PUNJAB ENGINEERING COLLEGE (DEEMED TO BE UNIVERSITY) CHANDIGARH

UNITED

SKYHACK 3.0

DATA-DRIVEN FRAMEWORK FOR FLIGHT DIFFICULTY SCORING AND OPERATIONAL DRIVER ANALYSIS

Data-driven insights for proactive planning and resource allocation

Table of Contents:

UNITED

SKYHACK 3.0

Cleaning Data

- Scope: Flights, PNR, Remarks, Airports, Bags
- Principles: Deduplicate; type-safe datetimes/numerics; canonical IDs (upper/trim); booleans→1/0
- Transforms: Flights (median-impute ops metrics); PNR/Remarks (counts→0, SSR "NONE", dates); Bags/Airports (titles/dates/codes)
- Harmonization: Trim flight_number, upper company_id, coerce scheduled_departure_date_local; join-ready flight_id basis
- Outcomes: Clean rows — 8,099 | 687,878 | 51,698 | 5,612 | 686,952; reliable joins/time math
- Next: Persist flight_id; add schema validation + quality stats

Data Cleaning Strategy

Deduplicate • Type-safe conversions • Canonical IDs • Boolean harmonization • Graceful coercion

TRANSFORMATION PIPELINE

Inputs
5 Raw Datasets

→

Cleaning
Dedup • Types • Impute

→

Harmonization
Canonical Keys

→

Outputs
Join-Ready Data

CLEANED DATA METRICS

8,099	687,878	51,698	5,612	686,952
Flights	PNR Flight	PNR Remarks	Airports	Bags

KEY PRINCIPLES

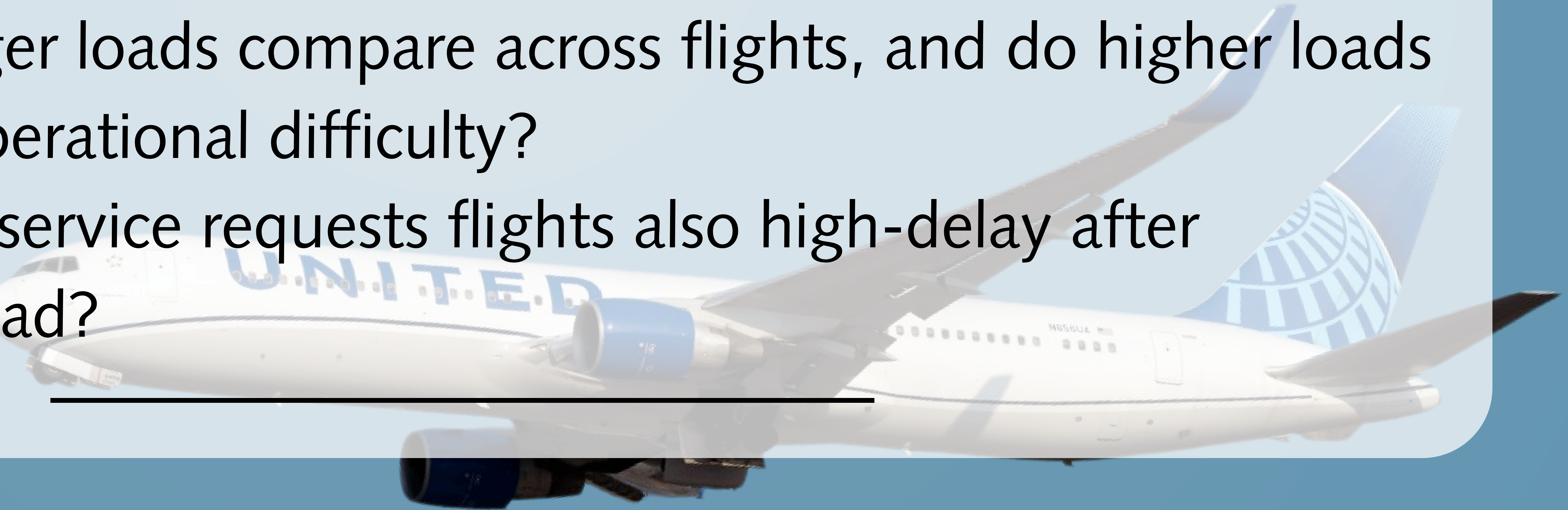
- **Imputation Strategy**
  Ops metrics→median | Count fields→zero
- **Join-Ready Keys**
  company_id + flight_number + scheduled_departure_date
- **Data Quality**
  100% type safety | Zero duplicates | Canonical identifiers

Source: Clean-Dataset.ipynb • Docs/clean-dataset.md

Total: 1.4M+ cleaned records

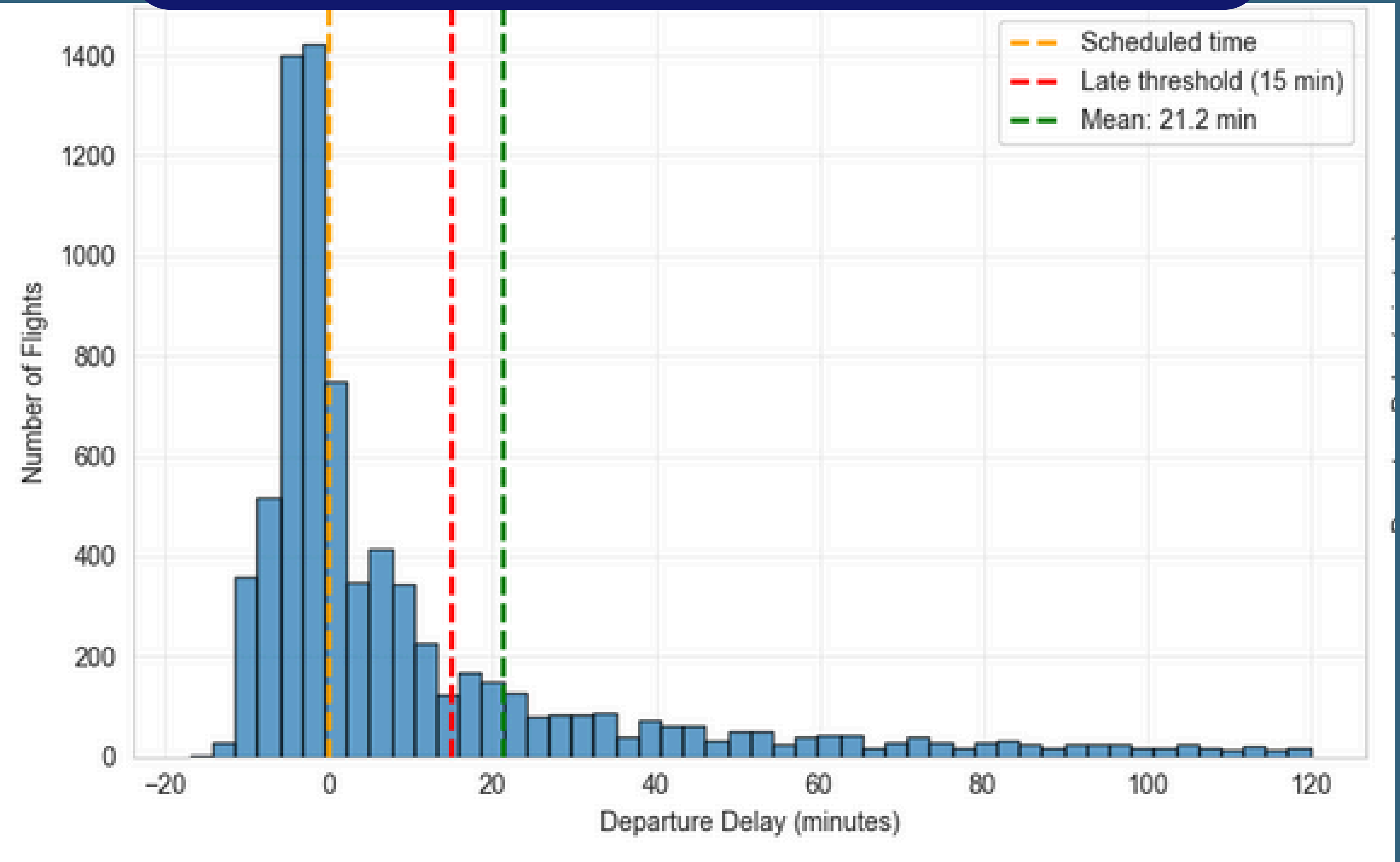
Deliverable 1:

- What is the average delay and what percentage of flights depart later than scheduled?
- How many flights have scheduled ground time close to or below the minimum turn mins?
- What is the average ratio of transfer bags vs. checked bags across flights?
- How do passenger loads compare across flights, and do higher loads correlate with operational difficulty?
- Are high special service requests flights also high-delay after controlling for load?



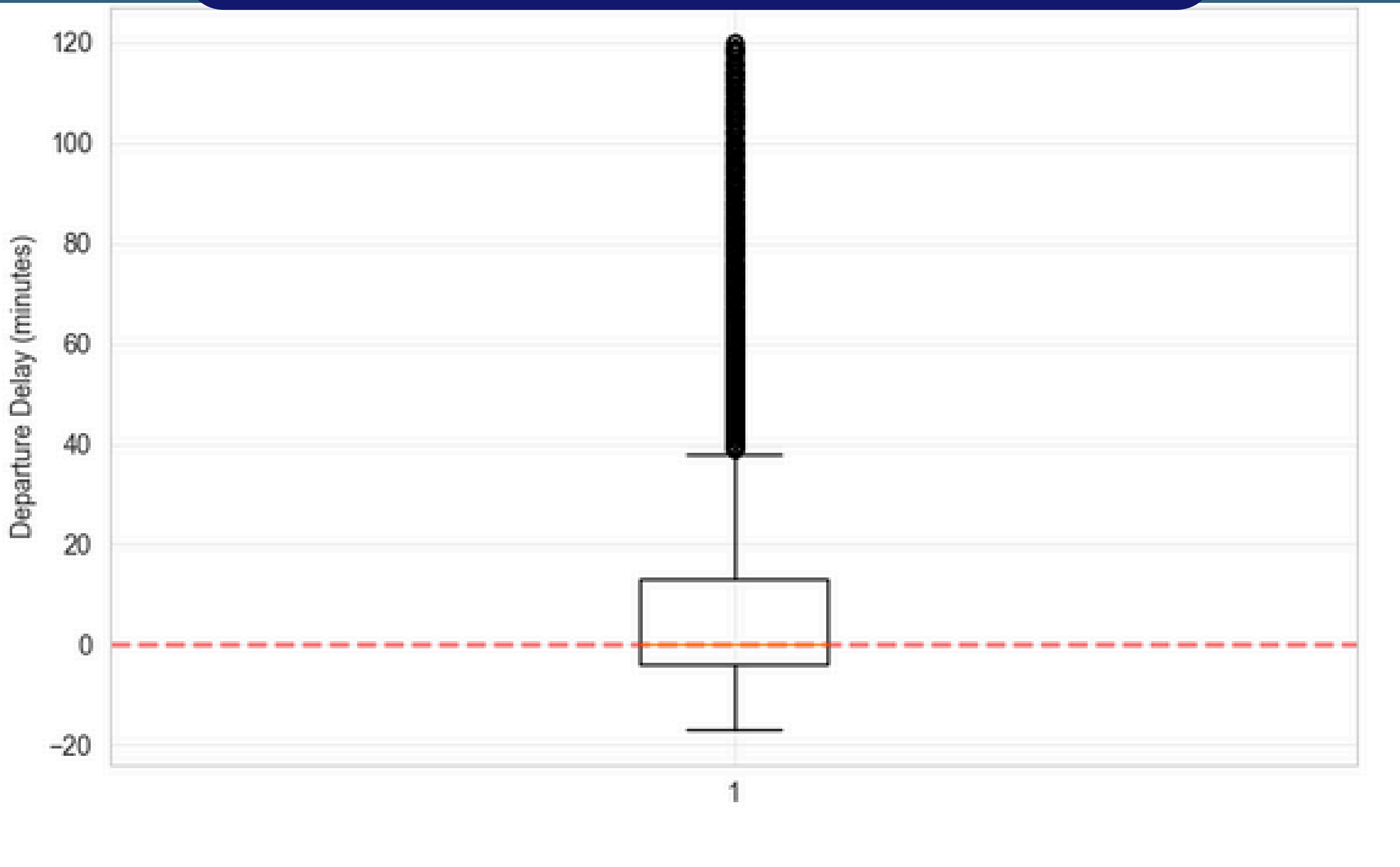
Average delay and percentage of flights which depart later than scheduled

Distribution of Department delays



The data shows most ORD flights depart on time, but about a quarter face delays over 15 minutes.

Box Plot of Department delays



Most delays cluster near zero minutes, with a few extreme outliers indicating occasional long delays.

Result:

- **Average delay: 21.18 minutes**
- **Flights analyzed: 8,099**
- **Late departures (>15 min): 2,163**
- **Percentage late: 26.71%**

Flights that have scheduled ground time below the minimum turn mins

```
=============================================================
Ground Time Risk by Fleet Type:
=============================================================
```

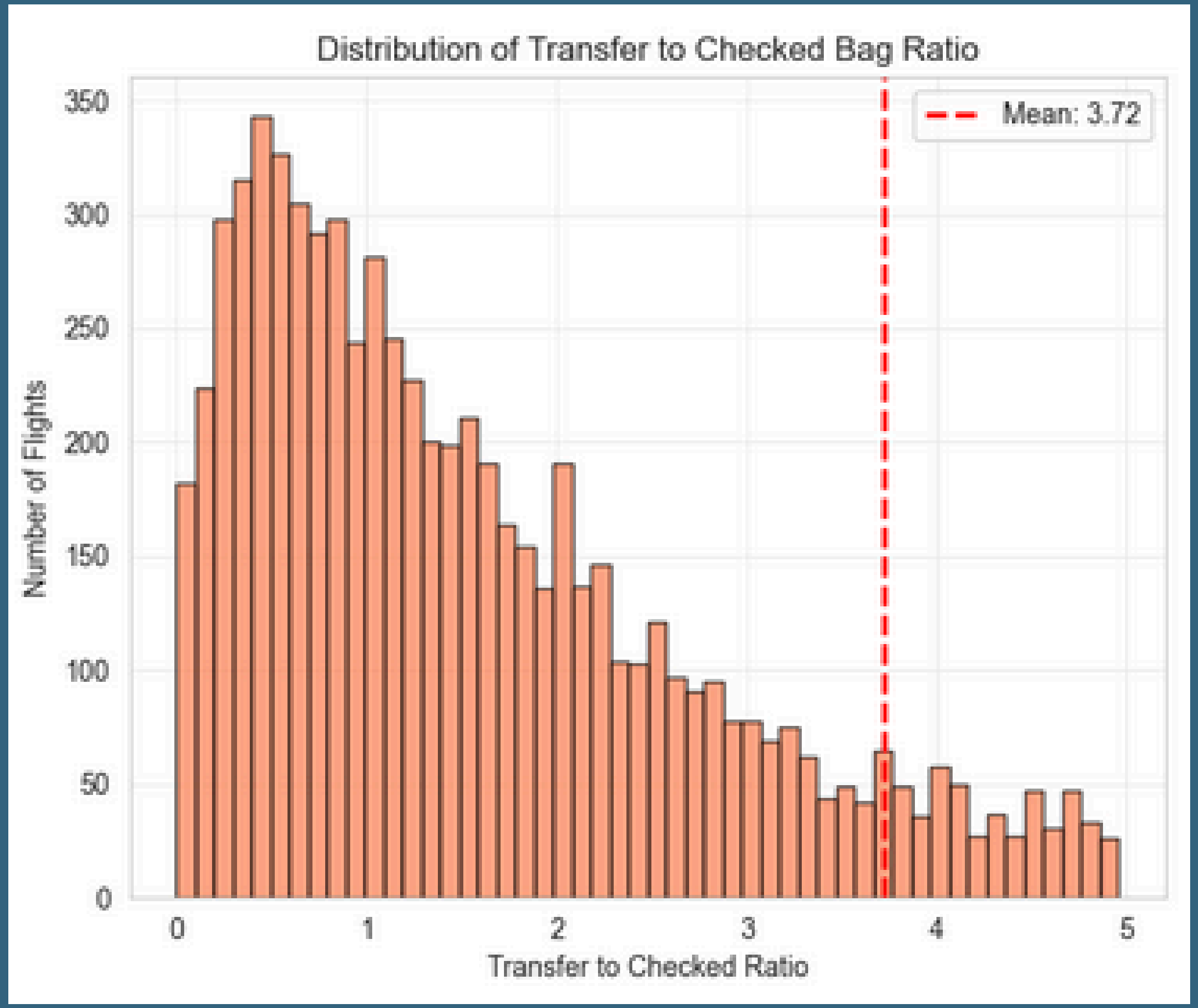
fleet_type	Risky_Flights	Total_Flights	Avg_Buffer_Minutes	Risk_Percentage
B767-400	2	2	-7.00	100.00
B737-800	121	674	92.10	17.95
B737-900	94	809	116.05	11.62
B737-700	41	395	108.83	10.38
B757-200	10	105	202.25	9.52
B767-300	5	60	455.75	8.33
B787-10	11	141	329.48	7.80
B757-300	12	156	332.51	7.69
CRJ-200	63	835	123.88	7.54
A319-100	37	499	123.93	7.41
A321-2NX	33	457	185.76	7.22
B787-8	5	73	593.77	6.85
A320-200	33	499	248.93	6.61
ERJ-175	82	1354	97.29	6.06
B737-MAX9	22	372	102.49	5.91
B737-MAX8	12	218	49.58	5.50
CRJ-550	69	1357	103.03	5.08
B777-2HD	0	37	244.95	0.00
B777-300	0	1	672.00	0.00
B787-9	0	4	622.25	0.00
ERJ-170	0	51	61.55	0.00

Out of all flights, 652 (8.05%) are at risk due to scheduled ground times below the minimum turnaround requirement. While the average ground buffer is 135.75 minutes, the median of only 29 minutes shows many tight connections. Risk levels vary by fleet type, indicating potential scheduling inefficiencies and operational vulnerabilities.

