**Technical Analysis Report: Decoding Flights Disruption**

**1. Introduction**

Air travel is a critical component of the modern economy, but its efficiency is frequently hampered by delays and cancellations. In recent years, particularly following the COVID-19 pandemic, the U.S. aviation industry has faced significant operational challenges, leading to widespread passenger dissatisfaction and financial losses for airlines. This report presents a data-driven analysis of U.S. airline flight disruptions between August 2019 and August 2023. The goal is to move beyond anecdotal evidence and pinpoint the root causes, be they are related to airline performance, specific routes, temporal patterns, or operational inefficiencies, to provide actionable insights for optimizing airline operations.

**2. Problem Statement**

U.S. airline flight delays and cancellations are causing severe passenger dissatisfaction and operational inefficiencies. This is evidenced by a 252% increase in passenger complaints compared to pre-pandemic 2019 levels (U.S. Department of Transportation, 2023).

**3. Objectives**

This project aims to accomplish the following:

* Analyze the overall airline performance metrics for delays and cancellations.
* Assess the impact of specific routes and airport locations on flight disruptions.
* Identify how time-related factors (time of day, day of week, and season) contribute to delays.
* Evaluate airport and airline operational efficiency, including gate-to-runway times and delay recovery.

**4. Target Audience**

The primary audience for this analysis is **Airline Operations Management** teams. These stakeholders are directly responsible for scheduling, resource allocation, and implementing strategic changes to improve on-time performance and operational resilience.

**5. Dataset**

* **Link:** [Flight Delay and Cancellation Dataset, 2019-2023 on Kaggle](https://www.kaggle.com/datasets/patrickzel/flight-delay-and-cancellation-dataset-2019-2023?utm_source=chatgpt.com&select=flights_sample_3m.csv)
* **Source:** U.S. Department of Transportation, Bureau of Transportation Statistics.
* **Time Period:** August 2019 – August 2023.
* **Description:** The dataset contains records of domestic flights within the USA, including variables such as airline, origin/destination airports, scheduled and actual departure/arrival times, and delay causes. The analysis uses a 3-million-row sample **Rows:** Each row represents a single flight that was either delayed or cancelled.
* **Dataset Contents Summary**

**Time Period:** January 2019 - August 2022

**Content:** Only delayed and cancelled flights (not successful on-time flights)

**Airlines:** Multiple major U.S. carriers (United, Delta, American, Spirit, Southwest, Republic)  
**Airports:** 380 different airports across the United States

