



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES  
**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

## **Desenvolvimento De Ferramental Para Auxiliar Na Mineração De Dados Ambientais Coletados Por Estações PCD**

### RELATÓRIO FINAL DE PROJETO DE INICIAÇÃO CIENTÍFICA (PIBIC/INPE/CNPq)

Guilherme Sallinas Furtado (UNIFESP, Bolsista PIBIC/CNPq)  
E-mail: [guilherme.sallinas@unifesp.br](mailto:guilherme.sallinas@unifesp.br)

Dr. Waldeir Amaral Vilela (INPE, Orientador)  
E-mail: [waldeir.vilela@inpe.br](mailto:waldeir.vilela@inpe.br)

Dr. Ricardo Toshiyuki Irita (INPE, Coorientador)



## **AGRADECIMENTOS**

A meus pais, que sempre me incentivaram a estudar e me apoiam em todas as minhas decisões.

Ao INPE por ter proporcionado a oportunidade de realizar esse trabalho e por mostrar, há anos, o quão importante é a pesquisa nacional.

Aos meus orientadores Waldeir Vilela e Ricardo Irita, que mesmo no contexto de pandemia e com uma maior dificuldade para nos encontrarmos sempre estiveram presentes para ajudar no que fosse preciso.

Ao CNPq por ter concedido a bolsa para essa pesquisa, valorizando e incentivando a ciência em nosso país.



## RESUMO

O setor de energias fotovoltaicas vem crescendo no Brasil e no mundo, significando que o melhor aproveitamento do potencial fotovoltaico representará impacto econômico significativo no setor. O Grupo de Dispositivos Fotovoltaicos, GDF/INPE, desenvolve pesquisas nesta área com intuito de contribuir e determinar tecnologias de células solares mais adequadas para cada região do Brasil. Este trabalho de Iniciação Científica está dentro deste contexto e tem como objetivo principal contribuir com a análise do potencial fotovoltaico na região Sudeste, utilizando dados disponíveis através de estações de coletas de dados localizados no IEE/USP em São Paulo e nos campus do INPE em Cachoeira Paulista (SP) e em São José dos Campos. Nestas estações, além de outros dados ambientais, são coletados dados irradiância global e espectral da radiação solar a cada minuto, o que gera grandes volumes de dados, e que, sendo obtidos simultaneamente e em diferentes formatos, dificultam ainda mais a utilização deles. Com objetivo de unificar o armazenamento de todas essas informações, vem sendo desenvolvido um banco de dados capaz de padronizar os dados de diferentes fontes e formatos, permitindo assim que o usuário consiga acessar facilmente as informações necessárias. Neste trabalho foi desenvolvido um Ferramental em Python para auxiliar o usuário na mineração dos dados. O programa desenvolvido é capaz de filtrar os dados a partir da data ou de quaisquer outros parâmetros lido do arquivo e, além disso, nele é possível selecionar quais dados serão visualizados de forma gráfica ou em forma de tabela. Para continuação deste trabalho, em atividades futuras, é recomendável acrescentar ao programa técnicas computacionais para aumentar a sua robustez e a velocidade de acesso, assim como, também, implementar mais funcionalidades, tal como, a validação dos dados armazenados no banco de dados.

**Palavras-Chaves:** Energia solar, Mineração de Dados, Python. Dados ambientais, Potencial fotovoltaico.



## LISTA DE FIGURAS

Figura 1 - Plataforma De Coleta De Dados.....	4
Figura 2 - Modelo de Dados.....	9
Figura 3 - Metodologia Do Projeto .....	13
Figura 4 – Funcionamento do ferramental.....	14
Figura 5 – Tela inicial do Ferramental .....	16
Figura 6 – Software com intervalo válido.....	17
Figura 7 – Software com intervalo inválido.....	17
Figura 8 - Dados Filtrados utilizando horas como parâmetro.....	18
Figura 9 – Interface de visualização dos dados .....	19
Figura 10 – Filtrar Dados a partir de um parâmetro.....	19
Figura 11 – Aba de filtros com somente um dos filtros marcado.....	20
Figura 12 – Dataframe com filtro aplicado.....	20
Figura 13 – Tela inicial da aba “grapher” .....	21
Figura 14 – Histograma da Temperatura Média.....	21
Figura 15 – Gráfico de linha temperatura média x umidade média.....	22
Figura 16 – Exportando Dados Como csv .....	23
Figura 17 – Exportar gráfico como png .....	23
Figura 18 – Exportar gráfico como html .....	24
Figura 19 – Gráfico aberto a partir de arquivo html .....	24





## LISTA DE SIGLAS E ABREVIATURAS

INPE	Instituto Nacional de Pesquisas Espaciais
GDF	Grupo de Dispositivos Fotovoltaicos
PCD	Plataforma de coleta de dados



## SUMÁRIO

1	INTRODUÇÃO.....	1
2	OBJETIVOS.....	3
3	FUNDAMENTAÇÃO TEÓRICA .....	4
3.1	Plataforma de coleta de dados .....	4
3.1.1	Sensores utilizados nas PCDs .....	5
3.1.1.1	Radiômetro .....	5
3.1.1.2	Barômetro.....	6
3.1.1.3	Pluviômetro.....	6
3.1.1.4	Termo-higrômetro.....	6
3.1.1.5	Anemômetro .....	7
3.1.1.6	Módulos Fotovoltaicos.....	7
3.1.1.7	Sistema de aquisição e armazenamento de dados .....	7
3.2	Banco de Dados Ambientais do GDF.....	8
3.3	Adaptação do banco de dados existente e proposição de ferramental .....	9
3.3.1	Complexidade de algoritmos.....	10
3.3.2	Notação $O(n)$ .....	10
4	MATERIAIS UTILIZADOS .....	11
4.1.	Plataforma de coleta de dados do GDF.....	11
4.2.	Computador .....	11
4.3.	Sistema de banco de dados atual (ICHIKURA, 2019).....	11
4.4.	Linguagens e plataformas de manipulação de dados .....	11
4.4.1.	Python .....	11
4.4.2.	Pandas .....	12
4.4.2.1.	PandasGUI .....	12
4.4.3.	NumPy.....	12
4.4.4.	Plotly.....	12
4.4.5	Tkinter .....	12
5	METODOLOGIA .....	13
5.1	Algoritmo Desenvolvido .....	14
5.1.1	Menu .....	14
5.1.2	Visualização Dos Dados .....	14
5.1.3	Filtrar Dados .....	15
5.1.4	Exportar Resultados .....	15
5.2.	Procedimento para validação do programa desenvolvido .....	15

6	RESULTADOS .....	16
6.1	Leitura do arquivo .....	16
6.2	Seleção do período desejado .....	16
6.3	Visualização Dos Dados .....	18
6.4	Filtrar Dados .....	19
6.5	Visualização Gráfica.....	20
6.6	Exportar Resultados .....	22
7	CONCLUSÃO .....	25
8	REFERÊNCIAS .....	26



## 1 INTRODUÇÃO

As mudanças climáticas e a preocupação mundial com o desenvolvimento sustentável tornam o setor de energia solar uma das principais escolhas para a geração de energia elétrica. Mesmo no Brasil, onde a fonte de energia é predominante as renováveis como as usinas hidrelétricas que são menos danosas que as usinas termoelétricas ou perigosas como as nucleares. Também se discute as desvantagens ambientais e sociais que as hidroelétricas promovem (imensa área alagada, emissão de metano, remoção da população ribeirinha e danos ao equilíbrio do ecossistema local) (PEREIRA, 2017). A energia solar, por outro lado, consegue ser convertida em eletricidade de forma direta a partir da incidência da radiação solar sem poluição, ruídos e é inesgotável (PV EDUCATION, 2019). E devido a esses motivos, ela é também uma das alternativas escolhidas na matriz energética do Brasil.

