ERICK DE OLIVEIRA

HUMBERTO DOS SANTOS MADEIRA

PEDRO AUGUSTO MIGLIARI MONTEIRO

A LEI GERAL DE PROTEÇÃO DE DADOS PESSOAIS E A ANONIMIZAÇÃO DE DADOS: UMA APLICAÇÃO DA TÉCNICA EM UMA BASE DE DADOS REAL

São Caetano Do Sul – São Paulo

2020

ERICK DE OLIVEIRA

HUMBERTO DOS SANTOS MADEIRA

PEDRO AUGUSTO MIGLIARI MONTEIRO

A Lei Geral de Proteção de Dados Pessoais e a Anonimização de Dados: Uma Aplicação da Técnica em uma Base de Dados Real

Trabalho de Conclusão de Curso apresentado à Faculdade de Tecnologia de São Caetano do Sul, sob a orientação do Prof. Doutor. Ricardo Baitz, como requisito parcial para a obtenção do diploma de Graduação no Curso de Segurança da Informação

São Caetano Do Sul – São Paulo

2020

Dedicamos esse trabalho a nossos professores que nos apoiaram ,nos guiaram e incentivaram a fazê-lo.

AGRADECIMENTOS

Agradecemos aos nossos professores que sempre nos mostraram o real caminho a ser seguido, através do esforço e do conhecimento, agradecemos também por sempre nos auxiliarem quando necessário, em especial ao nosso orientador Mestre Ricardo Baitz, por nos incentivar a realizar essa pesquisa.

“*A nova fonte de poder não é o dinheiro nas mãos de poucos, mas informação nas mãos de muitos.*”

John Naisbitt

RESUMO

DE OLIVEIRA, Erick.; MADEIRA, Humberto dos S.; MONTEIRO, Pedro A. M. A Lei Geral de Proteção de Dados Pessoais e a Anonimização de Dados: Aplicação eu uma Base de Dados Real 48f. Trabalho de Graduação – Faculdade de Tecnologia de São Caetano do Sul, São Caetano do Sul, 2020.

O presente estudo propõe explorar a técnica de anonimização nos dados publicados no cadastro imobiliário fiscal relativos ao Imposto Predial e Territorial Urbano – IPTU, do município de São Paulo. Em um primeiro momento será explorada a Lei Geral de Proteção de Dados – LGPD (Lei 13.709/2018), por seguinte, a pesquisa aprofundará as formas de proteção de dados, mais especificamente, na técnica anonimização, caracterizada por ocultar o dado totalmente ou parcialmente (pseudonimização). Por meio de ferramentas a base de dados cadastrais do site da prefeitura de São Paulo será manuseada para inferir se tais dados estão ou podem ser anonimizados ou pseudonimizados.

Palavras-chave: LGPD; anonimização; pseudoanonimização; IPTU.

ABSTRACT

DE OLIVEIRA, Erick.; MADEIRA, Humberto dos S.; MONTEIRO, Pedro A. M. The General Data Protection Law and the Data Anonymization: Application in a Real Data Base 48f. Trabalho de Graduação – Faculdade de Tecnologia de São Caetano do Sul, São Caetano do Sul, 2020

The present study aims to explore the anonymization technique on the data published in the tax register relative to the Predial Tax and Urban Territorial - IPTU, from São Paulo city.At first, the General Data Protection Law - LGPD (Law 13.709/2018) will be explored, followed by an indepth research on data protection forms, more specific on the anonymization technique, characterized by concealling the data entirely or patially (pseudonymization).Through the use of tools, the registration database from the website from São Paulo city hall will be handled to verify If such data are or can by anonymized or pseudonymized.

Key Words: LGPD; anonymization; pseudonymization; IPTU.

LISTA DE ILUSTRAÇÕES

LISTA DE ABREVIATURAS E SIGLAS

CEP	Código de Endereçamento Postal
CNPJ	Cadastro Nacional da Pessoa Jurídica
CPF	Cadastro de Pessoas Físicas
CPU	*Central Process Unit*
CSV	*Comma-separated values*
GB	*Gigabyte*
GDPR	*General Data Protection Regulation*
GHz	*Gigahertz*
IPTU	Imposto Predial e Territorial Urbano
JSON	*JavaScript Object Notation*
LGPD	Lei Geral de Proteção de Dados
LTS	*Long Term Supported*
RG	Registro Geral
RAM	*Random Access Memory*
SPL	*Search Processing Language*
SQL	*Structured Query Language*
TB	*Terabyte*

SUMÁRIO

INTRODUÇÃO

Com a chegada da LGPD e sua obrigatoriedade em todo o território brasileiro, é de suma importância que todos os afetados em sua abrangência estejam preparados a adaptar-se às suas exigências e com ajuda de seu estudo, termos uma melhor compreensão a respeito da importância de sua aplicação. A lei propõe mudanças grandes para companhias que estão acostumadas com o trabalho feito de uma forma específica e antiga, ainda mais para as que não possuem uma área de segurança da informação devidamente preparada para proteger seus dados de seus clientes.

É necessário que toda instituição que armazena dados pessoais de cidadãos faça uma análise de sua base de dados, pois alguns dados pessoais não possuem necessidade de armazenamento e apenas aumentam os riscos de vazamentos e outros incidentes com tais informações. Em casos em que a obtenção dos dados é necessária, analisa-se a melhor forma de protegê-los e deixá-los minimamente expostos somente a quem necessita ter acesso a esses dados. De acordo com a Lei, os dados pessoais devem receber toda cautela possível para que não sejam expostos desnecessariamente para terceiros, amenizando os riscos de mau uso.

No Brasil tem-se encontrado diversas dificuldades com relação ao cumprimento da nova lei LGPD, as corporações não estão adaptadas para seguir com seus importantes critérios, visto que a aplicação da segurança da informação ainda não recebe a devida importância em várias empresas brasileiras. É de suma importância que auxiliemos com o estudo das formas de se estar protegido e em acordo com a LGPD, visto que ainda há muita carência dos conhecimentos necessários para tal realização.

O objetivo geral deste trabalho demonstra a aplicação de técnicas de anonimização e pseudonimização em uma base de dados visando auxiliar a conformidade com a nova lei de proteção de dados pessoais. Para tanto, analisazamos uma base de dados (a de proprietários de imóveis urbanos da cidade de São Paulo) e demonstramos a aplicação da anonimização de dados, um dos mecanismos previstos na LGPD.

A partir desta questão da LGPD, verificaremos a forma como os dados pessoais dos cidadãos estão expostos na base e, se é possível aplicar os conceitos

de anonimização e pseudonimização, técnicas mencionadas na Lei. Utilizamos da ferramenta Splunk e de uma base de dados contendo informações de cidadãos da cidade de São Paulo para demonstrarmos a aplicação da anonimização neste trabalho.

1 LGPD (Lei Geral de Proteção de Dados)

No ano de 2018 foi publicada a lei nº 13.709, que também é conhecida por Lei Geral de Proteção de Dados ou LGPD. O propósito desta lei é regularizar o tratamento de dados pessoais por pessoas físicas, empresas e estabelecimentos comerciais, visando a proteção de tais informações, para que não sejam utilizadas de forma maliciosa e minimizar os riscos de vazamento. Isto encontra-se nos seguintes artigos, que reproduzimos a seguir:

> Art. 1º Esta Lei dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural.
>
> Parágrafo único. As normas gerais contidas nesta Lei são de interesse nacional e devem ser observadas pela União, Estados, Distrito Federal e Municípios.
>
> Art. 3º Esta Lei aplica-se a qualquer operação de tratamento realizada por pessoa natural ou por pessoa jurídica de direito público ou privado, independentemente do meio, do país de sua sede ou do país onde estejam localizados os dados.

O tratamento dos dados vai desde sua obtenção, até sua total utilização dentro do meio (compartilhamento, armazenamento, dentre qualquer outra forma de uso dos dados).

A obtenção deverá ser feita somente sob autorização de titular e em casos que a obtenção é justificavelmente necessária mediante os critérios existentes na lei, dentre alguns deles: para cumprimento de obrigação legal, para tutela de saúde, para proteção da vida, dentre outros.

A criação da lei já vinha sido discutida há muito tempo em vários congressos de vários países, entretanto, o processo foi acelerado devido ao ocorrido de março de 2018, onde a empresa de assessoria política Cambridge Analytica, utilizou dados pessoais de mais de 50 milhões de usuários do Facebook, sem seus

consentimentos, para propaganda política. O teste não apenas obteve informações de quem o fez, mas também daqueles em sua lista de amigos na rede social.

As informações foram obtidas por meio de um teste de personalidade denominado “Big Five”, disponibilizado gratuitamente no Facebook no ano de 2014, com o objetivo de mapear o perfil do usuário através do comportamento do mesmo na rede social, tais como conteúdos curtidos, compartilhados e postados na plataforma e, utilizar dessas informações para direcionar propaganda política de forma mais eficaz para a vítima.

