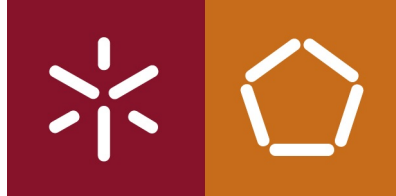


UNIVERSIDADE DO MINHO

MESTRADO EM ENGENHARIA INFORMÁTICA



Implementação de Sistema de Business Intelligence para o suporte à decisão clínica

Aplicações Informáticas na Biomedicina

André Soares, a67654
Caroline Rodrigues, pg39285
Hugo Nogueira, a81898
Mariana Marques, a85171

Janeiro, 2021

Conteúdo

1	Introdução	3
2	Base de dados	4
2.1	Estrutura	4
2.2	Dimensões	5
2.3	Factos	9
3	Tratamento dos Dados	11
4	Povoamento do <i>Data Warehousing</i>	12
5	Criação dos Indicadores	16
6	Trabalhos futuros	22
7	Conclusões	25

Lista de Figuras

1	Esquema do Data Warehouse	4
2	Dimensão Time	5
3	Dimensão Patient	5
4	Dimensão District	6
5	Dimensão External Cause	6
6	Dimensão Diagnosis	6
7	Dimensão Color	7
8	Dimensão Exam	7
9	Dimensão Reason Cancel	7
10	Dimensão Intervention	8
11	Dimensão Drug	8
12	Facto Procedure	9
13	Facto Urgency Episodes	10
14	Facto Exam	10
15	Funcionalidade de import do MySQL Workbench	12
16	Método auxiliar usado para facilitar o povoamento	12
17	Método auxiliar usado para facilitar o povoamento	13
18	Exemplo de criação da ligação aos ficheiros	13
19	Criação conexão à Base de Dados	14
20	Job para fazer o import	14
21	Quantidade de urgências por mês	16
22	Quantidade de cancelamentos por mês	17
23	Quantidade de cor	17
24	Quantidade de diagnósticos	18
25	Quantidade de medicamentos	18
26	Quantidade de exames	19
27	Quantidade de causas externas	19
28	Quantidade de género	20
29	Quantidade de género	20
30	Quantidade de escalas de dor	21
31	Secção de urgências	23
32	Secção de procedimentos	24

1 Introdução

O presente relatório foi elaborado no âmbito da entrega do Trabalho Prático da Unidade Curricular de Aplicações Informáticas na Biomedicina, que se insere no 2º semestre do 1º ano do Mestrado em Engenharia Informática. O enunciado do referido trabalho propõe que se criasse um sistema de Data *Warehousing*, bem como um sistema de *Business Intelligence* para o suporte à decisão clínica

Para desenvolvimento do projeto proposto é necessário realizar um trabalho de análise, planeamento, e implementação tendo como base o dataset fornecido. Deve ser desenvolvido um sistema de povoamento inicial, bem como estruturas necessárias à sua atualização de forma incremental e/ou diferencial. O sistema de *Business Intelligence* poderá ser desenvolvido na plataforma Microsoft *Power BI Desktop* ou *Tableau Desktop*, o grupo optou por utilizar a plataforma Microsoft *Power BI Desktop* para a realização desta tarefa.

2 Base de dados

2.1 Estrutura

Foi utilizado o MySQL para criação da base de dados do sistema. Foi criado uma estrutura dimensional que armazena as informações colectadas a partir dos arquivos disponibilizados para o trabalho.

É ainda de notar que para a realização deste esquema dimensional, optámos por retirar alguns atributos do dataset que, na nossa opinião não teriam grande importância para o estudo em causa e que mais tarde na parte da criação dos indicadores não teriam grande importância informativa. Estes atributos são:

Este é o esquema dimensional resultante do nosso planeamento, onde se encontram todas as tabelas de facto e tabelas de dimensão criadas:

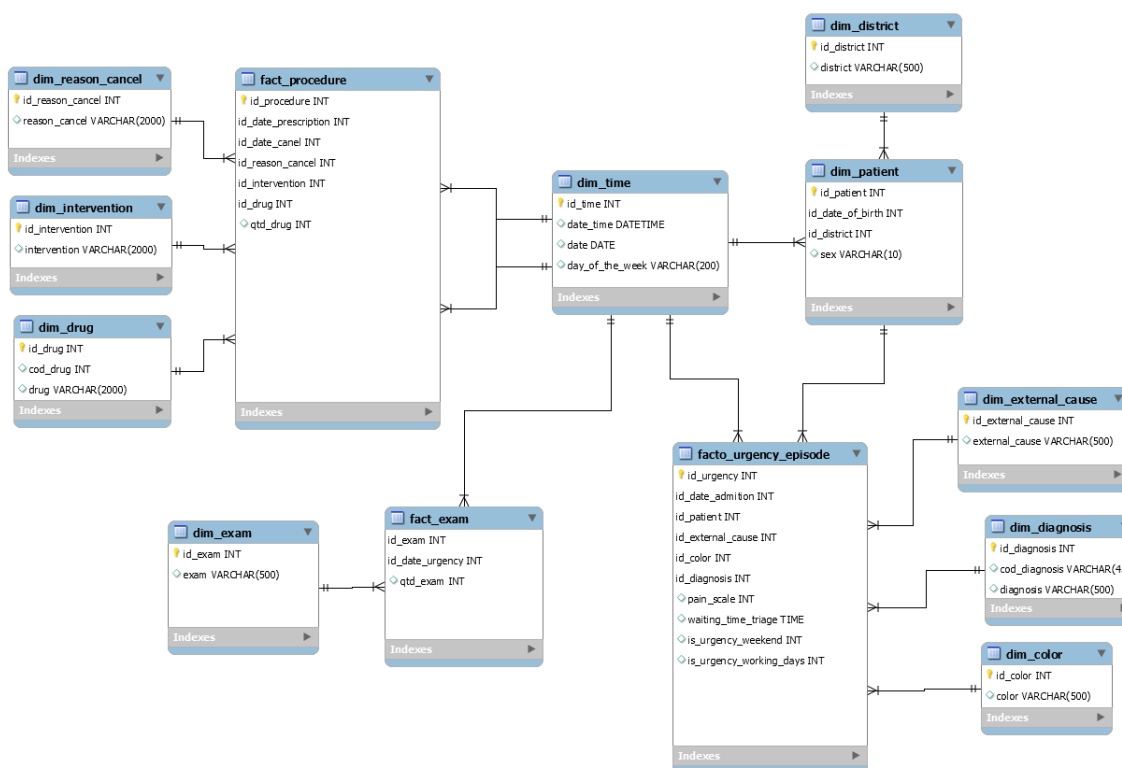


Figura 1: Esquema do Data Warehouse

2.2 Dimensões

O nosso modelo dimensional apresenta tabelas de dimensão, que apresentaremos de seguida.

