CENTRO UNIVERSITÁRIO CHRISTUS

SISTEMAS DE INFORMAÇÃO

DAVI BARRETO

PRESERVAÇÃO DE PRIVACIDADE DE DADOS EM SEGURANÇA VIÁRIA NA CIDADE DE FORTALEZA/CE

FORTALEZA

2022

DAVI BARRETO

PRESERVAÇÃO DE PRIVACIDADE DE DADOS EM SEGURANÇA VIÁRIA NA CIDADE DE FORTALEZA/CE

Trabalho de Conclusão de Curso (TCC) apresentado ao curso de Sistemas de Informação do Centro Universitário Christus, como requisito parcial para obtenção do grau de bacharel em Sistemas de Informação.

Orientador: Prof. MSc. Felipe Timbó Brito

FORTALEZA

2022

Dados Internacionais de Catalogação na Publicação
Centro Universitário Christus - Unichristus
Gerada automaticamente pelo Sistema de Elaboração de Ficha Catalográfica do
Centro Universitário Christus - Unichristus, com dados fornecidos pelo(a) autor(a)

B273p Barreto, Davi.
PRESERVAÇÃO DE PRIVACIDADE DE DADOS EM SEGURANÇA VIÁRIA NA CIDADE DE FORTALEZA/CE : PRESERVAÇÃO DE PRIVACIDADE DE DADOS EM SEGURANÇA VIÁRIA / Davi Barreto. - 2022.
26 f. : il. color.

Trabalho de Conclusão de Curso (Graduação) - Centro Universitário Christus - Unichristus, Curso de Sistemas de Informação, Fortaleza, 2022.
Orientação: Prof. Me. Felipe Timbó Brito.
Coorientação: Prof. Me. David Kenned Ferreira Andrade Viana.

1. Privacidade de Dados. 2. Privacidade Diferencial. 3. Segurança Viária. I. Título.

CDD 005

DAVI BARRETO

PRESERVAÇÃO DE PRIVACIDADE DE DADOS EM SEGURANÇA VIÁRIA NA CIDADE DE FORTALEZA/CE

Trabalho de Conclusão de Curso (TCC) apresentado ao curso de Sistemas de Informação do Centro Universitário Christus, como requisito parcial para obtenção do grau de bacharel em Sistemas de Informação.

Aprovada em:

BANCA EXAMINADORA

Prof. MSc. Felipe Timbó Brito (Orientador)
Centro Universitário Christus (Unichristus)

Prof. MSc. David Kenned Ferreira Andrade Viana
Centro Universitário Christus (Unichristus)

Prof. Felipe Cavalcante Monteiro
Centro Universitário Christus (Unichristus)

AGRADECIMENTOS

Gostaria de agradecer a minha mãe, e a toda minha família que, com muito carinho e apoio, não mediram esforços para que eu chegasse até está etapa de minha vida.

RESUMO

Segurança viária está relacionada ao conjunto de medidas de monitoramento necessárias para auxiliar na prevenção de acidentes em uma determinada região. O compartilhamento de dados relacionados à segurança viária é fundamental para evidenciar as consequências dos sinistros de trânsito que todos os dias provocam vítimas. Consequentemente, esses dados podem servir como base para elaboração de políticas públicas a fim de evitar que novas fatalidades ocorram nas ruas e avenidas. Contudo, o compartilhamento de dados de segurança viária deve ser feito com cautela, visto que esse compartilhamento envolve informações sensíveis de indivíduos. A Lei Geral de Proteção de Dados - LGPD, determina que o compartilhamento de informações deva garantir o anonimato dos indivíduos que tiveram seus dados utilizados. Assim, caso algum indivíduo possa ser reidentificado em um compartilhamento de dados de segurança viária por alguma entidade, essa entidade não estará em conformidade com a LGPD e consequentemente poderá sofrer sansões. Este trabalho propõe uma estratégia, baseada em privacidade diferencial, para publicar dados de segurança viária da cidade de Fortaleza/CE. Nossos resultados mostram que o ruído introduzido pela técnica proposta, para garantir a privacidade dos indivíduos, é bem baixo, o que torna os dados bastante úteis para análise.

Palavras-chave: Privacidade de Dados. Privacidade Diferencial. Segurança Viária.

ABSTRACT

Road safety is related to the set of monitoring measures needed to help prevent accidents in a given region. Sharing data related to road safety is essential to highlight the consequences of traffic accidents that cause victims every day. Consequently, these data can serve as a basis for the elaboration of public policies in order to prevent new fatalities from occurring on the streets and avenues. However, sharing road safety data should be done with caution, as this sharing involves sensitive information of individuals. The General Data Protection Law - LGPD, determines that the sharing of information must guarantee the anonymity of the individuals who had their data used. Thus, if any individual can be re-identified in a sharing of road safety data by any entity, that entity will not be in compliance with the LGPD and consequently may suffer sanctions. This work proposes a strategy, based on differential privacy, to publish road safety data in the city of Fortaleza/CE. Our results show that the noise introduced by the proposed technique, to ensure the privacy of individuals, is very low, which makes the data very useful for analysis.

Keywords: Data Privacy. Differential Privacy. Road Safety.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ALGORITMOS

SUMÁRIO

1 INTRODUÇÃO

Segurança viária pode ser definida como um conjunto de métodos, ações e normas necessárias para a circulação segura de pessoas e veículos nas ruas e rodovias, a fim de prevenir e reduzir o risco de acidentes de trânsito. Ela é fundamental para auxiliar na prevenção de acidentes em uma determinada região (MINISTÉRIO DA INFRAESTRUTURA, 2022).

De acordo com a Organização Mundial da Saúde, estima-se que 1,35 milhões de pessoas morrem devido a sinistros de trânsito por ano (ORGANIZATION, 2018). Isso representa uma média de 1 morte a cada 25 segundos no mundo. Para indivíduos com idade entre 5 e 29 anos, sinistros de trânsito já são a principal causa da morte. Consequentemente, tais indicadores mostram que o tema é uma emergência mundial.

