

Uma análise dos dados do Grammy Awards de 1958 a 2019

Beatriz Gonçalves Bento

Renato Silva de Assis

Ciências de Dados para negócios

Disciplina: Banco de Dados

Professor: Jorge Dias

Este trabalho apresenta uma análise detalhada dos dados do Grammy Awards de 1958 a 2019, aplicando conceitos de banco de dados para organizar e interpretar informações sobre as premiações ao longo dos anos.



Objetivo do Trabalho

Aplicação de Conceitos

Aplicar os conceitos de banco de dados estudados durante a disciplina, passando por todas as etapas de desenvolvimento de um sistema de banco de dados.

Análise de Dados

Utilizar como objeto de estudo a base de dados do Grammy Awards disponível no Kaggle, contendo informações detalhadas sobre edições do Grammy.

Implementação e Apresentação

Desenvolver desde a análise de dados brutos até a implementação final e apresentação de resultados.



Informações sobre o Banco de Dados Bruto



1

Edições Anuais

Informações sobre cada edição do evento, como o ano e o título (exemplo: "62nd Annual GRAMMY Awards (2019)").

2

Categorias de Prêmios

Cada categoria de premiação do Grammy, como "Record of the Year", "Album of the Year", "Best New Artist", entre outras.

3

Nomeados e Vencedores

Os nomeados e vencedores em cada categoria, além de informações sobre os profissionais envolvidos na produção das músicas.

4

Obras Premiadas

Detalhes sobre as músicas e álbuns indicados, como título da música, nome do álbum e nomes dos profissionais que participaram da sua criação.



Metodologia – Ferramentas Utilizadas



MySQL

Sistema gerenciador de banco de dados (SGBD) para implementar o modelo físico do banco de dados.



DBeaver

Interface gráfica para gerenciar o banco de dados MySQL, possibilitando a criação de tabelas e execução de scripts SQL.



VS Code

Editor de código escolhido para o desenvolvimento do script Python e escrita de scripts SQL.



Python

Linguagem utilizada para criar uma aplicação capaz de se conectar ao banco de dados, extrair informações e apresentar os resultados.

Etapas do Projeto

1

Análise e Normalização dos Dados

A partir dos dados brutos, foi realizado o processo de normalização para eliminar redundâncias e organizar as informações em um modelo E-R com pelo menos cinco tabelas.

2

Implementação do Banco de Dados

Implementamos o banco de dados normalizado no MySQL, utilizando o DBeaver como ferramenta para gerenciar a criação de tabelas e consultas SQL.

3

População do Banco de Dados

Desenvolvemos um script em Python para popular o banco de dados com os dados brutos normalizados.

4

Desenvolvimento da Aplicação

Criamos uma aplicação Python que se conecta ao banco de dados e apresenta os dados de forma visual (gráficos e tabelas), utilizando consultas SQL para extrair as informações.



Decisões Sobre o Banco de Dados

Remoção da Variável Winner

Identificamos que a variável Winner estava com o valor "True" para todos os registros, o que indica que ela não traria nenhuma informação relevante ou diferenciada para a nossa pesquisa. Por isso, decidimos removê-la do modelo.

Remoção da Variável img

A coluna img continha links para imagens, que não eram relevantes para os objetivos do nosso projeto, que está focado na análise de informações textuais e numéricas. Assim, essa variável foi excluída do banco de dados.

Remoção das Variáveis published_at e updated_at

Essas variáveis indicavam datas de publicação e atualização dos dados, que não apresentavam valor significativo para os objetivos do estudo. Optamos por retirá-las para simplificar o modelo e focar nas informações essenciais.



Diário de Bordo – Desafios Enfrentados

1 Definição de Variáveis para Remoção

Um dos primeiros passos foi analisar quais variáveis eram relevantes para o projeto e quais poderiam ser removidas.

2 Qualidade dos Dados Brutos

O banco de dados original apresentava uma estrutura confusa e desorganizada, com dados sujos e muitas redundâncias.

3 Confusão entre Nomeados e Vencedores

O dataset original rotulava os vencedores como "nomeados" (nominee), o que gerou confusão na interpretação dos dados.

