



**UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA E CIÊNCIAS ATUARIAIS**



Esdras Eliel Menezes Santos

ANÁLISE MULTIVARIADA DO PERFIL MUSICAL COM DADOS DO SPOTIFY

São Cristóvão - SE

2025

Esdras Eliel Menezes Santos

ANÁLISE MULTIVARIADA DO PERFIL MUSICAL COM DADOS DO SPOTIFY

Trabalho de Conclusão de Curso apresentado ao Departamento de Estatística e Ciências Atuariais da Universidade Federal de Sergipe, como parte dos requisitos para obtenção do grau de Bacharel em ESTATÍSTICA .

Orientador: Eucymara França Nunes Santos

São Cristóvão - SE

2025

Esdras Eliel Menezes Santos

ANÁLISE MULTIVARIADA DO PERFIL MUSICAL COM DADOS DO SPOTIFY

Trabalho de Conclusão de Curso apresentado ao Departamento de Estatística e Ciências Atuariais da Universidade Federal de Sergipe, como parte dos requisitos para obtenção do grau de Bacharel em ESTATÍSTICA .

Aprovado em 01/04/2025.

Eucymara França Nunes Santos
Orientadora

Esdras Adriano Barbosa dos Santos
Professor do DECAT

Cleber Martins Xavier
Professor do DECAT

São Cristóvão - SE
2025

Dedico este trabalho a todos os músicos, cantores e produtores que, com sua arte, transformam emoções em melodias e inspiram vidas. A música sempre esteve presente em minha trajetória, não apenas como um hobby, mas como uma expressão profunda do que sou. Cada acorde, cada nota e cada harmonia me ensinaram sobre disciplina, criatividade e paixão — elementos que também foram essenciais na construção deste trabalho.

Aos artistas que me influenciaram ao longo dos anos, sejam eles grandes ícones da música ou companheiros de jornada, sou grato por cada canção que me fez refletir, persistir e evoluir. A cada produtor que, nos bastidores, dá vida às ideias e transforma sons em experiências inesquecíveis, meu respeito e admiração.

Agradeço especialmente aos meus familiares e amigos, que sempre me incentivaram a buscar o conhecimento e a me desafiar. E, acima de tudo, agradeço a Deus, o maior Maestro de todos, que rege cada passo da minha caminhada.

Este trabalho é para vocês.

Agradecimentos

Agradeço, primeiramente, a Deus, por me conceder forças, sabedoria e oportunidades ao longo dessa caminhada. Sem Sua graça, nada disso seria possível.

À minha noiva, Bia, pelo amor, paciência e apoio incondicional em cada etapa dessa jornada. Seu incentivo nos momentos difíceis e sua confiança em mim foram fundamentais para que eu chegasse até aqui.

À minha família, por sempre acreditarem no meu potencial e por me ensinarem, com exemplos e valores, a importância da dedicação e do esforço.

Aos meus professores e colegas de curso, pelas trocas de conhecimento, pelas discussões enriquecedoras e pelo aprendizado compartilhado ao longo dos anos.

Agradeço também a todos que, direta ou indiretamente, contribuíram para a realização deste trabalho, seja por meio de apoio acadêmico, profissional ou emocional. Cada palavra de incentivo e cada gesto de apoio fizeram a diferença.

Por fim, dedico um agradecimento especial à música, que esteve presente em minha vida como fonte de inspiração e equilíbrio, tornando essa jornada mais leve e motivadora.

*“A ausência da evidência
não significa evidência da ausência.
(The absence of evidence
is not the evidence of absence.)”
(Carl Sagan)*

Resumo

Este trabalho analisa as características musicais e sua relação com a popularidade de músicas ao longo de várias décadas, utilizando técnicas de estatística multivariada, aprendizado de máquina supervisionado (Regressão Linear) e aprendizado de máquina não supervisionado (PCA e Clusters). Variáveis como valência, energia, dancabilidade, instrumentalidade, e outras, foram analisadas com o objetivo de compreender como elas influenciam o sucesso musical. Por meio da Análise de Componentes Principais (PCA), identificou-se que as principais dimensões do conjunto de dados estão associadas a atributos como energia, dancabilidade, acústica e instrumentalidade. A matriz de correlação revelou relações significativas entre variáveis, como a alta correlação entre energia e volume e a relação negativa entre instrumentalidade e popularidade. Além disso, a técnica de Análise Fatorial foi aplicada para identificar dimensões latentes que explicam as inter-relações entre as variáveis musicais. Os fatores latentes extraídos destacaram agrupamentos de características, como positividade emocional e dinâmica sonora, fornecendo insights sobre os padrões subjacentes que influenciam a popularidade das músicas. Além disso, a técnica de análise de agrupamentos também foi utilizada, através do método não hierárquico K-means, cujo resultado foi representado por cinco grupos com características similares, desde músicas introspectivas e acústicas até aquelas altamente dançantes e energéticas. A evolução da popularidade musical ao longo das décadas também foi analisada, destacando mudanças nos atributos preferidos pelo público. Os resultados fornecem insights relevantes para a indústria musical, destacando padrões de consumo e fatores que influenciam o sucesso de músicas. Este estudo contribui para o conhecimento da dinâmica da música popular e propõe uma base metodológica para futuras investigações. Como limitações, destaca-se a ausência de variáveis externas, como impacto midiático e contexto cultural. Trabalhos futuros podem integrar dados qualitativos e variáveis externas para ampliar a compreensão dos fenômenos analisados.

Palavras-chave: Popularidade musical, Análise de Componentes Principais (PCA), Análise de Agrupamentos, Análise Fatorial.

Abstract

This study analyzes musical characteristics and their relationship with the popularity of songs over several decades, using multivariate statistical techniques, supervised machine learning (Linear Regression), and unsupervised machine learning (PCA and Clustering). Variables such as valence, energy, danceability, instrumentality, and others were analyzed to understand how they influence musical success.

Through Principal Component Analysis (PCA), it was identified that the main dimensions of the dataset are associated with attributes such as energy, danceability, acoustics, and instrumentality. The correlation matrix revealed significant relationships between variables, such as the strong correlation between energy and loudness and the negative relationship between instrumentality and popularity. Additionally, Factor Analysis was applied to identify latent dimensions that explain the interrelationships between musical variables. The extracted latent factors highlighted groupings of characteristics, such as emotional positivity and sound dynamics, providing insights into the underlying patterns that influence song popularity.

Furthermore, cluster analysis was also employed using the non-hierarchical K-means method, resulting in five groups with similar characteristics, ranging from introspective and acoustic songs to highly danceable and energetic ones. The evolution of musical popularity over the decades was also analyzed, highlighting changes in attributes preferred by audiences. The results provide valuable insights for the music industry, emphasizing consumption patterns and factors that influence the success of songs.

This study contributes to understanding the dynamics of popular music and proposes a methodological foundation for future research. As limitations, the absence of external variables, such as media impact and cultural context, is noted. Future studies may integrate qualitative data and external variables to expand the understanding of the analyzed phenomena.

Keywords: Musical popularity, Principal Component Analysis (PCA), Cluster Analysis, Factor Analysis.

Lista de ilustrações

Figura 1 – Matriz de correlação entre as variáveis musicais.	34
Figura 2 – Análise de Componentes Principais (PCA).	36
Figura 3 – Método do cotovelo para determinar o número ideal de clusters.	39
Figura 4 – Agrupamento dos artistas com K-means baseado nos componentes principais.	40
Figura 5 – Scree Plot - Análise Fatorial mostrando os autovalores por número de fator. .	42
Figura 6 – Parallel Analysis Scree Plots comparando dados reais, simulados e reamostrados.	42
Figura 7 – Diagrama de Análise Fatorial mostrando a relação entre variáveis e fatores. .	43
Figura 8 – Mapa de Calor das Cargas Fatoriais.	45
Figura 9 – Distribuição das Músicas no Espaço Fatorial.	46
Figura 10 – Relação entre Fator 1 (Dimensão Energética/Dançante) e Popularidade. . . .	47
Figura 11 – Relação entre Fator 2 (Dimensão Acústica/Instrumental) e Popularidade. . .	47
Figura 12 – Relação entre Fator 3 (Dimensão Vocal/Textual) e Popularidade.	48

Lista de tabelas

Tabela 1 – Artistas mais populares por década (1960-2020)	30
Tabela 2 – Músicas mais populares por década (1960-2020)	32
Tabela 3 – Autovalores e Proporção de Variância Explicada pelos Componentes Principais	36
Tabela 4 – Matriz de Rotação (Autovetores dos Componentes Principais)	37
Tabela 5 – Comparação da Variância Explicada entre Análise Fatorial e PCA	48

Lista de abreviaturas e siglas

PCA	Análise de Componentes Principais
K-means	Algoritmo de K-médias
PC1	Primeiro Componente Principal
PC2	Segundo Componente Principal
R²	Coeficiente de Determinação
SSE	Soma dos Quadrados Dentro dos Clusters
MSA	Média de Similaridade dos Agrupamentos
ANOVA	Análise de Variância
BPM	Batidas por Minuto
R	Linguagem de Programação R

Lista de símbolos

Γ	Letra grega Gama
Λ	Lambda
ζ	Letra grega minúscula zeta
$\in$	Pertence
$\mathbf{X}$	Matriz de variáveis independentes
$\mathbf{Y}$	Vetor de variáveis dependentes ou componentes principais
β	Vetor de coeficientes em modelos de regressão
μ	Vetor de médias das variáveis originais
ϵ	Termo de erro ou ruído em modelos estatísticos
$\mathbf{W}$	Matriz de autovetores na Análise de Componentes Principais (PCA)
$\mathbf{C}$	Matriz de covariância
λ_k	Autovalor associado ao componente principal ou fator latente k
PC1, PC2	Primeiro e segundo componentes principais
SSE	Soma dos Quadrados Dentro dos Clusters (em agrupamentos K-means)
σ_X, σ_Y	Desvios-padrão das variáveis X e Y
$\text{cov}(X, Y)$	Covariância entre as variáveis X e Y
r_{xy}	Coefficiente de correlação de Pearson entre X e Y
$\ \cdot\ $	Norma euclidiana (distância)
$\bar{x}$	Média amostral
s	Desvio-padrão amostral
M_e	Mediana
P_k	Percentil k
Ψ	Matriz diagonal de variâncias específicas na Análise Fatorial
Σ	Matriz de covariância total

λ_{ij}	Carga fatorial, representando a contribuição do fator j na variável i
F	Vetor de fatores latentes

Sumário

1	INTRODUÇÃO	15
2	OBJETIVOS	17
2.1	Geral	17
2.2	Específicos	17
3	JUSTIFICATIVA	18
4	REVISÃO LITERÁRIA	20
5	METODOLOGIA	22
5.1	Tipo de Pesquisa	22
5.2	Coleta de Dados	22
5.3	Tratamento dos Dados	24
5.4	Métodos de Análise	24
5.4.1	Análise Descritiva	24
5.4.2	Análise de Componentes Principais (PCA)	25
5.4.3	Agrupamentos	25
5.4.4	Regressão Linear	26
5.4.5	Análise de Correlação	26
5.4.6	Análise Fatorial	26
5.5	Delimitação da Pesquisa	27
5.6	Justificativa dos Métodos	27
5.7	Softwares e Pacotes Utilizados	27
6	RESULTADOS	29
6.1	Artistas e Músicas Mais Populares por Década	29
6.1.1	Transformações Estilísticas e Culturais	31
6.1.2	Fenômenos Notáveis	31
6.2	Análise de Correlação	33
6.2.1	Correlação entre Energia e Volume	34
6.2.2	Correlação entre Instrumentalidade e Popularidade	34
6.2.3	Outras Relações	35
6.2.4	Relações Fracas e Interpretação	35
6.2.5	Implicações e Contextualização	35
6.3	Análise de Componentes Principais (PCA)	36
6.3.1	Autovalores e Variância Explicada	37

6.3.2	Autovetores (Matriz de Coeficientes)	37
6.3.3	Interpretação dos Componentes Principais	38
6.3.4	Implicações para a Indústria Musical	38
6.3.5	Conclusão	38
6.4	Agrupamento com K-means	38
6.4.1	Interpretação dos Clusters	40
6.4.2	Distribuição dos Clusters	40
6.4.3	Implicações para a Indústria Musical	41
6.4.4	Contribuições do K-means para o Estudo	41
6.5	Análise Fatorial	41
6.6	Importância do Modelo Fatorial	44
6.7	Comparação entre Análise Fatorial e PCA	48
7	CONCLUSÃO	50
	REFERÊNCIAS	52

1 INTRODUÇÃO

O estudo da música, além de explorar suas manifestações culturais e artísticas, abrange uma gama de dimensões científicas que podem ser analisadas e quantificadas. O desenvolvimento das tecnologias de coleta e análise de dados fez surgir novas possibilidades para investigar a música de maneira empírica, permitindo examinar e quantificar atributos específicos que contribuem para as preferências e reações do público. A análise multivariada é uma dessas abordagens que possibilita o estudo integrado de diversas variáveis, oferecendo uma visão mais abrangente sobre as características que constituem um perfil musical. Neste trabalho, utiliza-se essa técnica para explorar padrões em dados musicais, examinando atributos como dançabilidade, energia e valência que podem variar entre gêneros e ao longo das décadas.

