



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

Instituto de Ciências Exatas e Informática – Pontifícia Universidade Católica de Minas
Gerais (PUC-Minas) Belo Horizonte – MG – Brazil

1º Trabalho - Ciência de Dados

Belo Horizonte

2019

Ciência de Dados: Estudo dos dados da lista de 200 músicas mais escutadas diariamente no aplicativo Spotify

Gustavo de Castro Nogueira

Lucas Gabriel Padrão Rezende

Roberto Lafetá Mendonça Gomes

RESUMO

Foi estudado uma base de dados gerada a partir das 200 músicas mais escutadas diariamente no aplicativo Spotify entre janeiro de 2017 e julho de 2021. A base foi extraída do site Kaggle, com auxílio da metodologia orientada a dados é feita uma redução da tabela, com objetivo de facilitar o manuseamento das informações para identificar o problema encontrado. Destarte, com a premissa foi encontrado os gêneros mais tocados no aplicativo ao longo do período supracitado.

Palavras-chave: dados, spotify, estilo musical, métricas, top 200

ABSTRACT

A database generated from the 200 songs most listened to daily in the Spotify application between January 2017 and July 2021 was studied. The database was extracted from the Kaggle website, with the aid of the data-oriented methodology, a reduction of the table is made, in order to facilitate the handling of information to identify the problem found. Thus, with the premise, the most played genres in the app over the aforementioned period were found.

Keywords: database, spotify, genre, metrics, top 200

1 INTRODUÇÃO

No contexto de ciência de dados, uma pesquisa orientada a dados estuda a manipulação de uma central de informações com o objetivo de possibilitar a visualização de determinado problema identificado. Nessa perspectiva, com intuito de replicar essa metodologia, que é feita baseada em métricas, foi realizada uma investigação com a base de dados extraída do Kaggle gerada a partir de dados do Spotify com propósito de identificar qual a tendência dos estilos musicais mais populares com o passar do tempo. Por conseguinte, esse estudo visa estabelecer uma replicação de mercado, uma vez que, as averiguações desse aplicativo são feitas frequentemente, mostrando qual inserção musical está sendo mais popular no aplicativo. Dessa forma, podendo gerar informações

úteis que podem ser utilizadas por profissionais que pertencem ou não à indústria da música.

2 DESENVOLVIMENTO

Durante o estudo foram utilizados métodos de pesquisa de dados para parametrizar o resultado obtido. De tal forma, foi realizado seguindo quatro diretivas, sendo a primeira fase a identificação de uma base dados, a qual foi feita utilizando o site Kaggle buscando a base dos dados das 200 músicas mais escutadas diariamente no aplicativo Spotify entre janeiro de 2017 e julho de 2021.

A priori, foi feita uma redução dos dados fazendo um tratamento na coluna “GENRE” na qual existiam valores inválidos como a mensagem de erro 500 e linhas em branco o que não era interessante para o estudo.

Position		Track Name	Artist	Streams	Date	Genre
139	140	Lighthouse - Andrelli Remix	Hearts & Colors	439245	2017-01-01	['Sorry', "that one doesn't seem to be on any ...
304	105	Lighthouse - Andrelli Remix	Hearts & Colors	550199	2017-01-02	['Sorry', "that one doesn't seem to be on any ...
498	99	Lighthouse - Andrelli Remix	Hearts & Colors	616045	2017-01-03	['Sorry', "that one doesn't seem to be on any ...
692	93	Lighthouse - Andrelli Remix	Hearts & Colors	640563	2017-01-04	['Sorry', "that one doesn't seem to be on any ...
886	87	Lighthouse - Andrelli Remix	Hearts & Colors	656930	2017-01-05	['Sorry', "that one doesn't seem to be on any ...
...
320846	47	Build a Bitch	Bella Poarch	1726882	2021-07-16	['Sorry', "that one doesn't seem to be on any ...
321044	45	Build a Bitch	Bella Poarch	1642880	2021-07-17	['Sorry', "that one doesn't seem to be on any ...
321108	109	Faz Amor Comigo Só Hoje - Ao Vivo	Israel & Rodolffo	1016256	2021-07-17	['Sorry', "that one doesn't seem to be on any ...
321142	143	Só Não Divulga	Fernando & Sorocaba	899378	2021-07-17	['Sorry', "that one doesn't seem to be on any ...
321152	153	Batom de Cereja - Ao Vivo	Israel & Rodolffo	865327	2021-07-17	['Sorry', "that one doesn't seem to be on any ...
518 rows × 6 columns						

