Business Intelligence: Análise de dados de acidentes de trânsito fornecidos pelo Ministério da Justiça utilizando data warehouse

João Augusto Melo Manarin
Graduando em Banco de Dados pela Fatec Bauru
joao.manarin@fatec.sp.gov.br

Kevin Willians da Silva
Graduando em Banco de Dados pela Fatec Bauru
kevin.Willians@fatec.sp.gov.br

Rebeca Peixoto Motti
Graduando em Banco de Dados pela Fatec Bauru
rebeca.motti@fatec.sp.gov.br

Orientador: Luis Alexandre da Silva
Mestre em Ciência da Computação e Docente na Fatec Bauru
luis.silva51@fatec.sp.gov.br

RESUMO

O aumento significativo da frota de veículos nas vias urbanas e rodovias tem contribuído diretamente para a elevação no número de acidentes de trânsito, gerando impactos sociais, econômicos e sobre os sistemas de saúde. Diante desse cenário, este trabalho tem como objetivo aplicar tecnologias de Business Intelligence (BI) e data warehousing na análise de grandes volumes de dados de acidentes, com o intuito de extrair insights relevantes que auxiliem na formulação de políticas públicas mais eficazes e estratégias de prevenção. A metodologia baseia-se na construção de um data warehouse para organização dos dados, no desenvolvimento de dashboards interativos e na realização de análises multidimensionais que permitam identificar padrões, tendências e fatores críticos, como horários de maior incidência, condições climáticas e localizações geográficas. A pesquisa também inclui uma revisão de literatura, que oferece embasamento teórico para o uso dessas tecnologias no contexto da segurança viária. Os resultados obtidos demonstram que o uso de BI possibilita uma compreensão mais profunda dos fatores associados aos acidentes, apoiando a tomada de decisões mais assertivas. Conclui-se que a integração de ferramentas de BI à análise de dados de trânsito é uma estratégia promissora para reduzir a ocorrência de acidentes e promover maior segurança nas vias.

Palavras-chave: Acidentes de trânsito; Business Intelligence; Data Warehouse; Análise de dados; Políticas públicas; Segurança viária

1. INTRODUÇÃO

O presente estudo tem como objetivo geral investigar como a aplicação de BI e data warehousing pode otimizar a análise de dados de acidentes de trânsito, promovendo insights que subsidiem a formulação e a avaliação de políticas públicas voltadas à melhoria da segurança no trânsito. Para alcançar esse objetivo, os seguintes objetivos específicos foram estabelecidos: (1) Analisar os padrões e tendências dos acidentes de trânsito, identificando horários, dias da semana e regiões com maior incidência; (2) Avaliar os fatores críticos associados aos acidentes, como condições climáticas e tipo de veículo; (3) Estudar o perfil demográfico das vítimas; (4) Desenvolver um sistema de análise visual com dashboards interativos, utilizando ferramentas de BI; e (5) Propor recomendações para políticas públicas baseadas nas análises realizadas. Esses objetivos visam aproveitar o potencial dos dados para enfrentar os desafios relacionados à segurança no trânsito e contribuir para a formulação de políticas públicas mais eficazes e informadas.

A mobilidade urbana tem se expandido significativamente nos últimos anos, trazendo não apenas benefícios econômicos e sociais, mas também um aumento alarmante no número de acidentes de trânsito, que afetam diretamente a segurança e o bem-estar da população. Esse cenário torna imprescindível a análise dos dados relacionados a esses acidentes, especialmente os fornecidos pelo Ministério da Justiça, a fim de entender suas causas e dinâmicas. Nesse contexto, as tecnologias de Business Intelligence (BI) e data warehousing se apresentam como ferramentas poderosas para transformar grandes volumes de dados em informações estratégicas, permitindo uma análise detalhada e visual sobre os acidentes de trânsito. A integração e organização dessas informações possibilitam a identificação de padrões críticos, como os horários e locais mais afetados, além do perfil demográfico das vítimas. Isso não só permite uma avaliação da eficácia das políticas públicas atuais, como também oferece subsídios para o desenvolvimento de novas estratégias de segurança no trânsito.

Apesar da grande quantidade de dados disponíveis sobre acidentes de trânsito no Brasil, ainda existe uma lacuna significativa na forma como essas informações são utilizadas para subsidiar políticas públicas efetivas. Muitas vezes, os dados são armazenados de maneira descentralizada, pouco acessível ou sem estrutura adequada para análises aprofundadas. A ausência de um sistema eficiente de visualização e interpretação desses dados dificulta a identificação de padrões, fatores de risco e perfis de vítimas, limitando a atuação preventiva e estratégica dos órgãos responsáveis pela segurança no trânsito.

A aplicação de ferramentas de BI e data warehousing surge como uma solução promissora para enfrentar esse desafio, pois permite transformar dados brutos em conhecimento acionável. Por meio da construção de painéis interativos e análises multidimensionais, é possível proporcionar uma visão abrangente e dinâmica do cenário dos acidentes de trânsito. Esse tipo de abordagem não apenas melhora o processo decisório por parte dos gestores públicos, como também contribui para a transparência e o controle social. Nesse sentido, o presente estudo se justifica pela sua relevância social, ao propor um modelo de análise que pode orientar ações mais eficazes e baseadas em evidências.

Estudos anteriores já demonstraram a utilidade de sistemas baseados em BI para a análise de acidentes de trânsito. Um exemplo é o trabalho de Silva e Costa (2021), que desenvolveram um dashboard para análise de dados de acidentes em

rodovias federais brasileiras, permitindo a visualização de informações por tipo de acidente, localização e horário. Outro exemplo é a pesquisa de Fernandes et al. (2020), que aplicou técnicas de data warehousing para integrar dados de diferentes fontes e analisar a influência de fatores meteorológicos nos acidentes urbanos. Esses trabalhos demonstram o potencial das tecnologias de BI e warehousing para subsidiar estratégias de prevenção e planejamento no contexto do trânsito.

2. FUNDAMENTAÇÃO TEÓRICA

A análise de dados tem se tornado uma ferramenta essencial para diversos setores, inclusive para a gestão pública e a segurança no trânsito. Com o avanço das tecnologias de armazenamento e visualização de dados, tornou-se possível transformar grandes volumes de dados em informações úteis para o planejamento e a tomada de decisão.

