

Análise de dados do Spotify

Apresentação Final

Anna Carolina Fonseca - 1512996

Guilherme Bizzo - 1710563

Matheus Nogueira - 1810764

Roberto Mario L T L - 1711384

Relembrando

Fontes de dados

The new Spotify Charts is here

Unlock all of our global charts. Plus new charts for genres and cities. The Spotify Charts site you're on now will shut down starting June 3rd.

[Let's go](#)

[TOP 200](#)
[VIRAL 50](#)

Filter by







GLOBAL

DAILY

04/04/2022

TRACK

 STREAMS 

	1 – As It Was by Harry Styles	10,710,236
	2 – Heat Waves by Glass Animals	4,010,347
	3 – STAY (with Justin Bieber) by The Kid LAROI	3,220,392
	4 ▲ Enemy (with JID) - from the series Arcane League of Legends by Imagine Dragons	2,943,103
	5 ▼ Envolver by Anitta	2,841,752
	6 – Cold Heart - PNAU Remix by Elton John, Dua Lipa	2,622,408

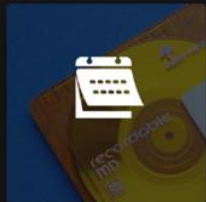
Categorias



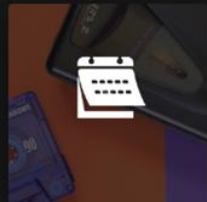
Anos 2010



Anos 2000



Anos 90



Anos 80



Anos 70



Anos 60



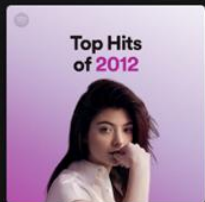
Anos 50

Hits da década



Top Hits de 2011

Relembre e descubra grandes sucessos em 201...



Top Hits de 2012

Relembre e descubra grandes sucessos em 201...



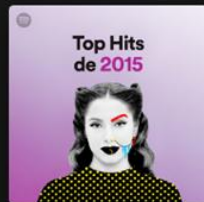
Top Hits de 2013

Relembre e descubra grandes sucessos em 201...



Top Hits de 2014

Relembre e descubra grandes sucessos em 201...



Top Hits de 2015

Relembre e descubra grandes sucessos em 201...



Top Hits de 2016

Relembre e descubra grandes sucessos em 201...



Top Hits de 2017

Relembre e descubra grandes sucessos em 201...

ENDPOINTS

Albums

[Get Album](#) GET[Get Several Albums](#) GET[Get Album Tracks](#) GET[Get Saved Albums](#) GET[Save Albums](#) PUT[Remove Albums](#) DELETE[Check Saved Albums](#) GET[Get New Releases](#) GET

Artists

[Get Artist](#) GET[Get Several Artists](#) GET[Get Artist's Albums](#) GET[Get Artist's Top Tracks](#) GET[Get Artist's Related Artists](#) GET

Shows

Episodes

Tracks

[Get Track](#) GET[Get Several Tracks](#) GET[Get User's Saved Tracks](#) GET[Save Tracks for Current User](#) PUT[Remove Tracks for Current User](#) DELETE

Spotify Web API

v1.0.0

You can use Spotify's Web API to discover music and podcasts, manage your Spotify library, control audio playback, and much more. Browse our available Web API endpoints using the sidebar at left, or via the navigation bar on top of this page on smaller screens.

In order to make successful Web API requests your app will need a valid access token. One can be obtained through [OAuth 2.0](#).

The base URI for all Web API requests is <https://api.spotify.com/v1>.

Need help? See our [Web API guides](#) for more information, or visit the [Spotify for Developers community forum](#) to ask questions and connect with other developers.

Questões a serem
respondidas

Análise descritiva

1. Comparar as características das músicas mais populares de diferentes regiões e global.
1. Analisar a evolução de algumas características de interesse ao longo do tempo.
1. Quantas músicas do Top 200 de cada país são cantadas no idioma dele?
1. Procurar músicas "anômalas" no Top 200.
1. Quais as características mais importantes para uma música ser popular?

Análise preditiva

1. Prever se uma música alcançará o Top 200 global/regional.
 1. Prever a quantidade de streams de uma música com base em suas características.
 1. Prever os tipos de playlist que cada música pode pertencer para aumentar as chances de viralizar.
-

Checkpoint 2

Análise descritiva

1. Comparar as características das músicas mais populares de diferentes regiões e global.
1. Analisar a evolução de algumas características de interesse ao longo do tempo.
1. Quantas músicas do Top 200 de cada país são cantadas no idioma dele?
1. Procurar músicas "anômalas" no Top 200.
1. Quais as características mais importantes para uma música ser popular?

