



FOM Hochschule für Oekonomie und Management

Hochschulzentrum Köln

Wirtschaftsinformatik

Usefulness of open-source tools for web analytics in E- Marketing

E-Marketing, Yvonne Romes, SS 20

Christian Frank (#473088)

July 9, 2020

ChilledCow: lofi hip hop radio - beats to relax/study to



This work is licensed under the Creative Commons Attribution 4.0 International License.

To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or
send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Table of contents

Table of contents	II
List of Figures	III
List of Abbreviations	IV
1 Introduction to Open-Source Web Analytics in E-Marketing	5
1.1 Pronouns	5
1.2 Web-Analytics	5
2 What is Open-Source and why is it important?	6
3 Open-Source Web Analytics tools	8
3.1 Matomo	8
3.2 Open Web Analytics	8
3.3 Plausible	8
4 Three questions to ask about our page visitors	10
4.1 Who are the users and how do they behave?	10
4.2 Which pages do they access?	12
4.3 Where do they come from?	13
5 Summary and recommendations	17
References	18

List of Figures

- Figure 1 - Weekly report 10
- Figure 2 - Devices by browser 11
- Figure 3 - Devices by OS 11
- Figure 4 - Devices by size..... 12
- Figure 5 - Top pages 13
- Figure 6 - Access by country 14
- Figure 7 - Top countries..... 15
- Figure 8 - Top referrers..... 15

List of Abbreviations

4P+	Product, Price, Place, Promotion, Participation
CPC	Cost-Per-Click
CRM	Customer Relationship Management
CTR	Click-Through-Rate
IT	Information Technology
OSOR	Open Source Observatory
SEA	Search Engine Advertising
SEO	Search Engine Optimization
SERP	Search Engine Results Page
SMM	Social Media Marketing
UCG	User-Generated Content

1 Introduction to Open-Source Web Analytics in E-Marketing

1.1 Pronouns

As we move towards a more gender-fluid society, it is time to rethink the usage of gendered pronouns in scientific texts. Two quite well-known professors from UCLA, Abigail C. Saguy, and Juliet A. Williams argue that it makes a lot of sense to use singular they/them instead: "The universal singular they is inclusive of people who identify as male, female or nonbinary."¹ Throughout this text, I will attempt to follow that suggestion, to support gender inclusivity through gender neutral language. Or, as Jane Roper writes on WBUR, "They Is Here to Stay. Get Over It"²

1.2 Web-Analytics

We cannot talk about Web Analytics without mentioning Google. Google is the predominant search engine of the Internet and drives most, if not all, traffic to websites. A complete industry has sprung up around Google search, from SEO (Search Engine Optimization) all the way to SEA (Search Engine Advertising). However, Google is not the subject of this essay. To understand the performance of a web site, one needs to understand its visitors, where they come from and how they behave on the web site. This process is called Web Analytics, and, in this paper, we want to have a look at a couple of open-source tools to perform this analysis. [Google Analytics](#) is a powerful tool, too; comparing it with the tools that we will be covering here could be an interesting subject but would greatly exceed the scope of this essay.

Before looking at the selected tools, we need look at open-source first, to explain why we have narrowed the selection of tools down to include only open-source tools.

¹ Saguy, Abigail and Williams, Juliet (Scientific American, 2019) Why we should all use they/them pronouns

² Roper, Jane (WBUR, 2019): They Is Here to Stay. Get Over It.

2 What is Open-Source and why is it important?

Let us look at an explanation of Open-Source: “The term open source refers to something people can modify and share because its design is publicly accessible.”³

Open source originated in the realm of software development and meant that the source code for a given tool was made available publicly, allowing everybody to read and access it. Also, having the source code publicly available meant that everybody was able to modify and re-use it. There are no hidden algorithms or backdoors in open source. Open source as a concept has meanwhile spread from software development into many more aspects of everyday life, from Academia (Open Access) to the Arts & Culture (Creative Commons) and Government (Open Data). Cornerstones of the open source movement are transparency, open exchange, and collaborative participation.⁴

One key reason to open-source something, a piece of code, a piece of art or a publication is the desire to share it with a wider audience and encourage participation.⁵ Unlike physical media, digital works do not diminish or lose quality when they are shared, quite the opposite, they grow and improve. Something legacy publishers still fail to understand (“Article 13”), but that is a different discussion.

