

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CAMPUS SÃO CARLOS

João Victor Pacheco

Programação de Módulo de *Business Intelligence* no
Block-Based Integrated Platform for Embedded Systems
(BIPES)

JOÃO VICTOR PACHECO

Programação de Módulo de *Business Intelligence* no
Block-Based Integrated Platform for Embedded Systems
(BIPES)

Trabalho de Conclusão de Curso submetido à Universidade Federal de São Carlos, como requisito necessário para obtenção do grau de Bacharel em Engenharia de Computação

São Carlos, Setembro de 2022

UNIVERSIDADE FEDERAL DE SÃO CARLOS

JOÃO VICTOR PACHECO

Esta Monografia foi julgada adequada para a obtenção do título de Bacharel em Engenharia de Computação, sendo aprovada em sua forma final pela banca examinadora:

Orientador(a): Profa. Dra. Marilde Terezinha Prado Santos
Universidade Federal de São Carlos (UFSCar)

Banca: Profa. Dra. Marcela Xavier Ribeiro
Universidade Federal de São Carlos (UFSCar)

Banca: Prof. Dr. Rafael Vidal Aroca
Universidade Federal de São Carlos (UFSCar)

São Carlos, 16 de Setembro de 2022

Dedico esse trabalho à minha família e amigos. Dedico especialmente ao meu pai, senhor João, que se foi antes que pudesse ver eu me formar, mas sem dúvidas, nesse momento ele seria uma das pessoas mais felizes do mundo. Mesmo que não esteja presente fisicamente, em minhas lembranças, ele sempre me acompanhará.

Agradecimentos

Meus mais profundos agradecimentos à Profa. Dra. Marilde Terezinha Prado Santos, por ter me orientado e dado suporte para a realização deste trabalho de pesquisa. À minha família, agradeço pelo incentivo, e ajuda durante todos esses anos. Aos meus amigos, me faltam palavras para agradecer pela parceria e grande ajuda com os empecilhos que surgiram ao longo da graduação, sempre me estendendo a mão e não deixando eu nunca desistir dos meus objetivos, mesmo frente às adversidades. Agradeço também toda equipe de professores e funcionários do Departamento de Computação da Universidade Federal de São Carlos.

Resumo

Com modernização das tecnologias e das atividades da sociedade, o processo de tomada de decisões passa a depender da análise de conjuntos de dados cada vez maiores. O aumento do uso da internet, o surgimento da Internet das Coisas (do inglês, “*Internet of Things*” - IoT), e a popularização do *Big Data*, impactaram de forma expressiva no aumento da geração de volumes de dados com maior variedade e velocidade de transmissão. Isso levou à demanda por ferramentas mais complexas que permitam que as empresas se tornem cada vez mais *Data Driven*, ou seja, tenham seus processos, tomada de decisão e planejamento estratégico orientados a dados. Nesse contexto surgem as ferramentas de “*Business Intelligence*” (BI), que são capazes de organizar e entregar de forma bem estruturada maneiras de manipular, analisar e visualizar esses dados. O *Block-Based Integrated Platform for Embedded Systems* (BIPES), por sua vez, é uma plataforma que permite o desenvolvimento sem código para sistemas embarcados, permitindo que usuários desenvolvam soluções através do uso de blocos de programação. Muitas das soluções desenvolvidas utilizando o BIPES são relacionadas a coleta de dados de sensores, e atualmente a plataforma oferece ferramentas para análise dos dados em tempo real, porém não oferece ao usuário uma maneira de analisar grandes volumes de dados históricos. Isso significa que o usuário não consegue de maneira fácil e visual ter *insights* e tomar decisões de forma que impactem o escopo do projeto como um todo. Deste modo, este presente trabalho se propõe a desenvolver, por meio de ferramentas de BI, um módulo capaz de auxiliar na análise de dados de sensores e processo de tomada de decisões, que seja integrável com a plataforma BIPES. Para isso, foi desenvolvido um *dashboard* utilizando a ferramenta Tableau Desktop, capaz de apresentar o comportamento histórico e previsão dos dados no futuro, ordenação de dispositivos por performance, alerta de não-conformidade, comportamento ao longo do dia, presença de *outliers* e indicadores de resultado médio diário e total.

Palavras-chave: *Business Intelligence*; Tomada de Decisões; BIPES; *Big Data*; Tableau Desktop; *Data Driven*; Internet das Coisas.

Abstract

With the modernization of technologies and the activities of the society, the decision-making process starts to depend on the analysis of increasingly larger data sets. The increase in the use of the internet, the emergence of the Internet of Things and the popularization of Big Data, affect in a expressive way, the increase in the generation of high volumes of data with greater variety at high speed. This lead the demand for more complex tools that allow companies to become more Data Driven, which means to have their processes, decision-making and strategic planning based on data. In this context Business Intelligence tools emerged, being able to organize and delivery well-structured ways to manipulate, analyse and visualize those data. The Block-based Integrated Platform for Embedded Systems (BIPES) is a platform that allows the no-code development for embedded systems, empowering users to develop solutions using programming blocks. Most solutions developed using BIPES are related to sensor data collection, and so far, the platform offer tools for real-time analysis of the data, however it does not offer a way to analyze large volumes of historical data. BIPES users, in a visual and easy way, cannot extract insights and make decisions that impact all the scope of the project. Thus, this present work proposes to develop, using BI tools, a module capable of assisting in the analysis of sensor data and the decision-making processs, which can be integrated with the BIPES. For this purpose , a dashboard was developed using the Tableau Desktop tool, capable of present the historical behavior and forecasting, device performance ranking, nonconformity alert, behavior throughout the day, outilers presence, and indicators of daily average and total results.

Keywords: Business Intelligence; Decision-making; BIPES; Tableau Desktop; Data Driven; Big Data; Internet of Things.

Lista de abreviaturas e siglas

AWS - Amazon Web Services

BI - do inglês, “*Business Intelligence*”

BIPES - do inglês, “*Block-Based Integrated Platform for Embedded Systems*”

CRM - do inglês, “*Customer Relationship Management*”

CSV - do inglês, “*Comma-Separated Values*”

DW - do inglês, “*Data Warehouse*”

ETL - do inglês, “*Extract, Transform and Load*”

IoT - do inglês, “*Internet of Things*”

JSON - do inglês, “*JavaScript Object Notation*”

KPIs - do inglês, “*Key Performance Indicators*”

MS - do inglês, “*Microsoft*”

NoSQL - do inglês, “*Not Only SQL*”

SAS - do inglês, “*Statistical Analysis System*”

SGBD - Sistema Gerenciador de Banco de Dados

SQL - do inglês, “*Structured Query Language*”

XLSX - do inglês, “*Microsoft Excel Open XML Spreadsheet*”

Lista de ilustrações

Figura 1 – Exemplo de solução para IoT criada com o BIPES. Extraído de [Junior et al. (2020)]	17
Figura 2 – Exemplo de monitor de tempo real no BIPES. Extraído de [BIPES (2022)]	17
Figura 3 – Ciclo de Vida dos Dados. Extraído de [Stobierski 2021] e traduzido . . .	20
Figura 4 – Importação de base de dados para o Tableau (Autoria Própria)	31
Figura 5 – Campo calculado <i>DateDay</i> (Autoria Própria)	32
Figura 6 – Parâmetro <i>Measure/Sensor</i> (Autoria Própria)	32
Figura 7 – Campo calculado <i>Sensor</i> (Autoria Própria)	33
Figura 8 – Gráfico de comportamento histórico dos dados (Autoria Própria)	34
Figura 9 – Sequencial de tons baseados no valor da variável <i>Sensor</i> (Autoria Própria)	34
Figura 10 – Parâmetro para definição do número de observações no <i>Ranking</i> (Autoria Própria)	35
Figura 11 – <i>Ranking</i> de Dispositivos (Autoria Própria)	35
Figura 12 – Campos calculados <i>UpperBand</i> e <i>LowerBand</i> (Autoria Própria)	36
Figura 13 – Campo calculado <i>It_is_Outlier</i> (Autoria Própria)	37
Figura 14 – Tabela de Alerta de não-conformidade (Autoria Própria)	37
Figura 15 – Gráfico de comportamento dos dados por Hora (Autoria Própria)	38
Figura 16 – Painel para configuração das opções de previsão (Autoria Própria)	39
Figura 17 – Gráfico com histórico e previsão do comportamento de dados (Autoria Própria)	39
Figura 18 – Tabela de dias com <i>outliers</i> (Autoria Própria)	40
Figura 19 – Campo Calculado <i>Sensor_AVG</i> (Autoria Própria)	41
Figura 20 – Indicador de média diária (Autoria Própria)	41
Figura 21 – Campo Calculado <i>Sensor_Total</i> (Autoria Própria)	42
Figura 22 – Indicador de Total Geral (Autoria Própria)	42
Figura 23 – <i>Dashboard</i> com visão de Dados Históricos + Previsão (Autoria Própria)	43
Figura 24 – Visão do <i>dashboard</i> com a presença de <i>outliers</i> (Autoria Própria)	44
Figura 25 – Módulo de BI disponibilizado no <i>Tableau Public</i> (Autoria Própria) . . .	45
Figura 26 – Conector de dados web para JSON (Autoria Própria)	45

Lista de tabelas

Tabela 1 – De-para de substituições de nomes de campos	29
Tabela 2 – Funções nativas do Tableau utilizadas	30

Sumário

1	INTRODUÇÃO	13
1.1	Contextualização e Problemática	13
1.2	Objetivo	14
1.2.1	Objetivo Geral	14
1.2.2	Etapas de Desenvolvimento	14
1.3	Organização	15
2	DESENVOLVIMENTO SEM CÓDIGO	16
2.1	<i>Block-Based Integrated Platform for Embedded Systems</i>	16
3	<i>BUSINESS INTELLIGENCE</i>	18
3.1	Etapas	19
3.1.1	ETL	19
3.1.1.1	Geração de Dados	19
3.1.1.2	Coleta de Dados	20
3.1.1.3	Processamento	20
3.1.1.4	Armazenamento	20
3.1.2	Desenvolvimento de BI	21
3.1.2.1	Gerenciamento	21
3.1.2.2	Análise	21
3.1.2.3	Visualização	21
3.1.2.4	Interpretação	22
3.2	Ferramentas	22
3.2.1	Ferramentas de ETL	22
3.2.1.1	Bancos de Dados Estruturados e o SQL	22
3.2.1.2	NoSQL	23
3.2.2	Ferramentas de Visualização	24
3.2.2.1	Grafana	25
3.2.2.2	Power BI	25
3.2.2.3	Tableau	25
4	REVISÃO BIBLIOGRÁFICA	27
4.1	A importância de BI	27
4.2	Implementações e usos de BI	28
5	PROPOSTA E DESENVOLVIMENTO	29

5.1	 Materiais e Métodos	29
5.2	 Construção do <i>dashboard</i>	30
5.2.1	Importação e Manipulação dos Dados	30
5.2.2	Comportamento histórico dos dados	31
5.2.3	<i>Ranking</i> de dispositivos	33
5.2.4	Alerta de não-conformidade	36
5.2.5	Comportamento dos dados em períodos do dia	38
5.2.6	Previsão de Comportamento	38
5.2.7	Indicador de dias com <i>outliers</i>	40
5.2.8	Indicador de média diária	40
5.2.9	Indicador de Total Geral	41
6	 RESULTADOS E DISCUSSÃO	43
7	 CONCLUSÃO	46
	 REFERÊNCIAS	48

1 Introdução

1.1 Contextualização e Problemática

Desde o surgimento dos seres humanos e sua necessidade por desempenhar funções, sejam ligadas a agricultura, criação de animais ou mesmo desenvolvimento de atividades domésticas, existe a necessidade de analisar e manipular dados. Mesmo em análises mais primitivas - como a temperatura, clima, e quantidade de água num local -, os seres humanos sempre fizeram uso de informações para fazer correlações e tomar decisões - sobre o melhor local para se estabelecer moradia, melhor época para o plantio, e qual o melhor momento para caçar determinado animal, por exemplo. Com o passar do tempo, a evolução das tecnologias e maior desenvolvimento do ser humano, tais correlações para tomada de decisões começaram a envolver cada vez mais parâmetros e maior volume de dados, tornando-se mais complexas. Com isso, também aumentou-se a necessidade por estabelecer normas e ferramentas que pudessem guiar a tomada de decisão de forma consistente e efetiva. Para tanto, foram desenvolvidas ferramentas tecnológicas que facilitam o processo de tirar *insights* desses grandes volumes de dados.