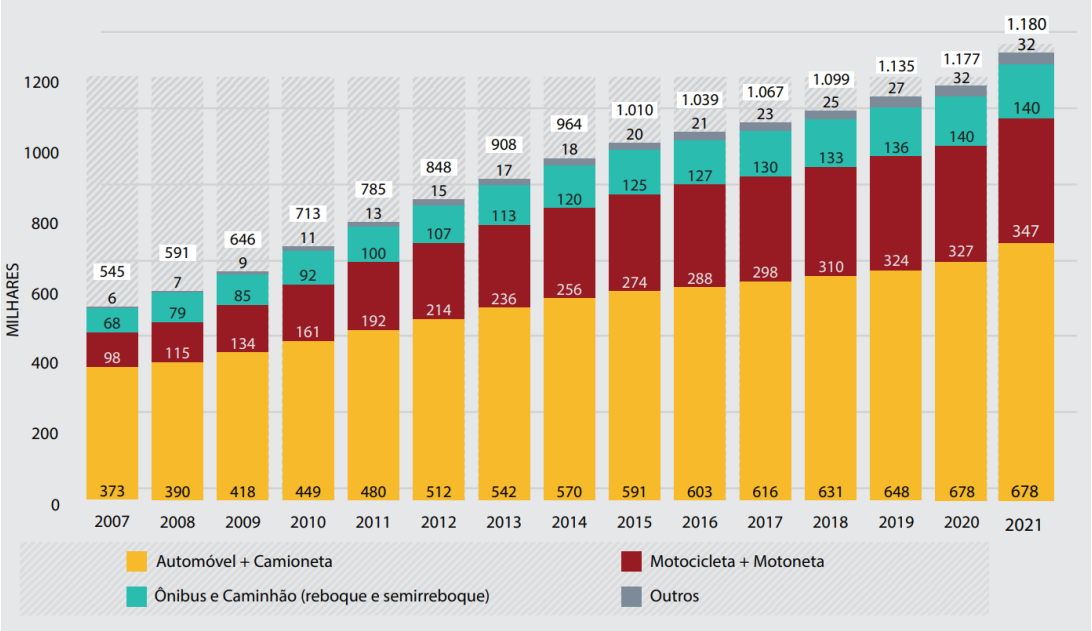
Já no Brasil, em 2020, 32.716 pessoas morreram em decorrência de sinistros de transporte terrestre. Isso representa 15,4 pessoas a cada 100.000 habitantes. Tal indicador é bem maior do que países da América Latina, como Argentina (12,3) e Chile (10,5), sendo quase seis vezes a taxa da Suécia (2,59) (ORGANIZATION, 2018). Para indivíduos com idade entre 5 a 14 anos, sinistros de trânsito são a principal causa de morte, e a segunda causa para pessoas entre 15 a 49 anos (MINISTÉRIO DA SAÚDE, 2022).

1.1 Contextualização e delimitação do tema

A cidade de Fortaleza, capital do Ceará, é a quinta maior cidade do país, com uma população estimada, em 2021, de 2,7 milhões de habitantes e uma área total de 313,8 km^2 (IBGE, 2022). Fortaleza possui aproximadamente 4,4 mil km de malha viária e uma frota de veículos estimada em 1,18 milhões, em dezembro de 2021 (DETRAN/CE, 2022). Nos últimos 10 anos, a cidade registrou um aumento de 65% da frota de veículos, enquanto para veículos de 2 ou 3 rodas (motocicletas, ciclomotores e motonetas) esse aumento foi de um pouco mais que o dobro para o mesmo período. A Figura 1 mostra a evolução da frota de veículos da cidade de Fortaleza nos últimos anos.

A prefeitura de Fortaleza tem dedicado esforços para construir um trânsito cada vez mais seguro, investindo em ações educativas, planejamento urbano e fiscalização preventiva. Para isso, a prefeitura utiliza dados do Sistema de Informação de Sinistros de Trânsito de Fortaleza - SIST, que integra informações coletadas pelos agentes de trânsito da AMC e as consolida com outras informações provenientes dos seguintes órgãos: Coordenadoria Integrada de Operações de

Figura 1 – Evolução da frota de veículos da cidade de Fortaleza.



Fonte: (PREFEITURA DE FORTALEZA, 2022)

Segurança - CIOPS; Perícia Forense do Ceará - PEFOCE; Instituto Dr. José Frota - IJF; Polícia Rodoviária Estadual do Ceará - PRE; Polícia Rodoviária Federal - PRF; Serviço de Atendimento Móvel de Urgência - SAMU e o Sistema de Informações de Mortalidade - SIM gerenciado pela Secretaria Municipal de Saúde - SMS (PREFEITURA DE FORTALEZA, 2022).

O SIST foi implementado em 2000 diante da necessidade de elaboração de estatísticas confiáveis para guiar o desenvolvimento de medidas a fim de reduzir o número de mortes no trânsito de Fortaleza. Atualmente, as estatísticas referentes à segurança viária da cidade de Fortaleza estão consolidadas no sistema VIDA (PREFEITURA DE FORTALEZA, 2022). O sistema é uma ferramenta fundamental para o acompanhamento da evolução dos sinistros de trânsito e a caracterização da problemática da violência do trânsito na cidade de Fortaleza.

1.2 Problematização

O compartilhamento de dados relacionados à segurança viária é fundamental para possibilitar um amplo conhecimento sobre a evolução dessa problemática na cidade de Fortaleza. Consequentemente, esses dados são úteis para a elaboração de políticas de prevenção de sinistros de trânsito, e visam evitar que novas fatalidades ocorram nas ruas e avenidas da cidade. Contudo, o compartilhamento de dados de segurança viária deve ser feito com cautela, uma vez que esse compartilhamento envolve informações sensíveis de indivíduos. A Lei Geral de Proteção de Dados - LGPD, determina que o compartilhamento de informações deva garantir

o anonimato dos indivíduos que tiveram seus dados utilizados. Assim, caso algum indivíduo possa ser reidentificado em um compartilhamento de dados de segurança viária pela prefeitura de Fortaleza, a mesma pode não estar em conformidade com a lei e consequentemente poderá sofrer sansões. Portanto, este trabalho visa auxiliar órgãos e entidades, como prefeituras e autarquias, a disponibilizarem dados de maneira privada sem violar a privacidade dos indivíduos que tiveram seus dados utilizados e ao mesmo tempo garantir que esses dados ainda sejam úteis para análises, tais como pesquisas científicas, busca de padrões e transparência.

1.3 Pressupostos

A privacidade diferencial (DWORK, 2006) é uma técnica de garantia de privacidade que visa publicar resultados de consultas por meio da adição de ruído. Em outras palavras, dada a resposta de uma consulta original sobre um conjunto de dados, um valor aleatório (ruído) é adicionado à essa resposta a fim de mascarar seu resultado. Este trabalho tem como pressuposto que a técnica de privacidade diferencial pode ser adotada para preservação da privacidade de indivíduos contidos em dados de segurança viária, em particular, da cidade de Fortaleza/CE.

