

Aula 2

Vamos lá?

#FADJ2020

#ANALISTADEDADOSJÚNIOR2020



FORMAÇÃO

ANALISTA DE DADOS JUNIOR

JANEIRO / 2020

Realização:



Patrocínio:



Apoio:



Organização:





DADOS
RECIFE

(PREVER)
**ACIDENTES
GRAVES**

Equipe:

Cloves Rocha (car2@cin.ufpe.br)

[Dados anonimizados — LGPD.](#)

[Dados anonimizados — LGPD](#)

[Dados anonimizados — LGPD](#)

CONTEXTO

Os acidentes de trânsito com vítima têm um alto custo para a sociedade.

Fatores sociais:

- ? Perda da qualidade de vida devido a ferimentos ou mutilações; e**
- ? A perda temporária ou definitiva para a família e comunidade do indivíduo.**

CONTEXTO

Fatores econômicos (Custos):

- ? Perda de produtividade devido a indisponibilidade da vítima;
- ? Os custos de hospitalização e tratamento médico;
- ? Custos de reparo dos veículos; e
- ? Possíveis investigações criminais;

Somados, esses custos podem chegar a mais de 1% do produto interno bruto (PIB) de um país.

OMS estimava em 2004 perdas anuais > US\$ 500 bilhões de dólares.

OBJETIVO

Prever acidentes graves, mobilizando recursos para áreas e eventos de maior probabilidade.

Específicos:

- ? Identificar quais fatores mais influenciam na ocorrência de acidentes graves.
- ? Tentar antecipar a gravidade de um acidente devido às condições em que ocorreu.

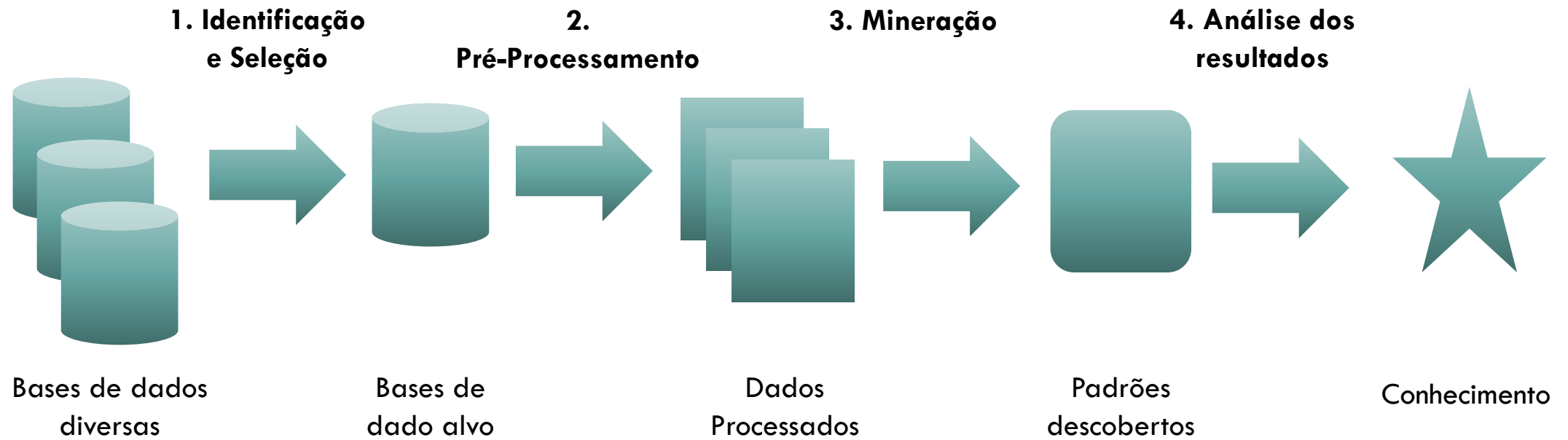
Acidentes graves: Acidentes com vítimas fatais e/ou que envolveu motocicletas.

DIAGRAMA DO FLUXO DECISÓRIO

Gestor Público:

- ? Realizar priorização e/ou identificação de pontos de alocação de recurso (financeiro, físico e/ou de qualquer outra natureza);
- ? Identificar os locais com maior incidência de acidentes graves, com o objetivo de prever;

METODOLOGIA



ENTENDIMENTO DOS DADOS



DADOS
RECIFE



Fontes de dados:

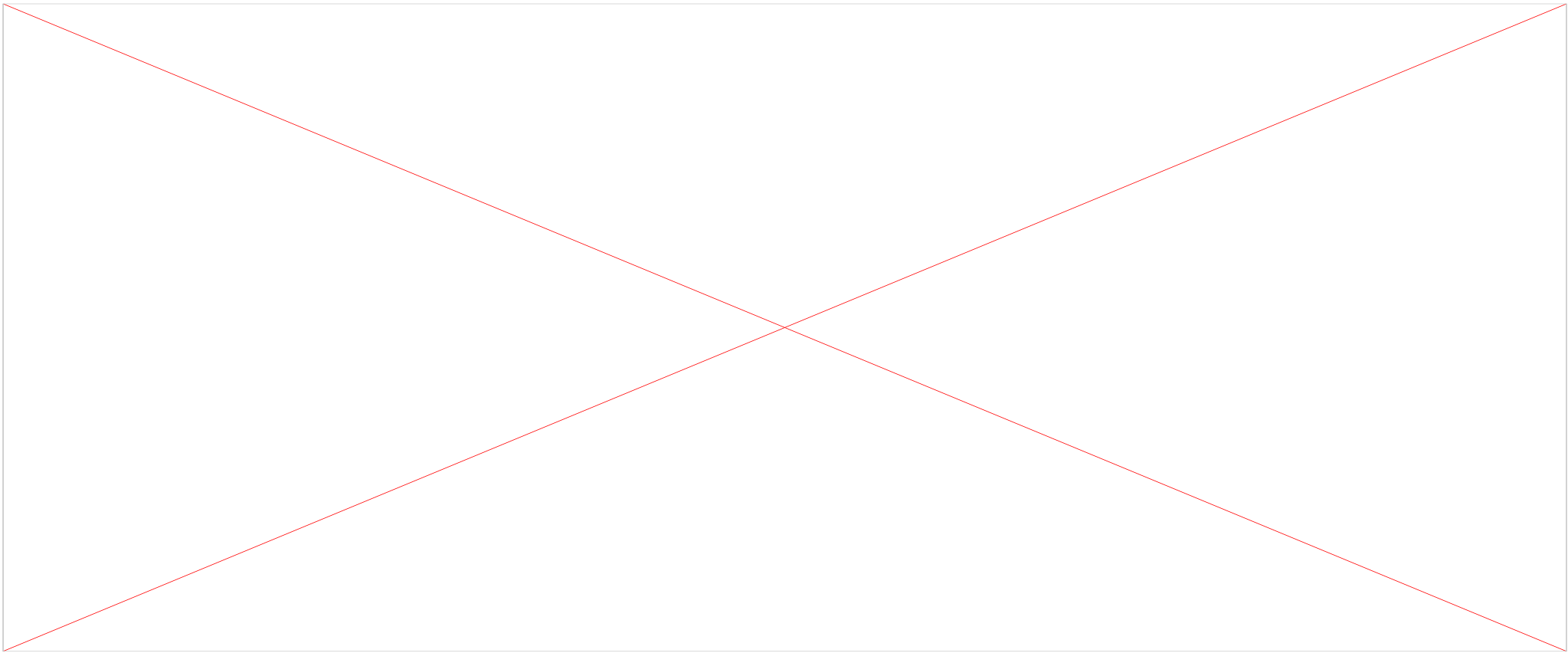
Base de Dados	Disponibilizado em
Acidentes de trânsito com e sem vítimas	dados.recife.pe.gov.br/dataset/acidentes-de-transito-com-e-sem-vitimas
Central de Atendimento de Serviços da Emlurb – 156	dados.recife.pe.gov.br/dataset/central-de-atendimento-de-servicos-da-emlurb-156
Equipamentos de Monitoramento e Fiscalização de Trânsito	dados.recife.pe.gov.br/dataset/equipamentos-de-monitoramento-e-fiscalizacao-de-transito
Registro de Velocidade das Vias	dados.recife.pe.gov.br/dataset/registro-de-velocidade-das-vias
Semáforos do Recife	dados.recife.pe.gov.br/dataset/localizacao-dos-semaforos/resource/ab6343e9-c3f2-4d62-9554-5778f9f33738
Monitoramento Pluviométrico	www.apac.pe.gov.br/meteorologia/monitoramento-pluvio.php

