

Escola de Artes, Ciências e Humanidades - EACH-USP - USP Leste

ACH2147 - Desenvolvimento de Sistemas de Informação Distribuídos

Relatório - Exercício Programa II

Ian Lourenço - 10430937
Mark Poll Herrmann - 11208291
Caio Fernandes Barbosa - 7557331
Guilherme Rodrigues Pisni - 11270851
Guilherme Kenzo Silva Oshiro - 11314988

Relatório - EP 2

Mark Poll Herrmann - N°USP 11208291

Guilherme Rodrigues Pisni - N°USP 11270851

Ian Lourenço - N°USP 10430937

Caio Fernandes Barbosa - N°USP 7557331

Guilherme Kenzo Silva Oshiro - 11314988

Implementação

Nosso programa foi desenvolvido em python utilizando principalmente a biblioteca “PySpark” que usa do software desenvolvido pela Apache, mas dentre outras bibliotecas que nos ajudam com cálculos, gráficos e outras coisas temos as bibliotecas: “os”(para auxiliar na seleção do diretório), “pandas” e “matplotlib” (para auxiliar na plotagem de gráficos) e “tkinter” (para criação de interface gráfica).

Consulta de dados

A consulta de dados é feita localmente na máquina, os arquivos correspondentes à cada ano devem estar em uma subpasta identificando o ano que por sua vez deverão estar dentro de uma pasta chamada “Data” no mesmo diretório de onde está armazenado o programa, assim, se quisermos consultar os arquivos do ano “1969” deveremos percorrer o seguinte caminho: ***LOCAL ATÉ O DIRETÓRIO DO PROGRAMA*/Data/1969**

A função que recebe os parâmetros da interface para consultar os dados chama-se “*processa_dados()*” e ela recebe o local da pasta “Data”, as datas que delimitam o intervalo de tempo, a coluna dos dados desejados e o tipo de agrupamento, podendo ser por dias da semana em inglês (Monday, Tuesday, etc.), por meses do ano em inglês (January, February,, etc.), por ano, no formato Mês/Ano ou Dia/Mês/Ano.

Ao entrar na função as datas serão ajustadas para podermos manipulá-las corretamente e vamos buscar o delimitador (variável que serve para indicar quais dados são considerados inválidos, ou como está na documentação dos dados, *missing*) baseado na coluna especificada, criamos um dataframe vazio que irá acumular os dados encontrados e partimos para o laço que fará a consulta.

Entrando no laço que seguirá até o ano determinado pelo usuário, teremos uma variável que será carregada com um dataframe contendo todos os dados do diretório (ou ano) atual, na linha seguinte, esses dados serão filtrados de acordo com o intervalo passado pelo usuário e por fim serão adicionados esses dados ao dataframe “dados” filtrando aquelas entradas com valores inválidos.

Saindo do laço nós teremos um dataframe carregado com todas as informações válidas para conseguirmos realizar as operações estatísticas que serão feitas nas próximas linhas entrando no *if/else* que se adeque ao tipo de agrupamento pedido pelo usuário.

Para agrupar os dados formatamos a coluna “DATE” de acordo com a vontade do usuário utilizando a função “date_format” do Spark e passando os parâmetros adequados para a formatação desejada, então nós iremos agrupar os dados (usando group by), usar a

função “agg” para especificar quais colunas iremos agrupar utilizando a coluna de datas como base.

Dentro dos parâmetros que passamos para a função “agg” devemos colocar as funções estatísticas, no nosso caso, **média** (calculada utilizando a função “mean” e passando a coluna como parâmetro), **desvio padrão** (calculado utilizando a função “stddev” e passando a coluna como parâmetro) e a **variância** (calculada utilizando a função “variance” e passando a coluna como parâmetro). Por fim será mostrada uma tabela ao usuário contendo as informações que pediu.

Predição dos dados

Então passaremos para a função que irá fazer uma predição de dados baseado no método dos quadrados mínimos, a função “predicao()” irá receber o caminho até a pasta “Data”, as datas iniciais e finais para delimitar um intervalo e por fim a coluna que servirá de base para a predição (coluna_um) e a coluna a ser predita (coluna_dois).

Essa função tem o mesmo início da função “processa_dados()”, basicamente ajusta as datas, pega os delimitadores e instancia um dataframe vazio para armazenar os dados válidos para então entrar em um laço para pegar os dados válidos e armazenar nesse dataframe.

