

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
ESCOLA POLITÉCNICA E DE ARTES
GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO



MATHEUS GOMES DOS SANTOS ARAÚJO

**USO DE FERRAMENTAS DE BUSINESS INTELLIGENCE PARA IDENTIFICAÇÃO
DE CONTRATOS COM FORNECEDORES INIDÔNEOS OU SUSPENSOS**

GOIÂNIA 2024

MATHEUS GOMES DOS SANTOS ARAÚJO

**USO DE FERRAMENTAS DE BUSINESS INTELLIGENCE PARA IDENTIFICAÇÃO
DE CONTRATOS COM FORNECEDORES INIDÔNEOS OU SUSPENSOS**

Trabalho de Conclusão de Curso apresentado à Escola Politécnica e de Artes da Pontifícia Universidade Católica de Goiás, como parte dos requisitos para a obtenção do título de Bacharel em Engenharia de Computação.

Orientador : André Luiz Alves

Banca Examinadora :

- Prof. Me. Ronaldo Lopes de
Oliveira

- Prof. Dr. Vicente Paulo de
Camargo

GOIÂNIA, 2024

MATHEUS GOMES DOS SANTOS ARAÚJO

**USO DE FERRAMENTAS DE BUSINESS INTELLIGENCE PARA IDENTIFICAÇÃO
DE CONTRATOS COM FORNECEDORES INIDÔNEOS OU SUSPENSOS**

Este Trabalho de Conclusão de Curso julgado adequado para obtenção do título de Bacharel em Engenharia da Computação, e aprovado em sua forma final pela Escola Politécnica e de Artes, da Pontifícia Universidade Católica de Goiás, em 17/06/2024.

Banca examinadora :

Orientadora: Prof. André Luiz Alves

Examinador: Prof. Me. Ronaldo Lopes de Oliveira

Examinador: Prof. Dr. Vicente Paulo de Camargo

GOIÂNIA, 2024

DEDICATÓRIA

Dedico este trabalho de conclusão de curso a todas as pessoas que acreditaram em mim e me apoiaram ao longo dessa jornada acadêmica. Agradeço em especial aos meus familiares e amigos, que estiveram sempre presentes, oferecendo suporte emocional e encorajamento.

Também dedico este trabalho aos meus professores e orientadores, que me guiaram e me ensinaram tanto durante esses anos de estudo. Seus conhecimentos, experiências e conselhos foram essenciais para o sucesso deste trabalho.

Por fim, dedico este trabalho à minha própria jornada de aprendizado e crescimento pessoal. Agradeço por todas as lições e desafios que encontrei ao longo do caminho, pois foram eles que me fizeram chegar até aqui.

RESUMO

Ano após ano, os dados se fazem mais presentes e úteis, crescendo em proporção geométrica, fazendo-se presente nos mais diversos segmentos com uma importância incalculável. Para extrair valor das massas de dados, muitas das vezes sem padrões e bagunçadas, se fez necessário o entendimento e estudo da ciência de dados, proporcionando o surgimento de técnicas, conceitos e ferramentas extremamente valiosas, para a extração dessas bases de dados. Nesse cenário, Big Data surge proporcionando o processamento de grandes bases de dados. Business Intelligence permite entender as regras dos mais variados negócios e enxergar padrões nessas massas de informações. Paralelo ao conhecimento dessas ferramentas, enxergou-se em órgãos públicos a criação de grandes massas de dados no segmento de contratos em licitações, com a necessidade de tratamento, para evitar prejuízos com dinheiro público. Nos últimos 5 anos, licitações públicas ultrapassaram 50 trilhões de reais em contratos. Mesmo evitando a mínima quantidade possível de contratos fraudados ou de risco, os valores resultantes ainda são grandes devido à proporção desses contratos. Portanto esse trabalho visa mostrar o processo de implementação de técnicas e ferramentas com o intuito de mapear contratos de licitações públicas com um fornecedor sancionado ou inidôneo de alguma maneira.

Palavras chaves: *Big Data*, *Business Intelligence*, Licitações Públicas, Ciência de Dados, *Data Mining*.

ABSTRACT

Year after year, data has become more prominent and useful, growing exponentially and proving to be invaluable across various sectors. To extract value from these often unstructured and chaotic data sets, understanding and studying data science became essential. This led to the development of highly valuable techniques, concepts, and tools for extracting insights from these data sets. In this context, Big Data emerged, enabling the processing of large volumes of data. Business Intelligence allows us to understand business rules and identify patterns within these vast information sets. Recognizing the accumulation of large data sets in public procurement, public agencies saw the need for proper data management to prevent financial losses. Over the past five years, public tenders have surpassed 50 trillion reais in contracts. Despite efforts to minimize fraudulent or risky contracts, the resulting amounts are still substantial due to the scale of these contracts. Therefore, this work aims to demonstrate the process of implementing techniques and tools to map public procurement contracts with potential risks.

Keywords: *Big Data, Business Intelligence, Public Tenders, Data Science, Data Mining.*

LISTA DE ABREVIATURAS

ACFE – Association of Certified Fraud Examiners

BI – Business Intelligence

CEIS – Cadastro de Empresas Inidôneas e Suspensas

CEO – *Chief Executive Officer*

CNPJ – Cadastro Nacional de Pessoa Jurídica

CPF – Cadastro de Pessoa Física

CSV – *Comma-Separated Values*

DDL – *Data Definition Language*

DML – *Data Manipulation Language*

DCL – *Data Control Language*

GAECO – Grupo de Atuação Especial de Repressão ao Crime Organizado

LGPD – Lei Geral de Proteção de Dados

SAS – *Statistical Analysis System*

SQL – *Structured Query Language*

Lista de Figuras

1.	Figura 1 - Representação Simplificada de Ciência de dados	14
2.	Figura 2 - Dashboard exemplo construído em Power BI	20
3.	Figura 3 - Evolução de Contratações	23
4.	Figura 4 -Fluxo do passo a passo da Metodologia	25
5.	Figura 5 -Trecho de Código - Importação de dados	27
6.	Figura 6 -Trecho de código - Unificação de tabelas	27
7.	Figura 7 - Trecho de Código - Tratamento de CNPJ e Join entre tabelas	28
8.	Figura 8 - Tela do Power Query	30
9.	Figura 9 - Editor do Power BI	31
10.	Figura 10 - Dashboard de Contratos	32
11.	Figura 11 - Dashboard de Contratos com fornecedor sancionado	33
12.	Figura 12 - Big Numbers do Dashboard	34
13.	Figura 13 - Gráficos visuais de informações aglutinadas	34
14.	Figura 14 - Dashboard filtrado pela seleção de gráfico	35

SUMÁRIO

1. 10

2. 14

2.1. **Error! Bookmark not defined.**

2.2. 19

2.3. 21

3. 23

4. 25

4.1. 25

4.2. 26

4.3. 27

4.4. 29

5. 31

5.1. 35

6. Conclusão

33

1. INTRODUÇÃO

Esse trabalho visa buscar uma interpretação e entendimento do uso de ferramentas de Business Intelligence, que de maneira resumida é uma metodologia de interpretação de cenários de negócios com implementação de técnicas e ferramentas, que auxiliam na produção de informações que auxiliam na tomada de decisão empresarial. Com ênfase na busca e evidenciação de contratos de órgãos públicos que contenham risco, pelo fato de estarem atribuídos a fornecedores inidôneos ou sancionados. Mas antes entendamos a importância do estudo de dados em nosso mundo.

A princípio, pensou-se na escrita como uma forma de registro, de modo a complementar a tradição oral e com isso preservar a memória e a história. Assim, a humanidade evoluiu para outros métodos mais eficientes de preservação de informação. A principal ideia é ter locais seguros para armazenar informações importantes que serão usadas posteriormente (Algar, 2023). O armazenamento de dados em bancos de dados ou sistemas tem como objetivo otimizar o acesso e controle das informações. Segundo Ajay Banga, Ex-CEO da Mastercard e presidente do Banco Mundial, os dados são o novo petróleo da Terra, com a diferença de que o petróleo vai acabar, os dados não, originalmente *“I believe that data is the new oil. I am saying it in this country because I believe that the prosperity that oil brought in the last 50 years, data will bring in the next 50, 100 years if you use it the right way,”* (Ajay Banga, 2006). Essa afirmação evidencia o valor e a importância dos dados para o presente e futuro, tendo como base a importância que o petróleo possui para o desenvolvimento econômico da sociedade atual, e os dados tendem a seguir o mesmo percurso com o passar do tempo.

Com o passar dos anos, o armazenamento de dados possuíram inúmeros mídias distintas. Desde a fita magnética, disco rígido, disquetes flexíveis, e mais recentemente popularizado o pen drive. Serviram de preparação para os armazenamentos estruturados e tecnológicos de hoje em dia. Com a chegada dos serviços de armazenamento em nuvem, a área de Ciência de Dados se encontra em meio a mais uma revolução, no entanto dessa vez com uma visão definitiva, devido a capacidade de escalabilidade da tecnologia.