ENTENDIMENTO DOS DADOS

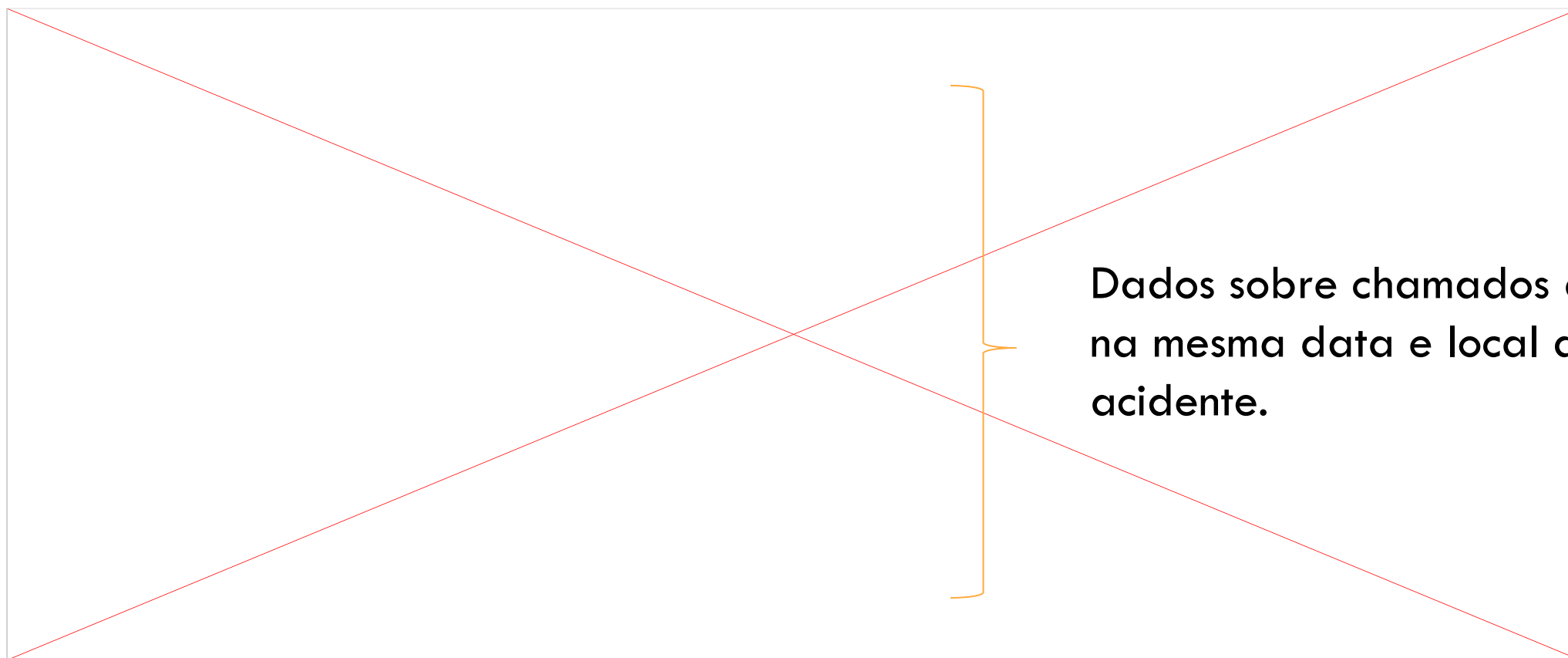
Pré-processamento:

1. Consolidar os dados de acidentes com e sem vítimas;
2. Buscar informações da EMLURB de chamados abertos na mesma data e local do acidente;
3. Buscar os equipamentos de monitoramento de trânsito nos locais dos acidentes;
4. Buscar históricos de velocidade a cada 10 km/h na mesma data e local do acidente;
5. Identificar a quantidade de semáforos nos locais dos acidentes;
6. Inserir informações sobre a data (Se é feriado e/ou próximo de algum feriado);
7. Inserir informações sobre a média pluviométrica (APAC);

ENTENDIMENTO DOS DADOS



ENTENDIMENTO DOS DADOS



ENTENDIMENTO DOS DADOS

Dados sobre os **Fotossensores**:

? Fotossensor – Usado para identificar, posteriormente, as informações sobre velocidades – Não foi considerado para análise;

? possuiFotossensor (Categorical) – Feature;

? qtd_0a10km (Categorical) – Feature;

? qtd_11a20km (Categorical) – Feature;

? qtd_21a30km (Categorical) – Feature;

? qtd_31a40km (Categorical) – Feature;

? qtd_41a50km (Categorical) – Feature;

? qtd_51a60km (Categorical) – Feature;

? qtd_61a70km (Categorical) – Feature;

? qtd_71a80km (Categorical) – Feature;

? qtd_81a90km (Categorical) – Feature;

? qtd_91a100km (Categorical) – Feature;

? qtd_acimade100km(Categorical) – Feature;

Quantidade de carros que passaram na mesma data e local até a hora do acidente, separando de 10 em 10 km/h.