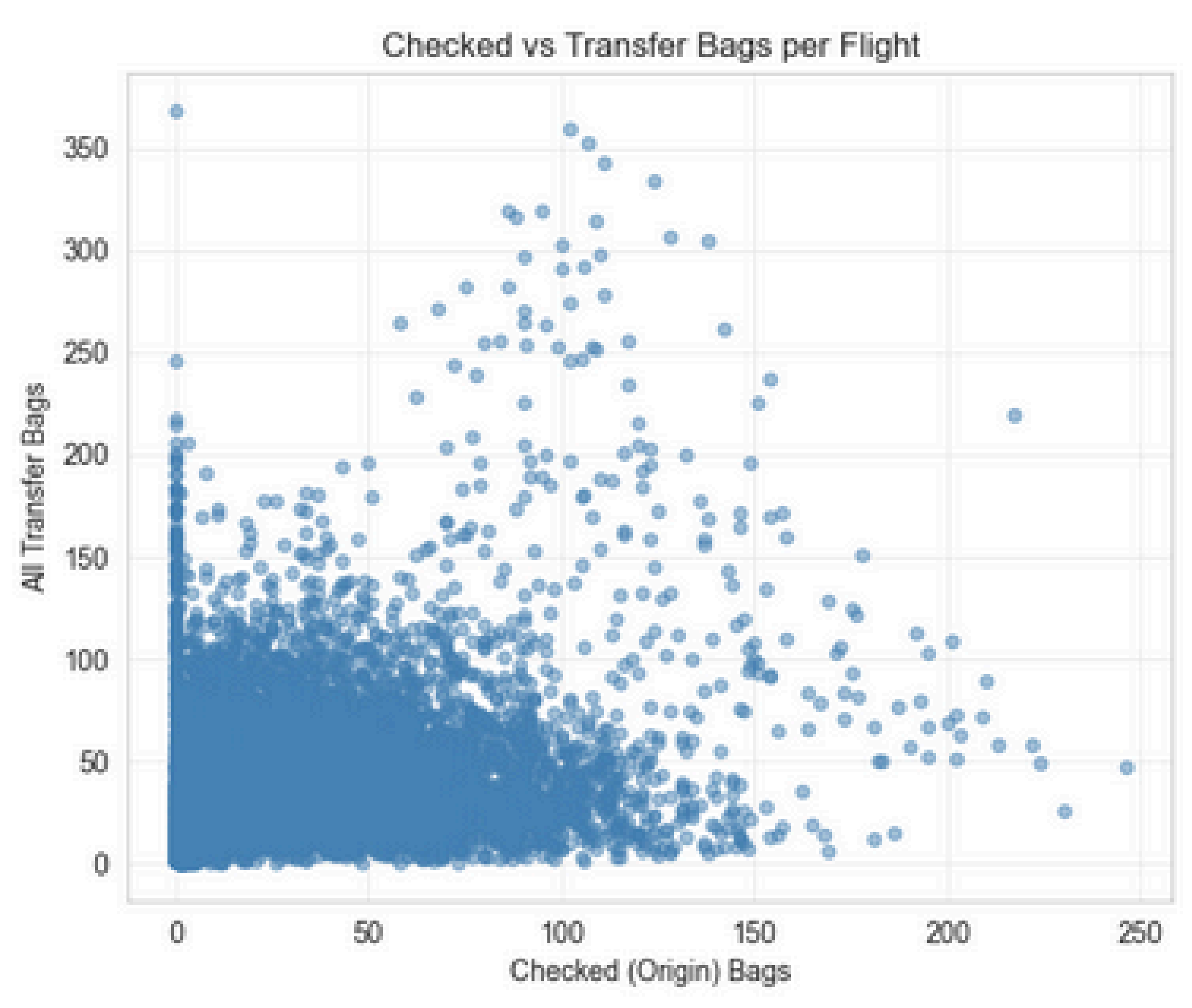
Result:

- **At-risk flights: 652**
- **Percent at risk: 8.05%**
- **Ground time buffer stats: avg 135.75 min; median 29.00 min**
- **Breakdown by fleet_type with risk rates**

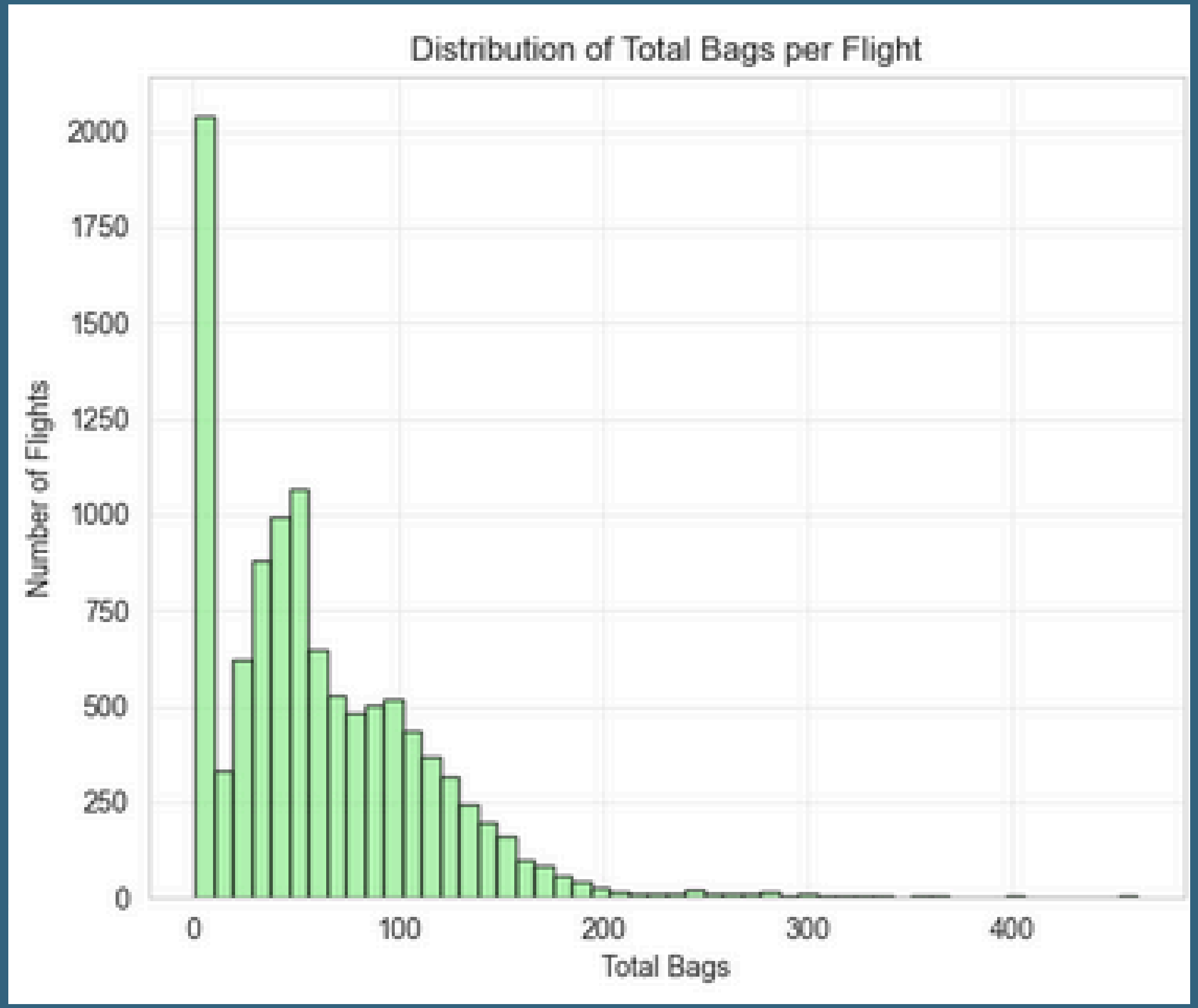
The average ratio of transfer bags vs. checked bags across flights



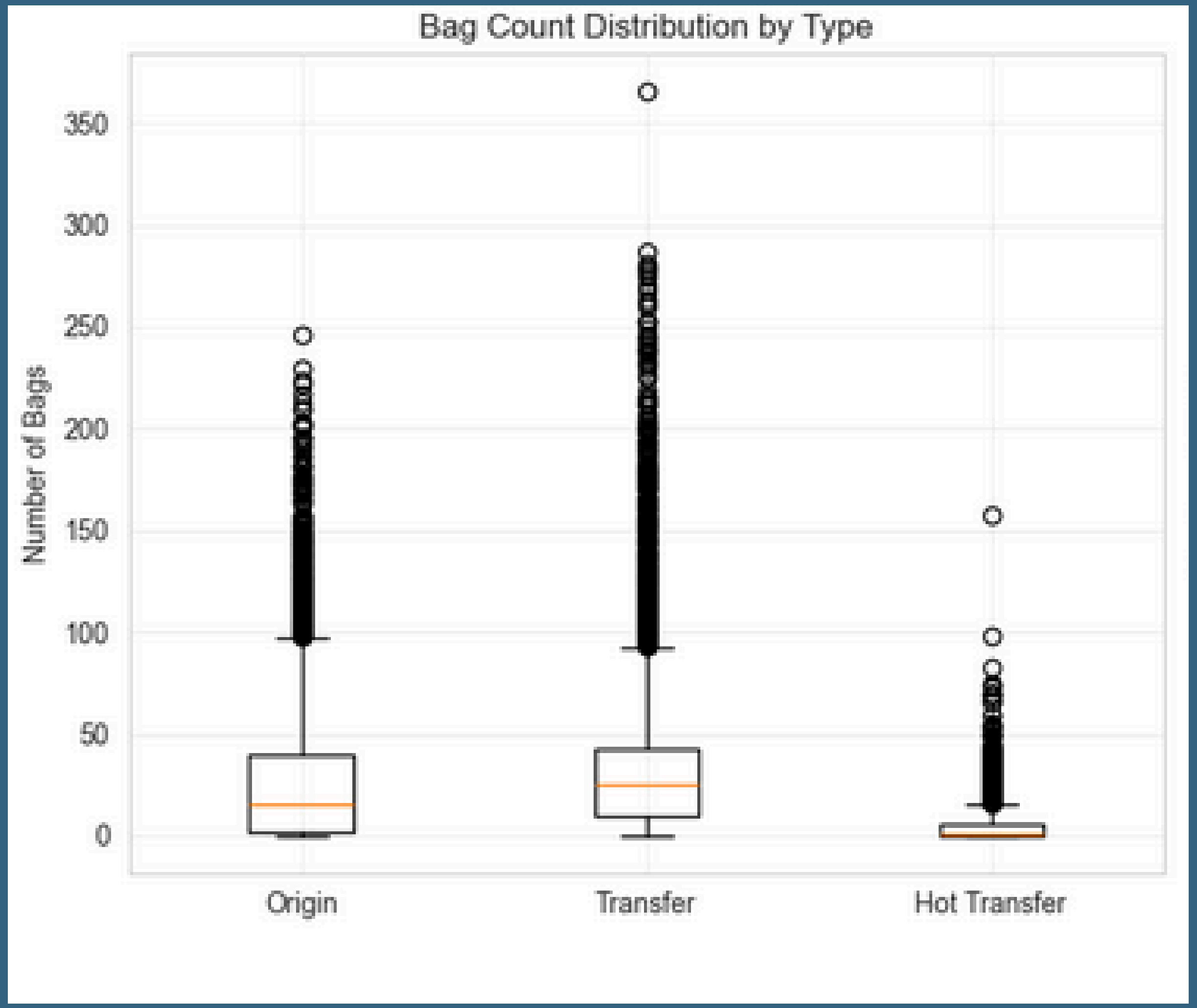
Most flights have low transfer-to-checked bag ratios, but a few with very high ratios skew the average upward to 3.72.



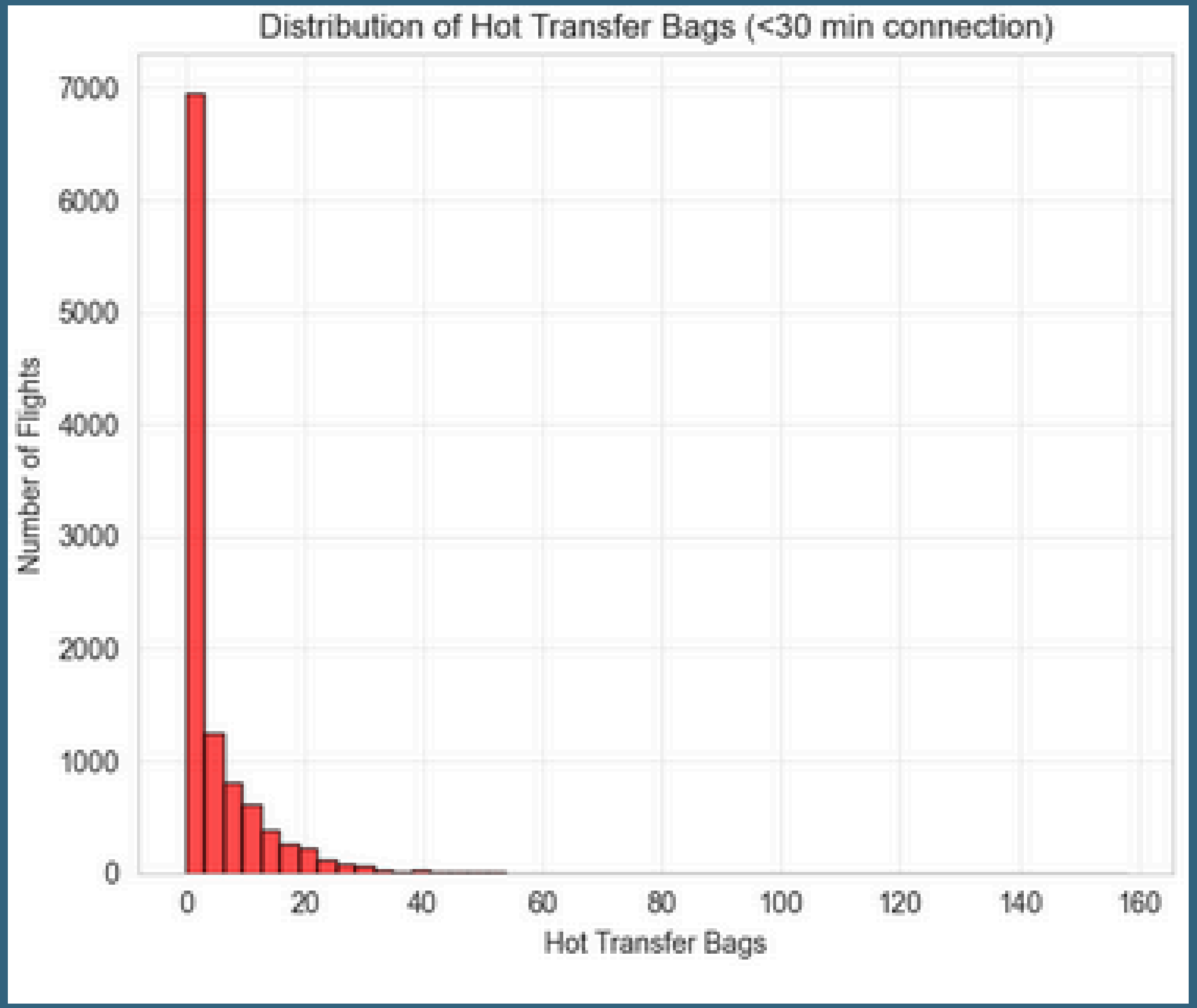
Most flights have under 100 checked bags and up to 350 transfer bags, with higher checked bag counts generally linked to fewer than 100 transfer bags.



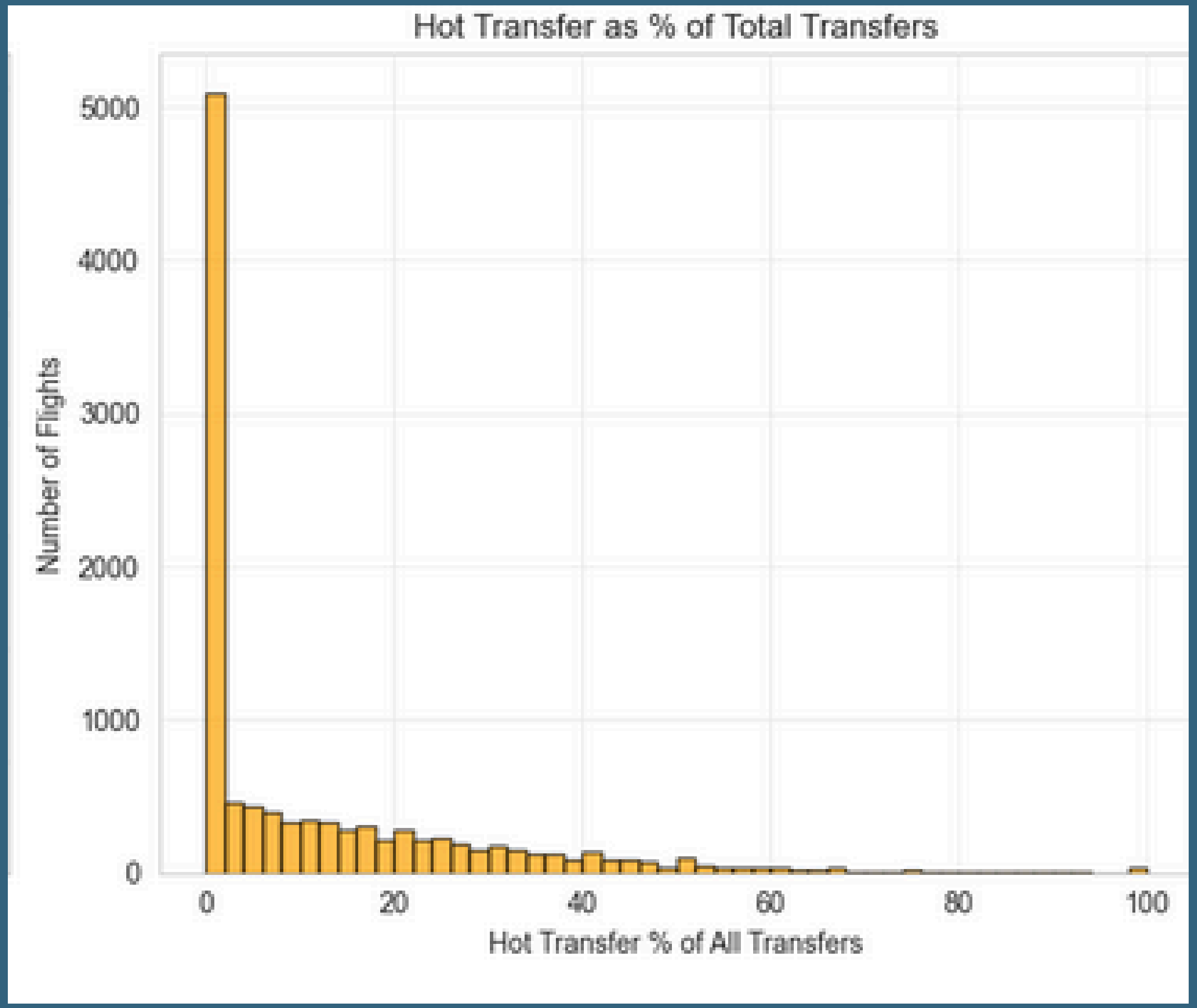
Most flights carry fewer than 100 total bags, with a sharp drop-off after that and only a few flights exceeding 200 bags.



Transfer and origin bags have similar median counts around 25–30, while hot transfer bags are fewer, typically under 10, with all types showing many high outliers.



Most flights have fewer than 10 hot transfer bags, with only a small number exceeding 20, showing a steeply right-skewed distribution.

