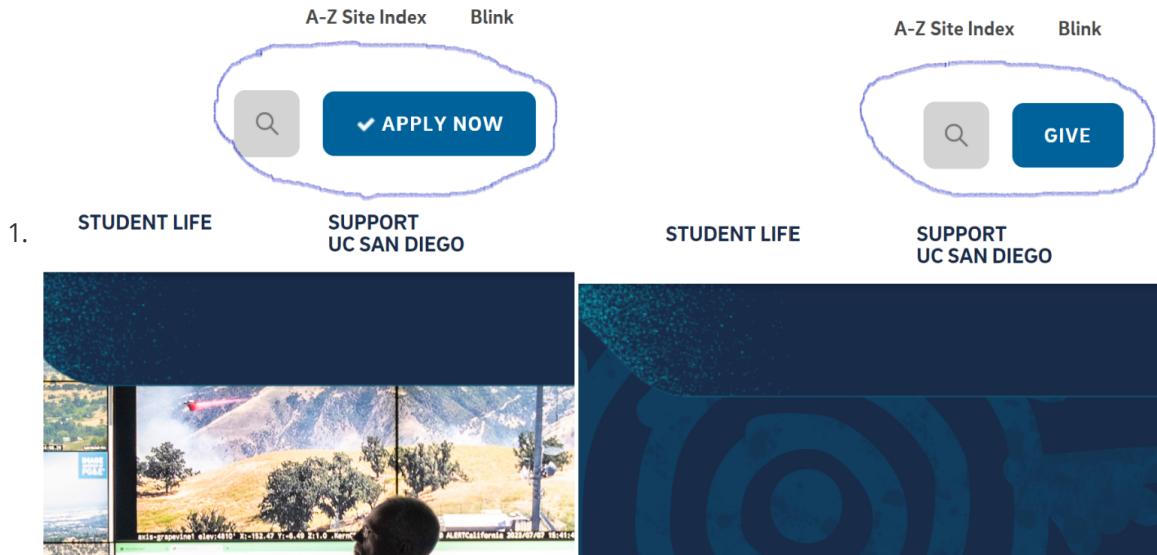
#### 3. What is troubling about the UCI response HTTP headers?

In the server section, it is exposing the server's information. This is an information disclosure issue that could be troubling to the UCI website if someone wants to do something malicious to the page.