O Business Intelligence (BI) é definido como um conjunto de estratégias, tecnologias e práticas voltadas à coleta, integração, análise e apresentação de informações que auxiliam no processo decisório. Segundo Kimball e Ross (2013), o BI permite que organizações compreendam melhor seus dados, identifiquem padrões e obtenham insights que conduzam a ações mais eficazes. No setor público, especialmente em questões de mobilidade urbana e segurança no trânsito, o BI pode ser usado para identificar áreas de risco, avaliar políticas públicas e propor intervenções direcionadas.

Além disso, o BI atua como uma ponte entre os dados brutos armazenados nos sistemas operacionais e as decisões estratégicas. Através de ferramentas analíticas, painéis interativos e visualizações dinâmicas, é possível transformar grandes volumes de dados em informações acessíveis e úteis. Um componente fundamental para esse processo é o Data Warehouse, uma estrutura de armazenamento orientada à análise, geralmente organizada em Data Marts, que são subconjuntos específicos voltados a temas ou áreas de negócio.

A aplicação de BI no setor público tem ganhado destaque pela sua capacidade de tornar a gestão mais eficiente, transparente e baseada em dados reais. No contexto do trânsito, estudos como o de Santos e Ferreira (2022) mostram que ferramentas de BI permitem identificar os principais fatores de risco, direcionar campanhas educativas e otimizar os recursos disponíveis para fiscalização e infraestrutura viária. A integração entre dados, tecnologia e políticas públicas é essencial para reduzir a acidentalidade e salvar vidas.

O Data Warehouse é um repositório de dados que centraliza informações oriundas de diferentes fontes, organizando-as de forma estruturada para análise. Sua arquitetura multidimensional permite consultas rápidas e flexíveis, facilitando a geração de relatórios e dashboards. De acordo com Kimball e Ross (2013), o uso de tabelas fato e dimensão viabiliza uma análise mais aprofundada e segmentada dos dados, sendo essencial para estudos como o presente, que envolve múltiplas variáveis como tempo, localização, tipo de acidente e perfil da vítima.

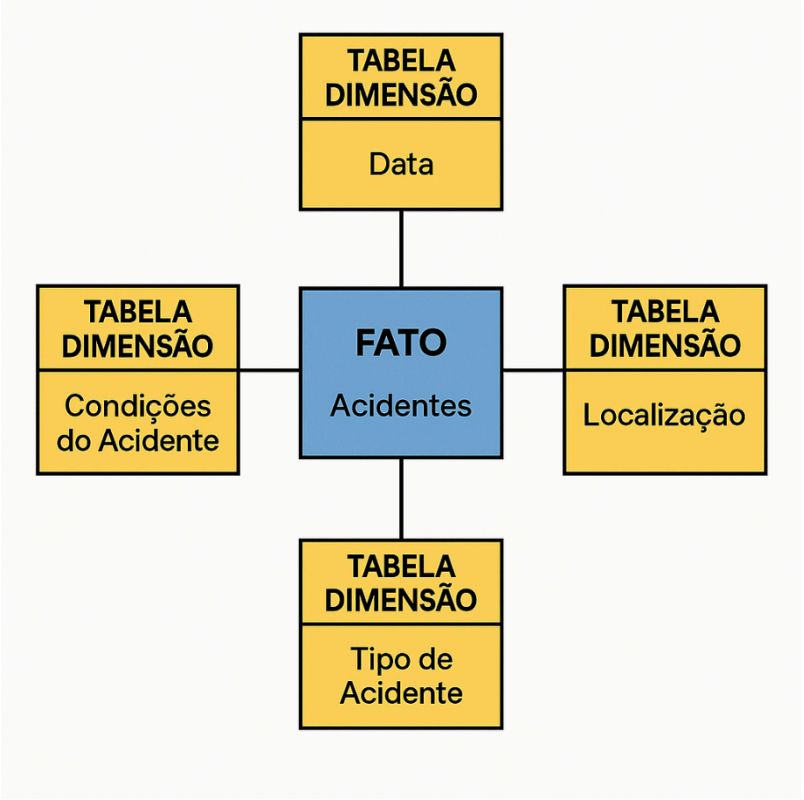
O modelo de Data Mart Estrela (Star Schema) é um dos mais populares e utilizados em projetos de BI, especialmente por sua simplicidade e eficiência na consulta de dados. Esse modelo foi proposto por Ralph Kimball e é amplamente adotado em ambientes onde a performance e a facilidade de compreensão são essenciais.

No esquema estrela, temos:

a) Tabela Fato (Fact Table): É o centro da estrela e contém os dados quantitativos (medidas), como número de acidentes, número de vítimas, entre outros. Essa tabela também possui chaves estrangeiras que se relacionam com as tabelas dimensão;
b) Tabelas Dimensão (Dimension Tables): Rodeando a tabela fato, estão as tabelas dimensão, que contêm dados descritivos, como data, localização, tipo de acidente, condição climática, entre outros. Essas dimensões fornecem o contexto necessário para análise dos fatos.

A estrutura forma visualmente uma estrela, com a tabela fato no centro e as dimensões ao redor, conforme o exemplo a seguir:

Figura 1 – Exemplo Star Schema



Fonte: Autores

Dentre as vantagens do Modelo estrela é possível ressaltar:

a) Facilidade de entendimento: É intuitivo para analistas e tomadores de decisão;
b) Performance: Otimizado para consultas e relatórios, pois evita junções complexas;
c) Organização lógica: Separação clara entre medidas e atributos descritivos.

No contexto deste projeto, que envolve a análise de dados de acidentes de trânsito fornecidos pelo Ministério da Justiça, o modelo estrela será utilizado para organizar e facilitar a análise multidimensional desses dados. A tabela fato conterá os registros com medidas como quantidade de acidentes e mascaramento de dados

descritivos com quantidade limitada sem a necessidade de criação de uma tabela à parte, como por exemplo, gênero. As tabelas dimensão incluirão dados como:

a) Dimensão Tempo (mês, ano, dia da semana);
b) Dimensão Região (nome do estado, UF);
c) Dimensão Tipo de Acidente (colisão, atropelamento, capotamento etc.).

Esse modelo permitirá criar painéis dinâmicos em ferramentas de BI, promovendo uma visão analítica e integrada sobre os acidentes de trânsito no Brasil.

A visualização de dados desempenha um papel fundamental na análise exploratória, pois permite identificar padrões, tendências, anomalias e correlações de forma clara, objetiva e acessível. Através de gráficos, mapas, tabelas dinâmicas e painéis interativos, a informação torna-se visualmente compreensível, favorecendo decisões mais rápidas e embasadas.

