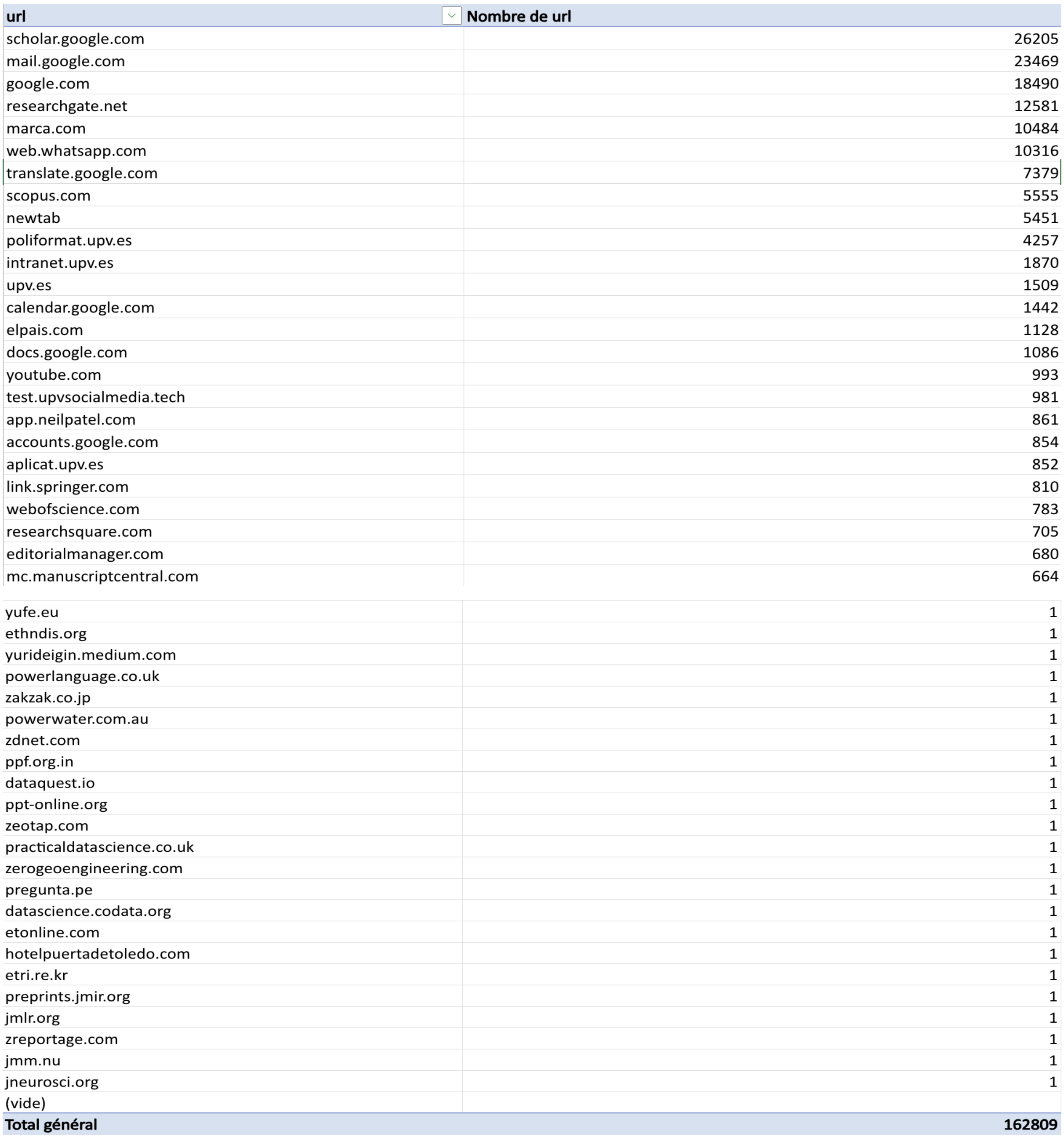
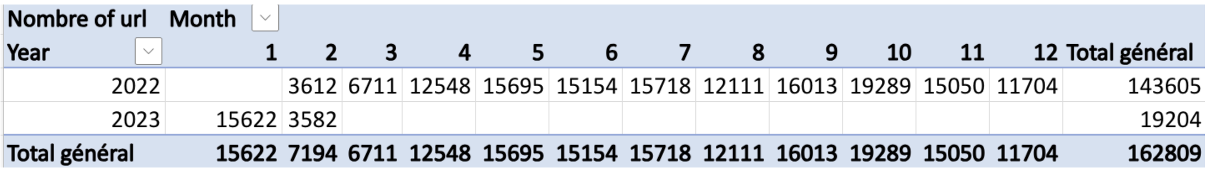
## a) Basic data :