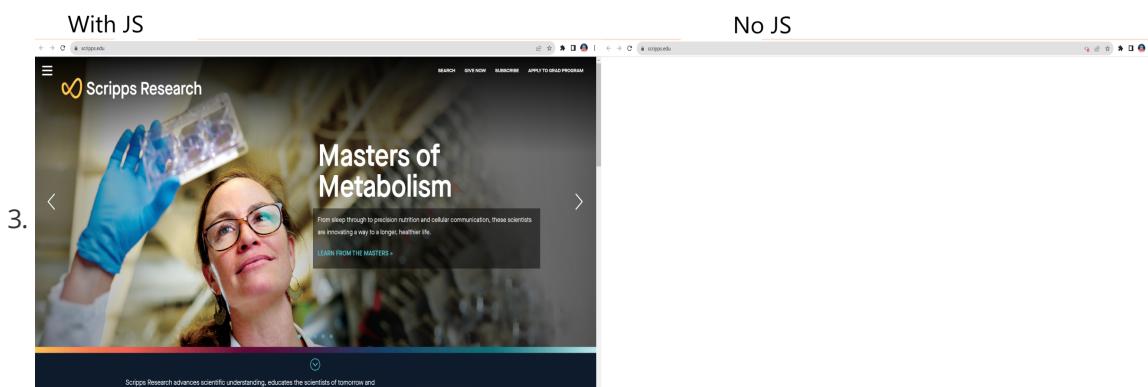
## Q2 JavaScript Off

With JS

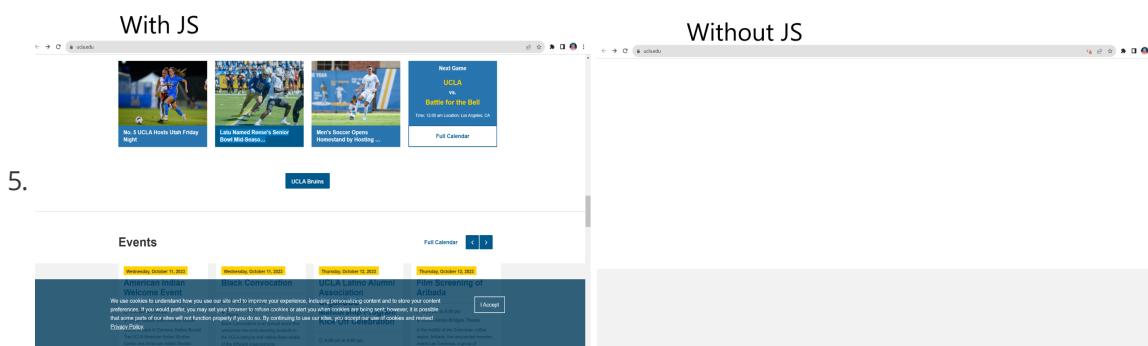
Without JS



1. STUDENT LIFE SUPPORT UC SAN DIEGO
2. One obvious broken feature is the search button located top right of the page. When JS is disabled, clicking the magnifying glass pattern does nothing. Also when the website is shrunk to point when the "topics" such as *ABOUT*, *ACADEMICS*, *ADMISSIONS AND AID*, etc, are not displayed and they are only visible if the drop down button is clicked, the drop down button does not show the "topics" with JS disabled.



3. Masters of Metabolism LEARN FROM THE MASTERS >
4. When JS is disabled, all of the contents in the Scripps page are gone, only a blank page is left.



- 
6. The website of UCLA is partially functioning when JavaScript is disabled: most of the content are still visible and clickable; clicking to the buttons or links can still redirect the page to the designated page; however, in the middle of the page, the page loses its visuals and leaves a blank in between the contents.
- 

## Q3 Custom vs. Default 404 Pages

1. Yes
  2. Link: <https://www.csuci.edu/programs>
  3. No
  4. Link: <https://pcatholic.edu/admissions>
  5. Users tend to have no idea about what a default 404 page means. They could think that the problem is on their end since the default 404 page is not informative; whereas a custom is way more informative and interactive, as seen from the csuci page. The users then would know what the problem is and continue their browsing on the page.
- 

## Q4 Search Engines - robots.txt

1. Yes.



The screenshot shows a browser window with the URL "nytimes.com/robots.txt" in the address bar. The page content is the robots.txt file for the New York Times. It contains rules for various user-agents, including search engines like Google and Bing, and specific crawlers like CCBot and GPTBot. It also includes sitemap URLs and a note about disclosing information to search engines.

```
User-agent: *
Disallow: /ads/
Disallow: /adx/bin/
Disallow: /puzzles/leaderboards/invite/*
Disallow: /svc/
Allow: /svc/crosswords
Allow: /svc/games
Allow: /svc/letter-boxed
Allow: /svc/spelling-bee
Allow: /svc/vertex
Allow: /svc/wordle
Disallow: /video/embedded/*
Disallow: /search
Disallow: /multiproduct/
Disallow: /hd/
Disallow: /int/
Disallow: /*?query=
Disallow: /*.pdf$*
Disallow: /*?login=
Disallow: /*?searchResultPosition=
Disallow: /*?campaignId=
Disallow: /*?mcubz=
Disallow: /*?smprod=
Disallow: /*?ProfileID=
Disallow: /*?ListingID=
Disallow: /wirecutter/wp-admin/
Disallow: /wirecutter/*_zip$
Disallow: /wirecutter/*_csv$
Disallow: /wirecutter/deals/beta
Disallow: /wirecutter/data-requests
Disallow: /wirecutter/search
Disallow: /wirecutter/*?*
Disallow: /wirecutter/*&id=
Disallow: /wirecutter/*?q=
Disallow: /wirecutter/*?l=
Disallow: /search
Disallow: /*?smid=
Disallow: /*?partner=
Disallow: /*?utm_source=
Allow: /wirecutter/*?utm_source=
Allow: /ads/public/
Allow: /svc/news/v3/all/pshb.rss

User-agent: CCBot
Disallow: /

User-agent: Google-Extended
Disallow: /

User-agent: GPTBot
Disallow: /

User-agent: ia_archiver
Disallow: /

User-Agent: omgili
Disallow: /

User-Agent: omgilobit
Disallow: /

User-agent: Twitterbot
Allow: /*?snid=>

Sitemap: https://www.nytimes.com/sitemaps/new/news.xml.gz
Sitemap: https://www.nytimes.com/sitemaps/new/sitemap.xml.gz
Sitemap: https://www.nytimes.com/sitemaps/new/collections.xml.gz
Sitemap: https://www.nytimes.com/sitemaps/new/video.xml.gz
Sitemap: https://www.nytimes.com/sitemaps/new/cooking.xml.gz
Sitemap: https://www.nytimes.com/sitemaps/new/recipe-collects.xml.gz
Sitemap: https://www.nytimes.com/sitemaps/new/regions.xml
Sitemap: https://www.nytimes.com/sitemaps/new/best-sellers.xml
Sitemap: https://www.nytimes.com/sitemaps/www.nytimes.com/2016_election_sitemap.xml.gz
Sitemap: https://www.nytimes.com/elections/2018/sitemap
Sitemap: https://www.nytimes.com/wirecutter/sitemapindex.xml
```

