

Universidade do Minho
Escola de Engenharia

Marta Alexandra Rolo Neiva Viana

**Extracção de Conhecimento
a partir do software *Open-Source*
de *Business Intelligence* Pentaho
em Unidade de Cuidados Intensivos**



Universidade do Minho

Escola de Engenharia

Marta Alexandra Rolo Neiva Viana

**Extracção de Conhecimento
a partir do software *Open-Source*
de *Business Intelligence* Pentaho
em Unidade de Cuidados Intensivos**

Dissertação de Mestrado
Mestrado Integrado em Engenharia Biomédica

Trabalho efetuado sob a orientação do
Professor Doutor José Machado

Novembro de 2012

DE ACORDO COM A LEGISLAÇÃO EM VIGOR, NÃO É PERMITIDA A REPRODUÇÃO DE QUALQUER PARTE DESTA TESE

Universidade do Minho, ___/___/_____

Assinatura: _____

Para a avó Clara.

Agradecimentos

Quero agradecer ao Professor José Machado pela oportunidade de realização deste projecto e por todo o apoio dado ao longo do seu desenvolvimento.

A todos os colegas que me acompanharam e ajudaram durante o percurso universitário, obrigada.

Aos meus amigos, pelos momentos partilhados em conjunto.

Finalmente, mas não menos importante, quero agradecer a toda a minha família por estar sempre presente.

Resumo

As organizações de saúde têm como principal objectivo a prestação de serviços de qualidade à população, e a tomada de decisões de forma rápida e eficaz é essencial para que tais objectivos sejam atingidos. Deste modo, neste sector, a adopção de ferramentas tecnológicas automatizadas que facilitam este processo tem vindo a aumentar ao longo dos anos. Neste contexto, surge o conceito de *Business Intelligence (BI)* que auxilia a tomada de decisão por parte dos profissionais de saúde, uma vez que estes sistemas se baseiam na *Extracção de Conhecimento (EC)* gerado pelos sistemas de informação transaccionais, sendo capazes de integrar uma enorme quantidade de dados provenientes de diversas fontes, normalmente de bases de dados que se encontram em diferentes tecnologias, plataformas e totalmente desintegradas. Assim, ultrapassando-se a heterogeneidade das bases de dados, através da estruturação dos dados, extrai-se informação que permitirá atingir conhecimento importante para as decisões clínicas.

Especificamente, a *Unidade de Cuidados Intensivos (UCI)* de um hospital é a unidade mais cara e que mais recursos exige, de tal forma que os sistemas de *BI* podem desempenhar um papel preponderante não só na racionalização dos custos, mas também na melhoria da qualidade dos cuidados prestados, através da monitorização dos dados clínicos dos pacientes. Deste modo, este projecto pioneiro incidiu na análise da aplicação do Pentaho, um software *Open-Source (OS)* de *BI*, nos processos de *EC* a estas unidades hospitalares, tendo como fonte os dados dos pacientes de um hospital localizado no Norte de Portugal, avaliando o conhecimento obtido e o seu impacto na tomada de decisão.

Este software disponibiliza ferramentas que analisam, sintetizam, assimilam e dão sentido às enormes quantidades de informação, sendo capaz de estabelecer ligações sofisticadas e discernir padrões, dando oportunidade para tirar conclusões e agir de forma preventiva.

Abstract

Health organizations have as their main objective the provision of quality services to the population, so making decisions in a fast and effective way is essential to achieve these objectives. Thus, in this sector, the adoption of technological and automated tools that facilitate this process has been increasing over the past few years. In this context, it arises the concept of BI which assists decision making by physicians, since these systems are based on the knowledge extraction generated by transactional information systems, being able to integrate a huge amount of data from various sources - typically databases which are in different technologies, platforms, and completely disintegrated. Thus, passing up the databases' heterogeneity, by structuring the data, it is possible to extract important and useful information for clinical decisions.

Specifically, the [Intensive Care Unit \(ICU\)](#) of a hospital is the most expensive unit and the one that requires more resources, such that BI systems can play an important role not only in streamlining costs, but also improving the quality of care provided, through constant patients monitoring. Thus, this pioneering project focused on reviewing the implementation of [Pentaho](#), an [OS BI](#) tool, in the knowledge extract processes at the [ICUs](#), having as data source clinical data of patients from an hospital located in the north of Portugal, assessing the knowledge gained and its impact on decision making.

This software provides tools that analyze, synthesize, assimilate and make sense of massive amounts of information, being able to discern and to liaise sophisticated standards, giving an opportunity to draw conclusions and act preventively.

Conteúdo

Figuras	VII
Tabelas	IX
Glossário	XII
1 Introdução	1
1.1 Motivação	1
1.2 Problema	2
1.3 Objectivos	3
2 Unidade de Cuidados Intensivos	5
2.1 Definição	5
2.2 Critérios	5
2.3 Objectivos	6
2.4 Recursos	6
3 <i>Business Intelligence</i>	9
3.1 BI	9
3.1.1 ETL	11
3.1.2 DW	12
3.1.3 Apresentação e visualização do conhecimento	15
4 Pentaho	21
4.1 <i>Open Source</i>	21
4.2 Módulos do Pentaho	25
4.2.1 BI-server	28
4.2.2 Pentaho Data Integration (PDI)	28
4.2.3 Pentaho Analysis	33
4.2.4 Community Dashboard Edition (CDE)	38
4.2.5 Pentaho Report-Designer (PRD)	42
4.2.6 Pentaho WEKA Data Mining	46

5	Resultados	49
5.1	Movimentos registados na UCI	50
5.2	Dias de internamento	61
5.3	Análise semestral dos movimentos diários registados na UCI	71
5.4	Reinternamentos	76
5.5	Cirurgias realizadas a pacientes da UCI	90
5.6	Data mining com WEKA	104
6	Discussão dos resultados	109
6.1	Resultados obtidos	109
6.2	Manipulação do software	109
6.2.1	Pentaho BI-server	110
6.2.2	PDI	110
6.2.3	Analysis View	111
6.2.4	CDE	111
6.2.5	PRD	112
6.2.6	Weka	112
7	Conclusões e trabalho futuro	115
	Referências	121
A	Relatório semestral dos movimentos da UCI	123

Lista de Figuras

1.1	Esquema representativo do problema do projecto.	3
1.2	Objectivo do projecto.	3
3.1	Diagrama representativo dos processos de BI [1].	10
3.2	Processo ETL [2].	11
3.3	Bases de dados operacionais vs. DW.	14
3.4	Etapas do processo de EC através do <i>data mining</i> [3].	18
4.1	Pentaho Open BI Suite [4].	25
4.2	Pentaho BI Suite Enterprise Edition [5].	27
4.3	Login para aceder ao Pentaho BI Server.	28
4.4	PDI: <i>Hops</i> e <i>steps</i>	30
4.5	Interface e página inicial do Spoon.	31
4.6	Componente “ <i>View</i> ”: visualização das ligações, <i>steps</i> , <i>hops</i> , <i>partition schemas</i> , <i>slave server</i> , <i>kettle cluster schemas</i> das trans- formações criadas.	31
4.7	Componente “ <i>Design</i> ”: <i>Steps</i> disponíveis para construção da transformação.	32
4.8	Cubo representativo da análise multidimensional. (Figura adap- tada de [6]).	34
4.9	Opção <i>New Analysis View</i>	34
4.10	Exemplo de um hipercubo, com as dimensões e medidas definidas.	35
4.11	Barra de ferramentas disponibilizada no módulo.	35
4.12	<i>OLAP Navigator</i>	36
4.13	Grelha de dados (<i>Data Grid</i>).	36
4.14	Editor MDX.	37
4.15	<i>Sort Options</i>	37
4.16	Exemplo de um gráfico com as configurações e propriedades ajustadas de acordo com o pretendido.	38
4.17	CDE: <i>Layout</i>	39
4.18	CDE: <i>Components</i>	39

4.19	CDE: <i>Data Sources</i>	40
4.20	Página inicial do PRD: <i>Welcome Page</i>	42
4.21	Área de trabalho do PRD.	43
4.22	Elementos gráficos disponíveis.	44
4.23	Painel estrutural (<i>Structure</i>).	44
4.24	Painel estrutural (<i>Data</i>).	45
4.25	Painel de propriedades.	45
4.26	Painel de mensagens.	45
4.27	WEKA: Pré-processamento dos dados.	47
5.1	Diferentes tipos de movimento registados na UCI e a sua evolução ao longo dos anos.	51
5.2	Comparação entre movimentos registados na UCI, ano a ano.	53
5.3	Registo das admissões mensais efectuadas na UCI de 2009 a 2012.	54
5.4	Altas registadas ao longo dos meses efectuadas na UCI de 2010 a 2012.	55
5.5	Óbitos registados ao longo dos meses na UCI de 2009 a 2012.	56
5.6	Número de pacientes transferidos da UCI para outras unidades hospitalares, desde 2009 a 2012.	57
5.7	Número de pacientes transferidos da UCI para outros hospitais, desde 2009 a 2012.	58
5.8	Dashboard com filtro: registo das admissões mensais efectuadas na UCI de 2009 a 2012.	61
5.9	Relatório elaborado através do PRD com informações estatísticas acerca dos dias ocorridos entre uma admissão e transferência interna, para os anos 2010, 2011 e 2012.	62
5.10	Relatório elaborado através do PRD com informações estatísticas acerca dos dias ocorridos entre uma admissão e transferência externa dos pacientes, para os anos 2010, 2011 e 2012.	63
5.11	Relatório elaborado através do PRD com informações estatísticas acerca dos dias ocorridos entre admissão até obtenção de alta, para os anos 2010, 2011 e 2012.	63
5.12	Relatório elaborado através do PRD com informações estatísticas acerca dos dias ocorridos entre as admissões e óbitos, para os anos 2010, 2011 e 2012.	64
5.13	Número de dias de internamento na UCI por paciente.	65
5.14	Número de pacientes por dias de internamento.	66
5.15	Transformação PDI criada para obtenção dos dias passados entre a admissão do paciente até a sua saída da UCI.	67

5.16	Resultado de uma consulta SQL feita à base de dados, relativa ao episódio=10004889.	68
5.17	Step <i>Calculator</i> , utilizado para cálculo dos dias decorridos. . .	68
5.18	Pré-visualização do relatório e configuração dos atributos disponíveis.	70
5.19	Transformação PDI: processo ETL para criação do DW.	71
5.20	Análise semestral dos movimentos diários registados na UCI, nos 1º e 2º trimestre de 2012.	73
5.21	Especialidades físicas ("especialidade_prev") de pacientes cuja especialidade responsável é a UCI.	74
5.22	Especialidades físicas ("especialidade_prev") de pacientes cuja especialidade responsável é a UCI. Ocultação da especialidade 30500 (UCI) no gráfico circular para melhor visualização das outras especialidades.	74
5.23	Número de reinternamentos ocorridos desde 2010 a 2012 e respectivas especialidades de destino.	76
5.24	Especialidades de reinternamentos ocorridos em 2010.	77
5.25	Especialidades de reinternamentos ocorridos em 2011.	77
5.26	Especialidades de reinternamentos ocorridos em 2012.	78
5.27	Relatório com análise anual do número total de readmissões, readmissões na UCI ou noutras especialidades.	79
5.28	Listagem de pacientes reinternados com tempo de internamento superior a 50 dias.	80
5.29	Especialidades de destino de reinternamento efectuado em menos de 24h, em menos de 48h e em menos de 72h, após prévio internamento na UCI.	81
5.30	Transformação criada através do PDI para extracção dos dados relativos aos reinternamentos ocorridos em 2010 e o seu carregamento para a tabela REINTERNAMENTOS_2010.	82
5.31	10 cores padrão reconhecidas pelo Pentaho.	83
5.32	Propriedades dos componentes no módulo CDE: "Colors".	84
5.33	Visualização gráfica interactiva com disponibilização dos dados referentes a cada barra.	85
5.34	Códigos de cores HTML introduzidos.	85
5.35	Elaboração do relatório relativo às readmissões hospitalares – PRD.	86
5.36	Esquema utilizado no PDI e resultado obtido disponibilizado num ficheiro de texto.	87
5.37	Step <i>Filter Rows</i> da transformação PDI	87
5.38	Transformação PDI com utilização do step <i>Merge Join</i> e resultado obtido.	88

5.39	Resultados da execução provenientes da transformação PDI com utilização do step <i>Filter Rows</i>	89
5.40	Resultados da execução provenientes da transformação PDI com utilização do step <i>Merge Join</i>	89
5.41	ETL: transformação criada através do PDI para saber quais os pacientes reinternados que faleceram.	90
5.42	Dashboard com o tipo de cirurgias (urgente, reoperação e reoperação em menos de 24h) efectuadas aos pacientes da UCI.	91
5.43	Dashboard com as patologias responsáveis pelas reoperações efectuadas em menos de 24h.	91
5.44	Tipo de intervenção cirúrgica a que os pacientes falecidos foram submetidos.	92
5.45	Tipo de cirurgias realizadas segundo o sexo dos pacientes da UCI.	93
5.46	N.º cirurgias realizadas, segundo a especialidade e sexo dos pacientes (1 – sexo masculino; 2 – sexo feminino).	94
5.47	Dashboard com algumas das patologias responsáveis pelas cirurgias urgentes realizadas a pacientes da UCI em 2012.	95
5.48	Análise qualitativa das especialidades das cirurgias com carácter urgente realizadas a pacientes da UCI, em 2012.	96
5.49	Cirurgia referente ao episódio 12000008.	96
5.50	Cirurgia referente ao episódio 12001252.	97
5.51	Relatório com os dados estatísticos acerca dos tempos de ocupação de sala, de cirurgia e anestesia, em segundos, relativos às cirurgias dos pacientes da UCI.	97
5.52	Configuração da base de dados e definição da consulta SQL para criação do cubo de análise multidimensional.	99
5.53	Modelo populado pelas dimensões e medidas por <i>default</i>	100
5.54	Escolha do <i>Schema</i> e <i>Cube</i> para proceder à análise multidimensional.	100
5.55	<i>OLAP Navigator</i>	101
5.56	<i>Data Grid</i>	101
5.57	Propriedades gráficas.	102
5.58	Excel: ficheiro .xls obtido da análise feita ao tipo de cirurgias realizadas, segundo o sexo dos pacientes.	103
5.59	Transformação criada através do PDI para cálculos estatísticos relativos aos tempos 5, 6 e 7.	104
5.60	Transformação PDI para obtenção do ficheiro .arff com os dados relativos à tabela bicirurgiasano.	105
5.61	Clusters obtidos através da aplicação do algoritmo <i>SimpleK-Means</i> , considerando como classe o atributo "sexo".	106

5.62 Clusters obtidos através da aplicação do algoritmo *SimpleK-Means*. 107

Lista de Tabelas

3.1	Comparação entre as tecnologias OLTP e OLAP.	17
-----	--	----

Glossário

API Application Programming Interface. 33

BI Business Intelligence. iii, v, 2, 3, 9, 10, 13, 15, 20, 21, 23–25, 109, 115, 116

CDE Community Dashboard Edition. 21, 38, 59, 69, 70, 75, 81–83, 86, 98, 111, 115

COSS Commercial Open-Source Software. 23

DM Data Mart. 12, 17, 20, 82

DSA Data Staging Area. 11

DSS Decision Support System. 9

DW Data Warehouse. 9, 11–15, 17, 18, 59, 67, 70, 87, 89, 98, 112

EC Extração de Conhecimento. iii, 1–4, 9, 18, 19, 46, 50, 104, 112, 115

ETL Extract, Transform and Load. 9, 11, 27–30, 67, 69, 75, 81, 82, 88, 89, 98, 109–111, 115

HOLAP Hybrid On Line Analytical Processing. 16

ICU Intensive Care Unit. v

KPI Key Performance Indicators. 20

MOLAP Multidimensional On Line Analytical Processing. 16

OLAP Online Analytical Processing. 13, 15, 16, 18, 20, 24–26, 28, 29, 33, 36, 92, 98, 110, 115

OLTP Online Transaction Processing. 13, 15, 16

OS Open-Source. iii, v, 3, 21–25, 115, 116

OSI Open-Source Initiative. 22

OSS Open-Source System. 21–23

PDI Pentaho Data Integration. 21, 28, 29, 50, 67, 69, 70, 81, 82, 86, 95, 98, 102, 104, 110, 111, 115

PRD Pentaho Report-Designer. 21, 26, 42, 67, 69, 78, 79, 81, 84, 88, 95, 102, 103, 110, 112, 115

ROLAP Relational On Line Analytical Processing. 16

UCI Unidade de Cuidados Intensivos. iii, 1, 3, 5, 6, 14, 15, 20, 49–52, 62, 64, 67, 71–73, 75–81, 84, 88, 90, 94, 97, 98, 104, 109, 113, 115, 116

WEKA Waikato Environment for Knowledge Analysis. 46, 47, 49, 104, 110, 112, 113

WYSIWYG What You See Is What You Get. 42, 112

Capítulo 1

Introdução

1.1 Motivação

“Knowledge is power.”

Francis Bacon

A área da saúde está altamente dependente da informação e há uma constante necessidade de tomada de decisões com vista a oferecer bem-estar ao doente. A tomada de decisão está intrinsecamente dependente da qualidade da informação disponível que, por sua vez, depende do conhecimento adquirido por parte dos profissionais de saúde que, conseqüentemente, influenciará a assistência oferecida ao paciente.

Deste modo, numa época em que a contenção da despesa pública e a obtenção da máxima eficácia com a utilização do mínimo de recursos é uma necessidade, as tecnologias de informação ganham uma relevância acrescida.

Assim, a necessidade de monitorização constante devido às situações de risco inerentes à UCI de um hospital são factores que motivam este trabalho, uma vez que os processos de EC a partir da simulação e monitorização de dados clínicos se tornam importantes neste ambiente, por forma a capacitar e ajudar os profissionais de saúde na tomada de decisões.

Os custos e o acesso aos cuidados intensivos são questões importantes. A UCI é a unidade que apresenta os custos mais elevados dos hospitais, onde o número de camas é escasso. Deste modo, fazer uma boa gestão dos recursos utilizados nestas unidades é crucial e, por isso, a possibilidade de intervenção, monitorização e correcção, em tempo real, é uma motivação para a realização deste projecto.

1.2 Problema

Durante muitos anos, o domínio médico tem sido uma área de pesquisa que apresenta grande potencial informático. A quantidade de informação aumenta, diariamente, de forma exponencial, o que implica uma gestão imperativa e eficiente do conhecimento por parte dos serviços de saúde, de maneira a diminuir custos e erros médicos. Há, assim, necessidade de automatizar essa informação, convertendo-a em conhecimento útil e, deste modo, apoiar a tomada de decisões [7].

A necessidade das instituições de saúde analisarem grandes volumes de dados de forma automatizada e dada a impossibilidade do ser humano conseguir explorar e reconhecer padrões nesses dados, bem como ser crucial a tomada de decisões de forma rápida e com base no melhor conhecimento possível, fez surgir um novo conceito: *BI* que pode ser definido como “o processo de transformar dados em informações e, por sua vez, informações em conhecimento”[8]. Por outras palavras, é um processo de extracção de informações das base de dados, onde os dados, que são a matéria “bruta”, permitem auferir relações de interesse, padrões, ou seja, extrair conhecimento útil que seria de difícil detecção por métodos tradicionais de análise. Esse conhecimento pode ser um suporte potencialmente útil à tomada de decisão.

A maioria dos dados clínicos que documentam operações diárias numa unidade de saúde são armazenados em Sistemas de Gestão de Bases de Dados Relacionais e, por forma a gerir a dispersão de informação, as instituições de saúde podem ser proprietárias de um sistema de *BI* para transformar dados brutos em conhecimento útil. Normalmente, e devido à extensa quantidade de informações, estas provêm de diversas fontes heterogéneas, grandes e complexas. Dessa forma, surge a necessidade de integrar todos os sistemas clínicos, médicos, financeiros e administrativos, o que num processo de tomada de decisão é considerado de extrema importância para se atender à qualidade dos dados que interfere directamente no sucesso do processo de *EC* [7, 9].

Como se vê pela figura abaixo (1.1), ultrapassando-se a heterogeneidade das bases de dados que contêm os dados dos pacientes, através da estruturação dos mesmos, extrai-se informação que permitirá atingir conhecimento importante para as decisões clínicas. Tudo isto através de um sistema de *BI*.

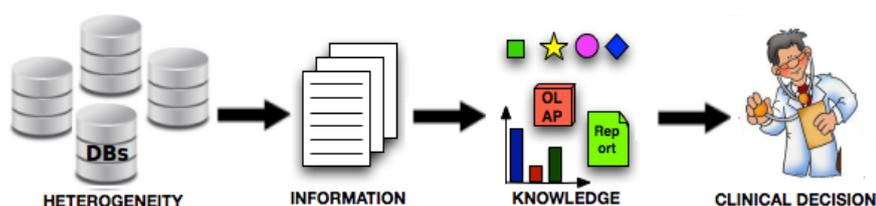


Figura 1.1: Esquema representativo do problema do projecto.

1.3 Objectivos

O grande objectivo passa por conseguir cruzar e combinar a enorme quantidade de dados provenientes das mais diversas fontes, por forma a ser possível a aplicação de ferramentas que permitam transformá-los em conhecimento que será útil ao utilizador final que, no contexto, se refere aos profissionais de saúde. Para tal, serão utilizadas as ferramentas disponibilizadas por um software *OS* de *BI*: Pentaho.

Este projecto é conceptual e genérico, na medida em que se pretende fazer uma análise descritiva e exploratória da aplicação do Pentaho na *EC* dos dados relativos a pacientes da *UCI*, perceber quais as vantagens e limitações na *EC* a partir das ferramentas disponibilizadas por ele e, deste modo, avaliar se o conhecimento obtido potencia a tomada de decisão nestas unidades hospitalares (figura 1.2). De realçar que este é um estudo pioneiro, na medida em que não se encontra documentado nenhum caso de utilização de ferramentas *OS* de *BI* aplicados a organizações hospitalares.



Figura 1.2: Objectivo do projecto.

Apesar de importantes, o principal foco do trabalho não é exactamente a essência dos resultados obtidos propriamente ditos, pois tudo isso depende

dos dados utilizados e da interpretação dos resultados - note-se que as conclusões extraídas dos resultados obtidos correspondem a interpretações pessoais dadas aos mesmos, podendo não corresponder à realidade clínica. Assim, o projecto foca-se no estudo do processo de *EC* a partir do Pentaho, das etapas envolventes e possibilidade de aplicação futura nas instituições de saúde deste tipo de software.

Fazer uma análise deste software na óptica do informático também é um dos objectivos, perceber as facilidades e barreiras encontradas durante a utilização do Pentaho no processo de *EC*.

Capítulo 2

Unidade de Cuidados Intensivos

Este capítulo define (2.1) e caracteriza (2.2) a UCI de um hospital, os objectivos (2.3) presentes nesta unidade e, por fim, aborda os recursos lá predominantes (2.4).

2.1 Definição

A UCI de um hospital corresponde a um departamento de prestação de cuidados a pessoas que se encontram em situação crítica, com risco ou falência das suas funções vitais, sendo por isso necessário um acompanhamento e monitorização constantes. Para tal, estas áreas são dotadas de um maior número de profissionais de saúde, bem como de equipamentos de tratamento específicos [10]. O internamento de um paciente na UCI é, por norma, um período transitório, cujo processo tem início no local e no momento em que se encontrava o doente quando foi vítima do acontecimento que lhe condicionou o risco de vida, e tem continuação depois da alta do serviço da UCI.

A prestação de cuidados intensivos não remete apenas para o suporte à vida, havendo um conjunto de características importantes que permitem distinguir cuidados intensivos de cuidados não intensivos, entre elas: tratamento de pacientes com elevado grau de severidade da doença que, de forma aguda ou potencial, ameaça a sua vida, integridade física ou outras funções vitais; existência de uma equipa multi-disciplinar capaz de responder de forma organizada e sustentável às necessidades dos pacientes; prestação dos cuidados numa localização definida geograficamente (neste caso, a UCI) [11].

2.2 Critérios

Uma UCI assenta e define-se segundo três critérios [10]:

- Prática - há um conjunto continuado de acções e procedimentos de monitorização, avaliação, diagnóstico e tratamento assegurados, em função das necessidades do doente, 24h por dia;
- Avaliação - existe um conjunto de variáveis analisadas, como a capacidade de recuperação das funções vitais, taxa de sobrevivência dos pacientes, entre outras, que permitem definir os recursos necessários a cada UCI e avaliar o seu desempenho;
- Investigação - a monitorização contínua e a avaliação dos resultados permitem compreender e promover a melhoria dos desempenhos organizacionais, na medida em que há acumulação de experiência e utilização do conhecimento útil adquirido dessas mesmas experiências.

2.3 Objectivos

São três os objectivos presentes dentro destas unidades, nomeadamente a preservação da vida humana, de modo a proteger e cuidar dos pacientes ameaçados por estado crítico, ferimento, ou como consequência do tratamento médico ou cirúrgico de que foram alvo - promove-se, assim, a qualidade de vida valorizável e apreciada pelo paciente; em segundo lugar, fornecer cuidados de reabilitação especializada, assim que os pacientes internados começam a recuperar da doença; e por último, para aqueles pacientes previamente internados mas já em estado terminal sem esperança de recuperação, ou para aqueles a quem se decide interromper o suporte à vida, a UCI presta um cuidado compassivo e atento aos moribundos e às suas famílias, tal como deve garantir que o sofrimento do paciente durante as suas horas finais é aliviado [11].

2.4 Recursos

Embora comumente referido, os recursos de uma UCI não se restringem apenas ao número de camas, mas também aos profissionais de saúde, às capacidades para monitorização fisiológica, diagnóstico invasivo e intervenções terapêuticas. Uma UCI encontra-se equipada com os mais modernos meios de suporte de vida, quer para monitorização do doente quer para assegurar a mais eficaz resposta terapêutica. Alguns exemplos de equipamentos que se encontram numa UCI são: ventiladores mecânicos para auxílio na respiração, através de um tubo endotraqueal ou uma traqueostomia; monitores

cardíacos, incluindo aqueles com telemetria; pacemakers externos; desfibriladores; equipamentos de diálise para problemas renais; equipamentos para a monitorização constante das funções corporais; ecógrafo e ecocardiógrafo; bombas de sucção; drenos; cateteres; entre outros [10, 11, 12].

Capítulo 3

Business Intelligence

Neste capítulo explicam-se todos os conceitos inerentes ao BI e à EC. Na secção 3.1, esta temática é introduzida, sendo abordada a sua aplicação nas instituições de saúde. Posteriormente, todas as etapas constituintes do processo de BI são explicadas em detalhe, nomeadamente: o processo de extracção, transformação e carregamentos dos dados - *Extract, Transform and Load (ETL)* (secção 3.1.1) - que remete para o pré-processamento dos dados, *Data Warehouse (DW)* (secção 3.1.2) cujo objectivo é o armazenamento dos dados e, por fim, são abordadas as técnicas para obtenção e visualização dos resultados/conhecimento (secção 3.1.3).

3.1 BI

O rápido crescimento das tecnologias e a globalização a que se assiste hoje em dia são factores que exigem que as empresas e organizações optimizem os seus processos para que estes sejam rentáveis. Para tal, há necessidade de se obter informações detalhadas, de maneira a que as organizações possam cruzar essa mesma informação e, assim, realizarem uma gestão eficaz e posicionarem-se competitivamente no mercado.

BI é um conceito que remonta às décadas de 60 e 70 (Era pré-BI), porém o seu termo é relativamente recente, tendo sido introduzido por Howard Dresner, do *Gartner Group*, em 1989, para se referir às tecnologias, aplicações e práticas utilizadas na recolha, integração, análise e apresentação de informações. O BI faz uso de sistemas que possibilitam a transformação dos dados em informações e o seu principal objectivo é o de apoiar a tomada de decisão e, por isso, é também descrito como um sistema de apoio à decisão (*Decision Support System (DSS)*). Assim, o BI suporta a tomada de decisões estratégicas que se referem à implementação e avaliação da missão, dos ob-

jectivos e metas a médio e longo prazo com impacto sobre a organização, contrariamente às decisões operacionais, mais direccionadas para a execução no dia-a-dia [13]. As principais características do BI remetem para o acesso à informação (tempos rápidos de resposta, dados coerentes, diferentes meios de consulta), apoio à tomada de decisão e orientação ao utilizador final.

A figura seguinte (3.1) faz uma representação global de todas as etapas envolvidas no BI: o processo inicia-se com a recolha de dados provenientes das mais diversas fontes (as ferramentas de BI permitem analisar dados provenientes de múltiplas fontes de forma contextualizada e integrada [1]) e culmina na apresentação aos utilizadores finais da informação e conhecimento útil obtidos, apresentados de acordo com as necessidades das organizações. As etapas intermédias são explicadas detalhadamente de seguida.

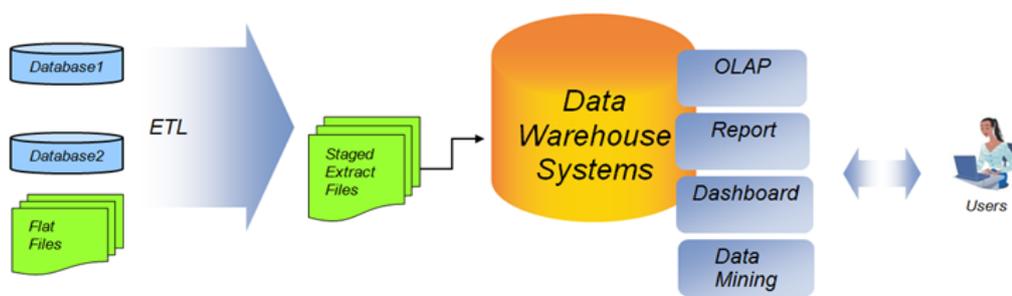


Figura 3.1: Diagrama representativo dos processos de BI [1].

BI nas instituições de saúde

O domínio sistematizado da informação permite monitorizar gastos e desempenhos, diminuindo o tempo que demoram determinados processos e os recursos gastos na sua execução. Numa instituição de saúde, é essencial que se adquiram informações de forma sistemática para que seja possível a tomada de decisões e a avaliação dos efeitos dessas mesmas decisões [14]. Durante a última década, a adopção de sistemas informáticos destinados a ajudar os profissionais de saúde e administradores nos processos de tomada de decisão, desde a análise de dados, diagnóstico a processos administrativos e de gestão, têm contribuído para a evolução e transformação dos processos clínicos, melhorando significativamente a qualidade da prestação de cuidados aos pacientes, bem como contribuído para a redução de custos.

Como já referido, o BI possibilita o acesso a ferramentas de suporte à tomada de decisão, contribuindo para a transformação do conhecimento numa vantagem competitiva [15]. Porém, os sistemas de apoio à decisão na área da saúde deparam-se com certos desafios como a existência de dados

complexos provenientes de diversas fontes, a não existência de uma terminologia padronizada em comparação com outras ciências ou a necessidade de haver um desempenho rigoroso para que não ocorram erros [16].

Ao fazer um cruzamento de informações provenientes das diversas bases de dados com os objectivos e intenções dos profissionais de saúde, através de ferramentas de apoio à decisão cada vez mais sofisticadas, sistemas de gestão e de prestação de cuidados podem ser desenvolvidos, de maneira a criar directrizes médicas mais eficazes. A partir de gráficos, por exemplo, os processos e fluxos de trabalho podem ser monitorizados, as dificuldades identificadas e os processos melhorados [17].

3.1.1 ETL

ETL é um processo que tem como principal objectivo extrair dados de base de dados, estando o seu processamento representado na figura 3.2. Estes dados são extraídos de diferentes fontes de dados, sendo processados, isto é, transformados/modificados e limpos e, em seguida, encaminhados para a área temporária de dados (**Data Staging Area (DSA)**) para posteriormente serem armazenados numa outra base de dados, o **DW**. Este é, assim, um repositório que armazena os dados de forma consolidada.

O processo de **ETL** é o mais lento e crítico aquando da construção de um **DW** devido à grande quantidade de dados a ser processada. Além disso, há uma complexa combinação de processos e tecnologias que consomem uma parcela significativa do tempo inerente a estas etapas [18, 2].

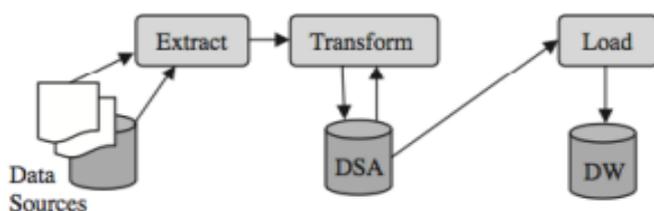


Figura 3.2: Processo ETL [2].

Uma ferramenta de **ETL** deve ser eficiente na comunicação entre bases de dados e ser capaz de ler diferentes formatos [18]. Também se deve ter conta a flexibilidade na sua alteração/modificação, porque uma vez que as bases de dados se alteram constantemente, o **DW** irá ser actualizado periodicamente e, conseqüentemente, o **ETL** também irá mudar e evoluir [2].

Tal como o nome sugere, este processo é realizado através de três fases que consistem em extrair dados de várias fontes, proceder à sua transformação e, finalmente, carregar essa informação para um DW [18].

- **Extracção:** os dados são provenientes de fontes distintas, de modo que o processo deve efectivamente integrar sistemas que possuam diferentes plataformas, linguagens, sistemas operacionais e de gestão de bases de dados e protocolos de comunicação [2, 18]. Esta fase consiste em duas etapas: extracção inicial e extracção de dados alterados. A extracção inicial refere-se aos primeiros dados extraídos e ao seu carregamento para o DW, enquanto a extracção de dados alterados corresponde à actualização contínua do DW [2].
- **Transformação:** nesta fase ocorre a limpeza, integração e optimização dos dados de entrada para obtenção de dados precisos, correctos, completos, coerentes e inequívocos. Os dados modificados não afectam as fontes de dados originais, apenas os dados contidos na extracção para o DW [2, 18].
- **Carregamento:** os dados extraídos e transformados são escritos numa estrutura multidimensional dirigida aos utilizadores finais e sistemas de aplicação. Na fase de carregamento são processados mapas sintácticos e semânticos entre os esquemas (*schemes*) [18, 2].

Depois de carregados, os dados permanecem disponíveis no DW para, então, serem processados e obter-se conhecimento útil para tomada de decisão. Às vezes, esses dados são armazenados numa etapa intermediária - em *Data Mart (DM)s* - que consolidam pequenas partes de informações relacionadas entre si, isto é, em cada DM há informação relativa a uma área específica. Seguidamente, há interligação dos DMs e é criado um DW com a informação relevante para um problema específico [18].

3.1.2 DW

Segundo Bill Inmon, considerado o pai do DW, este é definido como: "*a subject-oriented, integrated, time-varying, non-volatile collection of data that is used primarily in organizational decision making*". Detalhadamente, um DW é: orientado por temas, pois os dados relativos a diferentes assuntos são armazenados numa ou mais tabelas; integrado, isto é, os dados são padronizados e uniformizados; variável no tempo, onde cada entrada corresponde a um determinado momento específico no tempo; e não volátil, uma vez que os dados armazenados não são removidos.

Como já abordado, após procura de informações úteis em grandes bases de dados, provenientes de diferentes aplicações, quer internas ou externas, elas são estruturadas e armazenadas num repositório central, o DW. Algo que deve ser realçado é a diferença entre DW e *Data Warehousing*. DW é um repositório central de informações, podendo o BI estar-lhe subjacente, enquanto que o *Data Warehousing* corresponde ao desenvolvimento, gestão, métodos operacionais e práticas que definem a forma como os dados são recolhidos, integrados, interpretados, geridos e usados [19].

Como já referido, os dados são extraídos a partir de várias fontes, diferentes, heterogéneas e distribuídas e a informação armazenada é usada, mais tarde, em consultas (queries) e relatórios. Esta arquitectura de armazenamento de dados é parte de um modelo de dados multidimensional que permite uma análise de dados sob as mais variadas perspectivas e fornece um maior poder aos utilizadores, ao nível da tomada de decisão [2]. Desta forma, um DW fornece informação consistente, integrada e histórica, pronta para análise num sistema de BI e para ser usada nos processos de tomada de decisões de uma organização. Ao armazenar dados históricos, o DW pode fornecer informações sobre a evolução dessas informações num determinado período. Note-se que esta informação não é modificada como acontece em sistemas transaccionais, sendo incrementada, ao invés [20].

O DW suporta o *Online Analytical Processing (OLAP)*, cujos requisitos funcionais e de desempenho são bastante diferentes dos de processamento transaccional online (*Online Transaction Processing (OLTP)*), aplicações tradicionalmente apoiadas pelas bases de dados operacionais [21]. DWs tendem a ser de magnitude superior às bases de dados operacionais: enquanto as últimas rondam as centenas de megabytes, podendo atingir gigabytes de tamanho, os DWs geralmente são da ordem das centenas de gigabytes a terabytes. As principais diferenças entre bases de dados operacionais e um DW podem ser observadas na figura 3.3.

