Relatório de Trabalho

Este documento serve para a apresentação do trabalho prático realizado pelo   
Grupo G3H.

Projeto Base de dados

CC2005



Índice

[Elementos do grupo 2](#_Toc184652056)

[Universo considerado e modelo de classes UML 2](#_Toc184652057)

[Modelo relacional 3](#_Toc184652058)

[Povoamento de tabelas 3](#_Toc184652059)

[Interrogações SQL 4](#_Toc184652060)

[Aplicação Python 12](#_Toc184652061)

[Conclusão 12](#_Toc184652062)

[Referências 12](#_Toc184652063)

## Elementos do grupo

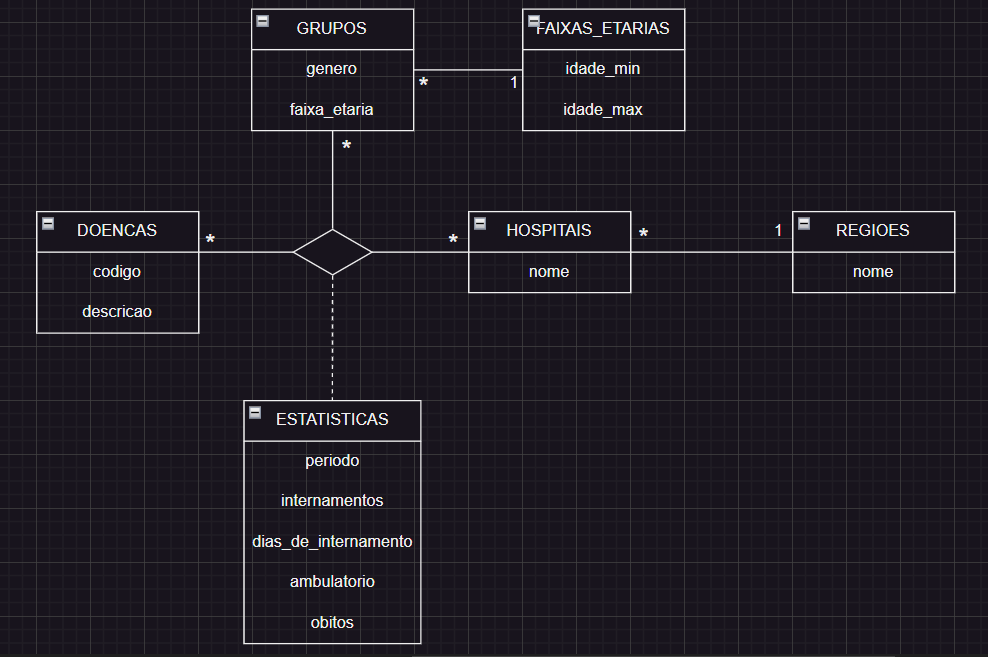
|  |  |
| --- | --- |
| **Nº mecanográfico** | **Nome** |
| 202304068 | Hugo Simas de Melo |
| 202303606 | Orlando Miguel Carvalho-Soares |
| 202208527 | Gonçalo Brito Machado Correia |
| 202303580 | |  | | --- | |  |   David Ventura Mendes de Sá |

## Universo considerado e modelo de classes UML

**Tema:** Morbilidade e mortalidade hospitalar

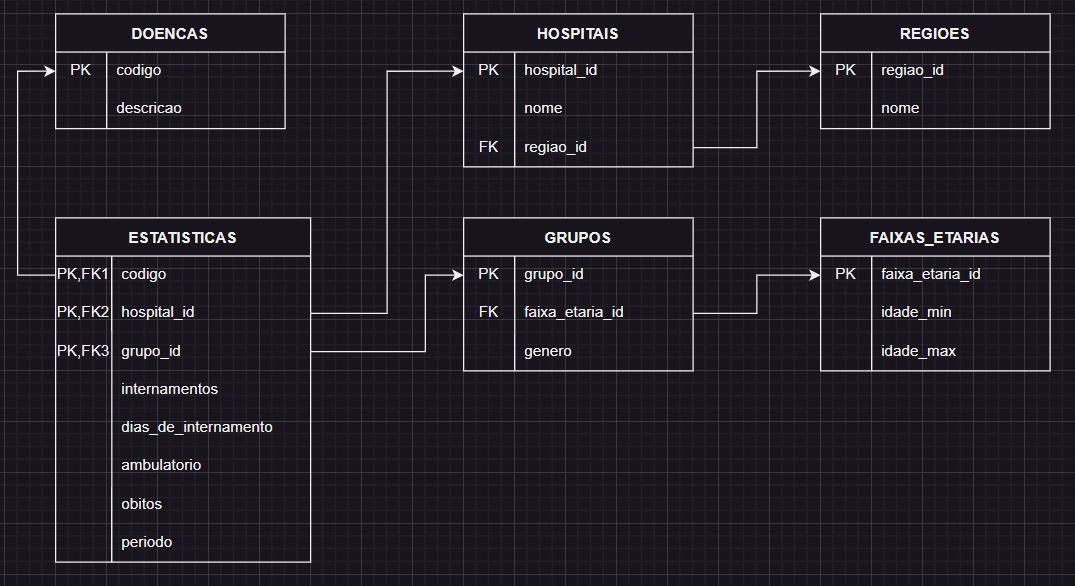
Registos de Morbilidade e mortalidade hospitalar, nos hospitais de todo o país de 2016 a 2020.

