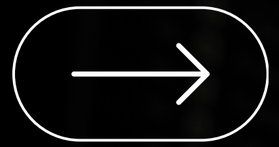
Análise de Violação de Dados

CIÊNCIA DE DADOS E BIG DATA



Integrantes

Ana Flavia Lorêdo - 23025092
Fernanda Mayumi Kuba Kato - 23024484
Guilherme Medeiros- 23024555
Kevin Makoto Shiroma - 20020925
Lorena Bernardo - 23025048
Matheus Sampaio Duarte - 23024588
Paulo Carvalho - 23024564
Renato Riichi Kato - 23024516

Agenda.

1. DESCRIÇÃO DO PROJETO
2. EXPLORANDO OS DADOS
3. TABELAS DE DADOS PROCESSADOS E SINTÉTICOS
4. ANÁLISES E IMPACTO
5. CONCLUSÃO
6. BIBLIOGRAFIA

Descrição do Projeto

A Data Safe AI é um projeto de Startup de Consultoria de Segurança da Informação com a missão de fortalecer a segurança cibernética de pequenas e médias empresas. Nossa proposta é desenvolver soluções personalizadas que auxiliem as PMEs na identificação e gestão de riscos, garantindo a proteção de seus dados e operações. Utilizando inteligência artificial, buscamos simplificar a segurança digital, tornando-a acessível e eficaz, adaptada às necessidades específicas de cada cliente.

Explorando os dados.

Dataset

ESTE CONJUNTO DE DADOS, DISPONÍVEL EM FORMATO CSV, ABRANGE O PERÍODO DE 2004 A 2022 E ANALISA VAZAMENTOS DE DADOS QUE COMPROMETEM 30.000 OU MAIS REGISTROS DE EMPRESAS DOS ESTADOS UNIDOS. ELE DESTACA DIVERSAS ORGANIZAÇÕES DE DIFERENTES TAMANHOS E SETORES QUE FORAM ALVO DE ATAQUES CIBERNÉTICOS. A MAIORIA DOS INCIDENTES É ATRIBUÍDA A INVASÕES, EVIDENCIANDO A VULNERABILIDADE GENERALIZADA ENFRENTADA POR ESSAS ENTIDADES.

TOTAL DE ATAQUES

ATAQUES

349

Ataques realizados entre 2004 e 2022.

