



**Atrasos nos voos americanos em 2009 e 2018**

**Gabriela Iannini Caruso (gabrielaic@al.insper.edu.br)**  
**Rafael Lucas Domingues de Andrade (rafaellda@al.insper.edu.br)**  
**Vitor Grando Eller (vitorge@al.insper.edu.br)**

**Relatório**  
**Versão Final**  
**Big Data para Dados Públicos**

**São Paulo - SP**

**Novembro de 2022**

# Sumário

<b>Novembro de 2022</b>	<b>1</b>
<b>Sumário</b>	<b>2</b>
<b>Introdução</b>	<b>3</b>
<b>Leitura de dados</b>	<b>3</b>
<b>Análises</b>	<b>5</b>
1. Atrasos de voos em 2009	5
1.1. Dado gerais do volume de atrasos ao longo do ano	5
Figura 1: Percentual de voos atrasados em cada mês de 2009.	5
Figura 2: Volume de atraso por mês no ano de 2009.	5
Figura 3: Motivos de atraso dos voos em 2009.	6
1.2. Companhias aéreas com mais voos atrasados	6
Figura 4: Atrasos em voos por companhias aéreas.	6
1.3. Relação entre tempo de voo e atraso na partida	7
Figura 5: Relação entre tempo de voo e atraso na partida.	7
1.4. Aeroportos com mais voos atrasados	8
Figura 6: Atrasos em voos por aeroportos, de Origem e de Destino dos voos.	8
2. Atrasos de voos em 2018	8
2.1. Dado gerais do volume de atrasos ao longo do ano	8
Figura 8: Percentual de voos atrasados em cada mês de 2018.	9
Figura 9: Motivos de atraso dos voos em 2018.	9
2.2. Companhias aéreas com mais voos atrasados	10
Figura 10: Atrasos em voos por companhias aéreas.	10
2.3. Relação entre tempo de voo e atraso na partida	10
Figura 11: Relação entre tempo de voo e atraso na partida.	10
2.4. Aeroportos com mais voos atrasados	11
Figura 12: Atrasos em voos por aeroportos, de Origem e de Destino dos voos	11
<b>Conclusões</b>	<b>11</b>
Pontos semelhantes entre os dois anos	11
Pontos diferentes entre os dois anos	12
Conclusão geral	12
<b>Repositório de análises</b>	<b>13</b>
<b>Referências</b>	<b>13</b>

# Introdução

A aviação potencializou a globalização mundial no passar da última década, não só no âmbito dos negócios, mas também da disseminação da cultura pelo globo. Porém, o acelerado avanço e crescimento na aviação comercial também trouxe consigo um aumento vertiginoso no número de passageiros e volume de voos.

Naturalmente, o número de atrasos nos voos também cresce, e isso impacta a economia de maneira estridente, como mostrado por um estudo da UC Berkeley, que estimou o impacto dos atrasos nos voos domésticos nos Estados Unidos no ano de 2007 em aproximadamente 32,9 bilhões de dólares.

Dada essa importância, buscamos analisar os dados de voos nos estados unidos nos anos de 2009 e 2018, para buscar entender melhor os comportamentos dos voos que apresentaram atrasos, bem como analisar onde se encontram as companhias aéreas e os aeroportos no espectro de atrasos em voos comerciais.

## Leitura de dados

Foi utilizada a base pública da *National Bureau of Transportation Statistics* (Escritório Nacional de Estatísticas de Transporte), obtida através da plataforma Kaggle. Tal base contém informações de todos os voos domésticos dos Estados Unidos, incluindo dados sobre o tempo estimado de voo, a origem e o destino, entre outros. Na tabela abaixo é possível visualizar as informações contidas na base de dados.

FL_DATE	OP_CARRIER	OP_CARRIER_FL_NUM	ORIGIN	DEST	CRS_DEP_TIME	DEP_TIME
Data do Voo	Companhia Responsável	Número do Voo	Aeroporto de Origem	Aeroporto de Destino	Hora estimada de partida	Hora de partida

DEP_DELAY	TAXI_OUT	WHEELS_OFF	WHEELS_ON	CRS_ARR_TIME	ARR_TIME	ARR_DELAY
Atraso na partida	Tempo de taxiamento na decolagem	Horário de decolagem	Horário de pouso	Hora estimada de chegada	Hora de chegada	Atraso na chegada