2. While robots.txt blocks search engine crawlers, it discloses information about the pages that the owner wants to hide from the others. Then robots.txt can reveal hidden information that

is not intended for the public.

3. The nytimes.com tries to block User-agents: CCBot, Google-Extended, GPTBot, ia\_archiver, omgili, omgilobit, and Twitterbot. Such bots attempt to extract contents from the internet to researchers, companies, and individuals for their purposes at no cost. Since NY Times require subscription for its content, such bots are then blocked.
- 

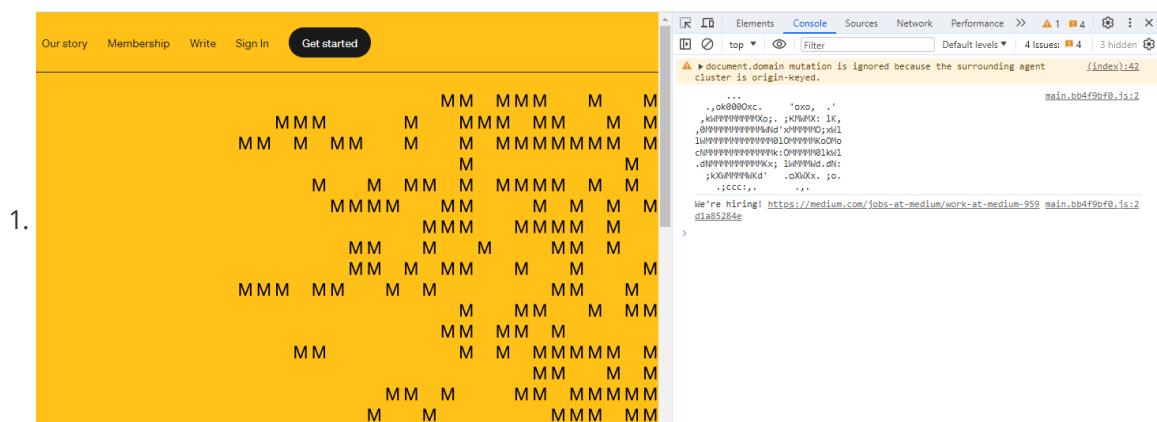
## Q5 Search Engines - Google Hacking

1. Google Hacking Database is a set of Google search queries that can be used to find sensitive or hidden information that is on the web. Individuals or organizations might use this to uncover the vulnerabilities and information that others do not intend to expose.
  2. Google bot is like the bots mentioned in the previous question, such a bot can index content that is private, sensitive, or even "secured" data behind the login screen. It is troubling that it can use Google search to discover such information.
  3. Developers should be aware of such bots and carefully measure the security level required for the web pages, directoris, or files so that confidential data are not exposed.
  4. The problems in Question 5 are very similar to those from Question 4. Both questions introduce some weaknesses of the current web development that can cause information disclosure and reveal the vulnerability of the data to all of the people who have access to the Internet.
- 

