# PROGRAMAÇÃO APLICADA

## Trabalho 2

Antenor Barros Leal  
Guilherme Montenegro Banharo

|  |
| --- |
| # Antes de tudo Descompacte o arquivo dados.zip. Após a descopactação e para existir as pastas aeroportos e voos. A pasta aeroportos com quatro arquivos e a pasta voos com 124 arquivos. |
| # Resumo Este trabalho detalha o processamento e análise de dados meteorológicos e de voos de aeroportos do sudeste brasileiro. A análise tem como objetivo identificar como as condições climáticas influenciam nos atrasos de voos. |
| Para responder esta pergunta usamos vários dataframes: um com as condições climáticas em um aeroporto e outros com as partidas e chegadas deste aeroporto. |
| O aeroporto escolhido será o do Galeão, por ter um maior movimento que o Santos Dumont, portanto mais dados para serem analisados. Este aeroporto possui código ICAO SBGL que será usado ao longo do código para se referir a este aeroporto. |
| Também serão comparados os atrasos com o aeroporto Santos Dumont, Congonhas e Guarulhos. |
| # Bases de dados |
| ## Base: Tempo |
| Possui as informações históricas metereológicas. É obtido acessando o endereço: http://a4barros.com/public/prog-aplicada/tempo.zip |
| São quatro arquivos no formato ‘dataset\_ICAO.xlsx’ onde |
| - ICAO=SBGL: Galeão - ICAO=SBGR: Guarulhos - ICAO=SBRJ: Santos Dumont - ICAO=SBSP: Congonhas |
| ### Descrição de colunas |
| - wind\_direction: Direção **de onde** o vento sopra em graus; - wind\_speed: Velocidade do vento em nós (milhas nauticas por hora); - temperature: Temperatura em graus Célsius; - dew\_point: Ponto de orvalho em graus Célsius; - qnh: Referência para o altímetro; - clouds\_few: Alturas em pés separadas por vírgulas das altitudes que existem nuvens few (1/8 a 2/8 do céu) presentes; - clouds\_scattered: O mesmo, mas para nuvens scattered (3/8 a 4/8 do céu); - clouds\_broken: O mesmo, mas para nuvens broken (5/8 a 7/8 do céu); - clouds\_overcast: O mesmo, mas para nuvens overcast (encoberto); - timestamp: Data e hora destas condições. |
| ## Base: Voos |
| Contém dados de pousos e decolagens em vários aeroportos do sudeste. Pode ser obtida em http://a4barros.com/public/prog-aplicada/voos.zip |
| São vários arquivos no formato: ‘YYYY-MM-DD-ICAO-arrivals.xlsx’ ou ‘YYYY-MM-DD-ICAO-departures.xlsx’. |
| Arrivals se refere as chegadas e departures as partidas. |
| Por exemplo: 2024-10-29-SBGL-arrivals.xlsx São as chegadas para o Galeão do dia 29 de outubro. |
| ### Descrição de colunas |
| - flight\_date: Data no formato YYYY-MM-DD. - flight\_status: status do voo pode ser: active, landed, diverted, scheduled, cancelled, unknown; - departure\_airport: Nome popular do aeroporto. - departure\_timezone: Fuso horário do aeroporto (ex.: America/Sao\_Paulo); - departure\_iata: Código IATA do aeroporto de partida. (ex.: SDU); - departure\_icao: Código ICAO do aeoporto de partida (ex.: SBRJ); - departure\_terminal: Terminal de partida do voo; - departure\_gate: Portão de embarque de onde o voo parte (ex.: C02); - departure\_scheduled: Horário programado para a partida do voo no formato de hora UTC (YYYY-MM-DDTHH:MM:SS+00:00); - departure\_estimated: Horário estimado para a partida do voo no formato de hora UTC; - arrival\_airport: Nome popular do aeroporto de chegada; - arrival\_timezone: Fuso horário do aeroporto de chegada, no formato de região. Ex.: America/Sao\_Paulo; - arrival\_iata: Código IATA do aeroporto de chegada (ex.: GRU); - arrival\_icao: Código ICAO do aeroporto de chegada (ex.: SBGR); - arrival\_terminal: Terminal de chegada do voo; - arrival\_gate: Portão de desembarque onde o voo chega (ex.: A02); - arrival\_baggage: Número da esteira onde as bagagens do voo serão disponibilizadas (ex.: 04); - arrival\_delay: Atraso na chegada do voo em minutos, considerando o horário programado. - arrival\_scheduled: Horário programado para a chegada do voo no formato de hora UTC; - arrival\_estimated: Horário estimado para a chegada do voo no formato de hora UTC; - airline\_name: Nome da companhia aérea operadora do voo (ex.: LATAM Airlines); - airline\_iata: Código IATA da companhia aérea (ex.: LA para LATAM); - airline\_icao: Código ICAO da companhia aérea (ex.: TAM para LATAM); - flight\_number: Número único do voo designado pela companhia aérea (ex.: 1234); - flight\_iata: Código IATA completo do voo, formado pelo código da companhia e o número do voo (ex.: LA1234); - flight\_icao: Código ICAO completo do voo, formado pelo código ICAO da companhia e o número do voo (ex.: TAM1234). |
| # Perguntas respondidas |
| 1. Quando os valores de vento não aparecem, significa que não há vento. Complete os valores ausentes de velocidade do vento com zero e os valores ausentes de direção com com a mediana das direções. Completar com a mediana é usada para que ouliers não afetem algum cálculo de média feito com a direção do vento. Mostre os 10 maiores e os 10 menores valores ordenados por velocidade de vento. |
| \* Objetivos: Preparar a coluna de vento para posterior análise. Ter uma ideia dos extremos de vento. |
| \* Requisitos atendidos: 2 (preenchimento de valores ausentes), 8 (medidas de sumarização: mediana), 1 (Concatenação) |

