

Olá!

Este documento foi criado com o intuito de documentar os passos feitos e as conclusões chegadas de acordo com a análise dos dados do *DataBase*, fornecido através de programa de mentoria.

As informações e conclusões as quais o presente documento chegaram não possuem valor mercadológico ou qualquer intuito comercial, mas tão somente acadêmico e para fins de demonstração de *skills* de Análise de Dados para Negócios e de uso do *SQL* com fito de gerar portfólio.

As informações apresentadas serão feitas na abordagem *top-down*, ou seja, primeiro serão mostradas as conclusões inferidas e após mostrar-se-á todo o projeto para chegar a tal conclusão, desde linhas de código até *prints* dos resultados.

O processo de ETL (Extração, Transformação e Limpeza) dos Dados e de criação de tabelas no *database* não será parte deste documento.

As ferramentas usadas para o presente projeto foram:

Oracle MySQL 8.0

Libre Office Calc

Libre Office Writer

Foxit PDF

Microsoft Windows 10

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:

- Quantas linhas e colunas existem na base?

Linhas: 304.493, Colunas: 10

- É possível classificar as colunas em entidades/grupos?

Sim, é possível agrupá-los por quase todas as colunas, exceto pela de cidade, onde todos os valores, embora não nulos, são zero, sendo mais interessante agrupar os resultados da base pelo Estado.

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

Exame mais recente: Eletroencefalograma de 03/2019

Exame mais antigo: Eletroencefalograma de 01/2018

- Quantos exames foram realizados, no total da base?

304.493, sendo 57 tipos diferentes de exames

- Quantos tipos de exames diferentes existem na base?

57 tipos diferentes de exames

- Qual a quantidade de médicos que existem nessa base?

20 médicos diferentes

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

286 clientes diferentes

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

São Paulo, seguido por Minas Gerais, Bahia, Maranhão e Pará

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

Alagoas, seguido por Acre, Espírito Santo, Roraima e Distrito Federal

- Distribuição percentual de exames em cada estado;

SP: 19,63% [...] (vide resultado da consulta, já que foram 26 resultados)

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

Electronics Consulting Application, North Studio South, Direct Power Solutions, Direct Systems Solutions, Network Vision

- Qual o exame mais caro e o mais barato?

Mais barato: Rx-Seios da Face, valor 4

Mais caro: Espirometria, valor 34

- Qual o top 5 médicos que realizaram mais exames?

Tara Williams, Raymond Collins, Laverne Hartson, Linda Hansen, Cheree Hwang

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

Eugene King, o desempenho dele nos meses de Janeiro e Fevereiro foram os mais baixos dos meses

O TOP 1 cliente que mais solicitou exames quer saber:

Top cliente = código 119

- Quantidade de exames que solicitou, por mês;

29.280 exames realizados

- Valor total dos exames que solicitou, por mês;

Mês 1: 145.533,00; Mês 2: 153.211,00; Mês 3: 127.036,00

- 5 principais especialidades que solicitou no último mês;

ELETROCARDIOGRAMA, ELETROENCEFALOGRAMA, ESPIROMETRIA, RX-TORAX (PA) COM LEITURA OIT, RX-COLUNA LOMBO-SACRA

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

23,90 dinheiros

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

Não é possível apurar cidades, pois as informações de cidade estão todas com o numeral zero, contudo é possível identificar o estado, que se tratou nesse caso de Minas Gerais.

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

RX-BACIA; Não.

- **EXTRA:** É possível correlacionar exames menos demandados com valor mais alto?

Não, pois o exame mais barato de toda a base e que tem a menor quantidade de pedidos é o mesmo.

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

ABAIXO

Exemplo de tabela de saída:

Especialidade	Qtd realizados	Total valor	Média valor	Qtd estados
ELETROENCEFALOGRAMA	10	2340	234	2
RX-TÓRAX	5	200	40	1
ELETROCARDIOGRAMA	2	700	350	1

CONSULTAS E RESULTADOS

-- 1) Quantas linhas e colunas existem na base?

