IndiGo Airlines: Maximising On-Time Arrival Rate (OTAR) through Operational Efficiency

IndiGo Airlines, a leading Indian low-cost carrier, has recently entered the U.S. domestic aviation market with a vision to replicate its high-efficiency, customer-first operations in a new and complex environment. Operating across a network of busy U.S. hubs, weather-sensitive regions, and constrained air traffic zones, IndiGo now faces operational challenges unfamiliar to its existing markets—including cascading flight delays, unpredictable seasonal disruptions, and systemic inefficiencies.

Problem Objective

The goal of the Project is to provide suggestions to IndiGo Airline that help the airline maximize OTAR across its new U.S. network.The analysis will focus on:

- Decomposing OTAR into departure delay, taxi-out, en-route, taxi-in and turnaround components.
- Explore temporal (time-of-day, day-of-week, season) and network (hubs, rotations, route structures) patterns.
- Formulating and validating hypotheses around key delay drivers.
- Uncovering unexpected insights that could yield a competitive advantage.

Dataset

This dataset contains information about IndiGo Airlines, offering detailed insights into flight schedules, delays, departures, arrivals, and operational metrics such as air time, distance, and delay causes. Using this data, our goal is to provide actionable suggestions to help IndiGo maximize On-Time Arrival Rate (OTAR) across its new U.S. network.

The dataset includes the following columns:
year, month, day, day_of_week, airline, flight_number, tail_number, origin_airport, destination_airport, scheduled_departure, departure_time, departure_delay, taxi_out, wheels_off, scheduled_time, elapsed_time, air_time, distance, wheels_on, taxi_in, scheduled_arrival, arrival_time, arrival_delay, diverted, cancelled, cancellation_reason, air_system_delay, security_delay, airline_delay, late_aircraft_delay, and weather_delay.

Data Cleaning and Preparation

Handling Missing Values

- Categorical Imputation: Replaced missing values in CANCELLATION_REASON with category 'E' (Unknown).
- Delay Columns: Filled NaNs in delay-related columns (LATE_AIRCRAFT_DELAY, WEATHER_DELAY, AIRLINE_DELAY, AIR_SYSTEM_DELAY, SECURITY_DELAY) with 0, assuming no delay.
- Preserving NaNs: Retained missing values in key timing fields (e.g., ARRIVAL_DELAY, WHEELS_OFF) to accurately reflect cancellations and diversions.
- Dropping Minimal NaNs: Removed rows with missing values in TAIL_NUMBER and SCHEDULED_TIME due to negligible impact.

Mapping and New Column Data Formulation

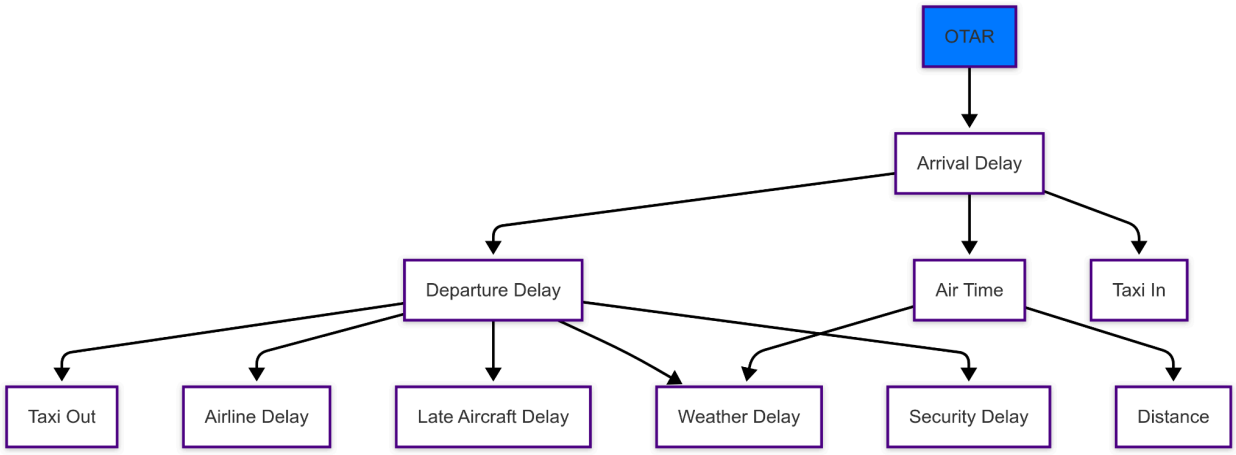
- AIRLINE Column: Mapped airline codes (e.g., AA) to full airline names (e.g., American Airlines).
- Airport Columns: Replaced ORIGIN_AIRPORT and DESTINATION_AIRPORT codes with full airport names using IATA mappings.
- Geolocation Data: Retrieved and added latitude and longitude for both origin and destination airports.
- Cancellation Reason: Mapped codes in CANCELLATION_REASON (e.g., A, B) to descriptive categories (e.g., Carrier, Weather).

