BOSSA: Um Dataset Multimodal de Popularidade Musical Regional no Spotify Brasileiro

Gabriel H. Silva[1], Leonam V. O. L. Almeida[1], Helen C. S. C. Lima[1], Filipe A. S. Moura[2]

[1]Instituto de Ciências Exatas e Aplicadas
Universidade Federal de Ouro Preto (UFOP)

[2]Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG)

{gabriel.hs,leonam.almeida}@aluno.ufop.edu.br

helen@ufop.edu.br, filipe.moura@dcc.ufmg.br

Abstract. *This paper introduces* ***BOSSA*** *(Brazilian Observatory of Streaming and Song Analytics), a multimodal dataset that covers the most popular tracks on Spotify Charts in 17 Brazilian cities from January 2022 to December 2024. BOSSA integrates weekly city-level ranking data for 16,148 tracks, shallow audio features (ReccoBeats API), deep audio representations (MERT-v1-330M), and automatically transcribed lyrics (Whisper v3). Three exploratory analyses validate the dataset: a UMAP projection of the deep feature space reveals unsupervised geographic clustering. City-level cosine similarity confirms that musical proximity mirrors cultural and geographic proximity (e.g., Campinas-São Paulo, Belém-São Paulo). Spearman correlation between shallow and deep feature distances confirms that both modalities are complementary. BOSSA is the first Brazilian dataset to combine all four modalities at city-level granularity.*

Resumo. *Este artigo apresenta o* ***BOSSA*** *(Brazilian Observatory of Streaming and Song Analytics), um conjunto de dados multimodal que cobre as faixas mais populares no Spotify Charts em 17 cidades brasileiras de janeiro de 2022 a dezembro de 2024. O BOSSA integra dados de ranking semanal por cidade para 16.148 faixas, shallow audio features (API ReccoBeats), representações profundas de áudio (MERT-v1-330M) e letras transcritas automaticamente (Whisper v3). Três análises exploratórias validam o repositório de dadost: uma projeção UMAP revela agrupamentos geográficos não supervisionados, a similaridade de cosseno entre cidades reflete proximidade cultural (e.g., Campinas-São Paulo, Belém-São Paulo) e uma correlação de Spearman confirma a complementaridade entre as duas modalidades de áudio. O BOSSA é o primeiro conjunto de dados brasileiro a oferecer essa combinação de modalidades com granularidade por cidade.*

1. Introdução

O *streaming* musical consolidou-se como o principal modelo de consumo de música na atualidade, transformando a forma como preferências culturais se manifestam e podem

ser estudadas em escala [Bello and Garcia 2021, Way et al. 2020]. Plataformas como o Spotify geram rankings de popularidade segmentados por cidade, constituindo registros das dinâmicas musicais locais[1]. No Brasil, país de dimensões continentais e grande diversidade cultural, diferentes regiões apresentam padrões de consumo musical distintos [Moura et al. 2024, Moura et al. 2025, Oliveira et al. 2025], evidenciando o potencial desses dados para a compreensão das preferências musicais locais.

A área de *Hit Song Science* [Seufitelli et al. 2023b], que investiga os fatores associados ao sucesso musical, tem se apoiado em *features* acústicas padronizadas como dançabilidade, energia e valência. Embora Bases de dados públicos voltados ao sucesso musical existam na literatura [Bertoni and Lemos 2021, Oliveira et al. 2021, Seufitelli et al. 2023a], a escassez de bases que integrem *features* de áudio com dados de *streaming* segmentados por região persiste [Sebastian et al. 2024], limitando estudos comparativos e dificultando a reprodução de experimentos. Esse cenário é agravado pela descontinuação do fornecimento de *audio features* pela API oficial do Spotify e por atualizações nos termos de uso do Spotify, que impedem coletas recentes dos Charts. A obtenção de letras musicais representa um desafio adicional, abordado por sistemas de reconhecimento automático de fala [Zhuo et al. 2023].

Diante desse cenário, este trabalho apresenta o **BOSSA**: *Brazilian Observatory of Streaming and Song Analytics*, um dataset com dados das músicas mais populares em 17 cidades brasileiras no Spotify Charts, cobrindo o período de janeiro de 2022 a dezembro de 2024. O dataset integra: dados de popularidade e posicionamento nos rankings semanais por cidade, *shallow audio features* extraídas via API do ReccoBeats, *deep features* extraídas com o modelo MERT [Li et al. 2024] e letras transcritas automaticamente com o Whisper v3 [Radford et al. 2023].

