

**DESENVOLVIMENTO DE UM FRAMEWORK PARA A ANÁLISE
INICIAL DE DADOS PROVENIENTES DO SISTEMA NACIONAL DE
DADOS AMBIENTAIS (SINDA)**

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

Glenda Paola Barboza Lami (Fatec Cruzeiro - Prof. Waldomiro May, Bolsista
PIBIC/CNPq)
E-mail: glenda.lami@inpe.br

Eugenio Sper de Almeida (CPTEC / SESSS)
E-mail: eugenio.almeida@inpe.br

Julho de 2020

GLENDPAOLA BARBOZA LAMI



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

DESENVOLVIMENTO DE UM FRAMEWORK PARA A ANÁLISE INICIAL DE DADOS PROVENIENTES DO SISTEMA NACIONAL DE DADOS AMBIENTAIS (SINDA)

Relatório final de projeto de
iniciação científica
(PIBIC/INPE/CNPq)

Orientador: Prof. Dr. Eugênio Sper de Almeida

Cachoeira Paulista, 2020



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

AGRADECIMENTOS

Ao CNPq, bolsa PIBIC, processo 102773/2019-0



RESUMO

Este trabalho, iniciado em julho de 2019, tem como objetivo a continuidade ao projeto de Iniciação Científica em andamento desde fevereiro de 2019, para o desenvolvimento de um framework que permite a visualização dos dados provenientes da base de dados do SINDA de forma interativa, melhorando o sistema atual. Inicialmente o trabalho realizado no começo de 2019 tratou do estudo de bibliotecas estatísticas da linguagem Python, assim como o desenvolvimento do código responsável por extrair os dados do SINDA, agrupá-los por variável de estudo (como bateria, velocidade do vento, etc), aplicação de métodos estatísticos e exibição dos dados na forma de gráficos interativos. O trabalho atual trata do desenvolvimento de uma plataforma web, ou seja, a criação de páginas html, onde é possível selecionar uma plataforma de coleta de dados, escolher o período de análise, e as variáveis de estudos e gerar gráficos, tabelas. Com o código gerado no primeiro trabalho e a plataforma deste trabalho os usuários do SINDA poderão acessar e visualizar as tabelas, gráficos e dados estatísticos gerados de forma interativa. Para dar continuidade a este projeto de Iniciação Científica estão programadas as atividades de inclusão de outras estaticas e gráficos sobre o periodo total de dados coletados e não somente do periodo selecionado pelo usuario.



DEVELOPMENT OF A FRAMEWORK FOR THE INITIAL ANALYSIS OF DATA FROM THE NATIONAL ENVIRONMENTAL DATA SYSTEM (SINDA)

ABSTRACT

This work, started in July 2019, aims to continue the Scientific Initiation project in progress since February 2019, for the development of a framework that allows the visualization of data from the SINDA database in an interactive way, improving the current system. Initially the work carried out in early 2019 dealt with the study of statistical libraries of the Python language, as well as the development of the code responsible for extracting the data from SINDA, grouping them by study variable (such as battery, wind speed, etc.), application of statistical methods and display of data in the form of interactive graphs. The current work deals with the development of a web platform, that is, the creation of html pages, where it is possible to select a data collection platform, choose the analysis period, and study variables and generate graphs, tables. With the code generated in the first job and the platform of this job, SINDA users will be able to access and view the tables, graphs and statistical data generated interactively. In order to continue this Scientific Initiation project, activities to include other statistics and graphics on the total period of data collected and not only the period selected by the user are scheduled.



LISTA DE FIGURAS

Figura 1: Esquema de comunicação entre elementos do projeto

Figura 2: Interface de resultado

Figura 3 – Distribuição das PCD's sobre o território brasileiro



LISTA DE TABELAS

Tabela 1 - Variáveis de estudo do SINDA



LISTA DE SÍMBOLOS E ABREVIATURAS

- INPE - Instituto Nacional de Pesquisa Espaciais
- CPTEC - Centro de Previsão de Tempo e Estudos Climáticos
- SINDA - Sistema Integrado de Dados Ambientais
- PCD - Plataforma de Coleta de Dados
- DCS - transportador de coleta de Dados
- ANA - Agência Nacional de Água
- INMET - Instituto Nacional de Meteorologia
- CETESB - Companhia Ambiental do Estado de São Paulo
- ID - Número de identificação



SUMÁRIO

1 INTRODUÇÃO	9
2 FUNDAMENTAÇÃO TEÓRICA	11
3 MATERIAIS E MÉTODOS UTILIZADOS	12
3.1. Base de dados	13
4 RESULTADOS	14
5 CONCLUSÕES	15
REFERÊNCIAS BIBLIOGRÁFICAS	17

1 INTRODUÇÃO

Os dados ambientais possuem importância fundamental em atividades de previsão meteorológica, monitoramento de recursos hídricos e de qualidade das águas, agricultura de precisão, prevenção de desastres naturais e diversas outras atividades (AEB, 2018). Os resultados das medições feitas pelos sensores dos sistemas de observação revelam características sobre o ambiente (Gunther, 1997).