A diferença entre essa função e a anterior está na parte que vem a seguir, baseado no dataframe de dados que pegamos durante o laço, iremos pegar os valores das colunas e calcular o desvio padrão agrupando os valores por dias (utilizando novamente a função “agg” e “stddev” do Spark).

Em seguida armazenaremos as médias de cada coluna em variáveis diferentes para podermos prosseguir com o cálculo da função, passamos os dados necessários para a função “calcula_b()” que como o nome diz calcula a variável “b” da função que desejamos, ao retornar calculamos a variável “a”, selecionamos os valores máximos e mínimos da coluna em que nos baseamos para a predição para obtermos y0 e y1.

Interface e interação esperada

A interação do usuário com o sistema é feita via interface gráfica. Ao executar o arquivo “interface.py” a aplicação será iniciada e aguardará os comandos do usuário como ilustrado abaixo.

The screenshot displays a graphical user interface for a weather application titled "SENHOR DO TEMPO". The interface is divided into two main sections: "Histórico de dados de tempo" (Historical weather data) and "Predição do tempo" (Weather prediction). Both sections feature date input fields for "Data de início" (Start date) and "Data de término" (End date), a "Método de Agrupamento" (Grouping method) dropdown menu, and two columns of radio button options for selecting weather parameters. A pink "clica" (click) button is positioned below the options in each section.

Histórico de dados de tempo

Data de início: [][][] Data de término: [][][]

Método de Agrupamento []

- ☐ Mean temperature
- ☐ Mean dew point
- ☐ Mean sea level pressure
- ☐ Mean station pressure
- ☐ Mean visibility
- ☐ Mean wind speed
- ☐ Maximum sustained wind speed
- ☐ Maximum wind gust
- ☐ Maximum temperature
- ☐ Minimum temperature
- ☐ Precipitation amount
- ☐ Snow depth

clica

Predição do tempo

Data de início: [][][] Data de término: [][][]

- ☐ Mean temperature
- ☐ Mean dew point
- ☐ Mean sea level pressure
- ☐ Mean station pressure
- ☐ Mean visibility
- ☐ Mean wind speed
- ☐ Maximum sustained wind speed
- ☐ Maximum wind gust
- ☐ Maximum temperature
- ☐ Minimum temperature
- ☐ Precipitation amount
- ☐ Snow depth

clica

Na tela inicial, o usuário encontra os campos de entrada para os dois tipos de execução do programa: “Histórico de dados de tempo” e “Predição do tempo” ambas execuções devem ter seus campos preenchidos de maneira integral, ou seja sem nenhum campo vazio, para a execução ser iniciada.

Para a função “Histórico de dados de tempo” devem ser fornecidos ao sistema o intervalo de tempo desejado utilizando os campos “data de início” e “data de término”, o método de

agrupamento de informações utilizando o menu dropdown e o tipo de informações que serão analisadas utilizando os radiobuttons.

Exemplo:

SENHOR DO TEMPO

Histórico de dados de tempo

Data de início: 01 01 1930

Data de término: 31 12 1934

Método de Agrupamento

- Método de Agrupamento
- ANO**
- MES
- D. SEMANA
- MES/ANO
- DIA/MES/ANO

☒ Mean temperature

☐ Mean dew point

☐ Mean sea level pressure

☐ Mean station pressure

☐ Mean visibility

☐ Mean wind speed

☐ Maximum sustained wind speed

☐ Maximum wind gust

☐ Maximum temperature

☐ Minimum temperature

☐ Precipitation amount

☐ Snow depth

clica

Para a função de “Predição de tempo” os dados requisitados são o intervalo de tempo desejado utilizando os campos “data de início” e “data de término”, e o tipo de informações a serem cruzadas para fazer a predição do tempo, utilizando os radiobuttons. Note que diferentemente da função anterior, é necessário que sejam escolhidos dois tipos de informações nos radiobuttons, podendo esses serem iguais ou não.

