

Panorama de Acidentes em Rodovias Federais no Brasil | 2017-2020

Projeto da Disciplina de BI

Prof. Anderson Nascimento
prof.anderson@ica.ele.puc-rio.br

Componentes do Projeto:

Marcos Alonso Guimarães
marcos.guimaraesalonso@gmail.com

Histórico de Versões

Data	Versão	Descrição	Autor	Aprovado por
16/01/2021	1.0	Workflow – Processo BI	Marcos Alonso	Marcos Alonso
30/01/2021	2.0	Modelo Transacional	Marcos Alonso	Marcos Alonso
16/02/2021	3.0	Modelo Multidimensional	Marcos Alonso	Marcos Alonso
28/02/2021	4.0	Atualização – Script SQL	Marcos Alonso	Marcos Alonso
06/03/2021	5.0	Inserindo tabelas e imagens – Pentaho	Marcos Alonso	Marcos Alonso
07/03/2021	6.0	Elaboração de Introdução	Marcos Alonso	Marcos Alonso
14/03/2021	7.0	Relatório Dashboards e Conclusão	Marcos Alonso	Marcos Alonso

Sumário

1	INTRODUÇÃO	4
2	ESTUDO DE CASO	5
2.1	DESCRIÇÃO DO ESTUDO DE CASO.....	5
2.2	ANÁLISE DOS DADOS COLETADOS	7
3	DESCRIÇÃO DO MODELO TRANSACIONAL.....	10
3.1	REPRESENTAÇÃO SIMULADA DE MODELO TRANSACIONAL	10
4	PROPOSTA DE PROCESSO DE BI.....	11
5	MODELO MULTIDIMENSIONAL.....	12
6	ELABORAÇÃO DO DATA WAREHOUSE.....	13
6.1	DEFINIÇÃO DO DW	13
7	PROJETO DE ETL	14
7.1	DESCRIÇÃO DO PROJETO DE ETL	14
8	DASHBOARD.....	21
8.1	DESCRIÇÃO DA ELABORAÇÃO	21
8.2	TELAS DO DASHBOARD.....	21
9	CONCLUSÃO.....	26
10	ANEXOS	27
11	ARQUIVOS.....	33

1 Introdução

Segundo dados do Relatório 2018 de Status Global em Segurança Rodoviária¹, da Organização Mundial da Saúde, cerca de 1,35 milhões de pessoas morrem anualmente devido a acidentes rodoviários. Mais surpreendente, é a principal causa de mortes de crianças e jovens adultos (5-29 anos) no mundo, sendo o número de acidentes considerado excessivamente alto.

A proporção continental do Brasil, associada ao contexto econômico histórico e logísticos de maior investimento no rodoviarismo como principal malha viária, a importância dessa discussão mostra-se ainda mais relevante. Embora o número de acidentes tenha caído no último ano, a letalidade nas estradas se mantém².

Nesse contexto, o setor público lida com o desafio de criar ações de fiscalização e políticas públicas capazes de estimular melhoras nos principais índices associados ao tráfego de veículos, como o número de acidentes, feridos e mortos nas rodovias. Por conseguinte, é fundamental que as tomadas de decisões sejam apoiadas por sólidas informações obtidas a partir dos acidentes registrados nas rodovias. Face à conhecida restrição orçamentária que envolve o setor público brasileiro e em prol do princípio da eficiência na prestação do serviço público, as decisões apoiadas em dados é uma condição precípua para eficácia dos objetivos.

Tratando-se das organizações privadas, todo conhecimento que possa agregar valor às empresas deve ser tratado também como condição excepcional. Empresas que atuam no segmento de logística e que possuem, nas rodovias federais, os meios de execução de seus serviços, devem utilizar estas informações como planos para otimizar suas rotas, definir planos de ações e de conscientização internas. Neste caso, o valor agregado suscita em fidelização de clientes, redução de custos e otimização dos lucros.

O presente relatório tem como finalidade técnica, portanto, apresentar um panorama geral dos acidentes em rodovias federais no Brasil, durante o período 2017-2020, através das ocorrências registradas pela Polícia Rodoviária Federal - PRF.

Para tanto, serão utilizadas técnicas tradicionais e que perpassam por todo o processo de elaboração de um projeto de *Business Intelligence*, o que envolve o processo de extração, modelagem de um banco de dados, tratamento e conexão do *Data Warehouse* criado com uma aplicação para visualização de relatórios dinâmicos.

¹ Disponível em: <<https://www.who.int/publications/i/item/9789241565684>>

² Disponível em: <<https://agenciabrasil.ebc.com.br/geral/noticia/2021-02/numero-de-acidentes-em-rodovias-federais-cai-mas-letalidade-aumenta>>

2 Estudo de Caso

2.1 Descrição do Estudo de Caso

O estudo de caso desenvolve-se a partir da premissa da necessidade constante de elaboração de ferramentas de análise e novas formas de apresentar informações que convertam em ações e conhecimentos.

Neste cenário, o Ministério da Infraestrutura e demais órgãos interessados, como o DNIT (Departamento Nacional de Trânsito), a ANTT (Agência Nacional de Transportes Terrestres) e o CNT (Conselho Nacional de Trânsito) ensejam pelo desenvolvimento de uma ferramenta que possibilite conexão direta com os dados de ocorrências registradas em rodovias federais, por intermédio de um sistema gerenciador de banco de dados.

A figura abaixo apresenta a infraestrutura rodoviária brasileira e dado sua extensão e nível de complexidade, manifesta-se importância imediata de medidas de controle e avaliação da malha especificada no presente estudo.

Segundo dados do Ministério da Infraestrutura, a malha rodoviária federal do Brasil possui atualmente extensão total de 75.533 km³, dos quais 87% correspondem a rodovias pavimentadas e 13% a rodovias não pavimentadas. Destacam-se nesse cenário a BR-116 e a BR-101, ambas com extensão acima de 4.400 km.

Figura 1: Mapa do Sistema Federal de Viação – Rodovias – Brasil⁴



Fonte: Ministério da Infraestrutura/DNIT

³ Disponível em: < <https://www.gov.br/infraestrutura/pt-br/assuntos/transporte-terrestre/rodovias-federais/rodovias-federais-informacoes-gerais-sistema-federal-de-viacao> >

⁴ Disponível em: < <http://servicos.dnit.gov.br/dnitcloud/index.php/s/RctBFGmJxDpiHE#pdfviewer> >

Através da consolidação dos dados obtidos, será possível apresentar as principais estatísticas das ocorrências no país, mapear rodovias com maior incidência de acidentes de trânsito, servindo como subsídio técnico para investigação das causas de acidentes em determinadas regiões. Tais propósitos suscitam eventualmente ações de cunho operacional e a proposição de políticas públicas mais assertivas para redução do número de acidentes de trânsito nas rodovias.

Por conseguinte, a proposta desse projeto é elaborar uma ferramenta dinâmica, com aplicação de técnicas de automatização e facilidade de atualização, a fim de se gerar subsídios técnicos para o processo de tomada de decisão do poder público, bem como de empresas privadas que, porventura, tenha sido atribuída a responsabilidade de fiscalização e controle de rodovias por meio de concessões, nos termos da Lei Federal nº 9277/1996. Ademais, este panorama abarca também as organizações que operam na cadeia logística brasileira e possuem no rodoviarismo uma de suas formas de transporte de passageiros e produtos.

2.2 Análise dos Dados Coletados

Para fins de analisar a base de referência e captar possíveis necessidades de ajustes e tratamentos prévios, foi realizado um processo de Data Analysis utilizando a *switch* do Jupyter Notebook, na plataforma de distribuição de linguagens (Anaconda) através da linguagem Python.

O primeiro passo realizado foi unir todas os arquivos com extensão .csv (*comma separated values*), correspondentes às ocorrências ocorridas no ano de 2017, 2018, 2019 e 2020. Através dos códigos elaborados, foi possível identificar um total de 30 colunas, com seus respectivos significados descritos no quadro a seguir:

Quadro 1: Estatística descritivas das colunas numéricas

(continua)

Coluna	Descrição ⁵
<i>id</i>	Identificador do acidente. A cada ano, uma nova contagem se inicia.
<i>data_inversa</i>	Data da ocorrência no formato dd/mm/aaaa.
<i>dia_semana</i>	Dia da semana da ocorrência.
<i>horario</i>	Horário da ocorrência no formato hh:mm:ss.
<i>uf</i>	Unidades da Federação.
<i>br</i>	Coluna com valores numéricos, representando o identificador da BR onde ocorreu o acidente.
<i>km</i>	Identificador do km registrado do acidente.
<i>municipio</i>	Nome do município.
<i>causa_acidente</i>	Descrição da causa principal do acidente.
<i>tipo_acidente</i>	Descrição do tipo de acidente segunda a sua dinâmica. Exemplo: Colisão frontal, Saída de pista.
<i>classificacao_acidente</i>	Classificação quanto à gravidade do acidente. Exemplo: sem vítimas fatais, com vítimas feridas etc.
<i>fase_dia</i>	Fase do dia. Exemplo: Amanhecer, Pleno dia etc.
<i>sentido_via</i>	Sentido da via considerando o ponto de colisão – crescente ou decrescente.
<i>condicao_meteorologica</i>	Condição meteorológica no momento do acidente.
<i>tipo_pista</i>	Tipo de pista de acordo com a quantidade de faixas.
<i>tracado_via</i>	Descrição do traçado da via.
<i>uso_solo</i>	Descrição sobre as características do local do acidente: Urbano = sim; Rural = não.
<i>peessoas</i>	Número de pessoas envolvidas no acidente.
<i>mortos</i>	Número de pessoas mortas.
<i>feridos_leves</i>	Número de feridos leves.
<i>feridos_graves</i>	Número de feridos graves.
<i>ilesos</i>	Número de ilesos.
<i>ignorados</i>	Número de ignorados. Envolvidos que não se soube o estado físico.
<i>feridos</i>	Número de feridos.

⁵ Informações obtidas através do dicionário de variáveis disponibilizados pela Polícia Rodoviária Federal.

Quadro 1: Estatística descritivas das colunas numéricas

(continuação)

Coluna	Descrição
<i>veiculos</i>	Número de veículos envolvidos na ocorrência.
<i>latitude</i>	Latitude do local de acidente.
<i>longitude</i>	Longitude do local de acidente.
<i>regional</i>	Regional de cobertura.
<i>delegacia</i>	Identificação da delegacia.
<i>uop</i>	Unidade Operacional (Posto de Fiscalização PRF).

Fonte: Elaborada pelo autor através do Bizagi Modeler

Ademais, além dos trinta tributos que caracterizam cada acidente, o período de 2017-2020, segundo informações extraídas da base, registrou um total de 289.751 acidentes. A tabela a seguir demonstra as principais estatísticas descritivas extraídas, sendo primordial para identificação de possíveis problemas na base e detecção de *outliers*.

Cumpramos esclarecer a decisão por não determinar previamente os tipos de dados em cada coluna, de forma que a própria biblioteca 'Pandas', com a função de leitura, interpretou o tipo de dado em cada coluna. No entanto, em outros tipos de casos e análises, que envolvam, por exemplo, CPF (Cadastro de Pessoa Física) a definição faz-se necessário, principalmente para fins da utilização para matches ou tratamento na coluna – isso porque o zero à esquerda nos CPF's precisa ser lido como *string* para ser identificado.