A análise quantitativa da música não é uma prática atual; porém, só atualmente ganhou amplitude devido à expansão de plataformas digitais e repositórios musicais, que fornecem dados volumosos e estruturados. O surgimento de plataformas de streaming, por exemplo, gerou um vasto banco de dados que contém informações detalhadas sobre cada faixa, como tempo de duração, tonalidade, e nível de energia. Essas plataformas tornaram-se não apenas meios de distribuição, mas também repositórios de dados ricos que permitem uma nova abordagem na análise musical. Ao mesmo tempo, a ciência de dados trouxe ferramentas e métodos que possibilitam extrair conhecimentos desses grandes volumes de dados, fornecendo insights que antes seriam inatingíveis. Nesse contexto tecnológico cria-se um ambiente propício para estudos como este, que combinam dados e algoritmos de análise para investigar o universo musical sob uma nova perspectiva.

Para realizar essa análise, foram utilizadas técnicas de análise multivariada como Análise de Componentes Principais (PCA) e Análise de Agrupamentos (Cluster), que são métodos adequados para identificar padrões e reduzir a complexidade dos dados. A PCA, em particular, ajuda a reduzir a dimensionalidade dos dados, facilitando a visualização de como as variáveis se relacionam em um espaço menor e permite a identificação de fatores subjacentes que caracterizam os diferentes estilos e décadas musicais. Já o clustering agrupa músicas com características semelhantes, revelando grupos que compartilham um perfil musical específico. Essas técnicas possibilitam uma compreensão mais estruturada dos padrões e tendências que caracterizam as músicas, ajudando a responder perguntas como: “Quais são as características comuns entre músicas populares de uma mesma década?” e “Como essas características mudaram ao longo do tempo?”

A aplicação de técnicas multivariadas para analisar perfis musicais é uma prática que transcende a esfera acadêmica e tem aplicações diretas em áreas como a recomendação de músicas, o desenvolvimento de playlists personalizadas e a criação de novos conteúdos musicais. Em plataformas de streaming, algoritmos de recomendação utilizam características como as

investigadas neste trabalho para sugerir faixas alinhadas aos gostos do usuário, proporcionando uma experiência mais satisfatória e engajante. A compreensão profunda dos elementos que influenciam a popularidade e aceitação de músicas pode contribuir significativamente para a criação de algoritmos de recomendação mais precisos e personalizados.

Além disso, a análise de padrões musicais fornece insights valiosos para artistas e produtores musicais que, ao entender quais elementos contribuem para o sucesso de uma música, podem ajustar suas produções para aumentar a atratividade de suas obras, considerando as preferências e tendências do mercado. A análise multivariada dos dados musicais também possui implicações culturais, pois revela como os gostos e preferências musicais evoluíram ao longo do tempo, refletindo mudanças na sociedade, na tecnologia e nos estilos de vida. Esses conhecimentos contribuem não apenas para a ciência e tecnologia, mas também para uma apreciação mais rica e informada da música como um fenômeno cultural.

Assim, este trabalho pretende não apenas identificar padrões musicais, mas também fornecer uma base para futuras pesquisas e aplicações práticas, promovendo um entendimento mais profundo das preferências musicais e da evolução dos gêneros musicais ao longo das décadas.

2 OBJETIVOS

2.1 Geral

Este trabalho visa compreender e identificar os principais padrões em músicas de diferentes gêneros ao longo das décadas, com foco na identificação de características musicais que influenciam a popularidade. Através da aplicação de técnicas de análise multivariada.

2.2 Específicos

- Analisar a relação entre características musicais: Examinar alguns atributos como dançabilidade, energia e valência se correlacionam, destacando como esses fatores podem influenciar o perfil geral de uma música.
- Identificar fatores determinantes da popularidade: Investigar variáveis que contribuem para o sucesso das músicas, destacando características que se mostram mais associadas à popularidade entre os ouvintes.
- Explorar a evolução dos estilos musicais: Analisar como estilos e características musicais variam ao longo dos anos, avaliando mudanças e tendências em diferentes períodos e gêneros.
- Aplicar Análise de Componentes Principais (PCA): Utilizar PCA para simplificar e interpretar as contribuições de cada característica musical, facilitando o entendimento das dimensões mais relevantes.
- Agrupar músicas com características similares: Empregar técnicas de clustering para identificar grupos de músicas com perfis semelhantes, proporcionando uma visão mais estruturada dos padrões existentes.
- Realizar a Análise Fatorial para identificar as dimensões latentes que explicam as inter-relações entre as variáveis musicais, como energia, dançabilidade e valência.

3 JUSTIFICATIVA

A análise da música através de dados quantitativos vem ganhando relevância com o avanço das tecnologias de coleta e processamento de dados, especialmente nas últimas décadas. Em um contexto onde o consumo musical é altamente influenciado por plataformas de streaming e recomendações automatizadas, compreender as características que determinam o sucesso e a popularidade de uma música tornou-se essencial para a indústria audiovisual. Além disso, há um crescente interesse acadêmico em decifrar padrões musicais e suas transformações ao longo do tempo, tanto como reflexo cultural quanto para o campo de estudo quantitativo.

A popularidade das músicas tem impacto significativo no comportamento e nos gostos culturais, além de moldar tendências artísticas e comerciais. Por meio de uma análise sistemática de dados, é possível compreender melhor as preferências e padrões de consumo musical de diferentes gerações, fornecendo insights que podem ser utilizados em áreas como marketing cultural, curadoria de conteúdo, e até em políticas culturais voltadas ao incentivo da diversidade musical.

Este estudo oferece uma abordagem estruturada para identificar quais características musicais contribuem para a popularidade, além de explorar como estilos e preferências evoluem ao longo das décadas. Esses resultados podem apoiar tanto criadores de conteúdo musical quanto empresas e plataformas de streaming, ajudando-os a entender o que torna certas músicas atraentes para públicos específicos. Dessa forma, a pesquisa busca fornecer respostas mais concretas sobre como as variáveis musicais interagem para gerar interesse e retenção de audiência.

Embora a análise quantitativa de músicas não seja um campo completamente novo, ele ainda está em estágio de expansão e aprimoramento. As técnicas de análise multivariada, como a Análise de Componentes Principais, Análise Fatorial e a Análise de Agrupamentos são ferramentas relativamente novas no contexto de estudos musicais e oferecem um aprofundamento sobre o tema, permitindo uma visão mais detalhada das relações entre as características das músicas. Dessa forma, a pesquisa contribui para o enriquecimento teórico e metodológico deste campo de estudos, promovendo uma maior compreensão da complexidade envolvida na análise de músicas.

Ao fornecer uma base de conhecimento mais sólida sobre os fatores que influenciam a popularidade e a evolução dos estilos musicais, este estudo pode fomentar mudanças na forma como a produção e a curadoria musical são realizadas, tanto por artistas quanto por empresas do setor. As conclusões geradas podem ajudar a aprimorar algoritmos de recomendação, incentivando um consumo musical mais diversificado e equilibrado. Além disso, os resultados podem inspirar novas estratégias para promover estilos musicais menos conhecidos, valorizando a pluralidade cultural e expandindo o acesso a diferentes gêneros.

Em síntese, este trabalho justifica-se pela relevância social e acadêmica de compreender os fatores que influenciam o sucesso musical e as mudanças de estilo ao longo dos anos, utilizando uma abordagem inovadora e quantitativa para expandir o conhecimento nesta área.

4 REVISÃO LITERÁRIA

A análise de dados musicais tem sido objeto de estudos recentes devido à ampla disponibilidade de informações ricas e detalhadas sobre músicas. Esses dados permitem explorar questões como a popularidade de músicas, características musicais e recomendações personalizadas .

Um estudo relevante foi conduzido por [Wang, Liu e Smith \(2020\)](#), que analisaram fatores de popularidade utilizando aprendizado de máquina. O trabalho destacou que características como energia (a intensidade e a potência da música), dançabilidade (o quanto a música é boa para dançar) e valência (o quão "feliz" ou "triste" a música soa) têm papel fundamental na previsão de sucesso de uma música, sugerindo que essas variáveis possuem uma correlação significativa com a popularidade percebida pelos usuários. O modelo utilizado combinou *Random Forest* e *Support Vector Machines*, alcançando resultados robustos. Além disso, os autores discutiram a importância de considerar variáveis temporais, como a duração da música e o ano de lançamento, para melhorar a precisão das previsões. Resultados semelhantes foram encontrados por [Tyson \(2006\)](#), que exploraram a relação entre preferências musicais e os cinco grandes fatores da personalidade, destacando a influência de aspectos emocionais e cognitivos na escolha musical.

Outro estudo, por [Yadav, Kumar e Singh \(2023\)](#), utilizou o *Spotify Dataset* para agrupar músicas com base em atributos como *danceability*, *tempo*, e *energy*. Aplicando técnicas de *clustering*, como K-means, os pesquisadores conseguiram identificar padrões emergentes que podem auxiliar em sistemas de recomendação. Apesar do sucesso, o estudo apontou limitações em capturar variações culturais amplas. Em um estudo complementar, [Pareto e Almeida \(2023\)](#) aplicaram a análise de Pareto para filtrar amostras representativas de músicas, permitindo uma análise mais focada em características dominantes.

De forma complementar, [Xu, Li e Chen \(2024\)](#) investigaram sistemas de recomendação aplicando *feature engineering* (processo de transformar dados brutos em características que melhoram o desempenho de modelos de aprendizado de máquina) avançado para melhorar a precisão de algoritmos baseados em conteúdo. A pesquisa enfatizou que variáveis como *tempo* e *key* são cruciais para sistemas de recomendação mais personalizados e precisos. Além disso, o trabalho argumenta que as técnicas de aprendizado profundo, como redes neurais recorrentes (RNNs), podem aprimorar os modelos existentes, embora tenham limitações em termos de custos computacionais. Outros estudos, como o de [Musical e Trends \(2024\)](#), exploraram a relação entre tendências musicais e preferências regionais, destacando a importância de integrar variáveis contextuais para enriquecer os modelos preditivos .

A análise de componentes principais (PCA) para reduzir a dimensionalidade de características musicais foi explorada por [Mavani \(2024\)](#). O estudo demonstrou como o PCA pode simplificar a visualização de padrões, mantendo informações relevantes. A aplicação dessa

técnica no contexto do *Spotify Dataset* permitiu compreender melhor a contribuição de cada variável na definição de grupos musicais. Além disso, outros trabalhos, como o de [Hair et al. \(2010\)](#), ressaltam a importância de técnicas de redução de dimensionalidade para interpretar grandes conjuntos de dados, especialmente em cenários multivariados. De forma complementar, [Silva e Santos \(2014\)](#) aplicaram Análise Fatorial para explorar as relações entre características subjetivas e objetivas de músicas, como gestos percussivos e emoções evocadas .