Figura 1: Exemplo de valor inválido no campo gênero

	Track Name	Artist	Streams		Genre	Year	Month	Day
308994	Bubblegum Bitch	MARINA	676327	500 Internal Server Error\n\nInternal Server E...		2021	04	04
309185	Bubblegum Bitch	MARINA	713763	500 Internal Server Error\n\nInternal Server E...		2021	04	05
309378	Bubblegum Bitch	MARINA	741283	500 Internal Server Error\n\nInternal Server E...		2021	04	06
309578	Bubblegum Bitch	MARINA	749757	500 Internal Server Error\n\nInternal Server E...		2021	04	07
309787	Bubblegum Bitch	MARINA	740742	500 Internal Server Error\n\nInternal Server E...		2021	04	08
309998	Bubblegum Bitch	MARINA	793846	500 Internal Server Error\n\nInternal Server E...		2021	04	09
310377	Bubblegum Bitch	MARINA	721813	500 Internal Server Error\n\nInternal Server E...		2021	04	11
310564	Bubblegum Bitch	MARINA	763784	500 Internal Server Error\n\nInternal Server E...		2021	04	12
310759	Bubblegum Bitch	MARINA	762403	500 Internal Server Error\n\nInternal Server E...		2021	04	13
310955	Bubblegum Bitch	MARINA	773013	500 Internal Server Error\n\nInternal Server E...		2021	04	14

Figura 2: Outro exemplo de valor inválido no campo gênero

Inicialmente esse campo apresentava um conjunto de gêneros em qual a música se encaixava, em ordem de relevância, ou seja, o principal gênero era o primeiro da lista. Foi extraído apenas o valor do principal gênero e o mesmo substituiu o valor inicial do campo. Por fim, a coluna “DATE”, campo de data (formato AAAA-MM-DD) foi subdividida em “YEAR”(AAAA), “MONTH”(MM) e “DAY”(DD), facilitando a captação do dado fragmentado para visualização gráfica, como exemplificado pela figura 3 e 4.

...

	Position	Track Name	Artist	Streams	Date	Genre
0	1	Starboy	The Weeknd	3135625	2017-01-01	['canadian pop', 'canadian contemporary r&b', ...
1	2	Closer	The Chainsmokers	3015525	2017-01-01	['pop', 'pop dance', 'tropical house', 'edm', ...
2	3	Let Me Love You	DJ Snake	2545384	2017-01-01	['pop', 'electronic trap', 'dance pop', 'edm', ...
3	4	Rockabye (feat. Sean Paul & Anne-Marie)	Clean Bandit	2356604	2017-01-01	['pop', 'uk dance', 'dance pop', 'uk funky', '...
4	5	One Dance	Drake	2259887	2017-01-01	['toronto rap', 'canadian pop', 'canadian hip ...

Figura 3: Base de dados original

	Track Name	Artist	Streams	Genre	Year	Month	Day
0	Starboy	The Weeknd	3135625	canadian pop	2017	01	01
1	Closer	The Chainsmokers	3015525	pop	2017	01	01
2	Let Me Love You	DJ Snake	2545384	pop	2017	01	01
3	Rockabye (feat. Sean Paul & Anne-Marie)	Clean Bandit	2356604	pop	2017	01	01
4	One Dance	Drake	2259887	toronto rap	2017	01	01
...
321195	Best Friend (feat. Doja Cat)	Saweetie	790257	trap queen	2021	07	17
321196	Wonderwall - Remastered	Oasis	789099	madchester	2021	07	17
321197	MI FAI IMPAZZIRE	BLANCO	788966	italian pop	2021	07	17
321198	Adore You	Harry Styles	788033	pop	2021	07	17
321199	Follow You	Imagine Dragons	785417	modern rock	2021	07	17

320622 rows × 7 columns

Figura 4: Base de dados pós tratamento

Em segundo lugar foi feita a definição do problema com base na pergunta orientada a dados, a qual consiste na identificação da evolução dos gêneros mais tocados com o passar do tempo. Concomitantemente à análise exploratória de dados, a partir da utilização de métodos estatísticos como no caso a percentil (demonstrado graficamente pelo Treemap da figura 5) e a moda (a frequência com que cada gênero aparecia ao longo dos anos, também representada na figura 5 pelo gráfico de barras e de linhas) dos dados obtidos com o objetivo de responder a pergunta problema.