Exemplo:

Predição do tempo

Data de início: 01 01 1930

Data de término: 31 12 1934

☐ Mean temperature

☒ Mean dew point

☐ Mean sea level pressure

☐ Mean station pressure

☐ Mean visibility

☐ Mean wind speed

☐ Maximum sustained wind speed

☐ Maximum wind gust

☐ Maximum temperature

☐ Minimum temperature

☐ Precipitation amount

☐ Snow depth

☐ Mean temperature

☐ Mean dew point

☐ Mean sea level pressure

☒ Mean station pressure

☐ Mean visibility

☐ Mean wind speed

☐ Maximum sustained wind speed

☐ Maximum wind gust

☐ Maximum temperature

☐ Minimum temperature

☐ Precipitation amount

☐ Snow depth

clica

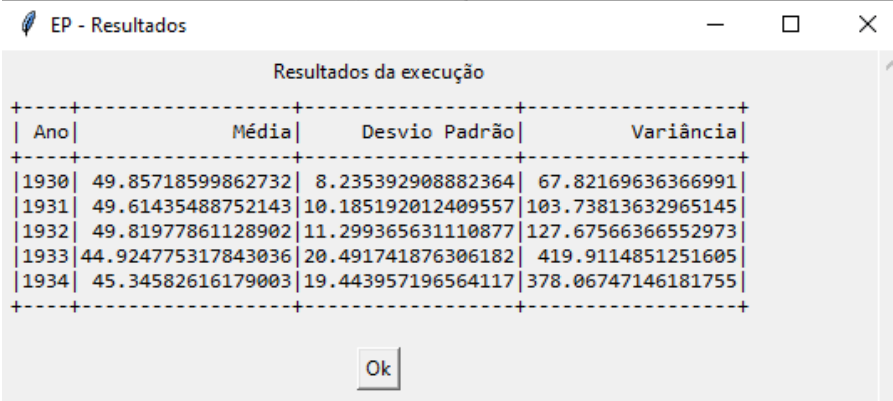
Após preencher os dados necessários na função escolhida, o programa será executado ao clicar no botão “clica” de sua referente função.

O processo de execução pode levar algum tempo dependendo do volume de dados escolhido, o andamento da execução pode ser monitorado através da barra de execução que é apresentada ao usuário através do terminal de seu sistema operacional como ilustrado na imagem abaixo.

```
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
21/07/31 18:10:40 WARN ProcsMetricsGetter: Exception when trying to compute pagesize, as a result reporting of ProcessTree metrics is stopped
[Stage 33:=====> (355 + 17) / 653]
```

O resultado, independentemente do tipo de execução, será exibido em uma nova janela devolvendo as informações solicitadas pelo usuário.

Exemplo: “Histórico de dados de tempo”



A janela intitulada "EP - Resultados" exibe um resumo da execução em uma tabela com 4 colunas: Ano, Média, Desvio Padrão e Variância. A tabela contém 5 linhas de dados para os anos 1930, 1931, 1932, 1933 e 1934. Abaixo da tabela, há um botão "Ok".

Ano	Média	Desvio Padrão	Variância
1930	49.85718599862732	8.235392908882364	67.82169636366991
1931	49.61435488752143	10.185192012409557	103.73813632965145
1932	49.81977861128902	11.299365631110877	127.67566366552973
1933	44.924775317843036	20.491741876306182	419.9114851251605
1934	45.34582616179003	19.443957196564117	378.06747146181755

Mais exemplos:



O formulário "Histórico de dados de tempo" permite selecionar um período de tempo e um tipo de dado. A data de início é 01/01/1930 e a data de término é 31/12/1935. Há uma opção para selecionar o formato da data (MES/ANO). Abaixo, há duas colunas de opções de dados, cada uma com um botão "clica" para executar a consulta.