O SINDA é parte integrante do Sistema Brasileiro de Coleta de Dados Ambientais. possui a finalidade de processar, armazenar e disseminar os dados das PCD's ambientais. (Santos et al., 2013)

O acesso aos dados da PCD (<http://sinda.crn.inpe.br>) é realizado através do número da estação/estado/ nome da estação. Em seguida o usuário fornece o intervalo de datas que deseja selecionar, a partir de informações disponibilizadas pelo SINDA. O sistema permite apenas a seleção de dados de um período de no máximo de um ano, fornecendo os dados solicitados em formato EXCELL.

Silva e Silva (2017) desenvolveram um *framework* para acesso e visualização das informações coletadas pelas PCD's, sem limitação de período de tempo. Apesar de não analisarem os dados, os gráficos gerados pelo *framework* permitiram verificar a existência de informações faltantes na base de dados do SINDA.

A obtenção da estatística descritiva das variáveis ambientais armazenadas no SINDA, antes do início da análise dos dados, pode ser bastante útil e permite uma visão global do conjunto de dados das PCD's. Outra forma de verificar a disponibilidade dos dados seria através da utilização de ferramentas gráficas (*histogram, box plot, etc.*).

Segundo Silvestre (2007), estatística descritiva é o ramo da estatística que visa sumarizar e descrever qualquer conjunto de dados, como medidas de posição e medidas de variabilidade ou dispersão. Medidas de posição incluem média, mediana e moda. Medidas de variabilidades incluem desvio padrão, variância, o valor de máximo e mínimo, obliquidade e curtose (Bussab, Morettin, 2017)

Desta forma, o objetivo deste trabalho é dar continuidade ao *framework* desenvolvido na etapa anterior e torná-lo um portal web onde é possível com informações de estatística descritiva e gráficos estatísticos referentes às diferentes variáveis ambientais coletadas pelas PCD's e armazenadas no SINDA.



A partir da informação fornecida pelo framework um pesquisador pode determinar *a priori* quais dados ambientais são mais adequados à sua pesquisa e os responsáveis pelas PCD's poderiam monitorar o funcionamento dos sensores e componentes das PCD's.

2 FUNDAMENTAÇÃO TEÓRICA

Nos dias atuais a dependência de informação tem aumentado cada vez mais, portanto a estatística se tornou uma importante ferramenta para a tomada de decisão. (SANTOS, 2007)

O objetivo principal da estatística descritiva é sintetizar uma série de valores de mesma natureza, permitindo uma visão global da variação desses valores, além de organizar e descrever os dados por meio de tabelas, de gráficos e de medidas descritivas. (GUEDES, 2015)

Na estatística existem muitos métodos e técnicas que possibilitam a análise de dados e servem para recolher, organizar, sintetizar e descrever os dados, além dos que são a base na Teoria das Probabilidades, permitindo a análise e a interpretação dos dados, assim como efectuar inferências sobre uma população com base no estudo de uma amostra. (SANTOS, 2007)

Com os métodos estatísticos não é possível adivinhar com precisão acontecimentos futuros, no entanto com eles é possível medir as variações ao longo do tempo e fazer previsões prováveis. (REIS, 1999)

Uma importante ferramenta da estatística é o gráfico, que é recursos visual utilizado nos meios de comunicação social, técnica e científica, devido à facilidade de interpretação e a eficiência com que resume informações dos mesmos. (GUEDES, 2015)

3 MATERIAIS E MÉTODOS UTILIZADOS

Para esta etapa do projeto foi utilizado no Backend o código desenvolvido na etapa anterior em linguagem Python, juntamente com as bibliotecas Pandas e Numpy para os cálculos científicos e estatísticos e a Bokeh para apresentação dos dados em gráficos dinâmicos.