4 Limpeza da Tabela Workers

A tabela workers apresentava uma grande variedade na forma de inserção dos dados, com inconsistências na identificação de pessoas e profissões.

Banco de Dados I

Alunos:
Beatriz Gonçalves Bento
Renato Silva de Assis

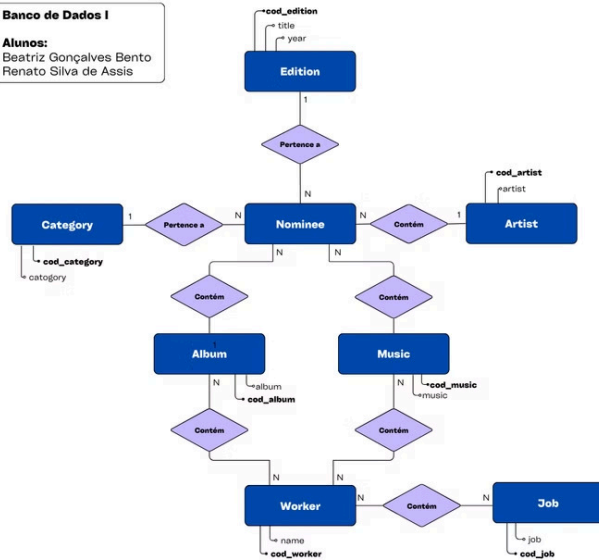


Tabela Original:

The_Grammy_Awards (year, title, published_at, updated_at, category, nominee, artist, workers, img, winner)

Observação: as colunas img e winner foram eliminadas, já que a primeira contém links sem acesso, e a segunda contém valores idênticos em todas as linhas (True) já que a tabela só possui ganhadores.

Após normalização:

Edition (cod_Edition, year, title)

Category (cod_Category, category)

Worker (cod_worker, name, job, cod_music, cod_album)

Job (cod_job, job)

Artist (cod_artist, artist)

Music (cod_music, music)

Album (cod_album, album)

WorkerMusicAlbum (cod_music, cod_album, cod_worker)

cod_music referencia Music

cod_album referencia Album

cod_worker referencia Worker

WorkerJob (cod_worker, cod_job)

cod_worker referencia Worker

cod_job referencia Job

Nominee (cod_artist, cod_music, cod_edition, cod_category, cod_album)

