



Análise Spotify

Vinícius – RA00319760
Vitória – RA00320578
Maria Eduarda – RA00318891



Mars Is a Cold Place
The 15th Planet

2:54



3:49



01

Tema e objetivo



Mars Is a Cold Place
The 15th Planet

2:54



3:49

Contexto

Estamos vivendo uma nova era do mercado musical. O streaming e as redes sociais ditam as regras do jogo e transformam as notas musicais em notas de dinheiro.

Em 2021:

- O mercado da música cresceu 32% no Brasil em relação ao ano anterior.
 - O streaming é responsável 85% desse volume, segundo relatório da Pró-música, uma associação de produtores fonográficos associados.
- R\$ 2,1 bilhões foram gerados pela indústria musical no Brasil.
 - Desses, R\$ 1,8 bilhão veio do streaming.

Joyce Affonso. "A indústria da música: conheça as cifras desse mercado bilionário". Nubank.
Disponível em: <https://blog.nubank.com.br/cifras-da-musica/>. Acesso em: 27/04/2024



Mars Is a Cold Place
The 15th Planet

2:54



3:49

Tema e objetivo

Área de interesse

Música e entretenimento.

OBJETIVO DO NEGÓCIO

Produzir músicas de sucesso.

OBJETIVO DO PROJETO

Prever a popularidade de uma música a partir de suas características como dançabilidade, energia, acústica, ao vivo, instrumentalidade, etc. usando a playlist Top50 Global do Spotify.



Mars Is a Cold Place
The 15th Planet

2:54



3:49



02

Planejamento



Mars Is a Cold Place
The 15th Planet

2:54



3:49

Notion



VISÕES

01

Fase CRISP

02

Andamento

03

Calendário

04

Pessoas



Projeto: Spotify

Consumo de API's para análise de dados - 29.04

Board

Andamento

Calendário

Pessoas



🔍

New

Geral

7



Entender o negócio

2

Entender os dados

4

Preparação dos dados

9

Modelagem

1

PDF do escopo do projeto

Entregáveis

Links úteis

PDF CRISP-DM

Configurando Windows

Aprender como usar git no VSCode

Relatório

+ New

DEFINIR OBJETIVOS

DEFINIR CRONOGRAMA

+ New

COLETAR DADOS INICIAIS

DESCREVER OS DADOS

Documentação da API: Explicar o que são os atributos de cada música.

EXPLORAR E VISUALIZAR OS DADOS (Gráficos)

+ New

SELECIONAR OS DADOS

DERIVAR OS DADOS

UNIFORMIZAR OS DADOS

CRIAR BANCO DE DADOS

Especificar tabela e campos do banco de dados

CRIAR API

Transformar requests da API em CSV

Editar requests para batches de 50 músicas

Criar DUMP do banco de dados (arquivo .db)

+ New

MODELAGEM: definir técnicas, testar e avaliar modelos

+ New



Mars Is a Cold Place

The 15th Planet

2:54



3:49



Requirements

```
ipykernel==6.29.4  
spotipy==2.22.1  
python-dotenv==1.0.1  
pandas==2.2.1  
Flask==3.0.3  
numpy==1.26.4  
matplotlib==3.8.4  
seaborn==0.13.2  
scikit-learn==1.4.2
```



Mars Is a Cold Place
The 15th Planet

2:54



3:49



03

Extração dos dados

Através da API



Mars Is a Cold Place
The 15th Planet

2:54



3:49

Requisições para a API



```
client_credentials_manager = SpotifyClientCredentials(  
    ... client_id=os.environ.get('SPOTIPY_CLIENT_ID'),  
    ... client_secret=os.environ.get('SPOTIPY_CLIENT_SECRET')  
)  
sp = spotipy.Spotify(client_credentials_manager=client_credentials_manager)
```

```
# Faz uma requisição na API  
playlist = sp.playlist_items(  
    ... playlist_id=playlist_id,  
    ... fields='items.track(album.release_date, artists.name, name, popularity, id)'  
)
```

Requisições das
playlists

```
for batch in batches:  
    track_features_list = sp.audio_features(tracks=batch['track_id'])  
    for track_dict in track_features_list:  
        df_list.append(pd.DataFrame([track_dict]))
```