DADOS

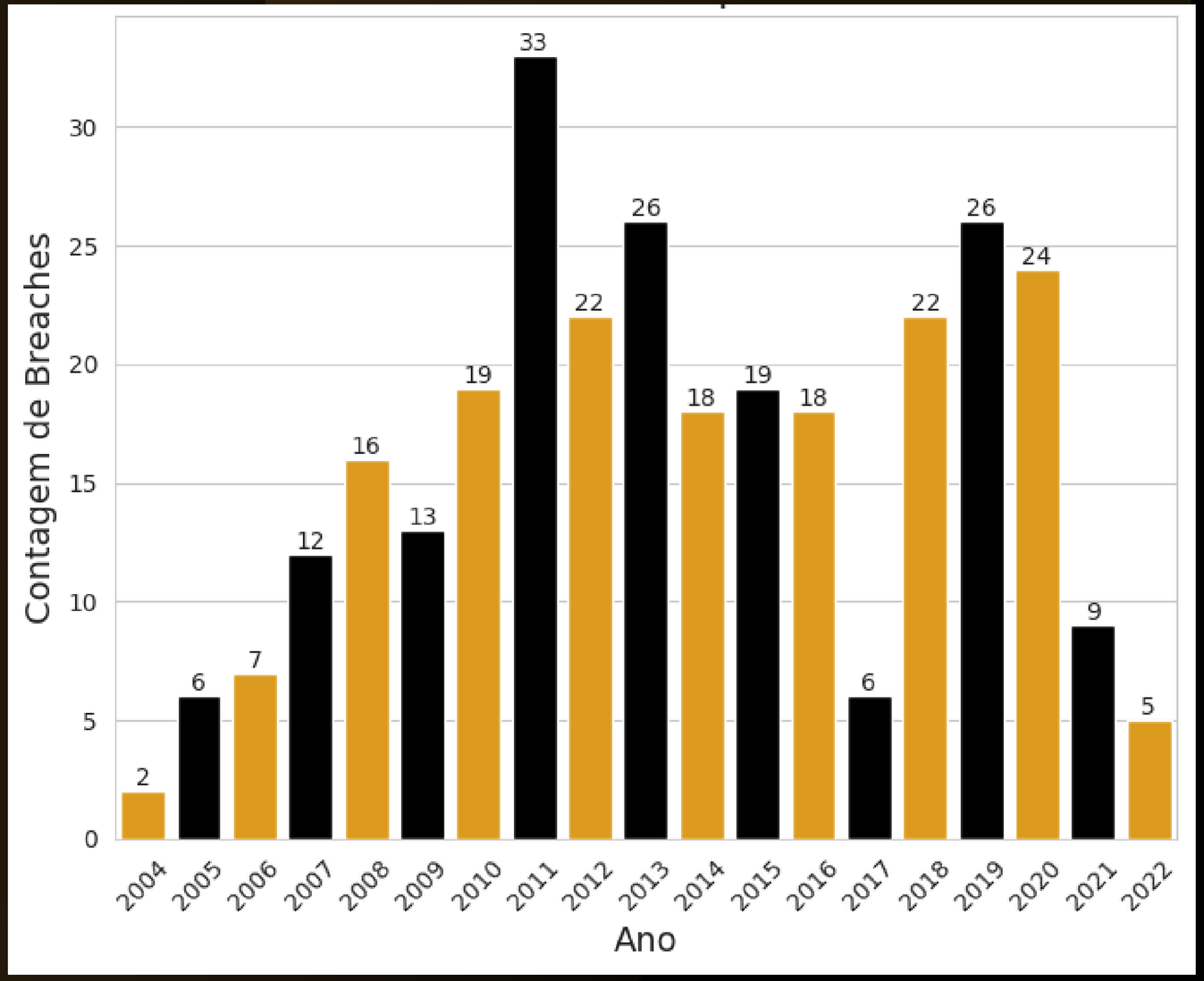
+13bi

Total de dados vazados.

Tabelas de dados processados e sintéticos.

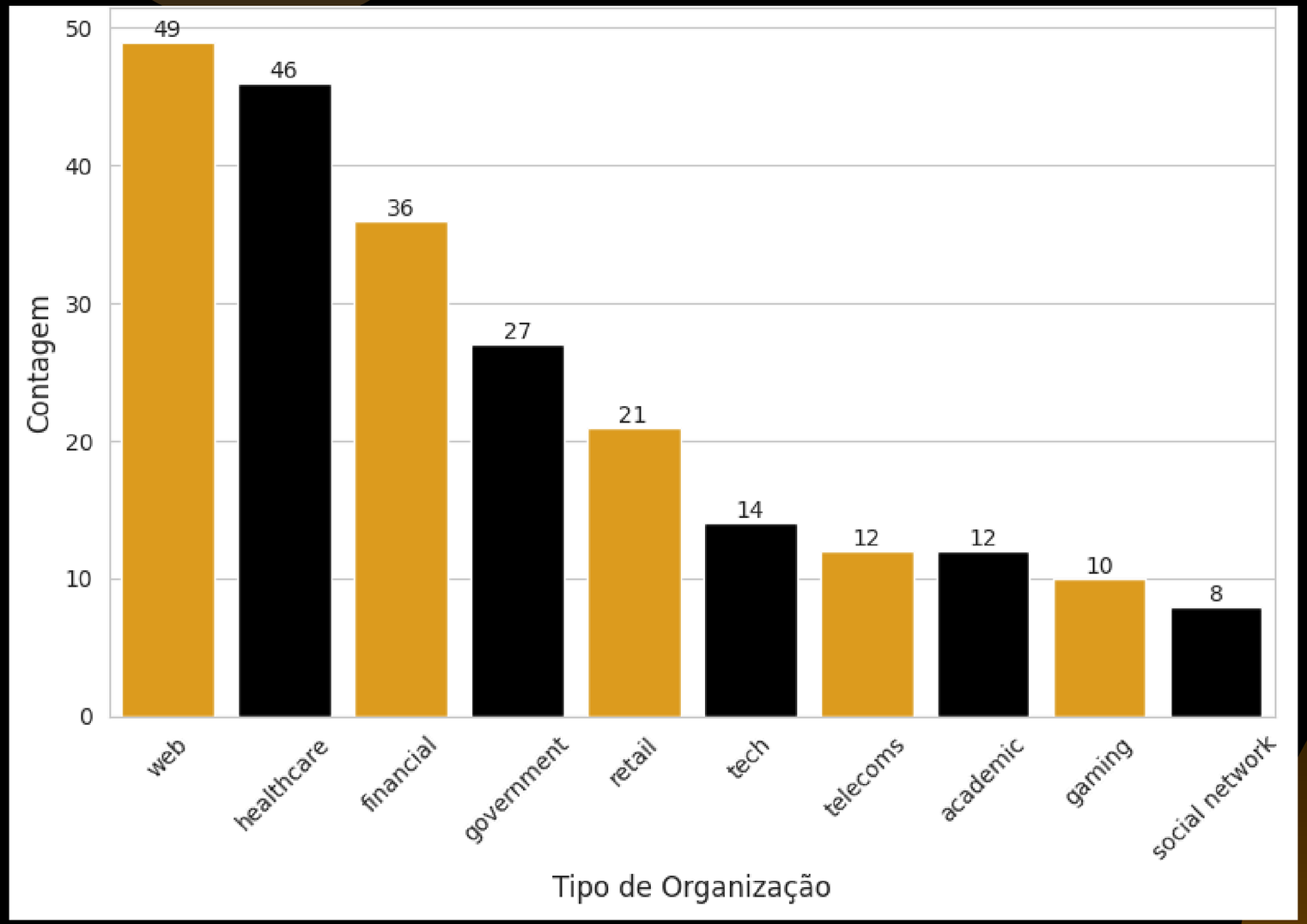
	Unnamed: 0	Entity	Year	Records	Organization type	Method	categoria	nivel_ameaca
0	0	21st Century Oncology	2016	2200000	healthcare	hacked	Incidentes relacionados a hacking	alto
1	1	500px	2020	14870304	social networking	hacked	Incidentes relacionados a hacking	alto
2	2	Accendo Insurance Co.	2020	175350	healthcare	poor security	Ameaças internas	baixo
3	3	Adobe Systems Incorporated	2013	152000000	tech	hacked	Incidentes relacionados a hacking	alto
4	4	Adobe Inc.	2019	7500000	tech	poor security	Ameaças internas	baixo
...	...	...	...	...	...	...	...	...
344	347	Zynga	2019	173000000	social network	hacked	Incidentes relacionados a hacking	alto
345	348	Unknown agency(believed to be tied to United S...	2020	200000000	financial	accidentally published	Exposição de dados/Má configuração	medio
346	349	National Health Information Center (NCZI) of S...	2020	391250	healthcare	poor security	Ameaças internas	baixo
347	350	50 companies and government institutions	2022	6400000	various	poor security	Ameaças internas	baixo
348	351	IKEA	2022	95000	retail	accidentally published	Exposição de dados/Má configuração	medio