1.4 Objetivos

O objetivo geral e os objetivos específicos deste trabalhos são descritos a seguir.

1.4.1 Objetivo geral

Este trabalho tem como principal objetivo principal aplicar a técnica de privacidade diferencial em dados de segurança viária da cidade de Fortaleza, a fim de preservar a privacidade dos indivíduos contidos nesses dados e, ao mesmo tempo, garantir que os dados permaneçam úteis para análises.

1.4.2 Objetivos específicos

Em particular, os objetivos específicos deste trabalho são:

- Aplicar o mecanismo de Laplace, oriundo da privacidade diferencial, nas contagens de veículos e de severidade na cidade de Fortaleza, no ano de 2021.
- Avaliar métricas de erro médio para garantir que dados ainda são úteis após aplicação do mecanismo proposto.

1.5 Justificativa

A *Global Road Safety Partnership* é uma organização sem fins lucrativos, fundada em 1999, que tem por objetivo criar e apoiar parcerias de segurança viária de primeira linha em países e comunidades em todo o mundo (GRSF, 2022). Em 2015, a cidade de Fortaleza foi uma das dez cidades contempladas com o projeto, e hoje conta com uma equipe de técnicos especializados, além de uma rede internacional de organizações, que dão suporte às ações do poder público municipal em melhorias no gerenciamento de dados, infraestrutura, fiscalização, educação e comunicação. Entretanto, a divulgação dos dados de segurança viária por parte da prefeitura de Fortaleza é realizada sem nenhum mecanismo de garantia de privacidade, o que pode acarretar em violações de privacidade dos indivíduos.

1.6 Estrutura do trabalho

A pesquisa está dividida em cinco capítulos, abordando o assunto de forma sucinta e direta. O Capítulo 2 apresenta o referencial teórico para este trabalho. O capítulo 3 detalha a metodologia utilizada. Já o capítulo 4 mostra os resultados obtidos. Por fim, o Capítulo 5 apresenta a conclusão e pontua os trabalhos futuros.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta conceitos relacionados a segurança viária, sinistros de trânsito, privacidade de dados e privacidade diferencial.

2.1 Segurança Viária

Segurança viária se caracteriza pelo conjunto de disposições, medidas e normas para garantir a segurança física de pedestres, ciclistas e motoristas em geral nas ruas e rodovias de uma determinada região (MINISTÉRIO DA INFRAESTRUTURA, 2022). Seu objetivo é evitar acidentes que levam a ferimentos sérios e óbitos oriundos do sinistros de trânsito.

2.1.1 Sinistros de trânsito

Recentemente, o termo "acidentes de trânsito"foi substituído pelo termo "sinistros de trânsito". Em 2020, a Associação Brasileira de Normas Técnicas (ABNT) oficializou a terminologia "sinistros de trânsito"a partir da NBR 10697. O termo acidente, por definição, destina-se a algo imprevisto ou fortuito. Dessa forma, a manutenção desse termo sugere que as mortes no trânsito não são evitáveis e nada poderia ser feito para evitá-las. Contudo, a mudança desse termo é importante para induzir a atitude de gestores, lideranças e da sociedade em geral sobre o problema da segurança viária e disseminar que mortes e ferimentos no trânsito são sim, evitáveis.

2.1.2 Segurança viária em Fortaleza

Os métodos de monitoramento em segurança das vias têm como objetivo garantir a segurança dos indivíduos, além de auxiliar na prevenção de acidentes por meio da coleta de dados de tráfego. A cidade de Fortaleza, especificamente, vem seguindo boas práticas internacionais de monitoramento e videomonitoramento das vias urbanas, tendo em vista a diminuição de acidentes fatais e óbitos (PREFEITURA DE FORTALEZA, 2022).

As mortes por sinistros de transporte terrestre, em 2016, representavam a 6ª causa de morte para os fortalezenses. Em 2019, foi a 15ª causa. No ano de 2021, não apareceu na lista das 15 principais causa de morte do munícipio (MINISTÉRIO DA SAÚDE, 2022). A Figura 2 mostra o ranking das principais causas de morte na cidade de Fortaleza. Infelizmente,

Figura 2 – Ranking das principais causas de morte na cidade de Fortaleza.

ORDEM*	2016	2017	2018	2019	2020	2021***
1ª	AVC	HOMICÍDIOS	HOMICÍDIOS	AVC	COVID-19	COVID-19
2ª	HOMICÍDIOS	AVC	AVC	PNEUMONIAS	HOMICÍDIOS	AVC
3ª	PNEUMONIAS	PNEUMONIAS	PNEUMONIAS	INFARTO AGUDO DO MIOCÁRDIO	AVC	PNEUMONIAS
4ª	INFARTO AGUDO DO MIOCÁRIO	INFARTO AGUDO DO MIOCÁRIO	INFARTO AGUDO DO MIOCÁRIO	HOMICÍDIOS	PNEUMONIAS	INFARTO AGUDO DO MIOCÁRDIO
5ª	CÂNCER DE BRÔNQUIOS E PULMÕES	CÂNCER DE BRÔNQUIOS E PULMÕES	CÂNCER DE BRÔNQUIOS E PULMÕES	CÂNCER DE BRÔNQUIOS E PULMÕES	INFARTO AGUDO DO MIOCÁRDIO	HOMICÍDIOS
6ª	SINISTROS DE TRÂNSITO	DIABETES MELLITUS	ALZHEIMER	ALZHEIMER	DIABETES MELLITUS	ALZHEIMER
7ª	DIABETES MELLITUS	ALZHEIMER	DIABETES MELLITUS	DIABETES MELLITUS	ALZHEIMER	DIABETES MELLITUS
8ª	OUTRAS DPOC	OUTRAS DPOC	OUTRAS DPOC	OUTRAS DPOC	CÂNCER DE BRÔNQUIOS E PULMÕES	CÂNCER DE BRÔNQUIOS E PULMÕES
9ª	ALZHEIMER	SINISTROS DE TRÂNSITO	CARDIOMIOPATIAS	DOENÇA ISQUÊMICA CRÔNICA DO CORAÇÃO	CÂNCER DE MAMA	PARADA CARDÍACA
10ª	CÂNCER DE MAMA	CÂNCER DE MAMA	CÂNCER DE MAMA	CARDIOMIOPATIAS	DOENÇA ISQUÊMICA CRÔNICA DO CORAÇÃO	DISPARO DE ARMA DE FOGO
11ª	INSUFICIÊNCIA CARDÍACA	INSUFICIÊNCIA CARDÍACA	DOENÇA ISQUÊMICA CRÔNICA DO CORAÇÃO	CÂNCER DE MAMA	PARADA CARDÍACA	INSUFICIÊNCIA CARDÍACA
12ª	CÂNCER DE ESTÔMAGO	DOENÇA ISQUÊMICA CRÔNICA DO CORAÇÃO	SINISTROS DE TRÂNSITO	CÂNCER DE ESTÔMAGO	OUTRAS DPOC	CÂNCER DE MAMA
13ª	DOENÇA ISQUÊMICA CRÔNICA DO CORAÇÃO	QUEDA	INSUFICIÊNCIA CARDÍACA	INSUFICIÊNCIA CARDÍACA	INSUFICIÊNCIA CARDÍACA	DOENÇA ISQUÊMICA CRÔNICA DO CORAÇÃO
14ª	OUTRA SEPTICEMIAS	PARADA CARDÍACA	CÂNCER DE ESTÔMAGO	DEMÊNCIA	SINISTROS DE TRÂNSITO	CÂNCER DE ESTÔMAGO
15ª	QUEDA	CÂNCER ESTÔMAGO	QUEDA	SINISTROS DE TRÂNSITO	DOENÇA CARDÍACA HIPERTENSIVA	DOENÇA CARDÍACA HIPERTENSIVA