Requisições das
features



Mars Is a Cold Place
The 15th Planet

2:54



3:49



04

Criação do banco de dados

Com sqlite



Mars Is a Cold Place
The 15th Planet

2:54



3:49

Criação do banco de dados



```
def create_db_from_csv(csv_files, db_name):
    """# Conecta com o banco de dados. Se inexistente, cria-o
    conn = sqlite3.connect(db_name)
    cur = conn.cursor()

    # Iteração nos CSVs
    for csv_file in csv_files:
        table_name = csv_file.split('.')[0] # extrai o nome do arquivo
        with open(csv_file, 'r', newline='', encoding='utf-8') as f:
            reader = csv.reader(f) # transforma cada linha em uma lista
            headers = next(reader) # Retorna a 1a linha e itera para a 2a
            column_names = ','.join(headers)
            placeholders = ','.join('?' * len(headers))
            cur.execute(f'CREATE TABLE IF NOT EXISTS {table_name} ({column_names})')
            cur.executemany(f'INSERT INTO {table_name} VALUES ({placeholders})', reader)

    conn.commit()
    conn.close()

# Example usage:
csv_files = ['top_songs.csv', 'top_songs_features.csv']
db_name = 'database.db'

create_db_from_csv(csv_files, db_name)
```

Tabela **top_songs**:

playlist_country
artist_name
track_name
release_date
popularity
track_id
playlist_id

Tabela **top_songs_features**:

danceability
energy
key
loudness
mode
speechiness
acousticness
instrumentalness
liveness
valence
tempo
type
id
uri
track_href
analysis_url
duration_ms
time_signature



Mars Is a Cold Place
The 15th Planet

2:54



3:49



05

Análise exploratória

Gráficos



Mars Is a Cold Place
The 15th Planet

2:54



3:49



Features

Popularity – Popularidade (01 – 50)

Danceability – Dançabilidade (0,0 – 1,0)

Energy – Energia (medida perceptiva de intensidade e atividade) (0,0 – 1,0)

Loudness – Intensidade (volume de uma faixa em decibéis) (-60 e 0 dB)

Speechiness – Fala (Detecta a presença de palavras faladas) (0,0 – 1,0)

Acousticness – Acústica (instrumentos que não usam eletricidade) (0,0 – 1,0)

Instrumentalness – Instrumental (prevê se uma faixa não contém vocais) (0,0 – 1,0)

Liveness – Vivacidade (detecta a presença de público na gravação) (0,0 – 1,0)

Valence – Valência (mede a positividade musical transmitida por uma faixa) (0,0 – 1,0)

Tempo – Representa o andamento geral estimado de uma faixa em batidas por minutos (BPM)

Duration_ms – Duração da faixa em milissegundos

Time_signature – Fórmula de compasso estimada (3 – 7)