Inclui o período, o código, a descrição e capítulo diagnóstico, a instituição, região, faixa etária,

****género, internamentos, duração do internamento, ambúlatório e os óbidoscontendo. Contém 431.456 registos únicos.

Criámos 3 tabelas principais - DOENCAS, HOSPITAIS e GRUPOS – e uma tabela associação ESTATISTICAS que liga as mesmas, pois sem elas não existe a estatística. Da tabela GRUPOS, surge a tabela FAIXAS\_ETARIAS, na qual a decidimos dividir em idade minima e máxima, para melhorias na pesquisa de informação. Da tabela HOSPITAIS surge a tabela REGIOES.

## Modelo relacional

****

## Povoamento de tabelas

Criamos uma folha de excel para cada tabela (tabelas\_excel).

Depois com um script em python (povoar.py), criamos e povoamos cada tabela na base de dados conforme apresentada na folha excel, usando as libraries ‘pandas’ e ‘sqlite3’.

Por fim, já dentro do SQLite Studio, definimos as Primary/Foreign Keys (e com quais tabelas se relacionavam).

Ao povoar as tabelas apercebemo-nos de que os códigos préviamente dados para cada doença estavam indevidademte associados, havendo várias doenças com o mesmo código. Para isso, tivemos de alterar a base de dados incial, usando ferramentas como ‘find and replace’, ‘remove duplicates’ e ‘sort by’ do excel.

## Interrogações SQL

1. Qual é o número de óbitos por género? Mostra o género e o número de óbitos.

SELECT GRUPOS.genero, SUM(ESTATISTICAS.obitos) as obitos

FROM ESTATISTICAS

NATURAL JOIN GRUPOS

GROUP BY GRUPOS.genero

ORDER BY obitos DESC

|  |  |
| --- | --- |
| **grupos** | **obitos** |
| M | 136733 |
| F | 119388 |
| I | 3 |

2. Qual é o número de óbitos por faixa etaria? Mostra a idade minima, idade máxima, e o número de óbitos. Ordena por idade mínima.

SELECT FAIXAS\_ETARIAS.idade\_min, FAIXAS\_ETARIAS.idade\_max, SUM(ESTATISTICAS.obitos) as obitos

FROM ESTATISTICAS

NATURAL JOIN GRUPOS

NATURAL JOIN FAIXAS\_ETARIAS

GROUP BY FAIXAS\_ETARIAS.idade\_min

ORDER BY obitos DESC

|  |  |  |
| --- | --- | --- |
| **Idade\_min** | **Idade\_max** | **obitos** |
| 65 | 120 | 214032 |
| 45 | 65 | 35428 |
| 25 | 45 | 4623 |
| 0 | 1 | 1080 |
| 15 | 25 | 578 |
| 5 | 15 | 243 |
| 1 | 5 | 140 |

3. Qual é o número de óbitos registados por região? Mostra o nome da região e o número de óbitos, ordenado pelo numero de óbitos descendente.

SELECT REGIOES.nome as regiao, SUM(ESTATISTICAS.obitos) as obitos

FROM ESTATISTICAS

NATURAL JOIN HOSPITAIS

JOIN REGIOES ON HOSPITAIS.regiao\_id = REGIOES.regiao\_id

GROUP BY REGIOES.nome

ORDER BY obitos DESC

|  |  |
| --- | --- |
| **regiao** | **obitos** |
| Região de Saúde LVT | 89571 |
| Região de Saúde do Norte | 81800 |
| Região de Saúde do Centro | 46954 |
| ... | … |

4. Qual é o número total de internamentos por hospital? Mostra o nome do hospital e o número de internamentos, ordenado pelo numero de internamentos descendente.