- **Dimensão *Time***

Esta dimensão pretende representar as várias datas. Apresentando assim, duas ligações à tabela de facto relativa aos procedimentos, duas ligações à tabela de factos relativa ao episódio de urgência e uma ligação à tabela de facto de exame.

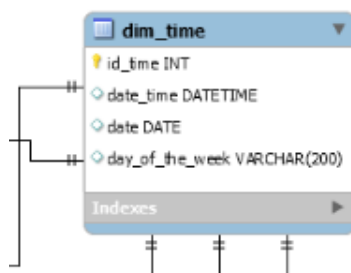


Figura 2: Dimensão Time

- **Dimensão *Patient***

Esta dimensão apresenta os atributos relativos ao paciente.

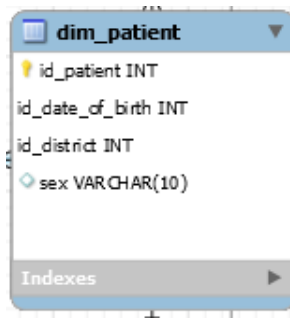


Figura 3: Dimensão Patient

- **Dimensão *District***

Esta dimensão apresenta o distrito do paciente.

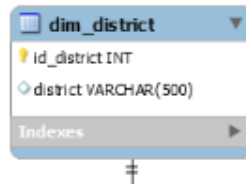


Figura 4: Dimensão District

- **Dimensão *External Cause***

Esta dimensão apresenta o atributo descrição da razão pela qual foi realizada o episódio nas urgências.

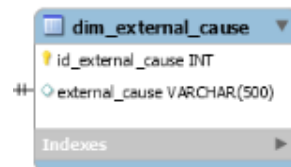


Figura 5: Dimensão External Cause

- **Dimensão *Diagnosis***

Esta dimensão apresenta o atributo da descrição do diagnóstico feito ao paciente e ainda o código de diagnóstico.

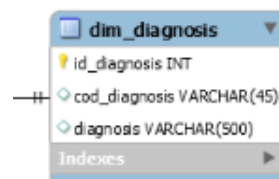


Figura 6: Dimensão Diagnosis

- **Dimensão *Color***

Esta dimensão apresenta o atributo da cor relativa ao pain scale representado na tabela de factos *fact_urgency_episodes*.



Figura 7: Dimensão Color

- **Dimensão *Exam***

Esta dimensão apresenta o atributo relativo à descrição do exame realizado.

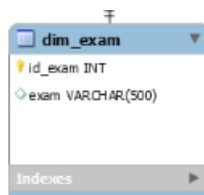


Figura 8: Dimensão Exam

- **Dimensão *Reason Cancel***

Esta dimensão apresenta o atributo descrição relativa à razão de cancelamento de um procedimento.



Figura 9: Dimensão Reason Cancel

- **Dimensão *Intervention***

Esta dimensão apresenta o atributo relativo à descrição da intervenção num procedimento.

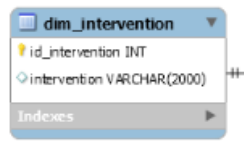


Figura 10: Dimensão Intervention

- **Dimensão *Drug***

Esta dimensão apresenta o atributo relativo à descrição do medicamento receitado e ainda o código de medicamento.

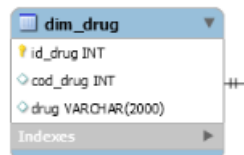


Figura 11: Dimensão Drug

2.3 Factos

Para a criação das tabelas de factos tivemos de definir que factos iríamos considerar, para futura avaliação, ou seja atributos que não sofrem alterações ao longo do tempo, para isto definimos duas tabelas de factos, sendo elas as seguintes:

- **Facto *Procedure***

Esta tabela apresenta seis chaves primárias relativas aos id únicos de: procedimento, data de prescrição, data de cancelamento do procedimento, razão para o cancelamento do procedimento e por último os id da intervenção e do medicamento receitado. Para além destas chaves primárias ainda tem outros atributos, sendo estes referentes à quantidade de medicamentos receitados.

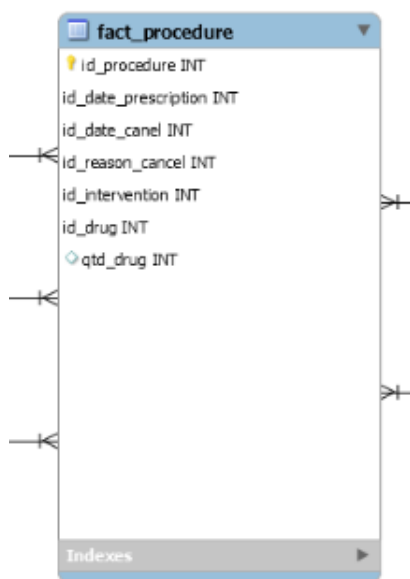


Figura 12: Facto Procedure

- **Facto *Urgency Episodes***

Para esta tabela de factos, optámos por escolher seis chaves primárias referentes a id, sendo estes : id de urgência, id de data de admissão na urgência, id relativo à causa que levou o paciente à urgência, id da cor da escala de dor do paciente e por último, id da do diagnostico feito ao paciente. Para além destas chaves primárias, esta tabela de factos possui ainda outros atributos, como a escala de dor, inteiros que indicam se o episódio de urgência ocorre num fim de semana ou num dia de semana.