In our line of work, in IT, and especially in technology startups, the use and creation of open source tools is already well established.⁶

On a societal level, the European Union fully endorses the use of open source software⁷ and has created an Open Source Observatory to foster and encourage adoption of open source software and principles across the member states and organizations.⁸

³ *Red Hat* (opensource.com, 2019): What is open source?

⁴ See *Red Hat* (opensource.com, 2019): What is open source?

⁵ See *Google Open Source* (Google, 2020): Why Open Source?

⁶ See *Balter, Ben* (opensource.com, 2019): 6 motivations for consuming or publishing open source software

⁷ See *European Pirate Party* (EPP, 2020): European Parliament strongly recommends

⁸ See *European Union* (EU, 2020): Open Source Observatory

When selecting a tool for web analytics in E-Marketing, choosing an open source tool is thus the somewhat obvious choice. However, using open source tools in marketing is not the same as open source marketing.⁹ Open source marketing is new concept of applying open source principles to the marketing process and offer an alternative to established marketing models and strategies.¹⁰

In this paper, we will not cover open source marketing itself but focus on the use of open source tools in (e-)marketing. Let us continue with a look at the selected tools.

⁹ See *Hunt, Ben* (OS Marketing, 2015): About Open Source Marketing

¹⁰ See *Reilly, Kenneth* (The Startup, 2019): What is Open-Source Marketing?

3 Open-Source Web Analytics tools

3.1 Matomo

Matomo¹¹, formerly known as Piwik, is the oldest tool of the selection and already well established. It claims millions of web sites as users and comes in a hosted and an on-premise version. Matomo is completely open source and free to use. A main selling point is data sovereignty – all collected web site and user data remains within the tool and is not shared with anybody else or used for any other purpose, setting it apart from Google Analytics and Google Analytics Premium; it also claims feature parity with the Google offerings.

You can find all available Matomo features on the [product feature](#) page.

3.2 Open Web Analytics

Open Web Analytics¹² is a bit newer than Matomo and has a much smaller installed base, it only claims thousands of web sites and only offers on-premise installation. Like Matomo, there is no sharing of data and full GDPR compliance. The main targets are CRM systems and Open Web Analytics comes with ready-to-use integrations for WordPress and Mediawiki.

You can have a look at the available features on the [product demo](#) page.

3.3 Plausible

Plausible¹³ is the newest kid on the block, started in 2019 by Uku Täht and Marko Saric. Of the three tools it is the most lightweight, with a single line of JavaScript code for analytics; in addition to being fully GDPR compliant, it also works without placing cookies. Plausible is, for the time being, only available as a hosted version. This is also how the company makes money, by selling the service. The product itself is fully open

¹¹ See *Matomo* (Matomo, 2020): Ethical Analytics, Powerful Insights

¹² See *Adams, Peter* (Open Web Analytics, 2020): Web Analytics. Open Sourced.

¹³ See *Täht, Utu* (Plausible, 2020): Simple and privacy-friendly alternative to Google Analytics

source and available on [GitHub](#), though; all algorithms are freely available and there's a public [roadmap](#).

There's also a demo available on the [product demo](#) page, and detailed comparisons with [Google Analytics](#) and [Matomo](#).

To support startups and fresh ideas, we will continue with Plausible as the tool of choice for the rest of this essay and continue with a list of questions we want to analyze.

