INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Identificação de Fatores de Risco na Gravidade dos Acidentes Rodoviários

Inês Mendes dos Santos Rodrigues de Carvalho

Mestrado em Sistemas Integrados de Apoio à Decisão

Orientador:
Doutor Fernando Manuel Marques Batista, Professor Associado, ISCTE-IUL - Departamento de Ciências e Tecnologias da Informação

Coorientador:
Doutor Luís Miguel Martins Nunes, Professor Associado, ISCTE-IUL - Departamento de Ciências e Tecnologias da Informação

Outubro, 2022

TECNOLOGIAS
E ARQUITETURA

Departamento de Ciências e Tecnologias da Informação

Identificação de Fatores de Risco na Gravidade dos Acidentes Rodoviários

Inês Mendes dos Santos Rodrigues de Carvalho

Mestrado em Sistemas Integrados de Apoio à Decisão

Orientador:
Doutor Fernando Manuel Marques Batista, Professor Associado, ISCTE-IUL - Departamento de Ciências e Tecnologias da Informação

Coorientador:
Doutor Luís Miguel Martins Nunes, Professor Associado, ISCTE-IUL - Departamento de Ciências e Tecnologias da Informação

Outubro, 2022

Agradecimentos

Desejo expressar o meu sincero agradecimento aos meus orientadores, Prof. Dr. Fernando Manuel Marques Batista e Prof. Dr. Luís Miguel Martins Nunes, bem como ao Prof. Dr. João C. Ferreira, por todo o apoio e aconselhamento na realização deste trabalho. Sem a sua ajuda persistente, o objetivo deste projeto não teria sido alcançado.

Agradeço também à ANSR, pela disponibilização dos dados que permitiram que este estudo fosse possível.

Por fim, gostaria de agradecer o apoio incondicional da minha família, em particular ao meu pai, Mário Carvalho, pela sua ajuda e genialidade.

Resumo

Os acidentes rodoviários são uma das principais causas mundiais de feridos e vítimas mortais e representam um grande custo económico e social. Por esta razão, tornaram-se num importante campo de investigação no sentido da utilização da análise de acidentes e de técnicas preditivas para a determinação dos fatores mais importantes que contribuem para a sua ocorrência. Neste contexto, este trabalho realiza uma análise aos dados dos acidentes rodoviários ocorridos entre 2010 e 2019 e reportados à ANSR pelas forças de segurança (GNR e PSP) através do Boletim Estatístico de Acidentes de Viação (BEAV), com o objetivo principal de determinar os fatores de risco que influenciam a proporção de acidentes graves. Foram colocadas duas hipóteses de trabalho que se prendem com o aumento da velocidade de circulação, condicionado pela densidade de tráfego, e com a estrutura da rede rodoviária. Para testar estas hipóteses, fez-se uma análise espacial e temporal da proporção de acidentes graves. Os resultados obtidos indicam de forma coerente que todas as situações que proporcionam uma redução da densidade de tráfego aumentam o risco de acidente grave. A análise dos dados revelou também que as características da infraestrutura rodoviária influencia a gravidade dos acidentes.

Palavras-chave: Segurança Rodoviária, Acidentes Rodoviários, Fatores de Risco, Análise de Dados, Modelos de Predição de Acidentes

Abstract

Road accidents are one of the world's leading causes of injuries and fatalities and represent a large economic and social cost. For this reason, they have become an important field of research towards the use of accident analysis and predictive techniques to determine the most important factors that contribute to their occurrence. In this context, this paper conducts an analysis of the data of road accidents that occurred between 2010 and 2019 and reported to ANSR by the security forces (GNR and PSP) through the Statistical Bulletin of Traffic Accidents (BEAV), with the main objective of determining the risk factors that influence the proportion of serious accidents. Two working hypotheses were put forward that are related to the increase in traffic speed, conditioned by traffic density, and to the structure of the road network. To test these hypotheses, a spatial and temporal analysis of the proportion of severe accidents was performed. The results obtained consistently indicate that all situations that provide a reduction in traffic density increase the risk of a severe accident. Data analysis also revealed that the characteristics of the road infrastructure influence the severity of accidents.

Keywords: Road Safety, Traffic Accidents, Risk Factors, Data Analysis, Crash Prediction Model

Índice

1 INTRODUÇÃO 1

1.1 ENQUADRAMENTO DO PROBLEMA E MOTIVAÇÃO 1

1.2 ABORDAGEM *SAFE SYSTEM* 2

1.3 OBJETIVOS E HIPÓTESES DE INVESTIGAÇÃO 5

1.4 METODOLOGIA 6

2 REVISÃO DA LITERATURA 7

2.1 METODOLOGIA 7

2.2 FATORES DE RISCO NOS ACIDENTES RODOVIÁRIOS 9

2.2.1 Metodologias utilizadas na identificação de fatores de risco 12

2.3 MEDIDAS DE SEGURANÇA 12

2.4 DADOS DOS ACIDENTES RODOVIÁRIOS 13

2.4.1 Qualidade dos Dados 15

2.5 MODELOS PREDITIVOS 16

3 ANÁLISE E VISUALIZAÇÃO DOS DADOS 19

3.1 QUALIDADE DOS DADOS 20

3.2 PREPARAÇÃO DOS DADOS 21

3.2.1 Dados dos acidentes 21

3.2.2 Dados dos veículos 21

3.3 VISUALIZAÇÃO DOS DADOS 22

3.3.1 Caracterização Geral 23

3.3.2 Análise Espacial e das Infraestruturas 25

3.3.3 Análise Temporal 33

4 MODELAÇÃO DE DADOS 37

4.1 PREDIÇÃO DE ACIDENTES EM FUNÇÃO DAS CARACTERÍSTICAS DOS VEÍCULOS 38

4.2 PREDIÇÃO DE ACIDENTES EM FUNÇÃO DAS CARACTERÍSTICAS DOS ACIDENTES 40

5 CONCLUSÕES E TRABALHO FUTURO 43

Índice de Figuras

Figura 1 - CRISP-DM [11] 6
Figura 2 - Metodologia da Revisão da Literatura 7
Figura 3 - Fatores que influenciam a segurança rodoviária [14] 9
Figura 4 - Fatores que contribuem para os acidentes rodoviários 10
Figura 5 - Fontes de dados dos acidentes rodoviários 14
Figura 6 – Fontes de dados alternativas 15
Figura 7 - Posicionamento dos dados dos acidentes com coordenadas entre 2010 e 2019 20
Figura 8 - Número total de acidentes por distrito 23
Figura 9 - Número de acidentes graves por distrito 24
Figura 10 - Proporção de acidentes graves no total de acidentes por distrito 25
Figura 11 - Número e proporção de acidentes graves dentro e fora das localidades nos distritos do Porto e Portalegre 25
Figura 12 - Características da rodoviária no Porto e em Portalegre - INE 26
Figura 13 - Acidentes graves por ações dos condutores 27
Figura 14 - Proporção de acidentes por tipologia de estrada no Porto e em Portalegre 28
Figura 15 - Índice de sinuosidade por distrito 29
Figura 16 - Cobertura da rede rodoviária por distrito 29
Figura 17 - Proporção de acidentes graves por estado de conservação das vias dentro e fora das localidades, respetivamente 30
Figura 18 -Proporção de acidentes graves por natureza do acidente dentro e fora das localidades, respetivamente 31
Figura 19 - Correlações entre as variáveis em todos os distritos 32
Figura 20 - Evolução da proporção de acidentes graves 33
Figura 21 - Nº de acidentes, nº de acidentes graves e proporção de acidentes graves por dia da semana, respetivamente 34
Figura 22 - Número e Proporção de acidentes graves por hora do dia 34
Figura 23 - Proporção de acidentes graves por hora do dia ao fim de semana e durante a semana 35
Figura 24 - Modelos preditivos na análise de acidentes rodoviários 37
Figura 25 - Correlações entre as características dos condutores/veículos e a gravidade 39
Figura 26 - Resultados da predição em função das características dos veículos 40
Figura 27- Matriz de confusão da predição em função das características dos veículos 40

Figura 28 - Resultados da predição em função das características dos acidentes 41
Figura 29 - Matriz de confusão da predição em função das características dos acidentes 41

Índice de Tabelas

Tabela 1 - Princípios chave do Safe System[8] 3
Tabela 2 - KPI's para avaliar o desempenho dos estados-membros no âmbito do Safe System 4
Tabela 3 - Revisão da literatura nos fatores de risco dos acidentes rodoviários (apenas foram incluídos os fatores mais relevantes) 11
Tabela 4 - Metodologias de identificação de fatores de risco 12
Tabela 5 - Variáveis independentes mais utilizadas em cada tipo de modelo 18
Tabela 6 - Organização dos dados 19
Tabela 7 - Dados dos acidentes 48
Tabela 8 - Dados dos veículos 49
Tabela 9 - Dados dos passageiros 51
Tabela 10 - Dados dos peões 52

1 Introdução

Os acidentes rodoviários são uma das principais causas mundiais de feridos e vítimas mortais e, por conseguinte, traduzem-se num importante campo de investigação no sentido da utilização da análise de acidentes e de técnicas preditivas para a determinação dos fatores mais importantes que contribuem para a sua ocorrência [1].

Os acidentes rodoviários têm um grande custo económico e social, especialmente os acidentes graves, uma vez que estes podem provocar mortes ou incapacidades permanentes.

Os meios de transporte rodoviários ocupam um lugar muito importante na deslocação das pessoas. O aumento da população com título de condução, o crescimento excessivo do número de veículos, o funcionamento da economia em espaço aberto, os desafios da competitividade empresarial e da sustentabilidade dos negócios, os recursos naturais atrativos para o turismo, o crescimento do poder de compra e a maior facilidade de acesso ao crédito, têm contribuído para o grande número de viaturas diariamente em circulação em Portugal [2].

Apesar das tentativas do governo para incentivar o uso dos transportes públicos e bicicletas, os veículos ligeiros continuam a circular em massa. Segundo os dados do PORDATA, Portugal ocupava, em 2019, a 13ª posição na União Europeia (UE) de veículos ligeiros de passageiros, com um valor de 5.452 milhares [3].

1.1 Enquadramento do Problema e Motivação

"Se no final de cada dia de trabalho conseguirmos salvar uma vida, estaremos a dar um enorme contributo para um futuro melhor" [4]. As lesões causadas por acidentes rodoviárias, embora negligenciadas, representam um dos maiores problemas na saúde pública, que requer esforços concentrados para uma prevenção eficaz e sustentável [5].

De acordo com o *Road Safety Manual* [6] mais de 1,24 milhões de pessoas morrem todos os anos na estrada e, entre 20 e 50 milhões de pessoas sofrem, anualmente, de lesões provocadas por acidentes rodoviários. Os acidentes rodoviários representam uma das principais causas de morte para as crianças e são a principal causa de morte nos jovens entre os 15 e os 29 anos. Atualmente representam a 9ª causa de morte a nível global e prevê-se que, a este ritmo, os acidentes rodoviários

sejam a 7ª causa de morte mais comum para a população global em 2030. O custo humano e social causado pela mobilidade é alto e insustentável. Em termos financeiros, o custo anual dos acidentes rodoviários na UE é de 280 bilhões de euros, ou cerca de 2% do PIB [3].

A nível nacional, os acidentes rodoviários registados no ano de 2019 tiveram um custo económico e social para o país estimado em 6.422,9 milhões de euros, o que representa 3,03% da riqueza criada em Portugal nesse ano. Desse custo total, a maior fatia (83,5% do total) é referente a acidentes com vítimas, totalizando 5.362,7 milhões de euros (2,53% do PIB), sendo que os restantes 1.060,1 milhões de euros (0,5% do PIB) dizem respeito a acidentes sem vítimas que geraram apenas danos patrimoniais. Entre os acidentes com vítimas, a maior componente do custo total, representando 64,7% desse valor, respeita aos custos humanos estimados em 3.471,1 milhões de euros (1,635% do PIB). A segunda componente mais expressiva (representando 26,8% do total) refere-se à perda bruta de produção, estimada em 1.438 milhões de euros (0,677% do PIB) em 2019. No seu conjunto, os custos humanos e a perda bruta de produção representam 91,5% do custo total dos acidentes de viação em Portugal [2].

1.2 Abordagem *Safe System*

A abordagem *Safe System* é uma boa prática e um paradigma orientador na abordagem da segurança rodoviária, centrando-se na prevenção de mortes e ferimentos graves.

"Não aceitamos mortes no ar, nem na ferrovia e também não devemos mais aceitá-las na estrada" [4]. Em resposta a este desafio de saúde pública, a estratégia da Visão Zero procura eliminar as mortes no trânsito e ferimentos graves [7].

A Visão Zero faz parte da Abordagem *Safe System,* abordagem pioneira na Suécia e na Holanda. Na Suécia, a estratégia Visão Zero foi adotada pelo Parlamento em 1997. Nos Países Baixos, uma política semelhante, intitulada Segurança Sustentável, foi desenvolvida na década de 1990 pelo Instituto Holandês de Investigação em Segurança Rodoviária (SWOV). A Abordagem *Safe System* foi aprovada pela Comissão Europeia em 2018 [7].

O *Safe System* parte do princípio de que as pessoas cometem erros e defende uma abordagem holística assente numa diversidade de medidas que visam evitar as mortes e os feridos graves, causadas por esses erros, e que têm em conta a vulnerabilidade humana [7]. É construído em torno da premissa de que a morte e os ferimentos graves são inaceitáveis e evitáveis e pretende acabar com a

visão passada de que as mortes são um preço a pagar pela mobilidade, tendo como objetivo reduzir as mortes a zero a longo prazo [6].

Consiste num plano estratégico a longo prazo com ações a serem desenvolvidas e implementadas em sucessivos planos intermédios, com responsabilidade partilhada entre os vários participantes do sistema, que apresenta como princípios chave:

Tabela 1 - Princípios chave do Safe System [7]

Os seres humanos cometem erros	As pessoas cometem erros, voluntária e involuntariamente. Esses erros podem levar à ocorrência de acidentes.
Os seres humanos são vulneráveis	O corpo humano tem uma capacidade limitada de tolerar as forças de uma colisão sem que ocorram lesões.
Responsabilidade partilhada	Existe uma responsabilidade compartilhada entre aqueles que projetam, constroem, gerem e utilizam a rede rodoviária e os veículos, incluindo aqueles que fornecem cuidados pós-acidentes, para prevenir acidentes que causam lesões graves ou morte.
Abordagem proativa	Deve ser adotada uma abordagem proativa em vez de esperar que os acidentes ocorram para reagir. Todas as partes do sistema devem ser reforçadas para multiplicar os seus efeitos; e se uma parte falhar, os utentes da estrada continuam a estar protegidos.
Nenhuma morte ou lesão grave é aceitável	Nenhuma morte ou lesão grave é aceitável no sistema de mobilidade. A falta de segurança não deve ser o preço a pagar por uma maior mobilidade e o sistema de mobilidade deve ser tão seguro quanto eficiente.

A Comissão Europeia definiu um conjunto de *KPI´s* (Tabela 2) para avaliar e monitorizar a desem dos estados-membros no âmbito do *Safe System* [8].

Em Portugal, a Autoridade Nacional de Segurança Rodoviária (ANSR) tem vindo, nos últimos anos, a desenvolver a sua estratégia tendo por base os princípios do *Safe System.* O seu mais recente quadro de política de segurança rodoviária, em vigor para a década de 2021 a 2030, baseia-se nesta abordagem. A Visão Zero 2030 pretende fixar o horizonte de médio prazo da política de segurança rodoviária em Portugal e definir os objetivos estratégicos e operacionais correspondentes, a concretizar através de planos de ação bienais a iniciar em 2021 e abrangendo um período total de 10 anos.

A ANSR, no cumprimento das suas atribuições de planeamento e coordenação de apoio à política do Governo em matéria de Segurança Rodoviária, está a definir a Estratégia Nacional de Segurança Rodoviária 2021-2030 – Visão Zero 2030, alinhada com a política de segurança rodoviária da Comissão Europeia 2021-2030, com a Declaração de Estocolmo e, com a adoção de uma visão virada para a criação de um *Safe System,* tem sido notória na redução das mortes a nível global.

Tabela 2 - KPI's para avaliar o desempenho dos estados-membros no âmbito do Safe System

Velocidade	Percentagem de veículos que viajam dentro do limite de velocidade.
Cinto de segurança	Percentagem de ocupantes do veículo que utilizam corretamente o cinto de segurança ou o sistema de retenção para crianças.
Equipamento de proteção	Percentagem de utilizadores com capacete de proteção em veículos de duas rodas.
Álcool	Percentagem de condutores que conduzem dentro do limite legal para o teor de álcool no sangue (BAC).
Distração	Percentagem de condutores que não utiliza um dispositivo móvel portátil.
Cuidados pós-acidente	O tempo que decorreu entre a chamada de emergência e a chegada ao local da colisão dos serviços de emergência.
Segurança dos veículos	Percentagem de automóveis novos de passageiros com uma classificação de segurança EuroNCAP igual ou superior a um limiar predefinido.
Infraestrutura	Percentagem de distância percorrida em estradas com uma classificação de segurança acima de um limiar acordado.

De acordo com o *Road Safety Anual Report* do *International Transport Forum* (ITF) de 2019 [9], o número de mortes no trânsito em 2019, na maioria dos países, foi inferior à média de 2016- 18. Vinte e dois dos 30 países (73%) com dados provisórios tiveram uma redução, enquanto oito países (27%) registaram um aumento. Em média, o número de mortes no trânsito caiu quase 5% nos países cobertos pelo banco de dados do IRTAD, em relação à média de 2016-18. Em todos os 34 países, o número de mortes no trânsito diminuiu, em média, 2,5% em 2018, em comparação com a média dos três anos anteriores.

Nos últimos 25 anos, o custo económico e social da sinistralidade rodoviária com vítimas em Portugal, passou de um valor anual que chegou a ultrapassar 7% do PIB para um valor que estabilizou, nos últimos anos, em torno dos 2,5% do PIB a preços correntes de 2019. Isto resultou, essencialmente, da redução significativa do número de vítimas com maior grau de gravidade, assinalando-se em particular a redução de 70% no número de vítimas mortais neste período e a diminuição em 80,7% no número de feridos graves [2].

O risco de fatalidade, medido por mortes nas estradas por distância percorrida, diminuiu desde o ano 2000 em todos os 21 países que coletam regularmente dados sobre os quilómetros de veículos percorridos [9]. As mortes no trânsito na faixa etária de 15 a 17 anos caíram, em média, quase um terço (32%) entre 2010 e 2018. A redução para a faixa etária de 18 a 24 anos foi de 25% [9]. Em 2020, o número de mortes por acidentes rodoviários diminuiu significativamente, o que se deveu maioritariamente à pandemia da Covid-19 [9].

É notório que os esforços mais recentes feitos por parte das autoridades têm vindo a contribuir de forma positiva para a redução da sinistralidade rodoviária, no entanto, existe ainda um longo caminho a percorrer no sentido de atingir a Visão Zero. Este trabalho, em conjunto com a ANSR, pretende dar mais um contributo para que seja possível estarmos um passo mais perto de atingir o objetivo da Visão Zero.

