



INSTITUTO FEDERAL DE ALAGOAS
CAMPUS MACEIÓ
CURSO BACHARELADO EM SISTEMAS DE INFORMAÇÃO

JOSÉ CARLOS GOMES FILHO
LEONARDO HENRIQUE DOS ANJOS SANTOS

**TRABALHO DE CONCLUSÃO DE CURSO DE FERRAMENTA QUE AUXILIA NA
ANÁLISE DOS ACIDENTES DE TRÂNSITO NAS RODOVIAS FEDERAIS
BRASILEIRAS POR MEIO DE ETL**

MACEIÓ, AL
2024

JOSÉ CARLOS GOMES FILHO
LEONARDO HENRIQUE DOS ANJOS SANTOS

TRABALHO DE CONCLUSÃO DE CURSO DE FERRAMENTA QUE AUXILIA NA
ANÁLISE DOS ACIDENTES DE TRÂNSITO NAS RODOVIAS FEDERAIS
BRASILEIRAS POR MEIO DE ETL

Especificação técnica apresentada ao Curso Bacharelado em Sistemas de Informação do Instituto Federal de Alagoas, *campus* Maceió, como requisito para a obtenção do grau de Bacharel em Sistemas de Informação.

Orientador(a): Prof. Dr. Edison Camilo de Moraes Júnior

MACEIÓ, AL

2024

FICHA CATALOGRÁFICA



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

005.74

G633e

Gomes Filho, José Carlos.

Especificação técnica de ferramenta que auxilia na análise dos acidentes de trânsito nas rodovias federais brasileiras por meio de ETL [recurso eletrônico] : relato de uma experiência de ensino-aprendizagem / José Carlos Gomes Filho, Leonardo Henrique dos Anjos Santos. – Dados eletrônicos (1 arquivo : 407 KB). – 2025.

Sistema requerido: Adobe Acrobat Reader.

Modo de acesso: Internet.

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, 2025.

1. Sistemas de Informação. 2. Rodovias federais – Acidentes. 3. Painel interativo. 4. ETL. 5. Interface de linha de comando I. Santos, Henrique dos Anjos. II. Título.

Franciane Monick Gomes de França
Bibliotecária – CRB 4/1831

JOSÉ CARLOS GOMES FILHO
LEONARDO HENRIQUE DOS ANJOS SANTOS

TRABALHO DE CONCLUSÃO DE CURSO DE FERRAMENTA QUE AUXILIA NA
ANÁLISE DOS ACIDENTES DE TRÂNSITO NAS RODOVIAS FEDERAIS
BRASILEIRAS POR MEIO DE ETL

Especificação técnica apresentada ao Curso Bacharelado em Sistemas de Informação do Instituto Federal de Alagoas, *campus* Maceió, como requisito para a obtenção do grau de Bacharel em Sistemas de Informação.

Aprovado em: 22/04/2024

BANCA EXAMINADORA

Prof. Dr. Edison Camilo de Moraes Júnior
Instituto Federal de Alagoas – IFAL

Prof. Jailton Cardoso da Cruz
Instituto Federal de Alagoas – IFAL

Prof. Flavio Mota Medeiros
Instituto Federal de Alagoas – IFAL

RESUMO

O trabalho apresenta uma abordagem técnica para lidar com a problemática dos acidentes de trânsito no Brasil, propondo o desenvolvimento de uma interface de linha de comando (CLI) em Python para a extração, transformação e carga (ETL) de dados de acidentes em rodovias federais. A arquitetura de Business Intelligence (BI) empregada no trabalho foi o esquema estrela na modelagem de dados. A CLI permite uma execução customizada do processo de ETL, incluindo seleção de conjuntos de dados e dimensões temporais. Os dados processados podem ser utilizados para alimentar um Dashboard interativo, construído com o PowerBI, oferecendo visualizações detalhadas dos dados e filtros por tipo de via e condições meteorológicas. O artigo visa fornecer uma solução técnica robusta para análise e prevenção de acidentes de trânsito, com potencial para embasar políticas públicas mais eficazes neste campo.

Palavras-chaves: acidentes em rodovias federais; painel interativo; ETL; interface de linha de comando

ABSTRACT

This work presents a technical approach to address the issue of traffic accidents in Brazil, proposing the development of a command-line interface (CLI) in Python for the extraction, transformation, and loading (ETL) of data from federal highway accidents. Business Intelligence (BI) architecture is employed, with emphasis on the star schema in data modeling. The CLI allows for customized execution of the ETL process, including selection of data sets and temporal dimensions. Processed data can be used to feed an interactive Dashboard, built with PowerBI, offering detailed data visualizations and filters by road type and weather conditions. The article aims to provide a robust technical solution for the analysis and prevention of traffic accidents, with the potential to support more effective public policies in this field.

Keywords: accidents on federal highways; interactive panel; ETL; command line interface.

LISTA DE ILUSTRAÇÕES

Figura 1 - Etapas da Arquitetura de Business Intelligence	8
Figura 2 - Modelagem do esquema em estrela	15
Figura 3 - Opções disponíveis na ferramenta	16
Quadro 1 - Descrição dos parâmetros da CLI	18
Figura 4 - Resultados da execução com dados referente ao ano de 2020	19
Figura 5 - Resultados da execução com dados referente ao ano de 2007	19
Figura 6 - Resultados da execução para gerar a dimensão tempo	20
Figura 7 - Resultados da execução com dados referente ao ano de 2008	20
Figura 8 - Resultados da execução com dados referente ao ano de 2008 e exportação. . .	21
Figura 9 - Visão geral do Dashboard	21
Figura 10 - Gráfico de número de veículos envolvidos em acidentes por mês	22
Figura 11 - Gráfico de número de pessoas envolvidas em acidentes por mês	23
Figura 12 - Gráfico de número de vítimas fatais por estado	24
Figura 13 - Gráfico de número de vítimas fatais por dia da semana	25
Figura 14 - Cartões de ilustração sobre o total de pessoas e veículos envolvidos	25
Figura 15 - Caixas de filtro de acidentes	26

LISTA DE ABREVIATURAS OU SIGLAS

ETL Extract Transform Load

CLI Command Line Interface

SUMÁRIO

1. INTRODUÇÃO.....	7
1.1. A ARQUITETURA DE BUSINESS INTELLIGENCE (BI).....	8
1.2. INTERFACE DE LINHA DE COMANDO.....	9
1.3 O PROCESSO DE EXTRAÇÃO, TRANSFORMAÇÃO E CARGA DE DADOS....	10
2. DESENVOLVIMENTO.....	11
2.2. BASE DE DADOS.....	11
2.3. PROCESSO DE ETL.....	12
2.3.1. Limpeza de Dados.....	12
2.3.2. Modelagem de Dados.....	13
2.4. INTERFACE DE LINHA DE COMANDO.....	14
2.5. DASHBOARD INTERATIVO.....	19
3. CONCLUSÃO.....	25
ANEXOS.....	27

1. INTRODUÇÃO

Os acidentes de trânsito representam uma trágica e evitável causa de morte, ceifando a vida de mais de 1,35 milhão de pessoas anualmente em todo o mundo. Este cenário é especialmente prevalente em países de renda média, onde os índices de motorização estão em ascensão. Surpreendentemente, embora ocupem a 8ª posição no ranking de causas de óbito para todas as faixas etárias, seu impacto equivale ao de diversas doenças transmissíveis, como a malária. De forma ainda mais preocupante, os acidentes de trânsito se configuram como a principal razão de óbito para indivíduos entre 5 e 29 anos, destacando a necessidade premente de medidas eficazes para reverter essa triste realidade [Conceição et al. (2021)].

