



CAGED

Metadados, Estatísticas e Sizing

TAMANHO DA BASE DE DADOS

The screenshot displays a database management interface. On the left, a tree view shows the 'public' schema with a 'Tables' folder. Three tables are listed: 'caged' (816M), 'caged_gv' (460M), and 'caged_outros_municipios' (356M). A red '1' is placed next to this list. In the center, a SQL query is shown in a text editor, enclosed in a blue box. The query is:

```
SELECT
  (select count(*) from caged c) as TOTAL,
  (select count(*) from caged_gv cg) as GRANDE_VITORIA,
  (select count(*) from caged_outros_municipios com) as OUTROS_MUNICIPIOS
```

 A red '2' is placed next to the query. Below the query, the 'Results' tab shows a table with three columns: 'total', 'grande_vitoria', and 'outros_municipios'. The first row contains the values 1,279,733, 724,410, and 555,323. A red '3' is placed next to the results table.

	total	grande_vitoria	outros_municipios
1	1,279,733	724,410	555,323

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:

```
SELECT
    pg_size_pretty (sum( pg_relation_size(C.oid) ))
FROM pg_class C
```

results 1

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

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.

```
ANALYZE caged;
SELECT * FROM pg_catalog.pg_stats WHERE schemaname = 'public' AND tablename = 'caged' order by avg_width DESC;
```

pg_stats 1

SELECT * FROM pg_catalog.pg_stats WHERE schemaname = 'pub' Enter a SQL expression to filter results (use Ctrl+Space)

