

Tendências Musicais 2019

Visualização de Informação

Martim Bravo
Instituto Superior Técnico
97036

Gonçalo Adolfo
Instituto Superior Técnico
97090

Francisco Delgado
Instituto Superior Técnico
97402

1 INTRODUÇÃO

Neste projeto tivemos como objetivo, produzir uma visualização que seja concisa e o mais fácil e intuitivo possível para o utilizador analisar e obter as informações requeridas. O tema deste projeto centra-se na área musical e na análise das tendências musicais, tendo como base o top 200 diário do Spotify, ao longo do ano 2019. Esta visualização poderá ser útil para os próprios artistas, podendo estes analisar as trends ao longo do ano de 2019 de modo a concluir de que forma poderão ser mais apelativos para o público. Adicionalmente esta visualização pode ser uma ferramenta interessante para o público, para estes analisarem a nossa visualização, perceber o que é mais tocado atualmente e os pontos comuns que estas músicas terão, para desta forma estarem “a par” das tendências musicais. Decidimos ter o nosso foco mais centrado sobre as características das músicas e qual a correlação destas com o número de streams, com a finalidade de tentar de certa forma encontrar razões para as músicas estarem no top. Definiu-se que a visualização deve responder ao seguinte conjunto de questões:

- Questão 1: Qual o impacto das características (energia, tempo etc) da música no número de streams;
- Questão 2: Quais os artistas que tiveram um melhor ano 2019;
- Questão 3: Se o quão feliz ou dançável são as músicas no top depende da altura do ano;
- Questão 4: O tamanho das músicas afeta o desempenho destas?;
- Questão 5: Será que o facto das letras serem mais explícitas afeta o desempenho das músicas;
- Questão 6: Qual a evolução do número de streams das diferentes músicas no top (ao longo do ano ou durante um período específico).

2 TRABALHO RELACIONADO

A inspiração para o desenvolvimento do trabalho advém da própria aplicação Spotify (fonte dos dados recolhidos). Esta apresenta uma página web (spotify charts) que lista o top 200 diário. A ideia trata-se de complementar esta informação (que se encontra ilustrada em formato de lista), de modo a que o utilizador possa compreender os motivos e as tendências.

Na documentação da Spotify API, existe uma secção relativa a trabalhos desenvolvidos utilizando esta API. Não foram encontradas visualizações semelhantes à que se pretende implementar. A mais aproximada (Artist Explorer) apresenta como objectivo desenhar grafos entre artistas semelhantes e dispõe informações acerca destes: a sua popularidade no geral e músicas populares.

Foram encontradas também aplicações (Soundcharts, Next Big Sound) cujo objectivo é essencialmente possibilitar a análise de músicas, nomeadamente a identificação de zonas geográficas onde a

música possui um maior crescimento; popularidade em função das diferentes plataformas; métricas relativas aos artistas; entre outras features. Estas aplicações, pelo que foi entendido, não possibilitam a análise das tendências relativas às características musicais. Por outro lado, relativamente à nossa visualização, os dados extraídos não contemplam informação geográfica e são apenas relativos à plataforma Spotify.

A visualização que mais se aproximou advém de um blog, publicado em Julho de 2019 e escrito por Scott Teal, no qual descreve a abordagem no desenvolvimento de uma visualização com o objectivo de “visualize audio features of top tracks”, tal como mencionado pelo autor. A figura 1 ilustra o estado da dashboard final. Apesar do objectivo induzir ligeiramente a erro, é possível identificar algumas semelhanças, nomeadamente quanto à identificação das músicas/artistas mais ouvidos(as). Por outro lado na visualização descrita, o foco é efectuado com recurso a um mapa e não possui nenhum mecanismo de filtragem temporal.