ENTENDIMENTO DOS DADOS

Dados sobre os **Semáforos**:

- ? Semaforo (Categorical) – Feature (Se possui semáforo ou não);
- ? qntdSemaforos (Categorical) – Feature (A quantidade de semáforos na rua);

Dados sobre a média pluviométrica (**APAC**):

- ? MediaPluviometrica (Categorical) – Feature (A média pluviométrica, proveniente da APAC. Média calculada para Recife);

Dados sobre **Feriados**:

- ? EFeriado (Categorical) – Feature (Verificação se a data do acidente caiu em algum feriado);
- ? ProximoAFeriado (Categorical) – Feature (Verificação se a data do acidente está próxima a algum feriado);

ENTENDIMENTO DOS DADOS

DataSet Final:

- ? Total de 39.381 linhas;
- ? 9483 acidentes potencialmente graves (24%);

ANÁLISE

1. Correlação entre as variáveis
2. Árvore de Decisão
3. Indução de Regras
4. Regressão Logística

ANÁLISE

1. Correlação entre as variáveis
2. Árvore de Decisão
3. Indução de Regras
4. Regressão Logística

CORRELAÇÕES ENTRE AS VARIÁVEIS

Nota-se que as variáveis relacionadas com as velocidades (de 10 em 10 km/h) aparecem na maioria das correlações.

Foi possível notar também que algumas variáveis relacionada com a média pluviométrica e/ou a quantidade de veículos no acidente também aparecem com diversas relações.

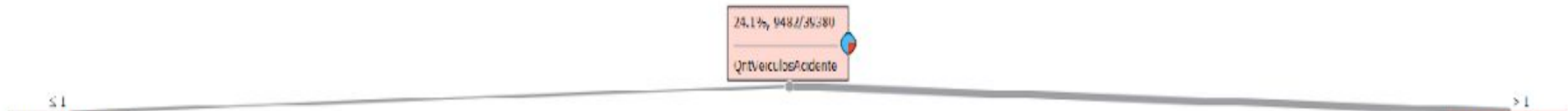
Por fim, a quantidade de semáforos aparece, também, algumas vezes.

+0.050	qtd_21a30km	qtd_acimade100km
+0.049	qtd_31a40km	qtd_acimade100km
+0.047	qtd_61a70km	qtd_acimade100km
+0.039	qtd_51a60km	qtd_acimade100km
+0.033	qtd_41a50km	qtd_acimade100km
+0.031	qntdSemaforos	qtd_acimade100km
+0.027	QntVeiculosAcidente	qntdSemaforos
+0.026	qntdSemaforos	qtd_71a80km
+0.025	qtd_11a20km	qtd_acimade100km
+0.021	MediaPluviometrica	qtd_71a80km
-0.019	MediaPluviometrica	QntVeiculosAcidente
+0.019	QntVeiculosAcidente	qtd_41a50km
+0.018	QntVeiculosAcidente	qtd_31a40km
+0.016	QntVeiculosAcidente	qtd_11a20km
+0.015	MediaPluviometrica	qtd_0a10km
+0.015	QntVeiculosAcidente	qtd_51a60km
+0.015	QntVeiculosAcidente	qtd_21a30km
+0.014	qntdSemaforos	qtd_81a90km
+0.014	MediaPluviometrica	qtd_61a70km
+0.013	MediaPluviometrica	qtd_11a20km
+0.012	QntVeiculosAcidente	qtd_61a70km
+0.012	MediaPluviometrica	qtd_81a90km
+0.011	QntVeiculosAcidente	qtd_0a10km
+0.010	MediaPluviometrica	qtd_51a60km
+0.010	MediaPluviometrica	qtd_31a40km
-0.009	MediaPluviometrica	qntdSemaforos
+0.008	MediaPluviometrica	qtd_41a50km
+0.007	QntVeiculosAcidente	qtd_71a80km
+0.006	MediaPluviometrica	qtd_21a30km
-0.005	QntVeiculosAcidente	qtd_acimade100km
+0.004	QntVeiculosAcidente	qtd_81a90km
-0.004	MediaPluviometrica	qtd_acimade100km

