

p.s: as questões grifadas de cinza eu fiz com auxílio, não fiz só

p.s2: as questões de amarelo as fiz sozinha

p.s3: as de vermelho não consegui chegar na resposta

## Análise exploratória de dados - SQL

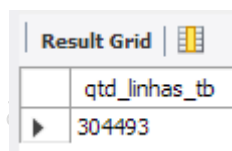
**Contexto:** Uma empresa de telemedicina vende laudos de exames de diversas especialidades para clínicas e hospitais. Cada laudo é realizado por um médico. Mensalmente é enviado para cada cliente um faturamento com a soma do valor de todos os laudos realizados.

Analise a base para responder às seguintes questões (Printe a saída do SQL para colar abaixo de cada questão; elas podem ser complementadas com texto também):

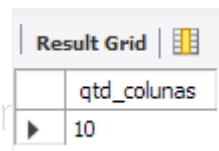
- Quantas linhas e colunas existem na base?

```
SELECT COUNT(client) AS qtd_linhas_tb FROM ocupacao; -- linhas
```

```
SELECT COUNT(*) AS qtd_colunas
FROM information_schema.COLUMNS
WHERE TABLE_NAME = 'ocupacao'; -- colunas
```



qtd_linhas_tb
304493



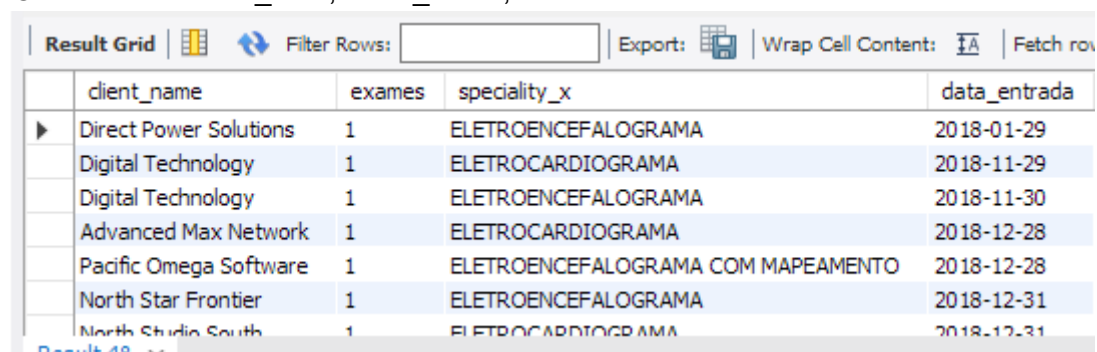
qtd_colunas
10

<https://www.github.com/iascfurtado>

- 
- É possível classificar as colunas em entidades/grupos?

Acredito que sim, se for assim??

```
SELECT client_name, COUNT(DISTINCT speciality_x) AS exames, speciality_x,
CAST(create_date AS DATE) AS data_entrada
FROM ocupacao
GROUP BY 4, speciality_x, client_name
ORDER BY create_date, client_name;
```



client_name	exames	speciality_x	data_entrada
Direct Power Solutions	1	ELETROENCEFALOGRAMA	2018-01-29
Digital Technology	1	ELETROCARDIOGRAMA	2018-11-29
Digital Technology	1	ELETROENCEFALOGRAMA	2018-11-30
Advanced Max Network	1	ELETROCARDIOGRAMA	2018-12-28
Pacific Omega Software	1	ELETROENCEFALOGRAMA COM MAPEAMENTO	2018-12-28
North Star Frontier	1	ELETROENCEFALOGRAMA	2018-12-31
North Studio South	1	ELETROCARDIOGRAMA	2018-12-31

o resultado ficou maior que esse, era isso?

- 
- Qual o exame mais recente e o mais antigo na base?

```
SELECT speciality_x, CAST(create_date AS DATE)
FROM ocupacao
GROUP BY create_date
ORDER BY create_date ASC
LIMIT 3; -- mais antigo
```

	speciality_x	CAST(create_date AS DATE)
▶	ELETROENCEFALOGRAMA	2018-01-29
	ELETROCARDIOGRAMA	2018-11-29
	ELETROENCEFALOGRAMA	2018-11-30

```
SELECT speciality_x, create_date
FROM ocupacao
ORDER BY create_date DESC
LIMIT 3; -- mais recente
```

	speciality_x	CAST(create_date AS DATE)
▶	ELETROENCEFALOGRAMA	2019-03-29
	RX-COTOVELO ESQUERDO	2019-03-29
	RX-COTOVELO DIREITO	2019-03-29

