



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Melodias e Dados: Análise e Previsão do Consumo Musical em Portugal

Tiago Augusto Mendes Teixeira Barbosa

Mestrado em Sistemas Integrados de Apoio à Decisão

Orientador:

Doutor Fernando Manuel Marques Batista, Professor Associado,

Iscte – Instituto Universitário de Lisboa

Co-Orientadora:

Doutora Elsa Alexandra Cabral da Rocha Cardoso, Professora Associada,

Iscte – Instituto Universitário de Lisboa

Outubro, 2024



TECNOLOGIAS
E ARQUITETURA

Departamento de Ciências e Tecnologias da Informação

Melodias e Dados: Análise e Previsão do Consumo Musical em Portugal

Tiago Augusto Mendes Teixeira Barbosa

Mestrado em Sistemas Integrados de Apoio à Decisão

Orientador:

Doutor Fernando Manuel Marques Batista, Professor Associado,

Iscte – Instituto Universitário de Lisboa

Co-Orientadora:

Doutora Elsa Alexandra Cabral da Rocha Cardoso, Professora Associada,

Iscte – Instituto Universitário de Lisboa

Outubro, 2024

Agradecimentos

Aos meus orientadores, Professor Doutor Fernando Batista e Professora Doutora Elsa Cardoso, expresso a minha mais sincera gratidão por toda a paciência e orientação ao longo deste percurso.

À minha família, especialmente aos meus pais e à minha avó, pelo cuidado constante e pelo apoio incondicional em todos os momentos. Ao meu irmão, por elevar sempre a minha autoestima, incentivar-me a nunca desistir dos meus objetivos e a manter o foco.

Aos meus amigos, Manuel Casimiro e José Seno, que compartilharam comigo a jornada deste mestrado e que sempre estiveram presentes, motivando-me e apoiando-me nos momentos mais desafiantes. Ao meu amigo, Miguel Soldado, agradeço pelo suporte e pela ajuda nas dúvidas mais complexas.

Por fim, à minha namorada, Catarina Fonseca, que desde o início da escolha deste tema desafiador até ao último momento da entrega esteve sempre ao meu lado, ajudando-me a superar cada obstáculo.

Resumo

A presente investigação explora a interseção dinâmica entre análise de dados e comportamento humano no contexto do consumo de música em Portugal. O cenário da indústria musical continua a transformar-se, tanto do ponto de vista económico quanto social.

Através da utilização de técnicas de estatística e aprendizagem automática, o estudo procura identificar padrões nestes conjuntos de dados para prever e compreender os fatores que influenciam os comportamentos individuais e coletivos de consumo de música na população portuguesa. Para uma melhor compreensão dos resultados, várias partes da análise de dados nesta pesquisa foram comparadas com a França, um país europeu com características socioculturais e linguísticas semelhantes às de Portugal.

A escolha do Spotify justifica-se pela sua disponibilização de *rankings* diários e pelo interesse no mercado global de *streaming*. Aplicando modelação preditiva, o estudo procura identificar correlações entre variáveis como localização geográfica, influências culturais e preferências musicais.

Os resultados desta investigação poderão contribuir para os campos da análise da indústria musical e comportamento do consumidor. Compreender as complexidades dos padrões de consumo de música em Portugal pode auxiliar intervenientes, incluindo artistas, editoras e serviços de *streaming*, na adaptação de conteúdo, estratégias de marketing e experiências do utilizador de acordo com as preferências e o contexto cultural do público português. Além disso, o estudo pode servir como um modelo para análises semelhantes em outros contextos culturais, promovendo uma compreensão mais detalhada do contexto global do consumo de música.

Palavras-chave: Spotify, Inteligência Artificial, Dados, Comportamento Humano, Música

Abstract

This research explores the dynamic intersection of data analysis and human behavior within the context of music consumption in Portugal.

Leveraging advanced statistical and machine learning techniques, the study aims to discern patterns in these datasets to predict and understand the nuanced factors influencing individual and collective music consumption behaviors in the Portuguese population. For a clearer understanding of the results, several parts of the data analysis in this research were compared with France, a European country with sociocultural and linguistic characteristics similar to those of Portugal.

The research methodology involves the collection and analysis of data from a streaming platform named “Spotify”. The choice of Spotify is justified by its daily rankings and the platform's relevance in the global streaming market. Utilizing predictive modeling, the study seeks to identify correlations between variables such as geographic location, cultural influences, and musical preferences. The findings of this research are expected to contribute valuable insights to the fields of music industry analytics and consumer behavior. Understanding the intricacies of music consumption patterns in Portugal can aid stakeholders, including artists, record labels, and streaming services, in tailoring their content, marketing strategies, and user experiences to align with the preferences and cultural context of the Portuguese audience. Moreover, the study may serve as a blueprint for similar analyses in other cultural contexts, fostering a more nuanced understanding of the global landscape of music consumption.

Keywords: Spotify, Artificial Intelligence, Data, Human Behavior, Music

Índice

Agradecimentos	i
Resumo	iii
Abstract	v
Índice	vii
Índice de Figuras	ix
Índice de Tabelas	xi
Capítulo 1	1
Introdução	1
1.1. Motivação e Enquadramento do Tema	1
1.2. Objetivos e Perguntas de Investigação.....	1
1.3. Metodologia	2
1.4. Estrutura e Organização da Dissertação.....	2
Capítulo 2	5
Estado da Arte	5
2.1. Evolução do Consumo de Música	5
2.2. Spotify e “ <i>streaming</i> ” de música.....	6
2.3. Quotas de Música na Rádio.....	7
2.3.1. Análise comparativa entre Portugal e França.....	7
2.3.2. O impacto das quotas de música nas preferências de consumo.....	10
2.3.3. Curadoria Musical e Impacto das Plataformas Digitais.....	10
2.4. Inteligência Artificial: evolução e potencial futuro preditivo na música	11
2.4.1. Evolução da Inteligência Artificial	11
2.4.2. O Futuro da Música: Aplicações Preditivas da Inteligência Artificial	12
Capítulo 3	15
Compreensão dos Dados	15
3.1. Base de Dados	15
3.1.1. Construção da Base de Dados - 1ª fase	15

3.1.2. Construção da Base de Dados - 2ª fase.....	15
3.1.3. Construção da Base de Dados - 3ª fase.....	18
3.2. Distribuições de frequência em relação às características de áudio.....	20
Capítulo 4.....	25
Modelação.....	25
4.1 Métodos de Previsão Temporal.....	25
4.1.1 Prophet.....	25
4.1.2 Neural Prophet.....	27
4.1.3 LSTM.....	27
4.1.4 XGBoost.....	28
Capítulo 5.....	31
Discussão de Resultados.....	31
5.1 Evolução Histórica do Consumo de Música: Comparação entre Portugal e França.....	31
5.2. Padrões e características recorrentes nas Músicas Populares em Portugal.....	33
5.3. Previsão do Consumo de Música Portuguesa no Spotify: Um Estudo Comparativo.....	37
Capítulo 6.....	41
Conclusões.....	41
6.1. Limitações.....	41
6.2. Futuras Análises.....	42
Referências Bibliográficas.....	43

Índice de Figuras

Figura 1 - Spotify API	16
Figura 2 - Genius API.....	19
Figura 3 - Flow chart da Construção da Base de dados.....	20
Figura 4 - Média de acousticness em Portugal e França.....	21
Figura 5 - Média de danceability em Portugal e França	22
Figura 6 - Média de energy em Portugal e França.....	23
Figura 7 - Média de valence em Portugal e França	23
Figura 8 - Correlação entre características de áudio.....	24
Figura 9 - Prophet workflow	27
Figura 10 - Línguas no Top 200 português.....	31
Figura 11 - Línguas no Top 200 francês.....	31
Figura 12 - Característica ‘acousticness’ ao longo do tempo.....	33
Figura 13 - Distribuições de frequência em relação à característica de áudio	34
Figura 14 - Característica ‘danceability’ ao longo do tempo	34
Figura 15 - Característica ‘energy’ ao longo do tempo	35
Figura 16 - Característica ‘tempo’ (BPM) ao longo do tempo.....	35
Figura 17 - Característica ‘duration_ms’ ao longo do tempo.....	36
Figura 18 - Previsão da Presença de Músicas em Língua Portuguesa no TOP 200 do Spotify (Modelo Prophet).....	37
Figura 19 - Previsão da Presença de Músicas em Língua Portuguesa no TOP 200 do Spotify (Modelo Neural-Prophet).....	38
Figura 20 - Previsão da Presença de Músicas em Língua Portuguesa no TOP 200 do Spotify (Modelo LSTM).....	39
Figura 21 - Previsão da Presença de Músicas em Língua Portuguesa no TOP 200 do Spotify (Modelo XGBoost).....	39

Índice de Tabelas

Tabela 1 - Base de dados de Portugal após recolha de dados	15
Tabela 2 - Descrição das características	16
Tabela 3 - Base de dados após a extração de características	18
Tabela 4 - Base de dados após a deteção do idioma.....	19
Tabela 5 - Características lag	28
Tabela 6 - Tabela de comparação de resultados dos modelos.....	40

Capítulo 1

Introdução

A presente dissertação explora a interseção entre a análise de dados e o comportamento humano no consumo de música em Portugal, com foco na utilização de técnicas de inteligência artificial e aprendizagem automática para prever tendências de consumo musical.

Neste capítulo será efetuado um breve enquadramento do tema e definição dos objetivos e questões de pesquisa. Posteriormente, será mencionado a metodologia que será aplicada assim como a estrutura do projeto em apreço.

1.1. Motivação e Enquadramento do Tema

A indústria musical tem sofrido várias revoluções ao longo dos anos que têm vindo a alterar o consumo e a produção musical. Com isto, novas músicas e artistas têm ganho popularidade nas *top charts*, graças às novas plataformas de *streaming*, como é o caso do *Spotify*, e torna-se crucial compreender o que leva uma música a obter sucesso e como cativam os ouvintes numa sociedade de consumo acelerado.

A atual pesquisa referente à temática “Análise de dados para previsão do comportamento humano no consumo de música portuguesa” é de elevada importância para os artistas e editoras musicais que poderão utilizar os dados e modelos de previsão para otimizar estratégias de promoção, decisões criativas e potencialmente reduzir os riscos associados ao lançamento de novas músicas. Desta forma, ao analisar profundamente os conjuntos de dados musicais disponíveis, esta pesquisa procura identificar, através da ciência dos dados e inteligência artificial, padrões e evoluções que influenciam as produções musicais para obter os dados referentes a este fenómeno multidimensional. Ainda, será estudada a presença da língua nacional nas *charts* musicais e o seu futuro no contexto do mercado digital e globalizado. Ao analisar o crescimento das faixas em português nos principais *rankings*, pretende-se compreender como fatores culturais e políticas de promoção da música local influenciam o comportamento dos ouvintes em Portugal e projetar a evolução dessa tendência para os próximos anos.

1.2. Objetivos e Perguntas de Investigação

Com o intuito de direcionar e orientar a investigação ao definir claramente os objetivos e propósitos do estudo, foram formuladas três questões de pesquisa. Estas questões guiam a análise, caracterizando-se por serem claras, concisas, mensuráveis e viáveis.

As questões de pesquisa definidas para este estudo são:

1. *Em comparação com o mercado francês, qual foi a evolução do consumo de música portuguesa no Spotify em Portugal ao longo dos anos?*
2. *Em comparação com França, é possível identificar um padrão de características recorrentes nas músicas populares em Portugal?*
3. *Com base em tendências passadas, é viável desenvolver um modelo de previsão para antecipar futuras mudanças no consumo de música portuguesa em Portugal?*

Com base nessas questões, os principais objetivos são:

- Explorar e analisar as tendências musicais ao longo do tempo;
- Procurar prever a evolução do mercado musical português;
- Identificar os elementos-chave que impulsionam a popularidade e aceitação de uma composição dentro do mercado musical de Portugal.

1.3. Metodologia

Nesta dissertação, foi utilizada a metodologia CRISP-DM (*Cross-Industry Standard Process for Data Mining*) para estruturar a investigação e orientar todas as etapas da análise e desenvolvimento dos modelos preditivos.

A CRISP-DM é uma abordagem amplamente reconhecida e adotada em projetos de mineração de dados, composta por seis fases principais: compreensão do negócio; compreensão dos dados; preparação dos dados; modelagem; avaliação; implementação.

Na primeira etapa, foi realizada uma análise aprofundada do mercado musical português, com a definição clara dos objetivos da dissertação e das questões de pesquisa a serem respondidas.

Nas segunda e terceira fases, foram definidos os dados a serem recolhidos e efetuou-se a limpeza e o processamento necessário dos mesmos para garantir sua adequação à análise.

Por fim, nas últimas fases, foram selecionados, testados e avaliados os modelos preditivos com o objetivo de prever futuras alterações no consumo de música portuguesa. A eficácia desses modelos foi medida e comparada, permitindo uma visão clara sobre as tendências futuras do mercado musical em Portugal.

1.4. Estrutura e Organização da Dissertação

Esta dissertação está organizada em seis capítulos, que se apresentam da seguinte forma:

Capítulo 2. Estado de Arte – Neste capítulo, apresentarei a pesquisa existente e as principais teorias, abordando a evolução de várias temáticas relevantes para o estudo, as quais ajudarão a explicar as possíveis conclusões do trabalho.

Capítulo 3. Compreensão dos Dados – Este capítulo descreve o processo de criação da base de dados e detalha os componentes considerados na sua construção.

Capítulo 4. Modelação – Aqui serão apresentados os diferentes métodos de previsão temporal que utilizei, explicando de que forma apliquei os modelos ao estudo.

Capítulo 5. Discussão dos Resultados – Neste capítulo, analiso e interpreto os resultados obtidos, estabelecendo relações com a literatura existente.

Capítulo 6. Conclusões – São apresentadas as principais conclusões do estudo, bem como as limitações encontradas e sugestões para futuras linhas de investigação.

Capítulo 2

Estado da Arte

O presente capítulo tem como objetivo apresentar a pesquisa e as teorias existentes face ao tema da dissertação, o que permitirá compreender as bases teóricas e os estudos anteriores.

Nesta secção, serão abordados temas como a evolução do consumo de música com a digitalização, o impacto das plataformas de *streaming* (como o Spotify), a influência das quotas de música na rádio e o papel crescente da Inteligência Artificial na previsão de tendências musicais.

2.1. Evolução do Consumo de Música

A digitalização da música, através das plataformas de *streaming*, como o Spotify, revolucionaram a forma como consumimos, produzimos e distribuimos música. Este acontecimento favoreceu as editoras ao permitir a reprodução infinita das músicas sem custos adicionais. Para os consumidores, esta reprodução também não gera custos elevados, “uma vez que apenas pagam uma taxa mensal fixa para subscrever uma plataforma como o Spotify” (Bello & Garcia, 2021, p.2).

O consumo de música sofreu uma transformação significativa com a digitalização e a tecnologia MP3. Cockrill *et al.* (2011) destacam que os consumidores passaram a ter acesso às suas canções preferidas em qualquer altura e lugar, sem precisar esperar pela transmissão na rádio ou televisão. Esta disponibilidade é um dos fatores cruciais no mercado atual para qualquer tipo de conteúdo, incluindo música e vídeo. Além disso, os consumidores modernos não necessitam de comprar álbuns inteiros, podendo escolher individualmente as músicas que desejam adquirir, o que mudou significativamente os padrões de consumo musical (Pınarbaşı, 2019, pp.267-268). O fenómeno do consumo de música pode ser descrito como o encontro entre a meta musical (canção, lista de reprodução, etc.) e o consumo em várias formas, incluindo discos de vinil, cassetes, CDs ou plataformas digitais. Esta perspetiva inclui fatores pessoais que afetam o consumo, como variáveis psicológicas (Pınarbaşı, 2019, p.267), sendo que segundo o autor “as expressões indicam que as canções com tons mais altos são consideradas mais alegres do que as canções com tons mais baixos”.

As plataformas de *streaming* levaram a um aumento da diversidade e quantidade de músicas existentes no mercado e, conseqüentemente, a uma convergência cultural cada vez mais proeminente. De forma a combater as conseqüências associadas a este fenómeno, os Governos europeus implementaram medidas como as quotas da rádio, onde é obrigatório que 30% da música que passa nas rádios seja música portuguesa. “Way et al. (2020) e os seus colaboradores analisam o comportamento auditivo dos utilizadores do Spotify e constataam que o “*home bias*” - a preferência por artistas nacionais - está a aumentar a nível mundial” (Bello & Garcia, 2021, p.2).