Total de Ataques por Ano



ESTE GRÁFICO MOSTRA A EVOLUÇÃO DO NÚMERO DE ATAQUES CIBERNÉTICOS AO LONGO DOS ANOS. PODEMOS OBSERVAR PICOS SIGNIFICATIVOS EM 2011, 2013 E 2019, INDICANDO QUE ESTES FORAM OS ANOS QUE HOUVERAM MAIOR NÚMERO DE REGISTROS DE ATAQUES REALIZADOS.

Top 10 Tipos de Organizações mais Afetadas



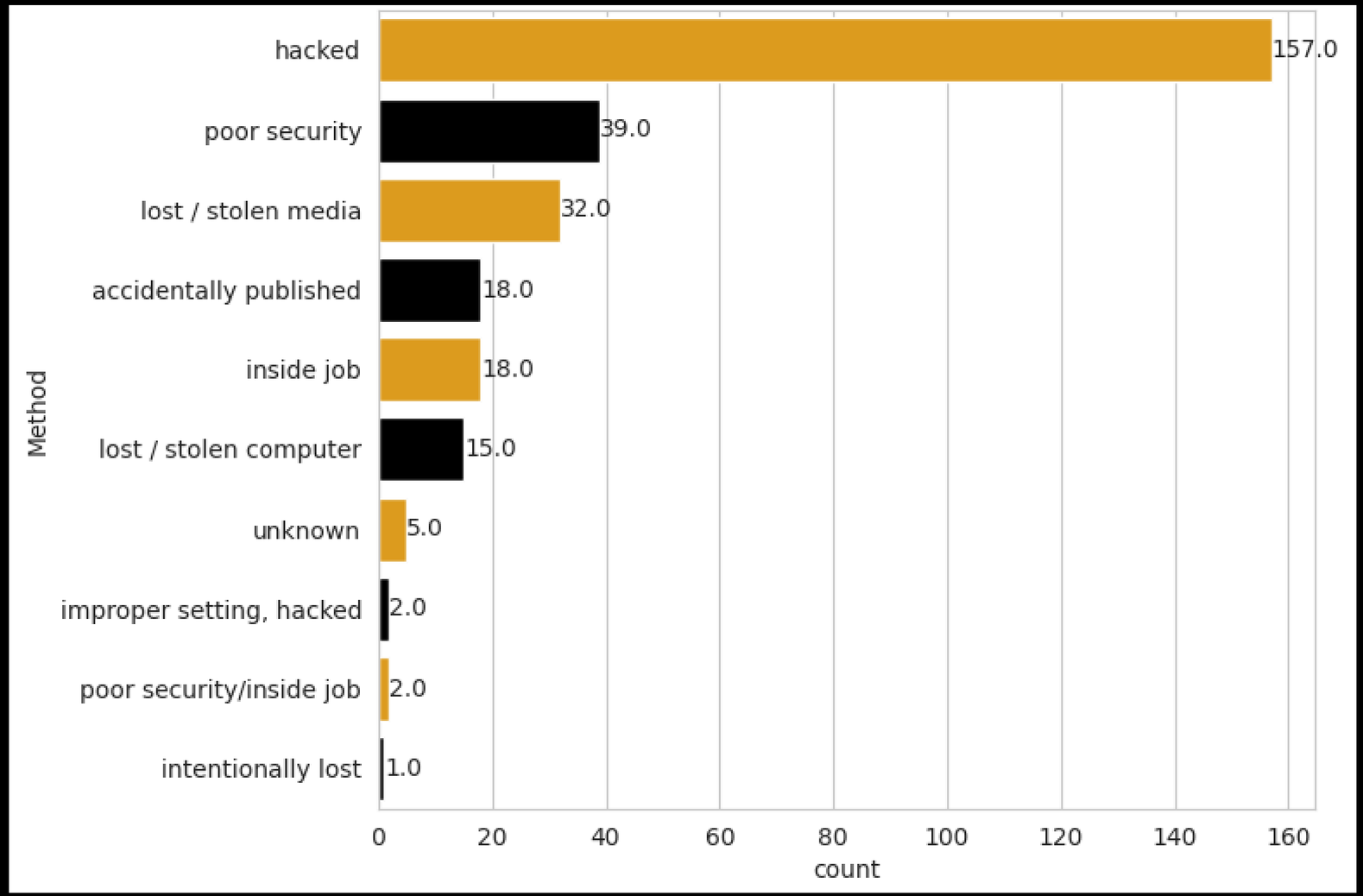
AS ÁREAS MAIS VULNERÁVEIS A ATAQUES CIBERNÉTICOS SÃO CLARAMENTE EVIDENCIADAS, COM O SETOR WEB SENDO O MAIS IMPACTADO COM 49 REGISTROS, SEGUIDO PELOS SETORES DE SAÚDE COM 46 E O FINANCEIRO COM 36. ESSES DADOS REVELAM QUE ESSAS ÁREAS TÊM SIDO OS PRINCIPAIS ALVOS DE ATAQUE.