Com as novas tecnologias, surgem também formas automatizadas para a coleta e armazenamento de dados. Tal coleta de dados pode ser feita de diversas maneiras, seja através de pesquisas na internet, que buscam pelos dados relacionados a determinado assunto, ou então em aplicações mais focadas, tais como coleta de informações através de sensores (temperatura, umidade, localização, etc.). Nos últimos anos, houve uma popularização do termo “Internet das Coisas” (do inglês, “*Internet of Things*” - IoT), que Singh et al. (2020) define como um conjunto de dispositivos computacionais, digitais e mecânicos interconectados, capazes de transmitir dados e informações através de uma rede definida, sem a necessidade de envolvimento humano em nenhum nível, podendo conter informações dos mais diversos sensores e nas mais diversas aplicações, tais como sistema de controle de segurança residencial, controles inteligentes de iluminação e monitoramento de tráfego. Tais dispositivos geram quantidades gigantescas de dados, que precisam ser tratados e utilizados para a tomada de decisões e ações baseadas na aplicação as quais pertencem.

De acordo com Chen, Chiang e Storey (2012), o termo “Inteligência” começou a ser utilizado por pesquisadores a partir de 1950 para retratar a Inteligência Artificial. O termo “Inteligência de Negócio” (do inglês, “*Business Intelligence*” - BI) só foi aparecer por volta de 1990. Além disso, mais recentemente os termos “*Big Data*” e “*Big Data Analytics*” começaram a ser utilizados para descrever a carga e técnicas de análise utilizadas em aplicações que são consideravelmente grandes (na casa de *terabytes* ou *exabytes*) e

complexas, como dados de redes sociais, que requerem formas mais avançadas de coleta, armazenamento, manipulação, análise e visualização.

Portanto, devido a essa crescente necessidade de manipulação e visualização de grandes volumes de dados, torna-se cada vez mais necessária a integração dessas bases de dados, com as mais modernas e robustas ferramentas de BI, facilitando assim o trabalho dos profissionais responsáveis pela análise, entendimento e tomada de decisões. E nesse contexto tem-se o "*Block-based Integrated Platform for Embedded Systems*" (BIPES) [Junior et al. (2020)], que é uma plataforma para o desenvolvimento de soluções sem o uso de código, fazendo-as através do uso de blocos de programação, que facilita o desenvolvimento e projeto de soluções para usuários que não tenham conhecimentos aprofundados em linguagens de programação, e até mesmo gera ganhos de tempo no desenvolvimento de códigos e programas, por já existirem soluções prontas para uso na ferramenta. Dentre as diversas soluções desenvolvidas, destaca-se a coleta de dados de sensores. Esses sensores podem coletar e armazenar grandes volumes de dados contínuos e históricos. Porém, não existe algum módulo integrado ao BIPES que permita a análise de forma histórica e integral dos sensores existentes no projeto, permitindo a análise do comportamento e a retirada de *insights* que permitam a tomada de decisões que afetem de maneira estratégica o escopo dos projetos e soluções desenvolvidas.

1.2 Objetivo

1.2.1 Objetivo Geral

O objetivo geral deste trabalho é desenvolver e implementar, por meio das principais ferramentas de BI, um módulo capaz de importar, processar, e gerar a visualização de dados coletados por sensores de modo a ser integrável com a ferramenta "*Block-based Integrated Platform for Embedded Systems*" (BIPES).

1.2.2 Etapas de Desenvolvimento

Considerando o objetivo geral apresentado, é possível enumerar as etapas desenvolvidas para alcançá-lo:

1. Pesquisar os principais métodos de BI;
2. Pesquisar e listar as principais ferramentas de BI, de forma a definir a ferramenta utilizada para o desenvolvimento do módulo proposto;
3. Obter uma base de dados de coleta de sensores disponível na internet;

4. Construir e implementar um *dashboard* de forma a gerar opções de visualização dos dados, fazendo os devidos tratamentos necessários;
5. Expor os resultados obtidos com *dashboard* construído no item anterior e listar as conclusões obtidas no final deste trabalho.

1.3 Organização

O restante deste trabalho está organizado da seguinte forma:

- Capítulo 2: Apresenta e descreve a plataforma BIPES;
- Capítulo 3: Apresenta os principais conceitos e ferramentas de BI;
- Capítulo 4: Apresenta uma revisão bibliográfica sobre a importância de BI e a implementação de técnicas de BI;
- Capítulo 5: Apresenta os principais aspectos da proposta de implementação;
- Capítulo 6: Apresenta e discute os resultados obtidos;
- Capítulo 7: Apresenta as conclusões e trabalhos futuros.

2 Desenvolvimento sem Código

O “Desenvolvimento sem Código” é uma abordagem que propõe o uso de ferramentas de interface gráfica para desenvolvimento de *software*, ao invés do uso de linguagens de programação convencionais. Assim, essas ferramentas de interface gráfica, ou plataformas, são construídas por programadores, para amplo uso por qualquer pessoa interessada em desenvolvimento de soluções de *software*, sem que esta última necessite ter experiência com linguagens de programação.

A importância do Desenvolvimento sem Código foi apresentada por Caballar (2020), que destacou o papel da tecnologia para permitir e facilitar a criação de soluções, de forma democrática e universal.

Nesse contexto, existem diferentes plataformas de Desenvolvimento sem Código disponíveis para uso, como o MIT App Inventor [Magnuson (2010)], o Google AppSheet [Google (2022)] e o *Block-Based Integrated Platform for Embedded Systems* (BIPES) [Junior et al. (2020)].

Na seção a seguir a plataforma BIPES será apresentada em detalhes, uma vez que ela constitui parte fundamental da proposta do presente projeto.

2.1 *Block-Based Integrated Platform for Embedded Systems*

O *Block-Based Integrated Platform for Embedded Systems* (BIPES) é uma plataforma de Desenvolvimento sem Código para Sistemas Embarcados. Os Sistemas Embarcados abrangem uma gama de componentes de *hardware* e *software* que são construídos juntos para desempenhar funções específicas conforme salientado por Baron, Geffroy e Motet (2013). Exemplos são os sistemas para satélites, impressoras, telefones celulares e de aviões. De acordo com Hodges et al. (2020), ainda existem muitos desafios a serem superados no desenvolvimento desses Sistemas Embarcados, principalmente no que tange a área de pesquisa e educação. Foi nesse contexto, e visando superar esses desafios, que o BIPES foi criado.

O BIPES é uma ferramenta web que não requer a instalação ou configuração de qualquer *software* ou *plug-ins*. Ele propõe o Desenvolvimento sem Código a partir de uma abordagem baseada em blocos. Esses blocos são aninhados e configurados pelo usuário, a fim de construir um solução. A Figura 1 apresenta um exemplo de solução para IoT criada com o BIPES.

Em Sistemas Embarcados, é de sumo interesse o acompanhamento dos dados, tanto aqueles de tempo real, quanto os dados históricos. O acompanhamento desses dados

```

repeat 100 times
do
  print " Reading ADC... "
  set value to Read ADC Input 0
  print " Sending to server... "
  set bipes to " http://www.bipes.net.br/sensors/send.php? "
  set url_envio to create text with
    bipes
    " id=950 "
    "&s1= "
    value
  set res to Make HTTP GET Request URL url_envio
  Execute Python Code " res = res.content "
  print create text with " Server response = "
    res
  delay 5
  
```

Figura 1 – Exemplo de solução para IoT criada com o BIPES. Extraído de [Junior et al. (2020)]

permite o monitoramento da saúde de uma determinada solução, bem como previsão de métricas e comportamento do sistema. Até o momento desse estudo, o BIPES só possui um módulo para acompanhamento de dados em tempo real, como apresentado na Figura 2. Portanto, existe a oportunidade de integração de um módulo de BI no BIPES, a fim de também monitorar os dados históricos.

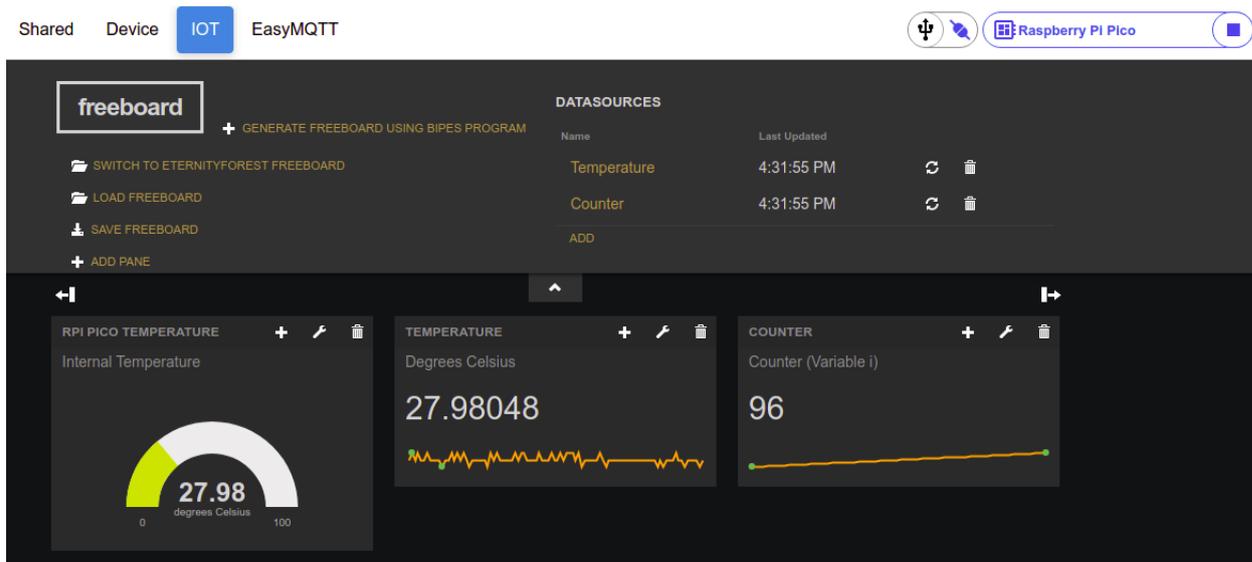


Figura 2 – Exemplo de monitor de tempo real no BIPES. Extraído de [BIPES (2022)]

3 *Business Intelligence*

“*Business Intelligence*” (BI) é o nome dado a um conjunto de operações e manipulações sobre conjuntos complexos de dados afim de auxiliar no planejamento e tomada de decisão. Negash e Gray (2008) ressaltam que BI é amplamente utilizado em empresas ou instituições privadas, geralmente extraíndo conjuntos de dados estruturados ou não-estruturados e transformando-os de forma a facilitar a tomada de decisões.

Grandes empresas costumam trabalhar com um grande volume de dados, sejam dados de clientes, vendas, faturamento e performance de funcionários, advindos das mais diversas fontes, como “*Data Warehouse*” (DW), planilhas, ou sistemas *online*. Porém, de forma isolada, os dados não transmitem uma mensagem ou representam algum conhecimento. Nesse ponto que atua o profissional de BI: Ele coleta, trata, organiza e tira *insights* a respeito dos dados, podendo ou não gerar *dashboards* para o acompanhamento dessas informações. Pode-se dizer, então, que o objetivo principal de BI é fornecer uma forma estruturada e possivelmente interativa de acesso aos dados tratados e que permita a utilização dos mesmos para a tomada de decisões. Isso pode ser feito analisando-se dados históricos e muitas vezes de tempo real, de forma a tirar percepções sobre determinado contexto a fim de solucionar problemas, ou propor melhorias dentro da estrutura ou negócio em que a proposição inicial estava contida.

Alguns exemplos de uso de BI para o auxílio do planejamento estratégico e tomada de decisões são expostos a seguir:

- **Criação de *Forecast* de Vendas:** Criação de *Forecast* baseado em dados históricos e atuais de performance de vendas de uma empresa, estimando os resultados das vendas nos próximos meses;
- **Apuração de *Salesperson Performance*:** Levantamento e apuração de indicadores da performance de vendedores e representantes de uma empresa, analisando resultados de vendas em comparação a uma meta predeterminada que pode ser utilizada para definição de remuneração variável;
- **Análise de Tendência de Mercado:** Apuração de informações a respeito de características e demandas dos consumidores de determinada empresa, afim de definir os melhores produtos a serem investidos, tanto já existentes no catálogo da empresa quanto produtos novos a serem desenvolvidos;
- **Prospecção de Novos Clientes:** Através do cruzamento de bases de dados internas, como a carteira de clientes atuais, com outras fontes de dados externas a respeito de uma população de pessoas de determinada região, é possível mapear

o perfil de possíveis clientes naquela localidade e quais produtos ou ferramentas podem ser mais úteis para os clientes em potencial. Assim, o time de *marketing* e vendas é capaz de trabalhar com maior assertividade na prospecção de clientes, aumentando a carteira de clientes e consequentemente o faturamento da empresa;

- **Análise Preditiva de Focos de Incêndio e Queimadas:** Análise de dados históricos de focos de queimada em uma região, fazendo a higienização e levantamento das maiores recorrências em determinados lugares e épocas do ano, ou características meteorológicas que impactam o surgimento de focos de incêndio e queimadas, prevendo futuros acontecimentos no objetivo de agir na prevenção destes;
- **Utilização por Povos Antigos:** De certa forma, pode-se dizer que mesmo povos antigos utilizavam métodos similares a BI, relacionando informações de períodos de chuvas e secas, posição dos astros no céu, comportamento das marés, períodos de frio e calor para tomar decisões de melhores momentos para plantio e colheita de alimentos, pesca e criação de animais;
- **Detecção de Fraudes:** Análise de dados de comportamento de clientes ou transações comprovadamente fraudulentas, afim de comparar com determinada rapidez e assertividade novas transações que são suspeitas de fraude, minimizando os danos à empresa.

