

STONE_DATA_CHALLENGE_2022

MURILO-FARIAS

Sumário

- **Análise Exploratória com Python.**
- **Banco de dados PostgreSQL.**
- **Power Bi.**
- **Conclusão.**

Análise Exploratória com Python

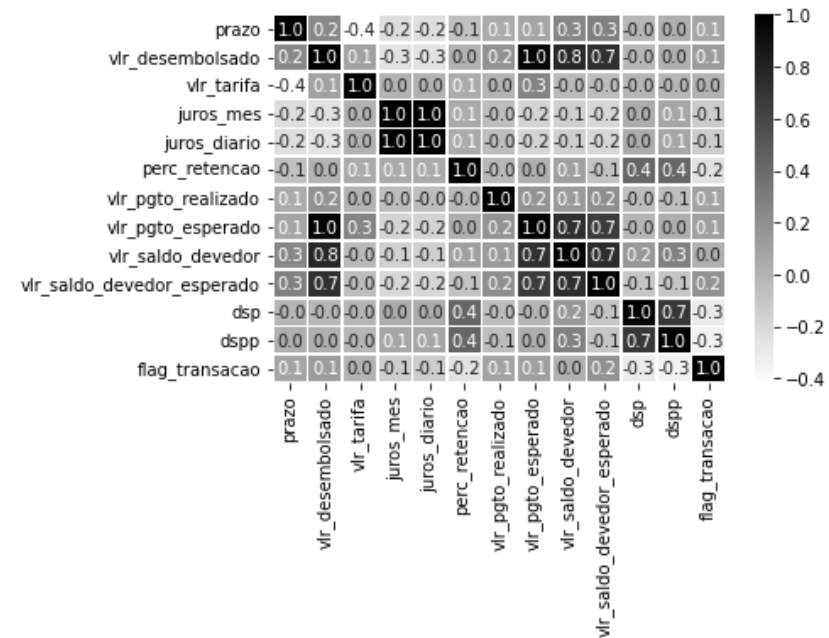
Arquivo: PORTFOLIO_GERAL.csv

```
In [7]: #verificando se existe alguma linha em branco de todas as colunas
PORTFOLIO_GERAL.isnull().sum()
```

```
Out[7]: contrato_id          0
dt_ref_portfolio            0
safra                       0
nr_documento                0
status_contrato             0
dt_contrato                 0
dt_desembolso               0
dt_vencimento               0
dt_wo                       0
prazo                      0
vlr_desembolsado            0
vlr_tarifa                  0
juros_mes                   0
juros_diario                0
perc_retencao               0
vlr_pgto_realizado          0
vlr_pgto_esperado           0
vlr_saldo_devedor           0
vlr_saldo_devedor_esperado  0
dsp                          0
dspp                        0
flag_transacao              0
dtype: int64
```

```
In [11]: #Criando uma copia do arquivo original para o arquivo tratado
PORTFOLIO_GERAL_TRATADO = PORTFOLIO_GERAL
```

```
In [10]: #Gerar tabela correlação para colunas numéricas
corr = PORTFOLIO_GERAL.corr()
plot = sns.heatmap(corr, annot = True, fmt=".1f", linewidths=.6, cmap = "gist_gray_r")
```



```
In [12]: #observação: com base em minha interpretação dos dados exclui 03 colunas

dropcoluns = ['vlr_tarifa', 'juros_mes', 'juros_diario']
PORTFOLIO_GERAL_TRATADO.drop(dropcoluns, axis='columns', inplace=True)
```

