

Sinistros de Trânsito no Brasil (2007)

- Rafael Augusto de Souza
- Luiz Guilherme Manfrim

Contexto

- Os acidentes de trânsito são uma das maiores causas de mortes no Brasil.
- Impactos sociais, econômicos e no sistema de saúde.

Perguntas de Pesquisa

1. Quais fatores (como horário, tipo de pista, condição meteorológica e uso do solo) estão mais associados à ocorrência de acidentes graves nas rodovias federais em 2007?
2. É possível prever a gravidade de um acidente de trânsito com base nas características da ocorrência (condições climáticas, tipo de via, horário e localização)?

Hipóteses Testáveis

1. H1: Fatores ambientais e temporais influenciam a gravidade do sinistro
2. H2: O número de veículos envolvidos e o tipo de pista são preditores significativos da gravidade
3. H3: Modelos de ensemble apresentam desempenho superior na predição da gravidade do acidente

Dados Utilizados

- Base de sinistros 2007 + População IBGE 2006
- Mais de 127 mil registros.

	id	data_inversa	dia_semana	horario	uf	br	km	municipio	causa_acidente	tipo_acidente	...	uso_s
0	10	11/06/2007	Segunda	15:30:00	MG	381	623.2	OLIVEIRA	Falta de aten ção	Colis ão frontal	...	Ru
1	10	11/06/2007	Segunda	15:30:00	MG	381	623.2	OLIVEIRA	Falta de aten ção	Colis ão frontal	...	Ru
...	uso_solo	ano	pessoas	mortos	feridos_leves	feridos_graves	ilesos	ignorados	feridos	veiculos	le	...
...	Rural	2007	9	0	4	0	5	0	4	6	ta	...
...	Rural	2007	9	0	4	0	5	0	4	6	co	...
...	Rural	2007	3	0	0	1	2	0	1	1	al	...
...	Urbano	2007	5	2	2	1	0	0	3	1

Limpeza de Dados

- Tratamento de '(null)'
- Padronização de municípios
- Conversão BR/KM
- Remoção de registros inválidos (127675 -> 79811)