SELECT HOSPITAIS.hospital\_id as id,

HOSPITAIS.nome as hospital,

SUM(ESTATISTICAS.internamentos) AS internamentos

FROM ESTATISTICAS

NATURAL JOIN HOSPITAIS

GROUP BY HOSPITAIS.hospital\_id

ORDER BY internamentos DESC

|  |  |  |
| --- | --- | --- |
| **id** | **nome** | **internamentos** |
| 4 | Centro Hospitalar e Universitário de Coimbra, E.P.E. | 277306 |
| 16 | Centro Hospitalar Universitário Lisboa Central, E.P.E. | 209208 |
| 1 | Centro Hospitalar Universitário de S. João, E.P.E. | 208681 |
| ... | … | … |

5. Quais as regiões com o número de ambulatorios maior que 200000? Mostra o nome da região e o número de ambulatorios. Ordenados pelo número de ambulatorios decrescente.

SELECT REGIOES.nome, SUM(ESTATISTICAS.ambulatorio) as ambulatorios

FROM ESTATISTICAS

NATURAL JOIN HOSPITAIS

JOIN REGIOES ON HOSPITAIS.regiao\_id = REGIOES.regiao\_id

GROUP BY REGIOES.regiao\_id

HAVING SUM(ESTATISTICAS.ambulatorio)>200000

ORDER BY ambulatorios DESC

|  |  |
| --- | --- |
| **regiao** | **ambulatorios** |
| Região de Saúde do Norte | 1445827 |
| Região de Saúde LVT | 1213020 |
| Região de Saúde do Centro | 550336 |

6. Qual é o número de óbitos por doença? Mostra o id da doença, a descrição da doença e o número de óbitos, ordenado pelo id da doença.

SELECT DOENCAS.doenca\_id as id,

DOENCAS.descricao as doenca,

SUM(ESTATISTICAS.obitos) as obitos

FROM ESTATISTICAS

NATURAL JOIN DOENCAS

GROUP BY DOENCAS.doenca\_id

ORDER BY obitos DESC

|  |  |  |
| --- | --- | --- |
| **id** | **doenca** | **obitos** |
| 11 | Doenças do aparelho respiratório | 63951 |
| 18 | Neoplasias | 50711 |
| 9 | Doenças do aparelho circulatório | 48347 |
| ... | … | … |

7. Qual a percentagem de óbitos por doença? Mostra o id da doença, a descrição da doença e, dos registados (óbitos+internamentos+ambulatório), a percentagem de quantos foram óbitos. Ordena pela percentagem descendente.

SELECT DOENCAS.doenca\_id as id,

DOENCAS.descricao as doenca,

AVG(ESTATISTICAS.obitos\*1.0/(ESTATISTICAS.obitos+ESTATISTICAS.ambulatorio+ESTATISTICAS.internamentos))\*100 AS percentagem

FROM ESTATISTICAS

NATURAL JOIN DOENCAS

WHERE (ESTATISTICAS.obitos+ESTATISTICAS.ambulatorio+ESTATISTICAS.internamentos)>0

GROUP BY DOENCAS.doenca\_id

ORDER BY DESC

|  |  |  |
| --- | --- | --- |
| **id** | **doenca** | **percentagem** |
| 21 | Códigos para fins especiais | 8.1294918147706 |
| 17 | Algumas doenças infecciosas e parasitárias | 7.0559621227801 |
| 25 | Doenças Infecciosas e Parasitárias | 6.00322907300675 |
| 18 | Neoplasias | 5.64537590148629 |
| 11 | Doenças do aparelho respiratório | 3.77291554556116 |
| 9 | Doenças do aparelho circulatório | 3.72973075955407 |
| 19 | Doenças do sangue e dos órgãos hematopoéticos e alguns transtornos imunitários | 2.94983874420775 |
| 30 | Doenças do Sangue e dos Órgãos Hematopoiéticos | 2.41194916139375 |
| 15 | Sintomas, sinais e achados anormais de exames clínicos e de laboratório, não classificados em outra parte | 2.11007090525253 |
| 13 | Doenças endócrinas, nutricionais e metabólicas | 1.59746954701145 |
| 12 | Doenças do aparelho digestivo | 1.47012650834106 |
| 28 | Doenças Endócrinas, Nutricionais, Metabólicas e da Imunidade | 1.4428766180392 |
| 1 | Doenças do aparelho geniturinário | 1.43894058163794 |
| 7 | Doenças do sistema nervoso | 1.43361308417182 |
| … | ... | … |

8. Qual a percentagem de óbitos por hospital? Mostra o código do hospital, o nome do hospital e, dos registados (óbitos+internamentos+ambulatório), a percentagem de quantos foram óbitos. Ordena pela percentagem descendente.