Em 2018 no Brasil foram registrados 69.206 acidentes, 58.885 feridos leves, 17.640 feridos graves e 5.269 óbitos no local do acidente. (POR VIAS SEGURAS, 2018). Em 2022, apresentou registros de cerca de 5.439 mortes em acidentes em rodovias federais. Muitos dos acidentes são evitáveis e passíveis de prevenção, o que torna ainda mais importante a realização de estudos e execução de ações que possam mitigá-los. [Rizzi et al. (2016)].

Desde 2009, o governo brasileiro tem implementado iniciativas voltadas para a promoção de políticas de disponibilização de dados governamentais visando seu livre acesso pela sociedade. Este movimento se alinha ao crescente impulso global em prol da democratização do acesso a informações e dados abertos. Em 2011, o Brasil promulgou a Lei de Acesso à Informação (12.527/2011) com o intuito de regulamentar o direito constitucional dos cidadãos ao acesso a informações públicas [Carvalho (2021)].

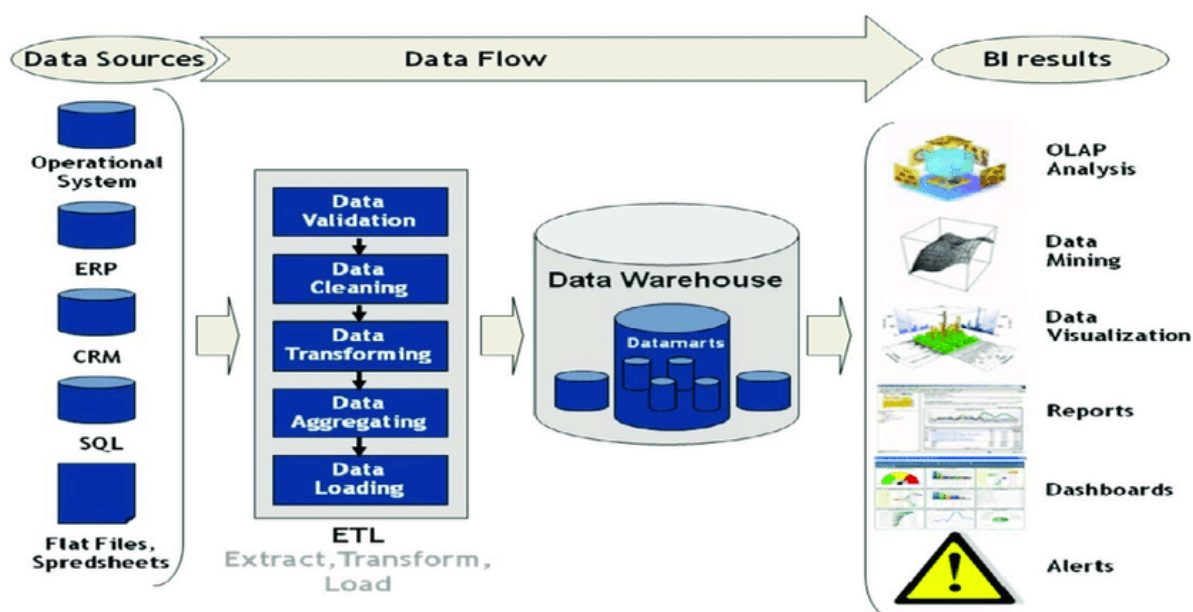
Devido a este movimento, os dados referentes a registros de acidentes em rodovias federais brasileiras são disponibilizados publicamente pela Polícia Rodoviária Federal - PRF anualmente e podem ser utilizados para gerar ferramentas de análise com o objetivo de entender como minimizar a taxa de acidentes em território nacional.

Com base nisso, o presente trabalho descreve o desenvolvimento técnico de uma interface de linha de comando (CLI) que viabiliza o processo de Extração, Transformação e Carregamento (ETL) com dados de acidentes em rodovias federais brasileiras para gerar insumo e alimentar um *Dashboard* interativo capaz de gerar análises sobre tais acidentes de maneira rápida e simples.

1.1. A ARQUITETURA DE BUSINESS INTELLIGENCE (BI)

A arquitetura de **Business Intelligence (BI)** é fundamental para organizar e gerenciar os dados de uma empresa, fornecendo uma estrutura robusta para a análise e tomada de decisões. Essa arquitetura compreende diversos componentes essenciais que trabalham em conjunto para transformar dados brutos em informações úteis e acionáveis [Guimarães, 2014].

Figura 1 - Etapas da Arquitetura de Business Intelligence



Fonte: https://www.researchgate.net/figure/Business-intelligence-system-BIS-architecture-with-key-components-and-functions_fig2_344444910

O **Esquema Estrela**, uma estrutura de organização de dados, destaca-se como uma abordagem eficaz para otimizar o desempenho de sistemas. Sua vantagem primordial reside na capacidade de escalabilidade, tornando-o apto a suportar uma ampla gama de usuários [Tavares, (2022)]. Além disso, sua flexibilidade é notável, pois permite a incorporação contínua de novos fatos e dimensões conforme o projeto evolui. Essa adaptabilidade garante que o esquema estrela permaneça relevante e eficiente ao longo do desenvolvimento do projeto, fornecendo uma base sólida para análises robustas e *insights* valiosos [Tavares, (2022)].

Além disso, a criação de **dimensões** e regras é essencial para garantir que as informações estejam estruturadas de forma adequada e em conformidade com as necessidades da empresa. As dimensões fornecem uma estrutura para organizar os dados, enquanto as regras garantem que as análises sejam feitas de acordo com as diretrizes estabelecidas.

O processo de Business Intelligence (BI) se desdobra em cinco etapas principais:

- **ETL (Extração, Transformação e Carregamento):** Nesta fase, os dados são coletados de várias fontes, transformados e armazenados centralmente para análise.
- **Modelagem de Dados:** Aqui, os dados passam por processos de transformação e relacionamento, utilizando ferramentas como o Power Query, para permitir uma compreensão mais profunda e encontrar padrões
- **Cálculos e Análises:** Os dados estruturados e tratados são utilizados para realizar cálculos e análises, visando obter insights valiosos para o negócio.
- **Relatórios:** Os resultados das análises são apresentados em relatórios, geralmente através de dashboards e visualizações de dados, para facilitar a compreensão e o compartilhamento com outros usuários.
- **Compartilhamento Online dos Relatórios:** Por fim, os relatórios podem ser compartilhados online, permitindo que várias pessoas tenham acesso aos insights e tomem decisões baseadas nos dados de forma colaborativa e em tempo real.

Essas etapas formam um ciclo contínuo, onde os dados são constantemente atualizados e refinados para fornecer informações relevantes e atualizadas para o negócio.

1.2. INTERFACE DE LINHA DE COMANDO

A Interface de linha de comando, também conhecido pelo termo CLI, oriundo da língua inglesa, é um programa que permite que os usuários digitem comandos de texto dando instruções a um computador para fazer funções específicas.