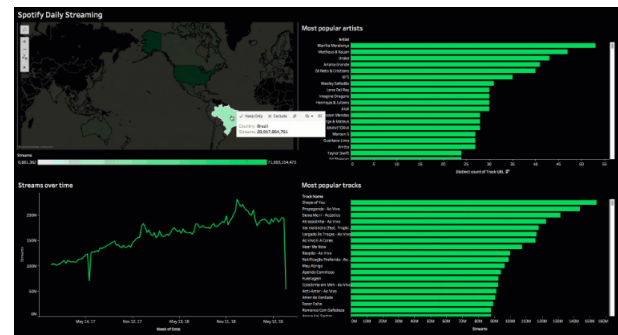


Figura 1: Visualização Relacionada

3 DADOS

Inicialmente pesquisámos datasets que permitiam responder às questões colocadas. Não conseguindo satisfazer os requisitos necessários, optámos por criar o nosso próprio dataset.

Como referido na introdução, o nosso dataset contém o top 200 das músicas mais tocadas diariamente ao longo do ano 2019. A informação foi buscada a partir do site Spotify Charts¹, onde tem uma listagem do top diário, sendo estes dados obtidos num formato CSV. Estes dados ainda estão incompletos para os nossos objetivos pois, este CSV apenas faz a listagem das músicas com poucas informações adicionais. Como se pode observar na figura 2, o Spotify Charts fornece além das listagens, o número de streams da música nesse dia em específico e o URL referente à música em questão. Para a nossa visualização precisávamos de mais informações, para tal, pegámos

¹<https://spotifycharts.com/>

na parte final do URL, referente ao id da música e fizemos pedidos à API do Spotify, que fornece os dados ao público, para obter as características de cada música. Fizemos dois tipos de pedidos, a dois diferentes endpoints (para /v1/tracks/{id} e /v1/audio-features/{id}). Onde os dados devolvidos são respetivamente os da figura 3 e 4. Do primeiro endpoint, obtivemos o URL da capa do álbum, se a música contém letras explícitas e o URL duma pequena amostra da música, sendo que ambos os URL acabaram por não ser utilizados. Do segundo endpoint seleccionámos todas exceto o type, o analysis URL, o track HRE, URI e o ID. Desta forma criamos um ficheiro JSON que tem todas estas informações dos endpoints e do CSV apenas adicionando uma variável *date* baseado no dia da listagem. Como o dataset foi criado por nós e já tínhamos esta visualização em mente, não foram necessárias variáveis derivadas para o pretendido. Apenas necessitamos duma variável adicional para esta visualização que é o intervalo de tempo para observar nos idiomas que é selecionado pelo utilizador e todos os cálculos são baseados nessa escolha.

```
(from Spotify Charts CSV)
Position,Track Name,Artist,Streams,URL
1,"Dance Monkey","Tones and I",5180626,https://open.spotify.com/track/lrgnBhdG2JDFtYkYRZAku
2,"HIGHEST IN THE ROOM","Travis Scott",4423871,https://open.spotify.com/track/3eekarcy7kvN4yt5ZFzltW
3,"Señorita","Shawn Mendes",3916024,https://open.spotify.com/track/6v3KW9xbzN5yKLT9YKDYA2
```

Figura 2: Resultado do Spotify Charts

```
(from Spotify API /v1/tracks/{id})
{ "album": { ... }, "artists": { ... }, "available_markets": { ... }, "disc_number": 1,
"duration_ms": 207959, "explicit": false, "external_ids": { ... }, "external_urls": { ... },
"href": "https://api.spotify.com/v1/tracks/1ldFghVXANMlKmJXaCbNl", "id":
"1ldFghVXANMlKmJXaCbNl", "is_local": false, "name": "Cut To The Feeling", "popularity":
63, "preview_url": "https://p.scdn.co/mp3-
preview/3eb16d18c2a700240e9dfb817b6f2d041f15eb1?cid=774b29d4f13844c495f206cafdad9c86",
"track_number": 1, "type": "track", "uri": "spotify:track:1ldFghVXANMlKmJXaCbNl" }
```

Figura 3: Resultado do pedido ao primeiro endpoint

```
(from Spotify API /v1/audio-features/{id})
{ "duration_ms": 255349, "key": 5, "mode": 0, "time_signature": 4, "acousticness": 0.514,
"danceability": 0.735, "energy": 0.578, "instrumentalness": 0.0902, "liveness": 0.159,
"loudness": -11.840, "speechiness": 0.0461, "valence": 0.624, "tempo": 98.002, "id":
"06AKEBrKucKwKREUWRnVt", "uri": "spotify:track:06AKEBrKucKwKREUWRnVt", "analysis_url":
"https://api.spotify.com/v1/tracks/06AKEBrKucKwKREUWRnVt", "analysis_uri":
"https://api.spotify.com/v1/audio-analysis/06AKEBrKucKwKREUWRnVt", "type":
"audio_features" }
```

