

Questão 2

1 - Indicadores Estatísticos

Para o cálculo dos principais indicadores estatísticos sobre cada atributo de cada tabela presente na base de dados, foi utilizada uma função fornecida no material de aula pelo docente. Essa função verifica o tipo de dado de cada atributo e calcula os indicadores mais adequados para cada um deles. A função foi modificada para que retornasse indicadores mais relevantes. Para atributos de ID, por exemplo, não foi retornada a variância e o desvio padrão, que seria o padrão da função, visto que esses indicadores não fornecem informações importantes no contexto do problema.

Os valores retornados pela função são: o tipo de variável do atributo; o número de *nulls*; a cardinalidade (valores distintos); a variância e o desvio padrão (atributos numéricos).

Assim, executamos a função para cada uma das três tabelas do banco de dados - pacientes, exames, desfechos. As saídas estão mostradas abaixo:

	nomeatrib	tipo	nulls	cardinalidade	variância	desviopadrao
1	id_paciente	bpchar	0	14673	<null>	<null>
2	ic_sexo	bpchar	0	2	<null>	<null>
3	aa_nascimento	int4	100	91	307.3418532976582	17.531168052861116
4	cd_pais	bpchar	0	2	<null>	<null>
5	cd_uf	bpchar	1195	5	<null>	<null>
6	cd_municipio	text	3714	9	<null>	<null>
7	cd_cepreduzido	varchar	14642	3	<null>	<null>

Figura 1 – Resultado da função estatística na tabela Pacientes.

	nomeatrib	tipo	nulls	cardinalidade	variância	desviopadrao
1	id_exame	int8	0	2952999	<null>	<null>
2	id_paciente	bpchar	0	14673	<null>	<null>
3	id_atendimento	bpchar	0	37364	<null>	<null>
4	dt_coleta	date	0	473	<null>	<null>
5	de_origem	varchar	0	113	<null>	<null>
6	de_exame	varchar	0	776	<null>	<null>
7	de_analito	varchar	0	1081	<null>	<null>
8	de_resultado	text	0	20440	<null>	<null>
9	cd_unidade	varchar	0	54	<null>	<null>
10	de_valor_referencia	varchar	0	814	<null>	<null>

Figura 2 – Resultado da função estatística na tabela Exames.

	nomeatrib	tipo	nulls	cardinalidade	variância	desviopadrazo
1	id_paciente	bpchar	0	14672	<null>	<null>
2	id_atendimento	bpchar	0	89937	<null>	<null>
3	dt_atendimento	date	0	475	<null>	<null>
4	de_tipo_atendimento	text	0	4	<null>	<null>
5	id_clinica	int4	0	51	<null>	<null>
6	de_clinica	text	0	51	<null>	<null>
7	dt_desfecho	varchar	0	476	<null>	<null>
8	de_desfecho	text	0	14	<null>	<null>

Figura 3 – Resultado da função estatística na tabela Desfechos.

Além dessas métricas principais, foram realizadas outras consultas mais específicas para cada tabela. Inicialmente a respeito da tabela de Pacientes, a primeira consulta mostra a quantidade de estrangeiros presentes na base de dados. O resultado está mostrado abaixo:

	n_estrangeiros
1	97

Figura 4 – Quantidade de estrangeiros cadastrados na tabela Pacientes.

Outra consulta realizada para a tabela de Pacientes mostra a quantidade de pessoas de cada estado (UF):

	cd_uf	count
1	SP	12151
2	GO	12
3	MT	5
4	PA	12
5	DF	1298

Figura 5 – Quantidade de pacientes em cada estado (UF).

Na mesma linha, temos duas últimas consultas para a tabela Pacientes: a quantidade de pessoas por município registrado na tabela e a quantidade de pessoas por CEP registrado na tabela. Os resultados são:

	cd_municipio	count
1	DIADEMA	5
2	SAO PAULO	9372
3	BELEM	5
4	SAO BERNARDO DO CAMPO	17
5	OSASCO	54
6	MOGI DAS CRUZES	5
7	GUARULHOS	197
8	BRASILIA	1283
9	CAMPOS DO JORDAO	21

Figura 6 – Quantidade de pacientes em cada município registrado.

	cd_cepreduzido	count
1	01307	5
2	01415	5
3	12460	21

Figura 7 – Quantidade de pacientes em cada CEP registrado.

