



Universidade Estadual de Campinas
FACULDADE DE ENGENHARIA ELÉTRICA E COMPUTAÇÃO

Trabalho Final de Graduação

**Análise de correlação entre métricas de alcance
e engajamento em plataformas musicais e
elementos de criação e distribuição de músicas,
dentro do nicho do *psy-trance***

Aluno: **Leonardo Mori Ferreira Luz**
RA: 172035

Orientador: **Prof. Dr. Tiago Tavares**

1º sem - 2021
Campinas

Conteúdo

1	Introdução	3
2	Objetivos	3
3	Metodologia	3
3.1	Levantamento de repertório a ser analisado	3
3.2	Levantamento de métricas relevantes para o estudo	4
3.3	Enriquecimento de dados e criação de dataset	5
3.4	Cálculos estatísticos	5
3.5	Análise dos resultados de correlação	6
4	Resultados e Discussão	6
5	Conclusões	28

1 Introdução

A música e o modo como é construída e consumida sofreu grandes evoluções desde seus primórdios. Desde os primeiros sons, na Idade da Pedra, passando pela escuta ativa, em teatros, até a criação da indústria fonográfica e a distribuição em massa, a sonoridade e o processo de criação sempre foi se moldando e evoluindo, adaptando-se ao tempo presente e seu contexto cultural e tecnológico.

No contexto do estudo presente, inteligência artificial é uma das tecnologias que vêm sendo cada vez mais explorada por diversos campos de pesquisa. Tais técnicas vêm sendo amplamente empregadas, por exemplo, na análise de opiniões de clientes sobre produtos ou empresas em comentários de redes sociais, no acompanhamento da saúde mental de pacientes no contexto médico, e na captura do senso comum em eventos comerciais, políticos e até econômicos.

A música e a tecnologia sempre tiveram seus caminhos interligados e hoje não é diferente. Na busca de alcançar a “música perfeita”, ou o “próximo grande *hit*”, alguns cientistas e músicos têm criado soluções de criação de músicas através de inteligência artificial. Porém, sabe-se também que a música se encontra diretamente relacionada à aspectos emocionais e uma série de outros fatores também influenciam o impacto que a música tem na vida das pessoas. Com isso em mente, o presente projeto busca estudar a relação entre o alcance e engajamento de músicas com outros aspectos de sua criação e distribuição, tentando chegar à uma conclusão sobre qual é o fator mais importante no quesito adesão do público.

2 Objetivos

Este trabalho tem como objetivo estudar a existência de correlação entre métricas relevantes de alcance e engajamento de músicas em plataformas de *streaming*, com diversos fatores de criação e distribuição como, características estruturais das mesmas, base de fãs dos artistas e base de fãs dos selos envolvidos. Caso exista correlação representativa, o atual trabalho busca aprofundar em quais são os fatores que têm maior influência na composição das métricas de sucesso nas diversas plataformas de *streaming* e realizar a análise dos resultados obtidos através da metodologia. Como foco deste estudo, buscou-se estudar o cenário musical do *psy-trance*.

3 Metodologia

Este trabalho foi dividido em cinco etapas principais sendo elas:

- Levantamento de repertório a ser analisado;
- Levantamento de métricas relevantes para o estudo;
- Enriquecimento de dados;
- Cálculos estatísticos;
- Análise dos resultados de correlação.

3.1 Levantamento de repertório a ser analisado

Para o levantamento de repertório a ser analisado nesta pesquisa levou-se em conta a seguinte lógica:

- Levantamento dos principais eventos do nicho escolhido;
- Levantamento dos artistas com maior recorrência em participações de eventos;
- Seleção dos artistas que seriam envolvidos na pesquisa;
- Levantamento das músicas mais relevantes de cada artista.

Para a etapa de levantamento dos principais eventos do nicho foi feita uma busca inicial por eventos da área em buscadores *online*. Ao levantar os principais eventos do setor foi feita uma segunda busca, mais específica, em relação a estes eventos e o *line up* das edições passadas de cada um. Com isso, obteve-se o levantamento da lista inicial de artistas envolvidos na pesquisa.

Tendo esta lista inicial em mãos, foi realizada a checagem das redes sociais dos mesmos com o intuito de buscar mais nomes de eventos, onde possivelmente agregaria a base e daria mais volume e confiabilidade aos dados. Com isso, descobriu-se que a rede social *Facebook* era o maior canal de divulgação dos eventos, não sendo visto divulgação dos mesmos nas outras redes sociais, ou, se visto, com uma intensidade muito menor.

A partir deste segundo momento foi possível fazer o levantamento de uma base suficientemente grande para iniciar a próxima etapa de seleção dos artistas que seriam objeto de estudo. Ao se analisar a recorrência de alguns nomes e utilizando do conhecimento prévio do autor em relação ao cenário musical escolhido, foi feita a listagem de artistas que seriam envolvidos.

Em seguida foi feita a pesquisa de alcance das principais músicas de cada artista pré selecionado, buscando músicas que tivessem número de *plays* próximos ou maiores do que 1 milhão, na plataforma *Spotify*. Com isso observou-se que nem todos os artistas tinham variadas músicas que atingiam a meta estipulada de *plays* na plataforma de *streaming*, fazendo com que o número de músicas variasse de artista para artista. Como critério de diversificação, foram escolhidas entre 1 e 4 músicas de cada artista, obtendo assim a lista final de músicas que compõem o repertório desta pesquisa.

3.2 Levantamento de métricas relevantes para o estudo

Para o levantamento das métricas que direcionaram os resultados do estudo, a linha de raciocínio incorporada foi a de levantar dados que tivessem relação com fatores de criação e distribuição do repertório, sendo aspectos estruturais e característicos das músicas, assim como aspectos indiretos como base de fãs do artista e base de fãs dos selos pelas quais foram lançadas. Como resultado, os critérios de análise foram:

Critérios característicos da música:

- É um *remix*, ou não;
- Ano de lançamento;
- Tonalidade da música;
- BPM da música;
- Duração da música.

Critérios estruturais da música:

- Número de *drops* da música;
- Número de *breaks* da música;
- Número de cadências de baixo;
- Uso, ou não, de vocal;
- Caso exista vocal, é usado de forma melódica, ou não;
- Caso exista vocal, ele foi criado por uma voz masculina, feminina, ou ambos;

Critérios da base de fãs dos artistas:

- Número de seguidores no *Deezer*;
- Número de seguidores no *Instagram*;
- Número de seguidores no *Facebook*;
- Número de seguidores no *Soundcloud*;
- Número de seguidores no *Spotify*;

- Ouvintes mensais no *Spotify*;
- Número de seguidores no *Twitter*;
- Número de seguidores no *Youtube*;

Critérios da base de fãs dos selos:

- Número de seguidores no *Instagram*;
- Número de seguidores no *Facebook*;
- Número de seguidores no *Soundcloud*;
- Número de seguidores no *Spotify*;
- Número de seguidores no *Twitter*;
- Número de seguidores no *Youtube*.

Todos estes critérios foram analisados em relação às métricas de sucesso, definidas neste estudo como:

- Número de plays no *Soundcloud*;
- Número de curtidas no *Soundcloud*;
- Número de plays no *Spotify*;
- Número de plays no *Youtube*;
- Número de curtidas no *Youtube*;

3.3 Enriquecimento de dados e criação de dataset

O enriquecimento de dados foi feito manualmente e teve um processo diferente para cada tipo de dado.

Para os dados, chamados na seção anterior de “Critérios característicos da música”, a fonte de referência foi o site *Beatport*, uma plataforma de compra de músicas. Buscando pela música, ou pelo autor, se obteve os dados relativos a ser um *remix*, ou não, o ano de lançamento, a tonalidade da música, seu BPM e seu tempo de duração. Além disso, a plataforma também entrega o dado relativo ao nome do selo que lançou a música.

Porém, nem todas as músicas se encontram dentro da plataforma, nestes casos específicos utilizou-se de outros recursos para a coleta de dados, dependendo de qual dado faltava. O *Google* e o *Youtube* foram utilizados para a descoberta do ano de lançamento, na questão de ser *remix* e no tempo de duração e o programa *Virtual DJ* foi usado para a coleta de tonalidade e BPM da música.

Para os “Critérios estruturais da música” o *Virtual DJ* foi o programa utilizado, realizando a escuta ativa do repertório e utilizando a visualização da forma de onda para auxiliar no entendimento do resultado sonoro e na coleta dos resultados.

Para a coleta de dados relativos às bases de fãs, tanto dos artistas quanto dos selos, se utilizou cada rede social individualmente para a obtenção dos números. Porém, as plataformas nem sempre retornavam o resultado desejado da pesquisa, levando em conta que algumas contas nas redes sociais não tinham tanto movimento e acabaram perdendo relevância para o algoritmo. Para solucionar isto, foi utilizado o *Google* e o *Spotify* para indicar as outras redes sociais dos artistas e selos.

Com isso, criou-se o *dataset* a ser analisado nesta pesquisa.

3.4 Cálculos estatísticos

Na etapa de cálculos estatísticos, após a base de dados já ter sido criada, foi necessário a realização de três tipos de cálculos diferentes: o cálculo de correlação direta, o T-teste e a ANOVA, de acordo com o tipo de resposta obtida. Mas antes da realização dos cálculos,

foi necessário extrair o logaritmo de base 10 dos números relativos às métricas de alcance e engajamento, pois as mesmas tinham discrepâncias muito grandes e dificultariam a visualização dos resultados.

Para os dados que retornavam respostas numéricas, o procedimento utilizado foi a correlação de dados, através da fórmula de correlação do *Google Sheets*, “=CORREL()” entre uma métrica e o critério de alcance, ou engajamento, analisado, obtendo assim a o grau de intensidade de ligação entre os parâmetros com cada métrica de sucesso definida.

Para os dados que retornavam resultados qualitativos utilizou-se o T-teste, quando o resultado era binário, e a ANOVA, quando o resultado era não binário. Para os dois casos, foram usadas calculadoras online que realizaram tais cálculos. Com isso, foi possível obter a média, o desvio padrão e o P valor dos resultados.

3.5 Análise dos resultados de correlação

Para extrair o máximo das análises dos resultados foi necessário utilizar tabelas dinâmicas, no *Google Sheets*, para concentrar as informações que seriam transformadas em gráficos, para dar suporte ao entendimento final dos resultados obtidos. Em seguida, foram gerados os gráficos para a visualização dos resultados e a facilitação das análises.

4 Resultados e Discussão

Como resultado do *dataset*, obteve-se os seguintes dados:

Foram levantados dados de 154 músicas, que foram compostas por 118 músicos, seja de forma individual, ou em colaboração. A seguir encontra-se a distribuição do repertório de acordo com os critérios de análise.

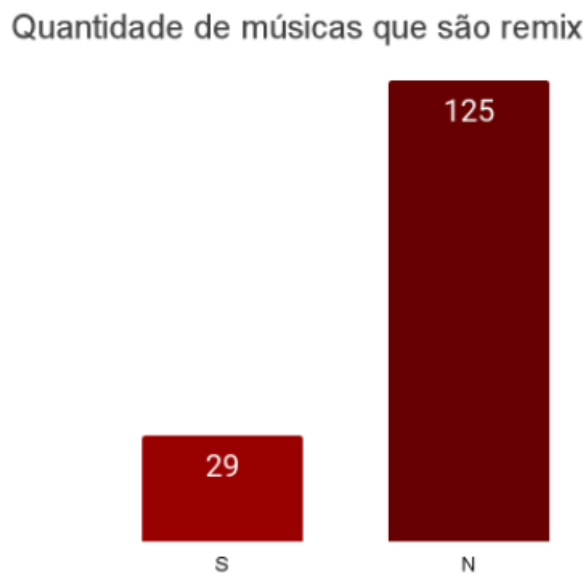


Figura 1: Quantidade de músicas que são *remix* e quantidade de músicas que são originais

Pode-se observar a predominância de músicas que são originais, foram 125, o que representa 81,17%. Contra 29 músicas remixadas.