CLIs são comumente utilizadas por desenvolvedores de software ou usuários que desejam executar ações sem a necessidade de utilização de uma interface gráfica padrão.

1.3. O PROCESSO DE EXTRAÇÃO, TRANSFORMAÇÃO E CARGA DE DADOS

O processo de extração, transformação e carga de dados, também conhecido pela sigla ETL, é o processo responsável pela coleta, limpeza, organização e carregamento de dados de diversas fontes com o objetivo de dar suporte à geração de relatórios, análise e tomada de decisões.

2. DESENVOLVIMENTO

2.1. REGISTRO DE SOFTWARE

Esta ferramenta passou pelo processo de Registro de Programa de Computador do Instituto Nacional de Propriedade Industrial (INPI).¹

2.2. BASE DE DADOS

A base de dados utilizada, intitulada Acidentes rodovias federais Brasil Jan/07 a Dez/20, foi encontrada na ferramenta de ciência de dados Kaggle². Os dados na mesma são referentes ao período entre Janeiro de 2007 e Dezembro de 2020. Originalmente, o conjunto de dados possui 26 colunas, descritas no quadro 1.

¹ Disponível em:

https://drive.google.com/file/d/152BKtw0qyS2RvkMuU87sxbQJ-XxkrKeO/view?usp=drive_link.

² Disponível em:

<https://www.kaggle.com/datasets/equeiroz/acidentes-rodovias-federais-brasil-jan07-a-jul19>

2.3. PROCESSO DE ETL

O processo de ETL utilizou dados disponibilizados pela PRF, referente a acidentes em rodovias federais brasileiras referente aos anos de 2007 a 2020. Tais dados ilustram tanto informações gerais a respeito dos acidentes, como a quantidade de veículos envolvidos, pessoas envolvidas e vítimas fatais, como também informações detalhadas a respeito da condição meteorológica, tipo de pista, além de características da via.

2.3.1. Limpeza de Dados

A limpeza de dados serviu para remover todas as colunas que não foram utilizadas no projeto e normalizar todos os registros de acidentes que possuem dados nulos, genericamente definidos e duplicações. As colunas *uop*, *delegacia*, *latitude*, *regional*, *uso_solo* e *horario* foram removidas, enquanto também houve renomeações das colunas *data_inversa* para *data*, *tipo_acidente* para *tipo*, *classificacao_acidente* para *classificacao*.

As seguintes colunas, relevantes para o escopo do projeto, foram devidamente normalizadas para remover excesso de espaço em branco:

- dia_semana
- fase_dia
- municipio
- uf
- tipo
- classificacao
- causa_acidente

A coluna *causa_acidente* também sofreu normalizações para substituir valores nulos pelo valor “Causa não registrada” e para definir um valor padrão para os acidentes em que a causa foi “Ingestão de álcool e/ou substâncias psicoativas pelo pedestre” e “Ingestão de álcool e/ou substâncias psicoativas pelo condutor” pois haviam registros onde a mesma causa estava registrada com palavras diferentes.

A execução padrão da ferramenta utiliza registros referentes ao ano de 2020. Para esses dados, após a limpeza, os 63530 registros foram categorizados em 55 causas únicas, 100 municípios e 10 condições meteorológicas.

2.3.2. Modelagem de Dados

A estratégia adotada para a modelagem de dados utilizada no processo de ETL se deu seguindo a abordagem do esquema em estrela. Nessa abordagem, o banco de dados é estruturado de forma que as tabelas que armazenam os dados são classificadas como dimensão ou fato.

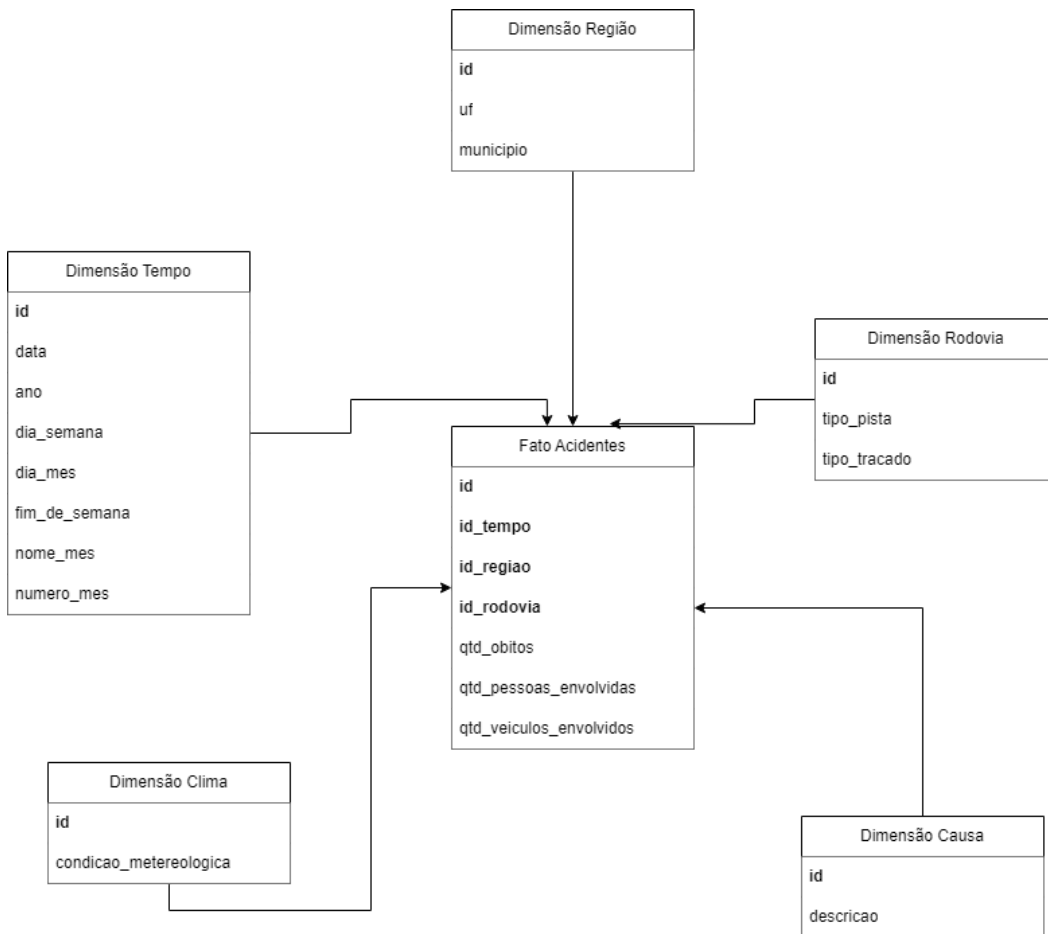
Tabelas de dimensões armazenam as entidades do negócio, isto é, os itens que você modela. Essas podem incluir produtos, pessoas, regiões e o próprio tempo. Tabelas de fatos, por sua vez, se relacionam com as tabelas de dimensões e armazenam os fatos do seu negócio, métricas e valores fundamentais para gerar informações a partir dos dados.

No cenário do presente trabalho, a estrutura de banco de dados apresenta as seguintes tabelas de dimensões: a) Região; b) Rodovia; c) Causa; d) Clima; e) Tempo; além da tabela de fatos.