1.3 Objetivos e Hipóteses de Investigação

Tendo em conta as mais recentes preocupações das autoridades responsáveis em evitar as mortes e feridos graves nos acidentes rodoviários, temos como objetivo principal determinar os fatores de risco que influenciam a proporção de acidentes graves.

Com base na revisão da literatura, percebeu-se que o aumento da velocidade aumenta a gravidade dos acidentes rodoviários. No entanto, os dados disponíveis sobre acidentes rodoviários em Portugal, não dispõem de informação sobre a velocidade de circulação dos veículos envolvidos no acidente, indicando apenas a velocidade média legal da estrada onde ocorreram os acidentes.

Neste sentido, importa perceber quais os fatores que poderão estar associados ao excesso de velocidade. Com o objetivo de dar resposta a esta questão colocou-se a seguinte hipótese:

> Hipótese um (H1): A redução da densidade de tráfego, aumentando a velocidade de circulação, aumenta o risco de acidentes graves;

Assumindo que a gravidade dos acidentes rodoviários não depende exclusivamente da velocidade, colocou-se outra hipótese:

> Hipótese dois (H2): A estrutura da rede rodoviária e das infraestruturas influencia a gravidade dos acidentes rodoviários.

A Autoridade Nacional de Segurança Rodoviária tem em desenvolvimento o projeto Sistema de Vigilância das Alterações da Sinistralidade – SiVig, que pretende potenciar as capacidades de análise de dados incluindo o recurso a modelos preditivos de ocorrência de acidentes e de gravidade das lesões resultantes. Com o objetivo de contribuir para o sucesso deste projeto e responder às necessidades e preocupações da ANSR, definiu-se como segundo objetivo desenvolver um modelo preditivo de ocorrência de acidentes graves.

1.4 Metodologia

A análise de dados é o processo de examinar grandes quantidades de dados em detalhes, a fim de extrair informações úteis que podem ser usadas para tomar decisões ou resolver problemas.

A metodologia CRISP-DM (*Cross Industry Standard Process for Data Mining*) aborda parte destes problemas, definindo um modelo que fornece padrões para a realização de projetos de mineração de dados. O modelo de processo CRISP-DM visa tornar os grandes projetos de mineração de dados menos dispendiosos, mais fiáveis, mais repetíveis e mais rápidos [10].

Tentou-se aderir às melhores práticas durante todo o processo de escrita dissertação, a fim de maximizar a probabilidade de sucesso. A metodologia considerada como um padrão para a realização de projetos de mineração de dados foi a abordagem CRISP-DM, ver Figura 1.

O ciclo de vida desta metodologia é composto por seis fases não sequenciais, contínuas e interativas, sendo elas: *Business/Problem Understanding, Data Understanding, Data Preparation, Modeling, Evaluation* e *Deployment*.

O modelo CRISP-DM é flexível e pode ser personalizado facilmente. Nesta dissertação, o foco será na exploração e visualização de dados para descobrir padrões nos dados.

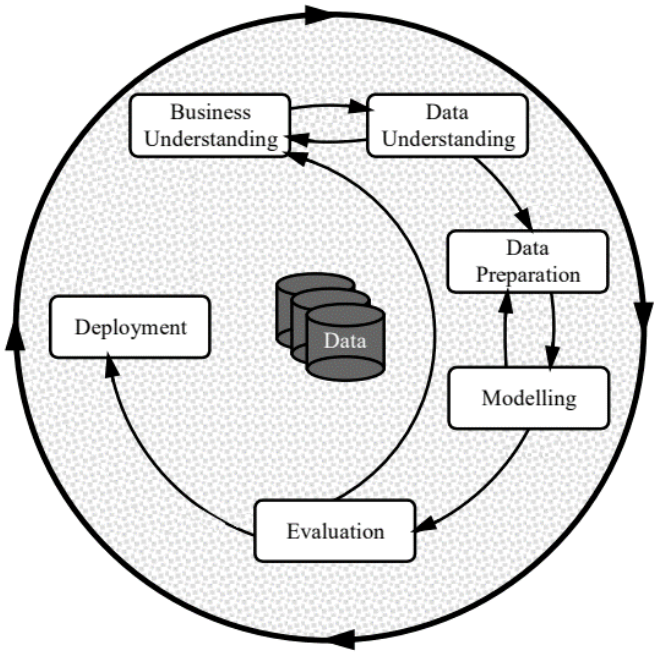


Figura 1 - CRISP-DM [10]

2 Revisão da Literatura

Este capítulo pretende analisar os estudos já publicados sobre o tema da segurança rodoviária, com o objetivo de construir um panorama sobre o que já se sabe sobre o tema e indicar eventuais orientações para a realização deste trabalho.

2.1 Metodologia

Para a realização da revisão de literatura foi utilizada a metodologia proposta por [11] que é realizada em três etapas: planeamento, pesquisa e reporte, ver Figura 2.

A maior parte dos trabalhos de pesquisa começam com uma revisão de literatura, no entanto, a menos que a revisão seja completa e justa, é de pouco valor científico.

Uma revisão sistemática da literatura (SLR) é um meio de avaliar e interpretar toda a investigação disponível relevante para uma determinada questão de investigação. O objetivo da SLR é identificar referências de boa qualidade de interesse real a um estudo [12].

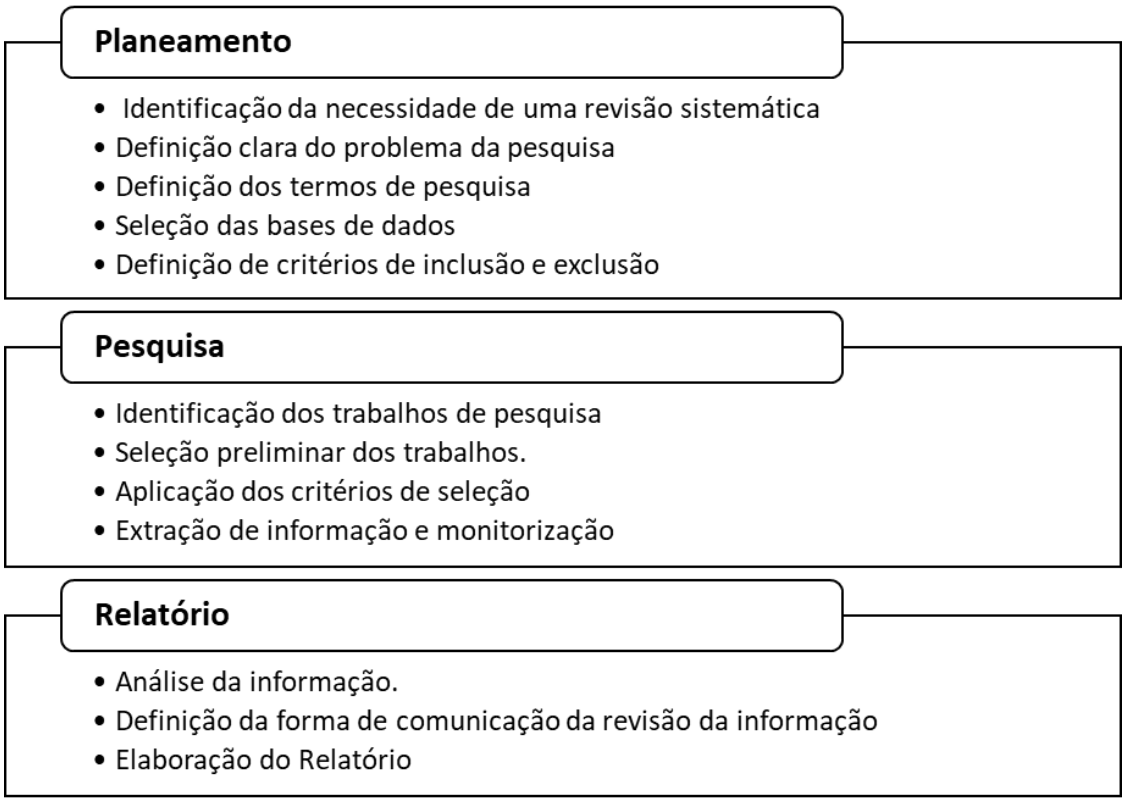


Figura 2 - Metodologia da Revisão da Literatura

As revisões sistemáticas visam apresentar uma avaliação justa de um tema de investigação utilizando uma metodologia fiável e rigorosa. É uma forma de estudo secundário que utiliza uma metodologia bem definida para identificar, analisar e interpretar todas as provas disponíveis relacionadas com uma questão de investigação específica de uma forma imparcial.

A necessidade de uma revisão sistemática decorre da exigência dos investigadores de resumirem todas as informações existentes sobre algum fenómeno de forma minuciosa e imparcial [11].

O problema da pesquisa corresponde ao tema definido para a dissertação: Identificação de fatores de risco na ocorrência de acidentes rodoviários e desenvolvimento de modelos para predição de acidentes rodoviários.

Definido o problema, foi necessário definir as palavras-chave da pesquisa. As questões de revisão impulsionam toda a metodologia de revisão sistemática. A pesquisa foi realizada em inglês por forma a alargar os resultados, sendo que as palavras-chave definidas foram: *Road Safety*, *Traffic Accidents*, *Risk Factors*, *Data Analysis*, *Crash Prediction Models*.

Para esta pesquisa foram utilizadas fontes diretas e indiretas. As fontes diretas resultam da pesquisa inicial e foram o *Science Direct, Scopus* e *Google Scholar*, bem como alguns livros e manuais relevantes para o tema e relatórios de instituições internacionais. As fontes indiretas resultaram das referências presentes nas fontes iniciais.

Os critérios de inclusão/exclusão surgiram em 3 níveis:

1. Critérios de elegibilidade: nesta fase, os resultados são aprovados ou rejeitados de acordo com os seguintes critérios:
    - Línguas- apenas artigos escritos em inglês ou português foram aceites;
    - Literatura científica exclusivamente a partir de artigos científicos ou revisões de literatura;
    - O tema é relevante para esta dissertação;
2. Leitura do resumo do artigo;
3. Leitura do artigo completo.

2.2 Fatores de Risco nos acidentes rodoviários

*"*Descrever os fatores que contribuem para sinistralidade rodoviária, listando os que foram estatisticamente associados à frequência de ocorrência de acidentes e à gravidade das lesões, representa apenas o rudimento de uma explicação científica*"* [13].

Determinar os fatores de risco que contribuem para os acidentes rodoviários é fundamental para permitir às autoridades competentes atuar de forma eficaz no sentido da sua prevenção.

O número de mortos ou feridos em acidentes rodoviários depende basicamente de três fatores: a exposição ao tráfego, a taxa de acidentes e a gravidade dos ferimentos [14], ver Figura 3.

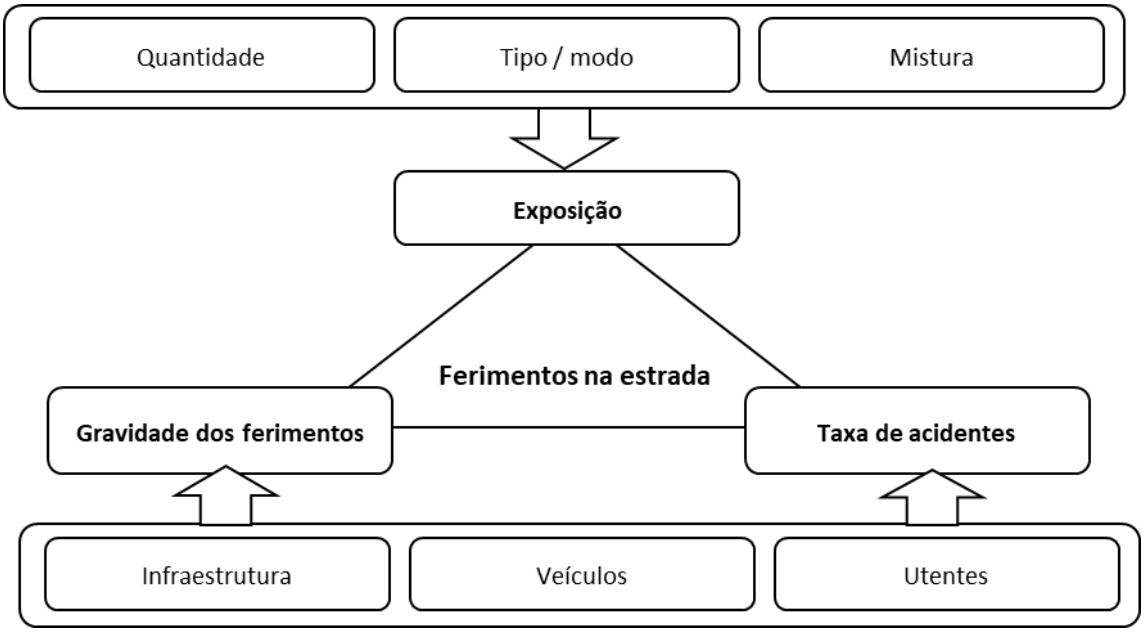


Figura 3 - Fatores que influenciam a segurança rodoviária [13]

Um fator de risco para sinistralidade é aquele que aumenta quer a probabilidade de ocorrência de acidentes (taxa de acidentes), quer a probabilidade de as lesões causadas pelo acidente serem mais graves (gravidade dos ferimentos).

Estes fatores de risco estão relacionados com os elementos do sistema de tráfego: infraestruturas, veículos e utentes da estrada, e dependem da sua interação.

Segundo [13], a taxa de acidentes depende, sobretudo, do tipo de estrada, do desenho da estrada, do meio ambiente, das características do condutor, do nível de álcool do condutor e da velocidade de marcha. Por sua vez, a gravidade dos ferimentos depende, essencialmente, do tipo de veículo, da

velocidade do impacto e da utilização de equipamento de proteção. Contudo, estas duas vertentes não podem ser olhadas de forma individual, estando intimamente relacionadas. Os fatores que influenciam a ocorrência do acidente também irão influenciar as características do mesmo e, evidentemente, as suas consequências.

A exposição ao tráfego está relacionada com o volume de tráfego e depende de:

1. Quantidade de tráfego – o volume de tráfego, expresso frequentemente pelo Tráfego Médio Anual (AADT), é um dos fatores mais determinantes na ocorrência de acidentes, sendo que quando o AADDT aumenta em 10%, o aumento estimado do número de acidentes é de 8,8% para um intervalo de confiança de 95%.
2. Tipo de tráfego – o tipo de tráfego influencia especialmente a gravidade dos acidentes, uma vez que diferentes tipos de veículos representam diferentes tipos de exposição ao perigo. Viajar de ciclomotor ou motociclo envolve um risco de lesão 10 vezes superior ao de viajar de automóvel, por exemplo.
3. Mistura – A proporção entre veículos e peões afeta a segurança rodoviária, sendo que, para uma determinada quantidade de tráfego pedonal, o risco de cada veículo a motor colidir com um peão diminui mais de 50% se o número de veículos a motor aumentar de 2.000 para 10.000. Este efeito ocorre uma vez que um tráfego mais denso, não só obriga o condutor a uma condução bem mais atenta no sentido de evitar acidentes, como obriga o condutor a praticar velocidades inferiores.

De acordo com o *Highway Safety Manual* [15], os fatores que contribuem para os acidentes rodoviários dividem-se em 3 categorias: Fatores Humanos, Veículos e Ambiente, ver Figura 4.

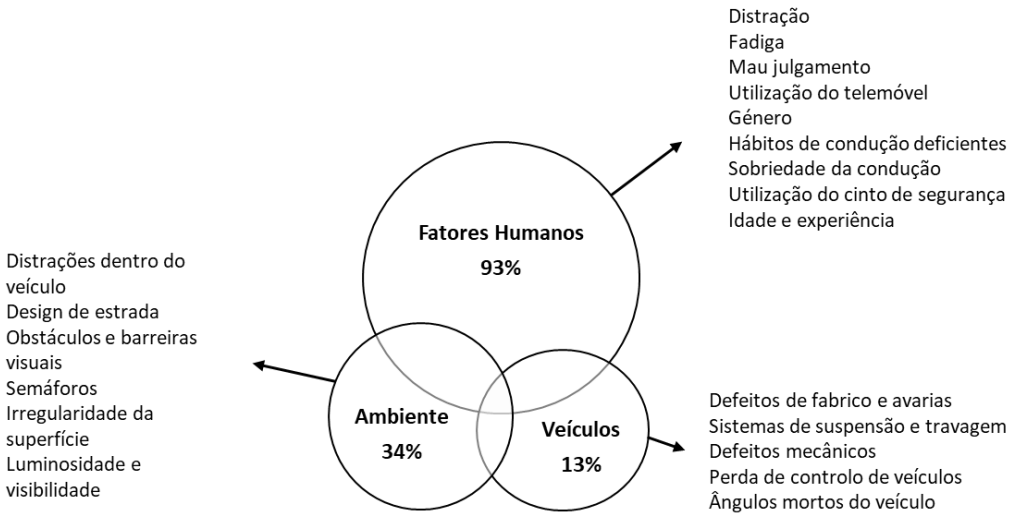


Figura 4 - Fatores que contribuem para os acidentes rodoviários

A gravidade dos ferimentos resultantes de um acidente rodoviário pode não resultar apenas das condições na qual o acidente sucedeu, mas também na rapidez com a qual os feridos foram socorridos. Segundo a OMS [16], os fatores de risco na gravidade dos ferimentos podem dividir-se em 2 grupos:

1. Fatores de risco que influenciam a gravidade do impacto, que incluem fatores de tolerância humana, prática de velocidades desadequadas, não utilização de equipamentos de segurança, condução sob o efeito de álcool, etc.
2. Fatores de risco que influenciam o resultado das lesões pós-acidente, como é o caso da demora na deteção do acidente e no transporte dos feridos para uma unidade de saúde, a falta de atendimento pré-hospitalar, a libertação de materiais perigosos, etc.

Alharbi et. al [17] elaboraram uma revisão de literatura que incluiu 80 estudos de investigação que examinam fatores de risco de mortalidade na sequência de um acidente de viação, com origem em 31 países de todo o mundo. A Tabela 3 apresenta o número de vezes que cada fator de risco foi identificado como estando associado à mortalidade na sequência de acidentes de viação.

Tabela 3 - Revisão da literatura nos fatores de risco dos acidentes rodoviários (apenas foram incluídos os fatores mais relevantes)

Fator de Risco	**Nº de artigos**
Idade mais avançada	29
Sexo masculino	28
Não utilização de cinto/capacete de segurança/ criança não contida ou inadequadamente contida	23
Álcool e/ou consumo de drogas	21
Excesso de Velocidade	16
Durante a noite/sem iluminação	16
Ferimentos na cabeça	13
Idade inferior a 55	11
Acidentes nas estradas rurais	11
Veículos pesados ou semipesados	11

Em suma, de acordo com o estado da arte, a ocorrência de acidentes rodoviários depende essencialmente de: fatores humanos, nomeadamente a idade dos condutores, o sexo dos mesmos, o consumo de álcool e/ou drogas, a prática de excesso de velocidade e a neglicência na utilização dos equipamentos de segurança; características dos veículos, nomeadamente do tipo de veículos; infraestruturas e ambiente, nomeadamente o tipo de estradas e a iluminação das mesmas. Estes fatores são, por sua vez, influenciados pela exposição de tráfego.