Metrics

Key Metrics

- ARRIVAL_DELAY
- DEPARTURE_DELAY
- TAXI_OUT
- TAXI_IN
- AIR_TIME
- ELAPSED_TIME
- DISTANCE
- AIRLINE_DELAY
- LATE AIRCRAFT DELAY
- WEATHER_DELAY
- SECURITY_DELAY

Metric Tree

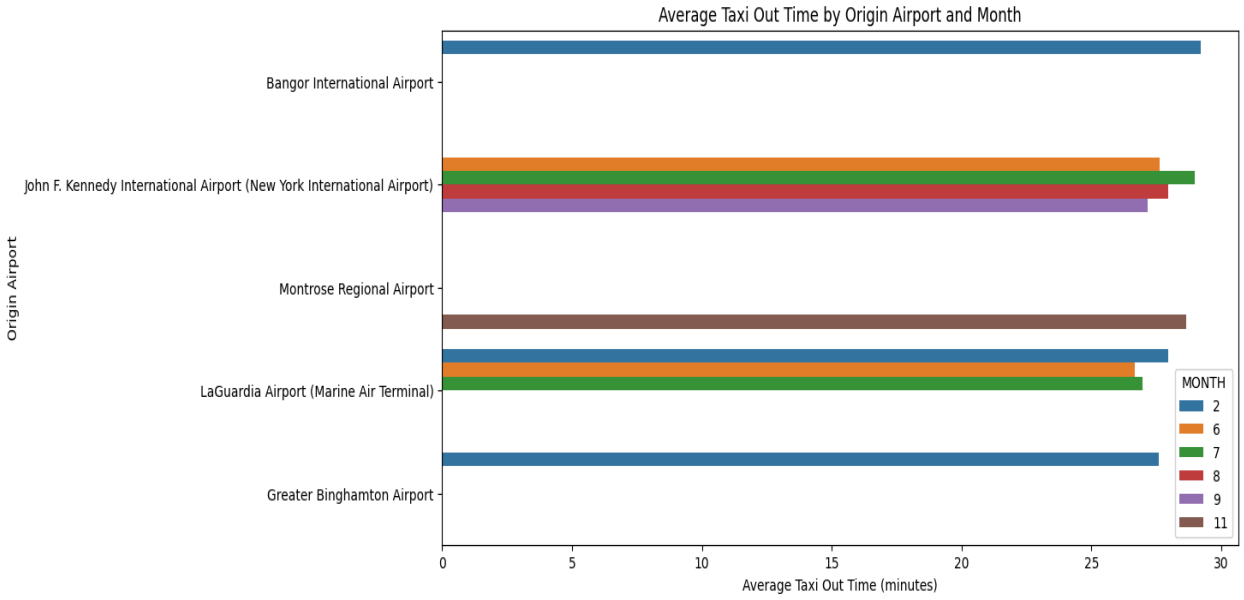


Hypothesis

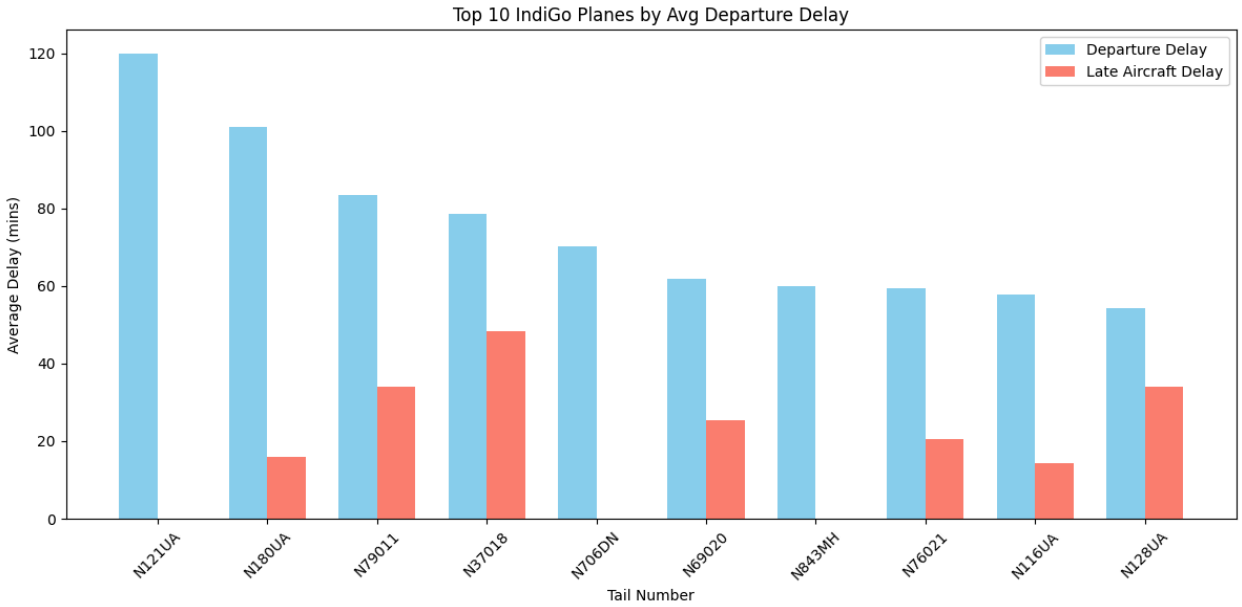
OATR Components

- **Depature Segment**