Como a linguagem python não é nativa da web foi preciso utilizar o framework flask para que o navegador pudesse interpretar a linguagem e possibilitar a comunicação com a linguagem javascript utilizada para manipulação da DOM, sigla que em inglês para *Document Object Model* que são os objetos de uma página html, como botões, *link*, formulários, entre outros que permitem com que os usuários interajam com os elementos em uma página web.

Do javascript também houve a utilização da biblioteca leaflet para gerar o mapa de PCD's em substituição à biblioteca folium do python pois esta apresentava limitações que não possibilitam uma maior customização do mapa, mesmo ela sendo baseada na leaflet.

A figura 1 demonstra como funciona a comunicação entre os arquivos do projeto e funções existentes neles e principalmente os papéis de cada uma das linguagens utilizadas.

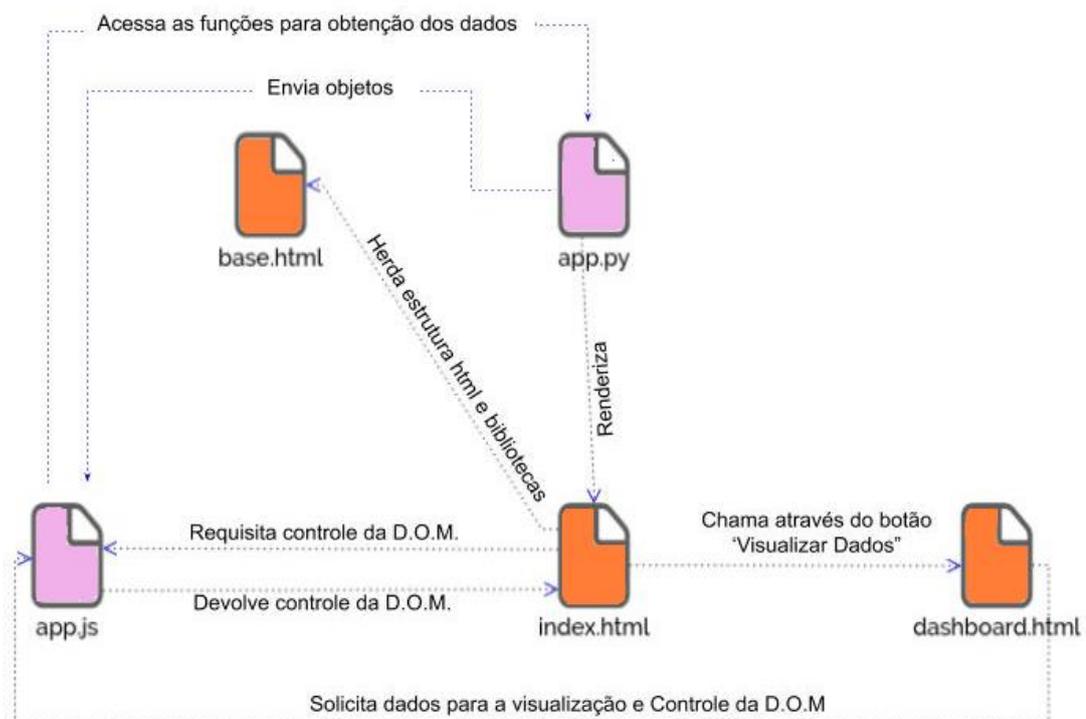


Figura 1: Esquema de comunicação entre elementos do projeto
Fonte: os autores



3.1. Base de dados

As bases de dados meteorológicos que foram utilizadas no projeto são as do INPE, onde o CPTEC é responsável por receber as informações de todas as PCDs do INPE e estão armazenadas no SINDA.

Os dados armazenados no SINDA são provenientes da coleta pelos sensores de mil e cinco PCD's e a Tabela 1 exibe quais são as variáveis de estudo do SINDA.