## Q6 Search Engines - Results Reality Check

1. About 2,770,000 results (0.34 seconds).
  2. I was unable to locate the 500 - 510th URLs for "why ucsd is awesome" as the browser does not order the list of pages.
  3. On the first page, Duckduckgo shows results from UCSD itself and some other sites that only try to introduce UCSD in the perspectives of campus lives and academics; whereas Google shows more questions about UCSD on the first page, the rest of the results from Google are forum/conversation based, where people ask and answer about UCSD. The only similarity is that they both include ucsd.edu although there is only one result from ucsd.edu from Google, Duckduckgo has multiple. I don't have a preference for search engines. I switch to others when the one I am using does not provide the results that I am looking for.
- 

## Q7 Chrome DevTools - JavaScript Console and Local Storage



1. Medium uses this simple but surprising method to attract talented people to join them.

Application		
	Key	Value
Manifest	device deviceld	"fee7333c63b5ab1f"
Service workers	read-next hide-meter	1669703696997
Storage	post-article posts-viewed-count	1
Local storage	uid list	["e45b3967fa66"]
https://medium.com	pvl 3998d5ac2684	1679030645326
Session storage	MUTEX_/_oh-noes X	"1669675116963:09682959919907907"
IndexedDB	prj 959d1a85284e	1697521925007
Web SQL	sign-in-confirmation-banner e45b3967fa66	"bfc23ca2d9a78e83d6dfbeb6872cd8c5cc92e83c24a7..."
Cookies	post-article posts-viewed-month-count	1
Private state tokens	sprig-attributes latest.post.clientRead.source	"post_page-----ceeb4f5eb18d-----"
Interest groups	pv 773ce67fc6a7	1686641454285
Shared storage	uid previous	"lo_06844de93e77"
Cache storage	post-article first-post-viewed-timestamp	1697521975220
Background services	pv fbc5af75ecbe	1682410773773
Back/forward cache	pv 850bb247cfb2	1678922166431
Background fetch	google-one-tap hide-at	1697515836325
Background sync	pv 7b165bdc13f4	1679367595380
Bounce tracking mitigations	prj 3998d5ac2684	1679030657808
Notifications	userleap.ids	{"WISfSM8eD3":{"vid":"15631250-7f78-4faa-a674-104..."}
Payment handler	lo-non-moc-membership-upsell dismissed-at	"2023-10-17T05:52:00.255Z"
Periodic background sync	prj e07cf8d3d22b	168973668096
Push messaging	post-article posts-viewed-today-count	1
Reporting API	MUTEX_/_batch X	"1669675116964:7928498295949367"
Preloading	post-article first-post-viewed-timestamp	1697521975220
3. Speculation rules	pv 65797a5e675b	1686706217618
Preloads	post-article first-post-viewed-month-timestamp	1697521975220
This page	sprig-attributes latest.post.clientRead.url	"https://medium.com/%E5%B0%8B%E6%89%BE%E9%..."
Frames	post-article first-non-moc-post-viewed-month-timestamp	1697521917424
top	pv 5bd55364d7ca	1679365883367
	pv 5ddd8b0350e	1684366884944
	sprig-attributes email	"w7hu@ucsd.edu"
	post-article non-moc-posts-viewed-month-count	1
	pv e07cf8d3d22b	1689773570368
	pv 20ccf7b6b5c8	1697521975220
	pv 30a9131807e0	1690470771012
	sprig-attributes latest.post.clientRead.postId	"ceeb4f5eb18d"
	viewer-status is-logged-in	false
	pv 959d1a85284e	1697521917425
	pv 4015ebf13ff	1682410799834
	branch_session_first	{"session_id":"1028079450084272159","identity_id":1...}
	EventReporter Intwx12d147vx6364vu	{"key":"susie.viewed","data":{"entryPoint":"follow_card", ...}}
	1	"fee7333c63b5ab1f"

I think these values are used for validating my interaction with medium.com, my identity, and my membership of medium.com as there are values such as deviceld, view-count, email, is-logged-in, etc.