Pipeline

- Dados → Limpeza → EDA → Features → Modelos → Avaliação

Feature Engineering

- Faixa horária
- Mês
- Final de semana
- Definição de gravidade

Definição de Gravidade

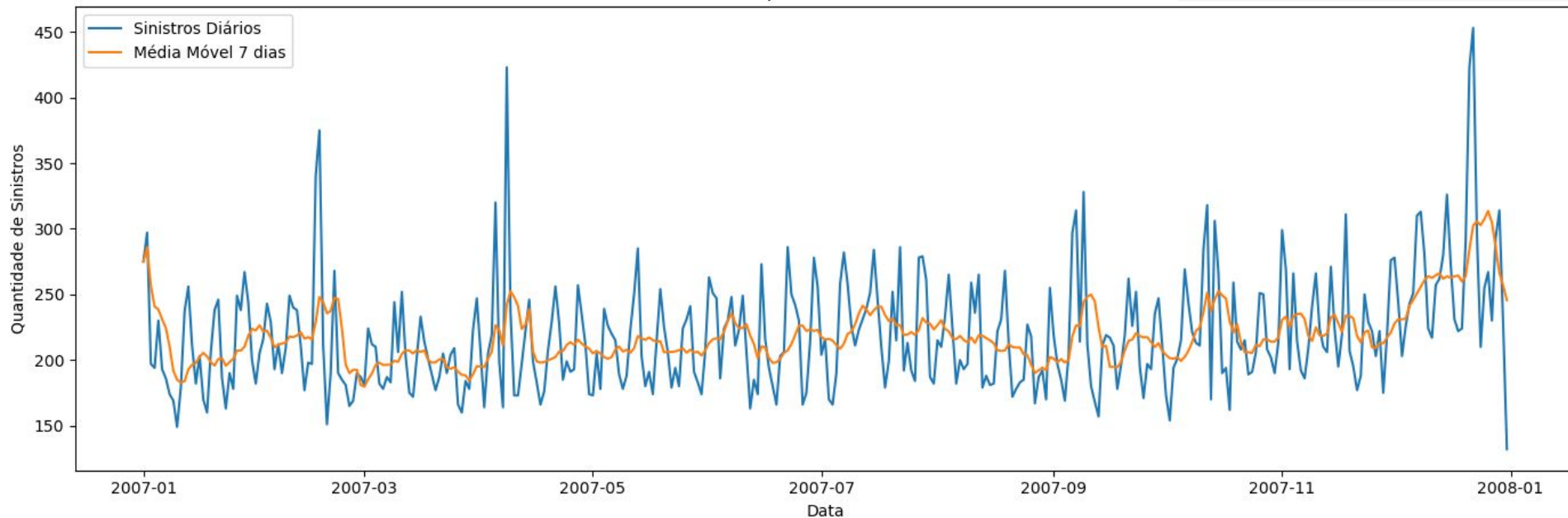
- Grave = mortos > 0 ou feridos > 0

Análise Temporal

- Picos em feriados e início do ano.

	dia	total_sinistros
0	2007-12-22	453
1	2007-12-21	423
2	2007-04-08	423
3	2007-02-17	375
4	2007-02-16	340
5	2007-09-09	328

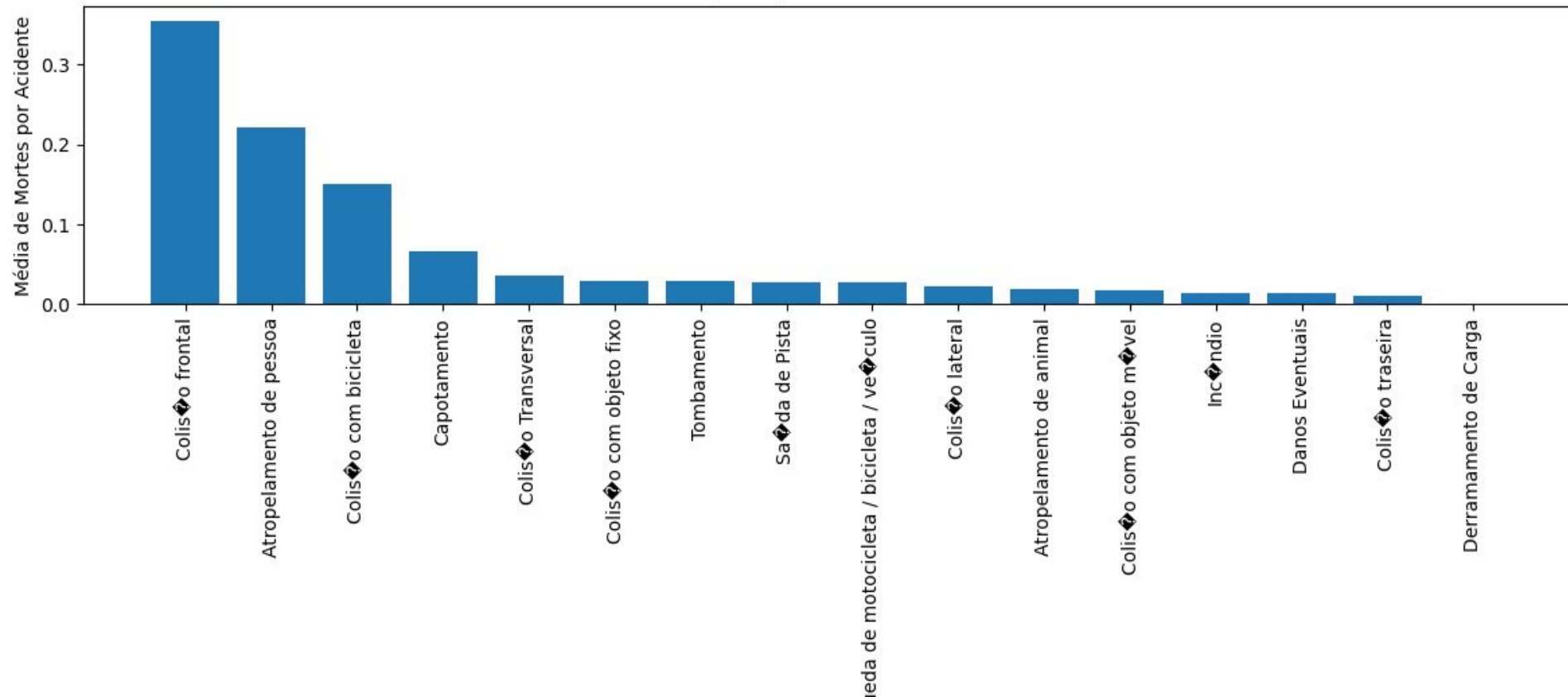
Tendência Temporal de Sinistros em 2007



Tipos de Acidente

- Colisão frontal é a mais letal.

