



**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E  
TECNOLOGIA DE ALAGOAS – IFAL – CAMPUS MACEIÓ  
COORDENAÇÃO DE INFORMÁTICA  
BACHARELADO EM SISTEMAS DE INFORMAÇÃO**

**LUCCA CARNAÚBA PEIXOTO ROSÁRIO  
MAYSON FERREIRA DE MEDEIROS**

**USO DE BUSINESS INTELLIGENCE PARA MENSURAÇÃO DE  
ACIDENTES RODOVIÁRIOS ATRAVÉS DA BASE DE DADOS DA  
POLÍCIA RODOVIÁRIA FEDERAL**

**MACEIÓ, AL  
2024**

LUCCA CARNAÚBA PEIXOTO ROSÁRIO  
MAYSON FERREIRA DE MEDEIROS

Monografia apresentada ao Instituto Federal de Educação, Ciência e Tecnologia de Alagoas – IFAL, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação. Área de concentração: Business Intelligence.

Orientador: Prof. Dr. Edison Camilo de Moraes Júnior

MACEIÓ, AL  
2024



**Dados Internacionais de Catalogação na Publicação**  
**Instituto Federal de Alagoas**  
***Campus Maceió***  
**Biblioteca Benevides Monte**

---

004

R789u Rosário, Lucca Carnáuba Peixoto.

Uso de *business intelligence* para mensuração de acidentes rodoviários através da base de dados da Polícia Rodoviária Federal [recurso eletrônico] / Lucca Carnáuba Peixoto Rosário, Mayson Ferreira de Medeiros. – Dados eletrônicos (1 arquivo em pdf: 415 MB). – 2024.

Trabalho acadêmico com 74 folhas.

Inclui figuras, quadros e referências.

Orientação: Prof. Dr. Edison Camilo de Moraes Júnior.

Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) – Instituto Federal de Alagoas, *Campus Maceió*, Maceió, 2024.

1. Sistemas de Informação. 2. *Business Intelligence* – Acidentes de trânsito. 2. Datatran – DPRF. I. Medeiros, Mayson Ferreira de. II. Título.

---

**Franciane Monick Gomes de França**  
**Bibliotecária – CRB 4/1831**

LUCCA CARNAÚBA PEIXOTO ROSÁRIO  
MAYSON FERREIRA DE MEDEIROS

Uso de Business Intelligence para mensuração de acidentes rodoviários através da base de dados abertos da Polícia Rodoviária Federal. Monografia de Bacharelado em Sistemas de Informação, do Instituto Federal de Alagoas – Campus Maceió, submetida ao corpo docente do Curso de Bacharel em Sistemas de Informação do Instituto Federal de Alagoas – Campus Maceió.

Aprovado em: 08/05/2024.

**BANCA EXAMINADORA:**

---

Prof. Dr. Edison Camilo de Moraes Júnior (Orientador)  
Instituto Federal de Alagoas - IFAL

---

Prof. Dr. Jailton Cardoso da Cruz  
Instituto Federal de Alagoas - IFAL

---

Prof. Dr. Flávio Mota Medeiros  
Instituto Federal de Alagoas - IFAL

## **AGRADECIMENTOS**

Gostaria de agradecer primeiramente a minha mãe, Irene Maria Mota Carnaúba, e meu pai, Francisco Rosário, que me apoiaram e me incentivaram em todos os momentos e nunca desistiram de mim, mesmo em momentos delicados.

Gostaria de agradecer também a todos os meus avós, que me ajudaram muito nesses últimos anos.

Agradeço a toda minha família que, mesmo de longe, estava me apoiando o tempo todo.

E por fim agradeço ao Prof. Dr. Edison Camilo por ter pegado na minha mão, me guiado em uma nova área de atuação e no Trabalho de Conclusão de Curso (TCC).

### **Lucca Carnaúba Peixoto Rosário**

Agradeço primeiramente ao senhor meu Deus, autor da vida e de toda sabedoria, por me conceder forças e resiliência para enfrentar e superar as dificuldades e me indicar o melhor caminho a seguir.

Agradeço aos meus pais, pelos ensinamentos de vida, cuidado e educação.

Agradeço eternamente a minha tia Nadir Medeiros da Conceição (in memoriam), pela acolhida nos momentos difíceis da vida, pelo amor, carinho e respeito que sempre teve por mim, neste momento tão especial tenho certeza que seu espírito se alegra.

Agradeço a todos os professores do Instituto Federal de Alagoas - IFAL que me ensinaram, inspiraram e contribuíram para a minha formação profissional e intelectual.

Agradeço aos meus orientadores Prof. Dr. Edison Camilo pelo auxílio na elaboração deste trabalho. Sempre motivando e encorajando para não desistir.

Aos amigos e familiares fundamentais na minha vida.

A todos que de alguma forma contribuíram para a conclusão deste trabalho.

### **Mayson F. Medeiros**

## LISTA DE FIGURAS

Figura 01: Crescimento da frota de veículos no Brasil.....	13
Figura 02 - Número de mortos por faixa etária (2010-2019).....	14
Figura 03 - Mapas das Rodovias Federais.....	26
Figura 04 - Site oficial Polícia Rodoviária Federal.....	27
Figura 05 - Estrutura de um BIFonte: Reprodução ManageBI (2020).....	29
Figura 06 - Modelagem Star Schema.....	31
Figura 07 - Exemplo de um Data Warehouse (DW).....	33
Figura 08 - Exemplo de Drill Up e Drill Down.....	34
Figura 09 - Exemplo de Slice and Dice.....	34
Figura 10 - Modelo Multidimensional com a Dimensão Tempo (DIM_PERIODO).....	35
Figura 11 - Processo de ETL.....	36
Figura 12 - Pentaho Data Integration.....	38
Figura 13 - Exemplo de uma carga de dados no Pentaho PDI.....	39
Figura 14 - Arquitetura do sistema PostgreSQL.....	41
Figura 15 - Exemplo do uso do Microsoft Power BI.....	42
Figura 16 - Arquitetura de BI.....	43
Figura 17 - Base de dados - PRF.....	44
Figura 18 - Modelagem Star Schema.....	46
Figura 19 - Carga Dim Causa Acidente Fonte: Autoria Própria (2024).....	48
Figura 20 - Carga Dim Condição Meteorológica.....	48
Figura 21 - Carga Dim Hora.....	49
Figura 22 - Carga Dim Localidade.....	49
Figura 23 - Carga Dim Tipo Acidente.....	50
Figura 24 - Carga Dim tipo Pista.....	50
Figura 25 - Carga Dim Data.....	51
Figura 26 - Carga Fato Acidente.....	51
Figura 27 - Job no Pentaho PDI.....	52
Figura 28 - Análise de Vítimas, Condições Meteorológicas e Fatores de Acidentes.....	66
Figura 29 - Análise Temporal.....	67
Figura 30 - Análise de BR, Pista e Uso Solo.....	68
Figura 31 - Análise Pré e Pós Pandemia.....	69

## LISTA DE QUADROS

Quadro 1: String de Busca gerada na ferramenta parsifal.....	18
Quadro 2 : Questões de Pesquisa QP1 e QP2.....	18
Quadro 3: Lista das fontes de buscas utilizadas nesta pesquisa. ....	19
Quadro 4: Critérios de Inclusão e Exclusão.....	20
Quadro 5: Strings de busca delimitada para cada base de dados.....	21
Quadro 6: Quantidade de estudos selecionados por fonte de dados.....	21
Quadro 7: Perguntas de Avaliação de Qualidade.....	22
Quadro 8: Estudos selecionados pelas perguntas de avaliação de qualidade.....	22
Quadro 9: Classificação de Acidentes de Trânsito.....	24
Quadro 10: Campos da base de dados datatran.....	44
Quadro 11: Script Oracle SQL Developer (Data Modeler).....	52
Quadro 12: Dicionário de Dados.....	62
Quadro 13: Dicionário de Dados das Chaves Estrangeiras.....	65

## **LISTA DE SIGLAS**

**AT**- Acidentes de Trânsito

**BAT** – Boletim de Acidente de Trânsito

**CNM** - Confederação Nacional dos Municípios

**DATATRAN** - Banco de Dados sobre acidentes da Polícia Rodoviária Federal

**MT** – Ministério dos Transportes

**OMS** – Organização Mundial da Saúde

**ONU** – Organização das Nações Unidas

**PDI** - Pentaho Data Integration

**PRF** – Polícia Rodoviária Federal

**SGBDOR** - Sistema de Gerenciamento de Banco de Dados Objeto-Relacional

**SQL** - Structured Query Language

**KPI** - Key Performance Indicator

## RESUMO

Ao longo dos anos, os acidentes de trânsito têm sido uma das principais causas de morte em todo o mundo, segundo o relatório oficial da ONU - Global Status Report On Road Safety(2018). Estima-se que 1,35 Milhões de pessoas morrem todos os anos em decorrência de acidentes no trânsito, sendo a principal causa de morte de jovens adultos. O avanço tecnológico, somado a capacidade de geração de um grande volume de dados, permite uma análise de tais fenômenos e suas causas, Assim, o presente trabalho traz à luz uma abordagem sobre as principais causas de Acidentes de Trânsito - AT, em Rodovias Federais em todo o Brasil, utilizando dados oficiais do Departamento Polícia Rodoviária Federal - DPRF, a base de dados abertos - DATATRAN. O presente estudo foi estruturado a partir de uma Revisão Sistemática de Literatura - RSL e pela utilização de um conjunto de ferramentas para tratamento de dados, junto as técnicas de Business Intelligence - BI, a fim de extrair informações relevantes, auxiliando na tomada de decisão dos usuários de rodovias federais. Foram obtidas conclusões por meio de uma série de análises, comparando os dados referentes ao período da pandemia com os dados oficiais do ano de 2019, ano pré pandemia, gerando gráficos interativos.

**Palavras-chaves:** Business Intelligence. Acidentes de Trânsito. Datatran DPRF. Pandemia

## **ABSTRACT**

Over the years, traffic accidents have been one of the main causes of death worldwide, according to the official UN report - Global Status Report On Road Safety (2018). It is estimated that 1.35 million people die every year as a result of traffic accidents, being the main cause of death for young adults. Technological advancement, combined with the capacity to generate a large volume of data, allows an analysis of such phenomena and their causes. Thus, the present work brings to light an approach to the main causes of Traffic Accidents - AT, on Federal Highways throughout Brazil, using official data from the Federal Highway Police Department - DPRF, the open database - DATATRAN. The present study was structured based on a Systematic Literature Review - RSL and the use of a set of tools for data processing, together with Business Intelligence - BI techniques, in order to extract relevant information, assisting in decision-making. users of federal highways. Conclusions were obtained through a series of analyses, comparing data relating to the pandemic period with official data from 2019, the pre-pandemic year, generating interactive graphs.

**Keywords:** Business Intelligence. Traffic Accidents. Datatran DPRF. Pandemic

# SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>13</b>
1.1. CONTEXTUALIZAÇÃO.....	13
1.2. MOTIVAÇÃO E JUSTIFICATIVA.....	14
1.3. PROBLEMA DA PESQUISA.....	15
1.4. OBJETIVO GERAL.....	16
1.5. OBJETIVOS ESPECÍFICOS.....	16
1.6 HIPÓTESE DE PESQUISA.....	17
1.7 DELIMITAÇÃO DO TRABALHO.....	17
1.8 ESTRUTURA DO TRABALHO.....	17
<b>2. REVISÃO SISTEMÁTICA DA LITERATURA.....</b>	<b>17</b>
2.1 INTRODUÇÃO.....	17
2.2 MÉTODO.....	18
2.2.1 QUESTÕES DE PESQUISA .....	18
2.2.2 FONTE DE DADOS .....	19
2.2.3 CRITÉRIO DE INCLUSÃO E EXCLUSÃO.....	19
2.4 CONDUÇÃO.....	20
2.4.1 BUSCA DOS TRABALHOS.....	20
2.4.2 ESTRATÉGIA DE SELEÇÃO.....	21
2.4.3 RESULTADOS ENCONTRADOS.....	22
<b>3. REFERENCIAL TEÓRICO.....</b>	<b>24</b>
3.1 CLASSIFICAÇÃO DOS ACIDENTES DE TRÂNSITO.....	24
3.2 BASE DE DADOS - DATATRAN.....	26
3.3 BUSINESS INTELLIGENCE (BI).....	27
3.3.1 MODELAGEM MULTIDIMENSIONAL.....	29
3.3.2 MODELO STAR SCHEMA.....	30
3.3.3 DATA WAREHOUSE.....	31
3.3.4 SISTEMAS OLAP.....	33
3.3.4.1 OPERAÇÕES TÍPICAS (DRILL UP, DRILL DOWN E SLICE AND DICE)...	33
3.3.4.2 ALGUNS TIPOS DE DIMENSÕES.....	34
3.3.5 ETL.....	35
3.3.6 SLOWLY CHANGING DIMENSION (SCD).....	36
3.4 ORACLE DATA MODELER.....	37
3.5 PENTAHO.....	38
3.5.1 PENTAHO DATA INTEGRATION (PDI).....	38

3.6 POSTGRESQL.....	39
3.7 MICROSOFT POWER BI (PBI).....	41
3.7.1 DASHBOARDS ANALÍTICOS.....	42
<b>4 BI-SAFEROAD: DADOS PARA SEGURANÇA NAS ESTRADAS.....</b>	<b>43</b>
4.1 AQUISIÇÃO DE DADOS.....	43
4.2 MODELAGEM MULTIDIMENSIONAL DE DADOS.....	45
4.3 PROCESSOS DE ETL (EXTRACT, LOAD, TRANSFORMATION) E PLANOS DE CARGA.....	47
4.4 CRIAÇÃO DO DATA WAREHOUSE NO POSTGRESQL.....	51
4.5 DICIONÁRIO DE DADOS.....	60
4.6 ANÁLISE DE DADOS.....	64
<b>5 RESULTADOS.....</b>	<b>68</b>
<b>6 CONSIDERAÇÕES FINAIS.....</b>	<b>70</b>
6.1 CONCLUSÃO.....	70
6.1.1 ANÁLISE DE VÍTIMAS, CONDIÇÃO METEOROLÓGICA E CAUSAS DE ACIDENTE.....	71
6.1.2 ANÁLISE TEMPORAL.....	71
6.1.3 ANÁLISE DE BR, PISTA E VIA PAVIMENTADA.....	72
6.1.4 ANÁLISE PRÉ E PÓS PANDEMIA.....	72
6.2 LIMITAÇÕES DA PESQUISA.....	72
6.3 TRABALHOS FUTUROS.....	73
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>73</b>

## 1. INTRODUÇÃO

### 1.1. CONTEXTUALIZAÇÃO

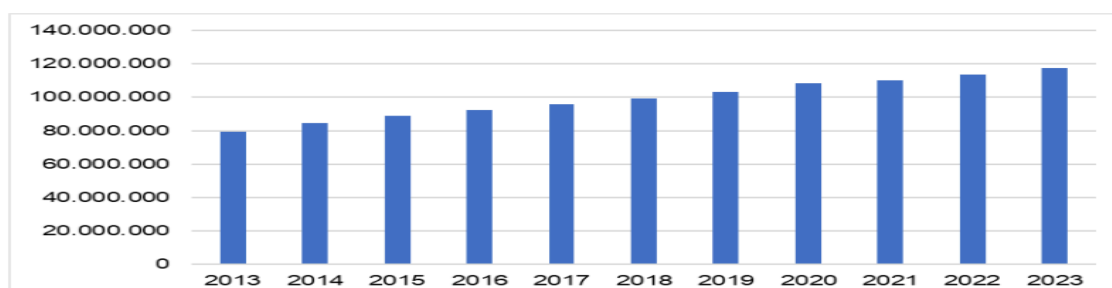
A extensão territorial do Brasil impõe grandes desafios em relação a mobilidade, nesse sentido, a rodovia é o principal meio de deslocamento entre as regiões do Brasil, sendo o 4º país no mundo com a maior malha rodoviária, cerca de 1.720.700 quilômetros de estradas e rodovias.

A extensão total da malha rodoviária federal, excluindo as vias planejadas, é de 75.257 mil km, dos quais 65.735 mil km (90,5%) correspondem a rodovias pavimentadas e 9.522 mil km (9,5%) correspondem a rodovias não pavimentadas. Da malha rodoviária federal pavimentada, 11.429 km são administrados por meio de concessão federal (ANTT, 2024).

Os acidentes de trânsito passaram a ocupar o segundo lugar dentre as principais causas de morte no Brasil a partir da década de 80, impactando principalmente jovens adultos em idade economicamente ativa, constituindo também um gravíssimo problema de saúde pública pela sua magnitude e número de vítimas e o impacto econômico causado todos os anos (ROCHA, 2009).

Um fator que contribui diretamente para o aumento no número de acidentes é a quantidade de veículos em circulação, de acordo com a Confederação Nacional dos Municípios - CNM (2023), a frota brasileira de veículos teve um aumento expressivo nos últimos 10 anos. Em 2013, o Brasil tinha 80 milhões de veículos, passando para aproximadamente 120 milhões em 2023, figura 1, representando um crescimento de 35%, tal incremento também refletiu no número de mortos e feridos, de acordo com dados do Datasus, no período de 2010-2019, verificou-se um aumento no sinistros de trânsito, totalizando 392,929 mil mortes, principalmente jovens adultos, com ênfase a faixa etária de 15 a 29 anos, figura 2. Ipea (2023).

Figura 01: Crescimento da frota de veículos no Brasil.



**Reprodução: Senatran / CNM (2023).**

Figura 02 - Número de mortos por faixa etária (2010-2019).

Faixa etária	Mortes	%
Menor de 15 anos	14.696	4
15 a 29 anos	125.293	32
30 a 49 anos	139.746	36
50 a 69 anos	80.893	21
70 anos ou mais	29.606	8
Ignorado	2.695	-
<b>Total</b>	<b>392.929</b>	<b>100</b>

**Reprodução: SIM/Datasus (2023).**

Em estudo realizado pelo CNT em 2023, com dados registrados em 2022, é possível constatar o impacto financeiro dos acidentes somente em rodovias federais, o custo estimado dos 64.447 acidentes registrados no Brasil alcançou o patamar de 12,92 bilhões de reais, em média, a cada 10 Km de rodovia 7 acidentes foram registrados. Diante desse cenário, e com o avanço do desenvolvimento tecnológico da última década, é relevante aprofundar o conhecimento sobre o tema, bem como buscar alternativas e meios de mitigar tais impactos que atingem o indivíduo, a sociedade e o sistema público de saúde.

## 1.2. MOTIVAÇÃO E JUSTIFICATIVA

Historicamente, os acidentes de trânsito são uma das principais causas de perda de indivíduos em idade produtiva no Brasil, para fins de avaliação somente em 2019, de acordo com dados oficiais do Ministério dos Transportes - MT (2021), foram registrados 1.007.603 acidentes em todo o Brasil, abrangendo áreas urbanas, rurais e rodovias estaduais e federais. Esses acidentes resultaram em 25.543 mortes e 1.595.659 pessoas envolvidas, entre feridos e ilesos.

Esses números destacam a necessidade urgente de estudar as causas, consequências e formas de prevenção de acidentes rodoviários, bem como de propor políticas públicas eficazes para melhorar a segurança viária.

Dado o exposto, o estudo em questão é motivado por um desejo de aprofundar o conhecimento sobre um tema de grande importância, devido ao seu significativo impacto, seja de natureza econômica ou social, sobre a sociedade. Este trabalho tem como objetivo analisar os dados fornecidos pelo Datatran, um sistema de informações que registra acidentes de trânsito ocorridos nas rodovias federais, mantido pela Polícia Rodoviária

Federal (PRF). O Datatran disponibiliza dados abertos sobre acidentes ocorridos desde 2007, agrupados por ocorrência, pessoa, causa e tipo de acidente.