As a basis for the data we will use analytics of my sed card, www.chfrank.net

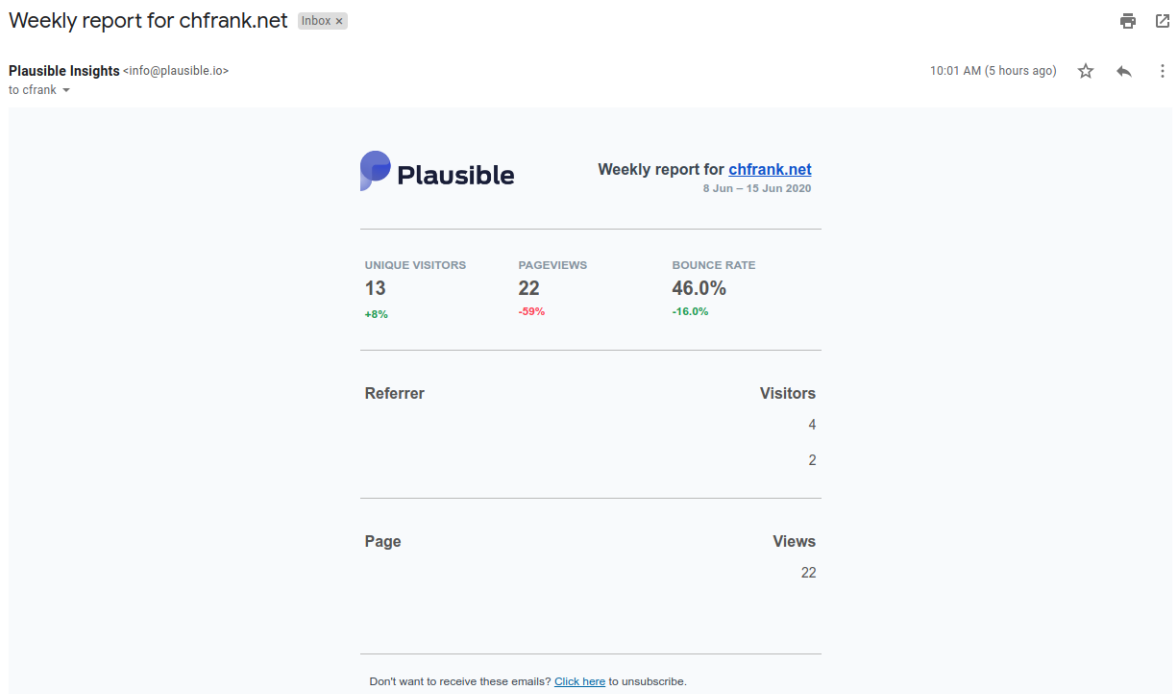
4 Three questions to ask about our page visitors

4.1 Who are the users and how do they behave?

The first question we want to ask is, who are our users and which devices do they use to access our page?

Let us start with a look at the weekly report:

Figure 1 - Weekly report



In the weekly report we can see that we had thirteen unique visitors this week, with a total number of twenty-two pageviews. More than half of our visitors stayed for a bit to read the information provided.

Which browser do our visitors use to access the page?

Figure 2 - Devices by browser

Devices		Size	Browser	OS
Browser		Visitors		
Chrome				43%
UC Browser				27%
Firefox				23%
Firefox Mobile				3%
Safari				3%

As expected, most of our visitors use Google's Chrome. Second place is the UC Browser, developed by Alibaba's UCWeb, tied with Firefox. While we can safely ignore our Mac visitors for now, we should make sure that our web page displays properly on Chrome, Firefox, and UC Browser.

Let us check the devices our visitors use:

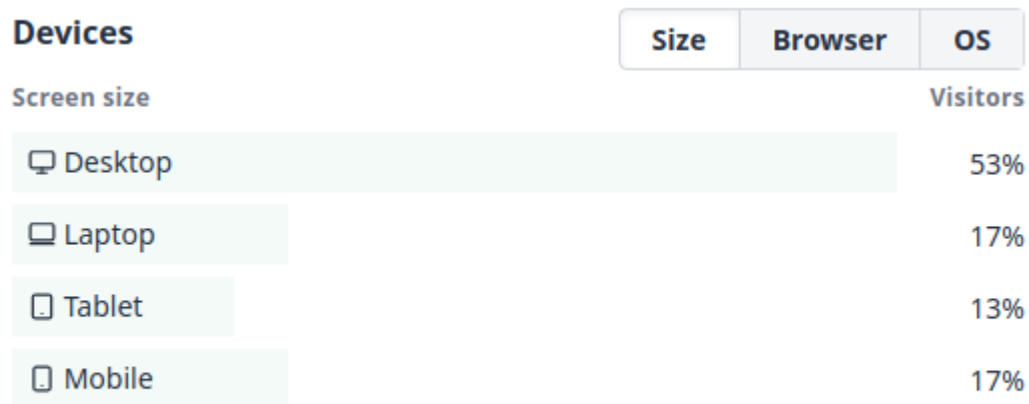
Figure 3 - Devices by OS

Devices		Size	Browser	OS
Operating system		Visitors		
Windows				40%
Android				40%
GNU/Linux				10%
Ubuntu				7%
Mac				3%

Android and Windows devices make out the bulk of our visitor's devices, trailed by Linux, which is expected for IT content. As before, Mac visitors are almost nonexistent. But we can already see that mobile access is particularly important. If our web site were not yet optimized for mobile viewing (it is), now would be a good time to change that.

