



NOVA

IMS

Information
Management
School

MGI

Mestrado em Gestão de Informação

Master Program in Information Management

Saúde em Portugal: Estudo das Urgências Hospitalares através do Data Mining

Daniela da Silva Alves

Dissertação apresentada como requisito parcial para
obtenção do grau de Mestre em Gestão de Informação

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa



NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

SAÚDE EM PORTUGAL: ESTUDO DAS URGÊNCIAS HOSPITALARES ATRAVÉS DO DATA MINING

por

Daniela da Silva Alves

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Gestão de Informação, Especialização em Gestão do Conhecimento e Business Intelligence

Orientador: Professor Doutor Roberto Henriques

Coorientador: Professor Doutor José Carlos Caiado

Novembro 2015

AGRADECIMENTOS

Em primeiro lugar, quero agradecer à minha família. Sem eles nada disto era possível. Um grande obrigado à minha mãe por todo o apoio, por tudo o que fez por mim; ao meu pai pela força que me deu, as palavras de conforto e pelo abraço no momento certo e ao meu irmão pela ajuda, pelo companheirismo e principalmente por todas as brincadeiras.

Quero agradecer também ao meu namorado, por ser a pessoa que é. Por todos os momentos que passamos e por tudo o que estará para vir. Obrigado pelo amor, carinho e confiança.

Aos meus amigos que estiveram sempre comigo. Um agradecimento especial à Ana, Sandra e Neuza por me apoiarem principalmente nesta fase final. Quando tudo parecia mais complicado, elas estavam lá para me ouvirem e apoiarem. Obrigado meninas, agora já tenho mais tempo para irmos passear!

Obrigada à Accenture, pelas pessoas que conheci e que me deram sempre um apoio especial nesta fase (Mariana e Inês, muito obrigada!)

Para terminar, agradeço também ao meu orientador e co-orientador, por ser uma aluna tão persistente, tirando-lhes muito do seu tempo ao longo deste ano.

RESUMO

Qualquer assunto relacionado com a saúde é sempre um tema sensível, pela importância que tem junto da população, já que interage diretamente com o bem-estar das pessoas e, essencialmente, com a sensação de segurança que as estas pretendem ter na prestação dos cuidados básicos de saúde. Dados estatísticos mostram que a população está cada vez mais envelhecida, reforçando a importância da existência de bons centros hospitalares e de um bom Sistema Nacional de Saúde (SNS) (Plano Nacional de Saúde, 2010). Em Portugal, caso os pacientes necessitem de cuidados mais urgentes, podem recorrer ao Serviço de Urgências disponibilizado para toda a população através do SNS. No entanto, a gestão e planeamento deste serviço é complexa, dado este serviço ser frequentemente utilizado por pacientes que não necessitam de cuidados urgentes, levando a que os hospitais deixem de conseguir dar a resposta esperada, implicando a prestação por vezes um serviço de menor qualidade.

Neste sentido, analisaram-se dados de um hospital do norte do país com o intuito de perceber o ponto de situação das urgências, de forma a encontrar padrões relevantes através da análise de *clusters* e de regras de associação. Começando pela análise de *clusters*, utilizaram-se apenas as variáveis que foram consideradas importantes para o problema, resultando da análise final 3 *clusters*. O primeiro *cluster* é constituído por elementos do sexo masculino de todas as idades, o segundo *cluster* por elementos do sexo masculino mais jovens e por elementos do sexo feminino até aos 60 anos e o terceiro *cluster* apenas por elementos do sexo feminino a partir dos 40 anos.

No final verificaram-se muitas semelhanças entre os *clusters* 1 e 3, pois ambos continham os pacientes mais idosos, havendo um padrão comum no seu comportamento. No ano 2012 não houve registo de nenhuma epidemia, não havendo por isso nenhuma doença que se destacasse comparativamente às restantes. Concluiu-se também que na maior parte dos casos houve a necessidade de uma intervenção urgente (pulseira de cor Amarela), no entanto a maioria dos pacientes observados conseguiu regressar às suas habitações após as consultas nas Urgências Hospitalares, sem intervenções médicas adicionais. Relativamente às regras de associação, houve a necessidade de transformar e eliminar algumas variáveis que enviesassem o estudo. Após o processo da criação das regras de associação, percebeu-se que as regras eram muito similares entre si, apresentando uma maior confiança nas variáveis que apareceram em maior número (“Pacientes com pulseira de cor Amarela”, “distrito do Porto” ou “Alta Médica para a Residência”).

PALAVRAS-CHAVE

Saúde; Urgências; Análise de *Clusters*; Regras de Associação

ABSTRACT

Health is a sensitive subject for its importance within population. This subject is usually related not only with the welfare of a given population but also with the sense of security that the population requires in the provision of basic health care. Statistics show us that the population is aging, supporting the idea that better hospitals are needed, as well as a better National Health Plan System. If the patients require urgent care, they can benefit from urgent care service in the hospital facility, usually called "Emergency Room". However, this service demands for a higher level of management to deal with the increasing number of patients. The use of this urgent care service in situations of urgency is not real, affects the hospital operation. This fact creates a higher affluence, which will slow down the answer and the service may be blocked. Considering all these issues and using a dataset from an Hospital located in the North of Portugal, we tried to manage and to find patterns in the information related with the urgent care service. We used Data Mining techniques including clustering and association rules. For clustering, three clusters were obtained. One of the clusters was composed by all male patients from 0 to 110 years old group, the second by female patients from 0 to 60 years old, and finally a cluster with all female patients over 60 years old. Clusters containing older people showed several similarities respecting the diseases. In 2012 there was no record of any epidemic crisis. It has also concluded that most of the cases had required urgent intervention (yellow bracelet) but despite the urgency, the majority was able to return to their residences on the same day. Those results highlight the fact that most patients did not require any other additional medical support after leaving the Emergency Room. Related with the association rules and, the rules with higher confidence were composed by variables that appeared very often ("Patients with Yellow bracelet", "District of Porto" or "Residence Medical Release").

KEYWORDS

Health; Emergency; Cluster Analysis; Association Rules

ÍNDICE

1. Introdução	1
2. Revisão da Literatura	3
2.1. Saúde e Urgências.....	3
2.2. Data Mining	5
2.3. Data Mining na Saúde	13
3. Metodologia	15
3.1. Sample.....	16
3.2. Explore.....	17
3.3. Modify	19
3.4. Model	24
3.5. Assess	26
4. Resultados e Discussão	27
5. Conclusões.....	54
6. Limitações e Recomendações para Trabalhos Futuros	57
7. Bibliografia.....	58
8. Anexos	61

ÍNDICE DE FIGURAS

Figura 1 – Conjunto de <i>itemsets</i> composta em árvore	8
Figura 2 – Conjunto de <i>itemsets</i> após o seu processo	8
Figura 3 - Diagrama do SAS <i>Enterprise Miner</i>	16
Figura 4 - Número de pacientes da base dados por mês	16
Figura 5 - Variável “ID Cliente” presente no WEKA	18
Figura 6 - Variável “Cor” presente no WEKA	18
Figura 7 - <i>Replacement</i> das variáveis de classe	19
Figura 8 - <i>Impute</i> na Variável “Idade”	20
Figura 9 - Filtro aplicado à variável “REP_Concelho”	20
Figura 10 - Filtro aplicado à variável “IMP_Idade”	21
Figura 11 - Número de variáveis excluídas	22
Figura 12 - Variáveis excluídas	22
Figura 13 - Variável “Código Postal” após a conversão	23
Figura 14 - Variáveis finais no <i>software</i> WEKA	23
Figura 15 – Escolha do número de <i>clusters</i> ideal.....	24
Figura 16 – Configurações no <i>software</i> WEKA para as regras de associação	25
Figura 17 – Número de pacientes em cada <i>cluster</i>	27
Figura 18 - Gráfico relativo ao número de pacientes em cada mês	28
Figura 19 - Gráfico referente à Variável “Alta Médica”	29
Figura 20 – Número de pacientes para “Alta para o Domicilio” e respetivos “Destino Final”	30
Figura 21 - Número de pacientes para “Hospital do Dia” e respetivos “Destino Final”	30
Figura 22 - Número de pacientes para “Consulta Externa” e respetivos “Destino Final”	31
Figura 23 - Número de pacientes para “Serviços de Internamento” e respetivos “Destino Final”	32
Figura 24 - Número de pacientes para “Outros” e respetivos “Destino Final”	33
Figura 25 - Número de pacientes para “Morte” e respetivos “Destino Final”	33
Figura 26 - Número de pacientes para “ARS / Centro de Saúde” e respetivos “Destino Final”	33
Figura 27 - Gráfico sobre o motivo dos pacientes recorrem às urgências.....	34
Figura 28 - Gráfico relativo à variável cor das pulseiras.....	35
Figura 29 – Gráfico referente ao Distrito dos pacientes	36
Figura 30 - Relação entre as variáveis “Data” e “Razão”	42
Figura 31 – Relação entre as variáveis “Cor” e “Razão”	48
Figura 32 - Regras criadas através do <i>software</i> WEKA.....	51

Figura 33 – Aumento do número de regras de associação	52
Figura 34 - <i>Missing Values</i>	61
Figura 35 - <i>Replacement</i> da variável Alta	61
Figura 36 - Variáveis após utilizar o nó <i>Replacement</i> e <i>Impute</i>	62
Figura 37 - Filtros nas variáveis de classe	63
Figura 38 - Filtros nas variáveis de classe	64
Figura 39 – Opção <i>NumericToNominal</i>	64
Figura 40 - Código Gráfico Cotovelo	66
Figura 41 - Configurações originais para as regras de associação	67
Figura 42 – Gráfico ampliado sobre o motivo dos pacientes recorrerem às urgências	71
Figura 43 – Figura ampliada referente à Variável Distrito	72
Figura 44 - <i>Items</i> para as regras de associação.....	81

ÍNDICE DE TABELAS

Tabela 1 – Tabela com as transações	7
Tabela 2 – Resultados de confiança para cada uma das regras	9
Tabela 3 – <i>Role</i> e <i>level</i> de cada variável	17
Tabela 4 – Número de pacientes em cada <i>cluster</i>	27
Tabela 5 – Tabela relativa ao número de pacientes em cada mês.....	68
Tabela 6 - Tabela de Anexo relacionando as Variáveis “Alta Médica” e “Destino Final”	69
Tabela 7 - Tabela dos Anexos referente à Variável “Alta Médica”	70
Tabela 8 - Tabela de Anexo sobre o motivo dos pacientes recorrem às urgências	70
Tabela 9 - Tabela dos Anexos referente à Variável “Cor”	70
Tabela 10 - Tabela dos Anexos referente à Variável Distrito	72
Tabela 11 - Tabela dos Anexos relacionando as Variáveis “Data” e “Razão”	76
Tabela 12 - Tabela dos Anexos relacionando as Variáveis “Cor” e “Razão”	79

LISTA DE SIGLAS E ABREVIATURAS

BD	Base de Dados
CART	<i>Classification And Regression Trees</i>
CHD	<i>Coronary Heart Disease</i>
DM	Data Mining
INEM	Instituto Nacional de Emergência Médica
PNS	Plano Nacional de Saúde
RBV	<i>Radial Basis Function</i>
ROC	<i>Receiver Operating Characteristics</i>
SEMMA	<i>Sample, Explore, Modify, Model e Assess</i>
SNS	Serviço Nacional de Saúde
SU	Serviços de Urgência
SUB	Serviço de Urgência Básico
SUMC	Serviço de Urgência Médico-Cirúrgico
SUP	Serviço de Urgência Polivalente
SVM	<i>Support Vector Machine</i>
UH	Urgências Hospitalares

1. INTRODUÇÃO

O tema da Saúde é um tema sensível para a sociedade, pelas implicações que tem a nível político, económico e social, estando sob escrutínio permanente, sendo as suas fraquezas motivo de análise e discussão frequentes. Analisando, por exemplo, os dados publicados na Pordata (uma plataforma onde se pode consultar informação sobre múltiplas áreas da sociedade portuguesa, municípios e países europeus), consegue-se perceber que a maior evolução nos últimos 50 anos foi a diminuição da mortalidade infantil, sendo este considerado como um indicador internacional para medir o desenvolvimento humano dos países (Rosa, 2014). Em 1961 verificaram-se 89 óbitos no primeiro ano de vida por cada mil nascimentos e em 2011 esse valor rondava apenas os 3.1, ou seja, 3 óbitos. Como consequência, a esperança média de vida dos portugueses aumentou de 63 anos em 1961 para 81 anos em 2011. Esta mesma fonte indica que o índice de envelhecimento é cada vez maior, apresentando mais de 10% de diferença no índice de longevidade entre 1961 e 2011 (Pordata, 2015).

Desta forma, a Saúde é umas das áreas que requer maior atenção em Portugal. A população portuguesa encontra-se cada vez mais envelhecida e existem cada vez menos nascimentos (em 1960 a taxa de natalidade era de 16,2% e em 2011 foi apenas de 7,9%) (Pordata, 2015). Como tal, para além de assegurar o adequado apoio aos utentes mais novos, é importante garantir o bem-estar dos idosos, pois são aqueles que necessitam de maiores cuidados e estão mais dependentes dos serviços de saúde.

Para além de se assegurar as condições nos serviços especializados, também se deve assegurar as condições nas Urgências Hospitalares (UH). As UH têm ganho cada vez mais importância, tendo que responder de forma eficaz e imediata a quaisquer problemas que surjam. No entanto, ao longo dos últimos anos o seu funcionamento tem causado preocupação ao Serviço Nacional de Saúde, não só pelos recursos que consome, mas também pelos pacientes que ali acorrem sem efetiva necessidade dos seus serviços, implicando um aumento do número de pacientes nos hospitais e como consequência, uma redução da eficiência e uma maior dificuldade em dar respostas adequadas às reais necessidades dos pacientes.

Para dar resposta a estes problemas, foi criado o Plano Nacional de Saúde (PNS). O PNS é uma estratégia adotada por vários países e que permite alinhar as políticas de saúde de forma estruturada e coerente. Existem diversos intervenientes na criação do PNS, como por exemplo instituições nacionais, organizações governamentais e não-governamentais, instituições do setor público e privado. O PNS tem como objetivo analisar o estado de saúde da população, de forma a identificar problemas, necessidades e possíveis soluções. Neste plano também está descrito a importância de todos os intervenientes na promoção da saúde, prevenção da doença, diagnóstico precoce, minimização e controlo da doença e reabilitação adequada (Plano Nacional de Saúde, 2010).

Atualmente está em vigor o Plano Nacional de Saúde 2011-2016, que foi estendido até 2020. Este documento define como metas a redução para menos de 20% da taxa de mortalidade prematura (abaixo dos 70 anos), o aumento de 30% da esperança média de vida saudável aos 65 anos e a redução de fatores de risco que levam a doenças, como o consumo e exposição ao tabaco e a obesidade infantil, para os menores de 15 anos (Plano Nacional de Saúde Revisão e Extensão a 2020, 2015).

Para além deste documento geral referido, existe um Plano de Saúde para cada região. Uma vez que os dados que iremos tratar referem-se às urgências de um hospital do norte do país, é importante estudar não só o Plano Nacional de Saúde geral, mas também o Plano de Saúde da Região Norte (Plano Nacional de Saúde, 2010).

O Plano de Saúde da Região do Norte para 2014-2016 tem como objetivo diminuir a taxa de mortalidade prematura (até aos 64 anos) causada por doenças como doenças cerebrovasculares, tumores na traqueia, brônquios, pulmões, estômago e mama; diminuir para 3/100 000 habitantes a taxa de pacientes com VIH ou diminuir a taxa de obesidade infantil para 14% (Sa & Norte, 2014). Este objetivo é semelhante ao do PNS, sendo que as medidas propostas para atingir o objetivo são, em regra, coincidentes.

No entanto, apesar deste planeamento para a área da saúde, continuam a surgir problemas nos hospitais, principalmente nas urgências. Em Portugal, no fim de 2014/início de 2015 surgiu uma grande epidemia de gripe, criando o caos nos Serviços de Urgência Hospitalar. As longas horas de espera para os pacientes serem atendidos associado à falta de camas para acolher os doentes foram os fatores mais relevantes. Como consequência desta situação, registaram-se mortes e mais horas de trabalho por parte dos médicos e enfermeiros (Leal, Ferreira & Carvalho, 2015; Santos, 2015).

São casos como estes da epidemia da gripe que permitem aferir deficiências no planeamento e gestão das Urgências Hospitalares. Apesar de existir o Plano Nacional de Saúde e a preocupação pelo bem-estar dos pacientes, ainda é preciso compreender o que se pode fazer para melhorar e evitar casos como o anterior.

Assim, com este trabalho pretendemos analisar os dados disponíveis de um hospital no norte do país, referentes a 2012, que permitirão perceber e compreender a atual situação nas Urgências Hospitalares. Com esta análise poderemos encontrar padrões e regras relevantes que ajudem no aperfeiçoamento da sua gestão.

Os problemas existentes nas Urgências Hospitalares continuam a ser um tema atual e de preocupação geral para a população, pelo que julgamos ser um tópico interessante para investigar e aprofundar. As Urgências Hospitalares são sempre o sítio onde quem precisa de uma consulta urgente ou não a consegue de outro modo se dirige, pelo que é necessário melhorar a sua gestão para garantir uma maior eficiência e eficácia no atendimento. O facto de ter acontecido o grande surto de gripe no inverno de 2014/2015 e quase ter causado o colapso nas urgências, aumentou a curiosidade sobre o tema, levando a querermos tentar perceber a situação atual dos serviços de urgência.

O objetivo deste trabalho é analisar os dados de um hospital do norte do país, utilizando as técnicas de Data Mining para obter resultados que possam suportar e apoiar o planeamento e gestão da Saúde, por forma a melhorar os seus serviços, mitigando as falhas existentes. No capítulo que se segue serão apresentados os conceitos teóricos que serão utilizados ao longo deste trabalho, tal como Data Mining, Saúde e Data Mining na Saúde. No capítulo seguinte poderá analisar-se todo o procedimento utilizado, focando-se na metodologia SEMMA (*Sample, Explore, Modify, Model, Assess*). Por último, será apresentada a discussão de resultados, onde se irá concluir se o trabalho realizado terá ou não um impacto positivo no planeamento e gestão das urgências.

2. REVISÃO DA LITERATURA

2.1. SAÚDE E URGÊNCIAS

O conceito de Saúde é tão abrangente, que se torna difícil defini-lo de uma forma simples. Boruchovitch & Mednick (2002) afirmam que a palavra Saúde apareceu aproximadamente no ano 1000 e é conhecida, ou pode ser definida, como a ausência de doenças, sintomas ou problemas, tanto a nível físico como psicológico. Sabe-se que para garantir uma boa saúde, é importante tomar certas precauções e levar um estilo de vida saudável. No entanto quando a saúde é posta em causa, é importante consultar um médico. Quando essa hipótese não é viável num curto prazo de tempo, seja por questões de dificuldade de acesso a consultas no Centro de Saúde ou económicas por falta de recursos para pagar uma consulta, as urgências são a opção mais viável.

Urgências são todas as situações onde existe intervenção de cuidados médicos num curto espaço de tempo, com risco de falha das funções vitais (Direcção-Geral da Saúde, 2001). Por outro lado, para além das Urgências também existem as Emergências. Emergências são todas as situações onde existe o risco de perda de vida ou de função orgânica, necessitando de intervenção médica o mais rapidamente possível (Direcção-Geral da Saúde, 2001). No entanto, no despacho 18459/2006 verifica-se que não existe distinção entre estes dois conceitos. É apenas definido como uma situação clínica onde se verifica de forma súbita o risco de falência de uma ou mais funções vitais (Campos, 2014).

Estes conceitos podem confundir-se entre si. Tanto Urgências como Emergências vem já do passado, desde que existem médicos a ajudarem pacientes em situações urgentes, destacando-se o período das duas grandes guerras mundiais e os surtos de epidemias, sendo socorro em emergência ou a prática da medicina de emergência fulcral para o tratamento dos doentes (Scarpelini, 2007; Totten & Bellou, 2013).

Para combater estas situações, foram criados dois modelos que foram adotados para o atendimento nas urgências no mundo: o Modelo Anglo-Americano e o Modelo Franco-Germânico. Muitos países desenvolvidos seguem o Modelo Anglo-Americano, embora o modelo Franco-Germânico também apresente vantagens (Arnold, 1999). No Modelo Anglo-Americano, os pacientes são transportados para o centro hospitalar, para que lhes sejam fornecidos os cuidados necessários. Durante o transporte, a intervenção é da responsabilidade de técnicos de emergências médica e de paramédicos especializados, que atuam sobre protocolos médicos rigorosos. Após a chegada ao centro hospitalar, os cuidados dos utentes ficam entregues aos médicos de emergência médica, preparados para lidar com estas situações. Com este modelo, a medicina de emergência é reconhecida como uma especialidade, cujos médicos são especialistas em casos de emergência (Arnold, 1999; Scarpelini, 2007). Por outro lado, no Modelo Franco-Germânico os médicos e as tecnologias são transportadas até ao paciente, de forma a proporcionar um elevado cuidado médico antes dos utentes chegarem ao hospital. A emergência médica é praticamente realizada no local onde os pacientes se encontram, onde os médicos prestam os seus serviços. Neste sentido, o departamento no hospital para as emergências médicas é menor, pois toda a triagem e diagnóstico são feitos em campo. Ao contrário do modelo anterior, neste modelo a medicina de emergência não é reconhecida como uma especialidade (Arnold, 1999; Scarpelini, 2007).

O Modelo Franco-Germânico é o que tem sofrido maiores críticas, sendo referido que os médicos não estão suficientemente treinados para atuar, não são supervisionados e não tem a mesma qualidade de serviços que o Modelo Anglo-Americano. Para além disto, este método é mais dispendioso e a eficácia dos médicos pode até estar em causa (Arnold, 1999).

Existem também algumas diferenças na forma como é gerido o atendimento hospitalar. Em alguns países, como por exemplo os Estados Unidos, Canadá e Inglaterra, existe a opção da formação de especialistas em emergências médicas. Estes especialistas são preparados para enfrentar situações críticas e de *stress*, trabalhando sempre em equipa. Por outro lado, existe um modelo multidisciplinar, onde um médico é responsável por várias áreas, de forma a responder prontamente às urgências que surjam (Scarpelini, 2007).

Relativamente à situação em Portugal, o Serviço Nacional de Saúde (SNS) está implementado desde 1979, pela Lei n.º 56/79, de 15 de setembro, onde foi constituída uma rede de instituições e serviços prestadores de cuidados globais de saúde a toda a população, financiada através de impostos, em que o Estado salvaguarda o direito à proteção da saúde (Portal da Saúde, 2015), nomeadamente, os estabelecimentos hospitalares, independentemente da sua designação, as unidades locais de saúde e os agrupamentos de centros de saúde.

Os estabelecimentos hospitalares, independentemente da sua designação, e as unidades locais de saúde devem garantir a presença de uma equipa intra-hospitalar de suporte em cuidados paliativos, com profissionais das áreas da medicina, enfermagem e psicologia (Diário da República, 2014). Os agrupamentos de centros de saúde são serviços públicos de saúde que agrupam um ou mais centros de saúde e que têm como objetivo principal a prestação de cuidados de saúde primários à população (Governo de Portugal, 2015).

Para além das entidades referidas anteriormente, as Urgências Hospitalares também têm um papel fundamental no SNS. Os Serviços de Urgência (SU) estão divididos em três níveis: Serviço de Urgência Polivalente (SUP), Serviço de Urgência Médico-Cirúrgico (SUMC) e Serviço de Urgência Básico (SUB). O SUP é o serviço mais diferenciado na resposta às urgências. Este deve dar resposta às necessidades de procura, resultantes da demografia regional. O SUMC é o serviço de segundo nível no acolhimento em situações de urgência, servindo de apoio ao SUP. O SUB é o serviço de primeiro nível de acolhimento. Este consiste na resolução de situações mais simples e mais comuns na urgência (Paiva, 2012).

As urgências têm ganho cada vez mais importância no ramo da saúde. Nos últimos 20 anos, registaram-se anualmente, em média, o acesso de mais de 10 milhões de utentes ao serviço de Urgências Hospitalares, implicando uma redução da eficiência e conseqüentemente, uma maior dificuldade nas respostas às necessidades reais, ou seja, “entopem” os serviços, atrasando ou dificultando as verdadeiras urgências. Para lutar contra esta situação, foram implementadas várias medidas corretivas, como sejam os incentivos a cuidados programados de saúde, penalizações na utilização das urgências de forma inadequada pela aplicação de taxas moderadoras, pela melhoria na acessibilidade às consultas médicas ou alargamento dos serviços de apoio por telefone (Direção-Geral da Saúde, 2001).

Devido ao aumento da procura, em 1981 foi criado o Instituto Nacional de Emergência Médica (INEM), tendo como objetivo principal responder rapidamente às situações críticas. Desde a

implementação até aos dias de hoje, a área da Emergência Médica foi uma das que mais evolução teve no apoio às necessidades da população, sendo considerada uma das áreas de maior qualidade no setor da Saúde, conforme dados obtidos através de um inquérito realizado, onde 93% dos inquiridos estão satisfeitos ou muito satisfeitos com o atendimento (Campos, 2014). Desta forma, dos dois modelos apresentados anteriormente, podemos dizer que Portugal se integra no Modelo Franco- Germânico, ainda que possa ter especificidades próprias, caracterizando-se pela rápida atuação do INEM em situações urgentes.

2.2. DATA MINING

O Data Mining é considerado como uma nova disciplina que incluiu a estatística, tecnologias de base de dados, reconhecimento de padrões, *machine learning*, entre outras áreas. O foco desta técnica consiste na análise de grandes bases de dados, com o objetivo de encontrar relações desconhecidas que possam trazer valor para a empresa (Hand, 1998). Muitas vezes o conceito de Data Mining aparece relacionado com a parte analítica de um processo mais amplo chamado de “Descoberta de Conhecimento em Bases de Dados”, onde se incluem todos os procedimentos de organização e preparação dos dados (Fayyad, Piatetsky-Shapiro & Smyth, 1996). Tendo em conta que o volume de dados cresce exponencialmente ao longo do tempo e estimando que os dados produzidos em 2020 serão 44 vezes maior que em 2009, é imprescindível a criação de técnicas para a análise desta informação (CSC, 2012).

Existem dois tipos de modelação que são muito aplicados em Data Mining: a Modelação Descritiva e a Modelação Preditiva.

A Modelação Descritiva, tal como o nome indica, pretende resumir e simplificar grandes quantidades de informação, de forma a aumentar o conhecimento do analista. Por outro lado, a Modelação Preditiva, consiste na utilização de um modelo que permita prever algum valor quantitativo ou qualitativo que seja de interesse para o negócio. Esta modelação utiliza a estatística e modelos matemáticos para prever futuros resultados, comportamentos e tendências. No entanto, não é de consenso geral as definições de ambas as modelações, pois alguns modelos preditivos podem ser considerados descritivos e vice-versa (Fayyad, Shapiro & Smyth, 1996).

Dentro da Modelação Preditiva, existem dois padrões muito conhecidos: a Regressão e a Classificação. A Regressão tem como principal função mapear um conjunto de dados a uma variável de valor real. No caso da Classificação, esta serve para classificar dados em classes já pré-definidas. Para a classificação dos dados, é criado um modelo de classificação, através de um algoritmo, e que é constituído por um conjunto de regras. A Classificação também se divide numa fase de treino e fase de teste, onde no treino é criado o tal modelo de classificação, para posteriormente ser testado na fase de testes (Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

O Data Mining é constituído por várias famílias de técnicas, seguindo-se uma explicação das mais importantes da Modelação Descritiva e Modelação Preditiva.

2.2.1 Modelação Descritiva

2.2.1.1 Regras de Associação

As regras de Associação são técnicas não supervisionadas que têm como objetivo a descoberta de padrões e relações, desconhecidos ou pouco óbvios, entre os dados. Como resultado, esta técnica gera um conjunto de regras que definam os padrões e relações encontrados. A aplicação desta técnica é muito comum no retalho, onde os dados são definidos como transações, isto é, um conjunto de itens. Por exemplo, numa cadeia de supermercados as compras de cada cliente são vistas como uma transação sendo os diversos produtos que a compõem itens individuais. As regras de associação geradas são do tipo “ $X \rightarrow Y$ ”, indicando que as transações que contêm itens “ X ” tendem a conter também itens “ Y ” (Srikant & Agrawal, 1997).

A partir desta técnica é possível gerar uma vasta quantidade de regras, no entanto muitas destas são óbvias e não acrescentam qualquer valor comercial. Kenneth Lai (2001) distingue dois tipos de métricas, objetivas e subjetivas, que permitem avaliar a qualidade e utilidade das regras geradas. Como métricas objetivas tem-se o Suporte e a Confiança, que envolvem análises matemáticas e estatísticas dos dados. As métricas subjetivas são a *Unexpectedness* e *Actionability*. A *Unexpectedness* afirma que estas regras apenas são uteis se o agente de negócio não tiver conhecimento dos padrões encontrados ou se os mesmos contradizem conhecimento anterior dado como certo. Por exemplo, encontrar a regra “Pão \rightarrow Leite” (se comprar pão, é muito provável comprar leite), não trás conhecimento inesperado ao negócio sendo uma regra óbvia e como tal, não seria considerada como *Unexpectedness*. A *Actionability* afirma que as regras apenas são uteis se forem utilizadas e trouxerem alguma vantagem prática para o problema. As métricas subjetivas são difíceis de serem aplicadas, sendo necessário analisar cada caso individualmente, estando por isso os resultados intrinsecamente relacionados com o agente de negócio que realizou a avaliação (Lai & Cerpa, 2001).

Existem inúmeros algoritmos de Regras de Associação, sendo um dos mais utilizados o Algoritmo Apriori (Agrawal & Srikant, 2008). Este algoritmo é utilizado maioritariamente em dados que consistam em transações, validando as regras utilizando métricas objetivas, nomeadamente de Suporte e Confiança. O Suporte consiste num indicador que representa a frequência com que os itens ou conjuntos dos mesmos aparecem no conjunto de dados. A Confiança representa o número de vezes que as regras geradas estão corretas (Lai & Cerpa, 2001). A confiança de uma regra do tipo $X \rightarrow Y$ é calculada através da fórmula que se segue. As regras que tenham uma Confiança e um Suporte alto são consideradas regras credíveis e “fortes” (Chen, Han & Yu, 1996).

$$\frac{\text{Suporte}(X \rightarrow B)}{\text{Suporte}(X)}$$

No algoritmo Apriori, um conjunto de itens é referido como um *itemset*, onde um *itemset* de k itens é referido como um *k-itemset*. O conjunto de dados é representado através de uma estrutura em árvore composta pelos vários *itemsets* gerados pela combinação dos itens existentes. A árvore é

composta por uma única folha com um *itemset* de tamanho n (com todos os itens existentes), sendo a raiz composta por tantos nós quanto o número de itens (vários 1-*itemset*). Os restantes nós da árvore representam os restantes *itemsets* possíveis de serem obtidas através da conjugação dos diversos itens (2-*itemsets*, 3-*itemsets*, etc). As relações entre os nós indicam as dependências, ou seja, o 3-*itemset* $\{X, Y, Z\}$, irá ter como dependência os 2-*itemsets* $\{X, Y\}$, $\{X,Z\}$ e $\{Y,Z\}$ (Hegland, 2005).

O algoritmo Apriori utiliza uma abordagem *bottom-up* (1-*itemsets* para k -*itemsets*) iterativa, composta por tantas iterações quantos *itemsets* com elevado suporte forem encontrados. O suporte de cada *itemset* consiste no número de transações onde esse *itemset* é encontrado, isto é, a frequência com que pode ser encontrado no conjunto de dados (Hegland, 2005).

A execução do algoritmo é baseada numa pesquisa *breadth-first* que, aliada à abordagem *bottom-up*, permite que o algoritmo seja mais eficiente limitando o número de nós analisados. É assumido que, se um dado *itemset* tem baixo suporte, então os seus descendentes também o terão e portanto não é necessário visitar os mesmos (Hegland, 2005).

Numa primeira iteração, é calculada o suporte de todos os 1-*itemsets*, eliminando aqueles que possuam um suporte inferior àquele definido aquando o início do algoritmo. Os *itemsets* com elevado suporte são utilizados como conhecimento, sendo numa segunda iteração calculado o suporte de todos os 2-*itemsets*, cujos parentes não tenham sido eliminados na iteração anterior. Este processo é repetido até que não restem mais *itemsets* válidos. Para cada *itemset* encontrado, é gerada um conjunto de regras a partir dos itens por ele composto, sendo calculada a sua confiança. As regras com uma confiança superior à definida no início do algoritmo são consideradas válidas. Por vezes mesmo regras com grande confiança e suporte podem não ser plausíveis, sendo necessária a aplicação de métricas subjetivas de forma a validar as mesmas (Hegland, 2005).

Exemplo:

Seja o conjunto de dados, composto por cinco transações perfazendo um total de quatro itens distintos. Seja o conjunto de dados definidos de acordo com a tabela 1:

Transação	Itens
1	a,b
2	a,b,d
3	a,c
4	a,b,d
5	a,c

Tabela 1 – Tabela com as transações

Um conjunto de *itemsets* é gerado a partir da combinação de todos os itens existentes, criando uma estrutura em árvore, como está representada na figura 1, composta na raiz por quatro nós e com uma única folha composta por um *itemset* contendo os quatro itens.

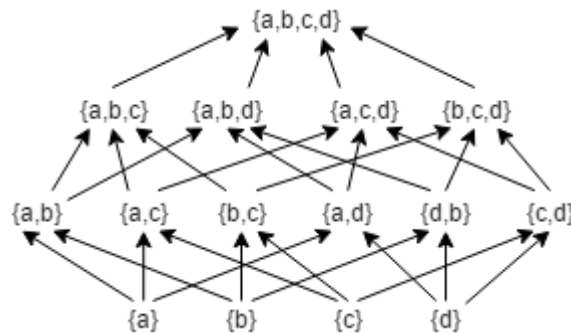


Figura 1 – Conjunto de *itemsets* composta em árvore

Pretende-se aplicar o algoritmo Apriori utilizando um suporte de 60% e uma confiança de 70%. Numa primeira iteração, é calculado o suporte de cada 1-*itemset*, sendo descartados aqueles que tenham um suporte inferior a 60% (que neste cenário equivale a uma frequência inferior a 3). O 1-*itemset* {c} é o único que têm um suporte de 2, sendo portanto descartado. Os restantes *itemsets* têm um suporte superior. Na iteração seguinte, é calculado o suporte dos 2-*itemsets* {a,b}, {a,c} e {b,c}. Destes, apenas {b,c} tem suporte inferior a 60%, sendo descartado. Não havendo mais *itemsets* válidos, esta fase do algoritmo termina. (Nenhum 3-*itemset* iria ter um suporte superior ao mínimo pois pelo menos um dos seus ascendentes foi eliminado usando a heurística Apriori).

A imagem da figura 2 demonstra, o estado da árvore após este processo. Todos os nós descendentes dos *itemsets* {d} e {b,c} são eliminados, não sendo visitados nas iterações seguintes.