* **Key Variables & Data Types/ Data Dictionary:**

|  |  |  |
| --- | --- | --- |
| Column Name | Description | Format / Unit |
| **FL\_DATE** | The date of the flight. | MM/DD/YYYY |
| **AIRLINE** | The full name of the airline. | Text |
| **AIRLINE\_DOT** | A concatenated string of the airline name and its two-letter code. | Text |
| **AIRLINE\_CODE** | The two-letter IATA code that uniquely identifies the airline. | Code |
| **DOT\_CODE** | The unique identifier assigned to the airline by the Department of Transportation. | Numeric Code |
| **FL\_NUMBER** | The flight number assigned by the airline. | Numeric |
| **ORIGIN** | The three-letter IATA code of the airport where the flight originated. | Code |
| **ORIGIN\_CITY** | The city and state of the origin airport. | Text |
| **DEST** | The three-letter IATA code of the destination airport. | Code |
| **DEST\_CITY** | The city and state of the destination airport. | Text |
| **CRS\_DEP\_TIME** | The scheduled departure time (from the origin). | Local Time (HHMM, 24-hr) |
| **DEP\_TIME** | The actual departure time (from the origin). | Local Time (HHMM, 24-hr) |
| **DEP\_DELAY** | The difference between actual and scheduled departure time. | Minutes |
| **TAXI\_OUT** | The time spent taxiing between the gate and runway for takeoff. | Minutes |
| **WHEELS\_OFF** | The time when the aircraft's wheels leave the ground. | Local Time (HHMM, 24-hr) |
| **WHEELS\_ON** | The time when the aircraft's wheels touch the ground on arrival. | Local Time (HHMM, 24-hr) |
| **TAXI\_IN** | The time spent taxiing between the runway and the gate after landing. | Minutes |
| **CRS\_ARR\_TIME** | The scheduled arrival time (at the destination). | Local Time (HHMM, 24-hr) |
| **ARR\_TIME** | The actual arrival time (at the destination). | Local Time (HHMM, 24-hr) |
| **ARR\_DELAY** | The difference between actual and scheduled arrival time. A primary measure of delay. | Minutes |
| **CANCELLED** | Indicates whether the flight was cancelled. | Boolean (1 = Yes, 0 = No) |
| **CANCELLATION\_CODE** | The reason for the cancellation (only present if CANCELLED = 1). | Code (A, B, C, D) |
| **DIVERTED** | Indicates whether the flight was diverted to an alternate airport. | Boolean (1 = Yes, 0 = No) |
| **CRS\_ELAPSED\_TIME** | The scheduled total time allocated for the flight (from scheduled departure to scheduled arrival). | Minutes |
| **ELAPSED\_TIME** | The actual total time of the flight (from actual departure to actual arrival). | Minutes |
| **AIR\_TIME** | The total time the aircraft was in the air (from wheels-off to wheels-on). | Minutes |
| **DISTANCE** | The distance between the origin and destination airports. | Miles |
| **DELAY\_DUE\_CARRIER** | The portion of the arrival delay (in minutes) caused by the airline (e.g., maintenance, crew). | Minutes |
| **DELAY\_DUE\_WEATHER** | The portion of the arrival delay (in minutes) caused by significant weather conditions. | Minutes |
| **DELAY\_DUE\_NAS** | The portion of the arrival delay (in minutes) caused by the National Aviation System (e.g., airport operations, heavy traffic, air traffic control). | Minutes |
| **DELAY\_DUE\_SECURITY** | The portion of the arrival delay (in minutes) caused by security-related issues (e.g., evacuation, security breach). | Minutes |
| **DELAY\_DUE\_LATE\_AIRCRAFT** | The portion of the arrival delay (in minutes) caused by the late arrival of the same aircraft from a previous flight. | Minutes |

**6. Data Handling**

**Created DAX Measures & Columns:**

**6.1. Performance Metrics**

* Arrived Delayed % = DIVIDE([Delayed Arrivals], [Total Flights])
* Arrived Early % = DIVIDE([Arrived Early Count], [Total Flights])
* Arrived On-Time % = DIVIDE([Arrived On-Time], [Total Flights])
* Cancelled % = DIVIDE([Cancellations], [Total Flights])
* Avg Departure Delay = AVERAGE(DEP\_DELAY)
* Avg Arrival Delay = AVERAGE(ARR\_DELAY)
* Significant Delay Rate = % of delays >15 minutes

**6.2. Time Analysis**

* Departure Hour = HOUR(DEP\_TIME)
* Arrival Hour = HOUR(ARR\_TIME)
* Month Name = FORMAT(FL\_DATE, "MMMM")
* Month Sort = MONTH(FL\_DATE)
* Day of Week = FORMAT(FL\_DATE, "dddd")

**6.3. Flight Categorization**

* Flight Duration Category = Groups by 0-60, 61-180, 181-300, 300+ min
* Taxi-Out Category = Normal (0-15), Concerning (16-30), Severe (31-45), Critical (45+)
* Official Delay Category = Early/On-Time, Minor, Significant, Extreme
* Time Status = Delayed, Early, On-Time (Actual vs Scheduled)
* Overall Status = Cancelled, Arrived Delayed, Arrived Early, Arrived On-Time

**6.4. Route & Location**

* Route = ORIGIN & " to " & DEST
* Adjusted Departure Delay = Filters outliers (-30 to 300 min

**6.5. Count Measures**

* Total Flights = COUNT(FL\_NUMBER)
* Operated Flights = Total - Cancelled
* Cancellations = COUNT where CANCELLED = 1
* Delayed Arrivals = COUNT where ARR\_DELAY > 15
* Delayed Departures = COUNT where DEP\_DELAY > 15
* On-Time Flights = COUNT where DEP\_DELAY ≤15 AND ARR\_DELAY ≤15