The screen size should support or findings:

Figure 4 - Devices by size



And indeed, almost half of our visitors use mobile devices (including laptops).

4.2 Which pages do they access?

The second question we want to ask is, which pages do our visitors access. As this is a small page, there is only so much data to be gathered:

Figure 5 - Top pages

Top pages

Page url	Pageviews	Bounce rate
/	79	65%
/page1.html	8	-
/index.html	1	-

We can see that most people only access the main page, either through “/” or directly through its canonical name “index.html”. Astonishing 8 views go to the mandatory legal information (“Impressum”) – as nobody ever reads these, this access will most likely be from a crawler or indexing bot.

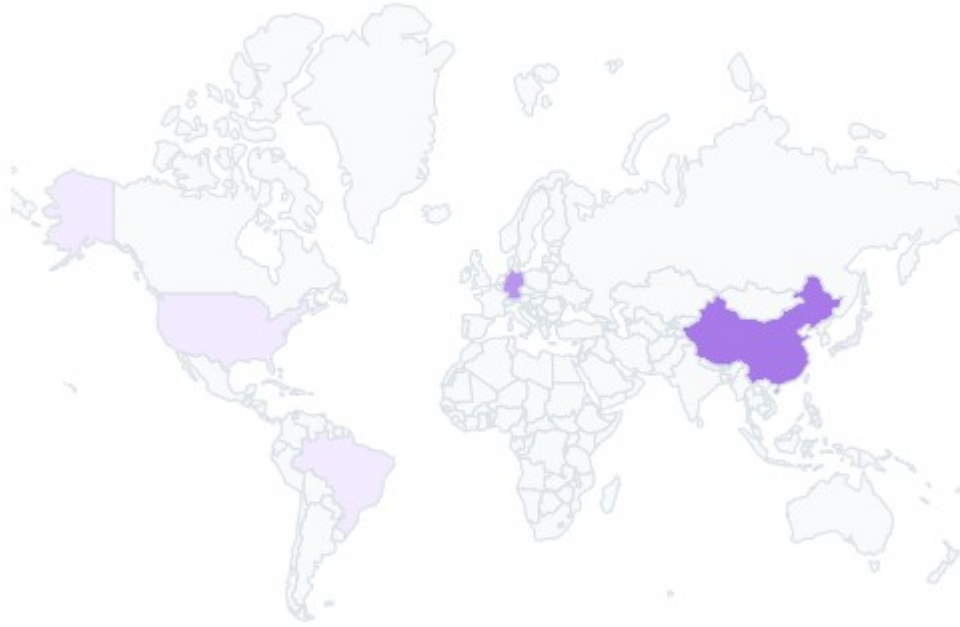
If the web site were more than just a comp card we could now see in detail, which pages receive the most visitors, and on which pages the users would stay on and interact with. This information would allow us to discern which information is useful for our visitors and enable us to adapt our content to better match their needs.

4.3 Where do they come from?

And lastly, we want to find out where our users come from. If we look at the overview provided, it seems that most of our users come from China:

Figure 6 - Access by country

Countries



[MORE](#)

From the browser information we already assumed that much, and if we look at the figures in more detail, we can verify that assumption:

Figure 7 - Top countries






Top countries

Country	Visitors
China	47%
Germany	17%
Switzerland	10%
Denmark	7%
Egypt	3%
United States	3%
Tunisia	3%
Albania	3%
Japan	3%
Brazil	3%

Almost half of all our visitors come from China. Let us check the referrers, do they corroborate this result too?

