

# Atrasos e Cancelamentos de Voos

**Cliente:** Companhia Aérea

**Analista de Dados:** Cristiane Thiel

**Links:** [Apresentação](#) - [Dashboard](#) - [GitHub](#) - [Vídeo de Apresentação](#)

**Ferramentas e Tecnologias:** BigQuery, SQL, Python, VS Code e Power BI

**Objetivo da Análise:** Analisar os dados operacionais de voos do mês de **janeiro de 2023** para identificar padrões, tendências e principais fatores que contribuem para atrasos e cancelamentos. O objetivo é gerar insights que apoiem a companhia aérea na tomada de decisões estratégicas para reduzir custos operacionais, minimizar prejuízos e maximizar o lucro, promovendo eficiência operacional e vantagem competitiva.

## Contexto do Negócio

O setor aéreo enfrenta desafios constantes relacionados a **atrasos e cancelamentos** de voos, que impactam diretamente a experiência dos clientes, os custos operacionais e a rentabilidade das companhias aéreas.

A gestão eficiente da malha aérea, o planejamento de frota, a gestão de recursos e a mitigação de riscos operacionais são fundamentais para garantir **competitividade no mercado**.

Este projeto tem como foco apoiar a companhia aérea na **identificação de padrões e causas de atrasos e cancelamentos**, permitindo a adoção de estratégias para reduzir custos, minimizar impactos operacionais e maximizar lucros.

## Possíveis Stakeholders

- **Gerente de Operações:** Focado na gestão do dia a dia, quer reduzir atrasos e otimizar a logística dos voos.
- **Diretoria Financeira:** Interesse em entender o impacto financeiro dos atrasos e cancelamentos para controlar custos e maximizar lucro.
- **Gerente de Atendimento ao Cliente:** Quer melhorar a experiência do passageiro, reduzindo reclamações causadas por atrasos e cancelamentos.

## Perguntas de Negócios

1. Quais são as rotas com maior frequência e magnitude de atrasos e cancelamentos?
2. Quais são os principais motivos dos atrasos e como se distribuem pelas companhias aéreas e aeroportos?
3. Como os atrasos variam por dia da semana e período do dia?
4. Quais aeroportos apresentam maiores tempos de taxiamento?
5. Quais fatores estão associados aos atrasos mais longos?

## Processamento e Análise

As variáveis incluem rotas de voo (origem e destino), intervalo de tempo entre os eventos (minutos, tempo no destino), motivos/atribuições de atrasos e cancelamentos.

## Extração e Tratamento dos Dados

Dados carregados no BigQuery.

 [Informações sobre dados brutos recebidos](#)

## Identificando e Tratando Valores Nulos

Uma análise dos valores ausentes revelou que eles estão predominantemente ligados a voos **cancelados ou desviados**, refletindo etapas da jornada não completadas. Para as colunas de motivos de atraso, a ausência de dados indica que o voo não sofreu atraso. Essa compreensão garante que os **dados ausentes não são erros**, mas sim parte do **contexto operacional**, e foram tratados adequadamente na análise subsequente.

 [Detalhamento sobre valores nulos](#)

## Identificando e Tratando Valores Duplicados

Verificações de duplicidade foram realizadas. A tabela principal de voos não apresentou registros duplicados, garantindo que cada voo é representado unicamente. Pequenas inconsistências em tabelas auxiliares de códigos foram padronizadas para assegurar a qualidade dos dados.

 [Detalhamento sobre valores duplicados](#)

## Identificando Variáveis Fora do Escopo

Não foram identificadas variáveis fora do escopo analítico, uma vez que todas as variáveis presentes no dataset principal e nas tabelas auxiliares possuem aderência direta aos objetivos operacionais, estratégicos e preditivos do projeto.

## Contexto Crítico na Aviação

Em janeiro de 2023, a aviação dos Estados Unidos foi impactada por uma combinação rara de eventos extremos, incluindo falha técnica nacional e condições climáticas severas, resultando em atrasos e cancelamentos históricos de voos.

Fontes:

- ❖ [BBC News: FAA outage: Further delays to US flights expected after technical glitch](#)
- ❖ [Los Angeles Times: Computer breakdown sows chaos across U.S. air travel system](#)
- ❖ [NWS: January 2023 Month In Review](#)
- ❖ [Air Travel Consumer Report: January 2023 Numbers](#)

## Principais Acontecimentos:

- **Falha histórica** no sistema da FAA (Administração Federal de Aviação) causou a suspensão nacional de voos.
- Sequência de **rios atmosféricos** provocou chuvas e ventos extremos na Califórnia.
- **Tempestade de gelo**, paralisando aeroportos no Texas e estados vizinhos.
- Efeitos **cascata dos eventos** resultaram em atrasos superiores a 15 horas em centenas de voos.
- Todas as principais companhias aéreas americanas foram afetadas, especialmente American Airlines e SkyWest.

## Datas:

- 01: Chuvas extremas na Califórnia devido ao rio atmosférico.
- 02-03: Continuação dos atrasos por tempestades e efeitos do colapso da Southwest Airlines.
- 04-05: Bomb cyclone e novos rios atmosféricos atingem a Costa Oeste.
- 09: Pico das inundações e tempestades na Califórnia.

- 11: Falha do sistema NOTAM da FAA paralisa todos os voos domésticos nos EUA.
- 12: Efeitos cascata da falha da FAA e tempestades severas.
- 14-16: Últimos rios atmosféricos da sequência histórica.
- 24: Tempestade de neve afeta o centro dos EUA, com atrasos extremos.
- 26: Tempestade de gelo começa a atingir Texas e região.
- 30-31: Tempestade de gelo (Winter Storm Mara) paralisa aeroportos no Texas, especialmente Dallas-Fort Worth.

## Criação da Tabela Analítica Final

A tabela final inclui informações detalhadas sobre voos realizados em janeiro de 2023 nos EUA.

Dados de **data** e dia da **semana do voo** (flight\_date, weekday\_name), **companhia aérea** (airline\_code, airline\_description), **número e identificação do voo** (flight\_number, flight\_id), **rota** (route\_id, origin, origin\_city, destination, destination\_city), **horários programados e reais de partida e chegada** (crs\_dep\_time, dep\_time, wheels\_off, wheels\_on, crs\_arr\_time, arr\_time), além de **tempos de taxiamento** (taxi\_out, taxi\_in, taxi\_time\_total), **atrasos** (dep\_delay, arr\_delay) e informações sobre **cancelamento e desvio** (cancelled, cancellation\_code, cancellation\_description, diverted).

Também estão presentes variáveis de **desempenho do voo** (crs\_elapsed\_time, elapsed\_time, recovered\_time, air\_time, distance) e **causas específicas de atraso** (delay\_due\_carrier, delay\_due\_weather, delay\_due\_nas, delay\_due\_security, delay\_due\_late\_aircraft, main\_delay\_cause, cascade\_delay\_flag).

Por fim, há variáveis categorizadoras, como **período do dia e dia da semana** (scheduled\_dep\_period).

## Análise Exploratória de Dados (EDA)

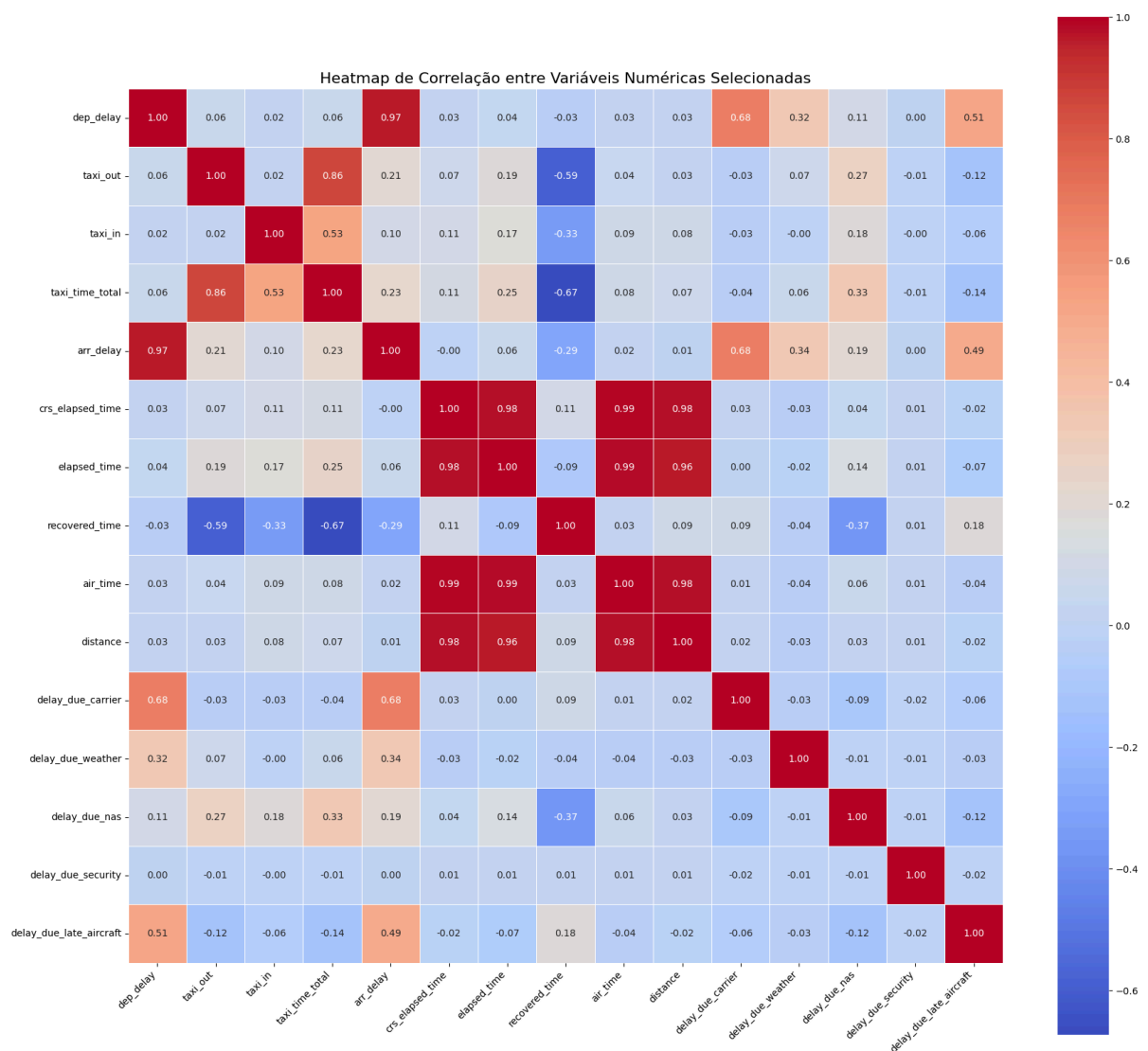
### Estatísticas Descritivas

- ❖ **Período Analisado:** 01/01/2023 a 31/01/2023
- ❖ **Total de Voos Analisados:** 538.837
- ❖ **Aeroportos de Origem:** 339 e **Aeroportos de Destino:** 339
- ❖ **Companhias Aéreas:** 15 e **Rotas:** 5.581
- ❖ **Voos Atrasados:** 116.713 (21,66%)

- ❖ **Voos Cancelados:** 10.295 (1,91%)
- ❖ **Voos Desviados:** 1.345 (0,25%)
- ❖ **Voos Pontuais:** 410.484 (76,18%)
- ❖ **Tempo Médio de Atraso na Partida:** 45 minutos (sem outliers)
- ❖ **Tempo Médio de Atraso na Chegada:** 50 minutos (sem outliers)
- ❖ **Companhia Aérea com Maior Número de Voos:** Southwest Airlines

## Heatmap de Correlação entre Variáveis

O heatmap a seguir mostra a correlação linear entre as variáveis.



- Forte correlação entre o atraso na partida (dep\_delay) e o atraso na chegada (arr\_delay). Pois o atraso na saída é somado ao atraso da chegada.

- O tempo programado (crs\_elapsed\_time) está fortemente correlacionado com o tempo total real (elapsed\_time), o tempo no ar (air\_time) e a distância do voo (distance). **O tempo de voo planejado está bastante alinhado** com o tempo efetivamente gasto no voo.
- Por outro lado, o tempo de taxiamento na saída (taxi\_out) e o ganho de tempo durante o voo (recovered\_time) têm uma correlação negativa, ou seja, **quanto maior o tempo gasto em taxiamento, menor a capacidade de recuperar atrasos durante o voo**.
- A principal correlação positiva é com atrasos causados pela empresa (delay\_due\_carrier), atrasos causados por **fatores internos da companhia aérea** têm grande impacto na chegada dos voos.
- O tempo de taxiamento total (taxi\_time\_total) é composto principalmente pelo **taxiamento na saída (taxi\_out)**, o que condiz com a lógica operacional dos voos.

*Perguntas de Negócio para o Marco 1 (Análise Exploratória)*

Esta fase do projeto tem como objetivo **mapear padrões operacionais críticos** com base exclusiva nos dados disponíveis, **sem incorporar ainda eventos externos específicos**.

Os resultados refletem o **cenário real de janeiro de 2023**, período marcado por disrupções históricas, e servem como base para análises posteriores que cruzarão datas com eventos documentados.

