



PUC Minas

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

NÚCLEO DE EDUCAÇÃO A DISTÂNCIA

Pós-graduação *Lato Sensu* em Gestão e Análise Estratégica de Dados

RELATÓRIO TÉCNICO

ESTUDO DE ACIDENTES ENVOLVENDO AUTOMÓVEIS NAS RODOVIAS
FEDERAIS

Mateus Cavalcanti Lira

Belo Horizonte

2021

SUMÁRIO

| | |
|---|-----------|
| 1. Introdução | 4 |
| 1.1. Contexto | 4 |
| 1.2. Objetivos | 4 |
| 1.3. Público alvo | 4 |
| 2. Modelos de Dados | 5 |
| 2.1. Modelo Dimensional | 5 |
| 2.2. Fatos e Dimensões | 5 |
| 3. Integração, Tratamento e Carga de Dados | 7 |
| 3.1. Fontes de Dados..... | 7 |
| 3.2. Processos de Integração e Carga (ETL)..... | 9 |
| 4. Camada de Apresentação | 13 |
| 4.1 Dashboard | 13 |
| 4.2 Análises avançadas | 15 |
| 5. Registros de Homologação | 21 |
| 5. Conclusões | 26 |
| 6. Links | 26 |
| REFERÊNCIAS | 27 |

1. Introdução

1.1. Contexto

No Brasil, o número de acidentes de trânsito é algo que preocupa toda a população. A quantidade de ocorrências, por mais que venha diminuindo, ainda está num patamar de dezenas de milhares de tragédias por ano [3].

Uma ferramenta como essa seria utilizada para analisar quais as principais causas de acidente e criar ações para minimizá-las no próximo ano, fazendo com que se repitam cada vez menos os desastres.

1.2. Objetivos

Este projeto tem como objetivo gerar inteligência através de dashboards criados em ferramentas de visualização de dados, para que analistas da Polícia Rodoviária Federal (PRF) possam atuar de forma eficaz no combate aos acidentes de automóveis nas rodovias federais.

1.3. Público alvo

Em uma visão mais tática e operacional [2], os dashboards e análises gerados pela ferramenta de visualização serão destinados aos próprios analistas da PRF, que poderão tomar decisões a curto e médio prazo sobre que planos de ação criar para remediar as situações avaliadas. Em um âmbito estratégico [2], os dashboard e análises gerados poderiam ser apresentados tanto para a diretoria como equipes fora da organização (nesse caso, outros órgãos do Governo), a fim de explanar o problema e garantir maiores recursos para solucionar os problemas já avaliados.

2. Modelos de Dados

2.1. Modelo Dimensional

O diagrama do modelo dimensional oferece uma visão sobre como os dados serão organizados para atender às necessidades identificadas por meio dos objetivos do projeto. Todas se conectam à tabela **f_acidente** através do **id_acidentes**.

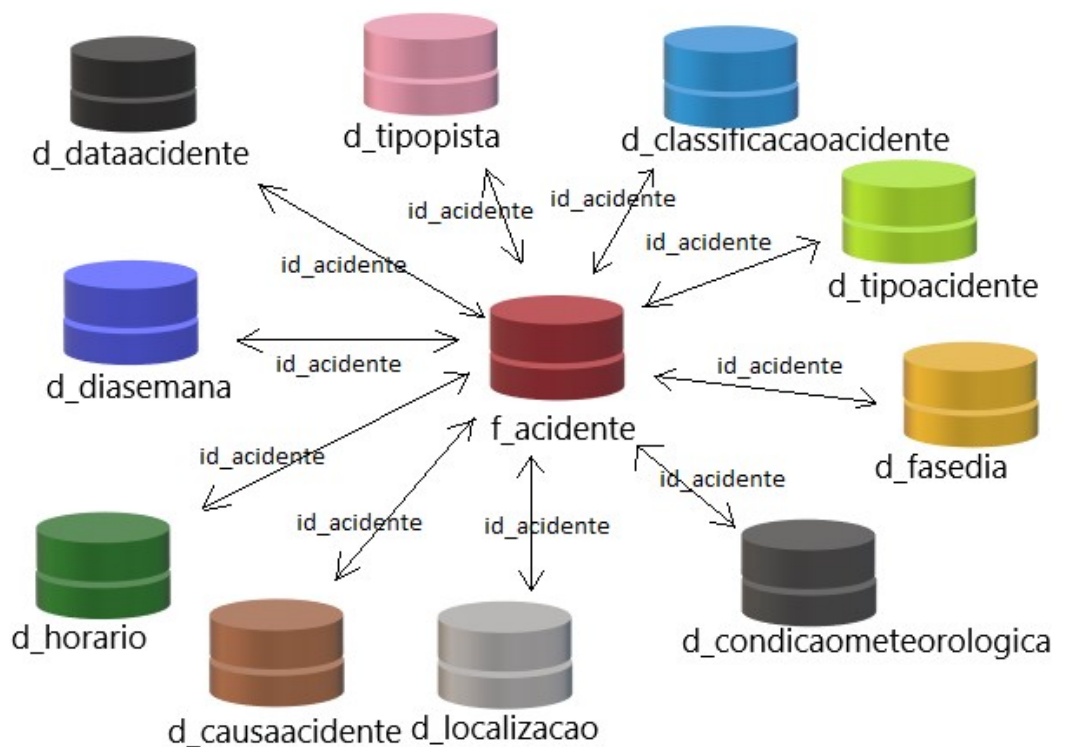


Figura 1: Modelo de dados

2.2. Fatos e Dimensões

| Tipo | Nome | Descrição | Exemplo |
|------|-------------------|-------------------------------------|---------|
| Fato | f_acidente | Define o acontecimento do acidente. | 331693 |

| | | | |
|----------|------------------------|--|---|
| Dimensão | d_dataacidente | tabela onde se encontra a data em que ocorreu o acidente | 01/01/2021 |
| Dimensão | d_diasemana | que dia da semana era quando ocorreu o acidente | sexta-feira, sábado |
| Dimensão | d_horario | em que horário do dia aconteceu o acidente | 21:56:00 |
| Dimensão | d_localizacao | tabela composta por cinco colunas que juntas especificam o local do acidente | <div> <div>d_uf</div> <div>d_br</div> <div>d_km</div> <div>d_latitude</div> <div>d_longitude</div> </div> <div> <div>-</div> <div>-</div> <div>2.843.029.</div> <div>4.884.925.</div> <div>SC</div> <div>101</div> <div>314,</div> <div>392</div> <div>073</div> </div> |
| Dimensão | d_causaacidente | tabela onde está descrita a causa do acidente | Ingestão de álcool pelo condutor Velocidade Incompatível Demais falhas mecânicas ou elétricas Manobra de mudança de faixa Reação tardia ou ineficiente do condutor Velocidade Incompatível Iluminação deficiente Condutor Dormindo Ausência de reação do condutor Etc. |
| Dimensão | d_tipoacidente | destaca qual tipo de acidente | Colisão traseira Colisão com objeto Incêndio Colisão lateral Etc. |

| | | | |
|----------|-------------------------------------|--|---|
| | | aconteceu | |
| Dimensão | d_classificaca oacidente | se houve ou não, vítima | Com Vítimas Feri- das Com Vítimas Fatais Sem Vítimas |
| Dimensão | d_fasedia | em que turno do dia aconteceu o acidente | Amanhecer Plena Noite Pleno dia Anoitecer |
| Dimensão | d_condicaom eteorologica | em que condição de tempo estava no momento do acidente | Nublado Chuva Céu Claro Etc |
| Dimensão | d_tipopista | qual tipo da pista onde ocorreu o acidente | Múltipla Simples Dupla |

3. Integração, Tratamento e Carga de Dados

3.1. Fontes de Dados

Foram baixadas duas planilhas em Excel direto da plataforma da PRF, uma para o ano de 2020 e outra para o ano de 2021[3]. O banco de dados foi gerado através de código em SQL, e os dados foram injetados através de planilhas CSV diretamente para o SQL Server por meio do Microsoft SQL Server Management Studio, obedecendo às regras da modelagem de dados do autor.

O nome do banco de dados criado é **db_acidentes** e ele vai servir de base para toda a análise do projeto. A figura abaixo mostra a apresentação no SGBD.

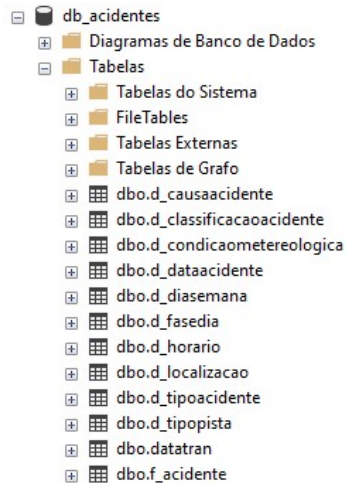


Figura 2: db_acidentes e tabelas

A forma como essas tabelas se relaciona está destacada na figura abaixo, no modelo ER (Entidade Relacionamento). Todas as dimensões estão vinculadas com o fato **f_acidente** e estão devidamente conectadas através de chave-estrangeira.

A conexão criada é de 1:N indicando que podem haver N acidentes em cada aspecto (cada dimensão) da base de dados, mas pode haver apenas um único acidente referenciando uma linha no banco de dados. A fonte dos dados se limita a indicar apenas se há ou não um boletim gerado, ou seja, se houve ou não o acidente. Dessa forma, mesmo se houvessem múltiplos automóveis, ainda seria apenas um acidente.