lasmy n C Furtado <https://www.github.com/iascfurtado>

---

- Quantos exames foram realizados, no total da base?

```
SELECT COUNT(speciality_x) AS qtd_exames_feitos,
COUNT(DISTINCT speciality_x) AS qtd_tipo_exames
FROM ocupacao;
```

	qtd_exames_feitos	qtd_tipo_exames
▶	304493	57

- 
- Quantos tipos de exames diferentes existem na base?

```
SELECT COUNT(DISTINCT speciality_x) AS qtd_tipo_exames
FROM ocupacao;
```

	qtd_tipo_exames
▶	57

- 
- Qual a quantidade de médicos que existem nessa base?

```
SELECT COUNT(DISTINCT physician) AS qtd_de_medicos
FROM ocupacao;
```

Result Grid	
	qtd_de_medicos
▶	20

- 
- Quantos clientes diferentes (clínicas/hospitais) temos na base?

```
SELECT COUNT(DISTINCT client_name) AS clientes_distintos
FROM ocupacao;
```

Result Grid	
	clientes_distintos
▶	286

- 
- Para qual estado a empresa realizou o maior número de laudos de exames?

```
SELECT state, COUNT(state) AS pedidos_por_UF
FROM ocupacao
GROUP BY state
ORDER BY 2 DESC
LIMIT 5;
```

Result Grid		
	state	pedidos_por_UF
▶	SP	59783
	MG	54765
	BA	30944
	MA	29699
	PA	26774

- 
- Para qual estado a empresa realizou o menor número de laudos de exames?

```
SELECT state, COUNT(state) AS pedidos_por_UF
FROM ocupacao
GROUP BY state
ORDER BY 2
LIMIT 5;
```

Result Grid			Filter R
	state	pedidos_por_UF	
▶	AL	1	
	AC	29	
	ES	384	
	RR	470	
	DF	584	

---

● Distribuição percentual de exames em cada estado;

```
SELECT state, ROUND((COUNT(state)*100/tb1.totalexame),2) AS percent
FROM ocupacao, (SELECT COUNT(speciality_x) AS totalexame FROM ocupacao) AS tb1
GROUP BY state
ORDER BY 2 DESC;
```

Result Grid			Filter R
	state	percent	
▶	SP	19.63	
	MG	17.99	
	BA	10.16	
	MA	9.75	
	PA	8.79	
	RJ	7.58	
	RS	6.61	
	SC	3.90	
	AM	2.38	
	MS	1.91	
	MT	1.85	
	CE	1.67	
	PR	1.49	
	SE	1.35	
	GO	1.30	
	RO	0.72	
	PE	0.63	
	TO	0.62	
	RN	0.59	
	PI	0.30	
	AP	0.30	
	DF	0.19	
	RR	0.15	
	ES	0.13	
	AC	0.01	
	AL	0.00	

---

● Quais os 5 clientes que mais solicitaram mais laudos de exames?

```
SELECT client_name, COUNT(client_name) AS pedidos_por_cliente
FROM ocupacao
GROUP BY client_name
```

ORDER BY 2 DESC  
LIMIT 5;

	client_name	pedidos_por_cliente
▶	Electronics Consulting Application	29280
	North Studio South	24895
	Direct Power Solutions	17300
	Direct Systems Solutions	9828
	Network Vision	8295

- 
- Qual o exame mais caro e o mais barato?

```
SELECT speciality_x, value
FROM ocupacao
GROUP BY speciality_x
ORDER BY value ASC
LIMIT 5; -- mais barato
```

	speciality_x	value
	RX-SEIOS DA FACE	4
	RX-MÃO	4
	RX-TORAX (PA) COM LEITURA OIT	4
	RX-COCCIX	4
▶	RX-PÉ	10

```
SELECT speciality_x, value
FROM ocupacao
GROUP BY speciality_x
ORDER BY value DESC
LIMIT 5; -- mais caro
```

	speciality_x	value
▶	ESPIROMETRIA	34
	ACUIDADE VISUAL	34
	RX-FÍBULA TÍBIA ESQUERDA	31
	RX-TORNOZELO DIREITO	31
	RX-ANTEBRAÇO DIREITO	31

- 
- Qual o top 5 médicos que realizaram mais exames?

```
SELECT name, COUNT(speciality_x) AS exames
FROM ocupacao
GROUP BY physician, name
```

ORDER BY exames DESC  
LIMIT 5;

Result Grid			Filter Rows
	name	exames	
▶	Tara Williams	65801	
	Raymond Collins	44853	
	Laverne Hartson	44091	
	Linda Hansen	42593	
	Cheree Hwang	32410	

- Qual o médico que realizou menos exames? Como foi o desempenho dele no mês?