Análise Exploratória com Python

Arquivo: PORTFOLIO_TPV.csv

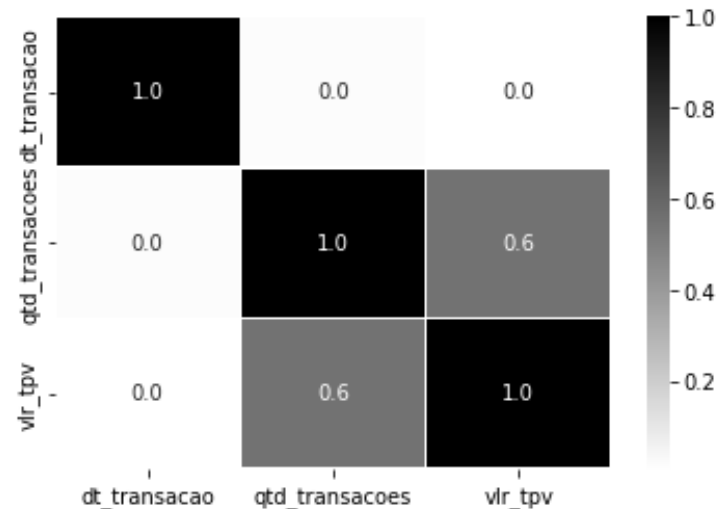
```
In [20]: #Verificando os nomes das colunas  
PORTFOLIO_TPV.columns
```

```
Out[20]: Index(['nr_documento', 'dt_transacao', 'qtd_transacoes', 'vlr_tpv'], dtype='object')
```

```
In [21]: # tipos das colunas do arquivo  
PORTFOLIO_TPV.dtypes
```

```
Out[21]: nr_documento      object  
dt_transacao      int64  
qtd_transacoes    int64  
vlr_tpv           float64  
dtype: object
```

```
In [25]: #Gerar tabela correlação para colunas numéricas  
corr_tpv = PORTFOLIO_TPV.corr()  
corr = PORTFOLIO_GERAL.corr()  
plot = sns.heatmap(corr_tpv, annot = True, fmt=".1f", linewidths=.6, cmap = "gist_gray_r")
```



Análise Exploratória com Python

Arquivo: PORTFOLIO_COMUNICADOS.csv

In [34]:

```
#Criando uma lista personalizada com as colunas escolhidas para contagem de valores
lista_com = ['tipo_acao', 'acao', 'status']

#Realizando a contagem de valores das colunas escolhidas no arquivo PORTFOLIO_GERAL_TRATADO
for c in lista_com:
    print(PORTFOLIO_COMUNICADOS[c].value_counts())
```

```
HSM      201870
EMAIL     201870
Name: tipo_acao, dtype: int64
campanhaobservacao      213334
campanhaparcelamento    89926
campanhaboletaquitado    68370
campanhaprenegativacao    21190
campanhanegativacao      10920
Name: acao, dtype: int64
NAO ENTREGUE    191207
ENTREGUE        137951
LIDO            71380
RESPONDIDO       3202
Name: status, dtype: int64
```