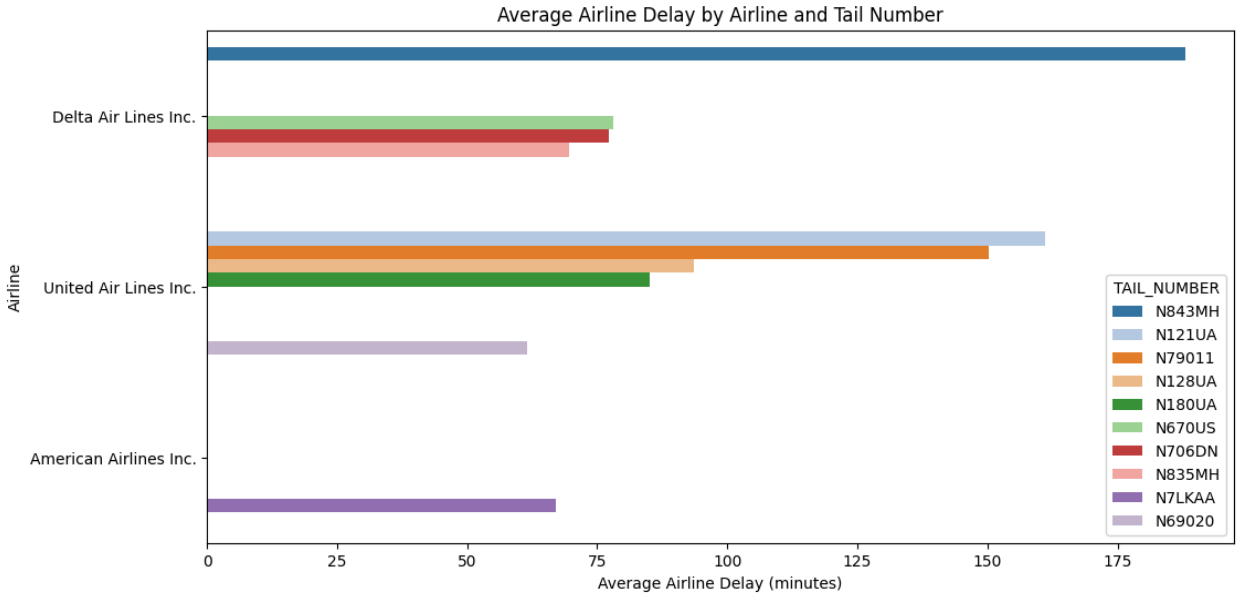
Hypothesis1: Bangor International Airport in February (29.21 min), John F. Kennedy International Airport (JFK) in June–September (27.17–28.99 min, peak 28.99 min in July), and LaGuardia Airport (LGA) in June–July (26.66–27.96 min) have the highest average taxi-out times compared to other airports