Ferramentas como Power BI, além de bibliotecas Python como Matplotlib, Seaborn e Plotly, têm sido amplamente utilizadas nesse processo. No Power BI, por exemplo, é possível criar dashboards interativos que conectam múltiplas fontes de dados, permitindo que o usuário filtre e explore informações com poucos cliques. Já em ambientes de programação, bibliotecas como Seaborn oferecem recursos para gerar visualizações estatísticas com alto nível de personalização e detalhamento, como gráficos de dispersão, heatmaps e histogramas.

Segundo Chen e Mao (2021), a análise visual de dados é uma etapa crucial na construção de conhecimento, pois torna a interpretação mais acessível e contribui para a construção de estratégias baseadas em evidências. A visualização transforma os dados em narrativas visuais que facilitam a comunicação entre técnicos, gestores e tomadores de decisão.

A Análise Exploratória de Dados (EDA – Exploratory Data Analysis) é a fase em que o analista examina os dados de forma livre, buscando entender sua estrutura, distribuição, variáveis mais relevantes e possíveis inconsistências. A EDA ajuda a:

a) Compreender a distribuição e comportamento dos dados;
b) Identificar valores ausentes ou discrepantes;
c) Verificar relações entre variáveis;
d) Formular hipóteses e direcionar análises futuras.

A combinação da análise exploratória com visualizações impactantes reforça o poder do BI como ferramenta de transformação de dados em inteligência, especialmente em temas sensíveis como a segurança no trânsito.

Os acidentes de trânsito representam um dos maiores desafios para a mobilidade urbana no Brasil. Dados disponibilizados por órgãos como o Ministério da Justiça e o DENATRAN permitem análises detalhadas sobre as causas e características desses eventos. Segundo Wang, Li e Xu (2023), o uso de big data e técnicas analíticas avançadas pode auxiliar na previsão e prevenção de acidentes, fornecendo subsídios para a formulação de políticas públicas mais eficazes.

A base de dados pública utilizada foi disponibilizada pelo Ministério da Justiça e Segurança Pública, contendo dados abertos. Dados Abertos são dados institucionais, disponibilizados em formato legível por máquina e sem restrição de licenças, patentes ou mecanismos de controle, que qualquer pessoa pode livremente usá-los, reutilizá-los e redistribuí-los. Os arquivos disponibilizados estão compactados (zipados) para diminuição do tamanho real e também no formato CSV (Comma-Separated Values) é um tipo de arquivo de texto simples que

armazena dados em formato de tabela, onde cada linha representa um registro e cada campo é separado por um delimitador ponto e vírgula, é um formato de texto simples, fácil de ler e manipular, amplamente suportado por diversos programas, como planilhas eletrônicas (Excel, LibreOffice Calc) e bancos de dados, facilitando a importação e exportação de dados entre diferentes aplicações.

Utilizamos aproximadamente 72 mil dados detalhados sobre acidentes de trânsito coletados pela Polícia Rodoviária Federal (PRF) em território nacional do ano de 2024. Os dados em sua fonte não passaram pela metodologia de tratamento de bases de dados, portanto foi realizado tratamento e estruturação dos dados para a realização do estudo. As informações estavam disponíveis em dois formatos: agrupados por ocorrência de acidente ou por pessoa envolvida, a opção da visão por ocorrência foi a escolhida assim tendo uma única visão de cada acidente registrado. As variáveis analisadas incluíram dados temporais (data, hora, dia da semana), geográficos (região e unidade federativa), características do evento (tipo e classificação do acidente, fase do dia, condições climáticas, tipo de pista e tipo de veículo envolvido), e o perfil das vítimas (faixa etária, gênero, estado físico).

3. MATERIAIS E MÉTODOS

Inicialmente, os dados brutos foram submetidos a um processo de limpeza e padronização utilizando ferramentas como Microsoft Excel e Google Sheets. Essa fase consistiu na remoção de duplicidades, no preenchimento de valores ausentes com base em regras consistentes e na padronização de formatos, como datas (convertidas para o formato ISO: AAAA-MM-DD) e categorias textuais. Além disso, foram criadas colunas auxiliares com base nos dados originais, como faixa etária, turnos do dia e classificações qualitativas, a fim de facilitar as análises subsequentes. Esse trabalho preliminar garantiu maior integridade, coerência e qualidade aos dados, reduzindo o risco de inconsistências nas etapas seguintes.

Após esse tratamento inicial, os dados foram importados para o sistema de gerenciamento de banco de dados MySQL, com o uso da interface phpMyAdmin, que permitiu a criação e o gerenciamento prático das tabelas. A modelagem seguiu a arquitetura dimensional do tipo estrela, ideal para ambientes de análise. O modelo foi composto por uma tabela fato que centralizou os registros dos acidentes de trânsito, e por tabelas dimensão, responsáveis por fornecer contexto às análises.

A partir dessa estrutura, foi possível montar o Data Mart com as principais necessidades a serem analisadas e realizar consultas analíticas utilizando a linguagem SQL. As instruções SQL possibilitaram o cruzamento eficiente entre a tabela fato e as dimensões, viabilizando análises segmentadas e detalhadas. Consultas como a contagem de acidentes por faixa etária e por região, a distribuição dos tipos de acidente por estado, e a frequência de determinados veículos em áreas específicas foram realizadas por meio de comandos como JOIN, GROUP BY, WHERE e ORDER BY. Esse processo permitiu a extração de padrões e a geração de tabelas intermediárias com informações estatísticas relevantes.

Simultaneamente às consultas em SQL, a linguagem Python foi empregada com o uso das bibliotecas Pandas, Matplotlib e Seaborn. Essa etapa teve como finalidade aprofundar a análise estatística e criar visualizações personalizadas, como histogramas, boxplots, gráficos de barras e dispersão. Através do Python, foi

possível explorar correlações entre variáveis, calcular medidas descritivas (como médias, medianas, desvios padrão) e representar visualmente tendências nos dados. Essa abordagem complementar permitiu o refinamento dos insights gerados e forneceu suporte estatístico mais robusto às interpretações realizadas.