Quantidade de músicas por ano de lançamento

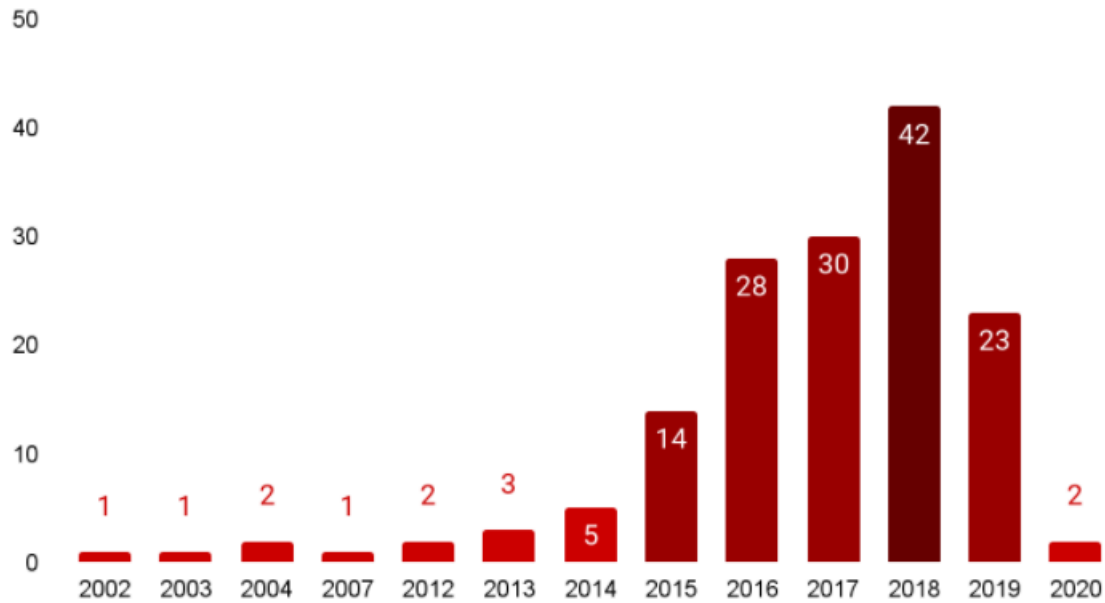


Figura 2: Distribuição das músicas por ano de lançamento

Pode-se observar uma predominância de músicas com menos de 5 anos de lançamento, que representam mais de 90% da amostra. Como destaque no ano de 2018 foram lançadas 42 músicas, o que representa 27,27% do total.

Quantidade de músicas por tonalidade

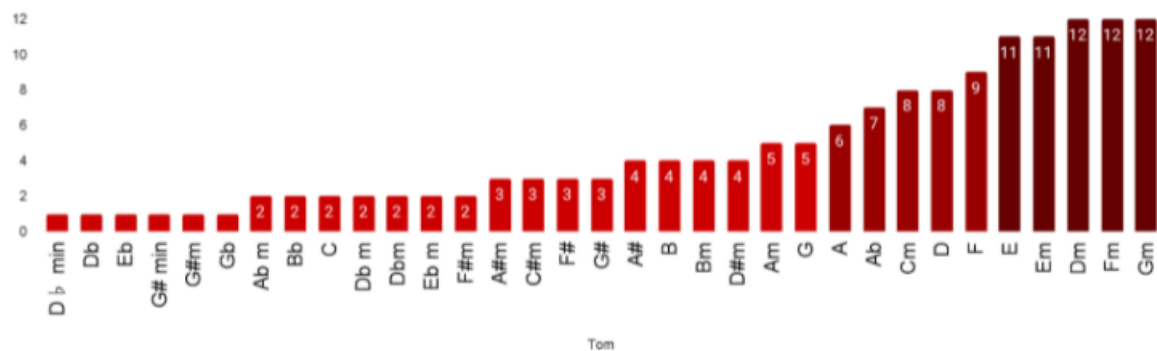


Figura 3: Distribuição das músicas por tom

Observa-se a predominância da escala menor entre os tons E (mi) até G (sol), representando 37,66% dos resultados.

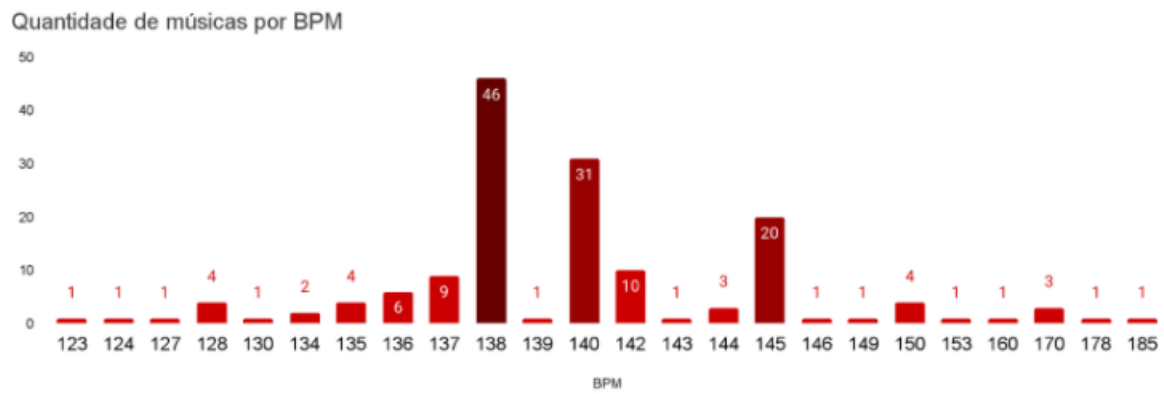


Figura 4: Distribuição das músicas por BPM

A análise da distribuição de músicas conforme seu BPM retorna que o valor mais utilizado foi 138 BPM, com 46 músicas, representando 29,87% da amostra, sendo seguido por 140 e 145 BPMs, com 31 e 20 músicas, respectivamente.

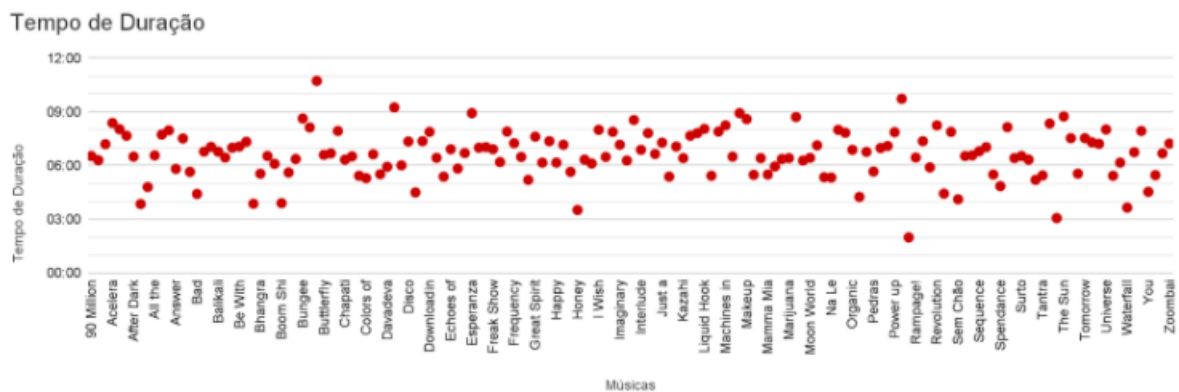


Figura 5: Distribuição das músicas por tempo de duração

Pode-se notar que a maioria das músicas se encontram entre a faixa de 4 à 9 minutos, podendo ter grandes variações dentro deste alcance.

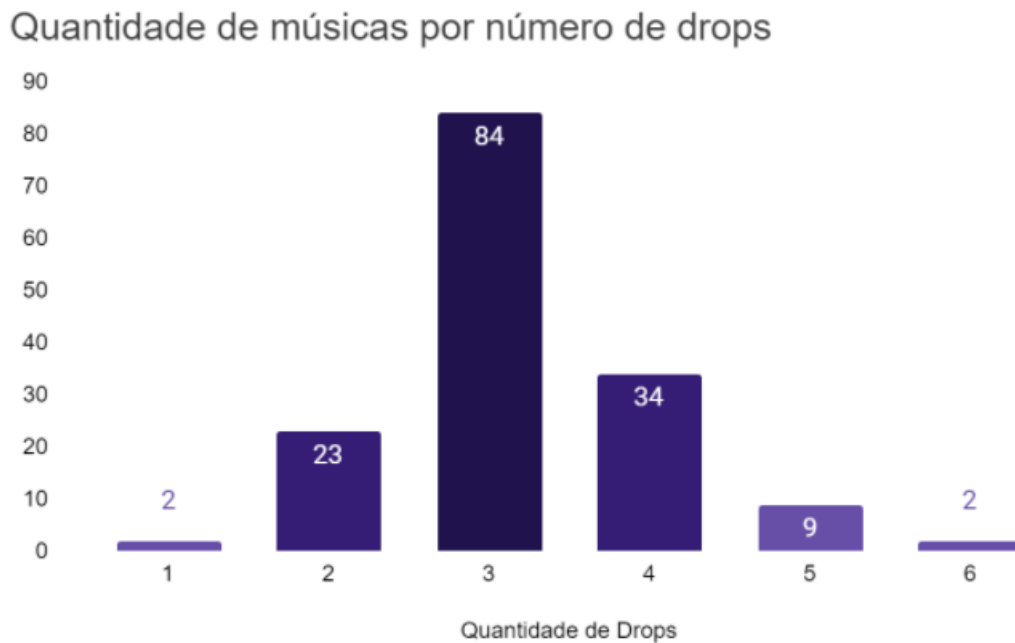


Figura 6: Distribuição das músicas de acordo com o número de *drops*

O *drop* na música eletrônica é caracterizado pela parte da música onde existe muita energia e pouca tensão, tendo o bumbo e o baixo como elementos característicos que fazem parte da seção. Ao analisar a amostra, a grande maioria (54,55%) têm 3 *drops* durante o decorrer da música.

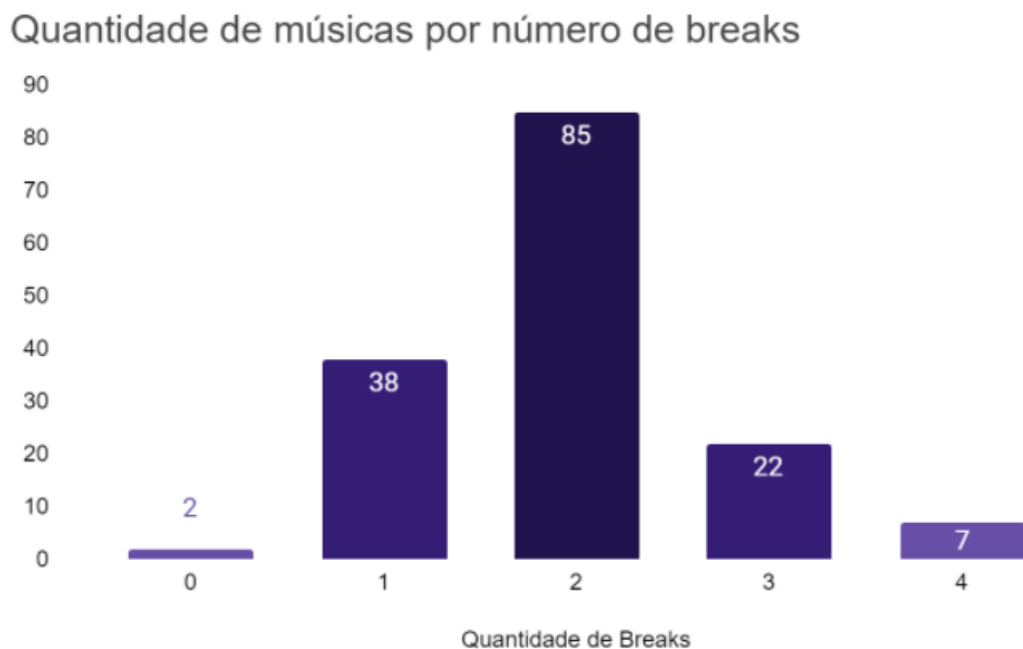


Figura 7: Distribuição das músicas de acordo com o número de *breaks*

O *break*, ao contrário do *drop*, é caracterizado pela seção da música com baixa energia e baixa tensão, geralmente acompanhado de ambiências e vocais. Em 55,19% dos casos desta amostra foram observados 2 *breaks* ao longo da música.

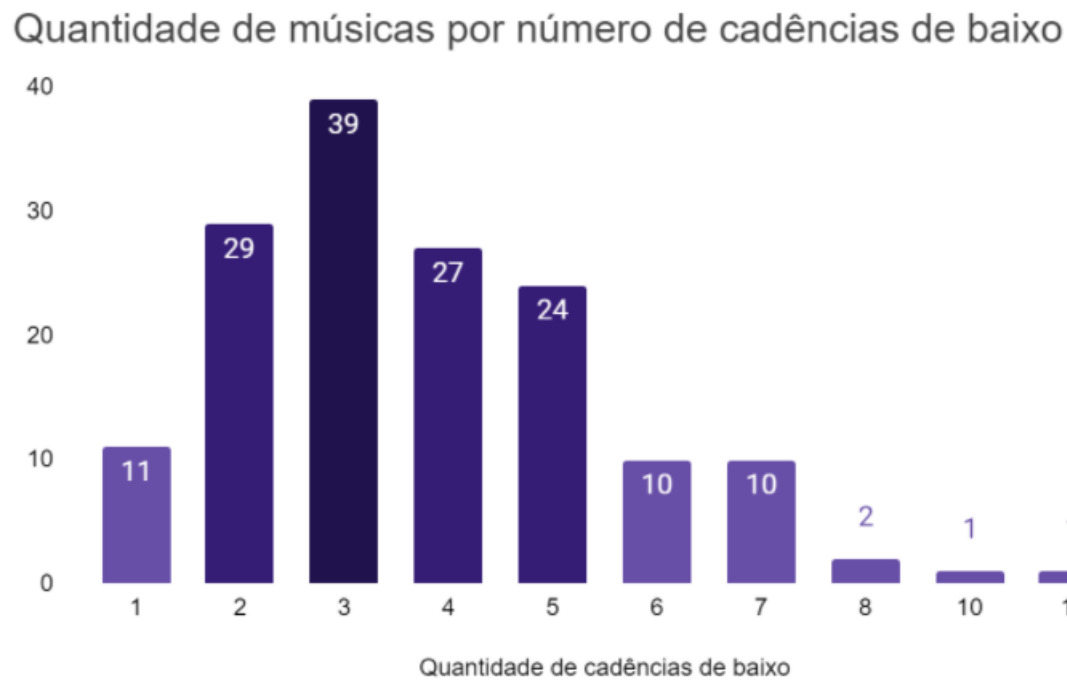


Figura 8: Distribuição das músicas de acordo com a quantidade de cadências de baixo utilizadas na música

Observou-se uma variação na distribuição das músicas conforme quantidade de cadências de baixo, tendo predominância de músicas com 3 cadências de baixo, seguida por 2, 4 e 5 cadências diferentes na mesma música.