## Q8 Chrome DevTools - Console and Source

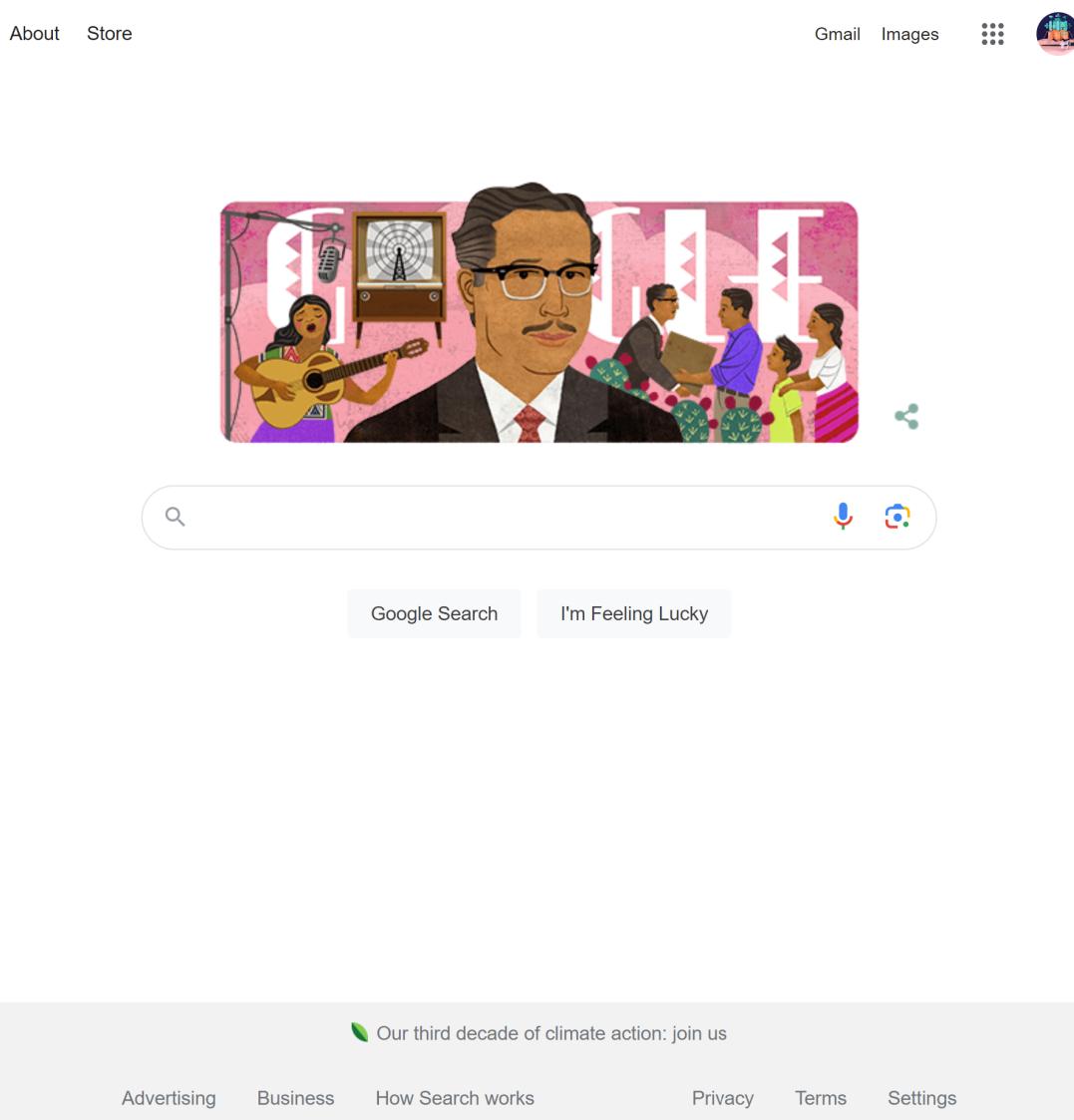
1. The name of the variable is datasetCount.
2. This message is probably there because the developers wanted to keep track of the number of datasets during developing and testing. Then they forgotto delete this line, or they didn't think that this is a huge problem.
3. The console message exposes the number of datasets that ca.gov uses. This can be problematic because it makes it easier for attackers to target a specific dataset for its vulnerabilities such that it can lead to data breach by web scraping.

