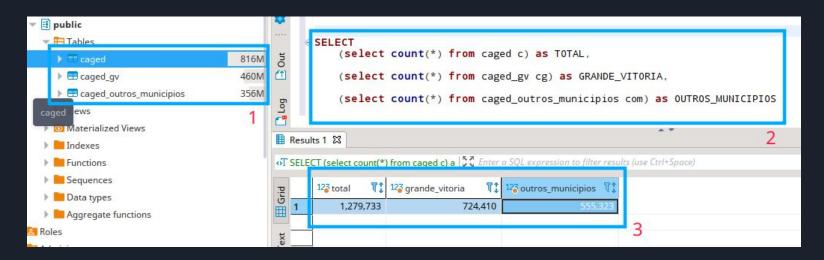
CAGED

Metadados, Estatísticas e Sizing

TAMANHO DA BASE DE DADOS



Criamos uma base de dados com três tabelas a fim de facilitar algumas tarefas, temos então a tabela caged, com todos os dados disponíveis, sendo então a tabela mais completa, em seguida temos a tabela caged_gv que armazena somente os dados relacionados a grande vitória e por fim a tabela caged_outros_municipios que armazena dados dos demais municípios.

Em 1, podemos ver que seus tamanhos são, 816MB, 460MB e 356MB, respectivamente.

Através da query apresentada em 2 podemos ver que a tabela caged possui 1.279.733 registros, enquanto a caged_gv 724.410 e a tabela referente aos demais municípios conta com 555.323 registros, como é mostrado em 3

TAMANHO DA BASE DE DADOS

O tamanho da base de dados também pode ser obtido através da seguinte query:

```
PSELECT

pg_size_pretty (sum( pg_relation_size(C.oid) ))

FROM pg_class C

esults 1 

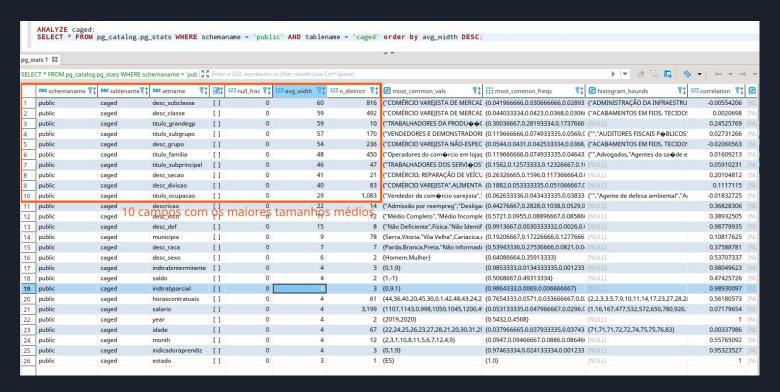
ELECT pg_size_pretty (sum( pg_relation_s | FROM | Enter a SQL expression to filter results

PS pg_size_pretty  

1640 MB
```

Vemos que o tamanho total da base é de 1640 MB, sendo aproximadamente o valor da somatória das três tabelas apresentadas anteriormente

Nulos, Tamanhos médios e o significado do n_distinct.

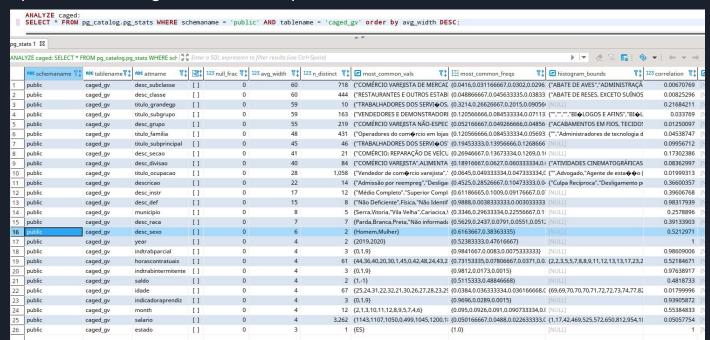


Nulos, Tamanhos médios e o significado do n_distinct.