3.1 Etapas

Como observado, o cerne de BI são os dados, e a forma como pode-se manipular esses dados para gerar interpretações. Dessa forma, existem algumas etapas nas quais o ciclo de vida desses dados está inserido, como pode ser observado na Figura 3. Essas etapas podem ser divididas entre duas macro etapas, o ETL e o desenvolvimento de BI.

3.1.1 ETL

ETL, sigla para “*Extract*”, “*Transform*” e “*Load*”, é a etapa inicial do ciclo de vida dos dados, que engloba a coleta, processamento e armazenamento dos dados. Além disso, pode-se adicionar uma etapa anterior a essa, que é a geração dos dados.

3.1.1.1 Geração de Dados

Dados podem ser gerados a partir de diversas fontes - como por exemplo, em cada venda ou compra feita, contratação ou demissão de funcionários, interação através de pessoas em um comunicador ou rede social. Quase tudo pode gerar dados, e esses dados futuramente podem ser utilizados para gerar grandes estudos ou *insights* acerca de um conteúdo ou tema específico [Stobierski 2021].

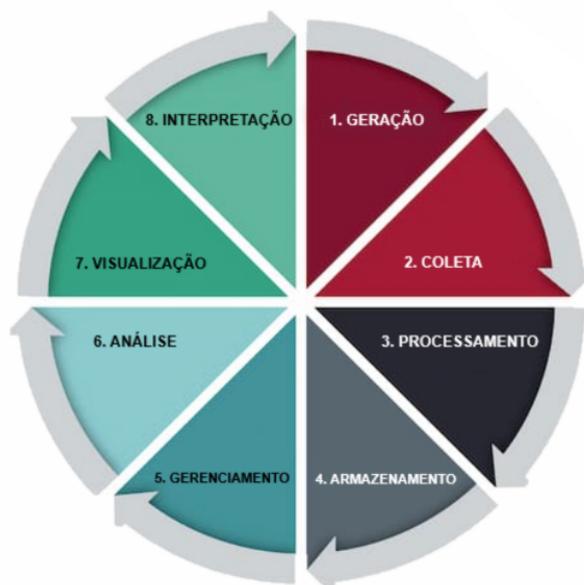


Figura 3 – Ciclo de Vida dos Dados. Extraído de [Stobierski 2021] e traduzido

3.1.1.2 Coleta de Dados

Tudo pode gerar dados, porém não são todos os dados que são coletados ou utilizados, e faz parte da etapa de coleta de dados, definir e identificar quais informações são relevantes para serem coletadas e quais são desnecessárias para o escopo definido. A coleta pode ocorrer de diversas formas e advindas dos mais diversos lugares, podendo esses dados ser coletados de *e-mails*, outros bancos de dados, pesquisas, formulários, entrevistas, equipamentos e sensores [Stobierski 2021].

3.1.1.3 Processamento

Nessa etapa, é feita a transformação dos dados crus da etapa anterior para uma forma de mais fácil acesso, com certa lógica e organização, onde os dados podem ser deduplicados, agrupados, encriptados e muitas vezes descartados caso não sejam úteis para o universo de aplicações utilizado, e assim são preparados para a próxima etapa, que será o armazenamento desses dados [Stobierski 2021].

3.1.1.4 Armazenamento

Na etapa de armazenamento, inicialmente se define onde os dados serão armazenados. Existem diversas possibilidades, podendo ser armazenados em *datasets*, DW físicos, servidores na nuvem e tantos outros métodos de armazenamento. Neste ponto é escolhido o melhor seja para um projeto específico ou para uma companhia inteira. Após a escolha, esses dados serão armazenados - preferencialmente com redundância. Isto é, quando os dados são guardados em mais de um local ou método de armazenamento, como por exemplo, *data centers* físicos espalhados em locais distantes numa mesma cidade, em países

diferentes, ou uma instância física e outra na nuvem. O objetivo é garantir que na ocorrência de problemas que possam comprometer um dos locais onde os dados estão disponíveis, exista um *backup* para atender prontamente e evitar a perda de dados [Stobierski 2021].

3.1.2 Desenvolvimento de BI

O desenvolvimento de BI é a etapa final do ciclo de vida dos dados, que engloba o gerenciamento, análise, visualização e interpretação dos dados. Porém, determinadas vezes durante a etapa de desenvolvimento de BI pode haver a necessidade de repetir uma das etapas do ETL, como um reprocessamento e novo armazenamento dos dados [Stobierski 2021].

3.1.2.1 Gerenciamento

Nessa etapa, tem-se o gerenciamento, organização, manutenção e entrega dos dados armazenados para as aplicações que os consultarem. São utilizadas ferramentas para fazer a extração dos dados de DW ou da nuvem, para que possam ser cruzados com outros dados e juntos serem analisados [Stobierski 2021].

3.1.2.2 Análise

Análise de dados refere-se ao processo de buscar *insights* ou retirar-se informações importantes baseadas nos dados que foram extraídos. Nesta etapa existem vários profissionais aptos a atuar com auxílio de suas respectivas ferramentas e métodos - como por exemplo, estatísticos utilizando modelagem estatística, analistas de dados utilizando algoritmos e linguagens de programação ou cientistas de dados aplicando métodos de inteligência artificial e mineração de dados. A depender da aplicação ou projeto, um ou mais profissionais podem atuar nessa etapa de análise [Stobierski 2021].

3.1.2.3 Visualização

Assim que os dados são analisados, organizados e *insights* são obtidos, é interessante trabalhar na forma como esses dados serão visualizados, normalmente através de *dashboards*. Na grande maioria das vezes, apenas olhar para uma tabela não é a forma mais eficiente de entender as informações, levando à necessidade de se criar visualizações mais diretas. “*Data visualization*” é a etapa onde se busca uma forma rápida de comunicar qual foi a análise feita nos dados, de modo que um certo público ou área de uma companhia que toma decisões seja capaz de entender. É também a etapa onde busca-se contar uma história desses dados, e essa visualização pode ser feita através de gráficos, histogramas, mapas de calor, semáforos, comparação com KPIs, curvas de variação histórica ou de tendência dos dados. Aqui existem centenas de possibilidades, sendo essa uma etapa muito mais conectada com os usuários finais, visto a construção dessas visualizações tem

que fazer sentido e ser de fácil compreensão desses usuários. E apesar de não ser sempre obrigatória, é uma etapa que vem crescendo cada vez mais quando se trata do ciclo de vida dos dados [Stobierski 2021].

3.1.2.4 Interpretação

Essa é geralmente a etapa final do ciclo de vida do projeto de BI, onde com o auxílio das análises e dos métodos de visualização, o usuário consegue entender e interpretar os dados de forma a obter um embasamento sólido que guie o processo de tomada de decisões relativas a um projeto ou aplicação [Stobierski 2021].

3.2 Ferramentas

Existem centenas de ferramentas que auxiliam BI, cada uma com determinada finalidade e complexidade. Aqui serão abordadas as principais ferramentas utilizadas tanto para a etapa de tratativa e análise quanto as principais da etapa de visualização.

3.2.1 Ferramentas de ETL

No mundo corporativo, muitas vezes a etapa de análise dos dados acaba se mesclando a etapa de ETL, pois por diversas vezes, pode-se ocorrer situações nas quais dados são consultados de um DW e carregados para uma máquina e nela são cruzados com dados de outras fontes de dados, como ferramentas de “*Customer Relationship Management*” (CRM), planilhas advindas do *Excel* [Gillis 2021] e outras ferramentas internas de gestão da empresa. Portanto, muitas vezes o analista ou cientista de dados acaba tendo que fazer novamente as etapas de ETL, porém agora guardando esses dados em algum *datamart* ou tabela na rede em que está conectado. Para a extração dos dados do DW, utilizam-se linguagens para consulta apropriadas a implementação do DW, essas podendo muitas vezes ser tanto linguagens “*Structured Query Language*” (SQL) como “*Not Only SQL*” (NoSQL).

3.2.1.1 Bancos de Dados Estruturados e o SQL

Bancos de dados estruturados são bancos que armazenam um conjunto específico de dados estruturados, ou seja, dados que são organizados em tabelas. Estas são compostas por uma coleção de linhas, também chamadas de tuplas, e colunas, também chamadas de atributos. Cada uma dessas colunas armazena um determinado tipo de informação, como datas, números inteiros, cadeias de caracteres ou *booleanos*.

SQL é uma linguagem de consulta a bancos de dados relacionais. No SQL existem vários comandos que permitem a criação, alteração, junção e exclusão de tabelas, além de consultas, agregações e ordenação de dados aplicando.

Os principais comandos do SQL são:

- “*Create*”: Cria novas tabelas em um banco de dados;
- “*Insert*”: Insere registros a uma tabela existente;
- “*Select*”: Resgata registros de uma tabela;
- “*Update*”: Atualiza registros de uma tabela;
- “*Delete*”: Exclui registros de uma tabela;
- “*Join*”: Combina duas tabelas em uma só;
- “*Drop*”: Exclui uma tabela de um banco de dados.

3.2.1.2 NoSQL

NoSQL é o termo utilizado para bancos de dados não-relacionais, esses bancos geralmente armazenam dados em um formato diferente do SQL, podendo armazená-los em forma de documentos “JavaScript Object Notation” (JSON), grafos, colunas e chave-valor. Geralmente esses bancos de dados são utilizados quando se deseja armazenar grandes volumes de dados que permitam escalabilidade e maior eficácia nas consultas, como plataformas de *e-commerce* ou jogos online. Os tipos de bancos de dados NoSQL e exemplos de sistemas gerenciadores desses bancos de dados são:

- **Orientado a Documentos:** Bancos de dados onde dados semiestruturados são armazenados na forma de documentos, geralmente do tipo JSON. Um exemplo de Sistema Gerenciador de Banco de Dados (SGBD) orientado a documentos é o MongoDB [MONGODB 2022];
- **Orientado a Colunas:** Os dados são armazenados em colunas, porém diferentemente dos bancos relacionais, este não utiliza tabelas e pode armazenar diversos tipos de dados em uma mesma coluna, facilitando assim com que o dado seja resgatado. Um exemplo de SGBD orientado a colunas é o Cassandra [Apache 2022];
- **Orientado a Grafos:** Nesse tipo de estrutura, os dados são armazenados em forma de grafos, distribuídos por nós conectados por arestas. Esse modelo proporciona grande velocidade nas consultas complexas e é muito utilizado em redes sociais e sistemas de detecção de fraude. Um exemplo de SGBD orientado a grafos é o Neo4j [Neo4j 2022];
- **Chave-valor:** Esse modelo, que aguenta uma maior carga de dados, e um dos tipos mais populares, tem como premissa armazenar chaves únicas, que por sua vez são

associadas a um valor ou dado de qualquer tipo. Para se fazer uma consulta, basta buscar pela chave. Um exemplo de SGBD orientado a chave-valor é o Riak [Riak 2022].

Em tratando-se de plataformas e ferramentas mais difundidas no mercado, podemos citar também a existência dos gigantes do mercado que são os serviços da Amazon e da Microsoft e ferramentas de *analytics* como o “*Statistical Analysis System*” (SAS) [SAS 2022] e o Microsoft SQL Server [Microsoft 2022a].

- **Amazon Web Services (AWS):** A AWS [Amazon 2022] é a plataforma de nuvem mais difundida e mais abrangente existente no mercado. Além de oferecer mais de 200 serviços, pode-se ressaltar os bancos de dados em AWS, que tem integração tanto a bancos relacionais, quanto NoSQL;
- **Microsoft Azure:** O Azure [Microsoft 2022b] é a plataforma de nuvem da Microsoft, e também pode oferecer suporte a vários serviços e dentre eles, os bancos relacionais e NoSQL.
- **Microsoft SQL Server:** É uma plataforma que fornece ferramentas para o gerenciamento dos dados e integração com BI. Através dela é possível fazer tanto a consulta em bancos de dados, quanto a manipulação de dados existentes em arquivos “*Comma-Separated Values*” (CSV) ou “*Microsoft Excel Open XML Spreadsheet*” (XLSX), o que permite uma gama gigantesca de possibilidades no gerenciamento e BI.
- **SAS:** Uma das maiores referências no mercado mundial em ferramentas para o processamento e análise estatística de dados. As ferramentas SAS integram a análise estatística com o acesso a bancos de dados, permitindo a criação de gráficos e geração de diversos tipos de relatórios desde a sua extração até uma parte básica de visualização, tudo isso em apenas um lugar. Permite também a manipulação de arquivos dos diversos tipos tais como CSV, XLSX e TXT.