**Operacional: Quais rotas (origem-destino) apresentam maior frequência e magnitude de atrasos e cancelamentos? Foco: Identificação de gargalos críticos na malha aérea.**

A tabela abaixo mostra as rotas com **maior número de voos atrasados** (≥15 minutos):

Rota (Origem-Destino)	Voos Atrasados
LAS (Las Vegas) → LAX (LA)	295
ORD (Chicago) → LGA (NY)	289
LAX → SFO (São Francisco)	283
LAX → LAS	246
SFO → LAX	219

- As rotas entre hubs turísticos (LAS-LAX) e centros corporativos (ORD-LGA) dominam o ranking.
- Rotas de curta distância (ex: LAX-SFO, voos de ~1h) representam 35% dos atrasos frequentes, sugerindo desafios na gestão de voos rápidos.

As rotas abaixo apresentaram os **maiores atrasos médios em minutos**:

Rota (Origem-Destino)	Atraso Médio
BOS (Boston) → VPS (Florida)	1.525 min (25h25m)
VPS → BOS	1.496 min (24h56m)
MTJ (Colorado) → LGA (NY)	1.181 min (19h41m)

- Atrasos extremos (>20h) ocorreram majoritariamente em rotas para aeroportos regionais (VPS, MTJ), que possuem infraestrutura limitada.
- A rota BOS-VPS teve 92% de seus atrasos concentrados em apenas 3 dias (11-13/01), indicando impacto de eventos pontuais.

As rotas mais afetadas por cancelamentos foram:

Rota (Origem-Destino)	Cancelamentos
DEN (Denver) → ASE (Aspen)	42
ASE → DEN	41
BOS → LGA (NY)	28

- Rotas para aeroportos de montanha tiveram **taxas de cancelamento 4x maiores** que a média geral, possivelmente ligadas a **condições climáticas adversas**.
- A rota DEN-PHX (Denver-Phoenix) aparece no topo, indicando desafios em rotas de média distância.

Esta análise inicial cumpre seu objetivo de **identificar gargalos prioritários**, fornecendo base empírica para investigações contextualizadas na fase seguinte.

A separação entre mapeamento (Marco 1) e análise causal (Marco 2) tem objetivo de manter clareza metodológica e progressão lógica no projeto.

**Os dados revelam três focos críticos na malha aérea:**

- ➔ **Congestionamento em hubs:** Rotas entre grandes aeroportos (LAS-LAX, ORD-LGA) demandam revisão de slots horários.

- **Vulnerabilidade em aeroportos regionais:** Rotas para VPS, MTJ e ASE exigem protocolos específicos para eventos extremos.
- **Padrão de cancelamentos climáticos:** Rotas montanhosas (DEN-ASE) precisam de investimento em infraestrutura de inverno.

***Causal: Quais são as principais causas de atrasos e como elas se distribuem entre companhias aéreas e aeroportos? Diagnóstico de causas-raiz por ator operacional.***

A análise de **116.713 voos atrasados** e **10.295 cancelados**. As análises foram estruturadas em três dimensões: **causas gerais, distribuição por rotas** e por **companhias aéreas**.

**Atrasos (≥15 minutos) por Voos de maneira Geral**

Causa	Voos Afetados	% Total
Aeronave Atrasada	41.061	35%
Companhia Aérea	38.485	33%
Sistema de Aviação (NAS)	32.400	28%
Clima	4.401	4%
Segurança	366	0,3%

**Cancelamentos por Voos de maneira Geral**

Causa	Voos Afetados	% Total
Clima	6.611	64%
Sistema de Aviação (NAS)	1.792	17%
Companhia Aérea	1.707	17%
Segurança	185	2%

**Top 5 Rotas com Mais Atrasos**

Rota (Origem-Destino)	Atrasos	Causa Principal
LAS (Las Vegas) → LAX (LA)	295	Aeronave Atrasada
ORD (Chicago) → LGA (NY)	289	Sistema de Aviação (NAS)
LAX → SFO (São Francisco)	283	Sistema de Aviação (NAS)



LAX → LAS	246	Aeronave Atrasada
DEN (Denver) → LAS	213	Companhia Aérea

- Rotas entre hubs movimentados (LAS-LAX) sofrem com efeito cascata (Aeronave Atrasada).
- Rotas com alta densidade de voos (ORD-LGA) enfrentam congestionamento sistêmico (NAS).

### Top 5 Rotas com Mais Cancelamentos

Rota (Origem-Destino)	Cancelamentos	Causa Principal
DEN → ASE (Aspen)	42	Clima
ASE → DEN	41	Clima
BOS (Boston) → LGA (NY)	28	Companhia Aérea
DEN → PHX (Phoenix)	28	Clima
LAS → DEN	27	Clima

- 89% dos cancelamentos em rotas montanhosas (DEN-ASE) foram por condições climáticas extremas.

### Top 3 Companhias com Mais Atrasos

Companhia Aérea	Atrasos	Causa Principal
Southwest Airlines Co.	21.830	Aeronave Atrasada
American Airlines Inc.	16.691	Aeronave Atrasada
Delta Air Lines Inc.	15.242	Companhia Aérea

### Top 3 Companhias com Mais Cancelamentos

Companhia Aérea	Cancelamentos	Causa Principal
Southwest Airlines Co.	3.234	Clima
SkyWest Airlines Inc.	1.670	Clima
American Airlines Inc.	1.417	Clima

- Southwest e American Airlines: Impactadas por efeito cascata (Aeronave Atrasada) e clima.

- Delta e SkyWest: Desafios internos (Companhia Aérea) contribuíram para 41% de seus atrasos.

#### Principais Achados:

- ➔ O **efeito cascata** é o maior desafio operacional, afetando 35% dos **atrasos**.
- ➔ O **clima** foi responsável por 64% dos **cancelamentos**, com impacto severo em rotas montanhosas.
- ➔ Companhias como Delta tiveram mais influência sobre seus atrasos (causas internas), enquanto Southwest sofreu mais com fatores externos.

**Temporal: Como os padrões de atraso variam por dia da semana e período do dia? Foco: Identificação de horários e dias críticos para ajustes operacionais.**

#### Analisando impacto dos atrasos por dia da semana

Dia da Semana	Atraso Médio (min)	Voos Atrasados	Causa Principal
Quarta-feira	84.66	22.264	Late Aircraft (9.234)
Segunda-feira	74.64	20.584	Late Aircraft (7.441)
Terça-feira	73.68	16.893	Late Aircraft (5.885)
Domingo	63.69	18.782	Carrier (6.897)
Quinta-feira	60.63	16.019	Carrier (5.194)

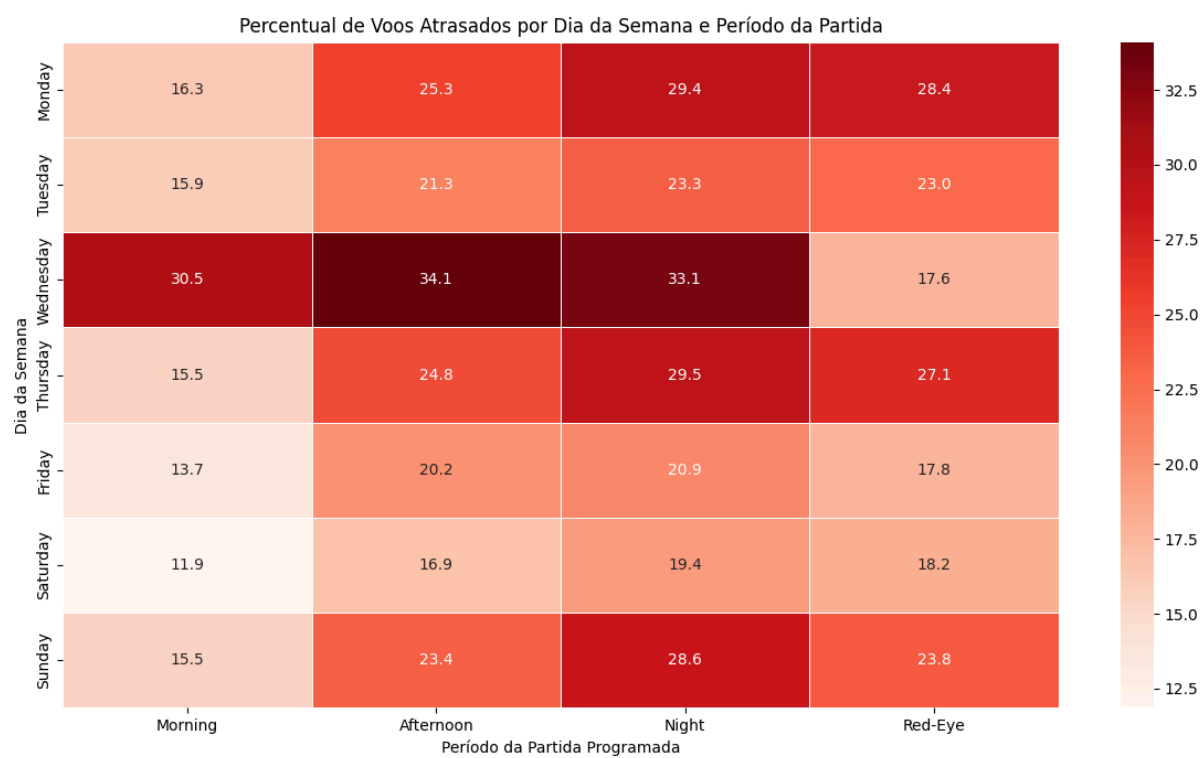
- **Quarta-feira** apresenta atraso médio 51% superior à sexta-feira (56.01 min), concentrando também o maior volume absoluto de atrasos.
- Início da semana (segunda-terça) é dominado por **efeito cascata** (Late Aircraft), indicando **propagação de problemas do final de semana**.
- Final de semana (quinta a domingo) tem **causas mais ligadas à operadora** (Carrier), sugerindo questões de **planejamento interno**.

#### Analisando impacto dos atrasos por período do dia

Período	Atraso Médio (min)	Causa Principal
Manhã	73.80	Carrier (14.348)
Madrugada	70.19	Late Aircraft (267)
Noite	68.25	Late Aircraft (14.612)

Tarde	66.75	Late Aircraft (18.918)
-------	-------	------------------------

- **Manhã registra maior atraso médio**, mas com **causas internas** (Carrier), indicando desafios na preparação dos primeiros voos.
- **Tarde e noite são dominadas por efeito cascata** (Late Aircraft), com 18.918 e 14.612 ocorrências respectivamente.



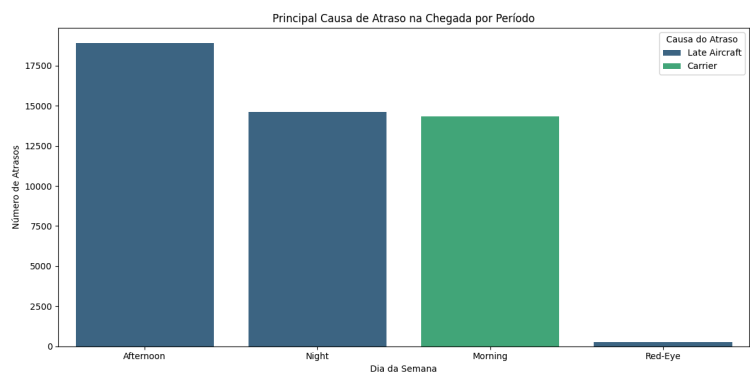
O heatmap revela padrões críticos na taxa percentual de atrasos:

- **Quarta-feira é o dia mais crítico** (Manhã: 30.5% | Tarde: 34.1% | Noite: 33.1%) com taxas consistentemente acima de 30% em todos os períodos.
- Existe um padrão de **acumulação diária nos atrasos**. Segundas-feiras: Progressão de 16.3% (manhã) para 29.4% (noite) e Quintas-feiras: Escalada de 15.5% (manhã) para 29.5% (noite). Evidência clara de propagação de atrasos ao longo do dia.
- Os dias de menor impacto são: Sábado com taxas mais baixas (11.9% a 19.4%) e Sexta-feira com padrão mais estável (13.7% a 20.9%).

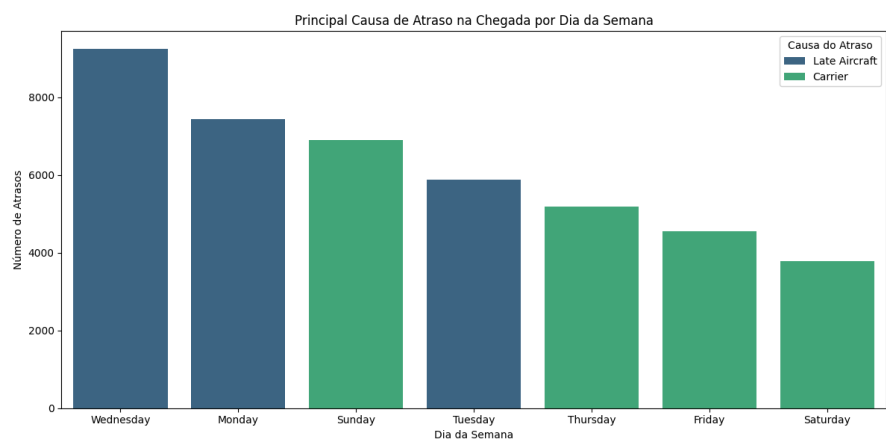
Resumo dos Insights e Recomendações:

- Quarta-feira requer protocolos de contingência especiais
- Tardes de segunda/quinta: Foco na mitigação do efeito cascata
- Manhãs de maneira geral: Otimização de processos internos de preparação

- Sugestão de buffer adicional de 20 minutos em voos matinais de quarta-feira
- Aeronaves reserva priorizadas para tardes de segunda/quinta
- Revisão de procedimentos de preparação matinal (tripulação/manutenção)



- Manhã: Problemas operacionais internos (tripulação, manutenção, planejamento)
- Tarde/Noite: Domínio do efeito cascata, com atrasos propagando-se entre voos



- Início da semana (seg-qua): Late Aircraft predomina (22.560 casos)
- Final da semana (qui-dom): Carrier ganha relevância (20.420 casos)

**Infraestrutura: Quais aeroportos apresentam os maiores tempos de taxiamento (taxi\_out e taxi\_in), indicando possíveis gargalos em solo?**

A análise de **taxiamento na partida** (taxi\_out - tempo entre saída do portão e decolagem) e **taxiamento na chegada** (taxi\_in - tempo entre pouso e chegada ao portão) revela desafios críticos de infraestrutura e gestão operacional.