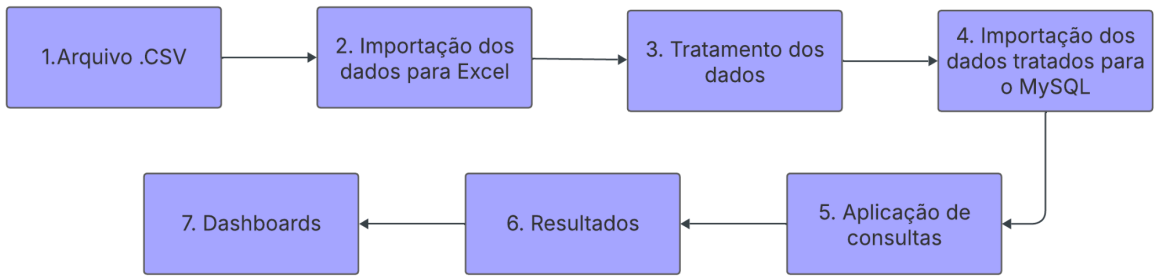
A construção dos dashboards permitiu a criação de gráficos dinâmicos e painéis com filtros, segmentações e indicadores-chave, assim como os gerados em ferramentas para BI como: Power BI, Tableau, Looker Studio, etc. Com isso, foi possível identificar com clareza regiões com maior incidência de acidentes, perfis demográficos mais vulneráveis, horários de maior risco e o comportamento dos diferentes tipos de veículos nas diversas localidades analisadas.

Ao longo da análise, observou-se, por exemplo, que determinadas regiões apresentaram uma maior frequência de vítimas do sexo masculino em faixas etárias mais jovens, enquanto outras revelaram predominância de vítimas femininas. Também foi possível identificar que motocicletas estavam fortemente associadas a acidentes no Nordeste, enquanto veículos de carga apresentavam maior incidência em áreas industriais do Sul. Tais achados destacam a importância de ações específicas, como campanhas educativas regionalizadas e investimentos em infraestrutura direcionada.

A modelagem dimensional adotada no banco de dados foi essencial para garantir agilidade, escalabilidade e confiabilidade nas análises. A integração entre as tabelas, bem como a consistência nas chaves de relacionamento, possibilitou responder a perguntas estratégicas com precisão, como a quantidade de acidentes por faixa etária em uma determinada região ou a associação entre tipos de veículos e gravidade dos acidentes. Assim, a estrutura criada não apenas atendeu aos objetivos técnicos do estudo, como também ofereceu subsídios valiosos para a formulação de políticas públicas voltadas à segurança viária.

Todo o processo metodológico descrito pode ser sintetizado por meio do fluxograma apresentado na Figura 2. O diagrama ilustra de forma clara e sequencial as etapas seguidas no trabalho, desde a obtenção e o tratamento inicial dos dados em formato .CSV, passando pela importação no MySQL e aplicação das consultas analíticas, até a geração dos resultados e a construção dos dashboards interativos, permitindo visualizar de maneira integrada o caminho percorrido durante a execução do projeto, reforçando a compreensão das etapas e a lógica adotada em cada fase da análise.

Figura 2 – Fluxograma do processo metodológico



Fonte: Autores

A estrutura representada na imagem abaixo corresponde ao modelo estrela desenvolvido com foco na análise de acidentes de trânsito. Trata-se de um modelo dimensional baseado na abordagem de fato e dimensões, com o objetivo de

otimizar consultas analíticas e permitir visualizações rápidas e segmentadas dos dados.

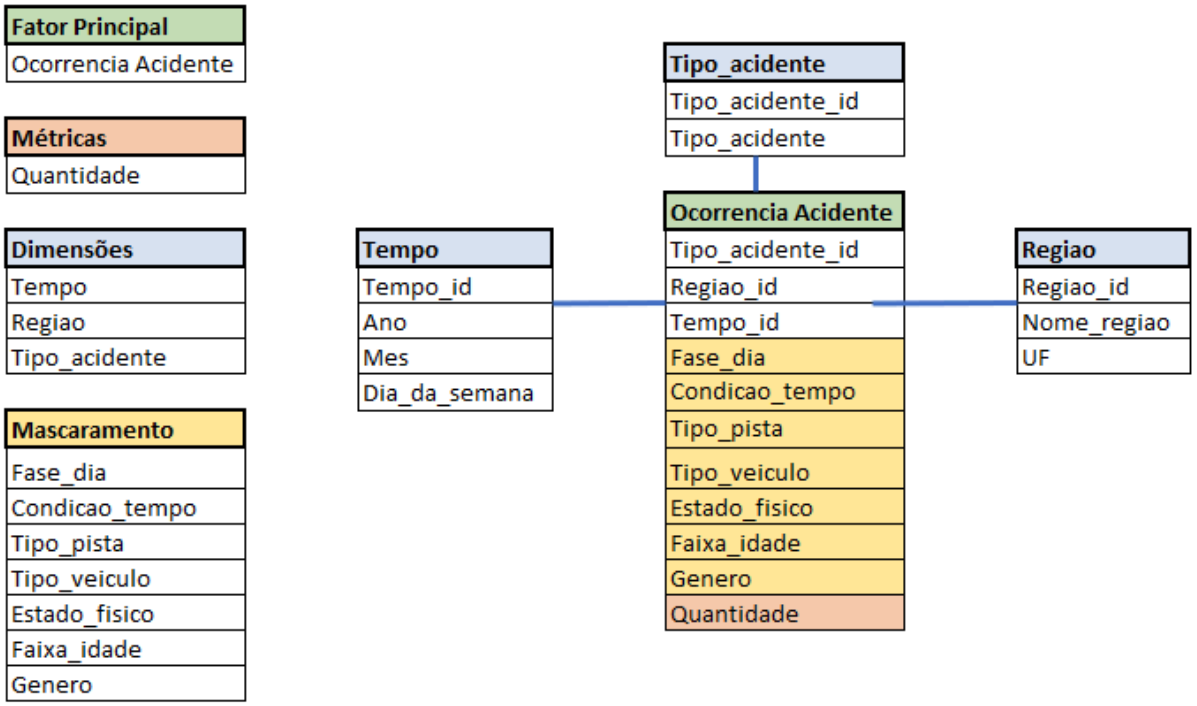
No centro do modelo está a tabela fato "Acidente", que armazena os principais dados quantitativos relacionados às ocorrências, como o número de acidentes. Esta tabela contém chaves estrangeiras que a conectam às tabelas dimensionais, permitindo a análise dos acidentes sob diversas perspectivas, as dimensões contempladas são:

a) Tempo: com os atributos de ano, mês e dia da semana, permite análises sazonais e temporais;
b) Região: composta pelos atributos nome da região e UF, possibilita o mapeamento geográfico dos acidentes;
c) Tipo do Acidente: classifica os acidentes conforme sua natureza, o que auxilia na compreensão da gravidade e perfil das ocorrências.