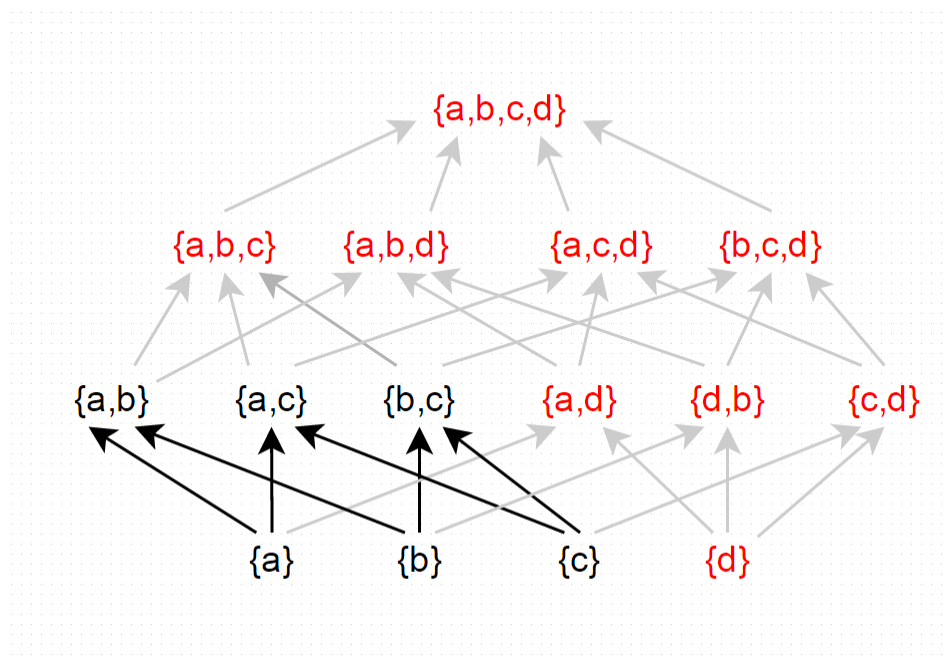


Figura 2 – Conjunto de *itemsets* após o seu processo

Para cada *itemset* com elevado suporte encontrados nesta primeira fase, são criadas regras do tipo $A \Rightarrow B$, calculando a sua confiança. Apenas as regras com elevada confiança serão consideradas fortes e válidas. No exemplo anterior os *itemsets* com suporte superior ao desejado são: $\{a,b\}$, $\{a,c\}$. A tabela 2 apresenta os resultados de confiança para cada uma das regras geradas a partir destes *itemsets*.

Itemset	Regra	Confiança
$\{a,b\}$	$a \Rightarrow b$	$3/5 = 60\%$
$\{a,b\}$	$b \Rightarrow a$	$3/3 = 100\%$
$\{a,c\}$	$a \Rightarrow c$	$3/5 = 60\%$
$\{a,c\}$	$c \Rightarrow a$	$3/3 = 100\%$

Tabela 2 – Resultados de confiança para cada uma das regras

As regras $a \Rightarrow b$ e $a \Rightarrow c$ possuem uma confiança inferior a desejada sendo portanto regras descartadas. As regras $c \Rightarrow a$ e $b \Rightarrow a$ são regras que possuem um suporte e uma confiança elevadas sendo portanto o resultado final deste algoritmo.

2.2.1.2 Clustering

A análise de *clusters* é uma família de métodos que agrupam uma população heterogénea de acordo com as suas características e similaridade em grupos homogéneos (Jain, 2009). Os dados em cada *cluster* têm que ser idênticos entre si e diferentes entre *clusters*. Quanto maior a similaridade dos dados no *cluster* e maior a diferença entre *clusters*, mais fácil será a sua análise (Gupta & Mishra, 2011).

Em Data Mining existem dois tipos de classificação: supervisionada e não-supervisionada. O *clustering* é classificado como não-supervisionada pois não depende de classes pré-definidas, ou seja, os registos são agrupados apenas com base nas suas semelhanças. Numa classificação supervisionada é fornecido um conjunto de dados pré-classificados, onde os padrões de treino já são conhecidos (Jain, Murty & Flynn, 1999).

Jain (2009) afirma que os *clusters* diferem entre si em termos de tamanho, forma e densidade. Um *cluster* ideal seria um conjunto de dados compactos e isolados, mas são raros os casos onde esta situação se verifica, pois a frequente deteção de “ruído” (*outliers*) influencia de forma negativa a criação dos *clusters*. Na realidade, um *cluster* é uma entidade subjetiva pois a sua interpretação e o seu significado depende do conhecimento do utilizador sobre o assunto. No entanto, enquanto o Homem consegue facilmente interpretar *clusters* a duas ou três dimensões, dificilmente o conseguirá para dimensões maiores, pelo que é importante criar um algoritmo que interprete dados a grandes dimensões. Assim, existem as chamadas técnicas de análise de *clusters*, que estão divididas em dois grandes métodos: Métodos Hierárquicos e Métodos de Partição:

- **Método Hierárquico**

Tal como o nome indica, este método agrupa os dados em subgrupos, com uma certa hierarquia e em diversas etapas. Este divide-se em aglomerativos e divisivos, dependendo de como é a sua decomposição. Um método hierárquico aglomerativo inicia com cada objeto separado formando um único grupo no final. Este algoritmo começa por juntar os objetos próximos entre eles, até que fique um único *cluster* no final. Um método hierárquico divisivo inicia com todos os objetos num único *cluster*. Em cada interação os objetos vão sendo divididos de forma a criar *clusters* mais pequenos, até que cada objeto corresponda a um único *cluster* (Zhao & Karypis, 2002).

Existem diversas regras de agrupamento, de forma a medir as distâncias entre *clusters*. Entre as mais conhecidas estão a *Single linkage* e a *Complete linkage*, no entanto existem outras, como por exemplo *Unweighted pair-group average* ou *Unweighted pair-group centroid*. *Single linkage* (vizinho mais próximo) afirma que a distância entre dois *clusters* é determinada pela distância dos objetos mais próximos de cada *cluster*. *Complete linkage* (vizinho mais afastado) é o oposto do anterior; a distância entre dois *clusters* é determinada pela distância dos objetos mais afastados de cada *cluster* (Gupta & Mishra, 2011). *Unweighted pair-group average* caracteriza-se como a média das distâncias entre todos os objetos, na distância entre dois *clusters*. Por último, o *Unweighted pair-group centroid* é igual ao anterior, com a diferença de não ser utilizado a média das distâncias, mas sim a distância entre os centróides dos dois *clusters*.

- **Método de Partição**

Este método consiste na formação de um conjunto de *clusters*, onde cada grupo contém pelo menos um objeto e cada objeto pertence apenas a um *cluster*. Inicialmente cria-se uma partição inicial e graças a uma técnica interativa de realocização, são movidos objetos entre grupos de forma a melhorar as diversas partições. Tal como foi explicado anteriormente, um critério de boa partição consiste nos objetos do mesmo *cluster* serem idênticos entre si e diferentes com os outros *clusters*. Sendo impossível uma enumeração exaustiva de todas as partições possíveis, existem dois algoritmos para realizar a partição: algoritmo *k-means*, onde cada *cluster* é representado pelas médias dos seus valores dos objetos, e o algoritmo *k-medoids*, onde cada *cluster* é representado por um dos objetos perto do centro do *cluster* (Jain, 2009). O *k-means* é o algoritmo mais utilizado; começa-se por seleccionar aleatoriamente *k-centróides* (ou objetos), sendo que há medida do processo, estes vão-se decompor e realocar-se em novos *k* grupos, baseado na similaridade dos objetos ao centróide (Yoo et al., 2012).

2.2.2 Modelação Preditiva

2.2.2.1 Árvores de Decisão

As árvores de decisão são dos algoritmos mais utilizados na Classificação, apesar de também serem utilizadas na Regressão. As árvores de decisão são compostas por nós, arcos e folhas. Os nós representam as interrogações que se colocam sobre o conjunto de dados; os arcos separam o conjunto de dados de acordo com o critério em causa e por ultimo as folhas representam os nós finais. As árvores de decisão são construídas com a raiz no topo e as folhas por baixo, tendo uma relação unidirecional (Rokach & Maimon, 2005).

A indução das árvores de decisão parte de um conjunto de variáveis de *input* e uma variável de output ou target. Utilizando as variáveis *input*, tenta-se criar regras que permitam isolar em subconjuntos as observações que possuem valores idênticos para a variável de output (Rokach & Maimon, 2005). Para a criação desses subconjuntos, a separação é da responsabilidade de vários algoritmos, como o C4.5 ou CART (Classification And Regression Trees), pois identificam as variáveis e o respetivo limite de separação, de forma a separa-los em dois ou mais ramos (ou subgrupos). (Hariz, Adnan, Husain, Aini, & Rashid, 2012).

O algoritmo C4.5 inicia-se com um único nó, incluindo todos os exemplos de treino. Caso estes sejam da mesma classe, este torna-se o nó final e são todos classificados de igual forma. Caso contrário, o algoritmo utiliza uma medida de entropia como forma de selecionar a variável de *input* que melhor faça a separação. A entropia varia entre um máximo 0.5 (entropia máxima) e 0 (quando todos os elementos pertencem à mesma classe) (Rokach & Maimon, 2005).

O algoritmo CART constrói uma árvore binária, ou seja, cada nó dá origem a apenas dois nós descendentes. O CART separa os registos em cada nó de acordo com uma função de um único campo de *input*. Assim, a primeira tarefa consiste em decidir qual, de entre as variáveis de *input*, produz a melhor partição. A melhor partição é definida como sendo a que produz a melhor separação dos registos em grupos onde uma única classe predomina (Rokach & Maimon, 2005).

Uma das maiores vantagens desta técnica é a fácil interpretação e leitura dos resultados. Esta permite que os utilizadores sem conhecimento da técnica consigam perceber os atributos com maior peso e que mais afetam os resultados. Por outro lado, quando existem muitos atributos, a árvore de decisão fica muito complexa e de difícil interpretação. Nestes casos, usam-se métodos estatísticos de forma a apurar os ramos que podem ser retirados da análise. (Yoo et al., 2012)

2.2.2.2 Redes Neurais

As redes neuronais artificiais baseiam-se nas redes neuronais do cérebro humano. Um neurónio é constituído por um núcleo, dois tipos de ramificações, o axónio e as dendrites. Os axónios transmitem sinais para os outros neurónios, através de sinapses, enquanto as dendrites recebem sinais dos neurónios anteriores, através de impulsos. Uma rede neuronal artificial atua, como base, com este princípio (Hariz et al., 2012).

Redes neuronais artificiais são técnicas computacionais, que permitem a criação de um modelo matemático através de uma aprendizagem automática utilizando um conjunto de dados previamente classificados (dados de treino). À semelhança com as redes neuronais humanas, as artificiais são compostas por vários nós que se ligam por canais de comunicação que têm atribuído um determinado peso. No processo de treino, estes pesos vão sendo ajustados de forma a ampliar ou atenuar o impacto que essa ligação terá no modelo matemático gerado. Os nós podem ser classificados em duas categorias: nós de entrada (camada de *input*) e nós de saída (camada de *output*) ou em três categorias: nós de entrada (camada de *input*), nós intermédios (camada escondida) e nós de saída (camada de *output*). No entanto, a camada de *input* nem sempre é considerada, pois esta só serve para introduzir os valores iniciais e passar à camada seguinte (Yoo et al., 2012). Após o processo de treino, é possível a classificação automática de dados utilizando o modelo matemático gerado.

As redes neuronais têm vantagens tais como a capacidade de classificação rápida após aprendizagem assim como a sua flexibilidade na resolução de diversos problemas. No entanto esta técnica acarreta algumas desvantagens, tais como a dificuldade em interpretar os resultados obtidos, a relação intrínseca entre a qualidade dos resultados e a qualidade dos dados de treino (dados de treino com pouca qualidade irão produzir um modelo matemático com fracos resultados), grande suscetibilidade de *overfitting* aos dados de treino e a complexidade de todo o processo. A dependência dos dados de treino resulta ainda na geração de algum “ruído” quando os dados a classificar são distintos dos dados de treino. Para além disto, este requer muitos parâmetros, incluindo um número ótimo de camadas escondidas e o seu processo de aprendizagem é muito lento e computacionalmente caro (Bellazzi & Zupan, 2008; Yoo et al., 2012).

Este algoritmo é dos mais utilizados para as áreas da saúde e da biomedicina, suportando casos como estudos sobre o cancro ou previsões de diagnósticos (Yoo et al., 2012).

2.3. DATA MINING NA SAÚDE

Tal como nas outras áreas, o armazenamento de informação na saúde é cada vez maior. As bases de dados médicas contêm registos dos pacientes e diagnósticos médicos que podem ser úteis na tomada de decisão e salvamento de vidas (Hariz et al., 2012). No entanto, existem alguns constrangimentos a nível ético e legal, que impedem a utilização dos dados na sua totalidade. A segurança dos dados dos pacientes é sempre tida em conta, garantindo a privacidade em questões mais sensíveis (Cios & Moore, 2002).

A interpretação das imagens, sinais ou outros dados clínicos pode ser semelhante entre médicos, mas a sua forma de relatar pode ser bem diferente. A falta de uniformização da escrita dificulta as análises, pois estes usam nomes diferentes (sinónimos) para descrever a mesma doença (Cios & Moore, 2002).

Apesar de todas as dificuldades encontradas, o Data Mining consegue ajudar na análise da informação e na tomada de decisão, tal como no diagnóstico ou prognóstico de diversos problemas. Desta forma, ajuda a melhorar o serviço de saúde, ao detetar irregularidades, padrões ou exceções nos dados (Hariz et al., 2012).

São vários os casos onde as técnicas de Data Mining foram úteis na área de Medicina. Os exemplos que se seguem são apresentados pelo Hariz et al. (2012), focando-se principalmente em previsões na área da saúde.

As doenças coronárias são uma das principais causas de morte na China, pelo que foi feito um estudo com o objetivo de identificar a síndrome de CHD (*Coronary Heart Disease*), em doentes com esta doença, através das técnicas de Data Mining. Obtiveram-se 1069 casos de CHD, de 5 clínicas diferentes. Foram utilizados quatro técnicas de DM: classificador bayesiano, árvores de decisão, redes neuronais e SVM (*Support Vector Machine*), onde a performance destas técnicas eram determinadas por parâmetros da matriz de confusão: a sua sensibilidade (proporção de casos positivos que foram corretamente identificados), especificidade (proporção de casos negativos que foram incorretamente classificados) e precisão (proporção do número total de previsões corretas). No final, os resultados apontavam para as redes neuronais como as mais precisas (88,6%) e as árvores de decisão como as menos precisas (80,4%). Com estes resultados, este estudo mostrou que as técnicas de DM são uteis na previsão de doenças coronárias (Hariz et al., 2012).

Outro caso analisado foi relativo ao cancro da mama, envolvendo 1035 pacientes com esta doença durante 10 anos. As técnicas de DM envolvidas foram o classificador bayesiano, árvores de decisão, SVM, entre outras técnicas menos conhecidas. Todos os resultados aqui alcançados foram comparados com previsões de médicos oncologistas. A média de precisão da previsão com os resultados dos oncologistas foi 0,65 e 0,63, para as classes de recorrência e não recorrência da doença. Como resultado final, o classificador bayesiano alcançou o melhor resultado em termos de precisão (0,678) e as árvores de decisão alcançaram o pior resultado (0,674) (Hariz et al., 2012).

O próximo caso refere-se a uma previsão sobre fertilização In-Vitro e teve como objetivo identificar o número de embriões a serem transferidas para o ventre da mulher e selecionar aqueles com maior probabilidade de fecundar. Este estudo tinha 2453 registos com 89% dos casos com resultado positivo. Foram utilizados 6 técnicas de DM, sendo elas o classificador bayesiano, os k vizinhos mais

próximos, árvores de decisão, SVM (*Support Vector Machine*), percepção multicamada e RBF (*Radial Basis Function*). Foi utilizada a análise ROC (*Receiver Operating Characteristics*), isto é uma técnica que permite visualizar, organizar e selecionar as técnicas perante a sua performance, para avaliar o modelo com melhor performance, mostrando que o classificador bayesiano e o RBF tiveram uma melhor prestação que os restantes. Neste caso a previsão é difícil de analisar, pois o caso com resultados positivos é significativamente maior que os resultados negativos. No entanto, os bons resultados das técnicas RBF e classificador bayesiano mostram como se pode executar as técnicas mesmo em previsões difíceis (Hariz et al., 2012).

Numa outra previsão sobre obesidade infantil foram utilizadas seis técnicas de DM (árvores de decisão, regras de associação, redes neuronais, classificador bayesiano, SVM e RBF SVM) e a regressão lógica. Este caso consistia na identificação de crianças obesas ou acima do peso normal com 3 anos de idade, usando dados desde o seu nascimento, com 6 semanas, 8 meses e 2 anos de idade. Obteve-se 16653 instâncias, onde apenas 20% dos casos eram crianças obesas. Numa primeira fase fez-se o pré-processamento dos dados, fazendo uma análise dos *outliers* e *missing values*. As variáveis que se selecionaram para a previsão foram o sexo, peso da criança quando nasceu, tempo de gestação, o peso ganho com 6 semanas, o peso ganho com 8 meses e o peso ganho com 2 anos. No final de todo o processo, o SVM e RBF SVM são as técnicas com maior sensibilidade (59,6% e 60% respetivamente). A maior precisão pertence ao classificador bayesiano (54,7%) (Hariz et al., 2012).

3. METODOLOGIA

O processo que irá ser utilizado ao longo deste trabalho será o processo SEMMA. O acrónimo SEMMA significa *Sample, Explore, Modify, Model* e *Assess* e é um processo próprio para projetos de Data Mining. Tal como o acrónimo indica, este é dividido em cinco fases diferentes (Azevedo & Santos, 2008):

- **Sample** – Esta fase consiste na recolha de dados que contenha o maior número de informação, mas que seja fácil de manipular.
- **Explore** – Esta fase corresponde à exploração e tratamento dos dados, procurando falhas e incorreções nos dados que influenciem os resultados.
- **Modify** – Esta fase compreende a modificação nos dados, transformando as variáveis perante o âmbito do projeto.
- **Model** – Esta fase consiste na modelação automática dos dados consoante as técnicas escolhidas.
- **Assess** – Esta fase pretende avaliar os resultados pela sua utilidade no projeto.

Neste sentido, os próximos capítulos são referentes a cada uma das fases do SEMMA, para a criação de *clusters*, através do *software* SAS Enterprise Miner, e regras de associação, através do *software* WEKA.

O SAS *Enterprise Miner* é um *software* indicado para diversas técnicas de Data Mining e estatística. Através de um diagrama de processos interativo e dinâmico, permite uma fácil análise e interpretação dos dados (SAS, 2015).

O *software* WEKA é um programa especializado para tarefas de Data Mining. Os algoritmos podem ser aplicados diretamente na base de dados ou ser “chamado” através de código de programação Java. Neste *software* pode-se fazer o pré-processamento dos dados, a classificação, regressão, *clustering*, regras de associação e a visualização dos dados (WEKA, 2015).

O *software* WEKA foi utilizado para a criação de regras de associação; a segunda parte do projeto. Não foi utilizado o *software* SAS *Enterprise Miner* pois o programa não inclui todas as variáveis na criação das regras. Apenas integra as variáveis designadas como Target e ID, que correspondem apenas a 2 variáveis.

O diagrama da figura 3 que se segue é referente a todo o processo que irá ser explorado nos próximos tópicos. Aqui estão presentes os nós utilizados e qual o seu percurso para chegar ao resultado final. Como se pode observar, começa-se por se selecionar os dados pretendidos (HOSPITAL), seguindo-se uma análise exploratória dos mesmos, através do *StatExplore*. Os nós *Replacement* e *Inpute* irão preencher os dados em falta, seguindo-se um tratamento dos dados, onde serão retirados os *outliers* (dados que enviesam o estudo) e incorreções que estes tenham, através

do nó *Filter*. Por último, faz-se a análise de *clusters* e o respetivo gráfico de cotovelo, com a apresentação dos dados e discussão de resultados.

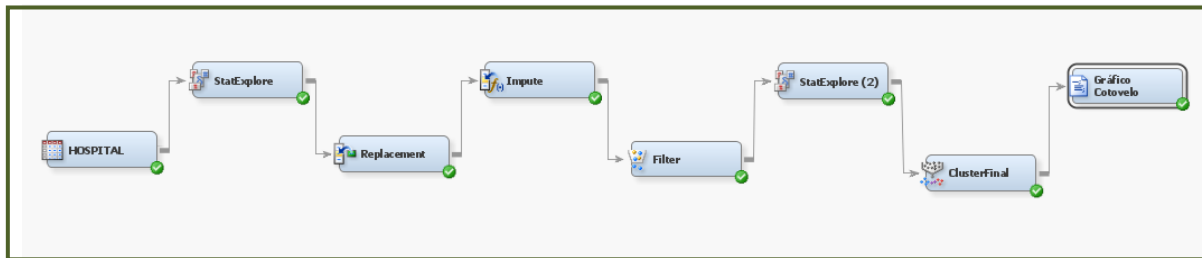


Figura 3 - Diagrama do SAS Enterprise Miner

3.1. SAMPLE

Recorreu-se a uma base de dados com informação da Urgência Hospitalar de um hospital do norte do país, relativa ao ano de 2012. Através da análise do gráfico da Figura 4, consegue-se perceber a variação de doentes que recorreram às urgências deste hospital. Na maior parte dos meses o número de pacientes nas urgências foi semelhante, com exceção nos meses de abril e novembro, onde se registou um número inferior à média (média representada pela linha do gráfico).

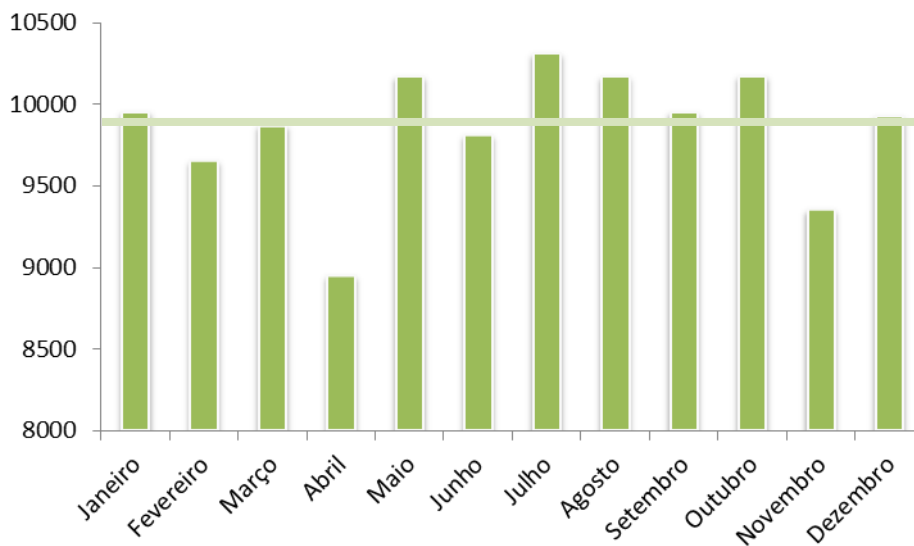


Figura 4 - Número de pacientes da base dados por mês

A base de dados tem um total de 118262 registos e é caracterizada por 13 variáveis:

- **ID Cliente** – valor único para cada paciente, sendo este a variável que os distingue dos restantes;
- **Data Episódio** – data em que ocorreu a ida às urgências;

- **Último Fluxograma / Razão** – razão pela qual o paciente se deslocou às urgências;
 - **Cor** – a cor da pulseira que foi atribuído a cada paciente, perante o grau de urgência no atendimento;
- Nesta BD irá encontrar-se as seguintes cores nas pulseiras: Vermelha (doentes emergentes), Laranja (doentes muito urgentes), Amarela (doentes urgentes), Verde (doentes pouco urgentes), Azul (doentes não urgentes) e Branca (doentes sem critério);
- **Localidade** – localidade onde cada paciente reside;
 - **Concelho** – concelho onde cada paciente reside;
 - **Código Postal** – código postal referente à localidade onde cada paciente reside;
 - **Distrito** – distrito onde cada paciente reside;
 - **País** – país onde cada paciente reside;
 - **Sexo** – sexo de cada paciente (masculino ou feminino);
 - **Alta** – alta médica de cada paciente;
 - **Destino** – qual o destino final de cada paciente, após a consulta nas urgências;
 - **Idade** – Idade de cada paciente;

3.2. EXPLORE

Já no ambiente *SAS Enterprise Miner*, o primeiro cuidado a ter foi na análise das variáveis. As decisões tomadas neste ponto têm influência em todos os *outputs* e resultados finais.

Desta forma, a primeira etapa consistiu na análise das variáveis para determinar corretamente o seu “papel” e a sua “natureza”. Como a tabela 3 demonstra, o primeiro passo a tomar foi então a definição do *role* e do *level* de cada uma das variáveis.

Name	Role	Level
Alta	Input	Nominal
Codigo_Postal	Input	Nominal
Concelho	Input	Nominal
Cor	Input	Ordinal
Data_Episodio	Input	Interval
Destino	Input	Nominal
Distrito	Input	Nominal
ID_Cliente	ID	Nominal
Idade	Input	Interval
Localidade	Input	Nominal
Pais	Input	Binary
Sexo	Input	Binary
Ultimo_Fluxograma	Input	Nominal

Tabela 3 – *Role* e *level* de cada variável

Todas as variáveis independentes são consideradas variáveis *input*, à exceção da variável “ID_Cliente”, sendo única e que toma valores exclusivos para cada paciente. As variáveis numéricas contínuas foram definidas com o tipo *Interval*, enquanto a variável “Cor” que pressupunha uma ordem clara e distinta foi definida como *Ordinal*. As variáveis não numéricas foram naturalmente definidas como *Nominal* e as binárias como *Binary*. País encontra-se como *Binary*, pois para esta BD só existem duas opções, “Portugal” e “Desconhecido”. Caso houvesse mais opções, esta seria uma variável Nominal.

Ao longo do projeto, as variáveis serão classificadas em variáveis intervalares e variáveis de classe. As primeiras, tal como o nome indica, apresentam as variáveis intervalares (“Data Episódio” e “Idade”), enquanto as variáveis de classe referem-se a variáveis nominais, binárias e ordinais.

No *software* WEKA, começou por se fazer uma análise das variáveis e perceber que dados eram necessários tratar. Este *software* atribui automaticamente o tipo de variáveis, havendo desta forma 3 variáveis numéricas (“IDCliente”, “Código Postal” e “Idade”) e 10 nominais (as restantes variáveis). Nas variáveis numéricas apresenta-se o valor mínimo e máximo, média e desvio padrão, dando uma perceção dos dados a tratar futuramente. Nas variáveis nominais, são apresentados todos os valores e a quantidade de vezes que aparecem. De seguida, segue-se o exemplo da variável “ID Cliente” como variável numérica e a variável Cor como variável nominal, nas figuras 5 e 6, respetivamente.

Selected attribute	
Name: ID Cliente	Type: Numeric
Missing: 0 (0%)	Distinct: 118262
	Unique: 118262 (100%)
Statistic	Value
Minimum	1
Maximum	118262
Mean	59131.5
StdDev	34139.443

Figura 5 - Variável “ID Cliente” presente no WEKA

Selected attribute		
Name: Cor		Type: Nominal
Missing: 0 (0%)		Distinct: 7
		Unique: 0 (0%)
No.	Label	Count
1	Laranja	12008
2	Amarelo	77489
3	Verde	22728
4	Azul	674
5	Branco	2685
6	Vermelho	949
7	Desconhecido	1729

Figura 6 - Variável “Cor” presente no WEKA

3.3. MODIFY

A fase Modify é uma das mais importantes, tratando-se da fase de pré-processamento dos dados, caracterizada pela limpeza dos dados, ou seja, a identificação e tratamento de *missing values* e a identificação e remoção de valores raros e *outliers*.

Começando pelo SAS Enterprise Miner, através da figura 34 dos Anexos, pode-se observar o número de *missing values* para cada variável. Esta situação acontece quando os pacientes não querem dar a respetiva informação ou quando existe uma falha no sistema. Para o tratamento destes *missing values* nas variáveis de classe utilizou-se o nó *Replacement*, que tem como objetivo principal preencher todas as observações em falta (Breu, Guggenbichler, & Wollmann, 2008). As opções que existem para o tratamento destes valores nas variáveis de classe são a exclusão da variável, a exclusão dos registos, o preenchimento automático com média, moda ou mediana e regressão linear. Para esta situação, optou-se por corrigir a existência de valores em branco por uma denominação determinada por nós, neste caso foi escolhida a palavra “Desconhecido”, tal como se pode ver na Figura 7. Na figura 35 dos Anexos é apresentado um exemplo do *Replacement* da variável “Alta”, exemplificando como fica o resultado final.

```
Replacement Values for Class Variables
```

Variable	Formatted Value	Type	Character Unformatted Value	Numeric Value	Replacement Value	Label
Alta		C		.	Desconhecido	
Concelho		C		.	Desconhecido	
Cor		C		.	Desconhecido	
Destino		C		.	Desconhecido	
Distrito		C		.	Desconhecido	
Localidade		C		.	Desconhecido	
Pais		C		.	Desconhecido	
Sexo		C		.	Desconhecido	
Ultimo_Fluxograma__Razao		C		.	Desconhecido	

```
*-----*  
* Report Output  
*-----*
```

Figura 7 - Replacement das variáveis de classe

Relativamente aos *missing values* das variáveis intervalares, utilizou-se o nó *Impute*. Este nó permite preencher os campos onde se observem *missing values*, utilizando a média, mediana ou desvio-padrão (Breu et al., 2008). Neste caso optou-se pela média para a variável intervalar “Idade”. De seguida apresenta-se o resultado final no tratamento destes *missing values*.

Após este processo, novas variáveis foram criadas contemplando todas as modificações realizadas, tendo sido automaticamente rejeitadas as variáveis originais do modelo. Estas variáveis possuem todos os valores originais com a exceção das alterações impostas pelo tratamento. Tal como se pode

observar na figura 36 dos Anexos e na figura 8 que se segue, as novas variáveis seguem uma nova nomenclatura onde é identificado o processo a partir do qual foram geradas, sendo “REP” indicativo da intervenção do nó *Replacement* nas variáveis de classe e “IMP” do nó *Impute* nas variáveis intervalares.

Imputation Summary							
Number Of Observations							
Variable Name	Impute Method	Imputed Variable	Impute Value	Role	Measurement Level	Label	Number of Missing for TRAIN
Idade	MEAN	IMP_Idade	54.3432	INPUT	INTERVAL		8

Figura 8 - Impute na Variável “Idade”

Seguiu-se o tratamento dos *outliers* sendo este realizado pelo nó *Filter*. Começando pelas variáveis de classe, os filtros presentes correspondem a pequenas incorreções que apareceram na conversão do ficheiro Excel para ficheiro SAS. Não é uma situação que costume acontecer regularmente nas conversões, mas que é inevitável de acontecer. Segue-se um exemplo na figura 9 de uma incorreção que surgiu na variável “REP_Concelho”. Neste exemplo aparece o valor 73, o que não faz sentido pois esta variável corresponde a todos os concelhos de onde pacientes do hospital pertencem. Neste sentido, foi aplicado um filtro para eliminar este valor dos registos. Na figura 37 dos Anexos estão apresentados os restantes filtros que foram aplicados às variáveis de classe.

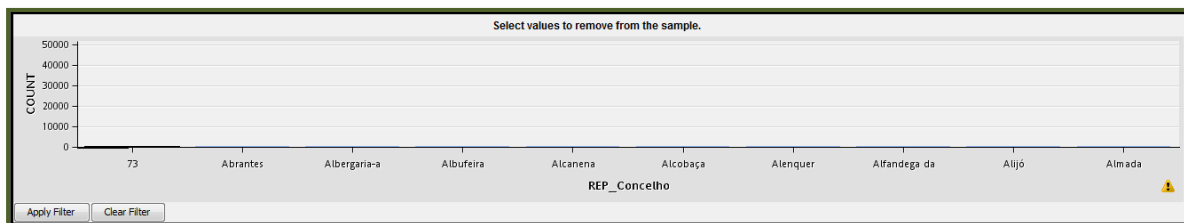


Figura 9 - Filtro aplicado à variável “REP_Concelho”

Nas variáveis intervalares, apenas a variável *Idade* é que apresentava *outliers*. Como se pode se observar na figura 10, na “*Idade*” aparecem valores menores que 0, o que não faz sentido numa variável deste género. Neste sentido, aplicou-se um filtro a indicar que só seriam considerados valores entre 0 e 107. Apesar dos extremos se apresentarem com menor número, não se considerou retirá-los, pois trata-se de urgências onde aparecem pessoas de todas as idades.

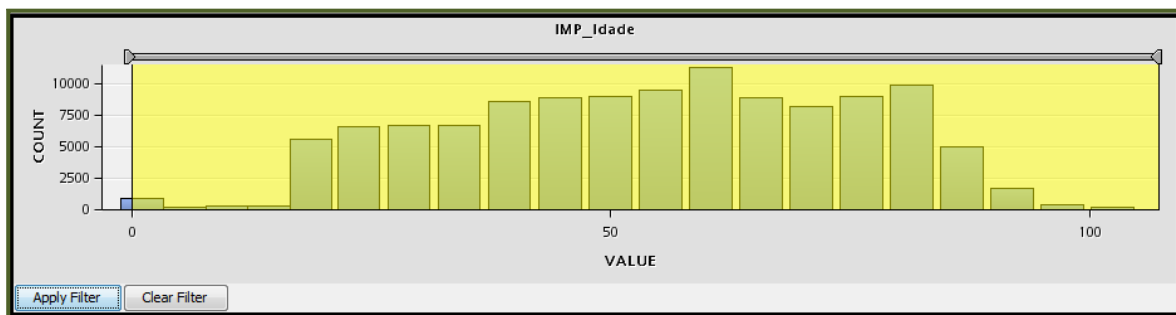


Figura 10 - Filtro aplicado à variável “IMP_Idade”

Após este tratamento inicial dos dados, colocaram-se outras questões que são importantes ter em conta. Na base de dados inicial, 33532 registos dos 118262 encontram-se como “Desconhecidos” para as variáveis “Localidade”, “Concelho”, “Código”, “Distrito” e “País”. Após alguma análise e pesquisa, acabou-se por se retirar estes registos “Desconhecidos” para as fases seguintes do projeto. Esta decisão baseou-se em diversos fatores, sendo que o primeiro teve em atenção o objetivo principal desta etapa: a formação de *clusters*. O objetivo é compreender como os dados se agrupam e se interligam entre si; estando estes registos presentes, a sua inserção na análise não trás mais-valias nem resultados conclusivos.

Para além deste ponto, considerou-se outros dois fatores como determinantes para a retirada destes 33532 registos “Desconhecidos”. Considerando os 33532 registos na análise, na fase de formação dos *clusters* o nó correspondente não consegue suportar os 118262 registos para a formação dos mesmos. Neste sentido, foi obrigatório criar uma amostra de 99999 registos, pois era o número máximo aceite para que a formação dos *clusters* se realizasse com sucesso. Visto não ser possível utilizar todos os dados, a utilização de dados “Desconhecidos” torna mais pobre a análise do problema e com menos resultados concretos.

Para além disso, experimentou-se utilizar uma amostra de 99999 registos (considerando os 33532 registos da BD total), e observou-se que todos os registos “Desconhecidos” ficaram num *cluster* à parte dos restantes. Neste sentido, este *cluster* não acrescenta valor aos resultados, perante o objetivo em questão.

Neste sentido, a opção mais viável foi a retirada destes 33532 registos “Desconhecidos”.

Por outro lado e em casos normais, apenas se devem retirar dados como *outliers* no máximo 5% do total de registos da BD total. No entanto, esta situação é de carácter excepcional, retirando-se assim 34858 registos (29,47% dos dados totais), como se pode observar na figura 11.

Number Of Observations			
Data	Filtered	Excluded	DATA
TRAIN	83410	34858	118268

Figura 11 - Número de variáveis excluídas

Segue-se o exemplo da variável “REP_Distrito”, na figura 12, onde se mostra o filtro aplicado nos registos “Desconhecidos”. Para além desta variável, como se pode ver anteriormente no nó *Replacement* foram descritas diversas situações como “Desconhecidas”, para o tratamento dos *missing values*, por isso esta situação repete-se para as restantes variáveis.