Existem algumas questões a considerar aquando da projecção e implementação de um DW numa organização: custos, tempo, os utilizadores, as pessoas, hardware e serviços. Os custos estão relacionados com quanto a organização quer gastar com hardware e software, e quais as ferramentas e serviços necessários. Tempo refere-se ao tempo disponível e necessário para a construção do DW. Utilizadores, ou seja, aquilo que os utilizadores finais precisam de obter do conjunto de dados armazenados, quais os objectivos que têm. As pessoas remete para quem vai construir e manter o DW. Por fim, hardware são as ferramentas utilizadas para construir o DW em questão, e os serviços correspondem à ajuda externa que é necessária ao longo de todo o processo [22].

Para além disto, para construir um modelo conceptual de um DW, de-

Operational Database	Data Warehouse
Time Critical	Historical data
Read/Write Accesses	Read access
Access fewer records at time	Access a lots of records at time
Data update in real-time	Data Updating periodical
Structured for OLTP(Online Transaction Processing)	Structured for OLAP(On-Line Analytical Processing)

Figura 3.3: Bases de dados operacionais vs. DW.

vemos analisar os requisitos de consulta, alterações da estrutura de dados, tempos de resposta, gestão de recursos, ferramentas de verificação, possuir um sistema de *backup* e recuperação de dados. O processo de design de um DW é muito demorado e sem as ferramentas apropriadas pode tornar-se extremamente longo.

As instituições de saúde são ambientes dominados por tecnologias da informação e, como tal, podem beneficiar de DWs clínicos. A elevada disponibilidade e utilização da tecnologia tornam os dados provenientes do DW cruciais. Um DW, combinado com ferramentas de consulta e relatórios, constitui-se promissor, apesar das dificuldades de implementação, além do *know-how* que é necessário por parte dos especialistas em saúde para lidar com estas tecnologias [23].

Dws nas UCIs

As UCIs são ambientes que possuem alto grau de tecnologia de informação e automação e, como tal, podem tirar proveito de DWs clínicos. A elevada disponibilidade e a dependência de tecnologia e de dados tornam os DW relevantes nestes ambientes. Os dispositivos na UCI suportam as funções vitais dos pacientes e, simultaneamente, produzem uma série de informações sobre o paciente: frequência cardíaca, pressão arterial, os resultados de laboratório, a ingestão de líquidos, medicação, etc.

Apesar de emergente, a utilização de DWs nas UCIs ainda não se encontra muito bem documentada, para além de haver pouca experiência na utilização de DWs com dados de pacientes da UCI para monitorização da qualidade dos serviços prestados e para pesquisa/investigação. Uma das razões é o facto dos sistemas de informação nestas unidades serem projectados para o momento dos cuidados ao paciente, o que dificulta a agregação de dados

para gestão e pesquisa. Além disso, a consulta a um DW de um sistema de informação clínico requer conhecimentos técnicos e da estrutura da base de dados, competências que a maioria dos profissionais de saúde não possui.

O custo e o acesso aos cuidados intensivos são questões importantes, ao nível de gestão. Num hospital, este é o departamento mais caro, os recursos são escassos e, por isso, para que haja boas políticas de admissão e de alta, bem como para controlar a sua aplicação, os dados sobre os pacientes da UCI são fundamentais. Além disso, uma vez que as UCIs só começaram a ser consideradas uma especialidade médica no século XXI, o conhecimento sobre o melhor tratamento para os pacientes que se encontram em estado crítico ainda está em desenvolvimento, e portanto os DWs podem ser uma poderosa fonte de dados para a pesquisa científica, especialmente para os hospitais universitários. Ora, por todas estas razões, a necessidade de analisar dados da UCI de diferentes perspectivas e a diferentes níveis - paciente, grupo de pacientes, departamento – torna-se crucial, sendo que os DWs poderão dar o seu real contributo [23].

3.1.3 Apresentação e visualização do conhecimento

Para além do OLAP já mencionado anteriormente, existem outras ferramentas de BI de exploração que fornecem a capacidade de obter e apresentar as informações que estão armazenadas no DW: consulta e relatórios; *data mining* e dashboards.

OLAP

Os dados presentes nas bases de dados OLTP são de enorme importância do ponto de vista operacional, porém, a forma como estão dispostos não ajudam na tomada de decisões. Neste contexto, surge a tecnologia OLAP para atender à necessidade de obtenção de informações que auxiliem no processo de tomada de decisão.

“On-Line Analytical Processing (OLAP) is a category of software technology that enables analysts, managers and executives to gain insight into data through fast, consistent, interactive access to a wide variety of possible views of information that has been transformed from raw data to reflect the real dimensionality of the enterprise as understood by the users.”

[Olap Council, 1997]

A tecnologia OLAP permite aos utilizadores acederem a dados organizados, agrupando subconjuntos de dados numa estrutura multidimensional

que pode responder a consultas específicas. Por outras palavras, baseia-se na análise multidimensional (análise do hipercubo) dos dados e permite ao utilizador ter uma visão mais rápida e interactiva dos mesmos. A grande vantagem das ferramentas OLAP é a comunicação entre os dados que é muito flexível e permite que o utilizador tenha uma visão analítica deles como um todo. O tempo de resposta de consulta é muito importante, já que a consulta à base de dados é o método principal desta tecnologia [21].

Numa aplicação OLAP, os dados não precisam de ser tão padronizados como no OLTP, e a base de dados não contém um grande número de tabelas. No entanto, estas são muito mais extensas. Isto porque numa base de dados OLAP a operação mais comum é o *insert*, enquanto o *delete* e *update* não são tão frequentes, sendo apenas utilizados quando ocorre algum erro no processo de carregamento dos dados [24]. Na tabela 3.1 figuram as principais diferenças entre as tecnologias OLTP e OLAP.

Dependendo da base de dados na qual os dados são armazenados, estes sistemas podem ser classificados como: **Relational On Line Analytical Processing (ROLAP)**, que armazena dados numa base de dados relacional; **Multidimensional On Line Analytical Processing (MOLAP)** que armazena dados em formatos multidimensionais; e **Hybrid On Line Analytical Processing (HOLAP)**, que é um sistema híbrido que combina os dois sistemas anteriores. As operações permitidas pelo OLAP incluem: *rollup* - aumenta o nível de agregação dos dados; *drill down* - diminui o nível de detalhe de agregação ou aumenta o detalhe ao longo de uma ou mais hierarquias dimensionais; *slice and dice* - selecção e projecção; por último, *pivot* - re-orienta a visão multidimensional dos dados [21].

	OLTP	OLAP
Fonte dos dados	Dados operacionais decorrentes de operações diárias	Dados consolidados
Base de dados	Altamente normalizada e com elevado número de tabelas	Pouco padronizada e com poucas tabelas
Objectivo	Execução de tarefas operacionais, normalmente diárias	Auxiliar a tomada de decisão
Velocidade de processamento	Elevada	Dependente da quantidade e cruzamento dos dados
Dados	Dados recentes	Séries de dados históricos

Tabela 3.1: Comparação entre as tecnologias OLTP e OLAP.

Consultas e relatórios

Consultas e relatórios são as ferramentas de análise de dados de um DW mais básicas, utilizadas para obter e disponibilizar as informações agregadas a partir de DWs e DMs [25].

As consultas e relatórios normalmente fornecem informações relativas a médias, desvios padrão e outras funções básicas de análise. Aqueles que o utilizador deve usar dependem das necessidades específicas da organização e das informações que se pretende extrair, tendo em conta o acesso aos detalhes da base de dados num curto período de tempo e com o menor custo possível [26].

Data mining

Data mining corresponde à previsão automatizada de tendências e comportamentos, permitindo a descoberta automática de padrões previamente desconhecidos. Citando Laudon & Laudon: “*Data mining is the analysis of large pools of data to find patterns and rules that can be used to guide decision making and predict future behavior*”. Assim, podem considerar-se quatro elementos chave neste processo, sendo eles: procura de padrões, grande quantidade de dados, análise preditiva e tomada de decisão [25].

Entre as técnicas utilizadas em *data mining*, podemos encontrar: redes

neurais artificiais, árvores de decisão, algoritmos genéticos, indução de regras, etc. O processo envolve a transformações dos dados, executa tarefas de classificação, regressão, regras de associação e algoritmos de clustering que podem ser usados para melhorar o desempenho futuro das organizações, através de análises preditivas [27].

Enquanto as consultas e a tecnologia OLAP exigem interação humana na busca de relações entre os dados, os programas de *data mining* possibilitam a derivação de muitas dessas relações de forma automática, através da análise e "aprendizagem" dos dados contidos nas bases de dados.

Os algoritmos aplicados aos dados pré-processados e transformados variam conforme o objectivo da EC. Os algoritmos, também denominados sistemas de aprendizagem, são os métodos utilizados para extrair o conhecimento. Nem sempre é fácil determinar qual é o mais adequado, porque nenhum método é óptimo em todas as áreas, no entanto, há certos parâmetros em que basear a escolha: o tipo e paradigmas de aprendizagem, linguagens de descrição, etc.

A abordagem é seleccionar um subconjunto de dados do DW, realizar a análise de dados no subconjunto seleccionado para identificar as principais características estatísticas e, em seguida, construir modelos preditivos. Finalmente, esses modelos de previsão são implementados na base de dados operacional [20].

Na figura 3.4 observam-se todas as etapas envolvidas no *data mining*.

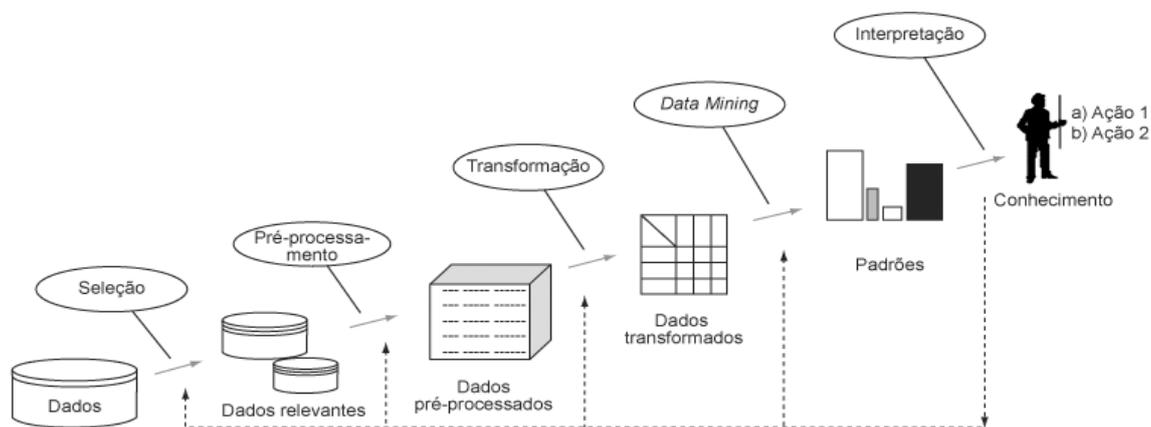


Figura 1. Etapas do processo KDD (Fayyad et al. (1996)).

Figura 3.4: Etapas do processo de EC através do *data mining* [3].

Primeiramente, tem de se analisar a base de dados de onde se pretende extrair o conhecimento e estabelecer os objectivos da aplicação, ou seja, fazer um planeamento das actividades, procurando compreender o significado e

a relevância dos atributos, avaliar a qualidade e a quantidade dos dados disponíveis, bem como documentar conhecimento prévio que possa existir.

Posto isto, procede-se à selecção de dados, escolhe-se uma amostra representativa para ser analisada e submetida ao *data mining*, uma vez que a busca de conhecimento em grandes bases de dados poderia culminar na obtenção de padrões pouco significativos ou imprecisos. Nesta fase tem-se em conta o tamanho da amostra, a homogeneidade e dinâmica dos dados, entre outros.

Antes do *data mining*, os dados seleccionados são tratados, isto é, informações inconsistentes são corrigidas ou eliminadas para que não comprometam a qualidade do conhecimento extraído.

Depois da limpeza dos dados, é efectuada a codificação dos dados, uma vez que estes têm de estar na forma correcta para serem usados como entrada dos algoritmos de *data mining*. Por fim, ainda relativamente ao pré-processamento, realiza-se um enriquecimento de dados, onde se efectuam pesquisas e consultas a base de dados externas, por exemplo, para que assim haja uma agregação de mais informações relativas aos registos já existentes, acrescentando-lhes valor. É de realçar que estas fases de selecção e preparação dos dados consomem cerca de 80% do tempo gasto em todo o processo de EC.

A segunda grande etapa corresponde ao *data mining* propriamente dito: aplicação de algoritmos específicos, que extraem padrões a partir dos dados pré-processados e transformados. Os algoritmos aplicados variam conforme o objectivo da EC.

Por último, o pós-processamento engloba: a avaliação do conhecimento extraído, tendo em conta critérios de desempenho, como a precisão e a representação desse conhecimento; interpretação dos padrões, através de ferramentas de visualização, permitindo a filtragem do conhecimento que possibilita a remoção de padrões redundantes e/ou irrelevantes; consolidação e utilização do conhecimento, através da sua incorporação num sistema inteligente, que será utilizado para ajuda à tomada de decisão [28, 29].

Todo este processo é dinâmico, interactivo e iterativo: aplicados os algoritmos de *data mining*, é gerado um *dataset* que revela padrões e, caso esses padrões sejam válidos e satisfaçam o utilizador, é atingido conhecimento, senão uma ou mais etapas serão repetidas tantas vezes quantas forem necessárias para que se chegue a um resultado satisfatório [30, 31].

Data mining na saúde

Actualmente, as ferramentas de *data mining* encontram diversas aplicações na área da saúde, essencialmente direccionadas para a previsão e identificação de tendências, como por exemplo, identificação dos pacientes

com maior probabilidade de serem readmitidos na UCI, ou detecção de pacientes com maior probabilidade de apresentarem determinadas complicações terapêuticas, análises financeiras, entre outros [27, 32, 33].

É, então, uma ferramenta capaz de ajudar os profissionais de saúde a actuar a nível preventivo, essencialmente.

Dashboards

Esta ferramenta possibilita a publicação de relatórios interactivos, com informações normalmente relativas aos desempenhos obtidos, em comparação aos objectivos ou valores-alvo estabelecidos previamente. Assim, alinham os objectivos das diferentes áreas com a estratégia da organização e a monitorização do seu progresso.

Existem dois tipos de dashboards: analíticos e integrais. Dashboards analíticos permitem obter, a partir de DMs, relatórios e indicadores chave de desempenho (*Key Performance Indicators (KPI)*), através da análise de áreas de negócios não relacionadas entre si. Na prática, é uma ferramenta de consulta que visa a obtenção e apresentação de indicadores de gestão. Já os dashboards integrais são desenvolvidos estrategicamente ao nível de toda a organização, e os diferentes níveis de gestão e liderança da organização têm uma visão estratégica com um conjunto de objectivos e indicadores que cobrem toda a organização [20].

Deste modo, através destas diferentes ferramentas de representação e visualização (*OLAP*, consultas e relatórios, *data mining* e dashboards), o utilizador final tem acesso ao conhecimento obtido e, com base nas informações disponibilizadas, poderá proceder à tomada de decisão, culminando assim o processo de BI.

Capítulo 4

Pentaho

Neste capítulo, primeiramente é abordada a temática OS, as suas vantagens, limitações e aplicações na área da saúde: na secção 4.1 é dado um *background* da evolução da aplicação de software OS nas instituições de saúde a que se assistiu ao longo dos anos e, posteriormente, particulariza-se a temática OS em relação às ferramentas de BI, onde se apresentam algumas das actualmente existentes, bem como as suas funcionalidades.

Posteriormente, na secção 4.2, é apresentado o software em estudo, o Pentaho, começando por se referir as suas características gerais, passando depois à exposição detalhada dos diferentes módulos que o constituem e que foram utilizados neste projecto. Assim, são apresentadas as interfaces, as ferramentas disponíveis em cada um deles, as suas funcionalidades e aplicações, começando pelo BI-server, seguindo-se o Pentaho Data Integration (PDI) que permite fazer o tratamento dos dados para posterior análise através do Pentaho Analysis, a sua visualização utilizando o Community Dashboard Edition (CDE) e representação fazendo uso do Pentaho Report-Designer (PRD). Por último, é apresentado o Pentaho Weka Data Mining.

4.1 *Open Source*

As instituições de saúde têm, ao longo destas duas últimas décadas, aumentado o investimento na implementação de sistemas informáticos e, nesta área, a utilização de sistemas OS - *Open-Source System (OSS)* - poderá ser um meio de acelerar a sua difusão e diminuir custos. No entanto, tudo isto depende das capacidades funcionais do software, do tipo de licença, entre muitas outras condicionantes.

Actualmente ainda é difícil perceber quais os OSS utilizados na área da saúde, por quem e como são utilizados. Ainda antes do termo *open-source*

estar difundido, Octo Barnett procedeu à distribuição do código fonte do COSTAR (*COmputer STored Ambulatory Record*), no final da década de 70 [34]. Hoje, já existe documentação relativa a aplicações OS utilizadas em imagem médica, ou de software que promove a interoperabilidade dos registos electrónicos médicos, como é o caso do *openEHR*. Também há registo de um projecto que está a ser desenvolvido na Europa, denominado *Care2X*, projectado para integrar os diferentes sistemas de informação existentes nos hospitais num sistema único e eficiente, integrando praticamente qualquer tipo de serviços, sistemas, processos, dados [35]. Segundo um estudo referido em [36], realizado em 2007, existiam cerca de 174 projectos OS associados à área da saúde, como sistemas de registo, apoio administrativo, de imagem e de apoio à decisão clínica.

Os OSS diferem do software proprietário em termos de processo de desenvolvimento e licenciamento do mesmo, dado que o primeiro é desenvolvido cooperativamente e de forma distribuída. A colaboração entre indivíduos e organizações dá origem a uma comunidade, muitas vezes referida como comunidade OS. A participação em projectos é totalmente voluntária e as equipas auto-organizam-se, sendo coordenadas por um núcleo de developers que têm a palavra final sobre todo o software desenvolvido. A qualidade é assegurada por um processo de revisão facilitada pelo acesso ao código fonte.

Todas as aplicações OS são certificadas através de uma licença concedida pela *Open-Source Initiative (OSI)*. As licenças são mais ou menos restritivas, podendo permitir o uso de forma livre, a redistribuição e a modificação do software. Deste modo, a escolha da licença para utilização de OSS nas instituições de saúde é de extrema importância e determinará os direitos dos utilizadores e até a qualidade do software desenvolvido.

Os OSS potenciam benefícios na área da saúde: é fácil de obter; reduzem a dependência a tecnologias proprietárias, aumentando a escolha, flexibilidade e interoperabilidade; os utilizadores têm a possibilidade de experimentar o software antes de adoptá-lo, sem qualquer compromisso; as organizações de saúde podem participar em projectos OS para desenvolver a sua experiência interna e influenciar a direcção do projecto para que o software vá de encontro às suas necessidades; a disponibilidade para aceder e modificar o código fonte oferece maior flexibilidade e facilidade na identificação e resolução de problemas; o próprio processo de implementação de novos módulos nos sistemas de saúde existentes acelera o progresso e induz a inovação; em termos de custos, as organizações economizam em taxas de licenciamento de software e, muitas vezes, reduzem a sua despesa em hardware porque evitam hardware proprietário [37].

Provavelmente nunca será feita a implementação total e única de OSS em instituições de saúde, uma vez que estas continuarão a obter os seus sistemas

de informação a partir de fornecedores de sistemas proprietários, sendo que os módulos e componentes OS têm de ser desenvolvidos de forma a ser possível a ligação e/ou intercomunicação com esses sistemas.

No entanto, há que considerar que as organizações ainda precisam de desenvolver os seus conhecimentos internos na adopção e gestão de OSS. A falta de experiência neste campo acaba por aumentar os custos de manutenção, implementação e suporte e, muitas vezes, poderá ser mais caro do que as soluções proprietárias. Também há necessidade de avaliar o OSS em determinados aspectos, tais como suporte de apoio e serviços de fornecedores de software, hardware compatível e a maturidade e a actividade do projecto, bem como a qualidade da produção e da taxa de adopção por outras empresas para perceber se vale a pena, ou não, a implementação do software [36, 38].

Há quem defenda que os OSS têm lacunas em termos de estabilidade, segurança e confiabilidade, o que se revela um entrave na adopção deste tipo de sistemas. Existem barreiras na implementação de OSS em instituições de saúde, tais como: os sistemas de informação nas unidades de saúde processam informação crítica, onde a segurança e a privacidade dos dados dos pacientes tem de ser salvaguardada; as instituições de saúde têm de possuir um sistema informático integrado e, muitas vezes, os sistemas OS concentram-se no desenvolvimento de componentes individuais fragmentados o que dificulta a implementação; a legislação e as políticas regulatórias são extremamente complexas [36].

Ferramentas OS de BI na área da saúde

As plataformas OS de BI normalmente são licenciadas como *Commercial Open-Source Software (COSS)*. As empresas COSS têm desenvolvido e imposto o conceito de licenciamento duplo que consiste na disponibilização de um modelo OS e um modelo comercial. Isto é, as empresas gerem a comunidade OS, através da qual se desenvolvem produtos que são distribuídos livremente e, em paralelo, oferecem um serviço de atendimento ou recursos adicionais para aqueles utilizadores que aceitam comprar tais serviços.

Desta forma, as organizações exploram todos os benefícios provenientes das comunidades OS, como o baixo custo em desenvolvimento, e têm a possibilidade de obter um alto retorno sobre o investimento. Quanto aos utilizadores, este tipo de licenciamento permite fazer o download e usar o software, sem custos e, se assim o pretenderem, pagar os serviços adicionais que necessitem.

A disponibilidade do código fonte constitui um meio propício a pesquisas e testes, bem como à criação de protótipos, o que inevitavelmente induz a

competitividade. Toda a inovação inerente ao conceito OS permite, assim, incorporar funcionalidades de BI em aplicações já existentes, bem como complementar as infra-estruturas para estender o uso de BI a mais utilizadores.

Não há casos registados de ferramentas OS de BI adoptadas por organizações de saúde, pelo que o estudo do software Pentaho neste campo é pioneiro.

Porém, existem diversas plataformas similares ao Pentaho, também OS, nomeadamente: Jasper Intelligence, BIRT (*Business Intelligence and Reporting Tools*), Vanilla e SpagoBI, entre outras. Todos estes software disponibilizam ferramentas BI com vista à melhoria dos processos das organizações e encontram-se em constante desenvolvimento, com lançamento de novos produtos e versões. De uma forma geral, as ferramentas por eles disponibilizadas vão desde plataformas para criação de relatórios, de visualização gráfica, análise dimensional (OLAP) e *data mining* [39].

Jasper Intelligence

Jasper Intelligence foi desenvolvido pela JasperSoft, tendo como lançamento inicial o JasperReports, em 2001. Posteriormente foram criadas outras funcionalidades, como o JasperServer, o JasperAnalysis e o JasperETL, disponíveis numa versão comercial. Em 2006, todas estas ferramentas foram unificadas e integradas numa só plataforma: o Jaspersoft Business Intelligence Suite.

Vanilla

Vanilla é uma plataforma de BI patrocinada pela BPM-Conseil e é constituída pelo módulos: BI Metadata Designer, OLAP Schema Designer, BI Dashboard Designer, BI Workflow Designer, KPI Metrics Designer, KPI Maps Designer, bem como pelo Report Designer Plugins para o iReport e BIRT.

SpagoBI

SpagoBI pertence à SpagoWorld e foi fundada e suportada pelo *Engineering Group*. O sistema foi desenvolvido como um framework para aplicações web em Java, por forma a atender projectos internos. Em 2004, foi hospedado no ProjectWeb, actualmente Ow2. Ao contrário do que normalmente acontece, não existem versões comerciais mais completas desta plataforma, sendo completamente livre. É constituído por diversos módulos: SpagoBiServer, SpagoBIStudio (ambiente de desenvolvimento integrado), SpagoBI Meta (ambiente focado nos metadados) e SpagoBI SDK

(camada de integração para usar o SpagoBI a partir de ferramentas externas).

4.2 Módulos do Pentaho

O Pentaho é uma plataforma de BI que inclui todos os componentes necessários à implementação de soluções baseadas em processos. Actualmente, existem duas edições disponíveis: a Community que é OS (código aberto) e a Enterprise que é uma versão não gratuita. Neste projecto foi utilizada a versão Community (Pentaho BI Server Community Edition).

O software foi desenvolvido pela Pentaho Corporation em 2004, em Java. É um projecto que está em contínuo desenvolvimento por uma comunidade OS, permitindo, assim, a inovação, bem como tirar partido das vantagens e riqueza deste tipo de tecnologias.

Como representado na Figura 4.1, o Pentaho BI Suite provê áreas de aplicação de dados e integração de diferentes aplicações, uma plataforma que fez com que este software OS se tornasse amplamente difundido a nível de relatórios, análises OLAP, dashboards, *data mining* e gestão de processos, ou seja, todas as ferramentas necessárias nas etapas do BI [4], já explicadas no capítulo 3.

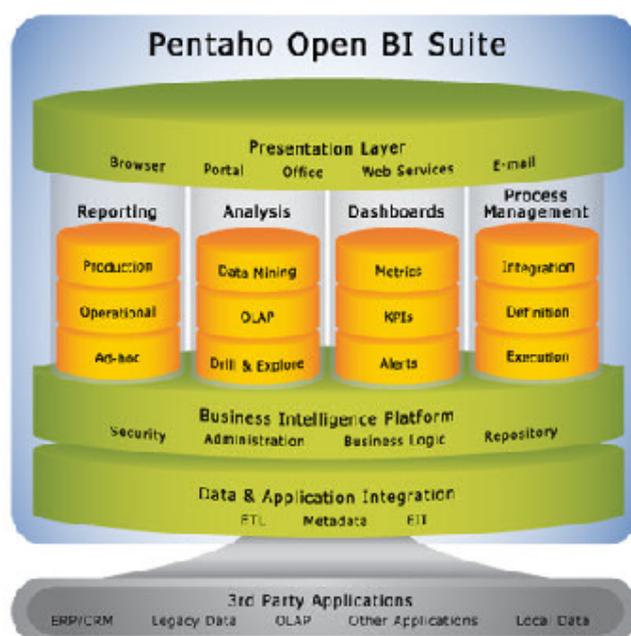


Figura 4.1: Pentaho Open BI Suite [4].

Entre as principais características do Pentaho contam-se:

- **Plataforma 100% J2EE** - Java2 Platform Enterprise Edition -, assegurando, desta forma a escalabilidade, integração e portabilidade;
- **Servidor**: qualquer um compatível com J2EE, como JBOSS, Tomcat, etc.;
- **Base de dados**: JDBC, IBM DB2, Microsoft SQL Server, MySQL, Oracle, PostgreSQL, NCR Teradata, Firebird;
- **Sistema operativo**: independente, uma vez que é uma linguagem interpretada;
- **Linguagem de programação**: Java, Java script, JSP, XSL (XSLT/XPath/ XSL-FO);
- **Repositório de dados baseado em XML**.
- Todos os componentes são disponibilizados através de **web services**.

Este software encontra-se estruturado em diversas componentes, abordadas de seguida. Cada um dos módulos será explicado detalhadamente, posteriormente.

- **Pentaho BI Platform (Bi-server)**: disponibiliza vários serviços para os utilizadores finais, tais como subscrições de agendamento, notificação e integração de ferramentas dos utilizadores finais, segurança centralizada integrada, entre outros;
- **Pentaho Reporting**: é possível desenvolver um relatório com facilidade, permitindo às organizações o acesso, a formatação e a distribuição das informações. Pentaho Reporting contém todas as funcionalidades gráficas para a construção de relatórios através do PRD, bem como relatórios direccionados para a área empresarial (Report Wizard), e ainda permite fazer consultas *ad hoc*;
- **Pentaho Analysis**: fornece análise OLAP, ajudando os utilizadores no processo de tomada de decisões e dando os *insights* necessários à compreensão do negócio. O Pentaho Analysis facilita a exploração interactiva de informações, através do cruzamento de dados, para além de oferecer uma completa integração com outros serviços disponibilizados no Pentaho BI Suite;

- **Pentaho Data Integration (PDI):** é uma poderosa ferramenta para o processo de ETL que utiliza uma abordagem inovadora, orientada a metadados.
- **Community Dashboard Edition (CDE):** dispõe de um ambiente gráfico que permite aos utilizadores terem acesso a informações críticas, essenciais ao entendimento e optimização do desempenho organizacional. É possível a completa integração com o Pentaho Reporting e Pentaho Analysis;
- **Pentaho Weka Data Mining:** é possível melhorar o desempenho futuro do negócio e/ou processos através da análise preditiva, fornecendo informações sobre padrões ocultos e relações entre os dados, bem como acerca de indicadores de desempenho. Este módulo contém uma interface gráfica para o pré-processamento de dados, classificação, regressão, *clustering*, regras de associação e visualização.

Através da figura seguinte(4.2), é possível perceber a estrutura do Pentaho BI Suite, bem como a integração e a inter-relação existentes entre os diferentes módulos.



Figura 4.2: Pentaho BI Suite Enterprise Edition [5].

4.2.1 BI-server

O BI-server é um servidor web, ao qual o utilizador acede através de um browser (figura 4.3). O servidor contém um repositório onde se tem acesso às diversas ferramentas disponíveis. Para tal, os utilizadores fazem login e após terem acesso podem procurar relatórios, executá-los, elaborar dashboards, realizar análises OLAP, ou criar relatórios *ad-hoc*, apesar desta última funcionalidade estar susceptível a remoção em versões futuras do software.



Figura 4.3: Login para aceder ao Pentaho BI Server.

4.2.2 Pentaho Data Integration (PDI)

O PDI, também conhecido como Kettle, dispõe de poderosas ferramentas de ETL, integradas num ambiente gráfico rico e atractivo, com ampla conectividade a qualquer tipo de dados e que proporciona escalabilidade. É composto por quatro ferramentas, nomeadamente a Spoon, Pan, Chef e Kitchen. O PDI permite tratar os dados para posterior análise (Analysis), visualização (Dashboards) e representação (Relatórios).

Aqui será abordado a ferramenta Spoon, pois foi a utilizada no projecto, sendo aquela que permite desenhar de forma gráfica a transformação ETL. Cada processo é criado usando uma ferramenta gráfica, onde se especifica o que se pretende fazer sem ser necessária a escrita de código para indicar como fazê-lo, ou seja, o PDI é orientado aos metadados.

Este módulo é utilizado para: transferir dados entre aplicações ou bases de dados; exportar dados de base de dados para ficheiros; carregar dados para as base de dados de forma massiva; limpeza de dados; integrar aplicações; entre outros. Deste modo, pode dizer-se que o PDI serve duas funções: o processo ETL e a integração de dados.

Concede três perspectivas diferentes: *Data Integration*, *Visualize* e *Model*. No primeiro, é desempenhado o processo ETL, onde se procede à construção gráfica do mesmo; através do segundo, os utilizadores podem visualizar os resultados obtidos após executarem uma transformação, ou no formato de relatório ou visualização analítica *ad hoc*; por último, a modelação permite aos utilizadores a criação ou modificação dos modelos OLAP e de relatórios subjacentes à integração dos dados.

O Kettle lida com quatro tipos básicos de componentes: *hops*, *steps*, *transformations* e *jobs*. Apesar de serem distintos, os dois últimos correspondem a conjuntos de itens interligados.

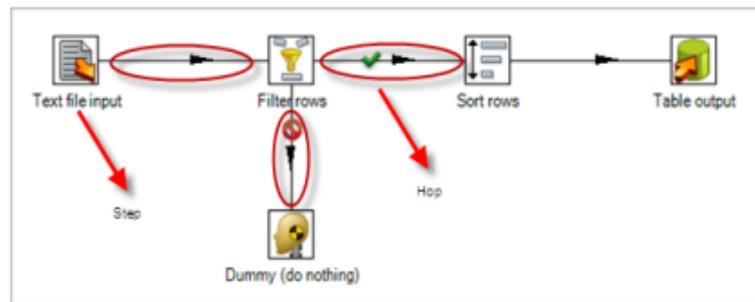
Os *hops* permitem a ligação entre *steps* e, que assim, os dados passem de uma etapa para outra. Um *hop* tem apenas uma origem e apenas um destino. Um *hop* preto indica que o próximo componente é sempre executado, um *hop* verde que o próximo componente é executado apenas se o anterior for executado com sucesso e, por último, um *hop* vermelho indica que o próximo componente será executado se o anterior falhar a sua execução.

Um *step* é a menor unidade dentro de uma transformação. Os *steps* estão agrupados em categorias, como *Input* e *Output*, entre outras. Cada *step* desempenha uma função específica, como ler um parâmetro ou normalizar um conjunto de dados, e quando os dados são transmitidos através dele, eles podem ser alterados e/ou filtrados, ou apenas instruídos por forma a corresponder a padrões. Há muitos *steps* diferentes disponíveis no PDI, mas caso estes não sejam suficientes para o utilizador fazer o que pretende, existem *steps* de script disponíveis, onde o SQL ou JavaScript podem ser usados. Um *stream* refere-se aos dados enviados de um *step* para outro.

Na figura 4.4 visualizam-se *hops* e *steps*.

Uma transformação corresponde a um conjunto de etapas (*steps*) e pode ser vista como um grande fluxo de dados, ou seja, tem um início (*input*) e um fim (*output*) e, no meio, todos os dados com que se está a lidar são canalizados através de toda a transformação pela ordem definida pelo utilizador. O fluxo dos *steps* da transformação ocorre de forma simultânea e assíncrona. Uma transformação é criada após construção do esqueleto da mesma, usando *steps* e *hops*, e após configuração dos *steps*, especificando seu comportamento.

Um *job* consiste no conjunto de uma ou mais transformações que serão executadas por ordem, dependendo de como ele está configurado. Os dados não fluem de uma transformação para outra, apenas são executadas

Figura 4.4: PDI: *Hops* e *steps*.

ordenadamente, ou seja, o fluxo dos *steps* do *job* ocorre de forma sequencial. Este processo pode ser usado para agendar tarefas, como por exemplo, correr todas as transformações actualizadas para tabelas de um DW, a determinado período. Neste projecto, porém, não foram criados *jobs*, apenas transformações.

Atendendo à interface, a página inicial pode ser visualizada na figura 4.5. No lado esquerdo do painel tem-se acesso aos componentes *View* e *Design*, como se pode ver nas figuras 4.6 e 4.7, sendo o primeiro uma forma de visualizar tudo o que está criado, e o segundo contém os meios para a criação e desenvolvimento dos processos ETL.

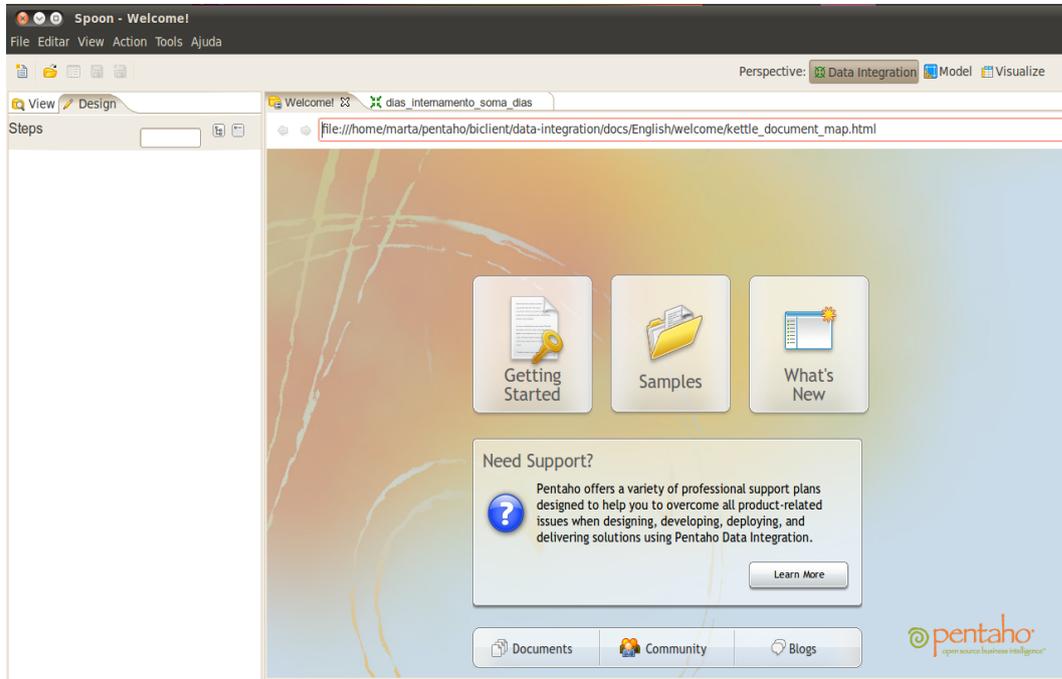
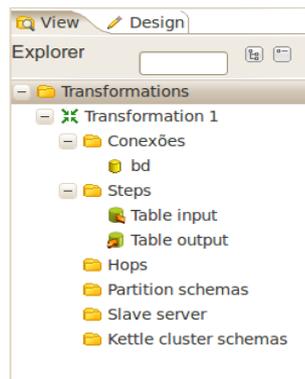


Figura 4.5: Interface e página inicial do Spoon.