Além das dimensões, a modelagem inclui atributos complementares na tabela fato que funcionam como elementos de mascaramento, colunas com dados de diversidade limitada onde não há a necessidade da criação de uma dimensão para a mesma, permitindo análises mais detalhadas. Entre eles, destacam-se: fase do dia, condição do tempo, tipo de pista, tipo do veículo, estado físico das vítimas, faixa etária e gênero.

Esse modelo foi projetado para ser compatível com ferramentas de Business Intelligence, como Power BI e Tableau, bem como ambientes de análise baseados em SQL ou Python, assegurando flexibilidade e escalabilidade para diferentes usos.

Figura 3 - Esquema estrela dos dados utilizados



Fonte: Autores

Em resumo, o modelo estruturado neste projeto é a base técnica que sustenta as análises desenvolvidas, viabilizando a extração de insights com rapidez

e precisão, fundamentais para a construção de soluções inteligentes voltadas à redução dos acidentes de trânsito no Brasil.

Dando continuidade à abordagem baseada em visualizações e modelagem dimensional, é importante destacar que as análises realizadas no estudo foram diretamente orientadas por necessidades específicas a serem analisadas, cujo objetivo central consiste na compreensão aprofundada das dinâmicas dos acidentes de trânsito para subsidiar decisões estratégicas.

A figura 4 ilustra uma versão simplificada e direcionada do modelo estrela, que podemos chamar de Data Mart, voltada especificamente para atender duas demandas principais de análise:

a) Quantidade de vítimas por região de acordo com o gênero e a faixa etária;
b) Quantidade de vítimas por região de acordo com o tipo de veículo.

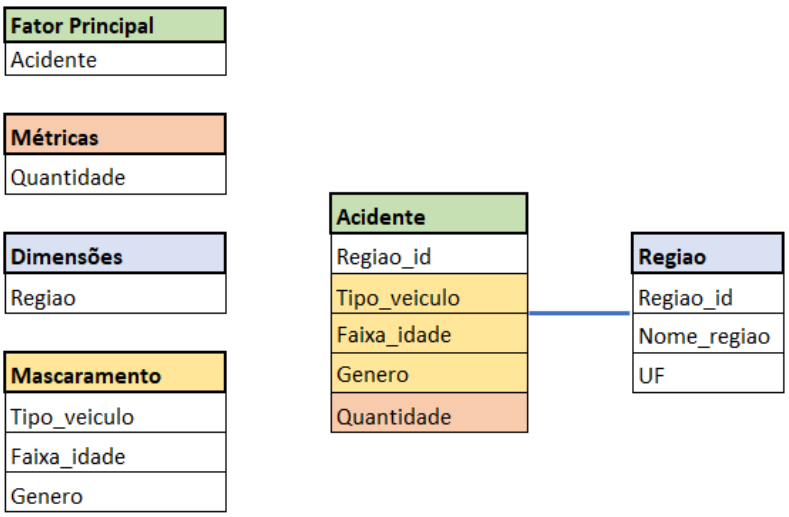
Para suprir essas necessidades, a estrutura é composta por duas entidades principais:

a) A tabela "Acidente": armazena os dados quantitativos, mascaramentos e chaves estrangeiras relevantes para a análise;
b) A tabela "Regiao": contém os dados geográficos (Nome_regiao e UF) e se relaciona com a tabela de acidente por meio da chave estrangeira Regiao_id.

Essa modelagem objetiva facilita a construção de consultas específicas eliminando dados desnecessários, além de alimentar ferramentas de BI com dados organizados de forma otimizada para análise a fim de garantir melhor performance.

Alinhando os requisitos de informação e as necessidades a serem avaliadas junto com uma modelagem bem estruturada, a solução proposta contribui para uma tomada de decisão mais informada, regionalizada e segmentada por perfil de vítima, possibilitando ações preventivas mais eficazes. Dessa forma, os dados deixam de ser apenas registros históricos e passam a atuar como insumos estratégicos para a gestão da segurança no trânsito.

Figura 4 - Data Mart



Fonte: Autores

4. DISCUSSÃO E RESULTADOS

A figura 5 representa o código utilizado dentro da plataforma MySQL para criação das tabelas com dados analíticos do Data Mart para atender às necessidades 1 e 2.

Figura 5 - Código SQL criação das tabelas

```
DROP TABLE IF EXISTS REGIAO;
CREATE TABLE REGIAO(
    ID_REGIAO INT PRIMARY KEY,
    UF CHAR(2),
    REGIAO VARCHAR(25));

DROP TABLE IF EXISTS OCORRENCIA;
CREATE TABLE OCORRENCIA(
    ID_OCORRENCIA INT PRIMARY KEY,
    DATA DATE,
    DIA_SEMANA VARCHAR(20),
    HORA INT,
    ID_REGIAO INT,
    TIPO_ACIDENTE VARCHAR(100),
    CLASSIFICACAO_ACIDENTE VARCHAR(50),
    FASE_DIA VARCHAR(50),
    CONDICAO_TEMPO VARCHAR(50),
    TIPO_PISTA VARCHAR(50),
    TIPO_VEICULO VARCHAR(50),
    ESTADO_FISICO VARCHAR(50),
    IDADE INT,
    FAIXA_IDADE VARCHAR(50),
    QTD INT,

    CONSTRAINT FK_OCORRENCIA_REGIAO FOREIGN KEY (ID_REGIAO) REFERENCES
REGIAO(ID_REGIAO));
```

Fonte: Autores

Referente à necessidade 1 que trata da análise da quantidade de vítimas por região de acordo com o gênero e a faixa etária, o código na figura 6 representa extração do resultado sintético que será posteriormente utilizado para gerar dashboards.