A preferência por artistas nacionais poderá ser influência de vários fatores: algoritmo de recomendação, a *playlist* “New Music Friday”, altura de adesão ao Spotify - “(...) os utilizadores

recentes têm uma preferência mais forte por música local do que aqueles que aderiram à plataforma desde o início” (Bello & Garcia, 2021, p.2).

A crescente uniformidade na música popular contemporânea tem sido objeto de debate, observando-se uma tendência para uma "receita" sonora e de estrutura musical que impulsiona o sucesso comercial. Bello & Garcia (2021) apontam que, embora haja um crescimento de artistas únicos nos *rankings* musicais, a produção musical muitas vezes reflete uma semelhança substancial. Esta convergência acústica pode ser um indício de uma padronização na busca pelo topo das *charts*. No entanto, é interessante notar que essa diversidade de artistas coexiste com uma distribuição mais equitativa de oportunidades para as editoras discográficas. Este fenômeno sugere que, embora a música possa parecer uniforme em certos aspectos, o mercado musical está a tornar-se mais inclusivo e acessível para uma variedade de agentes da indústria: “O aumento da diversidade parece ser impulsionado por uma segmentação do mercado da música e não por uma idiosincrasia uniformemente acentuada do consumo de música” (Bello & Garcia, 2021, p.5).

2.2. Spotify e “*streaming*” de música

A pesquisa sobre o consumo de música na era digital destaca as plataformas de *streaming online* como o principal método de consumo. Segundo Araújo *et al.* (2019), "as plataformas de *streaming online* tornaram-se o método mais importante de consumo de música" (p.859). O Spotify, um dos maiores serviços de *streaming*, desempenha um papel crucial, oferecendo oportunidades financeiras para artistas e editoras, representando, em 2022, “67% de todo o faturamento da indústria musical” (Peral, 2024, p.1).

Com milhões de utilizadores ativos, o Spotify é uma ferramenta valiosa para compreender o valor de um produto ou artista no mercado atual. Pınarbaşı (2019) afirma que "é importante compreender o que torna um produto ou artista valioso. Esse conhecimento pode ser utilizado para promover estratégias de marketing, para decidir o melhor momento para lançar um novo álbum, para alinhar o esforço dos artistas com o interesse do público, etc." (p.859).

Os serviços de *streaming* estão a transformar os padrões de consumo de música, com os consumidores preferindo faixas avulsas em vez de álbuns completos. O Spotify, lançado em 2008 e disponível em 79 mercados, é uma plataforma chave neste cenário, com milhões de utilizadores ativos mensais. De acordo com Pınarbaşı (2019), "o serviço Spotify, no qual os consumidores podem ouvir músicas a partir dos seus vários dispositivos e criar listas de reprodução é um serviço de transmissão *online* de conteúdos áudio" (p.266), sendo que os seus algoritmos de recomendação avançados sejam um dos aspectos mais diferenciadores face às restantes plataformas que acaba por “filtrar para os utilizadores aquilo que entende ser mais relevante para os mesmos, evitando assim uma espécie de “*choice overload*” (Peral, 2024, p.1). A análise de *rankings* e tabelas, como o "Top 50" do Spotify, é essencial para os artistas acompanharem a popularidade do seu trabalho.

A música digital representa uma parte significativa das receitas musicais globais, com os modelos de subscrição impulsionando o mercado. Pınarbaşı (2019) observa que "a música digital representa 58,9% das receitas musicais a nível mundial" (p.268). Além disso, os serviços de *streaming* desempenham papéis importantes na recolha de metadados musicais, no fornecimento de conteúdo aos utilizadores e na personalização da experiência do utilizador através das listas de reprodução, também denominadas de *playlists*. Conforme Pınarbaşı (2019) sugere, "os serviços de *streaming* de música desempenham vários papéis no consumo de música i) recolha de meta musical de artistas/produtores, ii) fornecimento de conteúdos áudio aos utilizadores finais, iii) melhorar a experiência do utilizador através da personalização" (p.268).

A análise dos atributos sonoros da música, como timbre e instrumentação, utilizando recursos como a API do Spotify, onde é possível extrair características de áudio como a acústica, *danceability*, energia, instrumentação, vivacidade, volume, etc., pode fornecer *insights* valiosos sobre o mercado musical.

Em resumo, a pesquisa existente destaca o impacto significativo dos serviços de *streaming*, especialmente do Spotify, no mercado e no consumo de música, oferecendo oportunidades para compreender o valor, promover estratégias de marketing e melhorar a experiência do consumidor.

2.3. Quotas de Música na Rádio

No contexto das políticas culturais, várias entidades reguladoras têm sentido a necessidade de implementar quotas de música. Essas quotas obrigam os meios de comunicação a incluir um mínimo de conteúdo produzido localmente ou na língua nacional, como uma forma de apoiar a cultura local e protegê-la contra uma hegemonia cultural internacional (Kyle and Niu, 2017, p.1). Esta medida visa não só preservar a identidade cultural, mas também promover a diversidade e apoiar a indústria musical nacional.

2.3.1. Análise comparativa entre Portugal e França

A comparação entre França e Portugal no que diz respeito à promoção e consumo de músicas nacionais oferece uma visão interessante das políticas culturais e das preferências dos ouvintes em contextos distintos. Ambos os países possuem regulamentações que incentivam a veiculação de música nacional, mas com características e intensidades diferentes.

França é conhecida por ter políticas robustas de quotas culturais que promovem a música francófona nas rádios e outros meios de comunicação. Estas quotas têm como objetivo garantir uma presença significativa da música nacional no mercado, ajudando a preservar a identidade cultural e a apoiar a indústria musical local. Em contraste, Portugal, embora também adote medidas regulatórias

semelhantes, apresenta um contexto de mercado e políticas diferentes devido à sua dimensão menor e características culturais específicas.

A análise comparativa entre estes dois países é particularmente relevante por várias razões. Primeiro, França e Portugal, apesar das suas diferenças, partilham algumas semelhanças culturais e regionais que permitem uma análise detalhada das políticas culturais e do impacto das quotas no consumo de música nacional. Segundo, ao observar como as políticas de quotas influenciam a oferta e a procura de música nacional em contextos culturais distintos, é possível obter *insights* valiosos sobre a eficácia dessas políticas e os desafios enfrentados em cada país.

Além disso, a comparação pode revelar como as preferências do público e a dinâmica do mercado de música internacional afetam a promoção da música nacional, oferecendo uma perspetiva abrangente sobre o impacto das políticas culturais na indústria musical. Portanto, o foco em França e Portugal proporciona uma análise rica e contextualizada das políticas de quotas culturais e das tendências de consumo de música nacional, contribuindo para uma compreensão mais profunda dessas dinâmicas.

Radiodifusão Sonora: o caso francês

A Radiodifusão Sonora, de acordo com a Autoridade Nacional de Comunicações (ANACOM, 2016), “entende-se como atividade de rádio a organização e fornecimento, com carácter de continuidade, de serviços de programas radiofónicos com vista à sua transmissão para o público em geral”.

Atualmente, cada país apresenta diversos grupos de Radiodifusão Sonora. No que diz respeito a França, a entidade reguladora pública é a Arcom. Existem 5 principais grupos de Radiodifusão Sonora nacionais: Radio France, RTL Group, NRJ Group, Lagardère Active, NextRadioTV. A Radio France engloba 5 estações de rádio transmitidas a nível nacional (France Inter, France Musique, France Culture, Franceinfo, France Bleu) e estações que transmitem localmente (programa FIP e Mouv’), segundo a Arcom (Les Radios En France | Arcom, n.d.).

As quotas de música na rádio foram estabelecidas pela legislação para promover a música de expressão francesa. Segundo a lei de 30 de setembro de 1986, as rádios privadas devem garantir que pelo menos 40% das músicas transmitidas sejam em francês, sendo que metade deverão ser novos talentos ou novas produções, especialmente durante as horas de maior audiência (ARCOM, 2024). Contudo, existem diversas alternativas consoante o tipo de rádio em questão. Segundo ARCOM (2024), as rádios que promovem jovens talentos devem garantir 35% de músicas em francês e 25% de novos talentos. As rádios "de património" têm uma obrigação mais alta, com pelo menos 60% de músicas em francês e um percentual de novas produções que pode chegar até 10%. Além disso, desde 2016, novas regras incluem um regime específico para rádios de "descoberta musical", exigindo pelo menos 15% de novas produções francófonas ou novos talentos. Também existem regras para impedir a repetição excessiva das mesmas músicas francófonas mais populares (ARCOM, 2024).

A implementação e o cumprimento destas quotas são monitorizados pela Arcom (*Autorité de régulation de la communication audiovisuelle et numérique*) que pode aplicar sanções em caso de incumprimento, incluindo avisos, multas ou mesmo a suspensão da licença de transmissão. Esta regulamentação visa assegurar a diversidade musical e o apoio contínuo à criação musical francesa.

Radiodifusão Sonora: o caso português

Em Portugal, existem 4 grupos de Radiodifusão Sonora nacionais, que funcionam por frequências FM, definidos pela Entidade Reguladora para a comunicação. Estes grupos são: RDP, Renascença, Media Capital e Controlinveste (Tomé, 2023, p.19). “O grupo RDP é constituído pelas estações Antena 1; Antena 2; Antena 3; Antena 1, Açores; Antena 1 Madeira; Antena 3 Madeira; RDP África; RDP Internacional e Rádio Lusitânia. Do grupo Renascença, fazem parte a RFM; a Rádio Renascença; a Mega FM e a Rádio SIM. A Media Capital engloba a Rádio Comercial; Cidade; Rádio Clube Português; Best; M80 e Romântica FM. A Controlinveste é constituída pela TSF”. Para além destes grupos, existem inúmeras rádios locais e universitárias.

Ao contrário de França, a lei da quota mínima obrigatória de música portuguesa nas rádios, Lei n.º 54/2010, de 24 de dezembro, entrou em vigor apenas em 2010. Porém, em 2021, a percentagem aumentou para 30% nas programações emitidas entre as 7h e as 20h de segunda a sexta-feira (Morais Leitão, 2024).

Para garantir o cumprimento destas quotas, a Entidade Reguladora para a Comunicação Social poderá exigir às estações de rádio, através de uma plataforma eletrónica, toda a informação necessária ao acompanhamento destas obrigações (incluindo o número de composições difundidas por cada serviço).

De acordo com o Governo Português (2023), “a fixação de um valor superior de quota permitiu, designadamente, verificar que o mesmo não influenciou negativamente as audiências de rádio no período em que vigorou. Além disso, a produção de música portuguesa apresenta hoje uma vitalidade que permite às rádios cumprirem o regime de quotas, sem comprometer a diversidade e a coerência do projeto editorial de cada serviço de programas”.

Contudo, as quotas variam consoante a rádio seja pública ou privada, uma vez que “as rádios públicas têm a quota mínima para a difusão de música portuguesa estabelecida de 60%” (Tomé, 2023, p.20). Esta situação constituiu um problema para os diretores e programadores de rádio, sendo que “(...) dificultam a programação dos programas com géneros mais específicos de música” (Tomé, 2023, p.21). Logo, a limitação e falta de adesão dos programadores de rádio que realizam as playlists tem uma influência direta na presença de música portuguesa nas rádios nacionais.

2.3.2. O impacto das quotas de música nas preferências de consumo

A correlação entre a rádio e a popularidade das músicas tem sido evidente ao longo do tempo. Kyle e Niu (2017, p. 1) afirmam que "a transmissão radiofônica tem historicamente desempenhado um papel crucial no impulsionamento das vendas de música".

O modelo de negócios das estações de rádio baseia-se no fornecimento de conteúdo gratuito para os ouvintes, que não têm a opção de personalizar a sua experiência e estão sujeitos à veiculação de anúncios. Consequentemente, as estações de rádio competem pela audiência através da seleção de playlists diferenciadas, com os ouvintes tendendo a preferir aquelas cujas *playlists* melhor correspondem aos seus gostos musicais (Kyle e Niu, 2017, pp. 4-5).

O impacto das quotas culturais sobre as estações de rádio é condicionado pelas preferências dos ouvintes em relação ao conteúdo local. De acordo com Kyle e Niu (2017), "se a procura por conteúdo local for elevada, as quotas terão uma influência reduzida. No entanto, se os ouvintes apresentarem uma forte preferência por conteúdo não local, as quotas podem afetar adversamente os rendimentos tanto das estações de rádio quanto dos artistas e produtores locais" (idem). Portanto, na ausência de quotas, o conteúdo nacional e local tende a ser menos promovido.

Este contexto afeta diretamente a escolha dos ouvintes entre a rádio e plataformas de *streaming*, uma vez que, "do ponto de vista das estações de rádio, as quotas limitam a sua capacidade de satisfazer as exigências dos ouvintes", levando-os a considerar alternativas competitivas, como o Spotify.

2.3.3. Curadoria Musical e Impacto das Plataformas Digitais

A curadoria musical refere-se ao processo de selecionar e organizar músicas para serem apresentadas ao público, seja em rádios, *playlists* digitais ou concertos.

Este processo envolve decisões que moldam como o público descobre e interage com a música. Na era digital, o papel da curadoria tornou-se mais complexo, uma vez que algoritmos e recomendações automáticas substituem, em muitos casos, a intervenção humana.

Simon Chambers (2022) realizou uma análise aprofundada sobre a curadoria musical em três plataformas principais: rádios, *playlists* digitais (como o Spotify) e concertos ao vivo. O estudo revelou que os concertos oferecem o maior grau de diversidade musical, expondo o público a uma ampla variedade de sons, enquanto as *playlists* digitais são as mais homogêneas, criando uma experiência musical menos variada. Isto resulta do facto de o público utilizar diversas plataformas digitais e o papel da curadoria mediada por algoritmos limita a descoberta de artistas menos conhecidos e estilos mais alternativos.

Esta questão é particularmente importante no contexto do consumo de música em Portugal. O crescente uso de plataformas como o Spotify tem vindo a redefinir a forma como os ouvintes descobrem música. Ao contrário da rádio, que está sujeita a quotas de música local (como discutido na secção 2.4),

as *playlists* digitais tendem a repetir músicas populares e familiarizar os ouvintes com um som padronizado, o que reduz a exposição a novas formas de música. Chambers (2022) observa que essa homogeneização da experiência auditiva pode ter impactos significativos no desenvolvimento de novas formas artísticas e na promoção de talentos emergentes, que têm menos oportunidades de serem descobertos em *playlists* automatizadas.

Portanto, a curadoria musical nas plataformas digitais assume uma função dupla: ao mesmo tempo que oferece aos utilizadores uma experiência de descoberta mais personalizada, também pode limitar a diversidade ao favorecer músicas que já são populares ou que seguem determinados padrões de sucesso. Isso gera um ciclo que prolonga os mesmos estilos musicais, restringindo o acesso a novas experiências sonoras. Para países como Portugal, onde a promoção da música local é crucial para manter a identidade cultural, o papel da curadoria torna-se ainda mais relevante, pois pode determinar o sucesso ou a invisibilidade de novos artistas.

2.4. Inteligência Artificial: evolução e potencial futuro preditivo na música

A Inteligência Artificial tem emergido como uma das tecnologias mais revolucionárias do século XXI, alterando profundamente diversos aspetos da nossa sociedade.

Esta secção explora a evolução da Inteligência Artificial e examina o potencial futuro preditivo desta tecnologia, abordando como as previsões baseadas em IA podem influenciar e modelar as tendências emergentes no setor musical.

2.4.1. Evolução da Inteligência Artificial

A Inteligência Artificial (IA) tem evoluído de forma rápida e demonstra um elevado nível de complexidade. Segundo Lu (2023), a sua evolução pode ser segmentada em várias fases distintas, refletindo os avanços tecnológicos e as aplicações práticas ao longo do tempo.

Nos seus primeiros anos, a IA era predominantemente baseada em regras e algoritmos simples, limitados pela tecnologia disponível. Durante esse período, os sistemas de IA eram capazes de realizar tarefas específicas, mas com uma flexibilidade reduzida. No entanto, têm surgido modelos cada vez mais sofisticados e versáteis.

