Parametric Analysis of Privacy Compliance Apps

The General Data Protection Regulation (GDPR) passed by the EU in 2018 (Wolfrod, 2019), and the California Consumer Privacy Act, passed in 2018 (CCPA, 2020), changed the experience for web users. Web Site owners are required to notify users how their data is gathered and used, resulting in an intrusive popup or modal. “Accept Cookies” has since become a staple in the web browsing experience.

Since then, numerous apps have come to market to assist content distributors with their obligations under the new Privacy Acts.

This study seeks to find if the addition of these apps to an origin website has a significant impact on the site’s User Experience. The User Experience measurement will be applied by publicly available data via Google’s Chrome Browser and will specifically look at website load times.

According to research by Forbes (Salz, 2019), nearly 70% of consumers say load times affect their experience and willingness to make purchases. While compliance to new Privacy laws should be content distributors’ first concern, compliance should not reduce the User Experience where possible.

This paper shows that significant differences in Privacy Compliance Apps exist for the Onload measurement, but not the First Contentful Paint measurement. Additionally, significant differences in the mean load times between the top 6 Privacy Compliance Apps were found with Cookiebot and Osano having the lowest mean onload time.

# Data Collection and Extraction

Google’s Chrome User Experience Report (ux-report) holds key user experience metrics from users who have opted-in to syncing their browsing history with usage statistics (Chrome UX Report, 2020). This data will be used to determine User Experience load times. The ux-report will be joined with the httparchive database (httparchive.org) to build samples of websites that employ privacy compliance 3rd-party systems.

## Ux-report Elements

### First Paint

First Paint refers to the time until an initial render is performed by the user’s browser. This could include initial steps such as a non-default background color (Chrome UX Report, 2020). According to some research (What Is First Paint (FP) in Modern Performance Metrics?, 2019), this may not be helpful in determining user experience and since this research focuses mainly on 3rd-party apps, it will be limited to when content is able to be consumed.

### First Contentful Paint

This is the first time the user could start consuming page content, according to Chrome UX Report (2020). Unlike first paint, the browser will have rendered a text, image, or non-white canvas.

### Onload

This measures the time until the page and all its dependent resources have finished (Chrome UX Report, 2020).

### Effective Connection Type

This reports connection type such as “3G” or “4G.” For this research, only “4G” connections will be sampled with the intent to reduce connection issues from the data set.

### Device Type

One of three classifications. “Phone” category will be used for this sample, for having the largest amount of data. User Experiences differ between Desktop, Tablet, and Phone, though this research is not meant to distinguish differences between these categories.

### Origin

Refers to the website address. This research will focus only on addresses beginning in ‘https://’ to reduce spam domains.

### Data Structure

First Contentful Paint and Onload are stored in Ux-report as histograms with bins corresponding to the load time in milliseconds. Start, End, and Density variables can be calculated to find the ratio of page request measurements in a certain range of times. For example, a density of .123 for start and end times of 1000 and 1200 indicates that 12.3% of page loads fell between 1000 and 1200 milliseconds. This data can be interpreted as 12.3% of page loads had a mean of 1.1 second. Density variables will sum to 1.0 for all origins. (Chrome UX Report, 2020)

## HttpArchive Elements

### Technologies

While HttpArchive contains tables with a wide range of website information. This study will focus on the technologies table by matching the origin in Ux-report with the url field to filter origins that use technology in the “Cookie Compliance” category.

## Queries via Google BigQuery

Ux-report and HttpArchive are freely available databases stored in Google’s BigQuery and can be accessed with a cloud console, command-line tool, or REST api (*What Is BigQuery?*, 2020).

The dataset for this research will be extracted using the cloud console. The cloud console allows Standard SQL queries to all public datasets. A user can then store the output of queries into a private database and query against it. Output can be downloaded from cloud console in various formats, including CSV (*Quickstart Using the Cloud Console*, 2020).

## Data Sampling

Analysis of the data in tests like ANOVA perform better with equal sample sizes. Having unequal sample sizes affects the assumption of equal variances and may affect Type I error rates (Glenn, 2020). To compensate for unequal counts of the top Privacy Compliance systems (top 6 determined by SQL query), the count of apps is used to adjust the sample size.

# Dataset Selection w/ Z-test

With the choice of analyzing the data from the Ux-report with either First Contenful Paint or Onload (or both), a determination is first made whether there is a significant difference between the population and origins with privacy compliant apps. In this section, a test of the null hypothesis will determine which dataset to continue with.

