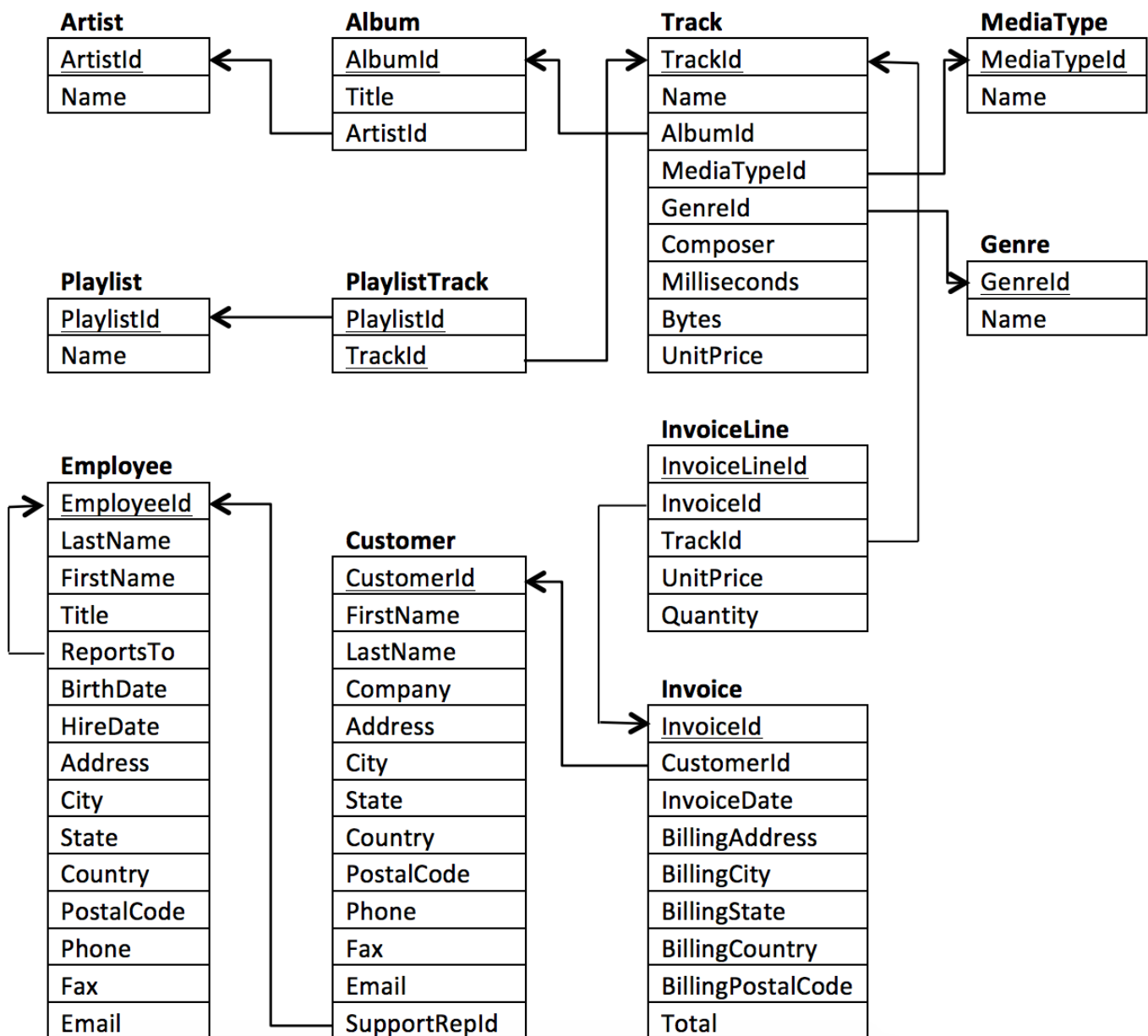


# Visão geral do projeto

## Banco de dados Chinook


### Introdução

Neste projeto você irá consultar o banco de dados Chinook. O banco de dados Chinook armazena informações sobre uma loja de música. Para este projeto, você ajudará a equipe Chinook a compreender a mídia em sua loja, seus clientes e funcionários e suas informações de fatura. Para ajudá-lo nas consultas a seguir, o esquema para o banco de dados Chinook é dado abaixo. Você pode ver as colunas que **conectam** tabelas usando as setas.



Todas as instruções abaixo são discutidas em detalhe conforme trabalhamos nesta aula para concluir este projeto. Abaixo temos uma rápida referência daquilo que você fará conforme avança pelo projeto.

