

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

Ferramentas e estruturas de tabelas

1. Exportação do banco de dados [residenciatechad](#), disponibilizado pela Harve, para o computador local.
2. Baixado os seguintes programas, nesta ordem:
 - a. [MySQL Workbench](#), verificado e baixado antes os dois programas;
 - i. [Microsoft .NET Framework 4.5](#)
 - ii. [Visual C++ Redistributable for Visual Studio 2019](#)
 - b. [Wampserver](#)
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**: 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.

```
-- Alterando a formatação dos números nas colunas 'valor_empenhado', 'valor_liquidado' e 'valor_pago'
-- Após isso será alterado o tipo de TEXT para DECIMAL.
UPDATE residenciatechad.recife_dados_despesas_2021 SET
    valor_empenhado = REPLACE (valor_empenhado, ',', '.'),
    valor_liquidado = REPLACE (valor_liquidado, ',', '.'),
    valor_pago = REPLACE (valor_pago, ',', '.');
```

- 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;
```

```
USE residenciatechad;
--
-- Estrutura da tabela `recife-dados-despesas-2021`
--
CREATE TABLE `recife-dados-despesas-2021` (
  `ano_movimentacao` bigint(20) DEFAULT NULL,
  `mes_movimentacao` bigint(20) DEFAULT NULL,
  `orgao_codigo` bigint(20) DEFAULT NULL,
  `orgao_nome` text DEFAULT NULL,
```

- 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.

```
-- Verificar a quantidade máxima de caracteres das colunas de valores.
SELECT MAX(length(orgao_codigo)),           -- 2
       MAX(length(fonte_recurso_codigo)),    -- 4
       MAX(length(grupo_despesa_codigo)),    -- 1
       MAX(length(credor_codigo)),           -- 7
       MAX(length(modalidade_licitacao_codigo)) -- 2
FROM residenciatechad.recife_dados_despesas_2021;
```

- 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:
- ◆ ÃÃÃÃÃÃÃ substituído por C (representando o Ç)
 - ◆ ÄÄÄÄÄÄÄ substituído por A (representando o ã)
 - ◆ ÅÅÅÅÅÅÅ 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:

Nome da nova coluna

Grupo_credores

	Nome da Coluna	Operador	Valor ①		Saída ①
Se	credor_codigo	igual a	0	Então	CREDOR NÃO INFORMADO
Senão...	Total do credor	é menor que ou i...	1000000	Então	Abaixo de R\$ 1 Milhão
Senão...	Total do credor	é menor que ou i...	10000000	Então	Entre R\$ 1 e R\$ 10 Milhões
Senão...	Total do credor	é menor que ou i...	50000000	Então	Entre R\$ 10 e R\$ 50 Milhões
Senão...	Total do credor	é menor que ou i...	100000000	Então	Entre R\$ 50 e R\$ 100 Milhões
Senão...	Total do credor	é menor que ou i...	500000000	Então	Entre R\$ 100 e R\$ 500 Milhões
Senão...	Total do credor	é menor que ou i...	1000000000	Então	Entre R\$ 500 Milhões e R\$ 1Bi
Senão...	Total do credor	é maior que ou ig...	1000000000	Então	Acima de R\$ 1Bi

Adicionar Cláusula

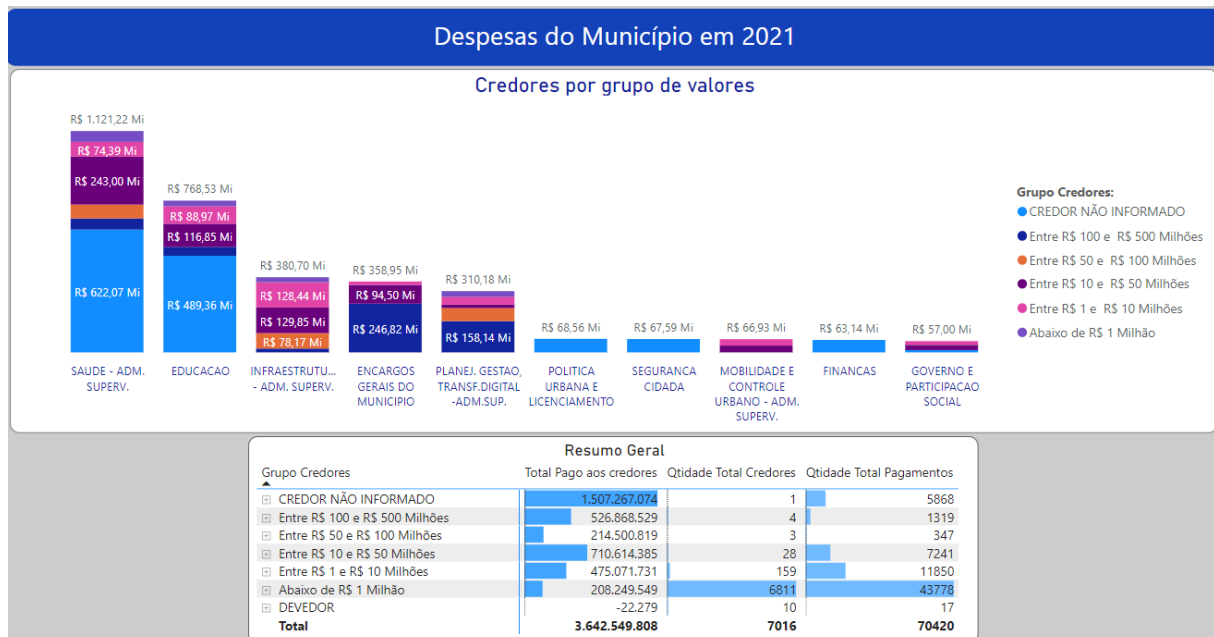
Senão ①

NAO CATEGORIZADO

- Os valores foram classificados nas seguintes categorias:

- ✓ (Selecionar Tudo)
- ✓ Abaixo de R\$ 1 Milhão
- ✓ CREDOR NÃO INFORMADO
- ✓ Entre R\$ 1 e R\$ 10 Milhões
- ✓ Entre R\$ 10 e R\$ 50 Milhões
- ✓ Entre R\$ 100 e R\$ 500 Milhões
- ✓ Entre R\$ 50 e R\$ 100 Milhões

- 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.docx:** 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

*** 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