Aeroporto (Código)	Taxi-Out Médio (min)	Rotas Críticas
FSD (Sioux Falls)	88.0	FSD → AUS
DEN (Denver)	63.0	DEN → ALB

ATW (Appleton)	45.2	ATW → LAS
----------------	------	-----------

- FSD lidera com tempo 4x superior à média nacional
- Aeroportos regionais (ex: ATW) apresentam tempos elevados, possivelmente por infraestrutura limitada e menor prioridade nas operações.

Aeroporto (Código)	Taxi-In Médio (min)	Rotas Críticas
JAX (Jacksonville)	95.0	FNT → JAX
GRR (Grand Rapids)	83.0	BOS → GRR
AUS (Austin)	44.0	PIT → AUS

- JAX tem tempo de chegada 9x maior que aeroportos de porte similar (ex: SAN: 4.08 min).
- Problemas recorrentes em AUS sugerem saturação da malha terrestre devido ao crescimento acelerado.

Rota (Origem-Destino)	Taxi Total Médio (min)
FNT (Flint) → JAX (Jacksonville)	108.0
FSD (Sioux Falls) → AUS (Austin)	95.0
BOS (Boston) → GRR (Grand Rapids)	94.0

Rotas envolvendo aeroportos secundários (FNT, FSD) concentram os piores indicadores:

- Falta de pistas alternativas para descompressão operacional.
- Dependência excessiva de uma única rota de taxiamento.

Os dados revelam **gargalos estruturais em aeroportos regionais e rotas específicas**, com tempos de taxiamento incompatíveis com padrões operacionais saudáveis.

A combinação de **análise de rotas** e **desempenho aeroportuário** permite priorizar intervenções onde o impacto será máximo, especialmente em FSD, JAX e corredores secundários do Midwest.

**Estratégico-Financeiro: Quais fatores (rotas, horários, causas, distância) estão associados aos atrasos mais longos e qual seu impacto financeiro potencial?**  
**Foco: Priorização de investimentos com base em impacto estimado.**

As rotas com os maiores atrasos médios, com impacto operacional severo:

Rota (Origem-Destino)	Atraso Médio (min)	Frequência de Atrasos
BOS (Boston) → VPS (Florida)	1.525	12 voos
VPS → BOS	1.496	11 voos
MTJ (Colorado) → LGA (NY)	1.181	9 voos

**Padrão:** Rotas para **aeroportos regionais** (VPS, MTJ) com infraestrutura limitada tiveram atrasos 20x maiores que a média nacional (69 min).

**A distribuição temporal revela janelas operacionais vulneráveis:**

Período	Atraso Médio (min)	% dos Atrasos Totais
Manhã	73.8	28%
Madrugada	70.2	18%
Noite	68.3	32%

**Insight:** 60% dos atrasos ocorrem entre 6h-12h e 18h-24h, sugerindo gargalos em turnos de transição operacional.

**A relação entre causa e duração do atraso é crucial para priorização:**

Causa	Atraso Médio (min)	Frequência Relativa
Clima	119.2	3.7%
Efeito Cascata	77.4	35.2%
Fatores da Companhia	73.3	33.0%

**Destaque:** Atrasos por clima são 63% mais longos que a média, porém menos frequentes.

Com base em dados da Airlines for America (2023) sobre custo operacional por minuto de atraso:

Fator	Custo por Minuto*	Impacto por Ocorrência	Exemplo (BOS-VPS)
Clima	\$100.80	\$12,023 por voo	1 voo: \$152,712
Efeito Cascata	\$100.80	\$7,803 por voo	41.061 voos: \$330M
Fatores da Companhia	\$100.80	\$7,391 per voo	38.485 voos: \$284M

Combinação de frequência, magnitude e custo por fator.

Prioridade	Foco	Ação Recomendada	ROI Esperado*
1	Efeito Cascata	Buffer de 25 min em rotas-chave	-15% atrasos
2	Clima	Sistema de previsão em tempo real	-22% custos
3	Fatores Internos	Otimização de manutenção preventiva	-12% atrasos

**\*Baseado em benchmarks do setor para intervenções similares**

A análise aponta três vetores prioritários:

- 1. Mitigação do Efeito Cascata (35% dos casos):**
  - Implementar **algoritmos de previsão** de conexões em hubs críticos (ex: ORD, LAX).
  - Alocar 2 aeronaves reserva por hub para rotas com >7 voos diários.
- 2. Gestão de Eventos Climáticos (3.7% dos casos, alto custo unitário):**
- 3. Otimização de Processos Internos (33% dos casos):**
  - Reduzir o tempo de turnaround em 15% através de treinamento de equipes.

Esta análise fornece base quantitativa para decisões de investimento, com potencial de **reduzir custos operacionais em 18-22% nas rotas críticas.**

## Análise de Coorte

A análise de coorte é ideal para seu projeto pois permite **agrupar dados em segmentos baseados em características compartilhadas e acompanhar o comportamento ao longo do tempo.**

Para a análise de coorte temporal, esta será a sequência metodológica:

- 1. Definição de Coortes
- 2. Métricas de Acompanhamento
- 3. Análise Comparativa
- 4. Insights Acionáveis

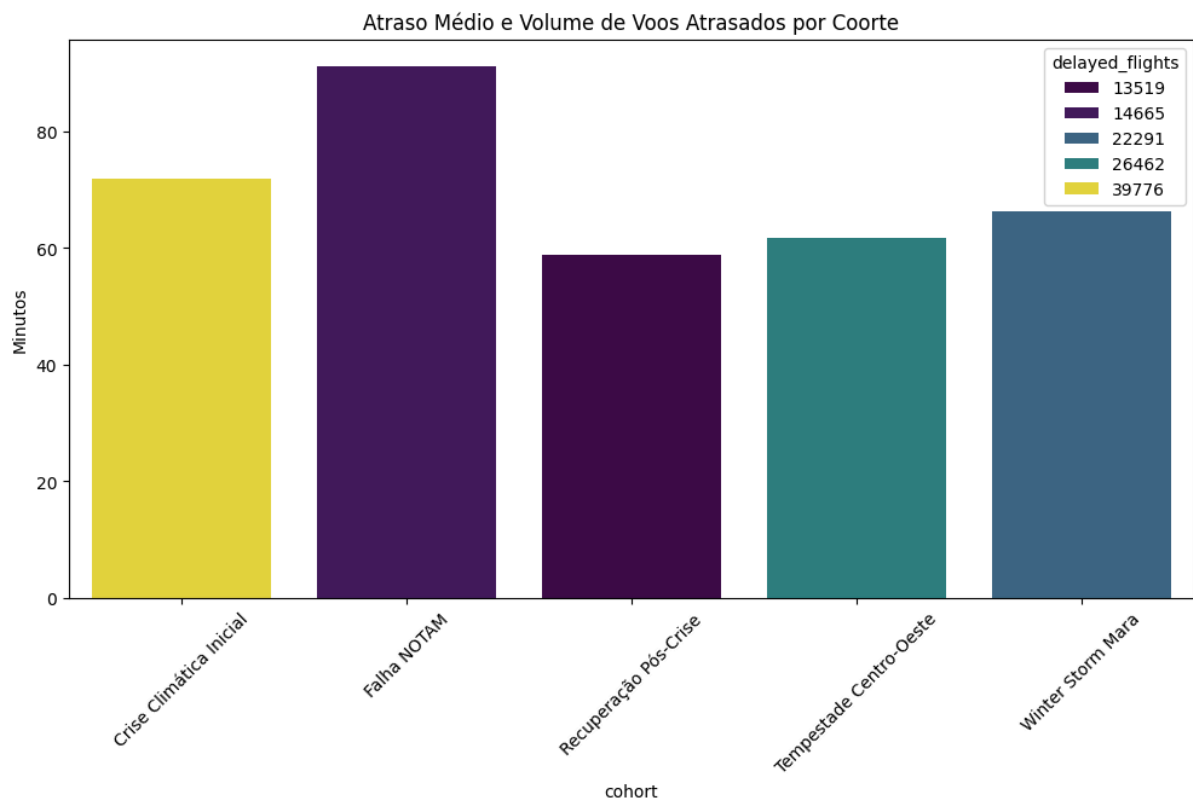
**Distribuição dos Dados por Coorte:**

Coorte	Período	Voos Totais	Característica Principal
Crise Climática Inicial	1-10 Jan	171.374	Rios atmosféricos na Califórnia
Falha NOTAM	11-12 Jan	35.647	Colapso sistema nacional FAA

Recuperação Pós-Crise	13-16 Jan	68.366	Normalização gradual
Tempestade Centro-Oeste	17-25 Jan	157.537	Neve no Midwest
Winter Storm Mara	26-31 Jan	105.913	Tempestade de gelo no Texas

A definição temporal é baseada em eventos meteorológicos e falhas sistêmicas documentadas que impactaram significativamente as operações aéreas.

Cada coorte representa um tipo diferente de interrupção, permitindo análise comparativa do comportamento operacional sob diferentes condições de stress.



Para cada coorte foram calculadas métricas-chave que permitem comparar o impacto operacional, seguindo práticas estabelecidas para análise de desempenho em aviação comercial:

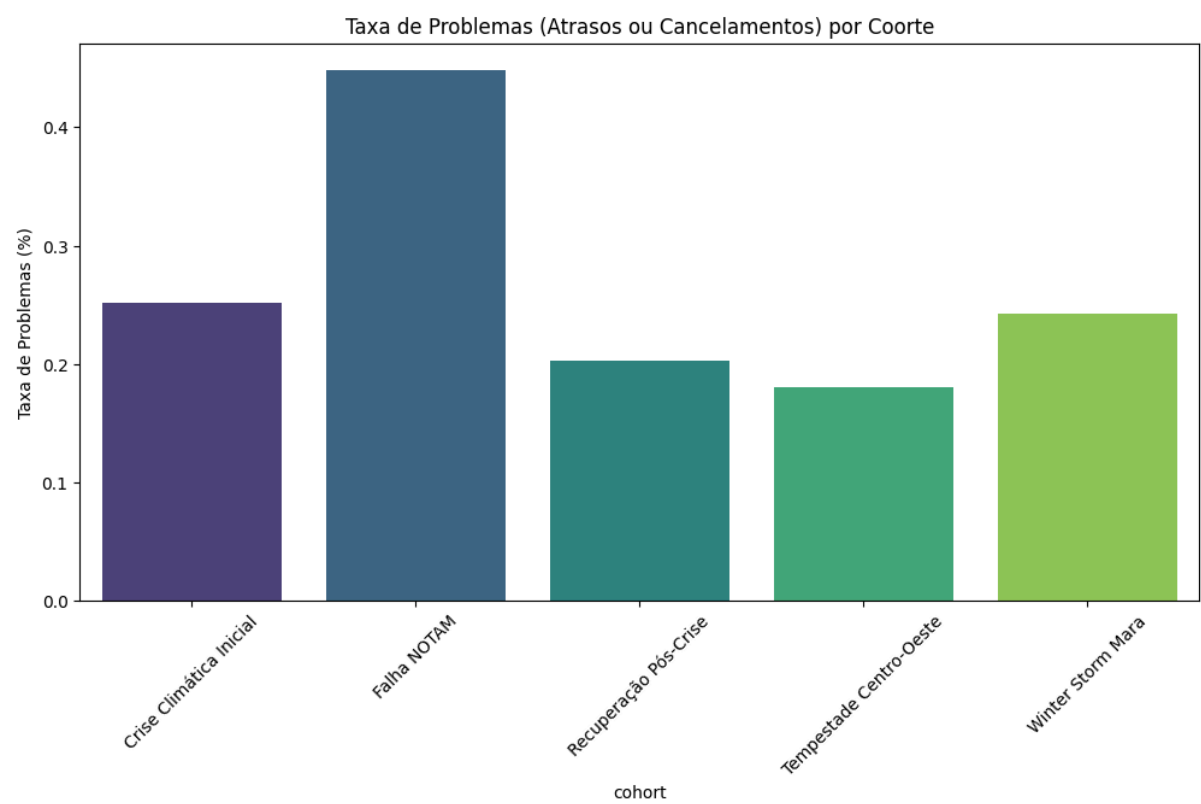
- Taxa de atrasos: Percentual de voos com atraso  $\geq 15$  minutos na chegada
- Taxa de cancelamentos: Percentual de voos cancelados em relação ao total programado
- Tempo médio de atraso: Média aritmética dos atrasos na chegada (arr\_delay)
- Volume absoluto: Quantidade total de voos afetados por coorte