	adc	schemaname	adc	tablename	adc	attname	123	null_frac	123	avg_width	123	n_distinct	most_common_vals	most_common_freqs	histogram_bounds	123	correlation
1	public	caged	adc	desc_subclasse	[]		0		60	816		816	{("COMERCIO VAREJISTA DE MERCAL", {0.0419666666,0.0306666666,0.02893	{("COMERCIO VAREJISTA DE MERCAL", {0.0419666666,0.0306666666,0.02893	{("ADMINISTRAÇÃO DA INFRAESTRU	-0.00554206	[NU]
2	public	caged	adc	desc_classe	[]		0		59	492		492	{("COMÉRCIO VAREJISTA DE MERCAL", {0.0440333334,0.0423,0.0368,0.0306	{("COMÉRCIO VAREJISTA DE MERCAL", {0.0440333334,0.0423,0.0368,0.0306	{("ACABAMENTOS EM FIOS, TECIDO!	0.0020698	[NU]
3	public	caged	adc	titulo_grandegp	[]		0		59	10		10	{("TRABALHADORES DA PRODU	{("TRABALHADORES DA PRODU	{(0.24525769	[NU]
4	public	caged	adc	titulo_subgrupo	[]		0		57	170		170	{("VENDEDORES E DEMONSTRADORI", {0.1196666666,0.074933335,0.0569,C	{("VENDEDORES E DEMONSTRADORI", {0.1196666666,0.074933335,0.0569,C	{("AUDITORES FISCAIS P	0.02731266	[NU]
5	public	caged	adc	desc_grupo	[]		0		54	236		236	{("COMÉRCIO VAREJISTA NÃO-ESPEC", {0.0544,0.0431,0.042533334,0.0368,	{("COMÉRCIO VAREJISTA NÃO-ESPEC", {0.0544,0.0431,0.042533334,0.0368,	{("ACABAMENTOS EM FIOS, TECIDO!	-0.02060563	[NU]
6	public	caged	adc	titulo_familia	[]		0		48	450		450	{("Operadores do comércio em lojas", {0.1196666666,0.074933335,0.04643	{("Operadores do comércio em lojas", {0.1196666666,0.074933335,0.04643	{("Advogados,"Agentes da sa	0.01609213	[NU]
7	public	caged	adc	titulo_subprincipal	[]		0		46	47		47	{("TRABALHADORES DOS SERVI	{("TRABALHADORES DOS SERVI	{(0.05910231	[NU]
8	public	caged	adc	desc_secao	[]		0		41	21		21	{("COMÉRCIO: REPARAÇÃO DE VEÍCUL", {0.26326665,0.1596,0.117366664,0,1	{("COMÉRCIO: REPARAÇÃO DE VEÍCUL", {0.26326665,0.1596,0.117366664,0,1	{(0.20104812	[NU]
9	public	caged	adc	desc_divisao	[]		0		40	83		83	{("COMÉRCIO VAREJISTA",ALIMENTA", {0.1882,0.053333335,0.051066667,C	{("COMÉRCIO VAREJISTA",ALIMENTA", {0.1882,0.053333335,0.051066667,C	{(0.1117115	[NU]
10	public	caged	adc	titulo_ocupacao	[]		0		29	1,083		1,083	{("Vendedor de comércio varejista", {0.062633336,0.043433335,0.03833	{("Vendedor de comércio varejista", {0.062633336,0.043433335,0.03833	{(","Agente de defesa ambiental",A	-0.01832725	[NU]
11	public	caged	adc	descricao	[]		0		22	14		14	{("Admissão por reempreg", "Desligam", {0.44276667,0.2828,0.1038,0.0529,0	{("Admissão por reempreg", "Desligam", {0.44276667,0.2828,0.1038,0.0529,0	{(0.36828306	[NU]
12	public	caged	adc	desc_instr	[]		0		19	72		72	{("Médio Completo", "Médio Incompleto", {0.5721,0.0955,0.08896667,0.08586	{("Médio Completo", "Médio Incompleto", {0.5721,0.0955,0.08896667,0.08586	{(0.38932505	[NU]