Figura 4.6: Componente “View”: visualização das ligações, *steps*, *hops*, *partition schemas*, *slave server*, *kettle cluster schemas* das transformações criadas.

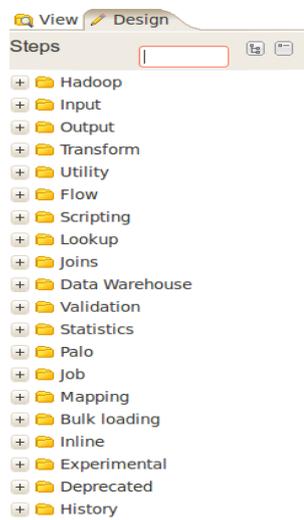


Figura 4.7: Componente “*Design*”: *Steps* disponíveis para construção da transformação.

4.2.3 Pentaho Analysis

Pentaho Analysis, também conhecido como Mondrian, é um servidor OLAP que permite aos utilizadores analisar grandes quantidades de dados em tempo real. O grande objectivo deste módulo visa analisar um grande volume de dados, de uma forma intuitiva, simples e rápida para o utilizador. Este módulo implementa três *Application Programming Interface (API)*s: as linguagens MDX, XML, bem como especificações JOLAP que interpreta instruções dadas em SQL. Depois de definido o modelo das consultas, é gerada uma página web, onde é feita a análise interactiva dos dados, através do cruzamento de informações, e que envolve diversas operações, como *drill up*, *drill down*, *slice and dice*, entre outras.

O Pentaho Analysis permite então fazer uma análise multidimensional das estruturas de dados utilizadas, concedendo controlo sobre as mesmas, além de dar a capacidade de trabalhar independentemente das fontes de dados originais, ou seja, é possível alterar os dados sem modificar directamente as bases de dados originais.

Para os mesmos dados, consegue-se obter múltiplas visualizações e assim satisfazer diversos requisitos, oferecendo, deste modo, várias soluções ao utilizador. O processo passa pela construção dos cubos que segmentam os dados em eixos de análise (figura 4.8), segundo as perspectivas pretendidas pelo utilizador, dada a possibilidade de decomposição dos valores para se obter a solução final. Um cubo de dados é, deste modo, uma estrutura que armazena os dados em formato dimensional. Numa análise multidimensional lida-se com dimensões, onde cada uma delas constitui uma unidade de análise com dados agrupados e representa um eixo principal no estudo dos dados, como se observa na figura 4.8. Cada dimensão pode possuir níveis hierárquicos e, posteriormente, há cruzamento entre uma ou mais dimensões, para se obter a(s) medida(s) - *measure(s)*. A dimensão *Measure* é especial, na medida em que contém todos os factos e números relativos ao conteúdo utilizado.

A interface disponibilizada pelo Pentaho Analysis consiste numa barra de ferramentas que permite configurar a forma como os dados são visualizados, bem como de interagir com os mesmos.

Como já referido, os dados introduzidos compõe-se de dimensões hierárquicas, que podem ser utilizadas para dividir e analisar os dados com o objectivo de obter diferentes perspectivas da informação ou até para obter detalhes adicionais acerca desses dados. A informação pode então ser cruzada entre as diferentes dimensões, através de diversas ferramentas disponibilizadas.

O primeiro passo para efectuar a análise OLAP é criar uma nova base de dados ou alterar uma já existente (opções *Create New* e *Manage Existing*),

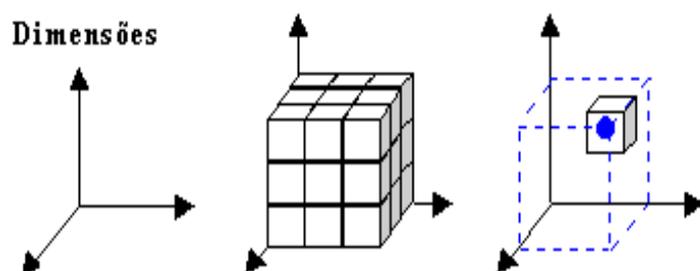


Figura 4.8: Cubo representativo da análise multidimensional. (Figura adaptada de [6]).

de maneira a definir o modelo pretendido sobre o qual se efectuará a análise. Clicando em “New Analysis”, aparecerão os *Schemas* e *Cubes* disponíveis, como se vê pela figura 4.9.

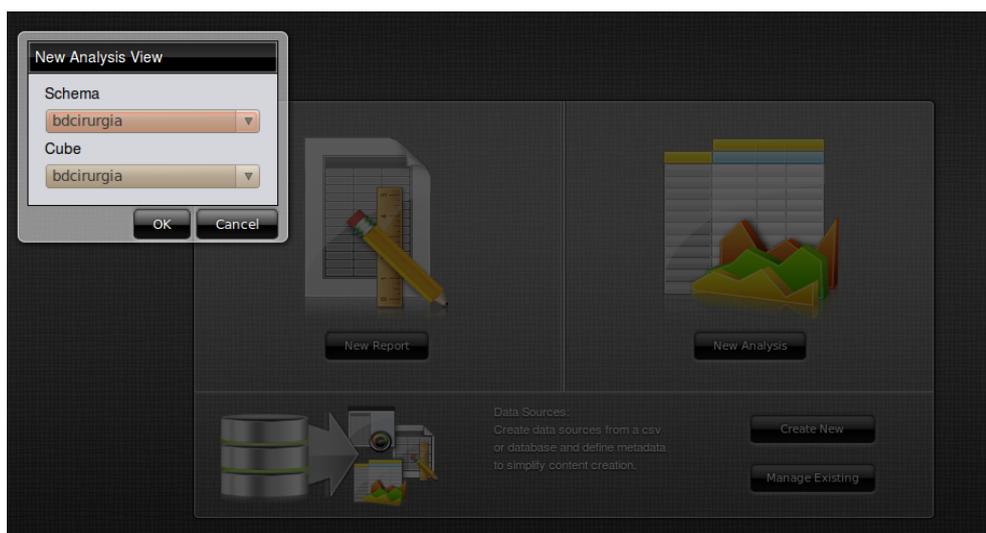


Figura 4.9: Opção *New Analysis View*.

Na figura 4.10, vê-se um exemplo de um hiper-cubo, com as respectivas medidas e dimensões para análise.

A barra de ferramentas do Analysis View tem o aspecto da figura 4.11.

O primeiro ícone corresponde ao "*OLAP Navigator*", onde se efectua a exploração do cubo. Aqui são exibidas todas as medidas e dimensões, permitindo seleccionar os items que se pretende visualizar, bem como a forma como são disponibilizados. Desta forma, definem-se que atributos são colunas e linhas e aplicam-se filtros, caso seja necessário. As *measures* têm de ser

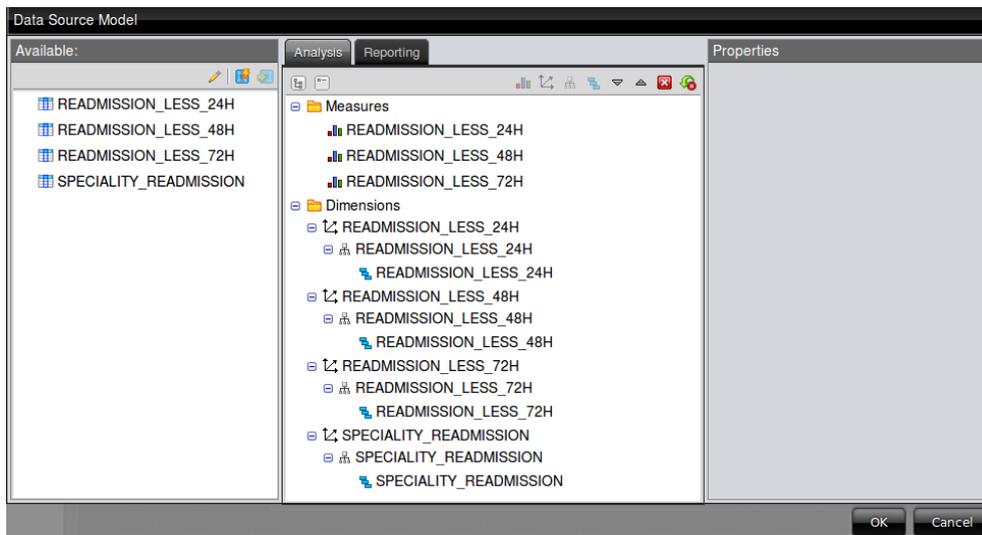


Figura 4.10: Exemplo de um hipercubo, com as dimensões e medidas definidas.



Figura 4.11: Barra de ferramentas disponibilizada no módulo.

seleccionadas também. Na figura 4.12, pode ver-se um exemplo de um cubo com as colunas e linhas definidas.

Com estas definições atribuídas, o *Data Grid* ou grelha de dados, através do qual se tem acesso à informação relativa às dimensões de forma detalhada, fica disponibilizada (figura 4.13).

O segundo ícone da barra de ferramentas direcciona o utilizador para o editor que permite escrever a consulta MDX – esta é uma linguagem semelhante ao SQL, mas projectada para fazer consultas ao cubo, em vez de ser feita à base de dados. O editor MDX mostra a consulta que está, no momento, a gerar o relatório Jpivot (figura 4.14).

Na figura 4.15 podem ver-se as opções oferecidas assim que se selecciona o terceiro ícone da barra de ferramentas: "*Config OLAP Table*". Encontram-se disponíveis diversas opções de ordenação: *Sort Mode*, onde se define a ordem de classificação, se ascendente ou descendente, juntamente com o tipo de exibição (coluna única – *keep*, ou coluna múltipla – *break*); *Number of rows for ranking* que estabelece o número de linhas que se pretende visualizar; e *Show Memberproperties* que permite mostrar ou ocultar todas as propriedades de

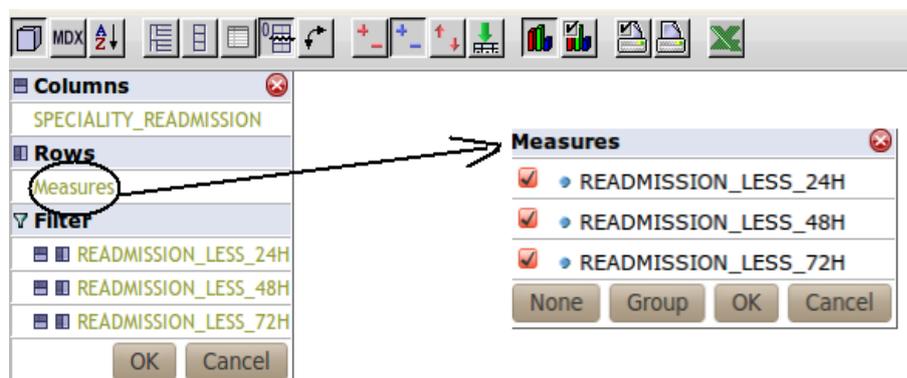


Figura 4.12: OLAP Navigator.

		SPECIALITY_READMISSION					
Measures	<input type="checkbox"/> All SPECIALITY_READMISSIONs	30301	30306	30400	30900	31500	32600
READMISSION_LESS_24H		4			1		3
READMISSION_LESS_48H		8		1	1	1	5
READMISSION_LESS_72H		17	1	1	1	1	12

Figura 4.13: Grelha de dados (*Data Grid*).

membros que foram definidas no modelo OLAP.

O ícone “*Show Parent Members*” permite que o utilizador escolha entre ter uma visão hierárquica dos dados ou uma visão plana. Basicamente, insere títulos de coluna acima dos dados quando se faz um *drill* (aprofundamento) na hierarquia.

A opção “*Hide Spans*” dá a possibilidade de alternar entre ter todos os descritores de linha visíveis ou não, isto é, impede a exibição de cabeçalhos repetidos quando múltiplas dimensões têm um acesso comum.

“*Show properties*” disponibiliza detalhes adicionais, caso estes estejam definidos no cubo. Esta opção está também disponível em *Show Member-properties* no ícone “*Config OLAP Table*”, como já referido.

“*Supress empty rows/columns*” permite ocultar dados nulos, o que se torna útil em casos de tabelas com elevado número de dados nulos.

“*Swap Axes*” possibilita a rotação em 90° dos eixos, isto é, há uma troca dos dados, as dimensões/membros da área de linhas passam a representar as dimensões/membros na área de colunas, e vice-versa.

Segue-se o “*Drill Options*” constituída por quatro opções: *drill member*, *drill position*, *drill replace* e *drill through*. Estas funcionalidades controlam o comportamento da query quando um utilizador clica num determinado membro para aprofundá-lo. Através desta opção, o utilizador escolhe o que pretende analisar e pode navegar por entre os níveis, do geral ao particular,

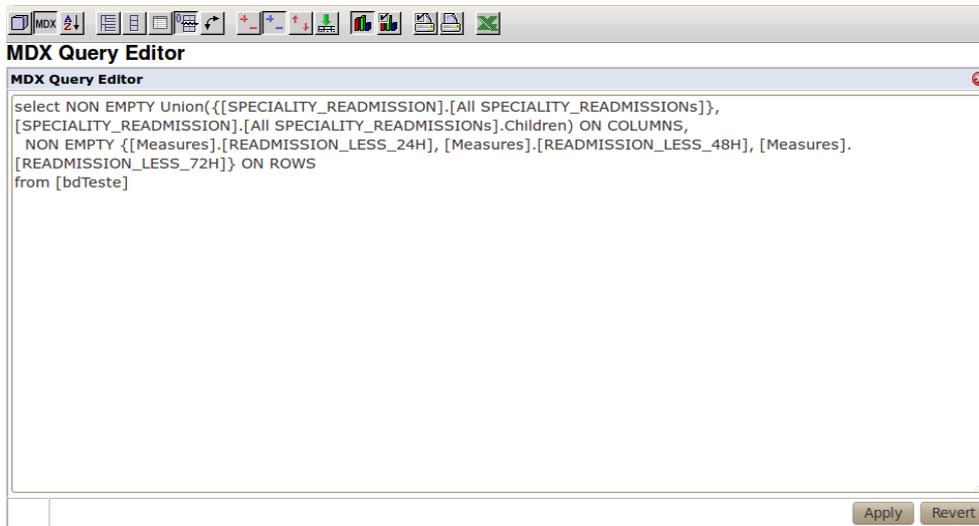
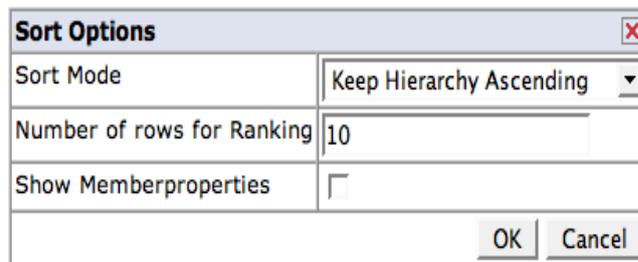


Figura 4.14: Editor MDX.

Figura 4.15: *Sort Options*.

ou seja, com maior ou menor detalhe.

A primeira, *drill member*, permite aprofundar vários membros através de um simples clique; o segundo tem a mesma função, mas apenas aprofunda o membro onde se clica; através do *drill replace*, o membro escolhido é substituído pelos seus membros-filho; por último, o *drill through* disponibiliza a informação em bruto para que o utilizador tenha acesso a detalhes por trás dos dados que estão a ser apresentados. Por outras palavras, o botão *drill through* permite que o utilizador tenha acesso aos detalhes individuais que resultaram na célula de valor agregado.

Relativamente às configurações gráficas, existem duas opções: visualização do gráfico ("*Show Chart*") e acesso às propriedades do mesmo ("*Chart Config*"), como título, tamanho da letra, entre outros.

"*Setting print properties*" permite definir configurações de impressão, como o título do relatório, a orientação da página, tamanho do papel, se se im-

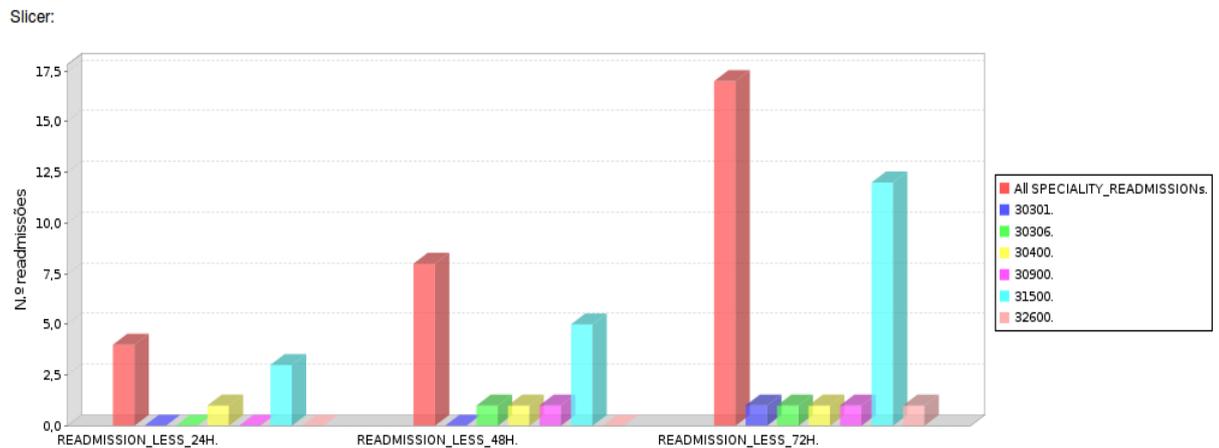


Figura 4.16: Exemplo de um gráfico com as configurações e propriedades ajustadas de acordo com o pretendido.

prime o gráfico numa página em separado ou não, etc.; "*Print as PDF*" gera uma versão em PDF do Analysis View com base nas configurações de impressão estabelecidas. Por último, "*Export to Excel*" inicia o Microsoft Excel e disponibiliza uma versão em Excel da análise criada.

4.2.4 Community Dashboard Edition (CDE)

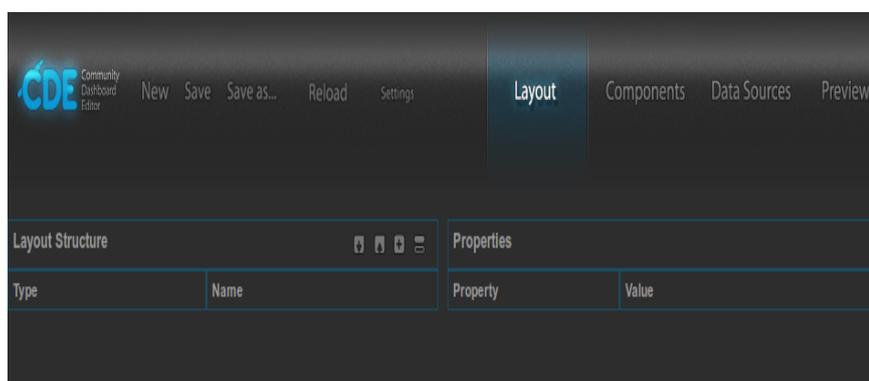
O **CDE** é um plugin utilizado para a criação de dashboards no servidor Pentaho. Este módulo permite criar e definir o layout do dashboard, especificando o conjunto de componentes associados a uma base de dados, por forma a ser possível a disponibilização desses componentes visuais no layout desejado. Assiste-se, assim, à flexibilização e parametrização dos componentes, uma vez que é possível a interação entre eles. O **CDE** possibilita um processo de desenvolvimento ágil, pois permite avançar e recuar durante as diferentes etapas de criação do dashboard.

Encontra-se dividido em *Layout*, *Components*, *Data Sources* e *Preview*.

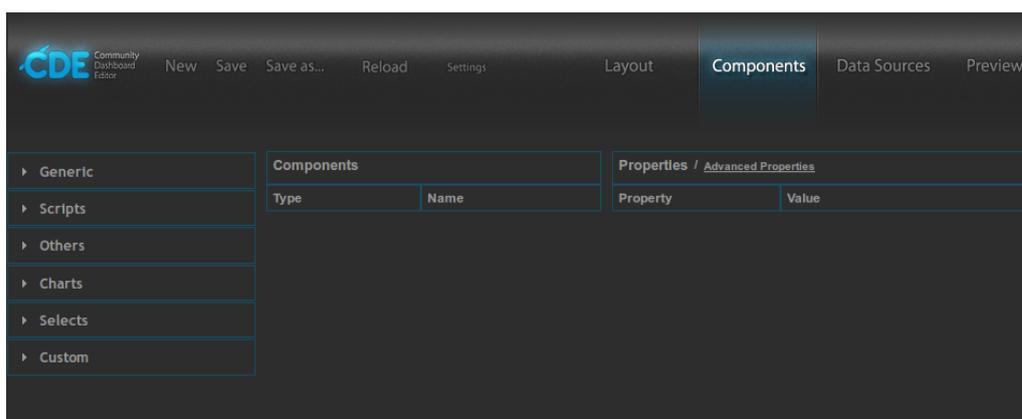
Layout

Aqui (figura 4.17) define-se a estrutura do layout, através de código HTML, tendo como grande vantagem a possibilidade de ser criada uma estrutura segundo a vontade do utilizador, podendo este incorporar funções em Javascript ou Css que permitem a formatação dos dashboards.

A aplicação de templates já pré-definidos também é possível, poupando tempo ao utilizador na formatação do layout desejado.

Figura 4.17: CDE: *Layout*.

Components

Figura 4.18: CDE: *Components*.

Nesta secção (figura 4.18) definem-se e personalizam-se os diferentes componentes (gráficos, tabelas, imagens, etc), sendo cada um deles atribuído a uma das regiões previamente definidas no *Layout*. Divide-se em diferentes categorias e sub-categorias que oferecem ao utilizador as mais diversas opções de disponibilização dos dados, nomeadamente:

- Generic: *Simple parameter*, *Custom parameter* e *Date parameter*;
- Scripts: *JavaScript Function*;
- Others: *PRPT Component*, *Pivot Component*, *Comments component*, *Navigation Menu component*, *Table Component*, *Execute Prpt Compo-*

nent, Traffic Component, Execute XAction Component, XAction Component, Freeform Component, Text Component, Related Content Component, Button Component, Query Component, Pivot link Component;

- Charts: *Protovis Component, CCC Dot Chart, Chart Component, CCC Pie Chart, TimePlot, OpenFlashChart Component, CCC Stacked Area Chart, CCC Stacked Line Chart, CCC Line Chart, Dial Chart Component, CCC Bar Chart;*
- Selects: *TextInput Component, Select Component, Date Range input Component, Check Component, Month Picker Component, RadioButton Component, Auto complete component, SelectMulti Component, Date input Component;*
- Custom: *Raphael Component.*

Data Sources

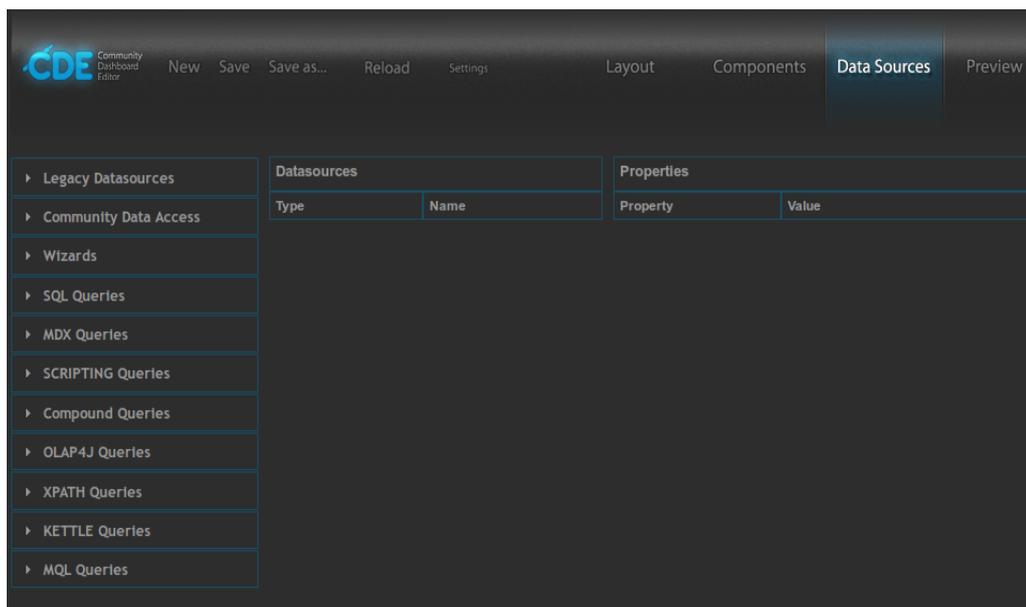


Figura 4.19: CDE: *Data Sources*.

Os componentes que se pretendem visualizar necessitam de estar associados a uma base de dados para que o seu conteúdo seja disponibilizado. Aqui (figura 4.19) são configuradas e editadas essas mesmas bases de dados, sendo atribuído um nome a cada uma delas pelos quais serão reconhecidas na secção “*Components*”. Podem ser:

- Legacy data sources: OLAP MDX query, SQL query, Xaction result set ou Kettle transformation;
- Community Data Access: CDA data source (inclui um ficheiro .cda no dashboard);
- Wizards: OLAP members wizard e OLAP Chart wizard;
- SQL queries: Sql over sqlJdbc (liga-se a uma base de dados através de Java DataBase Connectivity API - Jdbc), Sql over sqlJndi (associa-se a uma base de dados através de Java Naming e Directory Interface – Jdni);
- MDX queries: mdx over mondrianJndi, denormalizedMdx over mondrianJdbc, denormalizedMdx over mondrianJndi, mdx over mondrianJdbc;
- Scripting queries: Scriptable over scripting (actualmente a única linguagem de script que suporta é a BeanShell);
- Compound queries: Join (cria uma junção arbitrária de duas bases de dados, sendo por isso possível a junção de tabelas que residem em bases de dados diferentes), Union (concatena duas bases de dados que têm o mesmo formato de colunas);
- OLAP4J queries: Olap4J over olapJdbc, Oloap4J over olapJndi, denormalizedOlap4j over olapJdbc, denormalizedOlap4j over olapJndi;
- XPATH queries: Xpath over Xpath;
- KETTLE queries: Kettle over kettleTransFromFile;
- MQL Queries: Mql over metadata.

Preview

Permite a pré-visualização dos dashboards.

Porém, para que a pré-visualização seja possível é necessário o dashboard estar gravado (o Bi-server tem de receber os ficheiros do dashboard e posteriormente salvá-los para o disco) e sempre que seja feita uma alteração à base de dados, tem de se proceder ao *reload* do fiheiro .cda (*tools* → *reload* → *CDA cache*).

4.2.5 Pentaho Report-Designer (PRD)

Aplicação desktop que possibilita a criação de relatórios de uma forma rápida e sofisticada, além de prover suporte para tabelas e gráficos. Este módulo permite definir os campos a serem exibidos, bem como todas as informações que dizem respeito à formatação dos dados e aparência dos relatórios. A personalização dos relatórios de modo a adequá-los às necessidades do utilizador final é muito útil e vantajoso.

De realçar que esta ferramenta está voltada para utilizadores experientes, familiarizados com os conceitos e as fontes de dados utilizadas, uma vez que o PRD não é um editor *What You See Is What You Get* (WYSIWYG) e portanto tem de se desenvolver o relatório num ambiente que nada tem que ver com o layout final (figura 4.21). Consequentemente, isto requer algum conhecimento e compreensão da ferramenta por parte do utilizador.

Um relatório é constituído por dados fornecidos por uma base de dados e por uma definição de relatório que estabelecerá a forma como é visualizado pelo utilizador final. A página inicial concede ao utilizador a possibilidade de escolha entre iniciar a criação de um relatório de raiz ou através do Report Wizard (disponibiliza relatórios pré-definidos). A página inicial também dá acesso quer a exemplos de relatórios, quer a relatórios recentemente abertos, tal como se observa na seguinte figura:



Figura 4.20: Página inicial do PRD: *Welcome Page*.

Na figura 4.21, a grelha branca na área de trabalho constitui o local onde se contrói o relatório. Do lado esquerdo desta mesma área, encontram-se os elementos gráficos (figura 4.22) disponíveis para a construção do relatório,

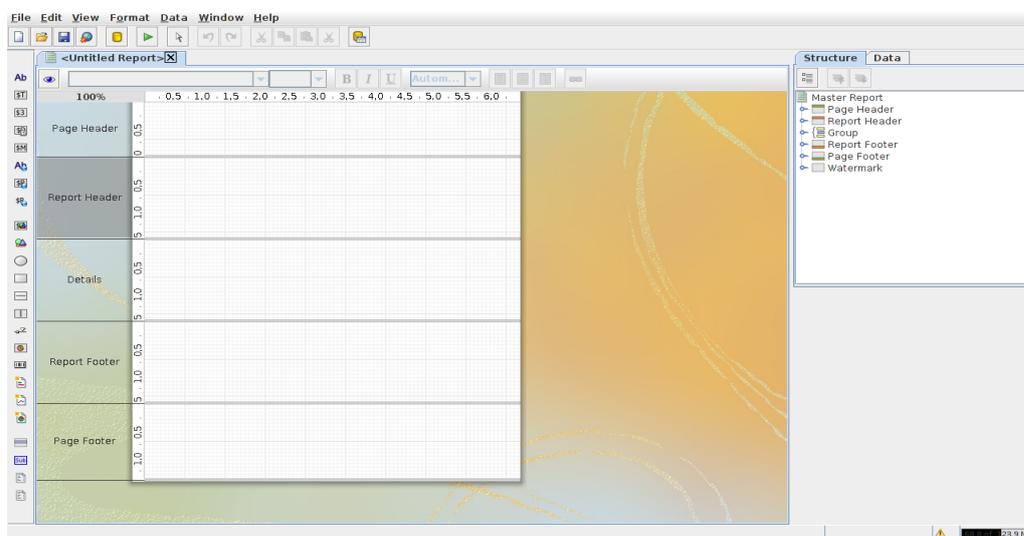


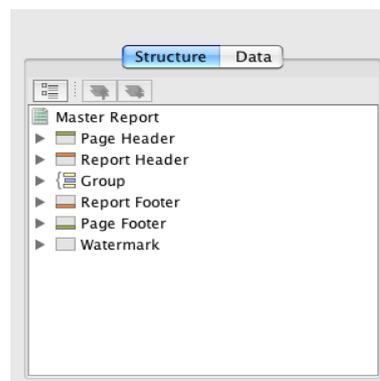
Figura 4.21: Área de trabalho do PRD.

desde áreas de texto a inserção de imagens, enquanto do lado direito o painel estrutural (figuras 4.23 e 4.24) disponibiliza a representação hierárquica do relatório. Aqui, a componente *Structure* fornece uma visão geral acerca dos componentes que constituem o design do relatório, enquanto na aba *Data* há uma visão centrada dos dados, onde se define a posição destes e a forma como são processados.

O painel de propriedades (figura 4.25) permite, tal como o nome sugere, o ajuste das propriedades da secção seleccionada, como tamanho de letra, fonte, alinhamento, etc. Para além disto, também é disponibilizada, na parte inferior, um painel de mensagens (figura 4.26) com informações úteis, dicas, avisos e erros contidos no relatório.



Figura 4.22: Elementos gráficos disponíveis.

Figura 4.23: Painel estrutural (*Structure*).

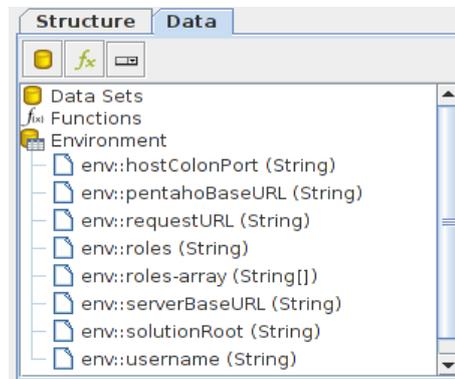
Figura 4.24: Painel estrutural (*Data*).

Figura 4.25: Painel de propriedades.

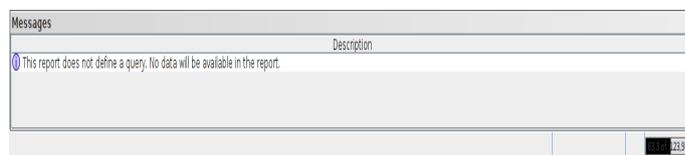


Figura 4.26: Painel de mensagens.

4.2.6 Pentaho WEKA Data Mining

Waikato Environment for Knowledge Analysis (WEKA), software desenvolvido em Java, pela Universidade de Waikato, Nova Zelândia, é uma colecção de algoritmos de aprendizado de máquina para realizar as tarefas de *data mining*. Em 2006, a *Pentaho Corporation* adquiriu uma licença exclusiva para usá-lo, sendo parte integrante do *Pentaho Business Intelligence Suite*. Os algoritmos podem ser aplicados directamente a um conjunto de dados ou chamados a partir de seu próprio código Java. WEKA contém ferramentas para pré-processamento de dados, classificação, regressão, agrupamento, regras de associação e visualização.

Pré-processamento

Esta é a primeira etapa de *data mining*. O software disponibiliza um conjunto de ferramentas (figura 4.27) para fazer o processamento dos dados antes de se aplicarem os algoritmos para EC: carregar dados; visualização dos atributos; obter informação acerca dos mesmos (tipo de atributo, valores, entre outros); remoção de atributos de modo a seleccionar os mais apropriados para a tomada de decisão; e aplicação de filtros que permitem transformar os dados de diferentes maneiras, estando divididos em supervisionados e não supervisionados e podem ser aplicados aos atributos ou às instâncias. Os filtros supervisionados consideram as classes, o que não acontece nos não supervisionados. Alguns dos filtros disponíveis incluem a discretização dos dados, a normalização, entre outros.

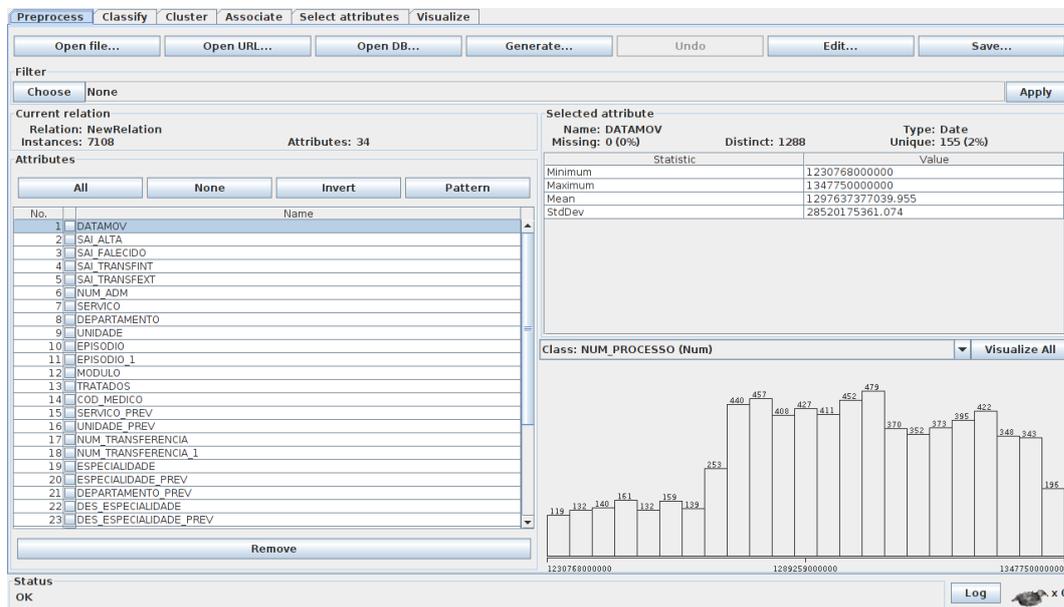


Figura 4.27: WEKA: Pré-processamento dos dados.

Classificação

A classificação é uma tarefa preditiva: define o valor de uma variável desconhecida (atributo classe) a partir de variáveis conhecidas, ou seja, é o processo de encontrar um conjunto de modelos (ou funções) que descrevem e distinguem classes de dados com o propósito de ser capaz de determinar a que classe de objectos pertence uma nova observação (classe desconhecida). Assim, permite determinar com que grupo de entidades, já classificadas anteriormente, essa nova observação apresenta mais semelhança.

O painel de classificação disponível no software permite a configuração, execução e aplicação ao *dataset* de qualquer um dos classificadores [WEKA](#), podendo-se optar por realizar uma validação cruzada ou um teste a um conjunto de dados. Os erros de classificação podem ser visualizados graficamente, e caso seja produzida uma árvore de decisão, esta pode ser visualizada através de uma ferramenta de visualização, que permite ter uma representação intuitiva, ajudando à interpretação.

Diversos classificadores podem ser aplicados, tais como redes neurais, métodos Bayesianos e árvores de decisão, entre outros.