2.2.1 Metodologias utilizadas na identificação de fatores de risco

Estimar os riscos da ocorrência de acidentes rodoviários nos diversos grupos de utentes da estrada tem sido de grande interesse prático e académico entre investigadores e profissionais de segurança rodoviária há muitos anos.

Haque et.al [18] identificam 4 metodologias utilizadas pelos investigadores neste sentido, ver Tabela 4.

Tabela 4 - Metodologias de identificação de fatores de risco

	Comparações descritivas	**Modelos de regressão**	**Técnicas de caso-controlo**	**Metodologia controlada pela exposição**
Metodologia	Dados comparados em várias circunstâncias através da contagem de frequências e da sua análise categórica.	Correlação entre a frequência de acidentes e as variáveis explicativas através de modelos de regressão.	Dados agrupados em categorias de caso e controlo, para que possam ser feitas comparações.	O risco é obtido dividindo a frequência de acidente pelo VMT (milhas viajadas pelo veículo).
Problemas	Falta de dados de exposição ou casos de controlo adequados.	Requerem uma recolha de dados em todo o sistema para a variável de exposição.	Falta de dados de exposição ou casos de controlo adequados.	Recolher dados de exposição exógenos coloca um desafio significativo.
Vantagens	Relativamente simples de aplicar e são úteis para obter uma compreensão rápida dos fatores de risco.	Calculam os riscos de acidente sob diferentes fatores influentes após a contabilização da exposição de acidentes.	É possível aplicar quando existe falta de dados exógenos.	Um dos métodos mais apelativos de avaliação de risco - fiável e de alta qualidade.

2.3 Medidas de Segurança

Depois de encontrados os fatores de risco, ou as causas dos acidentes, é possível aplicar medidas de segurança que diminuam os impactos destes mesmos fatores. O estudo destas medidas e do efeito de cada uma delas na prevenção da sinistralidade é, também, um tema que tem sido bastante estudado.

A velocidade de circulação, sendo um dos fatores de risco mais identificados em toda a literatura, é também um dos maiores alvos das medidas de segurança adotadas internacionalmente. São exemplos disso a instalação de lombas de velocidade, a adoção de zonas de 30km/h, a fiscalização policial do excesso de velocidade, o controlo médio das velocidades por secção, etc. Tratar os locais de alto risco na ocorrência de acidentes, através da identificação, análise e melhoria das estradas em locais com uma elevada concentração de acidentes, melhorando a conceção das estradas ou a regulamentação do tráfego nesses locais permite a redução de acidentes em 28%, segundo [19].

A avaliação dos efeitos das medidas implementadas pretende estimar, de forma quantitativa, o efeito de uma intervenção sobre o número de acidentes ou de vítimas. É através da avaliação e quantificação dos sucessos e insucessos resultantes da aplicação das medidas corretivas que se torna possível identificar as tipologias de medidas mais adequadas ao sistema de tráfego nacional, estimar a eficácia associada a cada tipo de intervenção e generalizar a sua aplicação ao país e uniformizar os seus parâmetros relevantes.

Os indicadores mais diretos e importantes de avaliação das medidas são o número de acidentes ou de vítimas, ou seja, a comparação desse indicador depois de implementada a medida com a situação expectável caso nada tivesse sido feito. Existem ainda outros indicadores que podem ser monitorizados, como por exemplo: a velocidade dos veículos; os tempos de viagens – aumentos significativos dos tempos de viagem podem levar a alterações da trajetória; o movimento dos peões, o comportamento dos utilizadores (condutores, peões, residentes); o meio envolvente (poluição, ruído, etc.); os fluxos de tráfego – verificar eventuais alterações nas trajetórias usuais.

2.4 Dados dos Acidentes Rodoviários

A caracterização e estudo dos acidentes rodoviários depende da informação recolhida no local e de informação existente sobre as características da via, dos veículos, do tráfego, entre outras [20].

Os dados de caracterização de acidentes são a base para se poder estudar padrões de sinistralidade e as suas causas, facilitando desta forma a tomada de decisão informada para ações que modifiquem os fatores ou influenciem comportamentos, com o objetivo final de reduzir a sinistralidade rodoviária [20].

Tradicionalmente, os dados utilizados na análise de acidentes rodoviários provêm de entidades governamentais. Apesar dessa fonte de informação ainda permanecer sólida, atualmente já se consideram outras fontes de informações que ajudam a completar a caracterização dos acidentes.

Recolher informação de diferentes fontes de informação não só torna a mesma mais completa como torna possível comparar/ hierarquizar as variáveis mencionadas.

As variáveis recolhidas dividem-se em 5 categorias, em função das suas características: Acidente, Ambiente e Estrada, Pessoas, Veículos e Tráfego [20].

Diferentes tipos de equipamentos eletrónicos são utilizados para recolher dados de acidentes da estrada, que incluem tecnologias que fazem parte da infraestrutura rodoviária, tais como câmaras de

deteção de velocidade, radares de veículos e outros vídeos de videovigilância foram utilizados em muitos estudos, tendo em conta a sua disponibilidade e acessibilidade [21].

Os equipamentos de bordo (OBE) instalados em veículos também podem ser utilizados para recolher os dados relacionados com as condições do veículo e as distrações do condutor. Estas OBEs incluem câmaras criadas para registar as condições rodoviárias e dos veículos, sistemas de posicionamento global, câmaras de vigilância para identificação de sonolência do condutor ou distrações [22].

As redes sociais surgiram como a mais recente fonte na recolha de dados de acidentes de viação e pesquisas relacionadas, e uma grande quantidade de dados é utilizada para o mesmo fim a partir do *Waze, Twitter, Google Maps, Bing*, e *Inrix* [1].

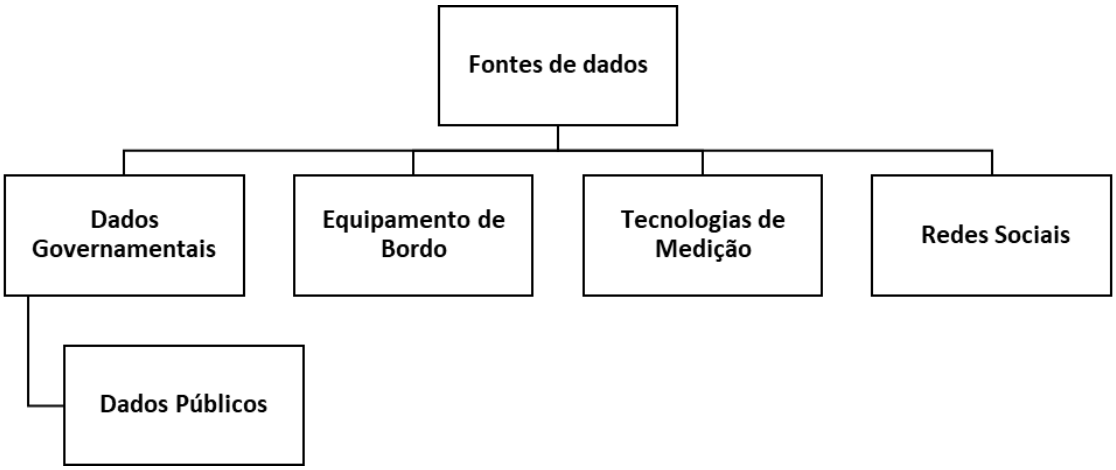


Figura 5 - Fontes de dados dos acidentes rodoviários

É importante ligar os dados hospitalares aos relatórios policiais para completar a informação com, por exemplo, período de internamento e reabilitação. Segundo [23], é também importante integrar informação das bases de dados de:

- Veículos (características da frota de veículos, de veículos recém-registados, aprovação / resultados da falha dos testes periódicos das inspeções periódicas de veículos)
- Estradas (características de projeto, pavimentação, sinalização e marcação, dispositivos de segurança, iluminação, estado de manutenção)
- Condutores (licenciamento de condutores, infrações de tráfego)
- Tráfego (fluxos de tráfego, mistura de tráfego)
- Acidentes de trabalho, de forma a identificar também se os acidentes ocorrem no caminho de ou para o trabalho.

Identificadas as variáveis independentes necessárias à boa caracterização do acidente, foi feita uma análise a que organismos possuem essa informação em Portugal, ver Figura 6.

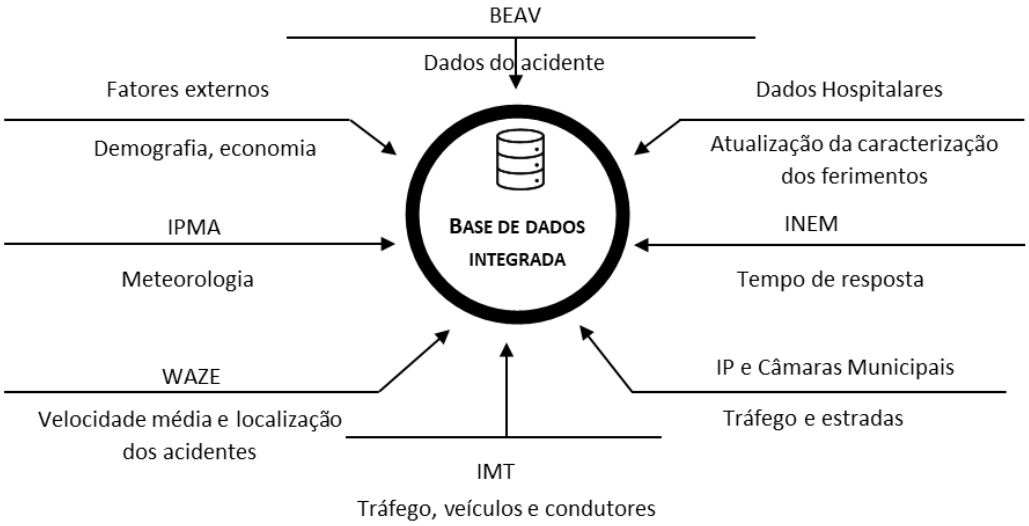


Figura 6 – Fontes de dados alternativas

2.4.1 Qualidade dos Dados

O tema da qualidade dos dados é essencial na medida em que cada vez mais as decisões são apoiadas em informação que depende dos *inputs* de grande quantidade de dados. É essencial que os dados sejam analisados e haja processos que garantam que a informação é relevante e adequada [24].

Dados com erros podem levar a uma má identificação de "zonas de acumulação de acidentes" e segmentos de estrada perigosos, a projeções de falsas estimativas no cálculo de taxas de acidentes e fatalidades e a falhas na deteção de parâmetros errados sobre as responsabilidades na ocorrência de acidentes [24].

A eficácia dos métodos de prevenção de acidentes depende principalmente da genuinidade dos dados recolhidos e estimados e da adequação dos métodos de análise [25].

Os erros presentes nos dados de acidentes rodoviários podem resultar quer de erros no reporte do acidente, quer de erros no registo do acidente. O primeiro resulta do facto de que, muitas vezes, os acidentes não são reportados por relutância em chamar a polícia, por fuga ao acidente, pela definição de acordos entre as partes ou simplesmente porque não houve feridos, o que leva a que sejam calculadas taxas de acidentes erradas. O segundo resulta, essencialmente, da falta de conhecimento e formação sobre como preencher o formulário de reporte e como recolher a informação, da falta de tempo e recursos policiais para efetuar o registo, da fraca troca de informação

sobre o estado das vítimas após a hospitalização, da identificação errada do local do acidente, etc. Este erro, por sua vez, tem como consequência uma ambiguidade na identificação de fatores responsáveis pelo acidente [24].

Embora estes erros sejam comuns e difíceis de eliminar, existem estratégias a adotar para tentar minimizá-los, que passam pelo reforço da necessidade do reporte de acidentes, tornando-o um requisito legal, pela melhoria da qualidade das ferramentas disponíveis para a recolha de dados, pela melhoria da formação das forças policiais ou pela auditoria e garantia da qualidade dos dados [6].

Uma vez que nem sempre é possível evitar estes erros, outra solução para melhorar a qualidade dos dados passa por recorrer a métodos para retificar/estimar os erros [24]:

1. Captura – Recaptura.

   Este método é utilizado na identificação de registos de acidente em falta em diversos registos. Pode ser feito cruzando os dados de diversas fontes como os da polícia com os dos hospitais ou estimando o número provável de acidentes numa determinada população. É Ferramenta muito utilizada para estimar o nível de erro devido a acidentes não reportados, nomeadamente para o número de mortos. É utilizado em diversos países, como Paquistão, Nicarágua, Canadá, Etiópia, Malawi, África do Sul, Austrália, França, Dinamarca e Nova Zelândia.
2. Comparação com os dados do setor da saúde.

   É mais fácil de utilizar, pelo que normalmente é usado mais prontamente para a análise de lesões graves e ligeiras assim como de mortes. Resulta da comparação entre os registos de entrada nos hospitais e da morgue com os registos de acidentes da polícia. É utilizado pelo Reino Unido, Austrália, Holanda, China, Dinamarca.
3. Ligação Probabilística.

   Consiste na utilização de mecanismos automáticos ou semiautomáticos para relacionar diferentes bases de dados. São utlizadas variáveis de ligação como nome, localização geográfica, data e hora do acidente, género e idade das vítimas, por exemplo. É utilizado na França, Nova Zelândia, EUA, Austrália.

2.5 Modelos Preditivos

Tradicionalmente, para a previsão do número de acidentes, ou para a classificação da severidade dos mesmos, são utilizadas técnicas de modelos estatísticos. No entanto, novas abordagens recorrendo a

Machine Learning (ML) estão a ser investigadas, uma vez que permitem não só a automatização das análises, o que possibilita a diminuição da intervenção humana, mas também permite aos sistemas aprenderem com os dados [12].

Existem diversos tipos de modelos preditivos, que têm vindo a ser investigados na ótica da predição de acidentes rodoviários.

Os Algoritmos Genéticos (GA) são heurísticos de pesquisa, usados para resolver problemas de otimização relacionados com o tráfego rodoviário na área da predição de acidentes; limitados e sem restrições, e baseiam-se na evolução natural [26].

As Redes Bayesianas (BN) são representadas através do Gráfico Acíclico Direcionado no qual cada nó se relaciona com uma variável aleatória exclusiva e cada borda se relaciona com uma dependência condicional. As relações que existem entre as variáveis podem ser representadas no gráfico através da utilização de fatores [27].

As *Suport Vector Machines* (SVM) podem lidar com questões que envolvem tarefas de classificação lineares e não lineares, no que diz respeito a problemas separáveis ou não separáveis. O modelo SVM será treinado, testado e comparado com dados recolhidos. Um modelo SVM mapeia os pontos de dados dos dados recolhidos para o espaço de características de alta dimensão, utilizando a função de núcleo. Os modelos SVM têm uma precisão de previsão consideravelmente maior em comparação com a abordagem tradicional e o melhor para a previsão de acidentes de viação em tempo real [28].

Redes Neuronais Convolucionais (CNN) e Redes Neuronais Recorrentes (RNN) são os dois tipos de quadros de *Deep Learning* e uma combinação dos dois são utilizados para descobrir relações e configurações desconhecidas em dados de acidentes rodoviários de alta dimensão, agrupamento de texto, reconhecimento de voz, processamento de imagem e visão computacional [29].

Uma rede neural artificial (ANN) consiste numa rede neural de alimentação multicamada, com camadas adicionais entre as camadas de entrada e saída. Através de linhas de coeficientes de peso, cada neurónio em cada camada está ligado a cada neurónio nas camadas adjacentes. Como resultado, qualquer alteração nos coeficientes de peso mudará o resultado da rede. Isto indicará os valores propostos de coeficientes ponderados para obter os resultados desejados. Assim, a rede neural pode ser treinada para desempenhar uma determinada função, regulando os valores dos pesos entre elementos [30].

As árvores de decisão (DT) são compostas por nós, onde o primeiro nó inicia a árvore e os seguintes representam o teste de um atributo, cujo critério para a ramificação é a utilidade do atributo para a classificação pretendida. O caminho da árvore corresponde a uma regra de associação. São muito úteis para problemas de classificação [12].

O algoritmo *K-Nearest Neighbor* (KNN) mede a distância entre os registos, identificando os vizinhos mais próximos e baseando a sua classificação nessa proximidade. É uma técnica simples e pioneira em *Machine Learning* e é usada maioritariamente em problemas de classificação [12].

Silva et al.[12] elaboraram um balanço sobre a literatura identificada no tema da utilização de técnicas de *Machine Learning* aplicadas à segurança rodoviária, onde são descritos os artigos em questão, o tipo de análise realizada (frequência/classificação), as variáveis dependentes calculadas e a técnica/modelo utilizado. Com base neste balanço, foi feito um resumo sobre as variáveis independentes mais utilizadas em cada tipo de análise, ver Tabela 5.

Tabela 5 - Variáveis independentes mais utilizadas em cada tipo de modelo

MODELO	**VARIÁVEIS**			
	AMBIENTE E ESTRADA	**PESSOAS**	**VEÍCULOS**	**ACIDENTE**
SEVERIDADE	Condições do tempo, luminosidade, segmento rodoviário, tipo de berma	Género, idade, uso de cinto de segurança	Género, idade, uso de cinto de segurança	Tipo de acidente, horas, dia da semana
FREQUÊNCIA	Volume de trafego, tamanho do segmento, largura da berma	Género		Ano, Estação do ano, veículos envolvidos no acidente
SEVERIDADE E FREQUÊNCIA	Volume de trafego, largura da berma, limite de velocidade, largura da via, condições do pavimento, tamanho do segmento			Horas, dia da semana

3 Análise e Visualização dos Dados

A fase de compreensão de dados da metodologia CRISP-DM é fundamental para alcançar o objetivo e antecipar o surgimento de desafios imprevistos nos passos seguintes.

Este processo implica a recolha de dados, visualização e análise dos mesmos, determinação da sua qualidade e identificação de funcionalidades relevantes para abordar as questões de investigação. O objetivo final da análise de acidentes rodoviários é reduzir o número e a gravidade das colisões e das vítimas mortais. Para atingir este objetivo é necessário dispor de dados de alta qualidade, fiáveis e precisos sobre os fatores que influenciam diretamente os acidentes rodoviários e as lesões que deles resultam.

Para esta dissertação utilizaram-se dados sobre acidentes rodoviários reportados à Autoridade Nacional de Segurança Rodoviária (ANSR) pelas forças de segurança (GNR e PSP), recolhidos através do Boletim Estatístico de Acidentes de Viação (BEAV), entre os anos de 2010 a 2019.

Estes dados apresentam-se no formato *xlsx* e estão organizados em ficheiros separados por ano e por secção, sendo elas: Acidentes, Veículos, Condutores, Passageiros e Peões.