Figure 8 - Top referrers

Top Referrers

Referrer	Visitors	Bounce rate
 baidu.com	8	70%
 sogou.com	2	100%
 Google	1	0%
 anti-crisis-seo.com	1	100%
 chfrank-cgn.github.io	1	100%

Yes indeed, the top referrer is Baidu, the Chinese search giant. And some visitors from China even stayed on the page, so it must contain some information that is deemed useful there. Interesting is the referral from my Cloud Wiki, which I reference on my comp card, but which is not indexed by itself, as far as I know.

Using web analytics gave us some interesting insights into the users of our page, without using any tracking cookies and without storing personally identifiable data.

One potential overall outcome could be to think about providing a Chinese version of the site, or at least include Chinese contact information.

5 Summary and recommendations

If you are concerned with data usage and the need for privacy, using an open source tool for your analytics needs is somewhat mandatory. There are a couple of great tools available and we had look at Plausible Analytics in a bit more detail.

Depending on the questions that you have regarding your web site, using a small and agile tool with fast load times makes a lot of sense. Bigger tools can deliver more data, but do you really need the answers the additional data can provide? Smaller tools are also better for the environment, as they consume less computing power.

In our example, we were able to answer all our theoretical marketer's question from the data collected by Plausible.

The big advantage of all three open source tools over Google Analytics (or Bing Webmaster Tools) is that there are no hidden algorithms – everything is out in the open and there's no need for guesswork or reverse engineering to make the website perform better.

Also, there are no hidden costs – you either pay for the hosting or provide on-premise compute capacity, there is no need for your customer's data to be sold or be repurposed for advertising.

This is by no means Google (or Microsoft) bashing – I extensively use the Google suite of tools and am aware what happens with my data. There is a difference though, between giving away your own data and making the users of your website give away their data.

Happy analytics!

References

Adams, Peter (Open Web Analytics, 2020): Web Analytics. Open Sourced. <<http://www.openwebanalytics.com/>> (2020) [Access: 9.6.2020]

Balter, Ben (opensource.com, 2019): 6 motivations for consuming or publishing open source software <<https://opensource.com/life/15/12/why-open-source>> (2015-12-09) [Access: 9.6.2020]

Buffer (Buffer, 2020): An Incomplete Guide to Inclusive Language for Startups and Tech <<https://open.buffer.com/inclusive-language-tech/>> (2020) [Access: 9.6.2020]

European Pirate Party (EPP, 2020): European Parliament strongly recommends <<https://european-pirateparty.eu/european-parliament-strongly-recommends-any-software-developed-by-and-for-the-eu-institutions-to-be-made-publicly-available-under-free-and-open-source-software-licence/>> (2020-05-15) [Access: 9.6.2020]

European Union (EU, 2020): Open Source Observatory <<https://joinup.ec.europa.eu/collection/open-source-observatory-osor>> (2020-06-04) [Access: 9.6.2020]

Google Open Source (Google, 2020): Why Open Source? <<https://opensource.google/docs/why/>> (2020) [Access: 9.6.2020]

Hunt, Ben (OS Marketing, 2015): About Open Source Marketing <<http://opensourcemarketingproject.org/>> (2015-07-31) [Access: 9.6.2020]

Matomo (Matomo, 2020): Ethical Analytics, Powerful Insights <<https://matomo.org/>> (2020) [Access: 9.6.2020]

Red Hat (opensource.com, 2019): What is open source? <<https://opensource.com/resources/what-open-source>> (2019) [Access: 9.6.2020]

Reilly, Kenneth (The Startup, 2019): What is Open-Source Marketing? <<https://medium.com/swlh/what-is-open-source-marketing-b1a2f81bd4f3>> (2019-08-22) [Access: 9.6.2020]

Roper, Jane (WBUR, 2019): They Is Here to Stay. Get Over It. <<https://www.wbur.org/cognoscenti/2019/08/13/gender-pronouns-jane-roper>> (2019-08-13) [Access: 9.6.2020]

Saguy, Abigail and Williams, Juliet (Scientific American, 2019) Why we should all use they/them pronouns <<https://blogs.scientificamerican.com/voices/why-we-should-all-use-they-them-pronouns/>> (2019-04-11) [Access: 25.5.2020]

Täht, Utu (Plausible, 2020): Simple and privacy-friendly alternative to Google Analytics
<<https://plausible.io/>> (2020) [Access: 9.6.2020]