O restante deste artigo está organizado da seguinte forma: a Seção 2 revisa os trabalhos relacionados; a Seção 3 descreve a metodologia de coleta e curadoria; a Seção 4 apresenta as análises exploratórias; e as Seções 5, 6 e 7 abordam considerações éticas, dicionário de dados e disponibilização do BOSSA.

2. Trabalhos Relacionados

Esta seção revisa os trabalhos relacionados ao BOSSA em duas frentes: Bases de dados musicais públicos e estudos sobre análise de consumo musical e predição de popularidade.

2.1. Datasets Musicais

Três datasets musicais previamente publicados servem como forte referência para construção do BOSSA. [Bertoni and Lemos 2021] apresentaram três datasets de canções brasileiras populares e não populares (2014–2019), focados em *features* acústicas do Spotify para classificação de sucesso. [Oliveira et al. 2021] introduziram o MUHSIC, com informações temporais de sucesso musical combinando metadados do Spotify e do YouTube, mas sem segmentação regional. [Seufitelli et al. 2023a] propuseram o MGD+, dataset de gêneros musicais enriquecido com redes de coocorrência baseadas em sucesso, também sem distinção por localidade.

O BOSSA se diferencia em três dimensões. Primeiro, é o único com granularidade **por cidade**. Segundo, resolve a descontinuação da API do Spotify combinando

[1] `https://charts.spotify.com/charts/overview/br`

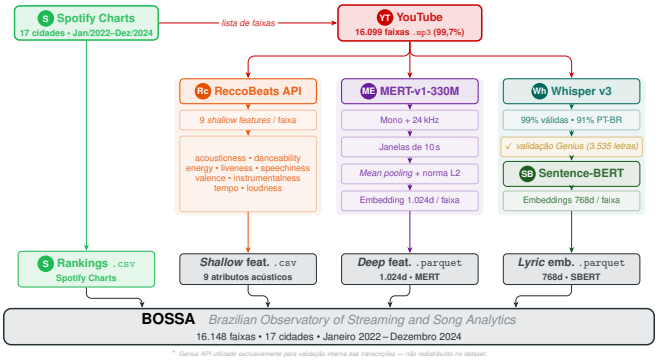


Figura 1. Metodologia de Coleta do BOSSA

ReccoBeats para *shallow features* e MERT para *deep features* extraídas diretamente do áudio. Terceiro, é o único a incorporar **letras transcritas automaticamente**, viabilizando análises textuais e multimodais.

2.2. Análise de Consumo Musical e Predição de Popularidade

O campo de *Hit Song Science* [Seufitelli et al. 2023b] combina *features* acústicas com dados de comportamento do público para prever sucesso musical. Estudos globais demonstraram que a proximidade geográfica e linguística condiciona as preferências nos charts [Bello and Garcia 2021, Way et al. 2020]. No contexto brasileiro, [Moura et al. 2025] identificaram perfis acústicos distintos por cidade a partir do Spotify Charts, ressaltando a escassez de estudos detalhados sobre distribuição geográfica de preferências musicais, trabalho que o BOSSA complementa ao adicionar *deep features* e letras. Do lado preditivo, [Sebastian et al. 2024] e [Rompolas et al. 2024] demonstraram que abordagens multimodais superam unimodais na predição de popularidade musical.

3. Coleta dos dados

Essa seção descreve o passo a passo da metodologia de coleta e curadoria do BOSSA. Todo processo é ilustrado pela Figura 1: desde a coleta dos rankings provenientes do Spotify Charts até a curadoria de todas as features computadas.

3.1. Coleta e processamento de dados do Spotify

Para analisar as músicas mais populares em cada cidade, utilizamos o *Spotify Charts*[2], serviço que fornece rankings semanais com as 100 músicas mais ouvidas, segmentados por cidade. Especialmente, coletamos o ranking "Local Pulse" – *playlist* de músicas mais populares naquele local – das 17 cidades brasileiras registradas na plataforma: Belém, Belo Horizonte, Brasília, Campinas, Campo Grande, Cuiabá, Curitiba, Florianópolis, Fortaleza, Goiânia, Manaus, Porto Alegre, Recife, Rio de Janeiro, Salvador, São Paulo e Uberlândia. O período de coleta vai de 01/01/2022 a 07/12/2024, totalizando 6.486 artistas únicos e 16.148 músicas distintas.

