INSTITUTO UNIVERSITÁRIO DE LISBOA

Licenciatura em Ciência de Dados

Projeto Aplicado em Ciência de Dados II

Base de Dados: Acidentes Rodoviários

Afonso Gião Santana Oliveira e Silva Nº 105208 | Turma: CDC2

Ana Reis Abreu Nº 98625 | Turma: CDCPL1

Francisco Ferreira Camilo Nº 99234 | Turma: CDPL1

Rui Chaves Nº 104914 | Turma: CDC2

Tomás Francisco Ribeiro Nº 105220 | Turma: CDC2

ISCTE – IUL | dezembro de 2023

Fernando Batista

Anabela Costa

Índice

1. Introdução

1.1. Motivação

O trânsito nas estradas representa um aspeto crucial do dia a dia de muitos indivíduos, atravessando diferentes partes de todo o território nacional. Incluídos neste contexto estão elementos críticos a considerar, como os acidentes que frequentemente ocorrem nas nossas vias.

Os acidentes de trânsito afetam significativamente a vida dos cidadãos, incluindo desde os condutores, passageiros a pedestres. Por isso, é vital promover a consciencialização da população sobre como adotar comportamentos responsáveis, tanto por parte da comunidade em geral quanto das respetivas autoridades, para reduzir a frequência e o impacto desses eventos. Isto envolve analisar os diversos tipos de acidentes ocorridos no país e suas causas principais, visando identificar padrões e implementar estratégias eficazes para sensibilizar a comunidade.

Posto isto, para avançar com este projeto, optou-se por utilizar como referência os dados disponibilizados pela Autoridade Nacional de Segurança Rodoviária (ANSR). Este é um dataset em que os dados incluem informações detalhadas sobre acidentes de trânsito ocorridos em várias áreas de Portugal Continental, cobrindo o período dos anos de 2010 a 2019.

1.2. Objetivo de Estudo

Neste projeto foi usada a metodologia *Cross-industry Standard Process for Data Mining*, ou CRISP DM1. Como o nome sugere, esta metodologia é normalmente usada na extração de dados, por isso será a base do projeto e do relatório. Foram realizadas análises, com suporte na base de dados, na tentativa de chegar a conclusões importantes para definir os nossos objetivos de estudo.

Assim, após esta análise e discussão de ideias, o grupo decidiu incidir o tema do projeto em algumas questões a serem respondidas ao longo deste relatório. A primeira questão levantada foi “que impacto têm as épocas festivas nos acidentes rodoviários?”. Em que queremos perceber em datas como o Natal e o Carnaval, o que acontece nas estradas portuguesas, sendo que por volta destas datas existe sempre um maior tráfego

rodoviário. Por sua vez, a segunda questão levantada como objeto de estudo focou-se em conseguirmos obter algo que conseguisse caracterizar um acidente como grave, isto através de um modelo de classificação, de forma a conseguirmos perceber quais as variáveis que poderiam influenciar mais nesta classificação.

O projeto é então baseado no tratamento e processamento de dados e a sua análise exploratória, trabalhados no RStudio com linguagem *R*, com a finalidade de obter respostas para as questões debatidas e propostas a estudo pelo grupo. As questões formuladas são:

- Quais as épocas do ano mais tendenciosas para acidentes?
- Que impacto têm as épocas festivas nos acidentes rodoviários?
- O que caracteriza um acidente como grave? (modelo de classificação)

2. Quais as épocas do ano mais tendenciosas para acidentes?

2.1. Metodologia

A nossa base de dados tem por fonte a Autoridade Nacional de Segurança Rodoviária com acidentes de vários anos, desde 2010 até ao ano de 2019. Iniciámos com uma base de dados com 309874 registos. Existiam originalmente 43 variáveis, algumas das principais como por exemplo data e hora do acidente, a sua localização, número de feridos, condições climáticas, estação do ano, entre outras. Verificou-se que existiam vários datasets repartidos em anos, foram então unidos todos os anos e lidos em um único dataframe 'acidentes'. Feita esta junção o dataset passou a ter 41 variáveis, em que foram eliminadas duas novas variáveis 'latitude_gps' e a 'longitude_gps' que não foram consideradas pertinentes para o estudo dos acidentes graves.

- **Tabela "Acidentes"**

Nome da Variável	Descrição	Tipo de Variável
Id.Acidente	Identificação do acidente	Quantitativa contínua
Cond_Aderencia	Condição de aderência da estrada	Qualitativa nominal
Tipos_Vias	Tipo de via do acidente	Qualitativa nominal
Fatores_Atmosfericos	Fatores atmosféricos que ocorrem durante o acidente	Qualitativa nominal
DataHora	Data e Hora quando ocorreu o acidente	Quantitativa contínua
Natureza	Tipo de acidente	Qualitativa nominal
Dia	Dia do acidente	Quantitativa discreta
Mês	Mês do acidente	Quantitativa discreta
Hora	Hora do acidente	Quantitativa contínua
Entidades Fiscalizadoras	Entidade que socorreu o acidente	Qualitativa nominal
Velocidade Local	Velocidade máxima permitida	Quantitativa contínua
Velocidade Geral	Velocidade máxima permitida	Quantitativa contínua
Dia da semana	Dia da semana do acidente	Qualitativa nominal

Num. Mortos a 30 dias	Número de mortos aquando 30 dias do acidente	Quantitativa discreta
Num. Feridos graves a 30 dias	Número de feridos graves aquando 30 dias do acidente	Quantitativa discreta
Num. Feridos ligeiros a 30 dias	Número de feridos ligeiros aquando 30 dias do acidente	Quantitativa discreta
Características Tecnicas1	Tipo de estrada	Qualitativa nominal
Distrito	Distrito onde ocorreu o acidente	Qualitativa nominal
Concelho	Concelho onde ocorreu o acidente	Qualitativa nominal
Freguesia	Freguesia onde ocorreu o acidente	Qualitativa nominal
Pov. Proxima	Povoação mais próxima	Qualitativa nominal
Nome Arruamento	Nome da rua onde ocorreu o acidente	Qualitativa nominal
Cod Via	Código da via onde ocorreu o acidente	Qualitativa nominal
Estado Conservação	Estado da estrada	Qualitativa nominal
KM	KM da estrada	Quantitativa discreta
Reg Circulação1	Quantidade de sentidos da estrada	Quantitativa discreta
Intersecção Vias	Tipo de interseção onde ocorreu o acidente	Qualitativa nominal
Localizações	Tipo de localização do acidente	Qualitativa nominal
Luminosidade	Altura do dia e se havia iluminação	Qualitativa nominal
Marca Via	Existência de marcas na via	Qualitativa nominal
Obstáculos	Existência de obstáculos na estrada quando houve o acidente	Qualitativa nominal
Sentidos	Sentido em que o carro circulava	Qualitativa nominal
Sinais	Sinais existentes	Qualitativa nominal
Sinais Luminosos	Existência de sinais luminosos	Qualitativa nominal
Tipo Piso	Estado e tipo do piso	Qualitativa nominal
Traçado 1	Tipo de traçado 1 da via	Qualitativa nominal
Traçado 2	Tipo de traçado 2 da via	Qualitativa nominal
Traçado 3	Tipo de traçado 3 da via	Qualitativa nominal
Traçado 4	Tipo de traçado 4 da via	Qualitativa nominal
Via Trânsito	Via de trânsito em que o veículo circulava	Qualitativa nominal

O DataSet continha 12767321 NA´s que correspondiam a 9% do dataset inicial, estes foram tratados com base no manual de preenchimento da beav. Foi assim realizada uma análise detalhada dos valores ausentes em várias colunas. Foram removidas linhas com valores NA em colunas específicas como: **velocidade_local, velocidade_geral, características_tecnicas1, marca_via, luminosidade,** entre outras que não consideramos relevantes tratar de outra forma os valores omissos sem ser eliminando os mesmos que pouco influenciavam na análise do nosso objetivo ou eram colunas com um alto teor de NA´s. Substituímos os valores NA em outras variáveis com valores predeterminados, refletindo sempre a interpretação do manual de preenchimento, estas colunas foram: **pov_proxima, nome_arruamento, km, factores_atmosfericos, obras_arte, sinais, sinais_luminosos, sentidos**. Em maior parte destas variáveis foram substituídos os valores nulos por "Outro" para permitir a inclusão do registo na análise evitando distorções na mesma e para não perder o restante da informação.

Para tratamento de outliers foram utilizados boxplots para as variáveis **'num_feridos_ligeiros_a_30_dias','num_feridos_graves_a_30_dias'** e **'num_mortos_a_30_dias'**. Usamos ainda *bloxplot.stats* para termos uma análise estatística mais detalhada do boxplot, incluindo assim a identificação dos valores de outliers que foram então removidos. Ao analisarmos variáveis como estas conseguimos analisar a relação entre a gravidade dos acidentes e fatores como a luminosidade, estado da via e perceber então o porquê de inúmeros acidentes.

