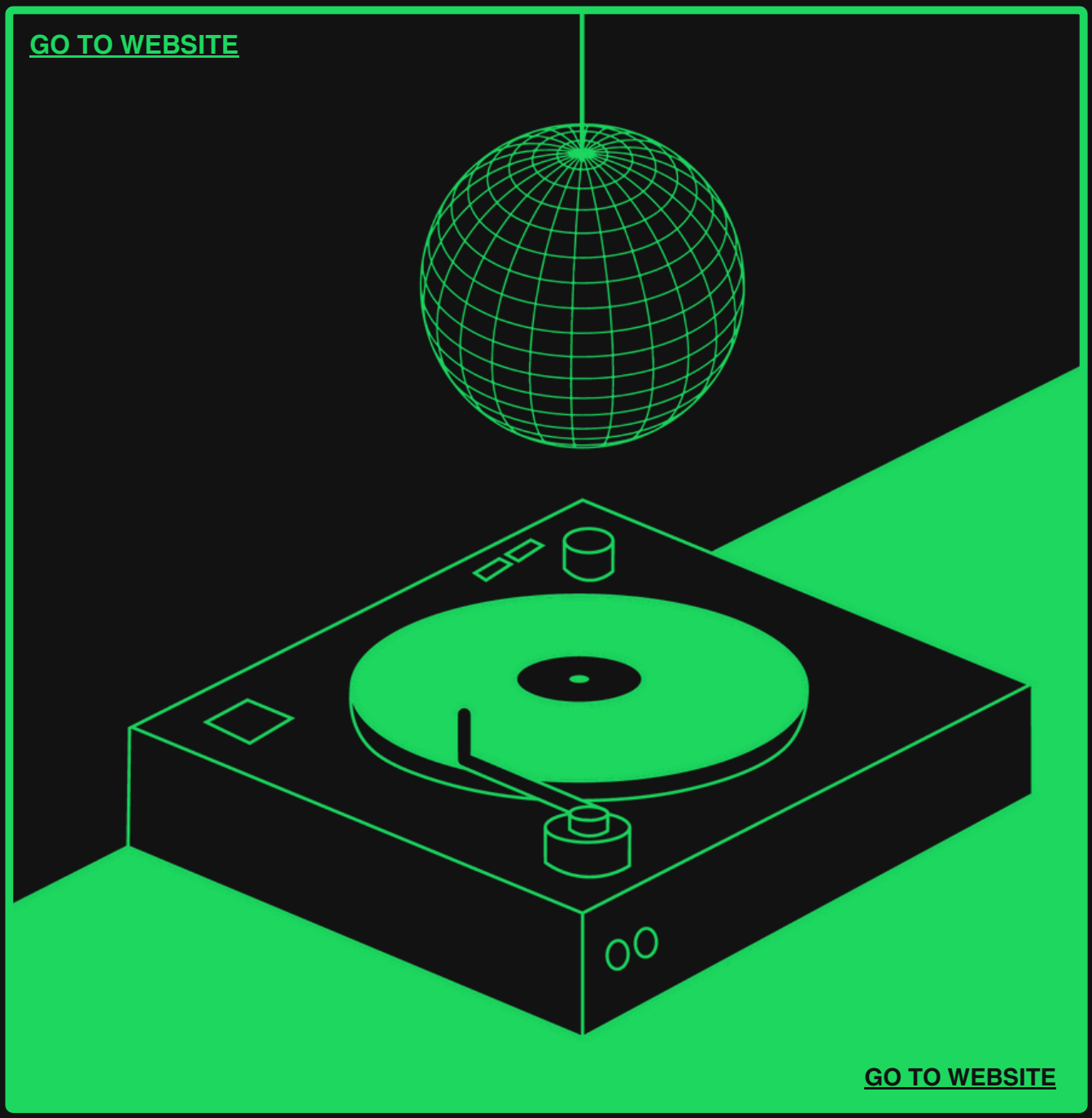



# SPOTIFY IN DATA

Yassine Turki, Kirill Zemlianskii, Levin Hertrich

## Vol. 0 Intro

Music plays an important role in most peoples lives. It motivates, helps in difficult times and makes the best moments even more beautiful. The important role of music is reflected in the Spotify charts, they show what songs are most popular at the moment and which artists are on the rise. But what is the world's music taste? Are there certain attributes in music that cause differences in music preferences across countries? This project aims to answer these questions and more. We analyze Spotify Chart Data to dive deep into the world of music and discover how the popularity of genres, songs and artists as well as music attributes differs between countries. We therefore created three main visualizations: Globe, Country Statistics and Artist Comparison as well as two extra visualizations for Music Discovery and a Dataset Overview.