De acordo com a matéria Cambridge Analytica se declara culpada em caso de uso de dados do Facebook (g1,2019), a propaganda política foi direcionada por forma de anúncios, específicos para o perfil traçado conforme os dados obtidos no “Big Five”, Cambridge Analytica ofereceu serviços para a campanha presidencial de Donald Trump em 2016, sendo acusada de manipulação e uso indevido de dados. O Facebook permaneceu alegando que a atitude da Cambridge Analytica não foi ilegal, visto que entre os anos de 2007 e 2014 a empresa autorizou gratuitamente o acesso a dados de usuários para acesso de pesquisadores para fins acadêmicos, tal acesso é consentido quando é criada uma conta no Facebook, através dos termos a serem aceitos. Entretanto, a rede social não permite que os dados sejam vendidos ou transferidos, atitude que foi cometida por Kogan, o criador do teste “Big Five”, que vendeu os dados para a empresa de consultoria política. O Facebook foi multado em 5 bilhões de dólares.

O ocorrido acelerou a criação da Lei GDPR europeia e consequentemente, da lei LGPD brasileira. A LGPD já estava em processo desde 2014, com a criação do Marco Civil da internet. A proposta inicial era de que a lei entrasse em vigor em agosto de 2020, porém foi adiada para 2022, com a justificativa de que apenas uma parcela das empresas havia iniciado o seu processo de adaptação em conformidade com a lei.

Para melhor entendimento da LGPD, é preciso compreender a sua definição de dados pessoais e dados pessoais sensíveis.

A lei manteve a definição de dado pessoal tal como a definição de informação pessoal, prevista na Lei 12.527/2011: IV - informação pessoal: aquela relacionada à pessoa natural identificada ou identificável. Ou seja, para determinados um dado

como pessoal, basta que este permita identificar um indivíduo, podemos citar como exemplos: RG, CPF, endereço residencial, telefone, entre outros.

Já os Dados Pessoais Sensíveis, de acordo com a lei, são aqueles sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural.

Os dados pessoais sensíveis são aqueles que além de pessoais, podem expor o indivíduo em questão a alguma forma de discriminação, como por exemplo: os dados de biometria facial expõem a etnia do sujeito, e, podem consequentemente, o expor a discriminações raciais no ambiente em que o tal dado é mostrado.

Agora que conhecemos a LGPD, discutiremos duas técnicas previstas na lei: a anonimização e pseudonimização.

1.2 Anonimização e Pseudonimização de Dados na LGPD

Segundo a Lei de Proteção de Dados Pessoais presente no Art. 5º, inciso III, define que um dado anonimizado é caracterizado por um:

“dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento.”

Por conseguinte, Art. 5º, inciso III, define que anonimização é caracterizada pela:

“utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo.”

Apesar da lei apenas ter estabelecido definição para anonimação, ela ainda possui em seu corpo alguns pontos aos quais são mencionados a utilização de pseudonimização. Seu conceito é aplicado de forma semelhando a definida na GDPR, no qual o dado sofre um processo de anonimização, mas que nesse caso, permite que o processo inverso seja realizado, ou seja, que os dados sejam reorganizados e reagrupados de modo que se torna possível identificar

indiretamente o indivíduo a quem os dados pertence. Cabe ressaltar que a pseudonimização é uma técnica utilizada para proteção de dados pessoais para preservação da identidade do denúncias anonimas, conforme previsto no § 4º do Art. 6º do Decreto nº 10.153/2019.

Tendo em vista que as análises realizadas neste estudo, tem como direção o que a lei determina, a seguir estão todos os parágrafos, artigos, alíneas e incisos na lei LGPD (LEI Nº 13.709, 2018). que façam menção para a utilização da anonimização e pseudominização:

> Art. 7, IV - para a realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais;
>
> Art. 11, II, c) realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais sensíveis;
>
> Art. 12. Os dados anonimizados não serão considerados dados pessoais para os fins desta Lei, salvo quando o processo de anonimização ao qual foram submetidos for revertido, utilizando exclusivamente meios próprios, ou quando, com esforços razoáveis, puder ser revertido.
>
> Art. 12, § 1º. A determinação do que seja razoável deve levar em consideração fatores objetivos, tais como custo e tempo necessários para reverter o processo de anonimização, de acordo com as tecnologias disponíveis, e a utilização exclusiva de meios próprios.
>
> Art. 12, § 3º. A autoridade nacional poderá dispor sobre padrões e técnicas utilizados em processos de anonimização e realizar verificações acerca de sua segurança, ouvido o Conselho Nacional de Proteção de Dados Pessoais.
>
> Art. 13. Na realização de estudos em saúde pública, os órgãos de pesquisa poderão ter acesso a bases de dados pessoais, que serão tratados exclusivamente dentro do órgão e estritamente para a finalidade de realização de estudos e pesquisas e mantidos em ambiente controlado e seguro, conforme práticas de segurança previstas em regulamento específico e que incluam, sempre que possível, a anonimização ou pseudonimização dos dados, bem como considerem os devidos padrões éticos relacionados a estudos e pesquisas.
>
> Art. 13, § 4º. Para os efeitos deste artigo, a pseudonimização é o tratamento por meio do qual um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo, senão pelo uso de informação adicional mantida separadamente pelo controlador em ambiente controlado e seguro.

> Art. 16, II - estudo por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais;
>
> Art. 16, IV - uso exclusivo do controlador, vedado seu acesso por terceiro, e desde que anonimizados os dados.
>
> Art. 18, IV - anonimização, bloqueio ou eliminação de dados desnecessários, excessivos ou tratados em desconformidade com o disposto nesta Lei;
>
> Art. 18, IX, § 6º O responsável deverá informar, de maneira imediata, aos agentes de tratamento com os quais tenha realizado uso compartilhado de dados a correção, a eliminação, a anonimização ou o bloqueio dos dados, para que repitam idêntico procedimento, exceto nos casos em que esta comunicação seja comprovadamente impossível ou implique esforço desproporcional.
>
> Art. 18, IX, § 7º A portabilidade dos dados pessoais a que se refere o inciso V do caput deste artigo não inclui dados que já tenham sido anonimizados pelo controlador.

Apesar do foco principal ser anonimização, a pseudonimização também é uma técnica mais simples e válida de ser utilizada, porém os dados pseudonimizados podem ser identificados em determinados casos.

A partir dessas citações, é evidente que a LGPD demonstra um incentivo significativo para o uso de dados anonimizados. Em relação a dados pseudo anonimizados, não são tão claras as vantagens jurídicas perante a lei. Porém, como a lei ainda sofre algumas alterações, pois ainda não entrou em vigor, podemos presumir que é possível que futuramente novas interpretações venham a reconhecer vantagens para a sua aplicação referente ao ponto de vista das obrigações regulatórias.

2 Anonimização

Segundo um dos pontos da LGPD, empresas que necessitem realizar a coleta de dados, o devem fazer de maneira que colete o mínimo possível de informações pessoais de indivíduos para minimizar os riscos de vazamentos ou uso indevido desses dados.

Entretanto, existem ocasiões em que a coleta de dados pessoais é necessária, a título de exemplo: suponha-se que uma instituição de ensino superior aleatório, prepara a realização do vestibular para a ingressão de seus cursos, e que uma das diretrizes de ingresso adotadas pela instituição fosse utilizado o sistema de cotas raciais. Diante disso faz-se necessário a coleta de dados étnicos de seus vestibulandos, um dado que é caracterizado como pessoal e sensível pela LGPD. Neste caso, como é estritamente necessária a coleta do dado, existem outras técnicas que podem ser utilizadas para amenizar os riscos de vazamento durante coleta e o manuseio desse dado, uma dessas técnicas é a anonimização de dados.

A anonimização consiste em utilizar técnicas visando impossibilitar a associação de um dado com o sujeito, com objetivo de proteger a identidade do titular e torná-lo não identificável. Segue um exemplo de seu uso:

Figura 1. Exemplo simples de anonimização de dados pessoais.

Figura Própria

Na imagem acima antes da anonimização, era possível identificar o titular dos dados devido ao nome completo e o documento OAB expostos. Após a anonimização, restando apenas os dados de gênero, nacionalidade e profissão, o

titular torna-se não identificável, visto que existem diversas pessoas que batem com as informações presentes.

A exclusão permanente sem possibilidade de recuperação, das informações que tornam possíveis a identificação da advogada em questão, garantem que o processo não será revertido e, sendo consequentemente mais seguro.

2.1 – Tipos de Anonimização

2.1.1 – De-identificação

A de-identificação ou desidentificação trata-se da remoção ou ofuscação de informações que impedem a identificação, ou seja, trata-se de um processo de criação de registros no qual os dados não possuem informações que permitam que possam identificar diretamente a identidade de uma entidade, como a identidade de um indivíduo. Segue um exemplo de uso:

Figura 2. Exemplo simples de de-identificação de dados pessoais.