Nos anos 2000, os dados armazenados digitalmente, eram estimados na faixa de 25%. No entanto, por volta de 2014 essa porcentagem já se apresentava na casa de 98% dos dados gerais armazenados, restando menos de 2% para os registros armazenados fisicamente em forma de papel (MACHADO, 2019).

Tendo em vista que os dados foram armazenados em diferentes mídias, foi necessário o surgimento de uma tecnologia, com o objetivo de unificar e trazer toda essa informação para um local apenas. Uma solução para isso foi encontrada com o surgimento do *Big Data*.

Aliado a isso o trabalho do Analista de Dados surgiu e se fez de alta valia. O analista de dados é responsável por decifrar esses grandes armazenamentos de informação, por meio de técnicas e análises a fim de traduzir o emaranhado de informações aparentemente desconexas em dados claros, que orientem em decisões estratégicas.

Segundo Bruna Giroldo da PUCPR (Giroldo,2020), *Big Data* é uma análise e união de grandes volumes distintos de dados, que contém muitas informações valiosas para muitas empresas. Essas informações ajudam na organização de processos, minimizando gastos e impulsionando lucros, conseqüentemente, por meio do trabalho de predição e construção de modelos.

Contudo, todo o poder dos dados e da tecnologia, dentro das empresas, abre-se margem para falhas ou fraudes humanas, com o trabalho de armazenamento feito de maneira não concisa e segura. Ano após ano, fraudes utilizando inconsistências de sistemas dos mais variados segmentos, representam perdas cada vez mais significativas para as empresas (Medrado, 2016). Segundo a pesquisa global de Identidade e Fraude de 2020, 57% das empresas estão enfrentando perdas crescentes devido a fraudes ano após ano (Malar, 2023).

Além disso, a ACFE (Associação de Examinadores de Fraude Certificados) gerou um relatório que afirma que empresas que produzem um trabalho preventivo dos dados, no quesito fraudes em possíveis rendas monetárias, consegue identificar possíveis perdas, 58% de maneira mais eficaz que organizações que não o fazem, e sofrem 52% menos prejuízos financeiros do que organizações sem essa preocupação (ACFE, 2018). Como mostram os números, compreender os seus dados e saber como

analisá-los de forma eficaz pode trazer benefícios significativos aos esforços de prevenção e detecção de fraudes das organizações.

O combate às fraudes se tornou algo primordial para as empresas, segundo estudo desenvolvido na Juniper Research, em todo o planeta até 48 bilhões de dólares podem ter sido alvos de prejuízo por fraudes ou falhas no ano de 2023. Sendo um valor maior que o dobro do registrado no ano de 2018 de apenas 22 bilhões de dólares.

Às estratégias de Business Intelligence, junto de uma equipe de Analistas de Dados dedicados a identificar e combater essas fraudes e falhas, é possível desenvolver modelos preditivos e de regressão de dados fraudulentos, para identificação desses registros em meio a grandes massas de dados. No entanto, mesmo com muitas empresas que já identificaram a importância dos dados em suas operações, infelizmente uma pequena parcela dessas organizações possuem algum trabalho na identificação e combate das falhas e fraudes.

Foi identificado que no Brasil somente por licitações de órgãos públicos, foram firmados mais de 50 trilhões de reais em contratos nos últimos três anos, segundo o portal de transparência do governo federal. Juntamente com essa informação muitos contratos estão com CNPJ de empresas que estão cadastrados no CEIS(Cadastro Nacional de Empresas Inidôneas e Suspensas), o que pode indicar um risco na resolução desses contratos e principalmente um prejuízo aos cofres públicos (PORTAL DE TRANSPARÊNCIA, 2024).

Portanto com esse estudo será desenvolvida uma pesquisa e demonstração do trabalho de instalação de ferramentas e de identificação de contratos de risco em licitações públicas, por meio de técnicas de *Business Intelligence*, percorrendo bases extensas de dados a fim de demonstrar contratos de licitações públicas com risco. No cenário estudado, surge então uma questão, se for analisado os contratos de licitações públicas ativos, encontraremos contratos com fornecedores cadastrados com algum tipo de sanção ?

Para sanar a questão anterior, é necessário adotar uma metodologia, de entender todos os conceitos, e principalmente depois conduzir o trabalho de desenvolvimento do painel com a massa de dados citada anteriormente, encontrada no portal de transparência do Governo Federal. Portanto para esse trabalho

seguiremos esse passo a passo :

- 1 - Explicar conceitos e vantagens do uso de *Big Data* e Business Intelligence.
- 2 - Conduzir uma análise de uma base de dados real de licitações públicas.
- 3 - Fazer entendimento de regras do processo licitatório Brasileiro.
- 4 - Conduzir um tratamento nos dados.
- 5 - Construir um painel intuitivo indicando os resultados.
- 6 - Identificar desafios e obstáculos na implementação de soluções de *Business Intelligence* para prevenção de fraudes e propor possíveis soluções.

2. CIÊNCIA DE DADOS

As inovações tecnológicas passaram a ser substanciais na vida do ser humano, visto que o grau de utilidade destes dispositivos se tornou indispensável, pois diversos indivíduos utilizam dispositivos tecnológicos como uma extensão de seus *Selves* (BELK, 1988). Nesse cenário surge a Ciência de Dados, um conceito novo, mas extremamente presente na modernidade.

Ciência de Dados se faz importante pois integra ferramentas, métodos e tecnologias para extrair significado dos dados (AWS, 2024). As organizações contemporâneas lidam com uma inundação de dados, provenientes de uma multiplicidade de dispositivos capazes de coletar e armazenar informações automaticamente. Portais online e sistemas de pagamento, por exemplo, capturam uma vasta gama de dados em setores como comércio eletrônico, saúde, finanças e outras esferas da vida cotidiana. Esses dados abrangem texto, áudio, vídeo e imagem em quantidades massivas.

Cientistas de Dados são bastante associados à Cientistas da computação e principalmente Estatísticos. E para efeitos práticos essa comparação é bem colocada, pois esse profissional soma o melhor dessas duas áreas, a visão e estratégias estatísticas com o poder do processamento computacional. Resumindo, Ciência de dados, é uma área interdisciplinar que atua na intersecção entre 3 áreas de estudo, com intuito de extrair o máximo de insights valiosos das massas de dados (Alura, 2023), a figura 1 exemplifica um pouco esse pensamento.

Figura 1 - Representação Simplificada de Ciência de dados



Fonte : Alura, 2023

A Ciência da Computação fornece as ferramentas e técnicas para gerenciar grandes volumes de dados, incluindo programação e tecnologias de machine learning. Matemática e Estatística são essenciais para analisar e interpretar os dados, aplicando modelos estatísticos e compreendendo padrões. O Conhecimento de Negócio é vital para aplicar essas análises no mundo real, ajudando a resolver problemas específicos e tomar decisões informadas.

Uma maneira de entender o trabalho de um cientista de dados, é por meio de uma comparação como um minerador, alguém que vai fazer escavações em um ambiente “hostil” em busca de algo bastante valioso. A tradução dessa analogia é o trabalho de extrair conhecimento de dados desorganizados.

Há mais de cinquenta anos, a Ciência de Dados tem evoluído com base em ideias e iniciativas focadas no aprendizado a partir dos dados. Atualmente, essa área do conhecimento se caracteriza por ser multidisciplinar, integrando métodos matemáticos e estatísticos com técnicas computacionais para a modelagem de problemas em diversos campos. A Ciência de Dados apresenta grandes perspectivas quanto às suas aplicações no contexto das organizações públicas (ENAP, 2022).

Nos últimos anos, os avanços tecnológicos têm impulsionado uma enorme produção de dados e o desenvolvimento de métodos analíticos. Paralelamente, surgem desafios complexos que exigem recursos diversificados para ampliar a compreensão e encontrar soluções. A Ciência de Dados surge como uma aliada fundamental nesse contexto, direcionando o uso crescente de dados, para produzir informações mais robustas e orientadas à ação estatal, também conhecidas como evidências para embasar políticas públicas. (ENAP, 2022)

Vários benefícios podem ser apontados no uso da Ciência de Dados para subsidiar a tomada de decisão nas políticas públicas, tais como a expansão dos dados simultaneamente analisados a respeito de diferentes dimensões e áreas de política (FILGUEIRAS, 2022)

Contudo é evidente que governos ao redor do mundo têm adotado a Ciência de Dados para aprimorar serviços e políticas públicas, visando reduzir riscos de fraudes e evasão fiscal, entre outros benefícios. (ENAP, 2022)

O uso de dados e informações gerado pelos analistas não ocorre de forma imediata e direta pelos tomadores de decisão, ele depende de várias etapas e condições para que seja bem-sucedido. Entende-se então que Ciência de Dados é o trabalho que soma técnicas computacionais e estatísticas com o foco em resolver problemas complexos, usando para isso grandes conjuntos de dados (ALURA, 2023).

2.1. Big Data

Big Data é uma análise e união de grandes volumes distintos de dados, que contém muitas informações estratégicas, úteis e principalmente valiosas para muitas empresas (Farol, 2020). *Big Data* é um conjunto de dados maior e mais complexo, especialmente de novas fontes de dados. Esses conjuntos de dados são tão volumosos que o software tradicional de processamento de dados simplesmente não consegue gerenciá-los.