A QUANTIDADE DE REGISTROS NESSES SETORES RESSALTA A NECESSIDADE DE MEDIDAS DE SEGURANÇA MAIS RIGOROSAS, ESPECIALMENTE EM SEGMENTOS QUE DEPENDEM FORTEMENTE DE TECNOLOGIAS ONLINE.

Métodos

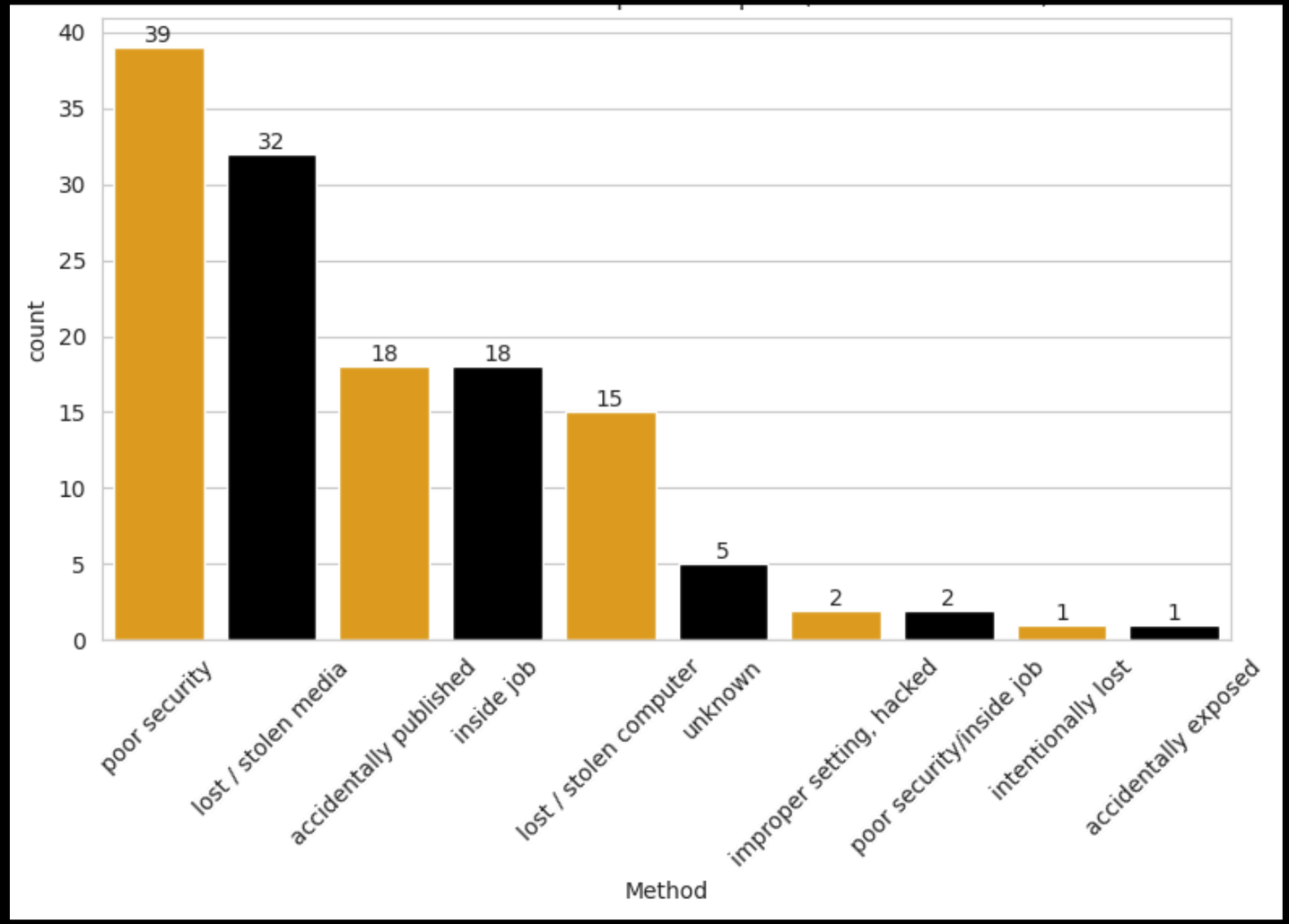
Métodos de ataque mais utilizados para a violação de dados.