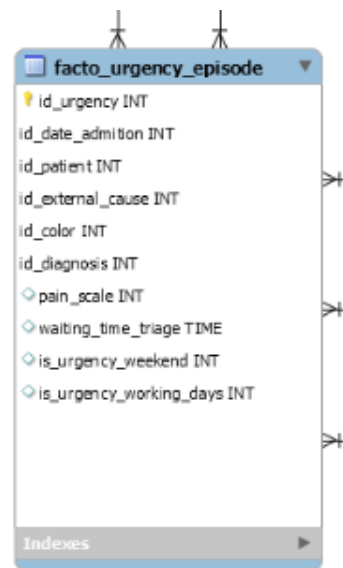


Figura 13: Facto Urgency Episodes

- **Facto *Exam***

Para esta tabela de facto, optámos por apresentar duas chaves primárias referentes aos id de exame do paciente e o id relativo à data da urgência.

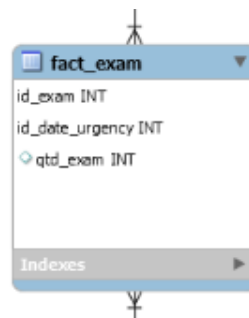


Figura 14: Facto Exam

3 Tratamento dos Dados

Nesta parte, o grupo reparou na análise ao dataset que no csv "urgency_procedures", existia uma data que raramente seria não nula devido a representar a data de cancelamento de um procedimento sendo esta "DT_CANCEL", assim optámos por realizar uma querie em SQL que permite substituir as datas nulas pela data standard "2000-01-01 00:00:00".

```
UPDATE media3.urgency_procedures SET DT_CANCEL = "2000-01-01 00:00:00"  
WHERE DT_CANCEL IS NULL;
```

4 Povoamento do *Data Warehousing*

Para proceder ao povoamento da nossa base de dados, optamos por fazê-lo através de vários passos, incluindo a funcionalidade de import do MySQL Workbench ou import através do Talend, tabelas auxiliares, e queries SQL.

Parte 1 (Imports) - Criámos tabelas auxiliares onde pudemos colocar todos os dados considerados úteis do dataset fornecido. Essas tabelas auxiliares são criadas aquando os imports e as tabelas são : "urgency_episodes_new", "urgency_episodes_procedures", "urgency_prescription" e, por último "urgency_exams".

Depois, foi feito o seguinte processo para os ficheiros seleccionados:

- Usando MySQL Workbench

- 1) Aproveitámos a funcionalidade do MySQL Workbench que facilita os imports.
- 2) Escolhemos o ficheiro a fazer o import.
- 3) Escolhemos a tabela onde adicionar os dados em questão.
- 4) Seleccionámos os atributos a inserir na respetiva tabela, e corrigimos os campos errados, que sendo preenchidos automaticamente podem estar errados quando o atributo a inserir não tem o mesmo nome que no ficheiro original.

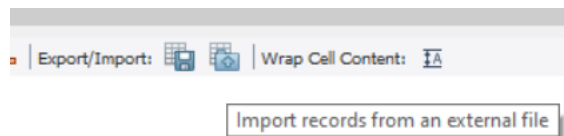


Figura 15: Funcionalidade de import do MySQL Workbench

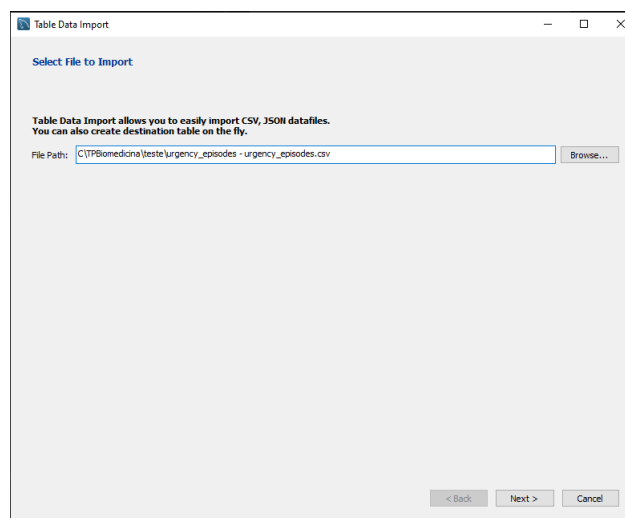


Figura 16: Método auxiliar usado para facilitar o povoamento

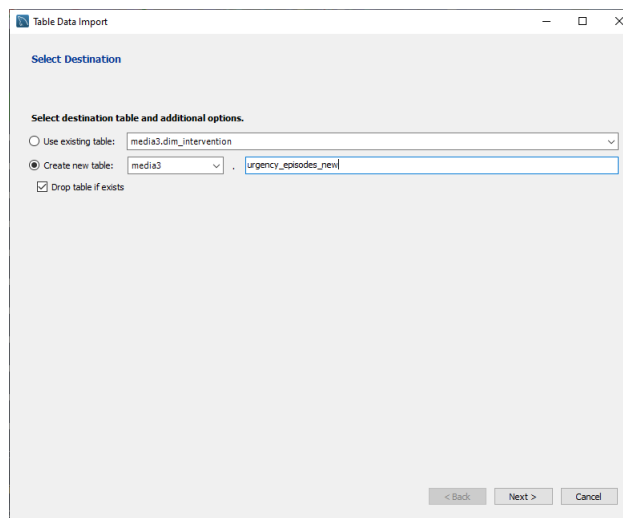


Figura 17: Método auxiliar usado para facilitar o povoamento

- Usando Talend

- 1) Criar as ligações no Talend para os ficheiros csv a importar.
- 2) Criar a ligação no Talend para a Base de Dados a usar.
- 3) Cria-se o job (ou jobs) para povoar
- 4) Seleciona-se o ficheiro pretendido, liga-se o mesmo a um nodo Tmap que depois se liga ao nodo de ligação da Base de Dados.
- 5) Executa-se o job para fazer o import.