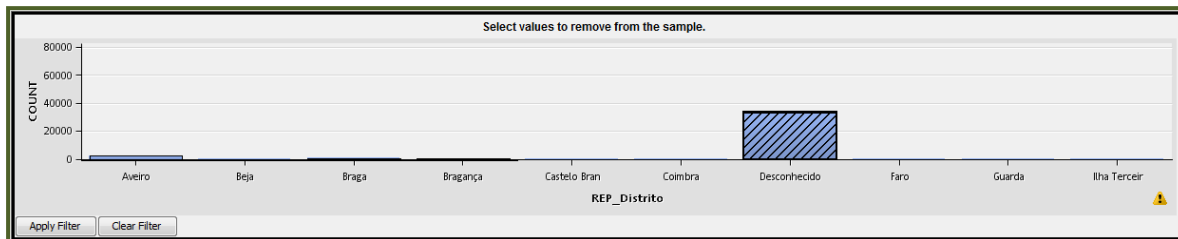


Figura 12 - Variáveis excluídas

Na figura 38 dos Anexos estão representadas todas as variáveis que foram excluídas através dos filtros.

Relativamente ao *software* WEKA, para a criação de regras de associação foi necessário converter todas as variáveis numéricas para nominais, pois só desta forma foi possível criar as regras. A conversão foi feita no próprio *software*, através de um filtro designado “NumericToNominal” (consultar figura 39 dos Anexos). Após a conversão, a variável “Código Postal” ficou da seguinte forma, como se pode observar na figura 13:

Selected attribute		
Name: Código Postal		Type: Nominal
Missing: 0 (0%)		Unique: 63 (0%)
		Distinct: 308
No.	Label	Count
1	-1	33532
2	1000	12
3	1050	2
4	1100	5
5	1170	1
6	1200	5
7	1300	5
8	1350	3
9	1495	4
10	1500	4
11	1600	3
12	1675	2

Figura 13 - Variável “Código Postal” após a conversão

Na fase seguinte recorreu-se a uma análise crítica para a escolha das variáveis na criação das regras. Nem todas as variáveis são uteis para as regras, pois estas poderão estar relacionadas (exemplo: as variáveis “Localidade”, “Conselho”, “Distrito” e “País”) ou simplesmente não trazerem quaisquer vantagens para a análise do problema.

A variável “País” foi logo retirada, uma vez que através das outras variáveis relativas à região consegue-se perceber qual é o país. Caso se mantivesse, iriam criar-se regras óbvias e que não traziam valor ao estudo. Havendo 4 variáveis referentes à localização dos pacientes, não foi necessário a presença de todas as variáveis. Neste sentido, manteve-se a variável “Distrito”, eliminando a “Localidade” e o “Concelho”. A variável “Código Postal” também não trazia mais-valias para o estudo e está interligada com todas as outras variáveis referidas anteriormente, sendo igualmente eliminada. A variável “ID Cliente” também foi retirada, pois é única para cada paciente e desta forma nunca seria utilizada nas regras. Por fim, a variável “Destino” também foi retirada da análise, pois através de análises posteriores, observou-se que esta estava altamente correlacionada com a variável Alta, criando apenas regras óbvias. Desta forma, e para evitar o aparecimento destas regras que iriam “introduzir ruído”, resolveu-se retirá-las da análise.

Com esta eliminação, fica-se com as seguintes variáveis, como se pode observar na figura 14:

No.	Name
1	<input checked="" type="checkbox"/> Data Episódio
2	<input type="checkbox"/> Último Fluxograma / Razão
3	<input type="checkbox"/> Cor
4	<input type="checkbox"/> Distrito
5	<input type="checkbox"/> Sexo
6	<input type="checkbox"/> Alta
7	<input type="checkbox"/> Idade

Figura 14 - Variáveis finais no *software* WEKA

3.4. MODEL

Após o pré-processamento dos dados no *SAS Enterprise Miner*, entra-se na última fase de análise neste *software* para a construção dos *clusters*. Para completar este processo, o nó a utilizar foi o nó *Cluster*.

Para compreender melhor a situação atual das urgências, criou-se um *cluster* com todas as variáveis importantes, tendo como objetivo principal perceber a existência de algum padrão interessante nos dados que ajudasse na resolução do problema apresentado. Para se chegar ao número de *clusters* ideal, utilizou-se o nó *SAS Code* para a criação do gráfico de cotovelo. Este gráfico tem como objetivo perceber a evolução da variabilidade intra-*cluster*, isto é, a soma da distância de todos os indivíduos do *centróide* ao *cluster*. O número ideal de *clusters* obtém-se quando o ponto a partir do qual o declive passa a ser menos acentuado. Na figura 40 dos Anexos pode-se ver o código utilizado para a criação do mesmo.

Neste sentido, foram selecionadas todas as variáveis exceto “ID_Cliente”, “IMP_Codigo Postal”, “REP_Concelho”, “REP_Localidade” e “REP_Pais”. Estas últimas não foram incluídas pois não traziam uma mais-valia ao *cluster*, visto a residência dos pacientes já estar representada pela variável “REP_Distrito”. Executou-se o nó *Cluster* para 1 a 10 *clusters*, obtendo o seguinte gráfico de cotovelo, abaixo. Através deste gráfico representado na figura 15, percebe-se que não foi uma escolha fácil, pois à medida que o número de *clusters* aumenta, a variabilidade *intra-clusters* não se apresenta sempre estável (por exemplo com 3 *clusters* a distância ao *centróide* é muito menor do que com 4 *clusters*). Conclui-se assim que o número ideal de *clusters* serão 3.

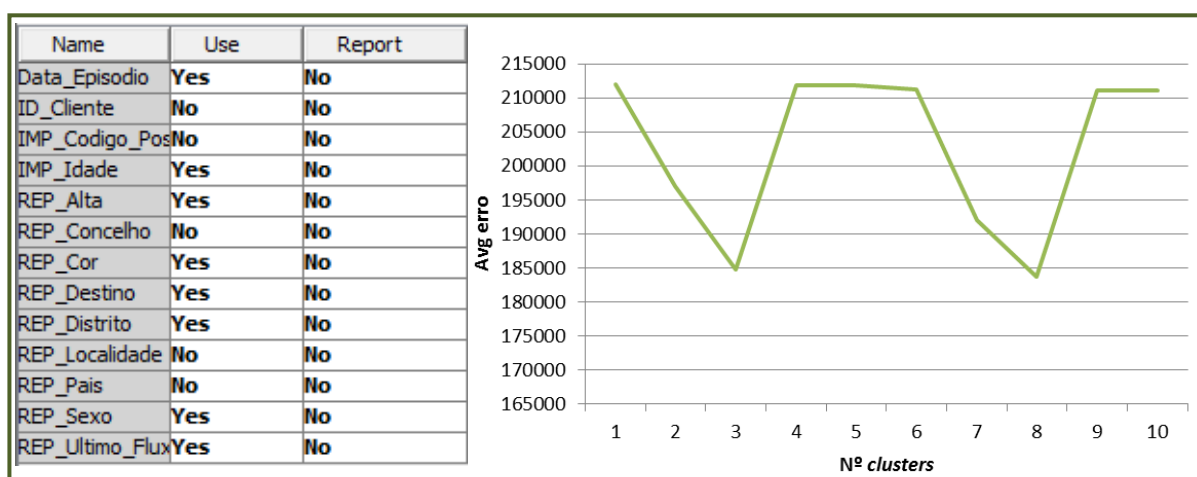


Figura 15 – Escolha do número de *clusters* ideal

Relativamente ao *software* WEKA, após o tratamento dos dados e das variáveis, estes estavam aptos a serem utilizados para a criação das regras de associação. Neste *software*, o separador *Associate* é o suposto para a formação dos mesmos.

Começou por se escolher o tipo de algoritmo a processar, estando já por defeito o algoritmo *Apriori*, havendo no entanto outras opções como o *FilteredAssociator*, *PredictiveApriori* e *Tertius*. Neste caso, foi utilizado o algoritmo por defeito, que é o mais utilizado e o mais conhecido.

Importante também foi escolher as configurações pretendidas, pois estas influenciavam o tipo de regras que iriam surgir. As configurações definidas foram as apresentadas na figura 16, tendo apenas alterado os parâmetros *minMetric*, *numRules*, *outputItemSets* e *removeAllMissingValues* (na figura 41 dos Anexos pode-se observar as configurações pré-definidas). Estes parâmetros têm uma grande influência na criação das regras. *MetricType* indica as opções de métricas que podem ser utilizadas para a criação das regras, podendo optar pela *Confidence*, *Leverage* e *Lift*; *minMetric* especifica qual o valor mínimo da métrica que poderá aparecer, *numRules* identifica o número de regras a criar e *removeAllMissingCols* afirma se se retira ou não todas as colunas a vazio.

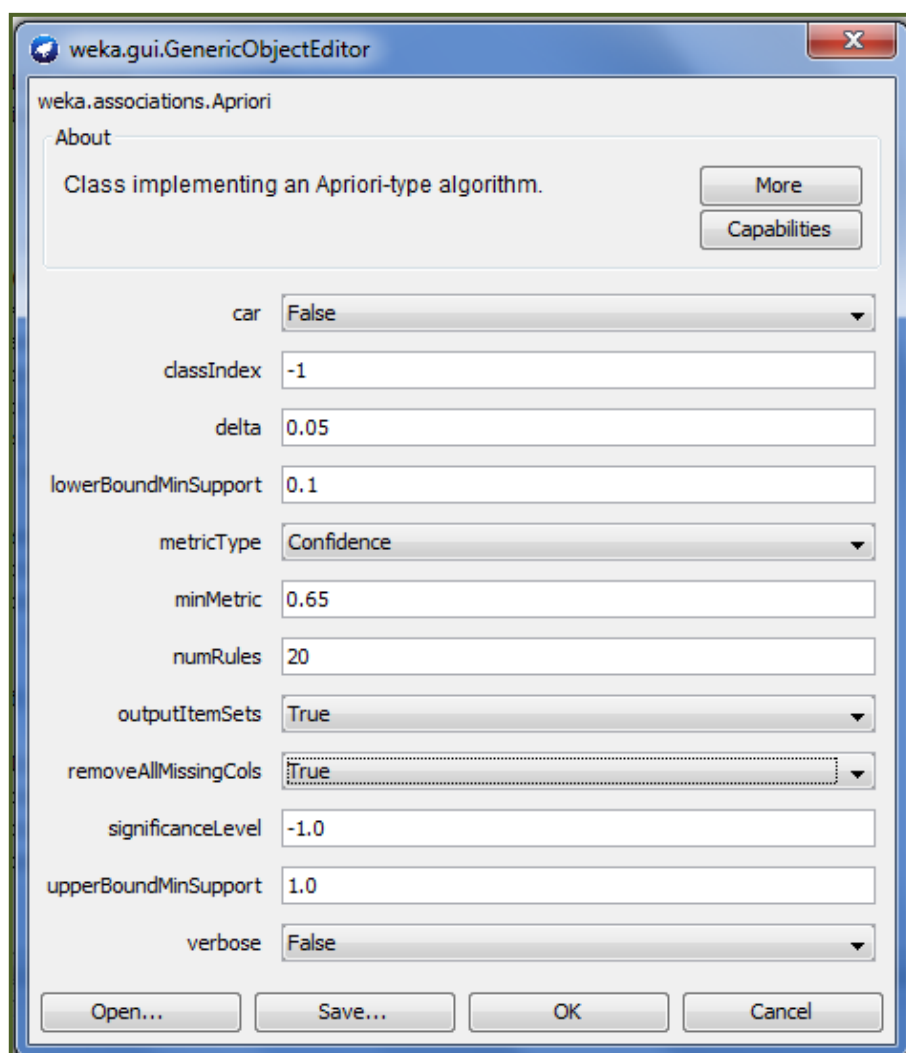


Figura 16 – Configurações no *software* WEKA para as regras de associação

Como se pode observar, o tipo de métrica escolhida foi a “Confiança” de forma a compreender número de vezes que as regras geradas estavam corretas. Sendo esta uma das métricas mais

utilizadas e que faz mais sentido no contexto do problema, foi a única utilizada. O valor mínimo na métrica foi 0.65, pois a maioria das regras encontradas estão com uma Confiança na média dos 0.67, como se poderá ver no capítulo da Discussão de Resultados. O número de regras geradas foram 20, uma vez que foi o limite estabelecido de regras necessárias para analisar. Por último, foi importante remover todas as colunas com *missing values* de forma a não enviesar os resultados, e expor os *ItemSets* que irão ajudar na sua análise de conclusões de resultados.

3.5. ASSESS

Este último subcapítulo refere-se à discussão dos resultados obtidos anteriormente. Esta análise será feita no próximo capítulo, próprio para a Discussão dos Resultados.

4. RESULTADOS E DISCUSSÃO

Neste capítulo são apresentados os resultados da análise de *clusters* e das regras de associação, aplicadas aos dados das urgências de um hospital no norte do país.

Relativamente à análise de *clusters*, analisou-se os dados perante os *clusters* formados, com o objetivo de perceber se foram alcançados novos padrões e que conclusões interessantes se retiram daí, tendo em vista o objetivo final.

Tal como se verificou anteriormente na Metodologia, executou-se o nó com todas as variáveis importantes ao problema, obtendo 3 *clusters* como número ideal de *clusters*. Segue-se o gráfico *Pie Chart* da figura 17 com o número de pacientes em cada *cluster*:

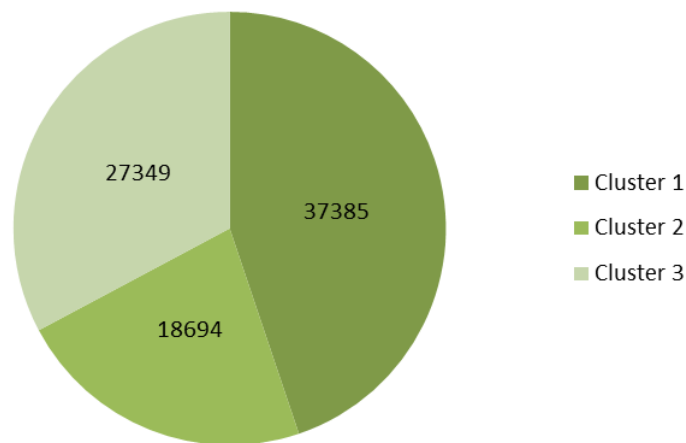


Figura 17 – Número de pacientes em cada *cluster*

O fator de separação dos dados em 3 *clusters* foi a variável relativa ao sexo dos pacientes (sexo feminino e masculino) e a sua idade. Com a ajuda das tabelas que se seguem, percebe-se que o *cluster 1* é constituído por pacientes do sexo masculino de todas as idades; o *cluster 2* em elementos do sexo masculino mais jovens e por elementos do sexo feminino dos 0 aos 60 anos de idade e por fim o *cluster 3* possui pacientes femininos a partir dos 40 anos, apesar de a sua maioria ser a partir dos 60 anos.

FEMININO	Cluster1	Cluster2	Cluster3
0-10 anos	-	100	-
11-20 anos	-	1675	-
21-30 anos	-	4652	-
31-40 anos	-	4578	-
41-50 anos	-	5349	383
51-60 anos	-	2265	4683
61-70 anos	-	-	7017
71-80 anos	-	-	7967
81-90 anos	-	-	6293
91-100 anos	-	-	986
101-110 anos	-	-	20

MASCULINO	Cluster1	Cluster2	Cluster3
0-10 anos	63	64	-
11-20 anos	1238	11	-
21-30 anos	4147	-	-
31-40 anos	4518	-	-
41-50 anos	5817	-	-
51-60 anos	6257	-	-
61-70 anos	5902	-	-
71-80 anos	5790	-	-
81-90 anos	3306	-	-
91-100 anos	343	-	-
101-110 anos	4	-	-

Tabela 4 – Número de pacientes em cada *cluster*

Começando por analisar a variável relacionada com a data da ida às urgências, com a ajuda do gráfico da figura 18 observa-se que o número de pacientes ao longo dos meses foram muito semelhantes entre si, havendo apenas uma pequena quebra maior no mês de Novembro para todos os *clusters*. Também se repara que os *clusters* 1 e 3 têm um número significativamente maior de pacientes comparado com o *cluster* 2, reforçando que existe uma maior percentagem de pessoas mais idosas que recorreram às Urgências Hospitalares. Para a consulta dos valores mais detalhados, consultar a tabela 5 dos Anexos.

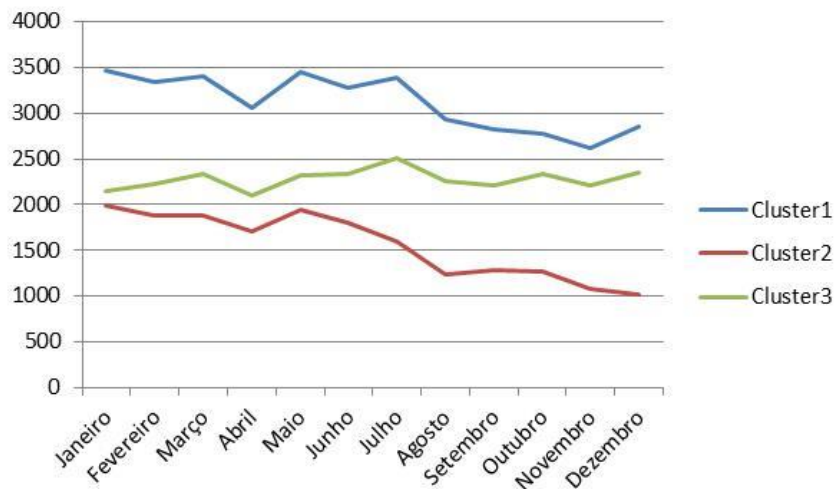


Figura 18 - Gráfico relativo ao número de pacientes em cada mês

Analisando agora as variáveis referentes à “Alta Médica” e ao “Destino Final”, presentes nas figuras 20 à 26, é possível observar um conjunto de características que se destacam. Estas variáveis podem ser analisadas em conjunto, pois estão relacionadas entre si. Desta forma, os seus gráficos aparecem também relacionados. Em Anexo, pode-se consultar as tabelas referentes a estes gráficos, com a informação mais pormenorizada (tabela 6).

Numa primeira análise apenas com a variável “Alta médica”, presente na figura 19 e com ajuda da tabela 7 dos Anexos, observa-se que a maior parte dos pacientes tiveram “Alta para o Domicílio”, com 52003 casos, indicando que o seu estado de saúde não necessitou de mais cuidados médicos imediatos. Também se verificou 10236 casos cujos pacientes foram transferidos para outros “Centros de Saúde” e 8419 casos tiveram mesmo que ser internados no hospital. No caso da transferência dos pacientes para o “Centro de Saúde”, os valores são muito idênticos entre *clusters*, mas no caso dos “Serviços de Internamento”, o sexo masculino destacou-se com 4272 casos. Um dado que salta à vista é o número de internamentos que se destacaram no *cluster* 1 e 3, significativamente maiores que no *cluster* 2. Esta é outra evidência que comprova que os pacientes mais idosos são os mais debilitados, estando mais expostos às doenças que levam ao internamento. Também se conclui que o *cluster* 2 é o que contém os elementos mais jovens, sendo normal, por ser previsível, haver menos pacientes que necessitam de internamento.

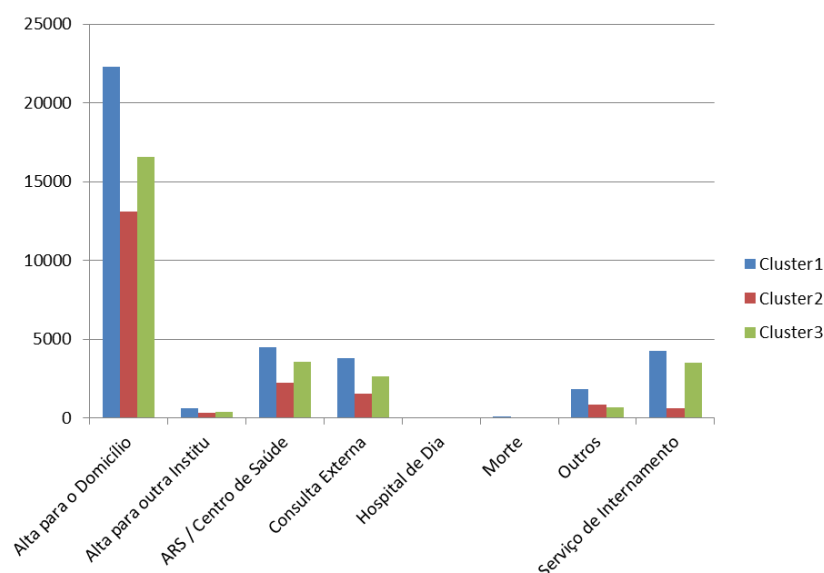


Figura 19 - Gráfico referente à Variável “Alta Médica”

Olhando ainda com mais pormenor para estes dados, observa-se que os pacientes que tiveram como destino final “Consulta Externa”, estavam na sua maioria relacionados com ortopedia e neurologia. Para além disso, existem diversas áreas presentes nestas consultas externas, tais como as consultas para o sono ou ventilo terapia, tendo sido apenas atendida uma pessoa em cada uma destas áreas. Relativamente ao “Hospital de Dia”, foram poucos os pacientes que tiveram como “Destino final” esta opção, tendo-se apenas verificado 79 casos; destes 25 estiveram em cirurgia ambulatória e 23 em hematologia. Verificou-se também a existência de uma grande percentagem de pacientes que abandonaram os serviços de urgências sem que lhes fosse dada a permissão de saída, 1458 casos, ou então que não responderam à chamada, com 1297 casos, presumindo-se que abandonaram as urgências após inscrição para atendimento. Por fim, no “Serviço de Internamento”, destaca-se a “Área de Decisão Clínica/SO”, um espaço onde os pacientes aguardam o resultado de exames complementares de diagnóstico, a observação de especialistas e a transferência para um serviço de internamento, com 2970 casos, sendo este o “Destino final” com maior número de pacientes.

Investigando a diferença de pacientes entre os *clusters*, observa-se um padrão comum que se repete para as várias “Altas Médicas” e respetivos “Destinos finais”. Para a maioria das doenças verifica-se um maior número de pacientes do *cluster 1* comparado com os restantes *clusters*, tirando alguns casos excecionais. No caso das “Consultas Externas” com consulta em Ortopedia, verificou-se uma maior afluência de pacientes do *cluster 3*, indicando que é uma especialidade com mais cuidados para o sexo feminino com mais idade. Outra exceção verificada são as “Consultas Externas” para Saúde Ocupacional, uma área de intervenção prioritária que valoriza o local de trabalho como espaço privilegiado para a prevenção, proteção e promoção da saúde e o acesso aos serviços de saúde dos trabalhadores, verificando-se uma maior afluência do *cluster 2*. Sendo este *cluster* constituído por elementos do sexo feminino com a idade correspondente à idade de trabalho, faz sentido uma maior afluência por parte destes pacientes.

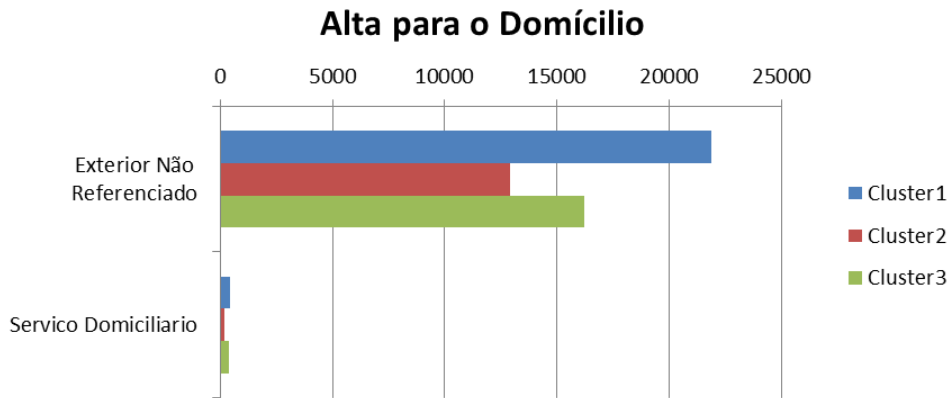


Figura 20 – Número de pacientes para “Alta para o Domicílio” e respetivos “Destino Final”

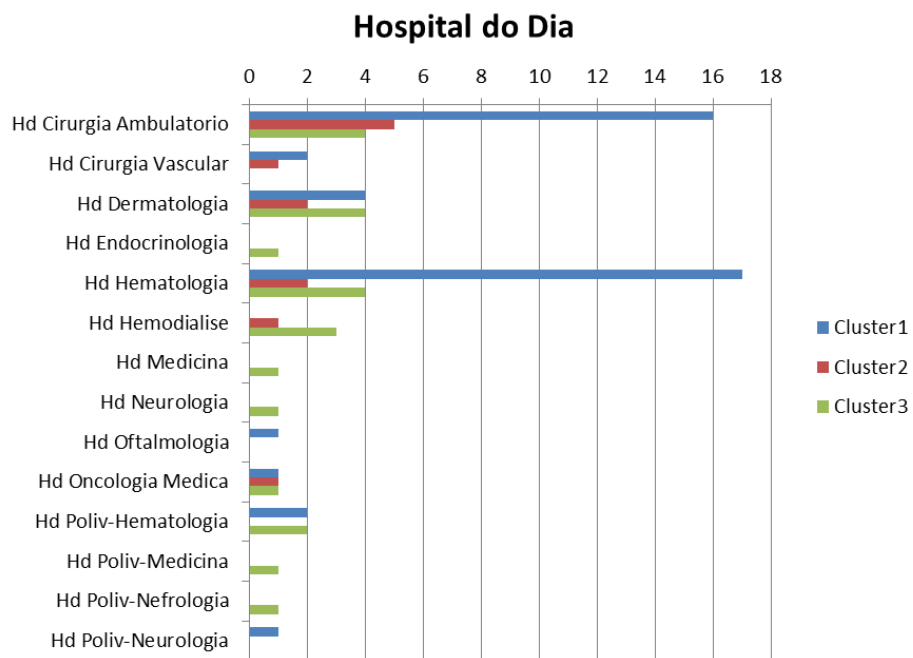


Figura 21- Número de pacientes para “Hospital do Dia” e respetivos “Destino Final”

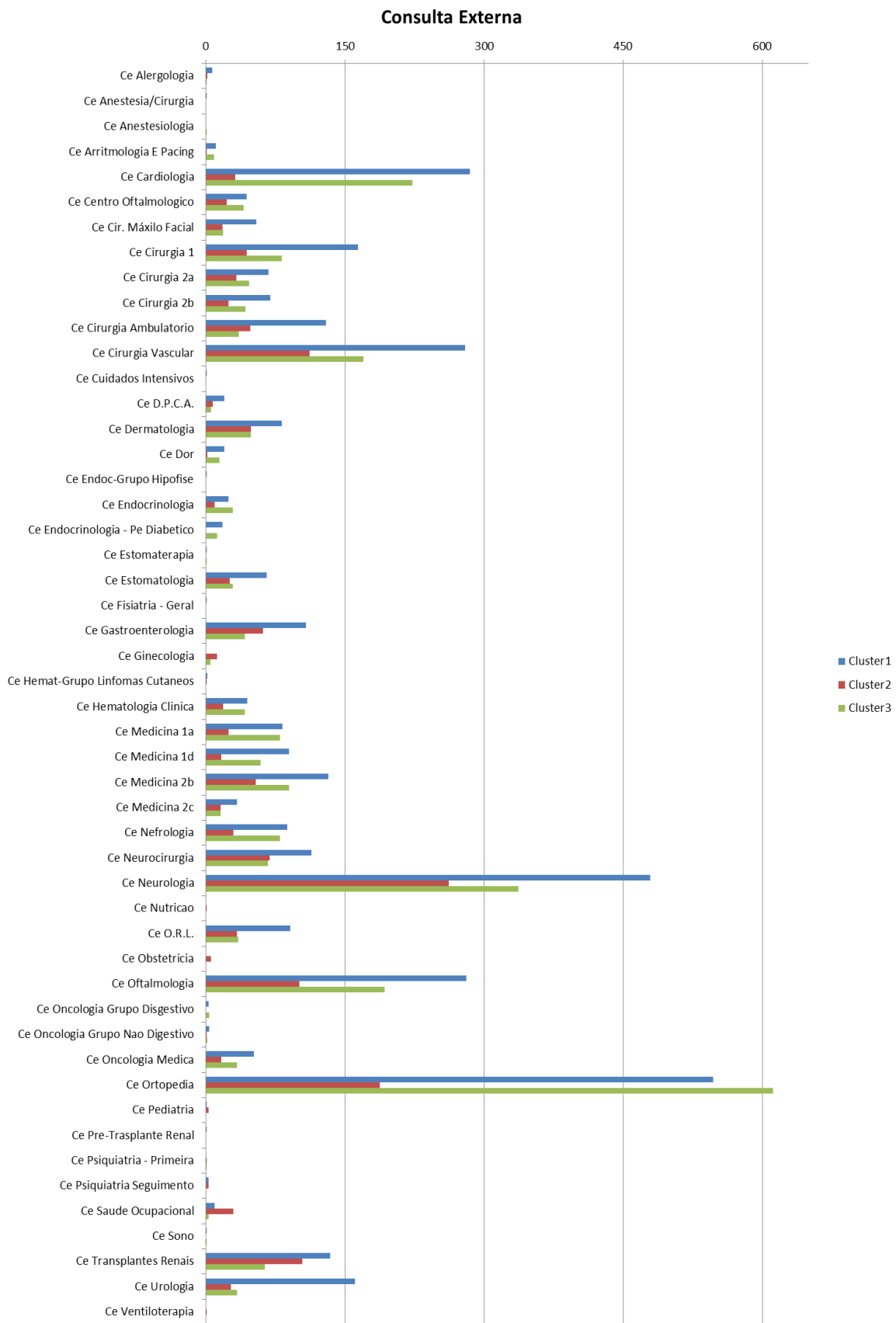


Figura 22 - Número de pacientes para “Consulta Externa” e respetivos “Destino Final”

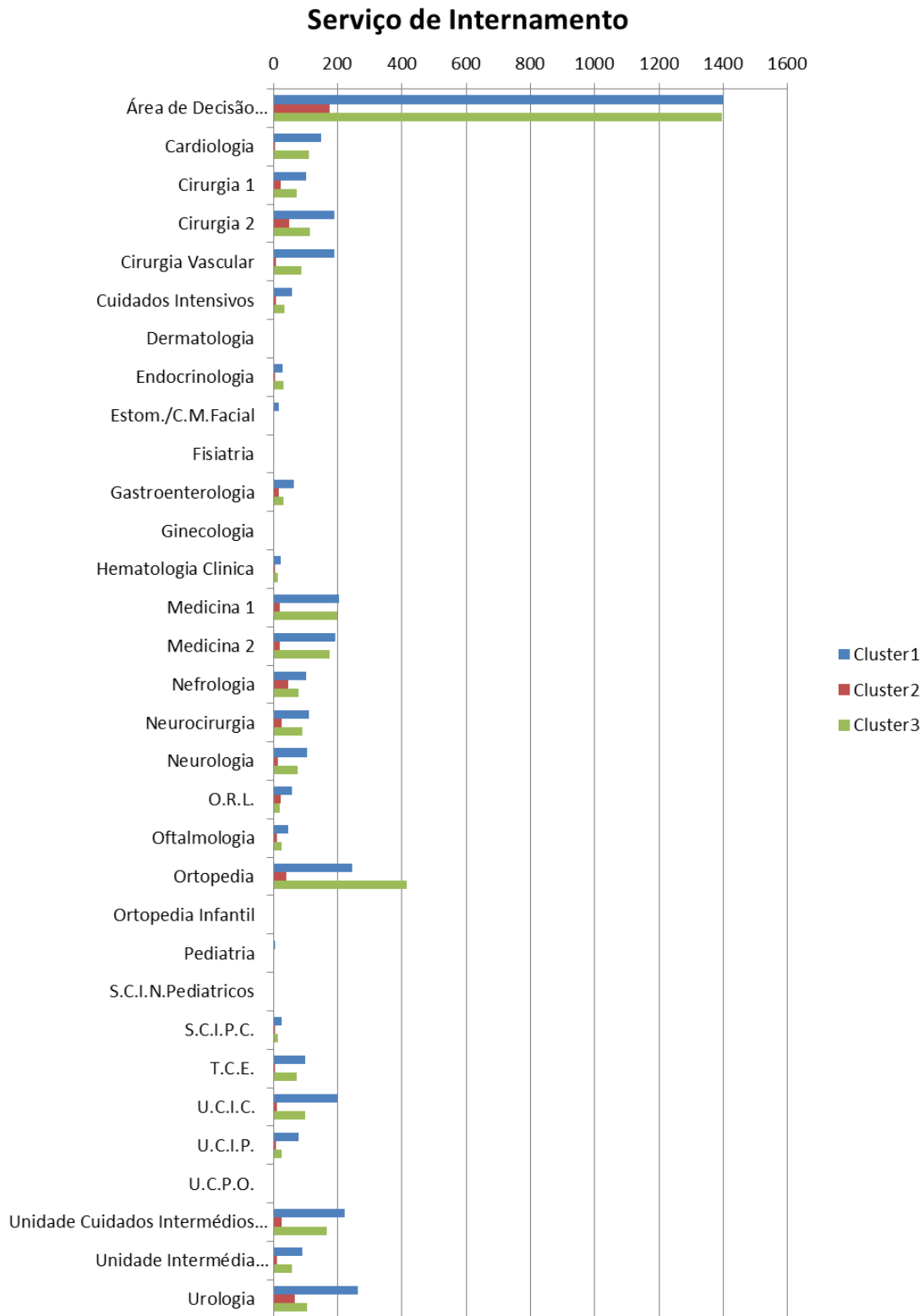


Figura 23 - Número de pacientes para “Serviços de Internamento” e respetivos “Destino Final”

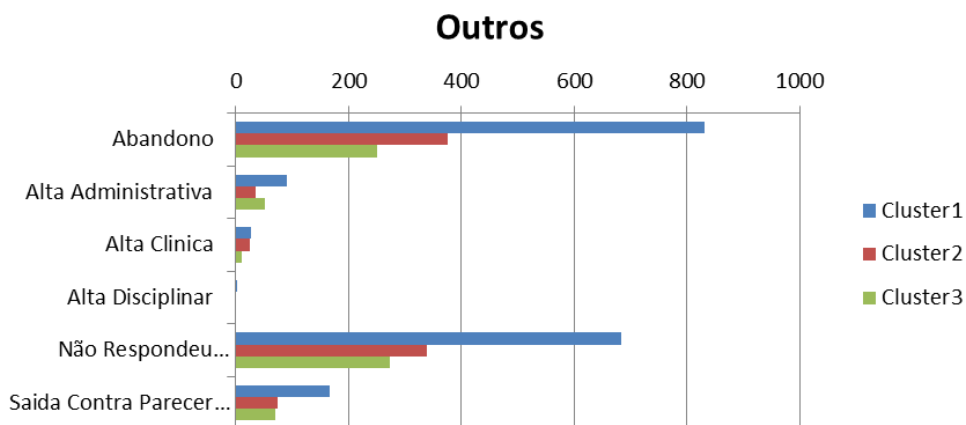


Figura 24 - Número de pacientes para “Outros” e respetivos “Destino Final”

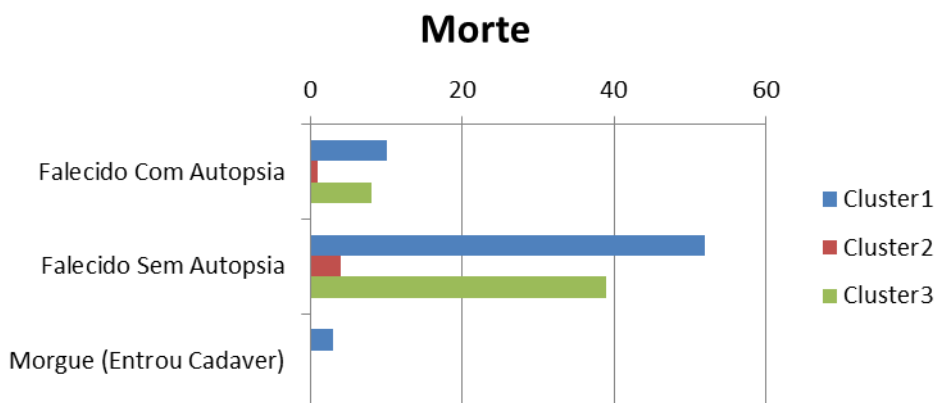


Figura 25 - Número de pacientes para “Morte” e respetivos “Destino Final”

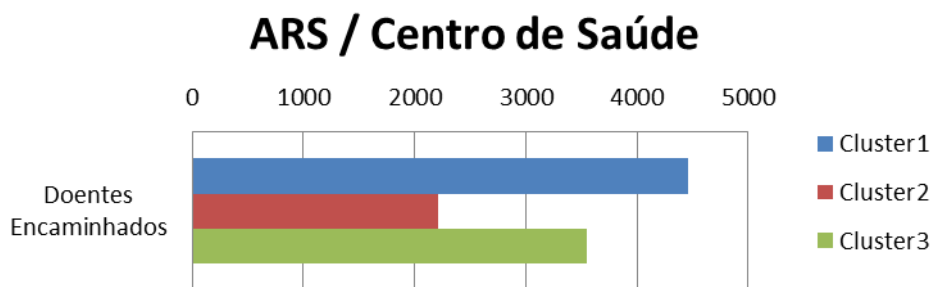


Figura 26 - Número de pacientes para “ARS / Centro de Saúde” e respetivos “Destino Final”

Relativamente à variável sobre o motivo dos pacientes recorrerem às urgências, através do gráfico da figura 27 com um esquema mais simplista (esta figura encontra-se ampliada na figura 42 em Anexo, para uma melhor perceção dos dados, e a informação mais pormenorizada na tabela 8 dos Anexos), observa-se um grande conjunto de doenças, havendo umas que se destacam bastante das restantes. Observa-se que em 2012 surgiu um grande surto de “Cefaleia”, com 2740 casos, “Dispneia”, com 5624 casos, “Dor Abdominal”, com 5682 casos, “Dor Lombar” e com 4399 casos. Também está registado na base de dados que 12011 casos como “Indisposição no adulto” e 14005 casos como “Problemas nos Membros”. Estas últimas duas razões que levam os pacientes a irem às urgências são muito vagas, não sendo perceptível identificar qual a verdadeira doença que os levou a recorrerem a estes serviços. Pode ser todavia o motivo para ter um valor tão elevado de pacientes, por abarcar um leque elevado de doenças não especificadas.