O estudo e implementação do *Big Data*, possui características bem claras que evidenciam boa parte das vantagens que esse conceito pode proporcionar, são chamados de os 7V's do *Big Data*, eles são:

1. **Volume:** Big Data refere-se à enorme quantidade de dados gerados e armazenados continuamente. A característica de volume é central para Big Data, pois envolve o processamento de grandes quantidades de informações. Esses dados são coletados de diversas fontes, como redes sociais, transações comerciais, sensores de IoT (Internet das Coisas), entre outros, resultando em uma vasta quantidade de dados heterogêneos. Para lidar com essa imensidão de dados, utiliza-se softwares e técnicas específicas que permitem a análise eficiente e rápida, possibilitando insights valiosos e tomada de decisões embasadas;
2. **Velocidade:** Milhares de novos dados chegam à rede a cada segundo. Pessoas conectadas sempre fizeram coisas como compartilhar e curtir coisas de interesse nas redes sociais, fazer compras e postar seus dados em sites. Esses novos dados alimentam o sistema, que ao ser cruzado e analisado gera novas informações sobre o mercado;
3. **Vínculo:** O vínculo em Big Data refere-se à capacidade de conectar e relacionar dados provenientes de diferentes fontes para criar um contexto mais amplo e significativo. Em um ambiente de Big Data, os dados não existem isoladamente; eles

estão interligados, e a habilidade de vincular essas informações permite uma compreensão mais profunda e abrangente. Por exemplo, relacionar dados de comportamento de clientes com dados de transações financeiras e interações em redes sociais pode revelar padrões e insights que não seriam visíveis ao analisar esses dados separadamente. Ferramentas e técnicas específicas são utilizadas para estabelecer essas conexões, possibilitando uma análise mais rica e informada;

4. Variedade: O *Big Data* é capaz de processar tanto informações estruturadas (como planilhas e tabelas) quanto informações não estruturadas (como imagens e vídeos). Esse tipo de dados não estruturados é muito comum na web, com isso se torna uma grande melhoria ao passar do tempo e o avanço da tecnologia;

5. Variabilidade: A variabilidade em Big Data refere-se à inconsistência dos dados, que podem vir em diferentes formatos e de várias fontes. Por exemplo, dados de redes sociais podem variar em linguagem e contexto. Essa característica torna essencial a utilização de ferramentas que possam lidar com a variabilidade, garantindo a consistência e a qualidade das análises;

6. Veracidade: A veracidade em Big Data diz respeito à qualidade e confiabilidade dos dados. Nem todos os dados coletados são precisos ou úteis, e a presença de informações errôneas pode levar a conclusões incorretas. Portanto, é crucial utilizar métodos que verifiquem a precisão e a integridade dos dados para assegurar que as análises sejam baseadas em informações confiáveis;

7. Visualização: A visualização em Big Data é a capacidade de representar graficamente grandes conjuntos de dados de maneira compreensível e intuitiva. Ferramentas de visualização ajudam a transformar dados complexos em gráficos, mapas e dashboards interativos, facilitando a identificação de padrões, tendências e insights que podem ser rapidamente compreendidos por diferentes públicos.

Sem um tratamento estratégico e organizado dos dados, ao invés de ser uma fonte de compreensão, se torna uma enorme fonte de dados dispersos e sem utilidade em qualquer tipo de armazenamento de dados, mas no *Big Data* esses problemas tomam uma escala ainda maior.

Com base no que foi afirmado, a integração e gerenciamento dos dados, se

mostra muito importante, a fim de prover uma coleta de dados úteis e validados com qualidade e confiabilidade, além disso é de muita importância uma boa organização para que os dados coletados sejam utilizados de maneira inteligente tendo em vista as necessidades do negócio(IBM, 2024).

Alguns dos casos mais valiosos de uso do Big Data, onde sistemas de gerenciamento de banco de dados tradicionais falhariam em processar eficientemente, incluem a manutenção preditiva, que prevê falhas em equipamentos antes que ocorram(Iberdrola, 2022); a eficiência operacional, que otimiza processos e melhora a produtividade analisando dados em tempo real (TOTVS, 2023); o machine learning, que utiliza grandes volumes de dados para treinar algoritmos e automatizar tarefas complexas(ENTERPRISE,2023); e, especialmente relevante para esta pesquisa, a verificação de fraude e conformidade, onde o Big Data detecta padrões suspeitos e anomalias em transações financeiras, ajudando a identificar fraudes e garantir a conformidade com regulamentações.

Existem inúmeras organizações que fazem uso dessa tecnologia para esses fins, no Brasil. Grandes empresas que lidam bastante com esse modelo de solução são os bancos. Diretor da EMC Brasil, declarou ao canal CDTV, que um dos principais produtos na época de crise, para a equipe de tecnologia, foram as soluções de *Big Data* para solução de fraudes e análise de seus clientes (CDTV,2015).

Às vezes o conceito de *Big Data* pode ser confundido com o de BI (*Business Intelligence*), mas são conceitos diferentes. BI é mais focado na análise de dados para ajudar na tomada de decisão estratégica dos seus usuários, fazendo uso de ferramentas e técnicas para organizar, analisar e apresentar esses dados de maneira clara ao cliente (DATAEX, 2021), e Big Data nesse âmbito serve como uma ferramenta que vai lidar com a extração e processamento dessas massas de dados, para o profissional que está conduzindo o processo.

Contudo não são conceitos antagônicos, o trabalho de *Business Intelligence* e *Big Data* são bastante agregadores um ao outro, pois muitos analistas de BI, utilizam de ferramentas de *Big Data* para conseguirem processar as mais distintas e extensas massas de dados em busca de suas análises.

2.2. Business Intelligence

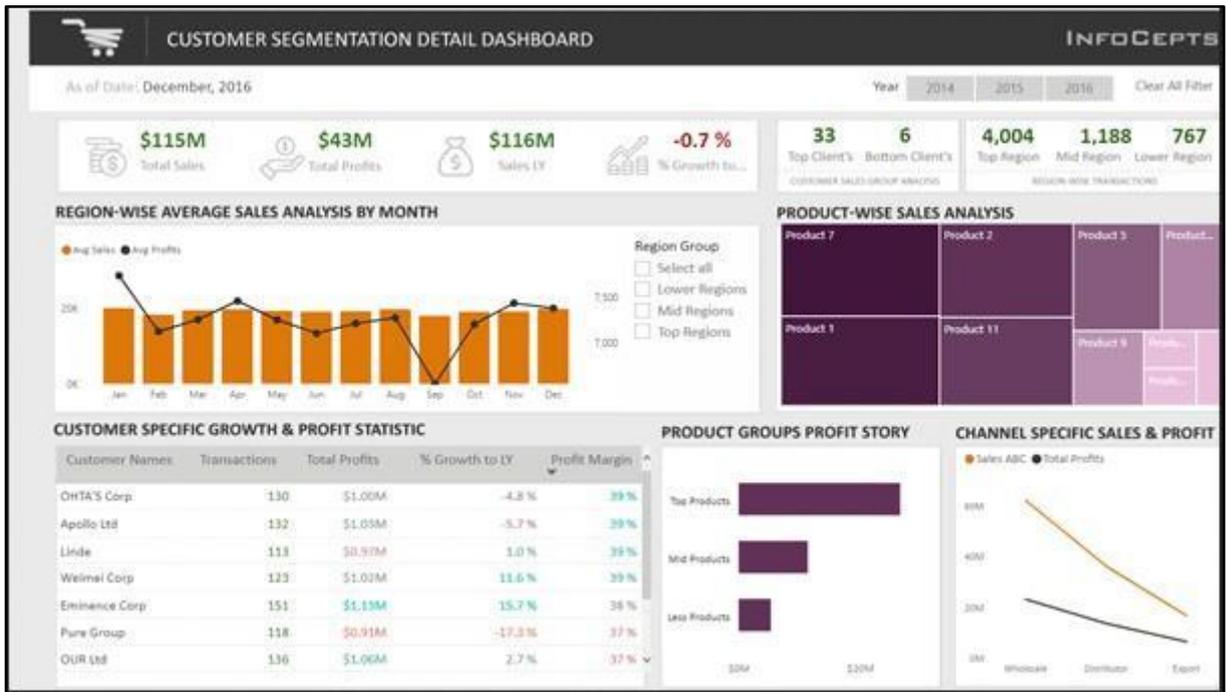
Business Intelligence, também conhecido como BI, é uma metodologia de tomada de decisões empresariais voltadas à análise racional de informações, pode ser usada por empresas para oferecer suporte a uma ampla gama de decisões de negócios, desde operacionais à estratégicas (JUCEMG, 2024).

O “BI” se faz característico, por suas produções possuir 4 requisitos extremamente valiosos, confiabilidade, consistência, disponibilidade e capacidade preditiva.