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

Result Grid	
	qtd_linhas_tb
▶	304493

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

Result Grid	
	qtd_colunas
▶	10

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

```
SELECT speciality_x AS exames,  
COUNT(speciality_x) AS qtd_exames,  
create_date,  
report_date,  
physician,  
client_name,  
state,  
value  
FROM ocupacao  
GROUP BY speciality_x  
ORDER BY speciality_x  
LIMIT 8;
```

Result Grid								
	exames	qtd_exames	create_date	report_date	physician	client_name	state	value
▶	ACUIDADE VISUAL	6785	2019-01-03 14:18:22	2019-01-03 17:24:26	7	General Telecom	BA	34
	ELETRCARDIOGRAMA	111456	2019-01-02 09:50:58	2019-01-02 13:11:08	87	Direct Pacific Analysis	MG	26
	ELETRONCEFALOGRAMA	74728	2019-01-02 08:56:15	2019-01-02 14:39:25	6	Direct Pacific Analysis	MG	26
	ELETRONCEFALOGRAMA COM MAPEAMENTO	2709	2019-01-02 12:07:51	2019-01-03 09:06:08	84	Pacific Internet Hardware	SP	25
	ESPIROMETRIA	50375	2019-01-02 09:01:24	2019-01-02 10:47:46	85	Telecom Graphics Systems	MS	34
	HOLTER	324	2019-01-07 08:13:14	2019-01-07 08:14:10	87	Data Universal Star	AM	18
	MAPA	283	2019-01-04 10:34:43	2019-01-07 08:53:35	87	Internet Digital Resource	SP	26
	RX-ANTEBRAÇO	49	2019-02-18 09:52:45	2019-02-18 10:51:50	8	Contract Bell Net	GO	10

-- 3) Qual o exame mais recente e o mais antigo da base?

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

Result Grid		
	speciality_x	mais_antigo
▶	ELETRONCEFALOGRAMA	2018-01-29
	ELETRCARDIOGRAMA	2018-11-29
	ELETRONCEFALOGRAMA	2018-11-30

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

Result Grid		
	speciality_x	mais_recente
▶	ELETRONCEFALOGRAMA	2019-03-29
	RX-COTOVELO ESQUERDO	2019-03-29
	RX-COTOVELO DIREITO	2019-03-29

-- 4) 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; -- exames total
```

Result Grid			Filter Rows:
	qtd_exames_feitos	qtd_tipo_exames	
▶	304493	57	

-- 5) Quantos tipos de exames diferentes existem na base

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

Result Grid		Filter Rows:
	qtd_tipo_exames	
▶	57	

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

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

Result Grid		Filter Rows:
	qtd_de_medicos	
▶	20	

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

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

Result Grid		Filter Rows:
	clientes_distintos	
▶	286	

-- 8) 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			Filter Rows:
	state	pedidos_por_UF	
▶	SP	59783	
	MG	54765	
	BA	30944	
	MA	29699	
	PA	26774	

-- 9) 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 Rows:
	state	pedidos_por_UF	
▶	AL	1	
	AC	29	
	ES	384	
	RR	470	
	DF	584	

-- 10) 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 Rows:
	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	

-- 11)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

-- 12)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

-- 13)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;
```

name	exames
Tara Williams	65801
Raymond Collins	44853
Laverne Hartson	44091
Linda Hansen	42593
Cheree Hwang	32410

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

```
SELECT name, COUNT(speciality_x) AS exames, MONTH(CAST(create_date AS DATE)) AS mes
FROM ocupacao
GROUP BY mes, physician
ORDER BY exames, mes
LIMIT 10;
```

name	exames	mes
Eugene King	4	1
Eugene King	5	2
Tara Williams	5	11
Linda Hansen	5	11
Jared Sells	7	12
Linda Hansen	8	12
Eugene King	13	3
Jared Sells	28	3
David Lillian	73	12
Barbara Shearin	108	1

-- O TOP CLIENTE QUE MAIS SOLICITOU EXAMES QUER SABER:

-- 1) 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

Result Grid		
	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
```

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

-- 2) 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;
```