---

## Q9 Chrome DevTools - User-Agent Header

1. Yes Google renders the pages differently.

**Default:**



**iPhone:**

The screenshot shows a Google search results page on an iPhone. At the top, there's a navigation bar with three horizontal lines, the word "ALL" underlined in blue, and "IMAGES". To the right are icons for a grid, a person, and a globe. Below the navigation is a large, colorful illustration of a man with glasses and a mustache, wearing a suit, standing in front of a microphone and a television set. To his left, a woman plays a guitar, and to his right, a group of people are gathered. Below the illustration is a search bar with a magnifying glass icon. Underneath the search bar, the text "Trending searches" is displayed, followed by a list of eight items, each with a small profile picture to the right:

- pastor ralph douglas west  
Ralph D. West — Writer
- lottery powerball jackpot
- tampa bay buccaneers baker mayfield  
Baker Mayfield — Football quarterback
- republican presidential debates
- jep sedgwick pitchbook
- durga navratri colours
- las vegas aces chelsea gray injury

2. With the default user agent, the page of google.com looks a lot cleaner. As a contrast, the page using an iPhone user agent fills with trending searches.

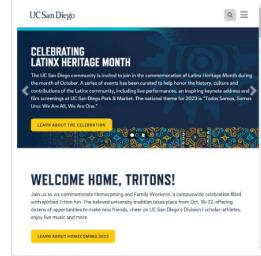
## Q10 Chrome DevTools - Extension Header

- *X-Content-Type-Options: nosniff* disables MIME sniffing, which ensures the browser to interpret files as it should be so that XSS attack is reduced.
- *X-Download-Options: noopener* prevents the downloaded file to run JS in the current site's context and removes the open button and replaces it with a save button when download
- *X-Frame-Options: DENY* prevents the page from being displayed in a frame or iframe.
- *X-Permitted-Cross-Domain-Policies: none* controls how data are shared across domains. Setting it to none prevents certain types of data to be shared.
- *X-Render-Origin-Server: Render* indicates that the origin server is responsible for rendering the page.
- *X-XSS-Protection: 0* disables XSS attacking filtering from the browser by setting the value to 0. Such a website can be vulnerable against XSS.

