

Análise Exploratória- Atrasos em Companhias Aéreas

Carlos Reis - RA: 2401937

Natalia Reis - RA: 2401504

Luigi Lago - RA: 2400428

Julio Dyna - RA: 2402440

Contextualização

- ▶ Esta apresentação aborda a análise exploratória de dados focada nos atrasos em companhias aéreas, Iniciando pela a configuração do ambiente, importação dos dados, geração de gráficos e análises e conclusões..

Questões de pesquisa

- ▶ Ao final deste estudo, com base nas análises dos gráficos e explorações feitos, 3 questões devem ser respondidas:

1.Qual(s) a(s) companhia(s) que mais registram atrasos?

2.A rota ou aeronave podem influenciar nos atrasos?

3.Existe algum padrão ou tendência nos atrasos ? Se sim, o que pode ser feito para reduzi-los?

Bibliotecas necessárias

- ▶ Para o estudo, é necessário importar as bibliotecas numpy, pandas e matplotlib.pyplot, utilizando os respectivos aliases np, pd e plt.

Leitura de dados Brutos

- Vamos importar dados brutos usando a biblioteca Pandas, ler um arquivo CSV, definir a coluna de índice e visualizar as primeiras linhas do DataFrame.

	year	month	day	dep_time	dep_delay	arr_time	arr_delay	carrier	tailnum	flight	origin	dest	air_time	distance	hour	minute
1	2013	1	1	517.0	2.0	830.0	11.0	UA	N14228	1545	EWR	IAH	227.0	1400	5.0	17.0
2	2013	1	1	533.0	4.0	850.0	20.0	UA	N24211	1714	LGA	IAH	227.0	1416	5.0	33.0
3	2013	1	1	542.0	2.0	923.0	33.0	AA	N619AA	1141	JFK	MIA	160.0	1089	5.0	42.0
4	2013	1	1	544.0	-1.0	1004.0	-18.0	B6	N804JB	725	JFK	BQN	183.0	1576	5.0	44.0
5	2013	1	1	554.0	-6.0	812.0	-25.0	DL	N668DN	461	LGA	ATL	116.0	762	5.0	54.0

Diante das informações importadas, os atributos são definidos como:

- **year:** Inteiro - Ano do voo.
- **month:** Inteiro - Mês do voo.
- **day:** Inteiro - Dia do voo.
- **dep_time:** Float - Hora de partida. Em HHmm (2 dígitos para as horas 2 dígitos para minutos)
- **dep_delay:** Float - Atraso na partida (em minutos).
- **arr_time:** Float - Hora de chegada. (2 dígitos para as horas 2 dígitos para minutos)
- **arr_delay:** Float - Atraso na chegada (em minutos).
- **carrier:** String - Código da companhia aérea.
- **tailnum:** String - Número de cauda da aeronave.
- **flight:** Inteiro - Número do voo.
- **origin:** String - Aeroporto de origem.
- **dest:** String - Aeroporto de destino.
- **air_time:** Float - Tempo de voo (em minutos).
- **distance:** Inteiro - Distância entre os aeroportos de origem e destino (em milhas).
- **hour:** Float - Hora de partida (extraída de dep_time).
- **minute:** Float - Minuto de partida (extraído de dep_time).

Quais companhias registram mais atrasos?

- Para determinar quais companhias aéreas registram mais atrasos, foram analisado os atrasos médios e a frequência de atrasos por companhia aérea. Foi feita a remoção dos valores nulos e menores/iguais a zero e acrescentou uma coluna com a frequência relativa do atraso (qtd de viagens atrasadas/qtd de viagens totais). O DataFrame fica da seguinte forma:

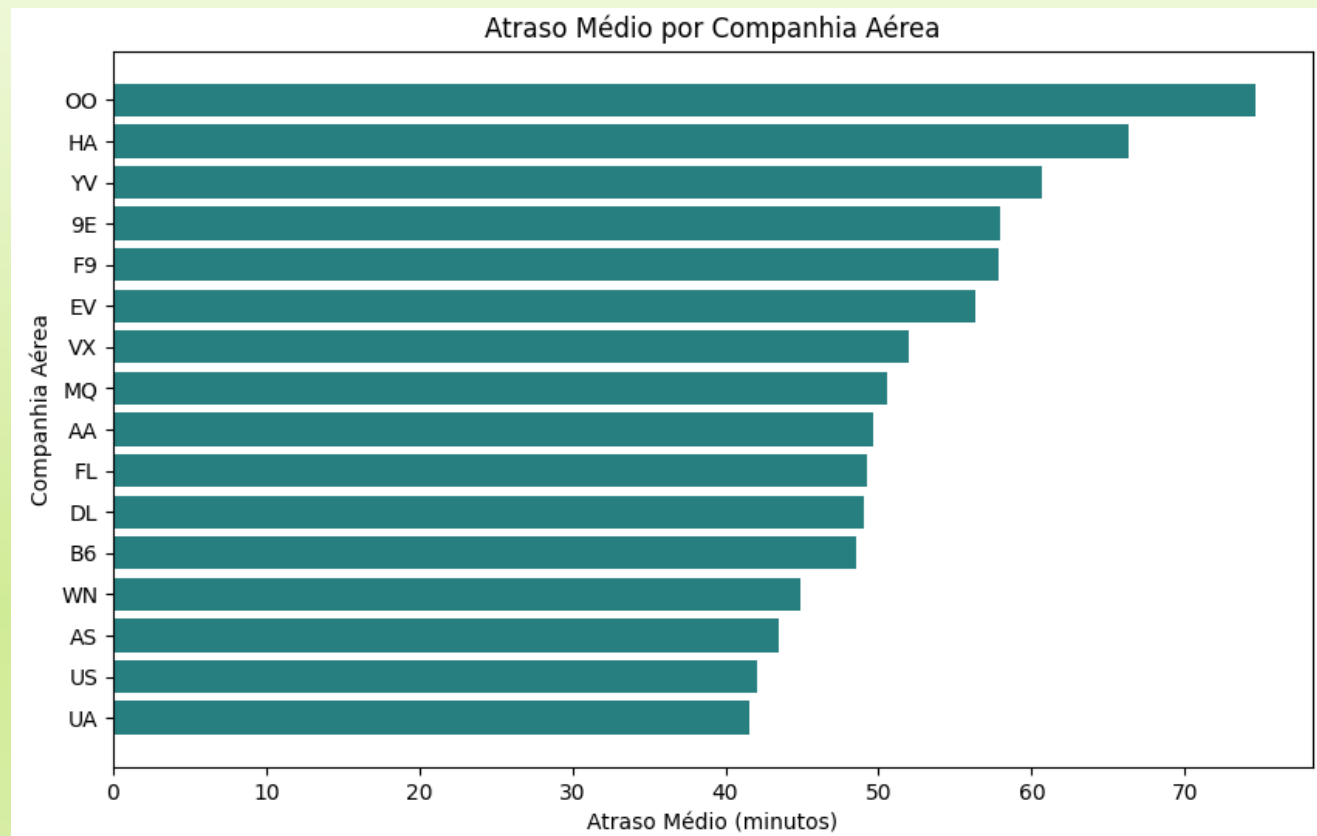
	Companhia	Atraso_Medio	Total_Voos_Atrasados	Soma_Atrasos	Frequencia_Relativa_Bruto
10	OO	74.625000	8	597.0	25.000000
8	HA	66.317073	41	2719.0	11.988304
15	YV	60.690821	207	12563.0	34.442596
0	9E	57.988536	5321	308557.0	28.824485
6	F9	57.815331	287	16593.0	41.897810

- Sendo os campos:

- **Companhia:** String - Código da companhia aérea.
- **Atraso_Medio:** Float - Tempo em minutos médios dos atrasos
- **Total_Voos_Atrasados:** Inteiro - Quantidade de voos atrasados registrados por companhia
- **Soma_Atrasos:** Float - Somatória do tempo total de minutos atrasados
- **Frequencia_Relativa_Bruto:** Float
Frequência relativa dos voos atrasados.

