



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

Instituto de Ciências Exatas e de Informática

Trabalho Sistemas Inteligentes: Ciência de Dados

Intelligent Systems Project: Data Science

Bárbara Luciano Araújo¹

Bernardo Souza Cardoso²

Erick Gonzaga Santos³

Frederico Malaquias Alves Caldeira⁴

Luisa Nogueira Campos Silva Souza⁵

Resumo

Este artigo foi criado para a disciplina de Introdução à Sistemas Inteligentes, do segundo período do curso de Ciência da Computação da PUC Minas. Dessa maneira, será abordado a identificação e descrição da base de dados escolhida pelo grupo. Além da definição do problema, análise de fatos e oportunidades, utilizando métodos estatísticos, como média, moda e mediana. Por fim, será apresentada a visualização de dados e a conclusão.

Palavras-chave: Sistemas Inteligentes. Base de dados.

Abstract

This present article was created for the subject of Introduction to Intelligent Systems, from the second period of Computer Science course, in PUC Minas. It will address the identification and description of the data base chosen by the group. Besides, it will present the problem definition, facts analyses and opportunities, utilizing statistics methods as arithmetic average, mode and median. Finally, it will approach the data visualization and the article conclusion.

Keywords: Intelligent Systems. Data base.

¹ Aluna de Graduação em Ciência da Computação, Brasil – barbara.araujo.1299141@sga.pucminas.br.

² Aluno de Graduação em Ciência da Computação, Brasil – bernardo.cardoso.1329186@sga.pucminas.br.

³ Aluno de Graduação em Ciência da Computação, Brasil – erick.gonzaga@sga.pucminas.br.

⁴ Aluno de Graduação em Ciência da Computação, Brasil – fmacaldeira@sga.pucminas.br.

⁵ Aluna de Graduação em Ciência da Computação, Brasil – lncssouza@sga.pucminas.br.

(PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS, 2022)

1 INTRODUÇÃO

1.1 Identificação

A base de dados escolhida foi da página de charts do Spotify, um aplicativo conhecido mundialmente utilizado para escutar músicas e podcasts de vários lugares do mundo. O charts consiste na divulgação semanal de uma lista de álbuns e músicas mais ouvidas nos EUA e ao redor do mundo. Essa lista é atualizada toda segunda-feira e postada na conta oficial do Spotify no Twitter.

1.2 Descrição dos Dados

A descrição dos dados, chamada de dicionário de dados (FIGURA 1), apresenta a descrição e explicação das variáveis da tabela de base de dados escolhida, sendo no caso a base de dados do SPOTIFY, como apresentado anteriormente.

Figura 1 – Dicionário de Dados

Variável	Nome Variável	Medidas	Valores	Descrição
título	title	caracteres	texto	título da música
artista	artist	caracteres	texto	nome do artista
genero	top genre	caracteres	texto	genero da música
ano	year	números	1900 - 2022	ano de criação
batidas por minuto	beats.per.minute	números	0 - 200	número de batidas da música
energia	energy	números	0 - 100	energia gasta na música
danceabilidade	danceability	números	0 - 100	quão dançável é a música
barulho	loudness.dB	números	-50 - +50	quantos decibéis a música produz
vivacidade	liveness	números	0 - 50	quão alegre é a música
balanço	balance	números	0 - 100	balanço da música
tamanho	length	números	0 - 500	tamanho da música em minutos
acústico	acousticness	números	0 - 100	quão acústica a música é
tonalidade	speechness	números	0 - 100	tom de voz da música (agudo, grave...)
popularidade	popularity	números	0 - 100	quão popular a música é

Fonte: OSF-Support (2022)

1.3 Exemplos

Figuras se encontram na página 4.

2 DEFINIÇÃO DO PROBLEMA

2.1 Pergunta

Qual ano foi o melhor na produção de músicas segundo a base de dados escolhida?

2.2 Explicação do problema

A pergunta gerada para analisar na base de dados do Spotify é qual foi o melhor ano para a música, de acordo com a tabela dos Top Hits Mundiais do Spotify. Para isso, será necessário avaliar e calcular quantas vezes cada ano apareceu na lista, buscando os dados da coluna "year".

3 ANÁLISE DOS DADOS

3.1 Tabela com cálculo de cada aparição de determinados anos na base de dados do Spotify

(FIGURA 5 - Pág. 5)

3.2 Tabela com o cálculo e com o resultado (grifado de vermelho) de cada aparição de determinados anos na base de dados do Spotify

(FIGURA 6 - Pág. 6)

3.3 Análise exploratória de dados

3.3.1 Média

Analisando o gráfico da relação dos anos, presente na figura 2, na página 4, e nas tabelas produzidas pelo grupo, podemos fazer o cálculo da média, que tem como valor 5,263157894736842, que arredondando resulta em 5,3.

Média: 5,3

3.3.2 Mediana

Para o cálculo da mediana, precisamos colocar os números em ordem crescente. Sendo mostrado abaixo:

0 0 0 0 0 1 1 1 1 3 3 3 4 6 12 13 16 16 20

Calculando a mediana:

0 0 0 0 0 1 1 1 1 **3** 3 3 4 6 12 13 16 16 20

Analisando o gráfico da relação dos anos, presente na figura 2, na página 4, podemos fazer o cálculo da mediana, que tem como valor 3.