O Grupo de Dispositivos Fotovoltaicos, GDF/INPE, vem desenvolvendo pesquisas na área de energia solar com intuito de contribuir com o setor na determinação do potencial fotovoltaico e na escolha de tecnologias de células solares mais adequadas para cada região do Brasil. E este trabalho de Iniciação Científica está dentro deste contexto que tem como objetivo principal contribuir na otimização de ferramentas que auxiliem a análise do potencial fotovoltaico na região Sudeste brasileira, utilizando dados disponíveis através de estações de coletas de dados localizados no IEE/USP em São Paulo e nos dois campus do INPE: em Cachoeira Paulista (SP) e em São José dos Campos.

Nestas estações, além de outros dados ambientais, são coletados dados de irradiância global e espectral da radiação solar a cada minuto, o que gera grandes volumes de dados, e que, sendo obtidos simultaneamente e em diferentes formatos, dificultam ainda mais a utilização deles.

Com intuito de unificar o armazenamento de todas essas informações, vem sendo desenvolvido um banco de dados capaz de padronizar os dados de

diferentes fontes e formatos, permitindo assim que o usuário consiga acessar facilmente as informações desejadas.

Em trabalhos anteriores, foi desenvolvido uma primeira versão de banco de dados e um sistema gerenciador para que os dados dessas diversas fontes ficassem armazenados de forma unificada e padronizada (ICHIKURA,2019) porém, os sistemas utilizados realizam medição dos parâmetros constantemente e gera um grande volume de dados de diferentes padrões de formatação, tornando a manipulação dos dados complexa com a necessidade de familiarização de ferramentas computacionais avançadas pelo usuário. Observou-se assim a necessidade de desenvolver um ferramental que auxiliasse nessa mineração de dados e que tornasse mais simples a manipulação deles por parte dos pesquisadores, independente da origem da fonte de dados. Este foi o foco deste trabalho de Iniciação Científica.

Este relatório, portanto, apresenta as atividades e os resultados obtidos durante o período de um ano de trabalho de Iniciação Científica. Ele inicia com o estudo básico da origem dos dados ambientais, das diferentes formas de armazenagem e do banco de dados proposto por ICHIKURA,2019. Após isso apresentamos a metodologia com os materiais e métodos essenciais para a implementação do ferramental proposto, bem como as modificações/adaptações necessárias no banco proposto por ICHIKURA,2019. E, por fim, os resultados são mostrados e discutidos com indicação possíveis atividades futuras para melhorias no ferramental e na adaptação para diferentes usuários.

## **2 OBJETIVOS**

Este trabalho de Iniciação Científica teve como principal objetivo implementar um ferramental em um banco de dados que, em linhas gerais, auxilie na mineração da fonte onde estão armazenados os dados ambientais, e que auxilie na seleção e visualização dos resultados da análise.

O ferramental tem as seguintes funções:

- a) Realizar a seleção de parte dos dados a partir de parâmetros informados pelo usuário.
- b) Ser de fácil utilização por parte dos pesquisadores mesmo que esses não tenham conhecimento prévio em programação.
- c) Dispor de visualização gráfica para que a análise dos dados seja completa.
- d) Conseguir exportar os resultados dessa análise.



### **3 FUNDAMENTAÇÃO TEÓRICA**

Para o desenvolvimento deste projeto foi necessário o conhecimento dos conceitos básicos sobre os sensores utilizados nas plataformas de coletas de dados, juntamente com o entendimento do sistema de coleta de dados, entendimento dos conceitos utilizados no banco de dados atualmente disponível no laboratório GDF para, finalmente, apresentar a proposição dos métodos para medir a eficiência dos algoritmos utilizados no ferramental.

#### **3.1 Plataforma de coleta de dados**

Para que fosse possível obter dados ambientais de diferentes localidades foram desenvolvidas estações de coletas automática de dados gerados por sensores que detectam as grandezas ambientais, como radiação solar, temperatura e umidade do ar, pressão atmosféricas, velocidade e direção dos ventos, chuvas. Essas estações são denominadas PCDs, plataformas de coletas de dados e estão espalhadas em diversas localidades com diferentes finalidades. As PCDs abordadas neste projeto são dispositivos capazes de obter a partir de sensores e painéis fotovoltaicos o conjunto de dados relacionados a radiação solar e o potencial fotovoltaico para serem analisados em determinada região (SINDA, 2021).

Figura 1 - Plataforma De Coleta De Dados



Fonte: SINDA, 2021.

A Plataforma utilizada neste trabalho consiste em um conjunto de sensores, painéis fotovoltaicos e um sistema de coleta e armazenamento de dados (data logger).

### **3.1.1 Sensores utilizados nas PCDs**

Por definição, sensor é um dispositivo que relaciona informações de alguma grandeza física a partir de sua sensibilidade com alguma forma de energia ou alteração de suas características, como mudança de cor, volume, geração de energia luminosa, elétrica, cinética, térmica, entre outras. Os sensores utilizados nas PCDs podem ser analógicos, que podem assumir qualquer valor de saída que esteja dentro da sua faixa de operação, ou digitais, que podem assumir apenas dois valores de saída (1 ou 0) (THOMAZINI, 2005).

Nas PCDs do INPE geralmente são coletados dados meteorológicos gerados através dos sensores do tipo: Radiômetro, Barômetro, Pluviômetro, Termohigrômetro e Anemômetro. Nos itens a seguir serão descritos os sensores em uso no GDF.

#### **3.1.1.1 Radiômetro**

A irradiância solar ( $W/m^2$ ) que incide na superfície terrestre é composta por duas componentes: direta e difusa. A irradiância direta é a que não sofreu os processos radioativos de espalhamento e absorção da atmosfera (direção de incidência na linha imaginária entre a superfície e o sol), já a irradiância difusa engloba a radiação que teve interação com a atmosfera e espalhamento. A irradiância global é a soma das duas irradiâncias (PEREIRA, 2017).

Os radiômetros são responsáveis pela medição da irradiância solar (global, direta e difusa) em  $W/m^2$  e existem diferentes tipos e modelos para essa função. Os modelos de termopilhas tem uma resposta uniforme em relação ao espectral solar, com resposta para os comprimentos de onda de 200 nm a 4000 nm. Basicamente o piranômetro mede a irradiância solar através da temperatura de uma superfície de área conhecida. Essa superfície absorve a radiação solar e converte em calor provocando um aumento da temperatura e essa temperatura gerando uma corrente elétrica em um conjunto de termopares. A corrente gerada pelos termopares é proporcional a potência da

radiação absorvida dividida pela área da superfície absorvedora. Já os modelos com elementos sensores de semicondutores utiliza-se o efeito fotovoltaico para produzir um sinal elétrico a radiação solar. Estes têm menor custo e tempo de resposta rápido, mas apresentam limitação em termos de resposta espectral. O pireliômetro tem o mesmo princípio do piranômetro, são instrumentos normalmente de custo elevado e são utilizados para medir a radiação solar direta. Para a medição da radiação solar difusa é usado a técnica de anel de sombreamento que impede a passagem da radiação direta permitindo medir apenas a componente difusa (PEREIRA, 2017).