Tabela 1: Estatística descritivas das colunas numéricas

	id	pessoas	mortos	feridos_leves	feridos_graves	ilesos	ignorados	feridos	veiculos
count	289751.000000	289751.000000	289751.000000	289751.000000	289751.000000	289751.000000	289751.000000	289751.000000	289751.000000
mean	166506.968683	2.343191	0.076404	0.826078	0.248662	1.058188	0.133860	1.074740	1.640050
std	96266.732051	1.858060	0.329376	1.086012	0.595844	1.431567	0.412079	1.220477	0.737317
min	8.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
25%	82260.500000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
50%	167388.000000	2.000000	0.000000	1.000000	0.000000	1.000000	0.000000	1.000000	2.000000
75%	250365.500000	3.000000	0.000000	1.000000	0.000000	1.000000	0.000000	1.000000	2.000000
max	334959.000000	80.000000	21.000000	58.000000	28.000000	73.000000	16.000000	66.000000	23.000000

Fonte: Elaborada pelo autor através do Bizagi Modeler

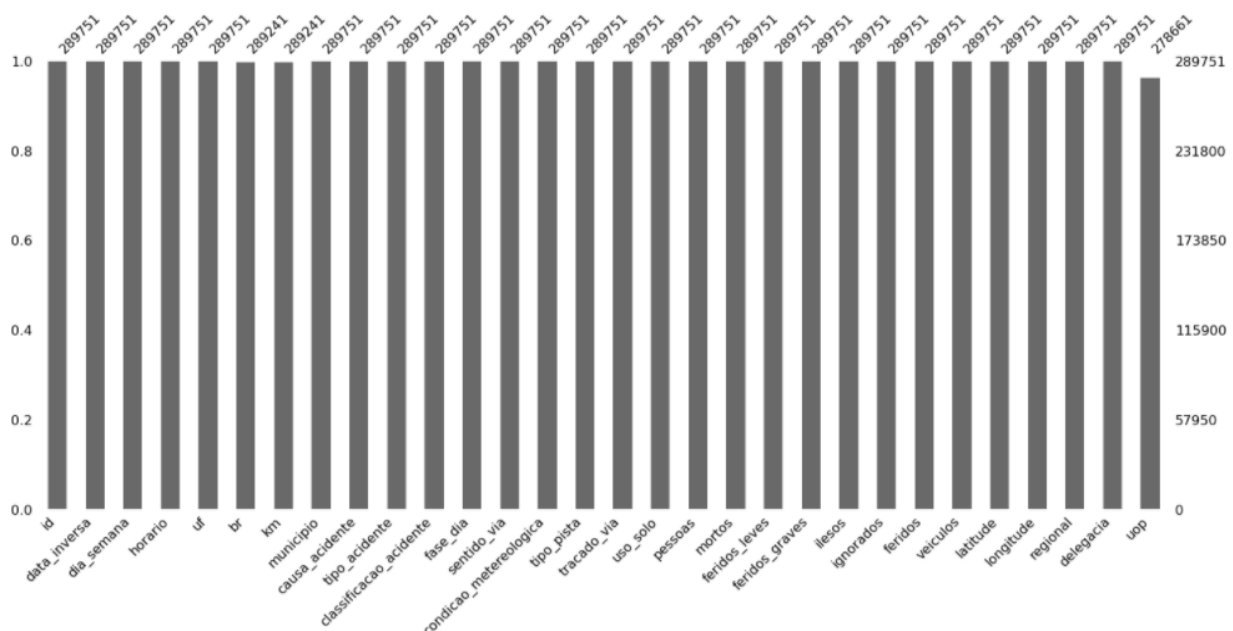
A partir das informações acima, identificou-se, por exemplo, a característica de recontagem dos identificadores de acidentes ao longo dos anos, o que impede, por exemplo, a utilização imediata desse identificador como chave primária nos casos em que se agregam diferentes períodos de observação.

Também foi identificado registro de acidentes com alto número de pessoas, envolvendo especificamente 80 indivíduos. Da mesma forma, nesse período, foi registrado acidente de grandes proporções, envolvendo 23 veículos. Cumpramos esclarecer que os dois registros não necessariamente envolvem o mesmo acidente, uma vez que as estatísticas descritivas são calculadas de acordo com cada um dos atributos (colunas) que caracterizam o acidente.

Através da aplicação da biblioteca 'Missingno', foi feita verificação de todas as variáveis envolvidas na base de dados em relação aos dados faltantes. Assim, a figura a seguir demonstra que as variáveis 'br', 'km', 'delegacia' e 'uop' estavam incompletas. Entre as quatro variáveis, apenas a variável 'br' possui características de interesse direta na elaboração do trabalho. Assim, por se tratar de uma variável categórica, optou-se por reunir todas as informações faltantes das rodovias em uma categoria 'Não Identifica' a ser criada durante o processo de ETL.

Essa decisão é sustentada por ser possível identificar tais acidentes por outras medidas de localização, como o estado e o município. Optou-se, portanto, pela não exclusão das observações com dados faltantes em variáveis consideradas importantes. A relação dos quantitativos de observações por coluna da base de dados pode ser observada a seguir:

Figura 2: Dados Faltantes – Base de Dados PRF



Fonte: Elaborada pelo autor através do Bizagi Modeler

Além disso, identificou-se através das análises descritivas um possível erro pontual em três localizações de acidentes com base nas colunas de latitude e longitude. Os valores não estavam separados por casas decimais, o que acusou valores excessivamente altos para as localizações. Como tratava-se de casos excepcionais, por uma questão de otimização, foi realizada alteração diretamente no arquivo compilado dos acidentes (.csv).

Faz-se necessário salientar que possíveis erros de localização são passíveis de ocorrer além daqueles identificados nas análises descritivas. No entanto, face à restrição temporal e impacto pequeno nos resultados gerais, esses ajustes não são de imediata atenção do trabalho.

Em suma, a análise de dados neste primeiro momento tem caráter avaliativo e corretivo da base, sendo as análises aprofundadas das informações e a maior desagregação do panorama estabelecidas nos *dashboards* finais, por intermédio do Power BI.

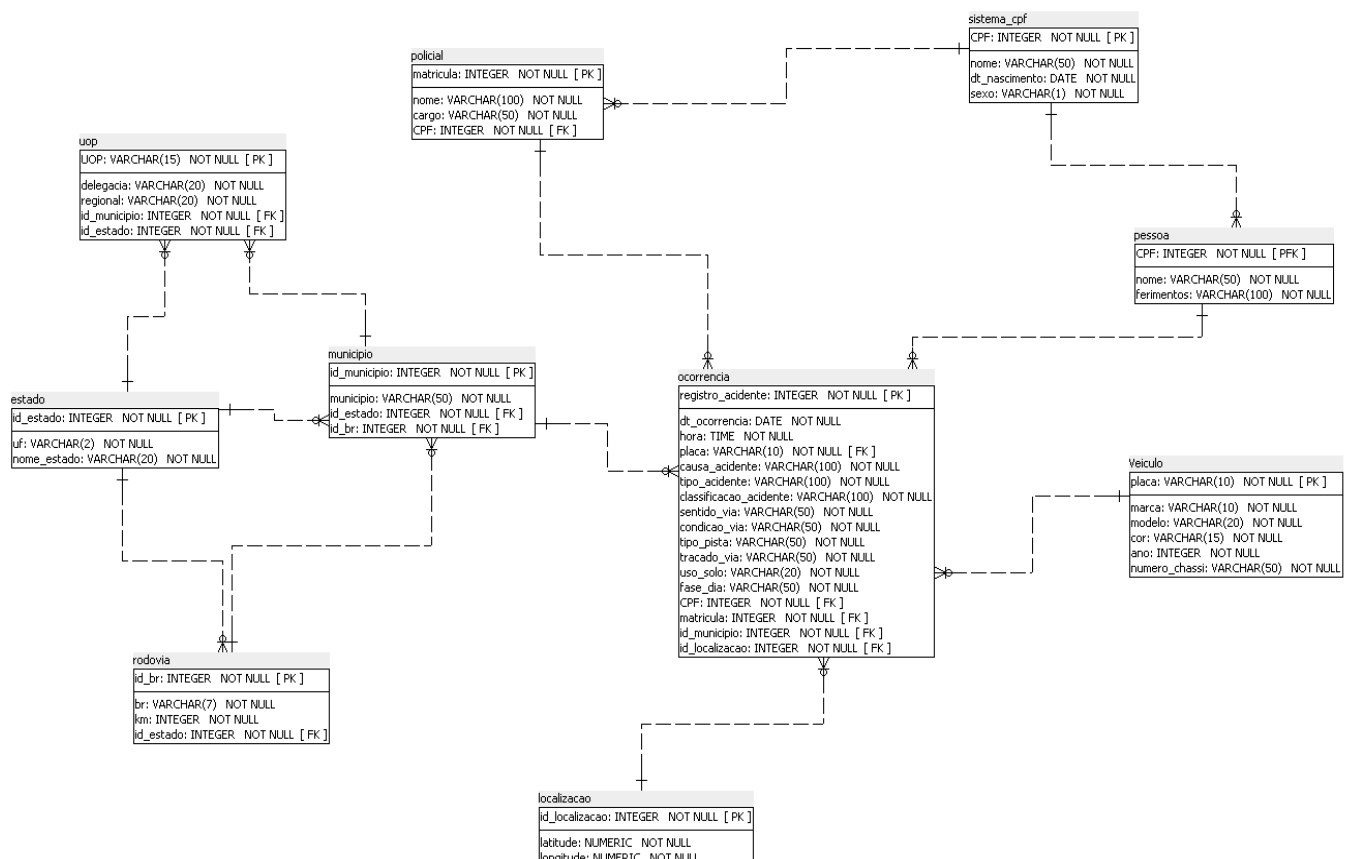
3 Descrição do Modelo Transacional

Os dados utilizados para o presente relatório foram coletados diretamente do sistema de dados abertos da Polícia Rodoviária Federal⁶, extraídos em planilhas com extensão .csv. Ressalta-se que estas informações são agrupadas por ocorrência e por pessoa, sendo que, por motivos de informações incompletas, em grande volume, na base agregada por pessoa, optou-se pela agregação por ocorrência. Embora o problema de dados faltantes ou incorretos seja uma realidade, a dificuldade de consultar e corrigir as informações problemáticas, bem como essa falta ocorrer em informações importantes para uma análise de perfil, sustentam a escolha definida.

3.1 Representação simulada de modelo transacional

Uma vez que os dados com registro das ocorrências foram obtidos diretamente do site do Governo Federal (Polícia Rodoviária Federal), com as informações necessárias ao estudo já detalhadas e divulgadas unificadamente, não existe um modelo transacional de referência. No entanto, para fins didáticos, foi realizada a seguir uma possível representação de um modelo transacional a partir do registro de ocorrência realizada pela equipe da PRF responsável.

Figura 3: Modelo Transacional Simulado



Fonte: Elaborada pelo autor através do SQL Power Architect.