## Vol. 1 Data

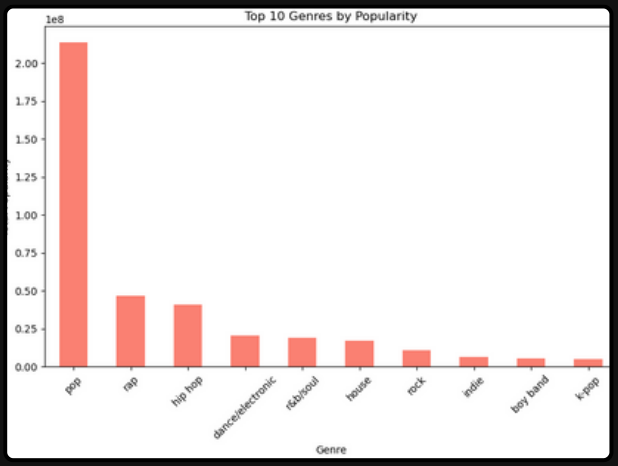
Each story needs a beginning. Ours is the Spotify Huge dataset. It contains all songs from Spotify's daily Top 200 charts in 34 countries in the period of 2017-2020. Even though it has already been cleaned from the creators, there was still some work left to make it useful for our project. With 151 columns it contained more information than we needed so we started by dropping columns not needed for our analysis like hot-encoded variables. We looked for missing values and found some columns containing missing values, so we decided to create two datasets: One where we drop columns with a lot of missing values and one where we keep all columns but drop rows with missing values. The idea was to minimize the dataloss by using the dataset without dropped rows whenever possible and the one with dataloss only when needing the dropped columns. But we realized that the dataloss when dropping rows with missing values would be too big so we ended up with the dataset with filtered columns for our analysis. In a last step we fixed columns with mixed types before then jumping into our analysis.

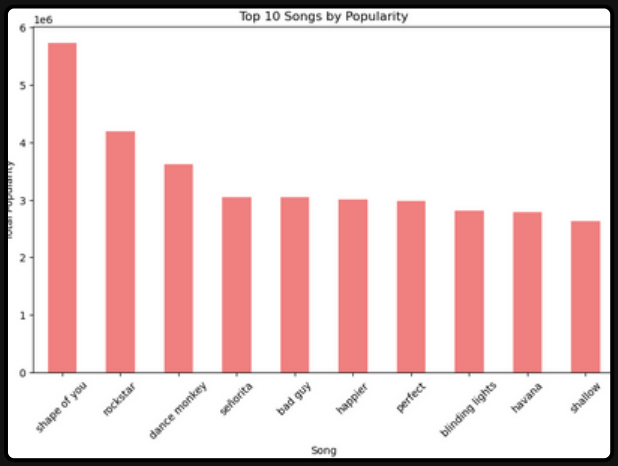
## Vol. 2 ~~EDM~~ EDA

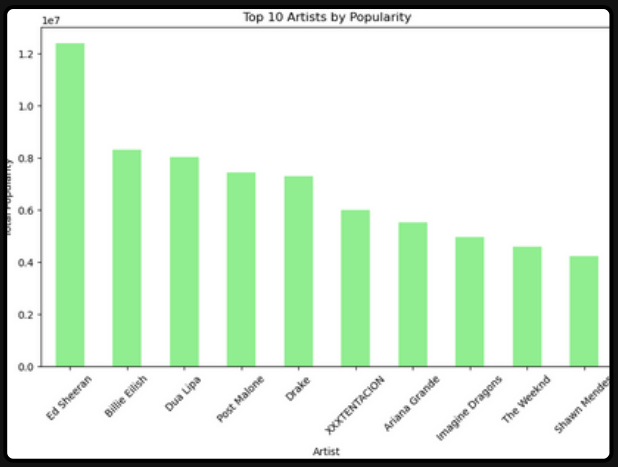
We started our project with an exploratory data analysis to get a better understanding of our data and to discover first insights. In a first step we calculated some basic statistics for the cleaned data:

| | |
|---|---|
| Number of Rows | 167356 |
| Number of unique Countries | 34 |
| Number of unique Artists | 25063 |
| Number of unique Albums | 33803 |
| Number of unique Genres | 1099 |
| Average Artist Followers | 8544889.08 |
| Average Popularity | 5475.70 |
| Average Song Duration | 206392.48 |
| Proportion of Explicit Songs | 36.82% |
| Average Days since Release | 1337.54 |