#### **3.1.1.2 Barômetro**

O barômetro é utilizado para medir a pressão atmosférica, que é a força exercida por metro quadrado de uma coluna de ar que está acima do local (JIMÉNEZ, 2018). A Unidade do SI para pressão atmosférica é Pascal (Pa), que é equivalente a 1 Newton por metro quadrado, e os dois principais tipos de barômetro são os de mercúrio e os aneroides sendo que os de mercúrio são mais precisos, porém devido a questões ambientais e legais não são mais comercializados. (World Meteorological Organization, 2008).

#### **3.1.1.3 Pluviômetro**

O pluviômetro é o sensor que determina o volume de precipitação por unidade de área. De forma geral os pluviômetros mais comuns possuem um sistema de balança que são recipientes abertos com volume previamente conhecidos que vão coletando a água até atingirem seu limite, quando isso ocorre ele se movimenta liberando a água gerando um pulso de tensão, que será lido e computado pelo sistema (FORNARI, 2013).

#### **3.1.1.4 Termo-higrômetro**

O termo-higrômetro tem como função medir a temperatura e a umidade relativa do ar. Para que os dados não tenham interferência este sensor é instalado dentro de um abrigo meteorológico com o objetivo de impedir que o vento, a chuva e a radiação solar atrapalhem a medição (FORNARI, 2013).

### **3.1.1.5 Anemômetro**

Esse sensor tem como objetivo medir a velocidade e a direção do vento. Na plataforma do GDF é utilizado um anemômetro mecânico com três conchas posicionadas a 120° entre si (FORNARI, 2013).

### **3.1.1.6 Módulos Fotovoltaicos**

A energia solar é convertida em energia elétrica através de células solares, que são componentes semicondutores capazes de transformar a radiação solar em eletricidade a partir do efeito fotovoltaico. Quando a radiação eletromagnética é absorvida pelo semicondutor os elétrons da banda de valência tem sua energia elevada fazendo com que ele vá para a banda de condução. Esses elétrons são removidos do semicondutor para um circuito externo. (PV EDUCATION, 2019).

Essas células são associadas eletricamente em série e paralelo para obter uma potência maior, e a esses conjuntos são chamados de módulos fotovoltaicos. Dentre as vantagens da utilização da energia gerada pelo efeito fotovoltaico tem-se que, não há emissão de poluentes durante sua utilização, é uma fonte inesgotável de energia, tem baixo impacto ambiental e as usinas podem ser construídas próximas ao consumidor. Em contrapartida, como desvantagens, tem-se que este tipo de geração de energia ainda é de elevado custo, e a geração é intermitente e também há necessidade de controle na manipulação de elementos tóxicos durante a fabricação desses dispositivos (NEVES, 2016).

### **3.1.1.7 Sistema de aquisição e armazenamento de dados**

Todos os sensores da plataforma são conectados a um sistema de aquisição e armazenamento de dados cuja a função é ler, armazenar e processar os sinais gerados nestes sensores. Na unidade de coleta de dados ambientais utilizada pelo GDF é utilizado o datalogger CR1000 da marca Campbell, que possui entradas analógicas sendo 16 entradas simples e 8 entradas diferenciais que podem ser configuradas individualmente e também 8 entradas digitais I/O ou 4 RS-232 (As portas I/O podem ser ajustadas em pares de

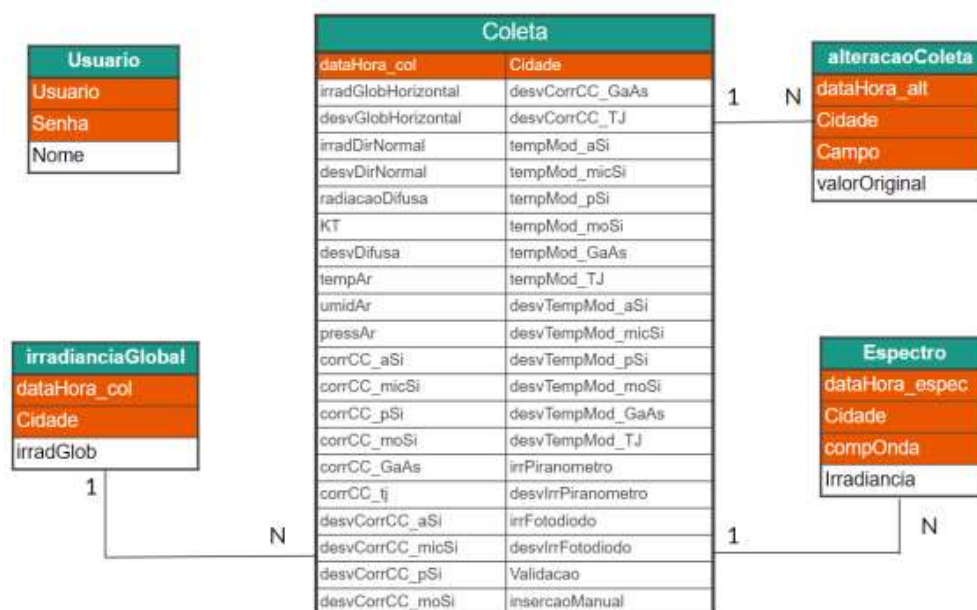
transmissor e receptor). Resolução analógica para os sinais de entrada de 0.33  $\mu$ V com conversor A/D de 13 Bits.

Além da escolha da escala de operação individualmente para cada entrada é possível programar qual será a periodicidade da aquisição até uma taxa de 100 Hz.

### **3.2 Banco de Dados Ambientais do GDF**

Para que todos os dados coletados pelos sistemas do GDF fossem armazenados de forma organizada, foi desenvolvido um banco de dados relacional. Este é formado por entidades e relacionamentos onde as entidades são análogas a tabelas e o relacionamento liga essas tabelas (ICHIKURA, 2019). O Software utilizado para modelar esse banco de dados foi o PostgreSQL, nele foram criadas 5 entidades: Usuário, Coleta, Espectro, irradianciaGlobal e alteracaoColeta. Na entidade “Usuário”, o administrador armazena o login das pessoas que poderão alterar informações do banco de dados. O usuário faz a importação e o armazenamento dos dados nas entidades “Coleta”, “Espectro” e “irradianciaGlobal”. Nestas entidades os dados são gerados pelos sensores que compõem o CR1000 e outros dados de outras fontes como por exemplo, da rede SONDA (Sistema de Organização Nacional de Dados Ambientais). Na funcionalidade “Espectro” são armazenados os dados de espectros da radiação solar que são coletados por um sistema dedicado a coleta de dados espectrais gerados por espectrorradiômetros solares. Em “alteracaoColeta” os dados são armazenados como um backup quando há alguma alteração de valor, neste caso os dados são armazenados e os dados originais são preservados (ICHIKURA, 2019). Na figura dois é possível ver a estrutura do modelo do banco de dados com suas entidades.

Figura 2 - Modelo de Dados



Fonte: Flavia Ichikura (2019)

Além do banco de dados propriamente dito, foi desenvolvido um programa em Java para auxiliar na consulta, cadastros e outras operações nesse banco, facilitando a utilização do banco de dados por usuários que não tem familiaridade com os códigos em SQL. Após verificar se o usuário está cadastrado na entidade “Usuario”, o programa irá oferecer as funcionalidades citadas anteriormente (ICHIKURA, 2019). O ferramental utilizará uma conexão com o banco de dados ou com um arquivo externo de acordo com o interesse do usuário.