cod_artist referencia Artist

cod_music referencia Music

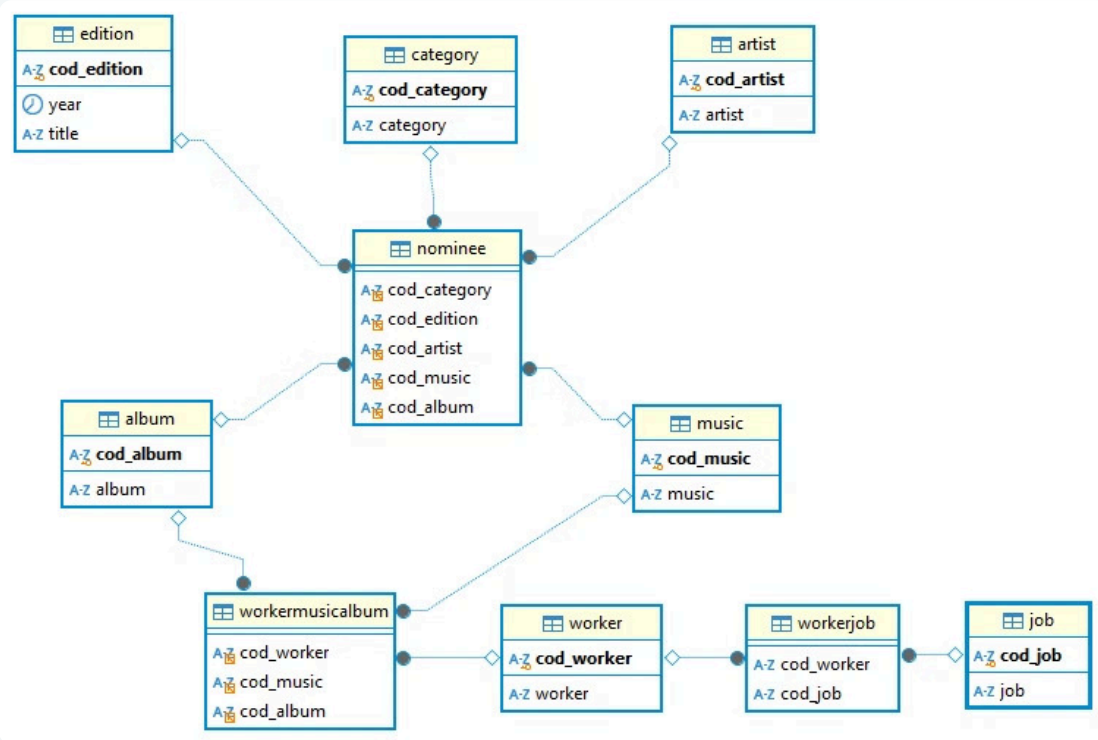
cod_edition referencia Edition

cod_category referencia Category

cod_album referencia Album

Modelo Conceitual

O modelo conceitual representa as entidades e relacionamentos do banco de dados do Grammy Awards, incluindo artistas, álbuns, músicas, categorias e edições.



Modelo Lógico

O modelo lógico detalha as tabelas, campos e relações entre as entidades do banco de dados do Grammy Awards, preparando a estrutura para implementação no SGBD.

Modelo Físico

1

WorkerMusicAlbum

Relaciona trabalhadores, músicas e álbuns.

3

Artist

Armazena dados dos artistas.

5

Edition

Guarda informações sobre as edições do Grammy.

7

Music

Armazena os dados das músicas nomeadas.

9

WorkerJob

Relaciona trabalhador com profissão

2

Album

Contém informações sobre os álbuns.

4

Category

Registra as categorias de premiação.

6

Job

Define os diferentes tipos de trabalho envolvidos.