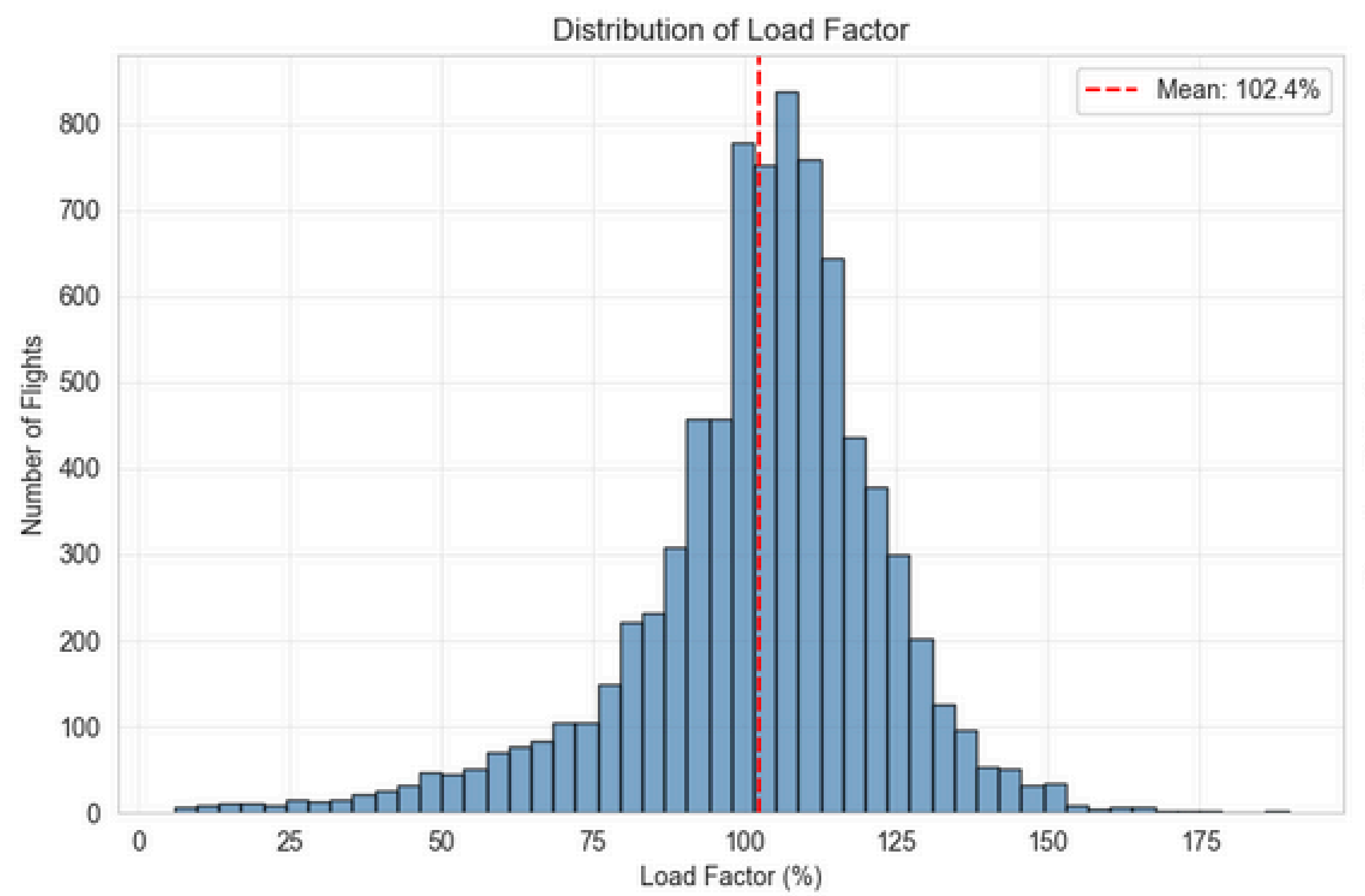


Most flights have less than 10% of their transfer bags classified as hot transfers, with very few reaching higher percentages.

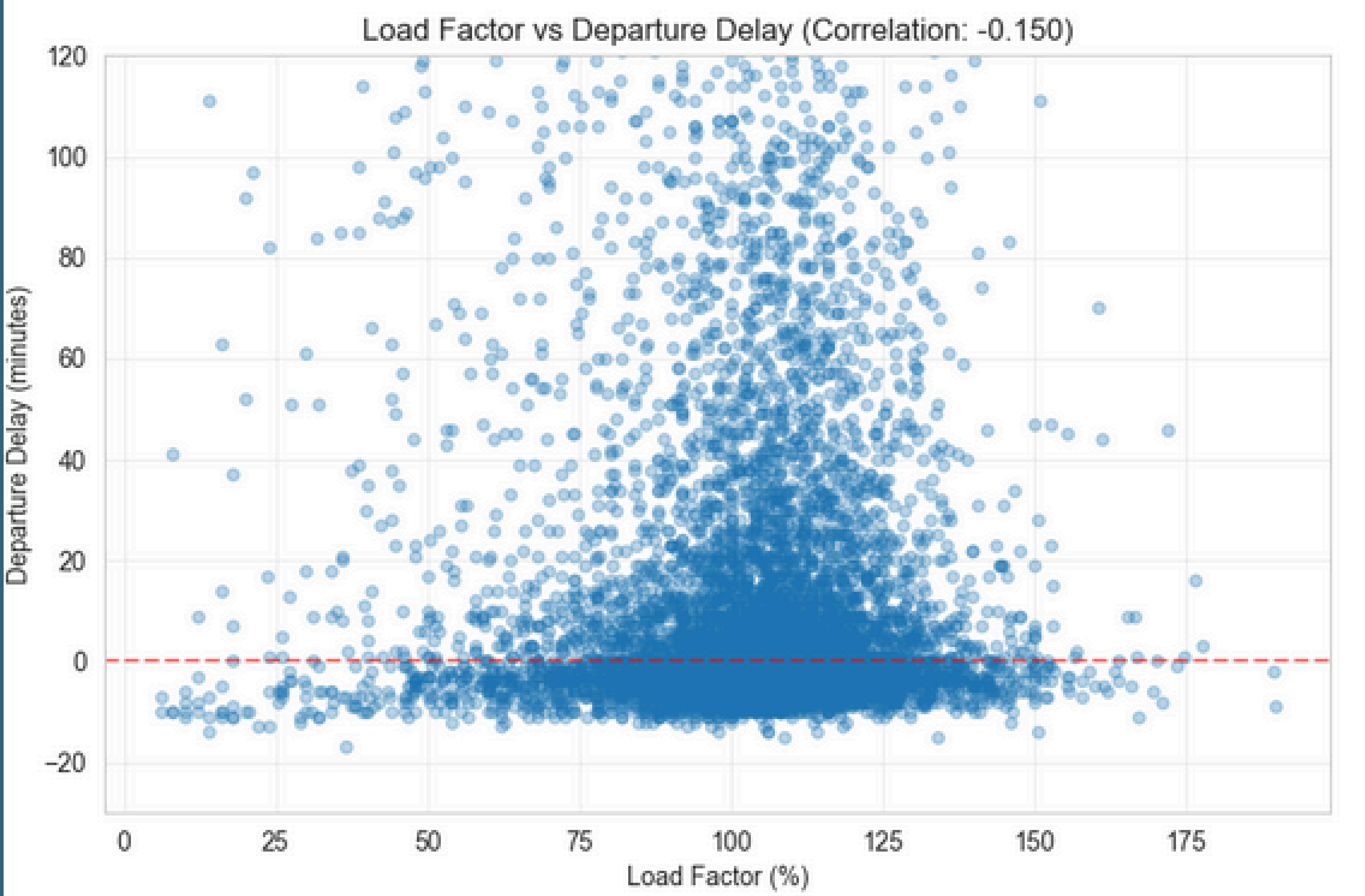
Result:

- **Average transfer:checked ratio: 3.7179**
- **Median ratio: 1.6000**
- **Flights with bag data: 10,863**
- **Totals: Checked 290,121; Regular Transfer 347,546; Hot Transfer 49,578; All Transfers 397,124; All Bags 687,245**
- **Average per flight: Checked 26.71; All Transfers 36.56; Hot Transfers 4.56**
- **Hot transfer share: 11.95%**

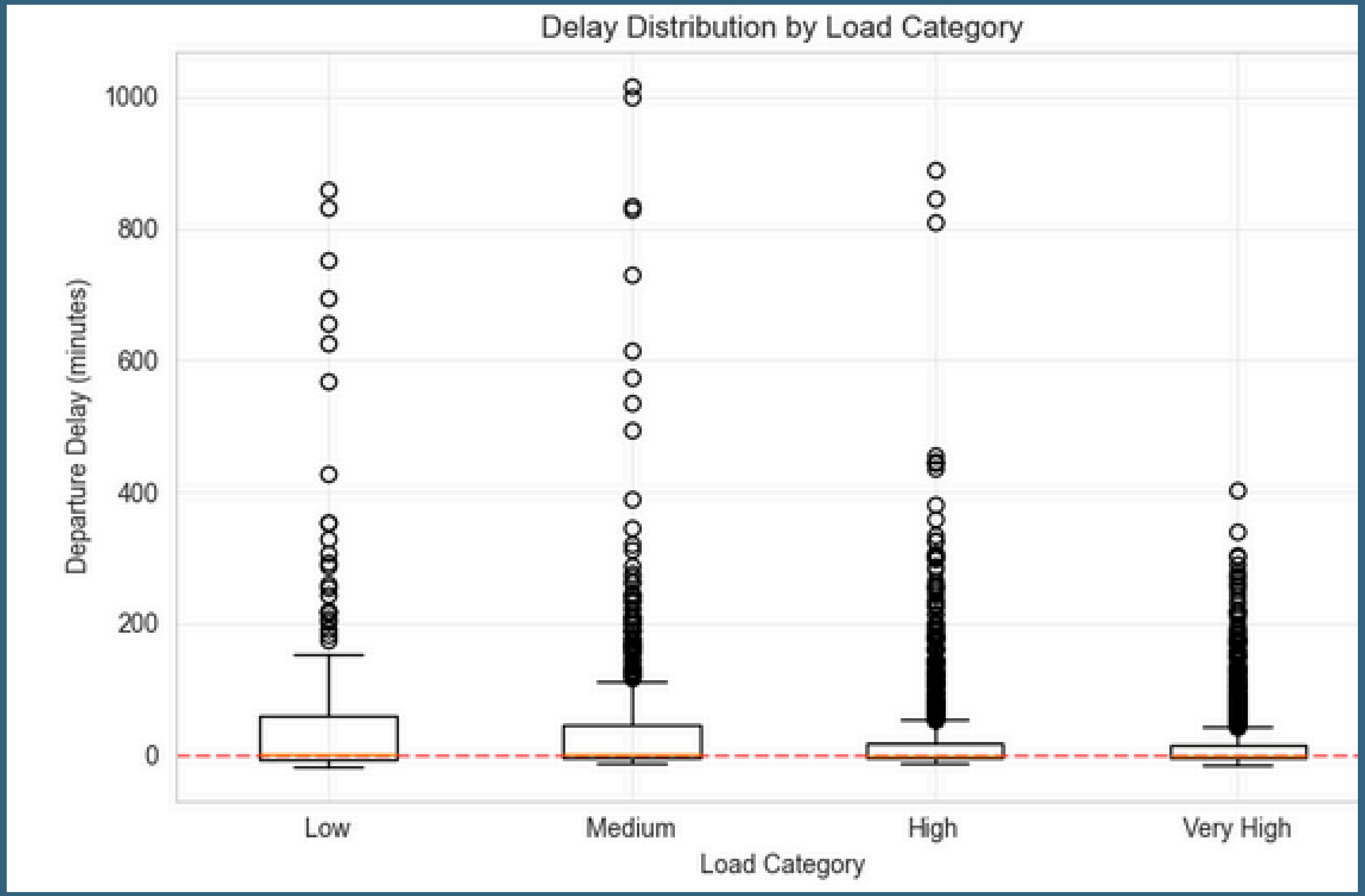
Passenger Load Analysis and its effect on operations



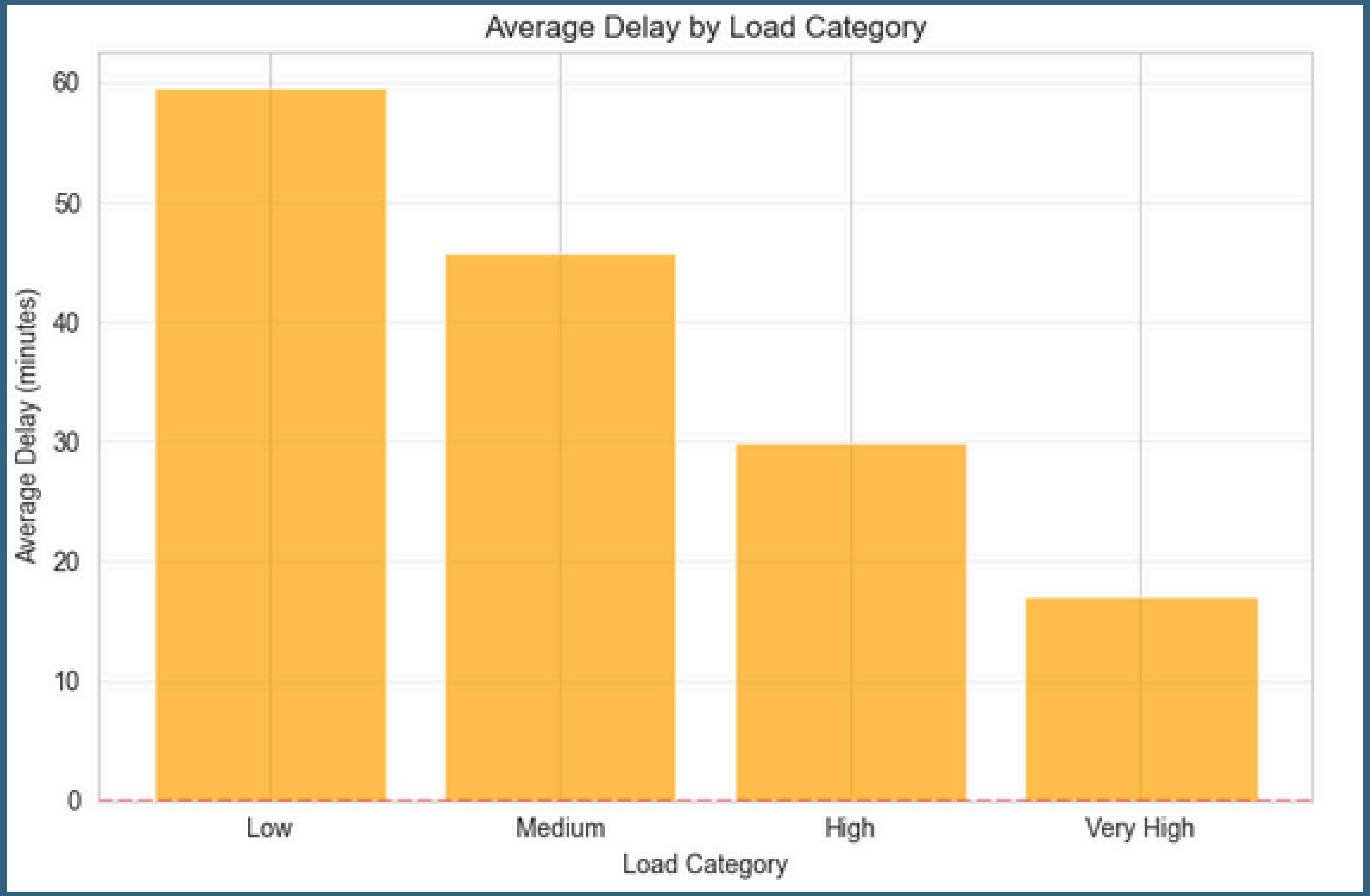
Most flights have load factors near 100%, with a mean of 102.4%, indicating many flights are operating at or slightly above capacity.



There is a very weak negative correlation (-0.15) between load factor and departure delay, indicating fuller flights don't significantly affect delays.



Across load categories, median departure delays are low, typically under 20 minutes, with numerous outliers causing delays up to 1000 minutes.

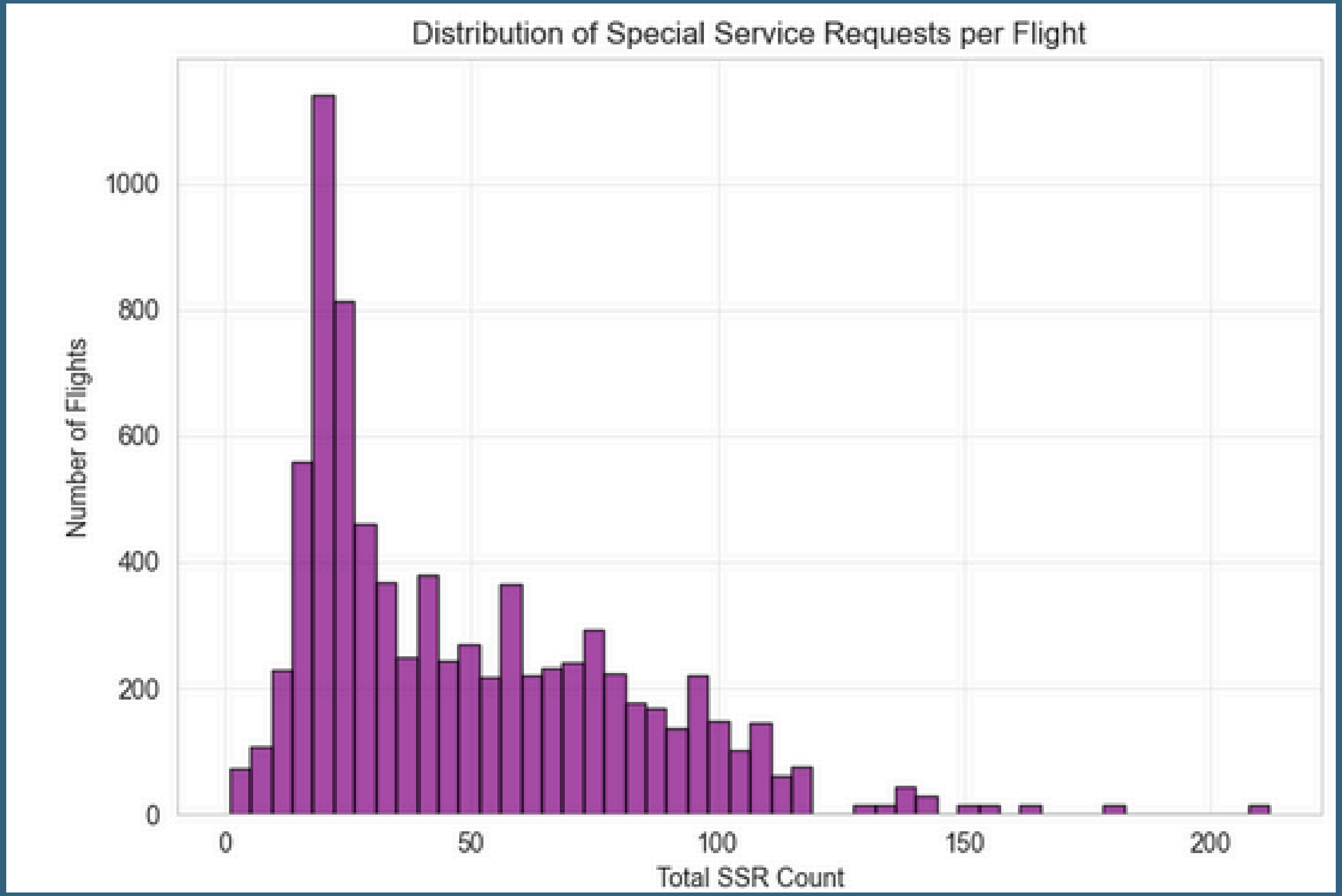


Average departure delays decrease by load category: Low is about 60 minutes, Medium about 45 minutes, High about 30 minutes, and Very High about 15 minutes.