Podemos observar as categorias que representam as diferentes maneiras pelas quais as informações (registros) foram expostas ou comprometidas.

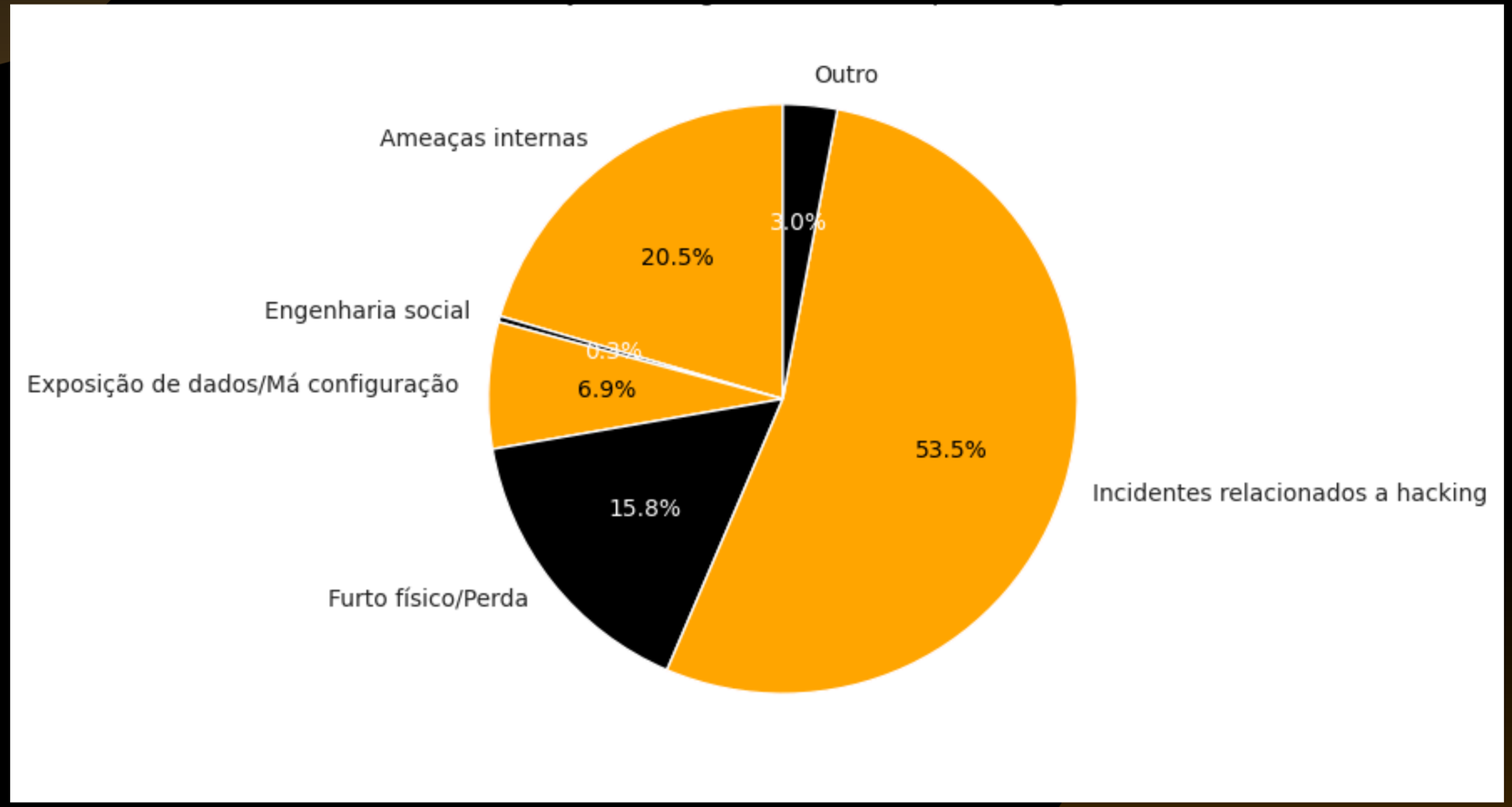


Métodos

Métodos, exceto Hacking, de ataque mais utilizados para a violação de dados.



