Flight Delay Patterns Analysis: Final Report

Dataset Description

This dataset contains over 7 million rows and 35 columns of domestic flight performance data from 2024, sourced from the BTS TranStats On-Time Performance database. It provides comprehensive information on scheduled and actual flight times, delays, cancellations, diversions, and airport distances. The data was cleaned and standardized with snake_case column names, ISO-formatted dates, binary indicators for cancellations and diversions, and zero-filled missing delay values, making it suitable for exploratory analysis, delay prediction, time series analysis, and airline/airport performance studies.

Research Question 1: How do delays build up throughout the day?

Analysis Approach

Flights were categorized into four time periods based on scheduled departure time: Morning (5:00 AM - 11:59 AM), Afternoon (12:00 PM - 4:59 PM), Evening (5:00 PM - 9:00 PM), and Night/Early Morning (9:01 PM - 4:59 AM). Average delays were calculated for each period, with particular focus on late aircraft delays to track cascade effects.

Results

time_period	num_flights	avg_dep_delay	avg_arr_delay	avg_late_aircraft_delay	avg_carrier_delay	avg_weather_delay	avg_nas_delay
Morning	2,907,959	6.48	0.60	2.73	4.43	0.62	2.07
Afternoon	2,033,975	14.31	9.28	7.00	4.92	0.94	3.36
Evening	1,570,965	19.73	14.77	10.13	6.06	1.24	3.67
Night/Early Morning	452,368	19.32	12.44	8.59	7.58	1.17	2.23

Conclusion

The analysis reveals a clear cascade effect where delays build up throughout the day. Morning flights start with an average late aircraft delay of just 2.73 minutes, but this more than triples to 10.13 minutes by evening. This pattern confirms that when aircraft are delayed on earlier flights, those delays ripple

forward to subsequent departures. Evening flights experience the worst delays overall, with average departure delays of 19.73 minutes and arrival delays of 14.77 minutes. The late aircraft delay metric—which specifically tracks delays caused by the previous flight arriving late—is the key evidence showing how morning disruptions compound into significant delays by the end of the day.

Research Question 2: How do delay patterns differ between hub-to-hub, hub-to-regional, and regional-to-regional routes?

Analysis Approach

A reference table of 36 major hub airports was created based on the primary hubs of United, American, Delta, Southwest, JetBlue, and Alaska Airlines. Routes were classified using LEFT JOINs: hub-to-hub (both origin and destination are hubs), hub-to-regional (one endpoint is a hub), and regional-to-regional (neither endpoint is a hub).

Results

route_type	num_flights	avg_dep_delay	avg_arr_delay	avg_late_aircraft_delay	avg_carrier_delay	avg_weather_delay	avg_nas_delay	on_time_pct
Hub-to-Hub	2,656,928	14.04	7.85	6.30	5.22	0.69	3.48	78.32
Hub-to-Regional	3,943,748	11.93	6.78	5.89	5.23	1.04	2.47	80.61
Regional-to-Regional	364,591	9.14	5.05	5.54	3.61	0.73	1.76	82.76

Conclusion

Hub-to-hub routes experience the most significant delays, while regional-to-regional routes perform best. Hub-to-hub flights averaged 14.04 minutes departure delay and 7.85 minutes arrival delay, with only 78.32% arriving on-time. This higher delay rate is likely due to congestion at major airports and the complexity of coordinating connecting flights. Hub-to-regional routes showed moderate performance with 11.93 minute average departure delays and 80.61% on-time arrivals. Surprisingly, regional-to-regional routes had the best performance with only 9.14 minute average departure delays, 5.05 minute arrival delays, and the highest on-time percentage at 82.76%. Interestingly, carrier delays and late aircraft delays remain fairly consistent across all route types (around 5-6 minutes), while NAS delays are notably higher for hub-to-hub routes (3.48 min vs 1.76 min for regional), suggesting that air traffic congestion at major hubs is a key contributing factor to delays.

Research Question 3: How do flight delays vary on major holidays compared to regular days?