* H0 Origins with Privacy Compliant apps show no significant difference in load times from the population of origins as determined by the Ux-report.
* H1 Origins with Privacy Compliant apps show a significant impact on load times as determined by the Ux-report.

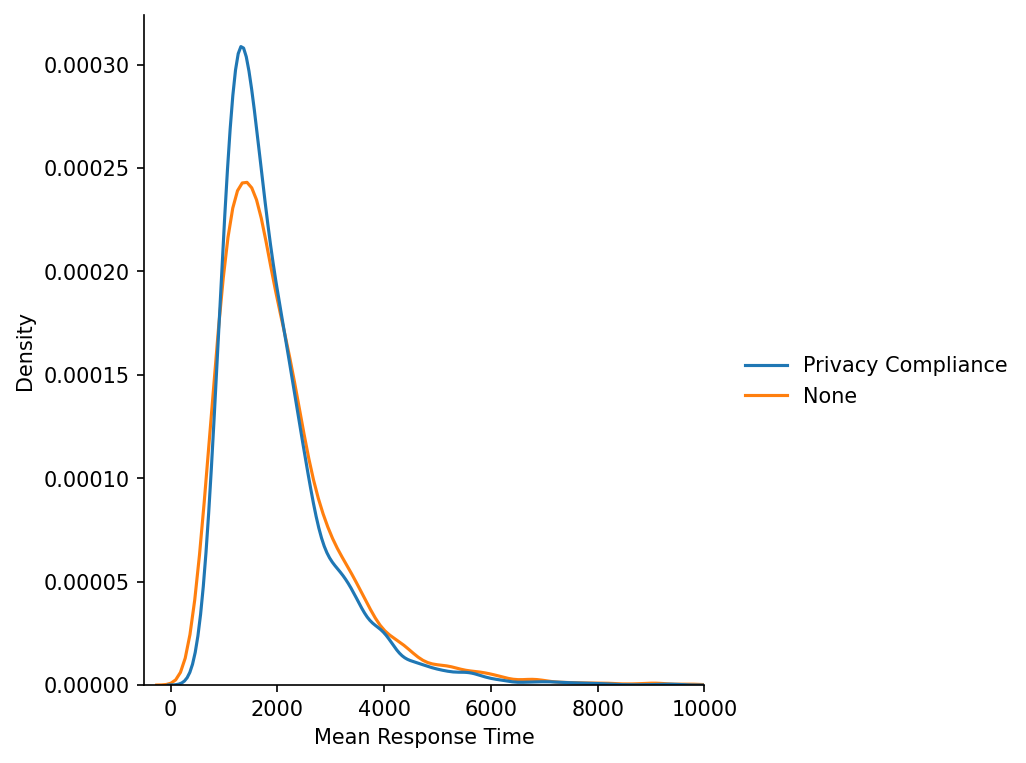


Figure 1 First Contentful Paint Distribution

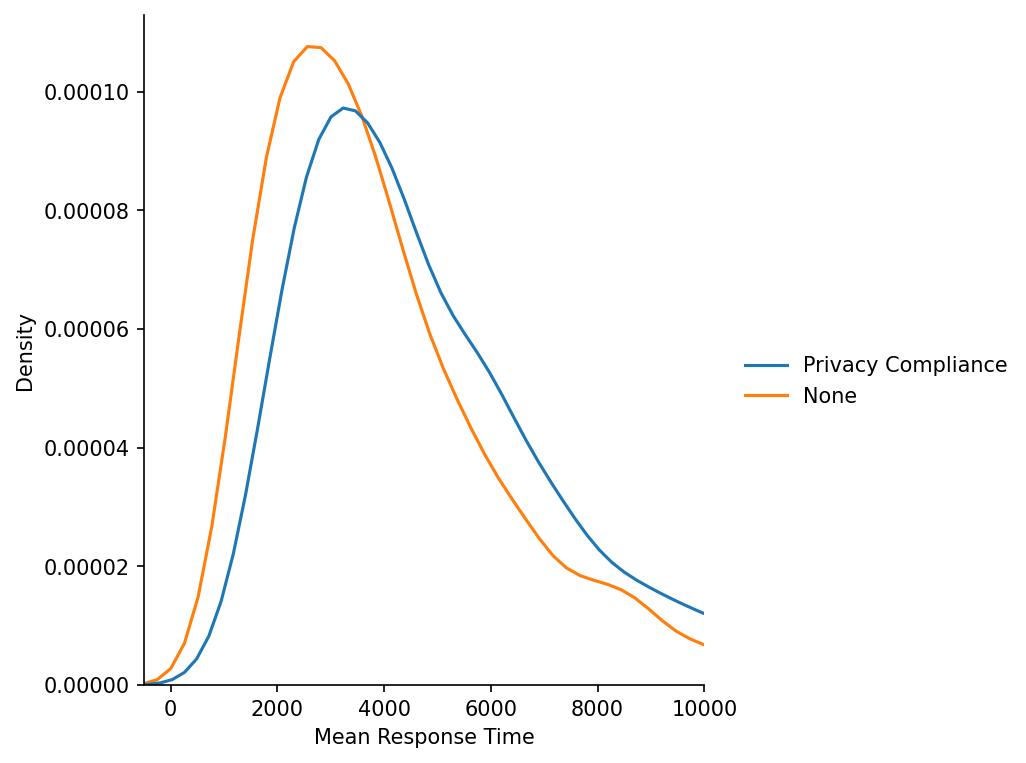


Figure 2 Onload Distribution

A distribution plot for Fcp and Onload mean response times between Privacy Compliant apps and those without suggest rejection of the null hypothesis only for Onload data. To support this conclusion, a Z-test is performed on the samples. A Z-test is a hypothesis test to determine if results are significant and repeatable. A Z-test can be performed on two sample means when the sample size is greater than 30, data points are independent, and data is normally distributed (Glenn, 2020b)

# Analysis of Individual Privacy Compliance Apps

* H0 Privacy App group means show no significant variation.
* H1 At least one group mean is significantly different.

Using the Onload Dataset as determined by the significance tests in the previous section, a sample of 1,200 origins for each of the top 6 Privacy Compliance apps is used for measurement of mean response time between apps. The result shows large differences between the apps.

Didomi 6538.852840

Iubenda 4756.082998

OneTrust 4602.652214

Cookiebot 3606.865404

Quantcast Choice 6837.868487

Osano 4142.180087

Analysis of Variance (ANOVA) will be used to test the null hypothesis. ANOVA compares the means of 2 or more groups using a variance-based F-test to check group mean equality. ANOVA assumes normally distributed data, homogeneity of variances, and independent samples (Bedre, 2018).

Figure 5 shows the distribution of Privacy App Mean Response times. The plot suggests right-skewed data like the previous tested Privacy Complaint Apps group and possible heterogeneity of variance.