Para a tabela de Exames, obtivemos a média de exames por atendimento:

	avg
1	1756.6187055938725343

Figura 8 – Média de exames por atendimento.

e a média de coletas por dia:

	media_coletas_dia
1	6907.6122684091664101

Figura 9 – Média de coletas por dia.

Para a tabela de Exames, ainda foi consultada a distribuição das origens dos exames:

	de_origem	count
1	Administração de Contratos	42
2	Administração - Itaim	39
3	Anatomia Patológica - Itaim	98
4	Anatomia Patológica - Jardins	383
5	Apoio Administrativo - Brasília II	2
6	Apoio Administrativo - Brasília III	1199
7	Apoio Administrativo e Financeiro - Brasília IV	876
8	Apoio Administrativo - Jardins	11225
9	Atendimento - Centro de Oncologia	58
10	Atendimento - Recepção / Internação	49

Figura 10 – Distribuição das origens dos exames.

Finalmente, para a tabela de Desfechos, três indicadores foram analisados: média de atendimentos por dia; a distribuição dos tipos de atendimentos; e a média de desfechos por dia. Os resultados estão mostrados abaixo.

	media_atendimentos
1	235.0950665465826078

Figura 11 – Média de atendimentos por dia.

	de_tipo_atendimento	count
1	Internado	4764
2	Pronto Atendimento	19342
3	Ambulatorial	39364
4	Externo	26467

Figura 12 – Distribuição dos tipos de atendimentos.

	media_desfechos
1	225.7042485295262239

Figura 13 – Média de desfechos por dia.

2. Tabela Pacientes

Para calcular a quantidade de pacientes presentes na base de dados, basta utilizar a tabela *pacientes*, e realizar um *count* da coluna *id_pacientes*. Como esta coluna é uma chave primária, não ocorrem repetições de ID's, logo o *count* retorna corretamente o número de pacientes na base.

	n_pacientes
1	14673

Figura 14 – Número de pacientes na base de dados.

A quantidade de homens e mulheres pode ser obtida ao realizar um *count* na coluna *ic_sexo* da tabela *pacientes*, em conjunto com um *group by ic_sexo*. Assim, pacientes do sexo feminino e masculino serão contabilizados separadamente, e a *query* retornará o número de pacientes homens e mulheres.

	sexo	n_pacientes
1	M	7381
2	F	7292

Figura 15 – Número de pacientes homens e mulheres.

Para obter a faixa etária por sexo, primeiro obteve-se os anos mínimo e máximo de nascimento, que, em seguida, foram subtraídos do ano atual, sendo agrupados pela coluna *ic_sexo*. Assim, os anos mínimo e máximo de cada um dos sexos serão selecionados separadamente, de forma que a *query* retorne as faixas etárias para cada um dos sexos dos pacientes atendidos.

	ic_sexo	idade_min	idade_max
1	M	1	91
2	F	1	91

Figura 16 – Faixa etária por sexo.

Quanto à distribuição dos quartis dentro de cada faixa etária, foi basicamente realizar um *percentile_count()* na coluna *aa_nascimento*, operação esta que retornará o ano de nascimento correspondente ao quartil pedido (por exemplo, retornará, para o primeiro quartil, o ano que é 25% maior que todos os demais anos (ordenados de forma ascendente)). Tendo essa informação, basta subtrair o ano obtido do ano atual, e assim a idade correspondente àquele quartil será obtida. Em seguida, só resta agrupar estas operações por sexo, e assim o primeiro, segundo e terceiro quartis de cada uma das faixas etárias calculadas anteriormente serão obtidos.

	sexo	"1o_quart"	"2o_quart"	"3o_quart"
1	F	53	41	32
2	M	56	43	33

Figura 17 – Quartis de cada faixa etária (homens e mulheres).