## Instruções

- Você precisará seguir as instruções nos próximos três tópicos para conseguir deixar o banco de dados Chinook em pleno funcionamento em sua própria máquina e verificar se ele está configurado corretamente. Haverá duas partes para este projeto.
1. A primeira parte é uma série de perguntas que vão assegurar que você dominou os tópicos principais ensinados pelas aulas sobre SQL. **Embora essas perguntas não vão receber uma nota de um revisor, elas ajudarão você a se auto-avaliar.**
  2. A segunda parte é uma **apresentação**. De maneira semelhante ao primeiro projeto, não há uma 'resposta correta' para a segunda porção do projeto. Você tem a habilidade de ser criativo(a) nas perguntas que você fizer. Em seguida, você irá escrever uma consulta SQL para extrair os dados necessários para responder com êxito a sua pergunta. Use os dados puxados para construir um gráfico (gráfico de barras, histograma ou outro plot) que responda à sua pergunta. O essencial no envio de seu projeto será discutido nas próximas seções. **Para que sua apresentação seja revista, você precisará salvar seus slides em .**

## Passo a passo do projeto

Este vídeo mostra um dos nossos instrutores dando algumas grandes dicas e sugestões sobre como começar seu projeto: [Vídeo de passo a passo](#)

# Vamos configurar seu ambiente local

Está tudo bem começar a aprender SQL aqui na sala de aula, mas a maneira de realmente dominar suas habilidades é fazer uma configuração local e aprender a trabalhar no seu próprio ambiente.

Os próximos problemas vão ajudar a assegurar que você está à vontade em trabalhar localmente. Quando você tiver configurado, você será capaz de usar esse espaço de trabalho não só para seu projeto... mas MUITO ALÉM!

O ambiente que usaremos vai ser bem rápido de configurar e, com sorte, você estará em pleno funcionamento rapidinho :)

Tudo o que você precisa fazer aqui é...

- Baixar seu novo banco de dados
  - Baixar DB Browser para SQLite
- Assim que você tiver DB Browser para SQLite conectado ao seu novo banco de dados, você está pronto(a) para a próxima página!

## Baixar DB Browser para SQLite

Há muitos navegadores de bancos de dados que funcionam com diferentes tipos de bancos de dados. Para esse curso, usaremos o DB Browser para SQLite. Os outros navegadores que você pode usar provavelmente serão muito similares a esse.

DB Browser para SQLite pode ser baixado aqui: <http://sqlitebrowser.org/>

**DB Browser for SQLite**

The Official home of the DB Browser for SQLite

View project on GitHub

### News

- 2016-12-17 - The v3.9.1 binary for OSX has been rebuilt using Qt 5.7.1, to fix an important colour display problem on macOS Sierra.
- 2016-12-15 - An initial DBHub.io server is online, running our latest development code. Testing and feedback is encouraged.
- Note - The data on this server will probably be wiped/reset every few days.
- 2016-10-20 - PortableApp version of DB4S 3.9.1 now available. Thanks John! :)

