



**Hugo André dos Santos Antunes**

Mestre em Engenharia Eletrotécnica

**Visualização e deteção offline de eventos de tráfego usando o processamento de eventos complexos**

Dissertação para obtenção do Grau de Mestre em  
**Engenharia Eletrotécnica e de Computadores**

Orientador: Ricardo Luís Rosa Jardim Gonçalves, Professor,  
Faculdade de Ciências e Tecnologia da  
Universidade Nova de Lisboa

Co-orientador: Ruben Duarte Dias da Costa, Professor,  
Faculdade de Ciências e Tecnologia da  
Universidade Nova de Lisboa

Júri

Presidente: Doutor Paulo da Costa Luís da Fonseca Pinto - FCT/UNL  
Arguente: Doutor João Francisco Alves Martins - FCT/UNL  
Vogal: Doutor Ruben Dias da Costa - FCT/UNL



FACULDADE DE  
CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE NOVA DE LISBOA

Setembro, 2017



## **Visualização e deteção offline de eventos de tráfego usando o processamento de eventos complexos**

Copyright © Hugo André dos Santos Antunes, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.



---



## AGRADECIMENTOS

Em primeiro, gostaria de deixar alguns agradecimentos as pessoas que me ajudaram na realização desta tese.

A níveis acadêmicos, gostaria de agradecer meu orientador Professor Ricardo Gonçalves por me ter permitido realizar este trabalho com ele e quero deixar um agradecimento especial ao meu co-orientador Ruben Costa, Paulo Figueira e ao Guilherme Guerreiro, por todo o tempo despendido e ajuda na realização da minha tese, foram impecáveis.

Quero também deixar um agradecimento especial aos meus amigos, não referindo nomes, porque para minha sorte teria de fazer muitas referências, mas apenas referi palavras que eles se irão enquadrar, Alverca, escuteiros e faculdade. Foi através dos meus amigos que grande parte deste caminho foi conseguido, foram eles que me ouviram, que me aturaram e descontraíram em todos os momentos de stress ao longo destes anos, um verdadeiro obrigado.

E por último, não podia deixar de agradecer ao meu pilar, a minha família, em particular à minha mãe, ao meu pai, meu irmão e meus avós, um obrigado pelo esforço financeiro que me permitiu financiar o meu curso e aventuras ao longo dele, sem vocês nada disto seria possível. Obrigado por estarem sempre ao meu lado nos meus 23 anos de existência, sempre com enorme amor e carinho, mas um verdadeiro e agradecido obrigado por me tornarem na pessoa que sou hoje.





## RESUMO

---

A criação de sistemas de transportes inteligentes com vista a melhorar a segurança e circulação de veículos nas estradas, têm sido alvo da pesquisa de diversos grupos nas últimas décadas. A combinação de várias ferramentas de processamento e análise de grandes volumes de dados têm vindo cada vez mais a vincar o seu papel na criação deste tipo de sistemas, demonstrando resultados positivos na aplicação destes nos sistemas de transportes. A combinação de grandes volumes de dados com a deteção de eventos complexos é um caminho inovador e desafiante na gestão de tráfego rodoviário, pois a utilidade deste tipo de sistemas depende da disponibilidade e qualidade de dados, dados estes adquiridos através de sensores, sujeitos a falhas.

O objetivo desta dissertação passa pela categorização e deteção de eventos de tráfego (acidentes, trânsito lento, etc), tendo por base um repositório de dados de tráfego rodoviário. Dados esses que são recolhidos por sensores de fluxo rodoviário colocados ao longo de estradas principais e autoestradas. Ao longo desta dissertação será desenvolvido um protótipo, utilizando técnicas de processamento complexo de eventos (CEP), que visa responder ao anteriormente descrito.

Todos os dados utilizados para validação do presente trabalho têm por base a rede viária eslovena, sendo este trabalho desenvolvido no contexto de um projeto de pesquisa europeu.

**Palavras-chave:** Processamento de eventos complexos (CEP), ferramentas de CEP (WSO2), sistema inteligente de transportes, base de dados e comportamento nas autoestradas.

---



## ABSTRACT

---

The development of intelligent transport systems to improve the safety and flow of vehicles in roads has been the focus of several research groups in recent years. The combination of several tools for processing and analysis of big data volumes have played an increasingly role in the creation of such systems, showing positive results in their application to transport systems. The combination of big data volumes with the detection of complex events is an innovative and challenging approach in road traffic management, since the efficiency of this type of systems is dependent on the availability and quality of data acquired through sensors normally exposed to failures.

The main objective of this study is to categorize and detect complex events based on a repository of road traffic data. This data is collected by road flow sensors placed along major roads and motorways. As an output of this study it will be developed a prototype, using complex event processing techniques (CEP), which aims to respond to the previously described.

All the data used to validate the present study is based on the Slovenian road network. This work has been carried out in the context of the OPTIMUM Project, funded by the H2020 European Research Framework Program.

**Keywords:** Complex event processing (CEP), CEP engines (WSO2), intelligent transport systems (ITS), data base, highway profiles

---



# ÍNDICE

<b>Lista de Figuras</b>	<b>xv</b>
<b>Lista de Tabelas</b>	<b>xix</b>
<b>Siglas</b>	<b>xxi</b>
<b>1 Introdução</b>	<b>1</b>
1.1 ITS - <i>Intelligent Transportation Systems</i>	2
1.1.1 Aplicações ITS	2
1.2 Problema do Modelo Atual Rodoviário	4
1.3 Motivação	5
1.4 Enquadramento do Problema	6
1.5 Abordagem	7
1.5.1 Contribuições	8
1.6 Estruturação do Documento	9
<b>2 Estado da Arte</b>	<b>11</b>
2.1 Fundamentação Teórica	11
2.1.1 Definição de CEP	12
2.1.2 Características das ferramentas CEP	12
2.1.3 Vantagens e Desvantagens do CEP	13
2.1.4 Técnicas de <i>Big Data</i>	13
2.2 Trabalho Relacionado	14
2.2.1 Área Financeira	14
2.2.2 Área Agrícola	15
2.2.3 Área de Transportes	16
2.2.4 Área Desportiva	17
2.2.5 Avaliação de ferramentas CEP	18
2.3 CRISP-DM	19
<b>3 Fontes de Dados</b>	<b>21</b>
3.1 Compreensão de Dados	22
3.1.1 Organização Estrutural dos Dados	22

3.1.2	Metadados . . . . .	23
3.1.3	Cobertura espaciotemporal . . . . .	23
3.1.4	Qualidade dos Dados . . . . .	27
3.2	Seleção de Dados . . . . .	29
3.3	Exploração dos Dados . . . . .	32
3.3.1	Relação ocorrência de eventos com a distribuição populacional . .	32
3.3.2	Valores Anómalos . . . . .	33
<b>4</b>	<b>Perfil das Autoestradas</b>	<b>37</b>
4.1	Comportamento Diário em Dias Úteis . . . . .	37
4.2	Comportamento Diário em Fins de Semana e Feriados . . . . .	41
4.3	Comparação comportamental diária de uma autoestrada entre dias úteis, feriados e fins de semana . . . . .	45
4.4	Comportamento Semanal . . . . .	46
4.5	Comportamento Sazonal . . . . .	49
<b>5</b>	<b>Deteção de Eventos em Autoestradas</b>	<b>53</b>
5.1	Comportamento de Eventos em Autoestradas . . . . .	53
5.1.1	Caraterização de eventos . . . . .	54
5.2	Modelação . . . . .	55
5.2.1	Recetores de eventos . . . . .	55
5.2.2	<i>Stream</i> de Eventos . . . . .	56
5.2.3	Processador de Eventos . . . . .	56
5.2.4	Divulgador de Eventos . . . . .	57
5.2.5	Visão Geral . . . . .	57
5.3	Aplicação de Regras . . . . .	59
5.4	Parametrização . . . . .	63
<b>6</b>	<b>Validação e Resultados</b>	<b>65</b>
6.1	Resultados . . . . .	65
6.2	Validação . . . . .	69
<b>7</b>	<b>Conclusões</b>	<b>75</b>
	<b>Bibliografia</b>	<b>77</b>
<b>I</b>	<b>Anexo 1 - Tabelas de Metadados</b>	<b>81</b>
<b>II</b>	<b>Annex 2 - Código do bloco processador de eventos</b>	<b>85</b>

## LISTA DE FIGURAS

1.1	Sistema de Painéis de Mensagens Variáveis[5]. . . . .	3
1.2	Modelo de comunicação de veículos[10]. . . . .	4
2.1	Painel de visualização de uma aplicação CEP de gestão de rotas . . . . .	17
2.2	Painel de visualização de informações em tempo real sobre um jogo de futebol, tendo por base uma aplicação CEP . . . . .	18
2.3	Diagrama do modelo de processo CRISP-DM . . . . .	19
3.1	Comparação estrutural entre dados de 4 de janeiro de 2016 (esquerda) e 04 de janeiro de 2017 (direita) . . . . .	22
3.2	Parte da estrutura de dados de eventos rodoviários . . . . .	23
3.3	Falhas na cobertura temporal de 15 sensores . . . . .	24
3.4	Disponibilidade mensal para cinco meses de 2016, tendo em consideração 9 sensores . . . . .	25
3.5	Disponibilidade diária do sensor 174 no mês de janeiro de 2017 . . . . .	26
3.6	Projeção espacial da localização dos sensores . . . . .	26
3.7	Projeção espacial da localização dos eventos . . . . .	27
3.8	Visualização de duplicados nos registos relativos a leituras de sensores . . . . .	28
3.9	Formatação dos dados informativos de cada sensor . . . . .	31
3.10	Formatação dos dados referentes às leituras dos sensores . . . . .	31
3.11	Densidade populacional em território esloveno [50] . . . . .	32
3.12	Projeção espacial da localização dos eventos . . . . .	33
3.13	Diagrama de extremos e quartis horário da ocupação de faixa para um sensor aleatório, destacando a vermelho possíveis valores anómalos no sensor . . . . .	34
3.14	Velocidade média e ocupação da via na ocorrência de um possível valor anómalo . . . . .	35
3.15	Diagrama de extremos e quartis horário da velocidade dos veículos para um sensor aleatório, destacando a vermelho possíveis valores anómalos no sensor . . . . .	35
3.16	Ocupação da via e velocidade média na ocorrência de um possível valor anómalo . . . . .	36
4.1	Fluxo rodoviário previsto para um dia útil . . . . .	38
4.2	Fluxo rodoviário diário num dia útil, próximo de uma zona com grandes dimensões populacionais (sensor 832) . . . . .	38

---

4.3	Fluxo rodoviário numa zona habitacional de menor população (sensor 258) .	39
4.4	Fluxo rodoviário entre duas zonas habitacionais com elevada população (sensor 836) . . . . .	40
4.5	Comparação dos fluxos rodoviários em diferentes localizações num dia útil (sensores: vermelho-174, azul-836 e verde-258) . . . . .	41
4.6	Fluxo rodoviário esperado ao fim de semana . . . . .	42
4.7	Fluxo rodoviário diário num fim de semana (sábado), perto de uma zona habitacional de elevada população (sensor 174) . . . . .	42
4.8	Comparação dos fluxos rodoviários em diferentes localizações, durante um fim de semana (sábado)(sensores: vermelho-174, azul-836 e verde-258) . . . .	43
4.9	Fluxo rodoviário num feriado, próximo de uma zona habitacional de elevada população (sensor 174) . . . . .	44
4.10	Comparação dos fluxos rodoviários em diferentes localizações num feriado (sensores: vermelho-174, azul-836 e verde-258) . . . . .	44
4.11	Comparação do fluxo rodoviário semanal, para um sensor perto de um grande centro populacional (sensor 174) . . . . .	45
4.12	Comparação do fluxo rodoviário semanal, para um sensor próximo de uma zona populacional menor (sensor 258) . . . . .	46
4.13	Fluxo rodoviário ao longo de uma semana, próximo de uma grande zona populacional (sensor 174) . . . . .	47
4.14	Fluxo rodoviário ao longo de uma semana, entre duas zonas habitacionais com elevada população (sensor 836) . . . . .	48
4.15	Fluxo rodoviário durante todas as segundas feiras nos meses janeiro e fevereiro (sensor 199) . . . . .	48
4.16	Comparação do fluxo de rodoviário entre três dias em abril e agosto, para um sensor localizado próximo de um grande centro populacional (sensor 199) .	49
4.17	Comparação do fluxo de rodoviário entre três dias em abril e agosto, para um sensor localizado próximo de um pequeno centro populacional (sensor 258)	50
4.18	Comparação do fluxo de rodoviário entre três dias em abril e agosto, para um sensor próximo a uma zona litoral (sensor 803) . . . . .	51
5.1	Evento anómalo (trânsito) para o sensor 174 . . . . .	54
5.2	Evento anómalo (acidente) para o sensor 174 . . . . .	55
5.3	Diagrama de blocos geral da ferramenta WSO2 CEP [51] . . . . .	56
5.4	Diagrama de blocos da aplicação de deteção pró-ativa de eventos . . . . .	57
5.5	Receção de dados no núcleo do bloco processador de eventos (parte 1) . . . . .	59
5.6	Receção de dados no núcleo do bloco processador de eventos (parte 2) . . . . .	60
5.7	Receção de dados no núcleo do bloco processador de eventos . . . . .	60
5.8	Operação lógica sobre os dados no núcleo do bloco processador de eventos .	61
5.9	Operação lógica para identificação de inicio e termino eventos sobre os dados no núcleo do bloco processador de eventos . . . . .	61



---

5.10	Código exemplo para identificação de início de evento . . . . .	62
5.11	Código exemplo para identificação de término de evento . . . . .	63
5.12	Verificação de repetições sobre os eventos no núcleo do bloco processador de eventos . . . . .	63
5.13	Exportação dos eventos para fora do núcleo do bloco processador de eventos . . . . .	63
6.1	Visualização de um início e fim de evento, detetado pela ferramenta desenvolvida . . . . .	65
6.2	Visualização de um início e fim de evento, entre outro evento . . . . .	66
6.3	Visualização de um início e fim de eventos, detetados pela ferramenta desenvolvida . . . . .	67
6.4	Ocorrência de início e fim de eventos na aplicação para 5 sensores próximos de um grande centro populacional . . . . .	68
6.5	Ocorrência de início e fim de eventos na aplicação para 5 sensores entre dois grandes centros populacionais . . . . .	68
6.6	Ocorrência de início e fim de eventos na aplicação para 5 sensores próximos de um pequeno centro populacional . . . . .	69
6.7	Comparação da ocorrência de eventos em 5 sensores próximos de um grande centro populacional, com uma taxa de variação da velocidade média de 20% . . . . .	70
6.8	Comparação da ocorrência de eventos em 5 sensores entre dois grandes centros populacionais, com uma taxa de variação da velocidade média de 20% . . . . .	70
6.9	Comparação da ocorrência de eventos em 5 sensores próximos de um pequeno centro populacional, com uma taxa de variação da velocidade média de 20% . . . . .	71
6.10	Número total de eventos por categoria para 15 sensores, com uma taxa de variação da velocidade média de 20% . . . . .	71



## LISTA DE TABELAS

2.1	Caraterização de ferramentas CEP . . . . .	18
5.1	Descrição da formatação dos dados recebidos no bloco "Receiver_Average_Reading_Points" . . . . .	58
5.2	Descrição da formatação dos dados recebidos no bloco "ReceiverCounterValues" . . . . .	58
6.1	Precisão, sensibilidade e F-Measure para eventos em 15 sensores com variações de velocidade superiores ou iguais 20% . . . . .	72
6.2	Tabela de contadores de eventos para 5 sensores com variação nos parâmetros das regras da aplicação concebida . . . . .	73
6.3	Tabela de precisão, sensibilidade e <i>F-measure</i> para eventos em 5 sensores com variação nos parâmetros das regras da aplicação concebida . . . . .	74
I.1	Descrição dos campos da base de dados de eventos rodoviários . . . . .	81
I.2	Descrição dos campos da base de dados de valores comportamentais de veículos, com cobertura temporal de dezembro de 2016 a maio de 2017 . . . . .	82
I.3	Descrição dos campos da base de dados de valores comportamentais de veículos, com cobertura temporal de janeiro de 2016 a novembro de 2016 . . . . .	83



## SIGLAS

ACM	Association for Computing Machinery's.
CEP	Processamento de Eventos Complexos.
CRISP-DM	Cross Industry Standard Process for Data Mining.
DEBS	Distributed Event-Based Systems.
DSMS	Data Stream Management System.
GPS	Global Positioning System.
ITS	Intelligent Transportation Systems.
LED	Light-Emitting Diode.
RFID	Radio-Frequency Identification.
V2I	Vehicle-to-Infrastructure.
V2V	Vehicle-to-Vehicle.



## INTRODUÇÃO

Nas últimas cinco décadas, tem-se assistido a um grande aumento do número de veículos a circular nas estradas, resultando este aumento na saturação das infraestruturas rodoviárias[1]. Esta situação afeta em particular o quotidiano das pessoas, especificamente nas áreas urbanas, onde as pessoas necessitam de se movimentar mais rapidamente entre locais. Desse modo, deste problema advém um conjunto de outros problemas, desde o aumento do congestionamento do tráfego, seguido, de um aumento no número de acidentes, atrasos nos transportes e num aumento da poluição por parte dos veículos, contribuindo todos estes fatores para a degradação da qualidade de vida das pessoas.

Até hoje, muitas foram as propostas concebidas para tentar combater estes problemas, desde a implementação de sistemas de segurança em veículos, como o *airbag*, até à construção de novas e melhoradas estradas e autoestradas, para tentar resolver os problemas de congestionamento. Contudo e principalmente esta última solução referida, apresentou uma rotura, pois a construção de novas estradas é algo dispendioso e necessita de novos espaços físicos dentro das cidades, o que na maioria das grandes cidades já não existe. Por outro lado, o sistema rodoviário e todos os benefícios, que este oferece, são essenciais tanto para a sociedade atual, como para o desenvolvimento da economia, sendo desse modo necessário encontrar uma solução aos problemas apresentados[1].

É então que se percebeu que a solução não passaria pela conceção de novas e melhoradas infraestruturas, mas sim pela utilização eficiente e segura das infraestruturas já existentes, levando a comunidade científica a virar a sua atenção para a área do [Intelligent Transportation Systems \(ITS\)](#) [2].

## 1.1 ITS - *Intelligent Transportation Systems*

Os *Intelligent Transportation Systems* são nos dias de hoje um fenómeno global, tornando-se vital na sociedade atual [2]. O ITS é uma área criada na base do desenvolvimento de novas de tecnologias, desde a eletrónica, ao processamento de informação, às comunicações e ao controlo. Este conjuga um conjunto de tecnologias com a interoperabilidade de diferentes campos do sistema rodoviário, desde gestão de transportes, controlo, infraestruturas, segurança, etc[2]. Produzindo desta maneira, soluções para a cooperação entre veículos (V2V) com veículos e veículos com infraestruturas (V2I).

Tendo objetivos muito específicos, o ITS pretende de uma forma geral reduzir os níveis de poluição, diminuindo os níveis de emissões de carbono, reduzir o tempo de viagem, aumentar a segurança rodoviária, diminuir a taxa de acidentes rodoviários, diminuir os níveis de congestionamento nas vias e aumentar a satisfação dos condutores. [1].

Como qualquer sistema, esta também apresenta as suas controvérsias, sendo de modo geral apontadas como desvantagens deste novo sistema, a vulnerabilidade que este mesmo possa apresentar perante ataques informáticos, deixando o sistema de controlo vulnerável e o custo na implementação inicial de equipamentos ITS.

### 1.1.1 Aplicações ITS

Atualmente, já existem algumas aplicações/sistemas que trabalham o conceito de ITS, sendo os Estados Unidos da América um dos países que mais tem evoluído e avançado no desenvolvimento destas aplicações[3].

Os sistemas ITS podem-se caracterizar consoante a sua área de aplicação, assumindo estes sistemas duas grandes categorias, uma denominada de estrutura inteligente e a outra de veículos inteligentes[3], tendo ambas como objetivo a resolução dos problemas que antevem do ITS, sendo apenas distintas no local onde são aplicadas, sendo uma referente a infraestruturas e a outra referente a veículos, respetivamente.

#### 1.1.1.1 Painéis de Mensagens Variáveis

Este tipo de aplicações consiste num sistema constituído por painéis **Light-Emitting Diode (LED)**, que informam os utilizadores sobre o estado de circulação da via em questão. Este sistema já tem alguns anos e é muito comum nas autoestradas ou vias rápidas portuguesas, onde os operadores de via deixam mensagens de prevenção ou de informação aos condutores[4].





Figura 1.1: Sistema de Painéis de Mensagens Variáveis[5].

#### 1.1.1.2 Semáforos Inteligentes

Os semáforos inteligentes são uma outra aplicação que veio em muito ajudar a resolver problemas de elevado tráfego nas cidades. Esta consiste num sistema de coordenação dos sinais de trânsito luminosos, ajustando o comprimento de fase do sinal com base nas condições de tráfego prevalentes. Este sistema permite ter mudanças na duração de um sinal luminoso ao longo do dia, tendo como fator decisivo o tráfego existente[6].

#### 1.1.1.3 Sistema de Notificação Automático

O sistema de notificação automático é um sistema aplicado em veículos que, com a ajuda de sensores, tecnologia [Global Positioning System \(GPS\)](#) e sistemas de comunicações sem fios instalados num veículo, notifica os meios de assistência da ocorrência de um acidente, podendo se encontrar várias versões deste sistema em automóveis da marca BMW, Toyota ou mesmo Ford. Numa versão ainda mais moderna e desenvolvida, é possível, consoante o tipo de acidente, que este sistema informe de imediato os serviços de emergência, da ocorrência de um possível acidente grave[7].

Como se pode observar pelos exemplos de aplicações descritos anteriormente, estas tecnologias ITS oferecem uma ampla variedade de benefícios tanto para o transporte pessoal, como para o transporte público. Além de proporcionarem melhorias na segurança, através de sistemas automáticos e pró-ativos em caso de acidentes ou emergências, permite ainda uma gestão de rotas prevendo uma redução do número de acidentes, tanto entre veículos como entre veículos e peões. É também através da gestão de rotas, que este tipo de tecnologias assegura um aumento da eficiência dos tempos de viagem e diminuição dos custos económicos associados a estas. Com a utilização destas tecnologias, e com a evolução e integração destas em veículos autónomos, prevê-se ainda minimizar o uso

de combustíveis e outros recursos, diminuindo significativamente os agentes poluidores, assegurando uma maior satisfação das pessoas sujeitas a este conjunto de problemas, obtendo-se uma melhor qualidade de vida em sociedade[8].

Hoje em dia, para além dos sistemas já implementados, muitos outros estão a ser desenvolvidos, como por exemplo, atualmente através de protocolos **Vehicle-to-Vehicle (V2V)** e **Vehicle-to-Infrastructure (V2I)**, preparam-se aplicações ITS que permitem a implementação de veículos conectados sem fios e com dispositivos à margem da estrada compartilhando dados de modo a evitar colisões e fornecer muitos outros benefícios para evitar a ocorrência de acidentes[9].



Figura 1.2: Modelo de comunicação de veículos[10].

## 1.2 Problema do Modelo Atual Rodoviário

Desde há muitos anos, que o transporte rodoviário é um dos maiores meios de transporte de pessoas e de mercadorias a nível nacional. Com o desenrolar dos tempos e com a criação de autoestradas, esta área tem sofrido um positivo investimento por parte do estado português e empresas privadas, levando a conceção de aproximadamente 3000 km de autoestrada, sendo que 84% destas sujeitas à cobrança de portagens por parte dos utilizadores [11].

Apesar da criação desta rede nacional de autoestradas e da reconhecida qualidade da mesma, a sua utilização tem estado muito aquém das expectativas. Independentemente da já referida qualidade da rede, os condutores optam frequentemente por vias gratuitas adjacentes àquelas que são sujeitas a taxas, levando a decisão do utilizador ser tomada praticamente em função do preço, excluindo fatores como a segurança ou o tempo da viagem.

Desse modo dá-se então a sobrelotação das vias secundárias ou adjacentes, que não apresentam as condições necessárias para o elevado fluxo de carros que se faz notar

nestas mesmas vias, contribuindo desse modo para um maior número de paragens e por consequente aumento generalizado dos tempos de viagens por parte dos utilizadores. Para além dos problemas referidos até agora, a sobrelotação de vias secundárias contribui para um aumento dos custos de manutenção destas e ainda, para um aumento do número de acidentes, tanto entre automóveis como acidentes com peões.

A manifestação destes problemas afetam em muito todas as grandes cidades mundiais, não sendo este um problema apenas detetado em Portugal, mas sendo observado em muitos outros países e tendo sido já mencionado por muitas das grandes organizações mundiais[2].

Com os problemas apresentados, concluiu-se que os custos associados ao melhoramento de todas as vias secundárias e até mesmo manutenção das mesmas, acarreta enormes custos, passando a solução do sistema de mobilidade rodoviário português pela melhor e eficiente utilização das autoestradas nacionais.

Para que seja possível uma melhor e eficiente utilização das autoestradas, tem de existir mecanismos capazes de analisar e monitorizar dados referentes as mesmas, como os níveis de tráfego (lotação da via, velocidade média, ...). Combinando estes dados com o tempo atual e condições climatéricas, permite-se assim uma previsão sobre possíveis ocorrências nas autoestradas.

Atualmente existem poucas formas de deteção de eventos rodoviários (acidentes, trânsito, obstruções na via, condicionamentos na via, condições climatéricas) por parte das empresas responsáveis pela conceção das autoestradas. Hoje em dia, a deteção de eventos de trânsito em "tempo-real" é efetuada nos centros de operação de tráfego, de forma ainda manual, exigindo a necessidade de existirem funcionários a analisarem continuamente várias câmeras de trânsito, manifestando-se num desperdício de recursos humanos para as concessionárias das autoestradas.

Contudo por vezes, alguns eventos de rodoviários poderão ser detetados de forma indireta, sempre que sejam comunicados às autoridades. Este tipo de deteção indireta ocorre normalmente em acidentes em autoestradas onde seja necessário a presença de meios de autoridade ou de socorro.

Como se pode desde já imaginar, estes métodos não são muito fiáveis e acarretam componente humana, nunca existindo em nenhum destes um método de deteção automática ou pró-ativa para a ocorrência deste tipo de eventos, estando-se sempre dependente de ações humanas.

### **1.3 Motivação**

Dos problemas apresentados anteriormente, parte da solução passa por modificar o sistema rodoviário atual, implementado um sistema de portajamento dinâmico de autoestradas, enquadrado num projeto Europeu de I&D H2020, denominado OPTIMUM[12].

Sendo esta abordagem um projeto pioneiro em Portugal [11], espera-se que este permita um uso eficiente das autoestradas, promovendo a rentabilização de todo o investimento realizado em novas infraestruturas rodoviárias e, ao mesmo tempo espera-se que garanta uma melhor qualidade de vida dos utentes, desde o alcance de maiores níveis de segurança, à redução do tempo de viagem, colmatando com uma redução do fluxo de tráfego em vias secundárias, levando por sua vez, a uma redução dos problemas causados pelo uso excessivo destas.

Olhando para a solução de portajamento dinâmico, que consiste na variação do valor cobrado em autoestradas através de sistemas de portagens, onde esta variação ocorre consoante fatores como a lotação das autoestradas ou o hora a que circula. Focando esta solução apresentada, o problema deste tipo de sistema é que pressupõem uma rápida análise e monitorização dos níveis de tráfego, algo que ainda não é possível através dos métodos de deteção implementados atualmente nas autoestradas. Todos os métodos são não proativos, ou seja, não executam de maneira automática a aquisição destas informações. Passando a solução deste problema pela implementação de uma ferramenta/aplicação capaz de executar estes requisitos de forma automática, ou seja, uma aplicação capaz de detetar, analisar e monitorizar fatores influenciadores do tráfego rodoviário, isto é, uma aplicação capaz de deter e identificar os diferentes eventos anómalos no sistema rodoviário.

Desse modo, assume-se como principal objetivo desta dissertação, a **criação de uma ferramenta de deteção automática de eventos anómalos em autoestradas**, tentando assim criar uma nova abordagem sobre a deteção de eventos anómalos em autoestradas, que até então apenas se realiza através de sistemas dependentes da componente humana (ação humana), desenvolvendo um novo sistema que realize a deteção dos eventos, sem a necessidade da componente humana, ou seja, automaticamente.

A presente dissertação, enquadrada na base do projeto europeu OPTIMUM, usará os dados disponibilizados e já trabalhados ao longo do mesmo, como forma de validação da presente dissertação.

Esta dissertação não tenciona solucionar todos os problemas associados ao congestionamento de tráfego, mas sim reduzir os problemas associados a este, como a sobrelotação das vias secundárias que, por conseguinte, se espera uma redução do número de acidentes sobre estas condições.

Como já foi referido anteriormente, a importância deste tema tem sido mencionado a nível mundial, sendo possível que este projeto seja o pioneiro na idealização de novos projetos, nesta área a nível mundial.

### 1.4 Enquadramento do Problema

Com o objetivo de desenvolver uma aplicação capaz de detetar eventos anómalos em autoestradas de forma automatizada, são alguns os problemas que se levantam na hora de desenvolver uma aplicação deste tipo.

Em primeiro lugar, é necessário recorrer a ferramentas capazes de processar enormes quantidades de dados e forma rápida, tendo de ser executado um estudo prévio e uma escolha assertiva na decisão da ferramenta por detrás da base deste tipo de aplicação.

Contudo, este não é o único problema que se levanta. Quando se pretende detetar anomalias ao normal funcionamento das autoestradas é necessário recorrer a estruturas, neste caso sensores, capazes de coletar dados sobre o comportamento dos utilizadores das autoestradas, desde velocidades médias, a ocupações na via, estruturas estas que requerem investimentos e manutenções por parte das operadores das autoestradas.

Outro critério a ter em consideração é o número de dados existentes, após implementação das estruturas (sensores), que permita desenvolver um estudo prévio do comportamento dos utilizadores nas autoestradas. Quando o objetivo é a deteção anómala de eventos, este requer o maior número possível de dados para criação de padrões comportamentais por parte dos utilizadores, sendo que a não existência de uma grande base de dados abrangendo diversos espaços temporais, pode induzir em erro à criação de padrões comportamentais que à posteriori serão usados para deteção dos eventos anómalos.

Um possível problema poderá passar também pela qualidade dos dados adquiridos. Para este tipo de aplicações é necessário um elevado grau de qualidade dos dados, para que as análises e processamentos destes mesmos dados não induzam em erro os resultados adquiridos. A má qualidade dos dados não só afetará os resultados finais, como todos os estudos previamente realizados para aquisição de padrões comportamentais em autoestradas.

Por fim, outro critério em ter em atenção no desenvolvimento deste tipo de aplicações, que recorrem a espaços temporais muito abrangentes é os períodos de sazonalidade. A ocorrência de períodos sazonais ao longo de um ano é algo comum, contudo se os dados utilizados na criação de padrões comportamentais rodoviários em autoestradas apenas utilizarem estes períodos sazonais, os estudos deles realizados poderão não corresponder aos comportamentos noutras alturas do ano.

## 1.5 Abordagem

Após serem abordados problemas a ter em atenção no desenvolvimento da presente dissertação, serão seguidas várias etapas de modo a eliminar os problemas apresentados anteriormente e desenvolver uma aplicação capaz de detetar e identificar de forma automática eventos em autoestradas.

Numa fase inicial é essencial avaliar os conceitos e traçar objetivos, segundo a visão empresarial atual, ou seja, segundo a presente dissertação, realizar uma abordagem aos temas ITS, *big data*, etc., com vista a enquadrar e aprofundar conhecimento sobre o projeto idealizado, traçando à posteriori um plano de objetivos para realização do mesmo.

Quando traçados os objetivos, que no presente caso consiste no desenvolvimento de uma aplicação capaz de categorizar e detetar eventos anómalos de forma automática no sistema rodoviário, é então necessário recorrer a uma colheita e perceção dos dados

para estudos e processamentos futuros. Nesta segunda fase, pretende-se colecionar um conjunto de dados, neste caso, dados relativos a comportamentos rodoviários em autoestradas, desde a velocidade dos seus utilizadores, à ocupação das vias, ou até mesmo recorrer a uma base de dados de ocorrências rodoviárias, de modo, a que estes conjuntos de dados possam ser interpretados para perceber se faz sentido o enquadramento futuro dos mesmos.

A terceira etapa realizará uma preparação e seleção dos dados anteriormente descritos, assim sendo, serão realizadas operações sobre os conjuntos de dados de modo a selecionar os dados, tendo em conta padrões de qualidade dos mesmos. É nesta etapa que serão executadas todas as transformações e operações necessárias sobre os dados.

Após os dados serem selecionados serão realizados estudos sobre os mesmos, criando padrões comportamentais rodoviários, ou seja, para o presente caso serão gerados padrões comportamentais das autoestradas, perfis das autoestradas, para que mais tarde, existam termos de comparação, para a deteção de eventos anómalos rodoviário. Para a deteção de eventos anómalos rodoviários, tem de existir previamente um conhecimento sobre o normal comportamento rodoviário, para que se permita uma comparação e posterior identificação deste tipo de eventos.

Quando realizado o estudo, a próxima abordagem a ser executada é o desenvolvimento de uma aplicação capaz de realizar os objetivos previamente traços de forma automática.

Por fim, é necessário recorrer a testes e avaliações sobre a aplicação desenvolvida de forma a identificar os parâmetros de otimização das mesmas, que serão utilizados mais tarde para implementação real da aplicação.