⁶ Disponível em: < <https://www.gov.br/prf/pt-br/acesso-a-informacao/dados-abertos/dados-abertos-acidentes> >

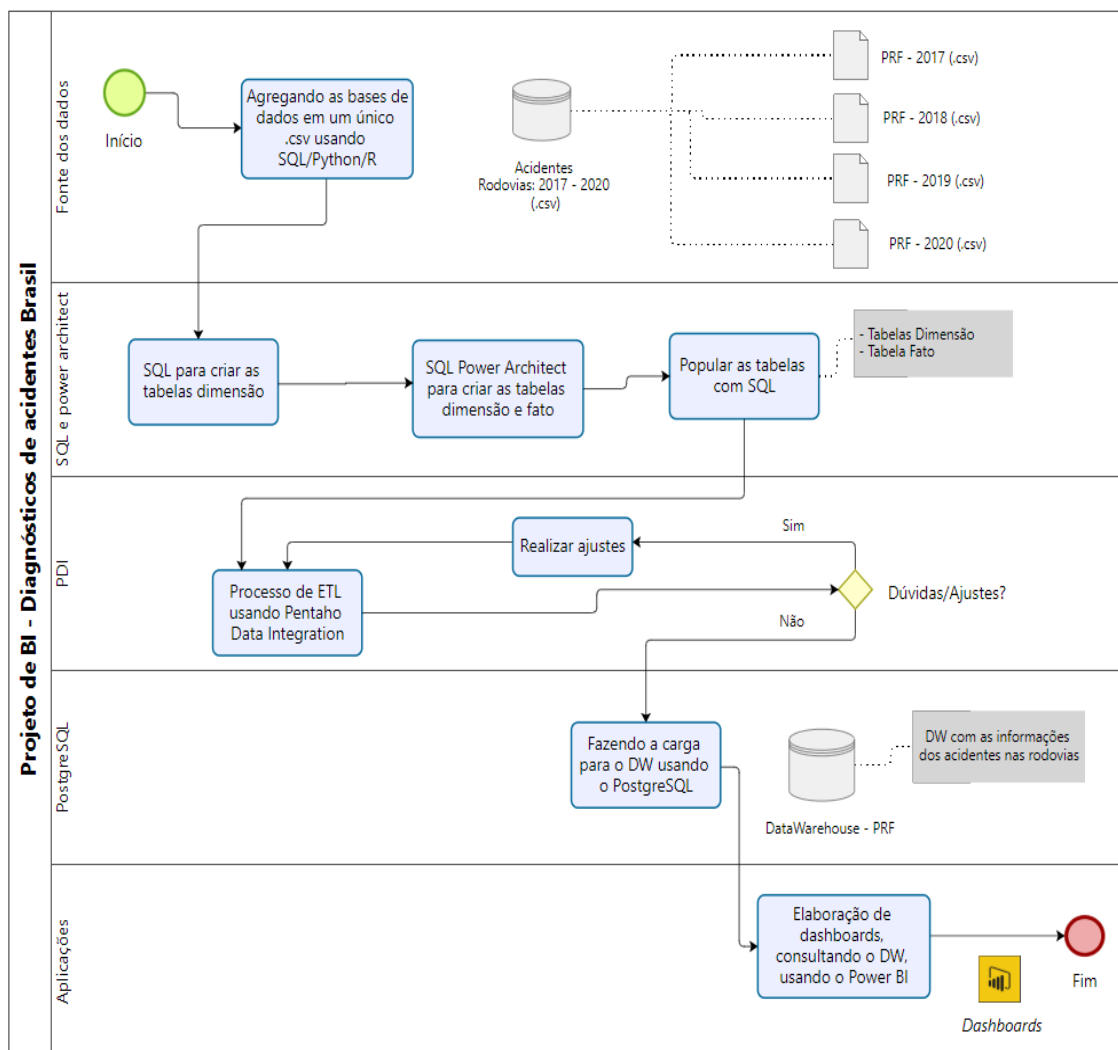
4 Proposta de Processo de BI

Esta seção apresenta o processo de BI proposto para o projeto. Por intermédio da elaboração do fluxo de trabalho, é possível determinar a sequência de atividades de um grupo técnico para elaboração do produto ou serviço acordado com o cliente. Assim, entende-se como atividade precípua a fim de estabelecer um planejamento com maior acurácia, bem como permitir a compreensão do projeto para todos os *stakeholders* envolvidos.

Assim, foi utilizado o software Bizagi Modeler para o mapeamento dos processos e definição do fluxo de trabalho – *workflow* – a ser desenvolvido no projeto de elaboração do panorama de acidentes no Brasil. A proposta de abordagem envolve a extração dos dados abertos, a criação do modelo multidimensional e a definição da arquitetura do *Data Warehouse*. Neste contexto, desenvolve-se naturalmente os processos de ETL – *Extract, Transform, Load*.

O resultado do fluxo de atividades pode ser verificado na figura a seguir:

Figura 4: Workflow do processo de BI



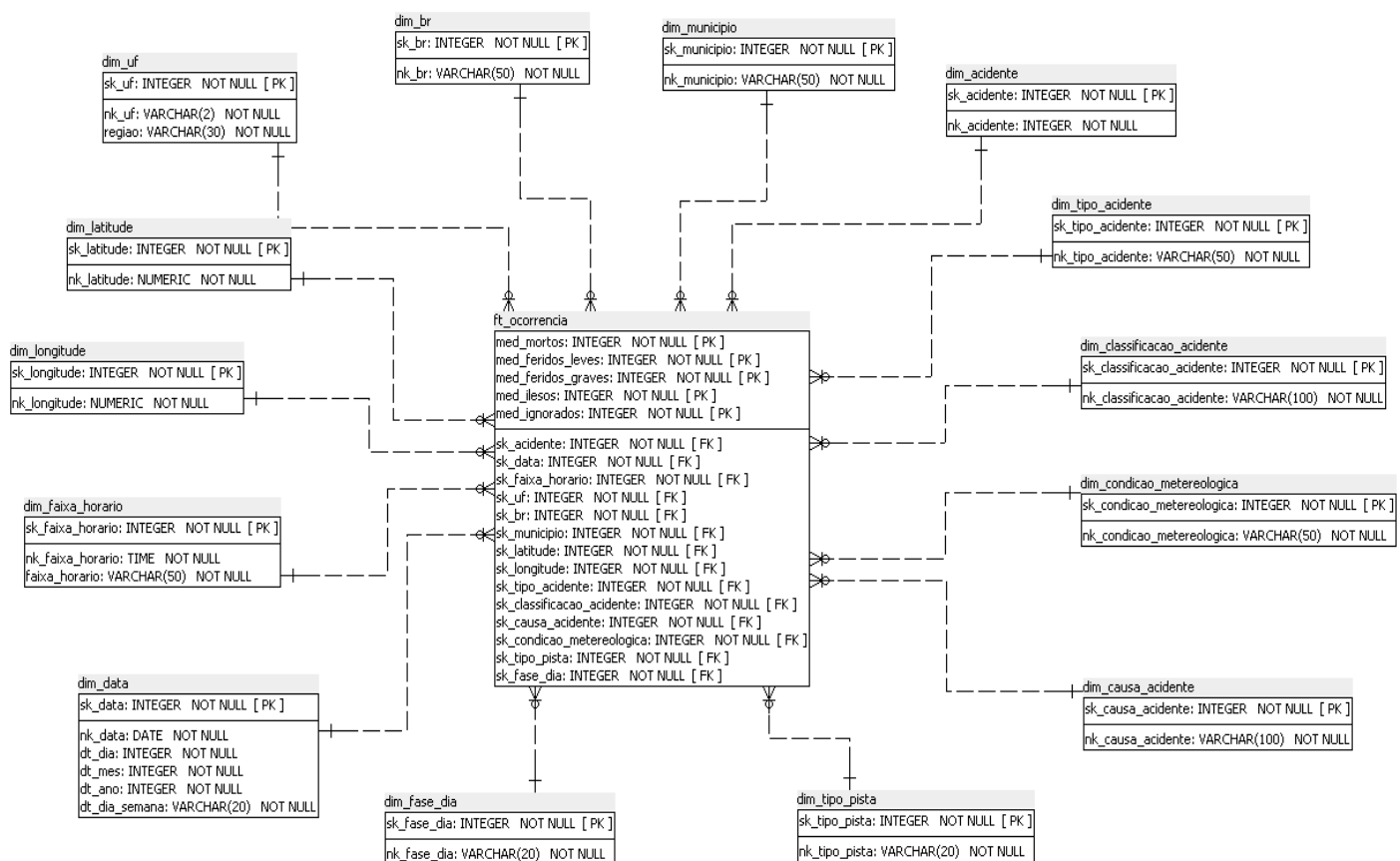
Fonte: Elaborada pelo autor através do Bizagi Modeler.

5 Modelo Multidimensional

A técnica de modelagem do banco de dados contendo as informações do Panorama de Acidentes em Rodovias Federais no Brasil | 2017-2020 a ser aplicada é a multidimensional. Neste tipo de modelagem, destacam-se os modelos Estrela (*Star Schema*) e Floco de Neve (*Snow Flake*), sendo o modelo estrela o escolhido para aplicar no banco de dados nas finalidades consultivas do *Data Warehouse*.

Embora a criação das tabelas diretamente na ferramenta que hospedará o banco de dados – neste caso, o PostgreSQL – seja relativamente simples, para efeito de otimização, foi utilizado o software SQL Power Architect para criação das tabelas. A figura abaixo representa as tabelas dimensão e a tabela fato responsável por registrar as ocorrências desenvolvidas nas rodovias federais.

Figura 5: Modelo Multidimensional – Star Schema



Fonte: Elaborada pelo autor através do SQL Power Architect.

6 Elaboração do Data Warehouse

O *Data Warehouse* será a fonte integradora de informações da empresa, a tecnologia será utilizada com o intuito de servir de base para a camada de aplicação que será responsável por fornecer dados para a tomada de decisão na organização.

Devido às características de elaboração do boletim de ocorrência, seria plenamente possível registrar cada pessoa no sistema envolvido em um acidente específico. Contudo, isso teria um impacto significativo no volume do banco de dados, além do aumento de possibilidades de registros equivocados sem ganho técnico significativo a respeito da contabilização de acidentes.

Após a verificação da atual base de dados, entendeu-se como a melhor opção a granulosidade alta justamente pelo alto número de registros de idade errôneos ou a falta deles. Assim, não seria possível uma análise mais adequada dos perfis dos usuários envolvidos. Optou-se, portanto, registrar apenas o identificador do acidente.

6.1 Definição do DW

6.1.1 Arquitetura

A arquitetura definida do DW para este projeto será a Global Centralizada, uma vez que o propósito é obter as informações diretamente em consulta ao sistema de ocorrências da PRF ou através de sua disponibilização via dados abertos, para as análises dos possíveis *stakeholders*, como por exemplo, a administração do DW pelo departamento de TI do Ministério da Infraestrutura.

6.1.2 Abordagem de Construção

Uma vez que o projeto consiste na elaboração de um único *Data Warehouse* consultivo para a tomada de decisões a nível de diagnósticos de acidentes nas rodovias brasileiras, a consulta dessas informações se dá diretamente do sistema de registro de ocorrências. Assim não se aborda os conceitos de abordagem e, portanto, não se aplica a definição de uma abordagem de construção do DW.

6.1.3 Arquitetura Física

No que diz respeito à arquitetura física do *Data Warehouse*, optou-se pela arquitetura em nuvem pelos motivos expostos a seguir.

Primeiramente, trata-se de um banco com propósito de expansão e a arquitetura por um *Virtual Private Service* permite rápido escalonamento. Ao mesmo tempo, a ideia de elaboração envolve gerar produtos com os recursos disponíveis atualmente, o que no caso de um Datacenter envolveria um vultuoso investimento inicial. Por fim, por ser um projeto piloto, a ideia é que seja flexível e seguro, o que um DW na nuvem oferece com maior facilidade.

