

# Manipulação de Dados com .CSV e MySQL Workbench

Cartilha de boas práticas para tratamento de dados referente à  
disciplina de 'Tópicos em Gestão'



Instituto Federal do Maranhão - Campus Monte Castelo

Autor: José Carlos Chaves Brasil Junior  
Graduando em Sistemas de Informação

Desenvolvido para a disciplina de 'Tópicos em Gestão'  
Ministrada pelo professor Abraão Neiver de Miranda Azevedo

São Luís, 2024  
Maranhão

## SUMÁRIO

<b>Sobre este guia.....</b>	<b>4</b>
<b>Contribua com este guia.....</b>	<b>5</b>
<b>1. Formulários/questionários/pesquisas.....</b>	<b>6</b>
<b>2. Tratando a planilha e exportando o arquivo ‘.csv’.....</b>	<b>7</b>
2.1 Tratando planilhas.....	7
2.2 Exportando os arquivos ‘.csv’.....	8
<b>3. ‘Charsets’ e como alterá-los.....</b>	<b>9</b>
<b>4. Importando dados.....</b>	<b>10</b>
<b>5. Manipulando dados.....</b>	<b>12</b>
<b>6. Dicas úteis.....</b>	<b>13</b>

## Sobre este guia

Esta cartilha tem como objetivo padronizar alguns procedimentos referentes à confecção de pesquisas que serão realizadas em sala de aula, tal como pontuar as condições mínimas que devem oferecer para que sua utilidade seja extraída ao máximo, além de propor um guia aos alunos, principalmente, sobre manipulação de arquivos **‘.csv’** em conjunto com a plataforma **MySQL Workbench**.

É válido ressaltar que este guia — ou cartilha, como preferir — foi estruturado com base em experimentos realizados em sala de aula, principalmente. Sendo assim, está aberta a mudanças e espera-se que mude, visto que alguns procedimentos apresentados até então são realizados sem automação alguma(\*); procedimentos esses que podem, supostamente, ser atribuídos a algum script de preferência.

Como essas questões, em conjunto com a utilização de recursos avançados como *‘procedures’*, *‘triggers’* e funções, além de estruturas lógicas através da utilização de linguagens de programação, não estão em foco, **vamos nos ater apenas ao tratamento das planilhas e arquivos ‘.csv’ e migração de dados entre planilhas e banco de dados de maneira equânime.**

Espera-se que aluno possua conhecimentos em e esteja habituado com:

- Google Forms
- Google Sheets
- MySQL Workbench e banco de dados relacional.
- Visual Studio Code (ou outro editor de código-fonte de sua preferência)

E que já possua:

- Pesquisa realizada através do Google Forms

É válido (muito válido, inclusive) ter em mente que, além dos dados obtidos pelas pesquisas, é importante que haja, no próprio banco de dados, dados sobre as próprias pesquisas, como as questões que elas possuem, por exemplo. Chamaremos, em alguns momentos e sem rigor, esses mesmos dados de **‘metadados da pesquisa’**. Isso é necessário ressaltar devido, por exemplo, à dificuldade que há em manipular pesquisas, importadas como grandes planilhas, sem um tratamento prévio referente às colunas e como são apresentadas (isso será explicado mais à frente).

## Contribua com este guia

Todas as informações referentes ao processo de padronização foram coletadas, principalmente, em sala de aula, **a fim de solucionar uma questão específica abordada na disciplina de 'Tópicos em Gestão'**. Sendo assim, é evidente que muitos tópicos aprofundados relacionados a banco de dados relacional, manipulação de arquivos '.csv' e planilhas, além de ferramentas diversas que ainda são desconhecidas por mim, não foram abordados aqui. Cabe a você, aluno, contribuir com esse guia.

Entre em contato através de:

- **GitHub:** [github.com/jjuniorbrasil](https://github.com/jjuniorbrasil)
- **E-mail:** [junior.brasil@acad.ifma.edu.br](mailto:junior.brasil@acad.ifma.edu.br)

para sugestões e dúvidas.