Análise Exploratória com Python

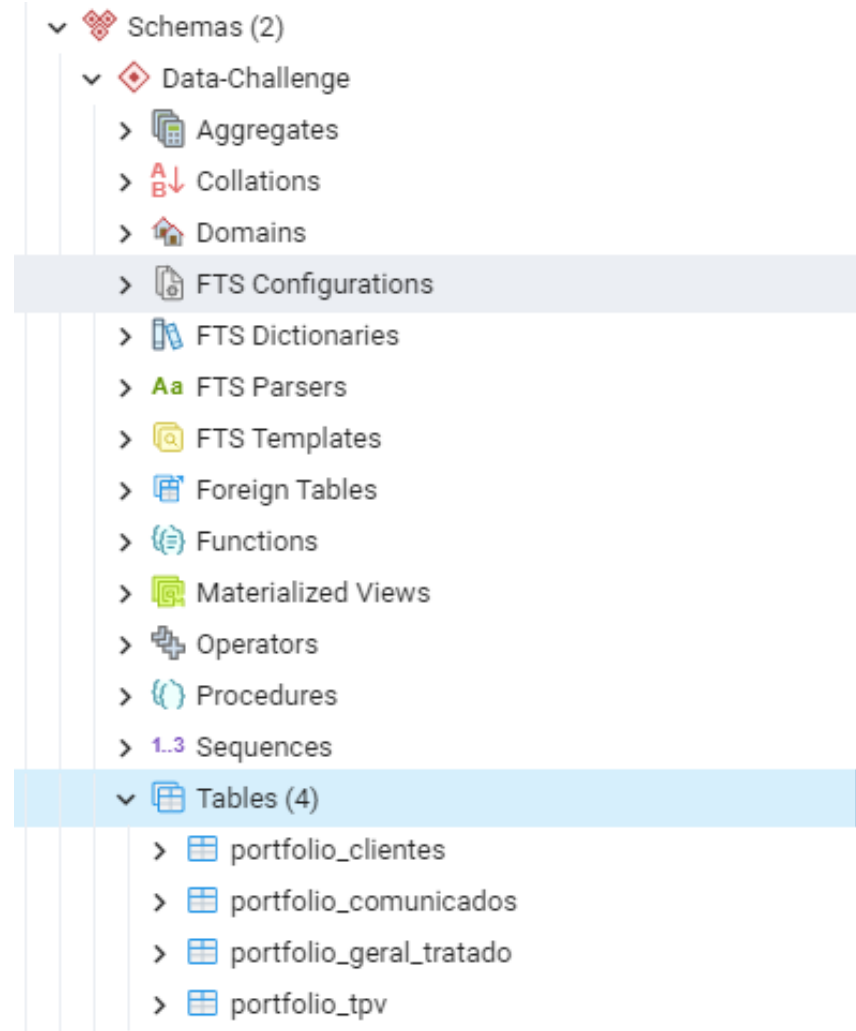
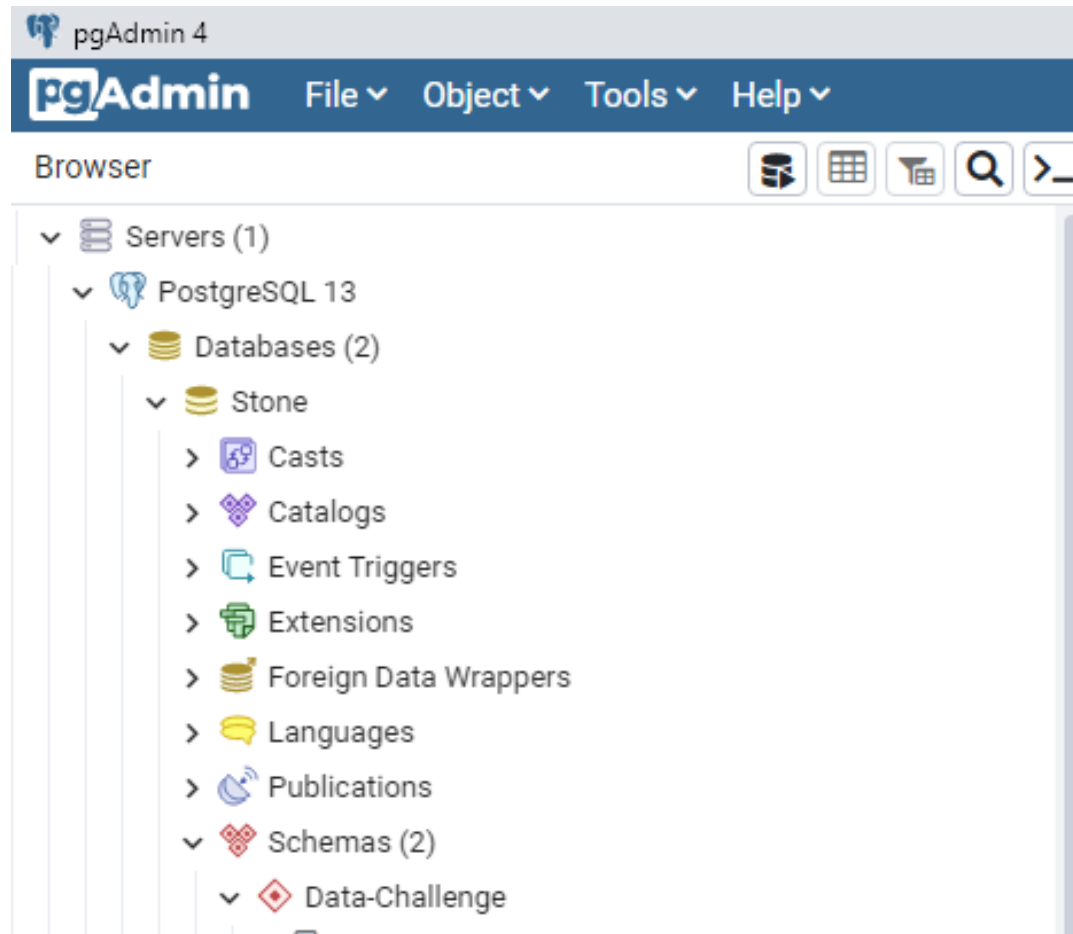
Arquivo: PORTFOLIO_CLIENTES.csv

```
In [36]: #Visualizando as primeiras 10 linhas do arquivo  
PORTFOLIO_CLIENTES.head(10)
```

```
Out[36]:
```