Criámos gráficos de linhas para observar as tendências nos nossos dados neste caso da ocorrência de acidentes de 2010 a 2019. Bem como na distribuição de acidentes ao longo de um ano, e um mês específicos, e até em épocas festivas importantes como Natal e Carnaval, revelando assim tendências sazonais ou relacionadas a eventos.

2.1.1. Tratamento de variáveis

Foi feita uma padronização dos nomes das colunas onde utilizamos a função **make_clean_names** durante a leitura dos dados de cada ano (de 2010 a 2019) com a função **read_excel**. Esta função transforma os nomes das colunas para um formato consistente e fácil de usar no R. Converte todos os caracteres para minúsculas, substitui

espaços e caracteres especiais por *underscores* (_) ainda garante que os nomes das colunas são únicos e legíveis.

De seguida convertemos algumas variáveis quanto ao seu tipo de dados:

- Data e Hora:

A coluna datahora é convertida para o formato de data e hora com a função **as_datetime**. Esta conversão é crucial para análises temporais, permitindo assim manipularmos e analisarmos as datas e os horários dos acidentes de uma forma mais eficaz.

- Conversão de Variáveis *Character* para *Factor*:

Foram convertidas também as variáveis do tipo *character* no dataframe acidentes para o tipo factor usando a função **mutate_if** do pacote *dplyr* juntamente com **is.character** e **as.factor**. A conversão para factor é importante para variáveis categóricas, pois assim otimizamos o armazenamento de dados e facilitamos a realização de certas análises estatísticas e visualizações.

2.2. Análise Exploratória

Como ponto de partida, e de forma a observarmos melhor de forma global o número de acidentes, fizemos uma análise geral através de um gráfico de linhas. O gráfico resultante mostra a tendência de ocorrências de acidentes rodoviários ao longo do tempo, de 2010 a 2019.

As variações na linha mostram como o número de acidentes mudou ao longo dos anos. O que nos é útil para identificar padrões temporais, como aumentos ou diminuições na frequência de acidentes ao longo do período estudado. Obtivemos o seguinte resultado: (Gráfico 1)

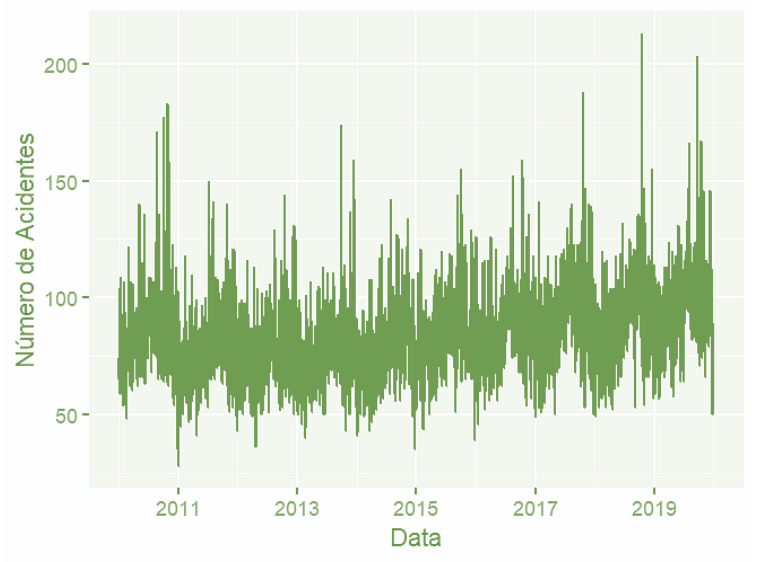


Gráfico 1 - Ocorrência de Acidentes 2010-2019

Verificámos que o padrão de tendência de acidentes de ano para ano é bastante similar, desde o ano 2011 que o número de acidentes tem vindo a aumentar de ano para ano. Para este aumento progressivo de ano para ano vemos como causa o aumento crescente de cidadãos com carta de condução, posto isto existem mais carros, com mais carros teremos mais circulação rodoviária o que consequentemente irá causar mais probabilidade de ocorrerem acidentes diariamente.

Posteriormente, quisemos analisar padrão geral do número de acidentes:

- Diariamente, ao longo dos anos (Gráfico 2)

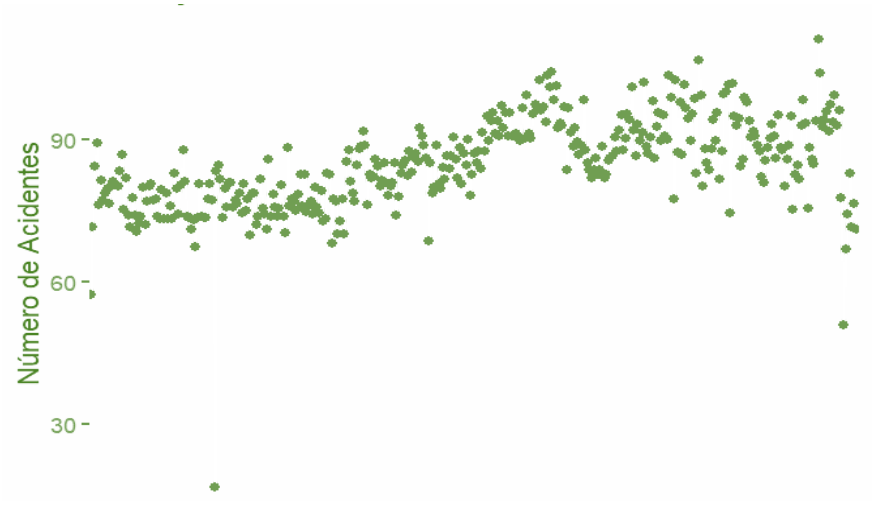


Gráfico 2 - Evolução do Nº de Acidentes ao longo dos anos

Neste gráfico conseguimos então obter a média do número de acidentes para cada dia do ano, e por isto existem 366 pontos no mesmo, que correspondem a cada dia do ano. Assim sendo, o primeiro ponto corresponde ao somatório/10 de acidentes ocorridos a 1 de janeiro na década de 2010. Podemos ver que o ponto correspondente ao dia 29 de fevereiro é o ponto mais baixo no gráfico pois para a média deste dia só existem 2 valores, pois na década de 2010 só ocorreu 2 vezes, assim tem uma média bastante baixa comparativamente aos restantes dias do ano.

Em análise geral ao gráfico conseguimos perceber que o primeiro semestre do ano tem uma média de acidentes diários mais baixa e que no segundo semestre existe um aumento da média diária do número de acidentes. Percebendo que no primeiro semestre são poucos os dias que estão acima dos 90 acidentes, e no segundo semestre já são praticamente equilibrados os dias acima dos 90 e os abaixo dos 90 acidentes. Com isto podemos perceber que existe um aumento do número de acidentes nos últimos meses do ano, e por sua vez a época com menos acidentes rodoviários foi nos primeiros meses do ano.

- Diariamente, ao longo de um mês (Gráfico 3)

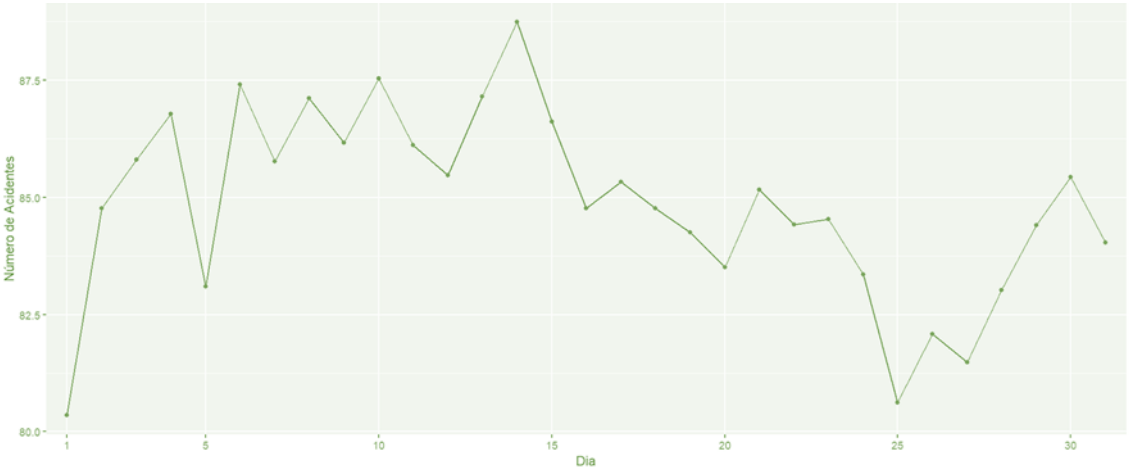


Gráfico 3 - Acidentes num mês padrão 2010 - 2019

Este gráfico, gerado a partir de um *ggplot,* fornece uma análise visual da evolução média diária de acidentes ao longo de um mês, com dados agregados de 2010 a 2019. Cada ponto no gráfico representa a média de acidentes para um dia específico,

calculada a partir dos dados de todos os meses correspondentes ao longo dos dez anos. Por exemplo**, o valor para o dia 1 é a média de todos os acidentes ocorridos nos dias 1 de cada mês entre 2010 e 2019**. A linha contínua conecta todos dias, o que nos mostra a tendência ao longo do mês, enquanto os pontos destacam os valores médios diários.