A partir da década de 2000, observou-se uma transição significativa para modelos baseados em redes neurais profundas. De acordo com Lu (2023), "os modelos baseados em redes neurais profundas têm demonstrado avanços notáveis na precisão e na eficiência". Estes modelos são capazes de aprender representações complexas de dados, que resolvem problemas anteriormente considerados como impossíveis.

No que toca às aplicações práticas, a Inteligência Artificial tem provocado um impacto transformador em várias áreas, incluindo saúde, finanças e transportes, por exemplo, na saúde, algoritmos de IA têm sido utilizados para diagnóstico assistido por computador, proporcionando maior

precisão e rapidez na análise de imagens médicas. No setor financeiro, a IA tem sido aplicada a algoritmos de negociação, otimizando estratégias de investimento e gestão de riscos.

Ainda neste setor, surge o conceito de IA explicável como uma **tendência futura**, que se refere ao desenvolvimento de sistemas de Inteligência Artificial que, além de fornecer previsões ou decisões, oferecem explicações claras e compreensíveis sobre o raciocínio subjacente. Lu (2023) afirma que "a busca por modelos de IA mais explicáveis e éticos será essencial para a adoção generalizada da tecnologia". Esta tendência reflete a necessidade crescente de garantir que os sistemas de IA sejam não apenas eficientes, mas também transparentes e responsáveis.

2.4.2. O Futuro da Música: Aplicações Preditivas da Inteligência Artificial

Existem diversos estudos que exploram previsões no mercado musical, especialmente na antecipação da popularidade de músicas e álbuns.

Segundo Araújo (2019), os métodos de *machine learning* tendem a apresentar resultados superiores. No seu estudo, o autor propõe uma metodologia para prever a popularidade de músicas utilizando dados obtidos das tabelas de popularidade do Spotify. O principal objetivo deste trabalho foi desenvolver modelos de classificação capazes de antecipar, com dois meses de antecedência, se uma música teria sucesso nas plataformas de *streaming*.

Para alcançar este objetivo, foram testados vários classificadores, destacando-se o *SVM* (*Support Vector Machine*) com *kernel RBF* (*Radial Basis Function*), que apresentou o melhor desempenho, com métricas de *AUC* (*Area Under the Curve*), precisão, valor preditivo negativo (*NPV*) e especificidade superiores a 80%. Uma das conclusões mais significativas do estudo foi a verificação de que a inclusão de características acústicas no modelo aumentou o desempenho em apenas 5,23%, um valor que pode não ser estatisticamente relevante. Esta descoberta sugere que as informações acústicas das músicas podem não ser essenciais para a previsão da sua popularidade, embora Araújo (2019) destaque a necessidade de mais estudos para validar esta hipótese.

Contudo, os estudos preditivos na música não se limitam apenas à previsão da popularidade de uma canção. Sharma (2022) vai mais longe ao tentar prever a posição das músicas nas tabelas de popularidade, bem como a duração, em semanas, que essas músicas permanecem nas tabelas, utilizando para isso características objetivas e bem definidas extraídas da API do Spotify. Este trabalho amplia a investigação existente sobre a classificação de músicas em grupos como o Top 10 e o Top 40, demonstrando que é possível prever as posições exatas das músicas com um erro quadrático médio (RMSE) de 28 posições, e estimar o número de semanas que uma música permanecerá nas tabelas com um RMSE de 7 semanas. O estudo revela ainda a existência de correlações claras entre características individuais das músicas e as suas posições nas tabelas de popularidade, evidenciando que metadados objetivos das músicas são bons preditores do seu desempenho nas tabelas da *Billboard*. O modelo desenvolvido por Sharma (2022) permite, assim, prever o sucesso de uma música com uma margem de

erro aceitável, reforçando o valor preditivo de dados objetivos para identificar tendências no desempenho musical.

Outro exemplo de *machine learning* que obteve resultados extraordinários em aplicações preditivas na música é o estudo de Gulmatico (2022), intitulado *SpotiPred: A Machine Learning Approach Prediction of Spotify Music Popularity by Audio Features*. Nesta pesquisa, propõe-se um modelo para prever a popularidade de músicas no Spotify com base em características áudio. Um dos modelos analisados foi o *Random Forest*, que agrega centenas de árvores de decisão, cada uma treinada em diferentes subconjuntos de características áudio e *clusters* de dados de treino. Este modelo alcançou uma impressionante taxa de precisão de 95,37%.

Desta forma, os avanços nas aplicações preditivas da inteligência artificial no mercado musical, como evidenciado pelos estudos de Araújo (2019), Sharma (2022) e Gulmatico (2022), demonstram o potencial da análise de dados e do *machine learning* para antecipar a popularidade e o desempenho das músicas, proporcionando assim novas ferramentas para artistas e profissionais da indústria musical.

Capítulo 3

Compreensão dos Dados

A pesquisa em questão será baseada numa análise quantitativa e qualitativa, uma vez que envolve o uso de dados numéricos e estatísticas para identificar padrões, tendências e relações, assim como a interpretação e compreensão de dados não numéricos.

3.1. Base de Dados

A análise comparativa entre Portugal e França foi baseada em dados recolhidos através do site **spotifycharts.com**, onde foram obtidos ficheiros *csv* com informações sobre as músicas mais populares de cada país ao longo de 7 anos (2017-2023).

Através de um programa desenvolvido em *python*, foram recolhidos dados semanais de cada mês, totalizando **67.179 músicas** para cada país. Deste conjunto, Portugal conta com **5.133 faixas únicas**, enquanto França possui **8.060 faixas únicas**. Com base nesta informação, já é possível retirar que França demonstra uma maior diversidade musical nas *charts*.

3.1.1. Construção da Base de Dados - 1ª fase

Tabela 1- Base de dados de Portugal após recolha de dados

rank	uri	artist_names	track_name	date
1	spotify:track:7qiZfU4dY1l	Ed Sheeran	Shape of You	01/01/2017
2	spotify:track:5aAx2yezTd8	The Weeknd, Daft Punk	Starboy	01/01/2017
3	spotify:track:6PCUP3dWrr	Ed Sheeran	Castle on the Hill	01/01/2017
4	spotify:track:5knuzwU65g	Clean Bandit, Sean Paul, Anne-Marie	Rockabye (feat. Sean Paul & Anne-Marie)	01/01/2017
5	spotify:track:5uCax9HTNI	James Arthur	Say You Won't Let Go	01/01/2017
6	spotify:track:5GXAXm5YO	The Weeknd, Daft Punk	I Feel It Coming	01/01/2017
7	spotify:track:3NdDpSvN91ZAYN	Taylor Swift	I Don't Wanna Live Forever (Fifty Shades Darker)	01/01/2017
8	spotify:track:6mlCuAdrwe	Shakira, Maluma	Chantaje (feat. Maluma)	01/01/2017
9	spotify:track:7BKLCZ1jbUf	The Chainsmokers, Halsey	Closer	01/01/2017
10	spotify:track:4TZy1wLyHec	Shawn Mendes	Mercy	01/01/2017

A Tabela 1 apresenta um excerto da base de dados de Portugal, obtido a partir da recolha de dados no *website* spotifycharts.com, exibindo as dez primeiras posições correspondentes à primeira semana de janeiro de 2017. Foram extraídos deste *website* o **ranking** das músicas (de 1 a 200) para cada semana, o **URI** (identificador único de cada música), o **nome do artista** e o **título da música**. A data correspondente a cada semana foi adicionada posteriormente, sendo estruturada no formato 'nº da semana/mês/ano', em que o número da semana varia de 1 a 4, mudando a cada sexta-feira.

3.1.2. Construção da Base de Dados - 2ª fase

Com o auxílio da **API do Spotify**, (ver Figura 1) foi possível extrair diversas características de cada música e artista utilizando o **Spotify ID** de cada faixa. Para isso, foi realizada uma requisição ao

endpoint `/audio-features/{id}` utilizando o método *GET*, onde `{id}` corresponde ao identificador único da música na plataforma.

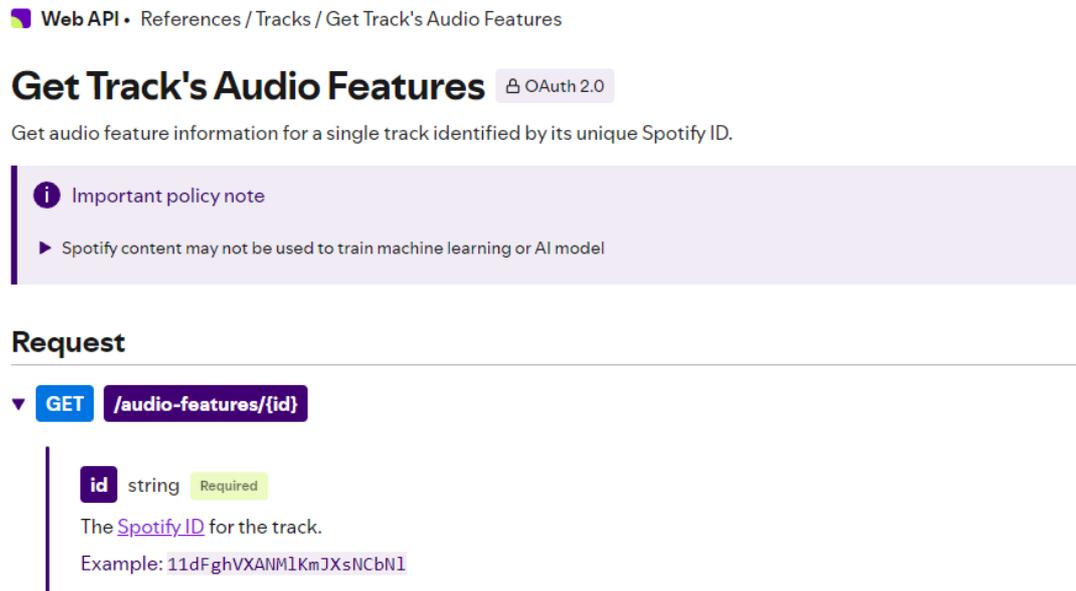


Figura 1 - Spotify API¹

Após a extração de características de cada música, foram adicionadas à base de dados **13** novas colunas: *acousticness*, *danceability*, *duration_ms*, *energy*, *instrumentalness*, *key*, *liveness*, *loudness*, *mode*, *speechiness*, *tempo*, *time_signature* e *valence*. Cada uma destas características tem um valor (*int* ou *float*) associado, de acordo com a música analisada. Na Tabela 2 é possível ver todas as características estudadas e a respetiva descrição.

Tabela 2 - Descrição das características

Características	Descrição
<i>Duration_ms</i>	Duração das músicas formato de milissegundos.
<i>Danceability</i>	Descreve o quão adequada é uma faixa para dançar, com base em uma combinação de elementos musicais, incluindo tempo, estabilidade do ritmo, força da batida e regularidade geral. Um valor de 0,0 é o menos em dançável, e 1,0 é o mais dançável.
<i>Energy</i>	Uma medida de 0,0 a 1,0 que representa uma medida percetual de intensidade e atividade. Tipicamente, faixas energéticas parecem rápidas, altas e barulhentas. Por exemplo, o <i>death metal</i> tem alta energia, enquanto um prelúdio de <i>Bach</i> pontua baixo na escala. Características percetuais que contribuem para este atributo incluem alcance dinâmico, volume percebido, timbre, taxa de início e entropia geral.

¹ Spotify. (n.d.). Get track's audio features. In Spotify for Developers. Retrieved February 20, 2024, from <https://developer.spotify.com/documentation/web-api/reference/get-audio-features>

<i>Key</i>	A tonalidade em que a faixa está. Inteiros mapeiam-se para notas usando a notação padrão da Classe de <i>Pitches</i> . Por exemplo, 0 = Dó (C), 1 = Dó#/Réb (C#/Db), 2 = Ré (D) e assim por diante. Se nenhuma tonalidade foi detetada, o valor é -1.
<i>Loudness</i>	O volume geral de uma faixa em decibéis (dB). Os valores de volume são calculados como média ao longo de toda a faixa e são úteis para comparar o volume relativo das faixas. Volume é a qualidade de um som que é o principal correlato psicológico da força física (amplitude). Os valores normalmente variam entre -60 e 0 dB.
<i>Mode</i>	Indica a modalidade (maior ou menor) de uma faixa, o tipo de escala do qual seu conteúdo melódico é derivado. Maior é representado por 1 e menor por 0.
<i>Speechiness</i>	Deteta a presença de palavras faladas em uma faixa. Quanto mais exclusivamente semelhante à fala for a gravação (por exemplo, <i>talk show</i> , audiolivro, poesia), mais próximo de 1,0 será o valor do atributo. Valores acima de 0,66 descrevem faixas que provavelmente são feitas inteiramente de palavras faladas. Valores entre 0,33 e 0,66 descrevem faixas que podem conter tanto música quanto fala, seja em seções ou em camadas, incluindo casos como música rap. Valores abaixo de 0,33 provavelmente representam músicas e outras faixas que não se assemelham à fala.
<i>Acousticness</i>	Uma medida de confiança de 0,0 a 1,0 sobre se a faixa é acústica. Um valor de 1,0 representa alta confiança de que a faixa é acústica.
<i>Instrumentalness</i>	Prediz se uma faixa contém ou não vocais. Sons como "Ooh" e "aah" são tratados como instrumentais neste contexto. Faixas de rap ou faladas são claramente "vocais". Quanto mais próximo o valor de <i>instrumentalness</i> for de 1,0, maior a probabilidade de que a faixa não contenha conteúdo vocal. Valores acima de 0,5 destinam-se a representar faixas instrumentais, mas a confiança é maior à medida que o valor se aproxima de 1,0.
<i>Liveness</i>	Deteta a presença de uma audiência na gravação. Valores mais altos de "liveness" representam uma probabilidade aumentada de que a faixa foi executada ao vivo. Um valor acima de 0,8 fornece uma forte probabilidade de que a faixa é ao vivo.
<i>Valence</i>	Positividade musical transferida por faixa. Valores mais altos sinalizam positivos, enquanto valores mais baixos sinalizam negativos.
Tempo	O tempo geral estimado de uma faixa em batidas por minuto (BPM). Em terminologia musical, tempo é a velocidade ou o ritmo de uma peça específica e deriva diretamente da duração média das batidas.
<i>Time_Signature</i>	Uma assinatura de tempo estimada. A assinatura de tempo (compasso) é uma norma musical para especificar quantas batidas existem em cada compasso (ou medida). A assinatura de tempo varia de 3 a 7, indicando assinaturas de tempo de "3/4" a "7/4".

Tabela 3 - Base de dados após a extração de características

track_name	acousticness	danceability	duration	energy	instrumentalness	key
Shape of You	0.581	0.825	233713	0.652	0.0	1
Starboy	0.165	0.681	230453	0.594	3.49e-06	7
Castle on the Hill	0.0232	0.461	261154	0.834	1.14e-05	2
Rockabye (feat. Sean Paul & Anne-Marie)	0.406	0.72	251088	0.763	0.0	9
Say You Won't Let Go	0.695	0.358	211467	0.557	0.0	10
I Feel It Coming	0.426	0.773	269187	0.819	0.0	0
I Don't Wanna Live Forever (Fifty Shades Darker)	0.0631	0.735	245200	0.451	1.3e-05	0
Chantaje (feat. Maluma)	0.187	0.852	195840	0.773	3.05e-05	8
Closer	0.414	0.748	244960	0.524	0.0	8
Mercy	0.137	0.469	208733	0.68	0.0	11

track_name	liveness	loudness	mode	speechiness	tempo	time_signature	valence
Shape of You	0.0931	-3.183	0	0.0802	95.977	4	0.931
Starboy	0.134	-7.028	1	0.282	186.054	4	0.535
Castle on the Hill	0.14	-4.868	1	0.0989	135.007	4	0.471
Rockabye (feat. Sean Paul & Anne-Marie)	0.18	-4.068	0	0.0523	101.965	4	0.742
Say You Won't Let Go	0.0902	-7.398	1	0.059	85.043	4	0.494
I Feel It Coming	0.0679	-5.946	0	0.118	92.99	4	0.585
I Don't Wanna Live Forever (Fifty Shades Darker)	0.325	-8.374	1	0.0585	117.973	4	0.0862
Chantaje (feat. Maluma)	0.159	-2.921	0	0.0776	102.034	4	0.907
Closer	0.111	-5.599	1	0.0338	95.01	4	0.661
Mercy	0.11	-4.921	0	0.117	147.734	4	0.374

Por exemplo, na Tabela 3, a faixa mundialmente conhecida '**Shape of You**', do artista **Ed Sheeran**, vibrante e contagiante, frequentemente tocada em discotecas e festas, apresenta um índice elevado de *danceability* (**0,825**), *energy* (**0,652**) e um *valence* particularmente alto (**0,931**), refletindo o seu caráter alegre e enérgico. Em contraste, a balada romântica e suave '**Say You Won't Let Go**', de **James Arthur**, transmite uma sensação mais introspetiva e melancólica, com valores significativamente mais baixos de *danceability* (**0,358**), *energy* (**0,557**) e *valence* (**0,494**), o que condiz com o seu tom mais sereno e emocional.