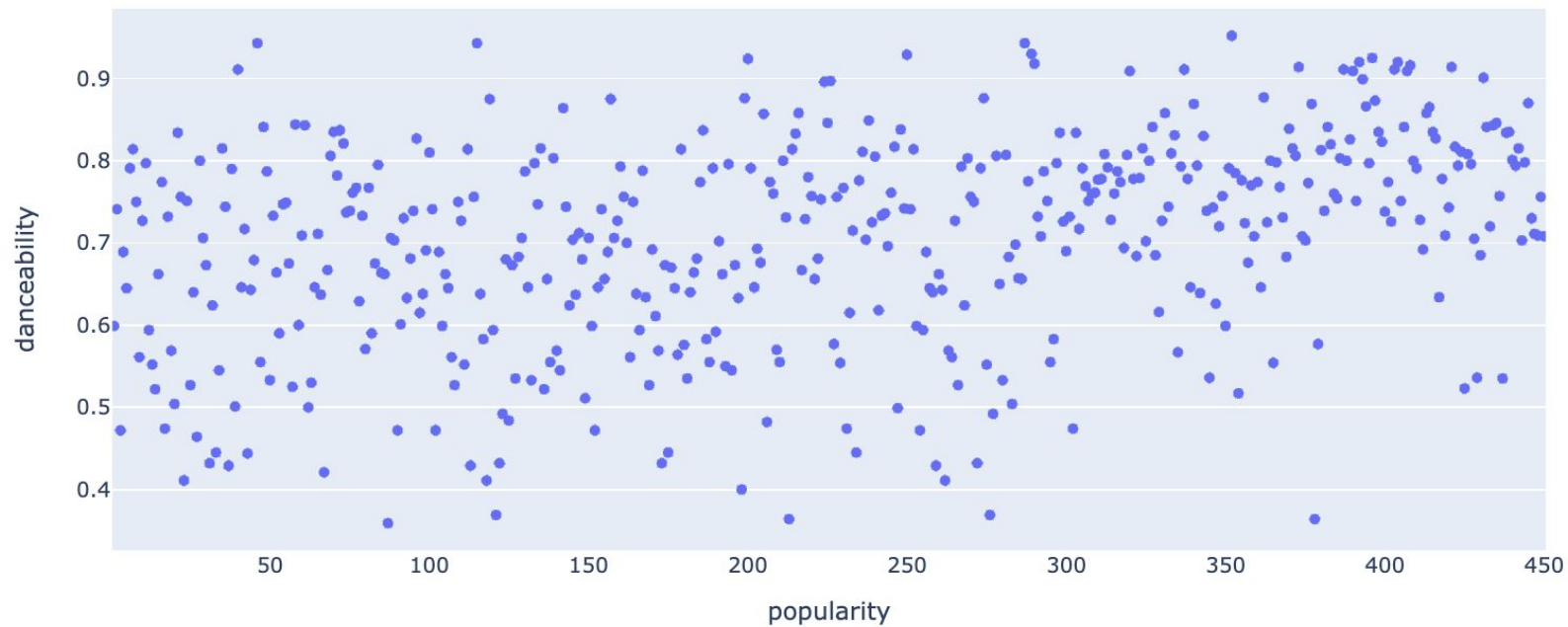


Mars Is a Cold Place
The 15th Planet

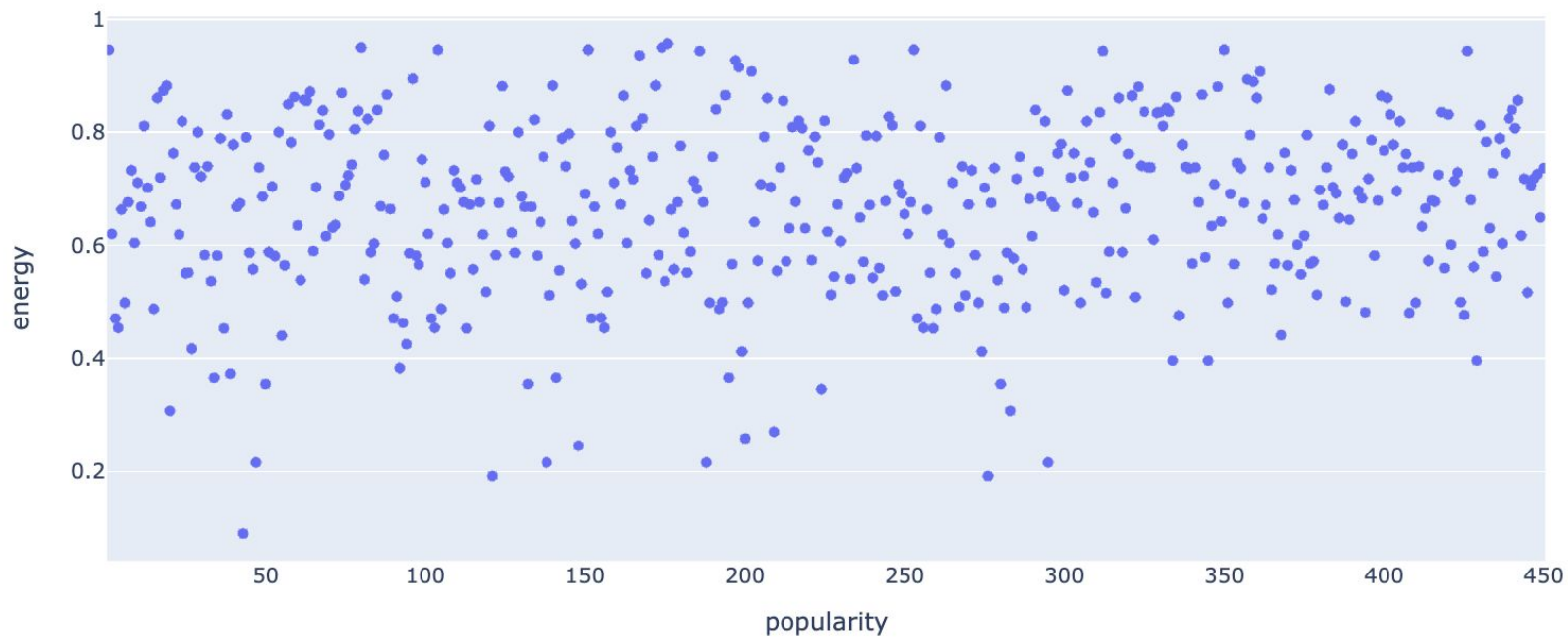