Análise preditiva

1. Prever se uma música alcançará o Top 200 global/regional.
 1. Prever a quantidade de streams de uma música com base em suas características.
 1. Prever os tipos de playlist que cada música pode pertencer para aumentar as chances de viralizar.
-

Novidades

Finalização da análise das músicas nos idiomas nativos de cada país

Faltaram os resultados, devido aos seguintes desafios:

- Permissão de utilizar as letras de um site.
- URLs e caracteres especiais.

Perguntas a serem respondidas

1. O quanto que as músicas em idioma nativo estão presentes em cada país?
2. Quais os idiomas estrangeiros das músicas que mais são ouvidos nesses países?

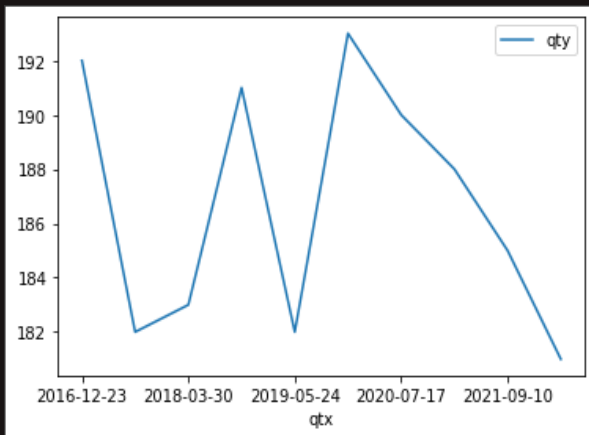
Como respondê-las?

1. Definir idiomas oficiais dos países.
2. Separar nome das músicas e artistas.
3. Gerar e buscar pelo URL das letras das músicas.
4. Extrair a letra.
5. Detectar o idioma.
6. Separar em listas de idioma oficial e não oficial.

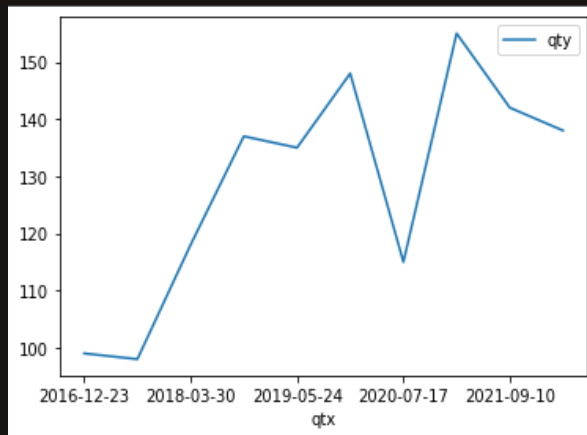
Fonte de Dados

1. Dados da API sobre as trilhas, os artistas e playlists.
 2. Site para a busca das letras.
-

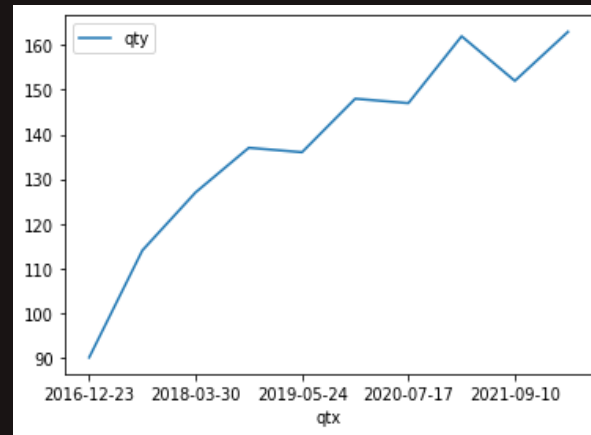
Quantidade de músicas no idioma nativo ao longo do tempo