CANCELLED	CANCELLATION_CODE	CRS_ELAPSED_TIME	ACTUAL_ELAPSED_TIME	AIR_TIME
Indica se o voo foi cancelado	Código de cancelamento	Tempo estimado de voo	Tempo de voo	Tempo no ar

DISTANCE	WHEATHER_DELAY	NAS_DELAY	SECURITY_DELAY	LATE_AIRCRAFT_DELAY
Distância do voo	Indica se houve atraso pelo tempo	Indica se houve atraso pela NAS	Indica se houve atraso por questões de segurança	Indica se houve atraso pela aeronave chegar atrasada

As informações foram extraídas como arquivo no formato .xlsx, possuindo um arquivo diferente para cada ano na janela entre 2009 e 2018. Optou-se por analisar informações das duas extremidades para poder comparar as mudanças que ocorreram neste intervalo de tempo. O arquivo de 2009 possui 6.429.338 registros, ocupando um total de 792,6MB de espaço em disco. O arquivo de 2018 por sua vez possui 7.213.446 registros, ocupando um total de 893MB de espaço em disco.

Foram utilizadas as bibliotecas *Dask* e *Plotly Express* do Python para realizar as análises. Com o *dask* foi possível paralelizar a leitura das informações, enquanto o *plotly* foi utilizado para criar as visualizações das informações.

Além das colunas listadas acima, foram criadas algumas variáveis extras, que seriam utilizadas como apoio no momento das análises. Sendo elas:

- **WAS\_DELAY**: Variável que indica se o voo sofreu atraso ou não.
- **MONTH\_OF\_YEAR**: Variável que indica o mês do voo, para facilitar o agrupamento de informações.
- **CAUSE\_OF\_DELAY**: Para casos de voo que sofreram atraso, indica qual foi a causa dessa situação. São elas: Tempo, Companhia Aérea, NAS (*National Airspace System*), Segurança, ou Atraso na Aeronave.
- **AIR\_TIME\_DELAY**: Variável que indica se o tempo de ar foi maior ou menor que o planejado.
- **DELAY\_ON\_DEPARTURE**: Variável que indica se houve atraso na partida do voo ou não.

Todas as análises foram realizadas em *Jupyter Notebooks* separados para os respectivos anos, facilitando a leitura das informações, que foram então consolidadas neste relatório. Elas podem ser encontradas no repositório do grupo: <https://github.com/VFermat/insper-2202-bigdata>.

# Análises

## 1. Atrasos de voos em 2009

### 1.1. Dado gerais do volume de atrasos ao longo do ano

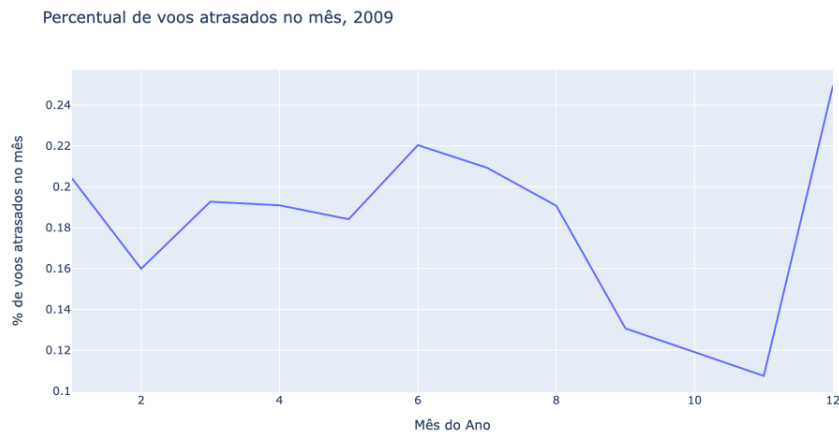


Figura 1: Percentual de voos atrasados em cada mês de 2009.