ANÁLISE

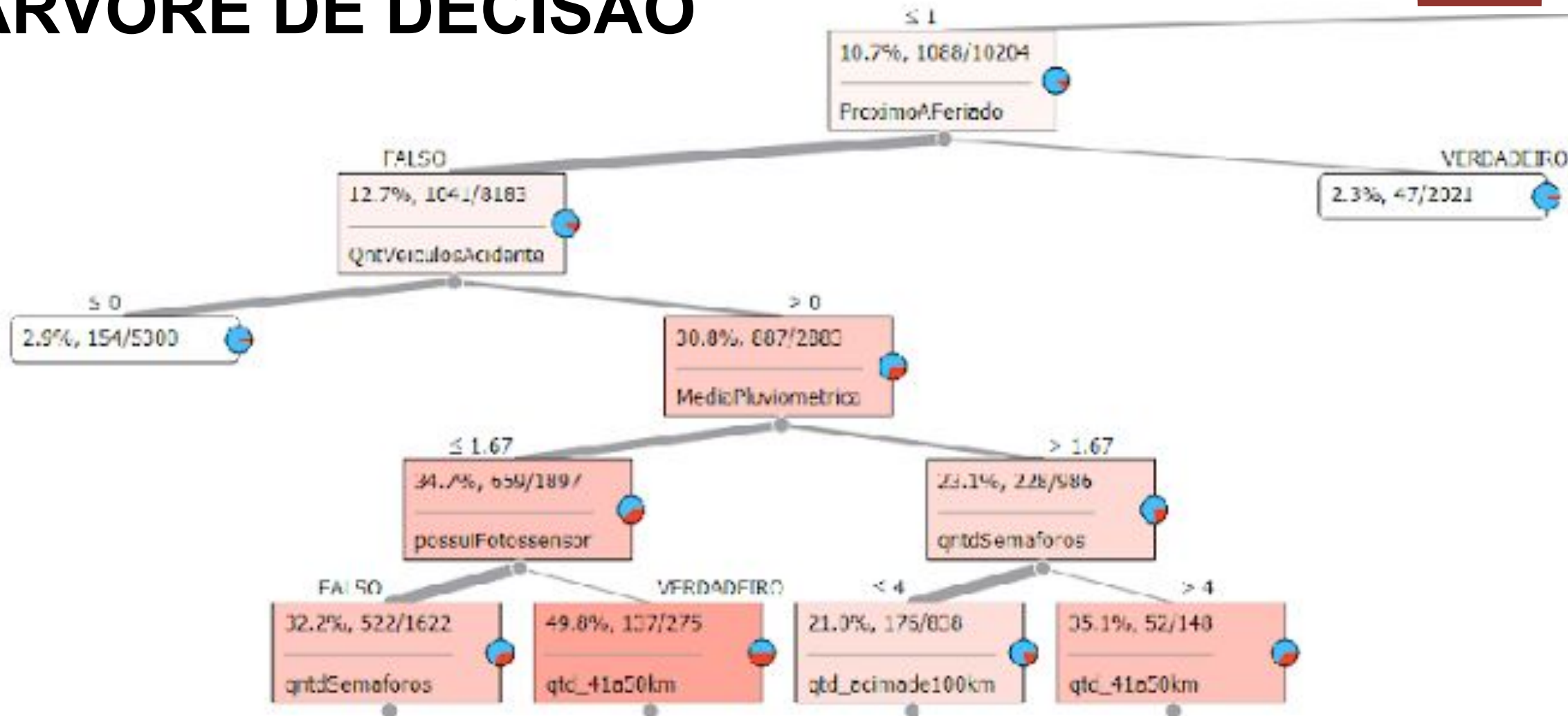
1. Correlação entre as variáveis
2. Árvore de Decisão
3. Indução de Regras
4. Regressão Logística

ÁRVORE DE DECISÃO



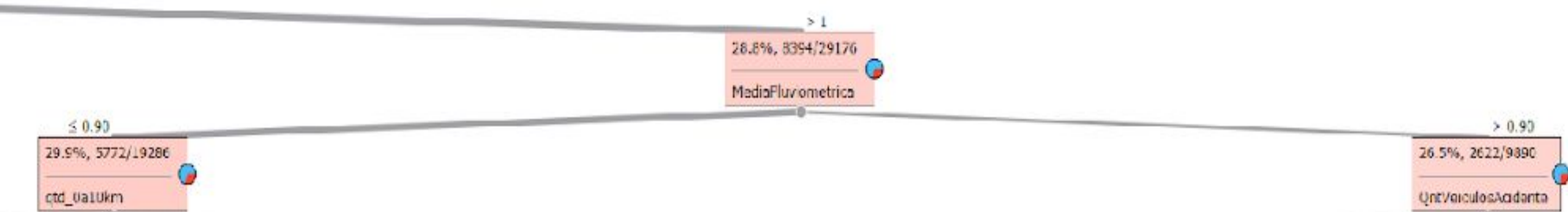
24% da base de dados possui acidentes graves (de um total de 39.381 registros).

ÁRVORE DE DECISÃO



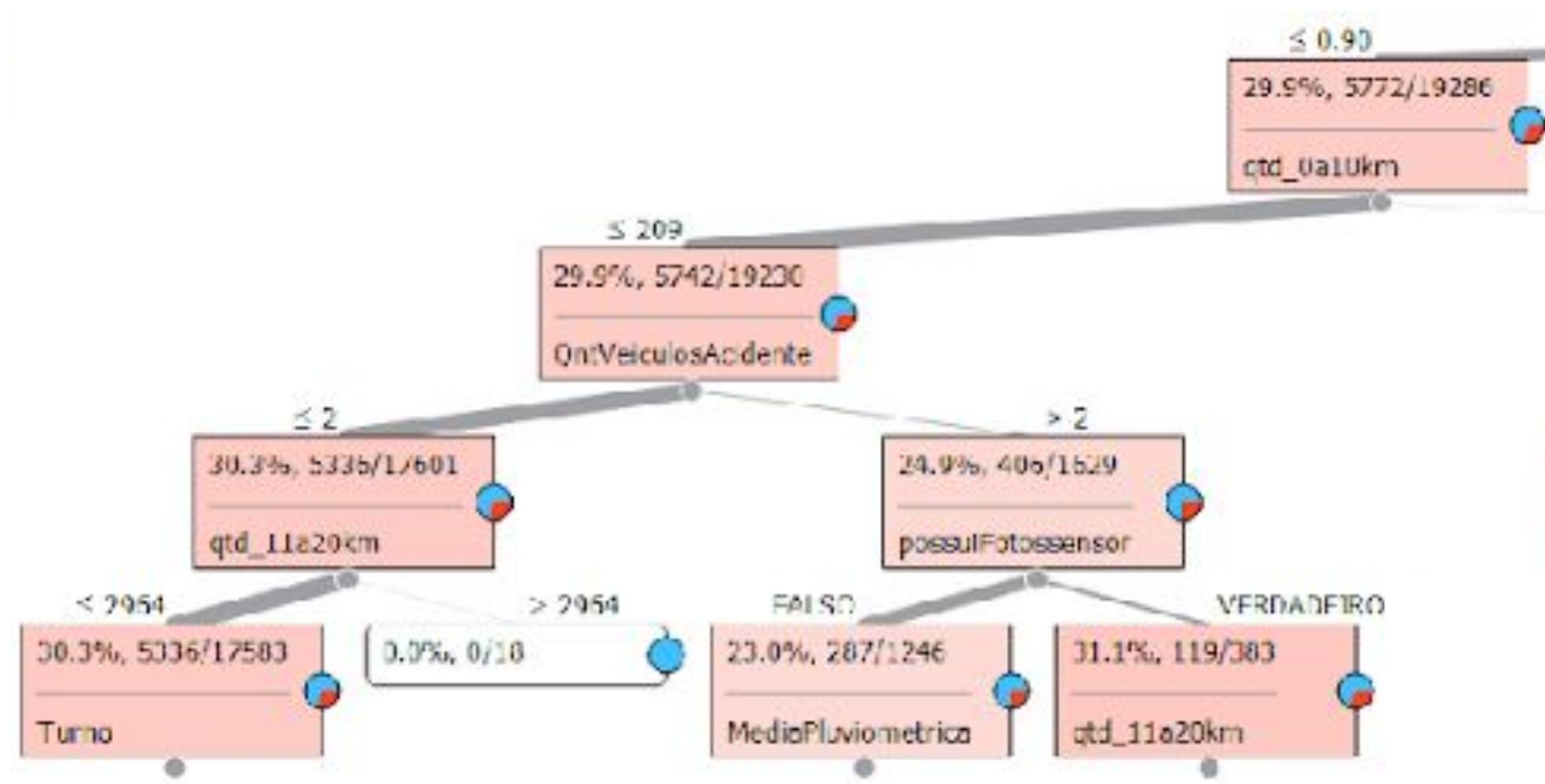
Quando não há proximidade de feriado, o aumento de veículos envolvidos no acidente, e o aumento da média pluviométrica incidem em pontos de atenção para a ocorrência de acidentes graves. Velocidades mais altas e a quantidade de semáforos também podem ser decisivos.