Tabela 6 - Organização dos dados

	Descrição	Número de Observações	Número de variáveis
Acidentes	Características do acidente	327 384	43
Veículos	Características e consequências dos veículos e nos condutores envolvidos no acidente.	495 982	40
Passageiros	Características e consequências dos passageiros	105 683	28
Peões	Características e consequências dos peões.	55 928	30

3.1 Qualidade dos dados

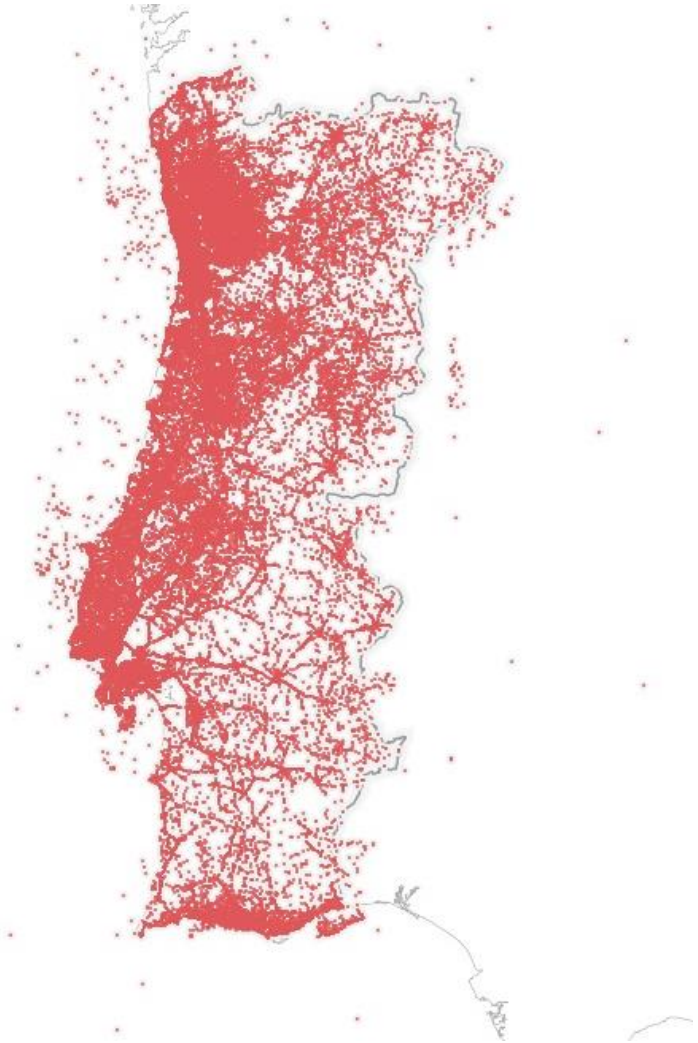


Figura 7 - Posicionamento dos dados dos acidentes com coordenadas entre 2010 e 2019

Com recurso à ferramenta *Tableau*, posicionaram-se, através das coordenadas, os acidentes ocorridos entre 2010 e 2019.

Como é possível observar na Figura 7, muitos dos acidentes registados encontram-se fora dos limites terrestres de Portugal. Esta visualização deixa patente os constrangimentos que as coordenadas impõem quando se pretende analisar a localização do acidente.

3.2 Preparação dos Dados

Antes de iniciar a análise foi necessário preparar os dados, limpando-os e organizando-os.

O primeiro passo foi juntar os *datasets* dos vários anos (2010 a 2019) para cada uma das folhas, de onde resultaram 4 *datasets,* acidentes, veículos, peões e passageiros.

Descartaram-se, à partida, os *datasets* dos peões e dos passageiros, por não serem considerados relevantes na análise procurada.

3.2.1 Dados dos acidentes

O *dataset* dos acidentes foi trabalhado com recurso à ferramenta *PowerBI*, onde se começou por eliminar as variáveis "Km" e "Nome Arruamento" por apresentarem muitos valores nulos e as variáveis "Latitude GPS" e "Longitude GPS" por apresentares muitos erros, como referido anteriormente. Foram também eliminadas as linhas com valores nulos nas restantes variáveis.

De seguida foi necessário alterar o formato da variável "datahora".

O objetivo do *dataset* dos acidentes seria ajudar a compreender o impacto das características dos acidentes e da sua envolvente na gravidade dos mesmos, através de ferramentas de visualização.

Para além dos dados do BEAV, fornecidos ao ISCTE pela ANSR, sentiu-se a necessidade, para responder às hipóteses colocadas*,* de procurar dados externos que complementassem a análise. Para isso, foram utilizados dados do INE, relativos à Extensão da rede nacional rodoviária (km) por Localização geográfica (Distrito) e Tipo de rede rodoviária, relativos ao ano de 2020, e do IMT, presentes no Relatório de Monitorização da Rede Rodoviária Nacional de 2012-2013.

3.2.2 Dados dos veículos

O *dataset* dos veículos foi trabalhado com recurso à linguagem *Python* com o objetivo de serem desenvolvidos modelos de *Machine Learning* para predição da gravidade dos acidentes.

O primeiro passo na preparação do *dataset* dos veículos foi eliminar as linhas que apresentavam variáveis com valores nulos.

De seguida, foi definida a variável “Gravidade”, que seria a variável dependente e resultaria da classificação da variável “Lesões a 30 dias”. Quando as “Lesões a 30 dias” equivalessem a 1 morto ou 1 ferido grave, a “Gravidade” seria igual a 1, caso contrário, seria igual a 0. Considerou-se que um acidente grave era aquele que envolvesse, pelo menos, 1 morto e/ou 1 ferido grave. O objetivo desta variável seria prever a probabilidade de um acidente ser grave, em função das características a priori dos veículos e dos condutores.

O passo seguinte foi criar a variável Grupo Etário, que assume valores do 1 ao 16, em função dos grupos etários definidos.

A variável Ano Matrícula foi discretizada em 4 classes:

1. <2005
2. 2005-2010
3. 2010-2015
4. 2015-2020

De seguida, eliminaram-se as variáveis que não interessavam incluir no modelo, ficando apenas com “Sexo”, “Categoria Veículos”, “Tipo Veiculo”, “Ano Matrícula”, “Inspeção Periódica”, “Seguros” e “Grupo Etário”.

Às variáveis categóricas, “Seguros”, “Inspeção Periódica”, “Sexo”, “Categoria Veículos”, “Tipo Veículo” foi aplicada a técnica *One-Hot-Encoding*, que permite converter variáveis de dados categóricos para que possam ser fornecidas a algoritmos de aprendizagem automática para melhorar as previsões.

Como já referido, estes dados serão trabalhados apenas no capítulo 4, referente à modelação.

3.3 Visualização dos Dados

Nesta fase procedeu-se a uma análise dos dados disponibilizados pela ANSR, recorrendo a ferramentas de visualização, nomeadamente o *PowerBI.* Foram também utilizados alguns dados de fontes externas, trabalhados no Excel.

Começou-se por caracterizar de forma geral os acidentes ao longo do território nacional continental, passando depois a uma análise mais específica a nível espacial e a nível temporal.

3.3.1 Caracterização Geral

Uma vez que o principal objetivo da Visão Zero é evitar os acidentes com feridos graves e mortos, dividiram-se os acidentes em graves e leves, sendo considerado grave qualquer acidente com pelo menos 1 ferido grave ou 1 morto.

O principal foco desta análise será caracterizar os acidentes graves e perceber as suas causas, de modo a facilitar a tomada de decisões no sentido de evitar a ocorrência dos mesmos.

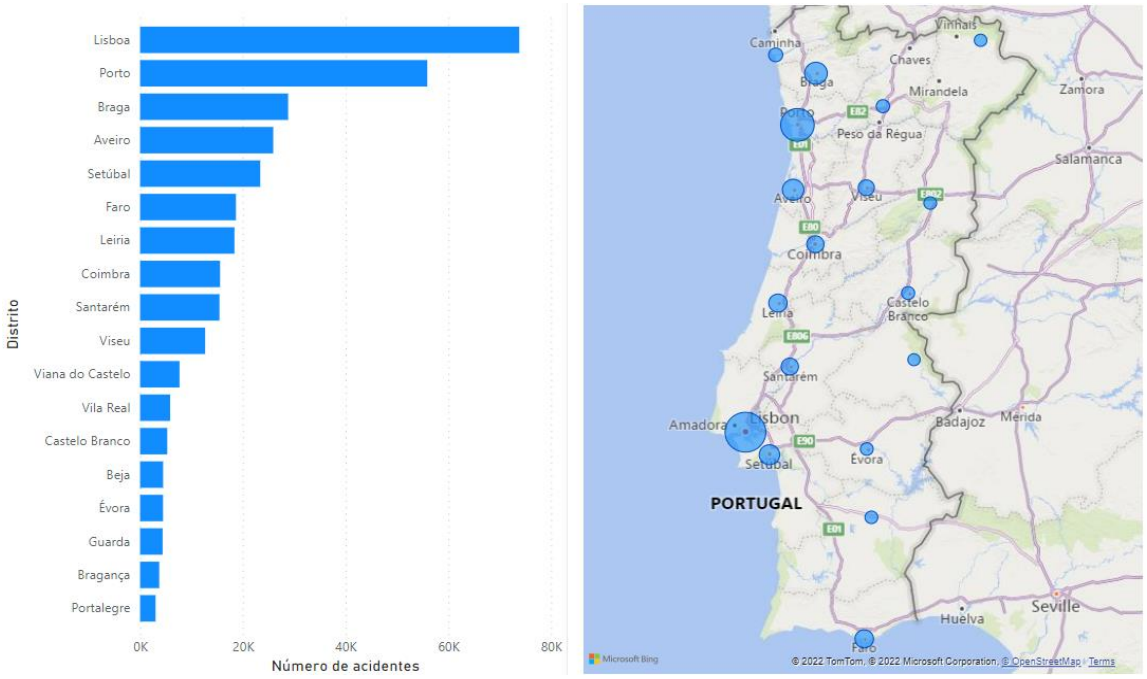


Figura 8 - Número total de acidentes por distrito

Tendo em conta estas restrições das coordenadas, optou-se por tentar compreender a dispersão geográfica dos acidentes em Portugal, utilizando a variável distrito para realizar esta análise.

Numa primeira análise, podemos verificar que o número de acidentes (Figura 8) é, como seria de esperar, consideravelmente superior em Lisboa, seguido do Porto. Este número seria fácil de prever uma vez que estes são os distritos com mais população e, consequentemente, mais tráfego.



Figura 9 - Número de acidentes graves por distrito

Também o número de acidentes graves (Figura 9) tem o mesmo comportamento, sendo superior nos distritos com maior população. Esta análise não é, no entanto, representativa do risco de acidentes graves no total de acidentes. É importante conhecer a proporção de acidentes graves no total de acidentes, por forma a perceber quais os distritos que apresentam maiores fatores causadores de acidentes graves.

Nesse sentido, fomos calcular esta proporção, dividindo o número de acidentes graves por distrito pelo número de acidentes total por distrito. Obtivemos a seguinte distribuição (Figura 10).

Podemos verificar que a proporção de acidentes graves em função do número de acidentes (Figura 10) foi maior no distrito de Portalegre: embora tenham ocorrido menos acidentes, aqueles que ocorreram foram, de um modo geral, mais graves. Por outro lado, o Porto foi o distrito que apresentou menor proporção de acidentes graves. Isto pode ajudar a corroborar a H1 de que um maior volume de tráfego (fruto de maior população), leva a menores níveis de velocidade e, consequentemente, menores níveis de gravidade dos acidentes.

Numa primeira abordagem, optou-se por comparar os distritos do Porto e o distrito de Portalegre, na tentativa de perceber quais os fatores que poderiam estar associados a estas diferenças e se poderiam estar de acordo com as hipóteses colocadas nesta dissertação, para numa fase posterior verificar a aderência destes fatores ao resto do território continental.

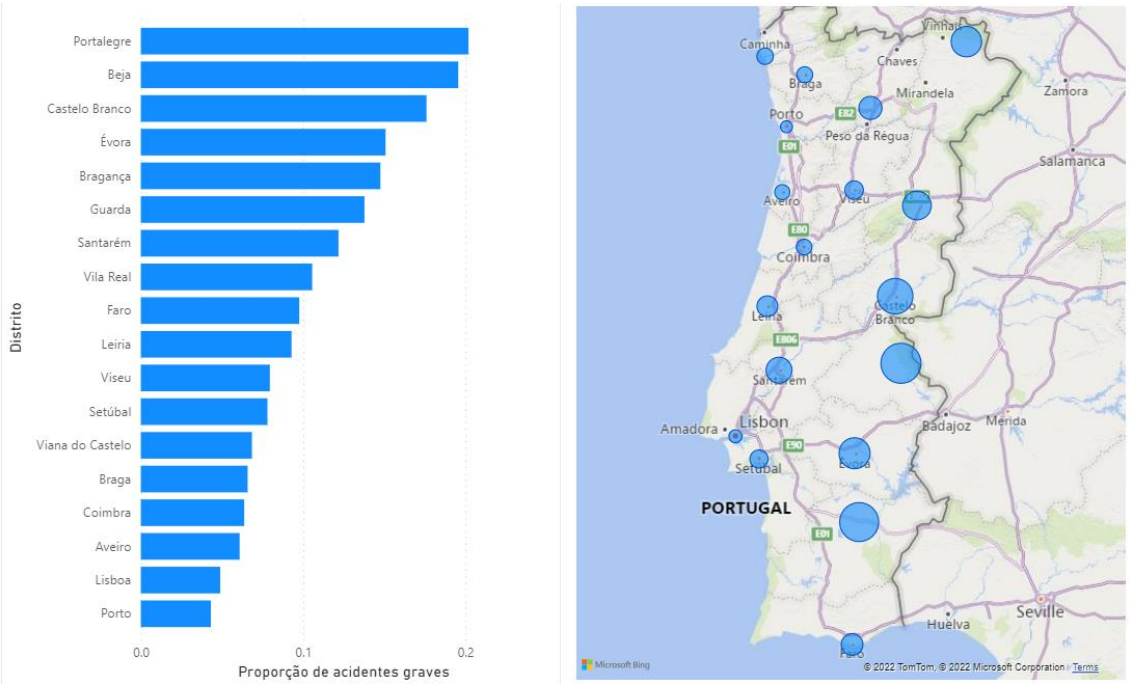


Figura 10 - Proporção de acidentes graves no total de acidentes por distrito

3.3.2 Análise Espacial e das Infraestruturas

Comparação Porto/Portalegre

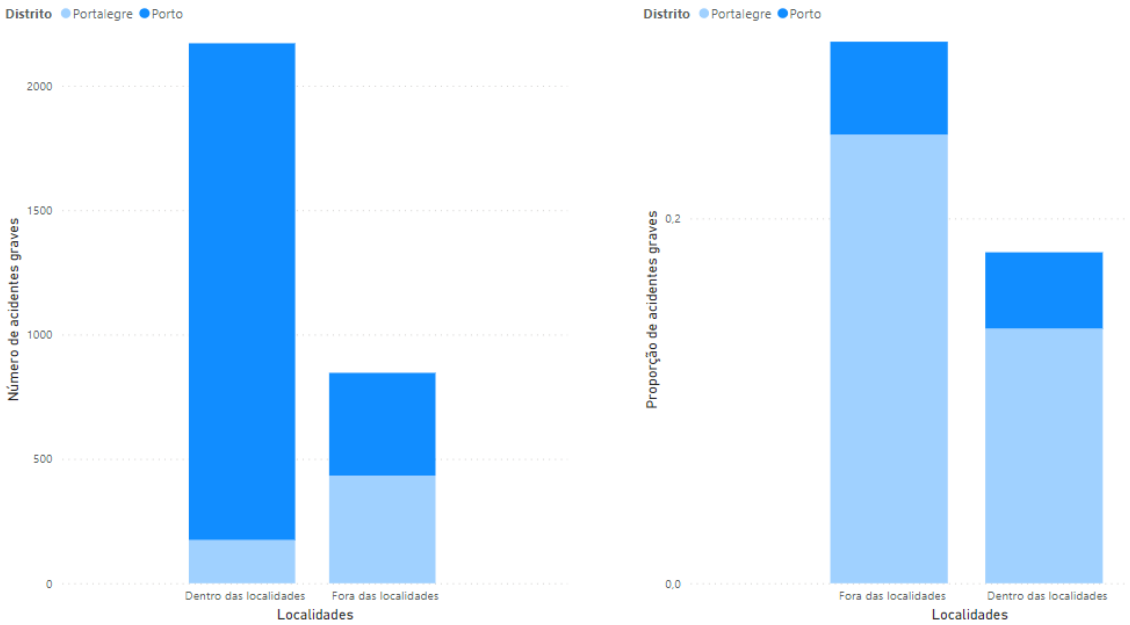


Figura 11 - Número e proporção de acidentes graves dentro e fora das localidades nos distritos do Porto e Portalegre

Calculando o número de acidentes graves e a sua proporção dentro e fora das localidades, nos distritos do Porto e Portalegre (Figura 11), verifica-se que, em Portalegre, o maior número de acidentes

graves ocorreu fora das localidades, enquanto no Porto aconteceram dentro das localidades. Provavelmente, isto deve-se a uma maior proporção de tráfego dentro de contexto urbano no Porto, por ser um distrito mais urbanizado. Segundo os dados do PORDATA, o município do Porto tinha, em 2021, uma densidade populacional de 5.597 habitantes por Km2, enquanto Portalegre apresentava um valor de 50 [31].

No entanto, a maior ruralidade do distrito de Portalegre, parece não ser a única justificação para uma maior incidência de acidentes graves fora das localidades. Neste distrito, a proporção de acidentes graves fora das localidades é muito maior do que no Porto. De facto, o mesmo se verifica dentro das localidades, onde a proporção de acidentes graves também é superior em Portalegre. No entanto, a desproporção entre acidentes dentro e fora das localidades é superior em Portalegre, sendo os acidentes fora das localidades acentuadamente mais graves em Portalegre.