Média de Mortes por Tipo de Acidente – Sinistros 2007



Pareto das Causas

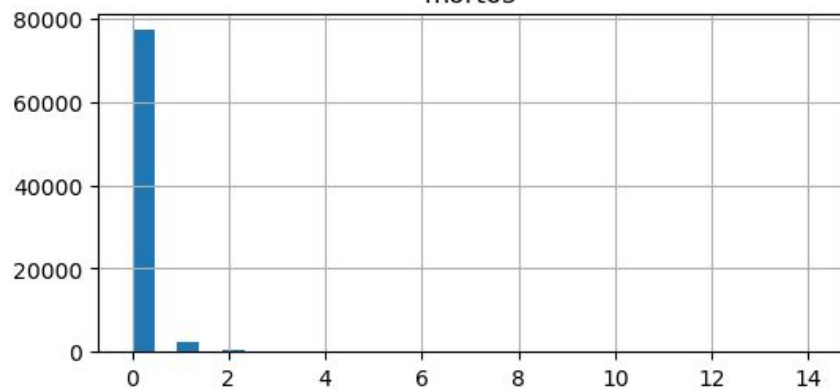
- 80% das mortes vêm de 5 causas principais.

	causa_acidente	mortos	acumulado	perc_acumulado
0	Falta de atenç�o	1436.0	1436.0	0.425
1	Velocidade incompat�vel	513.0	1949.0	0.577
2	Ultrapassagem indevida	494.0	2443.0	0.723
3	Dormindo	214.0	2657.0	0.786
4	Desobedi�ncia � sinaliza�o	209.0	2866.0	0.848
5	Ingest�o de �lcool	172.0	3038.0	0.899
6	Defeito mec�nico em ve�culo	117.0	3155.0	0.933
7	Animais na Pista	96.0	3251.0	0.962
8	N�o guardar dist�ncia de seguran�a	65.0	3316.0	0.981
9	Defeito na via	64.0	3380.0	1.000