O gráfico permite-nos identificar padrões e variações na frequência de acidentes. Podemos concluir através do gráfico que criando assim um "mês padrão", que grande parte dos acidentes tendem a ocorrer na primeira quinzena do mês. Na segunda quinzena do mês temos então uma descida ligeira no número de acidentes diários. O gráfico permite-nos identificar rapidamente tendências, picos e padrões diários, facilitando assim a compreensão de como a frequência de acidentes varia ao longo do tempo. Tendo o dia com mais acidentes da década de 2010 o dia 13, e o dia com menos acidentes o dia 25, neste mês criado como mês padrão.

- Mensalmente, durante um ano

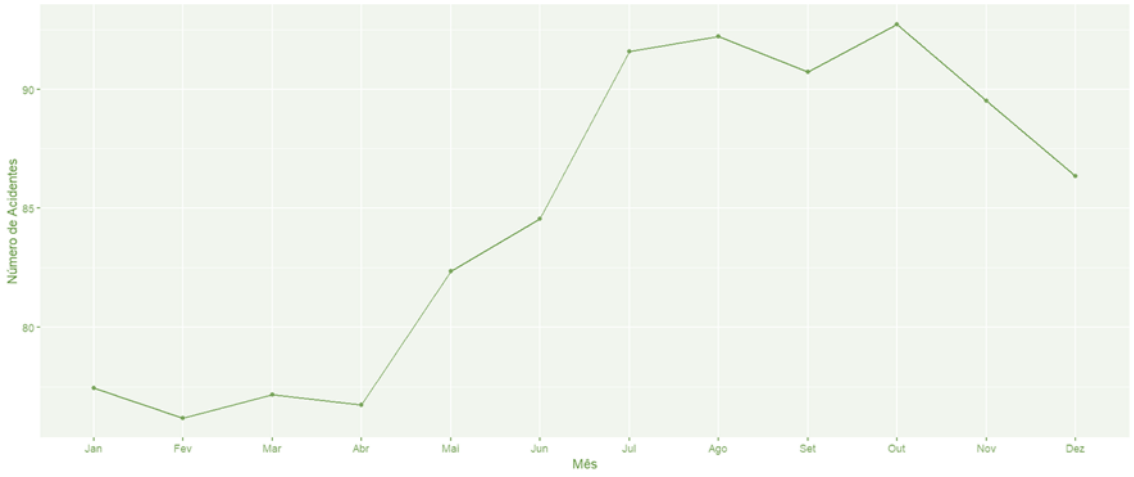


Gráfico 4 - Tendência de acidentes mensais 2010 - 2019

No gráfico 4 analisamos a média mensal de acidentes rodoviários ao longo de um ano, com base em dados recolhidos entre os anos disponíveis (2010-2019). A metodologia aplicada para este gráfico inclui a agrupação dos dados por respectivo mês, bem como o ajuste da média de acidentes para ser calculada conforme a duração variável dos dias dos meses. Este detalhe deve-se a nem todos os meses ter o mesmo número de dias, garantindo uma comparação precisa.

Observa-se neste gráfico uma tendência de maior número de acidentes nos meses de verão em comparação com os meses de inverno. Esta tendência pode ser atribuída a diversos fatores. Primeiramente, os meses de verão geralmente registam um aumento no tráfego rodoviário devido às férias escolares e viagens de lazer, levando a uma maior probabilidade de acidentes. Além disso, o comportamento dos condutores pode ser diferente durante o verão, com uma tendência a viagens mais longas e, em alguns casos, a condução mais imprudente, como o excesso de velocidade.

Outro aspeto a considerar é a influência das condições meteorológicas. Embora o inverno apresente condições de condução mais perigosas devido às chuvas, gelo e à neve em certas regiões do país, os meses de verão frequentemente têm melhores condições de visibilidade e estradas mais secas, o que pode incentivar velocidades mais altas por parte do condutor e, consequentemente, mais acidentes.

Este gráfico é particularmente relevante, pois não só apresenta informações de forma clara e concisa, mas também nos serviu como ponto de partida para análises mais profundas. Destaca-se assim a importância de considerar variáveis sazonais e comportamentais na análise de acidentes rodoviários. O que nos despertou a ideia de analisar com mais detalhe certas épocas do ano, que é a nossa questão a analisar **"Que impacto têm as épocas festivas nos acidentes?"**. (3)

2.3. Conclusões

A análise exploratória realizada forneceu insights fundamentais para decidir e abordar as duas questões centrais do nosso estudo: o impacto das épocas festivas nos acidentes rodoviários e os fatores que contribuem para a gravidade de um acidente. Os gráficos e análises serviram como base para uma compreensão mais aprofundada desses aspetos.

- **Variações Sazonais**
  - Aumento nos acidentes durante os meses de verão.
  - Épocas festivas no inverno, como Natal e Ano Novo.

- **Fatores que contribuem para a gravidade dos acidentes**
  - Condições meteorológicas

- Comportamento dos condutores

2.4. Metas de estudo

A partir da nossa análise exploratória, tornou-se evidente que diferentes épocas do ano, especialmente as festivas, apresentam tendências distintas em termos de acidentes rodoviários. Por exemplo, observamos um aumento no número de acidentes durante o verão, provavelmente devido ao aumento do tráfego e a mudanças nos comportamentos de condução durante as férias. Além disso, notamos picos no início de cada mês, que podem estar relacionados a alterações nas rotinas ou a eventos específicos. A gravidade dos acidentes também mostrou variação, muitas vezes influenciada por fatores como as condições climáticas adversas, o comportamento de risco dos condutores e as condições das estradas.

Existem diversas medidas que podem ser postas em prática pelas entidades responsáveis pela segurança rodoviária, bem como por todos nós cidadãos. Primeiramente, a implementação de aulas adaptativas nas escolas de condução é crucial. Estas aulas devem incluir simulações de diferentes cenários de trânsito e condições climáticas, focando especialmente em períodos de alto risco como as épocas festivas. Isso prepararia melhor os futuros condutores para situações reais, reduzindo potencialmente a frequência e a gravidade dos acidentes nestas épocas. Além disso, campanhas de sensibilização sazonais se fazem necessárias. Estas deveriam ser focadas em épocas específicas do ano, destacando os riscos aumentados e promovendo práticas de condução segura.

3. Que impacto têm as épocas festivas nos acidentes rodoviários?

3.1. Metodologia do objetivo de estudo

Tendo identificado a importância das épocas festivas nos padrões de acidentes rodoviários, direcionamos nosso estudo para investigar detalhadamente o impacto desses períodos. Esta análise visa compreender como as festividades específicas, tais como Natal, Ano Novo e Carnaval, influenciam a ocorrência e a natureza dos acidentes de trânsito. O objetivo é identificar se há um aumento significativo no número de acidentes ou na gravidade destes durante esses períodos festivos e entender as causas subjacentes a essas tendências.

Para abordar esta questão, inicialmente procedemos à coleta de conjuntos de dados pertinentes, focando especificamente em datas que incluem informações detalhadas sobre acidentes rodoviários durante as épocas festivas. Os dados serão analisados com o intuito de caracterizar os acidentes que ocorrem nestas épocas, incluindo a atenção a variáveis como o comportamento dos condutores, as condições da estrada, as condições meteorológicas e outros fatores relevantes. Ao compreender o papel que as épocas festivas desempenham nos acidentes rodoviários, podemos desenvolver estratégias mais eficazes para aumentar a segurança nas estradas e reduzir a incidência de acidentes durante estes períodos críticos.

3.2. Análise Exploratória – Épocas Festivas

Inicialmente foram debatidas várias épocas festivas como - Natal, Ano Novo, Páscoa, Halloween e Carnaval - nos acidentes rodoviários em Portugal. Reconhecendo a diversidade e a singularidade de cada uma destas festividades, o estudo foi cuidadosamente desenhado para abordar as especificidades relacionadas a cada período. Depois de analisar algumas datas e dados correspondentes às mesmas decidimos focar o estudo em apenas duas épocas: **Natal e Carnaval**.