1. Os valores de nuvens few (poucas), scatered (espalhadas), broken (muitas) e overcast (encoberto) são listas de números separados por vírgula com a altitude de cada nuvem. Por exemplo, few com valor “10000,12000” indicam poucas nuvens em 10 mil pés e 12 mil pés.

Crie uma coluna ‘nivel\_nuvem’ com o valor do tipo de nuvem mais encoberto seguindo a ordem few < scatered < broken < overcast. Para garantir que as nuvens realmente afetam o aeroporto, considere APENAS nuvens abaixo de 10 mil pés.

Qual o mais nebuloso (mais fechado) tipo de formação para cada valor de temperatura? Parece haver relação entre a nebulosidade e a temperatura?

* Objetivo: Filtrar os dados de nuvem para os que podem influenciar o aeroporto. Juntar dados de nuvem que estavam espalhados em quatro colunas em apenas uma coluna com o tipo de nuvem mais crítico.
* Requisitos atendidos: 3 (apply), 8 (medidas de sumarização (grupos simples)), 7 (gráfico barra)

1. A velocidade de vento está expressa em nós (milhas náuticas por hora), converta para km/h. Crie as seguintes categorias para a velocidade do vento:
   * **Calmo:** Menor ou igual à 2km/h
   * **Bafagem:** 2 à 5 km/h
   * **Brisa leve:** 6 a 11km/h
   * **Brisa fraca:** 12 a 19km/h
   * **Brisa moderada:** 20 a 28km/h
   * **Brisa forte:** 29 a 38km/h
   * **Vento fresco:** 39 a 49km/h
   * **Vento forte:** 50 a 61km/h
   * **Ventania:** 62 a 74km/h
   * **Ventania forte:** 75 a 88km/h
   * **Tempestade:** 89 a 102km/h
   * **Tempestade violenta**: 103 a 117km/h
   * **Furacao:** Maior que 118km/h

Esta é chamada de Escala de Beaufort.

3.1. Faça uma tabela de frequências destas categorias e mostre em um gráfico pizza. Qual é o tipo de vento mais presente?

3.2. Mostre uma tabela de frequência com o cruzamento das categorias de vento com os valores de temperatura. Em qual faixa de temperatura ocorrem mais ventos?

3.3. Parece haver relação entre velocidade do vento e temperatura?

* Objetivo: Discretizar as velocidades de vento em categorias comumente usadas na meteorologia e verificar a existência de relação entre a velocidade do vento e a temperatura.
* Requisitos atendidos: 4 (categorização com pd.cut), 3 (apply), 9 (cruzamento simples), 7 (gráfico pizza), 6 (tabela de frequência com valores absolutos)

1. Junte os dataframes de dados de voo do aeroporto do Galeão. Faça um Merge da tabela de condições meteorológicas com os atrasos. Crie as colunas atraso\_chegada e atraso\_partida.