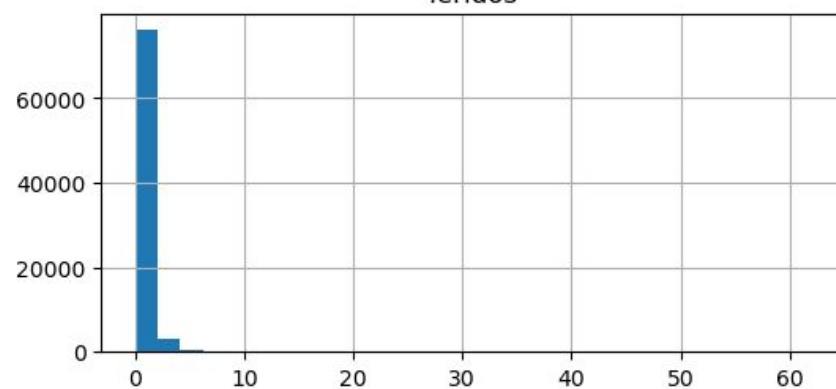
Análise Univariada

Distribuições Univariadas

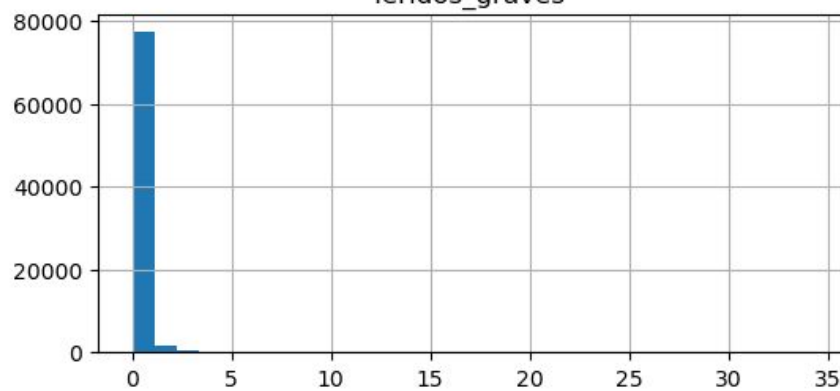
mortos



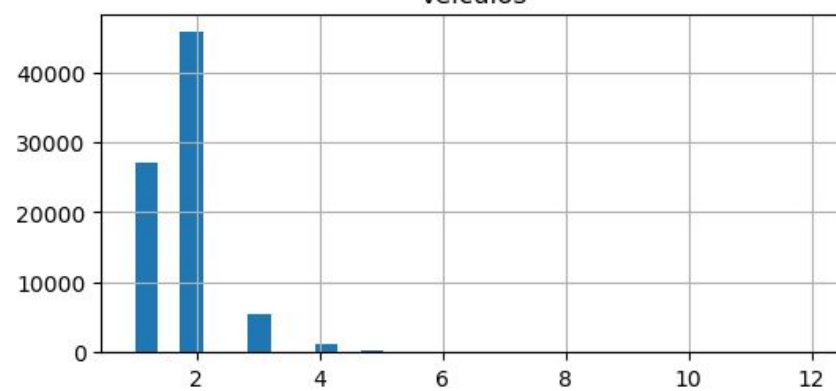
feridos



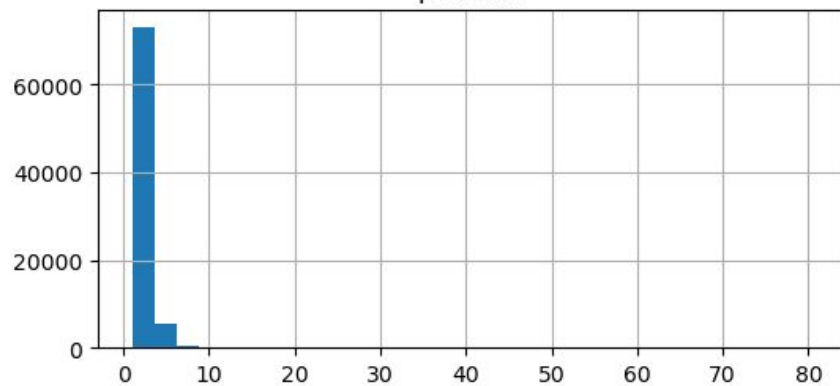
feridos_graves



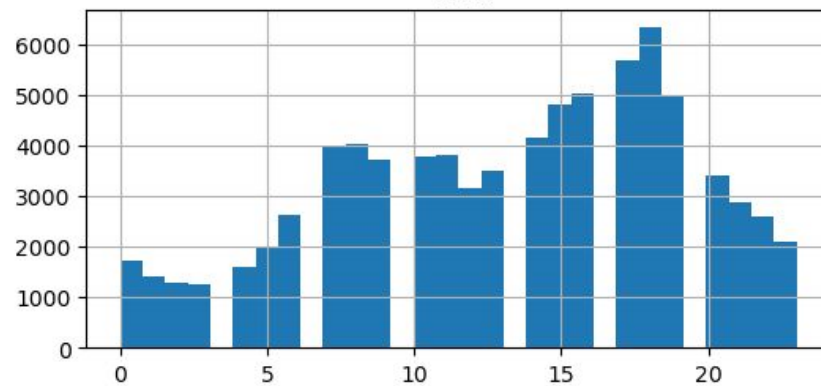
veiculos



pessoas

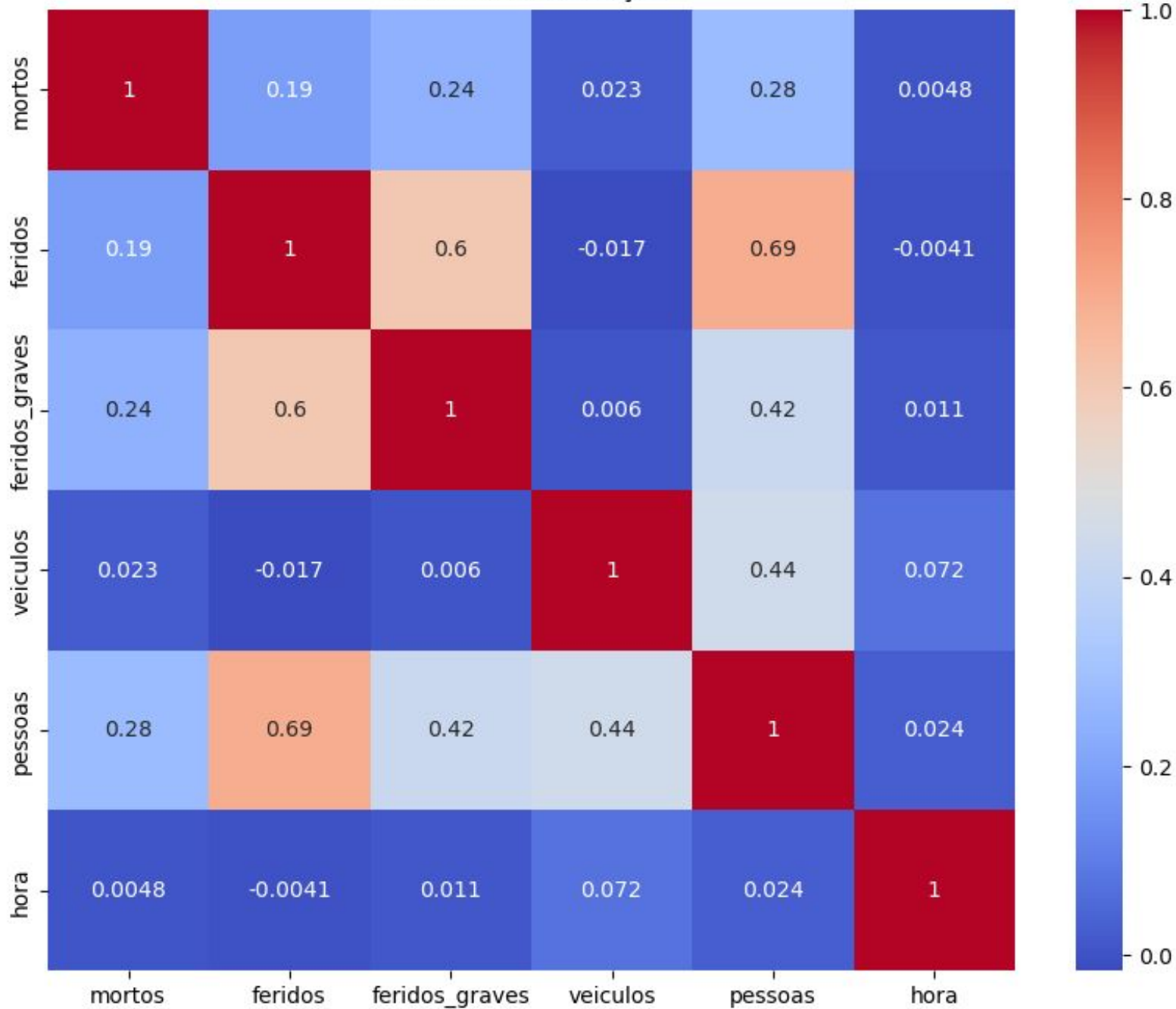


hora

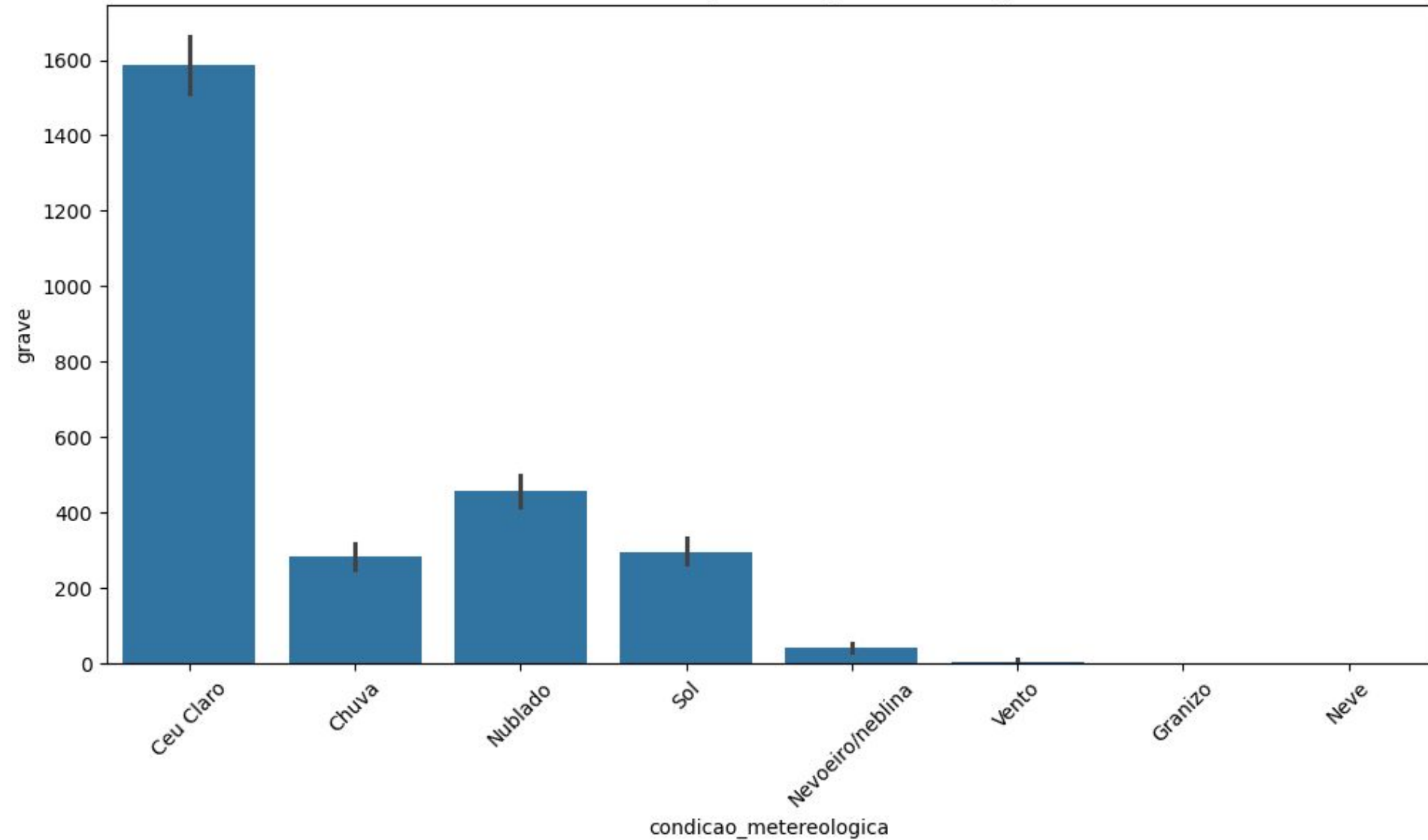


Análise Bivariada

Matriz de Correlação



Acidentes Graves por Condição Meteorológica



Teste de Hipótese

- H1: Fatores ambientais e temporais influenciam a gravidade do sinistro (aceita).
- Análise de variância para o número de mortos e a fase do dia: p-value de $3.8233613031366624e-47$.