After getting some basic statistics we performed some initial analysis to see if the results correspond with our personal expectations. Note that the genre is aggregated, as there are many genres in our dataset. For our project we used all genres to have more fine-grained analyses.




*The results met our expectations as Ed Sheeran, Billie Eilish and Dua Lipa were very popular in the time from 2017-2020. Also Pop being by far the most popular genre is expected and shows that our data probably provides a good representation of the world's music taste.*

## Vol. 3 Layout and Colors

The goal of our layout is to be intuitive, easy-to-use and fun to interact with. We split the different analysis and tools of our project in different pages for a clear and easy understandable structure. We choose our globe visualization (more on that below) as starting page as it contains a lot of our analysis in a fun and interactive way and serves as an eye-catcher to keep users interested.

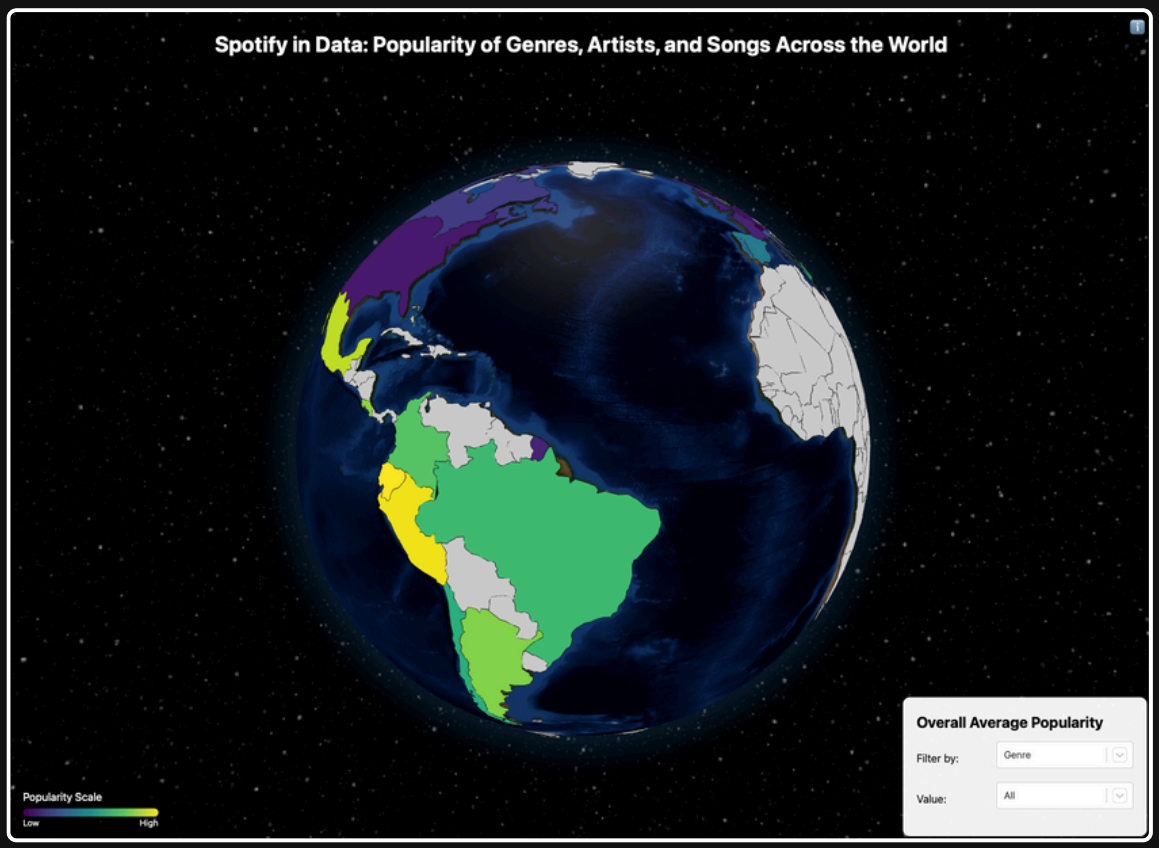
To explore further parts of our analysis the user can navigate between different pages. Therefore we provide two possibilities. First from the globe visualization, the page with more detailed country statistics can be accessed by just clicking on the corresponding country on the globe. Second we use a navigation bar on top of the page to enable a fast and easy navigation between all parts of our project. We chose to place the navigation bar on top of the page so it is easy to find and does not interfere with our visualizations.