Comparando o número de pacientes entre os 3 *clusters*, os motivos que os levaram às urgências também não variam muito entre si, tirando alguns casos excepcionais. O caso da “Dispneia” encontra-se em grande número nos *clusters* 1 e 3 (com 2545 e 2486 casos respetivamente, contra 592 casos do *cluster* 2), indicando que esta é uma doença que afeta maioritariamente a população mais idosa. Outros com casos semelhantes são a “Dor Torácica” e “Lesão Toráco-abdominal”, onde o *cluster* 2 tem os pacientes que menos sofrem com esta doença. Por outro lado, observa-se que os “Problemas Urinários” encontram-se na sua maioria nos pacientes do sexo masculino (*cluster* 1), apresentando uma grande diferença para os restantes *clusters* (1411 casos contra 570 e 532 casos para o *cluster* 2 e 3, respetivamente).

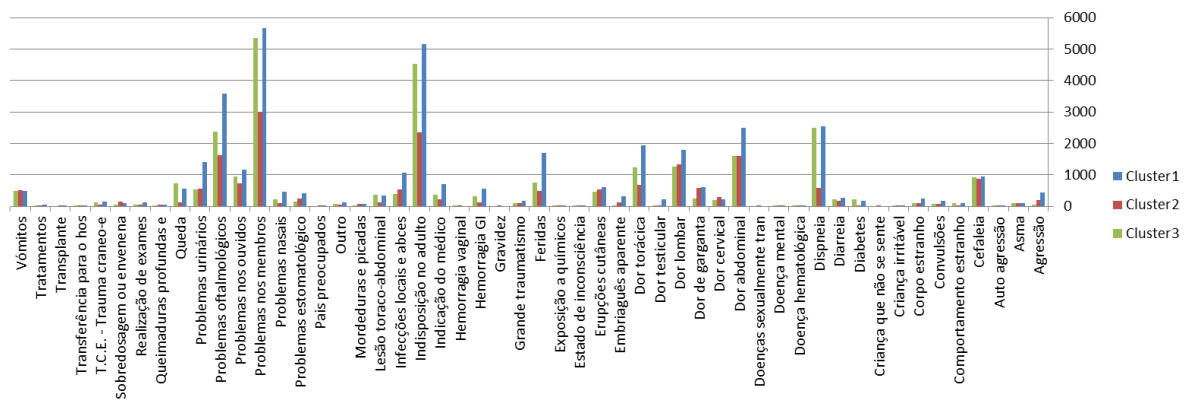


Figura 27 - Gráfico sobre o motivo dos pacientes recorrerem às urgências

Analisando agora o resultado com a variável “Cor” com ajuda da figura 28 (e com a tabela 9 em Anexos com a informação mais pormenorizada), observa-se que a “Cor Amarela” é a que predomina e se destaca mais, com 55224 casos registados, indicando que a maioria dos casos no Serviço de Urgência eram realmente urgentes, necessitando de uma rápida intervenção. Segue-se a “Cor Verde” com 16541 casos e a “Cor Laranja” com 8640 casos.

Observando com um maior detalhe os *clusters*, observa-se que a distribuição entre eles é muito idêntica e vai ao encontro das conclusões retiradas anteriormente: o *cluster 1* é o que tem o maior número de pacientes, indicando que a maior parte dos pacientes eram do sexo masculino com casos urgentes para tratar. Este padrão acontece com todas as cores, começando pelo *cluster 1* com maior número de pacientes, seguindo-se do *cluster 3* e terminando no *cluster 2*.

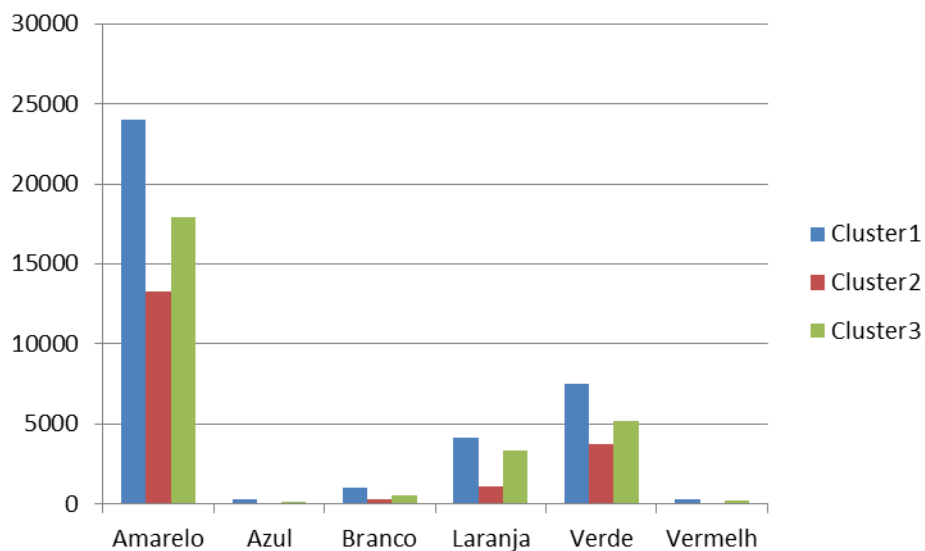


Figura 28 - Gráfico relativo à variável cor das pulseiras

Relativamente ao “Distrito” dos pacientes, pode-se observar na figura 29 (esta encontra-se com maior detalhe na figura 43 (Anexo) e com mais informação na tabela 10 (Anexos)) que a maior parte das doenças encontravam-se no distrito do Porto, tal como era expectável. Esta tabela revela que existe um número muito semelhante de pacientes nos diferentes *clusters*, independentemente do seu sexo e idade. No entanto, pode-se destacar aqui alguns casos que fogem à regra, como é o caso do distrito de Coimbra e Faro, onde apenas se verificam 5 e 2 casos respetivamente no *cluster 3*. Aveiro também se destaca pelo facto de haver um maior número de pacientes do sexo masculino de todas as idades, isto é, um maior número de pacientes no *cluster 1* e 3.

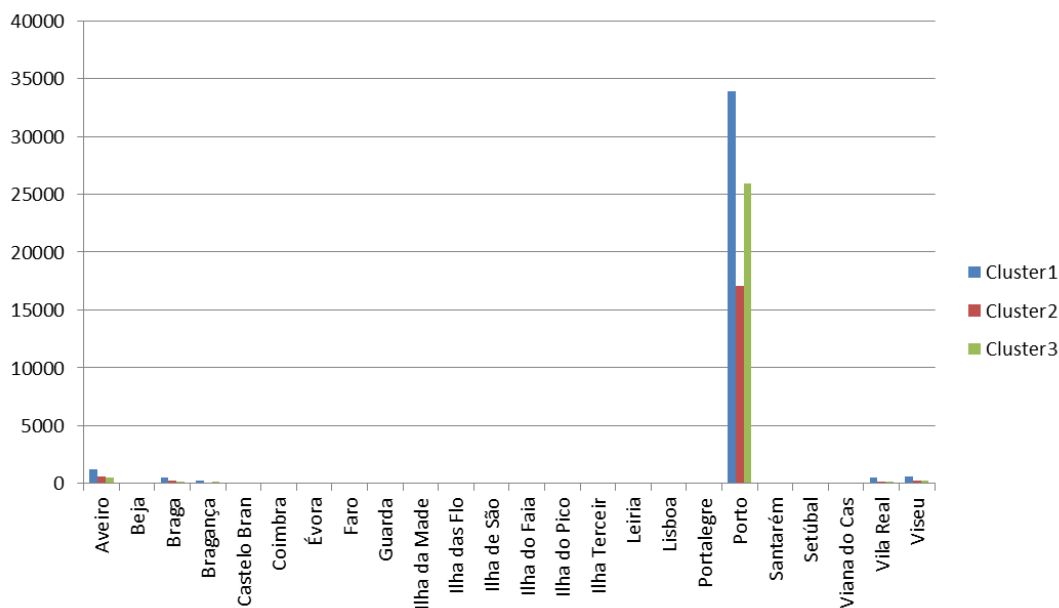
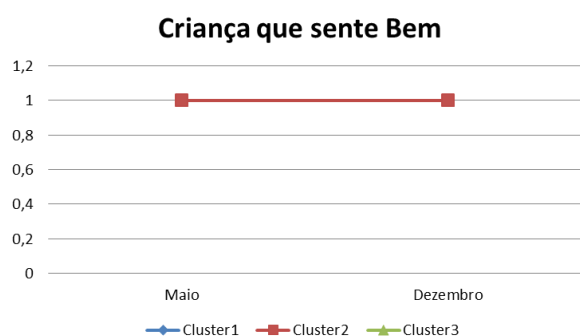
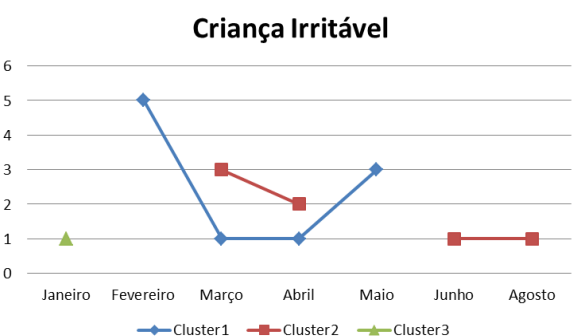
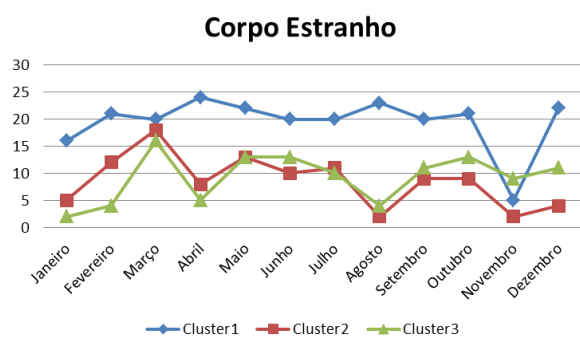
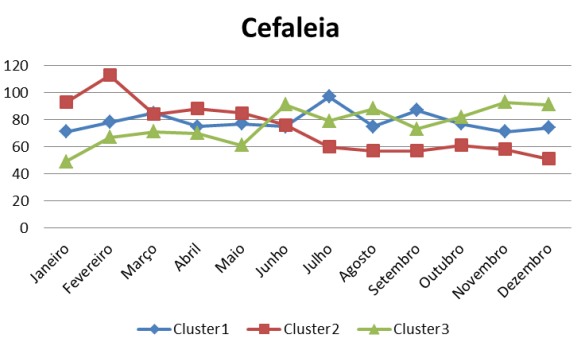
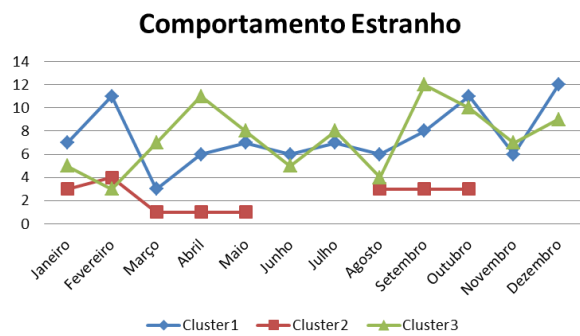
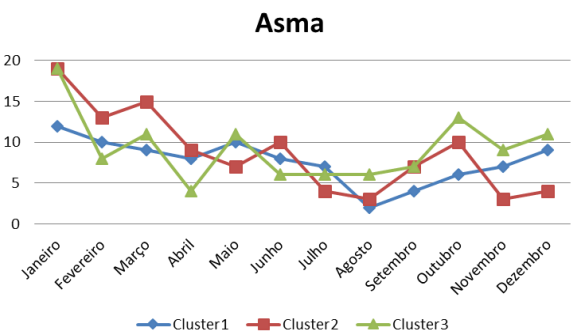
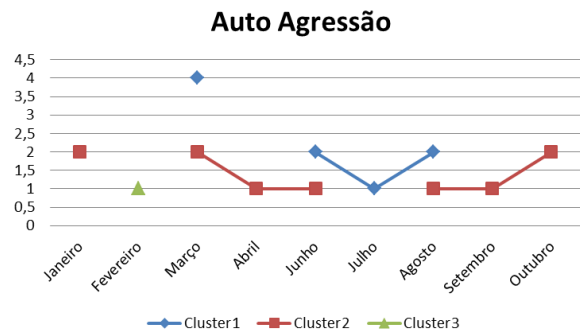
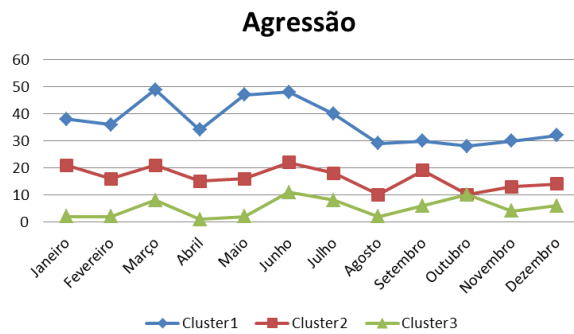


Figura 29 – Gráfico referente ao Distrito dos pacientes

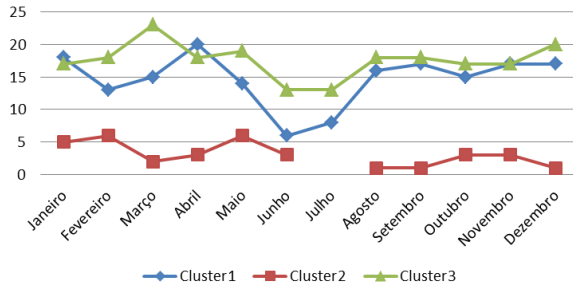
Relacionando agora as diversas variáveis entre si, para tirar conclusões ainda mais precisas, começou-se por analisar se haveria algum padrão interessante entre a razão dos pacientes recorrem às urgências e o mês em que essas idas ocorreram. Ao analisar a figura 30 que se seguem, com ajuda da tabela 11 em Anexo, verificaram-se alguns casos que se destacam. Um dos casos é o facto do número de pacientes com “Asma” ser superior em janeiro quando comparado com restantes meses (50 casos contra uma média de 25 casos nos restantes meses). Seria expectável que esta doença, pelas suas características, tivesse um maior número de casos nos meses da primavera (abril, maio e junho); porém, verificaram-se, neste período 21, 28 e 24 casos, respetivamente. Isto pode ser explicado não com a questão da “Asma Alérgica” mas sim com questões relacionadas com o inverno, onde acontecem as infeções respiratórias, o que provoca também estes sintomas. O número de pacientes entre os *clusters* variam muito nos diferentes meses, mas em janeiro houve um pico de pacientes do sexo feminino de todas as idades e pacientes do sexo masculino mais jovens. Outro caso interessante foi um maior número de casos de “Dispneia” nos meses de janeiro e fevereiro. Sendo a Dispneia conhecida como a “Falta de Ar”, esta vai de acordo com o que se verificou com o caso da “Asma”. O mês de fevereiro conseguiu ter o dobro dos casos que os restantes meses, com 843 casos registados. No entanto ao contrário da “Asma”, verifica-se um maior número de pacientes com “Dispneia” no *cluster* 1 e 3 independentemente dos meses, indicando que a “Dispneia” afeta em maior número os pacientes mais idosos. Estes factos poderão estar relacionados com o facto de ser nesta altura que ocorrem as síndromas gripais, interferindo com estes valores e influenciando-os.

Outro caso que se destaca são as “Erupções Cutâneas”. Apesar da diferença nos vários meses analisados não ser muito notória, existiu um ligeiro aumento de pacientes em maio, junho, julho e agosto. Visto estes serem meses de Verão e atendendo ao facto da pele estar mais exposta ao calor, podemos concluir que o resultado corresponde ao que seria expectável. Analisando ainda o número de pacientes com esta doença nos 3 *clusters*, não há nada de relevante a destacar, visto os valores serem muito idênticos entre si.

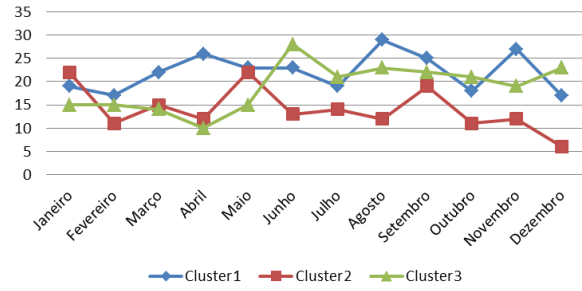
Por último e analisando os gráficos suportadas pelas tabelas, sobressaem as “Mordeduras e Picadas”. O expectável seria haver um aumento significativo nos meses de verão, junho, julho e agosto, com o aumento de insetos, no entanto isso também não se verificou. Existe um pequeno aumento nos meses de junho e julho, mas muito pouco significativo (apenas uma média a mais 5 casos que nos restantes meses). Curioso é constatar que no mês de julho, os pacientes mais afetados por “Mordeduras e Picadas” foram pacientes do *cluster 2* e no mês de julho foram pacientes do *cluster 3*.



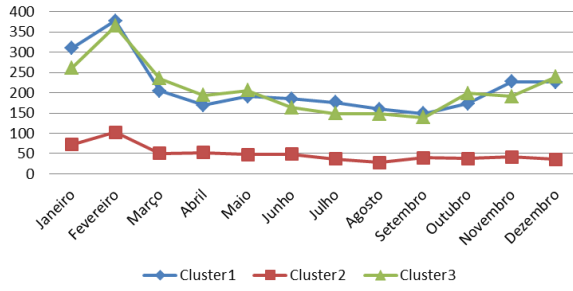
Diabetes



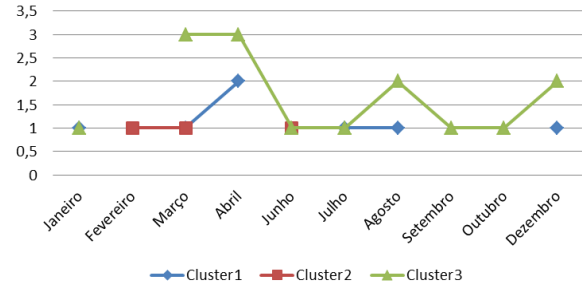
Diarreia



Dispneia



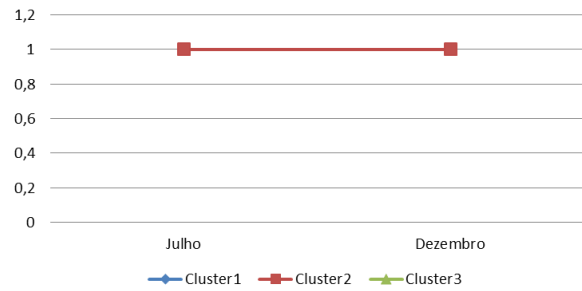
Doença Hematológica



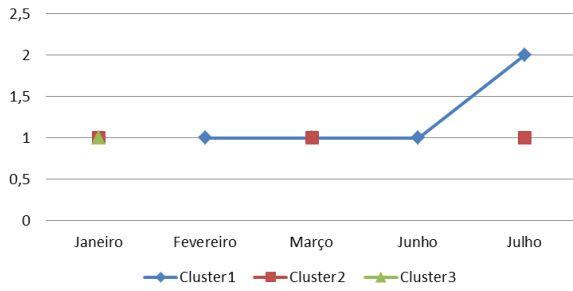
Doença Mental



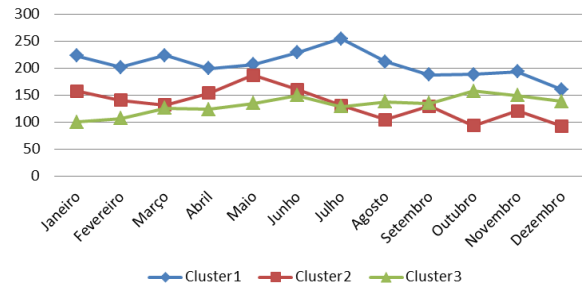
Doenças Sexualmente Transmissíveis



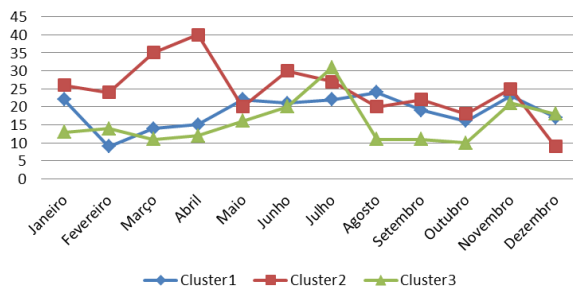
Exposição a Químicos



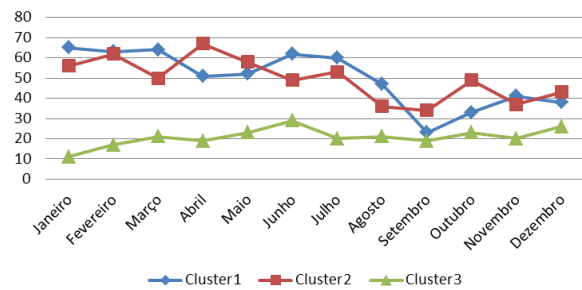
Dor Abdominal

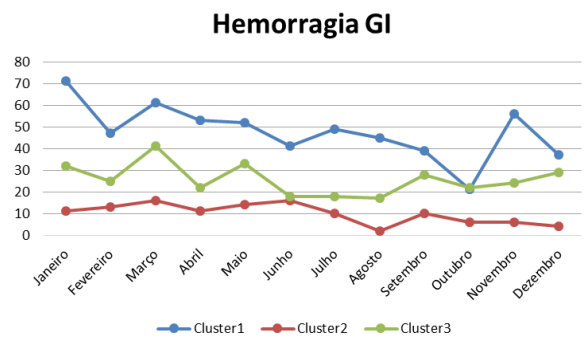
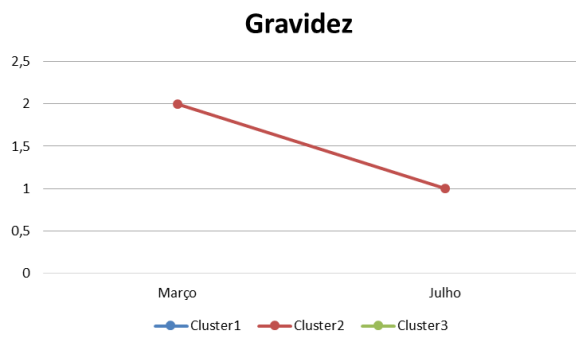
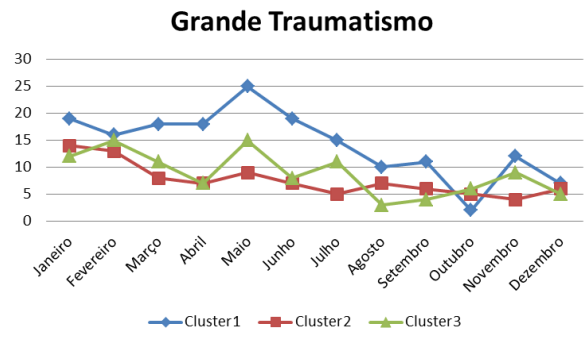
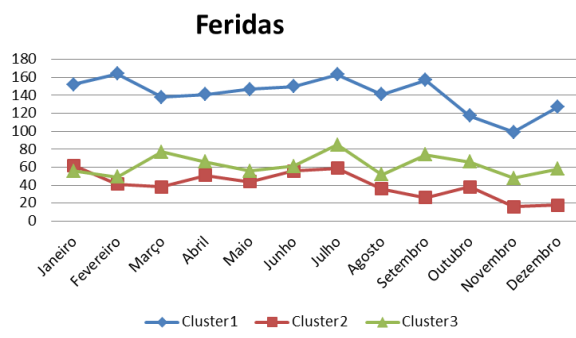
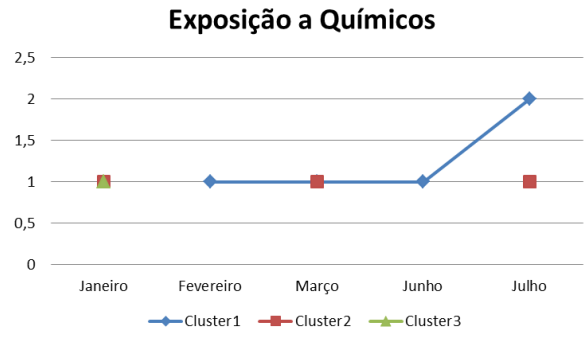
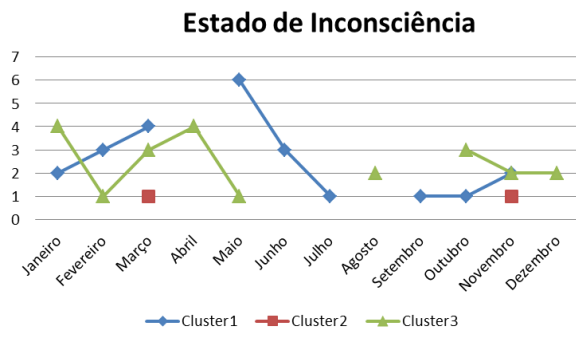
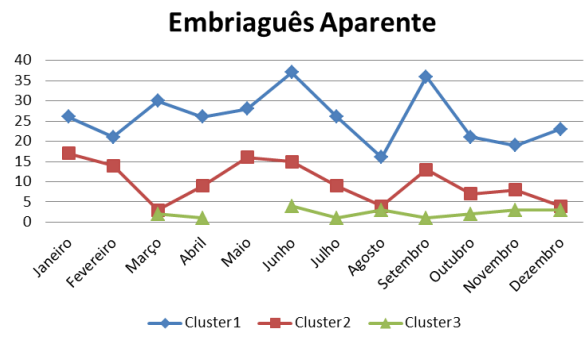
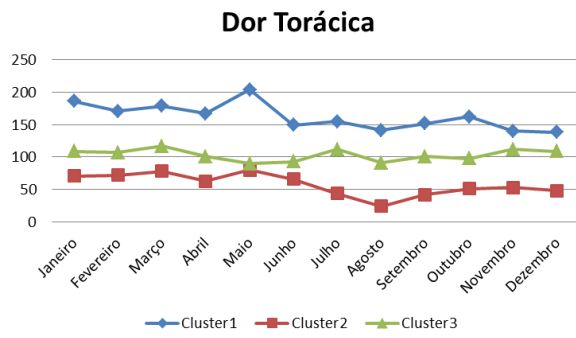
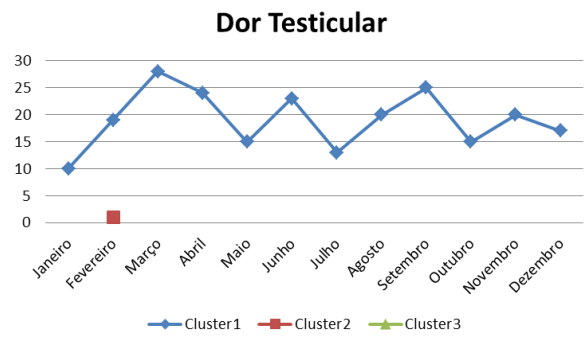
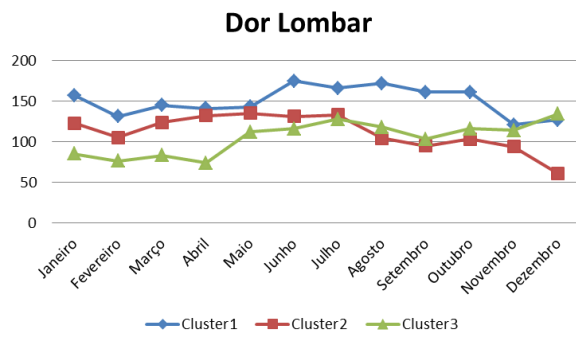