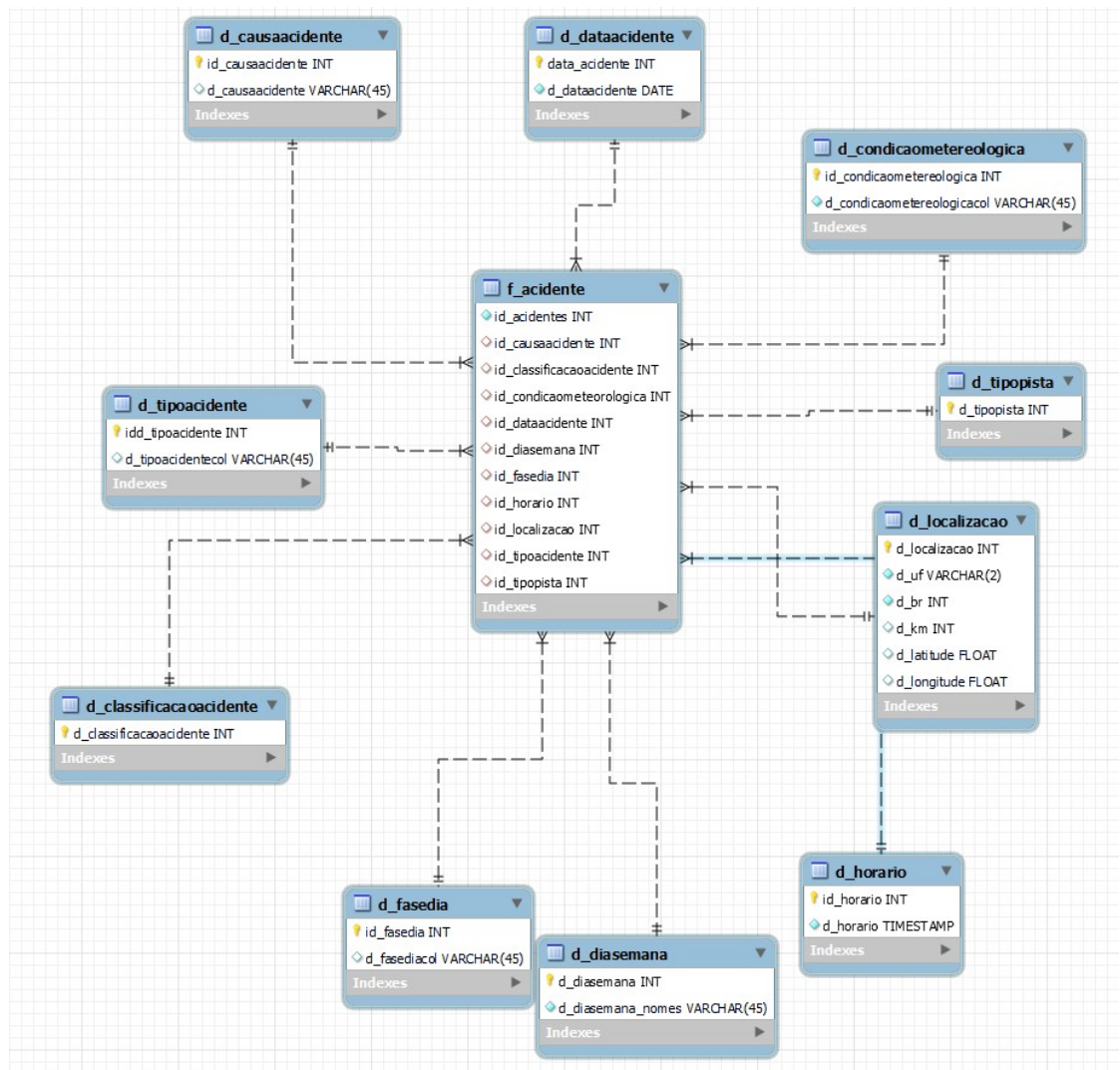


Figura 3: Modelo ER

3.2. Processos de Integração e Carga (ETL)

A extração do projeto foi realizada a partir de arquivos CSV baixados diretamente do site da PRF [3].

O processo de transformação de dados foi feito primariamente pelo próprio Microsoft Excel, a fim de facilitar a própria ingestão de dados no SQL Server, como pode ser acompanhado no capítulo 6 deste documento.