SELECT HOSPITAIS.hospital\_id as id, HOSPITAIS.nome as hospital,

AVG(ESTATISTICAS.obitos\*1.0/(ESTATISTICAS.obitos+ESTATISTICAS.ambulatorio+ESTATISTICAS.internamentos))\*100 AS percentagem

FROM ESTATISTICAS

NATURAL JOIN HOSPITAIS

WHERE (ESTATISTICAS.obitos+

ESTATISTICAS.ambulatorio+ESTATISTICAS.internamentos)>0

GROUP BY HOSPITAIS.hospital\_id

ORDER BY percentagem DESC

|  |  |  |
| --- | --- | --- |
| **id** | **hospital** | **percentagem** |
| 40 | Unidade Local de Saúde do Litoral Alentejano, E.P.E. | 4.34971826807521 |
| 35 | Instituto Português Oncologia F. Gentil - Porto, E.P.E. | 4.14101569162814 |
| 45 | Unidade Local de Saúde do Norte Alentejano E. P. E. | 3.59377093253023 |
| 36 | Serviço de Saúde da RAM, E.P.E. | 3.46844952383235 |
| 38 | Instituto Português Oncologia F. Gentil - Lisboa, E.P.E. | 3.28331277843944 |
| 18 | Hospital da Fundação Aurélio Amaro Diniz | 3.24031202773718 |
| 2 | Centro Hospitalar Universitário do Algarve, E.P.E. | 2.90179348441246 |
| 39 | Unidade Local de Saúde do Baixo Alentejo, E.P.E. | 2.83165453117765 |
| 23 | Centro Hospitalar Barreiro/Montijo, E.P.E. | 2.7871780387374 |
| 15 | Hospital Dr. Francisco Zagalo - Ovar | 2.58570158717971 |
| 20 | Unidade Local de Saúde da Guarda, E.P.E. | 2.57172481994531 |
| 13 | Hospital Distrital de Santarém, E.P.E. | 2.48669190368154 |
| 41 | Unidade Local de Saúde Nordeste, E.P.E. | 2.37195726330678 |
| 29 | Hospital Garcia de Orta, E.P.E. - Almada | 2.35944807694712 |
| 47 | Centro Hospitalar Tondela-Viseu, E.P.E. | 2.30135682924087 |
| … | ... | … |

9. Qual a média de dias de internamento, para cada doença? Mostra o id da doença, a descrição da doença, e a média de dias de internamento. Ordena por média descendente.

SELECT DOENCAS.doenca\_id as id,

DOENCAS.descricao as doenca, SUM(ESTATISTICAS.dias\_de\_internamento)/SUM(ESTATISTICAS.

internamentos) AS media

FROM ESTATISTICAS

NATURAL JOIN DOENCAS

GROUP BY DOENCAS.doenca\_id

ORDER BY media DESC

|  |  |  |
| --- | --- | --- |
| **id** | **doenca** | **media** |
| 6 | Transtornos mentais, comportamentais e de neurodesenvolvimento | 20 |
| 31 | Doenças Mentais | 17 |
| 21 | Códigos para fins especiais | 12 |
| 17 | Algumas doenças infecciosas e parasitárias | 12 |
| 25 | Doenças Infecciosas e Parasitárias | 11 |
| 20 | Lesões, envenenamento e algumas outras consequências de causas externas | 11 |
| 23 | Lesões e Intoxicações | 10 |
| 7 | Doenças do sistema nervoso | 10 |
| 3 | Doenças da pele e do tecido subcutâneo | 10 |
| 19 | Doenças do sangue e dos órgãos hematopoéticos e alguns transtornos imunitários | 9 |
| 18 | Neoplasias | 9 |
| 9 | Doenças do aparelho circulatório | 9 |
| 30 | Doenças do Sangue e dos Órgãos Hematopoiéticos | 8 |
| 28 | Doenças Endócrinas, Nutricionais, Metabólicas e da Imunidade | 8 |
| 13 | Doenças endócrinas, nutricionais e metabólicas | 8 |
| 11 | Doenças do aparelho respiratório | 8 |
| ... | ... | … |

10. Qual o número de óbitos de crianças, adolescentes, adultos e idosos? Mostra o número de óbitos por crianças(0<=idade\_min<15), jovens(idade\_min>=15,idade\_max<=25), adultos(idade\_min>=25, idade\_max<=65) e idosos(idade\_min<=65).