Por fim, [Oliveira e Pereira \(2023\)](#) desenvolveram uma escala para medir a preferência musical de jovens, utilizando análise fatorial para validar os constructos psicométricos. Os resultados indicaram que dimensões latentes, como positividade emocional e intensidade, são preditores importantes para a aceitação de músicas pelo público. Este tipo de abordagem complementa as análises quantitativas tradicionais, fornecendo insights qualitativos sobre o impacto das características musicais . Da mesma forma, [Souza e Lima \(2023\)](#) aplicaram análise fatorial para estudar comportamentos autorreguladores no processo de aprendizagem musical, destacando a influência de fatores latentes no engajamento dos alunos.

A Análise Fatorial é uma técnica estatística multivariada amplamente utilizada para identificar estruturas latentes subjacentes a um conjunto de variáveis observadas. Essa metodologia busca reduzir a dimensionalidade dos dados, agrupando variáveis que compartilham características comuns em fatores latentes não observados. Segundo [Hair et al. \(2010\)](#), a análise fatorial é particularmente útil em estudos exploratórios, onde o objetivo é entender como as variáveis originais podem ser explicadas por um número menor de dimensões.

No contexto da análise de características musicais, a Análise Fatorial pode ser empregada para identificar dimensões latentes que explicam atributos como energia, dancabilidade e valência. Por exemplo, [Silva e Santos \(2014\)](#) aplicaram essa técnica para observar, através de índices, a influência do gesto na performance da percussão, demonstrando a utilidade da análise fatorial em contextos específicos de comportamento humano.

Além disso, [Oliveira e Pereira \(2023\)](#) desenvolveram um estudo para construir e conhecer os parâmetros psicométricos de uma escala para medir a preferência musical de jovens. Os resultados da análise fatorial aplicada aos itens do questionário sugeriram três aspectos distintos dentre os comportamentos autorreguladores no processo [Souza e Lima \(2023\)](#). Esses estudos destacam a importância da análise fatorial na identificação de dimensões ocultas que influenciam fenômenos complexos, como a música.

Essa técnica é especialmente relevante no presente trabalho, pois permite identificar relações ocultas entre variáveis musicais e suas contribuições para a popularidade das faixas. A partir dessa análise, espera-se obter insights sobre quais dimensões latentes mais impactam a aceitação de músicas pelo público.

5 METODOLOGIA

Este estudo emprega uma abordagem quantitativa e exploratória, com foco em identificar padrões e relações entre características musicais, além de explorar a evolução dos estilos ao longo do tempo. A pesquisa baseia-se na análise de dados fornecidos pelo *Spotify Dataset* <<https://www.kaggle.com/datasets/vatsalmavani/spotify-dataset>>, disponível na plataforma Kaggle. Este capítulo detalha o tipo de pesquisa, os instrumentos utilizados para coleta e tratamento dos dados, além das técnicas aplicadas nas análises.

5.1 Tipo de Pesquisa

A pesquisa é classificada como:

- **Quantitativa:** Trabalha com dados numéricos, permitindo análises estatísticas detalhadas para identificar padrões e correlações.
- **Exploratória:** Busca insights sobre as características musicais e fatores que influenciam a popularidade das músicas, explorando o comportamento das variáveis de forma descritiva e multivariada.
- **Documental:** Baseada em dados estruturados obtidos de uma fonte de dados autenticados (*Spotify Dataset*), que contém informações sobre milhares de músicas.

Essa combinação de abordagens é ideal para compreender os fatores subjacentes às características musicais e à evolução temporal dos estilos.

5.2 Coleta de Dados

Os dados utilizados neste estudo foram extraídos do *Spotify Dataset*, disponibilizado gratuitamente no Kaggle. Este *dataset* contém informações detalhadas sobre músicas de diferentes estilos, lançadas em vários períodos. As variáveis presentes no conjunto de dados incluem:

- **valence** (Valência): representa a positividade emocional da música, variando de 0 (negativo) a 1 (positivo).
- **year** (Ano): Ano de lançamento da música.
- **acousticness** (Acústica): Mede o quão acústica é a música, variando de 0 a 1.
- **artists** (Artistas): Nome(s) dos artistas responsáveis pela música.

- **danceability** (Dançabilidade): mede o quanto uma música é adequada para dançar. É um valor que varia de 0 a 1, onde:
 - **0.0** → Música nada dançante (exemplo: música clássica lenta).
 - **0.5** → Música moderadamente dançante (exemplo: pop suave).
 - **1.0** → Música altamente dançante (exemplo: batidas eletrônicas ou funk).
- **duration_ms** (Duração): Duração da música em milissegundos.
- **energy** (Energia): Representa a intensidade da música em uma escala de 0 a 1.
- **explicit** (Explícita): Indica se a música contém conteúdo explícito (0 para não explícito, 1 para explícito).
- **id**: Identificador único de cada música no *dataset*.
- **instrumentalness** (Instrumentalidade): Mede a presença de vocais na música. Valores mais próximos de 1 indicam músicas com menos voz e mais instrumentos.
- **key** (Tom): Indica a tonalidade da música em notação numérica.
- **liveness** (Ao vivo): Probabilidade da música ter sido gravada ao vivo.
- **mode** (Modo): Indica se a música está em modo maior (1) ou menor (0). O modo musical influencia a sensação emocional da música:
 - **0** → Modo menor (música com tom mais triste ou melancólico).
 - **1** → Modo maior (música com tom mais alegre ou brilhante).

O modo é determinado pela **terceira nota** da escala musical:

- Se a **terça é maior**, a música está no **modo maior** (soando feliz).
Exemplo: Happy - Pharrell Williams
- Se a **terça é menor**, a música está no **modo menor** (soando triste).
Exemplo: Someone Like You - Adele

Por exemplo, na escala de Dó Maior ($C - E - G$), a nota E é uma terça maior $\Rightarrow$ **Modo Maior**.

Já na escala de Dó Menor ($C - E^b - G$), a nota E^b é uma terça menor $\Rightarrow$ **Modo Menor**.

- **name** (Nome): Título da música.
- **popularity** (Popularidade): Pontuação de popularidade da música, de 0 a 100.
- **release_date** (Data de Lançamento): Data de lançamento da música.
- **speechiness** (Fala): Mede a presença de palavras faladas na música.
- **tempo**: Ritmo da música, em batidas por minuto (BPM).

5.3 Tratamento dos Dados

Para garantir a qualidade das análises, os dados foram submetidos a um processo de tratamento rigoroso:

- **Limpagem de Dados:**

- Remoção de valores nulos e duplicados.
- Exclusão de registros inconsistentes, como músicas com variáveis fora de faixas esperadas (ex.: BPM extremamente baixos ou altos).

- **Normalização e Padronização:**

- As variáveis como *valence*, *energy* e *danceability*, foram normalizadas para escalas uniformes, garantindo a comparabilidade.
- Dados categóricos, como *key* e *mode*, foram codificados para aplicação em análises estatísticas.

- **Criação de Variáveis Derivadas:**

- A variável *decada* foi derivada de *year* para permitir análises temporais mais agregadas.
- Métricas adicionais, como categorias de popularidade, foram criadas com base em faixas de valores.

5.4 Métodos de Análise

5.4.1 Análise Descritiva

A análise descritiva foi realizada para explorar a distribuição das variáveis e identificar padrões preliminares. As principais medidas estatísticas calculadas incluem média ($\bar{x}$), mediana (M_e), desvio-padrão (s) e percentis (P_k). Essas medidas são definidas como:

- Média:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- Mediana:

$$M_e = \begin{cases} x_{(n+1)/2}, & \text{se } n \text{ é ímpar,} \\ \frac{x_{n/2} + x_{n/2+1}}{2}, & \text{se } n \text{ é par.} \end{cases}$$

- Desvio-padrão:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

- Percentil (P_k):

$$P_k = x_{[k \cdot n / 100]}$$

onde k é a posição do percentil na amostra ordenada.

Essas métricas foram aplicadas às variáveis contínuas do conjunto de dados, como energia, dançabilidade e valência.

5.4.2 Análise de Componentes Principais (PCA)

A Análise de Componentes Principais (PCA) foi aplicada para reduzir a dimensionalidade do conjunto de dados e identificar as variáveis mais influentes. A transformação PCA é dada por:

$$\mathbf{Y} = \mathbf{W}^T (\mathbf{X} - \boldsymbol{\mu})$$

onde $\mathbf{Y}$ são os componentes principais, $\mathbf{W}$ é a matriz de autovetores da matriz de covariância $\mathbf{C}$, e $\boldsymbol{\mu}$ é a média das variáveis originais.

A matriz de covariância $\mathbf{C}$ é calculada como:

$$\mathbf{C} = \frac{1}{N-1} \mathbf{X}^T \mathbf{X}$$

Os autovetores $\mathbf{w}_k$ e autovalores λ_k são obtidos via:

$$\mathbf{C} \mathbf{w}_k = \lambda_k \mathbf{w}_k$$

A proporção de variância explicada por cada componente principal é calculada como:

$$\text{Proporção de Variância} = \frac{\lambda_k}{\sum_{i=1}^p \lambda_i}$$

onde λ_k é o autovalor associado ao componente principal k .

O primeiro componente principal (PC1) é uma combinação linear das variáveis normalizadas:

$$\text{PC1} = w_{11}X_1 + w_{12}X_2 + \cdots + w_{1p}X_p$$

e o segundo componente principal (PC2) é dado por:

$$\text{PC2} = w_{21}X_1 + w_{22}X_2 + \cdots + w_{2p}X_p$$

5.4.3 Agrupamentos

As técnicas de clustering, como K-means, foram utilizadas para agrupar músicas com características semelhantes. O objetivo do K-means é minimizar a soma dos quadrados intra-clusters (SSE):

$$\text{SSE} = \sum_{k=1}^K \sum_{i \in C_k} \|\mathbf{x}_i - \boldsymbol{\mu}_k\|^2$$

onde K é o número de clusters, C_k é o conjunto de pontos no cluster k , e $\boldsymbol{\mu}_k$ é o centróide do cluster k . O número ideal de clusters foi determinado utilizando o método do Screen Plot.

5.4.4 Regressão Linear

Modelos de regressão linear foram empregados para investigar os preditores da popularidade musical. O modelo de regressão linear é definido por:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon$$

onde Y é a variável dependente (popularidade), $X_1, X_2, \dots, X_n$ são as variáveis independentes (características musicais), β_0 é o intercepto, $\beta_1, \beta_2, \dots, \beta_n$ são os coeficientes angulares, e ϵ representa o erro aleatório.

Em notação matricial, o modelo é representado por:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

5.4.5 Análise de Correlação

A matriz de correlação foi construída utilizando o coeficiente de Pearson (r), que mede a força da relação linear entre duas variáveis. O coeficiente de Pearson é dado por:

$$r_{xy} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

onde: - $\text{cov}(X, Y)$ é a covariância entre X e Y , definida como:

$$\text{cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- σ_X e σ_Y são os desvios-padrão de X e Y , respectivamente.

5.4.6 Análise Fatorial

A Análise Fatorial foi utilizada para identificar estruturas latentes subjacentes às variáveis observadas, reduzindo a dimensionalidade dos dados e agrupando variáveis que compartilham características comuns. Essa técnica é particularmente útil para entender como as variáveis originais podem ser explicadas por um número menor de fatores não observados.