Dor Cervical

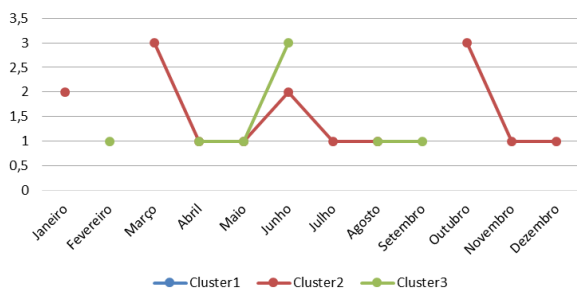


Dor de Graganta

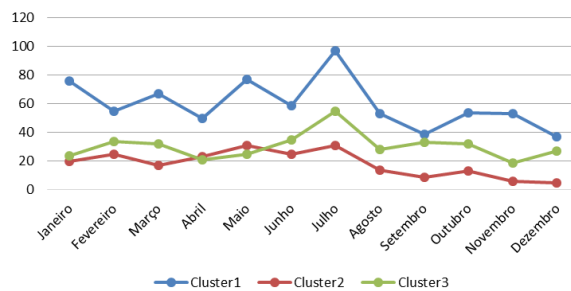




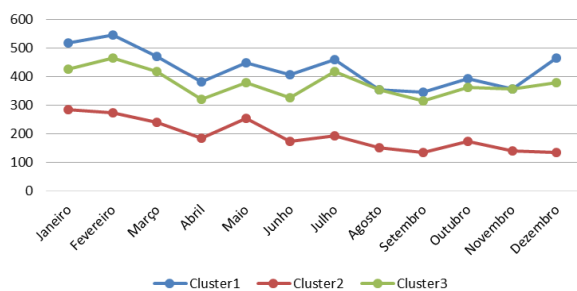
Hemorragia Vaginal



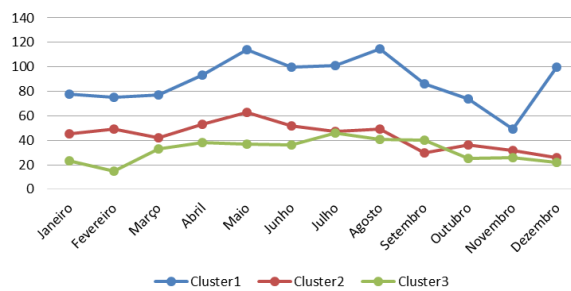
Indicação do Médico



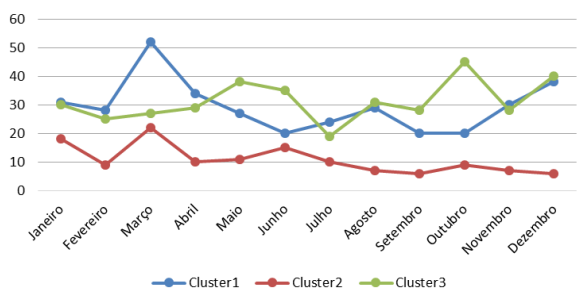
Indisposição no Adulto



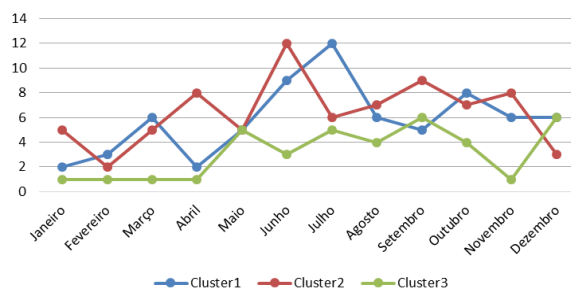
Infecções Locais e Abscessos



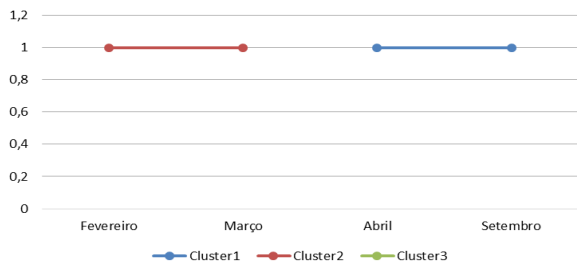
Lesão Toraco-Abdominal



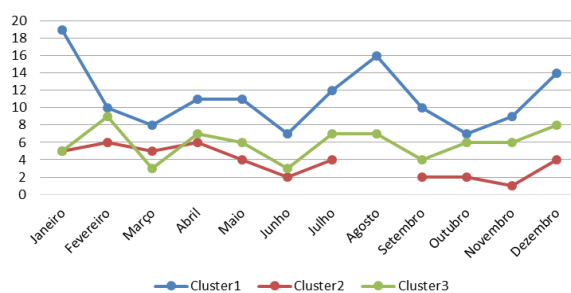
Mordeduras e Picadas



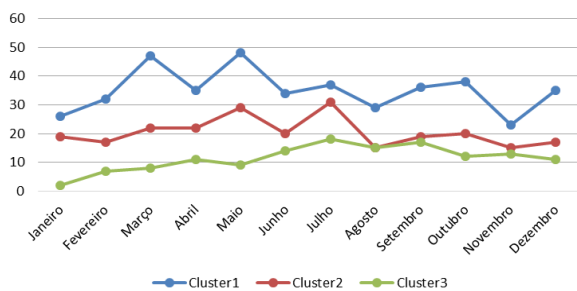
País Preocupados



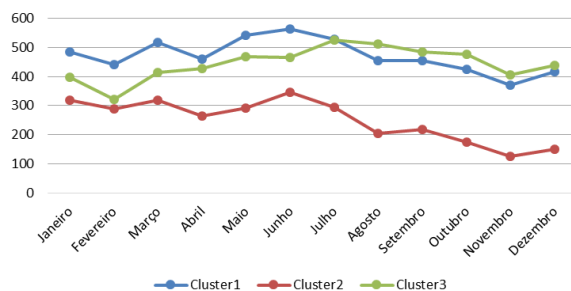
Outro



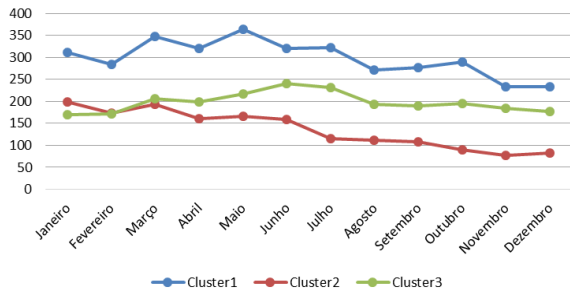
Problemas Estomatológico



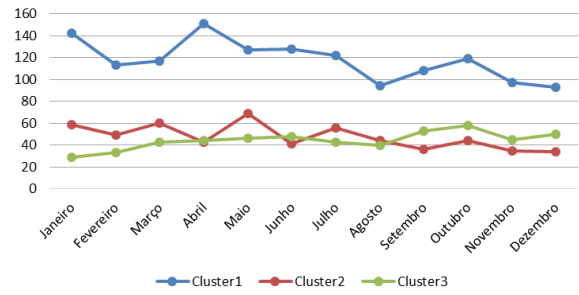
Problemas nos Membros



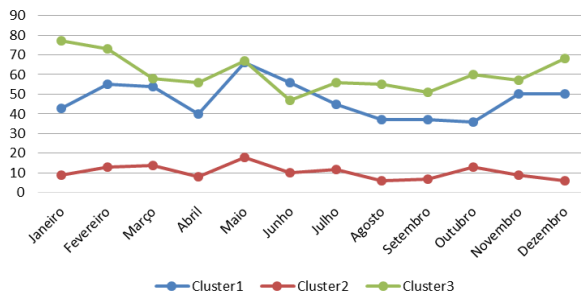
Problemas Oftalmológicos



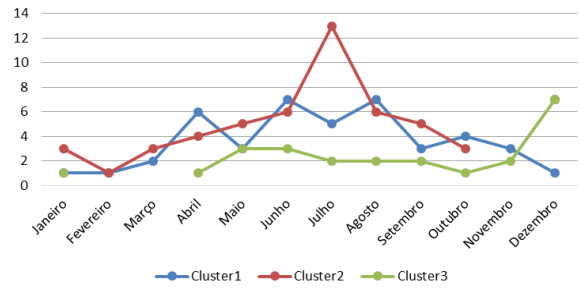
Problemas Urinários



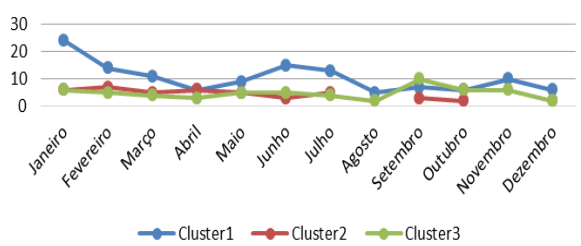
Queda



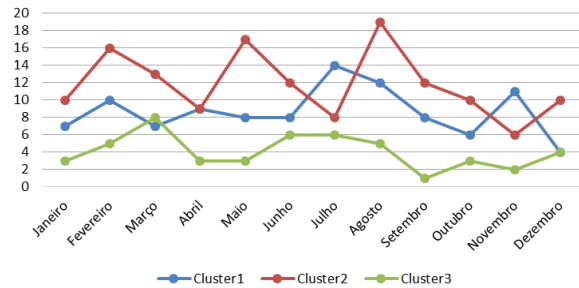
Queimaduras Profundas



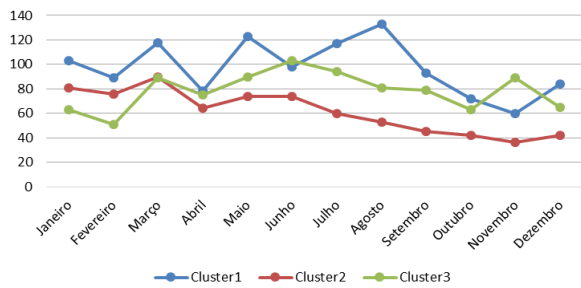
Realização de Exames



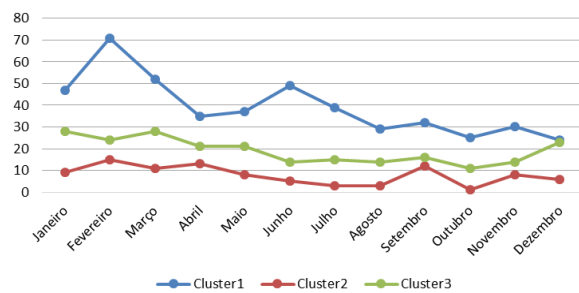
Sobredosagem ou Envenenamento



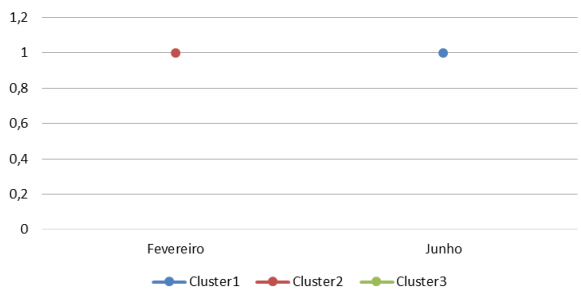
Problemas nos Ouvidos



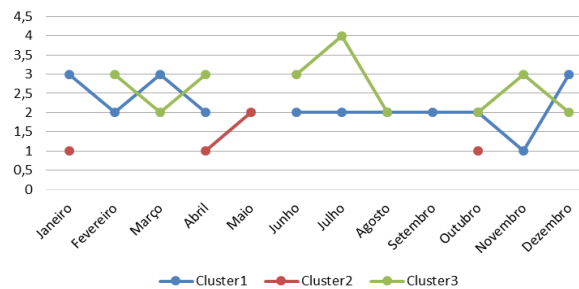
Problemas Nasais



Transplante



Transferência para Hospital



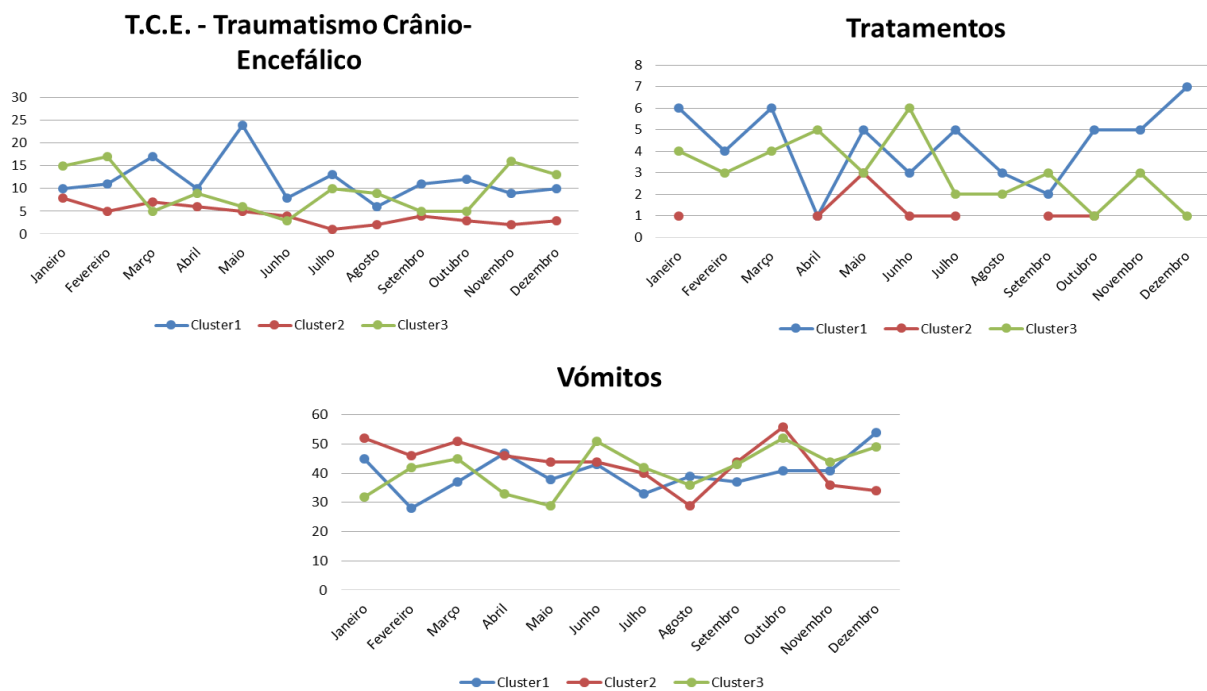


Figura 30 - Relação entre as variáveis “Data” e “Razão”

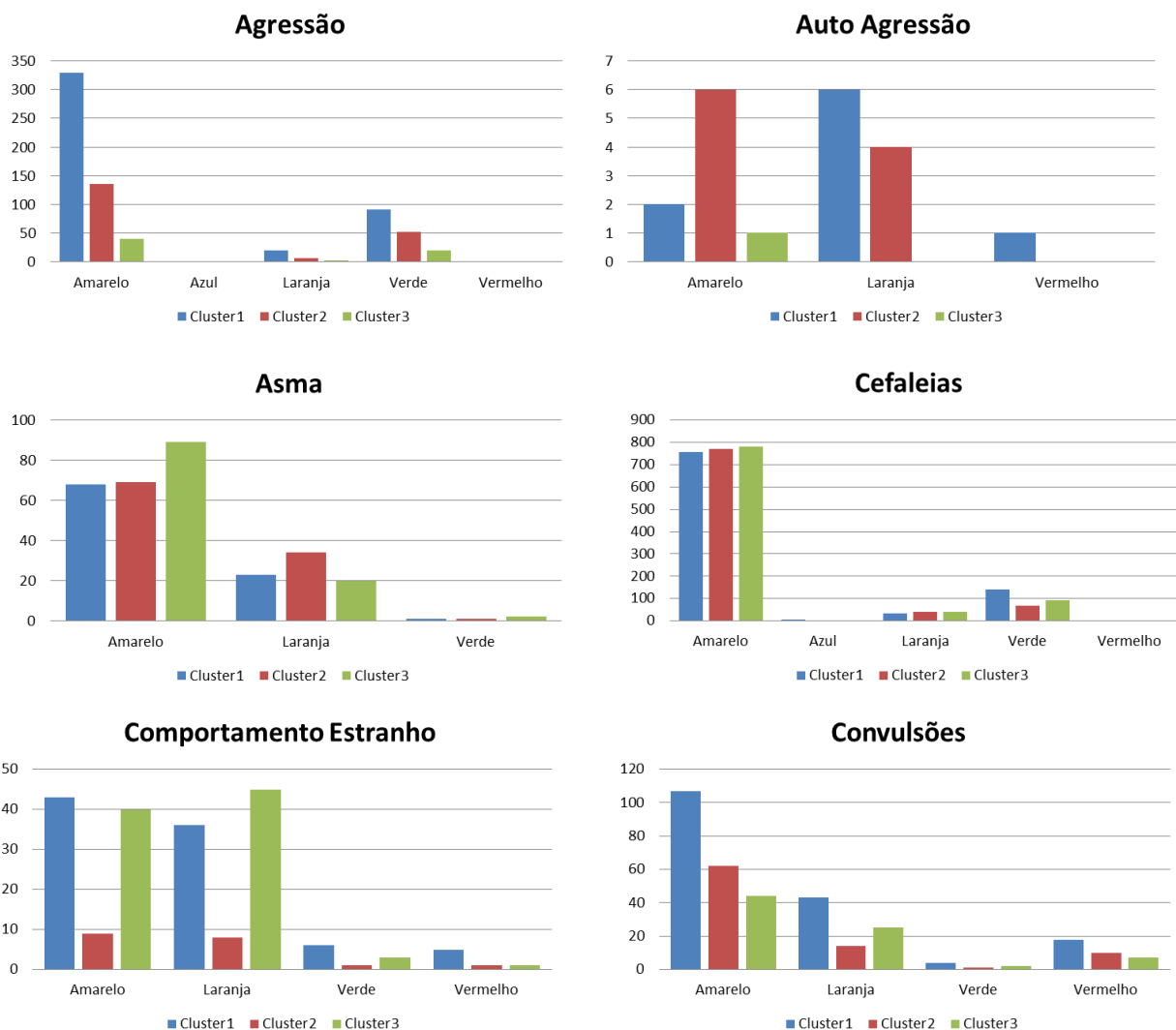
Outra análise foi perceber se haveria algum padrão entre as variáveis da “Razão que levaram os pacientes a recorrerem às urgências” e a “Cor das pulseiras” que lhes foram atribuídas. Numa análise pormenorizada às figuras 31 que se seguem, com ajuda das tabelas 12 em Anexo, observa-se que a maior parte das doenças apresenta pulseira de cor Amarela, como se constatou anteriormente.

Analisando os casos apresentados para cada “Cor” da pulseira, verificou-se que existe um padrão muito comum ao longo das doenças apresentadas, tal como foi referido anteriormente: a maior parte delas apresenta um grande número de pulseiras de cor Amarela, seguindo-se das pulseiras de cor Verdes e Laranjas. Nalguns casos essa diferença no número de pulseiras é bastante significativa, como por exemplo no caso da “Cefaleia”. Dos pacientes com esta doença, 2310 apresentaram cor Amarela, 302 cor Verde e 116 cor Laranja. Os pacientes com “Diarreia” também foram outro caso em destaque, com 609 pacientes apresentando cor amarela, 40 cor verde e 9 cor laranja; a “Dor Abdominal” mostrou 4663 pacientes com cor amarela, 609 com cor verde e 404 com cor laranja. Salientam-se ainda outras doenças com percentagens acima dos 70% com cor Amarela, como a “Dor Lombar”, as “Feridas”, “Vômitos”, “Indisposição no Adulto” e “Problemas nos Membros”. Entre *clusters*, o número de pacientes são muito semelhantes entre si, não havendo um grupo específico que fosse mais afetados por algum tipo de doença. Veja-se por exemplo os pacientes com pulseiras de cor Amarela com “Cefaleia”, verificando-se 757 casos para o *cluster 1*, 772 casos para o *cluster 2* e 781 casos para o *cluster 3*.

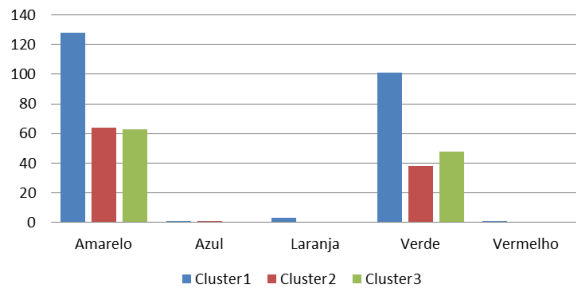
Da análise sobressaem algumas exceções, onde o número de pacientes com cor Amarela e Laranja por um lado, e cor Amarela e Verde por outro, são muito idênticos. No caso do “Comportamento Estranho” são apresentados 92 casos com cor Amarela e 89 casos com cor Laranja; os “Diabetes” têm 176 casos com a cor Amarela e 122 com cor Laranja e por fim, a “Dispneia” com 2963 casos com a

cor Amarela e 2325 com a cor Laranja. Com as pulseiras Verdes e Amarelas e com o número de pacientes idênticos, são os pacientes com “Erupções Cutâneas”, onde 887 casos têm cor Amarela e 603 cor Verde, e “Problema Oftalmológicos”, com 3958 casos com cor Amarela e 3448 casos com cor Verde. Relativamente ao número de pacientes entre os *clusters*, a situação neste caso é oposta à situação descrita no parágrafo anterior. Nos casos do “Comportamento Estranho”, “Diabetes”, “Dispneia” e “Problemas Oftalmológicos”, verifica-se que existe um maior número de pacientes no *cluster 1* e *cluster 3*, indicando, mais uma vez, que os mais idosos apresentam uma maior fragilidade e uma maior probabilidade de apanhar certas doenças. No caso dos pacientes com “Erupções Cutâneas”, os números não variam muito entre si, podendo qualquer faixa etária sofrer desta doença com igual probabilidade.

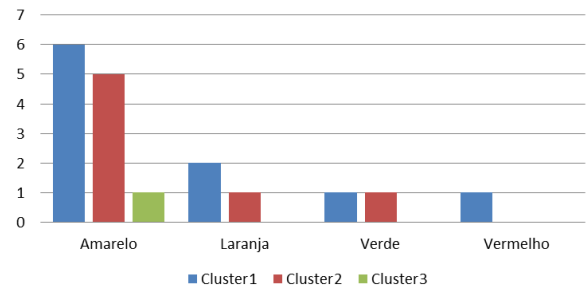
Casos ainda mais excecionais são aqueles onde outras cores se sobrepuseram à pulseira de cor Amarela. Verificou-se que 27 pacientes em “Estado de Inconsciência” obtiveram cor Vermelha (57% dos casos) e 97 pacientes que sofreram de “Mordeduras e Picadas” obtiveram cor Verde, representando 52% dos casos. Comparando o número de pacientes nos 3 *clusters*, os *clusters 1* e 3 obtiveram um maior número de pacientes com “Estado de Inconsciência”, indiciando mais uma vez que existirá uma maior fragilidade da população idosa. Constata-se a existência de um número bastante equilibrado nos pacientes que foram afetados com “Mordeduras e Picadas”.



