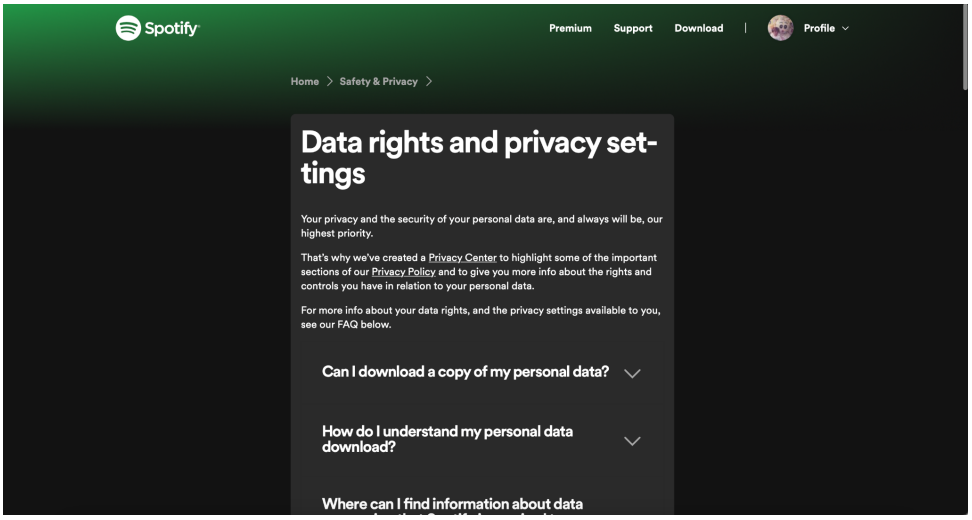
Data visualization - Process Book

Visualizing Data Ownership from Spotify

Jacopo Ferro, Mark Mouawad, Marcel Mauricio Moran Calderon

1- Introduction: Data rights

In May 2018, the European Union introduced the General Data Protection Regulation (GDPR), establishing a comprehensive framework to safeguard individuals rights and provide greater control over their personal information. It emerged as a response to growing concerns about data privacy and the need for a unified legal framework across the EU. It is known as one of the first attempts of modern data regulation in the world.



One of the main articles of the GDPR is article 15, which grants individuals the right to access their personal data held by organizations. This right empowers individuals to request and receive a copy of their data, enabling them to understand how their information is used and take control of their privacy. Article 15 serves as a cornerstone of transparency and user empowerment, ensuring that individuals have the means to explore the depths of their personal data and make informed decisions about its handling.

Since the enactment of GDPR, in order to comply with the EU regulation, every tech company offers tools so their users can understand their data and how its being processed, as stated in article 15. Users are able to receive their data in seconds or sometimes it may take as long as a week, especially if the amount of data is important. Google for example has Google Takeout, a service where users can choose which data from Google products they want to download. In our case, we chose to investigate Spotify's data processing as shown here.

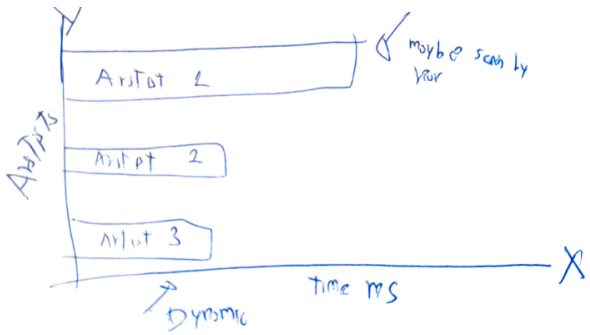
166.9 MB = Size of my personal spotify data

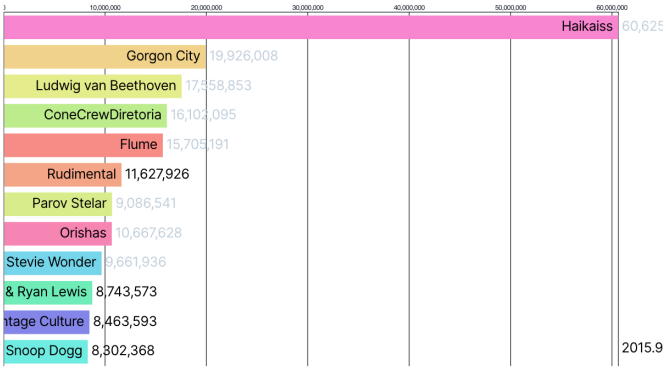
2- Our approach: the path to the final result

In the beginning, each member had the intention to work with different datasets, from personal Google datasets to medical ones. However, While exploring our personal information from Spotify, we realized how sensible and interesting the information could be, that is why we decided not only to develop fun visualization of people's data but to make users aware of the information that could be deduced with their data (without instilling fear or paranoia).

Once we decided on the topic, each of us started to think about what ideas could be brought for milestone one. At that point, we looked at how we could connect the Spotify data with the user's location to give multidimensional visualization for the user. In addition, we look at extra music features that could be extracted by using external API, so that we could analyze the sentiment of a person.

Once we received approval for the idea of our project, we started to develop milestone two. One of our teammates got the full dataset of Spotify (the dataset took 30 days to be sent) and with this dataset, we sketched some graphs that then we tried to prototype and implement on the website.