3.1.3. Construção da Base de Dados - 3ª fase

Um dos focos principais desta pesquisa foi investigar a presença da língua nacional nas músicas populares de cada país.

Através da **API do Genius** (ver Figura 2) uma plataforma conhecida pela sua vasta base de dados de letras de músicas e informações sobre artistas, foi retirada a letra de cada música presente na base de dados.

Resources

Annotations

An **annotation** is a piece of content about a part of a document. The document may be a *song* (hosted on Genius) or a *web page* (hosted anywhere). The part of a document that an annotation is attached to is called a *referent*.

Annotation data returned from the API includes both the substance of the annotation and the necessary information for displaying it in its original context.

GET /annotations/:id

Data for a specific annotation.

id ID of the annotation
10225840

text_format Format for text bodies related to the document. One or more of `dom`, `plain`, and `html`, separated by commas (defaults to `dom`). See details of each option [here](#)
Text Format

POST /annotations

Figura 2 - Genius API

Utilizando técnicas de Processamento de Linguagem Natural (NLP), o idioma de cada faixa foi identificado (ver coluna “*language*” na Tabela 4), com esta informação, torna-se possível realizar uma análise mais aprofundada sobre a prevalência da língua nacional nas faixas mais ouvidas no país.

Tabela 4 - Base de dados após a detecção do idioma

artist_names	track_name	language
Ed Sheeran	Shape of You	en
The Weeknd, Daft Punk	Starboy	en
Ed Sheeran	Castle on the Hill	en
Clean Bandit, Sean Paul, Anne-Marie	Rockabye (feat. Sean Paul & Anne-Marie)	en
James Arthur	Say You Won't Let Go	en
The Weeknd, Daft Punk	I Feel It Coming	en
ZAYN, Taylor Swift	I Don't Wanna Live Forever (Fifty Shades Darker)	en
Shakira, Maluma	Chantaje (feat. Maluma)	es
The Chainsmokers, Halsey	Closer	en
Shawn Mendes	Mercy	en

Para oferecer uma visão geral sobre a construção da base de dados, a Figura 3 apresenta um fluxograma que simplifica todo o processo descrito anteriormente. A base de dados francesa foi desenvolvida seguindo os mesmos procedimentos.

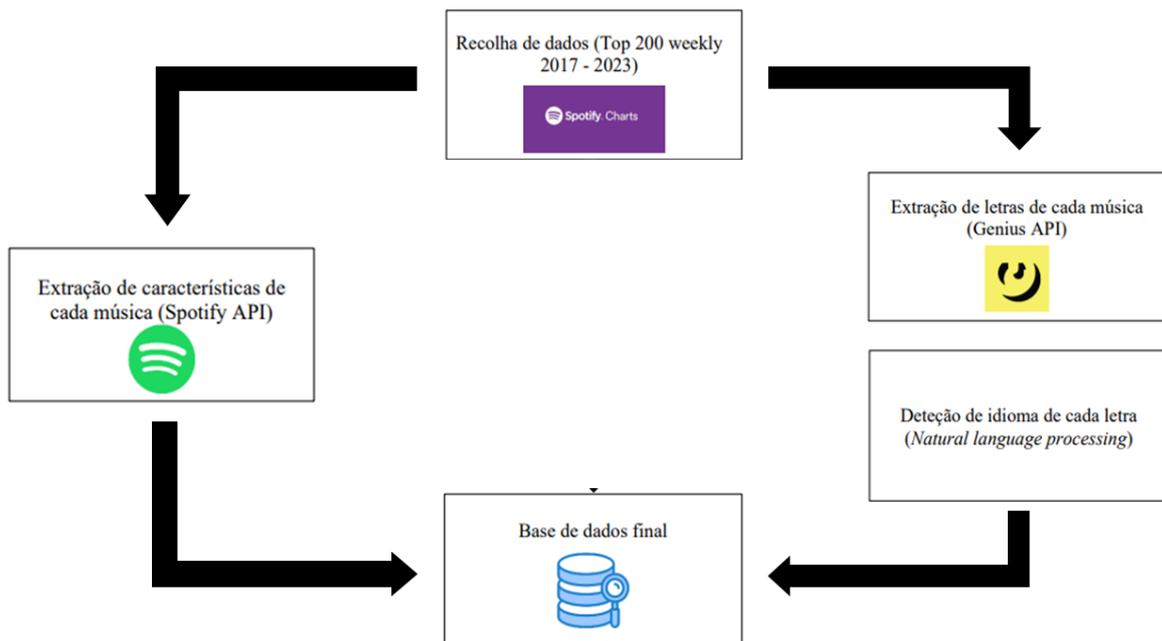


Figura 3 - Flow chart da Construção da Base de dados

3.2. Distribuições de frequência em relação às características de áudio

Os gráficos de distribuições de frequência, ao serem comparados entre dois países, como Portugal e França, oferecem uma perspetiva clara sobre as diferenças e semelhanças nas características de áudio das músicas que predominam no Top 200 semanal.

Nesta secção, serão analisadas as características mais pertinentes para a pesquisa, com o objetivo de elucidar as preferências musicais de cada país.

Nas figuras seguintes, o eixo x representa os **valores da característica em análise**, neste caso, avaliados numa escala de 0 a 1, enquanto o eixo y (“*count*”) apresenta a **densidade normalizada**. Este método de normalização permite comparar diretamente as distribuições entre os dois países, independentemente do número total de músicas em cada base de dados, garantindo que a área total sob cada histograma seja igual a 1.

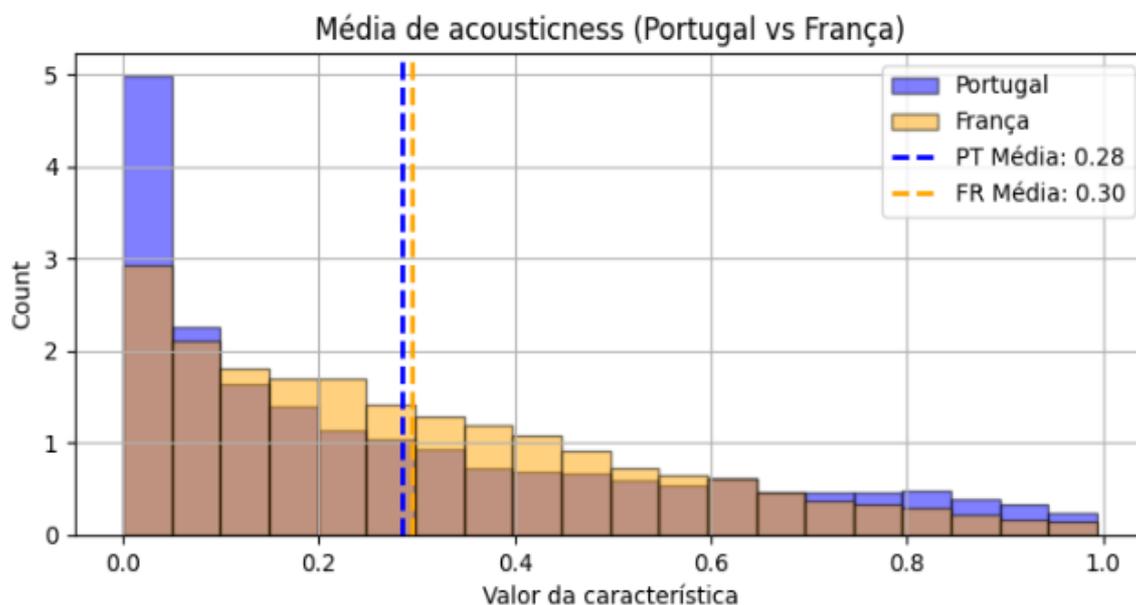


Figura 4 - Média de acousticness em Portugal e França

Na Figura 4, é possível observar que ambos os países apresentam uma tendência assimétrica à direita (*right-skewed*) para a característica de *acousticness*, o que sugere que a maioria das músicas tem um nível relativamente baixo de sonoridade acústica, com algumas exceções que apresentam valores mais altos. Contudo, Portugal demonstra, em média, um valor inferior (0,28) em comparação com França (0,30) ao longo do período de 2017 a 2023. Essa diferença média indica que as músicas em Portugal tendem a ser ligeiramente menos acústicas do que as em França, o que pode refletir as preferências musicais e tendências culturais distintas de cada país. Além disso, essa característica pode estar associada a diferentes estilos musicais predominantes em cada mercado, sugerindo uma inclinação maior para produções mais eletrônicas ou pop em Portugal. No entanto, a conclusão referente à característica de áudio '*acousticness*' será aprofundada no capítulo 5, onde analisaremos esta variável ao longo do tempo, proporcionando uma nova perspectiva sobre o seu impacto no consumo de música em Portugal.

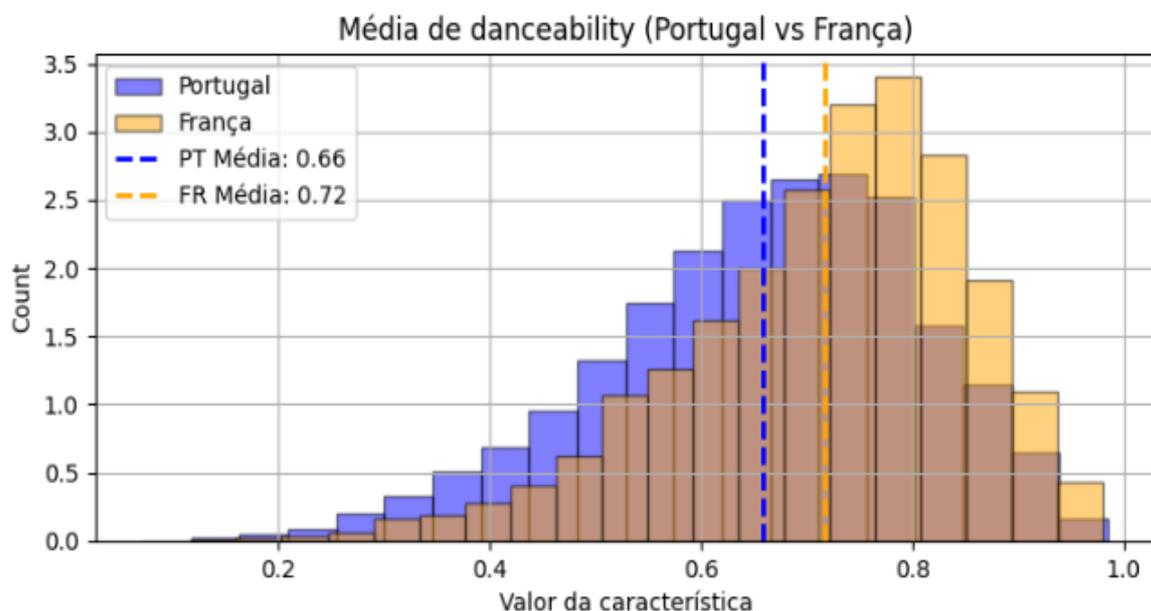


Figura 5 - Média de danceability em Portugal e França

A característica de *danceability* (ver Figura 5) nos dois países revela uma forte tendência de assimetria negativa (*left-skewed*), o que sugere que tanto os portugueses quanto os franceses demonstram uma clara preferência por músicas dançáveis, caracterizadas por ritmos cativantes e envolventes. Esta inclinação para a '*danceability*' pode ser atribuída a fatores culturais, sociais e mesmo a contextos festivos, onde a música desempenha um papel central nas interações e celebrações. Embora a diferença entre as médias dos dois países não seja significativa, com França a liderar com uma média de 0,72, essa leve superioridade pode refletir uma maior diversidade de estilos musicais que enfatizam a dança, além de uma cena musical mais robusta. Este fenómeno levanta questões interessantes sobre como a '*danceability*' pode influenciar o consumo de música e a experiência musical nos dois contextos culturais.