Faça o cruzamento de frequência entre o nível do vento e os atrasos e entre a pior formação de nuvens e os atrasos. Parece haver uma correlação?

* Objetivo: Verificar a possível relação entre a piora das condições de tempo com atrasos de voo.
* Requisitos atendidos: 1 (Concatenação), 2 (preenchimento de valores ausentes), 4 (categorização com pd.cut), 9 (cruzamento simples)

1. Calculando a diferença entre a temperatura e o ponto de orvalho temos um valor que quanto mais baixo, maior chance de chuva. Quando a diferença é zero, temos 100% de chance de chuva. Retire valores maiores de 10 graus, porque são outliers e filtre por tempo muito nebuloso ou visibiliade menor que 5km.

* Objetivo: Criar uma medida proporcional a chance a chuva e verificar se esta medida influencia nos atrasos em condições adversas de tempo.
* Requisitos atendidos: 9 (cruzamento estruturado), 5 (filtro)

1. Crie uma tabela no seguinte formato em que cada coluna é um aeroporto e cada linha é uma hora. Como valores, temos a média de atraso naquele aeroporto naquela hora. Mostre apenas as linhas que possuem em algum aeroporto atrasos maiores que 1h. Destes qual aeroporto tem o maior atraso acumulado?

* Objetivo: Ver qual aeroporto tem o pior somatório de atrasos extremos.
* Requisitos atendidos: 9 (cruzamento estruturado), 5 (filtro), 8 (medidas de sumarização)

1. Qual foi o pior atraso no aeroporto de congonhas no último dia de outubro?

* Objetivo: Ver o pior atraso de um dia específico em um aeroporto específico
* Requisitos atendidos: 5 (filtro de índice e filtro de valor)

1. Qual o tempo médio de atrasos médios diários do SBGL e qual a correlação com Nível de Nuvem?

* Objetivo: Encontrar o tempo médio de atrasos médios diários do SBGL e relacionar como nível das nuvens, por meio de uma análise gráfica e a correlação entre eles.
* Requisitos atendidos: 8 (Sumarização), 7 (Gráfico de linhas)

# Conclusões

## 1

O aeroporto do Galeão, em relação a velocidade de vento, teve um outlier em que o vento chegou a 63 nós no dia 29/10/2024 as 23h (UTC). O segundo vento mais veloz foi 19 nós dia 26/10 as 19h (UTC). *Nota:* O dado original está com a velocidade do vento em nós, mais na frente iremos converter para km/h.

wind\_direction wind\_speed temperature dew\_point ...  
timestamp   
2024-10-29 23:00:00+00:00 90.0 63.0 23 19 ...  
2024-10-26 19:00:00+00:00 210.0 19.0 32 20 ...  
2024-10-30 16:00:00+00:00 160.0 18.0 28 19 ...

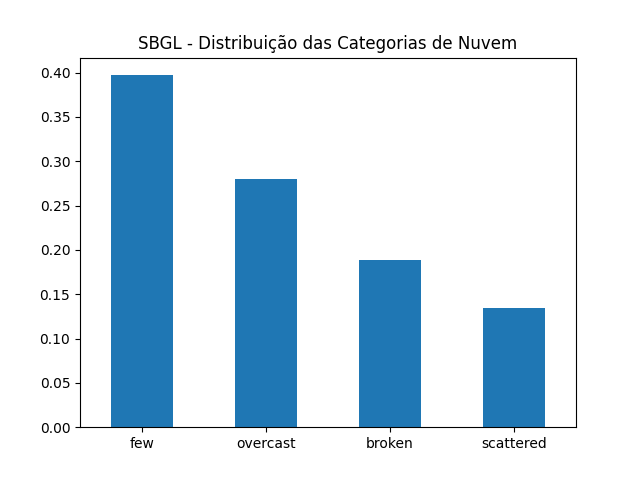
## 2

Para o aeroporto do Galeão temos uma correlação entre o pior tipo de nuvem e a temperatura de -0.5444268973056255, isto significa uma correlação inversa moderada.

Vendo por nível de temperatura, é fácil perceber esta correlação negativa: Para temperatura menores (20 a 26) temos nuvens encobertas e acima de 33 graus temos apenas nuvens esparsas.