Os áudios foram coletados via *youtube-dl*[3] com a *string* {*nome da música* + *nome do artista*}, em busca do vídeo com maior relevância na plataforma e realizando o download apenas do áudio `.mp3` do mesmo. Obtemos os vídeos de 16.099 faixas $(99,7\%)$,

[2] `https://charts.spotify.com/charts/overview/br`
[3] `https://github.com/ytdl-org/youtube-dl`

sendo os $0,3\%$ descartados por conta de falhas por restrição de idade do YouTube. Optamos não adotar uma *persona* no momento da recuperação dos vídeos musicais para evitar a interferência de algoritmos de recomendação da plataforma, o que também impede a recuperação das 49 músicas que possuem restrição de idade registradas.

3.2. Coleta das *shallow features* de áudio

Como a API do Spotify não fornece mais *audio features*, utilizamos a API do ReccoBeats[4], gratuita e de fácil acesso. As *features* extraídas são: ***acousticness***, ***danceability***, ***energy***, ***liveness***, ***speechiness***, ***valence***, ***instrumentalness***, ***tempo*** e ***loudness***. Para avaliar a confiabilidade desta API, comparamos as distribuições com uma base de dados de referência [Moura et al. 2024], considerando 1.274 faixas em comum. Como mostra a Figura 2, *instrumentalness* e *liveness* apresentam maior divergência em relação à referência, enquanto as demais *features* apresentam valores de *Kullback-Leibler* (KL) inferiores a 0.7, indicando que o ReccoBeats produz estimativas consistentes com a referência e adequadas para uso em análises posteriores.

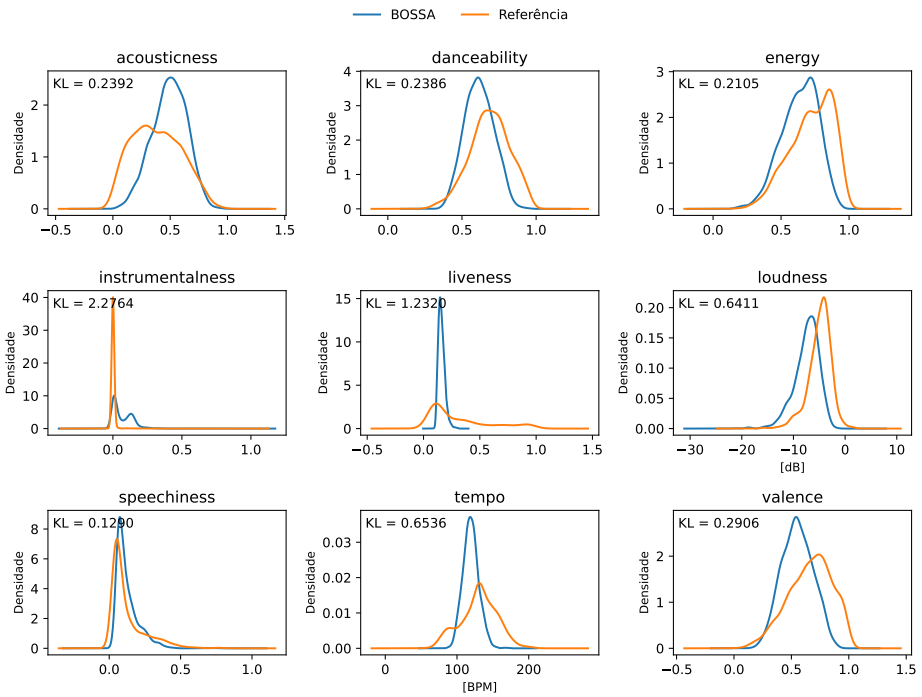


Figura 2. Comparação das distribuições das *shallow features* entre o BOSSA e o dataset de referência (faixas em comum). Para cada *feature*, foi calculada a divergência de *Kullback-Leibler* (KL), quantificando o quanto a distribuição do BOSSA diverge da referência.

3.3. Extração de *deep features* com MERT

Complementarmente às *shallow features*, extraímos representações de alto nível com o **MERT-v1-330M** [Li et al. 2024], modelo autossupervisionado com 330M parâmetros treinado via *Masked Autoencoders* sobre espectrogramas. O pipeline de extração segue

[4]https://reccobeats.com/docs/documentation/introduction

quatro etapas: (i) conversão do áudio para mono e reamostragem para 24 kHz, taxa nativa do modelo; (ii) fatiamento em janelas de 10 segundos para contornar a complexidade quadrática $O(N^2)$ do mecanismo de atenção dos *transformers*; (iii) *mean pooling* temporal sobre as ativações da última camada oculta, gerando um vetor de 1.024 dimensões por janela; e (iv) *mean pooling* entre janelas, produzindo um **embedding único de 1.024 dimensões** por faixa. Os vetores resultantes são centrados pela média do corpus e normalizados em norma L2, tornando o produto interno equivalente à similaridade de cosseno, que é a métrica padrão para recuperação por similaridade musical [Li et al. 2024].