### Screenshot

SQLite Database Browser - /Users/jc/tmp/example.db

New Database Open Database Write Changes Revert Changes

Database Structure Browse Data Edit Pragmas Execute SQL

Download 32-bit Windows .exe

Download 64-bit Windows .exe

Download PortableApp

Download Mac .dmg

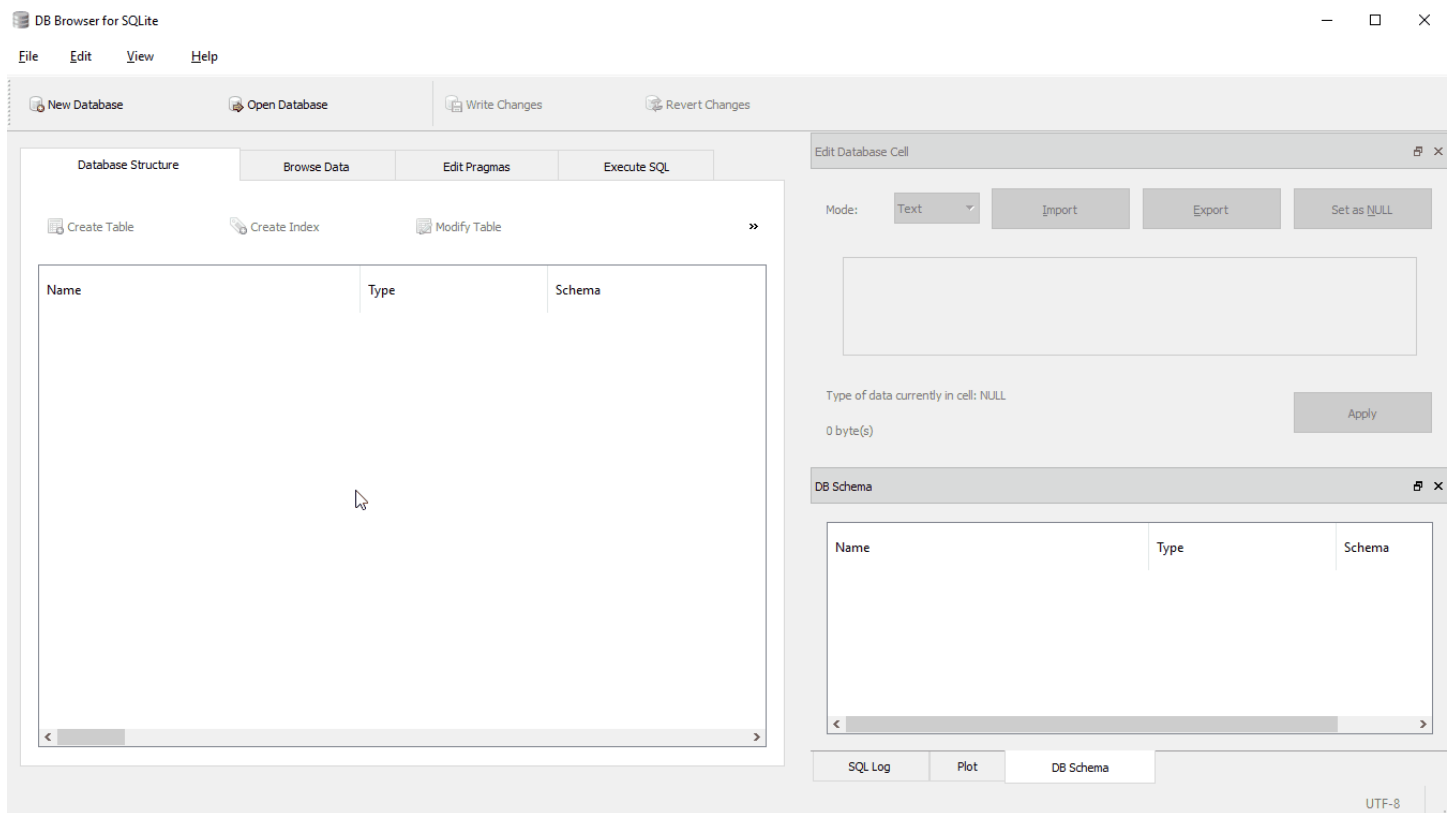
## Baixar banco de dados

Você pode baixar o banco de dados que usaremos para esse projeto na parte inferior da página.

## Conectar o navegador ao banco de dados

Aqui estão os passos:

- Abra o DB Browser para SQLite
- Clique em *Open Database*
- Navegue até o arquivo Chinook.db (provavelmente localizado em seus downloads)
- Clique em *Execute SQL*
- Comece a consultar seus dados



## Comece a consultar seus dados

O diagrama entidade-relacionamento (ERD) do banco de dados foi dado no tópico anterior, mas você também pode encontrá-lo na [página inicial do banco de dados Chinook](#).

Quando parecer que você está com tudo configurado, você pode começar a consultar seu banco de dados! Primeiro, nós poderíamos dar uma olhada em todos os dados da tabela **Invoice** (Fatura):

```
SELECT * FROM Invoice;
```

Sua primeira consulta, MAGNÍFICO!

Agora dê uma olhada no que temos na tabela **Employee** (Funcionário).

```
SELECT * FROM Employee;
```

**Parece que você está pronto(a) para dominar esse PROJETO e o MUNDO!** Tudo o que você vem estudando vai vir a calhar agora!

## Supporting Materials

[Chinook-Db](#)

## SQL: Conjunto de perguntas 1

### Pergunta 1: Quais países possuem mais faturas?

Use a tabela **Invoice** (Fatura) para determinar quais países possuem mais faturas. Forneça as tabelas de **BillingCountry** (país de cobrança) e **Invoices** (faturas) ordenadas pelo número de faturas para cada país. O país com mais faturas deve aparecer primeiro.

#### Verifique sua solução

Sua solução deve ter 2 colunas e 24 linhas. A imagem abaixo mostra um cabeçalho de sua tabela final. As colunas de **Invoices** (faturas) em uma contagem de número de faturas por país. Devem ser classificados de mais para menos faturas.