### 1.5.1 Contribuições

**Com o objetivo geral traçado, desenvolver uma aplicação capaz de detetar e categorizar automaticamente eventos anómalos no sistema rodoviário, eventos como a ocorrência de acidentes, a obstruções na via, condicionamento na via, elevado fluxo de tráfego, entre outros, pretende-se desse modo contribuir com:**

- Preparação e seleção de dados, tendo por base uma análise da qualidade e disponibilidade dos diferentes tipos de dados que afetam os níveis de utilização de autoestradas (dados de sensores contadores de veículos).
- Caracterização de diferentes tipos de perfis de utilização de autoestradas.
- Criação de padrões comportamentais em autoestradas.
- Análise e formalização de regras que caracterizam vários tipos de eventos de tráfego.
- Apresentação de um modelo para deteção de eventos anómalos de trânsito usando ferramentas CEP.



- Implementação e otimização de um protótipo CEP de detecção de eventos anómalos rodoviários.
- Escrita de um artigo científico[13].

Foi ainda incentivado a realizar uma pergunta de investigação, à qual concebi a seguinte *”Como detetar e caracterizar de forma automática eventos de tráfego rodoviários, considerados anómalos?(acidentes, trânsito lento, etc.)”*

Passando a hipótese da mesma, pela detecção e caracterização automática de eventos anómalos de tráfego rodoviário, tendo por base a implementação de uma abordagem CEP, capaz de processar fluxos de dados rodoviários provenientes de sensores contadores de veículos.

## 1.6 Estruturação do Documento

A presente tese é constituída por sete capítulos. Neste primeiro capítulo realizou-se uma contextualização dos temas abrangentes na tese, desde os sistemas de transportes inteligentes, aos problemas do modelo rodoviário, introduzindo posteriormente abordagem para tentativa de resolução dos problemas identificados, que passa pela criação de uma aplicação e posterior motivação pela qual é abordada esta solução.

No capítulo 2 introduzir-se-á os fundamentos teóricos relacionados com o tema geral da dissertação, desde o *big data*, aos processamento de eventos complexos, base da presente tese. Neste capítulo serão ainda abordados outros trabalhos relacionados com a utilização de processamento de eventos complexos em diversas áreas e por fim, será explicado o modelo assente na abordagem ao presente trabalho.

Seguidamente, o capítulo 3 apresentará uma descrição sobre os dados utilizados para desenvolvimentos de padrões e os dados utilizados na aplicação a ser criada. Será realizada uma avaliação sobre os dados de modo a perceber se estes se encontram aptos para análises futuras. No capítulo 4, descrever-se à um perfil comportamental sobre autoestradas, tendo por base os dados já estudados e selecionados no capítulo anterior.

O capítulo 5 abordará toda a conceção e metodologia utilizada na elaboração da aplicação proposta no capítulo 1, onde de forma pormenorizada serão descritos todos os processos e decisões tomadas na conceção da aplicação.

Em seguida, no capítulo 6 serão apresentados os resultados adquiridos da aplicação criada, sendo ainda realizada um posterior validação e otimização dos parâmetros e métodos utilizados na aplicação.

Por fim, no capítulo 7 encontrar-se as principais conclusões extraídas no decurso das análises efetuadas na presente dissertação.





## ESTADO DA ARTE

No presente capítulo irá ser realizado um introdução teórica aos temas de *big data* e **Processamento de Eventos Complexos (CEP)**, apresentando maior foco neste último tema.

Serão ainda explorados outros trabalhos de diversas áreas relacionados com processamento de eventos complexos e por último, será realizado uma avaliação de ferramentas de CEP e enquadrado o modelo de trabalho escolhido com os objetivos da presente tese.

### 2.1 Fundamentação Teórica

Com a enorme expansão tecnológica denotada nas últimas décadas, tem-se assistido a uma enorme avalanche de dados. Com a utilização de sensores, transações bancárias, chamadas telefônicas, redes sociais, estima-se que sejam gerados diariamente cerca de 2.5 exabytes de dados[14], dados estes que após serem processados e analisados, permitem retirar significado dos mesmos, com a criação de padrões, tendências, etc. Podendo à posteriori ser utilizados por empresas no apoio a decisões, retirando valor sobre os dados.

Estando na "presença" desta enorme quantidade de dados, entramos agora no campo *big data*, onde através de várias técnicas e tecnologias é possível analisar, processar, tratar e monitorizar grandes quantidades de dados.

*Big data* é um conceito recente que descreve dados complexos e variáveis, com grande quantidade de volume e grande rapidez, que requer técnicas e tecnologias avançadas para permitir captura, armazenamento, distribuição, gestão e análise da informação adquirida pelos dados[15].

Com esta tendência para a criação de dados, surge a necessidade de se criar ferramentas capazes de executar o processamento destes dados. Foi então que em 1995, David Luckham, professor da Universidade de Stanford, trabalhou no projeto Rapide que viria a ser a primeira ferramenta de análise e simulação de sistemas associados ao conceito de

processamento de eventos complexos(CEP), conceito englobado no termo *big data*. Após este avanço muitas foram as pesquisas que se seguiram neste ramo, denotando ainda os projetos Infosphere da Caltech e o Apama da Universidade de Cambridge[16].

### 2.1.1 Definição de CEP

O conceito de CEP é algo muito recente, demonstrando uma enorme expansão nos últimos anos e podendo ser aplicado a inúmeras áreas, tem sido redigidas várias definições sobre este conceito, existindo uma definição que, na minha opinião, penso ser a que melhor descreve o CEP. Esta foi escrita por Ivy Schmerken, jornalista financeira, onde esta assume o CEP como “um método de processamentos de eventos que combina dados de diferentes fontes para inferir eventos ou padrões que surgem em circunstâncias mais complicadas”[17].

Ao ser analisada a definição anteriormente descrita, é de salientar a introdução de um novo conceito, eventos. Um evento é uma ocorrência/acontecimento dentro de um sistema ou domínio. Sendo algo que aconteceu, ou está a acontecer num certo domínio[18]. Na sua forma mais simples pode-se considerar um evento como a ocorrência de chuva ou um telemóvel a tocar, ou na sua forma mais complexa como a ocorrência de um evento rodoviário, ou até uma transação multibanco.

Os eventos podem ainda ser classificados em dois tipos, consoante a sua ocorrência, eventos esperados ou eventos anómalos, sendo eventos esperados todos aqueles que ocorrem inúmeras vezes ao longo de um intervalo de tempo e eventos anómalos são eventos que não são muito comuns no nosso quotidiano. Contudo é de realçar que a classificação de um evento quanto à sua ocorrência tem em conta o quotidiano envolvente e a localização geográfica, podendo o mesmo evento para duas pessoas ser considerado com diferentes classificações[18].

Executando o processamento em simultâneo de vários eventos, originados por diferentes fontes, é então criado o conceito de processamento de eventos complexo. O conceito de complexo é apenas aplicado aos eventos e não ao processamento, este é um processamento simples aplicado a eventos que por vezes podem ser considerados complexos, na sua forma de se manifestar[19].

### 2.1.2 Caraterísticas das ferramentas CEP

Tal como qualquer outro tipo de ferramentas, as ferramentas CEP também apresentam as suas caraterísticas que lhes permitem diferenciar de outras ferramentas, existindo três caraterísticas fundamentais a qualquer aplicação CEP. Estas são o elevado desempenho, a baixa latência e a escalabilidade.

**Elevado desempenho:** com o quase incontável número de dados que são sujeitas este tipo de aplicações/sistemas, é fundamental que independentemente do número de dados que recebam estas sejam capazes de processar todos os dados.

**Baixa latência:** esta representa uma resposta rápida. Em sistemas CEP, após serem recebidos inúmeros dados, como já referido, é importante que a resposta a estes dados seja realizada em tempo real, ou seja, de forma imediata ou quase imediata, de modo a que é necessário que exista uma resposta bastante rápida, representada por uma baixa latência[20].

**Escalabilidade:** significa que uma ferramenta ou sistema tem de ser capaz de evoluir no número de eventos recebidos, ou seja, num dia o sistema tanto tem de conseguir processar dados vindos de cem fontes diferentes, como no dia seguinte tem de ser capaz de processar dados vindo de mil fontes diferentes.

### 2.1.3 Vantagens e Desvantagens do CEP

Após serem conhecidas as características apresentadas pelas ferramentas e aplicações CEP, falta ainda perceber quais são as vantagens e desvantagens deste tipo de ferramentas quando aplicadas em mundo empresarial ou em conceito de sociedade.

Vantagens[21]:

- Retirar informações de eventos em tempo real.
- Retirar valor comercial da ocorrência de eventos.
- Monitorização de grandes quantidades de dados.
- Prevenção e segurança automática perante comportamentos de risco conhecidos.
- Criação de padrões de comportamentos de risco.
- Notificação de comportamentos de risco

Desvantagens[21]:

- Não permite a evolução para padrões desconhecidos, ignorando possíveis comportamentos de risco.
- Criação de falsos positivos.

### 2.1.4 Técnicas de *Big Data*

O conceito de *big data* ainda é um conceito muito genérico, onde existem inúmeras ferramentas com variadas características e que trabalham conceitos diferentes dentro deste termo, sendo necessário recorrer a conceitos mais específico, na hora de escolha do sistema ou ferramentas.

Assim sendo, pretende-se para a presente tese ferramentas capazes de processar grandes quantidades de dados, em tempo real e produzir informação após processamento dos dados, sendo apenas referidas até então as ferramentas CEP. Contudo, existe um outro conjunto de ferramentas que de certa forma também apresenta as características pretendidas, são chamadas de ferramentas de DSMS (*Data Stream Management System*).

#### 2.1.4.1 DSMS - *Data Stream Management System*

*Data Stream Management System* (DSMS) é um conceito muito utilizado na gestão de grandes quantidades de fluxos. DSMS são programas de gestão de fluxos de dados, em tempo real, estes produzem respostas na forma de *queries*, que estão em constante atualização de forma a se adaptarem aos diferentes sinais de entrada. Normalmente estes sistemas não estão associados a processamento de dados, mas sim a uma estruturação de dados, de um modo simplificado. Os DSMS recebem continuamente fluxos de dados, dados estes que serão agregados em forma de *queries* para posteriormente outras ferramentas procederem ao processamento destas informações[22].

Destacando as diferenças entre DSMS e CEP, de uma forma muito superficial observamos desde já que uma das diferenças envolve o tipo de dados, enquanto que o DSMS avalia continuamente fluxos de dados, o CEP avalia eventos, sendo que o CEP só realiza o processamento de eventos ou padrões de eventos quando detetada uma anomalia comparativamente aos padrões considerados “normais”. Outra diferença está na análise dos dados adquiridos e na sua resposta, enquanto que as aplicações CEP procedem à análise dos dados adquiridos e numa fase final notificam outras aplicações ou utilizadores de comportamentos de risco, em aplicações DSMS este tipo de ações não ocorre, deixando a cabo de outras ferramentas a análise dos dados, centralizando a sua resposta final apenas numa agregação dos dados[23].

Tendo em consideração os conceitos descritos anteriormente e os objetivos do mesmo, e tendo em vista o desenvolvimento de uma aplicação de deteção de eventos anómalos no sistema rodoviário, a ferramenta que melhor se enquadra na resolução desta solução é o CEP. No modelo apresentado, o objetivo é a análise e notificação de certas informações, ocorridas devido a eventos anómalos, sendo todas estas características apresentadas apenas em ferramentas CEP.

## 2.2 Trabalho Relacionado

Como já foi referido anteriormente, atualmente o CEP encontra-se em inúmeras áreas. Áreas como a banca, a saúde ou a área dos transportes, são apenas alguns exemplos de áreas que mais tem sofrido positivamente com a implementação de sistemas CEP, aproveitando as capacidades trazidas no processamento de eventos, para solucionar lacunas encontradas nestas áreas e tornado alguns processos nestas áreas mais eficientes.

Em seguida serão expostos alguns exemplos, onde podemos observar a aplicação do CEP em algumas das áreas referidas anteriormente.

### 2.2.1 Área Financeira

Um dos grandes problemas que atravessa a área financeira são as fraudes[19]. Devido a este tipo de crime, só no ano de 2015 foi obtido um prejuízo de 21.84 biliões de dólares

para os bancos a nível mundial[24]. Todos os anos são exploradas as vulnerabilidades nos sistemas bancários, para que de alguma forma se consiga concretizar transações bancárias ilegais.

Percebeu-se então que a deteção de fraude necessitava de um rápido, eficiente e automático processamento de dados, onde estes iriam conter informações sobre as transações bancárias. A conjugação destas informações com informações de transações bancárias anteriores ou mesmo com comportamentos considerados de risco numa transação, permitem prevenir a ocorrência de fraudes. Para além destes fatores é importante que todo este processo ocorra em segundos, porque por vezes estes tipos de operações querem-se quase de forma instantânea.

Deste modo, seria apenas necessário um sistema capaz de processar e correlacionar os dados ocorridos das transições bancárias, num curto espaço de tempo e produzir um alerta em caso de perigo, sendo escolhidos para esta tarefa sistemas CEP, onde abordando padrões de eventos, são detetados possíveis inícios de fraudes através do comportamento das transações bancárias[25].

É através da monitorização e análise de mais de cinquenta tipos de eventos, que os sistemas CEP correlacionam informações sobre várias transações a ocorrer nos diversos sistemas bancários, que se detetam possíveis transações fraudulentas[19]. A deteção de fraudes por parte de sistemas CEP, não passa de uma probabilidade de possível ocorrência de fraude, tendo de ser avaliada esta suspeita à posteriori por um operador humano[26]. Um exemplo deste tipo de aplicações é a Feedzai, empresa que conjuga várias ferramentas *big data*, incluindo ferramentas CEP, na abordagem à deteção de fraudes bancárias.

### 2.2.2 Área Agrícola

Contudo, não é só na área financeira que as aplicações CEP têm vincado a sua importância, também na área agrícola, tem-se usado sistemas baseados em aplicações CEP para fazer monitorização dos produtos.

Um dos exemplos deste tipo de aplicações é demonstrado por Zhu Huaji, Wu Huarui, Sun Xiang, onde através da integração de um modelo CEP com um sistema de sensores [Radio-Frequency Identification \(RFID\)](#), conseguem monitorizar e detetar a ocorrência de eventos aplicados a produtos agrícolas [27].

Neste estudo foi posto à prova a capacidade de processamento das ferramentas CEP. Previamente existia um problema de processamento de eventos, havia uma elevada criação de eventos e as ferramentas utilizadas para o processamento de eventos não eram as preferíveis, perdendo-se em grande parte os eventos detetados e posteriores operações sobre os mesmos. Quando unificado o sistema anterior com um sistema CEP, para processamento de eventos e deteção de anomalias, foi desenvolvido um sistema aperfeiçoado de monitorização produtos agrícolas.

Com esta nova abordagem, foram conseguidos níveis de processamento mais elevado,

que por sua vez permitiram um acréscimo no número de dados processados, como também, garantiram uma gestão de qualidade e segurança sobre os produtos.

### 2.2.3 Área de Transportes

Nas empresas de logística rodoviária, uma das questões que mais tem vindo a preocupar os empresários é a gestão de rotas[28]. Esta tem de ser capaz de satisfazer o cliente e ao mesmo tempo minimizar os custos de viagem.

Uma boa gestão de rotas requer uma constante monitorização de todos os fatores capazes de influenciar tanto os custos de viagem, como os fatores de satisfação do cliente. É importante no ponto de vista do cliente que a viagem seja realizada no tempo pretendido e que a qualidade e segurança dos produtos se mantenha intacta[29].

É então com vista a salvaguardar a qualidade e segurança dos produtos transportados, que muitas empresas de logística começam a integrar sistemas CEP nos seus veículos de transporte. Os sistemas CEP por si só, não permitem manter a qualidade dos produtos, mas quando combinados com sensores localizados em contentores de transporte, permitem de uma maneira rápida monitorizar e filtrar todos os dados geridos pelos sensores, gerando alertas sempre que algum valor avaliado, não se encontrem normal[30]. Casos destes já são comuns, tendo a sua implementação demonstrado aumentos na qualidade e segurança dos produtos[29]. Para além de um aumento na satisfação por parte do cliente, este tipo de sistemas permite um aumento nos lucros, como se pode ver no estudo realizado por Benjamin Gaunitz, Martin Roth and Bogdan Franczyk[31].

Contudo, este não é a única abordagem na utilização de sistemas CEP para gestão de rotas. Nas empresas de logística, estes sistemas são também utilizados para uma gestão inteligente dos custos de viagem. Através da monitorização de dados de tráfego e da monitorização dos veículos de transporte, os sistemas CEP permitem a análise sobre os trajetos de modo a garantir melhores consumos de combustível e menores tempo de viagem, o que mais tarde se demonstra uma vantagem a nível económico para as empresas.

Um exemplo deste caso, foi o estudo realizado com base no tráfego rodoviário de Bilbao, onde ao serem avaliados continuamente os níveis de tráfego, se pôde influenciar as rotas escolhidas, obtendo-se resultados bastante satisfatório na gestão de rotas, usando as ferramentas CEP para processamento dos dados de tráfego[32].

Porém, este não é um caso único, outro exemplo da aplicação de sistemas CEP na área de transportes é o projeto MOBIS.

O projeto MOBIS é um projeto europeu com vista a criar uma plataforma de mobilidade, personalizada e inteligente, através da aplicação de novas tecnologias da Internet e métodos de inteligência artificial, sendo que uma das vertentes deste projeto, consistia ainda no desenvolvimento de uma plataforma de previsão de eventos rodoviários[33][34].

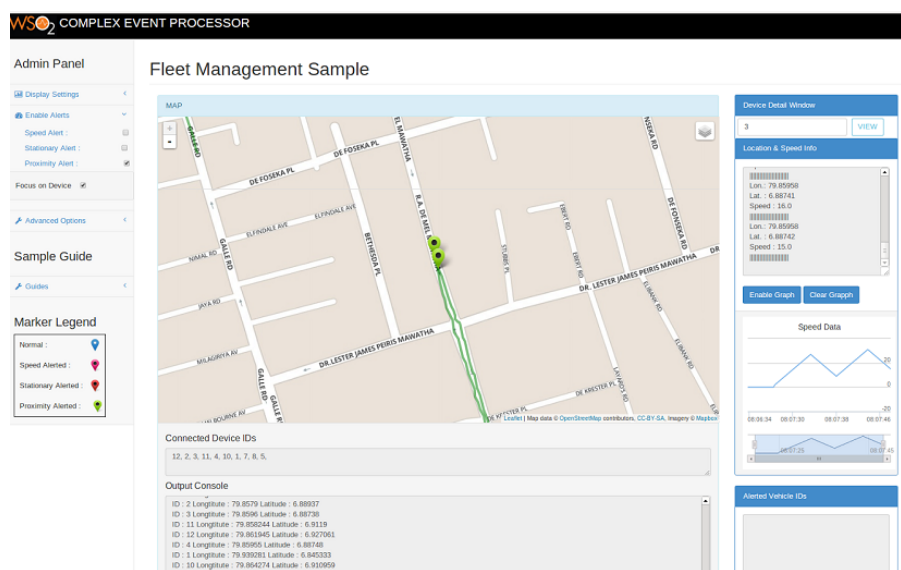


Figura 2.1: Painel de visualização de uma aplicação CEP de gestão de rotas

Nesta vertente de previsão de eventos rodoviária foram utilizadas ferramentas CEP de modo a analisar o comportamento dos utentes de autoestradas para criação de padrões comportamentais, que mais tarde permitiriam a identificação de eventos anómalos no sistema rodoviário[35].

Este projeto pode ser relacionado em alguns objetivos com os objetivos desta presente dissertação, contudo no projeto MOBIS apenas se pretendia a identificação da ocorrência de eventos anómalos no sistema rodoviário, enquanto que na presente dissertação a aplicação a ser desenvolvida para além da identificação dos eventos anómalos, prevê uma categorização do evento, ou seja, se este é um acidente, trânsito, etc.

#### 2.2.4 Área Desportiva

A área desportiva é uma das áreas que nos últimos tempos, mais tem promovido a introdução de novas tecnologias, tanto para estudo dos atletas como para salvaguardar decisões por parte das equipas de arbitragem[36].

Com vista a promover as ferramentas CEP no desporto, a [Association for Computing Machinery's \(ACM\) International Conference on Distributed Event-Based Systems \(DEBS\)](#), que é uma das principais conferências dedicadas à computação baseada em eventos[37], promoveu um desafio onde o objetivo era analisar um jogo de futebol utilizando um sistema baseado em eventos, processando fluxos de eventos gerados em sensores conectados a jogadores e à bola.

Tendo uma abordagem muito simples, este tipo de ferramentas pretendia calcular de forma automática o número de passes com sucesso de cada jogador, o número de passes falhados, o número de remates, posse de bola, entre outros parâmetros[38].

Ao se concluir este desafio, veio-se provar que a aplicações de ferramentas CEP aplicadas ao desporto também representa uma mais valia, podendo criar de forma proactiva



uma avaliação sobre os atletas e o jogo analisado. Na fig.2.2 pode-se observar uma das aplicações criadas para esta conferência.



Figura 2.2: Painel de visualização de informações em tempo real sobre um jogo de futebol, tendo por base uma aplicação CEP

### 2.2.5 Avaliação de ferramentas CEP

Muitas são as empresas que hoje em dia trabalham na área do CEP e que desenvolvem ferramentas que permitem o trabalho nesta área. De modo a selecionar a ferramenta que irá acompanhar a presente dissertação, foi realizado um breve estudo apresentado na tabela 2.1, onde são avaliadas algumas características das ferramentas CEP.

Tabela 2.1: Caracterização de ferramentas CEP

Nome	Baixa Latência/ Elevado Processamento	Escalabilidade	Licença	Linguagem de Programação	Apoio Técnico
Oracle CEP	✓	✓	Comercial	Java	Documentação Online e Comunidade Online
WSO2 Siddhi	✓	✓	Open Source	Java	Documentação Online
Esper	✓	✓	Open Source	Java	Documentação Online e Comunidade Online
StreamBase CEP	✓	✓	Comercial	SQL	Documentação Online e Comunidade Online
SASE	✓		Open Souce	SQL	Documentação
TIBCO BusinessEvents	✓	✓	Comercial	Java	Documentação Online e Comunidade Online

Com base no estudo realizado, e como podemos observar todas as ferramentas apresentam as características pretendidas, contudo as que mais se adequavam no contexto do trabalho apresentado são as ferramentas WSO2 Siddhi e ESPER, pois ambas apresentam licenças livres e têm como base de linguagem de programação o Java e uma aproximação ao SQL, sendo este último aspeto apenas uma escolha de preferência pessoal.

Aquando da escolha entre ambas, esta recaiu para a ferramenta WSO2 Siddhi, pois apesar de ambas as ferramentas assumirem a capacidade de escalabilidade, alguns autores



criticam a ferramenta ESPER por apesar de esta assumir-se como escalável, a escalabilidade apresentada por esta não é muito eficiente.

Outra diferença exposta entre ambas, consiste na documentação de apoio, apesar de a ferramenta ESPER apresentar documentação online e uma comunidade online, a documentação fornecida por parte do WSO2 Siddhi é mais clara e contém vários exemplos práticos, sendo esta uma das razões pela qual a escolha recaiu sobre esta ferramenta.

## 2.3 CRISP-DM

Após serem conhecidos os objetivos para a presente dissertação, para que esta cumpra uma linha de pensamento e de trabalho corretos, foi adotado o modelo de trabalho em análise e tratamentos de dados [Cross Industry Standard Process for Data Mining \(CRISP-DM\)](#).

O *Cross Industry Standard Process for Data Mining* é um modelo de processos de *data mining*, no qual se apresenta independente do sector da indústria ou da tecnologia usada.

Como se pode observar pela fig.2.3 o CRISP-DM é baseado em seis fases. Este pressupõe a execução destas seis fases, não existindo um ciclo na transição entre fases, ou seja, a transição de fases pode ocorrer para uma fase que já tenha ocorrido se assim necessário[39].

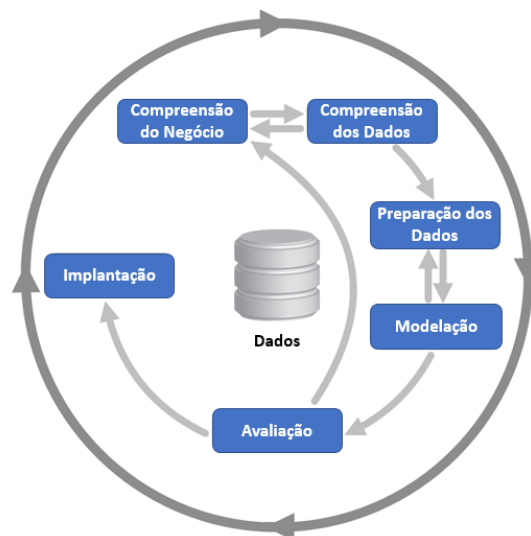


Figura 2.3: Diagrama do modelo de processo CRISP-DM

**Compreensão do Negócio:** Nesta fase inicial o modelo CRISP-DM foca-se com uma visão empresarial na compreensão dos objetivos e requisitos do projeto, sendo traçado um plano para os objetivos definidos.

Esta fase na presente dissertação representa uma introdução e aprofundamento do tema onde são analisadas as várias áreas que a englobam, desde os sistemas de transporte inteligentes, até à área e técnicas de *Big Data*, sendo ainda numa fase final criado um plano para desenvolvimento dos objetivos propostos.

**Compreensão dos Dados:** Esta fase é responsável pela colheita de dados, prosseguindo ainda com um início de atividades para enquadramento com os dados, onde serão estudados os dados rodoviários, identificando problemas com estes. É ainda realizada uma breve análise dos mesmos onde são detetados conjuntos de dados com informações ocultas.

**Preparação dos Dados:** Durante esta fase ocorrem um conjunto de transformações sobre os dados, serão executadas pequenas transformações para que estes, quando vindos de diferentes fontes possam ser uniformizados, para que numa fase seguinte possam ser modelados. Nesta fase são transformados dados brutos em dados capazes de serem tratados posteriormente.

Esta fase é muito importante, pois na presente dissertação são utilizadas diversas fontes de dados, na qual é necessário garantir coesão nos conjuntos de dados para modelações futuras.

**Modelação:** É na fase de modelação que são aplicadas várias técnicas de tratamento e moldagem de dados, são ainda calibrados os parâmetros para uma otimização da aplicação desenvolvida.

Esta fase passa pelo desenvolvimento da aplicação de deteção pró-ativa de eventos anómalos em autoestradas, onde através de vários estudos e análises serão criados padrões que permitam identificar o normal comportamento rodoviário para que posteriormente seja comparado com os dados dos sensores para identificação de eventos rodoviários anómalos.

**Avaliação:** Esta fase é uma fase de avaliação, no decorrer dela, já foi criado um modelo, neste uma aplicação, que aparenta apresentar grande qualidade e eficiência no tratamento e análise dos dados, contudo esta fase não prescinde de uma avaliação do modelo criado, verificando se o mesmo cumpre os objetivos do negócio.

No enquadramento da dissertação, esta fase passa pela validação de todos os processos e regras utilizados na criação da aplicação de deteção de eventos.

**Implementação:** Após serem cumpridos os objetivos do projeto na fase de avaliação, na implementação, o modelo desenhado terá de ser organizado de forma a apresentar-se como uma ferramenta útil para que o cliente a possa utilizar, consiste assim na implementação otimizada da aplicação desenvolvida.

## FONTES DE DADOS

Quando se fala em tecnologias de *big data*, estas pressupõem a utilização de enormes quantidades de dados, sendo através do rápido processamento e tratamento de dados que este tipo de tecnologias, tem vindo a vincar o seu papel dentro de inúmeras aplicações para tratamentos de dados [27][40][41][42][43][44].

Como tal, as aplicações CEP, enquadrando-se neste tipo de tecnologias, tendo vindo a destacar a sua relevância, sendo muito utilizados em aplicações que preveem a deteção de eventos[27][41][42][43]. Contudo para que a deteção de eventos seja fidedigna, este tipo de aplicações conjeturam uma análise prévia de dados, para criação de padrões sobre os dados em estudo[40][35].

Sendo ainda importante de se salientar que uma das razões pela qual este tipo de tecnologias têm vincado o seu papel junto de inúmeras aplicações, deve-se ao facto de o rápido processamento (elevado desempenho) ser executado em tempo , permitindo adquirir valor imediato sobre qualquer tipo de dados processados[40][42][41][35].

Como já foi referido anteriormente, esta dissertação objetiva a deteção de eventos anómalos em autoestradas. Para tal foi necessário recorrer a uma fonte múltipla de dados na qual existe uma aquisição de valores comportamentais de veículos rodoviários em autoestradas, onde esta recorre a vários sensores devidamente colocados em autoestradas. Aquando do avanço da presente tese foi também utilizado uma fonte de dados simples, contendo um repositório de eventos rodoviários (acidentes, trânsito, reparações na via,...), sendo de salientar, que introdução de dados/eventos nesta fonte de dados é realizada manualmente sempre que é detetada a ocorrência de um evento. A utilização fonte de dados deve-se por esta conter registo de eventos, facilitando o estudo sobre a identificação de eventos e a criação de padrões comportamentais sobre os mesmos.

Ao longo do presente capítulo serão abordados os conceitos relacionados com os dados,

não tendo em consideração o tipo de sensores que geram os dados, mas sim os diferentes tipos de dados gerados pelos vários sensores. Numa parte posterior deste capítulo realizar-se-á uma avaliação espaciotemporal e de qualidades dos dados, onde serão ainda abordadas todas as operações necessárias para a organização, limpeza e seleção dos dados até então estudados.

### 3.1 Compreensão de Dados

Para que se possa proceder ao tratamento de dados é necessário numa fase inicial uma compreensão e estudo dos mesmo. Sendo nesta fase essencial avaliar a formatação e descrição dos dados, passando posteriormente a uma apreciação da cobertura espaciotemporal e qualidade dos dados em estudo.

#### 3.1.1 Organização Estrutural dos Dados

Sem executar qualquer operação dos dados, os dados de sensores referentes às autoestradas apresentam duas formatações (esquemas de dados) distintas, uma formatação diferente consoante a sua cobertura temporal, tendo um esquema de dados alusivos aos meses de janeiro a novembro de 2016 e outro tipo de esquema de dados referentes aos meses de dezembro de 2016 até maio de 2017. Esta diferença pode ser observada na fig.3.1, onde não existe correspondência entre os campos dos diferentes conjuntos de dados.

```

1 {
2   "_id" : ObjectId("59679712911c3f2118ed41a9"),
3   "feed" : {
4     "xmlns_crs" : "http://www.dars.si/schemas/crs",
5     "updated" : NumberInt(1451949623),
6     "xmlns" : "http://www.w3.org/2005/Atom",
7     "subtitle" : "DRSC - Poročilo o trenutnem stanju promet",
8     "xmlns_stevci" : "http://www.dars.si/schemas/stevci/v1",
9     "crs_crsId" : "EPSG:2170",
10    "title" : "Števeci prometa",
11    "author" : {
12      "email" : "info@mikrobit.si",
13      "name" : "Mikrobit senzorika d.o.o."
14    },
15    "xmlns_georss" : "http://www.georss.org/georss",
16    "link" : {
17      "href" : "http://www.promet.si"
18    },
19    "entry" : [
20      {
21        "stevci_vmax" : NumberInt(100),
22        "stevci_smerOpis" : "Barjanska - Peruzzijska",

```

```

1 {
2   "_id" : ObjectId("5967958f911c3f2118ed3edc"),
3   "RoutingVersion" : NumberInt(1),
4   "updated" : NumberInt(1483571283),
5   "TileVersion" : NumberInt(0),
6   "ModifiedTime" : "2017-01-04T23:07:26.6209236Z",
7   "copyright" : "Prometno-informacijski center za državne ceste",
8   "ModelVersion" : NumberInt(1),
9   "Expires" : "2017-01-04T23:08:31.605085Z",
10  "Contents" : [
11    {
12      "Language" : "sl_SI",
13      "ModifiedTime" : "2017-01-04T23:06:06.0911122Z",
14      "IsModified" : true,
15      "ContentName" : "stevci",
16      "Expires" : "2017-01-04T23:16:06.0911122Z",
17      "ETag" : "1R7hHICiCxsdvhVLIUAzxNIHHq0BAEcqJIn2gsMfJXoeSq",
18      "Data" : {
19        "ContentName" : "stevci",
20        "Language" : "sl_SI",
21        "Items" : [
22          {

```

Figura 3.1: Comparação estrutural entre dados de 4 de janeiro de 2016 (esquerda) e 04 de janeiro de 2017 (direita)

Apesar de o tipo de formatação ser bastante distinto, esta não influência no tipo de dados gerados, sendo apenas distintos na maneira como os dados estão projetados na base de dados. A existência de campos díspares entre os vários conjuntos de dados, também não se demonstra um problema, pois estes campos são meramente informativos e sem qualquer utilização futura.

Relativamente aos dados de eventos rodoviários, podemos observar a sua formatação na fig.3.2

```

{
  "_id" : ObjectId("58b3a2f65a791a392480faa9"),
  "Cesta" : "A2-E61",
  "Description" : "A2, Karavanke - Ljubljana, med predorom L:",
  "SmerStacionaza" : NumberInt(0),
  "IsRoadClosed" : false,
  "SideContent" : "04:29",
  "PrioritetaCeste" : NumberInt(2),
  "Stacionaza" : NumberInt(4117),
  "y_wgs" : 46.31327317194146,
  "Odsek" : "0005",
  "isMejniPrehod" : false,
  "X" : 444081.714489242,
  "Y" : 130168.31201492,
  "Kategorija" : "A2",
  "CrsId" : "EPSG:2170",
  "x_wgs" : 14.269018667466547,
  "Updated" : ISODate("2017-02-27T03:30:29.000+0000"),
  "Title" : "Izredni dogodek",
}

```

Figura 3.2: Parte da estrutura de dados de eventos rodoviários

Em ambos os conjuntos de dados referidos até agora, é de referir que a formatação dos dados assenta sobre ficheiros do tipo *JSON*, implementado numa base de dados.