Devido ao facto de termos dados praticamente similares no Natal e Ano Novo, e a Páscoa e o Halloween praticamente não afetarem o número de acidentes em Portugal. Por percebermos que eram épocas festivas mais calmas e com pouco por analisar

decidimos focar apenas nas duas referidas anteriormente. Debatemos estes eventos comemorativos com base nos gráficos estudados anteriormente e pensamos analisar um ano por semanas para visualizar melhor as tendências das épocas que estavam em debate.

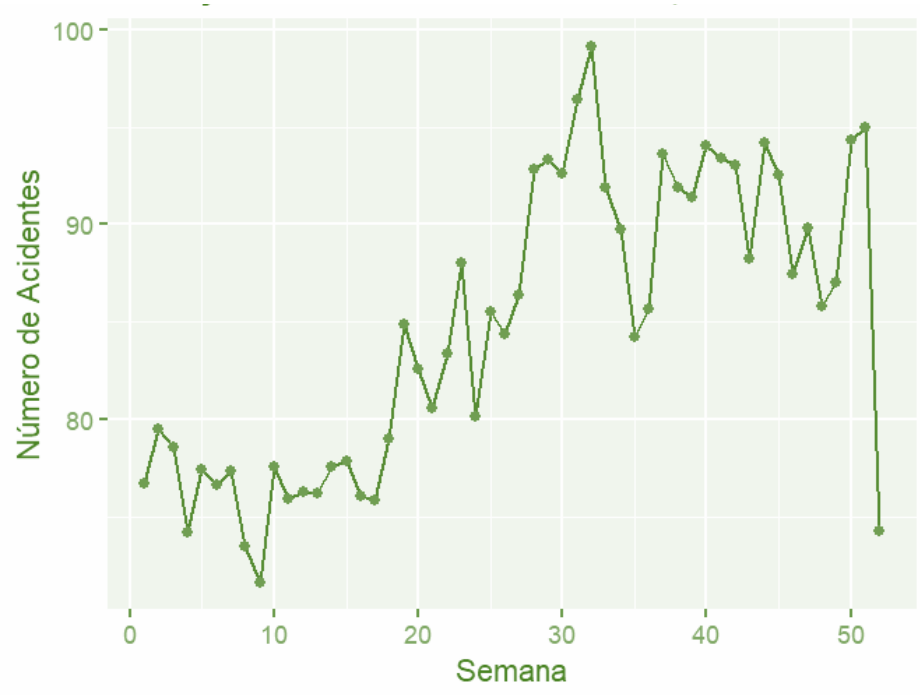


Gráfico 5 - Evolução do número de acidentes por semana ao longo de um ano

Selecionamos as épocas festivas para aprofundamento devido à sua relevância cultural e potencial influência no comportamento dos condutores e na segurança rodoviária. O Natal, com datas fixas (24/12 e 25/12), apresenta um desafio único, especialmente considerando o movimento intenso nas estradas devido às férias escolares e às viagens familiares.

Por outro lado, o Carnaval, com datas variáveis de ano para ano, exigiu uma abordagem mais flexível, especialmente nos locais onde sua celebração tem maior impacto, como em Torres Vedras e Sesimbra. Nestas áreas, o aumento do tráfego e as celebrações extensivas podem influenciar significativamente os padrões de acidentes nas mesmas.

Outro aspecto crucial do nosso estudo foi a consideração de que as causas dos acidentes podem variar de acordo com a época festiva. Por exemplo, os acidentes no período do Natal podem estar mais relacionados às deslocações familiares e às

condições climáticas do inverno, enquanto no Carnaval, a combinação de festividades e comportamento dos condutores em áreas específicas pode ter um impacto diferente. Assim, a análise foi orientada para entender essas causas variadas e seu impacto na segurança rodoviária, permitindo-nos desenvolver recomendações mais eficazes para a prevenção de acidentes nestes períodos festivos.

3.2.1. Natal

Durante o Natal, muitas pessoas viajam longas distâncias para reunir-se com familiares e vão mais para os centros comerciais para comprar antecipadamente os presentes, o que naturalmente conduz a um aumento do tráfego nas estradas. Esta maior densidade de veículos, combinada muitas vezes com condições climáticas adversas típicas do inverno, como chuva, neve ou gelo, pode aumentar significativamente o risco de acidentes. Além disso, fatores como o cansaço de longas viagens, a pressa para cumprir compromissos festivos e, em alguns casos, o consumo de álcool, podem influenciar o comportamento dos condutores e aumentar a probabilidade de ocorrência de acidentes. Neste contexto, a análise do período natalino e sua correlação com os acidentes rodoviários tornou-se fundamental. O objetivo foi compreender não apenas a frequência dos acidentes nesta época, mas também as suas causas, natureza e consequências, de forma a desenvolver estratégias eficazes para a prevenção de acidentes.

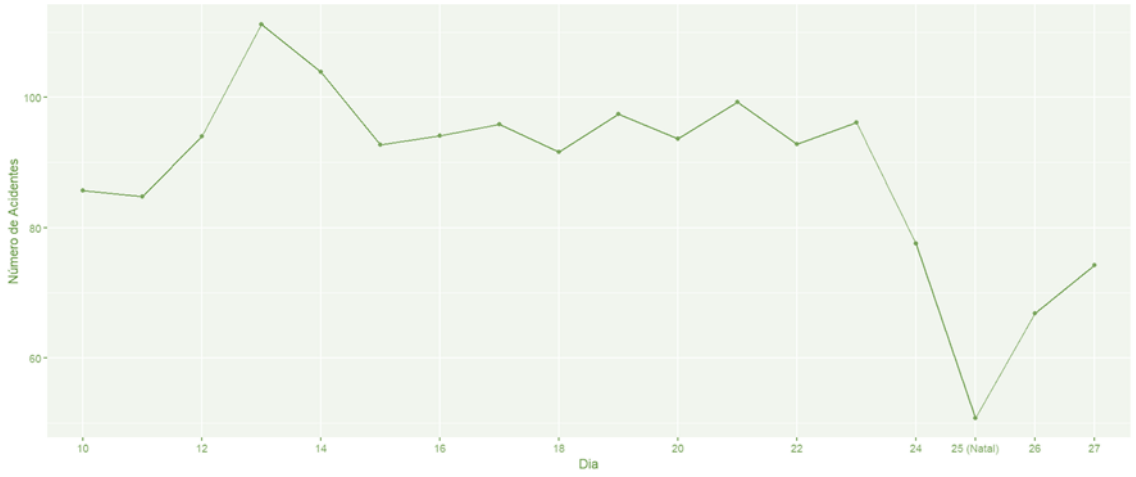


Gráfico 6 - Acidentes durante o período de natal

Foi realizada a análise detalhada do número de acidentes rodoviários durante o período natalício, escolhemos observar os dias de 10 a 27 de dezembro. Escolhemos este período de dias por corresponder desde o momento em que começam as férias escolares, e por isto começarem as viagens familiares nesta época. Primeiramente, identificamos que o dia com o maior número de acidentes durante este período foi o 13 de dezembro. Este pico, com mais de 100 acidentes num só dia, pode ser atribuído a vários fatores:

- **Aumento do tráfego rodoviário:** à medida que as pessoas começam as preparações para o Natal, não só como compras como deslocações para casa de familiares / viagens longas.
- **Condições climáticas adversas:** as condições climáticas típicas no inverno, como chuvas fortes, neve em diversas partes do país ou até mesmo gelo.

Além disso, observamos que a média de acidentes no dia de Natal foi de 50,8, um número relativamente baixo que pode estar relacionado às pessoas já estarem nas respectivas casas onde vão passar o Natal e à pouca movimentação nas estradas neste dia específico. Curiosamente, durante os 18 dias estudados do período natalício, a média diária de acidentes foi de 89, significativamente mais alta do que a média mensal de 70 acidentes observada em um mês considerado "normal".

Estas observações sugerem que a época do Natal é um **período particularmente crítico** em termos de segurança rodoviária, com um aumento notável na frequência de acidentes. Esta tendência pode ser causada por uma combinação de fatores:

- **Aumento do volume de tráfego**
- **Alterações nos padrões de comportamento dos condutores:** como o maior consumo de álcool, ou a pressa por parte dos condutores a chegar ao seu destino.
- **Condições meteorológicas desafiadoras**

A análise destes dados é crucial para entender melhor os padrões de acidentes rodoviários durante o Natal e pode servir como base para o desenvolvimento de

estratégias de prevenção e segurança rodoviária mais eficazes. Ao identificar os dias com maior incidência de acidentes e entender as causas subjacentes, as autoridades competentes e organizações responsáveis podem implementar medidas específicas para reduzir o risco de acidentes e garantir uma época festiva mais segura para todos.