Estados Unidos

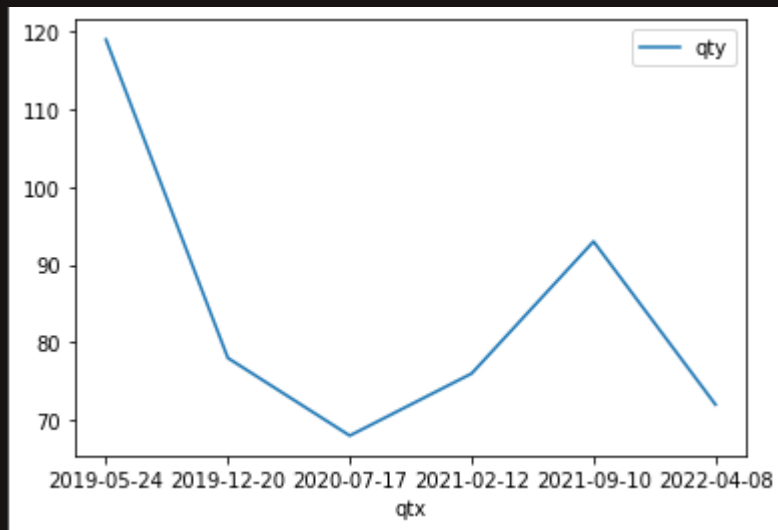


Brasil

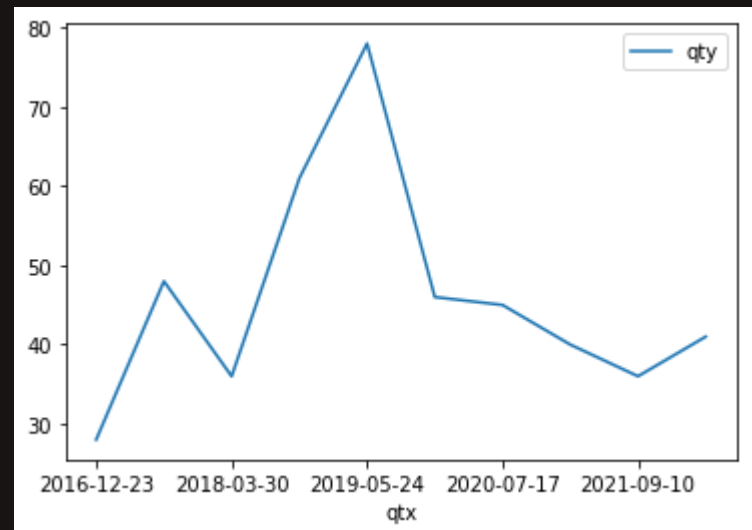


México

Quantidade de músicas no idioma nativo ao longo do tempo

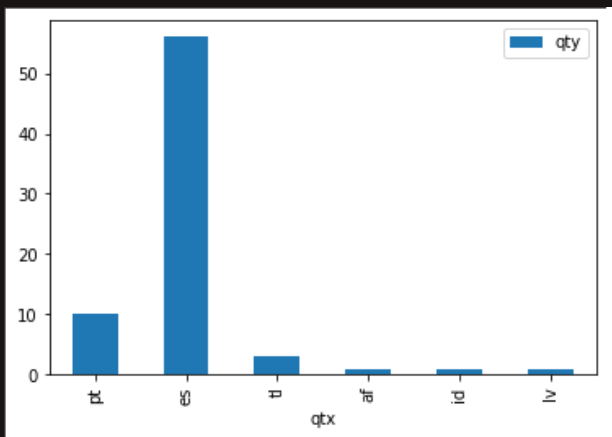


Índia

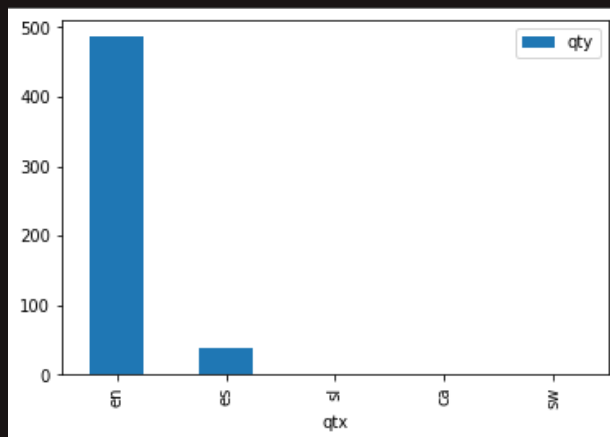


Alemanha

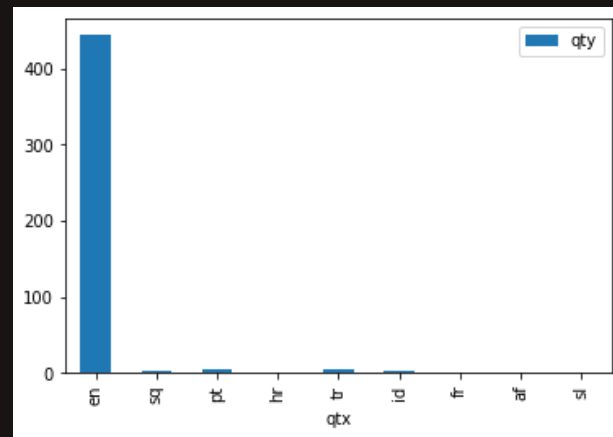
Tendências de idiomas estrangeiros em cada país