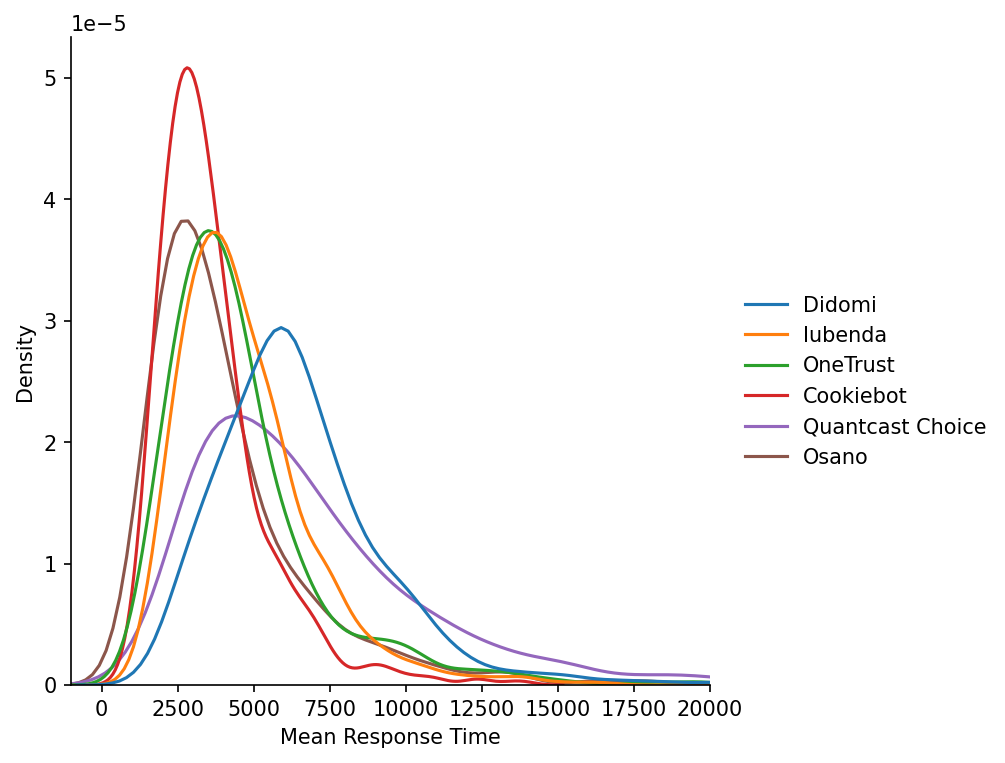


Figure 5 Privacy App Mean Response Distribution

## Outliers

Figure 6 shows how all 6 apps have right skewed data, with outliers ranging as far as 40 second load times. As the Privacy App’s effect on User Experience is the primary consideration of this study, most outliers will remain in the sample. However, since measurement errors could be the result of high load times, a threshold of 20 seconds load time is determined for tests of this sample data (Frost, 2020).

Uxplanet.org (2019) specifies that a 10 second load time increases the probability of a bounce to 123%. A 20 second load time, twice that amount, could be the result of confounding variables. An origin with a 20 second load time may employ a type of technology that affects the onload time in an unmeasurable way by the Chrome browser. While the 20 second threshold is somewhat arbitrary, origins with a load time over 20 seconds are unlikely to be representative of the population.

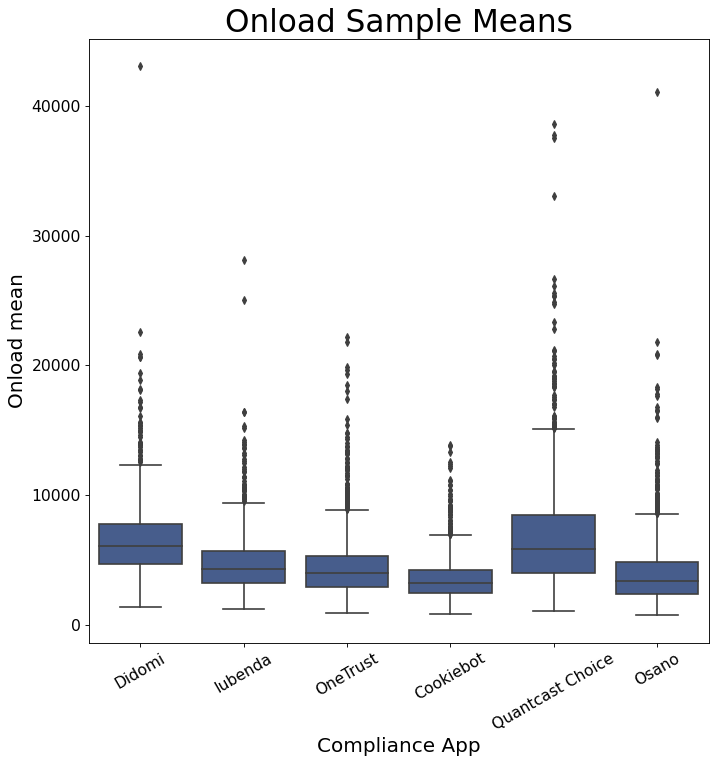


Figure 6 Privacy Compliance Mean Response Box Plot

## App Mean Response Distribution

Applying the log transformation here results in the appearance of a normal distribution in Figure 7.

However, it suggests the assumption of constant variance may not be fulfilled. There appears to be 3 groups within the 6 privacy compliance apps. Applying Levene’s test for unequal variances supports this conclusion with a low P-value. Levene’s test was ran with Python’s Scipy package.

stat, p = levene(stackedLogData['Didomi'],stackedLogData['Cookiebot'], stackedLogData['Quantcast Choice'],

stackedLogData['OneTrust'], stackedLogData['Iubenda'], stackedLogData['Osano'])