13	public	caged	adc	desc_def	[]		0		15	8		8	{("Não Deficiente",Física,"Não Identif", {0.9913667,0.003033332,0.0026,0,1	{("Não Deficiente",Física,"Não Identif", {0.9913667,0.003033332,0.0026,0,1	{(0.98779935	[NU]
14	public	caged	adc	municipio	[]		0		9	78		78	{("Serra,Vitoria,"Vila Velha",Caracica, {0.19206667,0.17226666,0.1277666	{("Serra,Vitoria,"Vila Velha",Caracica, {0.19206667,0.17226666,0.1277666	{(0.10817625	[NU]
15	public	caged	adc	desc_raca	[]		0		7	7		7	{("Parda,Branca,Preta,"Não informada", {0.53943336,0.27536666,0.0821,0,0	{("Parda,Branca,Preta,"Não informada", {0.53943336,0.27536666,0.0821,0,0	{(0.37588781	[NU]
16	public	caged	adc	desc_sexo	[]		0		6	2		2	{("Homem,Mulher", {0.64086664,0.35913333}, {0.64086664,0.35913333}, {	{("Homem,Mulher", {0.64086664,0.35913333}, {0.64086664,0.35913333}, {	{(0.53707337	[NU]
17	public	caged	adc	indrabintermitente	[]		0		4	3		3	{(0,1,9}, {0.9853333,0.013433335,0.001233}, {0.9853333,0.013433335,0.001233}, {	{(0,1,9}, {0.9853333,0.013433335,0.001233}, {0.9853333,0.013433335,0.001233}, {	{(0.98049623	[NU]
18	public	caged	adc	saldo	[]		0		4	2		2	{(1,-1}, {0.5068667,0.49313334}, {0.5068667,0.49313334}, {	{(1,-1}, {0.5068667,0.49313334}, {0.5068667,0.49313334}, {	{(0.47425726	[NU]
19	public	caged	adc	indrabparcial	[]		0		4	3		3	{(0,9,1}, {0.9864333,0.0069,0.006666667}, {0.9864333,0.0069,0.006666667}, {	{(0,9,1}, {0.9864333,0.0069,0.006666667}, {0.9864333,0.0069,0.006666667}, {	{(0.98930097	[NU]
20	public	caged	adc	horascontratuais	[]		0		4	61		61	{(44,36,40,20,45,30,0,1,42,48,43,24,2	{(44,36,40,20,45,30,0,1,42,48,43,24,2	{(2,2,3,3,5,7,9,10,11,14,17,23,27,28,2	0.56180573	[NU]
21	public	caged	adc	salario	[]		0		4	3,199		3,199	{(1107,1143,0,998,1050,1045,1200,4	{(1107,1143,0,998,1050,1045,1200,4	{(1,16,167,477,532,572,650,780,926,	0.07179654	[NU]
22	public	caged	adc	year	[]		0		4	2		2	{(2019,2020}, {0.5432,0.4568}, {0.5432,0.4568}, {	{(2019,2020}, {0.5432,0.4568}, {0.5432,0.4568}, {	{(1	[NU]
23	public	caged	adc	idade	[]		0		4	67		67	{(22,24,25,26,23,27,28,21,20,30,31,2	{(22,24,25,26,23,27,28,21,20,30,31,2	{(7,17,71,72,72,74,75,75,76,83}	0.00337986	[NU]
24	public	caged	adc	month	[]		0		4	12		12	{(2,3,1,10,8,11,5,6,7,12,4,9}, {0.0947,0.09466667,0.0886,0.08646	{(2,3,1,10,8,11,5,6,7,12,4,9}, {0.0947,0.09466667,0.0886,0.08646	{(0.55765092	[NU]
25	public	caged	adc	indicadoraprendiz	[]		0		4	3		3	{(0,1,9}, {0.97463334,0.024133334,0.001233}, {0.97463334,0.024133334,0.001233}, {	{(0,1,9}, {0.97463334,0.024133334,0.001233}, {0.97463334,0.024133334,0.001233}, {	{(0.95323527	[NU]
26	public	caged	adc	estado	[]		0		3	1		1	{(ES)}, {1,0}, {1,0}, {	{(ES)}, {1,0}, {1,0}, {	{(1	[NU]