Cluster

Esta ferramenta permite que o utilizador consiga agrupar os dados para determinar padrões, através da criação de um número específico de grupos,

dependendo das suas necessidades. Os algoritmos assumem sempre que não se conhece a que classe pertence a colecção de dados. Assim, e de uma forma muito rápida, todo o conjunto de dados é dividido em grupos, cujas entidades que lhes pertencem apresentam semelhanças entre si, havendo uma organização dos dados em grupos por forma a existir alta similaridade intra-grupo e baixa similaridade inter-grupos. A partir dos clusters, podem retirar-se conclusões e assim, tomar decisões com base nesses grupos.

A desvantagem, porém, remete para o facto do utilizador ter de saber à partida quantos clusters pretende obter, ou seja, exige que se tenha conhecimento real dos dados.

Associação

A associação permite gerar as regras que denotam padrões relevantes presentes no conjunto de dados. Estas regras são constituídas por antecedentes (subconjunto de atributos e os seus valores) e consequentes (subconjunto de valores que provêm do precedente). É definida uma métrica de similaridade, segundo a qual os dados são agrupados e que permite encontrar relações entre as instâncias. Esta medida determina a validade da regra obtida.

As regras apresentam um valor após o precedente que corresponde ao número de instâncias que a regra abrange, e um valor logo após o consequente que determina o número de instâncias correctamente previstas. Por último, é disponibilizado o valor correspondente ao valor da confiança da regra.

Seleccção atributos

Este painel permite aplicar métodos de pesquisa e avaliação para seleccionar os atributos mais pertinentes do *dataset*. Se um sistema de seleccção transformar os dados, estes podem ser visualizados.

Visualização

Este painel exhibe gráficos de dispersão para o conjunto de dados. O tamanho das células individuais e o tamanho dos pontos exibidos podem ser ajustados utilizando os controlos disponíveis. A visualização dos dados é feita a uma e duas dimensões.

Capítulo 5

Resultados

Neste capítulo são apresentados os resultados obtidos através da utilização das diferentes ferramentas do Pentaho para tratar, analisar e extrair conhecimento dos dados utilizados. Inicialmente, é feita uma referência aos dados utilizados para tal.

São seis os casos de estudos: na secção 5.1 estudam-se os movimentos registados na UCI ao longo dos anos; segue-se um caso em que se avaliam os dias de internamento dos pacientes (5.2); posteriormente, na secção 5.3, é feita uma análise semestral aos movimentos diários, por forma a detectar casos de sobrelotação; segue-se a secção 5.4, onde se estudam os casos de reinternamentos registados na UCI do hospital, dando-se especial atenção aos reinternamentos ocorridos em horas críticas (5.4). Na secção 5.5 é feito um estudo das cirurgias a que os pacientes internados na UCI foram submetidos e, por último, na secção 5.6 são apresentados os resultados obtidos pela execução das etapas de *data mining*, através do WEKA.

Em cada caso de estudo, após serem disponibilizados os resultados, é explicado o processo de desenvolvimento de cada um deles (5.1, 5.2, 5.3, 5.4, 5.5), sendo apresentados os passos para a obtenção dos mesmos, bem como as vantagens e inconvenientes das ferramentas encontradas aquando da sua utilização. Deste modo, é feita uma análise do ponto de vista informático, tal como estabelecido nos objectivos deste projecto.

Dados experimentais

Os dados utilizados para o desenvolvimento dos casos de estudo correspondem a dados clínicos reais de pacientes provenientes de uma base de dados de um hospital do Norte de Portugal.

A manipulação da base de dados foi realizada usando o Oracle SQL Developer (ambiente de desenvolvimento integrado gratuito).

A base de dados em questão contém vários tipos de dados, distribuídos por diversas tabelas. Para a monitorização dos dados na UCI, o processo de EC baseou-se essencialmente na manipulação de consultas SQL. Foram utilizadas tabelas independentes, cuja informação foi cruzada para se atingir o conhecimento útil. Ao longo do projecto, foram sendo acrescentadas tabelas à base de dados, devido à criação de novas tabelas provenientes dos processos executados no PDI.

Os dados em questão remetem para informação apenas dos pacientes internados na UCI, cujo código de especialidade é 30500, datas de movimentos, cirurgias a que foram submetidos, episódio, código do médico responsável pelo paciente, entre outros, utilizados de acordo com o que se pretendia.

5.1 Movimentos registados na UCI

Seja em que unidade hospitalar for, é sempre importante analisar os números inerentes aos movimentos dos pacientes. Na UCI concretamente, o número de óbitos e de altas, por exemplo, poderá ser um factor de qualidade dos cuidados prestados. É também importante perceber se há registo anormal de determinado movimento em determinado mês e, desta forma, tentar perceber as causas para tal.

Assim, neste caso de estudo fez-se uma análise ao tipo de movimentos ocorridos na UCI, desde 2009 a 2012: do número de admissões, do número de altas e óbitos, bem como das transferências internas e externas ocorridas nesta unidade, quer mensalmente, quer anualmente.

Começando pela análise global dos movimentos ao longo dos quatro anos, pelo gráfico de linhas da figura 5.1, conclui-se que em termos numéricos, as transferências internas estiveram sempre acima dos restantes movimentos, havendo um aumento em 2010 e 2011, talvez por causa de mais pacientes terem sido admitidos. Deve realçar-se, porém, que nesses mesmos anos, todos os movimentos, desde altas a óbitos, registaram aumentos. Conclui-se também que sempre houve mais óbitos do que altas, depreendendo-se que esta situação está inerente ao facto de a UCI ser uma unidade com situações bastante críticas, onde os pacientes têm sempre prognósticos reservados que infere uma maior probabilidade de morte.

A figura 5.2 corrobora já as conclusões tiradas, onde fazendo uma comparação entre movimentos, ano a ano, verifica-se que, de facto, as transferências internas (porção azul nos gráficos) representam grande percentagem dos movimentos registados na UCI: 81% em 2009, 90% em 2010, e cerca de

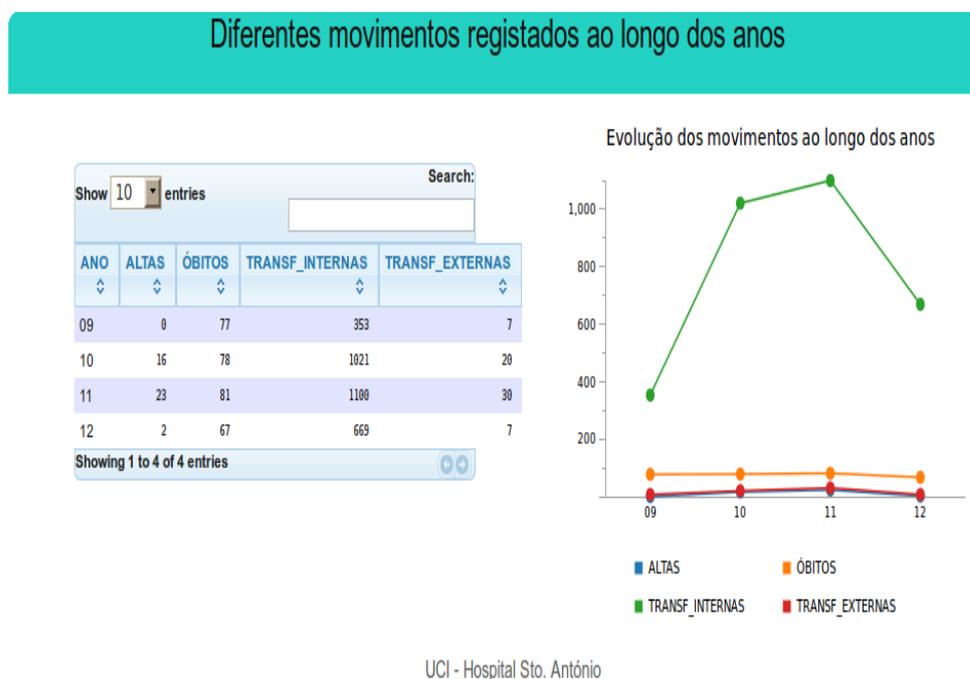


Figura 5.1: Diferentes tipos de movimento registados na UCI e a sua evolução ao longo dos anos.

98% e 90%, em 2011 e 2012, respectivamente.

Numa unidade como a UCI é normal os pacientes permanecerem nesta unidade durante algum tempo e, mais tarde, quando se encontram mais estáveis, serem transferidos para outra unidade, ou seja, é um período transitório no internamento de um doente, como já referido no capítulo 2.

Fazendo agora uma análise mensal ao longo dos anos, e relativamente às admissões efectuadas, observa-se, pela figura 5.3, que em 2009 os meses em que se registou maior número de pacientes admitidos foi em Maio e Julho.

Em 2010, destaca-se o aumento acentuado registado de Março para Abril, registando-se a partir daí admissões nunca abaixo dos 88 pacientes. Isto deve-se ao facto de ter ocorrido uma reestruturação na UCI, onde a lotação máxima passou de 10 para 24 camas o que, conseqüentemente, levou ao aumento de pacientes admitidos.

Nos meses de Março a Junho de 2011 registaram-se admissões acima dos 100 pacientes, enquanto nos restantes meses a média de admissões foi de cerca de 90 pacientes.

Por último, relativamente a 2012 e tendo em conta os registos efectuados

até Setembro, verifica-se uma diminuição do número de pacientes admitidos na unidade ao longo dos meses, isto por causa de nova reestruturação da unidade onde, a partir de Julho, a lotação máxima passou a ser de 12 camas.

Globalmente, não se pode extrair nenhuma tendência nas admissões efectuadas ao longo dos anos.

Tendo em conta as altas, observando a figura 5.4, verifica-se que no ano de 2009 não foram registadas nenhuma; em 2010, houve um total de 16 altas, não registadas em meses consecutivos, sendo Outubro o mês que mais altas teve; no ano seguinte, apenas os meses de Janeiro e Dezembro não tiveram ocorrência de altas. Por último, em 2012 e nos meses de Março e Maio, registaram-se 1 alta em cada. Mais uma vez, e como seria de esperar, não se verifica um padrão comportamental ao longo dos meses.

Analisando agora os óbitos ocorridos mensalmente de 2009 a 2012 (figura 5.5), conclui-se que todos os meses houve registo de falecimentos na UCI. Em 2009 o mês com maior número de óbitos ocorreu em Fevereiro. Em Abril e Maio de 2010, registaram-se 9 falecimentos, sendo esse o máximo nesse mesmo ano. No ano seguinte, em Maio registaram-se 11 mortes, mas deve ter-se em conta que este também foi o mês com maior número de pacientes admitidos, tal como já referido acima (figura 5.3). Em 2012, enquanto em Janeiro apenas se registou 1 óbito, no mês seguinte houve 17 pacientes falecidos. Este aumento foi único, pois nos meses seguintes registou-se novo decréscimo.

Na figura 5.6 pode ver-se o número de transferências internas ocorridas nos quatro anos em questão. Os pacientes são transferidos da UCI para outras especialidades hospitalares.

Por fim, na figura seguinte observam-se as transferências externas ocorridas desde 2009 a 2010, depreendendo-se que nos quatro anos tidos em conta, nunca ocorreram mais de 4 transferências externas no mesmo mês.

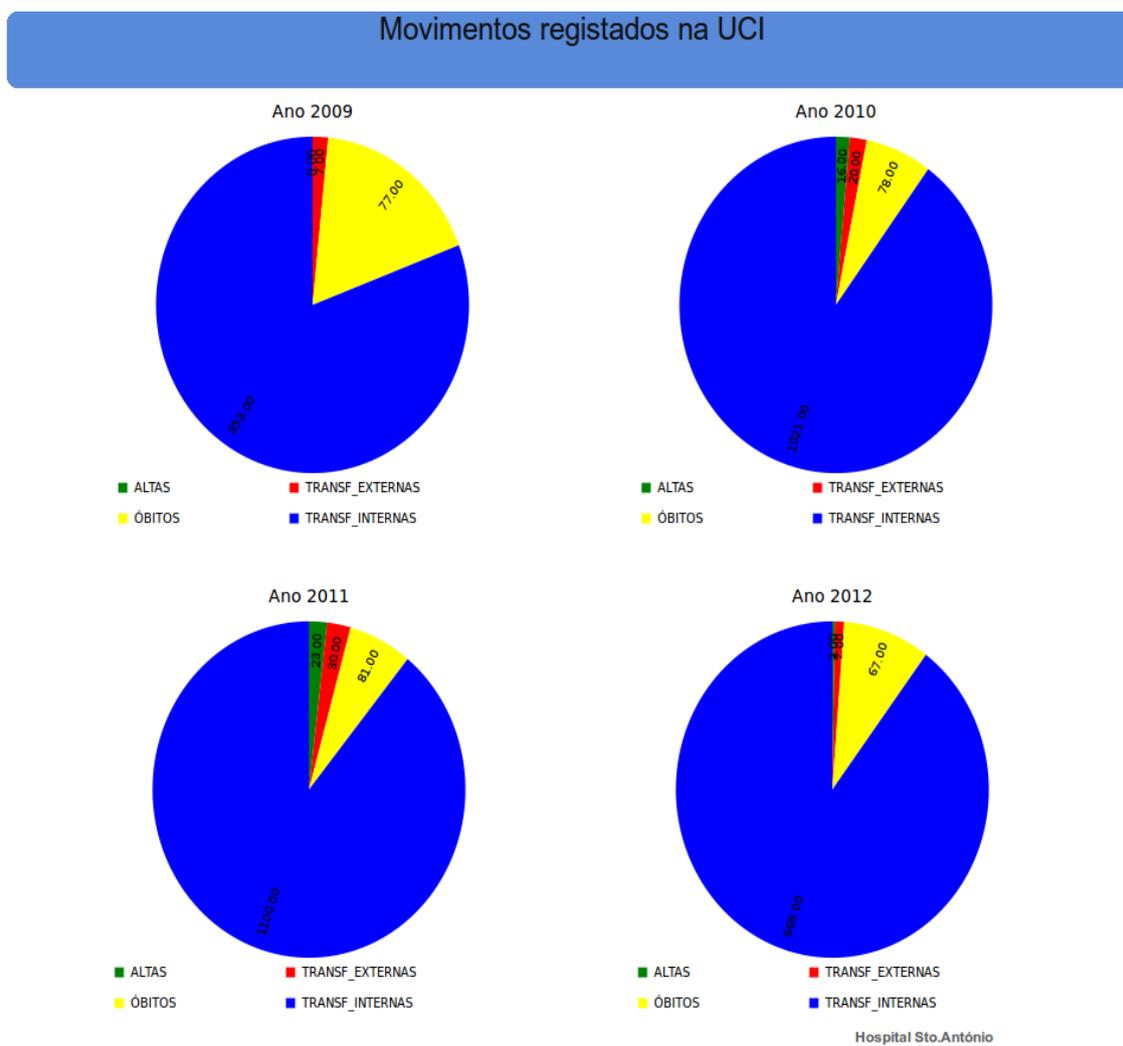


Figura 5.2: Comparação entre movimentos registados na UCI, ano a ano.

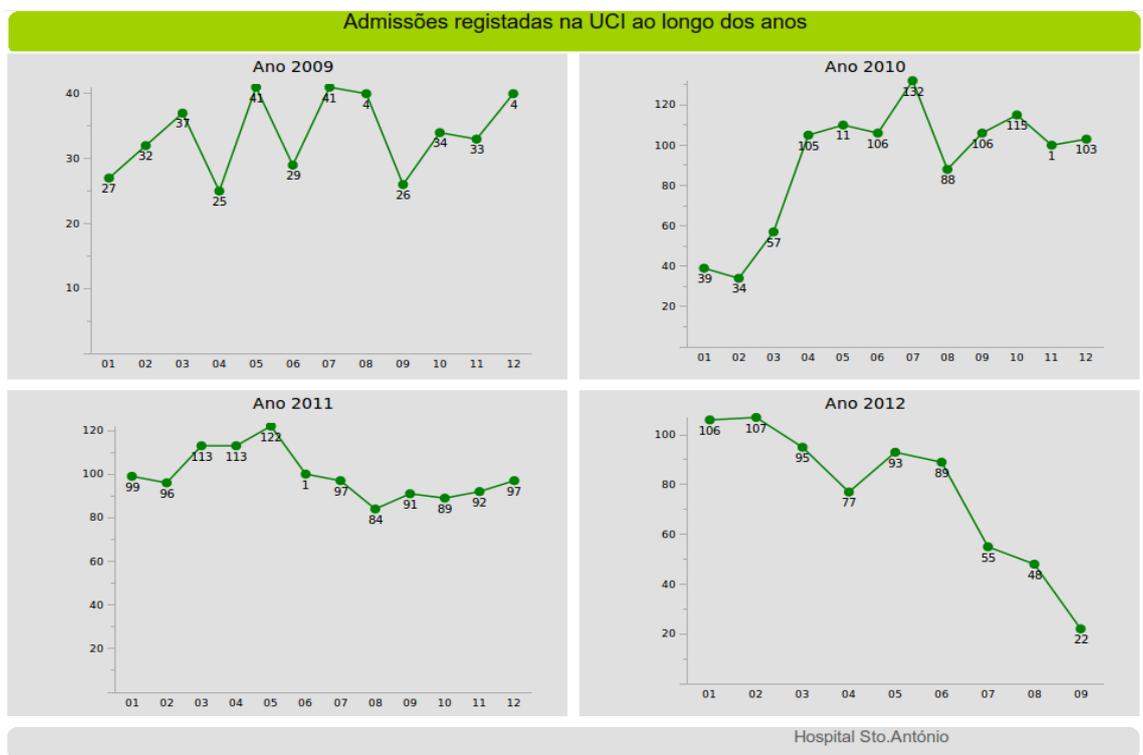


Figura 5.3: Registo das admissões mensais efectuadas na UCI de 2009 a 2012.

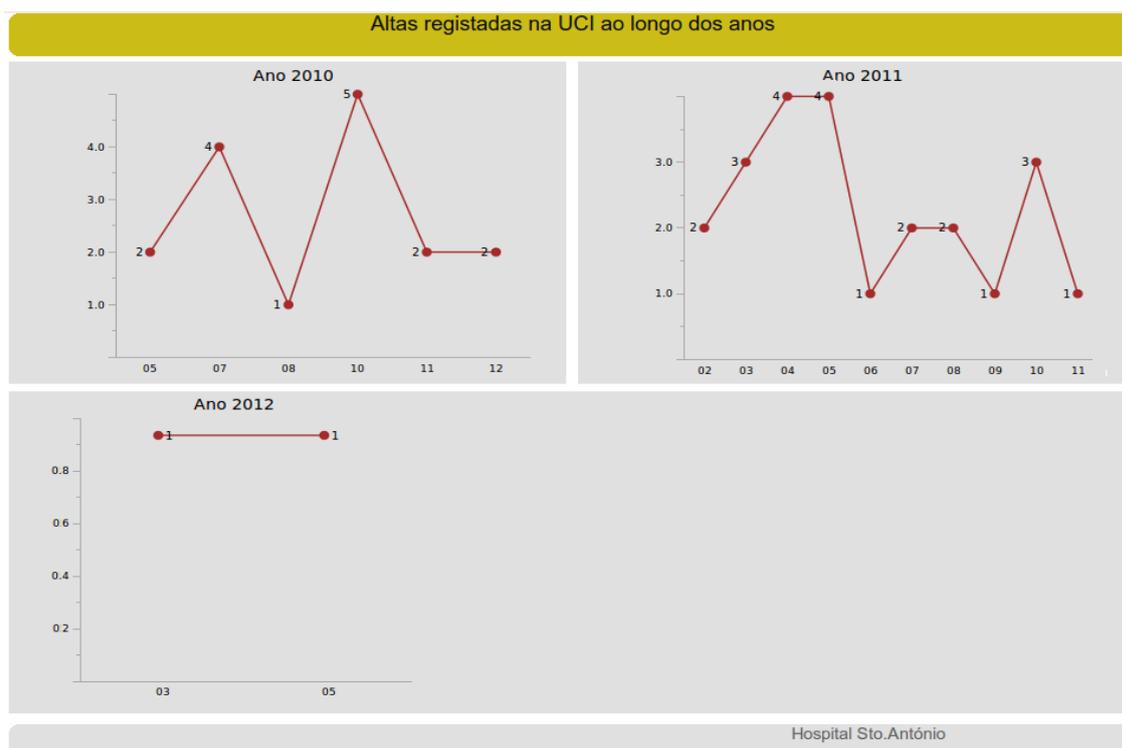


Figura 5.4: Altas registadas ao longo dos meses efectuadas na UCI de 2010 a 2012.

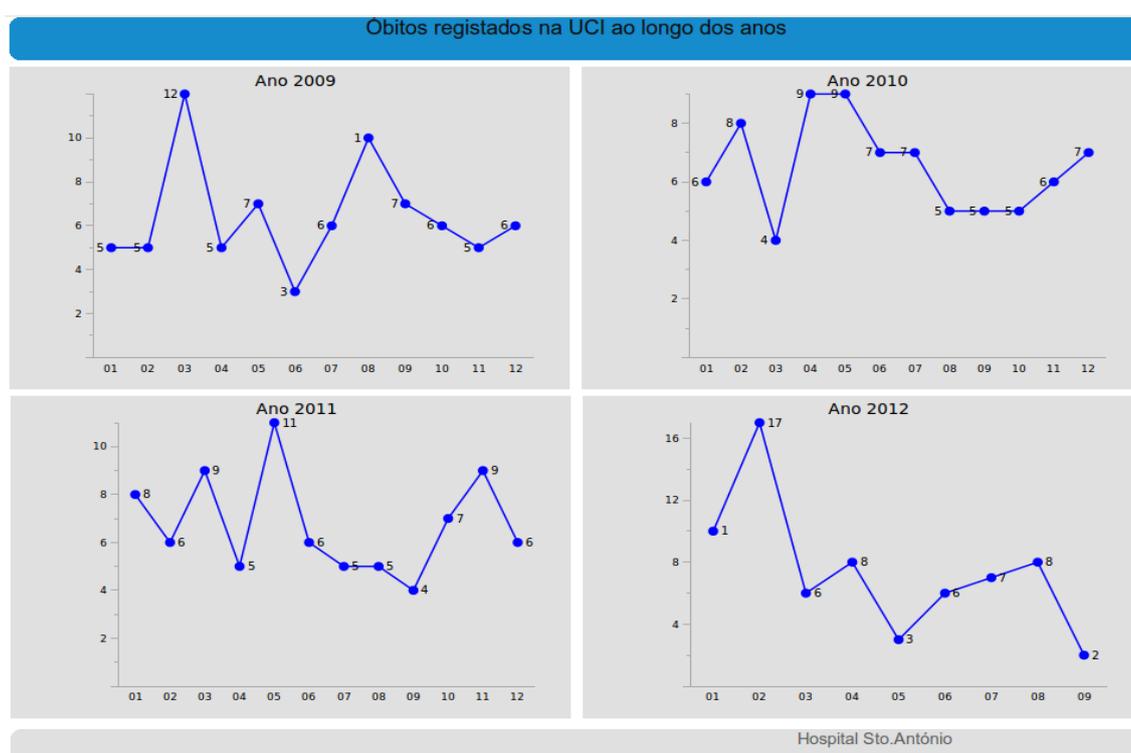


Figura 5.5: Óbitos registados ao longo dos meses na UCI de 2009 a 2012.

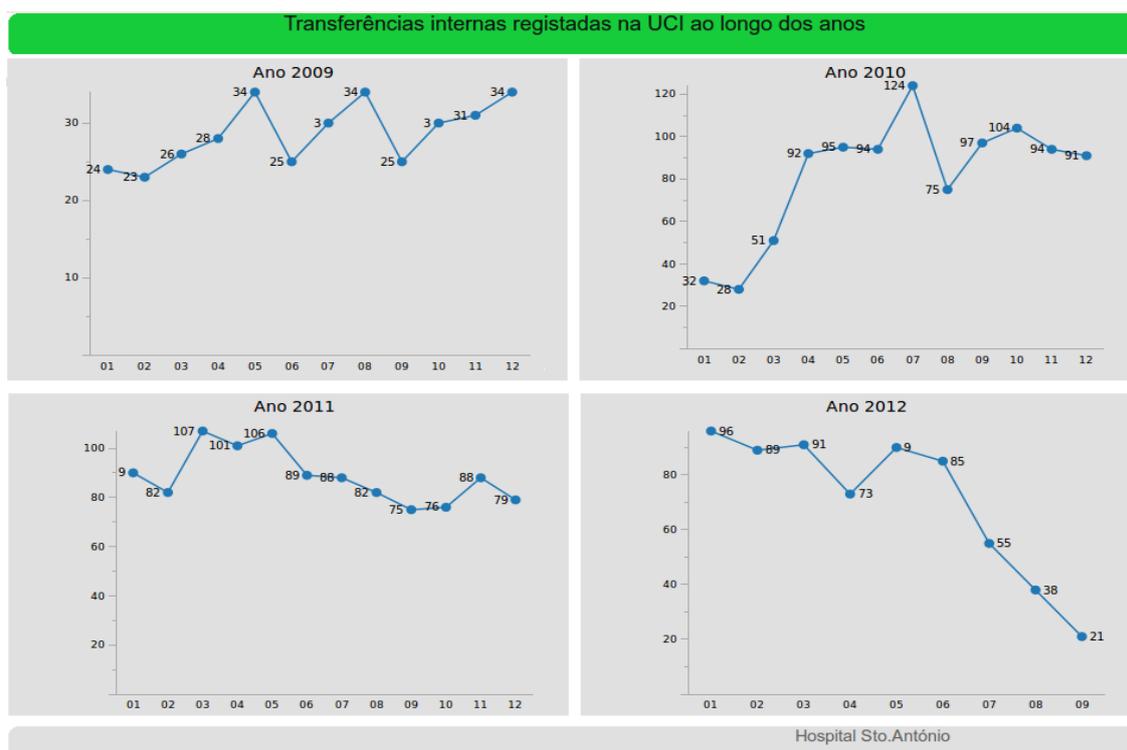


Figura 5.6: Número de pacientes transferidos da UCI para outras unidades hospitalares, desde 2009 a 2012.

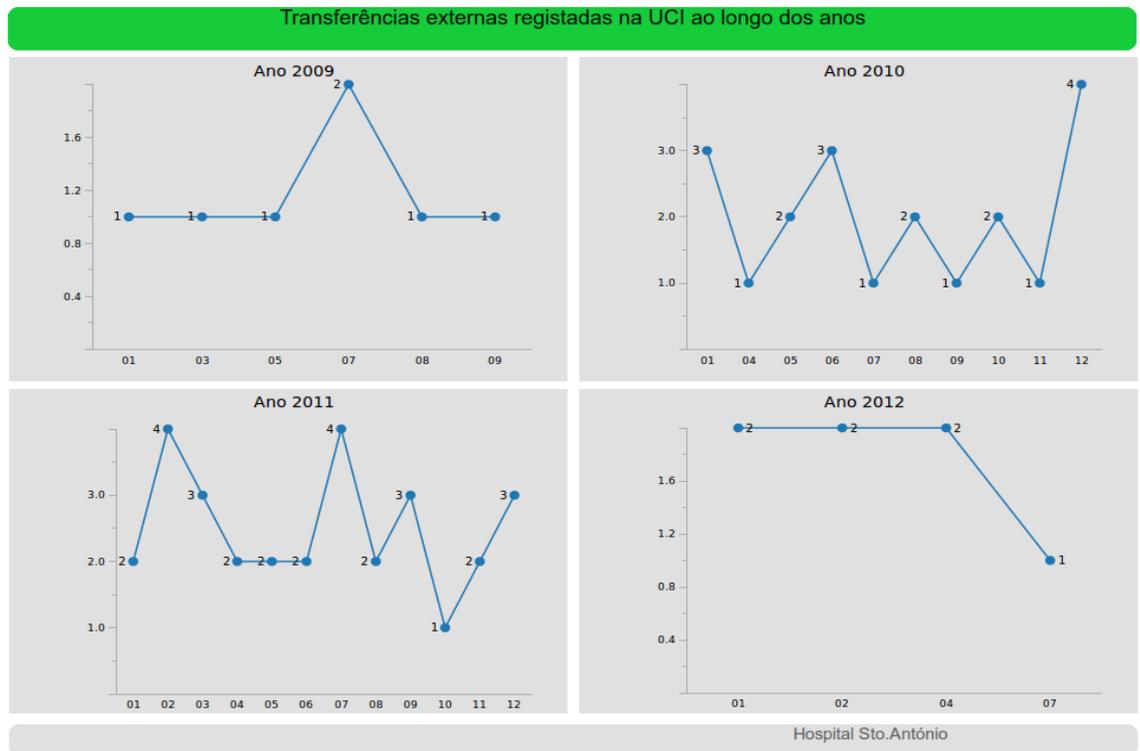


Figura 5.7: Número de pacientes transferidos da UCI para outros hospitais, desde 2009 a 2012.

Processo de desenvolvimento

Para este caso de estudo, uma vez que os dados já estavam armazenados no DW, sem haver necessidade de efectuar qualquer etapa de pré-processamento, apenas se fizeram consultas SQL ao DW para aceder aos resultados disponibilizados.

Assim, utilizou-se o módulo CDE que fornece ferramentas de visualização para disponibilizar os dados de forma atractiva e livre. Todas as figuras disponibilizadas neste caso de estudo são dashboards.

Layout

Tal como já previamente referido, o módulo CDE permite a aplicação de templates já criados em código HTML, que podem ser modificados segundo a vontade do utilizador, permitindo maior rapidez na criação e disponibilização da informação. Deste modo, para o primeiro dashboard criado, utilizou-se o “*Two Columns Template*” e, para os restantes, fez-se uso do “*Four Columns Template*”, que disponibilizam 2 e 4 painéis, respectivamente.

As propriedades do layout foram alteradas conforme necessário e pretendido, desde largura, altura, tamanhos de letra e cores, entre outros.

Components

O primeiro dashboard contém uma tabela e um gráfico de linhas como componentes, o segundo dashboard é constituído por gráficos circulares, sendo todos os outros compostos por gráficos de linhas.

Data Sources

Foram seleccionadas SQL Queries - sql over sqlJDBC, sendo o nome da base de dados *oracle.jdbc.driver.OracleDriver* e o URL *jdbc:oracle:thin:@gia-oracle.di.uminho.pt:1521:olia*.

Cada uma das bases de dados configurada apresenta um nome específico ao qual foi associado o respectivo componente.

Cada um dos gráficos está associado a uma query. Por exemplo, para o dashboard relativo à evolução das altas, no ano de 2011, a query definida foi:

```
select to_char(datamov, 'mm') as mes, count (sai_alta) as altas
from internados
where sai_alta = '1' and to_char(datamov, 'YY')='11'
group by to_char(datamov, 'mm')
```

order by mes

Visualização

Durante a realização deste caso de estudo foram encontrados dois problemas a ter em conta e explicados seguidamente.

Há determinados valores dos gráficos que não aparecem correctamente, nomeadamente nos dashboards relativos às análises mensais dos movimentos. Veja-se, por exemplo, a figura 5.3, onde o número de admissões em Junho de 2011 registado é de 100 pacientes, mas apenas é disponibilizado um '1', o que poderá induzir o utilizador em erro. A origem do bug não foi detectada e, por isso, não foi possível a sua correcção, sendo que a solução passa por escolher a opção de não disponibilizar os valores, mas isso implicaria uma percepção pouco clara para o utilizador acerca dos valores correspondentes.

Outro factor desvantajoso foi encontrado aquando da escolha de gráficos circulares (*CCC Pie Chart*) como componentes. Na figura 5.2, foram utilizados 4 gráficos circulares relativos a cada ano, sendo atribuída uma query SQL a cada um deles. Para o ano 2009:

```
select sum(num_adm), sum(sai_alta) as Altas,
sum(sai_transfext) as Transf_externas, sum(sai_falecido) as Óbitos,
sum(sai_transfint) as Transf_internas
from internados
where to_char(datamov, 'yy')='09'
```

O primeiro atributo *sum(num_adm)* não deveria estar na query, mas teve de se fazer uso dele como forma de solucionar o problema de o primeiro atributo escrito na query não ser disponibilizado graficamente. Teve então de proceder-se à selecção de um outro atributo ao calhas, para que *sum(sai_alta) as Altas* aparecesse no gráfico. O mesmo problema ocorreu para os restantes gráficos circulares utilizados e o mesmo raciocínio teve de ser seguido.

Outro tipo de gráfico elaborado para fornecer a mesma informação que o gráfico da figura 5.3 é o que se visualiza na figura 5.8. No dashboard criado fez-se uso de um filtro que o utilizador pode utilizar para, interactivamente, ter acesso às admissões mensais ocorridas nos diferentes anos. Esse filtro oferece a opção de escolha relativamente ao ano em que se pretende ver as admissões ocorridas. Na figura seguinte está disponibilizada a informação de 2009.

Este tipo de gráfico é vantajoso na medida em que permite que o utilizador

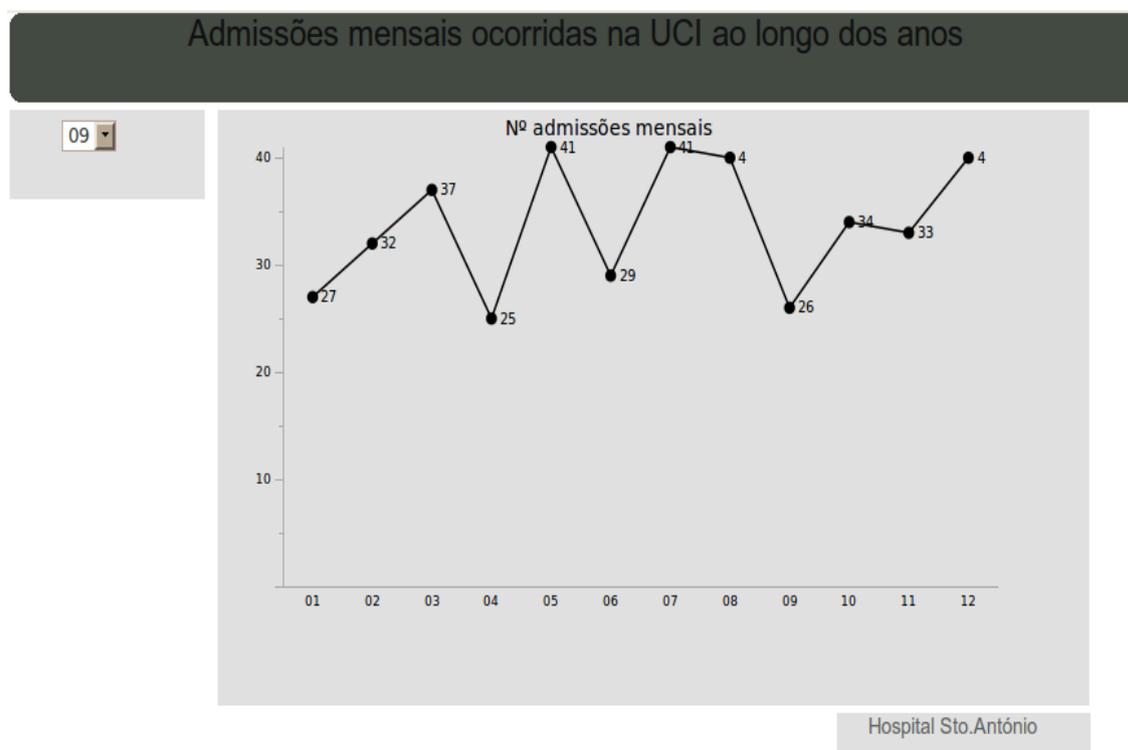


Figura 5.8: Dashboard com filtro: registo das admissões mensais efectuadas na UCI de 2009 a 2012.

se foque na informação que pretende e que tenha melhor visualização dos gráficos, visto que aparecem um a um.

Para a elaboração deste, o layout definido foi o “*Filter Template*” e fez-se uso de três componentes: *Simple Parameter*, *Select Component* e de um *CCC Line Chart*. Os dois primeiros referem-se ao filtro dos anos, enquanto o gráfico de barras é usado para disponibilização dos dados. Teve de se associar os diferentes componentes para que a informação disponibilizada estivesse em concordância para os diferentes anos.

5.2 Dias de internamento

Segue-se um relatório com informação estatística acerca dos dias de internamento dos pacientes, desde a sua admissão até ocorrer a sua transferência para outra unidade hospitalar. Atendeu-se aos três últimos anos, pois eram aqueles cujos dados disponíveis estavam mais completos e possuíam mais registos. Note-se que o número total de pacientes a ter em conta não é o

mesmo que o da figura 5.1, pois esta análise teve de ter em conta a diferença de datas entre pacientes admitidos e pacientes transferidos internamente e promover total concordância entre registos, isto é, calcular a diferença de dias em relação ao mesmo internamento. Os pacientes que foram transferidos para outra unidade, mas cujo registo de admissão não estava disponível na base de dados, não foram contabilizados na amostra.