3.2.2 Ferramentas de Visualização

Em se tratando de plataformas para a criação de visualizações, o que se busca são ferramentas de fácil uso, mas que tenham também um bom poder de processamento. Isto porque determinadas vezes é necessário que a mesma realize cálculos, e que seja altamente customizável, permitindo a criação de *dashboards* com uma vasta gama de formas de se visualizar um dado, cabendo ao desenvolvedor, escolher quais as mais adequadas para o projeto. No mercado existem três ferramentas que se destacam em visualização para BI: Grafana, Power BI e Tableau.

3.2.2.1 Grafana

Grafana é uma plataforma de visualização de dados *open source* desenvolvida pela Grafana Labs e lançada em 2014. Ela permite a criação de *dashboards* com a presença de gráficos e ferramentas que permitem a visualização de dados de forma a facilitar a interpretação. Pode se conectar com as mais diversas fontes de dados, seja com plataformas na nuvem, tais como AWS e Microsoft Azure, além de integração com MySQL, PostgreSQL, Splunk, entre outros. Alguns pontos que muitas vezes acabam fazendo as empresas optarem por não utilizar o Grafana, residem no fato deste ter algumas limitações em visualizações, organização e *design* para montagem de *dashboards* e suporte a uma quantidade limitada de tipos de dados. Apesar disso, por ser uma plataforma *open source*, o Grafana está cada vez mais evoluindo e a comunidade tem colaborado para cada vez mais inserir novos *plugins* que complementam as funcionalidades existentes na ferramenta [GrafanaLabs 2022].

3.2.2.2 Power BI

Lançado em 2013, o Power BI é a ferramenta de visualização de dados da Microsoft. Por conta da integração com outras ferramentas da Microsoft que já estão difundidas no mercado, o Power BI ganhou muita força. Muitos usuários acabam por construir gráficos e visualizações no Microsoft Excel, buscando as mais diferentes e improvisadas formas de construir essas visualizações. Porém, muitas vezes devido a complexidade para criar essas visualizações, o Excel não conseguia atender a todas as expectativas nesse tipo de aplicação. Ao perceber isso, a Microsoft viu a possibilidade de criar uma ferramenta específica para essas visualizações que fosse simples e permitisse que o usuário de Excel fosse capaz de operar nessa ferramenta e criar visualizações mais complexas, com maior volume de dados e de forma mais eficaz [Microsoft 2022c].

3.2.2.3 Tableau

Tableau é a ferramenta de análise e visualização de dados para BI mais difundida do mercado. A empresa Tableau Software foi fundada em 2003 na Califórnia e no ano de 2004 lançou a primeira versão do software Tableau Desktop, com a ideia de que a análise de dados deve ser mais acessível. Deste modo, os dados podem ser trabalhados de forma visual, e não apenas através de números e códigos, permitindo que com o uso do mouse, alguns cliques para arrastar e soltar já são o suficiente para que os dados estejam prontos para análise de forma visual.

O Tableau oferece algumas ferramentas para a manipulações mais complexas nos dados, como o Tableau Prep Builder e o Tableau Prep Conductor. Para visualização de dados, existe o Tableau Desktop, porém disponibilizado em duas versões: Tableau Desktop e Tableau Desktop Public.

- **Tableau Desktop:** É um *software* que permite modelar dados e criar visualizações através de planilhas, gráficos, *dashboards* e histórias. No fim, é possível publicá-las tanto em servidores locais, quanto na nuvem, permitindo a manipulação bases de dados de qualquer tamanho, das mais diversas fontes de dados. Além disso, oferece uma diversa gama de funções estatísticas, funções de modelagem preditiva e recursos de Inteligência Artificial, que permitem que o usuário faça perguntas em linguagem natural e obtenha diversas visualizações sobre seus dados [Tableau 2022a].
- **Tableau Desktop Public:** É a versão gratuita do Tableau Desktop, que foi inicialmente criado com o intuito de fortalecer a colaboração e o aprendizado da ferramenta no meio acadêmico, mas atualmente pode ser utilizado por qualquer pessoa ou empresa e conta com a grande maioria das funcionalidades da versão paga, permitindo que análises complexas sejam feitas sem custo algum. As principais diferenças em relação a versão paga são: Limite de dados de 10 milhões de registros por base de dados; e menor gama de integração com fontes de dados. Outro ponto importante sobre o Tableau Desktop Public é que ele não permite a publicação dos *dashboards* e bases de dados em servidores locais. Como a ferramenta tem o intuito de ser colaborativa, os *dashboards* só podem ser publicados na nuvem, o que pode ser um problema caso o projeto envolva dados confidenciais, porém é uma ótima forma de ter contato com a comunidade que busca sempre a evolução e a melhoria dos *dashboards*, trocando experiências e promovendo cada vez mais o estudo e aprendizado de ferramentas de visualização, tendo um forte incentivo da Tableau [Tableau 2022b].

4 Revisão Bibliográfica

4.1 A importância de BI

Para Romero et al. (2021), BI se tornou indispensável para tomada de decisão estratégica em companhias, indústrias e governos ao redor do mundo. Em análise feita por Romero et al. (2021) em mais de 39 estudos, foram encontrados diversos ambientes que foram positivamente impactados pelo uso de ferramentas BI, sendo o ambiente industrial o mais favorecido com cerca de 56% dos estudos indicando isso, seguidos pelos ambientes acadêmico e social, com 21% e 15% respectivamente. Nesse estudo é ressaltado que essas tecnologias BI, associada a *Big Data* e IoT, formam os pilares atuais do desenvolvimento das companhias, por suportarem na tomada de decisão, *forecasting* e economia corporativa. De acordo com Romero et al. (2021), num futuro próximo essas ferramentas se tornarão mais populares, suportando inclusive empresas menores em seu desenvolvimento.

Ferramentas de BI e *analytics*, facilitam a coleta e análise de dados, suportando a tomada de decisão, e empresas e executivos acreditam que quão melhor for feita a análise de dados, e suporte a tomada de decisões, mais valor é agregado a sua empresa, como resalta Rikhardsson e Yigitbasioglu (2018). Eles defendem a importância não apenas do uso, mas também do estudo e entendimento desses métodos na academia, que segundo Granlund (2011) ainda é muito limitado. Rikhardsson e Yigitbasioglu (2018) analisaram o impacto de BI em gerenciamento de contas através de 60 artigos e identifica que por mais que nos artigos retratados fique claro a importância do uso dessas ferramentas de BI em contabilidade gerencial, ainda são poucos artigos focados nas aplicações de BI e *analytics* para essa área. A maioria dos artigos encontrados são muito conceituais e não endereçam bem as principais atividades dessas aplicações, confirmando então a real necessidade desse assunto ter uma maior relevância e ser mais abordado na academia.

Nas últimas duas décadas, o ambiente corporativo tem se tornado mais complexo, competitivo e variável, criando a necessidade de ações e respostas rápidas a essa alta mudança na dinâmica do mercado para que as empresas sejam capazes de sobreviver e permanecer competitivas, o que aumentou a demanda para a tomada de decisões e de ações de forma efetiva, eficiente, rápida e com qualidade. Para Tripathi, Bagga e Aggarwal (2020), é nesse cenário em que se faz necessário a análise contínua da performance e dos resultados das empresas, onde as técnicas e ferramentas de BI se encaixam - o que foi comprovado através da análise feita em 553 artigos publicados dentre os anos de 2007 e 2018. Através desse estudo, foi possível analisar com o passar dos anos, quais eram as tendências de abordagem nos artigos, como a adoção, implementação e benefícios, e a tendência pelos métodos de BI mais difundidos e pesquisados através dos anos.

4.2 Implementações e usos de BI

No trabalho realizado por Kajava (2018), um módulo de BI foi implementado em uma empresa e foram analisados seus impactos. Para o estudo, foi implementado um *dashboard* utilizando a ferramenta Microsoft Power BI para analisar as oportunidades de vendas da empresa em que o estudo foi aplicado. Com esse estudo, foi possível notar um aumento significativo na produtividade e na capacidade da empresa de analisar e visualizar seu progresso em tempo real. Com isso a empresa conseguiu identificar quais produtos eram mais vantajosos, quais tinham gargalos na produção e quais eram as principais necessidades do mercado a respeito de seus produtos e definir assim estratégias para aumentar as vendas.

No trabalho de [Saputra e R (2021)] é feita a implementação de um *dashboard* em Tableau para o acompanhamento dos resultados de vendas para uma grande indústria de cosméticos focadas no *Business-to-Business*. Nesse trabalho, foram desenvolvidas visualizações que expusessem as variáveis de valor total de vendas no último ano, análise de Pareto, principais produtos vendidos por mês, análise da disseminação dos produtos em mapa, e quantidade de clientes por trimestre. Este *dashboard* ajudou o time de operação de vendas a analisar os resultados periodicamente, permitindo tomar decisões estratégicas para aumentar as vendas.

Fireteanu (2020) integra o Tableau a módulos de IoT para manipular e gerar a visualização de disponibilidade de um total de 56 estações de estacionamento em uma cidade, contando com 1000 sensores, que coletam dados em tempo real e fazem o *upload* desses dados para um servidor. Foram criadas visualizações utilizando mapas, que mostram através de um padrão de cores qual a disponibilidade das vagas naquela região. Através desse projeto, pôde-se identificar o alto poder de processamento do Tableau quando integrado com bases de dados muito grandes, permitindo fácil acesso aos dados e escalabilidade, visto que os projetos podem ser expandidos, sem impactar a forma como o *dashboard* dispõe os dados.

5 Proposta e Desenvolvimento

Neste capítulo serão apresentados os diversos aspectos a cerca das ferramentas, recursos e implementação do módulo BI proposto nesse trabalho através de um *dashboard* em Tableau.

5.1 Materiais e Métodos

Para o desenvolvimento desse *dashboard*, é necessária a utilização de uma base de dados modelo, para que seja feita a criação de campos calculados, gráficos, visões e filtros. Os dados utilizados foram obtidos através da plataforma Kaggle e representam a coleta de sensores presentes em uma planta de painéis solares na Índia, a cada 15 minutos, entre os dias 15 de Maio de 2020 a 17 de Junho de 2020. Eles estão disponíveis através do link <<https://www.kaggle.com/datasets/anikannal/solar-power-generation-data>> e foram extraídos no formato CSV, que também é um formato totalmente integrável com a plataforma BIPES. Para que esse módulo de BI fique o mais intercambiável possível, foi feita a mudança do nome de alguns campos da base de dados utilizada para um nome genérico, conforme mostrado na Tabela 1.

Substituições de nomes de campos	
Base modelo	Nome no Tableau
DATE_TIME	Date Time
SOURCE_KEY	Device Name
AC_POWER	Sensor1
DC_POWER	Sensor2
DAILY_YIELD	Sensor3

Tabela 1 – De-para de substituições de nomes de campos

Para a criação do *dashboard* presente nesse módulo, foi escolhido o software Tableau, já que o mesmo fornece ferramentas para fazer a importação, tratativa, correção de *outliers*, aplicação de métodos estatísticos e principalmente a possibilidade de criar-se os mais diversos tipos visualizações para os dados, podendo ser escolhidas as que melhores se encaixarem com a aplicação ou com o modelo de dados abordado. Para a manipulação e definição das visualizações a serem criadas no módulo, esse trabalho se baseou em algumas perguntas fundamentais que podem ser utilizadas para se definir a análise de conjuntos de sensores. Essas são:

1. É possível visualizar o comportamento dos dados de sensores de maneira histórica dispostos por períodos de tempo?
2. É possível ordenar e classificar os dispositivos que contêm determinado sensor por uma ordem de performance, deixando a cargo do usuário definir a quantidade de dispositivos que deseja visualizar?
3. É possível criar um alerta que mostre os dispositivos cujo valores dos sensores estejam operando fora da normalidade, indicando possível falha ou defeito (*outlier*)?
4. É possível visualizar o comportamento médio dos sensores dispostos em períodos do dia?
5. É possível visualizar uma previsão (*forecasting*) do comportamento dos dados por um período de tempo futuro baseado no comportamento histórico?

Tais perguntas serão respondidas, respectivamente, nas subseções 5.2.2 a 5.2.6

5.2 Construção do *dashboard*

Para a construção do dashboard em Tableau, foram utilizadas funções nativas do Tableau para a manipulação dos campos, sendo essas expostas na Tabela 2.