O modelo de Análise Fatorial assume que cada variável observada X_i pode ser expressa como uma combinação linear de fatores latentes F_j mais um erro específico ϵ_i :

$$X_i = \lambda_{i1} F_1 + \lambda_{i2} F_2 + \cdots + \lambda_{im} F_m + \epsilon_i$$

onde: - X_i : Variável observada i , - F_j : Fator latente j , - λ_{ij} : Carga fatorial, que representa a contribuição do fator F_j na variável X_i , - ϵ_i : Erro específico associado à variável X_i .

A matriz de covariância $\boldsymbol{\Sigma}$ das variáveis observadas pode ser decomposta como:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi}$$

onde: - Λ : Matriz de cargas fatoriais, - Ψ : Matriz diagonal de variâncias específicas (erros).

Os fatores latentes são obtidos através da extração dos componentes principais ou pelo método de máxima verossimilhança. A proporção de variância explicada por cada fator é dada por:

$$\text{Proporção de Variância} = \frac{\lambda_k}{\sum_{i=1}^p \lambda_i}$$

onde λ_k é o autovalor associado ao fator k .

A interpretação dos fatores latentes é baseada nas cargas fatoriais, que indicam a correlação entre as variáveis observadas e os fatores. Valores altos de carga fatorial (em módulo) sugerem que a variável está fortemente associada ao fator correspondente.

No contexto deste estudo, a Análise Fatorial foi aplicada para identificar dimensões latentes que explicam as características musicais, como energia, dançabilidade e valência.

5.5 Delimitação da Pesquisa

A pesquisa está delimitada às músicas presentes no *Spotify Dataset*, que abrange diversos estilos e períodos. Essa delimitação garante consistência nos dados analisados, mas restringe a generalização dos resultados para outros contextos musicais ou bases de dados.

5.6 Justificativa dos Métodos

A escolha dos métodos quantitativos e das técnicas estatísticas baseou-se na capacidade de identificar padrões em grandes volumes de dados. A combinação de análises descritivas, PCA, *clustering* e regressão fornece uma visão abrangente sobre as relações entre características musicais e popularidade.

5.7 Softwares e Pacotes Utilizados

Para a realização da análise, foram empregados softwares e pacotes estatísticos que facilitaram o processamento e a interpretação dos dados. A seguir, listam-se os principais:

- R e Rstudio: Utilizado para a aplicação de métodos estatísticos, incluindo a Análise Fatorial e a PCA.
- Pacote FactoMineR (R): Utilizado para realizar Análise de Componentes Principais e Análise Fatorial de maneira eficiente.
- Pacote psych (R): Aplicado na execução de Análise Fatorial Exploratória e testes de confiabilidade dos fatores.
- ggplot2 (R): Usado para criar visualizações gráficas avançadas dos resultados.

- tidyverse (R): Conjunto de pacotes para manipulação e transformação eficiente de dados.
- corrplot (R): Aplicado para a criação de mapas de correlação entre variáveis.
- cluster (R): Utilizado para análise de agrupamentos, incluindo o método K-means.
- nFactors (R): Empregado para determinar o número ótimo de fatores na Análise Fatorial.

A escolha desses softwares e pacotes foi baseada na sua robustez e ampla utilização na comunidade acadêmica e científica para análises estatísticas e modelagem de dados.

6 RESULTADOS

Neste capítulo, são apresentados os resultados das análises realizadas sobre as características musicais do dataset. A investigação abordou como essas características influenciam a popularidade das músicas, a evolução dos estilos ao longo das décadas e o agrupamento de músicas por similaridade. Além disso, aplicaram-se técnicas estatísticas e de aprendizado de máquina para explorar os dados de maneira robusta e interpretativa.

6.1 Artistas e Músicas Mais Populares por Década

A análise dos artistas mais populares por década, conforme apresentado na Tabela 1, fornece um panorama da evolução dos estilos musicais predominantes em diferentes períodos históricos.

Tabela 1 – Artistas mais populares por década (1960-2020)

Década	Artistas	Média de Popularidade
1960	Frank Sinatra, Nancy Sinatra	71
	The Beach Boys, Mark Linett, Sweet, Larry Walsh	69
	Bob Dylan, Johnny Cash	67
	Bobby “Boris” Pickett	64
	Astrud Gilberto, Antônio Carlos Jobim	62
1970	Blue Swede, Björn Skifs	75
	Earth, Wind & Fire, The Emotions	74
	John Travolta, Olivia Newton-John	68
	Chaka Khan	67
	Aliotta Haynes Jeremiah	66
1980	Grover Washington, Jr., Bill Withers	79
	Katrina & The Waves	78
	Michael Sembello	74
	John Waite	71
	Stan Getz, João Gilberto, Astrud Gilberto	69
1990	Frank Sinatra, B. Swanson Quartet	80
	Fools Garden	78
	2Pac, Big Syke	77
	Blackstreet, Dr. Dre, Queen Pen	76
	Lou Bega	76
	Perry Como, The Fontane Sisters, Mitchell Ayres & His Orchestra	76
2000	Shakira, Wyclef Jean	83
	Usher, Lil Jon, Ludacris	82
	Flo Rida, T-Pain	80
	JAY-Z, Alicia Keys	80
	Usher, Alicia Keys	80
2010	Topic, A7S	92
	Billie Eilish, Khalid	89
	KAROL G, Nicki Minaj	88
	Surf Mesa, Emilee	88
	JP Saxe, Julia Michaels	87
	Regard	87
2020	Bad Bunny, Jhay Cortez	100
	24kGoldn, iann dior	99
	Cardi B, Megan Thee Stallion	96
	Justin Bieber, Chance the Rapper	95
	Justin Bieber, benny blanco	95

Fonte: Elaborado pelo autor.

A Tabela 1 destaca os artistas mais populares em cada década, com base na popularidade média de suas músicas.

6.1.1 Transformações Estilísticas e Culturais

A análise dos artistas por década também ilustra importantes transformações estilísticas na música popular:

- **Anos 1960-1970:** Predominância de artistas do jazz vocal (Frank Sinatra), rock clássico (The Beach Boys) e colaborações importantes como Bob Dylan e Johnny Cash. A bossa nova brasileira também figura entre os mais populares com Astrud Gilberto e Antônio Carlos Jobim.
- **Anos 1980-1990:** Transição para o pop mainstream (Katrina & The Waves) e fusões como R&B com jazz (Grover Washington Jr.). Nos anos 1990, nota-se a diversificação com hip-hop (2Pac) e R&B contemporâneo (Blackstreet, Dr. Dre).
- **Anos 2000:** Consolidação do pop latino (Shakira) e R&B urbano (Usher, JAY-Z), com crescente presença de colaborações entre artistas de diferentes gêneros.
- **Anos 2010-2020:** Explosão da música eletrônica, pop alternativo (Billie Eilish) e definitiva globalização com a ascensão do reggaeton e trap latino (Bad Bunny, KAROL G).

6.1.2 Fenômenos Notáveis

Alguns fenômenos importantes são evidenciados pelos dados:

1. **Longevidade artística:** Frank Sinatra aparece como artista popular tanto nos anos 1960 quanto nos 1990, demonstrando sua influência duradoura.
2. **Aumento de colaborações:** Percebe-se um crescimento constante nas parcerias entre artistas, especialmente a partir dos anos 2000, como reflexo de estratégias mercadológicas e criativas.
3. **Globalização musical:** A presença crescente de artistas latinos (Shakira, Bad Bunny) e fusões culturais nas décadas recentes evidencia a internacionalização do mercado musical.
4. **Representatividade:** A representação feminina e de artistas de diferentes origens étnicas e culturais aumenta significativamente nas décadas mais recentes.

A evolução das músicas mais populares por década, conforme representado na Tabela 2, evidencia como certos estilos e artistas transcendem gerações, enquanto outros dominam períodos específicos.

Tabela 2 – Músicas mais populares por década (1960-2020)

Década	Música	Popularidade
1960	Rockin' Around The Christmas Tree	85
	Here Comes The Sun - Remastered 2009	82
	Fortunate Son	81
	Brown Eyed Girl	79
	California Dreamin' - Single Version	79
1970	Highway to Hell	84
	Rocket Man (I Think It's Going To Be A Long, Long Time)	82
	Sweet Home Alabama	82
	Landslide	80
	September	80
1980	Back In Black	84
	Every Breath You Take	84
	Africa	83
	Livin' On A Prayer	83
	I Wanna Dance with Somebody (Who Loves Me)	82
1990	Thunderstruck	83
	Let It Snow! Let It Snow! Let It Snow! (with The B. Swanson Quartet)	80
	November Rain	80
	Still D.R.E.	80
	All Star	79
2000	Hips Don't Lie (feat. Wyclef Jean)	83
	Last Christmas - Remastered	83
	Fix You	82
	Yeah! (feat. Lil Jon & Ludacris)	82
	Feel Good Inc.	81
2010	Breaking Me	92
	Sofia	90
	The Box	89
	lovely (with Khalid)	89
	HIGHEST IN THE ROOM	88
2020	Dakiti	100
	Mood (feat. iann dior)	99
	Blinding Lights	96
	WAP (feat. Megan Thee Stallion)	96
	What You Know Bout Love	96

Fonte: Elaborado pelo autor.

Observa-se que a década de 1960 foi dominada por clássicos do rock e folk, como *Here Comes The Sun* e *Fortunate Son*, refletindo o impacto do rock psicodélico e do folk rock na

contracultura da época. Já nos anos 1970, o rock se consolidou com canções icônicas como *Highway to Hell* e *Sweet Home Alabama*, refletindo a popularização do hard rock e do southern rock.

Nos anos 1980, a crescente adoção de sintetizadores e produção eletrônica resultou em sucessos como *Africa*, do Toto, e *I Wanna Dance with Somebody*, de Whitney Houston, que exemplificam a transição para uma sonoridade mais pop e radiofônica.

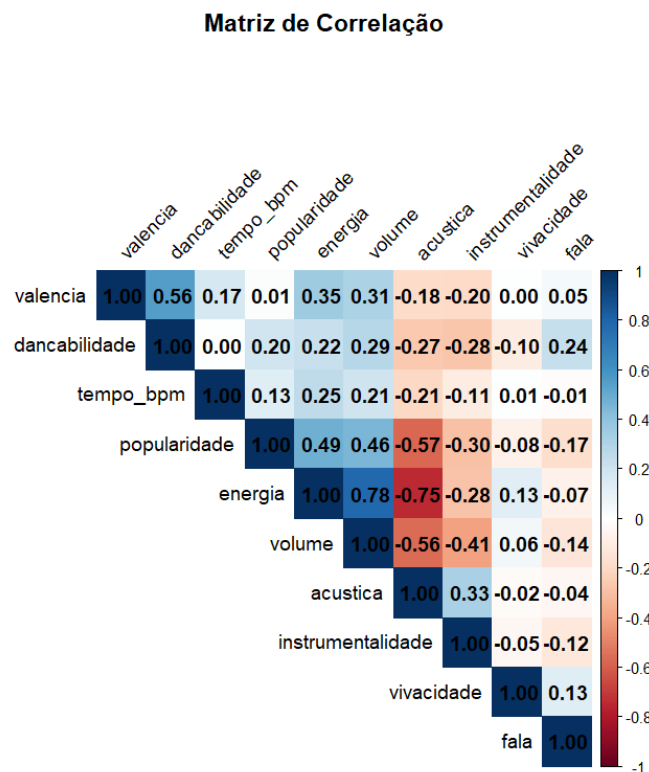
O impacto da transição dos formatos físicos para o digital fica evidente nos anos 2000 e 2010, com o surgimento de hits como *Hips Don't Lie* e *Feel Good Inc.*, os quais se beneficiaram da expansão dos meios digitais. Finalmente, nos anos 2020, observa-se a supremacia das plataformas de streaming, refletida no domínio de artistas como The Weeknd e Bad Bunny, cujas canções atingiram recordes de popularidade graças à ampla disseminação em serviços como Spotify e YouTube.