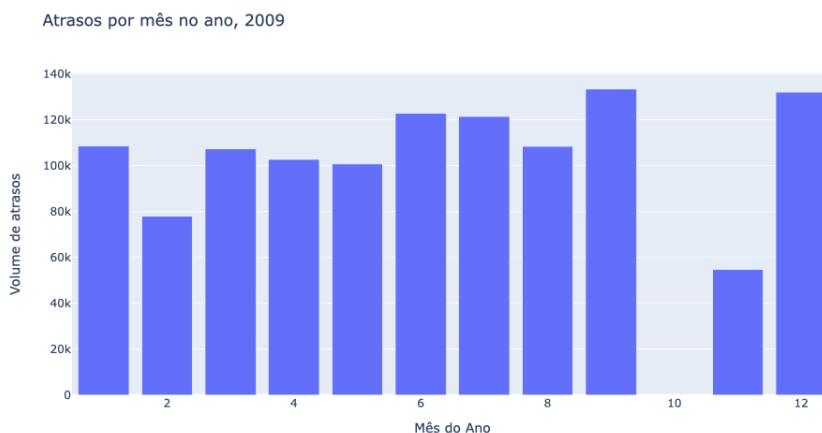


Figura 2: Volume de atraso por mês no ano de 2009.

Nas Figuras 1 e 2 pode-se identificar que alguns meses se destacam quanto a quantidade de voos atrasados em relação aos demais meses, principalmente quando observamos o valor percentual na Figura 1.

Uma possível interpretação para esse cenário é que meses como junho, julho, dezembro e janeiro geralmente constituem o período de férias escolares e universitárias, nos quais muitas pessoas costumam viajar. Além disso, o mês de dezembro conta com feriados internacionais como Natal e Ano Novo.

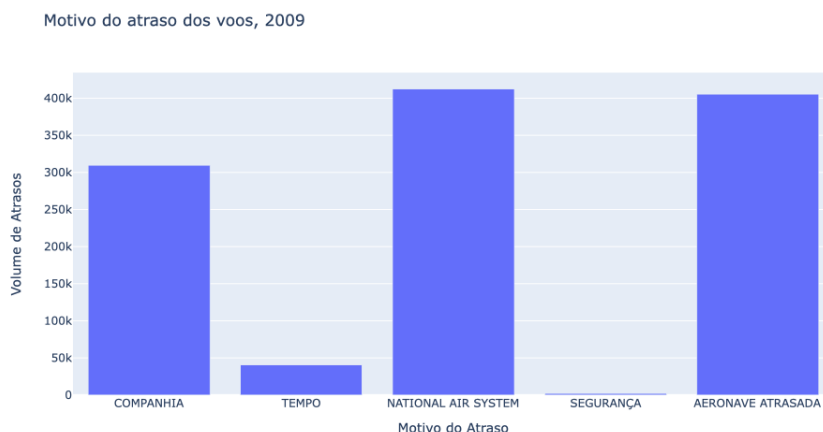


Figura 3: Motivos de atraso dos voos em 2009.

Na Figura 3 são identificadas 3 principais causas de atraso dos voos de 2009. Sendo elas: Companhia aérea, *National Air System* (controle de tráfego aéreo) e aeronave atrasada. Todas essas razões contam com a participação humana e, conseqüentemente, estão sujeitas a erros dos mesmos. Podendo, assim, justificar seus valores elevados em comparação com os outros fatores. Uma vez que o tempo é uma questão climática variável e a segurança em todos os aeroportos passou por diversas melhorias e implementações de novas medidas após o atentado terrorista em 11 de setembro de 2001.

## 1.2. Companhias aéreas com mais voos atrasados

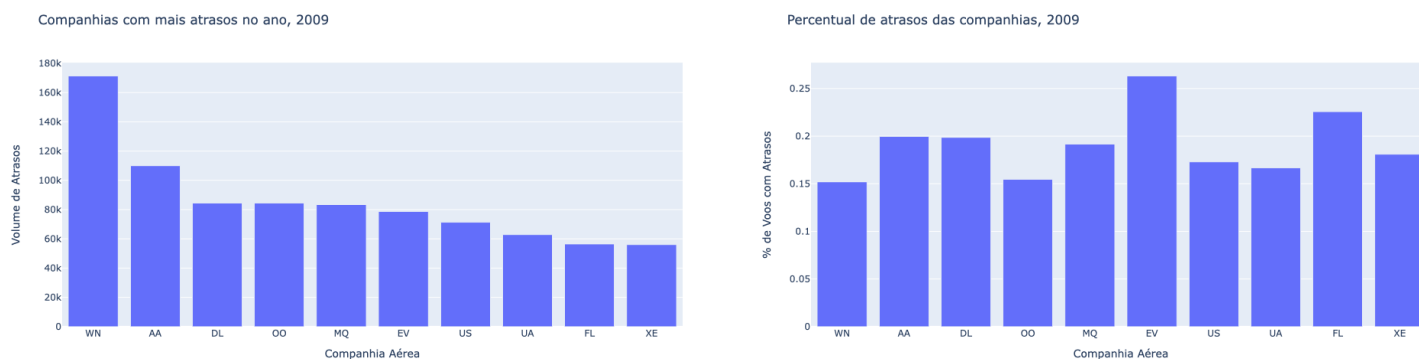


Figura 4: Atrasos em voos por companhias aéreas.