Função	Descrição
DATE()	Converte a variável para Data
DATETIME()	Converte a variável para Data e Hora
FIXED()	Calcula um valor utilizando dimensões especificadas
HORA()	Retorna a hora do campo de Data
MAX()	Retorna o máximo valor
SUM()	Realiza a soma
WINDOW_AVG()	Retorna a média da expressão existente na janela
WINDOW_STDEV()	Retorna o desvio padrão da expressão existente na janela
ZN()	Retorna zero se nulo

Tabela 2 – Funções nativas do Tableau utilizadas

5.2.1 Importação e Manipulação dos Dados

O Tableau permite conectar diretamente uma base de dados CSV no *software* para uso. Após a importação, os dados são mostrados já com uma detecção automática do tipo

de dado - tais como data, inteiro, cadeia de caracteres, etc. - como mostrado na Figura 4.

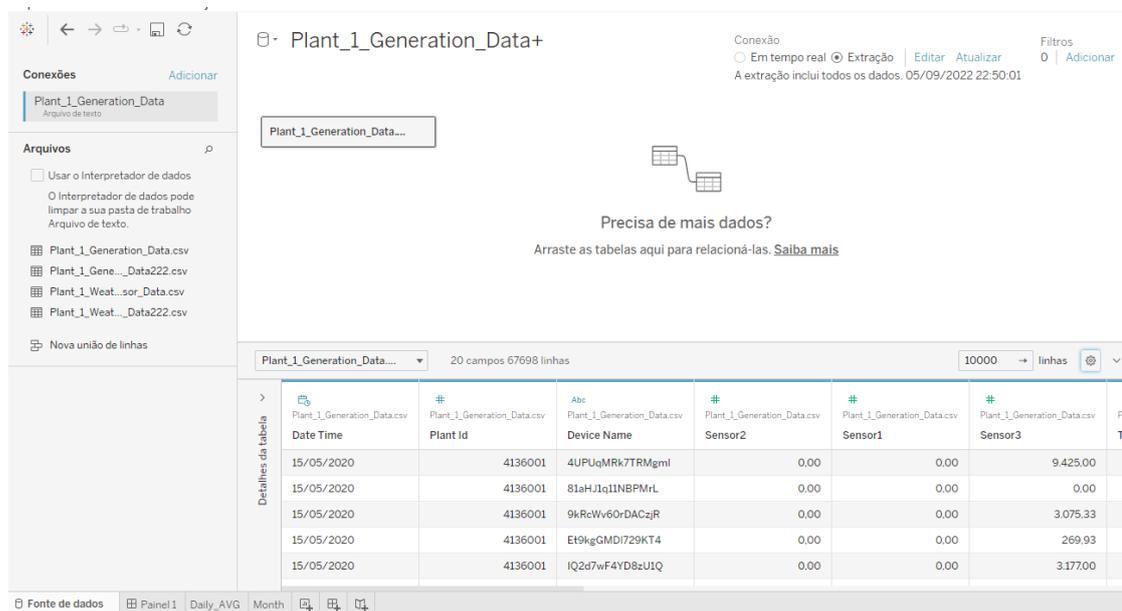


Figura 4 – Importação de base de dados para o Tableau (Autoria Própria)

Caso a detecção automática de tipo de dados falhe em alguma das colunas, basta clicar no símbolo que representa tal tipo de dado, e realizar a alteração manualmente, sem afetar a base de dados original, visto que o Tableau faz uma cópia dessa base de dados, evitando assim, sobrescrever a fonte.

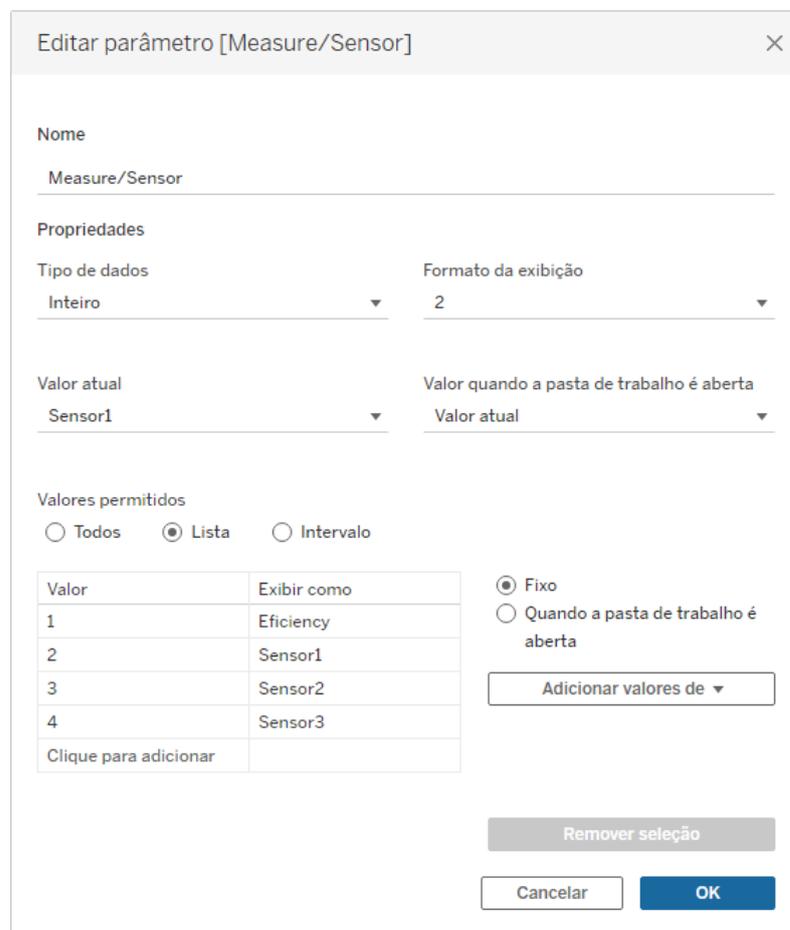
5.2.2 Comportamento histórico dos dados

Para a construção de uma visão dos dados históricos de sensores, duas variáveis serão majoritariamente utilizadas, uma de Tempo e outra de Valor. Para essa visualização, foi criado um gráfico de linhas, que indicará no eixo X, a passagem de tempo e no eixo Y, os valores do sensor escolhido. O campo de Data da base de dados utilizada, traz as informações da data completa com dia, mês, ano, hora e minutos. Para melhor visualização nesse gráfico, foi definida uma visão diária, portanto, através de uma funcionalidade do Tableau foi criado um campo calculado *DateDay*. O campo calculado é de uma nova variável, em que pode-se aplicar através de um código, funções e cálculos sobre um campo da base de dados. Foi aplicada no campo *Date Time* as funções *DATETIME()* e *DATE()*. A Figura 5 mostra como esse campo foi criado.

Para ser possível selecionar dentre diferentes tipos de Sensores, foi criado um parâmetro *Measure/Sensor*, como pode ser observado na Figura 6. Um “parâmetro” é uma funcionalidade do Tableau que permite a inserção ou seleção em lista de valores. Ele pode

Figura 5 – Campo calculado *DateDay* (Autoria Própria)

substituir uma constante em um cálculo ou filtro. Esse parâmetro permitirá que todas as visualizações do *dashboard* sejam filtradas e alteradas para a análise de um determinado sensor ou medida. Cada um dos valores selecionados nesse parâmetro será utilizado para se referir a um sensor ou medida específicos.

Figura 6 – Parâmetro *Measure/Sensor* (Autoria Própria)

Para a utilização dos valores dos sensores, foi criado um campo calculado chamado *Sensor*, como exposto na Figura 7. Esse campo calculado utiliza o valor retornado pelo parâmetro *Measure/Sensor* para definir qual será seu valor e segue as seguintes regras:

- Caso *Measure/Sensor* seja 1: Retorne o calculo de eficiência;
- Caso *Measure/Sensor* seja 2: Retorne a soma de *Sensor1*;
- Caso *Measure/Sensor* seja 3: Retorne a soma de *Sensor2*;
- Caso *Measure/Sensor* seja 4: Retorne o máximo valor de *Sensor3* (devido ao fato deste campo trazer a informação acumulada do dia);



Figura 7 – Campo calculado *Sensor* (Autoria Própria)

Por fim, a visualização recebeu como filtros de exibição a Data, presença de *outliers*, tipo de Gráfico e nome do Dispositivo. O resultado da visualização histórica é exposto na Figura 8.

5.2.3 *Ranking* de dispositivos

Para a criação do *ranking* de dispositivos baseados na performance, foram majoritariamente utilizados os campos *Device Name* e *Sensor*. Para essa visualização, foi criado um gráfico de barras horizontais, que indicará no eixo Y, os nomes dos dispositivos existentes no campo *Device Name*, e no eixo X, os valores da variável *Sensor*. Para que os valores aparecessem em frente a cada barra, a variável *sensor*, foi definida nesse gráfico como Rótulo de linhas.

Para auxiliar na visualização, foi aplicada uma escala de cores nas barras tendo como referência a variável *Sensor*, onde valores mais baixos aparecem em tons mais claros, e valores mais altos aparecem com tons mais escuros, como exposto na Figura 9.

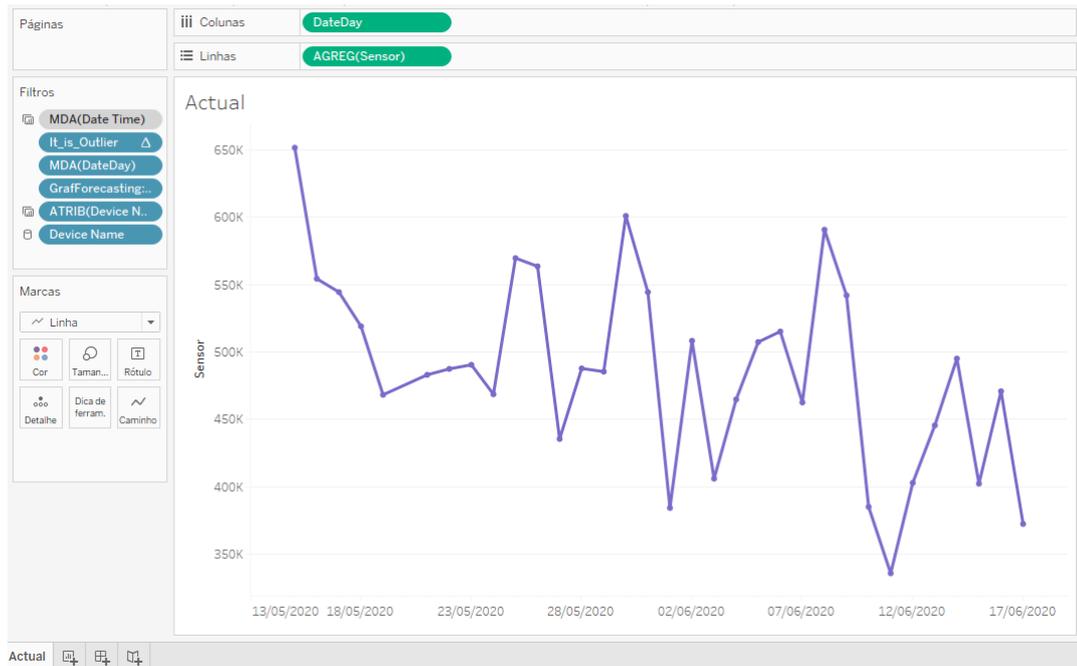


Figura 8 – Gráfico de comportamento histórico dos dados (Autoria Própria)

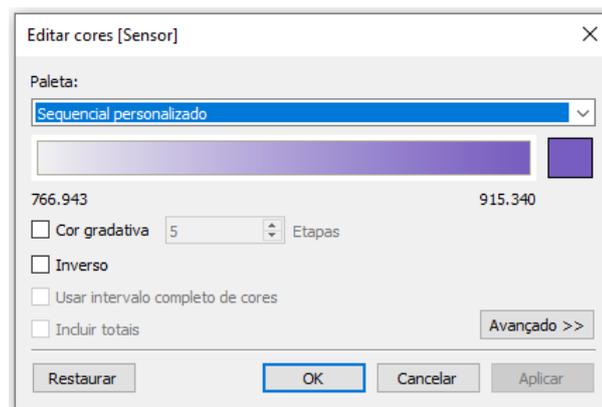


Figura 9 – Sequencial de tons baseados no valor da variável *Sensor* (Autoria Própria)

Para a seleção da quantidade de dispositivos que esse ranking deve mostrar, um parâmetro chamado *TOP_DEVICES* foi criado, como mostrado na Figura 10. Esse parâmetro aceita números inteiros, sendo no mínimo 1 e sem limites máximos de dispositivos a apresentar.

Por fim, a visualização recebeu como filtros de exibição a Data e nome do Dispositivo. O resultado da visualização do *ranking* de dispositivos é exposto na Figura 11.