7 Projeto de ETL

7.1 Descrição do Projeto de ETL

Para o processo de Extração, Transformação e Carga, foi utilizado a switch do Pentaho - Pentaho Data Integration (PDI). Primeiramente, faz-se necessário observar que o processo de transformação envolvendo a carga dos dados ocorreu de duas maneiras diferentes, por justificativas técnicas e de aprendizagem.

A justificativa técnica esteve na necessidade de realizar uma análise dos dados, por intermédio do Jupyter Notebook e a linguagem Python, bem como a alteração local do arquivo 'dados17_20_datatran.csv' em relação aos dados de localização (latitude e longitude), conforme comentado anteriormente. No entanto, para fins de aprendizado, optou-se também pelo processo de carga individual de cada arquivo anual com a relação de acidentes (datatran2017.csv, datatran2018.csv, datatran2019.csv, e datatran2020.csv) e a realização de um *merge* desses dados.

Ademais, é necessário salientar que após a realização do processo de ETL e a carga no *Data Warehouse* criado no PostgreSQL, com maior experiência adquirida, identificou-se processos que podem ser plenamente otimizados como, por exemplo, em um mesmo arquivo para criação de dimensão, ser gerado uma estrutura da tabela fato. Isso ocorreria pela criação de uma ramificação – com output de um arquivo com a base tratada – antes do processo de remover duplicatas para gerar a respectiva dimensão.

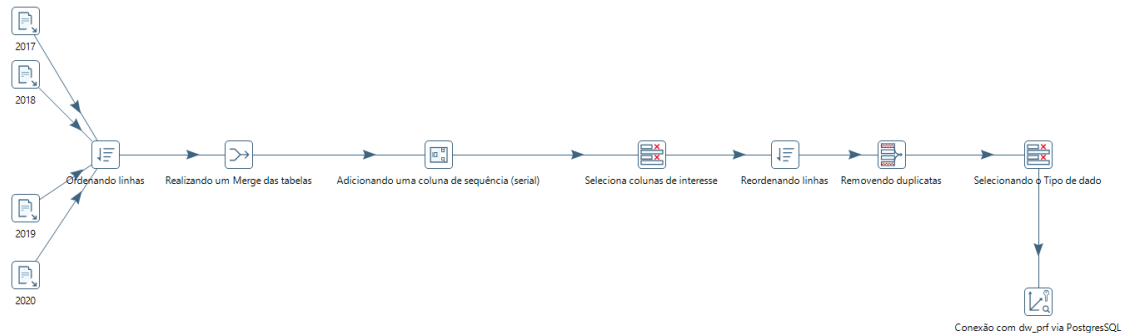
Concomitantemente, algumas atividades podem ser realizadas em um único passo, o que otimiza a leitura dos *steps* pelo PDI, isso economiza recursos de tempo e de processamento da máquina.

Isso posto, todas essas etapas de reflexão supramencionadas tiveram efeito de aprendizagem e, por esse motivo, aqui explicitadas. Desde seu princípio, o processo de ETL foi pensado para ser eficiente e eficaz. Notou-se após maior experiência com esse tipo de trabalho pontos de melhoria na eficiência, mas o processo foi eficaz no sentido de gerar as tabelas dimensão e a tabela fato, via conexão local, com o DW criado no PostgreSQL.

A seguir, serão explicitas de forma sucinta o processo de ETL nas tabelas dimensão e fato. Todos os arquivos com extensão ".ktr" serão anexadas, bem como o script em Python para a concatenar as tabelas anuais de ocorrências e o script em SQL para realização de *queries* e aprendizagem da linguagem.

- a) Dimensão Acidente – Para esta tabela, a intenção era gerar uma coluna de sequência com o registro de ID's das ocorrências. Neste caso, não seria possível usar a referência de ID's dos arquivos obtidos junto ao site da Polícia Rodoviária Federal, pois estes valores reiniciam a contagem a cada ano.

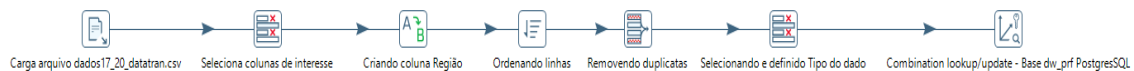
Figura 6: Dimensão – dim_acidente



Fonte: Elaborada pelo autor através do PDI.

- b) Dimensão UF – Para este caso, foi realizada a criação de uma coluna Região, que por definição, corresponde a região da respectiva Unidade Federativa.

Figura 7: Dimensão – dim_uf



Fonte: Elaborada pelo autor através do PDI.

- c) Dimensão Causa do Acidente – Neste caso, foi realizada apenas a seleção das colunas de interesse e a remoção das duplicatas para geração da tabela dimensão.

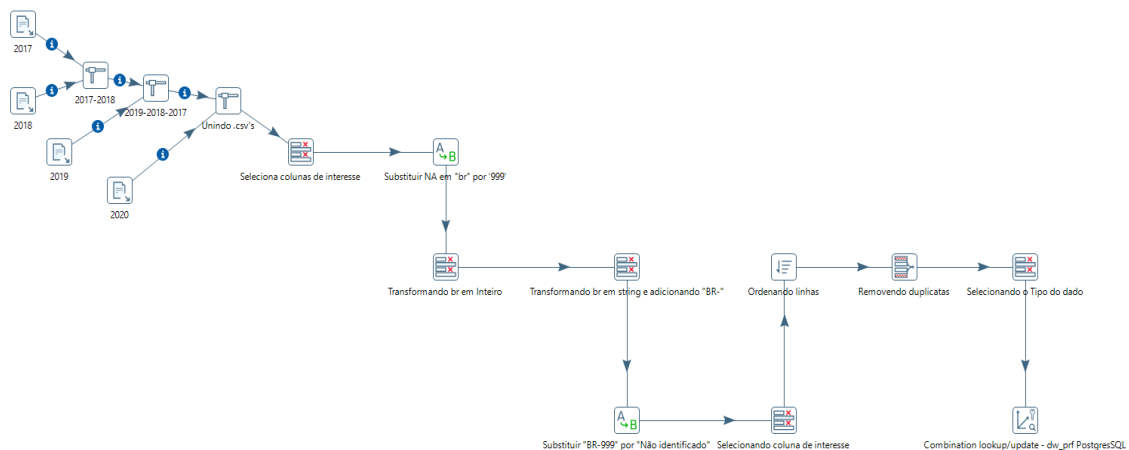
Figura 8: Dimensão – dim_causa_acidente



Fonte: Elaborada pelo autor através do PDI.

- d) Dimensão BR – Para este caso, foi criada a dimensão contendo as respectivas rodovias federais. Para tanto, utilizou-se um processo de *merge* de dois arquivos através do passo *Append Streams*, que difere do passo *Multiway Merge Join* principalmente por realizar o merge de apenas dois arquivos. Neste processo, foi realizada a substituição de valores vazios por um número de rodovia não existente (999), cujo objetivo está na alteração do formato da coluna para ter três números (a coluna br, por exemplo, tratava a BR-040 como 40. Contudo, foi realizada a substituição de BR-999 por “Não Identificado” mais para frente.

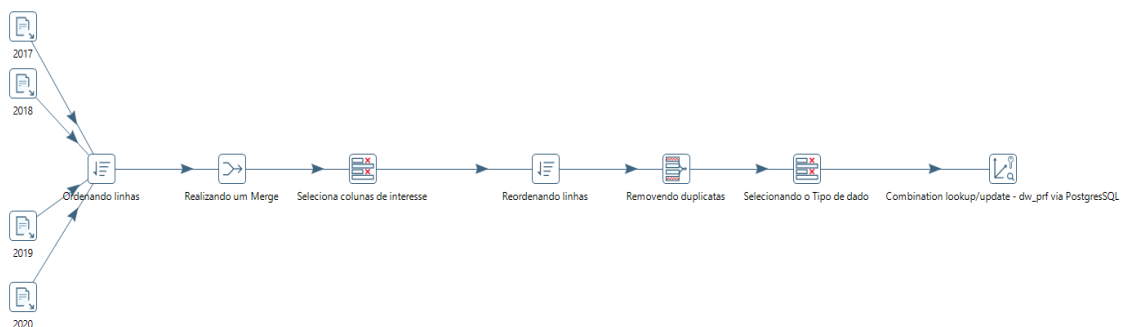
Figura 9: Dimensão – dim_br



Fonte: Elaborada pelo autor através do PDI.

- e) Dimensão Classificação do Acidente - Neste caso, foi realizada apenas a seleção das columnas de interesse e a remoção das duplicatas para geração da tabela dimensão.

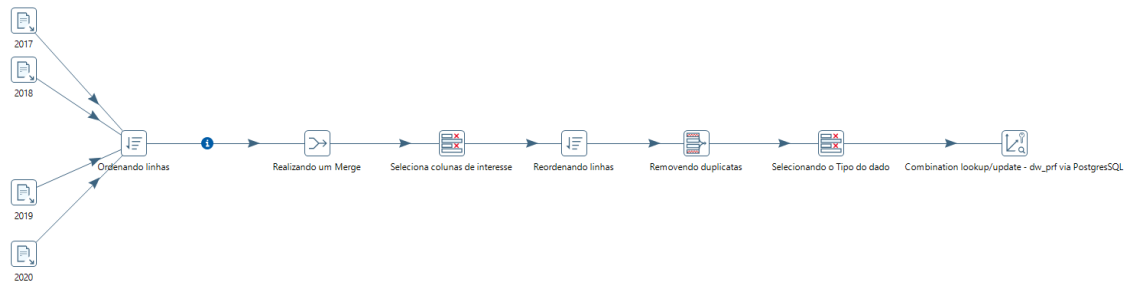
Figura 10: Dimensão – dim_classificacao_acidente



Fonte: Elaborada pelo autor através do PDI.

- f) Dimensão Condição Meteorológica - Neste caso, foi realizada apenas a seleção das colunas de interesse e a remoção das duplicatas para geração da tabela dimensão.

Figura 11: Dimensão – dim_condicao_meteorologica



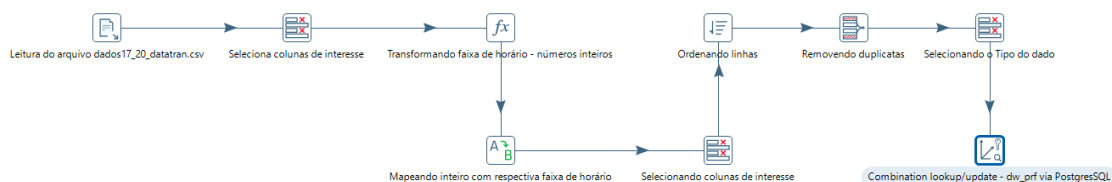
Fonte: Elaborada pelo autor através do PDI.

- g) Dimensão Faixa de Horário – Para a criação de uma coluna com faixa de horário, foi utilizada uma fórmula para extrair da coluna horário o número inteiro correspondente ao valor da hora registrada no horário. Assim, foi criada uma coluna a partir do mapeamento do valor do número inteiro da hora distribuída em quatro faixas:

- 00:00 – 06:00 se o número inteiro entre 0 e 6;
- 06:00 – 12:00 se entre 6 e 12;
- 12:00 – 18:00 se entre 12 e 18; e
- 18:00 – 23:00 se entre 18 e 23.