8

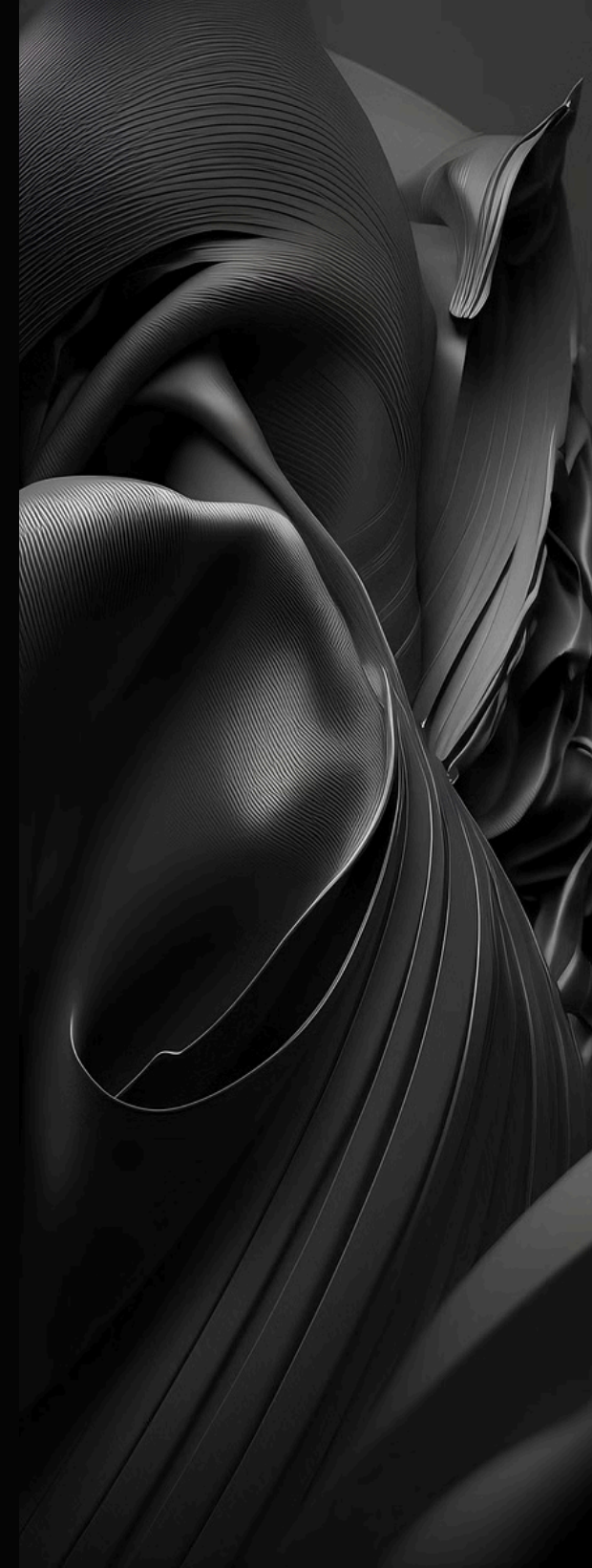
Nominee

Relaciona vencedores com categoria, edição, artistas e músicas

10

Worker

Informações sobre os trabalhadores





Implementação no SGBD

1

Criação de Tabelas

Todas as tabelas foram criadas com suas respectivas chaves primárias e estrangeiras.

2

Estabelecimento de Relações

Foram estabelecidas as relações necessárias entre as entidades do banco de dados.

3

Criação de Views

Foram criadas views para facilitar a consulta e análise dos dados.

4

Desenvolvimento de Rankings

Foram criados rankings para permitir uma visão aprofundada sobre os vencedores e categorias de premiação ao longo dos anos.

Views Criadas

1 View de Artistas Nomeados por Categoria

Permite visualizar os artistas nomeados em cada categoria de premiação.

2 View de Trabalhos e Suas Funções

Mostra os diferentes trabalhos e funções associados às premiações.

3 View de Músicas por Edição

Apresenta as músicas premiadas em cada edição do Grammy.

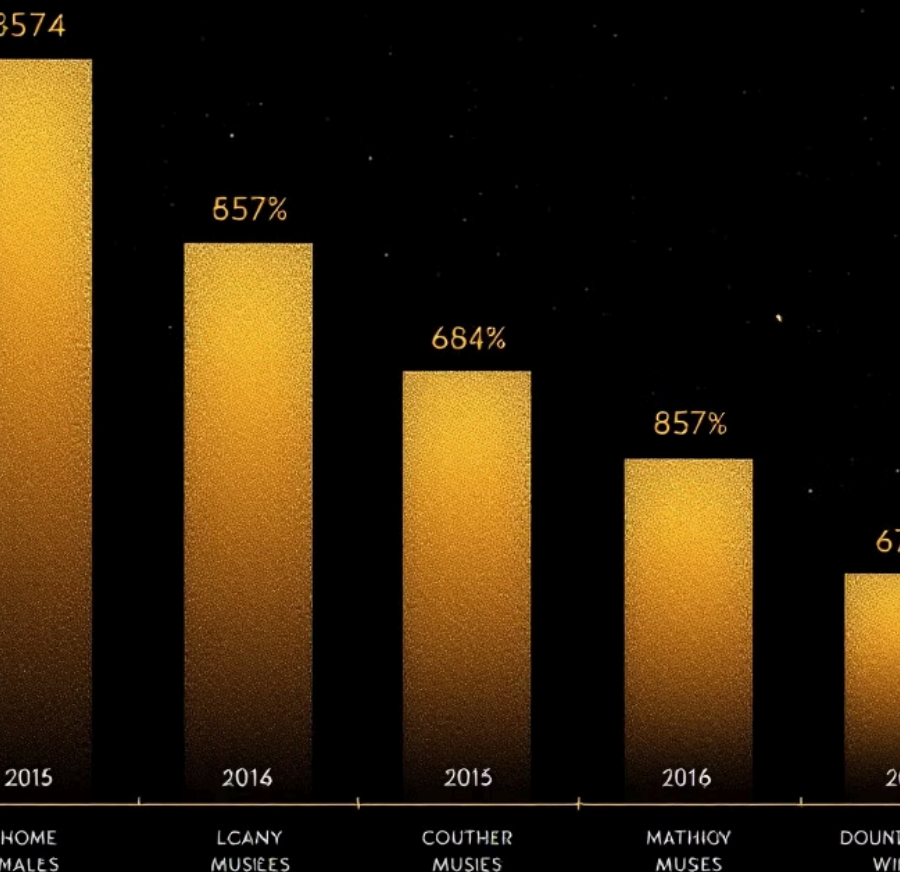
4 View de Trabalhadores e Funções por Categoria de prêmio

Exibe os profissionais e suas funções em cada categoria de premiação.