Analysis Approach

Seven major US holidays were analyzed: New Years, MLK Day, Memorial Day, July 4th, Labor Day, Thanksgiving, and Christmas. For each holiday, a 6-day window was created (3 days before through 2 days after the holiday). Flights were classified as either "Holiday Period" or "Regular Day" using CTEs and date range comparisons.

Results

Overall Comparison:

day_type	num_flights	avg_dep_delay	avg_arr_delay	avg_late_aircraft_delay	avg_carrier_delay	avg_weather_delay	avg_nas_delay	cancellation_rate	on_time_pct
Regular Day	6,255,562	12.29	6.78	5.91	5.07	0.81	2.78	0.00	80.08
Holiday Period	709,705	15.27	9.90	7.01	5.80	1.61	3.14	0.00	77.83

Individual Holiday Breakdown:

holiday_name	num_flights	avg_dep_delay	avg_arr_delay	avg_late_aircraft_delay	on_time_pct
MLK Day	92,936	33.09	31.62	13.92	60.67
Memorial Day	113,067	20.65	17.53	9.70	72.28
Christmas	111,059	17.26	9.72	7.81	77.84
July 4th	114,692	11.01	4.65	5.43	81.88
Labor Day	110,604	9.68	4.09	4.84	82.97
Thanksgiving	112,700	6.67	0.38	3.25	85.12

New Years	54,647	7.75	-0.02	3.57	84.59

Conclusion

Holiday periods experience noticeably worse delays compared to regular days, with average departure delays of 15.27 minutes versus 12.29 minutes, arrival delays of 9.90 minutes versus 6.78 minutes, and on-time performance dropping from 80.08% to 77.83%. Late aircraft delays increase by nearly 20% during holidays, weather delays more than double, and NAS delays increase by 13%, suggesting that higher passenger volumes create airport congestion while making it harder to recover from disruptions. However, the impact varies dramatically by holiday: MLK Day is by far the worst with severe delays averaging 33.09 minutes for departures and 31.62 minutes for arrivals (only 60.67% on-time), likely due to winter weather conditions in mid-January. Memorial Day and Christmas also show significant disruptions with 17-20 minute departure delays, while July 4th and Labor Day experience moderate impacts. Surprisingly, Thanksgiving and New Years perform exceptionally well—Thanksgiving actually outperforms regular days with only 6.67 minute departure delays and 85.12% on-time performance, and New Years shows minimal impact. This suggests that winter weather, rather than passenger volume alone, is the primary driver of holiday-related delays.

Research Question 4: Which day of the week experiences the largest flight delays?

Analysis Approach

Flights were grouped by day_of_week (1=Monday through 7=Sunday), and aggregate functions calculated average delays for each day. Window functions were used to compare each day's performance against the weekly average and rank days by delay severity.

Results

day_name	num_flights	avg_dep_delay	avg_arr_delay	avg_late_aircraft_delay	weekly_avg_dep_delay	diff_from_weekly_avg	delay_rank
Monday	1,056,398	12.64	7.03	6.04	12.54	0.10	4
Tuesday	961,064	10.36	4.39	5.15	12.54	-2.18	7
Wednesday	958,699	10.33	4.83	5.01	12.54	-2.21	6

Thursday	1,024,394	13.08	8.18	6.37	12.54	0.54	3
Friday	1,037,087	14.59	9.61	6.99	12.54	2.05	1
Saturday	898,255	12.51	6.45	5.77	12.54	-0.03	5
Sunday	1,029,370	14.28	8.77	6.68	12.54	1.74	2

Conclusion

Friday experiences the largest flight delays with average departure delays of 14.59 minutes and arrival delays of 9.61 minutes, ranking as the worst day of the week (2.05 minutes above the weekly average). Sunday follows closely as the second-worst day with 14.28 minute departure delays. Tuesday and Wednesday are the best-performing days with average departure delays around 10.3 minutes—more than 2 minutes below the weekly average—and the lowest arrival delays at 4.39 and 4.83 minutes respectively. The pattern reveals that weekends and Fridays are significantly worse than mid-week travel, with late aircraft delays peaking on Friday (6.99 min) and Sunday (6.68 min), suggesting that the combination of higher leisure travel volume and accumulated delays from the week compound to create the worst performance. The Tuesday-Wednesday window offers travelers the most reliable on-time performance.