### 3.3 Adaptação do banco de dados existente e proposição de ferramental

Como o banco de dados e sua interface para utilização do mesmo foram feitos em 2019, ocorreram problemas com a versão do código ao ser utilizado. Com isso foi necessário atualizar dependências do código como por exemplo de bibliotecas e do SGBD (Sistema de gerenciamento de banco de dados).

Além disso para que o ferramental tivesse uma boa performance, foram adotados os conceitos de complexidade de algoritmos e notação assintótica citados nos próximos tópicos.

### 3.3.1 Complexidade de algoritmos

Para definir se um código é eficiente ou não, foram criadas formas de categorizar a partir do tempo de execução com relação ao tamanho da entrada  $n$  (CORMEN, 2012)

### 3.3.2 Notação $O(n)$

A notação “ $O$ ” é a mais frequente quando se deseja analisar a eficiência de um código, sua definição diz que dada uma função  $g(n)$ , denotamos por  $O(g(n))$  o conjunto das funções  $\{f(n): \exists \text{ constantes } c \text{ e } n_0 \text{ tais que } 0 \leq f(n) \leq cg(n) \text{ para } n \geq n_0.\}$  (CORMEN, 2012).

## **4 MATERIAIS UTILIZADOS**

### **4.1. Plataforma de coleta de dados do GDF**

O primeiro contato com os dados é realizado pela plataforma PCD, onde os sensores ambientais ou dos experimentos em campo são instalados. Nela são obtidos os dados que serão convertidos pelo sistema de aquisição de dados para um tipo de informação que poderá ser inserido no banco de dados.

### **4.2. Computador**

Foi utilizado no laboratório GDF um computador com um processador Intel core i7 e 8 GB de RAM a fim de manter o banco de dados e o SGBD PostgreSQL além do código para gerenciar o banco e o ferramental desenvolvido.

### **4.3. Sistema de banco de dados atual (ICHIKURA, 2019)**

Sistema de banco de dados atual disponível no GDF é baseado no modelo ER (Entidade e Relacionamento) que consiste basicamente numa lógica em que entidades são semelhantes a tabelas e elas se relacionam entre si, fazendo com que, mesmo que as tabelas sejam diferentes entre si seja possível analisá-las de forma conjunta. O SGBD utilizado foi o PostgreSQL, que tem fácil integração com a linguagem utilizada no software de gerenciamento do banco, Java.

Foram também utilizados códigos em python para leitura de arquivos específicos e para a criação do banco na primeira vez que é utilizado no computador.

### **4.4. Linguagens e plataformas de manipulação de dados**

#### **4.4.1. Python**

Python é uma linguagem de alto nível, ou seja, sua sintaxe é similar a nossa língua, orientada a objetos assim como a linguagem Java utilizada em projetos anteriores, mas o seu grande diferencial para análise de dados são seus pacotes de bibliotecas que auxiliam e facilitam a manipulação desde a extração



do banco de dados, até a exportação dos mesmos. A seguir serão mostradas as principais bibliotecas utilizadas no desenvolvimento do ferramental.

#### **4.4.2. Pandas**

Pandas é utilizada para manipulação de tabelas criando um objeto chamado “DataFrame”, com isso ao retirar os dados do banco de dados estes são transformados em DataFrames para que possam ser manipulados pelas outras bibliotecas.

##### **4.4.2.1. PandasGUI**

PandasGUI é uma biblioteca que faz a integração do Dataframe do Pandas com uma interface gráfica amigável, assim facilitando a análise dos dados.

#### **4.4.3. NumPy**

Numpy cria um tipo de dado suportando vetores multidimensionais de forma integrada ao Pandas, além disso essa oferece diversas funções matemáticas para auxiliar em cálculos estatísticos.

#### **4.4.4. Plotly**

Para a criação de gráficos e visualização de dados foi utilizado a biblioteca plotly, facilitando a análise dos mesmos. Este pacote tem integração com o tipo de dados usado pelo NumPy e Pandas, portanto, além de auxiliar na visualização dos resultados desejados, essa previne conflitos e erros a partir do momento que não é necessário fazer nenhuma alteração no tipo do dado.

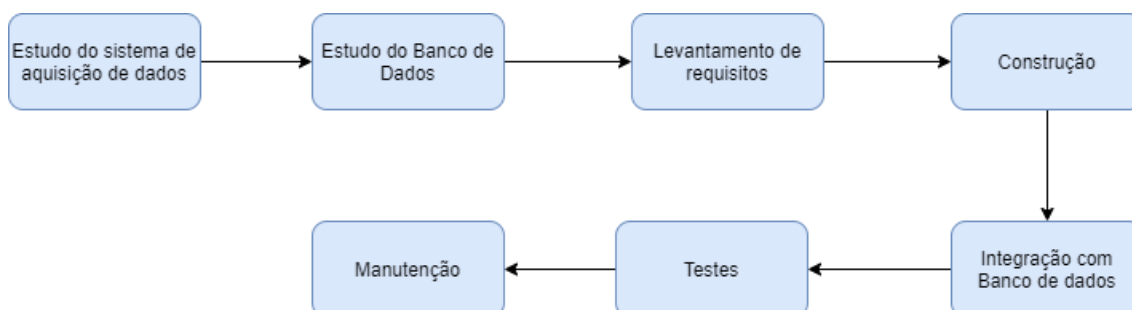
#### **4.4.5 Tkinter**

Para que o ferramental ficasse amigável aos pesquisadores que irão utilizá-lo, foi necessário desenvolver uma interface gráfica que facilitasse o uso do mesmo, com isso o pacote Tkinter, biblioteca padrão do Python, foi usado para o desenvolvimento da parte visual do programa, exceto na parte em que foi utilizado a biblioteca pandasGUI.

## 5 METODOLOGIA

Na Figura 3 são mostrados os passos seguidos para a construção do ferramental proposto. Inicialmente foi realizado um estudo da bibliografia para que fosse compreendido como era realizado a aquisição dos dados pelas plataformas de coleta de dados.

Figura 3 - Metodologia Do Projeto



Fonte: O Autor.

Após o estudo sobre o sistema de aquisição de dados, foi analisado o funcionamento do banco de dados e sua aplicação em Java. A análise permitiu decidir quais tecnologias seriam utilizadas para desenvolver o ferramental considerando que ele deveria ler o tipo de arquivo de interesse do usuário.

Com o conhecimento das tecnologias que já estavam sendo utilizadas foi levantado quais funcionalidades seriam necessárias no ferramental e qual seria a melhor maneira de implementá-las.

Na fase de desenvolvimento do trabalho realizou-se a codificação do ferramental. Essa etapa consistiu em duas partes, a primeira sendo o *back-end*, que consiste nos algoritmos que manipularão os dados e a lógica presente no programa, nesse está concentrada a essência do ferramental para que funcione. A segunda parte consiste no *front-end* que é a interface que é a interface que o irá visualizar,, A interface com o usuário deve ser interativa, por isso é importância que a utilização do ferramental seja simples.

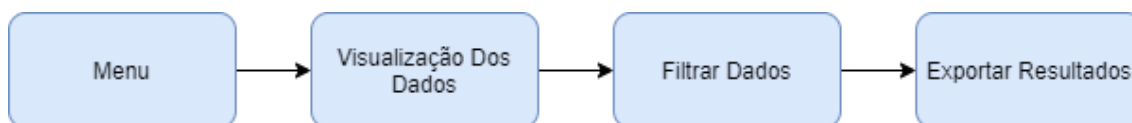
Com o ferramental desenvolvido a próxima etapa abordada foi a fase de testes, nessa fase foi verificada se todas as manipulações e cálculos com os dados foram feitos corretamente.