	BillingCountry	Invoices
1	USA	91
2	Canada	56
3	Brazil	35
4	France	35
5	Germany	28
6	United Kingdom	21
7	Czech Republic	14
8	Portugal	14
9	India	13
10	Argentina	7
11	Australia	7

### Pergunta 2: Qual cidade tem os melhores clientes?

Gostaríamos de lançar um festival de música promocional na cidade que nos gerou mais dinheiro. Escreva uma consulta que retorna a cidade que possui a maior soma dos totais de fatura. Retorne tanto o nome da cidade quanto a soma de todos os totais de fatura.

#### Verifique sua solução

A cidade com mais dólares em fatura foi **Prague** (Praga) com uma quantia de 90.24.

### Pergunta 3: Quem é o melhor cliente?

O cliente que gastou mais dinheiro será declarado o melhor cliente. Crie uma consulta que retorna a pessoa que mais gastou dinheiro. Eu encontrei essa informação ao linkar três tabelas: **Invoice** (fatura), **InvoiceLine** (linha de faturamento), e **Customer** (cliente). Você provavelmente consegue achar a solução com menos tabelas!

#### Verifique sua solução

O cliente que mais gastou de acordo com as faturas foi o **Cliente 6** com 49.62 em compras.

## SQL: Conjunto de perguntas 2

### Pergunta 1

Use sua consulta para retornar o e-mail, nome, sobrenome e gênero de todos os ouvintes de Rock. Retorne sua lista ordenada alfabeticamente por endereço de e-mail, começando por [A](#). Você consegue encontrar um jeito de lidar com e-mails duplicados para que ninguém receba vários e-mails? Eu escolhi linkar as informações das tabelas **Customer** (cliente), **Invoice** (fatura), **InvoiceLine** (linha de faturamento), **Track** (música) e **Genre** (gênero), mas você pode encontrar outra forma de obter a informação.

### Verifique sua solução

Com a minha consulta, descobri que todos os clientes possuem uma conexão ao Rock (você poderia ver isso olhando para o comprimento original da tabela de clientes. Sua tabela final deve ter 59 linhas e 4 colunas (se você quiser checar a conexão com 'Rock'). O cabeçalho da tabela é fornecido abaixo.

	Email	FirstName	LastName	Name
1	aaronmitchell@yahoo.ca	Aaron	Mitchell	Rock
2	alero@uol.com.br	Alexandre	Rocha	Rock
3	astrid.gruber@apple.at	Astrid	Gruber	Rock
4	bjorn.hansen@yahoo.no	Bjørn	Hansen	Rock
5	camille.bernard@yahoo.fr	Camille	Bernard	Rock
6	daan_peeters@apple.be	Daan	Peeters	Rock
7	diego.gutierrez@yahoo.ar	Diego	Gutiérrez	Rock
8	dmiller@comcast.com	Dan	Miller	Rock
9	dominiquelefebvre@gmail.com	Dominique	Lefebvre	Rock
10	edfrancis@yachoo.ca	Edward	Francis	Rock
11	eduardo@woodstock.com.br	Eduardo	Martins	Rock

### Pergunta 2: Quem está escrevendo as músicas de rock?

Agora que sabemos que nossos clientes amam rock, podemos decidir quais músicos convidar para tocar no show.

Vamos convidar os artistas que mais escreveram as músicas de rock em nosso banco de dados. Escreva uma consulta que retorna o nome do **Artist** (artista) e a contagem total de músicas das dez melhores bandas de rock.

Você precisará usar as tabelas **Genre** (gênero), **Track** (música), **Album** (álbum), and **Artist** (artista).

### Verifique sua solução

As dez melhores bandas são mostradas abaixo junto do número de canções que cada uma possui em registro.

	ArtistId	Name	Songs
1	22	Led Zeppelin	114
2	150	U2	112
3	58	Deep Purple	92
4	90	Iron Maiden	81
5	118	Pearl Jam	54
6	152	Van Halen	52
7	51	Queen	45
8	142	The Rolling Stones	41
9	76	Creedence Clearwater Revival	40
10	52	Kiss	35

### Pergunta 3

Primeiro, descubra qual artista ganhou mais de acordo com **InvoiceLines** (linhas de faturamento). Agora encontre qual cliente gastou mais com o artista que você encontrou acima.