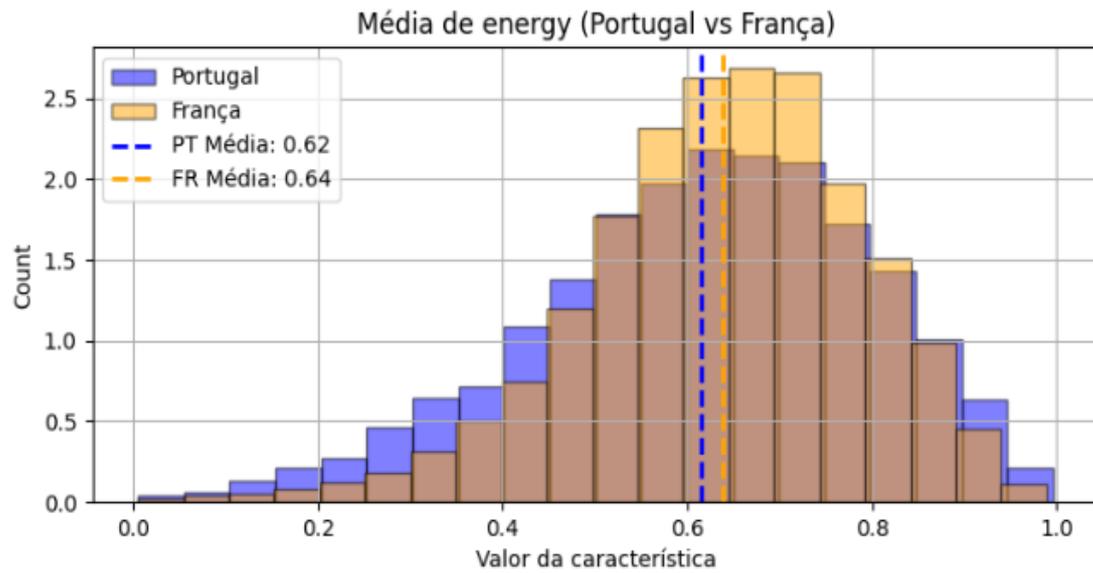


Figura 6 - Média de energy em Portugal e França

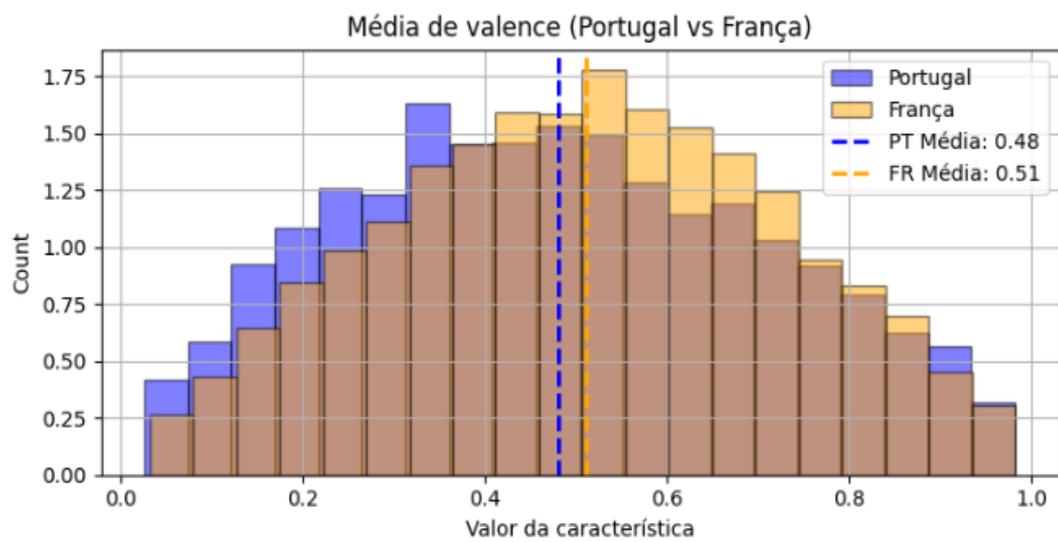


Figura 7 - Média de valence em Portugal e França

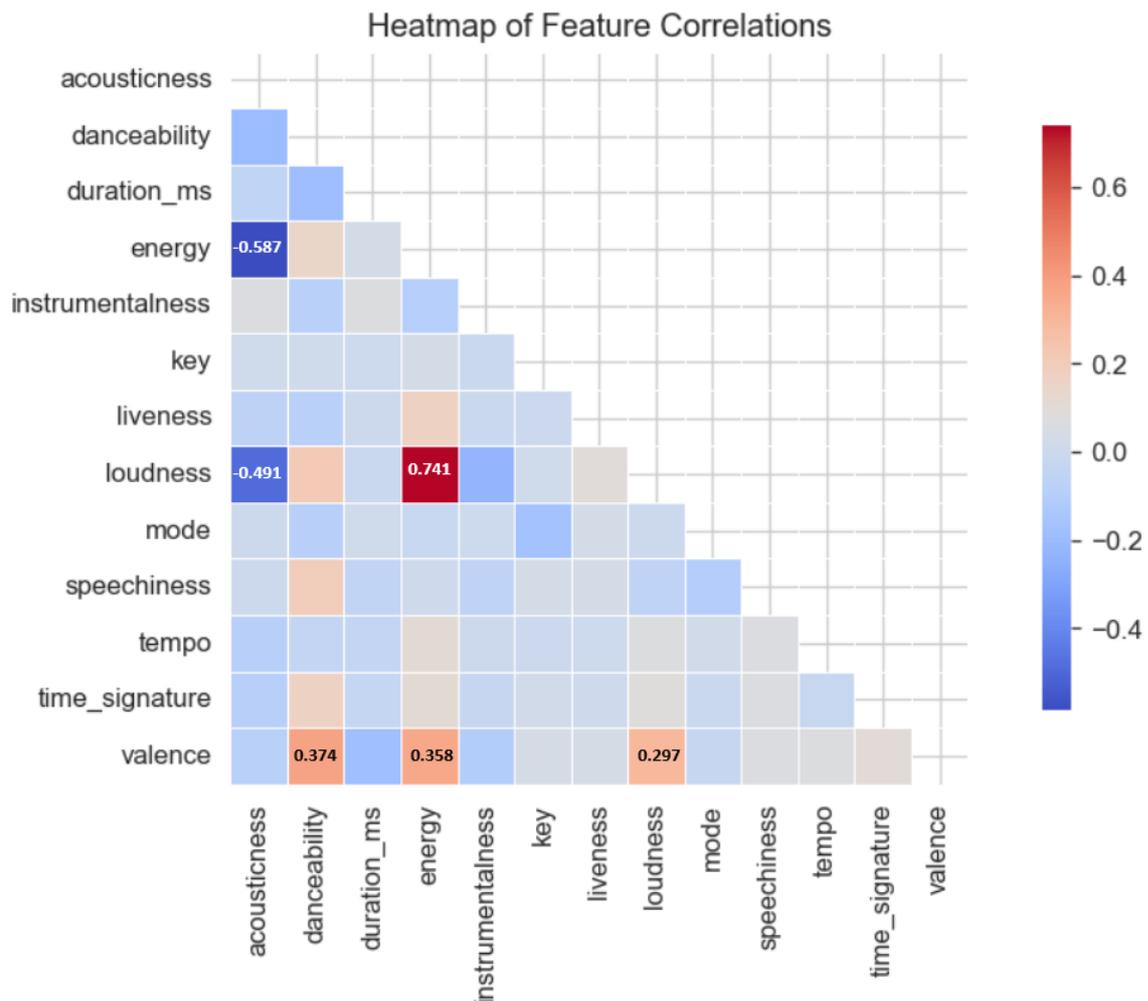


Figura 8 - Correlação entre características de áudio

As categorias *energy* e *valence* (ver Figura 6 e Figura 7) revelam ser proporcionais à categoria anteriormente analisada, *danceability*.

Conforme evidenciado no *heatmap* apresentado na Figura 8, as categorias analisadas exibem uma correlação mais robusta em comparação com as demais. Essa inter-relação entre as três categorias é significativa, pois, para que uma música seja considerada dançável, é provável que apresente níveis elevados de energia e uma tonalidade positiva. Adicionalmente, a Figura 8 revela uma forte relação entre várias outras variáveis, como *energy* e *loudness*: músicas com valores elevados de *loudness* tendem a apresentar também um alto nível de energia. Em contrapartida, observa-se uma relação inversa entre *acousticness* e *loudness*, assim como entre *acousticness* e *energy*; isto é, à medida que os valores de *acousticness* aumentam, os níveis de *energy* e *loudness* tendem a diminuir.

Capítulo 4

Modelação

A análise de séries temporais evoluiu significativamente desde a sua origem, sendo alvo de inúmeras pesquisas para atender a diferentes objetivos.

A previsão de séries temporais baseia-se na ideia de que padrões passados podem se repetir no futuro, permitindo que a história dos dados seja utilizada para prever eventos futuros. Nesta abordagem, dois elementos-chave desempenham um papel fundamental: **sazonalidade** (*seasonality*) e a **tendência** (*trend*). A sazonalidade refere-se a padrões que se repetem em intervalos regulares, como aumentos de consumo em determinados períodos do ano. Já a tendência, o fator que me levou a estudar esta questão, refere-se à direção geral que os dados seguem ao longo do tempo, seja de crescimento ou declínio. Ambos os componentes, tendência e sazonalidade, têm um impacto direto na previsão de séries temporais. Ao modelar esses fatores corretamente, os algoritmos conseguem fazer previsões mais robustas, capturando tanto a direção de longo prazo quanto os ciclos regulares presentes nos dados.

Com o objetivo de melhorar os resultados desta fase da análise, decidi utilizar uma base de dados diária em vez de semanal, permitindo aos modelos captar padrões mais detalhados e variações temporais que uma base semanal poderia não refletir com precisão.

Nesta **nova base de dados**, são analisadas duas variáveis principais: a data e o total de músicas portuguesas no **top 200 em cada dia**. É com base nesses dados que os modelos preditivos serão aplicados, buscando estimar com maior precisão a dinâmica da presença de músicas portuguesas no top diário

4.1 Métodos de Previsão Temporal

Na análise de séries temporais, a escolha do método de previsão adequado é essencial para garantir a precisão das estimativas futuras. Para esta pesquisa, foram aplicados quatro métodos distintos: **Prophet**, **NeuralProphet**, **XGBoost** e **LSTM**, cada um com características e vantagens específicas.

4.1.1 Prophet

O *Prophet* é um modelo de previsão de séries temporais desenvolvido pelo Facebook, projetado para lidar com dados que apresentam sazonalidade e tendências de forma intuitiva e acessível. “Na sua essência, o *Prophet* baseia-se num modelo de regressão aditiva” (Kumar Jha & Pande, 2021, p.547), que permite a incorporação de diferentes componentes para capturar a dinâmica temporal dos dados.

Este modelo é estruturado em torno de várias características importantes, como “curva modular regressiva ou linear para a tendência de crescimento” (Kumar Jha & Pande, 2021, p.547), que possibilita a modelagem de tendências crescentes ou decrescentes ao longo do tempo. Além disso, o *Prophet* utiliza uma “componente sazonal baseada na série de Fourier” (idem), que permite a identificação de padrões

sazonais complexos em diversas escalas. Para a sazonalidade semanal, o modelo inclui “uma componente sazonal semanal” (idem), tornando-o especialmente útil para dados que apresentam comportamentos recorrentes em ciclos semanais.

De acordo com Kumar Jha & Pande (2021, p.549), o modelo de previsão do *FB Prophet* é baseado num modelo aditivo regressivo que pode ser formulado como:

$$y(t) = g(t) + h(t) + s(t) + et$$

Equação 1 - Modelo de regressão aditiva

Na Equação 1, $y(t)$ representa o modelo de regressão aditiva; $g(t)$ é o fator de tendência; $h(t)$ é o componente de feriado; $s(t)$ é o componente de sazonalidade, e et o termo de erro. O fator de tendência $g(t)$ pode ser modelado de duas maneiras:

a) Modelo de crescimento logístico – Este modelo representa o crescimento em várias etapas. Na primeira etapa, o crescimento é observado de forma aproximadamente exponencial, e, após isso, alcança-se uma fase de saturação, a partir da qual o crescimento se torna linear. O modelo pode ser formulado na equação abaixo.

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$

Equação 2 - Modelo de crescimento logístico

Neste modelo matemático (ver Equação 2), L representa o valor máximo da curva do modelo; k é a taxa de crescimento; x_0 é o valor x no ponto sigmoide.

b) Modelo Linear por Partes - Este é uma versão modificada do modelo linear, onde diferentes intervalos de x têm diferentes relações lineares. O modelo pode ser formulado conforme mostrado na equação abaixo (ver Equação 3)

$$y = \beta_0 + \beta_1 + \beta_2(x - c)^+ + \varepsilon$$

Equação 3 - Modelo Linear por Partes

Na Figura 9 é possível observar o *Prophet workflow*

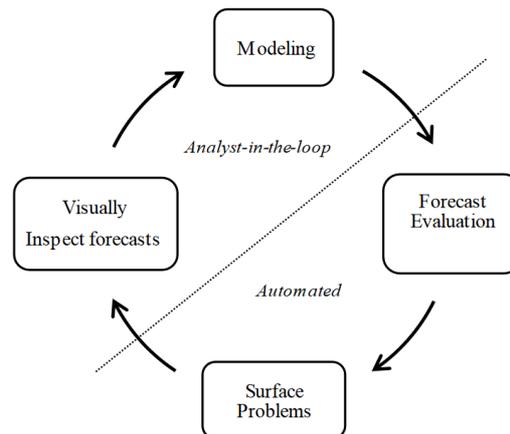


Figura 9 - *Prophet workflow*²

Para aplicar o modelo *Prophet* na base de dados diária, foi necessário inicializá-lo com a função **m = Prophet(seasonality_mode='multiplicative')**. Este parâmetro permite que a sazonalidade seja modelada de forma multiplicativa, o que é particularmente adequado quando se espera que o impacto da sazonalidade varie em função do nível da série temporal.

Após a inicialização, o modelo é ajustado (**fit**) utilizando os dados de treino. Em seguida, as previsões são geradas com base nas datas do conjunto de teste, utilizando a função **predict**.

4.1.2 Neural Prophet

NeuralProphet é uma ferramenta de previsão que combina a simplicidade do *Prophet* com o poder de técnicas de *deep learning*. Segundo Triebe (2021, p.1) “o contexto local é introduzido com módulos de auto-regressão e co-variados, que podem ser configurados como regressão linear clássica ou como redes neurais. Fora isso, o *NeuralProphet* mantém a filosofia de design do *Prophet* e oferece os mesmos componentes básicos do modelo”. Este modelo acaba por ser aplicado com as mesmas funções que o anterior.

4.1.3 LSTM

Adicionalmente aos modelos anteriormente apresentados, o modelo LSTM (*Long Short-Term Memory*), introduzido em 1997 por Hochreiter e Schmidhuber, utiliza uma variação das redes recorrentes (RNNs), desenvolvidas para lidar com problemas como o desaparecimento do gradiente, comum nas RNNs tradicionais.

Conforme explicado por Patterson & Gibson (2017), "o conteúdo da célula de memória é modulado pelas *input gates* e *forget gates*" (p.150) e, quando ambas estão fechadas, "o conteúdo da célula de memória permanece inalterado entre um passo temporal e o seguinte" (p.150). Esta arquitetura

² Kumar Jha & Pande (2021, p.547)

permite que as redes LSTM retenham informação ao longo de muitos passos temporais, permitindo também que os gradientes fluam ao longo de várias interações, evitando assim o problema da *vanishing gradient*.

Para implementar este modelo foi necessário criar características de *lag* (ver Tabela 5) a partir dos dados de entrada, permitindo que o modelo utilize informação anterior para obter previsões.

Tabela 5 - Características lag

date	count	count(t-1)	count(t-2)	count(t-3)	count(t-4)
2017-01-05	20	20.0	22.0	22.0	23.0
2017-01-06	31	20.0	20.0	22.0	22.0
2017-01-07	33	31.0	20.0	20.0	22.0
2017-01-08	32	33.0	31.0	20.0	20.0
2017-01-09	35	32.0	33.0	31.0	20.0
...
2023-12-27	112	113.0	79.0	57.0	91.0
2023-12-28	112	112.0	113.0	79.0	57.0
2023-12-29	120	112.0	112.0	113.0	79.0
2023-12-30	121	120.0	112.0	112.0	113.0
2023-12-31	117	121.0	120.0	112.0	112.0

Após isso, os dados são normalizados usando a função **MinMaxScaler**, garantindo que estejam na mesma escala. A fase de treinamento e validação são realizados em *loops*, calculando a perda com a função **MSE (Mean Squared Error)** e ajustando os parâmetros da rede.

4.1.4 XGBoost

O *XGBoost (Extreme Gradient Boosting)* é uma técnica de *machine learning* amplamente utilizada em problemas de classificação e regressão, conhecida pela sua eficiência e precisão. Embora originalmente não tenha sido desenvolvido especificamente para a previsão de séries temporais, o *XGBoost* pode ser adaptado para essa tarefa por meio de uma abordagem conhecida como *supervised learning*, onde a sequência temporal é transformada em variáveis preditoras.

De acordo com Dairu & Shilong (2021, p.480), “em comparação com outros métodos de aprendizado de conjunto, o *XGBoost* é dez vezes mais rápido e usa muito menos recursos”. Para além da rapidez, este método destaca-se pela sua escalabilidade e eficiência. Este modelo é iniciado com mil estimadores a partir da instrução:

```
model = XGBRegressor(objective="reg:squarederror", n_estimators = 1000)
```

após isto, a função **xgb_predict** treina o modelo e faz previsões para as datas específicas.

Com ajuda de uma função de validação, para cada instância no conjunto de teste, o modelo prevê o valor da variável alvo, calcula o erro quadrático médio (RMSE) e retorna os resultados.

Capítulo 5

Discussão de Resultados

Neste capítulo, são interpretados os dados apresentados na pesquisa, ligando-os às questões e hipóteses iniciais, de forma crítica. Desta forma, é dividido de acordo com as questões de pesquisa e respetivas hipóteses.

5.1 Evolução Histórica do Consumo de Música: Comparação entre Portugal e França

H1: O consumo de música portuguesa no Spotify em Portugal tem crescido de forma menos consistente ao longo dos anos em comparação com o consumo de música francesa.

Para sustentar a Hipótese 1, foi conduzido um estudo comparativo entre dois países, **Portugal e França**. Esta abordagem permite obter conclusões mais robustas e objetivas ao considerar contextos musicais distintos. Com base na análise das bases de dados, foram elaborados dois gráficos que ilustram a evolução das línguas predominantes no Top 200 semanal de cada país, apresentados numa escala mensal para melhor visualização (ver Figura 10 e Figura 11).