Esses padrões indicam como a evolução tecnológica e cultural impacta diretamente a forma como o público consome e valoriza a música ao longo das décadas.

6.2 Análise de Correlação

A matriz de correlação das características musicais, apresentada na Figura 1, oferece uma visão detalhada das inter-relações entre as variáveis analisadas. A correlação de Pearson, utilizada na construção da matriz, varia de -1 a 1, onde valores próximos de -1 indicam uma forte relação inversa, 0 indica ausência de correlação e valores próximos de 1 indicam uma forte relação direta. Serão analisadas as correlações que tiverem os coeficientes mais significativos.

Figura 1 – Matriz de correlação entre as variáveis musicais.



Fonte: Elaborado pelo autor.

6.2.1 Correlação entre Energia e Volume

Um dos destaques da matriz é a forte correlação positiva entre *energia* e *volume* ($r = 0.78$). Essa relação sugere que músicas com maior volume geralmente são percebidas como mais energéticas. Isso pode estar relacionado a aspectos psicológicos da audição, em que volumes elevados frequentemente transmitem intensidade e emoção, características marcantes de gêneros como rock e música eletrônica.

Essa relação também aponta para uma possível estratégia de produção musical: aumentar o volume pode ser uma forma de criar a percepção de maior energia, independentemente de outros fatores como ritmo ou tonalidade. Isso corrobora com os achados de Wang, Liu e Smith (2020), que destacaram essa relação em sistemas de recomendação musical. Essa forte associação sugere que músicas com alta energia tendem a ter volumes mais altos, um padrão observado em diversas plataformas de streaming.

6.2.2 Correlação entre Instrumentalidade e Popularidade

Outro ponto significativo é a correlação negativa entre *instrumentalidade* e *popularidade* ($r \approx -0.30$), que, mesmo sendo relativamente fraca, indica que músicas predominantemente instrumentais tendem a ser menos populares em relação às músicas com vocais. Tal padrão reflete as preferências do público contemporâneo, que valoriza músicas com letras e vocais capazes de

transmitir histórias ou emoções, algo menos presente em faixas instrumentais.

Por exemplo, músicas instrumentais geralmente têm maior destaque em nichos específicos, como trilhas sonoras ou música clássica, enquanto gêneros populares, como pop e hip-hop, dependem fortemente da presença de vocais para alcançar grandes audiências.

6.2.3 Outras Relações

A matriz também revela correlações moderadas que merecem destaque:

- A relação positiva entre *valência* e *dancabilidade* ($r = 0.56$) sugere que músicas mais alegres também tendem a ser mais dançantes. Essa conexão é intuitiva, pois músicas alegres geralmente possuem ritmos que incentivam o movimento.
- A correlação moderada entre *energia* e *dancabilidade* ($r = 0.46$) reforça que músicas com maior intensidade sonora também tendem a ter características que favorecem a dança. Essa relação é especialmente evidente em gêneros como música eletrônica e funk.
- A correlação negativa entre *instrumentalidade* e *energia* ($r \approx -0.28$) indica que músicas altamente instrumentais frequentemente apresentam níveis mais baixos de energia, característica comum em composições acústicas ou minimalistas.

6.2.4 Relações Fracas e Interpretação

Algumas variáveis apresentam correlações muito baixas ou quase inexistentes com a popularidade, como *vivacidade* e *acústica*. Esses resultados indicam que essas características têm impacto limitado na aceitação geral das músicas, podendo ser mais relevantes apenas em contextos específicos ou em nichos musicais.

6.2.5 Implicações e Contextualização

A análise da matriz de correlação destaca que, embora algumas variáveis individuais apresentem relações significativas, a popularidade de uma música é influenciada por uma interação complexa de múltiplos fatores. Além disso, as correlações observadas fornecem insights importantes para artistas, produtores e gravadoras que buscam criar músicas que ressoem com o público. Por exemplo, focar em características como *energia* e *dancabilidade* pode aumentar as chances de uma música alcançar maior popularidade, enquanto a instrumentalidade reduzida pode ampliar o apelo comercial.

Por fim, a matriz de correlação serve como base para análises mais avançadas, como a Análise de Componentes Principais (Figura 2), os Agrupamentos via *K-means* (Figura 4) e a Análise Fatorial (Figura 8), que exploram a interação simultânea entre as variáveis.

6.3 Análise de Componentes Principais (PCA)

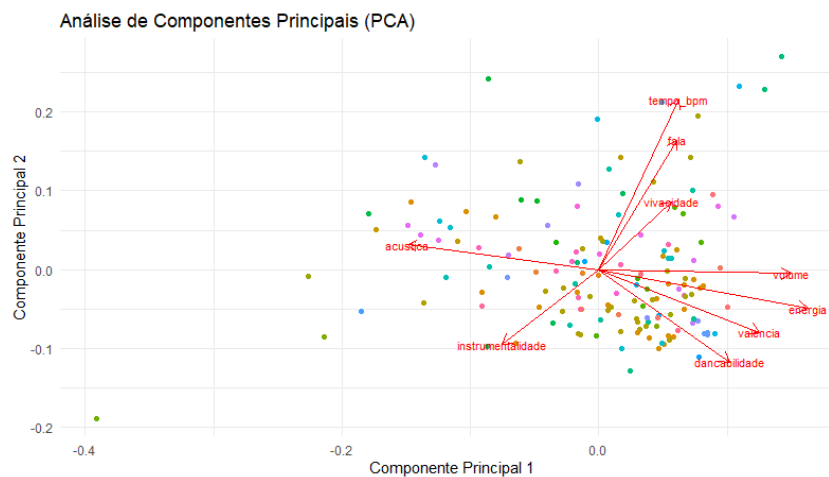
A Análise de Componentes Principais (PCA) é uma técnica estatística que permite reduzir a dimensionalidade do conjunto de dados, mantendo a maior parte da variabilidade explicada pelas variáveis originais. A PCA é apresentada na Figura 2, onde os eixos PC1 (Primeiro Componente Principal) e PC2 (Segundo Componente Principal) representam as dimensões que melhor resumem a estrutura dos dados.

Tabela 3 – Autovalores e Proporção de Variância Explicada pelos Componentes Principais

Componente	Desvio Padrão	Proporção da Variância
PC1	1.830	37.20%
PC2	1.092	13.26%
PC3	1.044	12.12%
PC4	0.981	10.70%
PC5	0.907	9.15%
PC6	0.857	8.16%
PC7	0.610	4.14%
PC8	0.596	3.94%
PC9	0.348	1.35%

Fonte: Elaborado pelo autor.

Figura 2 – Análise de Componentes Principais (PCA).



Fonte: Elaborado pelo autor.

O gráfico da PCA apresenta dois eixos principais:

- Componente Principal 1 (PC1): Captura a maior variação dos dados.
- Componente Principal 2 (PC2): Captura a segunda maior variação.

Os pontos coloridos representam diferentes músicas, posicionadas de acordo com suas características transformadas pela PCA. As setas vermelhas indicam as variáveis originais e sua influência sobre os componentes principais:

- tempo_bpm: Relacionado ao tempo da música em batidas por minuto, fortemente alinhado com PC1.
- energia e volume: Variáveis correlacionadas, indicando que músicas mais energéticas tendem a ter maior volume.
- dançabilidade e valência: Associadas ao caráter positivo da música, sugerindo que músicas mais dançantes tendem a ter um tom mais alegre.
- acústica e instrumentalidade: Apontam para uma direção oposta, indicando que músicas mais acústicas e instrumentais possuem características distintas das demais.

6.3.1 Autovalores e Variância Explicada

A Tabela 3 apresenta os autovalores dos componentes principais, juntamente com a proporção da variância explicada por cada um deles. O Primeiro Componente Principal (PC1) explica 37,20% da variância total dos dados, seguido pelo PC2, que explica 13,26%, acumulando 50,46% da variância explicada. Isso indica que os dois primeiros componentes capturam grande parte da estrutura dos dados, conforme discutido por [Hair et al. \(2010\)](#). Pode-se também acrescentar o PC3 e o PC4, que, juntos ao PC1 e PC2, acumulam 73,28% da variância total dos dados.

6.3.2 Autovetores (Matriz de Coeficientes)

A Tabela 4 apresenta os autovetores dos componentes principais, indicando a contribuição de cada variável em cada um dos componentes. Por exemplo, o PC1 é fortemente influenciado por acústica, energia e volume, enquanto o PC2 está associado principalmente à tempo e fala. Já o PC3 está associado à Vivacidade, e o PC4 à Dançabilidade.

Tabela 4 – Matriz de Rotação (Autovetores dos Componentes Principais)

Variável	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Valência	0.372	-0.236	0.236	-0.025	-0.377	0.476	0.421	0.393	-0.223
Acústica	-0.443	0.094	0.030	0.274	-0.277	-0.001	-0.237	0.671	0.368
Dançabilidade	0.302	-0.348	-0.126	0.600	-0.100	0.304	-0.472	-0.267	0.125
Energia	0.482	-0.146	0.101	-0.271	0.134	-0.173	0.058	0.094	0.776
Instrumentalidade	-0.223	-0.282	0.598	-0.302	0.390	0.317	-0.403	0.047	-0.070
Vivacidade	0.166	0.255	0.683	0.170	-0.420	-0.415	-0.128	-0.206	-0.070
Volume	0.445	-0.016	-0.175	-0.083	0.171	-0.366	-0.424	0.491	-0.425
Fala	0.179	0.487	0.210	0.487	0.587	0.174	0.227	0.152	0.024
Tempo (BPM)	0.186	0.641	-0.136	-0.357	-0.225	0.464	-0.361	-0.077	0.077

Fonte: Elaborado pelo autor.

6.3.3 Interpretação dos Componentes Principais

Os componentes principais são combinações lineares das variáveis originais e são ordenados de forma a capturar a maior parte da variância dos dados. Com base na Tabela 3 e Tabela 4, pode-se interpretar os principais componentes da seguinte forma:

- PC1 (37,20% de explicação): Altamente influenciado por Acústica, Energia e Volume. Esse componente diferencia músicas animadas e dançantes de músicas mais introspectivas.
- PC2 (13,26% de explicação, 50,46% acumulado): Está associado principalmente ao Tempo e Fala, separando músicas com mais fala e com maior ou menor andamento (BPM).
- PC3 (12,12% de explicação, 62,58% acumulado): Tem forte correlação com vivacidade e instrumentalidade, sugerindo que diferencia músicas mais vivas de outras mais monótonas.
- PC4 (10,70% de explicação, 73,28% acumulado): Tem correlação com a Dançabilidade. Sugerindo que diferencia músicas mais agitadas de músicas mais lentas.

6.3.4 Implicações para a Indústria Musical

A PCA fornece insights valiosos para a segmentação de músicas em categorias ou gêneros. Para produtores musicais e plataformas de streaming, os componentes principais podem ajudar na criação de algoritmos de recomendação, agrupando músicas com características semelhantes ou sugerindo playlists personalizadas.

Além disso, ao entender as características que mais contribuem para a variação entre músicas, artistas e produtores podem direcionar seus esforços para criar faixas alinhadas às tendências de popularidade.

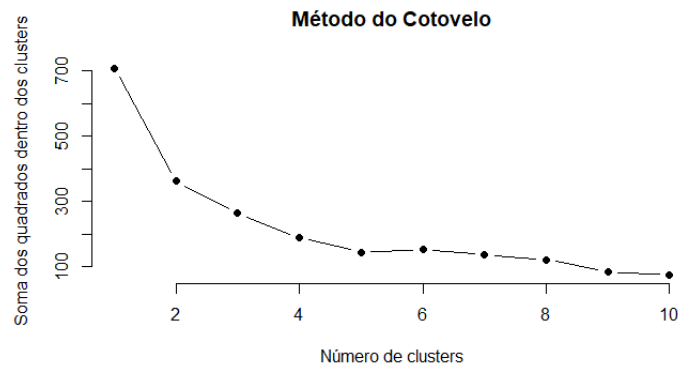
6.3.5 Conclusão

A PCA demonstrou ser uma ferramenta eficiente para resumir e interpretar as variáveis musicais, destacando as principais características que diferenciam as músicas. Este método simplifica a análise e permite identificar tendências que podem ser exploradas tanto em estudos acadêmicos quanto em estratégias de marketing musical.