SELECT

CASE

WHEN FAIXAS\_ETARIAS.idade\_min >= 0 AND FAIXAS\_ETARIAS.idade\_max <= 15 THEN 'Crianças'

WHEN FAIXAS\_ETARIAS.idade\_min >= 15 AND FAIXAS\_ETARIAS.idade\_max <= 25 THEN 'Jovens'

WHEN FAIXAS\_ETARIAS.idade\_min >= 25 AND FAIXAS\_ETARIAS.idade\_max <= 65 THEN 'Adultos'

WHEN FAIXAS\_ETARIAS.idade\_min >= 65 THEN 'Idosos'

END AS faixa\_etaria, SUM(ESTATISTICAS.obitos) AS obitos

FROM ESTATISTICAS

NATURAL JOIN GRUPOS

NATURAL JOIN FAIXAS\_ETARIAS

GROUP BY FAIXAS\_ETARIAS

ORDER BY obitos DESC;

|  |  |
| --- | --- |
| **faixa\_etaria** | **obitos** |
| Idosos | 214032 |
| Adultos | 40051 |
| Crianças | 1463 |
| Jovens | 578 |

11. Como variou o número de óbidos ao longo do tempo? Mostra o periodo e o número de óbidos. Ordena pelo crescente.

SELECT ESTATISTICAS.periodo, SUM(ESTATISTICAS.obitos) AS obidos

FROM ESTATISTICAS

GROUP BY ESTATISTICAS.periodo

ORDER BY periodo

|  |  |
| --- | --- |
| **periodo** | **obitos** |
| 2016-01 | 4923 |
| 2016-02 | 4470 |
| 2016-03 | 4794 |
| … | … |

12. Como variaram os obitos ao longo dos anos?

SELECT SUBSTRING(ESTATISTICAS.periodo,1,4) AS ano, SUM(ESTATISTICAS.obitos) AS obitos

FROM ESTATISTICAS

GROUP BY ano

ORDER BY obitos DESC;

|  |  |
| --- | --- |
| **periodo** | **obitos** |
| 2016 | 52894 |
| 2017 | 53928 |
| 2018 | 54295 |
| 2019 | 52547 |
| 2020 | 42460 |

## Aplicação Python

|  |  |
| --- | --- |
| **“Endpoint”** | **Funcionalidade** |
| **/** | **Página de entrada** |
| **/hospitals/** | **Página com a lista de hospitais** |
| **/hospitals/<int:id>/** | **Página com estatísticas sobre um hospital** |
| **/hospitals/<int:id\_h>/<int:id\_d>/** | **Página com estatísticas sobre uma doença num hospital** |
| **/regions/** | **Página com a lista das regiões** |
| **/regions/<int:id>/** | **Página com estatísticas sobre uma região** |
| **/regions/<int:r\_id>/<int:d\_id>/** | **Página com estatísticas sobre uma doença numa região** |
| **/diseases/** | **Página com a lista de doenças** |
| **/diseases/<int:id>/** | **Página com estatísticas sobre uma doença** |
| **/queries** | **Página com uma lista de 12 queries relevantes** |
| **/query-result/<int:file\_number>** | **Página com o resultado de uma das 12 queries** |

## Conclusão

De todos os datasets, achamos que este tenha sido um dos mais relevantes devido à sua aplicabilidade real, embora o mais trabalhoso. A forma como as informações se interligavam (regiões, hospitais, grupos, doenças...) dificultou um pouco no desenvolvimento de um website completo.

Apesar disso, com a realização deste trabalho obtemos competências não só em SQLite, mas também em excel (manipulação da informação), html e css (no website) e em python (como intermediário entre a base de dados e o website).

## Referências

- Slides da matéria fornecidos no Moodle da cadeira: <https://moodle2425.up.pt/course/view.php?id=5451>

- Ferramenta SQLite;

- Ferramenta VSCode;

- Repositório no GitHub para partilha interna de ficheiros - <https://github.com/Mantorras17/BD>