Figura 4: Resultado do pedido ao segundo endpoint

4 VISUALIZAÇÃO

4.1 Descrição geral

A nossa solução funciona à base duma tabela no lado esquerdo do dashboard que contém uma listagem de todas as músicas relativas ao intervalo de tempo escolhido (Figura 5). Como tal esta tabela é o ponto essencial do nosso projeto e onde todos os idiomas irão buscar os dados necessários para mostrar. Em relação aos idiomas, utilizámos seis tipos diferentes de idiomas para responder às questões.

Rank	Track Name	Artist	Streams
1	Dance Monkey	Tones and I	5180626
2	ROXANNE	Arizona Zervas	4352027
3	Memories	Maroon 5	4191138
4	everything i wanted	Billie Eilish	4055300
5	Circles	Post Malone	3777019
6	Don't Start Now	Dua Lipa	3493376
7	RITMO	Bad Boys For	3466034

Figura 5: Listagem das músicas

Na tabela da esquerda, para além da seleção do intervalo de tempo para visualizar, temos um botão para caso o utilizador queira fazer uma filtragem pelas características das músicas ou pelo seu número de streaming. Adicionalmente podemos procurar diretamente as músicas e os artistas utilizando a barra de pesquisa e escrever o pretendido. Relativamente à listagem, esta está ordenada pelo número de streams, onde em cada entrada da tabela se pode observar o nome e artista da música, o número de streams e se tem letras explícitas ou não. Caso o utilizador queira pode seleccionar até três músicas para visualizar com mais detalhe nos idiomas, ficando estas *highlighted* quer na listagens quer nos idiomas onde serão utilizadas. Quando a visualização é carregada, as três faixas com mais streams estão seleccionadas por omissão (Figura 6). O mesmo acontece quando deixa de haver alguma música seleccionada. Por último temos no final da tabela o número de músicas que contém letras explícitas.

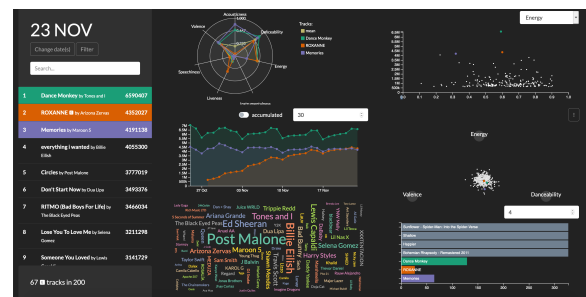


Figura 6: Visualização com nenhuma faixa selecionada

Relativamente aos diferentes idiomas, decidimos utilizar seis tipos diferentes para responder às questões escolhidas.

Inicialmente, criamos um radial chart (Figura 7) para facilitar a análise das músicas que o utilizador deseja, pois este idioma

demonstra o valor das diferentes características das músicas. Permitindo assim uma análise mais simples entre as diversas músicas. As faixas selecionadas pelo utilizador aparecerão com a cor correspondente como uma entrada no idioma. Colocámos adicionalmente uma entrada no radial chart que contém a média das características das músicas que estão na listagem corrente. Temos uma feature adicional que caso o utilizador clique numa das características demonstradas neste idioma, este irá definir a característica do scatter plot. Ainda temos a possibilidade de o utilizador fazer *hover* numa das músicas na legenda, isto fará com que a opacidade das outras entradas no idioma diminuam, realçando a que fazemos *hover*.

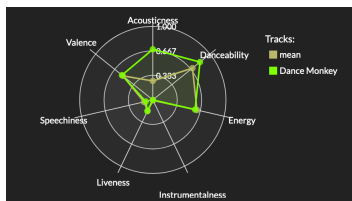


Figura 7: Radial Chart

De seguida, decidimos utilizar um scatter plot (Figura 8), pois achámos este o mais correto para analisar correlações entre diferentes características das músicas com a sua popularidade. A faixa selecionada pelo utilizador ficará *highlighted* no idioma, aumentando também o tamanho do ponto associado à faixa para uma percepção mais simples do utilizador. A característica a ser demonstrada neste idioma pode ser alterada para a desejada utilizando a drop down list no canto superior direito do idioma. O utilizador pode clicar na faixa que desejar no scatter plot e esta ficará como selecionada, atualizando os restantes idiomas.