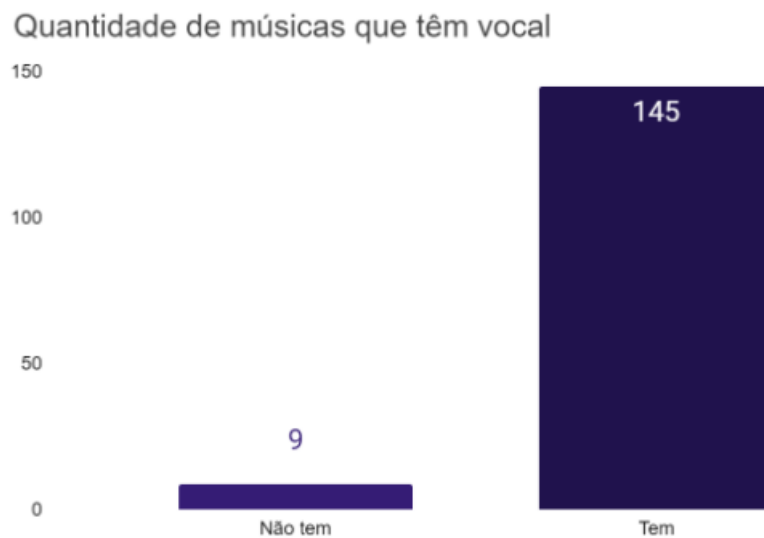


Figura 9: Quantidade de músicas que possuem vocal

Os resultados mostram que 94,16% das músicas, ou seja 145 delas, têm vocal. Indicando que este é um fator bem representativo do gênero musical.

Quantidade de músicas que tem vocal melódico

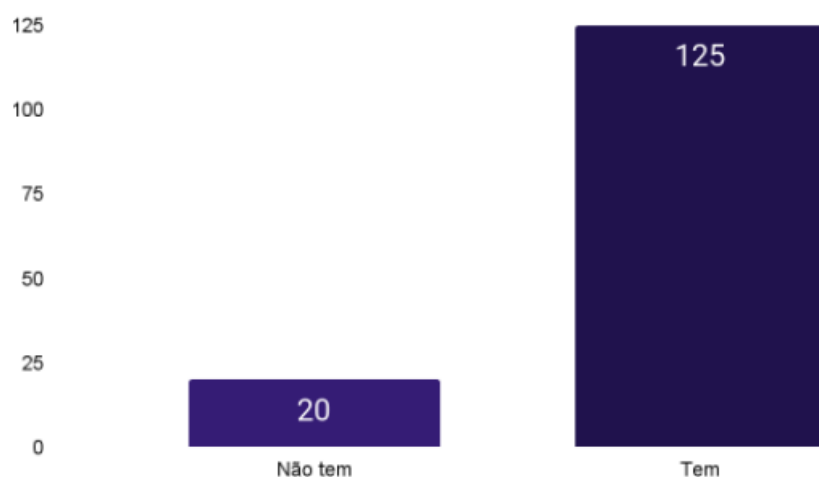


Figura 10: Quantidade de músicas que utilizam o vocal de forma melódica

Dentre o repertório que utiliza vocal, 125 músicas utilizam o recurso de forma melódica.

Dos vocais melódicos, qual gênero se destaca?



Figura 11: Quantidade de músicas que utilizam vocal de forma melódica por gênero do vocal

Dentre as músicas que utilizam vocal melódico, a maior parte delas, 61 músicas, possuem vocais masculinos, porém também existem quantidades significativas, 40 e 43, que utilizam vocais femininos, ou que possuem os dois gêneros em seus vocais, respectivamente.

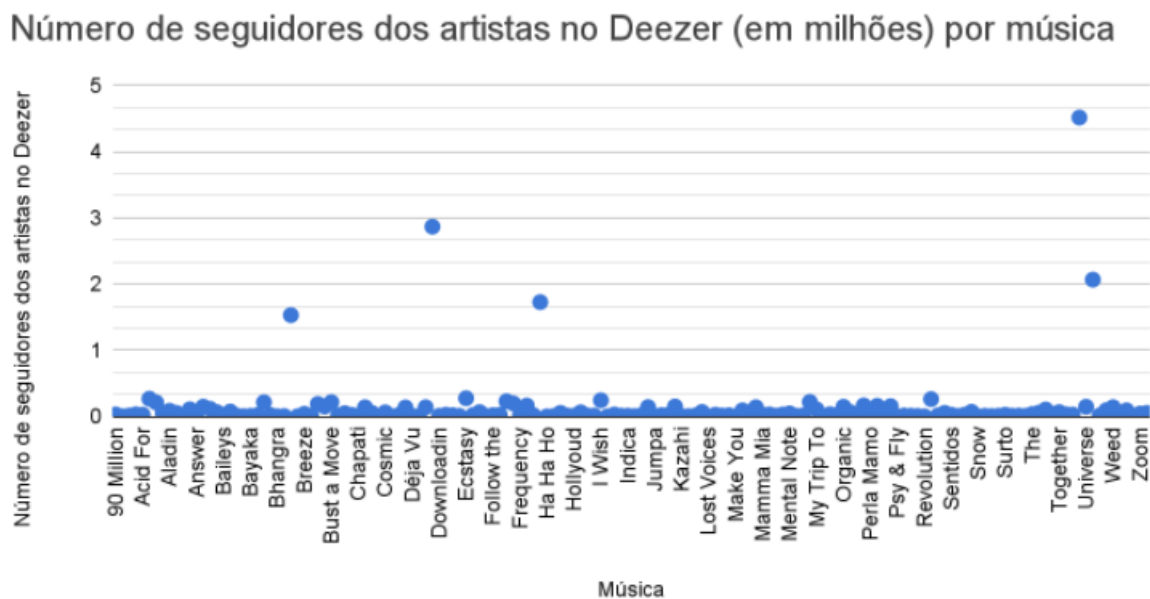


Figura 12: Número de seguidores dos artistas no *Deezer* (em milhões) por música

O *Deezer* é a segunda plataforma com menor expressão desta pesquisa. Em 96,75% das músicas, a base de fãs têm até 300 mil seguidores, tendo apenas 5 casos que ultrapassam essa marca.

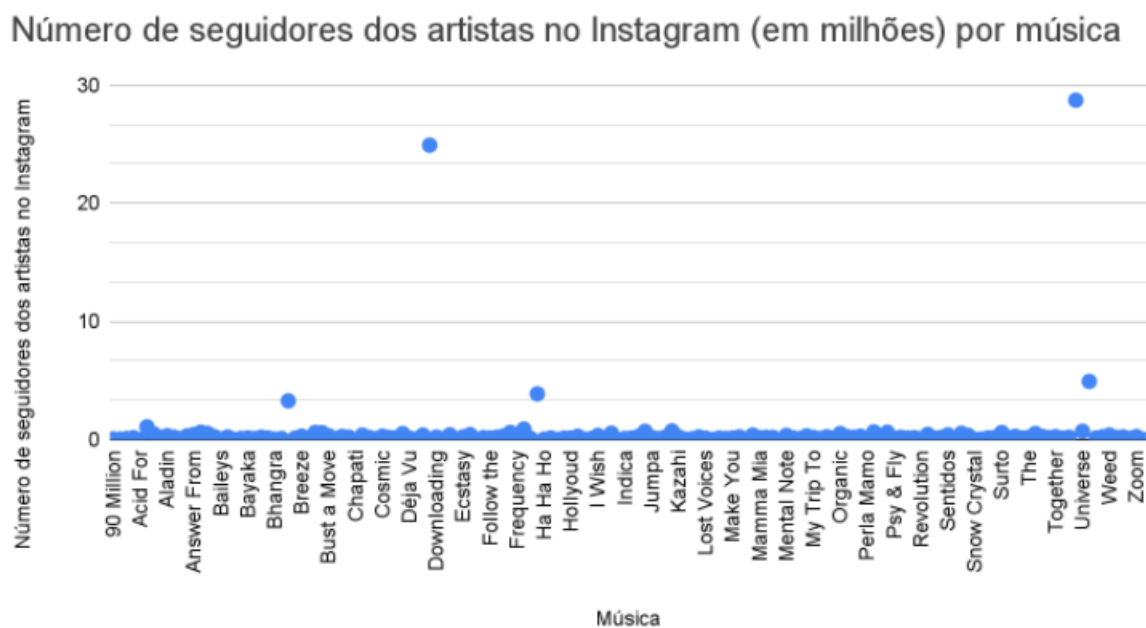


Figura 13: Número de seguidores dos artistas no *Instagram* (em milhões) por música

Em 96,75% dos casos, a soma das bases de fãs dos artistas no *Instagram* têm até 1 milhão de seguidores, tendo apenas 5 casos que ultrapassam essa marca.

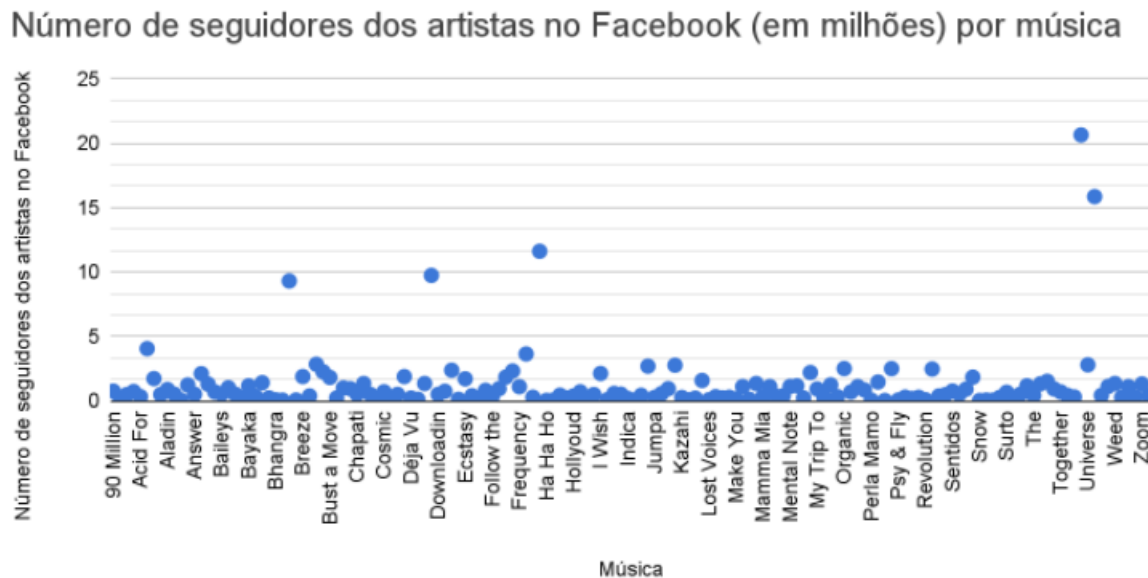


Figura 14: Número de seguidores dos artistas no *Facebook* (em milhões) por música

Em 96,75% dos casos, a soma das bases de fãs dos artistas no *Facebook* têm até 5 milhões de seguidores, tendo apenas 5 casos que ultrapassam essa marca.