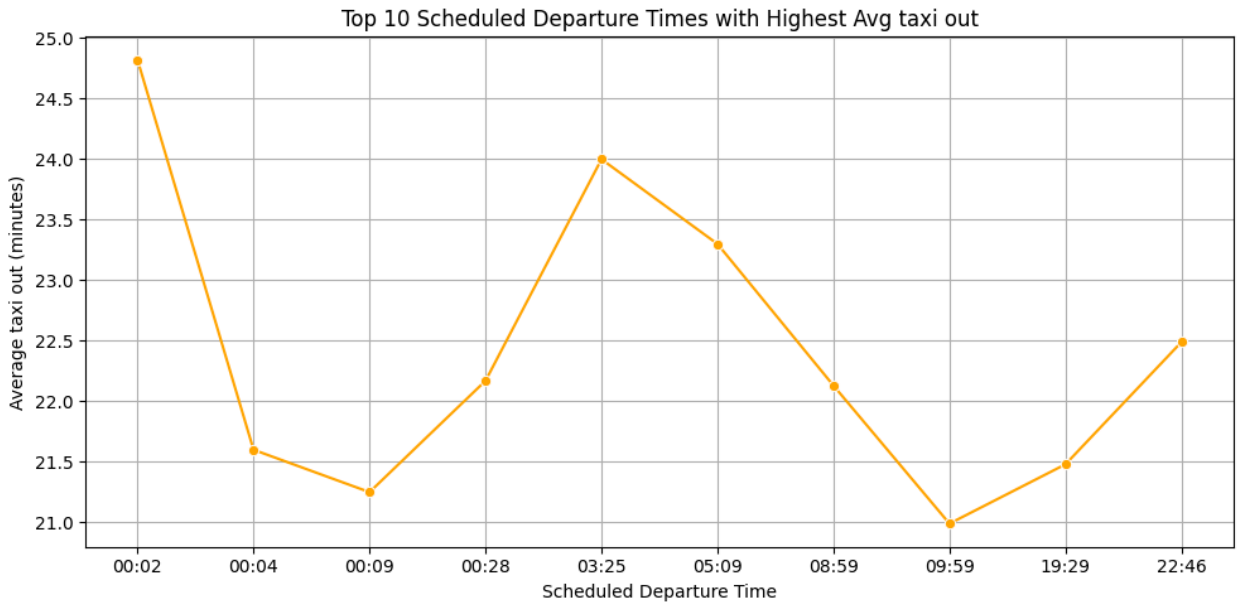
Hypothesis2: The aircraft with tail number N121UA, N180UA has the highest average departure delay among, indicating it may have more frequent or severe operational issues compared to other aircraft.



Hypothesis3: The aircraft with tail number N843MH from Delta Air Lines Inc. has the highest average airline delay of approximately 175 minutes, indicating a significantly higher delay compared to other aircraft across all airlines.