	nr_documento	tipo_empresa	cidade	estado	subsegmento	segmento
0	094aaed740bbd7fcddb6c678c18d883a	PJ	Caucaia	CE	Alimentação Rápida	Alimentação
1	eaade7915611fcdcbba747e3a83365c6	PF	Cidade Ocidental	GO	Lojas Diversas	Varejo
2	1f21958fd296f1b04a1f0c2658913856	MEI	Manaus	AM	Alimentação Rápida	Alimentação
3	fcdfe15db4520db1a1fe8b6a72878655	MEI	São Bernardo do Campo	SP	Alimentação Rápida	Alimentação
4	ba57e8837452b5e0114d7e273de039b4	PJ	Foz do Iguaçu	PR	Lojas Diversas	Supermercado/Farmácia
5	fccfcf9e822150a18a853dc8b8513aa9	PJ	Paranavaí	PR	Automotivo	Bens duráveis
6	f2e61554da65b70f98e36fe7e8a9633a	PJ	Mongaguá	SP	Lazer & Turismo	Viagens e entretenimento
7	ad8c01c0291171a4e9cbe11ef55dc9ca	PF	Rio de Janeiro	RJ	Alimentação Rápida	Alimentação
8	c7892d0f843cc3a0af025986ae7ca7b3	PJ	Pontal do Paraná	PR	Postos de Gasolina	Posto
9	f53fbbf2af08ed10141ed4a8e6e4dc13	PJ	Apucarana	PR	Bares e Restaurantes	Alimentação

Banco de dados PostgreSQL



Banco de dados PostgreSQL







portfolio_clientes

General Columns Advanced Constraints Parameters Security SQL

Inherited from table(s)

Select to inherit from...

Columns

	Name	Data type	Length/Precision	Scale	Not NULL?	Primary key?	Default
 	nr_documento	character varying v	100		<input checked="" type="checkbox"/>	<input type="checkbox"/>	
 	tipo_empresa	character varying v	100		<input checked="" type="checkbox"/>	<input type="checkbox"/>	
 	cidade	character varying v	100		<input checked="" type="checkbox"/>	<input type="checkbox"/>	
 	estado	character varying v	100		<input checked="" type="checkbox"/>	<input type="checkbox"/>	
 	subsegmento	character varying v	100		<input checked="" type="checkbox"/>	<input type="checkbox"/>	
 	segmento	character varying v	100		<input checked="" type="checkbox"/>	<input type="checkbox"/>	

Stone/postgres@PostgreSQL 13 v

Query Editor Query History

Scratch Pad

```
1 select * from "Data-Challenge".portfolio_clientes limit 10;
```

Data Output Explain Messages Notifications

	nr_documento character varying (100)	tipo_empresa character varying (100)	cidade character varying (100)	estado character varying (100)	subsegmento character varying (100)	segmento character varying (100)
1	094aad740bbd7fcd6b6c678c18d883a	PJ	Caucaia	CE	Alimentação Rápida	Alimentação
2	eaade7915611fcdcbba747e3a83365c6	PF	Cidade Ocidental	GO	Lojas Diversas	Varejo
3	1f21958fd296f1b04a1f0c2658913856	MEI	Manaus	AM	Alimentação Rápida	Alimentação
4	fcdfe15db4520db1a1fe8b6a72878655	MEI	São Bernardo do Campo	SP	Alimentação Rápida	Alimentação
5	ba57e8837452b5e0114d7e273de039b4	PJ	Foz do Iguaçu	PR	Lojas Diversas	Supermercado/Farmácia
6	fccfc9e822150a18a853dc8b8513aa9	PJ	Paranavaí	PR	Automotivo	Bens duráveis
7	f2e61554da65b70f98e36fe7e8a9633a	PJ	Mongaguá	SP	Lazer & Turismo	Viagens e entretenimento
8	ad8c01c0291171a4e9cbe11ef55dc9ca	PF	Rio de Janeiro	RJ	Alimentação Rápida	Alimentação
9	c7892d0f843cc3a0af025986ae7ca7b3	PJ	Pontal do Paraná	PR	Postos de Gasolina	Posto
10	f53fbf2af08ed10141ed4a8e6e4dc13	PJ	Apucarana	PR	Bares e Restaurantes	Alimentação

Banco de dados PostgreSQL

portfolio_comunicados

General

Columns

Advanced

Constraints

Parameters

Security

SQL

Inherited from table(s)

Select to inherit from...