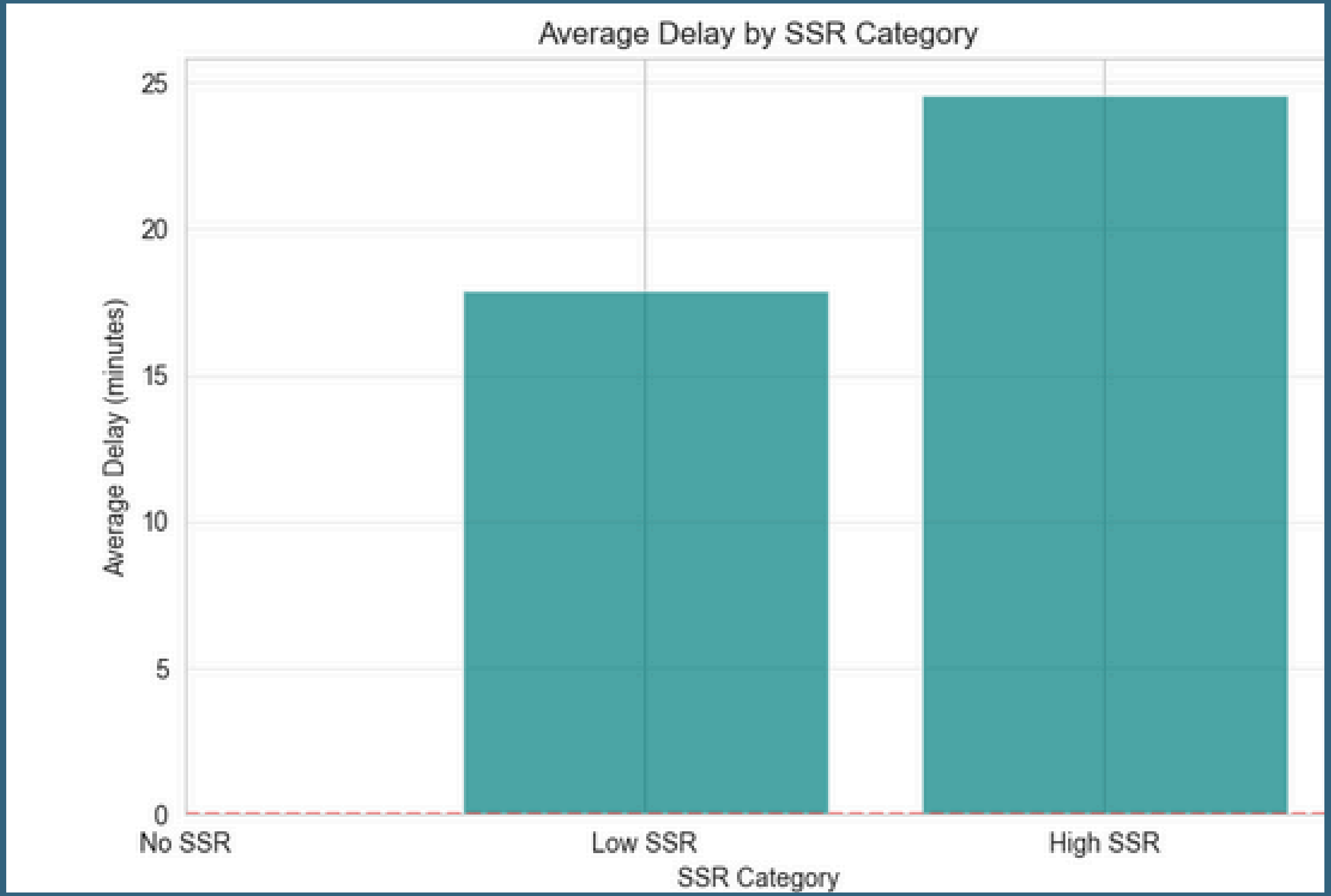
Result (from this analysis run):

- **Average load factor: 102.42%**
- **Correlation (load vs delay): weak/near zero**
- **Distribution of flights heavily weighted >90% LF**

SSR vs delays controlling for load



Most flights have between 10 to 70 special service requests, with peak frequency near 20 requests, and the number of flights sharply decreases beyond 100 requests.



Flight data shows a 50% jump in delay time for high special service requests (appx. 24 min) as compared to low request frequency (appx. 16 mins)

Result (aligned with the approach):

- **High SSR flights show higher delays even within the same load bands**

Summary

Deliverable 1: Exploratory Data Analysis

Flight Operations Performance Metrics | August 2025 | 8,099 Flights Analyzed

DEPARTURE DELAYS

Avg Delay	**21.2 min**
Late Flights (>15m)	**2,163**
Late %	**26.7%**

GROUND TIME RISK

At-Risk Flights	**652**
At-Risk %	**8.05%**
Below Minimum Turn	**High Risk**

BAGGAGE COMPLEXITY

Transfer:Checked	**3.72x**
Hot Transfer %	**11.9%**
Connection Load	**Very High**

KEY INSIGHTS

Load Factor Analysis

102.4% average | Weak correlation with delays | High utilization standard

Special Service Requests

High SSR flights show +7.71 min more delay (controlling for load) | Resource planning critical

Operational Risk Profile

26.7% of flights delayed >15min | 8% tight turnarounds | Transfer-heavy baggage operations

8,099	**Aug 1-15, 2025**	**290,121**	**397,124**	**1.04M**
Total Flights	Analysis Period	Checked Bags	Transfer Bags	Total Passengers

Source: Deliverable-1.ipynb | Flight Operations Analytics

Deliverable 2:

Build a systematic daily-level scoring approach that resets every day. Below are the outputs required:

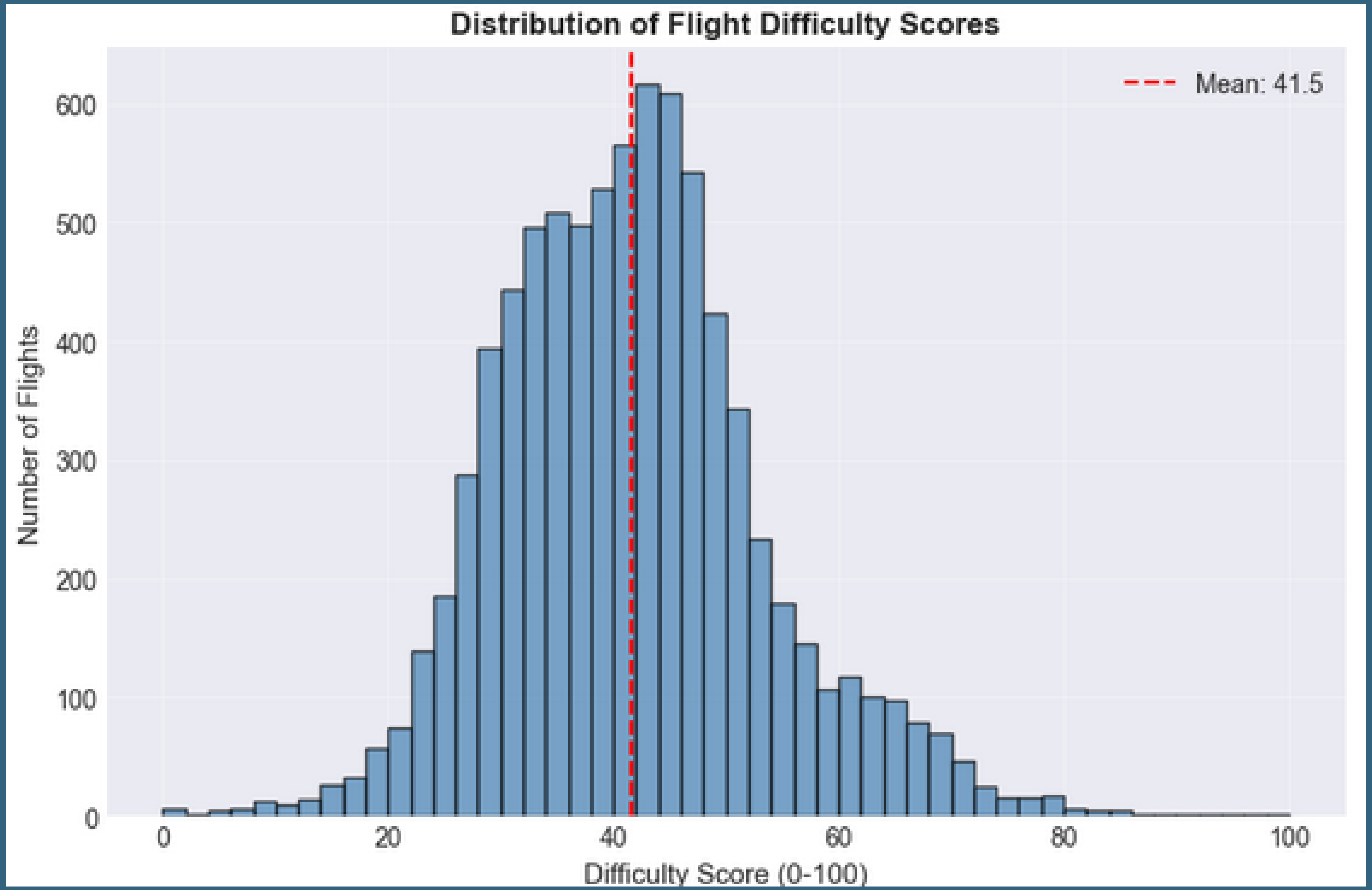
- Ranking: Within each day, order flights by their difficulty score so that the highest-ranked flights represent the most difficult to manage.
- Classification: Group flights into three categories — Difficult, Medium, or Easy — based on rank distribution.

Flight Operational Difficulty Scoring

Objective:
 Quantify flight-level operational difficulty using schedule, passenger, baggage, and service data to improve turnaround, staffing, and planning.

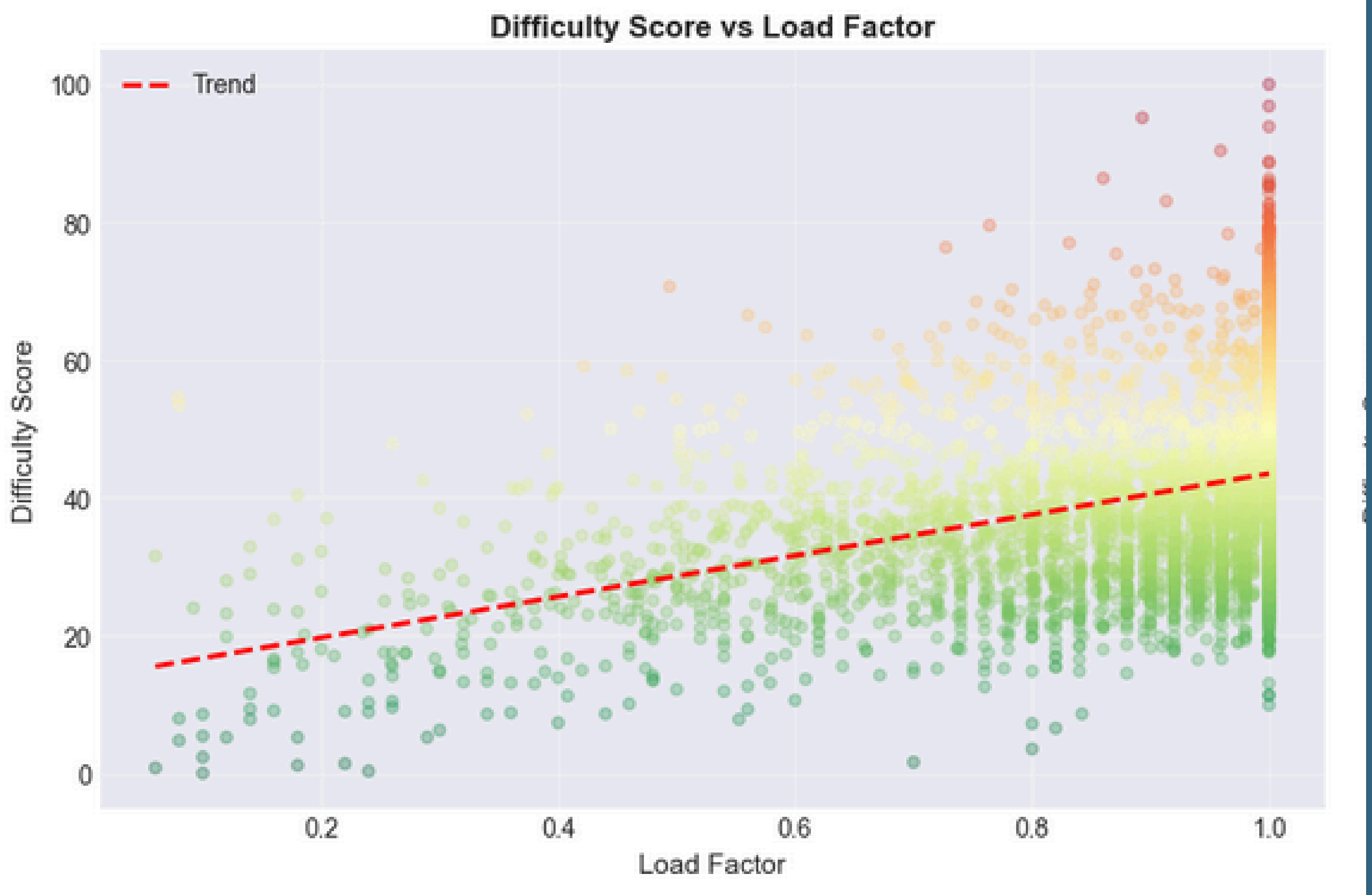
Data Inputs (≈1M rows):

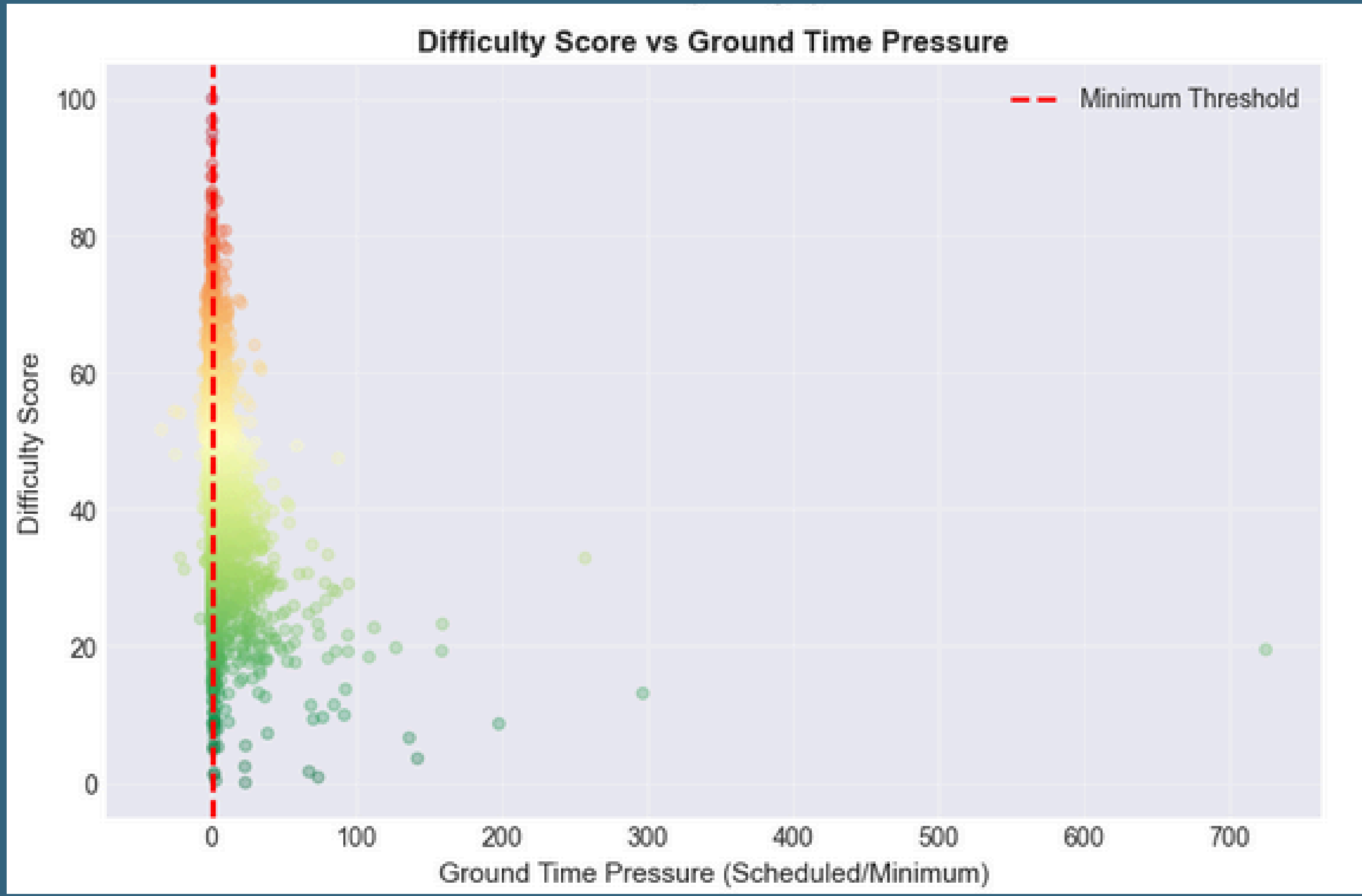
- Flights (8,099)
- PNR Flights (687k), PNR Remarks (52k)
- Bags (687k), Airports (5.6k)



The graph of flight difficulty score versus the number of flights gives us roughly a bell curve, with a sharper decline on the right side of the curve, suggesting that majority of the flights are moderately difficult, with a bias towards easy flight management.