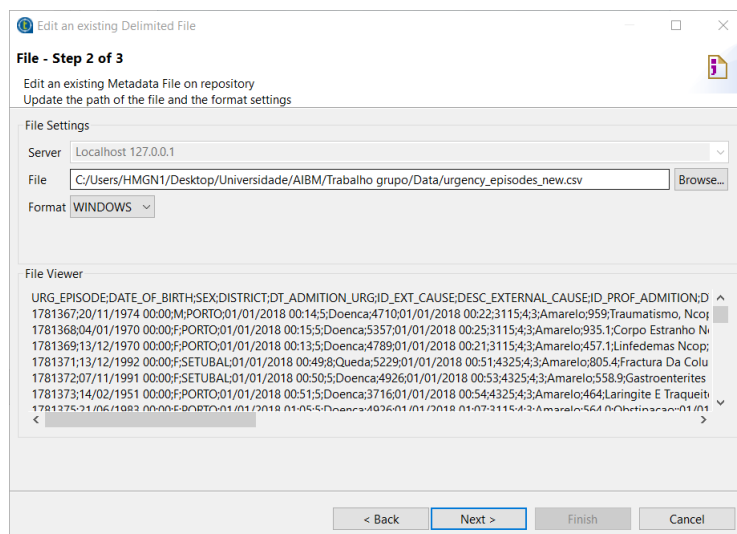


Figura 18: Exemplo de criação da ligação aos ficheiros

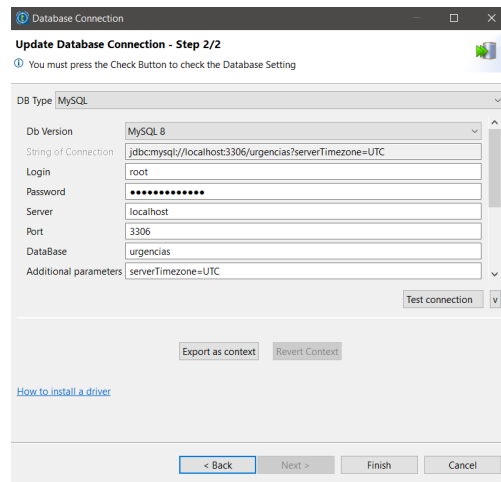


Figura 19: Criação conexão à Base de Dados

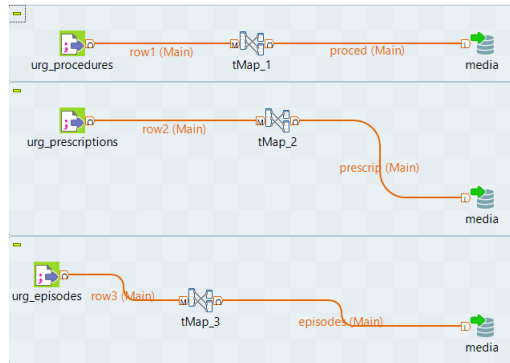


Figura 20: Job para fazer o import

Parte 2 (Inserts) - Para procedermos aos inserts dos atributos do dataset fornecido, realizamos queries baseadas na seleção de todos os valores distintos existentes na dada coluna das tabelas auxiliar (todas as queries realizadas pelo grupo serão enviadas para a equipa docente juntamente com o trabalho, desta forma aqui apresentaremos algumas a título de exemplo) :

```
INSERT INTO dim_color
SELECT DISTINCT ID_COLOR, DESC_COLOR FROM media3.urgency_episodes_new;
```

Parte 3 (Inserts Complexos) - Para as tabelas de facto foram feitos inserts mais complexos devido ao maior número de atributo:

```
INSERT INTO facto_urgency_episode
SELECT med.URG_EPISODE, ti.id_time, id_patient, id_external_cause, ID_COLOR,
id_diagnosis,
PAIN_SCALE, TIMEDIFF(DT_ADMITION_TRAIGE, DT_ADMITION_URG),
CASE WHEN weekday(DT_ADMITION_URG) IN (5,6) THEN 1 ELSE 0 END AS weekend,
```

```

CASE WHEN weekday(DT_ADMITION_URG) IN (5,6) THEN 0 ELSE 1 END AS week_day
FROM media3.urgency_episodes_new AS med
INNER JOIN dim_time AS ti ON med.DT_ADMITION_URG = ti.date_time
INNER JOIN dim_district AS dis ON med.DISTRICT = dis.district
INNER JOIN dim_time AS ti1 ON med.DATE_OF_BIRTH = ti1.date_time
INNER JOIN dim_patient AS pat ON med.SEX = pat.sex
AND ti1.id_time = pat.id_date_of_birth AND dis.id_district = pat.id_district
INNER JOIN dim_external_cause AS exC ON med.DESC_EXTERNAL_CAUSE = exC.external_cause
INNER JOIN dim_diagnosis AS dia ON med.DIAGNOSIS = dia.diagnosis;

INSERT INTO fact_procedure (id_date_prescription, id_date_canel,
id_reason_cancel, id_intervention, id_drug, qtd_drug)
SELECT dp.id_time,
dc.id_time,
rc.id_reason_cancel,
i.id_intervention,
IF(isnull(d.id_drug), 0, d.id_drug),
QT
FROM media3.urgency_procedures aux
JOIN media3.urgency_prescriptions aux2 ON aux.URG_EPISODE = aux2.URG_EPISODE
INNER JOIN dim_time dc ON dc.date_time = aux.DT_CANCEL
INNER JOIN dim_time dp ON dp.date_time = aux.DT_PRESCRIPTION
INNER JOIN dim_reason_cancel rc ON rc.reason_cancel = aux.NOTE_CANCEL
INNER JOIN dim_intervention i ON i.id_intervention = aux.ID_INTERVENTION
INNER JOIN dim_drug d ON d.cod_drug = aux2.COD_DRUG AND d.drug = aux2.DESC_DRUG;

INSERT INTO fact_exam
SELECT id_exam, id_time, count(*) FROM media3.urgency_episodes_new as med
INNER JOIN media3.urgency_exams as ex ON ex.URG_EPISODE = med.URG_EPISODE
INNER JOIN dim_time AS ti ON ti.date_time = med.DT_ADMITION_TRAIGE
INNER JOIN dim_exam AS dexm ON dexm.exam = ex.DESC_EXAM
group by id_exam, id_time;

```


5 Criação dos Indicadores

Por último, e já com a implementação terminada, temos de passar à criação de diversos indicadores referentes ao nosso modelo dimensional. Para isso, foram criadas diferentes dashboards com o intuito de fazer uma análise mais cuidada de cada indicador, podendo servir propósitos diferentes, como simplesmente para fins estatísticos.