| id | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | AA | AB | AC | | | | | | | | | | | |
|--------|-------|------------|----------|------|---------|-------|------------|------------|-------------|-------|-------|----------|-----------|-------------|-----------|-----------|---------|------------|-------|-------|---------|------------|----------|--------|---------|----|---------|----|-------|----------|----------|----------|----------|-------------|----------|-----------|----|--|--|--|
| id | data | irve | dia | sema | horario | uf | br | km | municipio | causa | aci | tipo | acid | classificac | fase | dia | sentido | v_condicao | tipo | pista | tracado | v_uso_solo | personas | mortos | feridos | lf | feridos | gr | ileso | ignorado | feridos | veiculos | latitude | longitude | regional | delegacia | uc | | | |
| 260031 | ##### | quarta-fei | 01:00:00 | TO | 153 | 678.1 | GURUPI | Animais n | Atropelan | Com | Vitim | Plena | Noi | Crescente | Nublado | Simplex | Reta | Não | | | | | 5 | 0 | 0 | 5 | 0 | 0 | 5 | 1 | -11,7746 | -49,1074 | SR-TO | UOP01/TO | | | | | | |
| 260036 | ##### | quarta-fei | 01:00:00 | RJ | 116 | 178 | NOVA IGU | Falta de A | Atropelan | Com | Vitim | Plena | Noi | Decrescer | Céu | Claro | Dupla | Reta | Sim | | | | 2 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | -22,7522 | -43,4379 | SR-RJ | DEL5/1 | | | | | | |
| 260037 | ##### | quarta-fei | 01:52:00 | SC | 101 | 206.9 | SAO JOSE | Ingestão c | Colisão | la | Com | Vitim | Plena | Noi | Crescente | Nublado | Simplex | Reta | Sim | | | | 2 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | -27,9319 | -48,6182 | SR-SC | DEL8/1 | | | | | | |
| 260038 | ##### | quarta-fei | 01:15:00 | RO | 364 | 236 | CACAOI | Velocidade | Saída de li | Com | Vitim | Plena | Noi | Decrescer | Nublado | Simplex | Reta | Sim | | | | | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | -11,4462 | -61,4376 | SR-RO | DEL21/2 | | | | | | |
| 260039 | ##### | quarta-fei | 01:30:00 | PR | 153 | 360 | REBOUCA | Ingestão c | Capotame | Com | Vitim | Plena | Noi | Crescente | Garoa/Chi | Simplex | Curva | Não | | | | | 6 | 0 | 2 | 4 | 0 | 0 | 6 | 1 | -25,675 | -50,7509 | SR-PR | DEL21/3 | | | | | | |
| 260040 | ##### | quarta-fei | 00:40:00 | RO | 319 | 64 | PORTO VE | Ingestão c | Colisão | tr | Sem | Vitim | Plena | Noi | Decrescer | Céu | Claro | Dupla | Reta | Sim | | | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 2 | -8,76752 | -63,8843 | SR-RO | DEL21/1 | | | | | | |
| 260041 | ##### | quarta-fei | 02:20:00 | SC | 101 | 89.9 | BARRA VE | Falta de A | Colisão | co | Sem | Vitim | Plena | Noi | Crescente | Garoa/Chi | Dupla | Reta | Sim | | | | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | -26,6633 | -48,609 | SR-SC | DEL8/3 | | | | | | |
| 260042 | ##### | quarta-fei | 02:15:00 | PR | 369 | 155 | LONDRINA | Depedib | Colisão | tr | Com | Vitim | Plena | Noi | Crescente | Céu | Claro | Múltipla | Não | Infon | Sim | | | 3 | 0 | 1 | 0 | 2 | 0 | 1 | 2 | -23,2929 | -51,1615 | SR-PR | DEL7/7 | | | | | |
| 260043 | ##### | quarta-fei | 02:30:00 | AL | 104 | 37 | UNIAO DC | Animais n | Atropelan | Com | Vitim | Plena | Noi | Decrescer | Nublado | Simplex | Reta | Não | | | | | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | -9,17491 | -36,0294 | SR-AL | DEL13/1 | | | | | | |
| 260044 | ##### | quarta-fei | 00:20:00 | RS | 290 | 97 | PORTO AL | Falta de A | Tombame | Com | Vitim | Plena | Noi | Decrescer | Céu | Claro | Dupla | Reta | Não | | | | 2 | 0 | 2 | 0 | 0 | 0 | 2 | 1 | -29,9978 | -51,2078 | SR-RS | DEL9/1 | | | | | | |
| 260045 | ##### | quarta-fei | 02:50:00 | SC | 470 | 116.3 | IBIRAMA | Condutor | Colisão | fr | Com | Vitim | Plena | Noi | Decrescer | Nublado | Simplex | Curva | Sim | | | | 4 | 0 | 2 | 2 | 0 | 0 | 4 | 2 | -27,0776 | -49,4945 | SR-SC | DEL8/4 | | | | | | |
| 260046 | ##### | quarta-fei | 02:15:00 | SC | 282 | 309 | XANXERE | Animais n | Saída de li | Com | Vitim | Plena | Noi | Decrescer | Céu | Claro | Simplex | Reta | Não | | | | 3 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | -26,8944 | -52,4614 | SR-SC | DEL7 | | | | | | |
| 260050 | ##### | quarta-fei | 03:40:00 | PE | 101 | 49.5 | ABREU E L | Falta de A | Colisão | la | Com | Vitim | Plena | Noi | Crescente | Céu | Claro | Múltipla | Reta | Sim | | | 4 | 0 | 0 | 0 | 2 | 1 | 1 | 2 | -7,90052 | -34,9018 | SR-PE | DEL11/1 | | | | | | |
| 260051 | ##### | quarta-fei | 02:30:00 | MT | 364 | 265.3 | SAO PEDR | Falta de A | Queda de | Com | Vitim | Plena | Noi | Crescente | Céu | Claro | Simplex | Não | Infon | Sim | | | 2 | 0 | 2 | 0 | 0 | 0 | 2 | 1 | -16,0044 | -54,92 | SR-MT | DEL2/1 | | | | | | |
| 260052 | ##### | quarta-fei | 03:10:00 | BA | 101 | 174 | FEIRA DE S | Animais n | Atropelan | Com | Vitim | Plena | Noi | Crescente | Nublado | Simplex | Reta | Não | | | | | 2 | 0 | 1 | 1 | 0 | 0 | 2 | 1 | -12,38 | -38,88 | SR-BA | DEL10/1 | | | | | | |
| 260053 | ##### | quarta-fei | 04:00:00 | SC | 101 | 132 | BALNEARI | Falta de A | Atropelan | Com | Vitim | Plena | Noi | Crescente | Céu | Claro | Dupla | Curva | Sim | | | | 2 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | -26,9965 | -48,6556 | SR-SC | DEL8/4 | | | | | | |
| 260054 | ##### | quarta-fei | 00:05:00 | AP | 156 | 406 | FERREIRA | Animais n | Atropelan | Sem | Vitim | Plena | Noi | Crescente | Céu | Claro | Simplex | Reta | Não | | | | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1,021969 | -51,0902 | SR-AP | UOP01/AP | | | | | | |
| 260055 | ##### | quarta-fei | 03:20:00 | GO | 60 | 164.8 | GOIANIA | Falta de A | Saída de li | Com | Vitim | Amanheci | Crescente | Nublado | Dupla | Reta | Não | | | | | | 4 | 0 | 0 | 4 | 0 | 0 | 0 | 4 | -16,7036 | -49,342 | SR-GO | DEL1/1 | | | | | | |
| 260057 | ##### | quarta-fei | 04:55:00 | DF | 20 | 21.8 | BRASILIA | Ingestão c | Atropelan | Com | Vitim | Amanheci | Crescente | Garoa/Chi | Dupla | Reta | Sim | | | | | | 3 | 0 | 2 | 0 | 1 | 0 | 0 | 2 | -15,6034 | -47,6847 | SR-DF | UOP01/DEL01 | | | | | | |
| 260060 | ##### | quarta-fei | 05:00:00 | SP | 116 | 99.6 | PINDAMO | Falta de A | Colisão | co | Com | Vitim | Amanheci | Crescente | Céu | Claro | Dupla | Reta | Não | | | | 2 | 0 | 2 | 0 | 0 | 0 | 2 | 1 | -22,9824 | -45,4644 | SR-SP | DEL6 | | | | | | |
| 260061 | ##### | quarta-fei | 03:00:00 | BA | 324 | 616 | SALVADO | Falta de A | Queda de | Com | Vitim | Plena | Noi | Decrescer | Nublado | Dupla | Reta | Sim | | | | | 2 | 0 | 2 | 0 | 0 | 0 | 2 | 1 | -12,8796 | -38,4321 | SR-BA | DEL10/1 | | | | | | |
| 260063 | ##### | quarta-fei | 03:30:00 | SC | 101 | 252.2 | PAULO LO | Ingestão c | Colisão | la | Sem | Vitim | Plena | Noi | Crescente | Céu | Claro | Dupla | Reta | Não | | | | 5 | 0 | 0 | 0 | 5 | 0 | 0 | 2 | -27,9547 | -48,6768 | SR-SC | DEL6/2 | | | | | |
| 260064 | ##### | quarta-fei | 05:45:00 | PB | 101 | 32 | MAMANGA | Ingestão c | Colisão | co | Com | Vitim | Pleno | dia | Crescente | Céu | Claro | Dupla | Reta | Não | | | | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | -6,76718 | -35,1335 | SR-PB | DEL14/1 | | | | | |
| 260065 | ##### | quarta-fei | 06:00:00 | PI | 373 | 210.4 | IPIRANGA | Condutor | Colisão | co | Com | Vitim | Amanheci | Crescente | Céu | Claro | Simplex | Reta | Não | | | | | 4 | 0 | 2 | 2 | 0 | 0 | 4 | 1 | -25,1451 | -50,4877 | SR-PI | DEL7/3 | | | | | |
| 260066 | ##### | quarta-fei | 02:45:00 | PI | 407 | 445.3 | GEMINIA | Falta de A | Colisão | tr | Com | Vitim | Pleno | dia | Decrescer | Céu | Claro | Simplex | Reta | Não | | | | 2 | 0 | 0 | 1 | 1 | 0 | 1 | 2 | -21,4407 | -41,2873 | SR-PI | DEL17/4 | | | | | |
| 260067 | ##### | quarta-fei | 06:00:00 | MG | 381 | 671.5 | PERDÕES | Ingestão c | Saída de li | Com | Vitim | Pleno | dia | Crescente | Chuva | Dupla | Não | Infon | Não | | | | | 4 | 0 | 4 | 0 | 0 | 4 | 1 | -21,0895 | -45,0686 | SR-MG | DEL4/4 | | | | | | |
| 260068 | ##### | quarta-fei | 05:40:00 | PA | 316 | 94 | SAO FRAN | Falta de A | Saída de li | Com | Vitim | Pleno | dia | Decrescer | Céu | Claro | Simplex | Reta | Não | | | | | 2 | 0 | 2 | 0 | 0 | 0 | 2 | 1 | -1,31019 | -47,7446 | SR-PA | DEL19/1 | | | | | |
| 260069 | ##### | quarta-fei | 05:45:00 | SP | 153 | 110.1 | JOSE BON | Falta de A | Colisão | co | Com | Vitim | Pleno | dia | Decrescer | Céu | Claro | Simplex | Reta | Sim | | | | 5 | 0 | 3 | 0 | 2 | 0 | 3 | 1 | -21,0885 | -49,6747 | SR-SP | DEL6/9 | | | | | |
| 260070 | ##### | quarta-fei | 06:15:00 | AL | 101 | 15.4 | MATRIZ DI | Falta de A | Colisão | co | Com | Vitim | Pleno | dia | Crescente | Céu | Claro | Dupla | Curva | Não | | | | 2 | 0 | 2 | 0 | 0 | 0 | 2 | 1 | -9,02218 | -35,6797 | SR-AL | DEL13/1 | | | | | |
| 260071 | ##### | quarta-fei | 06:20:00 | GO | 153 | 417 | PIRENOPC | Ingestão c | Saída de li | Com | Vitim | Amanheci | Crescente | Nublado | Simplex | Reta | Não | | | | | | 2 | 0 | 2 | 0 | 0 | 0 | 2 | 1 | -16,151 | -49,0447 | SR-GO | DEL1/2 | | | | | | |
| 260073 | ##### | quarta-fei | 06:00:00 | MG | 262 | 804 | UBERABA | Falta de A | Colisão | tr | Com | Vitim | Pleno | dia | Decrescer | Céu | Claro | Dupla | Reta | Sim | | | | 4 | 0 | 1 | 0 | 3 | 0 | 1 | 2 | -19,7675 | -47,9673 | SR-MG | DEL13 | | | | | |
| 260074 | ##### | quarta-fei | 05:50:00 | RN | 101 | 96 | NATAL | Pista Esco | Queda de | Com | Vitim | Pleno | dia | Crescente | Nublado | Dupla | Viaduto | Sim | | | | | 2 | 0 | 1 | 0 | 1 | 0 | 1 | 2 | -5,8495 | -35,2093 | SR-RN | DEL15/1 | | | | | | |
| 260075 | ##### | quarta-fei | 05:30:00 | MG | 381 | 548.5 | RIO MANI | Condutor | Capotame | Sem | Vitim | Amanheci | Decrescer | Céu | Claro | Dupla | Reta | Não | | | | | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | -20,3102 | -44,4499 | SR-MG | DEL4/4 | | | | | | |
| 260076 | ##### | quarta-fei | 07:00:00 | SC | 101 | 155 | PORTO BE | Falta de A | Tombame | Com | Vitim | Pleno | dia | Crescente | Sol | Dupla | Reta | Não | | | | | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | -27,1629 | -48,6113 | SR-SC | DEL4/4 | | | | | | |
| 260077 | ##### | quarta-fei | 06:40:00 | GO | 40 | 10.1 | LUZIANIA | Falta de A | Colisão | co | Sem | Vitim | Pleno | dia | Crescente | Céu | Claro | Dupla | Reta | Sim | | | | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | -16,1313 | -47,9579 | SR-DF | UOP01/DF | | | | | |
| 260078 | ##### | quarta-fei | 06:00:00 | MT | 381 | 325 | NOVA FRI | Velocidade | Saída de li | Com | Vitim | Amanheci | Crescente | Céu | Claro | Simplex | Reta | Não | | | | | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | -19,761 | -43,0367 | SR-MG | DEL4/1 | | | | | | |

Figura 4: Dados não tratados

Inicialmente, haviam 30 colunas em cada um dos arquivos, tanto o arquivo de 2020 quanto o de 2021. Após uma análise, foi decidido quais colunas eram mais relevantes para o cenário proposto, e então as colunas foram renomeadas e extraídas para formar arquivos distintos. As planilhas foram integradas (2020 e 2021).












| Nome | Data de modificação |
|--|---------------------|
|  d_causaacidente | 11/12/2021 15:25 |
|  d_classificacaoacidente | 11/12/2021 15:31 |
|  d_condicaoometereologica | 11/12/2021 15:35 |
|  d_dataacidente | 11/12/2021 14:38 |
|  d_diasemana | 11/12/2021 15:44 |
|  d_fasedia | 11/12/2021 15:45 |
|  d_horario | 11/12/2021 15:47 |
|  d_localizacao | 11/12/2021 15:49 |
|  d_tipoacidente | 11/12/2021 15:50 |
|  d_tipopista | 11/12/2021 15:52 |
|  f_acidente | 15/12/2021 17:16 |