- a) armazena dados da região em que o acidente ocorreu por meio de duas informações fundamentais: município e estado (UF).
- b) armazena dados da rodovia em que o acidente ocorreu, tais como tipo da pista (simples, dupla, etc) e tipo do traçado da via (curva, reta, etc).
- c) armazena dados da causa do acidente por meio de uma breve descrição textual.
- d) armazena dados a respeito do clima no momento do acidente através de uma breve descrição da condição meteorológica.
- e) armazena dados temporais como data, ano, número que representa o mês, nome que representa o mês, dia da semana e se o determinado dia é final de semana ou não.

Como ilustra a figura abaixo, todas as tabelas de dimensões são associadas com a tabela de fatos. A tabela de fatos, por sua vez, armazena métricas a respeito da quantidade de veículos envolvidos em acidentes, pessoas envolvidas e vítimas fatais. Essas métricas ilustram o somatório de todos os registros de acidentes do conjunto de dados.

Figura 2 - Modelagem do esquema em estrela



Fonte: Elaboração autoral

2.4. INTERFACE DE LINHA DE COMANDO

A CLI foi construída com a linguagem de programação *Python - versão 3.10.11* e bibliotecas como *Pandas - versão 1.5.2*, *mysql-connector-python - versão 8.1.0*, *SQLAlchemy - versão 2.0.21*, para a realização de tratamento e organização de dados. A interface interage com dois bancos de dados: a) *SQLite - versão oriunda na versão 3.10.11 do Python*, b) *MYSQL - versão 8.1.0 disponível no Docker*. O primeiro é utilizado durante a execução da CLI para a realização das operações necessárias e para armazenar os resultados do processo

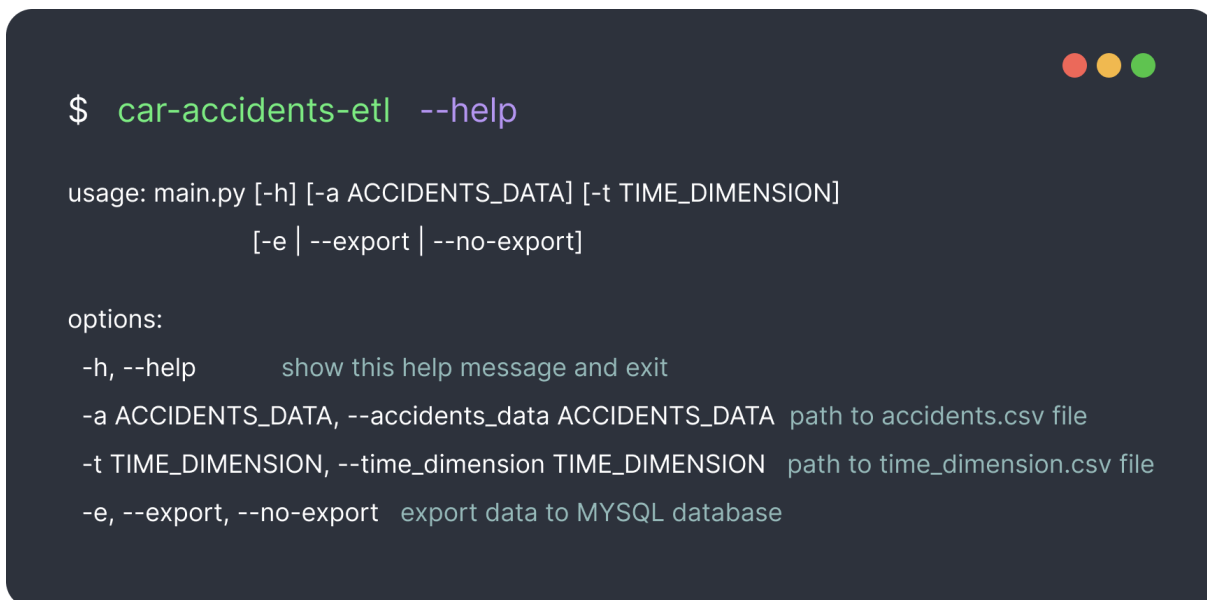
de ETL momentaneamente, enquanto o segundo é utilizado para armazenar os resultados do processo de ETL de forma definitiva e para servir o Dashboard interativo.

A interface de linha de comando disponibiliza as seguintes funcionalidades:

- 1) Realização do processo de ETL com dados referentes a um período específico
- 2) Exportação de insumo para o Dashboard
- 3) Geração de dados referentes à dimensão tempo

A instrução de uso da CLI está disponível por meio do comando de ajuda da ferramenta, onde é apresentado todos os possíveis parâmetros e funcionalidades. Para exibir tal instrução é necessário executar o seguinte comando:

Figura 3 - Opções disponíveis na ferramenta

A terminal window with a dark background and three colored window control buttons (red, yellow, green) in the top right corner. The prompt '\$' is followed by the command 'car-accidents-etl --help'. The output shows the usage and options for the tool.

```
$ car-accidents-etl --help

usage: main.py [-h] [-a ACCIDENTS_DATA] [-t TIME_DIMENSION]
              [-e | --export | --no-export]

options:
  -h, --help            show this help message and exit
  -a ACCIDENTS_DATA, --accidents_data ACCIDENTS_DATA path to accidents.csv file
  -t TIME_DIMENSION, --time_dimension TIME_DIMENSION path to time_dimension.csv file
  -e, --export, --no-export export data to MYSQL database
```

Fonte: Elaboração autoral

Os parâmetros disponíveis na ferramenta de linha de comando estão diretamente relacionados com a funcionalidade principal, isto é, a execução do processo de ETL. De forma mais descritiva, há um parâmetro que determina o conjunto de dados que será utilizado no processo, outro que determina os dados da dimensão tempo e outro que determina se o resultado do processo de ETL vai ser exportado para o painel interativo ou não, tendo como comportamento padrão a não exportação dos dados. Todos esses parâmetros possuem uma versão estendida e uma versão curta, oferecendo ao usuário o poder de escolha de sintaxe.

Como detalha o Quadro 1, há ainda o parâmetro de ajuda padrão, responsável por retornar um resumo de todos os parâmetros existentes.