→ Não consegui chegar

→ consultas salvas no arquivo

o TOP 1 cliente que mais solicitou exames quer saber:

- Quantidade de exames que solicitou, por mês;

```
SELECT client, COUNT(speciality_x) AS exames
FROM ocupacao
GROUP BY client
ORDER BY exames DESC
LIMIT 1; -- saber quem é o cliente top
```

Result Grid			Filter Rows
	client	exames	
▶	119	29280	

```
SELECT client_name, COUNT(speciality_x), MONTH(CAST(create_date AS DATE)) AS
mes
FROM ocupacao
WHERE client = 119
GROUP BY client, mes
ORDER BY mes; -- saber quantos exames por mes solicitou
```

Result Grid				Filter Rows	Export
	client_name	COUNT(speciality_x)	mes		
▶	Electronics Consulting Application	10008	1		
	Electronics Consulting Application	10536	2		
	Electronics Consulting Application	8736	3		

- Valor total dos exames que solicitou, por mês;

```
SELECT client_name, SUM(value) AS valor_ttl_exames, MONTH(CAST(create_date AS
DATE)) AS mes
FROM ocupacao
```

WHERE client = 119  
 GROUP BY client, mes  
 ORDER BY mes;

client_name	valor_ttl_exam	mes
Electronics Consulting Application	145533	1
Electronics Consulting Application	153211	2
Electronics Consulting Application	127036	3

● 5 principais especialidades que solicitou no último mês;

SELECT client\_name, speciality\_x AS tipos\_exam, COUNT(speciality\_x) AS qtd\_exam,  
 MONTH(CAST(create\_date AS DATE)) AS mes  
 FROM ocupacao  
 WHERE client = 119 AND MONTH(CAST(create\_date AS DATE)) = 3  
 GROUP BY speciality\_x, mes  
 ORDER BY qtd\_exam DESC  
 LIMIT 5;

client_name	tipos_exam	qtd_exam	mes
Electronics Consulting Application	ELETRCARDIOGRAMA	2880	3
Electronics Consulting Application	ELETRONCEFALOGRAMA	1920	3
Electronics Consulting Application	ESPIROMETRIA	1560	3
Electronics Consulting Application	RX-TORAX (PA) COM LEITURA OIT	1392	3
Electronics Consulting Application	RX-COLUNA LOMBO-SACRA	600	3

● Média de valores das 5 principais especialidades que solicitou no último mês disponível na base;

-- primeiro pegar as cinco principais especialidades solicitadas no mes, questao  
 acima: eletrocardiograma, eletroencefalograma, espirometria, rx torax pa, rx coluna

SELECT speciality\_x, value  
 FROM ocupacao  
 WHERE speciality\_x IN('ELETRCARDIOGRAMA', 'ELETRONCEFALOGRAMA',  
 'ESPIROMETRIA', 'RX-TORAX (PA) COM LEITURA OIT', 'RX-COLUNA LOMBO-  
 SACRA')  
 GROUP BY speciality\_x;

-- pegar o valor dessas cinco 26,29,26,4,34 a média deveria ser a soma desses valores  
 dividido pelos valores, que daria  $119/5 = 23,8$

-- pegar a media desses valores considerando apenas os exames acima  
 SELECT client\_name, ROUND(AVG( DISTINCT value),2) AS media\_valores  
 FROM ocupacao  
 WHERE client = 119 AND speciality\_x IN('ELETRCARDIOGRAMA',  
 'ELETRONCEFALOGRAMA', 'ESPIROMETRIA', 'RX-TORAX (PA) COM LEITURA OIT',

'RX-COLUNA LOMBO-SACRA')

GROUP BY client\_name; -- resultado nao bate com calculo acima -- *resultado nao bate com calculo acima*

Result Grid	Filter Rows:
client_name	media_valores
▶ Electronics Consulting Application	15.94

- 
- 5 principais cidades que solicitaram exames para esse cliente, no último mês;

```
SELECT client_name, state, speciality_x
FROM ocupacao
WHERE client = 119 AND YEAR(CAST(create_date AS DATE)) = 2019 AND
MONTH(CAST(create_date AS DATE)) = 3
GROUP BY state, speciality_x
ORDER BY state DESC
LIMIT 5; -- city só retorna zero
```

client_name	state	speciality_x
▶ Electronics Consulting Application	MG	ELETRCARDIOGRAMA
Electronics Consulting Application	MG	ELETRORNEFALOGRAMA
Electronics Consulting Application	MG	ESPIROMETRIA
Electronics Consulting Application	MG	RX-COLUNA CERVICAL
Electronics Consulting Application	MG	RX-COLUNA DORSAL