Figura 5: Dados tratados

Cada arquivo contém apenas o que lhe é necessário. Assim, a tabela Fato f_acidente contém apenas o id_acidente, referente ao id, que será a chave primária no banco de dados. E as dimensões terão o id_acidente além do campo que define a dimensão

f_acidente - Excel

Arquivo **Página Inicial** Inserir Layout da Página Fórmulas Dados Revisão Exibir Ajuda Power Pivot Diga-me o que você deseja

Recortar Copiar Colar Pincel de Formatação

Calibri 11 A A

N I S

Quebrar Texto Automaticamente

Mesclar e Centralizar

Área de Transferência Fonte Alinhamento Número

Formato Condicional

POSSÍVEL PERDA DE DADOS Alguns recursos podem ser perdidos se você salvar esta pasta de trabalho no formato (.csv) delimitada por vírgulas. Para preservar estes recursos, salve-a em

A1 id_acidente

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
|----|-------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | id_acidente | | | | | | | | | | | | | | | |
| 2 | 331693 | | | | | | | | | | | | | | | |
| 3 | 331694 | | | | | | | | | | | | | | | |
| 4 | 331696 | | | | | | | | | | | | | | | |
| 5 | 331699 | | | | | | | | | | | | | | | |
| 6 | 331701 | | | | | | | | | | | | | | | |
| 7 | 331702 | | | | | | | | | | | | | | | |
| 8 | 331703 | | | | | | | | | | | | | | | |
| 9 | 331704 | | | | | | | | | | | | | | | |
| 10 | 331706 | | | | | | | | | | | | | | | |
| 11 | 331707 | | | | | | | | | | | | | | | |
| 12 | 331708 | | | | | | | | | | | | | | | |
| 13 | 331709 | | | | | | | | | | | | | | | |
| 14 | 331710 | | | | | | | | | | | | | | | |
| 15 | 331711 | | | | | | | | | | | | | | | |
| 16 | 331712 | | | | | | | | | | | | | | | |
| 17 | 331714 | | | | | | | | | | | | | | | |
| 18 | 331715 | | | | | | | | | | | | | | | |
| 19 | 331716 | | | | | | | | | | | | | | | |
| 20 | 331717 | | | | | | | | | | | | | | | |
| 21 | 331718 | | | | | | | | | | | | | | | |
| 22 | 331719 | | | | | | | | | | | | | | | |
| 23 | 331721 | | | | | | | | | | | | | | | |
| 24 | 331722 | | | | | | | | | | | | | | | |
| 25 | 331724 | | | | | | | | | | | | | | | |
| 26 | 331725 | | | | | | | | | | | | | | | |
| 27 | 331726 | | | | | | | | | | | | | | | |
| 28 | 331727 | | | | | | | | | | | | | | | |
| 29 | 331728 | | | | | | | | | | | | | | | |
| 30 | 331729 | | | | | | | | | | | | | | | |
| 31 | 331730 | | | | | | | | | | | | | | | |
| 32 | 331731 | | | | | | | | | | | | | | | |
| 33 | 331732 | | | | | | | | | | | | | | | |
| 34 | 331733 | | | | | | | | | | | | | | | |
| 35 | 331777 | | | | | | | | | | | | | | | |
| 36 | 331778 | | | | | | | | | | | | | | | |
| 37 | 331779 | | | | | | | | | | | | | | | |

f_acidente

Figura 6: Tabela Fato tratada

POSSÍVEL PERDA DE DADOS Alguns recursos podem ser perdidos se você salvar esta pasta de trabalho no formato (.csv) delimitada por vírgulas. Para preservar estes recursos, salve-a em um formato de pasta de trabalho do Excel.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|----|------------|----------------|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | f_acidente | d_dataacidente | | | | | | | | | | | | | |
| 2 | 331693 | 01/01/2021 | | | | | | | | | | | | | |
| 3 | 331694 | 01/01/2021 | | | | | | | | | | | | | |
| 4 | 331696 | 01/01/2021 | | | | | | | | | | | | | |
| 5 | 331699 | 01/01/2021 | | | | | | | | | | | | | |
| 6 | 331701 | 01/01/2021 | | | | | | | | | | | | | |
| 7 | 331702 | 01/01/2021 | | | | | | | | | | | | | |
| 8 | 331703 | 01/01/2021 | | | | | | | | | | | | | |
| 9 | 331704 | 01/01/2021 | | | | | | | | | | | | | |
| 10 | 331706 | 01/01/2021 | | | | | | | | | | | | | |
| 11 | 331707 | 01/01/2021 | | | | | | | | | | | | | |
| 12 | 331708 | 01/01/2021 | | | | | | | | | | | | | |
| 13 | 331709 | 01/01/2021 | | | | | | | | | | | | | |
| 14 | 331710 | 01/01/2021 | | | | | | | | | | | | | |
| 15 | 331711 | 01/01/2021 | | | | | | | | | | | | | |
| 16 | 331712 | 01/01/2021 | | | | | | | | | | | | | |
| 17 | 331714 | 01/01/2021 | | | | | | | | | | | | | |
| 18 | 331715 | 01/01/2021 | | | | | | | | | | | | | |
| 19 | 331716 | 01/01/2021 | | | | | | | | | | | | | |
| 20 | 331717 | 01/01/2021 | | | | | | | | | | | | | |
| 21 | 331718 | 01/01/2021 | | | | | | | | | | | | | |
| 22 | 331719 | 01/01/2021 | | | | | | | | | | | | | |
| 23 | 331721 | 01/01/2021 | | | | | | | | | | | | | |
| 24 | 331722 | 01/01/2021 | | | | | | | | | | | | | |
| 25 | 331724 | 01/01/2021 | | | | | | | | | | | | | |
| 26 | 331725 | 01/01/2021 | | | | | | | | | | | | | |
| 27 | 331726 | 01/01/2021 | | | | | | | | | | | | | |
| 28 | 331727 | 01/01/2021 | | | | | | | | | | | | | |
| 29 | 331728 | 01/01/2021 | | | | | | | | | | | | | |
| 30 | 331729 | 01/01/2021 | | | | | | | | | | | | | |
| 31 | 331730 | 01/01/2021 | | | | | | | | | | | | | |
| 32 | 331731 | 01/01/2021 | | | | | | | | | | | | | |
| 33 | 331732 | 01/01/2021 | | | | | | | | | | | | | |
| 34 | 331733 | 01/01/2021 | | | | | | | | | | | | | |
| 35 | 331777 | 01/01/2021 | | | | | | | | | | | | | |
| 36 | 331778 | 01/01/2021 | | | | | | | | | | | | | |
| 37 | 331779 | 01/01/2021 | | | | | | | | | | | | | |

Figura 7: Exemplo de Dimensão após tratamento de dados

Após isso, as colunas (dimensões) foram extraídas em formato CSV UTF-8, levando consigo o id_acidente.

A carga é feita no Power BI diretamente pelo SQL Server. Há um vídeo dentre os materiais do projeto que indica como essa etapa de carga foi realizada.

4. Camada de Apresentação

4.1 Dashboard

Como mencionado previamente, foi utilizado o software Power BI para a criação do Dashboard deste trabalho. A primeira aba (estratégica) passa uma visão geral dos principais KPIs relacionados ao problema.

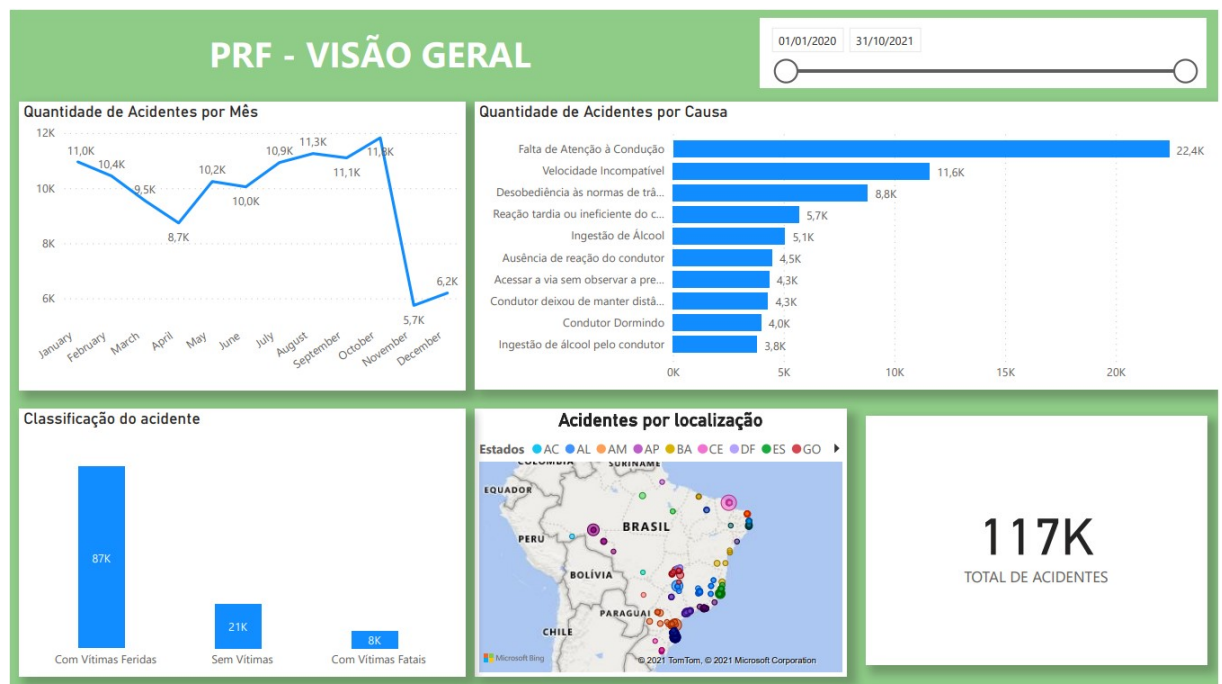


Figura 8: Dashboard - Visão Geral

Nessa visão geral, é possível observar a quantidade total de acidentes no período estudado, quantos desses aconteceram em cada mês do ano, além de seus respectivos motivos. Fora isso, temos a quantidade de vítimas (graves ou não). Por último, temos um mapa de onde aconteceram todas as ocorrências.