Por último como consequência da fase de testes veio a “manutenção”, com correções dos erros encontrados na etapa anterior e também inserindo melhorias.

## 5.1 Algoritmo Desenvolvido

Para que todas as necessidades do ferramental fossem cumpridas foi desenvolvido um programa em Python com auxílio das bibliotecas citadas no capítulo anterior. Visando a eficiência do código, todos os algoritmos que não são executados diretamente pelas bibliotecas também têm complexidade  $O(n)$ . A figura 3 mostra o funcionamento do ferramental a partir de um diagrama de blocos.

Figura 4 – Funcionamento do ferramental



Fonte: O autor.

A descrição de cada etapa é apresentada nos itens a seguir.

### 5.1.1 Menu

Ao iniciar o programa será solicitado ao usuário inserir qual o período desejado para ser analisado. Esta informação é baseada na data da primeira coleta e da última coleta conforme o arquivo lido e transformado em *Dataframe* pela biblioteca pandas.

### 5.1.2 Visualização Dos Dados

Após ser selecionado o intervalo desejado, o programa com auxílio da biblioteca PandasGUI abrirá uma outra janela que consiste em uma interface para que o usuário possa ver os dados e manipulá-los. Além disso é possível incluir mais arquivos dentro da interface para complementar a análise desejada.

### **5.1.3 Filtrar Dados**

Dentro dessa segunda interface é possível filtrar os dados de acordo com os parâmetros desejados na aba “filters”, facilitando que sejam removidos períodos que não servem para ser estudado, como por exemplo, os momentos em que não são coletados dados de irradiância.

### **5.1.4 Exportar Resultados**

Além de filtrar os dados o ferramental com auxílio da interface consegue gerar diferentes tipos de gráficos como de linhas, histogramas, de dispersão, mapa de calor dentre outros. Esses gráficos podem ser exportados como um arquivo de imagem (png) ou como um arquivo HTML que pode ser aberto no navegador e analisado de forma mais precisa mostrando os valores ponto a ponto de acordo com onde o cursor do mouse é colocado.

## **5.2. Procedimento para validação do programa desenvolvido**

Para validar o programa foi obedecido a seguinte lógica como teste do programa: Leitura do arquivo → Filtrar por período → Visualização em formato de tabela → Filtrar por parâmetro desejado → Visualizar graficamente → Exportar Resultados que serão demonstrados nos resultados.

## 6 RESULTADOS

Como resultados do desenvolvimento realizados neste projeto é apresentado a interface do programa com o usuário e o funcionamento do ferramental. Também é mostrando o procedimento prático previsto na metodologia para sua validação.

### 6.1 Leitura do arquivo

Ao iniciar o programa, com auxílio da biblioteca pandas o software irá ler o arquivo automaticamente.

### 6.2 Seleção do período desejado

Na figura 4, é mostrado a tela inicial do programa, solicitando o período em que o usuário deseja analisar.

Figura 5 – Tela inicial do Ferramental



GDF Ferramental

**Ferramental Mineração De Dados**

Insira o intervalo desejado:

Arquivo começa na data: 2018-12-01 e termina na data: 2019-01-08

Data Inicio:

Data Fim:

**Visualizar**

AGUARDANDO...

Fonte: O autor.

Em seguida quando é inserido um intervalo válido o programa se prepara para abrir a segunda interface onde os dados serão manipulados, mostrado na figura 6.

Figura 6 – Software com intervalo válido



The screenshot shows a window titled 'GDF Ferramental' with a light gray background. At the top, the title 'Ferramental Mineração De Dados' is displayed in bold blue text. Below it, the instruction 'Insira o intervalo desejado:' is centered. A line of text states 'Arquivo começa na data: 2018-12-01 e termina na data: 2019-01-08'. Two input fields are shown: 'Data Início:' with the value '2019-01-01' and 'Data Fim:' with the value '2019-01-07'. A pink button labeled 'Visualizar' is positioned below the fields. At the bottom, the text 'Abrindo...' is displayed in green.

Fonte: O autor.

Caso o usuário informe um intervalo em que o arquivo não suporte, ou seja, uma data menor que a data inicial do arquivo ou uma data maior que a data final do arquivo, será solicitado ao usuário para inserir um intervalo válido. A figura 7 mostra esta ocorrência.

Figura 7 – Software com intervalo inválido



The screenshot shows the same window as Figure 6. The title and instruction are identical. The file range text is also the same. However, the 'Data Início:' field contains '2017-01-01' and the 'Data Fim:' field contains '2018-01-01'. The 'Visualizar' button is present. Below it, the text 'Intervalo de data invalido' is displayed in red.

Fonte: O autor.

O usuário também pode inserir em seguida da data escolhida as horas minutos e segundos se desejados. Como exemplo, a Figura 8 ilustra esta situação, onde são selecionados os dados do dia 01/01/2019 das 05:00:00 as 19:00:00.

Figura 8 - Dados Filtrados utilizando horas como parâmetro



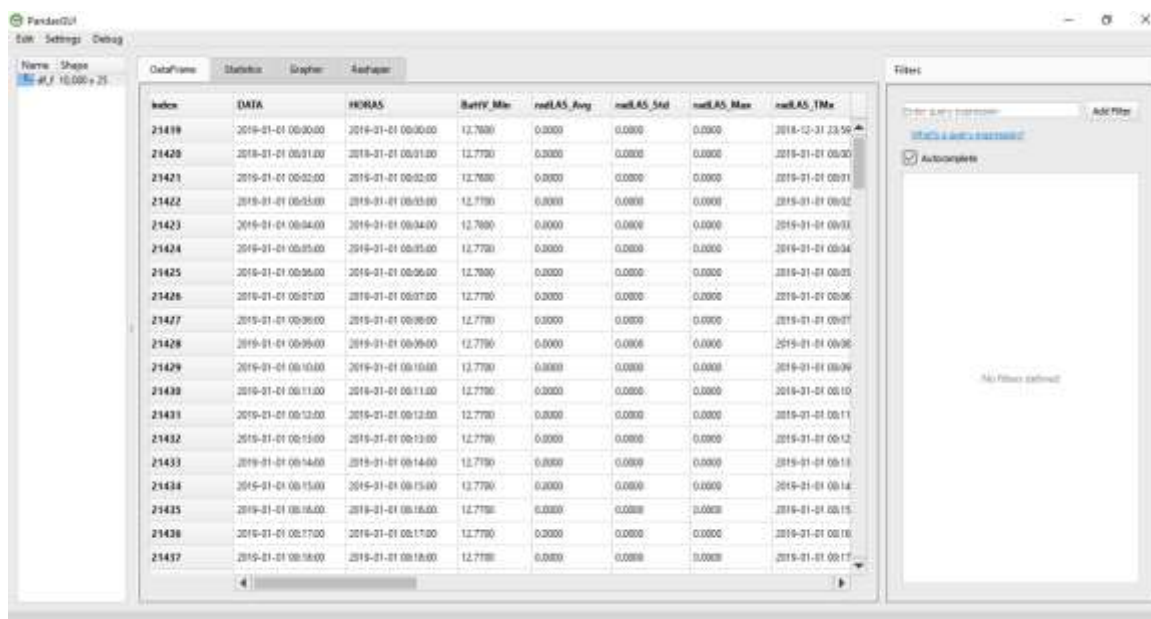
Fonte: O Autor.

