



Modelo Preditivo de Acidentes Fatais em Rodovias



28/05/2020



Pós-graduação Análise de Dados, Data Mining e Inteligência Artificial

Nome do Aluno:

Letícia Otero Dias Thomaz

Coordenadores:

Prof.^a Dr.^a Alessandra de Ávila Montini

Prof. Dr. Adolpho Walter Pimazoni Canton

Agenda

- 1. Objetivo do Trabalho
- 2. Contextualização do Problema
- 3. Base de Dados
 - i. Bases originais
 - ii. Principais variáveis
- 4. Análise Exploratória de Dados
- 5. Modelagem com Estatística Tradicional
- 6. Modelagem com Inteligência Artificial
- 7. Conclusões

1. Objetivo do Trabalho

O objetivo do trabalho é **predizer o tipo de acidente** ocorrido nas rodovias brasileiras, considerando as categorias:

- Sem vítimas;
- Com vítimas feridas;
- Com vítimas fatais.

A predição será realizada utilizando **dados históricos de acidentes, modelos estatísticos e algoritmos de Inteligência Artificial**, que selecionarão as **características mais relevantes** que explicam o evento de interesse.

Desta forma, podemos identificar as características que mais resultam em acidentes com vítimas fatais e realizar **ações preventivas** em prol de diminuir esses números.



2. Contextualização do Problema

Em 2019, os acidentes nas rodovias brasileiras mataram em média 14 pessoas por dia.

A Polícia Rodoviária Federal atende cerca de 70 mil quilômetros de rodovias federais e uma de suas responsabilidades é atender acidentes. Estes são registrados no sistema BR-Brasil, que coleta informações como: estado físico dos envolvidos, informações sobre os veículos, causa do acidente, etc). Em atendimento às diretrizes do Programa de Dados Abertos Governamentais, estes dados são disponibilizados em arquivos de formato .csv.

O objetivo deste estudo é tentar prever a classificação dos acidentes entre: sem vítimas, com vítimas feridas ou com vítimas fatais. E com isso, identificar quais características mais contribuem para acidentes com vítimas (especialmente vítimas fatais).

Este resultado pode servir para orientar campanhas ou ações nas rodovias para que o número de acidentes diminua cada vez mais.

Fonte: <https://portal.prf.gov.br/dados-abertos-dicionario-acidentes>



3.i. Base original



1.184.445 acidentes

Visão da base

- Acidente

Filtros de inclusão

- Acidentes ocorridos de 01/01/2011 a 31/03/2020 em rodovias brasileiras

Filtros de exclusão

- Acidentes sem classificação



3.ii. Principais variáveis

A base de dados apresenta **24 variáveis explicativas** e a **variável resposta**.



Variáveis Acidente

- Causa do Acidente
- Tipo de Acidentes
- Nº Pessoas Envolvidas
- Nº Feridos
- Nº Veículos



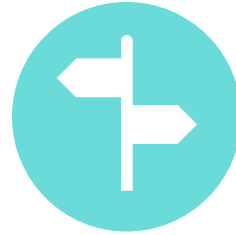
Variáveis Data

- Data do acidente
- Ano do acidente
- Dia da Semana
- Horário
- Fase do Dia



Variável Localização

- UF
- BR (Rodovia)
- Km
- Município



Variável Via

- Sentido Via
- Tipo de Pista
- Traçado da Via
- Uso do Solo



Variável Clima

- Condição Meteorológica



Variável Resposta

Classificação do Acidente:

- Sem vítimas;
- Com vítimas feridas;
- Com vítimas fatais



4. Análise Exploratória

ANÁLISE EXPLORATÓRIA | Variável Resposta

8



Variáveis Acidente

- Causa do Acidente
- Tipo de Acidentes
- Nº Pessoas Envolvidas
- Nº Feridos
- Nº Veículos



Variáveis Data

- Data do acidente
- Ano do acidente
- Dia da Semana
- Horário
- Fase do Dia



Variável Localização

- UF
- BR (Rodovia)
- Km
- Município



Variável Via

- Sentido Via
- Tipo de Pista
- Traçado da Via
- Uso do Solo



Variável Clima

- Condição Meteorológica



Variável Resposta

Classificação do Acidente:

- Sem vítimas;
- Com vítimas feridas;
- Com vítimas fatais



Classificação dos Acidentes

ANÁLISE EXPLORATÓRIA | Variável Resposta

9

A base total conta com 1.184.445 acidentes ocorridos de 2011 à Março de 2020.

Na tabela abaixo, observa-se a quantidade de cada tipo de acidente.
Nota-se que apenas **5%** dos acidentes têm vítimas fatais.

Acidentes sem vítimas correspondem a **50%** do total de acidentes.
E com vítimas feridas, têm-se **45%** das observações.

Classificação do Acidente	Quantidade	(%)
Com Vítimas Fatais	54.080	5%
Com Vítimas Feridas	533.618	45%
Sem Vítimas	596.747	50%
Total	1.184.445	100%



4. Análise Exploratória

ANÁLISE EXPLORATÓRIA | Acidente

10



Variáveis Acidente

- Causa do Acidente
- Tipo de Acidentes
- Nº Pessoas Envolvidas
- Nº Feridos
- Nº Veículos



Variáveis Data

- Data do acidente
- Ano do acidente
- Dia da Semana
- Horário
- Fase do Dia



Variável Localização

- UF
- BR (Rodovia)
- Km
- Município



Variável Via

- Sentido Via
- Tipo de Pista
- Traçado da Via
- Uso do Solo



Variável Clima

- Condição Meteorológica



Variável Resposta

Classificação do Acidente:

- Sem vítimas;
- Com vítimas feridas;
- Com vítimas fatais



Causa do Acidente

Quantidade de acidentes por causa

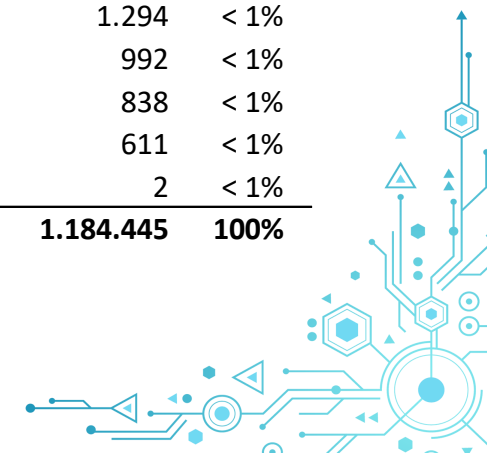
11

A principal causa dos acidentes analisados é a **Falta de Atenção (33%)**.

Velocidade Incompatível ocupa a terceira posição com **10%** do total de acidentes.

Ingestão de álcool foi a causa de 61.954 (**5%**) acidentes de 2011 a 2020.

Causa do Acidente	Quantidade	(%)
Falta de Atenção à Condução	395.451	33%
Outras	238.248	20%
Velocidade incompatível	119.008	10%
Não guardar distância de segurança	110.207	9%
Desobediência às normas de trânsito	65.110	5%
Ingestão de álcool	61.954	5%
Defeito Mecânico no Veículo	51.525	4%
Condutor Dormindo	35.776	3%
Animais na Pista/Fenômenos da Natureza	31.081	3%
Defeito na Via/Pista Escorregadia	27.468	2%
Ultrapassagem Indevida	27.193	2%
Desobediência às normas ou Falta de Atenção do pedestre	7.630	1%
Avarias e/ou desgaste excessivo no pneu	3.659	< 1%
Mal Súbito	2.520	< 1%
Restrição de Visibilidade	2.027	< 1%
Objeto estático sobre o leito carroçável	1.851	< 1%
Carga excessiva e/ou mal acondicionada	1.294	< 1%
Sinalização da via insuficiente ou inadequada	992	< 1%
Agressão Externa	838	< 1%
Deficiência ou não Acionamento do Sistema de Iluminação/Sinalização do Veículo	611	< 1%
Não informado	2	< 1%
Total	1.184.445	100%



Cinco Principais Causas dos Acidentes

Quantidade de acidentes por causa e classificação

12

Neste slide, exibimos as cinco principais causas de cada tipo de acidente.

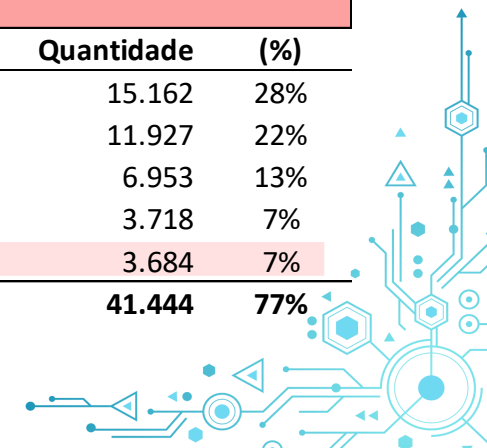
Não guardar distância de segurança corresponde a **13%** dos acidentes **sem vítimas**. Esta causa não aparece entre as cinco principais de acidentes com vítimas.

Ultrapassagem indevida aparece em quinto lugar (**7%**) quando observamos acidentes com **vítimas fatais**.