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_grandep, 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

```
ANALYZE caged;
SELECT * FROM pg_catalog.pg_stats WHERE schemaname = 'public' AND tablename = 'caged_gv' order by avg_width DESC;
```

pg_stats 1

ANALYZE caged; SELECT * FROM pg_catalog.pg_stats WHERE scl

	asc schemaname	asc tablename	asc attname	123 null_frac	123 avg_width	123 n_distinct	most_common_vals	most_common_freqs	histogram_bounds	123 correlation
1	public	caged_gv	desc_subclasse	[]	0	60	718 {"COMÉRCIO VAREJISTA DE MERCAD	{0.0416,0.031166667,0.0302,0.0296	{{"ABATE DE AVES","ADMINISTRAÇÃ	0.00670769
2	public	caged_gv	desc_classe	[]	0	60	444 {"RESTAURANTES E OUTROS ESTABI	{0.048866667,0.045633335,0.03833	{{"ABATE DE RESES, EXCETO SUÍNOS	0.00825296
3	public	caged_gv	titulo_grandepg	[]	0	59	10 {"TRABALHADORES DOS SERVIÇOS	{0.3214,0.26626667,0.2015,0.09056	{[NULL]}	0.21684211
4	public	caged_gv	titulo_subgrupo	[]	0	59	163 {"VENDEDORES E DEMONSTRADORI	{0.120566666,0.084533334,0.07113	{{"","","BIBLOGOS E AFINS","BIBL	0.033769
5	public	caged_gv	desc_grupo	[]	0	55	219 {"COMÉRCIO VAREJISTA NÃO-ESPEC	{0.052166667,0.049266666,0.04856	{{"ACABAMENTOS EM FIOS, TECIDO: d	0.01250097
6	public	caged_gv	titulo_familia	[]	0	48	431 {"Operadores do comércio em lojas	{0.120566666,0.084533334,0.05693	{{"","Administradores de tecnologia d	0.04538747
7	public	caged_gv	titulo_subprincipal	[]	0	45	46 {"TRABALHADORES DOS SERVIÇOS	{0.19453333,0.13956666,0.1268666	{[NULL]}	0.09956712
8	public	caged_gv	desc_secao	[]	0	41	21 {"COMÉRCIO; REPARAÇÃO DE VEÍCUL	{0.26946667,0.13673334,0.1269,0.11	{[NULL]}	0.17302386
9	public	caged_gv	desc_divisao	[]	0	40	84 {"COMÉRCIO VAREJISTA",ALIMENTA	{0.18916667,0.0627,0.060333334,0.1	{{"ATIVIDADES CINEMATOGRAFICAS	0.08362997
10	public	caged_gv	titulo_ocupacao	[]	0	28	1,058 {"Vendedor de comércio varejista";	{0.0645,0.049333334,0.047333334,C	{{"Advogado","Agente de estaçã	0.01999313
11	public	caged_gv	descricao	[]	0	22	14 {"Admissão por reemprego","Desliga	{0.4525,0.28526667,0.10473333,0.0	{{"Culpa Recíproca","Desligamento p	0.36600357
12	public	caged_gv	desc_instr	[]	0	17	12 {"Médio Completo","Superior Compl	{0.61186665,0.1009,0.09176667,0.0	{[NULL]}	0.39606768
13	public	caged_gv	desc_def	[]	0	15	8 {"Não Deficiente","Física","Não Identif	{0.9888,0.003833333,0.003033333	{[NULL]}	0.98317939
14	public	caged_gv	municipio	[]	0	8	5 {"Serra,Vitoria","Vila Velha","Cariacica,	{0.3346,0.29633334,0.22556667,0.1	{[NULL]}	0.5278896
15	public	caged_gv	desc_raca	[]	0	7	7 {"Parda,Branca,Preta,"Não informada	{0.5629,0.2437,0.0791,0.0551,0.051	{[NULL]}	0.39133903
16	public	caged_gv	desc_sexo	[]	0	6	2 {"Homem,Mulher}	{0.6163667,0.38363335}	{[NULL]}	0.5212971
17	public	caged_gv	year	[]	0	4	2 {"2019,2020}	{0.52383333,0.47616667}	{[NULL]}	1
18	public	caged_gv	indtrabparcial	[]	0	4	3 {"0,1,9}	{0.9841667,0.0083,0.007533333}	{[NULL]}	0.98609006
19	public	caged_gv	horascontratuais	[]	0	4	61 {"44,36,40,20,30,1,45,0,42,48,24,43,2	{0.73153335,0.07806667,0.0371,0.0	{2,2,3,5,7,8,8,9,11,12,13,13,17,23,2	0.52184671
20	public	caged_gv	indtrabintermitente	[]	0	4	3 {"0,1,9}	{0.9812,0.0173,0.0015}	{[NULL]}	0.97638917
21	public	caged_gv	saldo	[]	0	4	2 {"1,-1}	{0.5115333,0.48846668}	{[NULL]}	0.4818733
22	public	caged_gv	idade	[]	0	4	67 {"25,24,31,22,32,21,30,26,27,28,23,25	{0.0384,0.036333334,0.036166668,C	{69,69,70,70,70,71,72,72,73,74,77,8	0.01799996
23	public	caged_gv	indicadoraprendiz	[]	0	4	3 {"0,1,9}	{0.9696,0.0289,0.0015}	{[NULL]}	0.93905872
24	public	caged_gv	month	[]	0	4	12 {"2,1,3,10,11,12,8,9,5,7,4,6}	{0.095,0.0926,0.091,0.090733334,0.1	{[NULL]}	0.55384833
25	public	caged_gv	salario	[]	0	4	3,262 {"1143,1107,1050,0,499,1045,1200,1	{0.050166667,0.0488,0.022633333,C	{1,17,42,469,525,572,650,812,954,11	0.05057754
26	public	caged_gv	estado	[]	0	3	1 {"ES}	{1,0}	{[NULL]}	1