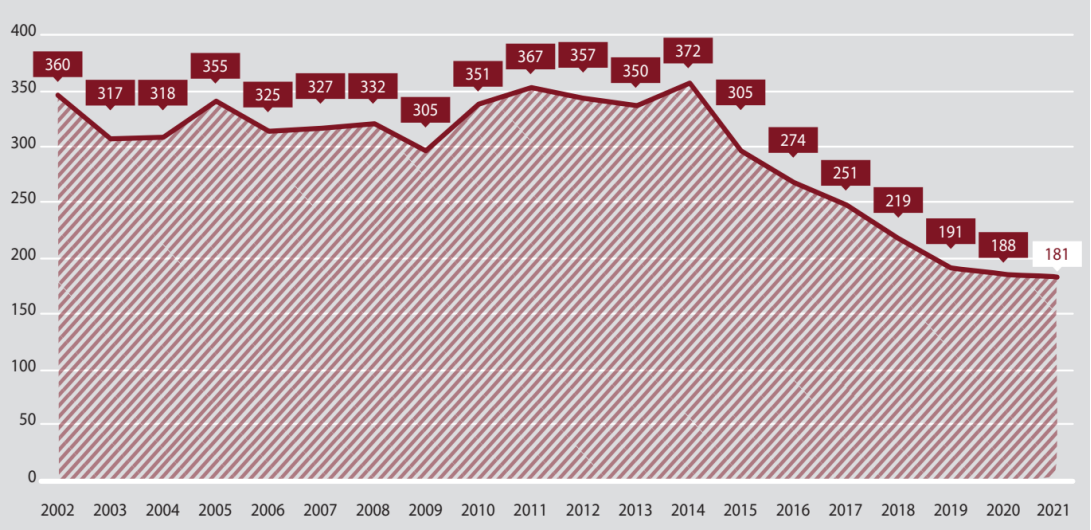
Fonte: Secretaria Municipal de Saúde (SMS, 2020)

esta evolução na cidade de Fortaleza não segue a tendência mundial. A Organização Mundial da Saúde prevê que, até 2050, as lesões por sinistros de trânsito serão a 5ª principal causa de morte para a população global, caso políticas públicas para a segurança no trânsito não sejam implementadas.

Em 2021, a Autarquia Municipal de Trânsito de Fortaleza - AMC, registrou 10.130 sinistros de trânsito com vítimas, sendo 1.110 atropelamentos, onde o principal veículo atropelador foi a motocicleta, aparecendo em 41,2% dos casos. Os meses com mais ocorrências com vítimas foram dezembro e janeiro. Semanalmente, os dias que apresentaram maior número de ocorrências foram o sábado, seguido da sexta-feira. As faixas horárias que concentram maior número de ocorrências com vítimas são de 17h às 20h e de 7h às 9h (PREFEITURA DE FORTALEZA, 2022).

Entre 2020 e 2021, houve uma redução de 4,26% no número de sinistros com vítimas fatais. Em relação aos sinistros com vítimas, observou-se um aumento de 3,5%. Esse crescimento pode ter relação com o aumento do fluxo de veículos, ao final do período de restrição de circulação, adotado para prevenir a proliferação do vírus da Covid-19 (PREFEITURA DE FORTALEZA, 2022). Houveram também medidas de restrições de circulação em 2021, porém, bem menos intensas as de 2020. Com mais veículos circulando, espera-se um aumento nas

Figura 3 – Número de sinistro com vítimas fatais na cidade de Fortaleza nos últimos 20 anos.



Fonte: (PREFEITURA DE FORTALEZA, 2022)

ocorrências de sinistros. A Figura 3 mostra o número de sinsitros com vítimas fatais na cidade de Fortaleza nos últimos 20 anos.

2.2 Privacidade de dados

A preocupação com a privacidade de indivíduos remonta os primórdios da história, com a separação do ambiente público e o privado e as delegações de o que deveria, e poderia, ser feito em cada espaço. O mesmo ocorre com nossos dados pessoais: o que deve ser, ou não, compartilhado?

A privacidade de dados está associada às regras de controle de acesso efetivas, que permitem a revelação da informação apenas por usuários autorizados. Contudo, a privacidade não está garantida apenas com o controle de acesso eficiente, visto que os usuários com acesso àquelas informações podem ser maliciosos, mas sim com a capacidade de divulgar informações pessoais acerca de indivíduos sem reidentificá-los (BRITO; MACHADO, 2017).

2.2.1 Modelos tradicionais de anonimização

Diante da perda de controle sobre dados compartilhados, várias técnicas foram desenvolvidas para proteger a privacidade de indivíduos. Uma das formas de proteção é a anonimização de dados, que consiste em modificar os dados originais de tal forma que, os dados anonimizados não se assemelham aos dados originais, mas ambos possuem semântica e sintaxe semelhantes (BRITO; MACHADO, 2017).

Uma técnica convencional para anonimizar dados é a remoção dos identificadores explícitos de indivíduos, como nome, CPF, e-mail, etc. do conjunto de dados, antes de qualquer compartilhamento. Contudo, essa estratégia não é suficiente para proteger a privacidade dos indivíduos (SWEENEY, 2002). Consequentemente, várias técnicas foram propostas para superar esse problema. *K-anonimity* (SWEENEY, 2002) é um dos modelos de privacidade mais conhecidos que consistem em formar classes de registros de tamanho k. Em uma classe, cada registro é idêntico aos outros registros $(k-1)$. Em outras palavras, cada registro não pode ser vinculado a um indivíduo com probabilidade inferior a $1/k$.