Na Figura 4 são apresentadas as siglas das companhias aéreas presentes na base de dados em relação ao volume de atrasos (percentual e absoluto).

A partir desses gráficos é possível identificar a WN como companhia aérea com o maior volume de atrasos absoluto e a EV como companhia aérea com o maior percentual de atrasos.

## 1.3. Relação entre tempo de voo e atraso na partida

No decorrer das análises, foi levantada a hipótese de que voos com atraso na partida poderiam ter seus tempos de voo encurtados, com o intuito de compensar um possível tempo perdido. Por isso, cruzamos a informação de tempo de voo previsto contra executado nos cenários onde houve e não houve atraso na partida do avião.

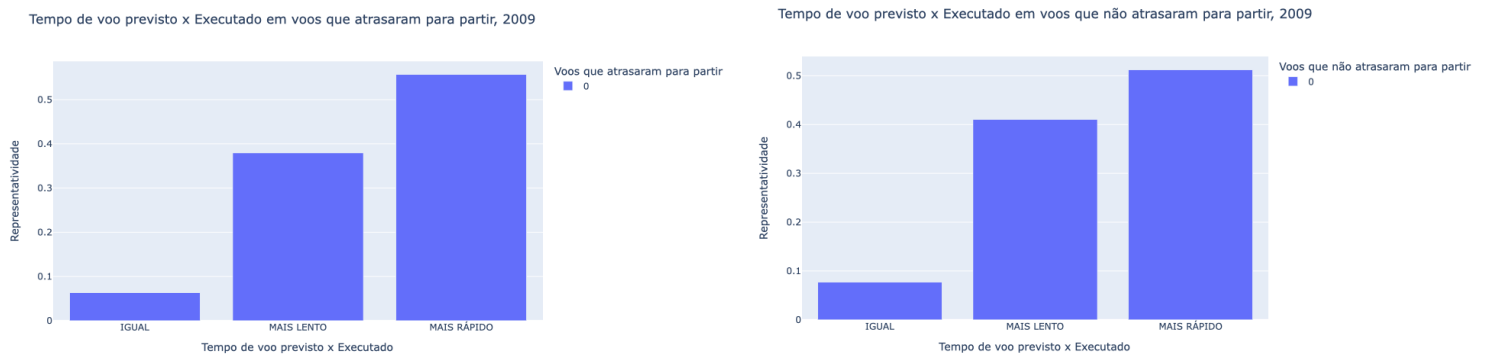


Figura 5: Relação entre tempo de voo e atraso na partida.

Nota-se que, naturalmente, os voos tendem a ser mais rápidos do que o planejado. Porém, essa métrica apresenta melhora no caso do avião atrasar para decolar. No cenário onde não há atraso na decolagem, os voos são mais rápidos que o previsto com frequência de 10 pontos percentuais a mais que os voos mais lentos. Já no caso de existir atraso para a decolagem, essa margem aumenta para aproximadamente 18 pontos percentuais. Um ganho de aproximadamente 80% na métrica.

Existem, portanto, fortes indícios de que o tempo perdido pelo atraso na decolagem seja compensado por um voo mais rápido do que o planejado.

Analisou-se também o mesmo caso no ano de 2018, o que pode ser visto mais abaixo no relatório.