Sem Vítimas		
Causa do Acidente	Quantidade	(%)
Falta de Atenção à Condução	204.178	34%
Outras	125.991	21%
Não guardar distância de segurança	76.867	13%
Velocidade incompatível	55.393	9%
Defeito Mecânico no Veículo	29.929	5%
Total	492.358	83%

Com Vítimas Feridas		
Causa do Acidente	Quantidade	(%)
Falta de Atenção à Condução	179.346	34%
Outras	97.095	18%
Velocidade incompatível	56.662	11%
Desobediência às normas de trânsito	39.522	7%
Ingestão de álcool	34.551	6%
Total	407.176	76%

Com Vítimas Fatais		
Causa do Acidente	Quantidade	(%)
Outras	15.162	28%
Falta de Atenção à Condução	11.927	22%
Velocidade incompatível	6.953	13%
Desobediência às normas de trânsito	3.718	7%
Ultrapassagem Indevida	3.684	7%
Total	41.444	77%



Tipos de Acidentes

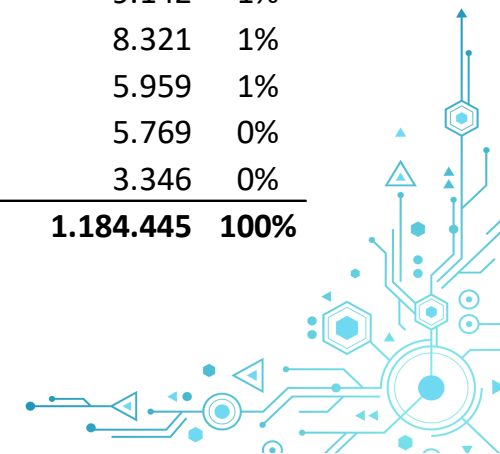
Quantidade de acidentes por tipo de acidente

13

Na tabela ao lado vemos que Colisão traseira é o tipo de acidente mais comum observado (**26%**). Seguido por Colisão lateral (**15%**) e Saída de Pista (**12%**).

Colisão frontal corresponde a **4%** do total de acidentes, porém, nos próximos slides, veremos que este tipo ocupa o primeiro lugar nos acidentes com vítimas fatais.

Tipo de Acidente	Quantidade	(%)
Colisão traseira	310.866	26%
Colisão lateral	182.448	15%
Saída de Pista	137.083	12%
Colisão transversal	128.215	11%
Colisão com objeto fixo	71.745	6%
Tombamento	56.053	5%
Capotamento	55.506	5%
Colisão frontal	52.589	4%
Saída de leito carroçável	40.069	3%
Atropelamento de Pedestre	36.591	3%
Queda de motocicleta / bicicleta / veículo	35.149	3%
Atropelamento de animal	25.489	2%
Queda de ocupante de veículo	10.627	1%
Colisão com objeto móvel	9.478	1%
Colisão com bicicleta	9.142	1%
Incêndio	8.321	1%
Danos eventuais	5.959	1%
Derramamento de carga	5.769	0%
Engavetamento	3.346	0%
Total	1.184.445	100%



Tipos de Acidentes – Vítimas Fatais

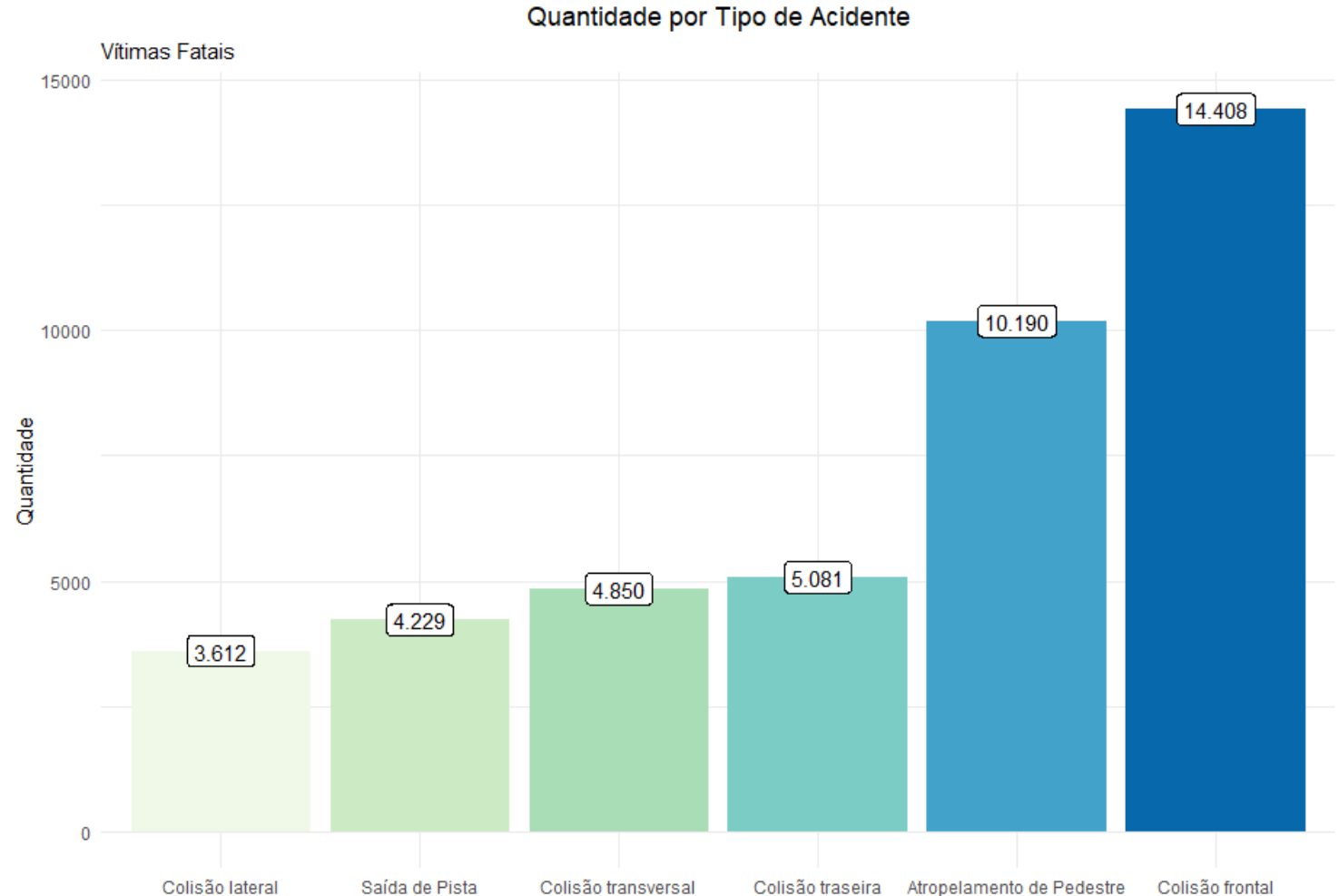
Quantidade de acidentes por tipo de acidente

14

Colisão Frontal é o principal tipo de acidente que envolve vítimas fatais (27%).

Em seguida, observamos Atropelamento de Pedestre com 19%.

Tipo de Acidente	Quantidade	(%)
Colisão frontal	14.408	27%
Atropelamento de Pedestre	10.190	19%
Colisão traseira	5.081	9%
Colisão transversal	4.850	9%
Saída de Pista	4.229	8%
Colisão lateral	3.612	7%
Total	42.370	78%



Quantidade de Veículos

Quantidade de veículos por acidente

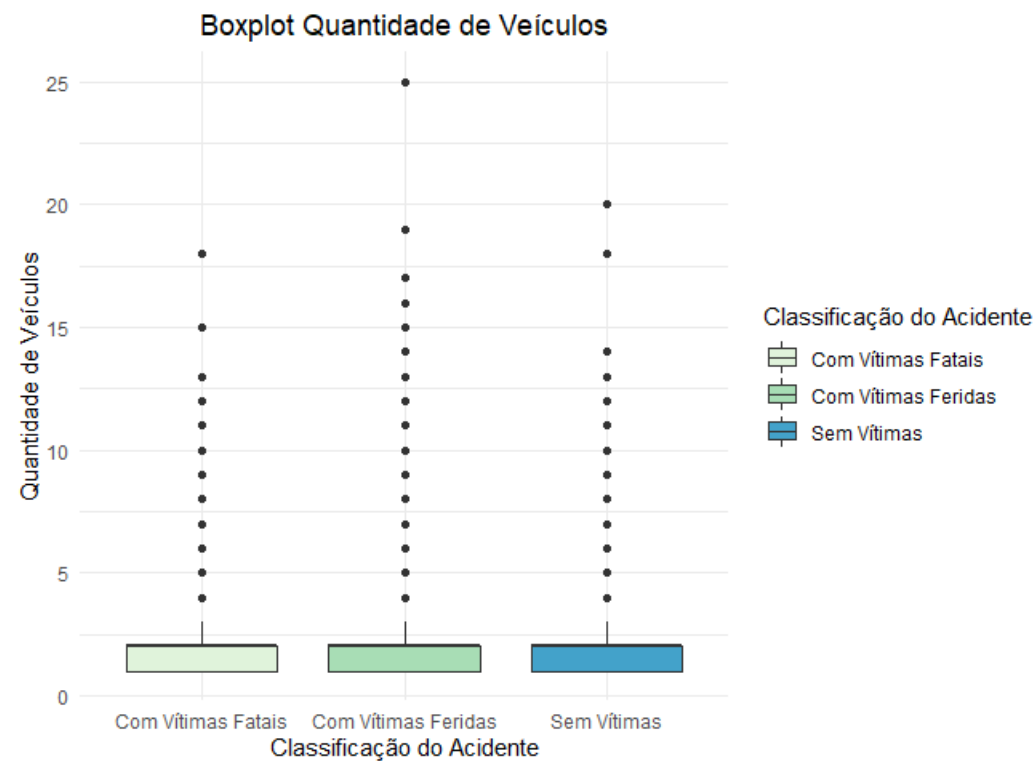
15

Em média, há **1,74 veículo** envolvido por acidente. O máximo observado na base de dados foi um acidente que envolveu **25 veículos**.

Quando observamos o *boxplot* da quantidade de veículos por classificação do acidente, nota-se que aparentemente não há diferença nas distribuições.

Pelo 3º quartil (75% dos dados), observa-se que a maior parte dos acidentes envolve até **2 veículos**.