Note que a primeira linha começa no horário indicado na “Data Início”, ou seja, as 05:00:00.

### 6.3 Visualização Dos Dados

Após inserir um intervalo válido, será aberto em outra janela a interface da biblioteca pandasGUI, contendo uma aba chamada “DataFrame”, que serve para ver os dados que foram previamente filtrados pela data, e que atualiza de acordo com as mudanças que são feitas ao longo do programa. A figura 9 mostra essa interface.

Figura 9 – Interface de visualização dos dados



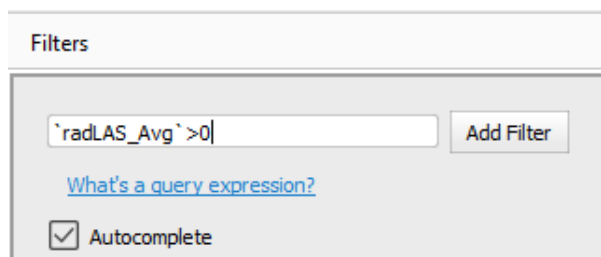
Fonte: O Autor

Na primeira aba a esquerda é possível arrastar outros arquivos à interface, para que durante a visualização gráfica sejam utilizados.

## 6.4 Filtrar Dados

No canto direito da figura 8, a aba “filters” é utilizada para que sejam filtrados os dados a partir de alguma coluna do DataFrame. Na figura 10, para exemplificar a utilização, é utilizado a coluna “radLAS\_Avg”.

Figura 10 – Filtrar Dados a partir de um parâmetro



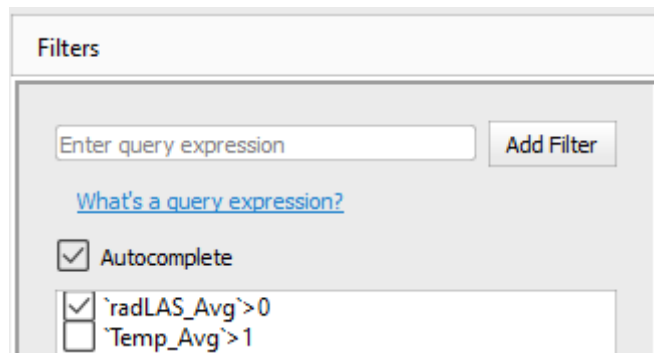
Fonte: O Autor.

Após clicar em “Add Filter” o filtro será aplicado, porém ele pode ser desmarcado caso não queira que seja utilizado, como é mostrado na Figura 11, a qual somente um dos dois filtros está marcado. Quando o usuário começa a



digitar o nome da coluna é mostrado uma aba para que seja completado automaticamente, para facilitar a escolha e também evite erros de digitação.

Figura 11 – Aba de filtros com somente um dos filtros marcado



Fonte: O Autor.

A figura 12 mostra a aba “DataFrame” após o filtro ter sido utilizado. Note que como o radiômetro só começa seu funcionamento a partir de um determinado horário. As linhas em que a hora da coleta tinha o valor daquela coluna menor ou igual a 0 foram excluídas.

Figura 12 – Dataframe com filtro aplicado

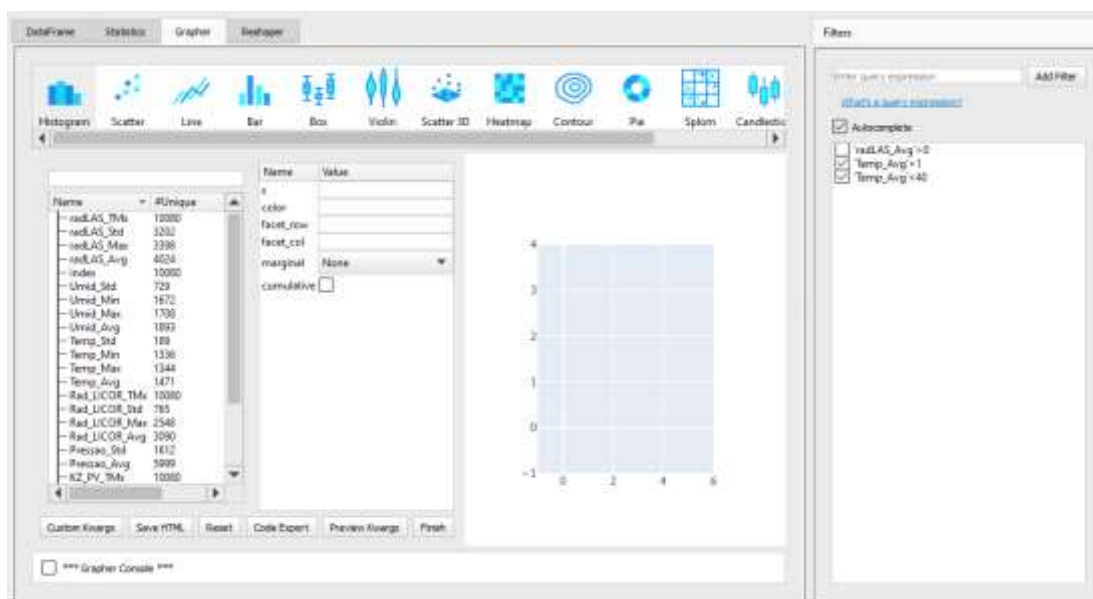
Index	DATA	HORAS	BattV_Min	radLAS_Avg	radLAS_Std	radLAS_Max	radLAS_TMe
21718	2019-01-01 04:59:00	2019-01-01 04:59:00	12.6900	0.0370	0.0630	0.2220	2019-01-01 04:59
21719	2019-01-01 05:00:00	2019-01-01 05:00:00	12.6900	0.2220	0.0000	0.2220	2019-01-01 04:59
21720	2019-01-01 05:01:00	2019-01-01 05:01:00	12.6900	0.2220	0.0000	0.2220	2019-01-01 05:00
21721	2019-01-01 05:02:00	2019-01-01 05:02:00	12.6900	0.2220	0.0000	0.2220	2019-01-01 05:01
21722	2019-01-01 05:03:00	2019-01-01 05:03:00	12.6900	0.2220	0.0000	0.2220	2019-01-01 05:00
21723	2019-01-01 05:04:00	2019-01-01 05:04:00	12.6900	0.2580	0.0630	0.4450	2019-01-01 05:03
21724	2019-01-01 05:05:00	2019-01-01 05:05:00	12.6900	0.4450	0.0000	0.4450	2019-01-01 05:04
21725	2019-01-01 05:06:00	2019-01-01 05:06:00	12.6900	0.4820	0.0630	0.6670	2019-01-01 05:06

Fonte: O Autor.

## 6.5 Visualização Gráfica

A aba “Grapher” é responsável por gerar e mostrar os dados por meio de gráficos. A Figura 12 mostra essa aba aberta.