- 
- Qual a especialidade de exames que foi menos demandada? Essa característica se repete nos meses anteriores?

```
SELECT client_name, speciality_x AS tipos_exames, COUNT(speciality_x) AS qtd_exames,
MONTH(CAST(create_date AS DATE)) AS mes
FROM ocupacao
WHERE client = 119
GROUP BY tipos_exames
ORDER BY qtd_exames, mes;
-- RX BACIA
-- NÃO
```



Result Grid    Filter Rows: <input type="text"/>   Export:    Wrap Cell Content:				
	client_name	tipos_exam	qtd_exam	mes
▶	Electronics Consulting Application	RX-BACIA	24	2
	Electronics Consulting Application	RX-TÓRAX	24	3
	Electronics Consulting Application	RX-OMBRO ESQUERDO	24	3
	Electronics Consulting Application	RX-COLUNA DORSAL	72	3
	Electronics Consulting Application	RX-COLUNA TORÁCICA	192	1
	Electronics Consulting Application	RX-COLUNA CERVICAL	264	1
	Electronics Consulting Application	RX-COLUNA LOMBO-SACRA	1440	1
	Electronics Consulting Application	ESPIROMETRIA	4296	1
	Electronics Consulting Application	RX-TORAX (PA) COM LEITURA OIT	4944	1
	Electronics Consulting Application	ELETOENCEFALOGRAMA	7272	1
	Electronics Consulting Application	ELETOCARDIOGRAMA	10728	1

- **EXTRA:** É possível correlacionar exames menos demandados com valor mais alto? Não, pois o exame mais barato é o RX-COCCIX e ele custa 4, embora seja o menos demandado também é o mais barato.

```
SELECT speciality_x, COUNT(speciality_x) AS qtd_exam, value
FROM ocupacao
GROUP BY speciality_x
ORDER BY qtd_exam ASC, value DESC
LIMIT 10;
```

lasmy n C Furtado <https://www.github.com/iascfurtado>

Result Grid    Filter Rows: <input type="text"/>			
	speciality_x	qtd_exam	value
▶	RX-COCCIX	26	4
	RX-BRACO DIREITO	31	23
	RX-PUNHO	40	27
	RX-ARCOS COSTAIS	44	12
	RX-ANTEBRAÇO	49	10
	RX-PÉ	49	10
	RX-CALCANEIO DIREITO	57	29
	RX-CALCANEIO ESQUERDO	57	29
	RX-CLAVICULA ESQUERDA	63	31
	RX-FÍBULA TÍBIA ESQUERDA	63	31

Entregar para o cliente TOP 1 em valores gastos o relatório abaixo, utilizando os dados do mês mais recente disponível na base:

- especialidades
- qtd laudos de exames realizados
- total valor que será pago à empresa de telemedicina
- média de valor de cada exame
- qtd de estados diferentes onde realizou laudos de exames

**Exemplo de tabela de saída:**

Especialidade	Qtd realizados	Total valor	Média valor	Qtd estados
ELETRONECEFA LOGRAMA	10	2340	234	2
RX-TÓRAX	5	200	40	1
ELETRCARDI OGRAMA	2	700	350	1

```
SELECT
DISTINCT(speciality_x) AS 'Especialidades',
COUNT(speciality_x) AS 'Qtd Realizados',
SUM(value) AS 'Total Valor',
ROUND(AVG(DISTINCT value),2) 'Média Valor',
COUNT(DISTINCT state) AS 'Qtd Estados'
FROM ocupacao
WHERE client = 119 AND YEAR(CAST(create_date AS DATE)) = 2019 AND MONTH(CAST(create_date
AS DATE)) = 3
GROUP BY speciality_x
ORDER BY speciality_x;
```

Result Grid					
Filter Rows:		Export:		Wrap Cell Content:	
	Especialidades	Qtd Realizados	Total Valor	Média Valor	Qtd Estados
▶	ELETRCARDIOGRAMA	2880	41880	15.94	1
	ELETRONECEFALOGRAMA	1920	27920	15.94	1
	ESPIROMETRIA	1560	22685	15.94	1
	RX-COLUNA CERVICAL	168	2443	15.94	1
	RX-COLUNA DORSAL	72	1047	15.94	1
	RX-COLUNA LOMBO-SACRA	600	8725	15.94	1
	RX-COLUNA TORÁCICA	96	1396	15.94	1
	RX-OMBRO ESQUERDO	24	349	15.94	1
	RX-TÓRAX	24	349	15.94	1
	RX-TORAX (PA) COM LEITURA OIT	1392	20242	15.94	1