2:54



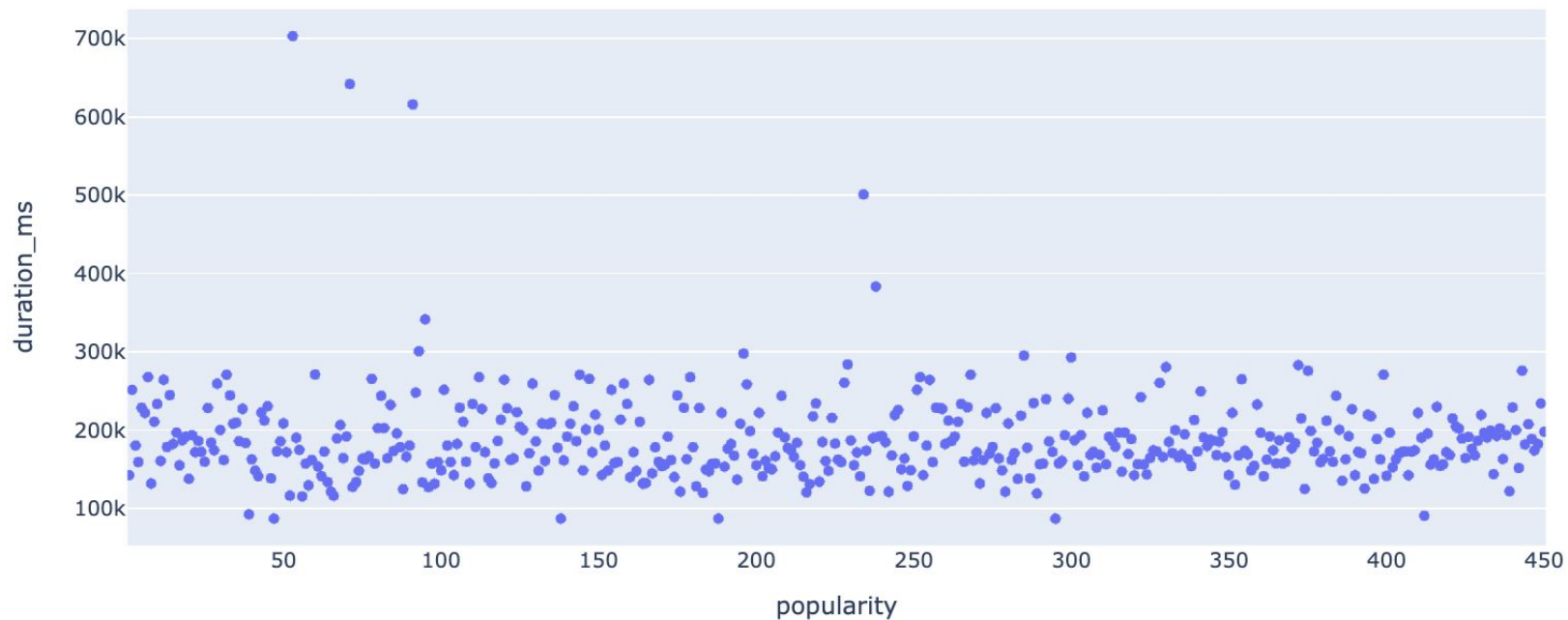
3:49



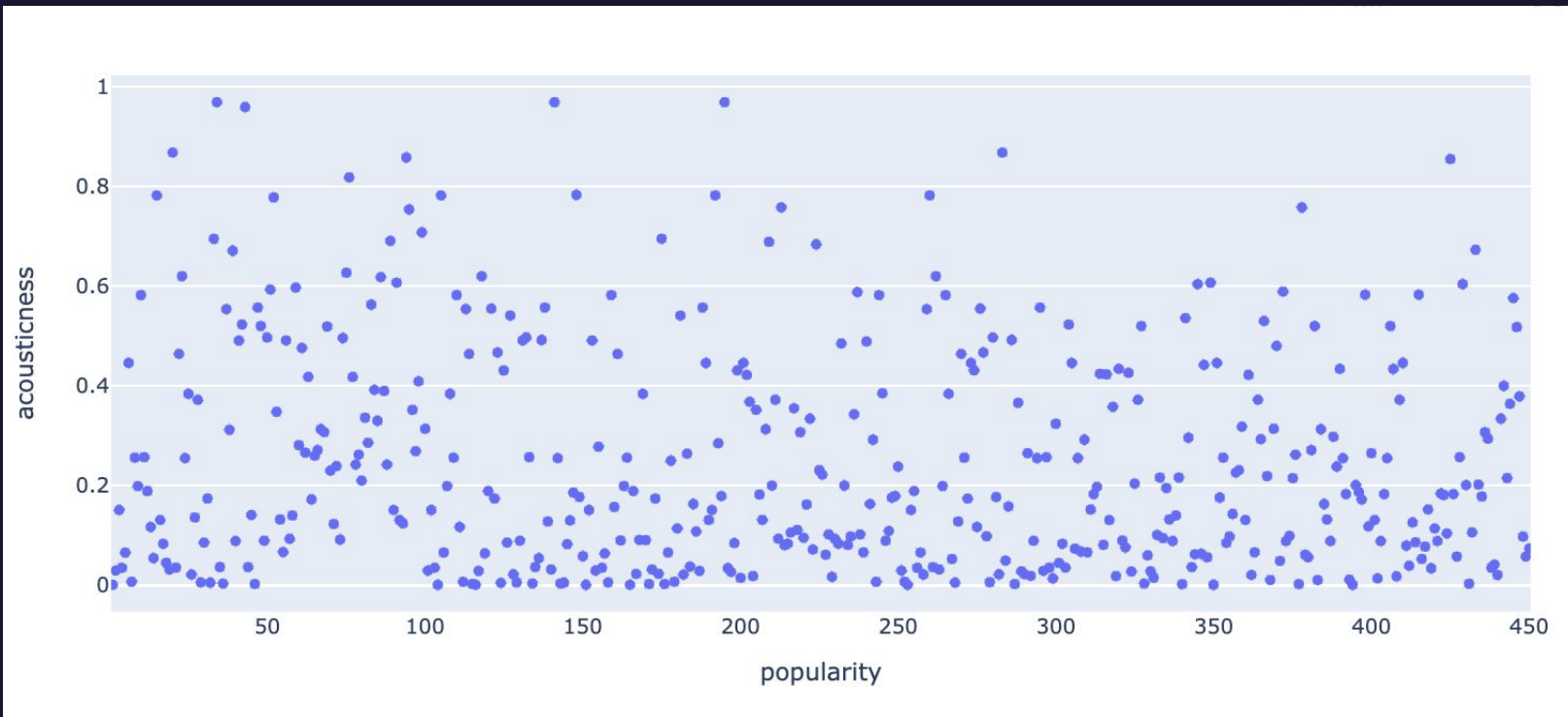