----- Pior nível de nuvem abaixo de 10 mil por temperatura -----  
 nivel\_nuvem  
temperature   
20 overcast  
21 overcast  
22 overcast  
23 overcast  
24 overcast  
25 overcast  
26 overcast  
27 broken  
28 broken  
29 broken  
30 scattered  
31 broken  
32 broken  
33 scattered  
34 scattered  
35 scattered  
36 few

Porém na maior parte do tempo tivemos poucas nuvens como mostra o gráfico de frequência.



Galeão Distribuição das categorias de nuvem

## 3

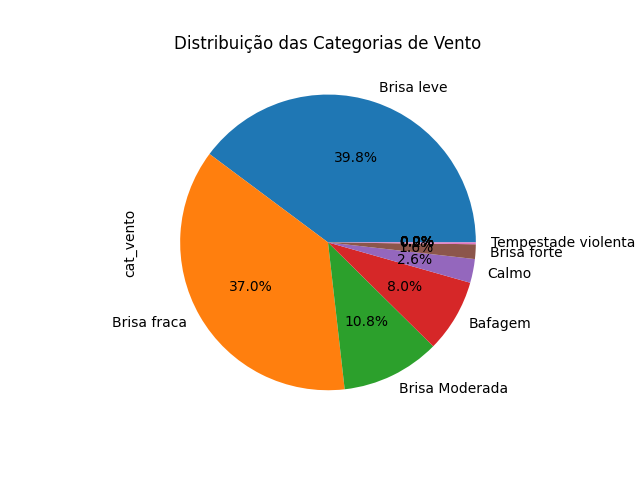
### 3.1

Para este aeroporto temos a grande predominância de ventos leves como mostra a tabela de frequência abaixo:

----- tabela de frequencia numérica de tipos de vento -----  
Brisa leve 199  
Brisa fraca 185  
Brisa Moderada 54  
Bafagem 40  
Calmo 13  
Brisa forte 8  
Tempestade violenta 1  
Vento fresco 0  
Vento forte 0  
Ventania 0  
Ventania fote 0  
Tempestade 0  
Furacao 0  
Name: cat\_vento, dtype: int64

Os tipos de ventos mais presentes são os mais fracos.

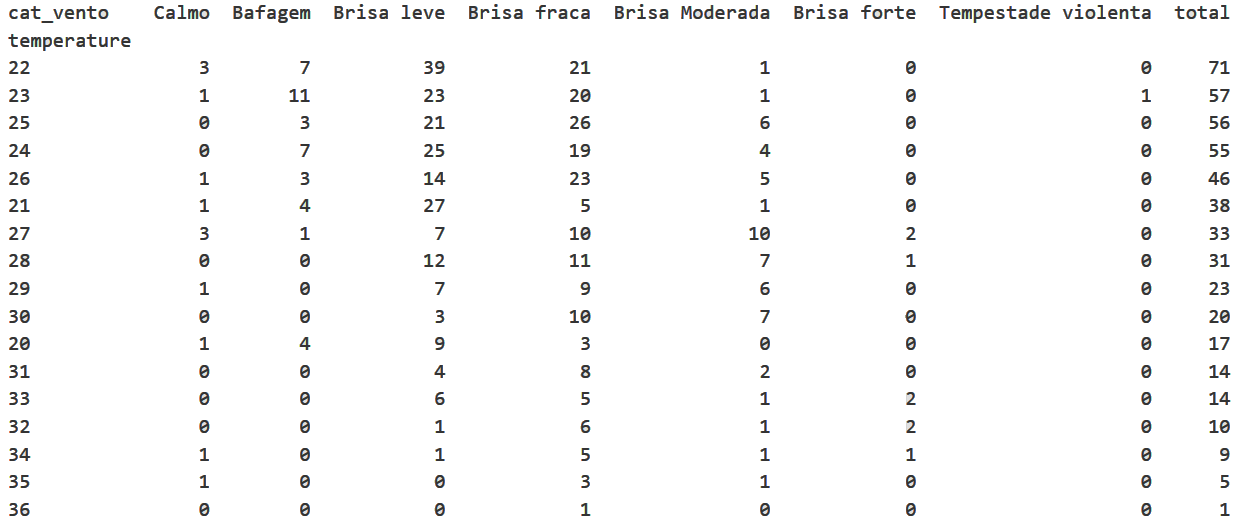
Vendo a mesma informação em forma de gráfico pizza temos:



Galeão Distribuição das categorias de vento

### 3.2

A maior quantidade de ventos de qualquer tipo ocorre em 22 graus e diminui monotonicamente com o aumento da temperatura.