ÁRVORE DE DECISÃO



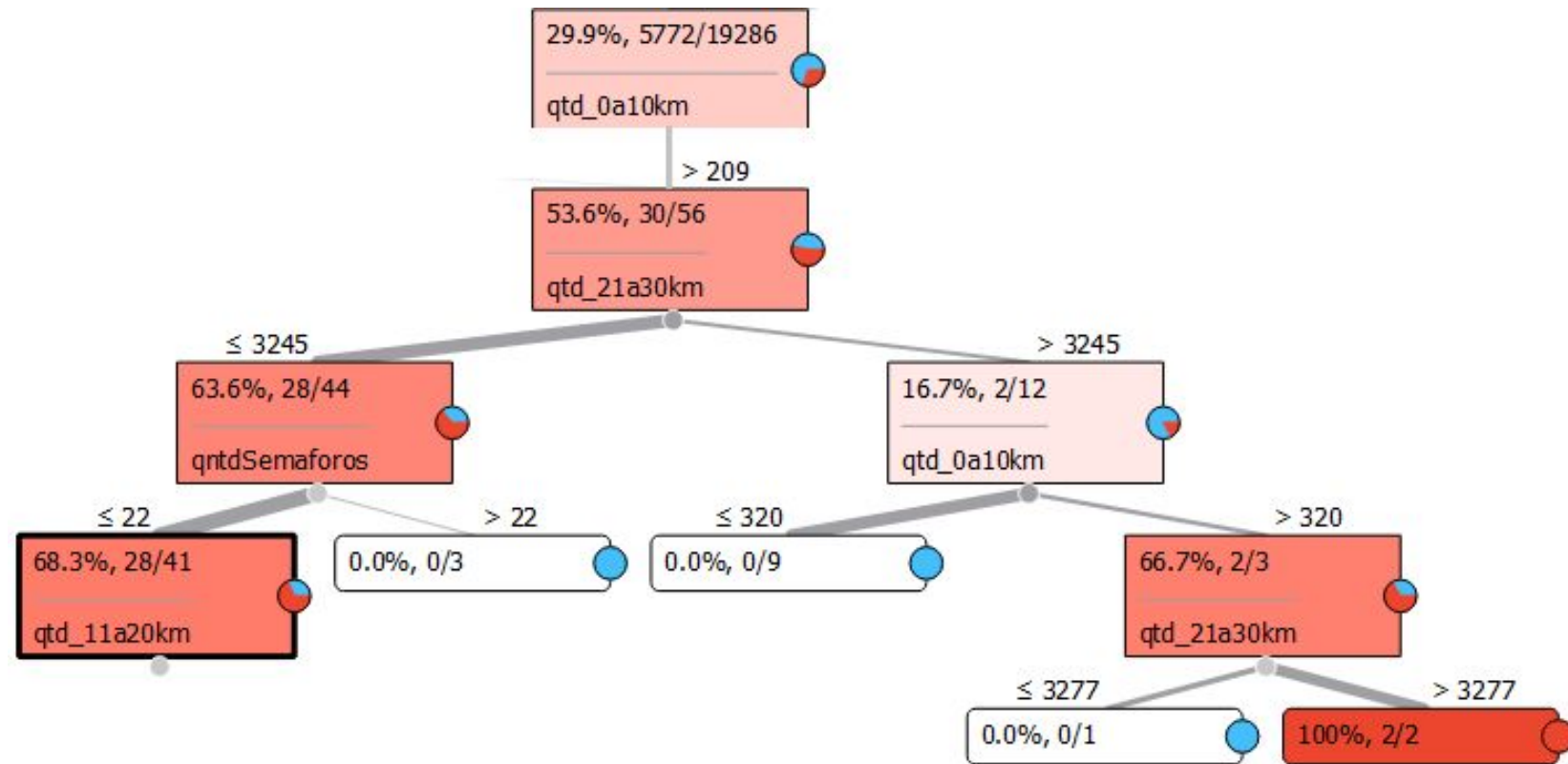
Quanto maior o número de veículos envolvidos no acidente, maior a chance do acidente ser grave.

ÁRVORE DE DECISÃO



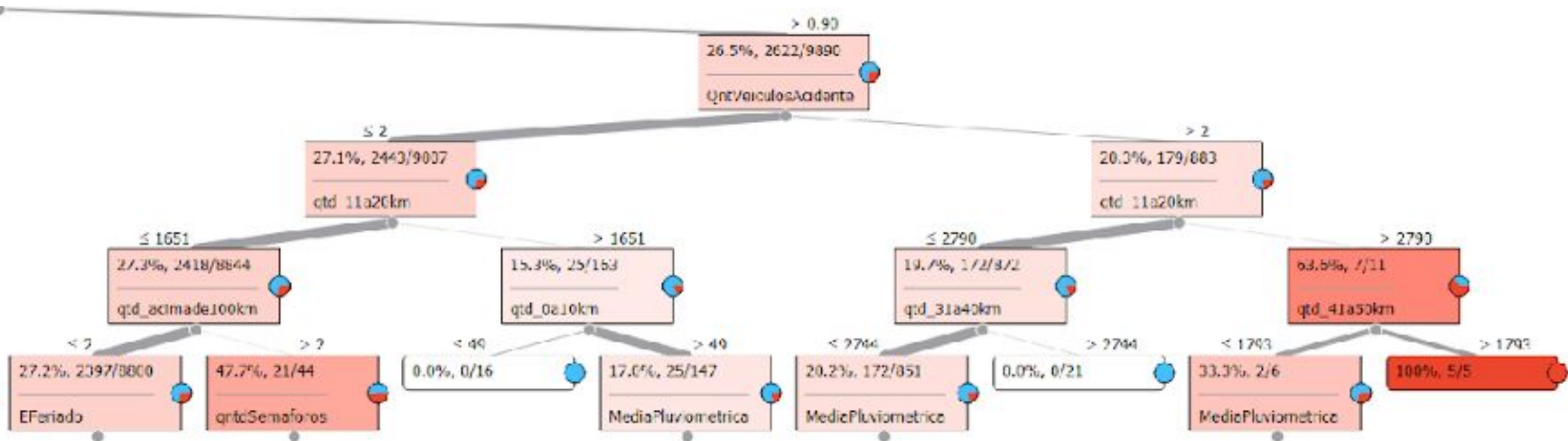
Quando há uma menor incidência de chuva (menos média pluviométrica), há uma chance maior de acidente potencialmente grave quando o turno é noturno, e há também uma diferença de veículos parados e outros com deslocamento (ainda que leve).