We structure each part of our project by using titles and clear textual descriptions of visualizations and possible interactive tools that can be used. Furthermore we provide information through legends for example for the different plots of our country statistics.

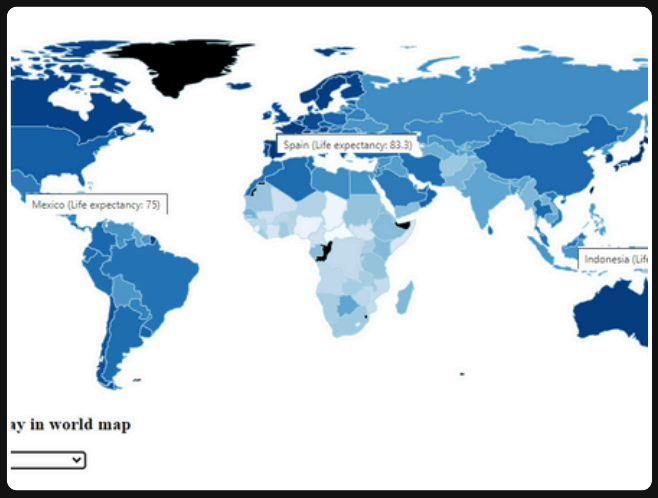
As the whole project analyses and visualizes Spotify data we oriented our color and design on Spotify's Design & Brand Guidelines. Overall we choose a dark color theme for a calm atmosphere and a design that reminds the user of his Spotify app. We then use the Spotify green for colored accents, for example in our navigation bar, to again follow Spotify's design and to make the website more exciting and vivid. We use white and black font depending on the background following our overall simple and minimal design and to ensure readability of all our texts. To represent popularity in our globe visualization, we decided to use the Viridis color scheme from D3.js to ensure accessibility for color-blind users.

## Vol. 4 Visualizations

The heart of this project is the visualization of our dataset and our analysis. This project consist of three main parts: Popularity of genres, artists and songs across countries, country statistics showing the most popular artists, songs and genre as well as the music taste in form of song characteristics, and the comparison of artists on their music characteristics. For each part we use different task-specific visualizations.

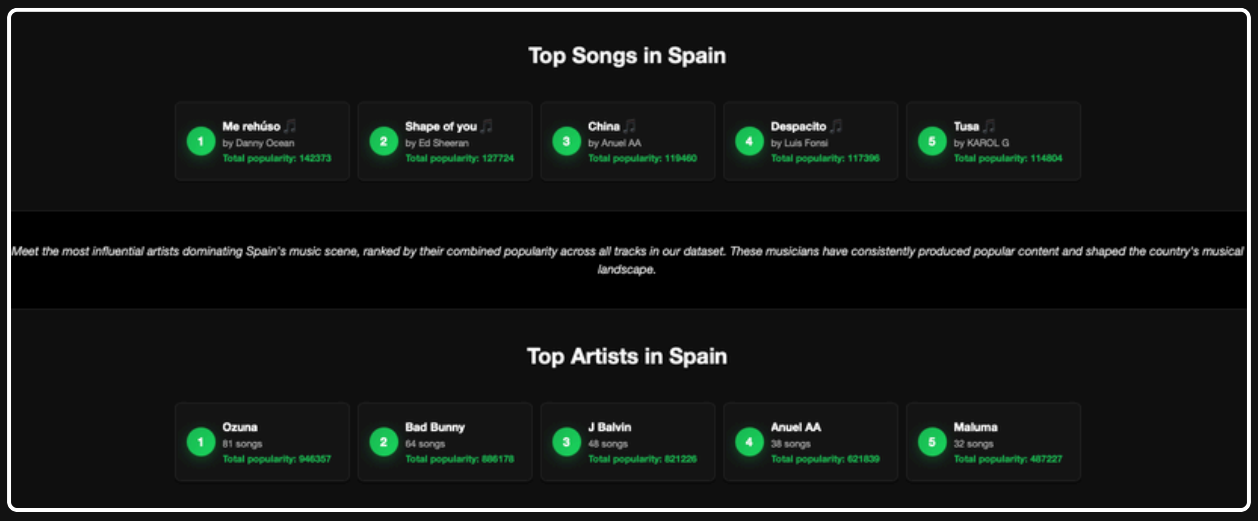


Our project starts with the globe visualization that can be seen above. This visualization shows the popularity of different genres, artists and songs across the countries in our dataset with a color scale. The visualization can be filtered with the filter panel in the lower right corner. Due to the interactive character of this visualization and the nice representation of key data of our project the globe is the perfect introduction to our project as it keeps users interested and engaged.



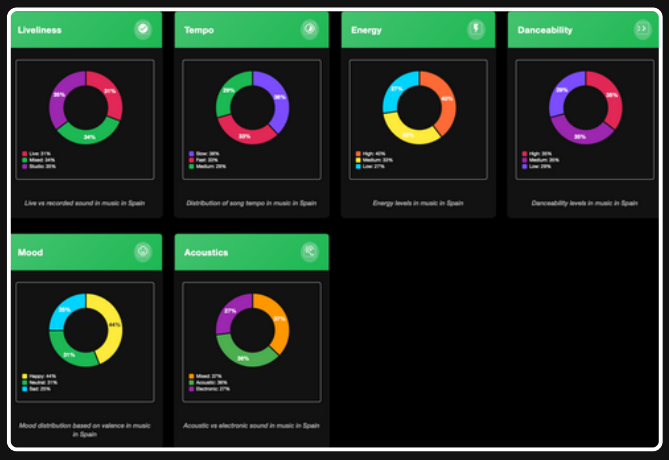

*Originally we planned to use a world map visualization as shown on the left, but after discovering the globe visualization, we decided to use it instead because of its beautiful visual design and interactiveness.*

Our second main visualization are the country statistics where we show popular artists, songs and genre for each country as well as the music profile for each country.

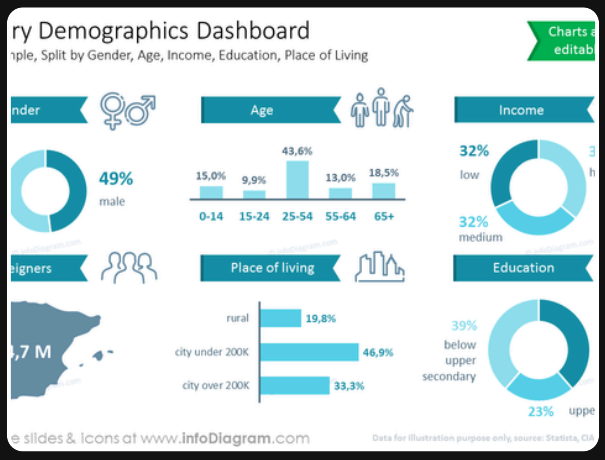



Above we show a snippet of the country statistics dashboard. The goal is to provide a good overview of the most important statistics for each country. We orient the design again on Spotify's design to align with the project data and topic. The clean dashboard enables user to gather the most important information quickly. All components change slightly when hovered, providing an interactive and enjoyable user experience. Our dashboard provides the most important information like the most popular songs and artists fast and in an easy understandable format.

To provide a deeper analysis of the dominant music characteristics for a specific country we visualize them with pie charts. We planned such a dashboard in the beginning of the project as can be seen below and adapted the idea during the project to first show the key findings as described above and then use charts to provide more detailed insights. The pie charts shown below provide a more detailed view on the music taste, in form of song characteristics, for a specific country. Again all components have a visual reaction to hover, enabling the user to highlight a specific characteristic and providing an interactive feedback.