- **Confiabilidade:** Diretamente ligada à qualidade dos dados de origem. Quanto melhores forem os dados, melhores serão os resultados e insights.
- **Consistência:** O BI tem como função fornecer análises precisas ao longo do tempo. Isso é alcançado a partir de um processo robustos de coleta, transformação e atualização de dados
- **Disponibilidade:** Um sistema de BI eficaz deve garantir que os usuários autorizados tenham acesso rápido e fácil às análises e relatórios relevantes
- **Capacidade preditiva:** O BI fornece insights, não somente do passado e presente, por meio de técnicas, também é capaz de prever tendências do futuro

Ferramentas de *Business Intelligence* que se destacam de maneira positiva, com suas peculiaridades, existem *Power BI* e SAS(Statistical Analysis System). *Power BI* é uma plataforma de análise de dados da Microsoft que permite visualizar, transformar e compartilhar insights poderosos por meio de painéis interativos e relatórios dinâmicos (4infra, 2024) como demonstrado o exemplo na figura 2. SAS é uma poderosa suíte de software para análise avançada de dados, oferecendo soluções abrangentes para mineração, manipulação e interpretação de informações complexas, com foco em análises estatísticas e preditivas (MARCELA,2023).

Figura 2 - Dashboard exemplo construído em Power BI



Fonte : FBS Consultoria, 2024

Quando se trata de falha na segurança, não necessariamente está se tratando de hackers mal-intencionados, mas sim podendo ser especialistas e funcionários da própria área. Os cenários de segurança e requisitos de conformidade estão evoluindo constantemente. *Business Intelligence* pode ser útil na identificação de padrões de grandes massas de dados, mesmo que esses padrões sejam a exceção como no caso das fraudes (Becker, 2018) . A identificação da fraude, falha ou risco, vai variar do cliente ou produto, por isso é tão importante o entendimento e levantamento de requisitos e regras de maneira bastante criteriosa.

Somado com isso ferramentas de programação como R e Python, também se tornam úteis para o trabalho de *Data Mining*. Por meio do *Data Mining* é possível navegar nos bancos de dados, para encontrar dados que serão úteis para o trabalho de BI (CETAX, 2022). *Data Mining* nada mais é do que técnicas para explorar grandes conjuntos de dados em busca de padrões, relações e insights valiosos, como :

- Identificar padrões complexos;
- Prever comportamentos futuros;
- Detectar Anomalias ou outliers;
- Otimização de processos.

Data Mining ou mineração de dados, se trata de uma fase de descoberta das informações, onde existem 3 fases bastante importantes: coleta de dados, pré-processamento e pós-processamento de dados. Contudo, para uma melhor performance do trabalho é importante o atendimento de um pré-requisito : conhecer o contexto em que os dados estão inseridos e como eles ocorrem nesse contexto (SILVA, 2016).

2.3. Linguagens de Consulta

Para percorrer, controlar e filtrar essas massas de dados extensas, é necessário um conhecimento técnico específico que automatize ações de consulta de maneira eficiente, precisa e padronizada.

Nesse Contexto surge SQL, (Structured Query Language) é uma linguagem para interface com bancos de dados relacionais, isto é, todos os usuários e programas que desejarem realizar alguma tarefa no banco de dados devem fornecer comandos escritos nesta linguagem. (GONÇALVES, 2015).

Dentre os conceitos dominados por SQL estão DDL (Data Definition Language), DML (Data Manipulation Language) e DCL (Data Control Language), que somados permitem um controle e manipulação das bases de dados (CALASAN,2022), por mais extensas que forem, da maneira que o usuário desejar.

- DDL : são responsáveis por definir tabelas e outros objetos pertencentes ao banco de dados. Dentre os principais comandos da DDL estão: CREATE TABLE, CREATE INDEX, ALTER TABLE, DROP TABLE, DROP VIEW E DROP INDEX (GONÇALVES, 2015);
- DML: Esta parte da SQL possui comandos para que os usuários façam acesso e armazenamento dos seus dados. Fazem parte da DML os seguintes comandos: INSERT, DELETE, UPDATE, SELECT E JOIN' (GONÇALVES, 2015);
- DCL: Para finalizar, temos a DCL, que é subdividida em três grupos de comandos: comandos para controle de transação, sessão e sistema. Para os comandos de controle, podem-se citar os mais usados, que são: COMMIT, ROLLBACK, GRANT e REVOKE.(GONÇALVES, 2015).

Esses conceitos e linguagem, são pontos primordiais para o trabalho que será

conduzido no projeto. Com eles será percorrida, bases de dados de contratos públicos, tratada, filtrada e unificada com informações distintas, originadas de diferentes fontes.

3. CENÁRIO DE ANÁLISE

Em uma análise de contratos firmados por órgãos públicos no Brasil no portal de transparência demonstrado na figura 3, foram encontrados valores muito altos. Com a surpresa dos valores surge uma dúvida, quantos desses contratos tiveram algum tipo de complicação na entrega de suas obrigações por parte do fornecedor ? Existe alguma maneira de mapear possíveis riscos nesses contratos ? Portanto foi iniciado um estudo para identificar regras para diferenciar os contratos que fogem dos padrões e estão “fora da curva”.

Figura 3 - Evolução de Contratações

Evolução histórica das contratações	
ANO	VALOR CONTRATADO
2013	R\$ 232.850.829.276,65
2014	R\$ 183.981.180.504,82
2015	R\$ 177.300.826.047,12
2016	R\$ 114.457.330.836,92
2017	R\$ 156.055.285.983,86
2018	R\$ 78.205.498.497,32
2019	R\$ 40.885.324.627,68
2020	R\$ 74.233.463.259,68
2021	R\$ 54.520.835.100.592,89
2022	R\$ 87.489.706.220,51
2023	R\$ 81.633.616.587,07
2024	R\$ 16.147.288.252,75
Total	R\$ 55.764.075.450.687,27

Fonte : Portal de Transparência, 2024

Esse trabalho de detecção de Anomalias, representa de maneira sucinta o trabalho desse projeto em identificação de fraudes e falhas, com foco em órgãos públicos. Pois, nas bases de dados que serão exploradas a maioria dos dados estarão dentro de um padrão, os registros que fogem da curva, estarão de alguma maneira fraudulentos ou com risco por isso denominados *outliers* (pontos fora da curva).

Fraude é um ato astuto, enganoso e malicioso destinado a prejudicar ou enganar outra pessoa, a fim de dar ao fraudador algum tipo de vantagem (geralmente financeira) sobre a vítima. O artigo 171 do Código Penal Brasileiro afirma que a fraude abrange uma área complexa com infrações e penalidades de natureza diversa. Pode causar danos irreparáveis à vítima, sejam eles financeiros, psicológicos ou mesmo de

imagem (Tchilian, 2023) .

Usando como referência, organizações públicas, são fechados milhares de contratos para que nos últimos 5 anos ultrapassem a casa dos 50 trilhões de reais segundo o portal de transparência do governo (<https://portaldatransparencia.gov.br>). Em contraponto, são cadastrados milhares de fornecedores por algum descumprimento básico ou grave de contratos firmados no CEIS (Cadastro de empresas Inidôneas e Suspensas), e ainda assim existem contratos ativos com esses fornecedores suspensos. Isso abre uma área de estudo para interpretar se esses contratos possuem algum tipo de fraude ou erro (PORTAL DE TRANSPARÊNCIA, 2024).

Estudando notícias de investigações e operações federais, entende-se que os valores somados de desvios ou contratos fraudados, somam centenas de milhões de reais nos últimos anos. Exemplo, em 2023 GAECO, cumpriu mandado de busca e apreensão por fraudes em licitações, que somavam R\$15 Milhões em uma única operação no município de Anápolis (PORTELA,2023).

Devido a alta demanda e geração de dados em licitações públicas de todos os estados brasileiros, o *Big Data* se mostra ainda mais propício para soluções neste âmbito, pelo fato de conseguir processar grandes massas de dados. Além do trabalho de *data mining* e técnicas de BI para interpretar a necessidade, o funcionamento e o fluxo, para que seja possível identificar padrões nesses *outliers*, os contratos fechados por licitação pública com alguma fraude ou apenas um risco em sua finalização.

A combinação dos conhecimentos de Ciência de dados, *Big Data* e *Business Intelligence*, visa construir um painel que expõe de maneira compacta e intuitiva maneiras de enxergar padrões nos dados coletados, e o usuário por meio desse ambiente conseguir tirar conclusões ou caminhos para a solução, e não para apontar de fato um órgão, fornecedor ou funcionário como criminoso.

4. APLICAÇÃO

Este capítulo visa mostrar o processo de aplicação das técnicas e uso de ferramentas, além de explicar regras que foram utilizadas no processo com trechos de código e imagens das ferramentas .

4.1. Metodologia

Passo a Passo da Metodologia:

Figura 4 -Fluxo do passo a passo da Metodologia



Fonte : O Autor

- Preparação da base de dados:

Coletar os dados, organizá-los e modificar caso necessário, excluindo colunas inúteis para o processo, para evitar processamento desnecessário. Esse trabalho é produtivo pois evita muito retrabalho posteriormente nas fases futuras. Esse trabalho foi melhor conduzido e explicado nas seções 4.2 e 4.3.