- Como visto na imagem anterior, não foram encontrados campos nulos para esta base de dados
- Destacamos os 10 campos com os maiores tamanhos médios, que são: desc_subclasse, desc_classe, titulo_grandegp, titulo_subgrupo, desc_grupo, titulo_familia, titulo_subprincipal, desc_secao, desc_divisao, titulo_ocupacao e possuem os seguintes valores para n_distinct: 816.0, 492.0, 10.0, 170.0, 236.0, 450.0, 47.0, 21.0, 83.0, 1083.0, respectivamente.
- Como podemos observar, todos os valores de n_distinct apresentados para estas colunas são inteiros positivos, isto é, maiores que zero, isto significa que a n_distinct está apresentando a quantidade estimada de valores distintos na coluna em questão.

Nulos, Tamanhos médios e o significado do n_distinct. (CAGED_GV)

Apresentamos a seguir a análise feita para as demais tabelas de nossa base de dados



Nulos, Tamanhos médios e o significado do n_distinct. (CAGED_OUTROS_MUNICIPIOS)

Apresentamos a seguir a análise feita para as demais tabelas de nossa base de dados

| ANALYZE caged: SELECT * FROM pg_catalog.pg_stats WHERE schemaname = 'public' AND tablename = 'caged_outros_municipios' order by avg_width DESC; | | | | | | | | | | | |
|--|------------------|------------------|---------------------|----|-----------------|-------------------------------|------------------|---|------------------------------------|--------------------------------|-----------------------|
| • • • • • • • • • • • • • • • • • • • | | | | | | | | | | | |
| pg_stats 1 🗵 | | | | | | | | | | | |
| ANALYZE caged; SELECT * FROM pg_catalog.pg_stats WHERE sct 💸 Enter a SQL expression to filter results (use Ctrl+Space) | | | | | | | | | | | |
| | asc schemaname 🟋 | ABC tablename 1: | ABC attname 73 | | 123 null_frac 🏋 | 123 avg_width \(\tau\) | 123 n_distinct 📆 | ■ most_common_vals | most_common_freqs | ☐ histogram_bounds | T: 123 correlation T: |
| 1 | oublic | caged_outros_mui | titulo_grandegp | [] | 0 | 60 | 8 | {"TRABALHADORES DA PRODU��0 | {0.34083334,0.24116667,0.1343333 | [NULL] | 0.27944553 |
| 2 | oublic | caged_outros_mui | desc_subclasse | [] | 0 | 60 | 718 | {"CULTIVO DE CAFÉ","COMÉRCIO VA | {0.0589,0.0448,0.0263333334,0.0247 | ({"ADMINISTRAÇÃO DE CONSÓ | RCIO: -0.00399933 |
| 3 | oublic | caged_outros_mui | desc_classe | [] | 0 | 58 | 454 | {"CULTIVO DE CAFÉ", "COMÉRCIO VA | {0.0589,0.0454,0.0414,0.032066666 | , {"ACABAMENTOS EM FIOS, TEC | IDO! -0.0053635 |
| 4 | oublic | caged_outros_mui | titulo_subgrupo | [] | 0 | 55 | 169 | {"VENDEDORES E DEMONSTRADOR | {0.11986667,0.0752,0.057533335,0. | ("ADVOGADOS, PROCURADORE | S, TA 0.00180629 |
| 5 1 | oublic | caged_outros_mui | desc_grupo | [] | 0 | 53 | 224 | {"PRODUÇÃO DE LAVOURAS PERMA | {0.079733334,0.0596,0.047966667, | ("ACABAMENTOS EM FIOS, TEC | IDO: -0.05501676 |
| 6 | oublic | caged_outros_mui | titulo_familia | [] | 0 | 49 | 440 | {"Operadores do com@rcio em lojas | {0.11986667,0.057533335,0.037366 | {Administradores,Advogados,"A | Analis -0.0255292 |
| 7 | oublic | caged_outros_mui | titulo_subprincipal | [] | 0 | 48 | 45 | {"VENDEDORES E PRESTADORES DE | {0.12466667,0.1208,0.1165,0.11346 | ("DIRETORES E GERENTES EM E | MPR 0.02296585 |