Tabela 1 - Variáveis de estudo do SINDA

Parâmetro	Sigla	Unidade	Descrição
Temp. do Ar	TempAr	°C	valor instantâneo a cada 3 H
Temp. Máx. do Ar últ. 24 H	TempMax	°C	valor a cada 3 H com a Máx. das últ. 24 H, amostragem a cada 1 minuto.
Temp. Mín. do Ar últ. 24 H	TempMin	°C	valor a cada 3 H com a Mín. das últ. 24 H, amostragem a cada 1 minuto.
Umidade Relativa do Ar	UmidRel	%	valor instantâneo a cada 3 H
Pressão Barométrica	PressaoAtm	mB	valor instantâneo a cada 3 H
Veloc. do Vento	VelVento	m/s	valor a cada 3 H, calculado da média de 200 amostras com 3 seg de intervalo, 10 min antes de cada 3 H
Dir. do Vento	DirVento	° NV	valor a cada 3 H, calculado da média de 200 amostras com 3 seg de intervalo, 10 min antes de cada 3 H
Veloc. Máx. do Vento (Rajada)	VelVentoMax	m/s	valor máximo(rajada) cada 3 H, amostras cada 3 seg
Dir. do Vento na Veloc. Máx.	DirVelVentoMax	° NV	valor(Dir. da rajada) cada 3 H, amostras cada 3 seg
Radiação Solar Global	RadSolAcum	MJ/m ²	valor acumulado a cada 3 H, integração de 1080 amostras de 10s de intervalo
Radiação Solar Líquida	RadSolLiq	W/m ²	valor instantâneo a cada 3 H
Precipitação Acumulada	Pluvio	mm	valor acumulado mensal a cada 3 H (zera o acum. automat. todo dia 01 de cada mês)
Temp. do Solo 100mm, 200mm, 400mm (*)	TempSolo100 TempSolo200 TempSolo400	°C	valor instantâneo a cada 12 H amostrado às 06 e 18 Hs GMT
Conteúdo Água no Solo 100mm, 200mm, 400mm (*)	ContAguaSolo100 ContAguaSolo200 ContAguaSolo400	(m ³ /m ³)	valor instantâneo a cada 12 H amostrado às 06 e 18 Hs GMT
Fluxo de Calor no Solo	FluxCalSolo	W/m ²	valor instantâneo a cada 3 H
Bateria	Bateria	Volts	valor instantâneo a cada 3 H
Corrente Painel Solar	CorrPSol	Lógico	valor instantâneo a cada 3 H
Umidade Interna	UmidInt	%	valor instantâneo a cada 3 H

Nota: O horário de Coleta das PCDs é sincronizado com a Hora Universal GMT (Greenwich Mean Time) = Hora de Brasília + 3 horas (horário normal) ou Hora de Brasília + 2 horas (horário de verão).

Fonte - Centro de Pesquisa Tecnológica e Estudos Climáticos (CPTEC)

4 RESULTADOS

Como resultado obtido tem-se a interface exibida na figura 2, onde o usuário escolhe uma única PCD e o período que deseja visualizar os dados, que pode ser de um mês até todo o período de coleta desta PCD.