Para essa consulta, você precisará usar as tabelas **Invoice** (fatura), **InvoiceLine** (linha de faturamento), **Track** (música), **Customer** (cliente), **Album** (álbum) e **Artist** (artista). Observe que essa consulta é complicada porque a quantia **Total** gasta na tabela **Invoice** (fatura) pode não ser em um só produto, então você precisa usar a tabela **InvoiceLine** (linha de faturamento) para descobrir quanto de cada produto foi comprado e, então, multiplicar isso pelo preço de cada artista.

### Verifique sua solução

Os artistas com maior faturamento são mostrados na tabela abaixo. Em primeiro lugar, temos **Iron Maiden**.

	Name	AmountSpent
1	Iron Maiden	138.6
2	U2	105.93
3	Metallica	90.08999999999999
4	Led Zeppelin	86.12999999999999
5	Lost	81.59

Continuação da solução com o maior comprador

Então, os maiores compradores são mostrados na tabela abaixo. O cliente com a maior quantia total em fatura é o cliente **55, Mark Taylor**.

	Name	AmountSpent	CustomerId	FirstName	LastName
1	Iron Maiden	17.82	55	Mark	Taylor
2	Iron Maiden	15.84	35	Madalena	Sampaio
3	Iron Maiden	13.86	16	Frank	Harris
4	Iron Maiden	13.86	36	Hannah	Schneider
5	Iron Maiden	8.91	5	František	Wichterlová
6	Iron Maiden	8.91	27	Patrick	Gray



## SQL (Avançado): Conjunto de perguntas 3

### SQL Avançado:

Para resolver as perguntas que estão aqui, você precisa escrever uma consulta que se estende além do conteúdo abordado nessas aulas. Essas perguntas estão aqui simplesmente para mostrar que há extensões do material que já abordamos, mas você definitivamente tem os blocos de construção necessários para enfrentar esses tópicos mais difíceis! **Essas perguntas são dadas como material adicional para desafiar você!** Cada uma das perguntas abaixo requer as ferramentas com as quais você já está familiarizado(a), mas elas também usam um novo método conhecido como uma **SUBCONSULTA**.

### Pergunta 1

Queremos descobrir o gênero musical mais popular em cada país. Determinamos o gênero mais popular como o gênero com o maior número de compras. Escreva uma consulta que retorna cada país juntamente a seu gênero mais vendido. Para países onde o número máximo de compras é compartilhado retorne todos os gêneros.

Para essa consulta você precisará usar as tabelas **Invoice** (fatura), **InvoiceLine** (linha de faturamento), **Track** (música), **Customer** (cliente) e **Genre** (gênero).

### Verifique sua solução

Embora existam apenas 24 países, a consulta deve retornar 25 linhas. As primeiras 11 linhas são mostradas na imagem abaixo. Observe que Argentina possui 2 gêneros que compartilham o máximo.

	Purchases	Country	Name	GenreId
1	9	Argentina	Alternative & Punk	4
2	9	Argentina	Rock	1
3	22	Australia	Rock	1
4	15	Austria	Rock	1
5	21	Belgium	Rock	1
6	81	Brazil	Rock	1
7	107	Canada	Rock	1
8	9	Chile	Rock	1
9	25	Czech Republic	Rock	1
10	21	Denmark	Rock	1
11	18	Finland	Rock	1

## Pergunta 2

Retorne todos os nomes de músicas que possuem um comprimento de canção maior que o comprimento médio de canção. Embora você possa fazer isso com duas consultas. Imagine que você queira que sua consulta atualize com base em onde os dados são colocados no banco de dados. Portanto, você não quer fazer um hard code da média na sua consulta. Você só precisa da tabela **Track**(música) para completar essa consulta.

Retorne o **Name** (nome) e os **Milliseconds** (milissegundos) para cada música. Ordene pelo comprimento da canção com as músicas mais longas sendo listadas primeiro.

### Verifique sua solução

Abaixo temos uma imagem de como as primeiras dez linhas de sua tabela devem ser. Deve haver apenas **494** das 3503 músicas em sua tabela.

	Name	Milliseconds
1	Occupation / Precipice	5286953
2	Through a Looking Glass	5088838
3	Greetings from Earth, Pt. 1	2960293
4	The Man With Nine Lives	2956998
5	Battlestar Galactica, Pt. 2	2956081
6	Battlestar Galactica, Pt. 1	2952702
7	Murder On the Rising Star	2935894
8	Battlestar Galactica, Pt. 3	2927802
9	Take the Celestra	2927677
10	Fire In Space	2926593

### Pergunta 3

Escreva uma consulta que determina qual cliente gastou mais em músicas por país. Escreva uma consulta que retorna o país junto ao principal cliente e quanto ele gastou. Para países que compartilham a quantia total gasta, forneça todos os clientes que gastaram essa quantia.