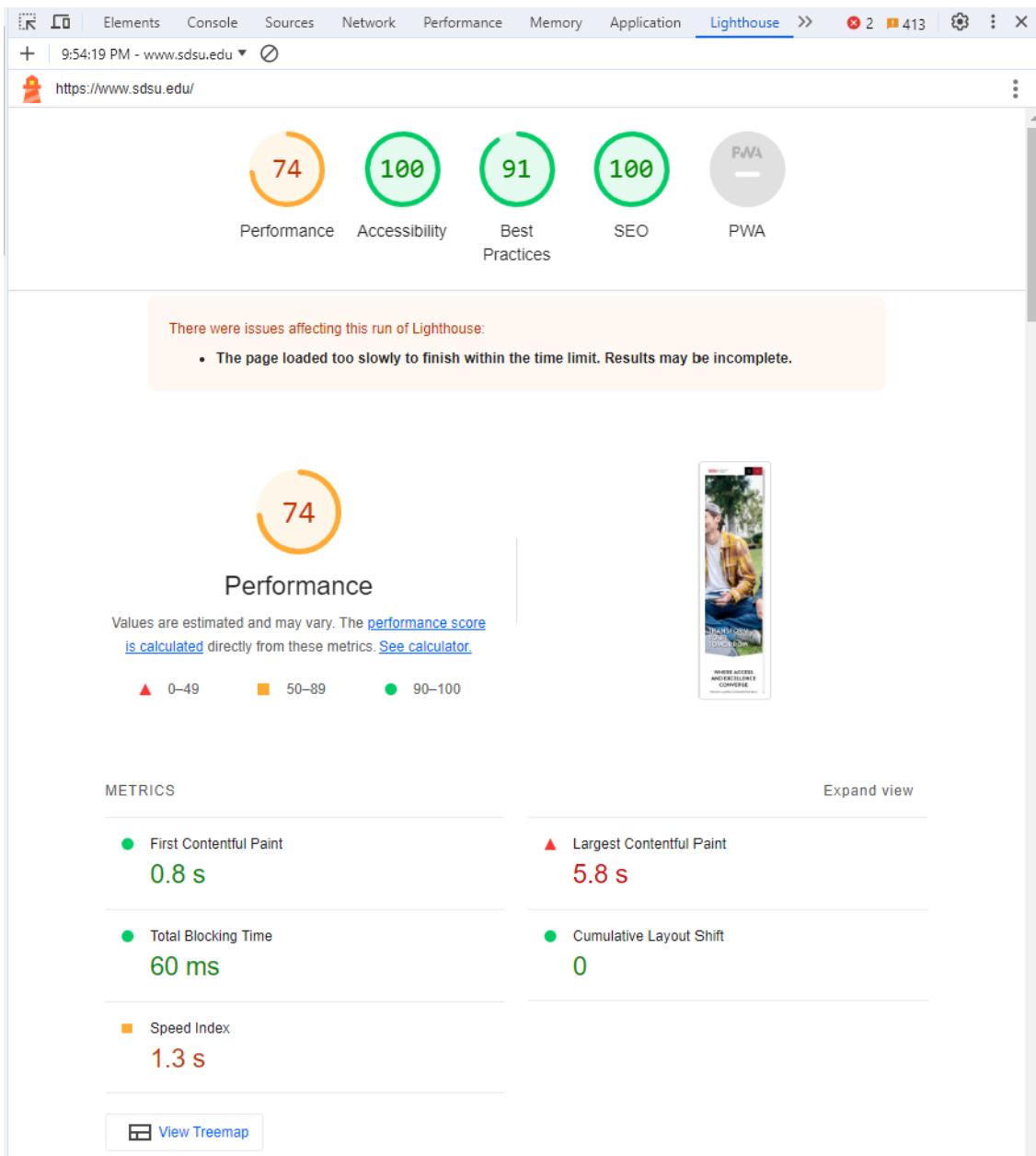
## Q11 Chrome DevTools - Performance Test

1. + 5:16:44 PM - ucsd.edu ▾ ⊖

https://ucsd.edu/



2.
  1. Reduce unused CSS: reduce the use of those redundant rules from stylesheet that consumed by the network activity. To do this, double check the stylesheet to find out which rules are not used or not necessary.
  2. Eliminate render-blocking resources: Running some JS code and compiling the CSS styles blocks the page from displaying its first look. As suggested from the audit, consider delivering critical JS/CSS inline and deferring all non-critical JS/styles.



The scores from sdsu.edu are way higher than from ucsd.edu.