Key Results

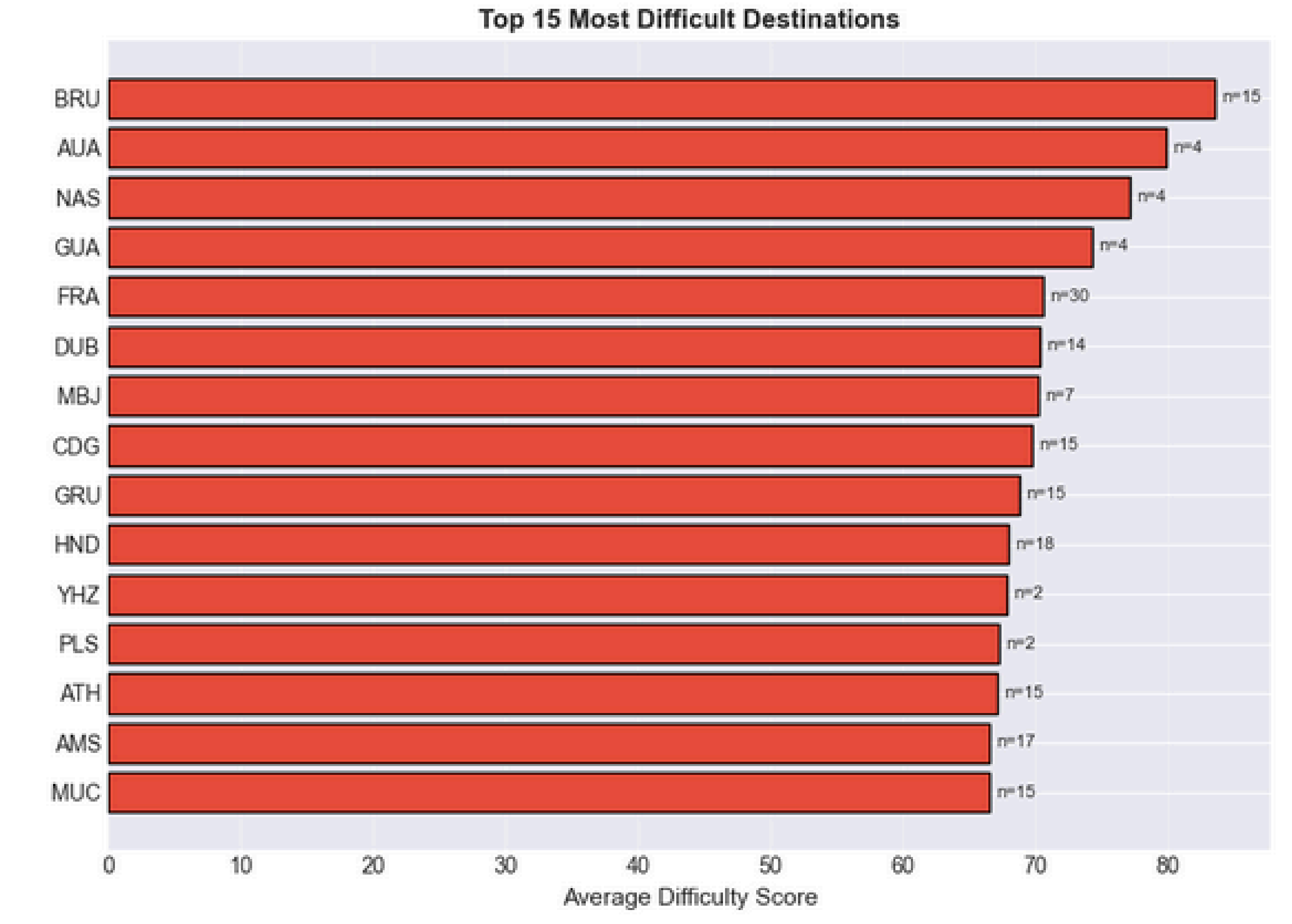


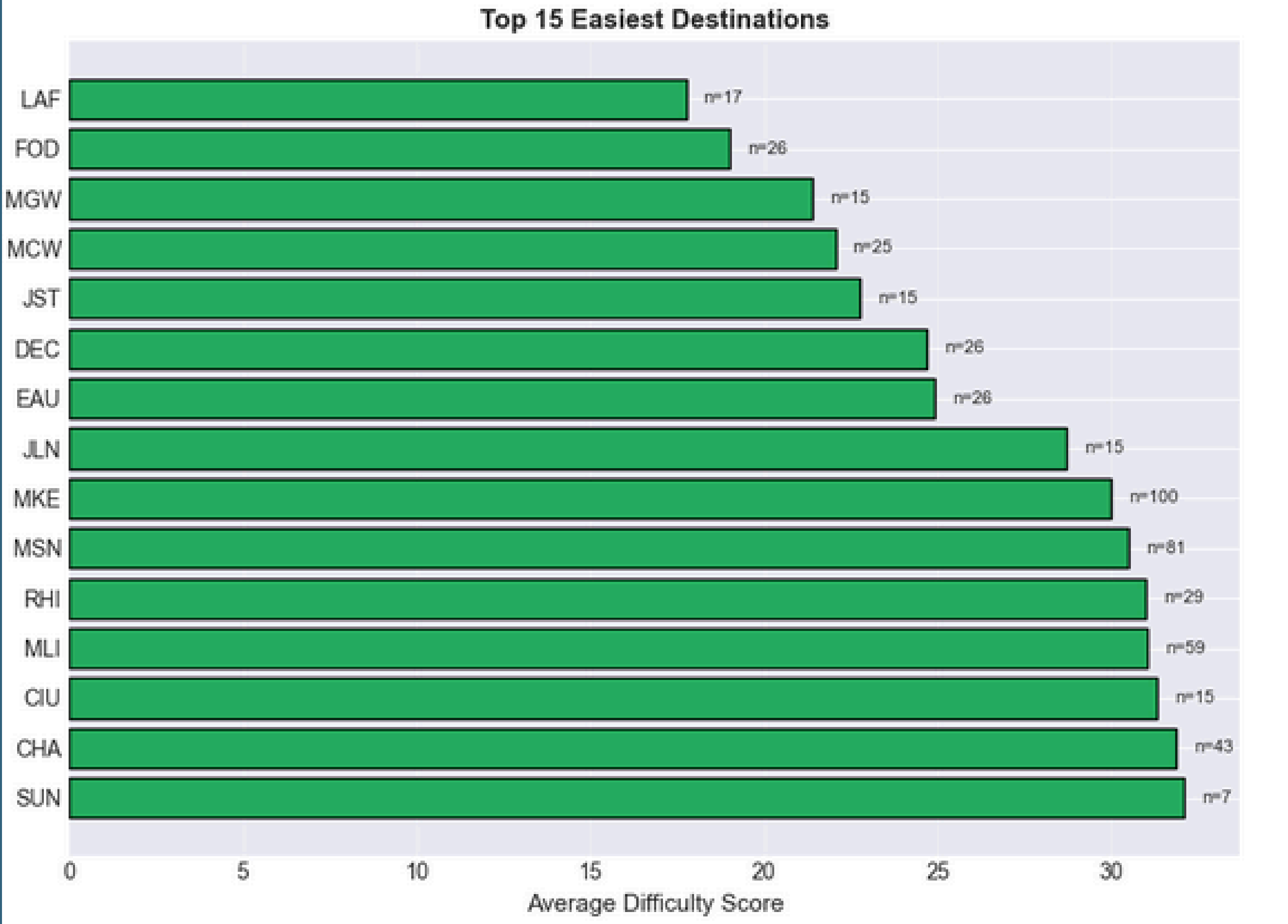


The graph shows that as Load Factor increases, the Difficulty Score generally rises, indicating a positive correlation. Higher Load Factors tend to be associated with higher and more variable Difficulty Scores. The trend line highlights this upward relationship. Overall, it suggests increasing difficulty with increasing load.

The graph shows most Difficulty Scores are high when Ground Time Pressure is near zero. As Ground Time Pressure increases, Difficulty Scores are generally lower and more spread out. The red dashed line marks the minimum threshold, emphasizing the concentration at low values. This suggests difficulty is greatest when Ground Time Pressure is very low.

Key Results



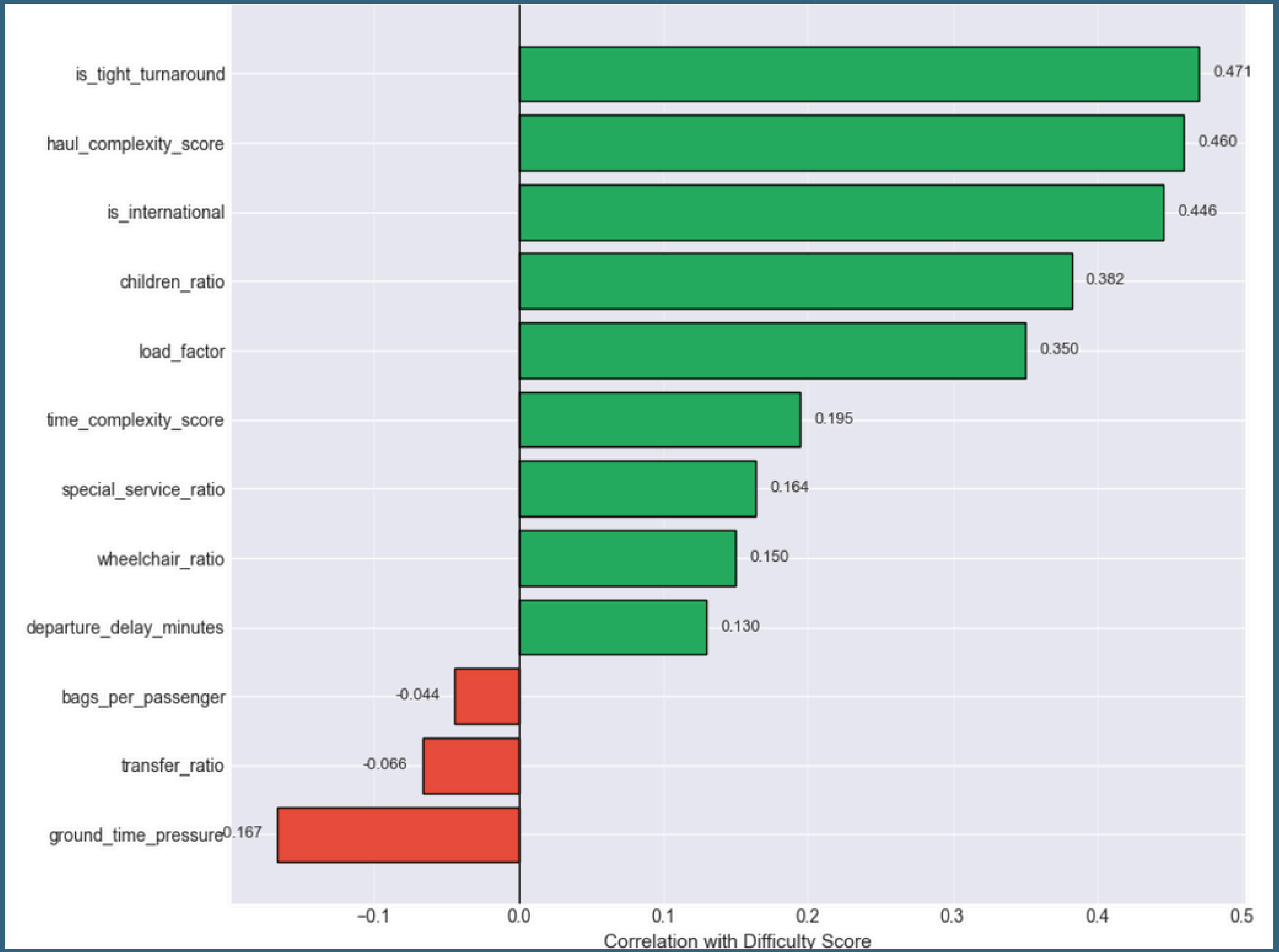


- The easiest destinations likely showcase factors such as efficient processes, favorable conditions, or fewer operational barriers.
- The hardest destinations might indicate locations with more constraints, higher risk, or challenging operational environments.
- This comparison can reveal opportunities for improvement, best practices to adopt, and strategies to mitigate risks in difficult locations.
- Such insights help prioritize resource allocation, training, and support for consistently difficult places while leveraging strengths from easier ones.

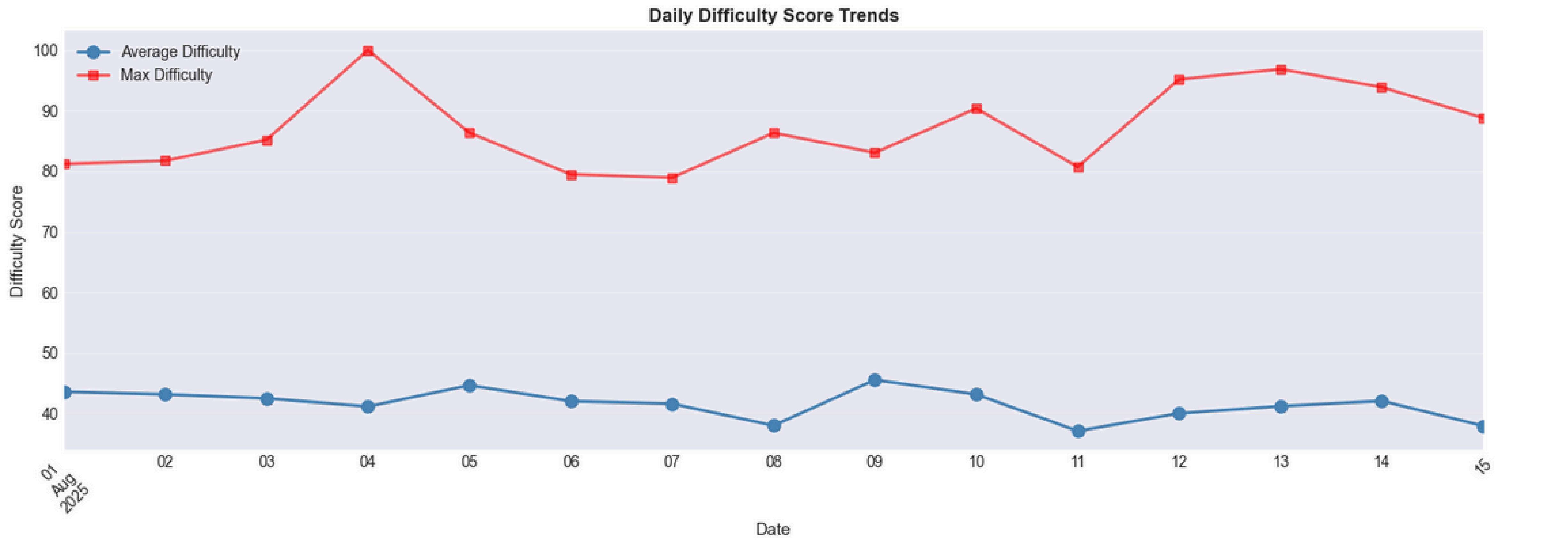
Flight Operational Difficulty Scoring

Methodology

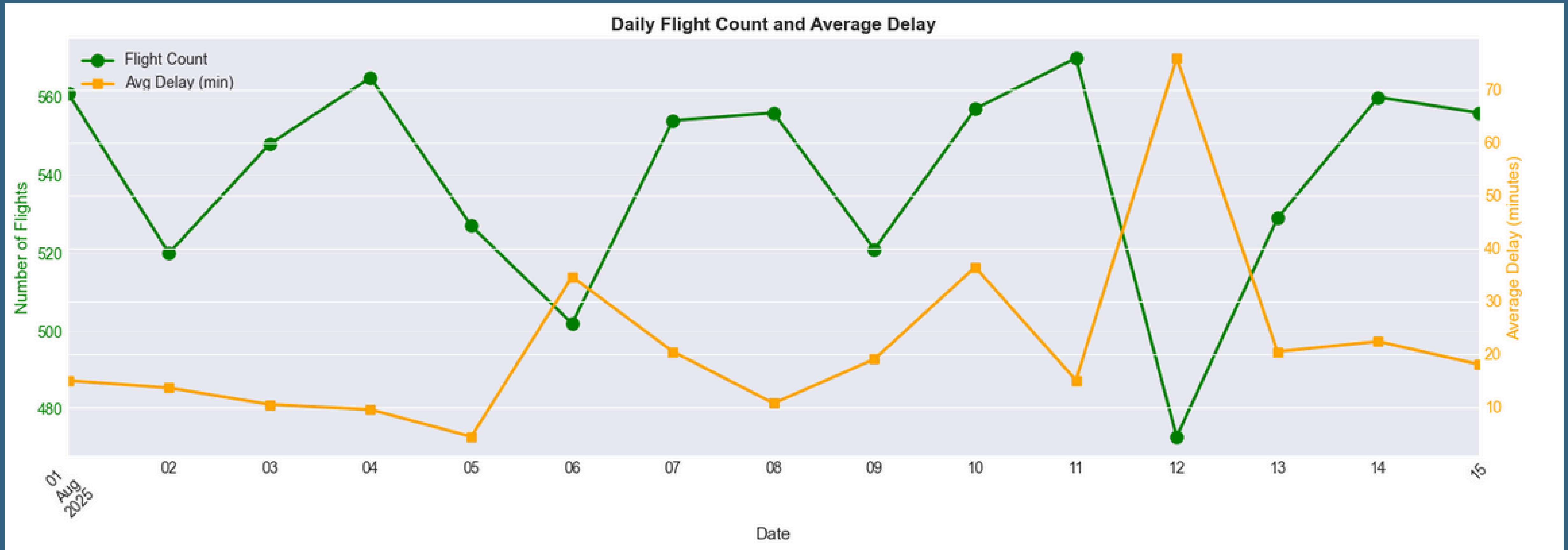
1. Build flight dataset: Standardize times, create flight_id.
2. Compute metrics: Delay, haul/fleet/time/carrier complexity, international flag, ground-time pressure.
3. Aggregate:
   Bags: Origin, transfer, hot ratios.
   PNR: Passengers, children, SSR, wheelchairs.
4. Score:
   Normalize features by day.
   Weighted sum → 0–100 Difficulty Score.
   Classify daily: Easy (40%), Medium (35%), Difficult (25%).



Feature Importance: Correlation with Difficulty score:
The correlation analysis identifies which operational factors most strongly influence the flight difficulty score. Higher correlation means stronger association with operational complexity and potential risk of delays or service strain.



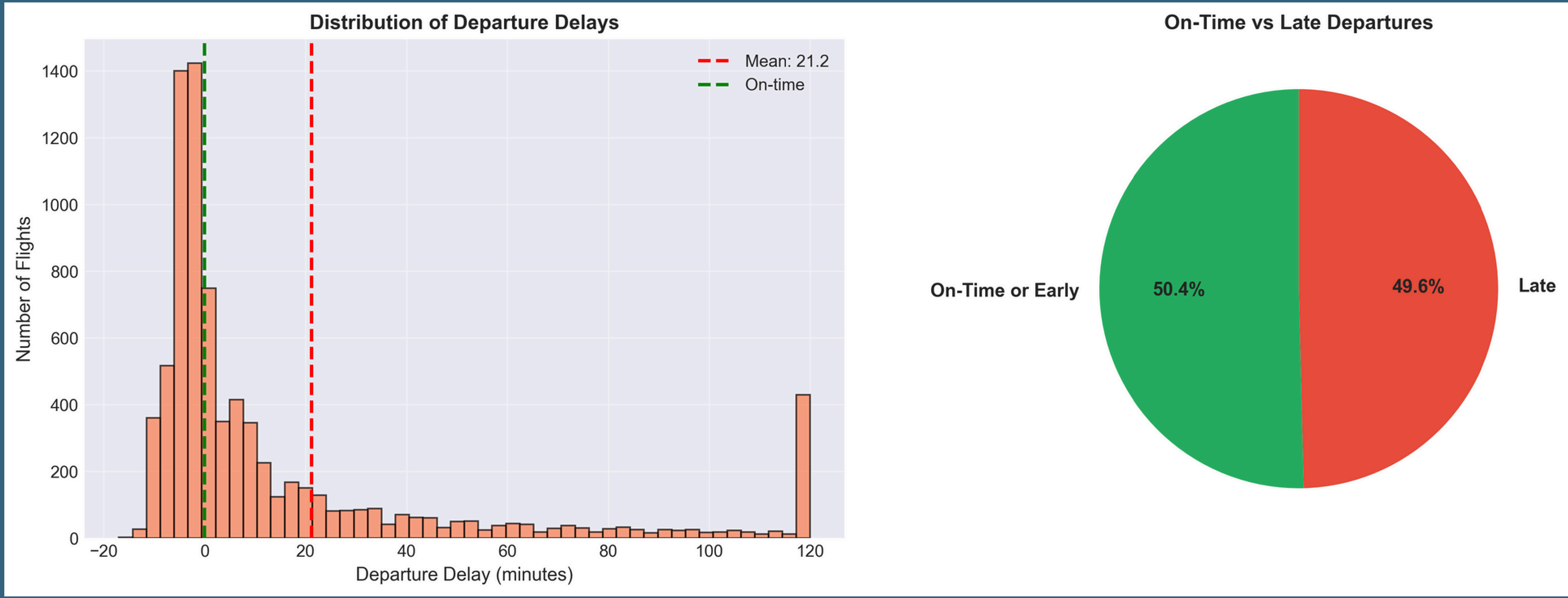
The average difficulty stays fairly stable, hovering between 38 and 46 — not much variation.
The max difficulty fluctuates significantly, with a clear spike around August 4 (near 100), and a few other peaks around August 10–13.
Overall, the max difficulty seems to vary independently of the average difficulty — the spikes in max difficulty don't align with big changes in the average.



The flight count generally ranges between 500–560 flights per day, with one sharp drop around August 12 (close to zero).
The average delay is mostly low (under 30 minutes) but spikes sharply on August 12, reaching over 70 minutes.
That day (August 12) shows an inverse relationship — very few flights but very high delays.