- Teste qui-quadrado para a condição meteorológica e a gravidade: p-value de $3.5852705469560646e-17$.

Teste de Hipótese

- H2: O número de veículos envolvidos e o tipo de pista são preditores significativos da gravidade (aceito).
- Regressão logística.
- Tipo de pista simples teve $p = 0.000$.
- Tipo de pista múltipla teve $p = 0.002$.
- Número de veículos teve $p = 0.045$.

Modelos

- Regressão Logística, Random Forest e MLP

Métricas

- F1, Acurácia, Precisão e Recall

	Modelo	F1	Acurácia	Precisão	Recall
0	Regressão Logística	0.382192	0.640105	0.576948	0.285737
1	Random Forest	0.433106	0.640857	0.562404	0.352147
2	LR Reduzido	0.364231	0.634780	0.565910	0.268532
3	RF Reduzido	0.435332	0.630771	0.538516	0.365332
4	MLP Completo	0.415441	0.649001	0.591503	0.320148
5	MLP Reduzido	0.426936	0.648688	0.585646	0.335906
6	Ensemble RF+MLP (Reduzido)	0.537313	0.634906	0.530657	0.544139
7	Ensemble LR+RF+MLP (Reduzido)	0.544674	0.634780	0.529537	0.560701
8	Ensemble LR+RF+MLP (Reduzido, tunado)	0.595025	0.566560	0.467789	0.817334

Importância

- Identificar as principais causas do aumento da mortalidade dos acidentes de trânsito.

Recomendações

- Duplicação de vias.
- Reforço da prontidão de ambulâncias em alguns períodos do ano.

Conclusão

- As nossas hipóteses foram aceitas.
- Dados salvam vidas.