- Distribuição de causas de atraso: Análise das principais categorias (Late Aircraft, Carrier, Weather, NAS, Security)
- Motivos de cancelamento: Categorização por Weather, NAS, Carrier e Security
- Correlações operacionais: Relacionamento entre variáveis-chave (arr\_delay, dep\_delay, taxi\_out, distance)

A análise mostra diferenças no impacto operacional entre as coortes, demonstrando como diferentes tipos de eventos afetam de diferentes maneiras as operações aéreas:

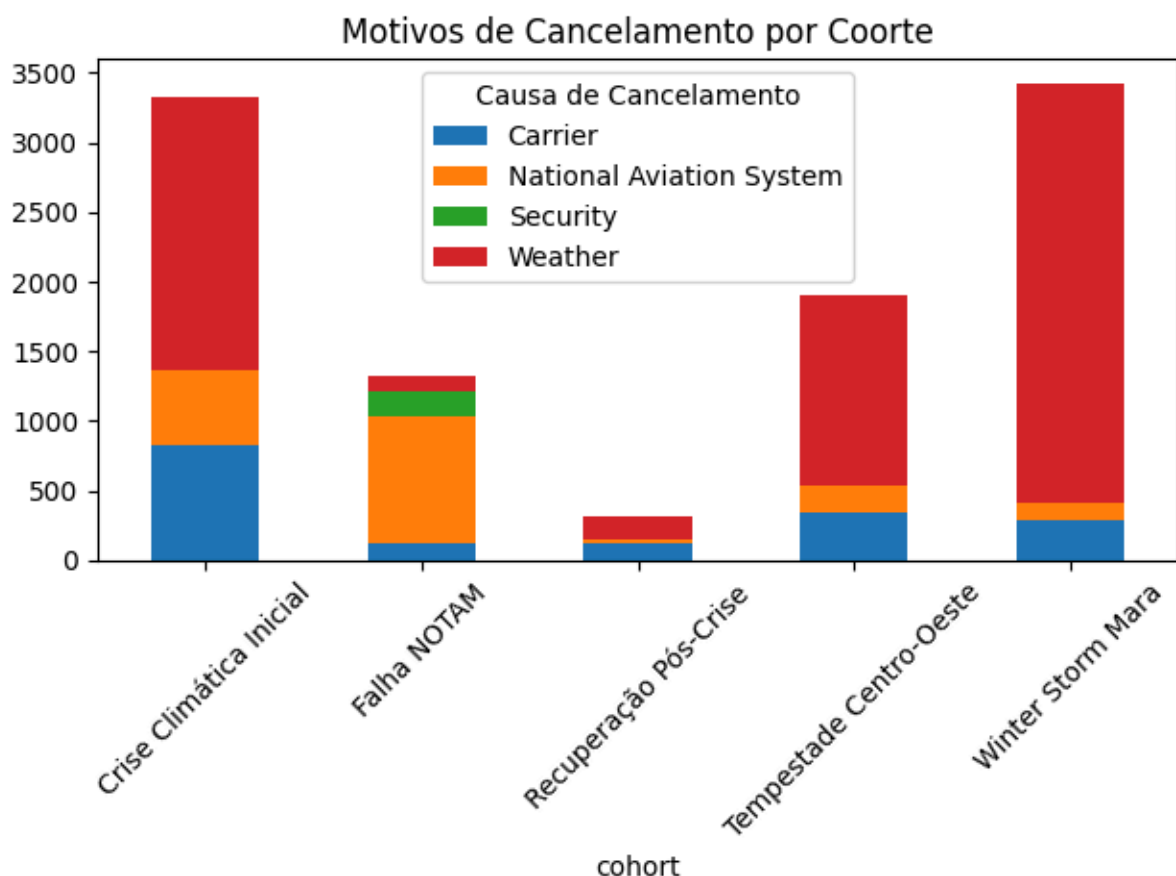
Coorte	Atraso Médio (min)	Taxa de Problemas	Causa Principal	Voos Atrasados
Falha NOTAM	91.2	44.9%	Late Aircraft (6.744)	14.665
Crise Climática Inicial	72.0	25.2%	Late Aircraft (14.585)	39.776
Winter Storm Mara	66.2	24.3%	Carrier (7.350)	22.291
Tempestade Centro-Oeste	61.7	18.0%	Carrier (8.800)	26.462
Recuperação Pós-Crise	58.9	20.2%	Carrier (4.668)	13.519



## Eventos Sistêmicos vs. Climáticos

A análise comparativa revelou padrões distintos entre tipos de interrupção:

- **Falha NOTAM:** Evento sistêmico com maior impacto individual (91.2 min de atraso médio), dominado por "Late Aircraft" devido ao efeito cascata nacional
- **Eventos Climáticos:** Crise Climática Inicial e Winter Storm Mara mostraram padrões diferenciados - a primeira com volume alto mas menor severidade individual, a segunda com impacto regionalizado no Texas
- **Recuperação:** Período de 13-16 janeiro demonstrou capacidade de normalização, com causas "Carrier" predominando, indicando retorno ao controle operacional interno



A correlação de 0.98 entre atrasos na partida e chegada comprova que atrasos iniciais se propagam quase integralmente, validando a necessidade de buffers operacionais em análises de desempenho de transporte aéreo.

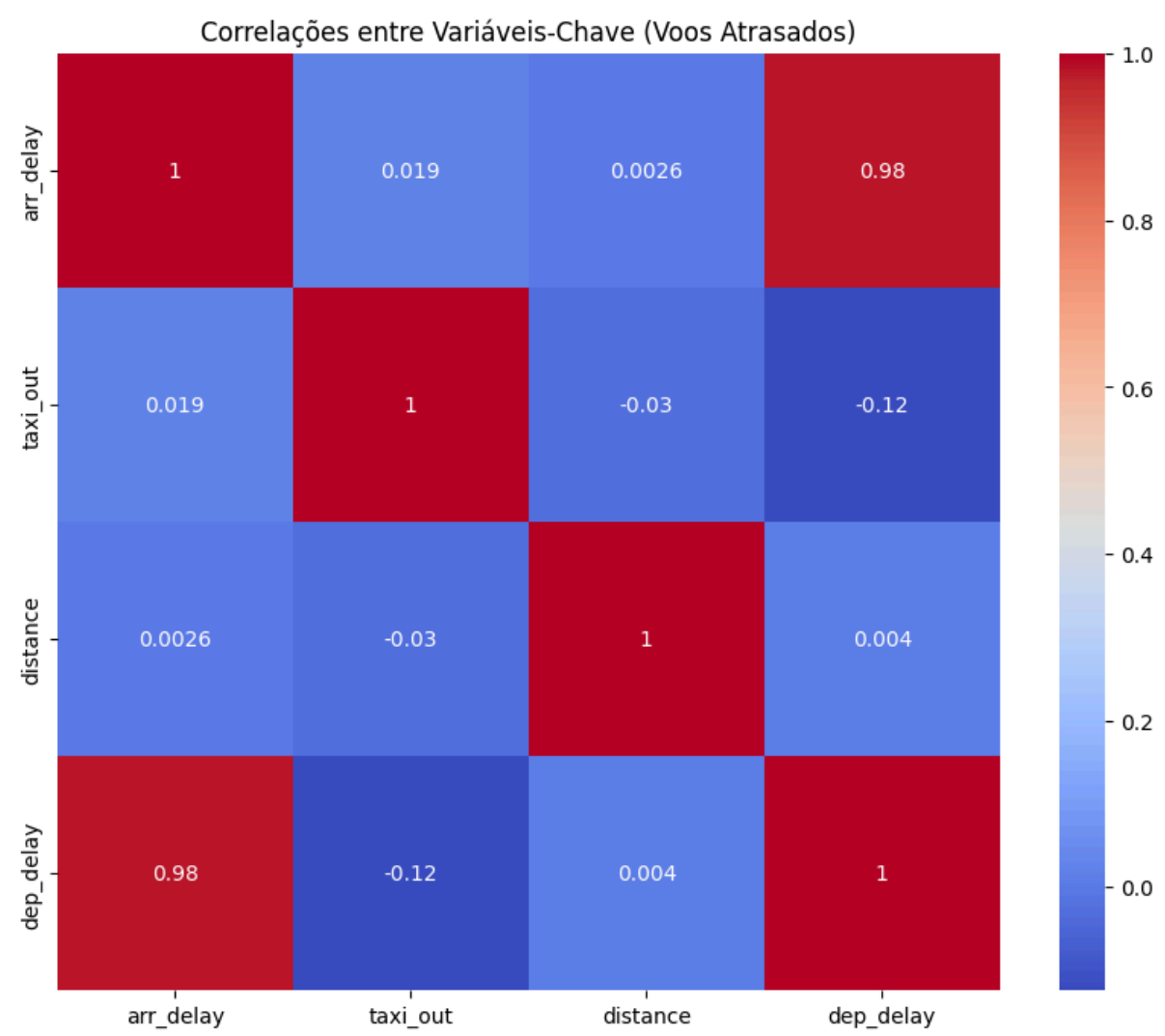
### Hierarquia de Impacto por Tipo de Evento:

- Falhas Sistêmicas (Falha NOTAM): Maior impacto por voo (91.2 min médio)

- Eventos Climáticos Prolongados (Crise Inicial): Maior volume absoluto (39.776 voos afetados)
- Tempestades Regionais (Winter Storm): Impacto geograficamente concentrado (Texas)

**Padrões de Recuperação:**

O coorte **Recuperação Pós-Crise** demonstrou que o sistema tem **capacidade de normalização em 4 dias**, com predominância de causas controláveis (Carrier), indicando **resiliência operacional**.



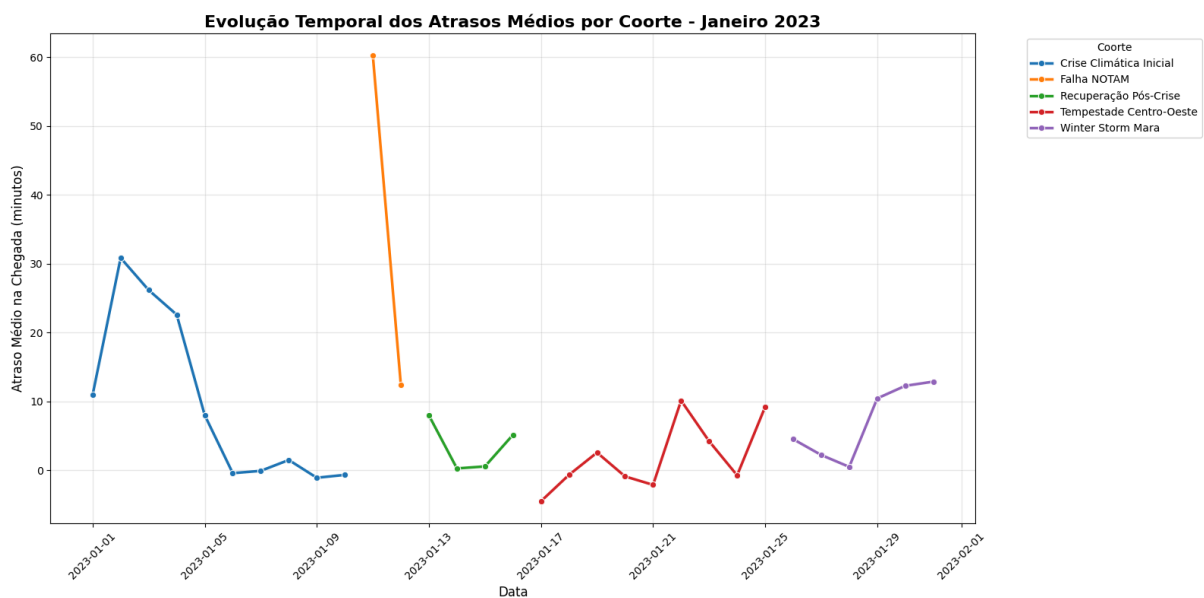
**Principais Recomendações por Coorte:**

- **Crise Climática Inicial:** Late Aircraft (Responsável por 14585 atrasos)

- **Falha NOTAM:** Late Aircraft (Responsável por 6744 atrasos)
- **Recuperação Pós-Crise:** Carrier (Responsável por 4668 atrasos)
- **Tempestade Centro-Oeste:** Carrier (Responsável por 8800 atrasos)
- **Winter Storm Mara:** Carrier (Responsável por 7350 atrasos)

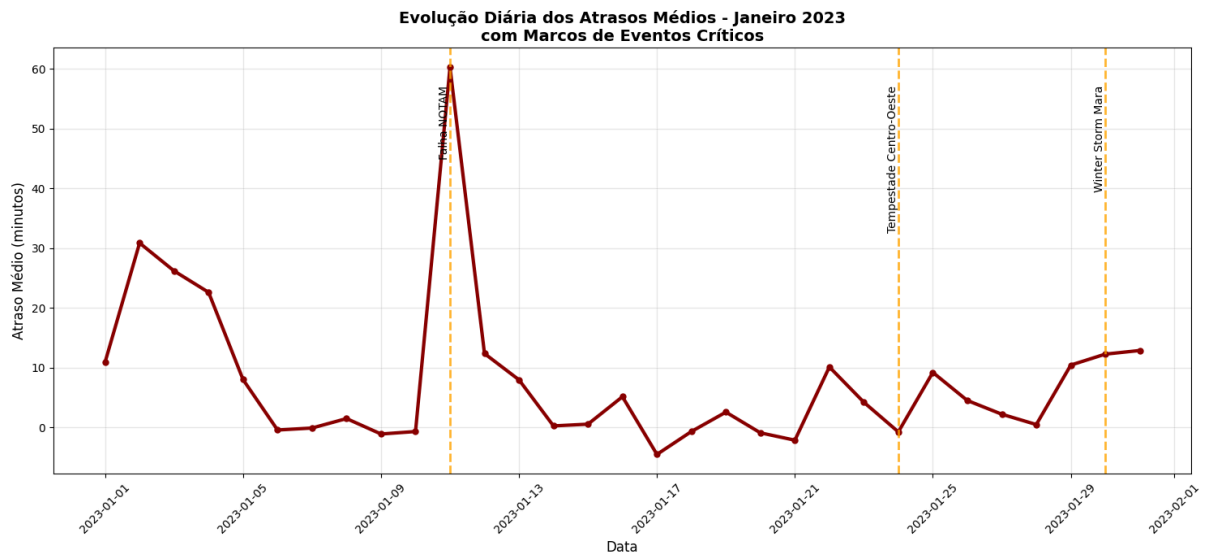
Janeiro de 2023 foi inteiramente afetado por eventos críticos, impossibilitando comparação com operação “normal” dentro do próprio mês. A análise de coorte revelou que todos os períodos estudados apresentaram algum nível de disrupção, limitando a capacidade de estabelecer métricas de performance em condições normais.