A partir desses dados, pretende-se identificar os principais fatores de risco, as características dos envolvidos, as tendências e os padrões dos acidentes, bem como as suas implicações para a saúde e mobilidade. Polícia Rodoviária Federal (2023).

A relevância deste estudo se justifica pela escassez de literatura acadêmica que utilize o Datatran como fonte de dados para analisar os acidentes de trânsito nas rodovias federais durante a pandemia. A maioria dos estudos existentes baseia-se em casos específicos. Além disso, este trabalho contribui para o avanço do conhecimento científico sobre o tema, que é de grande interesse social, bem como para a formulação de propostas de intervenção que possam reduzir os índices de acidentalidade e mortalidade nas rodovias federais.

As razões para acidentes nas estradas federais são as mais variadas possíveis, incluindo embriaguez ao dirigir, motorista sonolento, animais na via, entre outras. A partir do Datatran, é possível realizar uma análise minuciosa dos principais fatores de risco de acidentes, bem como dos pontos que requerem maior atenção por parte dos órgãos responsáveis, bem como dos condutores de veículos.

### **1.3. PROBLEMA DA PESQUISA**

Os acidentes de trânsito são um grave problema de saúde pública e segurança viária, que afetam milhões de pessoas no mundo todo. No Brasil, as rodovias federais são cenários frequentes de acidentes, que resultam em mortes, ferimentos, danos materiais e ambientais, além de custos sociais e econômicos elevados.

Segundo o Datatran, em 2022, foram registrados 63.447 acidentes em rodovias federais, com 5.287 mortes e 76.481 feridos. Esses números revelam a necessidade de se compreender melhor os fatores que influenciam a ocorrência e a gravidade dos acidentes de trânsito em rodovias federais, bem como as medidas de prevenção e intervenção que podem ser adotadas para reduzi-los.

O presente trabalho tem como objetivo analisar os dados do Datatran, referentes aos acidentes de trânsito em rodovias federais no período de 2019 a 2023, buscando identificar as principais variáveis associadas aos acidentes, tais como: tipo de acidente, causa presumida, condição da pista, condição meteorológica, tipo de veículo, sexo e idade do

condutor, entre outras. A partir dessa análise, pretende-se responder ao seguinte problema de pesquisa: Quais são os fatores que mais contribuem para a ocorrência dos acidentes de trânsito em rodovias federais no Brasil?

Diante do contexto citado, surgem os seguintes questionamentos:

- Quais foram os principais fatores causadores e tipos de acidentes no período de estudo?
- Quais estados e rodovias registraram os maiores índices de acidentes e vítimas fatais?
- Quais os horários, meses, dias de semana e anos com maior ocorrência de acidentes?
- Como as condições climáticas influenciaram na ocorrência de vítimas fatais de trânsito?
- Como as condições das vias (como pavimentação, por exemplo) influenciam a frequência e gravidade dos acidentes de trânsito?

#### **1.4. OBJETIVO GERAL**

O objetivo geral deste estudo é desenvolver dashboards que apresentem os principais indicadores (KPIs) e gráficos dos acidentes de trânsito nas rodovias federais brasileiras.

#### **1.5. OBJETIVOS ESPECÍFICOS**

Dos objetivos específicos, temos:

- Construir o Banco De Dados para o dashboard;
- Criar o processo de *ETL (Extract, Load, Transformation)* no Banco De Dados;
- Realização dos testes;
- Apontar os principais tipos e causas de acidentes em rodovias federais;
- Apontar quais estados e rodovias possuem maiores registro de acidentes e vítimas fatais;
- Realizar análise temporal do acidentes;
- Analisar a influência das condições meteorológicas no número de acidentes e vítimas fatais;

- Avaliar como as condições da via impactam na frequência e gravidade dos acidentes;

## 1.6 HIPÓTESE DE PESQUISA

A restrição de circulação de pessoas está positivamente associada à diminuição de acidentes em rodovias federais.

## 1.7 DELIMITAÇÃO DO TRABALHO

Este trabalho será limitado a analisar apenas a malha de rodovias federais no período de 2019 até 2023, comparando os números de casos de acidentes de trânsito no período de pandemia com os números registrados no ano pré-pandemia de 2019, e com base na revisão de literatura.

## 1.8 ESTRUTURA DO TRABALHO

O presente trabalho foi dividido em 05 capítulos, conforme descrito abaixo:

**O Capítulo 01** trata da introdução, onde consta a contextualização do problema, os objetivos, a justificativa, objetivo geral, objetivos específicos, a delimitação e a estrutura do trabalho.

**Já no Capítulo 02** tem-se a Revisão Sistemática de Literatura (RSL), um relatório de trabalhos primários relevantes com a questão pesquisada.

**No Capítulo 03** trata-se do referencial teórico, a qual apresenta as ferramentas utilizadas para o desenvolvimento do trabalho.

**O Capítulo 04** apresenta a ferramenta desenvolvida ao longo do trabalho, **BI-SafeRoad: Dados para Segurança nas Estradas**; onde são abordados a aquisição dos dados, processo de ETL (*Extract, Transformation, Load*), armazenamento dos dados, carregamento dos dados na ferramenta Power BI e realizar a análise dos dados.

Por fim, **no Capítulo 05**, são apresentadas as conclusões, considerações finais, limitações da pesquisa e trabalhos futuros.

## 2. REVISÃO SISTEMÁTICA DA LITERATURA

### 2.1 INTRODUÇÃO

Segundo Kitchenham e Charters (2007), Revisão Sistemática de Literatura - RSL é um método de pesquisa cujo objetivo é identificar, avaliar e interpretar todos os artigos

relevantes em uma área específica do conhecimento, tendo como base uma questão particular de pesquisa específica, área ou fenômeno de interesse do pesquisador, desta forma a RSL é denominada estudo secundário, enquanto que os artigos analisados são chamados de estudos primários.

Este trabalho segue as diretrizes específicas para RSL de Kitchenham e Charters (2007), definidas em três fases: Planejamento, Condução e Documentação.

## 2.2 MÉTODO

Para realizar a Revisão Sistemática da Literatura - RSL, foi realizada buscas nas bases de dados Google Scholar, Periódico CAPES e INEP (Instituto Nacional de Estudos e Pesquisas Educacionais), usando as palavras-chave "**Acidentes em rodovias**", "**Business Intelligence**", "**Datatrán PRF**" e seus sinônimos "**Acidente em Rodovia Federal**", "**BI**", "**Inteligência de Negócios**", "**Base de dados da Polícia Rodoviária Federal**", "**Dados Abertos Acidentes PRF**" foi gerada um string de busca que foi utilizadas nas bases mencionadas. É importante destacar que o termo "**BI**" ou "**Inteligência de Negócios**" foi inserido a fim de verificar os trabalhos realizados utilizando essa tecnologia, para verificar os trabalhos mais recentes no tema abordado foi definido o período de 2019 a 2023 para seleção das publicações.

Quadro 1: String de Busca gerada na ferramenta parsifal.

**("Acidentes em Rodovias" OR "Acidente em Rodovia Federal" OR "Datatrán PRF" OR "Base de dados da Polícia Rodoviária Federal" OR "Dados Abertos Acidentes PRF") AND ("Business Intelligence" OR "BI" OR "Inteligência de Negócios")**

Fonte: Autoria própria.

### 2.2.1 QUESTÕES DE PESQUISA

Foram definidas 02 (seis) questões de pesquisa (QP) a serem respondidas pelo estudo (Quadro 1).

Quadro 2 : Questões de Pesquisa QP1 e QP2.

<b>QP1</b>	<b>Quais ferramentas são utilizadas em estudos sobre acidentes de trânsito em rodovias?</b>
	Esta questão de pesquisa, através da literatura existente, responde quais ferramentas são utilizadas na avaliação dos acidentes de trânsito em rodovias federais, de forma que, através dessas informações, o pesquisador possa conhecer as ferramentas frequentemente utilizadas na produção de estudos na área de interesse.

<b>QP2</b>	<b>Quais processos de tratamento os dados recebem para gerar informação?</b>
	Esta questão de pesquisa busca relacionar quais processos de tratamento de dados são utilizados na avaliação de acidentes de trânsito, de modo que o pesquisador aprofunde seu conhecimento na área de estudo em relação às formas de tratamento de dados mais comumente utilizados.

**Fonte:** Autoria própria (2024).

### 2.2.2 FONTE DE DADOS

As fontes de busca são bancos de dados confiáveis que possuem publicações científicas relevantes ao tema de pesquisa do artigo e que sejam relevantes para o meio acadêmico (Quadro 2).

Quadro 3: Lista das fontes de buscas utilizadas nesta pesquisa.

<b>Fonte de busca</b>	<b>Link de acesso</b>	<b>Tipo de busca</b>
Biblioteca Digital da Sociedade Brasileira de Computação	<a href="https://sol.sbc.org.br/index.php/indice">https://sol.sbc.org.br/index.php/indice</a>	Automática
Google Scholar	<a href="https://scholar.google.com.br">https://scholar.google.com.br</a>	Automática
Periódico CAPES	<a href="https://www-periodicos-capes-gov-br.ezl.periodicos.capes.gov.br/index.php?">https://www-periodicos-capes-gov-br.ezl.periodicos.capes.gov.br/index.php?</a>	Automática

**Fonte:** Autoria própria (2024).

### 2.2.3 CRITÉRIO DE INCLUSÃO E EXCLUSÃO

De modo geral, e delineando os critérios de inclusão para seleção dos estudos, foram selecionados trabalhos que abordem a temática em questão, escritos em português, que apresentem o texto completo disponível gratuitamente online. Delimitados pelos critérios de exclusão. Foram excluídas resenhas, sites e textos duplicados.

Desta forma, os critérios de inclusão tem a finalidade de selecionar os artigos que satisfazem as questões de pesquisa elaboradas. Os critérios de exclusão, por sua vez, indicam trabalhos que não são relevantes para a pesquisa e foram encontrados nas buscas automáticas. O quadro 4, apresenta maiores detalhes sobre tais critérios.

Quadro 4: Critérios de Inclusão e Exclusão.

<b>C. I.</b>	<b>C. E.</b>
<b>a.</b> Trabalhos de conclusão de curso - TCC que trate sobre acidentes de trânsito em rodovias federais usando a base de dados DATATRAN.	<b>a.</b> Estudos em qualquer idioma diferente do português.
<b>b.</b> Estudos que apresentem no título, resumo, palavras-chaves ou abordem primariamente o tópico citado na alínea <i>a</i> .	<b>b.</b> Estudos que não estejam disponíveis livremente/gratuitamente para consulta na web.
<b>c.</b> Trabalhos como: artigos de periódicos, relatórios, dissertação de pós-graduação, tese de mestrado, tese de doutorado.	<b>c.</b> Estudos publicados antes do ano de 2019 .
	<b>d.</b> Estudos duplicados.
	<b>e.</b> Estudos secundários
	<b>f.</b> Manuais
	<b>g.</b> Resenhas
	<b>h.</b> Short papers
	<b>i.</b> Trabalhos fora de contexto
	<b>j.</b> Trabalhos de Pós Graduação fora do contexto

## 2.4 CONDUÇÃO

### 2.4.1 BUSCA DOS TRABALHOS

Para criar a RSL com apoio computacional, deve-se procurar estudos em fontes confiáveis, na modalidade automática, usando strings de busca (busca ativa executada pelo pesquisador). Neste trabalho, foi utilizada busca automática.

Para localizar trabalhos nas fontes de busca automática, foi elaborada uma *string* de busca a partir das palavras-chave e seus sinônimos, extraídos das questões de pesquisa. (Quadro 5).

Quadro 5: Strings de busca delimitada para cada base de dados.

<b>Fonte de busca</b>	<b>String de busca (adequada à base)</b>
Periódicos Capes	("Acidentes em Rodovias" OR "Acidente em Rodovia Federal" OR "Datatran PRF" OR "Base de dados da Polícia Rodoviária Federal" OR "Dados Abertos Acidentes PRF") AND ("Business Intelligence" OR "BI" OR "Inteligência de Negócios")
Google Scholar	("Acidentes de Trânsito" OR "Acidentes em Rodovias" OR "Sinistros" OR "Datatran PRF" OR "Base de dados da Polícia Rodoviária Federal") AND ("Pandemia" OR "Coronavírus" OR "Covid-19" OR "Sars-Cov-2") AND ("Business Intelligence" OR "BI" OR "Inteligência de Negócios")
Biblioteca Digital da Sociedade Brasileira de Computação	("Acidentes de Trânsito" OR "Acidentes em Rodovias" OR "Sinistros" OR "Datatran PRF" OR "Base de dados da Polícia Rodoviária Federal") AND ("Pandemia" OR "Coronavírus" OR "Covid-19" OR "Sars-Cov-2") AND ("Business Intelligence" OR "BI" OR "Inteligência de Negócios")

**Fonte:** Autoria própria (2024).

#### 2.4.2 ESTRATÉGIA DE SELEÇÃO

A primeira busca retornou um total de 25 trabalhos das três bases de dados selecionadas. A estratégia para aplicação dos critérios de inclusão (CI) e exclusão (CE) e seleção dos trabalhos se deu em 4 etapas: leitura do título de cada trabalho; leitura do resumo; leitura da metodologia e leitura dos resultados e conclusão.

Após a aplicação dos CI e CE sobraram 3 estudos que se adequam ao objetivo final da RSL e estão relacionados no Quadro 5.

Quadro 6: Quantidade de estudos selecionados por fonte de dados.

<b>Fonte de busca</b>	<b>String de busca aplicada</b>	<b>CI e CE</b>	<b>Seleção 4 etapas</b>	<b>Estudos selecionados</b>
Periódicos Capes	0	0	07	04
Google Scholar	25	06	6	3
Biblioteca Digital da	0	0	0	0

Sociedade Brasileira de Computação				
------------------------------------	--	--	--	--

**Fonte:** Autoria própria (2024).

### 2.4.3 RESULTADOS ENCONTRADOS

Nessa fase, são expostos às perguntas de avaliação de qualidade bem como os resultados encontrados e selecionados, após a análise do material.

Para verificar a qualidade dos artigos, foi definido uma série de pesquisas com seus respectivos pesos, a fim de avaliar o grau de relevância do estudo, a seguir temos a tabela com as perguntas que orientaram a seleção dos trabalhos mais relevantes.

Quadro 7: Perguntas de Avaliação de Qualidade.

<b>Perguntas</b>	<b>Pesos</b>
O objetivo de pesquisa está claramente definido?	SIM / PARCIALMENTE / NÃO
A metodologia está descrita adequadamente?	SIM / PARCIALMENTE / NÃO
O estudo recebeu mais de 10 citações?	SIM / PARCIALMENTE / NÃO
Os autores descrevem as limitações do estudo de forma clara?	SIM / PARCIALMENTE / NÃO
O procedimento de análise de dados é bem descrito?	SIM / PARCIALMENTE / NÃO

**Fonte:** Autoria própria (2024).

Cada peso teve um nota específica atribuída, SIM - 1.0, PARCIALMENTE - 5.0 e NÃO - 0.0, após a leitura dos trabalhos e respectiva atribuição de valores para os quesitos elencados, chegamos ao seguintes trabalhos selecionados.

Quadro 8: Estudos selecionados pelas perguntas de avaliação de qualidade.

<b>Item</b>	<b>Título</b>	<b>Escopo</b>	<b>Score de Qualidade</b>
ES01	PREVENDO A GRAVIDADE DE ACIDENTES RODOVIÁRIOS NO BRASIL: A INFLUÊNCIA DO	O objetivo é avaliar a capacidade	4.0

	<p>AMBIENTE E CARACTERÍSTICAS DO VEÍCULO</p>	<p>preditiva de modelos de classificação em relação à gravidade dos acidentes, com foco nas características dos veículos e nos fatores do ambiente. Por meio da aplicação de algoritmos de classificação e de técnicas de explicabilidade de aprendizado de máquina, pretende-se adquirir conhecimentos relevantes relacionados a esse conjunto de dados, contribuindo para o entendimento e a prevenção de acidentes.</p>	
ES02	<p>UTILIZAÇÃO DE SISTEMA DE INFORMAÇÕES GEOGRÁFICAS NA TOMADA DE DECISÕES DO SISTEMA VIÁRIO COM BASE EM DADOS DE ACIDENTES DE TRÂNSITO NO MUNICÍPIO DE PATO BRANCO PR</p>	<p>O objetivo deste estudo é desenvolver um método, baseado em SIG, que viabilize o gerenciamento e análise de dados de acidentes de trânsito. A técnica empregada para identificar a área geográfica com maior tendência à ocorrência de acidentes de trânsito é o da Estimativa da Densidade Kernel</p>	3.5

		(KDE), incorporada de forma estática a um SIG.	
ES03	ELABORAÇÃO DE MODELO DA PREVISÃO DA VELOCIDADE OPERACIONAL DE RODOVIAS EM PISTAS SIMPLES CONSIDERANDO OS VEÍCULOS DE CARGA	O presente trabalho tem por objetivo geral propor um modelo de estimativa do perfil de velocidade operacional para diferentes tipos de veículos de carga em função das características geométricas de rodovias de pistas simples.	4.0

**Fonte:** Autoria própria (2024).

### 3. REFERENCIAL TEÓRICO

O referencial teórico que fundamenta essa pesquisa está centrado em conceitos relacionados ao presente trabalho. Neste capítulo iremos falar sobre as ferramentas e técnicas utilizadas ao longo do desenvolvimento da ferramenta **BI-SafeRoad: Dados para Segurança nas Estradas**. Foram escolhidas principalmente ferramentas livres e de código aberto (*Open-Source*) para a construção da ferramenta OLAP.

#### 3.1 CLASSIFICAÇÃO DOS ACIDENTES DE TRÂNSITO

O Código de Trânsito Brasileiro - CTB, estabelece que "considere-se trânsito a utilização das vias por pessoas, veículos e animais, isolados ou em grupos, conduzidos ou não, para fins de circulação, parada, estacionamento e operação de carga e descarga" (Art. 1º, § 1º da Lei 9.503/97 CTB - Código de Trânsito Brasileiro). Os acidentes de trânsito no Brasil são classificados segundo as normas técnicas NBR 10.697 e NBR 6.067 da Associação Brasileira de Normas Técnicas - ABNT e o Código Brasileiro de trânsito, conforme mostrado a seguir:

Quadro 9: Classificação de Acidentes de Trânsito

<b>Atropelamento de animal(is)</b>
------------------------------------

Sinistro de trânsito em que animal(is) sofre(m) o impacto de um veículo em movimento.
<b>Atropelamento de Pessoa(s)</b> Sinistro de trânsito em que pessoa(s) sofre(m) o impacto de um veículo em movimento.
<b>Capotamento</b> Sinistro de trânsito em que o veículo gira sobre si mesmo, em qualquer sentido, ficando em algum momento com as rodas para cima, imobilizando-se em qualquer posição.
<b>Choque</b> Sinistro de trânsito em que há impacto de um veículo contra qualquer objeto fixo ou objeto sem movimento.
<b>Colisão</b> Sinistro de trânsito em que o veículo em movimento sofre impacto de outro veículo também em movimento.
<b>Colisão Frontal</b> Colisão que ocorre lateralmente, quando os veículos transitam na mesma direção, podendo ser no mesmo sentido ou em sentidos opostos.
<b>Colisão Transversal</b> Colisão que ocorre na frente contra traseira ou na traseira contra traseira, quando os veículos transitam no mesmo sentido ou em sentidos opostos, podendo pelo menos um deles estar em marcha à ré.
<b>Engavetamento</b> Sinistro de trânsito em que há impacto em três ou mais veículos, em um mesmo sentido de circulação, resultado de uma sequência de colisões traseiras, laterais ou transversais.
<b>Queda</b> Sinistro de trânsito em que há impacto em razão de queda livre do veículo, queda de pessoas ou cargas transportadas em razão de movimento do veículo.
<b>Tombamento</b> sinistro de trânsito em que veículo sai da posição de sua posição normal, imobilizando-se sobre uma de suas laterais, sua frente ou sua traseira.
<b>Outros Sinistros de Trânsito</b> Qualquer sinistro de trânsito que não se enquadre nas definições atropelamento de animal(is) ou atropelamento de pessoa(s).
<b>Sequência</b> Combinação de sinistros de trânsito em que atropelamento de animal(is) a outros sinistros de trânsito, que ocorrem em sequência no mesmo evento.