3.4. Coleta das letras das músicas

As letras foram transcritas com o *Whisper v3* [Radford et al. 2023] aplicado diretamente sobre os arquivos `.mp3`. Embora tenham sido exploradas fontes de letras já disponíveis, nenhuma delas apresentou cobertura suficiente para todo o conjunto de músicas. Assim, optou-se pelo uso do *Whisper v3* como uma estratégia padronizada para obtenção das transcrições. Do total processado, 99% geraram transcrições válidas com idioma detectado pelo próprio modelo. Dentre estas, $91,0\%$ das músicas estão em português, $5,7\%$ em inglês e $3,3\%$ foram classificadas em outros idiomas.

Para avaliar a qualidade das transcrições, coletamos letras oficiais via API do Genius[5], obtendo 3.535 letras verificadas do total de 16.148 faixas, utilizadas como *ground truth*. A avaliação foi conduzida com duas métricas complementares: similaridade de cosseno com vetorização TF-IDF e similaridade semântica com Sentence-BERT. O TF-IDF mede a sobreposição direta de termos entre a transcrição e a letra oficial, sendo sensível a variações de vocabulário e grafias, enquanto o Sentence-BERT captura o significado semântico do texto, sendo mais robusto a paráfrases e diferenças superficiais de escrita. As estatísticas de ambas as métricas estão na Tabela 1. Cerca de 44,6% das transcrições atingiram alta similaridade TF-IDF ($\geq$0,70) e 19,0% apresentaram similaridade muito baixa ($<$0,20), reflexo das dificuldades da tarefa de ASR (*Automatic Speech Recognition*) em músicas com estilo vocal distante da fala convencional [Berendes et al. 2024]. No espaço semântico, o Sentence-BERT registrou média de 0,80 e mediana de 0,89, indicando que o conteúdo das transcrições preserva o sentido das letras originais mesmo nos casos em que há divergências lexicais pontuais. Esses resultados evidenciam que as transcrições constituem um dado valioso no BOSSA: apenas 21,9% das faixas tiveram sua letra coletada oficialmente via Genius, e para as demais o *Whisper v3* fornece inferências de qualidade, ampliando substancialmente a cobertura textual do conjunto de dados.

Tabela 1. Estatísticas descritivas das métricas de qualidade das transcrições

Métrica	TF-IDF	Sentence-BERT
Média	0,5745	0,8048
Mediana	0,6742	0,8954
Desvio padrão	0,2889	0,2170
Mínimo	0,0000	−0,0230
Máximo	0,9999	1,0000
P25 / P75	0,47 / 0,78	0,73 / 0,96

[5]`https://genius.com/developers`

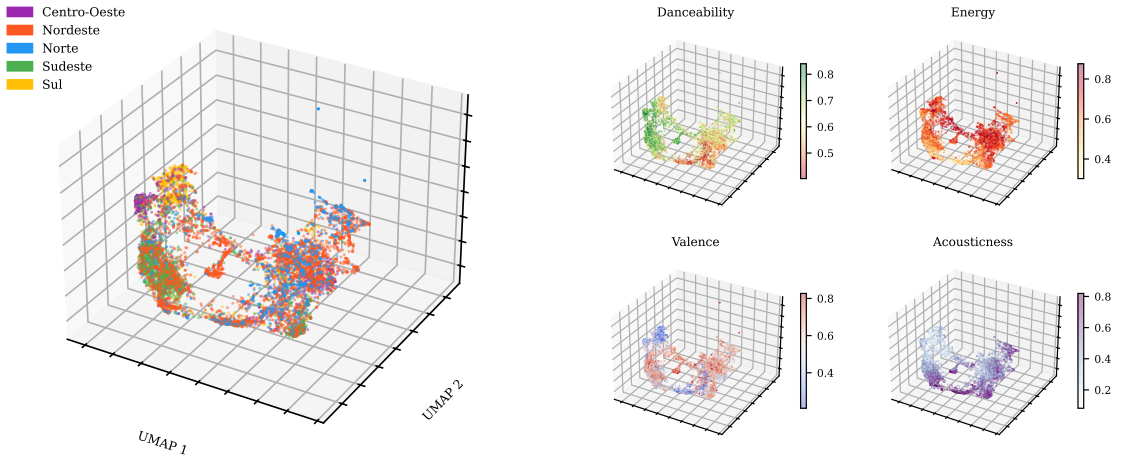
4. Aplicação

Para demonstrar a utilidade do BOSSA, realizamos três análises exploratórias sobre as *shallow features* e *deep features* do MERT, considerando apenas faixas que aparecem exclusivamente em uma única cidade , garantindo que cada ponto no espaço de *features* seja representativo de uma identidade regional específica.