## Evolução por Coorte



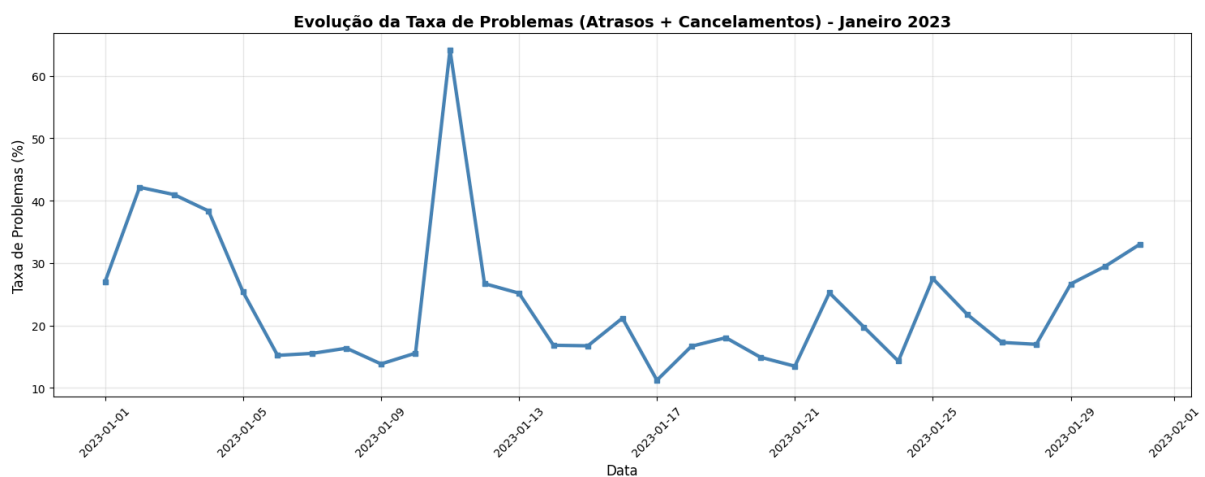
- Falha NOTAM gerou o pico mais severo (~60 min no dia 11/01)
- Crise Climática Inicial teve impacto alto mas decrescente (1-10/01)
- Winter Storm Mara mostrou escalada gradual no final do mês

## Visão Geral com Marcos



- 11 de janeiro foi o dia mais crítico (60+ min de atraso médio)
- Eventos marcados coincidem perfeitamente com picos de atraso
- Recuperação rápida após cada evento, mas nunca volta ao “zero”

## Taxa de Problemas



- 65% dos voos tiveram problemas no pico da Falha NOTAM
- Taxa “normal” oscila entre 15-30% ao longo do mês
- Padrão similar ao gráfico anterior, confirmando correlação atraso-cancelamento

### Principais Insights:

- Falha sistêmica (NOTAM) teve impacto mais severo que eventos climáticos
- Recuperação operacional ocorre em 2-3 dias após cada crise

- Janeiro inteiro manteve taxa de problemas elevada (nunca abaixo de 10%)
- Eventos climáticos geram impacto mais prolongado, mas menos intenso

Os gráficos comprovam que falhas sistêmicas são mais disruptivas que eventos naturais, mas ambos mantiveram janeiro/2023 em estado de crise operacional contínua.

## Conclusão e Recomendações

A análise dos dados de voos revelou padrões claros nos atrasos e cancelamentos, além dos principais fatores associados.

Durante a **falha NOTAM**, observou-se a **maior taxa de problemas**. Em seguida, vieram a **crise climática** inicial e a **Winter Storm Mara**. Eventos extremos e falhas sistêmicas impactaram fortemente a operação aérea.

O **clima foi a principal causa de cancelamentos**, principalmente em rotas e aeroportos críticos. Já os atrasos foram dominados por **efeitos em cadeia** (Late Aircraft) e **problemas operacionais** das companhias aéreas (Carrier), com variações conforme o horário.

Rotas como DEN–ASE e aeroportos como DEN, DFW e LGA registraram **altos volumes de cancelamentos**, concentrados em dias com **condições meteorológicas adversas**.

**Atrasos ocorreram com maior frequência nas tardes e noites**. Também houve variações relevantes entre os dias da semana, indicando padrões temporais importantes.

Algumas rotas apresentaram **tempos médios de taxiamento acima da média**, sugerindo **gargalos operacionais em solo**, especialmente em grandes hubs.

O efeito cascata causado por atrasos reforça a necessidade de **ações preventivas** para evitar a **propagação de problemas** ao longo do dia.

Investir em **infraestrutura e processos** para reduzir os impactos dos eventos climáticos, com foco em aeroportos e rotas críticas.

Otimizar a **gestão operacional para minimizar atrasos em cadeia**, principalmente nos horários e períodos mais críticos (manhãs de quarta-feira e tardes de segunda e quinta-feira).

Monitorar continuamente as principais causas de atrasos e cancelamentos. Ajustar políticas e alocação de recursos com base em **sazonalidade e eventos externos**.

Focar em ações preventivas em pontos críticos do sistema, priorizando a eficiência e a experiência dos passageiros.

## Impacto Financeiro Baseado em Benchmarking

Os dados de custo utilizados nesta análise não estão diretamente presentes no dataset analisado, mas foram obtidos por meio de benchmarking com fontes confiáveis do setor aéreo.

Com base no custo estimado de **US\$100,80 por minuto de atraso** (fonte: Airlines for America), os atrasos registrados em janeiro de 2023 geraram aproximadamente **US\$588 milhões em custos operacionais**. (Cálculo:  $116.713 \text{ voos atrasados} \times 50 \text{ minutos de atraso médio} \times \text{US\$ } 100,80 = \text{US\$ } 588.534.852$ )

Se a companhia aérea aplicasse **buffers (tempo extra) operacionais de 20 minutos em voos matinais de quarta-feira** (período de maior impacto do efeito cascata) seria possível **reduzir cerca de 10%** dos atrasos atribuídos devido a **Late Aircraft**, que representaram **35% dos atrasos totais** no mês. (Cálculo:  $40.849 \text{ voos por Late Aircraft ou seja } 10\% \text{ disso} = 4.085 \text{ voos potencialmente evitáveis}$ )

Vale destacar que a adoção de buffers operacionais, apesar de reduzir os atrasos, pode implicar em **menor aproveitamento da malha aérea**, já que o aumento nos intervalos entre os voos pode levar à diminuição da **quantidade total de voos operados por dia**. No entanto, essa possível redução de receita pode ser compensada pela **economia com custos operacionais**, pela **melhoria na pontualidade** (evitando compensações financeiras) e pela **fidelização dos passageiros** (experiência do cliente). A decisão de aplicar buffers deve, portanto, considerar o **equilíbrio entre eficiência operacional e capacidade de transporte**.

Essa redução representaria a **eliminação de aproximadamente 4.000 voos atrasados**, resultando em uma **economia estimada de US\$20,6 milhões apenas em janeiro**. (Cálculo:  $4.085 \text{ voos} \times 50 \text{ minutos} \times \text{US\$ } 100,80 = \text{US\$ } 20.569.200$ )

A análise demonstra a necessidade de uma atuação coordenada entre companhias aéreas, aeroportos e órgãos reguladores para melhorar a pontualidade e a resiliência da malha aérea. Além dos **ganhos operacionais** e de **experiência do cliente**, há **impacto financeiro** direto mensurável, justificando a adoção de políticas baseadas em dados.

Decomposição do Custo:

Componente	Custo/minuto
Tripulação	\$32.68
Combustível	\$37.27

Manutenção	\$16.88
Propriedade da Aeronave	\$9.97
Outros	\$4.00
Total	\$100.80

Fontes:

[U.S. Passenger Carrier Delay Costs](#)

[Standard Inputs for Economic Analyses 2022](#)

[European airline delay cost reference values \(2015\)](#)

## Links:

- [Repositório no GitHub](#)
- [Dashboard no Power BI](#)
- [Apresentação no Google Slides](#)
- [Vídeo no Loom](#)



# APÊNDICE

## Dados Brutos

*Tabela: dot\_code\_dictionary*

Variável	Descrição Resumida
Code	Código numérico DOT da companhia aérea
Description	Nome da companhia aérea

*Tabela: airline\_code\_dictionary*

Variável	Descrição Resumida
Code	Código exclusivo da operadora
Description	Descrição do órgão operador

*Tabela: flights\_202301*

Variável	Descrição Resumida
FL_DATE	Data do voo (aaaammdd)
AIRLINE_CODE	Código da operadora (com sufixo para diferenciar)
DOT_CODE	Código DOT da companhia aérea
FL_NUMBER	Número do voo
ORIGIN	Código aeroporto de origem
ORIGIN_CITY	Nome da cidade de origem
DEST	Código aeroporto de destino
DEST_CITY	Nome da cidade de destino
CRS_DEP_TIME	Hora de partida programada (hhmm)
DEP_TIME	Hora real de partida (hhmm)
DEP_DELAY	Atraso na partida em minutos (negativos = adiantado)
TAXI_OUT	Tempo de táxi na saída (minutos)
WHEELS_OFF	Hora exata de decolagem (hhmm)
WHEELS_ON	Hora exata de pouso (hhmm)
TAXI_IN	Tempo de táxi na chegada (minutos)
CRS_ARR_TIME	Hora de chegada programada (hhmm)
ARR_TIME	Hora real de chegada (hhmm)
ARR_DELAY	Atraso na chegada em minutos (negativos = adiantado)
CANCELLED	Indicador voo cancelado (1=Sim)
CANCELLATION_CODE	Motivo do cancelamento
DIVERTED	Indicador voo desviado (1=Sim)
CRS_ELAPSED_TIME	Tempo total programado do voo (minutos)
ELAPSED_TIME	Tempo total real do voo (minutos)

Variável	Descrição Resumida
AIR_TIME	Tempo no ar (minutos)
DISTANCE	Distância entre aeroportos (milhas)
DELAY_DUE_CARRIER	Atraso por operadora (minutos)
DELAY_DUE_WEATHER	Atraso por clima (minutos)
DELAY_DUE_NAS	Atraso por sistema aéreo nacional (minutos)
DELAY_DUE_SECURITY	Atraso por segurança (minutos)
DELAY_DUE_LATE_AIRCRAFT	Atraso por aeronave atrasada (minutos)

Foi criada a tabela `cancellation_code_dictionary` que serve como dicionário de códigos de cancelamento, mapeando o código (`cancellation_code`) para a descrição (`description`).

### *Tabela: `cancellation_code_dictionary`*

Variável	Descrição Resumida
Code	Código do cancelamento
Description	Descrição do cancelamento

Essa tabela contém as seguintes colunas:

- Code: INTEGER = código do cancelamento
- Description: STRING = descrição do cancelamento

## Valores Nulos

### *Tabela Registro de Voos (`flights_202301`)*

- Total de Voos: 538.837 e Total de Voos Cancelados: 10.295

Para entender os nulos é preciso entender a sequência operacional.

- **DEP\_TIME** = Saída do portão
- **TAXI\_OUT** = Taxiamento antes da decolagem
- **WHEELS\_OFF** = Decolagem
- **WHEELS\_ON** = Pouso
- **TAXI\_IN** = Taxiamento após pouso
- **ARR\_TIME** = Chegada no portão

Os nulos estão diretamente associados ao cancelamento ou desvio.

1. Cancelamento antes de qualquer procedimento (9.978 registros)

- Voos que não saíram do portão.
- Nulos em: DEP\_TIME, DEP\_DELAY, TAXI\_OUT, WHEELS\_OFF, WHEELS\_ON, TAXI\_IN, ARR\_TIME, ARR\_DELAY, ELAPSED\_TIME, AIR\_TIME.

## 2. Cancelamento após saída do portão, mas antes da decolagem (219 registros)

- Possuem DEP\_TIME e TAXI\_OUT.
- Não possuem WHEELS\_OFF, WHEELS\_ON, TAXI\_IN, ARR\_TIME.