Fonte: ABNT NBR 10697 (2020)

### 3.2 BASE DE DADOS - DATATRAN

Os dados abertos são um método de publicação de dados administrativos e institucionais em formato legível e reutilizável, sem restrições de licenças, patentes ou mecanismos de controle que visam aumentar a transparência e participação política dos cidadãos. Os dados considerados abertos podem ser usados de diversas maneiras, seja pelo governo ou sociedade, como, por exemplo, no desenvolvimento de aplicações desenvolvidas em parceria com a sociedade para exibir informações de forma gráfica ou interativa.

Disponibilizado pelo Departamento de Polícia Rodoviária Federal - DPRF, órgão responsável pela segurança viária, prevenção e repressão ao crime em rodovias e estradas federais, o Datatran é uma base de dados aberta que compila informações coletadas nos sistemas de informação da instituição, contém os registros de acidente nos mais de 75 mil quilômetros de rodovias e estradas em que atua, em todo o país.

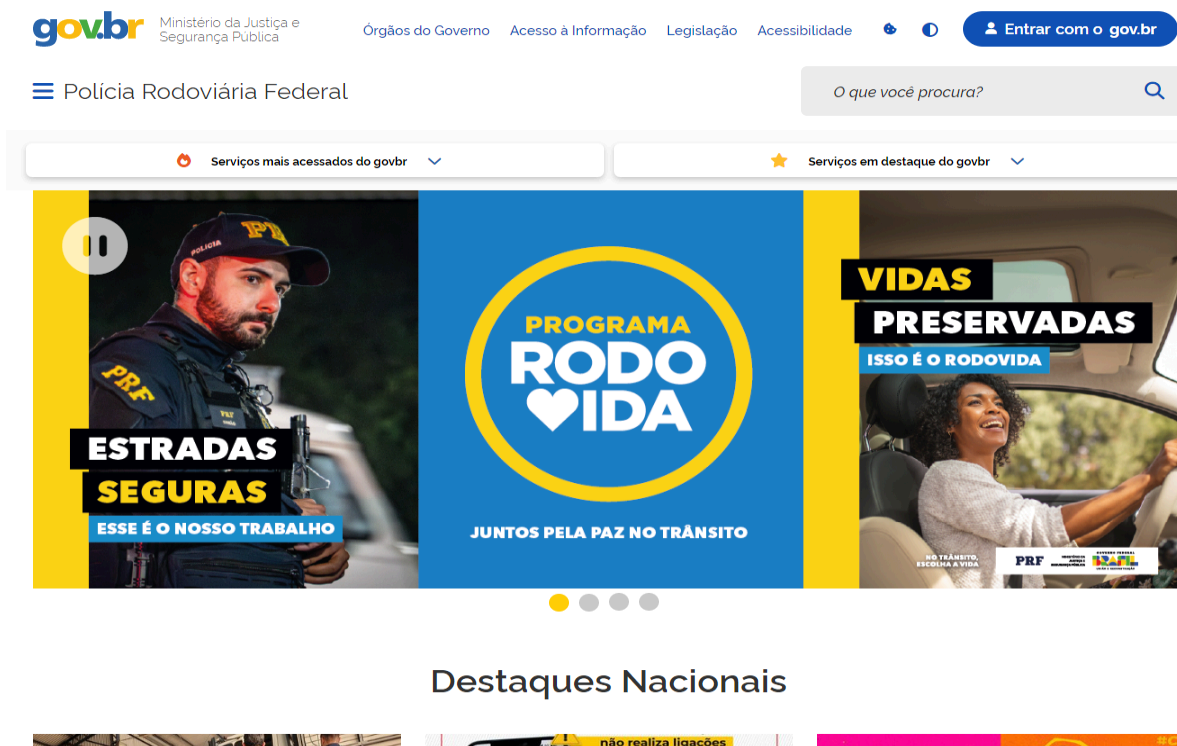
Figura 03 - Mapas das Rodovias Federais



Reprodução: Departamento Nacional de Infraestrutura de Transportes - DNIT.

Para o presente trabalho foi utilizado os arquivos disponibilizado no site oficial que condensa os dados de acidentes de trânsito provenientes de Boletins de Acidentes de Trânsito - BAT, no período de 2007 até 2023.

Figura 04 - Site oficial Polícia Rodoviária Federal



**Reprodução:** Departamento de Polícia Rodoviária Federal

### 3.3 BUSINESS INTELLIGENCE (BI)

Segundo Detoni Junior et al. (2018) e Rouhani et al. (2016), o Business Intelligence (BI) é um conceito que abrange as atividades e técnicas de coleta, armazenamento, análise e uso de dados relevantes para os negócios, visando melhorar o desempenho e a tomada de decisão das organizações. De acordo com Wang (2016) e Nguyen et al. (2016), o BI utiliza ferramentas e tecnologias que permitem transformar dados brutos em informações úteis, que auxilia as organizações dando suporte ao processo decisório com o intuito de gerar vantagem de mercado, compreender melhor seus clientes, concorrentes, processos e resultados.

Business Intelligence (BI), ou Inteligência de Negócios, é uma área essencial para as organizações nos dias atuais. Por meio de análises e processamento de dados, o BI

capacita as empresas a tomar decisões estratégicas embasadas em informações sólidas.

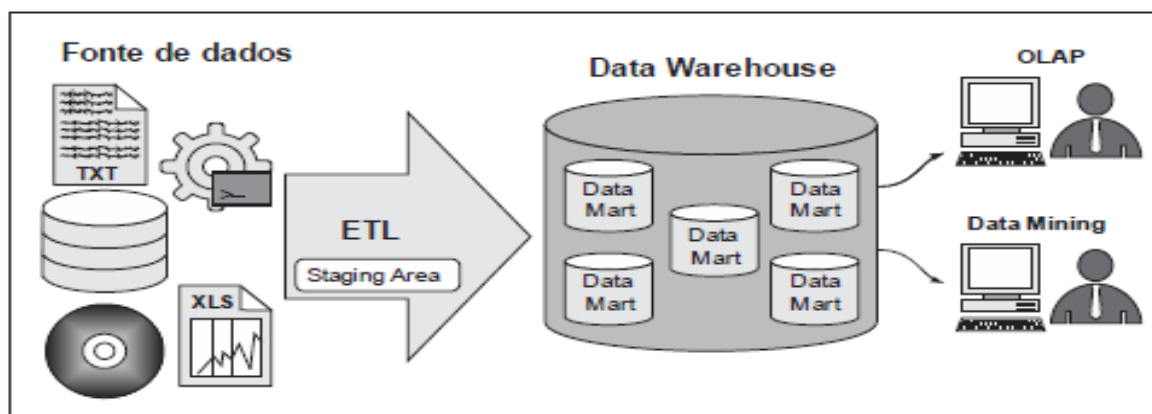
Com o BI podem:

- **Contribuição do BI para a Controladoria:** Um estudo de caso conduzido por pesquisadores investigou a contribuição das ferramentas de Business Intelligence para a área de controladoria. Essas ferramentas permitem que a controladoria forneça informações confiáveis, úteis e tempestivas ao processo decisório. A flexibilidade e a dinamicidade do BI melhoram os resultados das áreas organizacionais e da empresa como um todo. Reginato e Nascimento (2007);
- **Impacto nas Decisões Empresariais:** O BI não apenas fornece dados históricos, mas também permite análises preditivas e prescritivas. Através de modelos de negócios, visualizações e painéis, as organizações podem identificar tendências, oportunidades e riscos. Isso é fundamental para a tomada de decisões informadas e ágeis. Reginato e Nascimento (2007);
- **Evolução e Adoção do BI:** A evolução do BI ao longo das décadas reflete sua importância contínua. Inicialmente, o foco era compartilhar informações internamente. Hoje, o BI é uma ferramenta estratégica que abrange desde a coleta de dados até a análise avançada. A adoção de sistemas de BI também é observada no setor público, onde ele auxilia na gestão e na alocação eficiente de recursos. Pontes, Pontes, de Andrade (2021);
- **Desafios e Oportunidades:** Embora o BI ofereça inúmeras vantagens, também enfrenta desafios, como a integração de dados de várias fontes e a garantia de qualidade. No entanto, as organizações que investem em BI colhem benefícios significativos, como otimização de processos, redução de custos e vantagem competitiva. Reginato e Nascimento (2007)

O BI pode ser aplicado em diversos segmentos e setores de qualquer organização, desde o marketing, as vendas, o financeiro, o operacional, o logístico, o jurídico, o humano,

entre outros. O BI pode auxiliar na definição de produtos, preços, promoções, canais, segmentos, orçamentos, metas, planos, projetos, processos, políticas, entre outras questões relevantes para os negócios, nesse sentido o (BI) é de fundamental importância para as organizações que tem como intuito incrementar o desempenho estratégico e operacional. AUDZEYEVA E HUDSON (2016).

Figura 05 - Estrutura de um BI



Fonte: Reprodução ManageBI (2020)

### 3.3.1 MODELAGEM MULTIDIMENSIONAL

De acordo com Machado (2000), a modelagem multidimensional é uma forma de representar os dados de uma organização em uma estrutura lógica que facilita a análise e a tomada de decisão. Essa estrutura é composta por dimensões, que são as categorias de análise, e fatos, que são as medidas ou indicadores de desempenho.

Segundo Kimball e Ross (2013), a modelagem multidimensional é baseada no conceito de cubo, que é uma matriz tridimensional que contém os dados de uma determinada área de negócio. De acordo com Machado (2000), um fato é uma coleção de dados que contém informações de medidas e de contexto. Cada fato (evento) representa um elemento, uma transação ou um evento de negócio, e é utilizado para análise do processo de negócios de uma organização.

As dimensões são elementos que estão presentes em um fato (evento), como um assunto de negócios. As dimensões são as maneiras de se visualizar dados de forma sistemática: "por mês", "por região", "por produto", "por estado", etc.

Nesse sentido, as dimensões são importantes para a compreensão do contexto de um tema de negócios, por exemplo, um banco de dados que analise as vendas de produtos de uma organização Machado(2000). As dimensões, geralmente, não têm valor numérico,

pois são descritivas (ou qualitativas) e classificam os elementos que estão envolvidos em um fato (evento).

Ainda segundo Machado(2000), a modelagem multidimensional oferece diversas vantagens para a análise de dados, tais como:

- Permite que os dados sejam visualizados de diferentes ângulos e níveis de detalhe, de acordo com as suas necessidades de análise.
- Auxilia na busca de relações, tendências e padrões, que podem gerar insights e conhecimento para tomar melhores decisões.
- Melhora a velocidade e a qualidade de consultas analíticas, pois os dados são organizados e calculados de forma eficiente e consistente.
- Aumenta a satisfação e a confiança dos seus usuários, pois os dados são apresentados de forma simples, intuitiva e personalizada.
- A estrutura multidimensional é especialmente adequada para análises complexas, como análises de tendências, análises de mercado e análises de desempenho.
- Permite a integração de dados de várias fontes, proporcionando uma visão unificada e abrangente das informações da organização.

### **3.3.2 MODELO STAR SCHEMA**

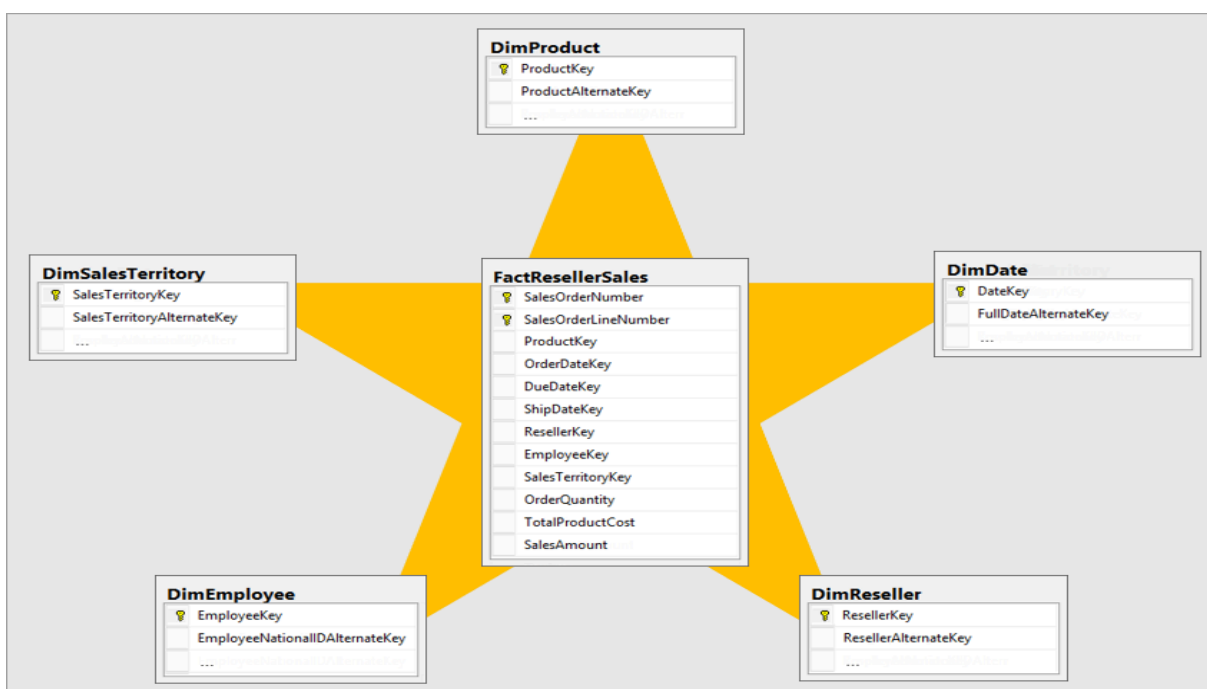
O modelo star schema é uma técnica de projeto físico de um banco de dados relacional que suporta a modelagem multidimensional. Esse modelo é chamado assim porque possui uma estrutura em forma de estrela, onde uma tabela central representa os fatos e as tabelas periféricas representam as dimensões. Kimball, R. (1996).

Segundo Kimball e Ross (2013), a tabela de fatos contém as medidas ou indicadores de desempenho de uma determinada área de negócio, que podem ser numéricos, como quantidade, valor, lucro, etc. A tabela de fatos também contém as chaves estrangeiras que se referem às tabelas de dimensões, formando os relacionamentos entre elas. As tabelas de dimensões contêm os atributos ou categorias de análise dos fatos, que podem ter vários níveis de granularidade, formando uma hierarquia. As tabelas de dimensões também contêm as chaves primárias que identificam cada registro de forma única.

Algumas vantagens do modelo star schema de acordo com Kimball e Ross (2013), para o armazenamento e a consulta dos dados multidimensionais são:

- **Simplicidade:** O modelo é fácil de entender e de navegar, pois possui poucas tabelas e relacionamentos, e segue a lógica do negócio
- **Desempenho:** É rápido e eficiente, pois diminui o número de junções e usa índices nas chaves;
- **Flexibilidade:** É flexível e adaptável, pois permite mudar ou adicionar dimensões e fatos sem mexer na estrutura existente;
- **Escalabilidade:** O modelo permite o crescimento dos dados sem comprometer a qualidade e a consistência.

Figura 06 - Modelagem Star Schema



**Disponível em:** <<https://learn.microsoft.com/en-us/power-bi/guidance/star-schema>>.

### 3.3.3 DATA WAREHOUSE

Segundo Inmon (2002), um data warehouse é um conjunto de dados orientado por assunto, integrado, não volátil e variante no tempo, que serve de suporte à tomada de decisão gerencial. Essas características significam que:

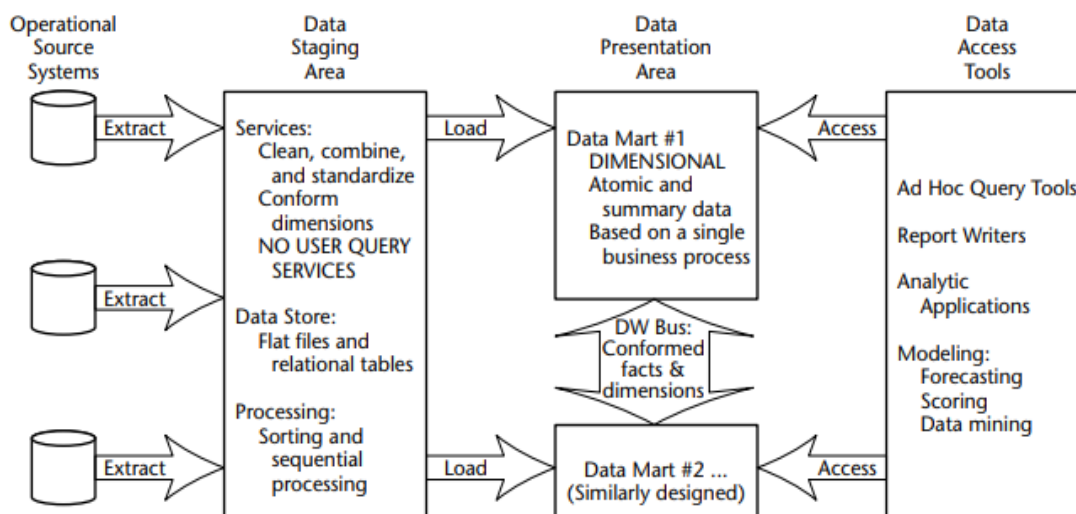
- **Orientado por assunto:** o data warehouse contém os dados relacionados a uma área específica de interesse da organização, como vendas, finanças, marketing, etc;
- **Integrado:** o data warehouse unifica os dados de diferentes fontes, sistemas e formatos, aplicando padrões de qualidade, consistência e integridade;

- **Não volátil:** o data warehouse não sofre alterações nos dados já armazenados, pois eles representam um registro histórico e confiável dos eventos ocorridos na organização;
- **Variante no tempo:** o data warehouse mantém os dados com a respectiva marcação temporal, permitindo a análise de tendências, comparações e evoluções ao longo do tempo.

Para realizar um data warehouse, existem diferentes metodologias e ferramentas disponíveis no mercado. Uma das mais utilizadas é a proposta por Kimball e Ross (2013), que consiste em quatro etapas principais:

- **Planejamento do projeto:** definição do escopo, objetivos, requisitos, recursos, riscos e o cronograma do projeto de data warehouse.
- **Modelagem dimensional:** definição do modelo lógico do data warehouse, utilizando o conceito de fatos e dimensões, que representam as medidas e as categorias de análise dos dados.
  - **Projeto físico:** implementação do modelo lógico em um banco de dados relacional ou multidimensional, utilizando um SGBD (Sistema Gerenciador de Banco de Dados), como o PostgreSQL, o SQL Server, o MySQL, etc.
- **Implementação e manutenção:** Carga dos dados das fontes operacionais para o data warehouse, utilizando um processo de ETL, bem como monitorar e atualizar o sistema.