Columns

	Name	Data type	Length/Precision	Scale	Not NULL?	Primary key?	Default
	contrato_id	character varying	100		<input checked="" type="checkbox"/>	<input type="checkbox"/>	
	dt_ref_portfolio	date			<input checked="" type="checkbox"/>	<input type="checkbox"/>	
	data_acao	date			<input checked="" type="checkbox"/>	<input type="checkbox"/>	
	tipo_acao	character varying	100		<input checked="" type="checkbox"/>	<input type="checkbox"/>	
	acao	character varying	100		<input type="checkbox"/>	<input type="checkbox"/>	
	status	character varying	100		<input type="checkbox"/>	<input type="checkbox"/>	

Query Editor

Query History

Scratch Pad

Data Output

Explain

Messages

Notifications

	contrato_id character varying (100)	dt_ref_portfolio date	data_acao date	tipo_acao character varying (100)	acao character varying (100)	status character varying (100)
1	84c0045f56d2a9f856b9a35412bc3bb0	2020-11-29	2020-12-01	HSM	campanhaobservacao	ENTREGUE
2	8ad0a8c4ccdb9a95ad996c9a868a9adc	2021-04-06	2021-04-06	HSM	campanhaobservacao	NAO ENTREGUE
3	efbb4f441497b34790c2d87a57002a10	2020-07-15	2020-07-15	HSM	campanhaobservacao	NAO ENTREGUE
4	3a4311a23a7813ce06f920641c63aa9b	2021-03-13	2021-03-15	HSM	campanhaobservacao	NAO ENTREGUE
5	1f3eb24d3aef7bc084046dcaa23e3fa2	2021-03-13	2021-03-15	HSM	campanhaobservacao	NAO ENTREGUE
6	92ea2a1156797ce309dcd633bfb6e169	2021-04-03	2021-04-05	HSM	campanhaobservacao	ENTREGUE
7	ba71e57ce82961a7491523f7fb4345d1	2021-04-03	2021-04-05	HSM	campanhaobservacao	NAO ENTREGUE
8	d3234114bbdca0a89e4d959732d56bf8	2020-07-04	2020-07-06	HSM	campanhaobservacao	NAO ENTREGUE
9	ae240450263b958ab09e6f3c61353fb7	2021-04-01	2021-04-01	HSM	campanhaobservacao	LIDO
10	c9ed1ed80f05c03aa83e4070d0facb07	2022-02-14	2022-02-14	HSM	campanhaobservacao	NAO ENTREGUE

Banco de dados PostgreSQL















portfolio_geral_tratado

General Columns Advanced Constraints Parameters Security SQL

Inherited from table(s)

Select to inherit from...

Columns

	Name	Data type	Length/Precision	Scale	Not NULL?	Primary key?	Default
 	contrato_id	character varying v	100		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
 	dt_ref_portfolio	date v			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
 	safra	character varying v	8		<input checked="" type="checkbox"/>	<input type="checkbox"/>	
 	nr_documento	character varying v			<input checked="" type="checkbox"/>	<input type="checkbox"/>	
 	status_contrato	character varying v			<input checked="" type="checkbox"/>	<input type="checkbox"/>	
 	dt_contrato	date v			<input checked="" type="checkbox"/>	<input type="checkbox"/>	
 	dt_desembolso	date v			<input checked="" type="checkbox"/>	<input type="checkbox"/>	

Query Editor Query History

Scratch Pad

```
1 select * from "Data-Challenge".portfolio_geral_tratado limit 10;
```

Data Output Explain Messages Notifications

	contrato_id [PK] character varying (100)	dt_ref_portfolio [PK] date	safra character varying (8)	nr_documento character varying (100)	status_contrato character varying (100)	dt_contrato date	dt_desembolso date
1	160b4432db8e63c6cc8f69f31a9ab800	2022-04-15	2020-09	92e92d5c9cc76de10722eae05cacf078	Settled	2020-09-17	2020-09-21
2	2dcdb54958f189f59db82deda9568bc3	2022-04-15	2020-02	318b0de2519b64c9d647734d73c739fd	Settled	2020-01-29	2020-02-03
3	c6ee25d320ba5524a4efd4f72bbcc243	2022-03-05	2020-06	83e0560a08a9783208d39104cdf8904	Active	2020-06-09	2020-06-15
4	8d54d1f09f21419c8ac08ed0d6d533ac	2022-03-05	2020-09	aaf45c4d57edca94191209558efb9d27	Settled	2020-09-02	2020-09-04
5	aeb3150bf06e1bf6a15aeb8b38c6a7c4	2022-03-05	2020-12	b5d826b2e2f8005b2d77d5362eb3bd95	Settled	2020-12-23	2020-12-28
6	73f5a1b5be874931fa15290c16a55d28	2020-06-18	2020-04	e91e5bffc250c90e0fe439d69dc28004	Active	2020-04-02	2020-04-06
7	a9ec6117c5541ac55197eb3f0d0665df	2022-03-24	2020-07	4f38f54483bc58f43272497a2c335325	Active	2020-07-27	2020-07-29
8	2e39ec419eb45263c7377adfcacab511	2022-03-24	2020-07	c23655b6a55af9e2d9c962231e609ab1	Settled	2020-07-06	2020-07-08
9	5980679bac790e361521e5874a977834	2022-03-24	2020-06	2786f080d1d99700b83e0c0260177e3b	Settled	2020-06-02	2020-06-05
10	ac281020193babef26308e8e539db38c	2022-03-24	2020-11	4600b69a26c7c240222e0dcc8560a9f	Active	2020-11-17	2020-11-19