Figura 6 - Extração dos dados sintéticos da necessidade 1

```
SELECT GENERO, FAIXA_IDADE, R.REGIAO, SUM(QTD) AS QTD
FROM OCORRENCIA O
INNER JOIN REGIAO R ON R.ID_REGIAO = O.ID_REGIAO
GROUP BY GENERO, FAIXA_IDADE, R.REGIAO;
```

Fonte: Autores

Essa consulta retorna a soma da quantidade de vítimas agrupadas por gênero, faixa etária e região. A análise permite identificar as regiões mais afetadas, os grupos demográficos mais vulneráveis e os padrões de gênero. As regiões com maior incidência de vítimas podem ser priorizadas para ações preventivas. Além disso, é possível observar concentrações de vítimas masculinas ou femininas em determinadas faixas etárias, o que fornece subsídios importantes para a elaboração de políticas públicas direcionadas. Como exemplo, destacam-se os jovens do gênero masculino, que são mais afetados em áreas urbanas.

Já a necessidade 2, que trata da análise da quantidade de vítimas por região de acordo com o tipo do veículo, o código na figura 7 representa extração do resultado sintético que será posteriormente também utilizado para gerar dashboards.

Figura 7 - Extração dos dados sintéticos da necessidade 2

```
SELECT TIPO_VEICULO, R.REGIAO, SUM(QTD) AS QTD
FROM OCORRENCIA O
INNER JOIN REGIAO R ON R.ID_REGIAO = O.ID_REGIAO
GROUP BY TIPO_VEICULO, R.REGIAO;
```

Fonte: Autores

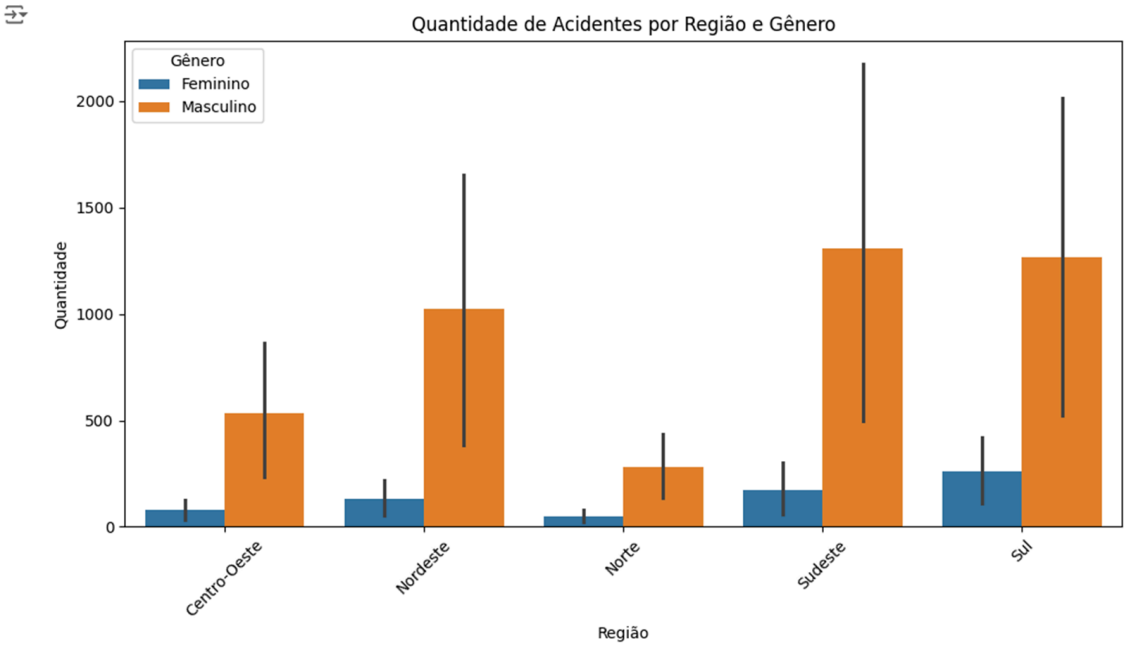
Essa consulta calcula a soma das vítimas associadas a diferentes tipos de veículos, segmentadas por região, permitindo identificar padrões relevantes. Observa-se, por exemplo, a prevalência de acidentes envolvendo motocicletas em determinadas regiões, como no Nordeste, enquanto no Sul há uma alta incidência de veículos de carga, possivelmente relacionada à maior presença de áreas industriais ou rodoviárias. A partir desses resultados, é possível direcionar ações específicas: veículos mais frequentemente envolvidos, como motocicletas e automóveis, podem ser foco de campanhas educativas ou de aprimoramentos na legislação de trânsito. Já as regiões com predominância de veículos de carga podem demandar investimentos em infraestrutura rodoviária, com o objetivo de reduzir os riscos e melhorar a segurança no trânsito.

A visualização de dados por meio de gráficos configura-se como uma estratégia essencial na análise de padrões e tendências relacionadas aos acidentes de trânsito. As representações visuais não apenas facilitam o acesso e a interpretação das informações por diferentes públicos, como também contribuem para a identificação de correlações complexas entre variáveis, tais como região geográfica, gênero, faixa etária e tipo de veículo envolvido. Essa abordagem permite uma compreensão mais aprofundada dos fatores associados à ocorrência de acidentes, servindo como base para a formulação de políticas públicas mais eficazes e fundamentadas em evidências.

Neste estudo, foram utilizadas ferramentas de análise de dados como Python, com o apoio das bibliotecas especializadas Pandas, Matplotlib e Seaborn, que possibilitaram a geração de gráficos com elevado grau de qualidade e capacidade analítica. Tais visualizações permitiram a identificação de padrões relevantes, fornecendo subsídios concretos para o planejamento de ações de segurança no trânsito mais direcionadas.

De acordo com a figura 8, observa-se a distribuição da quantidade de acidentes nas diferentes regiões do Brasil, discriminados por gênero (feminino e masculino). Os dados revelam uma clara predominância de vítimas do gênero masculino em todas as regiões, quando comparado ao gênero feminino. As regiões Sudeste e Sul se destacam por apresentarem os maiores quantitativos de acidentes para ambos os gêneros, com uma diferença expressiva a favor do gênero masculino. Isso pode estar relacionado ao maior número de veículos, densidade populacional e circulação urbana mais intensa nessas regiões, como dito anteriormente, possivelmente relacionada à maior presença de áreas industriais ou rodoviárias. Este achado reforça a necessidade de campanhas educativas voltadas principalmente ao público masculino, principalmente nas regiões Sudeste e Sul, onde a incidência é maior, dada sua maior exposição a comportamentos de risco no trânsito.