### 3.3

A correlação entre a temperatura e a velocidade do vento é de 0.3029027092833759. Ou seja, há uma correlação, mas ela é leve.

### 4

Para partidas, nuvem do tipo few (poucas) parece influenciar muito atraso médio (10 a 30 min). Para chegadas o mesmo tipo few incluencia baixo atraso (menor que 10 min).

Para as categorias de vento a brisa leve parece causar atrasos médios nas partidas. Nas chegadas ela causa atrasos baixos.

----- Crosstab nível de nuvem x atraso partida -----  
nivel\_nuvem broken few overcast scattered  
atraso\_partida   
baixo atraso 82 394 192 141  
médio atraso 178 1387 0 342  
alto atraso 0 253 0 6  
altíssimo atraso 12 82 0 39  
----- Crosstab nível de nuvem x atraso chegada -----  
nivel\_nuvem broken few overcast scattered  
atraso\_chegada   
baixo atraso 251 1951 167 455  
médio atraso 6 120 15 73  
alto atraso 15 29 0 0  
altíssimo atraso 0 16 10 0  
----- Crosstab categoria do vento x atraso partida -----  
cat\_vento Calmo Bafagem Brisa leve Brisa fraca Brisa Moderada Brisa forte  
atraso\_partida   
baixo atraso 0 90 528 286 18 0  
médio atraso 242 248 1569 350 63 3  
alto atraso 0 48 149 71 6 0  
altíssimo atraso 0 38 0 103 0 0  
----- Crosstab categoria do vento x atraso chegada -----  
cat\_vento Calmo Bafagem Brisa leve Brisa fraca Brisa Moderada Brisa forte  
atraso\_chegada   
baixo atraso 242 420 2068 703 77 3  
médio atraso 0 4 141 78 6 0  
alto atraso 0 0 15 29 0 0  
altíssimo atraso 0 0 22 0 4 0

### 5

Uma maior chance de chuva influencia na quantidade de atrasos como mostra a tabela abaixo. Mas os mais longos atrasos e a maior quantidade de atrasos se concentram quando a diferença é de 4 graus.

nivel\_nuvem overcast total\_atrasos  
atraso 2.0 4.0 5.0 6.0 10.0 12.0 124.0 126.0   
diff\_temp   
3 0 6 0 0 0 12 0 0 18  
4 24 6 0 0 0 0 8 2 40  
6 0 1 4 1 2 0 0 0 8

### 6

Nota-se que existem mais atrasos superiores a uma hora nas partidas. Vide as tabelas Atraso médio por hora das partidas e Atraso médio por hora das chegadas no final da página. Para as horas que não apareceram nestas tabelas foi devido a todos os quatro aeroportos não terem tido atrasos.

Em atraso durante todo o período analisado nas partidas o aeroporto de Congonhas possui o maior somatório. Nas chegadas é o Santos Dumont.

atraso\_partida\_total pior\_atraso\_partida atraso\_chegada\_total pior\_atraso\_chegada  
ICAO   
SBGL 16670.0 245.0 3572.0 244.0  
SBGR 19574.0 1042.0 420.0 22.0  
SBRJ 13691.0 260.0 6581.0 95.0  
SBSP 27915.0 162.0 4008.0 123.0