Banco de dados PostgreSQL


portfolio_tpv

General Columns Advanced Constraints Parameters Security SQL

Inherited from table(s)

Select to inherit from...

Columns

	Name	Data type	Length/Precision	Scale	Not NULL?	Primary key?	Default
 	nr_documento	character varying v	100		<input checked="" type="checkbox"/>	<input type="checkbox"/>	
 	dt_transacao	date v			<input checked="" type="checkbox"/>	<input type="checkbox"/>	
 	qtd_transacoes	character varying v	100		<input checked="" type="checkbox"/>	<input type="checkbox"/>	
 	vlr_tpv	double precision v			<input checked="" type="checkbox"/>	<input type="checkbox"/>	

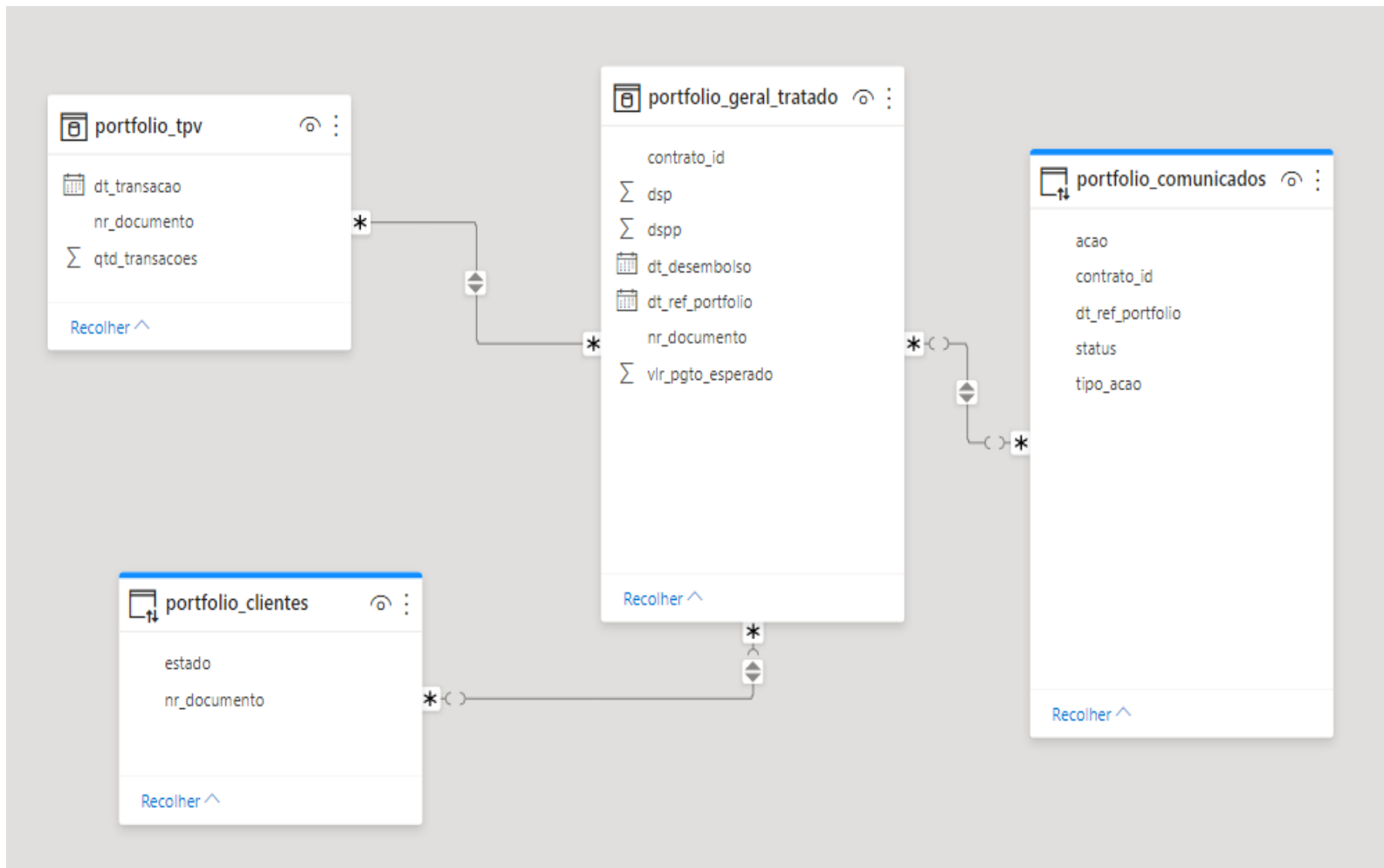
Query Editor Query History

```
1 select * from "Data-Challenge".portfolio_tpv limit 10;
```