Verifica-se que o número máximo de dias de internamento registado ocorreu em 2010 e foi de 99 dias, ou seja, cerca de 3 meses. O estado dos pacientes admitidos numa UCI, por regra, é grave e crítico, e a sua transferência para outro departamento normalmente ocorre após estabilização do paciente, quando já não é necessário fazer-lhe uma monitorização contínua e intensiva ou fornecer um suporte terapêutico especializado. O número de dias de internamento de um paciente ganha destaque, uma vez que os custos hospitalares desta unidade são tanto maiores quanto mais elevado for o número de procedimentos terapêuticos utilizados para monitorização e tratamento dos pacientes. A média de dias de internamento até ocorrer transferência dos doentes encontra-se entre 3 a 4 dias, o que não se pode considerar um número muito elevado.

Hospital Sto. António UCI

Dias de internamento na UCI até ocorrer transferência interna dos pacientes

Ano	Total de pacientes	Média	Mínimo	Máximo
2010	1017	3	0	99
2011	1091	3	0	58
2012	654	4	0	64

Figura 5.9: Relatório elaborado através do PRD com informações estatísticas acerca dos dias ocorridos entre uma admissão e transferência interna, para os anos 2010, 2011 e 2012.

Nas UCIs, a avaliação das necessidades de cuidados tem sido de enorme importância quando se procura aliar qualidade da assistência, optimização de recursos e redução de custos. As transferências internas são um parâmetro a ponderar na avaliação do modelo de financiamento dos serviços hospitalares, pois alertam para a necessidade dos hospitais terem meios humanos e materiais adequados para o apoio ao doente transferido das UCI, ou seja, para prover cuidados pós-UCI.

O mesmo tipo de análise feita para as transferências externas revela que a média de internamento dos pacientes é superior comparativamente aos pacientes que são transferidos internamente. Porém, é de realçar que a amostra de pacientes é bastante menor, pois enquanto de 2010 a 2012, se tem em conta 2762 pacientes transferidos internamente, para o mesmo período, apenas 54 foram transferidos externamente.

Hospital de Sto. António UCI

Dias de internamento na UCI desde admissão até ocorrer transferência externa

Ano	Total de pacientes	Média	Mínimo	Máximo
2010	18	9	0	42
2011	30	9	0	41
2012	6	14	0	33

Figura 5.10: Relatório elaborado através do PRD com informações estatísticas acerca dos dias ocorridos entre uma admissão e transferência externa dos pacientes, para os anos 2010, 2011 e 2012.

Hospital de Sto. António UCI

Dias de internamento dos pacientes na UCI desde admissão até obterem alta

Ano	Total de pacientes	Média	Mínimo	Máximo
2010	16	2	1	16
2011	23	2	1	30
2012	2	7	5	9

Figura 5.11: Relatório elaborado através do PRD com informações estatísticas acerca dos dias ocorridos entre admissão até obtenção de alta, para os anos 2010, 2011 e 2012.

Observando o relatório disponibilizado na figura 5.11, percebe-se que em 2012 a média de dias de internamento dos pacientes sobe consideravelmente, mas isto apenas se deve ao facto de a amostra só ser composta por dois pacientes que obtiveram alta.

Por último, e pelo relatório da figura 5.12, constata-se que nos três anos analisados, dos 222 pacientes falecidos, a média dos três anos é de 11 dias de internamento na UCI. Em relação aos máximos registados, verifica-se que todos eles são superior a 1 mês de internamento.

Hospital de Sto. António UCI

Dias de internamento dos pacientes falecidos na UCI desde a sua admissão até ao óbito

Ano	Total de pacientes	Média	Mínimo	Máximo
2010	77	8	0	65
2011	79	10	0	75
2012	66	5	0	28

Figura 5.12: Relatório elaborado através do PRD com informações estatísticas acerca dos dias ocorridos entre as admissões e óbitos, para os anos 2010, 2011 e 2012.

Uma outra análise realizada remete para o número máximo de dias de internamento de cada paciente na UCI. Aqui deve ter-se em conta que a soma dos dias pode ser não consecutiva, ao contrário do que acontecia nos casos descritos acima. Por outras palavras, um paciente pode ter estado internado nesta unidade durante 2 dias e ter sido transferido, sendo reinternado posteriormente durante 5 dias, sendo o dias máximo de internamento aqui considerado de 7 dias.

Assim, na figura 5.13, observa-se que num total de 2882 pacientes, no período de aproximadamente 4 anos, o máximo de dias de ocupação registado foi de 122 dias, sendo o mínimo 1 dia.

Como complemento de visualização deste gráfico acima, criou-se um outro dashboard (figura 5.14) que disponibiliza o número de pacientes discriminado por dias de internamento, constatando-se que, de facto, a maioria dos pacientes apenas fica internado nesta unidade num período inferior a 10 dias mas, acima de tudo, verifica-se que em 37% dos casos (1070 pacientes), os internados mantêm-se só 1 dia na UCI.

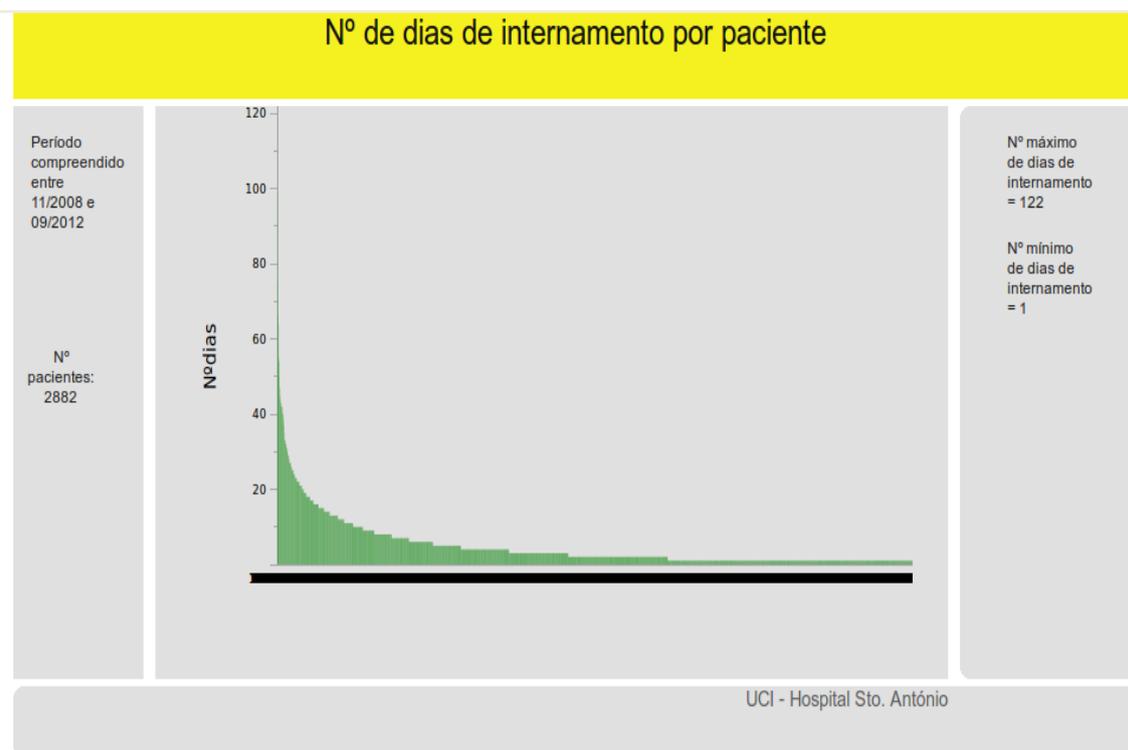


Figura 5.13: Número de dias de internamento na UCI por paciente.

Dias de internamento na UCI

Show <input type="text" value="25"/> entries		Search: <input type="text"/>	
DIAS_INTERNAMENTO	▲	NR_PACIENTES	↕
1		1070	
2		478	
3		260	
4		228	
5		119	
6		112	
7		76	
8		81	
9		51	
10		49	
11		41	
12		34	
13		39	
14		27	
15		25	
16		24	
17		15	
18		19	
19		10	
20		9	
21		11	
22		13	
23		6	
24		6	
25		7	

Showing 1 to 25 of 55 entries Hospital Sto. António

Figura 5.14: Número de pacientes por dias de internamento.

Processo de desenvolvimento

Para a análise estatística dos dias ocorridos entre a admissão de um paciente até a sua saída da UCI quer por ter tido alta, ter sido transferido para outra unidade hospitalar ou para outro hospital, ou ainda por ter falecido, foram utilizados dois módulos do Pentaho: o PDI e PRD.

Assim, foi necessário efectuar processos de ETL para extracção, transformação e carregamento dos dados desejados. A figura seguinte ilustra a transformação criada através do PDI.

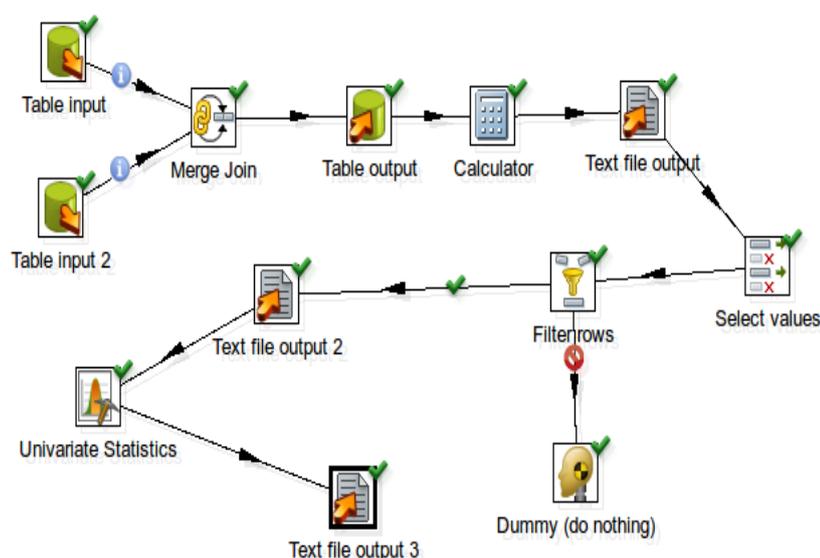


Figura 5.15: Transformação PDI criada para obtenção dos dias passados entre a admissão do paciente até a sua saída da UCI.

A mesma transformação foi utilizada para todos os casos, apenas alterando e configurando as consultas SQL conforme desejado. Será aqui explicado em detalhe o processo para o caso das transferências internas, sendo que o mesmo raciocínio foi seguido para os restantes.

Assim, para criação do DW, começou por se extrair informação da base de dados referente aos pacientes admitidos e aos pacientes que foram transferidos para outra unidade hospitalar. Para tal, utilizaram-se duas tabelas de entrada, com as seguintes consultas SQL:

```
select distinct episodio, cod_medico, datamov as TRANSF_INT
from internados
where sai_alta='0' and sai_falecido='0' and sai_transfint='1'
```

and sai_transfext='0' and to_char(datamov, 'yy')='10'
order by episodio

select distinct episodio, cod_medico, datamov as ADMITIDOS
from internados
where sai_alta='0' and sai_falecido='0' and
sai_transfint='0' and sai_transfext='0' and to_char(datamov, 'yy')='10'
order by episodio

Para o mesmo episódio, pode haver ocorrência de mais do que uma transferência interna, em diferentes períodos. Veja-se o exemplo do paciente com episódio=10004889, sujeito a 5 transferências internas:

	DATAMOV	SAI_ALTA	SAI_FALECIDO	SAI_TRANSFINT	SAI_TRANSFEXT	NUM_ADM	SERVICO	DEPARTAMENTO	UNIDADE	EPISODIO
1	10.09.10	0	0	1	0	0 (null)	(null)	(null)	(null)	10004889
2	10.09.02	0	0	0	0	1 (null)	(null)	(null)	(null)	10004889
3	10.08.09	0	0	1	0	0 (null)	(null)	(null)	(null)	10004889
4	10.07.07	0	0	0	0	1 (null)	(null)	(null)	(null)	10004889
5	10.07.07	0	0	1	0	0 (null)	(null)	(null)	(null)	10004889
6	10.04.08	0	0	1	0	0 (null)	(null)	(null)	(null)	10004889
7	10.03.27	0	0	0	0	1 (null)	(null)	(null)	(null)	10004889
8	10.03.13	0	0	1	0	0 (null)	(null)	(null)	(null)	10004889
9	10.05.23	0	0	0	0	1 (null)	(null)	(null)	(null)	10004889
10	10.02.17	0	0	0	0	1 (null)	(null)	(null)	(null)	10004889

Figura 5.16: Resultado de uma consulta SQL feita à base de dados, relativa ao episódio=10004889.

Ora, a forma encontrada para garantir que os dias de internamento calculados se referiam a períodos consecutivos, foi fazer o “*Merge Join*” quer do episódio, quer do código do médico, das duas tabelas de entrada. Tomando como exemplo ainda o episódio 10004889, considerando o mesmo código do médico para os períodos de internamento, evitou-se que se calculasse, por exemplo, os dias de internamento decorridos desde 10.02.17 a 10.09.10, mas antes de 10.02.17 a 10.03.13, de 10.03.27 a 10.04.08, e assim sucessivamente.

Através do step “*Calculator*”, e como se pode ver pela figura 5.17, definiu-se o cálculo dos dias decorridos desde admissão até à transferência.

▲	#	Novo campo	Cálculo	Campo A	Campo B
	1	Dias	Date A - Date B (in days)	TRANSF_INT	ADMITIDOS

Figura 5.17: Step *Calculator*, utilizado para cálculo dos dias decorridos.

Seguidamente seleccionaram-se apenas determinados campos desejados, ou seja, houve uma limpeza dos dados. O step “*Select Values*” permite remover os campos e, neste caso, todos os atributos que não fossem "Dias" (dias decorridos), "Episódio" e "Cod_Medico" foram excluídos da análise.

Através do “*Filter rows*” estabeleceu-se a condição de que apenas os dias calculados com valor igual ou superior a 0 são tidos em conta.

Por fim, procedeu-se aos cálculos estatísticos, neste caso, o número total de pacientes (amostra N), a média, os mínimos e máximos registados. Para isso, fez-se uso do step “*Univariate Statistics*”, tendo o resultado final culminado na obtenção de um ficheiro de texto.

Esta transformação PDI foi repetida para os diferentes anos: 2010, 2011 e 2012.

Com os resultados obtidos, procedeu-se à elaboração do relatório através do PRD para visualização dos mesmos. Utilizou-se o Report Wizard para isso, uma vez que a sua grande vantagem é a rapidez na criação dos relatórios. Neste caso, porém, apresentou uma limitação, pois pretendia-se elaborar um relatório onde houvesse uma separação mais nítida entre os diferentes anos. Veja-se a figura 5.18: atendendo à pré-visualização do relatório, pretendia-se destacar os diferentes anos, através da selecção do atributo “ANO” como “*Group item*”. Porém, dado que a ferramenta Report Wizard está essencialmente voltada para a análise de custos, por definição, os valores são apresentados com configurações monetárias. Existe um conjunto de formatações-tipo disponibilizadas, mas nenhuma delas aquilo que se pretendia. Tentou formatar-se este item de acordo com o desejado, mas tal só foi possível para os restantes itens que se encontram no campo “*Selected*”. Assim, optou-se por colocar o atributo ano juntamente com os restantes, obtendo-se o relatório disponibilizado no caso de estudo.

Para a criação do dashboard da figura 5.13, fez-se uso do módulo CDE. Estabelecido o layout e o componente a utilizar (*CCC Bar Chart*), foi feita uma consulta SQL à base de dados para disponibilização dos dados:

```
select distinct (int_episodio), sum(ncamasocup) as soma_camas_ocupadas
from ocupacao
where int_episodio is not null
group by int_episodio
order by soma_camas_ocupadas desc
```

A figura 5.14 foi criada após se efectuar o processo ETL e posterior utilização do módulo CDE para disponibilização da tabela em questão. Assim,

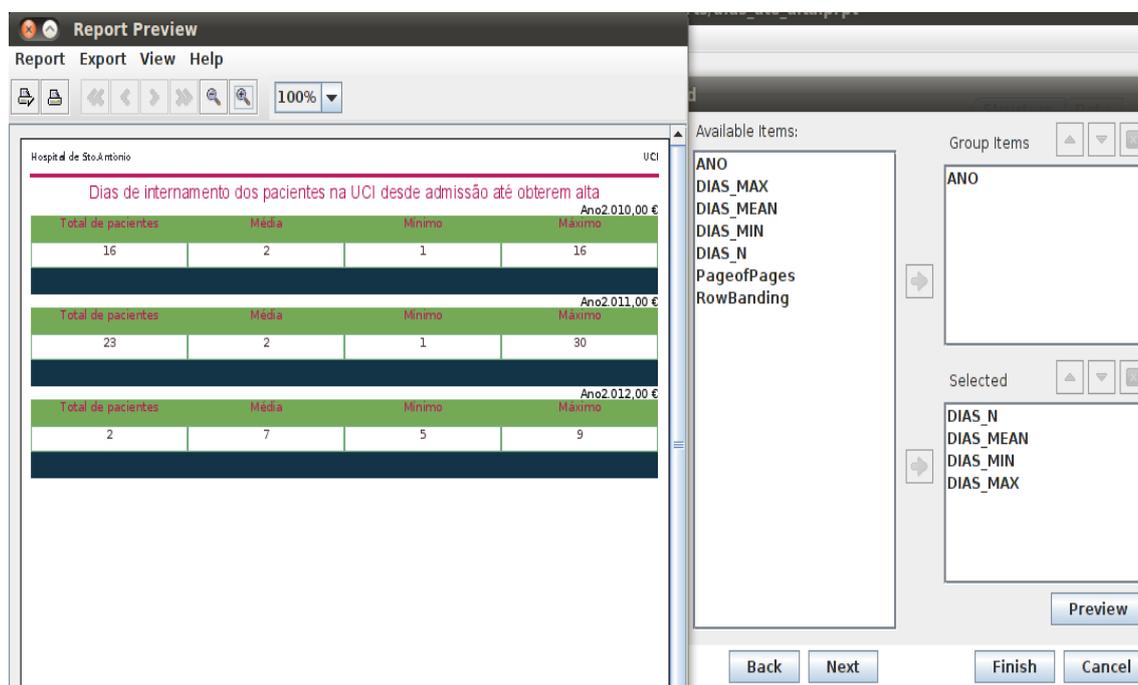


Figura 5.18: Pré-visualização do relatório e configuração dos atributos disponíveis.

inicialmente foi necessário criar uma nova tabela (SOMA_CAMAS) na base de dados, com os atributos "int_episodio" e "soma_camapas_ocupada", por forma a ser possível o armazenamento dos dados relativos a cada episódio ocorrido e o total de dias decorridos, ou seja, os dias de internamento por paciente.

Depois criou-se uma transformação PDI (figura 5.19) para extrair a informação necessária da base de dados e proceder ao seu armazenamento na nova tabela, ou seja, criar o DW.

Posto isto e tendo já o repositório com os dados armazenados, criou-se o dashboard através do módulo CDE, tendo como componente uma tabela cuja informação nela disponibilizada foi extraída a partir da seguinte consulta SQL:

```
select soma_camapas_ocupadas as Dias_internamento,
count(soma_camapas_ocupadas) as Nr_pacientes
from soma_camapas
group by soma_camapas_ocupadas
order by soma_camapas_ocupadas desc
```

5.3. ANÁLISE SEMESTRAL DOS MOVIMENTOS DIÁRIOS REGISTRADOS NA UCI¹

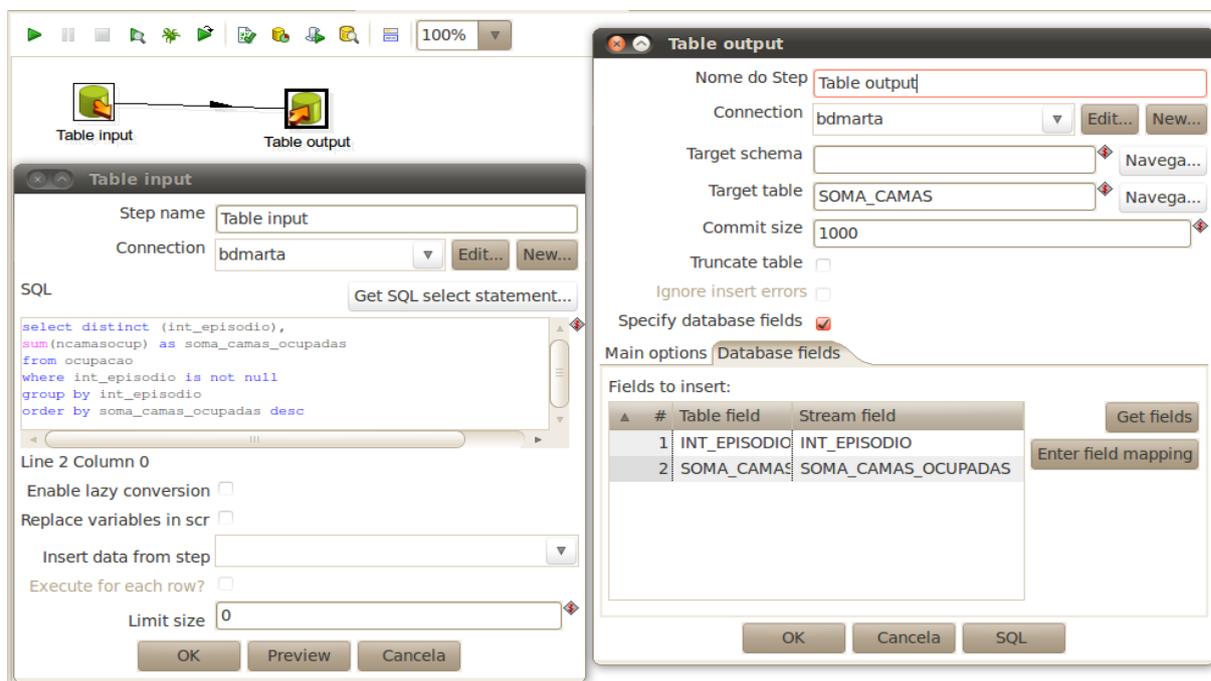


Figura 5.19: Transformação PDI: processo ETL para criação do DW.

5.3 Análise semestral dos movimentos diários registados na UCI

No contexto hospitalar, é crucial ter em conta o número de camas ocupadas pelos pacientes, pois sendo a UCI uma unidade em que os recursos são escassos, tem de se fazer uma gestão equitativa e justa dos mesmos. Para tal, deve proceder-se à realização de um censo hospitalar diário para avaliar o registo diário de movimentos da UCI.

A lotação máxima das camas na UCI do hospital em causa tem sofrido reestruturações. Em 2009 o número de camas disponível era 10, mas em 2010 aumentou para 24 camas. Assim continuou no ano seguinte até que em Julho de 2012 sofreu uma redução, sendo agora a lotação máxima 12 camas. Diariamente, o serviço desta unidade nunca poderá exceder os 100%, uma vez que isso significaria maior número de pacientes para o número de camas disponíveis.

O dashboard que se segue (figura 5.20) provê uma análise semestral (Janeiro a Junho de 2012)¹ do número de movimentos diários registados na UCI, período em que o número de camas disponível na unidade era de

¹No Apêndice A encontra-se o relatório integral desta análise, criado através do PRD.

24 camas. Pela análise do gráfico, percebe-se que no primeiro trimestre se registou, mais vezes que no segundo trimestre, um número de movimentos acima do valor 20, apesar de nunca se ter verificado um registo diário acima dos 22 movimentos. Caso se tivesse verificado a existência de dias em que os movimentos excedessem o valor 24, isto implicaria que teoricamente diferentes pacientes ocupassem a mesma cama, no mesmo dia. Porém, apenas significaria que um paciente admitido na UCI teve alta no mesmo dia em que foi admitido, cedendo lugar a outro paciente, nesse mesmo dia.

Contudo, a necessidade de camas disponíveis nas UCIs em determinadas alturas pode, porém, exceder a oferta, havendo necessidade de priorizar as admissões. Tendo em conta os princípios para a distribuição justa dos recursos existentes numa UCI, definidos pela *American Thoracic Society*, aqueles pacientes que necessitem de cuidados intensivos, devem ser admitidos seguindo a regra *first-come, first-served*. Além disso, pacientes que continuam a cumprir os critérios de necessidade de cuidados médicos intensivos devem continuar a recebê-los, não devendo ter alta prematuramente, apenas com o objectivo de libertar camas para novas admissões. Este princípio deve ser seguido, mesmo que pacientes em concorrência difiram em termos de necessidades ou potenciais benefícios (incluindo factores como a idade, duração prevista de sobrevivência, qualidade de vida prevista após saída da UCI), bastando exceder os limites estabelecidos para receberem cuidados intensivos [40].

Ora, neste contexto, a cada paciente é atribuída uma especialidade responsável e uma especialidade física, onde a primeira corresponde à especialidade incumbida do internamento do paciente e dos cuidados que lhe são prestados, e a segunda à especialidade onde o paciente está fisicamente alocado. Por outras palavras, quando na especialidade responsável não há recursos suficientes para lhe prestar tratamento, o paciente é alocado na especialidade física. Casos destes acontecem quando há sobrelotação de camas, por exemplo. Fez-se uma análise desses casos ocorridos na UCI. Veja-se a figura 5.21, onde num total de 7108 registos referentes à UCI efectuados de 01/01/2009 até 16/09/2012, observa-se que 7073 pacientes foram alocados fisicamente nessa mesma unidade, ou seja, em 99.5% dos casos.

Porém, há ocorrência de 35 casos em que os pacientes tiveram de ser internados noutra especialidade hospitalar, como se analisa na tabela da figura 5.22. Isto implica situações de sobrelotação na UCI (0.5%). Seguidamente, está um gráfico onde se tem melhor visualização das especialidades físicas em que esses internamentos ocorreram, obtido pela ocultação da especialidade UCI.

Conclui-se que a especialidade que mais “acolheu” pacientes provenientes da UCI nos casos de sobrelotação foi a 31700, com 7 registos, seguida pelas

5.3. ANÁLISE SEMESTRAL DOS MOVIMENTOS DIÁRIOS REGISTRADOS NA UCI73

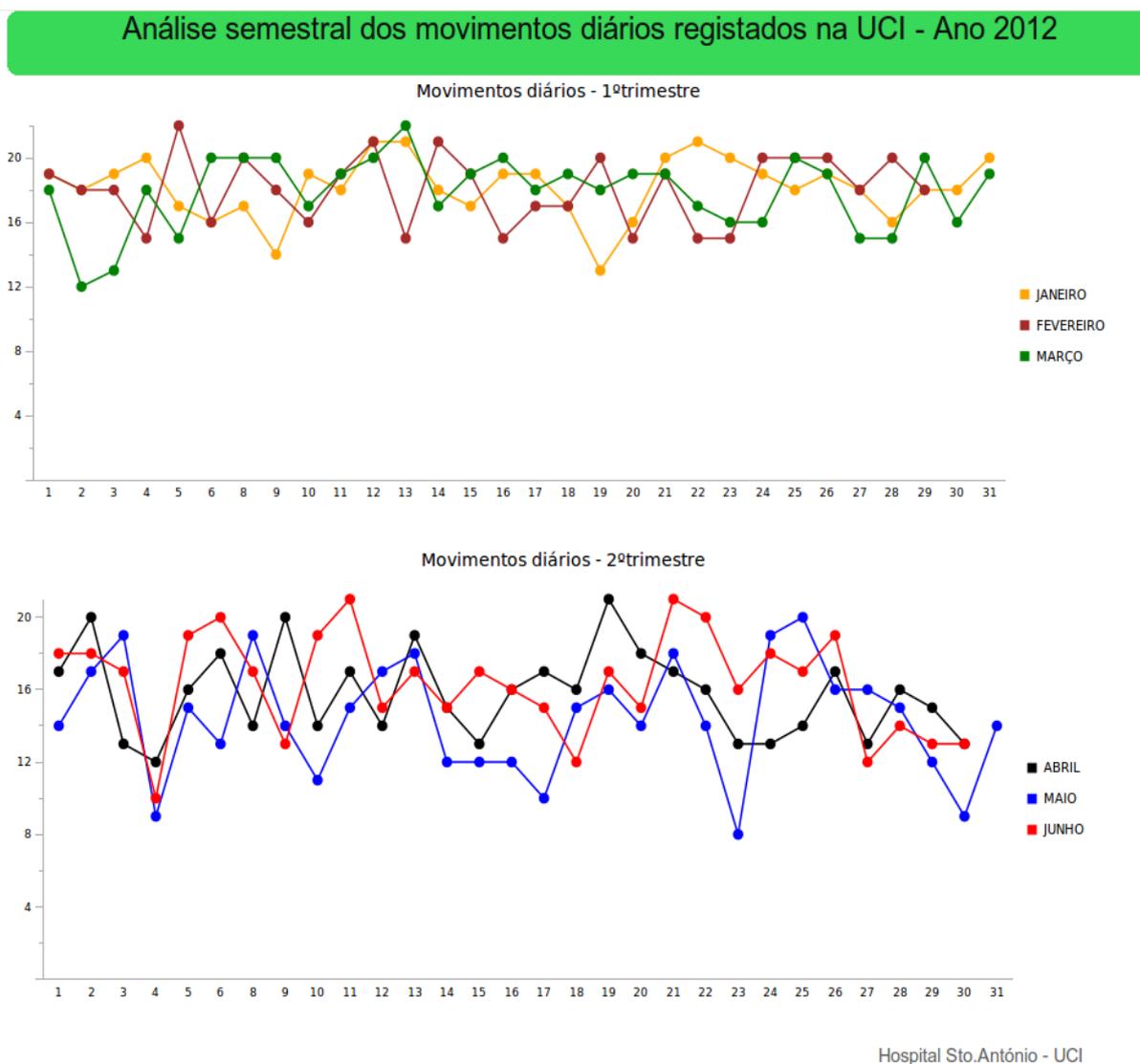


Figura 5.20: Análise semestral dos movimentos diários registados na UCI, nos 1º e 2º trimestre de 2012.

especialidades 30400, 30502 e 30307, com 6 registos cada.

Tendo em conta que se está a analisar um período de quatro anos, constata-se que não ocorreram muitos casos de sobrelotação na UCI, o que significa que o número de camas desta unidade encontra-se adequado ao movimento de pacientes registado e isto é, sem dúvida, positivo.

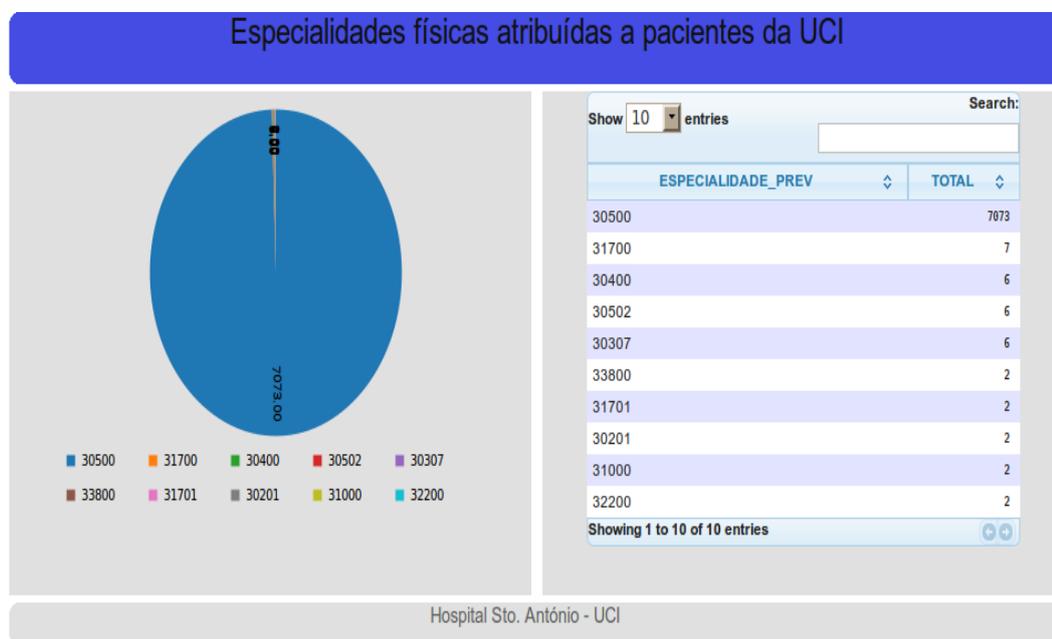


Figura 5.21: Especialidades físicas ("especialidade_prev") de pacientes cuja especialidade responsável é a UCI.

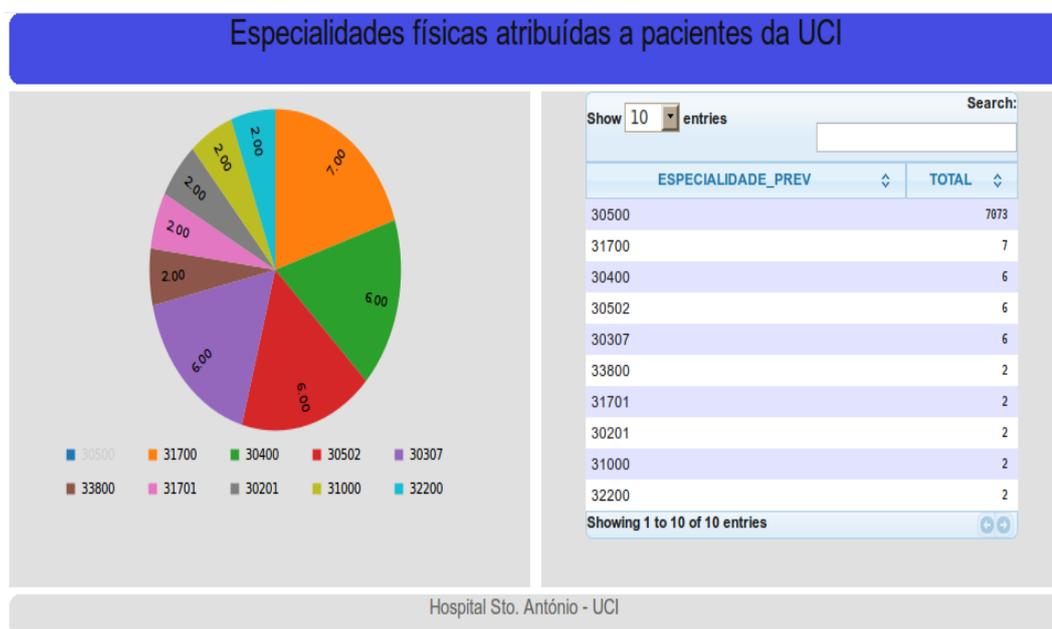


Figura 5.22: Especialidades físicas ("especialidade_prev") de pacientes cuja especialidade responsável é a UCI. Ocultação da especialidade 30500 (UCI) no gráfico circular para melhor visualização das outras especialidades.

5.3. ANÁLISE SEMESTRAL DOS MOVIMENTOS DIÁRIOS REGISTADOS NA UCI75

Processo de desenvolvimento

Para este caso de estudo, tirou-se partido do módulo [CDE](#). Não foi desempenhado qualquer tipo de etapa de [ETL](#), pois o repositório já estava criado, com os dados necessários lá armazenados.

Relativamente ao dashboard da figura [5.20](#) optou-se por se criar um template de raiz, ao invés de se utilizar um previamente criado. Definiu-se o código html pretendido, bem como a base de dados associada, com as seguintes queries SQL, uma para cada trimestre:

```
Select dia, janeiro, fevereiro, março  
from analise_semestral  
order by dia
```

```
Select dia, abril, maio, junho  
from analise_semestral  
order by dia
```

O componente escolhido foi um *CCC Line Chart*, cujas propriedades foram configuradas de acordo com o que se pretendia.

Para o dashboard da figura [5.21](#) dos pacientes cuja especialidade responsável é a [UCI](#), estabeleceram-se dois componentes: um gráfico circular (*CCC Pie Chart*) e uma tabela (*Table Component*), o primeiro associado ao painel da esquerda e o segundo ao painel da direita. Ambos fazem a mesma consulta SQL à base de dados:

```
select especialidade_prev, count(episodio) as total  
from internados  
group by especialidade_prev  
order by total desc
```

De destacar a interactividade concedida durante a visualização dos dashboards, sendo possível ocultar os atributos nos gráficos, conforme desejado, o que é vantajoso para o utilizador, na medida em que lhe dá flexibilidade e aumenta a compreensão dos resultados. Exemplo disso é o dashboard da figura [5.22](#).

5.4 Reinternamentos

Muitos dos pacientes que têm alta após internamento na UCI permanecem em risco de terem complicações posteriormente. Tal facto pode resultar em readmissão na UCI ou noutras especialidades hospitalares.

Pacientes readmitidos na UCI têm um risco acrescido de morte e de prolongamento do internamento o que, conseqüentemente, implica custos hospitalares superiores [32].

Através do software em estudo, fez-se uma análise do número de pacientes readmitidos e qual a especialidade de destino, após prévio internamento na UCI, para os anos 2010, 2011 e 2012. Os resultados são disponibilizados na figura seguinte, porém para melhor visualização dos dados, optou-se por inserir as figuras 5.24, 5.25 e 5.26.