6.4 Agrupamento com K-means

O algoritmo K-means foi aplicado com o objetivo de agrupar as músicas em categorias com características semelhantes, utilizando os dois primeiros componentes principais (PC1 e PC2) gerados pela Análise de Componentes Principais (PCA). O processo para determinar o número ideal de clusters foi guiado pelo método do cotovelo, apresentado na Figura 3.

Figura 3 – Método do cotovelo para determinar o número ideal de clusters.

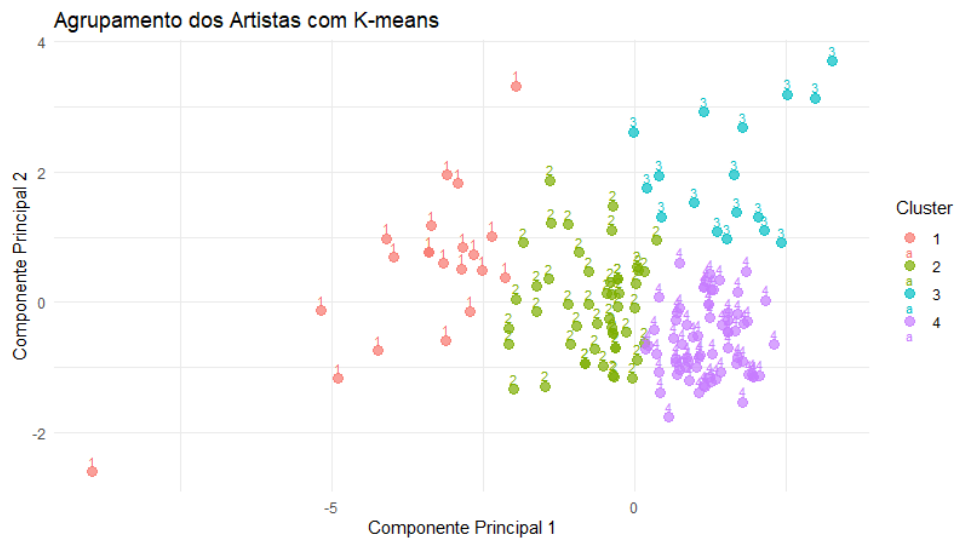


Fonte: Elaborado pelo autor.

O método do cotovelo é uma técnica utilizada para determinar o número ideal de clusters em um agrupamento. A ideia principal é analisar como a soma dos quadrados dentro dos clusters (WCSS - Within-Cluster Sum of Squares) se comporta à medida que o número de clusters k aumenta. Inicialmente, quando há poucos clusters, a WCSS tem valores elevados, pois os pontos estão mais espalhados dentro de cada grupo. Conforme k cresce, a WCSS diminui, pois os clusters ficam menores e os pontos mais próximos de seus centroides. No entanto, essa redução não ocorre de forma linear: há um ponto em que a diminuição começa a desacelerar significativamente, formando um “cotovelo” no gráfico de WCSS versus k . Esse ponto representa o número ideal de clusters, pois adicionar mais grupos após esse limite traz ganhos marginais insignificantes em termos de organização dos dados.

Esse método revelou que quatro clusters eram suficientes para capturar a estrutura subjacente dos dados. Após a definição do número ideal, os artistas e suas músicas foram agrupados, e a visualização desses clusters é apresentada na Figura 4.

Figura 4 – Agrupamento dos artistas com K-means baseado nos componentes principais.



Fonte: Elaborado pelo autor.

6.4.1 Interpretação dos Clusters

Os quatro clusters identificados representam grupos distintos de músicas baseados em suas características musicais predominantes. Cada cluster reflete um conjunto de atributos que compartilham semelhanças significativas, conforme descrito a seguir:

- *Cluster 1:* Este cluster agrupa músicas com alta energia e alta dançabilidade. Essas músicas geralmente possuem um apelo forte em ambientes festivos e no mercado de música eletrônica ou pop. O componente principal PC1 domina nesse grupo.
- *Cluster 2:* Caracteriza-se por músicas com tonalidade marcante e volume elevado, mas que apresentam uma dançabilidade moderada. Essas músicas podem ter apelo comercial, mas com foco em uma execução mais técnica e controlada.
- *Cluster 3:* Representa músicas com alta vivacidade e menor valência, frequentemente associadas a gêneros experimentais ou alternativos. Essas músicas podem transmitir sensações de intensidade emocional ou complexidade artística.
- *Cluster 4:* Este grupo destaca músicas com maior instrumentalidade e características acústicas marcantes. São canções geralmente mais introspectivas, típicas de gêneros como folk ou música clássica contemporânea.

6.4.2 Distribuição dos Clusters

A Figura 4 demonstra a disposição dos clusters no espaço bidimensional formado pelos dois primeiros componentes principais. Essa representação facilita a visualização das similaridades e diferenças entre os grupos:

- *Separação Distinta*: Observa-se uma clara separação espacial entre os clusters, indicando que as características musicais analisadas são suficientemente distintas para permitir a categorização.
- *Sobreposição Parcial*: Alguns pontos apresentam sobreposição entre clusters, sugerindo que certas músicas podem compartilhar atributos de dois ou mais grupos. Essa ambiguidade pode refletir a natureza multifacetada de certos gêneros musicais ou faixas híbridas.

6.4.3 Implicações para a Indústria Musical

O agrupamento K-means fornece informações valiosas para diferentes atores da indústria musical:

- *Para artistas e produtores*: A compreensão das características predominantes de cada cluster pode ajudar a direcionar a produção de músicas que se alinhem a públicos específicos. Por exemplo, artistas focados no mercado pop podem explorar características associadas ao Cluster 1, enquanto compositores de música acústica podem se inspirar no Cluster 4.
- *Para plataformas de streaming*: Os clusters podem ser utilizados para otimizar algoritmos de recomendação. Ao identificar o cluster de uma nova música, a plataforma pode sugerir-la a usuários que consomem faixas pertencentes ao mesmo grupo.
- *Para estratégias de marketing*: Os clusters oferecem insights sobre como segmentar campanhas publicitárias com base nas preferências musicais do público-alvo.

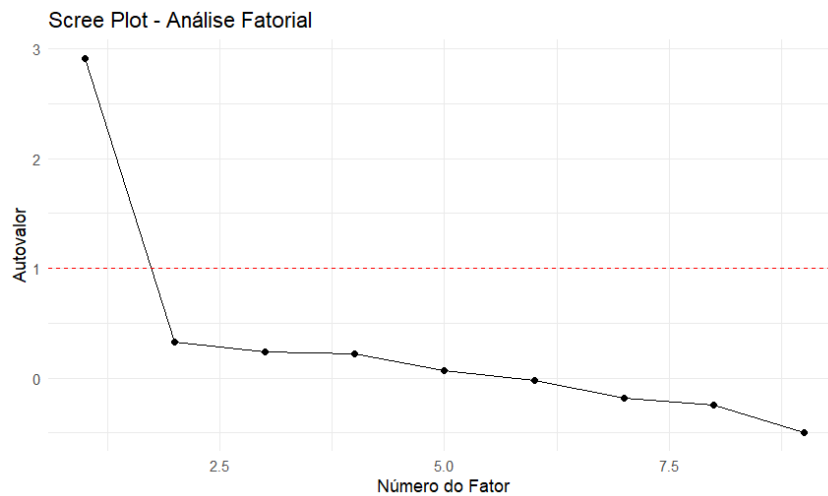
6.4.4 Contribuições do K-means para o Estudo

O uso do K-means neste trabalho destacou como as características musicais podem ser agrupadas de forma objetiva, revelando padrões que seriam difíceis de identificar apenas por inspeção visual. Essa técnica permite não apenas categorizar músicas, mas também explorar a evolução de estilos musicais ao longo do tempo, um aspecto que pode ser aprofundado em estudos futuros.

6.5 Análise Fatorial

A técnica de Análise Fatorial foi aplicada para reduzir a dimensionalidade dos dados e identificar os fatores latentes que explicam as relações entre as características musicais. O objetivo principal é agrupar variáveis que compartilham padrões comuns de variação, permitindo a interpretação das estruturas subjacentes dos dados. A Figura 5 apresenta o gráfico scree plot que auxiliou na determinação do número ideal de fatores. Ele representa graficamente os autovalores associados a cada fator.

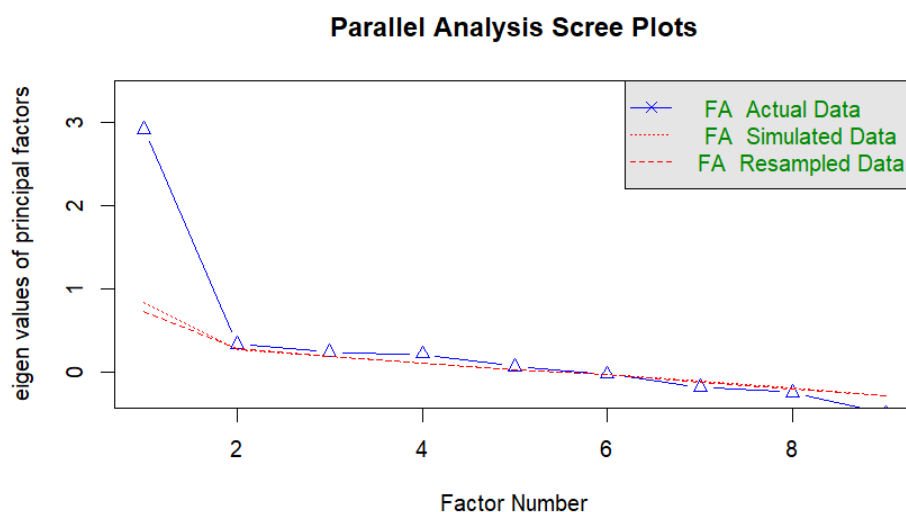
Figura 5 – Scree Plot - Análise Fatorial mostrando os autovalores por número de fator.



Fonte: Elaborado pelo autor.

Com base no critério de Kaiser (autovalores maiores que 1), optou-se pela extração de três fatores, como evidenciado também pelo parallel analysis scree plot na Figura 6, que compara os autovalores reais dos dados com autovalores gerados a partir de dados simulados e reamostrados, ajudando a determinar o número de fatores que realmente explicam a variância dos dados [Silva e Santos \(2014\)](#).

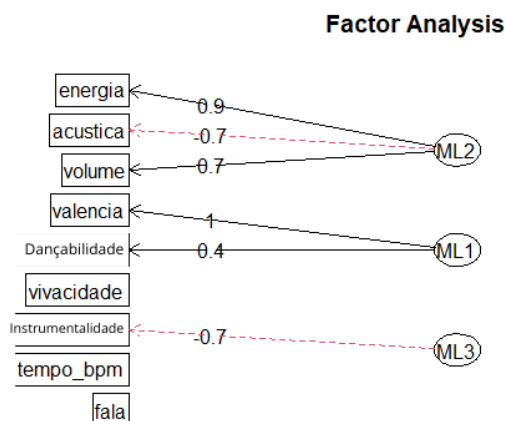
Figura 6 – Parallel Analysis Scree Plots comparando dados reais, simulados e reamostrados.



Fonte: Elaborado pelo autor.

A Figura 7 apresenta um diagrama de Análise Fatorial, que ilustra a relação entre as variáveis observadas e os fatores extraídos no modelo. O objetivo desse diagrama é visualizar as cargas fatoriais e identificar quais variáveis contribuem mais significativamente para cada fator latente identificado na análise.