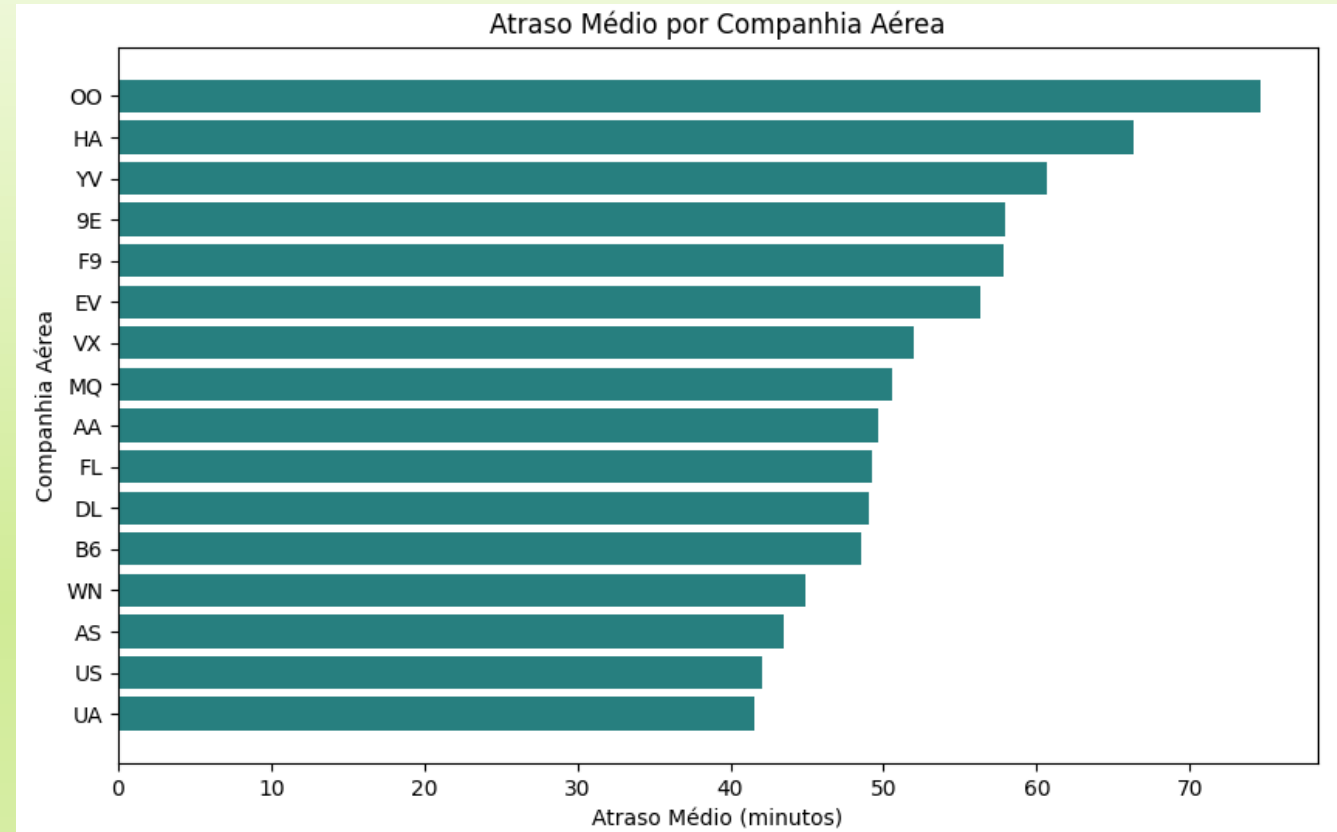
Resultados Quantitativos e Qualitativos

- ▶ A companhia **OO** tem o maior atraso médio por voo, com o valor de **74,625** minutos em media de atraso, mas apresenta apenas 8 voos atrasados, totalizando 25 % dos voos que a empresa fez. Embora o número total de voos atrasados seja baixo (8 voos), o impacto por voo é alto.
- ▶ **HA** segue com um atraso médio de **66.32** minutos por voo, em 41 voos registrados, aproximadamente 12% dos seus voos com atraso. Apesar de poucos dos seus voos atrasarem, o impacto por atraso é alto.
- ▶ **YV** também se destaca com um atraso médio de **60.69** minutos em um total de 207 voos.
- ▶ **9E** segue com um atraso médio de **57.99** minutos em 5.321 voos, mostra que o impacto total do atraso é substancial, especialmente em termos absolutos.
- ▶ **F9** tem um atraso médio de **57.82** minutos em 287 voos, completando a lista das companhias com os maiores atrasos médios.



Resultados Quantitativos e Qualitativos

- **Impacto Absoluto vs. Relativo:** Companhias como **9E** e **YV** operam um número significativamente maior de voos, o que significa que, mesmo com um atraso médio menor do que **OO** e **HA**, o impacto absoluto do atraso é muito maior. Isso sugere que estratégias de mitigação de atrasos devem priorizar essas companhias para maximizar a redução de atrasos em termos gerais.
- **Pequeno Número de Voos com Alto Impacto:** **OO** e **HA** têm menos voos, mas seus atrasos médios são extremamente altos. Isso pode indicar problemas específicos com rotas ou operações dessas companhias que, se resolvidos, poderiam melhorar significativamente a pontualidade.
- **Mitigação Focada em Alta Frequência:** Para companhias com um grande número de voos, como **9E**, qualquer redução no atraso médio poderia ter um impacto positivo considerável na experiência do passageiro e na eficiência operacional.



A rota ou aeronave podem influenciar nos atrasos?

- Para estudar se a rota poderia interferir nos atrasos, fez-se uma tabela levantando as rotas, os atrasos médios, o total de voos atrasados, a soma dos tempos de atraso, o total de voos geral daquela rota e a frequência relativa que o atraso representa daquele total.

	Origin	Destination	Atraso_Medio	Total_Voos_Atrasados	Soma_Atrasos	origin	dest	Total_Voos_Bruto	Frequencia_Relativa_Bruto
141	JFK	SDF	84.916667	12	1019.0	JFK	SDF	46	26.086957
83	EWB	TYS	83.757396	169	14155.0	EWB	TYS	323	52.321981
216	LGA	TVC	78.041667	24	1873.0	LGA	TVC	77	31.168831
23	EWB	DSM	73.401198	167	12258.0	EWB	DSM	411	40.632603
81	EWB	TUL	71.846154	156	11208.0	EWB	TUL	315	49.523810

A análise por rota (origem e destino) indica que algumas rotas específicas apresentam os maiores atrasos médios, particularmente aquelas partindo de JFK, mostrado no gráfico nos próximos slides:

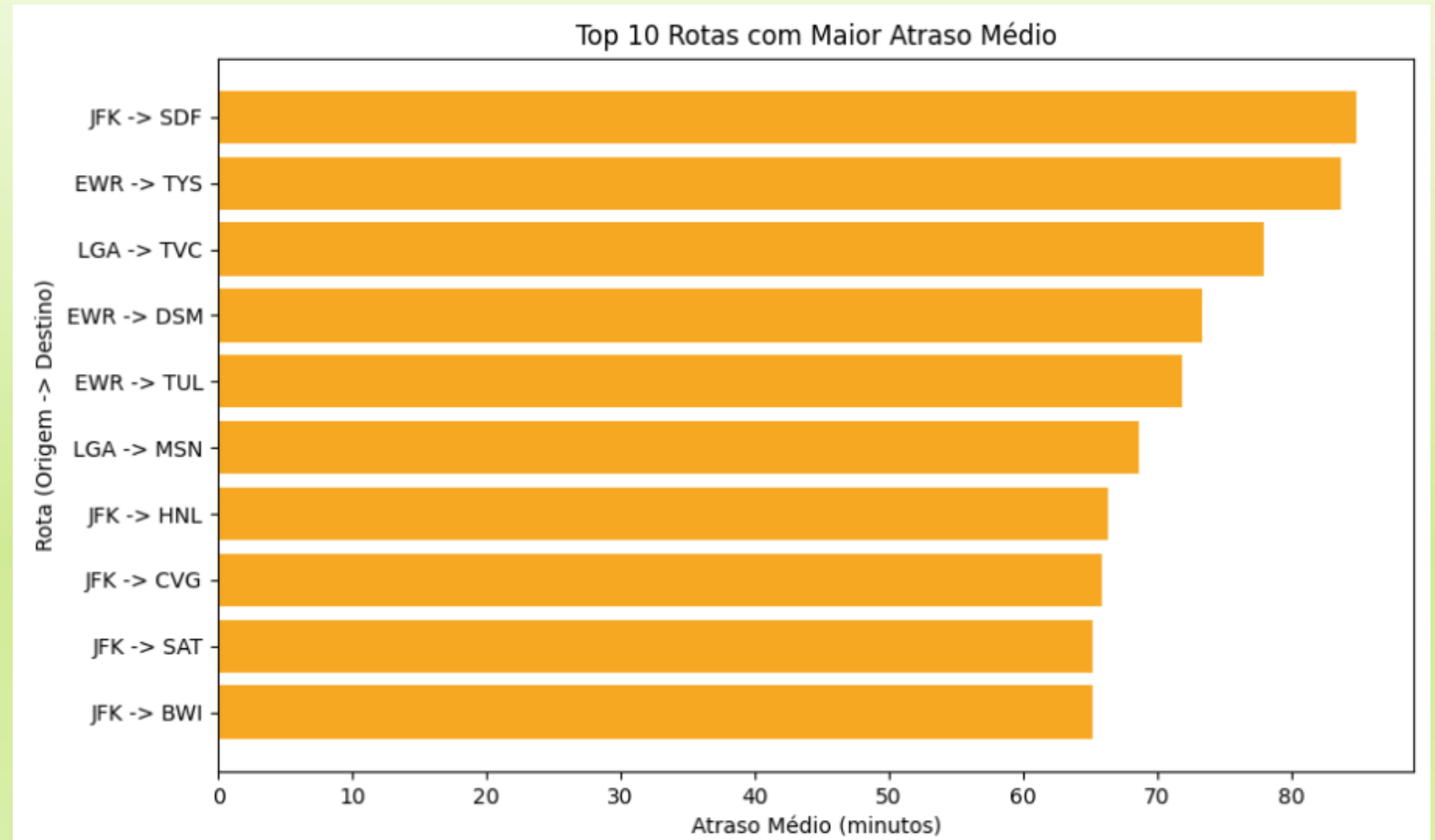
- **JFK -> SDF:** Atraso médio de **84.92** minutos.
- **EWB -> TYS:** Atraso médio de **83.58** minutos.
- **LGA -> TVC:** Atraso médio de **78.41** minutos.
- **EWB -> DSM:** Atraso médio de **73.04** minutos.
- **EWB -> TUL:** Atraso médio de **71.85** minutos.

Esses atrasos podem estar relacionados a fatores locais nos aeroportos de origem, como clima, tráfego aéreo intenso, ou problemas operacionais específicos que afetam essas rotas com maior frequência.

A rota ou aeronave podem influenciar nos atrasos?

Rotas Críticas: É evidente que certas rotas, especialmente aquelas partindo de aeroportos como JFK e EWR, são mais propensas a maiores atrasos.

Os destinos não tiveram tanto impacto nas 10 primeiras rotas, portanto os dados podem estar indicando um problema grave nestes aeroportos de origem.



A rota ou aeronave podem influenciar nos atrasos?

- Para estudar se a aeronave poderia interferir nos atrasos, fez-se uma tabela levantando as aeronaves, os atrasos médios, o total de voos atrasados, a soma dos tempos de atraso, o total de voos geral daquela aeronave e a frequência relativa que o atraso representa daquele total.

	Numero_Aeronave	Atraso_Medio	Total_Voos_Atrasados	Soma_Atrasos	tailnum	Total_Voos_Bruto	Frequencia_Relativa_Bruto
3216	N844MH	320.0	1	320.0	N844MH	1	100.000000
3466	N911DA	294.0	1	294.0	N911DA	1	100.000000
1499	N452UW	277.0	1	277.0	N452UW	7	14.285714
3533	N922EV	276.0	1	276.0	N922EV	1	100.000000
1067	N384HA	270.4	5	1352.0	N384HA	33	15.151515

A análise por aeronave (número da cauda) mostra que certas aeronaves específicas apresentam atrasos médios bastante elevados:

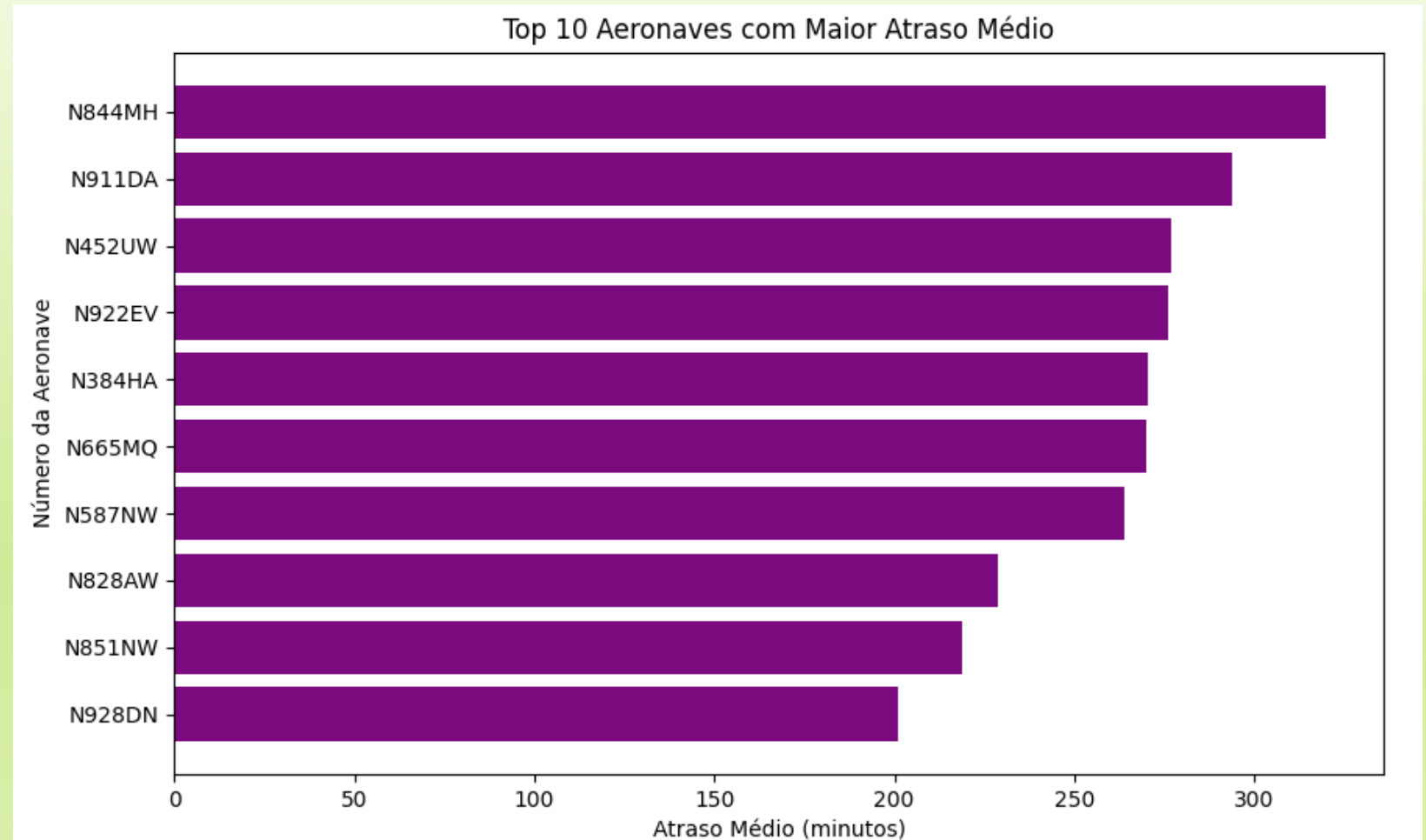
- **N844MH:** Atraso médio de **320** minutos.
- **N911DA:** Atraso médio de **294** minutos.
- **N545UW:** Atraso médio de **277** minutos.
- **N922EV:** Atraso médio de **276** minutos.
- **N384HA:** Atraso médio de **219** minutos.

Esses valores podem indicar problemas operacionais ou técnicos associados a essas aeronaves específicas. A alta variabilidade nos atrasos pode ser causada por questões de manutenção, problemas mecânicos ou até fatores externos que impactam repetidamente essas aeronaves.

A rota ou aeronave podem influenciar nos atrasos?

Rotas Críticas: É evidente que certas rotas, especialmente aquelas partindo de aeroportos como JFK e EWR, são mais propensas a maiores atrasos.

Os destinos não tiveram tanto impacto nas 10 primeiras rotas, portanto os dados podem estar indicando um problema grave nestes aeroportos de origem.



Existe algum padrão ou tendência nos atrasos?

- Para verificar se existe algum padrão ou tendência, deve-se observar o que foi levantado anteriormente além de novas informações, como um padrão de atrasos por mês, por dia do mês e por hora do dia.

Mes Atraso_Medio		
5	6	62.295127
6	7	62.180162
3	4	54.308961
4	5	48.957598
2	3	48.288027

Dia Atraso_Medio		
7	8	67.965732
9	10	63.035442
27	28	59.286864
21	22	57.023104
22	23	55.784437

Hora Atraso_Medio		
3	3.0	287.727273
2	2.0	224.650794
1	1.0	195.959276
0	0.0	128.763636
22	23.0	117.460899

Existe algum padrão ou tendência nos atrasos?