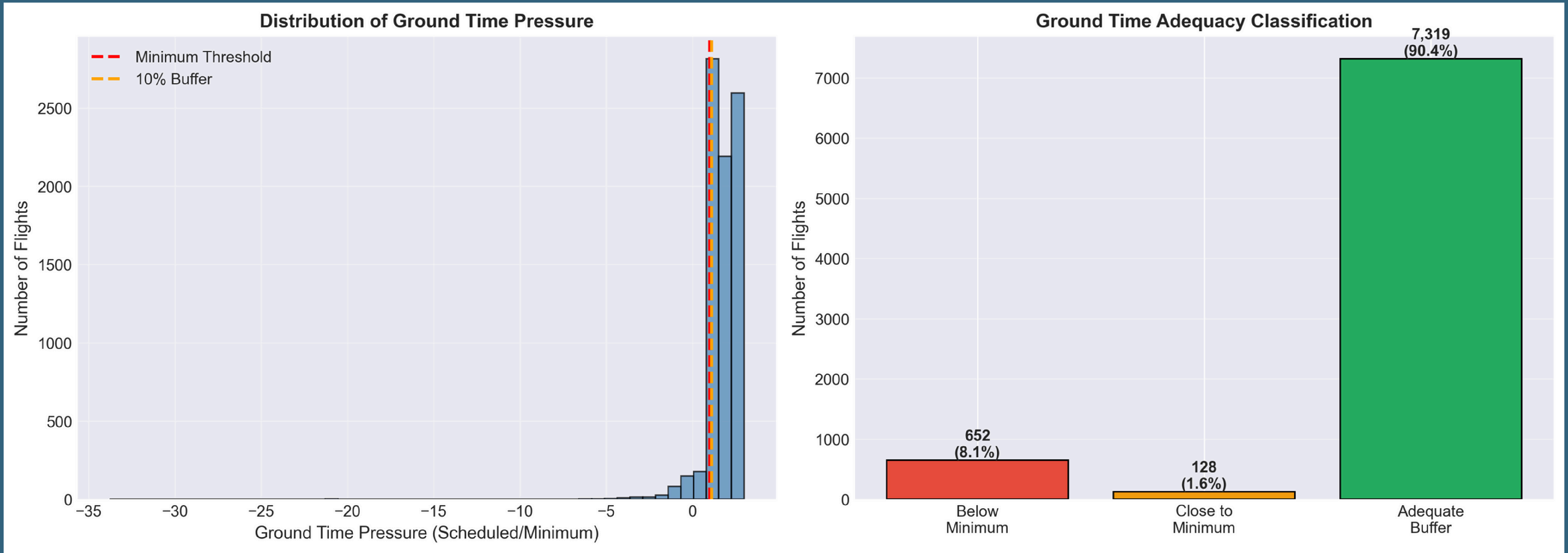
The flight count generally ranges between 500–560 flights per day, with one sharp drop around August 12 (close to zero). The average delay is mostly low (under 30 minutes) but spikes sharply on August 12, reaching over 70 minutes. That day (August 12) shows an inverse relationship — very few flights but very high delays.

Special Service Requests vs Delays (Load-Controlled)



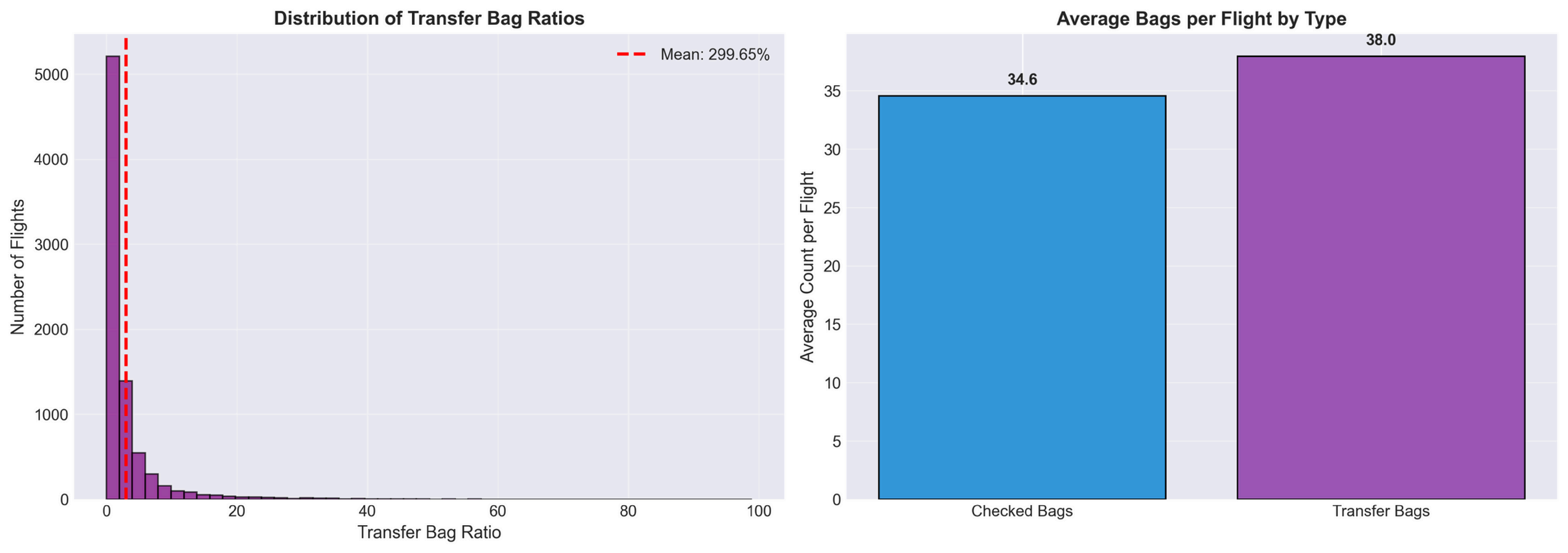
Flights with high special service requests (top 25%) experience 7.71 minutes more delay on average, even after controlling for passenger load factors. The impact is most pronounced in low-load scenarios (<60% load) where high-SSR flights face 57.5 additional delay minutes. Across all load categories, high-SSR flights consistently show elevated difficulty scores, indicating the need for dedicated ground staff to handle wheelchair assistance and mobility services efficiently.

Ground Time Pressure Analysis



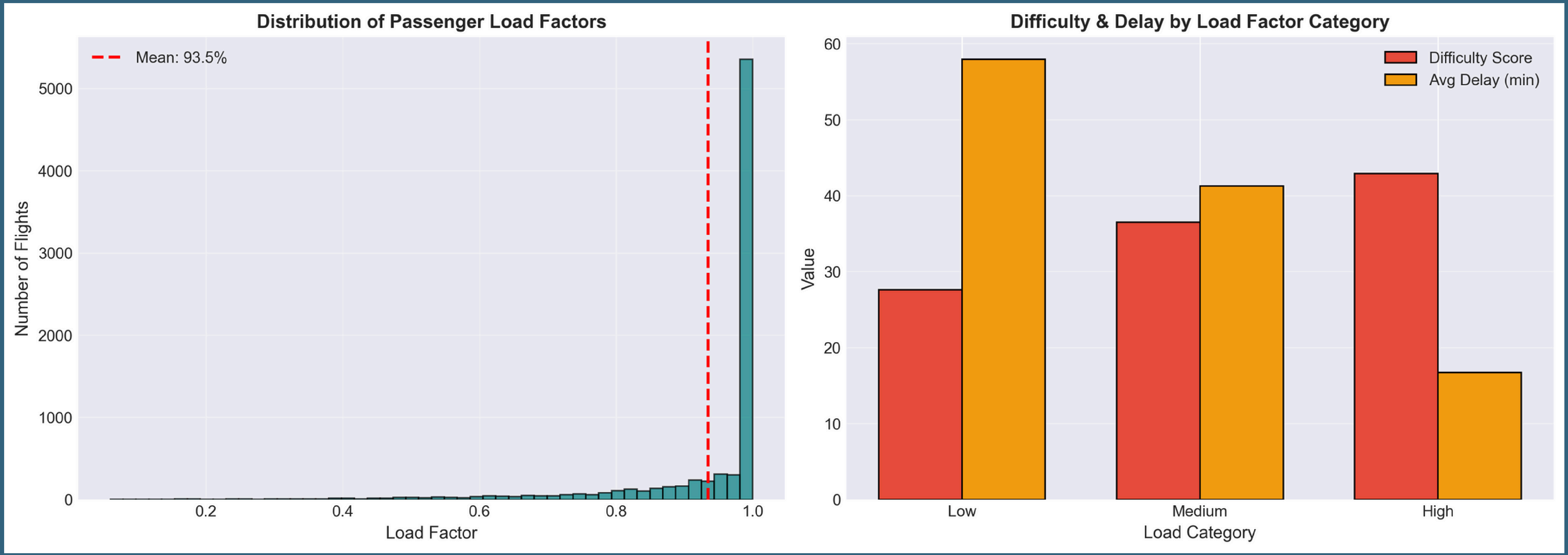
Ground time analysis identifies that 8.1% of flights operate below minimum turnaround thresholds, creating operational risk. The ground time pressure ratio (scheduled/minimum) averages 3.96, with 652 flights having insufficient buffer time. Most flights (90.4%) maintain adequate buffer, but the 9.6% at-risk flights require enhanced ground crew coordination and priority gate assignments to prevent cascading delays.

Transfer Bags Analysis



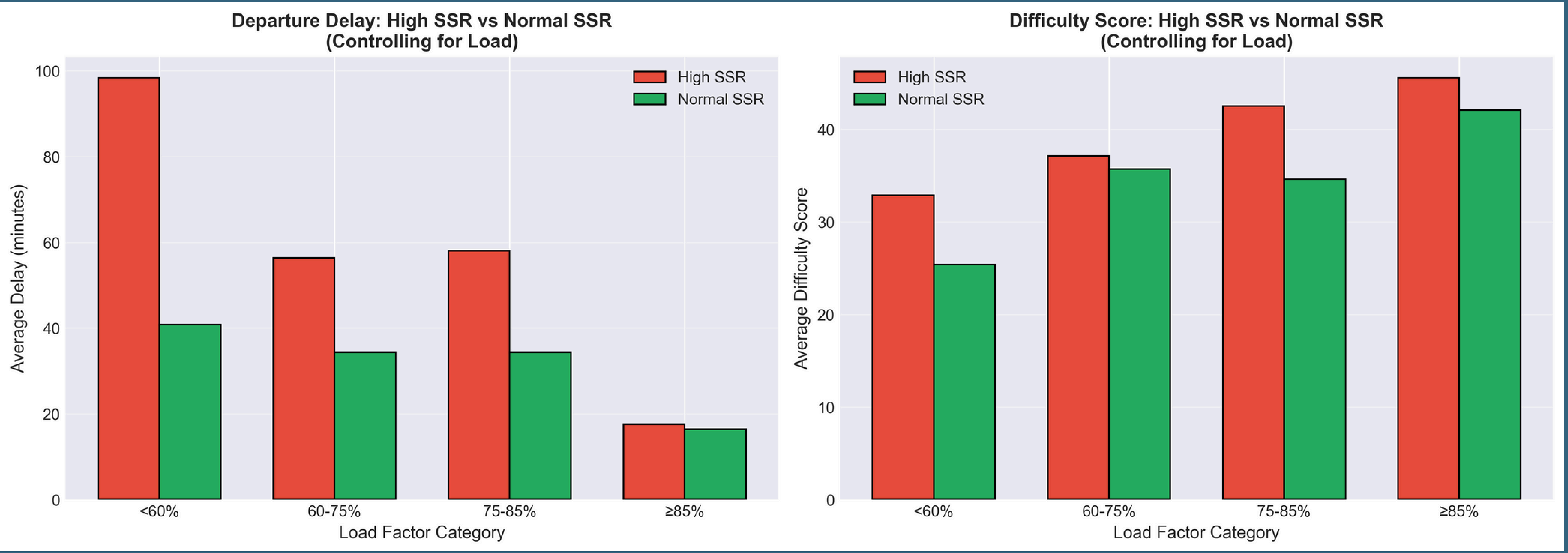
Transfer bags significantly outnumber checked bags, with an average ratio of 299.65%, indicating ORD's role as a major connection hub. On average, each flight handles 38 transfer bags compared to 34.6 checked bags, totaling 78.3 bags per flight. The high transfer volume (80% of flights have >50% transfer ratio) demands specialized baggage handling resources and efficient transfer coordination systems.

Passenger Load Factor Analysis



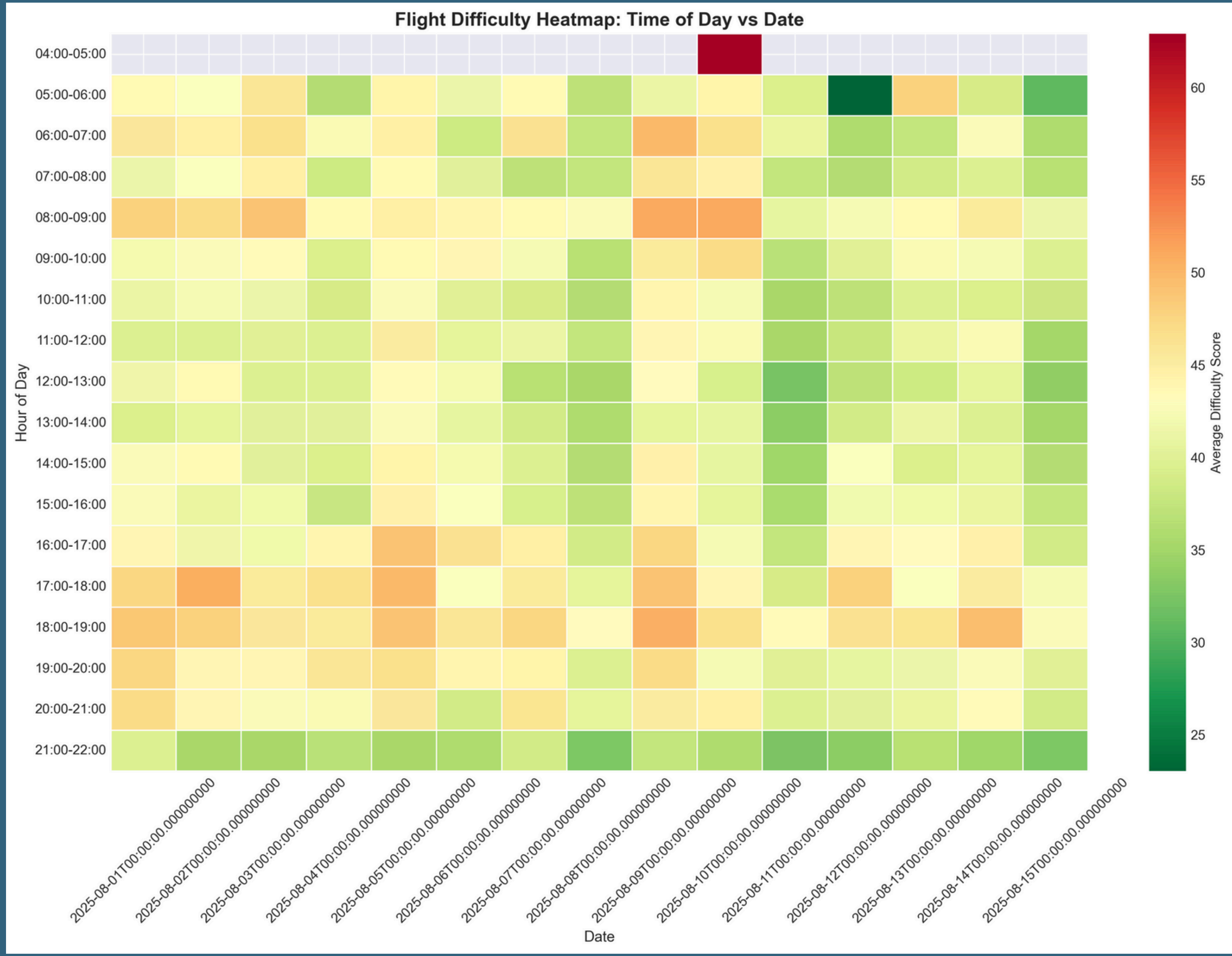
The airline operates with exceptionally high load factors, averaging 93.5% with 84.8% of flights exceeding 85% capacity. Higher load factors positively correlate with operational difficulty (correlation: 0.350), as fuller flights require more boarding/deplaning time and baggage handling. Interestingly, high-load flights show lower average delays (16.7 min) compared to low-load flights (57.9 min), suggesting better operational efficiency with fuller aircraft.

Departure Delays Analysis



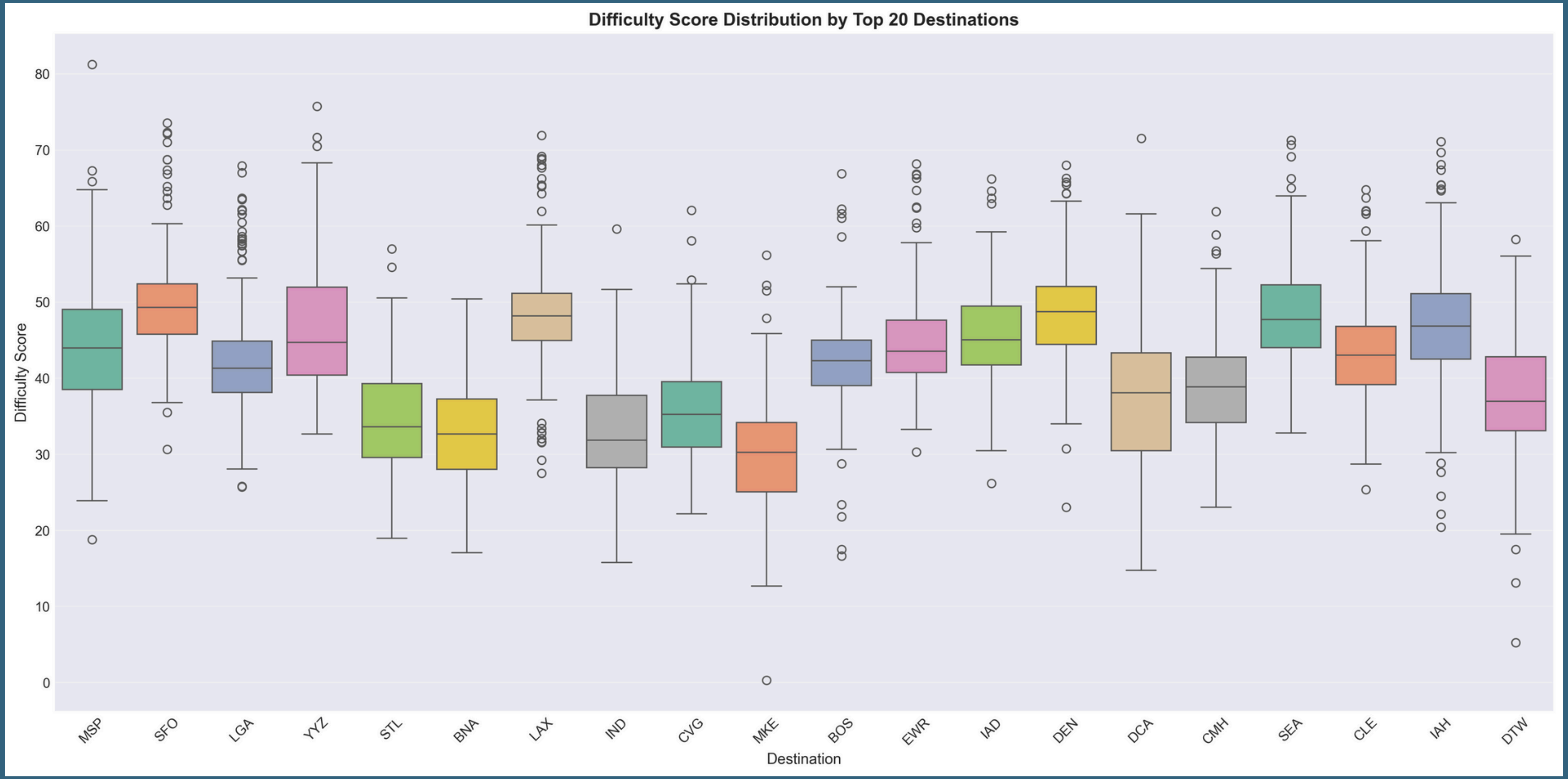
Analysis reveals that 49.6% of flights experience departure delays, with an average delay of 21.2 minutes across the dataset. The distribution shows most flights cluster around on-time performance, but a significant tail extends to 60+ minute delays. Approximately 26.7% of flights are delayed by more than 15 minutes, indicating operational challenges that require attention during peak congestion periods.