**7. Analysis and Findings**

The analysis yielded several key insights across different dimensions:

* **Overall Performance:** Of the analyzed flights, 33% arrived delayed, while only 2% arrived exactly on time. The average departure delay (10.12 minutes) was more than double the average arrival delay (4.26 minutes), indicating significant delay recovery occurs in-flight.
* **Airline-Specific Disruptions:** Southwest Airlines Co. had the highest number of delayed flights (107K) and cancellations (19K), notably due to a major operational meltdown in December 2022.
* **Route & Location Analysis:**
  + Routes like LAS-MDW and LAS-DAL had extremely high delay rates (≈60%), primarily due to tight turnaround times and a lack of recovery buffers.
  + Smaller airports (e.g., Pago Pago, TT; Williamsport, PA) suffered from the longest average delays due to limited ground resources and infrastructure, despite lower flight volumes.
* **Time Contribution:**
  + **Time of Day:** Departure delays spiked during peak travel hours due to runway and gate congestion.
  + **Day of Week:** Fridays and Thursdays experienced the worst delays, while Tuesdays had the best performance.
  + **Seasonality:** June and December (holiday seasons) consistently showed high disruption rates.
* **Operational Efficiency:**
  + **Gate-to-Runway Time:** Flights with severe ground delays (31+ minutes) showed high departure delays, though the data suggests that some may get runway priority, masking the full ground wait.
  + **Delay Recovery:** On average, flights recovered 6 minutes in the air. Long-haul flights could effectively reduce delays by increasing speed, while short-haul flights lacked the time to recover.
  + **Scheduling Accuracy:** A scatter plot of Scheduled vs. Actual Time revealed that many flights, especially long-haul, arrived earlier than scheduled, suggesting potential schedule padding.

**8. Recommendations**

Based on the findings, the following actions are recommended:

1. **Right-Size Schedules:**
   * Add 10–15-minute buffers to short-haul flights to account for poor delay recovery.
   * Tighten schedules for long-haul flights that consistently arrive early to free up gates and reduce idle time.
   * Cap scheduled flights during peak periods to match real airport and airline capacity.
2. **Reinforce Operations:**
   * Strategically increase staffing and ground resources during peak hours and high-risk days (Thursdays, Fridays).
   * Mandate "meltdown-proof" contingency plans for major airlines to prevent systemic failures.
   * Boost operational capacity and reinforce staffing ahead of known holiday rushes.
3. **Target Key Airports:**
   * Upgrade ground resources (e.g., baggage handlers, gate crew) at regional airports with long delays.
   * Optimize arrival sequencing and taxiway management at congested hubs like DEN, LAS, and MDW to reduce cascade effects.

**9. Limitations and Assumptions**

* **Data Scope:** The analysis uses a 3-million-row sample, which, while large, may not capture the full nuance of the complete 29-million-record dataset.
* **Causal Attribution:** While the analysis identifies correlations (e.g., between short flights and poor recovery), it relies on domain knowledge and external reports (e.g., from the DOT) to infer causation.
* **Data Granularity:** The dataset provides high-level delay causes but may lack granular details on specific incidents (e.g., specific mechanical issues, detailed weather data at the minute-level), which could affect root cause analysis.
* **Assumption of Representativeness:** It is assumed that the 3-million-row sample is representative of the overall population of U.S. domestic flights in terms of delay patterns and causes.
* **External Factors:** The analysis period includes the anomalous COVID-19 pandemic years (2020-2021), which severely disrupted normal travel patterns and may skew some historical comparisons.

**10. References**

* U.S. Department of Transportation. (2023). *Air Travel Consumer Report*.
* U.S. Bureau of Transportation Statistics (BTS). (2023). *On-Time Performance Data*.
* Federal Aviation Administration (FAA). (2022). *Administrative Data*.
* U.S. Department of Transportation, Office of Inspector General. (2023). *Reports on Airline Operations*.
* Kaggle Dataset: "Airline Flight Delay and Cancellation Data, August 2019 - August 2023".