Atrasos por Mês

- ▶ A análise dos atrasos por mês revela um padrão interessante:
- ▶ **Junho** e **Julho** são os meses com os maiores atrasos médios, com **62.30** minutos e **62.18** minutos respectivamente.
- ▶ **Abril** e **Maio** também apresentam atrasos médios elevados, com **54.31** minutos e **48.96** minutos, possivelmente devido à transição sazonal e às condições climáticas variadas.

Atrasos por Dia do Mês

- ▶ Os dias **8**, **10**, e **28** do mês mostram atrasos médios mais elevados. Isso pode coincidir com eventos específicos, finais de semana ou períodos de maior tráfego, o que aumenta a probabilidade de atrasos.
- ▶ **Dia 8:** Atraso médio de **67.97** minutos.
- ▶ **Dia 10:** Atraso médio de **63.04** minutos.
- ▶ **Dia 28:** Atraso médio de **59.29** minutos.
- ▶ **Dia 22:** Atraso médio de **57.02** minutos.
- ▶ **Dia 23:** Atraso médio de **55.78** minutos.
- ▶ A análise diária pode ser crucial para identificar picos de atraso dentro de cada mês e ajustar a operação para reduzir esses atrasos.

Atrasos por Hora do Dia

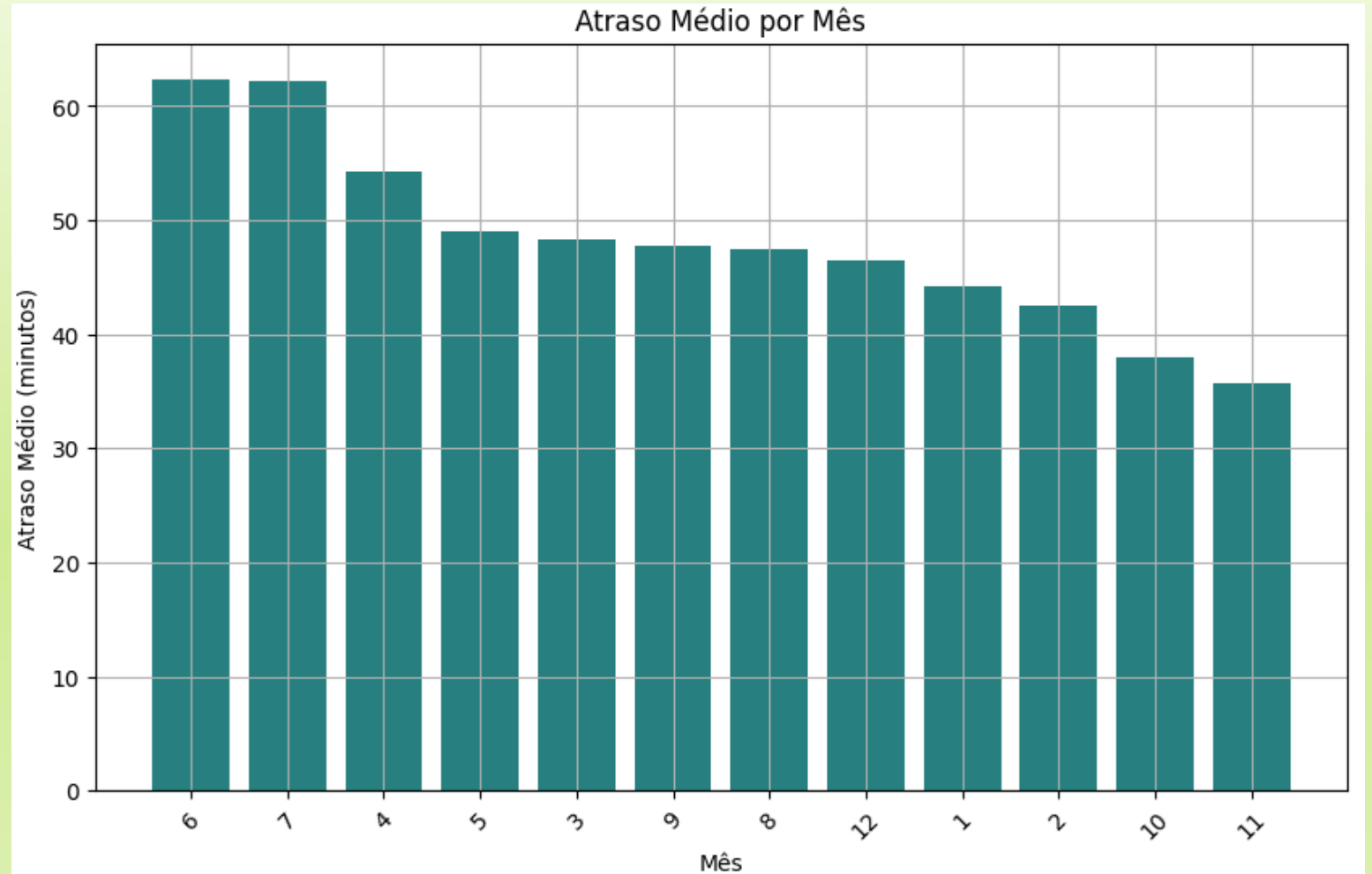
- ▶ Os atrasos variam significativamente conforme o horário do dia:
- ▶ As horas da madrugada, especialmente **03:00**, **02:00** e **01:00**, apresentam atrasos médios extremamente elevados, com **287.73** minutos, **224.65** minutos, e **195.96** minutos respectivamente. Isso pode indicar que os voos noturnos ou de madrugada estão mais suscetíveis a atrasos significativos, possivelmente devido à menor operação nos aeroportos, menos disponibilidade de pessoal e menor frequência de voos para diluir os atrasos.
- ▶ O horário **23:00** também apresenta atrasos médios altos, com **117.46** minutos, o que pode ser devido ao acúmulo de atrasos ao longo do dia.