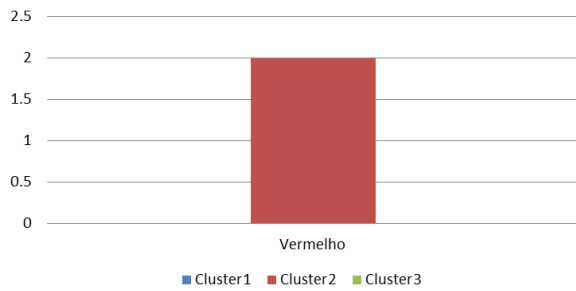
Corpo Estranho



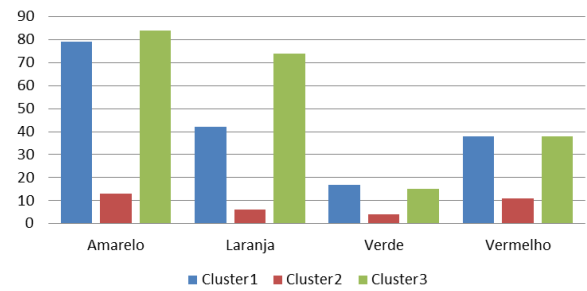
Criança Irritável



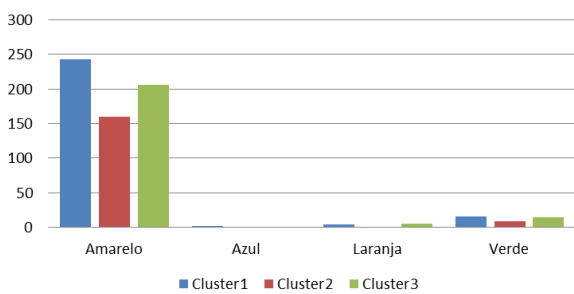
Criança que não se Sente Bem



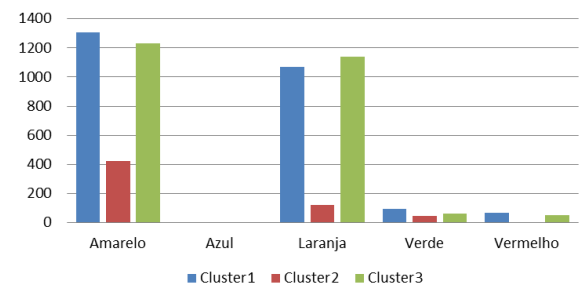
Diabetes



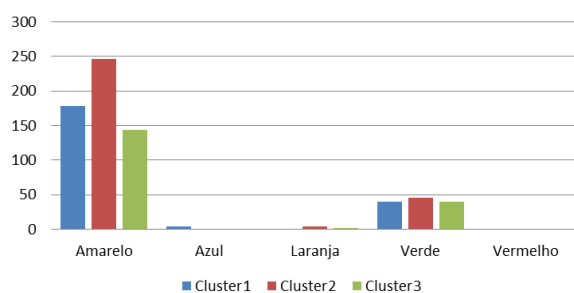
Diarreia



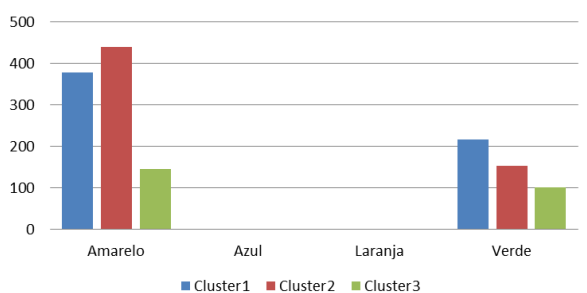
Dispneia



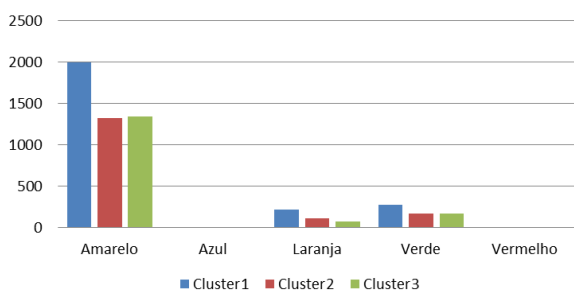
Dor Cervical



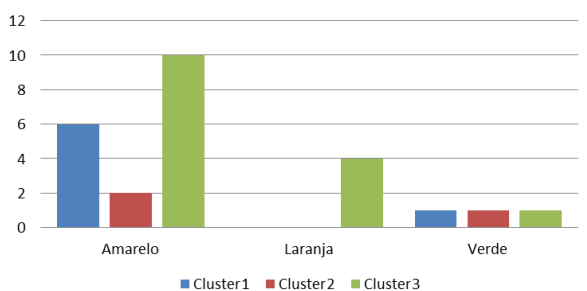
Dor de Garganta



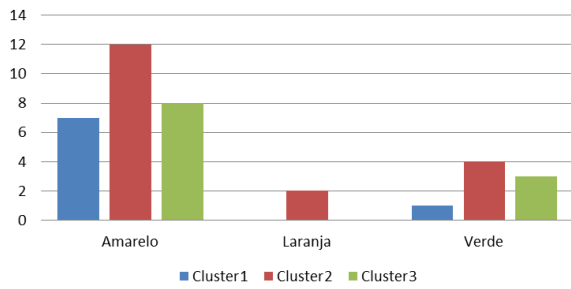
Dor Abdominal



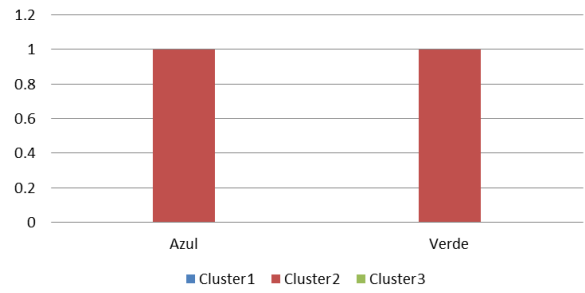
Doença Hematológica



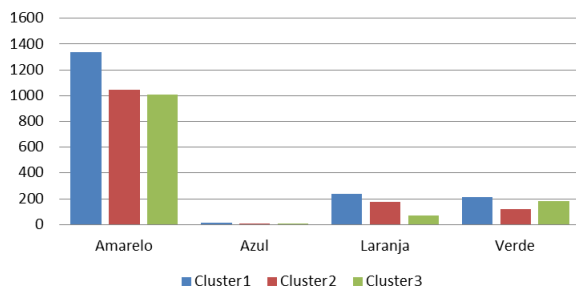
Doença Mental



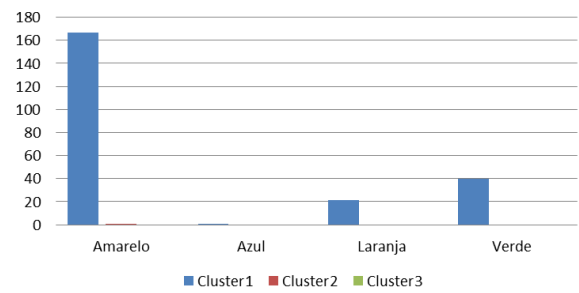
Doenças Sexualmente Transmissíveis



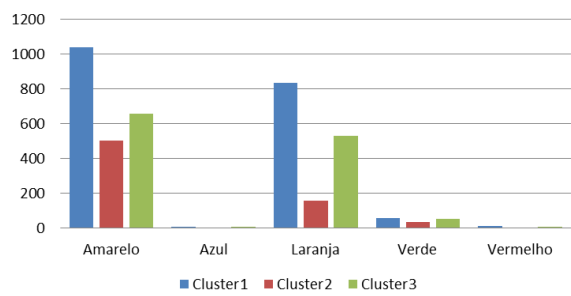
Dor Lombar



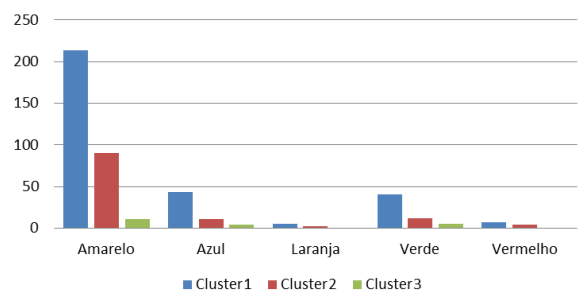
Dor Testicular



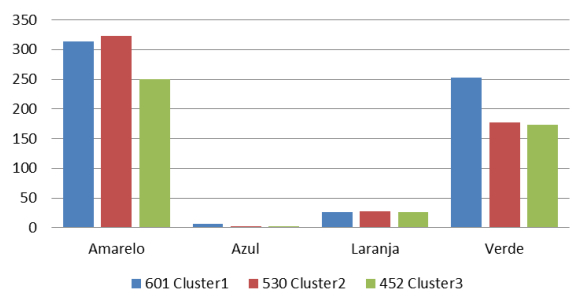
Dor Torácica



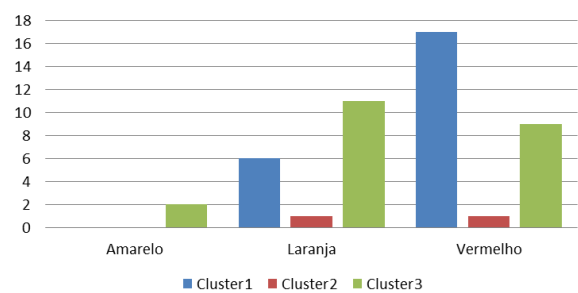
Embriaguês Aparente



Erupções Cutâneas



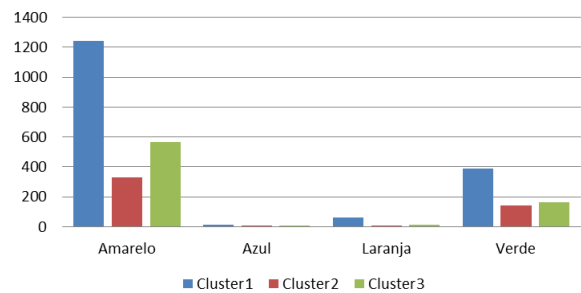
Estado de Inconsciência



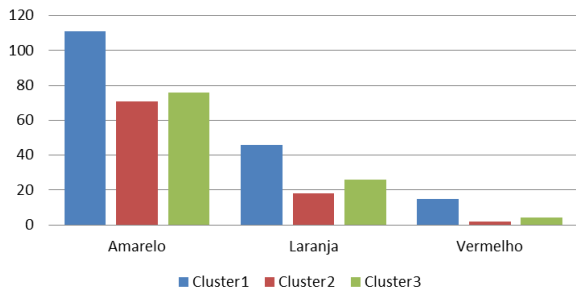
Exposição a Químicos



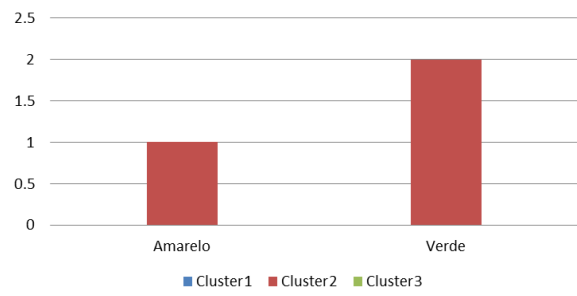
Feridas



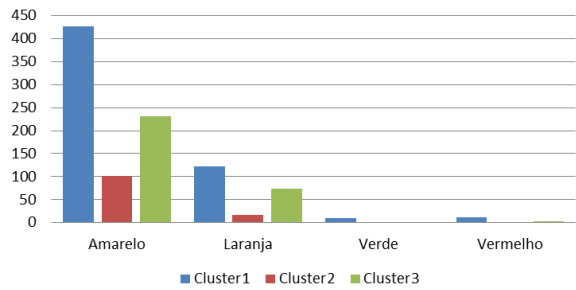
Grande Traumatismo



Gravidez



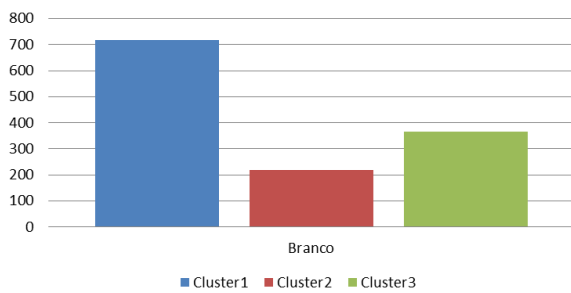
Hemorragia GI



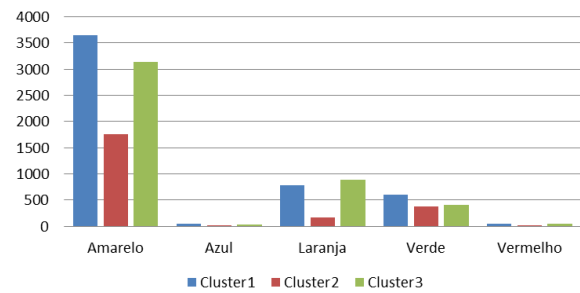
Hemorragia Vaginal



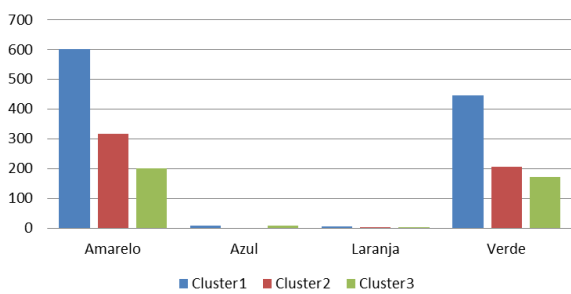
Indicação do Médico



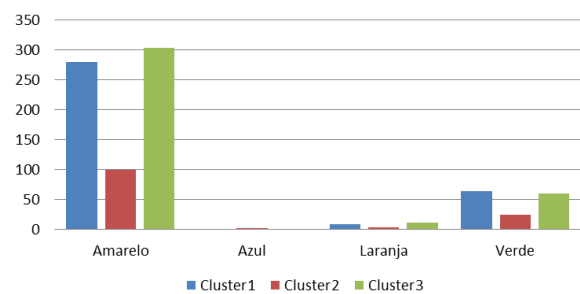
Indisposição no Adulto



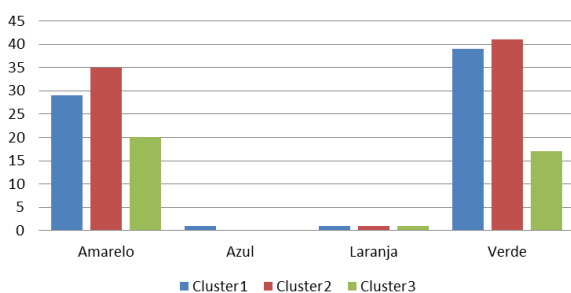
Infeções locais e abscessos



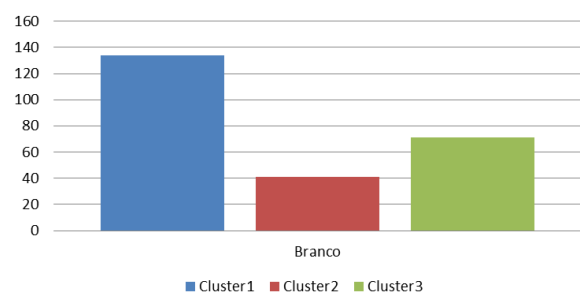
Lesão Torácico-abdominal



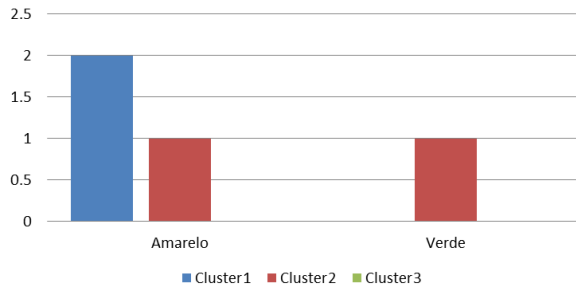
Mordeduras e Picadas



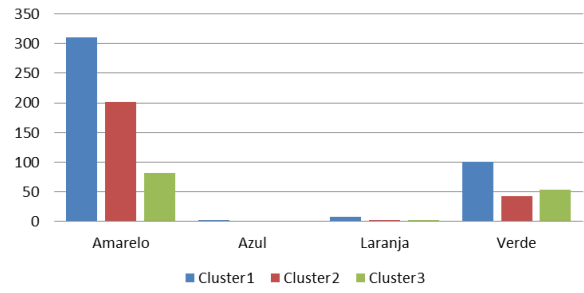
Outro



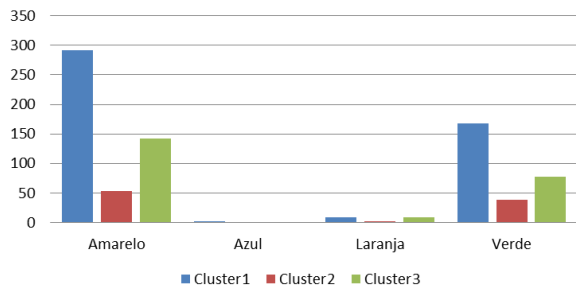
Pais Preocupados



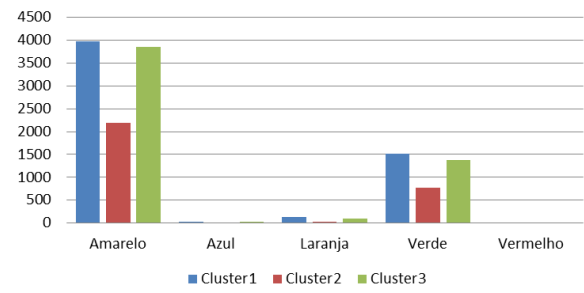
Problema Estomatológico



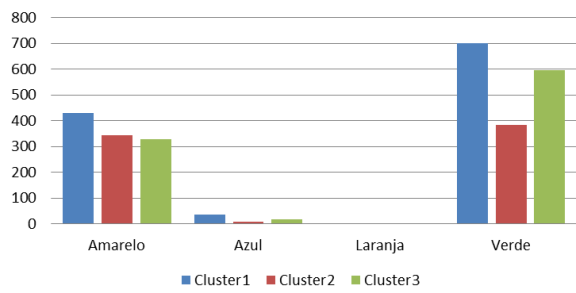
Problemas Nasais



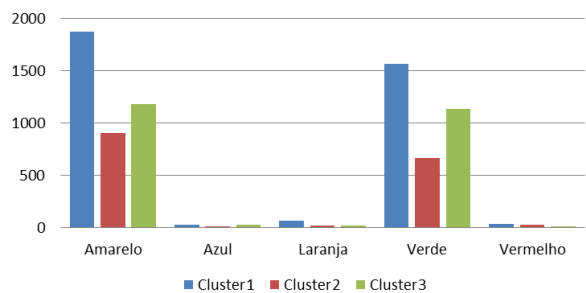
Problemas nos Membros



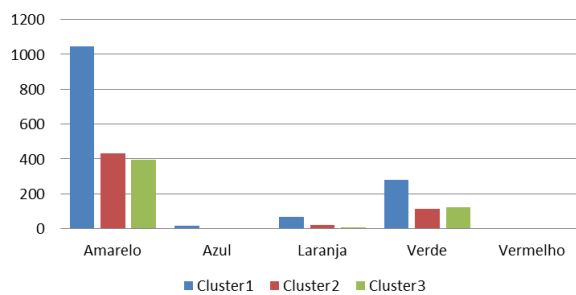
Problemas nos Ouidos



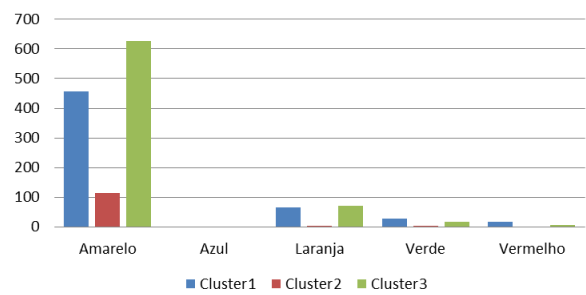
Problemas Oftalmológicos



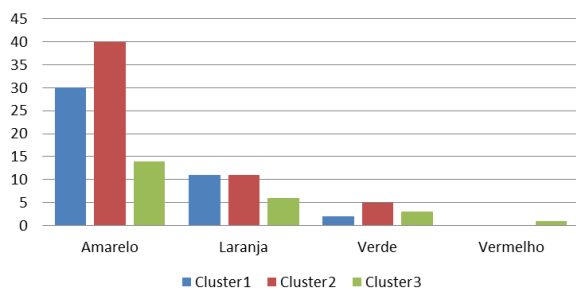
Problemas Urinários



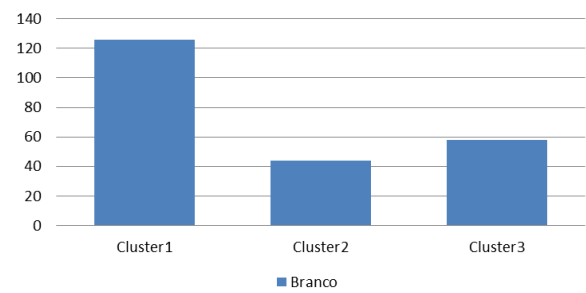
Queda



Queimaduras Profundas



Realização de Exames



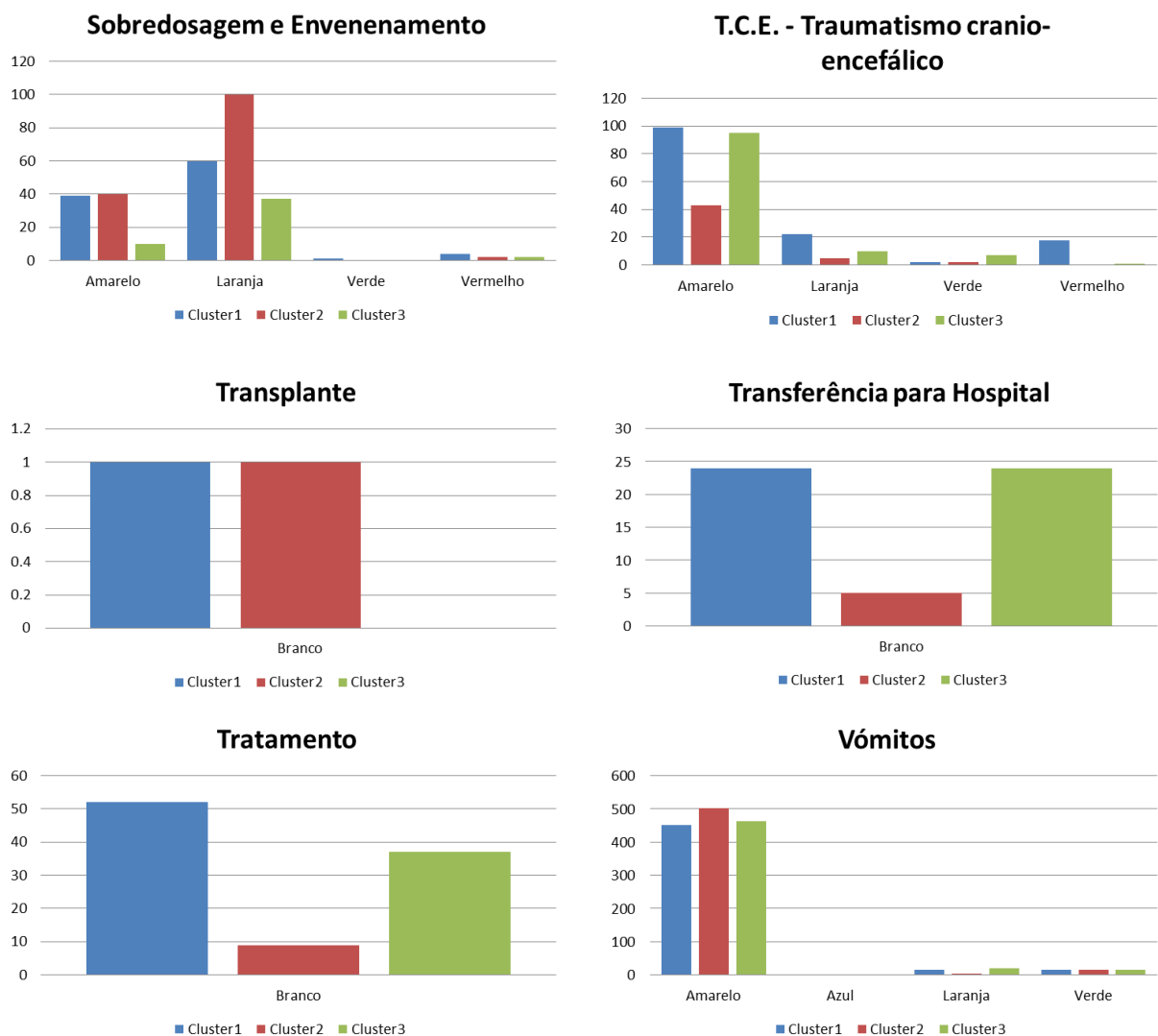


Figura 31 – Relação entre as variáveis “Cor” e “Razão”

Para finalizar, analisou-se se haveria ou não algum padrão entre as variáveis correspondentes ao motivo dos pacientes irem às urgências, a sua “Alta médica” e o seu “Destino final”.

Começando por analisar com mais pormenor a relação sobre o motivo que levou os pacientes a irem às urgências e a sua “Alta médica”, observa-se que a maioria dos pacientes teve “Alta para o Domicílio”, o que indica que após a ida às Urgências Hospitalares não precisavam de mais serviços médicos. As doenças que tiveram um maior destaque com “Alta para o Domicilio” foram a “Cefaleia” com 1591 dos 2740 casos, “Dispneia” com 2140 dos 5623 casos e “Dor Abdominal” com 3079 dos 5682 casos. Para além destes, também sobressaiu a “Dor de Garganta”, “Dor Lombar”, “Dor Torácica”, “Problemas nos Membros”, “Indisposição no Adulto”, “Problemas nos Ouvidos” e “Problemas Oftalmológicos”. Pode assim concluir-se que a maioria destas doenças não revelaram problemas graves e que os pacientes tiveram autorização para regressar às suas residências. Comparando o número de pacientes entre os *clusters* para a “Alta para o Domicílio”, observa-se que o *cluster 1* apresenta um maior número de pacientes, indicando que o sexo masculino teve uma

maior percentagem de pacientes a regressarem às suas habitações. No entanto, analisando o *cluster 2* verifica-se através das tabelas apresentadas, o destaque dos pacientes com “Asma” e “Dor Cervical” e no *cluster 3* destacam-se os pacientes com “Lesão Torácico-Abdominal”.

Para além da “Alta Médica” “Alta para o Domicílio” a destacar-se no número de pacientes, também existem casos onde outras “Altas Médicas” também se destacam. Por exemplo, dos pacientes que revelaram um “Comportamento Estranho”, a maioria teve “Alta para o Domicílio” (61 dos 198 casos); porém, houve também um número significativo de pacientes que tiveram que ficar nos “Serviços de Internamento” (45 casos) ou serem transferidos para outro “Centro de Saúde” (30 casos). No caso dos pacientes com “Convulsões”, a sua maioria necessitou de “Consultas Externas” (122 dos 337 casos), havendo ainda casos em que tiveram que ser transferidos para outros “Centros de Saúde”. Os pacientes com “Diabetes” tiveram um número muito idêntico para “Alta para o Domicílio”, “Serviço de Internamento” ou “Consulta num Centro de Saúde”, sendo 140, 105 e 111 respetivamente. Por último, a “Sobredosagem ou Envenenamento” teve um maior número de pacientes a ter “Alta para outra Instituição”, com 141 num total de 295 casos. Em todos estes casos, verificou-se que existe um maior número de pacientes nos *clusters 1* e *3* para os pacientes com “Comportamento Estranho” e com “Diabetes”; já para os casos com “Convulsões e Sobredosagem” e “Envenenamento” verificou-se que foram as classes mais jovens de ambos os sexos que foram mais afetadas (*cluster 2*). Também sobressaíram situações mais graves, que obrigaram os pacientes a ficarem internados. Exemplo disso é o caso de “Estado de Inconsciência”, em que 23 casos tiveram que ficar nos “Serviços de Internamento” e 7 acabaram mesmo por falecer. Este caso verificou-se apenas no intervalo superior das idades, em ambos os sexos, tanto no *cluster 1* como no *3*.

Fazendo uma análise ainda mais específica, incluindo já “Destino final”, observa-se que para a maioria dos casos onde foi definido que os pacientes iam ter “Alta para o Domicílio”, aparece como “Exterior não Referenciado”. Isto indica que após terem tido “Alta médica”, não foi especificado nenhum local que implicasse uma deslocação obrigatória. Um exemplo disso é a “Cefaleia”, onde se apresentava 1591 casos de “Alta para o Domicílio” e 1575 são considerados “Exterior não Referenciado”, e a “Dor Abdominal”, onde 3048 dos 3079 casos são considerados “Exterior não Referenciado”.

Observa-se também que dentro do “Serviço de Internamento”, a maior parte das decisões finais correspondem à “Área de Decisão Clínica”. Um exemplo disso é o caso do “Comportamento Estranho”, onde dos 45 casos internamento, 19 vão diretamente para a “Área de Decisão Clínica”. No entanto, há casos que se podem observar no “Destino final” dos pacientes com “Dor Lombar”, no “Serviço de Internamento”. A maioria dos pacientes ficou internada em Urologia, sendo que mais de 50% correspondiam ao *cluster 1*. Isto é um indício de infeção urinária, que afetou em maior número o sexo masculino.

Outro ponto a destacar são as “Consultas Externas”. Como seria de esperar constatou-se que nas doenças mais específicas, houve um maior número de pacientes que foram atendidos por consultas externas correspondentes à área dessas doenças apresentadas. Por exemplo, os pacientes com “Problemas Estomatológicos”, foram na sua maioria a consultas de Estomatologia, e os pacientes com “Problemas Oftalmológicos”, foram na sua maioria a uma consulta de Oftalmologia. Neste ponto não surgiram surpresas, sendo as consultas esperadas para cada tipo de doença.

Neste sentido, pode-se concluir que existe uma tendência comum que se repete pelas várias doenças. Por exemplo no caso dos doentes com “Diabetes”, observa-se que a maioria dos casos “Alta para o Domicílio” vai para o “Exterior não Referenciado”, o “Serviço de Internamento” vai para a “Área de Decisão Clínica” e o “Centro de Saúde” não tem nenhuma opção alternativa senão “Centro de Saúde”. O mesmo acontece para a maior parte das outras doenças.

Observando agora numa perspetiva diferente, comparando o número de pacientes entre os três *clusters*, observa-se um padrão comum que se repete em todas as análises. O número de pacientes não varia muito entre os *clusters*, havendo sempre uma uniformidade. Verifica-se também um maior número de pacientes do sexo feminino, mas estes estão divididos por dois *clusters* diferentes, dependendo da sua idade. Pode-se concluir que o *cluster* 1 se destaca dos restantes, porque o somatório de todas as idades do sexo masculino se sobrepõe às pacientes do sexo feminino a partir dos 60 anos. No entanto há casos que fogem a esta regra.

Observa-se que os pacientes com “Asma” são maioritariamente femininos, pois o *cluster* 1 é o que apresenta o menor número de pacientes. Após as urgências, a sua maioria “Regressou às Suas Residências” (144 com 307 casos) ou foram transferidos para um “Centro de Saúde” (84 casos).

Outro caso são os pacientes que sofreram de “Convulsões”. Observa-se que a maioria dos pacientes teve necessidade de uma “Consulta Externa”, ao contrário do que acontece com a maioria das doenças, que conseguem ter logo “Alta Para o Domicílio”. Dos 122 pacientes que foram a esta consulta, 112 foram observados por “Neurologistas”, sendo a consulta indicada para a doença apresentada. Os pacientes mais afetados foram os pacientes do sexo masculino, pois existe um maior número nos *clusters* 1 e 2, quando comparados com o *cluster* 3.

No caso dos doentes com “Diabetes”, observa-se que os pacientes mais afetados são os do sexo feminino a partir dos 50 anos. Observa-se uma clara diferença entre o *cluster* 3 e os restantes, sendo que a maioria foi para os “Serviços de Internamento” e para outros “Centros de Saúde”. Esta é uma situação a ter em atenção pelo elevado número de casos que apresenta, e por implicar a necessidade de cuidados redobrados.

A “Dispneia” é outro caso a ter em atenção, pois o número de pacientes no *cluster* 1 e 3 são muito idênticos e elevados, indicando que a população mais idosa é a mais afetada. No entanto, a maioria consegue recuperar e regressar às suas habitações; existem porém situações em que existe a necessidade de ficarem nos “Serviços de Internamento” (2140 e 1788 contra 5623 casos na totalidade).

Também se constata que os pacientes mais jovens são os mais afetados pela “Dor Cervical”. Apesar de a diferença não ser muito significativa, existe um número maior de pacientes no *cluster* 2 relativamente ao *cluster* 1 e 3. No entanto, a maioria regressou para as suas residências, com uma elevada percentagem de pacientes com alta para o domicílio.

Outros casos de realçar são os casos sobre “Agressão e Embriaguez Aparente”. Para ambos os casos verificou-se que o *cluster* 1 tinha um elevado número de pacientes, quando comparado com os restantes. Pode-se concluir que casos como estes evidenciam que os problemas relacionados com álcool e violência estão associados ao sexo masculino.

Também se verificou que os pacientes femininos com mais idade foram os mais afetados pela “Queda”. No entanto, a maioria teve “Alta para o Domicílio”, conseguindo recuperar facilmente.

Para concluir, os casos “Sobredosagem ou Envenenamento” e “Queimaduras e Picadas” tiveram um maior impacto nos pacientes mais jovens, pois o número é significativamente maior no *cluster 2* que nos restantes.

Por fim, esta análise permitiu compreender melhor o que se entende por “Problemas nos Membros” e “Indisposição no Adulto”. Começando pelos “Problemas nos Membros”, pode-se observar que dentro da “Consulta Externa”, a maior parte das consultas corresponde a “Ortopedia” e “Cirurgia Vascular”, havendo também uma relação dessas áreas da medicina no “Serviço de Internamento”. Outras áreas que aparecem em menor número são “Neurologia”, “Hematologia Clínica” e “Dermatologia”. No caso da “Indisposição do Adulto”, não é tão fácil uma análise, pois as consultas estão melhor distribuídas, havendo um pequeno destaque na “Neurologia e Cardiologia”. No entanto, o número de pacientes está igualmente distribuído pela “Alta para o Domicílio”, “Centro de Saúde”, “Consultas Externas” e “Serviços de Internamento”.

Estando concluída a análise dos *clusters*, a fase seguinte consiste na discussão de resultados das regras de associação. Como se observou anteriormente na Metodologia, alterou-se alguns parâmetros que estavam definidos por defeito no WEKA, de forma a alcançar o objetivo pretendido. Foram executadas 20 regras com a métrica “Confiança”, sendo o seu valor mínimo 0.65. Também foram removidas todas as colunas com *missing values*, de forma a não enviesar a análise.

Na figura 44 dos Anexos pode-se observar o conjunto todos os *itemsets* que apresentam um maior número de vezes na base de dados, e partir destas foram criadas as regras de associação. Segue-se a figura 32 com as 20 regras de associação criadas, com a respetiva “Confiança”.

Best rules found:

1. Cor=Verde Distrito=Porto 15324 ==> Alta=Alta para o Domicílio 11860 conf:(0.77)
2. Cor=Verde 22728 ==> Alta=Alta para o Domicílio 17540 conf:(0.77)
3. Último Fluxograma / Razão=Problemas nos membros 19185 ==> Cor=Amarelo 13867 conf:(0.72)
4. Último Fluxograma / Razão=Problemas nos membros 19185 ==> Alta=Alta para o Domicílio 13400 conf:(0.7)
5. Sexo=feminino Alta=Alta para o Domicílio 40416 ==> Distrito=Porto 27985 conf:(0.69)
6. Sexo=feminino Alta=Alta para o Domicílio 40416 ==> Cor=Amarelo 27836 conf:(0.69)
7. Cor=Amarelo Sexo=feminino Alta=Alta para o Domicílio 27836 ==> Distrito=Porto 19169 conf:(0.69)
8. Distrito=Porto Sexo=feminino Alta=Alta para o Domicílio 27985 ==> Cor=Amarelo 19169 conf:(0.68)
9. Sexo=feminino 63892 ==> Distrito=Porto 43556 conf:(0.68)
10. Distrito=Desconhecido Alta=Alta para o Domicílio 20404 ==> Cor=Amarelo 13853 conf:(0.68)
11. Último Fluxograma / Razão=Problemas nos membros 19185 ==> Distrito=Porto 13014 conf:(0.68)
12. Cor=Amarelo Sexo=feminino 42887 ==> Distrito=Porto 29057 conf:(0.68)
13. Cor=Verde Alta=Alta para o Domicílio 17540 ==> Distrito=Porto 11860 conf:(0.68)
14. Cor=Verde 22728 ==> Distrito=Porto 15324 conf:(0.67)
15. Alta=Alta para o Domicílio 72479 ==> Cor=Amarelo 48749 conf:(0.67)
16. Sexo=feminino 63892 ==> Cor=Amarelo 42887 conf:(0.67)
17. Distrito=Porto Alta=Alta para o Domicílio 48577 ==> Cor=Amarelo 32582 conf:(0.67)
18. Alta=Alta para o Domicílio 72479 ==> Distrito=Porto 48577 conf:(0.67)
19. Cor=Amarelo Alta=Alta para o Domicílio 48749 ==> Distrito=Porto 32582 conf:(0.67)
20. Distrito=Porto Sexo=feminino 43556 ==> Cor=Amarelo 29057 conf:(0.67)

Figura 32 - Regras criadas através do software WEKA

Como se pode observar na figura anterior, as regras com maior Confiança apresentam um valor de 0.77. Isto significa que para as duas primeiras regras, 77% dos casos com pulseira de cor Verde podem acabar por ter “Alta para o Domicilio”. Na primeira ainda indica uma estimativa de 15324 pacientes que possam residir no distrito do Porto. Esta regra faz sentido, pois a cor Verde indica que os casos não são muito urgentes de tratar, podendo desta forma ter alta médica e não necessitar de mais consultas.

Também se observa com uma confiança elevada os pacientes com “Problemas nos Membros”, indicando que 72% podem acabar por ter pulseira de cor Amarela, apresentados casos urgentes a tratar e 70% podem ter “Alta para o Domicilio”. Sendo os casos que originaram “Problemas nos Membros” subjetivos e de algum modo vagos, percebe-se desta regra que os casos que apareceram nas urgências são casos graves e que tiveram urgência de serem tratados.

As restantes regras são abaixo dos 0,7 de Confiança, sendo algumas de pouca relevância para o problema. Um exemplo é a regra nº9, onde indica que 68% do sexo feminino podiam pertencer ao Distrito do Porto, não trazendo valor para o problema em questão.

Para perceber se estes campos apareciam com a mesma regularidade nas seguintes regras, aumentou-se o número de regras a serem criadas para 40. O resultado apresenta-se na figura 33, comprovando que as regras acabam por ser muito similares entre si, apresentando sempre as mesmas variáveis. Para além disso, não são criadas mais que 26 regras, pois o limite de Confiança estabelecido foi 0.65, havendo apenas 26 que alcançam essa confiança.

Best rules found:

1. Cor=Verde Distrito=Porto 15324 ==> Alta=Alta para o Domicilio 11860 conf:(0.77)
2. Cor=Verde 22728 ==> Alta=Alta para o Domicilio 17540 conf:(0.77)
3. Último Fluxograma / Razão=Problemas nos membros 19185 ==> Cor=Amarelo 13867 conf:(0.72)
4. Último Fluxograma / Razão=Problemas nos membros 19185 ==> Alta=Alta para o Domicilio 13400 conf:(0.7)
5. Sexo=feminino Alta=Alta para o Domicilio 40416 ==> Distrito=Porto 27985 conf:(0.69)
6. Sexo=feminino Alta=Alta para o Domicilio 40416 ==> Cor=Amarelo 27836 conf:(0.69)
7. Cor=Amarelo Sexo=feminino Alta=Alta para o Domicilio 27836 ==> Distrito=Porto 19169 conf:(0.69)
8. Distrito=Porto Sexo=feminino Alta=Alta para o Domicilio 27985 ==> Cor=Amarelo 19169 conf:(0.68)
9. Sexo=feminino 63892 ==> Distrito=Porto 43556 conf:(0.68)
10. Distrito=Desconhecido Alta=Alta para o Domicilio 20404 ==> Cor=Amarelo 13853 conf:(0.68)
11. Último Fluxograma / Razão=Problemas nos membros 19185 ==> Distrito=Porto 13014 conf:(0.68)
12. Cor=Amarelo Sexo=feminino 42887 ==> Distrito=Porto 29057 conf:(0.68)
13. Cor=Verde Alta=Alta para o Domicilio 17540 ==> Distrito=Porto 11860 conf:(0.68)
14. Cor=Verde 22728 ==> Distrito=Porto 15324 conf:(0.67)
15. Alta=Alta para o Domicilio 72479 ==> Cor=Amarelo 48749 conf:(0.67)
16. Sexo=feminino 63892 ==> Cor=Amarelo 42887 conf:(0.67)
17. Distrito=Porto Alta=Alta para o Domicilio 48577 ==> Cor=Amarelo 32582 conf:(0.67)
18. Alta=Alta para o Domicilio 72479 ==> Distrito=Porto 48577 conf:(0.67)
19. Cor=Amarelo Alta=Alta para o Domicilio 48749 ==> Distrito=Porto 32582 conf:(0.67)
20. Distrito=Porto Sexo=feminino 43556 ==> Cor=Amarelo 29057 conf:(0.67)
21. Distrito=Desconhecido 33532 ==> Cor=Amarelo 22195 conf:(0.66)
22. Cor=Amarelo Distrito=Porto Sexo=feminino 29057 ==> Alta=Alta para o Domicilio 19169 conf:(0.66)
23. Cor=Amarelo 77489 ==> Distrito=Porto 50946 conf:(0.66)
24. Distrito=Porto 78073 ==> Cor=Amarelo 50946 conf:(0.65)
25. Sexo=masculino Alta=Alta para o Domicilio 32021 ==> Cor=Amarelo 20879 conf:(0.65)
26. Distrito=Porto Sexo=masculino Alta=Alta para o Domicilio 20580 ==> Cor=Amarelo 13403 conf:(0.65)

Figura 33 – Aumento do número de regras de associação

Não foi estabelecido um valor de Confiança menor, pois o objetivo é perceber o maior número de regras criadas correctamente e com um maior nível de confiança. Caso esse valor fosse menor, não trazia vantagens para o problema.

Contata-se nesta fase do projeto como os dados aqui apresentados coincidem com os resultados expostos anteriormente na análise dos *clusters*. Pode-se observar que na maioria das regras está presente o Distrito do Porto, pois é o distrito com mais número de pacientes. Para além deste distrito estar exposto na maior parte das regras, também se observa em quase todas as regras a pulseira de cor Amarela, pacientes do sexo feminino, a “Alta médica” como “Alta para o Domicilio” e os “Problemas nos membros” como a maior razão de pacientes que leva os pacientes a recorrerem às urgências. Estes campos correspondem a todos os valores mais elevados de diferentes variáveis, fazendo sentido a sua correlação entre eles.

Como conclusão, as regras de associação acabam por salientar os resultados alcançados anteriormente, mas de uma forma diferente. Com um nível de Confiança elevado, apenas são apresentados os campos com um elevado número de casos (pulseira de cor Amarela, distrito do Porto, entre outros); por haver esse enorme número, existem muitas relações entre essas variáveis. No entanto, o valor de confiança não é tão elevado como se esperaria, pois o melhor valor apresentado é uma confiança de 0.77. Numa experiência anterior alcançou-se uma confiança de 1 quando a variável “Destino” ainda estava inserida. No entanto essa variável estava relacionada com a variável “Alta”, trazendo apenas resultados óbvios e evidentes.

O ponto mais importante a retirar daqui é que os pacientes com “Problemas no Membros” tiveram um elevado nível de confiança com pulseira de cor Amarela, indicando que os casos apresentados têm uma maior probabilidade de serem graves. Esta razão apresentada é muito vaga, tal como disse anteriormente, mas reflete a possível gravidade da situação apresentada nas urgências. No entanto, apesar desta aparente gravidade, existe uma tendência para acabarem por ter “Alta médica” para as suas residências, mostrando melhorias no seu estado de saúde.

5. CONCLUSÕES

Este projeto pretendeu analisar os dados relacionados com os problemas que surgem nas Urgências Hospitalares (UH), de modo a poderem ser retiradas conclusões que permitam compreender o que leva as pessoas a recorrerem às mesmas e quais as consequências de tal afluxo. Numa primeira análise, verificou-se que a população tinha uma média de 56 anos de idade, correspondendo 55% ao sexo feminino. Estudos indicam que as mulheres têm uma longevidade maior relativamente aos homens, o que é comprovado pela média de idades para cada um dos sexos: média de 58 anos para as mulheres e média de 54 anos para os homens.

Olhando para os resultados mais específicos, observa-se na formação dos *clusters* que houve uma clara separação em três *clusters* diferentes. No primeiro observa-se apenas pacientes do sexo masculino de todas as idades, no segundo pacientes do sexo masculino mais jovens e sexo feminino até aos 60 anos e o terceiro apenas por pacientes do sexo feminino a partir dos 40 anos. Também se repara que existe muitas semelhanças entre o *cluster* 1 e 3, evidenciando que o comportamento entre o sexo feminino com mais idade e o sexo masculino no geral é muito parecido. No entanto, o *cluster* com todos os pacientes do sexo masculino tem maioritariamente pacientes mais idosos, podendo generalizar-se que os comportamentos são muito idênticos em ambas as idades, independentemente do seu sexo.

Como referi anteriormente, houve muitas semelhanças entre os *clusters* 1 e 3, assumindo que a população idosa é a que está presente em maior número nas Urgências Hospitalares, sendo a mais fragilizada e portanto a que é mais suscetível a doenças. Exemplo disso é a “Diabetes”, doença que afeta maioritariamente a população mais idosa. A Diabetes em si não é uma doença que requeira normalmente uma ida às Urgências Hospitalares; porém, os dados mostram que a maioria das pessoas que lá se dirige tem esta doença. Isto acontecerá porque existem muitas doenças associadas à Diabetes e essas sim, poderão necessitar de uma ação imediata.

Observando a análise das variáveis, uma das conclusões que se retirou foi a não existência de picos de doenças em certas alturas do ano, como por exemplo “Asma” nos meses da Primavera ou um grande surto de gripe nos meses de Inverno. Verificou-se apenas que a “Asma” foi ligeiramente superior no mês de janeiro, sendo conhecida como “Asma Alérgica”. Este tipo de “Asma” está relacionado com as doenças respiratórias típicas do Inverno, considerando este aspeto como fator de justificação para o aumento das incidências nesta altura e não na primavera. Este aumento não justifica um reforço ou uma prevenção para esta doença pois registaram-se apenas alguns casos. De referir também o caso das “Mordeduras e Picadas”, que atingiram em maior número a população mais jovem, o que se poderá explicar, eventualmente, pela maior exposição e permanência no exterior.

Outra conclusão que se retirou desta análise foi o elevado número de casos com pulseira de cor Amarela, refletindo a urgência com que os pacientes precisaram de ser atendidos. As doenças que mais se destacaram nos pacientes a quem foi atribuída esta pulseira foram a “Cefaleia”, “Diarreia” e “Dor Abdominal”. Destas três doenças, apenas a “Dor Abdominal” mostra um padrão comum ao longo dos meses: os pacientes masculinos são os que mais sofreram com esta doença, apresentando na maior parte dos meses o dobro dos pacientes que nos restantes *clusters*. Por outro lado, com casos menos graves, apresentam-se apenas os pacientes com “Mordeduras e Picadas” e com

“Problemas nos Ouvidos”, onde o número de pulseiras de cor Verde se sobrepõe às restantes. Assim, pode concluir-se que a maioria dos pacientes recorreu às urgências apenas quando foi necessário, pois os valores apresentados com pulseiras de cor Azul e Brancas são menos de 5% do total (2416 em 83428 casos). Esta conclusão contraria o estudo apresentado anteriormente da Direção Geral da Saúde (Direcção-Geral da Saúde, 2001), que indica ter havido um aumento no número de pacientes que vão às Urgências Hospitalares, sem necessidade de o fazer como primeira opção de ato médico. Refira-se que esta opinião é veiculada pela comunicação social, sendo percebida pela população como ocorrendo de facto (Porto Canal, 2015).

Observou-se também que a maioria dos pacientes que recorreram às urgências conseguiu ter “Alta para o Domicílio”. Para além de indicar um fator positivo, por não ser necessário recorrer a mais consultas no hospital nesta fase, também permite que o hospital esteja mais livre para outras situações de emergência que possam surgir. No entanto, em algumas doenças o segundo “Destino final” dos pacientes é o “Serviço de Internamento” ou transferência dos mesmos para outros “Centros de Saúde”, para consultas especializadas. Também se observou neste ponto que nos “Serviços de Internamento” existe um maior número de pacientes acima dos 50 anos, quando comparado com o *cluster* dos pacientes mais jovens do sexo feminino. Este é outro indício de como a população mais idosa está mais fragilizada, sendo por vezes necessário outros recursos médicos.

Com esta análise de *clusters* verifica-se que não existe um padrão fixo que ajude de forma direta no problema apresentado. Percebe-se que existe uma grande percentagem de casos graves a aparecer em qualquer altura do ano, sendo importante garantir reforços para responder a essas necessidades. Os comportamentos entre os *clusters* 1 e 3 são muito idênticos, sendo apenas o sexo e a idade o seu fator de separação. Importante a reter e a ter em atenção são as doenças que mais afetam estes pacientes, como a “Dispneia”, “Dor Torácica” e “Problemas Oftalmológicos”, o que obriga a cuidados reforçados.

Outra técnica utilizada neste trabalho consistiu na aplicação de regras de associação. Observou-se que a confiança não foi tão elevada como seria de esperar. Com a variável “Destino” incluída, a confiança conseguia atingir o seu valor máximo, mas também não era vantajoso mantê-la pelos resultados que eram atingidos. Como se observou, todas as regras com confiança 1 estavam sempre relacionadas com a “Alta Médica” e o seu “Destino Final”.

Desta forma, ao retirar a variável destino, a confiança máxima observada foi 0,77%, isto é, apenas 77% dos casos continham campos iguais. Os resultados obtidos vêm a reforçar o que foi dito anteriormente, pois são apresentadas regras onde se verificou maior número de pacientes.

Pode-se garantir com uma confiança de 77% dos casos que os pacientes com pulseira de cor Verde conseguiram “Alta Para o Domicílio”. Observa-se também que o número de pulseiras de cor Amarela é bastante superior ao número de pulseiras Verdes, mas a sua confiança é menor (apenas de 67%). Também faz sentido este resultado, pois sendo a cor Verde para casos pouco urgentes, mais rapidamente têm alta para as suas residências, quando comparado com os casos urgentes.

Desta análise também se pode afirmar com maior precisão o que se entende por “Problemas nos Membros”. Com estas regras observa-se que os pacientes tiveram pulseira de cor Amarela mas que

conseguiram ter “Alta Para o Domicílio”, com uma confiança acima dos 70%. Com ajuda da análise de *clusters*, percebe-se que esta doença também está principalmente associada a doenças relacionadas com ossos ou músculos (Ortopedia) ou coração (Cirurgia Vascular).

Os dados não permitiram retirar mais informação, pois a maioria das regras eram muito similares entre si. Na abordagem aos dados disponíveis, verificou-se que não valeu a pena diminuir o valor da confiança para a criação de novas regras, pois o objetivo foi criar regras com a maior confiança possível, não sendo útil as regras se iriam ser criadas.

A análise final deste trabalho mostra diversos resultados, com diferentes níveis de importância. Ao contrário do que seria espectável, não foi encontrado um padrão fixo que ajudasse no problema inicialmente apresentado. Este ano em estudo (2012) foi bastante estável, não apresentando casos que pudessem destacar-se perante os restantes, introduzindo uma estabilidade nos resultados que não faz sobressair nenhum evento, como aconteceu em 2014 com o surto de *Legionella* (Borja-Santos, 2014) ou em 2015 com a Bactéria Multirresistente (Campos, 2015), onde as Urgências Hospitalares tiveram um papel muito importante na deteção dos surtos e na reação imediata que tiveram. Desta forma, os pontos-chave e conclusões mais importantes a reter deste trabalho são a confirmação de que a maioria dos utentes que se dirigem às Urgências Hospitalares o faz de facto por situações urgentes e são na sua maioria pacientes mais idosos. Assim, é importante garantir que os hospitais e principalmente as Urgências Hospitalares estejam preparadas para responderem às solicitações diversas, esperadas ou inesperadas, devendo ter em atenção que a maioria dos utentes são idosos, necessitando por isso de acompanhamento mais diferenciado e acompanhado.

6. LIMITAÇÕES E RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Ao longo deste trabalho, fui confrontada com algumas dificuldades durante o projeto. Na conversão da base dados em Excel para o formato correto em SAS, deparei-me com alguns dados que não faziam sentido. Analisando a BD original, concluiu-se que o problema estava mesmo na conversão. Tentou fazer-se outra vez a conversão, mas os resultados foram os mesmos, ou seja, apareceram os mesmos erros. Neste sentido, os dados tiveram que ser tratados e excluídos, como foi explicado na Metodologia. Outra dificuldade encontrada foi a falta de conteúdo. Apesar de haver 118262 registos na base dados, um terço destes registos tinham campos como “Desconhecidos”, o que empobreceu a análise e as conclusões para o objetivo final.

Para futuros estudos ou trabalhos na área recomenda-se a utilização de uma base de dados mais completa, tanto em termos de conteúdo como em termos de variáveis. Esta BD tinha um total de 13 variáveis, mas 5 delas correspondiam apenas à localização da residência do paciente, o que limitou bastante a análise para chegar ao objetivo final.