Estados Unidos

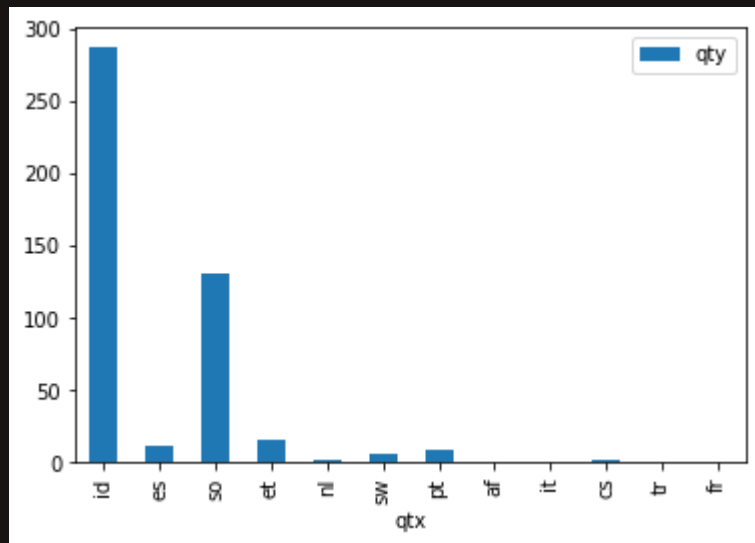


Brasil

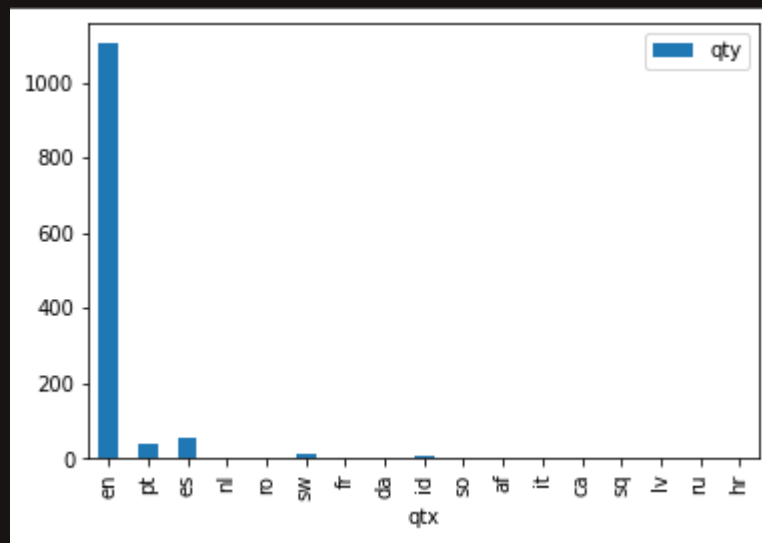


México

Tendências de idiomas estrangeiros em cada país

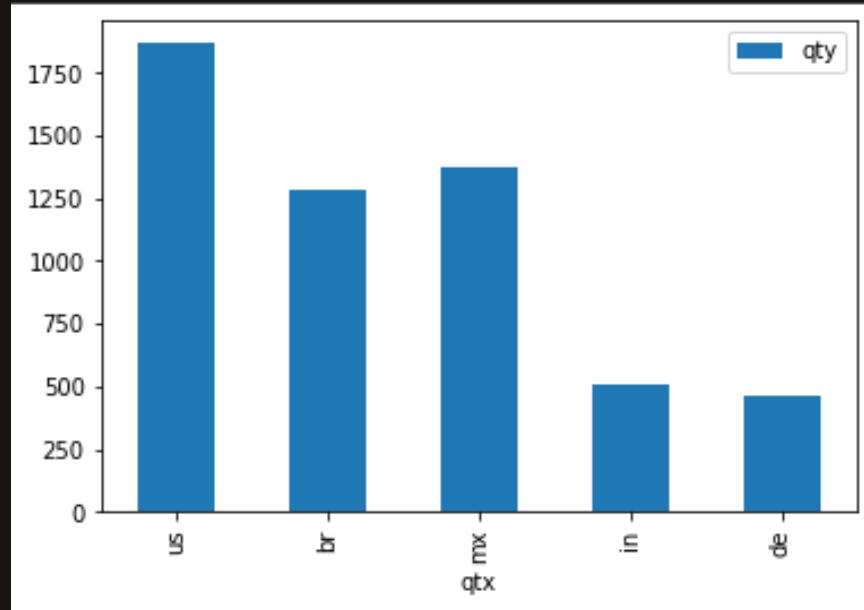


Índia



Alemanha

Total de músicas nos idiomas nativos ouvidas em cada país



Análise descritiva

1. Comparar as características das músicas mais populares de diferentes regiões e global.
1. Analisar a evolução de algumas características de interesse ao longo do tempo.
1. Quantas músicas do Top 200 de cada país são cantadas no idioma dele?
1. Procurar músicas "anômalas" no Top 200.
1. Quais as características mais importantes para uma música ser popular?

Análise preditiva

1. Prever se uma música alcançará o Top 200 global/regional.
 1. Prever a quantidade de streams de uma música com base em suas características.
 1. Prever os tipos de playlist que cada música pode pertencer para aumentar as chances de viralizar.
-

Perguntas a serem respondidas

1. Será que todas as músicas do Top Charts Global apresentam características semelhantes? Ou será que existem músicas populares que são **anômalas**?

Como respondê-las?