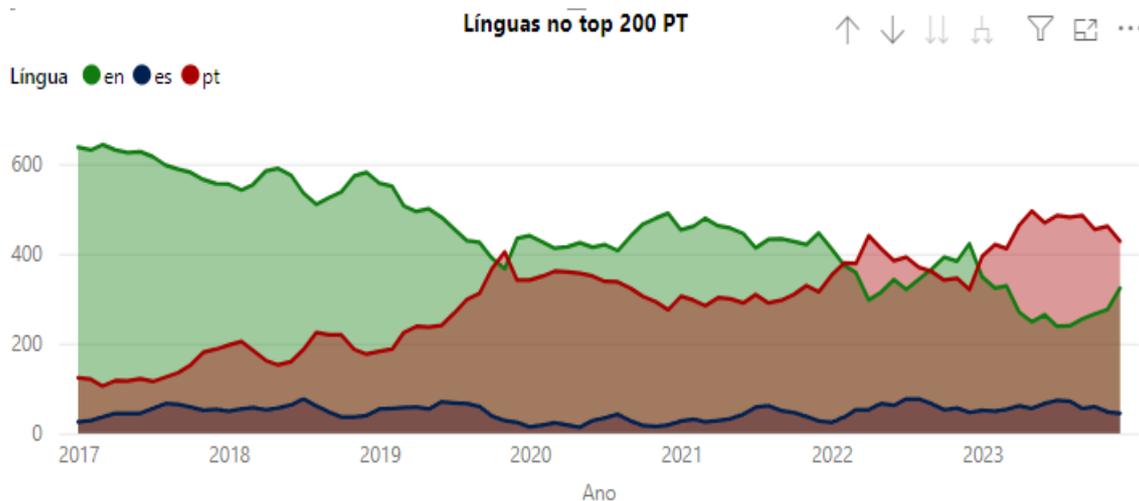


Figura 10 - Línguas no Top 200 português

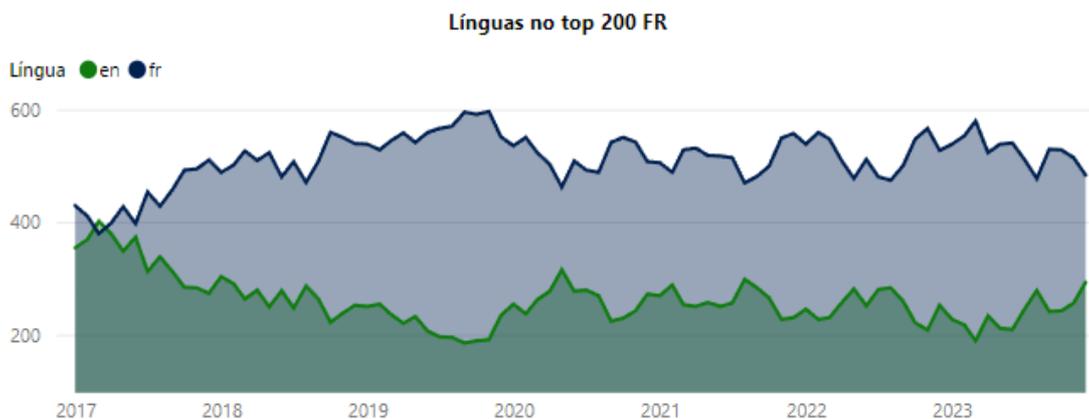


Figura 11 - Línguas no Top 200 francês

A Figura 10 ilustra a evolução das línguas das músicas presentes no Top 200 do Spotify em Portugal (PT) entre os anos de 2017 e 2023. De 2017 a 2019, a música em língua portuguesa (indicada pela cor vermelha) manteve uma participação relativamente modesta no Top 200, com o mercado sendo predominantemente dominado por músicas em inglês (representadas pela cor verde). Esse cenário sugere que, até 2019, a maioria dos consumidores de música em Portugal preferia canções em línguas estrangeiras, especialmente em inglês. No entanto, em 2019, observa-se um crescimento significativo da música em português, representado por 45 músicas em português no Top 200. Ao longo do ano, essa quantidade cresceu de forma notável e, em dezembro de 2019, a música em língua portuguesa superou, pela primeira vez, a quantidade de músicas em inglês, com 103 faixas no Top 200. Esse aumento representa um crescimento histórico de 29% no consumo de música em português. Esse crescimento pode ser atribuído a diversos fatores, como o surgimento de novos artistas locais, um fortalecimento do sentimento de identidade cultural (potencialmente impulsionado por eventos como a primeira edição dos Prêmios *PLAY*), e mudanças nas preferências musicais dos ouvintes portugueses. Entre 2019 e 2022, a música em português permaneceu abaixo da música em inglês em termos de popularidade neste *ranking*. No entanto, a partir de 2022, ocorre uma mudança expressiva: a presença da língua portuguesa começa a crescer de maneira consistente, sugerindo um aumento significativo no consumo de músicas em português. Em 2023, essa tendência continua culminando no pico de popularidade da música em português no Top 200, onde ela finalmente ultrapassa a quantidade de músicas em inglês, consolidando-se como a principal língua no consumo musical em Portugal.

A Figura 11 oferece uma visão sobre o mercado musical francês durante o mesmo período (2017 a 2023). Este gráfico evidencia uma notável estabilidade na presença de músicas em francês (representadas pela cor verde) no Top 200 do Spotify na França. Ao longo de todo o período, a música em francês manteve uma participação consistentemente alta, com algumas flutuações menores, mas sem mudanças significativas. Isso sugere que o consumo de músicas em língua francesa é sustentado por uma base de ouvintes sólida e fiel, que permaneceu estável ao longo dos anos. Desta forma, o crescimento observado no consumo de música em português em Portugal aponta para um fortalecimento da identidade cultural e musical do país, possivelmente impulsionado por políticas de promoção cultural, o surgimento de novos talentos locais, ou uma mudança nas plataformas de *streaming* que passou a favorecer conteúdos locais.

Por outro lado, a estabilidade do consumo de música em francês em França reflete uma continuidade na valorização da cultura e da língua francesas, mesmo num mercado globalizado onde o inglês costuma dominar as preferências musicais.

Portanto, em termos de evolução, o mercado português demonstrou uma dinâmica mais marcante e transformadora no consumo de música local, enquanto o mercado francês permaneceu mais estável e consistente na preservação e valorização da música em sua língua nativa.

5.2. Padrões e características recorrentes nas Músicas Populares em Portugal

H2: As músicas populares em Portugal tendem a partilhar características recorrentes, o que influencia diretamente a sua popularidade nas plataformas de streaming.

Para validar a hipótese H2, nesta secção, analisei as características mais relevantes para o estudo ao longo do tempo, abrangendo o período de 2017 a 2023. Esta análise temporal permite observar como as variáveis evoluíram ao longo dos anos, identificando padrões e tendências que podem refletir mudanças nas preferências musicais dos ouvintes. Além disso, ao examinar esses dados ao longo de um período prolongado, é possível entender melhor a dinâmica do consumo musical em Portugal e na França, assim como as influências culturais e sociais que podem ter impactado estas características ao longo do tempo.

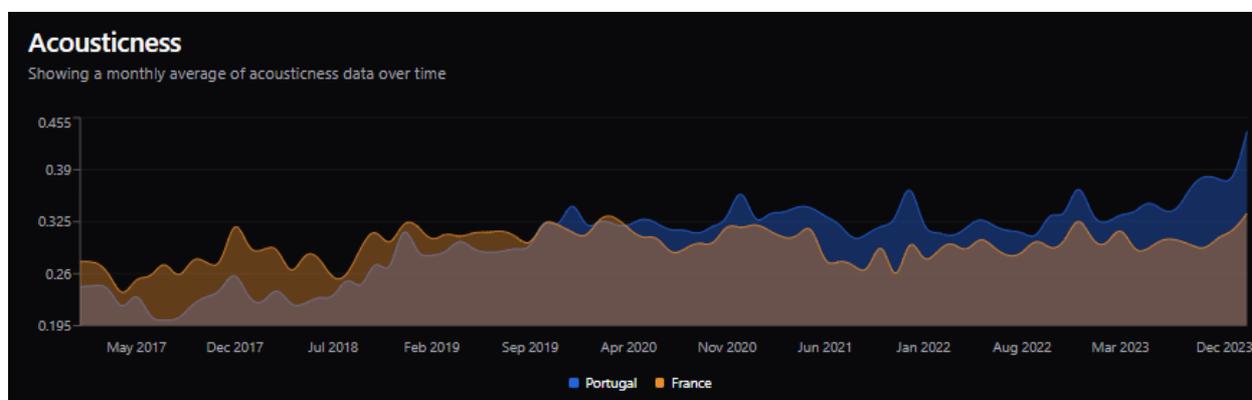


Figura 12 - Característica 'acousticness' ao longo do tempo

No capítulo 3, na secção 3.2, foi observado que a média geral da categoria *acousticness* apresentava uma tendência de assimetria positiva (*right-skewed*), indicando que os valores tendiam a ser relativamente baixos. No entanto, ao analisar essa variável com maior detalhe ao longo do tempo (ver Figura 12), verifica-se que, ao contrário de França, que manteve um nível de *acousticness* bastante estável, **Portugal** tem experimentado um crescimento gradual na presença de elementos acústicos nas músicas populares. Particularmente a partir de julho de 2018, observa-se um aumento considerável nesta variável, culminando em picos frequentes durante o mês de dezembro. Esses picos podem ser atribuídos a músicas específicas associadas à época festiva. Esse crescimento na *acousticness* pode estar intimamente relacionado com a ascensão da música de língua portuguesa no Top 200 do Spotify, conforme analisado na secção 5.1. Essa correlação sugere que a evolução do consumo musical em Portugal está a abraçar uma maior diversidade de estilos, destacando a importância dos elementos acústicos na criação de uma identidade musical mais rica e representativa da cultura portuguesa.

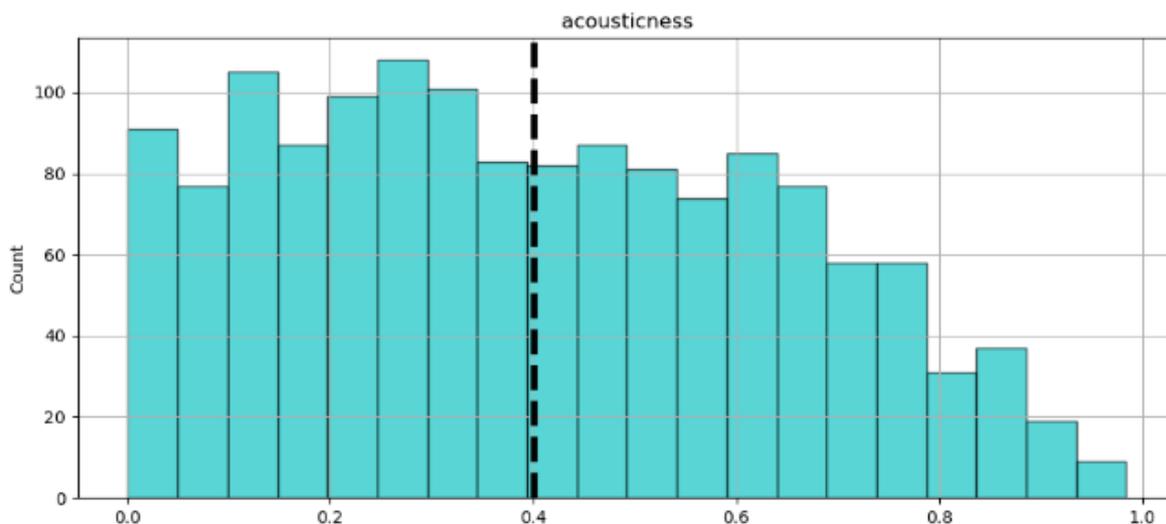


Figura 13 - Distribuições de frequência em relação à característica de áudio

Se retirarmos da base de dados apenas as músicas de língua portuguesa (ver Figura 13), é possível observar um aumento significativo na média de 'acousticness' para estas faixas. Ao comparar com a Figura 4, apresentado no Capítulo 3, verificamos que a média sobe de 0,28 para 0,408. Este aumento substancial comprova que as músicas portuguesas incorporam mais elementos acústicos. Assim, pode-se concluir que a tendência de crescimento na 'acousticness' pode ser atribuída ao aumento da presença de música de língua portuguesa nas *charts*.

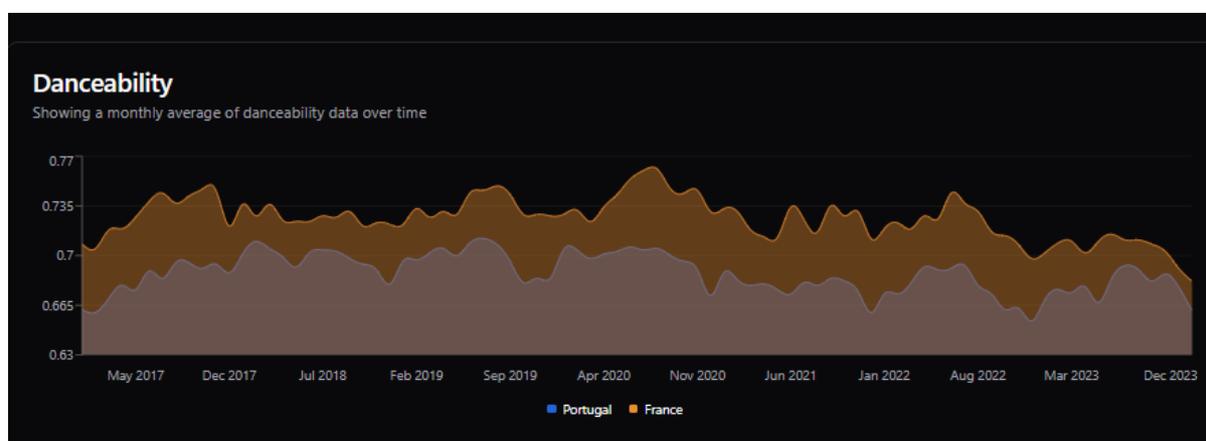


Figura 14 - Característica 'danceability' ao longo do tempo

Na Figura 14, observa-se que, desde 2017, a França manteve consistentemente uma média de *danceability* superior à de Portugal. Durante o período de 2017 a 2023, essa característica permaneceu em níveis elevados, variando entre 0.65 e 0.77 em ambos os países, com um pico notável na França no verão de 2020. Esses dados permitem concluir que, tanto em Portugal como na França, as músicas populares tendem a seguir um padrão de batidas fortes e ritmos festivos, sugerindo que a *danceability* é uma característica valorizada nas canções de sucesso nesses mercados.

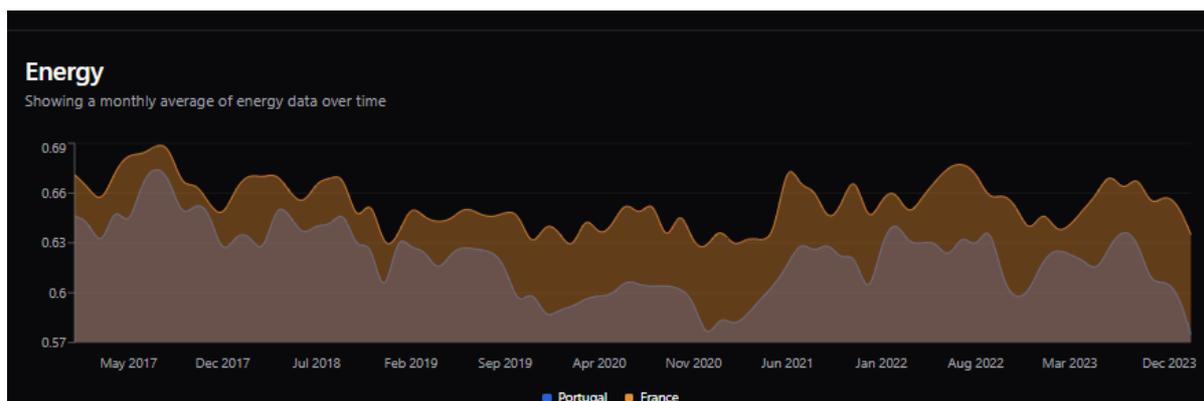


Figura 15 - Característica 'energy' ao longo do tempo

Na Figura 15, observa-se que, entre 2017 e 2023, a variável *energy* mantém-se em níveis elevados e relativamente constantes. Esse padrão sugere que as músicas populares em Portugal e França tendem a ser enérgicas, reforçando a conclusão anterior que existe uma preferência dos mercados por canções com intensidade e ritmo elevados. Curiosamente, Portugal e França apresentam picos e quedas semelhantes em diversos momentos ao longo desse período. Um exemplo disso ocorre entre julho e dezembro de 2017, quando a variável *energy* segue quase o mesmo padrão em ambos os países, ainda que com médias diferentes. Esse fenómeno também é visível em dezembro de 2018 e agosto de 2019. Em Portugal, observa-se uma queda acentuada de *energy* em 2020, assim como nos últimos meses de 2023. Essa diminuição pode estar relacionada ao aumento da variável *acousticness*, estudada anteriormente, uma vez que, como discutido no capítulo 3, há uma forte correlação negativa entre *energy* e *acousticness*.