## 1. Formulários/questionários/pesquisas



É de se esperar que você já possua uma pesquisa realizada através da plataforma **Forms**, do Google, com campos que podem ser utilizados como '**chave primária**' dentro do MySQL (por exemplo, **exigir e-mail ou matrícula** dentro do questionário). Optamos por essa abordagem devido à facilidade que é trabalhar com o ambiente Google e suas aplicações, tais quais o **Google Drive** e o **Google Sheets** (esse último em especial) e pela necessidade de utilizarmos uma chave primária para correlacionar diferentes pesquisas ao utilizar **MySQL Workbench**.

Novamente, garanta que sua pesquisa:

- **Tenha sido feita através do Google Forms**
- **Exija informação(ões) que pode ser utilizada como chave primária em uma tabela no banco de dados**

Feito isso, gere uma planilha com as respostas através das opções do seu próprio formulário/questionário.

## 2. Tratando a planilha e exportando o arquivo '.csv'



### 2.1 Tratando planilhas

Arquivos **'.csv'**, quando importados como tabelas para o **MySQL Workbench**, funcionam como entidades. Todo arquivo **'.csv'** terá um **título**, **colunas** com respectivos nomes e **valores**, que são todas as tuplas com exceção da primeira. É possível ver esse formato desde o Google Sheets, quando a planilha com os dados obtidos pela pesquisa é criada. Em termos de banco de dados, a primeira linha de um arquivo **'.csv'**, mesmo quando ainda é uma planilha, sempre se refere ao nome das colunas da respectiva tabela.

	A	B	C	D	E	F
1	Carimbo de data	Email	Nome	Qual (ou quais) :	Quais tecnologias	Você costuma cr D
2	07/03/2024 18:0	fe	ac M	Desenvolvimento	HTML, CSS, Jav	Sim, possuo/tenl E
3	07/03/2024 18:0	h	@ H	s Desenvolvimento	JavaScript	Não, programo v E
4	07/03/2024 18:0	ju	ci Jo	na Desenvolvimento	React, TailwindC	Sim, possuo/tenl A
5	07/03/2024 18:0	g	ci Zi	Desenvolvimento	SQL, Python par	Sim, possuo/tenl E
6	07/03/2024 18:0	ru	@ R	el Desenvolvimento	Python, JavaScri	Sim, possuo/tenl E
7	07/03/2024 18:2	ra	za R	o Desenvolvimento	back-end, Segu	Não, programo v M
8	07/03/2024 18:2	s	ac S	s Desenvolvimento	Desenvolvimento	Sim, possuo/tenl E
9	07/03/2024 18:2	le	@ La	Desenvolvimento	front-end (web	Sim, possuo/tenl E
10	12/03/2024 17:1	fr	ac F	s Segurança da Informação		Sim, possuo/tenl E
11	12/03/2024 18:0	le	il. P	Desenvolvimento	Todas	Sim, possuo/tenl E
12	12/03/2024 18:0	p	da P	Segurança da In	Inteligência artifi	Sim, possuo/tenl E
13	12/03/2024 18:1	j	da P	Desenvolvimento	react	Sim, possuo/tenl M

Figura 2.1.1 - Exemplo de pesquisa em formato de planilha Google Sheets

Entretanto, a planilha ainda precisa ser tratada, pois os nomes das colunas precisam ser **identificadores simples e práticos**. Isso pode ser facilmente resolvido adicionando uma linha ao início da tabela (a fim de que não se percam os metadados da pesquisa) e adotando nomes **significativos** para cada coluna da seguinte maneira:

	A	B	C	D
1	Carimbo de data	Nome	Qual (ou quais)	Quais tecnologia
2	07/03/2024 18:0	////////////////////////////////	Desenvolvimento	HTML, CSS, Jav

	A	B	C	D
1				
2	Carimbo de data	Nome	Qual (ou quais)	Quais tecnologia
3	07/03/2024 18:0	////////////////////////////////	Desenvolvimento	HTML, CSS, Jav

	A	B	C	D
1	data	nome	q1	q2
2	Carimbo de data	Nome	Qual (ou quais)	Quais tecnologia
3	07/03/2024 18:0	////////////////////////////////	Desenvolvimento	HTML, CSS, Jav

Figura 2.1 2- Tratamento da planilha pré-'.csv'.