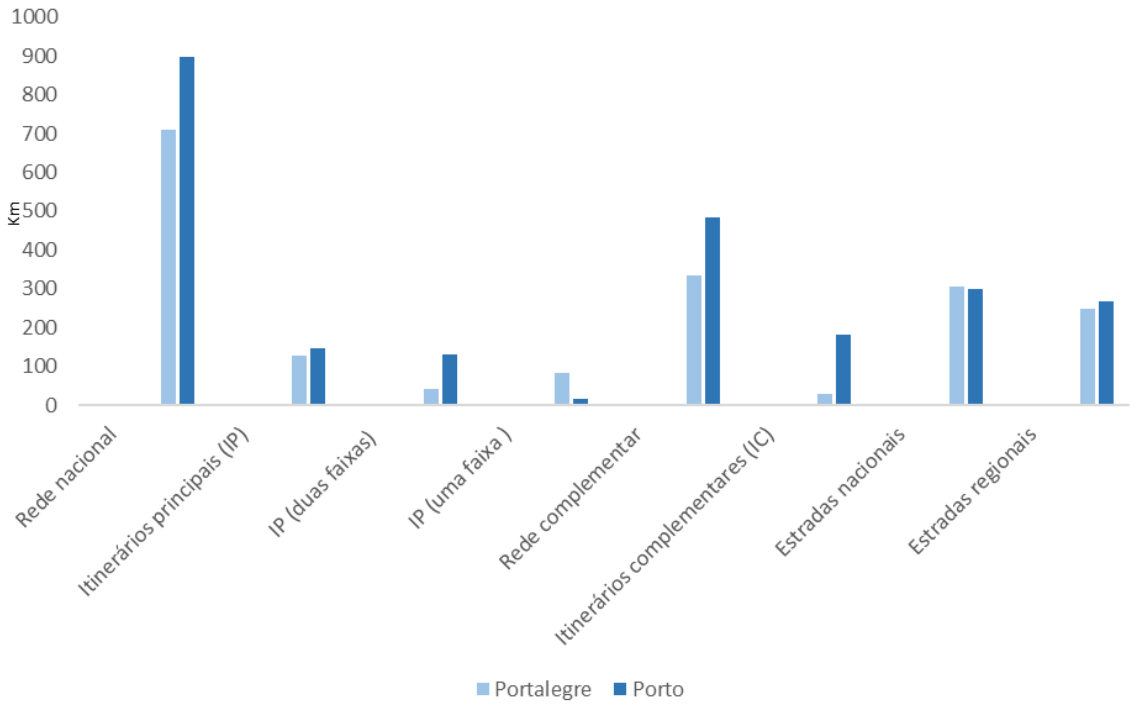


Figura 12 - Características da rodoviária no Porto e em Portalegre - INE

A maior gravidade dos acidentes fora das localidades no distrito de Portalegre, comparativamente com o Porto, parece poder justificar-se, pelo menos em parte, pelas diferenças nas infraestruturas rodoviárias entre os dois distritos (Figura 12). Nos dois distritos, a rede de itinerários principais (IP) é idêntica, mas, no distrito do Porto, uma elevada percentagem destas vias tem duas faixas de rodagem em cada sentido, enquanto em Portalegre a maior parte da rede de IP tem apenas uma faixa de rodagem. No distrito do Porto, os Itinerários complementares (IC) têm uma muito maior

representatividade. Já nas estradas nacionais (EN) e regionais (ER) existe uma situação semelhante nos dois distritos.

Na tentativa de melhor perceber a importância da estrutura das infraestruturas rodoviárias no risco de acidente grave, analisaram-se os tipos de manobras envolvidas em acidentes graves (Figura 13).

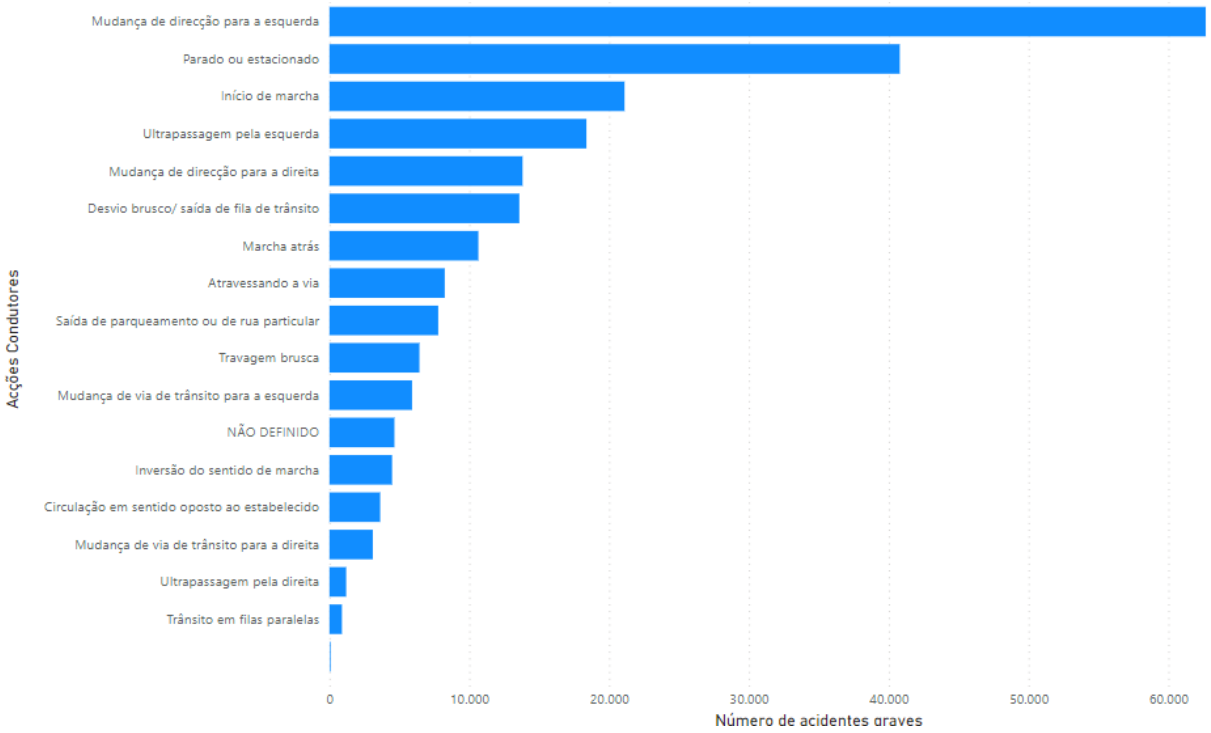


Figura 13 - Acidentes graves por ações dos condutores

Se considerarmos as manobras/ações dos condutores envolvidos em acidentes graves (Figura 13), percebemos que o maior número de acidentes graves ocorre em situações de desvio brusco/saída de fila de trânsito, mudança de direção para a esquerda e ultrapassagem pela esquerda. Estas manobras são mais prováveis de ocorrer em IPs com 1 faixa, o que ajuda a complementar a conclusão anterior.

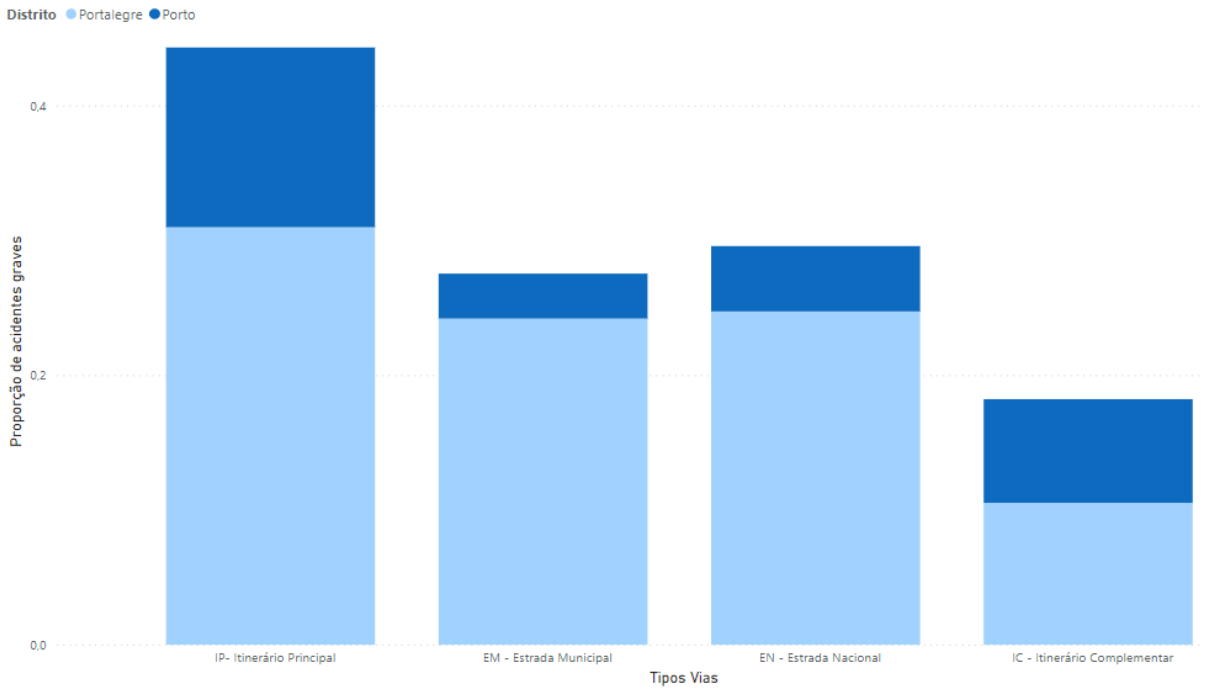


Figura 14 - Proporção de acidentes por tipologia de estrada no Porto e em Portalegre

Quando se compara a proporção de acidentes por tipologia de estrada nos dois distritos, fora das localidades (Figura 14), verifica-se que em Portalegre há uma maior proporção de acidentes graves nos IP, indicando novamente que a existência de IP's com 1 faixa de rodagem aumenta o risco de acidente grave.

Nas EM e EN, continua a verificar-se um maior risco de acidente grave em Portalegre. A diferença entre esta proporção nas estradas nacionais, poderá estar relacionada com o facto de que o distrito de Portalegre apresenta um maior grau de sinuosidade comparativamente com as estradas no distrito do Porto.

O Índice de Sinuosidade (IS) consiste num indicador "geométrico", que identifica a acessibilidade geográfica entre dois pontos i e j. O Indicador de Sinuosidade traduz-se pelo quociente entre a distância real de deslocação e a distância em linha reta entre dois pontos, de acordo com a seguinte expressão:

$$IS_{i,j} = \frac{S_{i,j}}{L_{i,j}} \geq 1$$

O IS em Portalegre (1,39) é superior ao IS no Porto (1,28), o que também poderá justificar a maior proporção de acidentes graves no distrito de Portalegre (Figura 15).

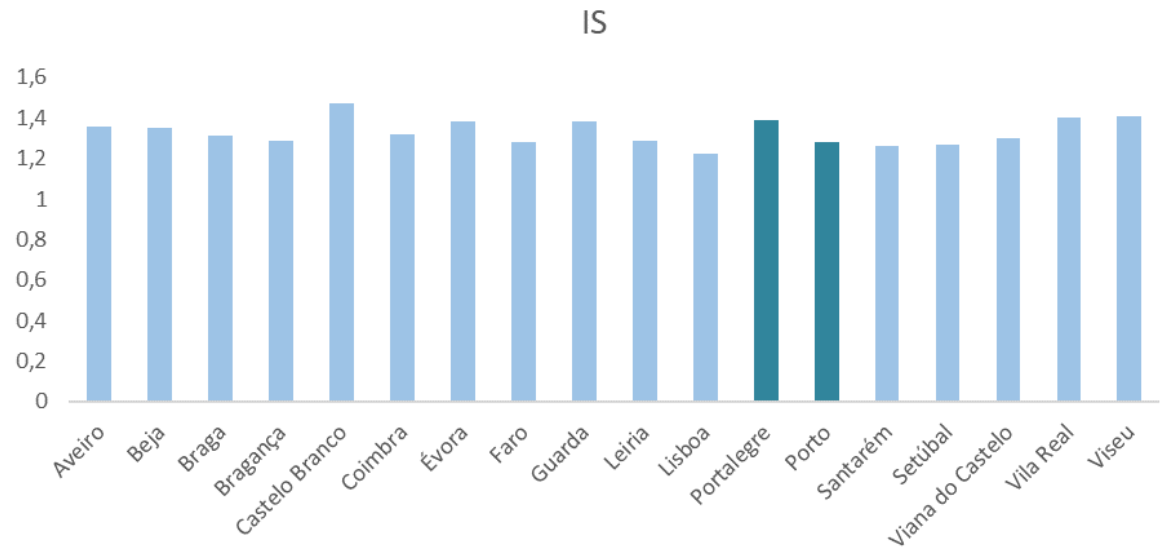


Figura 15 - Índice de sinuosidade por distrito

Estas justificações ajudam a corroborar a H2 de que a estrutura da rede rodoviária influencia a gravidade dos acidentes, quer pelo tipo de vias quer pela sinuosidade das mesmas.

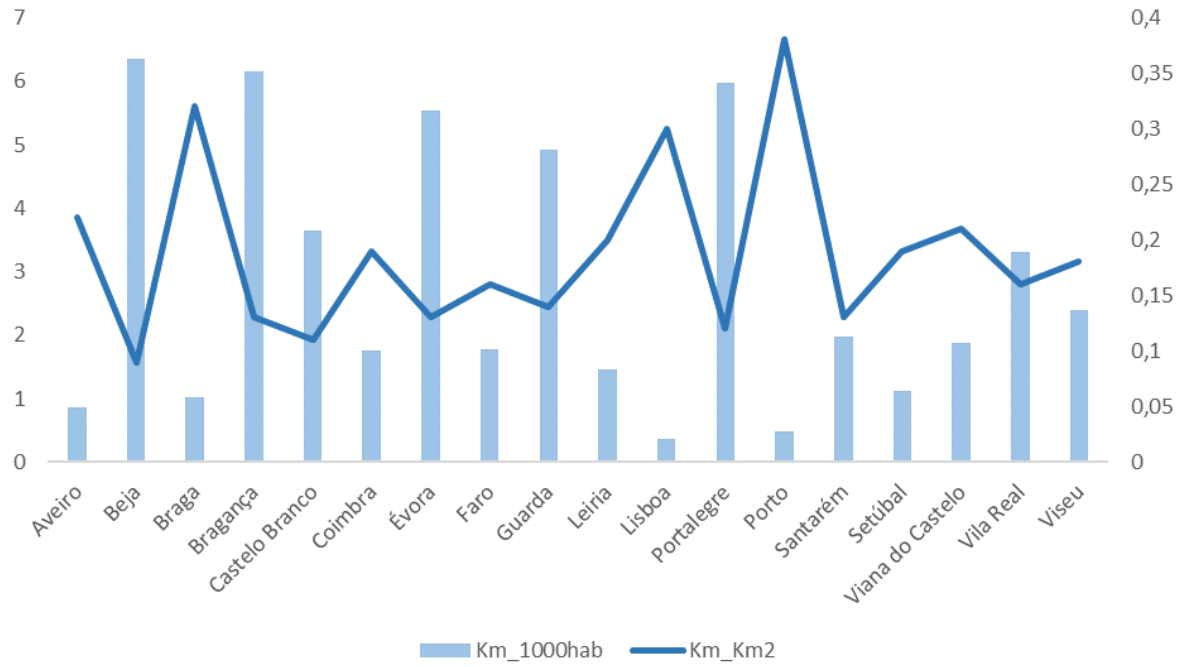


Figura 16 - Cobertura da rede rodoviária por distrito

Para além da sinuosidade, a densidade da rede rodoviária poderá ter influência na sinistralidade. Analisando a cobertura proporcionada pela rede rodoviária nacional (Figura 16), verificamos que a densidade rodoviária de km/km2 no Porto é muito mais elevada do que em Portalegre, o que significa que no Porto uma percentagem muito maior do território está coberta por estradas, comparativamente com Portalegre. No entanto, verificamos que em Portalegre existe uma densidade

rodoviária de km/1000 habitantes, muito superior à verificada no Porto, o que significa que o volume de tráfego por km será muito inferior. Naturalmente, isto permite maiores velocidades de circulação, resultantes de um menor volume de tráfego, o que poderá justificar o aumento da proporção dos acidentes graves em Portalegre.

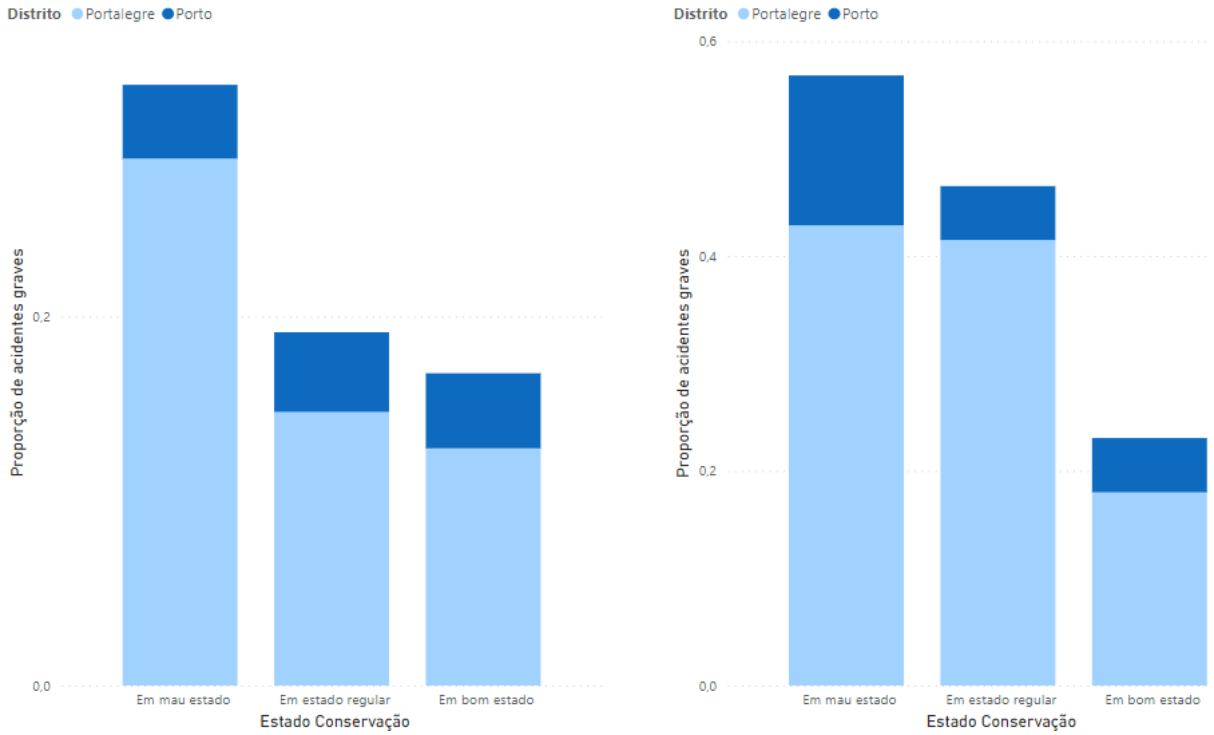


Figura 17 - Proporção de acidentes graves por estado de conservação das vias dentro e fora das localidades, respetivamente

A proporção de acidentes graves em Portalegre é sempre superior à verificada no Porto (Figura 17), independentemente do estado de conservação das estradas. Este facto parece reforçar as hipóteses de que a maior proporção de acidentes graves em Portalegre poderá estar eventualmente mais relacionada com o tipo de estrada (uma ou duas faixas e índice de sinuosidade) e uma maior velocidade de circulação em função de uma menor densidade de tráfego.