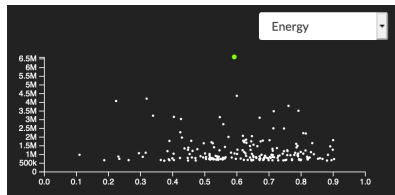


Figura 8: Scatter Chart

Próximo idioma foi o area chart (Figura 9), que foi utilizado principalmente para fazer uma análise da popularidade de diferentes faixas, podendo esta análise ser relativa à popularidade diária ou à popularidade acumulada, para saber o preciso dia onde uma faixa começou a ter mais visualizações. Sendo esta escolha efetuada num *switch*. Adicionalmente o utilizador pode escolher quantos dias quer ver representados no area chart, por exemplo se o utilizador escolher 5 dias vai ver a evolução das streams nos 5 dias anteriores.

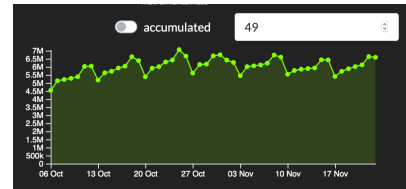


Figura 9: Area Chart

O quarto idioma foi o multi-foci force layout (Figura 10), onde temos um ponto referente a cada uma das músicas da listagem. Neste idioma podemos ter até quatro pólos, onde cada um destes pólos é uma característica da música. Os pontos do idioma serão "puxados" para o pólo consoante o valor da característica da música, por exemplo, se uma música tiver uma energia muito alta o ponto referente a essa música vai-se aproximar mais ao pólo da energia. Decidimos utilizar este idioma, pois demonstra as tendências das músicas em relação as características escolhidas, permitindo assim responder algumas perguntas sem termos de fazer uma análise muito extensiva. Tal como o scatter plot a faixa selecionada vai ficar *highlighted* no idioma com a cor associada à faixa. O utilizador pode clicar na faixa que desejar neste idioma e esta ficará como escolhida, alterando os restantes idiomas.

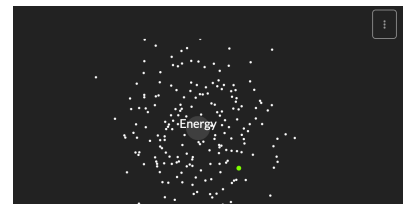


Figura 10: Multi-Foci Force Layout

Seguidamente decidimos implementar um word cloud (Figura 11) para mostrar os artistas mais tocados no intervalo previamente escolhido. Ficando os artistas mais populares com maior destaque pois o seu nome estará maior. Este idioma serve principalmente para ter uma ideia visual e não muito detalhada desta popularidade, pois este idioma não permite esse nível de detalhe. Adicionalmente podemos clicar num dos nomes do artista e este irá para a barra de pesquisa, mostrando apenas as músicas desse artista no top.



Figura 11: Word Cloud

Por último, implementámos um bar chart horizontal (Figura 12), para demonstrar as músicas que estão à mais dias consecutivos no top e ainda para mostrar à quantos dias consecutivos estão as

músicas selecionadas. O número de músicas que estão à mais tempo no top, que são mostradas neste idioma, dependem do número que o utilizador queira ver, sendo esta opção controlada a partir do idioma, podendo mostrar as 15 músicas à mais tempo ou 0 (mostrando apenas as selecionadas pelo utilizador). Tal como no scatter plot e multi-foci podemos clicar numa faixa diretamente neste idioma para a selecionar.