- Regras de Negócio:

Analisar as leis de licitações públicas, e identificar questões que se repitam de alguma maneira, para auxiliar na organização da identificação de pontos fora da curva. Nesse trabalho em específico foi uma questão mais teórica, de entender o processo licitatório e cadastro de sanções, um exemplo do que foi feito no código foi a comparação da vigência do contrato, com a vigência da sanção do fornecedor, para entender se de fato o fornecedor estava sancionado no período de vigência do contrato, mas pode ser melhor entendido com as informações da seção 4.4.

- Identificação de Anomalias :

Incluir filtros na tabela de dados colhida e adaptada, a fim de exibir aqueles que atendem às regras levantadas na atividade anterior. Esse trabalho foi uma simples aplicação de filtro no painel construído, com base nos dados que foram inseridos na tabela final.

- Relatórios e Dashboards :

Montar relatórios e dashboards com a tabela de dados encontrada de registros com algum risco de fraudulência, e apresentar esses dados de maneira ordenada e intuitiva. Esse trabalho está melhor representado na seção 4.4.

4.2. Extração dos dados

Para iniciar o trabalho foi feito pesquisas com intuito de localizar informações de contratos licitatórios, a princípio para entender suas regras, leis e dados armazenados.

Foi encontrado portanto informações sobre contratos no site do portal de transparência do governo brasileiro : <https://portaldatransparencia.gov.br>. Com essas informações foram extraídas 2 fontes de dados. A primeira com informações de contratos e a segunda de fornecedores cadastrados com sanções no CEIS (Cadastro de Empresas Inidôneas e Suspensas) .

Esses dados estavam todos em planilhas excel, a princípio os dados de contratos estavam em planilhas separadas, cada uma representando contratos de anos diferentes, e apesar de possuírem muitas informações úteis, essas planilhas estavam desorganizadas e possuíam milhares de registros. Os arquivos de fornecedores vieram em formato .CSV (é um tipo de arquivo que possui suas informações separadas por “,”, por isso o nome “Comma-separated values”) e apresentavam as mesmas características dos dados de contrato.

Dito isso, evidencia-se a necessidade ferramentas que processam *Big Data* para consumir e unificar os dados em um local apenas de trabalho, além de técnicas de *Business Intelligence* com intuito de extrair e construir regras sobre o negócio em questão, busca de padrões de contratos com alguma espécie de risco. As ferramentas escolhidas portanto que permitem a abordagem citada, foram o Power BI e SAS (Statistical Analysis System).

4.3. Tratamento de Dados

Com posse dos registros, citados, todos foram importados na plataforma SAS (Statistical Analysis System), onde seria construído o código para tratamento das informações. O código desenvolvido utiliza linguagem SQL e trechos SAS.

A linguagem SQL (Linguagem de Consulta Estruturada) “é uma linguagem padrão para manipulação de registros em bancos de dados relacionais” (CALASAN,2022), e linguagem SAS é própria para ambientes SAS, possui o mesmo objetivo de SQL, mas formatos e termos distintos.

Para Importar os dados foi utilizado formatos seguindo os padrões da Figura 5:

Figura 5 -Trecho de Código - Importação de dados

```
filename csv3 "/home/servico/Git/Matheus-Gomes//sancoes_2.csv" encoding="utf-8";  
  
proc import datafile=csv3 out=work.IMPORT2  
  (RENAME=(VAR1=CPF_CNPJ VAR2=NOME_RAZAO VAR3=CADASTRO  
          VAR4=UF VAR5=ORGAO VAR6=CATEGORIA VAR7=DATA_PUB  
          VAR8=VALOR_MULTA VAR9=QUANTIDADE) DROP= VAR10) dbms=csv  
  replace;  
  getnames=no;  
  delimiter=",";  
  datarow=2;  
  GUESSINGROWS=MAX;  
run;
```

Fonte : O Autor

O trecho da Figura 5 seleciona o arquivo original na pasta em que está armazenado, aplica um decodificador padrão, e em seguida atribui para cada coluna da planilha o nome que será mais útil e claro no desenvolvimento da solução.

Em seguida, com todas as planilhas importadas no sistema, para as tabelas semelhantes foram utilizados no trecho de código SAS a seguir, que nada mais faz do que unificar as diversas tabelas em uma só.

Figura 6 -Trecho de código - Unificação de tabelas

```
1 ⊖ DATA WORK.SancoesCEIS;  
2     SET WORK.IMPORT WORK.IMPORT1 WORK.IMPORT2;  
3     RUN;
```

Fonte : O Autor

Com as tabelas disponíveis de Contratos e Fornecedores com cadastro no CEIS seria necessário unificá-los com identificador único, portanto foi utilizado o registro de CPF ou CNPJ de fornecedor, que está disponível em ambas as tabelas

Figura 7 - Trecho de Código - Tratamento de CNPJ e Join entre tabelas

```
PROC SQL;
  CREATE TABLE WORK.CnpjLimpo as
    select unique compress(a.CPF_CNPJ,,"kd") as CPF_CNPJ,
           a.*
  FROM WORK.SancoesCEIS as a;
quit;

PROC SQL;
  CREATE TABLE WORK.IMPORT3 as
    select unique compress(a.CnpjFornecedor,,"kd") as CnpjFornecedor,
           a.*
  FROM WORK.IMPORT3 as a;
quit;

PROC SQL;
  CREATE TABLE WORK.ContratosCEIS AS
  SELECT
    *
  FROM
    IMPORT3 AS A
  LEFT JOIN
    CnpjLimpo AS B
  ON
    A.CnpjFornecedor=B.CPF_CNPJ
  WHERE CnpjFornecedor is not missing and CPF_CNPJ is not missing;
QUIT;
```

Fonte : O Autor

No entanto, como dificuldade, foram encontrados formatos distintos no CPF ou CNPJ, pois na tabela de Fornecedores com sanções encontrava-se dados com caracteres especiais, tipo : XX. XXX. XXX/0001-X, enquanto nos contratos os mesmos dados apresentavam-se no formato XXXXXXXX0001XX.

Foi implementado então código para manter somente os números nessas colunas, seguindo então um único formato e portanto poderão ser comparados. O código citado é uma função padrão da linguagem : "compress(a.CPF_CNPJ,,"kd")".

Após uma padronização da chave única que será utilizada para unificação das tabelas por meio do CNPJ ou CPF do fornecedor do contrato. É possível então unificar

as tabelas, para então podermos visualizar todos os contratos que possuem como fornecedor, um representante com algum tipo de sanção ativa, ou seja, podem representar algum tipo de risco na entrega ou conclusão desse contrato firmado.

Essa informação é de alta valia, pois, se for possível guardar-se de contratos com possibilidade de não conclusão, seriam evitados portanto prejuízos aos cofres públicos.

Ao fim desse processo de tratamento, foi construída uma tabela única, com informações cruzadas das diferentes origens de dados, que possuem informações necessárias para a construção de um dashboard para uma melhor visualização das informações, além do tratamento registro a registro, para uma melhor apresentação de textos, com isso foi exportado um arquivo excel que posteriormente será disponibilizado no *Power BI* para construção do painel.

4.4. Construção de painel

Com a planilha disponível, foi feita a importação na plataforma *Power BI*, e utilizado a ferramenta *Power Query* visualizado na *Figura 8* (é uma ferramenta criada para extração, carregamento e transformação de dados) para um último tratamento antes de construir o painel de fato, então fora permitido ser construído mais uma coluna que aponta e define se o registro em questão possui fornecedor com sanção. E adequação de algumas informações, como formatar a data em um melhor formato, colocar máscara no CNPJ, e caixa alta em algumas colunas para uma melhor visualização do painel.