- **Quantidade de urgências por mês**

Com a análise deste indicador, conseguimos concluir que a maioria dos episódios de urgência ocorreram no mês de agosto. Através dos KPIs criados é possível verificar que a quantidade de urgências no total foi de aproximadamente 65 mil. O dashboard também dispõe dos KPIs que indicam a quantidade de urgências no final de semana e a quantidade de urgências em dias úteis, como já é esperado, as ocorrências em dias úteis é maior que as do final de semana, conforme é possível verificar a pela figura 21.

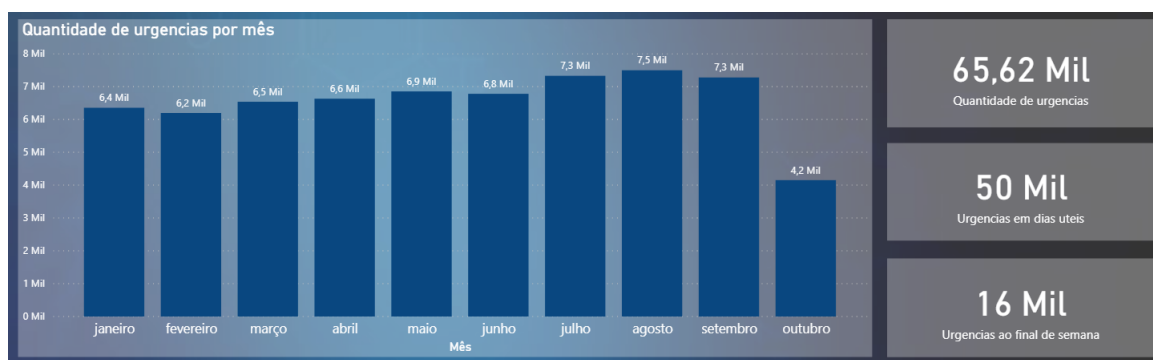


Figura 21: Quantidade de urgências por mês

- **Quantidade de cancelamentos de procedimentos por mês**

Com a análise deste indicador, conseguimos concluir que a maioria dos cancelamentos foram realizados no mês de Janeiro.

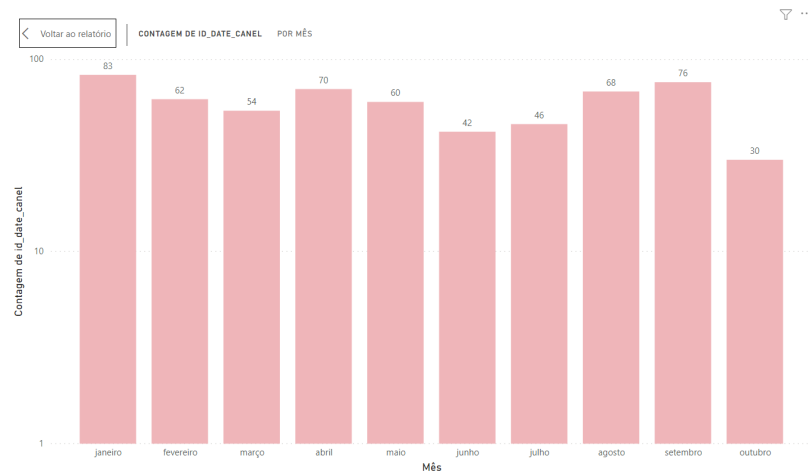


Figura 22: Quantidade de cancelamentos por mês

• Quantidade de Cor relativamente à dor

Com este indicador conseguimos perceber que a maior parte dos pacientes que frequentam esta urgência consideraram a sua dor numa fasquia de cor amarela, sendo portanto este o nível de dor intermédio.

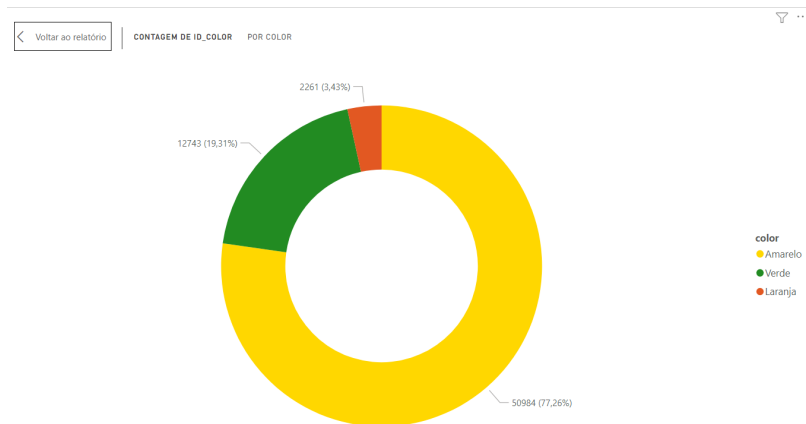


Figura 23: Quantidade de cor

• Quantidade de diagnósticos

Através da análise deste indicador, conclui-se que de entre todos os diagnósticos feitos aquele mais recorrente é "Lumbago", esta informação será importante não só para efeitos estatísticos mas também para auxiliar futuros diagnósticos.