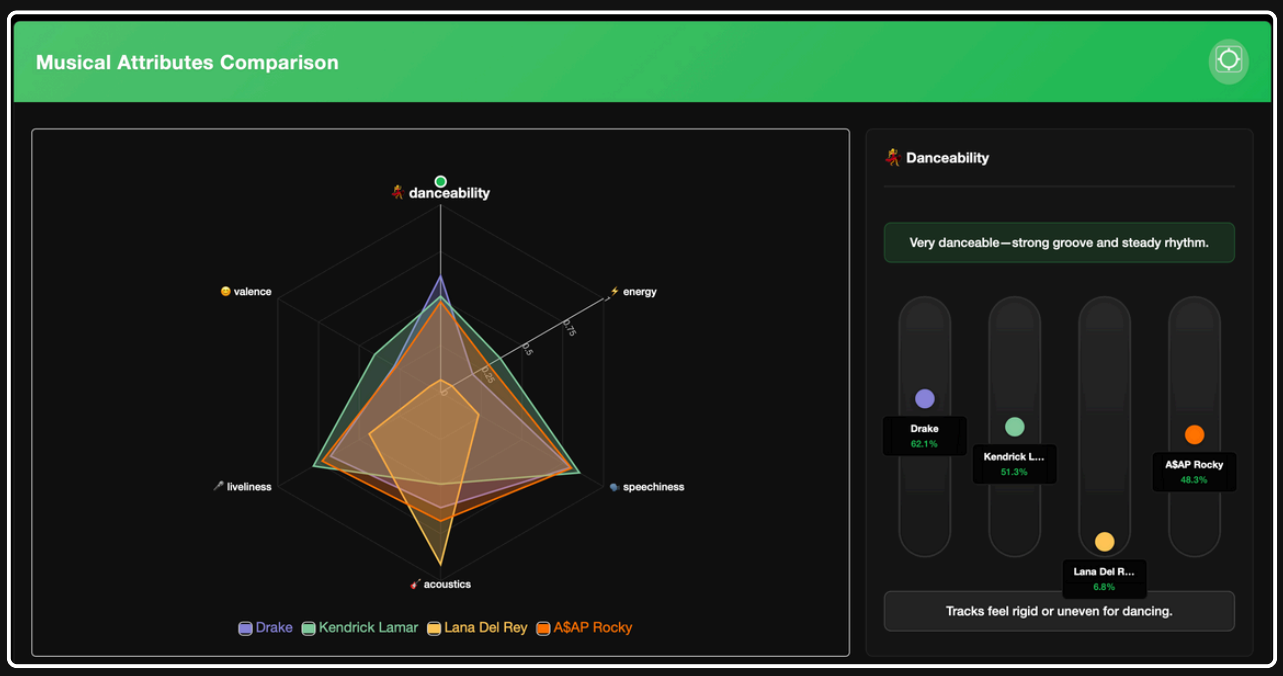



Implementation of the country statistics

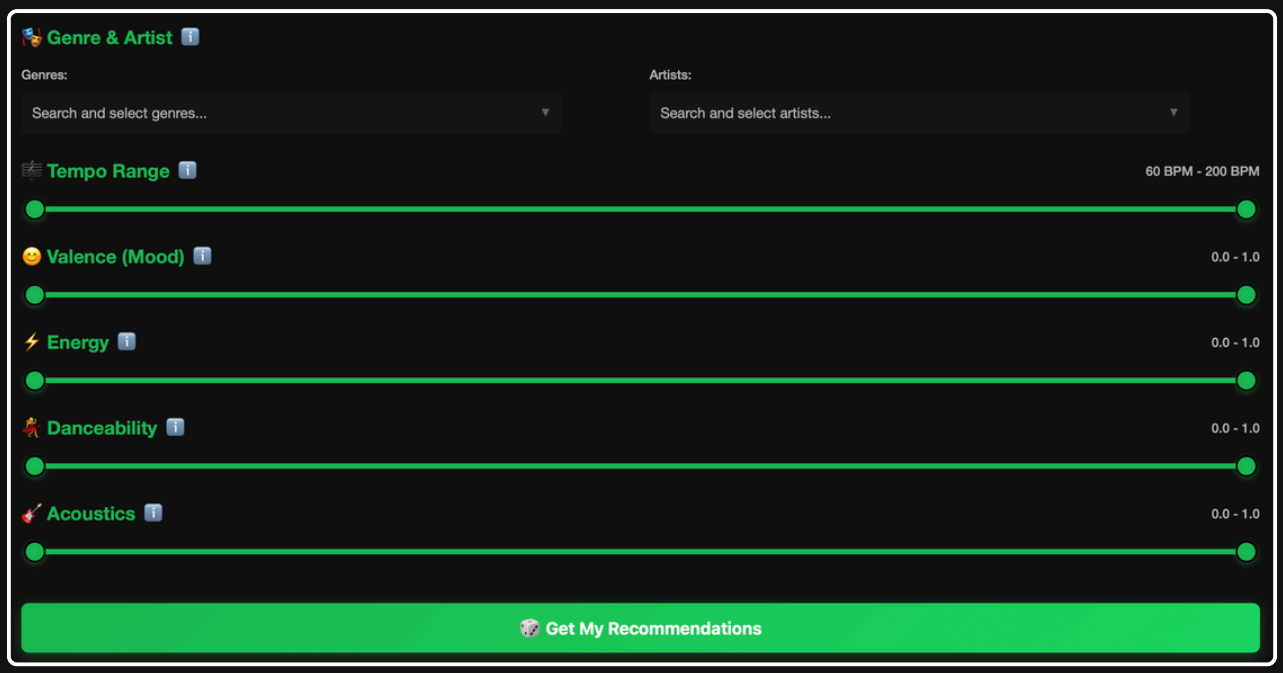



Orientation for the country statistics

The third main visualization of our project is the artist comparison. It enables user to compare their favorite artists on the average music characteristics of selected artists. The goal is that after discovering the different popularities of artists around the world and the different music characteristics dominant in different countries, user can use these discoveries to compare artists. We added a detailed view for each characteristic, to ensure the clarity of our analysis.



In addition to our three main visualizations described above, we considered adding more parts to our project. Initially we thought of a search tool for the whole dataset or user specific recommendations. We also thought about an Analysis Board which we implemented as a Dataset Overview to enable users to deepen the understanding of our dataset. As it is not a major feature in our project we don't provide a dedicated image here, but it can be found on our website. For the first two ideas we decided to combine them into a Music Discovery tool where users can find new songs based on preferences on genre and artists as well as song characteristics. This tool allows some pre-filtering and then recommends 5 random songs that fulfill the filter criteria, so users can discover new songs based on their music taste. The main visualization for this tool can be seen below.




## Vol. 5 The Story

As the opening of our story, we present an interactive globe that displays the popularity of genres, artists and songs around the world. A concise title and brief introductory text invite users to engage with global music trends and explore the question: What is the world's music taste?

Typically the next step in our story is the country statistics view, accessible by clicking on any nation. Here we reveal in-depth insights into each country's music taste, highlighting top genres, leading tracks and key song characteristics, providing a key part of our analysis.