1. Utilizando uma combinação de modelos de detecção de anomalias.

Modelos utilizados:

1. ABOD
2. Autoencoders
3. Isolation Forest
4. Local Outlier Factor

Fonte de Dados

1. Features fornecidas pela Spotify Web API para as músicas do Top Charts Global.

Resultados individuais dos modelos

1. **ABOD** - 0 músicas classificadas como anômalas
2. **Autoencoder** - 258 músicas classificadas como anômalas
3. **Isolation Forest** - 61 músicas classificadas como anômalas
4. **Local Outlier Factor** - 258 músicas classificadas como anômalas

Individualmente, esses resultados podem ser questionáveis.

Para aumentar nossa confiança nos resultados, combinamos os outputs individuais dos modelos.

Resultados combinados

Estratégia: calcular a probabilidade de uma música ser anômala utilizando as probabilidades individuais dos modelos como peso de uma média ponderada.

Resultado final: 25 músicas classificadas como anomalia.
Vejam algumas delas...

Resultados combinados

Nome	Daceab	Energy	Loudness	Speech	Accoust	Instrument	Liveness	Valence	Duration
Chromatica I	0.231	0.4570	-10.773	0.0318	0.01260	0.875000	0.3270	0.0588	60400
Beautiful Trip	0.331	0.5130	-15.392	0.6320	0.97200	0.953000	0.8820	0.4200	37013
To Be Loved	0.356	0.1760	-8.291	0.0391	0.84300	0.000000	0.2310	0.2890	403962
Dawn FM	0.280	0.4780	-8.755	0.0370	0.67400	0.000058	0.4830	0.0807	96240
FEAR.	0.588	0.4790	-7.039	0.2810	0.60400	0.000007	0.5270	0.4340	460573
Valores Médios	0.684269	0.635878	-6.309993	0.127005	0.229244	0.011145	0.178445	0.490334	203397

Análises Preditivas

Análise descritiva

1. Comparar as características das músicas mais populares de diferentes regiões e global.
1. Analisar a evolução de algumas características de interesse ao longo do tempo.
1. Quantas músicas do Top 200 de cada país são cantadas no idioma dele?
1. Procurar músicas "anômalas" no Top 200.
1. Quais as características mais importantes para uma música ser popular?

Análise preditiva

1. Prever se uma música alcançará o Top 200 global/regional.
 1. Prever a quantidade de streams de uma música com base em suas características.
 1. Prever os tipos de playlist que cada música pode pertencer para aumentar as chances de viralizar.
-

Perguntas a serem respondidas

1. Será possível prever se uma música entrará ou não, em algum momento, no Top 200 Global do Spotify Charts?

Como respondê-las?

1. Utilização de 2 modelos de classificação e algumas técnicas de apoio

Modelos:

1. Regressão Logística
2. Random Forest

Fonte de Dados

1. Features fornecidas pela Spotify Web API para as músicas
 - a. do Spotify Charts Top 200 Global
 - b. aleatórias que não entraram no Top 200 Global.

Modelos Utilizados

- Regressão Logística
- Random Forest

Para ambos os modelos, foram implementadas 4 variações:

1. Uma versão com todas as variáveis disponíveis
2. Uma sem variáveis sobre o artista da música
3. Uma utilizando PCA com 5 componentes
4. Uma retirando música anômalas

Em todas as versões, foi utilizado um *Grid Search Cross Validation* para seleção de hiper-parâmetros.

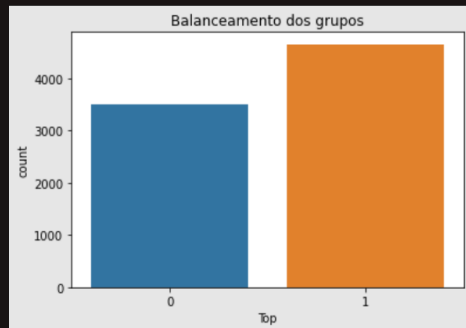
Variáveis utilizadas

- danceability
- energy
- key
- loudness
- mode
- speechiness
- acousticness
- instrumentalness
- liveness

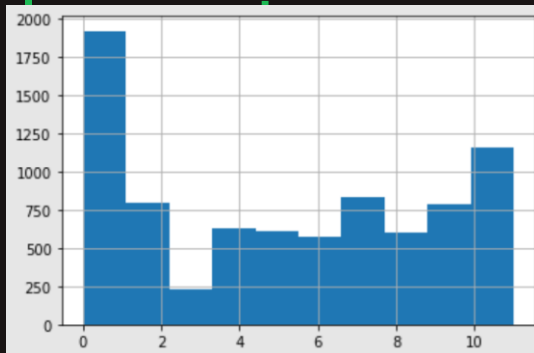
- valence
- tempo
- duration_ms
- ano_lancamento
- mes_lancamento
- dia_semana_lancamento
- Popularidade Artista
- Seguidores
- Top [Target]

Pré processamento

Balanceamento:



Variável Key, Mes e Dia da semana do



	key1	key2	key3	key4	key5	key6	key7	key8	key9	key10	key11
0	0	0	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0	0
2	0	0	0	0	0	0	0	0	0	0	1
3	1	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	1	0	0	0
...
8155	0	0	0	0	0	0	0	0	1	0	0
8156	0	0	0	0	1	0	0	0	0	0	0
8157	0	0	0	1	0	0	0	0	0	0	0
8158	1	0	0	0	0	0	0	0	0	0	0

	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez
0	0	1	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	1	0	0	0	0
3	0	0	0	0	0	1	0	0	0	0	0
4	0	0	1	0	0	0	0	0	0	0	0

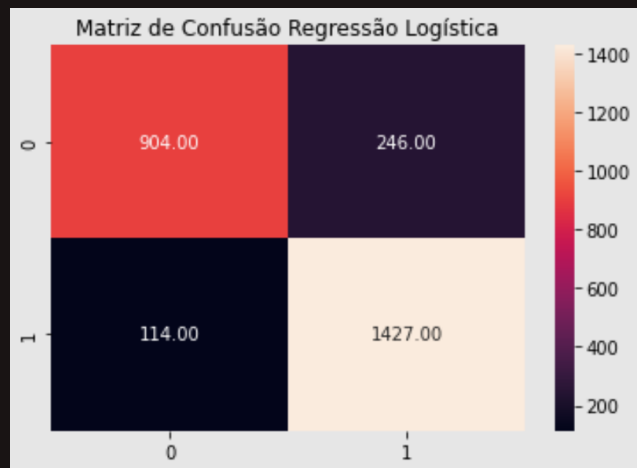
	ter	qua	qui	sex	sab	dom
0	0	0	1	0	0	0
1	0	0	0	1	0	0
2	0	0	1	0	0	0
3	0	0	0	1	0	0
4	0	0	1	0	0	0

Treinamento

- Separação de **33%** dos dados para teste.
- Escolha de hiperparâmetros:
 - Regressão Logística:
 - solver = ['newton-cg', 'lbfgs', 'liblinear', 'sag', 'saga']
 - max_iter = [10,50,100,200]
 - Random Forest:
 - n_estimators: [10,50,100,200],
 - criterion: ['gini', 'entropy', 'log_loss']
 - max_depth: [None,5,10,20,50]
 - bootstrap: [False,True]
 - max_samples:[10,20,50,100]

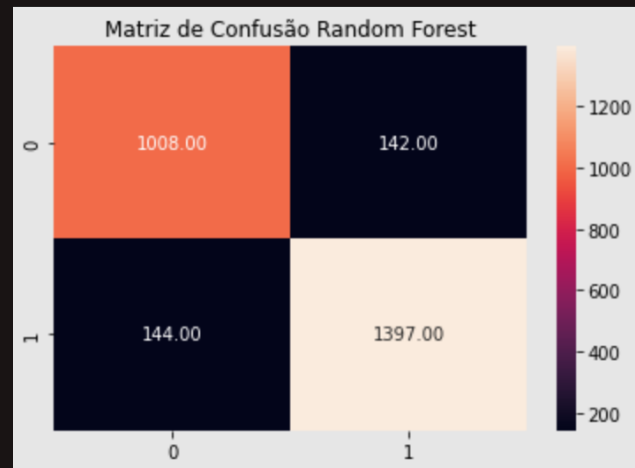
Resultados versão default

Regressão Logística



Acurácia = 0,87
Precisão = 0,85
Recall = 0,93

Random Forest



Acurácia = 0,89
Precisão = 0,91
Recall = 0,91

Importância de variáveis

Técnica utilizada: *permutation importance* no modelo treinado

Regressão Logística

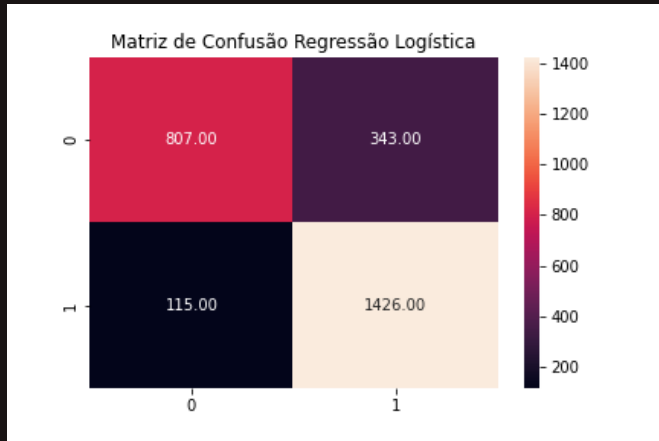
	Variável		Variável
1	Pop Artista	6	Instrum
2	Ano lançamento	7	Loudness
3	Sexta	8	Speech
4	Danceability	9	Duration
5	Quinta	10	Energy