Após o processo, foi feita a seleção das colunas de interesse, o ordenamento das linhas para a remoção das duplicatas, gerando a respectiva tabela dimensão.

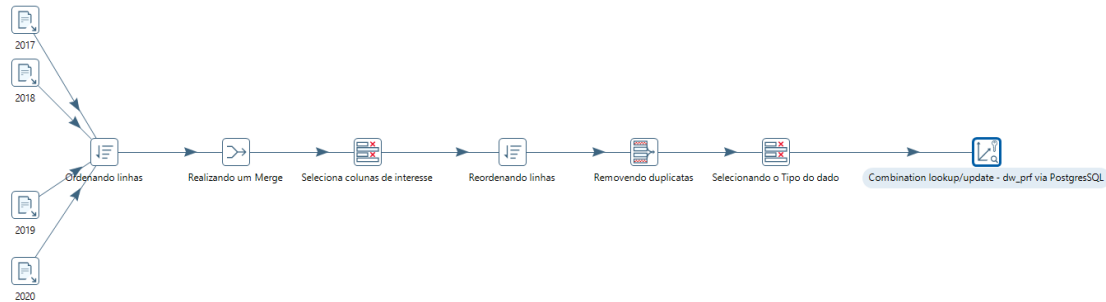
Figura 12: Dimensão – dim_faixa_horario



Fonte: Elaborada pelo autor através do PDI.

- h) Dimensão Fase do Dia – Neste caso, foi realizada apenas a seleção das colunas de interesse e a remoção das duplicatas para geração da tabela dimensão.

Figura 13: Dimensão – dim_fase_dia



Fonte: Elaborada pelo autor através do PDI.

- i) Dimensão Latitude – Neste caso, foi realizada apenas a seleção das colunas de interesse e a remoção das duplicatas para geração da tabela dimensão. Cumpre esclarecer, conforme mencionado anteriormente, que a identificação de três observações divergentes em relação à latitude ocorreu no processo de análise de dados e sua correção realizada localmente no arquivo fonte.

Figura 14: Dimensão – dim_latitude



Fonte: Elaborada pelo autor através do PDI.

- j) Dimensão Longitude – Neste caso, foi realizada apenas a seleção das colunas de interesse e a remoção das duplicatas para geração da tabela dimensão. Cumpre esclarecer, conforme mencionado anteriormente, que a identificação de três observações divergentes em relação à longitude ocorreu no processo de análise de dados e sua correção realizada localmente no arquivo fonte.

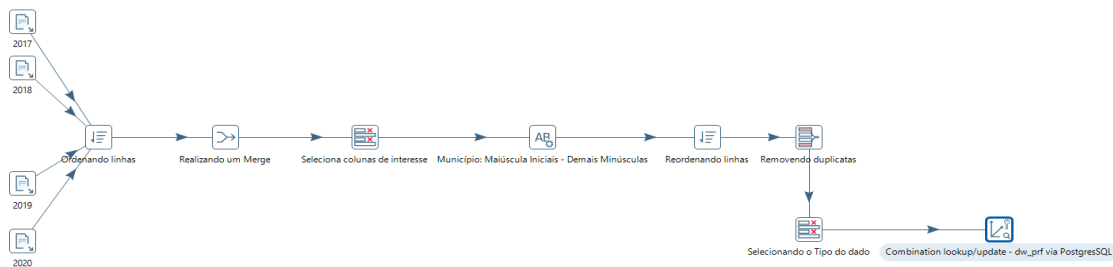
Figura 15: Dimensão – dim_longitude



Fonte: Elaborada pelo autor através do PDI.

- k) Dimensão Município – Para a criação da tabela dimensão município, o processo de transformação envolveu a necessidade de transformar os nomes dos municípios. A coluna município estava com a descrição em letras maiúsculas, sendo realizada a transformação para que apenas as letras iniciais tivessem essa característica. Após o processo, foi realizado o ordenamento das linhas e remoção das duplicatas para a criação da tabela.

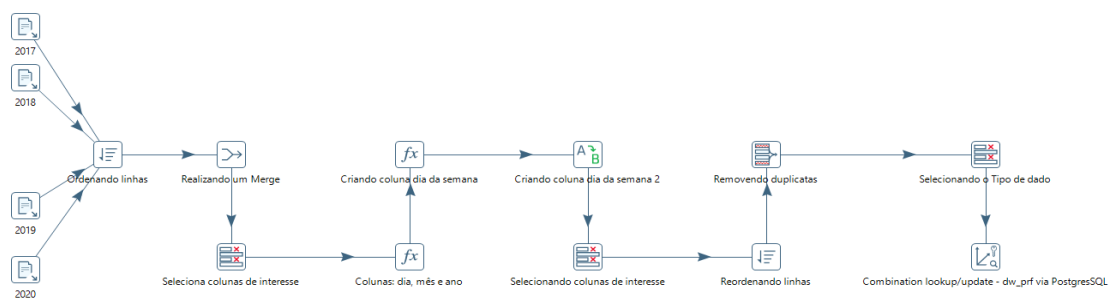
Figura 16: Dimensão – dim_municipio



Fonte: Elaborada pelo autor através do PDI.

- l) Dimensão Tempo – Na elaboração da tabela dimensão relativa ao tempo/data, foram aplicadas fórmulas para criação das colunas separando as datas em dia, mês e ano. Em seguida, foi criada uma coluna especificando o dia da semana, sendo primeiro necessário gerar o número inteiro correspondente ao dia, aplicando posteriormente um passo de mapeamento para definição do dia da semana escrito. Após os referidos processos descritos, ordenou-se as linhas e foram realizadas as remoções das duplicatas para a geração da tabela.

Figura 17: Dimensão – dim_tempo



Fonte: Elaborada pelo autor através do PDI.

- m) Dimensão Tipo de Acidente - Neste caso, foi realizada apenas a seleção das colunas de interesse e a remoção das duplicatas para geração da tabela dimensão.

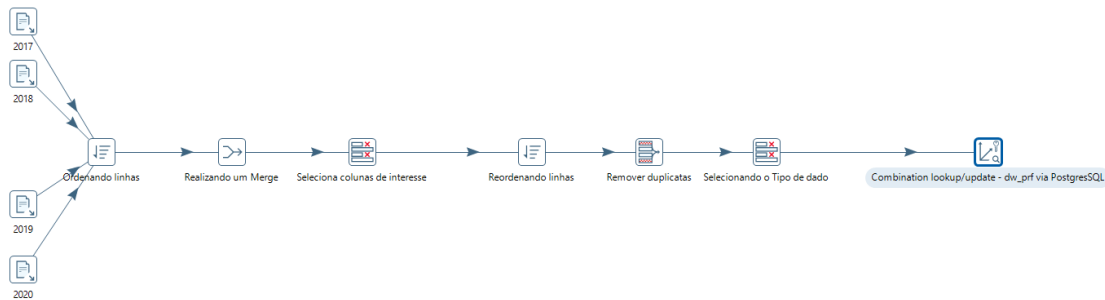
Figura 18: Dimensão – dim_tipo_acidente



Fonte: Elaborada pelo autor através do PDI.

- n) Dimensão Tipo de Pista - Neste caso, foi realizada apenas a seleção das colunas de interesse e a remoção das duplicatas para geração da tabela dimensão.

Figura 19: Dimensão – dim_tipo_pista

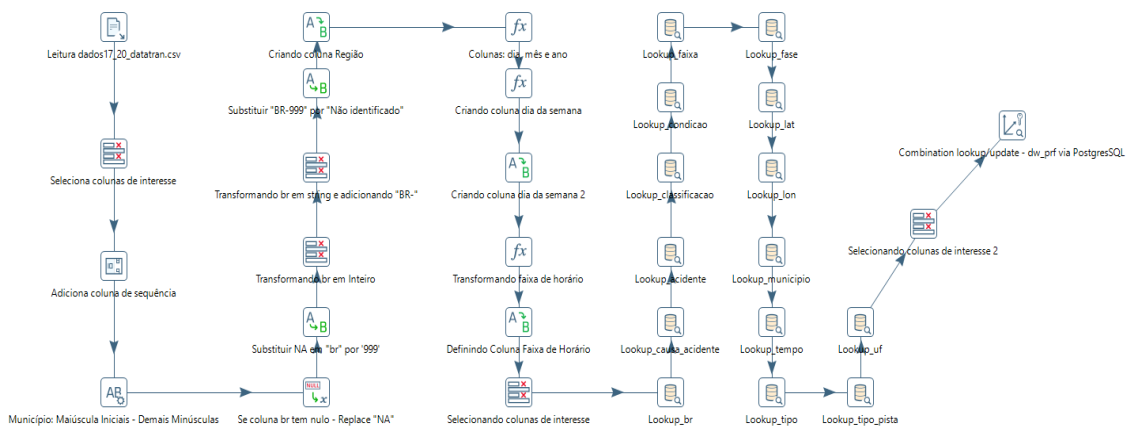


Fonte: Elaborada pelo autor através do PDI.

- o) Fato Ocorrência – Para a elaboração da tabela fato ocorrência, foram feitas todas as alterações comuns nas tabelas dimensão, sendo apenas registrados todos os passos em uma única tabela. Em termos de otimização, essas alterações poderiam ser realizadas no processo de criação das tabelas dimensão, conforme relatado anteriormente. Ainda assim, o registro abaixo abarca todas as transformações realizadas e tiveram como resultado o preenchimento completo das informações no DW dw_prf. Através do passo *Lookup*, a fato foi elaborada a partir da construção das *Surrogate Key*, chaves primárias de cada tabela dimensão.

Ademais, foram selecionadas as colunas de feridos leves, graves, mortos, ilesos e ignorados (todas tabelas com números inteiros), a partir da tabela dados17_20_datatran.csv, sendo utilizadas como medidas.

Figura 20: Fato – ft_ocorrencia



Fonte: Elaborada pelo autor através do PDI.

8 Dashboard

8.1 Descrição da Elaboração

Com o *Data Warehouse* produzido e o processo de carga do DW no PostgreSQL, foi feita a conexão desse SGBD com o software Power BI, ferramenta escolhida para elaboração dos *dashboards* de acidentes nas rodovias federais brasileiras. A escolha desta ferramenta está fundamentada em sua popularização e fácil utilização para diferentes níveis de usuários, o alto volume de possibilidades de fornecer visualizações interativas e seu custo-benefício em relação às outras ferramentas conhecidas do mercado.

Para isso, foi feita uma conexão local entre Power BI e a porta local do PostgreSQL. Após a identificação do DW e a obtenção de dados, foi realizada a carga dos dados para início da elaboração dos relatórios. Uma vez pensado em termos de otimização, a carga dos valores no Power BI exigiu poucas alterações diretas no Power BI, estando mais restritas as atividades de elaboração de medidas por meio da linguagem DAX e o desenvolvimento dos gráficos, tabelas e cartões de interesse, além do processo de interatividade do usuário.

8.2 Telas do Dashboard

A seguir, será explicado de maneira sucinta todas as telas elaboradas no Painel de Acidentes em Rodovias Federais, detalhando as informações obtidas a partir dos dados compilados.

- Menu – Tela de apresentação inicial com dois botões: início do relatório e informativo de identificação da fonte dos dados e do autor.