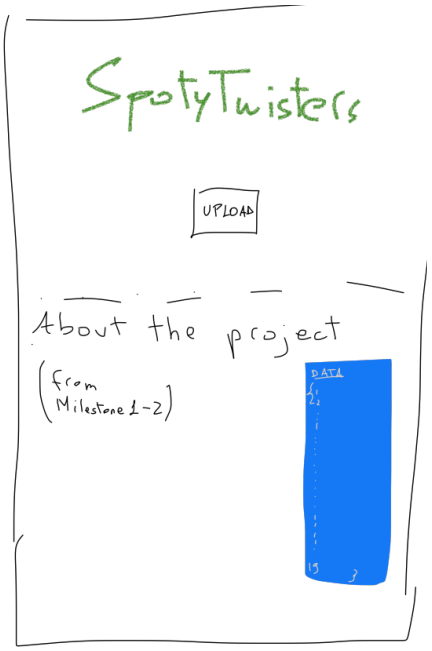


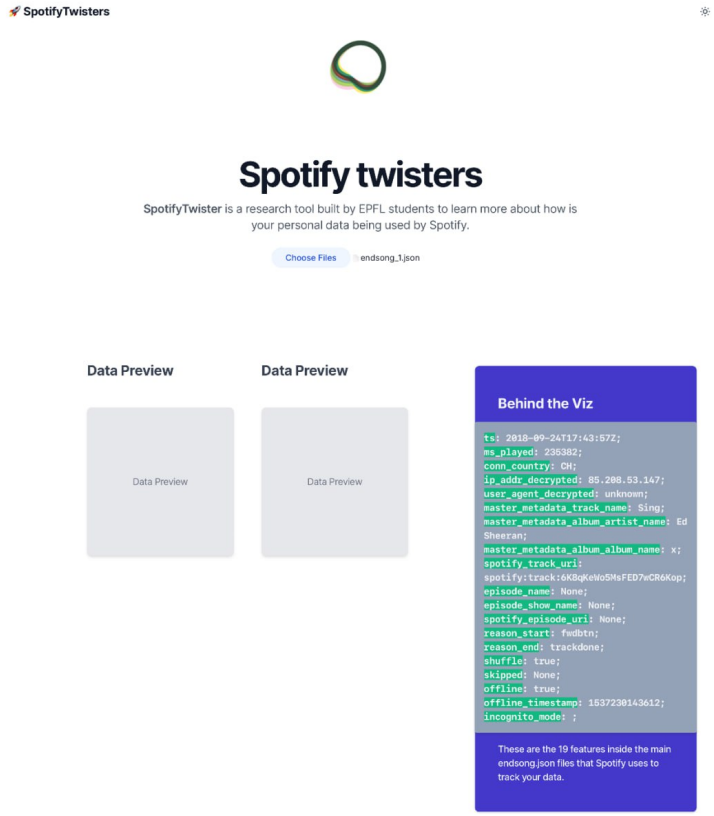
Upon receiving feedback and evaluating the time constraints for the project, we decided to go for the MVP shown during milestone two.

3- The design:

We intended to create a data story that embraced a playful approach. Our goal was for users to not only learn about data rights but also visualize their Spotify data using our tool. Additionally, we aimed to make the website interactive, allowing users to utilize their personal information from Spotify and engage with our tool. Therefore, we faced the challenge of providing file uploads for users to accomplish this.

We had two options to go with: use file uploads with a web server, which would mean we would have to probably keep their files or at least, read them in our servers and then do whatever would need to be done regarding data processing. The other option was to do it fully locally, so the users' files wouldn't even have to leave their computers and all data processing and rendering would be done in their browsers. We decided to go with the latter, as the former option was very cumbersome regarding development as it would mean having to maintain a dedicated web server. In addition, it would be very invasive to ask users to send us their personal Spotify's data and kind of paradoxical, as our project stands for data privacy.





Initially, our intention was to leverage Spotify's API in order to have access to audio features and information of the tracks that were played by the users. But, with our design choice of going fully client-side and offering every user visualizations without a static dataset, we

couldn't afford to send thousands of requests to the Spotify's API every time a user would use our system.

During the development of the website, we realized that the goal of our project wasn't only about the visualizations, but more about how the Spotify data could be leveraged to create those visualizations. With this idea in mind, we chose to add small boxes of text called "Behind the viz" next to the visualizations. It displays which fields from the data were used and how we processed the data to obtain them.

The **Visited Cities** visualization for example uses the **decrypted_ip_addr** field from the data, the **ts** field (date) and the **country** field to show cities the user has gone to. In order to get the location associated with the IP address, we use this API from ip-api.com. However, their system is throttled to 40 requests per minute, so we had to choose wisely which rows of the data we would send to their api. We have decided to group by country the rows, and then take the first rows sorted by date. This way we are sure of different cities, as the user has switched countries. However we are limited to spotting different cities in different countries. We didn't not conceive of a way of spotting rows of different cities in the same country.

4- The stack

We knew we wanted to leverage the power of D3.js but also have a more user friendly framework than plain HTML, CSS and Javascript. We opted to use Astro, which is a web framework that enabled us to modularize our code with the usage of Astro Components and to use Tailwind, which is a CSS library for bootstrapping css classes.

5 - Peer assessment

All the participants of the team contributed equally to the project, please find the breakdown of all the contributions of team members:

Jacopo Ferro

- Process book
- Map Visualization
- Research and generation of dataset
- Screencast
- Deployment of website
- Layout of the webpage

Mark Mouawad

- Proces book
- Spider Visualization
- Map Visualization
- Bar chart visualization
- Screencast
- Uploading and merging functionality for raw json files

Marcel Mauricio Moran Calderon

- Proces book
- Word cloud
- Sketches for visualizations
- Screencast
- Top artist visualization