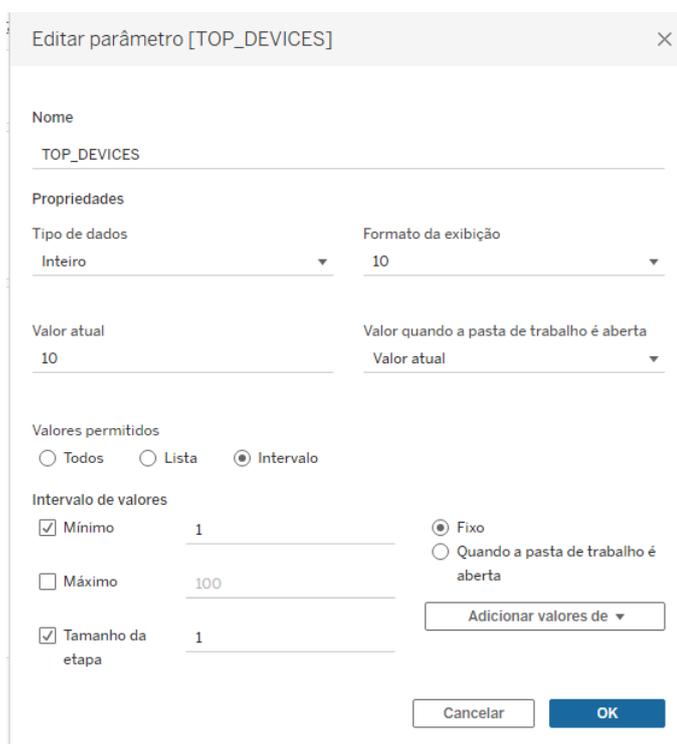


Figura 10 – Parâmetro para definição do número de observações no *Ranking* (Autoria Própria)

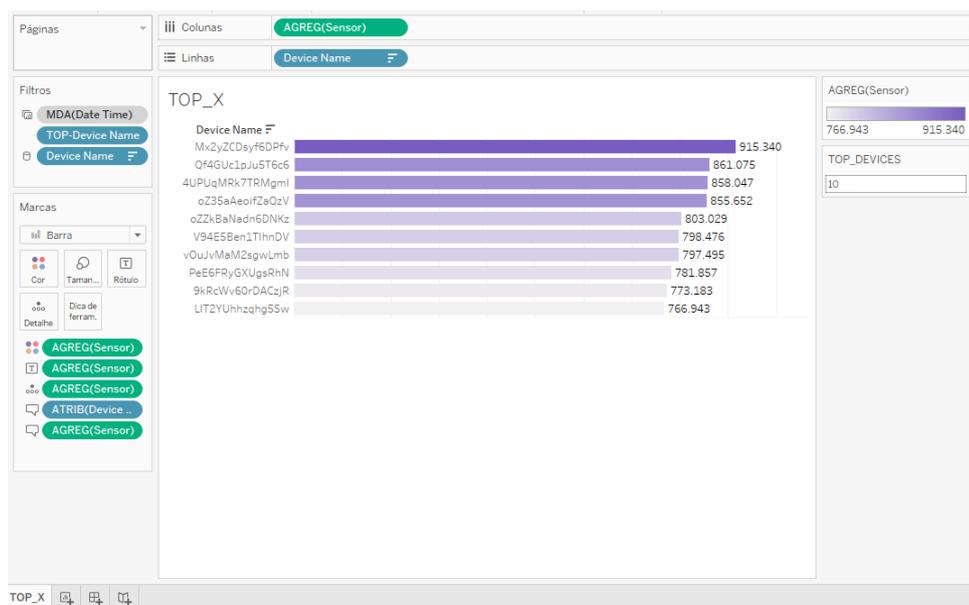


Figura 11 – *Ranking* de Dispositivos (Autoria Própria)

5.2.4 Alerta de não-conformidade

Para que se analise a não conformidade de sensores de um dispositivo, inicialmente deve-se definir um intervalo considerado de bom funcionamento de um sensor, para que então valores fora desses intervalos sejam classificados como *outliers*. Para que esse cálculo de *outliers* seja feito de forma mais automatizada possível, foram definidos dois campos calculados, *UpperBand* e *LowerBand*, que definem as fronteiras superiores e inferiores de normalidade respectivamente. Para a definição dessas fronteiras, o cálculo feito foi baseado na média e desvio padrão dos valores da variável *Sensor*. Ambos os campos calculados criados para esses limites superiores e inferiores são mostrados na Figura 12.

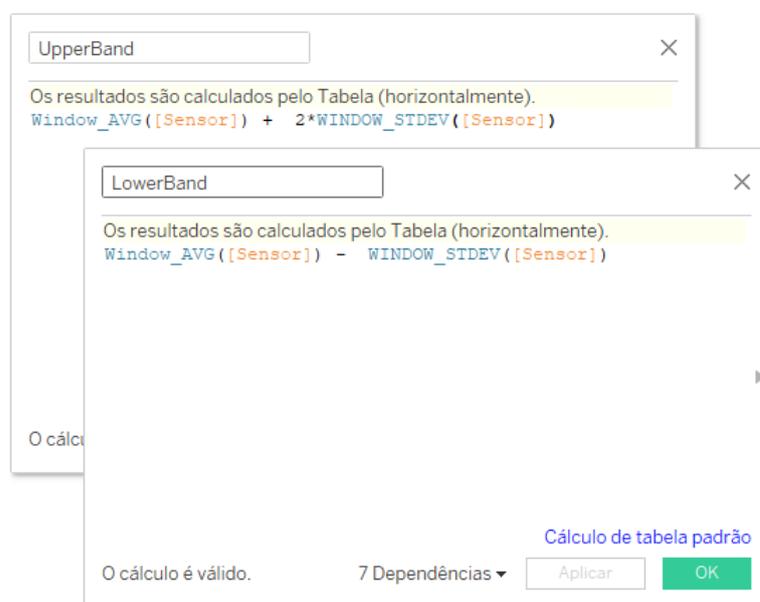


Figura 12 – Campos calculados *UpperBand* e *LowerBand* (Autoria Própria)

A partir dessas duas bandas de valores aceitáveis, foi criado o campo calculado *It_is_Outlier* (Figura 13), que analisa se os valores estão fora do intervalo definido pelas duas bandas e os classifica como *Outliers* ou *In Bounds*.

Para a visualização dos dispositivos em não conformidade e que são motivos de alerta, foi criada uma tabela, contendo o *Device Name* nas linhas, e seu valor correspondente da variável *Sensor* que é considerado um *outlier*. O filtro *It_is_Outlier* é aplicado a essa tabela, como filtro, fazendo com que apenas os valores fora da faixa determinada sejam exibidos. Outros filtros aplicados foram os de Data e nome do dispositivo, e na Figura 14 é observado o resultado desta visualização.



Figura 13 – Campo calculado *It_is_Outlier* (Autoria Própria)

Device Name	
mqwcsP2rE7J0TFp	625.458
NgDI19wMapZy17u	620.397
LYwnQax7tkwH5Cb	608.497
Et9kgGMDI729KT4	580.090
Quc1TzYxW2pYoWX	539.677

Figura 14 – Tabela de Alerta de não-conformidade (Autoria Própria)

5.2.5 Comportamento dos dados em períodos do dia

Para que seja possível a análise do comportamento dos dados de sensores em períodos do dia, foi feita a criação de um gráfico de linha, que mostra os valores distribuídos por hora de cada dia. Foi aplicada a função `HORA()` na variável `Date Time` que foi disposta no eixo X, e os valores da variável `Sensor` foram dispostos no eixo Y. Como filtros, foram aplicados a Data e nome do dispositivo e esta visualização é apresentada na Figura 15.

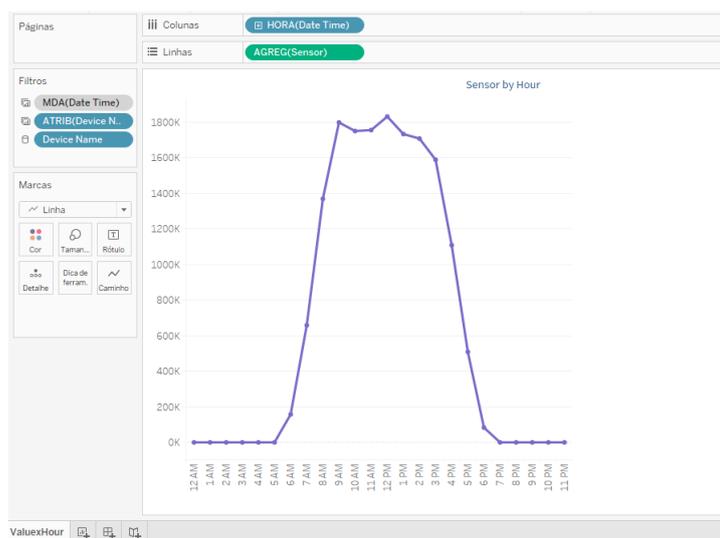


Figura 15 – Gráfico de comportamento dos dados por Hora (Autoria Própria)

5.2.6 Previsão de Comportamento

Para essa visualização de previsão de comportamento, foi criada inicialmente uma visualização análoga a presente na visualização do comportamento histórico dos dados na seção 5.2.2. Para prever os dados, usa-se a funcionalidade de Previsão do Tableau. Essa funcionalidade adiciona uma modelagem preditiva dos dados à visualização, permitindo a personalização da forma que o Tableau fará a predição desses dados. Caso as configurações de previsão estejam definidas como “Automático”, o Tableau analisará os dados e definirá a duração da previsão, o intervalo de agregação temporal e escolherá o melhor dentre os modelos para aplicar à visualização. Na Figura 16 pode-se observar a janela de seleção das opções de previsão.

Após definidas as opções e configurações de Previsão como Automático, o Tableau mostra na visualização em cores distintas da visualização histórica, qual a tendência do comportamento desses dados nos próximos dias, bem como uma faixa de intervalos de previsão, como pode ser observado na Figura 17. Para essa visualização, foram aplicados filtros de data e nome do dispositivo.

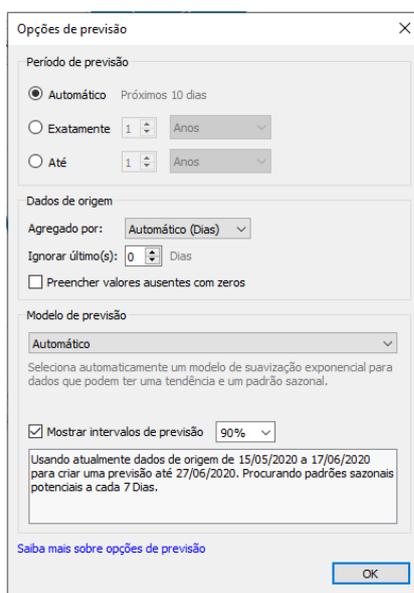


Figura 16 – Painel para configuração das opções de previsão (Autoria Própria)

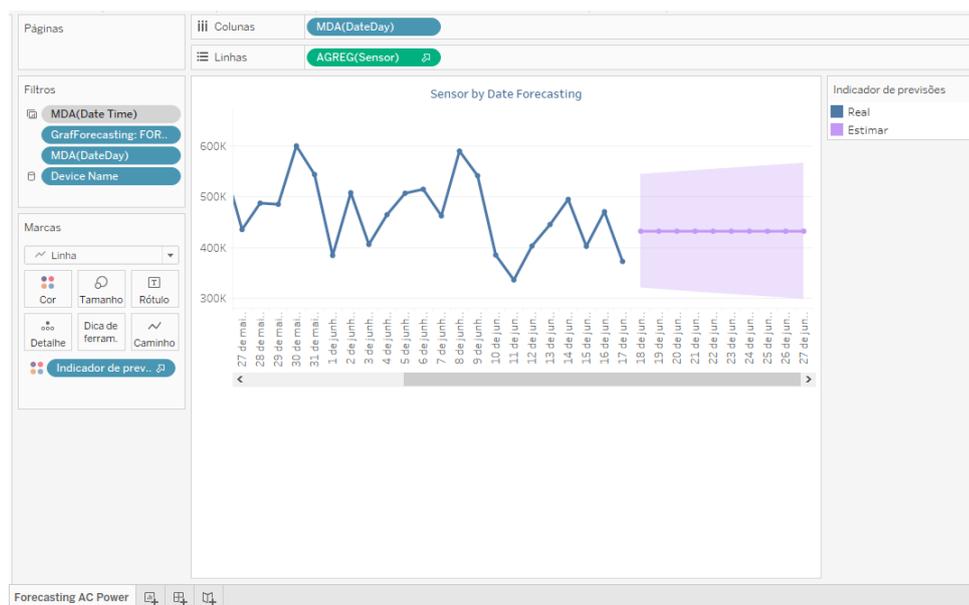


Figura 17 – Gráfico com histórico e previsão do comportamento de dados (Autoria Própria)

5.2.7 Indicador de dias com *outliers*

Para o usuário do *dashboard*, pode ser interessante visualizar os dados históricos eliminando-se os dias com *outlier*. Para isso, foi implementado um indicador de dias com *outliers*. Tem-se uma lista com os valores de *Date Time*, mostrado na Figura 18, e para que apenas os dias com *outlier* apareçam nessa lista, foi aplicado o filtro *It_is_outlier*, e um filtro de nome do dispositivo.