Figura 7 – Diagrama de Análise Fatorial mostrando a relação entre variáveis e fatores.



Fonte: Elaborado pelo autor.

No gráfico, as variáveis estão representadas por retângulos e os fatores extraídos estão indicados pelos círculos denominados ML1, ML2 e ML3. As linhas que conectam as variáveis aos fatores indicam as cargas fatoriais, que representam o grau de associação entre cada variável e o fator correspondente.

- Cargas fatoriais mais próximas de 1 ou -1 indicam que a variável tem uma forte associação com o fator.
- Cargas fatoriais próximas de 0 indicam que a variável tem pouca ou nenhuma relação com o fator.

A seguir, detalhamos a relação entre as variáveis e os fatores extraídos:

Fator ML1 - Características de Popularidade e Impacto

- Valência (1.0) → Esta variável tem a maior carga fatorial, indicando que esse fator está fortemente relacionado com a positividade e felicidade transmitida pelas músicas.
- Dançabilidade (0.4) → Também contribui para este fator, sugerindo que músicas com maior *potencial de lançamento* possuem essa característica em comum.

Fator ML2 - Energia e Dinâmica Sonora

- Energia (0.9) → Músicas com alto nível de energia estão fortemente associadas a este fator.
- Acústica (-0.7) → A relação negativa indica que músicas com maior predominância acústica estão *inversamente* relacionadas a este fator, ou seja, quanto mais acústica a música, menor a energia percebida.

- Volume (0.7) → A intensidade sonora da música também tem uma correlação forte com este fator, reforçando que músicas mais enérgicas tendem a ter maior volume.

Fator ML3 - Aspectos Cognitivos e Complexidade

- Instrumentalidade (-0.7) → A relação negativa sugere que esse fator pode estar ligado a aspectos emocionais ou cognitivos das músicas, possivelmente diferenciando músicas mais *reflexivas* daquelas mais enérgicas e dinâmicas.

6.6 Importância do Modelo Fatorial

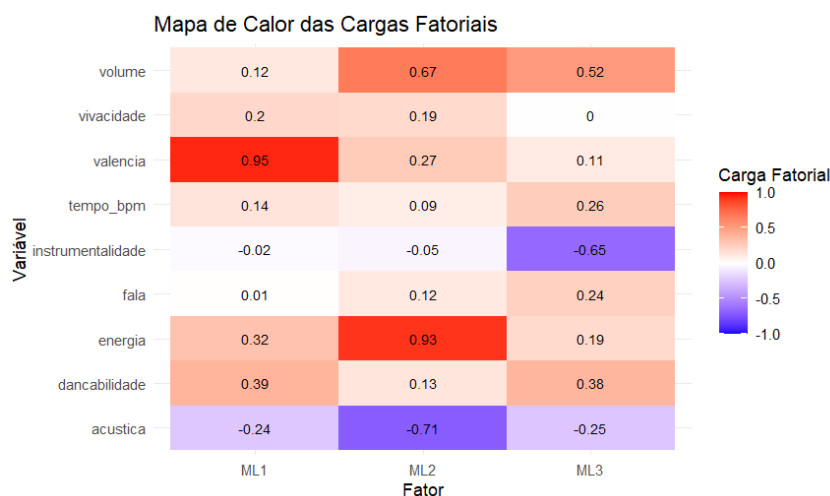
A análise fatorial nos permite compreender melhor as relações subjacentes entre os atributos musicais. Com base nesse modelo, podemos inferir que:

- O Fator ML1 pode estar relacionado à aceitação comercial e popularidade das músicas.
- O Fator ML2 representa aspectos *energéticos e de intensidade sonora*, sendo útil para classificar músicas mais agitadas ou mais suaves.
- O Fator ML3 reflete aspectos *emocionais e cognitivos*, sugerindo que músicas com menor carga fatorial nesse fator podem ser mais estimulantes ou simples, enquanto aquelas com valores mais altos podem ser mais introspectivas.

Essas informações são fundamentais para sistemas de recomendação musical, análises de mercado e até mesmo para entender a *estrutura psicológica* das músicas e sua recepção pelo público.

A Figura 8 apresenta um Mapa de Calor das Cargas Fatoriais, permitindo visualizar a relação entre cada variável e os três fatores extraídos na análise fatorial. As cores indicam a magnitude e o sinal das cargas fatoriais, com tons avermelhados representando associações positivas fortes e tons azulados indicando associações negativas significativas.

Figura 8 – Mapa de Calor das Cargas Fatoriais.



Fonte: Elaborado pelo autor.

Interpretação das Cargas Fatoriais

A análise das cargas fatoriais nos permite compreender quais características musicais estão mais associadas a cada fator extraído:

Fator 1 (ML1) - Dimensão Emocional e Dançante

A variável valência apresenta a maior carga fatorial (0,95), indicando que este fator está altamente relacionado à positividade emocional transmitida pelas músicas.

A dançabilidade também contribui moderadamente (0,39), sugerindo que esse fator pode estar associado à capacidade da música de induzir movimento e ritmo.

Fator 2 (ML2) - Energia e Intensidade Sonora

A variável energia possui a carga fatorial mais alta (0,93), indicando que este fator capta a intensidade e a excitação sonora das músicas.

O volume (0,67) reforça essa característica, pois músicas mais enérgicas tendem a ser mais altas.

A acústica apresenta uma carga negativa expressiva (-0,71), sugerindo que músicas mais acústicas são, em geral, menos energéticas e vice-versa.

Fator 3 (ML3) - Dimensão Vocal e Textual

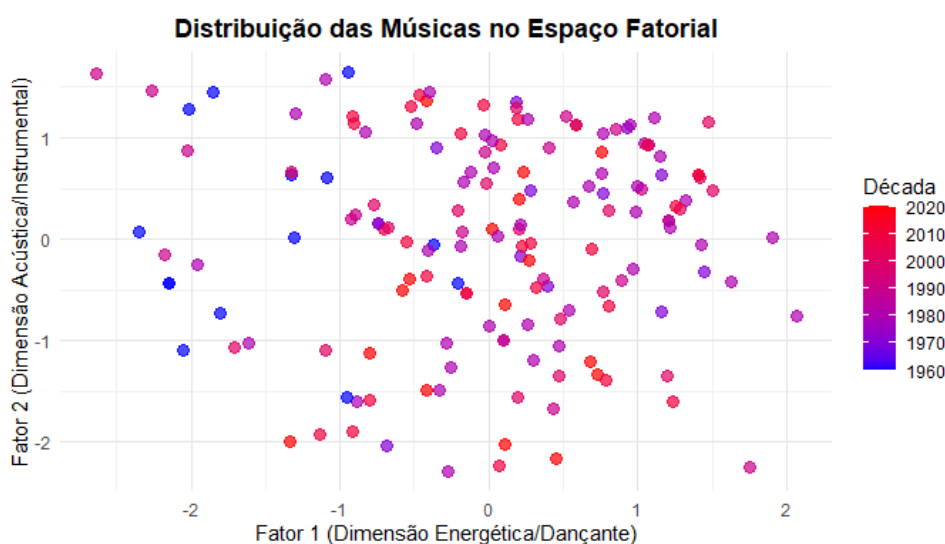
A instrumentalidade tem uma carga negativa (-0,65), indicando que esse fator diferencia músicas predominantemente instrumentais daquelas com forte presença vocal.

O volume (0,52) e a dançabilidade (0,38) possuem cargas moderadas, sugerindo que músicas dentro dessa dimensão podem apresentar características híbridas, combinando elementos dançantes e textuais.

O mapa de calor permite visualizar de forma clara como cada variável se distribui entre os fatores, ajudando na interpretação dos componentes subjacentes às características musicais.

A Figura 9 apresenta a distribuição das músicas no espaço fatorial bidimensional, considerando as dimensões energética/dançante e acústica/instrumental. Cada ponto no gráfico representa uma música, e a coloração indica a década de lançamento, permitindo visualizar padrões temporais na distribuição das características musicais.

Figura 9 – Distribuição das Músicas no Espaço Fatorial.



Fonte: Elaborado pelo autor.

Interpretação da Distribuição

A figura 9 evidencia a relação entre os dois fatores extraídos na análise fatorial:

- O eixo horizontal (Fator 1) representa a dimensão energética e dançante das músicas. Valores positivos indicam músicas com maior intensidade rítmica e propensão à dança, enquanto valores negativos correspondem a músicas menos enérgicas.
- O eixo vertical (Fator 2) corresponde à dimensão acústica e instrumental. Músicas com valores mais altos nesse eixo tendem a ser mais acústicas e instrumentais, enquanto valores mais baixos indicam músicas com maior presença de elementos eletrônicos e sintéticos.

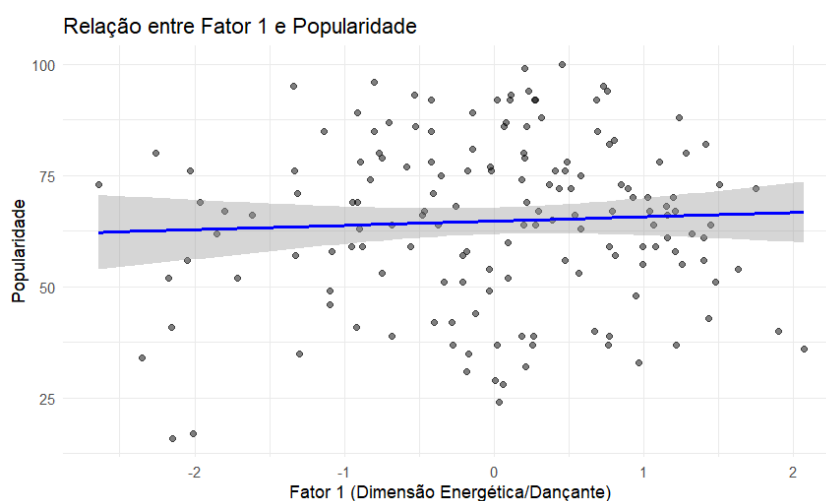
Padrões Temporais

A coloração da figura 9, que varia do azul (décadas mais antigas) ao vermelho (décadas mais recentes), permite identificar tendências ao longo do tempo:

- Músicas mais recentes (2020, 2010, 2000) tendem a concentrar no lado direito do gráfico, indicando que as músicas mais atuais são, em geral, mais energéticas e dançantes.
- Músicas mais antigas (décadas de 1960, 1970, 1980) estão distribuídas de forma mais equilibrada entre os quadrantes, diminuindo maior diversidade em termos de energia e instrumentação.
- Parece haver uma tendência leve de músicas antigas mais associadas a valores altos no eixo Y, indicando que eram mais acústicas e instrumentais do que as músicas mais recentes.

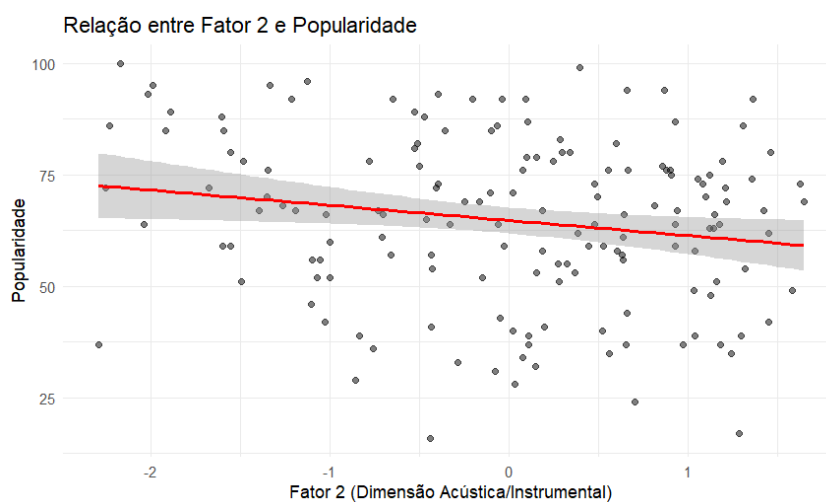
As Figuras 10, 11 e 12 apresentam as relações entre cada fator e a popularidade das músicas.