A partir do *k-anonimity*, outros modelos de privacidade foram propostos para evitar a reidentificação de indivíduos em publicações de conjuntos de dados, tais como *l-diversity* (MACHANAVAJJHALA *et al.*, 2007), *t-closeness* (LI *et al.*, 2007; LI *et al.*, 2009) e δ*-presence* (NERGIZ *et al.*, 2007). No entanto, todas essas abordagens assumem que um usuário malicioso possui conhecimento limitado, o que não é verdadeiro em situações do mundo real. Consequentemente, muitos ataques foram desenvolvidos para esses modelos de privacidade tradicionais (GANTA *et al.*, 2008; CORMODE *et al.*, 2010; JIN *et al.*, 2010; XIAO *et al.*, 2010; WONG *et al.*, 2011). Dessa forma, é necessária uma técnica mais robusta para garantir que a privacidade dos indivíduos não seja violada, e que os dados ainda permaneçam úteis para analistas e pesquisadores.

2.2.2 Privacidade Diferencial

A privacidade diferencial (DWORK, 2006) é um modelo matemático proposto para prover garantias de privacidade para indivíduos em um conjunto de dados compartilhado. Essa técnica tem sido aplicada na indústria por grandes empresas, tais como Apple (TEAM, 2017), Google (ERLINGSSON *et al.*, 2014) e Uber (JOHNSON *et al.*, 2018).

A privacidade diferencial visa publicar resultados de consultas por meio da adição de ruído. Esse ruído é introduzido aleatoriamente à resposta de uma determinada consulta por um mecanismo. A definição formal de privacidade diferencial é apresentada a seguir:

Definição 1 *(DWORK, 2006) Um mecanismo M garante o* ε*-Privacidade Diferencial se, para todos os conjuntos de dados vizinhos D e* D'*, i.e., que diferem em no máximo um elemento (registro), e para todo O contido no conjunto de todos os resultados possíveis provenientes de M, isto é, para todo* $O \subseteq Imagem(M)$*:*

$$Pr[M(D) \in O] \leq exp(\varepsilon) \times Pr[M(D') \in O],$$

onde Pr é a probabilidade sobre a aleatoriedade do mecanismo M.

Com base na definição acima, o conceito de sensibilidade de uma consulta mede o impacto de um indivíduo no resultado da consulta ao ser removido ou inserido no conjunto de dados.

Definição 2 *(DWORK, 2006) Considere $\mathscr{D}$ o domínio de todos os conjuntos de dados e f uma função de consulta que mapeia conjuntos de dados a vetores de números reais. A sensibilidade da função f é dada por:*

$$\Delta f = \max_{D,D' \in \mathscr{D}} \| f(D) - f(D') \|_1$$

para todo D e D' diferindo vizinhos.

Vale ressaltar que a sensibilidade Δf para consultas de contagem é igual a 1, visto que a adição ou remoção de qualquer registro do conjunto de dados altera qualquer contagem em no máximo uma unidade, independente do dado de entrada. A adição e a remoção simulam a presença ou a ausência de qualquer indivíduo no dado.

Os mecanismos mais comuns existentes na literatura para garantir a privacidade diferencial são o Exponencial (MCSHERRY; TALWAR, 2007) e o de *Laplace* (DWORK *et al.*, 2014). O mecanismo Exponencial é aplicado sobre consultas não numéricas, enquanto que o de *Laplace*, que é empregado neste trabalho, atua sobre consultas numéricas.

No mecanismos de Laplace, o ruído a ser adicionado se baseia na geração de uma variável aleatória x, a qual segue uma distribuição de *Laplace* com média μ e escala b, através da fórmula:

$$Laplace(x|\mu,b) = \frac{1}{2b}\exp(-\frac{|x-\mu|}{b})$$

Assim, o mecanismo de *Laplace* é definido formalmente por:

Definição 3 *Dada uma função de consulta $f : D \to \Re$, o mecanismo de Laplace M:*

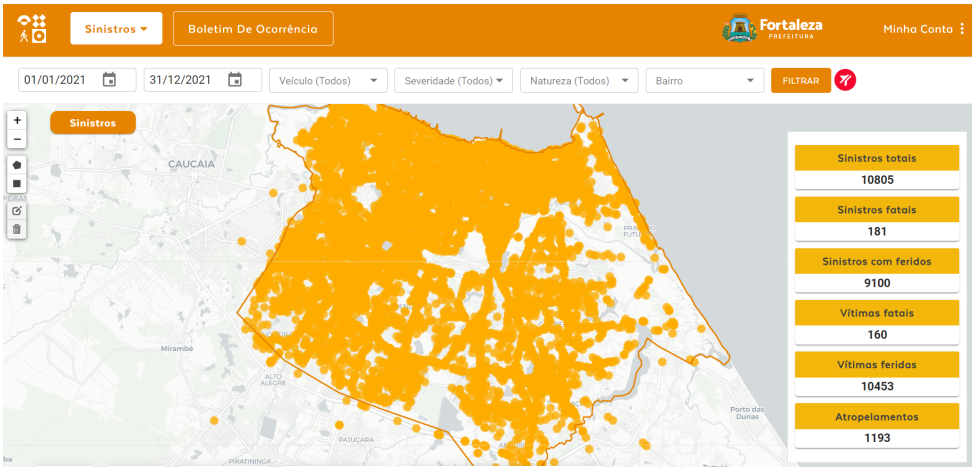
$$M = f(D) + Laplace(0, \frac{\Delta f}{\varepsilon})$$

fornece ε*-Privacidade Diferencial. Onde* $Laplace(0, \frac{\Delta f}{\varepsilon})$ *retorna uma variável aleatória da distribuição de Laplace com média zero e escala* $\frac{\Delta f}{\varepsilon}$.

3 METODOLOGIA

Este capítulo apresenta a metodologia utilizada para o desenvolvimento deste trabalho. Inicialmente, os dados de sinistro de trânsito da cidade de Fortaleza são coletados da plataforma VIDA (PREFEITURA DE FORTALEZA, 2022). A Figura 4 mostra o funcionamento dessa plataforma.

Figura 4 – Exemplo de utilização do sistema VIDA.



Fonte: (PREFEITURA DE FORTALEZA, 2022)

São filtrados sinistros de 01/01/2021 a 31/12/2021, para todos os tipos de veículos, severidade e natureza. A Tabela 1 mostra as contagens dos sinistros de trânsito consultados por severidade.