Histórico de dados de tempo

Data de inicio: 01/01/1930 Data de término: 31/12/1935

MES/ANO

☐ Mean temperature ☐ Maximum sustained wind speed

☒ Mean dew point ☐ Maximum wind gust

☐ Mean sea level pressure ☐ Maximum temperature

☐ Mean station pressure ☐ Minimum temperature

☐ Mean visibility ☐ Precipitation amount

☐ Mean wind speed ☐ Snow depth

clica

Resultados da execução			
Mes/Ano	Média	Desvio Padrão	Variância
01/1930	40.690147783251234	4.792603416830868	22.969047511018914
02/1930	35.30688405797102	5.465077022449563	29.86706686130618
03/1930	38.45183333333333	6.110113401718387	37.333485781858634
04/1930	42.0567084078712	4.596334901021521	21.126294522348516
05/1930	46.22357019064125	4.69000108190569	21.996110148276543
06/1930	52.0629963898917	4.004439540377241	16.03553603253669
07/1930	53.473851590106	3.185301698255438	10.146146908908976
08/1930	54.69841827768013	3.8519860837656856	14.837796789524505
09/1930	53.09877192982455	4.36449751512037	19.04883855949189
10/1930	47.62512479201332	5.581005079081575	31.147617692734343
11/1930	42.08830313014827	6.469844616380343	41.858889360105714
12/1930	41.353037766830866	5.384877228254508	28.99690276337395
01/1931	36.456258234519105	7.35990060423395	54.16813690420326
02/1931	35.477999999999994	7.98693149326521	63.791074678111634
03/1931	34.826585695006756	9.235973021349315	85.30319765109238
04/1931	40.086345903771125	6.134780311367022	37.63552946873646
05/1931	47.512280701754385	6.603879089685434	43.611219031184525
06/1931	51.24713064713064	5.326565423794382	28.372299213961824



SENHOR DO TEMPO



Histórico de dados de tempo

Data de início:

01 01 1930

Data de término:

31 12 1931

DIA/MES/ANO

- ☐ Mean temperature
- ☒ Mean dew point
- ☐ Mean sea level pressure
- ☐ Mean station pressure
- ☐ Mean visibility
- ☐ Mean wind speed

- ☐ Maximum sustained wind speed
- ☐ Maximum wind gust
- ☐ Maximum temperature
- ☐ Minimum temperature
- ☐ Precipitation amount
- ☐ Snow depth

clica

Resultados da execução			
Dia/Mes/Ano	Média	Desvio Padrão	Variância
01/01/1930	43.09499999999999	5.279800135770692	27.87628947368422
02/01/1930	43.230000000000004	5.713151955195827	32.640105263157906
03/01/1930	45.300000000000004	5.613236237023308	31.50842105263159
04/01/1930	41.636842105263156	4.952912782317064	24.531345029239763
05/01/1930	40.16	2.4843510218968663	6.172000000000004
06/01/1930	39.575	2.9314761864253565	8.593552631578952
07/01/1930	44.95	2.6069744512506614	6.796315789473688
08/01/1930	37.485	2.7906375427018175	7.787657894736839
09/01/1930	38.130000000000001	2.97659289509968	8.860105263157893
10/01/1930	38.785000000000004	2.857727289850783	8.166605263157901
11/01/1930	35.58888888888889	3.646414659536276	13.296339869281054
12/01/1930	36.71052631578947	3.600446690935585	12.963216374269003
13/01/1930	40.365	4.059203318256843	16.477131578947365
14/01/1930	43.468421052631584	5.706921005973452	32.56894736842104
15/01/1930	40.59473684210526	5.1251529716152024	26.26719298245614
16/01/1930	41.07777777777778	5.988965888891056	35.86771241830064
17/01/1930	43.82	3.6506091777155567	13.326947368421054
18/01/1930	46.136842105263156	2.1273902965098377	4.525789473684216
19/01/1930	45.575	2.5042858000677666	6.271447368421053

SENHOR DO TEMPO

Histórico de dados de tempo

Data de início: 01/01/1930

Data de término: 31/12/1931

D. SEMANA

- ☐ Mean temperature
- ☐ Mean dew point
- ☐ Mean sea level pressure
- ☐ Mean station pressure
- ☐ Mean visibility
- ☒ Mean wind speed

- ☐ Maximum sustained wind speed
- ☐ Maximum wind gust
- ☐ Maximum temperature
- ☐ Minimum temperature
- ☐ Precipitation amount
- ☐ Snow depth

clica

EP - Resultados

Resultados da execução

Dia da semana	Média	Desvio Padrão	Variância
Wednesday	8.92101378323416	5.910521391382759	34.93426311799318
Tuesday	8.721079920005925	5.825050631824243	33.931214863316015
Friday	9.005964095245247	6.035971065083932	36.43294669853046
Thursday	9.098275472550643	6.168237509785854	38.04715397712919
Saturday	9.132365747460087	6.224539994127074	38.74489813848748
Monday	8.972157423432913	5.917949272231205	35.022123588701845
Sunday	9.149074730293725	6.121800398474719	37.47644011876523

Ok

SENHOR DO TEMPO

Histórico de dados de tempo

Data de início:

01

01

1930

Data de término:

16

01

1930

DIA/MES/ANO

Mean temperature

Mean dew point

Mean sea level pressure

Mean station pressure

Mean visibility

Mean wind speed

Maximum sustained wind speed

Maximum wind gust

Maximum temperature

Minimum temperature

Precipitation amount

Snow depth

clica

Resultados da execução

Dia/Mes/Ano	Média	Desvio Padrão	Variância
01/01/1930	38.94285714285714	6.102423406202771	37.23957142857142
02/01/1930	44.10952380952381	5.99282110211082	35.913904761904746
03/01/1930	43.957142857142856	5.899285671042844	34.80157142857142
04/01/1930	40.834999999999994	5.798754403454153	33.625552631578934
05/01/1930	41.804761904761904	2.3559448615101726	5.550476190476187
06/01/1930	39.825	3.351649240535016	11.233552631578949
07/01/1930	42.60952380952381	3.6644105613187996	13.42790476190476
08/01/1930	38.74285714285714	2.7737287950647644	7.693571428571429
09/01/1930	37.480000000000004	3.778693731408974	14.278526315789476
10/01/1930	38.68571428571429	4.026075721380029	16.20928571428572
11/01/1930	36.315	3.592429906286127	12.905552631578953
12/01/1930	34.60952380952381	3.3284988751544966	11.07890476190475
13/01/1930	39.74285714285715	4.677453519659114	21.87857142857143
14/01/1930	42.81	5.92336142121311	35.08621052631579
15/01/1930	39.77894736842105	6.433987786000946	41.39619883040936
16/01/1930	38.268421052631574	6.5743654219821375	43.22228070175437

Ok

Histórico de dados de tempo

Data de início: 01 01 1930

Data de término: 31 12 1931

MES —

- ☐ Mean temperature
- ☐ Mean dew point
- ☐ Mean sea level pressure
- ☐ Mean station pressure
- ☐ Mean visibility
- ☐ Mean wind speed

- ☐ Maximum sustained wind speed
- ☐ Maximum wind gust
- ☐ Maximum temperature
- ☐ Minimum temperature
- ☒ Precipitation amount
- ☐ Snow depth

clica

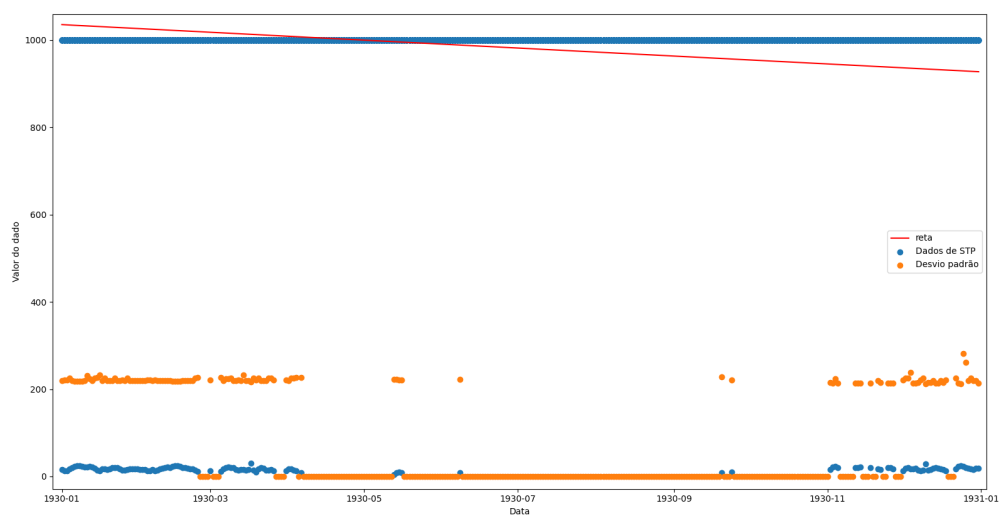
Resultados da execução			
Mês	Média	Desvio Padrão	Variância
July	18.867920566354673	39.08556233617389	1527.6811831349353
November	27.20444012716352	44.465359913374755	1977.1682322259546
February	23.06352335586967	42.0859226928163	1771.2248889057096
January	26.485865551067477	44.09049212262066	1943.9714956148741
March	24.59999711441349	43.048170107906344	1853.1449496392415
October	22.730417649040064	41.85808728460722	1752.099471125797
May	23.32581166549541	42.26604104462239	1786.4182255857047
August	18.651419009370795	38.909583780697794	1513.9557099871406
April	28.03252572772711	44.884035260466	2014.5766212627555
June	19.773677069199447	39.795151470293696	1583.6540805436186
December	25.541844380403443	43.577282489409555	1898.9795491618008
September	19.139662106863078	39.29301836463319	1543.9412922034014