Você só precisará usar as tabelas **Customer** (cliente) e **Invoice** (fatura).

#### Verifique sua solução

Embora existam apenas 24 países, a consulta deve retornar 25 linhas. As **últimas** 11 linhas são mostradas na imagem abaixo. Observe que o Reino Unido tem 2 clientes que compartilham o máximo.

	Country	TotalSpent	FirstName	LastName	CustomerId
15	Ireland	45.62	Hugh	O'Reilly	46
16	Italy	37.62	Lucas	Mancini	47
17	Netherlands	40.62	Johannes	Van der Berg	48
18	Norway	39.62	Bjørn	Hansen	4
19	Poland	37.62	Stanisław	Wójcik	49
20	Portugal	39.62	João	Fernandes	34
21	Spain	37.62	Enrique	Muñoz	50
22	Sweden	38.62	Joakim	Johansson	51
23	USA	47.62	Richard	Cunningham	26
24	United Kingdom	37.62	Emma	Jones	52
25	United Kingdom	37.62	Phil	Hughes	53

# Datas em SQLite (o projeto) diferenciam-se do Postgres (sala de aula)

No projeto você está trabalhando com uma sintaxe SQL um pouco diferente daquela na sala de aula. Embora a maioria dos comandos e lógicas vá transitar diretamente de um a outro, há algumas diferenças entre SQLite (usado para esse projeto) e PostgreSQL (usado em sala de aula). Especificamente, você pode ser impactado graças às diferenças de funcionalidade de datas .

[Aqui](#) temos um link útil para ajudar no trabalho com datas no SQLite. Se você tiver de trabalhar em outro ambiente SQL no futuro, como o Microsoft SQL Server, Oracle, MySQL ou qualquer outro ambiente SQL; novamente, você poderá encontrar diferenças sutis. Com suas habilidades atuais, uma rápida pesquisa no Google provavelmente irá ajudar você a transferir o que você sabe para trabalhar com esses ambientes muito rapidamente.

## Envio do projeto

### *Apresentações*

Agora você está na parte do projeto na qual você precisa enviá-lo para um revisor. Para ter seu projeto aprovado, siga as instruções abaixo para criar uma apresentação!

Sua apresentação deve incluir:

- Quatro slides
- Uma visualização por slide
- Uma explicação de uma ou duas linhas sobre cada slide
- A consulta SQL usada para criar os dados usados na visualização.

**Observe, você pode decidir usar consultas que são motivadas pelas perguntas nos tópicos anteriores ou pode optar por quatro novas perguntas.** O modelo de envio é um arquivo Google Slides. Faça uma cópia do modelo de apresentação para concluir o seu projeto. Sugerimos que você use o layout fornecido, embora não seja uma exigência.

### *Consultas*

Por favor, inclua um arquivo de texto que inclui cada uma das consultas utilizadas para criar as visualizações. Você deve formatar suas consultas para facilitar a leitura. Use esta ferramenta para ajudá-lo: <http://www.sql-format.com/>.