Data Output Explain Messages Notifications

	nr_documento character varying (100)	dt_transacao date	qtd_transacoes character varying (100)	vlr_tpv double precision
1	7f0464355620bc7c137b3546eb90867a	2021-05-05	4	116.5
2	7f0464355620bc7c137b3546eb90867a	2021-05-06	20	537
3	7f0464355620bc7c137b3546eb90867a	2021-05-07	17	440
4	7f0464355620bc7c137b3546eb90867a	2021-05-08	29	1328
5	7f0464355620bc7c137b3546eb90867a	2021-05-09	35	969.5
6	7f0464355620bc7c137b3546eb90867a	2021-05-10	18	688
7	7f0464355620bc7c137b3546eb90867a	2021-05-11	7	310.5
8	7f0464355620bc7c137b3546eb90867a	2021-05-12	15	497.5
9	7f0464355620bc7c137b3546eb90867a	2021-05-13	11	396.5
10	7f0464355620bc7c137b3546eb90867a	2021-05-14	22	644.5

Power Bi



Filtros

Pesquis...

Filtros nesta página ...

dspp
não é 0

Tipo de filtro ⓘ

Filtragem básica ▾

☒ Selecionar tudo

<input type="checkbox"/> 0	4750802
<input checked="" type="checkbox"/> 1	852989
<input checked="" type="checkbox"/> 2	677828
<input checked="" type="checkbox"/> 3	211792
<input checked="" type="checkbox"/> 4	119142
<input checked="" type="checkbox"/> 5	81247
<input checked="" type="checkbox"/> 6	64881

☐ Requer seleção única

Adicionar os campos de d...

Filtros em todas as páginas ...

Adicionar os campos de d...

Visualizações

Pesquisar

Valores

Adicionar os campos de da...

Drill-through

Relatório cruzado

Desativado ☐

Manter todos os filtros

Ativado ☒

Adicionar os campos de dr...

Campos

Pesquisar

- portfolio_clientes
 - ☐ estado
 - ☐ nr_documento
- portfolio_comunicados
 - ☐ acao
 - ☐ contrato_id
 - ☐ dt_ref_portfolio
 - ☐ status
 - ☐ tipo_acao
- portfolio_geral_tratado
 - ☐ contrato_id
 - ☐ dsp
 - ☐ dspp
 - ☐ dt_desembolso
 - ☐ dt_ref_portfolio
 - ☐ nr_documento
 - ☐ vlr_pgto_esperado
- portfolio_tpv
 - ☐ dt_transacao
 - ☐ nr_documento
 - ☐ qtd_transacoes

Power Bi

STONE_DATA_CHALLENGE_2022

Murilo Farias

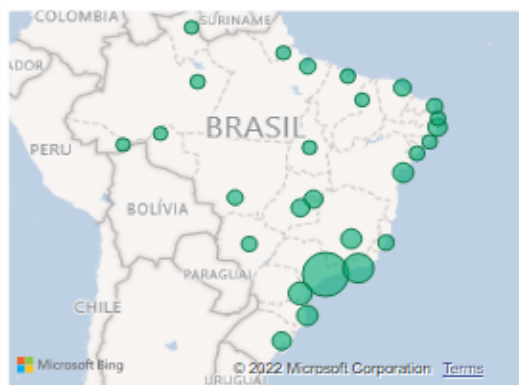
Nº total de contratos:

14756

Valor pagamento esperado:

R\$ 346.268.520,24

Pagamento esperado por estado :



estado vlr_pgto_esperado

SP	R\$ 91.383.051,66
RJ	R\$ 48.006.608,77
PR	R\$ 27.641.383,17
SC	R\$ 20.145.333,31
MG	R\$ 19.385.481,64
BA	R\$ 19.259.856,07
PE	R\$ 14.730.635,03
Total	R\$ 346.268.520,24

campanhaboletaquitado

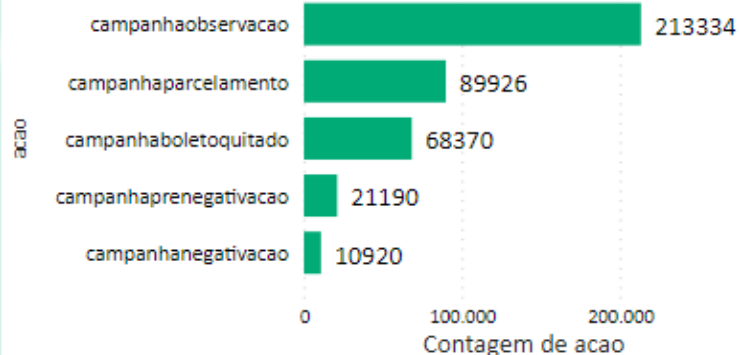
campanhanegativacao

campanhaobservacao

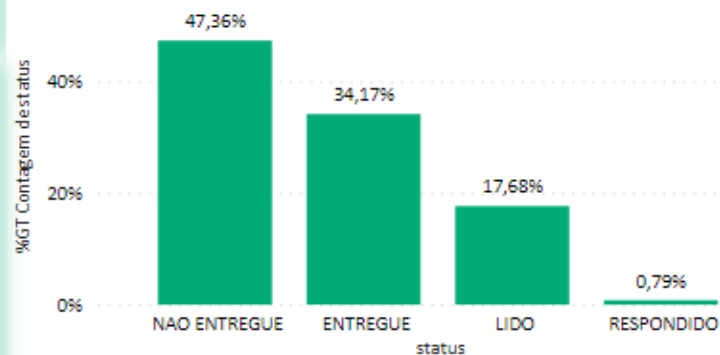
campanhaparcelamento

campanhaprenegativacao

Quantidade de acionamentos:

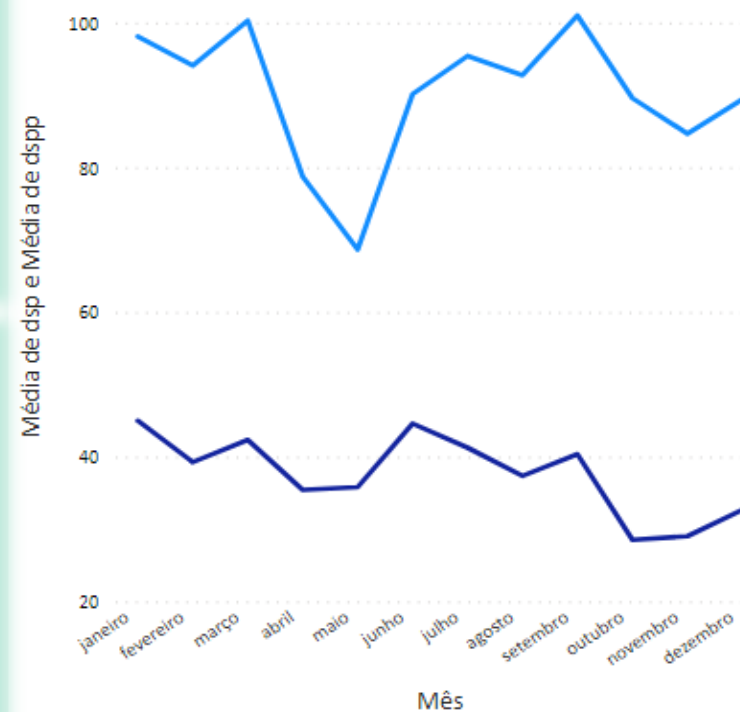


Contagem de status:



Média de dsp e Média de dspp por Mês

● Média de dsp ● Média de dspp



Conclusão