### 3.1.2 Metadados

Posteriormente à exploração estrutural, é importante proceder-se a descrição de cada atributo nos diferentes conjuntos de dados. Ao conceito de descrição de dados associamos o termo metadados, onde este significa “dados de dados” [45].

Este termo é muito utilizado na área da informática e dos bancos de dados[45], através deste conceito consegue-se compreender o significado de cada atributo associado aos diferentes conjuntos de dados.

Nas tabelas, tab.I.1, tab.I.2, tab.I.3, encontradas no anexo I, pode-se observar a descrição de cada campo de atributos associado aos diferentes conjuntos de dados até então considerados.

### 3.1.3 Cobertura espaciotemporal

Após se perceber a organização dos dados é essencial compreender a sua cobertura espaciotemporal, onde de certo modo se consegue avaliar se os dados a ser estudados demonstrarão fiabilidade para que se possa executar um estudo sobre os mesmos.

A nível temporal existem dois parâmetros a ter em consideração na presente tese, a sua escala temporal e a disponibilidade temporal, onde a escala temporal representa o espaço temporal abrangido pela aquisição de dados, e a disponibilidade é um valor percentual que indica a quantidade de dados ao longo de um mês, de uma semana ou de um dia, para cada sensor. Desse modo, quanto maior for a escala temporal e disponibilidade dos dados,

mais fiável será o estudo realizado e mais realista serão os níveis padrão para disparo de eventos.

Considerando estes dois parâmetros, foram definidos valores mínimos para considerar a utilização dos presentes dados fiáveis. A nível de escala temporal, o pré-requisito apontado para este parâmetro é uma dispersão temporal de seis meses, ou seja que existam registos ao longo de seis meses, contudo a este não é o único pré-requisito, sendo ainda necessário garantir uma disponibilidade diária e mensal dos dados superior a 80% por sensor. A disponibilidade dos dados é uma característica dos dados, dada pelo número de registos recebidos (eq.3.1), sobre o número de registos esperados, ou seja, se um sensor em funcionamento e sem falhas produz 144 registos por dia, considerando então a ocorrência de falhas este, tem de apresentar pelo menos 116 registos diários (80%) para que o presente dia seja considerado para estudos futuros, sendo aplicada a mesma lógica para a disponibilidade mensal.

$$Disponibilidade = \frac{Número\ de\ registos\ recebidos}{Número\ de\ registos\ esperados} \times 100 \quad (3.1)$$

$$Disponibilidade = \frac{116}{144} \times 100 = 80\%$$

Examinando ao pormenor a escala temporal dos dados em estudo, ambos os dois conjuntos de dados referidos até então, apresentam uma dispersão temporal de um ano e cinco meses. Apesar de se denotar falhas temporais, nos dados referentes a leituras dos sensores das autoestradas, fig.3.3, ilustradas pelos espaços em branco, estas não são suficientemente grandes para que tenha influência futura no estudo dos mesmos. Para a fig.3.3, foram escolhidos 15 sensores de forma aleatória, entre os 355 sensores existentes, apenas tendo em consideração a dispersão espacial dos mesmos ao longo de todo território esloveno.

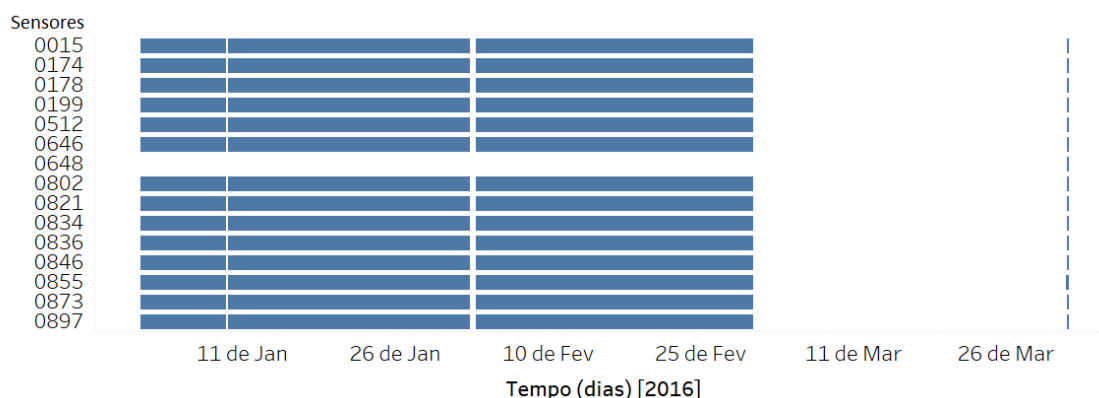


Figura 3.3: Falhas na cobertura temporal de 15 sensores

O segundo parâmetro em consideração foi então a disponibilidade. Relativamente aos dados de sensores em autoestradas, estes apresentaram uma disponibilidade superior a 80% para a maioria dos dias e meses, à exceção do mês de março, respeitando assim o

pré-requisito imposto. Quanto aos dados de eventos em autoestradas, estes não foram sujeitos a qualquer tipo de avaliação de disponibilidade, pois este conjunto de dados, contrariamente ao anteriormente referido, não apresentam qualquer periodicidade, não podendo-se definir assim um valor de disponibilidade sobre os mesmos.

De forma a ilustrar a disponibilidade mensal, foram escolhidos aleatoriamente 9 sensores dispersos ao longo do território esloveno, estudo este possível de se observar fig.3.4. A fig.3.5 representa a disponibilidade diária para um sensor escolhido de forma aleatório, exemplificando um estudo mais profundo sobre a disponibilidade diária de cada sensor.

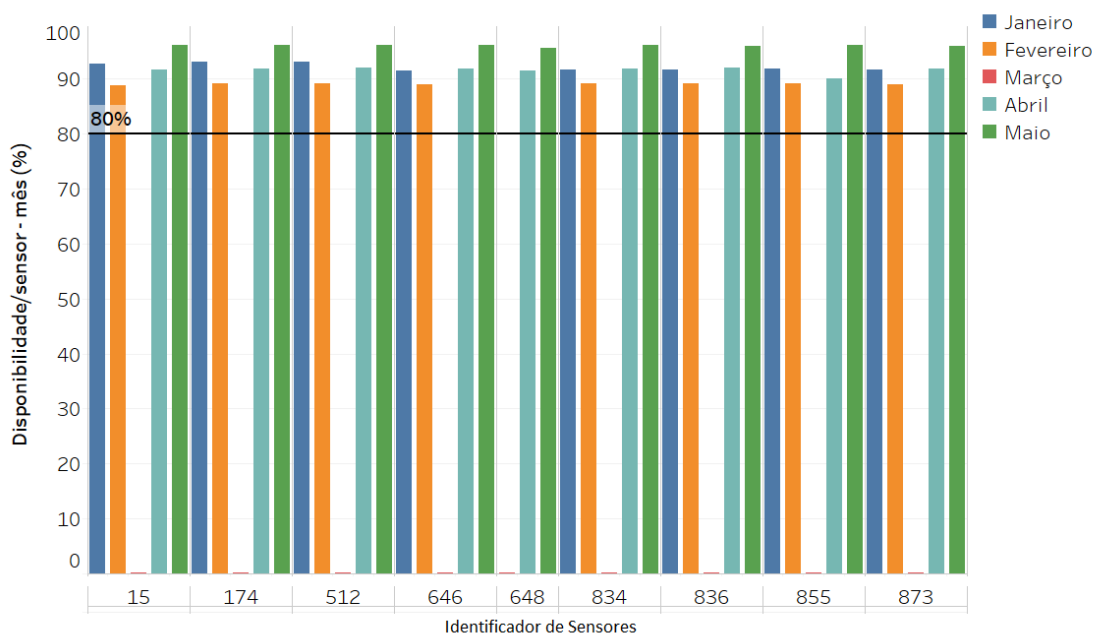


Figura 3.4: Disponibilidade mensal para cinco meses de 2016, tendo em consideração 9 sensores

Terminando a apreciação temporal dos dados, falta realizar uma avaliação espacial. Não existindo nenhum pré-requisito relativamente a esta avaliação, tendo apenas como único objetivo, que os dados cobram a maior quantidade de área territorial possível. Para se perceber melhor a cobertura espacial dos dados observe-se as seguintes figuras, fig.3.6 e fig.3.7.

É de salientar que ao longo desta dissertação serão utilizados um total de 355 sensores, dispersos ao longo de várias autoestradas eslovenas, cobrindo em grande parte todo o território esloveno, como se pode observar pela fig.3.6. Estes sensores transmitem informações com intervalos de tempo aproximados de 10 minutos, onde em cada conjunto de dados são apresentados valores de velocidade média, ocupação da faixa, o registo do tempo, entre outros, para mais informações sobre os dados que os sensores transmitem, consultar as tabelas I.2 e I.3 do anexo I.

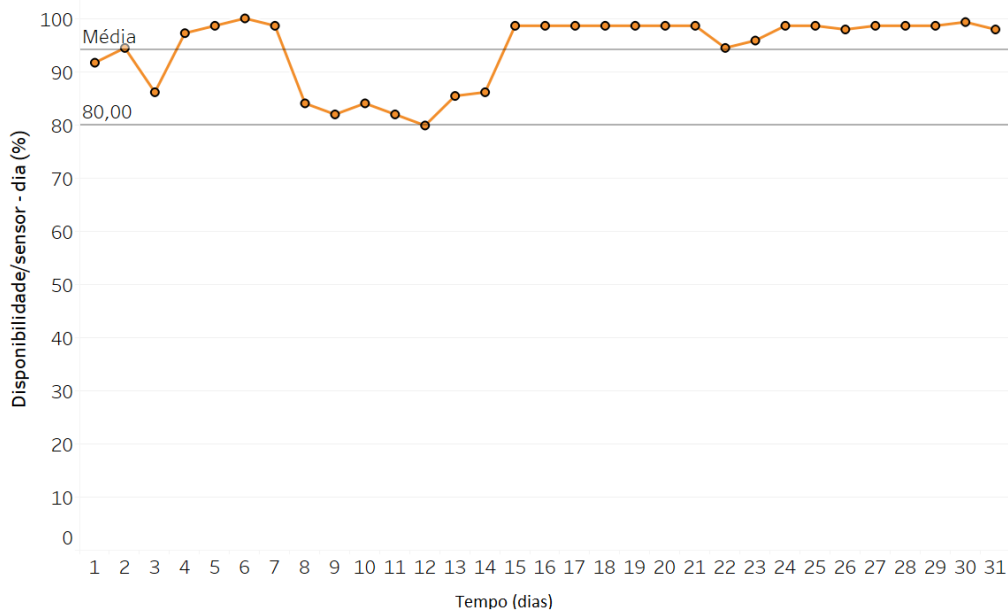


Figura 3.5: Disponibilidade diária do sensor 174 no mês de janeiro de 2017

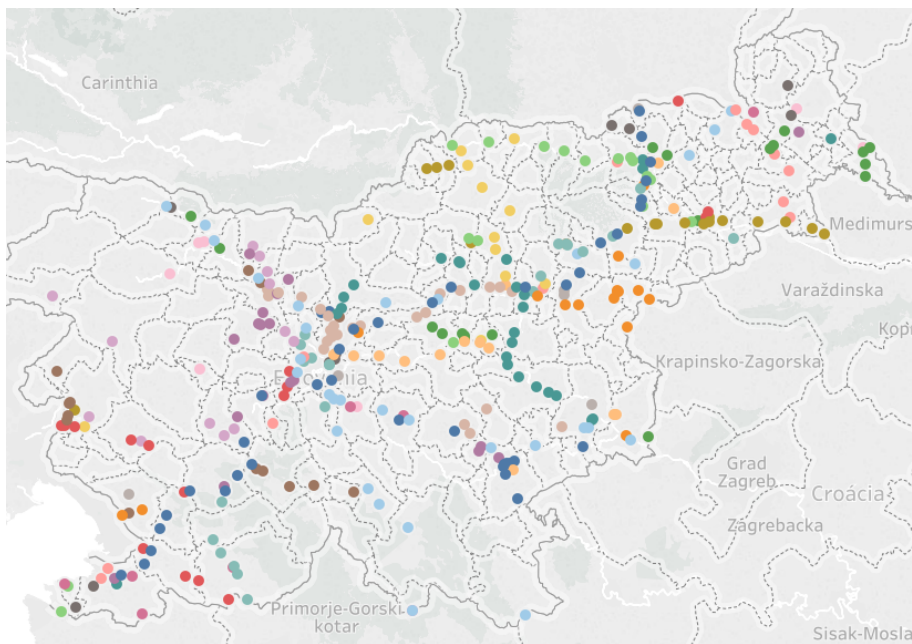


Figura 3.6: Projeção espacial da localização dos sensores



Mais uma vez, a fig.3.7, referentes à localização de eventos em autoestradas, apresentam uma elevada dispersão espacial, cobrindo muitas e diferentes áreas de possível estudo.

Desta forma apesar da não existência de pré-requisitos relativamente à dispersão espacial, esta apresenta uma dimensão bastante interessante, devido as várias zonas cobertas tanto pelos diversos sensores, como pela base de dados de eventos.

Deste modo, conclui-se que tanto os pré-requisitos da componente temporal dos dados, como a enorme dispersão espacial apresentada pelo vários sensores e eventos, representa pontos bastante positivos, demonstrando que todos os conjuntos de dados e fontes de dados avaliados até agora, apresentam características espaço-temporais que permitirão a realização de estudos futuros, segundo estas características.

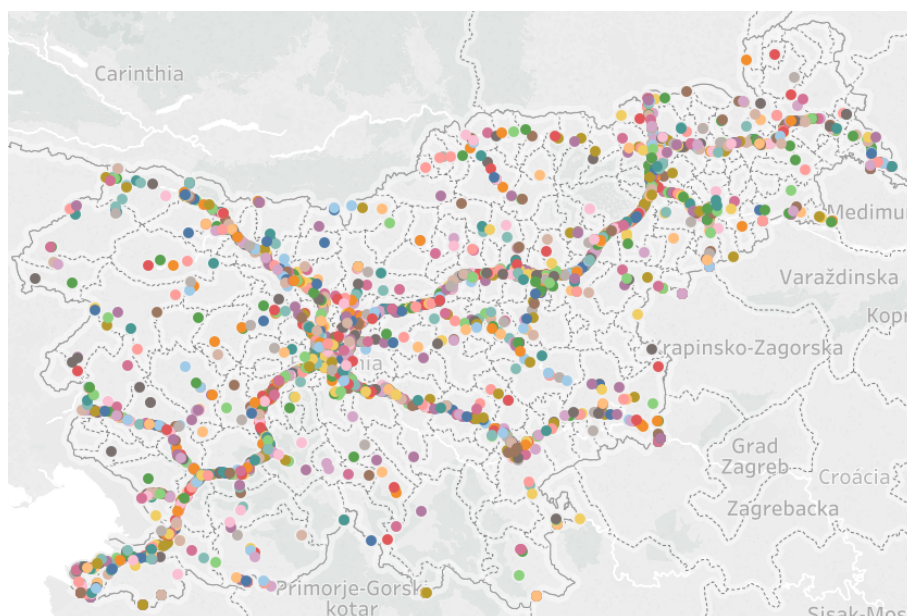


Figura 3.7: Projeção espacial da localização dos eventos

### 3.1.4 Qualidade dos Dados

Após realizada a descrição dos dados, relativamente a sua formatação, cobertura espaciotemporal e descrição de atributos, de seguida abordar-se-á o conceito de qualidade dos dados.

O conceito de qualidade dos dados pode ser definido como "aptidão para o uso", o que por sua vez implica que este conceito seja subjetivo. Assim, dados considerados com qualidade apropriada para um determinado uso podem não possuir qualidade suficiente para um outro uso distinto[46].

Os problemas associados a qualidade de dados podem denotar-se consoante o tipo de fonte de dados utilizadas, podendo surgir em fontes simples, mas sendo mais críticos quando utilizadas múltiplas fontes de dados que necessitem de ser integradas[47].

Desse modo, para se proceder a avaliação da qualidade dos dados, esta avaliação integrará duas vertentes, uma a nível estrutural e outra a nível dos atributos, consoante o tipo de fonte de dados utilizada.

A avaliação de qualquer conjunto de dados a nível de estrutural é facilitada pela utilização de bases de dados, estas introduzem restrições a nível do modelo de dados utilizados, tendo os dados de respeitar o modelo pré-definido para que se proceda a introdução dos mesmos na base de dados. Não existindo qualquer problema a nível individual com a estruturação dos dados, considerando nesta fase todos os dados como dados provenientes de fontes simples. Contudo quando se analisam os dados relativos a valores comportamentais de veículos rodoviários, estes apresentam uma formatação diferente consoante a sua aquisição temporal, o que gera um problema a nível estrutural aquando da junção destes conjuntos de dados.

Outro tipo de avaliação estrutural necessário de se realizar em base de dados é a existência de duplicação de dados. Este problema é resolvido por um identificador único que permita a não duplicação de dados, contudo e apesar de existir um identificador único em ambos os conjuntos de dados estudados, os dados provenientes da fonte de dados relativas a valores comportamentais de veículos rodoviários, ostentam duplicação dos mesmos. Este erro deve-se pelo facto de o identificador único de dados não ter sido criado de forma correta. O identificador deveria apresentar-se como uma junção do campo "ModifiedTime" e do campo "Id", a qual não existiu e permitiu a repetição de dados com identificadores diferentes.



Figura 3.8: Visualização de duplicados nos registos relativos a leituras de sensores

Através da fig.3.8 é possível observar a existência de duplicados, onde seria de se esperar apenas um registo por cada instante temporal, o que não ocorre, sendo as anomalias

identificados pelas ovas a vermelho.

Quando se avalia os atributos de uma fonte de dados, é importante perceber a existência de erros na introdução de campos, ou seja, se existe valores de atributos que possam ser considerados anormais ou erros, normalmente este tipo de erros apresenta-se sobre a forma de inconsistências na representação de determinados atributos, sendo o exemplo mais comum um parâmetro classificado como inteiro que depois apresenta o seu valor na forma de *strings*. Do estudo prévio realizado nas fontes de dados até então estudadas, apenas se sinalizou um erro na introdução de atributos, onde de alguma maneira nos dados de valores comportamentais de veículos, existiu a repetida introdução de um conjunto de dados referentes ao mês de agosto de 2015, fora da gama temporal estudada, onde todos os atributos alusivos a esta gama temporal se apresentavam com o valor do atributo a nulo. Na fonte de dados de eventos rodoviários, não se sinalizou qualquer anomalia/erro nesta formatação, assumindo que os dados nesta avaliação se encontram válidos.

### 3.2 Seleção de Dados

A realização de uma boa preparação de dados é um pré-requisito fundamental para o sucesso em “data mining”, estando esta na base de aplicações CEP. Inúmeros estudos sugerem que a preparação dos dados num projeto com base em “data mining”, leve cerca de 60 a 80% do tempo envolvido [48]. Esta fase é também associada à terceira fase do modelo de trabalho CRISP-DM (“Preparação dos Dados”) apresentado no capítulo 2.

Todo o processo de seleção de dados é um processo demorado[48], na qual se pretende executar ações que visam detetar e remover anomalias nos dados, aumentando/melhorando a qualidade associada a estes[49]. Ao processo de deteção e remoção de anomalias de forma a melhorar a qualidade dos dados dá-se o nome de limpeza de dados, esta assume um papel muito importante no contexto da gestão de grandes repositórios de informação.

Sendo este um processo seleção, de seguida serão abordadas apenas as ações que permitam melhorar/solucionar os problemas anteriormente identificados no estudo da qualidade de dados, referenciados no cap.3.1.4.

Um dos problemas identificados anteriormente foi a duplicação de dados. Para se solucionar este problema foi filtrado o conjunto de dados com base no tempo de aquisição de dados (“ModifiedTime”) e no identificador do sensor (“Id”). Desse modo, criou-se um conjunto de dados únicos, com base no tempo de aquisição de dados e identificador do sensor.

Após realização deste processo de limpeza de dados, todos os conjuntos de dados apresentam apenas um único registo por cada leitura dos sensores, procedendo-se assim a eficaz eliminação de duplicados.

Outro dos problemas identificado aquando da avaliação da qualidade de dados, foi a introdução errada de atributos num conjunto de registo, onde para além da cobertura temporal não representar a cobertura temporal estudada (de janeiro de 2016 a maio de

2017), todos os campos referentes a este conjunto de dados, apresentava como atributo, o valor nulo, independentemente do tipo da variável (inteiro ou *string*). Para resolver este problema, mais uma vez realizou-se uma filtragem dos dados e posteriormente remoção dos mesmos, não sendo considerados daqui em diante em nenhuma coleção de dados. A eliminação destes dados foi a decisão tomada, porque estes representam um erro dos sensores ou um erro na passagem dos dados para base de dados, não revelando qualquer benefício a utilização destes dados na criação futuramente de padrões comportamentais dos dados.

Por fim o último erro detetado que até então não obteve qualquer resolução, foi a incompatibilidade de estruturas/modelos, apresentando entre os dados de valores comportamentais de veículos rodoviários com projeções temporais diferentes. Para resolução deste problema, foi necessário criar-se um novo conjunto de dados, onde foram transcritos todos os campos em comum entre os dois conjuntos diferentes de dados, uniformizando desse modo estes conjuntos.

De modo a facilitar o posterior estudo dos dados, denotou-se aquando da seleção final de dados, a existência de campos em comum que eram repetidos a cada nova aquisição de dados. Estes campos são campos identificadores, descritivos ou de localização do sensor, que se iriam repetir sempre que o sensor realiza uma nova aquisição de dados. Assim, de modo a diminuir a repetição de informações, dividiu-se o conjunto de dados referido, em dois novos conjuntos, um informativo e caracterizador do sensor e outro relativo aos dados adquiridos pelo sensor. Estes novos conjuntos de dados serão unificados pelo identificador único dos dados informativos do sensor.

Esta solução ao permitir a diminuição da repetição de certo número de atributos nos dados, permitiu por sua vez diminuir o tamanho de cada conjunto de dados, o que mais à frente na presente dissertação vai possibilitar um processamento mais rápido dos dados.

Assim alcançamos com a seleção final de dados, dois novos conjuntos de dados, um caracterizador dos sensores (fig.3.9), com campos que representam a localização dos sensores, através de coordenadas, e que identificam a autoestrada em que os sensores estão instalados. O outro conjunto de dados, representa os valores comportamentais dos veículos rodoviários em cada sensor, como por exemplo a velocidade média, a ocupação, entre outros. A formatação deste conjunto de dados e todos os seus campos pode ser visualizado na fig.3.10.

```

"_id" : ObjectId("5919e285452fbb232810f84c"),
"y_wgs" : 46.01962915242322,
>Description" : "AC-A1, LJ (južna obvoznica)",
>Title" : "AC-A1, LJ (južna obvoznica)",
>ContentName" : "stevci",
>x_wgs" : 14.51818035950113,
>CrsId" : "EPSG:2170",
>stevci_cestaOpis" : "AC-A1",
>Y" : 97385.0,
>X" : 463074.0,
>Id" : "0178",
>stevci_lokacijaOpis" : "LJ (južna obvoznica)",
>location" : {
  >type" : "Point",
  >coordinates" : [
    14.51818035950113,
    46.01962915242322
  ]
}
}

```

Figura 3.9: Formatação dos dados informativos de cada sensor

```

"_id" : ObjectId("592f249dd2fb3d2470a0d59d"),
>ReadingPoint_id" : "5919e285452fbb232810f84c",
>ModifiedTime" : ISODate("2016-02-01T00:00:00.000+0000"),
>Data" : [
  {
    >properties" : {
      >stevci_gap" : "999,9",
      >stevci_statOpis" : "Ni prometa",
      >stevci_hit" : "0",
      >stevci_stev" : "0",
      >stevci_pasOpis" : "(v)",
      >stevci_smerOpis" : "Barjanska - Peruzzijska",
      >stevci_stat" : "6"
    },
    >Id" : "0178-21"
  },
]
}

```

Figura 3.10: Formatação dos dados referentes às leituras dos sensores

É de salientar que relativamente ao conjunto de dados, referentes a eventos rodoviários, não foi realizada nenhuma operação sobre os mesmos.

### 3.3 Exploração dos Dados

A exploração de dados é uma abordagem semelhante à análise inicial de dados, pelo que se explora visualmente o conjunto de dados, de modo a obter-se uma caracterização sobre os conjuntos de dados. Nesta abordagem estão incluídos todos os processos de seleção e qualidade de dados explorados nas secções anteriores, 3.1.4 e 3.2.

Na presente secção a abordagem de exploração de dados incidirá sobre a maximização global dos dados tendo em consideração a densidade populacional, irão ainda ser verificados padrões comportamentais, de modo a perceber o comportamento horário rodoviário para que sejam analisadas anomalias no comportamento médio de fluxo rodoviário.

#### 3.3.1 Relação ocorrência de eventos com a distribuição populacional

A primeira abordagem a realizar-se na exploração de dados correlacionará a ocorrência de eventos com a distribuição populacional no território esloveno, identificando a influência da densidade populacional na ocorrência de eventos.

Através da fig.3.11 é exposta visualmente a densidade populacional no território esloveno, onde é visível uma maior concentração populacional no centro e nordeste esloveno.

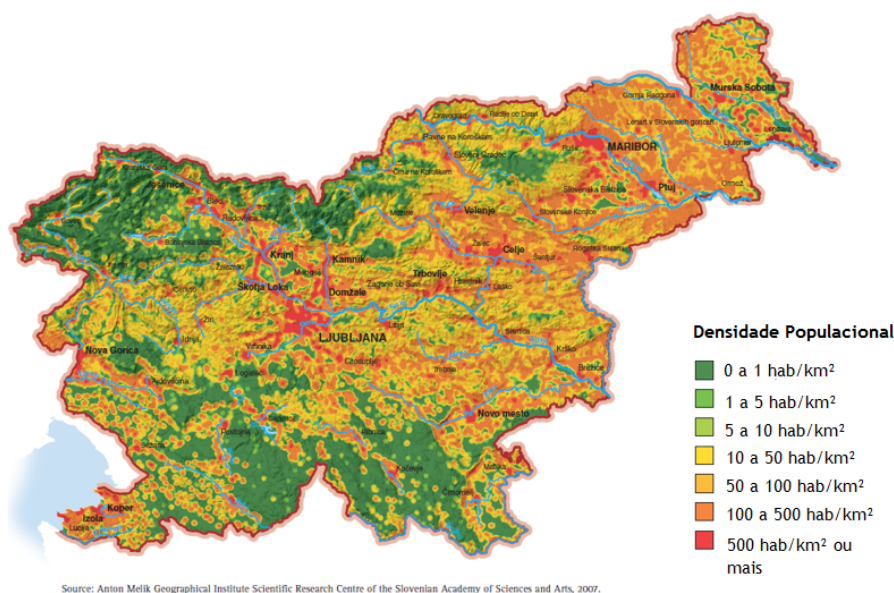


Figura 3.11: Densidade populacional em território esloveno [50]



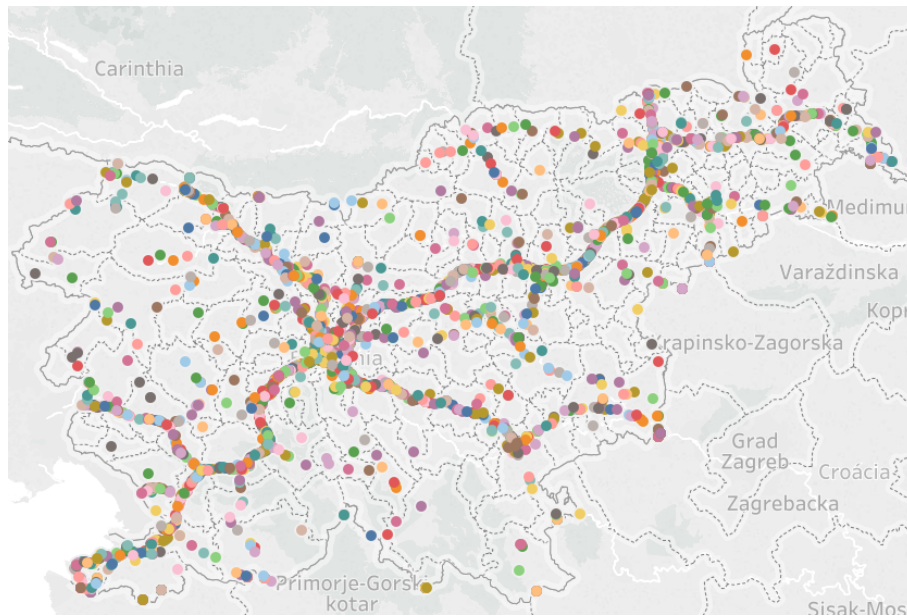


Figura 3.12: Projeção espacial da localização dos eventos

Comparando a fig.3.11 com a ocorrência de eventos, ilustrado pela fig.3.12, é visível um maior número de eventos junto dos maiores centros populacionais eslovenos, sendo estes dados esperados pelo maior número de veículos em circulação nestas zonas. Outra observação relevante é a ocorrência de um número também considerável de eventos entre a ligação dos dois grandes centros populacionais eslovenos (Liubliana e Maribor), sendo deste modo importante avaliar o comportamento rodoviário em três categorias diferentes de zonas, zonas populacionais com elevado número de habitantes, ou seja, entre zonas com densidade populacional superior a  $100 \text{ hab/km}^2$ , zonas populacionais com baixo número populacional, com densidade populacional inferior a  $100 \text{ hab/km}^2$  e zonas que liguem dois centros populacionais de elevado número de habitantes.

### 3.3.2 Valores Anómalos

A realização de médias é fundamental na exploração de dados. Através da criação de médias sobre os atributos de um conjunto de dados é possível prever o padrão comportamental de cada atributo e identificar possíveis anomalias na atribuição destes valores.

Na presente secção 3.3.2, avaliar-se-á numa primeira fase a disposição temporal da velocidade e ocupação de via, de modo a possibilitar a criação de padrões comportamentais, para que seja possível a identificação de valores discrepantes no "normal" comportamento rodoviário numa autoestrada. Este estudo terá como espaço temporal três meses e avaliará o comportamento por horas, sendo escolhido aleatoriamente um sensor para demonstração dos valores anómalos.

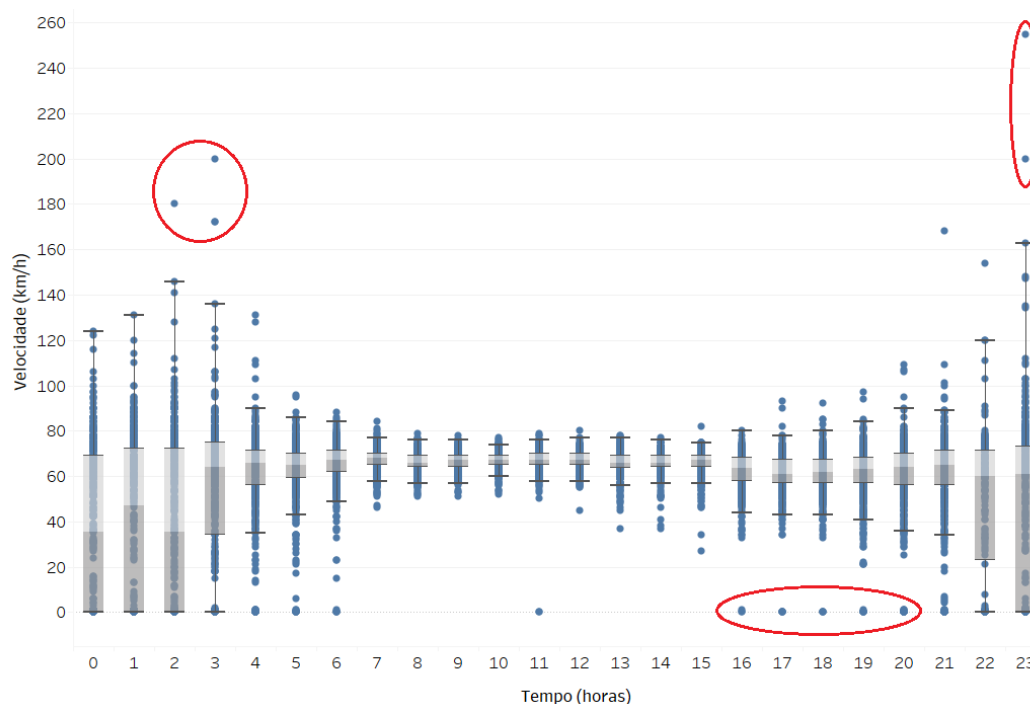


Figura 3.13: Diagrama de extremos e quartis horário da ocupação de faixa para um sensor aleatório, destacando a vermelho possíveis valores anómalos no sensor

Quando expresso o diagrama de extremos e quartis das velocidade para um sensor, ilustrado na figura 3.13, é visível, através das ovas vermelhas, a ocorrência de vários valores discrepantes, localizados fora dos extremos. Estes valores expressam a ocorrência de anomalias no comportamento do sensor, sendo necessária uma análise pormenorizada sobre estes valores, de modo a identificar-se possíveis valores anómalos no sensor.

Para uma análise pormenorizada destes valores é necessário recorrer a leitura comportamental durante a ocorrência destes valores, sendo necessário expressar uma comparação entre valores de velocidade e ocupação, permitindo uma comparação de comportamentos, de modo a averiguar a possível manifestação desta ocorrência.

Na fig.3.14 é representada a velocidade e o fluxo rodoviário para a ocorrência de um valor discrepante da velocidade. Este valor é visível no gráfico da velocidade como um pico onde é alcançada uma velocidade de 255 km/h. A manifestação deste valor apesar da sua possível ocorrência, quando analisado em conjunto com o número de veículos em circulação por hora (12 veículos), torna a sua realização pouco irreal, podendo-se assumir este valor como uma possível irregularidade ou falha no sensor.