ID	Nome	Sobrenome	Gênero	Nacionalidade	Profissão	Data de Nasc.
1	Fulana	de Tal	Fem	Brasileira	Cirurgiã	21/03/1986
2	Cicrano	da Silva	Masc	Brasileiro	Contador	23/08/1995
3	Beltrano	dos Santos	Masc	Brasileiro	Balconista	07/12/2000

ID	Nome	Sobrenome	Gênero	Nacionalidade	Profissão	Data de Nasc.
1	Fulana	AFCD86AB	Fem	Brasileira	Cirurgiã	3BC51DAF
2	Cicrano	BDCG4AF3	Masc	Brasileiro	Contador	4F2ACD6A
3	Beltrano	BDGAC3A5	Masc	Brasileiro	Balconista	32FABEE1

Figura Própria

Como pode ser observado, após a desidentificação informação imprescindível para identificar a identidade do indivíduo foi ofuscado na tabela.

A desidentificação irreversível refere-se à incapacidade de re-identificar um registro de dados para um indivíduo específico associado a esse registro por meio de “engenharia reversa, incluindo mas não limitado a decodificação, decifração ou

descriptografia, remoção, generalização ou substituição de informações pessoalmente identificáveis explícitas.

2.1.2 Criptografia

A criptografia em si não é um mecanismo mencionado na LGPD, contudo é uma técnica viável para garantir a segurança das informações.

Existem duas formas de criptografia:

- Simétrica: método mais antigo, o remetente e o destinatário utilizam de uma única chave para codificar e decodificar a mensagem ou dado.
- Assimétrica: é considerado o meio mais seguro, porém mais é mais lento e requer maior capacidade de processamento. Utiliza-se de duas chaves, uma pública que é utilizada para criptografar a mensagem ou dado, e uma chave privada que serve para decodificação.

Segundo o GARRET(TECMUNDO,2020): “criptografia é o nome que se dá a técnicas que transformam informação inteligível em algo que um agente externo seja incapaz de compreender” .

Conforme mencionado anteriormente, a pseudonimização diferencia-se da anonimização por conter os dados identificáveis do titular em um local separado dos demais dados, isso significa que para que o processo de anonimização seja feito corretamente utilizando o método da criptografia, deve-se descartar a chave utilizada no processo, impossibilitando assim a reversão (descriptografia).

2.1.3 Supressão de atributos

Quando não é possível realizar uma anonimização adequada de um atributo ou quando um atributo dos conjuntos dos dados não tem muita relevância, podemos utilizar-se da técnica supressão de atributos, no qual consiste em realizar a remoção completa de uma seção (também conhecido como "coluna" em uma base de dados) do conjunto de dados.

Conforme exemplo a seguir, a coluna CPF foi totalmente removida, aplicando-se assim a técnica de supressão:

Figura 3. Exemplo simples de supressão de atributos.

Figura Própria

2.1.4 Encobrimento de caracteres

A técnica encobrimento de caracteres consiste em realizar uma troca parcial dos caracteres de um determinado valor no conjunto de dados por um símbolo constante (por exemplo "*").

Figura 4. Exemplo simples de encobrimento de caracteres.

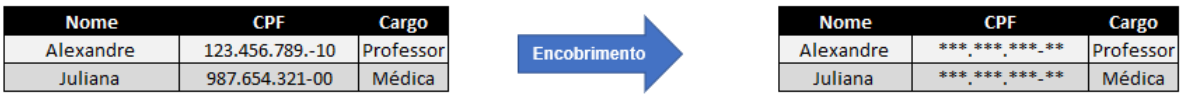


Figura Própria

2.1.5 Agregação de Dados

Trata-se da conversão de dados para uma versão resumida baseada em conjuntos ou intervalos de valores. Através de uma análise prévia que identifique que os valores individuais detalhados podem ser agrupados desde que possam ser suficientes para o seu propósito. Podemos utilizar como exemplo a base de dizimistas de uma igreja que ao invés de utilizar uma base com todos os contribuintes incluindo o valor de sua renda, agrupa-os por meio de faixas salariais.

Figura 5. Exemplo simples de agregação de dados.

Figura Própria.

2.1.6 Generalização

Quando um determinado valor no conjunto dos dados não precisa ser exato mas que ainda assim possui uma utilidade em seu objetivo, pode-se utilizar a técnica de generalização (também conhecida como recodificação), no qual consiste em reduzir deliberadamente a precisão dos dados (substituir a idade de um indivíduo pela sua faixa etária ou substituir o nome da rua onde mora pelo nome do bairro, por exemplo).

Figura 6. Exemplo simples de generalização.

Figura Própria

2.2 Pseudonimização

O ato de separar os dados que possam identificar o titular separados, sem que haja exclusão destes, denomina-se pseudonimização, ou seja, substituir dados reais por pseudônimos. Palavras ou código gerados artificialmente. Representação mascaradas dos dados originais. A ideia é permitir manter os dados e atributos de uma base relacional permitindo guardar a estrutura e sintaxe dos dados.

Figura 7.Exemplo de pseudonimização.

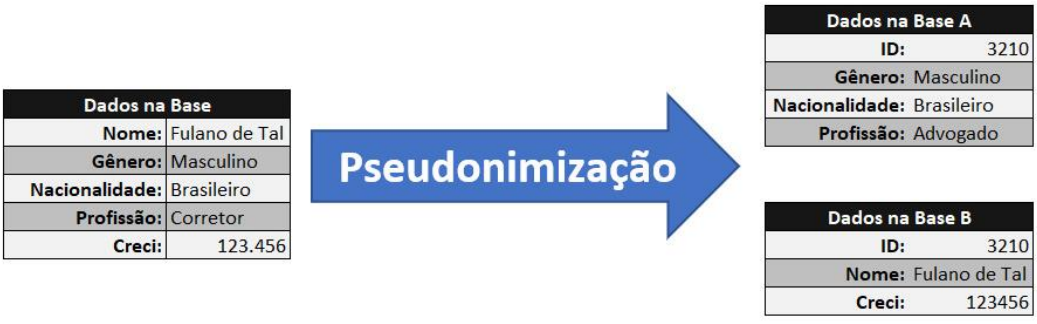


Figura Própria

No próximo capítulo explicaremos mais sobre a anonimização de dados pessoais.

3 Obtendo a base de dados e importando-a ao Splunk

Neste capítulo, na seção 3.1 e 3.2 tem como objetivo apresentar a ferramenta splunk e sua linguagem de pesquisa. A seção 3.3 mostra como obter os dados cadastrais dos imóveis no município de São Paulo. A seção 3.4 mostra requisitos de hardware e software utilizados. Finalmente da seção 3.5 são apresentadas as análises realizadas dos dados obtidos na seção 3.3.

3.1 O que é a ferramenta Splunk

O Splunk é uma empresa multinacional americana, que criou e mantém uma software, de mesmo nome, que permite coletar, indexar e analisar, as informações dos componentes da infraestrutura ou negócios de TI de uma organização. Esta ferramenta facilita a pesquisa, monitoramento e análise de grandes quantidades de dados, onde é possível gerar gráficos, relatórios, alertas, dashboards (painel visual que apresenta, de maneira centralizada, um conjunto informações: indicadores e suas métricas).

A título de exemplo, a partir da coleta e indexação dos dados contidos no syslog de uma máquina, é possível visualizar problemas de TI que ocorre no ambiente da organização. Tendo isso em vista, pode-se realizar a implementação de sistemas de alertas baseados na ocorrência de um determinado evento, sendo ele desejado ou não, e/ou pela quantidade x de erros inesperados.

Para este projeto, foi utilizado a versão gratuita do Splunk Enterprise, versão 8.0.4. O Splunk Enterprise pode indexar qualquer tipo de dados. Em particular, todo e qualquer fluxo de TI, máquina e dados históricos, como logs de eventos do Windows, logs de servidores da Web, logs de aplicativos ativos, status de rede, métricas, monitoramento de alterações, filas de mensagens, aquivos em formato csv, json e raw, arquivos compactados etc. Esses dados são indexados com três campos por padrão, denominados pela ferramenta de host, source e sourcetype. Os campos que determinam o tempo da coleta é chamado de timeout e timestamp.

Para obter a versão gratuita do Splunk Enterprise, basta entrar no site da Splunk (ou seja, acessar www.splunk.com), realizar um cadastro. Ele está disponível

para sistemas operacionais Windows, Linux e Mac OS, em arquiteturas de 32-bits e 64-bits.