Existe algum padrão ou tendência nos atrasos?

Os meses com o maior atraso médio, sendo estes Junho e Julho estão também coincidindo com os meses de verão da américa do norte (fonte da base de dados) o que provavelmente ocasiona um aumento no numero de viagens e consequentemente um aumento dos atrasos.

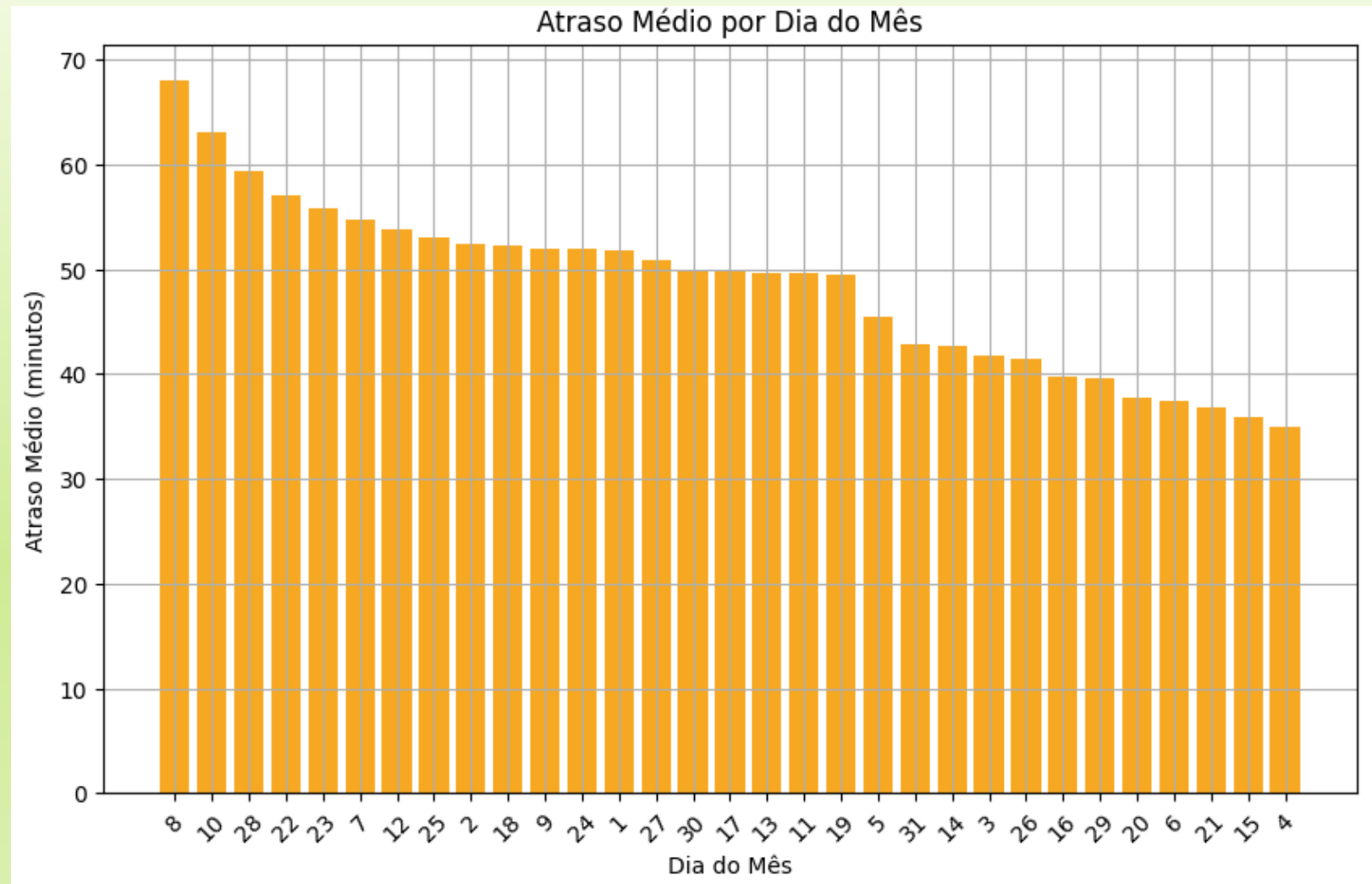
Os 5 primeiros meses correspondem ao período de primavera e verão na américa do norte, quando ocorrem férias escolares de verão (Summer Vacations) e o final do inverno, que é marcado pelo recesso de primavera (Spring Break)

Provavelmente aí está o padrão para os atrasos.