7. BIBLIOGRAFIA

- Agrawal, R. & Srikant, R. (2008). *Fast Algorithms for Mining Association Rules*.
- Arnold, J., L. (1999, January). International Emergency Medicine and the Recent Development of Emergency Medicine Worldwide. *Annals of Emergency Medicine*, 33(1), 97-99.
- Azevedo, A., & Santos, M. F. (2008). KDD, SEMMA and CRISP-DM: a parallel overview. *IADIS European Conference Data Mining*, 182–185. Retrieved from <http://recipp.ipp.pt/handle/10400.22/136>
- Bellazi, R., & Zupan, B. (2008). Predictive data mining in clinical medicine: Current issues and guidelines. *International Journal of Medical Informatics*, 77(2008), 81-89.
- Borja-Santos (2014, November). *OMS considera surto de Legionella em Portugal como “grande emergência de saúde pública”*. Retrieved from <http://www.publico.pt/sociedade/noticia/oms-considera-surto-de-legionaella-em-portugal-como-grande-emergencia-de-saude-publica-1675862>
- Boruchovitch, E., & Mednick, B., R. (2002). The meaning of health and illness: some considerations for health psychology. *Psico-USF*, 7(2), 175-176.
- Breu, F., Guggenbichler, S., & Wollmann, J. (2008). *Getting Started with SAS Enterprise Miner 5.3. Vasa*. Retrieved from <http://medcontent.metapress.com/index/A65RM03P4874243N.pdf>
- Campos, A. (2015, October). *Hospital de Gaia diz que o surto de bactéria multirresistente está controlado*. Retrieved from <http://www.publico.pt/sociedade/noticia/hospital-de-gaia-diz-que-surto-de-bacteria-multiresistente-esta-controlado-1711960>
- Campos, L. (2014). Plano Nacional de Saúde: Roteiro de Intervenção em Cuidados de Emergência e Urgência Luis.
- Chen, M. S., Han, J., & Yu, P. S. (1996, December). Data Mining: An Overview from a Database Perspective. *IEEE Transactions on Knowledge and Data Engineering*, 8(6), 866-871.
- Cios, K., J. & Moore, G., W. (2002). Uniqueness of medical data mining. *Artificial Intelligence in Medicine*, 26(2002), 2-10.
- CSC. (2012,). *Big Data Universe Beginning to Explode*. Retrieved from http://www.csc.com/insights/flxwd/78931-big_data_universe_beginning_to_explode
- Diário da Republica (August, 2014). *Diário da República, 2.ª série — N.º 154 — 12 de agosto de 2014* . Retrieved from https://www.ers.pt/uploads/document/file/4916/Despacho_n._10429_2014_DR_154_S_ri_e_II_12_de_agosto_2014_Parte_C.pdf
- Direcção-Geral da Saúde (2001, November). Rede de Referenciação Hospitalar de Urgência/Emergência, 5-8.

- Fayyad, U., Shapiro, G. P., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *American Association for Artificial Intelligence*, 17(3) 37-44.
- Furnas, A. (2012, April 2). *Everything You Wanted to Know About Data Mining but Were Afraid to Ask*. Retrieved from <http://www.theatlantic.com/technology/archive/2012/04/everything-you-wanted-to-know-about-data-mining-but-were-afraid-to-ask/255388/>
- Governo de Portugal. *Serviço Nacional de Saúde*. Retrieved from <http://www.portugal.gov.pt/pt/o-governo/arquivo-historico/governos-constitucionais/gc19/os-ministerios/ms/quero-saber-mais/sobre-o-ministerio/servico-nacional-de-saude/composicao-do-servico-nacional-de-saude.aspx>
- Gupta, E., A. & Mishra, E., A. (2011, November). Research Paper on Cluster Techniques of Data Variations. *International Journal of Advance Technology & Engineering Research*, 1(1), 39-43.
- Hand, D. J. (1998, May). Data Mining: Statistics and More? *The American Statistician*, 52(2), 112-115.
- Hariz, M., Adnan, M., Husain, W., Aini, N., & Rashid, A. (2012). Data Mining for Medical Systems: A Review. *Proc. Of the International Conference on Advances in Computer and Information Technology*, 17-20.
- Jain, A., K. (2009, September 9). Data Clustering: 50 years beyond K-mean. *Pattern Recognition Letters*, 31, 651-658.
- Jain, A., K., Murty, M., N., & Flynn, P., J. (1999, September). Data Clustering: A Review. *ACM Computing Surveys*, 31(3), X-X
- Lai, K., & Cerpa, N. (2001). Support vs Confidence in Association Rule Algorithms, OPTIMA Conference, 1–14.
- Leal, A., Ferreira, J., P. & Carvalho, R. (Abril, 2015). *Repórter TVI na íntegra: caos nas urgências mesmo depois da gripe*. Retrieved from <http://www.tvi24.iol.pt/sociedade/reportagem/reporter-tvi-na-integra-caos-nas-urgencias-mesmo-depois-da-gripe>
- Paiva, J. A. O. D. C., Silva, A. M. Da, Almeida, A. L., Seco, C. M. D. S., Gomes, C. M. P. R., Ribeiro, E. D. P. R. G., ... Silva, R. P. F. Da. (2012). *Reavaliação da Rede Nacional de Emergência e Urgência*, 117. Retrieved from <http://www.portaldasaude.pt/NR/rdonlyres/0323CC90-45A4-40E4-AA7A-7ACBC8BF6C75/0/ReavaliacaoRedeNacionalEmergenciaUrgancia.pdf>
- Plano Nacional de Saúde (Fevereiro, 2010). *Estado Saúde dos Portugueses*. Retrieved from <http://pns.dgs.pt/planeamento-saude/perfil-saude/>
- Plano Nacional de Saúde (Fevereiro, 2010). *Natureza e Objetivos dos PNS*. Retrieved from <http://pns.dgs.pt/planeamento-saude/natureza-e-objectivos-dos-planos-nacionais-de-saude/>
- Pordata (2015). *BI em Portugal*. Retrieved from <http://www.pordata.pt/Portugal>

- Pordata (2015). *Esperança de vida à nascença: total e por sexo – Europa*. Retrieved from <http://www.pordata.pt/Europa/Esperan%C3%A7a+de+vida+%C3%A0+nascen%C3%A7a+total+e+por+sexo-1260>
- Pordata (2015). *Indicadores de envelhecimento segundo os Censos em Portugal*. Retrieved from <http://www.pordata.pt/Portugal/Indicadores+de+envelhecimento+segundo+os+Censos+-+525>
- Portal da Saúde (2015). *História do Serviço Nacional de Saúde*. Retrieved from <http://www.portaldasauade.pt/portal/conteudos/a+saude+em+portugal/servico+nacional+de+sauade/historia+do+sns/historiadosns.htm>
- Portal dos Diabetes (2014). *Diabetes Factos e Números*. Retrieved from <http://www.apdp.pt/index.php/diabetes/factos-e-numeros/diabetes-factos-e-numeros-2014>
- Porto Canal (2015). *Menos camas e fecho de SAP aumentaram afluência às urgências*. Retrieved from <http://portocanal.sapo.pt/noticia/71471/>
- Rosa, L., V. (Setembro, 2014). *A saúde em Portugal*. Retrieved from <http://visao.sapo.pt/a-saude-em-portugal=f795952>
- Santos, R., B. (Janeiro, 2015). *Caos nas urgências entre “o que acontece todos os anos” e “espelho dos problemas de todo o SNS”*. Retrieved from <http://www.publico.pt/sociedade/noticia/ps-diz-que-crise-nas-urgencias-e-o-espelho-dos-problemas-de-todo-o-sns-1681593>
- SAS (2015). SAS Enterprise Miner. *Reveal valuable insights with powerful data mining software*. Retrieved from http://www.sas.com/en_us/software/analytics/enterprise-miner.html
- Scarpelini, S. (2007). A Organização do Atendimento às Urgências e Trauma. *Medicina, Ribeirão Preto, 40(3)*, 315-317.
- Srikant, R., & Agrawal, R. (1997). Mining generalized association rules. *Future Generation Computer Systems, 13*, 161-162.
- Totten, V., & Bellou, A. (2013). Development of Emergency Medicine in Europe. *Academic Emergency Medicine, 20(5)*, 514-517.
- Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J., & Hua, L. (2012). Data Mining in Healthcare and Biomedicine: A Survey of the Literature. *Springer Science+Business Media, 36*, 2431-2438.
- Zhao, Y., & Karypis, G. (2002). Comparison of Agglomerative and Partitional Document Clustering Algorithms. 1-5.
- WEKA (2015). Weka 3: *Data Mining Software in Java*. Retrieved from <http://www.cs.waikato.ac.nz/ml/weka/>

8. ANEXOS

Class Variable Summary Statistics
(maximum 500 observations printed)

Data Role=TRAIN

Data Role	Variable Name	Role	Number of Levels	Missing	Mode	Mode Percentage	Mode2	Mode2 Percentage
TRAIN	Alta	INPUT	10	4	Alta para o Domicilio	61.42	ARS / Centro de Saúde	12.17
TRAIN	Codigo_Postal	INPUT	304	2	-1	27.02	4050	8.59
TRAIN	Concelho	INPUT	214	1	Porto	42.86	Desconhecido	27.02
TRAIN	Cor	INPUT	10	0	Amarelo	65.05	Verde	19.74
TRAIN	Destino	INPUT	113	4	Exterior Nao Referenciado	60.23	Ars / Centro De Saúde	12.17
TRAIN	Distrito	INPUT	25	4	Porto	67.23	Desconhecido	27.02
TRAIN	Localidade	INPUT	1165	0	Porto	42.86	Desconhecido	27.02
TRAIN	Pais	INPUT	3	4	Portugal	72.98	Desconhecido	27.02
TRAIN	Sexo	INPUT	4	11	feminino	54.10	masculino	45.88
TRAIN	Ultimo_Fluxograma__Razao	INPUT	55	0	Problemas nos membros	16.32	Indisposição no adulto	13.68

Interval Variable Summary Statistics
(maximum 500 observations printed)

Data Role=TRAIN

Variable	Role	Mean	Standard Deviation	Non Missing	Missing	Minimum	Median	Maximum	Skewness	Kurtosis
Data_Episodio	INPUT	19147.78	88.92888	100000	0	18993	19149	19305	-0.02664	-1.20847
Idade	INPUT	54.3984	20.68599	99995	5	-1	55	104	-0.14737	-0.86505

Figura 34 - Missing Values

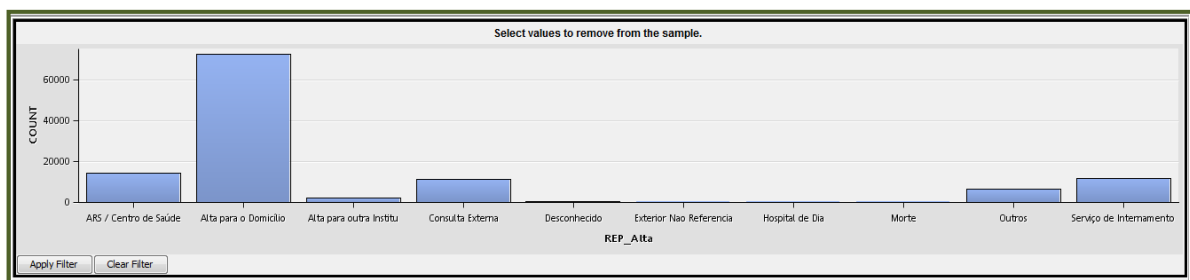
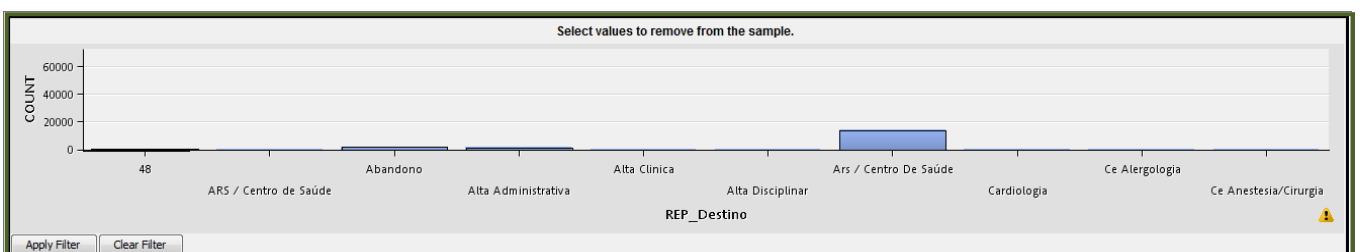
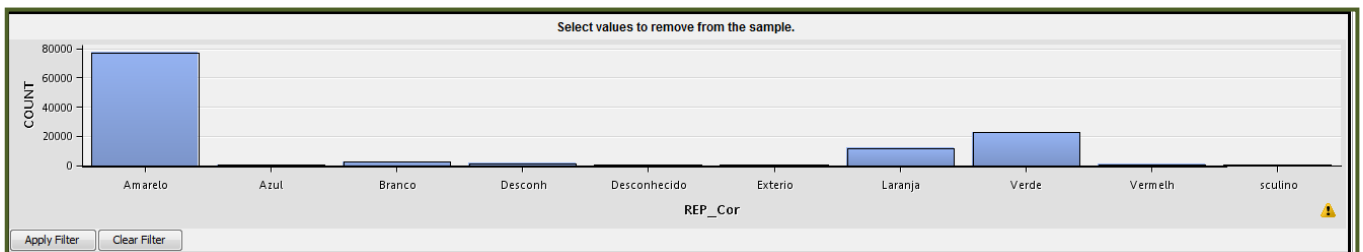


Figura 35 - Replacement da variável Alta

Name	Use	Report
Alta	Default	No
Codigo_Postal	Default	No
Concelho	Default	No
Cor	Default	No
Data_Episodio	Default	No
Destino	Default	No
Distrito	Default	No
ID_Cliente	Default	No
IMP_Idade	Default	No
Localidade	Default	No
Pais	Default	No
REP_Alta	Default	No
REP_Concelho	Default	No
REP_Cor	Default	No
REP_Destino	Default	No
REP_Distrito	Default	No
REP_Localidade	Default	No
REP_Pais	Default	No
REP_Sexo	Default	No
REP_Ultimo_Fluxo	Default	No
Sexo	Default	No
Ultimo_Fluxograma	Default	No

Figura 36 - Variáveis após utilizar o nó Replacement e Impute



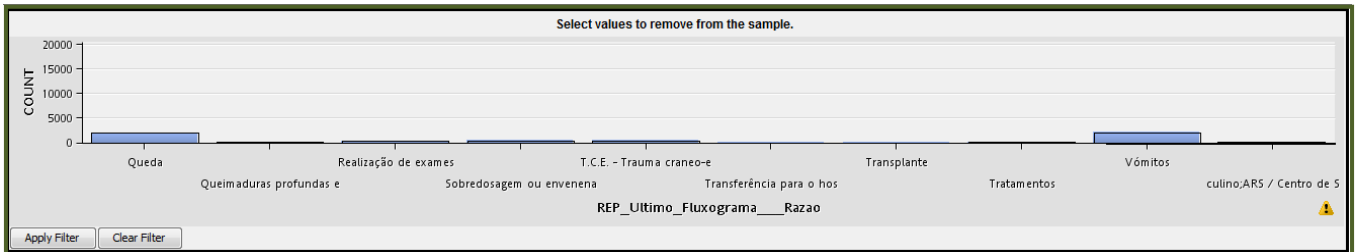
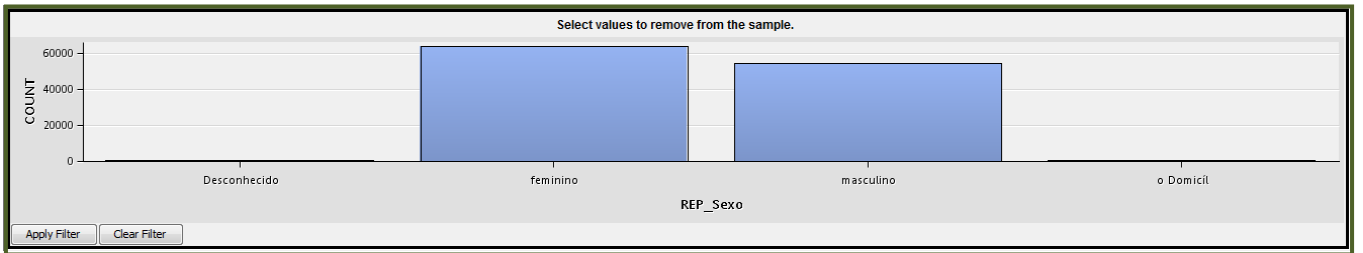
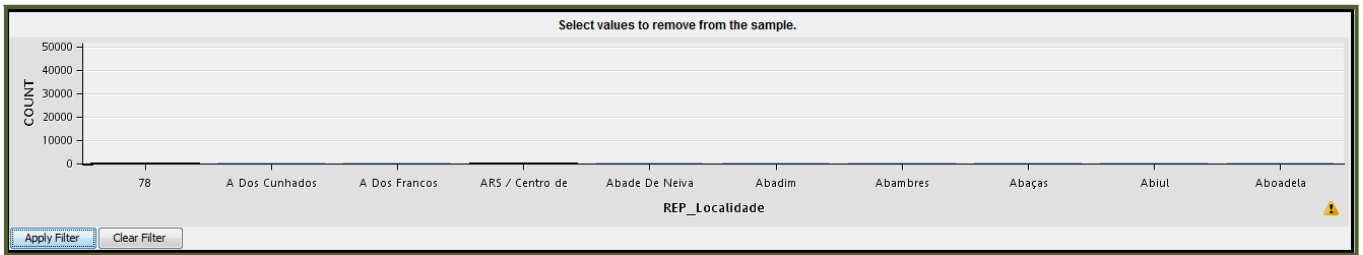


Figura 37 - Filtros nas variáveis de classe

Variable	Role	Level	Train Count	Train Percent	Label	Filter Method
Concelho	REJECTED		4	0.003382		MANUAL
Concelho	REJECTED	73	1	0.008455		MANUAL
Cor	REJECTED		3	0.002537		MANUAL
Cor	REJECTED	Exterio	1	0.008455		MANUAL
Cor	REJECTED	sculino	1	0.008455		MANUAL
Cor	REJECTED	úde	1	0.008455		MANUAL
Destino	REJECTED		7	0.005919		MANUAL
Distrito	REJECTED		7	0.005919		MANUAL
Localidade	REJECTED		3	0.002537		MANUAL
Localidade	REJECTED	78	1	0.008455		MANUAL
Pais	REJECTED		7	0.005919		MANUAL
REP_Alta	INPUT	Desconhecido	7	0.005919	Replacement: Alta	MANUAL
REP_Alta	INPUT	_BLANK_	0	0	Replacement: Alta	MANUAL
REP_Concelho	INPUT	73	1	0.008455	Replacement: Concelho	MANUAL
REP_Concelho	INPUT	Ars / Centro	1	0.008455	Replacement: Concelho	MANUAL
REP_Concelho	INPUT	Desconhecido	33536	28.35594	Replacement: Concelho	MANUAL
REP_Concelho	INPUT	_BLANK_	0	0	Replacement: Concelho	MANUAL
REP_Cor	INPUT	Desconhecido	3	0.002537	Replacement: Cor	MANUAL
REP_Cor	INPUT	Exterio	1	0.008455	Replacement: Cor	MANUAL
REP_Cor	INPUT	_BLANK_	0	0	Replacement: Cor	MANUAL
REP_Cor	INPUT	sculino	1	0.008455	Replacement: Cor	MANUAL
REP_Destino	INPUT	48	1	0.008455	Replacement: Destino	MANUAL
REP_Destino	INPUT	Desconhecido	7	0.005919	Replacement: Destino	MANUAL
REP_Destino	INPUT	_BLANK_	0	0	Replacement: Destino	MANUAL
REP_Distrito	INPUT	Desconhecido	33539	28.35847	Replacement: Distrito	MANUAL
REP_Distrito	INPUT	_BLANK_	0	0	Replacement: Distrito	MANUAL
REP_Localidade	INPUT	78	1	0.008455	Replacement: Localidade	MANUAL
REP_Localidade	INPUT	Ars / Centro De	1	0.008455	Replacement: Localidade	MANUAL
REP_Localidade	INPUT	Desconhecido	33535	28.35509	Replacement: Localidade	MANUAL
REP_Localidade	INPUT	_BLANK_	0	0	Replacement: Localidade	MANUAL
REP_Pais	INPUT	Desconhecido	33539	28.35847	Replacement: Pais	MANUAL
REP_Pais	INPUT	_BLANK_	0	0	Replacement: Pais	MANUAL
REP_Sexo	INPUT	Desconhecido	71	0.060033	Replacement: Sexo	MANUAL
REP_Sexo	INPUT	_BLANK_	0	0	Replacement: Sexo	MANUAL
REP_Sexo	INPUT	o Domicil	1	0.008455	Replacement: Sexo	MANUAL
REP_Ultimo_Fluxograma__Ra...	INPUT	Desconhecido	1732	1.464471	Replacement: Ultimo_Fluxogram...	MANUAL
REP_Ultimo_Fluxograma__Ra...	INPUT	_BLANK_	0	0	Replacement: Ultimo_Fluxogram...	MANUAL
REP_Ultimo_Fluxograma__Ra...	INPUT	culino;ARS / Centro de S	1	0.008455	Replacement: Ultimo_Fluxogram...	MANUAL

Figura 38 - Filtros nas variáveis de classe

The screenshot shows the Weka Explorer interface with the 'NumericToNominal' filter selected in the 'Filter' list. The 'Selected attribute' panel for 'Cor' displays the following data:

No.	Label	Count
1	Laranja	12008
2	Amarelo	77489
3	Verde	22728
4	Azul	674
5	Branco	2685
6	Vermelho	949
7	Desconhecido	1729

The bar chart below the table visualizes these counts, with the 'Amarelo' bar being the tallest at 77489.

Figura 39 – Opção NumericToNominal

```

%Let Segmentacao=ClusterPonto1;

Data &EM_EXPORT_VALIDATE.;

Set &EM_IMPORT_DATA.(Keep=%EM_ID_Segment_Label_);

  Rename _Segment_Label_=Seg_&Segmentacao.;

Run;

%Macro Ficheiro;

Proc Sql;

  Select Sum(Distance),

    Max(_SEGMENT_) into:Distance,

      :NCluster From &EM_Import_data.;

Quit;

%Let Distance=&Distance.;

%Let NCluster=&NCluster.;

%Put =&NCluster.=&Distance.;;

  %If %sysfunc(exist(&Em_Lib..&Segmentacao.,DATA)) ne %str(1) or &NCluster. eq 1 %Then %Do;

    Data &Em_Lib..&Segmentacao.;

      NumClusters=&NCluster.;

      ErroTotal=&Distance.;

      Output;

      Run;

  %End;

  %Else %Do;

    Data New;

      NumClusters=&NCluster.;

      ErroTotal=&Distance.;

      Output;

      Run;

```

```

Data &Em_Lib..&Segmentacao.;

Set &Em_Lib..&Segmentacao. New;

Run;

%End;

%EM_Register(Key=SEG,Type=DATA);

Proc Sql;

Create Table &EM_USER_SEG As

Select NumClusters,

    Avg(ErroTotal) As AvgErro,

    Min(ErroTotal) As MinErro,

    Max(ErroTotal) As MaxErro

From &Em_Lib..&Segmentacao.

Group By NumClusters;

Quit;

%EM_Report(Key=SEG,

    Viewtype=DATA,

    Autodisplay=Y,

    Block=Cotovelo,

    Description=Erro Total);

%Mend Ficheiro;

%Ficheiro;

```

Figura 40 - Código Gráfico Cotovelo

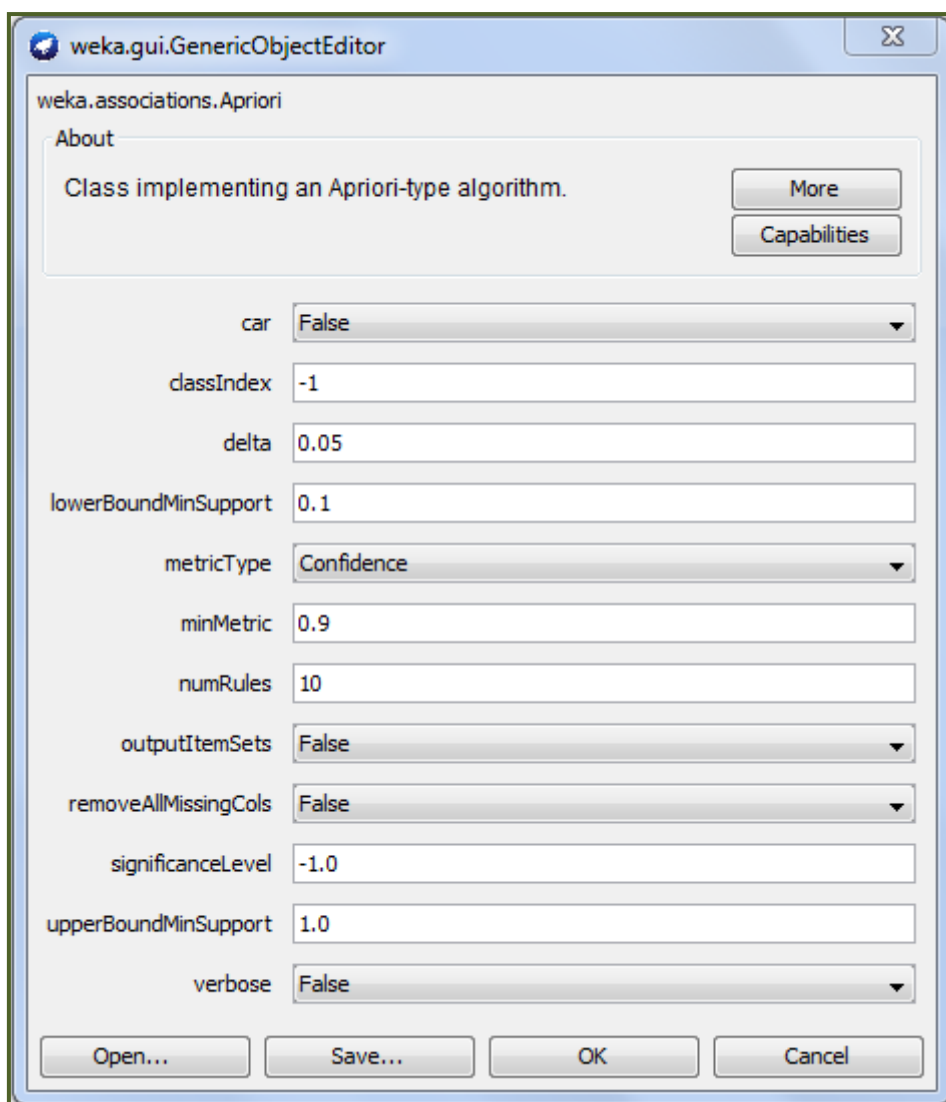


Figura 41 - Configurações originais para as regras de associação

	Cluster1	Cluster2	Cluster3	Total Geral
Janeiro	3466	1995	2150	7611
Fevereiro	3346	1885	2224	7455
Março	3404	1881	2343	7628
Abril	3051	1704	2098	6853
Mai	3449	1946	2322	7717
Junho	3284	1795	2339	7418
Julho	3383	1603	2503	7489
Agosto	2927	1240	2251	6418
Setembro	2826	1277	2207	6310
Outubro	2781	1269	2342	6392
Novembro	2617	1083	2211	5911
Dezembro	2851	1016	2359	6226

Tabela 5 – Tabela relativa ao número de pacientes em cada mês

	Cluster1	Cluster2	Cluster3
Morte	65	5	47
Falecido Com Autopsia	10	1	8
Falecido Sem Autopsia	52	4	39
Morgue (Entrou Cadaver)	3		

	Cluster1	Cluster2	Cluster3
Alta para o Domicílio	22309	13116	16578
Exterior Nao Referenciado	21874	12905	16209
Exterior Não Referenciado	1		
Servico Domiciliario	434	211	369

	Cluster1	Cluster2	Cluster3
Outros	1802	851	659
Abandono	832	376	250
Alta Administrativa	90	35	53
Alta Clinica	28	25	12
Alta Disciplinar	3		
Não Respondeu à Chamada	683	340	274
Saida Contra Parecer Do Medico	166	75	70

	Cluster1	Cluster2	Cluster3
Alta para outra Instituto	615	343	358
Hospital Nao Pertencente Sns	38	7	14
Outro Hospital	577	336	344

	Cluster1	Cluster2	Cluster3
ARS / Centro de Saúde	4465	2217	3555
Ars / Centro De Saúde	4465	2217	3555

	Cluster1	Cluster2	Cluster3
Consulta Externa	3814	1517	2614
Ce Alergologia	7	2	1
Ce Anestesia/Cirurgia	1		
Ce Anestesiologia			1
Ce Arritmologia E Pacing	11	1	9
Ce Cardiologia	285	32	223
Ce Centro Oftalmologico	44	23	41
Ce Cir. Máxilo Facial	55	18	19
Ce Cirurgia 1	164	44	82
Ce Cirurgia 2a	68	33	47
Ce Cirurgia 2b	70	25	43
Ce Cirurgia Ambulatorio	130	48	36
Ce Cirurgia Vasculuar	280	112	170
Ce Cuidados Intensivos	1		
Ce D.P.C.A.	20	8	6
Ce Dermatologia	82	49	49
Ce Dor	20	2	15
Ce Endoc-Grupo Hipofise	1		
Ce Endocrinologia	25	10	29
Ce Endocrinologia - Pe Diat	18		12
Ce Estomaterapia	1		1
Ce Estomatologia	66	26	29
Ce Fisiatria - Geral	1		
Ce Gastroenterologia	108	62	42
Ce Ginecologia		12	5
Ce Hemat-Grupo Linfomas	2	1	
Ce Hematologia Clinica	45	19	42
Ce Medicina 1a	83	25	80
Ce Medicina 1d	90	17	59
Ce Medicina 2b	132	54	90
Ce Medicina 2c	34	16	16
Ce Nefrologia	88	30	80
Ce Neurocirurgia	114	69	67
Ce Neurologia	479	262	337
Ce Nutricao		1	
Ce O.R.L.	91	34	35
Ce Obstetricia		6	
Ce Oftalmologia	281	101	193
Ce Oncologia Grupo Disges	3		4
Ce Oncologia Grupo Nao Di	4	1	2
Ce Oncologia Medica	52	17	34
Ce Ortopedia	547	188	612
Ce Pediatria	1	3	
Ce Pre-Transplante Renal	1		
Ce Psiquiatria - Primeira		1	1
Ce Psiquiatria Seguimento	3	3	
Ce Saude Ocupacional	10	30	3
Ce Sono	1		1
Ce Transplantes Renais	134	104	64
Ce Urologia	161	27	34
Ce Ventiloterapia		1	

	Cluster1	Cluster2	Cluster3
Serviço de Internamento	4271	633	3515
Área de Decisão Clínica/SO	1400	175	1395
Cardiologia	150	4	112
Cirurgia 1	101	23	72
Cirurgia 2	191	50	115
Cirurgia Vasculuar	189	8	86
Cuidados Intensivos	59	9	34
Dermatologia	2		2
Endocrinologia	28	6	31
Estom./C.M.Facial	17	3	1
Fisiatria			1
Gastroenterologia	64	16	31
Ginecologia			2
Hematologia Clinica	23	5	13
Medicina 1	205	19	198
Medicina 2	194	20	176
Nefrologia	101	47	79
Neurocirurgia	110	27	90
Neurologia	105	15	75
O.R.L.	57	22	20
Oftalmologia	45	10	26
Ortopedia	246	41	416
Ortopedia Infantil	1		
Pediatria	4	3	
S.C.I.N.Pediatricos		1	
S.C.I.P.C.	25	4	14
T.C.E.	98	5	72
U.C.I.C.	199	11	99
U.C.I.P.	79	7	25
U.C.P.O.			1
Unidade Cuidados Intermédi	223	25	165
Unidade Intermédia Médica	91	11	59
Urologia	264	66	105

	Cluster1	Cluster2	Cluster3
Hospital de Dia	44	12	23
Hd Cirurgia Ambulatori	16	5	4
Hd Cirurgia Vasculuar	2	1	
Hd Dermatologia	4	2	4
Hd Endocrinologia			1
Hd Hematologia	17	2	4
Hd Hemodialise		1	3
Hd Medicina			1
Hd Neurologia			1
Hd Oftalmologia	1		
Hd Oncologia Medica	1	1	1
Hd Poliv-Hematologia	2		2
Hd Poliv-Medicina			1
Hd Poliv-Nefrologia			1
Hd Poliv-Neurologia	1		

Tabela 6 - Tabela de Anexo relacionando as Variáveis “Alta Médica” e “Destino Final”

	Cluster1	Cluster2	Cluster3	Total Geral
Alta para o Domicílio	22309	13116	16578	52003
Alta para outra Institu	615	343	358	1316
ARS / Centro de Saúde	4465	2217	3555	10237
Consulta Externa	3814	1517	2614	7945
Hospital de Dia	44	12	23	79
Morte	65	5	47	117
Outros	1802	851	659	3312
Serviço de Internamento	4271	633	3515	8419

Tabela 7 - Tabela dos Anexos referente à Variável “Alta Médica”

	Cluster1	Cluster2	Cluster3	Total		Cluster1	Cluster2	Cluster3	Total
Agressão	441	195	62	698	Grande traumatismo	172	91	106	369
Asma	92	104	111	307	Gravidez		3		3
Auto agressão	9	10	1	20	Hemorragia GI	572	119	309	1000
Cefaleia	942	883	915	2740	Hemorragia vaginal		16	8	24
Comportamento estranho	90	19	89	198	Indicação do médico	717	219	365	1301
Convulsões	172	87	78	337	Indisposição no adulto	5146	2341	4524	12011
Corpo estranho	234	103	111	448	Infecções locais e abces	1062	524	382	1968
Criança irritável	10	7	1	18	Lesão toraco-abdominal	353	130	375	858
Criança que não se sente bem		2		2	Mordeduras e picadas	70	77	38	185
Diabetes	176	34	211	421	Outro	134	41	71	246
Diarreia	265	169	226	660	Pais preocupados	2	2		4
Dispneia	2546	592	2486	5624	Problemas estomatológico	420	246	137	803
Doença hematológica	7	3	15	25	Problemas nasais	470	94	229	793
Doença mental	8	18	11	37	Problemas nos membros	5662	2999	5344	14005
Doenças sexualmente tran		2		2	Problemas nos ouvidos	1168	737	942	2847
Dor abdominal	2484	1607	1591	5682	Problemas oftalmológicos	3577	1634	2371	7582
Dor cervical	224	296	188	708	Problemas urinários	1411	570	532	2513
Dor de garganta	599	594	249	1442	Queda	569	125	725	1419
Dor lombar	1800	1340	1259	4399	Queimaduras profundas e	43	56	24	123
Dor testicular	229	1	1	231	Realização de exames	126	44	58	228
Dor torácica	1944	692	1240	3876	Sobredosagem ou envenen	104	142	49	295
Embriaguês aparente	309	119	20	448	T.C.E. - Trauma craneo-e	141	50	113	304
Erupções cutâneas	601	530	452	1583	Transferência para o hos	24	5	24	53
Estado de inconsciência	23	2	22	47	Transplante	1	1		2
Exposição a químicos	5	3	1	9	Tratamentos	52	9	37	98
Feridas	1696	485	748	2929	Vômitos	483	522	498	1503

Tabela 8 - Tabela de Anexo sobre o motivo dos pacientes recorrem às urgências

	Cluster1	Cluster2	Cluster3	Total Geral
Amarelo	24035	13300	17889	55224
Azul	281	77	130	488
Branco	1054	319	555	1928
Laranja	4130	1144	3367	8641
Verde	7546	3778	5217	16541
Vermelho	339	76	191	606

Tabela 9 - Tabela dos Anexos referente à Variável “Cor”