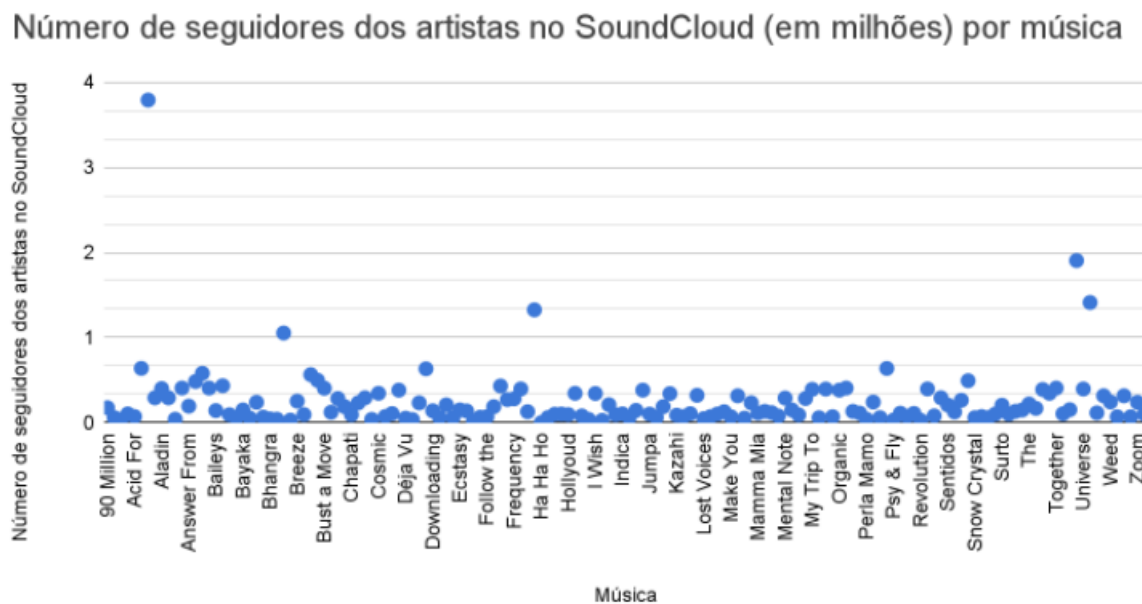


Figura 15: Número de seguidores dos artistas no *Soundcloud* (em milhões) por música

Pode-se observar que a grande maioria, 96,75%, da soma dos artistas envolvidos nas músicas apresentam uma base de fãs de até 650 mil seguidores no *Soundcloud*, tendo apenas 5 casos onde este número é ultrapassado.

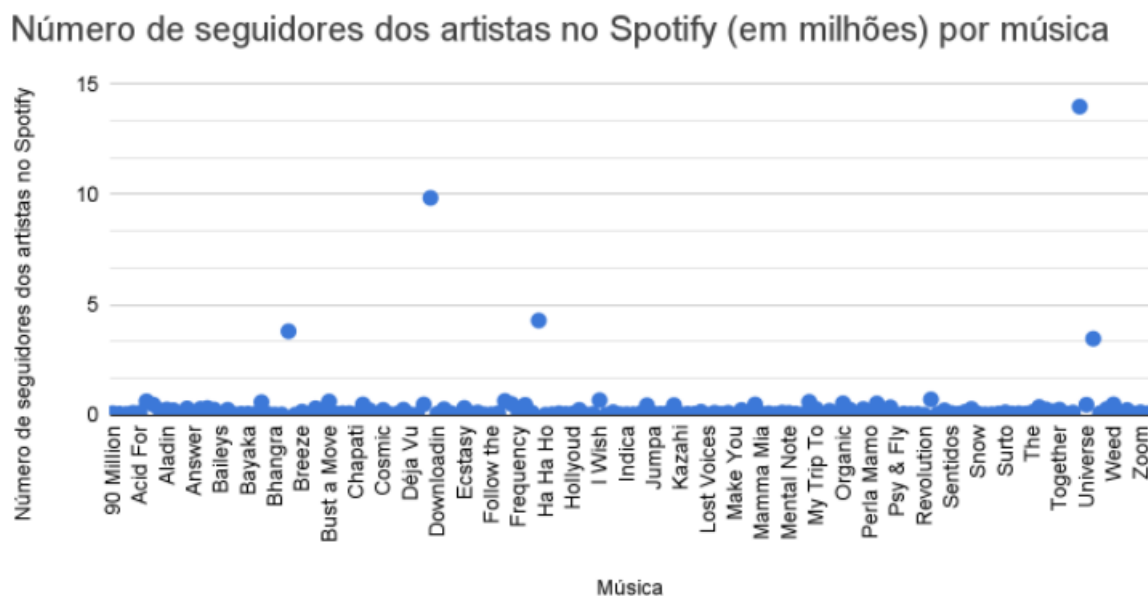


Figura 16: Número de seguidores dos artistas no *Spotify* (em milhões) por música

Dentre as músicas analisadas, 96,75% da soma dos seguidores dos artistas têm até 700 mil no *Spotify*, tendo apenas 5 casos onde este número é ultrapassado.

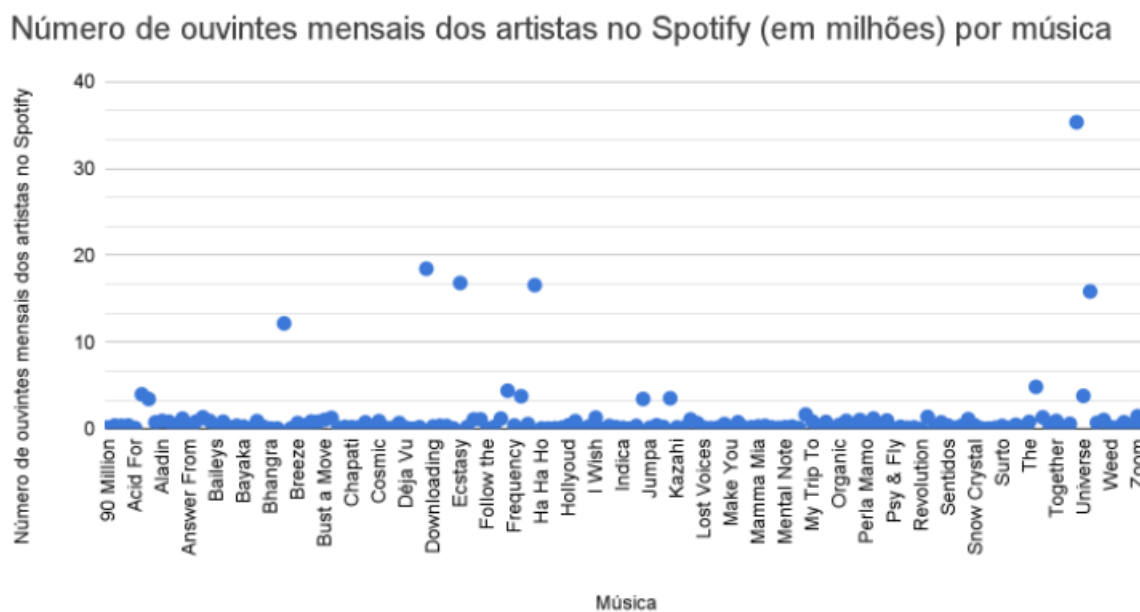


Figura 17: Número de ouvintes mensais dos artistas no *Spotify* (em milhões) por música

Com relação a quantidade de ouvintes mensais dos artistas no *Spotify*, por música. Têm-se que 96,10%, ou 148 músicas, têm até 5 milhões de ouvintes mensais. Em apenas 6 casos, este número é ultrapassado.

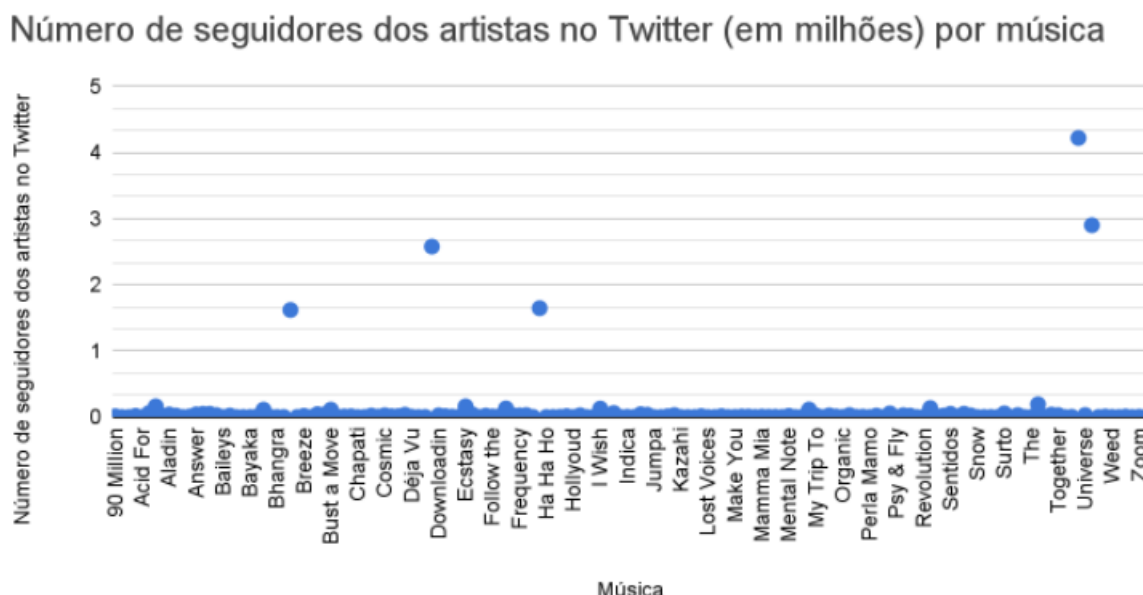


Figura 18: Número de seguidores dos artistas no *Twitter* (em milhões) por música

O *Twitter* é a plataforma com menor expressão desta pesquisa. Em 96,75% dos casos a base de fãs dos artistas têm até 200 mil seguidores, tendo apenas 5 casos que ultrapassam essa marca.

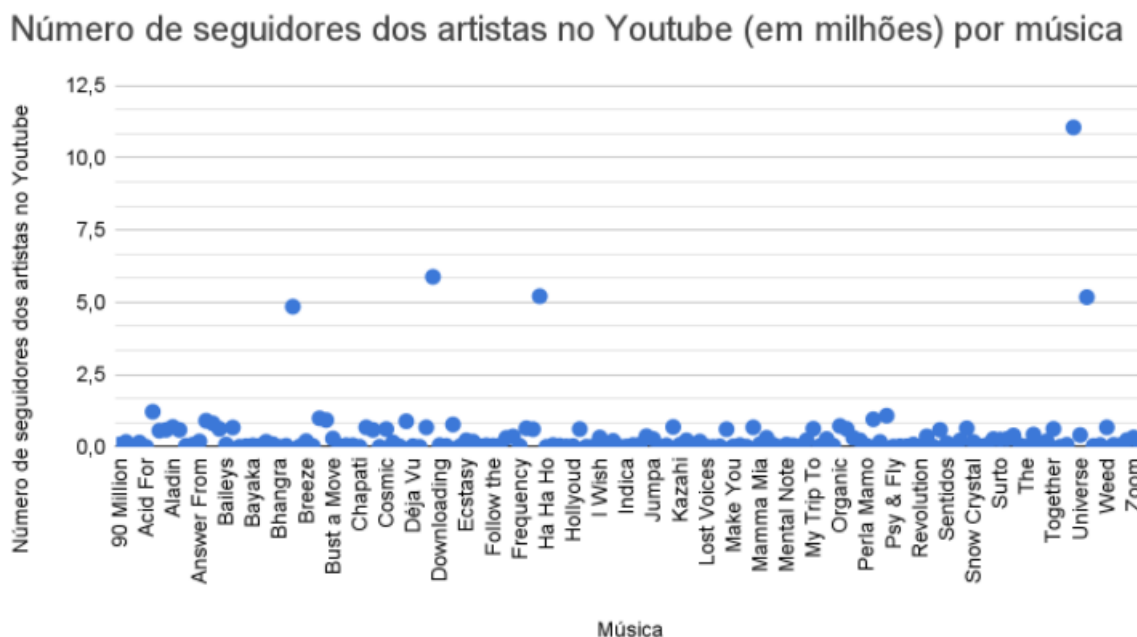


Figura 19: Número de seguidores dos artistas no *Youtube* (em milhões) por música

Ao se analisar a soma de seguidores dos artistas envolvidos em cada música pode-se observar que a grande maioria, 95,45%, têm até 1 milhão de seguidores no *Youtube*. Tendo apenas 7 casos onde este número é ultrapassado.

Número de seguidores dos selos no Instagram (em milhões) por música

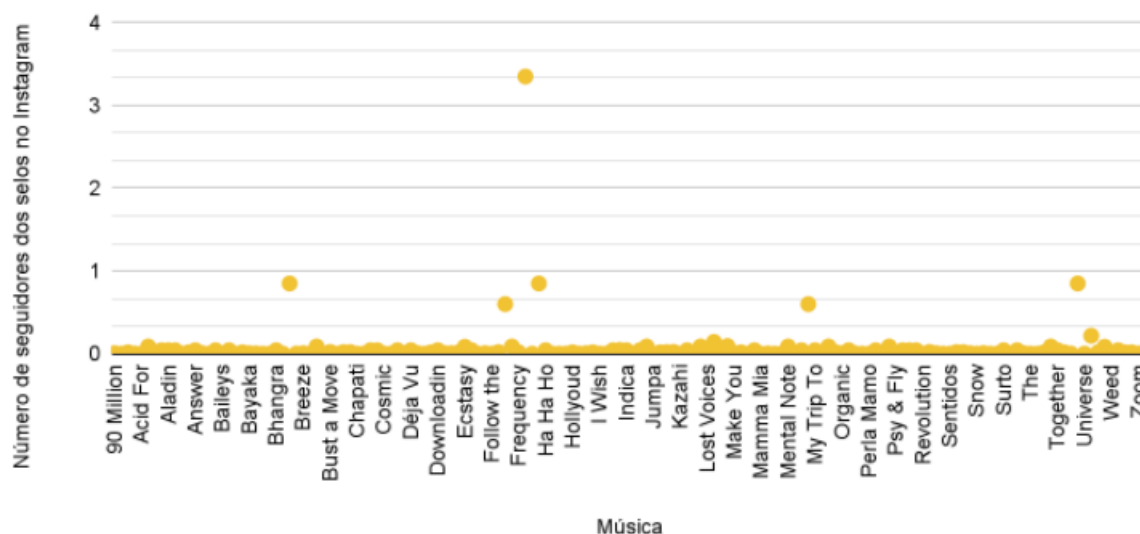


Figura 20: Número de seguidores dos selos no *Instagram* (em milhões) por música

O número de seguidores dos selos no *Instagram*, por música, não ultrapassa 250 mil em 96,10% dos casos. Em apenas 6 casos, este número é superado.