4.1. Espaço latente das *deep features*

Aplicamos o UMAP [McInnes et al. 2020] com métrica de cosseno sobre os embeddings de 1.024 dimensões, reduzindo-os para três componentes. Os parâmetros foram selecionados por maximização do *trustworthiness score* [Venna and Kaski 2006] sobre uma amostra de 3.000 faixas ($n\_neighbors \in \{15, 30, 50, 100\}$, $min\_dist \in \{0{,}0, 0{,}1, 0{,}25, 0{,}5\}$). A melhor configuração obtida para essa representação foi $n\_neighbors = 15$ e $min\_dist = 0{,}0$.

A Figura 3 apresenta a projeção colorida por região geográfica (a) e por *shallow features* (b). Em (a), percebe-se uma polarização entre dois grandes grupos: Em uma região do espaço, temos grande concentração de músicas do Sudeste, Sul e Centro-Oeste. Na outra, Norte e Nordeste tendem a ser mais próximos entre si e com similaridades esporádicas com as outras regiões do Brasil. Na região central do espaço, observamos um vale no qual seletas músicas do Norte e Nordeste fazem o intermédio entre os dois grupos então polarizados. A interpretação deste espaço latente indica que o MERT captura a identidade musical regional de forma não supervisionada. Em (b), gradientes contínuos e coerentes em *danceability*, *energy*, *valence* e *acousticness* confirmam que o espaço latente é consistente com atributos musicais interpretáveis.



(a) Colorido por região geográfica

(b) Colorido por *shallow features*

Figura 3. Projeção UMAP 3D dos *embeddings* MERT-v1-330M. (a) Separação regional sem supervisão geográfica. (b) Gradientes coerentes validam consistência com atributos interpretáveis.

4.2. Similaridade musical entre cidades

Calculamos o centróide de cada cidade no espaço de *embeddings* e computamos a matriz de similaridade de cosseno entre os 17 centróides. A Figura 4 apresenta essa matriz

reordenada por agrupamento hierárquico (*Ward linkage*).

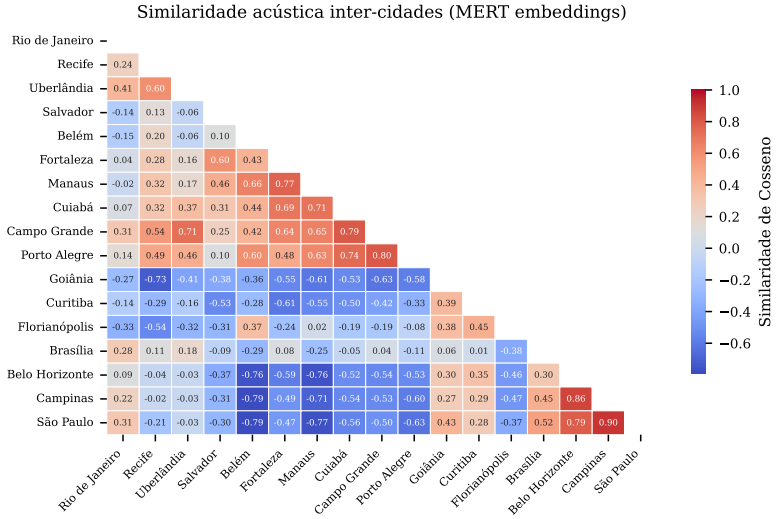


Figura 4. Similaridade de cosseno entre perfis musicais das 17 cidades (embeddings MERT, Ward linkage). Tons mais escuros indicam maior similaridade.

Os resultados revelam padrões consistentes com a geografia e cultura brasileira. O par mais similar é Campinas–São Paulo ($\rho_c = 0{,}905$), cidades com forte integração cultural e econômica. Na sequência, Belo Horizonte–Campinas ($\rho_c = 0{,}860$) e Campo Grande–Porto Alegre ($\rho_c = 0{,}801$) também exibem alta similaridade apesar da distância geográfica. No extremo oposto, os pares menos similares envolvem Belém: Belém–São Paulo ($\rho_c = -0{,}794$), Belém–Campinas ($\rho_c = -0{,}787$) e Manaus–São Paulo ($\rho_c = -0{,}767$). As preferências musicais da região Norte aparentam ter estruturas rítmicas e tímbricas fortemente distintas das que dominam os charts do Sudeste, diferenças capturadas pelo MERT sem supervisão. Esses achados corroboram e aprofundam [Moura et al. 2025], que identificaram perfis acústicos distintos para clusters de cidades a partir de *shallow features*.