After exploring the music taste of different countries, users can compare artists across the music characteristics uncovered earlier, representing the final part of our main story. While this order represents our intended story line, users are free to navigate in any order through our navigation bar and can furthermore explore our additional tools, like the song recommendations.

## Vol. 6 Outro - Challenges and Adaptations

During our project we faced multiple challenges and made adaptions accordingly.
In the beginning of our project we started with the D3.js framework for our visualizations, as intended by the course in which we do this project. Even though we were interested in learning a new framework, none of us had prior knowledge in it. As we had one team member with prior knowledge in React, we decided to switch to React to be able to develop faster and therefore be able to make the most out of our project.

After switching to React, we swapped our planned world map for an interactive globe visualization while keeping the other charts with minimal changes. We originally used a red-to-green scale for popularity but, after a color-blind team member struggled to distinguish values, we adopted the Viridis palette, reminding us that accessibility is essential for inclusive data exploration.

A major challenge was our dataset's limited coverage: while Spotify is Europe's dominant streaming platform, famously via "Spotify Wrapped", it sees low usage in Africa and Asia, leaving some large gaps on the globe. Attempts to merge other sources failed due to missing music attributes or data quality issues, so we kept the original dataset for its consistency and the rich information it provides across 34 countries.

## The Artists (Peer Assessment)

Yassine Turki:

- Implementation and design of the country statistics
- Implementation and design of the music discovery
- Exploratory data analysis
- Planning of visualizations

Kirill Zemlianskii

- Implementation and design of the artist comparison
- Overall project structure and adaption to react
- Exploratory data analysis
- Search for suitable datasets and related work

Levin Hertrich

- Implementation and design of the globe visualization
- Implementation and design of the dataset overview
- Writing and design of process book
- Planning of visualizations and overall design