ÁRVORE DE DECISÃO



Por outro lado, quando há uma combinação de veículos lentos, outros em velocidades superiores, pode ocorrer uma incidência maior de acidentes graves. Destacando que, neste caso, a quantidade de semáforos na via influencia (provavelmente em casos de vias longas, com grande engarrafamento, e veículos mais rápidos).

ÁRVORE DE DECISÃO



Nos casos de uma maior incidência de chuva (média pluviométrica maior), a quantidade de veículos envolvidos no acidente também influenciou. Neste caso, a diferença de velocidades entre os veículos mostrou uma grande influencia. Destaque para os casos em que há carros lentos, e outros acima de 100 km/h e/ou entre 41 a 50 km/h.

ANÁLISE

1. Correlação entre as variáveis
2. Árvore de Decisão
3. Indução de Regras
4. Regressão Logística

INDUÇÃO DE REGRAS

CONDIÇÃO 1	CONDIÇÃO 2	CONDIÇÃO 3	CONDIÇÃO 4	COBERTURA	CONFIANÇA	LIFT
Quantidade de veículos entre 31 a 40 km/h \leq 37	Turno matutino	Quantidade de veículos envolvidos \geq 1	Quantidade de semáforos na rua \leq 19	0,1%	61,0%	2,53
Próximo a feriado = FALSO	Possui fotossensor = FALSO	Quantidade de veículos envolvidos \geq 1		0,1%	59,5%	2,47
Turno noturno	Chamado aberto para iluminação pública = FALSO	Quantidade de veículos envolvidos \geq 1		0,2%	58,6%	2,43
Turno noturno				0,1%	55,9%	2,32
Turno matutino				0,7%	50,0%	2,08

ANÁLISE

1. Correlação entre as variáveis
2. Árvore de Decisão
3. Indução de Regras
4. Regressão Logística

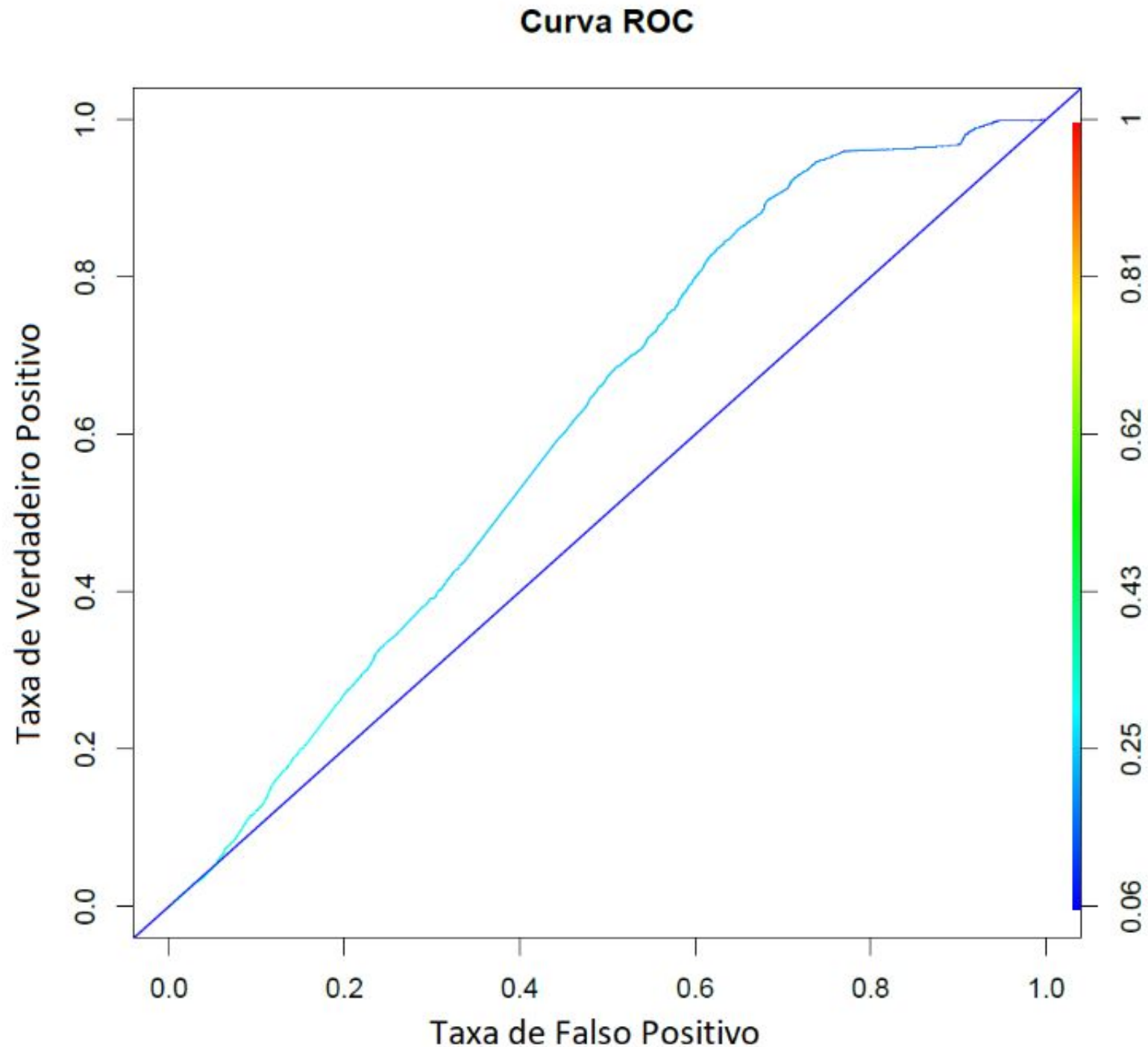
REGRESSÃO LOGÍSTICA

Avaliação

Method	AUROC	CA	F1	Precision	Recall	LogLoss
Neural Network	0.651	0.757	0.664	0.681	0.757	0.514
Logistic Regression	0.614	0.757	0.656	0.628	0.757	0.53

