Observações e passo a passo do trabalho no banco de dados:

# Ferramentas e estruturas de tabelas

1. Exportação do banco de dados [residenciatechad](http://ip-45-79-142-173.cloudezapp.io/phpmyadmin/db_structure.php?server=1&db=residenciatechad), disponibilizado pela Harve, para o computador local.
2. Baixado os seguintes programas, nesta ordem:
   1. [MySQL Workbench](https://dev.mysql.com/downloads/workbench/), verificado e baixado antes os dois programas;
      1. [Microsoft .NET Framework 4.5](https://www.microsoft.com/en-us/download/details.aspx?id=30653)
      2. [Visual C++ Redistributable for Visual Studio 2019](https://learn.microsoft.com/en-us/cpp/windows/latest-supported-vc-redist?view=msvc-170#visual-studio-2015-2017-2019-and-2022)
   2. [Wampserver](https://www.wampserver.com/)
3. Conexão do Visual Studio Code com MySQL e o banco de dados.
4. Conexão com o repositório no GitHub , onde serão colocadas todas as tabelas e arquivos referentes a este projeto:

<https://github.com/LucyMenges/Tech_Dados_Projeto_Setor_Publico>

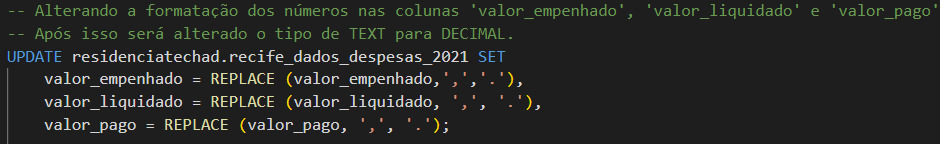
# Observações gerais:

* [**ANEXO 1 - TABELAS**](#_heading=h.5kxycebki5wd)**:**  compilado as informações das tabelas: nome de colunas, tipo, Collation e Valores Nulos por arquivo. Foi utilizado para analisar quais as colunas seriam necessárias para realizar as demandas. E a verificação da necessidade da alteração do Tipo destas colunas ou não.
* Após a visualização e análise inicial dos quatro arquivos, foi definido que estes seriam preparados no MySQL Workbench, na estrutura de tabelas fato e dimensões e depois o resultado de cada consulta exportado e salvo como .csv, para ser utilizado diretamente no Power BI.
* O objetivo foi deixar as tabelas mais enxutas e limpas, apenas com as informações necessárias para responder às demandas solicitadas.
* Não foram automatizadas as conexões e consultas, no primeiro momento, já que não havia a necessidade de atualizações do banco de dados neste projeto. Entretanto foi testado com automatização num outro momento, após a finalização das demandas, quando foi conectado o banco de dados direto no Power BI e inserida as consultas SQL diretamente para criação das tabelas.
  + Durante esta fase de preparação das tabelas como fato e dimensão ficou claro que não haveria a necessidade desta estrutura para este projeto, devido a baixa complexidade das informações solicitadas.
  + Poderia ter feito a seleção de algumas colunas, mas sem desmembrar a tabela em outras com as informações complementares (dimensão). Teria reduzido o tempo de execução desta fase, sem alterar, ao meu ver, o resultado final.
* As consultas feitas em SQL foram salvas no repositório do Github, arquivo:
  + “Script Projeto Setor Publico.sql”

# Observações sobre as tabelas:

**\* Tabela “residenciatechad.recife\_dados\_despesas\_2021”**

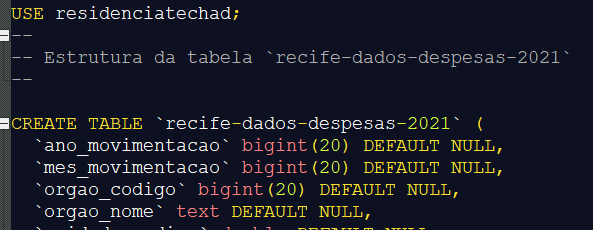
* Nesta tabela seria necessário alterar três colunas de valores que estão como TEXT para DECIMAL por serem valor monetário.
* --Alterando o tipo das colunas
* ALTER TABLE residenciatechad.recife\_dados\_despesas\_2021
* CHANGE COLUMN `valor\_empenhado` `valor\_empenhado` DECIMAL NULL DEFAULT NULL,
* CHANGE COLUMN `valor\_liquidado` `valor\_liquidado` DECIMAL NULL DEFAULT NULL,
* CHANGE COLUMN `valor\_pago` `valor\_pago` DECIMAL NULL DEFAULT NULL;
* O tipo das colunas foi alterado e ao conferir os valores apresentados delas, se percebeu que estavam diferentes do original, o que geraria resultados errados.
* Isso ocorreu porque o sistema operacional do computador utilizado para o projeto, está no sistema americano de valores, ou seja o ‘.’ (ponto) é o separador de centavos e não a vírgula como no sistema brasileiro.
* Por esse motivo foi necessário fazer uma alteração nas colunas com valores monetários para que os valores não ficassem sem os centavos e aparecessem corretamente. Exemplo: número original na tabela 4788,95, para o MySQL calcular corretamente o decimal deve ser separado por ‘.’ (ponto) e não vírgula.



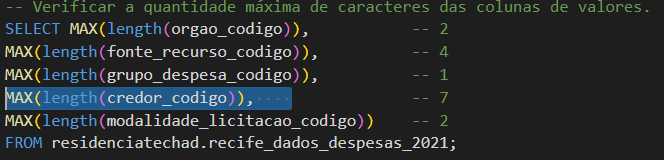
* Porém infelizmente não foi mais possível retornar ao valor inicial original. Então foi deletada a tabela completa “recife\_dados\_despesas\_2021” do banco de dados e inserida novamente através dos comandos iniciais dela.

--Deletando tabela completa do banco de dados

DROP TABLE residenciatechad.recife\_dados\_despesas\_2021;



* O tipo também foi alterado em outras colunas, de bigint(20) para smallint(), pois eram colunas com quantidade baixa de caracteres, como mostra a consulta abaixo.



* Porém a coluna ‘credor\_codigo’ não pode ser alterada por dar um erro, o qual dizia que estava fora do padrão estipulado, na linha 1142. Acabei mantendo a coluna ‘credor\_codigo’ com o tipo original, bigint().

--Alterando o tipo de algumas colunas de BIGINT(20) para INT

ALTER TABLE residenciatechad.recife\_dados\_despesas\_2021

CHANGE COLUMN `ano\_movimentacao` `ano\_movimentacao` YEAR(4) NULL DEFAULT NULL,

CHANGE COLUMN `mes\_movimentacao` `mes\_movimentacao` SMALLINT (2) NULL DEFAULT NULL,

CHANGE COLUMN `orgao\_codigo` `orgao\_codigo` SMALLINT (3) NULL DEFAULT NULL,

CHANGE COLUMN `fonte\_recurso\_codigo` `fonte\_recurso\_codigo` SMALLINT (4) NULL DEFAULT NULL,

CHANGE COLUMN `grupo\_despesa\_codigo` `grupo\_despesa\_codigo` SMALLINT (2) NULL DEFAULT NULL,

CHANGE COLUMN `credor\_codigo` `credor\_codigo` SMALLINT (8) DEFAULT NULL,

CHANGE COLUMN `modalidade\_licitacao\_codigo` `modalidade\_licitacao\_codigo` SMALLINT(2) NULL DEFAULT NULL;

**\* Problemas de acentuação nas tabelas:**

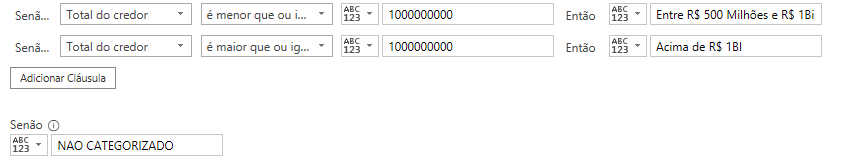
* Nas tabelas “recife\_dados\_despesas\_2021” e “recife\_dados\_receitas\_2021” estavam com distorções nas palavras acentuadas, problema comum. Isso ocorreu provavelmente porque não houve o devido tratamento das palavras antes de gerar o arquivo para o banco de dados (charset utf-8).
* Estas tabelas foram mantidas desta forma no banco de dados. E o tratamento foi feito no Power BI diretamente. No Power Query foi substituído os caracteres desconhecidos pelas letras correspondentes sem acento, por exemplo:
  + ÃÂÃÂÃÂÃÂ substituido por C (representando o Ç)
  + ÃÂÃÂÃÂÃÂ substituído por A (representando o Ã)
  + ÃÂÃÂÃÂÃÂ substituído por U (representando o Ú)
* Foi uma forma um tanto quanto trabalhosa de se alterar as letras acentuadas. Porque é necessário colocar em cada coluna os caracteres individualmente para alteração. E como foram criadas várias tabelas complementares, foi necessário entrar em cada tabela para fazer a alteração.
  + Aqui com toda certeza é necessário se criar um código para tratar as palavras de outra forma, sugestão com o Python.

# Observações Power BI:

Demanda 8. Quem é o principal credor para cada órgão?

* Ao todo o município de Recife tem 7016 credores diferentes. O com maior valor a receber seria o código 0, que corresponde a “Credor não informado”. Ou seja, não foi informado o nome do credor para o pagamento.
* A partir desta lista de 7016 credores, da dificuldade de mostrar num gráfico quais seriam os principais credores e analisando o valor a receber de cada um, surgiu a ideia de categorizar os valores a receber, ou seja traçar um perfil do credor por total a receber.
* Foi criada uma nova tabela referência, da tabela “recife\_dados\_despesas\_2021”, com duas colunas apenas, “credor\_codigo” e “Total do credor”, que foram agrupadas primeiro pelo “credor\_codigo”. Assim gerando uma lista com o total a receber de cada credor.
* Na segunda etapa foi criada uma nova coluna, chamada de “grupo\_credores”, com o agrupamento por faixas de valores aleatórios, como mostro abaixo:

Interface gráfica do usuário, Aplicativo

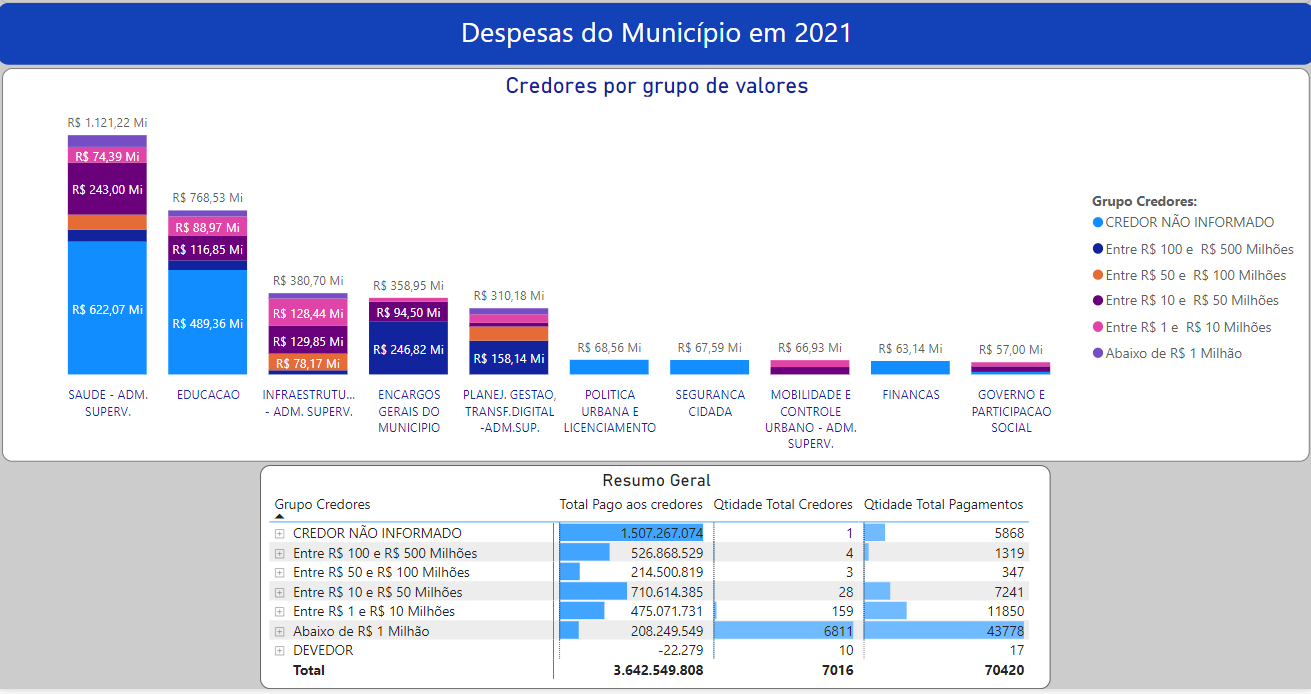
Descrição gerada automaticamente

* Os valores foram classificados nas seguintes categorias:

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

* As faixas foram escolhidas totalmente aleatórias, o que pode levar a um resultado enviesado. Já que não houve nem uma quantidade nem valor padrão para a categorização.
* Neste ponto é necessário se fazer um estudo sobre as formas (técnicas) de se estabelecer uma categorização. Como este ponto não estava dentro desta ou outra demanda, foi deixado para ser resolvido posteriormente.
* Mesmo assim foi gerado um gráfico e uma tabela utilizando esta coluna “grupo\_credores”.



* A partir do resultado apresentado na tabela podemos ver que 6.811, ou 97,1% dos credores receberam até R$ 1 milhão do município. Em contrapartida, apenas 4 credores receberam entre R$ 100 e R$ 500 milhões, ou seja, a faixa superior de pagamentos.
* No somatório em reais dos valores pagos, o segundo maior valor seria com credores que receberam entre R$ 10 e R$ 50 milhões, ou seja totalizando R$ 710.641.385.
* Entretanto permanece a dúvida se estes resultados são confiáveis ou podem estar de alguma forma enviesados, matematicamente falando.

# Arquivos existentes no Github

**Pastas:**

* **Banco Dados Completo**: código SQL para criação do banco de dados "residenciatechad" e os quatro arquivos dele.
* **Tabelas\_Preparadas\_para\_Power\_BI**: tabelas geradas em SQL e salvas como .csv para utilizar no Power BI.

**Arquivos:**

* **Análise de Dados do Setor Público.doc**x: Passo a passo geral da estrutura, Buggs e conclusões de erros.
* **Apresentação Setor Público.docx**: Respostas e explicações dos gráficos produzidos para cada demanda proposta.
* **Projeto 2 Setor Público Recife.pbix**: Visualização das respostas para cada demanda proposta.
* **Projeto 2 Setor Público Recife.pdf**: A visualização gerada no Power BI, salva em pdf.
* **Script Projeto Setor Publico.sql**: Script SQL final, o que foi gerado no MySQL

## 

## 

## ANEXO 1 TABELAS

- Algumas colunas foram alteradas o “Tipo”para facilitar a manipulação das mesmas posteriormente.

**Tabela: recife-dados-receitas-2021**

(\*) Tabelas dimensões, com as colunas complementares (\*\*)

 Colunas necessárias para a análise posterior

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| # | Nome | Tipo | **Alterar Tipo para:** | Agrupamento (Collation) | Atributos | Nulo | Predefinido |
| 1 | ano | bigint(20) | YEAR(4) |  |  | Sim | NULL |
| 2 | mes | bigint(20) | SMALLINT(2) |  |  | Sim | NULL |
| 3 | mes\_nome | text |  | utf8\_general\_ci |  | Sim | NULL |
| 4 | orgao\_codigo (\*) | bigint(20) | SMALLINT(3) |  |  | Sim | NULL |
| 5 | orgao\_nome (\*\*) | text |  | utf8\_general\_ci |  | Sim | NULL |
| 6 | unidade\_codigo | double |  |  |  | Sim | NULL |
| 7 | unidade\_nome | text |  | utf8\_general\_ci |  | Sim | NULL |
| 8 | fonte\_recurso\_codigo (\*) | bigint(20) | SMALLINT(4) |  |  | Sim | NULL |
| 9 | fonte\_recurso\_nome (\*\*) | text |  | utf8\_general\_ci |  | Sim | NULL |
| 10 | categoria\_receita\_codigo (\*) | bigint(20) |  |  |  | Sim | NULL |
| 11 | categoria\_receita\_nome (\*\*) | text |  | utf8\_general\_ci |  | Sim | NULL |
| 12 | fonte\_origem\_receita\_codigo | bigint(20) |  |  |  | Sim | NULL |
| 13 | fonte\_origem\_receita\_nome | text |  | utf8\_general\_ci |  | Sim | NULL |
| 14 | subfonte\_receita\_codigo | bigint(20) |  |  |  | Sim | NULL |
| 15 | subfonte\_receita\_nome | text |  | utf8\_general\_ci |  | Sim | NULL |
| 16 | rubrica\_receita\_codigo | bigint(20) |  |  |  | Sim | NULL |
| 17 | rubrica\_receita\_nome | text |  | utf8\_general\_ci |  | Sim | NULL |
| 18 | alinea\_receita\_codigo | bigint(20) |  |  |  | Sim | NULL |
| 19 | alinea\_receita\_nome | text |  | utf8\_general\_ci |  | Sim | NULL |
| 20 | subalinea\_receita\_codigo | bigint(20) |  |  |  | Sim | NULL |
| 21 | subalinea\_receita\_nome | text |  | utf8\_general\_ci |  | Sim | NULL |
| 22 | receita\_local\_codigo (\*) | bigint(20) |  |  |  | Sim | NULL |
| 23 | receita\_local\_nome (\*\*) | text |  | utf8\_general\_ci |  | Sim | NULL |
| 24 | receita\_prevista | text | DECIMAL(15,2) | utf8\_general\_ci |  | Sim | NULL |
| 25 | receita\_prevista\_acrescimo | text | DECIMAL(15,2) | utf8\_general\_ci |  | Sim | NULL |
| 26 | receita\_prevista\_atualizada | text | DECIMAL(15,2) | utf8\_general\_ci |  | Sim | NULL |
| 27 | receita\_arrecadada | text | DECIMAL(15,2) | utf8\_general\_ci |  | Sim | NULL |

**Tabela: recife-dados-despesas-2021**

(\*) Tabelas dimensões, com as colunas complementares (\*\*)

 Colunas necessárias para a análise posterior

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| # | Nome | Tipo | Alterar Tipo para: | Agrupamento (Collation) | Atributos | Nulo | Predefinido |
| 1 | ano\_movimentacao | bigint(20) | YEAR(4) |  |  | Sim | NULL |
| 2 | mes\_movimentacao | bigint(20) | SMALLINT(2) |  |  | Sim | NULL |
| 3 | orgao\_codigo (\*) | bigint(20) | SMALLINT(3) |  |  | Sim | NULL |
| 4 | orgao\_nome (\*\*) | text |  | utf8\_general\_ci |  | Sim | NULL |
| 5 | unidade\_codigo | double |  |  |  | Sim | NULL |
| 6 | unidade\_nome | text |  | utf8\_general\_ci |  | Sim | NULL |
| 7 | categoria\_economica\_codigo (\*) | bigint(20) |  |  |  | Sim | NULL |
| 8 | categoria\_economica\_nome (\*\*) | text |  | utf8\_general\_ci |  | Sim | NULL |
| 9 | grupo\_despesa\_codigo (\*) | bigint(20) | SMALLINT(2) |  |  | Sim | NULL |
| 10 | grupo\_despesa\_nome (\*\*) | text |  | utf8\_general\_ci |  | Sim | NULL |
| 11 | modalidade\_aplicacao\_codigo (\*) | bigint(20) |  |  |  | Sim | NULL |
| 12 | modalidade\_aplicacao\_nome (\*\*) | text |  | utf8\_general\_ci |  | Sim | NULL |
| 13 | elemento\_codigo | bigint(20) |  |  |  | Sim | NULL |
| 14 | elemento\_nome | text |  | utf8\_general\_ci |  | Sim | NULL |
| 15 | subelemento\_codigo | bigint(20) |  |  |  | Sim | NULL |
| 16 | subelemento\_nome | text |  | utf8\_general\_ci |  | Sim | NULL |
| 17 | funcao\_codigo | double |  |  |  | Sim | NULL |
| 18 | funcao\_nome | text |  | utf8\_general\_ci |  | Sim | NULL |
| 19 | subfuncao\_codigo | double |  |  |  | Sim | NULL |
| 20 | subfuncao\_nome | text |  | utf8\_general\_ci |  | Sim | NULL |
| 21 | programa\_codigo | double |  |  |  | Sim | NULL |
| 22 | programa\_nome | text |  | utf8\_general\_ci |  | Sim | NULL |
| 23 | acao\_codigo | double |  |  |  | Sim | NULL |
| 24 | acao\_nome | text |  | utf8\_general\_ci |  | Sim | NULL |
| 25 | fonte\_recurso\_codigo (\*) | double | SMALLINT(4) |  |  | Sim | NULL |
| 26 | fonte\_recurso\_nome (\*\*) | text |  | utf8\_general\_ci |  | Sim | NULL |
| 27 | empenho\_ano | double |  |  |  | Sim | NULL |
| 28 | empenho\_modalidade\_nome (\*\*) | text |  | utf8\_general\_ci |  | Sim | NULL |
| 29 | empenho\_modalidade\_codigo (\*) | double |  |  |  | Sim | NULL |
| 30 | empenho\_numero | double |  |  |  | Sim | NULL |
| 31 | subempenho | double |  |  |  | Sim | NULL |
| 32 | indicador\_subempenho | text |  | utf8\_general\_ci |  | Sim | NULL |
| 33 | credor\_codigo (\*) | double | SMALLINT(7) |  |  | Sim | NULL |
| 34 | credor\_nome (\*\*) | text |  | utf8\_general\_ci |  | Sim | NULL |
| 35 | modalidade\_licitacao\_codigo (\*) | double | SMALLINT(2) |  |  | Sim | NULL |
| 36 | modalidade\_licitacao\_nome (\*\*) | text |  | utf8\_general\_ci |  | Sim | NULL |
| 37 | valor\_empenhado | text | DECIMAL(15,2) | utf8\_general\_ci |  | Sim | NULL |
| 38 | valor\_liquidado | text | DECIMAL(15,2) | utf8\_general\_ci |  | Sim | NULL |
| 39 | valor\_pago | text | DECIMAL(15,2) | utf8\_general\_ci |  | Sim | NULL |