## 1.4. Aeroportos com mais voos atrasados

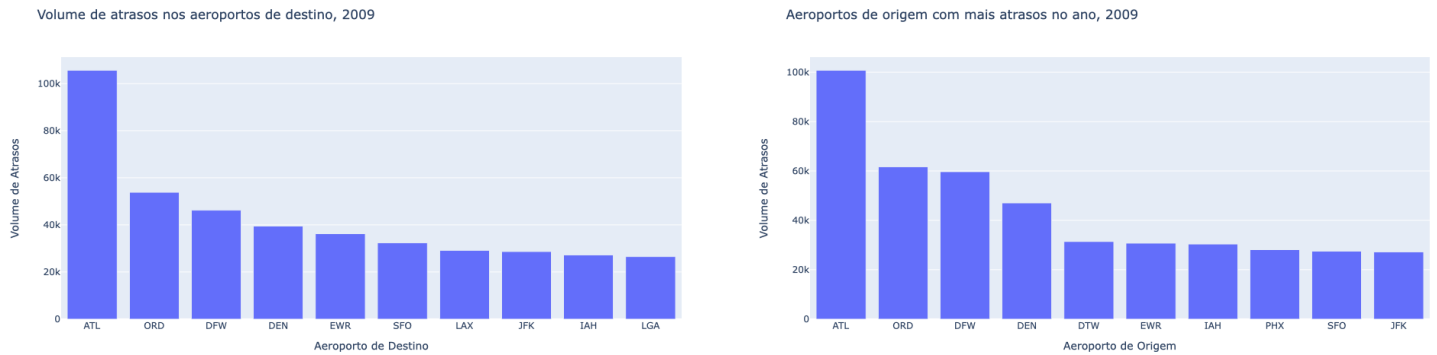


Figura 6: Atrasos em voos por aeroportos, de Origem e de Destino dos voos.

Na Figura 6 são apresentadas as siglas dos aeroportos presentes na base de dados em relação ao volume de atrasos absoluto, tanto em aeroportos de destino como de origem.

A partir desses gráficos é possível identificar o aeroporto de Atlanta (ATL) como local com o maior volume de atrasos absoluto em ambos os casos.

## 2. Atrasos de voos em 2018

### 2.1. Dado gerais do volume de atrasos ao longo do ano

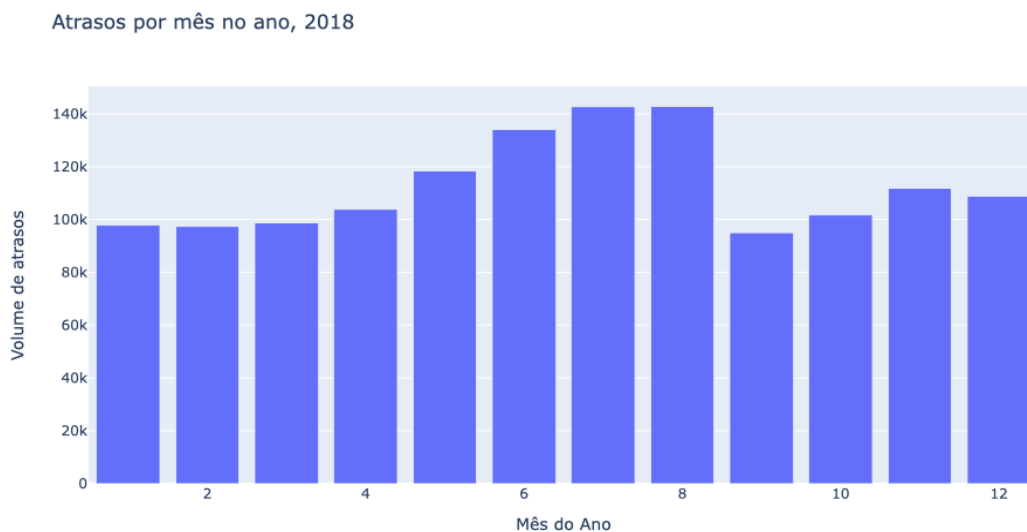


Figura 7: Volume de atraso por mês no ano de 2018.





Figura 8: Percentual de voos atrasados em cada mês de 2018.

É possível notar que existe um volume significativamente maior de atrasos nos meses de junho, julho e agosto. Este período corresponde às férias escolares e universitárias no hemisfério norte, o que pode ser uma possível justificativa para o aumento no volume de voos e um consequente aumento de atrasos. Para validar esta hipótese, seriam necessárias análises futuras.