3.2 SPL

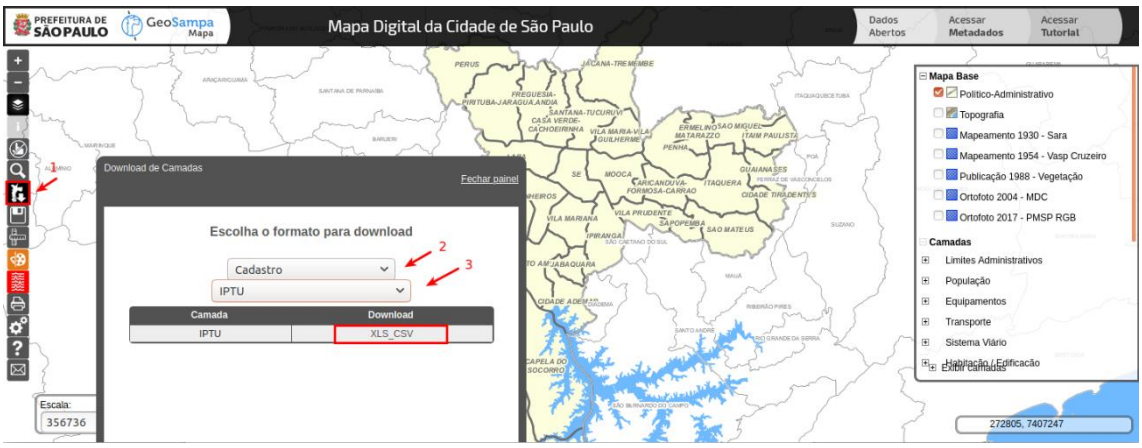
Como o Splunk Enterprise trabalha com dados estruturados e não estruturados, o uso da linguagem de consulta estruturada, conhecida como SQL (Structured Query Language) não está de acordo com a ferramenta. Por conseguinte, fez-se necessária a criação de uma linguagem própria para a manipulação dos dados, a Linguagem de Processamento de Busca (SPL, do inglês Search Processing Language). Esta linguagem foi desenvolvida pela própria empresa e sua sintaxe teve como base os processos de Unix e da linguagem SQL.

A partir do uso da SPL, os usuários tem a condição de manusear os dados indexados a fim de alcançar uma melhor visualização e/ou aplicabilidade destes. Os comandos em SPL são categorizados conforme funções de filtragem agrupamento, informação e modificação que são usados na barra de pesquisa e separados por barras verticais (|).

3.3 Base de dados do IPTU da prefeitura de São Paulo

A coleta de todos os dados constantes do cadastro imobiliário fiscal relativos ao Imposto Predial e Territorial Urbano – IPTU, do município de São Paulo.

Figura 8. Mapa Digital da cidade de São Paulo.



Fonte: http://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx

Para obter os dados, clique no ícone “download de arquivo” indicado na seta 1, em seguida selecione a opção cadastro apontada pela seta 2, e por último selecione a opção “IPTU” referenciada pela seta 3. Um link irá aparecer “XLS_CSV”, circulado, onde poderá ser executado o download de um arquivo em formato ZIP, contendo os dados do ano corrente em formato CSV.

O arquivo em formato CSV contido no download possui 3.498.644 de imóveis registrados até o ano de 2020. O registro possui 35 campos:

- Ano da construção corrigido,
- ano de início da vida do contribuinte,
- ano do exercício, área construída,
- área do terreno,
- área ocupada,
- bairro do imóvel,
- CEP do imóvel,
- codlog do imóvel,
- complemento do imóvel,
- CPF/CNPJ do contribuinte 1,
- CPF/CNPJ do contribuinte 2,
- data do cadastramento,
- fase do contribuinte,
- fator de obsolescência,
- fração ideal,
- mês de início da vida do contribuinte,
- nome de logradouro do imóvel,
- nome do contribuinte 1,
- nome do contribuinte 2,
- número da NL,
- número do condomínio,
- número do contribuinte,
- número do imóvel,
- quantidades de esquinas frentes,
- quantidades de pavimentos,

- referência do imóvel,
- testada para cálculo,
- tipo de contribuinte 1,
- tipo de contribuinte 2,
- tipo de padrão da construção,
- tipo de terreno,
- tipo de uso de imóvel,
- valor do M2 de construção,
- valor do M2 do terreno, para cada imóvel.

Figura 9 .Campos encontrados.

Type	Field	Value	Actions
Event	ANO DA CONSTRUCAO CORRIGIDO	1986	
	ANO DE INICIO DA VIDA DO CONTRIBUINTE	1990	
	ANO DO EXERCICIO	2020	
	AREA CONSTRUIDA	291	
	AREA DO TERRENO	500	
	AREA OCUPADA	203	
	BAIRRO DO IMOVEL	VILA SANTO ESTEFANO	
	CEP DO IMOVEL	04152-100	
	CODLOG DO IMOVEL	14118-6	
	CPF_CNPJ DO CONTRIBUINTE 1	XXXXXX5797XXXX	
	CPF_CNPJ DO CONTRIBUINTE 2	XXXXXX7383XXXX	
	DATA DO CADASTRAMENTO	11/01/20	
	FASE DO CONTRIBUINTE	0	
	FATOR DE OBSOLESCENCIA	0,64	
	FRACAO IDEAL	1,0000	
	MES DE INICIO DA VIDA DO CONTRIBUINTE	1	
	NOME DE LOGRADOURO DO IMOVEL	R CALOGERO CALIA	
	NOME DO CONTRIBUINTE 1	MARTA REGINA ALVES PUGLIESE	
	NOME DO CONTRIBUINTE 2	MARIA HELENA ALVES DE SOUZA LEAO	
	NUMERO DA NL	1	
	NUMERO DO CONDOMINIO	00-0	
	NUMERO DO CONTRIBUINTE	0480370153-9	
	NUMERO DO IMOVEL	186	
	QUANTIDADE DE ESQUINAS_FRENTES	0	
	QUANTIDADE DE PAVIMENTOS	2	
	TESTADA PARA CALCULO	10,00	
	TIPO DE CONTRIBUINTE 1	PESSOA FISICA (CPF)	
	TIPO DE CONTRIBUINTE 2	PESSOA FISICA (CPF)	
	TIPO DE PADRAO DA CONSTRUCAO	Residencial horizontal - padr\xE3o C	
	TIPO DE TERRENO	Normal	
	TIPO DE USO DO IMOVEL	Resid\xEAncia	
	VALOR DO M2 DE CONSTRUCAO	1368,00	
	VALOR DO M2 DO TERRENO	1952,00	

Imagem do Sistema

3.4 Importando os arquivos csv para o Splunk

A análise dos dados foi realizada através de notebook Dell Intel® Core™ i5-7200U CPU, 2.50GHz, 12GB de memória RAM, 1TB de armazenamento de disco, arquitetura 64-bits com sistema operacional Linux Ubuntu 18.04.4 LTS.

O arquivo contendo os dados foram indexados através do método de monitoram de arquivo, em um index criado com nome de “iptu”.

Figura10 - Dados indexados

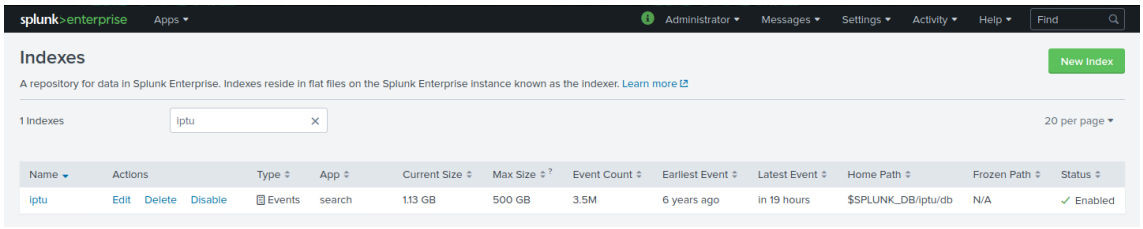


Imagem do Sistema

3.5 Analisando os dados no Splunk

Após inserirmos a base de dados no Splunk por meio dos métodos acima mencionados, iniciamos uma série de pesquisas para estudar a base, analisando os dados expostos nela.