Número de seguidores dos selos no Facebook (em milhões) por música

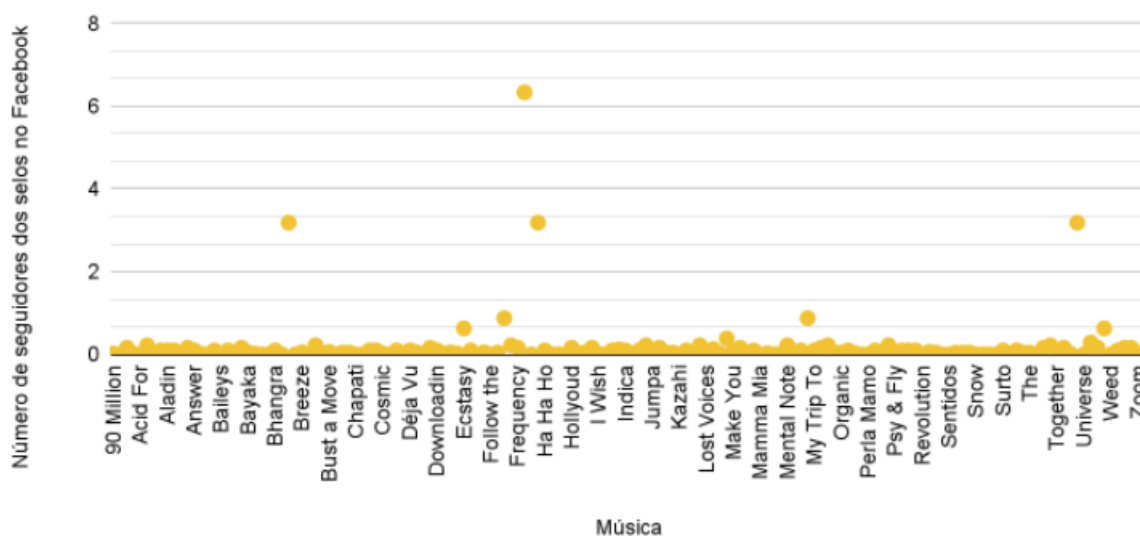


Figura 21: Número de seguidores dos selos no *Facebook* (em milhões) por música

Em apenas 4 casos o número de seguidores dos selos no *Facebook* ultrapassa 900 mil. Dentre os outros 146, ou seja 94,81%, têm menos de 400 mil seguidores.

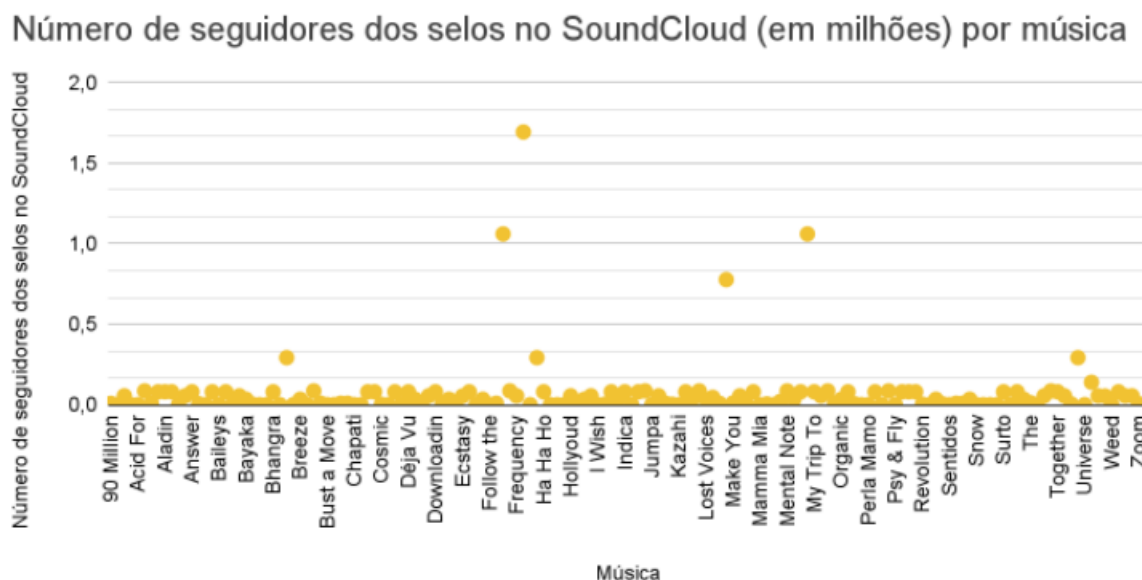


Figura 22: Número de seguidores dos selos no *Soundcloud* (em milhões) por música

Em 94,81% dos casos os selos da música apresentam uma base de fãs com menos de 100 mil seguidores no *Soundcloud*, tendo apenas 8 faixas que superam este valor.

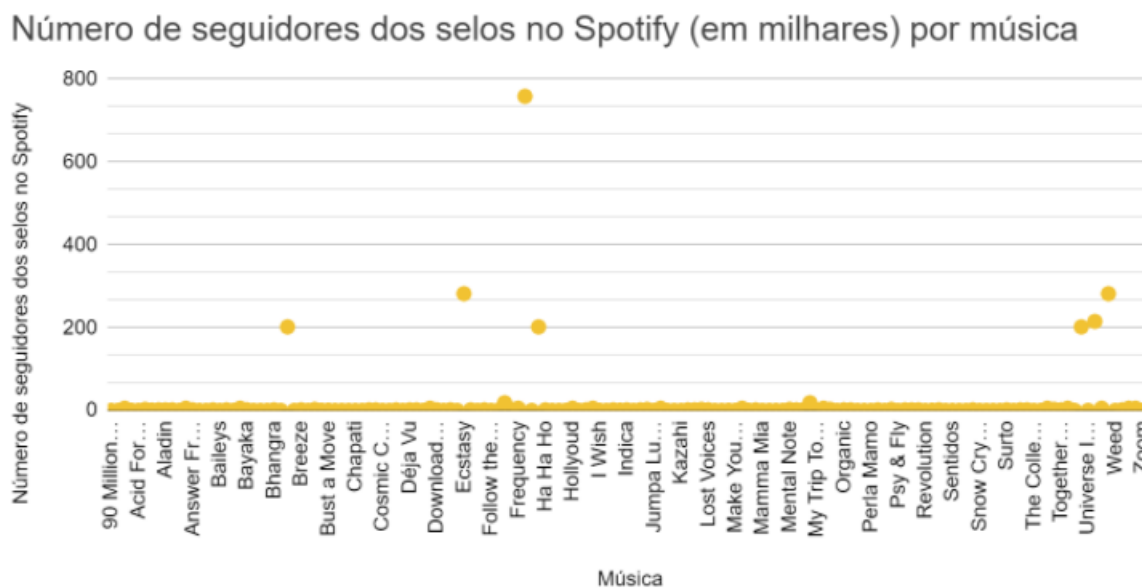


Figura 23: Número de seguidores dos selos no *Spotify* (em milhares) por música

A grande maioria, 94,16%, dos selos possuem menos de 5,5 mil seguidores no *Spotify*. 2 ultrapassam os 18 mil seguidores e 7 faixas superam o número de 200 mil seguidores.

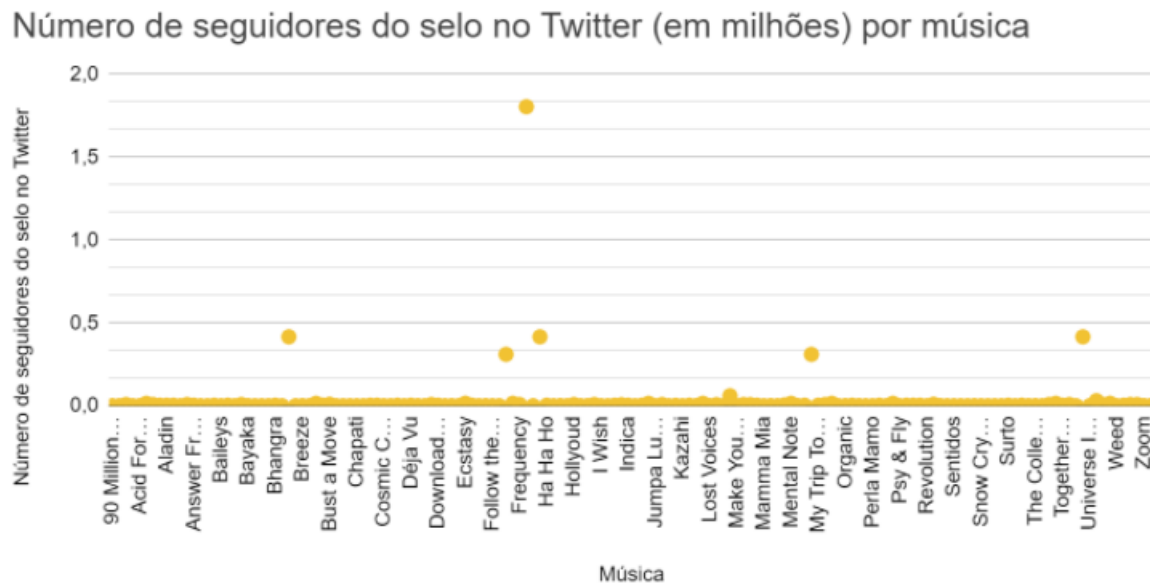


Figura 24: Número de seguidores do selo no *Twitter* (em milhões) por música

O número de seguidores dos selos no *Twitter*, por música, não ultrapassa 60 mil em 96,10% dos casos. Em apenas 6 casos, este número é superado.

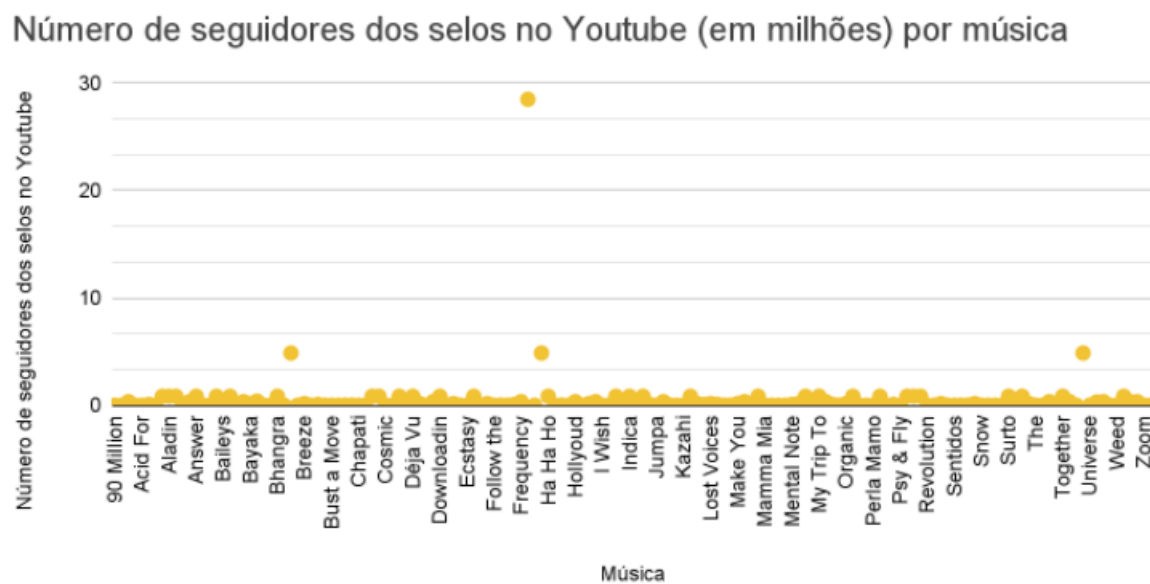


Figura 25: Número de seguidores dos selos no *Youtube* (em milhões) por música

Em apenas 4 casos, o número de seguidores do selo no *Youtube* ultrapassa 900 mil. Os outros 97,4% possuem menos de 900 mil seguidores.