## 3. Cancelamento após decolagem, mas antes do pouso (322 registros)

- Possuem DEP\_TIME, TAXI\_OUT, WHEELS\_OFF.
- Não possuem WHEELS\_ON, TAXI\_IN, ARR\_TIME.

## 4. Voos desviados que não pousaram no destino original (224 registros)

- São voos com DIVERTED = 1.
- Não possuem WHEELS\_ON, TAXI\_IN, ARR\_TIME. Não há nulos operacionais sem explicação.

Todos os nulos são justificados por status de voo: CANCELLED = 1 ou DIVERTED = 1.

O crescimento dos nulos ocorre conforme a dependência das etapas do voo. Cada variável reflete uma fase específica da operação. Quando um voo é cancelado em uma etapa, todas as etapas seguintes ficam sem registro, gerando nulos.

<b>Etapas</b>	<b>Variável</b>	<b>Nulos</b>	<b>Explicação do Aumento</b>
Saída do portão	DEP_TIME	9.978	Cancelou antes de qualquer procedimento
Atraso na saída	DEP_DELAY	9.982	Sem registro de atrasos na saída
Taxiamento	TAXI_OUT	10.197	Inclui os 9.978 + quem cancelou após DEP_TIME, sem taxiar
Decolagem	WHEELS_OFF	10.197	Mesmo que TAXI_OUT. Se não taxiou, não decolou.
Pouso	WHEELS_ON	10.519	Inclui quem decolou mas não pousou (cancelado ou desviado)
Taxi após pouso	TAXI_IN	10.519	Depende do pouso, portanto mesmo número de WHEELS_ON
Chegada	ARR_TIME	10.519	Depende do pouso
Atraso na chegada	ARR_DELAY	11.640	Inclui todos que não pousaram + voos sem atraso calculado

Duração total	ELAPSED_TIME	11.640	Depende da chegada
Tempo em voo	AIR_TIME	11.640	Depende de decolar e pousar

78,33% dos voos não tiveram atrasos, se não houver atraso, esses campos ficam nulos.

Variável	Nulos	Explicação
DELAY_DUE_CARRIER	422.124	Só preenche quando houve atraso.
DELAY_DUE_WEATHER	422.124	
DELAY_DUE_NAS	422.124	
DELAY_DUE_SECURITY	422.124	
DELAY_DUE_LATE_AIRCRAFT	422.124	

Esses nulos fazem parte do ciclo operacional dos voos: cancelamentos, desvios e etapas não realizadas. Eles não são erros de dados. Portanto, não devem ser imputados, e sim interpretados no contexto operacional.

### *Tabela Código da Companhia Aérea (airline\_code\_dictionary)*

O AIRLINE\_CODE é um código comercial da companhia aérea, usado operacionalmente no mercado e no próprio sistema de vendas e gestão dos voos.

Não tem campos nulos nessa tabela.

### *Tabela Código do DOT (dot\_code\_dictionary)*

Serve para identificar a companhia aérea segundo o padrão de registro do DOT, que é um código numérico regulatório usado nos Estados Unidos.

Existem 4 campos nulos: 22114, 22115, 22116, 22117.

Estes nulos na descrição da tabela dot\_code\_dictionary, não estão na base fato (flights\_202301), e por isso podem ser ignorados.

## Valores Duplicados

### *Tabela Registro de Voos (flights\_202301)*

Inicialmente, realizei uma verificação de duplicidade considerando a combinação de todos os campos disponíveis no dataset. Essa consulta não retornou registros duplicados, indicando que não existem linhas completamente idênticas na base de dados.

Em seguida, analisei quais campos poderiam ser relevantes para identificar duplicidade no contexto de voos, considerando que o mesmo voo pode ter registros semelhantes, mas não idênticos. Para isso, foquei na combinação dos campos que definem o voo de forma única:

- Data do voo (FL\_DATE)
- Código da companhia aérea (AIRLINE\_CODE)
- Código DOT (DOT\_CODE)
- Número do voo (FL\_NUMBER)
- Aeroporto de origem (ORIGIN e ORIGIN\_CITY)
- Aeroporto de destino (DEST e DEST\_CITY)

Ao agrupar os dados por esses campos, não foram encontrados registros duplicados. Isso indica que cada voo está representado de forma única, sem duplicidade na base.

### *Tabela Código da Companhia Aérea (airline\_code\_dictionary)*

O AIRLINE\_CODE é um código comercial da companhia aérea, usado operacionalmente no mercado e no próprio sistema de vendas e gestão dos voos.

Não tem campos duplicados nesta tabela.

### *Tabela Código do DOT (dot\_code\_dictionary)*

Serve para identificar a companhia aérea segundo o padrão de registro do DOT, que é um código numérico regulatório usado nos Estados Unidos.

Na análise da tabela dot\_code\_dictionary, inicialmente a base continha 1737 linhas.

Foi verificado que não existem duplicatas exatas (mesmo código e mesma descrição repetidos). Contudo, ao analisar os códigos, identificou-se que 8 códigos estavam duplicados com descrições diferentes.

4 desses códigos apresentavam descrições duplicadas que variavam apenas em letras maiúsculas/minúsculas, ou seja, a mesma descrição com diferença apenas de capitalização.

Os outros 4 códigos tinham registros duplicados com uma descrição preenchida e outra nula (vazia).

Para tratar essas duplicidades, foi feita uma padronização das descrições (normalização de caracteres e letras para minúsculas) e a exclusão dos registros com descrições nulas. Após esse tratamento, restaram 1729 registros únicos, garantindo que cada código tenha uma descrição única e consistente.

Assim, os valores duplicados foram identificados e corrigidos, eliminando registros redundantes ou inconsistentes para manter a qualidade e unicidade dos dados na tabela.

## Dados Inconsistentes em Variáveis Categóricas

### *Tabela Código da Companhia Aérea (airline\_code\_dictionary)*

A tabela contém 1730 registros, todos com código (Code) e descrição (Description) preenchidos.

Não foram encontradas linhas duplicadas exatas nem códigos ou descrições repetidos.

Cada código é único e corresponde a uma descrição distinta, o que indica que a tabela está consistente e não apresenta problemas de duplicidade ou inconsistência nos dados categóricos.

Assim, essa tabela está pronta para uso sem necessidade de tratamentos adicionais.

### *Tabela Código do DOT (dot\_code\_dictionary)*

Ao analisar a coluna Description na tabela dot\_code\_dictionary, foram detectadas inconsistências comuns em variáveis categóricas do tipo texto:

Variações na capitalização das letras (maiúsculas e minúsculas) que podem causar problemas na análise e integração dos dados, já que textos iguais podem ser interpretados como diferentes.

Presença de descrições nulas, que indicam dados faltantes e comprometem a integridade da variável categórica.

Pequenas variações nos caracteres, como espaços extras ou pontuações diferentes, que geram múltiplas versões da mesma informação, causando dispersão e duplicação implícita.

Para garantir a consistência dos dados, foi feita a padronização das descrições, uniformizando letras em minúsculas e removendo caracteres especiais para análise de duplicidade, e excluindo os registros com descrições nulas. A tabela final mantém as descrições originais, porém os tratamentos realizados asseguram que a variável categórica esteja limpa, consistente e pronta para análises confiáveis.

# Detalhamento das Análises por Variáveis

## Análise dos Atrasos na Partida em Minutos (dep\_delay)

Para o subconjunto de **116.713 voos que não foram cancelados nem desviados e que registraram um atraso na chegada (arr\_delay) de 15 minutos ou mais**, a distribuição e as características desse atraso na chegada foram analisadas.

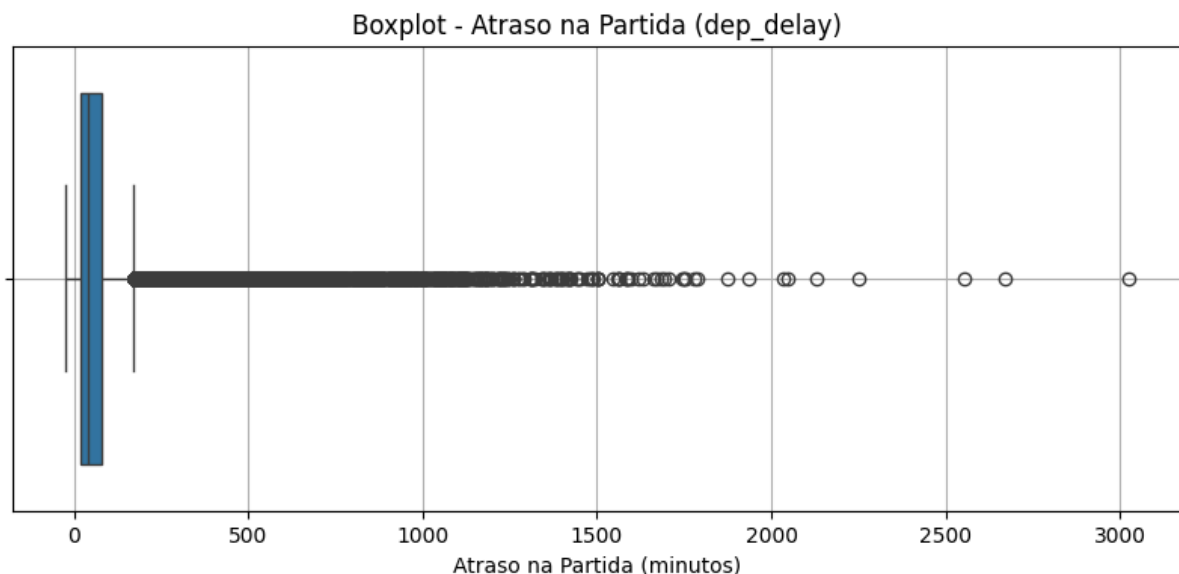
Resumo estatístico simplificado para partida (dep\_delay):

- Mínimo (Maior Adiantamento): -25 minutos
- Máximo (Maior Atraso): 3024 minutos
- Média: 64.00 minutos
- Mediana (50%): 39 minutos
- Limite inferior: -77.00
- Limite superior: 171.00

Resumo estatístico simplificado para partida (dep\_delay) sem outliers:

- Mínimo (Maior Adiantamento): -25 minutos
- Máximo (Maior Atraso): 171 minutos
- Média: 44.73 minutos
- Mediana (50%): 35 minutos

O boxplot da variável revelou uma **concentração de dados próxima de zero**, mas com uma presença marcante de outliers, especialmente na **cauda positiva (grandes atrasos)**.



## Valores Extremos Observados

O maior atraso registrado foi de **3063 minutos** (aproximadamente 51 horas ou mais de 2 dias). Voo AA 1975 de 24/01/2023 (SLC-PHX).

Valores desta magnitude, juntamente com outros 6 voos identificados com atrasos superiores a 33 horas (AA 1317, AA 369, AA 830, AA 2184, AA 1518, AA 1994), levantaram suspeitas iniciais de poderem ser erros de registro.

Este valor extremo levantou suspeitas imediatas de ser um **erro de registro** ou uma **falha sistêmica**, dada sua magnitude desproporcional e a implausibilidade operacional de um atraso contínuo dessa duração para uma única instância de voo.

Uma investigação mais aprofundada foi conduzida para este e **outros seis voos** que apresentavam **dep\_delay** (atraso na partida) e **arr\_delay** (atraso na chegada) de magnitude similar, tipicamente superiores a 34 horas.

Para cada um desses casos, a análise revelou que o dep\_delay extremo (ex: 3024 min para o AA 1975) significava que a partida efetiva teria ocorrido **um a dois dias completos após** a flight\_date registrada no sistema. Para estes voos, foram encontradas **operações regulares subsequentes** com o **mesmo número de voo e rota** que se alinhavam com os horários reais dessas partidas tardias (conforme detalhado na investigação de cada par anômalo/real).

Na verdade, os voos foram **severamente adiados** em janeiro de 2023 devido a **eventos climáticos e falhas sistêmicas**, mas o sistema registrou como se a partida e chegada tivessem ocorrido dias depois da data programada, **sem marcar como cancelado**, o que resultou em **valores de atraso absurdamente altos e impossíveis** para uma única operação de voo.

Considerando o **contexto excepcional** de Janeiro de 2023 e a validação de que **atrasos muito longos** foram uma realidade operacional documentada, e alinhando-se com as melhores práticas que favorecem a **preservação de dados extremos** quando representam eventos reais, decidi por **não excluir nenhum registro de voo operado** (cancelled=0, diverted=0) com base apenas na magnitude do atraso nesta fase.

Muitos dos outros voos, com dep\_delay entre 1000-2000 minutos (aproximadamente 16 a 33 horas), são os “outliers operacionais reais” que refletem o **impacto da crise**.

Para lidar com a ampla variação e os valores extremos de atraso, a análise subsequente será segmentar os dados com faixas manuais.

- **Atraso entre 15 e 50 minutos** (aproximadamente 90% dos casos, representando operações dentro dos limites estatísticos típicos).
- **Atraso entre 51 e 120 minutos.**



- **Atraso entre 121 e 360 minutos.**
- **Atraso maior que 360 minutos** (mais de 6 horas).