[**Tabela: dados\_abertos\_aditivos**](http://ip-45-79-142-173.cloudezapp.io/phpmyadmin/sql.php?db=residenciatechad&table=dados_abertos_aditivos)

\*\*\* Tabelas exportadas completas

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| # | Nome | Tipo | Agrupamento (Collation) | Atributos | Nulo | Predefinido |
| 1 | num\_contrato | text | utf8\_general\_ci |  | Sim | NULL |
| 2 | orgaocontratante | text | utf8\_general\_ci |  | Sim | NULL |
| 3 | objetocontrato | text | utf8\_general\_ci |  | Sim | NULL |
| 4 | cnpjcontratado | text | utf8\_general\_ci |  | Sim | NULL |
| 5 | cpfcontratado | text | utf8\_general\_ci |  | Sim | NULL |
| 6 | razaonomecontratado | text | utf8\_general\_ci |  | Sim | NULL |
| 7 | num\_termoaditivo | bigint(20) |  |  | Sim | NULL |
| 8 | justificativa\_termoaditivo | text | utf8\_general\_ci |  | Sim | NULL |
| 9 | data\_inicio\_vigencia | text | utf8\_general\_ci |  | Sim | NULL |
| 10 | data\_fim\_vigencia | text | utf8\_general\_ci |  | Sim | NULL |
| 11 | valor\_termoaditivo | double |  |  | Sim | NULL |
| 12 | situacao\_termoaditivo | text | utf8\_general\_ci |  | Sim | NULL |

**Tabela: dados\_abertos\_contratos**

\*\*\* Tabelas exportadas completas

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| # | Nome | Tipo | Agrupamento (Collation) | Atributos | Nulo | Predefinido |
| 1 | numerocontrato | text | utf8\_general\_ci |  | Sim | NULL |
| 2 | anocontrato | double |  |  | Sim | NULL |
| 3 | orgaocontratante | text | utf8\_general\_ci |  | Sim | NULL |
| 4 | objetocontrato | text | utf8\_general\_ci |  | Sim | NULL |
| 5 | origemcontrato | text | utf8\_general\_ci |  | Sim | NULL |
| 6 | sequencialcompra | double |  |  | Sim | NULL |
| 7 | cnpjcontratado | text | utf8\_general\_ci |  | Sim | NULL |
| 8 | cpfcontratado | text | utf8\_general\_ci |  | Sim | NULL |
| 9 | razaonomecontratado | text | utf8\_general\_ci |  | Sim | NULL |
| 10 | logradouroendcontratado | text | utf8\_general\_ci |  | Sim | NULL |
| 11 | numeroendcontratado | double |  |  | Sim | NULL |
| 12 | complementoendcontratado | text | utf8\_general\_ci |  | Sim | NULL |
| 13 | bairroendcontratado | text | utf8\_general\_ci |  | Sim | NULL |
| 14 | cidadeendcontratado | text | utf8\_general\_ci |  | Sim | NULL |
| 15 | estadoendcontratado | text | utf8\_general\_ci |  | Sim | NULL |
| 16 | datainiciovigencia | text | utf8\_general\_ci |  | Sim | NULL |
| 17 | datafimvigencia | text | utf8\_general\_ci |  | Sim | NULL |
| 18 | valorcontrato1 | double |  |  | Sim | NULL |
| 19 | valorcontrato2 | double |  |  | Sim | NULL |
| 20 | situacaocontrato | text | utf8\_general\_ci |  | Sim | NULL |