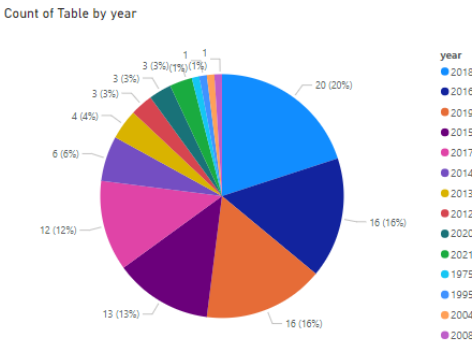
Média: 3

4 CONCLUSÃO

Com base nos dados apresentados e nos cálculos feitos pela equipe, foi possível identificar que o melhor ano, ou seja, aquele que mais apareceu na tabela das Melhores Músicas do Spotify, é o ano de 2018, com um total de 20 aparições. Essas aparições são com as músicas: Attention - Charlie Puth, Better Now - Post Malone, Cold Water - Major Lazer, Eastside (with Halsey Khalid) - benny blanco, Girls Like You (feat. Cardi B) - Cardi B Version - Maroon 5, God's Plan - Drake, Happier - Marshmello, Havana (feat. Young Thug) - Camila Cabello, I Like It - Cardi B, I Like Me Better - Lauv, lovely (with Khalid) - Billie Eilish, Lucid Dreams - Juice WRLD, One Kiss (with Dua Lipa) - Calvin Harris, Psycho (feat. Ty Dolla Sign) - Post Malone, rockstar (feat. 21 Savage) - Post Malone, SAD! - XXXTENTACION, Shallow - Lady Gaga, SICKO MODE - Travis Scott, Sunflower - Spider-Man: Into the Spider-Verse - Post Malone, Youngblood - 5 Seconds of Summer.

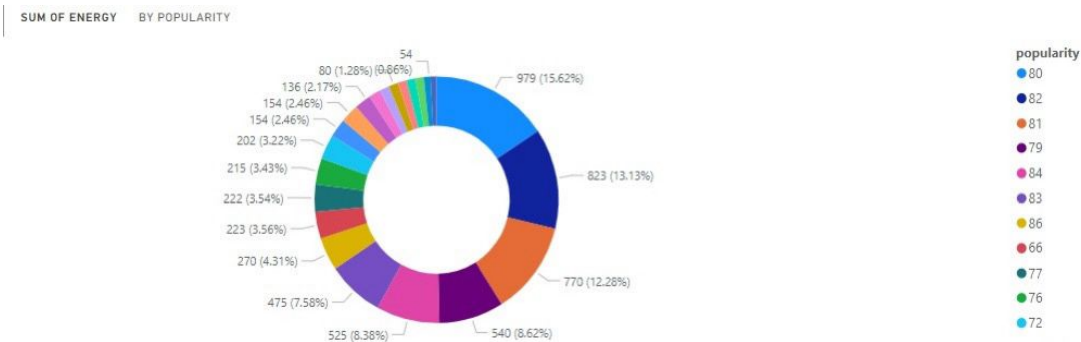
5 FIGURAS

Figura 2 – Contabilidade de anos



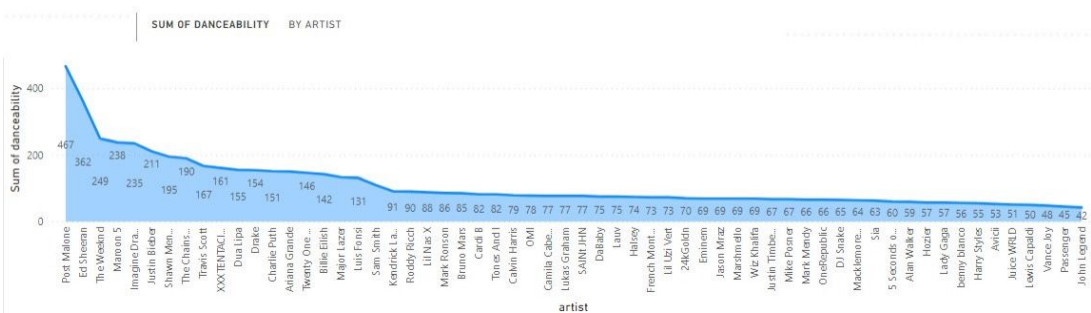
Fonte: Power-BI (2022)

Figura 3 – Somatório de Energia



Fonte: Power-BI (2022)

Figura 4 – Somatório de Danceabilidade



Fonte: Power-BI (2022)

Figura 5 – Tabela

Ano	Quantidade que aparece na tabela
1975	1
1980	0
1985	0
1990	0
1995	1
2000	0
2004	1
2008	1
2012	3
2013	4
2014	6
2015	13
2016	16
2017	12
2018	20
2019	16
2020	3
2021	3
2022	0

Fonte: Power-BI (2022)

Figura 6 – Tabela com resultado

Ano	Quantidade que aparece na tabela	
1975	1	
1980	0	
1985	0	
1990	0	
1995	1	
2000	0	
2004	1	
2008	1	
2012	3	
2013	4	
2014	6	
2015	13	
2016	16	
2017	12	
2018	20	
2019	16	
2020	3	
2021	3	
2022	0	

Fonte: Power-BI (2022)

REFERÊNCIAS

OSF-SUPPORT. **How to Make a Data Dictionary**. 2022. Disponível em: <<https://help.osf.io/article/217-how-to-make-a-data-dictionary>>.

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS. **Padrão PUC Minas de Normalização**: normas da ABNT para apresentação de teses, dissertações, monografias e trabalhos acadêmicos. 9. ed. rev. ampl. atual. Belo Horizonte: PUC Minas, 2022. Disponível em: <<http://www.pucminas.br/biblioteca/>>. Acesso em: 6 de set. 2013.

POWER-BI. **Visualização de Dados**. 2022. Disponível em: <<https://app.powerbi.com/groups/me/reports/2a7a377b-63c3-4162-829d-40f6160d2de5/ReportSection>>.