Distribuição de Registros Afetados por Categoria

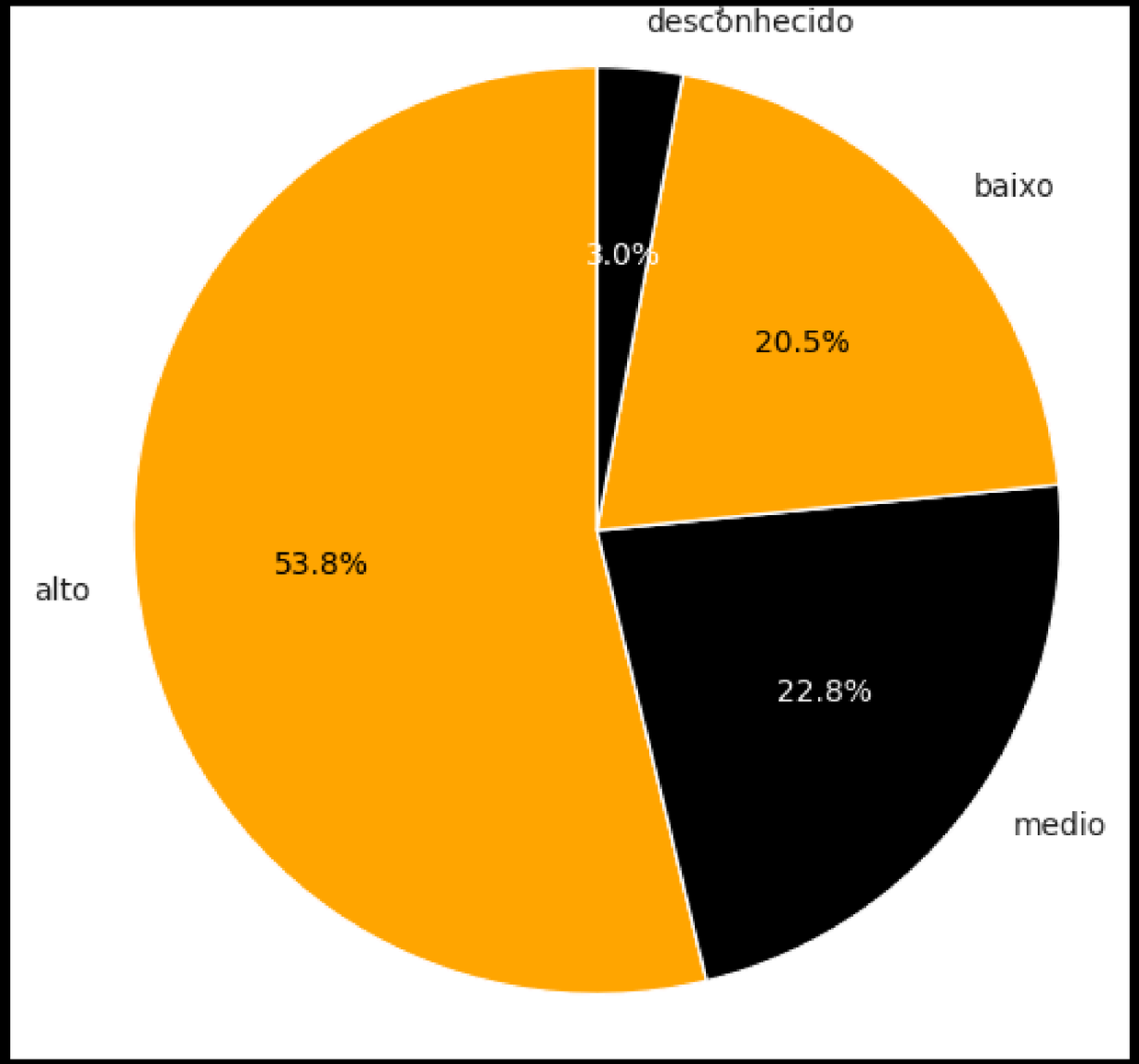


O gráfico mostra que 53,5% dos registros afetados são relacionados a ataques de hacking, indicando que essa categoria representa a maior parte das violações de segurança. Essa alta porcentagem destaca a urgência de implementar medidas eficazes de proteção contra ataques cibernéticos.