Figura 07 - Exemplo de um Data Warehouse (DW)



Fonte: Reprodução Data Warehouse Toolkit (2013)

### 3.3.4 SISTEMAS OLAP

Segundo Machado (2000), o Sistema OLAP é um conjunto de ferramentas que permite a análise e exploração dos dados de um *Data Warehouse*. A análise denominada multidimensional apresenta os dados como sendo dimensões em vez de tabelas.

Ao combinar essas dimensões, o usuário tem uma perspectiva dos dados de um *Data Warehouse*, podendo realizar operações básicas, como slice and dice, que é uma forma de alterar as dimensões que serão exibidas, e drill up, que é a navegação entre os níveis de detalhamento (granularidade) dos dados de um *Data Warehouse*.

Ainda de acordo com Machado (2000), a análise multidimensional requer a utilização de operações típicas, tais como: os dez maiores municípios (por exemplo, os dez com o maior número de registros de acidentes), médias, somas e outras funções estatísticas.

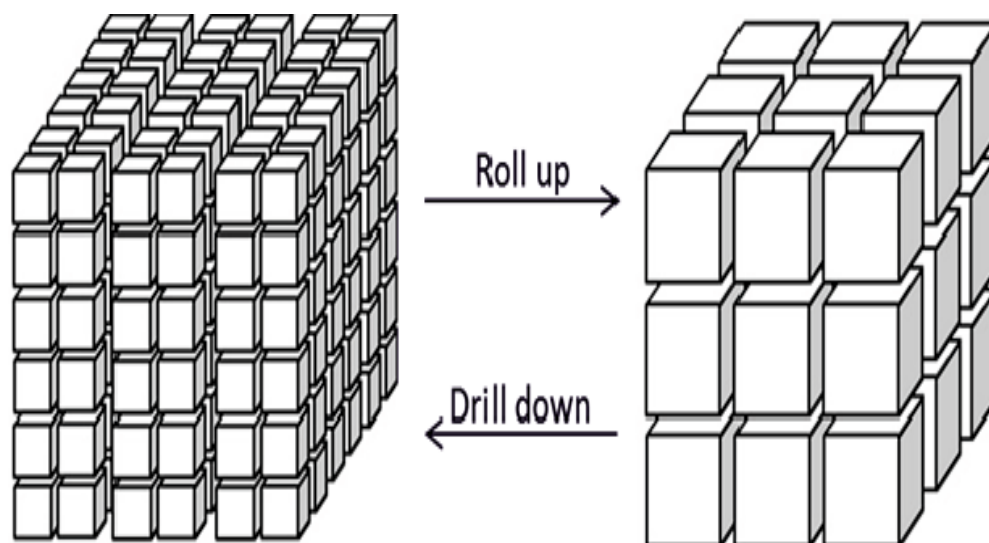
#### 3.3.4.1 OPERAÇÕES TÍPICAS (*DRILL UP, DRILL DOWN E SLICE AND DICE*)

Para Machado (2000), as operações mais utilizadas em um Sistema OLAP são:

**Drill Down:** Ocorre quando o usuário aumenta o nível de detalhe da informação, diminuindo a granularidade dos dados.

**Drill Up:** Ocorre quando o usuário diminui o nível de detalhe da informação, aumentando a granularidade dos dados.

Figura 08 - Exemplo de Drill Up e Drill Down

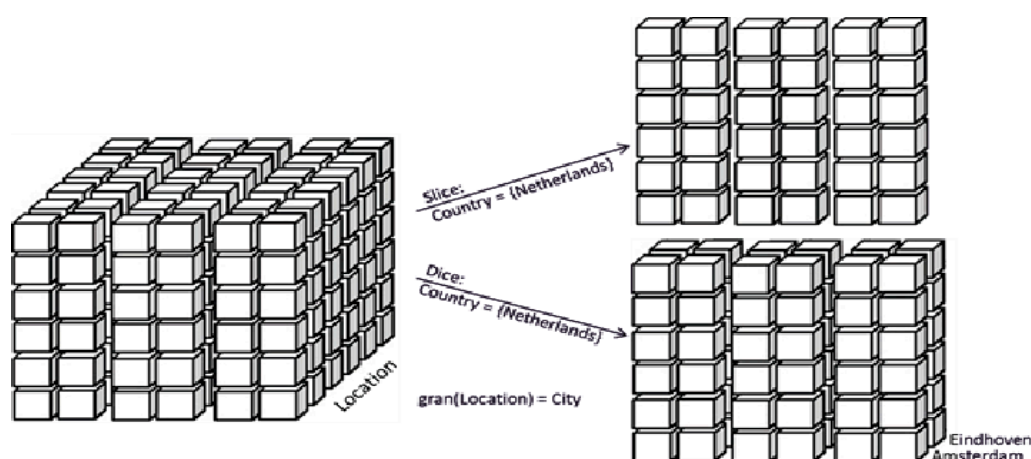


**Fonte:** Multidimensional Process Mining Using Process Cubes (2015)

**Slice and Dice:** São procedimentos que permitem navegar através dos dados na visualização de um cubo. *Slice and Dice* é a abreviação do escopo dos dados em análise, bem como a alteração da ordem das dimensões, alterando, dessa forma, a forma como os dados são vistos. Essa característica dos Sistemas OLAP é crucial, pois permite analisar as informações de diferentes ângulos, limitados apenas pela imaginação.

*Slice and Dice* significa definir um pedaço da base, uma “mordida”, como universo para uma consulta. *Slice and Dice* é o mesmo que realizar filtros.

Figura 09 - Exemplo de Slice and Dice

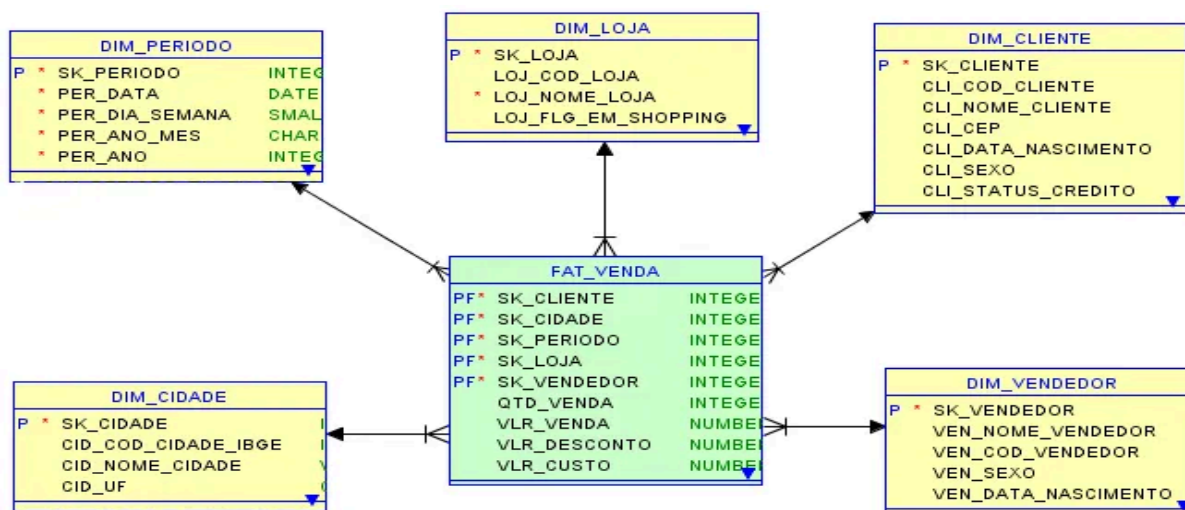


**Fonte:** Multidimensional Process Mining Using Process Cubes (2015)

### 3.3.4.2 ALGUNS TIPOS DE DIMENSÕES

Para Machado (2000), a Dimensão Tempo é crucial em todo modelo de dados de um *Data Mart*. É a única dimensão presente e obrigatória em todos os projetos de *Data Warehouse*. Esse fato crucial da Dimensão Tempo deve-se à relação da dimensão com a granularidade dos dados de um *Data Mart*. A Dimensão Tempo estabelece os limites para as janelas de tempo que podem ser criadas durante a existência de dados.

Figura 10 - Modelo Multidimensional com a Dimensão Tempo (DIM\_PERIODO)



Disponível em: <<https://encurtador.com.br/cfHZ5>>

De acordo com Machado (2000), a Dimensão Tempo é uma hierarquia de espaços de tempo. Por Exemplo:

- Ano
- Semestre
- Trimestre
- Mês
- Semana
- Dia

Ao analisar um fato (evento) em um *Data Mart*, não é relevante manter as informações de cada transação, mas sim todas as transações. O grão de análise é a unidade de tempo mínima que nos permitirá ter acesso, armazenar e sumarizar os dados.

### 3.3.5 ETL

O termo Extract, Transformation, Load (ETL) diz respeito a uma etapa crucial no processo de desenvolvimento de um *Data Warehouse* (DW). É responsável por mover dados oriundos de fontes diversas (Oracle, SQL Server, DB2, arquivo CSV, etc), adequá-los para um formato específico por meio de transformações, por fim fazer o

carregamento em um destino centralizado, comumente um data warehouse, business intelligence (BI), data lake ou banco de dados.

Nessa etapa ocorre o processo de extração de dados oriundos de fontes diversas, podendo ser composto por bancos de dados operacionais (OLTP), arquivos CSV, entre outras. Em seguida, é feita a realização de modificações, como filtragem, conversão de tipos de dados, renomear colunas, etc. Por fim, é realizada a carga dos dados em um Data Warehouse (DW). Kimball;Ross (2013).

Figura 11 - Processo de ETL



**Reprodução:** Indicium Academy

### 3.3.6 SLOWLY CHANGING DIMENSION (SCD)

É um conceito de modelagem de dados em ambientes Data Warehouse e Business Intelligence para o gerenciamento de mudanças em atributos de uma entidade ao longo do tempo, essas alterações ocorrem em uma frequência muito menor em comparação com as mudanças na tabela fato. Kimball e Ross (2013). Geralmente essas técnicas são aplicadas para correção de valores e mudança em alguma descrição, por exemplo.

Para Kromer (2022), existem vários tipos de Slowly Changing Dimension, porém somente algumas são utilizadas em projetos de BI. Os tipos de Slowly Changing Dimension (SCD) mais utilizados são:

**Tipo 0 (Fixo):**

No Tipo 0, as dimensões permanecem inalteradas ao longo do tempo. Os atributos não sofrem mudanças. Por exemplo, a data de nascimento e a pontuação de crédito original são consideradas dimensões do Tipo 0.

#### **Tipo 1 (Sobrescrita):**

Nesse método, os dados antigos são substituídos pelos novos. Não há histórico mantido no data warehouse. Por exemplo, se um fornecedor muda de estado, a entrada é atualizada com o novo estado, mas não há registro histórico.

#### **Tipo 2 (Adicionar Nova Linha):**

Este método rastreia dados históricos criando várias entradas para uma mesma chave natural nas tabelas dimensionais. Cada inserção preserva o histórico. Por exemplo, se um fornecedor muda de localização, uma nova linha é adicionada com um número de versão diferente.

### **3.4 ORACLE DATA MODELER**

O Oracle SQL Developer Data Modeler é uma ferramenta gráfica gratuita que aumenta a produtividade e simplifica as tarefas de modelagem de dados. Com o Oracle SQL Developer Data Modeler, os usuários têm a capacidade de criar, navegar e editar modelos lógicos, relacionais, físicos, multidimensionais e de tipo de dados. Essa ferramenta oferece suporte ao desenvolvimento colaborativo por meio do controle integrado de código-fonte e pode ser usada tanto em ambientes tradicionais quanto em ambientes de nuvem. Oracle (2014).

Alguns aspectos importantes do Oracle SQL Developer Data Modeler:

1. **Modelagem de Dados Multidimensional:** O Data Modeler permite a criação de modelos multidimensionais, essenciais para sistemas de data warehousing e análise OLAP. Esses modelos ajudam a representar dados complexos, como cubos e dimensões, facilitando a análise de grandes volumes de informações.
2. **Engenharia Direta e Reversa:** O Data Modeler oferece recursos de engenharia direta, permitindo que os usuários criem modelos a partir do

zero. Além disso, a engenharia reversa possibilita a importação de modelos a partir de bancos de dados existentes, facilitando a compreensão da estrutura de dados legada.

3. **Controle de Versão Integrado:** A versão mais recente do Oracle SQL Developer Data Modeler (20.3) inclui suporte para controle de versão por meio do Subversion, uma ferramenta de gerenciamento de código-fonte aberto. Isso permite o acesso colaborativo, permitindo que vários usuários trabalhem no mesmo design simultaneamente.

### 3.5 PENTAHO

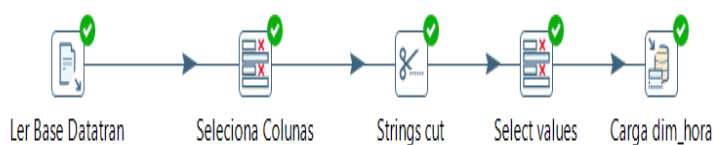
O Pentaho é uma suíte de ferramentas de código aberto desenvolvida em Java e de propriedade da Hitachi Vantara. Uma de suas funcionalidades é o Pentaho Data Integration (PDI), que se concentra nas tarefas de ETL (Extract, Transformation, Load). Vantara (2023).

Seus recursos são usados para gerenciar os dados em todas as etapas, desde a extração de dados até o desenvolvimento de dashboards.

#### 3.5.1 PENTAHO DATA INTEGRATION (PDI)

O Pentaho Data Integration (PDI), conhecido também como Kettle, é a ferramenta da suíte do Pentaho que executa as tarefas de Extração, Transformação e Carga (ETL). O PDI possui diversas funcionalidades para facilitar tarefas, como limpeza, migração e transformação de grandes volumes de dados. O PDI é composto por três componentes: Spoon, Kitchen e Pan. Vantara (2017).

Figura 12 - Pentaho Data Integration



Fonte: Autoria Própria (2024)

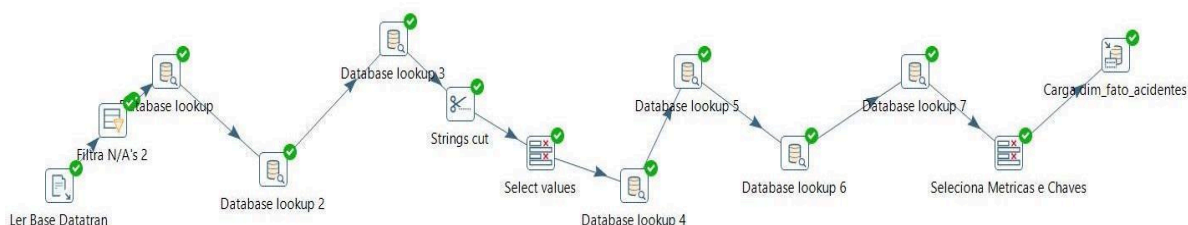
O Spoon é a aplicação que desenvolve as rotinas de ETL (Extract, Transformation, Load), além da geração de automatização de cargas (Jobs). Nessa ferramenta é onde estão localizadas todas as funcionalidades utilizadas para alteração e limpeza dos dados. O PDI possui uma interface gráfica que mostra os arquivos criados pelo Spoon, como por exemplo, a criação de transformações de dados e a criação de Jobs. Hatle et al. (2013).

O processo se inicia com a construção de um job, onde os "steps" são agregados e conectados, formando uma estrutura coesa para manipulação e movimentação dos dados entre diferentes fontes e destinos. Cada "step" é cuidadosamente escolhido e configurado de acordo com a sua função específica no processo ETL (Extract, Transformation, Load). Isso envolve a definição de parâmetros como conexões de banco de dados, consultas SQL e mapeamento de campos. CASTERS; BOUMAN; DONGEN (2010).

A essência do processo reside na habilidade de conectar os "steps" de maneira a estabelecer a direção do fluxo de dados. A ordem de execução é determinada por essa conexão, garantindo a lógica sequencial desejada para o processamento eficiente dos dados. A execução do fluxo de trabalho é iniciada através do botão "Run", sendo acompanhada por recursos de monitoramento e depuração que permitem a identificação e resolução eficaz de problemas durante a execução.

A flexibilidade do PDI é evidenciada pela capacidade de incorporar iterações e ciclos, permitindo a execução repetida de determinados "steps" com base em condições específicas, oferecendo adaptabilidade aos diversos cenários de integração.

Figura 13 - Exemplo de uma carga de dados no Pentaho PDI



Fonte: Autoria Própria (2024)

### 3.6 POSTGRESQL

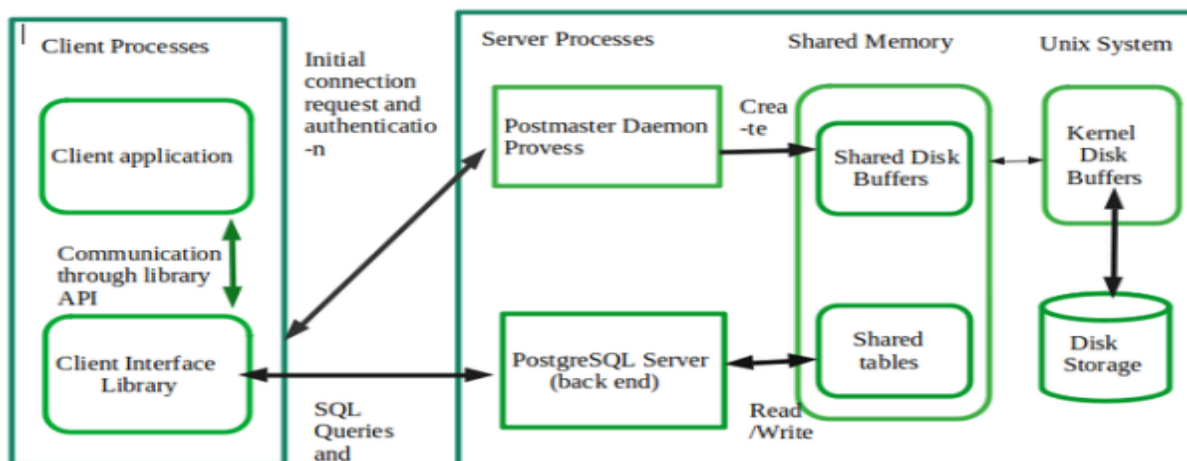
Para Stonebraker, Wong e Rowe (1991), o PostgreSQL, também conhecido como Postgres, é um Sistema de Gerenciamento de Banco de Dados Objeto-Relacional (SGBDOR) de código aberto e multiplataforma, que combina características de bancos de dados relacionais tradicionais com recursos avançados. Ele foi desenvolvido na Universidade da Califórnia, Berkeley, por Michael Stonebraker e sua equipe no final dos anos 80. Desde então, o PostgreSQL evoluiu continuamente, ganhando reconhecimento por sua robustez, escalabilidade e suporte ao padrão SQL99.

**Arquitetura e Componentes:** O PostgreSQL segue uma arquitetura cliente-servidor, com um processo de servidor principal (postmaster) que gerencia conexões de clientes e executa consultas SQL - *Structured Query Language*. Seus componentes incluem o mecanismo de armazenamento, o otimizador de consultas, o planejador e o executor. Esses componentes trabalham em conjunto para processar consultas de maneira eficiente. PostgreSQL Documentation (2023).