4.3. Complementaridade entre representações rasas e profundas

Calculamos a correlação de Spearman entre distâncias par a par no espaço das *deep features* ($1 -$ similaridade de cosseno) e no espaço das *shallow features* (distância euclidiana padronizada), sobre 15.000 pares aleatórios. O resultado ($\rho = 0{,}30$, $p < 10^{-300}$), ilustrado na Figura 5, indica que as representações são complementares e não redundantes. Uma parcela substancial da estrutura capturada pelo MERT não está contida nas nove *shallow features*, e vice-versa. Esse resultado justifica a arquitetura multimodal do BOSSA e sugere que abordagens combinadas têm potencial de superar abordagens unimodais em tarefas de predição de popularidade e recuperação por similaridade.

5. Considerações Éticas

Direitos autorais e uso dos dados. Os áudios foram obtidos via YouTube exclusivamente para fins de pesquisa e não são redistribuídos, seguindo a Lei dos Direitos Au-

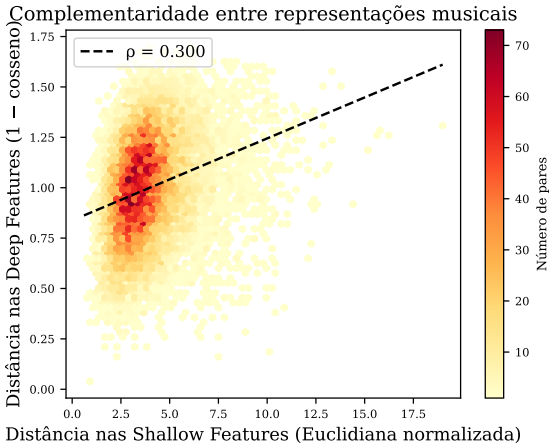


Figura 5. Correlação de Spearman entre distâncias nos espaços de *deep* e *shallow features* (15.000 pares). $\rho = 0{,}30$ ($p < 10^{-300}$) indica complementaridade.

torais (Lei nº 9.610/1998)[6]. O BOSSA disponibiliza apenas vetores de *features* derivados , representações numéricas que não permitem reconstrução dos áudios , e metadados públicos do Spotify Charts. As letras transcritas automaticamente podem ser consideradas obras derivadas. Recomendamos que pesquisadores utilizem os *embeddings* disponibilizados ou então façam uso do link de download em suas pesquisas. O conteúdo original não deve ser disponibilizado.

Disponibilidade dos dados e mudanças nas políticas do Spotify. O BOSSA cobre o período em que os dados do Spotify Charts eram publicamente acessíveis e não existiam restrições a respeito da raspagem de dados, momento no qual foi realizada a coleta. Em 2024, o Spotify descontinuou as *audio features* via API oficial. Em 2025, seus Termos de Uso foram atualizados proibindo explicitamente o *scraping* automatizado dos Charts[7] [Policy 2025]. Isso torna o BOSSA um **recurso irreproduzível**, reforçando seu valor como registro histórico do consumo musical regional brasileiro.

Vedação ao uso em treinamento de IA. Em conformidade com os Termos de Uso do Spotify [Policy 2025], os dados do BOSSA **não devem ser utilizados para treinar, ajustar ou avaliar modelos de aprendizado de máquina ou inteligência artificial**. O conjunto de dados destina-se exclusivamente a análises descritivas e exploratórias no âmbito da pesquisa acadêmica.

Limitações e vieses. Os áudios coletados via YouTube podem diferir das versões originais do Spotify em qualidade (0,3% descartados). O BOSSA cobre apenas as 17 cidades do Spotify Charts, excluindo municípios menores e populações com menor acesso à plataforma. O MERT-v1-330M foi treinado predominantemente em música ocidental, podendo gerar *embeddings* de menor qualidade para gêneros regionais brasileiros como tecnobrega, carimbó, forró eletrônico e outros gêneros especialistas.

[6] https://www.planalto.gov.br/ccivil_03/leis/l9610.htm
[7] https://www.spotify.com/us/legal/user-guidelines/

6. Dicionário dos dados

O BOSSA é composto por quatro arquivos principais, integráveis via `uri` (Spotify) e pelo padrão de nome {`track_name`} + {`artist_names`}`.mp3`.**Rankings por Cidade:** Organizados em uma pasta por cidade com arquivos semanais `.csv`. Os campos estão descritos na Tabela 2. **Features de Áudio e Transcrições:** As *shallow features*, os *embeddings* MERT e as transcrições de letras estão, respectivamente, em arquivos `.csv`, `.parquet` e `.csv`. Os esquemas estão descritos nas Tabelas 3, 4 e 5.