Tabela 1 – Contagens dos sinistros de trânsito por severidade consultados entre 01/01/2021 a 31/12/2021.

Mortos	Feridos	Ilesos
181	11179	1663

Fonte: Elaborado pelo autor.

Já a Tabela 2 exibe as contagens dos sinistros de trânsito consultados por tipo de veículo. Os tipos de veículos existentes no conjunto de dados são: motocicleta; automóvel; micro-ônibus; caminhão; ônibus; ciclomotor; bicicleta; outros; trêm; e tração animal.

Tabela 2 – Contagens dos sinistros de trânsito por tipo de veículo consultados entre 01/01/2021 a 31/12/2021.

Motoc	Autom	Mic-ôn	Camin	Ônib	Ciclom	Bicic	Outros	Trêm	Anim
8192	6697	18	280	241	29	696	772	4	2

Fonte: Elaborado pelo autor.

Em seguida, o mecanismo de Laplace é aplicado em cada contagem de sinistro de trânsito por severidade e por tipo de veículo. A sensibilidade Δf da consulta de contagem em questão é igual a 1, pois a adição ou remoção de qualquer registro do conjunto de dados altera a contagem em no máximo 1. Como duas consultas foram realizadas sobre o mesmo conjunto de dados, o orçamento de privacidade ε é dividido por dois. Dessa forma, para cada consulta de contagem, o ruído adicionado r é igual a:

$$r = \frac{\Delta f}{\varepsilon} = \frac{1}{\varepsilon/2} = \frac{2}{\varepsilon}.$$

O algoritmo 1 mostra o funcionamento da privacidade diferencial na adição de ruído de Laplace.

Algoritmo 1: Adição de ruído de Laplace

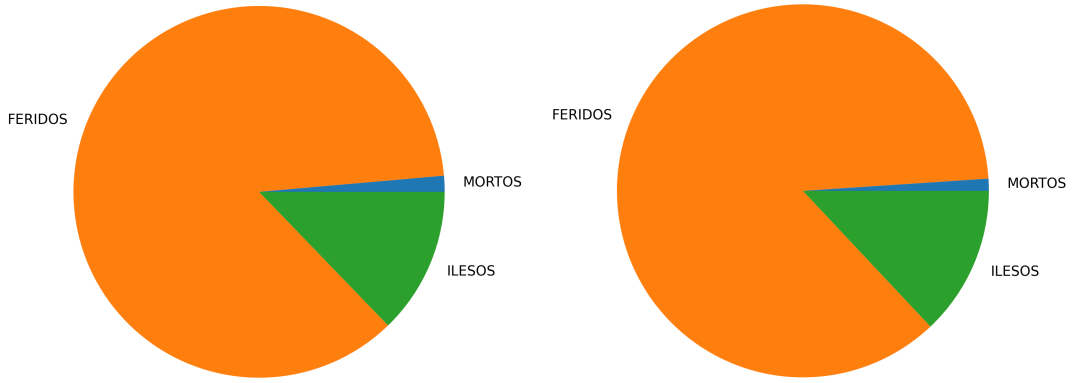
Entrada: Conjunto de dados D, Orçamento de privacidade ε
Saída: Contagens de sinistros de trânsito diferencialmente privadas S' (por severidade) e V' (por tipo de veículo)
início
  $S \leftarrow obter\_Contagens\_Originais\_Por\_Severidade(D)$;
  $V \leftarrow obter\_Contagens\_Originais\_Por\_Tipo\_Veiculo(D)$;
  $S' \leftarrow S + Laplace(\frac{2}{\varepsilon})$;
  $V' \leftarrow V + Laplace(\frac{2}{\varepsilon})$;
  retorne S' e V'
fim

4 RESULTADOS

Este capítulo apresenta os resultados obtidos para a aplicação da privacidade diferencial sobre os dados de segurança viária na cidade de Fortaleza. Os experimentos foram implementados na linguagem de programação Python, em um sistema operacional Windows, e uma máquina com 8GB de RAM e 4 cores de CPU.

A Figura 5 mostra a diferença entre a contagem original e a com ruído de Laplace, dos sinistros de trânsito por severidade, considerando $\varepsilon = 0.1$.

Figura 5 – Diferença entre a contagem original e a com ruído de Laplace, dos sinistros de trânsito por severidade, considerando $\varepsilon = 0.1$.



(a) Original (b) Com ruído

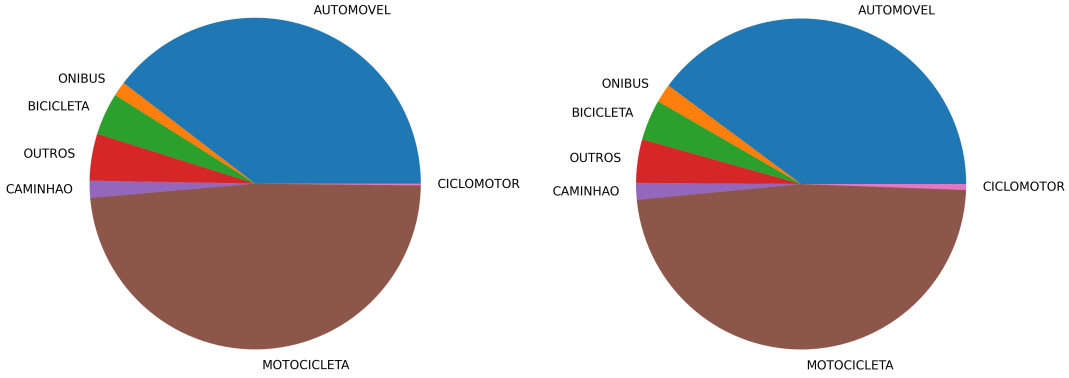
Fonte: Elaborado pelo autor.

Nota-se que em ambas as imagens (Figuras 5a e 5b) as contagens são bastante semelhantes. Em particular, para o ruído adicionado nesse exemplo, o número de feridos original passou de 11.179 para 11.208, enquanto o número de ilesos passou de 1.663 para 1.641 e o número de mortos passou de 181 para 174.

Já a Figura 6 mostra a diferença entre a contagem original e a com ruído de Laplace, dos sinistros de trânsito por tipo de veículo, considerando $\varepsilon = 0.1$.