Figura 8 - Tela do Power Query

NumeroProcesso	Objeto	TipoAquisicao	Situacao	ModalidadeProcesso	NomeOrgao
84177519	MANDADOS JUDICIAIS NA AREA DA SAUDE, MEDICAMENTOS.	OUTRAS COMPRAS	CELEBRADO	PREGÃO ELETRÔNICO	SECRETARIA DE ESTADO DA SA
84177519	MANDADOS JUDICIAIS NA AREA DA SAUDE, MEDICAMENTOS.	OUTRAS COMPRAS	CELEBRADO	PREGÃO ELETRÔNICO	SECRETARIA DE ESTADO DA SA
2023-599C23	AQUISIÇÃO DE CERTIFICADO DIGITAL	OUTROS SERVIÇOS	CELEBRADO	PREGÃO ELETRÔNICO	SECRETARIA DE ESTADO DA SA
2023-599C134	RESSONÂNCIA NUCLEAR MAGNÉTICA DE ALTO CAMPO DE QUALQUER...	OUTRAS COMPRAS	CELEBRADO	PREGÃO ELETRÔNICO	HOSPITAL E MATERINIDADE SII
84177519	MANDADOS JUDICIAIS NA AREA DA SAUDE, MEDICAMENTOS.	OUTRAS COMPRAS	CELEBRADO	PREGÃO ELETRÔNICO	SECRETARIA DE ESTADO DA SA
84177519	MANDADOS JUDICIAIS NA AREA DA SAUDE, MEDICAMENTOS.	OUTRAS COMPRAS	CELEBRADO	PREGÃO ELETRÔNICO	SECRETARIA DE ESTADO DA SA
2023-599C23	AQUISIÇÃO DE CERTIFICADO DIGITAL	OUTROS SERVIÇOS	CELEBRADO	PREGÃO ELETRÔNICO	SECRETARIA DE ESTADO DA SA
2023-599C134	RESSONÂNCIA NUCLEAR MAGNÉTICA DE ALTO CAMPO DE QUALQUER...	OUTRAS COMPRAS	CELEBRADO	PREGÃO ELETRÔNICO	HOSPITAL E MATERINIDADE SII
84177519	MANDADOS JUDICIAIS NA AREA DA SAUDE, MEDICAMENTOS.	OUTRAS COMPRAS	CELEBRADO	PREGÃO ELETRÔNICO	SECRETARIA DE ESTADO DA SA
84177519	MANDADOS JUDICIAIS NA AREA DA SAUDE, MEDICAMENTOS.	OUTRAS COMPRAS	CELEBRADO	PREGÃO ELETRÔNICO	SECRETARIA DE ESTADO DA SA
2023-599C23	AQUISIÇÃO DE CERTIFICADO DIGITAL	OUTROS SERVIÇOS	CELEBRADO	PREGÃO ELETRÔNICO	SECRETARIA DE ESTADO DA SA
2023-599C134	RESSONÂNCIA NUCLEAR MAGNÉTICA DE ALTO CAMPO DE QUALQUER...	OUTRAS COMPRAS	CELEBRADO	PREGÃO ELETRÔNICO	HOSPITAL E MATERINIDADE SII
2023-599C419	AQUISIÇÃO DE MATERIAIS DE LABORATORIO PARA A LIUM, CONFORM...	AQUISIÇÃO DE MATERIAIS	CELEBRADO	PREGÃO ELETRÔNICO	SECRETARIA DE ESTADO DA SA
2023-599C468	REGISTRO DE PREÇOS PARA AQUISIÇÃO DE PAPEL		null	null	null
2023-599C470	Aquisição de CADEIRAS DO TIPO EXECUTIVA	OUTRAS COMPRAS	CELEBRADO	PREGÃO ELETRÔNICO	SECRETARIA DE ESTADO DA SA
2023-599C468	REGISTRO DE PREÇOS PARA AQUISIÇÃO DE PAPEL		null	null	null
2023-599C470	Aquisição de CADEIRAS DO TIPO EXECUTIVA	OUTRAS COMPRAS	CELEBRADO	PREGÃO ELETRÔNICO	SECRETARIA DE ESTADO DA SA
2023-599C587	AQUISIÇÃO DE GS GLP	AQUISIÇÃO DE EQUIPAMENTOS	ENCERRADO	-	SECRETARIA DE ESTADO DA SA
2023-599C537	AQUISIÇÃO DE BALSAS DE SALVAMENTO AQUATICO	COMPRAS DE ENTREGA IMEDIATA E INTEG.	CELEBRADO	PREGÃO ELETRÔNICO	SECRETARIA DE ESTADO DA SA
2023-599C507	AQUISIÇÃO DE MATERIAL MÉDICO HOSPITALAR PARA LIUM ATRAVÉS...	OUTRAS COMPRAS	CELEBRADO	DISPENSA DE LICITAÇÃO	SECRETARIA DE ESTADO DA SA
85886040	AQUISIÇÃO DE MATERIAL HOSPITALAR, SCALP.	OUTRAS COMPRAS	CELEBRADO	PREGÃO ELETRÔNICO	SECRETARIA DE ESTADO DA SA
85886040	AQUISIÇÃO DE MATERIAL HOSPITALAR, SCALP.	OUTRAS COMPRAS	CELEBRADO	PREGÃO ELETRÔNICO	SECRETARIA DE ESTADO DA SA
2023-599C472	AMBIENTANTANA 10 MG E 0,5 MG.	COMPRAS DE ENTREGA IMEDIATA E INTEG.	CELEBRADO	PREGÃO ELETRÔNICO	SECRETARIA DE ESTADO DA SA
2023-599C605	AQUISIÇÃO DE MEDICAMENTOS ATRAVÉS DA ARP 0556/2020	OUTRAS COMPRAS	CELEBRADO	PREGÃO ELETRÔNICO	HOSPITAL DE SÃO JOSÉ DO CA
2023-599C521	MEDICAMENTOS DESTINADOS A ASSISTÊNCIA SOCIAL	OUTRAS COMPRAS	CELEBRADO	-	HOSPITAL DE SÃO JOSÉ DO CA
2023-599C458	AQUISIÇÃO DE MEDICAMENTOS BASA A LIUM ATRAVÉS DA ARP 000...	OUTRAS COMPRAS	ENCERRADO	PRESENCIA DE LICITAÇÃO	SECRETARIA DE ESTADO DA SA

Fonte : O Autor

Na figura 8, conseguimos visualizar a tela de uso do Power Query, onde foram feitas 2 adaptações, a primeira foi a padronização dos registros, colocando por exemplo dados que deverão aparecer no painel todas maiúsculas; outra adaptação foi a filtragem na coluna que armazenava se o fornecedor estava sancionado, essa coluna, foi criada na seção anterior.

Após isso por meio da análise dos registros e colunas foram escolhidos alguns insights a serem mostrados em gráficos no dashboard, para uma melhor e fácil visualização foram escolhidos gráficos de barra, rosca, lista e mapa para o painel.

1. Gráfico de rosca:

- É melhor usado para mostrar a proporção das partes em relação ao todo.
- Útil para destacar a distribuição de classes em um conjunto de dados.
- Pode ser usado para expressar uma porcentagem ou porção de um total, como fatias de bolo.

2. Gráfico de barras:

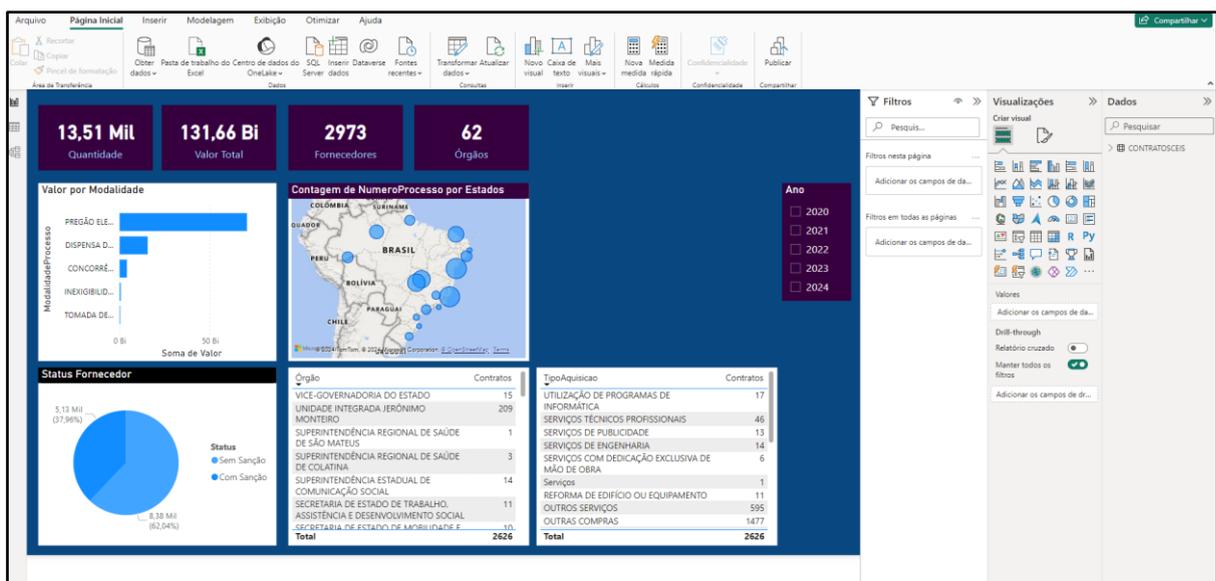
- É ideal para comparar quantidades entre diferentes categorias.
- Consegue mostrar claramente as diferenças de tamanho entre os dados.
- Útil para mostrar tendências ao longo do tempo ou comparações entre grupos.

3. Mapa:

- É mais adequado para representar dados geográficos e espaciais.
- Útil para visualizar padrões regionais, distribuição geográfica de dados ou analisar dados por localização.
- Pode ser usado para exibir dados como densidade populacional, vendas por região, distribuição de recursos, etc.

Com essas noções foram construídos gráficos para melhor expor os insights necessários para que a mensagem fosse transmitida, utilizando o editor do *Power BI*, conforme demonstrados na figura 9.