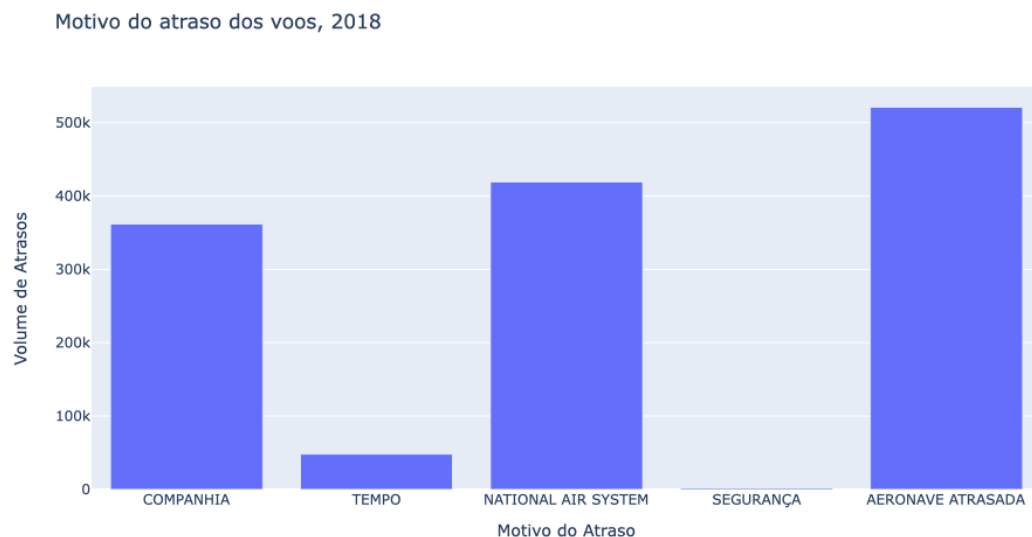


Figura 9: Motivos de atraso dos voos em 2018.

Na Figura 9 são identificadas as 3 principais causas de atraso dos voos em 2018. Sendo elas as mesmas apresentadas em 2009: Companhia aérea, *National Air System* (controle de tráfego aéreo) e aeronave atrasada. Assim, pode-se interpretá-las da mesma maneira: Essas razões contam com a participação humana e, consequentemente, estão

sujeitas a erros dos mesmos. Diferentemente, do tempo e da segurança, a qual apresenta melhorias muito significativas que dificultam eventuais infrações da mesma.

## 2.2. Companhias aéreas com mais voos atrasados

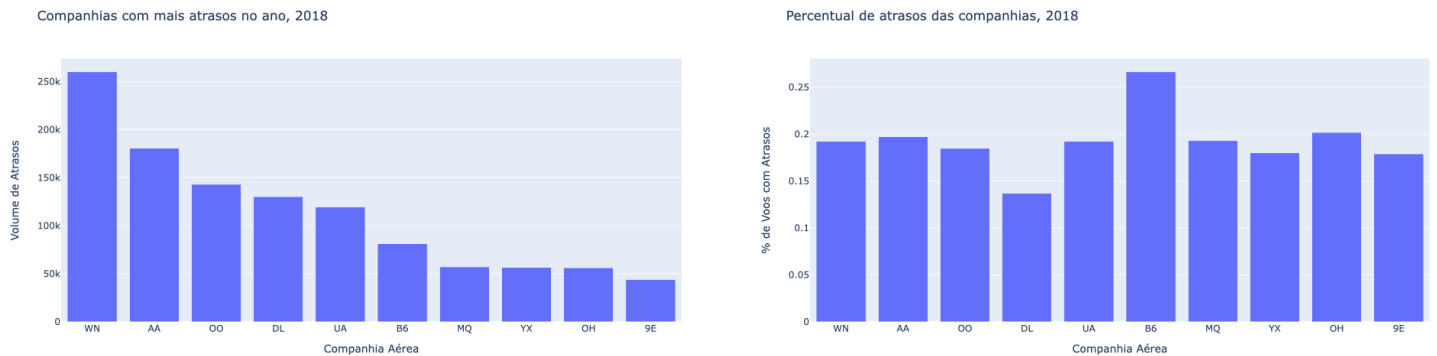


Figura 10: Atrasos em voos por companhias aéreas.

Na Figura 10 é possível visualizar mudanças no ranking das companhias aéreas que têm mais voos atrasados, dada a mudança no mercado de aviação ao longo dos anos. A ExpressJet por exemplo, com o nome EV no ano de 2009, estava em fim de operações em 2018, explicando a mudança no gráfico. Apesar disso, o topo do ranking permanece o mesmo.

Esse fato pode estar associado à baixas mudanças nos principais players do segmento, que já são extremamente consolidados. Outro ponto interessante é a entrada da United Airlines (UA) entre as empresas com maior volume de atrasos, o que pode também ser explicado pelo aumento no volume total de voos da companhia após a fusão com a United em 2015.