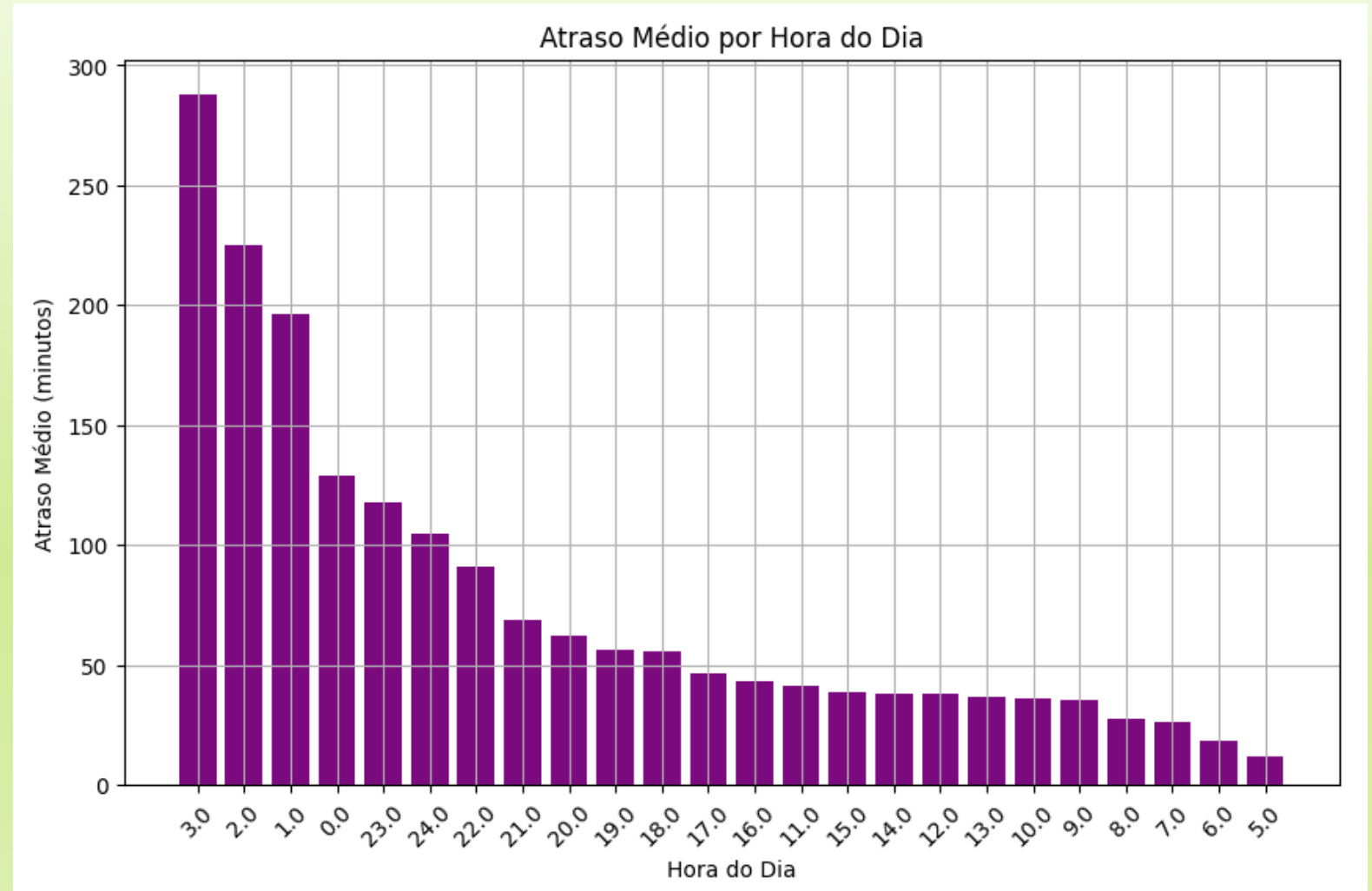
Existe algum padrão ou tendência nos atrasos?

A princípio, os dias do mês aparentemente não seguem um padrão para os atrasos.



Existe algum padrão ou tendência nos atrasos?

Neste ponto claramente observa-se um padrão para as 2 ultimas horas do dia e para as primeiras 3 horas do dia que existem atrasos.



Conclusão

É difícil afirmar qual seria a causa justificável para os atrasos nos voos e encontrar uma possível solução. Possivelmente se os aeroportos se organizarem melhor em momentos de pico, como nos meses de primavera e verão, os consumidores vão experimentar menos atrasos. Além disso deveria-se ter observações sobre o clima da região em cada situação, provavelmente com chuvas o atraso seja maior e mau tempo também, principalmente em meses de tempestades de verão ou em nevascas de inverno.

Com base no que foi levantado, é possível dizer que se o consumidor for marcar uma viagem de avião, partindo do aeroporto JFK, durante os meses de verão, no início da madrugada, pela companhia FL ou F9 (cada uma com respectivamente 45% e 41% dos voos com atraso) essa pessoa pode esperar enfrentar algum atraso na sua viagem.

GITHUBS

- ▶ Carlos Reis: <https://github.com/CarlosAReis/Analise-Exploratoria-Voos-NYC>
- ▶ Luigi Lago: <https://github.com/luigi-lago/Analise-Exploratoria-Atrasos>
- ▶ Natalia Reis: <https://github.com/NStrass/Analise-Exploratoria-Atrasos>