Nível da Ameaça

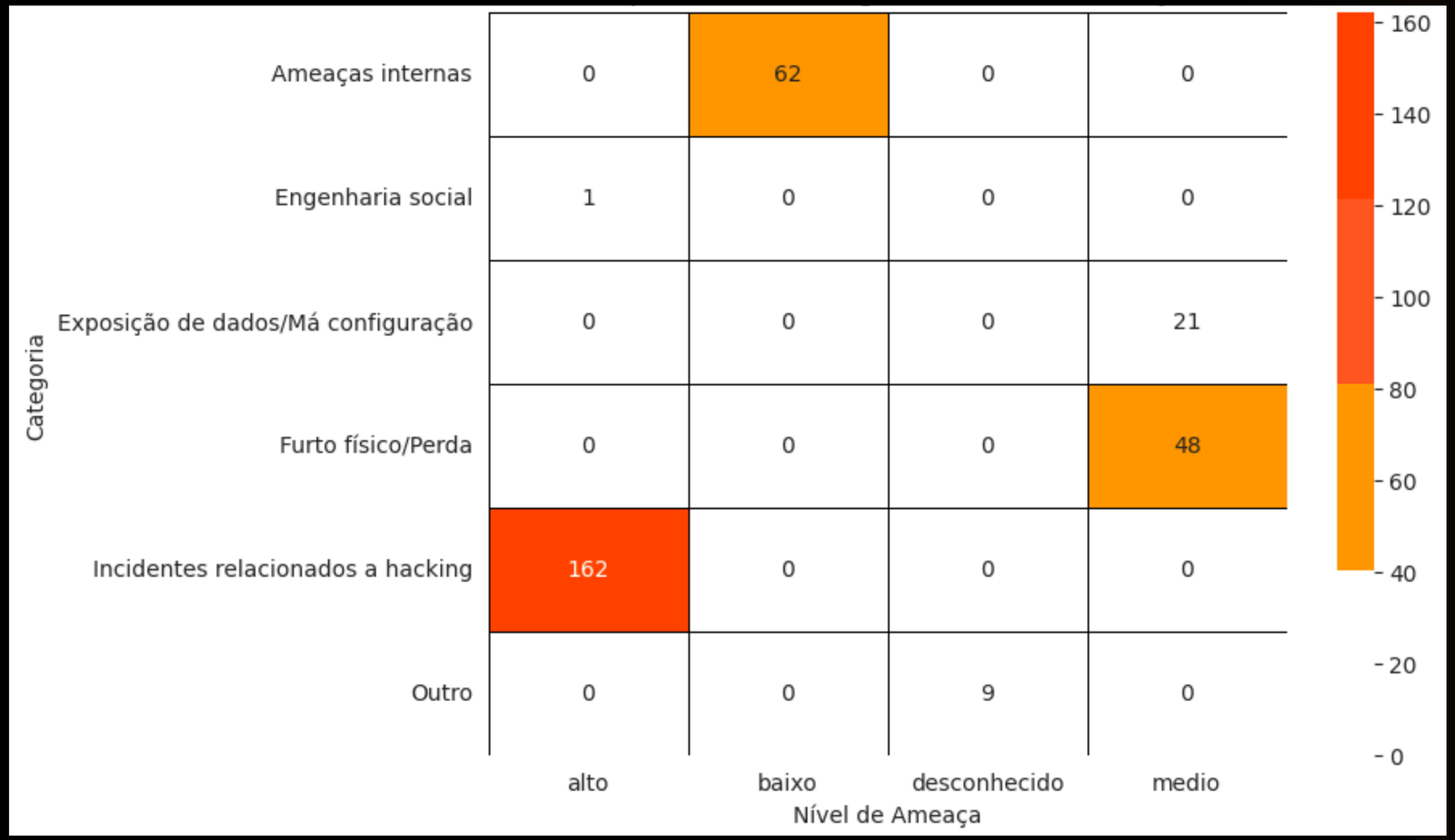
O gráfico de pizza mostra a distribuição dos níveis de ameaça, com 53,8% dos casos classificados como alto, 22,8% como médio, 20,5% como baixo e 3,0% como desconhecido.

A alta porcentagem de ameaças de nível elevado destaca a necessidade de medidas eficazes de mitigação e resposta para proteger sistemas e dados críticos.



Mapa de Calor

Categoria x Nível de Ameaça



O mapa de calor mostra que a maior parte dos incidentes está relacionada a Hacking no nível alto, com 162 casos. Ameaças internas são mais preocupante no nível baixo, com 62 ocorrências, enquanto Furto físico/Perda se destaca no nível médio, com 48 incidentes.

Já a Engenharia social e Outros apresentam baixa incidência. Isso indica que os esforços de segurança devem ser focados em ataques de hacking e nas ameaças identificadas.

Análise de impacto

Como analisados anteriomente, as empresas mais afetadas em **números de ataques cibernéticos**, foram do setor "Web (53)", "Healthcare (47)" e "Financial (38)" durante o período de 2004-2022.

Mas, quando analisamos mais a fundo o impacto de **Records** (Registros) em **números vazados**, nos seguintes setores "Web", "'Social networking" e "Tech".

Cerca de **5 bilhões** de Records foram vazados no **setor Web,** que se destaca nas duas analises, tanto em número de ataque, quanto em número de Records vazados. O mesmo abrange empresas como por exemplo: **Yahoo, Wattpad, Ebay** e etc. Ferramentas onde são armazenados dados de milhares de pessoas, como e-mail, nome completo e no caso do Ebay, também pode conter dados de cartão de crédito e documentos de identificação governamentais dos usuários.

Esses dados vazados podem ser comercializados sem autorização dos usuários, **impactando gravemente suas vidas a curto e longo prazo quando caem em mãos erradas.**