Figura 21: Tela Menu Inicial



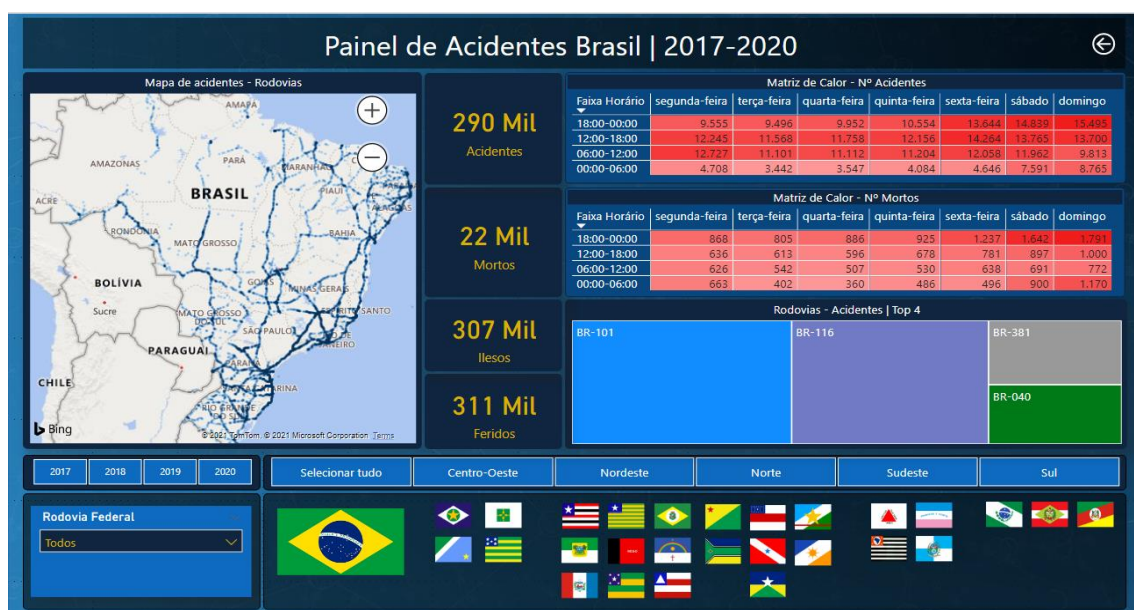
Fonte: Elaborada pelo autor através do Power BI.

- Brasil – Rodovias – Após clicar no botão de início do relatório, abre-se o painel de acidentes com resumos das ocorrências no Brasil. No painel é possível obter detalhes por ano, por rodovia e região, sendo ainda explicitados duas matrizes de calor, com informativos das faixas de acidentes e dias da semana com maiores registros de acidentes.

É possível perceber, por exemplo, que o maior número de acidentes, se considerado os dados de todas as UF's, ocorre durante os fins de semana, entre 18:00 e 00:00. É possível perceber também que nesse intervalo, o número de mortalidades é mais elevado, tornando-se o período mais intenso e perigoso devido ao fluxo de veículos e características de luminosidade do horário.

Percebe-se também que, através dos cartões explicitados e dos filtros por ano, o número de acidentes está em tendência de queda, mas o número de mortos não apresenta o mesmo comportamento. Ademais, através da criação de indicadores, é possível obter cortes detalhados das ocorrências por estados, por meio de elaboração de uma única tela, conforme a ser demonstrado na tela seguinte.

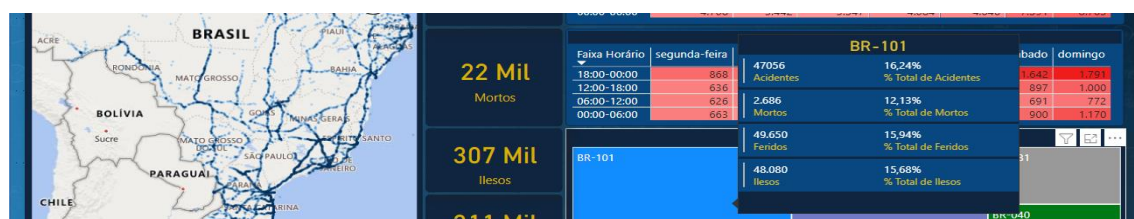
Figura 22: Tela Brasil - Rodovias



Fonte: Elaborada pelo autor através do Power BI.

Foi elaborada também uma ferramenta de interação dos dados, com informações do quantitativo de acidentes, mortos, feridos e ileso, bem como respectivos percentuais em relação ao volume total. Essas ferramentas de interação também estão presentes em outras telas do painel.

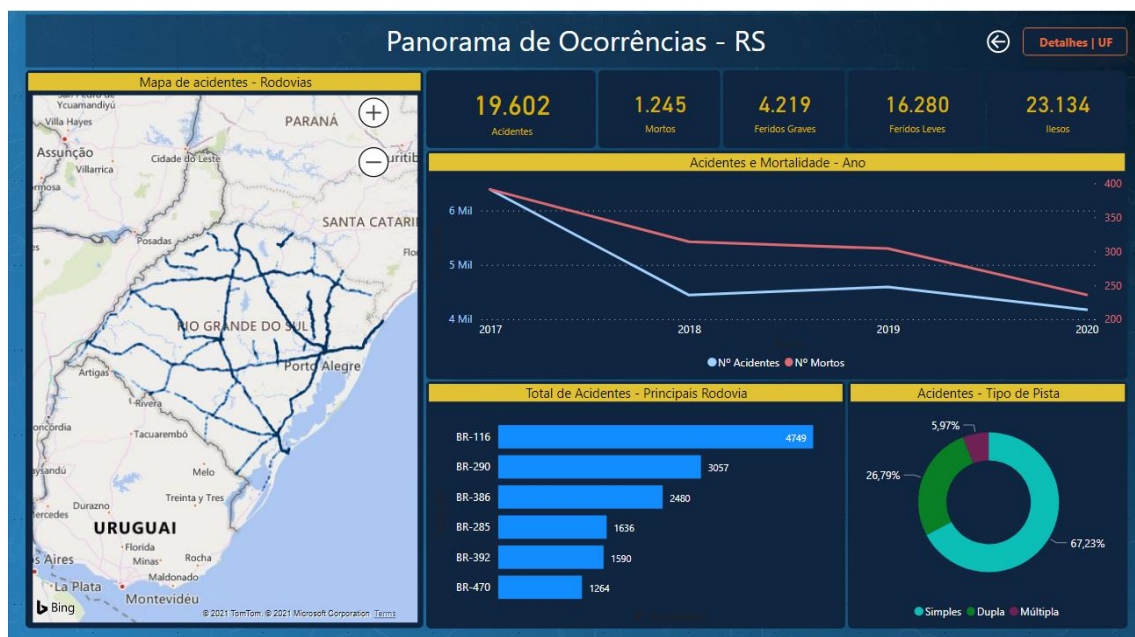
Figura 23: Tela Brasil - Rodovias



Fonte: Elaborada pelo autor através do Power BI.

- UF – Ao clicar na respectiva bandeira (que pode ser identificada passando o cursor por cima da bandeira), é possível obter um corte detalhado do estado em relação aos quantitativos de acidentes, uma série temporal dos registros de ocorrências e números de mortos. Um mapa interativo permite a identificação dos trechos com rodovias federais, sendo ainda possível obter informações das rodovias com maiores registros de acidente e características do tipo de pista. É possível perceber, por exemplo, que o estado do Rio Grande do Sul, unidade federativa com alto índice de acidentes e mortalidade, possui mais de 67% de seus acidentes em trechos com pista Simples.

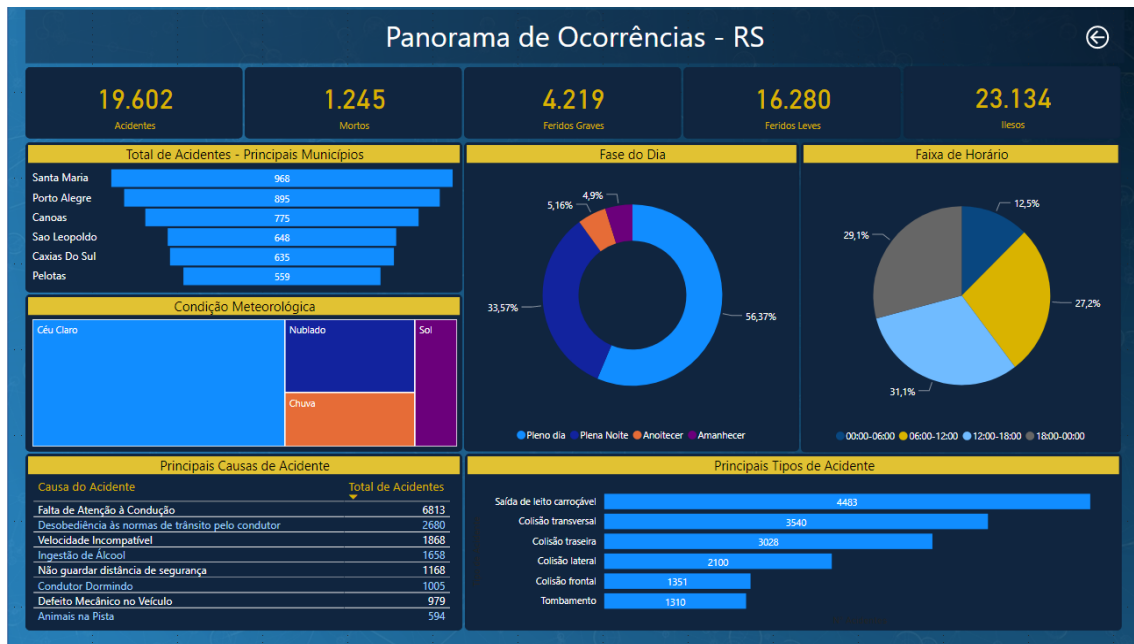
Figura 24: Tela UF



Fonte: Elaborada pelo autor através do Power BI.

- Detalhes UF – Ao clicar no botão Detalhes | UF, é possível continuar a análise dos dados no estado, mas sob um ponto de vista preponderante da característica do acidente. Para tanto, foi utilizado o botão de segmentação de dados de sincronização, na guia Exibição. É possível conferir o resultado na Figura 25 a seguir.

Figura 25: Tela Detalhes UF

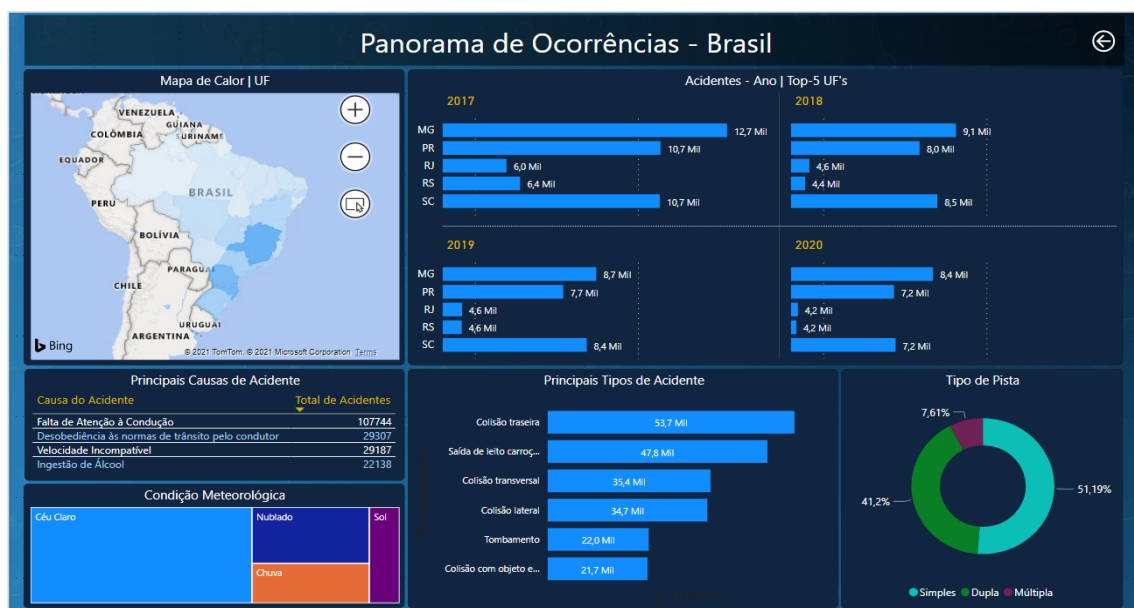


Fonte: Elaborada pelo autor através do Power BI.