Nota-se que as respostas não são muito diferentes mesmo com um valor de ε baixo (Figuras 6a e 6b), o que garante a utilidade dos dados mesmo após aplicação da privacidade diferencial. Particularmente, para o ruído aleatório adicionado nesse exemplo, o número de sinistros envolvendo motocicleta variou de 8.192 para 8.199. Envolvendo automóveis variou de 6.697 para 6.671. Envolvendo caminhões variou de 280 para 298. Envolvendo ônibus passou de 241 para 261. Envolvendo ciclomotores passou de 29 para 42. Já envolvendo bicicletas passou de 696 para 661. Finalmente, envolvendo outros veículos, a contagem passou de 772 para 741.

Figura 6 – Diferença entre a contagem original e a com ruído de Laplace, dos sinistros de trânsito por tipo de veículo, considerando $\varepsilon = 0.1$.

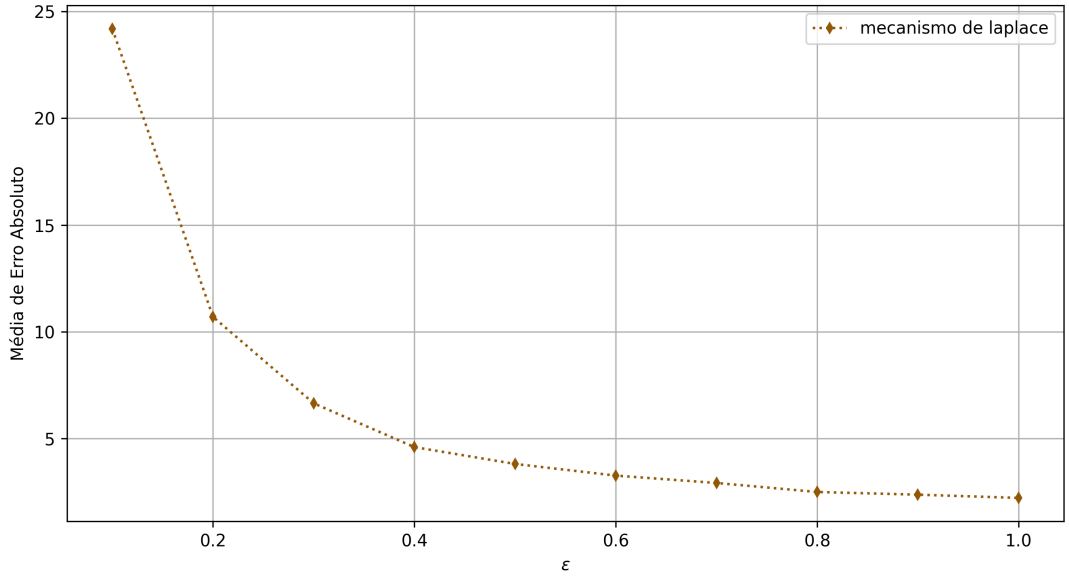


(a) Original (b) Com ruído

Fonte: Elaborado pelo autor.

A Figura 7 mostra o erro absoluto médio das contagens de severidade e de tipo de veículo para valores de ε variando entre 0.1 e 1.0. Foram realizadas 10 execuções e foi retirada a média entre elas.

Figura 7 – Erro absoluto médio das contagens de severidade e de tipo de veículo.



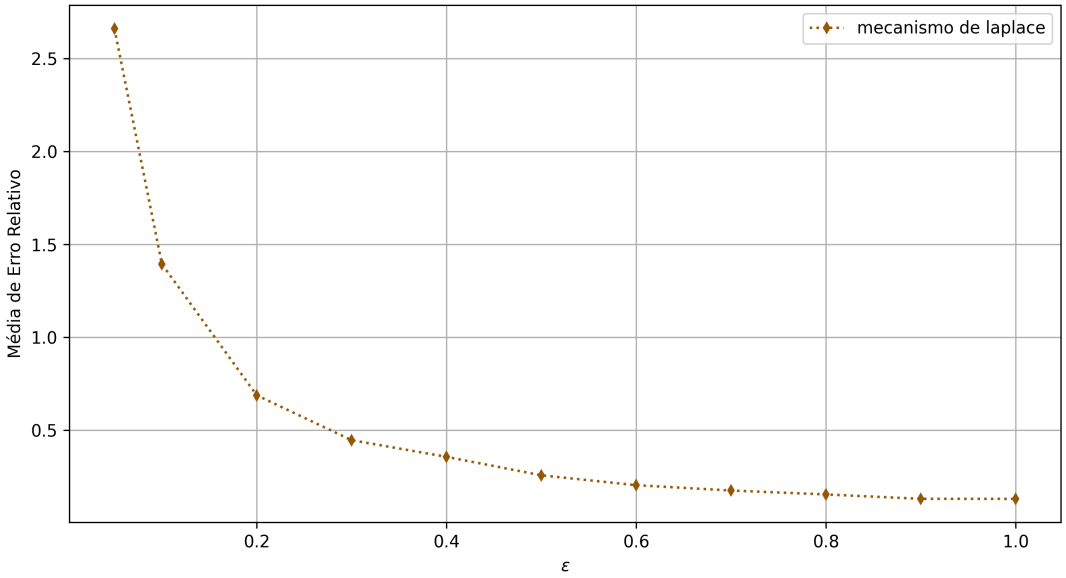
Fonte: Elaborado pelo autor.

Observa-se que para $\varepsilon = 0.1$, em média uma contagem varia por volta de 24 unidades. Já para $\varepsilon = 0.5$, a variação já cai bastante, para valores em torno de 4 unidades de contagem. Quando $\varepsilon = 1.0$ a variação é em torno de 2 unidades.

Por fim, a Figura 8 mostra o erro relativo médio das contagens de severidade e de tipo de veículo para valores de ε variando entre 0.1 e 1.0. Neste experimento, também foram

realizadas 10 execuções e foi retirada a média entre elas.

Figura 8 – Erro relativo médio das contagens de severidade e de tipo de veículo.



Fonte: Elaborado pelo autor.

Observa-se que para $\varepsilon = 0.1$, em média uma contagem varia por volta de 2.7 vezes o valor original. Já para $\varepsilon = 0.5$, a variação já cai bastante, para valores em torno de 0.3 vezes a contagem original. Quando $\varepsilon = 1.0$ a variação é em torno de 0.2 vezes a contagem original, o que torna o resultado bastante útil para analistas de dados e pesquisadores.