Random Forest

	Variável		Variável
1	Ano lançamento	6	Instrum
2	Pop Artista	7	Tempo
3	Seguidores	8	Loudness
4	Speech	9	Accoustic
5	Danceability	10	Energy

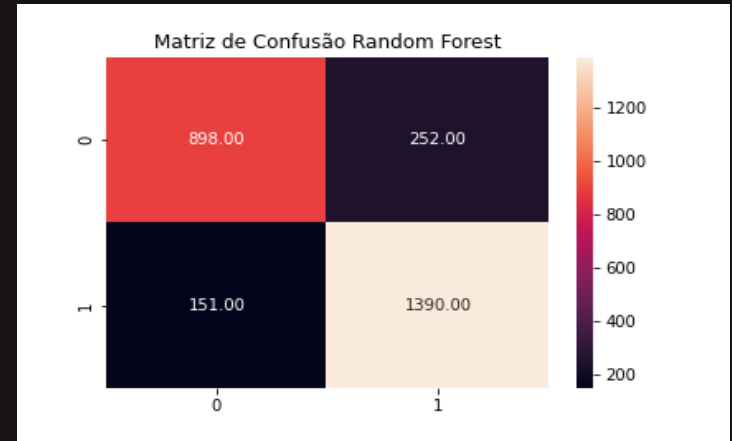
Resultados versão sem artista

Regressão Logística



Acurácia = 0.83 (0,87)
Precisão = 0.81 (0,85)
Recall = 0.93 (0,93)

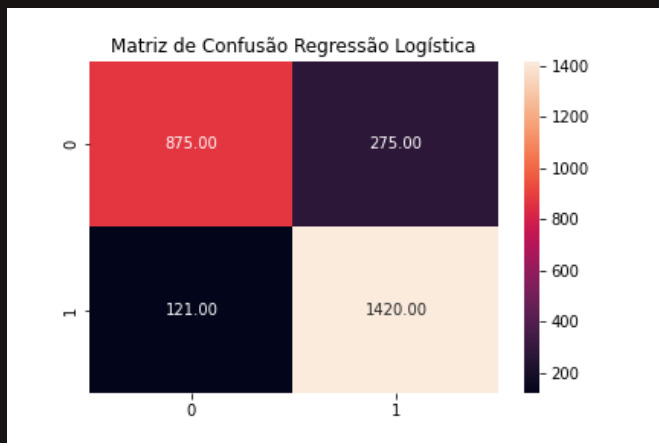
Random Forest



Acurácia = 0,85 (0,89)
Precisão = 0,85 (0,91)
Recall = 0,90 (0,91)

Resultados versão PCA

Regressão Logística

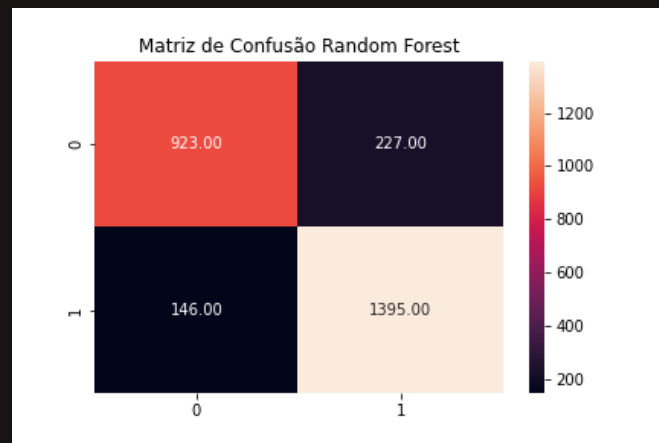


Acurácia = 0.85 (0,87)

Precisão = 0.84 (0,85)

Recall = 0.93 (0,93)

Random Forest



Acurácia = 0,86 (0,89)

Precisão = 0,86 (0,91)

Recall = 0,91 (0,91)

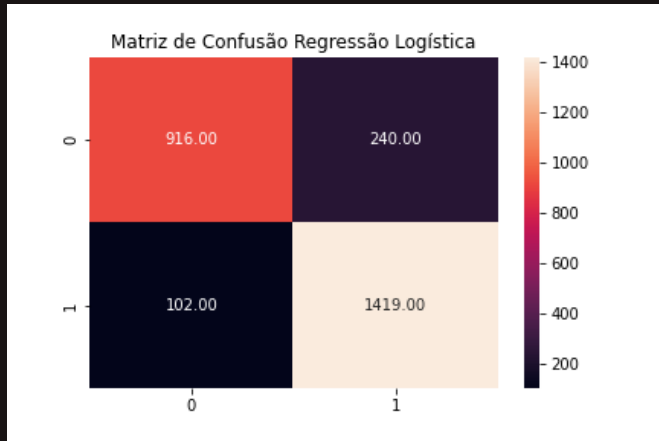
Remoção de Anomalias

Estratégia: foi utilizado o mesmo procedimento apresentado anteriormente para identificação de anomalias.