Dançabilidade X Popularidade



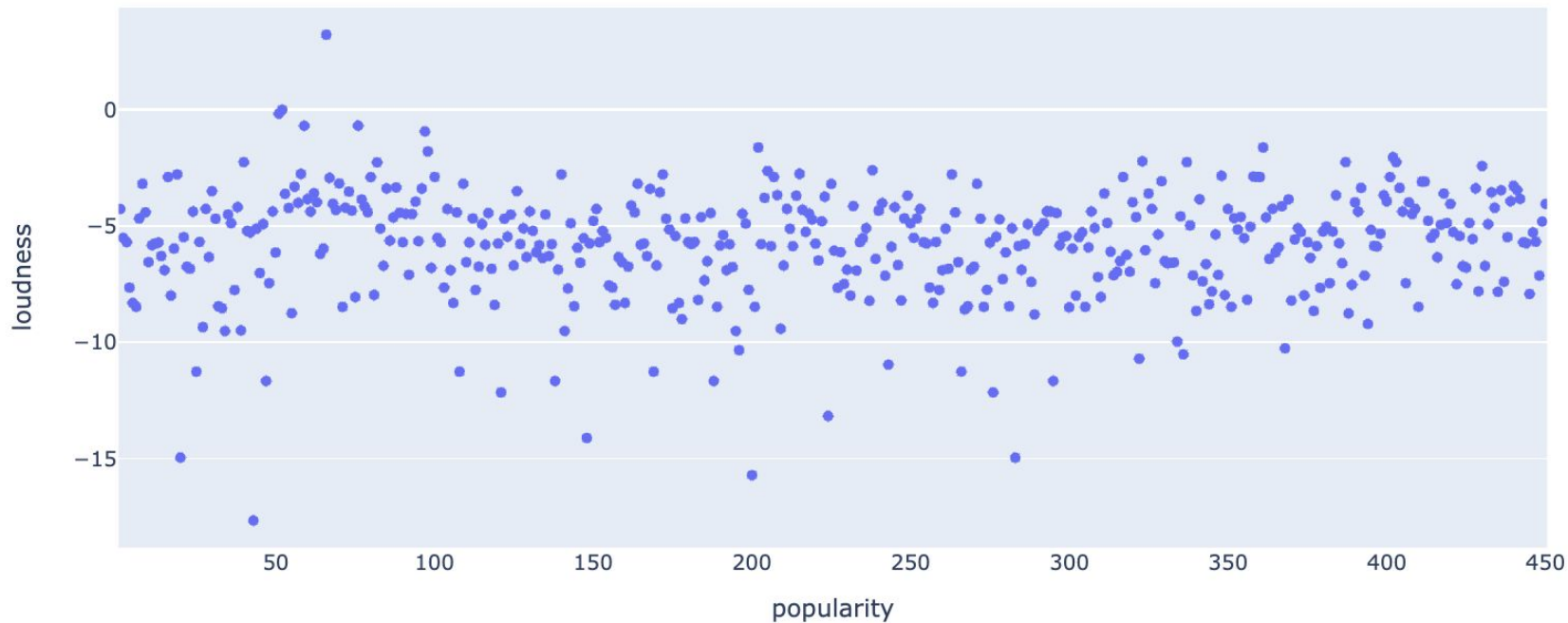
Energia X Popularidade