Figura 8 – Gráfico referente à necessidade 1



Fonte: Autores

A figura 9 ilustra a quantidade de acidentes de trânsito por faixa etária e gênero. O gráfico revela que a maioria dos acidentes está concentrada no gênero masculino, independentemente da faixa etária. A discrepância entre os gêneros é bastante evidente, especialmente nas faixas mais produtivas e economicamente ativas da população.

Observa-se que as faixas entre os 21 e 50 anos possuem os maiores registros de acidentes, sendo que o pico ocorre na faixa de 31 a 40 anos. Isso sugere que adultos jovens e de meia-idade do gênero masculino estão mais expostos ou mais envolvidos em situações de risco no trânsito, revelando sua vulnerabilidade, possivelmente relacionados à maior frequência de deslocamento para trabalho, deslocamento urbano, lazer ou estilo de direção com comportamentos arriscados.

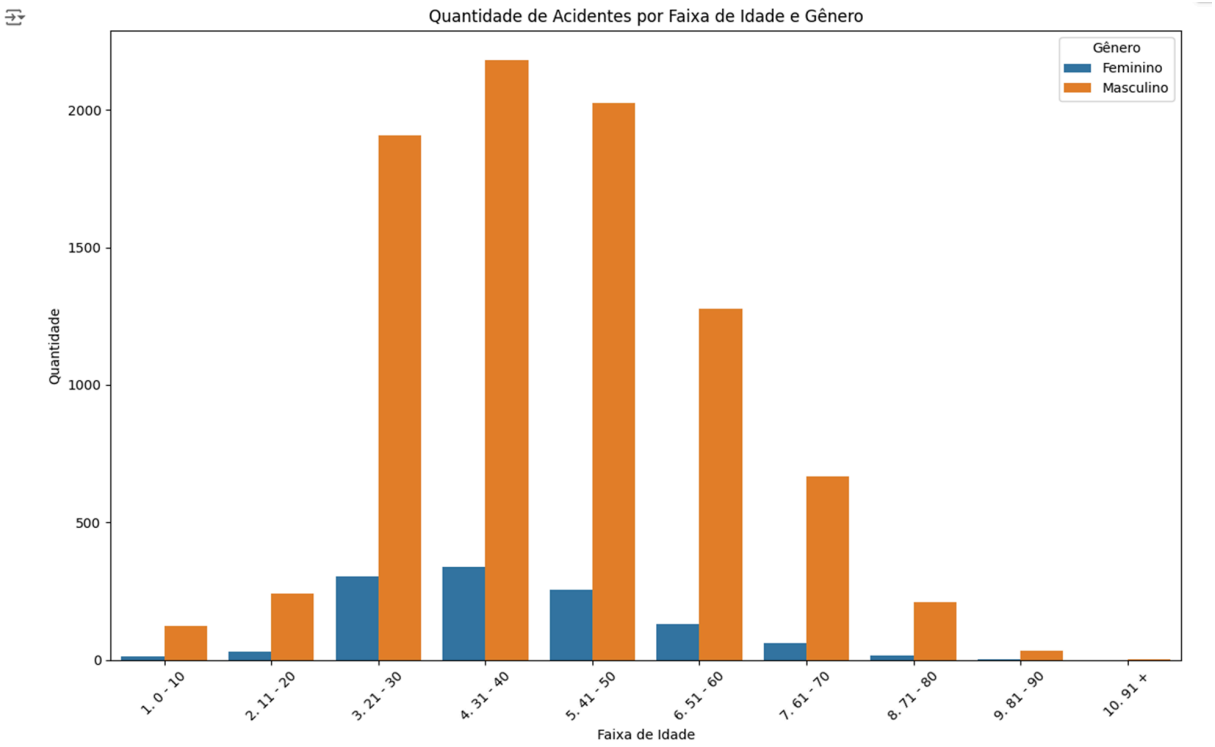
Tal constatação reforça a necessidade de campanhas educativas específicas voltadas à juventude, com ênfase em práticas de direção segura e consciente.

O gênero feminino mantém uma participação significativamente menor em todas as faixas etárias, demonstrando uma direção segura, embora haja um ligeiro aumento nas faixas de 21 a 50 anos, que seguem o mesmo padrão das faixas mais críticas do masculino, porém em proporções bem menores.

Os dados mostram que políticas públicas de segurança viária podem ser mais efetivas se direcionadas às faixas etárias de 21 a 50 anos, onde há clara concentração dos casos independente do gênero.

A utilização de gráficos como estes demonstrou uma ferramenta valiosa na análise de dados, permitindo identificar com clareza os perfis mais suscetíveis aos acidentes. Essas informações são fundamentais para a construção de políticas públicas mais direcionadas, sensíveis às variáveis etárias, de gênero e às especificidades regionais, promovendo maior eficácia nas ações de prevenção e segurança no trânsito

Figura 9 – Gráfico 2 referente à necessidade 1



Fonte: Autores

A figura 10 demonstra a quantidade de acidentes distribuídos por tipo de veículo (automóvel, caminhão, moto e ônibus) nas diferentes regiões do Brasil. Observa-se que os acidentes envolvendo automóveis apresentam os maiores índices em todas as regiões. É possível perceber que os automóveis registraram os maiores índices de ocorrência nas regiões Sudeste e Sul, refletindo o elevado volume da frota particular e a intensa circulação urbana nessas áreas.

Os acidentes com caminhões também possuem números elevados, especialmente nas regiões Sudeste e Sul, seguidos do Nordeste, o que pode estar associado ao intenso tráfego de transporte de cargas nessas regiões, que possuem corredores logísticos e industriais mais desenvolvidos