Quantidade de Veículos						
Qtde	Média	Mín	Máx	1º Quartil	Mediana	3º Quartil
1.184.445,00	1,74	1,00	25,00	1,00	2,00	2,00



Quantidade de Pessoas

Quantidade de pessoas por acidente

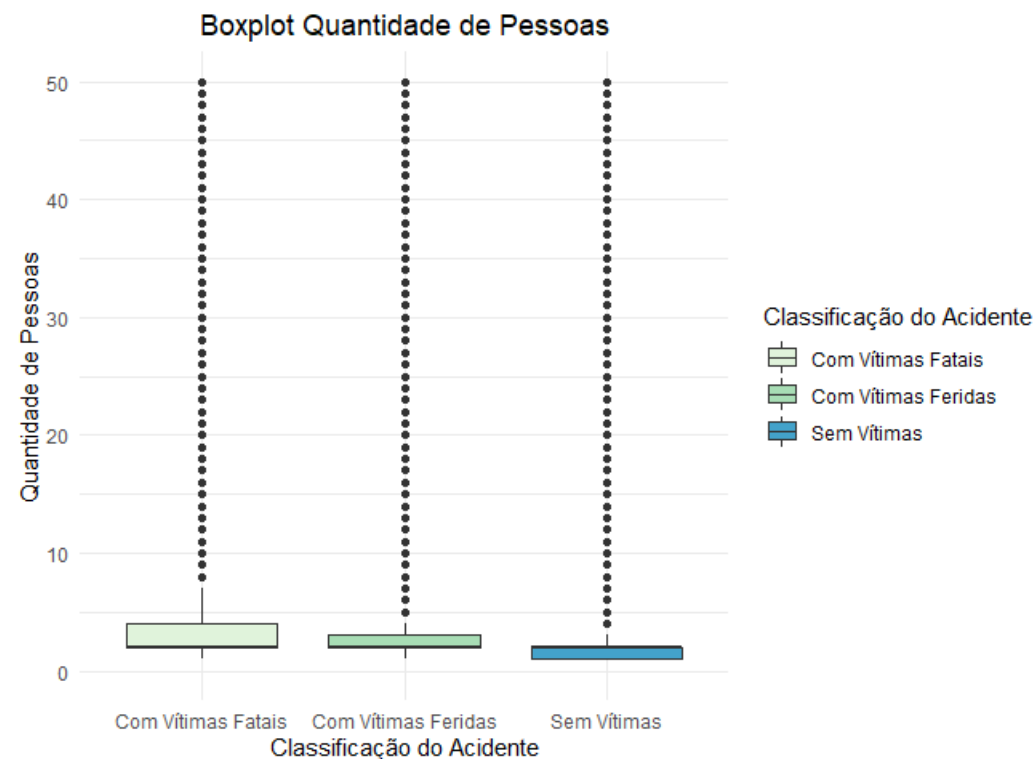
16

Em média, há **2,22 pessoas** envolvidas por acidente. O máximo observado na base de dados foi um acidente que envolveu **248 pessoas**.

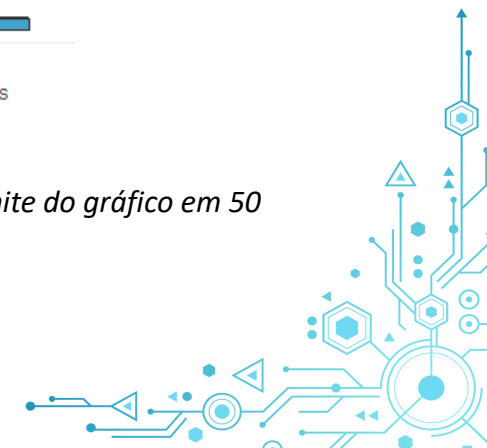
Quando observamos o *boxplot* da quantidade de pessoas por classificação do acidente, nota-se que aparentemente a acidentes sem vítimas envolvem menos pessoas. Porém, as três classificações apresentam muitos *outliers*.

Pelo 3º quartil (75% dos dados), observa-se que a maior parte dos acidentes envolve até **2 pessoas**.

Quantidade de Pessoas						
Qtde	Média	Mín	Máx	1º Quartil	Mediana	3º Quartil
1.184.445,00	2,22	1,00	248,00	1,00	2,00	2,00



Para facilitar a visualização, definimos o limite do gráfico em 50



4. Análise Exploratória

ANÁLISE EXPLORATÓRIA | Data

17



Variáveis Acidente

- Causa do Acidente
- Tipo de Acidentes
- Nº Pessoas Envolvidas
- Nº Feridos
- Nº Veículos



Variáveis Data

- Data do acidente
- Ano do acidente
- Dia da Semana
- Horário
- Fase do Dia



Variável Localização

- UF
- BR (Rodovia)
- Km
- Município



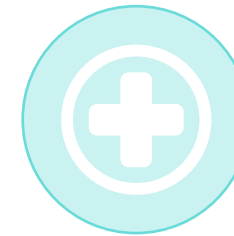
Variável Via

- Sentido Via
- Tipo de Pista
- Traçado da Via
- Uso do Solo



Variável Clima

- Condição Meteorológica



Variável Resposta

Classificação do Acidente:

- Sem vítimas;
- Com vítimas feridas;
- Com vítimas fatais

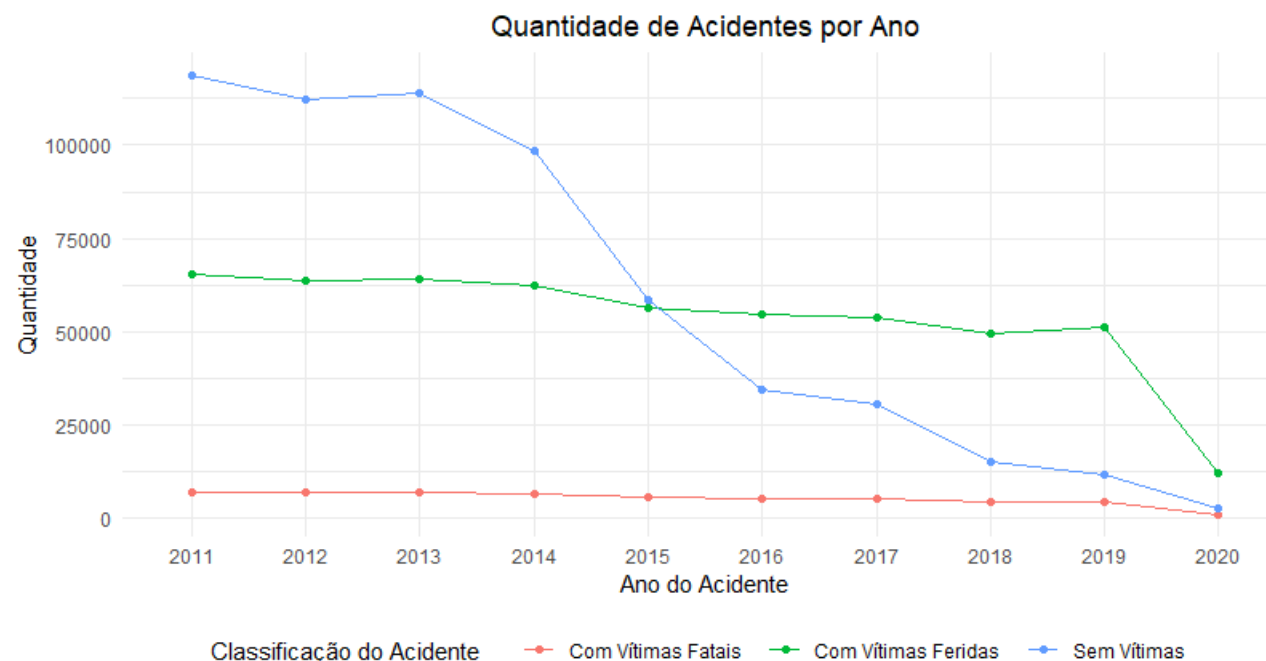




De 2011 a 2020, observou-se uma queda acentuada nos acidentes **sem vítimas**.

Acidentes com **vítimas feridas** apresentou uma queda menos agressiva, e com **vítimas fatais** manteve-se no mesmo patamar ao longo do tempo.

A queda apresentada em 2020 deve-se ao fato de os dados disponibilizados iam apenas até Março.



Dia da semana

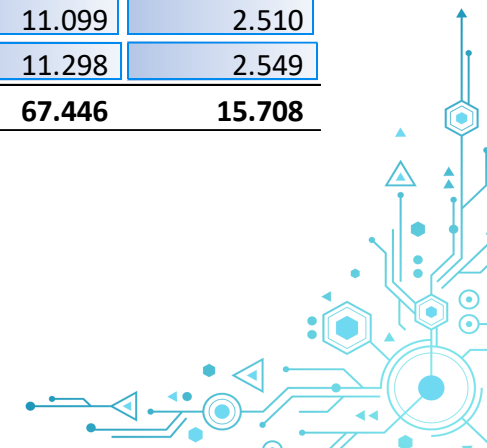
Quantidade de acidentes por dia da semana



Na tabela abaixo, observa-se que nas **sextas e finais de semana**, há um leve aumento na quantidade de acidentes quando comparamos com os demais dias.

Nota-se também, que há uma queda no número de acidentes ao longo do tempo.