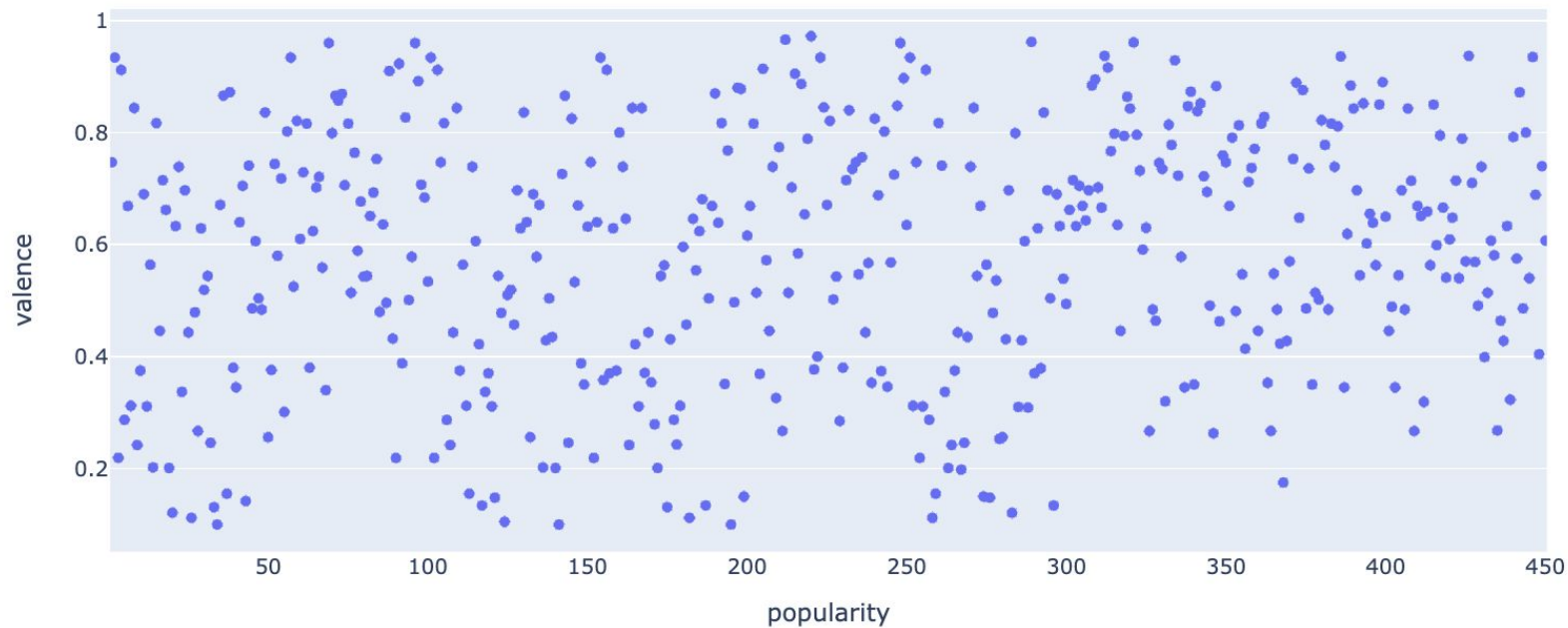
Duração da música X Popularidade
180k ms = 3 minutos