Envie suas consultas em um arquivo de texto sem formatação, usando o bloco de notas, notepad++, ou [atom](#), formatados como a imagem abaixo:

```
/* Query 1 - query used for first insight */
SELECT t1.col, COUNT(*) ct
FROM table t1
JOIN table2 t2
ON t1.col = t2.col
GROUP BY t1.col;

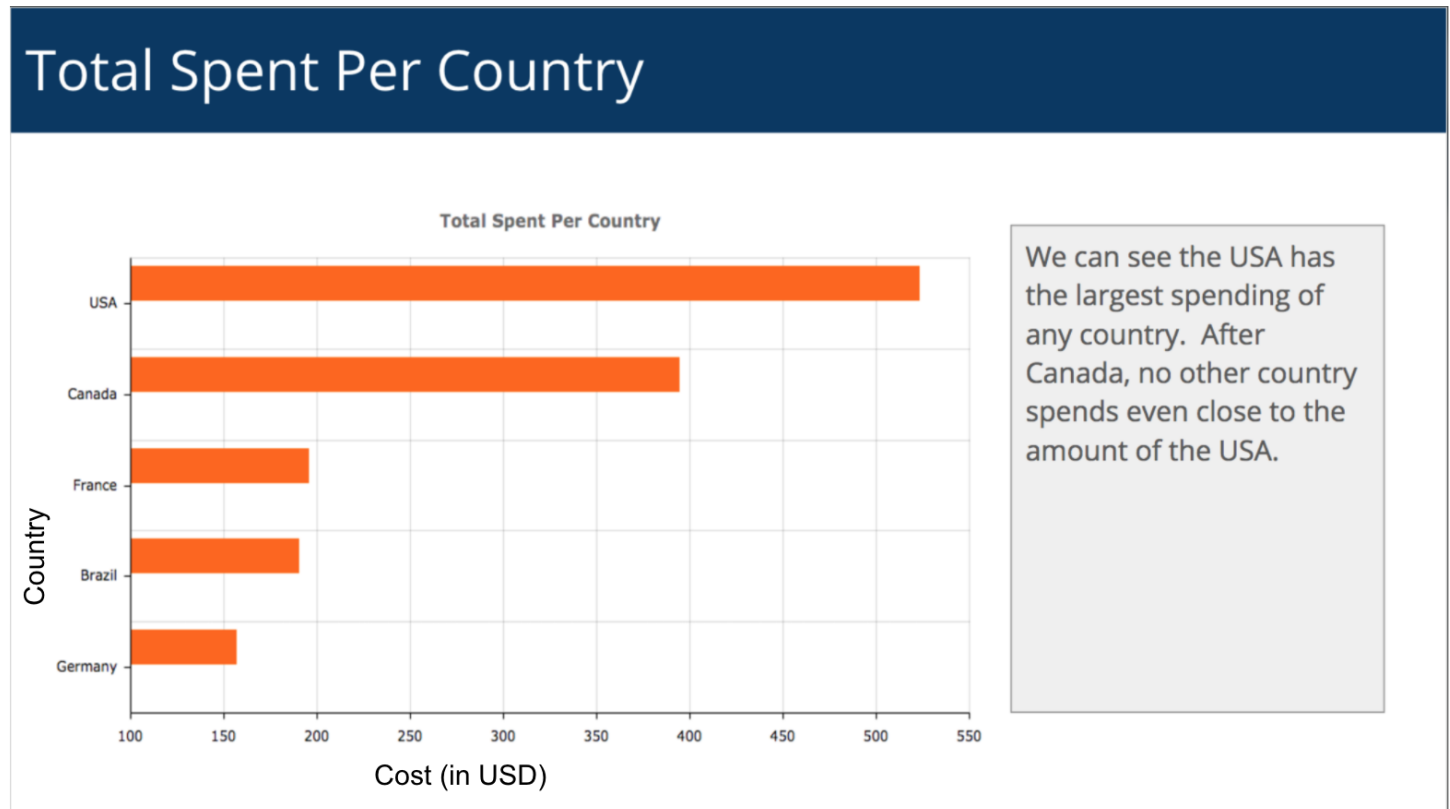
/* Query 2 - the query used for second insight*/
SELECT t1.col, COUNT(*) ct
FROM table t1
JOIN table2 t2
ON t1.col = t2.col
GROUP BY t1.col;
```

**Coloque seu arquivo de texto e sua apresentação em uma pasta, e crie um arquivo compactado (um zip) desta pasta. Envie este arquivo zip.** Um modelo de slide é fornecido abaixo:

[MODELO PARA ENVIO](#)

### Visualizações

Sugerimos que você use um aplicativo de planilha, como o Excel ou Google Sheets para criar suas visualizações. No entanto, fique à vontade para usar qualquer ferramenta que você gostar. Suas visualizações poderiam ser qualquer uma sobre a qual você aprendeu na aula anterior. Abaixo está um exemplo, e foi dado um link para um slide de exemplo.



Você deve ter quatro slides que são semelhantes ao envio abaixo, mas as perguntas a serem feitas são de sua escolha, e todas as suas quatro consultas finais enviadas devem conter uma **JOIN** e uma **AGREGAÇÃO**. Dê uma olhada na [rubrica](#) para verificar se você atingiu todos os critérios para esse projeto.

[EXEMPLO DE SLIDE PARA ENVIO](#)

### Como colocar dados no Excel

Para exportar resultados de suas consultas do DB Browser, use o botão abaixo e à direita da janela de resultados. Abaixo mostramos como você pode exportar seus dados para um software de planilhas.

Para criar as visualizações como aquelas mostradas no link acima, você precisará mover seus dados fora do SQL para o Excel (ou outro aplicativo de planilhas).