A manifestação de valores discrepantes ou anómalos não é caraterístico apenas de atributos de velocidade, sendo também necessário verificar a existência de possíveis valores discrepantes no atributo de ocupação de via. Quando concebido o diagrama de extremos e quartis sobre a ocupação da via, representado na fig.3.15, é notado também a ocorrência de valores anormais às médias e estudos sobre a ocupação da via, mais uma vez alguns dos valores anómalos encontram-se destacados a vermelho para melhor identificação.



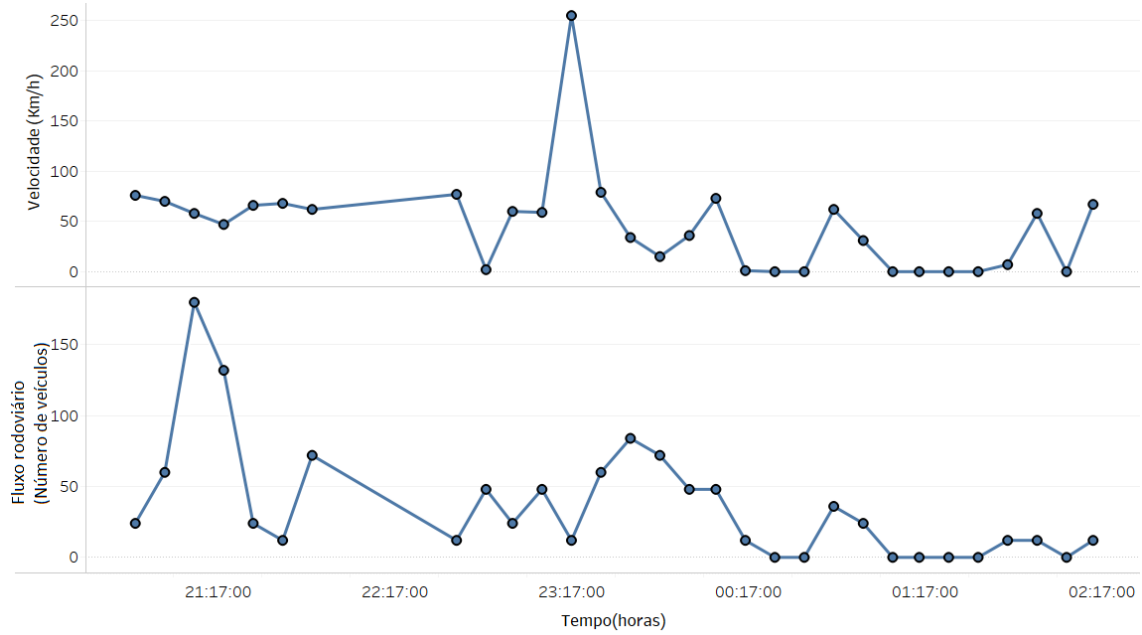


Figura 3.14: Velocidade média e ocupação da via na ocorrência de um possível valor anômalo

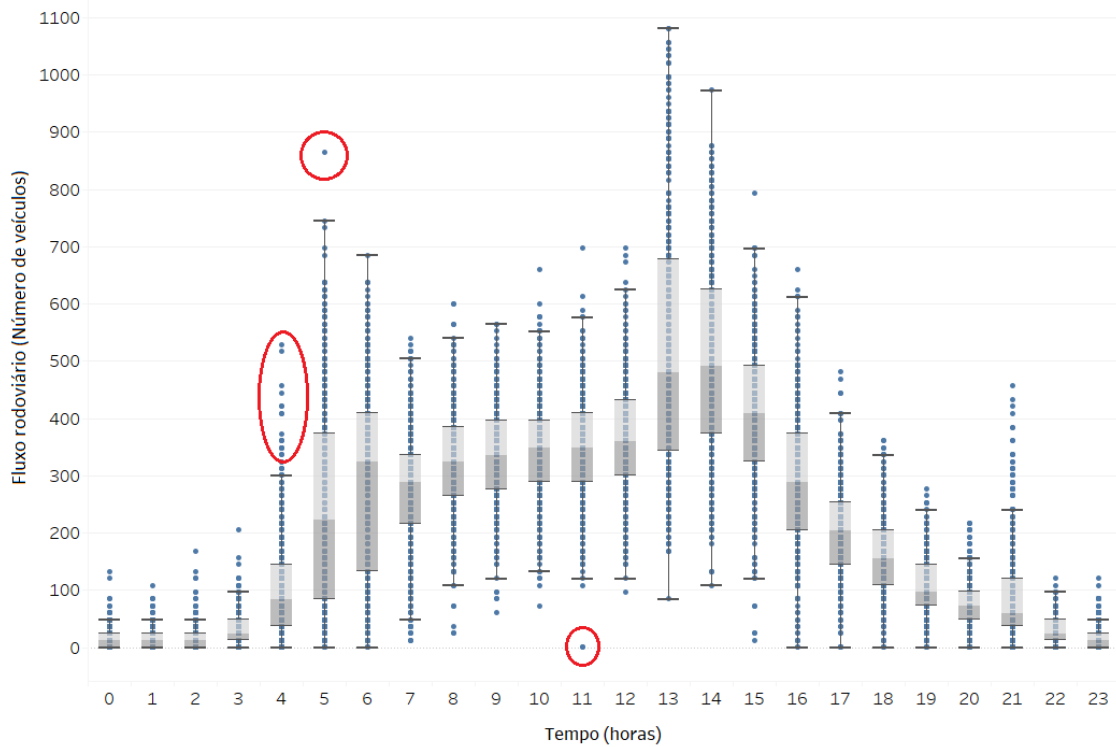


Figura 3.15: Diagrama de extremos e quartis horário da velocidade dos veículos para um sensor aleatório, destacando a vermelho possíveis valores anômalos no sensor

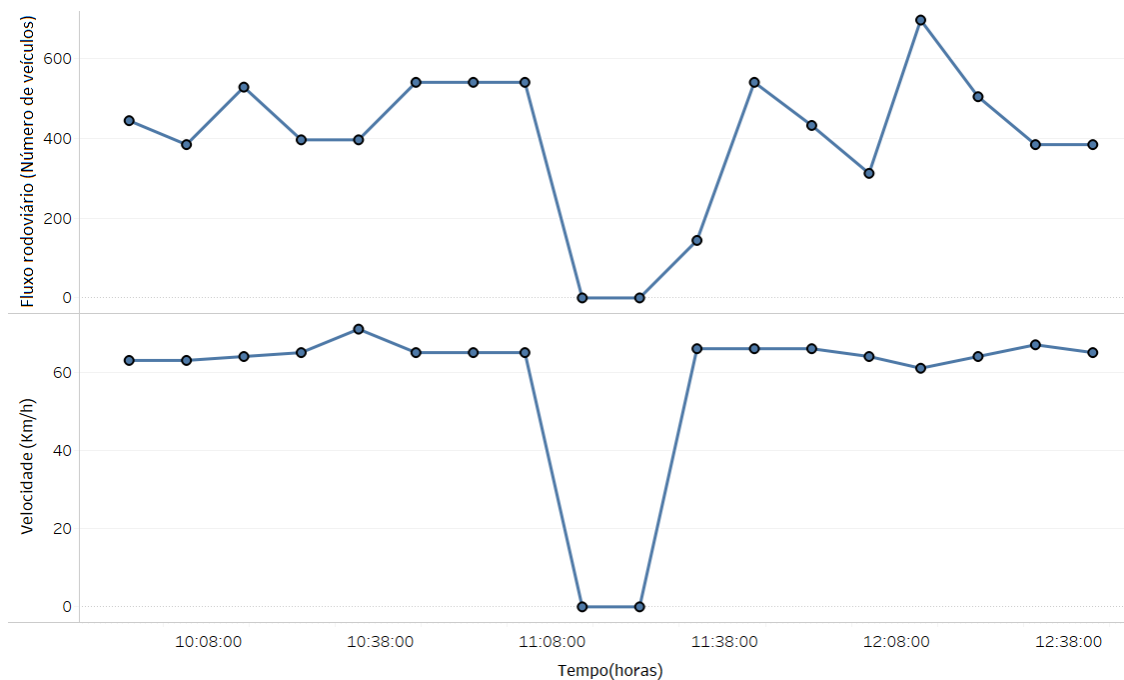


Figura 3.16: Ocupação da via e velocidade média na ocorrência de um possível valor anômalo

Mais uma vez, é analisada a velocidade e ocupação da via na ocorrência de um possível valor anômalo. Esta análise demonstra um valor nulo tanto na ocupação da via, como no atributo de velocidade, numa hora pouco viável para a ocorrência deste comportamento. Contudo, nada indica a impossibilidade de ocorrência destes valores, tempo de ser assumidos como plausíveis. A velocidade e ocupação da via expressa numa linha temporal contínua para verificação destes valores pode ser visualizada através da fig.3.16.

É de realçar, que a ocorrência destes valores discrepantes, não prevê a exclusão dos mesmos em análise ou estudos futuros, pois apesar da sua ocorrência ser pouco plausível, não foram identificadas provas concretas da impossibilidade de ocorrência deste valores, sendo esta a razão pela qual estes valores são considerados anômalos e como erros.

## PERFIL DAS AUTOESTRADAS

O perfil de uma autoestrada pode ser concebido de acordo com o comportamento do fluxo rodoviário. De seguida, no presente capítulo ir-se-á estudar e avaliar todas os fatores capazes de influenciar o comportamento do fluxo rodoviário, tais como sazonalidade, as rotinas diárias, ou as condições climatéricas e desse modo construir o perfil das autoestradas.

A caracterização do comportamento do fluxo rodoviário nas estradas precede de vários fatores, desde a localização da estrada, ao horário laboral em vigor no país, passando pela sazonalidade, à ocorrência de eventos nas autoestradas (trânsito, acidentes, obras, ...), podendo ser influenciado até pelo dia da semana (segunda, terça, ...). Todos estes fatores são decisivos no comportamento de tráfego rodoviário, sendo estes capazes de influenciar o fluxo de circulação de veículos e descrevendo o perfil de cada autoestrada.

Em seguida, e tendo em conta os fatores expostos anteriormente, examinar-se-á numa primeira fase, o comportamento diário das autoestradas, tendo em consideração a localização espacial dos sensores, estuda anteriormente na sec.3.3, observando a influência que estes demonstrem no perfil das autoestradas. À posteriori recorrer-se-á a uma análise semanal para comparação dos perfis traçados e ainda a uma análise mensal, de modo a garantir que os perfis adquiridos se manifestam ao longo dos meses e não como eventos únicos.

### 4.1 Comportamento Diário em Dias Úteis

Quando considerado o comportamento diário do fluxo rodoviário de uma autoestrada, num dia útil, é esperado que este apresente uma curva não homogénea, ou seja, deverá ostentar dois máximos de tráfego (picos), causados pelo horário laboral. É ainda de se prever, que exista uma maior afluência de tráfego nas horas que perfazem o início e o fim

do horário laboral, existindo uma quebra entre estes dois picos, sendo ainda visível uma menor utilização destas vias no período noturno. Desse modo, o comportamento esperado para dias úteis que se irá analisar, deverá assemelha-se ao apresentado na fig.4.1.

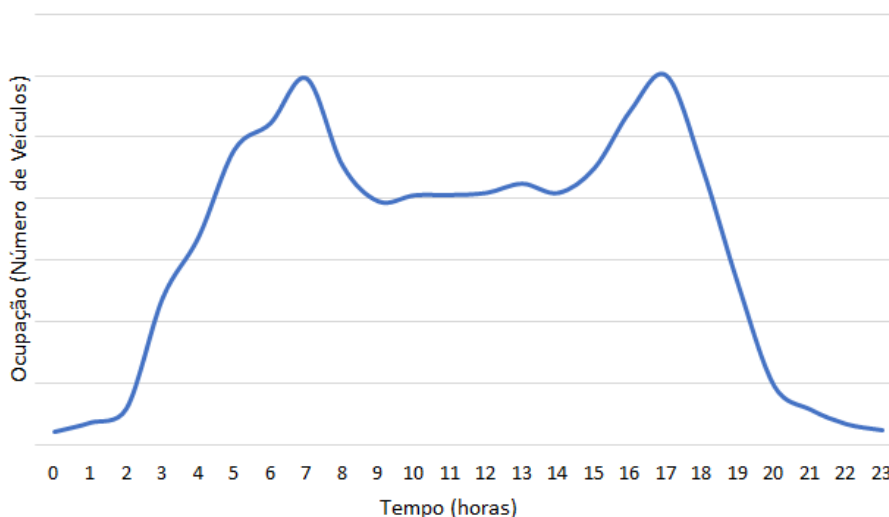


Figura 4.1: Fluxo rodoviário previsto para um dia útil

Comparando agora o fluxo previsto, desenvolvido a partir do comportamento esperado para uma autoestrada, representado pela fig.4.1, com o fluxo obtido numa autoestrada, exibido na fig.4.2, observar-se que em ambos os casos o comportamento do fluxo rodoviário é bastante semelhante, existindo um desfasamento temporal relativamente à ocorrência dos máximos, facto este que pode ser justificado pela hora laboral, ou localização do sensor, sendo de realçar também a diferença de tempos entre máximos que se esperava ser superior.

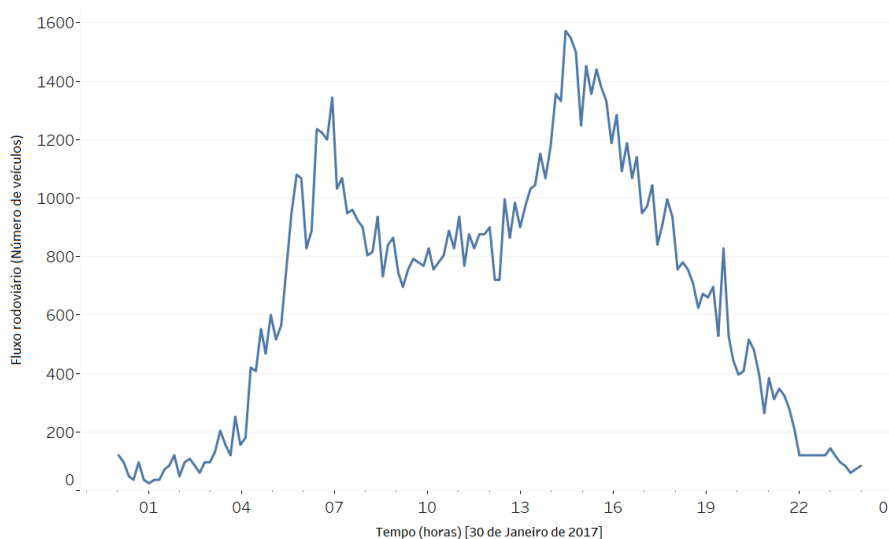


Figura 4.2: Fluxo rodoviário diário num dia útil, próximo de uma zona com grandes dimensões populacionais (sensor 832)

Contudo, a observação dos máximos é mais esclarecedora quando avaliado troços da autoestrada junto de grandes centros populacionais, como é exposto na fig.4.2, podendo por vezes o estudo desta curva ser mais difícil em sensores localizados em troços longe dos grandes centros populacionais, o que remete para outro fator influenciador de fluxo rodoviário, a localização dos sensores.

A localização é outro fator importante a considerar na análise e observação de variações de fluxo rodoviário, sendo este um fator diretamente influenciador no número de veículos a circular e ainda, um fator decisivo na criação de um perfil da autoestrada.

Desta forma é de se esperar, que sensores localizados perto de centros populacionais com um maior número de habitantes, adquiram uma curva comportamental mais saliente e mais próxima do esperado, enquanto que no oposto desta situação a curva apresentada seja mais suscetível a variações.

A razão pela qual existe uma menor aproximação do comportamento esperado relativamente ao fluxo rodoviário em zonas com menor número populacional, deve-se pelo facto de nestas zonas o número de veículos a circular em autoestradas ou próximo delas ser moderadamente menor, o que promove maiores oscilações nos padrões comportamentais adquiridos pelos sensores.

Na fig.4.3 pode-se observar a curva de fluxo rodoviário numa autoestrada localizada próximo de um centro populacional de menores dimensões, onde a curva apresentada não é tão saliente e sujeita a maiores oscilações comparativamente com a curva padrão.

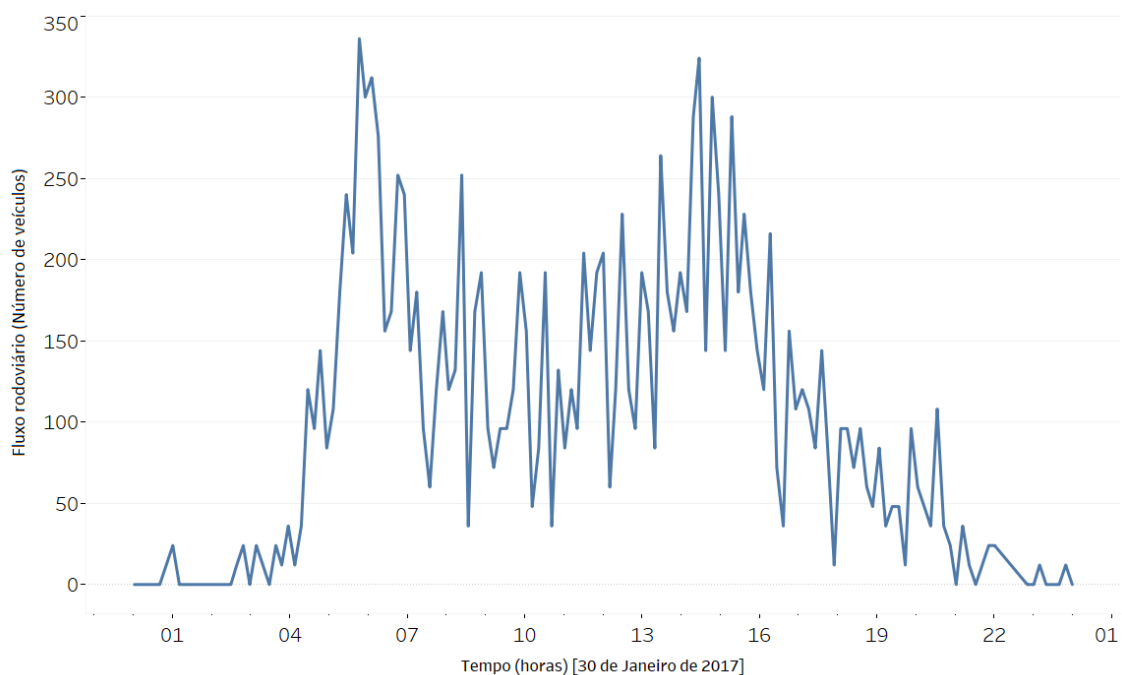


Figura 4.3: Fluxo rodoviário numa zona habitacional de menor população (sensor 258)

Até agora, foram definidos dois padrões comportamentais de acordo com a localização do sensor, tendo de ser considerado ainda um terceiro comportamento segundo uma nova localização. A utilização de três zonas distintas é esclarecido no capítulo 3, onde através da comparação da densidade populacional com a localização de eventos, foram definidas três zonas com possíveis comportamentos rodoviários distintos.

Este terceiro padrão, corresponde a sensores localizados em zonas populacionais de menor dimensão, mas localizados em autoestradas na qual permitem a ligação entre dois grandes centros populacionais. Neste terceiro caso, teremos um valor intermédio no número de carros a circular, comparativamente aos dois casos anteriores. É também neste presente caso que a curva apresenta, ostenta maiores desvios relativamente à curva padrão, sendo de difícil visualização a ocorrência dos máximos da curva. A fig.4.4 ilustra o caso descrito.

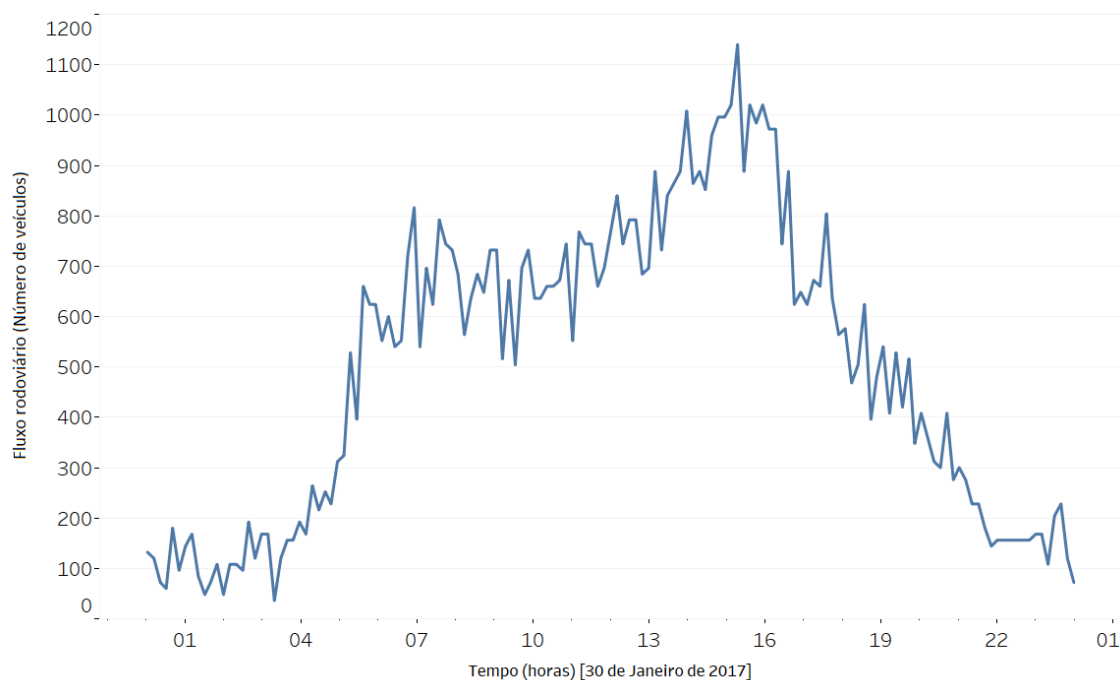


Figura 4.4: Fluxo rodoviário entre duas zonas habitacionais com elevada população (sensor 836)

Na fig.4.5 é possível observar-se a comparação entre os três padrões comportamentais dos sensores, segundo as diferentes localizações descritas. Através desta figura é ainda fácil de se observar a grande discrepância no fluxo rodoviário, representado pelo número de veículos a circular por hora.

## 4.2. COMPORTAMENTO DIÁRIO EM FINS DE SEMANA E FERIADOS

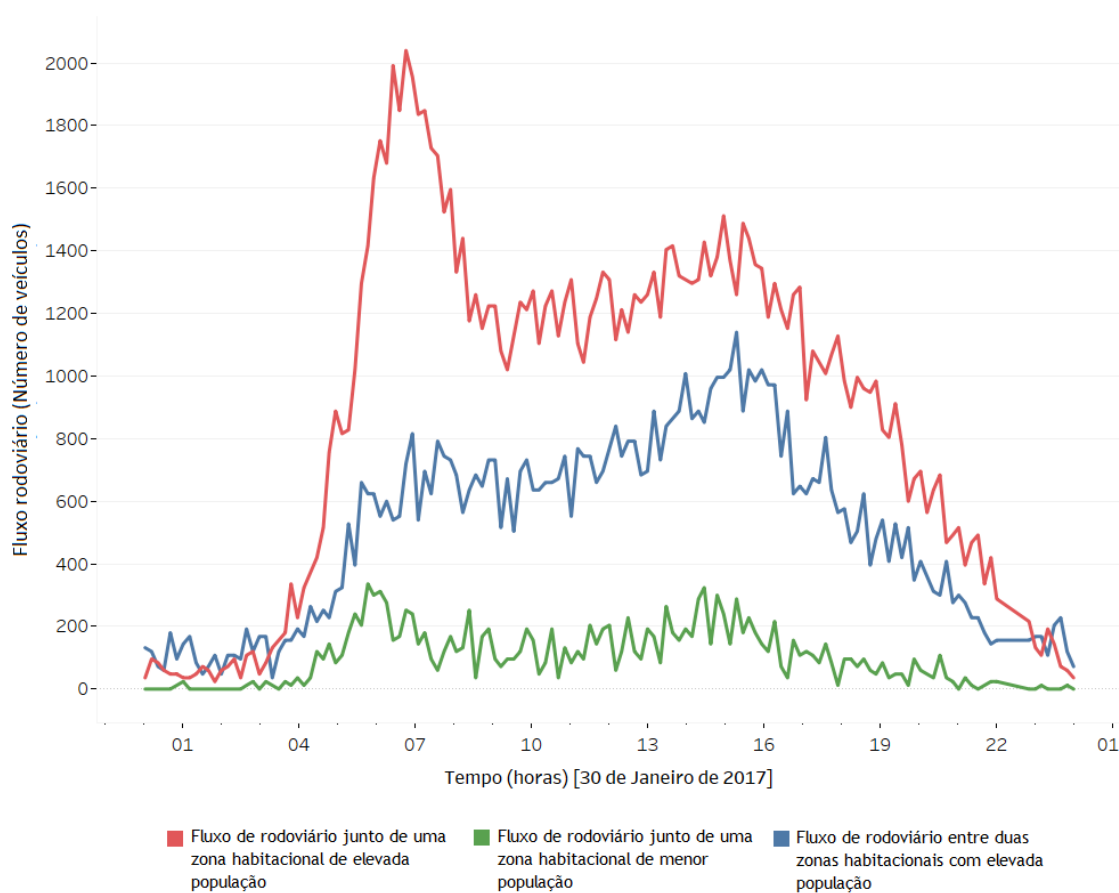


Figura 4.5: Comparação dos fluxos rodoviários em diferentes localizações num dia útil (sensores: vermelho-174, azul-836 e verde-258)

## 4.2 Comportamento Diário em Fins de Semana e Feriados

Até agora, o estudo realizado sobre o comportamento do fluxo rodoviário apenas incidiu sobre dias úteis, sendo necessário recorrer a uma posterior análise para os fins de semana e feriados. Nestes dias existe uma interrupção do "normal" horário laboral, sendo esperado uma diminuição significativa no número de carros em circulação e ainda um fluxo mais homogéneo no período diurno, não existindo a presença de picos máximos tráfego, como os visualizados anteriormente. A representação deste efeito é ilustrado na fig.4.6.

Comparando agora a curva esperada com o comportamento obtido a um sábado, os resultados obtidos são muito semelhantes, existindo uma maior homogeneidade no número de veículos a circular, no período diurno, como se pode observar pela fig.4.7.

Mais uma vez, é importante avaliar o comportamento rodoviário consoante a localização do sensor, tendo por base os três tipos de localização descrito anteriormente. Neste caso, contrariamente ao obtido na análise de dias úteis, a localização do sensor em nada influenciou a curva de fluxo rodoviário, existindo apenas o desnível no número de carros a circular em cada uma das situações como era esperado.

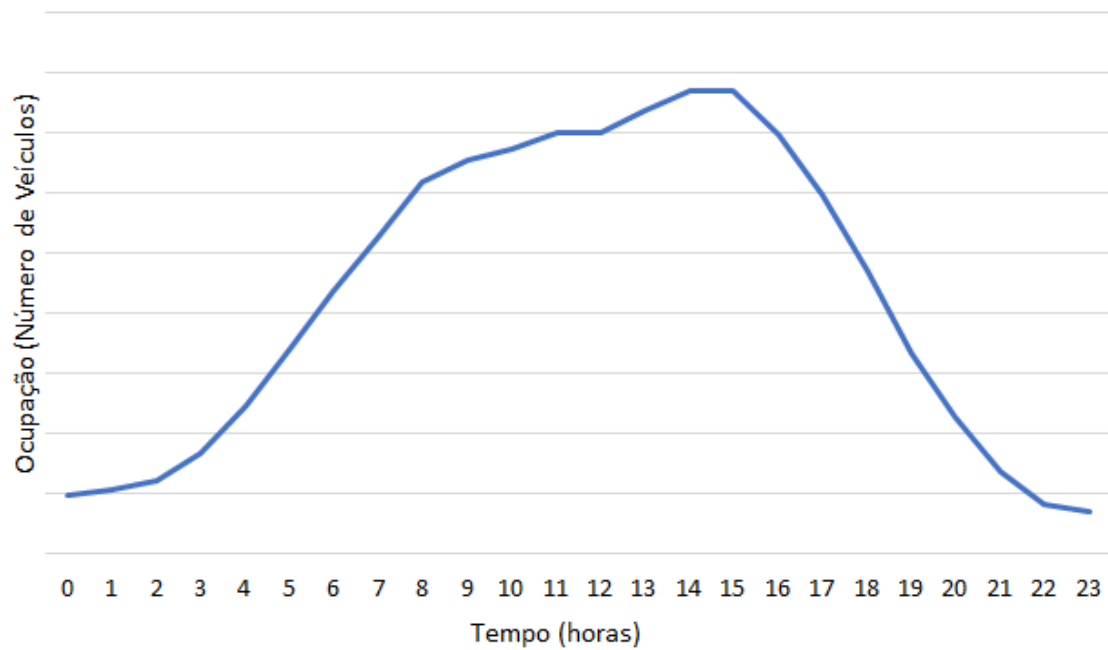


Figura 4.6: Fluxo rodoviário esperado ao fim de semana

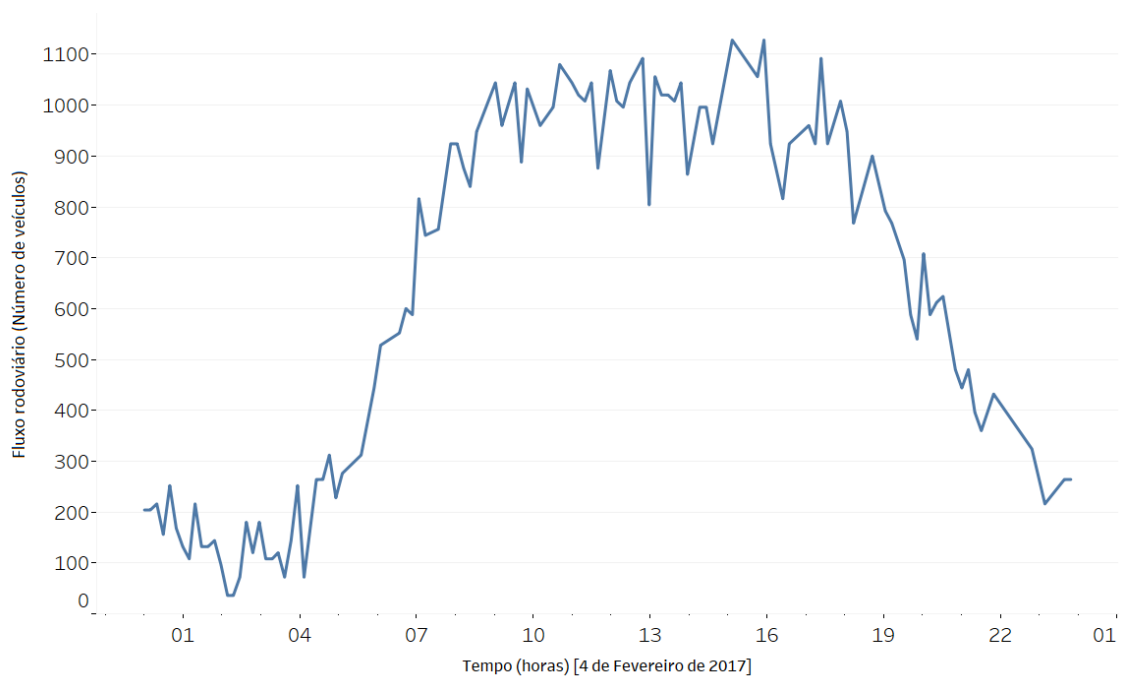


Figura 4.7: Fluxo rodoviário diário num fim de semana (sábado), perto de uma zona habitacional de elevada população (sensor 174)



## 4.2. COMPORTAMENTO DIÁRIO EM FINS DE SEMANA E FERIADOS

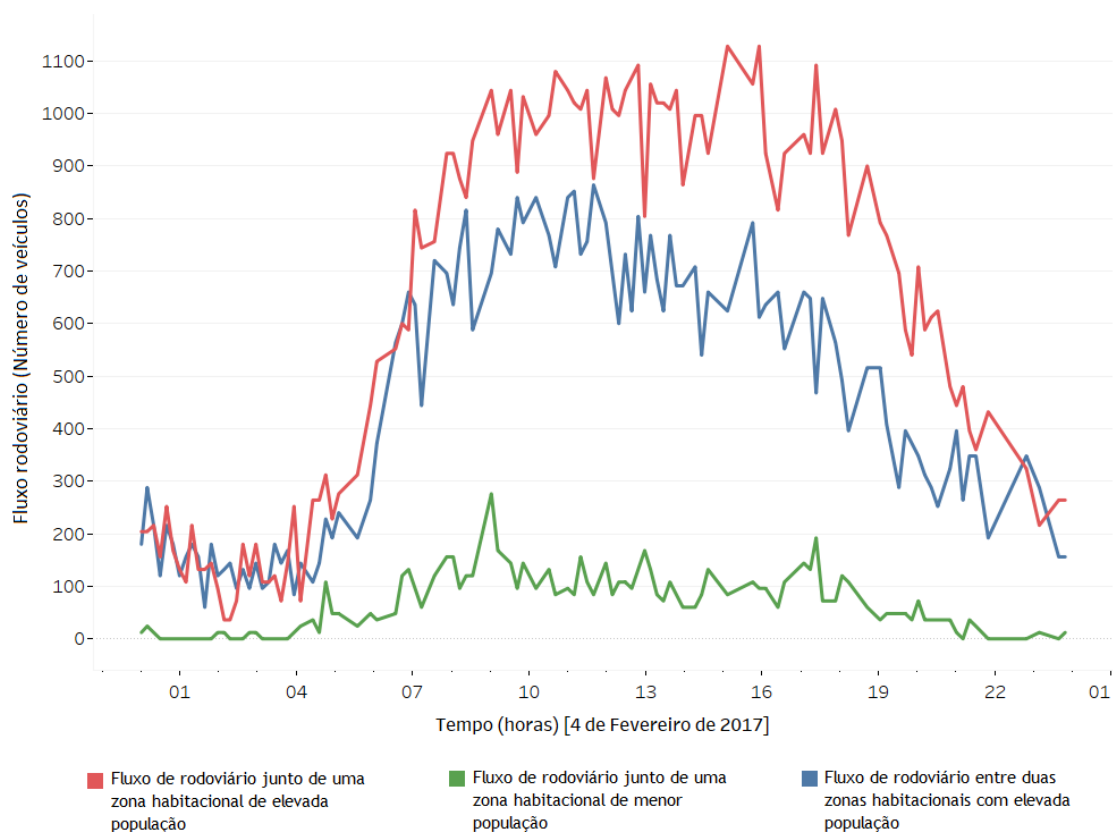


Figura 4.8: Comparação dos fluxos rodoviários em diferentes localizações, durante um fim de semana (sábado)(sensores: vermelho-174, azul-836 e verde-258)

Quando avaliado o comportamento do fluxo rodoviário num feriado, este apresenta um comportamento muito idêntico ao manifestado nos fins de semana, onde mais uma vez é observado a uniformidade no número de veículos a circular, como é ilustrado na fig.4.9, não sendo visível alturas de acréscimo de movimento (picos) ao longo do dia.