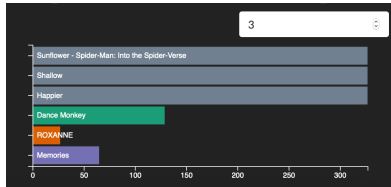


Figura 12: Bar Chart

4.2 Fundamentação

Inicialmente tínhamos uma visualização à base de idiomas muito simples, utilizando múltiplos Area charts e bar charts, que pioravam o nosso modelo quer em termos de visual quer em termos de análise do utilizador. Deste modo, introduzimos mais variabilidade nos idiomas para tornar mais apelativo mas também facilitar a análise. Desde o checkpoint 3, onde definimos o nosso prototipo inicial, que os nossos idiomas nunca se alteraram, apenas tendo de fazer alguns pequenos ajustes.

Para a resposta da pergunta "Quais os artistas que tiveram um melhor ano 2019" pensamos em ter um botão na tabela que iria agregar as diferentes entradas de cada artista e fazer uma listagem somente dos artistas ordenados pelo número de streams, deste modo sabendo qual o mais popular. Esta abordagem tinha um problema que era que tornava a word cloud desnecessária e tinha uma desvantagem extra de ser mais difícil de analisar do que o idioma, como tal, decidimos eliminar esta ideia da nossa visualização final.

Os nossos dados tiveram de ser trabalhados numa maneira diferente a inicialmente pensada, pois temos de fazer alguns cálculos previamente para melhorar a performance da nossa visualização, como tal, antes de carregar a visualização calculamos alguns dados auxiliares para melhorarem a performance.

Os visual encodings utilizados para cada um dos idiomas são os seguintes

Radial Chart:

-Marks:

- Ponto - Uma característica da faixa selecionada.

-Channels:

- Cor - Uma das faixas selecionadas; (categórica)
- Distância para o centro - Valor da característica. (ratio)

Scatter Plot:

-Marks:

- Ponto - Uma das faixas da listagem.

-Channels:

- Cor - A faixa selecionada; (categórica)
- Posição Vertical - Número de Streams; (ratio)

- Posição Horizontal - Valor da característica selecionada. (ratio)

Area Chart:

-Marks:

- Ponto - O número de streams de uma faixa, quer seja acumulada ou diária.

-Channels:

- Cor - A faixa selecionada; (categórica)
- Posição Horizontal - Dia; (ordinal)
- Posição Vertical - Número de streams. (ratio)

Multi-Foci Force Layout:

-Marks:

- Ponto - Uma faixa de listagem.

-Channels:

- Cor - A faixa selecionada; (categórica)
- Distância para o pólo - Valor da característica do pólo. (ratio)

Word Cloud:

-Marks:

- Palavra - Artista.

-Channels:

- Tamanho - Popularidade desse artista. (ratio)

Bar Chart:

-Marks:

- Barra - O número consecutivo de dias que a faixa está no topo.

-Channels:

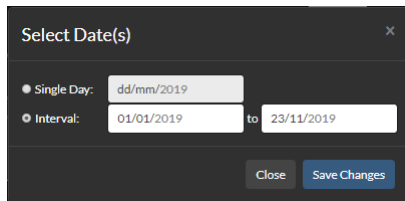
- Cor - A faixa selecionada; (categórica)
- Comprimento - Número de dias que está no topo. (ratio)

Para todos os idiomas exceto o word cloud, utilizamos como channel a cor para distinguir as faixas, pois estas têm uma natureza categórica.

4.3 Demonstrações

No início do trabalho foram colocadas um conjunto de questões requisito. Nesta secção irão ser descritos os passos necessários de modo a que estas sejam respondidas.

As figuras 13, 14, 15 ilustram os passos para retirar conclusões perante a primeira pergunta: "como é que as características musicais (energy, loudness, tempo, etc) afectam o número de streams?". Primeiramente, tal como em outras questões, selecciona-se o dia ou intervalo temporal no qual se pretende analisar (Figura 13). De seguida, selecciona-se a característica na drop-down list. Por fim, interpreta-se no scatter plot a correlação entre a feature selecionada e o número de streams. As figuras 14 e 15 representam dois exemplos com a energia e liveness, respectivamente, das músicas para todo o ano.