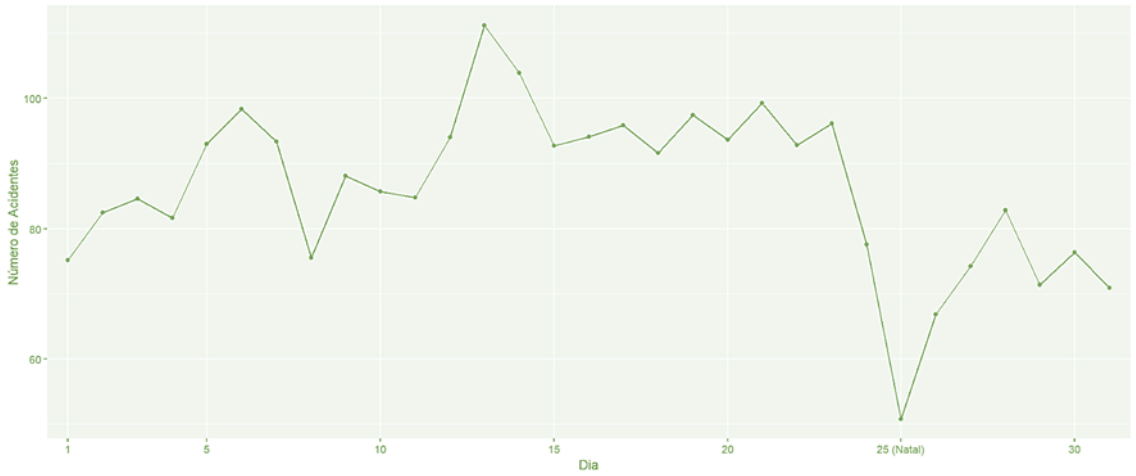


Gráfico 7 - Acidentes durante o mês de Dezembro

Na continuação do nosso estudo sobre a segurança rodoviária, realizou-se uma análise detalhada da evolução do número de acidentes ao longo do mês de dezembro. Através do gráfico, conseguimos obter informações importantes sobre os padrões de acidentes neste mês específico de Natal.

O gráfico criado, revelou que o dia com o maior número de acidentes foi como visto já anteriormente o 13 de dezembro. Este dado reforça a ideia de que a preparação para o Natal, com o aumento do fluxo de pessoas nas estradas e as condições climáticas adversas típicas do inverno, contribui significativamente para o aumento dos acidentes.

Outra observação relevante foi a confirmação que o período em torno do Natal regista uma maior incidência de acidentes. Um fato interessante mostrou que após o dia de Natal, existe uma diminuição acentuada no número de acidentes. Este declínio pode ser devido a uma redução no tráfego rodoviário, uma vez que muitas pessoas já teriam chegado aos seus destinos de Natal e Ano Novo e as viagens tendem assim a diminuir.

3.2.2 Conclusões Época de Natal

Para combater e diminuir os acidentes nesta época do ano, algumas medidas podem ser implementadas tanto pelas autoridades competentes como por nós cidadãos e também condutores.

- **Campanhas de Sensibilização:** Desenvolver e promover campanhas de conscientização sobre os riscos aumentados de conduzir durante o mês de dezembro, focando especialmente nos dias que antecedem o Natal.
- **Fiscalização e Controle de Velocidade:** Aumentar a fiscalização nas estradas, especialmente em pontos conhecidos por altas taxas de acidentes, e implementar controles de velocidade mais rigorosos. Para esse efeito existem, operações STOP e/ou uma maior aposta em alertas nas estradas.
- **Educação sobre Condução em Condições Adversas:** Reforçar a educação dos condutores sobre como dirigir de forma segura em condições climáticas adversas, como chuva, neve e gelo.
- **Planear e Preparar as Viagens:** Incentivar os condutores a planear as viagens familiares com antecedência, para evitar assim horários de pico e garantir que os veículos encontram-se em boas condições para viagens de longa distância.

Estas medidas, juntamente com uma conscientização geral sobre os riscos associados à condução durante a época natalícia, podem contribuir significativamente para a redução dos acidentes rodoviários, garantindo um período mais seguro para todos.

3.2.3 Carnaval

O Carnaval, uma festa popular amplamente celebrada em várias regiões de Portugal, esta época festiva é marcada por um aumento significativo nas atividades sociais e eventos públicos, com especial destaque para regiões como Torres Vedras, onde as celebrações do Carnaval são particularmente famosas e atraem grandes multidões. Em regiões como Torres Vedras, durante o Carnaval, observa-se um aumento do tráfego rodoviário, devido à chegada de visitantes de outras áreas e ao movimento local associado à festividade. Este aumento no volume da circulação rodoviária, juntamente com o ambiente festivo, pode levar a um comportamento de condução mais imprudente, incluindo o consumo de álcool, o que eleva o risco de acidentes.

Além disso, as alterações temporárias nas configurações do tráfego e as restrições de estacionamento comuns durante grandes festivais como o Carnaval, podem criar situações de condução mais desafiadoras, aumentando assim o potencial para acidentes rodoviários.

Portanto, ao analisarmos a segurança rodoviária durante o Carnaval, é crucial considerar esses fatores específicos, especialmente em regiões com celebrações extensivas. A compreensão detalhada de como a época do Carnaval influencia os padrões de acidentes rodoviários pode ser fundamental para desenvolver estratégias eficazes de prevenção e garantir a segurança de condutores e peões durante este período.

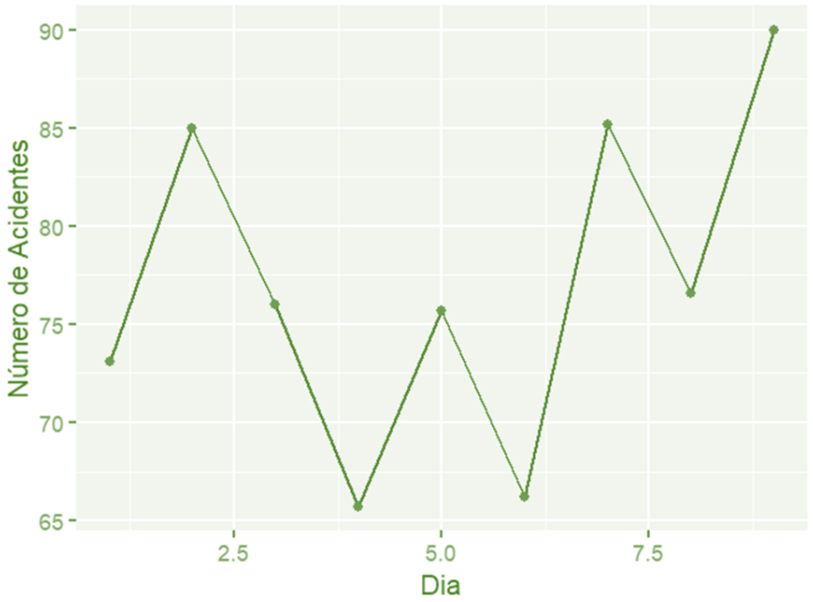


Gráfico 8 - Acidentes durante a época de Carnaval

Para o Carnaval focamos na análise do número de acidentes ao longo de um período de nove dias em torno desta festividade, abrangendo toda a década de 2010. O nosso estudo teve em especial atenção às variações das datas do Carnaval para cada ano, um fator crítico para uma análise precisa e correta. Foi então criado um gráfico que nos mostrasse a evolução dos acidentes nesta época, tendo assim como dia 1 o primeiro dia de Carnaval de cada ano respectivo. Tendo assim o gráfico criado conseguimos observar vários aspectos como:

- **Acidentes no Dia de Carnaval:** Observou-se que a média de acidentes no dia específico do Carnaval foi de 66,7. Este dado pode ser indicativo do aumento do movimento e das atividades festivas que tipicamente ocorrem neste dia.
- **Média de Acidentes Durante a Época do Carnaval:** A média de acidentes durante os nove dias da época de Carnaval foi de 77. O que nos confirma assim a tendência do aumento de acidentes nestes dias de Carnaval.

Depois de vistos estes padrões através do gráfico criado, foi pensado estudar o que acontece numa região típica de celebrações de Carnaval: **Torres Vedras**.

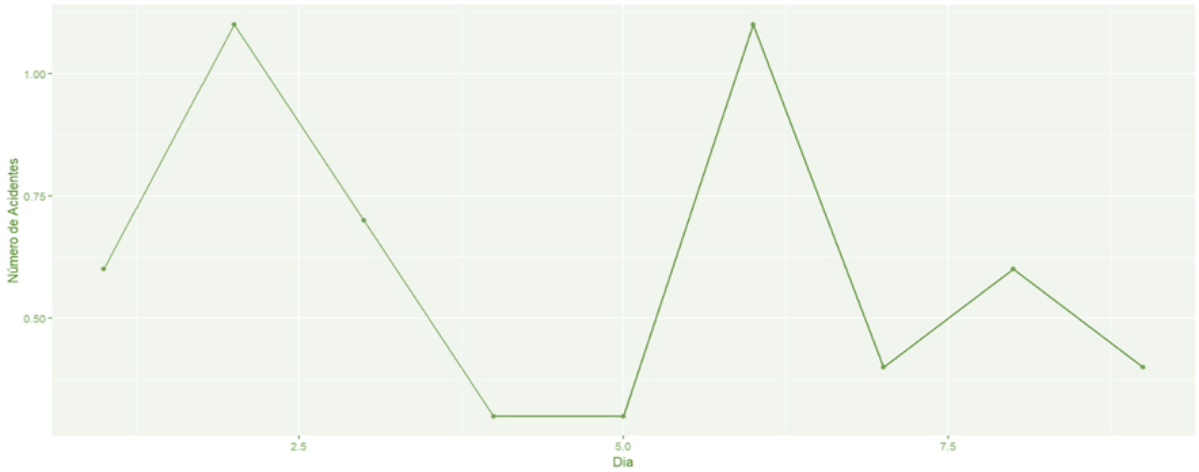


Gráfico 9 - Acidentes durante a época do Carnaval em Torres Vedras

- **Aumento de Acidentes em Torres Vedras:** Em regiões como Torres Vedras, onde o Carnaval é celebrado com grande entusiasmo, notou-se um aumento ligeiro no número de acidentes. Isso pode estar relacionado ao aumento de circulação nesta área, à atração de grandes multidões e possivelmente a comportamentos de condução mais imprudentes e inconscientes associados às festividades.

No âmbito do nosso estudo sobre a segurança rodoviária e o impacto do Carnaval nos acidentes, expandimos a análise para incluir o mês de fevereiro inteiro, com o objetivo de verificar se o aumento nos acidentes era específico dos dias de Carnaval ou se distribuía uniformemente ao longo do mês. Para isso, utilizamos dados de todos os dias de fevereiro, de 2010 a 2019.

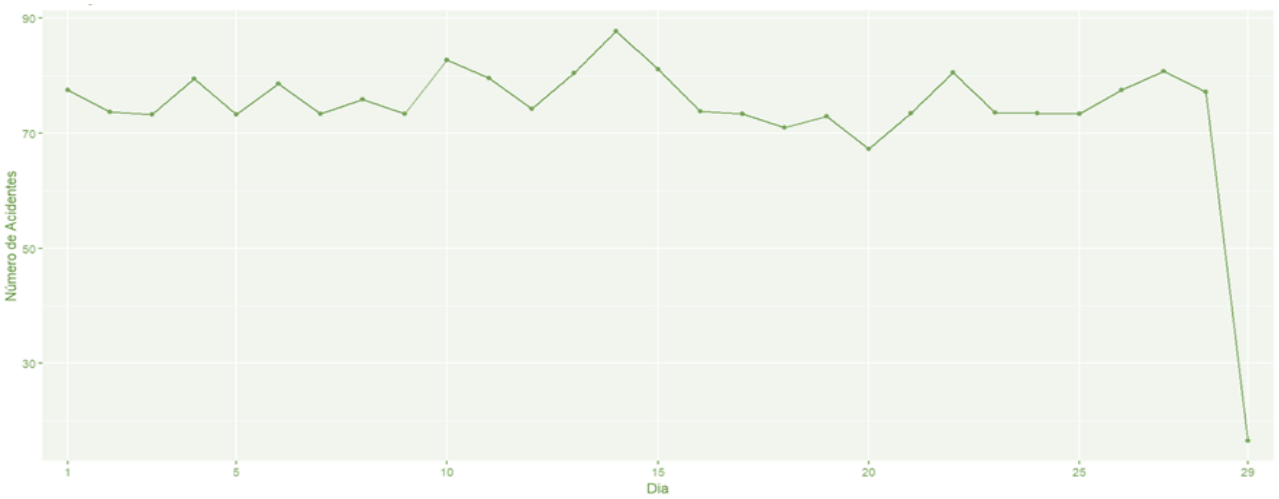


Gráfico 10 - Média Diária de Acidentes em Fevereiro

O gráfico que obtivemos foi construído com base nos dados de acidentes ao longo deste mês, ao longo de uma década. Através da análise do gráfico, observamos que os dias com mais acidentes dentro do mês de fevereiro coincidem com a época do Carnaval. Isto sugere que, de facto, há uma relação direta entre as celebrações do Carnaval e um aumento na incidência de acidentes rodoviários.

Além disso, constatamos que a média diária de acidentes durante a época do Carnaval é de aproximadamente 77 acidentes, em comparação com uma média de 71 acidentes nos outros dias do mês. Este aumento na média durante os dias de Carnaval reforça a ideia de que esta festividade tem um impacto significativo na segurança rodoviária. Estes dados são então importantes para compreender como eventos culturais e festividades podem influenciar o comportamento dos condutores e, consequentemente, a segurança nas estradas.

A identificação do Carnaval como um período de maior risco para acidentes rodoviários sugere a necessidade de medidas preventivas específicas durante esta época.

3.2.4. Conclusões Época Carnaval

Com base nestas observações, várias medidas podem ser recomendadas para melhorar a segurança rodoviária durante a época do Carnaval:

- **Reforço da Fiscalização de Trânsito:** Aumentar a presença policial e a fiscalização nas estradas, especialmente em áreas com grandes celebrações como Torres Vedras, Sesimbra ou Grande Lisboa para controlar o comportamento dos condutores e garantir a adesão às regras de trânsito.
- **Campanhas de Conscientização:** Promover campanhas educativas focadas na segurança rodoviária durante o Carnaval, alertando sobre os riscos de condução imprudente e consumo de álcool.
- **Sinalização adequada:** Garantir sinalização adequada em áreas de festividades.
- **Promover o uso do Transporte Público e Alternativas de Deslocação:** Incentivar o uso de transporte público ou serviços de transporte alternativos durante os dias de maior movimento, para reduzir o número de veículos nas estradas. Criação de shuffles nas zonas mais típicas de celebrações.

Estas estratégias, combinadas com uma compreensão aprofundada dos padrões de acidentes durante o Carnaval, são essenciais para desenvolver abordagens eficazes que garantam a segurança rodoviária durante este período festivo.

4. O que caracteriza um acidente como grave? (modelo de classificação)

4.1 Metodologia do objetivo de estudo

Aprofundando o nosso estudo sobre segurança rodoviária, decidimos ampliar o escopo da nossa análise, mudando o foco para um aspecto crítico: **a gravidade dos acidentes de trânsito**. Para isso, selecionamos variáveis específicas relacionadas ao número de feridos e fatalidades, incluindo o número de mortos a 30 dias, o número de feridos ligeiros a 30 dias, e o número de mortos no local. A escolha dessas variáveis é

estratégica, pois acreditamos que podem oferecer dados valiosos sobre a severidade dos acidentes.

Com base nesses dados, o nosso objetivo foi desenvolver um modelo de classificação capaz de prever a gravidade dos acidentes. Este modelo não apenas ajudará a entender melhor as variáveis que mais impactam a gravidade dos acidentes, mas também fornecerá uma ferramenta útil para antecipar a mesma gravidade.

A aplicação prática deste modelo pode ser extremamente benéfica para as equipas de saúde, INEM, ou autoridades que tenham de prestar socorro ao local do acidente. Com uma previsão mais precisa da gravidade dos acidentes, os profissionais podem aprimorar as estratégias de resposta e recorrer aos melhores recursos, torna assim o atendimento de socorro mais eficaz e adaptado à situação específica. Esta abordagem orientada pelos dados disponíveis representa um avanço significativo na gestão de resposta a emergências rodoviárias, com o potencial de salvar vidas e melhorar a eficiência dos serviços de saúde.

Portanto, este novo foco não apenas complementa a nossa análise anterior sobre a frequência dos acidentes, mas também traz uma contribuição vital para a compreensão e prevenção das consequências mais graves desses incidentes.