#### **Recursos Avançados do PostgreSQL:**

- **Tipos de Dados:** O PostgreSQL oferece uma ampla variedade de tipos de dados, incluindo tipos geométricos, de texto, de data/hora e até mesmo tipos personalizados definidos pelo usuário;
- **Extensibilidade:** Os usuários podem criar suas próprias funções, operadores e tipos de dados personalizados;
- **Índices Avançados:** Além dos índices B-tree padrão, o PostgreSQL suporta índices hash, GIN (Generalized Inverted Index) e GiST (Generalized Search Tree);
- **Transações e Controle de Concorrência:** O PostgreSQL segue o padrão ACID (Atomicidade, Consistência, Isolamento e Durabilidade) e oferece níveis de isolamento configuráveis.

Figura 14 - Arquitetura do sistema PostgreSQL



Disponível em: <<https://acervolima.com/postgresql-arquitetura-do-sistema/>>

### 3.7 MICROSOFT POWER BI (PBI)

Segundo Rigon et al. (2021), o Power BI, desenvolvido pela Microsoft, é uma ferramenta de business intelligence (BI) que permite a análise, visualização e compartilhamento de dados de maneira eficiente. Lançado em 2015, o Power BI se tornou uma escolha popular para empresas que buscam insights acionáveis a partir de seus dados.

#### Arquitetura e Componentes do Microsoft Power BI:

- **Power Query:** Essa ferramenta permite a conexão com várias fontes de dados, transformação e limpeza dos dados antes de serem carregados no Power BI;
- **Power Pivot:** O Power Pivot é um mecanismo de modelagem de dados que permite criar modelos dimensionais complexos e relacionamentos entre tabelas;
- **Power View e Power Map:** Essas ferramentas permitem a criação de visualizações interativas e mapas geográficos;
- **Power BI Service:** A versão online do Power BI, onde os relatórios e painéis são compartilhados e acessados por usuários autorizados.

#### Recursos Avançados do Microsoft Power BI. Microsoft (2024):

- **DAX (Data Analysis Expressions):** A linguagem DAX é usada para criar cálculos personalizados e medidas no Power BI.
- **Integração com Azure:** O Power BI pode se integrar ao Azure Data Lake, Azure SQL Database e outras soluções da Microsoft.
- **Gateway de Dados:** Permite a conexão segura entre o Power BI Service e fontes de dados locais.
- **M:** A **linguagem M** (também conhecida informalmente como “M”) é usada para **filtrar, combinar e transformar dados** durante o processo de importação.

Figura 15 - Exemplo do uso do Microsoft Power BI



Disponível em: <<https://xperiun.com/blog/exemplos-de-relatorios-com-power-bi/>>

### 3.7.1 DASHBOARDS ANALÍTICOS

Um painel (ou dashboard) é uma representação visual das informações mais importantes necessárias para atingir uma ou mais metas, combinadas e organizadas em uma tela para rastreamento rápido.

Segundo Fraga (2019), Dashboard significa painel de indicadores em tradução livre para o português, é uma maneira de visualizar e analisar informações de forma intuitiva, rápida e eficiente para diferentes níveis estratégicos de uma organização. O Power BI

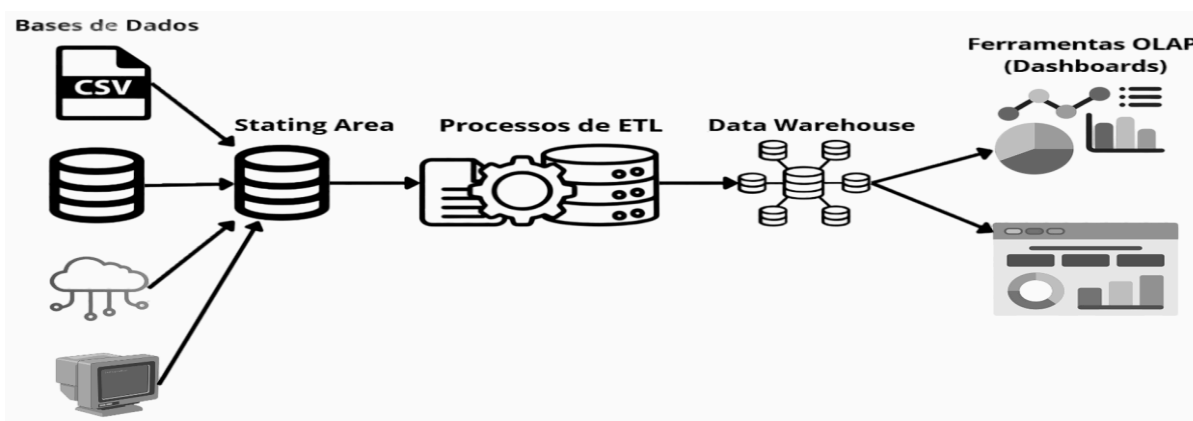
disponibiliza uma ampla gama de ferramentas gratuitas que são otimizadas para a criação e automatização de painéis.

Também de acordo com da Costa & Barros (2013), Um dashboard é um painel que cabe em uma tela e permite ver todas as informações necessárias para monitorar e analisar um processo ou negócio ao mesmo tempo. Permitem uma visão mais detalhada do estado da empresa ou do processo ao longo do tempo, tanto a nível global como específico, o que facilita os processos de tomada de decisão estratégica, analítica ou operacional. O Dashboard não procura informações detalhadas em toda a base de dados, mas sim reduzir as informações para que os profissionais possam entender e analisar melhor.

#### 4 BI-SAFEROAD: DADOS PARA SEGURANÇA NAS ESTRADAS

Neste tópico será abordada toda a construção da ferramenta BI-SafeRoad, utilizando a arquitetura de Business Intelligence (BI) a seguir:

Figura 16 - Arquitetura de BI



Fonte: Autoria Própria (2024)

#### 4.1 AQUISIÇÃO DE DADOS

O Governo Federal do Brasil disponibiliza vários arquivos com dados abertos para toda população brasileira. Para o presente trabalho, foram selecionados os dados referentes aos acidentes de trânsito em rodovias federais - Datatran, disponibilizado no portal da Polícia Rodoviária Federal - PRF, organizados por ano.

Figura 17 - Base de dados - PRF

gov.br Ministério da Justiça e Segurança Pública

Órgãos do Governo Acesso à Informação Legislação Acessibilidade

Entrar com o gov.br

Polícia Rodoviária Federal

O que você procura?

> Acesso à Informação > Dados Abertos > Acidentes

## Acidentes

Publicado em 05/08/2020 15h59 | Atualizado em 21/03/2023 20h13

Compartilhe: f X in

Observação: para melhor desempenho, e um download mais rápido, os arquivos estão compactados. Após baixá-los, é necessário descompactar os arquivos, para visualização do conteúdo. É preciso ter instalado um aplicativo no computador para descompactar os arquivos. Abaixo, algumas sugestões de aplicativos:

- WinRAR
- WinZip
- 7-zip
- PeaZip

Agrupados por ocorrência

- 2023 (Fevereiro)
- 2022
- 2021
- 2020
- 2019
- 2018

### Reprodução: Departamento de Polícia Rodoviária Federal (2024)

Os arquivos foram obtidos por meio de download diretamente no site, em formato CSV, para delimitação do trabalho foram selecionados os anos de 2019 a 2023, que condensam os dados de acidentes de trânsito provenientes de Boletins de Acidentes de Trânsito - BAT, temos a seguir a relação do campos do arquivo e suas respectivas descrições.

Quadro 10: Campos da base de dados datatran

Atributo	Descrição
Id	Campo em formato numérico, representando o identificador do acidente.
Data Inversa	Dia da ocorrência do acidente.
Dia Semana	Representa o dia do acidente
Horário	Horário do acidente.
UF	Unidade Federativa.
Br	Identificador da BR onde o acidente ocorreu.
Km	Quilômetro onde o acidente ocorreu.

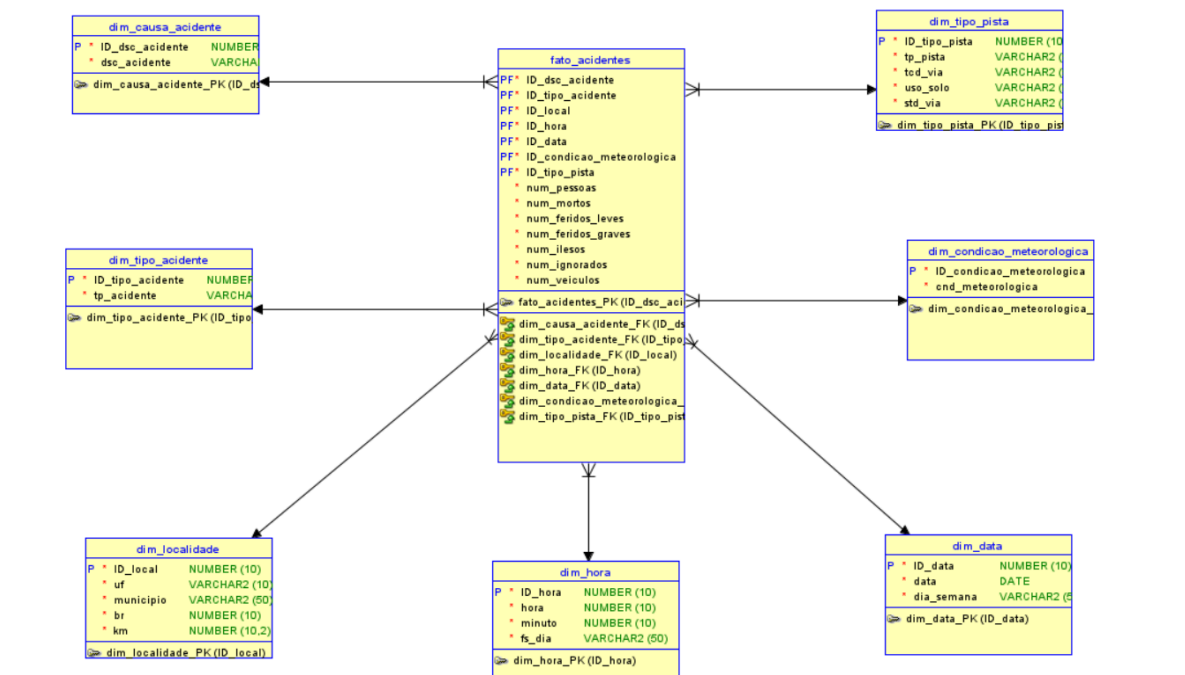
Município	Município do acidente.
Causa Acidente	Motivo do acidente.
Tipo Acidente	Tipo do acidente. Ex: Atropelamento
Classificação Acidente	Classificação quanto à gravidade do acidente: Ex: Com Vítimas
Fase Dia	Período do dia em que o acidente ocorreu.
Sentido Via	Sentido da via referente ao ponto do acidente.
Condição Meteorológica	Condição climática no momento do acidente.
Tipo Pista	Quantidade de faixas. Ex: pista simples.
Traçado Via	Define o tipo de via. Ex: via curva.
Uso Solo	Caracteriza o local do acidente. Ex: Urbano ou Rural.
Pessoas	Total de indivíduos envolvidos no acidente.
Mortos	Total de indivíduos mortos no acidente.
Feridos Leves	Total de indivíduos com ferimentos leves no acidente.
Feridos Graves	Total de indivíduos com ferimentos graves no acidente.
Ilesos	Total de indivíduos sem ferimentos no acidente.
Ignorados	Total de indivíduos envolvidos no acidente sem a informação do estado de saúde.
Feridos	Total de indivíduos envolvidos no acidente. (Feridos Leves + Feridos Graves)
Veículos	Total de veículos envolvidos no acidente.
Latitude	Latitude da ocorrência.
Longitude	Longitude da ocorrência.
Regional	Região de atuação.
Delegacia	Identificação da Delegacia Responsável.
Uop	Unidade Operacional Policial.

**Tabela 01:** Dicionário de dados (Original) - Datatran 2023.

## 4.2 MODELAGEM MULTIDIMENSIONAL DE DADOS

Para a construção da modelagem multidimensional dos dados foi escolhido o esquema estrela. Por ser um modelo que tem maior flexibilidade, desempenho e consultas analíticas simplificadas, o modelo em questão tem maior utilização nos sistemas de Business Intelligence (BI). A modelagem multidimensional foi desenvolvida no Oracle Data Modeler.

Figura 18 - Modelagem Star Schema.



Fonte: Autoria Própria (2024)

Para a criação do Data Warehouse usamos o SQL Data Modeler. O modelo contém 8 tabelas dimensões que são dim\_data, dim\_localidade, dim\_tipo\_acidente, dim\_tipo\_pista, dim\_condicao\_meteorologica, dim\_hora e dim\_causa\_acidente, além da tabela fato\_acidente.

- A dimensão dim\_localidade contém dados relacionados às localidades dos acidentes registrados.
- A dimensão dim\_data contém um período de 4 anos começando da data 01/01/2020.
- A dimensão dim\_tipo\_acidente contém os tipos de acidentes registrados.
- A dimensão dim\_tipo\_pista contém dados relacionados ao tipo da pista do acidente.
- A dimensão dim\_causa\_acidente contém dados relacionados às causas dos acidentes registrados.
- A dimensão dim\_condicao\_meteorologica contém dados relacionados às condições meteorológicas do dia do acidente.
- A dimensão dim\_hora contém dados relacionados às horas e fase do dia dos acidentes registrados.

A tabela fato\_acidentes contém os indicadores número de pessoas, número de mortos, número de feridos leves, número de feridos graves, número de ilesos, número de ignorados e número de veículos envolvidos nos acidentes.

### **4.3 PROCESSOS DE ETL (EXTRACT, LOAD, TRANSFORMATION) E PLANOS DE CARGA**

O processo de ETL se fez necessário para extração de dados contidos nos arquivos CSV fornecidos pela Polícia Rodoviária Federal -PRF, além de dados contidos em planilhas do Google Sheets, transformação desses dados e carga. Para o desenvolvimento dos processos de ETL (*Extract, Transformation, Load*), foi utilizada a ferramenta Pentaho PDI, ferramenta essa que é de código aberto (*Open Source*) e disponível de forma gratuita.

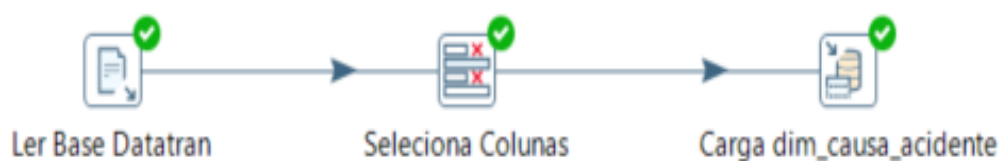
Foi utilizada a ferramenta Pentaho para criar as *transformations* que são responsáveis pelas cargas de dados das tabelas do banco de dados.

A transformation da dim\_causa\_acidente tem sua carga iniciada através de um arquivo .CSV contendo dados dos acidentes registrados nas estradas brasileiras.

Antes de entrar na parte das transformações, é importante frisar que as colunas UOP, Latitude e Longitude foram descartadas do processo de *ETL (Extract, Load, Transformation)* por não contribuírem para o propósito deste estudo, pois o foco do estudo é voltado para análise de fatores relacionados aos acidentes.

O *Step* Ler Base Datatran é o gatilho para Dados CSV, que é onde o *endpoint* dos dados é informado, no *step* seguinte foi selecionadas e renomeadas colunas com a *step* Select Values e foram gravados os dados em dim\_causa\_acidente com o *step* Insert/Update (Carga dim\_causa\_acidente):

Figura 19 - Carga Dim Causa Acidente



**Fonte:** Autoria Própria (2024)

A *transformation* da *dim\_condicao\_meteorologica* tem sua carga iniciada através de um arquivo .CSV contendo dados dos acidentes registrados nas estradas brasileiras. O *Step* Ler Base Datatran é o gatilho para arquivos CSV, que é onde o *endpoint* dos dados é informado. No step seguinte foi selecionadas e renomeadas colunas com a step *Select Values* e foram gravados os dados em *dim\_condicao\_meteorologica* com o step *Insert / Update* (Carga *dim\_condicao\_meteorologica*):

Figura 20 - Carga Dim Condição Meteorológica



**Fonte:** Autoria Própria (2024)

A *transformation* *dim\_hora* tem sua carga iniciada através de um arquivo .csv contendo dados dos acidentes registrados nas estradas brasileiras. O *Step* Ler Base Datatran é o gatilho para Dados .csv, que é onde o *endpoint* dos dados é informado, no step seguinte foi selecionadas e renomeadas colunas com a *step Select Values*, logo depois foi utilizado o *step String Cut* para transformar a hora e minuto para dar a carga no banco de dados de forma correta, depois foi novamente utilizado o *step Select Values* para selecionar os novos campos criados através do *step String Cut*, campos criados foram hora e minuto, e em seguida os dados foram gravados na dimensão *dim\_hora* com o *step Insert /Update* (Carga *dim\_hora*):

Figura 21 - Carga Dim Hora



**Fonte:** Autoria Própria (2024)

A *transformation* `dim_localidade` tem sua carga iniciada através de um arquivo .CSV contendo dados dos acidentes registrados nas estradas brasileiras. O *Step* Ler Base Datatran é o gatilho para dados CSV, que é onde o *endpoint* dos dados é informado, no *step* seguinte foi inserido o *step filter rows*, pois em algumas linhas da base da dados Datatran foram encontrados campos *string* onde a coluna era predominante campos numéricos, e com isso daria problema no plano de carga. No *step* seguinte foi selecionadas e renomeadas colunas com a *step Select Values* e foram gravados os dados em `dim_localidade` com o *step Insert/Update* (Carga `dim_localidade`):

Figura 22 - Carga Dim Localidade



**Fonte:** Autoria Própria (2024)

A *transformation* `dim_tipo_acidente` tem sua carga iniciada através de um arquivo .CSV contendo dados dos acidentes registrados nas estradas brasileiras. O *Step* Ler Base Datatran é o gatilho para Dados CSV, que é onde o *endpoint* dos dados é informado, no *step* seguinte foi selecionadas e renomeadas colunas com a *step Select Values* e foram gravados os dados em `dim_tipo_acidente` com o *step Insert/Update* (Carga `dim_tipo_acidente`):

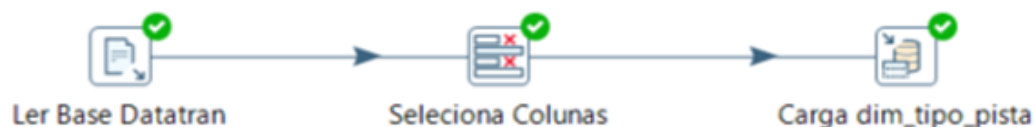
Figura 23 - Carga Dim Tipo Acidente



**Fonte:** Autoria Própria (2024)

A *transformation* `dim_tipo_pista` tem sua carga iniciada através de um arquivo .CSV contendo dados dos acidentes registrados nas estradas brasileiras. O *Step* `Ler Base Datatran` é o gatilho para dados .CSV, que é onde o endpoint dos dados é informado, no *step* seguinte foi selecionadas e renomeadas colunas com a *step* `Select Values` e foram gravados os dados em `dim_tipo_pista` com o *step* `Insert/Update` (Carga `dim_tipo_pista`):

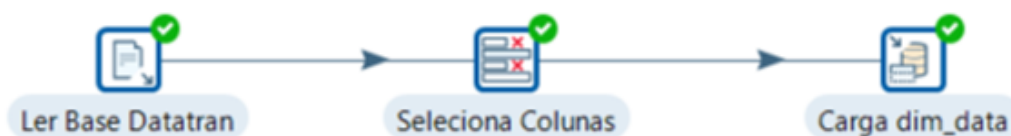
Figura 24 - Carga Dim tipo Pista



**Fonte:** Autoria Própria (2024)

A *transformation* `dim_data` tem sua carga iniciada através de um arquivo .csv contendo dados dos acidentes registrados nas estradas brasileiras. O *Step* `Ler Base Datatran` é o gatilho para Dados .csv, que é onde o endpoint dos dados é informado, no *step* seguinte foi selecionadas e renomeadas colunas com a *step* `Select Values` e foram gravados os dados em `dim_data` com o *step* `Insert /Update` (Carga `dim_data`):

