# PROJETO SQL - ANÁLISE DE DADOS DE UMA LOJA DE MÚSICA

## Estrutura do projeto

1. **Criar uma pasta no computador ou na nuvem:** Projeto\_LojaDeMusica √
2. **Criar um repositório GitHub:** Projetos\_SQL √
3. **Criar um arquivo dos tópicos para resolver os problemas de negócio**: Os problemas serão explicados através do arquivo README. √
4. **Criar um arquivo com as definições do planejamento do projeto:** 
   1. **Quais serão as fontes de dados para encontrar as respostas?**

Os dados da Loja de Música Chinook foram baixados através do DATASET do KAGGLE pelo link: <https://www.kaggle.com/datasets/samaxtech/chinook-music-store-data>. O arquivo CHINOOK.DB foi analisado, através do site <https://sqliteonline.com/> e importados para o Microsoft SQL Server manualmente e separados por tabelas. Os scripts em SQL para as criações das tabelas foram salvos na pasta DATASET junto com o arquivo CHINOOK.DB.

* 1. **Quais serão as ferramentas utilizadas para encontrar as respostas?**

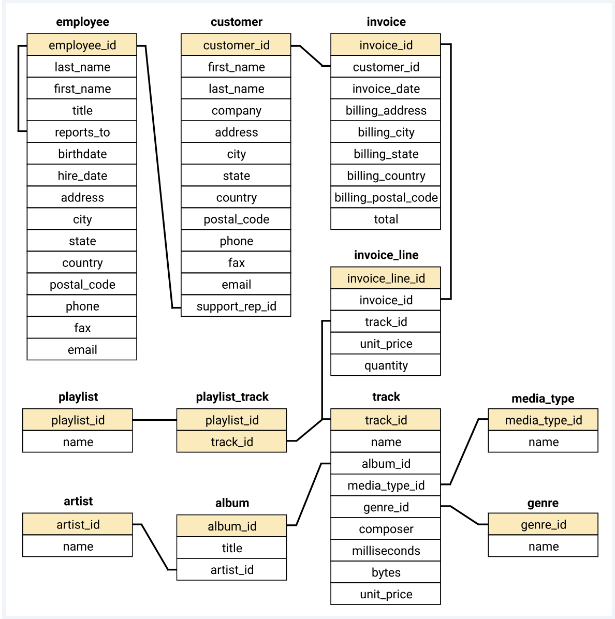
Foram utilizadas as seguintes ferramentas: Microsoft SQL Server e ChatGPT √

* 1. **Como será o Processo?**
  2. **Qual será o Produto Final?**

1. **Criar as subpastas que serão utilizadas no projeto:** 
   1. **Pasta Dados (DATASETS) √**
   2. **Pasta Relatórios (REPORTS)** √
   3. **Pasta Referência ou Documentos (REFERENCES ou DOCS)** √

## Escopo do projeto

O banco de dados Chinook modela um armazenamento de mídia digital fictício. O banco de dados mantém um inventário e rastreia as vendas de uma pequena loja de música. Há dois aspectos principais cobertos no banco de dados: Inventário e Faturamento.