TOP GRAMMY WINS

THE MOST GRAMMY OUR ATISIES



20212

SELARIT AN



GLOBAL GRAMMYS
FOUNDATION

Rankings Desenvolvidos

1

Top 10 Artistas que mais venceram em 2019

Ranking dos artistas com mais vitórias na edição de 2019 do Grammy.

2

Artistas que mais venceram na história do Grammy

Lista dos artistas com maior número de vitórias em todas as edições do Grammy.

3

Artistas que mais venceram em categoria Gospel

Ranking dos artistas com mais vitórias na categoria Gospel.

4

Artistas que mais venceram em categoria Rock

Lista dos artistas com maior número de vitórias na categoria Rock.

5

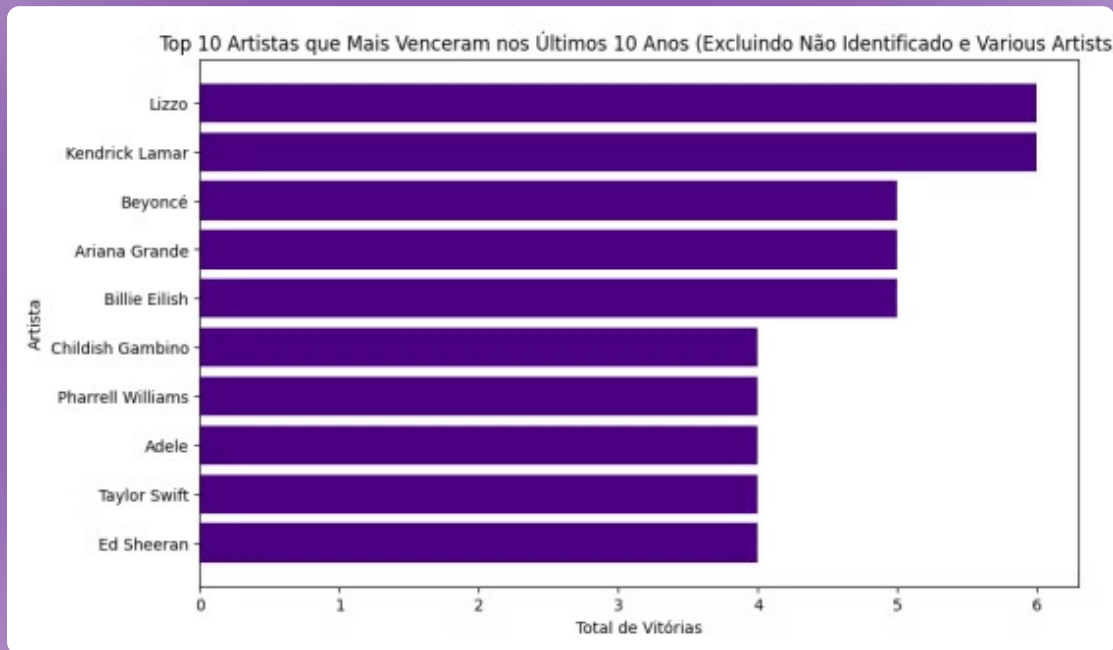
Ver resultados no DBeaver


```
import mysql.connector
import pandas as pd
import matplotlib.pyplot as plt

# Conexão com o banco de dados
conn = mysql.connector.connect(
    host='localhost',
    user='novo_usuario',
    password='Rairocha123@',
    database='projeto_banco_de_dados'
)
```

Integração do MySQL com Python

O código acima demonstra a integração entre MySQL e Python, utilizando as bibliotecas mysql-connector-python para conexão com o banco de dados e pandas para manipulação dos dados extraídos.