## 2.3. Relação entre tempo de voo e atraso na partida

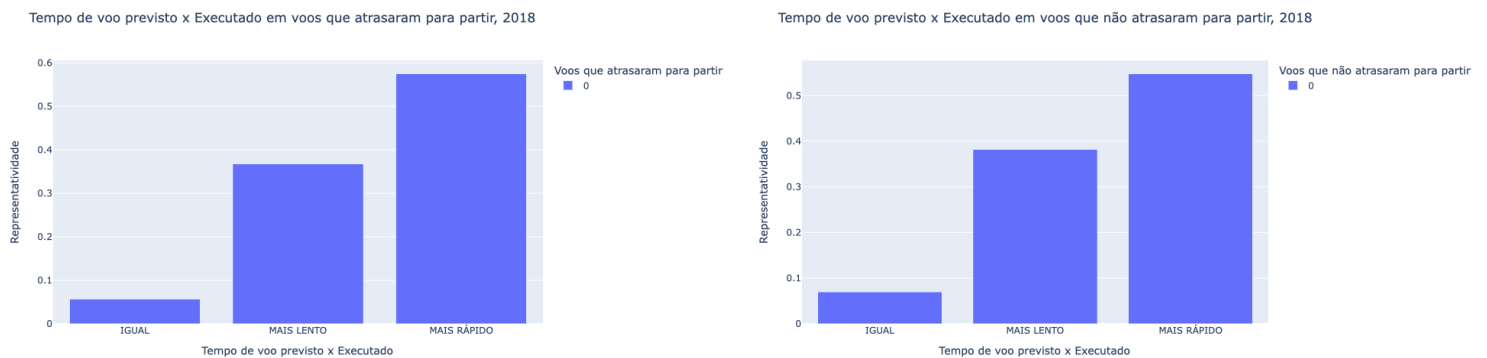


Figura 11: Relação entre tempo de voo e atraso na partida.

Assim como visto no ano de 2009, os voos que apresentaram atraso na partida tendem a ter menor tempo de voo do que o previsto. No ano de 2018, a diferença entre voos mais rápidos e mais lentos quando não há atraso na partida é de aproximadamente 15 pontos percentuais. Já no cenário onde há atraso na partida, a diferença fica na casa de 20 pontos percentuais. Apesar de diminuir em relação a 2009, a diferença entre os dois ainda é de consideráveis 33,33%.

## 2.4. Aeroportos com mais voos atrasados

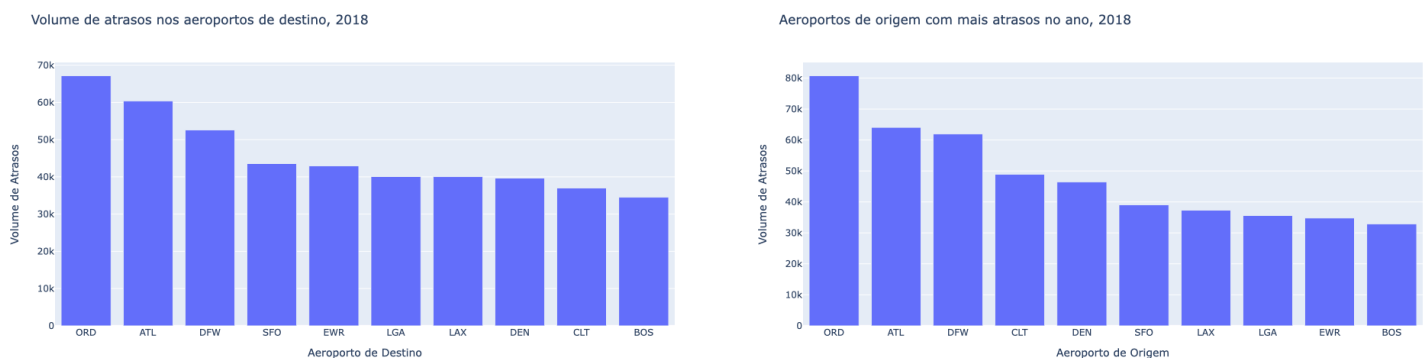


Figura 12: Atrasos em voos por aeroportos, de Origem e de Destino dos voos

O principal ponto notado na Figura 12 é a queda do volume de atrasos nas partidas do aeroporto de Atlanta (ATL), grande hub da aviação americana. Anteriormente, em 2009, na casa de 100 mil voos no ano, os atrasos caíram para em torno de 60 mil, apesar do aumento de voos que ocorre naturalmente com o passar dos anos. O mesmo ocorre com o aeroporto John Kennedy, de Nova York (JFK), que inclusive deixa de estar presente na lista de aeroportos com maior volume de atrasos.