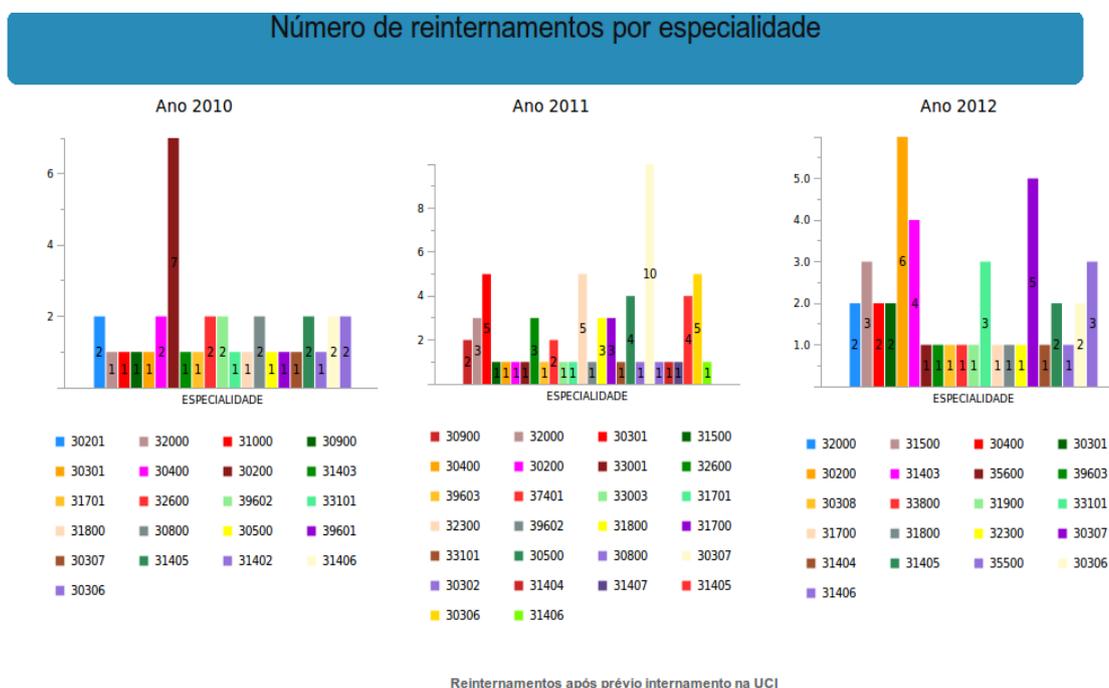


Figura 5.23: Número de reinternamentos ocorridos desde 2010 a 2012 e respectivas especialidades de destino.

Pela observação da figura 5.24, dos 35 reinternamentos efectuados em 2010, destacam-se os 7 registados na unidade 30200. Neste ano, apenas houve uma readmissão na UCI.

Em 2011, como se verifica pela figura 5.25, registaram-se 63 reinternamentos, 10 dos quais na especialidade 30307. 4 pacientes foram reinternados na UCI.

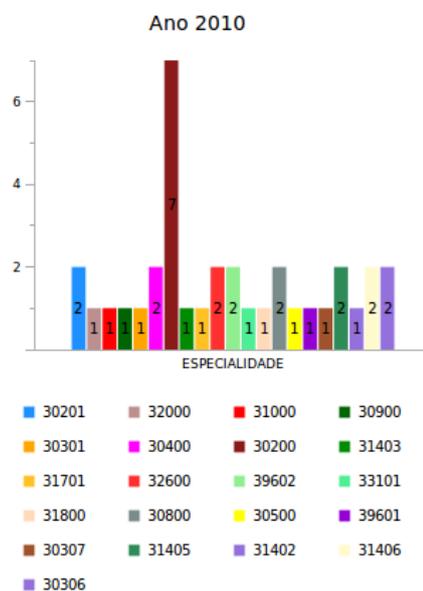


Figura 5.24: Especialidades de reinternamentos ocorridos em 2010.

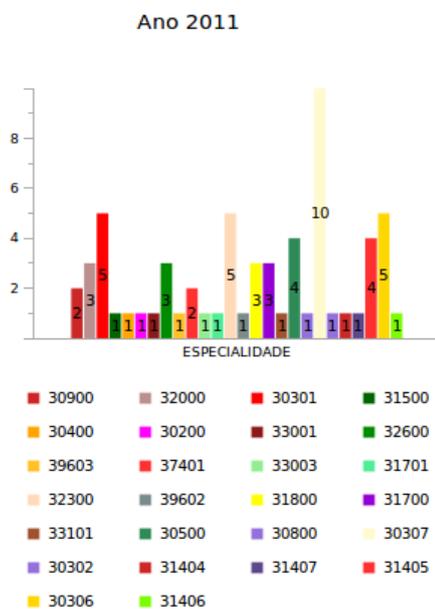


Figura 5.25: Especialidades de reinternamentos ocorridos em 2011.

Por fim, analisando o ano 2012 (figura 5.26), das 21 especialidades que admitaram pacientes previamente internados na UCI, destacam-se a 30200 com 6 reinternamentos, a 30307 e a 31403, com 5 e 4 reinternamentos, respectivamente. Verifica-se que não ocorreram reinternamentos na UCI.

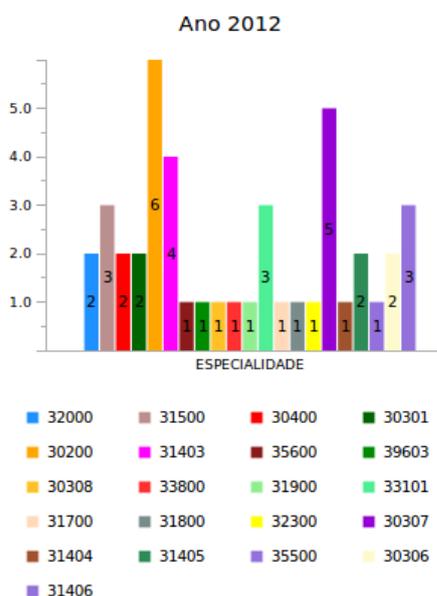


Figura 5.26: Especialidades de reinternamentos ocorridos em 2012.

Ora, pelos gráficos disponibilizados, conclui-se que as unidades hospitalares que readmitem mais pacientes previamente internados na UCI são as 30200, 30307 e 31403.

Fez-se uso do PRD para elaborar um relatório com uma análise anual das readmissões hospitalares, das readmissões registadas na UCI e noutras especialidades (figura 5.27). Para esta análise, está a ter-se em conta 7108 registos contidos na base de dados. Assim, em 3 anos, para a totalidade desses registos, apenas 1,99% correspondem a reinternamentos de pacientes previamente internados na UCI. Como já referido previamente, só 5 pacientes voltaram à UCI (4 deles no ano 2011), representando 3,52% do total de reinternamentos, sendo os restantes 96,48% relativos a reinternamentos noutras especialidades.

Se se analisar os reinternamentos ano a ano, em 2010 representam 1,53% do total de movimentos registados na UCI nesse ano, em 2011 constituem 2,54%, e por último, 3,84% para o ano de 2012.

Deste modo, tendo em conta os resultados obtidos, onde apenas 1,99% dos registos correspondem a reinternamentos e só 5 pacientes, em 3 anos, terem sido reinternados na UCI, e dado que não é possível fazer uma análise mais profunda acerca dos reinternamentos ocorridos, nomeadamente das suas causas, a única conclusão fidedigna a retirar, é que a UCI eventualmente não apresenta custos acrescidos com os reinternamentos, nem tem implicações no aumento de custos para as outras unidades hospitalares, uma vez que os rein-

Hospital Sto. António

Hospital readmissions			
Year	Number of readmissions	ICU readmissions	Readmissions in other specialties
2010	35	1	34
2011	63	4	59
2012	44	0	44
TOTAL:	142	5	137

Figura 5.27: Relatório com análise anual do número total de readmissões, readmissões na UCI ou noutras especialidades.

ternamentos noutras especialidades apenas representaram, nestes três anos, 1.93% dos registos ocorridos.

Por último, fazendo uma análise aos dias de internamento dos pacientes reinternados, através do relatório 5.28, criado com o PRD, verifica-se que no total de 142 readmissões, apenas 3 pacientes estiveram internados mais de 50 dias. Porém, desses 3, é de realçar que dois deles faleceram. Readmissões e um tempo longo de internamento estão relacionados com taxas de mortalidade superiores. Além disso, o período de internamento de um paciente é um factor determinante nos custos e utilização dos recursos da UCI, sendo que a sua maior utilização ocorre para a pequena proporção de pacientes com internamentos mais longos [41].

Hospital Sto. António

UCI

Listagem de pacientes reinternados com período de estadia superior a 50 dias	
Episódio	Dias de internamento
10005586	67
Episódio	Dias de internamento
10030236	66
Episódio	Dias de internamento
11026552	54

Figura 5.28: Listagem de pacientes reinternados com tempo de internamento superior a 50 dias.

Reinternamentos em horas críticas

Apesar de ser um assunto controverso, que não gera consenso entre os especialistas, as taxas de readmissão precoces (até às 48 horas) são encaradas como uma variável de qualidade dos cuidados prestados na UCI, sendo sem dúvida, um grupo que merece especial atenção. Isto porque estão directamente relacionadas com elevados números de óbitos, bem como com complicações terapêuticas que poderiam ter sido evitadas.

Uma das causas que poderá estar na origem destas readmissões remete para pacientes que obtiveram alta precocemente devido à limitação dos recursos da UCI. É, desta forma, crucial fazer um estudo deste tipo de situação para se poder actuar preventivamente. Assim, fazer um prévio reconhecimento e gestão dos pacientes com alto risco de readmissão nas UCIs é importante quer para otimizar os cuidados prestados aos pacientes, quer para minimizar os custos hospitalares [33].

Analisando conjuntamente os períodos relativos a 2010, 2011 e 2012, destaca-se o facto de não terem ocorrido reinternamentos em horas críticas na UCI. Observa-se pela figura 5.29 que ocorreram 2 reinternamentos em menos de 24 horas: 1 na especialidade 30201 e outro na especialidade 30306.

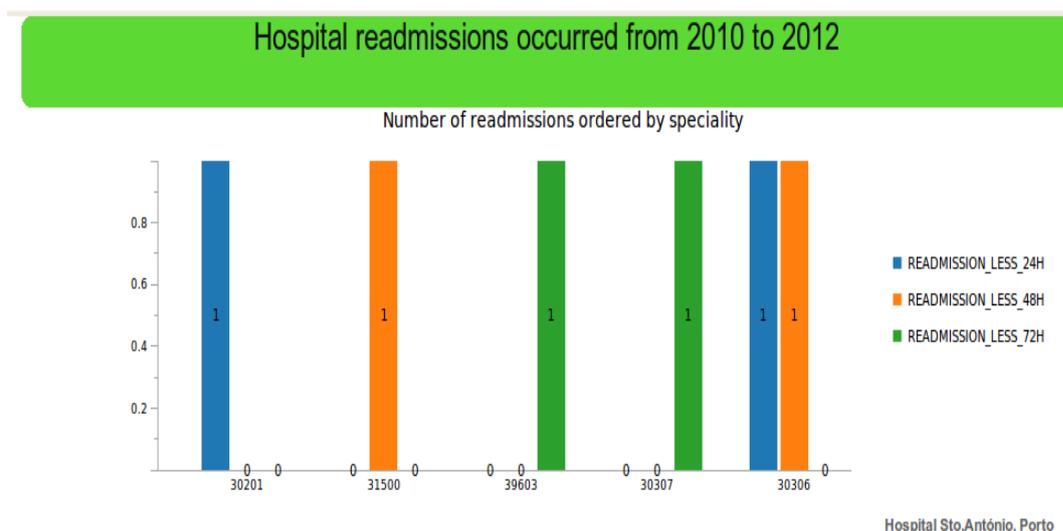


Figura 5.29: Especialidades de destino de reinternamento efectuado em menos de 24h, em menos de 48h e em menos de 72h, após prévio internamento na UCI.

Quanto aos reinternamentos em menos de 48 horas, 1 deles ocorreu na especialidade 31500 e outro na especialidade 30306. Por fim, após as 48h e até às 72h, foram readmitidos 2 pacientes, 1 na especialidade 39603 e outro na 30307.

Tal como já referido anteriormente, reinternamentos em horas críticas apresentam maior probabilidade de morte dos pacientes. Porém, nenhum destes 6 casos de reinternamentos culminou na morte dos pacientes.

Pelos dados disponíveis não é possível determinar as causas de reinternamento destes pacientes, no entanto conclui-se que o facto de nenhum deles ter sido novamente internado na UCI poderá ser um factor positivo, tal como nenhum deles ter falecido, pois como já referido, um paciente reinternado na UCI apresenta risco acrescido de morte.

Processo de desenvolvimento

Para a elaboração deste caso de estudo, foram utilizadas três componentes integrados no Pentaho, nomeadamente do PDI para a realização das etapas de ETL, do módulo CDE para a elaboração dos dashboards, e também do PRD para a criação do relatório.

Assim, para o desenvolvimento deste caso de estudo, inicialmente procedeu-

se ao processo de ETL, através do PDI, onde os dados relativos aos reinternamentos foram extraídos da base de dados onde estavam armazenados conjuntamente com todos os registos, e armazenados e transferidos para essa mesma base de dados, mas para tabelas de destino novas - DMs - criadas para tal. Assim, os dados relativos aos reinternamentos de 2010, 2011 e 2012 apenas foram extraídos e carregados para as três novas tabelas: REINTERNAMENTOS_2010, REINTERNAMENTOS_2011 e REINTERNAMENTOS_2012, respectivamente, ou seja, não houve qualquer transformação efectuada aos dados. Criaram-se as novas tabelas e, seguidamente, foi criada uma transformação no PDI para cada um dos anos, com esquema representado na figura 5.30.

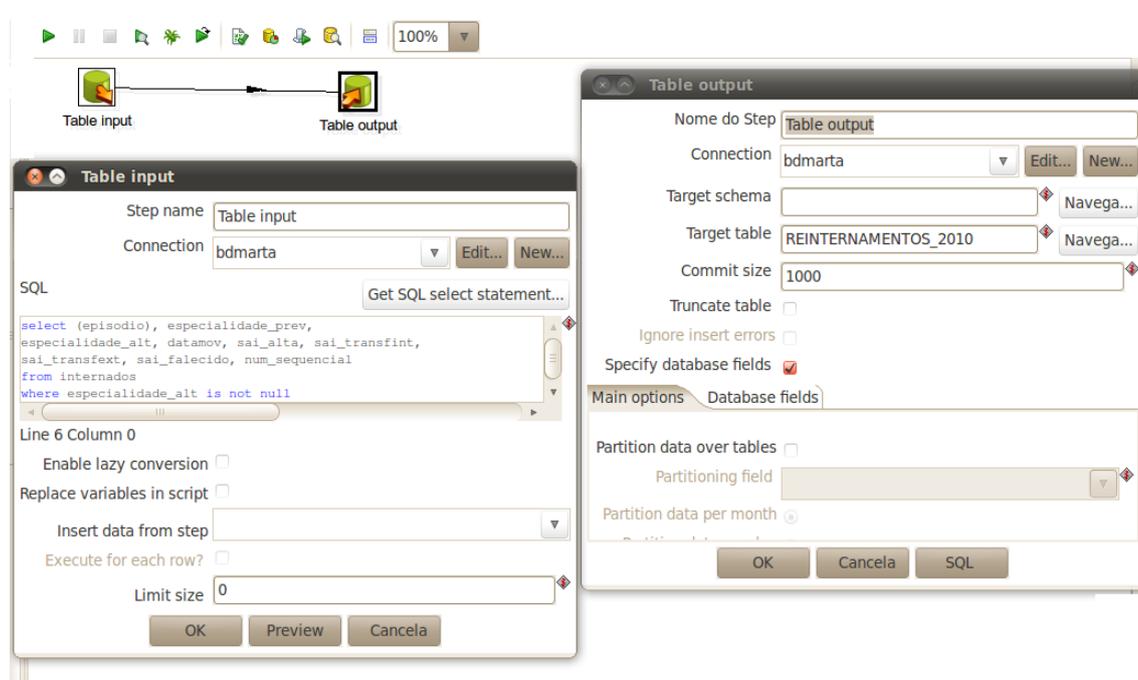


Figura 5.30: Transformação criada através do PDI para extracção dos dados relativos aos reinternamentos ocorridos em 2010 e o seu carregamento para a tabela REINTERNAMENTOS_2010.

Como se vê, a transformação é constituída por uma tabela de entrada e uma de saída para que se fossem buscar determinados atributos (referidos no campo SQL) relativos ao ano 2010 e armazenados na tabela "internados" e os inserisse numa nova tabela (*Table output*: REINTERNAMENTOS_2010). O mesmo procedimento foi efectuado para os anos 2011 e 2012.

Posto isto, procedeu-se à criação do primeiro dashboard (figura 5.23), através do módulo CDE. Definiu-se o layout pretendido e a base de dados

a utilizar, tal como as consultas SQL necessárias (uma para cada ano) e atribuiu-se a cada uma delas um componente. Neste caso, optou-se por utilizar gráficos de barras (*CCC Bar Chart*). As propriedades gráficas e os ajustes ao layout foram feitos conforme pretendido.

No entanto, é de realçar a desvantagem encontrada aquando da definição das cores das barras do gráfico. Ou seja, observando a figura 5.31 correspondente ao gráfico inicialmente criado, percebe-se que há barras compostas pelas mesmas cores, porque o módulo CDE, por defeito, apenas disponibiliza dez cores diferentes. Isto torna-se ambíguo aquando da interpretação do gráfico e poderá levar a que o utilizador final tire conclusões erradas acerca das informações disponibilizadas.

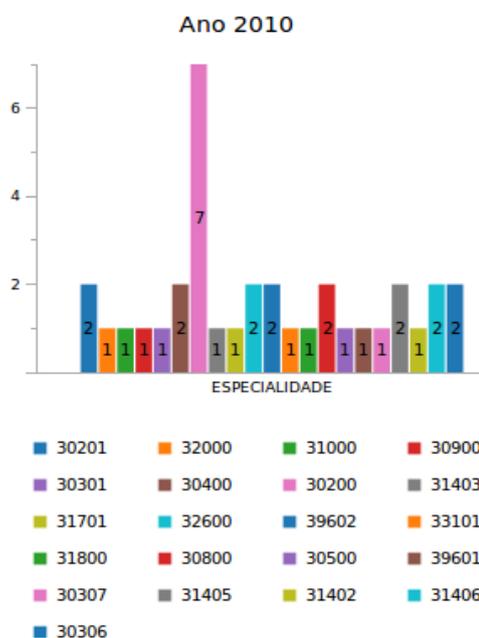


Figura 5.31: 10 cores padrão reconhecidas pelo Pentaho.

Para tentar solucionar o problema, introduziram-se as cores pretendidas como argumentos, tal como se vê na figura 5.32. Contudo, esta tentativa mostrou-se infrutífera, uma vez que era necessário definir cerca de 21 cores (21 especialidades diferentes, logo 21 barras) e as cores atribuídas não eram reconhecidas pelo software. Deste modo, a legenda dos gráficos teria de ser lida atentamente, por ordem de aparecimento de cada barra.

Ainda em relação a este assunto, sendo a visualização gráfica feita na interface web do Pentaho, torna-se mais fácil o entendimento das legendas, pois esta visualização é interactiva e dinâmica, e caso se passe o cursor por

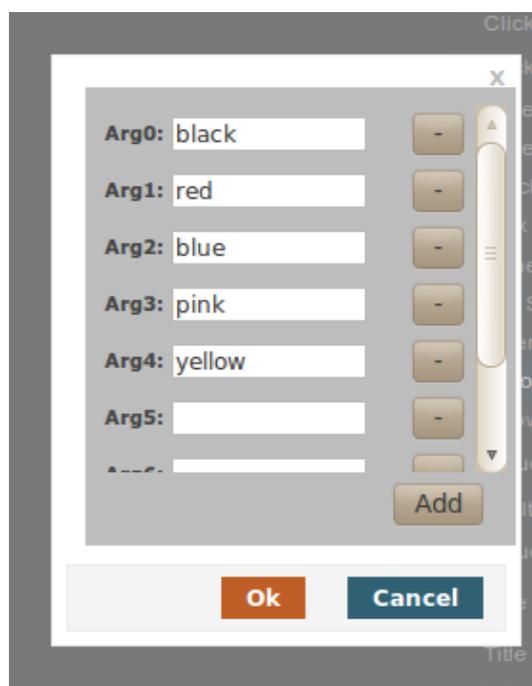


Figura 5.32: Propriedades dos componentes no módulo CDE: “Colors”.

cima de cada barra do gráfico é disponibilizada a informação relativamente a ele, como se observa na figura 5.33.

A solução encontrada e aquela que foi disponibilizada no caso de estudo, passou por se definir manualmente os códigos de cores em html, como se vê na figura 5.34. No entanto, isto apesar de resultar, não é prático nem rápido.

Relativamente ao relatório com as informações do número total de readmissões, readmissões na UCI ou noutras especialidades (figura 5.27), este foi elaborado através do PRD. Primeiro, escolheu-se o layout (*Look and feel*), seguidamente a base de dados com a consulta SQL pretendida foi associada (*Data source*), seleccionaram-se os atributos desejados resultantes dessa mesma consulta (*Layout Step*) e formataram-se esses atributos (*Format Step*) quanto ao nome, alinhamento, formato e agregação, como se observa na figura 5.35. Por fim, procedeu-se a pequenos ajustamentos, como tamanho de letra, cores, cabeçalhos, entre outros.

No caso de estudo é depois apresentado o dashboard relativo às readmissões em horas críticas (figura 5.29). Ora, também aqui foi necessário utilizar

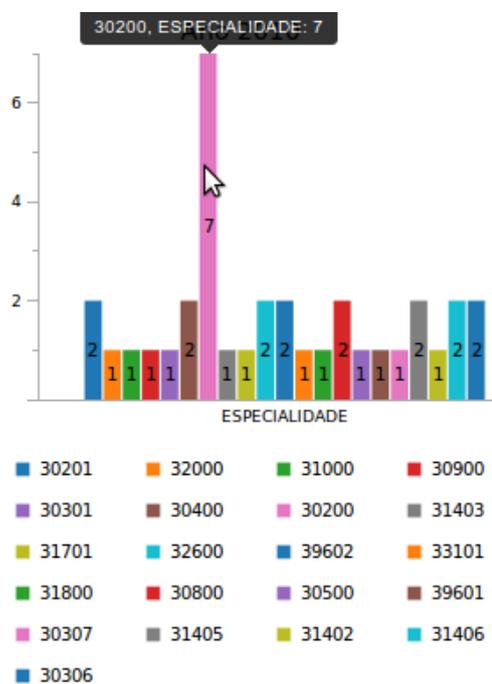


Figura 5.33: Visualização gráfica interactiva com disponibilização dos dados referentes a cada barra.

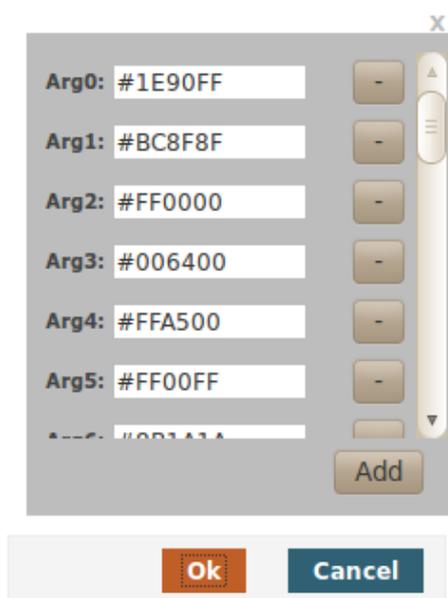


Figura 5.34: Códigos de cores HTML introduzidos.

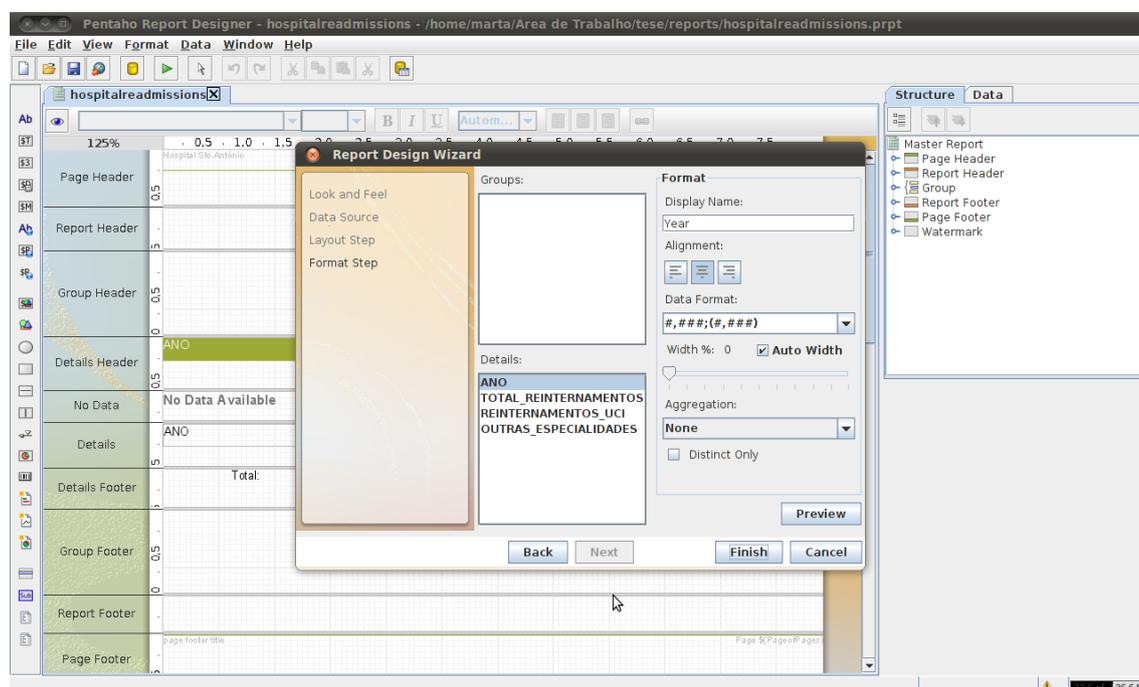


Figura 5.35: Elaboração do relatório relativo às readmissões hospitalares – PRD.

o PDI para agregar os dados relativos a reinternamentos em horas críticas. Seguiu-se o mesmo raciocínio que acima, com extração dos dados da base de dados e a sua introdução em nova tabela.

Posteriormente e mais uma vez, foi utilizado o módulo CDE para a disponibilização gráfica dos dados e seguiu-se o mesmo processo de elaboração que para os dashboards já explicados anteriormente. A consulta SQL efectuada à tabela criada através do processo PDI foi a seguinte:

```
select especialidade_ alt as speciality_ readmission,
sum(h24) as readmission_ less_ 24h, sum(h48) as readmission_ less_ 48h,
sum(h72) as readmission_ less_ 72h
from horas
group by especialidade_ alt
```

O componente associado a esta consulta foi um gráfico de barras (*CCC Bar Chart*).

Como se observa na figura 5.36, foi construída uma transformação PDI

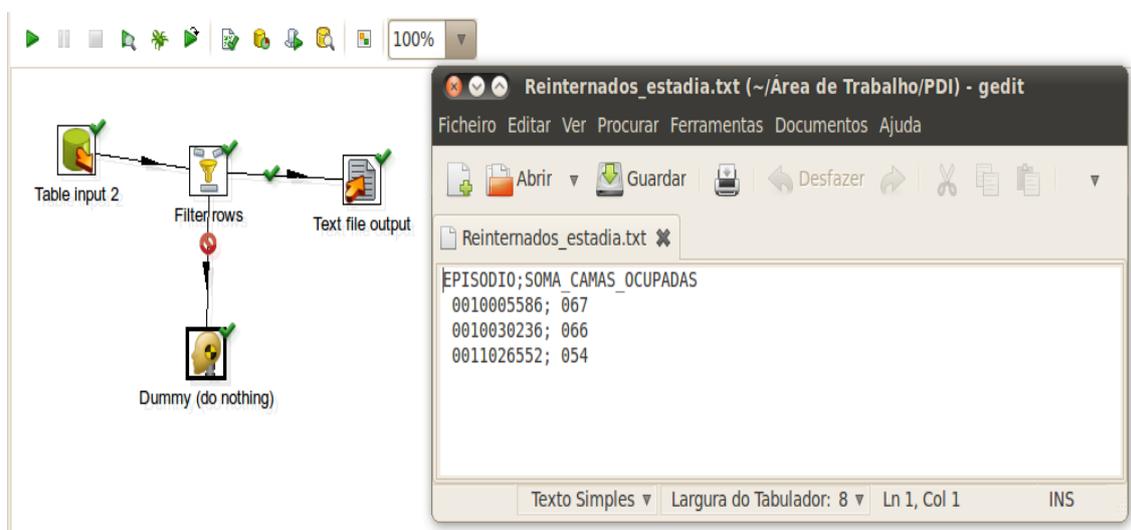


Figura 5.36: Esquema utilizado no PDI e resultado obtido disponibilizado num ficheiro de texto.

para se obter o episódio relativo aos pacientes reinternados com período de estadia acima dos 50 dias. Utilizou-se uma tabela como entrada, com uma consulta SQL feita ao DW, e filtraram-se as linhas relativas aos pacientes com período de internamento igual ou superior a 50 dias. Para tal foi imposta uma condição (*soma_camias_ocupadas* >= '50'), como se observa na figura 5.37.

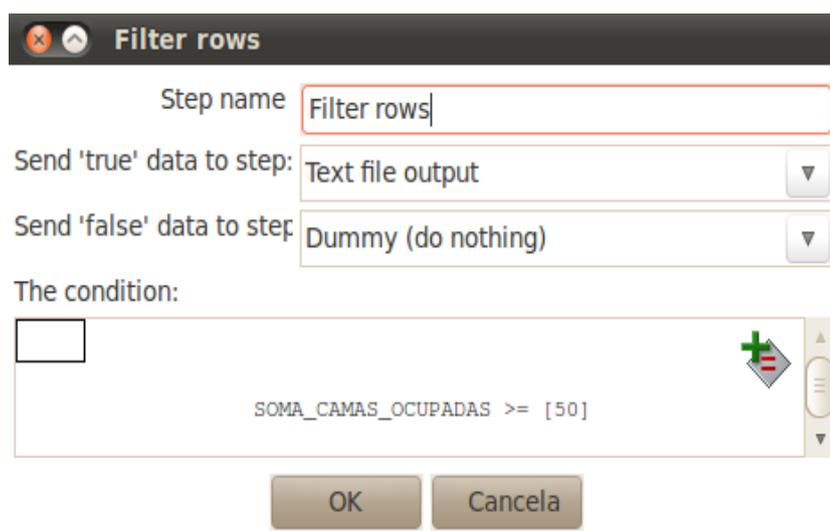


Figura 5.37: Step *Filter Rows* da transformação PDI

Ainda relativamente à etapa ETL, como forma de experimentar e analisar as diferentes ferramentas concedidas pela plataforma Pentaho, construiu-se outra transformação, com o mesmo objectivo que a anterior, mas fazendo uso de outros *steps*. Assim, ao invés de se usar um filtro de linhas, optou-se por utilizar o *step* “*Merge Join*”, que iria encontrar os episódios em comum (*Inner Join*) das tabelas de entrada: uma com os episódios de todos os pacientes reinternados e outra com os episódios dos pacientes que permaneceram internados na UCI mais de 50 dias. Porém, como resultado apenas se obteve um ficheiro de texto vazio, como se observa na figura 5.38.

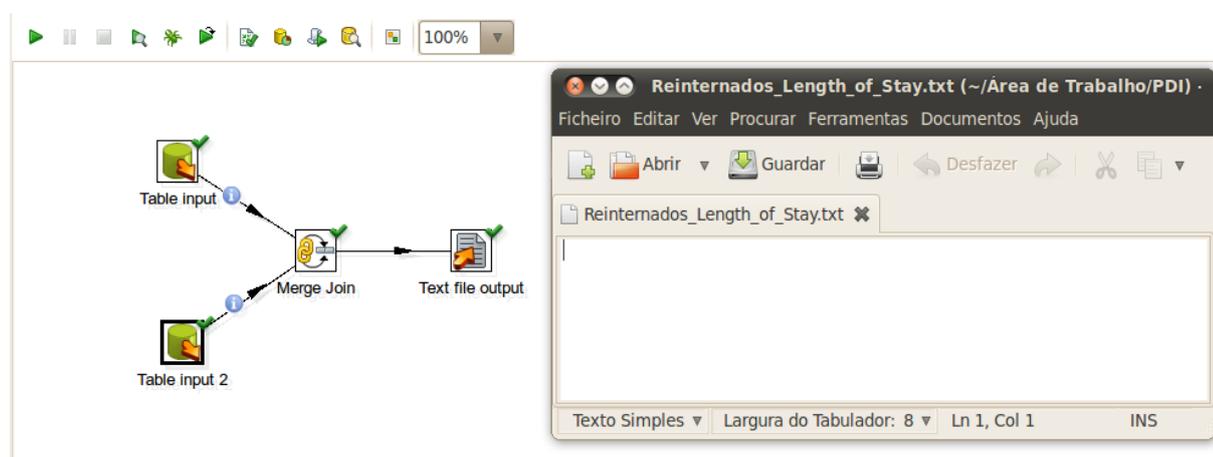


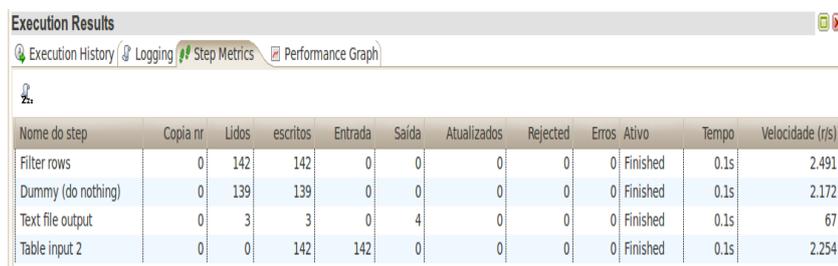
Figura 5.38: Transformação PDI com utilização do *step Merge Join* e resultado obtido.

Tentou perceber-se o porquê dos resultados não serem disponibilizados, sem sucesso. Inicialmente, pensou-se que poderia ser alguma discrepância existente no tipo de atributos, mas em ambas as tabelas o atributo *episódio* é *number(10,0)*.

Analisando os resultados de execução das duas transformações 5.39 e 5.40, verifica-se que, de facto, não há resultados no ficheiro de texto para a transformação com o *Merge Join* e, neste *step*, apesar de serem lidos 152 registos (soma dos 142 do *Table input* com os 10 do *Table input 2*), estes não são escritos.

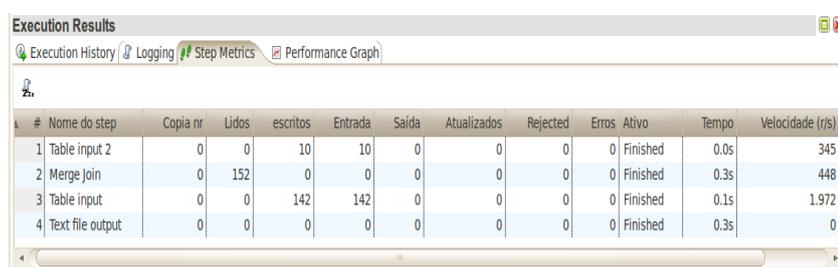
Por via das dúvidas, testaram-se os outros tipos de *join* disponibilizados (*Full Join*, *Right Join* e *Left Join*) e, de facto, os resultados obtidos eram válidos e estavam de acordo com o esperado.

Posto isto, após se terem obtido estes resultados, procedeu-se à elaboração do relatório da figura 5.28, com o PRD, mas desta vez, criando-se um de raiz. Este módulo fornece imensas funcionalidades e, desta forma, caso



Nome do step	Copia nr	Lidos	escritos	Entrada	Saída	Atualizados	Rejected	Erros	Ativo	Tempo	Velocidade (r/s)
Filter rows	0	142	142	0	0	0	0	0	Finished	0.1s	2.491
Dummy (do nothing)	0	139	139	0	0	0	0	0	Finished	0.1s	2.172
Text file output	0	3	3	0	4	0	0	0	Finished	0.1s	67
Table input 2	0	0	142	142	0	0	0	0	Finished	0.1s	2.254

Figura 5.39: Resultados da execução provenientes da transformação PDI com utilização do step *Filter Rows*.