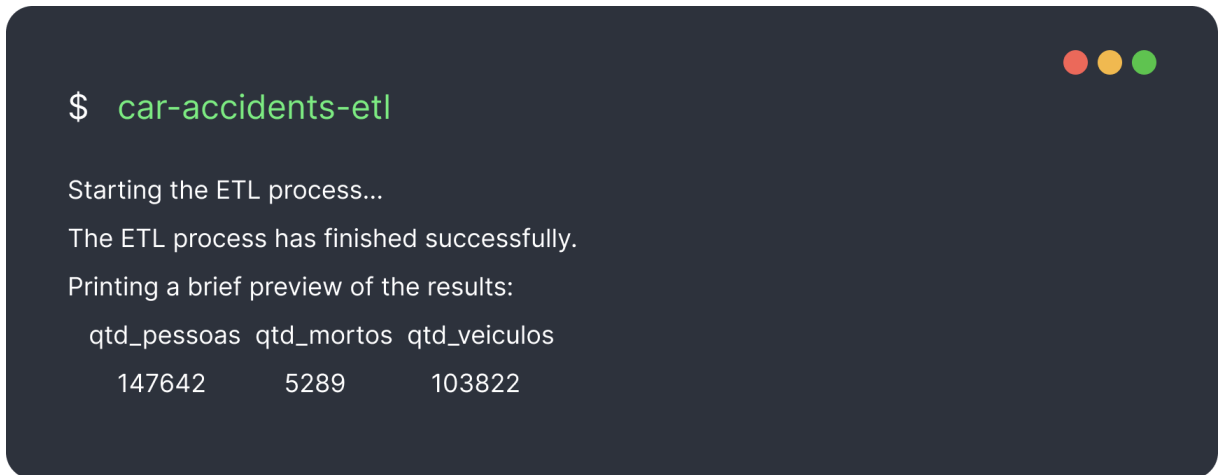
Quadro 1 - Descrição dos parâmetros da CLI

Parâmetro principal	Parâmetro alternativo	Descrição
-h	--help	Imprime o guia de uso da CLI
-a <Caminho do arquivo>	--accidents_data <Diretório>	Informa o caminho do arquivo que contém o conjunto de dados de acidentes
-t <Caminho do arquivo>	--time_dimension <Caminho do arquivo>	Informa o caminho do arquivo que contém a dimensão tempo que será utilizada
-e	--export	Determina que os resultados devem ser exportados para o banco de dados MYSQL
Não determinado	--no-export	Determina que os resultados não devem ser exportados para o banco de dados MYSQL. Esse é o comportamento padrão da ferramenta, caso o parâmetro anterior não tenha sido informado.

Fonte: Elaboração autoral

Para utilizar a ferramenta da maneira mais simples, é possível utilizar o comando ilustrado na Figura 4, onde somente é necessário informar o nome da interface de linha de comando. Dessa forma, será utilizado o conjunto de dados padrão (referente aos dados de acidentes que ocorreram em 2020) e a dimensão tempo padrão, que envolve o intervalo desde 01 de janeiro de 2006 a 01 de janeiro de 2022. O resultado desta execução não será exportado para o painel interativo.

Figura 4 - Resultados da execução com dados referente ao ano de 2020 (execução padrão)



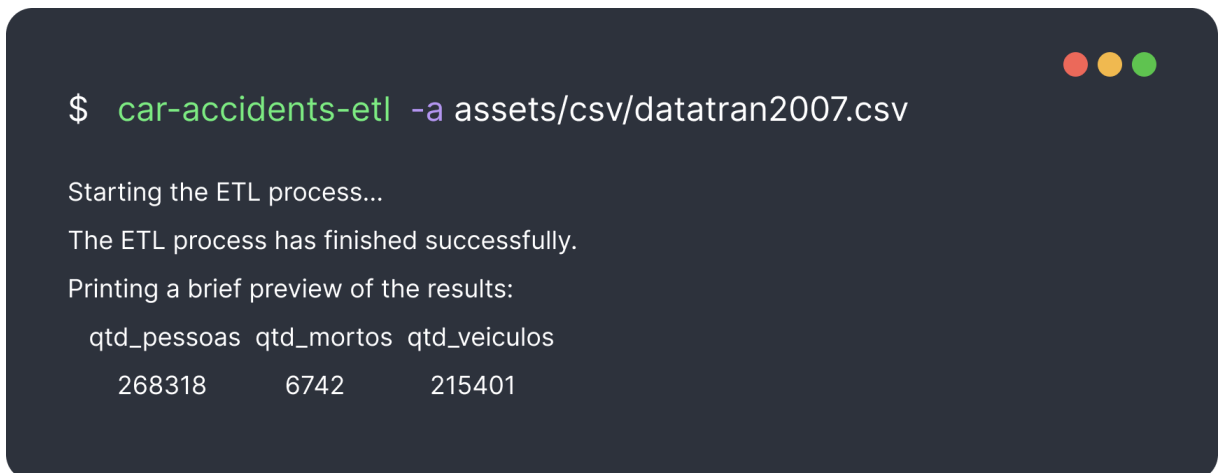
```
$ car-accidents-etl

Starting the ETL process...
The ETL process has finished successfully.
Printing a brief preview of the results:
qtd_pessoas qtd_mortos qtd_veiculos
147642      5289      103822
```

Fonte: Elaboração autoral

Para termos uma execução customizada, o parâmetro que determina o conjunto de dados que irá ser utilizado na execução pode ser informado como ilustra a Figura 5. Entretanto, o resultado dessa execução não será exportado para o painel interativo, já que o parâmetro de exportação não foi informado.

Figura 5 - Resultados da execução com dados referente ao ano de 2007



```
$ car-accidents-etl -a assets/csv/datatran2007.csv

Starting the ETL process...
The ETL process has finished successfully.
Printing a brief preview of the results:
qtd_pessoas qtd_mortos qtd_veiculos
268318      6742      215401
```

Fonte: Elaboração autoral

Os dados da dimensão tempo podem ser gerados através do seguinte comando, ilustrado na figura abaixo, onde é necessário informar a data inicial e a data final. Após a execução, um arquivo CSV será salvo com a dimensão tempo gerada e a mesma já pode ser utilizada na próxima execução.

Figura 6 - Resultados da execução para gerar a dimensão tempo

```

$ car-accidents-etl generate-time-dimension -s 2020-01-01
-e 2024-01-01

Starting generating time dimension with start date: 2020-01-01
and end date: 2024-01-01

Generating time dimension...

Converting time dimension to csv...

Done! The CSV file for your custom time dimension was saved at
output/2020-01-01-2024-01-01-time_dimension.csv

```

Fonte: Elaboração autoral

Caso o parâmetro que determina os dados da dimensão tempo não seja informado, a ferramenta irá executar o processo considerando a dimensão tempo padrão, porém caso haja interesse de execução levando em consideração um intervalo diferente do padrão, esse parâmetro pode ser informado como ilustra a Figura 7, desde que a nova dimensão tempo esteja no mesmo formato que a padrão.