Além da visão geral, foi realizada uma visão mais aprofundada do acidente, denominada temporal. Nela, é possível observar a quantidade de acidentes pelo horário que ele acontece e filtrar pelo Estado brasileiro que ele ocorreu. Além disso, é possível observar em que fase do dia, e dia da semana, há uma maior ocorrência de acidentes. Como há um filtro por estado na aba, essa filtragem de fase do dia e dia da semana também pode ser realizada pelo estado.

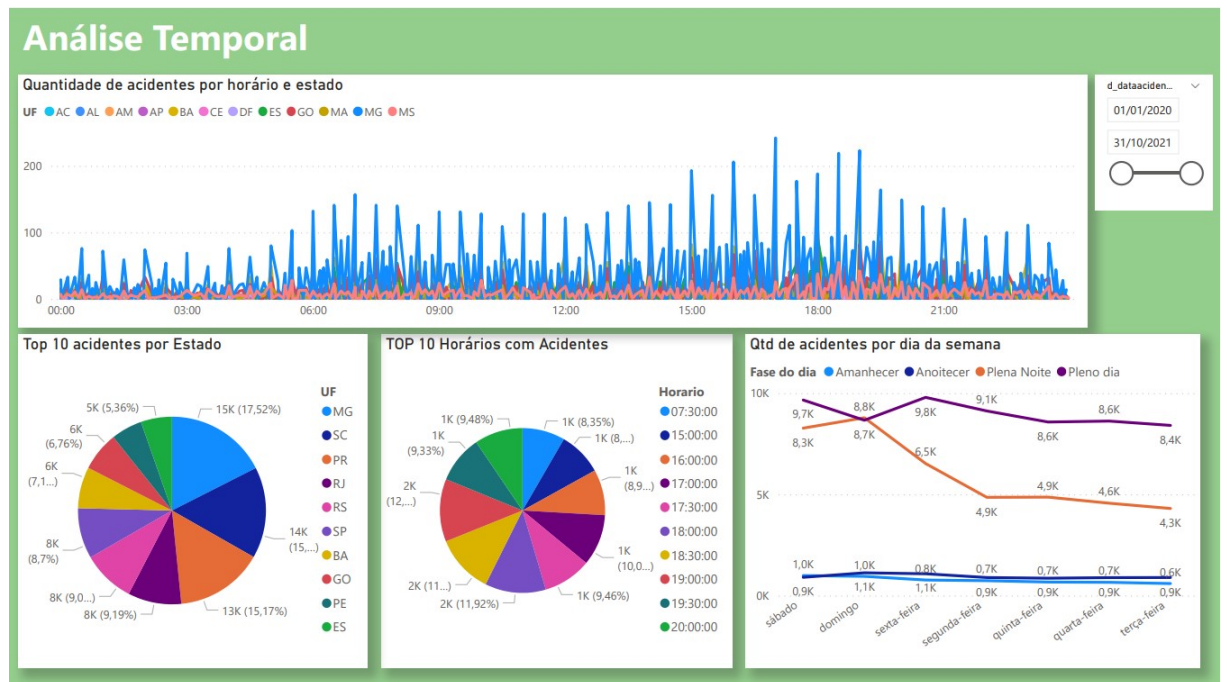


Figura 9: Dashboard: Análise Temporal

O último dashboard traz um detalhamento ainda maior sobre as informações que definem o momento do acidente. Nele, é possível observar acidentes pelo tipo de pista, quais as condições meteorológicas no momento do desastre, em qual rodovia aconteceu, além de uma contagem do tipo do acidente.

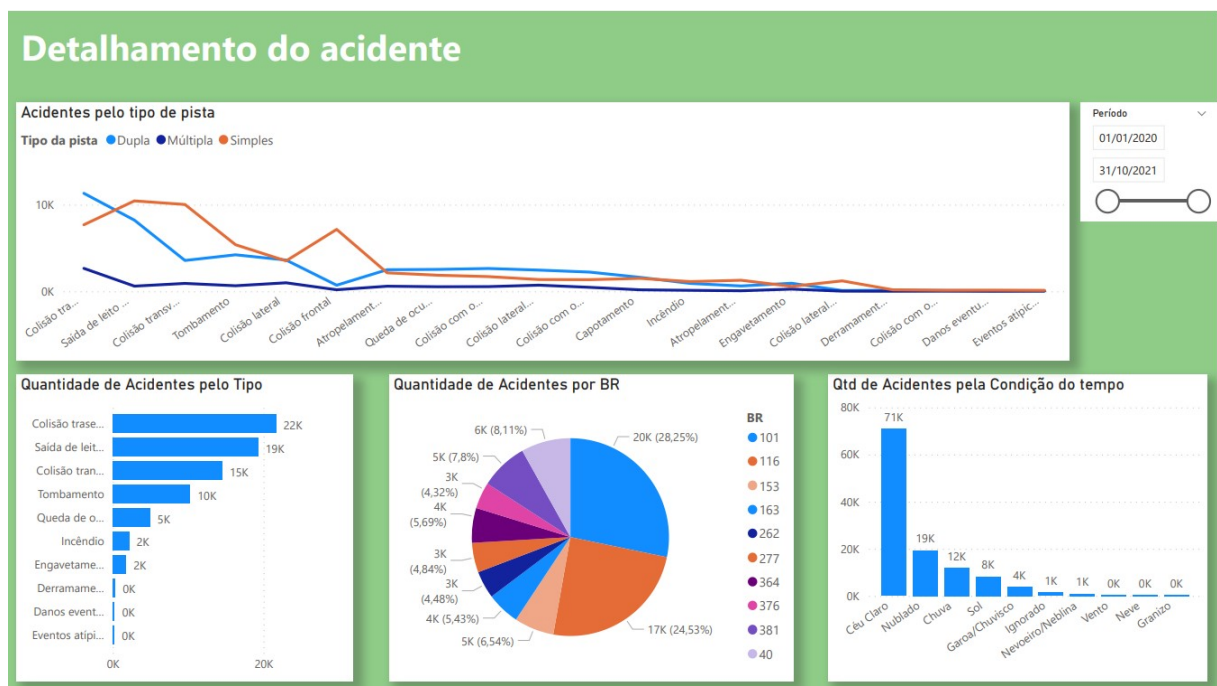


Figura 10: Dashboard- Detalhamento do acidente

4.2 Análises avançadas

Para fazer as análises aqui apresentadas, foi utilizado um caderno Jupyter, codificado em Python. As codificações e imagens se encontram no mesmo link disponibilizado na seção 6.

Na análise exploratória, as duas primeiras imagens se encontram com uma normalização, para que se possa comparar visualmente em percentuais.

Pode-se observar que mais da metade dos desastres acontecem em pleno dia, o que de fato ocorreu (os números indicam 53.7%). Além disso, na Figura 8 observa-se uma alta taxa de acidentes com condição meteorológica “Céu claro” (taxa de 60,6%).