#	Nome do step	Copia nr	Lidos	escritos	Entrada	Saída	Atualizados	Rejected	Erros	Ativo	Tempo	Velocidade (r/s)
1	Table input 2	0	0	10	10	0	0	0	0	Finished	0.0s	345
2	Merge Join	0	152	0	0	0	0	0	0	Finished	0.3s	448
3	Table input	0	0	142	142	0	0	0	0	Finished	0.1s	1.972
4	Text file output	0	0	0	0	0	0	0	0	Finished	0.3s	0

Figura 5.40: Resultados da execução provenientes da transformação PDI com utilização do step *Merge Join*.

o utilizador opte por criar um relatório ao seu critério tem toda a liberdade para o fazer, podendo moldá-lo das mais diversas maneiras, consoante o que pretende.

Por último, é dito neste caso de estudo que os pacientes readmitidos em horas críticas não faleceram. Para tal constatação ser feita, teve de proceder-se às etapas de ETL. A transformação criada está disponibilizada na figura 5.41. Assim, através de consultas SQL à base de dados, extraiu-se o número sequencial dos pacientes falecidos e dos pacientes reinternados, definindo um *INNER Join* para que a informação comum fosse carregada para o DW - a tabela estabelecida como output (Pacientes_reinternados_falecidos). Porém, as tabelas de entrada não possuíam nenhum número sequencial em comum, o que implica que não ocorreu carregamento de dados para a tabela de saída e, deste modo, conclui-se que de facto, nenhum paciente sujeito a reinternamento faleceu.

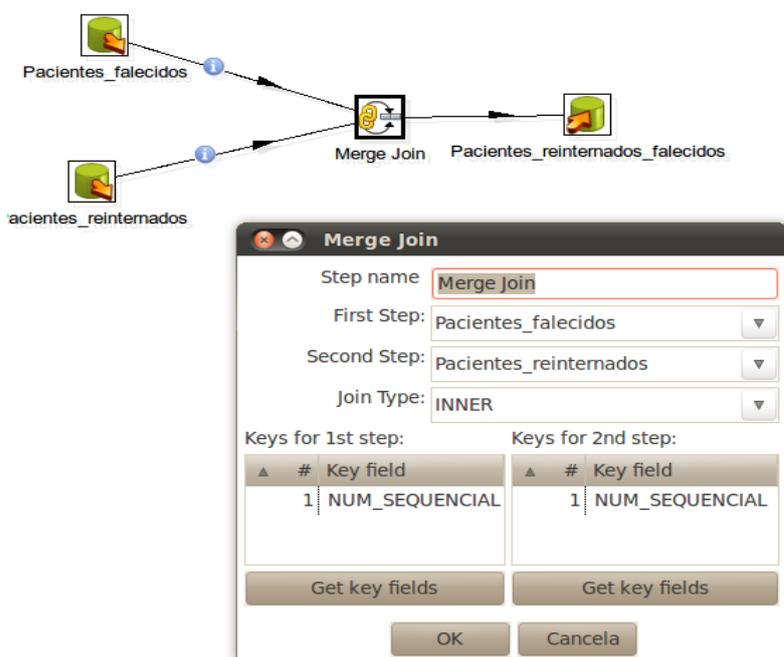


Figura 5.41: ETL: transformação criada através do PDI para saber quais os pacientes reinternados que faleceram.

5.5 Cirurgias realizadas a pacientes da UCI

Conhecer os pacientes cirúrgicos admitidos nas UCIs, perceber qual a gravidade do seu estado, os recursos aplicados, o tipo de cirurgias a que foram submetidos, entre outros aspectos, é importante quer para saber a quantidade de recursos utilizados, quer para, conseqüentemente, ajustar e gerir esses mesmos recursos da melhor forma possível.

Além disso, a análise deste tipo de dados poderá contribuir para saber em que medida os tratamentos e cuidados instituídos aos pacientes contribuíram para a recuperação dos pacientes.

Para este caso de estudo apenas estavam disponíveis 80 registos de pacientes cirúrgicos, sendo todos eles relativos aos primeiros três meses de 2012. Destes registos, 42 são relativos a pacientes do sexo masculino e 38 do sexo feminino.

Começando por uma análise mensal às cirurgias realizadas (5.42), verifica-se que Janeiro foi o mês com maior número de cirurgias urgentes, num total de 25, seguido por 23 realizadas em Fevereiro e 18 em Março.

Relativamente às reoperações, foram efectuadas 17 em Janeiro, 18 em

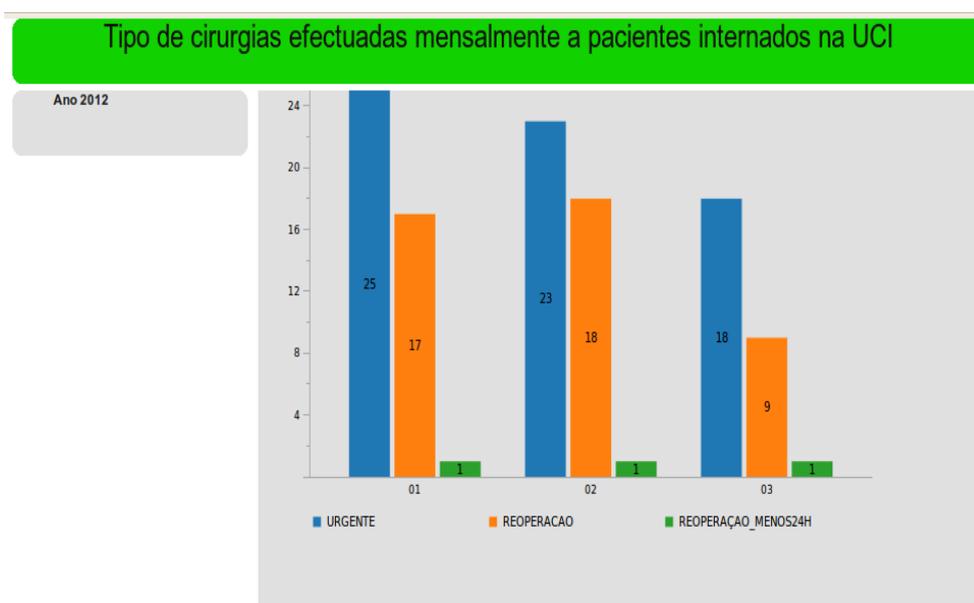


Figura 5.42: Dashboard com o tipo de cirurgias (urgente, reoperação e reoperação em menos de 24h) efectuadas aos pacientes da UCI.

Fevereiro e 9 em Março, sendo que destes valores, 1 reoperação foi realizada em menos de 24h relativamente à cirurgia anterior para cada mês. No dashboard que se segue (5.43) visualizam-se as patologias responsáveis por essas três reoperações realizadas em horas críticas:

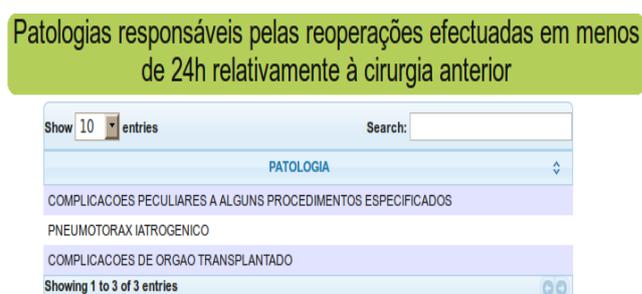


Figura 5.43: Dashboard com as patologias responsáveis pelas reoperações efectuadas em menos de 24h.

Percebe-se então que todas as patologias resultam de complicações provenientes da cirurgia anterior, como seria de esperar.

Como se vê pela figura 5.44, de todas as 80 cirurgias realizadas, há um

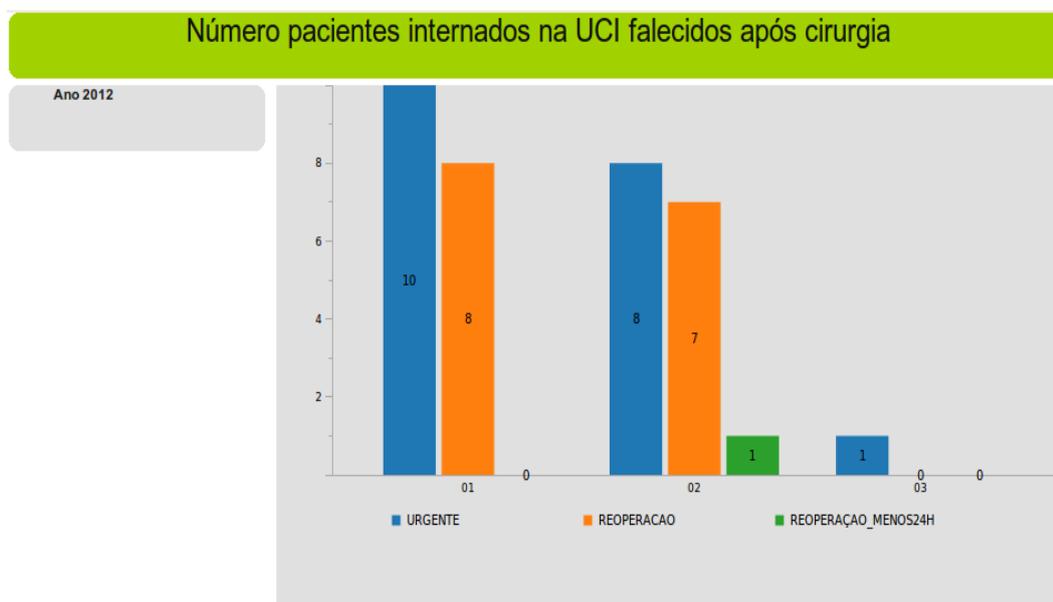


Figura 5.44: Tipo de intervenção cirúrgica a que os pacientes falecidos foram submetidos.

registo de 34 pacientes falecidos após serem submetidos a cirurgias: 19 sujeitos a cirurgias urgentes, 15 depois de reoperação, 1 dos quais sujeito a uma reoperação em menos de 24 horas. Ora, isto corresponde a 42.5% de pacientes falecidos na totalidade de cirurgias submetidas, o que é considerado um número elevado mas, apesar de tudo, compreensível, dado o estado crítico dos pacientes e o tipo de cirurgias em questão.

Ainda relativamente ao tipo de cirurgias realizadas, conclui-se que as operações urgentes foram efectuadas em maior número a mulheres do que a homens, como se pode visualizar no gráfico da figura 5.45. A barra vermelha corresponde aos dois sexos, a barra azul a homens e a verde a mulheres.

Num total de 44 reoperações, 23 foram efectuadas a homens e 21 a mulheres. Verifica-se também que das 3 reoperações realizadas em menos de 24h, todos os pacientes eram homens.

Através do gráfico 5.46, construído com recurso à análise OLAP, verifica-se que as cirurgias realizadas a pacientes do sexo masculino (corresponde ao 1 no gráfico) estão distribuídas por mais especialidades, apesar das mulheres (corresponde ao 2 no gráfico) serem operadas em maior número (excepto nos transplantes hepáticos e especialidade de urologia).

Deste modo, as especialidades de Cirurgia Vascular, Estomat/cir,max facial, Oftalmologia, Prog transp reno-pancreático e Prog transplante pan-

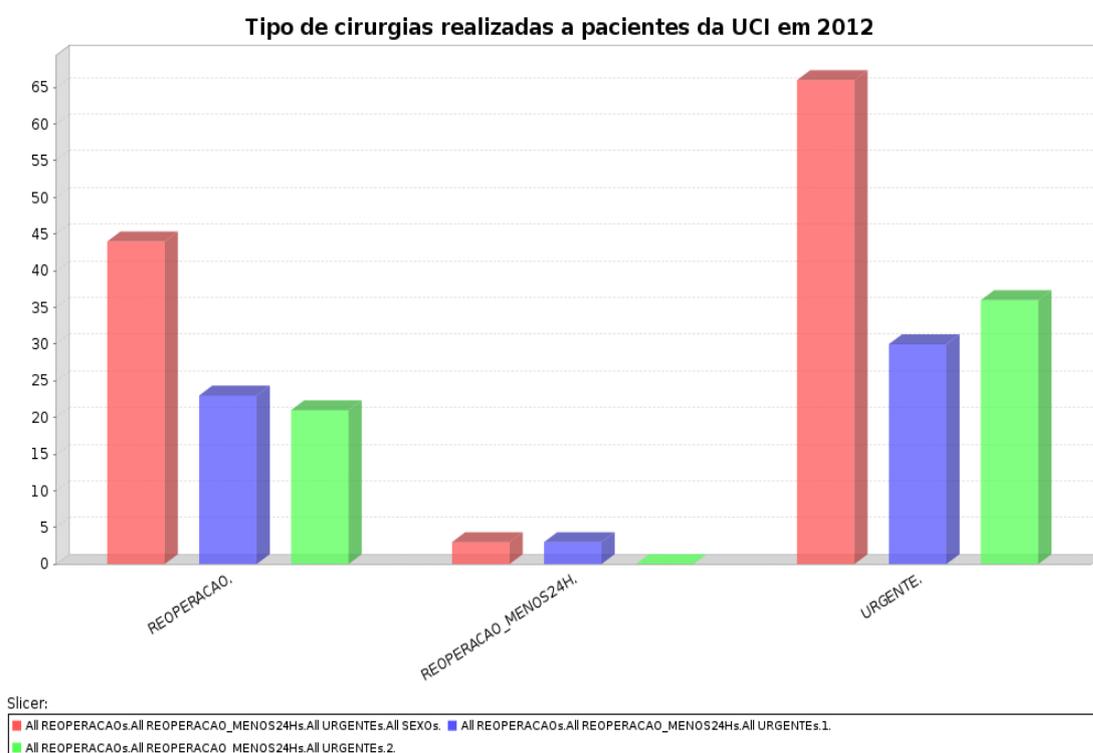


Figura 5.45: Tipo de cirurgias realizadas segundo o sexo dos pacientes da UCI.

creático só foram realizadas a homens e, como é óbvio, Obstetrícia MJD é uma especialidade exclusivamente feminina. Em 80 cirurgias analisadas, destacam-se as cirurgias vasculares realizadas a pacientes masculinos e as Cirurgias 1 a pacientes do sexo feminino.

Voltando às patologias, mas desta vez em relação às cirurgias urgentes, verifica-se que a fascíte necrotizante é responsável pelo maior número de cirurgias urgentes, seguida pelas complicações provenientes de órgãos transplantados, como se pode ver pelo dashboard da figura 5.47. Apenas são apresentadas as primeiras 10 entradas por serem as mais representativas, dado o número de ocorrências.

A fascíte necrotizante é uma infecção bacteriana grave e pouco frequente caracterizada por extensa necrose dos tecidos, sendo uma infecção das camadas mais fundas da pele e tecidos subcutâneos, espalhando-se facilmente pelo plano facial dentro do tecido subcutâneo. Existem muitos tipos de bactérias causadores deste tipo de infecção.

Ainda relativamente às operações de carácter urgente, o gráfico da figura

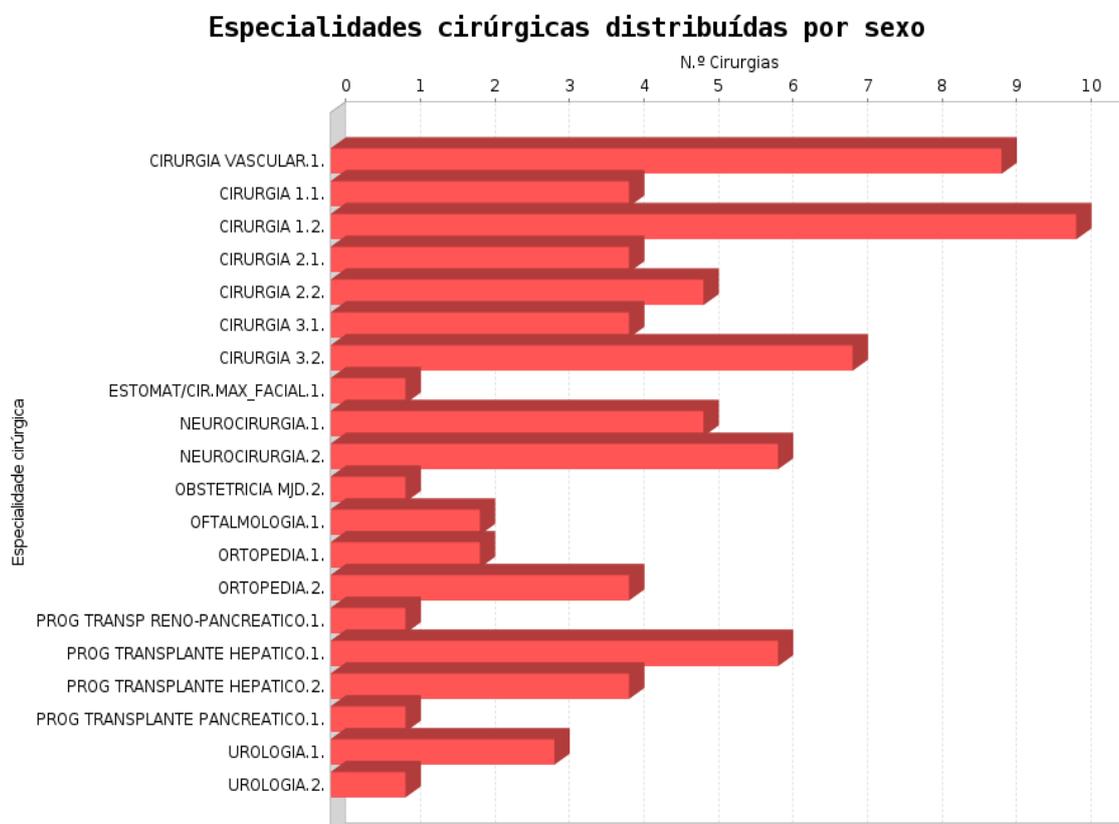


Figura 5.46: N.º cirurgias realizadas, segundo a especialidade e sexo dos pacientes (1 – sexo masculino; 2 – sexo feminino).

5.48 permite fazer uma análise qualitativa das especialidades responsáveis por essas operações a pacientes da UCI.

Percebe-se que não há uma especialidade que se destaque, sendo que a maior percentagem corresponde à Cirurgia1 com 12,18%, seguida pela Cirurgia2 com 9,14%. As especialidades Prog Transplante Pancreático, Estomat/Cir.Max_Facial, Obstetícia e Prog.Transplante Reno-Pancreático têm relevância pouco significativa neste campo, representando cada um apenas 1,2% das cirurgias urgentes realizadas.

Caso se pretenda fazer uma pesquisa por episódio, por forma a saber a data da cirurgia, a patologia e a especialidade relativa às cirurgias a que cada paciente esteve sujeito, pode-se fazer uma pesquisa interactiva através do Pentaho, como exemplificam as figuras 5.49 e 5.50, que são relativas às cirurgias dos pacientes com episódios 12000008 e 12001252, respectivamente.

Cirurgias urgentes realizadas a pacientes da UCI em 2012

Show 10 entries		Search:
PATOLOGIA	URGENTE	
FASCEITE NECROTIZANTE	5	
COMPLICACOES DE ORGAO TRANSPLANTADO	4	
HIDROCEFALIA OBSTRUTIVA	3	
HEMOPERITONEU (NAO TRAUMATICO)	3	
PERITONITE	3	
HEMORRAGIA COMPLICANDO UM PROCEDIMENTO	2	
INFECCAO POS-OPERATORIA NCOP	2	
PERFURACAO NAO TRAUMATICA DO ESOFAGO	2	
TRANSTORNOS DO FIGADO NCOP	2	
HEMORRAGIA SUBARACNOIDEIA	2	

Showing 1 to 10 of 59 entries

Figura 5.47: Dashboard com algumas das patologias responsáveis pelas cirurgias urgentes realizadas a pacientes da UCI em 2012.

Outro aspecto muito importante que se deve analisar remete para os tempos cirúrgicos. Estes tempos estão intrinsecamente ligados aos custos hospitalares e portanto deve haver uma gestão rigorosa dos recursos utilizados: quer das salas, quer dos profissionais de saúde. Em termos de gestão, os tempos de cirurgia (segundos em que a cirurgia propriamente dita tem início) e de ocupação de sala (tem início desde que o paciente dá entrada na sala) são aqueles sobre os quais deve haver um controlo rigoroso para que não haja recursos ocupados desnecessariamente.

O software Pentaho permite fazer análises estatísticas, tendo-se feito uso dos módulos PDI para tal, e posteriormente do PRD para disponibilizar os resultados sob a forma de relatório.

Analisando o relatório da figura 5.51, verifica-se então que o tempo médio de ocupação de sala pelos 80 pacientes submetidos a cirurgia foi de aproximadamente 2,24h (8057 segundos), o tempo de anestesia foi de cerca de 2h (7177segundos) e o tempo cirúrgico foi de 1,6h (5811segundos).

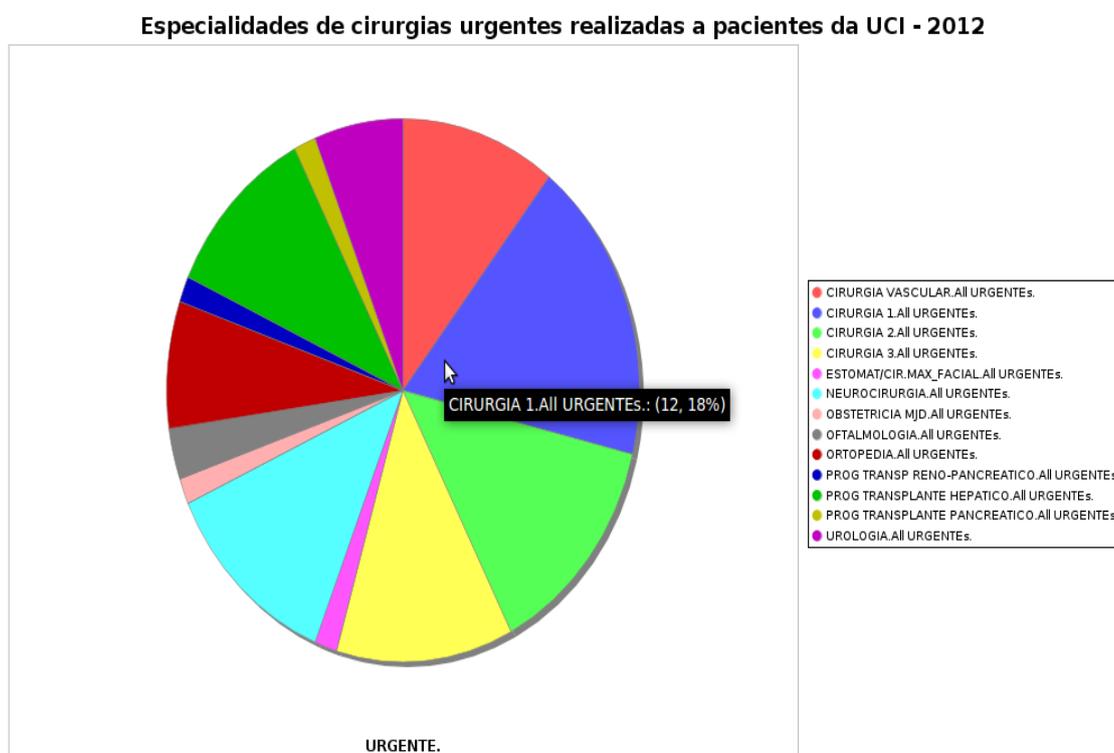


Figura 5.48: Análise qualitativa das especialidades das cirurgias com carácter urgente realizadas a pacientes da UCI, em 2012.

Tipo de cirurgias realizadas na UCI por episódio - 2012

DATA_CIRURGIA	PATOLOGIA	ESPECIALIDADE
2012-01-01 00:00:00.0	INSUFICIENCIA VASCULAR AGUDA DO INTESTINO	CIRURGIA 3

Showing 1 to 1 of 1 entries

Figura 5.49: Cirurgia referente ao episódio 12000008.

Observa-se também que o tempo máximo de ocupação de sala foi de 10,55h, enquanto o tempo máximo anestésico foi de 10,2h e de cirurgia foi 8,5h. A cirurgia que demorou mais tempo corresponde a uma cirurgia vascular, referente a uma fractura da diáfise e teve carácter urgente.

Sendo a mediana uma medida da tendência central, verifica-se que metade da amostra foi submetida a cirurgias acima das 1,27h (4560 segundos) e a outra metade abaixo deste valor.

Tipo de cirurgias realizadas na UCI por episódio - 2012

DATA_CIRURGIA	PATOLOGIA	ESPECIALIDADE
2012-01-20 00:00:00.0	HEMORRAGIA INTRACRANIANA NCOE AS NAO ESPECIFICADAS CONSEQUENTES A TRAUMATISM	NEUROCIRURGIA

Showing 1 to 1 of 1 entries

Figura 5.50: Cirurgia referente ao episódio 12001252.

Pela análise do percentil, conclui-se que 75% dos casos registraram um tempo cirúrgico abaixo das 2,20h, ou seja, 60 pacientes, enquanto 20 pacientes foram submetidos a cirurgias com tempo superior a esse valor.

A maior variação no conjunto de valores foi registada para os tempos de anestesia ($\text{stdDev}=5579$ segundos, aproximadamente 1,55h) e a menor para os tempos de cirurgia ($\text{stdDev}=4650$ segundos, aproximadamente 1,30h).

Hospital Sto. António

Dados estatísticos relativos às cirurgias realizadas a pacientes da UCI

Valores estatísticos	Tempo ocupação de sala	Tempo de anestesia	Tempo de cirurgia
N	80	80	80
Mean	8057	7177	5811
stdDev	5460	5579	4650
min	0	0	0
max	37980	36720	30600
median	6870	6180	4560
75th percentile	10800	9810	7920

Figura 5.51: Relatório com os dados estatísticos acerca dos tempos de ocupação de sala, de cirurgia e anestesia, em segundos, relativos às cirurgias dos pacientes da UCI.

Pode concluir-se que os tempos cirúrgicos das operações realizadas a pacientes da UCI, em média, não são muito elevados, sendo que em 75% dos casos têm duração inferior às 2,20h, como já referido.

Futuramente, seria de interesse analisar em detalhe os tempos de ocupação de sala e de cirurgia para perceber se existem “desperdícios” de tempo que levem ao não aproveitamento dos recursos da melhor maneira. Porém, isto não seria fácil, uma vez que há cirurgias que até podem não ter uma

duração muito elevada, mas cujo tempo de ocupação de sala exige bastante tempo, por causa da preparação do doente, por exemplo.

Processo de desenvolvimento

Para este caso de estudo foram utilizados o PDI para processos de ETL, o módulo CDE para criação dos dashboards e o Analysis View para efectuar análise OLAP. Primeiro são aqui explicados os processos de desenvolvimento dos dashboards e, posteriormente, explicam-se os gráficos obtidos com o Analysis View, terminando com a explicação dos procedimentos efectuados para a análise estatística dos tempos cirúrgicos.

Antes de mais, teve de ser criado o DW através do PDI, onde o objectivo foi extrair de uma tabela relativa a todas as cirurgias todos os dados da UCI e introduzi-los numa nova tabela, contida noutra base de dados.

Assim, para o dashboard da figura 5.42, o layout escolhido foi o “*Filter Template*”, com um *CCC Bar Chart* como componente e a seguinte query SQL:

```
select mes, sum(urgente) as Urgente, sum(reoperacao) as Reoperacao,
sum(reoperacaomenos24) as Reoperacao_Menos24h
from bicirurgiasano
group by mes
order by mes
```

O mesmo procedimento foi seguido para os restantes dashboards, com pequenas diferenças a nível de layout, componentes utilizados e com diferentes consultas feitas ao repositório.

De referir uma construção mais complexa para o dashboard que permite consultar as cirurgias por episódio (figuras 5.49 e 5.50). Foram utilizados três componentes: *table component*, *Simple Parameter* e um *Select Component*. Estes dois últimos parâmetros conjugados permitem filtrar os dados por episódio e, para tal, teve de ser configurada uma base de dados (*Data Source*) à qual é feita uma consulta SQL que vai pesquisar todos os episódio distintos existentes na tabela *bicirurgiasano*:

```
select distinct (episodio)
from bicirurgiasano
```

De seguida, foi configurada uma segunda base de dados associada ao *table component* que permite disponibilizar a data de cirurgia, a patologia e especialidade para cada episódio, especificando-se que o episódio corresponde

ao parâmetro $pEpisódio$ previamente definido, através da linha `episodio=` $\{pEpisodio\}$. A query é:

```
select datamov as Data_Cirurgia, patologia as Patologia, especialidade as
Especialidade
from bicirurgiasano
where episodio=  $\{pEpisodio\}$ 
```

A vantagem da utilização de *Table Components* é a disponibilização de um filtro de procura que permite fazer pesquisas segundo o critério que se pretende, para além de possibilitar a escolha do número de entradas que se pretende visualizar.

Relativamente ao Analysis View e à figura 5.45, o primeiro passo foi a criação do cubo dimensional sujeito à análise. Primeiro foi configurada a base de dados (bdNOVA) e atribuída uma query SQL, tal como se pode ver na figura 5.52.

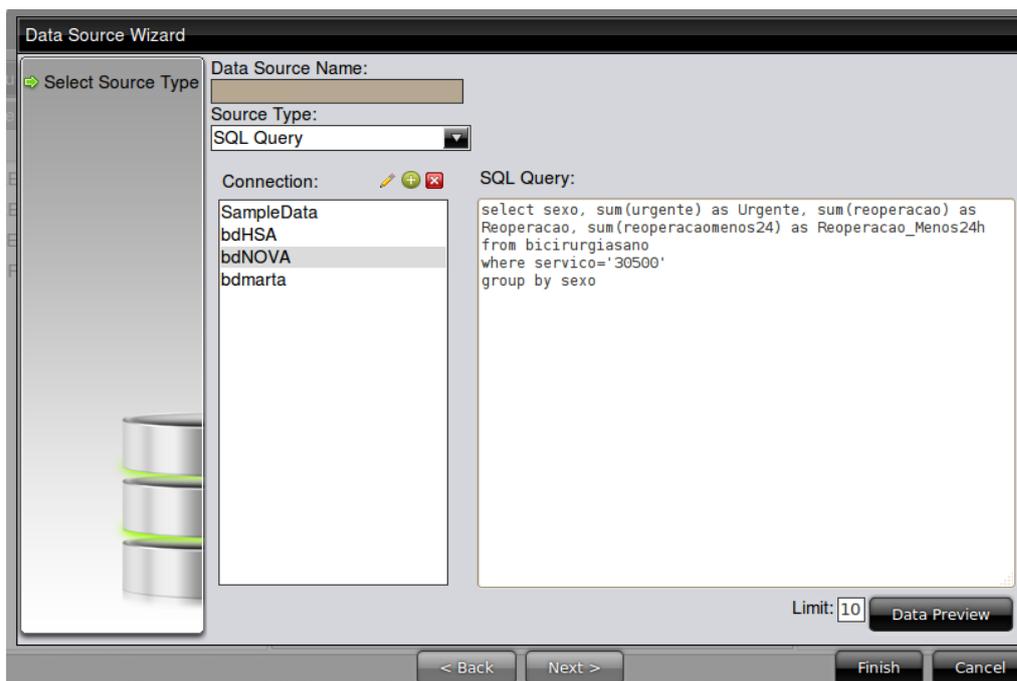


Figura 5.52: Configuração da base de dados e definição da consulta SQL para criação do cubo de análise multidimensional.

Após isto, e como se observa na figura 5.53, tem-se acesso às medidas e dimensões provenientes da base de dados. Neste passo, pode-se popular o

modelo de forma automática, com as medidas e dimensões por *default*, ou então o utilizador pode fazê-lo segundo os seus critérios.

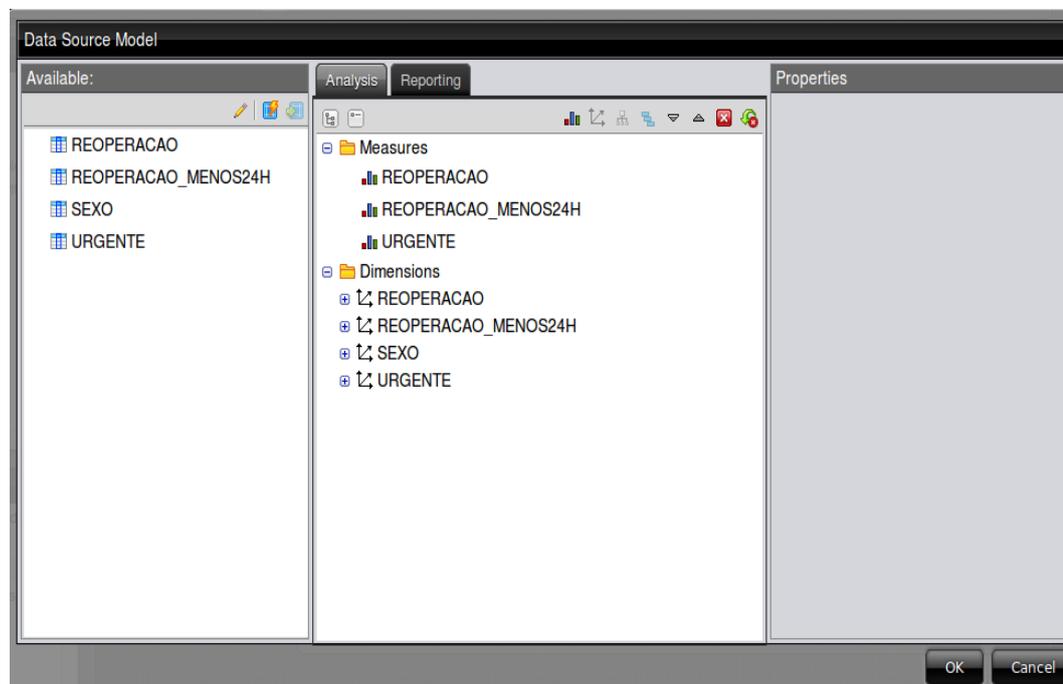


Figura 5.53: Modelo populado pelas dimensões e medidas por *default*.

Por fim, o *schema* e o cubo ficam disponíveis (figura 5.54).

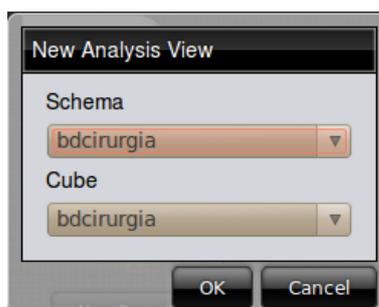


Figura 5.54: Escolha do *Schema* e *Cube* para proceder à análise multidimensional.

Passando à análise multidimensional propriamente dita, começou por se utilizar o *OLAP Navigator* para estabelecer quais as dimensões (linhas e colunas) a mostrar e também os membros filtrados, caso seja pretendido. Neste caso, as colunas são constituídas pelos atributos "Reoperação", "Reoperação_Menos24h", "Urgente" e "Sexo". As linhas são compostas pelas

measures que, neste caso, são Reoperação, Reoperação_Menos24h e Urgente, pois são estes os tipos de cirurgias que se pretende analisar. Segue-se, abaixo, a figura referente ao *OLAP Navigator* (5.55)



Figura 5.55: *OLAP Navigator*.

O resultado obtido pode ser visto através da grelha (*Data Grid*), na figura 5.56. Fez-se uso do *Drill position* e optou-se por suprimir as linhas/colunas com valores nulos.

	REOPERACAO	REOPERACAO_MENOS24H	URGENTE	SEXO
Measures	All REOPERACAOs	All REOPERACAO_MENOS24Hs	All URGENTESs	All SEXOs
REOPERACAO	44	23	21	
REOPERACAO_MENOS24H	3	3		
URGENTE	66	30	36	

Figura 5.56: *Data Grid*.

As propriedades gráficas podem ser escolhidas e editadas, de acordo com a vontade do utilizador, como se observa na figura 5.57.

O facto de não se poder alterar as legendas revelou-se desvantajoso, uma vez que a distinção entre sexos fica limitada aos números 1 e 2, o que não é esclarecedor na óptica do utilizador final. Além disso, a coluna referente aos dois sexos (barra vermelha) não é possível ocultar, o que apesar de não ser necessariamente uma desvantagem, poderá limitar o utilizador quanto à disponibilização da informação de acordo com o que ele pretende.

Chart Properties	
Chart Type	Vertical Bar 3D
Enable Drill Through	<input checked="" type="checkbox"/>
Chart Title	Tipo de cirurgias realizadas a pacientes
Chart Title Font	SansSerif Bold 18
Horizontal axis label	
Vertical axis label	
Axes Label Font	SansSerif Plain 12
Axes Tick Label font	SansSerif Plain 12 30°
Show Legend	<input checked="" type="checkbox"/> Bottom
Legend Font	SansSerif Plain 10
Show Slicer	<input checked="" type="checkbox"/> Bottom Left
Slicer Font	SansSerif Plain 12
Chart Height	600
Chart Width	1000
Background (R, G, B)	255 255 255
OK Cancel	

Figura 5.57: Propriedades gráficas.

Uma opção disponibilizada neste módulo é a exportação para Excel, como já previamente referido. Assim, é possível exportar a análise feita para esse programa, o que é extremamente vantajoso caso se pretenda fazer outro tipo de análise, cálculos, etc. Na figura 5.58 encontra-se o ficheiro .xls.