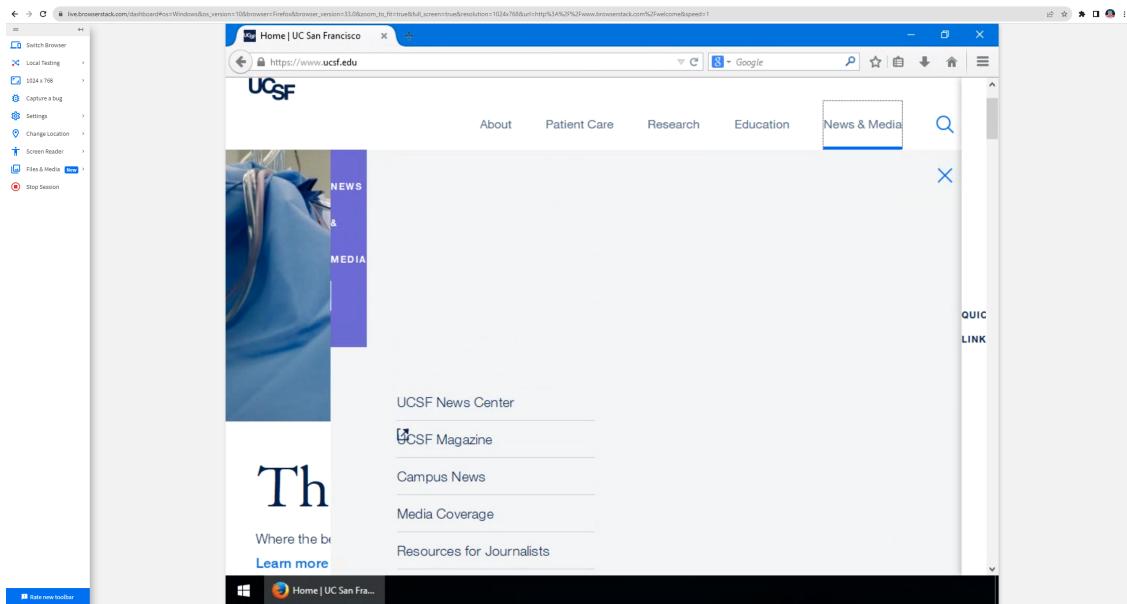
4. UCSD has 57 as the score for performance, whereas SDSU has 74 as the score for performance. UCSD has more work to improve the performance on its page.

## Q12. Browsers Versions

1. Chrome Platform Status (Chromium): <https://www.chromestatus.com/>
  2. Firefox Platform Status (Mozilla): <https://github.com/mozilla/platform-status> (there is no actual status page like the other three so far)
  3. Edge Platform Status (Microsoft Edge): <https://developer.microsoft.com/en-us/microsoft-edge/platform/status/>
  4. Safari (Webkit) Feature Status (Apple): <https://webkit.org/status/>
2. In version 53, Chrome introduced Shadow DOM v1.  
 3. The version was released on August 30th, 2016.  
 4. Google released the non beta version of Chrome on December 11th, 2008 for Windows XP.

## Q13 Testing Different or Older Browsers

The oldest version I saw for firefox was v32. My free trial session for v32 has ended before I can see anything, so I was testing on v33.



1.

The website works but the scales and visibilities of texts are off. It is difficult to use.

2. Since a website cannot support all the browsers from time to time, the site should have a document that lists the browsers and the versions it supports. What we can do is to try our best to make it as compatible as possible to address this problem.

---

## Q14 UCSD.edu

1. *X-Content-Type-Options: nosniff* disables MIME sniffing, which ensures the browser to interpret files as it should be so that XSS attack is reduced.
  2. Cloudflare provides data centers in more than 300 cities all over the world to provide fast internet service by putting data centers closer to people; so when users try to load content from a page, the content does not have to travel across the world to get to the users because it is taking a "long" time even with the speed of light. Instead, the data comes from a "local" server that can reach the user instantly.
- 

## Q15 Cookies

1. There are 16 cookies stored on ucsd.edu.
2. There are 4 different domains shown for the cookies on UCSD.
3. 1. .linkedin.com  
2. .myfonts.net  
3. .youtube.com  
4. .ucsd.edu
4. There are 29 cookies stored on sdsu.edu.
5. There are 13 different domains.
6. 1. 8a5e7a58-cf02-4302-a62b-bcc1af878097.rlets.com  
2. .clarity.ms  
3. .simpli.fi  
4. .linkedin.com  
5. .snapchat.com

6. .sc-static.net
7. .sdsu.edu
8. .doubleclick.net
9. [www.sdsu.edu](http://www.sdsu.edu)
10. .youtube.com
11. [www.clarity.ms](http://www.clarity.ms)
12. .google.com
13. 66356343.blob.al.siteimproveanalytics.io