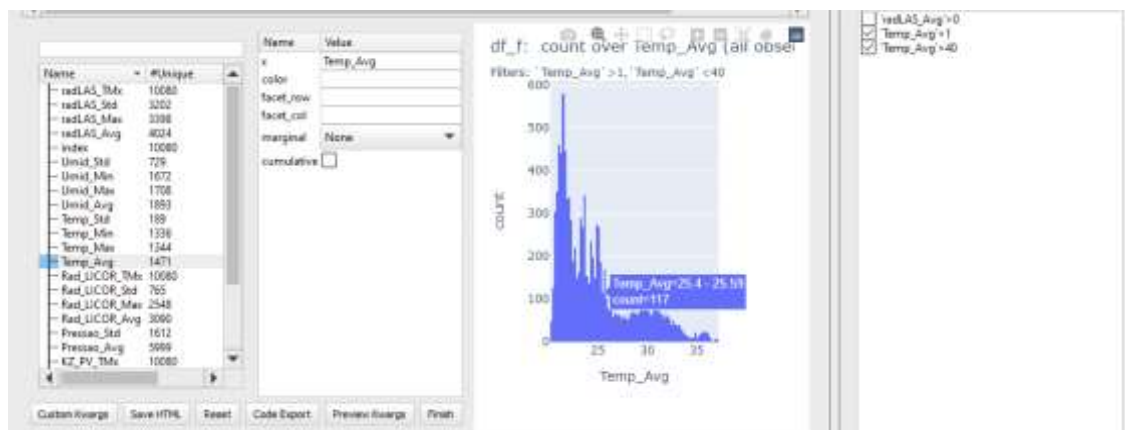
Figura 13 – Tela inicial da aba “grapher”



Fonte: O Autor.

Nessa aba é possível selecionar qual tipo de gráfico será utilizado através da seleção do formato do gráfico na parte superior da aba. No canto esquerdo são mostradas todas as colunas do DataFrame, que podem ser arrastadas para as linhas em branco na aba consecutiva a direita. A figura 14 mostra um histograma quando arrastado a coluna “Temp\_Avg” para a linha “x”.

Figura 14 – Histograma da Temperatura Média



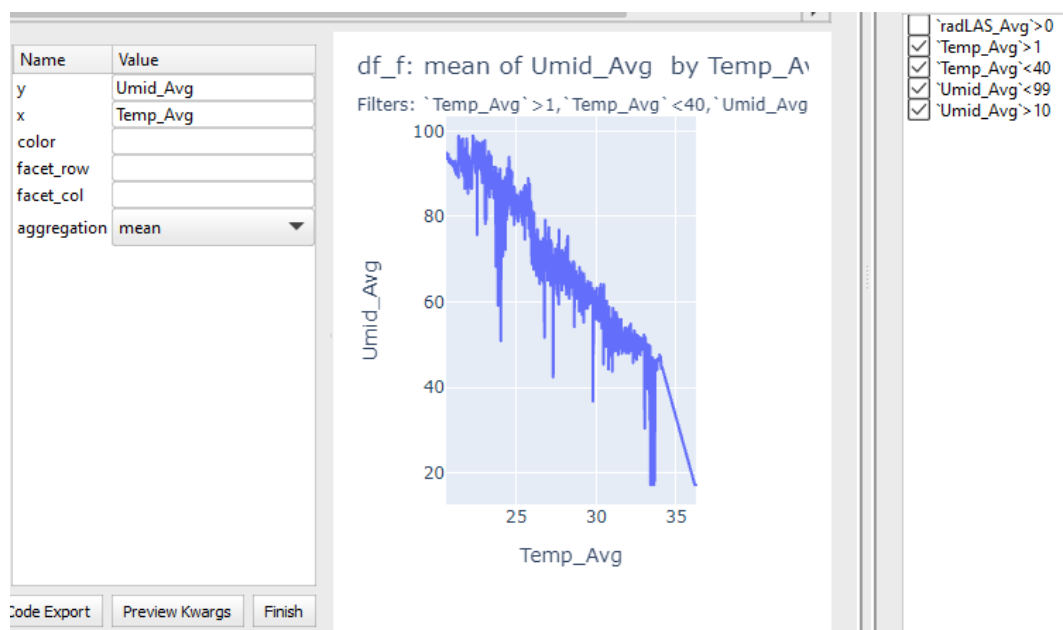
Fonte: O Autor.

Note que a etiqueta em azul mostra a faixa da coluna onde o cursor está apontado, e o número de linhas em que está naquela faixa. Além disso acima do gráfico estão dispostas algumas funcionalidades que auxiliam na

visualização do gráfico como aumentar e diminuir o zoom, exportar o gráfico como png, redimensionar e outras.

A Figura 14 mostra como um exemplo um gráfico de linhas entre umidade média e temperatura, com filtros aplicados.

Figura 15 – Gráfico de linha temperatura média x umidade média



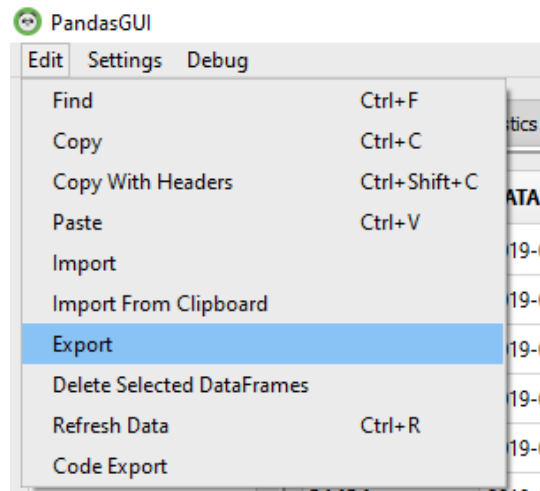
Fonte: O Autor.

Todas as funcionalidades aplicadas no histograma também servem para os outros tipos de gráficos. Além disso a interface contém outras funcionalidades que podem ser exploradas pelo usuário como por exemplo a cor da linha que pode diferenciar o tipo de dado caso eles tenham uma classificação diferente.

## 6.6 Exportar Resultados

Ao finalizar a análise dos dados e gerar suas visualizações o programa disponibiliza uma opção de exportar os dados filtrados no formato "csv" e duas opções de exportar os gráficos criados: Como imagem PNG e como arquivo HTML. A figura 15 mostra como os dados após ser filtrados podem ser exportados.

Figura 16 – Exportando Dados Como csv

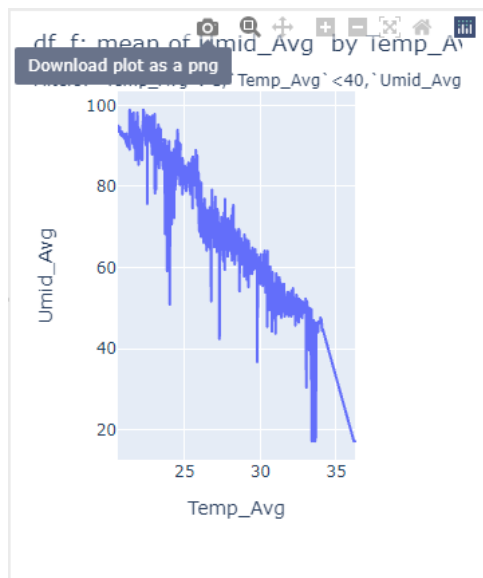


Fonte: O Autor.

Na aba “Edit” do canto superior esquerdo da janela, tem-se a opção “Export”, que ao ser selecionada irá solicitar um local e um nome para ser salvo o DataFrame no formato csv.

As Figuras 16 e 17 mostram respectivamente, como exportar o gráfico no formato png e no formato arquivo html.