## 2.2 Exportando os arquivos '.csv'

Garantindo o tópico anterior, a planilha está pronta para ser exportada como um arquivo '.csv', através das opções:

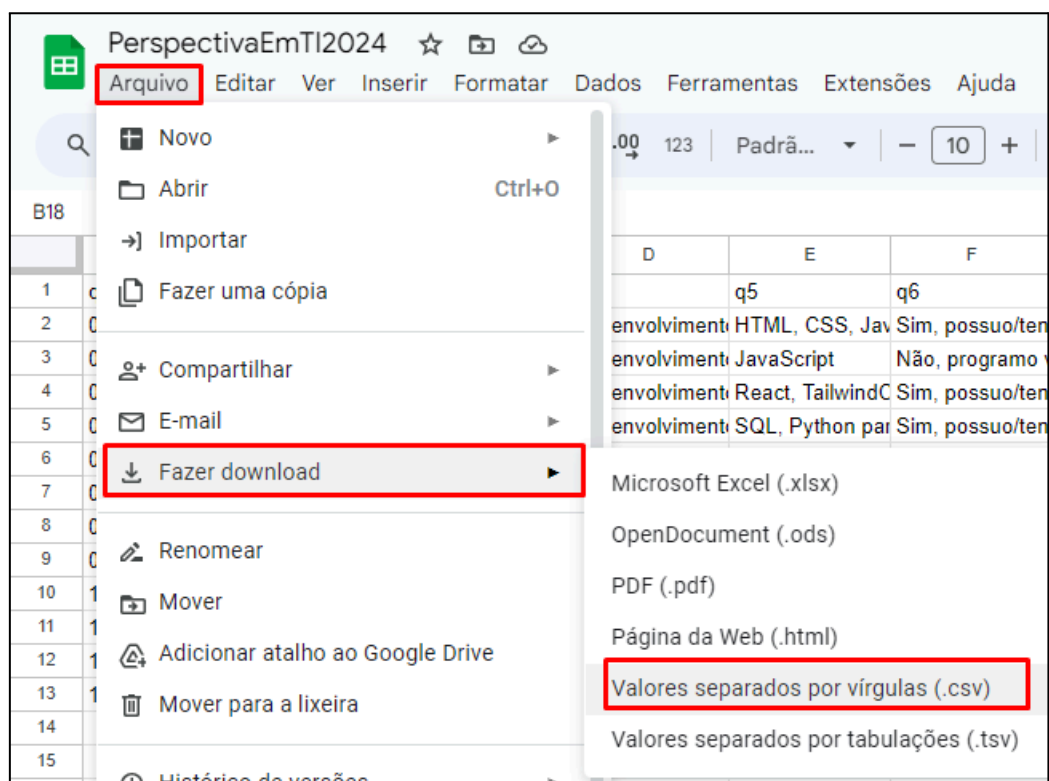


Figura 2.2.1 - Exportação dos dados em arquivos '.csv'



### 3. 'Charsets' e como alterá-los

Planilhas do Google Sheets exportadas como arquivos '.csv' podem (e vão) gerar inconsistências quando importadas pelo banco de dados. Principalmente, relacionadas ao '**charset**' que o Google Sheets utiliza em relação à maneira que o MySQL Workbench interpreta. Cabe mudar o *charset* manualmente, através de um editor de código-fonte, como o **Visual Studio Code**. Para isso:

1. Abra o arquivo '.csv' que você acabou de exportar no Visual Studio Code
2. Clique na opção 'Select Encoding', na barra inferior:

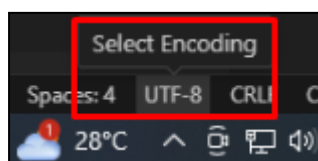


Figura 3.1 - Opção 'Select Encoding'. Usualmente, estará como padrão 'UTF-8'

3. Salve o arquivo novamente utilizando o formato 'iso88591'

:

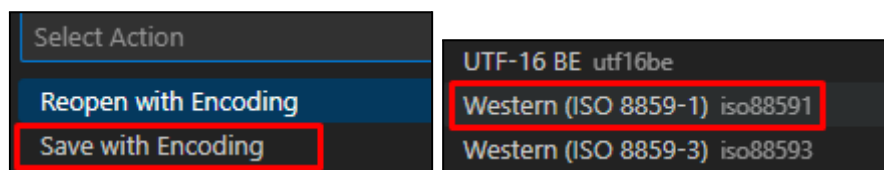


Figura 3.2 - Alterando o 'charset' do arquivo '.csv'

Dessa maneira, é possível evitar inconsistências diversas na maneira que os dados se apresentam durante as consultas em SQL.

## 4. Importando dados

Importar os dados através de arquivos '.csv', então, se torna mais fácil após o tratamento realizado nos capítulos anteriores. Tendo um banco de dados, **importe o arquivo '.csv' da pesquisa como uma tabela** ao seu próprio banco de dados através da ferramenta **'Table Data Import Wizard'**:

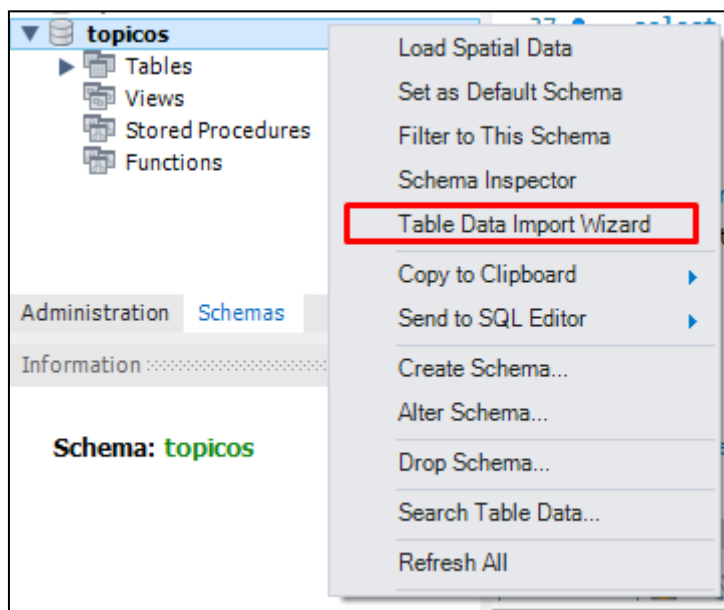


Figura 4.1 - Importação dos dados da pesquisa.

Como dito anteriormente, é esperado que o aluno esteja habituado com a plataforma MySQL Workbench e consiga localizar o seu próprio banco de dados no navegador (aba 'Schemas'). Garanta, então, que seja criada uma nova tabela com a opção **'Create new table'** marcada. Dessa maneira, as pesquisas ficam separadas umas das outras:

 A screenshot of a dialog box titled 'Select destination table and additional options.' It contains three radio button options: 'Use existing table:' (unselected), 'Create new table:' (selected), and 'Drop table if exists' (unselected). The 'Use existing table:' option has a dropdown menu showing 'topicos.pesqfran'. The 'Create new table:' option has two input fields: the first is a dropdown menu showing 'topicos' and the second is a text box containing 'pesqjj'.

Figura 4.2 - Opções de tabela

Caso seja necessário, faça configurações em relação à maneira que os dados serão importados. Após isso, prossiga:

**Configure Import Settings**

Detected file format: csv

Encoding:

Columns:

<input checked="" type="checkbox"/>	Source Column	Dest Column
<input checked="" type="checkbox"/>	q1	<input type="text" value="q1"/>
<input checked="" type="checkbox"/>	q2	<input type="text" value="q2"/>
<input checked="" type="checkbox"/>	q3	<input type="text" value="q3"/>
<input checked="" type="checkbox"/>	q4	<input type="text" value="q4"/>
<input checked="" type="checkbox"/>	q5	<input type="text" value="q5"/>
<input checked="" type="checkbox"/>	q6	<input type="text" value="q6"/>