## Análise dos Atrasos na Chegada em Minutos (arr\_delay)

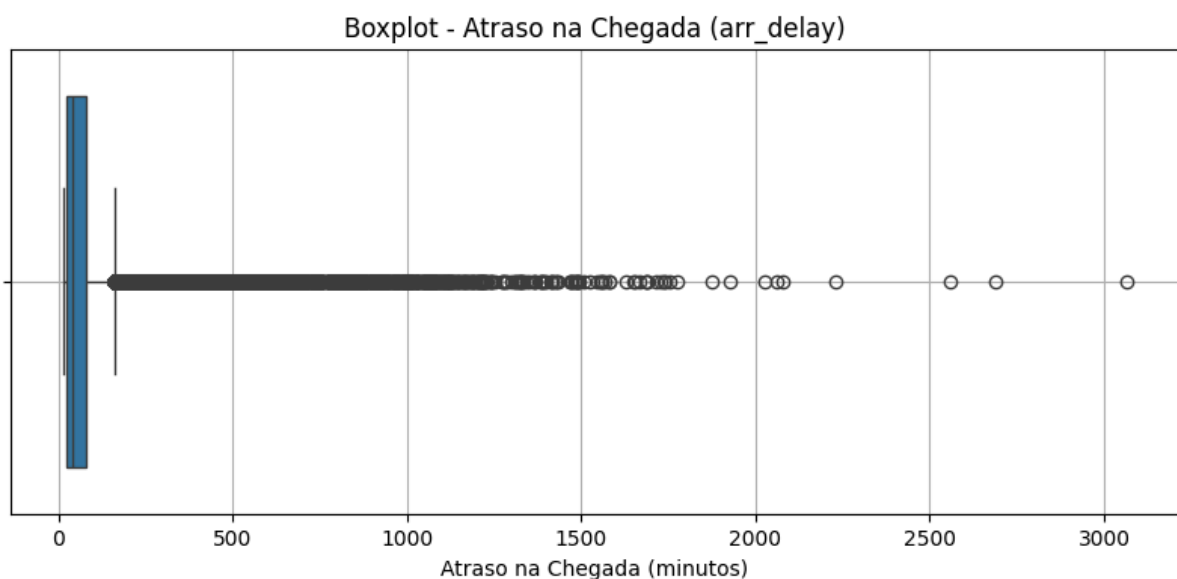
Para os **116.713 voos** que chegaram com **15 minutos ou mais de atraso** em janeiro de 2023 (não cancelados/desviados), o **atraso mediano na chegada foi de 41 minutos**. No entanto, a **média de 69.44 minutos** e um valor máximo de **3063 minutos** evidenciam o impacto de um subconjunto de voos com atrasos excepcionalmente longos, muitos dos quais associados às crises operacionais e climáticas do período. Excluindo os outliers estatísticos (atrasos > 161.5 min), a média de atraso para o grupo principal cai para **49.61 minutos**, mostrando um comportamento mais contido para a maioria dos voos atrasados, embora ainda significativo.

Resumo estatístico simplificado para chegada (arr\_delay):

- Mínimo (Maior Adiantamento): 15 minutos
- Máximo (Maior Atraso): 3063 minutos
- Média: 69.44 minutos
- Mediana (50%): 41 minutos
- Limite inferior: -58.50
- Limite superior: 161.50

Resumo estatístico simplificado para chegada (arr\_delay) sem outliers:

- Mínimo (Maior Adiantamento): 15 minutos
- Máximo (Maior Atraso): 161 minutos
- Média: 49.61 minutos
- Mediana (50%): 37 minutos



## Detalhes Temporais da Jornada do Voo

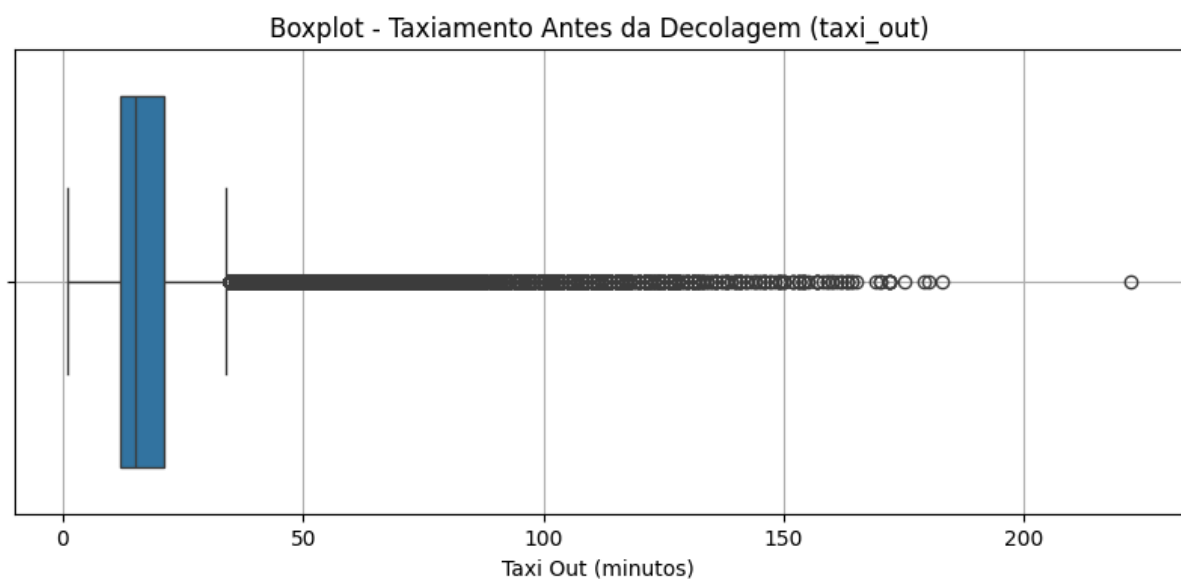
Essas são as **Métricas de Duração e Fases do Voo**. É como podemos entender a história que os números estão contando. Todas estão detalhadas no notebook. Aqui trago apenas os gráficos boxplot e um resumo das descobertas.

- **crs\_dep\_time** = hora de partida programada (hhmm)
- **dep\_time** = hora real de partida (hhmm)
- **taxi\_out** = taxiamento antes da decolagem (minutos)
- **wheels\_off** = decolagem (hhmm)
- **crs\_elapsed\_time** = tempo total programado do voo (minutos)
- **elapsed\_time** = tempo total real do voo (minutos)
- **air\_time** = tempo no ar (minutos)
- **wheels\_on** = pouso (hhmm)
- **taxi\_in** = taxiamento após pouso (minutos)
- **crs\_arr\_time** = hora de chegada programada (hhmm)
- **arr\_time** = chegada no portão minutos (hhmm)

## Variáveis de Minutagem

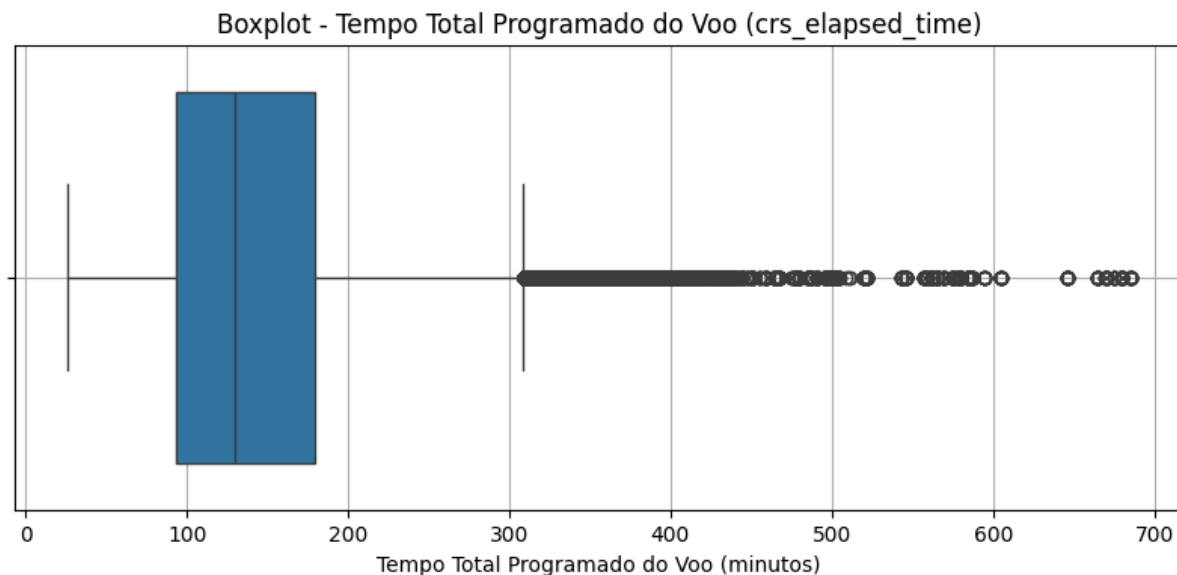
### *Taxiamento Antes da Decolagem (taxi\_out)*

Mediana de 15 min, média de 18.33 min. Os outliers chegam a 222 minutos. Mesmo sem outliers estatísticos, o máximo ainda é 34 minutos, mostrando que mesmo os tempos “normais” podem variar bastante.



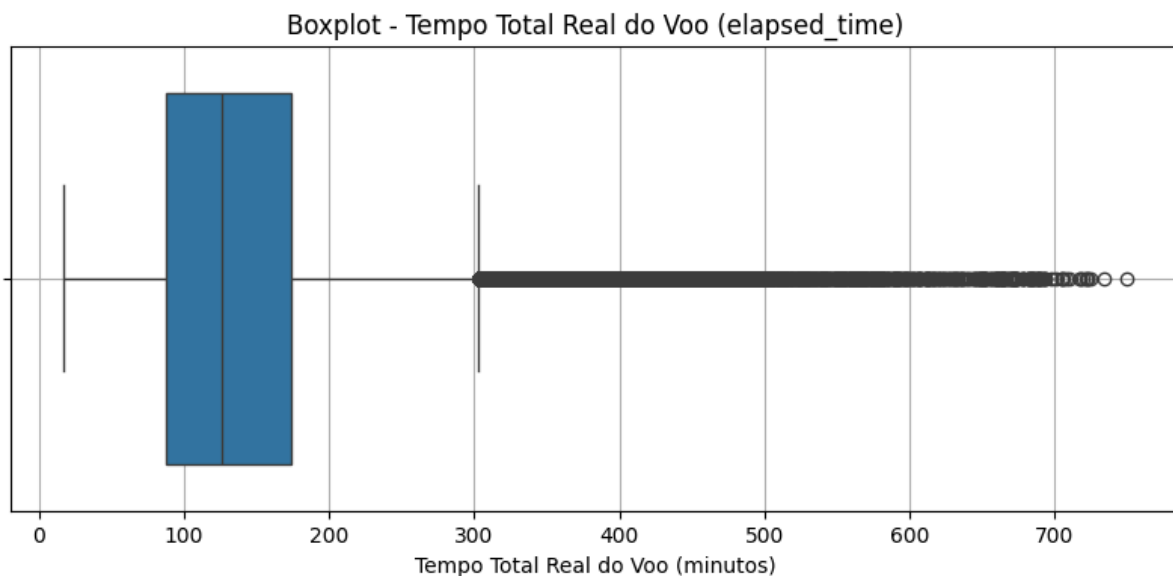
### *Tempo Total Programado do Voo (crs\_elapsed\_time)*

Mediana de 130 min, média de 147.24 min. Mostra a distribuição dos tempos de voo planejados.



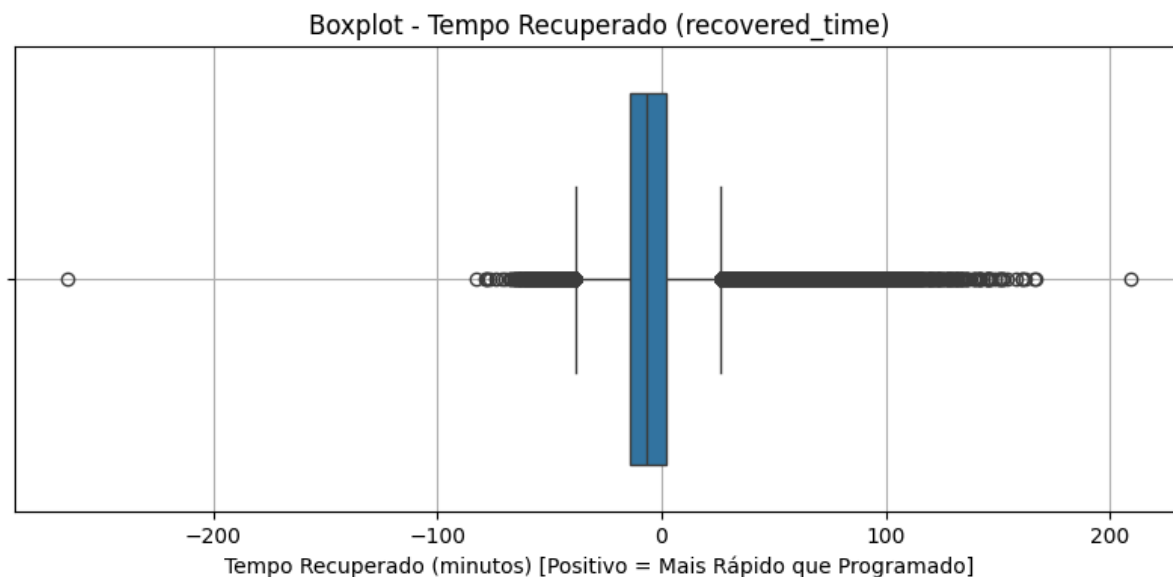
### *Tempo Total Real do Voo (elapsed\_time)*

Mediana de 126 min, média de 142.17 min. Interessante notar que, em média e na mediana, o tempo real de voo (elapsed\_time) é menor que o programado (crs\_elapsed\_time). Isso sugere que, no geral, há uma tendência de “ganhar tempo” em relação ao total programado. Isso se conecta à variável recovered\_time.