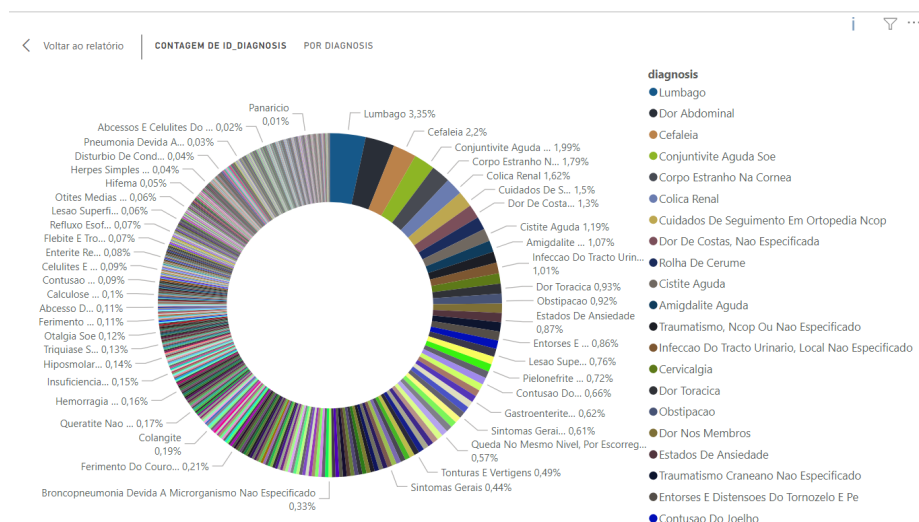


Figura 24: Quantidade de diagnósticos

• Quantidade de medicamentos

Com este indicador podemos concluir que o medicamento mais receitado nesta urgência foi o "Paracetamol 100mg, 18 unidades".

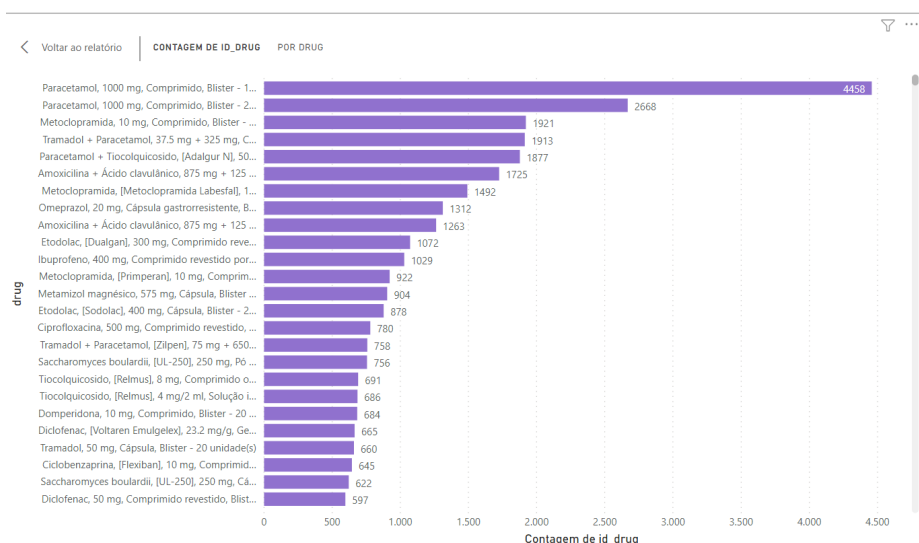


Figura 25: Quantidade de medicamentos

• Quantidade de exames

Ao analisar este indicador podemos perceber que o tipo de exame mais recorrente na urgência do nosso dataset é exames ao "torax, uma incidência", esta informação é útil ate para o próprio hospital saber gerência recursos.

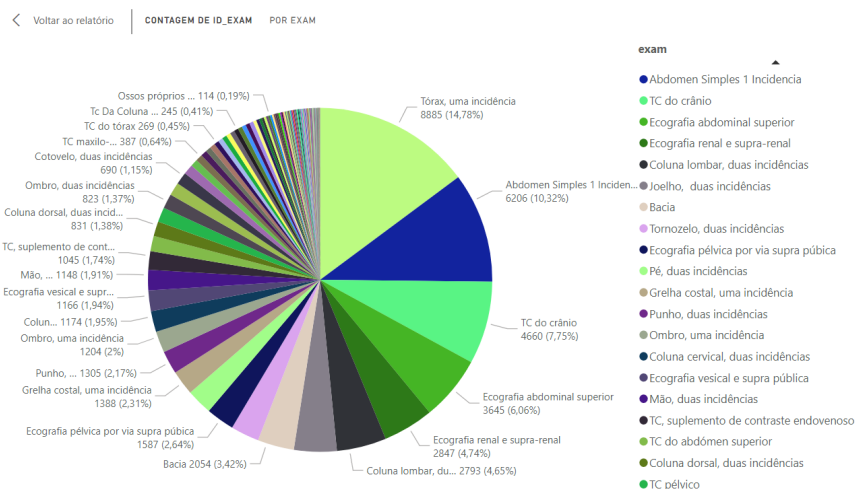


Figura 26: Quantidade de exames

• Quantidade de causas externas para ida às urgências

Através da análise deste indicador conclui-se que, a causa mais frequente para os pacientes visitarem as urgências é por motivos de doença.

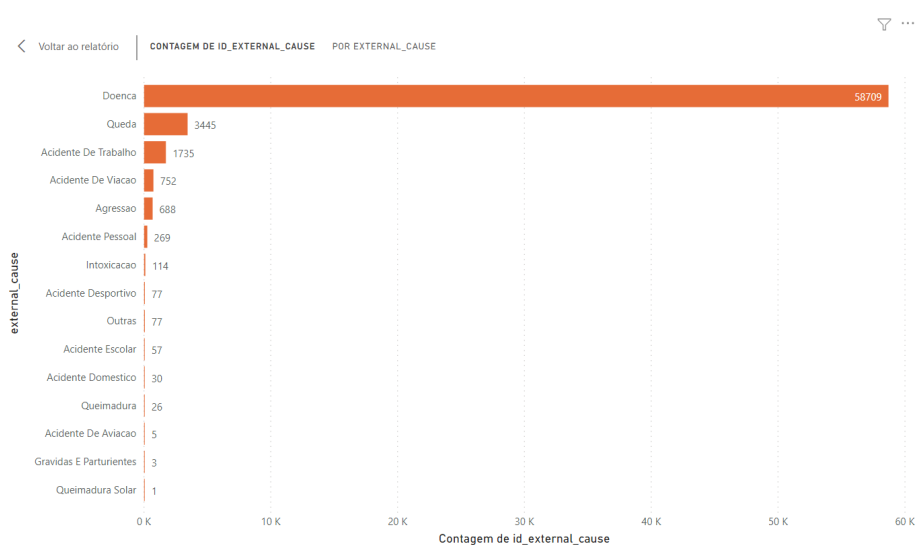


Figura 27: Quantidade de causas externas

- **Quantidade de gênero**

Com este indicador podemos concluir que a maioria dos visitantes da urgência deste dataset são do gênero masculino.