Em seguida, foi feita a visualização de dados, utilizando-se da ferramenta Microsoft PowerBI para a construção de gráficos a partir dos dados tratados. Tornando assim

possível descrever os dados em elementos visuais para uma análise mais simples e precisa, como representado na figura 5.

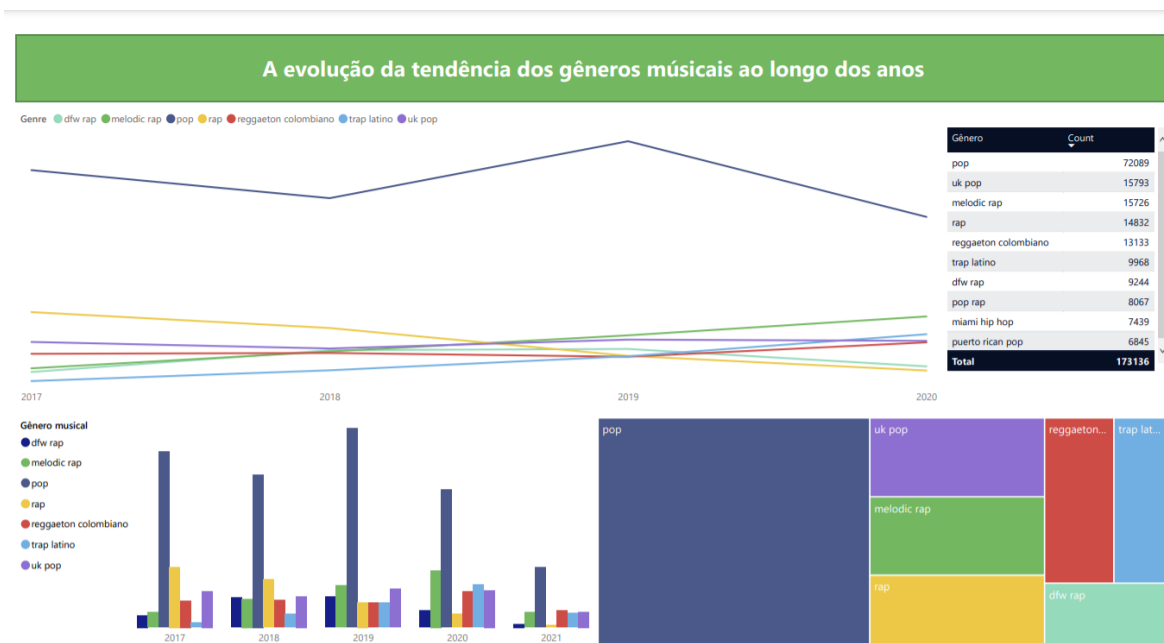


Figura 5: Visualização gráfica com PowerBi

3 CONSIDERAÇÕES FINAIS

A partir da análise dos dados e dos gráficos gerados podemos concluir que o gênero Pop foi o mais escutado dentre os anos de 2017 e 2021, com seu auge em 2019 e declínio em 2020. Esse seguiu uma tendência inversa dos outros gêneros populares, que cresceram com o desenrolar do período. Outra informação interessante que pode se identificar é a constante queda nos números de reprodução do gênero rap tradicional (apenas “rap” no gráfico), inicialmente um dos mais escutados, acompanhado da forte ascensão na popularidade de suas vertentes, como o rap melódico e o trap latino. É importante salientar também a constância na relevância dos gêneros reggaeton colombiano e pop britânico.

Essa análise de dados é de grande valor para a indústria musical, assim como as outras as quais se utilizam da mesma como produtoras musicais ou marcas que queiram se associar com algum estilo musical, pois com ela pode-se estipular quais as tendências dos próximos anos. Seria interessante, por exemplo, investimentos nos artistas do gêneros de pop por sua alta popularidade, em músicos de reggaeton colombiano ou pop britânico devido sua constância com o passar dos anos e nos gêneros de rap melódico e trap latino devido ao recente crescimento das mesmas. Por outro lado, o investimento no

rap tradicional, uma vez que apresenta uma grande queda de seus números, deixa de ser interessante.

REFERÊNCIAS

No contexto de ciência de dados