- Brasil – Além do recorte detalhado por unidade federativa, foi feita uma tela que permitisse uma análise mais detalhada também em relação aos dados compilados do Brasil inteiro, em relação às principais causas de acidente, condições do tempo, tipos de acidente e características das pistas. Esse detalhamento foi feito olhando para os cinco estados com maior número de ocorrências registradas.

É possível perceber que o estado de Minas Gerais possui o maior número de registros no Brasil, sendo seguido pelos três estados do Sul. Embora possuam pequena abrangência territorial, possuem complexa malha rodoviária e ponto extremo das maiores rodovias do país.

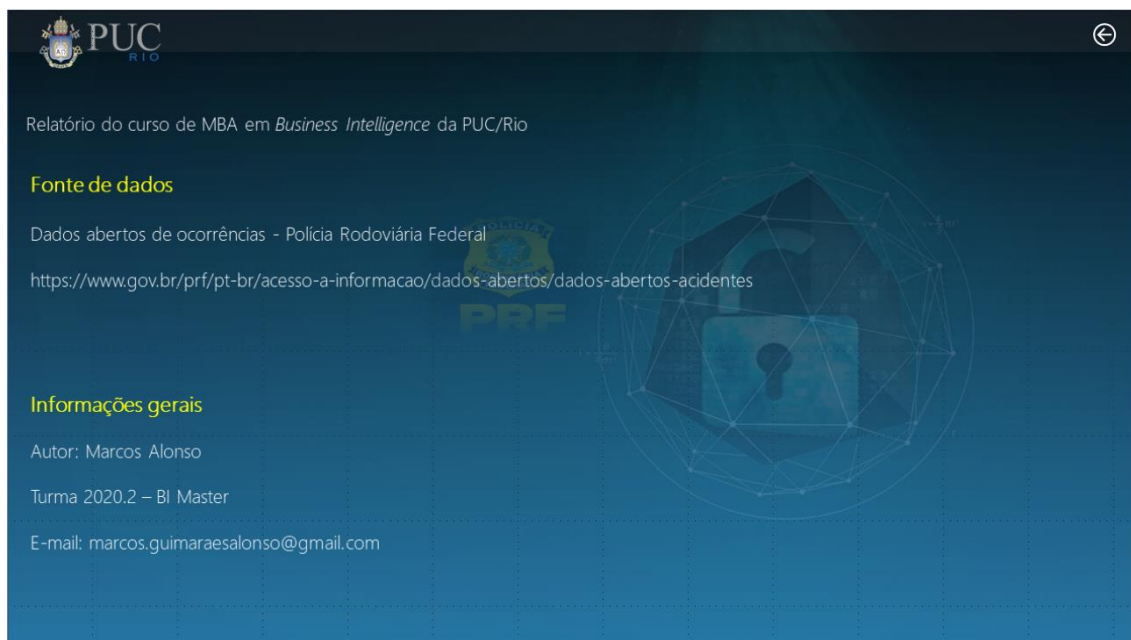
Figura 26: Tela Brasil



Fonte: Elaborada pelo autor através do Power BI.

- Tela Informativo – Painel informativo com relação de fonte dos dados para a realização do presente estudo técnico e informações de contato do autor. Acréscimo de botão para retorno ao Menu Inicial.

Figura 27: Tela Informativo



Fonte: Elaborada pelo autor através do Power BI.

- Tela de Finalização – Tela de finalização do relatório, com botão para voltar à Tela de Menu.

Figura 28: Tela de Finalização



Fonte: Elaborada pelo autor através do Power BI.

9 Conclusão

O presente relatório teve como objetivo apresentar um panorama dos acidentes nas rodovias federais brasileiras, por meio de um painel dinâmico das ocorrências registradas. O objetivo com esse relatório é apresentar ferramentas técnicas de tomada de decisão para órgãos públicos – na realização de políticas públicas mais assertivas – e empresas do setor privado, cuja atividade de logística configura-se como fator de produção principal, no intuito de avaliar estratégias, treinamentos e disseminação de informações de forma a preservar vidas e otimizar seus lucros.

Para tanto, foi elaborado um projeto completo de *Business Intelligence*, com foco na principal linguagem de programação para um sistema de gerenciador de banco de dados (SGBD), o SQL. O SGBD de referência aplicado foi o PostgreSQL, sendo realizadas conexões com as ferramentas SQL Power Architect – para otimização no processo de elaboração das tabelas – e a switch do Pentaho especializada no processo de ETL, o Pentaho Data Integration (PDI).

Após a definição da arquitetura do *Data Warehouse* e sua respectiva criação no PostgreSQL, foi realizada uma conexão com o Power BI, ferramenta muito utilizada no mercado para elaboração dos *dashboards* do panorama de acidentes nas rodovias federais brasileiras.

Com os dados, foi possível obter informações quantitativas dos acidentes nos estados, municípios e rodovias, bem como informações qualitativas dos acidentes. Concomitantemente, foi possível observar que dentro da série temporal de análise, o ano de 2017 registrou um alto índice de acidentes nas estradas, e por mais que o ano de 2020 registrou a tendência de queda nos acidentes, o número de fatalidades nas estradas teve comportamento inversamente proporcional.

A partir de informações detalhadas e dinâmicas, o Ministério da Infraestrutura, junto com os demais órgãos públicos responsáveis e empresas privadas de interesse, poderá analisar possíveis contextos externos que influenciem nas tendências verificadas. Assim ações mais assertivas podem manter uma tendência de queda nos acidentes e a redução das fatalidades nas estradas, por ordem de prioridade dados os cenários apresentados.

Por fim, cumpre esclarecer que todo o projeto foi desenvolvido de forma com que esse trabalho possa ser atualizado de maneira mais automatizada possível, servindo de referência para as divulgações anuais da Polícia Rodoviária Federal.

10 Anexos

SCRIPT SQL – Criação das tabelas do DW dw_prf através do SQL Power Architect

```
CREATE SEQUENCE public.dim_faixa_horario_sk_faixa_horario_seq;
```

```
CREATE TABLE public.dim_faixa_horario (  
    sk_faixa_horario INTEGER NOT NULL DEFAULT  
nextval('public.dim_faixa_horario_sk_faixa_horario_seq'),  
    nk_faixa_horario TIME NOT NULL,  
    faixa_horario VARCHAR(50) NOT NULL,  
    CONSTRAINT sk_faixa_horario PRIMARY KEY (sk_faixa_horario)  
);
```

```
ALTER SEQUENCE public.dim_faixa_horario_sk_faixa_horario_seq OWNED BY  
public.dim_faixa_horario.sk_faixa_horario;
```

```
CREATE SEQUENCE public.dim_acidente_sk_acidente_seq;
```

```
CREATE TABLE public.dim_acidente (  
    sk_acidente INTEGER NOT NULL DEFAULT  
nextval('public.dim_acidente_sk_acidente_seq'),  
    nk_acidente INTEGER NOT NULL,  
    CONSTRAINT sk_acidente PRIMARY KEY (sk_acidente)  
);
```

```
ALTER SEQUENCE public.dim_acidente_sk_acidente_seq OWNED BY  
public.dim_acidente.sk_acidente;
```

```
CREATE SEQUENCE public.dim_tipo_pista_sk_tipo_pista_seq;
```

```
CREATE TABLE public.dim_tipo_pista (  
    sk_tipo_pista INTEGER NOT NULL DEFAULT  
nextval('public.dim_tipo_pista_sk_tipo_pista_seq'),  
    nk_tipo_pista VARCHAR(20) NOT NULL,  
    CONSTRAINT sk_tipo_pista PRIMARY KEY (sk_tipo_pista)  
);
```

```
ALTER SEQUENCE public.dim_tipo_pista_sk_tipo_pista_seq OWNED BY  
public.dim_tipo_pista.sk_tipo_pista;
```

```
CREATE SEQUENCE public.dim_condicao_meteorologica_sk_condicao_meteorologica_seq;
```

```
CREATE TABLE public.dim_condicao_meteorologica (  
    sk_condicao_meteorologica INTEGER NOT NULL DEFAULT  
nextval('public.dim_condicao_meteorologica_sk_condicao_meteorologica_seq'),  
    nk_condicao_meteorologica VARCHAR(50) NOT NULL,  
    CONSTRAINT sk_condicao_meteorologica PRIMARY KEY (sk_condicao_meteorologica)  
);
```

```
ALTER SEQUENCE public.dim_condicao_meteorologica_sk_condicao_meteorologica_seq OWNED  
BY public.dim_condicao_meteorologica.sk_condicao_meteorologica;
```

```
CREATE SEQUENCE public.dim_tipo_acidente_sk_tipo_acidente_seq;
```

```

CREATE TABLE public.dim_tipo_acidente (
    sk_tipo_acidente INTEGER NOT NULL DEFAULT
nextval('public.dim_tipo_acidente_sk_tipo_acidente_seq'),
    nk_tipo_acidente VARCHAR(50) NOT NULL,
    CONSTRAINT sk_tipo_acidente PRIMARY KEY (sk_tipo_acidente)
);

ALTER SEQUENCE public.dim_tipo_acidente_sk_tipo_acidente_seq OWNED BY
public.dim_tipo_acidente.sk_tipo_acidente;

CREATE SEQUENCE public.dim_causa_acidente_sk_causa_acidente_seq;

CREATE TABLE public.dim_causa_acidente (
    sk_causa_acidente INTEGER NOT NULL DEFAULT
nextval('public.dim_causa_acidente_sk_causa_acidente_seq'),
    nk_causa_acidente VARCHAR(100) NOT NULL,
    CONSTRAINT sk_causa_acidente PRIMARY KEY (sk_causa_acidente)
);

ALTER SEQUENCE public.dim_causa_acidente_sk_causa_acidente_seq OWNED BY
public.dim_causa_acidente.sk_causa_acidente;

CREATE SEQUENCE public.dim_latitude_sk_latitude_seq;

CREATE TABLE public.dim_latitude (
    sk_latitude INTEGER NOT NULL DEFAULT nextval('public.dim_latitude_sk_latitude_seq'),
    nk_latitude NUMERIC NOT NULL,
    CONSTRAINT sk_latitude PRIMARY KEY (sk_latitude)
);

ALTER SEQUENCE public.dim_latitude_sk_latitude_seq OWNED BY public.dim_latitude.sk_latitude;

CREATE SEQUENCE public.dim_longitude_sk_longitude_seq;

CREATE TABLE public.dim_longitude (
    sk_longitude INTEGER NOT NULL DEFAULT
nextval('public.dim_longitude_sk_longitude_seq'),
    nk_longitude NUMERIC NOT NULL,
    CONSTRAINT sk_longitude PRIMARY KEY (sk_longitude)
);

ALTER SEQUENCE public.dim_longitude_sk_longitude_seq OWNED BY
public.dim_longitude.sk_longitude;

CREATE SEQUENCE public.dim_data_sk_data_seq;

CREATE TABLE public.dim_data (
    sk_data INTEGER NOT NULL DEFAULT nextval('public.dim_data_sk_data_seq'),
    nk_data DATE NOT NULL,
    dt_dia INTEGER NOT NULL,
    dt_mes INTEGER NOT NULL,
    dt_ano INTEGER NOT NULL,
    dt_dia_semana VARCHAR(20) NOT NULL,
    CONSTRAINT sk_data PRIMARY KEY (sk_data)
);