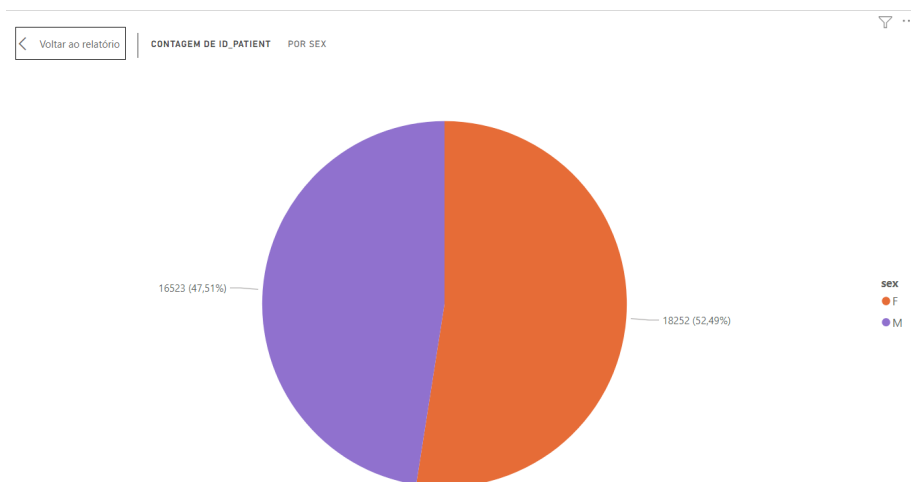


Figura 28: Quantidade de gênero

- **Quantidade de intervenções**

A partir da análise deste indicador, concluímos que nesta urgência a maioria das intervenções realizadas são do tipo "injecção Por Via IV"

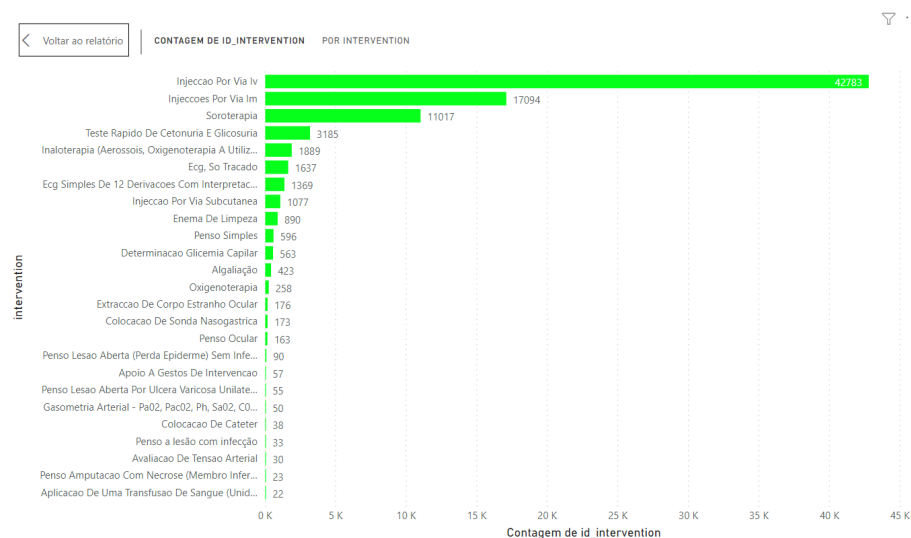


Figura 29: Quantidade de gênero

- **Quantidade de escalas de dor**

Com este indicador, podemos concluir que os pacientes destas urgências valiam a sua dor em 4 numa escala de 1 a 10, o que podemos considerar um valor intermédio de dor.

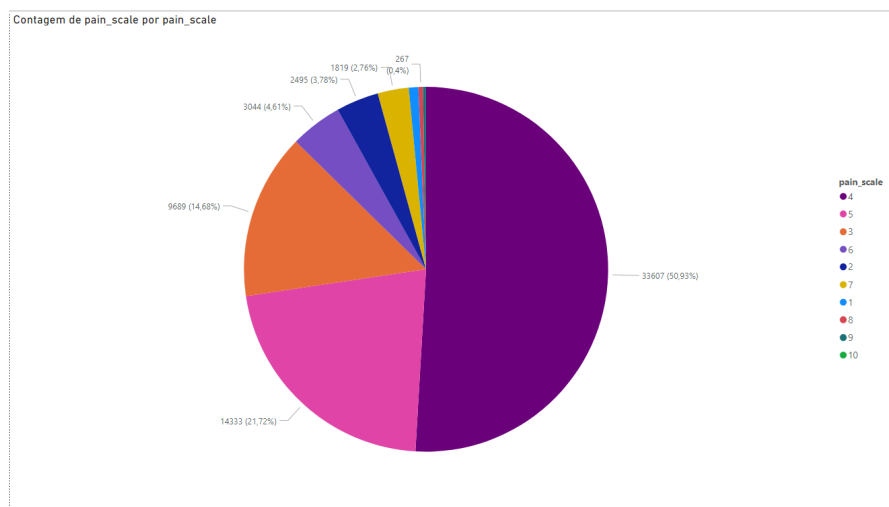


Figura 30: Quantidade de escalas de dor

6 Trabalhos futuros

Foi iniciado a criação de um dashboard com os indicadores dispostos em duas secções, uma para gestão das urgências e outra para acompanhamentos dos indicadores de procedimentos, contudo não foi possível finalizar a elaboração do mesmo. Apesar do layout estar praticamente concluído não foi possível avançar com a validação dos indicadores e correção dos bugs, por isso essa parte foi designada para a secção de trabalhos futuros.