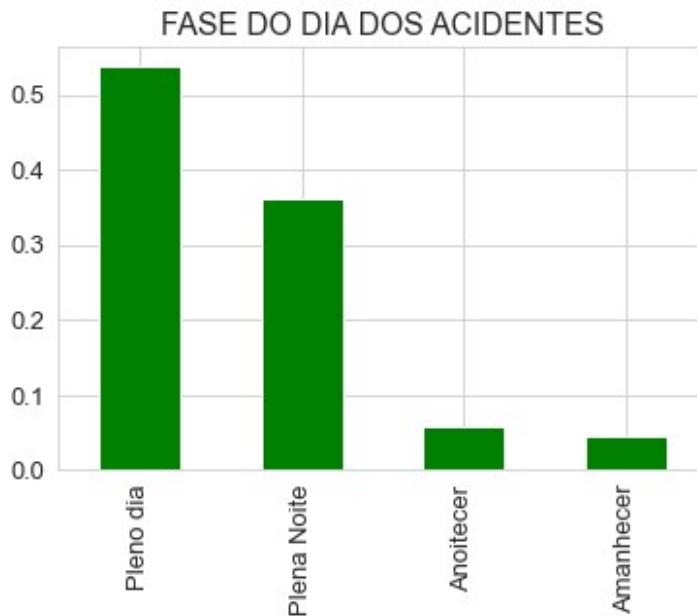


Figura 11: Análise Exploratória - Fase do dia dos acidentes

CONDIÇÃO METEOROLÓGICA NO MOMENTO DO ACIDENTE

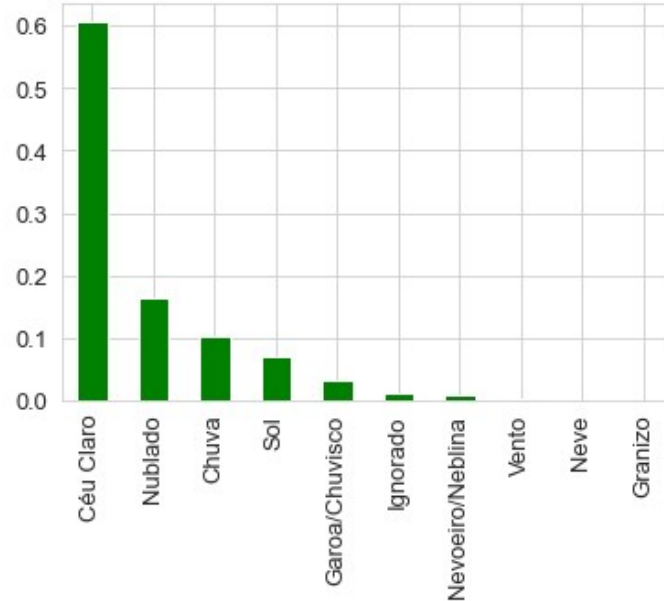


Figura 12: Análise exploratória - Condição meteorológica no momento do acidente

CAUSAS MAIS FREQUENTES



Figura 13: Análise exploratória - Causas mais frequentes



Figura 14: Análise exploratória - Acidentes por Estado

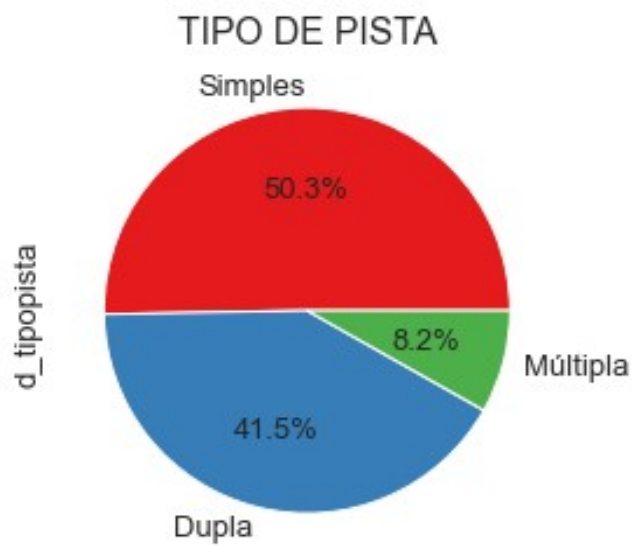


Figura 15: Análise exploratória - Tipo de Pista do Acidente

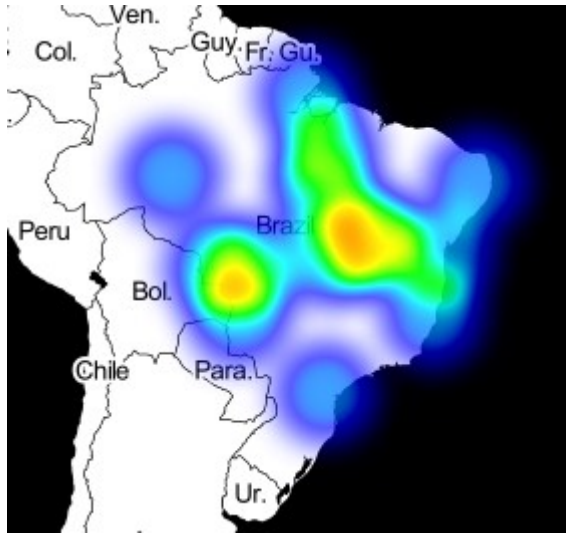


Figura 16: Heat map - Mapa brasileiro com a localização de onde houve mais acidentes

Análise Preditiva

Quanto à análise preditiva, foram escolhidas as dimensões referentes à d_localização: d_br e d_km, além da condição meteorológica (d_condicaometeorologica). O que elas têm em comum: estão no dashboard que define o acidente em detalhes, por isso devem aumentar a confiança na predição.

Além dessas dimensões, foram escolhidas as dimensões d_fasedia e d_diasemana, do dashboard temporal, trazendo uma nova perspectiva do acidente e fortalecendo o treinamento da máquina.

Fora isso, foi criado o campo “risco”, que será o campo aonde a máquina trabalhará em torno. Foi criada uma classificação que indica se houveram ou não vítimas no banco a ser trabalhado, como segue:

```
df['risco'] = df.d_classificacaoacidente.apply(lambda x: 1 if x in ['Com Vítimas Feridas', 'Com Vítimas Fatais'] else 0)
```

Essa classificação alimenta a variável para que ela compreenda como risco todo aquele acidente que teve vítimas, sejam elas feridas ou vítimas fatais.

Foram utilizados cinco diferentes modelos de Machine Learning[1]: Regressão Logística, KNN (Classificador KNeighbors), Árvore de Decisão, Gradient Boosting e o Random Forest.

Para medir qual previsão se saiu melhor, foi utilizado Receiver Operating Characteristics (ROC) unido de Area Under The Curve (AUC). Foram obtidas matrizes de confusão para cada um dos modelos, e o modelo que obteve o melhor resultado foi o Gradiente Boosting, com o coeficiente 55.8%. A Regressão Logística ficou logo atrás com 55.2%, como pode ser observado na Figura abaixo:

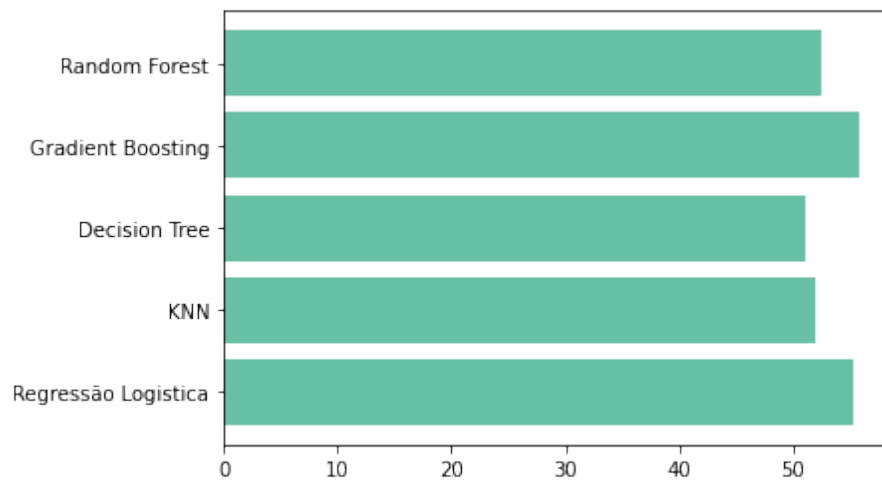


Figura 17: Comparativo dos resultados dos modelos escolhidos

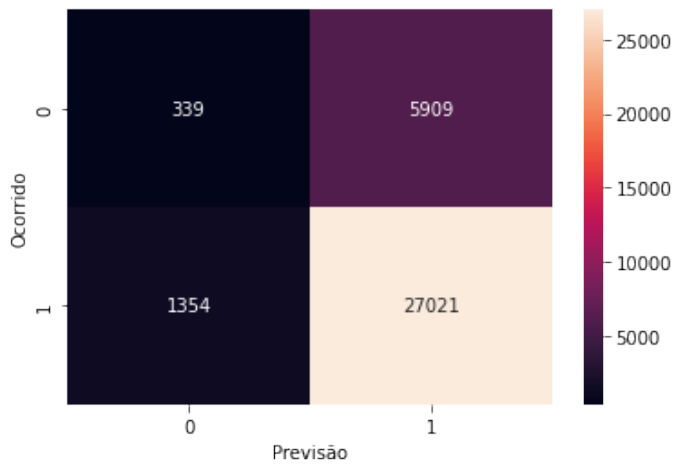


Figura 18: Matriz de confusão: KNN

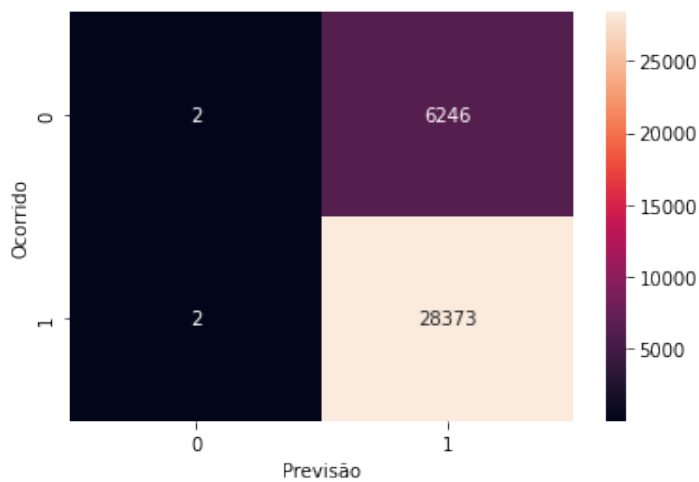


Figura 19: Matriz Confusão - Gradient Boosting

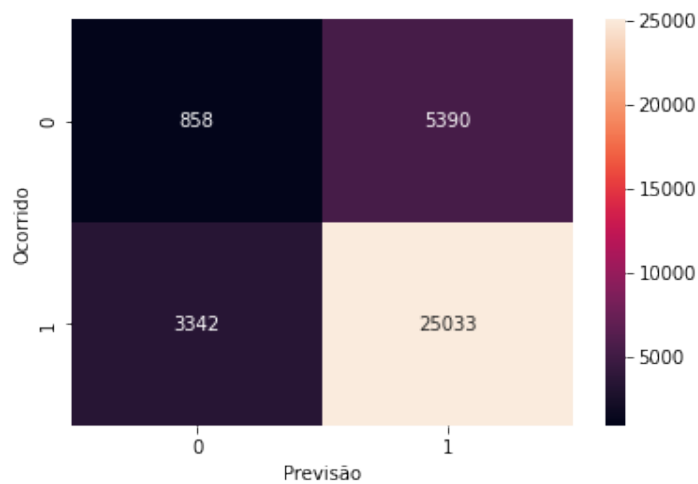


Figura 20: Matriz Confusão - Árvore de Decisão

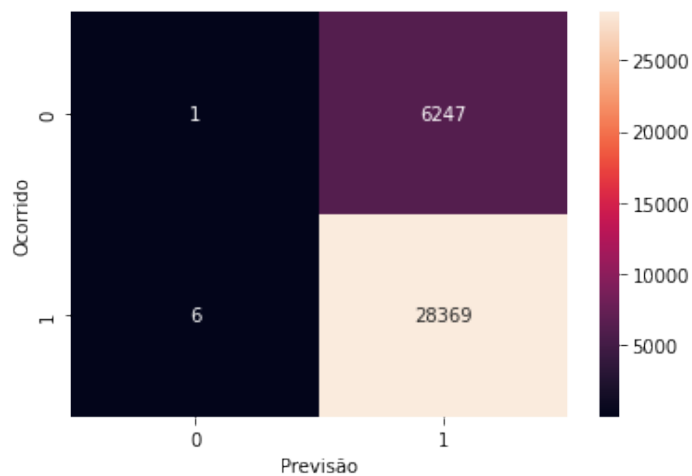


Figura 21: Matriz Confusão - Regressão Logística

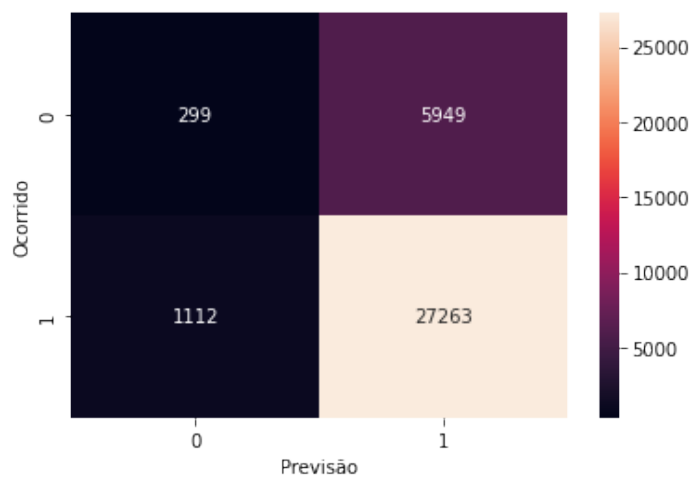
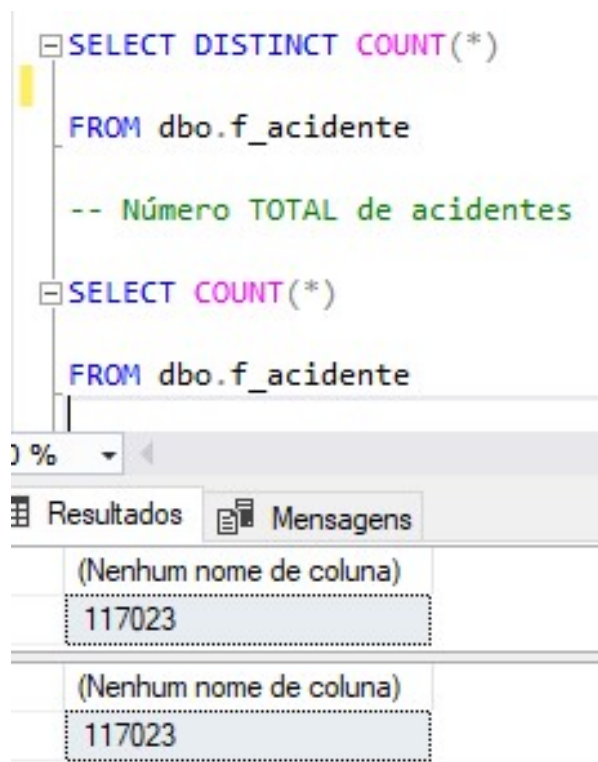


Figura 22: Matriz Confusão - Random Forest

5. Registros de Homologação

Abaixo encontra-se figuras comprovando o funcionamento do Banco de Dados db_acidentes e suas respectivas tabelas. São feitas consultas SQL para mostrar que os dados estão sendo apresentados no dashboard do Power BI conforme deveriam ser.

Nas figuras, pode-se observar o código da consulta, o resultado da consulta (que seriam as tabelas) e também a forma como está sendo apresentado na ferramenta Power BI).



117K
TOTAL DE ACIDENTES

Figura 23: Número total de acidentes - Comprovação de que não há distinção entre pesquisar com ou sem "distinct" no banco de dados, pois não há desastres repetidos no SGBD