Resultado: 22 anomalias entre as músicas do Top Charts Global e 20 anomalias entre as músicas fora do Top Charts foram retiradas.

Resultados versão sem músicas anômalas

Regressão Logística

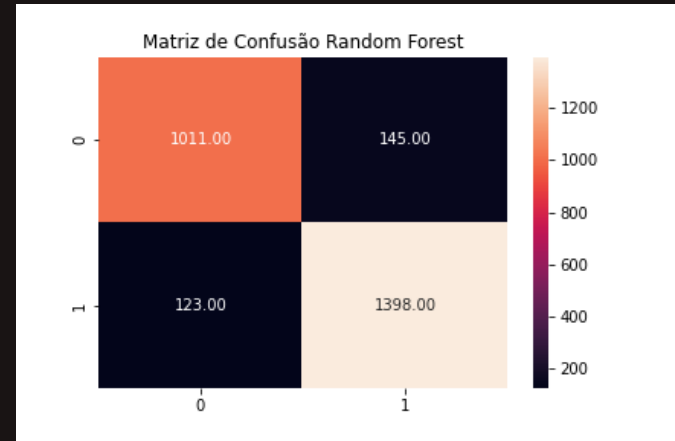


Acurácia = 0.87 (0,87)

Precisão = 0.86 (0,85)

Recall = 0.93 (0,93)

Random Forest



Acurácia = 0,90 (0,89)

Precisão = 0,91 (0,91)

Recall = 0,92 (0,91)

Curiosidade

Os resultados do *permutation importance* da regressão logística indicaram que a música ser lançada no final da semana (quinta ou sexta feira) é importante para definir se ela alcançará o Top 200 Global.

	Modelo Default	Sem Artista	PCA	Sem anomalias
Rank Sexta	3	4	3	3
Rank Quinta	5	5	4	5

What is the best day to release a song?

A. What is the best day to release a song?

For starters, if all the world's most famous artists schedule their music releases on a Friday, it's not by chance. Did you know that all labels release their new music on a Friday? Again, this is not a coincidence.

Let's mention that the [Billboard](#) (charts that reflect the success of the most popular musical titles and albums in the United States) compiles its rankings for the week starting on Friday. Therefore, posting **on a Friday** will be the most profitable way to increase your chances of appearing at the top of the **Billboard charts**. There is also the [SNEP](#), the equivalent of Billboard in France, which also compiles the same day.

<https://www.wiseband.com/blog/what-is-the-best-day-to-release-a-song/>



Why Does So Much New Music Drop at the Same Time Each Friday?



BY ERIC SKELTON
Head of music. Contact: Instagram / Twitter

May 07, 2018



<https://www.complex.com/pigeons-and-planes/2018/05/why-does-so-much-music-release-friday-each-week>

What's the best day to release music?

Madeleine Amos · in Artists Advice · Support / FAQs

September 9, 2021



You want to give your new music the best possible chance of success, so make sure you pick the best day of the week to release music online.

What's the best day to release music? Friday is day of the week that most of the music industry agrees is the best time to release new music.

Generally, Friday is the best day of the week to release music online, especially for established artists. It's when most record labels schedule releases, to maximise on a full week of sales, and when you'll find the most listeners eagerly seeking out new music to stream and add to their playlists.

<https://routenote.com/blog/whats-the-best-day-to-release-music/#:~:text=Generally%2C%20Friday%20is%20the%20best,and%20add%20to%20their%20playlists>

Análise descritiva

1. Comparar as características das músicas mais populares de diferentes regiões e global.
1. Analisar a evolução de algumas características de interesse ao longo do tempo.
1. Quantas músicas do Top 200 de cada país são cantadas no idioma dele?
1. Procurar músicas "anômalas" no Top 200.
1. Quais as características mais importantes para uma música ser popular?

Análise preditiva

1. Prever se uma música alcançará o Top 200 global/regional.
 1. Prever a quantidade de streams de uma música com base em suas características.
 1. Prever os tipos de playlist que cada música pode pertencer para aumentar as chances de viralizar.
-

Perguntas a serem respondidas

1. Será possível prever a quantidade de streams que uma música terá com base nas suas características?

Como respondê-las?

1. Utilização de 2 modelos de regressão e algumas técnicas de apoio

Modelos:

1. Regressão Linear
2. Random Forest

Fonte de Dados

1. Features fornecidas pela Spotify Web API para as músicas do Spotify Charts Top 200 Global.

Modelos Utilizados

- Regressão Linear
- Random Forest

Para os dois modelos, foram implementados:

1. Uma versão com todas as variáveis + análise de importância de variáveis
2. Uma versão com PCA (5 componentes)

Vamos apresentar essa pergunta diretamente pelo notebook.

Obrigado pela
atenção!