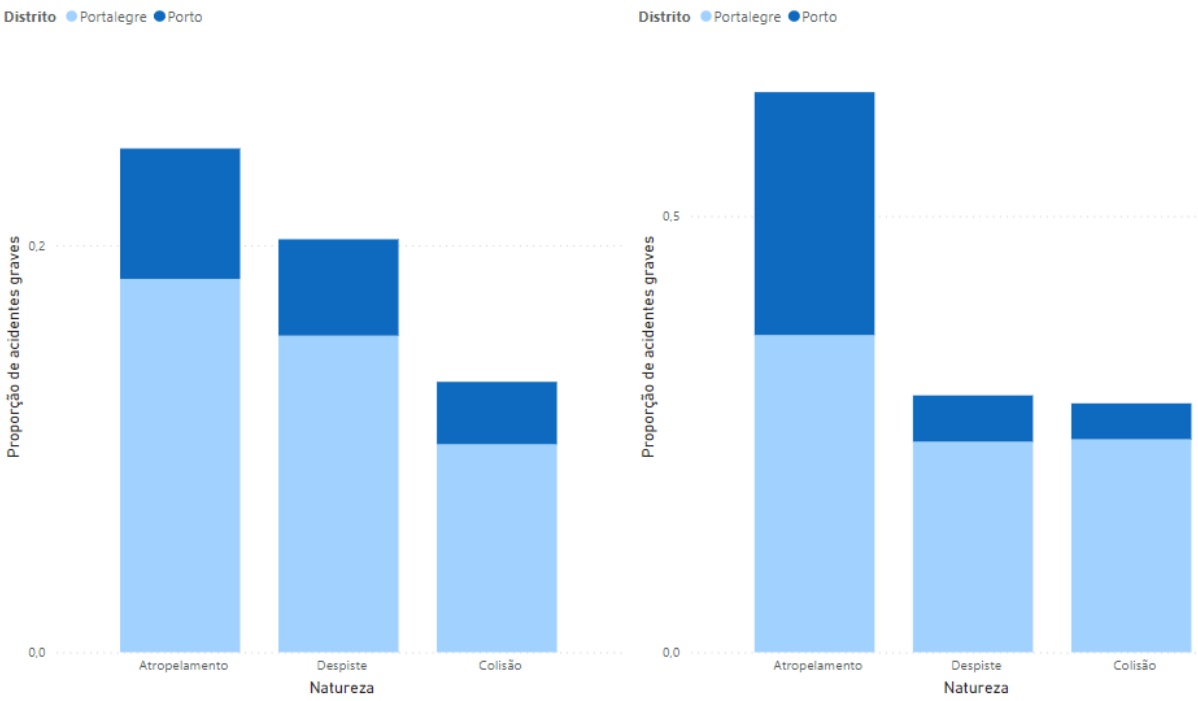


Figura 18 -Proporção de acidentes graves por natureza do acidente dentro e fora das localidades, respetivamente

A análise da natureza dos acidentes graves fora das localidades (Figura 18), indica-nos que os despistes e as colisões são a causa da maior proporção de acidentes graves no distrito de Portalegre, uma vez que a proporção de acidentes graves por atropelamento é relativamente idêntica em ambos os distritos. Mais uma vez, o tipo de estrada e uma maior velocidade de circulação poderão ser explicações possíveis para as diferenças encontradas.

Avaliação no território continental

Após a comparação entre os distritos do Porto e Portalegre e a identificação de diferenças espaciais e nas infraestruturas nestes distritos, que poderão estar envolvidas no risco de acidente grave, importa verificar se esta análise é também válida para o conjunto do território nacional continental.

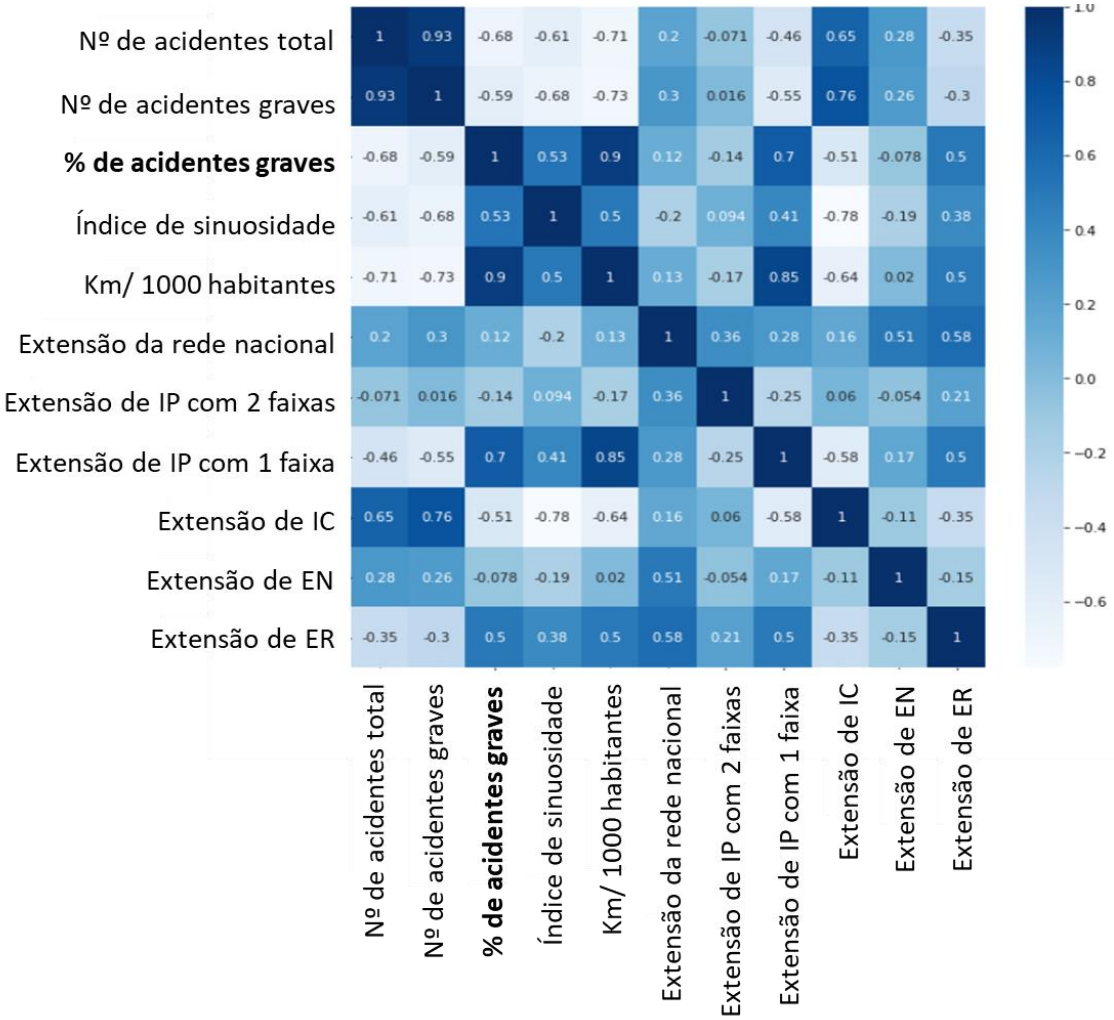


Figura 19 - Correlações entre as variáveis em todos os distritos

Observando as correlações entre as variáveis em análise, para todos os distritos (Figura 19), verificamos que, por ordem decrescente de correlação, as variáveis que mais influenciaram a percentagem de acidentes graves foram:

- De forma positiva: a Densidade da Rede Rodoviária por habitante (km/1000hab), a Extensão de IPs com 1 faixa de rodagem em cada sentido, o Índice de Sinuosidade e a Extensão de Estradas Regionais (ER).
- De forma negativa: o Número Total de Acidentes, o Número de Acidentes Graves e a Extensão de Itinerários Complementares (IC).

As restantes variáveis, tem pouca influência na proporção de acidentes graves.

A nível nacional, a densidade da rede rodoviária por 1000hab apresentou a maior correlação positiva com a proporção de acidentes graves, confirmando o que se verificou anteriormente na comparação entre o Porto e Portalegre. A H1 parece ficar assim confirmada, uma vez que uma

densidade de km/1000hab mais baixa permite velocidades superiores e, consequentemente, uma maior proporção de acidentes graves.

Também o IS e a extensão de IPs com 1 faixa se correlacionam fortemente, de uma forma positiva, com a proporção de acidentes graves. Estas duas variáveis estão relacionadas com a estrutura da rede rodoviária e, em conjunto, ajudam a corroborar a H2, de que a rede rodoviária influencia a gravidade dos acidentes. Pelo contrário, os IP com 2 faixas relacionam-se negativamente, comprovando que a presença de 2 faixas previne os acidentes graves. A importância da estrutura da rede rodoviária é também confirmada pelo efeito positivo da extensão de ER e negativo da extensão de IC, na taxa de acidentes graves.

Por último, o facto de a proporção apresentar uma correlação negativa com o número de acidentes e o número de acidentes graves, também poderá estar relacionado com o facto de um maior número de acidentes está associado a um maior volume de tráfego que, reduzindo a velocidade de circulação, resulta em acidentes menos graves.

Em resumo, a análise nacional confirmou que as variáveis identificadas com maior peso na gravidade dos acidentes, na comparação entre os distritos do Porto e Portalegre, são aplicáveis a todo o território nacional continental.

3.3.3 Análise Temporal

O estudo do efeito das variáveis temporais na proporção de acidentes graves poderá também ajudar a testar as hipóteses colocadas nesta dissertação. Neste estudo, optou-se por avançar diretamente para a avaliação a nível nacional, uma vez que não se verificam diferenças entre os distritos.

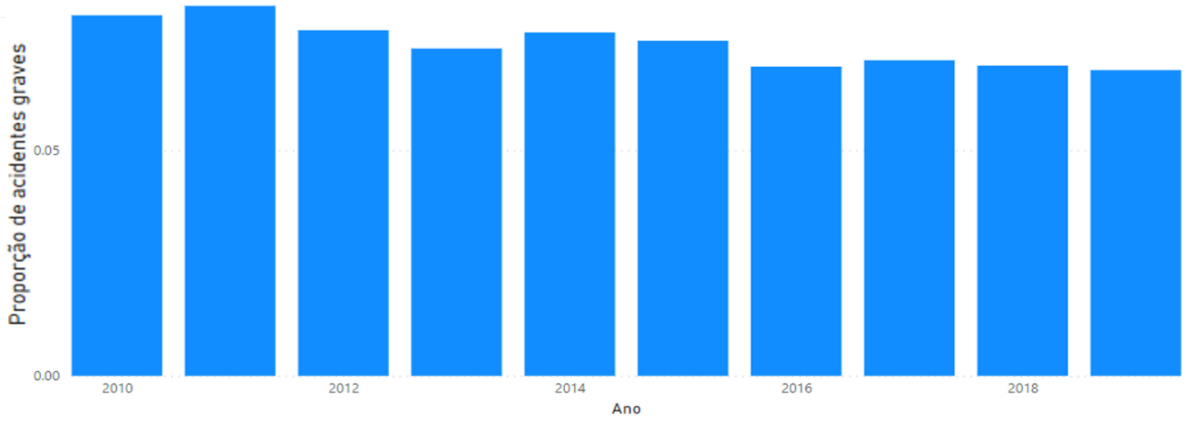


Figura 20 - Evolução da proporção de acidentes graves

A proporção dos acidentes graves apresenta uma ligeira diminuição ao longo dos anos em estudo (Figura 20), o que poderá significar um maior investimento nas infraestruturas ou na fiscalização.

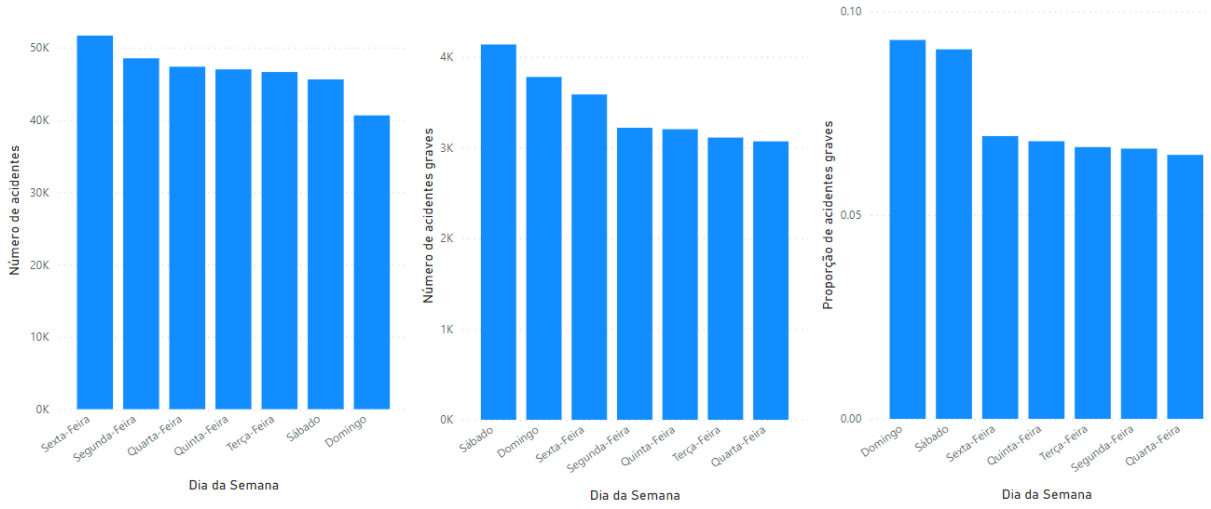


Figura 21 - Nº de acidentes, nº de acidentes graves e proporção de acidentes graves por dia da semana, respetivamente

Verificamos que o maior número de acidentes rodoviários ocorre às sextas-feiras. Verificamos ainda que ao fim de semana ocorrem menos acidentes do que durante a semana, o que provavelmente estará relacionado com a quantidade de tráfego, que deverá ser superior durante a semana (Figura 21).

O número de acidentes graves, por outro lado, é superior ao fim de semana. O domingo, embora seja o dia com menos acidentes, é o dia da semana com maior proporção de acidentes graves (Figura 21).

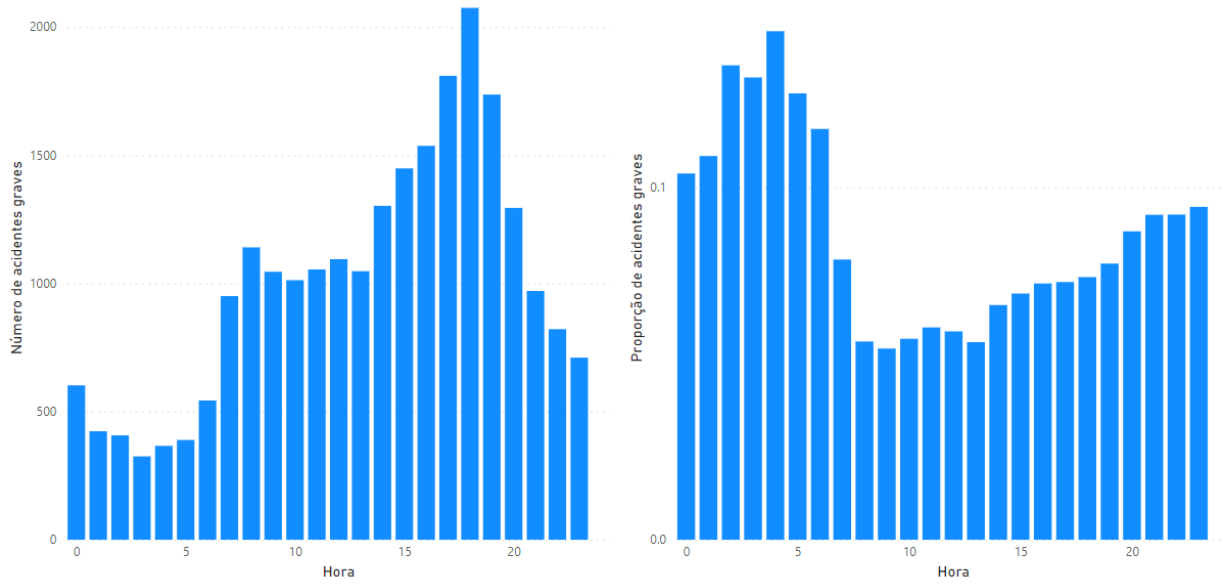


Figura 22 - Número e Proporção de acidentes graves por hora do dia

Verificamos que há mais acidentes graves ao final do dia o que, novamente, poderá ser justificado por um maior volume de tráfego a estas horas, que correspondem com a hora de regresso a casa do trabalho. No entanto, se considerarmos a proporção de acidentes graves, verificamos que os acidentes de madrugada são mais graves (Figura 22).

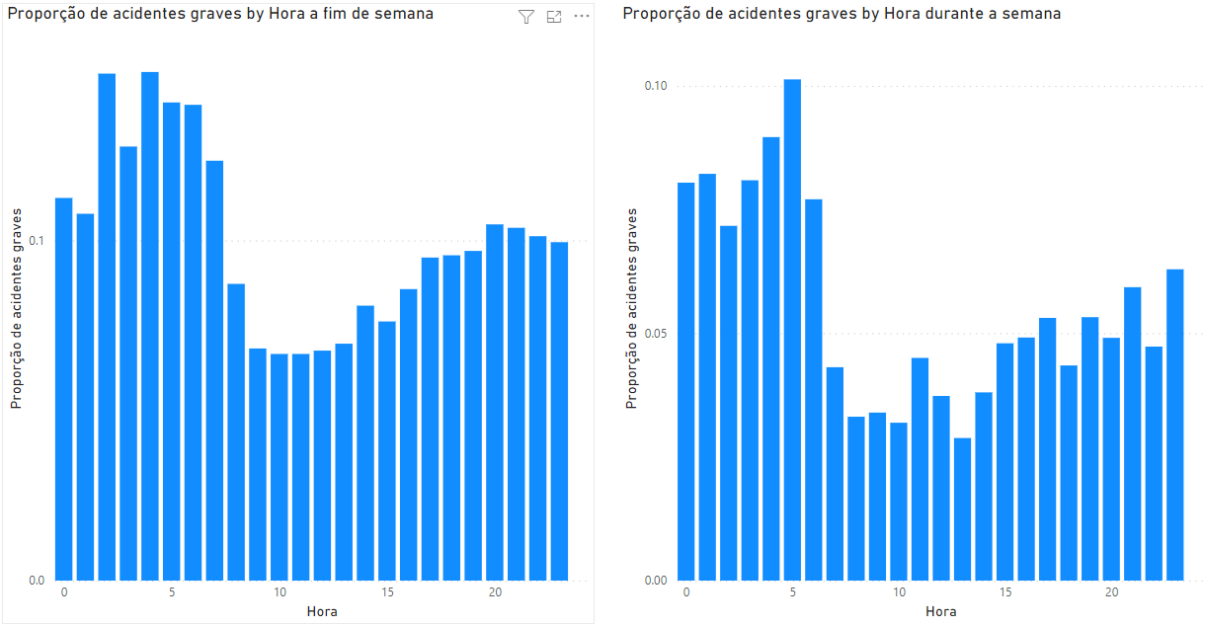


Figura 23 - Proporção de acidentes graves por hora do dia ao fim de semana e durante a semana

Ao fim de semana, bem como de madrugada, é de admitir uma menor densidade de tráfego, e ainda um aumento das deslocações fora das localidades ao fim de semana. Estas situações potenciam uma maior velocidade de deslocação. Esta análise temporal vem corroborar a análise espacial e a H1 de que uma redução da densidade de tráfego aumenta a velocidade e, consequentemente, o risco de acidente grave.