Plays no SoundCloud (em milhões) por música

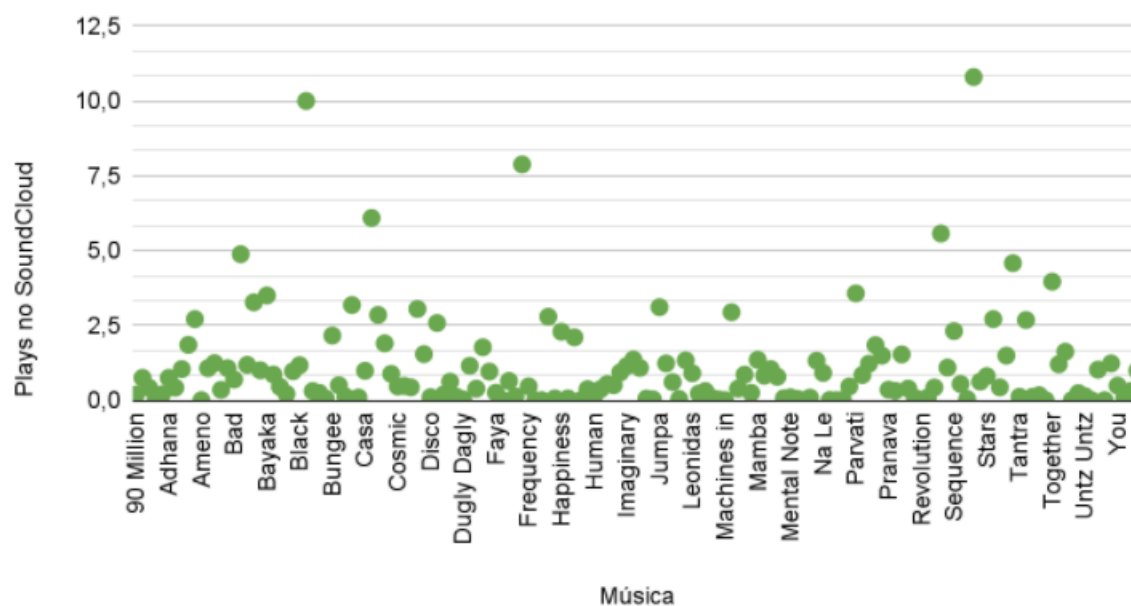


Figura 26: *Plays no Soundcloud* por música

Todas as métricas foram correlacionadas com as métricas de sucesso escolhidas. 149 músicas, ou seja, 96,75% possuía menos de 5 milhões de *plays* no *Soundcloud*, no período de coleta de dados desta pesquisa.

Número de curtidas no SoundCloud (em milhares) por música

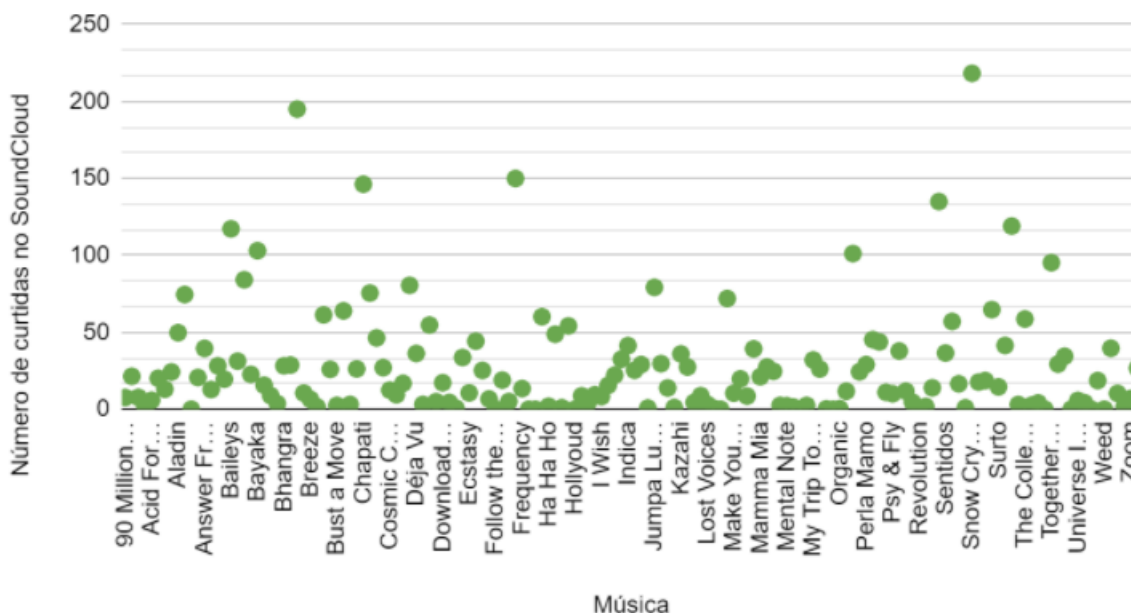


Figura 27: Número de curtidas no *Soundcloud* (em milhares) por música

O número de curtidas no *Soundcloud* possui uma distribuição mais homogênea comparado com as outras métricas, variando desde dezenas até uma faixa próxima aos 200 mil.

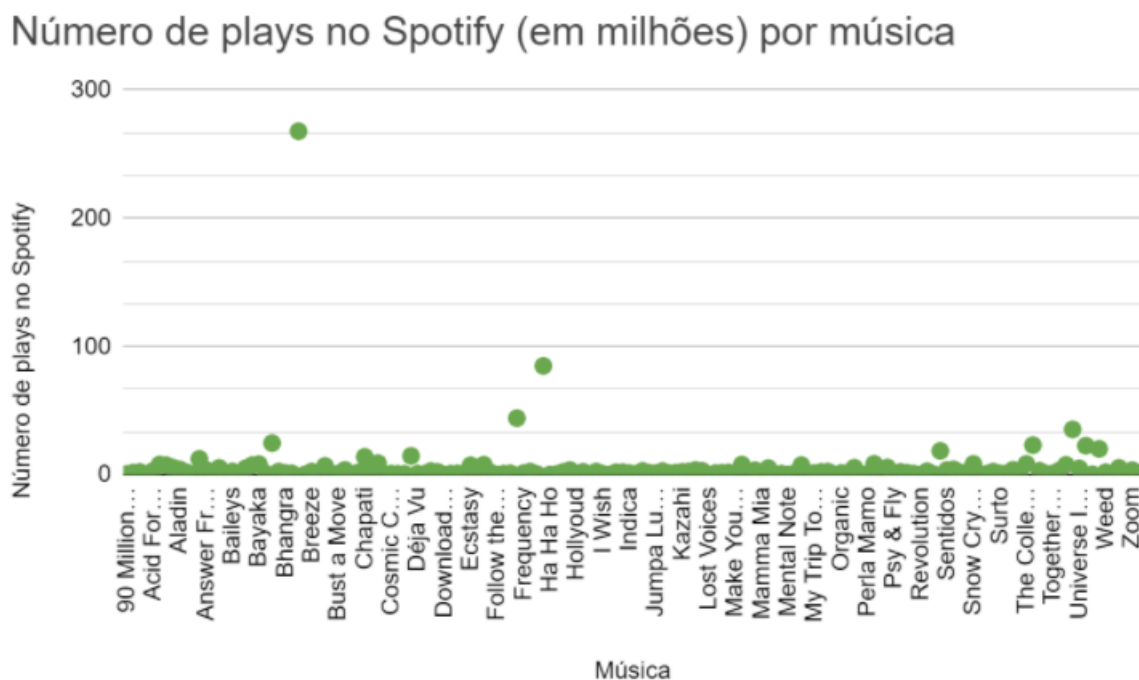


Figura 28: Número de *plays* no *Spotify* (em milhões) por música

151 músicas, ou seja 98,05% dos casos, possui até 40 milhões de *plays* no *Spotify*, tendo 3 exceções.

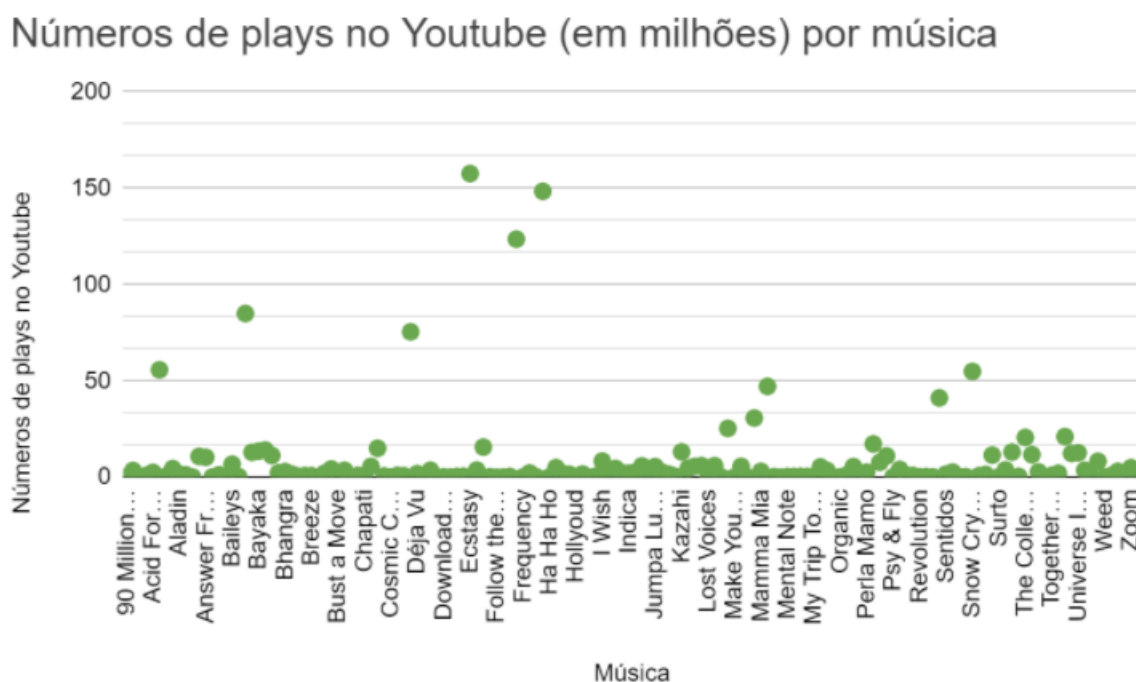


Figura 29: Número de *plays* no *Youtube* (em milhões) por música

Do repertório da pesquisa, 125 músicas têm até 10 milhões de *plays* no *Youtube*, 16 músicas têm entre 10 e 20 milhões de *plays*, 10 têm de 20 a 100 milhões *plays* e 3 têm mais de 100 milhões de *plays* na plataforma.