Análise de impacto

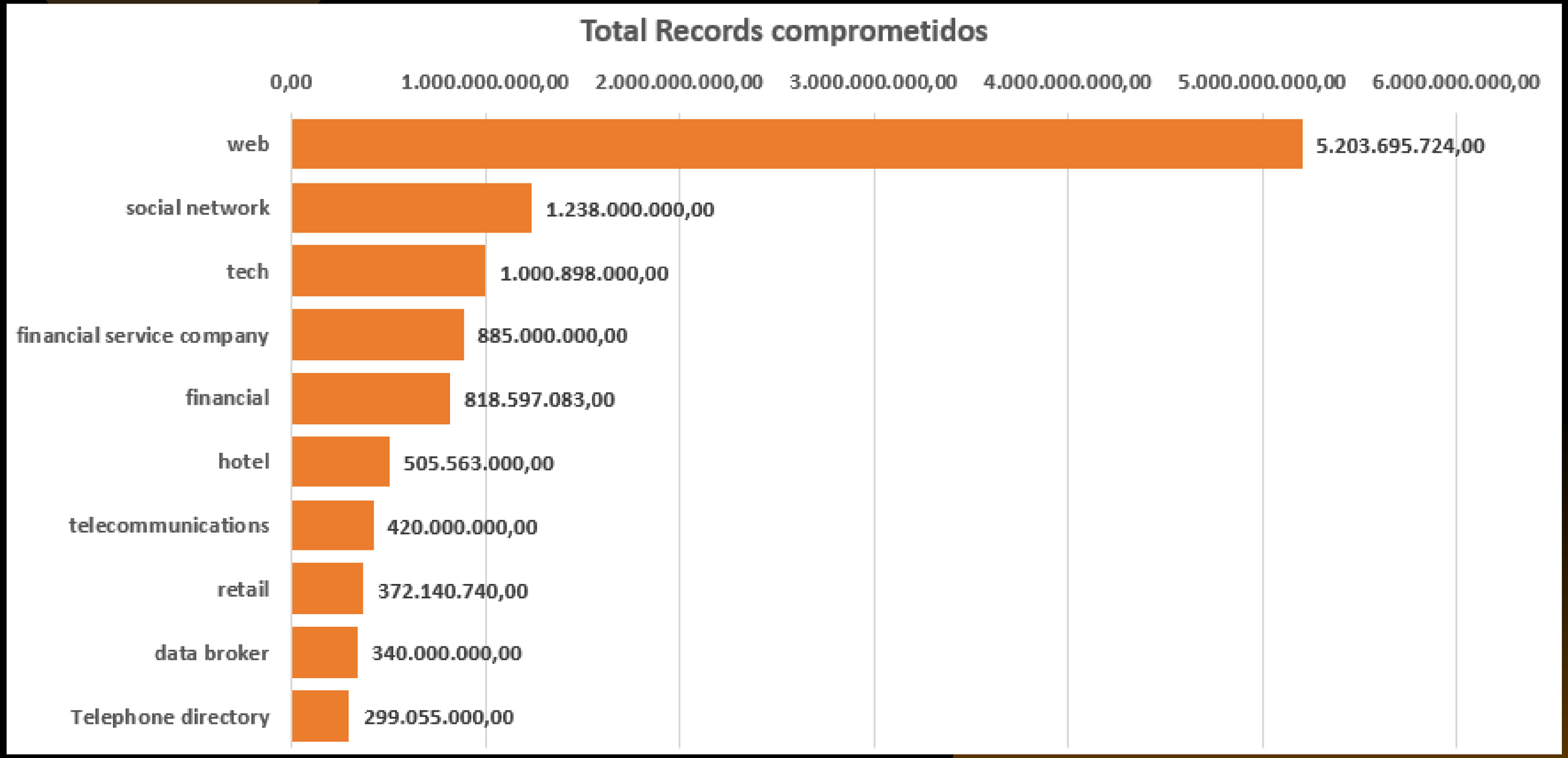


*Gráfico extraído da base de dados do Excel utilizando tabela dinâmica.

Análise de impacto

Perda de Confiança dos Clientes: Um ataque cibernético pode resultar na exposição de dados pessoais dos clientes, como nomes, endereços e informações de pagamento. Essa violação de dados causa um profundo impacto na reputação da empresa, levando à perda de confiança e fidelidade dos clientes.

Dificuldades Financeiras: Os custos associados a um ataque cibernético podem ser significativos, incluindo gastos com recuperação de dados, reparo de sistemas, notificação de clientes, multas e indenizações. Além disso, a interrupção das operações pode levar à perda de receita e danos à cadeia de suprimentos.

Danos à Reputação da Marca: A exposição negativa na mídia e nas redes sociais pode causar danos irreparáveis à reputação da empresa. A imagem de uma empresa comprometida pode levar à perda de oportunidades de negócios e investimentos.

Análise de impacto

Aferimos a partir destes dados que, a maior consequência relacionada a estes ataques em geral, é o **impacto humano** aos afetados. Para a empresa os danos são principalmene em relação à **reputação da marca** e ao setor **financeiro** da organização, quando arca com processos e indenizações, ocasionando também uma exposição à mídia sobre o caso.

Yahoo revela acordo para indenizar consumidores em US$ 47 milhões por vazamentos de dados

Consumidores prejudicados se juntaram em processo contra a empresa nos Estados Unidos. Companhia ocultou violação por dois anos.

Conclusão.

Com este estudo concluímos que referente aos ataques, foram registrados mais de 300 incidentes de violação de dados, resultando no comprometimento de mais de 13 bilhões de registros. O intervalo entre 2011 e 2020 destacou-se pela intensificação de ataques cibernéticos, especialmente no setor “web”, que é vulnerável devido à sua dependência de tecnologias online e ao armazenamento de grandes volumes de dados sensíveis dos usuários.

Destaca-se o ataque hacker, correspondendo a mais de 50% dos dados catalogados em nossa base, impactando na perda de confiabilidade dos clientes, resultando em questões financeiras, processos, multas, gastos em recuperação de dados e indenizações que podem ser aplicadas como no caso do Yahoo, além de ter a grande possibilidade da queda de reputação da marca com a exposição negativa, com possíveis danos irreparáveis

Bibliografia.

Violação de Dados - Kagle

https://www.kaggle.com/datasets/thedevastator/data-breaches-a-comprehensive-list?resource=download

https://g1.globo.com/economia/tecnologia/blog/altieres-rohr/post/2018/09/21/yahoo-revela-acordo-para-indenizar-consumidores-em-us-47-milhoes-por-vazamentos-de-dados.ghtml