REGRESSÃO LOGÍSTICA

Curva ROC

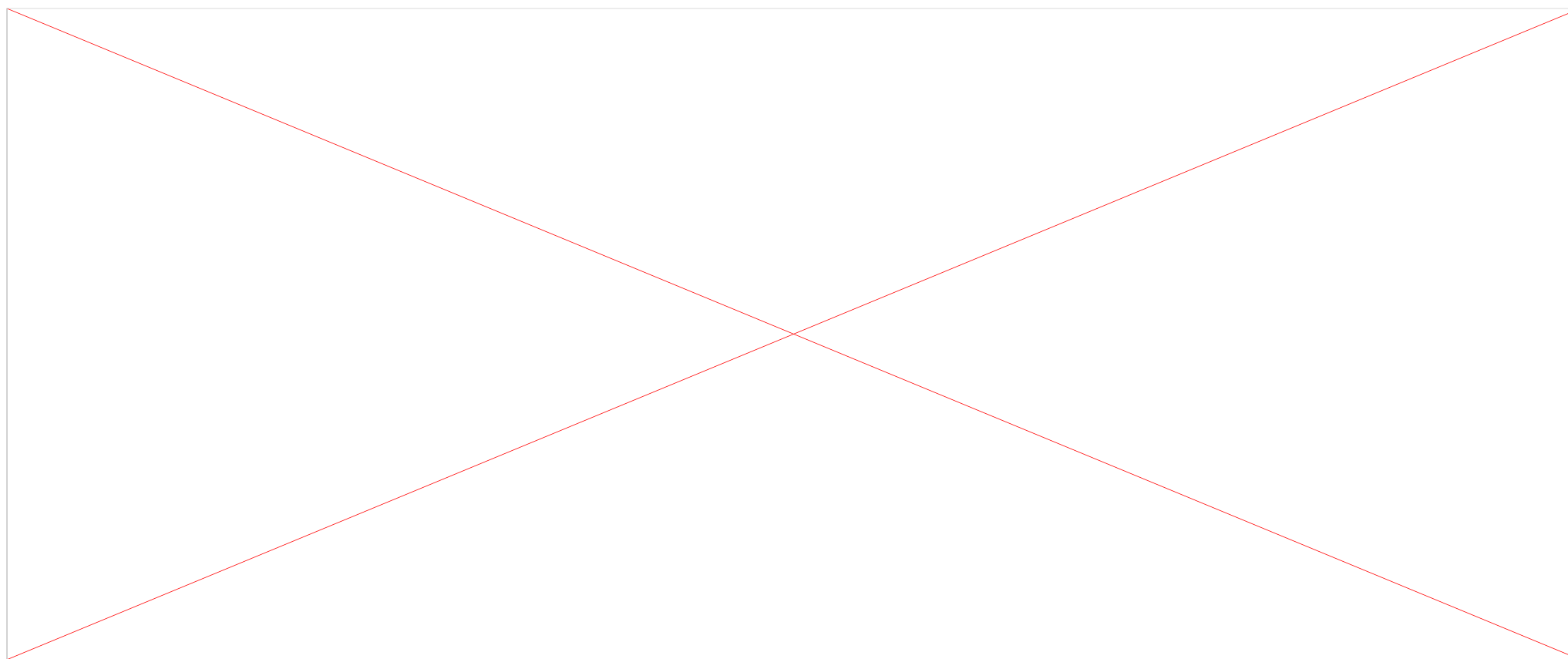


REGRESSÃO LOGÍSTICA

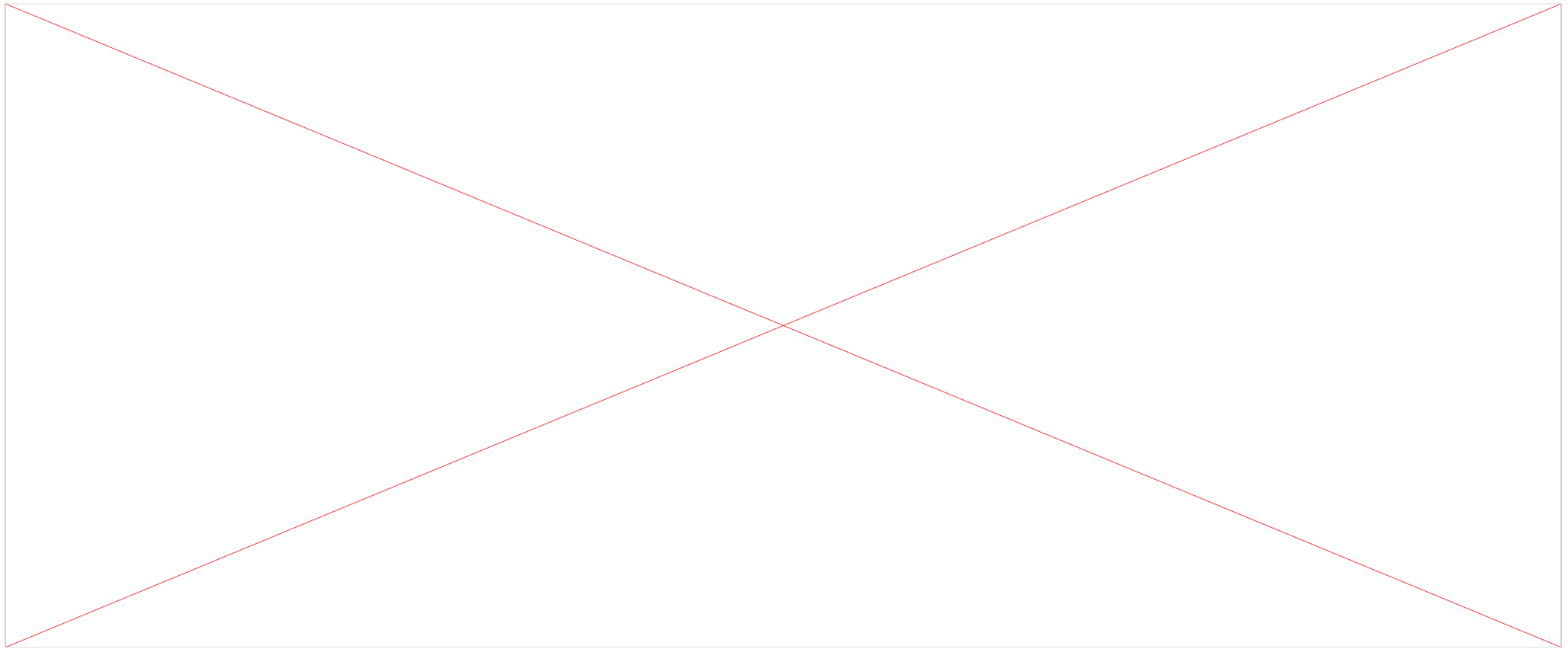
Variáveis mais relevantes:

Feature	Coeficiente Beta	P - valor	Tipo
Turno = Noturno	0.374	0.000	Categórica
Chamado de Drenagem aberto	0.364	0.004	Binária
Rua com Fotossensor	0.313	0.000	Binária
Quantidade de Veículos Envolvidos	0.285	0.000	Numérica
Turno = Manhã	0.240	0.000	Categórica
Possui Semáforo	0.178	0.000	Binária
Data próxima a feriado	-0.675	0.000	Binária

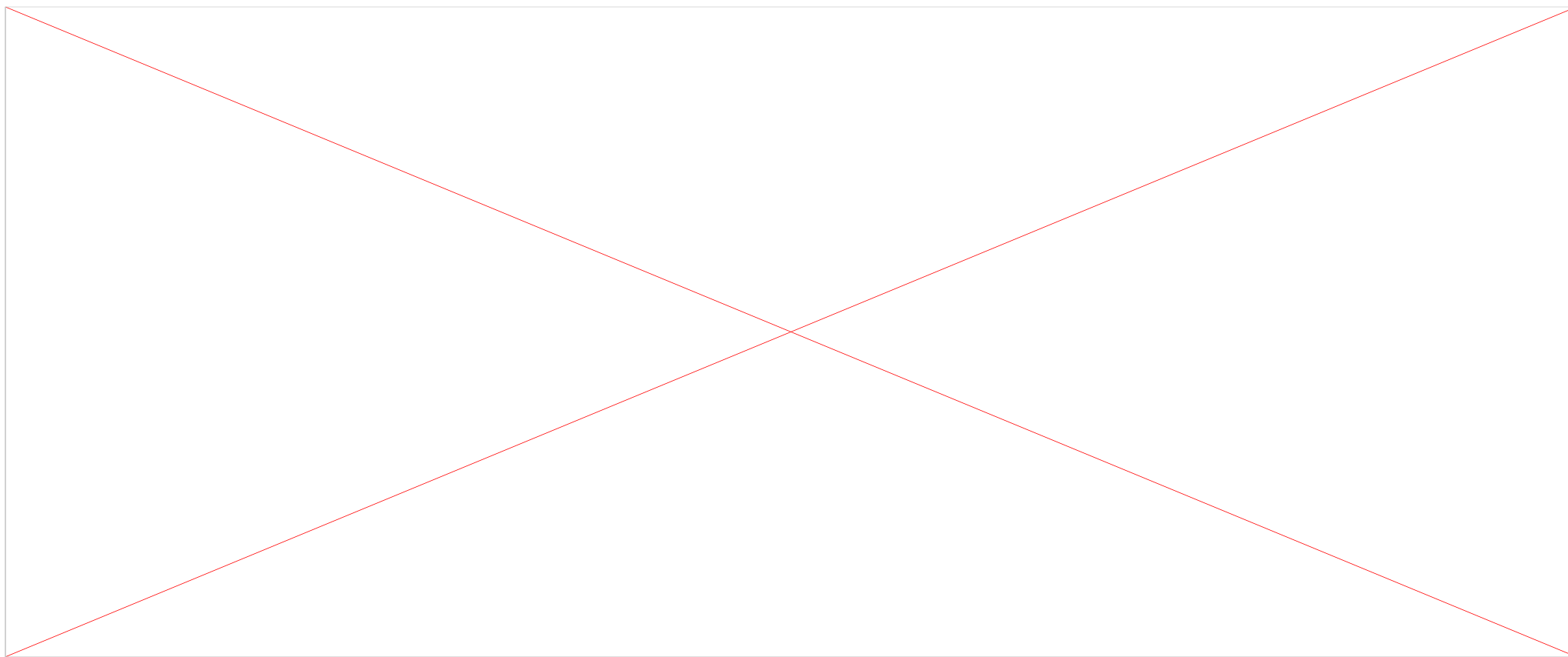
DISCUSSÃO SOBRE OS RESULTADOS



DISCUSSÃO SOBRE OS RESULTADOS



AMEAÇAS À VALIDADE



CONCLUSÕES E TRABALHOS FUTUROS

- ? Quanto mais veículos envolvidos no acidente, maior a influência na possibilidade do acidente ter sido grave.
 - ? Existe a necessidade de um atendimento mais rápido por parte de órgãos relacionados ao atendimento hospitalar, a exemplo do SAMU (Serviço de Atendimento Médico de Urgência)
- ? A quantidade de sinais e foto-sensores nas vias podem ter influência para a diminuição da gravidade dos acidentes.
- ? Grande concentração de carros em uma velocidade baixa, enquanto outros vários carros estão em velocidades maiores aumenta a chance de acidente grave.

CONCLUSÕES E TRABALHOS FUTUROS

Possíveis recomendações:

- i. Aumento de foto-sensores, buscando uma maior quantidade de informações em tempo real;
- ii. Instalação de painéis de alerta para alertar em trechos com concentração de carros em diferentes faixas de velocidade (principalmente com uma grande diferença);
- iii. Estudo para a distribuição dos sinais de trânsito, ao longo das vias da cidade e sua região metropolitana.

Dúvidas?

DESAFIO

Introdução
Alfabetização dos Dados

Pesquise dados abertos sobre acidentes graves na cidade de São Paulo...

- Apresente 5 causas de acidentes graves;
- Custos hospitalares relacionados;
- Duração de 30min.



Muito Obrigado!
Thank you!

#ANALISTADEDADOSJÚNIOR2020



FORMAÇÃO
**ANALISTA
DE DADOS
JUNIOR**
JANEIRO / 2020

Realização:



Patrocínio:



Apoio:



Organização:

