

PROGRAMAÇÃO APLICADA

Trabalho 2

Antenor Barros Leal
Guilherme Montenegro Banharo

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 meteorológicas. 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 aeroporto 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 outliers não afetem algum cálculo de média feito com a direção do vento. Mostre os 20 primeiros 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)

-
2. Os valores de nuvens `few` (poucas), `scattered` (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 < scattered < 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)

3. 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. Para cada faixa de vento mostre temperatura mínima, média, máxima e desvio padrão. 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 as categorias de vento e a temperatura.
- Requisitos atendidos: 4 (categorização com pd.cut), 3 (apply), 9 (cruzamento simples), 8 (medidas de sumarização (grupos simples)), 7 (gráfico pizza), 6 (tabela de frequência com valores absolutos)

4. Junte os dataframes de dados de voo de um mesmo aeroporto. 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)
-

5. 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 visibilidade menor que 5km.

- Objetivo: Criar uma medida proporcional a chance a chuva e verificar se esta medida influencia nos atrasos.
 - Requisitos atendidos: 9 (cruzamento estruturado), 5 (filtro)
-

6. 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 atrasos maiores que 1h. Destes qual aeroporto tem o maior atraso acumulado?

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

7. 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)

Conclusões

1

O aeroporto do Galeão, em relação a velocidade de vento, teve um outlier em que o vento chegou a 63 km/h no dia 29/10/2024 as 23h (UTC). O segundo vento mais veloz foi 19km/h dia 26/10 as 19h (UTC).

	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 analisado, temos uma correlação entre o a temperatura e o tipo mais nebuloso de nuvem. De 20 graus até 26 (inclusivo), temos apenas formações totalmente encobertas. De 27 à 35 temos tanto nuvens (broken) como nuvens espalhadas (scattered). Na temperatura mais quente (36) temos poucas nuvens.

```
----- 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. Porém elas são seguidas por nuvens encobertas.

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
```

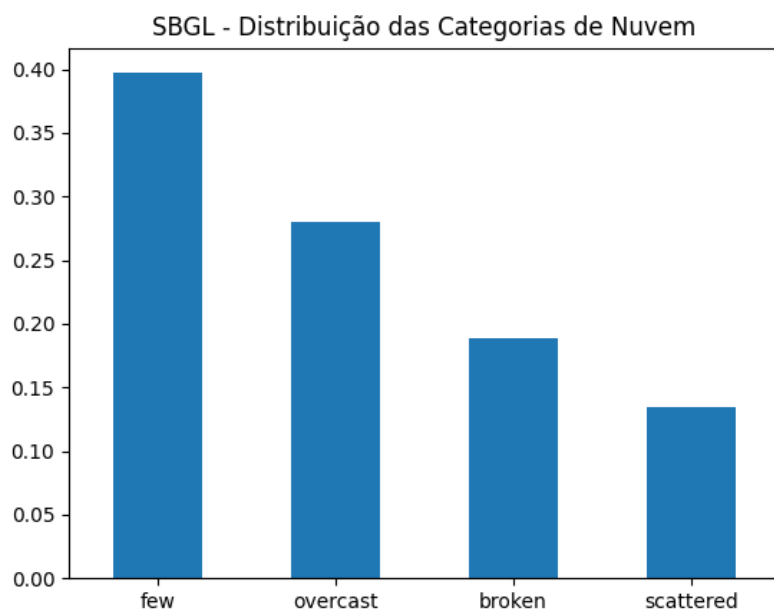


Figure 1: Galeão Distribuição das categorias de nuvem

```
Vento forte          0
Ventania             0
Ventania fote        0
Tempestade           0
Furacao              0
Name: cat_vento, dtype: int64
```

O tipo de vento mais presente é Brisa leve.

Vendo a mesma informação em forma de gráfico pizza temos:

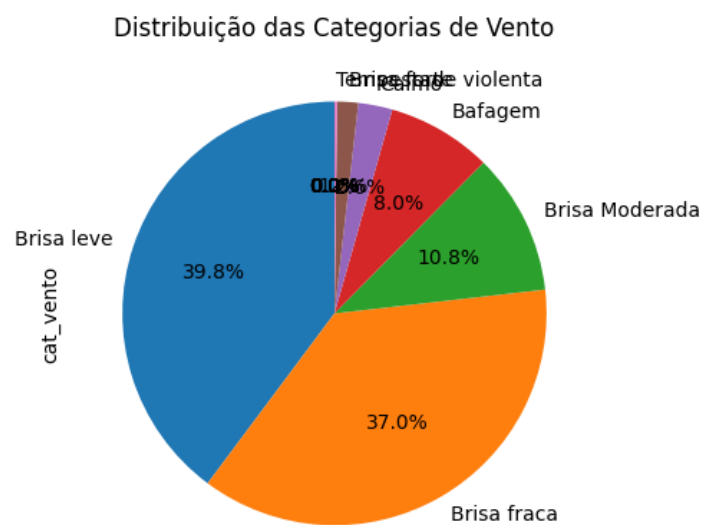


Figure 2: Galeão Distribuição das categorias de vento

3.2

A maior quantidade de ventos de qualquer tipo ocorre em 22 graus e em outras temperaturas mais baixas.

cat_vento	Calmo	Bafagem	Brisa leve	Brisa fraca	Brisa Moderada	Brisa forte	Tempestade violenta	total
temperature								
22	3	7	39	21	1	0	0	71
23	1	11	23	20	1	0	1	57
25	0	3	21	26	6	0	0	56
24	0	7	25	19	4	0	0	55
26	1	3	14	23	5	0	0	46
21	1	4	27	5	1	0	0	38
27	3	1	7	10	10	2	0	33
28	0	0	12	11	7	1	0	31
29	1	0	7	9	6	0	0	23
30	0	0	3	10	7	0	0	20
20	1	4	9	3	0	0	0	17
31	0	0	4	8	2	0	0	14
33	0	0	6	5	1	2	0	14
32	0	0	1	6	1	2	0	10
34	1	0	1	5	1	1	0	9
35	1	0	0	3	1	0	0	5
36	0	0	0	1	0	0	0	1

3.3

O tipo de vento mais forte neste aeroporto (Tempestade violenta) ocorreu em uma temperatura mais baixa onde a amplitude estava zero. Porém não parece haver uma relação significativa entre a temperatura e o tipo de vento.

cat_vento	temperature		amplitude	
	min	max	mean	
Calmo	20	35	25.769231	15
Bafagem	20	27	22.975000	7
Brisa leve	20	34	24.266332	14
Brisa fraca	20	36	26.194595	16
Brisa Moderada	21	35	27.555556	14
Brisa forte	27	34	30.750000	7
Tempestade violenta	23	23	23.000000	0

4

Para partidas, nuvem do tipo few (poucas) parece influenciar atraso médio (10 a 30 min).

----- Crosstab nível de nuvem x atraso partida -----				
nivel_nuvem	broken	few	overcast	scattered
atraso_partida				
baixo atraso	8	63	18	14
médio atraso	31	312	41	58
alto atraso	12	99	0	22
altíssimo atraso	0	25	0	1

Para chegadas o mesmo tipo de nuvem influencia baixo atraso.

----- Crosstab nível de nuvem x atraso chegada -----				
nivel_nuvem	broken	few	overcast	scattered
atraso_chegada				
baixo atraso	45	456	50	87
médio atraso	1	32	7	8

alto atraso	5	7	0	0
altíssimo atraso	0	4	2	0

Ao contrário do que eu achada, ventos muito fortes não parecem causar mais atrasos. A maioria dos atrasos se concentram em Brisa leve (6 a 11km/h). Mas o vento um pouco mais forte (Brisa fraca, 12 a 19km/h) foi o único tipo que causou altíssimo atraso (mais que uma hora).

----- Crosstab categoria do vento x atraso partida -----						
cat_vento	Calmo	Bafagem	Brisa leve	Brisa fraca	Brisa Moderada	Brisa forte
atraso_partida						
baixo atraso	0	11	79	40	1	0
médio atraso	41	12	370	110	13	1
alto atraso	0	43	57	32	3	0
altíssimo atraso	0	0	0	30	0	0

Para as chegadas, temos menos atrasos em geral que foram causados principalmente por brisa leve e fraca.

----- Crosstab categoria do vento x atraso chegada -----						
cat_vento	Calmo	Bafagem	Brisa leve	Brisa fraca	Brisa Moderada	Brisa forte
atraso_chegada						
baixo atraso	41	64	463	190	13	1
médio atraso	0	2	34	15	2	0
alto atraso	0	0	5	7	0	0
altíssimo atraso	0	0	4	0	2	0

5

Uma maior chance de chuva influencia na quantidade de atrasos como mostra a tabela abaixo. Em três graus de diferença temos bem mais atrasos que em diferenças maiores.

nivel_nuvem	overcast	total_atrasos						
atraso	10.0	12.0	14.0	16.0	17.0	32.0	246.0	
diff_temp								
3	0	0	0	31	0	6	0	37
4	6	0	0	0	0	0	2	8
5	0	9	0	0	0	0	0	9
6	0	1	2	1	1	0	0	5

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