DB Browser for SQLite - C:/Users/sam/Box Sync/\_Udacity/SQL/Northwind.sqlite3

File Edit View Help

New Database Open Database Write Changes Revert Changes

Database Structure Browse Data Edit Pragma Execute SQL

SQL 1

```
1 select * from products
```

	ProductID	ProductName	SupplierID	CategoryID	QuantityPerUnit	UnitPrice	UnitsInStock	UnitsOnOrder
1	1	Chai	1	1	10 boxes x 20 bags	18.0	39	0
2	2	Chang	1	1	24 - 12 oz bottles	19.0	17	40
3	3	Aniseed Syrup	1	2	12 - 550 ml bottles	10.0	13	70
4	4	Chef Anton's Cajun Seasoning	2	2	48 - 6 oz jars	22.0	0	0

77 rows returned in 4ms from: select \* from products

Export Database Cell

Mode: Text Import Export Set as NULL

Type of data currently in cell: NULL  
0 byte(s) Apply

DB Schema

Name	Type	Schema
Categories	CREATE	
CustomerCustomer...	CREATE	
CustomerDemogra...	CREATE	
Customers	CREATE	
EmployeeTerritories	CREATE	
Employees	CREATE	
Order Details	CREATE	

Selecione Export to CSV e, então, selecione as configurações da mesma forma que mostramos abaixo. Certifique-se de sua configuração em **New line characters** está colocada como `Unix: LF (\n)`.

Export data as CSV

Column names in first line ☒

Field separator

Quote character

New line characters

OK Cancel

## Orientações adicionais

- Não deve haver qualquer preparação de dados adicional (classificação, filtragem, renomeação, etc.) entre o output da consulta e a visualização.
  - Todas as suas quatro consultas devem incluir pelo menos uma join e uma agregação.
  - Revise seu projeto comparando-o com a [Rubrica](#). Os revisores vão usar isso para avaliar o seu trabalho.
  - A primeira parte deste projeto visa ajudá-lo a entender o banco de dados, para que você possa fazer perguntas interessantes na segunda parte. Fique à vontade para usar e se expandir sobre as consultas que você escreveu na primeira parte.
  - Quando você terminar seu projeto, sua apresentação deve estar em um arquivo PDF, e suas consultas em um arquivo .txt. Não se esqueça de enviar estes arquivos dentro de uma pasta compactada (zip).
  - Não tenha medo de desafiar a si mesmo(a)! Tente combinar os conceitos de SQL que você conhece!
- Para revisar sua apresentação, você precisará salvar seus slides como um PDF.** Você pode fazer isso no Google Slides, selecionando **Arquivo > Baixar como > Documento PDF**.

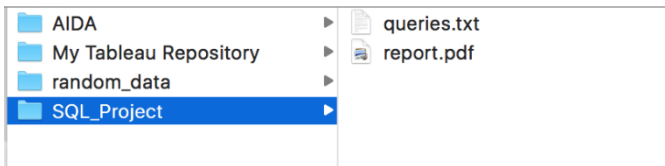
## Project Submission

Para submeter seu projeto, por favor faça o seguinte:

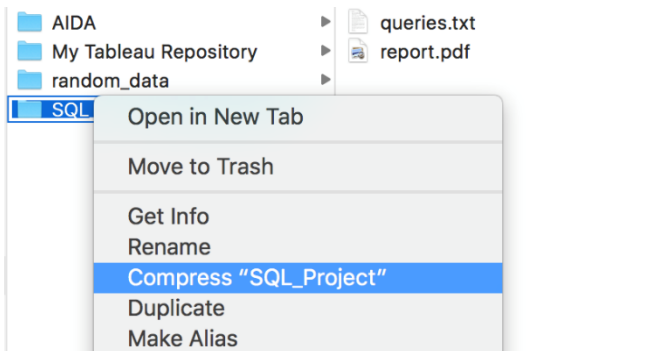
- Revise seu projeto de acordo com a [Rubrica](#). Revisores irão usar esta rubrica para avaliar seu trabalho.
- Crie seus slides com qualquer software de apresentação que você preferir (por exemplo: Google Slides, PowerPoint, Keynote, etc.)

**Para ter sua apresentação avaliada, você precisa salvar seus slides como um arquivo **PDF**.** Você pode fazer isto diretamente do Google Slides ao selecionar **Arquivo > Download como > Documento PDF**.

- Crie um arquivo de texto separado com cada uma das queries SQL usadas para criar as visualizações.
- Salve a apresentação como um PDF e as queries SQL em um arquivo de texto (.txt) na mesma pasta.



- Comprima (zipe) a pasta e submeta esta pasta comprimida com ambos os arquivos nela.



- Submeta o arquivo comprimido.