É de salientar ainda, que quando analisado o comportamento rodoviário num feriado, mais uma vez, tal como ocorre nas curvas comportamentais referentes a fins de semana, a localização dos sensores apenas apresenta influência o número de carros em circulação, obtendo-se uma curva comportamental muito idêntica independentemente da localização do sensor, como se pode observar pela fig.4.10.

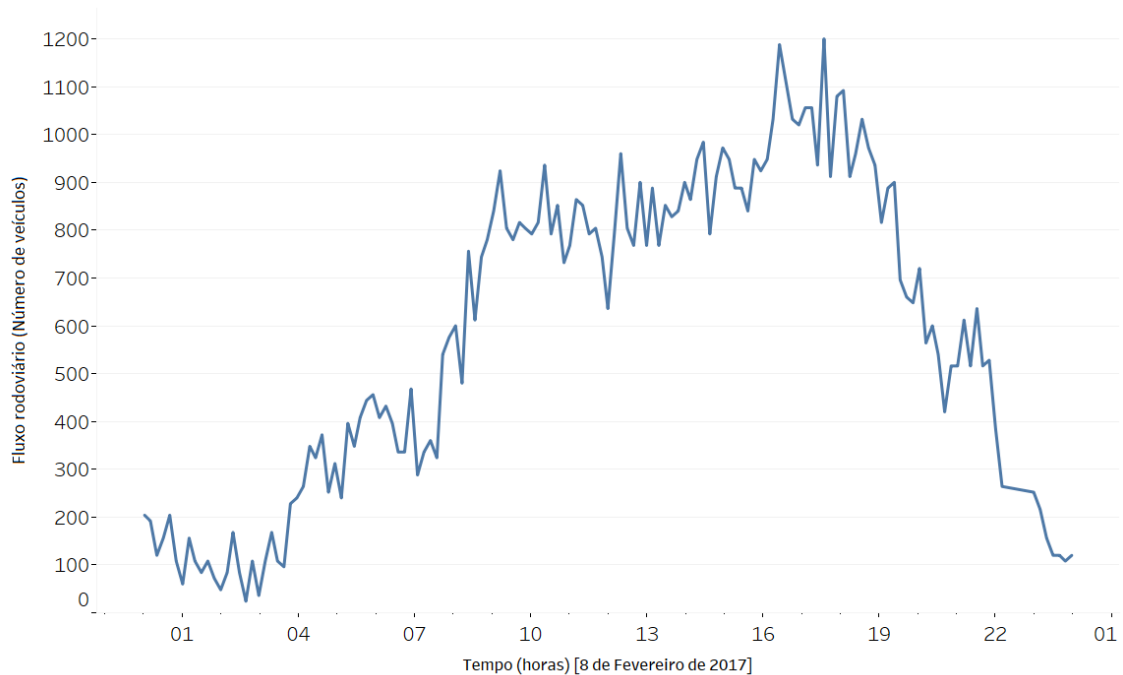


Figura 4.9: Fluxo rodoviário num feriado, próximo de uma zona habitacional de elevada população (sensor 174)

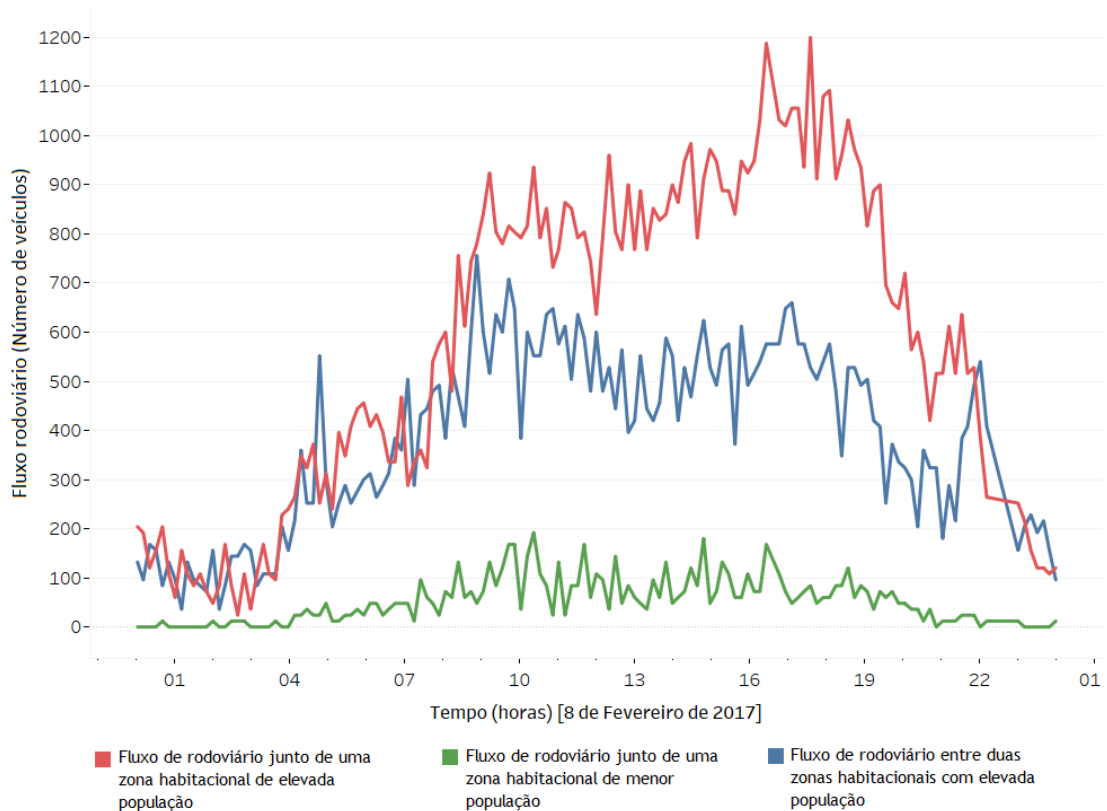


Figura 4.10: Comparação dos fluxos rodoviários em diferentes localizações num feriado (sensores: vermelho-174, azul-836 e verde-258)

### 4.3 Comparação comportamental diária de uma autoestrada entre dias úteis, feriados e fins de semana

Após realizado um estudo sobre a variação comportamental do fluxo rodoviário consoante as diferentes categorizações de dias de semana, é necessário realizar-se agora uma comparação entre estes padrões comportamentais criados, de modo a perceber-se a influência destes na variação do número médio de veículos a circular.

Numa primeira instância espera-se que exista uma quebra no número de veículos a circular, aquando da ocorrência de um feriado ou de um fim de semana, sendo também possível constatar as diferentes curvas até então analisadas.

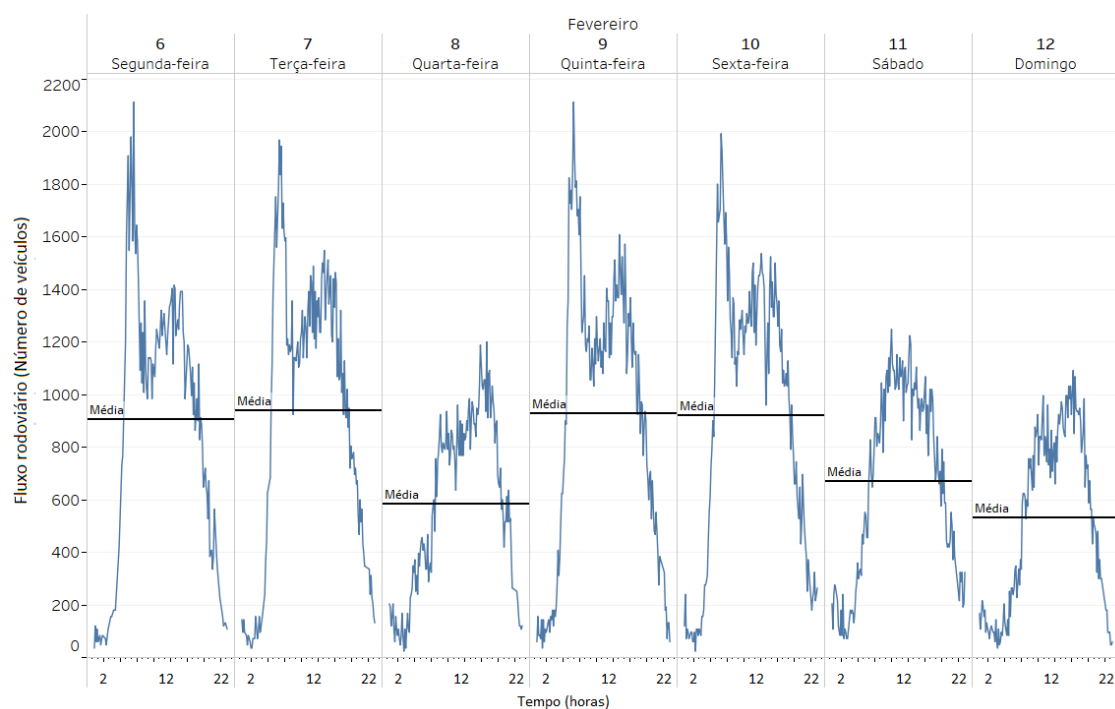


Figura 4.11: Comparação do fluxo rodoviário semanal, para um sensor perto de um grande centro populacional (sensor 174)

A fig4.11 ilustra o fluxo rodoviário ao longo de uma semana, sendo que no dia 8 de fevereiro, quarta-feira ocorre um feriado.

Ao analisar-se esta figura é possível de se constatar os padrões estudados anteriormente, onde é bastante visível a existência de máximos de fluxo nos dias úteis, sendo ainda observável um comportamento mais homogêneo nos restantes dias. Contudo a diferença que mais se destaca, é o número de carros a circular por hora, enquanto que nos dias úteis a média de veículos por hora por dia, ronda os 950 veículos, nos restantes dias esse valor apresenta uma redução de cerca de 1/3, chegando no dia com menor média a alcançar valores perto dos 580 veículos.

Mais uma vez é necessário ter em consideração a localização do sensor, sendo avaliado anteriormente apenas localizações junto de grandes centros populacionais, faltando ainda

avaliar zonas com menor população e sensores entre duas zonas populacionais grandes.

Para estes dois casos, a redução no número de veículos ainda é maior, chegando mesmo a uma redução de 1/2, como se pode observar pela fig.4.12.

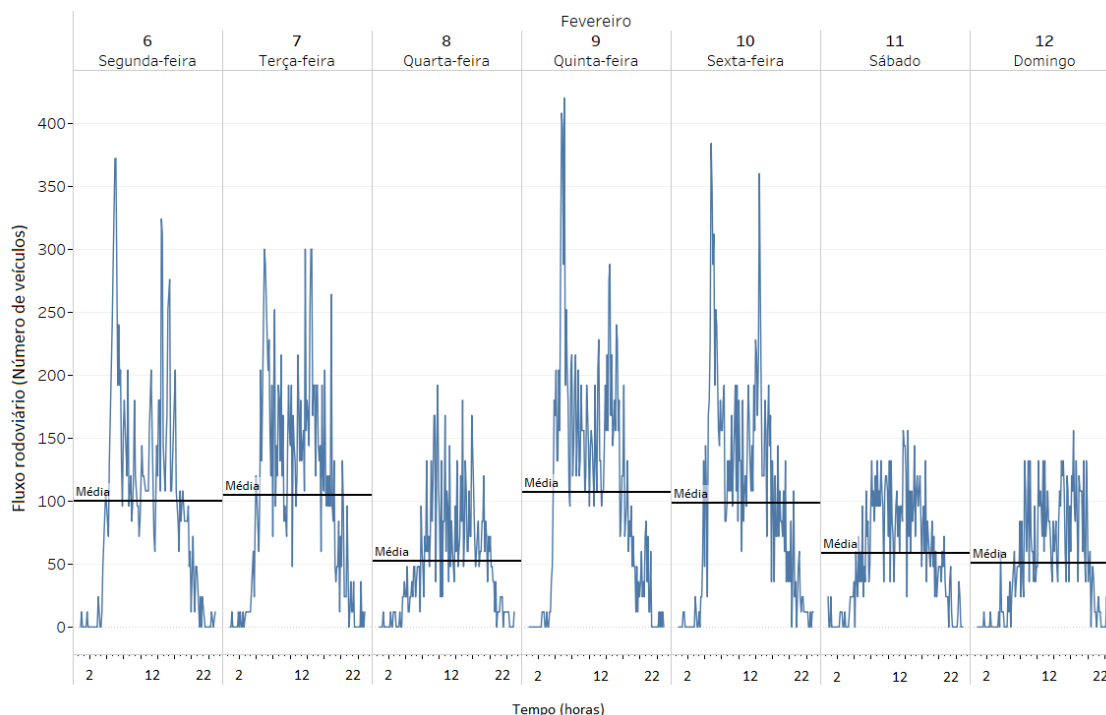


Figura 4.12: Comparação do fluxo rodoviário semanal, para um sensor próximo de uma zona populacional menor (sensor 258)

#### 4.4 Comportamento Semanal

Até ao momento, o estudo realizado sobre o comportamento do fluxo rodoviário incidiu sobre uma categorização geral dos dias, dias úteis, fins de semana e feriados, tendo sido assumido que segundo a categorização atribuída, o comportamento para estes dias se mantém. Desse modo pressupõe-se que comportamento adquirido numa segunda-feira é semelhante ao de uma quinta-feira por serem ambos categorizados como dias úteis, contudo nem sempre estas características se mantêm, sendo necessária uma análise mais profunda e detalhado sobre cada dia.

Para que seja possível analisar-se a dependência das características comportamentais rodoviárias diárias, recorreu-se a uma análise semanal do fluxo rodoviário, exposto na fig.4.13.

Como ilustrado na fig.4.13 o comportamento para os dias úteis é muito idêntico, ambos os dias descrevem a curva de fluxo esperada, tendo a amplitude máxima dos picos bem definida e apresentando-se muito idêntica no número de veículos em circulação por hora para todos os dias. É observável também pela média diária, que o número de veículos em circulação ao longo de cada dia útil é muito similar, havendo apenas uns

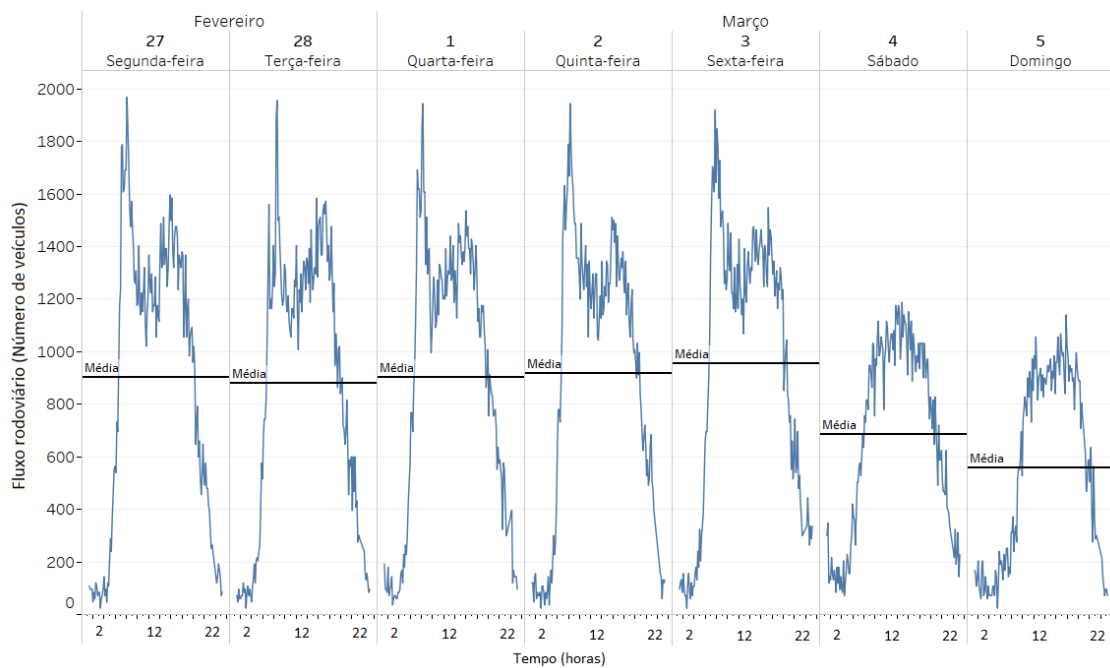


Figura 4.13: Fluxo rodoviário ao longo de uma semana, próximo de uma grande zona populacional (sensor 174)

desvios insignificantes nestes valores. Porém, quando avaliado o comportamento no fim de semana os desvios apresentados entre estes dois dias já são mais discrepantes. Apesar de a curva de fluxo manifestada ser a esperada, observam-se grandes diferenças a nível das médias, tendo o domingo uma média de número de carros a circular por hora ligeiramente inferior à apresentada para sábado, tendo de ser considerado para estudos futuros, o comportamento destes dois dias em separado.

Porém quando avaliado a presente característica para diferentes localizações dos sensores, a igualdade comportamental expressa para os dias úteis deixa de se manifestar, como é observável pela fig.4.14. Nesta figura é visível um desvio na média do número de veículos em circulação por hora nos dias úteis, apesar de a curva de fluxo apresentada ser a mesma.

Desta forma, todas as análises e criação de padrões comportamentais futuros, independentemente da localização do sensor, terão em consideração o dia da semana e não a categorização anteriormente utilizada. Apesar de os padrões de curvas de fluxo estudados serem corretos, quando avaliado em pormenor o número de carros em circulação por hora, estes valores apresentam uma maior dependência consoante o dia da semana, executando-se uma análise mais fidedigna diferenciando-se os vários dias.

A figura seguinte, 4.15, comprova que a consideração de aproximações comportamentais por dia de semana é mantido com menos oscilações que a aproximação comportamental por dia útil.

## CAPÍTULO 4. PERFIL DAS AUTOESTRADAS

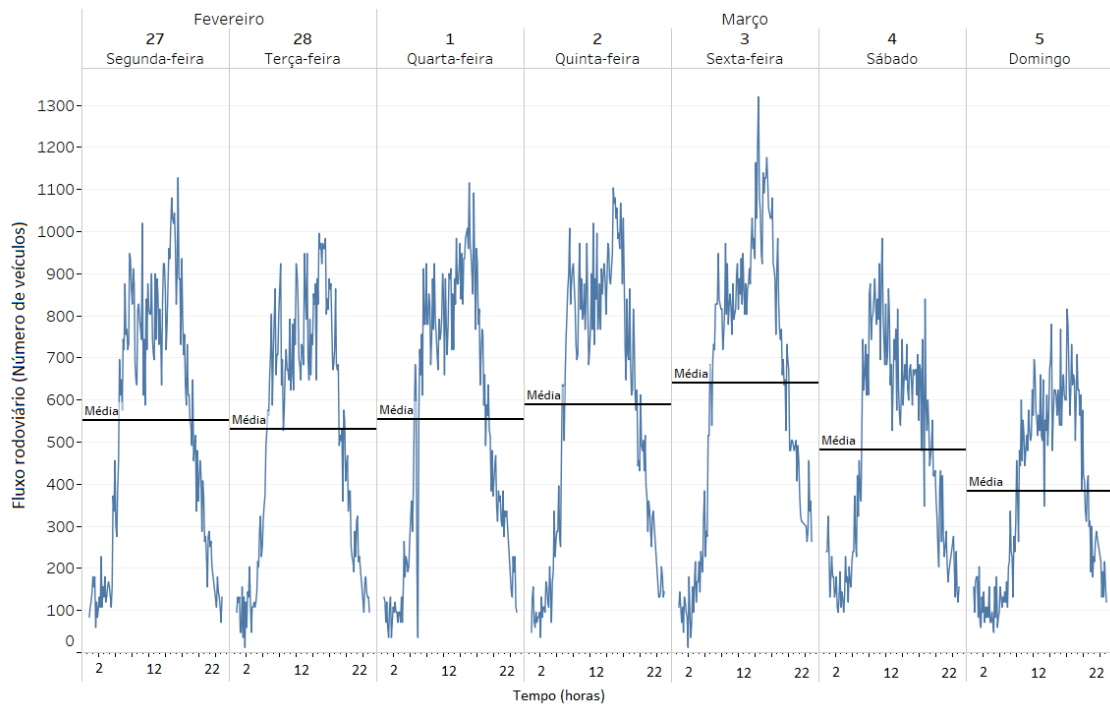


Figura 4.14: Fluxo rodoviário ao longo de uma semana, entre duas zonas habitacionais com elevada população (sensor 836)

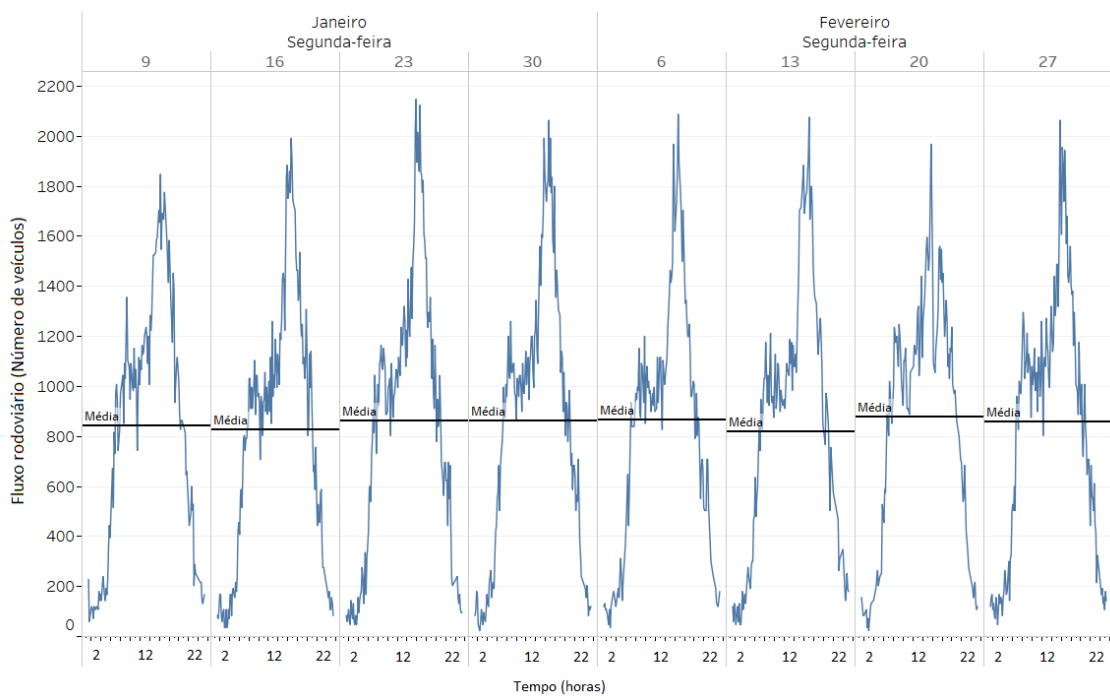


Figura 4.15: Fluxo rodoviário durante todas as segundas feiras nos meses janeiro e fevereiro (sensor 199)

## 4.5 Comportamento Sazonal

Quando referido no início deste capítulo 4 os fatores capazes de influenciar o fluxo rodoviário, a sazonalidade foi um dos fatores apontados, ou seja, eventos sazonais como as férias de verão, ou férias da Páscoa são eventos capazes de gerar grande influência no "normal" fluxo rodoviário.

Estes períodos são caracterizados por uma avalanche migratória de pessoas, proporcionando uma quebra das rotinas nas grandes e pequenas cidades, passando a existir um aumento do número normal de pessoas junto de zonas balneares ou no litoral. Todos estes fatores são fatores a ter em consideração na avaliação do comportamento do fluxo rodoviário, esperando-se nestas alturas um acréscimo do número médio de veículos por hora em determinadas zonas, enquanto que noutros locais, este número pode diminuir.

Desta forma, para a avaliação deste fator ter-se-á em consideração três localizações diferentes, de forma a que se possa avaliar a influência que estes períodos temporais apresentam sobre o comportamento do fluxo rodoviário. Para tal, serão analisados sensores localizados perto de grandes centros populacionais interiores, sensores juntos a zonas balneares ou litorais e sensores próximos de centros populacionais de menores dimensões, tendo sempre em consideração o país estudado.

Na seguinte avaliação será ainda considerada a forma da curva de fluxo, o número médio de carros a circular por hora e a amplitude máxima dos máximos (picos) da curva de fluxo, sempre que estes existam.

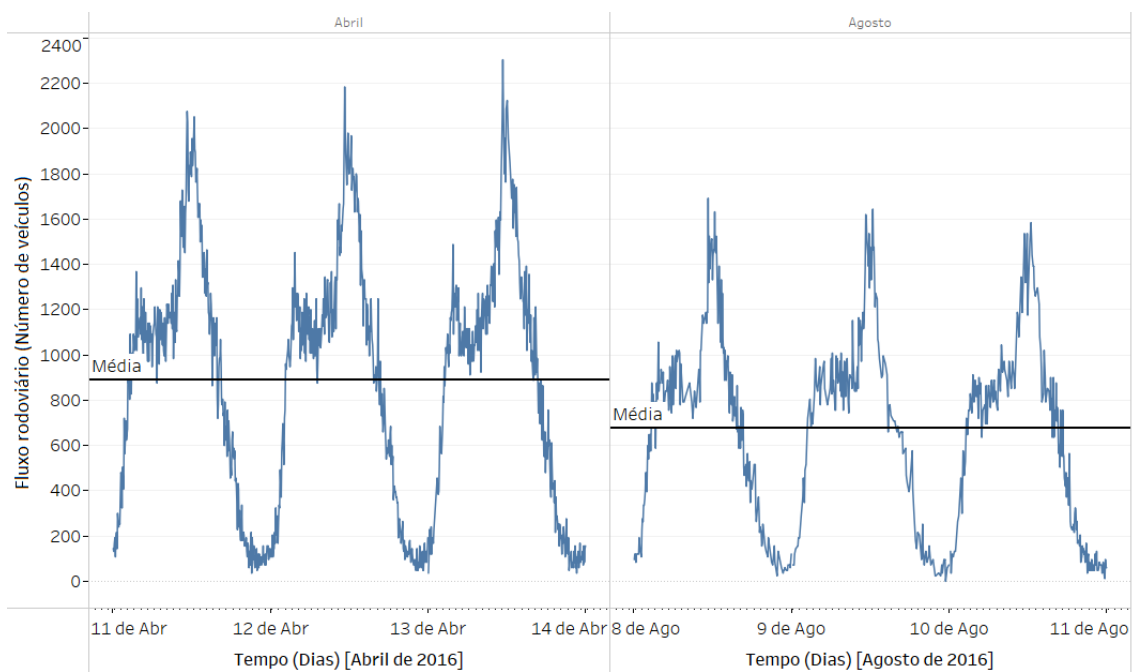


Figura 4.16: Comparação do fluxo de rodoviário entre três dias em abril e agosto, para um sensor localizado próximo de um grande centro populacional (sensor 199)

No fluxo rodoviário junto de um grande centro populacional, é visível uma pequena

diminuição no número médio de carros a circular, sendo mantido a forma da curva esperada para os dias estudados. Contudo a diferença que mais se destaca, corresponde à amplitude dos picos da curva, onde é notório uma redução elevada nos valores máximos de amplitude da curva.

Para o comportamento próximo de um centro populacional de menores dimensões, os resultados obtidos são muito idênticos aos anteriormente descritos, apesar de não se constatar uma diminuição tão significativa no número médio de carros a circular, mas mais uma vez é visível uma redução considerável na amplitude máxima dos picos na curva de fluxo, sendo a forma da curva mantida. Este comportamento é exposto na fig.4.17.

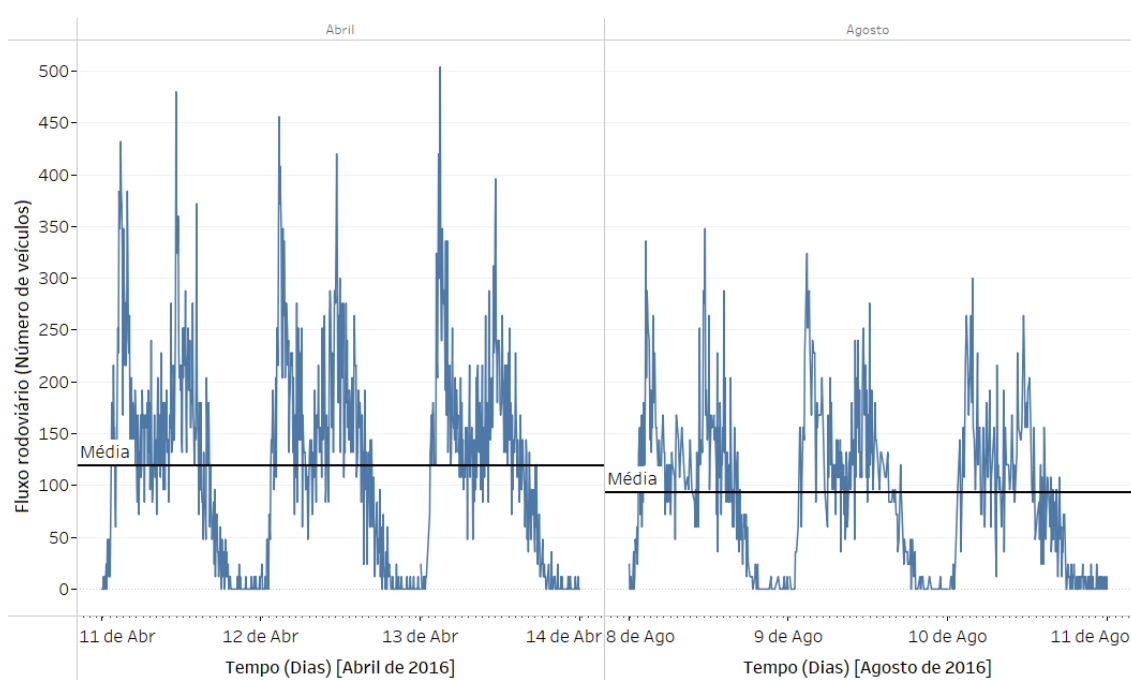


Figura 4.17: Comparação do fluxo de rodoviário entre três dias em abril e agosto, para um sensor localizado próximo de um pequeno centro populacional (sensor 258)

Até ao momento, apenas se têm assistido a uma diminuição dos valores médios de circulação dos veículos e numa grande redução da amplitude dos picos da curva de fluxo. Todavia, junto de zonas litorais ou balneares este comportamento não se verifica.

Quando analisado as curvas comportamentais de fluxo rodoviário nestas zonas, é observável um aumento significativo no número de carros a circular por hora, constatando-se ainda uma mudança na curva de fluxo, torna-se esta mais homogênea mesmo em dias de semana, perdendo-se ligeiramente a formação de picos na curva, como ilustrado na fig.4.18.

Esta modificação no comportamento rodoviário é justificável pela elevada migração de pessoas para estes locais, nestas alturas específicas do ano, onde existe um comportamento horário diferente do habitual.