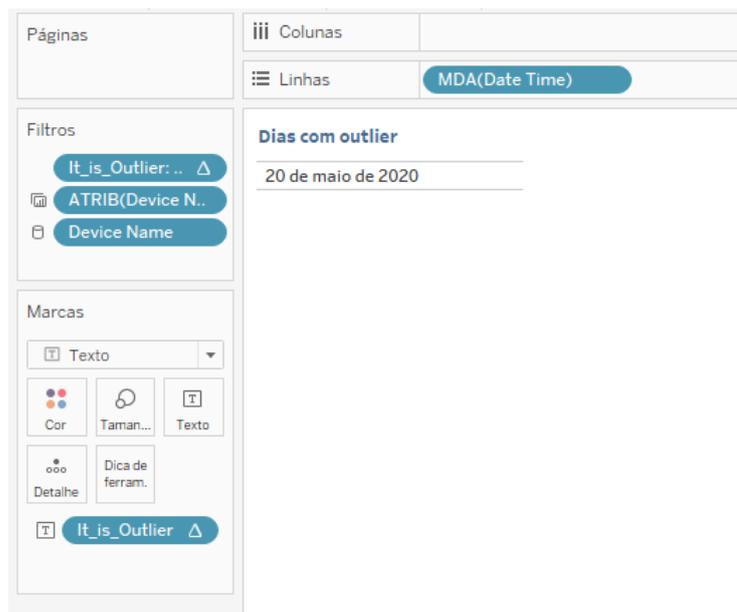


Figura 18 – Tabela de dias com *outliers* (Autoria Própria)

5.2.8 Indicador de média diária

Para a criação de um indicador de média diária, foi criada um campo calculado chamado *Sensor_AVG* nesse sensor, é calculada a média sobre os mesmos cálculos realizados para a criação do campo calculado *Sensor*. Para que essa visualização esteja sempre calculando médias diárias, é necessário fixar a agregação de data utilizada, e para isso é utilizada a expressão de nível de detalhe *FIXED*, garantindo que o cálculo sempre será feito por dias. A criação desse campo calculado pode ser observada na Figura 19.

A visualização final desse indicador é exposta na Figura 20. Os filtros aplicados nesse indicador, para que seja exposto a média diária, foram os filtros de data e nome do dispositivo.

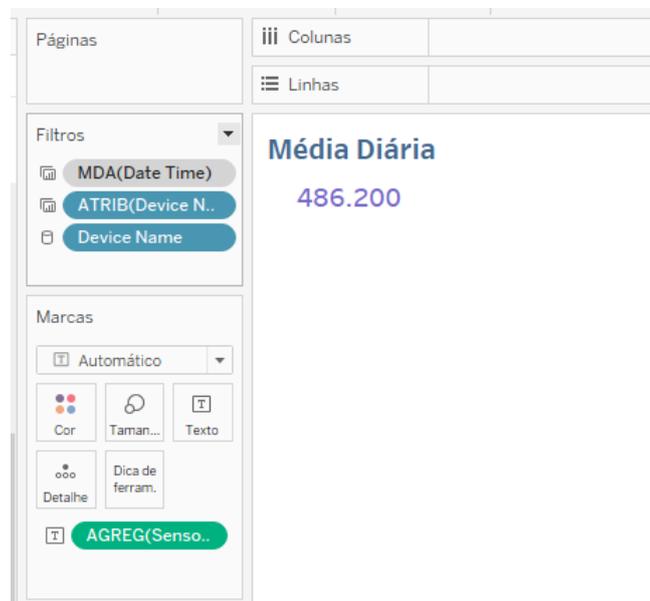
Figura 19 – Campo Calculado *Sensor_AVG* (Autoria Própria)

Figura 20 – Indicador de média diária (Autoria Própria)

5.2.9 Indicador de Total Geral

Para a criação de um indicador de total geral em todo o período de tempo das medições, foi criada um campo calculado chamado *Sensor_Total*, como pode ser visto na Figura 21, onde utilizam-se os mesmos cálculos realizados para a criação do campo calculado *Sensor*. A única exceção é no caso do *Sensor3* que é um campo com valores acumulados e para que este calcule sempre o resultado mensal, é necessário fixar a agregação de data utilizada, e para isso é utilizada expressão de nível de detalhe *FIXED* sobre uma função de data, definindo que independente do nível de agregação ou detalhe aplicado, o cálculo sempre será feito nessa variável sem influencia de filtros de data.

Figura 21 – Campo Calculado *Sensor_Total* (Autoria Própria)

Para que essa visualização apresente o valor total, não é aplicado filtro de data a esse indicador, sendo aplicado apenas filtro de nome do dispositivo. O resultado desse indicador é mostrado na Figura 22.

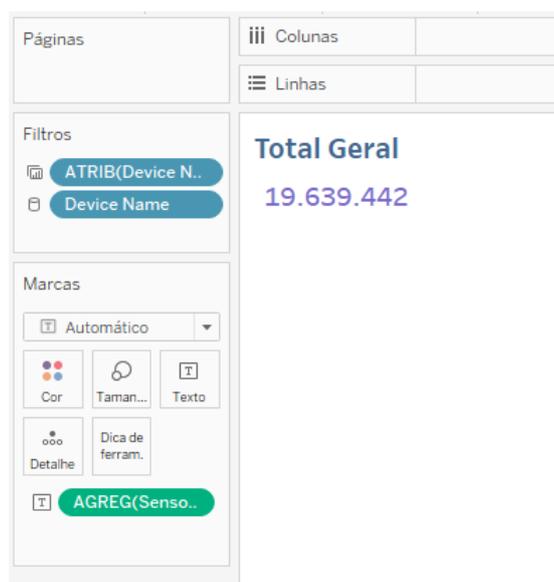


Figura 22 – Indicador de Total Geral (Autoria Própria)

6 Resultados e Discussão

O BIPES atualmente tem vários usuários, e cada um deles com aplicações distintas utilizando diferentes tipos de sensores. Devido a essa diversidade, durante a execução desse projeto houve um foco em desenvolver um módulo de BI que pudesse fazer análises temporais e fosse universal, de modo que possa ser utilizado independente do conjunto e tipo de dados coletado. Tendo isso como premissa, buscou-se não se ater especificamente aos tipos de dados na criação de cálculos e análises, ao mesmo tempo permitindo que seja customizável pelo usuário no momento de importar uma nova base de dados.

Além disso, outro foco importante, foi em dar a liberdade ao usuário de escolher a forma como deseja analisar as informações, disponibilizando filtros e parâmetros que personalizam a visão para que esta faça sentido para quem está a analisando. O usuário pode optar por ter uma visão histórica dos dados, ou uma visão histórica incluindo a previsão dos próximos dias, além disso, pode decidir qual sensor deseja visualizar, quais dispositivos, e definir se quer observar todo o período de dados, ou dias específicos. Na Figura 23 pode-se observar o *dashboard* completo e disposto na visão *Actual + Forecast*.

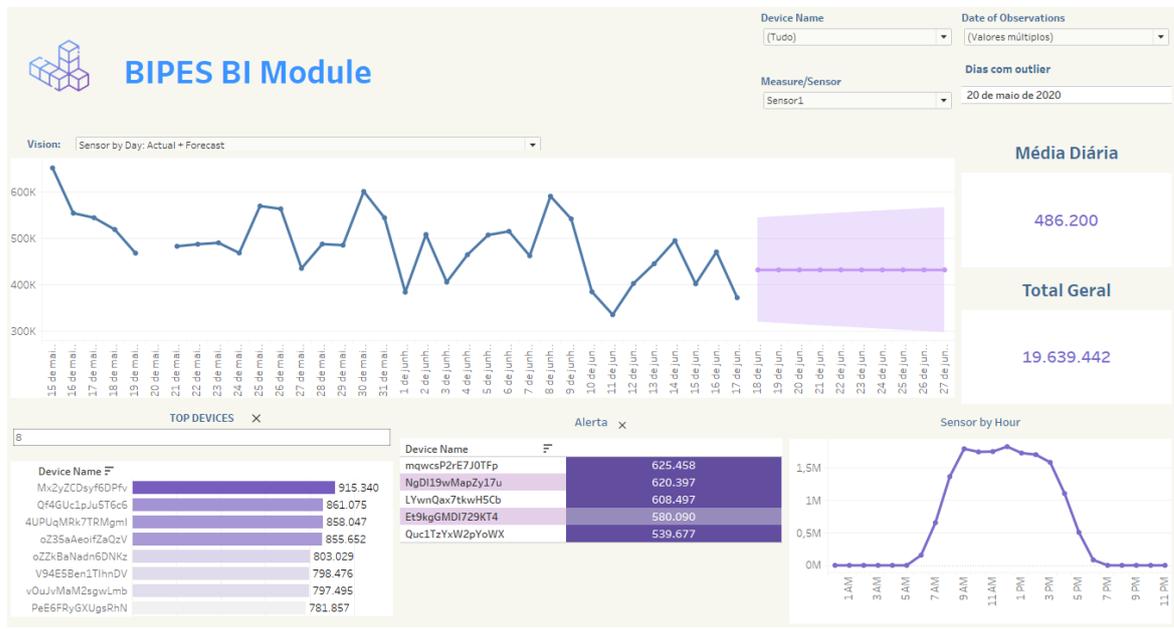


Figura 23 – *Dashboard* com visão de Dados Históricos + Previsão (Autoria Própria)

Outro ponto que foi de extrema importância, foi a exposição de clara de quais são e onde estão os *outliers*, o que permite que o usuário opte se eles devem ou não fazer parte de suas visualizações e análises no *dashboard* e identifique como essas escolhas afetam a previsão de comportamento. Nessa base de dados modelo utilizada, existem *outliers* muito expressivos e que são capazes de ilustrar, conforme visto na Figura 24, como essas

medições, que provavelmente são provenientes de falha ou mal-funcionamento, afetam a previsão de dados futuros e atrapalha nas visualizações quando comparado com a Figura 23 onde esses *outliers* estão filtrados.

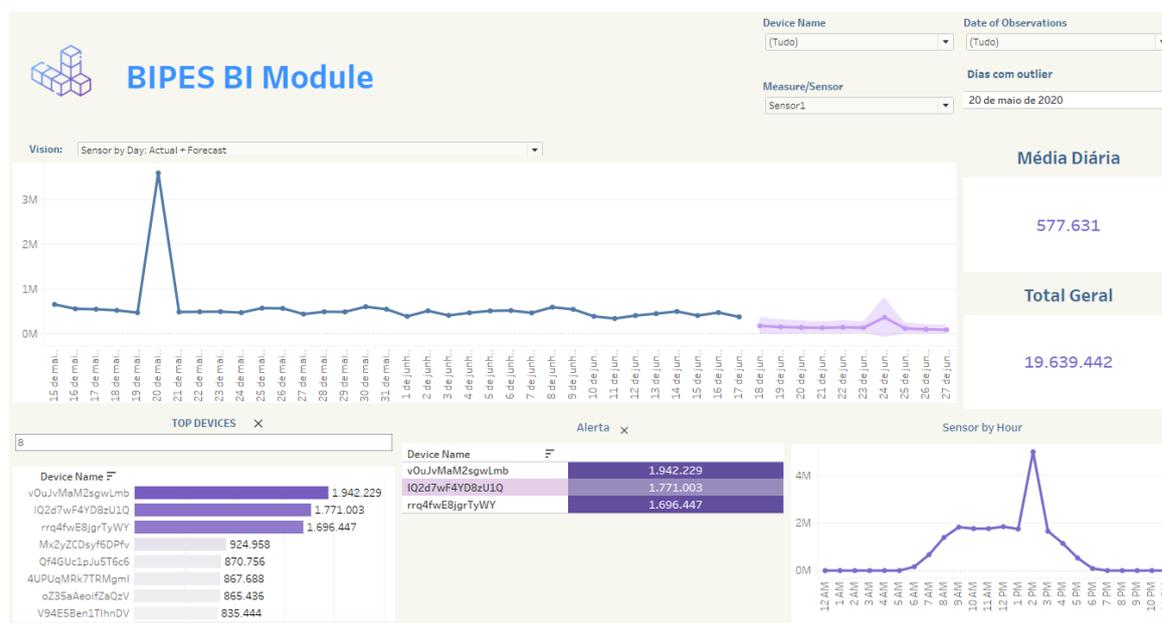


Figura 24 – Visão do *dashboard* com a presença de *outliers* (Autoria Própria)

Para integrar esse módulo BI ao BIPES é necessário que seja feita a disponibilização deste módulo *online*. Dessa forma, o *dashboard* foi disponibilizado na plataforma *online* do Tableau Public, de forma que os usuários possam visualizar, fazer *download* para o uso e edição. O resultado está representado na Figura 25 e pode ser acessado através do link <<https://public.tableau.com/views/BIPESBIModule/Painel1>>.