Levene's Test: stat: 36.413354354782186 p: 5.793557508427283e-37

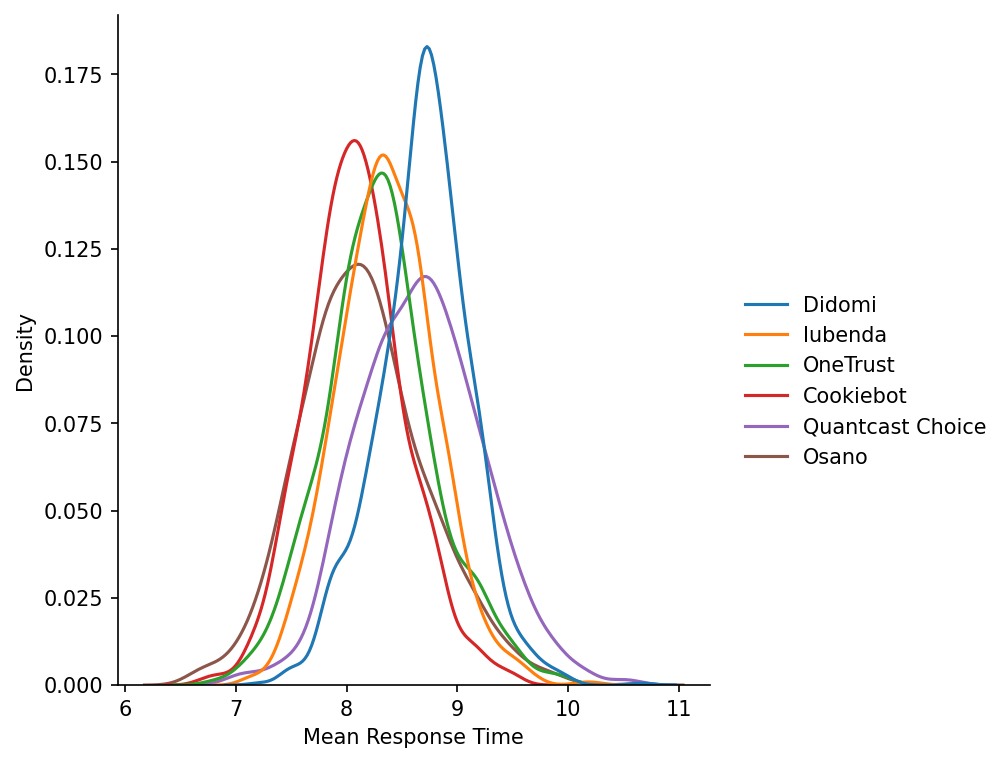


Figure 7 App Mean Response Distribution

Some research suggests a rule of thumb that the ratio between the largest treatment variance and smallest not exceed 3 ( Smax / Smin < 3). For ratios less than 3, the assumption of equal variance is probably satisfied for ANOVA (Ford, 2013).

Using Python’s Numpy package (np.var), the ratio is 1.7 (.33 / .19) and supports continuing with ANOVA.

## ANOVA Results

Using Python’s Scipy package, the differences found in the means of origins using the top 6 Privacy Compliance apps is significant (P-value < .0001)

## Tukey’s Post-Hoc Test

After ANOVA determined a significance level exists between the Privacy Compliance Samples, Tukey’s Range test is performed to determine which apps are significantly different. Tukey uses pairwise comparisons between the means of each sample using a studentized range distribution and can suggest which apps are significant from one another (Lee & Lee, 2018).

Python’s Statsmodels package is used to run the pairwise Tukey.

# Summary

This study of User Experience load times for the top 6 Privacy Compliance apps as measured by Chrome’s UX Report found two apps performed better than their piers when determined by Chrome’s Onload times.

Cookiebot and Osano had mean onload times of 3606ms and 4142ms, respectfully, and were significant results over the closest pier at 4602ms.

Didomi and Quantcast Choice appeared the laggards in the study with onload times over 6500ms.

This study found it was possible to find significant measurements in the Onload times between origins based on the technology choice for Privacy Compliance and could be applied to other technologies.

However, it found no significance in the load times of First Contentful Paint, which may be a more important consideration in the User Experience and is less likely to have confounding variables.

# Recommendations and Limitations

Further studies could be done to determine the cause of the significant differences in Onload times when they are not significant in First Contentful Paint times. Onload times may be less interesting in the User Experience when First Contentful Paint could indicate when a consumer is able to interact with the origin content. FCP and Onload measurements only relate to how the Chrome Browser manages content and is therefor subject to any biases in that data collection. Furthermore, this study does not include any user experience metrics collected directly from end users, such as a survey.

Another course of study could determine why the significant difference exists between the Onload times of the Top 6 apps. Because of the limitations of the measurement used in this study, many reasons besides the app performance itself could result in significantly different mean onload times. For example, if an app’s target market is in a specific region where end users typically had a slower or faster network connection than another app’s target market.