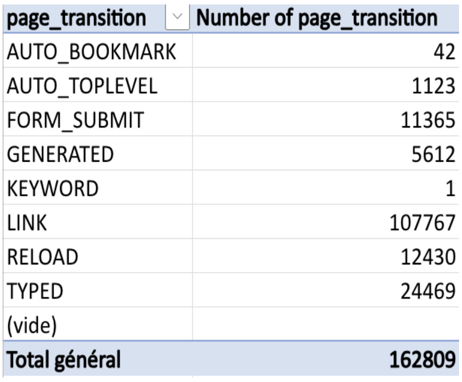
| url | Nombre de url |
|---|---|
| scholar.google.com | 26205 |
| mail.google.com | 23469 |
| google.com | 18490 |
| researchgate.net | 12581 |
| marca.com | 10484 |
| web.whatsapp.com | 10316 |
| translate.google.com | 7379 |
| scopus.com | 5555 |
| newtab | 5451 |
| poliformat.upv.es | 4257 |
| intranet.upv.es | 1870 |
| upv.es | 1509 |
| calendar.google.com | 1442 |
| elpais.com | 1128 |
| docs.google.com | 1086 |
| youtube.com | 993 |
| test.upvsocialmedia.tech | 981 |
| app.neilpatel.com | 861 |
| accounts.google.com | 854 |
| aplicat.upv.es | 852 |
| link.springer.com | 810 |
| webofscience.com | 783 |
| researchsquare.com | 705 |
| editorialmanager.com | 680 |
| mc.manuscriptcentral.com | 664 |
| yufe.eu | 1 |
| ethndis.org | 1 |
| yurideigin.medium.com | 1 |
| powerlanguage.co.uk | 1 |
| zakzak.co.jp | 1 |
| powerwater.com.au | 1 |
| zdnet.com | 1 |
| ppf.org.in | 1 |
| dataquest.io | 1 |
| ppt-online.org | 1 |
| zeotap.com | 1 |
| practicaldatascience.co.uk | 1 |
| zerogeoengineering.com | 1 |
| pregunta.pe | 1 |
| datascience.codata.org | 1 |
| etonline.com | 1 |
| hotelpuertadetoledo.com | 1 |
| etri.re.kr | 1 |
| preprints.jmir.org | 1 |
| jmlr.org | 1 |
| zreportage.com | 1 |
| jmm.nu | 1 |
| jneurosci.org | 1 |
| (vide) | |
| **Total général** | **162809** |

Since this first table is too long, we will analyze only the most important lines but for the rest, we will analyze all the data. It is noticeable that sites like scholar.google.com, mail.google.com and google.com are in the lead with more than 18,000 to 26,000 visits. This indicates a high activity related to academic research, emails and general searches, while sites like researchgate.net, whatsapp.com, and scopus.com show activities related to work, communication and scientific research. On the other hand, we see sites with one-off searches or occasional visits like jmm.nu or jmlr.org.