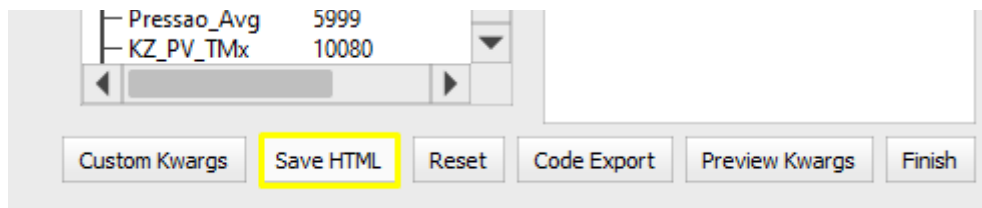
Figura 17 – Exportar gráfico como png



Fonte: O Autor.

Ao clicar no ícone da câmera será aberto uma janela para que seja salvo o arquivo png com o gráfico gerado.

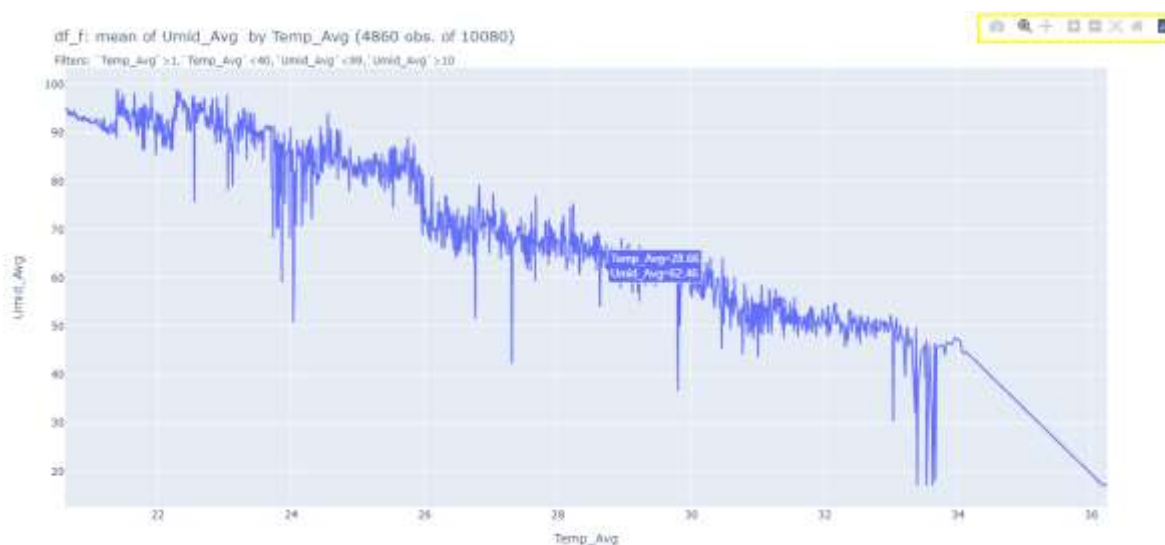
Figura 18 – Exportar gráfico como html



Fonte: O Autor.

A clicar em “Save HTML” será aberta uma janela para que seja salvo o arquivo HTML. A figura 18 mostra o arquivo html aberto no navegador, que pode auxiliar na análise do gráfico de forma mais detalhada.

Figura 19 – Gráfico aberto a partir de arquivo html



Fonte: O Autor

As ferramentas no canto superior direito auxiliam na visualização do gráfico da mesma maneira quando utilizado dentro da interface do programa e o cursor mostra o valor no ponto em que está apontado.

## 7 CONCLUSÃO

A mineração de dados ambientais é de fundamental importância nas pesquisas que visam estudar o potencial fotovoltaico e também em estudos de clima. A quantidade de dados gerados em uma unidade de coleta de dados ambientais é grande e de diversos tipos e formatos. Para que possa facilitar a mineração do grande volume de dados obtidos pelas plataformas de coleta de dados nos projetos do Grupo de Dispositivos Fotovoltaicos do INPE, está sendo desenvolvido um banco de dados com uma série de recursos computacionais que vise facilitar as atividades dos pesquisadores. Esse projeto teve como objetivo principal criar um ferramental que auxiliasse os pesquisadores na difícil tarefa de manipular esses dados.

O programa desenvolvido é capaz de ler arquivos de dados ambientais, filtrá-los de acordo com a data desejada e por alguns outros parâmetros e visualizar na tela os dados. Além disso é possível gerar diversos tipos de visualizações gráficas e exportar os dados filtrados com os seus respectivos gráficos.

O programa poderá ser melhorado em versões futuras com inserção de novas funcionalidades, conforme as necessidades dos usuários, com maiores velocidades de busca e visualizações. Além disso é necessário fazer uma melhor integração com o banco de dados que foi desenvolvido anteriormente para que o funcionamento seja mais automático.

Os projetos de Iniciação Científica com apoio econômico do CNPq são fundamentais na promoção do conhecimento científico, além de contribuir de forma precisa com os pesquisadores da área e no desenvolvimento do país. Os conhecimentos adquiridos neste trabalho vão além do que aqui foi apresentado, pois o contato intenso, apesar da pandemia, com os pesquisadores do INPE promove uma aprendizagem enriquecedora.

## 8 REFERÊNCIAS

Ichikura, F. Y., Vilela, W. A., Neves, G. M. **Desenvolvimento de uma metodologia de armazenamento e análise de dados espectrais utilizando ferramentas computacionais**. Relatório final de projeto de iniciação científica. INPE. SJC-SP, 2019.

Neves, G. M. **Influência do Espectro da Radiação Solar em Módulos Fotovoltaicos**, Dissertação de Mestrado - INPE. São José dos Campos - SP, 2016.

Vilela, W. A. Estudo, **Desenvolvimento e Caracterização de Radiômetros para Medidas da Radiação Solar**; Tese de Doutorado - INPE. São José dos Campos - SP, 2010.

Fornari C. I. ; **Unidade de Monitoramento Ambiental para o Estudo da Irradiância Solar e de Sensores**, Dissertação de Mestrado - INPE, São José dos Campos – SP, 2013.

PV EDUCATION. **Solar Energy**. Disponível em: <<https://www.pveducation.org/pvcdrom/introduction/solar-energy> > Acesso em: 20 mar. 2021

World Meteorological Organization. **Guide to Meteorological Instruments and Methods of Observation**. Geneva 2, Suíça: World Meteorological Organization, 2008.

PEREIRA, E. B.; MARTINS, F. R.; GONÇALVES, A. R.; COSTA, R. S.; LIMA, F. J. L.; RÜTHER, R.; ABREU, S. L.; TIEPOLO, G. M.; PEREIRA, S. V.; SOUZA, J. G. **Atlas brasileiro de energia solar**. São José dos Campos: INPE, 2017. 88 p. ISBN 978-85-17-00090-4. IBI: <8JMKD3MGP3W34P/3PERDJE>. Disponível em: <<http://urlib.net/rep/8JMKD3MGP3W34P/3PERDJE>>.

THOMAZINI, Daniel. Sensores Industriais: Fundamentos e **Aplicações**. 8ª edição, Érica, São Paulo, SP, 2009.

CORMEN, Thomas H. et al. **Algoritmos: teoria e prática**. Rio de Janeiro: Elsevier, 2012.

SINDA, **O QUE É UMA PCD?** Disponível em:

<http://sinda.crn.inpe.br/PCD/SITE/novo/site/index.php> 22 Ago. 2021