Tabela 2. Esquema do arquivo de rankings por cidade

Campo	Tipo	Descrição
`rank`	Inteiro	Posição no ranking semanal (1–100)
`uri`	String	Identificador único no Spotify
`track_name`	String	Nome da faixa
`artist_names`	String	Nome do(s) artista(s)
`peak`	Inteiro	Melhor posição já atingida
`prev`	Inteiro	Posição na semana anterior
`streak`	Inteiro	Semanas consecutivas no chart
`city`	String	Nome da cidade
`week`	Data	Início da semana (`YYYY-MM-DD`)

Tabela 3. Esquema do arquivo de *shallow features*.

Campo	Intervalo/Tipo	Descrição
`uri`	String	Identificador da faixa.
`youtube_url`	String	URL do vídeo no YouTube.
`track_name`	String	Nome da música.
`artist_names`	String	Nome(s) do(s) artista(s).
`acousticness`	[0,0–1,0]	Grau em que a faixa é acústica.
`danceability`	[0,0–1,0]	Adequação da faixa para dança.
`energy`	[0,0–1,0]	Intensidade e atividade percebida da música.
`instrumentalness`	[0,0–1,0]	Probabilidade de a faixa ser instrumental.
`liveness`	[0,0–1,0]	Probabilidade de gravação ao vivo.
`loudness`	dB	Volume médio da faixa.
`speechiness`	[0,0–1,0]	Predominância de palavras faladas.
`tempo`	BPM	Velocidade da música.
`valence`	[0,0–1,0]	Positividade percebida da música.

7. Disponibilização do Dataset

O dataset BOSSA-preview está disponível no Zenodo[8]. O repositório inclui os arquivos de ranking por cidade, as *shallow features* em `.csv`, os *embeddings* MERT em `.parquet`, as transcrições em `.csv` e os notebooks de extração e análise. As transcrições de letras são disponibilizadas exclusivamente na forma de *embeddings* textuais de 768 dimensões, obtidos com o modelo Sentence-BERT aplicado sobre as transcrições do *Whisper v3*.

[8] `https://doi.org/10.5281/zenodo.20006480`

Tabela 4. Esquema do arquivo de transcrições de letras.

Campo	Descrição
`filename`	Nome do arquivo de áudio.
`uri`	Link de download.
`lyric_embedding`	Vetor textual (768 dimensões).
`detected_lang`	Idioma detectado.
`status`	Status do processamento (`ok` ou `erro`).

Tabela 5. Esquema do arquivo de *deep features* (`mert_embeddings.parquet`)

Campo	Tipo	Descrição
`track_id`	String	Nome da faixa e artista(s) no formato {`track_name`} – {`artist_names`}
`duration_sec`	Float	Duração total do áudio em segundos
`n_chunks`	Inteiro	Número de janelas de 10 s processadas pelo MERT
`embedding`	Lista	Vetor de 1.024 dimensões L2-normalizado, representando a faixa no espaço latente do MERT-v1-330M

O texto bruto das transcrições não são redistribuídas, uma vez que as letras musicais constituem obras protegidas por direito autoral. As letras oficiais coletadas via API do Genius foram utilizadas exclusivamente como *ground truth* na validação interna e não integram o dataset público.

8. Conclusão

Este trabalho apresentou o **BOSSA**, dataset multimodal que integra dados de popularidade semanal por cidade, *shallow audio features*, *deep features* (MERT-v1-330M) e letras transcritas automaticamente para 16.148 faixas em 17 cidades brasileiras (jan. 2022 – dez. 2024).

Três análises exploratórias demonstram a utilidade do conjunto de dados. A projeção UMAP revela que o espaço latente do MERT organiza as músicas de forma coerente com a identidade cultural regional sem supervisão geográfica. A análise de similaridade entre cidades confirma que a proximidade musical reflete a proximidade cultural: Campinas–São Paulo formam o par mais similar ($\rho_c = 0{,}905$), enquanto Belém contrasta fortemente com os centros do Sudeste ($\rho_c = -0{,}794$), aprofundando no espaço de *deep features* os achados de [Moura et al. 2025]. A correlação de Spearman ($\rho = 0{,}30$, $p < 10^{-300}$) confirma que *shallow* e *deep features* são complementares, justificando a arquitetura multimodal do BOSSA. Uma dimensão não explorada neste trabalho são as letras musicais, que também podem apresentar informações discriminativas para identificação de padrões regionais e narrativas presentes na produção musical brasileira.