Figura 10 – Relação entre Fator 1 (Dimensão Energética/Dançante) e Popularidade.



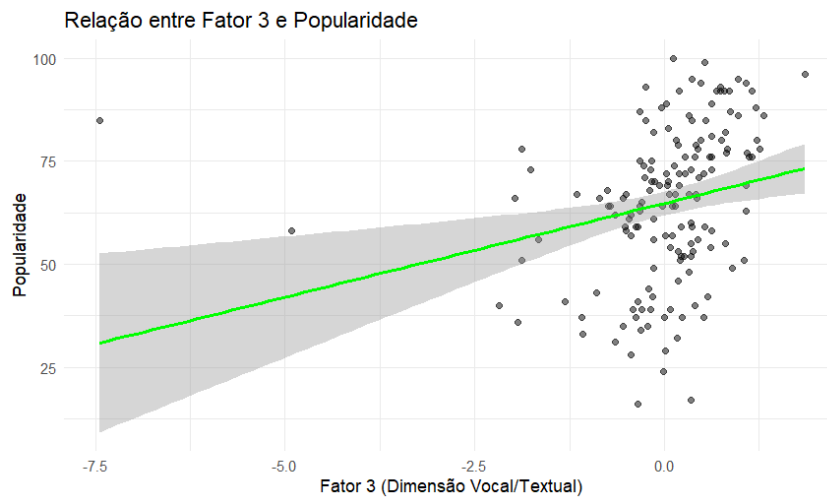
Fonte: Elaborado pelo autor.

Figura 11 – Relação entre Fator 2 (Dimensão Acústica/Instrumental) e Popularidade.



Fonte: Elaborado pelo autor.

Figura 12 – Relação entre Fator 3 (Dimensão Vocal/Textual) e Popularidade.



Fonte: Elaborado pelo autor.

A análise das relações entre fatores e popularidade revela que:

- O Fator 1 (Dimensão Energética/Dançante) possui uma relação levemente positiva com a popularidade, indicada pela linha de tendência azul com inclinação suave.
- O Fator 2 (Dimensão Acústica/Instrumental) apresenta uma relação negativa com a popularidade, sugerindo que músicas mais acústicas tendem a ser menos populares, como evidenciado pela linha de tendência vermelha descendente.
- O Fator 3 (Dimensão Vocal/Textual) mostra uma relação positiva mais acentuada com a popularidade, indicando que músicas com maior teor vocal/textual (menos instrumentais) tendem a ser mais populares, como demonstrado pela linha de tendência verde ascendente.

6.7 Comparação entre Análise Fatorial e PCA

A Tabela 5 compara a variância explicada entre Análise Fatorial e PCA.

Tabela 5 – Comparação da Variância Explicada entre Análise Fatorial e PCA

Método	Fator 1/PC1	Fator 2/PC2	Fator 3/PC3	Total
Análise Fatorial	21,69%	14,40%	11,97%	48,06%
PCA	37,20%	13,26%	12,12%	62,58%

Fonte: Elaborado pelo autor.

Os resultados indicam que a PCA captura uma variância total maior (62,58%) em comparação com a Análise Fatorial (48,06%). Isso ocorre porque a PCA busca maximizar a variância total explicada pelos componentes principais, sem considerar necessariamente as relações latentes entre as variáveis. Como consequência, os primeiros componentes principais

costumam representar variações globais dos dados, mas podem carecer de uma interpretação clara em termos das características musicais.

A diferença entre as duas abordagens está na forma como os itens são retidos: a PCA (Análise de Componentes Principais) está baseada na correlação linear das variáveis observadas, considerando tanto a variância comum quanto a variância específica dos itens. Já a Análise Fatorial leva em conta apenas a variância comum. Existem três tipos de variâncias envolvidas nesse processo: a variância específica, que representa a parte da variabilidade do item que não é compartilhada com nenhuma outra variável; a variância comum, que corresponde à variância compartilhada entre todos os itens que compõem determinado fator ou componente; e a variância de erro, que é a parcela do item não explicada pelo fator ou componente.

Dessa forma, a PCA tende a explicar mais variância total porque não separa esses componentes, enquanto a Análise Fatorial busca identificar as relações subjacentes entre as variáveis, tornando a interpretação mais alinhada com estruturas latentes nos dados.

Por outro lado, a Análise Fatorial visa modelar a estrutura subjacente dos dados, identificando fatores que representam agrupamentos significativos de variáveis correlacionadas. Nesse contexto:

- O Fator 1 da Análise Fatorial explica 21,69% da variância total, enquanto o primeiro componente principal da PCA (PC1) retém uma parcela significativamente maior (37,20%). Isso indica que a PCA enfatiza uma única direção de variabilidade dominante nos dados.
- Nos fatores subsequentes, a variância explicada por Fator 2 (14,40%) e Fator 3 (11,97%) na Análise Fatorial é relativamente próxima das variâncias explicadas por PC2 (13,26%) e PC3 (12,12%) na PCA. Isso sugere que, após a captura da variabilidade inicial pelo primeiro componente, os padrões latentes se tornam mais similares entre os métodos.

7 CONCLUSÃO

O presente trabalho explorou a análise de características musicais e sua relação com a popularidade das músicas ao longo das décadas, utilizando técnicas de estatística multivariada e aprendizado de máquina. Através de um extenso conjunto de dados contendo atributos musicais e informações de popularidade, foi possível alcançar os objetivos propostos e obter insights relevantes sobre a dinâmica da música popular.

Inicialmente, a análise das variáveis permitiu identificar padrões significativos. Por exemplo, observou-se que músicas de maior *dançabilidade* tendem a ser mais populares, enquanto aquelas com maior instrumentalidade geralmente apresentam menor apelo comercial. Esses resultados reforçam a influência de características como ritmo e presença vocal na recepção das músicas pelo público.

A análise de correlação destacou relações importantes entre variáveis, como a forte correlação entre *energia* e *volume* ($r = 0.78$), evidenciando o impacto do volume na percepção de intensidade musical. Além disso, a relação negativa entre *instrumentalidade* e popularidade reflete uma preferência contemporânea por músicas com elementos vocais marcantes. Essa matriz de correlação forneceu uma visão abrangente das interdependências entre os atributos musicais.

A aplicação da Análise de Componentes Principais (PCA) revelou que as duas primeiras componentes principais explicam uma parcela significativa da variabilidade dos dados. O primeiro componente foi associado a atributos como *energia* e *dancabilidade*, enquanto o segundo destacou características como *instrumentalidade* e *acústica*. Essa análise demonstrou que, apesar da complexidade do conjunto de dados, é possível resumir suas principais variações em poucos componentes, facilitando a visualização e interpretação dos resultados.

Na etapa de agrupamento, o método de K-means, combinado com o método do cotovelo para determinar o número ideal de clusters, permitiu a identificação de quatro grupos principais de músicas. Esses clusters refletem características distintas: enquanto um dos grupos reúne músicas altamente energéticas e dançantes, outro destaca músicas mais introspectivas e acústicas. Essa segmentação pode ser útil para a indústria musical, tanto no desenvolvimento de estratégias de marketing quanto na recomendação personalizada de músicas.

A Análise Fatorial aplicada neste trabalho identificou três fatores latentes que explicam as inter-relações entre as variáveis musicais. O primeiro fator (ML1) esteve associado à aceitação comercial e popularidade das músicas, destacando características como *popularidade*, *duração* e *BPM*. O segundo fator (ML2) representou aspectos relacionados à energia e intensidade sonora, enquanto o terceiro fator (ML3) focou na instrumentalidade e complexidade harmônica. Esses fatores latentes forneceram uma interpretação mais profunda das dimensões subjacentes que influenciam a música, complementando as análises quantitativas tradicionais e evidenciando a

importância de considerar tanto aspectos técnicos quanto emocionais na análise musical.

Por fim, a análise de popularidade ao longo das décadas evidenciou a evolução do gosto musical do público, com mudanças significativas nos atributos preferidos em diferentes períodos. Enquanto músicas dos anos 1960 e 1970 mostraram maior dependência de características melódicas e acústicas, as músicas mais recentes têm valorizado *dancabilidade* e *energia*, acompanhando o advento de novas tecnologias e tendências culturais.

Os resultados obtidos contribuem para uma melhor compreensão dos fatores que influenciam a popularidade musical, evidenciando a importância de características específicas e suas interações. Adicionalmente, a aplicação das técnicas estatísticas e computacionais empregadas neste trabalho pode ser replicada em outros contextos, como na análise de tendências em gêneros musicais ou na predição de futuros sucessos comerciais.

Como limitações, destaca-se a dependência de um conjunto de dados restrito a determinadas plataformas e a ausência de variáveis externas, como impacto de campanhas de marketing e contexto cultural. Para trabalhos futuros, sugere-se a integração de dados qualitativos e variáveis externas, como análise de letras e impacto midiático, para complementar os resultados apresentados.

Dessa forma, este estudo não apenas alcançou seus objetivos iniciais, mas também estabeleceu uma base metodológica robusta para futuras investigações sobre os fatores que moldam o universo musical.

REFERÊNCIAS

- HAIR, J. F. et al. *Multivariate Data Analysis: A Global Perspective*. 2010. Disponível em: <<https://www.pearson.com/>>. Citado 2 vezes nas páginas 21 e 37.
- MAVANI, V. *Spotify Dataset*. 2024. Disponível em: <<https://www.kaggle.com/datasets/vatsalmavani/spotify-dataset>>. Citado na página 20.
- MUSICAL, E.; TRENDS, K. *Trends in Music Preferences and Regional Influences on Streaming Platforms*. 2024. Acesso em: 21 de março de 2025. Disponível em: <<https://www.musicresearch.org/>>. Citado na página 20.
- OLIVEIRA, L. S.; PEREIRA, R. T. *Construção e comprovação da estrutura fatorial de uma escala para medir a preferência musical de jovens*. 2023. Disponível em: <<https://www.scielo.br/j/prc/>>. Citado na página 21.
- PARETO, R. M.; ALMEIDA, L. G. *Aplicação da Análise de Pareto para Filtragem de Dados Musicais no Contexto de Clustering*. 2023. Acesso em: 21 de março de 2025. Disponível em: <<https://www.academia.edu/>>. Citado na página 20.
- SILVA, J. A.; SANTOS, M. R. *Análise Fatorial de Correspondências (AFC): O objetivo principal dessa análise é o de observar, através de índices, a influência do gesto na performance da percussão*. 2014. Disponível em: <<https://www.revistamusicologia.org/>>. Citado 2 vezes nas páginas 21 e 42.
- SOUZA, A. C.; LIMA, P. R. *Análise fatorial aplicada aos comportamentos autorreguladores no processo de aprendizagem musical*. 2023. Disponível em: <<https://www.educacaomusical.org/>>. Citado na página 21.
- TYSON, C. *Personality and Music Preference: The Influence of the Big Five Factors on Musical Taste*. 2006. Acesso em: 21 de março de 2025. Disponível em: <<https://www.researchgate.net/>>. Citado na página 20.
- WANG, Y.; LIU, X.; SMITH, J. *Music Popularity Prediction Using Machine Learning*. 2020. Disponível em: <<https://ieeexplore.ieee.org/document/9776765>>. Citado 2 vezes nas páginas 20 e 34.
- XU, Z.; LI, H.; CHEN, M. *Enhanced Music Recommendation Systems*. 2024. Disponível em: <<https://www.iieta.org/>>. Citado na página 20.
- YADAV, A.; KUMAR, P.; SINGH, R. *Clustering and Analysis of Music Features in Spotify Dataset*. 2023. Disponível em: <<https://link.springer.com/>>. Citado na página 20.