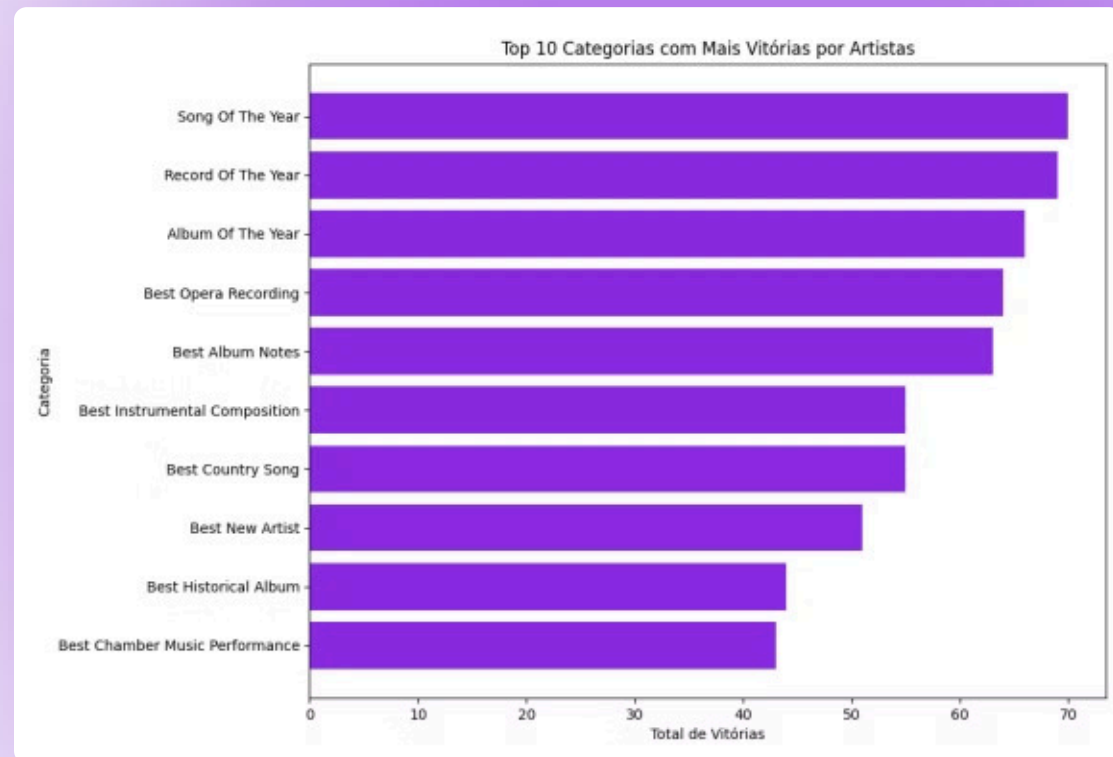


Visualização dos Resultados – Exemplo 1

Gráfico de Barras para o Top 10 Artistas que Mais Venceram nos Últimos 10 Anos, mostrando a dominância de artistas como Lizzo, Kendrick Lamar e Beyoncé no período recente.