Figura 7 - Resultados da execução com dados referente ao ano de 2008

```

$ car-accidents-etl -a assets/csv/datatran2008.csv
-t assets/csv/time_dimension.csv

Starting the ETL process...

The ETL process has finished successfully.

Printing a brief preview of the results:

qtd_pessoas qtd_mortos qtd_veiculos
298804      6914      242719

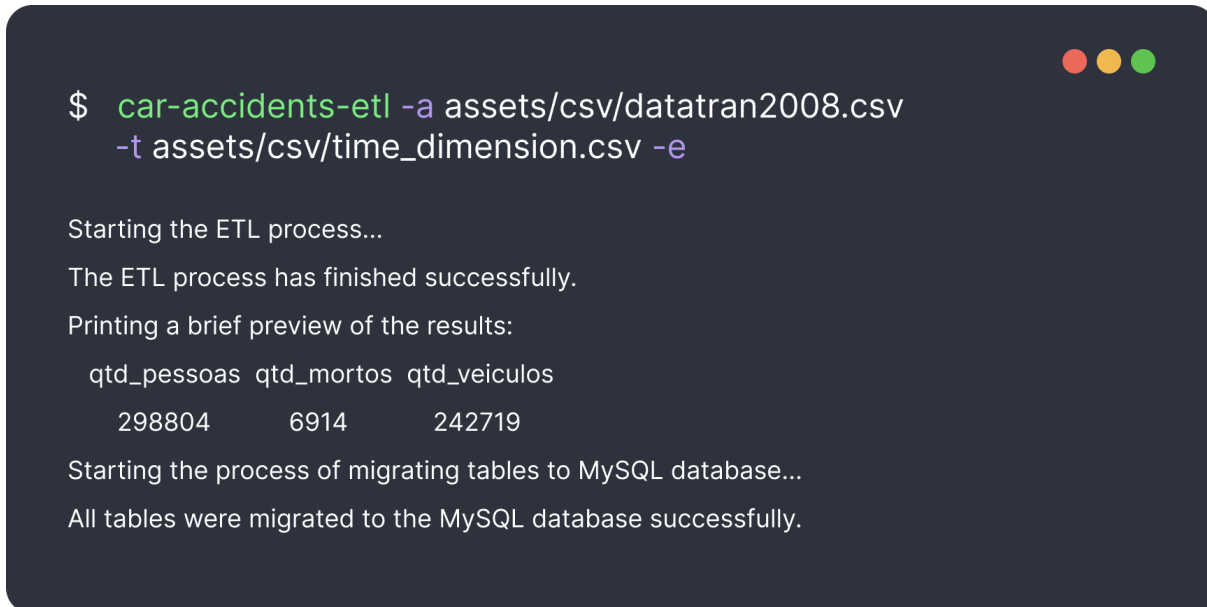
```

Fonte: Elaboração autoral

Para exportar os dados para o painel interativo é necessário fazer uso do parâmetro de exportação. Esse, quando informado, irá exportar o resultado do processo de ETL para o

banco de dados que é utilizado pelo painel interativo, atualizando-o de forma rápida. Por padrão, a ferramenta não irá exportar os dados.

Figura 8 - Resultados da execução com dados referente ao ano de 2008 e exportação

A terminal window with a dark background and three colored window control buttons (red, yellow, green) in the top right corner. The terminal shows the execution of a command and its output. The command is: \$ car-accidents-etl -a assets/csv/datatran2008.csv -t assets/csv/time_dimension.csv -e. The output includes: Starting the ETL process..., The ETL process has finished successfully., Printing a brief preview of the results:, a table with three columns: qtd_pessoas, qtd_mortos, qtd_veiculos, and three rows of data: 298804, 6914, 242719. The final output is: Starting the process of migrating tables to MySQL database... All tables were migrated to the MySQL database successfully.

```
$ car-accidents-etl -a assets/csv/datatran2008.csv
-t assets/csv/time_dimension.csv -e

Starting the ETL process...
The ETL process has finished successfully.
Printing a brief preview of the results:
  qtd_pessoas  qtd_mortos  qtd_veiculos
          298804         6914        242719

Starting the process of migrating tables to MySQL database...
All tables were migrated to the MySQL database successfully.
```

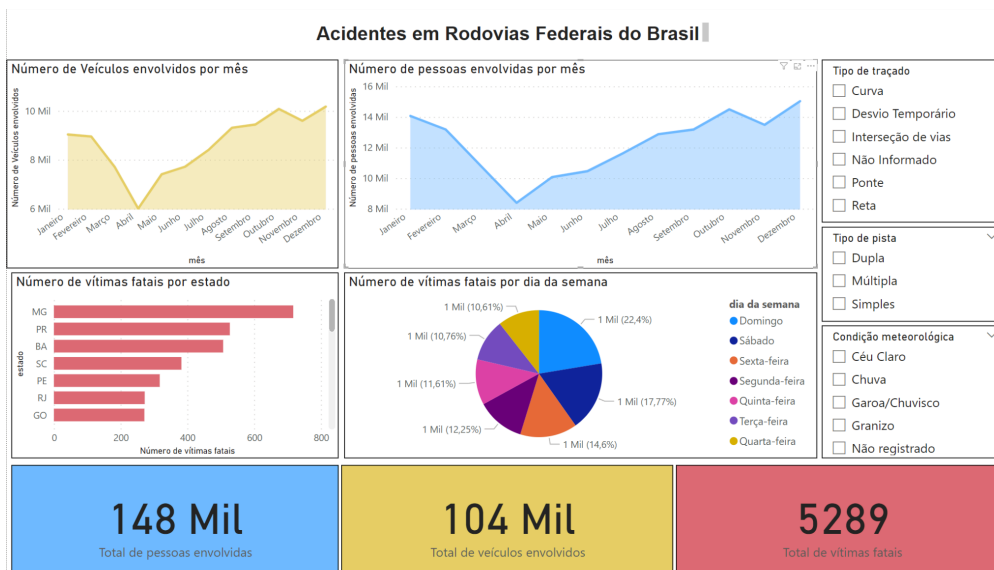
Fonte: Elaboração autoral

A execução ilustrada na Figura 8 processou 141.043 registros referente ao ano de 2008, onde houve a limpeza e normalização de dados e a organização dos mesmos dentro dos limites do modelo de ETL. Após isso, esses dados também foram exportados para o banco de dados que alimenta um Dashboard interativo que pode ser consultado para uma melhor visualização de informações.

2.5. DASHBOARD INTERATIVO

O dashboard interativo foi desenvolvido a partir do PowerBI, uma ferramenta da Microsoft para visualização e análise de dados. Este ilustra informações como o número de veículos por mês, número de pessoas envolvidas por mês, número de vítimas fatais por estado, número de vítimas fatais por dia da semana, além de opção de filtragem por tipo do traçado da via, tipo da pista e condição meteorológica. É importante ressaltar que todo o painel se adapta a medida das seleções realizadas pelo usuário.

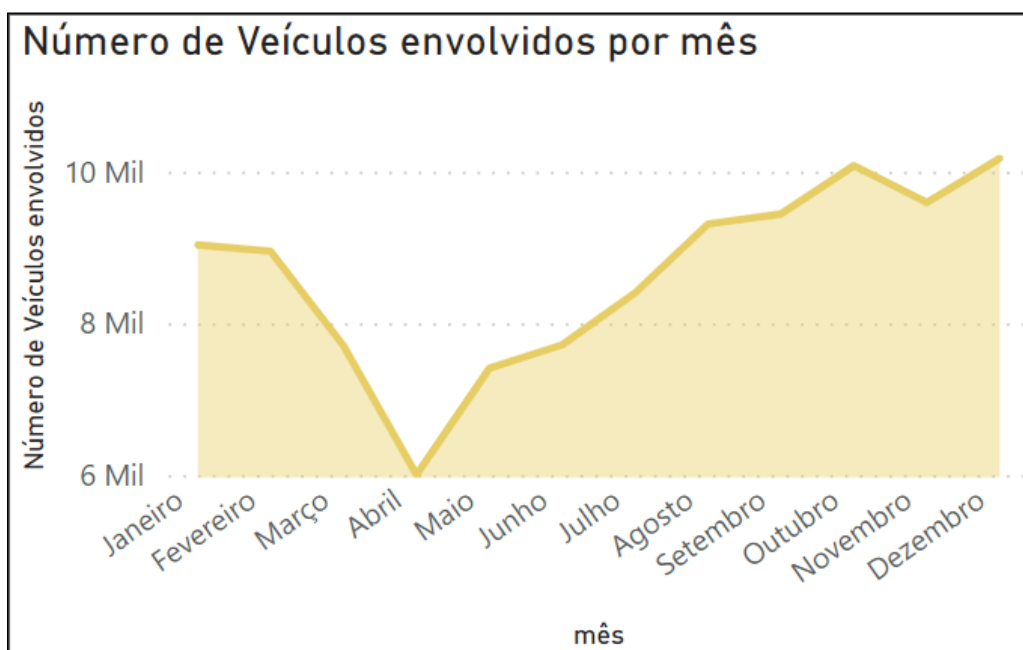
Figura 9 - Visão geral do Dashboard



Fonte: Elaboração autoral

O primeiro gráfico, localizado no topo superior esquerdo do painel interativo, ilustra por meio de um gráfico de área empilhada, o número de veículos que foram envolvidos em acidentes por mês. Ilustrado na Figura 8, tivemos cerca de 10 mil veículos envolvidos em acidentes de acordo com os registros do ano de 2020.