Geral										
Dia da Semana	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Segunda	26.465	25.358	26.127	23.434	16.913	12.721	12.175	9.422	9.134	2.112
Terça	24.616	23.196	24.218	21.563	15.277	11.518	10.765	8.665	8.405	1.978
Quarta	25.044	24.102	24.687	22.474	15.936	11.948	11.029	8.824	8.374	2.075
Quinta	26.279	25.413	25.678	23.113	16.781	12.612	11.787	9.230	8.755	2.072
Sexta	30.629	30.632	30.713	27.669	18.784	14.882	13.849	10.632	10.381	2.412
Sábado	30.080	27.634	27.678	25.423	18.729	15.872	15.174	11.239	11.099	2.510
Domingo	28.048	26.859	26.109	23.853	18.213	15.153	14.784	11.283	11.298	2.549
Total	191.161	183.194	185.210	167.529	120.633	94.706	89.563	69.295	67.446	15.708



Dia da semana – Sem vítimas

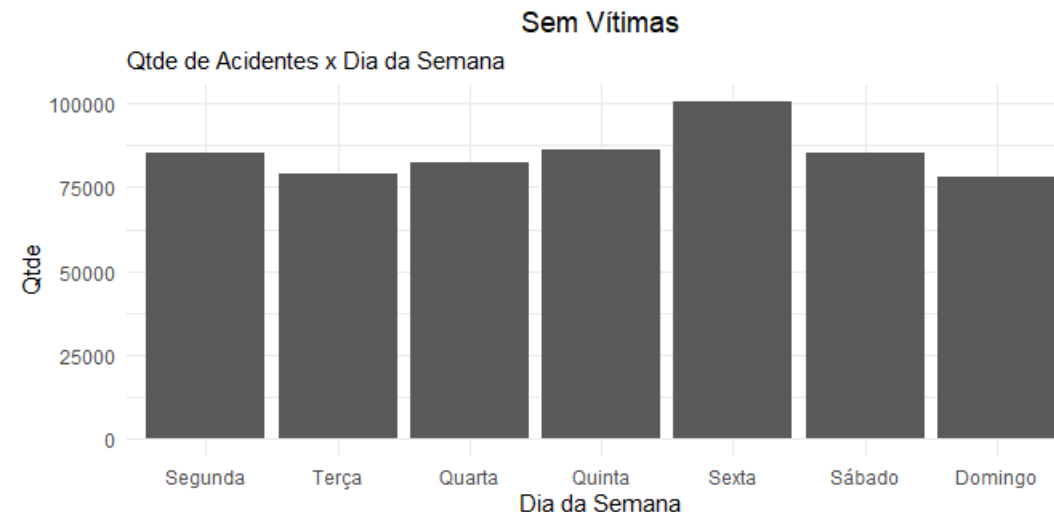
Quantidade de acidentes por dia da semana



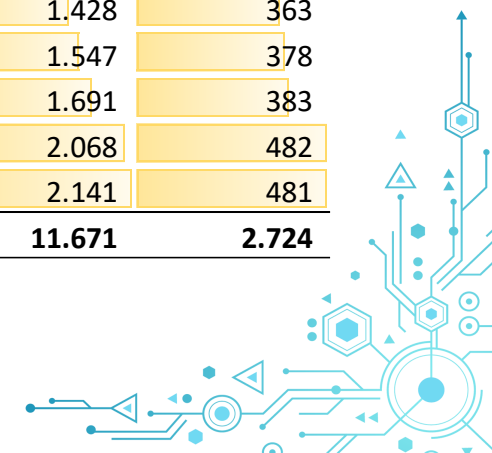
20

Observando a tabela nota-se que em anos anteriores a 2017, a **Sexta** se destacava na quantidade de acidentes.

De 2018 em diante, **Sábado e Domingo** apresentaram valores superiores que os demais dias da semana.



Sem Vítimas										
Dia da Semana	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Segunda	17.134	16.043	16.570	14.357	8.493	4.692	4.282	1.997	1.462	310
Terça	16.169	14.870	15.568	13.255	7.783	4.324	3.693	1.925	1.334	327
Quarta	16.355	15.614	15.873	13.927	8.126	4.608	3.883	1.970	1.428	363
Quinta	17.209	16.555	16.725	14.200	8.593	4.802	4.227	2.043	1.547	378
Sexta	19.737	19.652	19.779	16.966	9.466	5.663	4.902	2.310	1.691	383
Sábado	17.001	15.438	15.541	13.592	8.243	5.377	4.985	2.483	2.068	482
Domingo	15.140	14.251	13.999	12.015	7.707	5.012	4.709	2.519	2.141	481
Total	118.745	112.423	114.055	98.312	58.411	34.478	30.681	15.247	11.671	2.724



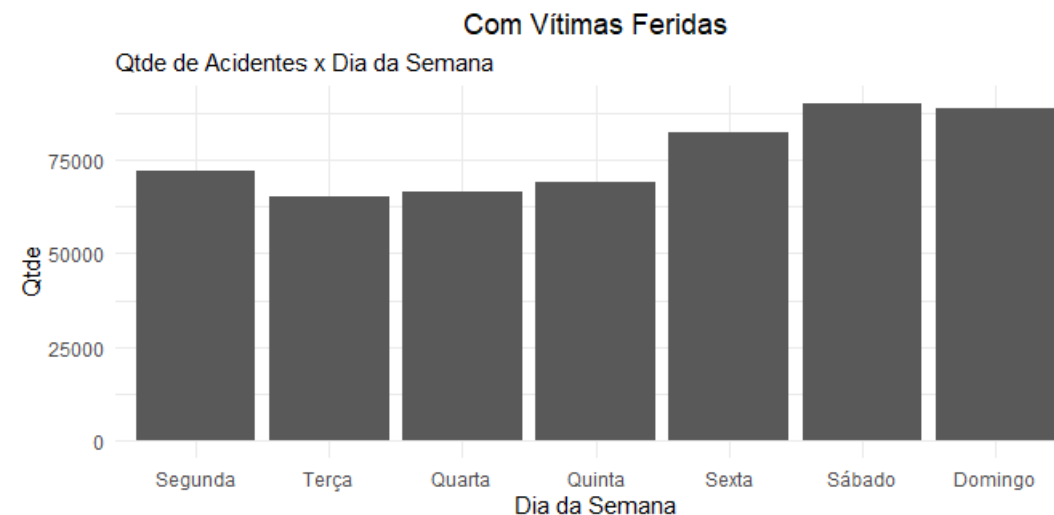
Dia da semana – Vítimas Feridas

Quantidade de acidentes por dia da semana

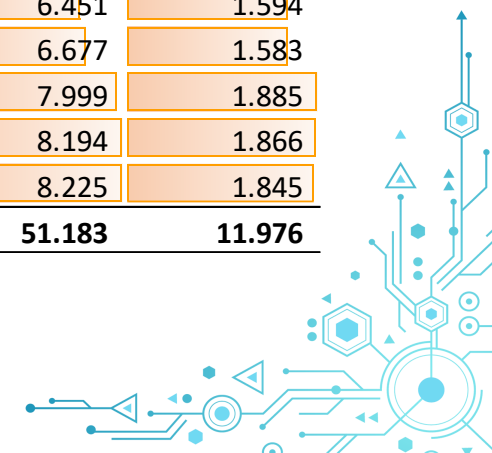


21

Pelo gráfico e pela tabela abaixo, vê-se que acidentes com vítimas feridas ocorrem mais às sextas e finais de semana.



Vítimas Feridas										
Dia da Semana	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Segunda	8.465	8.456	8.732	8.269	7.768	7.408	7.242	6.868	7.052	1.669
Terça	7.712	7.598	7.886	7.591	6.906	6.644	6.559	6.235	6.585	1.534
Quarta	7.931	7.738	8.053	7.789	7.198	6.740	6.599	6.350	6.451	1.594
Quinta	8.196	8.014	8.184	8.091	7.491	7.192	6.932	6.638	6.677	1.583
Sexta	9.878	9.969	9.958	9.752	8.499	8.452	8.232	7.682	7.999	1.885
Sábado	11.710	10.869	10.816	10.549	9.415	9.460	9.185	7.918	8.194	1.866
Domingo	11.366	11.124	10.639	10.435	9.297	8.977	8.950	7.852	8.225	1.845
Total	65.258	63.768	64.268	62.476	56.574	54.873	53.699	49.543	51.183	11.976



Dia da semana – Vítimas Fatais

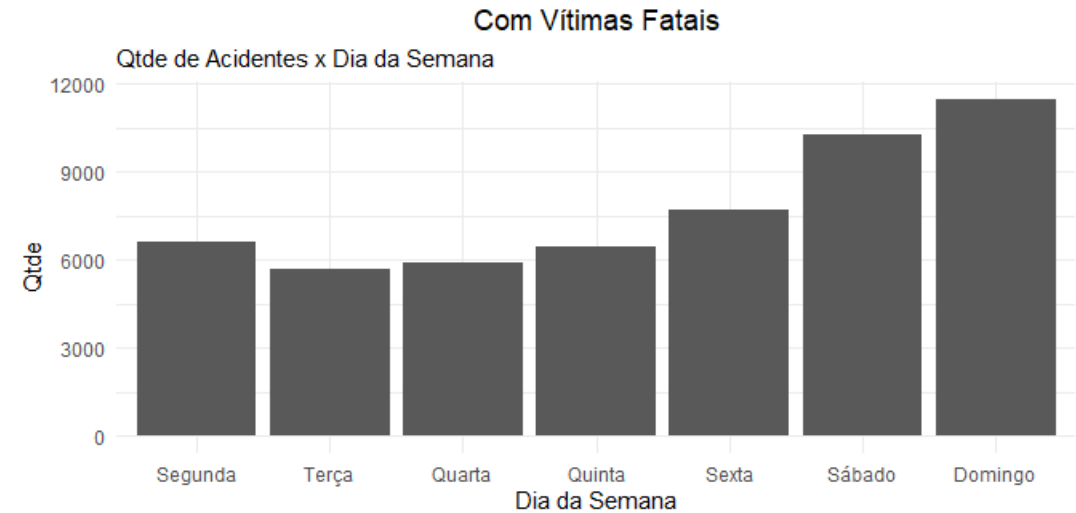
Quantidade de acidentes por dia da semana



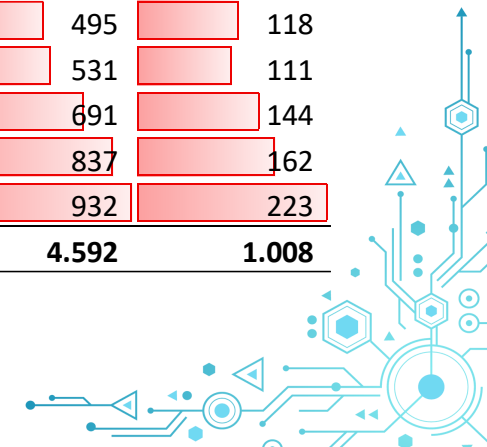
22

Acidentes com vítimas fatais ocorrem com maior frequência aos **finais de semana**.