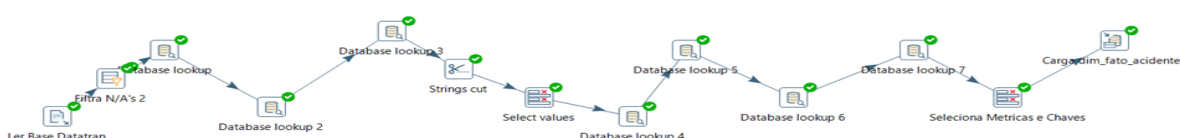
Figura 25 - Carga Dim Data



**Fonte:** Autoria Própria (2024)

A transformação `fato_acidentes` tem sua carga iniciada através de um arquivo CSV contendo dados dos acidentes registrados nas estradas brasileiras. O *Step Ler Base Datatran* é o gatilho para Dados CSV, que é onde o *endpoint* dos dados é informado, após a leitura do arquivo CSV, são inseridos os *Database Lookup*, ao total de 7 *steps* de *Database Lookup*, *step* utilizado para encontrar e realizar a inserção das chaves primárias das dimensões criadas na tabela `fato_acidentes`. Nesse processo, foram utilizados outros *steps*, como o *filter rows* para filtrar dados não informados em algumas colunas do arquivo CSV, *string cut* para novamente criar novos campos para carga correta da hora no banco de dados, *select values* para selecionar as colunas, e por último, o *step Insert/Update* para inserção dos dados na tabela `fato` (Carga `fato_acidentes`):

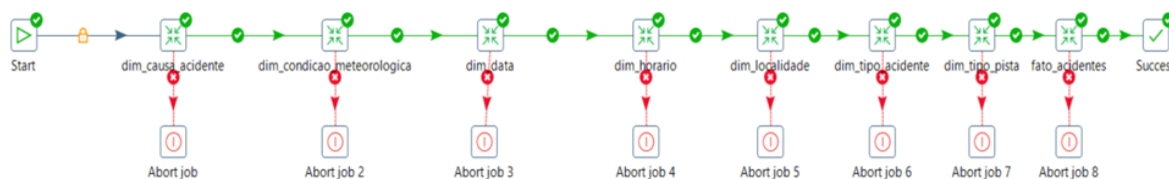
Figura 26 - Carga Fato Acidente



**Fonte:** Autoria Própria (2024)

Feitas todas as transformações, partimos para o *job*, que é onde ocorre o processo de execução dos planos de carga de todas as transformações realizadas no Pentaho, que foram as dimensões `dim_causa_acidente`, `dim_condicao_meteorologica`, `dim_data`, `dim_horario`, `dim_localidade`, `dim_tipo_acidente`, `dim_tipo_pista` e `fato_acidentes` dos acidentes registrados pela Polícia Rodoviária Federal (PRF) na base de dados aberta Datatran. Ao chegar no último *step* do *job*, o *Success*, o *job* é concluído e junto com ele, todo o processo de ETL também é finalizado.

Figura 27 - Job no Pentaho PDI



**Fonte:** Autoria Própria (2024)

O *job* se inicia no *start* (*step* padrão de um *job*), e depois disso todas as transformações são testadas, e se caso alguma transformação ocorrer uma falha no seu processo, o *job* é abortado, e se caso a transformação ocorrer bem, o *job* é um sucesso e segue para o próximo *step*.

#### 4.4 CRIAÇÃO DO DATA WAREHOUSE NO POSTGRESQL

Diante das limitações do sistema OLTP para o uso de Business Intelligence, visando a facilidade no processo de análise de dados e tomada de decisão, criamos um Data Warehouse que nos dará a possibilidade de analisar por exemplo: regiões com maior número de acidentes, data com menores números de acidentes, estados com maior média de acidentes.

Foi utilizado o PostgreSQL para a criação do nosso banco de dados baseado no modelo acima, e o Oracle data modeler para criação da modelagem do Data Warehouse.

Serão disponibilizados dois scripts SQL para criação do Data Warehouse, um script proveniente do data modeler é um script adaptado para o SGBD utilizado no projeto, o PostgreSQL.

Quadro 11: Script Oracle SQL Developer (Data Modeler)

— Criação da tabela dimensão `dim_causa_acidente`

```
CREATE TABLE dim_causa_acidente (
  id_dsc_acidente NUMBER(10) NOT NULL,
  dsc_acidente VARCHAR(50) NOT NULL
);
ALTER TABLE dim_causa_acidente ADD CONSTRAINT dim_causa_acidente_pk PRIMARY KEY (
id_dsc_acidente );
```

**— Criação da tabela dimensão dim\_condicao\_meteorologica**

```
CREATE TABLE dim_condicao_meteorologica (  
    id_condicao_meteorologica NUMBER(10) NOT NULL,  
    cnd_meteorologica VARCHAR2(50) NOT NULL  
);  
  
ALTER TABLE dim_condicao_meteorologica ADD CONSTRAINT dim_condicao_meteorologica_pk  
PRIMARY KEY ( id_condicao_meteorologica );
```

**— Criação da tabela dimensão dim\_data**

```
CREATE TABLE dim_data (  
    id_data NUMBER(10) NOT NULL,  
    data DATE NOT NULL,  
    dia_semana VARCHAR2(50) NOT NULL  
);  
  
ALTER TABLE dim_data ADD CONSTRAINT dim_data_pk PRIMARY KEY ( id_data );
```

**— Criação da tabela dimensão dim\_hora**

```
CREATE TABLE dim_hora (  
    id_hora NUMBER(10) NOT NULL,  
    minuto NUMBER(10) NOT NULL,  
    hora NUMBER(10) NOT NULL,  
    fs_dia VARCHAR2(50) NOT NULL  
);  
  
ALTER TABLE dim_hora ADD CONSTRAINT dim_hora_pk PRIMARY KEY ( id_hora );
```

**— Criação da tabela dimensão dim\_localidade**

```
CREATE TABLE dim_localidade (  
    id_local NUMBER(10) NOT NULL,  
    uf VARCHAR2(10) NOT NULL,  
    municipio VARCHAR2(50) NOT NULL,  
    br NUMBER(10) NOT NULL,  
    km NUMBER(10, 2) NOT NULL  
);
```

```
ALTER TABLE dim_localidade ADD CONSTRAINT dim_localidade_pk PRIMARY KEY ( id_local
);
```

— Criação da tabela dimensão dim\_tipo\_acidente

```
CREATE TABLE dim_tipo_acidente (
id_tipo_acidente NUMBER(10) NOT NULL,
tp_acidente VARCHAR2(50) NOT NULL
);
ALTER TABLE dim_tipo_acidente ADD CONSTRAINT dim_tipo_acidente_pk PRIMARY KEY (
id_tipo_acidente );
```

— Criação da tabela dimensão dim\_tipo\_pista

```
CREATE TABLE dim_tipo_pista (
id_tipo_pista NUMBER(10) NOT NULL,
tp_pista VARCHAR2(50) NOT NULL,
tcd_via VARCHAR2(50) NOT NULL,
uso_solo VARCHAR2(50) NOT NULL,
std_via VARCHAR2(50) NOT NULL
);
ALTER TABLE dim_tipo_pista ADD CONSTRAINT dim_tipo_pista_pk PRIMARY KEY (
id_tipo_pista );
```

— Criação da tabela fato fato\_acidentes

```
CREATE TABLE fato_acidentes (
id_dsc_acidente NUMBER(10) NOT NULL,
id_tipo_acidente NUMBER(10) NOT NULL,
id_local NUMBER(10) NOT NULL,
id_hora NUMBER(10) NOT NULL,
id_data NUMBER(10) NOT NULL,
id_condicao_meteorologica NUMBER(10) NOT NULL,
id_tipo_pista NUMBER(10) NOT NULL,
num_pessoas NUMBER(10) NOT NULL,
num_mortos NUMBER(10) NOT NULL,
num_feridos_leves NUMBER(10) NOT NULL,
```

```
num_feridos_graves    NUMBER(10) NOT NULL,  
num_ilesos            NUMBER(10) NOT NULL,  
num_ignorados         NUMBER(10) NOT NULL,  
num_veiculos          NUMBER(10) NOT NULL  
);
```

— Criação de chaves estrangeiras para a Tabela Fato fato\_acidentes:

```
ALTER TABLE fato_acidentes
```

```
    ADD CONSTRAINT fato_acidentes_pk PRIMARY KEY ( id_dsc_acidente,  
id_tipo_acidente,  
id_local,  
id_hora,  
id_data,  
id_condicao_meteorologica,  
id_tipo_pista );
```

```
ALTER TABLE fato_acidentes
```

```
    ADD CONSTRAINT dim_causa_acidente_fk FOREIGN KEY ( id_dsc_acidente )  
    REFERENCES dim_causa_acidente ( id_dsc_acidente );
```

```
ALTER TABLE fato_acidentes
```

```
    ADD CONSTRAINT dim_condicao_meteorologica_fk FOREIGN KEY(  
id_condicao_meteorologica)  
    REFERENCES dim_condicao_meteorologica (id_condicao_meteorologica);
```

```
ALTER TABLE fato_acidentes
```

```
    ADD CONSTRAINT dim_data_fk FOREIGN KEY ( id_data )  
    REFERENCES dim_data (id_data);
```

```
ALTER TABLE fato_acidentes
```

```
    ADD CONSTRAINT dim_hora_fk FOREIGN KEY (id_hora)
```

```
REFERENCES dim_hora (id_hora);
```

```
ALTER TABLE fato_acidentes
```

```
ADD CONSTRAINT dim_localidade_fk FOREIGN KEY (id_local)
```

```
REFERENCES dim_localidade (id_local);
```

```
ALTER TABLE fato_acidentes
```

```
ADD CONSTRAINT dim_tipo_acidente_fk FOREIGN KEY (id_tipo_acidente)
```

```
REFERENCES dim_tipo_acidente (id_tipo_acidente);
```

```
ALTER TABLE fato_acidentes
```

```
ADD CONSTRAINT dim_tipo_pista_fk FOREIGN KEY (id_tipo_pista)
```

```
REFERENCES dim_tipo_pista (id_tipo_pista)
```

**Fonte:** Autoria própria (2024).

**Script adaptado PostgreSQL:**

— **Criação da tabela dimensão dim\_causa\_acidente**

```
CREATE TABLE dim_causa_acidente (
```

```
id_dsc_acidente NUMBER(10) NOT NULL,
```

```
dsc_acidente VARCHAR2(50) NOT NULL
```

```
);
```

```
ALTER TABLE dim_causa_acidente ADD CONSTRAINT dim_causa_acidente_pk PRIMARY KEY (
id_dsc_acidente );
```

— **Criação da tabela dimensão dim\_condicao\_meteorologica**

```
CREATE TABLE dim_condicao_meteorologica (
```

```
id_condicao_meteorologica NUMBER(10) NOT NULL,
```

```
cond_meteorologica VARCHAR2(50) NOT NULL
```

```
);
```

```
ALTER TABLE dim_condicao_meteorologica ADD CONSTRAINT dim_condicao_meteorologica_pk  
PRIMARY KEY ( id_condicao_meteorologica );
```

— **Criação da tabela dimensão dim\_data**

```
CREATE TABLE dim_data (  
    id_data NUMBER(10) NOT NULL,  
    data DATE NOT NULL,  
    dia_semana VARCHAR2(50) NOT NULL  
);  
ALTER TABLE dim_data ADD CONSTRAINT dim_data_pk PRIMARY KEY ( id_data );
```

— **Criação da tabela dimensão dim\_hora**

```
CREATE TABLE dim_hora (  
    id_hora NUMBER(10) NOT NULL,  
    hora NUMBER(10) NOT NULL,  
    minuto NUMBER(10) NOT NULL,  
    fs_dia VARCHAR2(50) NOT NULL  
);  
ALTER TABLE dim_hora ADD CONSTRAINT dim_hora_pk PRIMARY KEY ( id_hora);
```

— **Criação da tabela dimensão dim\_localidade**

```
CREATE TABLE dim_localidade (  
    id_local NUMBER(10) NOT NULL,  
    uf VARCHAR2(10) NOT NULL,  
    municipio VARCHAR2(50) NOT NULL,  
    br NUMBER(10) NOT NULL,  
    km NUMBER(10, 2) NOT NULL  
);
```

```
ALTER TABLE dim_localidade ADD CONSTRAINT dim_localidade_pk PRIMARY KEY ( id_local
);
```

— **Criação da tabela dimensão dim\_tipo\_acidente**

```
CREATE TABLE dim_tipo_acidente (
    id_tipo_acidente NUMBER(10) NOT NULL,
    tp_acidente   VARCHAR2(50) NOT NULL
);
```

```
ALTER TABLE dim_tipo_acidente ADD CONSTRAINT dim_tipo_acidente_pk PRIMARY KEY (
id_tipo_acidente );
```

— **Criação da tabela dimensão dim\_tipo\_pista**

```
CREATE TABLE dim_tipo_pista (
    id_tipo_pista NUMBER(10) NOT NULL,
    tp_pista   VARCHAR2(50) NOT NULL,
    tcd_via   VARCHAR2(50) NOT NULL,
    uso_solo  VARCHAR2(50) NOT NULL,
    std_via   VARCHAR2(50) NOT NULL
);
```

```
ALTER TABLE dim_tipo_pista ADD CONSTRAINT dim_tipo_pista_pk PRIMARY KEY (
id_tipo_pista );
```

— **Criação da Tabela Fato fato\_acidentes:**

```
CREATE TABLE fato_acidentes (
    id_dsc_acidente   NUMBER(10) NOT NULL,
    id_tipo_acidente  NUMBER(10) NOT NULL,
    id_local           NUMBER(10) NOT NULL,
    id_hora            NUMBER(10) NOT NULL,
    id_data            NUMBER(10) NOT NULL,
```

```

id_condicao_meteorologica NUMBER(10) NOT NULL,
id_tipo_pista          NUMBER(10) NOT NULL,
num_pessoas           NUMBER(10) NOT NULL,
num_mortos            NUMBER(10) NOT NULL,
num_feridos_leves    NUMBER(10) NOT NULL,
num_feridos_graves   NUMBER(10) NOT NULL,
num_ilesos            NUMBER(10) NOT NULL,
num_ignorados        NUMBER(10) NOT NULL,
num_veiculos          NUMBER(10) NOT NULL
);

```

— Criação da Tabela Fato fato\_acidentes:

```

CREATE TABLE fato_acidentes (
  id_dsc_acidente      NUMBER(10) NOT NULL,
  id_tipo_acidente     NUMBER(10) NOT NULL,
  id_local             NUMBER(10) NOT NULL,
  id_hora              NUMBER(10) NOT NULL,
  id_data              NUMBER(10) NOT NULL,
  id_condicao_meteorologica NUMBER(10) NOT NULL,
  id_tipo_pista        NUMBER(10) NOT NULL,
  num_pessoas          NUMBER(10) NOT NULL,
  num_mortos           NUMBER(10) NOT NULL,
  num_feridos_leves   NUMBER(10) NOT NULL,
  num_feridos_graves  NUMBER(10) NOT NULL,
  num_ilesos           NUMBER(10) NOT NULL,
  num_ignorados        NUMBER(10) NOT NULL,
  num_veiculos         NUMBER(10) NOT NULL
);

```

```
);
```

— Criação de chaves estrangeiras para a Tabela Fato fato\_acidentes:

```
ALTER TABLE fato_acidentes
```

```
ADD CONSTRAINT fato_acidentes_pk PRIMARY KEY ( id_dsc_acidente,  
id_tipo_acidente,  
id_local,  
id_hora,  
id_data,  
id_condicao_meteorologica,  
id_tipo_pista );
```

```
ALTER TABLE fato_acidentes
```

```
ADD CONSTRAINT dim_causa_acidente_fk FOREIGN KEY ( id_dsc_acidente )  
REFERENCES dim_causa_acidente ( id_dsc_acidente );
```

```
ALTER TABLE fato_acidentes
```

```
ADD CONSTRAINT dim_condicao_meteorologica_fk FOREIGN KEY(  
id_condicao_meteorologica)  
REFERENCES dim_condicao_meteorologica (id_condicao_meteorologica);
```

```
ALTER TABLE fato_acidentes
```

```
ADD CONSTRAINT dim_data_fk FOREIGN KEY ( id_data )  
REFERENCES dim_data (id_data);
```

```
ALTER TABLE fato_acidentes
```

```
ADD CONSTRAINT dim_hora_fk FOREIGN KEY (id_hora)
```

```
REFERENCES dim_hora (id_hora);

ALTER TABLE fato_acidentes
ADD CONSTRAINT dim_localidade_fk FOREIGN KEY (id_local)
REFERENCES dim_localidade (id_local);

ALTER TABLE fato_acidentes
ADD CONSTRAINT dim_tipo_acidente_fk FOREIGN KEY (id_tipo_acidente)
REFERENCES dim_tipo_acidente (id_tipo_acidente);

ALTER TABLE fato_acidentes
ADD CONSTRAINT dim_tipo_pista_fk FOREIGN KEY (id_tipo_pista)
REFERENCES dim_tipo_pista (id_tipo_pista);
```

**Fonte:** Autoria própria (2024).

#### 4.5 DICIONÁRIO DE DADOS

O dicionário de dados é uma ferramenta essencial em projetos de gestão de dados. Ele atua como uma ponte entre o conhecimento técnico e o acesso à informação, fornecendo um catálogo organizado de todos os elementos de dados relevantes para um sistema, projeto ou banco de dados.

Definição básica para o dicionário de dados do Datatran:

1. **Tabela de Dados:** Uma lista de todas as tabelas presentes no Datatran, juntamente com suas descrições e campos associados.
2. **Campos:** Cada campo em uma tabela deve ser definido, incluindo seu nome, tipo de dados, tamanho e significado.
3. **Relações:** Se houver relações entre tabelas (por exemplo, chaves estrangeiras), essas relações devem ser documentadas.
4. **Metadados:** Informações adicionais, como data de criação, autor, versão etc., também podem ser incluídas.