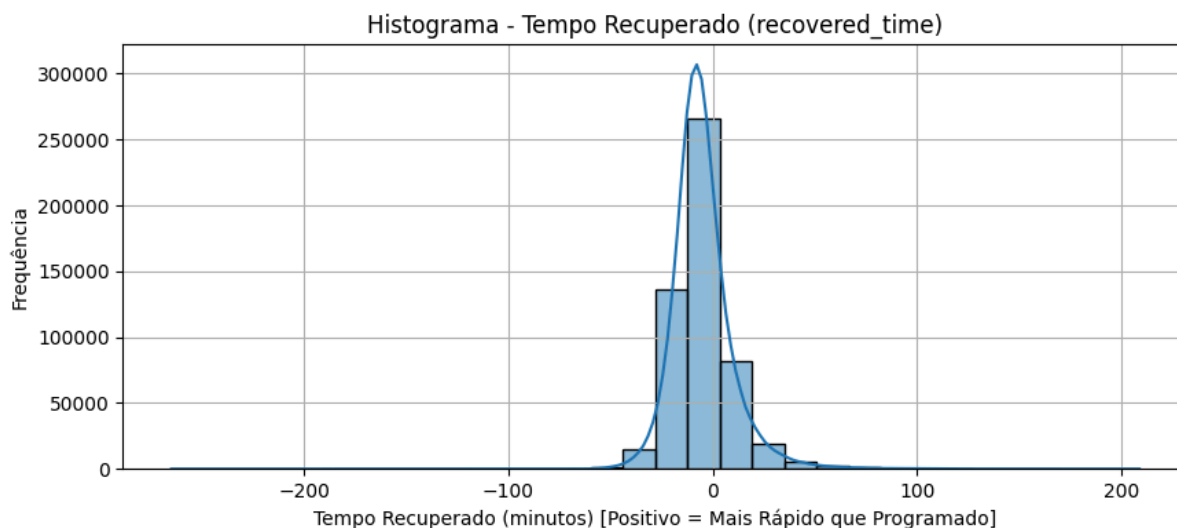


## Análise da Variação do Tempo Real de Voo em Relação ao Programado

A análise da capacidade dos voos de recuperar tempo em relação ao planejado (recovered\_time) mostrou que, para os 527.197 voos operados em janeiro de 2023, a maioria tende a **levar um pouco mais de tempo que o previsto** (média de -5.06 min, mediana de -7 min).

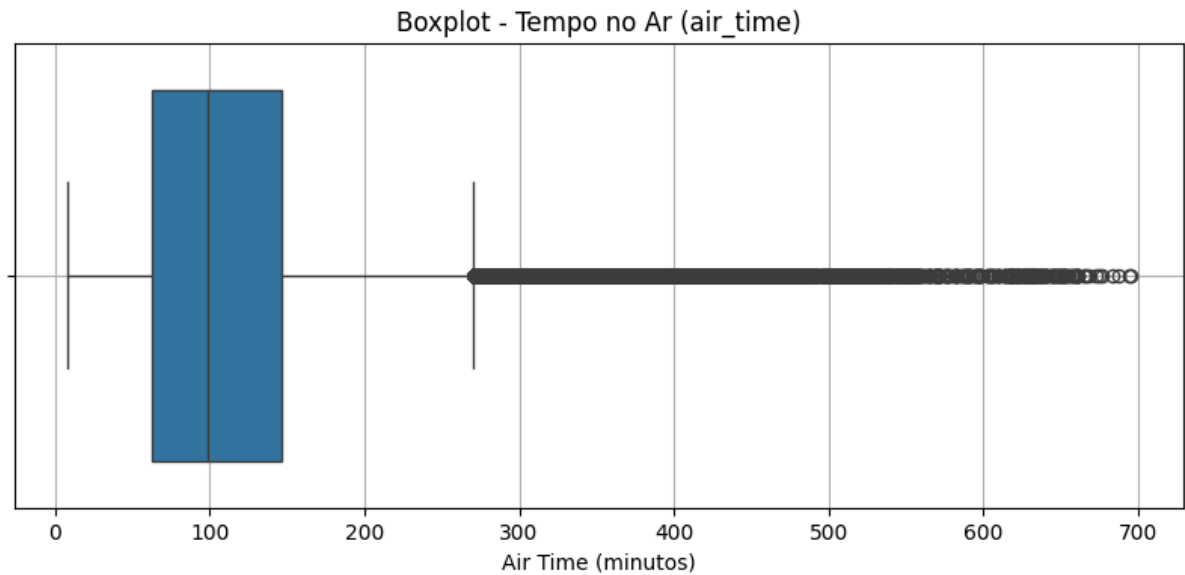


Apesar da existência de voos que ganharam tempo significativo (máx. 209 min), o resultado geral sugere que as eficiências em rota não foram, em média, suficientes para compensar outros fatores que podem estender a duração total da viagem, alinhando-se com o contexto operacional desafiador do período.



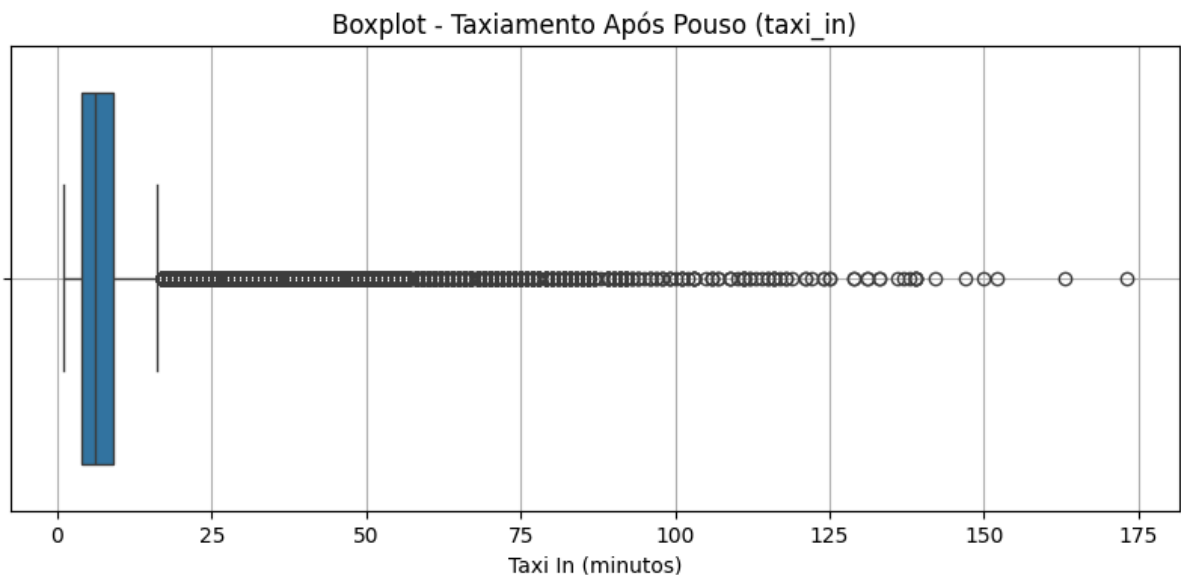
### *Tempo no Ar (air\_time)*

Mediana de 99 min, média de 115.81 min. Esta é a parte que conta o tempo apenas de voo realmente.



### *Taxiamento Após Pouso (taxi\_in)*

Mediana de 6 min, média de 8.03 min. Geralmente mais curto e com menos variabilidade que o taxi\_out, mas ainda com outliers significativos (até 173 min).



## Variáveis de Horário (hhmm)

As colunas de horário estavam originalmente armazenadas no formato INTEGER (por exemplo, 645 representando 06:45). Para possibilitar análises temporais mais precisas e garantir a correta interpretação por ferramentas de visualização, essas colunas foram transformadas para o tipo de dado TIME do BigQuery.

Este processo de conversão envolveu a extração das componentes de hora e minuto do valor inteiro, sua formatação para uma string padrão “HH:MM:SS”, e em seguida o uso da função PARSE\_TIME para converter para o tipo TIME. (crs\_dep\_time, dep\_time, wheels\_off, wheels\_on, crs\_arr\_time, arr\_time)

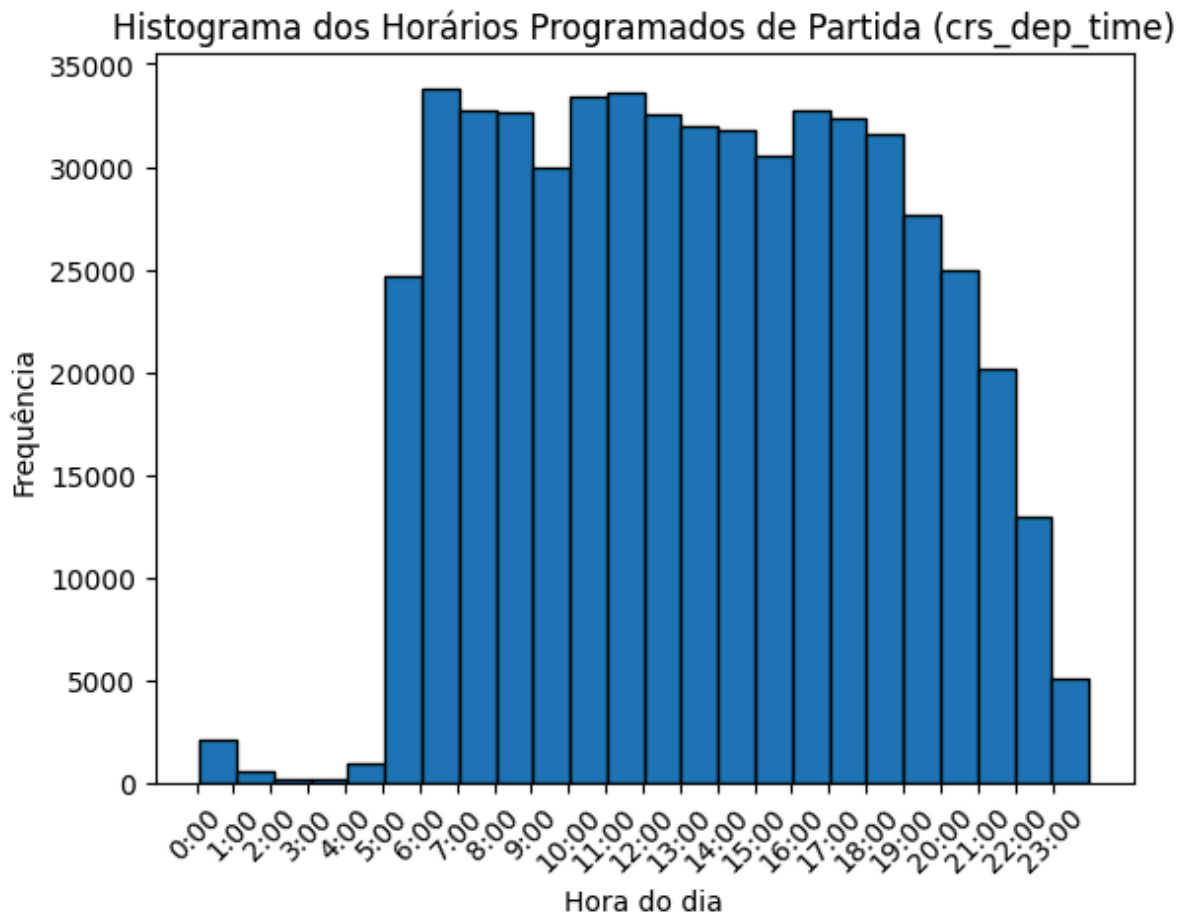
### *Hora de Partida Programada (crs\_dep\_time)*

A análise da distribuição dos horários programados de partida (crs\_dep\_time), visualizada por meio de um histograma da hora do dia, revela um padrão operacional claro. Observa-se uma **baixa atividade de voos programados durante a madrugada (00:00-04:00)**.

A partir das 05:00, o volume de partidas aumenta significativamente, atingindo um **primeiro pico entre 06:00 e 09:00**.

A alta atividade se mantém ao longo do dia, com um possível **segundo pico no final da tarde/início da noite (aproximadamente 16:00-19:00)**.

Após as 21:00, o número de partidas programadas diminui consideravelmente. Este padrão sugere os períodos de maior intensidade operacional e potencial congestionamento nos aeroportos.



### *Hora Real de Partida (dep\_time)*

Os dados mostram que há **baixa atividade na madrugada** (00h-04h), com o menor volume de voos e um **pico isolado de atraso às 04h** (32,25 min).

A operação se intensifica fortemente a partir das 05h, com o maior volume de voos entre **06h e 08h**, mantendo-se alta até o fim da tarde, com picos secundários entre **10h-11h** e **16h-18h**.

À noite, os voos diminuem progressivamente após 19h. A distribuição dos horários reais de partida segue padrão semelhante ao programado, porém com deslocamentos e caudas mais largas, refletindo atrasos.

O **atraso médio é menor no início da manhã** (7-9 min) e cresce ao longo do dia, **chegando a 18,18 min às 19h**, indicando efeito cascata típico de acúmulo operacional.

Histograma dos Horários Reais de Partida (dep\_time)