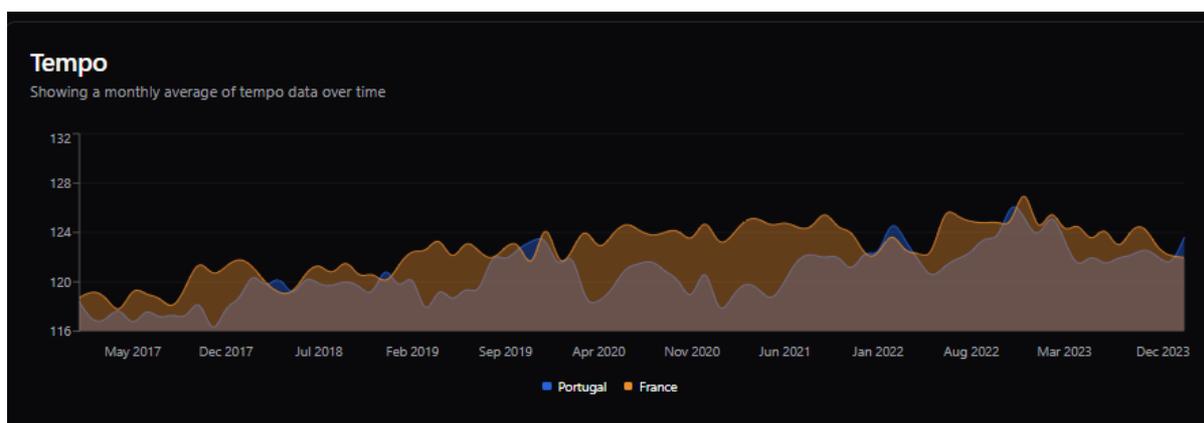


Figura 16 - Característica 'tempo' (BPM) ao longo do tempo

Na Figura 16, observa-se que a média de BPM (*Beats per Minute*) tem vindo a aumentar gradualmente desde 2017 nos mercados musicais de Portugal e França. No entanto, as tendências de crescimento em cada país diferem ligeiramente em ritmo e consistência. Em Portugal, a subida de BPM não foi constante, apresentando várias oscilações ao longo do período. Diversos picos e quedas são visíveis, refletindo uma possível variação nas preferências musicais ou influências temporárias de

diferentes estilos e géneros ao longo dos anos. Esses picos e descidas podem indicar períodos em que músicas de ritmos mais acelerados ou mais lentos dominaram temporariamente as listas de popularidade. Já em França, o aumento na média de BPM foi mais significativo e constante, revelando uma tendência mais clara de aceleração no ritmo das músicas populares. Este crescimento na França teve uma breve interrupção no início de 2022, onde ocorreu uma queda notável no BPM médio. No entanto, essa diminuição foi rapidamente compensada, com uma recuperação acentuada no verão do mesmo ano, sugerindo uma rápida adaptação ou mudança nas preferências do público francês para músicas com ritmos mais acelerados.

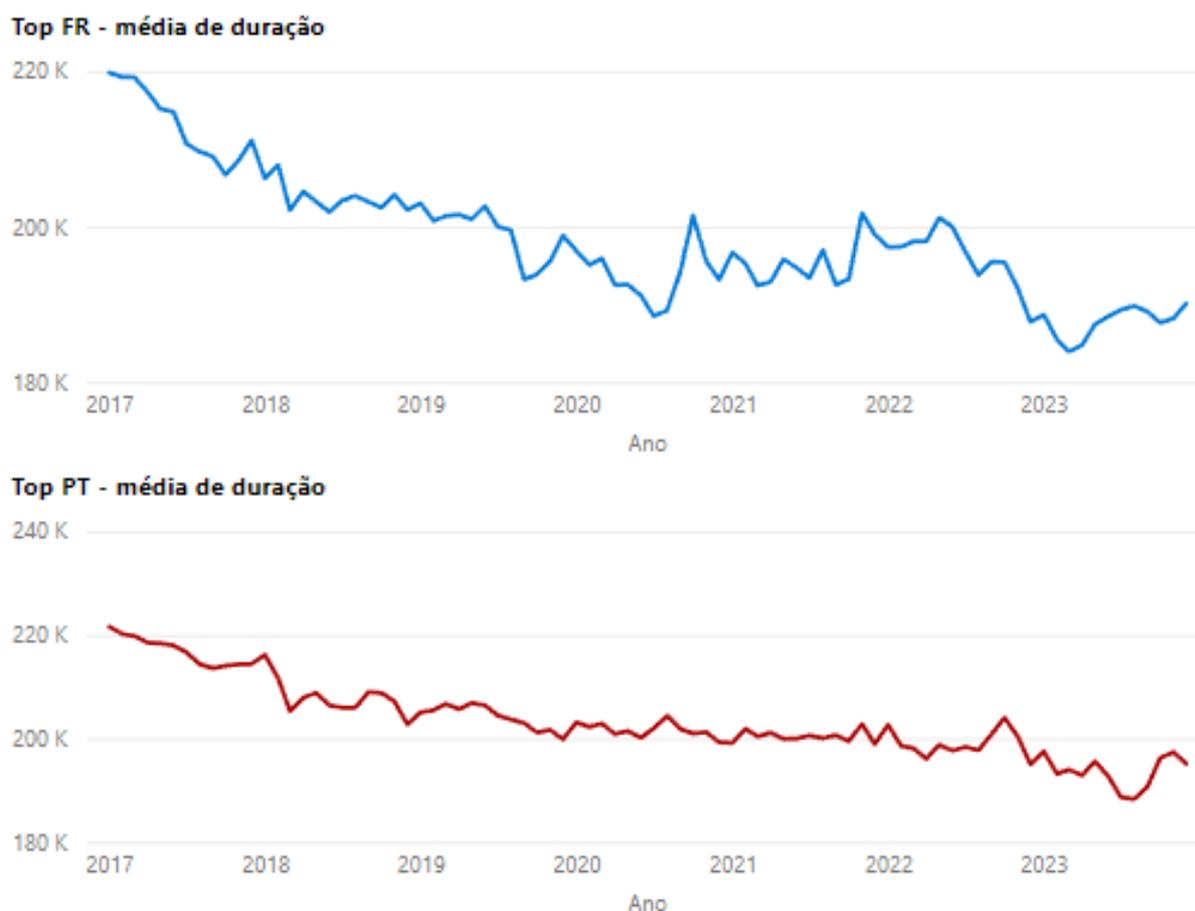


Figura 17 - Característica 'duration_ms' ao longo do tempo

Ao contrário do aumento observado no BPM, a duração das músicas populares tem vindo a diminuir consistentemente em ambos os países ao longo dos últimos anos (ver Figura 17). Essa redução no tempo médio das canções pode estar diretamente ligada à adaptação da indústria musical às novas plataformas digitais, como o TikTok, que desempenha um papel cada vez mais central na promoção e popularização de músicas. O TikTok, sendo uma aplicação onde se utilizam excertos curtos de músicas em vídeos, tem incentivado a criação de faixas mais curtas e impactantes. Em vez de longas introduções

e construções graduais, muitas músicas populares são agora projetadas para capturar a atenção rapidamente, com refrões ou *hooks* cativantes posicionados logo no início.

A análise das características das músicas populares em Portugal e França permite concluir que é possível identificar padrões e evoluções significativas ao longo dos últimos anos, refletindo tanto as preferências dos ouvintes quanto as adaptações da indústria musical.

5.3. Previsão do Consumo de Música Portuguesa no Spotify: Um Estudo Comparativo

H3: Com base em tendências passadas, é viável desenvolver um modelo de previsão para antecipar futuras mudanças no consumo de música portuguesa em Portugal?

Para responder a esta questão, foi criada uma base de dados derivada da original. Nesta nova base, extraiu-se o total de músicas portuguesas presentes no top 200 por dia, garantindo assim que todos os modelos fossem treinados com a maior quantidade de dados possível. A base de dados foi, então, dividida em conjuntos de treino e teste: o treino inclui todos os dados até 1 de janeiro de 2023, e o teste contém os dados posteriores a essa data.

Os modelos preditivos foram avaliados com base na comparação entre os valores previstos e os valores de teste.

Na Figura 18, observa-se a previsão feita pelo modelo **Prophet**. A linha laranja representa o conjunto de treino, enquanto a linha azul mostra as previsões do modelo. A área sombreada ao redor da linha azul indica o intervalo de confiança das previsões, refletindo a incerteza associada às mesmas.

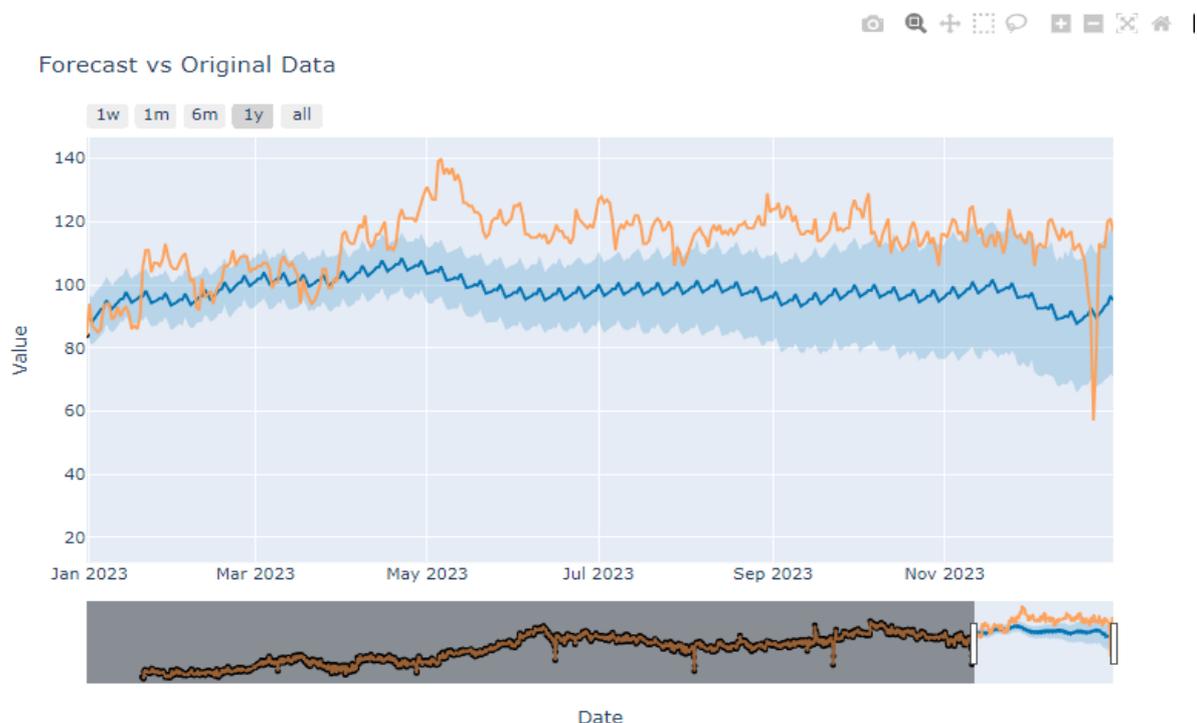


Figura 18 - Previsão da Presença de Músicas em Língua Portuguesa no TOP 200 do Spotify (Modelo Prophet)

Embora o modelo *Prophet* apresentado na Figura 18 não tenha apresentado os melhores resultados, capturou a tendência geral dos dados. No entanto, é visível que o modelo teve dificuldade em prever picos abruptos, o que é esperado, dado que o *Prophet* tende a suavizar os dados para identificar tendências de longo prazo e padrões sazonais.

Na Figura 19, o modelo *Neural-Prophet* apresentou um desempenho semelhante ao *Prophet*, também capturando a tendência geral, mas com dificuldades em prever picos e quedas bruscas.

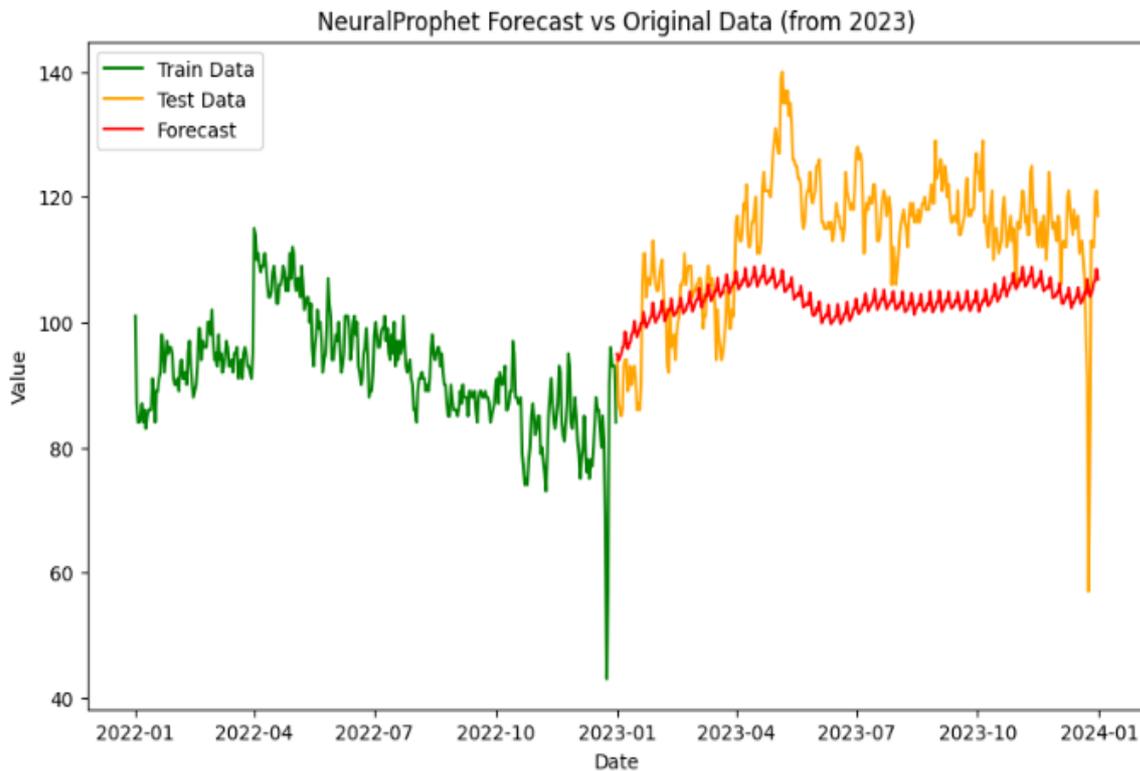


Figura 19 - Previsão da Presença de Músicas em Língua Portuguesa no TOP 200 do Spotify (Modelo Neural-Prophet)

Já na Figura 20, que apresenta as previsões feitas pelo modelo **LSTM**, nota-se uma diferença significativa em relação aos modelos anteriores. O **LSTM**, além de capturar a tendência geral, conseguiu modelar de forma mais eficaz as variações abruptas dos dados reais.