Figura 9 - Editor do Power BI



Fonte : O Autor

5. RESULTADOS OBTIDOS

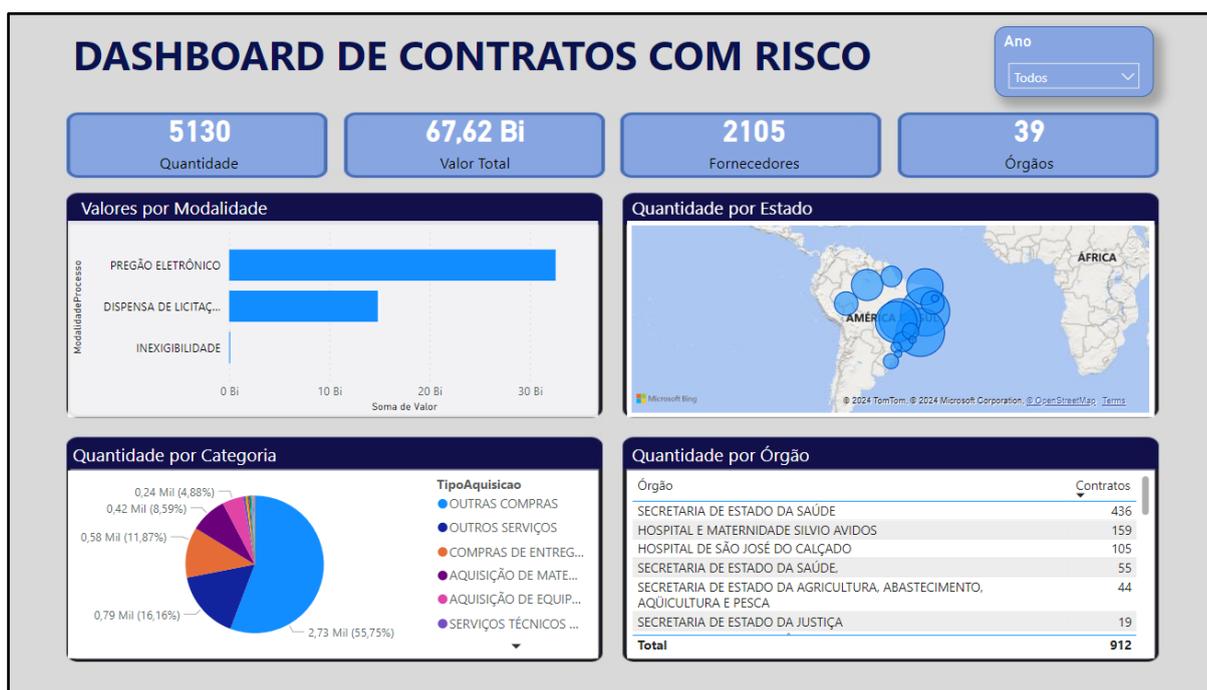
Neste trabalho foi desenvolvido um dashboard utilizando a plataforma *Power BI*, integrando habilidades em programação com códigos SAS/SQL para a análise e tratamento de dados. A trabalho teve como foco a identificação de contratos relacionados a licitações públicas, com foco aos fornecedores sancionados no Cadastro Nacional de Empresas Inidôneas e Suspensas (CEIS). Através da análise detalhada dos dados e da aplicação de técnicas avançadas de visualização, foi possível extrair conclusões valiosas sobre a relação entre contratos públicos e fornecedores sancionados, contribuindo assim para o aprimoramento das práticas de transparência e compliance no setor público. Foi desenvolvido portanto um Dashboard em *Power BI* com as seguintes telas disponíveis.

Figura 10 - Dashboard de Contratos



Fonte : O Autor

Figura 11 - Dashboard de Contratos com uma espécie de risco



Fonte : O Autor

Nessas telas, é possível identificar diversas informações agrupadas de diversos contratos, permitindo identificar padrões e regras de suas ocorrências.

Como primeira funcionalidade o software identifica o ano dos registros dos contratos e lista todos os anos distintos que existem por toda a massa de dados, após isso é possível filtrar os dados, a partir do campo suspenso, permitindo uma identificação de padrões e ocorrências ano a ano.

Outro componente importante, são os *Big Numbers*, que aglutinam as informações recorrentes e retornam um somatório, seja de quantidade ou valor do registro em questão descrito em cada componente. Eles são :

- Quantidade : Quantidade Geral de Contratos ;
- Valor Total : Valor somado de todos os contratos descritos na base de dados ;
- Fornecedores: Quantidade de Fornecedores distintos que estão designados a todos os contratos;
- Órgãos : Quantidade de Órgãos distintos que fizeram a contratação de todos os contratos.

Figura 12 - Big Numbers do Dashboard



Fonte : O Autor

Depois foram selecionados 4 tipos de gráficos, para apontar e descrever de maneira clara cada informação para que se possa ter uma visão clara e precisa dos contratos licitados. Nesses gráficos é possível identificar as seguintes informações :

- Comparação e Quantidade de contratos por Modalidade contratada;
- Comparação e Quantidade de contratos por Estado, de maneira mais visual e clara, utilizando um mapa;
- Quantidade de contratos por tipo de Categoria do Objeto contratado;
- Quantidade de contratada por cada Órgão.

Figura 13 - Gráficos visuais de informações aglutinadas



Fonte : O Autor

Por fim, foi desenvolvido filtros de seleção para todos os gráficos, permitindo uma visualização e identificação mais detalhada dos contratos. Isso é feito em todos os gráficos e influenciam diretamente os demais gráficos do dashboard.

Figura 14 - Dashboard filtrado pela seleção de gráfico



Fonte : O Autor

No Dashboard foi feita uma seleção por categoria “Outros Serviços”, com isso o painel se atualiza, fornecendo informações mais detalhadas e precisas desses contratos, podendo ainda fazer a comparação com os dados antes da filtragem.

5.1. Dificuldades Encontradas

Durante o desenvolvimento desse projeto, algumas dificuldades foram verificadas, e portanto necessário de contorná-las para finalizar o processo.

A princípio notou-se a dificuldade na extração dos dados, devido às limitações de *download* do site do portal de transparência, limitando os dados extraídos a 1000 registros por arquivo.

Como solução foi necessário desenvolver códigos que unifiquem as planilhas com dados distintos em planilhas semelhantes. Isso gerou um pequeno atraso no processo, devido a limitação de espaço na máquina onde o projeto foi desenvolvido.

Após a unificação das tabelas de contratos e fornecedores sancionados, foi notado uma quantidade não significativa de contratos com risco, pelo fato da amostra disponível ser ainda baixa tendo como base a quantidade de contratos que são firmados em todo o Brasil.

Portanto foram inseridos registros de exemplo, pelo fato de o processo ser acadêmico e com intuito de mostrar o processo e suas dificuldades na implementação,

e não uma acusação de fraude em contratos licitatórios.

Em seguida a plataforma SAS possui ferramentas para desenvolvimento visual, dashboard e painel. No entanto, é necessário possuir licenças pagas para utilizar tais recursos, contudo, foi extraído o arquivo tratado com código SAS e SQL, para importar na ferramenta *Power BI*, que possui recursos de desenvolvimento visual de dashboards gratuitos.

Por final foi necessário anonimizar os fornecedores representados no painel, para que o dashboard se adeque às normas de LGPD (Lei Geral de Proteção de Dados).

6. CONCLUSÃO

A proposta apresentada por esse trabalho, expõe resultados obtidos de maneira parcialmente satisfatória, previstos no projeto, porém de extrema valia, que permite a identificação de padrões em contratos firmados por órgãos públicos e fraudes ou irregularidades em fornecedores que estão de posse desses processos.

Foram consumidas informações disponibilizadas no portal de transparência do governo federal, de fornecedores e contratos públicos. Em posse desses dados foram desenvolvidas duas telas de painel em base de conceitos, técnicas e ferramentas de BI.

Como finalização, concluo que Business Intelligence pode sim auxiliar na identificação de padrões em contratos com um certo grau de risco ou não. Porém o trabalho que foi desenvolvido não atua de maneira ativa na prevenção de futuros casos, somente como uma ferramenta de auxílio na visualização e entendimento desses insights.

Esse trabalho pode servir como base para futuros trabalhos que busquem evidenciar uma gama maior de situações fraudulentas em contratos, até auxiliando em um possível trabalho de prevenção desse tipo de situações. Além de que o conhecimento aqui contido, pode ser utilizado para auxiliar inúmeros trabalhos que envolvem o estudo de dados, e implementação das ferramentas e técnicas utilizadas.

Concluo que esse trabalho acrescentou na minha formação como Engenheiro da computação, permitindo revisão de conceitos que sinto que estavam “adormecidos”, além de me permitir desenvolver uma habilidade de estudo autodidata para conceitos que não os conhecia, explorando mais ainda documentos e artigos com bastante conteúdo. Sem contar um crescimento como cidadão brasileiro, acredito que muitos de nós não possuímos muita noção de como funcionam os trâmites em órgãos públicos, esse trabalho me fez mergulhar em documentos, leis e cenários, que antes de seu início não tinha noção ao menos de sua existência.

Referências

Bancos usam o big data para acelerar o combate às fraudes. Disponível em: <<https://www.youtube.com/watch?v=AiGManTzjcc&t=252s>>. Acesso em: 11 out. 2023.