Apesar de todos os acidentes ocorrerem em maior quantidade a partir de sexta, neste caso o final de semana é muito mais acentuado que para os demais tipos de acidentes (sem vítimas ou com vítimas feridas).



Vítimas Fatais										
Dia da Semana	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Segunda	866	859	825	808	652	621	651	557	620	133
Terça	735	728	764	717	588	550	513	505	486	117
Quarta	758	750	761	758	612	600	547	504	495	118
Quinta	874	844	769	822	697	618	628	549	531	111
Sexta	1.014	1.011	976	951	819	767	715	640	691	144
Sábado	1.369	1.327	1.321	1.282	1.071	1.035	1.004	838	837	162
Domingo	1.542	1.484	1.471	1.403	1.209	1.164	1.125	912	932	223
Total	7.158	7.003	6.887	6.741	5.648	5.355	5.183	4.505	4.592	1.008



Fase do dia

Quantidade de acidentes por fase do dia

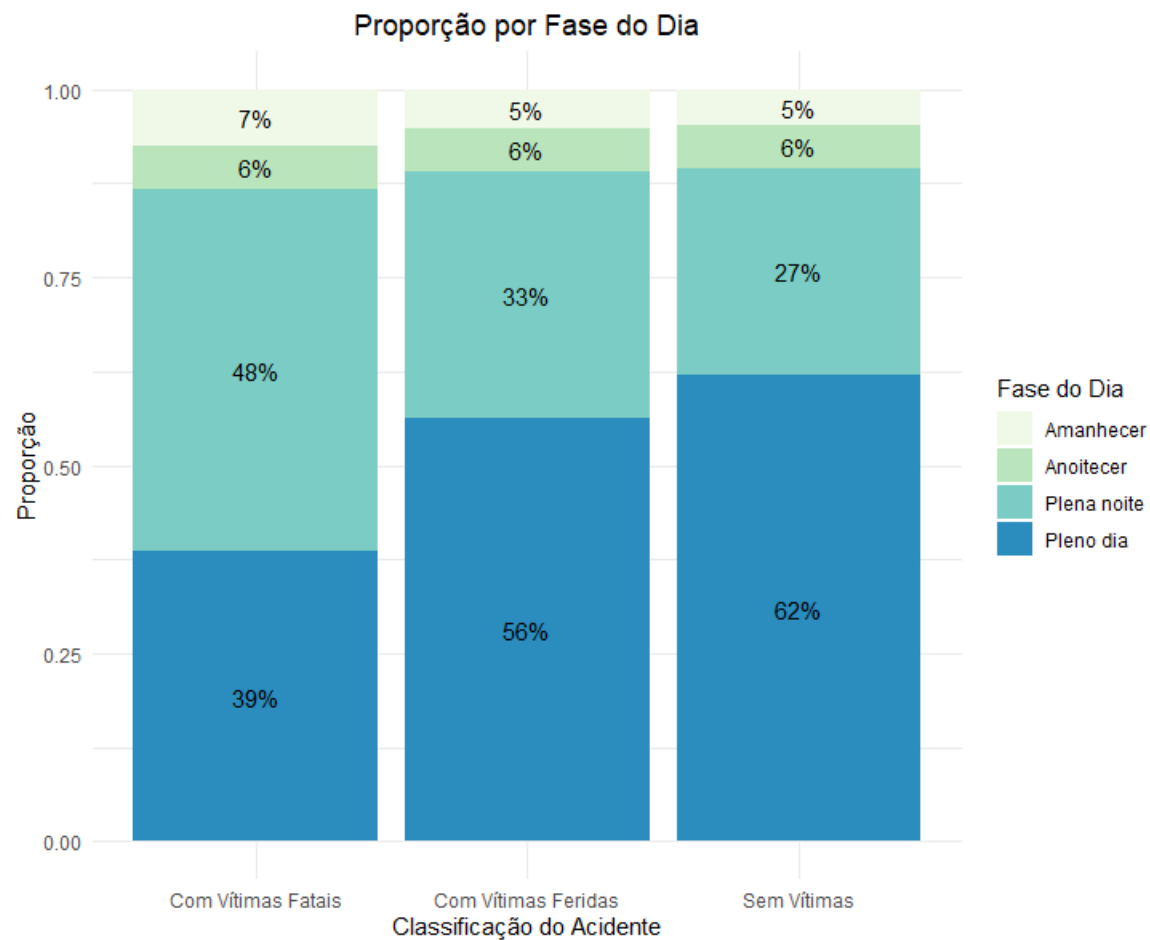


23

58% dos acidentes observados ocorreram em **pleno dia**, como pode ser observado na tabela abaixo.

No entanto, quando observamos o gráfico ao lado, a maior parte dos acidentes com vítimas fatais (**48%**) ocorreram em **plena noite**.

Fase do Dia	Quantidade	(%)
Pleno dia	691.293	58%
Plena noite	364.616	31%
Anoitecer	68.808	6%
Amanhecer	59.702	5%
Sem informação	26	0%
Total	1.184.445	100%



4. Análise Exploratória

ANÁLISE EXPLORATÓRIA | Local

24



Variáveis Acidente

- Causa do Acidente
- Tipo de Acidentes
- Nº Pessoas Envolvidas
- Nº Feridos
- Nº Veículos



Variáveis Data

- Data do acidente
- Ano do acidente
- Dia da Semana
- Horário
- Fase do Dia



Variável Localização

- UF
- BR (Rodovia)
- Km
- Município



Variável Via

- Sentido Via
- Tipo de Pista
- Traçado da Via
- Uso do Solo



Variável Clima

- Condição Meteorológica



Variável Resposta

Classificação do Acidente:

- Sem vítimas;
- Com vítimas feridas;
- Com vítimas fatais





Na primeira tabela, vemos o total de acidentes (2011 a 2020) em cada estado. Os três estados que mais se destacam são: **Minas Gerais, Paraná e Santa Catarina**.

Nas tabelas menores, selecionamos as três rodovias em que mais ocorrem esses acidentes. Em Minas e no Paraná essas rodovias detêm **65%** de todos os acidentes ocorridos.

Já em Santa Catarina, as 3 principais rodovias detêm **84%** de todos os acidentes ocorridos.

UF	Quantidade	(%)	UF	Quantidade	(%)
MG	164.150	14%	RN	24.212	2%
PR	130.668	11%	PA	22.961	2%
SC	128.045	11%	RO	22.420	2%
RJ	99.514	8%	PI	19.321	2%
RS	91.875	8%	MA	19.256	2%
SP	84.873	7%	AL	12.721	1%
BA	66.062	6%	DF	11.121	1%
GO	53.366	5%	SE	10.255	1%
ES	48.390	4%	TO	8.536	1%
PE	48.266	4%	AC	3.536	0%
MT	32.896	3%	RR	2.557	0%
CE	25.888	2%	AP	1.909	0%
PB	25.826	2%	AM	1.505	0%
MS	24.312	2%			

UF	Rodovia	Qtde	
MG	BR 381	57.397	
MG	BR 40	31.616	65%
MG	BR 116	18.424	

UF	Rodovia	Qtde	
PR	BR 376	33.490	
PR	BR 277	32.892	64%
PR	BR 116	16.967	

UF	Rodovia	Qtde	
SC	BR 101	63.063	
SC	BR 282	22.387	84%
SC	BR 470	21.981	





Quando realizamos essa mesma análise levando em conta o tipo de acidente, Minas Gerais, Paraná e Santa Catarina continuam ocupando as primeiras posições para acidentes sem vítimas ou com vítimas feridas.

Porém, para acidentes com vítimas **fatais**, a **Bahia** acaba ocupando a 3ª posição.