Embora seja visível que, quer ao fim de semana, quer durante a semana, a proporção de acidentes graves é superior de madrugada, esta diferença é mais marcada durante a semana (Figura 23). Isto poderá significar que ao fim de semana o tráfego é inferior, pelo que a velocidade ao longo do dia não será tão condicionada pela densidade de tráfego. Nos dias de semana, uma vez que se verificam elevados volumes de tráfego durante o dia, a velocidade praticada é menor e a proporção de acidentes graves nestas horas também é menor.

4 Modelação de Dados

A seleção de métodos estatísticos errados pode conduzir a conclusões inválidas da análise de acidentes rodoviários [1].

A seleção do método certo de análise de dados ajudará a descobrir as razões da ocorrência de acidentes numa determinada zona ou local de estudo e a prever com precisão razoável a probabilidade de ocorrência de acidentes por dia ou a relativa segurança de várias categorias de utentes da estrada naquela região [32]. Assim, é fundamental selecionar o método adequado para obter uma análise o mais precisa possível.

Os modelos de análise / previsão são divididos em frequência (previsão da ocorrência de um determinado número de acidentes num segmento de estrada) e modelos de classificação (onde é explorada a relação entre severidade dos ferimentos e a contribuição de fatores). Para que seja possível predizer a frequência dos acidentes rodoviários, é necessário que existam dados referentes ao volume de tráfego. Assim, uma vez que não temos esses dados disponíveis, optou-se por utilizar apenas modelos de classificação (Figura 24).

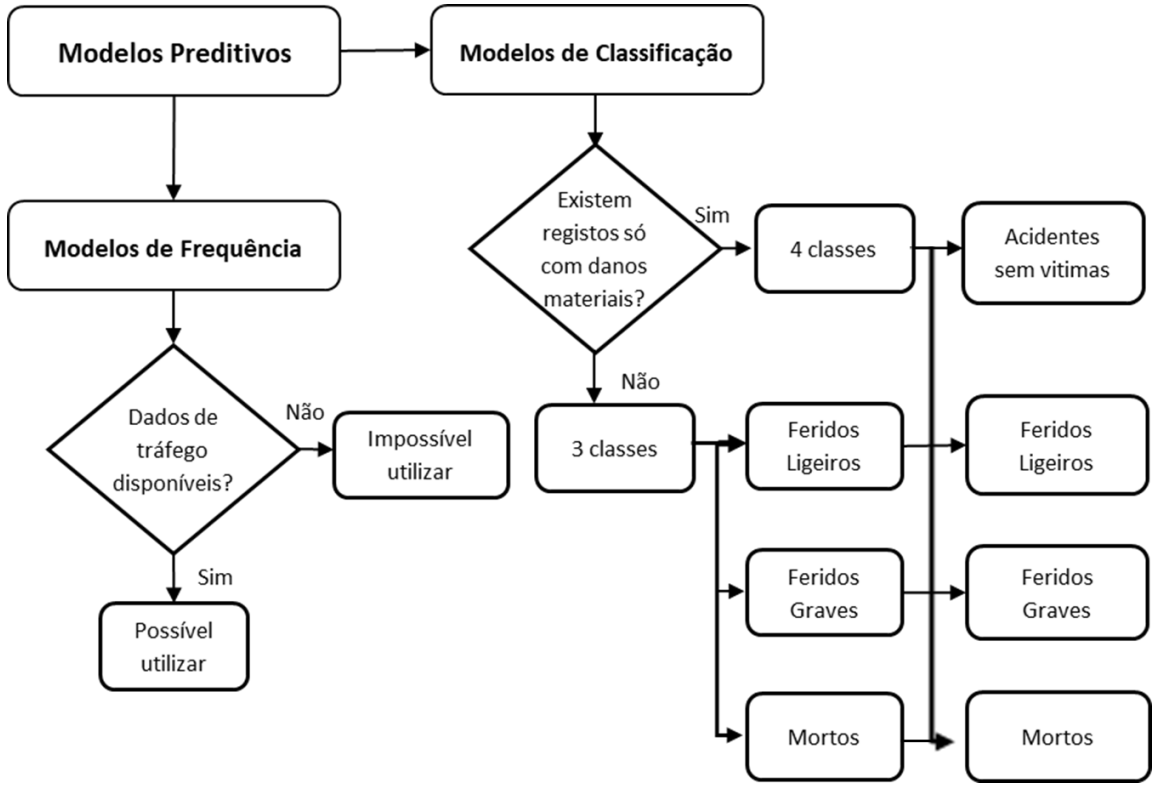


Figura 24 - Modelos preditivos na análise de acidentes rodoviários

Atualmente os dados de acidentes sem vítimas são pouco completos, porque as forças policiais nem sempre são chamadas para tomar a ocorrência destes acidentes. Utilizando modelos de classificação, terão de ser selecionados modelos de 3 classes.

Numa tentativa de dar resposta ao segundo objetivo proposto e apoiar a ANSR no seu objetivo de potenciar as capacidades de análise de dados incluindo o recurso a modelos preditivos de ocorrência de acidentes e de gravidade das lesões resultantes, foram realizadas duas tentativas de criação de modelos predição de acidentes rodoviários: em função das características dos veículos e em função das características dos acidentes.

4.1 Predição de acidentes em função das características dos veículos

Com o objetivo de entender a probabilidade de determinados condutores estarem envolvidos num acidente grave, tentou elaborar-se um modelo de classificação que prevê, com base nas características dos condutores e respetivos veículos, a probabilidade e, consequentemente, o nível de risco do mesmo se envolver num acidente do qual resultem feridos graves ou vítimas mortais a 30 dias.

Como referido no Capítulo 3, foi definida a variável dependente como uma variável binária "Gravidade", onde "1" indica a existência de feridos graves ou vítimas mortais a 30 dias e "0" representa a ausência dos mesmos.

O primeiro passo para o desenvolvimento deste modelo foi então selecionar as variáveis independentes a considerar. Como referido no capítulo 3, após a preparação dos dados, obtivemos as variáveis "Sexo", "Categoria Veículos", "Tipo Veículo", "Ano Matrícula", "Inspeção Periódica", "Seguros" e "Grupo Etário". Para determinar as variáveis mais relevantes para esta análise elaborou-se um mapa de correlações que nos permite visualizar as correlações entre as variáveis independentes e a variável dependente, ver Figura 25.

Observando estas correlações foi possível perceber que, isoladamente, nenhuma das variáveis disponíveis se correlaciona com a Gravidade dos acidentes. Nesse sentido, optou-se por manter todas elas e perceber qual o seu efeito combinado na variável dependente. Foram, assim, utilizadas como variáveis independentes os atributos "Sexo", "Categoria Veículos", "Tipo Veículo", "Ano Matrícula", "Inspeção Periódica", "Seguros" e "Grupo Etário".

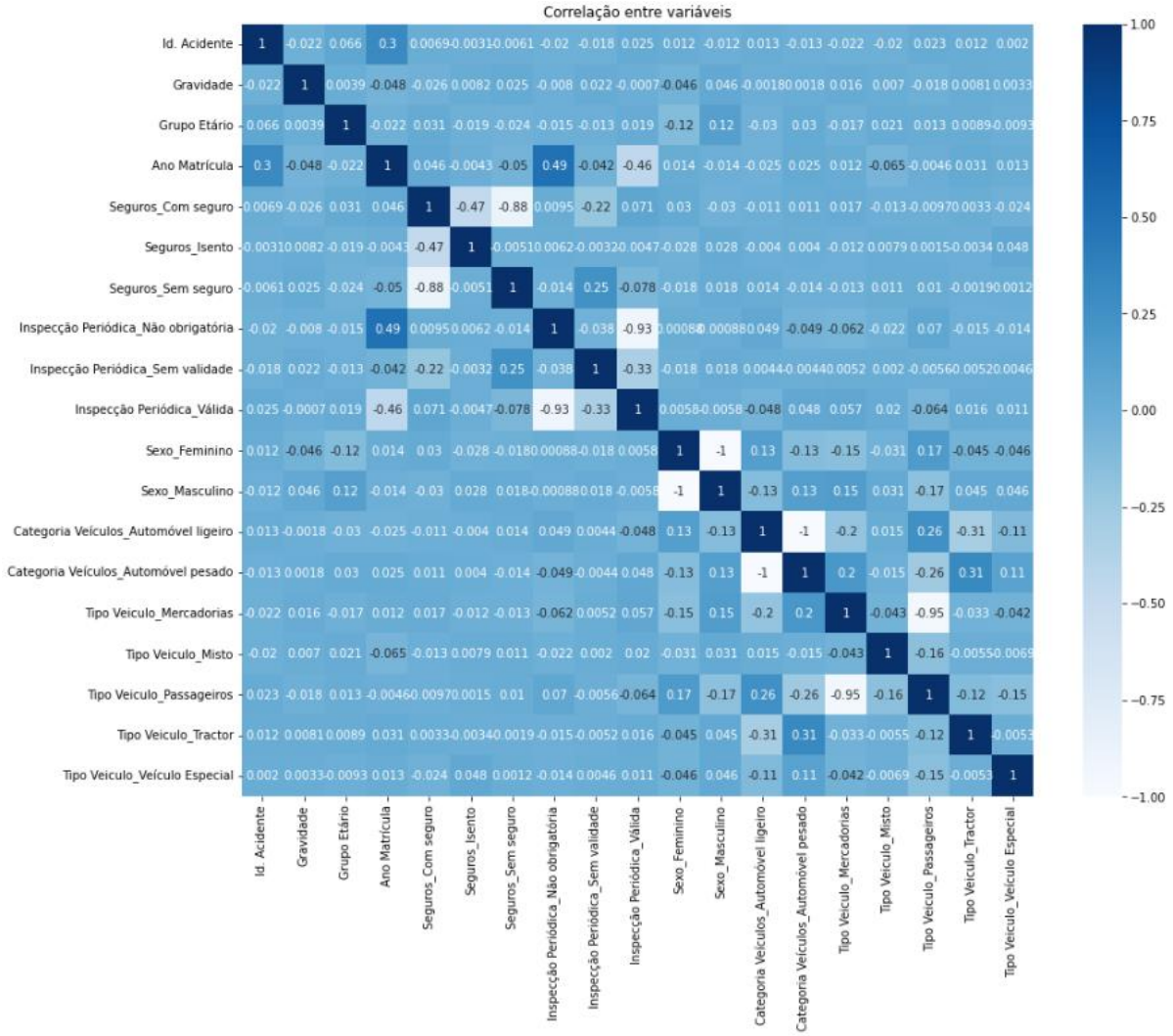


Figura 25 - Correlações entre as características dos condutores/veículos e a gravidade

Para a criação deste modelo, foram incluídos 70% dos dados no conjunto de treino e os 30% restantes no conjunto de teste.

Importa salientar que as classes da variável dependente não se encontravam balanceadas nos registos de treino, onde cerca 97,3% correspondiam à classe “0” e 2,7 % correspondiam à classe “1”. Isto faz com que os modelos funcionem muito bem para os casos em que a variável faz parte da classe mais significativa, e por isso apresentarem altos valores de precisão. Para que a criação destes modelos fosse mais realista e assertiva seria necessário recolher mais acidentes com maiores valores de Gravidade. Neste sentido, tornou-se necessário proceder à técnica de *SMOTE Oversample*, por forma a equilibrar os dados de treino.

Após o balanceamento dos dados, obtivemos a seguinte distribuição:

Classe 0, n=97.050 (50%)

Classe 1, n=97.050 (50%)

Foram testados 2 modelos: Regressão Logística e *Random Forest*, sendo que o modelo que obteve o melhor desempenho na classificação da gravidade dos acidentes foi o *Random Forest,* produzindo os seguintes resultados:

	precision	recall	f1-score	support
0	0.981	0.542	0.699	29133
1	0.035	0.618	0.067	788
accuracy			0.544	29921
macro avg	0.508	0.580	0.383	29921
weighted avg	0.956	0.544	0.682	29921

Figura 26 - Resultados da predição em função das características dos veículos

```
array([[15753, 13380],
       [  300,   488]])
```

Figura 27- Matriz de confusão da predição em função das características dos veículos

Os resultados (Figura 26, Figura 27), percebemos que para os acidentes não graves (classe 0), o modelo acerta 15753 vezes e erra 13380 vezes, enquanto para os acidentes graves acerta 488 e erra 300. Percebemos que existem no conjunto de teste muito mais dados com acidentes não graves, pelo que o teste poderá não ser o mais assertivo.

Podemos concluir que, embora algumas variáveis pareçam influenciar ligeiramente a gravidade dos acidentes, de um modo geral, as características dos veículos e dos condutores, a priori, não têm uma influência muito significativa na previsão da gravidade do mesmo. Concluímos ainda que, com o conjunto de dados disponível, bastante desequilibrado, torna-se difícil para o modelo fazer a predição da ocorrência de acidentes graves.

4.2 Predição de acidentes em função das características dos acidentes

Considerando as conclusões anteriores, de que as características dos veículos e dos condutores, por si só, não são suficientes para predizer a gravidade dos acidentes, tentou-se prever esta gravidade com base nas características dos acidentes.

Novamente, começou-se por determinar as variáveis independentes a incluir no modelo. Como, mais uma vez, nenhuma das variáveis isoladamente se correlacionava com a gravidade dos acidentes, optou-se por utilizar para a análise aquelas que, segundo a revisão da literatura e o trabalho de

visualização anteriormente realizado, foram consideradas mais relevantes para determinar a gravidade dos acidentes. Consideraram-se, nesta fase, as Características Técnicas, as Condições de Aderência do Piso, os Tipos de Vias, o Estado de Conservação das Vias, os Fatores Atmosféricos, o Tipo de Interseção das Vias, a Localização dos Acidentes (dentro/fora da localidade), a Luminosidade da Via, o Tipo de Piso, os Sinais de Trânsito e Sinais Luminosos, a Curvatura e Inclinação da Via e o Tipo de Berma. Foi importante optar apenas por variáveis que não dependessem do comportamento do condutor, uma vez que não é possível prever o mesmo e, consequentemente, tomar decisões por forma a agir sobre ele.

Procedeu-se novamente ao balanceamento das classes utilizando a técnica *SMOTE oversample*, com a qual se obteve a seguinte distribuição:

Classe 0, n=191.218 (50.000%)

Classe 1, n=191.218 (50.000%)

Voltaram a aplicar-se os modelos de Regressão Logística e *Random Forest* e, novamente, verificou-se que o modelo que melhor resultado obteve foi o de Regressão Logística, ver Figura 28.

	precision	recall	f1-score	support
0.0	0.93	0.89	0.91	81985
1.0	0.13	0.22	0.17	6541
accuracy			0.84	88526
macro avg	0.53	0.55	0.54	88526
weighted avg	0.88	0.84	0.85	88526

Figura 28 - Resultados da predição em função das características dos acidentes

```
array([[72660,  9325],
       [ 5111,  1430]])
```

Figura 29 - Matriz de confusão da predição em função das características dos acidentes

Observando os resultados, percebemos que o modelo prevê de forma assertiva os acidentes não graves, no entanto, apresenta muitos erros ao classificar os acidentes graves. Isto resulta do facto de existirem poucos registos de acidentes graves no conjunto de teste. Para que este modelo fosse eficaz, seria necessário termos mais dados de acidentes graves.

Adicionalmente, considerando os objetivos da ANSR, seria interessante tentar perceber a influência destas mesmas condições na ocorrência ou não de acidentes. Para isso, seria, no entanto,

necessário obter dados sobre as condições das estradas e do ambiente em situações de não ocorrência de acidentes, o que me parece bastante complicado.

Outra possibilidade seria tentar predizer a frequência de acidentes em função do tráfego, uma vez que é mais provável que ocorram mais acidentes em situações de maior tráfego, mesmo que as condições à partida pareçam mais favoráveis. No entanto, mais uma vez, para que esta análise fosse possível seriam necessários dados sobre o volume de tráfego para cada estrada.

5 Conclusões e Trabalho Futuro

Neste trabalho pretende-se averiguar as causas relacionadas com o aumento da proporção de acidentes graves. Foram colocadas duas hipóteses de trabalho que se prendem, por um lado, com o aumento da velocidade de circulação, condicionado pela densidade de tráfego (Hipótese um (H1): A redução da densidade de tráfego, aumentando a velocidade de circulação, aumenta o risco de acidentes graves), e por outro com a estrutura da rede rodoviária (Hipótese dois (H2): A estrutura da rede rodoviária e das infraestruturas influencia a gravidade dos acidentes rodoviários). Para testar estas hipóteses, fez-se uma análise espacial e temporal da proporção de acidentes graves.

Os resultados obtidos indicam de forma coerente que todas as situações que proporcionam uma redução da densidade de tráfego aumentam o risco de acidente grave. É o caso do aumento da cobertura da rede rodoviária por 1000 habitantes, que apresenta uma correlação de 0,9 com o aumento da percentagem de acidentes graves, assim como os períodos da semana ou do dia em que o tráfego diminui, que apresentam percentagens de acidentes graves superiores. A relação entre o aumento da velocidade e a gravidade dos acidentes está referida na literatura por [13], [15], [16],[17], sendo que neste trabalho se conseguiu identificar fatores que potenciam este aumento da velocidade.

A estrutura da rede rodoviária influencia também a gravidade dos acidentes, conforme é reconhecido por diversos autores [13], [15], [17]. A análise dos dados para Portugal Continental revelou que as características da infraestrutura rodoviária que maior influência apresentaram no aumento da taxa de acidentes graves foram a sinuosidade das estradas e a extensão de IPs com 1 só faixa de rodagem.

Atendendo aos custos económicos e sociais dos acidentes graves, é extremamente importante refletir-se sobre como se poderá atuar sobre os fatores identificados no sentido de reduzir este flagelo nacional.

Esta análise terá forçosamente de considerar os recursos disponíveis e uma relação custo/benefício. Não é razoável procurar uma redução da cobertura da rede viária por 1000 habitantes, até porque esta cobertura aumenta nos distritos com menor densidade populacional e não por um aumento exagerado da construção de infraestruturas rodoviárias. Assim, a solução que parece mais óbvia para reduzir a gravidade dos acidentes em locais em que a densidade da rede rodoviária permite maiores velocidades de circulação, será um aumento da fiscalização da velocidade nestes locais. O mesmo se pode dizer em relação aos períodos do dia/semana em que a redução da densidade de tráfego, permitindo um aumento da velocidade de circulação, aumenta o risco de acidentes graves.

Está fora do âmbito deste trabalho identificar as soluções com maior custo/benefício para a fiscalização da velocidade de circulação, mas conseguir-se uma redução desta velocidade parece ser essencial para reduzir a gravidade do problema. Como identifica [19] esta redução da velocidade pode ser conseguida através, por exemplo, da instalação de lombas de velocidade, a adoção de zonas de 30km/h, a fiscalização policial do excesso de velocidade, o controlo das velocidades médias por troço, etc.