BELK, R. W. **Possessions and the extended self.** *Journal of consumer research*, v. 15, n. 2, p. 139-168, 1988.

BI e big data: veja quais são as principais diferenças – Farol Business Intelligence. Disponível em: <<https://farolbi.com.br/bi-e-big-data-veja-quais-sao-as-principais-diferencas/>>. Acesso em: 22 jun. 2024.

Big Data: o que é e qual a importância para o marketing? Disponível em: <<https://posdigital.pucpr.br/blog/o-que-e-big-data>>.

BOWE.COM.BR, BOWE-; TELECOM, A. **O que é Armazenamento de dados, Importância e Principais Tipos.** Disponível em: <<https://blog.algartelecom.com.br/tecnologia/armazenamento-de-dados/>>. Acesso em: 2 out. 2023.

CALASAN, M. **What Are DDL, DML, DQL, and DCL in SQL?** Disponível em: <<https://learnsql.com/blog/what-is-dql-ddl-dml-in-sql/>>.

CDTV. **Bancos usam o big data para acelerar combate às fraudes.** Disponível em: <<https://www.youtube.com/watch?v=AiGManTzjcc&t=268s>>. Acesso em: 22 jun. 2024.

CETAX. **Data Mining (Mineração de Dados): O que é, conceito definição.** Disponível em: <<https://cetax.com.br/data-mining/>>.

Cientista de Dados programa? Disponível em: <<https://www.alura.com.br/artigos/cientista-de-dados-programa>>.

Conceitos de Business Intelligence Guia Definitivo Termos e Condições Aviso Legal. [s.l: s.n.]. Disponível em: <<https://www.direcaoentido.com.br/wp-content/uploads/2014/04/Conceitos-de-Business-Intelligence-Guia-Definitivo.pdf>>.

Contratos | Portal da Transparência do Governo Federal. Disponível em: <<https://portaldatransparencia.gov.br/contratos?ano=2021>>. Acesso em: 17 mar. 2024.

“Dados são o novo petróleo”, diz CEO da Mastercard – exceto por um pequeno detalhe. Disponível em: <<https://epocanegocios.globo.com/Empresa/noticia/2019/07/dados-sao-o-novo-petroleo-diz-ceo-da-mastercard.html>>.

DATAEX. **Big Data e Business Intelligence: você conhece a diferença?** Disponível em: <<https://www.dataex.com.br/big-data-e-business-intelligence/#:~:text=O%20Big%20Data%2C%20por%20exemplo>>. Acesso em: 22 jun. 2024.

Detalhamento das Sanções Vigentes - Portal da transparência. Disponível em: <<https://portaldatransparencia.gov.br/sancoes/consulta?cadastro=1&ordenarPor=nomeSancionado&direcao=asc>>.

Descubra 5 das vantagens do big data na saúde! - Know Solutions. Disponível em <<https://www.knowsolution.com.br/vantagens-big-data-saude/>>. Acesso em: 13 set. 2023.

DIAS, R. **Big Data: O que é e como pode impactar no marketing digital.** Disponível em: <<https://blog.e-goi.com/br/big-data/>>. Acesso em: 13 set. 2023.

FILGUEIRAS, F., LUI, L. **Designing data governance in Brazil: an institutional analysis.** *Policy Design and Practice* 1-16, 2022. Disponível em: . Acesso em: 10 out. 2022.

FRANCIELE, C.; MEDRADO. **INDICADORES ECONÔMICO-FINANCEIROS COMO RED FLAGS DE RISCOS DE FRAUDES OU MANIPULAÇÃO CONTÁBIL.** [s.l.: s.n.]. Disponível em: <[https://repositorio.cgu.gov.br/bitstream/1/41960/5/9_Dissertacao_Franciele_Demat te.pdf](https://repositorio.cgu.gov.br/bitstream/1/41960/5/9_Dissertacao_Franciele_Demat%20te.pdf)>. Acesso em: 7 dez. 2023.

GIL, A. C. **Como elaborar projetos de pesquisa.** 4. ed. São Paulo: Atlas, 2008.

Junta Comercial do estado de Minas Gerais - JUCEMG. Disponível em: <https://jucemg.mg.gov.br/adminlte/bower_components/kcfinder/upload/files/Planejamento_estrat%C3%A9gico_jucemg_2024_v1.pdf>. Acesso em: 22 jun. 2024.

MACHADO, F. N. R. **Big Data: o futuro dos dados e aplicações.** São Paulo: Erica, 2018.

MARCELA. **Domine Sas e Sql: Aprenda as Melhores Técnicas Em um Curso Completo.** Disponível em: <https://awari.com.br/domine-sas-e-sql-aprenda-as-melhores-tecnicas-em-um-curso-completo/?utm_source=blog&utm_campaign=projeto+blog&utm_medium=Domine%20Sas%20e%20Sql:%20Aprenda%20as%20Melhores%20T%C3%A9cnicas%20Em%20um%20Curso%20Completo>. Acesso em: 22 jun. 2024.

Manutenção preditiva: a técnica baseada em dados-chave para se antecipar aos erros. Disponível em: <<https://www.iberdrola.com/inovacao/manutencao-preditiva#:~:text=A%20manuten%C3%A7%C3%A3o%20preditiva%20%C3%A9%20uma>>.

SQL: O que é e como usar os principais comandos básicos SQL. Disponível em: <<https://blog.betrybe.com/sql/>>.

GONÇALVES, E. **PL/SQL.** [s.l.] Editora Casa do Código, 2015.

Tipos de gráficos: escolha o gráfico ideal para os seus dados. Disponível em: <<https://pt.venngage.com/blog/tipos-de-graficos/>>.

ENAP, E. N. DE A. **Ciência de dados em políticas públicas: uma experiência de formação.** repositorio.enap.gov.br, 2022.

O que é ciência de dados? – Explicação sobre ciência de dados – AWS. Disponível em: <<https://aws.amazon.com/pt/what-is/data-science#:~:text=A%20ci%C3%Aancia%20de%20dados%20%C3%A9%20importante%20porque%20combina%20ferramentas%2C%20m%C3%A9todos>>. Acesso em: 22 jun. 2024.

O que é confiabilidade de dados? | IBM. Disponível em: <<https://www.ibm.com/br-pt/topics/data-reliability>>. Acesso em: 22 jun. 2024.

O que é Power BI da Microsoft | Business Intelligence. Disponível em: <<https://4infra.com.br/o-que-e-power-bi/#:~:text=Power%20BI%20%C3%A9%20uma%20solu%C3%A7%C3%A3o>>. Acesso em: 22 jun. 2024.

TOTVS, E. **Eficiência operacional: o que é, como medir e como melhorar.** Disponível em: <<https://www.totvs.com/blog/gestao-logistica/eficiencia-operacional#:~:text=A%20efici%C3%Aancia%20operacional%20%C3%A9%20a>>. Acesso em: 22 jun. 2024.

WEBI. **Conheça a história do armazenamento de dados e a sua evolução.** Disponível em: <<https://acervonet.com.br/blog/historia-do-armazenamento-de-dados/>>. Acesso em: 2 out. 2023.

ZOOP, R. **Fraudes na internet: 7 tipos para ficar de olho e evitar golpes!** Disponível em: <<https://zoop.com.br/blog/gestao-empresarial/fraudes-internet/>>. Acesso em: 29 nov. 2024.

78% das empresas reconhecem que não conseguem responder de forma eficaz a fraudes. Disponível em: <<https://exame.com/future-of-money/78-empresas-reconhecem-nao-conseguem-responder-fraudes/>>. Acesso em: 7 dez. 2023.

RESOLUÇÃO n^o 038/2020 — CEPE

ANEXO I APÊNDICE

ao TCC

Termo de autorização de publicação de produção acadêmica

O(A) estudante Matheus Gomes dos Santos Araujo
do Curso de Engenharia da computação, matrícula 20181003300775,
telefone: (62)982080102 e-mail matheusgsaraujo@gmail.com, na qualidade de
titular dos direitos autorais, em consonância com a Lei n^o 9.610/98 (Lei dos Direitos do
autor), autoriza a Pontificia Universidade Católica de Goiás (PUC Goiás) a
disponibilizar o Trabalho de Conclusão de Curso intitulado

Uso de Big Data e Business Intelligence, na identificação de Contratos com possíveis
irregularidades, gratuitamente, sem ressarcimento dos direitos autorais,
por 5 (cinco) anos, conforme permissões do documento, em meio eletrônico, na rede
mundial de computadores, no formato especificado (Texto (PDF); Imagem (GIF ou
JPEG); Som (WAVE, MPEG, AIFF, SND); Vídeo (MPEG, MWV, AVI, QT); outros,
específicos da área; para fins de leitura e/ou impressão pela internet, a título de
divulgação da produção científica gerada nos cursos de graduação da PUC Goiás.

Goiânia, 18 de Junho de 2024

Assinatura do(s) autor(es): _____

Nome completo do autor: Matheus Gomes dos Santos Araujo

Assinatura do professor-orientador: _____

Nome completo do professor-orientador: André Luiz Alves