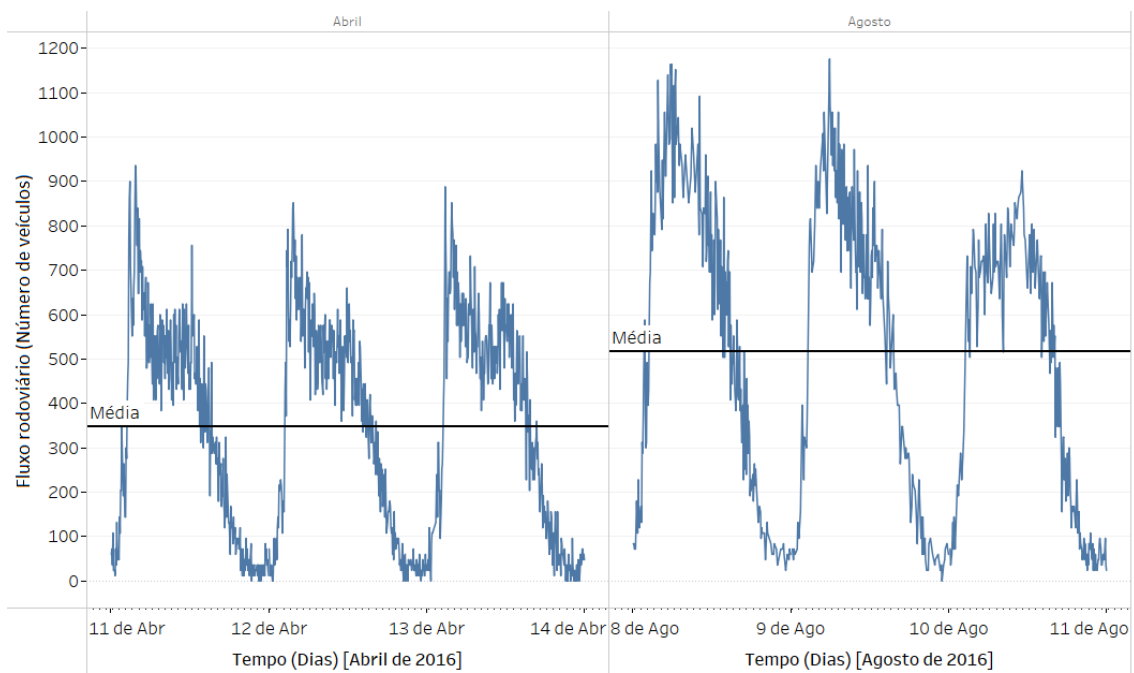


Figura 4.18: Comparação do fluxo de rodoviário entre três dias em abril e agosto, para um sensor próximo a uma zona litoral (sensor 803)

Após se analisar o perfil das autoestradas, pode-se concluir que os padrões comportamentais das estradas são influenciados por diversos fatores, desde o horário laboral, localização, passando pela sazonalidade e até mesmo pelo dia da semana em estudo.

A necessidade de se realizar esta prévia análise do perfil das autoestradas deve-se ao facto de, para se desenvolver uma aplicação de deteção de eventos anómalos em autoestradas, ter de existir padrões comportamentais comparativos de forma a detetar anormalidades ao normal funcionamento do sistema rodoviário.

Desse modo, este estudo permitiu decidir em que moldes iriam assentar os padrões comportamentais, estando a criação deste padrões estipulado pelo dia da semana (segunda, terça, ...), pois o estudo revelou que a maior semelhança no comportamento rodoviário em autoestradas reflete-se pelo dia da semana.



## DETEÇÃO DE EVENTOS EM AUTOESTRADAS

Após realizada a seleção e exploração dos dados nos capítulos anteriores, permitindo adquirir conhecimento sobre o comportamento rodoviário ao longo de diversas autoestradas, no presente capítulo procede-se-á a uma breve introdução e caracterização de eventos em autoestradas, sendo ainda abordado o processamento dos dados selecionados. Serão abordados todos os processos, regras e métodos utilizados na construção da aplicação de deteção automática de eventos anómalos em autoestradas.

### 5.1 Comportamento de Eventos em Autoestradas

A ocorrência de eventos no sistema rodoviário influencia o normal funcionamento das autoestradas, criando alterações ao normal comportamento rodoviário. Antes de se proceder a modelação e descrição dos processos utilizados na criação de uma aplicação CEP para identificação de eventos anómalos/inesperados em autoestradas, abordar-se-á em seguida a identificação e comparação dos eventos anómalos detetados nas autoestradas eslovenas.

Dependo da ocorrência dos eventos (transito, acidentes, reparações na via, ...) este pode tomar diferentes variações no comportamento rodoviário, influenciando de diferentes formas a ocupação da via, bem como a velocidade dos veículos. Para um estudo mais preciso, foram apenas considerados eventos de curta duração (acidentes e trânsito), não sendo do interesse da presente dissertação identificar eventos previamente conhecidos, como reparações nas vias, pois estes são eventos agendados pelos concessionários das autoestradas, sendo estes por norma publicitados previamente, considerados assim eventos esperados. Focando a identificação de eventos inesperados como acidentes e trânsito, pois estes apresentam um maior impacto, principalmente quando se está a processar dados em tempo-real, como é objetivo da aplicação idealizada.

### 5.1.1 Caraterização de eventos

Todo o processo de identificação de eventos inesperados, foi um processo automático tendo sido utilizados eventos na base de dados de eventos eslovenos. De modo a abranger apenas os eventos de curta duração, foi utilizado um tempo máximo de 5 horas por evento, sendo apenas detetados eventos classificados como trânsito lento e acidentes, que serão caraterizados em seguida.

A ocorrência deste tipo de eventos denota-se especialmente por uma diminuição da velocidade, como se pode observar pelas figuras 5.2 e 5.1, sendo a variação de ocupação da via um critério mais difícil de avaliar, pois consoante a localização dos sensores, existem variações diferentes deste comportamento. Contudo, quando se compara a ocorrência de eventos do tipo trânsito (exemplo na fig.5.1) com eventos do tipo acidente (exemplo na fig.5.2, as diferenças entre estes não são esclarecedoras, não se verificando qualquer diferença que permita, distinguir a identificação deste dois tipos de eventos na criação regras. Este fato pode ser justificado pela ocorrência de um evento do tipo acidente, gerar à posterior, normalmente um evento do tipo trânsito, estando ambos os eventos relacionados.

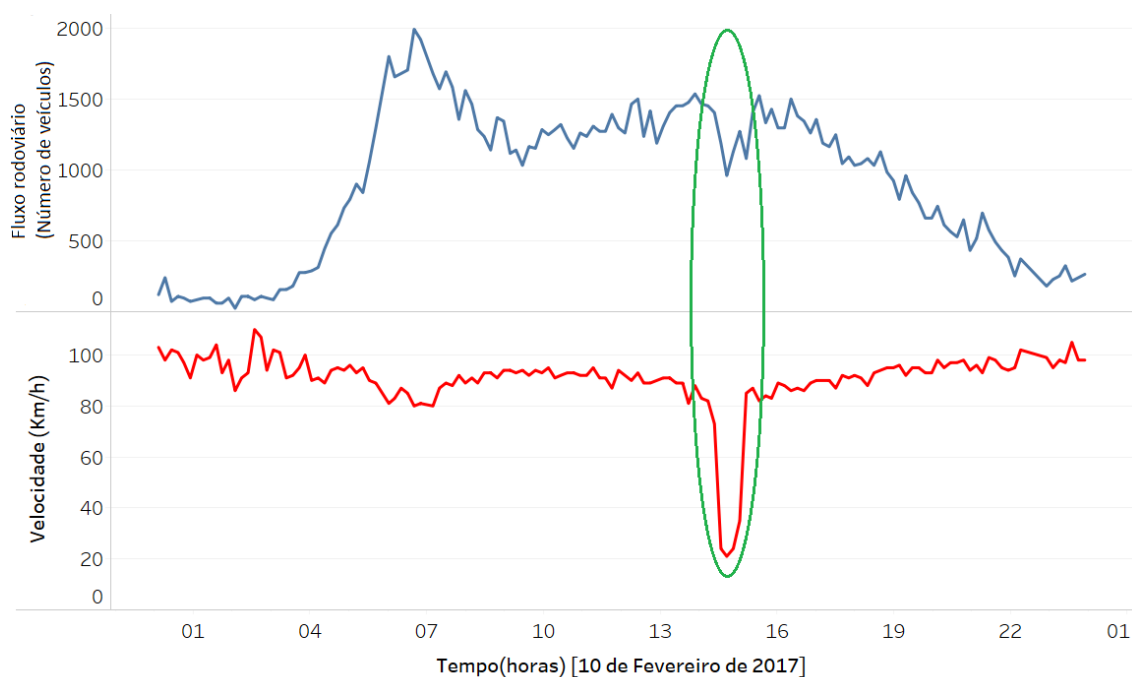


Figura 5.1: Evento anómalo (trânsito) para o sensor 174

Da análise comportamental deste dois tipos de eventos, acidente e trânsito, exemplos ilustrados nas figuras 5.2 e 5.1 respetivamente, conclui-se que existe uma enorme quebra na velocidade média dos veículos e um quebra ligeira no número de veículos a circular por hora, sendo à posteriori este um dos princípios para a criação de regras para deteção de eventos anómalos em autoestradas.

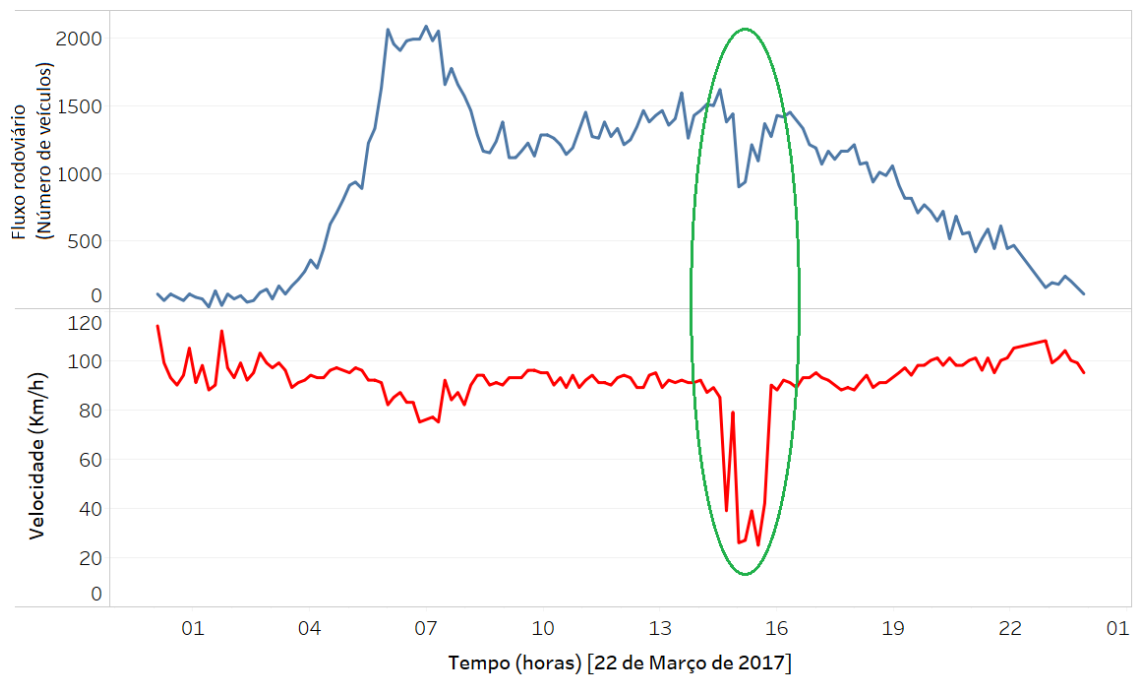


Figura 5.2: Evento anómalo (acidente) para o sensor 174

Em seguida, após realizada uma breve introdução relativamente a identificação de eventos em autoestradas, abordar-se-à todas a ferramentas e processos necessários para conceber uma aplicação capaz de detetar eventos anómalos em autoestradas de forma pro-ativa.

## 5.2 Modelação

Antes da criação de regras que permitam a identificação automática de situações anómalas ao normal comportamento rodoviário, tem que se compreender o funcionamento por detrás da ferramenta CEP utilizada, que para a presente dissertação será a ferramenta WSO2 CEP, como explicito na secção 2.2.5.

A ferramenta WSO2 CEP, é constituída por quatro grandes blocos, como ilustrado na figura 5.3, sendo estes os recetores de eventos, os *stream* de eventos, os divulgadores de eventos e por fim os processadores de eventos.

### 5.2.1 Recetores de eventos

Os recetores de eventos, tal e qual como o seu nome indica são os recetores de dados, ou seja, a primeira entrada de dados no nosso sistemas. Estes podem receber dados sobre diversos formatos, bem como, podem ser utilizados diversos modos de comunicação para receção dos dados na ferramenta, desde a utilização da base de dados à utilização de protocolos de comunicação.

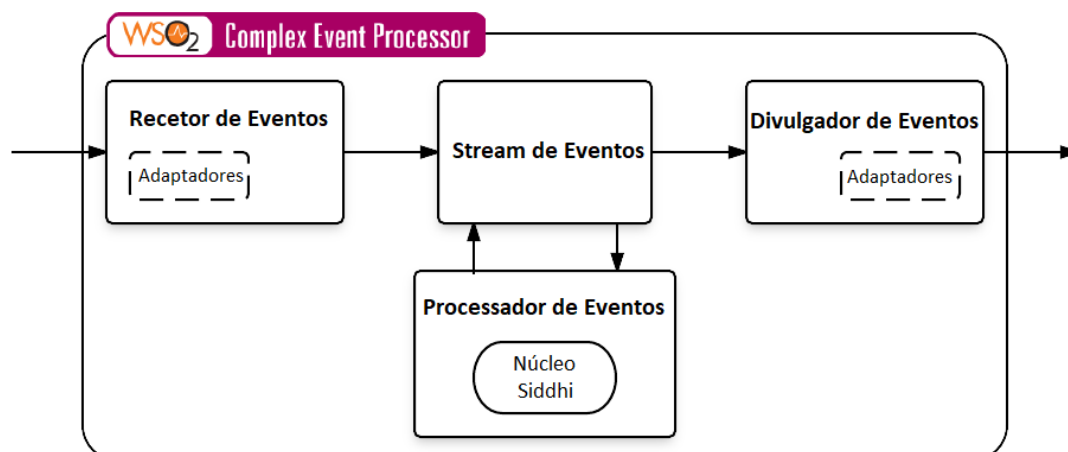


Figura 5.3: Diagrama de blocos geral da ferramenta WSO2 CEP [51]

É de reforçar que tanto o formato do ficheiro como os diversos modos de comunicação com a ferramenta são de livre escolha do utilizador, contudo quando enviados os dados se a formatação apresentada não for a pré-estabelecida ou existir uma mínima mudança na sua formatação, o sistema CEP não irá executar qualquer ação a partir dessa deteção.

### 5.2.2 *Stream de Eventos*

O *stream* de eventos representa a ligação entre os vários blocos do circuito, contendo fluxos exclusivos de dados, podendo existir já processamento sobre estes dados, mas nunca executado nestes blocos.

No presente trabalho os blocos desta categoria utilizados, apenas serviram para ligações entre os restantes blocos, pois a ferramenta WSO2 CEP só permite a interação de blocos através de *stream* de eventos. Mais a frente secção 5.2.5 pode ser observado a interação destes blocos.

### 5.2.3 *Processador de Eventos*

O processador de eventos é o bloco responsável pelo processamento de dados na ferramenta CEP. Estes blocos através de planos de execução processam os dados com base em várias operações lógicas.

Conhecido como o núcleo de operações da ferramenta WSO2 CEP, é nele que são criadas todas as regras capazes de detetar eventos em qualquer aplicação CEP, este é capaz de desencadear novos eventos através da comparação de várias fontes de dados, ou seja, é no processador de eventos que se podem realizar ações sobre *streams* de conjuntos de dados diferentes.

Normalmente, no interior deste bloco existe outro diagrama de blocos que descreve todas as operações realizadas no mesmo, para mais informações sobre os blocos e as regras criadas no desenvolvimento deste bloco consultar a sec.5.3.

### 5.2.4 Divulgador de Eventos

Este bloco é responsável pela extração de resultados da ferramenta WSO2 CEP. O Divulgador de eventos armazena em base de dados, ou transmite para um visualizador, todas as informações enviadas pela *stream* de eventos diretamente ligada a este bloco. Este bloco apenas funciona como saída da ferramenta CEP.

### 5.2.5 Visão Geral

Posteriormente após se compreender a função de cada bloco analisar-se-à especificamente cada bloco utilizado na aplicação de deteção automática de eventos. Desde já, é ilustrado na fig.5.4 o diagrama de blocos que compõe a parte responsável pela deteção de eventos.

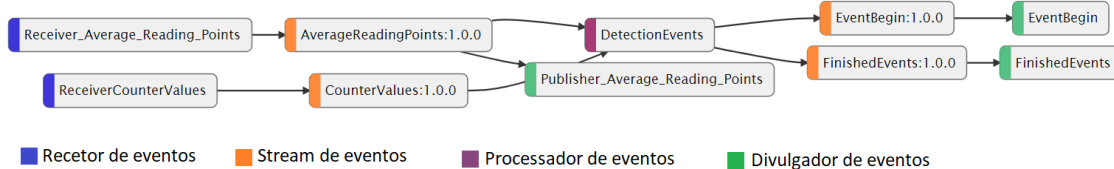


Figura 5.4: Diagrama de blocos da aplicação de deteção pró-ativa de eventos

Como é observável e descrito na imagem anterior, a aplicação apresentada é constituída por dois recetores de eventos, quatro *streams* de eventos, um processador de eventos e por três divulgadores de eventos.

Numa primeira fase, a aquisição de dados por parte da aplicação, é realizada pelos recetores de eventos. No presente trabalho foram utilizados dois recetores, um responsável por receber inicialmente e uma única vez dados referentes a médias horárias por dia da semana das autoestradas, previamente calculadas com base nos registos dos sensores, denominado "Receiver\_Average\_Reading\_Points" e outro recetor designado por "Receiver-CounterValues", responsável pela receção contínua dos dados dos sensores, sendo possível neste recetor trabalhar com a aquisição de dados em tempo-real ou apenas com intervalos de tempo predefinidos. Para ambos os recetores, foi usado o protocolo de comunicação HTTP, sendo utilizados ficheiros .txt previamente formatados com os dados. As tabelas com a descrição da formatação dos dados pode ser consultadas em seguida nas tab.5.1 e 5.2.

Posteriormente à receção dos dados, foram utilizados quatro *streams* de eventos, onde estes na presente aplicação apenas serviam para interligação entre os vários blocos do sistema.

Tabela 5.1: Descrição da formatação dos dados recebidos no bloco "Receiver\_Average\_Reading\_Points"

Dados	Descrição
Id	Identificador único do sensor
Road	Identificador da faixa
dayOfTheWeek	Dia da semana, expresso por inteiros (1-Domingo, 2-Segunda,...)
hour	Hora
velocity_data0	Velocidade média dos veículos para a faixa 0 (mais direita), expresso em km/h
velocity_data1	Velocidade média dos veículos para a faixa 1, expresso em km/h
ocupacation_data0	Ocupação média da faixa 0 na última hora, expresso em número de veículos
ocupacation_data1	Ocupação média da faixa 1 na última hora, expresso em número de veículos

Tabela 5.2: Descrição da formatação dos dados recebidos no bloco "ReceiverCounterValues"

Dados	Descrição
ModifiedTime	Descrição da data da recolha de dados, sobre o formato de DD MMM YYYY hh mm ss, exemplo: 28 fev 2017 22 54 21
Day	Dia do mês
Hour	Hora
Id	Identificador único do sensor que realizou a recolha
Hit_11	Velocidade média para a faixa 1 e via 1 (mais à direita), expresso em km/h
Hit_12	Velocidade média para a faixa 1 e via 2, expresso em km/h
Hit_21	Velocidade média para a faixa 2 e via 1, expresso em km/h
Hit_22	Velocidade média para a faixa 2 e via 2, expresso em km/h
Occu_11	Ocupação média para a faixa 1 e via 1, expresso em número de veículos
Occu_12	Ocupação média para a faixa 1 e via 2, expresso em número de veículos
Occu_21	Ocupação média para a faixa 2 e via 1, expresso em número de veículos
Occu_22	Ocupação média para a faixa 2 e via 2, expresso em número de veículos



Mais uma vez, como é observável pela fig.5.4, a aplicação apresenta ainda um processador de eventos e três divulgadores de eventos. O processador de eventos não será abordado nesta secção, para mais informações consultar a secção 5.3. Relativamente aos divulgadores de eventos, dois são responsáveis pela extração dos resultados finais da aplicação, após processamento e identificação de eventos, sendo um responsável pela extração da identificação do início de novos eventos, enquanto que o outro é responsável pela extração da identificação de término de eventos, sendo denominados respetivamente de "EventBegin" e "FinishedEvents". Quanto ao terceiro divulgador este tem apenas características informativas, denominado de "Publisher\_Average\_Reading\_Points", apenas é utilizado para garantir que as médias dos vários sensores foram introduzidas corretamente na aplicação.

### 5.3 Aplicação de Regras

Até agora apenas foram especificados blocos gerais de entrada, saída e transporte de dados, não tendo sido referenciado qualquer operação lógica sobre os dados.

O bloco responsável pela realização de operações lógicas sobre os dados é o bloco processador de eventos, é neste que são executadas todas as operações sobre os conjuntos de dados e são identificados os eventos pretendidos. Sendo este um bloco muito particular, pode ser expresso sobre a forma de um diagrama de blocos, como o ilustrado pelas fig.5.5 e fig.5.6.

Não esquecendo o objetivo principal da presente dissertação de criar uma aplicação capaz de detetar de forma pró-ativa eventos anómalos em autoestradas, é nesta presente secção que irão ser abordados todas as técnicas e operações de identificação de eventos anómalos, sendo o bloco processador de eventos o responsável por esta identificação.

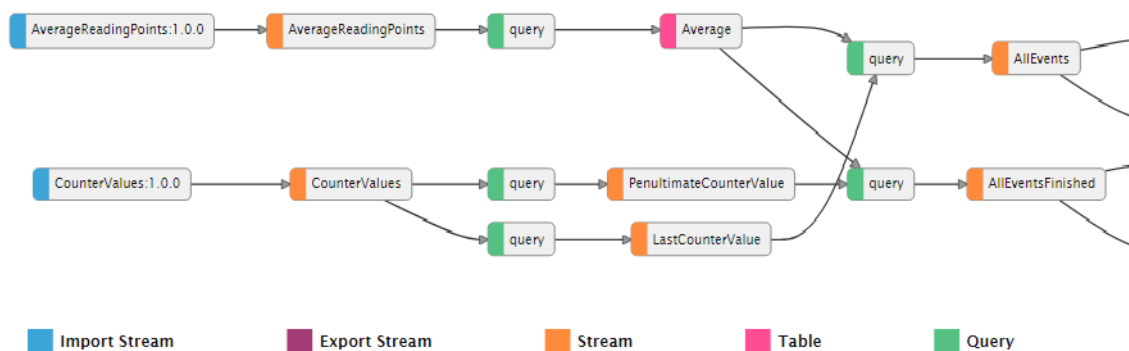


Figura 5.5: Receção de dados no núcleo do bloco processador de eventos (parte 1)

Para que se perceba o significado de cada bloco ilustrado na fig.5.5 e fig.5.6, de seguida analisar-se-à de forma mais pormenorizada cada um dos blocos de modo a perceber a sua função dentro da aplicação.

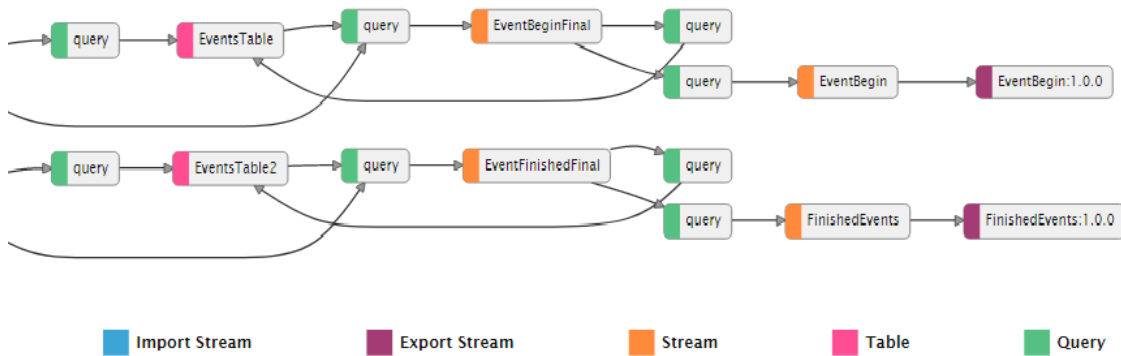


Figura 5.6: Recepção de dados no núcleo do bloco processador de eventos (parte 2)

Numa primeira instância, a operação realizada consiste na introdução das *streams* de dados externas ao bloco, no interior do núcleo do processador de eventos. É uma operação muito simples onde através dos recetores de *streams* se associam os dados a novas *streams*. Os conjuntos de dados externos utilizados neste processo, foram os dados adquiridos pelo vários sensores ao longo de diversos espaços temporais, conforme escolha do utilizador e também os dados referentes as médias horárias por dia da semana de cada sensor, sendo as suas *streams* denominadas respetivamente de "CounterValues" e "AverageReadingPoints". A formatação dos dados nas presentes *stream* mantêm-se as mesmas que as visualizadas e descrita nas tab.5.1 e 5.2.

A utilização deste passo é essencial pois mais uma vez, a ferramenta WSO2 CEP não permite a execução imediata dos dados sem que estes passem primeiramente por *streams* de dados. Este primeiro passo é ilustrado na fig.5.7.



Figura 5.7: Recepção de dados no núcleo do bloco processador de eventos

De seguida após os dados já se encontrarem em *streams*, estes são submetidos a operações. No caso da ferramenta WSO2 sempre que se execute uma operação sobre os dados, independentemente da sua natureza (comparações, transformações, ...), este constrói um bloco *query* que representa essas operações.

Sendo assim, nesta segunda fase (fig.5.8), existem três operações sobre as *streams* anteriormente referidas. Inicialmente sobre a *stream* "CounterValues" são realizadas duas ações, estas ações em ambas muito idênticas, servem para colocar nas *streams* seguintes,

tanto os valores atuais dos dados bem como, os últimos e penúltimos valores associados aos dados de cada sensor. Desse modo, na *stream* "PenultimateCounterValue" encontra-se os valores a serem processados e os valores processados nos dois intervalos de tempo anteriores a este, enquanto que na *stream* "LastCounterValue" encontram-se apenas os dados a ser processados e anteriores a estes, juntamente em ambas as *streams* encontra-se também um campo com a identificação única de cada sensor, bem como a via e faixa rodoviária a ser avaliada.

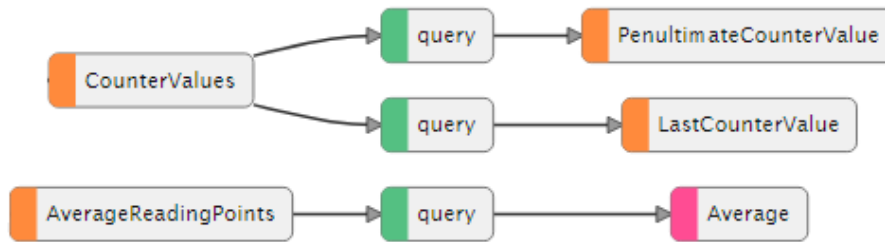


Figura 5.8: Operação lógica sobre os dados no núcleo do bloco processador de eventos

Relativamente à *stream* "AverageReadingPoints" esta é copiada na íntegra para uma tabela de eventos, tabela esta que é a única forma de se conseguir armazenar dados na ferramenta WSO2 CEP, sendo que os blocos de *streams* não apresentam esta capacidade. A utilização deste armazenamento de dados foi optado, pois não faz sentido estar continuamente a sobrecarregar a ferramenta com a introdução dos valores médios dos sensores, dados estes que serão utilizados continuamente.

A operação que seguinte ilustrada na fig.5.9 representa a seleção de eventos anómalos. Esta operação é dividida em dois passos, a identificação de início de evento e a identificação de fim de evento.

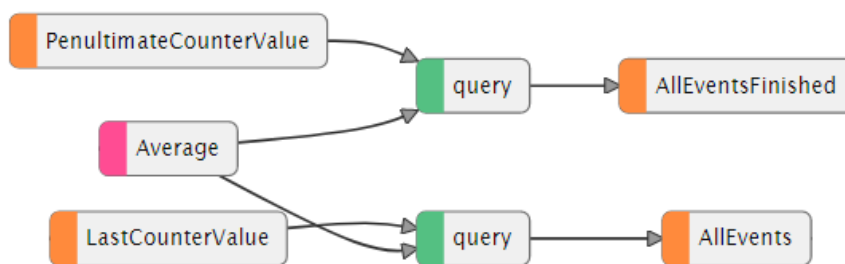


Figura 5.9: Operação lógica para identificação de início e término eventos sobre os dados no núcleo do bloco processador de eventos

A operação de identificação de início de evento consiste numa comparação lógica entre a *stream* "LastCounterValues" e a tabela "Average", onde através da equivalência de identificadores de sensores, faixas rodoviárias, horas e dias da semana são comparados os valores de velocidade média nas faixas de modo a identificar desvios percentuais relativamente aos valores médios comportamentais, armazenados na tabela "Average", ou seja, se a velocidade média instantânea de um sensor apresentar um desvio de 20% (valor

exemplo) relativamente às médias para esse sensor, à hora e dia da semana estudados, é identificado esse caso como um possível evento. São também utilizados valores pré-definidos de ocupação de faixas de modo eliminar os períodos de menor número de carros, onde as variações de velocidade são maiores, sendo maior a probabilidade do desvio percentual ser quebrado, induzindo em erro a aplicação no processo de identificação de eventos. Esta estratégia permite a eliminação de falsos positivos.

As comparações referidas anteriormente para identificação de início de eventos, não são só aplicadas às velocidades do último conjunto de dados processados, sendo comparados também com os valores registados anteriormente para cada sensor. A razão pela qual se adotou esta estratégia, deve-se ao facto de por vezes os sensores apresentarem erros únicos em leituras, o que poderia ativar deste modo, mais uma vez uma identificação falsa de eventos, sendo necessário para a identificação de início de eventos que se prolonguem a comparação pelo menos por dois intervalos de tempo contínuos, eliminando possíveis falsos positivos nesta deteção. Esta operação pode ser visualizada na fig.5.10.

Relativamente as opções percentuais e valores de ocupação de via utilizados para cancelamento de períodos com menor número de carros, este são considerados parâmetros variáveis consoante a localização das autoestradas, sendo este tema abordado à posteriori na sec.5.4.

```
on (LastCounterValue.Day == Average.dayOfTheWeek
and LastCounterValue.Hour == Average.hour
and LastCounterValue.Id == Average.Id
and LastCounterValue.Hit_11 < (Average.velocity_data*0.9)
and LastCounterValue.Hit11_last < (Average.velocity_data*0.9)
and Average.Road == 1 and LastCounterValue.Occu_11>300
and LastCounterValue.Occu11 last>300)
```

Figura 5.10: Código exemplo para identificação de início de evento

Quanto à operação de deteção de término de eventos esta é igualmente parecida com a de início de evento, tendo uma pequena modificação no número de registos comparados. Esta operação compara se os dados a serem processados no momento se encontram acima do desvio percentual escolhido em comparação com as médias dos sensores, enquanto que os outros dois registos, relativos aos dois intervalos de tempos imediatamente anteriores são analisados de modo a se compreender se estes se encontram abaixo do desvio percentual com as médias, ou seja, se ocorreu uma passagem de evento anómalo para término desse mesmo evento.

Em ambas as operações apenas são transpostos para as *streams* seguintes todos os dados, já identificados pelas regras anteriores, dados estes que cumpram os requisitos descritos anteriormente, logo identificados como possíveis eventos anómalos.

Após identificação dos eventos anómalos, em ambos os casos, tanto os resultados das operações de identificação de início e término de evento são colocados em tabelas de eventos (fig.5.12). A utilização destas tabelas permite a não repetição de identificação das operações anteriormente descritas em intervalos inferiores a três horas, ou seja, permite

```

on (PenultimateCounterValue.Day == Average.dayOfTheWeek
and PenultimateCounterValue.Hour == Average.hour
and PenultimateCounterValue.Id == Average.Id
and Average.Road == 1
and ( PenultimateCounterValue.Hit_11 >= (Average.velocity_data0*0.9)
and PenultimateCounterValue.Hit11_last < (Average.velocity_data0*0.9)
and PenultimateCounterValue.Hit11_pen < (Average.velocity_data0*0.9)
and PenultimateCounterValue.Occu_11>300
and PenultimateCounterValue.Occu11_last>300
and PenultimateCounterValue.Occu11_pen>300)

```

Figura 5.11: Código exemplo para identificação de término de evento

eliminar a constante identificação de eventos em intervalos imediatamente posteriores à ocorrência do evento, eliminando as continuas identificações de eventos mesmo quando estes já se encontra sinalizados.

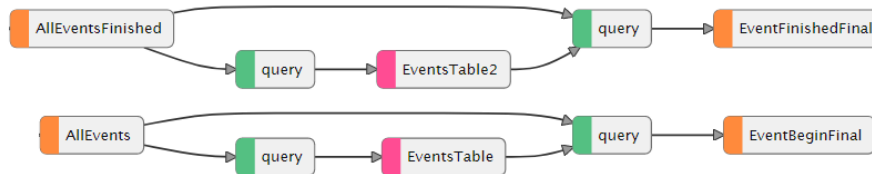


Figura 5.12: Verificação de repetições sobre os eventos no núcleo do bloco processador de eventos

Por fim, o último passo realizado no bloco de processador de eventos, corresponde a operações exportação dos eventos para fora do processador de eventos. Como particularidade, são necessárias executar algumas acções de identificação de eventos, para que não se proceda a uma exportação repetitiva de eventos, ou seja, é realizada uma operação que apenas permite que cada evento na tabela de eventos seja passado para o exterior da aplicação uma única vez.