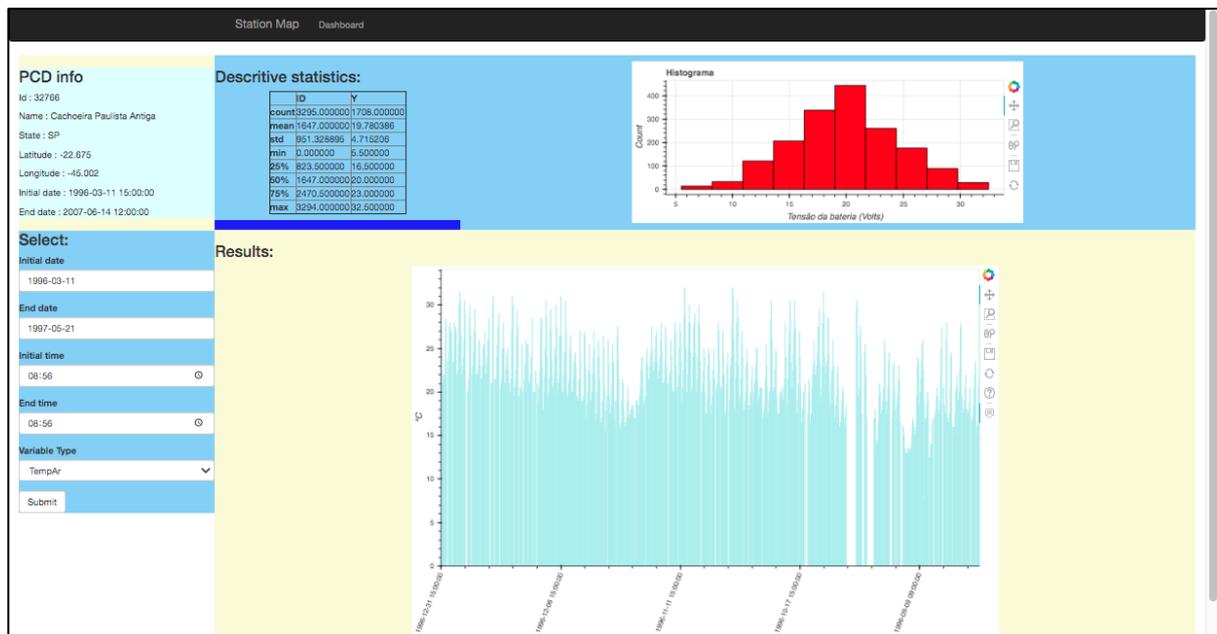


Figura 2: Interface de resultado

Fonte: os autores

O usuário decide a forma de escolha da PCD que pode ser através de localização no mapa ou escrevendo no campo de pesquisa o código ou nome dela. Se o usuário optar por utilizar a localização ele precisa clicar no link na barra superior “Station Map” onde aparecerá o mapa do Brasil como na figura 2 com a localização de todas as PCD, para selecionar basta clicar na PCD pretendida.



Figura 3 – Distribuição das PCD's sobre o território brasileiro
Fonte: os autores

Após a escolha o usuário informa o período e a variável de estudo que deseja e após confirmar sua escolha, é apresentado em tela os dados estatísticos em forma de tabela, um gráfico de histograma da variável e um outro com os valores reais, ambos dinâmicos.

5 CONCLUSÕES

Este projeto foi desafiador desde o começo exigindo muito estudo para desenvolver as funcionalidades do framework, assim como compreender a gama de possibilidade e também as limitações das bibliotecas utilizadas para se chegar ao resultado esperado.

Mesmo sendo desafiador foi possível chegar a uma solução que atende ao esperado, um portal web onde os usuários do Sinda podem visualizar os dados da PCD desejada e também ter acesso a uma análise estatística inicial.

Acredita-se que através deste portal possa auxiliar os pesquisadores na escolha de quais variáveis seriam mais interessantes às suas pesquisas.

REFERÊNCIAS BIBLIOGRÁFICAS

AEB Dados ambientais. Disponível em <http://www.aeb.gov.br/servicos/dados-ambientais/>, acessado em junho/2018.

Bussab, W.O.; Morettin, P.A. Estatística Básica, v., 2017.

CPTEC. Informações. Disponível em: <<http://www.cptec.inpe.br/sobreocptec/pt>> Acesso em: 12 de Fevereiro de 2019.

GUEDES, Terezinha Aparecida et al. Estatística descritiva. Projeto de ensino aprender fazendo estatística, p. 1-49, 2005.

Gunther, O. Environmental information system. ACM SIGMOD Record 26(1) 3-4, 1997.

Bussab, W.O.; Morettin, P.A. Estatística Básica, v., 2017.

CPTEC. Informações. Disponível em: <<http://www.cptec.inpe.br/sobreocptec/pt>> Acesso em: 12 de Fevereiro de 2019.

GUEDES, Terezinha Aparecida et al. Estatística descritiva. Projeto de ensino aprender fazendo estatística, p. 1-49, 2005.

Gunther, O. Environmental information system. ACM SIGMOD Recohttps://jupyter-notebook.readthedocs.io/en/stable/notebook.html> Acesso em 20 de junho de 2019.

REIS, Elizabeth et al. Estatística aplicada. Lisboa: Edições Sílabo, 1999.

Sanner, Michel F. et al. Python: a programming language for software integration and development. J Mol Graph Model, v. 17, n. 1, p. 57-61, 1999.

SANTOS, Carla. Estatística descritiva. Manual de auto-aprendizagem, v. 2, 2007.

Santos, M. A. F.; Francisco, M. F.M.; Yamaguti, W. O Sistema Nacional de Dados Ambientais e a Coleta de Dados por Satélite. Anais XVI Simpósio Brasileiro de Sensoriamento Remoto, 2013.

Silva, J.V.O.; Silva, W.C.D. Proposta de Sistema de Apoio a Análise de Dados Meteorológicos - Fatec Cruzeiro, 2010

Silvestre, A.L. Análise de dados e estatística descritiva. Escolar editora, 2007.

SINDA. Informações. Disponível em:

<http://sinda.crn.inpe.br/PCD/SITE/novo/site/sobre.php>> Acesso em: 15 Fevereiro de 2019.

TOSI, Sandro. Matplotlib for Python Developers. Reino Unido: Packt Publishing, 2009.