Flight Difficulty Heatmap: Time of Day vs Date



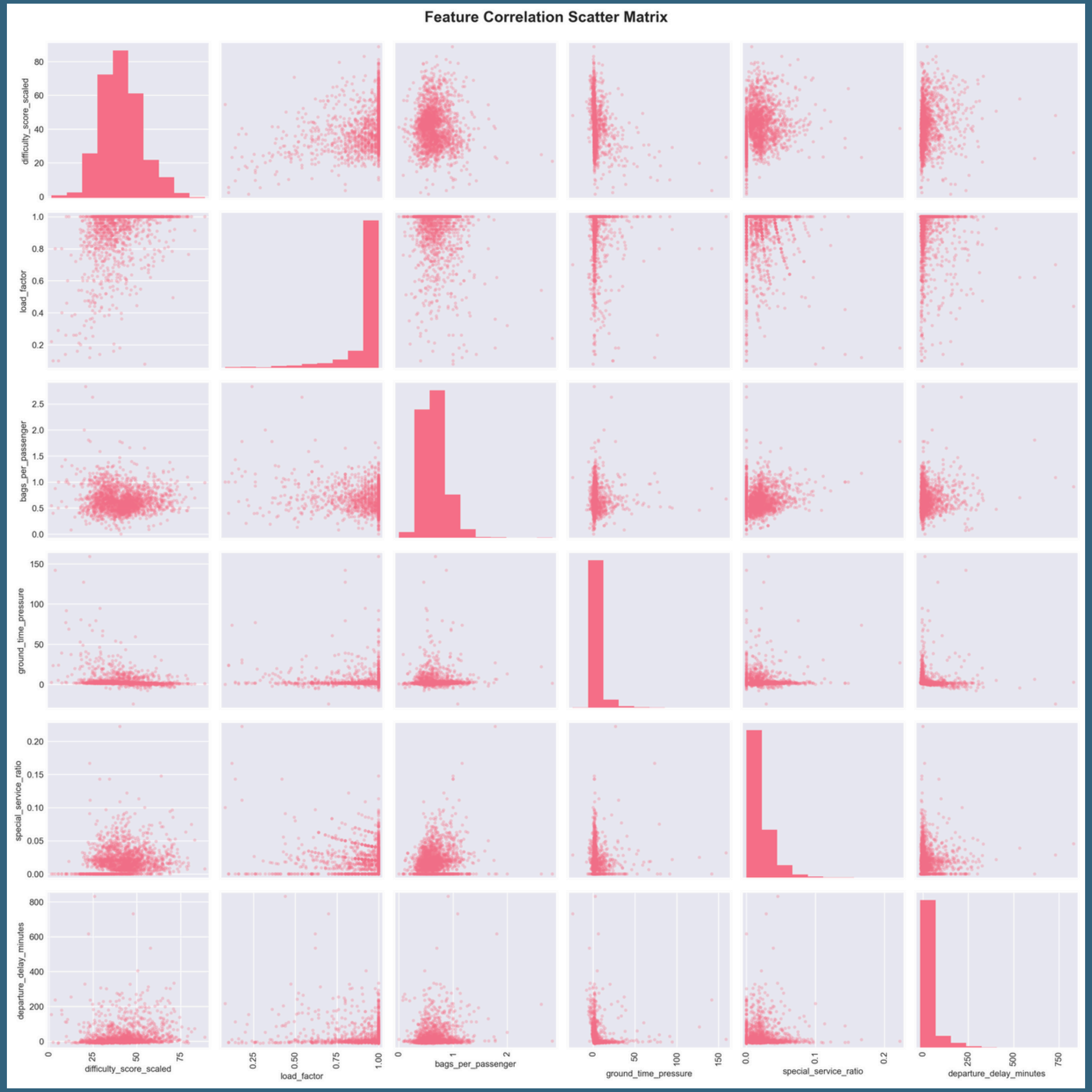
The hourly heatmap identifies early morning (4:00-8:00 AM) and evening departure windows (17:00-19:00) as the most operationally challenging periods, with difficulty scores peaking at 55-60. Mid-day operations (10:00-15:00) and late-night flights (21:00-22:00) consistently show lower difficulty (30-40), offering opportunities for crew training and buffer recovery. Certain dates (Aug 4th, Aug 9th) display elevated difficulty across multiple time slots, indicating system-wide operational stress requiring enhanced coordination.

Difficulty Score Distribution by Top 20 Destinations



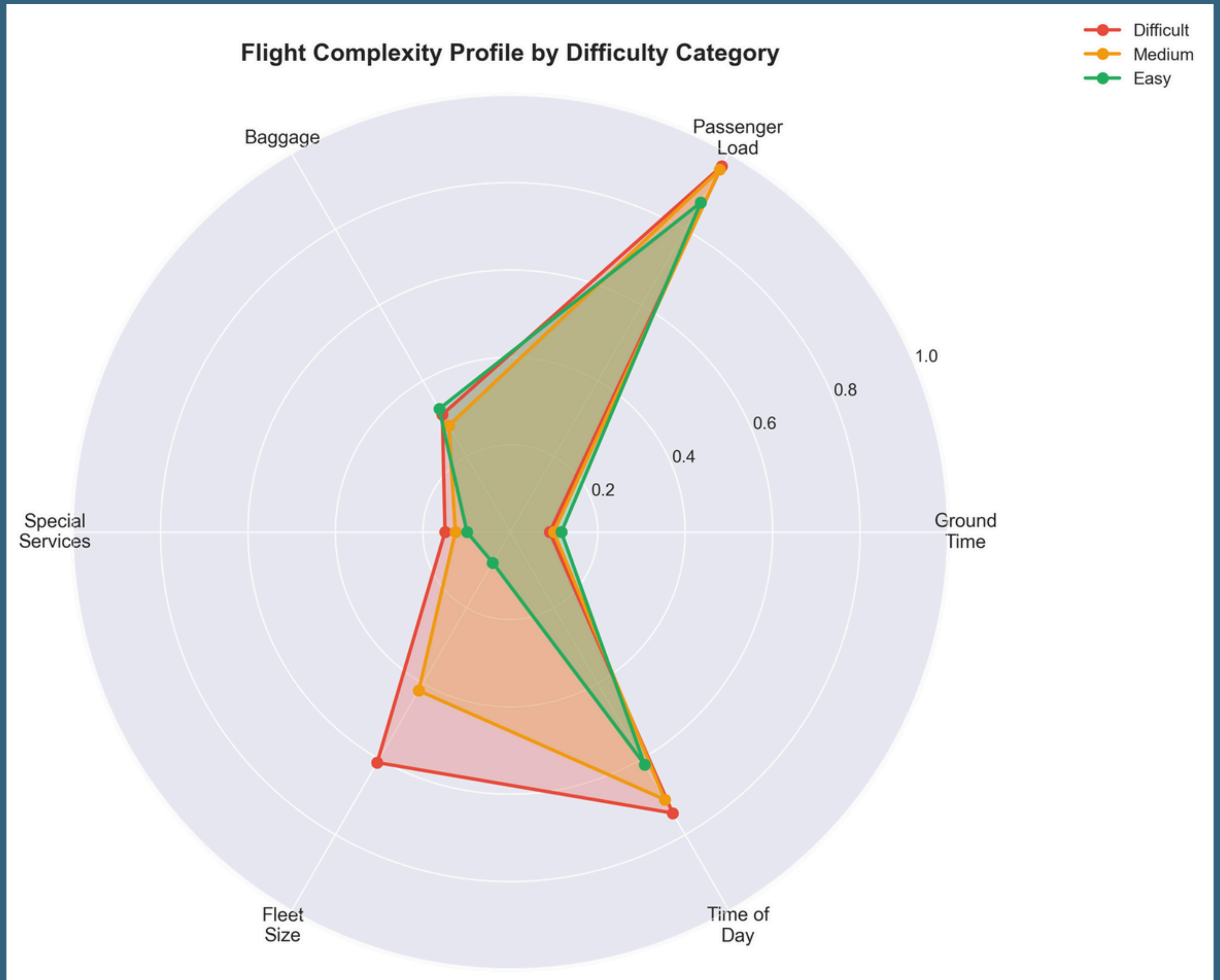
Destination-level analysis shows significant variability in operational difficulty, with high-frequency domestic routes (MSP, SFO) displaying consistent mid-range difficulty scores around 45-50. International and smaller regional destinations exhibit wider score distributions and higher outliers, reflecting schedule variability and special handling requirements. The box plot reveals that even within the same destination, difficulty scores can vary by 30-40 points depending on flight-specific factors like timing, load, and turnaround constraints.

Feature Correlation Scatter Matrix



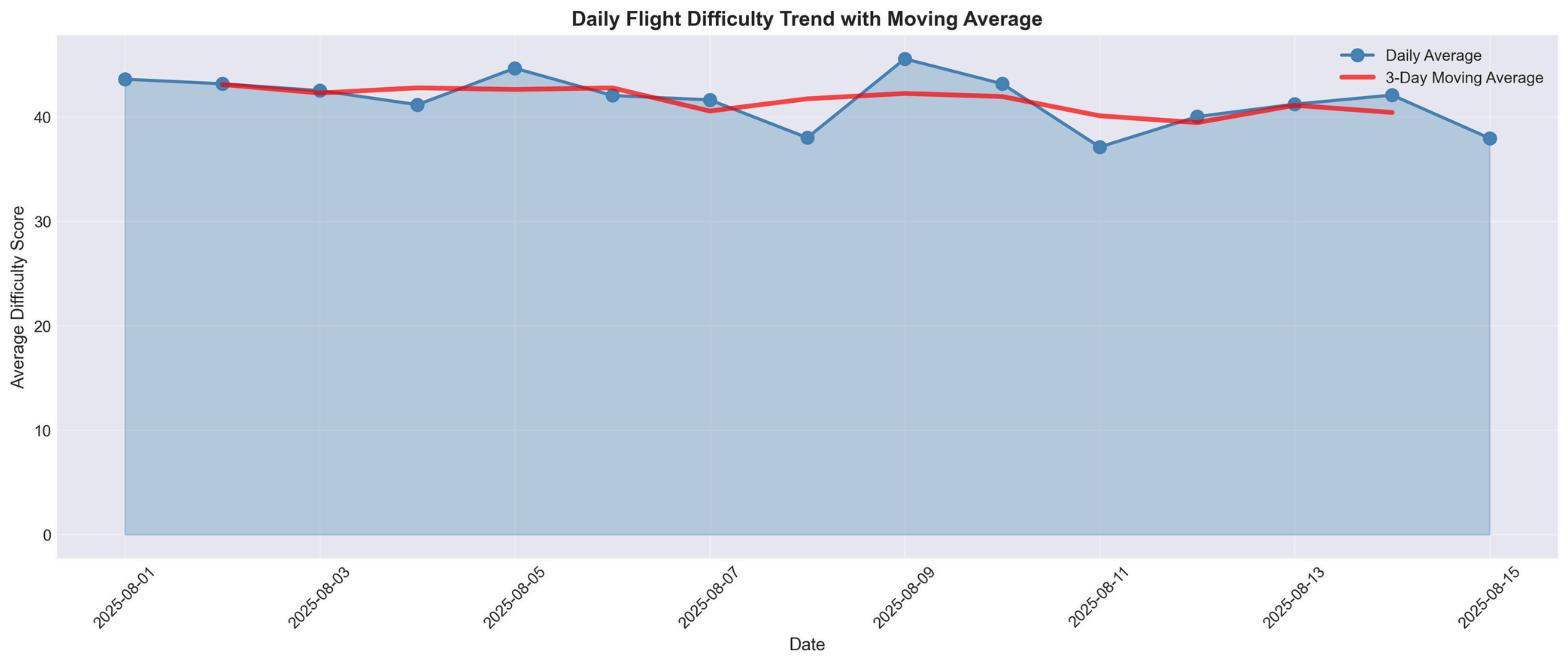
The scatter matrix visualizes pairwise relationships among six key operational metrics, revealing strong positive correlation between difficulty scores and load factor (visible upward trend). Ground time pressure shows inverse relationship with difficulty—flights with higher ground time ratios (more buffer) tend to have lower difficulty scores. Departure delays cluster heavily near zero but show scattered high-delay outliers across all difficulty levels, suggesting delays are influenced by external factors beyond the core difficulty model. The diagonal histograms reveal that most flights operate near 100% load factor and minimal ground time pressure.

Flight Complexity Profile by Difficulty Category



The radar chart reveals distinct operational profiles across difficulty categories, with "Difficult" flights showing peak complexity in ground time constraints, special services, and time-of-day factors. All categories converge on passenger load and baggage metrics, indicating these are universal operational challenges. Easy flights maintain adequate ground time buffers and operate during off-peak hours, while difficult flights face compressed turnarounds and peak-time departures that compound operational pressure.

Daily Flight Difficulty Trend with Moving Average

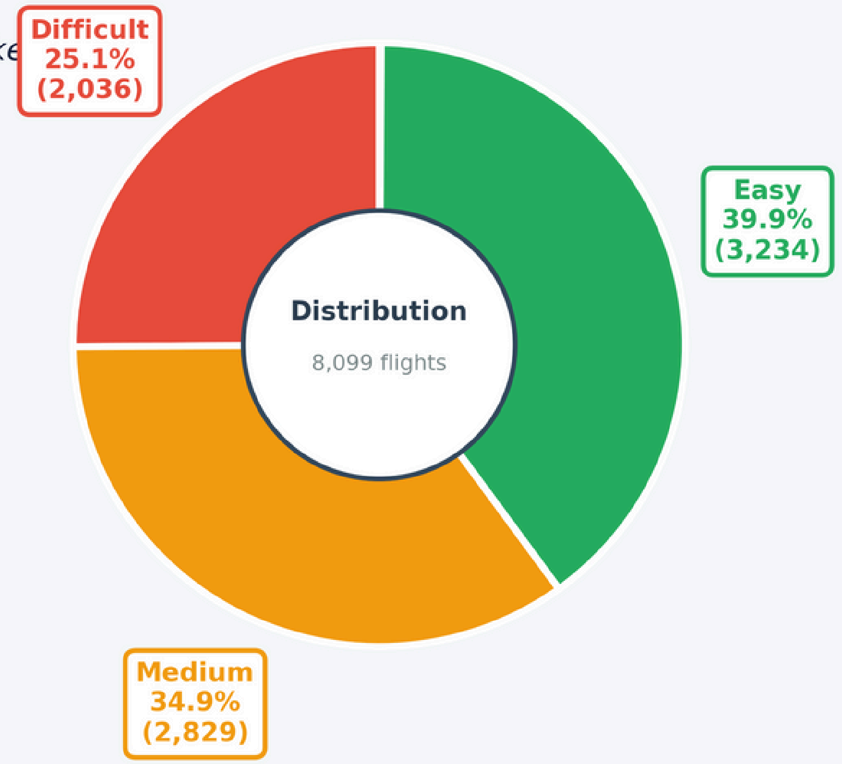


The 15-day operational window shows difficulty scores fluctuating between 37-45, with notable peaks on August 1st, 5th, and 9th coinciding with higher flight volumes and potential weather/operational disruptions. The 3-day moving average (red line) smooths daily volatility, revealing a relatively stable baseline around 41-42 difficulty points. August 8th, 11th, and 15th show dips below 40, suggesting these days had more favorable operational conditions with better ground time buffers and lower service demands.

Summary

Deliverable 2: Flight Difficulty Scoring

Weighted Operational Complexity Analysis | Daily Normalized Scoring | 8,099 Flights Ranke



SCORING METHODOLOGY

- **Ground Time (25%)**
  Turnaround pressure & constraints
- **Passenger Service (30%)**
  Load, SSR, wheelchair, children
- **Baggage (20%)**
  Volume, transfers, hot connections
- **Flight Characteristics (15%)**
  Fleet size, haul type, time of day
- **Additional Factors (10%)**
  International, carrier type, strollers

TOP DIFFICULTY DRIVERS

1. Tight Turnaround
Correlation: 0.471
47.1%

2. Haul Complexity
Correlation: 0.460
46.0%

3. International
Correlation: 0.446
44.6%

OPERATIONAL IMPACT COMPARISON

Difficult Flights
Avg Delay: 37.1 min

Easy Flights
Avg Delay: 15.1 min

Source: Deliverable-2.ipynb | Normalized by day, weighted scoring | Percentile-based classification

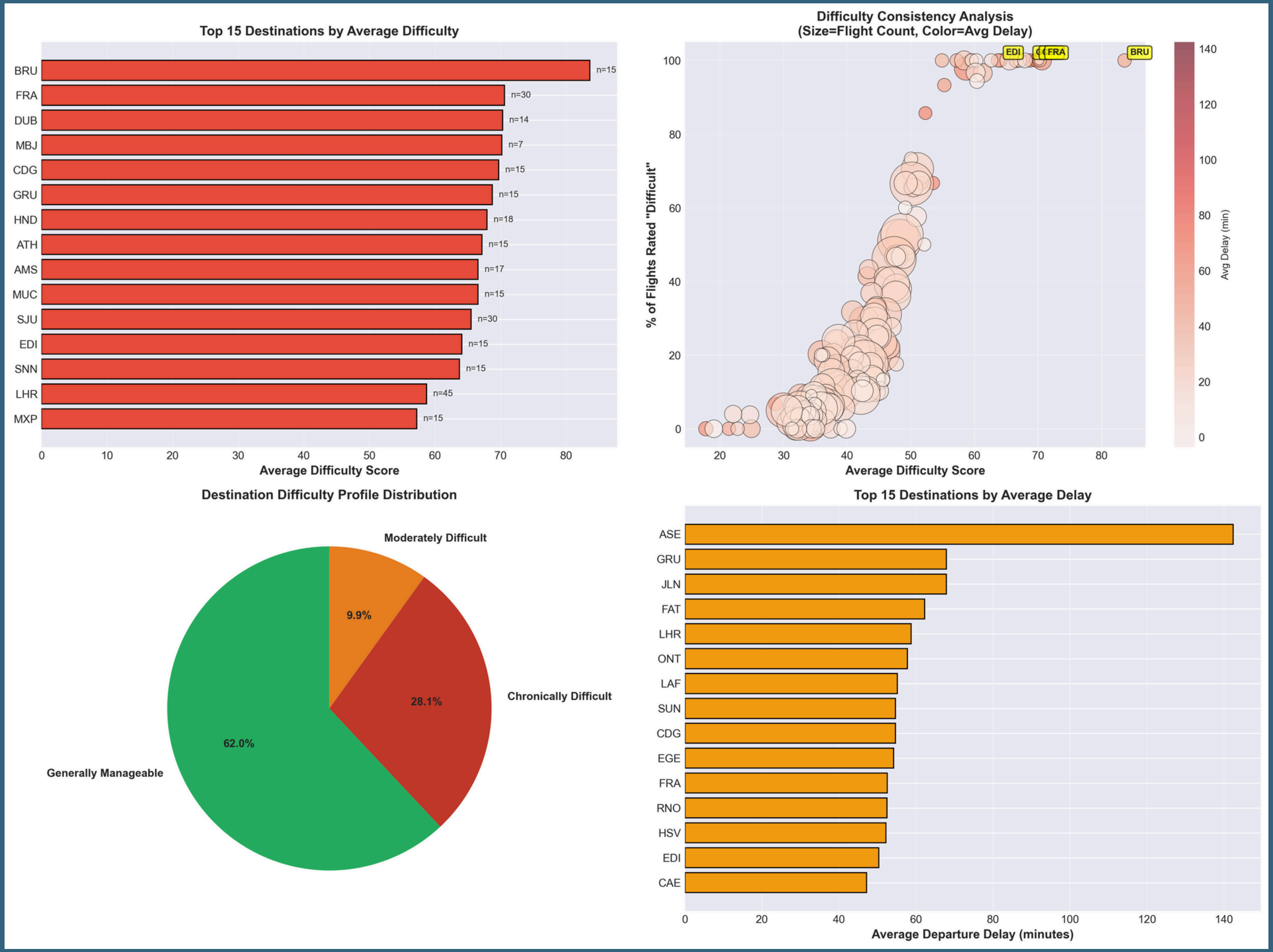
▲ 22.0 min difference

Deliverable 3:

Post-Analysis & Operational Insights

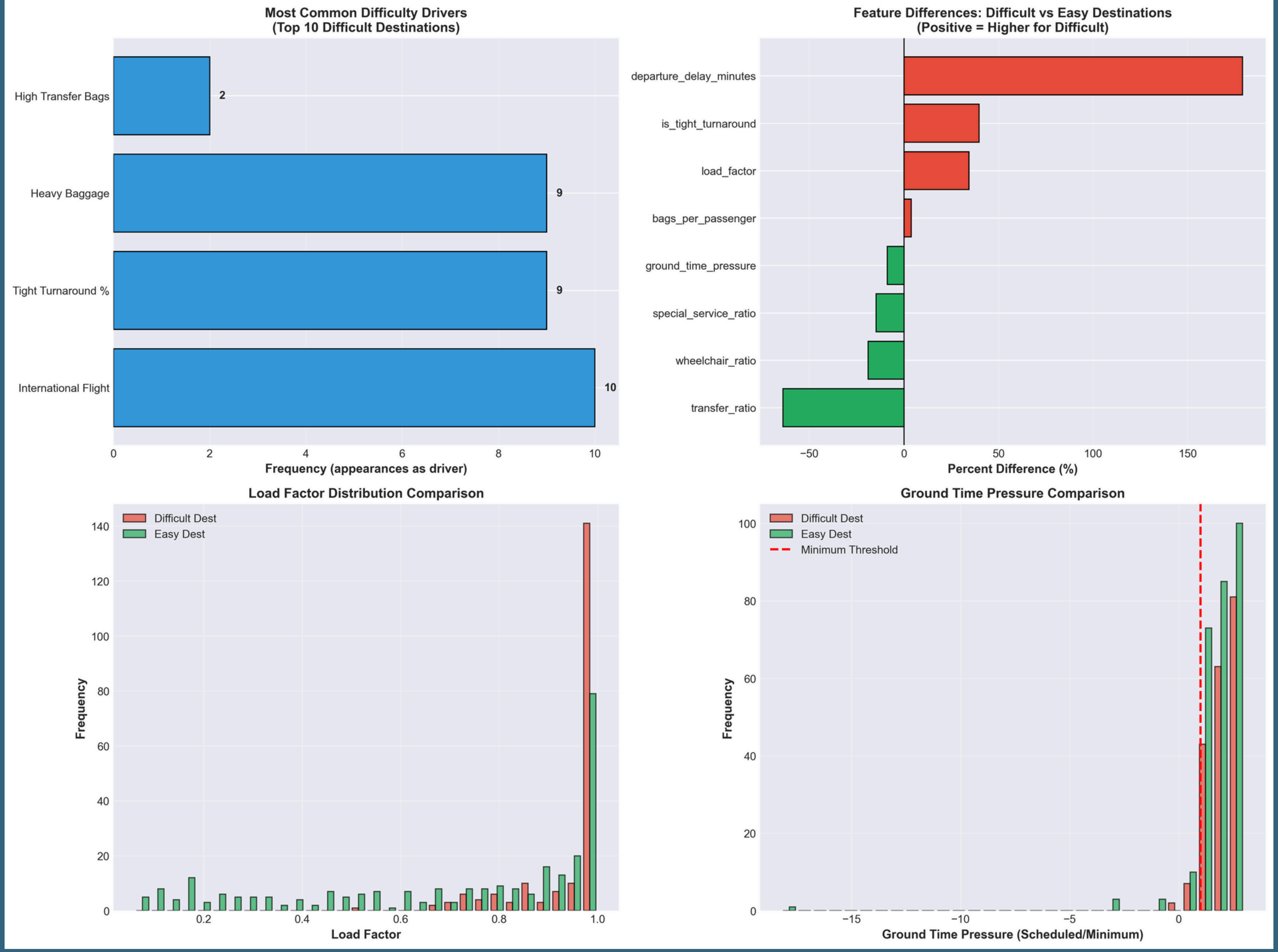
- Summarize which destinations consistently show more difficulty.
- What are the common drivers for those flights?
- What specific actions would you recommend based on the findings for better operational efficiency?

Graph Results



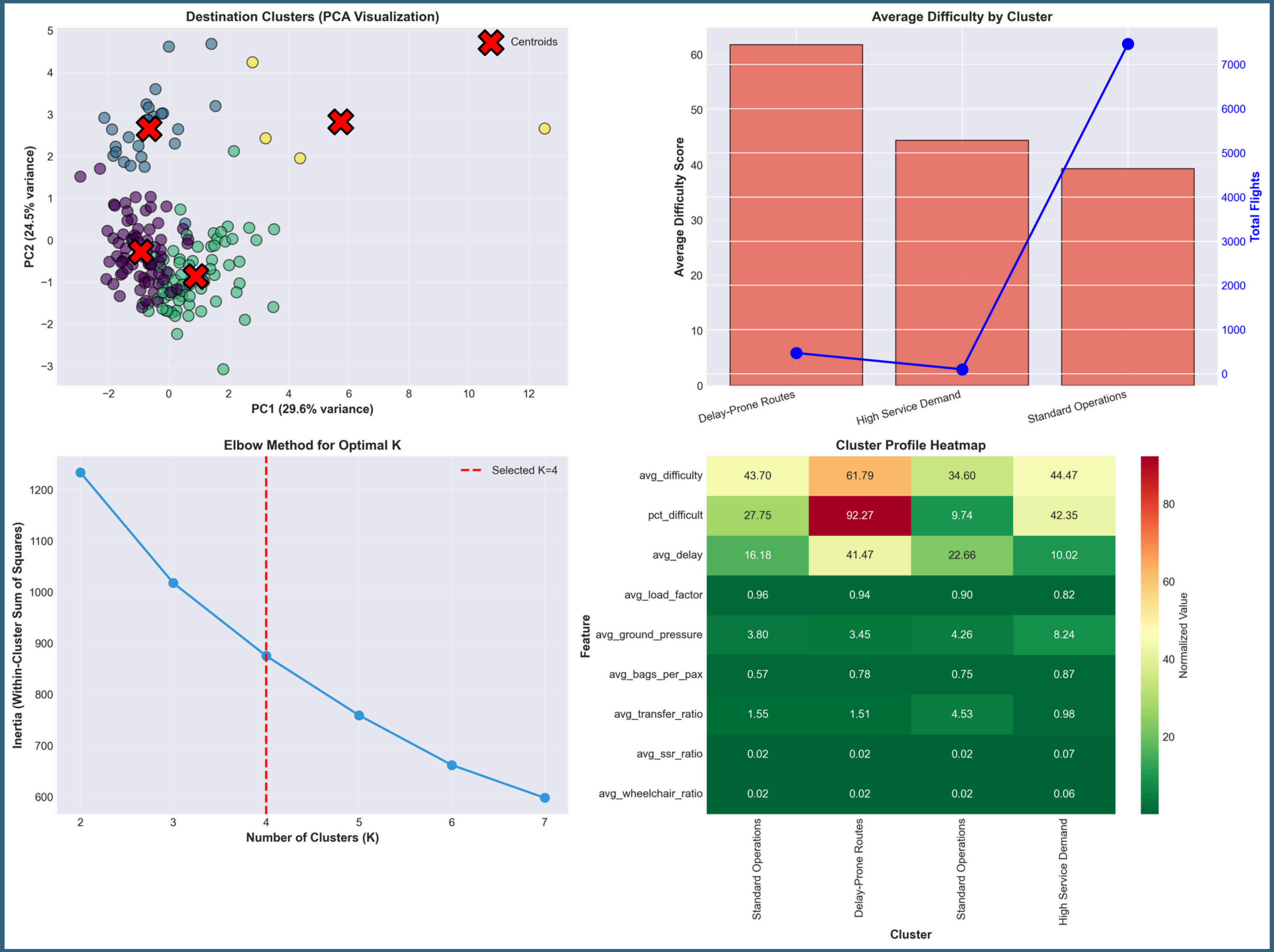
Difficulty is broken down geographically and statistically. Major European hubs (BRU, FRA, DUB) score highest for operational difficulty and are consistently hard across flights, while others face intermittent challenges. The destination profile shows ~30% of locations are chronically difficult — persistent problem points. Correlation between difficulty and delay highlights predictable hotspots. Destinations like ASE and GRU also appear among top delay centers. The insight: difficulty and delay strongly overlap, suggesting unified solutions can target both efficiency and reliability.

Graph Results



This figure examines what makes certain destinations "difficult." The key drivers — international operations, tight turnarounds, and heavy baggage — all point to intensive timing and handling stress. Comparative feature analysis shows that difficult routes have higher departure delays, tighter ground time, and near-max load factors. Ground pressure distributions confirm that minimal buffer time is a core issue. Overall, difficulty stems less from randomness and more from structural operational constraints, signaling that optimizing scheduling and ground operations could yield large gains.

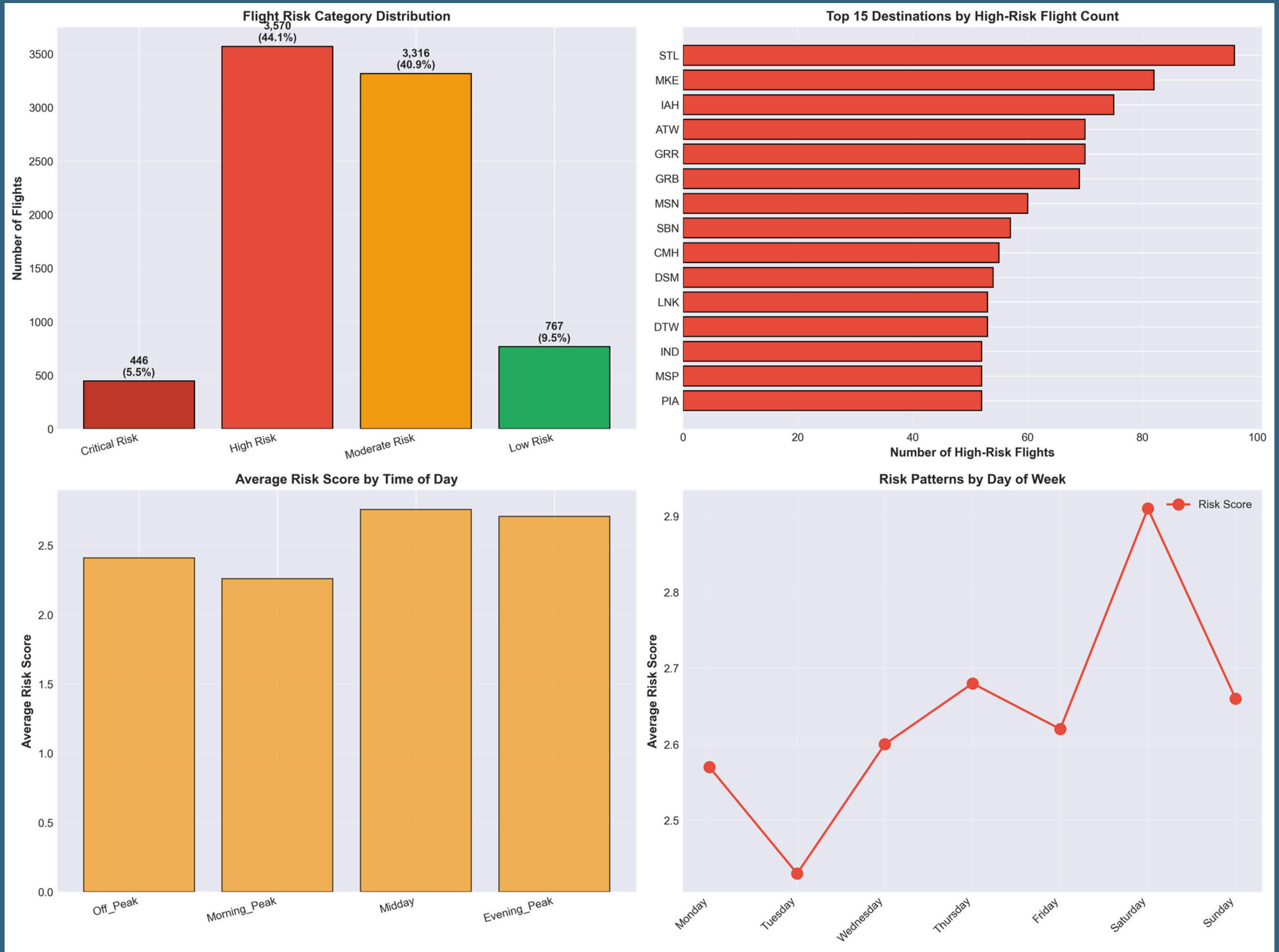
Graph Results



This analysis groups destinations by operational behavior into four clusters:

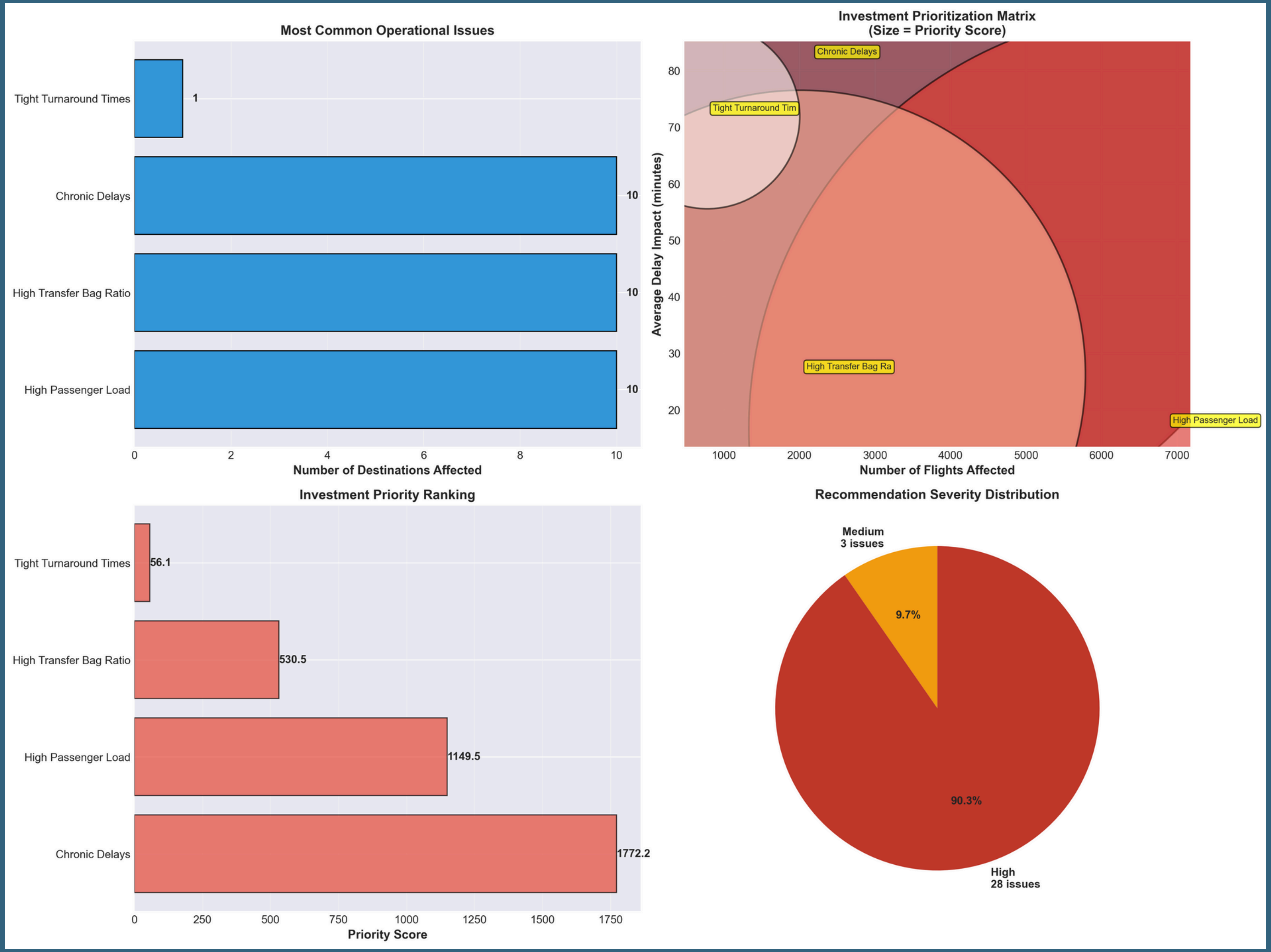
1. Delay-Prone Routes – high difficulty and long delays.
2. High Service Demand – complex but fewer flights.
3. Standard Operations (2 clusters) – stable, efficient routes.
4. Cluster heatmaps confirm these operational signatures, with Delay-Prone routes showing extreme strain. The PCA plot validates clear separability, and the elbow method confirms K=4 as optimal. Broadly, this segmentation reveals where distinct operational strategies should apply — tailored delay mitigation for one cluster, service optimization for another, and monitoring for stable ones.

Graph Results



This dashboard highlights risk concentration across flights. Most operations fall into High or Moderate Risk, with Critical flights forming a small but impactful minority. Risk is highest mid-day and on Saturdays, pointing to schedule-driven congestion. Airports like STL, MKE, and IAH lead in high-risk activity, indicating local operational challenges. The pattern suggests risk is systematic, not random, driven by timing, demand, and specific destination dynamics — implying that targeted interventions could reduce high-risk clusters without broad operational changes.

Graph Results



This view identifies root operational pain points. Chronic delays and high passenger loads dominate both in scale and impact, making them top investment priorities. High transfer bag ratios further amplify delays. The prioritization matrix and ranking confirm where resources will have the most effect. Over 90% of issues are high severity, underscoring urgency. Together, these insights argue for strategic improvements in turnaround efficiency, delay mitigation, and baggage handling to enhance network stability

Summary

Deliverable 3: Operational Recommendations

Data-Driven Action Plan | Destination Segmentation | Investment Priorities | Improvement Roadmap

10	4	237	50%
Problem Destinations	Operational Clusters	Chronic Delay Routes	High-Risk Flights

ANALYSIS PIPELINE

Destination Ranking → **Driver Analysis** → **Clustering (K=4)** → **Risk Scoring** → **Investment Priority**

KEY RECOMMENDATIONS

Immediate Actions (0-30 days)

- Pre-position SSR teams at high-difficulty destinations
- Surge staffing for midday & evening peak periods
- Flag 237 chronic-delay routes for schedule review

Short-Term (1-3 months)

- Implement cluster-specific operational playbooks
- Optimize ground time for tight turnaround flights
- Deploy baggage handling surge capacity

Long-Term (3-12 months)

- Invest in infrastructure at top-10 problem destinations
- Schedule optimization for delay-prone routes
- Predictive staffing models based on difficulty scores

PROJECTED IMPACT: 30% reduction in difficult-flight delays = **22,682 minutes saved**

(15-day period) | Annualized: ~9,200 hours recovered | Equivalent to 383 days of operations

Source: Deliverable-3.ipynb | KMeans clustering | Risk scoring | Investment prioritization matrix

Just like United, our ideas are ready to soar – thanks for flying with us on this journey!