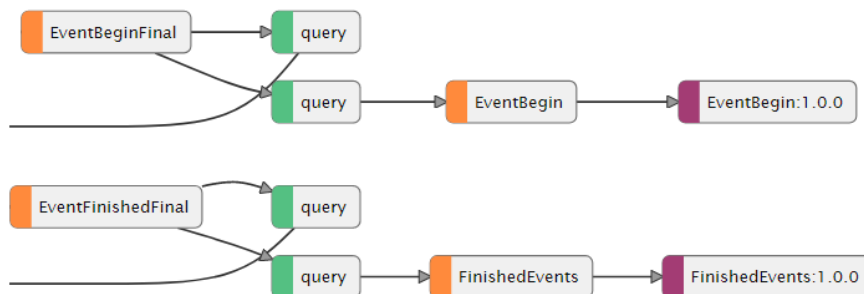


Figura 5.13: Exportação dos eventos para fora do núcleo do bloco processador de eventos

## 5.4 Parametrização

Na secção anterior de aplicação de regras, quando trabalhado o bloco de processador de eventos foram referidos parâmetros variáveis para identificação de eventos. Estes

parâmetros são apresentados sobre a forma de desvios percentagens sobre os valores médios, dos padrões das autoestradas calculados previamente e ainda sobre o número de veículos em circulação por hora em cada autoestrada.

A classificação destes parâmetros como variáveis deve-se ao facto estes poderem assumir diferentes valores consoante a otimização da aplicação.

Contudo a utilização destes dois parâmetros é diferente, pois enquanto que o desvio percentual pode ser aplicado de igual modo, independentemente da localização dos sensores, ou seja, para os vários sensores podem ser utilizados vários desvios percentuais (10%, 20%,...), no caso do parâmetro relativo à ocupação da via, sendo este um valor numérico, apresenta uma dependência relativa a categorização do sensor segundo a dispersão populacional abrangida pelos mesmos, sendo necessário utilizar diferentes valores numéricos para identificação de eventos em sensores próximos de grandes centros populacionais, ou próximos de pequenos centros populacionais. Esta ocorrência acontece porque a utilização deste parâmetro assiste na eliminação de falsos positivos em horas que as variações de velocidade sejam muito elevadas, como é o caso dos períodos noturnos.

Assim, para os parâmetro pode ser usada uma gama de percentagens entre 0% e 100%, enquanto que no parâmetro de ocupação da via, os valores utilizados variam entre os 70 veículos por hora e os 300 veículos por hora. A utilização destes valores teve por detrás um estudo prévio dos valores médios do número de veículos em circulação por hora para as diferentes categorias de localização de sensores.

## VALIDAÇÃO E RESULTADOS

Neste capítulo será avaliado o potencial da aplicação de detecção de eventos concebida. Esta análise corresponde à fase de avaliação no modelo CRISP-DM (2.3), onde serão abordadas todas as regras de aplicadas na detecção de eventos trabalhadas no capítulo 5 e avaliados os resultados segundo o método de precisão e sensibilidade.

### 6.1 Resultados

Tendo como principal objetivo de dissertação a elaboração de uma aplicação capaz de detetar eventos anómalos em autoestradas, é plausível de se assumir a realização do presente objetivo, sendo possível através da aplicação gerada identificar a ocorrência de vários eventos de diferentes autoestradas, bem como em diferentes espaços temporais.

Contudo a capacidade da distinção de eventos por parte da aplicação não foi alcançada, pois o foco dos eventos centrou-se em eventos com duração até 5 horas e em todos os eventos com esta característica, manifestam-se de uma forma muito idêntica, não sendo possível categorização dos vários eventos.

De seguida, irão ser apresentados alguns exemplos de detecção de eventos, tendo como base as regras clarificadas no capítulo anterior.

```
[2017-08-21 14:55:18,309] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: EventBegin,
Event: ModifiedTime:2017-02-20 14:39:52,
Id:199,
Road:1
[2017-08-21 14:55:18,479] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: FinishedEvents,
Event: ModifiedTime:2017-02-20 16:28:14,
Id:199,
Road:1
```

Figura 6.1: Visualização de um início e fim de evento, detetado pela ferramenta desenvolvida

Para que seja possível obter-se valor sobre a identificação de eventos, estes têm de apresentar um conjunto de parâmetros que permitam identificar a data e hora, bem

como a sua localização, possibilitando diferenciar os diversos eventos. Deste modo, após detecção de um evento, os resultados apresentados terão quatro parâmetros, um parâmetro identificador de início ou término de cada evento, outro parâmetro referente à data e hora da ocorrência do evento (ModifiedTime), outro representando o identificador único do sensor (Id) e por fim um parâmetro detalhado da via em que se detetou o evento (Road). Na fig.6.1 é ilustrado um exemplo da ocorrência de um evento, onde estão especificados todos os parâmetros referidos anteriormente.

A apresentação de resultados por vezes pode ser um pouco mais complexa de se avaliar quando o número de sensores em estudo é superior, pois nem sempre existe uma continuidade na visualização de início e fim de eventos, ou seja, é possível que entre a ocorrência de um evento exista um evento menor, em outro sensor, que se manifestará entre o início e o fim do primeiro evento. Como pode ser visualizado na fig.6.2, onde a vermelho está representado o início e fim de um evento e entre este mesmo evento dá-se a ocorrência de outro evento.

A utilização deste método de visualização foi adotado a pensar no processamento em tempo-real, onde interessa adquirir de imediato o início ou o fim de cada evento, para que seja possível retirar-se instantaneamente valor sobre a detecção de eventos, não sendo viável a publicação de início e fim de eventos ocorrer em simultâneo ou de seguida no visualizador de eventos. Assim, apesar de a visualização apresentada para início e término de eventos ser mais confusa a nível visual, é possível identificar instantaneamente o início ou o final do evento, permitindo retirar em tempo-real valor sobre esta operação, o que não era possível se não fosse adotado este método de visualização.

```
[2017-08-26 18:08:44,916] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: EventBegin,
Event: ModifiedTime:2017-02-01 14:33:21,
Id:174,
Road:2
1
[2017-08-26 18:08:45,168] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: EventBegin,
Event: ModifiedTime:2017-02-01 15:12:44,
Id:199,
Road:1
1
[2017-08-26 18:08:45,199] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: FinishedEvents,
Event: ModifiedTime:2017-02-01 15:32:26,
Id:199,
Road:1
2
[2017-08-26 18:08:45,326] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: FinishedEvents,
Event: ModifiedTime:2017-02-01 16:02:00,
Id:174,
Road:2
2
```

Figura 6.2: Visualização de um início e fim de evento, entre outro evento

Assim sendo, a figura 6.3 ilustra parte dos eventos detetados ao longo do mês de Fevereiro, onde é possível de se observar a detecção detalhada de cada evento. Esta detecção de eventos teve apenas em consideração cinco sensores dispersos ao longo do território esloveno, servindo apenas para uma demonstração mais clara dos resultados adquiridos pela aplicação criada.



```

[2017-08-26 18:08:42,082] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: EventBegin,
Event: ModifiedTime:2017-02-01 05:41:34,
Id:834,
Road:1
[2017-08-26 18:08:42,167] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: FinishedEvents,
Event: ModifiedTime:2017-02-01 06:11:06,
Id:834,
Road:1
[2017-08-26 18:08:44,916] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: EventBegin,
Event: ModifiedTime:2017-02-01 14:33:21,
Id:174,
Road:2
[2017-08-26 18:08:45,168] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: EventBegin,
Event: ModifiedTime:2017-02-01 15:12:44,
Id:199,
Road:1
[2017-08-26 18:08:45,199] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: FinishedEvents,
Event: ModifiedTime:2017-02-01 15:32:26,
Id:199,
Road:1
[2017-08-26 18:08:45,326] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: FinishedEvents,
Event: ModifiedTime:2017-02-01 16:02:00,
Id:174,
Road:2
[2017-08-26 18:08:51,877] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: EventBegin,
Event: ModifiedTime:2017-02-02 14:21:17,
Id:199,
Road:1
[2017-08-26 18:08:51,964] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: FinishedEvents,
Event: ModifiedTime:2017-02-02 14:50:50,
Id:199,
Road:1
[2017-08-26 18:08:58,955] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: EventBegin,
Event: ModifiedTime:2017-02-03 14:19:02,
Id:199,
Road:1
[2017-08-26 18:08:59,381] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: FinishedEvents,
Event: ModifiedTime:2017-02-03 16:07:23,
Id:199,
Road:1
[2017-08-26 18:09:21,454] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: EventBegin,
Event: ModifiedTime:2017-02-06 14:32:05,
Id:199,
Road:1
[2017-08-26 18:09:21,583] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: FinishedEvents,
Event: ModifiedTime:2017-02-06 15:11:28,
Id:199,
Road:1
[2017-08-26 18:09:28,549] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: EventBegin,
Event: ModifiedTime:2017-02-07 15:29:00,
Id:199,
Road:1
[2017-08-26 18:09:28,593] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: FinishedEvents,
Event: ModifiedTime:2017-02-07 15:48:42,
Id:199,
Road:1
[2017-08-26 18:09:50,353] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: EventBegin,
Event: ModifiedTime:2017-02-10 14:23:17,
Id:199,
Road:1
[2017-08-26 18:09:50,386] INFO {org.wso2.carbon.event.output.adapter.logger.LoggerEventAdapter} - Unique ID: EventBegin,
Event: ModifiedTime:2017-02-10 14:33:09,
Id:174,
Road:2

```

Figura 6.3: Visualização de um início e fim de eventos, detetados pela ferramenta desenvolvida

Para um estudo mais profundo da aplicação, foram testados aleatoriamente cinco sensores por zona, tendo em consideração a classificação de dispersão populacional já estuda. Este estudo além de permitir uma primeira contagem sobre o número de eventos, serviu ainda para avaliar se as regras de deteção de início e fim de evento estariam a responder de maneira correta e uniforme. Sendo o estudo ilustrado nas figuras 6.4, 6.5 e 6.6.

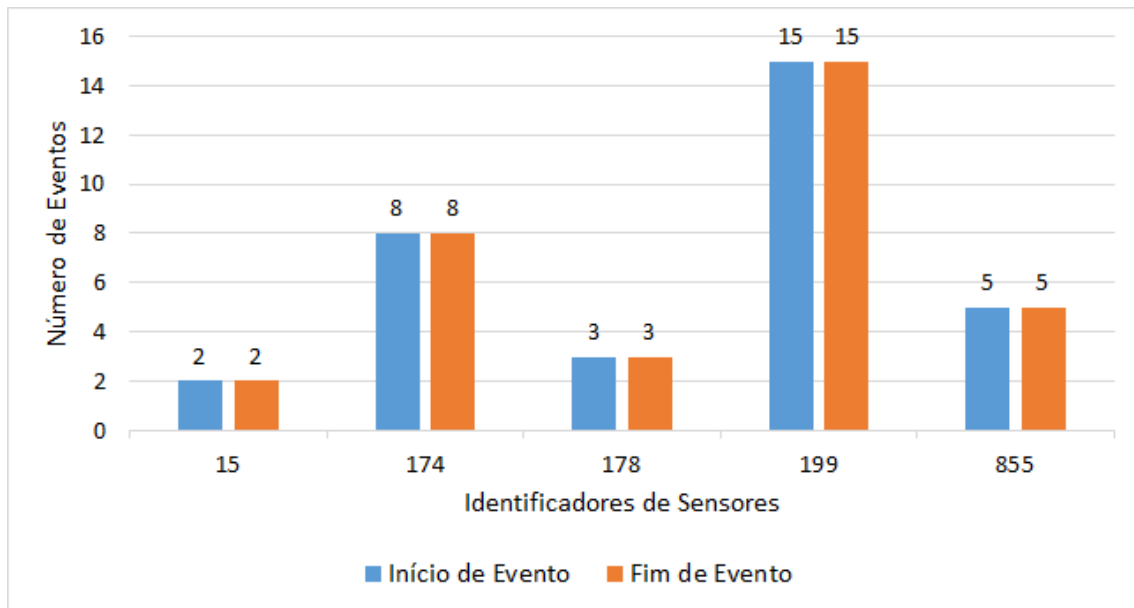


Figura 6.4: Ocorrência de início e fim de eventos na aplicação para 5 sensores próximos de um grande centro populacional

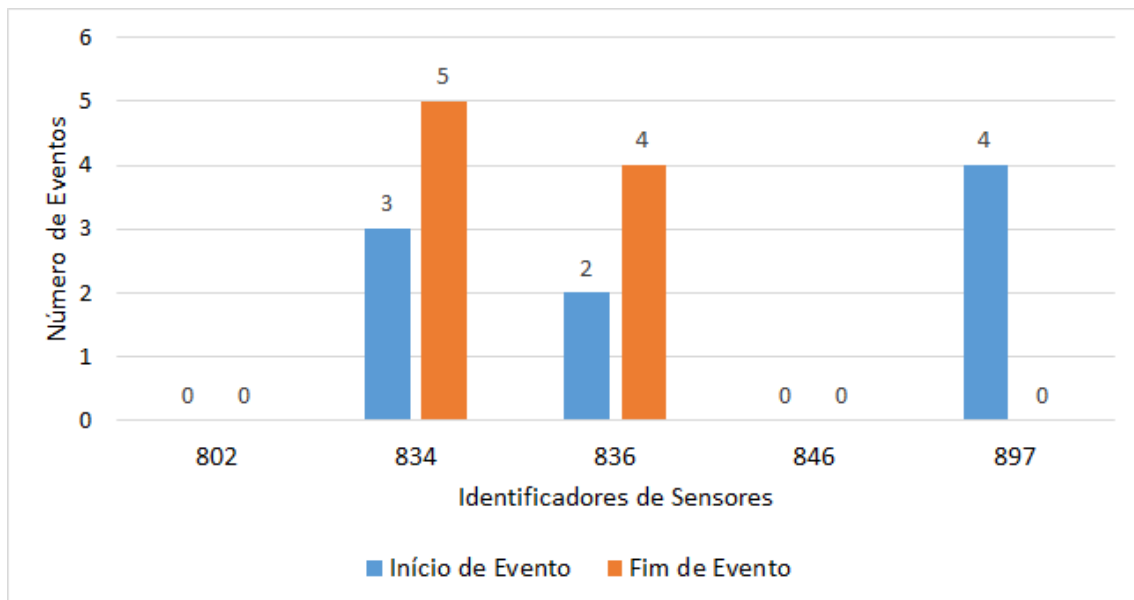


Figura 6.5: Ocorrência de início e fim de eventos na aplicação para 5 sensores entre dois grandes centros populacionais

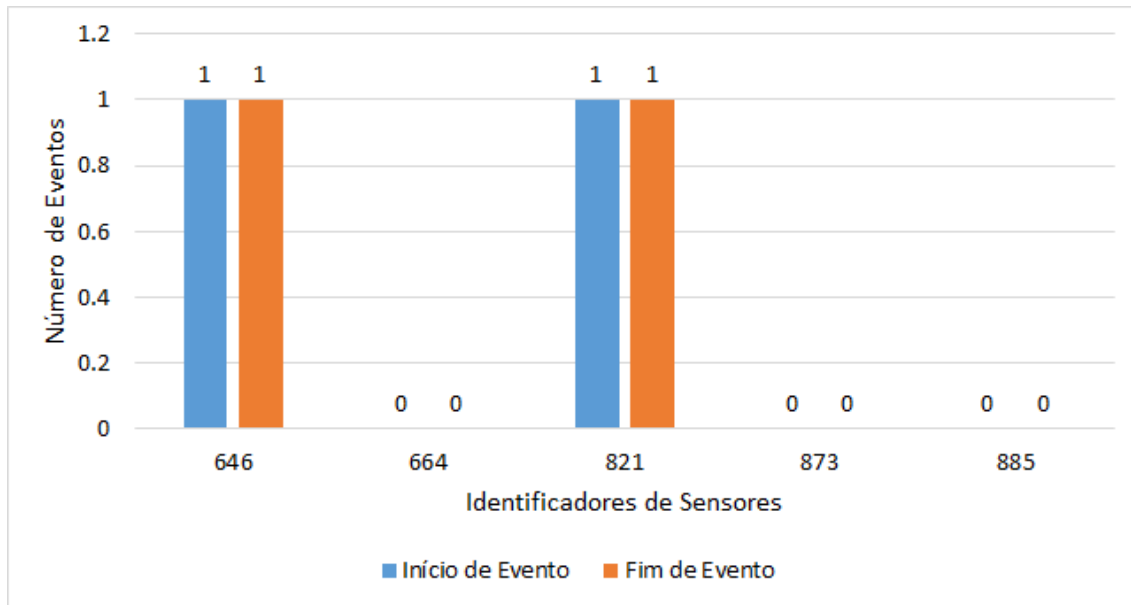


Figura 6.6: Ocorrência de início e fim de eventos na aplicação para 5 sensores próximos de um pequeno centro populacional

É de realçar que aplicação de regras de forma geral teve um comportamento uniforme, excepto nos sensores localizados entre dois grandes centros populacionais, onde se observa uma discrepância na esperada uniformidade entre início e termino de eventos.

## 6.2 Validação

Após serem observados os resultados, é importante executar uma atribuição de valores sobre os resultados, de forma classificar a relevância dos dados adquiridos. A validação de resultados terá por base três conceitos a precisão, sensibilidade e o *F-measure*.

$$\text{Precisão} = \frac{\text{eventos que correspondem}}{\text{eventos detetados}} \quad (6.1)$$

$$\text{Sensibilidade} = \frac{\text{eventos que correspondem}}{\text{eventos na base de dados}} \quad (6.2)$$

$$F - \text{measure} = 2 \frac{\text{precisão} \times \text{sensibilidade}}{\text{precisão} + \text{sensibilidade}} \quad (6.3)$$

É de salientar que a validação de resultados é sempre uma medida comparativa, sendo que no presente caso terá como comparação um conjunto de eventos detetados de forma não pro-ativa, não sendo este conjunto de dados o mais indicado para o presente estudo.

Numa primeira fase, a validação terá apenas em consideração as diferentes localizações já referidas anteriormente e à posteriori comparar-se-à os vários resultados obtidos tendo em consideração variações nos parâmetros das regras pré-estabelecidas de identificação de eventos, ambos os testes serão realizados sobre um espaço-temporal de um mês. Para os seguintes resultados, o valor base de variação de velocidade média para identificação de eventos utilizado foi de 20% comparativamente com os padrões previamente criados.

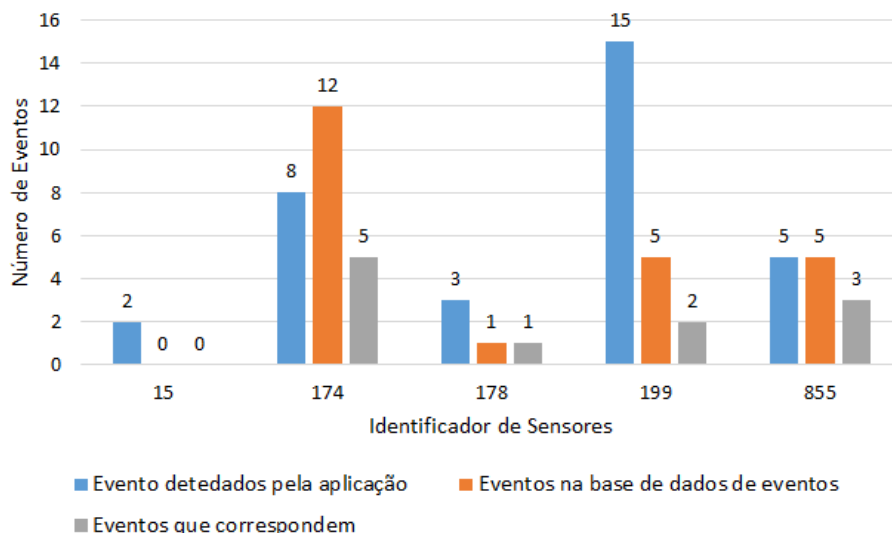


Figura 6.7: Comparação da ocorrência de eventos em 5 sensores próximos de um grande centro populacional, com uma taxa de variação da velocidade média de 20%

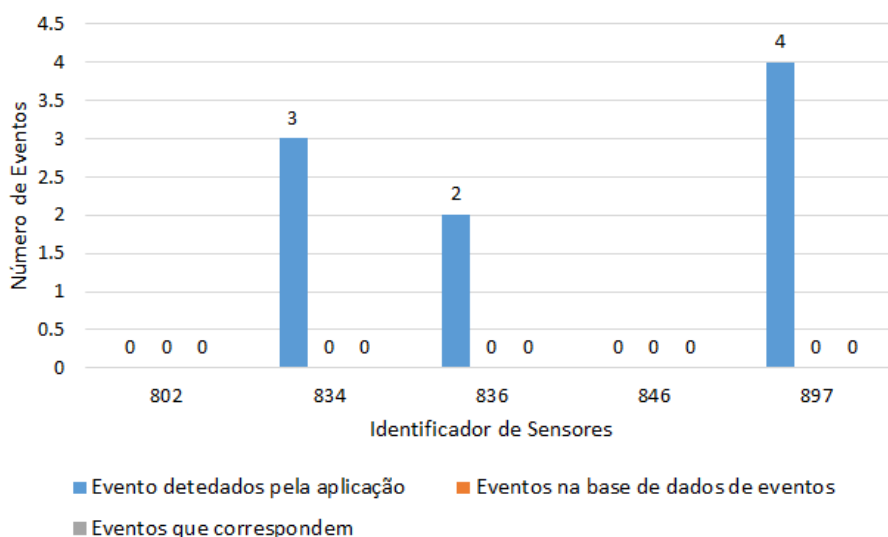


Figura 6.8: Comparação da ocorrência de eventos em 5 sensores entre dois grandes centros populacionais, com uma taxa de variação da velocidade média de 20%

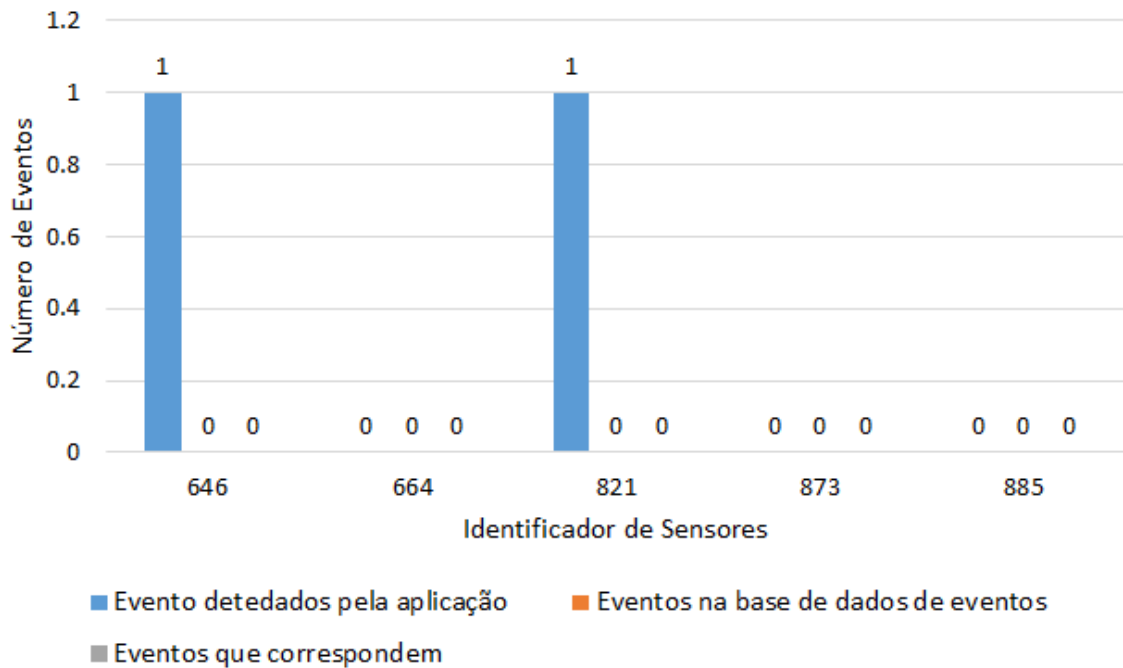


Figura 6.9: Comparação da ocorrência de eventos em 5 sensores próximos de um pequeno centro populacional, com uma taxa de variação da velocidade média de 20%

Numa primeira análise é de realçar um enorme desfasamento entre o número de eventos na base de dados e os eventos detetados pela aplicação criada, mais uma vez e como já referido este desfasamento deve-se pelo facto de a base de dados de eventos apresentada ser não pro-ativa, estando a identificação de eventos dependente da ação humana.

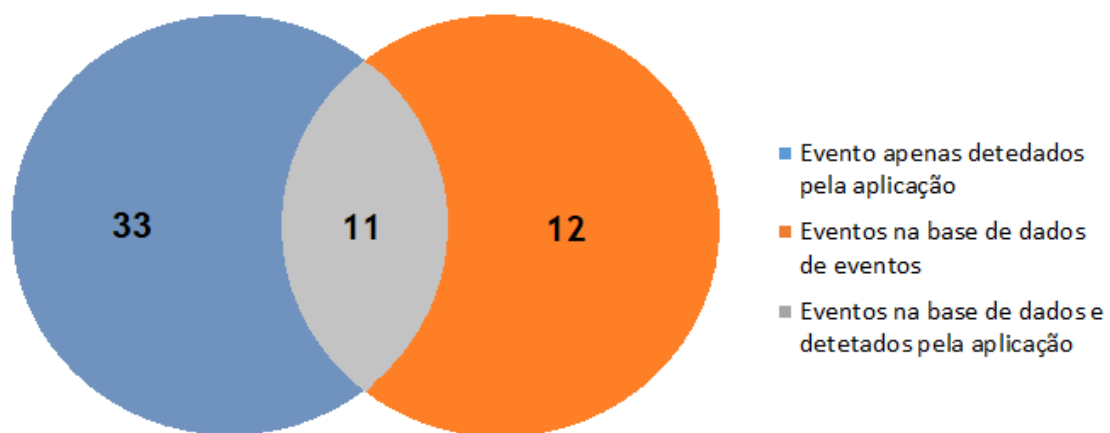


Figura 6.10: Número total de eventos por categoria para 15 sensores, com uma taxa de variação da velocidade média de 20%

A figura 6.10 ilustra todos os resultados adquiridos através dos 15 sensores estudados, onde se evidencia os 44 eventos detetados pela aplicação, contra apenas 23 eventos previamente identificados e colocados na base de dados. Para que a aplicação criada funciona-se a 100% de eficiência era de se esperar que pelo menos todos os eventos na base de dados apresentassem correspondência por parte da aplicação, o que não é verificado, sendo alcançando apenas cerca de metade desse valor (representado pelos 11 eventos), este facto pode ser justificado pela calibração da aplicação ou através da não influência do evento no tráfego da autoestrada em análise. É de realçar, relativamente a este último ponto que a seleção na base de dados de eventos, teve apenas como carácter seletor o tempo do evento e a sua localização, podendo existir eventos que apesar de se apresentarem próximos dos sensores em estudo (máximo 1 km), podem não causar qualquer influência no tráfego, mas sendo contabilizados de igual forma para a validação dos resultados.

Aposta esta contagem dos resultados, para que seja possível classificar as regras e a aplicação criada foram submetidos os resultados anteriormente visualizados ao método de precisão, sensibilidade e *F-measure*, como se pode visualizar pela tabela 6.1.

Tabela 6.1: Precisão, sensibilidade e F-Measure para eventos em 15 sensores com variações de velocidade superiores ou iguais 20%

Localização	Identificador de Sensor	Precisão	Sensibilidade	F-measure
	15	0	*	*
Grandes centros populacionais	174	0.625	0.417	0.5
	178	0.333	1	0.5
	199	0.133	0.4	0.2
	855	0.6	0.6	0.6
Entre dois	646	0	*	*
	664	*	*	*
grandes centros populacionais	821	0	*	*
	873	*	*	*
	885	*	*	*
	802	*	*	*
Pequenos centros populacionais	834	0	*	*
	836	0	*	*
	846	*	*	*
	897	0	*	*

(a) \* - indeterminação

É então possível de se observável que metade da tabela apenas apresenta zeros ou indeterminações, facto este justificável por tanto a sensibilidade como a precisão serem valores dependentes de eventos correspondentes entre a base de dados de eventos e os eventos detetados pela aplicação, sendo observável pelas figuras, 6.7, 6.8 e 6.9, que para a maioria dos sensores este valor é zero o que provoca tais resultados na tabela. Quanto ao valor *F-measure*, este representa a combinação dos valores de precisão e sensibilidade, sendo estes na sua maioria indeterminações, não sendo possível realizar o cálculo deste

valor de igual modo.

Após uma breve análise sobre o comportamento da aplicação para detecção de eventos com variações superiores a 20% da velocidade média padrão, é então necessário de seguida alterar esta percentagem de modo a perceber o comportamento na detecção e identificação de eventos da aplicação, bem como, identificar diferenças nos valores de precisão, sensibilidade e *F-measure* de modo a poder-se aprimorar os parâmetros de detecção de eventos da aplicação elaborada.

Para os seguintes testes apenas foram utilizados os sensores próximos de grandes centros populacionais, pois são estes os únicos que apresentam valores comparáveis na base de dados de eventos, permitindo dessa forma elaborar uma comparação entre as várias variações dos parâmetros da aplicação.

Tabela 6.2: Tabela de contadores de eventos para 5 sensores com variação nos parâmetros das regras da aplicação concebida

Nível de variação da velocidade	Identificador de Sensor	Evento detetados pela aplicação	Eventos na base de dados de eventos	Eventos que correspondem
10%	15	8	0	0
	174	8	12	5
	178	3	1	1
	199	22	5	2
	855	14	5	3
20%	15	2	0	0
	174	8	12	5
	178	3	1	1
	199	15	5	2
	855	5	5	3
30%	15	0	0	0
	174	6	12	5
	178	3	1	1
	199	10	5	1
	855	3	5	1

(a) \* - indeterminação

Como se atenta pela tabela 6.2, ao se variar os parâmetros de detecção de eventos obtém-se valores diferentes no número de eventos gerados pela aplicação. Quanto menor a percentagem aplicada sobre as variações no comportamento padrão das autoestradas, mais suscetível é a aplicação a variações no comportamento do tráfego rodoviário, sendo maior o número de eventos detetados pela aplicação, enquanto que quanto maior a percentagem aplicada na aplicação, menos suscetível é esta a variações e por sua vez, menores são os números de eventos detetados.

Quando comparado a precisão, sensibilidade e *F-measure* para os vários sensores com diferentes variações de percentagem de variação dos níveis de velocidade como parâmetro de identificação de eventos na aplicação gerada, chega-se à conclusão que quanto menor

for a percentagem, maior é a sensibilidade, e quanto maior for a percentagem, maior é a precisão. Contudo esta avaliação teve três níveis de variação diferentes, 10%, 20% e 30% na qual o valor intermédio de 20% foi o valor mais eficaz em todos os parâmetros, alcançando desse modo também um maior valor de *F-measure* o qual representa uma mediação harmónica entre os dois valores (precisão e sensibilidade) calculados anteriormente.

Tabela 6.3: Tabela de precisão, sensibilidade e *F-measure* para eventos em 5 sensores com variação nos parâmetros das regras da aplicação concebida

Nível de variação da velocidade	Identificador de Sensor	Precisão	Sensibilidade	F-measure
10%	15	0	*	*
	174	0.625	0.417	0.5
	178	0.333	1	0.5
	199	0.091	0.4	0.148
	855	0.214	0.6	0.316
20%	15	0	*	*
	174	0.625	0.417	0.5
	178	0.333	1	0.5
	199	0.133	0.4	0.2
	855	0.6	0.6	0.6
30%	15	*	*	*
	174	0.833	0.417	0.556
	178	0.333	1	0.5
	199	0.1	0.2	0.133
	855	0.333	0.2	0.25

(a) \* - indeterminação

Em suma, os resultados obtidos pela aplicação na deteção de eventos foram bastantes positivos, sendo capaz de detetar eventos em tempo real, bem como através de dados rodoviários históricos. Foi ainda concluído que a aplicação apesar de não detetar na totalidade todos os eventos na base de dados de eventos, apresenta resultados bastante positivos, sendo através do desvio de 20% dos padrões comportamentais criados, que se obtêm melhores resultados, comparativamente a base de dados de eventos não pro-ativos.



## CONCLUSÕES

A rentabilidade e qualidade de vida, são dois fatores tidos em enorme consideração por parte das empresas e pessoas. Sendo a presente tese englobada num projeto europeu inovador (OPTIMUM), que visa a criação de um sistema de portajamento dinâmico em autoestradas, ambos os fatores expostos anteriormente enquadram-se na solução apresentada neste projeto. A implementação deste sistema, apresenta-se como uma mais valia, pois apresenta-se com perspetivas bastantes positivas e inovadoras tanto as empresas como os seus utilizadores. A nível empresarial, espera-se uma maior rentabilidade das autoestradas, tendo por detrás uma maior e melhor utilização das mesmas por parte dos condutores. A nível social a implementação deste sistema também apresenta vantagens, esperando-se um aumento na qualidade de vida da sociedade, com a diminuição da sobrelotação das vias secundárias e por sua vez, todos os riscos associados a esta.

Este sistema pressupõem de uma ferramenta capaz de analisar eventos anómalos em autoestradas de modo, a que a qualidade de serviço prestado, não seja afetado, devido as variações de preços. É então neste ponto que se enquadra a presente tese, tendo como principal objetivo o desenvolvimento de uma aplicação (protótipo) proativa de categorização e deteção de eventos complexos, tendo por base um repositório de dados de tráfego rodoviário, sendo através desta criada um ponto de partida para o sistema de portajamento dinâmico.

A criação da aplicação de deteção de eventos em autoestrada é um projeto inovador e que visa beneficiar em muito o meio rodoviário, bem como as concessionárias das autoestradas e a sociedade. Esta pode vir a substituir as atuais ferramentas de deteção anómala de eventos, que até então apresentam dependência de componente humana, ou seja, não são proativas, sendo esta a maior diferença desta aplicação para as atuais. A capacidade proativa desta ferramenta permite a libertação/redução dos meios humanos necessários para a deteção de eventos, como ainda permite a otimização de todo este processo, sendo

detetado um maior número de eventos e sendo coberta uma maior área de deteção, com a utilização desta ferramenta.