## Dicionário dos dados

* **Tabela Empregados ([employee])** - Armazena os dados dos funcionários, tendo como colunas: ID do funcionário: [employee\_id], Nome: [first\_name], Sobrenome: [last\_name], Título do cargo: [title], Especifica quem se reporta a quem: [reports\_to], Níveis: [levels], Data de nascimento: [birthdate], Data de contratação: [hire\_date], Endereço: [address], Cidade: [city], Estado: [state], País [country], Código postal: [postal\_code], Telefone: [phone], Fax: [fax] e E-mail: [email]
* **Tabela Clientes ([customer])** - Armazena os dados dos clientes, tendo como colunas: ID do cliente: [customer\_id], Nome: [first\_name], Sobrenome: [last\_name], Empresa: [company], Endereço: [address], Cidade: [city], Estado: [state], País: [country], Código postal: [postal\_code], Telefone: [phone], Fax: [fax], Email: [email], ID do representante de suporte : [support\_rep\_id]
* **Tabela Fatura ([invoice])** - A tabela Fatura armazena os dados do cabeçalho da fatura, tendo como colunas: ID dos itens da fatura: [invoice\_line\_id], ID da Fatura: [invoice\_id], ID da Trilha: [track\_id], Preço unitário: [unit\_price] e Quantidade: [quantity]
* **Tabela Itens da Fatura ([invoice\_line]) -** Armazena os dados dos itens de linha da fatura, tendo como colunas: ID da Fatura: [invoice\_id], ID do Cliente: [customer\_id], Data da fatura: [invoice\_date], Endereço da cobrança: [billing\_address], Cidade de cobrança : [billing\_city], Estado da cobrança: [billing\_state], País da cobrança: [billing\_country], Código postal da cobrança: [billing\_postal\_code] e Total da Fatura: [total]
* **Tabela Artista ([artist])** - Armazena os dados dos artistas, tendo como colunas: ID do artista: [artist\_id] e o Nome: [name]
* **Tabela Album ([album])** - Armazena os dados sobre uma lista de faixas, tendo como colunas: ID do album: [album\_id], Título: [title] e o ID do artista: [artist\_id]. Cada álbum pertence a um artista. No entanto, um artista pode ter vários álbuns.
* **Tabela Tipos de Mídia ([media\_type])** - Armazena os tipos de mídia, tendo como colunas: ID do tipo de míedia: [media\_type\_id] e o Nome: [name]
* **Tabela Gênero ([genre])** - Armazena os tipos de música, tendo como colunas: ID do genero: [genre\_id] e o Nome do gênero: [name]
* **Tabela Faixa ([track])** - Armazena os dados das faixas/músicas disponíveis na loja digital, tendo como colunas: ID da Faixa: [track\_id], Nome: [name], ID do Album: [album\_id], ID do tipo de mídia: [media\_type\_id], ID dos gêneros: [genre\_id], Nome do compositor: [composer], Duração em milisegundos: [milliseconds], tamanho em bytes: [bytes] e Preço unitário: [unit\_price]. Cada faixa pertence a um álbum.
* **Tabela Listas de Reprodução ([playlist])** - Armazena os dados sobre as listas de reprodução, tendo como colunas: ID da lista de reprodução: [playlist\_id] e Nome: [name]. Cada lista de reprodução contém uma lista de faixas. Cada faixa pode pertencer a várias listas de reprodução. A relação entre a tabela de listas de reprodução e a tabela de faixas ([track]) é de muitos para muitos. A tabela **Faixa da Lista de Reprodução ([playlist\_track])** é usada para refletir esse relacionamento., tendo como colunas: ID da lista de reprodução: [playlist\_id] e o ID da Faixa: [track\_id]

**# Análise de dados de vendas de uma loja**

Este é um projeto de análise de dados que tem como objetivo explorar os dados de vendas de uma loja e obter insights relevantes que possam ajudar a melhorar o desempenho da empresa.

**## Dados**

Os dados utilizados neste projeto foram coletados ao longo de um ano e estão disponíveis no arquivo `vendas.csv`. O conjunto de dados contém as seguintes colunas:

- `data`: data da venda.

- `produto`: nome do produto vendido.

- `quantidade`: quantidade vendida.

- `preco\_unitario`: preço unitário do produto.

- `total\_venda`: valor total da venda.

**## Análise exploratória**

Antes de iniciar a análise, fizemos uma análise exploratória dos dados para entender melhor as características do conjunto de dados. Alguns dos insights que obtivemos incluem:

- A maioria das vendas ocorreu nos meses de dezembro e janeiro, o que sugere que a loja pode ter uma forte dependência das vendas de fim de ano.

- A maioria dos produtos vendidos tem preços entre R$ 100 e R$ 500, o que sugere que a loja pode estar perdendo oportunidades de vendas de produtos mais caros ou mais baratos.

- A maioria das vendas foi de produtos de uma categoria específica, o que sugere que a loja pode estar subutilizando o potencial de outras categorias.

**## Análise de vendas por categoria**

Com base nos insights obtidos na análise exploratória, decidimos analisar as vendas por categoria para entender melhor como a loja está performando em cada uma delas. Alguns dos insights que obtivemos incluem:

- A categoria mais vendida é a de eletrônicos, seguida de perto pela categoria de móveis.

- A categoria de brinquedos é a que apresentou o menor desempenho em termos de vendas.

**## Conclusão**

Com base na análise realizada, recomendamos que a loja:

- Invista em outras categorias além da de eletrônicos e móveis para diversificar sua base de vendas.

- Considere a possibilidade de oferecer produtos com preços mais altos ou mais baixos para atender a uma gama mais ampla de clientes.

- Analise com mais detalhes o desempenho da categoria de brinquedos e busque identificar possíveis oportunidades de melhoria.

**## Tecnologias utilizadas**

- Python 3.8

- Jupyter Notebook

- Pandas

- Matplotlib

**## Como executar o projeto**

1. Clone este repositório.

2. Instale as dependências do projeto com o comando `pip install -r requirements.txt`.

3. Execute o arquivo `analise\_vendas.ipynb` no Jupyter Notebook.