<

q1    q2    q3    q4    q5

Figura 4.3 - Opções de importação

Detected file format: csv

Options:

Field Separator

Line Separator

Enclose Strings in

null and NULL word as SQL keyword

Figura 4.4 - Opções de importação (após clicar na chave)

De acordo com a figura 4.4, garanta que pelo menos as opções de 'Field Separator' e 'Enclose Strings in' estejam de acordo com a maneira que seu arquivo .csv está formatado (você pode abri-lo pelo **Visual Studio Code** novamente ou até mesmo pelo bloco de notas, para conferir). Por padrão, o Google Sheets exporta os dados separados por vírgula e entre aspas duplas.

Até aqui, parte essencial do processo está feito: **a pesquisa foi realizada e os dados obtidos já estão no banco de dados**. Entretanto, ainda nos resta inserir os metadados da pesquisa ao banco de dados, mas isso fica a seu cargo, de acordo com a necessidade. **(veja as dicas no capítulo 6)**

## 5. Manipulando dados

A partir daqui, a manipulação de dados se dá de maneira natural, visto que garantimos uma chave primária, seja qual for, às pesquisas. Por exemplo, a partir das tabelas *'pesqjj'* e *'pesqfran'*, que se referem a pesquisas de alunos diferentes, temos, por exemplo:

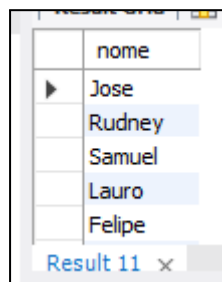
```

50      -- Nome de pessoas que programam fora de horário de aula/trabalho
51      -- e estão satisfeitas com o curso.
52
53 •    select nome from pessoa as p
54          inner join pesqjj as p1 on p.email = p1.q9
55          inner join pesqfran as p2 on p1.q9 = p2.q2
56      WHERE p1.q5 LIKE 'Sim%' AND p2.q8 > 7;

```

Figura 5.1 - Exemplo de consulta correlacionando pesquisas diferentes

Observe que, com nomes simples para as colunas que contém os dados das pesquisas, como 'q1' e semelhantes, declarar uma consulta torna-se extremamente mais fácil. Novamente, a escolha quanto aos nomes das colunas é arbitrário. Caso tenha uma ideia que considere mais eficiente para o seu caso, não hesite em usá-la.



nome
Jose
Rudney
Samuel
Lauro
Felipe

Figura 5.2 - Resultado da consulta em SQL apresentada na figura 5.1.

Se preferir, utilize uma tabela *'pessoa'*, como a do exemplo (ver o próximo capítulo), para guardar dados importantes que podem ser utilizados no futuro, como e-mails, datas de nascimento e até mesmo matrículas em âmbito acadêmico (que podem ser utilizadas como chave primária em outras tabelas, adicionando uma nova coluna).

## 6. Dicas úteis

- a) **Crie uma tabela para dados dos entrevistados:** Como dito anteriormente, de maneira simples: na tabela 'pessoa', por exemplo, irão ser armazenados dados sobre indivíduos pesquisados. É uma tabela que considero importante, visto que correlacionar dados a longo prazo, em um mesmo banco de dados, se torna uma tarefa complexa à medida que as pesquisas mudam de forma em um mesmo banco de dados.
- b) **Crie uma tabela para dados sobre a pesquisa:** Na tabela 'pesquisa' se registrarão as próprias pesquisas, cada qual com seu '*id*' único, nome da tabela de acordo com '*titulopesquisa*' e breve descrição armazenada em '*descpesquisa*'. A tabela 'questoes\_pesquisas' armazenará apenas questões, cada qual com seu '*id*'.
- c) **Existe a opção de adicionar comentários às colunas,** evitando a necessidade de um máximo rigor quanto ao tratamento da planilha antes de ser exportada, para armazenar perguntas, de maneira exclusiva. Isso também elimina a necessidade de que a primeira linha de cada pesquisa contenham os 'metadados da pesquisa'. Cabe a você optar pela maneira que armazena esse dado, de acordo com a necessidade. De qualquer maneira, não o perca. É imprescindível que estejam no banco de dados.