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 sch

	asc schemaname	asc tablename	asc attname	123 null_frac	123 avg_width	123 n_distinct	most_common_vals	most_common_freqs	histogram_bounds	123 correlation
1	public	caged_outros_mui_titulo_grandepg	[]	0	60	8	{'TRABALHADORES DA PRODUÇÃO', 'COMÉRCIO VA	{0.34083334,0.24116667,0.1343333}	[NULL]	0.27944553
2	public	caged_outros_mui_desc_subclasse	[]	0	60	718	{'CULTIVO DE CAFÉ','COMÉRCIO VA	{0.0589,0.0448,0.02633334,0.0247	{'ADMINISTRAÇÃO DE CONSÓRCIO'	-0.00399933
3	public	caged_outros_mui_desc_classe	[]	0	58	454	{'CULTIVO DE CAFÉ','COMÉRCIO VA	{0.0589,0.0454,0.0414,0.032066666,	{'ACABAMENTOS EM FIOS, TECIDO'	-0.0053635
4	public	caged_outros_mui_titulo_subgrupo	[]	0	55	169	{'VENDEDORES E DEMONSTRADORI	{0.11986667,0.0752,0.057533335,0.1	{'ADVOGADOS, PROCURADORES, T'	0.00180629
5	public	caged_outros_mui_desc_grupo	[]	0	53	224	{'PRODUÇÃO DE LAVOURAS PERMA	{0.079733334,0.0596,0.047966667,C	{'ACABAMENTOS EM FIOS, TECIDO'	-0.05501676
6	public	caged_outros_mui_titulo_familia	[]	0	49	440	{'Operadores do comércio em lojas	{0.11986667,0.057533335,0.037366	{Administradores,Advogados,"Analís	-0.0255292
7	public	caged_outros_mui_titulo_subprincipal	[]	0	48	45	{'VENDEDORES E PRESTADORES DE	{0.12466667,0.1208,0.1165,0.11346}	{'DIRETORES E GERENTES EM EMPR	0.02296585
8	public	caged_outros_mui_desc_secao	[]	0	43	21	{'COMÉRCIO, REPARAÇÃO DE VEÍCUL	{0.2624,0.22376667,0.13356666,0.0	[NULL]	0.26011273
9	public	caged_outros_mui_desc_divisao	[]	0	40	82	{'COMÉRCIO VAREJISTA','AGRICULT	{0.2008,0.123,0.052533332,0.05073}	{'ATIVIDADES ARTÍSTICAS, CRIATIV	0.17312422
10	public	caged_outros_mui_titulo_ocupacao	[]	0	31	1,039	{'Vendedor de comércio varejista',	{0.06553333,0.0368,0.036266666,0.1	{'Acabador de calçados',"Agente d	-0.05579066
11	public	caged_outros_mui_descricao	[]	0	22	13	{'Admissão por reempreg',"Desligai	{0.44023332,0.28143334,0.0951333}	[NULL]	0.35827065
12	public	caged_outros_mui_desc_instr	[]	0	17	12	{'Médio Completo',"Fundamental C	{0.5326,0.10723333,0.0971,0.08583}	[NULL]	0.37647223
13	public	caged_outros_mui_desc_def	[]	0	15	8	{'Não Deficiente',"Física,"Não Identif	{0.9943333,0.0023,0.0011666666,0.1	[NULL]	0.99162418
14	public	caged_outros_mui_municipio	[]	0	12	73	{Aracruz,Linhares,"Cachoeiro de Itaj	{0.13403334,0.12626667,0.0947,0.0	[NULL]	0.03839273
15	public	caged_outros_mui_desc_raca	[]	0	7	7	{Parda,Branca,Preta,"Não informad	{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}	[NULL]	0.56500894
17	public	caged_outros_mui_year	[]	0	4	2	{2019,2020}	{0.57233334,0.42766666}	[NULL]	1
18	public	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	public	caged_outros_mui_indicadoraprendiz	[]	0	4	3	{0,1,9}	{0.984,0.015566667,0.00043333}	[NULL]	0.96912462
20	public	caged_outros_mui_indtrabparcial	[]	0	4	3	{0,9,1}	{0.98826665,0.006966667,0.004766	[NULL]	0.99045074
21	public	caged_outros_mui_indtrabintermitente	[]	0	4	3	{0,1,9}	{0.989,0.010533334,0.00046667}	[NULL]	0.98303467
22	public	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}	[NULL]	0.48405066
24	public	caged_outros_mui_idade	[]	0	4	63	{21,23,20,22,25,24,26,28,27,29,19,3	{0.03863333,0.038466666,0.038366	{68,68,68,70,71,72,73,82}	-0.02430368
25	public	caged_outros_mui_salario	[]	0.00003333	4	2,828	{1107,1143,0.998,1045,1023,1050,1	{0.0584,0.052833334,0.041633334,C	{1,16,300,477,533,572,624,741,899,	0.08610336
26	public	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