UF	Quantidade	(%)	UF	Quantidade	(%)
MG	7.219	13%	PA	1.369	3%
PR	5.095	9%	PB	1.323	2%
BA	4.862	9%	MS	1.254	2%
SC	3.689	7%	RN	1.240	2%
RJ	3.451	6%	AL	985	2%
RS	3.103	6%	RO	864	2%
PE	3.088	6%	TO	745	1%
GO	2.931	5%	SE	554	1%
SP	2.467	5%	DF	428	1%
MA	2.036	4%	RR	248	0%
MT	1.833	3%	AC	179	0%
CE	1.738	3%	AM	148	0%
ES	1.634	3%	AP	110	0%
PI	1.486	3%			



4. Análise Exploratória

ANÁLISE EXPLORATÓRIA | Via

27



Variáveis Acidente

- Causa do Acidente
- Tipo de Acidentes
- Nº Pessoas Envolvidas
- Nº Feridos
- Nº Veículos



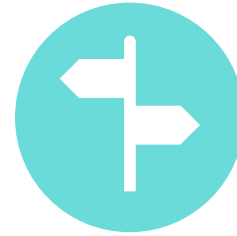
Variáveis Data

- Data do acidente
- Ano do acidente
- Dia da Semana
- Horário
- Fase do Dia



Variável Localização

- UF
- BR (Rodovia)
- Km
- Município



Variável Via

- Sentido Via
- Tipo de Pista
- Traçado da Via
- Uso do Solo



Variável Clima

- Condição Meteorológica



Variável Resposta

Classificação do Acidente:

- Sem vítimas;
- Com vítimas feridas;
- Com vítimas fatais



Tipo de Pista

Quantidade de acidentes por tipo de pista

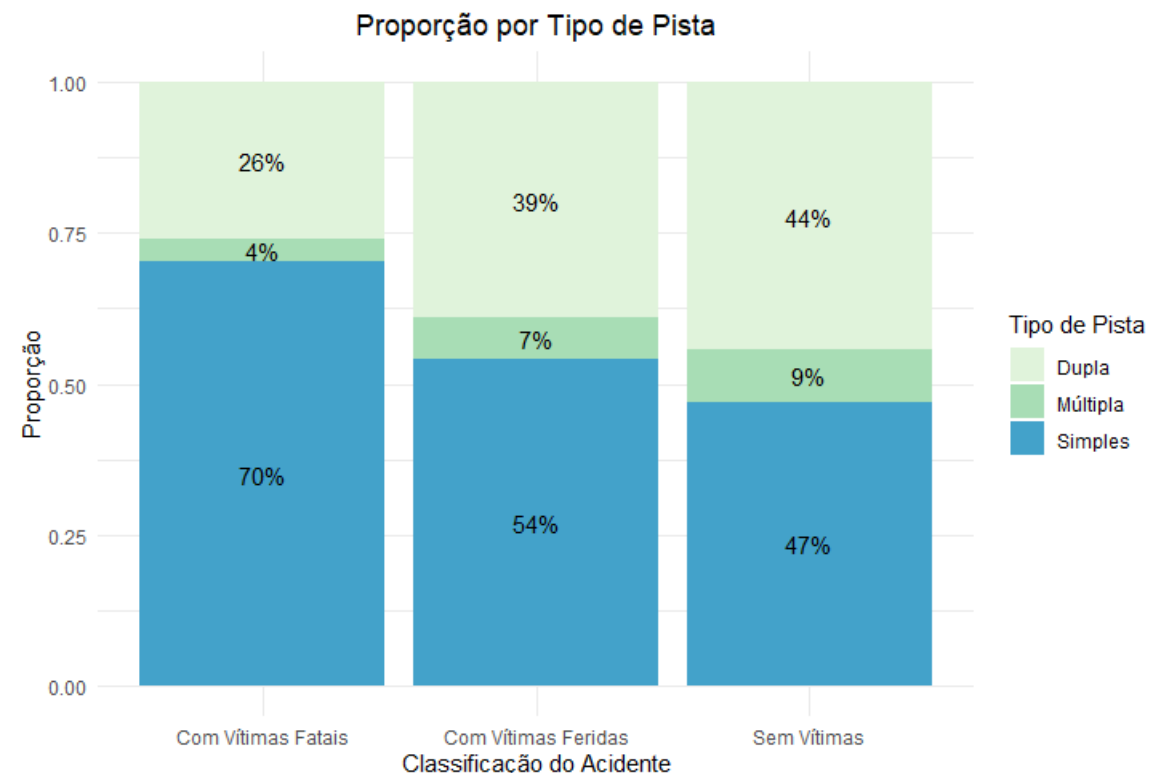
28

51% dos acidentes observados ocorreram em **pista simples**, como pode ser observado na tabela abaixo.

No entanto, quando observamos o gráfico ao lado, enquanto **70%** dos acidentes com vítimas fatais ocorreram em pistas simples, apenas **47%** dos sem vítimas ocorreram neste tipo de pista.

A pista dupla aparece em **44%** dos acidentes sem vítimas, e em **39%** dos acidentes com vítimas feridas.

Tipo da Pista	Quantidade	(%)
Simples	606.882	51%
Dupla	487.054	41%
Múltipla	90.499	8%
Não informado	10	0%
Total	1.184.445	100%



Traçado da Via

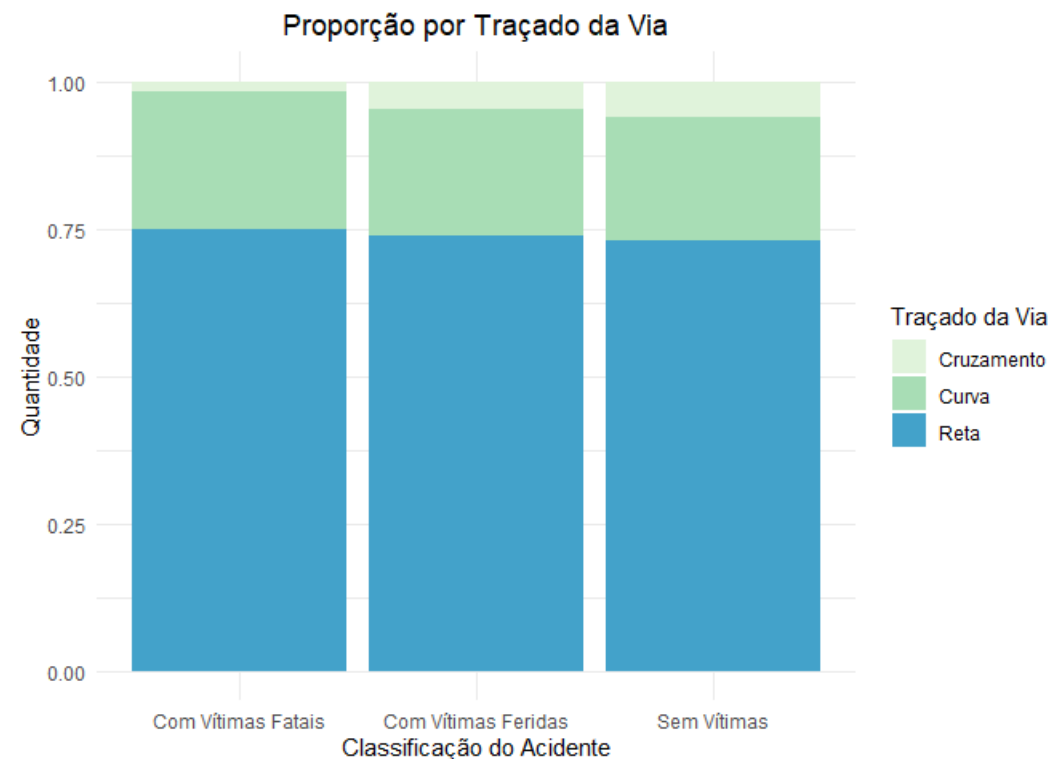
Quantidade de acidentes por traçado da via

29

A maior parte dos acidentes (**75%**) ocorreu em uma **reta**. E esse percentual se mantém em todos os tipos de acidentes, como pode ser visto no gráfico ao lado.

Ou seja, considerando os tipos de traçados, não há um que possa diferenciar as classificações dos acidentes.

Traçado da Via	Quantidade	(%)
Reta	829.840	70%
Curva	240.992	20%
Cruzamento	60.892	5%
Não Informado	23.977	2%
Interseção de vias	11.069	1%
Desvio Temporário	6.638	1%
Rotatória	4.978	0%
Retorno Regulamentado	2.736	0%
Viaduto	1.711	0%
Ponte	1.353	0%
Túnel	259	0%
Total	1.184.445	100%



4. Análise Exploratória

ANÁLISE EXPLORATÓRIA | Clima

30



Variáveis Acidente

- Causa do Acidente
- Tipo de Acidentes
- Nº Pessoas Envolvidas
- Nº Feridos
- Nº Veículos



Variáveis Data

- Data do acidente
- Ano do acidente
- Dia da Semana
- Horário
- Fase do Dia



Variável Localização

- UF
- BR (Rodovia)
- Km
- Município



Variável Via

- Sentido Via
- Tipo de Pista
- Traçado da Via
- Uso do Solo



Variável Clima

- Condição Meteorológica



Variável Resposta

Classificação do Acidente:

- Sem vítimas;
- Com vítimas feridas;
- Com vítimas fatais



Condição Meteorológica

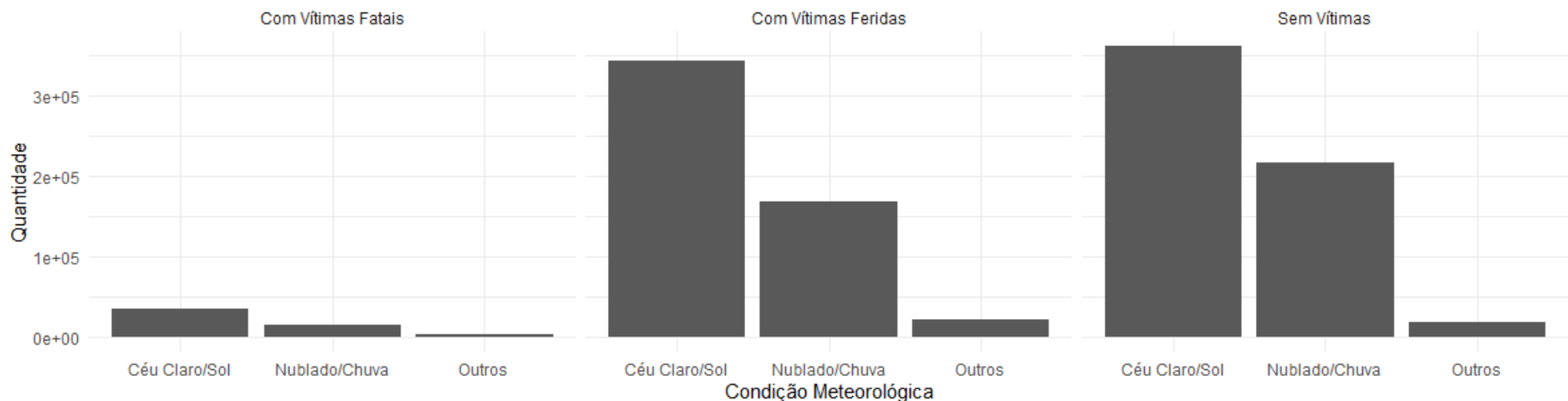
Quantidade de acidentes por UF



A tabela ao lado exibe a quantidade de acidentes por condição meteorológica. Apesar do tempo chuvoso ou nublado nos causar uma impressão mais perigosa, **62%** dos acidentes ocorreram com o **céu claro** ou com **sol**.