Figura 10 - Gráfico de número de veículos envolvidos em acidentes por mês



Fonte: Elaboração autoral

O segundo gráfico, localizado no centro superior, ilustra por meio de um gráfico de área empilhada, o número de pessoas envolvidas em acidentes por mês. Mais de 14 mil pessoas foram envolvidas em acidentes (Figura 9) de acordo com os registros do ano de 2020.

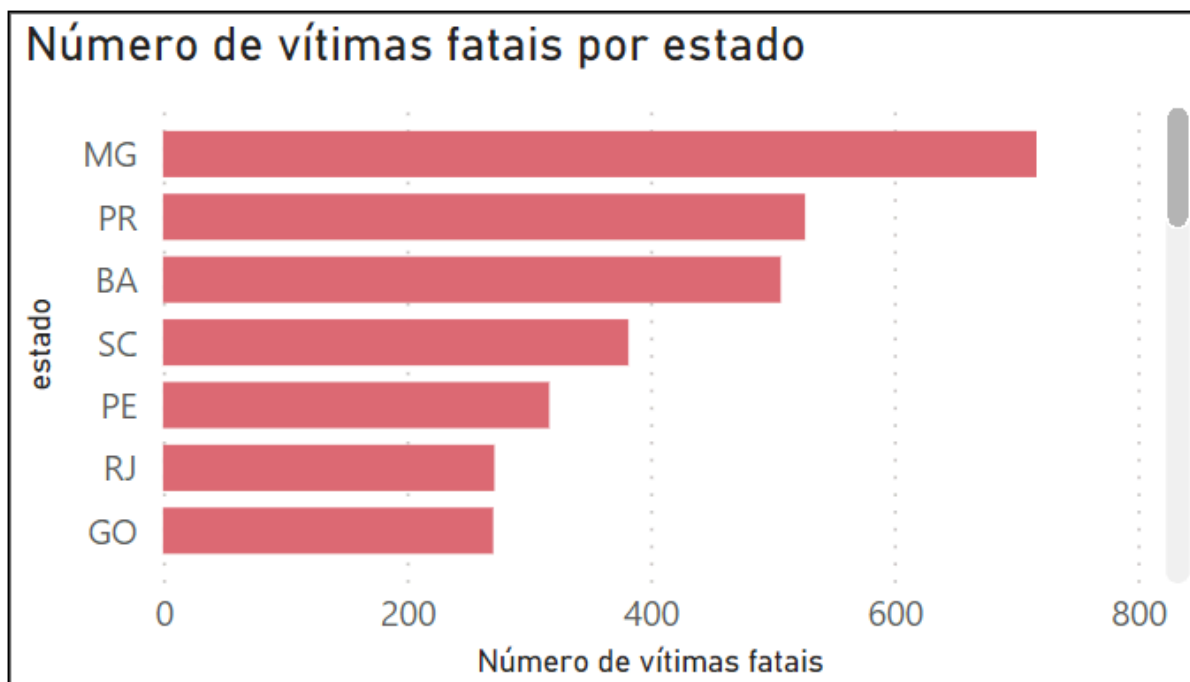
Figura 11 - Gráfico de número de pessoas envolvidas em acidentes por mês



Fonte: Elaboração autoral

O terceiro gráfico, localizado na região inferior esquerda, ilustra por meio de um gráfico de barras empilhadas, o número de vítimas fatais por estado. Como ilustra a Figura 12, para os registros de acidentes do ano de 2020, Minas Gerais foi o estado com mais vítimas fatais, logo abaixo temos Paraná e Bahia.

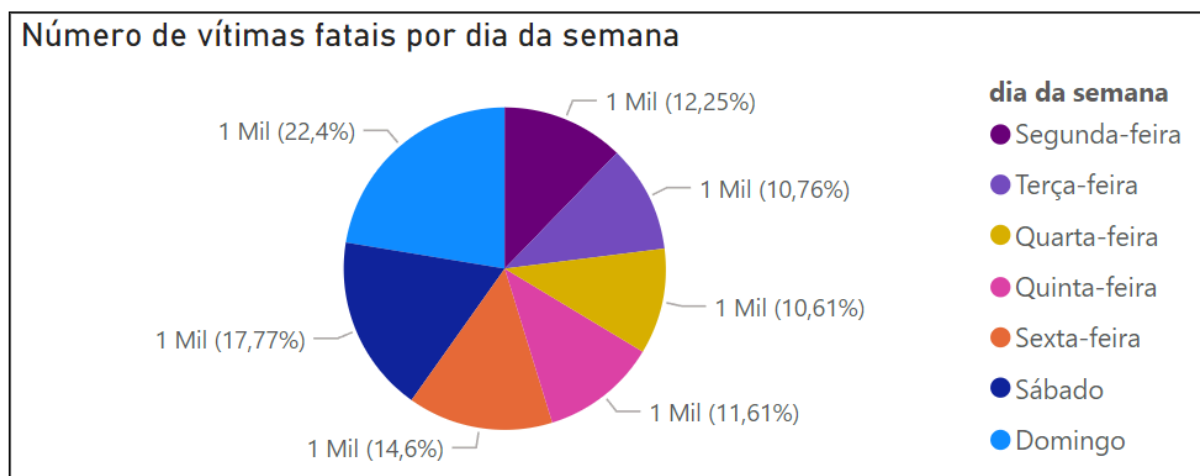
Figura 12 - Gráfico de número de vítimas fatais por estado



Fonte: Elaboração autoral

O quarto e último gráfico, ilustra por meio de um gráfico pizza, o número de vítimas fatais por dia da semana. Referente aos registros de acidentes do ano de 2020, Domingo foi o dia da semana com mais vítimas fatais, cerca de 22,4%, enquanto Quarta-feira e Quinta-feira estão empatados com a porcentagem de menos vítimas fatais dentre os demais dias da semana, cerca de 10,61%.

Figura 13 - Gráfico de número de vítimas fatais por dia da semana



Fonte: Elaboração autoral

Há também, no lado inferior do painel interativo, ilustrado através de cartões, as métricas totais de pessoas envolvidas em acidentes, veículos envolvidos em acidentes e vítimas fatais em acidentes (Figura 13). Essas informações, assim como todos os gráficos do painel, atualizam de acordo com as seleções feitas no painel.