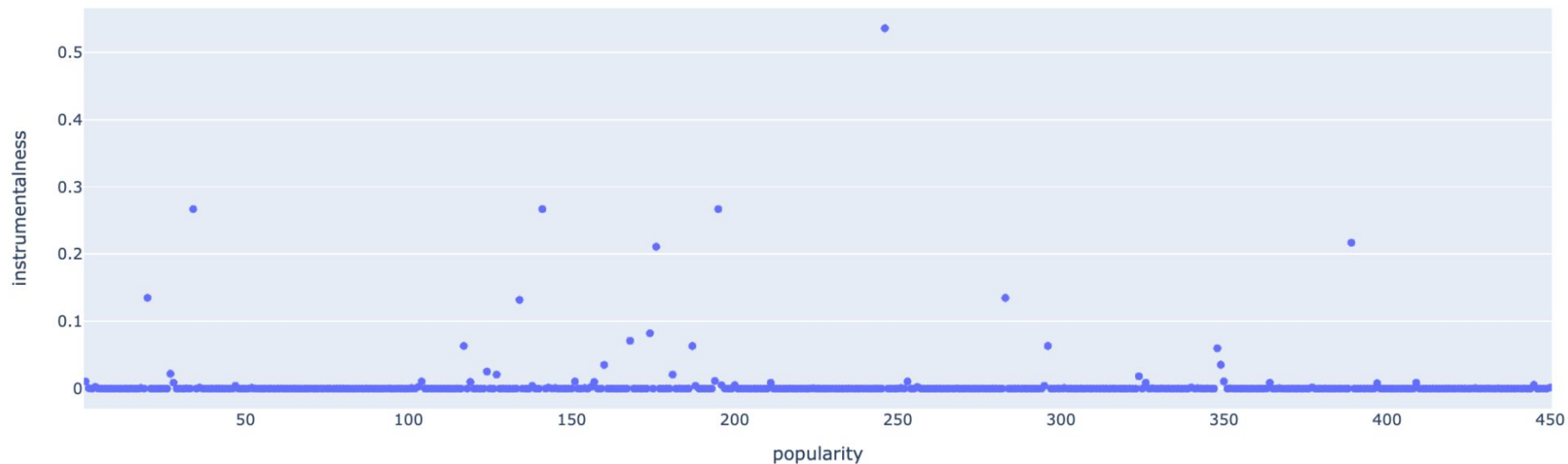
Acusticidade X Popularidade



Intensidade X Popularidade



Valência X Popularidade



Instrumental X Popularidade

1,0 : valores próximos significa maior probabilidade da faixa não conter conteúdo vocal. 0,5 : valores acima representam faixas instrumentais.