# References

Bedre, R. (2018, October 22). *ANOVA using Python*. Renesh Bedre. https://reneshbedre.github.io/blog/anova.html

*California Consumer Privacy Act (CCPA)*. (2020, July 20). State of California - Department of Justice - Office of the Attorney General. https://oag.ca.gov/privacy/ccpa

Calvano, P. (2018, July 31). *Analyzing 3rd Party Performance via HTTP Archive + CrUX*. Paul Calvano. <https://paulcalvano.com/2018-05-15-analyzing-3rd-party-performance-via-http-archive-crux/>

*Chrome User Experience Report | Chrome UX Report |*. (2020). Google Developers. <https://developers.google.com/web/tools/chrome-user-experience-report>

*Floating Point Arithmetic: Issues and Limitations — Python 3.9.1 documentation*. (2020). Python Docs. https://docs.python.org/3/tutorial/floatingpoint.html

Ford, C. (2013, July 29). *A Rule of Thumb for Unequal Variances | University of Virginia Library Research Data Services + Sciences*. University of Virginia Library. https://data.library.virginia.edu/a-rule-of-thumb-for-unequal-variances/

Frost, J. (2020, June 6). *Guidelines for Removing and Handling Outliers in Data*. Statistics By Jim. https://statisticsbyjim.com/basics/remove-outliers/

Glenn, S. (2020, July 9). *Unequal Sample Sizes*. Statistics How To. https://www.statisticshowto.com/unequal-sample-sizes/

Glenn, S. (2020a, July 9). *Tukey Test / Tukey Procedure / Honest Significant Difference*. Statistics How To. https://www.statisticshowto.com/tukey-test-honest-significant-difference/

Glenn, S. (2020b, December 8). *Z Test: Definition & Two Proportion Z-Test*. Statistics How To. https://www.statisticshowto.com/z-test/

Lee, S., & Lee, D. (2018, October 1). *What is the proper way to apply the multiple comparison test?* PubMed Central (PMC). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6193594/

MARSZAŁKOWSKI, J., MARSZAŁKOWSKI, J. M., & DROZDOWSKI, M. (2014). EMPIRICAL STUDY OF LOAD TIME FACTOR IN SEARCH ENGINE RANKING. *Journal of Web Engineering*, *14*, 1. <https://www.riverpublishers.com/journal/journal_articles/RP_Journal_1540-9589_1316.pdf>

Nielson, J. (1997). *The Need for Speed*. Nielsen Norman Group. <https://www.nngroup.com/articles/the-need-for-speed-1997/>

Nielson, J. (2010, June 20). *Website Response Times*. Nielsen Norman Group. <https://www.nngroup.com/articles/website-response-times/>

*Quickstart using the Cloud Console | BigQuery |*. (2020). Google Cloud. https://cloud.google.com/bigquery/docs/quickstarts/quickstart-web-ui

*R vs Python: For Statistics and Data Science in 2020. (2020). Code Conquest.* [*https://www.codeconquest.com/versus/r-vs-python-for-statistics-and-data-science-in-2020/*](https://www.codeconquest.com/versus/r-vs-python-for-statistics-and-data-science-in-2020/)

Radečić, D. (2020, January 4). *Top 3 Methods for Handling Skewed Data - Towards Data Science*. Medium. <https://towardsdatascience.com/top-3-methods-for-handling-skewed-data-1334e0debf45>

Salz, P. A. (2019, August 24). *How Fast Is Fast Enough? Mobile Load Times Drive Customer Experience And Impact Sales*. Forbes. https://www.forbes.com/sites/peggyannesalz/2019/08/22/how-fast-is-fast-enough-mobile-load-times-drive-customer-experience-and-impact-sales/?sh=3cd58a0eef31

uxplanet.org. (2019, December 29). *How page speed affects Web User Experience - UX Planet*. Medium. https://uxplanet.org/how-page-speed-affects-web-user-experience-83b6d6b1d7d7

Wolford, B. (2019, February 13). *What is GDPR, the EU’s new data protection law?* GDPR.Eu. <https://gdpr.eu/what-is-gdpr/>

*What is BigQuery? |*. (2020). Google Cloud. <https://cloud.google.com/bigquery/docs/introduction>

*What is First Paint (FP) in Modern Performance Metrics?* (2019, February 1). LoadFocus. <https://loadfocus.com/blog/2019/02/what-is-first-paint-fp-in-modern-performance-metrics/>