Quanto aos outros gráficos, nomeadamente os da figura 5.46 e da figura 5.48, foram seguidos os mesmos procedimentos, alterando-se alguns parâmetros, desde a consulta SQL ao tipo de gráfico utilizado para cada um deles.

Para a figura 5.46 teve-se o mesmo problema com as legendas e, além disso, como o *measures* é constituído pelo *fact count* que calcula o número de cirurgias por cada especialidade, não há distinção de cores para as barras de distintas especialidades ou para distinção do sexo, o que não é tão agradável para visualização e intuitivo para a sua compreensão.

Por último, para a criação do relatório com as análises estatísticas, utilizaram-se dois módulos: o PDI e o PRD. Inicialmente, o PDI foi utilizado para se obterem os resultados estatísticos num ficheiro de texto. A transformação criada pode ser visualizada na figura 5.59.

Definiu-se então uma tabela de entrada (*table input*) associada à base de dados previamente configurada e com a seguinte query atribuída:

```
select tempo5, tempo6, tempo7
from bicirurgiasano
```

	SEXO	1	2	
REOPERACAO		44	23	21
REOPERACAO_MENOS24H		3	3	
URGENTE		66	30	36

Figura 5.58: Excel: ficheiro .xls obtido da análise feita ao tipo de cirurgias realizadas, segundo o sexo dos pacientes.

where $servico = '30500'$

O passo seguinte foi tornar possível os cálculos estatísticos pretendidos e tal foi conseguido fazendo uso do *Univariate Statistics*. Introduziram-se os campos desejados (*tempo5*, *tempo6* e *tempo7*) e, para cada um deles, indicou-se os resultados estatísticos esperados (N, mean, std dev, min, max, median e percentile). Por fim, os resultados foram direccionados para um ficheiro de texto (.txt) – *text file output*.

Posto isto, procedeu-se à criação do relatório, através do PRD. Aqui definiu-se que os valores relativos ao "tempo5" indicam o tempo de ocupação de sala, os do "tempo6" os tempos de anestesia e os do "tempo7" são relativos aos tempos de cirurgia.

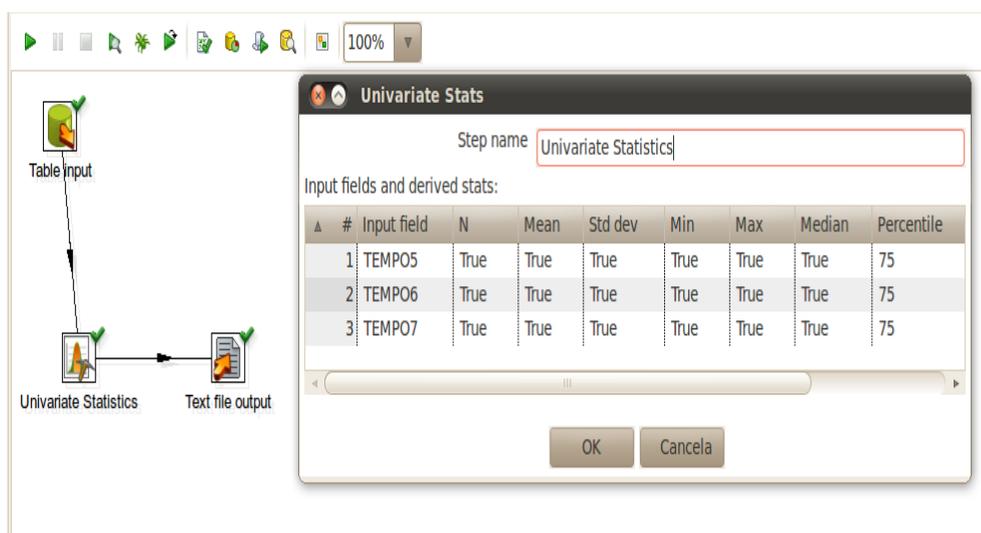


Figura 5.59: Transformação criada através do PDI para cálculos estatísticos relativos aos tempos 5, 6 e 7.

5.6 Data mining com WEKA

O **WEKA** foi o módulo do Pentaho menos explorado neste projecto, pois o processo de *data mining* é extremamente demorado e requer alguma experiência na manipulação dos dados, de tal modo que o tempo disponível não permitiu aprofundar este processo de **EC**.

Antes de tudo, foi necessário fazer uso do **PDI** para poder utilizar o **WEKA**, exportando os dados para um formato reconhecível pelo módulo, nomeadamente ficheiros **.arff** onde estão presentes uma série de informações como domínio do atributo, valores que os atributos podem representar e atributo classe, entre outros.

Assim, os dados contidos nas diferentes tabelas, nomeadamente a "bicirurgiasano", "internados" e "ocupacao", foram extraídos da base de dados e carregados para ficheiros **.arff**. Tal foi executado através de uma simples transformação **PDI**. A figura abaixo representa a extracção dos dados da tabela bicirurgiasano, tendo sido o mesmo efectuado para as restantes tabelas.

Uma vez que os restantes resultados não foram satisfatórios, apenas se vão considerar aqueles que se obtiveram fazendo uso da tabela relativa às cirurgias dos pacientes da **UCI** ("bicirurgiasano") e através da etapa de *clustering*. O dataset é constituído por 80 instâncias.

Primeiramente, procedeu-se ao pré-processamento dos dados: remoção de atributos cujos valores eram nulos e aplicação de um filtro não-supervisionado

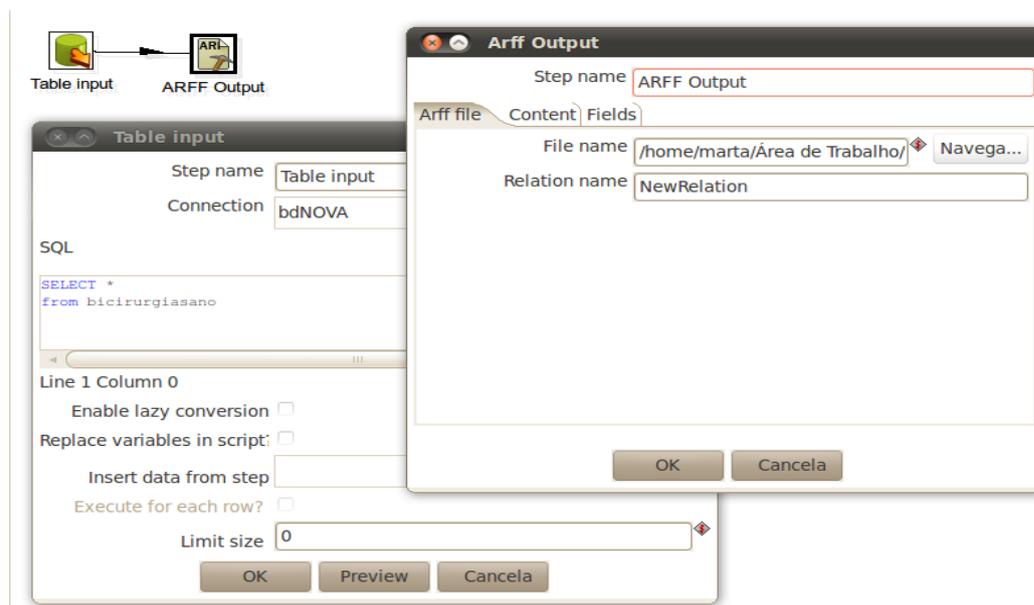


Figura 5.60: Transformação PDI para obtenção do ficheiro .arff com os dados relativos à tabela bicirurgiasano.

para que os atributos numéricos passassem a binários.

Cluster

O algoritmo utilizado foi o *SimpleKMeans* que tem como característica realizar o agrupamento de acordo com um número de centróides e é uma técnica não-supervisionada, pois fornece uma classificação automática dos dados (sem a necessidade de intervenção humana). Foi utilizada a distância Euclidiana como métrica de proximidade entre instâncias e grupos. Tomou-se em consideração o "sexo" dos pacientes como classe para avaliação do cluster (*class to cluster evaluation*) e, em 72 atributos, apenas se tiveram em conta "patologia", "programada", "urgente", "reoperacao" e "reoperacaomais24", tendo os restantes sido ignorados.

Assim, como se observa na figura 5.61, obtiveram-se 2 clusters em 0.01 segundos, após 3 iterações, verificando-se a existência de 1 cluster estatisticamente mais relevante que outro, representando ele 88% do dataset, ou seja, 70 dos 80 pacientes submetidos a cirurgias. Neste caso, foram agrupadas incorrectamente cerca de 45% das instâncias.

```

Time taken to build model (full training data) : 0.01 seconds          Full Data          0
                                                                    (80)                (10)
=== Model and evaluation on training set =====
Clustered Instances          COMPLICACOES DE ORGAO TRANSPLANTADO  ATEROSCLEROSE DE ARTERIAS ESPECIFICADAS  NCOP          COMPLICACOES DE ORGAO TRAN
0          10 ( 13%)          0          1          0
1          70 ( 88%)          1          1          0
                                                                    0          0
Class attribute: SEXO
Classes to Clusters:
  0 1 <-- assigned to cluster
  8 34 | 1
  2 36 | 2
Cluster 0 <-- 1
Cluster 1 <-- 2
Incorrectly clustered instances :          36.0          45          %

```

Figura 5.61: Clusters obtidos através da aplicação do algoritmo *SimpleKMeans*, considerando como classe o atributo "sexo".

O cluster 0 (menos relevante), constituído por 10 pacientes, é composto por 8 pacientes do sexo masculino e 2 do sexo feminino, enquanto ao cluster 1, com um total de 70 pacientes, foram atribuídos 34 pacientes do sexo masculino e 36 do sexo feminino.

Atendendo ao cluster 1, a patologia referente às cirurgias corresponde a complicações de órgão transplantados, tendo as mesmas carácter urgente e sendo relativas a reoperações. O cluster composto por apenas 10 pacientes refere-se a cirurgias programadas a Aterosclerose de artérias especificadas.

Seguidamente e continuando com o mesmo pré-processamento, mas atendendo aos atributos "patologia", "programada", "urgente", "reoperacao", "reoperacaomais24", "especialidade" e "sexo", e aplicando novamente o al-

goritmo *SimpleKMeans* a todo o dataset, obtiveram-se 2 clusters, após 3 iterações: o cluster 0 constituído por 56 pacientes e o cluster 1 por 24 pacientes.

Cluster centroids:			
Attribute	Full Data (80)	Cluster#	
		0 (56)	1 (24)
PATOLOGIA		COMPLICACOES DE ORGAO TRANSPLANTADO	FASCEITE NECROTIZANTE
PROGRAMADA_binarized	0	0	0
URGENTE_binarized	1	1	1
REOPERACAO_binarized	1	1	0
REOPERACAOAIS24_binarized	0	0	0
ESPECIALIDADE	CIRURGIA 1	CIRURGIA VASCULAR	CIRURGIA 1
SEXO	1	1	2

Figura 5.62: Clusters obtidos através da aplicação do algoritmo *SimpleKMeans*.

Atendendo primeiro ao cluster 1 (ver figura 5.62) que representa 30% do dataset, verifica-se que é composto por pacientes do sexo feminino submetidos a cirurgias urgentes, da especialidade Cirurgia1 e patologia fasceíte necrotizante.

Quanto ao cluster 0, com mais peso estatístico que o cluster 1, e representado por pacientes do sexo masculino, conclui-se que as cirurgias realizadas são urgentes e correspondem a reoperações de cirurgias vasculares e derivadas de complicações de órgãos transplantados.

Ora, analisando estes resultados conjuntamente com os anteriores, verifica-se que, de facto, as cirurgias vasculares apenas foram realizadas a homens. A tendência aqui encontrada é de que estas cirurgias resultam de complicações de órgãos transplantados.

Além disso, estes resultados corroboram o que previamente já foi analisado na secção 5.5, pois as mulheres foram submetidas a mais cirurgias urgentes que os homens e sabe-se que a patologia mais representativa dessas cirurgias é a fasceíte necrotizante.

Capítulo 6

Discussão dos resultados

6.1 Resultados obtidos

Durante a exploração do software, grande parte do tempo foi dedicado a perceber o que era possível fazer com cada módulo, a aprender a trabalhar e dominar os seus componentes para posteriormente se extrair conhecimento útil dos dados utilizados. Os resultados obtidos nos diferentes casos de estudo e o conhecimento deles extraído acabam por, futuramente, não ter um papel preventivo na UCI, uma vez que os dados disponíveis permitiram fazer análises retrospectivas, mas sem grande impacto caso se pretenda actuar futuramente, pois em nenhum dos casos de estudo foram detectados padrões relevantes. Ainda assim, considera-se que as análises elaboradas são necessárias e importantes de se realizar em ambiente hospitalar.

Além disso, o software revelou-se perfeitamente capaz e eficiente na realização das diversas etapas de BI, desde os processos de ETL à representação e visualização dos resultados. Mostrou, desta maneira, ser uma solução flexível para a adaptação e implementação de BI.

6.2 Manipulação do software

Atendendo à sua usabilidade e considerando todos os módulos analisados, o software apresenta interfaces agradáveis e atractivas ao utilizador; a qualidade de uso do mesmo é elevada, pois foi possível atingir as metas com eficiência, apesar de, obviamente, requerer conhecimentos na manipulação das ferramentas para que se consiga produzir os resultados a disponibilizar ao utilizador final. Aqui surge o conceito de facilidade de aprendizado: considera-se que o software tem um nível de uso intermediário, pois perceber como interagir com o sistema e entender a utilização das diferentes porções com

determinado nível de competência e desempenho foi o que consumiu mais tempo neste projecto; revela facilidade de memorização, onde operações anteriores são fáceis de lembrar e ser novamente executadas; é produtivo na obtenção de resultados, ou seja, revela-se satisfatório, pois permite ao utilizador obter aquilo que pretende; é flexível e útil, pois dispõe de diversas ferramentas que permitem realizar as mais variadas tarefas; é seguro, na medida em que o utilizador tem acesso à plataforma Bi-server através de um login e pode controlar quem acede aos seus dados. Também o PDI e o PRD, para se aceder às bases de dados exigem passwords, porém o mesmo não se verifica no WEKA, o que é uma questão a ter em conta, uma vez que se está a lidar com dados sensíveis relativos a pacientes.

6.2.1 Pentaho BI-server

Esta plataforma apresenta uma interface agradável e bastante fácil de manipular. Dado que oferece diversas ferramentas, como elaboração de relatórios, dashboards, análise OLAP, considera-se que é organizada, na medida em que o software separa bem estas diferentes funcionalidades, mas em simultâneo é possível a sua integração.

Durante a manipulação da plataforma, detectaram-se alguns aspectos menos positivos, como o facto de ao se fazer logout ou fechar determinada aplicação, o utilizador não é questionado se pretende gravar o trabalho até ali desenvolvido. Isto é desvantajoso, pois caso acidentalmente se pressione o ícone de fecho, o trabalho é perdido e impossível de recuperar.

6.2.2 PDI

Este módulo exige ter o objectivo final bem presente, pois é necessário saber exactamente o que se pretende obter e qual o caminho a seguir. Esta foi uma ferramenta que mostrou ser extremamente poderosa e rica para se proceder às etapas de ETL, com extenso conjunto de recursos e amplo suporte a bases de dados. Durante a realização do projecto, apesar de se tentar utilizar o maior número de *steps* disponíveis, não foi possível, de todo, experimentá-los todos, dado o tipo de dados utilizados e os objectivos que se tinha.

A grande vantagem encontrada remete para os aspectos gráficos concedidos às transformações criadas, conseguido pela utilização de ícones gráficos, o que permite uma percepção mais rápida e fácil do processo que se está a implementar, sem se escrever uma única linha de código. Outro factor positivo é a disponibilização dos erros, caso estes ocorram, através da consola de execução.

Verificou-se que, de facto, o ETL é um processo que demora bastante tempo, cujas transformações exigem ser pensadas por forma a se conseguir obter aquilo que se pretende, podendo até, tornar-se bastante complexo.

Porém, houve aspectos do PDI que ficaram por compreender, como se verificou no caso de estudo dos reinternamentos (explicado na secção 5.4), aquando da utilização do *step* "Merge JOIN", não se entendendo o porquê de não se terem obtido os resultados esperados.

6.2.3 Analysis View

O Analysis View mostrou ser uma ferramenta capaz de atender às necessidades do utilizador relativamente à análise multidimensional dos dados. Apresenta ferramentas muito úteis e que concedem uma análise rica e completa dos dados, das mais diversas formas, consoante a necessidade do utilizador, ou seja, destaca-se a flexibilidade concedida durante a análise que, através do cruzamento de dados, permite obter conhecimento aprofundado acerca dos mesmos e, desta forma, atingir conhecimento útil.

Inicialmente, para que os conceitos inerentes à análise OLAP fossem assimilados, foi necessário fazer uma pesquisa aprofundada, mas após o seu entendimento, a ferramenta torna-se bastante fácil de utilizar.

Permite a exportação para Excel o que se torna vantajoso caso se pretenda efectuar outro tipo de cálculos. Porém, tal como no caso do CDE abordado de seguida, não permite a exportação para formato pdf. Uma outra desvantagem encontrada remete para a impossibilidade de alterar as legendas dos gráficos, o que limita a representação dos dados de acordo com o que o utilizador pretende.

6.2.4 CDE

O CDE revelou ser uma ferramenta muito útil na disponibilização gráfica dos resultados, pois permite a criação de dashboards extremamente atractivos para visualização de dados, com grande variedade de componentes visuais. Para obter aquilo que pretende, o utilizador dispõe de variados caminhos, ou seja, há flexibilidade de implementação, podendo fazer uso de html, css, javascript, etc. Mais uma vez, dada a grande variedade de componentes a utilizar, não foi possível a utilização de todos eles, sendo dado ênfase aos gráficos circulares, de barras e tabelas.

De todos os módulos analisados, este foi aquele que se revelou ser mais fácil de manipular, bastante intuitivo. A maior dificuldade foi perceber como configurar as base de dados. Porém, houve aspectos que só após um conhecimento mais aprofundado da ferramenta foram ficando claros, nomeadamente

nas definições avançadas dos componentes utilizados ou do layout.

Neste módulo encontraram-se alguns *bugs* já referidos nos casos de estudo, nomeadamente aquando da utilização de gráficos circulares (discrepância entre a query e os dados disponibilizados no gráfico) e apresentação dos valores nos gráficos. Além disso, não é possível exportar os dashboards para pdf, html, ou outro formato, o que implica que a sua visualização apenas seja possível na plataforma, o que nem sempre é conveniente, pois implicou fazer *print screens* aos dashboards, o que não é prático.

6.2.5 PRD

O PRD permite a elaboração livre de relatórios (analíticos, financeiros, operacionais), através do acesso a dados de diferentes tipos de fontes (XML, OLAP,...), permitindo representar a informação segundo os critérios desejados (os utilizadores especificam exactamente onde cada elemento do relatório será exibido, por exemplo), bem como apresenta inúmeras opções ao utilizador, desde utilização de funções (somas, médias, etc) até utilização de imagens e criação de gráficos que podem ser incorporados nos relatórios. Além disso, é possível a utilização do Report Wizard que se revelou um meio bastante eficaz e rápido de se representar a informação de forma agradável para disponibilizar ao utilizador final, estando mais voltado para a análise de custos. Para iniciantes, é o mais recomendado.

O facto de ser não ser uma ferramenta WYSIWYG exige que o utilizador esteja bastante familiarizado com a plataforma e as suas funcionalidades. A manipulação da ferramenta exige algum tempo de aprendizagem, dada a necessidade de explorar todas as características e funções disponibilizadas. Uma desvantagem está na definição da consulta SQL que permitirá obter os dados armazenados no DW, pois não é possível a utilização de mais do que uma consulta por relatório. Desta maneira, caso se pretenda elaborar um relatório com diversos tipos de dados, cujo acesso à base de dados só seja possível através de consultas independentes, isto não é possível.

6.2.6 Weka

A EC através do WEKA exigiu imenso tempo e, no final, os resultados obtidos não foram satisfatórios, apesar da ferramenta ser extremamente intuitiva e de fácil utilização. Consideram-se não satisfatórios, na medida em que não permitiu extrair padrões e tendências ocultas nos dados que permitiriam, eventualmente, a tomada de decisões a nível preventivo.

O processo em questão revelou ser não trivial, exigindo perícia e experiência por parte do utilizador, quer no pré-processamento dos dados, quer

no conhecimento e aplicação dos algoritmos disponíveis. Além disso, o utilizador deve ter os objectivos acerca do conhecimento que pretende obter bem definidos e clarificados previamente, para saber exactamente o percurso a percorrer e, desta forma, poupar tempo, visto que já por si, este processo é muito demorado.

Optou-se por apenas disponibilizar resultados que se consideraram minimamente satisfatórios, mas deve ter-se em conta que o *dataset* apenas é constituído por 80 instâncias, o que poderá ser pouco significativo num universo hospitalar em que se consideram milhares de registos diários. Mesmo tendo em conta apenas a UCI, detectar tendências para posterior acção preventiva baseando-se num número tão reduzido de pacientes, e considerar apenas um período de três meses, não é eficaz.

De qualquer maneira, o processo de *data mining* desenvolvido com a ferramenta WEKA é interessante e útil, na medida em que o utilizador consegue efectuar todas as etapas, com inúmeros algoritmos disponíveis, e suportado por uma interface gráfica extremamente intuitiva e agradável. Porém, a memória *default* disponível revela-se insuficiente para grandes bases de dados.

Capítulo 7

Conclusões e trabalho futuro

Actualmente é imprescindível a adopção, por parte das instituições de saúde, de mecanismos tecnológicos automatizados para o tratamento dos dados. A utilização de sistemas de BI permite ter a informação correcta no menor tempo possível, permitindo tomar decisões eficazes, corrigir processos, antecipar-se às necessidades dos pacientes.

O software Pentaho mostrou ser extremamente eficaz no processo de EC, disponibilizando ferramentas poderosas que auxiliam a tomada de decisão. Este software analisa, sintetiza, assimila e dá sentido às enormes quantidades de informação, sendo capaz de estabelecer ligações sofisticadas e discernir padrões, dando oportunidade para tirar conclusões e agir em tempo real. A utilização das ferramentas, em conjunto ou em separado, potenciam a capacidade humana para a tomada de decisão. Todo o processo de BI foi desempenhado com sucesso através da utilização dos diferentes módulos do Pentaho: desde os processos de ETL com o PDI, análise OLAP com o Analysis View, visualização e representação das informações através do CDE e do PRD, respectivamente. No entanto, a inserção de ferramentas de BI exige formação específica e intensiva dada ao utilizador, o que nem sempre facilita o processo de implementação.

Do ponto de vista informático, pode-se concluir que os módulos, no geral, apresentam dificuldade intermédia na sua manipulação e evidenciam usabilidade. Os dados utilizados eram referentes à UCI, mas sendo aplicado a outras unidades, as conclusões não divergiram, pois a facultação de meios para obter gráficos, dashboards, entre outros, faz com que este software apresente diversas áreas de aplicação. Porém, foi constatado que o Pentaho encontra-se maioritariamente direccionado para a área empresarial, ou seja, para lidar com dados monetários, com custos.

Como seria de esperar, dado o facto de ter sido utilizada uma versão OS, foram encontrados *bugs* que ainda precisam de ser corrigidos. De referir que

a comunidade responsável pelo desenvolvimento dos produtos deste software é extremamente activa e participativa, destacando-se a sua disponibilidade e cooperação nos fóruns, facilitando a resolução de problemas.

A aplicação de um software OS de BI às instituições de saúde não iria ser um processo burocraticamente fácil, uma vez que a regulamentação legal encontra-se atrasada em relação à tecnologia. Incorrer-se-ia em riscos de manipulações, o que implicaria a salvaguarda da segurança dos dados dos pacientes ao máximo.

Concluindo, os objectivos deste projecto foram atingidos e o uso do software OS de BI Pentaho como ferramenta de auxílio na tomada de decisão dentro da UCI torna-se viável e um diferencial na obtenção de conhecimento útil.

Futuramente, seria interessante comparar o Pentaho a outro(s) software de BI existente(s), também OS, para entender as vantagens e desvantagens das ferramentas por ele disponibilizadas, comparativamente às funcionalidades oferecidas por outros software vingentes no mercado.

Além disso, apesar de se ter conseguido manipular os módulos constituintes do Pentaho BI Suite eficazmente, dado que não houve tempo disponível para explorar todas as funcionalidades de cada um deles, seria vantajoso fazê-lo no futuro.

Avaliar o impacto da utilização do software junto de profissionais de saúde que dessem o seu parecer acerca do conhecimento obtido seria extremamente útil para perceber o real impacto da aplicação do Pentaho nas instituições de saúde.

Por fim, proceder à aplicação propriamente dita do Pentaho no hospital: criação de um repositório de variáveis clínicas, de forma automática, através de agentes, com armazenamento em DWs, para monitorização constante e permanente com as ferramentas disponibilizadas pelo software.

Referências

- [1] S. Chaudhuri, U. Dayal, and V. Narasayya, “An overview of business intelligence technology,” *Commun. ACM*, vol. 54, pp. 88–98, Aug. 2011.
- [2] S. H. A. El-Sappagh, A. M. A. Hendawi, and A. H. E. Bastawissy, “A proposed model for data warehouse etl processes,” *Journal of King Saud University - Computer and Information Sciences*, vol. 23, no. 2, pp. 91 – 104, 2011.
- [3] F. et al, *Etapas do processo KDD*, 1996 (accessed July 23, 2012). <http://litolima.files.wordpress.com/2010/11/utilbi1.jpg>.
- [4] A. B. Rumi, “An implementation of pentaho in reporting management module,” (Malaysia), Centre for Advanced Software Engineering- Universiti Teknologi Malaysia, 2008.
- [5] J. Greenbaum, “Realizing the pentaho agile bi opportunity: Bi for the masses and customer success,” 2010.
- [6] *Cubo representativo da análise multidimensional*, (accessed July 27, 2012). <http://www.scielo.br/img/revistas/gp/v13n2/31177f1.gif>.
- [7] J. Machado, V. Alves, A. Abelha, and J. Neves, “Ambient intelligence via multiagent systems in the medical arena,” *Engineering Intelligent Systems for Electrical Engineering and Communications*, vol. 15, no. 3, pp. 151–157, 2007.
- [8] S. E. Reed, D. Y. Na, T. C. Mayo, L. W. Shapiro, J. B. Duty, J. H. Conklin, and D. E. Brown, “Implementing and analyzing a data mart for the arlington county initiative to manage domestic violence offenders,” in *Proceedings of the 2010 IEEE Systems and Information Engineering Design Symposium*, (Chalottesville, VA, USA), University of Virginia, 2010.

- [9] F. Portela, P. Gago, M. F. Santos, Á. M. Silva, F. Rua, J. Machado, A. Abelha, and J. Neves, “Knowledge discovery for pervasive and real-time intelligent decision support in intensive care medicine,” in *KMIS*, pp. 241–249, 2011.
- [10] P. M. da Saúde. Direcção-Geral da Saúde, *Cuidados Intensivos Recomendações para o seu desenvolvimento*. 2002.
- [11] A. J., “Fair allocation of intensive care unit resources,” *American Thoracic Society - MEDICAL SECTION OF THE AMERICAN LUNG ASSOCIATION*, vol. 156, no. 4, pp. 1282–1301, 1997.
- [12] F. Aleixo, *Unidade de Cuidados Intensivos – Manual de Enfermagem – CHBA*. 2007.
- [13] M. Ghazanfari, M. Jafari, and S. Rouhani, “A tool to evaluate the business intelligence of enterprise systems,” *Scientia Iranica*, vol. 18, no. 6, pp. 1579 – 1590, 2011.
- [14] M. F. Santos, F. Portela, M. Vilas-Boas, J. Machado, A. Abelha, J. Neves, A. Silva, and F. Rua, “Information modeling for real-time decision support in intensive medicine,” *Proceedings of the 8th Wseas International Conference on Applied Computer and Applied Computational Science*, pp. 360–365, 2009. Chen, SY Li, Q 8th WSEAS International Conference on Applied Computer and Applied Computational Science MAY 20-22, 2009 Hangzhou, PEOPLES R CHINA.
- [15] R. Bose, “Knowledge management-enabled health care management systems: capabilities, infrastructure, and decision-support,” *Expert Systems with Applications*, vol. 24, no. 1, pp. 59 – 71, 2003.
- [16] T. A. Spil, R. A. Stegwee, and C. J. Teitink, “Business intelligence in healthcare organizations,” in *Proceedings of the 35th Annual Hawaii International Conference on System Sciences, HICSS*, IEEE, 2002.
- [17] T. P. Clemmer, “Computers in the icu: Where we started and where we are now,” *Journal of Critical Care*, vol. 19, no. 4, pp. 201 – 207, 2004.
- [18] J. Ferreira, M. Miranda, A. Abelha, and J. Machado, “O processo etl em sistemas data warehouse,” in *INForum 2010 – II Simpósio de Informática* (L. S. Barbosa and M. P. Correia, eds.), pp. 757–765, 2010.
- [19] S. T. March and A. R. Hevner, “Integrated decision support systems: A data warehousing perspective,” *Decision Support Systems*, vol. 43, no. 3, pp. 1031 – 1043, 2007.

- [20] S. Sharma, K.-M. Osei-Bryson, and G. M. Kasper, “Evaluation of an integrated knowledge discovery and dataminingprocess model,” *Elsevier Expert Systems with Applications*, vol. 39, no. 13, pp. 11335–11348, 2012.
- [21] T. Thalhammer, M. Schrefl, and M. Mohania, “Active data warehouses: complementing olap with analysis rules,” *Data & Knowledge Engineering*, vol. 39, no. 3, pp. 241 – 269, 2001.
- [22] C. Gameiro, “Implementation of business intelligence tools using open source approach - master in open source software,” (Lisbon, Portugal), ISCTE, 2011.
- [23] M. Mul, P. Alons, P. van der Velde, I. Konings, J. Bakker, and J. Hazelzet, “Development of a clinical data warehouse from an intensive care clinical information system,” *Computer Methods and Programs in Biomedicine*, vol. 105, no. 1, pp. 22 – 30, 2012.
- [24] A. G. Júnior and C. C. Bernardino, “Proposta de um sistema de business intelligence para exploração de indicadores de gerência de redes,” (Brasília), Faculdade de Tecnologia, Universidade de Brasília, 2009.
- [25] D. Q. van Ufford, “Business intelligence – the umbrella term,” (Amsterdam), Universiteit Amsterdam, Faculteit der Exacte Wetenschappen Studierichting Bedrijfskunde & Informatica, 2002.
- [26] A. F. Silva, “Business intelligence®: auxílio na tomada de decisão,” (Brasília), Departamento de Administração, Faculdade de Economia, Administração e Contabilidade, Universidade de Brasília, 2010.
- [27] Álvaro Rebuge and D. R. Ferreira, “Business process analysis in health-care environments: A methodology based on process mining,” *Information Systems*, vol. 37, no. 2, pp. 99 – 116, 2012. <ce:title>Management and Engineering of Process-Aware Information Systems</ce:title>.
- [28] L. Soibelman and H. Kim, “Data preparation process for construction knowledge generation through knowledge discovery in databases,” *Journal of Computing in Civil Engineering*, vol. 16, no. 1, pp. 39 – 48, 2002.
- [29] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “The kdd process for extracting useful knowledge from volumes of data,” *Commun. ACM*, vol. 39, pp. 27–34, Nov. 1996.
- [30] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “From data mining to knowledge discovery in databases,” *AI Magazine - American Association for Artificial Intelligence*, vol. 17, no. 3, pp. 37–54, 1996.

- [31] S. Hui and G. Jha, “Data mining for customer service support,” *Information & Management*, vol. 38, no. 1, pp. 1 – 13, 2000.
- [32] A. S. Fialho, F. Cismondi, S. M. Vieira, S. R. Reti, J. M. C. Sousa, and S. N. Finkelstein, “Data mining using clinical physiology at discharge to predict icu readmissions,” *Expert Syst. Appl.*, vol. 39, pp. 13158–13165, Dec. 2012.
- [33] A. J. C. amd J. A. Cook, G. Adey, and B. H. Cuthbertson, “Predicting death and readmission after intensive care discharge,” *British Journal of Anaesthesia*, vol. 100, no. 5, pp. 656–662, 2008.
- [34] C. J. McDonald, G. Schadow, M. Barnes, P. Dexter, J. Overhage, B. Mamlin, and J. McCoy, “Open source software in medical informatics—why, how and what,” *International Journal of Medical Informatics*, vol. 69, no. 2–3, pp. 175 – 184, 2003. <ce:title>Working Conference on Health Information Systems</ce:title>.
- [35] Care2x, “Care2x - the open source hospital information system,” 2010.
- [36] B. Janamanchi, E. Katsamakos, W. Raghupathi, and W. Gao, “The state and profile of open source software projects in health and medical informatics,” *International Journal of Medical Informatics*, vol. 78, no. 7, pp. 457 – 472, 2009.
- [37] M. Goulde and E. Brown, *Open Source Software: A Primer for Health Care Leaders*. ihealthreports: California Healthcare Foundation, 2006.
- [38] T. Waring and P. Maddocks, “Open source software implementation in the uk public sector: Evidence from the field and implications for the future,” *International Journal of Information Management*, vol. 25, no. 5, pp. 411 – 428, 2005.
- [39] E. Damiani, F. Frati, and C. Monteverdi, “Open source bi adoption,” (Italy), Department of Information Technology, University of Milan, 2009.
- [40] G. Iapichino, D. Radrizzani, A. Pezzi, E. Assi, P. D. Mauro, G. Mistraletti, and F. Porta, “Evaluating daily nursing use and needs in the intensive care unit: a method to assess the rate and appropriateness of icu resource use,” *Health Policy*, vol. 73, no. 2, pp. 228 – 234, 2005.
- [41] J. Zimmerman, A. A. Kramer, D. S. McNair, F. M. Malila, and V. L. Shaffer, “Intensive care unit length of stay: Benchmarking based on

acute physiology and chronic health evaluation (apache) iv,” *Critical Care Medicine*, vol. 34, pp. 2517–2529, Oct. 2006.

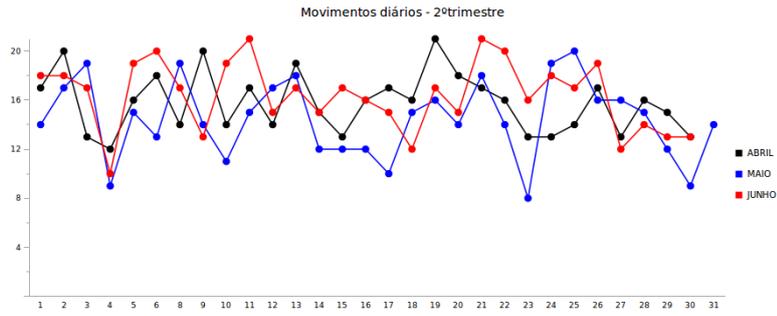
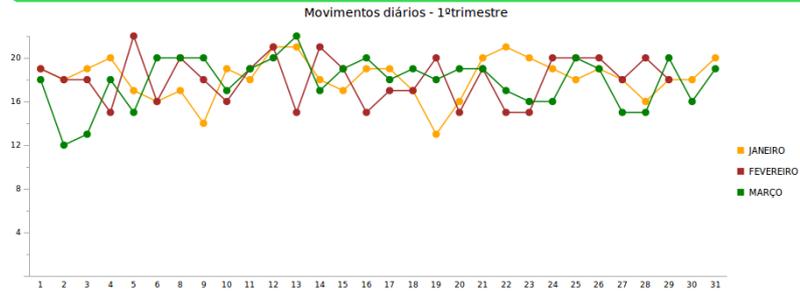
Apêndice A

Relatório semestral dos movimentos da UCI

Análise semestral dos movimentos registados - Ano 2012

Dia	Janeiro	Fevereiro	MARÇO	Abril	Maio	Junho
1	19	19	18	17	14	18
2	18	18	12	20	17	18
3	19	18	13	13	19	17
4	20	15	18	12	9	10
5	17	22	15	16	15	19
6	16	16	20	18	13	20
8	17	20	20	14	19	17
9	14	18	20	20	14	13
10	19	16	17	14	11	19
11	18	19	19	17	15	21
12	21	21	20	14	17	15
13	21	15	22	19	18	17
14	18	21	17	15	12	15
15	17	19	19	13	12	17
16	19	15	20	16	12	16
17	19	17	18	17	10	15
18	17	17	19	16	15	12
19	13	20	18	21	16	17
20	16	15	19	18	14	15
21	20	19	19	17	18	21
22	21	15	17	16	14	20
23	20	15	16	13	8	16
24	19	20	16	13	19	18
25	18	20	20	14	20	17
26	19	20	19	17	16	19
27	18	18	15	13	16	12
28	16	20	15	16	15	14
29	18	18	20	15	12	13
30	18		16	13	9	13
31	20		19		14	
TOTAL MOVIMENTOS	545	506	536	457	433	474

Análise semestral dos movimentos diários registados na UCI - Ano 2012



Hospital Sto.António - UCI