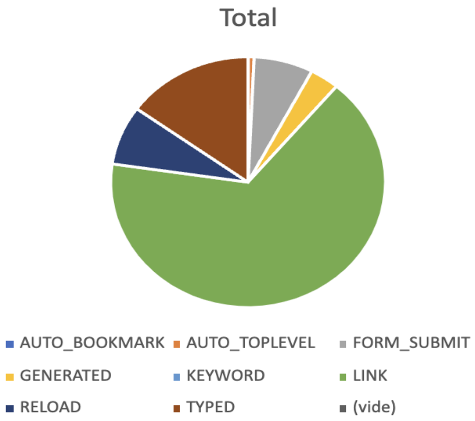


| Nombre of url | Month | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | Total général |
| 2022 | | 3612 | 6711 | 12548 | 15695 | 15154 | 15718 | 12111 | 16013 | 19289 | 15050 | 11704 | 143605 |
| 2023 | 15622 | 3582 | | | | | | | | | | | 19204 |
| **Total général** | **15622** | **7194** | **6711** | **12548** | **15695** | **15154** | **15718** | **12111** | **16013** | **19289** | **15050** | **11704** | **162809** |

From the table above, we can compare the two years 2022 and 2023, we notice that in 2022, the total visits are 143,605, covering the whole year, this number is much lower than that of 202 or only the data of January and February are available, with 19,204 visits in total. In addition, we notice that October has the highest number of visits (19,289), followed by September (16,013) and July (15,718) and January and December are the least active (3,612 and 11,704 visits respectively). I think it would have been interesting to see if the trend continues for the following months in 2023



| page_transition | Number of page_transition |
|---|---|
| AUTO_BOOKMARK | 42 |
| AUTO_TOPLEVEL | 1123 |
| FORM_SUBMIT | 11365 |
| GENERATED | 5612 |
| KEYWORD | 1 |
| LINK | 107767 |
| RELOAD | 12430 |
| TYPED | 24469 |
| (vide) | |
| **Total général** | **162809** |



The majority of visits come from link clicks (107,767), followed by direct URLs typed in (24,469) and page reloads (12,430). Form submissions (11,365) and auto-generated pages (5,612) are also notable. Automatic tab openings and bookmarking are rare. This distribution shows a link-dominated navigation, with a significant share of direct visits to familiar sites and frequent use of reloads, indicating repeat visits to certain pages.

## b) Practical questions :

**Are search histories "unique" to define a user?**
Search histories can reflect a user's preferences and habits, but they are not completely unique. They can be biased by shared device usage, browsing in private mode, or casual searches. To better define a user, cross-reference this data with other online behaviors.

**Are web search histories and “interest profiles” stable over the time? In other words, are these data steady behavioural fingerprints?**

No, search histories and interest profiles are not completely stable. They evolve over time based on changes in the user's situation, interests or specific needs. They are therefore dynamic behavioral footprints.

**Can we trust on these data to depict users’ profiles?**

Search data can be useful for creating user profiles, but it is not completely reliable. It can be biased by casual searches, temporary behaviors, or shared device usage. Additional analysis is therefore necessary for an accurate representation.

## c) Critical comment :

Access to search and browsing data stored by Google has major implications in terms of security and privacy. As a data scientist, this data can be very useful to personalize services and improve user experience. However, as a citizen, I feel monitored especially when I have ads that target my interests and there is also the risk of my data being leaked.

I find that the fact that third parties have access to data is both good and bad. Bad because companies can profile users and influence their decisions (targeted advertising, manipulation of opinions) and if this data is compromised, it can be exploited by cybercriminals. Good because Google for example can give you more relevant results and then it saves time.

Concerning sensitive data to be deleted include private searches, such as those related to health, finances or personal relationships. Likewise, location data, which can be used to trace my movements. Finally, browsing history related to non-representative or temporary searches should also be deleted I think. Finally, what matters most to me is the transparency of data controls to preserve user trust and offer them a secure framework.

**Documentation :**

[https://www.eff.org/issues/online-tracking](https://www.eff.org/issues/online-tracking)

[https://arxiv.org/abs/1802.08232](https://arxiv.org/abs/1802.08232)

[https://es.wikipedia.org/wiki/Wikipedia:Portada](https://es.wikipedia.org/wiki/Wikipedia:Portada)

**Marwa Chiguer**