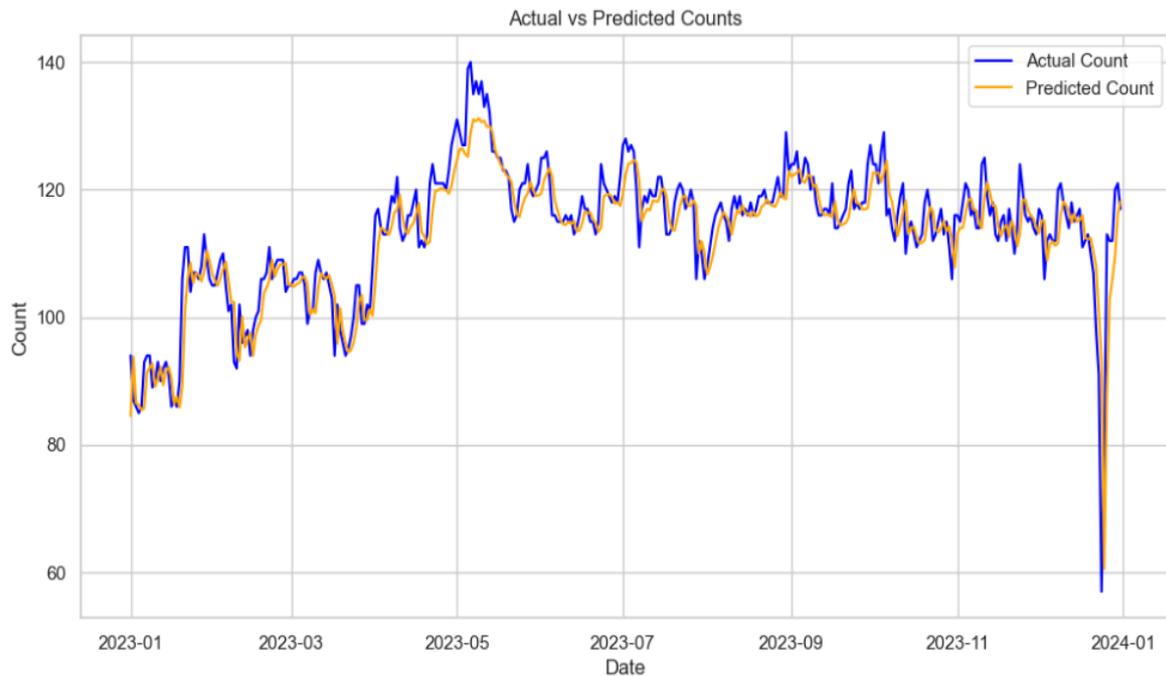


Figura 20 - Previsão da Presença de Músicas em Língua Portuguesa no TOP 200 do Spotify (Modelo LSTM)

O modelo **XGBoost**, representado na Figura 21, foi o que obteve os melhores resultados. Comparado com os modelos anteriores, o **XGBoost** demonstrou maior precisão tanto na previsão da tendência geral como nas variações presentes no conjunto de teste.

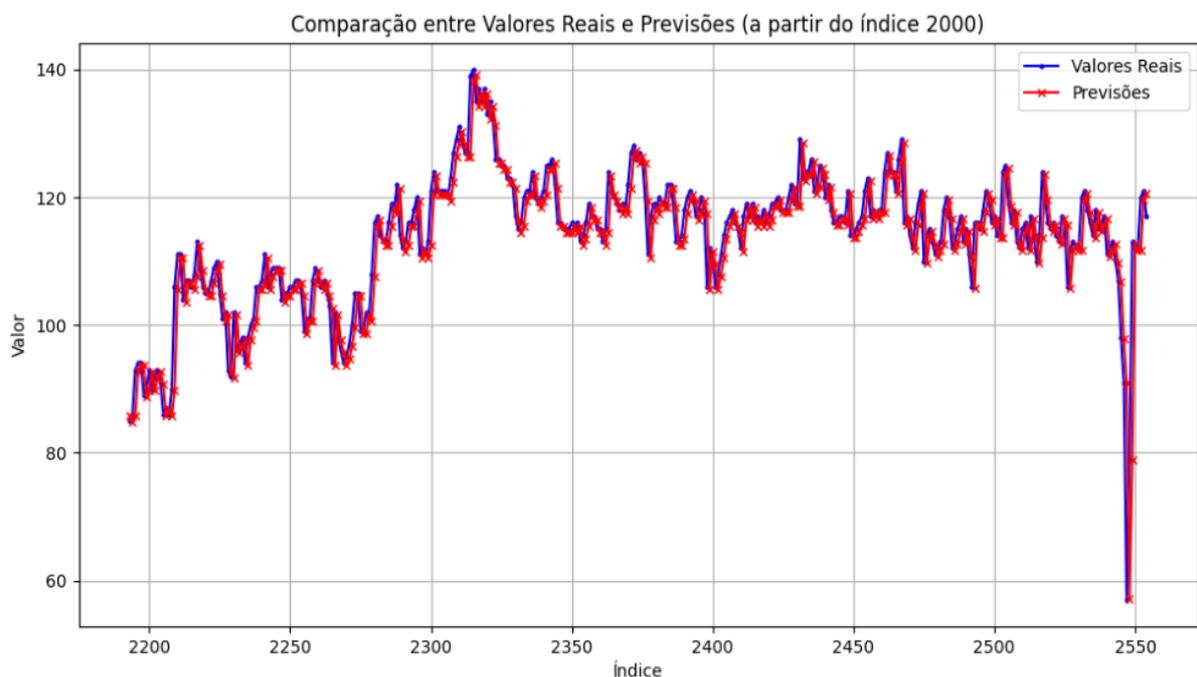


Figura 21 - Previsão da Presença de Músicas em Língua Portuguesa no TOP 200 do Spotify (Modelo XGBoost)

Para uma comparação mais objetiva dos resultados, segue-se uma tabela comparativa entre os quatro modelos estudados (ver Tabela 6).

Tabela 6 - Tabela de comparação de resultados dos modelos

Modelo	RMSE
Prophet	18.4881
Neural-Prophet	13.9303
LSTM	4.7164
XGBoost	4.6634

Com base no resultado do modelo XGBoost que apresentou uma RMSE de 4.6634, é possível afirmar que a construção de um modelo de previsão para antecipar futuras mudanças no consumo de música portuguesa em Portugal é viável. A RMSE relativamente baixa indica que o modelo XGBoost conseguiu captar bem os padrões históricos do consumo, sugerindo que ele pode fornecer previsões precisas.

Este modelo poderá servir como uma ferramenta útil para prever tendências futuras, mas deve ser continuamente ajustado para manter a sua eficácia.

Capítulo 6

Conclusões

Esta investigação abordou três questões principais sobre o mercado musical português, com foco nas suas características e previsões para o futuro. A análise inicial revelou padrões e tendências nas músicas do Top 200 semanal do Spotify em Portugal, destacando a evolução das preferências musicais ao longo dos anos.

Comparando com o mercado francês, observou-se uma clara diferença na adoção da língua nacional. Enquanto a música em francês sempre dominou o mercado local, em Portugal, a música em língua portuguesa só começou a ganhar destaque em novembro de 2019. Desde então, o consumo de músicas em português cresceu, alcançando mais de 50% do Top 200 e consolidando-se como a língua dominante em 2023.

Na segunda parte, ao examinar as características das músicas no Top 200 de Portugal, ficou evidente que o mercado privilegia canções com altos níveis de "danceability", energia e duração mais curta. Em contrapartida, as músicas em português apresentaram uma maior tendência para elementos acústicos, com uma média de "acousticness" significativamente superior.

Por fim, testaram-se quatro modelos preditivos para avaliar a viabilidade de prever o consumo de música portuguesa. O modelo XGBoost obteve o melhor desempenho, com um RMSE de 4,6634, comprovando a possibilidade de criar um modelo preditivo eficaz para o consumo de música portuguesa em Portugal. Este modelo pode servir como uma ferramenta valiosa para editoras e artistas na definição de estratégias de marketing e no planeamento de lançamentos futuros.

6.1. Limitações

Uma das maiores limitações encontradas nesta dissertação foi o difícil acesso a dados sobre o mercado musical português. Ao contrário de países como os Estados Unidos, que dispõem de grandes bases de dados focadas no seu mercado musical, Portugal ainda carece de estudos abrangentes nesta área. O peso cultural e a dimensão do mercado musical em países maiores facilitam a criação de bases de dados robustas e de fácil acesso, o que não acontece em Portugal, tornando a recolha de dados extensa e detalhada um grande desafio para este estudo.

Embora o website *spotifycharts.com* tenha sido uma fonte essencial para a obtenção dos dados necessários, o seu uso não foi particularmente prático. Para conseguir compilar uma base de dados abrangente de 2017 a 2023, foi necessário atualizar manualmente a página para cada data desejada, o que tornou o processo de recolha de dados significativamente mais moroso e complexo.

Após a construção da base de dados, outro desafio relevante foi o processamento das letras das músicas para a identificação da língua. Devido ao facto de as músicas em português e francês não terem o mesmo alcance global que o inglês, muitas letras não eram automaticamente reconhecidas pelas ferramentas de processamento de texto disponíveis. Como resultado, foi necessário corrigir

manualmente uma pequena parte dos dados, correspondente a 34 músicas das duas bases de dados, exigindo um esforço adicional para garantir a precisão da análise.

6.2. Futuras Análises

Para futuras investigações, um dos aspetos que considero importante aprofundar é a influência das rádios no desempenho das músicas no Top nacional.

Embora este tema tenha sido brevemente abordado nesta dissertação, um estudo mais detalhado sobre o impacto das rádios no sucesso das músicas seria extremamente relevante. Contactar rádios nacionais e obter dados específicos sobre a frequência com que certas músicas são tocadas poderia fornecer insights valiosos.

A exposição de uma música nas rádios desempenha, de facto, um papel crucial na sua popularidade? Este é um ponto que mereceria uma análise mais profunda, dado que as rádios podem ser um veículo fundamental para amplificar o sucesso de determinados artistas e faixas. Além disso, uma recolha de dados mais extensa poderia melhorar significativamente a qualidade da investigação. O treino de modelos preditivos tem uma forte correlação com a quantidade de dados disponíveis, e com uma base de dados maior seria possível capturar padrões e características que possam ter surgido antes de 2017. Este aumento de dados permitiria, ainda, uma melhor compreensão da evolução do consumo de música em Portugal ao longo de um período mais alargado.

No que diz respeito aos modelos preditivos, embora os resultados obtidos nesta dissertação já sejam bastante promissores, há sempre margem para melhorias. O modelo XGBoost mostrou um bom desempenho, mas a sua eficácia poderia ser otimizada. Para futuras iterações, seria interessante explorar uma combinação do XGBoost com o modelo LSTM (Long Short-Term Memory), que é altamente eficaz para séries temporais, ajustando os parâmetros de ambos os modelos para maximizar a precisão das previsões.

A integração de diferentes abordagens pode, potencialmente, oferecer resultados ainda mais robustos e precisos, elevando a capacidade preditiva do modelo.

Referências Bibliográficas

- ANACOM. (2016, May 13). *Serviço de radiodifusão*. Consultado a 28 de julho de 2024. Disponível em <https://www.anacom.pt/render.jsp?categoryId=385914>
- Araujo, C., Cristo, M., & Giusti, R. (2019). *Predicting music popularity on streaming platforms*. <https://doi.org/10.5753/sbcm.2019.10436>
- ARCOM. (s.d.). *Fun Radio: Changement de régime de diffusion de chansons d'expression française*. Consultado a 28 de julho de 2024. Disponível em <https://www.arcom.fr/nos-ressources/espace-juridique/decisions/fun-radio-changement-de-regime-de-diffusion-de-chansons-dexpression-francaise>
- ARCOM. (2024). *Les quotas de chansons à la radio*. Consultado a 28 de julho de 2024. Disponível em <https://www.arcom.fr/nous-connaître-nos-missions/promouvoir-et-protéger-la-creation/les-quotas-de-chansons-la-radio>
- ARCOM. (s.d.). *Les radios en France*. Consultado a 28 de julho de 2024. Disponível em <https://www.arcom.fr/radio-et-audio-numerique/les-radios-en-france>
- ARCOM. (2023). *Quotas de chansons d'expression française au premier trimestre 2023*. Consultado a 14 de julho de 2024. Disponível em <https://www.arcom.fr/nos-ressources/espace-juridique/decisions/quotas-de-chansons-dexpression-francaise-au-premier-trimestre-2023>
- Bello, P., & 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. <https://doi.org/10.1057/s41599-021-00855-1>
- Caramujo, C. (2017). *Music streaming market in Portugal: The introduction of a new music streaming service* (Dissertação de mestrado, ISCTE - Instituto Universitário de Lisboa, pp. 1–106). Consultado a 20 de dezembro de 2023. Disponível em <http://hdl.handle.net/10071/14160>
- Chambers, S. (2022). The curation of music discovery: The presentation of unfamiliar classical music on radio, digital playlists, and concert programmes. *Empirical Studies of the Arts*, 41(1), 304–326. <https://doi.org/10.1177/02762374221128729>
- Dairu, X., & Shilong, Z. (2021). Machine learning model for sales forecasting using XGBoost. In *Proceedings of the 2021 International Conference on Computer Engineering and Communication Technology (ICCECE)* (pp. 480–483). IEEE. <https://doi.org/10.1109/ICCECE51280.2021.9342304>

- Gulmatico, J. S., Susa, J. A. B., Malbog, M. A. F., Acoba, A., Nipas, M. D., & Mindoro, J. N. (2022). SpotiPred: A machine learning approach prediction of Spotify music popularity by audio features. In *ICPC2T 2022 - 2nd International Conference on Power, Control and Computing Technologies, Proceedings*. Scopus. <https://doi.org/10.1109/ICPC2T53885.2022.9776765>
- Interiano, M., Kazemi, K., Wang, L., Yang, J., Yu, Z., & Komarova, N. L. (2018). Musical trends and predictability of success in contemporary songs in and out of the top charts. *Royal Society Open Science*, 5(5), 171274. <https://doi.org/10.1098/rsos.171274>
- Kumar Jha, B., & Pande, S. (2021). Time series forecasting model for supermarket sales using FB-Prophet. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 547–554). IEEE. <https://doi.org/10.1109/ICCMC51019.2021.9418033>
- Kyle, M., & Niu, D. (2017). Strategic responses to cultural quotas: Evidence from French radio. Consultado a 14 de julho de 2024. Disponível em <https://www.lancaster.ac.uk/staff/desilvad/Strategic%20responses%20to%20cultural%20quotas%20evidence%20from%20French%20radio.pdf>
- Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications, and future trends. *Journal of Management Analytics*, 6(1), 1–29. <https://doi.org/10.1080/23270012.2019.1570365>
- Morais Leitão. (2024, February 7). Legal alert | Amendment to the Portuguese Radio Act. Consultado a 14 de julho de 2024. Disponível em <https://www.mlgs.pt/en/knowledge/legal-alerts/Legal-Alert-Amendment-to-the-Portuguese-Radio-Act/24823/>
- Patterson, J., & Gibson, A. (2017). *Deep learning: A practitioner's approach*. O'Reilly Media.
- Pınarbaşı, F. (2019). Demystifying musical preferences at the Turkish music market through audio features of Spotify charts. *Turkish Journal of Musicology*, 4(3), 265–279. <https://doi.org/10.30685/tujom.v4i3.62>
- Portugal.gov.pt. (2023). *Em setembro, 30% da música nas rádios volta a ser obrigatoriamente portuguesa*. Consultado a 14 de julho de 2024. Disponível em <https://www.portugal.gov.pt/pt/gc23/comunicacao/noticia?i=em-setembro-30-da-musica-nas-radios-volta-a-ser-obrigatoriamente-portuguesa>
- Ribeiro, F., & Costa Alves, T. (2021). Do I sound like a broken record? *Medijska istraživanja*, 27(1), 75–98. <https://doi.org/10.22572/mi.27.1.4>
- Sharma, D., Khetarpaul, S., Mohit Kumar, S., Parthasarathy, A., & Agarwalla, S. (2022). Performance prediction of songs on online music platforms. In W. Hua, H. Wang, & L. Li (Eds.), *Databases*

theory and applications (pp. 209–216). Springer International Publishing.
https://doi.org/10.1007/978-3-031-15512-3_19

Soares Araujo, C. V., Pinheiro de Cristo, M. A., & Giusti, R. (2019). Predicting music popularity using music charts. In *2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 859–864). IEEE. <https://doi.org/10.1109/ICMLA.2019.00149>

Spotify. (n.d.). *Spotify charts - Spotify charts are made by fans*. Consultado a 3 de agosto de 2024. Disponível em <https://charts.spotify.com/home>

Spotify. (2024, July 23). *Spotify reports second quarter 2024 earnings*. Consultado a 3 de agosto de 2024. Disponível em <https://newsroom.spotify.com/2024-07-23/spotify-reports-second-quarter-2024-earnings/>

Spotify for Developers. (n.d.). *Web API | Spotify for developers*. Consultado a 3 de agosto de 2024. Disponível em <https://developer.spotify.com/documentation/web-api>

Triebe, A., Erb, D., & Cichos, F. (2021). NeuralProphet: Explainable forecasting at scale. *arXiv*. <https://doi.org/10.48550/arXiv.2111.15397>