As duas características identificadas com maior impacto na gravidade dos acidentes foram a sinuosidade das estradas e a existência de IPs com 1 só faixa de rodagem. A alteração destas características exigirá elevados investimentos, que deverão ser ponderados em relação a outras alternativas. Será preciso averiguar se a sinalização das estradas identifica corretamente as zonas de maior perigo e, eventualmente, a correção do traçado nos locais de maior incidência de acidentes graves. Também um reforço da fiscalização nestas zonas poderá contribuir para uma melhoria da situação. No entanto, a localização exata dos acidentes é uma falha importante nos dados que não permite identificar estas zonas com a precisão necessária, conforme se pode verificar na Figura 7.

Em resumo, com este trabalho foi possível concluir que:

- A redução da densidade de tráfego aumenta o risco de acidentes graves.
- O aumento da cobertura da rede rodoviária por 1000 habitantes aumenta o risco de acidentes graves.
- Aos fins de semana e de madrugada, o risco de acidentes graves aumenta.
- Estas conclusões indicam que todas estas situações contribuem para um aumento da velocidade de circulação e que esta será causa do aumento da gravidade dos acidentes.
- A estrutura da rede rodoviária influencia o risco de acidentes graves.
- A extensão de IPs com 1 faixa aumenta o risco de gravidade dos acidentes, tendo sido, dentro das características da estrutura da rede rodoviária, aquela que maior correlação apresentou.
- O índice de sinuosidade das estradas aumenta a proporção de acidentes graves.
- É importante medir o tráfego e a velocidade de circulação para ajudar a confirmar estas hipóteses.
- A qualidade dos dados no que toca às coordenadas dos acidentes é uma grande limitação na identificação dos locais de maior risco, pelo que deve ser uma prioridade.

Referências

[1] A. Chand, S. Jayesh, and A. B. Bhasi, "Road traffic accidents: An overview of data sources, analysis techniques and contributing factors," in *Materials Today: Proceedings*, 2021, vol. 47, pp. 5135–5141.

[2] J. M. Silva, C. M. P. da, Bravo, J. M. V., & Gonçalves, "Impacto Económico e Social da Sinistralidade Rodoviária em Portugal," 2021.

[3] PORDATA, "Número de Veículos Ligeiros de Passageiros," 2019. https://www.pordata.pt/europa/veiculos+registados+por+tipo+de+veiculo-3070-3886 (accessed Oct. 30, 2022).

[4] ANSR, "Visão Zero 2030," *Estratégia Nacional de Segurança Rodoviária 2021 / 2030*, 2021. https://visaozero2030.pt/ (accessed Oct. 30, 2022).

[5] M. M. Peden, World Health Organization., and World Bank., *World report on road traffic injury prevention*. World Health Organization, 2004.

[6] PIARC, "Road Safety Manual, A Manual for Practitioners and Decision Makers on Implementing Safe System Infrastructure," 2019. http://roadsafety.piarc.org/en%0Ahttps://trid.trb.org/view/1374814 (accessed Oct. 30, 2022).

[7] International Transport Forum, "New Directions for Data-Driven Transport Safety Corporate Partnership Board Report," 2019.

[8] European Commission, "EU road safety policy framework 2021-2030 - Next steps towards 'Vision Zero,'" Luxembourg, 2020. doi: 10.2832/391271.

[9] International Traffic Safety Data and Analysis Group, "Road Safety Annual Report," 2019.

[10] R. Wirth and J. Hipp, "CRISP-DM: towards a standard process model for data mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 29-39," 2000.

[11] B. Kitchenham, "Guidelines for performing Systematic Literature Reviews in Software Engineering," 2007.

[12] P. B. Silva, M. Andrade, and S. Ferreira, "Machine learning applied to road safety modeling: A systematic literature review," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 7, no. 6. Chang'an University, pp. 775–790, 2020.

[13] R. Elvik, A. Hoye, T. Vaa, and M. Sorensen, *The Handbook of Road Safety Measures*, Second Edi. Emerald Group Publishing Limited, 2009.

[14] G. Nilsson, "Traffic Safety Dimensions and the Power Model to Describe the Effect of Speed on Safety," 2004.

[15] American Association of State Highway Transportation Officials, “Highway Safety Manual,” First Edit., 2009.

[16] D. Mohan, G. Tiwari, M. Khayesi, and F. M. Nafukho, “Road traffic injury prevention training manual,” 2006.

[17] R. J. Alharbi, V. Lewis, and C. Miller, “A state-of-the-art review of factors that predict mortality among traumatic injury patients following a road traffic crash,” *Australasian Emergency Care*. 2021.

[18] Md. M. Haque, S. Washington, and B. Watson, “A Methodology for Estimating Exposure-controlled Crash Risk Using Traffic Police Crash Data,” *Procedia Soc Behav Sci*, vol. 104, pp. 972–981, 2013.

[19] S. Daniels, H. Martensen, and A. Schoeters, “A systematic cost-benefit analysis of 29 road safety measures,” *Accid Anal Prev*, vol. 133, 2019.

[20] C. Gutierrez-Osorio and C. Pedraza, “Modern data sources and techniques for analysis and forecast of road accidents: A review,” *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 7, no. 4. Chang’an University, pp. 432–446, 2020.

[21] D. Topolšek, E. A. Herbaj, and M. Sternad, “The Accuracy Analysis of Measurement Tools for Traffic Accident Investigation,” *J Transp Technol*, vol. 04, no. 01, pp. 84–92, 2014.

[22] J. L. Mazher Iqbal and M. Id, “Automatic Vehicle Accident Detection and Reporting With Black Box Article in,” *International Journal of Applied Engineering Research*, vol. 10, no. 87, pp. 139–145, 2015.

[23] P. Pete Thomas, R. Author, and E. Dupont, “Consultation of a panel of experts on the needs for data and technical tools in road safety policy-making, Deliverable 1.1/4.1 of the EC FP7 project DaCoTA. Grant,” 2010.

[24] A. Ahmed, A. F. M. Sadullah, and A. S. Yahya, “Errors in accident data, its types, causes and methods of rectification-analysis of the literature,” *Accid Anal Prev*, vol. 130, pp. 3–21, 2019.

[25] D. Miguel and C. Borges, “Classificação automática de locais de acidentes rodoviários em Lisboa - no âmbito dos desafios do LxDataLab,” 2021.

[26] R. W. David D. Clarke, Richard Forsyth, “Behavioural factors in accidents at road junctions: The use of a genetic algorithm to extract descriptive rules from police case files,” in *Accident Analysis & Prevention, Volume 30, Issue 2*, 1998, pp. 223–234.

[27] Y. J. K. Y. Castro, “Data mining on road safety: avaliação de fatores em veículos com modelos de classificação,” *International Journal of Crashworthiness*, no. 104–111, 2016.

[28] Z. Li, P. Liu, W. Wang, and C. Xu, “Using support vector machine models for crash injury severity analysis,” *Accid Anal Prev*, vol. 45, pp. 478–486, 2012.

[29] B. Y. Naseer, A., Nour, M. K., & Alkazemi, “Towards Deep Learning based Traffic Accident Analysis,” 2020, pp. 0817–0820.

[30] F. R. Moghaddam, S. Afandizadeh, and M. Ziyadi, “Prediction of accident severity using artificial neural networks,” *International Journal of Civil Engineering*, vol. 9, no. 1, pp. 41–49, 2011.

[31] PORDATA, “Censos 2021: Conheça o seu Município,” 2021. https://www.pordata.pt/Municipios/Quadro+Resumo (accessed Oct. 30, 2022).

[32] M. A. Q. Peter T. Savolainen, Fred L. Mannering, Dominique Lord, *The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives, Volume 43, Issue 5*. 2011.

ANEXOS

Tabela 7 - Dados dos acidentes

Variável	Descrição	Tipo
Id. Acidente	Id. do acidente	Numérica
Datahora	Data e hora	Numérica
Dia	Dia	Numérica
Mês	Mês	Numérica
Hora	Hora	Numérica
Entidades Fiscalizadoras	Entidades fiscalizadoras	Categórica
Velocidade local	Velocidade local	Numérica
Velocidade geral	Velocidade geral	Numérica
Dia da Semana	Dia da semana	Categórica
Latitude GPS	Latitude em graus decimais	Numérica
Longitude GPS	Longitude em graus decimais	Numérica
Num. Mortos a 30 dias	Número de vítimas cujo óbito ocorra no período de 30 dias após o acidente	Numérica
Num. Feridos graves a 30 dias	Número de vítimas cujos danos corporais obrigam a um período de hospitalização superior a 24 horas e que não venham a falecer nos 30 dias após o acidente	Numérica
Num. Feridos ligeiros a 30 dias	Número de vítimas que não sejam consideradas feridos graves e que não venham a falecer nos 30 dias após o acidente	Numérica
Características Tecnicas1	Caracterização técnica da via	Categórica
Cond Aderência	Condições de aderência da estrada	Categórica
Distrito	Distrito	Categórica
Concelho	Concelho	Categórica
Freguesia	Freguesia	Categórica
Pov. Proxima	Povoação mais próxima	Categórica
Nome arruamento	Nome do arruamento	Categórica
Tipos Vias	Tipo de via	Categórica
Cod Via	Código da via	Categórica
Estado Conservação	Estado de conservação da via	Categórica
Km	Km da via onde ocorreu o acidente	Numérica
Fatores Atmosféricos	Fatores atmosféricos	Categórica
Reg Circulação1	Sentidos de circulação disponíveis	Categórica
Intersecção Vias	Tipo de interseção da via	Categórica

Localizações	Localização do acidente relativamente a localidades	Categórica
Luminosidade	Condições de luminosidade	Categórica
Marca Via	Marcas rodoviárias na via	Categórica
Natureza	Natureza do acidente	Categórica
Obras Arte	Obra de arte na qual o acidente ocorreu	Categórica
Obstáculos	Obstáculos na via	Categórica
Sentidos	Sentido da quilometragem	Categórica
Sinais	Sinais de trânsito	Categórica
Sinais Luminosos	Sinais luminosos da estrada	Categórica
Tipo Piso	Tipo de piso	Categórica
Traçado 1	Curvatura da via	Categórica
Traçado 2	Perfil longitudinal da via	Categórica
Traçado 3	Tipo de berma	Categórica
Traçado 4	Zona em que ocorreu o acidente	Categórica
Via Trânsito	Via de trânsito	Categórica

Tabela 8 - Dados dos veículos

Variável	**Descrição**	**Tipo**
Id. Acidente	Id. do Acidente	Numérica
Datahora	Data e hora	Numérica
Sexo	Género do Condutor	Categórica
Lesões a 30 dias	Lesões do condutor a 30 dias	Categórica
Licença Condução	Licença de condução	Categórica
Teste Alcool	Teste de Álcool	Categórica
Acções Condutores	Ação do condutor no momento do acidente	Categórica
Inf. Comp. a Acções e Manobras	Ação ou manobra do condutor associado ao acidente	Categórica
Nomeoutrosfactores	Estado do condutor	Categórica
Tempo Condução Continuada	Tempo de condução continuado do condutor	Categórica
Acessórios Condutores	Acessório de segurança do condutor	Categórica
Categoria Veículos	Categoria do veículo	Categórica
Tipo Veiculo	Tipo de veículo	Categórica
Tipo Serviço	Tipo de serviço	Categórica
Veiculo Especial	Tipo de veículo	Categórica
Ano matricula	Ano da matrícula do veículo	Numérica
Inspecção Periódica	Inspeção periódica do veículo	Categórica
Certificado Adr	Certificado de condutor de mercadorias perigosas	Categórica
Carga Lotação	Carga do veículo	Categórica
Pneus	Estado dos pneus	Categórica
Seguros	Seguro	Categórica
Distrito	Distrito	Categórica
Concelho	Concelho	Categórica

Condutor Gr.Etario(<=5) SUM	Condutor tem menos de 5 anos de idade	Categórica
Condutor Gr.Etario(6-9) SUM	Idade do condutor situa-se no intervalo 6 a 9 anos	Categórica
Condutor Gr.Etario(10-14) SUM	Idade do condutor situa-se no intervalo 10 a 14 anos	Categórica
Condutor Gr.Etario(15-17) SUM	Idade do condutor situa-se no intervalo 15 a 17 anos	Categórica
Condutor Gr.Etario(18-20) SUM	Idade do condutor situa-se no intervalo 18 a 20 anos	Categórica
Condutor Gr.Etario(21-24) SUM	Idade do condutor situa-se no intervalo 21 a 24 anos	Categórica
Condutor Gr.Etario(25-29) SUM	Idade do condutor situa-se no intervalo 25 a 29 anos	Categórica
Condutor Gr.Etario(30-34) SUM	Idade do condutor situa-se no intervalo 30 a 34 anos	Categórica
Condutor Gr.Etario(35-39) SUM	Idade do condutor situa-se no intervalo 35 a 39 anos	Categórica
Condutor Gr.Etario(40-44) SUM	Idade do condutor situa-se no intervalo 40 a 44 anos	Categórica
Condutor Gr.Etario(45-49) SUM	Idade do condutor situa-se no intervalo 45 a 49 anos	Categórica
Condutor Gr.Etario(50-54) SUM	Idade do condutor situa-se no intervalo 50 a 54 anos	Categórica
Condutor Gr.Etario(55-59) SUM	Idade do condutor situa-se no intervalo 55 a 59 anos	Categórica
Condutor Gr.Etario(65-69) SUM	Idade do condutor situa-se no intervalo 65 a 69 anos	Categórica
Condutor Gr.Etario(70-74) SUM	Idade do condutor situa-se no intervalo 70 a 74 anos	Categórica
Condutor Gr.Etario(>=75) SUM	Condutor tem 75 anos ou mais	Categórica
Condutor Gr.Etario(Não Def.) SUM	Idade do condutor desconhecida	Categórica

Tabela 9 - Dados dos passageiros

Variável	Descrição	Tipo
Id. Acidente	Id. do acidente	Numérica
Datahora	Data e hora	Categórica
Entidades Fiscalizadoras	Entidades fiscalizadoras	Categórica
Id. Passageiro	Id. do passageiro	Numérica
Lesões a 30 dias	Lesões do passageiro a 30 dias	Categórica
Sexo	Género do passageiro	Categórica
Posição Veiculo	Posição que o passageiro ocupava no veículo	Categórica
Acessórios Passageiro	Acessório de segurança do passageiro	Categórica
Distrito	Distrito	Categórica
Concelho	Concelho	Categórica
Passageiro Gr.Etario(<=5) SUM	Passageiro tem menos de 5 anos de idade	Categórica
Passageiro Gr.Etario(6-9) SUM	Idade do passageiro situa-se no intervalo 6 a 9 anos	Categórica
Passageiro Gr.Etario(10-14) SUM	Idade do passageiro situa-se no intervalo 10 a 14 anos	Categórica
Passageiro Gr.Etario(15-17) SUM	Idade do passageiro situa-se no intervalo 15 a 17 anos	Categórica
Passageiro Gr.Etario(18-20) SUM	Idade do passageiro situa-se no intervalo 1 a 20 anos	Categórica
Passageiro Gr.Etario(21-24) SUM	Idade do passageiro situa-se no intervalo 21 a 24 anos	Categórica
Passageiro Gr.Etario(25-29) SUM	Idade do passageiro situa-se no intervalo 25 a 29 anos	Categórica
Passageiro Gr.Etario(30-34) SUM	Idade do passageiro situa-se no intervalo 30 a 34 anos	Categórica
Passageiro Gr.Etario(35-39) SUM	Idade do passageiro situa-se no intervalo 35 a 39 anos	Categórica
Passageiro Gr.Etario(40-44) SUM	Idade do passageiro situa-se no intervalo 40 a 44 anos	Categórica
Passageiro Gr.Etario(45-49) SUM	Idade do passageiro situa-se no intervalo 45 a 49 anos	Categórica
Passageiro Gr.Etario(50-54) SUM	Idade do passageiro situa-se no intervalo 50 a 54 anos	Categórica
Passageiro Gr.Etario(55-59) SUM	Idade do passageiro situa-se no intervalo 55 a 59 anos	Categórica
Passageiro Gr.Etario(60-64) SUM	Idade do passageiro situa-se no intervalo 60 a 64 anos	Categórica
Passageiro Gr.Etario(65-69) SUM	Idade do passageiro situa-se no intervalo 65 a 69 anos	Categórica
Passageiro Gr.Etario(70-74) SUM	Idade do passageiro situa-se no intervalo 70 a 74 anos	Categórica
Passageiro Gr.Etario(>=75) SUM	Passageiro tem 75 anos ou mais	Categórica
Passageiro Gr.Etario(Não Def.)SUM	Idade do passageiro desconhecida	Categórica

Tabela 10 - Dados dos peões

Variável	Descrição	Tipo
Id. Acidente	Id. do acidente	Numérica
Freguesia	Freguesia	Categórica
Tipo Natureza	Tipo de acidente	Categórica
Natureza	Tipo de acidente mais detalhado	Categórica
Datahora	Data e hora	Numérica
Entidades Fiscalizadoras	Entidades fiscalizadoras	Categórica
Id. Peao	Id. do peão	Numérica
Sexo	Género do peão	Categórica
Acções Peão	Ação do peão no momento do acidente	Categórica
Distrito	Distrito	Categórica
Concelho	Concelho	Categórica
Lesões a 30 dias	Lesões do peão a 30 dias	Categórica
Peao Gr.Etario(<=5) SUM	Peão tem menos de 5 anos de idade	Categórica
Peao Gr.Etario(6-9) SUM	Idade do peão situa-se no intervalo 6 a 9 anos	Categórica
Peao Gr.Etario(10-14) SUM	Idade do peão situa-se no intervalo 10 a 14 anos	Categórica
Peao Gr.Etario(15-17) SUM	Idade do peão situa-se no intervalo 15 a 17 anos	Categórica
Peao Gr.Etario(18-20) SUM	Idade do peão situa-se no intervalo 18 a 20 anos	Categórica
Peao Gr.Etario(21-24) SUM	Idade do peão situa-se no intervalo 21 a 24 anos	Categórica
Peao Gr.Etario(25-29) SUM	Idade do peão situa-se no intervalo 25 a 29 anos	Categórica
Peao Gr.Etario(30-34) SUM	Idade do peão situa-se no intervalo 30 a 34 anos	Categórica
Peao Gr.Etario(35-39) SUM	Idade do peão situa-se no intervalo 35 a 39 anos	Categórica
Peao Gr.Etario(40-44) SUM	Idade do peão situa-se no intervalo 40 a 44 anos	Categórica
Peao Gr.Etario(45-49) SUM	Idade do peão situa-se no intervalo 45 a 49 anos	Categórica
Peao Gr.Etario(50-54) SUM	Idade do peão situa-se no intervalo 50 a 54 anos	Categórica
Peao Gr.Etario(55-59) SUM	Idade do peão situa-se no intervalo 55 a 59 anos	Categórica
Peao Gr.Etario(60-64) SUM	Idade do peão situa-se no intervalo 60 a 64 anos	Categórica
Peao Gr.Etario(65-69) SUM	Idade do peão situa-se no intervalo 65 a 69 anos	Categórica
Peao Gr.Etario(70-74) SUM	Idade do peão situa-se no intervalo 70 a 74 anos	Categórica
Peao Gr.Etario(>=75) SUM	Peão tem 75 anos ou mais	Categórica
Peao Gr.Etario(Não Def.) SUM	Idade do peão desconhecida	Categórica