Ok

Exemplo: “Predição do tempo”

Figure 1

— □ ×



Mais exemplos:

SENHOR DO TEMPO

clica

Predição do tempo

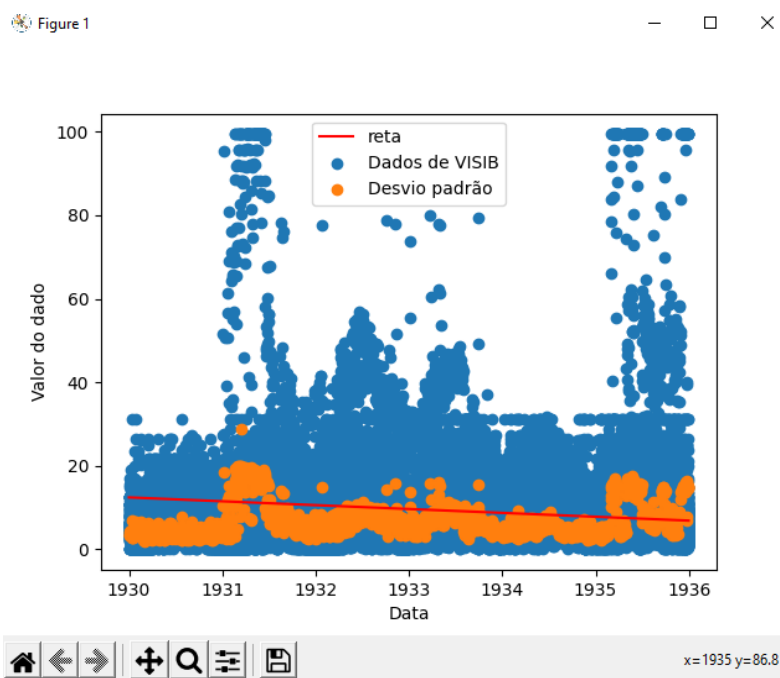
Data de início: 01 01 1930

Data de término: 31 12 1935

☐ Mean temperature
☒ Mean dew point
☐ Mean sea level pressure
☐ Mean station pressure
☐ Mean visibility
☐ Mean wind speed
☐ Maximum sustained wind speed
☐ Maximum wind gust
☐ Maximum temperature
☐ Minimum temperature
☐ Precipitation amount
☐ Snow depth

☐ Mean temperature
☐ Mean dew point
☐ Mean sea level pressure
☐ Mean station pressure
☒ Mean visibility
☐ Mean wind speed
☐ Maximum sustained wind speed
☐ Maximum wind gust
☐ Maximum temperature
☐ Minimum temperature
☐ Precipitation amount
☐ Snow depth

clica



Predição do tempo

Data de início:

Data de término:

☐ Mean temperature

☒ Mean dew point

☐ Mean sea level pressure

☐ Mean station pressure

☐ Mean visibility

☐ Mean wind speed

☐ Maximum sustained wind speed

☐ Maximum wind gust

☐ Maximum temperature

☐ Minimum temperature

☐ Precipitation amount

☐ Snow depth

☐ Mean temperature

☐ Mean dew point

☐ Mean sea level pressure

☐ Mean station pressure

☐ Mean visibility

☐ Mean wind speed

☐ Maximum sustained wind speed

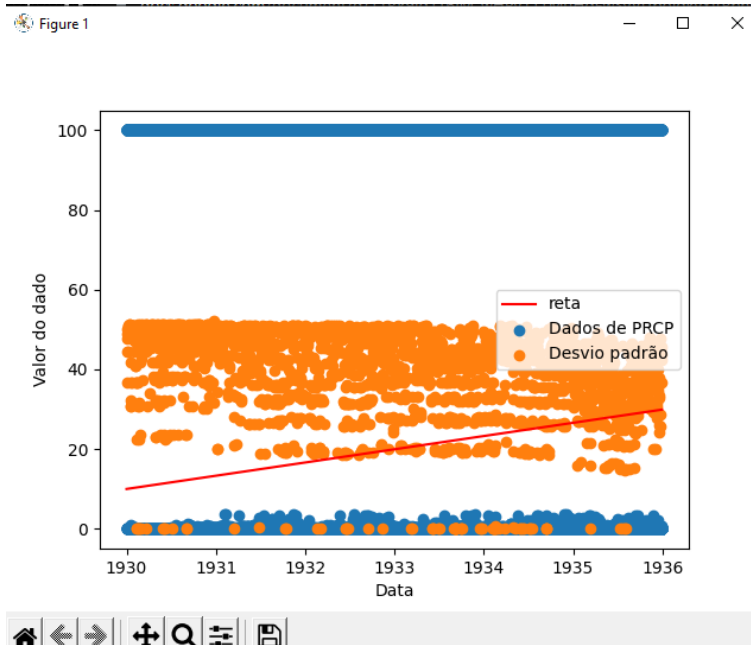
☐ Maximum wind gust

☐ Maximum temperature

☐ Minimum temperature

☒ Precipitation amount

☐ Snow depth



Predição do tempo

Data de início:

Data de término:

☐ Mean temperature

☐ Mean dew point

☐ Mean sea level pressure

☒ Mean station pressure

☐ Mean visibility

☐ Mean wind speed

☐ Maximum sustained wind speed

☐ Maximum wind gust

☐ Maximum temperature

☐ Minimum temperature

☐ Precipitation amount

☐ Snow depth

☐ Mean temperature

☐ Mean dew point

☐ Mean sea level pressure

☐ Mean station pressure

☐ Mean visibility

☐ Mean wind speed

☐ Maximum sustained wind speed

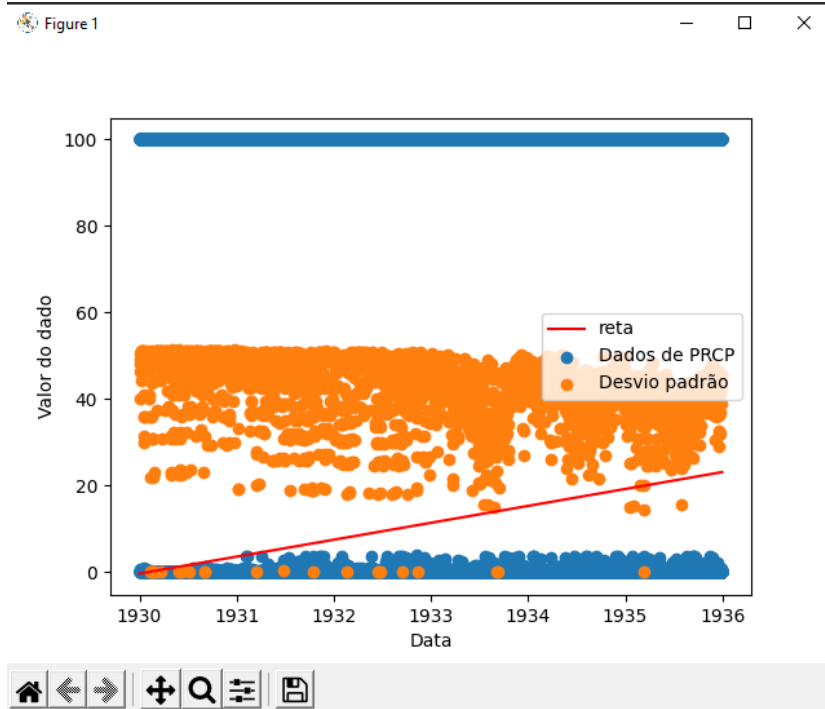
☐ Maximum wind gust

☐ Maximum temperature

☐ Minimum temperature

☒ Precipitation amount

☐ Snow depth



Predição do tempo

Data de início: 01 01 1930

Data de término: 01 06 1930

☐ Mean temperature

☐ Mean dew point

☐ Mean sea level pressure

☒ Mean station pressure

☐ Mean visibility

☐ Mean wind speed

☐ Maximum sustained wind speed

☐ Maximum wind gust

☐ Maximum temperature

☐ Minimum temperature

☐ Precipitation amount

☐ Snow depth

☐ Mean temperature

☐ Mean dew point

☐ Mean sea level pressure

☐ Mean station pressure

☐ Mean visibility

☐ Mean wind speed

☐ Maximum sustained wind speed

☐ Maximum wind gust

☐ Maximum temperature

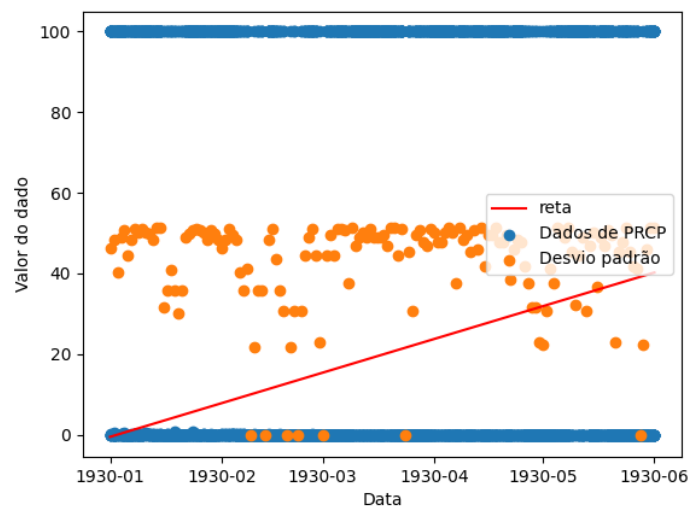
☐ Minimum temperature

☒ Precipitation amount

☐ Snow depth

clica

Figure 1



Predição do tempo

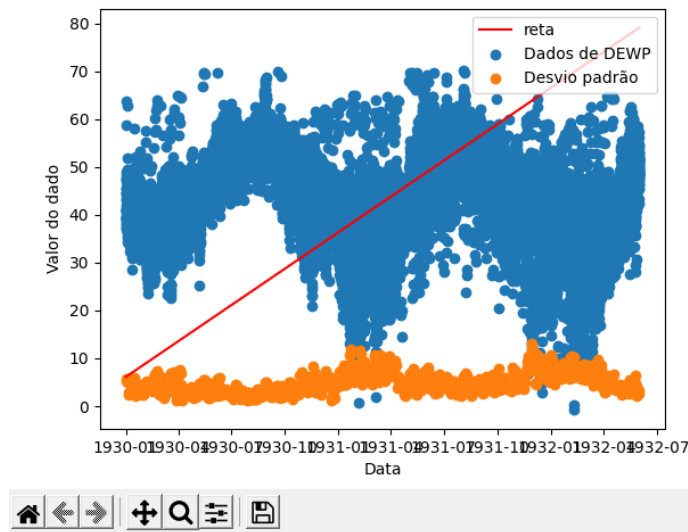
Data de início: 01 01 1930 Data de término: 01 06 1934

☒ Mean temperature
☐ Mean dew point
☐ Mean sea level pressure
☐ Mean station pressure
☐ Mean visibility
☐ Mean wind speed
☐ Maximum sustained wind speed
☐ Maximum wind gust
☐ Maximum temperature
☐ Minimum temperature
☐ Precipitation amount
☐ Snow depth

☐ Mean temperature
☒ Mean dew point
☐ Mean sea level pressure
☐ Mean station pressure
☐ Mean visibility
☐ Mean wind speed
☐ Maximum sustained wind speed
☐ Maximum wind gust
☐ Maximum temperature
☐ Minimum temperature
☐ Precipitation amount
☐ Snow depth

[clica](#)

Figure 1



Predição do tempo

Data de início:

Data de término:

☒ Mean temperature

☐ Mean dew point

☐ Mean sea level pressure

☐ Mean station pressure

☐ Mean visibility

☐ Mean wind speed

☐ Maximum sustained wind speed

☐ Maximum wind gust

☐ Maximum temperature

☐ Minimum temperature

☐ Precipitation amount

☐ Snow depth

☐ Mean temperature

☒ Mean dew point

☐ Mean sea level pressure

☐ Mean station pressure

☐ Mean visibility

☐ Mean wind speed

☐ Maximum sustained wind speed

☐ Maximum wind gust

☐ Maximum temperature

☐ Minimum temperature

☐ Precipitation amount

☐ Snow depth

[clica](#)

Figure 1