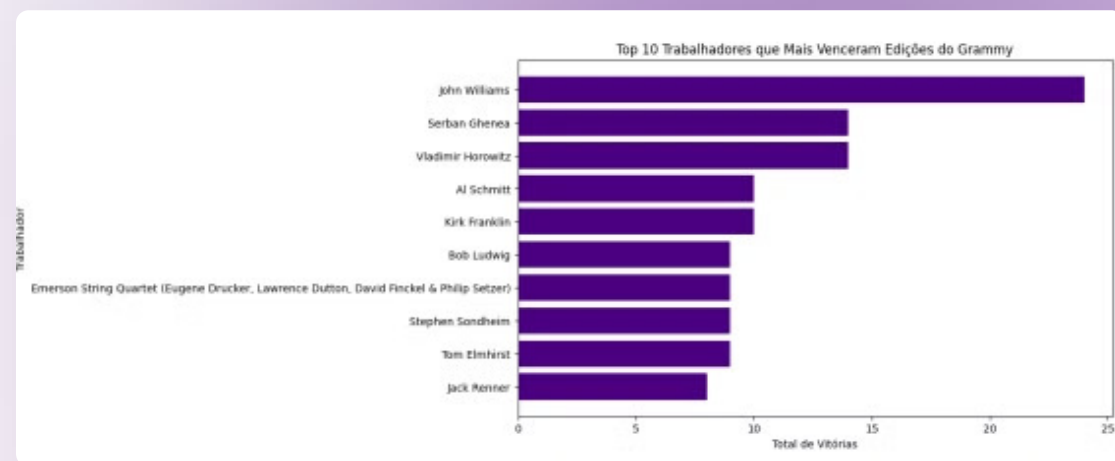
Visualização dos Resultados – Exemplo 2

Gráfico de 10 Categorias com Mais Vitórias por Artistas, destacando as categorias mais competitivas e prestigiadas do Grammy Awards.

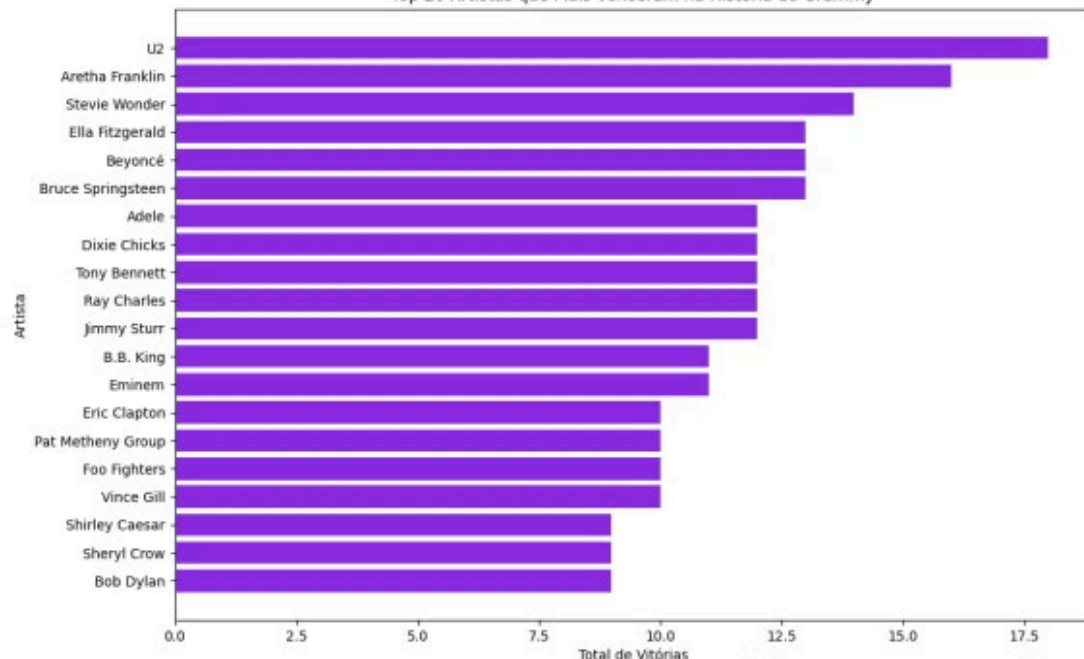


Visualização dos Resultados – Exemplo 3

Gráfico de Barras para os 10 Trabalhadores que Mais Venceram Edições do Grammy, evidenciando a importância de profissionais como John Williams e Serban Ghenea na indústria musical.



Top 20 Artistas que Mais Venceram na História do Grammy



Visualização dos Resultados – Exemplo 4

Gráfico de Artistas com Mais Vitórias na História do Grammy, mostrando a dominância histórica de artistas como U2, Aretha Franklin e Stevie Wonder.

Conclusão

Análise Aprofundada

O projeto permitiu uma análise detalhada das premiações do Grammy ao longo dos anos, revelando tendências e padrões importantes.

Estrutura Sólida

A implementação do modelo conceitual, lógico e físico no MySQL, seguida pela integração com Python, proporcionou uma base robusta para consultas e visualizações avançadas.

Insights Visuais

A geração de gráficos foi fundamental para transformar os dados em insights visuais, facilitando a compreensão e a comunicação dos resultados.

Relevância do Banco de Dados

O projeto demonstrou a importância de um banco de dados bem estruturado para organizar e interpretar grandes volumes de dados históricos.