Por fim, para a distribuição de pacientes homens e mulheres por década de vida, primeiro calculou-se as décadas através de $(aa_nascimento/10)*10$, o que retorna como *decada* as décadas correspondentes aos anos de nascimento dos pacientes. Com isso, basta realizar um *count* de *id_pacientes*, agrupando a contagem por década e sexo. Assim, para cada década e gênero, haverá o número correspondente de pacientes que atendem à categoria.

	ic_sexo	decada	n_pacientes
1	F	1930	219
2	F	1940	279
3	F	1950	518
4	F	1960	928
5	F	1970	1521
6	F	1980	1984
7	F	1990	1204
8	F	2000	428
9	F	2010	164
10	F	2020	6

Figura 18 – Distribuição de pacientes (mulheres) por década de vida.

3. Tabela Exames

Para obter a maior quantidade de exames solicitados para um único paciente, primeiro conta-se a quantidade de exames por paciente, através de *count(id_exame)* agrupado por *id_paciente*. Depois de ter feito isto, bastava selecionar o maior valor obtido pelo *count*.

	n_max_exames
1	18428

Figura 19 – Maior quantidade de exames solicitados para um único paciente.

Para calcular a média de exames pedidos para homens e mulheres foi necessário realizar inicialmente uma junção entre as tabelas *pacientes* e *exames*. Em seguida, foi feita uma

contagem do número de exames, sendo este valor agrupado por *id_paciente*. Foram feitas duas *query*'s, uma correspondente ao sexo masculino e outra ao feminino, pois a contagem era realizada separadamente para cada um dos gêneros. Ao fim da contagem, calcula-se a média do número de exames calculado por paciente.

	media_exames_m
1	254.4756808020593416

Figura 20 – Média de exames pedidos para homens.

	media_exames_f
1	147.3826110806363138

Figura 21 – Média de exames pedidos para mulheres.

Com os tratamentos do Exercício 1 aplicados na base, é simples calcular quantos exames de Covid-19 foram solicitados, pois basta realizar uma contagem na coluna *de_analito* onde os valores correspondem a "DETECCAO COVID", "COVID 19 - ANTICORPOS IGM" ou "COVID 19 - ANTICORPOS IGG", que são os exames deste hospital em específico que são usados para detectar a doença.

	qnt_exames_covid
1	11548

Figura 22 – Número de exames de Covid-19 solicitados.

Já para contabilizar quantos dos exames acima deram resultado positivo, basta adicionar à *query* uma comparação relacionada à coluna *de_resultado*, de forma que a contagem só seja realizada nos exames cujo valor desta coluna seja "DETECTADO". Assim, somente exames de detecção de Covid-19 com resultado positivo são contabilizados.

	qnt_exames_covid_positivo
1	6597

Figura 23 – Número de exames de Covid-19 com resultado positivo.

Por fim, para mostrar para cada idade os resultados do exame de Covid-19: primeiramente, foi necessário realizar uma junção das tabelas *exames* e *pacientes*, para ser possível acessar tanto o ano de nascimento quanto o ID dos exames. Para o cálculo da idade, bastou

subtrair do ano atual o ano de nascimento do paciente. Quanto ao resultado do exame, fez-se uma contagem, agrupando-a pelo resultado do exame e pelo ano de nascimento. Dessa forma, tem-se para cada ano de nascimento duas colunas, correspondentes número de pacientes com determinado resultado (positivo, negativo ou inconclusivo) no exame.

	idade	de_resultado	count
1	91	DETECTADO	59
2	91	INCONCLUSIVO	3
3	91	NAO DETECTADO	58
4	90	DETECTADO	7
5	90	NAO DETECTADO	10
6	89	DETECTADO	11
7	89	NAO DETECTADO	5
8	88	DETECTADO	11
9	88	NAO DETECTADO	15
10	87	DETECTADO	16

Figura 24 – Resultados dos exames de Covid-19 por idade.