Declaração de uso de IA Generativa

Neste trabalho, LLMs (Grammarly, ChatGPT) foram utilizadas como assistentes para revisão gramatical e auxilio técnico no posicionamento de tabelas e figuras ao longo do texto. Não houve geração de conteúdo textual por IA.

Referências

Bello, P. and Garcia, D. (2021). Cultural divergence in popular music: The increasing diversity of music consumption on Spotify across countries. *Humanities and Social Sciences Communications*, 8(1):182.

Berendes, H.-U., Schwär, S., and Müller, M. (2024). Lyrics transcription in western classical music with Whisper: A case study on Schubert's winterreise. In *Proceedings of the 3rd Workshop on NLP for Music and Audio (NLP4MusA)*, pages 11–16.

Bertoni, A. and Lemos, R. (2021). Três datasets criados a partir de um banco de canções populares brasileiras de sucesso e não-sucesso de 2014 a 2019. In *Anais do III Dataset Showcase Workshop*, pages 11–20, Porto Alegre, RS, Brasil. SBC.

Li, Y., Yuan, R., Zhang, G., Ma, Y., Chen, X., Yin, H., Xiao, C., Lin, C., Ragni, A., Benetos, E., Gyenge, N., Dannenberg, R., Liu, R., Chen, W., Xia, G., Shi, Y., Huang, W., Wang, Z., Guo, Y., and Fu, J. (2024). {MERT}: Acoustic music understanding model with large-scale self-supervised training. In *The Twelfth International Conference on Learning Representations*.

McInnes, L., Healy, J., and Melville, J. (2020). Umap: Uniform manifold approximation and projection for dimension reduction.

Moura, F. A. S. et al. (2024). Characterization of the Brazilian musical landscape: A study of regional preferences based on the Spotify charts. In *Proceedings of the 30th Brazilian Symposium on Multimedia and the Web (WebMedia)*, pages 80–88. SBC.

Moura, F. A. S., Ferreira, C. H. G., and Lima, H. C. S. C. (2025). Beyond acoustic features: A data-driven analysis of music consumption in Brazil. *Journal on Interactive Systems*, 16(1).

Oliveira, G., Barbosa, G., Melo, B., Silva, M., Seufitelli, D., and Moro, M. (2021). Muhsic: An open dataset with temporal musical success information. In *Anais do III Dataset Showcase Workshop*, pages 65–76, Porto Alegre, RS, Brasil. SBC.

Oliveira, G., Vassio, L., Silva, A., and Moro, M. (2025). Modeling music popularity as an epidemic: insights from the brazilian market. In *Anais do XIV Brazilian Workshop on Social Network Analysis and Mining*, pages 79–92, Porto Alegre, RS, Brasil. SBC.

Policy, M. T. (2025). AI implications of spotify's updated terms of use: Your data is their new oil.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., and Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, pages 28492–28518.

Rompolas, G., Smpoukis, A., Kafeza, E., and Makris, C. (2024). Predicting song popularity through machine learning and sentiment analysis on social networks. In *Artificial Intelligence Applications and Innovations (AIAI 2024)*, volume 715 of *IFIP Advances in Information and Communication Technology*. Springer.

Sebastian, N., Mayer, F., et al. (2024). Beyond beats: A recipe to song popularity? a machine learning approach.

Seufitelli, D., Oliveira, G., Silva, M., and Moro, M. (2023a). Mgd+: An enhanced music genre dataset with success-based networks. In *Anais do V Dataset Showcase Workshop*, pages 36–47, Porto Alegre, RS, Brasil. SBC.

Seufitelli, D. B., Oliveira, G. P., Silva, M. O., Scofield, C., and Moro, M. M. (2023b). Hit song science: A comprehensive survey and research directions. *Journal of New Music Research*, 52(1):41–72.

Venna, J. and Kaski, S. (2006). Local multidimensional scaling. *Neural Networks*, 19(6):889–899. Advances in Self Organising Maps - WSOM'05.

Way, L. D., Garcia, D., Gathright, J., et al. (2020). Local trends in global music streaming. In *Proceedings of the International AAAI Conference on Web and Social Media*.

Zhuo, L., Yuan, R., Pan, J., Ma, Y., Li, Y., Zhang, G., et al. (2023). LyricWhiz: Robust multilingual zero-shot lyrics transcription by whispering to ChatGPT. In *Proceedings of the 24th International Society for Music Information Retrieval Conference (ISMIR)*, pages 343–351.