O projeto conta com um layout melhor organizado para utilização no suporte à decisão clínica e foi utilizado para algumas análises.

- **Secção de Urgências**

Conforme é possível validar na figura 31, a página dispõe de filtros que melhoram a utilização pelo usuário, são eles: filtros de Data da urgência, Distrito, Cor, Causas externas e Diagnostico.

Há os gráficos de quantidade de urgências agrupados por mês, quantidade de urgências por causas externas, quantidade de urgências por sexo, quantidade de urgências por cor, quantidade de urgências por diagnostico, quantidade de urgências por escala de dor, quantidade de urgências por tempo de espera até a triagem, quantidade de urgências por tempo até o diagnostico e quantidade de urgências por distrito.

Conforme referido anteriormente, algumas informações ainda precisam ser corrigidas, como por exemplo o mapa com a quantidade de urgências por distrito e a as informações de tempo de espera.

- **Secção de Procedimentos**

Assim como na secção de Urgências, essa também disponibiliza os mesmos filtros. Na figura 32 é possível verificar que a página contém os indicadores de quantidade de prescrições por mês, quantidade de procedimentos por tipo medicamento, quantidade de procedimento por tipo de intervenção, quantidade de exames feitos por mês, quantidade de cancelamento agrupado por tipo, quantidade de exames agrupados por tipo.

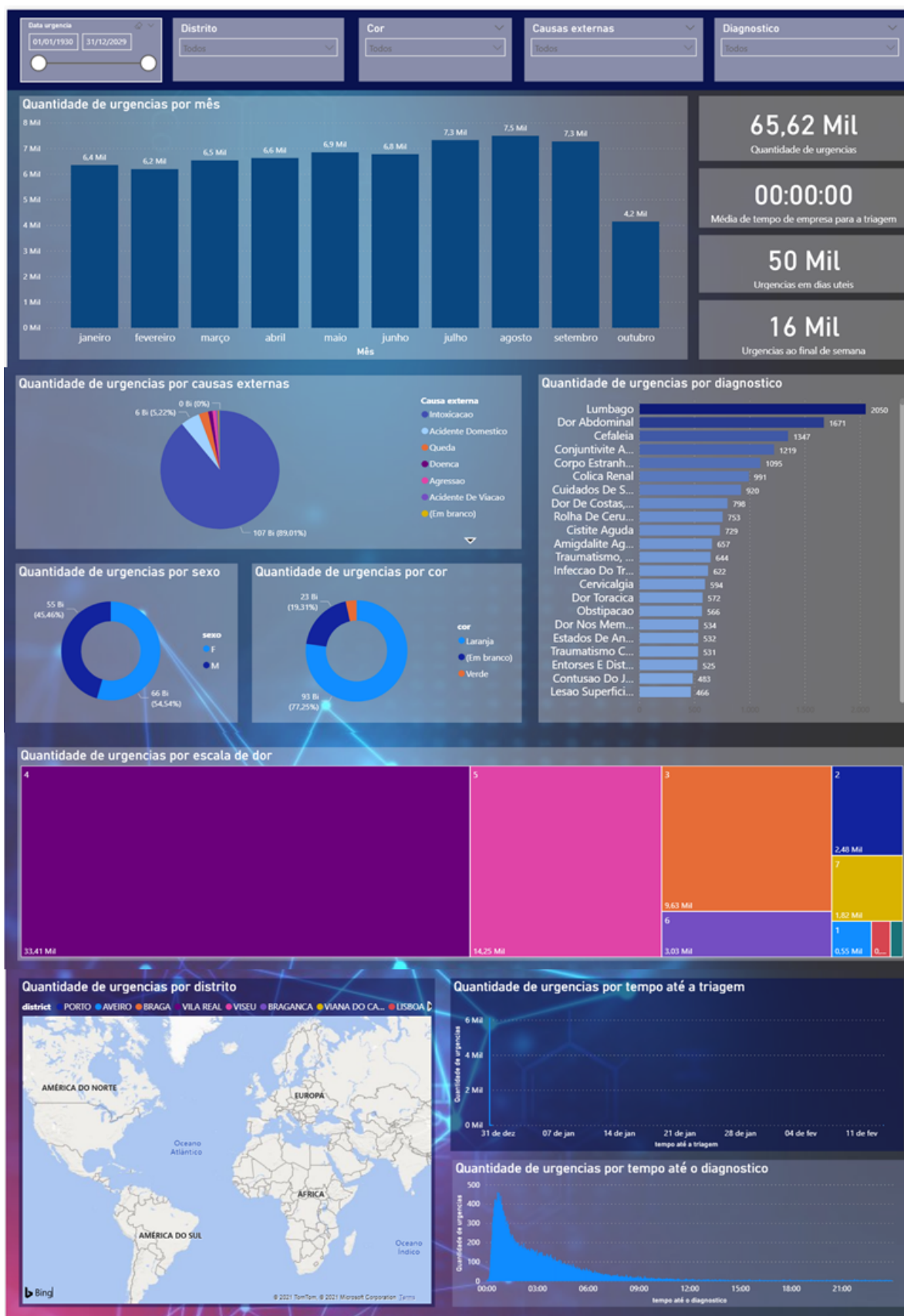


Figura 31: Secção de urgências
23

7 Conclusões

Inicialmente na realização deste trabalho enfrentámos algumas dificuldades pois o dataset fornecido pela equipa docente era bastante complexo, pelo que tornou um pouco complicada tanto a tarefa de desenvolvimento do modelo como a tarefa de povoamento Data Warehousing.

Este trabalho foi importante, pois ajudou-nos a perceber a dificuldade da implementação de sistemas de *Business Intelligence* para o suporte à decisão. Foi útil ainda para nos habituarmos a utilizar plataformas como o *Microsoft Power BI Desktop*. Permitiu-nos também aplicar os conhecimentos sobre um sistema de *Data Warehousing* com a criação de uma base de dados usando um modelo sob a forma de floco de neve.