4.2. Análise Exploratória

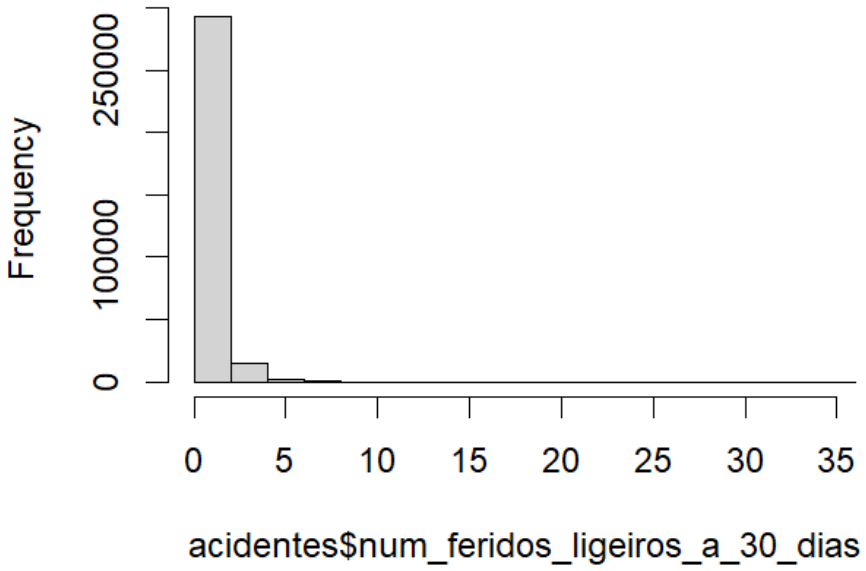


Gráfico 11 - Número de acidentes com feridos ligeiros

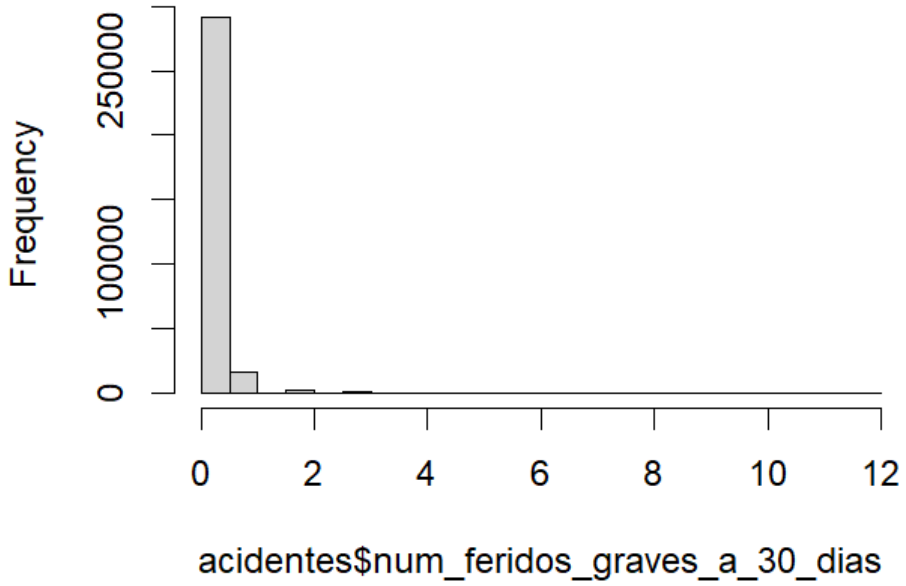


Gráfico 12 - Número de acidentes com feridos graves

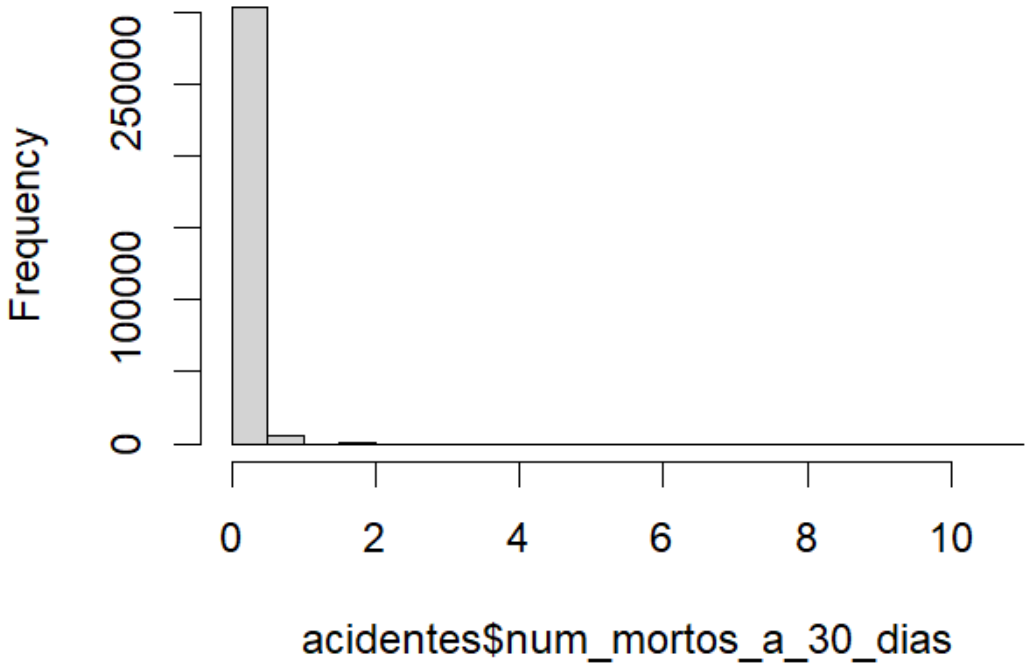


Gráfico 13 - Número de acidentes com mortos

Após uma análise preliminar sobre a extensão dos ferimentos em diferentes categorias de acidentes rodoviários, optámos por simplificar a nossa abordagem, criando uma variável única, dividida em apenas duas categorias: "Grave" e "Não grave". Esta variável será o alvo central do nosso estudo, orientando a modelação e a análise preditiva.

Na escolha das variáveis preditoras para os nossos modelos, decidimos adotar uma abordagem dupla. Primeiramente, utilizamos o V de Cramér como medida estatística para avaliar a correlação entre as variáveis disponíveis e a gravidade dos acidentes. Em segundo lugar, baseamo-nos na nossa compreensão intuitiva do que poderia influenciar a gravidade de um acidente, garantindo assim uma seleção de variáveis tanto empiricamente fundamentada quanto logicamente coerentes.

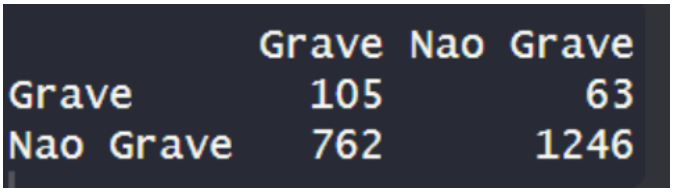
Dada a natureza desequilibrada dos nossos dados - com mais de 90% dos casos reportando acidentes não graves - foi necessário aplicar uma técnica de *undersampling*. Assim, dividimos o conjunto de dados numa proporção de 90% para treino e 10% para teste, uma distribuição que nos permite equilibrar os dados e melhorar a precisão e a relevância do nosso modelo preditivo.

Com os conjuntos de dados devidamente preparados para a modelação, podíamos avançar para a fase de previsão. Nesta etapa, aplicamos técnicas de modelação estatística para prever a gravidade dos acidentes rodoviários, com objetivo de aplicação prática na prevenção e resposta a acidentes rodoviários.

4.3. Modelo SVM

Iniciámos a nossa análise preditiva focando-nos inicialmente no algoritmo Support Vector Machine (SVM), uma técnica robusta e eficiente para a classificação de dados. Esta escolha baseou-se na capacidade do SVM de lidar com conjuntos de dados complexos e na sua eficiência em categorizar dados em grupos distintos, neste caso, em acidentes "Graves" e "Não graves".

Após a implementação do modelo com estas variáveis, procedemos à avaliação do seu desempenho através de uma matriz de confusão. Esta matriz é uma ferramenta crucial na análise de modelos de classificação, pois permite uma visualização clara da eficácia do modelo em prever corretamente os acidentes classificados como "Graves" e "Não graves". Através desta matriz, conseguimos quantificar tanto os acertos (verdadeiros positivos e verdadeiros negativos) quanto os erros (falsos positivos e falsos negativos) do nosso modelo, proporcionando uma compreensão detalhada da precisão e confiabilidade.



	Grave	Nao Grave
Grave	105	63
Nao Grave	762	1246

Imagem 1 - Matriz de Confusão

Nesta matriz de confusão, várias métricas podem ser extraídas, mas focamo-nos particularmente na sensibilidade, uma vez que é esta medida que avalia a eficácia do nosso modelo em identificar corretamente a classe em minoria, neste caso, os acidentes classificados como graves. A sensibilidade obtida foi de aproximadamente 63%, o que, considerando que o algoritmo foi aplicado a um conjunto de 2643 dados, pode ser visto como um resultado positivo. No entanto, é importante salientar que podem existir outros algoritmos com desempenhos superiores, os quais pretendemos explorar em seguida.

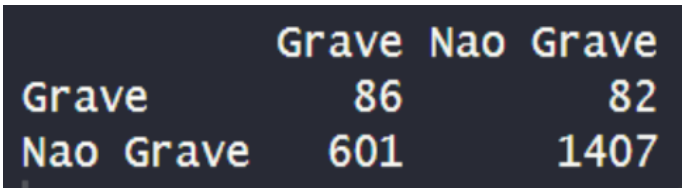
Importa também referir que o modelo alcançou uma precisão (accuracy) de 62%. Este valor indica que a proporção de previsões corretas (verdadeiros positivos e verdadeiros negativos) em relação ao total de previsões não apresenta uma discrepância significativa. Esta métrica é relevante, pois oferece uma visão geral da eficácia do modelo em classificar tanto os acidentes graves quanto os não graves.

Em suma, apesar dos bons resultados alcançados com o modelo SVM, a nossa investigação continuará no sentido de explorar e testar outros algoritmos de classificação, procurando aprimorar ainda mais a capacidade de previsão da gravidade dos acidentes rodoviários.

4.4. Modelo Random Forest

Para continuar o nosso estudo, testámos o algoritmo de Random Forest para a classificação dos acidentes rodoviários. Utilizámos 1000 árvores de decisão e selecionamos variáveis preditoras como "mês", "hora", "distrito", "localizações", "luminosidade" e "natureza", visando captar diferentes aspetos que influenciam a gravidade dos acidentes.

Através da matriz de confusão resultante, pudemos avaliar o desempenho do Random Forest, especialmente na sua capacidade de distinguir entre acidentes graves e não graves. Este passo é crucial para verificar se este modelo supera o desempenho do anteriormente testado SVM, proporcionando uma análise mais refinada e precisa da previsibilidade da gravidade dos acidentes rodoviários.



	Grave	Nao Grave
Grave	86	82
Nao Grave	601	1407

Imagem 2 - Random Forest

No que respeita à sensibilidade do modelo Random Forest, obtivemos um valor aproximado de 51%, o que pode ser considerado modesto, especialmente tendo em conta que o modelo foi treinado com um conjunto de dados onde a proporção entre as categorias da variável alvo é relativamente equilibrada. Este resultado indica que o modelo tem uma eficiência limitada na identificação correta de acidentes classificados como graves, que representam a categoria minoritária no nosso estudo.

Por outro lado, o algoritmo demonstrou uma precisão (accuracy) de cerca de 69%, superando o modelo SVM anterior. Esta melhoria na precisão sugere que o Random Forest é mais eficaz na previsão da categoria majoritária, ou seja, os acidentes Não Graves. Este aspeto é particularmente relevante, pois indica uma maior capacidade do modelo em identificar corretamente a maioria dos acidentes, o que pode ser importante em aplicações práticas, como a gestão de recursos de emergência e criar medidas de segurança rodoviária.

Contudo, considerando o contexto do nosso problema, que quer prever a gravidade dos acidentes de forma eficiente, a sensibilidade mais baixa do modelo Random Forest pode limitar a sua utilidade prática. Idealmente, procuramos um modelo que equilibre uma boa precisão com uma sensibilidade elevada, garantindo assim uma previsão confiável tanto para acidentes graves como para os não graves. Os resultados indicam a necessidade de continuar a explorar e ajustar modelos de classificação para alcançar um equilíbrio ideal que atenda às necessidades específicas do nosso estudo.

4.5. Gradient Boosting

Decidimos passar para o algoritmo de Gradient Boosting, uma técnica que, embora fosse menos familiar para nós inicialmente, apresentava potencial para a análise em questão. Optámos por configurar o modelo com 500 árvores de decisão, esperando que esta abordagem oferecesse um equilíbrio entre precisão e eficiência computacional. As variáveis preditoras selecionadas para o algoritmo incluíam "mês", "hora", "distrito", "localizações", "luminosidade", "natureza" e uma adição que consideramos muito importante: "velocidade_geral". A inclusão da variável "velocidade_geral" foi estratégica, dado que a velocidade tem um papel significativo na gravidade dos acidentes. Com a implementação do Gradient Boosting, esperávamos que o resultado fosse obter uma visão mais abrangente e talvez mais precisa sobre a classificação dos acidentes.

	Grave	Nao Grave
Grave	103	65
Nao Grave	727	1281

Imagem 3 - Gradient Boosting

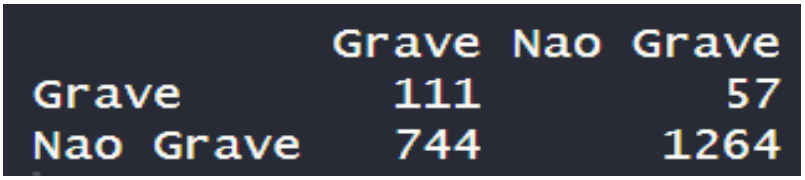
A sensibilidade do modelo situou-se em torno dos 61%, o que, apesar de representar uma melhoria em comparação com o modelo SVM anterior, ainda não atingiu um nível que consideramos satisfatório. Por esta razão, decidimos que a nossa pesquisa ainda não estava concluída neste ponto.

Interessante notar que, neste algoritmo, as variáveis que se revelaram mais influentes foram "natureza" e "distrito". Curiosamente, a "luminosidade" mostrou-se uma das variáveis menos relevantes, o que nos leva a ponderar sobre a complexidade dos fatores que afetam a gravidade dos acidentes rodoviários.

No que toca à precisão (accuracy) do modelo, esta situou-se em aproximadamente 64%. Este valor, embora razoável, não é suficientemente elevado para compensar a sensibilidade. Esta conclusão reforça a nossa determinação em continuar a explorar e a aperfeiçoar outros modelos, na busca de um equilíbrio mais efetivo entre sensibilidade e precisão, que nos permita prever com maior acuidade a gravidade dos acidentes rodoviários.

4.6. Regressão Logística

Na etapa final da nossa análise, optámos por uma abordagem mais simplificada, recorrendo ao algoritmo de regressão logística. Para este modelo, as variáveis preditoras incluídas foram "mês", "hora", "velocidade geral", "distrito", "localização", "luminosidade" e "natureza". A matriz de confusão gerada para este modelo foi a seguinte:



	Grave	Nao Grave
Grave	111	57
Nao Grave	744	1264

Imagem 4 - Regressão Logística

Com este algoritmo, alcançamos os melhores resultados dentro desta análise, nomeadamente uma sensibilidade de 66%, um valor que nos deixou satisfeitos e que marcou este modelo como o mais adequado para prever o nosso objetivo de estudo. Este destaque deve-se, sobretudo, à sua elevada sensibilidade, ou seja, à capacidade do modelo em identificar corretamente os casos de acidentes graves.

Na prática, isto significa que, ao apresentarmos um caso grave ao modelo, existe uma maior probabilidade de este ser corretamente classificado, o que é essencial para a finalidade da nossa questão. Esta eficácia na deteção de casos graves é crucial, pois permite uma resposta mais adequada e rápida em situações de emergência, contribuindo assim para a melhoria da segurança rodoviária e para a eficácia dos serviços de emergência.

5. Conclusão

5.1. Pergunta 1 - Épocas Festivas

É de se notar que as épocas festivas têm um impacto sobre os acidentes rodoviários. Pelo facto que as pessoas gostam de celebrar eventos anuais, é normal que sempre que estes aconteçam o tráfego aumente e os acidentes respetivamente. Por exemplo, na época natalícia o número de acidentes costuma aumentar cerca de 2 semanas antes do natal, o que faz sentido porque é nessa altura que as pessoas começam as compras de presentes, festas familiares e mesmo as condições climáticas do inverno. E no carnaval as pessoas têm a tendência de sair de casa para celebrar, o que aumenta os acidentes rodoviários respetivamente.

5.2. Pergunta 2 - Acidentes Graves

Concluindo, o estudo dos casos graves é essencial, sobretudo para criar estratégias de prevenção de acidentes graves. A adoção de modelos que tenham uma boa especificidade leva a que se possam tirar algumas conclusões, relativamente às principais variáveis que possam influenciar a gravidade do acidente. O nosso objetivo com a construção destes modelos de classificação foi o de entender o que caracteriza um acidente grave, para que no futuro se possam tomar medidas que minimizem o risco de acidente grave. Com base nos resultados que obtivemos, podemos constatar que as variáveis mais impactantes e que descrevem melhor um acidente grave vão ser o "mês", o "distrito", a "luminosidade", a "localização" e a "natureza" do acidente.