Select Date(s)

Single Day: dd/mm/2019

Interval: 01/01/2019 to 23/11/2019

Close Save Changes

Figura 13: Selecção do intervalo de tempo pretendido

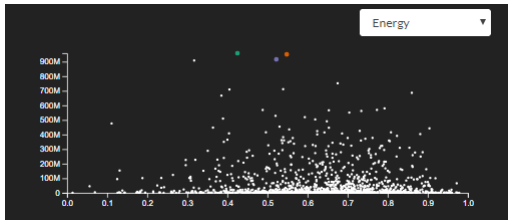


Figura 14: Scatter plot com streaming numbers em função da feature energy

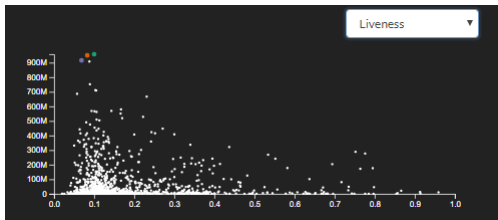


Figura 15: Scatter plot com streaming numbers em função da feature liveness

Em relação à pergunta "existe alguma correlação entre a happiness/valence e a altura do ano?", efectua-se o seguinte conjunto de passos:

- (1) seleccionar os meses de uma altura do ano em concreto, da mesma forma dos exemplos anteriores (Figura 13);
- (2) interpretar multi-foci e radar chart.

As figuras 16 e 17 ilustram o multi-foci para alturas do ano diferentes. Tal como se pode observar, não existe grande impacto no peso das características inerentes à questão. O mesmo se conclui observando a média presente no radar chart.

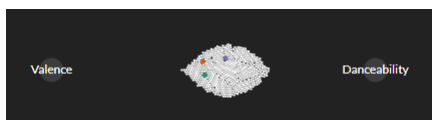


Figura 16: Danceability/valence Jun, Jul, Aug



Figura 17: Danceability/valence Jan, Feb, Mar

Relativamente às outras perguntas, são respondidas da seguinte forma:

- "A popularidade das músicas é afectada pela sua duração?": seleccionar a altura do ano; seleccionar a duração como característica do scatter plot; analisar possível correlação
- "Que artistas tiveram um bom ano 2019? Que músicas foram mais tocadas durante um período?": seleccionar altura do ano; a lista de músicas demonstra as faixas mais tocadas nesse período; o word cloud demonstra os artistas em destaque de acordo com o número de streams acumulado das suas músicas
- "Qual a evolução de uma faixa num intervalo de tempo?": seleccionar altura do ano; analisar area chart (por omissão demonstra o número de streams nos últimos 30 dias mas é possível alterar período)
- "Que faixas duraram mais tempo no top 200?": seleccionar altura do ano pretendida; analisar bar chart que demonstra o número de dias consecutivos que uma faixa esteve no top de acordo com o período de tempo seleccionado

5 DETALHES DE IMPLEMENTAÇÃO

A visualização foi implementada com auxílio à biblioteca React.js. Esta é utilizada para implementação de interfaces e segue o paradigma components-based. Adicionalmente, foi utilizada a library bootstrap, de modo a facilitar a implementação do grafismo e organização dos elementos presentes nos componentes. Os componentes da dashboard estão ilustrados na figura 18. Tal como se pode observar, existe uma componente (Tracks) para o menu de configuração (definição do intervalo temporal, parâmetros de filtragem, lista de músicas) e uma componente para o conjunto dos idiomas (Visualizations). Cada idioma também se trata de um componente.