Medidas ideais

1. Dançabilidade e Energia: em torno de 0.7
2. Duração: Em torno de 180k ms (3 minutos)
3. Pouca acusticidade, preferível instrumentos eletrônicos
4. Muito falada
5. Sem ser ao vivo
6. Intensidade: Em torno de -5 dB
7. Fórmula do compasso 4



Mars Is a Cold Place
The 15th Planet

2:54



3:49



06

Criação da API

Com o Flask



Mars Is a Cold Place
The 15th Planet



2:54

3:49

```

# Retorna todas as músicas no banco de dados
# ex.: http://127.0.0.1:5000/all_playlists
@app.route("/all_playlists", methods = ['GET'])
def all_playlists():
    return jsonify(execute_query("SELECT DISTINCT playlist_country FROM top_songs"))

# Retorna todas as músicas de um país
# ex.: http://127.0.0.1:5000/playlist?country=Brazil
@app.route("/playlist", methods = ['GET'])
def country():
    country = request.args.get('country')
    return jsonify(execute_query(f"SELECT * FROM top_songs WHERE playlist_country='{country}'"))

# Retorna as músicas mais populares
# ex.: http://127.0.0.1:5000/popular
@app.route("/popular", methods = ['GET'])
def popular():
    return jsonify(execute_query(f"SELECT * FROM top_songs ORDER BY popularity DESC"))

# Retorna as músicas menos populares
# ex.: http://127.0.0.1:5000/unpopular
@app.route("/unpopular", methods = ['GET'])
def unpopular():
    return jsonify(execute_query(f"SELECT * FROM top_songs ORDER BY popularity ASC"))

# Retorna as músicas mais antigas
# ex.: http://127.0.0.1:5000/old_songs
@app.route("/old_songs", methods = ['GET'])
def old_songs():
    return jsonify(execute_query(f"SELECT * FROM top_songs ORDER BY release_date ASC"))

# Retorna as músicas mais recentes
# ex.: http://127.0.0.1:5000/new_songs
@app.route("/new_songs", methods = ['GET'])
def new_songs():
    return jsonify(execute_query(f"SELECT * FROM top_songs ORDER BY release_date DESC"))

```

Requisições que abrangem **todo o conjunto de músicas**. Ao contrário da API original, que é seccionada por playlists.



Mars Is a Cold Place
The 15th Planet

2:54



3:49



Modelos



Mars Is a Cold Place
The 15th Planet

2:54



3:49



01 Random Forest

Composto por um conjunto de árvores de decisão, onde cada árvore é construída de forma independente usando uma amostra aleatória do conjunto de dados de treinamento e um subconjunto aleatório das características. Isso introduz aleatoriedade e diversidade no processo de treinamento.

02 KNN

KNN é um algoritmo de aprendizado supervisionado que se baseia na proximidade dos exemplos de treinamento para tomada de decisões de classificação ou regressão

03 Naive Bayes(Gaussian)

Naive é uma técnica de classificação de classificação estatística baseada no teorema de Bayes. É um dos algoritmos de aprendizagem supervisionada.



Mars Is a Cold Place
The 15th Planet

2:54



3:49

Métricas



- **Mean Squared Error (MSE):**
Média dos quadrados das diferenças entre as previsões do modelo e os valores reais. Fornece uma medida da qualidade geral do modelo, onde valores menores indicam um melhor ajuste aos dados.
- **Mean Absolute Error (MAE):**
Média das diferenças absolutas entre as previsões do modelo e os valores reais. Ele mede a magnitude média dos erros do modelo, sem considerar sua direção.
- **R^2 (R-squared):**
O R^2 é uma medida da proporção da variância nos valores de resposta que é explicada pelo modelo. Ele varia de 0 a 1, onde valores mais próximos de 1 indicam um melhor ajuste do modelo aos dados.
- **Root Mean Squared Error (RMSE):**
Raiz quadrada do MSE e fornece uma interpretação na mesma unidade dos valores de destino. É uma medida comum de erro que penaliza mais fortemente grandes erros.



Mars Is a Cold Place
The 15th Planet

2:54



3:49

**RF****Knn****Gaussian****MSE**

22.1355

62.3538

116.9667

MAE

3.1534

6.3378

8.7000

R²

0.6678

0.0643

-0.7551

RMSE

4.7048

7.8964

10.8151

**Mars Is a Cold Place**
The 15th Planet

2:54

3:49



Obrigado!



Mars Is a Cold Place
The 15th Planet

2:54



3:49