----- Atraso médio por hora das partidas -----   
ICAO SBGL SBGR SBRJ SBSP  
row\_0   
2024-10-30 17:00:00+00:00 245.000000 0.000000 39.833333 38.285714  
2024-10-31 02:00:00+00:00 237.500000 0.000000 0.000000 0.000000  
2024-10-31 04:00:00+00:00 0.000000 135.666667 0.000000 0.000000  
2024-10-31 05:00:00+00:00 0.000000 102.294118 0.000000 0.000000  
2024-10-31 07:00:00+00:00 18.000000 7.333333 65.600000 15.250000  
2024-11-01 01:00:00+00:00 14.500000 61.789474 0.000000 0.000000  
2024-11-01 23:00:00+00:00 65.000000 0.000000 0.000000 0.000000  
2024-11-02 06:00:00+00:00 7.000000 63.300000 5.222222 6.923077  
2024-11-03 20:00:00+00:00 95.000000 0.000000 22.888889 13.500000  
2024-11-04 09:00:00+00:00 12.000000 0.000000 10.428571 61.000000  
2024-11-04 10:00:00+00:00 20.000000 0.000000 44.333333 82.666667  
2024-11-04 11:00:00+00:00 20.000000 0.000000 38.500000 65.375000  
2024-11-04 13:00:00+00:00 0.000000 0.000000 28.800000 62.666667  
2024-11-04 14:00:00+00:00 0.000000 0.000000 31.625000 67.666667  
2024-11-04 16:00:00+00:00 10.000000 0.000000 2.000000 82.666667  
2024-11-04 18:00:00+00:00 36.000000 0.000000 50.250000 72.875000  
2024-11-05 01:00:00+00:00 81.000000 42.200000 0.000000 0.000000  
2024-11-06 10:00:00+00:00 63.600000 0.000000 26.333333 20.166667  
2024-11-06 12:00:00+00:00 14.000000 0.000000 132.000000 30.666667  
2024-11-06 14:00:00+00:00 45.000000 0.000000 26.571429 72.000000  
2024-11-06 18:00:00+00:00 86.000000 0.000000 28.250000 41.444444  
2024-11-06 21:00:00+00:00 74.333333 22.000000 13.666667 15.000000  
2024-11-07 07:00:00+00:00 70.000000 13.166667 16.000000 42.000000  
2024-11-07 11:00:00+00:00 14.000000 0.000000 36.000000 65.571429  
2024-11-07 12:00:00+00:00 77.000000 0.000000 6.666667 83.400000  
2024-11-07 13:00:00+00:00 0.000000 0.000000 59.600000 77.333333  
2024-11-07 14:00:00+00:00 0.000000 0.000000 31.666667 85.666667  
2024-11-07 16:00:00+00:00 17.000000 23.000000 24.000000 103.500000  
2024-11-07 17:00:00+00:00 0.000000 0.000000 31.250000 65.000000  
2024-11-07 18:00:00+00:00 31.666667 0.000000 51.666667 85.400000  
  
----- Atraso médio por hora das chegadas -----   
ICAO SBGL SBGR SBRJ SBSP  
row\_0   
2024-10-29 10:00:00+00:00 73.000000 0.0 2.00 0.000000  
2024-10-29 17:00:00+00:00 0.000000 0.0 69.00 0.000000  
2024-10-30 10:00:00+00:00 244.000000 0.0 0.00 19.500000  
2024-10-31 14:00:00+00:00 0.000000 0.0 15.75 123.000000  
2024-11-01 19:00:00+00:00 64.000000 0.0 7.00 0.000000  
2024-11-03 09:00:00+00:00 0.000000 0.0 82.00 0.000000  
2024-11-03 18:00:00+00:00 198.000000 0.0 0.00 14.000000  
2024-11-06 08:00:00+00:00 4.375000 1.0 82.00 0.000000  
2024-11-07 10:00:00+00:00 2.428571 0.0 62.00 17.571429

### 7

O pior atraso no aeroporto de congonhas no último dia de outubro foi do TAP5239 com 64 minutos de atraso.

### 8

Por meio da análise do gráfico abaixo, e a correlação de 0.7048088948027401, podemos ver que existe uma correlação forte entre o atraso dos voos, com o nível das nuvens, sugerindo que condições meteorológicas relacionadas ao tipo de nuvem podem estar associadas a aos atrasos em voos. Isso é visto, à medida que quando o nível de nuvens se torna mais carregado(overcast), os atrasos tendem a ser maiores.

Aqui está o atraso médio por dia no aeroporto Galeão:

timestamp\_Dia  
2024-10-29 00:00:00+00:00 15.50  
2024-10-30 00:00:00+00:00 14.43  
2024-10-31 00:00:00+00:00 10.61  
2024-11-01 00:00:00+00:00 11.69  
2024-11-02 00:00:00+00:00 9.71  
2024-11-03 00:00:00+00:00 9.81  
2024-11-04 00:00:00+00:00 12.80  
2024-11-05 00:00:00+00:00 12.79  
2024-11-06 00:00:00+00:00 13.82  
2024-11-07 00:00:00+00:00 9.02  
Name: atraso\_medio, dtype: float64

Gráfico que mostra o atraso médio e o ponto máximo do nível das nuvens no dia. Mostrando que os maiores atrasos foram dias com nuvens mais carregadas e o de menor atraso com o céu mais limpo.