Quadro 12: Dicionário de Dados

Nome do Atributo	Tipo de Dado	Tamanho	Descrição	Restrição
id_dsc_acidente	SERIAL	4	Identificador único da causa do acidente	Primary Key, Not Null
dsc_acidente	VARCHAR	50	Descrição da causa do acidente, como colisão, atropelamento,etc..	Not Null
id_condicao_meteorologica	SERIAL	4	Identificador único da condição meteorológica	Primary Key, Not Null
cnd_meteorologica	VARCHAR	50	Descrição da causa do acidente, como ensolarado, chuvoso,etc..	Not Null
id_data	SERIAL	4	Identificador único da data	Primary Key, Not Null
data	DATE	4	A data do acidente no formato AAAA-MM-DD	Not Null
dia_semana	VARCHAR	50	O nome do dia da semana correspondente à data do acidente, como segunda-feira	Not Null
id_hora	SERIAL	4	Identificador único da hora	Primary Key, Not Null
hora	INTEGER	4	A hora do acidente, no formato 24 horas, de 0 à 23	Not Null
minuto	INTEGER	4	O minuto do acidente, de 0 à 59	Not Null
horario	TIME	8	O horário do acidente no formato HH:MM:SS	Not Null
fs_dia	VARCHAR	50	A faixa do dia correspondente à hora do acidente, como manhã, tarde, noite etc..	Not Null

id_local	SERIAL	4	Identificador único da local	Primary Key, Not Null
uf	VARCHAR	10	A sigla da unidade federativa onde ocorreu o acidente, como AL, PE, SE, etc...	Not Null
município	VARCHAR	50	Nome do município onde ocorreu o acidente, como Rio Largo, Escada, etc..	Not Null
br	NUMERIC	10	Número da rodovia onde ocorreu o acidente, como 101, 118, 126, etc..	Not Null
km	DECIMAL	10,2	Quilômetro da rodovia federal onde ocorreu o acidente, com duas casas decimais, como 12.34, 56.78, etc..	Not Null
id_tipo_acidente	SERIAL	4	Identificador único do tipo do acidente	Primary Key, Not Null
tp_acidente	VARCHAR	50	A descrição do tipo do acidente, como capotamento, saída de pista, tombamento, etc...	Not Null
id_tipo_pista	SERIAL	4	Identificador único do tipo da pista	Primary Key, Not Null
tp_pista	VARCHAR	50	A descrição do tipo da pista, como simples, dupla, etc...	Not Null
tcd_via	VARCHAR	50	A descrição do tipo de condição da via, como pavimentada, não pavimentada, etc..	Not Null
uso_solo	VARCHAR	50	A descrição do uso do solo, na área do acidente, como urbano, rural, etc...	Not Null

std_via	VARCHAR	50	A descrição do estado de via no momento do acidente, como seca, molhada, escorregadia, etc...	Not Null
ID_dsc_acidente	INTERGER	4	Identificador único do acidente	Primary Key, Not Null
ID_tipo_acidente	INTERGER	4	Identificador único do tipo do acidente	Primary Key, Not Null
ID_local	INTERGER	4	Identificador único do local do acidente	Primary Key, Not Null
ID_hora	INTERGER	4	Identificador único da hora do acidente	Primary Key, Not Null
ID_data	INTERGER	4	Identificador único da data do acidente	Primary Key, Not Null
ID_condicao_meteorologica	INTERGER	4	Identificador único da condição meteorológica	Primary Key, Not Null
ID_tipo_pista	INTERGER	4	Identificador único do tipo de pista	Primary Key, Not Null
num_pessoas	NUMERIC	10	Número de pessoas envolvidas no acidente	Not Null
num_mortos	NUMERIC	10	Número de mortos envolvidos no acidente	Not Null
num_feridos_leves	NUMERIC	10	Número de feridos leves no acidente	Not Null
num_feridos_graves	NUMERIC	10	Número de feridos graves no acidente	Not Null
num_ilesos	NUMERIC	10	Número de ilesos envolvidos no acidente	Not Null
num_ignorados	NUMERIC	10	Número de ignorados nos acidente	Not Null
num_veiculos	NUMERIC	10	Número de veículos envolvidos no acidente	Not Null

Quadro 13: Dicionário de Dados das Chaves Estrangeiras

Nome da Restrição	Tipo da Restrição	Coluna da Tabela Fato	Coluna da Tabela Dimensão	Tabela Dimensão
dim_causa_acidente_fk	FOREIGN KEY	ID_dsc_acidente	id_dsc_acidente	dim_causa_acidente
dim_condicao_meteorologica_fk	FOREIGN KEY	ID_condicao_meteorologica	id_condicao_meteorologica	dim_condicao_meteorologica
dim_data_fk	FOREIGN KEY	ID_data	id_data	dim_data
dim_hora_fk	FOREIGN KEY	ID_hora	id_hora	dim_hora
dim_localidade_fk	FOREIGN KEY	ID_localidade	id_localidade	dim_localidade
dim_tipo_acidente_fk	FOREIGN KEY	ID_tipo_acidente	id_tipo_acidente	dim_tipo_acidente
dim_tipo_pista_fk	FOREIGN KEY	ID_tipo_pista	id_tipo_pista	dim_tipo_pista

**Fonte:** Autoria própria (2024).

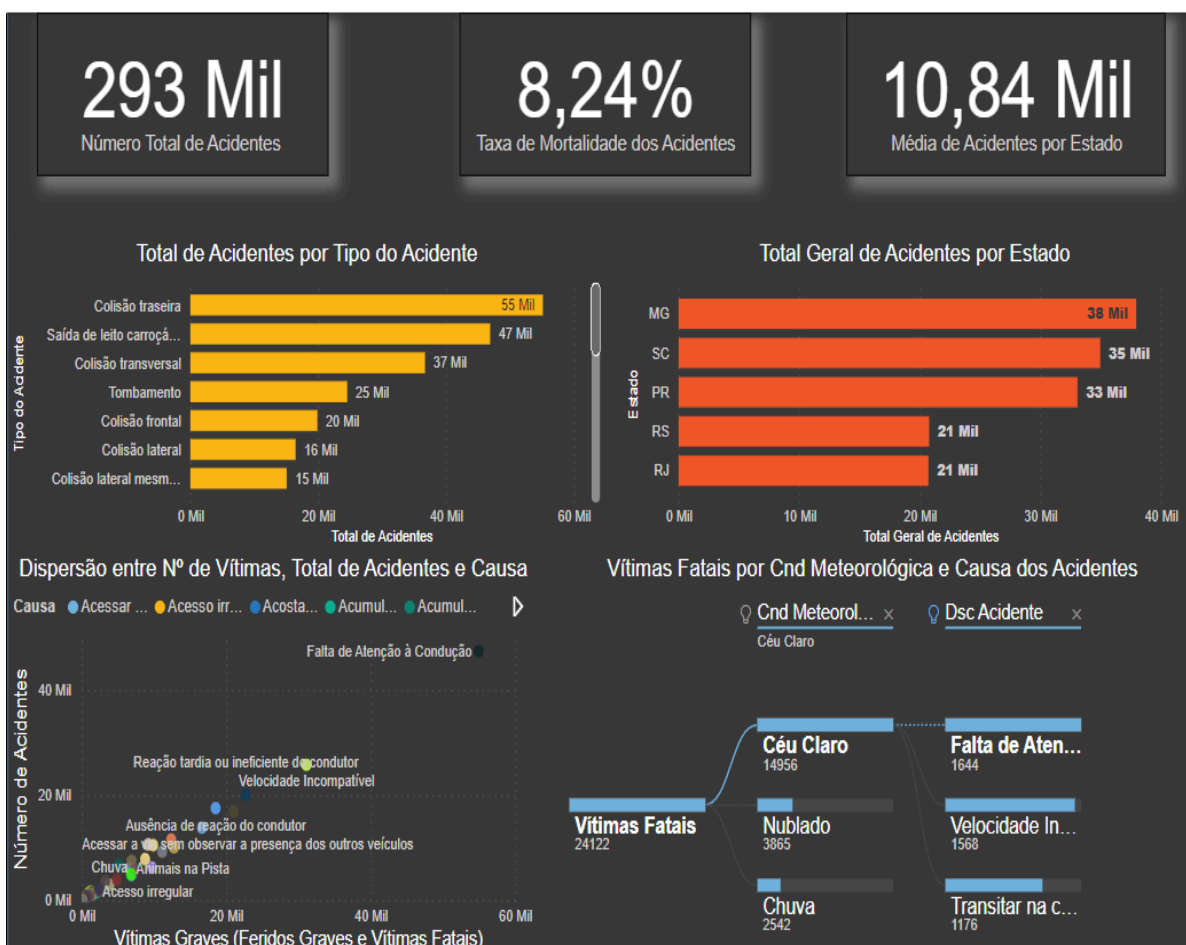
#### 4.6 ANÁLISE DE DADOS

Após o processo de *ETL (Extract, Load, Transformation)* e conexão do banco de dados PostgreSQL com a ferramenta de análise de dados Microsoft Power BI, foi possível realizar as análises dos dados através da construção de painéis analíticos com gráficos interativos e *KPIs (Key Performance Indicators)* para melhor análise e identificação de tendências e padrões dos acidentes.

A primeira análise realizada foi Análise de Vítimas, Condição Meteorológica e Causas de Acidente. Os *KPIs* utilizados foram “Número Total de Acidentes”, “Taxa de Mortalidade dos Acidentes” e “Média de Acidentes por Estado”. Os gráficos trabalhados foram “Total de Acidentes por Tipo de Acidente”, utilizando gráfico de barras empilhadas, “Total de Acidentes por Estado”, utilizando gráfico de barras clusterizado, “Dispersão entre Número de Vítimas Graves (Feridos Graves e Mortos), Total de Acidentes e Causa dos Acidentes”, utilizando um gráfico de dispersão em bolhas, e “Número de Vítimas Fatais”, Condição Meteorológica e Causa do Acidente”, utilizando uma árvore hierárquica,

analisando por “Número de Mortos” e explicando por “Condição Meteorológica” e “Causa do Acidente”.

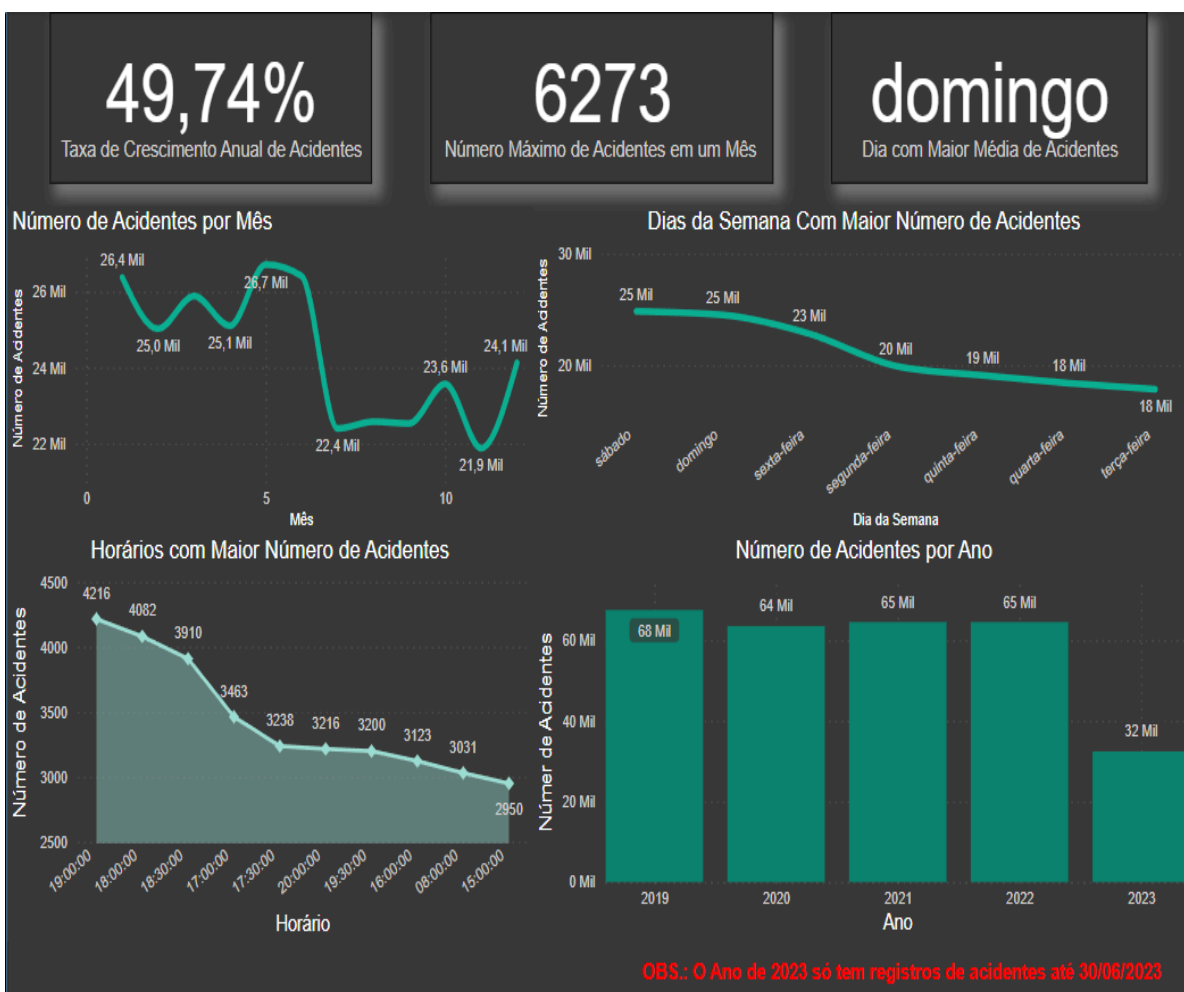
Figura 28 - Análise de Vítimas, Condições Meteorológicas e Fatores de Acidentes



Fonte: Autoria Própria (2024)

A segunda análise realizada foi a análise temporal. Os KPIs utilizados foram “Taxa de Crescimento Anual de Acidentes”, “Número Máximo de Acidentes em um Mês” e “Dia com Maior Média de Acidentes”. Os gráficos trabalhados foram “Número de Acidentes por Mês”, onde foi utilizado um gráfico de linhas, “Dias da Semana com Maior Número de Acidentes”, onde também foi utilizado um gráfico de linhas para mostrar os dias da semana de forma decrescente de acordo com o número de acidentes, “Horários com Maior Número de Acidentes”, onde foi utilizado um gráfico de área e um filtro de horários onde é possível identificar os 10 horários com maiores ocorrências de ocorrências, e para finalizar um gráfico de evolução do Número de Acidentes ao Longo dos Anos, utilizando um gráfico de colunas agrupadas e linha.

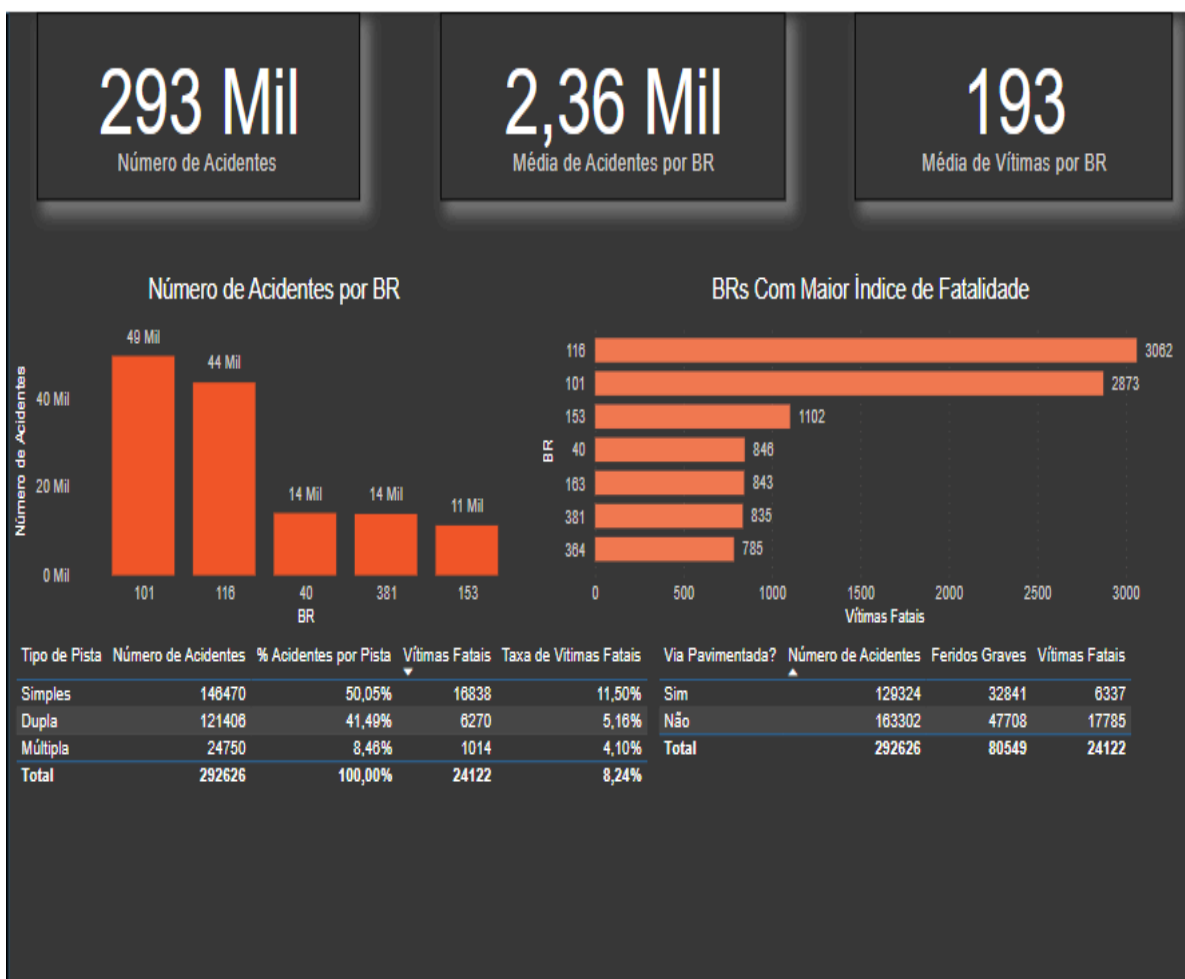
Figura 29 - Análise Temporal



**Fonte:** Autoria Própria (2024)

A terceira análise realizada foi a análise de BRs, Pista e Uso Solo. Os *KPIs* utilizados foram “Total de Acidentes”, “Média de Acidentes por BR” e “Média de Vítimas Fatais por BR”. Os gráficos desenvolvidos foram “Número de Acidentes por BR”, “BRs com Maior Índice de Fatalidade”, tabela com “Tipo de Pista, Número de Acidentes por Tipo de Pista, Taxa (%) de acidentes por Tipo de Pista”, “Vítimas Fatais” e “Taxa (%) de Vítimas Fatais”, e por último uma tabela contendo se a via é pavimentada, número de acidentes, feridos graves e vítimas fatais.