5 CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho apresentou uma estratégia, baseada em privacidade diferencial, para publicar dados de segurança viária da cidade de Fortaleza/CE sem violar a privacidade dos indivíduos pertencentes ao dado. Em particular, este trabalho visou publicar contagens de sinistros de trânsito por severidade e por tipo de veículo. Os resultados provenientes de avaliação experimental mostraram que o ruído introduzido pelo mecanismo de Laplace, para garantir a privacidade dos indivíduos, atingiu valores bem baixos devido à sensibilidade da consulta de contagem também ser baixa, isto é, igual a 1. Dessa forma, os dados, mesmo após perturbação, continuam sendo bastante úteis para analistas de dados e pesquisadores.

Como trabalhos futuros, pretende-se avaliar outros tipos de mecanismos, como o mecanismo Geométrico (GHOSH *et al.*, 2009), uma variante discreta do mecanismo de Laplace para consultas com respostas inteiras. Além disso, espera-se utilizar dados de anos anteriores da cidade de Fortaleza/CE para o desenvolvimento de um *Dashboard* de dados de segurança viária todo construído a partir de técnicas de privacidade diferencial.

REFERÊNCIAS

BRITO, F. T.; MACHADO, J. C. Preservação de privacidade de dados: Fundamentos, técnicas e aplicações. **Jornadas de A tualização em Informática**, 2017.

CORMODE, G.; SRIVASTAVA, D.; LI, N.; LI, T. Minimizing minimality and maximizing utility: analyzing method-based attacks on anonymized data. **Proceedings of the VLDB Endowment**, VLDB Endowment, v. 3, n. 1-2, p. 1045–1056, 2010.

DETRAN/CE. **Estatísticas da Frota de Veículos**. 2022. Disponível em: <https://www.detran.ce.gov.br/estatisticas/>. Acesso em: 08/11/2022.

DWORK, C. Differential privacy. In: SPRINGER. **International Colloquium on Automata, Languages, and Programming**. [S.l.], 2006. p. 1–12.

DWORK, C.; ROTH, A. *et al.* The algorithmic foundations of differential privacy. **Foundations and Trends® in Theoretical Computer Science**, Now Publishers, Inc., v. 9, n. 3–4, p. 211–407, 2014.

ERLINGSSON, Ú.; PIHUR, V.; KOROLOVA, A. Rappor: Randomized aggregatable privacy-preserving ordinal response. In: ACM. **Proceedings of the 2014 ACM SIGSAC conference on computer and communications security**. [S.l.], 2014. p. 1054–1067.

GANTA, S. R.; KASIVISWANATHAN, S. P.; SMITH, A. Composition attacks and auxiliary information in data privacy. In: **Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. [S.l.: s.n.], 2008. p. 265–273.

GHOSH, A.; ROUGHGARDEN, T.; SUNDARARAJAN, M. Universally utility-maximizing privacy mechanisms. In: **Proceedings of the forty-first annual ACM symposium on Theory of computing**. [S.l.: s.n.], 2009. p. 351–360.

GRSF. **Global Road Safety Facility**. 2022. Disponível em: <https://www.roadsafetyfacility.org/>. Acesso em: 08/11/2022.

IBGE. **Fortaleza - Panorama**. 2022. Disponível em: <https://cidades.ibge.gov.br/?codmun=230440>. Acesso em: 08/11/2022.

JIN, X.; ZHANG, N.; DAS, G. Algorithm-safe privacy-preserving data publishing. In: **Proceedings of the 13th International Conference on Extending Database Technology**. [S.l.: s.n.], 2010. p. 633–644.

JOHNSON, N.; NEAR, J. P.; SONG, D. Towards practical differential privacy for sql queries. **Proceedings of the VLDB Endowment**, VLDB Endowment, v. 11, n. 5, p. 526–539, 2018.

LI, N.; LI, T.; VENKATASUBRAMANIAN, S. t-closeness: Privacy beyond k-anonymity and l-diversity. In: IEEE. **2007 IEEE 23rd International Conference on Data Engineering**. [S.l.], 2007. p. 106–115.

LI, N.; LI, T.; VENKATASUBRAMANIAN, S. Closeness: A new privacy measure for data publishing. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 22, n. 7, p. 943–956, 2009.

MACHANAVAJJHALA, A.; KIFER, D.; GEHRKE, J.; VENKITASUBRAMANIAM, M. l-diversity: Privacy beyond k-anonymity. **ACM Transactions on Knowledge Discovery from Data (TKDD)**, ACM New York, NY, USA, v. 1, n. 1, p. 3–es, 2007.

MCSHERRY, F.; TALWAR, K. Mechanism design via differential privacy. In: IEEE. **Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on**. [S.l.], 2007. p. 94–103.

MINISTÉRIO DA INFRAESTRUTURA. **Segurança Viária**. 2022. Disponível em: <https://www.gov.br/infraestrutura/pt-br/assuntos/transporte-terrestre/rodovias-federais/seguranca-viaria>. Acesso em: 08/11/2022.

MINISTÉRIO DA SAÚDE. **DATASUS Brasil**. 2022. Disponível em: <http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sim/cnv/ext10uf.def>. Acesso em: 08/11/2022.

NERGIZ, M. E.; ATZORI, M.; CLIFTON, C. Hiding the presence of individuals from shared databases. In: **Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data**. [S.l.: s.n.], 2007. p. 665–676.

ORGANIZATION, W. H. Global status report on road safety 2018: Summary (no. who/nmh/nvi/18.20). In: . [S.l.: s.n.], 2018.

PREFEITURA DE FORTALEZA. **Plataforma VIDA**. 2022. Disponível em: <https://vida.centralamc.com.br/>. Acesso em: 08/11/2022.

SMS. **Secretaria Municipal de SaÚde de Fortaleza**. 2020. Disponível em: <SMS/CEVEPI/SistemadeInformaÃğÃčodeMortalidade>. Acesso em: 08/11/2022.

SWEENEY, L. k-anonymity: A model for protecting privacy. **International journal of uncertainty, fuzziness and knowledge-based systems**, World Scientific, v. 10, n. 05, p. 557–570, 2002.

TEAM, A. D. P. Learning with privacy at scale. In: . [S.l.: s.n.], 2017.

WONG, R. C.-W.; FU, A. W.-C.; WANG, K.; YU, P. S.; PEI, J. Can the utility of anonymized data be used for privacy breaches? **ACM Transactions on Knowledge Discovery from Data (TKDD)**, ACM New York, NY, USA, v. 5, n. 3, p. 1–24, 2011.

XIAO, X.; TAO, Y.; KOUDAS, N. Transparent anonymization: Thwarting adversaries who know the algorithm. **ACM Transactions on Database Systems (TODS)**, ACM New York, NY, USA, v. 35, n. 2, p. 1–48, 2010.