Result Grid			
	client_name	valor_ttl_exames	mes
▶	Electronics Consulting Application	145533	1
	Electronics Consulting Application	153211	2
	Electronics Consulting Application	127036	3

-- 3) 5 principais especialidades que solicitou no último mês;

```
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 AND MONTH(CAST(create_date AS DATE)) = 3
```

```
GROUP BY speciality_x, mes
```

```
ORDER BY qtd_exames DESC
```

```
LIMIT 5;
```

Result Grid				
	client_name	tipos_exames	qtd_exames	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


```
-- 4) 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('ELETROCARDIOGRAMA', 'ELETROENCEFALOGRAMA', '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(119/5, 2) AS media_valores
FROM ocupacao
WHERE client = 119 AND speciality_x IN('ELETROCARDIOGRAMA', 'ELETROENCEFALOGRAMA', 'ESPIROMETRIA', 'RX-TORAX (PA) COM LEITURA OIT', 'RX-COLUNA LOMBO-SACRA')
GROUP BY client_name;
```

Result Grid			Filter Rows:
	speciality_x	value	
▶	ELETROENCEFALOGRAMA	26	
	RX-COLUNA LOMBO-SACRA	29	
	ELETROCARDIOGRAMA	26	
	RX-TORAX (PA) COM LEITURA OIT	4	
	ESPIROMETRIA	34	

Result Grid			Filter Rows:
	client_name	media_valores	
▶	Electronics Consulting Application	23.80	

```
-- 5) 5 principais cidades que solicitaram exames para esse cliente, no último mês;
SELECT client_name, state,
COUNT(DISTINCT state) AS qtd_estados,
COUNT(speciality_x) AS qtd_exames,
COUNT(DISTINCT MONTH(CAST(create_date AS DATE))) AS qtd_meses
FROM ocupacao
WHERE client = 119
AND YEAR(CAST(create_date AS DATE)) = 2019
AND MONTH(CAST(create_date AS DATE)) = 3
GROUP BY state; -- city só retorna zero
```

Result Grid			Filter Rows:	Export:	Wrap Cell Content:
	client_name	state	qtd_estados	qtd_exames	qtd_meses
▶	Electronics Consulting Application	MG	1	8736	1

```
-- 6) 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;
```

Result Grid			Filter Rows:	Export:	Wrap Cell Content:
	client_name	tipos_exames	qtd_exames	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	ELETROENCEFALOGRAMA	7272	1	
	Electronics Consulting Application	ELETROCARDIOGRAMA	10728	1	

-- EXTRA: É possível correlacionar exames menos demandados com valor mais alto?

```
SELECT speciality_x, COUNT(speciality_x) AS qtd_exames, value
```

```
FROM ocupacao
```

```
GROUP BY speciality_x
```

```
ORDER BY qtd_exames ASC, value DESC
```

```
LIMIT 10;
```

Result Grid			
	speciality_x	qtd_exames	value
▶	RX-COCCIX	26	4
	RX-BRAÇO 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

```
SELECT DISTINCT(speciality_x) AS 'Especialidades',
```

```
COUNT(speciality_x) AS 'Qtd Realizados',
```

```
SUM(value) AS 'Total Valor',
```

```
ROUND(SUM(DISTINCT value)/COUNT(speciality_x),2) AS 'Média de Valores',
```

```
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 COUNT(speciality_x) DESC;
```

Result Grid					
	Especialidades	Qtd Realizados	Total Valor	Média de Valores	Qtd Estados
▶	ELETROCARDIOGRAMA	2880	41880	0.09	1
	ELETROENCEFALOGRAMA	1920	27920	0.13	1
	ESPIROMETRIA	1560	22685	0.16	1
	RX-TORAX (PA) COM LEITURA OIT	1392	20242	0.18	1
	RX-COLUNA LOMBO-SACRA	600	8725	0.42	1
	RX-COLUNA CERVICAL	168	2443	1.52	1
	RX-COLUNA TORÁCICA	96	1396	2.66	1
	RX-COLUNA DORSAL	72	1047	3.54	1
	RX-OMBRO ESQUERDO	24	349	10.62	1
	RX-TÓRAX	24	349	10.62	1