```

```

ALTER SEQUENCE public.dim_data_sk_data_seq OWNED BY public.dim_data.sk_data;

CREATE SEQUENCE public.dim_fase_dia_sk_fase_dia_seq;

CREATE TABLE public.dim_fase_dia (
    sk_fase_dia INTEGER NOT NULL DEFAULT
nextval('public.dim_fase_dia_sk_fase_dia_seq'),
    nk_fase_dia VARCHAR(20) NOT NULL,
    CONSTRAINT sk_fase_dia PRIMARY KEY (sk_fase_dia)
);

ALTER SEQUENCE public.dim_fase_dia_sk_fase_dia_seq OWNED BY
public.dim_fase_dia.sk_fase_dia;

CREATE SEQUENCE public.dim_classificacao_acidente_sk_classificacao_acidente_seq;

CREATE TABLE public.dim_classificacao_acidente (
    sk_classificacao_acidente INTEGER NOT NULL DEFAULT
nextval('public.dim_classificacao_acidente_sk_classificacao_acidente_seq'),
    nk_classificacao_acidente VARCHAR(100) NOT NULL,
    CONSTRAINT sk_classificacao_acidente PRIMARY KEY (sk_classificacao_acidente)
);

ALTER SEQUENCE public.dim_classificacao_acidente_sk_classificacao_acidente_seq OWNED BY
public.dim_classificacao_acidente.sk_classificacao_acidente;

CREATE SEQUENCE public.dim_municipio_sk_municipio_seq;

CREATE TABLE public.dim_municipio (
    sk_municipio INTEGER NOT NULL DEFAULT
nextval('public.dim_municipio_sk_municipio_seq'),
    nk_municipio VARCHAR(50) NOT NULL,
    CONSTRAINT sk_municipio PRIMARY KEY (sk_municipio)
);

ALTER SEQUENCE public.dim_municipio_sk_municipio_seq OWNED BY
public.dim_municipio.sk_municipio;

CREATE SEQUENCE public.dim_br_sk_br_seq;

CREATE TABLE public.dim_br (
    sk_br INTEGER NOT NULL DEFAULT nextval('public.dim_br_sk_br_seq'),
    nk_br VARCHAR(50) NOT NULL,
    CONSTRAINT sk_br PRIMARY KEY (sk_br)
);

ALTER SEQUENCE public.dim_br_sk_br_seq OWNED BY public.dim_br.sk_br;

CREATE SEQUENCE public.dim_uf_sk_uf_seq;

CREATE TABLE public.dim_uf (
    sk_uf INTEGER NOT NULL DEFAULT nextval('public.dim_uf_sk_uf_seq'),
    nk_uf VARCHAR(2) NOT NULL,
    regiao VARCHAR(30) NOT NULL,
    CONSTRAINT sk_uf PRIMARY KEY (sk_uf)

```

);

ALTER SEQUENCE public.dim_uf_sk_uf_seq OWNED BY public.dim_uf.sk_uf;

```
CREATE TABLE public.ft_ocorrencia (  
    med_mortos INTEGER NOT NULL,  
    med_feridos_leves INTEGER NOT NULL,  
    med_feridos_graves INTEGER NOT NULL,  
    med_ilesos INTEGER NOT NULL,  
    med_ignorados INTEGER NOT NULL,  
    sk_acidente INTEGER NOT NULL,  
    sk_data INTEGER NOT NULL,  
    sk_faixa_horario INTEGER NOT NULL,  
    sk_uf INTEGER NOT NULL,  
    sk_br INTEGER NOT NULL,  
    sk_municipio INTEGER NOT NULL,  
    sk_latitude INTEGER NOT NULL,  
    sk_longitude INTEGER NOT NULL,  
    sk_tipo_acidente INTEGER NOT NULL,  
    sk_classificacao_acidente INTEGER NOT NULL,  
    sk_causa_acidente INTEGER NOT NULL,  
    sk_condicao_metereologica INTEGER NOT NULL,  
    sk_tipo_pista INTEGER NOT NULL,  
    sk_fase_dia INTEGER NOT NULL,  
    CONSTRAINT sk_ocorrencia PRIMARY KEY (med_mortos, med_feridos_leves,  
med_feridos_graves, med_ilesos, med_ignorados)  
);
```

```
ALTER TABLE public.ft_ocorrencia ADD CONSTRAINT dim_faixa_horario_ft_ocorrencia_fk  
FOREIGN KEY (sk_faixa_horario)  
REFERENCES public.dim_faixa_horario (sk_faixa_horario)  
ON DELETE NO ACTION  
ON UPDATE NO ACTION  
NOT DEFERRABLE;
```

```
ALTER TABLE public.ft_ocorrencia ADD CONSTRAINT dim_acidente_ft_ocorrencia_fk  
FOREIGN KEY (sk_acidente)  
REFERENCES public.dim_acidente (sk_acidente)  
ON DELETE NO ACTION  
ON UPDATE NO ACTION  
NOT DEFERRABLE;
```

```
ALTER TABLE public.ft_ocorrencia ADD CONSTRAINT dim_tipo_pista_ft_ocorrencia_fk  
FOREIGN KEY (sk_tipo_pista)  
REFERENCES public.dim_tipo_pista (sk_tipo_pista)  
ON DELETE NO ACTION  
ON UPDATE NO ACTION  
NOT DEFERRABLE;
```

```
ALTER TABLE public.ft_ocorrencia ADD CONSTRAINT  
dim_condicao_metereologica_ft_ocorrencia_fk  
FOREIGN KEY (sk_condicao_metereologica)  
REFERENCES public.dim_condicao_metereologica (sk_condicao_metereologica)  
ON DELETE NO ACTION  
ON UPDATE NO ACTION  
NOT DEFERRABLE;
```

```
ALTER TABLE public.ft_ocorrencia ADD CONSTRAINT dim_tipo_acidente_ft_ocorrencia_fk
```

```
FOREIGN KEY (sk_tipo_acidente)
REFERENCES public.dim_tipo_acidente (sk_tipo_acidente)
ON DELETE NO ACTION
ON UPDATE NO ACTION
NOT DEFERRABLE;
```

```
ALTER TABLE public.ft_ocorrencia ADD CONSTRAINT dim_causa_acidente_ft_ocorrencia_fk
FOREIGN KEY (sk_causa_acidente)
REFERENCES public.dim_causa_acidente (sk_causa_acidente)
ON DELETE NO ACTION
ON UPDATE NO ACTION
NOT DEFERRABLE;
```

```
ALTER TABLE public.ft_ocorrencia ADD CONSTRAINT dim_latitude_ft_ocorrencia_fk
FOREIGN KEY (sk_latitude)
REFERENCES public.dim_latitude (sk_latitude)
ON DELETE NO ACTION
ON UPDATE NO ACTION
NOT DEFERRABLE;
```

```
ALTER TABLE public.ft_ocorrencia ADD CONSTRAINT dim_longitude_ft_ocorrencia_fk
FOREIGN KEY (sk_longitude)
REFERENCES public.dim_longitude (sk_longitude)
ON DELETE NO ACTION
ON UPDATE NO ACTION
NOT DEFERRABLE;
```

```
ALTER TABLE public.ft_ocorrencia ADD CONSTRAINT dim_data_ft_ocorrencia_fk
FOREIGN KEY (sk_data)
REFERENCES public.dim_data (sk_data)
ON DELETE NO ACTION
ON UPDATE NO ACTION
NOT DEFERRABLE;
```

```
ALTER TABLE public.ft_ocorrencia ADD CONSTRAINT dim_fase_dia_ft_ocorrencia_fk
FOREIGN KEY (sk_fase_dia)
REFERENCES public.dim_fase_dia (sk_fase_dia)
ON DELETE NO ACTION
ON UPDATE NO ACTION
NOT DEFERRABLE;
```

```
ALTER TABLE public.ft_ocorrencia ADD CONSTRAINT dim_classificacao_acidente_ft_ocorrencia_fk
FOREIGN KEY (sk_classificacao_acidente)
REFERENCES public.dim_classificacao_acidente (sk_classificacao_acidente)
ON DELETE NO ACTION
ON UPDATE NO ACTION
NOT DEFERRABLE;
```

```
ALTER TABLE public.ft_ocorrencia ADD CONSTRAINT dim_municipio_ft_ocorrencia_fk
FOREIGN KEY (sk_municipio)
REFERENCES public.dim_municipio (sk_municipio)
ON DELETE NO ACTION
ON UPDATE NO ACTION
NOT DEFERRABLE;
```

```
ALTER TABLE public.ft_ocorrencia ADD CONSTRAINT dim_br_ft_ocorrencia_fk
FOREIGN KEY (sk_br)
REFERENCES public.dim_br (sk_br)
ON DELETE NO ACTION
ON UPDATE NO ACTION
```

NOT DEFERRABLE;

```
ALTER TABLE public.ft_ocorrencia ADD CONSTRAINT dim_uf_ft_ocorrencia_fk  
FOREIGN KEY (sk_uf)  
REFERENCES public.dim_uf (sk_uf)  
ON DELETE NO ACTION  
ON UPDATE NO ACTION  
NOT DEFERRABLE;
```


11 Arquivos

Segue abaixo uma relação dos arquivos anexos a este relatório:

Dados abertos - Polícia Rodoviária Federal:

datatran2017.csv

datatran2018.csv

datatran2019.csv

datatran2020.csv

Python:

Script_01_DataAnalysis_datatran.ipynb

dados17_20_datatran.csv

SQL:

Alterações_dim_causa_acidente.sql

SQL Power Architect:

multidimensional_sql_pa.architect

transacional_exemplo.architect

PDI:

Criação_dim_acidente.ktr

Criação_dim_br.ktr

Criação_dim_causa_acidente.ktr

Criação_dim_classificacao_acidente.ktr

Criação_dim_condicao_meteorologica.ktr

Criação_dim_faixa_horario.ktr

Criação_dim_fase_dia.ktr

Criação_dim_latitude.ktr

Criação_dim_longitude.ktr

Criação_dim_municipio.ktr

Criação_dim_tempo.ktr

Criação_dim_tipo_acidente.ktr

Criação_dim_tipo_pista.ktr

Criação_dim_uf.ktr

Criação_ft_ocorrencia_teste.ktr

Power BI

Relatório_Final_PRF.pbix