A integração do módulo projetado com as bases de dados geradas pelo BIPES, pode ser feita através do uso de um tipo de conexão específica que o Tableau oferece, que é a de extrair dados de um Conector de dados Web. Os dados do BIPES podem ser acessados no servidor do BIPES que contem os dados armazenados em arquivos do tipo JSON, como é o caso da implementação de um projeto de monitoramento de chocadeira feito utilizando o BIPES e que tem seus dados de temperatura armazenados em <<http://bipes.net.br/easymqtt/gettopic.php?session=chocadeira&topic=temperatura>>. Para conectar esses dados ao Tableau utiliza-se um conector JSON/XML cedido pela heroukapp, o qual conecta ao site e integra os dados ao Tableau conforme pode ser observado na Figura 26. Após a conexão ser realizada, é necessário adequá-los para o formato de dados existente no *dashboard*, tais como formato de data e nome de variável, repetindo passos que foram explicitados no Capítulo 5.

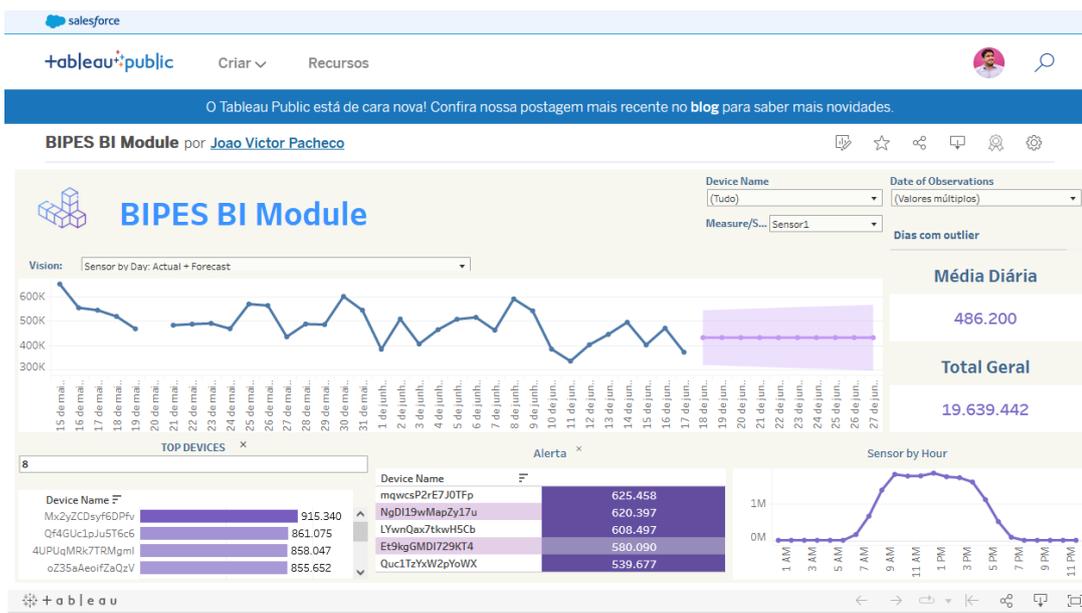


Figura 25 – Módulo de BI disponibilizado no *Tableau Public* (Autoria Própria)



Figura 26 – Conector de dados web para JSON (Autoria Própria)

Em se tratando da implementação de um módulo capaz de permitir análises claras e de forma visual a respeito de dados de sensores em determinados períodos de tempo, o módulo desenvolvido atende bem a tais requisitos, mostrando eficiência, clareza e bom grau de personalização da visão desejada, permitindo que o usuário consiga utilizar esse módulo no acompanhamento de indicadores e no auxílio a tomada de decisões.

7 Conclusão

O trabalho vigente propôs a implementação de um módulo de BI capaz de importar, processar e gerar visualizações de dados coletados por sensores e que fosse integrável com a plataforma BIPES. Pode-se concluir que ele atendeu a esse objetivo, permitindo auxiliar no acompanhamento de indicadores e processo de tomada de decisões. Foi muito construtivo conhecer sobre a existência e a crescente de ferramentas de “Desenvolvimento sem Código”, que permitem implementações sem a necessidade de se escrever um código e sobre as características e funcionalidades de uma ferramenta tão promissora e que já está disponível, que é o BIPES.

Além disso, através dos estudos sobre as etapas existentes em BI e dos itens da revisão bibliográfica, pode-se observar a importância dessas ferramentas não só para o meio corporativo, como também nota-se a importância de que esse tema seja mais abordado e mais estudado pela academia, frente as evoluções tecnológicas e crescente necessidade por métodos mais complexos para a análise de dados e tomada de decisões, atividades frequentes em todos os campos do conhecimento.

O módulo desenvolvido pode gerar as mais diversas interpretações, tais como análise de quais sensores tem uma melhor performance, permitindo com que o usuário busque quais são os aspectos que os tornam melhores e busque por dispositivos similares ou então dispositivos que estejam performando abaixo do desejado e identificar a partir disso se estes necessitam de uma manutenção ou substituição. Através do monitor dos sensores através de períodos do dia, é possível identificar momentos do dia em que os sensores não performam corretamente ou não coletam informações relevantes, permitindo que o usuário encontre, por exemplo, momentos do dia em que os sensores podem permanecer desligados, e com isso ocorrer uma economia de energia e redução dos custos envolvidos.

Um dos pontos mais interessantes a respeito da criação de visualizações e *dashboards* como o desenvolvido neste projeto, reside no fato de por conta da pluralidade de linhas de raciocínio do ser humano, cada usuário pode acabar extraindo novos e distintos entendimentos e análises através de um gráfico ou tabela já existente, que por diversas vezes o desenvolvedor das visualizações não previu a princípio, e aí mora a importância dessas ferramentas, elas buscam auxiliar o usuário a extrair da forma que for mais clara ou relevante a ele o que os dados querem dizer.

Como trabalhos futuros, pode-se ressaltar tanto a criação de novos gráficos e visualizações, que permitam ao usuário uma gama ainda maior de possibilidades de entendimento e análise a respeito dos dados. Outro ponto importante como trabalho futuro, é a implementação deste módulo em ferramentas que sejam assim como o BIPES, puramente

open source, permitindo que usuários ao redor do mundo colaborem mais ativamente sem estarem presos a uma ferramenta de uma companhia privada, pois apesar de que este módulo esteja atualmente publicado no Tableau Public que é uma plataforma pública, colaborativa e sem custos, a ferramenta e seu uso ainda podem estar de certa forma presos aos interesses de sua empresa desenvolvedora.

Referências

- AMAZON. Serviços de computação em nuvem- amazon web services (aws). [Online]. Disponível em: <<https://aws.amazon.com/pt/>>. Acesso em: 03 de Setembro de 2022, 2022. Citado na página 24.
- APACHE, S. F. Apache cassandra: Apache cassandra documentation. [Online]. Disponível em: <https://cassandra.apache.org/_/index.html>. Acesso em: 08 de Agosto de 2022, 2022. Citado na página 23.
- BARON, J.; GEFFROY, J.; MOTET, G. *Embedded System Applications*. [S.l.]: Springer US, (2013). Citado na página 16.
- BIPES. Bipes: Features. [Online]. Disponível em: <<https://bipes.net.br/wp/features/>>. Acessado em: 03 de Agosto de 2022, (2022). Citado 2 vezes nas páginas 9 e 17.
- CABALLAR, R. D. Programming without code: The rise of no-code software development. *IEEE Spectr. Tech Talks* [Online]. Disponível em: <<https://spectrum.ieee.org/tech-talk/>>. Acessado em: 02 de Novembro de 2020, (2020). Citado na página 16.
- CHEN, H.; CHIANG, R. H. L.; STOREY, V. C. Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, Management Information Systems Research Center, University of Minnesota, v. 36, n. 4, p. 1165–1188, (2012). ISSN 02767783. Citado na página 13.
- FIRETEANU, V.-V. Integrating tableau with internet of things acquiring projects. *2020 12th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, p. 1–5, (2020). Citado na página 28.
- GILLIS, A. S. *What is excel? Everything You Need to Know*. 2021. Citado na página 22.
- GOOGLE. Appsheet: The fastest way to build apps and automate work. [Online]. Disponível em: <<https://about.appsheet.com/home/>>. Acessado em: 03 de Agosto de 2022, (2022). Citado na página 16.
- GRAFANALABS. Grafana features | grafana labs. [Online]. Disponível em: <<https://grafana.com/grafana/?plcmt=footer>>. Acesso em: 03 de Setembro de 2022, 2022. Citado na página 25.
- GRANLUND, M. Extending ais research to management accounting and control issues: A research note. *International Journal of Accounting Information Systems*, v. 12, n. 1, p. 3–19, (2011). ISSN 1467-0895. Citado na página 27.
- HODGES, S. et al. Physical computing: A key element of modern computer science education. v. 53, n. 4, p. 20–30, (2020). Citado na página 16.
- JUNIOR, A. G. D. S. et al. Bipes: Block based integrated platform for embedded systems. *IEEE Access*, v. 8, (2020). Citado 4 vezes nas páginas 9, 14, 16 e 17.

- KAJAVA, E. Improving company performance through implementation of business intelligence tools : Implementation of a microsoft power bi in a case study company. (2018). Citado na página 28.
- MAGNUSON, B. D. Building blocks for mobile games: A multiplayer framework for app inventor for android. In: . [S.l.: s.n.], (2010). Citado na página 16.
- MICROSOFT. Sql server 2019. [Online]. Disponível em: <<https://www.microsoft.com/pt-br/sql-server/sql-server-2019>>. Acesso em: 03 de Setembro de 2022, 2022a. Citado na página 24.
- MICROSOFT. Serviços de computação em nuvem | microsoft azure. [Online]. Disponível em: <<https://azure.microsoft.com/pt-br/>>. Acesso em: 03 de Setembro de 2022, 2022b. Citado na página 24.
- MICROSOFT. Microsoft power bi - visualização de dados. [Online]. Disponível em: <<https://powerbi.microsoft.com/pt-br/>>. Acesso em: 03 de Setembro de 2022, 2022c. Citado na página 25.
- MONGODB. Mongodb: A plataforma de aplicação de dados. [Online]. Disponível em: <<https://www.mongodb.com/pt-br>>. Acesso em: 03 de Setembro de 2022, 2022. Citado na página 23.
- NEGASH, S.; GRAY, P. Business intelligence. Springer, (2008). Citado na página 18.
- NEO4J. Neo4j graph data platform. [Online]. Disponível em: <<https://neo4j.com/>>. Acesso em: 08 de Agosto de 2022, 2022. Citado na página 23.
- RIAK. Riak enterprise nosql database. [Online]. Disponível em: <<https://riak.com/>>. Acesso em: 08 de Agosto de 2022, 2022. Citado na página 24.
- RIKHARDSSON, P.; YIGITBASIOGLU, O. Business intelligence & analytics in management accounting research: Status and future focus. *International Journal of Accounting Information Systems*, v. 29, p. 37–58, (2018). ISSN 1467-0895. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1467089516300616>>. Citado na página 27.
- ROMERO, C. A. T. et al. Business intelligence: Business evolution after industry 4.0. *Sustainability*, v. 13, n. 18, (2021). Citado na página 27.
- SAPUTRA, E. A.; R, D. A. The implementation of business intelligence on smart sales dashboard using tableau (study case: Pt. derma konsep estetika. *International Research Journal of Advanced Engineering and Science*, v. 6, p. 223–226, (2021). Citado na página 28.
- SAS. Sas: Analytics, inteligência artificial e gerenciamento de dados. [Online]. Disponível em: <https://www.sas.com/pt_br/home.html>. Acesso em: 03 de Agosto de 2022, 2022. Citado na página 24.
- SINGH, R. P. et al. Internet of things (iot) applications to fight against covid-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, v. 14, n. 4, p. 521–524, (2020). Citado na página 13.

STOBIERSKI, T. *8 steps in the data life cycle*. 2021. Disponível em: <<https://online.hbs.edu/blog/post/data-life-cycle>> Acesso em: 01 de setembro de 2022. Citado 5 vezes nas páginas 9, 19, 20, 21 e 22.

TABLEAU. Tableau desktop - conecte, analise e visualize qualquer tipo de dado. [Online]. Disponível em: <<https://www.tableau.com/pt-br/products/desktop>>. Acesso em: 03 de Setembro de 2022, 2022a. Citado na página 26.

TABLEAU. Descubra mais - tableau public. [Online]. Disponível em: <<https://public.tableau.com/app/discover>>. Acesso em: 07 de Setembro de 2022, 2022b. Citado na página 26.

TRIPATHI, A.; BAGGA, T.; AGGARWAL, R. Strategic impact of business intelligence : A review of literature. *Prabandhan: Indian Journal of Management*, v. 13, p. 35, 03 (2020). Citado na página 27.