3.5.1 Erros de grafia nos nomes

Realizamos pesquisas por alguns possíveis erros de grafia que poderiam ser cometidos nos nomes ou sobrenomes inseridos na base, encontramos os seguintes resultados:

Figura 11 - Erro de grafia no nome "Thiago Predo" ao invés de "Thiago Pedro"

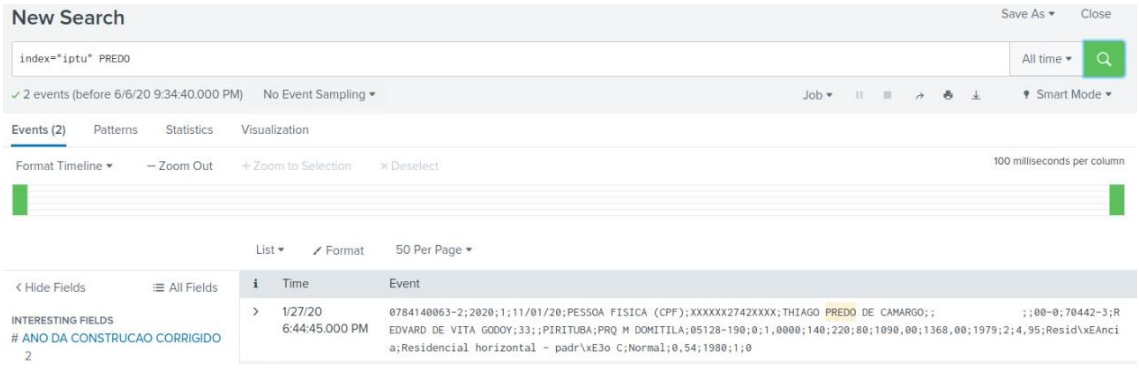


Imagem do Sistema

Figura 12 - Nome "Humberto" escrito como "Hmberto"

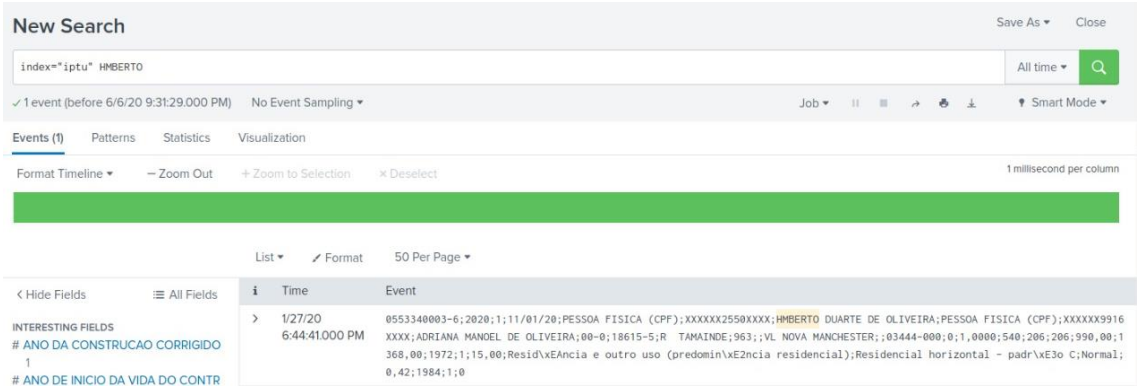


Imagem do Sistema

Figura 13 - Diversos casos de erro de grafia do nome "Albuquerque"

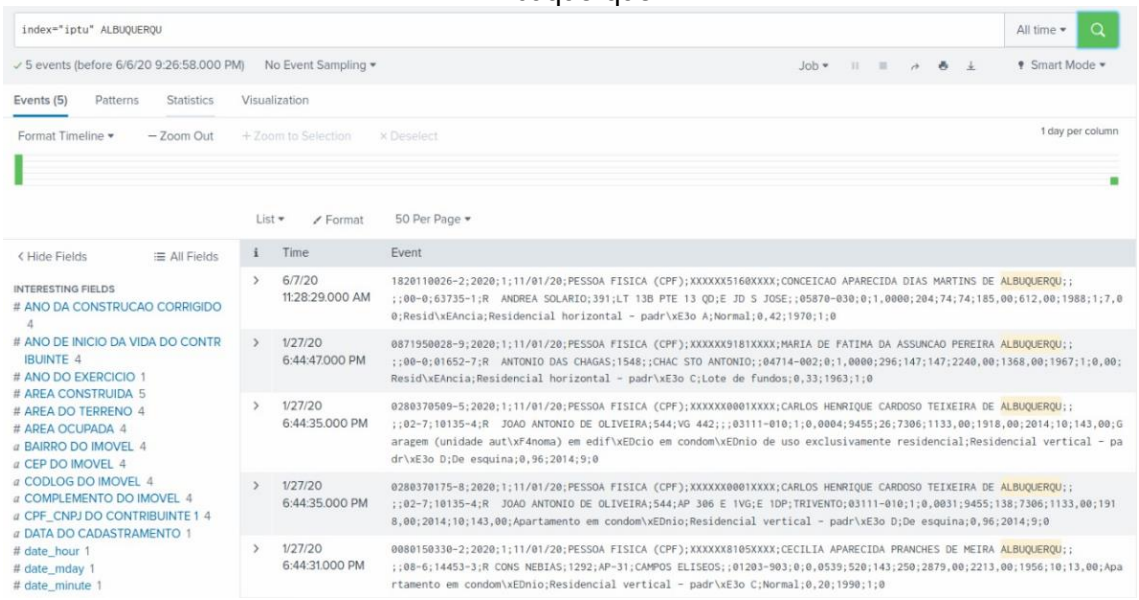


Imagem do Sistema

Os erros na inserção dos nomes dos cidadãos podem indicar um problema na comunicação e envio de informações, desde o contribuinte do IPTU, ao cartório até a prefeitura criadora da base.

O ocorrido pode acarretar em futuros problemas para os contribuinte, visto que o nome não irá ser o mesmo de seus documentos de identidade e precisará de mais informações para comprovar sua propriedade.

3.5.2 Exemplo de pesquisa por bairro

Figura 14 - Resultado de pesquisa de um proprietário conhecido

Type	Field	Value	Actions
Event	ANO DA CONSTRUCAO CORRIGIDO	1986	
	ANO DE INICIO DA VIDA DO CONTRIBUINTE	1990	
	ANO DO EXERCICIO	2020	
	AREA CONSTRUIDA	291	
	AREA DO TERRENO	500	
	AREA OCUPADA	203	
	BAIRRO DO IMOVEL	VILA SANTO ESTEFANO	
	CEP DO IMOVEL	04152-100	
	CODLOG DO IMOVEL	14118-6	
	CPF_CNPJ DO CONTRIBUINTE 1	XXXXXX5797XXXX	
	CPF_CNPJ DO CONTRIBUINTE 2	XXXXXX7383XXXX	
	DATA DO CADASTRAMENTO	11/01/20	
	FASE DO CONTRIBUINTE	0	
	FATOR DE OBSOLESCENCIA	0,64	
	FRACAO IDEAL	1,0000	
	MES DE INICIO DA VIDA DO CONTRIBUINTE	1	
	NOME DE LOGRADOURO DO IMOVEL	R CALOGERO CALIA	
	NOME DO CONTRIBUINTE 1	MARTA REGINA ALVES PUGLIESE	
	NOME DO CONTRIBUINTE 2	MARIA HELENA ALVES DE SOUZA LEAO	
	NUMERO DA NL	1	
	NUMERO DO CONDOMINIO	00-0	
	NUMERO DO CONTRIBUINTE	0480370153-9	
	NUMERO DO IMOVEL	186	
	QUANTIDADE DE ESQUINAS_FRENTES	0	
	QUANTIDADE DE PAVIMENTOS	2	
	TESTADA PARA CALCULO	10,00	
	TIPO DE CONTRIBUINTE 1	PESSOA FISICA (CPF)	
	TIPO DE CONTRIBUINTE 2	PESSOA FISICA (CPF)	
	TIPO DE PADRAO DA CONSTRUCAO	Residencial horizontal - padr\xE3o C	
	TIPO DE TERRENO	Normal	
	TIPO DE USO DO IMOVEL	Resid\xEAncia	
	VALOR DO M2 DE CONSTRUCAO	1368,00	
	VALOR DO M2 DO TERRENO	1952,00	

Imagem do Sistema

Realizamos uma pesquisa filtrando pelo bairro Saúde e incluindo o nome de uma pessoa conhecida do grupo. Em seguida, utilizamos do método “group by” para localizar outras pessoas proprietárias de imóveis no mesmo bairro.