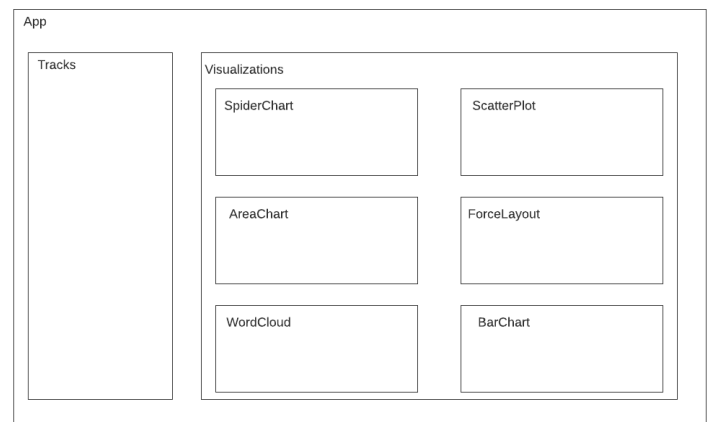


Figura 18: Componentes da Dashboard

Os handlers dos eventos relativos à mudança de intervalo temporal ou música selecionada são definidos na componente App. Isto deve-se ao facto da informação da lista de músicas "total"/selecionadas serem necessárias em ambos os componentes filhos: Tracks e Visualizations. Todos os componentes apresentam uma função render, responsável por retornar os elementos HTML que definem o componente. O código seguinte demonstra um excerto da renderização da componente App. Tal como se pode observar, esta ao definir a componente Visualizations passa-lhe um conjunto de propriedades, no qual são recebidas no seu construtor.

```
1 <Visualizations tracks={this.state.tracks}
  selectedTracks={this.state.selectedTracks}
  />
2 allTracksMap={allTracksMap} onTrackClick={
  this.handleOnTrackClick}
3 date={this.state.date} />
```

Cada componente apresenta um objecto state, no qual define propriedades que o caracterizam. Quando estas se alteram, o componente volta a ser renderizado. Adicionalmente, se houver propriedades recebidas do componente pai e estas se alterarem, o componente também volta a ser renderizado. A seguinte lista apresenta propriedades no estado de algumas componentes exemplo:

- App: maxDate, tracks, selectedTracks, nSelectedTracks, numExplicit, date;
- Tracks: showDateForm, search;
- AreaChart: accumulated, nDays, minDate, maxDate;
- BarChart: nTracks;
- etc

Relativamente ao código que desenha os idiomas não foi implementado nenhum algoritmo de raiz. Em termos de organização, cada idioma é representado por uma função que por sua vez está implementada num ficheiro js com um nome ilustrativo. Estas são importadas no componente relativo a esse idioma. De forma bastante resumida e genérica, cada função irá:

- (1) criar o svg inerente ao idioma;
- (2) criar os eixos;
- (3) adicionar marcas (pontos, linhas, palavras, barras);
- (4) definir eventos e interacção.

Durante a implementação, foi encontrado um problema depois da criação do idioma do area chart, notámos que este tinha algumas músicas duplicadas, isto deve-se ao facto de haver a mesma música (com as mesmas características, artista e nome da música), mas que este tem um ID diferente. Para solucionar este problema, quando mostramos as músicas na listagem da esquerda, verificamos essas redundâncias e agrupamos na mesma entrada as diferentes versões da mesma música com as mesmas características.

6 CONCLUSÕES E TRABALHO FUTURO

O processo de desenvolvimento da dashboard incorporou diferentes etapas: definição das perguntas alvo, extracção e processamento dos dados, desenho da interface de utilizador e por fim a sua implementação. Esta sequência de blocos não foi exactamente linear, uma vez que foram identificados problemas em blocos anteriores. De acordo com o produto final, e tal como mencionado anteriormente, todas

as questões conseguem ser respondidas. Com base na sequência descrita, realça-se especialmente o desenho da visualização: escolha dos idiomas adequados em função das perguntas alvo e do tipo de dados com que estamos a lidar.

O desenvolvimento da visualização permitiu a percepção da sua importância para os utilizadores. A eficiência da interpretação das questões colocadas torna-se significativamente superior, transformando as tarefas alvo pré-atentivas e bastante mais apelativas. Note-se também que a visualização permite responder a mais questões do que as inicialmente colocadas, como por exemplo:

- identificar quando é que uma track ultrapassou outra relativamente ao número de streams acumulados;
- verificar o quanto uma track se afasta do tradicional relativamente às suas características;
- comparar tracks quanto à sua evolução;
- entre outras cujos idiomas permitam responder.

Como trabalho futuro, e tal como observado no trabalho relacionado, seria interessante incorporar:

- informações geográficas das músicas: permitir a observação de zonas geográficas relevantes consoante características musicais especificadas;
- dados provenientes de diferentes plataformas: permitir a visualização do quanto certa faixa é utilizada nas diferentes redes sociais;
- ferramentas de report automático;
- aplicação de técnicas Data Science, nomeadamente problemas de regressão e forecasting.