De modo geral, o objetivo principal foi concluído com sucesso, foi realizado o desenvolvimento de um protótipo otimizado para deteção de eventos anómalos rodoviários, na qual se obtiveram resultados bastante positivos na deteção de eventos já previamente conhecidos e na deteção de novos eventos, sendo realizados todos os objetivos proposto a exceção da categorização de eventos, explicito mais a frente.

Para o desenvolvimento desta aplicação, foi necessário realizar um estudo prévio sobre os diversos perfis de utilização das autoestradas, o qual também foi concluído com sucesso e utilizado para criação de padrões comportamentais em autoestradas.

É de salientar ainda a seleção e uniformização de dados, ao longo desta presente dissertação foram utilizados dados de diversas fontes de dados, os quais não se encontravam com a formatação correta ou valores de qualidade exigidos, sendo realizada uma limpeza e seleção que permitiu uma uniformização da formatação dos dados, podendo estes ser usados em trabalhos futuros.

Contudo, nem todos os objetivos pretendidos foram alcançados, pois para além da deteção de eventos, era também pretendido uma distinção entre os diversos eventos rodoviários detetados, ou seja, aplicar uma categorização sobre os eventos (trânsito, acidente, ...). Categorização esta que não foi realizada pois quando estudados os padrões comportamentais dos diversos eventos anómalos de curta duração, acidentes e trânsito, ambos estes se manifestavam de forma muito semelhante, não sendo encontrado nenhum padrão diferenciador para estes. Apesar da ocorrência deste ponto negativo, todos os outros objetivos esperados foram concluídos com sucesso, sendo criada uma aplicação capaz de detetar eventos anómalos de forma automática em autoestradas.

No geral, este foi um projeto muito interessante e desafiante, em que foram desenvolvidos estudos e ferramentas, nas quais revejo elevados padrões de qualidade e eficiência. Relativamente aos estudos realizados, estes podem ser utilizados para outras análises ou até como base de desenvolvimento de outras aplicações na área do ITS, tendo ainda sido elaborado um artigo científico. Quanto à ferramenta desenvolvida, foi criado um protótipo que pode vir a ser o ponto de partida para uma aplicação por parte de uma concessionária das autoestradas.

Em suma, num futuro próximo penso que este tipo de aplicações irão contribuir para os problemas causados pela sobrelotação das vias secundárias, sendo a meu ver, este um ponto de partida para a criação do sistema de portajamento dinâmico, bem como para o desenvolvimento de uma próxima geração de aplicações e ferramentas ITS com base em sistemas automáticos, proativos, com vista a melhorar as ferramentas atuais e reduzindo cada vez mais a componente humano e o erro a ela associado neste tipo de sistemas.

## BIBLIOGRAFIA

- [1] L. Figueiredo, I. Jesus, J. T. Machado, J. R. Ferreira e J. M. De Carvalho. “Towards the development of intelligent transportation systems”. Em: *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*. IEEE. 2001, pp. 1206–1211.
- [2] K. N. Qureshi e A. H. Abdullah. “A survey on intelligent transportation systems”. Em: *Middle-East Journal of Scientific Research* 15.5 (2013), pp. 629–642.
- [3] U. D. of Transportation. *Adaptive Signal Control*. Accessed: 2017-02-10. 2009. URL: [www.itsoverview.its.dot.gov/default.asp](http://www.itsoverview.its.dot.gov/default.asp).
- [4] U. D. of Transportation. *Information Dissemination*. Accessed: 2017-02-10. URL: [/Options.asp?System=IMS{&}SubSystem=ID{&}Tech=Information](http://Options.asp?System=IMS{&}SubSystem=ID{&}Tech=Information).
- [5] *Qual a importância da sinalização de trânsito*. Accessed: 2017-12-23. URL: <http://seplaque.com.br/2017/04/25>.
- [6] U. D. of Transportation. *Adaptive Signal Control*. Accessed: 2017-02-10. 2009. URL: [/Options.asp?System=AM{&}SubSystem=TC{&}Tech=Adaptive](http://Options.asp?System=AM{&}SubSystem=TC{&}Tech=Adaptive).
- [7] U. D. of Transportation. *Advanced Automated Collision Notification*. Accessed: 2017-02-10. 2009. URL: [/Options . asp ? System = CNS { & } SubSystem = AA { & } Tech = Advanced](http://Options.asp?System=CNS{&}SubSystem=AA{&}Tech=Advanced).
- [8] U. S. D. of Transportation. *ITS ePrimer- Module 1:Introduction to ITS*. Accessed: 2016-02-10. URL: [www.pcb.its.dot.gov/eprimer/module1.aspx{&}benefits](http://www.pcb.its.dot.gov/eprimer/module1.aspx{&}benefits).
- [9] F. Ferreira. *Adaptive Signal Control*. Accessed: 2017-02-10.
- [10] T. Staff. *U.S. DoT To Mandate Vehicle to Vehicle Telematics for Crash Avoidance, Sparking Privacy Concerns*. Accessed: 2017-12-23. URL: <http://www.thetruthaboutcars.com>.
- [11] A. Ramos e S. Machado. “Portagens dinâmicas no caminho da gestão eficiente da infraestrutura rodoviária”. Em: *Carga e Mercadorias Transportes* 163 (2016), pp. 32–36.
- [12] *OPTIMUM: A short description*. Accessed: 2017-12-23. URL: <http://www.optimumproject.eu/about/project.html>.

- [13] P. Figueiras, R. Costa, G. Guerreiro, H. Antunes, A. Rosa e R. Jardim-Gonçalves. “User Interface Support for a Big ETL Data Processing Pipeline - An application scenario on highway toll charging models”. Em: In Proceedings of the 23rd ICE/ITMC conference, 2017, Funchal, Portugal.
- [14] A. McAfee, E. Brynjolfsson, T. H. Davenport, D. Patil e D. Barton. “Big data”. Em: *The management revolution*. *Harvard Bus Rev* 90.10 (2012), pp. 61–67.
- [15] A. Gandomi e M. Haider. “Beyond the hype: Big data concepts, methods, and analytics”. Em: *International Journal of Information Management* 35.2 (2015), pp. 137–144.
- [16] N. Leavitt. “Complex-event processing poised for growth”. Em: *Computer* 42.4 (2009).
- [17] I. Schmerken. “Deciphering the myths around complex event processing”. Em: *Wall Street & Technology* (2008).
- [18] O. Etzion e P. Niblett. *Event processing in action*. Manning Publications Co., 2010.
- [19] D. C. Luckham. *Event processing for business: organizing the real-time enterprise*. John Wiley & Sons, 2011.
- [20] S. Oberoi. “Introduction to complex event processing & data streams”. Em: *SOA World Magazine*, 5 (2007), pp. 20–24.
- [21] E. Wu, Y. Diao e S. Rizvi. “High-performance complex event processing over streams”. Em: *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*. ACM. 2006, pp. 407–418.
- [22] J. Alferes. *Complex Event Processing, Processamento de Streams*, FCT-NOVA. 2015 - 2016.
- [23] R. Buyya e A. V. Dastjerdi. *Internet of Things: Principles and paradigms*. Elsevier, 2016.
- [24] T. N. Report. *Payment Card Fraud 2015*. Accessed: 2017-02-23. URL: <https://www.nilsonreport.com/>.
- [25] N. P. Schultz-Møller, M. Migliavacca e P. Pietzuch. “Distributed complex event processing with query rewriting”. Em: *Proceedings of the Third ACM International Conference on Distributed Event-Based Systems*. ACM. 2009, p. 4.
- [26] I. Correia, F. Fournier e I. Skarbovsky. “The uncertain case of credit card fraud detection”. Em: *Proceedings of the 9th ACM International Conference on Distributed Event-Based Systems*. ACM. 2015, pp. 181–192.
- [27] Z. Huaji, W. Huarui e S. Xiang. “Research on the ontology-based complex event processing engine of RFID technology for agricultural products”. Em: *Artificial Intelligence and Computational Intelligence, 2009. AICI'09. International Conference on*. Vol. 1. IEEE. 2009, pp. 328–333.

- [28] V. Simast. “A importância da gestão de frotas para a logística e os negócios”. Em: *Mundo Logística* (2016).
- [29] S. Nechifor, A. Petrescu, D. Damian, D. Puiu e B. Tarnauca. “Predictive analytics based on CEP for logistic of sensitive goods”. Em: *Optimization of Electrical and Electronic Equipment (OPTIM), 2014 International Conference on*. IEEE. 2014, pp. 817–822.
- [30] S. Nechifor, B. Târnaucă, L. Sasu, D. Puiu, A. Petrescu, J. Teutsch, W. Waterfeld e F. Moldoveanu. “Autonomic monitoring approach based on cep and ml for logistic of sensitive goods”. Em: *Intelligent Engineering Systems (INES), 2014 18th International Conference on*. IEEE. 2014, pp. 67–72.
- [31] B. Gaunitz, M. Roth e B. Franczyk. “Dynamic and scalable real-time analytics in logistics combining Apache Storm with Complex Event Processing for enabling new business models in logistics”. Em: *Evaluation of Novel Approaches to Software Engineering (ENASE), 2015 International Conference on*. IEEE. 2015, pp. 289–294.
- [32] J. Dunkel, A. Fernández, R. Ortiz e S. Ossowski. “Event-driven architecture for decision support in traffic management systems”. Em: *Expert Systems with Applications* 38.6 (2011), pp. 6530–6539.
- [33] R. Costa, P. Figueiras, P. Maló, M. Jermol e K. Kalaboukas. “MobiS-Personalized Mobility Services for energy efficiency and security through advanced Artificial Intelligence techniques”. Em: *Neves-Silva, R., Watada, J., Phillips-Wren, G., Jain, LC, Howlett, RJ (eds.) Intelligent Decision Technologies. Frontiers in Artificial Intelligence and Applications* 241 (2013), pp. 296–306.
- [34] P. Figueiras, R. Costa, P. Malo, L. Bradesko e M. Jermol. “Knowledge base approach for developing a mobile personalized travel companion”. Em: *ITS Telecommunications (ITST), 2013 13th International Conference on*. IEEE. 2013, pp. 97–103.
- [35] “Workpackage 3 Modelling and Prediction”. Em: IEEE. 2013, pp. 97–103.
- [36] J. Ramos. *Como a tecnologia invadiu o desporto*. Accessed: 2017-08-10. 2015.
- [37] S. Ravindra. *WSO2 CEP in Action - An Analysis of Use in Real-World Applications of Different Domains*. Accessed: 2017-08-10. 2014. URL: <http://wso2.com/library/articles>.
- [38] S. Perera. *Solving DEBS 2013 Grand Challenge with WSO2 CEP/Siddhi*. Accessed: 2017-08-10. 2013. URL: <http://srinathsvi.blogspot.pt/2013/05/>.
- [39] G. Harper e S. D. Pickett. “Methods for mining HTS data”. Em: *Drug Discovery Today* 11.15 (2006), pp. 694–699.
- [40] A. Adi, D. Botzer, G. Nechushtai e G. Sharon. “Complex event processing for financial services”. Em: *Services Computing Workshops, 2006. SCW’06. IEEE*. IEEE. 2006, pp. 7–12.

- [41] D. Wang, E. A. Rundensteiner, H. Wang e R. T. Ellison III. “Active complex event processing: applications in real-time health care”. Em: *Proceedings of the VLDB Endowment* 3.1-2 (2010), pp. 1545–1548.
- [42] S. Meinecke. “Real-time monitoring of clinical processes using complex event processing and transition systems.” Em: *MIE*. 2014, pp. 604–608.
- [43] Y. Zhang, Q. Dai e R. Zhong. “An extensible event-driven manufacturing management with complex event processing approach”. Em: *International Journal of Control and Automation* 2.3 (2009), pp. 1–12.
- [44] Y. Wang e K. Cao. “A proactive complex event processing method for large-scale transportation internet of things”. Em: *International Journal of Distributed Sensor Networks* 10.3 (2014), p. 159052.
- [45] J. Greenberg. “Understanding metadata and metadata schemes”. Em: *Cataloging & classification quarterly* 40.3-4 (2005), pp. 17–36.
- [46] G. K. Tayi e D. P. Ballou. “Examining data quality”. Em: *Communications of the ACM* 41.2 (1998), pp. 54–57.
- [47] P. Oliveira, F. Rodrigues e P. Henriques. “Limpeza de dados: Uma visão geral”. Em: *Data Gadgets* (2004), pp. 39–51.
- [48] P. Jermyn, M. Dixon e B. J. Read. “Preparing clean views of data for data mining”. Em: *ERCIM Work. on Database Res* (1999), pp. 1–15.
- [49] E. Rahm e H. H. Do. “Data cleaning: Problems and current approaches”. Em: *IEEE Data Eng. Bull.* 23.4 (2000), pp. 3–13.
- [50] “Population density”. Em: European Environment Agency. 2011.
- [51] W. Documentation. *Architecturei*. Accessed: 2017-09-23. 2015. URL: <https://docs.wso2.com/display/CEP400/Architecture>.

## ANEXO 1 - TABELAS DE METADADOS

Tabela I.1: Descrição dos campos da base de dados de eventos rodoviários

Metadados	Descrição
_id	Identificador único do documento
Cesta	Nome da estrada, especificando o trecho
Description	Descrição da estrada
IsRoadClosed	Identificador de estrada fechada. 0=normal, 1=estrada fechada
PrioritetaCeste	Estado numérico da classificando a prioridade da estrada
y_wgs	Campo correspondente ao valor de latitude, sobre a projeção WGS
X	Campo correspondente ao valor de longitude, sobre a projeção ESPG:2170
Y	Campo correspondente ao valor de latitude, sobre a projeção ESPG:2170
Kategorija	Nome da estrada
x_wgs	Campo correspondente ao valor de longitude, sobre a projeção WGS
Updated	Data da última atualização dos dados. yyyy-MM-dd'T'hh:mm:ss.SSS+XXXX
Title	Categorização do evento
VeljavnostOd	Data de início de evento. yyyy-MM-dd'T'hh:mm:ss.SSS+XXXX
Id	Número identificador do evento
VeljavnostDo	Data de fim de evento. yyyy-MM-dd'T'hh:mm:ss.SSS+XXXX
Prioriteta	Estado numérico da nível do acidente

Tabela I.2: Descrição dos campos da base de dados de valores comportamentais de veículos, com cobertura temporal de dezembro de 2016 a maio de 2017

<b>Metadados</b>	<b>Descrição</b>
_id	Identificador único do documento
updated	Representação numérica da última data de atualização do ficheiro
copyright	Descrição do direito autorial dos dados
Language	Idioma dos dados
ModifiedTime	Data da criação do dados. yyyy-MM-dd'T'hh:mm:ssSSSSSS'Z'
IsModified	Identificador de modificação dos dados
ContentName	Categoria do sensor
Etag	Identificador de cache em memória
y_wgs	Campo correspondente ao valor de latitude, sobre a projeção WGS
Description	Descrição da estrada
Title	Nome completo da estrada
ContentName	Descrição do tipo sensor
x_wgs	Campo correspondente ao valor de longitude, sobre a projeção WGS
CrsId	Código identificador do sistema de referência de coordenadas
stevci_cestaOpis	Nome da estrada
Y	Campo correspondente ao valor de latitude, sobre a projeção ESPG:2170
X	Campo correspondente ao valor de longitude, sobre a projeção ESPG:2170
stevci_hit	Velocidade média dos veículos, expresso em Km/h
stevci_stev	Números de veículos, expresso em Veículos/h
stevci_pasOpis	Descrição da faixa segundo a localização do sensor
stevci_smerOpis	Direção da estrada
stevci_stat	Estado numérico do tráfego
Id	Identificar do sensor



Tabela I.3: Descrição dos campos da base de dados de valores comportamentais de veículos, com cobertura temporal de janeiro de 2016 a novembro de 2016

Metadados	Descrição
_id	Identificador único do documento
updated	Representação numérica da última data de atualização do ficheiro
title	Tipo de sensor
stevci_vmax	Velocidade máxima na estrada, expresso em Km/h
stevci_smerOpis	Descrição da direção da estrada
id	Identificador único do sensor. Contem também concatenado a direção expressa numericamente
stevci_ura	Tempo da última medição de dados do sensor
stevci_gap	Diferença de tempo entre veículos, expresso em segundos
georss_point	Coordenadas do sensor dentro da Eslovénia
stevci_datum	Data da última medição do sensor
stevci_cestaOpis	Nome da estrada
stevci_geoY_wgs	Latitude do sensor segundo a projeção WGS
stevci_regija	Nome da região onde se localiza o sensor
updated	Data da última medição do sensor, expressa num long
stevci_stacionaza	Distância desde o início da estrada até à localização do sensor
stevci_lokacijaOpis	Identificador único do sensor
stevci_odsek	Número da secção da estrada
stevci_stev	Números de veículos, expresso em Veículos/h
stevci_smer	Direção da estrada
stevci_stat	Estado numérico do tráfego
stevci_geoX_wgs	Longitude do sensor segundo a projeção WGS
stevci_occ	Ocupação da estrada. Gama de valores 0-100
stevci_geoY	Campo correspondente ao valor de latitude, sobre a projeção ESPG:2170
stevci_geoX	Campo correspondente ao valor de longitude, sobre a projeção ESPG:2170
stevci_statOpis	Codificação HTML em linguagem natural da descrição do estado do tráfego na estrada.
stevci_hit	Velocidade média dos veículos, expresso em Km/h
summary	Sumarização do tráfego
stevci_pasOpis	Descrição da faixa segundo a localização do sensor
stevci_lokacija	Identificador único do sensor
title	Nome completo da estrada
copyright	Descrição do direito autorial dos dados





## ANNEX 2 - CÓDIGO DO BLOCO PROCESSADOR DE EVENTOS

```
1
2 /* Enter a unique ExecutionPlan */
3 @Plan:name('DetectionEvent')
4
5 /* Enter a unique description for ExecutionPlan */
6 -- @Plan:description('ExecutionPlan')
7
8 /* define streams/tables and write queries here ... */
9
10 @Plan:statistics('true')
11
12 @Plan:trace('true')
13
14
15
16
17
18
19
20
21
22
23 @Export('FinishedEvents:1.0.0')
24 define stream FinishedEvents (ModifiedTime string, Id int, Road int);
25
26 @Export('EventBegin:1.0.0')
27 define stream EventBegin (ModifiedTime string, Id int, Road int);
28
29 --@Export('FinalTeste1:1.0.0')
```

## ANEXO II. ANNEX 2 - CÓDIGO DO BLOCO PROCESSADOR DE EVENTOS

```

30 --define stream FinalTeste1 (ModifiedTime string, Id int, Road int, Hour int);
31
32 define stream LastCounterValue (ModifiedTime string, Hit_11 double,
33     Hit_12 double, Hit_13 double, Hit_21 double,
34     Hit_22 double, Hit_23 double, Hit11_last double,
35     Hit12_last double, Hit13_last double,
36     Hit21_last double, Hit22_last double,
37     Hit23_last double, Occu_11 double,
38     Occu_12 double, Occu_13 double, Occu_21 double,
39     Occu_22 double, Occu_23 double,
40     Occu11_last double, Occu12_last double,
41     Occu13_last double, Occu21_last double,
42     Occu22_last double, Occu23_last double,
43     Id int, Day int, Hour int);
44
45 define stream PenultimateCounterValue (ModifiedTime string, Hit_11 double,
46     Hit_12 double, Hit_13 double,
47     Hit_21 double, Hit_22 double,
48     Hit_23 double, Hit11_last double,
49     Hit12_last double, Hit13_last double,
50     Hit21_last double, Hit22_last double,
51     Hit23_last double, Hit11_pen double,
52     Hit12_pen double, Hit13_pen double,
53     Hit21_pen double, Hit22_pen double,
54     Hit23_pen double, Occu_11 double,
55     Occu_12 double, Occu_13 double,
56     Occu_21 double, Occu_22 double,
57     Occu_23 double, Occu11_last double,
58     Occu12_last double, Occu13_last double,
59     Occu21_last double, Occu22_last double,
60     Occu23_last double, Occu11_pen double,
61     Occu12_pen double, Occu13_pen double,
62     Occu21_pen double, Occu22_pen double,
63     Occu23_pen double, Id int, Day int,
64     Hour int);
65
66
67 @Import('CounterValues:1.0.0')
68 define stream CounterValues (ModifiedTime string, Day int, Hour int, Id int,
69     Hit_11 double, Hit_12 double, Hit_13 double,
70     Hit_21 double, Hit_22 double, Hit_23 double,
71     Occu_11 double, Occu_12 double, Occu_13 double,
72     Occu_21 double, Occu_22 double, Occu_23 double);
73
74 @Import('AverageReadingPoints:1.0.0')
75 define stream AverageReadingPoints (Id int, Road int, dayOfTheWeek int,
76     hour int, velocity_data0 double,
77     velocity_data1 double,
78     ocupacation_data0 double,
79     ocupacation_data1 double);

```

```

80
81 define table Average (Id int, Road int, dayOfTheWeek int, hour int,
82     velocity_data0 double, velocity_data1 double,
83     ocupacation_data0 double, ocupacation_data1 double);
84
85 define table EventsTable(ModifiedTime string, Id int, Road int, Day int,
86     Hour int, publisher bool);
87
88 define table EventsTable2(ModifiedTime string, Id int, Road int, Day int,
89     Hour int, publisher bool);
90
91 from AverageReadingPoints
92 select *
93 insert into Average;
94
95
96
97 from every( e1=CounterValues) -> e2=CounterValues[e1.Id==Id]<1:> ->
98 e3=CounterValues[e1.Id==Id]<1:>
99 select e3[last].ModifiedTime, e3[last].Hit_11, e3[last].Hit_12,
100 e3[last].Hit_13, e3[last].Hit_21, e3[last].Hit_22,
101 e3[last].Hit_23, e2[last].Hit_11 as Hit11_last,
102 e2[last].Hit_12 as Hit12_last, e2[last].Hit_13 as Hit13_last,
103 e2[last].Hit_21 as Hit21_last, e2[last].Hit_22 as Hit22_last, e2[last].Hit_23
104 as Hit23_last, e1.Hit_11 as Hit11_pen, e1.Hit_12 as Hit12_pen, e1.Hit_13
105 as Hit13_pen, e1.Hit_21 as Hit21_pen, e1.Hit_22 as Hit22_pen, e1.Hit_23
106 as Hit23_pen, e3[last].Occu_11, e3[last].Occu_12, e3[last].Occu_13,
107 e3[last].Occu_21, e3[last].Occu_22, e3[last].Occu_23, e2[last].Occu_11 as
108 Occu11_last, e2[last].Occu_12 as Occu12_last, e2[last].Occu_13 as
109 Occu13_last, e2[last].Occu_21 as Occu21_last,
110 e2[last].Occu_22 as Occu22_last, e2[last].Occu_23 as Occu23_last,
111 e1.Occu_11 as Occu11_pen, e1.Occu_12 as Occu12_pen, e1.Occu_13 as Occu13_pen,
112 e1.Occu_21 as Occu21_pen, e1.Occu_22 as Occu22_pen,
113 e1.Occu_23 as Occu23_pen,
114 e1.Id, e1.Day, e1.Hour
115 insert into PenultimateCounterValue;
116
117
118
119 from every( e1=CounterValues) -> e2=CounterValues[e1.Id==Id]<1:>
120 select e2[last].ModifiedTime, e2[last].Hit_11, e2[last].Hit_12,
121 e2[last].Hit_13, e2[last].Hit_21, e2[last].Hit_22, e2[last].Hit_23,
122 e1.Hit_11 as Hit11_last, e1.Hit_12 as Hit12_last, e1.Hit_13 as Hit13_last,
123 e1.Hit_21 as Hit21_last, e1.Hit_22 as Hit22_last, e1.Hit_23 as Hit23_last,
124 e2[last].Occu_11, e2[last].Occu_12, e2[last].Occu_13, e2[last].Occu_21,
125 e2[last].Occu_22, e2[last].Occu_23, e1.Occu_11 as Occu11_last,
126 e1.Occu_12 as Occu12_last, e1.Occu_13 as Occu13_last,
127 e1.Occu_21 as Occu21_last, e1.Occu_22 as Occu22_last,
128 e1.Occu_23 as Occu23_last, e2[last].Id, e2[last].Day,
129 e2[last].Hour

```

## ANEXO II. ANNEX 2 - CÓDIGO DO BLOCO PROCESSADOR DE EVENTOS

```

130 insert into LastCounterValue;
131
132
133 --Detection of start of the event in road 1 (11,12,13)
134
135 from LastCounterValue
136 join Average
137     on (LastCounterValue.Day == Average.dayOfTheWeek
138         and LastCounterValue.Hour == Average.hour
139         and LastCounterValue.Id == Average.Id
140         and LastCounterValue.Hit_11 < (Average.velocity_data0*0.9)
141         and LastCounterValue.Hit11_last < (Average.velocity_data0*0.9)
142         and Average.Road == 1 and LastCounterValue.Occu_11>300
143         and LastCounterValue.Occu11_last>300)
144 or (LastCounterValue.Day == Average.dayOfTheWeek
145     and LastCounterValue.Hour == Average.hour
146     and LastCounterValue.Id == Average.Id
147     and LastCounterValue.Hit_12 < (Average.velocity_data1*0.9)
148     and LastCounterValue.Hit12_last < (Average.velocity_data1*0.9)
149     and Average.Road == 1 and LastCounterValue.Occu_12>300
150     and LastCounterValue.Occu12_last>300)
151 --Detection of start of the event in road 2 (21,22,23)
152 or (LastCounterValue.Day == Average.dayOfTheWeek
153     and LastCounterValue.Hour == Average.hour
154     and LastCounterValue.Id == Average.Id
155     and LastCounterValue.Hit_22 < (Average.velocity_data1*0.9)
156     and LastCounterValue.Hit22_last < (Average.velocity_data1*0.9)
157     and Average.Road == 2 and LastCounterValue.Occu_21>300
158     and LastCounterValue.Occu21_last>300)
159 or (LastCounterValue.Day == Average.dayOfTheWeek
160     and LastCounterValue.Hour == Average.hour
161     and LastCounterValue.Id == Average.Id
162     and LastCounterValue.Hit_22 < (Average.velocity_data1*0.9)
163     and LastCounterValue.Hit22_last < (Average.velocity_data1*0.9)
164     and Average.Road == 2 and LastCounterValue.Occu_22>300
165     and LastCounterValue.Occu22_last>300)
166 select convert(time:dateFormat(LastCounterValue.ModifiedTime,
167     'yyyy-MM-dd_HH:mm:ss', 'dd_MMM_yyyy_HH_mm_ss'), 'string')
168     as ModifiedTime, LastCounterValue.Id, Average.Road,
169     LastCounterValue.Hour
170 insert into AllEvents;
171
172
173 --Detection of end of the event in road 1 (11,12)
174 from PenultimateCounterValue
175 join Average
176     on (PenultimateCounterValue.Day == Average.dayOfTheWeek
177         and PenultimateCounterValue.Hour == Average.hour
178         and PenultimateCounterValue.Id == Average.Id
179         and Average.Road == 1

```

```

180     and ( PenultimateCounterValue.Hit_11 >= (Average.velocity_data0*0.9)
181         and PenultimateCounterValue.Hit11_last < (Average.velocity_data0*0.9)
182         and PenultimateCounterValue.Hit11_pen < (Average.velocity_data0*0.9)
183         and PenultimateCounterValue.Occu_11>300
184         and PenultimateCounterValue.Occu11_last>300
185         and PenultimateCounterValue.Occu11_pen>300)
186 or (PenultimateCounterValue.Day == Average.dayOfTheWeek
187     and PenultimateCounterValue.Hour == Average.hour
188     and PenultimateCounterValue.Id == Average.Id
189     and Average.Road == 1)
190     and PenultimateCounterValue.Hit_12 >= (Average.velocity_data1*0.9)
191     and PenultimateCounterValue.Hit12_last < (Average.velocity_data1*0.9)
192     and PenultimateCounterValue.Hit12_pen < (Average.velocity_data1*0.9)
193     and PenultimateCounterValue.Occu_12>300
194     and PenultimateCounterValue.Occu12_last>300
195     and PenultimateCounterValue.Occu12_pen>300)
196     --Detection of end of the event in road 2 (21,22)
197 or (PenultimateCounterValue.Day == Average.dayOfTheWeek
198     and PenultimateCounterValue.Hour == Average.hour
199     and PenultimateCounterValue.Id == Average.Id and Average.Road == 2
200     and ( PenultimateCounterValue.Hit_21 >= (Average.velocity_data0*0.9)
201         and PenultimateCounterValue.Hit21_last < (Average.velocity_data0*0.9)
202         and PenultimateCounterValue.Hit21_pen < (Average.velocity_data0*0.9))
203     and PenultimateCounterValue.Occu_21>300
204     and PenultimateCounterValue.Occu21_last>300
205     and PenultimateCounterValue.Occu21_pen>300)
206 or (PenultimateCounterValue.Day == Average.dayOfTheWeek
207     and PenultimateCounterValue.Hour == Average.hour
208     and PenultimateCounterValue.Id == Average.Id
209     and Average.Road == 2
210     and PenultimateCounterValue.Hit_22 >= (Average.velocity_data1*0.9)
211     and PenultimateCounterValue.Hit22_last < (Average.velocity_data1*0.9)
212     and PenultimateCounterValue.Hit22_pen < (Average.velocity_data1*0.9)
213     and PenultimateCounterValue.Occu_22>300
214     and PenultimateCounterValue.Occu22_last>300
215     and PenultimateCounterValue.Occu22_pen>300)
216 select convert(time:dateFormat(PenultimateCounterValue.ModifiedTime ,
217     'yyyy-MM-dd_HH:mm:ss', 'dd_MMM_yyyy_HH_mm_ss'), 'string')
218     as ModifiedTime, PenultimateCounterValue.Id,
219     Average.Road, PenultimateCounterValue.Hour
220 insert into AllEventsFinished;
221
222
223
224
225 from AllEvents[not (( time:extract('day',time:dateFormat(AllEvents.ModifiedTime ,
226     'yyyy-MM-dd_HH:mm:ss', 'yyyy-MM-dd_HH:mm:ss'),
227     'yyyy-MM-dd_hh:mm:ss') == EventsTable.Day
228     and Hour <= EventsTable.Hour+3
229     and Id == EventsTable.Id) in EventsTable)]

```

## ANEXO II. ANNEX 2 - CÓDIGO DO BLOCO PROCESSADOR DE EVENTOS

```

230 select AllEvents.ModifiedTime, AllEvents.Id, AllEvents.Road,
231        time:extract('day',time:dateFormat(AllEvents.ModifiedTime,
232        'yyyy-MM-dd_LHH:mm:ss', 'yyyy-MM-dd_LHH:mm:ss'),'yyyy-MM-dd_Lhh:mm:ss')
233        as Day, AllEvents.Hour,
234        False as publisher
235 insert into EventsTable;
236
237
238 from AllEventsFinished[not (( time:extract('day',time:dateFormat
239        (AllEventsFinished.ModifiedTime,
240        'yyyy-MM-dd_LHH:mm:ss', 'yyyy-MM-dd_LHH:mm:ss'),
241        'yyyy-MM-dd_Lhh:mm:ss') == EventsTable2.Day
242        and Hour <= EventsTable2.Hour+3
243        and Id == EventsTable2.Id) in EventsTable2)]
244 select AllEventsFinished.ModifiedTime, AllEventsFinished.Id,
245        AllEventsFinished.Road, time:extract('day',
246        time:dateFormat(AllEventsFinished.ModifiedTime,
247        'yyyy-MM-dd_LHH:mm:ss', 'yyyy-MM-dd_LHH:mm:ss'),
248        'yyyy-MM-dd_Lhh:mm:ss') as Day, AllEventsFinished.Hour,
249        False as publisher
250 insert into EventsTable2;
251
252
253
254 from AllEvents
255 join EventsTable
256     on EventsTable.publisher == False
257 select EventsTable.ModifiedTime, EventsTable.Id, EventsTable.Road,
258 EventsTable.Hour
259 insert into EventBeginFinal;
260
261
262
263 from AllEventsFinished
264 join EventsTable2
265     on EventsTable2.publisher == False
266 select EventsTable2.ModifiedTime, EventsTable2.Id, EventsTable2.Road,
267 EventsTable2.Hour
268 insert into EventFinishedFinal;
269
270
271 from EventFinishedFinal
272 select EventFinishedFinal.ModifiedTime, EventFinishedFinal.Id,
273 EventFinishedFinal.Road, time:extract('day',
274        time:dateFormat(EventFinishedFinal.ModifiedTime,
275        'yyyy-MM-dd_LHH:mm:ss', 'yyyy-MM-dd_LHH:mm:ss'),
276        'yyyy-MM-dd_Lhh:mm:ss') as Day,
277        EventFinishedFinal.Hour, True as publisher
278 update EventsTable2
279     on EventsTable2.publisher== False;

```



---

```
280
281
282 from EventBeginFinal
283 select EventBeginFinal.ModifiedTime, EventBeginFinal.Id,
284 EventBeginFinal.Road, time:extract('day',
285     time:dateFormat(EventBeginFinal.ModifiedTime,
286     'yyyy-MM-dd\HH:mm:ss', 'yyyy-MM-dd\HH:mm:ss'),
287     'yyyy-MM-dd\hh:mm:ss') as Day, EventBeginFinal.Hour,
288     True as publisher
289 update EventsTable
290     on EventsTable.publisher== False;
291
292
293 from EventBeginFinal
294 select FinalTeste1.ModifiedTime, FinalTeste1.Id, FinalTeste1.Road
295 insert into EventBegin;
296
297 from EventFinishedFinal
298 select FinalTeste2.ModifiedTime, FinalTeste2.Id, FinalTeste2.Road
299 insert into FinishedEvents;
```