Figura 15 - Proprietários de imóveis do bairro da Saúde

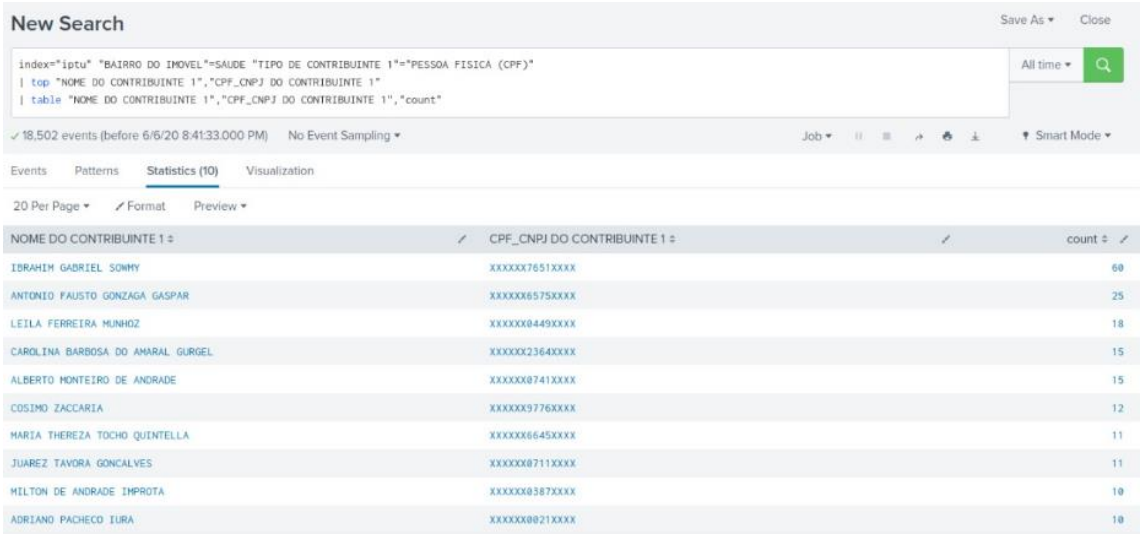


Imagem do Sistema

Figura 16 - Continuação dos proprietários da Saúde

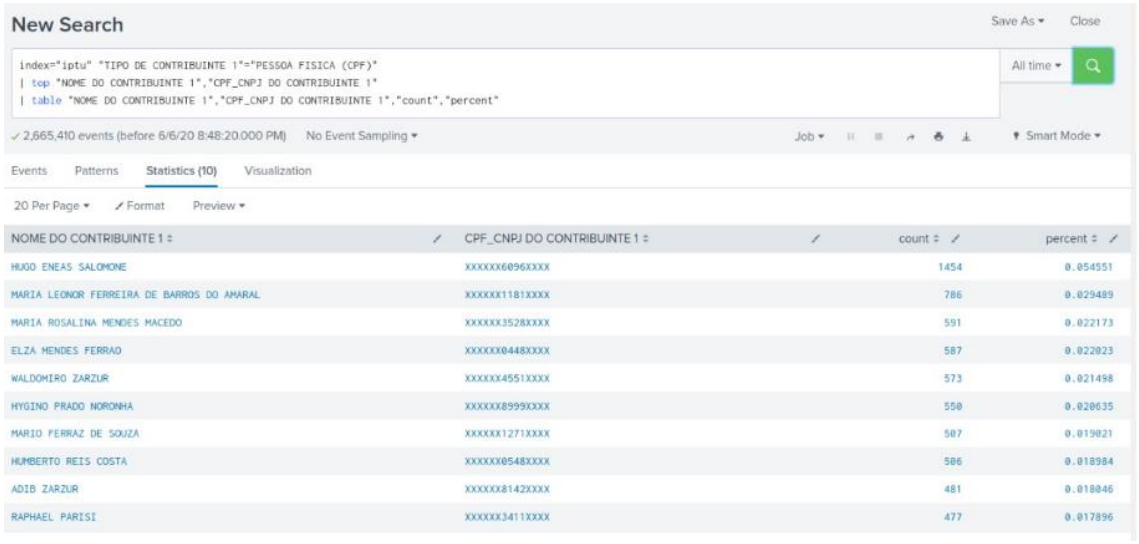


Imagem do Sistema

Conforme mostrado acima, encontramos nome completo e CPF descaracterizado de contribuintes.

3.5.3 Pesquisas externas com dados da base de contribuintes

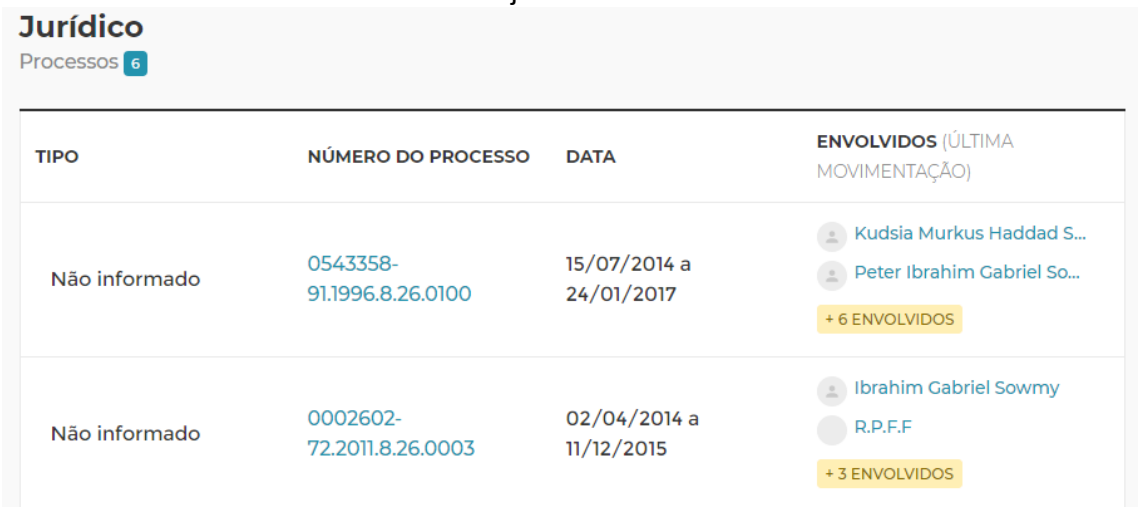
Utilizamos de dados localizados na base para verificarmos a possibilidade de encontrar mais informações sobre os cidadãos ali mencionados, consequentemente verificando o nível de segurança e o quanto estão realmente protegidos.

No resultado mostrado na figura 15 na página 33, na primeira linha de resultado encontramos um nome considerado incomum: “Ibrahim Gabriel Sowmy” e seu CPF descaracterizado.

Não podemos afirmar com certeza se a descaracterização do CPF trata-se de uma pseudonimização, visto que não podemos visualizar se o gerenciador da base de dados possui o dado de CPF completo armazenado em algum local.

Utilizamos dessas informações para identificar o cidadão e, encontramos este mesmo nome envolvido em processos jurídicos, entretanto, por não ter informações completas ou parciais do CPF no site detectado, não foi possível confirmar se trata-se do mesmo cidadão.

Figura 17 - Nome encontrado na base também foi encontrado em um site de processos
jurídicos.



Fonte: https://www.escavador.com/sobre/127101101/ibrahim-gabriel-sowmy

Realizamos a mesma busca, porém dessa vez com uma pessoa jurídica ao invés de pessoa física.

Figura 18 - Pesquisa por CNPJ aleatório

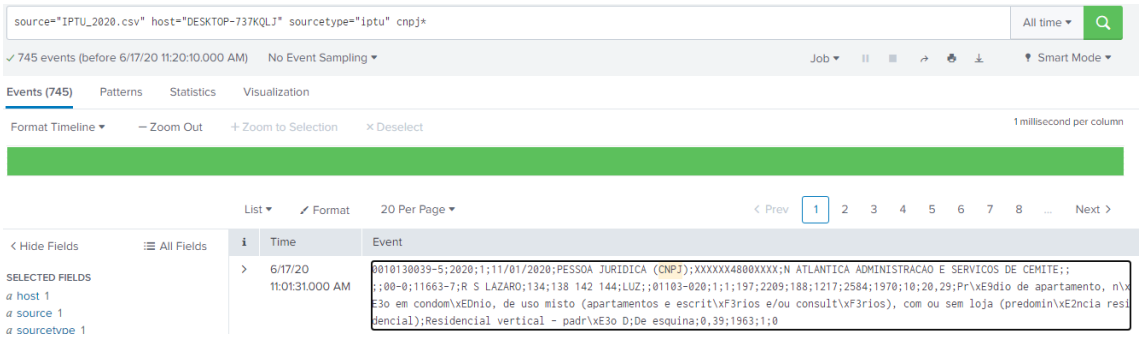


Imagem do Sistema

Podemos notar que o CNPJ encontra-se descaracterizado e o nome da empresa exposto.

Através de uma pesquisa simples no Google, foi facilmente possível localizar o CNPJ completo da empresa, sem qualquer segurança.

Figura 19 - Identificação do CNPJ da empresa.

CNPJ e telefone de N. Atlantica, Administracao e Servicos de Cemiterio Ltda.

Consulte o cartão CNPJ da empresa "N. Atlantica, Administracao E Servicos De Cemiterio Ltda." realizando a busca do CNPJ pelo nome. Descubra o telefone, endereço, email, sócios e demais informações cadastrais da empresa. Dados atualizados com a Receita Federal.