Razão das Urgências

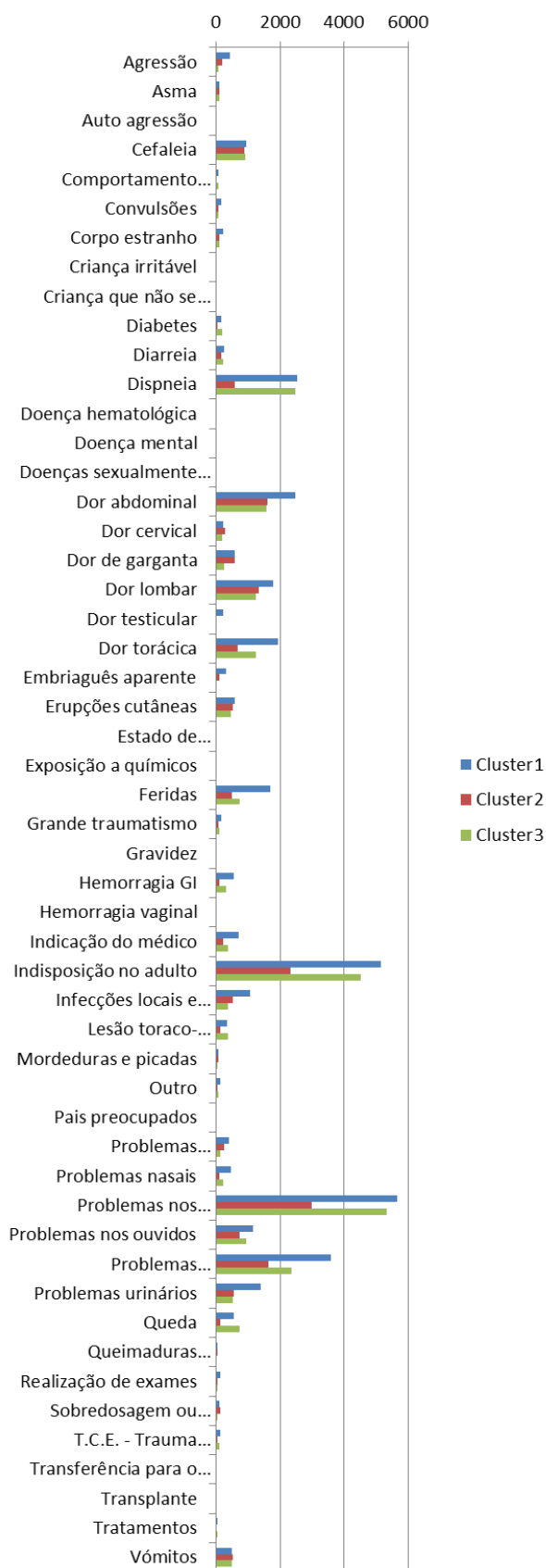


Figura 42 – Gráfico ampliado sobre o motivo dos pacientes recorrerem às urgências

	Cluster1	Cluster2	Cluster3	Total Geral
Aveiro	1260	629	532	2421
Beja	4	2		6
Braga	496	289	162	947
Bragança	283	107	134	524
Castelo Branco	7	4	3	14
Coimbra	35	31	5	71
Évora		3		3
Faro	18	10	2	30
Guarda	19	16	11	46
Ilha da Madeira	15	15	6	36
Ilha das Flores	1			1
Ilha de São Miguel	13	9	1	23
Ilha do Faial	1			1
Ilha do Pico			1	1
Ilha Terceira	4	3		7
Leiria	23	23	2	48
Lisboa	68	43	21	132
Portalegre	6	3	2	11
Porto	33933	17040	25941	76914
Santarém	11	13	4	28
Setúbal	13	14	4	31
Viana do Castelo	107	74	35	216
Vila Real	484	144	196	824
Viseu	584	222	287	1093

Tabela 10 - Tabela dos Anexos referente à Variável Distrito

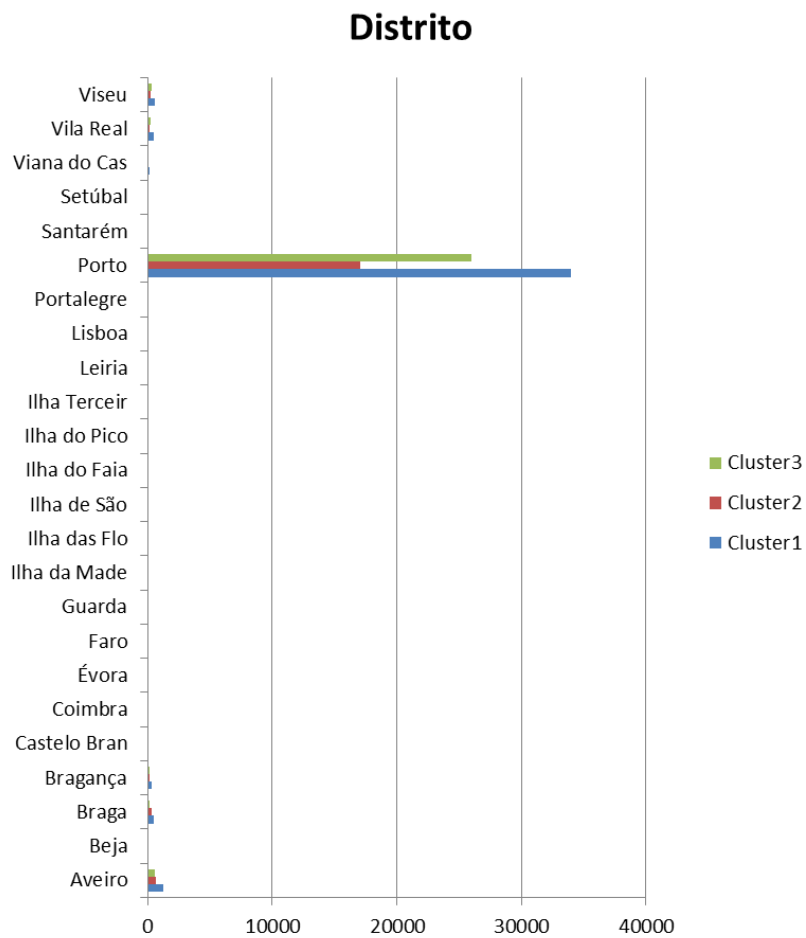


Figura 43 – Figura ampliada referente à Variável Distrito

	Cluster1	Cluster2	Cluster3
Agressão	441	195	62
Janeiro	38	21	2
Fevereiro	36	16	2
Março	49	21	8
Abril	34	15	1
Maio	47	16	2
Junho	48	22	11
Julho	40	18	8
Agosto	29	10	2
Setembro	30	19	6
Outubro	28	10	10
Novembro	30	13	4
Dezembro	32	14	6

	Cluster1	Cluster2	Cluster3
Asma	92	104	111
Janeiro	12	19	19
Fevereiro	10	13	8
Março	9	15	11
Abril	8	9	4
Maio	10	7	11
Junho	8	10	6
Julho	7	4	6
Agosto	2	3	6
Setembro	4	7	7
Outubro	6	10	13
Novembro	7	3	9
Dezembro	9	4	11

	Cluster1	Cluster2	Cluster3
Auto agressão	9	10	1
Janeiro		2	
Fevereiro			1
Março	4	2	
Abril		1	
Junho	2	1	
Julho	1		
Agosto	2	1	
Setembro		1	
Outubro		2	

	Cluster1	Cluster2	Cluster3
Cefaleia	942	883	915
Janeiro	71	93	49
Fevereiro	78	113	67
Março	85	84	71
Abril	75	88	70
Maio	77	85	61
Junho	75	76	91
Julho	97	60	79
Agosto	75	57	88
Setembro	87	57	73
Outubro	77	61	82
Novembro	71	58	93
Dezembro	74	51	91

	Cluster1	Cluster2	Cluster3
Comp. Estranho	90	19	89
Janeiro	7	3	5
Fevereiro	11	4	3
Março	3	1	7
Abril	6	1	11
Maio	7	1	8
Junho	6		5
Julho	7		8
Agosto	6	3	4
Setembro	8	3	12
Outubro	11	3	10
Novembro	6		7
Dezembro	12		9

	Cluster1	Cluster2	Cluster3
Convulsões	172	87	78
Janeiro	17	9	5
Fevereiro	21	10	9
Março	16	4	6
Abril	17	4	5
Maio	15	12	8
Junho	12	9	5
Julho	11	5	10
Agosto	14	8	6
Setembro	9	6	4
Outubro	17	8	8
Novembro	7	5	8
Dezembro	16	7	4

	Cluster1	Cluster2	Cluster3
Corpo estranho	234	103	111
Janeiro	16	5	2
Fevereiro	21	12	4
Março	20	18	16
Abril	24	8	5
Maio	22	13	13
Junho	20	10	13
Julho	20	11	10
Agosto	23	2	4
Setembro	20	9	11
Outubro	21	9	13
Novembro	5	2	9
Dezembro	22	4	11

	Cluster1	Cluster2	Cluster3
Criança irritável	10	7	1
Janeiro			1
Fevereiro	5		
Março	1	3	
Abril	1	2	
Maio	3		
Junho		1	
Agosto		1	
Cluster1 Cluster2 Cluster3			
Criança que não se sente	2		
Maio	1		
Dezembro	1		

	Cluster1	Cluster2	Cluster3
Diabetes	176	34	211
Janeiro	18	5	17
Fevereiro	13	6	18
Março	15	2	23
Abril	20	3	18
Maio	14	6	19
Junho	6	3	13
Julho	8		13
Agosto	16	1	18
Setembro	17	1	18
Outubro	15	3	17
Novembro	17	3	17
Dezembro	17	1	20

	Cluster1	Cluster2	Cluster3
Diabetes	176	34	211
Janeiro	18	5	17
Fevereiro	13	6	18
Março	15	2	23
Abril	20	3	18
Maio	14	6	19
Junho	6	3	13
Julho	8		13
Agosto	16	1	18
Setembro	17	1	18
Outubro	15	3	17
Novembro	17	3	17
Dezembro	17	1	20

	Cluster1	Cluster2	Cluster3
Diarreia	265	169	226
Janeiro	19	22	15
Fevereiro	17	11	15
Março	22	15	14
Abril	26	12	10
Maio	23	22	15
Junho	23	13	28
Julho	19	14	21
Agosto	29	12	23
Setembro	25	19	22
Outubro	18	11	21
Novembro	27	12	19
Dezembro	17	6	23

	Cluster1	Cluster2	Cluster3
Dispneia	2546	592	2486
Janeiro	310	72	260
Fevereiro	377	102	364
Março	204	50	235
Abril	169	52	194
Maio	191	47	205
Junho	185	48	163
Julho	176	37	149
Agosto	159	28	148
Setembro	149	40	139
Outubro	173	38	199
Novembro	227	42	191
Dezembro	226	36	239

	Cluster1	Cluster2	Cluster3
Doença hematológica	7	3	15
Janeiro	1		1
Fevereiro		1	
Março	1	1	3
Abril	2		
Junho		1	1
Julho	1		1
Agosto	1		2
Setembro			1
Outubro			1
Dezembro	1		2

	Cluster1	Cluster2	Cluster3
Doença mental	8	18	11
Janeiro		1	
Fevereiro			1
Março		4	
Abril			1
Maio		4	1
Junho		1	1
Julho			1
Agosto	2	4	
Setembro	3	1	3
Outubro			1
Novembro	2	1	1
Dezembro	1	2	1

	Cluster1	Cluster2	Cluster3
Doenças sexualmente transmissíveis	2		
Julho	1		
Dezembro		1	

	Cluster1	Cluster2	Cluster3
Exposição a queimaduras	5	3	1
Janeiro		1	1
Fevereiro	1		
Março	1	1	
Junho	1		
Julho	2	1	

	Cluster1	Cluster2	Cluster3
Dor abdominal	2484	1607	1591
Janeiro	223	158	100
Fevereiro	202	141	107
Março	224	132	126
Abril	200	154	124
Mai	207	187	135
Junho	229	161	150
Julho	255	131	129
Agosto	212	105	138
Setembro	188	130	135
Outubro	189	94	158
Novembro	194	121	150
Dezembro	161	93	139

	Cluster1	Cluster2	Cluster3
Dor cervical	224	296	188
Janeiro	22	26	13
Fevereiro	9	24	14
Março	14	35	11
Abril	15	40	12
Mai	22	20	16
Junho	21	30	20
Julho	22	27	31
Agosto	24	20	11
Setembro	19	22	11
Outubro	16	18	10
Novembro	23	25	21
Dezembro	17	9	18

	Cluster1	Cluster2	Cluster3
Dor de garganta	599	594	249
Janeiro	65	56	11
Fevereiro	63	62	17
Março	64	50	21
Abril	51	67	19
Mai	52	58	23
Junho	62	49	29
Julho	60	53	20
Agosto	47	36	21
Setembro	23	34	19
Outubro	33	49	23
Novembro	41	37	20
Dezembro	38	43	26

	Cluster1	Cluster2	Cluster3
Dor lombar	1800	1340	1259
Janeiro	157	123	85
Fevereiro	131	105	76
Março	145	124	83
Abril	141	132	74
Mai	143	135	112
Junho	175	131	116
Julho	166	133	128
Agosto	172	104	118
Setembro	161	95	103
Outubro	161	103	116
Novembro	121	94	114
Dezembro	127	61	134

	Cluster1	Cluster2	Cluster3
Dor testicular	229	1	1
Janeiro	10		
Fevereiro	19	1	
Março	28		1
Abril	24		
Mai	15		
Junho	23		
Julho	13		
Agosto	20		
Setembro	25		
Outubro	15		
Novembro	20		
Dezembro	17		

	Cluster1	Cluster2	Cluster3
Dor torácica	1944	692	1240
Janeiro	186	71	109
Fevereiro	171	72	107
Março	179	78	117
Abril	167	63	101
Mai	204	80	90
Junho	149	66	93
Julho	155	44	112
Agosto	141	24	91
Setembro	152	42	101
Outubro	162	51	98
Novembro	140	53	112
Dezembro	138	48	109

	Cluster1	Cluster2	Cluster3
Embraguês ap	309	119	20
Janeiro	26	17	
Fevereiro	21	14	
Março	30	3	2
Abril	26	9	1
Mai	28	16	
Junho	37	15	4
Julho	26	9	1
Agosto	16	4	3
Setembro	36	13	1
Outubro	21	7	2
Novembro	19	8	3
Dezembro	23	4	3

	Cluster1	Cluster2	Cluster3
Erupções cutâne	601	530	452
Janeiro	44	49	14
Fevereiro	38	36	16
Março	50	48	30
Abril	60	42	39
Mai	64	61	37
Junho	51	62	50
Julho	66	48	58
Agosto	45	51	59
Setembro	50	32	54
Outubro	41	38	41
Novembro	43	34	27
Dezembro	49	29	27

	Cluster1	Cluster2	Cluster3
Estado de incons	23	2	22
Janeiro	2		4
Fevereiro	3		1
Março	4	1	3
Abril			4
Mai	6		1
Junho	3		
Julho	1		
Agosto			2
Setembro	1		
Outubro	1		3
Novembro	2	1	2
Dezembro			2

	Cluster1	Cluster2	Cluster3
Exposição a qu	5	3	1
Janeiro		1	1
Fevereiro	1		
Março	1	1	
Junho	1		
Julho	2	1	

	Cluster1	Cluster2	Cluster3
Feridas	1696	485	748
Janeiro	152	62	56
Fevereiro	164	41	49
Março	138	38	77
Abril	141	51	66
Mai	147	44	56
Junho	150	56	61
Julho	163	59	85
Agosto	141	36	52
Setembro	157	26	74
Outubro	117	38	66
Novembro	99	16	48
Dezembro	127	18	58

	Cluster1	Cluster2	Cluster3
Grande traumati	172	91	106
Janeiro	19	14	12
Fevereiro	16	13	15
Março	18	8	11
Abril	18	7	7
Mai	25	9	15
Junho	19	7	8
Julho	15	5	11
Agosto	10	7	3
Setembro	11	6	4
Outubro	2	5	6
Novembro	12	4	9
Dezembro	7	6	5

	Cluster1	Cluster2	Cluster3
Gravidez		3	
Março		2	
Julho		1	

	Cluster1	Cluster2	Cluster3
Hemorragia GI	572	119	309
Janeiro	71	11	32
Fevereiro	47	13	25
Março	61	16	41
Abril	53	11	22
Mai	52	14	33
Junho	41	16	18
Julho	49	10	18
Agosto	45	2	17
Setembro	39	10	28
Outubro	21	6	22
Novembro	56	6	24
Dezembro	37	4	29

	Cluster1	Cluster2	Cluster3
Hemorragia vaginal	16	8	
Janeiro		2	
Fevereiro			1
Março		3	
Abril		1	1
Mai		1	1
Junho		2	3
Julho		1	
Agosto		1	1
Setembro			1
Outubro		3	
Novembro		1	
Dezembro		1	

	Cluster1	Cluster2	Cluster3
Indicação do mé	717	219	365
Janeiro	76	20	24
Fevereiro	55	25	34
Março	67	17	32
Abril	50	23	21
Mai	77	31	25
Junho	59	25	35
Julho	97	31	55
Agosto	53	14	28
Setembro	39	9	33
Outubro	54	13	32
Novembro	53	6	19
Dezembro	37	5	27

	Cluster1	Cluster2	Cluster3
Indisposição no	5146	2341	4524
Janeiro	519	284	427
Fevereiro	545	275	465
Março	472	241	417
Abril	382	184	321
Maio	448	255	380
Junho	407	174	327
Julho	460	192	418
Agosto	354	151	354
Setembro	345	134	315
Outubro	392	175	363
Novembro	356	140	358
Dezembro	466	136	379

	Cluster1	Cluster2	Cluster3
Infecções locais	1062	524	382
Janeiro	78	45	23
Fevereiro	75	49	15
Março	77	42	33
Abril	93	53	38
Maio	114	63	37
Junho	100	52	36
Julho	101	47	46
Agosto	115	49	41
Setembro	86	30	40
Outubro	74	36	25
Novembro	49	32	26
Dezembro	100	26	22

	Cluster1	Cluster2	Cluster3
Lesão toraco-abd	353	130	375
Janeiro	31	18	30
Fevereiro	28	9	25
Março	52	22	27
Abril	34	10	29
Maio	27	11	38
Junho	20	15	35
Julho	24	10	19
Agosto	29	7	31
Setembro	20	6	28
Outubro	20	9	45
Novembro	30	7	28
Dezembro	38	6	40

	Cluster1	Cluster2	Cluster3
Mordeduras e p	70	77	38
Janeiro	2	5	1
Fevereiro	3	2	1
Março	6	5	1
Abril	2	8	1
Maio	5	5	5
Junho	9	12	3
Julho	12	6	5
Agosto	6	7	4
Setembro	5	9	6
Outubro	8	7	4
Novembro	6	8	1
Dezembro	6	3	6

	Cluster1	Cluster2	Cluster3
Outro	134	41	71
Janeiro	19	5	5
Fevereiro	10	6	9
Março	8	5	3
Abril	11	6	7
Maio	11	4	6
Junho	7	2	3
Julho	12	4	7
Agosto	16	7	7
Setembro	10	2	4
Outubro	7	2	6
Novembro	9	1	6
Dezembro	14	4	8

	Cluster1	Cluster2	Cluster3
Pais preocupado	2	2	
Janeiro		1	
Fevereiro		1	
Março		1	
Abril	1		
Setembro	1		

	Cluster1	Cluster2	Cluster3
Transplante	1	1	
Janeiro		1	
Fevereiro		1	
Junho	1		

	Cluster1	Cluster2	Cluster3
Problemas esto	420	246	137
Janeiro	26	19	2
Fevereiro	32	17	7
Março	47	22	8
Abril	35	22	11
Maio	48	29	9
Junho	34	20	14
Julho	37	31	18
Agosto	29	15	15
Setembro	36	19	17
Outubro	38	20	12
Novembro	23	15	13
Dezembro	35	17	11

	Cluster1	Cluster2	Cluster3
Problemas nasa	470	94	229
Janeiro	47	9	28
Fevereiro	71	15	24
Março	52	11	28
Abril	35	13	21
Maio	37	8	21
Junho	49	5	14
Julho	39	3	15
Agosto	29	3	14
Setembro	32	12	16
Outubro	25	1	11
Novembro	30	8	14
Dezembro	24	6	23

	Cluster1	Cluster2	Cluster3
Problemas nos n	5662	2999	5344
Janeiro	486	320	398
Fevereiro	441	288	321
Março	519	320	413
Abril	460	265	427
Maio	541	293	470
Junho	563	347	467
Julho	529	295	527
Agosto	455	204	513
Setembro	454	218	485
Outubro	426	174	478
Novembro	371	126	407
Dezembro	417	149	438

	Cluster1	Cluster2	Cluster3
Problemas nos	1168	737	942
Janeiro	103	81	63
Fevereiro	89	76	51
Março	118	90	89
Abril	78	64	75
Maio	123	74	90
Junho	98	74	103
Julho	117	60	94
Agosto	133	53	81
Setembro	93	45	79
Outubro	72	42	63
Novembro	60	36	89
Dezembro	84	42	65

	Cluster1	Cluster2	Cluster3
Problemas oftal	3577	1634	2371
Janeiro	312	198	169
Fevereiro	285	174	171
Março	347	193	206
Abril	320	160	198
Maio	364	166	217
Junho	321	158	241
Julho	323	116	231
Agosto	272	111	193
Setembro	276	108	189
Outubro	289	90	195
Novembro	234	77	184
Dezembro	234	83	177

	Cluster1	Cluster2	Cluster3
Problemas uriná	1411	570	532
Janeiro	142	59	29
Fevereiro	113	49	33
Março	117	60	43
Abril	151	43	44
Maio	127	69	46
Junho	128	41	48
Julho	122	56	43
Agosto	94	44	40
Setembro	108	36	53
Outubro	119	44	58
Novembro	97	35	45
Dezembro	93	34	50

	Cluster1	Cluster2	Cluster3
Queda	569	125	725
Janeiro	43	9	77
Fevereiro	55	13	73
Março	54	14	58
Abril	40	8	56
Maio	66	18	67
Junho	56	10	47
Julho	45	12	56
Agosto	37	6	55
Setembro	37	7	51
Outubro	36	13	60
Novembro	50	9	57
Dezembro	50	6	68

	Cluster1	Cluster2	Cluster3
Queimaduras pr	43	56	24
Janeiro	1	3	1
Fevereiro	1	1	
Março	2	3	
Abril	6	4	1
Maio	3	5	3
Junho	7	6	3
Julho	5	13	2
Agosto	7	6	2
Setembro	3	5	2
Outubro	4	3	1
Novembro	3	2	2
Dezembro	1	7	7

	Cluster1	Cluster2	Cluster3
Realização de ex	126	44	58
Janeiro	24	6	6
Fevereiro	14	7	5
Março	11	5	4
Abril	6	6	3
Maio	9	5	5
Junho	15	3	5
Julho	13	5	4
Agosto	5	2	2
Setembro	7	3	10
Outubro	6	2	6
Novembro	10	6	6
Dezembro	6	2	2

	Cluster1	Cluster2	Cluster3
Sobredosagem	104	142	49
Janeiro	7	10	3
Fevereiro	10	16	5
Março	7	13	8
Abril	9	9	3
Maio	8	17	3
Junho	8	12	6
Julho	14	8	6
Agosto	12	19	5
Setembro	8	12	1
Outubro	6	10	3
Novembro	11	6	2
Dezembro	4	10	4

	Cluster1	Cluster2	Cluster3
T.C.E. - Trauma	141	50	113
Janeiro	10	8	15
Fevereiro	11	5	17
Março	17	7	5
Abril	10	6	9
Maio	24	5	6
Junho	8	4	3
Julho	13	1	10
Agosto	6	2	9
Setembro	11	4	5
Outubro	12	3	5
Novembro	9	2	16
Dezembro	10	3	13

	Cluster1	Cluster2	Cluster3
Transferência pa	24	5	24
Janeiro	3	1	
Fevereiro	2		3
Março	3		2
Abril	2	1	3
Maio		2	
Junho	2		3
Julho	2		4
Agosto	2		2
Setembro	2		
Outubro	2	1	2
Novembro	1		3
Dezembro	3		2

	Cluster1	Cluster2	Cluster3
Tratamentos	52	9	37
Janeiro	6	1	4
Fevereiro	4		3
Março	6		4
Abril	1	1	5
Maio	5	3	3
Junho	3	1	6
Julho	5	1	2
Agosto	3		2
Setembro	2	1	3
Outubro	5	1	1
Novembro	5		3
Dezembro	7		1

	Cluster1	Cluster2	Cluster3
Vômitos	483	522	498
Janeiro	45	52	32
Fevereiro	28	46	42
Março	37	51	45
Abril	47	46	33
Maio	38	44	29
Junho	43	44	51
Julho	33	40	42
Agosto	39	29	36
Setembro	37	44	43
Outubro	41	56	52
Novembro	41	36	44
Dezembro	54	34	49

Tabela 11 - Tabela dos Anexos relacionando as Variáveis “Data” e “Razão”

	Cluster1	Cluster2	Cluster3
Agressão	441	195	62
Amarelo	329	136	40
Azul	1		
Laranja	20	6	2
Verde	91	52	20
Vermelh		1	

	Cluster1	Cluster2	Cluster3
Asma	92	104	111
Amarelo	68	69	89
Laranja	23	34	20
Verde	1	1	2

	Cluster1	Cluster2	Cluster3
Auto agressão	9	10	1
Amarelo	2	6	1
Laranja	6	4	
Vermelh	1		

	Cluster1	Cluster2	Cluster3
Cefaleia	942	883	915
Amarelo	757	772	781
Azul	5	1	
Laranja	35	42	39
Verde	142	68	92
Vermelh	3		3

	Cluster1	Cluster2	Cluster3
Comportam	90	19	89
Amarelo	43	9	40
Laranja	36	8	45
Verde	6	1	3
Vermelh	5	1	1

	Cluster1	Cluster2	Cluster3
Convulsões	172	87	78
Amarelo	107	62	44
Laranja	43	14	25
Verde	4	1	2
Vermelh	18	10	7

	Cluster1	Cluster2	Cluster3
Corpo estrar	234	103	111
Amarelo	128	64	63
Azul	1	1	
Laranja	3		
Verde	101	38	48
Vermelh	1		

	Cluster1	Cluster2	Cluster3
Criança irrita	10	7	1
Amarelo	6	5	1
Laranja	2	1	
Verde	1	1	
Vermelh	1		

	Cluster1	Cluster2	Cluster3
Criança que não se sente		2	
Vermelh		2	

	Cluster1	Cluster2	Cluster3
Diabetes	176	34	211
Amarelo	79	13	84
Laranja	42	6	74
Verde	17	4	15
Vermelh	38	11	38

	Cluster1	Cluster2	Cluster3
Diarreia	265	169	226
Amarelo	243	160	206
Azul	2		
Laranja	4		5
Verde	16	9	15

	Cluster1	Cluster2	Cluster3
Dispneia	2546	592	2486
Amarelo	1309	421	1233
Azul	4		2
Laranja	1068	120	1137
Verde	95	48	62
Vermelh	70	3	52

	Cluster1	Cluster2	Cluster3
Doença hem	7	3	15
Amarelo	6	2	10
Laranja			4
Verde	1	1	1

	Cluster1	Cluster2	Cluster3
Doença men	8	18	11
Amarelo	7	12	8
Laranja		2	
Verde	1	4	3

	Cluster1	Cluster2	Cluster3
Doenças sexualmente tra		2	
Azul		1	
Verde		1	

	Cluster1	Cluster2	Cluster3
Dor abdomir	2484	1607	1591
Amarelo	1992	1324	1347
Azul		2	1
Laranja	216	115	73
Verde	274	166	169
Vermelh	2		1

	Cluster1	Cluster2	Cluster3
Dor cervical	224	296	188
Amarelo	179	246	144
Azul	4		1
Laranja	1	4	2
Verde	40	46	40
Vermelh			1

	Cluster1	Cluster2	Cluster3
Dor de gargant	599	594	249
Amarelo	379	439	146
Azul	2	2	1
Laranja	2		
Verde	216	153	102

	Cluster1	Cluster2	Cluster3
Dor lombar	1800	1340	1259
Amarelo	1337	1043	1005
Azul	13	3	6
Laranja	239	177	69
Verde	211	117	179

	Cluster1	Cluster2	Cluster3
Dor testicula	229	1	1
Amarelo	167	1	1
Azul	1		
Laranja	21		
Verde	40		

	Cluster1	Cluster2	Cluster3
Dor torácia	1944	692	1240
Amarelo	1038	503	656
Azul	2		1
Laranja	835	156	531
Verde	58	33	51
Vermelh	11		1

	Cluster1	Cluster2	Cluster3
Embraguês	309	119	20
Amarelo	213	90	11
Azul	43	11	4
Laranja	5	2	
Verde	41	12	5
Vermelh	7	4	

	Cluster1	Cluster2	Cluster3
Erupções cut	601	530	452
Amarelo	314	323	250
Azul	7	2	2
Laranja	27	28	27
Verde	253	177	173

	Cluster1	Cluster2	Cluster3
Estado de incor	23	2	22
Amarelo			2
Laranja	6	1	11
Vermelh	17	1	9

	Cluster1	Cluster2	Cluster3
Exposição a c	5	3	1
Amarelo	3	1	
Laranja	2	2	1

	Cluster1	Cluster2	Cluster3
Feridas	1696	485	748
Amarelo	1240	331	567
Azul	10	5	4
Laranja	60	6	12
Verde	386	143	165

	Cluster1	Cluster2	Cluster3
Grande trauma	172	91	106
Amarelo	111	71	76
Laranja	46	18	26
Vermelh	15	2	4

	Cluster1	Cluster2	Cluster3
Gravidez		3	
Amarelo		1	
Verde		2	

	Cluster1	Cluster2	Cluster3
Hemorragia	572	119	309
Amarelo	427	102	231
Laranja	123	16	74
Verde	10	1	1

	Cluster1	Cluster2	Cluster3
Hemorragia vaginal	16	8	8
Amarelo		8	6
Laranja		5	2
Verde		3	

	Cluster1	Cluster2	Cluster3
Indicação do	717	219	365
Branco	717	219	365

	Cluster1	Cluster2	Cluster3
Indisposição	5146	2341	4524
Amarelo	3643	1763	3141
Azul	55	15	29
Laranja	783	170	887
Verde	608	385	417
Vermelh	57	8	50

	Cluster1	Cluster2	Cluster3
Infecções locais	1062	524	382
Amarelo	602	317	200
Azul	9		7
Laranja	6	1	3
Verde	445	206	172

	Cluster1	Cluster2	Cluster3
Lesão toraco	353	130	375
Amarelo	280	100	304
Azul		2	
Laranja	9	3	11
Verde	64	25	60

	Cluster1	Cluster2	Cluster3
Mordeduras	70	77	38
Amarelo	29	35	20
Azul	1		
Laranja	1	1	1
Verde	39	41	17

	Cluster1	Cluster2	Cluster3
Outro	134	41	71
Branco	134	41	71

	Cluster1	Cluster2	Cluster3
Pais preocup	2	2	
Amarelo	2	1	
Verde		1	

	Cluster1	Cluster2	Cluster3
Problemas e	420	246	137
Amarelo	310	201	82
Azul	2	1	
Laranja	7	2	2
Verde	101	42	53

	Cluster1	Cluster2	Cluster3
Problemas nase	470	94	229
Amarelo	291	53	142
Azul	2		
Laranja	9	2	9
Verde	168	39	78

	Cluster1	Cluster2	Cluster3
Problemas n	5662	2999	5344
Amarelo	3969	2194	3848
Azul	35	9	24
Laranja	137	32	101
Verde	1520	764	1371
Vermelh	1		

	Cluster1	Cluster2	Cluster3
Problemas n	1168	737	942
Amarelo	430	345	328
Azul	35	7	18
Laranja	1	1	
Verde	702	384	596

	Cluster1	Cluster2	Cluster3
Problemas ofta	3577	1634	2371
Amarelo	1875	904	1179
Azul	30	9	25
Laranja	69	21	22
Verde	1564	670	1135
Vermelh	39	30	10

	Cluster1	Cluster2	Cluster3
Problemas u	1411	570	532
Amarelo	1044	432	395
Azul	16	5	5
Laranja	69	21	7
Verde	281	112	125
Vermelh	1		

	Cluster1	Cluster2	Cluster3
Queda	569	125	725
Amarelo	457	115	626
Azul	1		
Laranja	65	4	73
Verde	28	5	19
Vermelh	18	1	7

	Cluster1	Cluster2	Cluster3
Queimaduras p	43	56	24
Amarelo	30	40	14
Laranja	11	11	6
Verde	2	5	3
Vermelh			1

	Cluster1	Cluster2	Cluster3
Realização de	126	44	58
Branco	126	44	58

	Cluster1	Cluster2	Cluster3
Sobredosage	104	142	49
Amarelo	39	40	10
Laranja	60	100	37
Verde	1		
Vermelh	4	2	2

	Cluster1	Cluster2	Cluster3
T.C.E. - Trauma	141	50	113
Amarelo	99	43	95
Laranja	22	5	10
Verde	2	2	7
Vermelh	18		1

	Cluster1	Cluster2	Cluster3
Realização de	126	44	58
Branco	126	44	58

	Cluster1	Cluster2	Cluster3
Sobredosage	104	142	49
Amarelo	39	40	10
Laranja	60	100	37
Verde	1		
Vermelh	4	2	2

	Cluster1	Cluster2	Cluster3
T.C.E. - Trauma	141	50	113
Amarelo	99	43	95
Laranja	22	5	10
Verde	2	2	7
Vermelh	18		1

	Cluster1	Cluster2	Cluster3
Transferênci	24	5	24
Branco	24	5	24

	Cluster1	Cluster2	Cluster3
Transplante	1	1	
Branco	1	1	

	Cluster1	Cluster2	Cluster3
Tratamentos	52	9	37
Branco	52	9	37

	Cluster1	Cluster2	Cluster3
Vômitos	483	522	498
Amarelo	451	503	463
Azul		1	
Laranja	16	3	19
Verde	16	15	16

Tabela 12 - Tabela dos Anexos relacionando as Variáveis “Cor” e “Razão”

Size of set of large itemsets L(1): 11

Large Itemsets L(1):

Último Fluxograma / Razão=Problemas nos membros 19185
Último Fluxograma / Razão=Indisposição no adulto 16302
Cor=Laranja 12008
Cor=Amarelo 77489
Cor=Verde 22728
Distrito=Porto 78073
Distrito=Desconhecido 33532
Sexo=masculino 54306
Sexo=feminino 63892
Alta=ARS / Centro de Saúde 14206
Alta=Alta para o Domicílio 72479

Large Itemsets L(2):

Último Fluxograma / Razão=Problemas nos membros Cor=Amarelo 13867
Último Fluxograma / Razão=Problemas nos membros Distrito=Porto 13014
Último Fluxograma / Razão=Problemas nos membros Alta=Alta para o Domicílio 13400
Cor=Amarelo Distrito=Porto 50946
Cor=Amarelo Distrito=Desconhecido 22195
Cor=Amarelo Sexo=masculino 34553
Cor=Amarelo Sexo=feminino 42887
Cor=Amarelo Alta=Alta para o Domicílio 48749
Cor=Verde Distrito=Porto 15324
Cor=Verde Sexo=feminino 12120
Cor=Verde Alta=Alta para o Domicílio 17540
Distrito=Porto Sexo=masculino 34498
Distrito=Porto Sexo=feminino 43556
Distrito=Porto Alta=Alta para o Domicílio 48577
Distrito=Desconhecido Sexo=masculino 16285
Distrito=Desconhecido Sexo=feminino 17202
Distrito=Desconhecido Alta=Alta para o Domicílio 20404
Sexo=masculino Alta=Alta para o Domicílio 32021
Sexo=feminino Alta=Alta para o Domicílio 40416

Large Itemsets L(3):

Cor=Amarelo Distrito=Porto Sexo=masculino 21874
Cor=Amarelo Distrito=Porto Sexo=feminino 29057
Cor=Amarelo Distrito=Porto Alta=Alta para o Domicílio 32582
Cor=Amarelo Distrito=Desconhecido Alta=Alta para o Domicílio 13853
Cor=Amarelo Sexo=masculino Alta=Alta para o Domicílio 20879
Cor=Amarelo Sexo=feminino Alta=Alta para o Domicílio 27836
Cor=Verde Distrito=Porto Alta=Alta para o Domicílio 11860
Distrito=Porto Sexo=masculino Alta=Alta para o Domicílio 20580
Distrito=Porto Sexo=feminino Alta=Alta para o Domicílio 27985

Size of set of large itemsets L(4): 2

Large Itemsets L(4):

Cor=Amarelo Distrito=Porto Sexo=masculino Alta=Alta para o Domicílio 13403

Cor=Amarelo Distrito=Porto Sexo=feminino Alta=Alta para o Domicílio 19169

Figura 44 - *Items* para as regras de associação