Os dois aeroportos em questão são exemplos de locais que passaram por melhorias estruturais e de gestão, o que pode ter sido um fator contribuinte no avanço na operação, acarretando em um menor volume de atraso. Esse fator leva, assim, à uma melhoria na rentabilidade da operação do aeroporto.

## Conclusões

### Pontos semelhantes entre os dois anos

1. Foi observada uma sazonalidade dos atrasos no decorrer do ano. Em temporadas de alta como nos meses de julho, agosto e setembro, a quantidade de atrasos relativa do ano orbitaram entre 18% e 24% por mês. Isso se dá ao fato de que em muitos países do norte, a época de verão é caracterizada por férias escolares, fato que aumenta o volume de voos e consequentemente o de atrasos.

2. Existe uma relação entre o atraso na partida do voo e um menor tempo de voo. Em 2009, os voos que partiram com atraso mostraram um diferencial entre tempos de voos menores que o esperado e maiores que o esperado de 80%. Já em 2018, essa diferença foi de 33,33%.  
Apesar da diferença entre os anos, o padrão se manteve o mesmo: voos que decolam com atraso tendem a ter uma compensação deste tempo perdido durante o voo, com o tempo em ar sendo menor que o tempo esperado.
3. Além disso, existe uma similaridade entre os motivos de atraso de 2009 e 2018. Os principais 3 fatores se mantêm em ambos os anos: Companhia aérea, *National Air System* (controle de tráfego aéreo) e aeronave atrasada, razões que contam com a participação humana e, consequentemente, estão sujeitas a erros dos mesmos.

## Pontos diferentes entre os dois anos

1. As principais mudanças se deram nas listas de representantes com maiores volumes de atrasos, tanto no âmbito das companhias aéreas quanto no dos aeroportos. Isso se deve majoritariamente a mudanças de mercado, evoluções tecnológicas, e até mesmo a quebra de algumas companhias.  
Um exemplo seria o aeroporto de Atlanta, que, em 2009, atingiu a marca de mais de 100 mil atrasos. Mesmo com uma infraestrutura datada, presente desde 1980, o aeroporto já era um dos mais movimentados do mundo. Então, em 2016, foi criado um plano de expansão e modernização de mais de 6 bilhões de dólares. Com isso, o aeroporto reduziu o volume de atrasos no ano em aproximadamente 30 mil, enquanto aumentou o volume total de voos.

## Conclusão geral

Nota-se que, no geral, os comportamentos se mantiveram constantes ao longo dos anos, com a sazonalidade anual causando maior volume de atrasos nos meses de férias escolares, o tempo de voo sendo reduzido quando há atraso na decolagem, e as causas dos atrasos se mantendo similares. Estes fatores estão profundamente conectados às causas comportamentais, sendo difíceis de se modificarem com o passar do tempo.

No entanto, os pontos relacionados ao mercado, que é mais dinâmico, demonstraram variações nos comparativos entre o ano de 2009 e o ano de 2018. Consolidações de mercado causaram variações nas companhias com maiores representatividades nos atrasos de voos americanos, como a United Airlines por exemplo, que se aproveitou de uma fusão para expandir seu volume de viagens diárias, o que também acarretou no aumento dos atrasos. Além disso, melhorias nas operações dos aeroportos, bem como reformas que melhoraram suas estruturas, promoveram avanços significativos na eficácia da operação dos mesmos, o que pode ser visualizado na redução de voos com atrasos nos aeroportos de Atlanta e de Nova York.

# Repositório de análises

<https://github.com/VFermat/insper-2202-bigdata>

## Referências

Airline Delay and Cancellation Data, 2009-2018:

<https://www.kaggle.com/datasets/yuanyuwendymu/airline-delay-and-cancellation-data-2009-2018>

Expansão do aeroporto de Atlanta:

<https://www.ajc.com/news/atlanta-airport-blog/hartsfield-jacksons-master-plan-grows-in-price-complexity/UJ2XB7PYNNDIZHZUAP52YNNYCI/>

Situação dos aeroportos em 2009:

<https://www.airport-technology.com/analysis/feature72860/>

Flight delays cost \$32.9 billion, passengers foot half the bill:

<https://vcresearch.berkeley.edu/news/flight-delays-cost-329-billion-passengers-foot-half-bill>