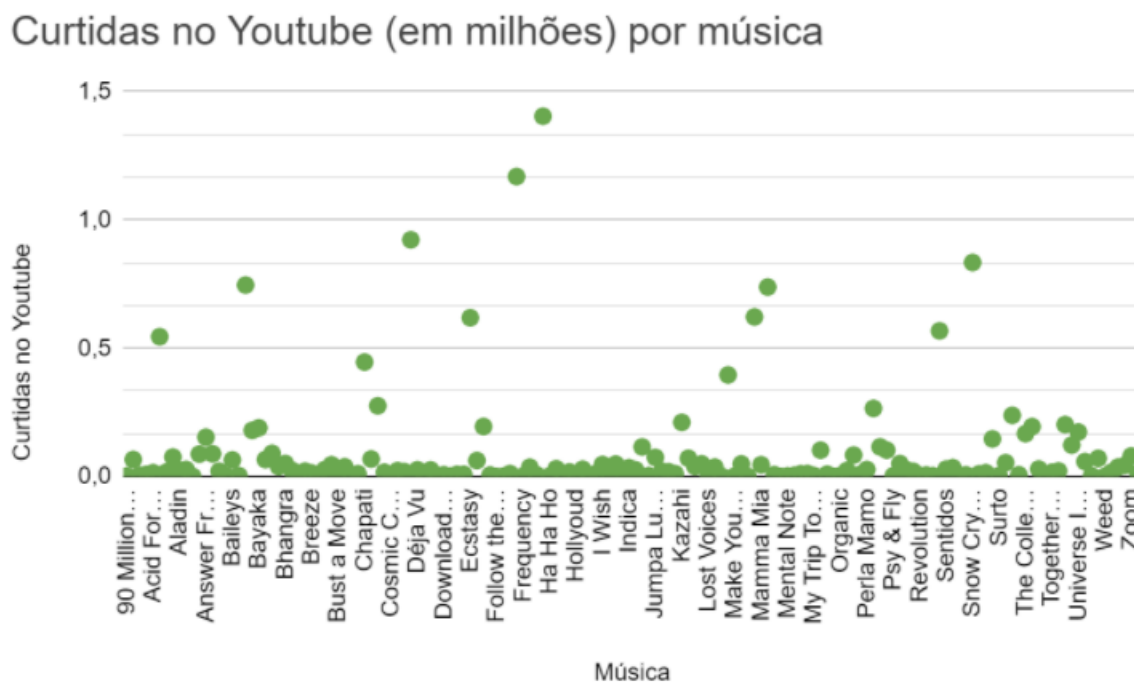
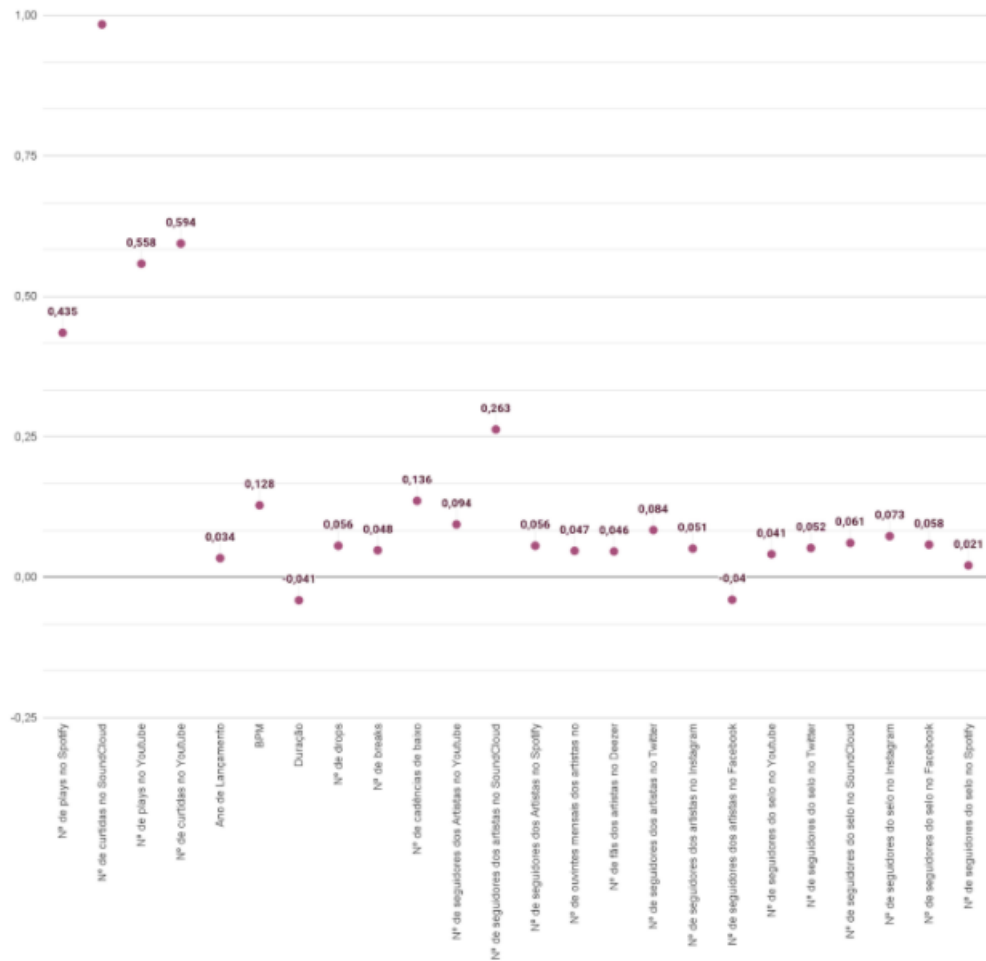


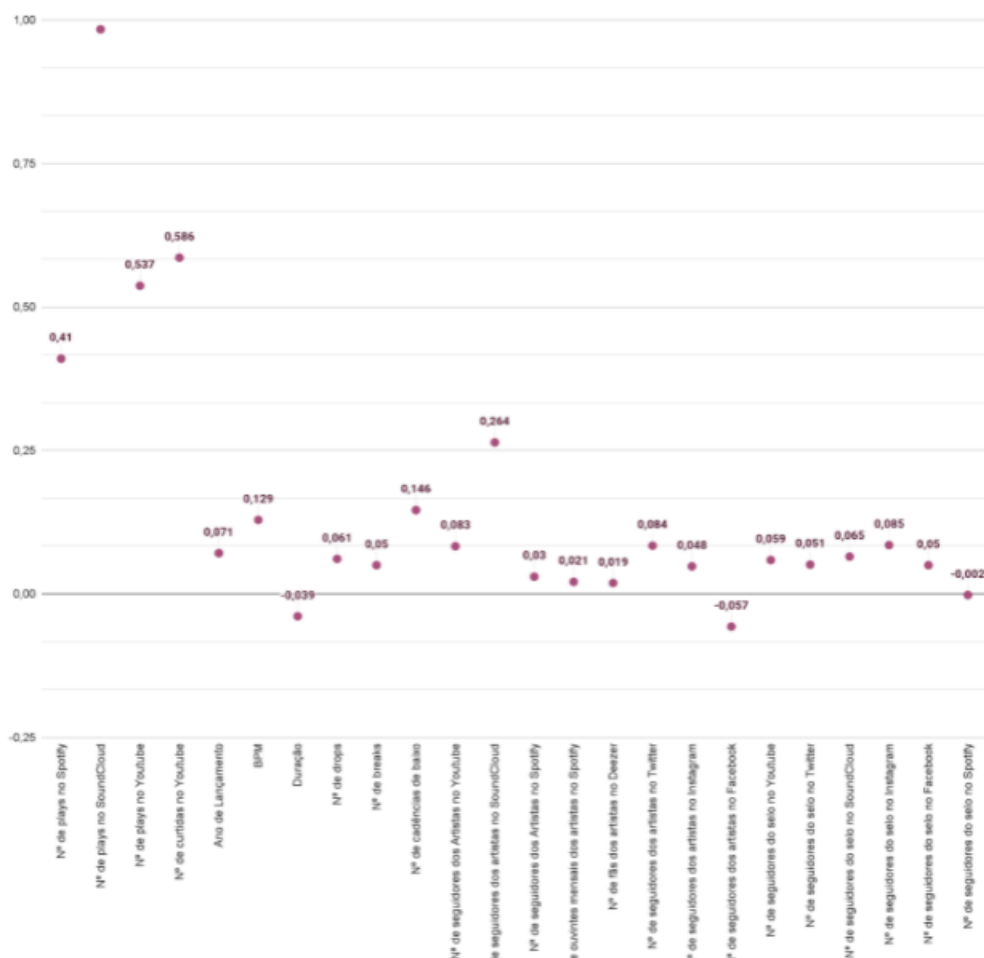
Figura 30: Número de curtidas no *Youtube* (em milhões) por música

Das 154, 144 músicas têm menos de 500 mil curtidas no *Youtube*, tendo apenas 10 que superam o valor.

Correlações com plays no Soundcloud

Figura 31: Correlações com *plays* no *Soundcloud*

Correlações com curtidas no Soundcloud

Figura 32: Correlações com curtidas no *Soundcloud*

Ao analisar as correlações com *plays* e curtidas no *Soundcloud*, pode-se observar que ambas exercem forte correlação entre si. Logo em seguida a plataforma *Youtube* se mostra o critério de maior relevância para ambas as métricas de sucesso. Deve-se dar destaque para o fato de que curtidas no *Youtube* retornou correlação mais forte do que *plays* nos dois casos, indicando que ela realmente é uma métrica relevante de engajamento tanto dentro, quanto fora da plataforma. Em seguida, *plays* no *Spotify* aparece como o quarto fator de maior correlação tanto com *plays*, quanto com curtidas, no *Soundcloud*. As outras métricas retornaram valores menores do que 0,3 indicando que não existe correlação significativa entre as métricas.

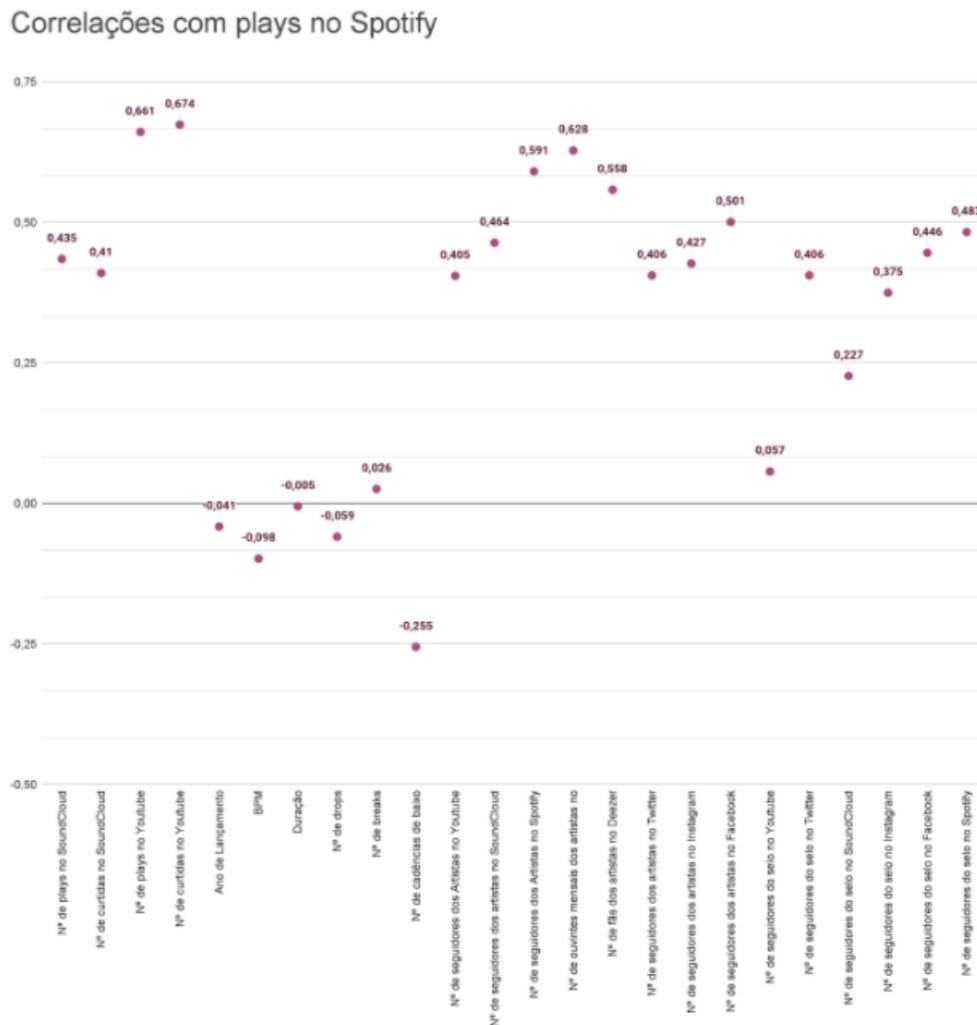
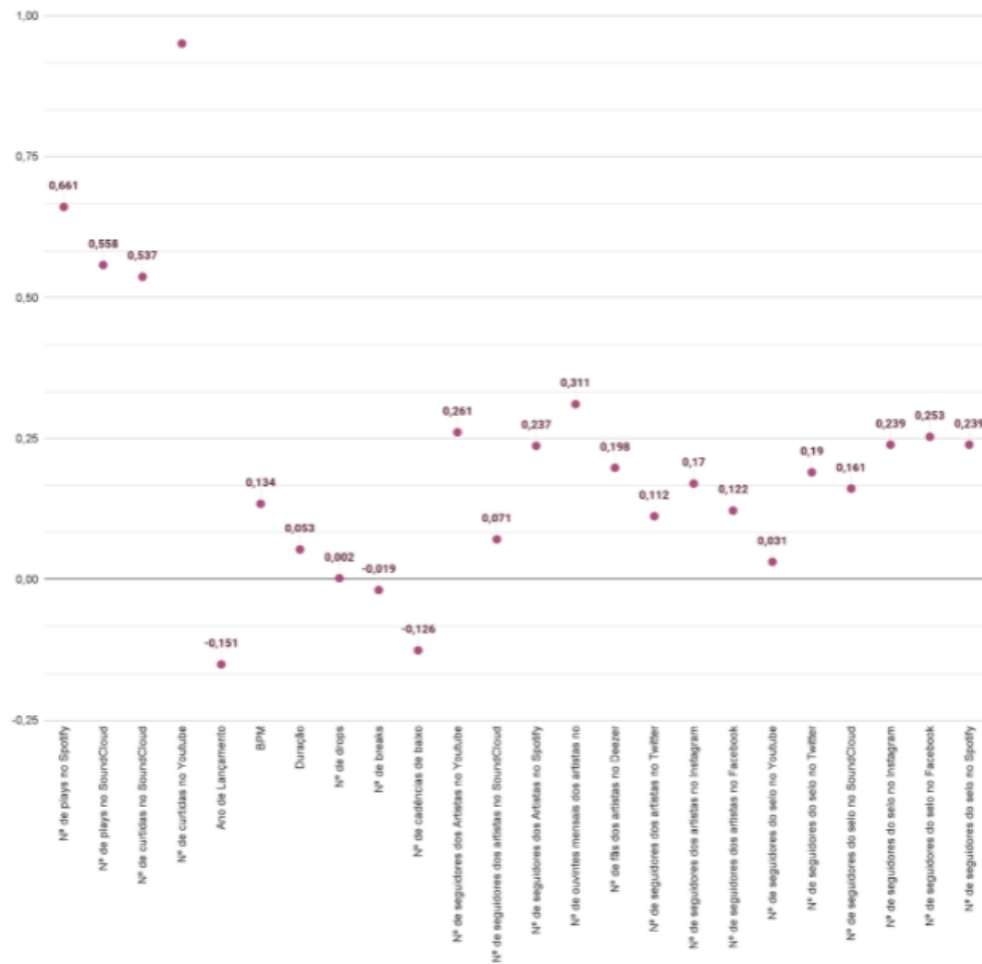