4. Tabela Desfechos

Para descobrir o desfecho da maioria dos casos registrados, calculou-se a moda dos resultados da coluna *de_desfechos* da tabela de desfechos. Como a moda retorna o valor com a maior ocorrência, a *query* retornará corretamente qual a maioria dos desfechos dos casos analisados.

	desfecho_maioria
1	Alta Administrativa

Figura 25 – Desfecho da maioria dos casos.

Já para descobrir a maioria dos desfechos por gênero e por década, pode-se apenas complementar o que foi feito na *query* acima. Entretanto, é necessário realizar a junção das tabelas *pacientes* e *desfechos*, uma vez que aquela tabela é necessária para ter acesso a idade e gênero de cada paciente. Para encontrar a maioria dos desfechos para cada gênero, basta selecionar a moda de *de_desfecho* agrupada por gênero. Assim, será retornado separadamente a moda dos desfechos para cada gênero. Quanto à década, o procedimento é semelhante: calcula-se as décadas através de $(aa_nascimento/10)*10$, e calcula-se a moda dos desfechos, agrupando-a por década; assim, cada década terá, separadamente, qual foi a maioria dos desfechos dos pacientes que nasceram nela.

	de_desfecho	contagem
1	Alta Administrativa	66908
2	Alta médica melhorado	21702
3	Desistência do atendimento	401
4	Alta a pedido	261
5	Alta médica Inalterado	227
6	Óbito após 48hs de internação sem necrópsia	172
7	Alta médica curado	129
8	Alta por abandono	65
9	Transferência Inter-Hospitalar Externa - Serviço de Ambulância	27
10	Óbito nas primeiras 48hs de internação sem necrópsia não agônico	27
11	Assistência Domiciliar	11
12	Transferência Inter-Hospitalar Externa - Transporte Próprio	4
13	Óbito nas primeiras 48hs de internação sem necrópsia agônico	2
14	Óbito após 48hs de internação com necrópsia	1

Figura 26 – Número de desfechos de cada tipo.

	de_desfecho	contagem
1	Alta Administrativa	66908
2	Alta médica melhorado	21702
3	Desistência do atendimento	401
4	Alta a pedido	261
5	Alta médica Inalterado	227
6	Óbito após 48hs de internação sem necrópsia	172
7	Alta médica curado	129
8	Alta por abandono	65
9	Óbito nas primeiras 48hs de internação sem necrópsia não agônico	27
10	Transferência Inter-Hospitalar Externa - Serviço de Ambulância	27
11	Assistência Domiciliar	11
12	Transferência Inter-Hospitalar Externa - Transporte Próprio	4
13	Óbito nas primeiras 48hs de internação sem necrópsia agônico	2
14	Óbito após 48hs de internação com necrópsia	1

Figura 27 – Número de desfechos de cada tipo - Windows functions.

	ic_sexo	desfecho_maioria
1	F	Alta Administrativa
2	M	Alta Administrativa

Figura 28 – Maioria de desfechos por sexo.

	decada	desfecho_maioria
1	1930	Alta Administrativa
2	1940	Alta Administrativa
3	1950	Alta Administrativa
4	1960	Alta Administrativa
5	1970	Alta Administrativa
6	1980	Alta Administrativa
7	1990	Alta Administrativa
8	2000	Alta Administrativa
9	2010	Alta Administrativa
10	2020	Alta Administrativa
11	<null>	Alta Administrativa

Figura 29 – Maioria de desfechos por década de nascimento.

	decada	desfecho_maioria
1	0	Alta Administrativa
2	1	Alta Administrativa
3	2	Alta Administrativa
4	3	Alta Administrativa
5	4	Alta Administrativa
6	5	Alta médica melhorado
7	6	Alta Administrativa
8	7	Alta Administrativa
9	8	Alta Administrativa
10	9	Alta Administrativa

Figura 30 – Maioria de desfechos por década de vida.