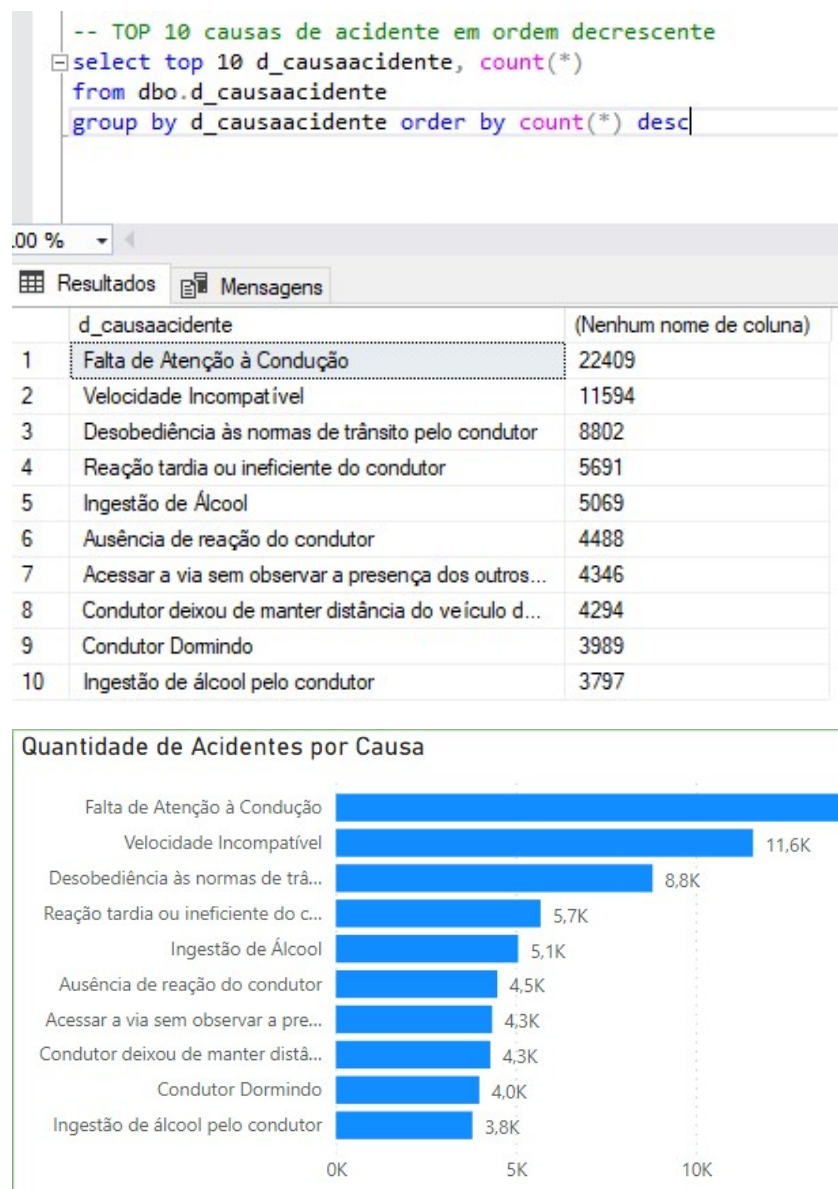


Figura 24: TOP 10 Quantidade de acidentes por causa em ordem decrescente

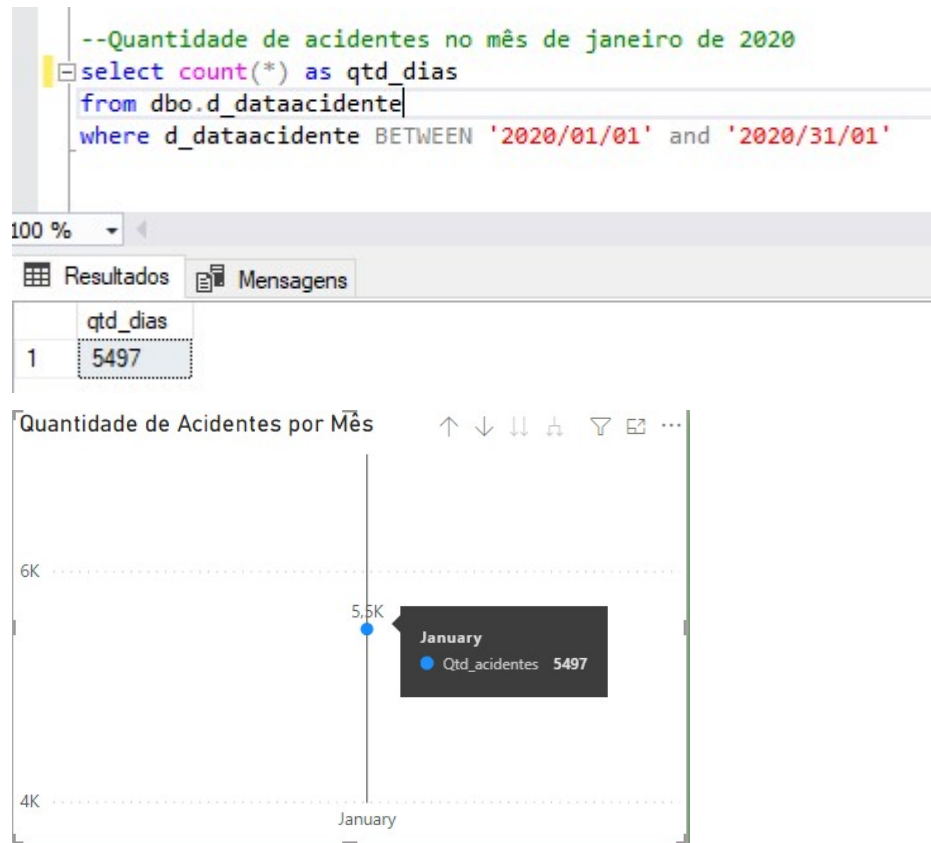


Figura 25: Quantidade de acidentes ocorridos no mês de janeiro de 2020

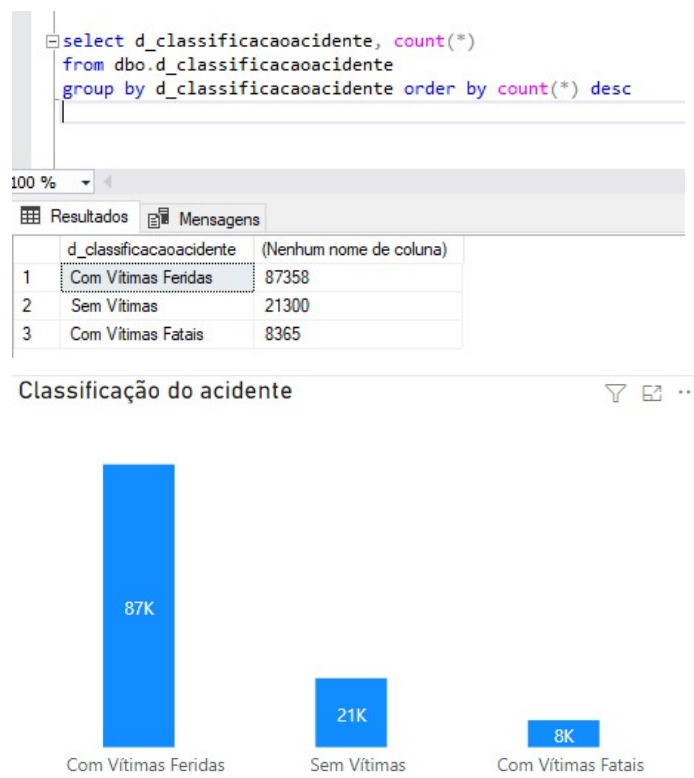


Figura 26: Quantidade de vítimas nos acidentes

```

select top 10 d_localizacao.d_uf, count(*)
from d_localizacao
group by d_localizacao.d_uf order by count(*) desc

```

100 %

Resultados Mensagens

| | d_uf | (Nenhum nome de coluna) |
|----|------|-------------------------|
| 1 | MG | 15279 |
| 2 | SC | 13709 |
| 3 | PR | 13223 |
| 4 | RJ | 8016 |
| 5 | RS | 7917 |
| 6 | SP | 7585 |
| 7 | BA | 6218 |
| 8 | GO | 5895 |
| 9 | PE | 4677 |
| 10 | ES | 4669 |

Top 10 acidentes por Estado

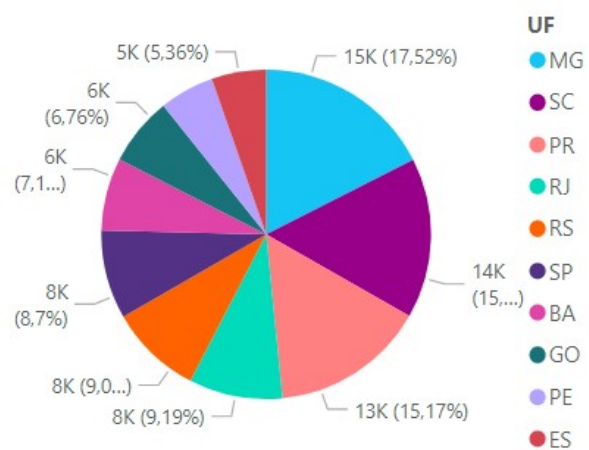


Figura 27: Top 10 Estados com mais acidentes

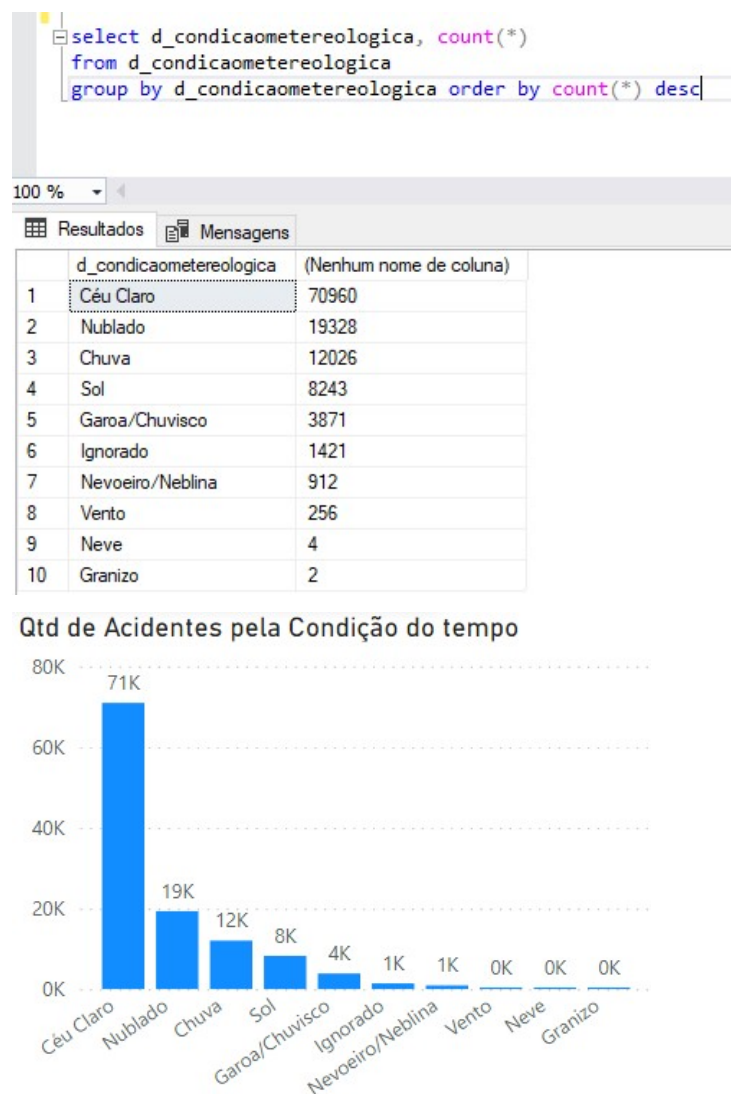


Figura 28: Quantidade de acidentes pela condição do tempo

5. Conclusões

Os acidentes automobilísticos que acontecem nas rodovias brasileiras são bastante difíceis de prever. Isso foi possível confirmar através da análise preditiva realizada, onde o melhor modelo para este projeto teve acurácia de 55.8%.

Existem alguns tipos de acidentes que não há uma forma clara nem de prever, nem de se prevenir. Infelizmente, estes são a maioria deles, pois são os acidentes que acontecem em plena luz do dia, com sol, e com o motivo do acidente sendo falta de atenção à condução. Essa conexão de fatores torna o trabalho dos gestores de prevenção ao acidente uma missão um tanto desafiadora, pois tais falhas humanas serão sempre possíveis de acontecer.

Entretanto, há algo com o que se possa trabalhar, que são os fatores externos que causam os acidentes. É possível observar que a taxa de colisão frontal, por exemplo, sobe muito quando se filtra os acidentes apenas por pistas simples. Uma forma então de quase erradicar esse tipo de acidente seria tornar todas as BRs como pistas duplas.

Esse projeto serve para indicar a outros que possam vir mais adiante para mostrar que, embora não haja uma forma de impedir acidentes, há sim formas de amenizá-los, fazer com que se tornem menos letais e em menor quantidade. Para projetos futuros, seria interessante se trabalhar com não apenas os dados aqui estudados como também os automóveis envolvidos nos acidentes. Pois, algo que limita esse projeto é a falta de dados relacionando os acidentes aos tipos de carros. Possivelmente, se for feita essa relação, pode ser útil para governo indicar quais automóveis mais seguros para compra ou até mesmo o consumidor na hora de compra do seu veículo saber quais carros mais seguros.

6. Links

Dados:

<https://drive.google.com/drive/folders/1HhUVbOSdeMTebtM5ktXSKY160pAYrRgV?usp=sharing>

Dados não tratados:

<https://drive.google.com/drive/folders/1bX5CV7xYyPZAZ7sZMDFqQmRn0wGf5ne3?usp=sharing>

Dados tratados:

<https://drive.google.com/drive/folders/1SwkTH4zdxhS61e08DUosrmGqfD8akC5q?usp=sharing>

Vídeo de apresentação: <https://www.youtube.com/watch?v=1nGKidPuRA4>

REFERÊNCIAS

- 1- CAMILO, C. O.; SILVA, J. C. Mineração de dados: Conceitos, tarefas, métodos e ferramentas. **Universidade Federal de Goiás (UFG)**, v. 1, n. 1, 2009.
- 2- DRAGOMIRESCU, S. E.; SOLOMON, D. C. The Role Of The Performance Dashboard In The Management Of Modern Enterprises. **Studies and Scientific Researches. Economics Edition**, n. 18, 2013.
- 3- PRF. **Conjunto de Dados Públicos de acidentes automobilísticos**, 2021. Disponível em: <<https://www.gov.br/prf/pt-br/aceso-a-informacao/dados-abertos/dados-abertos-acidentes>>. Brasil: Acesso em: 23 dez. 2021.