Enquanto **chuva** e tempo **nublado** correspondem a **34%** do total.

Condição Meteorológica	Quantidade	(%)
Céu Claro	596.138	50%
Nublado	214.091	18%
Chuva	187.461	16%
Sol	143.685	12%
Ignorada	19.952	2%
Nevoeiro/Neblina	11.227	1%
Garoa/Chuvisco	8.386	1%
Vento	3.292	0%
Granizo	160	0%
Neve	53	0%
Total	1.184.445	100%



4. Conclusão - Análise Exploratória

ANÁLISE EXPLORATÓRIA | Conclusão

32

Pela análise exploratória dos dados é possível observar que algumas variáveis podem ser importantes na hora de diferenciar a classificação dos acidentes.

- Alguns **rodovias** são mais perigosas (maior frequência de acidentes fatais) que outras.
- O período do dia também é importante, pois observa-se que a fase do dia 'Plena Noite' é mais frequente em acidentes fatais que nos demais tipos de acidentes.
- O tipo de acidente também nos mostrou que **colisão traseira** ocupa a terceira posição de acidentes com vítimas fatais, enquanto que para os outros, **colisão traseira** é mais comum (primeira posição).

Ou seja, com os dados que são disponibilizados aparentemente é possível ajustar um modelo para tentar prever a classificação do acidente, que veremos nos próximos slides.



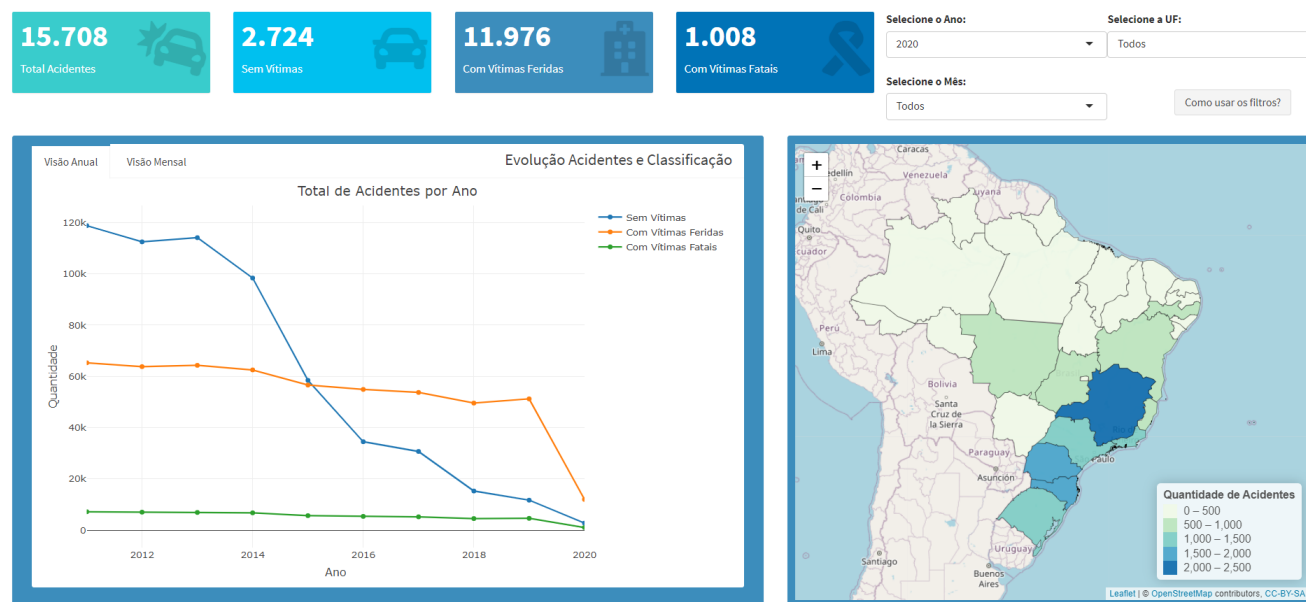
4. Análise Exploratória - Dashboard

ANÁLISE EXPLORATÓRIA | Dashboard

Por conta de um outro curso que realizei, montei essa mesma análise exploratória via dashboard. Utilizando o pacote **shiny** do R.

Para visualizar basta acessar: <https://leticiaoterodiasthomaz.shinyapps.io/Acidentes/>

Ou escanear o QR Code abaixo (a visualização por celular não é tão boa quanto pelo computador)



5. Modelagem com Estatística Tradicional

BALANCEAMENTO | Base Treino e Teste

34

A ABT utilizada para a modelagem inicial possuía 1.184.445 acidentes. Selecionou-se 80% da base para treino (948 mil) e 20% para teste (237 mil).

Como as categorias da variável de interesse estavam muito desbalanceadas como pode ser visto abaixo (Tabela 1), realizou-se mistura entre balanceamento oversampling e undersampling sobre a base de Treino (Tabela 2).

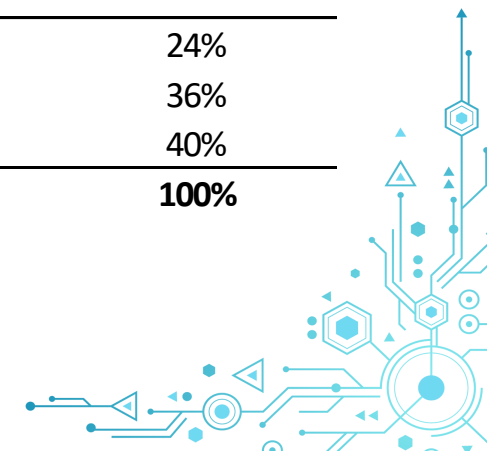
Após isso, um modelo de Regressão Logística e uma Árvore de Decisão foram ajustados.

Tabela 1: Base Treino **sem** balanceamento

Classificação do Acidente	Quantidade	(%)
Com Vítimas Fatais	43.454	5%
Com Vítimas Feridas	426.939	45%
Sem Vítimas	477.417	50%
Total	947.810	100%

Tabela 2: Base Treino **com** balanceamento

Classificação do Acidente	Quantidade	Amostra	(%)
Com Vítimas Fatais	86.908	200%	24%
Com Vítimas Feridas	128.325	30%	36%
Sem Vítimas	143.447	30%	40%
Total	358.680		100%



5 e 6. Modelagem com Estatística Tradicional e Inteligência Artificial

MODELOS | Comparação de Resultados

Estatística Tradicional

Inteligência Artificial

35

Modelo 1: RegLog Multinomial sem Balanceamento

Acurácia: 0,66

Classificação Acidente	Precisão	Recall
Sem Vítimas	0,66	0,80
Com Vítimas Feridas	0,66	0,61
Com Vítimas Fatais	0,00	0,00

Modelo 2: RegLog Multinomial com Balanceamento

Acurácia: 0,63

Classificação Acidente	Precisão	Recall
Sem Vítimas	0,67	0,80
Com Vítimas Feridas	0,67	0,46
Com Vítimas Fatais	0,25	0,49

Modelo 3: Árvore de Decisão com Balanceamento

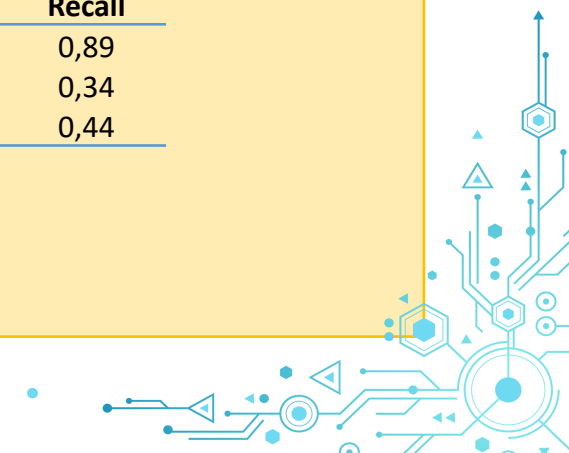
Acurácia: 0,63

Classificação Acidente	Precisão	Recall
Sem Vítimas	0,63	0,88
Com Vítimas Feridas	0,73	0,36
Com Vítimas Fatais	0,29	0,42

Modelo 4: Random Forest com Balanceamento

Acurácia: 0,62

Classificação Acidente	Precisão	Recall
Sem Vítimas	0,63	0,89
Com Vítimas Feridas	0,73	0,34
Com Vítimas Fatais	0,28	0,44