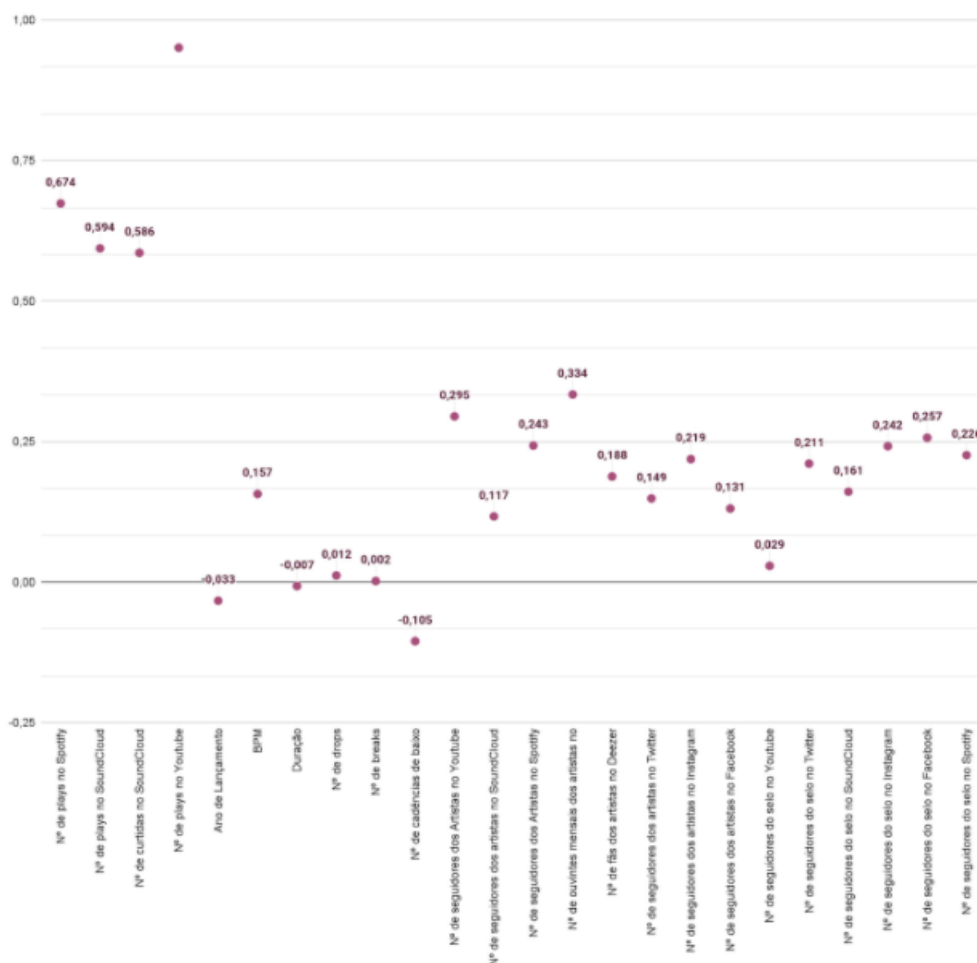
Figura 33: Correlações com *plays* no *Spotify*

As correlações com *plays* no *Spotify* foram as análises que retornaram maior número de métricas com correlações significativas. O *Youtube* se destacou com as correlações mais fortes, tanto em *plays* como em curtidas. Em seguida, com fator bem próximo ao *Youtube*, o número de ouvintes mensais no *Spotify* aparece como terceiro fator de correlação mais forte, vindo antes do número de seguidores dos artistas na própria plataforma, que representa o quarto fator de correlação mais forte. A base de fãs no *Deezer* se mostra mais impactante do que quantidade de *plays* e curtidas no *Soundcloud*, vindo como quinto fator de maior influência. Os critérios de *plays* e curtidas no *Soundcloud* e critérios relativos à base de fãs, tanto dos artistas quanto dos selos, possuem ligações semelhantes com a métrica de sucesso analisada, com exceção de seguidores dos selos no *Youtube* e *Soundcloud*, que apresentam resultados insignificantes. Também é de se destacar que os critérios característicos e estruturais das músicas não apresentaram resultados significativos.

Correlação com Plays no Youtube

Figura 34: Correlações com *plays* no *Youtube*

Correlações com curtidas no Youtube

Figura 35: Correlações com curtidas no *Youtube*

Assim como no *Soundcloud*, *plays* e curtidas dentro da plataforma *Youtube* têm correlações muito fortes entre si. Em seguida, *plays* no *Spotify* é a métrica correlacionada mais forte. Em terceiro e quarto lugar, *plays* e curtidas no *Soundcloud* são, respectivamente, as outras métricas que indicam correlações significativas. Da mesma forma que ocorreu no *Soundcloud*, as outras métricas possuem correlações praticamente insignificantes demonstrando que, de fato, as plataformas se relacionam entre si e com o público de maneira expressiva e critérios como base de fãs dos artistas, dos selos e aspectos estruturais das músicas não possuem tanta interferência em relação às métricas de sucesso selecionadas neste estudo.

A seguir estão os resultados relativos aos T-testes e ANOVAs realizados para entender aspectos qualitativos da pesquisa. Vale ressaltar que os dados a seguir são o resultado do logaritmo de base 10 dos dados quantitativos coletados. Isto foi feito para se adequar ao formato de T-teste e ANOVA, onde é necessário se adequar os dados de maneira o mais próxima possível de uma distribuição gaussiana.

Tabela 1: T-teste relativo a ser, ou não um *remix*

	É um remix?	Sim	Não	
	Nº de músicas	29	124	
Plays no Spotify	Média	6,45	6,26	P valor = 1,007
	Desvio padrão	0,53	0,56	
Plays no Soundcloud	Média	5,56	5,69	P valor = 0,4327
	Desvio padrão	0,77	0,78	
Curtidas no Soundcloud	Média	4,00	4,14	P valor = 0,3702
	Desvio padrão	0,72	0,74	
Plays no Youtube	Média	5,70	6,20	P valor = 0,0101
	Desvio padrão	1,19	0,84	
Curtidas no Youtube	Média	4,01	4,40	P valor = 0,0215
	Desvio padrão	1,02	0,75	

Com relação a ser *remix*, *plays* no *Spotify* se destaca como o único critério onde ser *remix* possui média de *plays* maior do que as músicas originais. Em todos os outros critérios, ser original apresentou uma média maior, indicando que possuem mais sucesso. Porém, ao realizar o T-teste, o *Youtube* foi a única plataforma que retornou P valores menores do que 0,05, ou 5%, indicando que a análise é realmente confiável. Como os outros critérios deram resultados acima do limite, as análises se mostraram inconclusivas, não havendo confiabilidade nos dados.

Tabela 2: T-teste relativo a ter, ou não, vocal

	Tem vocal?	Sim	Não	
	Nº de músicas	114	9	
Plays no Spotify	Média	6,29	6,36	P valor = 0,7258
	Desvio padrão	0,56	0,55	
Plays no Soundcloud	Média	5,66	5,73	P valor = 0,8052
	Desvio padrão	0,78	0,76	
Curtidas no Soundcloud	Média	4,10	4,20	P valor = 0,7118
	Desvio padrão	0,74	0,67	
Plays no Youtube	Média	5,97	5,97	P valor = 0,6567
	Desvio padrão	0,93	0,99	
Curtidas no Youtube	Média	4,33	4,18	P valor = 0,5891
	Desvio padrão	0,83	0,77	

Em relação a ter ou não vocal, músicas sem vocal possuíram melhores resultados em números de *plays* no *Spotify* e número de *plays* e curtidas no *Soundcloud*. A análise retornou valores iguais para *plays* no *Youtube*, indicando ser indiferente para o critério e curtidas no *Youtube* possuiu resultados melhores em músicas com vocal. No entanto, todas retornaram P valores maiores que 5%, indicando que os resultados não possuem precisão.

Tabela 3: T-teste relativo a ter um vocal melódico, ou não

	O vocal é melódico?	Sim	Não	
	Nº de músicas	126	19	
Plays no Spotify	Média	6,25	6,56	P valor = 0,0248
	Desvio padrão	0,54	0,63	
Plays no Soundcloud	Média	5,68	5,62	P valor = 0,7703
	Desvio padrão	0,78	0,78	
Curtidas no Soundcloud	Média	4,12	4,05	P valor = 0,7419
	Desvio padrão	0,74	0,73	
Plays no Youtube	Média	6,14	5,93	P valor = 0,3669
	Desvio padrão	0,92	0,97	
Curtidas no Youtube	Média	4,38	4,02	P valor = 0,075
	Desvio padrão	0,81	0,88	

No critério do vocal ser melódico ou não, *plays* no *Spotify* foi o único critério que retornou valores maiores nas músicas que não utilizam o vocal de forma melódica e retornou um P valor dentro da margem de confiabilidade. Em todos os outros critérios ser melódico apresentou resultados melhores comparados a não ser, mas novamente retornaram P valores maiores do que os limites de confiabilidade, sendo, portanto, um resultado inconclusivo.

Os testes relativos às ANOVAs são extensos demais para apresentar no corpo deste relatório. Por isso, os dados e resultados numéricos podem ser encontrados através do seguinte link: <https://github.com/leo-mori/TCC2021>, no arquivo "Resultado Final - TFC Leonardo Mori". As ANOVAs podem ser achadas nas abas "Qual o gênero do vocal?", "Qual o BPM?" e "Qual o tom?". Além disso o arquivo também contém todos os outros dados relativos ao presente trabalho, ou seja, o *dataset*, as correlações, os T-testes. Para a finalidade do trabalho será discutidos os resultados e análises das ANOVAs a seguir.

O teste ANOVA para gênero do vocal o *Spotify* foi o único a retornar que o vocal feminino obteve maior sucesso do que as outras duas categorias. Em todos os outros critérios, o vocal masculino apresentou melhores resultados se comparado à vocais femininos e músicas que possuem os dois gêneros no vocal. É de se observar que o P valor para todos os critérios foi alto, com exceção do *Spotify*, que mesmo assim não chegou ao limite de confiabilidade.

No teste ANOVA para BPM das músicas, o *Spotify* foi o único que retornou 146 como o BPM de maior sucesso. Em todos os outros critérios, o BPM com melhor desempenho foi 185. Mas, assim como o gênero do vocal, o P valor de todos os critérios foi alto e os BPM com melhores resultados apresentaram apenas uma amostra. Sendo assim, os resultados se mostram com baixíssima confiabilidade.

No teste ANOVA para tom da música, o G menor obteve unanimidade como melhor resultados, porém, assim como nos outros casos, o P valor ficou longe da margem de 5% e novamente o tom analisado foi relativo a uma única amostra, indicando que esses resultados são inconclusivos.

5 Conclusões

Este estudo teve como objetivo analisar a correlação entre métricas de alcance e engajamento, diretamente relacionadas com vendas de músicas, e outros aspectos como estrutura da música e base de fãs dos envolvidos no repertório, focando no nicho do *psy-trance*.

As etapas cumpridas pelo projeto foram realizadas de maneira satisfatória, desde a coleta de dados, o levantamento do *dataset* e, por fim, a análise de correlação e confiabilidade dos dados. É relevante ressaltar que a pesquisa possuiu limitações, sendo que a coleta de dados foi feita manualmente e que, por serem dados voláteis, eles tendem a crescer com o tempo de forma desigual entre as músicas. Além disso, o grau de confiabilidade relativo aos dados qualitativos foi baixo, segundo os testes realizados estatisticamente.

As análises indicaram que as plataformas exercem forte influência entre si, porém, também demonstrou que critérios estruturais da música e base de fãs dos artistas e selos não têm tanta relevância. O que nos leva a concluir que existem aspectos não abordados nesta pesquisa que possivelmente exercem maior influência nos resultados, sendo o marketing a principal hipótese de fator relevante para ser analisado em um estudo futuro. Além disso, os resultados são base argumentativa contra aqueles que desejam fazer sucesso através da criação de músicas através de inteligência artificial, visto que resultados estruturais da música não tiveram correlação significativa com nenhum critério de sucesso analisado no presente estudo.

Como aprimoramento do modelo seria interessante desenvolver a coleta automatizada dos dados nas diferentes plataformas analisadas, além de um espaço amostral maior, com a classificação dos diferentes estágios de maturidade dos artistas.