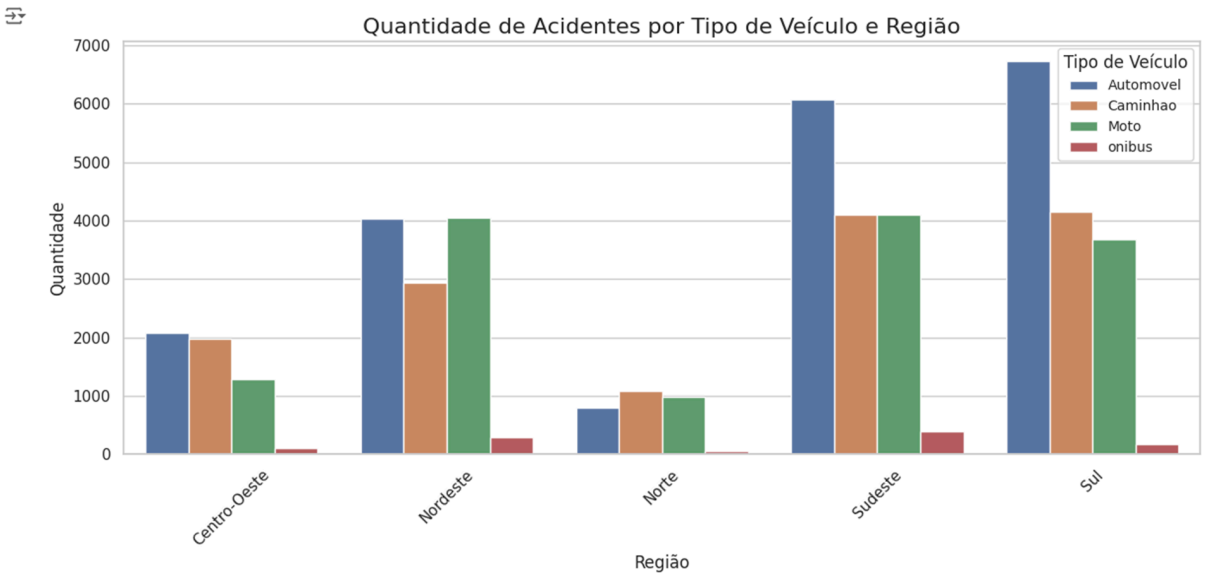
Já no que se refere às motos, há uma incidência muito alta na região Nordeste, praticamente equivalente aos acidentes com automóveis. Isso sugere uma dependência maior do transporte por motocicleta, possivelmente por razões socioeconômicas, urbanização dispersa ou oferta limitada de transporte público.

Por outro lado, os acidentes envolvendo ônibus possuem incidência baixa em todas as regiões, sendo um indicativo de que esse tipo de transporte, embora essencial, indica menor exposição ao risco, apresenta menor participação nas estatísticas de acidentes, possivelmente devido a medidas de segurança mais rigorosas e fiscalização específica desses veículos.

Esses dados reforçam a importância da formulação de políticas públicas segmentadas, que considerem o tipo de veículo como uma variável estratégica para o planejamento de ações de fiscalização, educação no trânsito e regulamentação. A adoção de medidas específicas voltadas a cada categoria

veicular pode contribuir significativamente para a redução dos índices de acidentes e a promoção de um trânsito mais seguro.

Figura 10 – Gráfico referente à necessidade 2



Fonte: Autores

8 CONSIDERAÇÕES FINAIS

Este estudo evidenciou o papel estratégico do Business Intelligence (BI) e do Data Warehousing na análise de acidentes de trânsito, ao transformar grandes volumes de dados em informações úteis para a formulação de políticas públicas. A aplicação dessas ferramentas permitiu identificar padrões relevantes, como a alta incidência de acidentes com motociclistas no Nordeste e com jovens do sexo masculino em todo o país.

A modelagem multidimensional, aliada à visualização de dados, facilitou a compreensão de dinâmicas complexas, como a influência do tipo de veículo, faixa etária e condições climáticas na gravidade dos acidentes. Esses achados apontam para intervenções práticas, como campanhas educativas segmentadas, reforço na fiscalização e melhorias na infraestrutura viária.

Apesar dos avanços, o estudo apresenta limitações relacionadas à qualidade e abrangência dos dados. A inclusão de variáveis socioeconômicas e comportamentais pode enriquecer futuras análises. Além disso, recomenda-se a aplicação de projetos-piloto para validar os insights obtidos antes de ações em larga escala.

Conclui-se que a integração entre tecnologia e gestão pública é promissora para a segurança no trânsito. O uso de BI e Data Warehouse não apenas qualifica o planejamento de ações, como também contribui para a redução de acidentes e para a construção de cidades mais seguras e sustentáveis.

9 REFERÊNCIAS

ABREU, J. R.; SOUZA, G. C. Análise preditiva e visualização interativa: uma aplicação em segurança viária. **Revista de Engenharia de Tráfego e Transportes**, v. 29, n. 4, p. 34-50, 2022.

ALMEIDA, R. P.; OLIVEIRA, T. A. Análise de dados de acidentes de trânsito por meio de Business Intelligence. **Revista Brasileira de Engenharia de Transportes**, v. 7, n. 2, p. 45-60, 2023.

BRASIL. Ministério da Infraestrutura – Secretaria Nacional de Trânsito (SENATRAN). **Base de dados de acidentes de trânsito e estatísticas regionais**. Disponível em: https://dados.gov.br/dataset/acidentes-transito. Acesso em: 15 nov. 2024.

BRASIL. Ministério da Justiça e Segurança Pública – Departamento Nacional de Trânsito (DENATRAN). **Acidentes de trânsito no Brasil – Banco de dados nacional.** Disponível em: https://www.gov.br/infraestrutura/pt-br/assuntos/transito. Acesso em: 16 nov. 2024.

CHEN, C.; MAO, H. Visualizing traffic accident patterns with spatial-temporal analysis tools and machine learning techniques. **Journal of Transport Geography**, v. 92, p. 102890, 2021. DOI: 10.1016/j.jtrangeo.2021.102890.

FERNANDES, A. C.; MOURA, T. S.; LIMA, R. M. Integração de dados para análise de acidentes de trânsito urbanos utilizando data warehousing. R**evista de Engenharia e Pesquisa Aplicada, v. 8, n. 1, p. 22–37, 2020**.

KIMBALL, R.; ROSS, M. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. **3ª ed. New York: John Wiley & Sons**, 2013.

SANTOS, M. F.; FERREIRA, E. R. Utilização de ferramentas de BI para análise de acidentes de trânsito. **Anais do Congresso Brasileiro de Informática na Educação, 2022.** Disponível em: https://anais.cbie2022.org.br. Acesso em: 20 nov. 2024.

SILVA, M. R.; COSTA, J. L. Sistema de Business Intelligence para análise de acidentes em rodovias federais. R**evista Brasileira de Informática Aplicada, v. 13, n. 2, p. 45–59, 2021**.

WANG, Z.; LI, X.; XU, C. Big data analytics for traffic accident prediction: A review of current techniques and future directions. **Transportation Research Part C: Emerging Technologies, v. 144, p. 103822, 2023**.