5 e 6. Modelagem com Estatística Tradicional e Inteligência Artificial

MODELOS | Matriz de Confusão

Estatística Tradicional

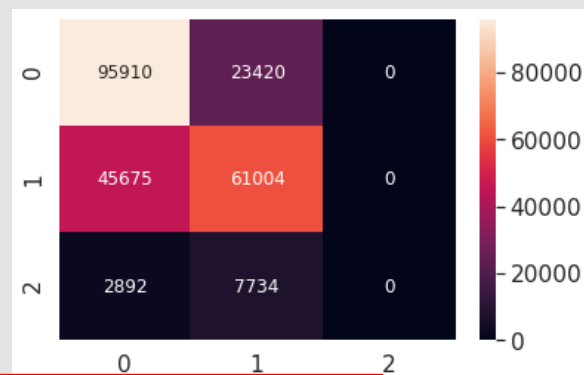
Inteligência Artificial

36

Modelo 1: RegLog Multinomial **sem** Balanceamento

Acurácia: 0,66

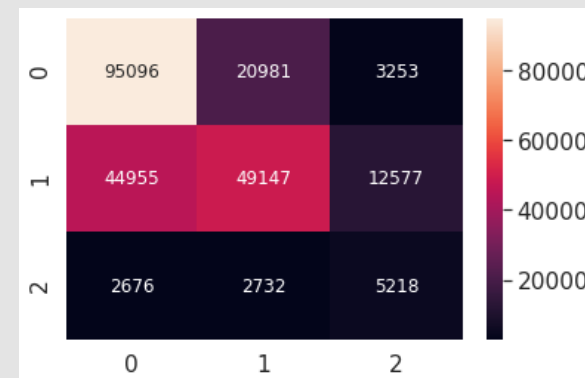
Não realizou
nenhuma
previsão para a
menor classe
(Vítimas Fatais)



Modelo 2: RegLog Multinomial **com** Balanceamento

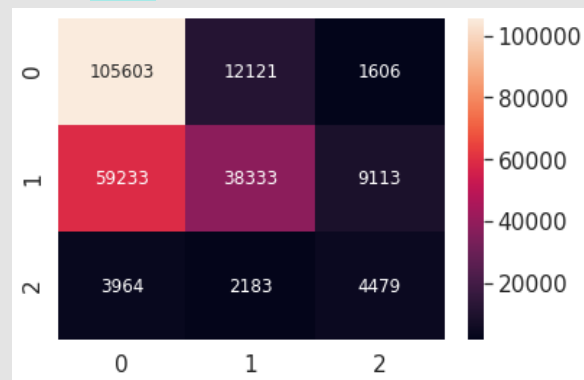
Acurácia: 0,63

Acurácia é menor
quando comparada
com primeiro
modelo, mas ele
passou a prever
casos com vítimas
fatais.



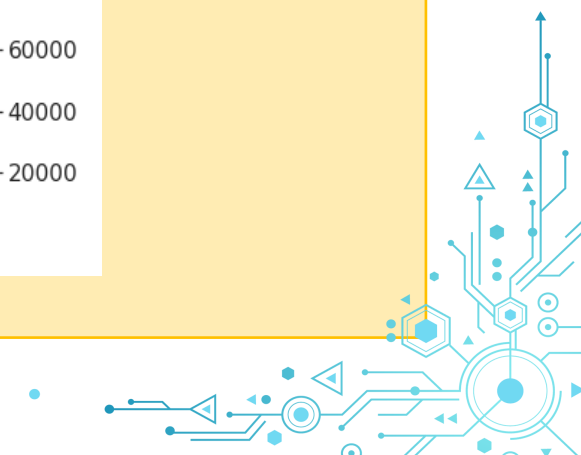
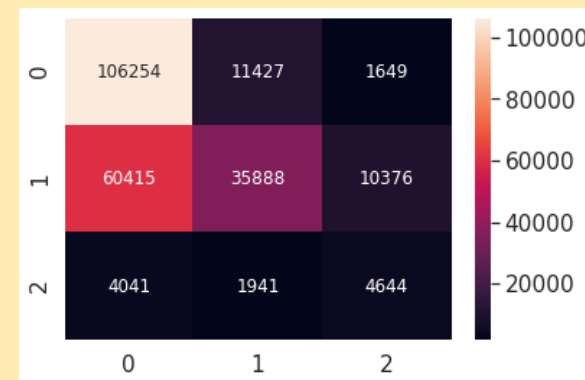
Modelo 3: Árvore de Decisão **com** Balanceamento

Acurácia: 0,63



Modelo 4: Random Forest **com** Balanceamento

Acurácia: 0,62



5 e 6. Modelagem com Estatística Tradicional e Inteligência Artificial

MODELOS | Escolha do melhor modelo

Estatística Tradicional

Inteligência Artificial

37

Modelo 1: RegLog Multinomial **sem** Balanceamento

Acurácia: 0,66

Classificação Acidente	Precisão	Recall
Sem Vítimas	0,66	0,80
Com Vítimas Feridas	0,66	0,61
Com Vítimas Fatais	0,00	0,00

Não realizou nenhuma previsão para a classe de maior interesse

Modelo 2: RegLog Multinomial **com** Balanceamento

Acurácia: 0,63

Classificação Acidente	Precisão	Recall
Sem Vítimas	0,67	0,80
Com Vítimas Feridas	0,67	0,46
Com Vítimas Fatais	0,25	0,49

Árvore de Decisão e Random Forest apresentam um recall menor para acidentes **com vítimas** (feridas ou fatais)

Modelo 3: Árvore de Decisão **com** Balanceamento

Acurácia: 0,63

Classificação Acidente	Precisão	Recall
Sem Vítimas	0,63	0,88
Com Vítimas Feridas	0,73	0,36
Com Vítimas Fatais	0,29	0,42

Árvore de Decisão e Random Forest apresentaram um recall maior para acidentes sem vítimas, porém ...

Modelo 4: Random Forest **com** Balanceamento

Acurácia: 0,62

Classificação Acidente	Precisão	Recall
Sem Vítimas	0,63	0,89
Com Vítimas Feridas	0,73	0,34
Com Vítimas Fatais	0,28	0,44

Sem Vítimas

Fatores que aumentam a probabilidade

Tipo de Acidente: colisão com objeto fixo, lateral ou traseira

Causa do Acidente: não guardar distância segura, carga excessiva ou mal condicionada

Fatores que diminuem a probabilidade

Tipo de Acidente: colisão frontal, atropelamento de pedestre, queda de motocicleta/bicicleta

Causa do Acidente: Mal súbito

Vítimas Feridas

Fatores que aumentam a probabilidade

Tipo de Acidente: capotamento, tombamento, queda de motocicleta/bicicleta

Causa do Acidente: Avarias ou desgaste excessivo do pneu

Fatores que diminuem a probabilidade

Tipo de Acidente: colisão frontal, colisão traseira

Causa do Acidente: deficiência ou não acionamento do sistema de iluminação do veículo, carga excessiva ou mal condicionada

Vítimas Fatais

Fatores que aumentam a probabilidade

Tipo de Acidente: atropelamento de pedestre, colisão frontal

Causa do Acidente: desobediência às normas e falta de atenção do condutor, deficiência ou não acionamento do sistema de iluminação do veículo

Fatores que diminuem a probabilidade

Tipo de Acidente: colisão traseira, tombamento

Causa do Acidente: não guardar distância segura

Obs: o modelo tenta prever tipos de acidentes, então fatores que diminuem a probabilidade de determinado tipo de acidente ocorrer, possivelmente irá aumentar a probabilidade em outra categoria.



7. Conclusão

O modelo final tenta prever a classificação dos acidentes entre sem vítimas, com vítimas feridas ou com vítimas fatais. O modelo apresentou uma **acurácia de 62%**.

O meu principal objetivo era identificar os principais fatores que contribuíam para cada tipo de acidente. Com isso, analisar se é possível pensar em campanhas para diminuir o número de acidentes em rodovias ou não.

Por exemplo:

Um dos fatores que aumentam a probabilidade de um acidente com vítima ferida é **o desgaste excessivo do pneu**. Sabendo disso, em dias próximos à finais de semana ou feriados, em que as rodovias são mais movimentadas, os condutores poderiam receber mensagens ou alertas em aplicativos sugerindo uma revisão e conscientizando sobre a importância da estabilidade e segurança do automóvel.

Por fim, o modelo consegue classificar os acidentes razoavelmente bem, e a partir disso pode-se identificar as principais variáveis de cada tipo de acidente.



Obrigada!



- Anexo -

Descrição das variáveis

Variável	Descrição
id	Identificador do acidente.
data_inversa	Data da ocorrência no formato dd/mm/aaaa.
dia_semana	Dia da semana da ocorrência
horario	Horário da ocorrência no formato hh:mm:ss.
uf	Unidade da Federação. Ex.: MG, PE, DF, etc.
br	Variável com valores numéricos representando o identificador da BR do acidente.
municipio	Nome do município de ocorrência do acidente.
causa_acidente	Identificação da causa presumível do acidente. Ex.: Falta de atenção, Velocidade incompatível, etc.
tipo_acidente	Identificação do tipo de acidente. Ex.: Colisão frontal, Saída de pista, etc.
classificação_acidente	Classificação quanto à gravidade do acidente: Sem Vítimas, Com Vítimas Feridas, Com Vítimas Fatais e Ignorado.
fase_dia	Fase do dia no momento do acidente. Ex. Amanhecer, Pleno dia, etc.
sentido_via	Sentido da via considerando o ponto de colisão: Crescente e decrescente.
condição_meteorologica	Condição meteorológica no momento do acidente: Céu claro, chuva, vento, etc.
tipo_pista	Tipo da pista considerando a quantidade de faixas: Dupla, simples ou múltipla.
tracado_via	Descrição do traçado da via: reta, curva ou cruzamento.
uso_solo	Descrição sobre as características do local do acidente: Urbano ou rural.
peessoas	Total de pessoas envolvidas na ocorrência.
mortos	Total de pessoas mortas envolvidas na ocorrência.
feridos_leves	Total de pessoas com ferimentos leves envolvidas na ocorrência.
feridos_graves	Total de pessoas com ferimentos graves envolvidas na ocorrência.
llesos	Total de pessoas ilesas envolvidas na ocorrência.
Ignorados	Total de pessoas envolvidas na ocorrência que não se soube o estado físico.
feridos	Total de pessoas feridas envolvidas na ocorrência (é a soma dos feridos leves com os graves).
veiculos	Total de veículos envolvidos na ocorrência.