Dados do Cadastro Nacional de Pessoa Jurídica do Brasil

Número de inscrição do CNPJ

07.078.848/0001-82

Aberta em

3/9/2004

Razão social (nome empresarial)

N. Atlantica, Administracao E Servicos De Cemiterio Ltda.

Fonte: https://www.cnpjreceita.com/empresa/n-atlantica-administracao-e-servicos-de-cemiterio-ltda/07078848000182

Figura 20 - Informações encontradas pelo nome da empresa

Capital social

R$880.404,00 (Oitocentos e oitenta mil e quatrocentos e quatro reais).

Sócios da empresa

Nome (M)
Qualificação (22-Sócio)
Qualificação do Representante Legal (05-Administrador)
Nome do Representante Legal (Marina Cesar Jaguaribe Ekman Helito)
Nome (R)
Qualificação (22-Sócio)
Qualificação do Representante Legal (05-Administrador)
Nome do Representante Legal (Marina Cesar Jaguaribe Ekman Helito)
Nome (Marina Cesar Jaguaribe Ekman Helito)
Qualificação (05-Administrador).

Fonte: https://www.cnpjreceita.com/empresa/n-atlantica-administracao-e-servicos-de-cemiterio-ltda/07078848000182

Nesse mesmo site, também estavam inseridas outras informações da empresa como capital social e nome do representante legal da mesma.

Juntando as informações já encontradas, localizamos o CPF descaracterizado da representante legal da empresa e seu nome envolvido em processos judiciais.

Figura 21 - Identificação do CPF descaracterizado da representante legal e de outro administrador da empresa.

Sócios

Código	Nome	Data de entrada	Qualificação
CPF***007348**	Marina Cesar Jaguaribe Ekman Helito	2012-07-02	Administrador
Representante CPF**7007348**	N. Atlantica, Administracao e Servicos de Cemiterio Ltda.	2010-12-13	Sócio Representante: Administrador
Representante CPF**9325218**	M. Bela Vista, Incorporacoes e Administracao de Bens e Servicos de Cemiterios Ltda	2009-08-18	Sócio Representante: Administrador

Fonte: http://cnpj.info/MANTIQUEIRA-EMPREENDIMENTOS-IMOBILIARIOS-E-SERVICOS-DE-CEMITERIO-LTDA-Av-Papa-Joao-Xxiii-4734-Maua-SP-09370800

Localizamos também outras empresas representadas por esta mesma pessoa e seus CNPJs, além de seu capital social.

Figura 22 - Identificação de outros CNPJs da mesma representante.

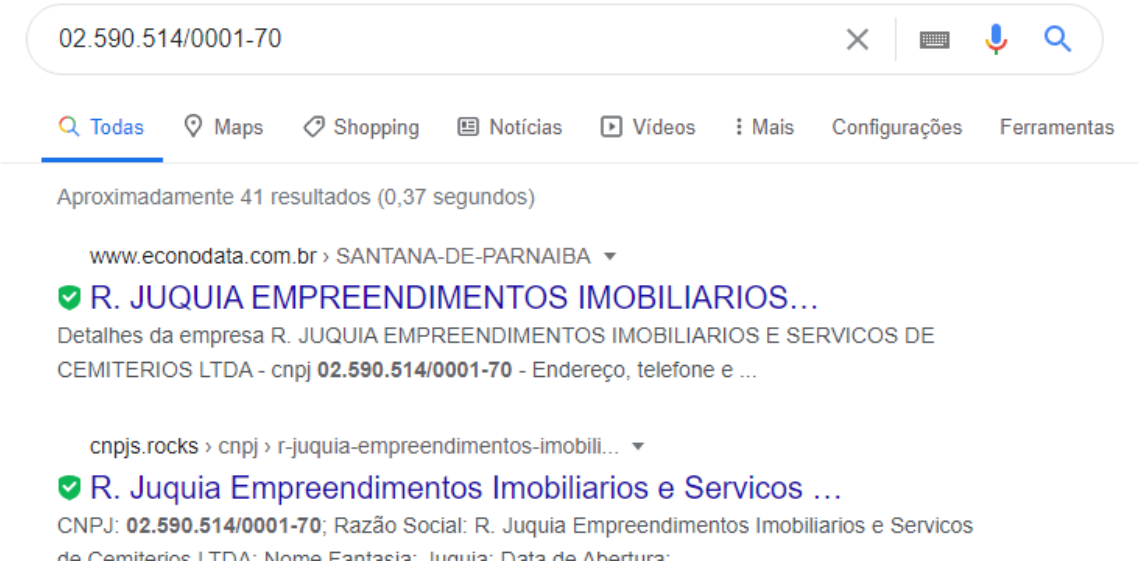
Fonte: https://www.empresascnpj.com/s/socio/marina-cesar-jaguaribe-ekman-helito

Com o nível de informações coletadas, conseguimos perceber que a descaracterização do CPF dos contribuintes da base foi suficiente para manter sua privacidade, pois não obtivemos sucesso em confirmar outros dados do cidadão exemplificado, mesmo que este tivesse um nome incomum e pouco utilizado.

Já com o nome da empresa exposto, foi possível localizar mais informações da empresa e de sua representante legal, o CNPJ que foi descaracterizado em sua inserção na base acabou por não ser de muito auxilio para proteger a privacidade da corporação, visto que este foi facilmente localizado através do nome exposto na base.

Para melhorar a proteção no caso do CNPJ, seria necessário pseudonimizar ou anonimizar também o nome da empresa, já que mesmo se fosse feito o inverso, protegendo o nome e expondo o CNPJ, também seria possível localizar as mesmas informações encontradas posteriormente, conforme imagem:

Figura 23 - Identificação da empresa através de seu CNPJ



Fonte: buscador, google.com

4 Aplicação da anonimização e pseudonimização

Neste capítulo, iremos demonstrar as aplicações das técnicas de anonimização e pseudonimização na base de dados obtida em questão, com base na análise que tivemos neste capítulo, verificaremos mais profundamente a eficácia das medidas já tomadas pelo autor da base e informaremos melhores aplicações caso necessário.

4.1 – Aplicação do encobrimento de caracteres

Conforme discutido no capítulo anterior, a medida de proteção aplicada para pessoa física foi suficiente. A base continha o nome completo de pessoas físicas e seus CPFs descaracterizados, com apenas alguns números visíveis, ao pesquisarmos somente o nome e a parte do CPF informada na base, não pudemos localizar quaisquer informação que podemos afirmar ser sobre o cidadão informado na base.

Já no caso de pessoa jurídica, o nome da empresa estava completamente explícito, enquanto o CNPJ estava descaracterizado. Apenas com o nome da empresa foi suficiente para encontrar muitas outras informações da empresa e de seu representante legal, por isso concluímos que precisamos de uma forma mais segura de mascarar os dados de pessoa jurídica na base.

Criamos uma tabela para demonstrar a atual forma que os dados (nome da empresa e número de CNPJ) do mesmo exemplo de pessoa jurídica utilizado no capítulo anterior, apresenta-se na base.

N ATLANTICA ADMINISTRACAO E SERVICOS DE CEMITE	XXXXXX48000XXXX

Agora aplicaremos o conceito de anonimização de encobrimento de caracteres (substituir um dado por outro(s) caractere(s) não relacionados) no nome

da empresa, ocultaremos completamente o nome desta sem armazenar a informação em outro local.

X XXXXXXX XXXXXXXXX X XXXXXX XX XXXXXX	XXXXXX48000XXXX

Se buscarmos somente pelo número 48000, a única informação explícita, não encontraremos nada que nos leve à empresa em questão.

Analisaremos agora a base com as demais informações detectadas pelo Splunk no capítulo anterior, tais como CEP e nome da rua e do bairro.

X XXXXXXX XXXXXXXXX X XXXXXX XX XXXXXX	XXXXXX48000XXXX
R S LAZARO LUZ	01103-020

Com estas informações, localizamos diversas empresas encontradas neste endereço, sem o número específico do estabelecimento não é possível ter certeza de qual das encontradas é a empresa na base.

No próximo sub capítulo aplicaremos a de-identificação.

4.2 Aplicação da de-identificação

Faremos agora a aplicação da técnica de de-identificação no nome da empresa, substituindo o nome por códigos não identificáveis para os usuários que irão acessar a base online.

Dados expostos na base atualmente:

Nome	CNPJ	Endereço	CEP
n atlantica administracao e servicos de cemite	xxxxxx48000xxxx	r s lazaro luz	01103-020

Dados expostos após a aplicação da de-identificação:

Nome	CNPJ	Endereço	CEP
RLSM34OM15	xxxxxx48000xxxx	r s lazaro luz	01103-020

Com a técnica aplicada, não é mais possível identificar a pessoa jurídica em questão com os dados expostos.

Considerações Finais

As consequências que as inovações tecnológicas trouxeram ao mundo atual ou a ocorrência de inúmeros escândalos de vazamentos de dados pessoais que ocorreram nos últimos anos, resultou em uma mudança de postura das instituições governamentais para esse tema, assim como a tecnologia muda, novas legislações tendem acompanhar essas mudanças. A GDPR representa o início de uma nova cultura de proteção da privacidade e dos dados pessoais no continente europeu, tal qual a LGPD está sendo representa no Brasil.