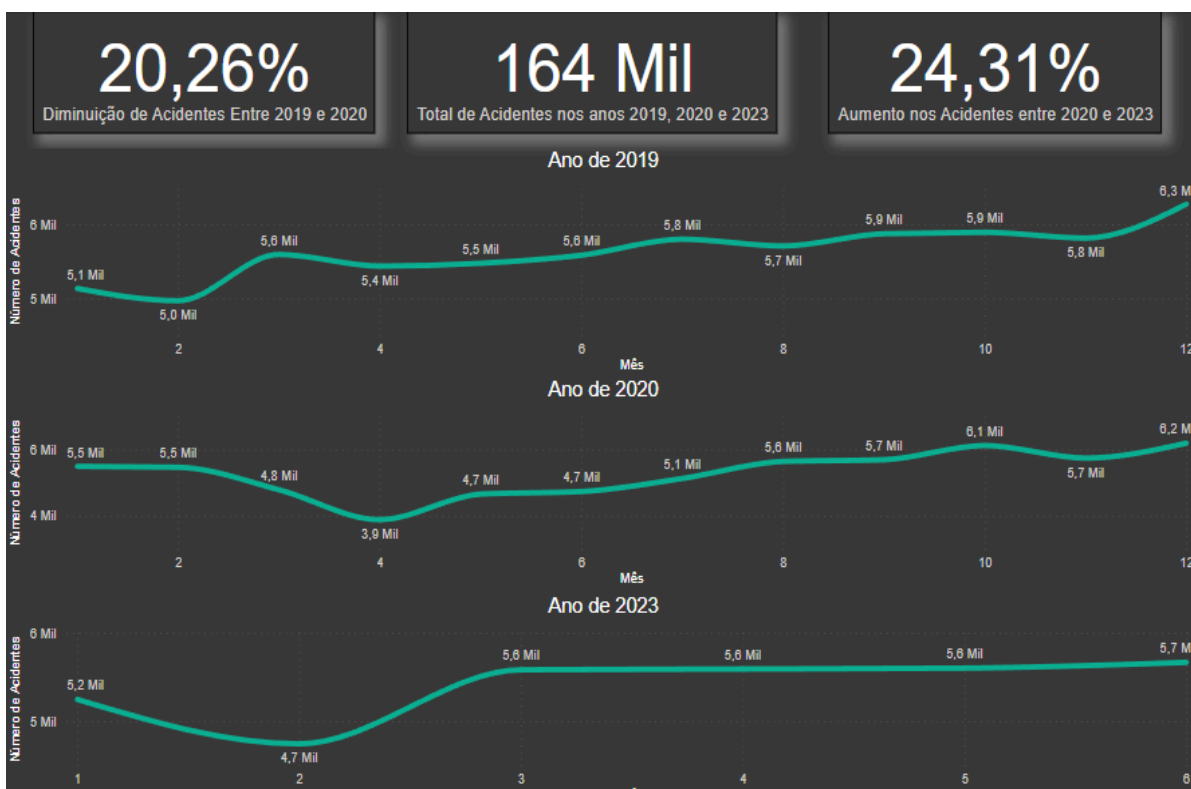
Figura 30 - Análise de BR, Pista e Uso Solo



**Fonte:** Autoria Própria (2024)

E como última análise foi realizada uma análise para identificação de redução do número de acidentes no período de *lockdown* total no Brasil, período que se reflete ao início da pandemia global de Covid-19. Foram comparados os anos de 2019, ano que precede a pandemia de Covid-19, 2020, ano que se inicia a pandemia global de Covid-19, e 2023, ano em que a Organização Mundial da Saúde (OMS) declarou o fim da pandemia. O período de *lockdown* total no Brasil se iniciou no período de Março/2020 e se estendeu até o fim de Julho/2020, após esse período houve o “relaxamento” das medidas de proteção contra o Covid-19.

Figura 31 - Análise Pré e Pós Pandemia



Fonte: Autoria própria (2024)

## 5 RESULTADOS

A primeira análise, análise de vítimas, condição meteorológica e causa dos acidentes, após a elaboração dos gráficos foi possível identificar que “Colisão Traseira” foi o tipo de acidente com maior número de ocorrências (aproximadamente 55 mil acidentes registrados).

Para explicar melhor essa análise, foi utilizado o gráfico de dispersão, onde é possível verificar que a causa com maior registro de ocorrências foi a “Falta de Atenção do Condutor”, registrando 47433 acidentes. Continuando a análise do gráfico de dispersão, a “Falta de Atenção do Condutor” também foi responsável por alavancar o número de vítimas (não fatais), quase 55 mil pessoas. Em uma análise feita pela Unidade Federativa (UF), foi identificado que Minas Gerais é o estado com maior registro de ocorrências de acidentes entre as 27 Unidades Federativas (incluindo o Distrito Federal). E por fim, a análise de árvore hierárquica mostrou que dos 24 mil registros de vítimas fatais em um período de 5 anos (2019-2023), aproximadamente 15 mil registros foram de acidentes em boas condições climáticas (Céu Claro), causado principalmente pela Falta de Atenção do Condutor (1644 registros) e Velocidade Incompatível (1568 registros).

O KPI “Taxa de Mortalidade dos Acidentes “ indica que 8,24% dos acidentes resultaram em vítimas fatais. Outro KPI, agora “Média de Acidentes por Estado”, indica que cada estado possui, em média, aproximadamente 11 mil acidentes por ano. O último KPI dessa análise, “Número Total de Acidentes”, indica o total de acidentes nas rodovias federais.

A segunda análise, análise de BR, Pista e Uso Solo, após a elaboração dos gráficos foi possível verificar que, visualizando de forma geral para todo o período da base de dados (2019-2023), as BRs 101 e 116, por serem as maiores do Brasil, apresentam o maior índice de registros de acidentes, com 49 mil e 46 mil registros aproximadamente. Além disso, as BRs 101 apresentaram os maiores índices de mortalidade, porém dessa vez com destaque para a BR 116, que apresentou 3062 registros de vítimas fatais, enquanto a BR 101 apresentou 2873 registros.

Na análise de tipo de pistas foi identificado que a pista simples possui maior índice de acidentes e de mortalidade, sendo 146470 registros de acidentes e 16838 registros de vítimas fatais, sendo a taxa de vítimas fatais em relação ao número de acidentes superior a 11%. Os acidentes em uso solo, acidentes em via asfaltada ou via rural, foram distribuídos da seguinte maneira: 163 mil para não uso solo e 129 mil para uso solo, ou seja, 163 mil acidentes foram registrados em vias rurais, enquanto 129 mil registros de acidentes foram em vias de asfalto.

O KPI dessa análise, “Número Total de Acidentes”, indica o total de acidentes nas rodovias federais. Outro KPI, “Média de Acidentes por BR”, indica o número médio de acidentes por BR. O último KPI, “Média de Vítimas por BR”, indica a média de vítimas fatais por BR.

A terceira análise, análise temporal, após a elaboração dos gráficos foi possível identificar que, olhando de forma geral para o período de 5 anos presente na base de dados (2019-2023), o ano que mais houve registro de acidentes nas rodovias federais do Brasil foi o ano de 2019, com aproximadamente 68 mil registros.

Os meses que mais ocorrem registros de acidentes são os meses de Janeiro e Maio, com Maio tendo o maior número de registros entre todos os meses. Seguindo para as demais análises, os dados apontam que os dias da semana que mais ocorrem registros de acidentes

são Sexta-Feira, Sábado e Domingo, ou seja, os acidentes ocorrem mais aos fins de semana, com destaque para Sábado.

Os horários com picos de registros de acidentes são os horários a partir das 17:00, período em que começa a transição da tarde para a noite. 2019 foi o ano com maior número de ocorrências de acidentes entre os 5 anos da base de dados (2019-2023). O KPI “Taxa de Crescimento Anual” indica o crescimento dos acidentes ao longo do tempo. Já o KPI “Número Máximo de Acidentes em um Mês”, indica o mês “campeão” de acidentes no Brasil, que foi o mês de Dezembro/2019. E o KPI “Dia com Maior Média de Acidentes” é o dia que tem a maior média de acidentes, que é o dia de domingo.

Olhando para os 3 anos em questão e com o período de lockdown total em evidência, 2020 com relação à 2019, ano que precede a pandemia, houve uma redução de pouco mais de 20% no número de acidentes. Já comparando o mesmo período entre os anos de 2020 e 2023, ano que foi declarado o fim da pandemia, houve um aumento de pouco mais de 24% no número de acidentes, em números totais, um aumento de aproximadamente 4400 acidentes.

O KPI “Diminuição de Acidentes Entre 2019 e 2020” indica a porcentagem de diminuição do número de acidentes no período de *lockdown* total no Brasil. Já o KPI “Total de Acidentes nos Anos de 2019, 2020 e 2023” indica o número total de acidentes nos anos de comparação (2019, 2020 e 2023). E o KPI “Aumento de Acidentes entre 2020 e 2023” indica a taxa de aumento do número de acidentes pós pandemia no Brasil, analisando o mesmo período do *lockdown* total.

## **6 CONSIDERAÇÕES FINAIS**

### **6.1 CONCLUSÃO**

Este trabalho teve como objetivo apresentar as etapas do processo de desenvolvimento de Business Intelligence focadas nos acidentes de trânsito em rodovias federais no Brasil. Ao todo foram criados 4 painéis (dashboards), dividido em quatro análises diferentes: Análise de Vítimas, Condições Meteorológicas e Fatores dos Acidentes, Análise Temporal, Análise de BR, Pista e Via Pavimentada, e por último Análise Pré e Pós Pandemia.

Neste trabalho buscamos compreender o fenômeno dos acidentes em rodovias e apresentar uma ferramenta para conscientizar o condutor e auxiliar no processo de tomada de decisão em relação às rotas mais problemáticas e aos principais fatores que influenciam na ocorrência de acidentes, demonstrando fatores de risco correlacionados.

### **6.1.1 ANÁLISE DE VÍTIMAS, CONDIÇÃO METEOROLÓGICA E CAUSAS DE ACIDENTE**

Foi possível observar que a falta de atenção do condutor causa um número elevado de acidentes, gerando inúmeras vítimas, incluindo vítimas com ferimentos graves e vítimas fatais. De modo geral, a condição meteorológica não influencia diretamente no aumento do número dos acidentes, conseqüentemente no número de vítimas fatais.

Os estados que mais apresentaram registros de acidentes foram: 1º Minas Gerais, 2º Santa Catarina, 3º Paraná, 4º Rio Grande do Sul e 5º Rio de Janeiro. Colisão traseira foi o tipo de acidente mais comum. Os KPIs indicam que nos 5 anos selecionados para estudos houve aproximadamente 293 mil acidentes, com pouco mais de 8% de taxa de mortalidade nos acidentes e 10,840 em média por estado.

### **6.1.2 ANÁLISE TEMPORAL**

Na análise temporal foi identificado que o mês com maior registro de acidentes foi o mês de Maio, com cerca de 26,7 mil registros de acidentes, com o *KPI* “Número Máximo de Acidentes em um Mês”, indicando o registro máximo de acidentes em um mês, mês esse que foi dezembro de 2019. Os dias com maior registro de acidentes foram os dias de final de semana: sexta, sábado e domingo, com o *KPI* “Dia com Maior Média de Acidentes” indicando Domingo como dia com maior média de acidentes ao longo de 5 anos.

Os horários com maior ocorrência de acidentes são principalmente os horários a partir das 17:00, período que começa a fase de anoitecer. O Ano “campeão” de acidentes foi o ano de 2019, com aproximadamente 68 mil registros de acidentes. O *KPI* “Taxa de Crescimento Anual de Acidentes” mostra a evolução dos acidentes analisando ano a ano.

### **6.1.3 ANÁLISE DE BR, PISTA E VIA PAVIMENTADA**

Na análise de BR, Pista e Uso Solo mostra que as BRs 101 e 116, por serem as maiores em questão de extensão territorial, apresentam os maiores índices de acidentes, com cerca de 49 mil e 44 mil registros, respectivamente. Apesar da BR 116 apresentar menos

ocorrências de acidentes em relação a BR 101, a BR 116 apresentou um maior número de vítimas fatais, 3062 vítimas fatais na BR 116 e 2873 na BR 101. O *KPI* “Média de Acidentes por BR” indica o número médio de acidentes em cada BR.

O tipo de pista simples apresenta um número maior de acidentes com relação aos outros tipos, com aproximadamente 147 mil registros de acidentes, representando 50,05% dos registros, 16838 vítimas fatais e 11,50% de taxa de mortalidade. O *KPI* “Número de Acidentes” indica o total de acidentes registrados. As vias rurais carregam os maiores números de acidentes, com aproximadamente 163 mil registros, além de elevar o número de vítimas graves, como feridos graves e vítimas fatais.

#### **6.1.4 ANÁLISE PRÉ E PÓS PANDEMIA**

Fazendo uma comparação entre os anos pré e pós pandemia (2019, 2020 e 2023), no período de *lockdown* total no Brasil (Março/2020-Julho/2020) houve uma diminuição significativa no número de acidentes, em relação aos anos de 2019 e 2020 verifica-se uma diminuição no número de acidentes de forma geral de pouco mais de 20%, como indica o *KPI* (Diminuição de Acidentes Entre 2019 e 2020). Porém, comparando o mesmo período (Março-Julho) de 2020 com 2023, houve um aumento no número de acidentes de mais de 24%, como indica o *KPI* “Aumento nos Acidentes entre 2020 e 2023.

Após a conclusão do estudo, considero um recurso interessante que pode ser disponibilizado à sociedade, bem como utilizado em instituições de forma estratégica.

#### **6.2 LIMITAÇÕES DA PESQUISA**

A Polícia Rodoviária Federal (PRF) não disponibiliza os registros do ano de 2023 de forma completa, apenas do dia 01 de janeiro de 2023 até 30 de junho de 2023, dificultando uma análise mais geral dos acidentes no ano em questão.

A PRF também não disponibiliza informações que poderiam aprofundar ainda mais as análises, como tipo de veículo, modelo do veículo, ano de fabricação, faixa etária do condutor, sexo do condutor, entre outros campos para melhores análises.

Além disso, não é possível identificar o fluxo de veículo em determinada rodovia e trecho de rodovia. Essas informações não estão presentes em nenhuma base de dados da Polícia Rodoviária Federal (PRF).

### 6.3 TRABALHOS FUTUROS

A proposta de trabalhos futuros é fundamentada no aperfeiçoamento da solução desenvolvida por meio da implementação de novas funcionalidades, fontes de dados e aprimoramento das análises.

Para trabalhos futuros o dashboard pode ser incrementado cruzando dados do DataSus, a fim de aumentar a acurácia dos dados disponibilizados, incluir um *KPI* acidentes/km de rodovia, visto que, algumas BRs têm quilometragem bem distinta uma da outra, assim como, um kpi levando em consideração o número de veículo na rodovia (fluxo de veículos), entre outras informações.

### REFERÊNCIAS BIBLIOGRÁFICAS

DISTRITO FEDERAL. CNM. **Análise do impacto da frota de veículos nos Municípios brasileiros**. Brasília, 2023. Disponível em: [https://cnm.org.br/storage/biblioteca/2023/Estudos\\_tecnicos/202309\\_ET\\_MOB\\_Impacto\\_frota\\_veiculos.pdf](https://cnm.org.br/storage/biblioteca/2023/Estudos_tecnicos/202309_ET_MOB_Impacto_frota_veiculos.pdf). Acesso em: 29 fev. 2024.

ROCHA, G. S. **Caracterização dos acidentes de trânsito e vítimas no município de Rio Branco-Acre**. 2009.

CARVALHO, C. H. R.; GUEDES, E. P. **Balanço da 1ª década de ação pela segurança no trânsito no Brasil e perspectivas para a 2ª década**. Brasília: Ipea, 2023.

SANTOS, M. R. **Aplicação de Business Intelligence para Análise de Indicadores das Redes de Referência no Estado do Paraná**. Universidade Estadual de Londrina, 2009.

AUDZEYEVA, A.; HUDSON, R. **How to get the most from a business intelligence application during the post implementation phase? Deep structure transformation at a U.K. retail bank**. *European journal of information systems: an official journal of the Operational Research Society*, v. 25, n. 1, p. 29–46, 2016.

KITCHENHAM, B; CHARTERS, S. M. **Guidelines for performing Systematic Literature Reviews in Software Engineering**. Keele: Keele University, 2007. Disponível em: [https://legacyfileshare.elsevier.com/promis\\_misc/525444systematicreviewsguide.pdf](https://legacyfileshare.elsevier.com/promis_misc/525444systematicreviewsguide.pdf)

SCHOLZ, J. C. W; PINHEIRO, Y. P. **Preveno a gravidade de acidentes rodoviários no Brasil: a influência do ambiente e características do veículo**. Paraná: RIUT, 2023.

LEGRAMANTI, G. *et al.* **Utilização de sistema de informações geográficas na tomada de decisões do sistema viário com base em dados de acidentes de trânsito no município de Pato Branco**. Pato Branco: RIUT, 2021.

D'AZEVEDO, Z. S. **Elaboração de modelo da previsão da velocidade operacional de rodovias em pistas simples considerando os veículos de carga.** Porto Alegre. 2020.

**Registro Nacional de Acidentes e Estatísticas de Trânsito. Ministério dos Transportes.**

Disponível em:

<https://www.gov.br/transportes/pt-br/assuntos/transito/arquivos-senatran/docs/renaest>

Acesso em 12 de fev. 2024.

MURRAY, C. **Oracle SQL Developer User's Guide.** Austin: Oracle, 2014.

MACHADO, F. N. R. **Projeto de Data Warehouse: Uma visão multidimensional.** São Paulo: Érica, 2000.

DETONI, W. **BUSINESS INTELLIGENCE: UMA REVISÃO DE LITERATURA.**

Revista Científica Semana Acadêmica. Fortaleza, ano MMXVIII, Nº. 000132, 19/09/2018.

WANG, Y. **What are the biggest obstacles to growth of SMEs in developing countries? -**

An empirical evidence from an enterprise survey. 3. ed. Borsa: ScienceDirect, 2016.

167-176 p. v. 16.

NGUYEN, H; DURMAS, M; HOFSTEDE, A. H. M. T; LA ROSA, M.; MAGGI, F. M .

**Business Process Performance Mining with Staged Process Flows.** 28. ed. Switzerland:

Cham: Springer, 2016. 167-185 p. v. 9694.

KIMBALL, R.; ROSS, M. **The data warehouse toolkit: the definitive guide to dimensional modeling.** 3. ed. Indianapolis: Wiley, 2013. Disponível em:

<https://acervo-digital.espm.br/E-BOOKS/2020/382023.pdf>. Acesso em 12 de fev. 2024

Documentação Pentaho Data Integration and Analytics. **Hitachi Vantara**, 2023. Disponível

em: [https://www.hitachivantara.com/en-us/products/pentaho-platform/](https://www.hitachivantara.com/en-us/products/pentaho-platform/data-integration-analytics.html)

[data-integration-analytics.html](https://www.hitachivantara.com/en-us/products/pentaho-platform/data-integration-analytics.html). Acesso em 13 de fev. 2024.

INMON, W. H. **Building the Data Warehouse.** Nova Iorque: John wiley & sons, 2005.

Disponível

em: <https://fit.hcmute.edu.vn/Resources/Docs/SubDomain/fit/ThayTuan/DataWH/Bulding%20the%20Data%20Warehouse%204%20Edition.pdf>. Acesso em 13 de fev. 2024

STONEBRAKER, M. ; ROWE, L. A. **The Design of Postgres.** 2. ed. Berkeley: ACM SIGMOD Record, 1986. 340 - 355 p. v. 15.

FRAGA, A. **Microsoft Power BI: gráficos, banco de dados e configuração de relatórios.**

Rio de Janeiro: Alta Books, 2019.

Documentação Entenda o esquema em estrela e a importância para o Power BI. **Microsoft Learn**, 2024. Disponível em:

<https://learn.microsoft.com/pt-br/power-bi/guidance/star-schema>. Acesso em: 20 fev. 2024.

Documentação Modelagem de Dados com o Oracle SQL Developer. **Oracle**, 2024.

Disponível em: [www.oracle.com/br/database/sqldeveloper/technologies/sql-data-modeler](http://www.oracle.com/br/database/sqldeveloper/technologies/sql-data-modeler).

Acesso em: 14 fev. 2024.

BOLT, A; AALST, W. M. P. V. D. **Multidimensional process mining using process cubes.**

16. ed. Stockholm: Charm: Springer, 2015. 102 - 116 p. v. 214.

Documentação Pentaho Data Integration and Analytics. **Hitachi Vantara**, 2017. Disponível em: [https://help.pentaho.com/Documentation/7.1/0D0/Pentaho\\_Data\\_Integration](https://help.pentaho.com/Documentation/7.1/0D0/Pentaho_Data_Integration). Acesso em: 13 fev. 2024.

HATLE, D. G. N. G. S., Sayeed, A., Gupta, D., Neela, G. P., & Winter, S. U. S. A. N. **Pentaho data integration tool. Business Intelligence Tool**. Orlando: Pentaho, p. 2-18, 2013.

RIGON, A. J. *et al.* **O desenvolvimento psíquico e o uso do Microsoft Power BI como ferramenta de análise de dados**. Salvador: Revista Brasileira de Psicologia, v. 78, n. 2, p. 1–15, 2021.