| 8 | oublic | caged_outros_mui | desc_secao | [] | 0 | 43 | 21 | {"COMÉRCIO; REPARAÇÃO DE VEÍCU | {0.2624,0.22376667,0.13356666,0.0 | [NULL] | 0.26011273 |
| 9 1 | oublic | caged_outros_mu | desc_divisao | [] | 0 | 40 | 82 | {"COMÉRCIO VAREJISTA", "AGRICULT | {0.2008,0.123,0.052533332,0.05073 | ("ATIVIDADES ARTÍSTICAS, CRI | ATIV: 0.17312422 |
| 10 | oublic | caged_outros_mui | titulo_ocupacao | [] | 0 | 31 | 1,039 | {"Vendedor de com�rcio varejista", | {0.06553333,0.0368,0.036266666,0. | ("Acabador de cal@ados","Age | nte d -0.05579066 |
| 11 | public | caged_outros_mui | descricao | [] | 0 | 22 | 13 | {"Admissão por reempreg","Desliga | {0.44023332,0.28143334,0.0951333 | [NULL] | 0.35827065 |
| 12 | oublic | caged_outros_mui | desc_instr | [] | 0 | 17 | 12 | {"Médio Completo", "Fundamental Co | {0.5326,0.10723333,0.0971,0.08583 | [NULL] | 0.37647223 |
| 13 | oublic | caged_outros_mui | desc_def | [] | 0 | 15 | 8 | {"Não Deficiente", Física, "Não Identif | {0.9943333,0.0023,0.0011666666,0. | [NULL] | 0.99162418 |
| 14 | oublic | caged_outros_mui | municipio | [] | 0 | 12 | 73 | {Aracruz,Linhares,"Cachoeiro de Ita | {0.13403334,0.12626667,0.0947,0.0 | [NULL] | 0.03839273 |
| 15 | oublic | caged_outros_mui | desc_raca | [] | 0 | 7 | 7 | {Parda,Branca,Preta,"Não informada | {0.50276667,0.32836667,0.0850666 | [NULL] | 0.35666418 |
| 16 | public | caged_outros_mui | desc_sexo | [] | 0 | 6 | 2 | {Homem,Mulher} | {0.6806,0.3194} | | 0.56500894 |
| 17 | oublic | caged_outros_mui | year | [] | 0 | 4 | 2 | {2019,2020} | {0.57233334,0.42766666} | | 1 |
| 18 | oublic | caged_outros_mui | month | [] | 0 | 4 | 12 | {5,2,3,7,1,10,8,6,4,9,11,12} | {0.10803334,0.09663333,0.0917333 | [NULL] | 0.52480167 |
| 19 | oublic | caged_outros_mui | indicadoraprendiz | [] | 0 | 4 | 3 | {0,1,9} | {0.984,0.015566667,0.00043333} | | 0.96912462 |
| 20 | public | caged_outros_mui | indtrabparcial | [] | 0 | 4 | 3 | {0,9,1} | {0.98826665,0.006966667,0.004766 | [NULL] | 0.99045074 |
| 21 | oublic | caged_outros_mui | indtrabintermitente | [] | 0 | 4 | 3 | {0,1,9} | {0.989,0.010533334,0.00046667} | | 0.98303467 |
| 22 | oublic | caged_outros_mui | horascontratuais | [] | 0 | 4 | 61 | {44,36,40,20,45,48,0,1,42,43,22,30,2 | {0.81913334,0.031166667,0.023166 | [NULL] | 0.66387236 |
| 23 | public | caged_outros_mui | saldo | [] | 0 | 4 | 2 | {1,-1} | {0.5084,0.4916} | | 0.48405066 |
| 24 | public | caged_outros_mui | idade | [] | 0 | 4 | 63 | {21,23,20,22,25,24,26,28,27,29,19,30 | {0.03863333,0.038466666,0.038366 | {68,68,68,70,71,72,73,82} | -0.02430368 |
| 25 | oublic | caged_outros_mui | salario | [] | 0.00003333 | 4 | 2,828 | {1107,1143,0,998,1045,1023,1050,1 | {0.0584,0.052833334,0.041633334, | {1,16,300,477,533,572,624,741, | 899,9 0.08610336 |
| 26 | oublic | caged_outros_mui | estado | [] | 0 | 3 | 1 | {ES} | {1.0} | [NULL] | 1 |
| | | | | | | | | | | | |

Nulos, Tamanhos médios e o significado do n_distinct.

Como visto, os resultados são similares, visto que as demais tabelas são subsets da primeira, por tanto a forma de análise é a mesma