Figura 14 - Cartões de ilustração sobre o total de pessoas envolvidas, veículos envolvidos e vítimas fatais



Fonte: Elaboração autoral

A aba de filtros (Figura 14) pode ser localizada no lado direito do painel interativo e disponibiliza filtragem de acordo com o tipo de traçado da via, o tipo de rodovia e condições meteorológicas.

Figura 15 - Caixas de filtro de acidentes

<p>Tipo de traçado</p> <p><input type="checkbox"/> Curva</p> <p><input type="checkbox"/> Desvio Temporário</p> <p><input type="checkbox"/> Interseção de vias</p> <p><input type="checkbox"/> Não Informado</p> <p><input type="checkbox"/> Ponte</p> <p><input type="checkbox"/> Reta</p>
<p>Tipo de pista</p> <p><input type="checkbox"/> Dupla</p> <p><input type="checkbox"/> Múltipla</p> <p><input type="checkbox"/> Simples</p>
<p>Condição meteorológica</p> <p><input type="checkbox"/> Céu Claro</p> <p><input type="checkbox"/> Chuva</p> <p><input type="checkbox"/> Garoa/Chuveisco</p> <p><input type="checkbox"/> Granizo</p> <p><input type="checkbox"/> Ignorado</p>

Fonte: Elaboração autoral

3. CONCLUSÃO

A interface de linha de comando desenvolvida torna a execução do processo de ETL, para os dados de acidentes em rodovias federais, mais simples e amigável aos desenvolvedores. Além das funcionalidades específicas para a base de dados da PRF, a ferramenta também dispõe da funcionalidade para gerar dados genéricos para alimentar a dimensão tempo dado um intervalo de datas, que pode ser utilizado em qualquer outro caso de uso.

A ferramenta oferece para desenvolvedores a capacidade de executar o complexo processo de ETL, com os dados públicos de acidentes da PRF, por meio de ferramentas que são comuns em seu cotidiano. Que busca, da forma que foi idealizada, proporcionar uma redução na curva de aprendizado sobre o que é necessário para executar esses passos e possibilitar a geração de Dashboard de análises de uma forma mais rápida e precisa.

O Dashboard interativo, que é alimentado a partir da execução da CLI, oferece cartões, gráficos e filtros para facilitar a visualização e análise dos dados da PRF, estes que podem ser segmentados ou agrupados por ano, a depender do conjunto de dados informado, e desde que estejam no formato estabelecido pela PRF. Por meio dessas análises, é possível entender a quantidade de acidentes, pessoas e carros envolvidos, em que condições meteorológicas, dia da semana, região, tipo de pista, tipo de traçado eles acontecem com mais frequência, para assim auxiliar na formulação de estratégias para evitá-los.

REFERÊNCIAS

ORACLE. Oracle Brasil: **o que é ETL?**, 2023. Disponível em: <https://www.oracle.com/br/integration/what-is-etl/>. Acesso em: 27 set. 2023.

CONCEIÇÃO, GLEICE MARGARETE DE SOUZA; ALENCAR, GIZELTON PEREIRA; LATORRE, MARIA DO ROSÁRIO DIAS DE OLIVEIRA. **Tendência temporal das internações por acidentes de trânsito na cidade de São Paulo**, Brasil, 2000-2019. Cadernos de Saúde Pública. Disponível em: <https://www.scielo.br/j/csp/a/8TRcknLktGxJmQNYGPYgxnz/#>. Acesso em: 12 out. 2023.

CARVALHO, SIANDRO RODRIGO DE. **Análise das ocorrências de acidentes de trânsito na cidade do Recife**. Recife, 2021. Disponível em: <https://repositorio.ifpe.edu.br/xmlui/bitstream/handle/123456789/752/An%c3%a1lise%20das%20ocorr%c3%aaancias%20de%20acidentes%20de%20tr%c3%a2nsito%20na%20cidade%20do%20Recife.pdf?sequence=1&isAllowed=y>. Acesso em: 14 fev. 2024.

GUIMARÃES, LEANDRO. **Arquitetura de BI: como funciona e como trabalhar esse conceito?** Rio de Janeiro. Disponível em: <https://www.knowsolution.com.br/arquitetura-bi-como-funciona-como-trabalhar-conceito/>. Acesso em: 14 fev. 2024.

TAVARES, JOSAFÁ. **Data Warehouse: como estruturar no modelo estrela?**. Rio de Janeiro, 2022. Disponível em: <https://www.mindtek.com.br/2022/10/como-estruturar-um-data-warehouse-em-modelo-estrela/>. Acesso em: 14 fev. 2024.

ANEXOS

ANEXO A - REGISTRO DE PROGRAMA DE COMPUTADOR DO INSTITUTO NACIONAL DE PROPRIEDADE INTELECTUAL



REPÚBLICA FEDERATIVA DO BRASIL
MINISTÉRIO DO DESENVOLVIMENTO, INDÚSTRIA, COMÉRCIO E SERVIÇOS
INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL
DIRETORIA DE PATENTES, PROGRAMAS DE COMPUTADOR E TOPOGRAFIAS DE CIRCUITOS

Certificado de Registro de Programa de Computador

Processo Nº: **BR512023001711-6**

O Instituto Nacional da Propriedade Industrial expede o presente certificado de registro de programa de computador, válido por 50 anos a partir de 1º de janeiro subsequente à data de 14/06/2023, em conformidade com o §2º, art. 2º da Lei 9.609, de 19 de Fevereiro de 1998.

Título: Software de linha de comando que executa o processo de ETL para gerar dashboards de conhecimento sobre Acidentes em Rodovias Federais do Brasil

Data de publicação: 14/06/2023

Data de criação: 12/01/2023

Titular(es): JOSÉ CARLOS GOMES FILHO

Autor(es): JONATAN LESSA DOS SANTOS; LEONARDO HENRIQUE DOS ANJOS SANTOS; KLEISSON DA SILVA MORAIS; JOSÉ CARLOS GOMES FILHO; ANDERSON FERNANDES DE ARAÚJO LEITE; ALEXANDRE CASTRO REBELO DA SILVA FILHO

Linguagem: SQL; PYTHON

Campo de aplicação: ED-06; IF-07; UB-04

Tipo de programa: AT-01; DS-02; FA-04; GI-01; GI-03; GI-04; SO-04; TI-02; UT-01

Algoritmo hash: SHA-512

Resumo digital hash:
c162ad4a143b97fcb2592852b238835849aae4e74f5058e9df5f87acdf6c9253fb70809661e0eb7e4c4b708c005f6f3b9fe
bc797a43a776045272ff39e08aef

Expedido em: 27/06/2023

Aprovado por:
Carlos Alexandre Fernandes Silva
Chefe da DIPTO

ANEXO B - DESCRIÇÃO DAS COLUNAS DA BASE DE DADOS

Coluna	Descrição
feridos_leves	Total de pessoas com ferimentos leves envolvidas na ocorrência.
feridos_graves	Total de pessoas com ferimentos graves envolvidas na ocorrência.
ilesos	Total de pessoas ilesas envolvidas na ocorrência.
ignorados	Total de pessoas envolvidas na ocorrência.
feridos	Total de pessoas feridas envolvidas na ocorrência (é a soma dos feridos leves com os graves).
veículos	Total de veículos envolvidos na ocorrência.