Através desse estudo foi possível concluir que existem diversas técnicas de anonimização que podem ser utilizadas na proteção dos dados pessoais para que estes estejam em melhor conformidade com a LGPD. Apesar do exemplo utilizado para a realização deste estudo já possuía uma camada de anonimização aplicada em um de seus campos, foi demonstrado que não era o suficiente para manter a privacidade de seus titulares.

Vale ressaltar que até a publicação desse estudo, a lei vem sofrendo alterações e adiamentos para início de seu vigor, na qual seja possível culminar em uma inaplicabilidade dessas técnicas, ou o oposto, em uma intensificação de sua aplicação. Dada à importância do assunto, torna-se necessário o desenvolvimento de outros estudos que visam suprir essas modificações graduais na lei.

É possível notar as atitudes de empresas brasileiras em seu cotidiano, onde as próprias ainda não dão o devido valor à segurança da informação em seu ambiente profissional, além do mau uso e compartilhamento sem consentimento dos dados de clientes ser algo comum em seu dia a dia. É essencial que haja um maior estudo e maior valorização por parte das corporações e também dos próprios cidadãos titulares de tais dados para que seja feita a adaptação para com a LGPD.

REFERÊNCIAS

AGÊNCIA O GLOBO. Brasil é o país mais vulnerável a vazamento de informações, diz pesquisador. [S. l.], 2017. Disponível em: https://revistapegn.globo.com/Tecnologia/noticia/2017/09/brasil-e-o-pais-mais-vulneravel-vazamento-de-informacoes-diz-pesquisador.html. Acesso em: 3 jul. 2019.

BRITO, Carlos; ALVARENGA, Darlan. População brasileira chegará a 233 milhões em 2047 e começará a encolher, aponta IBGE. [S. l.], 2018. Disponível em: https://g1.globo.com/economia/noticia/2018/07/25/populacao-brasileira-chegara-a-233-milhoes-em-2047-e-comecara-a-encolher-aponta-ibge.ghtml. Acesso em: 5 jul. 2019.

GARRET, Filipe. TECMUNDO - O que é Criptografia? Disponível em: https://www.techtudo.com.br/artigos/noticia/2012/06/o-que-e-criptografia.html. Acesso em 28 de Junho de 2020.

PRESSE , France. Cambridge Analytica se declara culpada em caso de uso de dados do Facebook,2019. Disponível em:

https://g1.globo.com/economia/tecnologia/noticia/2019/01/09/cambridge-analytica-se-declara-culpada-por-uso-de-dados-do-facebook.ghtml Acesso em 28 de Junho de 2020.

POZZI, Sandro. EUA multam Facebook em 5 bilhões de dólares por violar privacidade dos usuários. [S. l.], 2019. Disponível em: https://brasil.elpais.com/brasil/2019/07/12/economia/1562962870_283549.html. Acesso em: 5 jul. 2019.

ENCRIPT IT SOLUTIONS. O caso Cambridge Analytica e a influência na criação da LGPD. [S. l.], 2019. Disponível em: https://encript.com.br/o-caso-cambridge-analytica-e-a-influencia-na-criacao-da-lgpd/. Acesso em: 5 jul. 2019.

LLP, Latham; LLP, Watkins. A New Era for Data Protection in Brazil. [S. l.], 2019. Disponível em: https://www.globalprivacyblog.com/legislative-regulatory-developments/a-new-era-for-data-protection-in-brazil/. Acesso em: 5 nov. 2019.

JOTA. GDPR: a nova legislação de proteção de dados pessoais da Europa. [S. l.], 2018. Disponível em: https://www.jota.info/opiniao-e-analise/artigos/gdpr-dados-pessoais-europa-25052018. Acesso em: 15 ago. 2019.

ALVES, Paulo. Facebook e Cambridge Analytica: sete fatos que você precisa saber. Techtudo, 2018. Disponível em: https://www.techtudo.com.br/noticias/2018/03/facebook-e-cambridge-analytica-sete-fatos-que-voce-precisa-saber.ghtml. Acesso em: 10 dez. 2019.

BBC NEWS. Como os dados de milhões de usuários do Facebook foram usados na campanha de Trump. [S. l.], 2018. Disponível em: https://www.bbc.com/portuguese/geral-43705839. Acesso em: 10 dez. 2019.

DONEDA, Danilo. A PROTEÇÃO DOS DADOS PESSOAIS COMO UM DIREITO FUNDAMENTAL. [S. l.], 2011. Disponível em: http://editora.unoesc.edu.br/index.php/espacojuridico/article/view/1315/658. Acesso em: 13 dez. 2019.

DONEDA, Danilo. Da privacidade à proteção de dadospessoais. [S. l.], 2006. Disponível em: http://egov.ufsc.br/portal/sites/default/files/anexos/29536-29552-1-PB.pdf. Acesso em: 13 dez. 2019.

PRESIDÊNCIA DA REPÚBLICA CASA CIVIL SUBCHEFIA PARA ASSUNTOS JURÍDICOS. Lei nº LEI Nº 12.737, de 30 de novembro de 2012. Dispõe sobre a tipificação criminal de delitos informáticos; altera o Decreto-Lei nº 2.848, de 7 de dezembro de 1940 - Código Penal; e dá outras providências. Lei Carolina Dieckmann, [S. l.], 2012. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12737.htm. Acesso em: 10 jul. 2019.

G1. Lei ‘Carolina Dieckmann‘, que pune invasão de PCs, entra em vigor. [S. l.], 2013. Disponível em: http://g1.globo.com/tecnologia/noticia/2013/04/lei-carolina-

dieckmann-que-pune-invasao-de-pcs-passa-valer-amanha.html. Acesso em: 10 jul. 2019.

SRPEO.GOV.BR. O que são dados pessoais, segundo a LGPD. [S. l.], 2019. Disponível em: https://www.serpro.gov.br/lgpd/menu/protecao-de-dados/dados-pessoais-lgpd. Acesso em: 20 jan. 2020.

SERPRO.GOV.BR. Fique por dentro das palavras e termos-chave que dão suporte à Lei Geral de Proteção de Dados Pessoais. [S. l.], 2019. Disponível em: https://www.serpro.gov.br/lgpd/menu/a-lgpd/glossario-lgpd. Acesso em: 20 jan. 2020.

MORAES, José Ricardo Maia. Anonimização de dados é fundamental para que PME's estejam em conformidade com a LGPD. [S. l.], 2019. Disponível em: https://cryptoid.com.br/identidade-digital-destaques/anonimizacao-de-dados-e-fundamental-para-que-pmes-estejam-em-conformidade-com-a-lgpd/. Acesso em: 23 fev. 2020.

MICROSOFT. Descrição da criptografia simétrica e assimétrica. [S. l.], 2018. Disponível em: https://support.microsoft.com/pt-br/help/246071. Acesso em: 6 mar. 2020.

PRESIDÊNCIA DA REPÚBLICA CASA CIVIL SUBCHEFIA PARA ASSUNTOS JURÍDICOS. Lei nº LEI Nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º , no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. Acesso às informações públicas, [S. l.], 2011. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 6 mar. 2020.

PRESIDÊNCIA DA REPÚBLICA CASA CIVIL SUBCHEFIA PARA ASSUNTOS JURÍDICOS. Lei nº LEI Nº 13.709, de 14 de agosto de 2017. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). Lei Geral de Proteção de Dados Pessoais (LGPD). Lei Geral de Proteção de Dados Pessoais (LGPD), [S. l.], 2017. Disponível em:

http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em: 14 abr. 2020.

KHALIL, Mohammad; EBNER, Martin. De - Identification in Learning Analytics. Journal of Learning Analytics, [s. l.], 2016. Disponível em: https://epress.lib.uts.edu.au/index.php/JLA/article/view/4519/5435. Acesso em: 20 mar. 2020.

DIRECTORATE C (FUNDAMENTAL RIGHTS AND UNION CITIZENSHIP) OF THE EUROPEAN COMMISS. ARTICLE 29 DATA PROTECTION WORKING PARTY. DE-IDENTIFICATION AND LINKAGE OF DATA RECORDS, [S. l.], p. 3,37, 10 abr. 2014. Disponível em: https://epress.lib.uts.edu.au/index.php/JLA/article/view/4519/5435. Acesso em: 20 mar. 2020.

GILBERT, Eric S.; EVANS, Kathi S.; CLARK, Troy S. DE-IDENTIFICATION AND LINKAGE OF DATA RECORDS. [S. l.], 10 abr. 2012. Disponível em: https://patentimages.storage.googleapis.com/15/b6/bb/923028a652ae22/US20020073138A1.pdf. Acesso em: 22 mar. 2020.