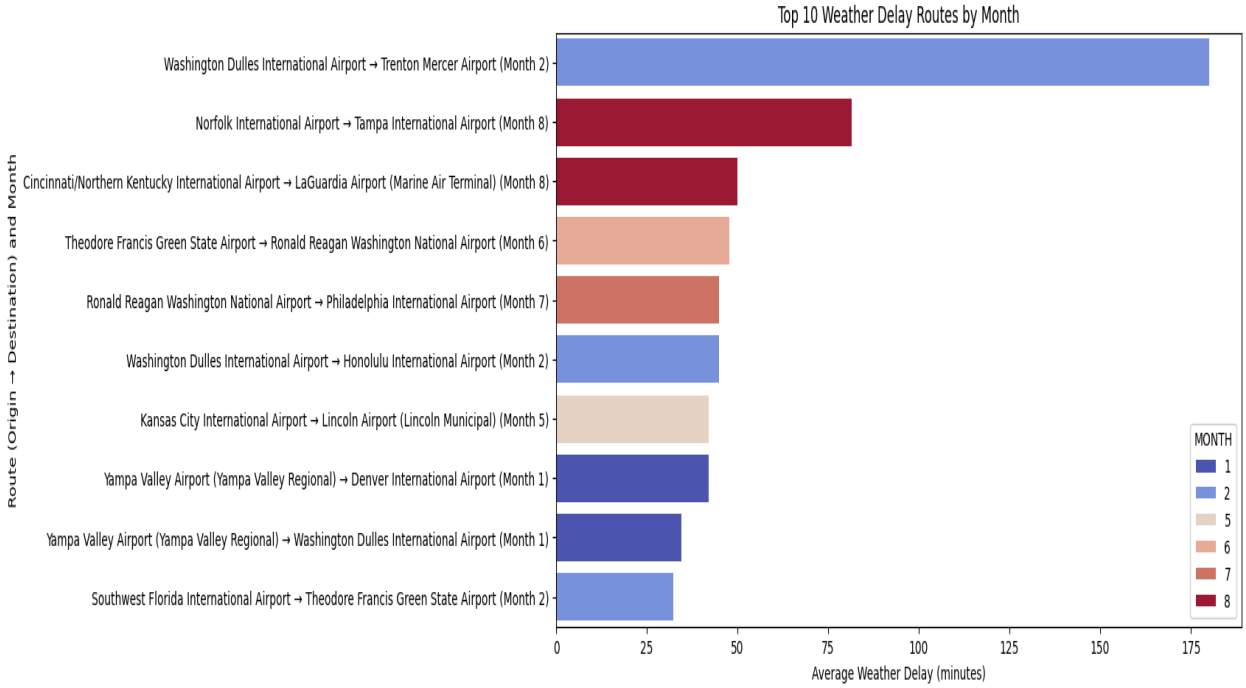


Hypothesis4: Flights scheduled around midnight to early morning (00:00 AM to 05:30 AM) and late evening (around 10:40 PM) have the highest average taxi-out times among all scheduled departure times.

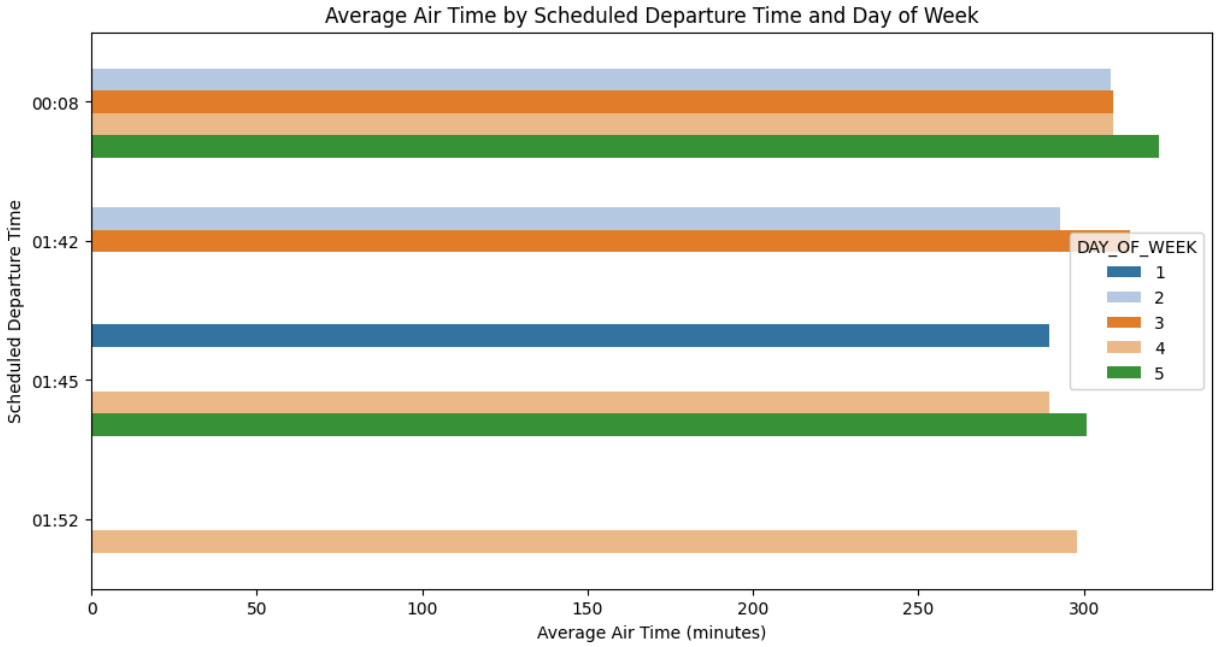


- **In flight Segement**

Hypothesis5: In the month of February (Month 2), the route from Washington Dulles International Airport to Trenton Mercer Airport recorded the highest average weather delay, exceeding 175 minutes, among all observed routes.

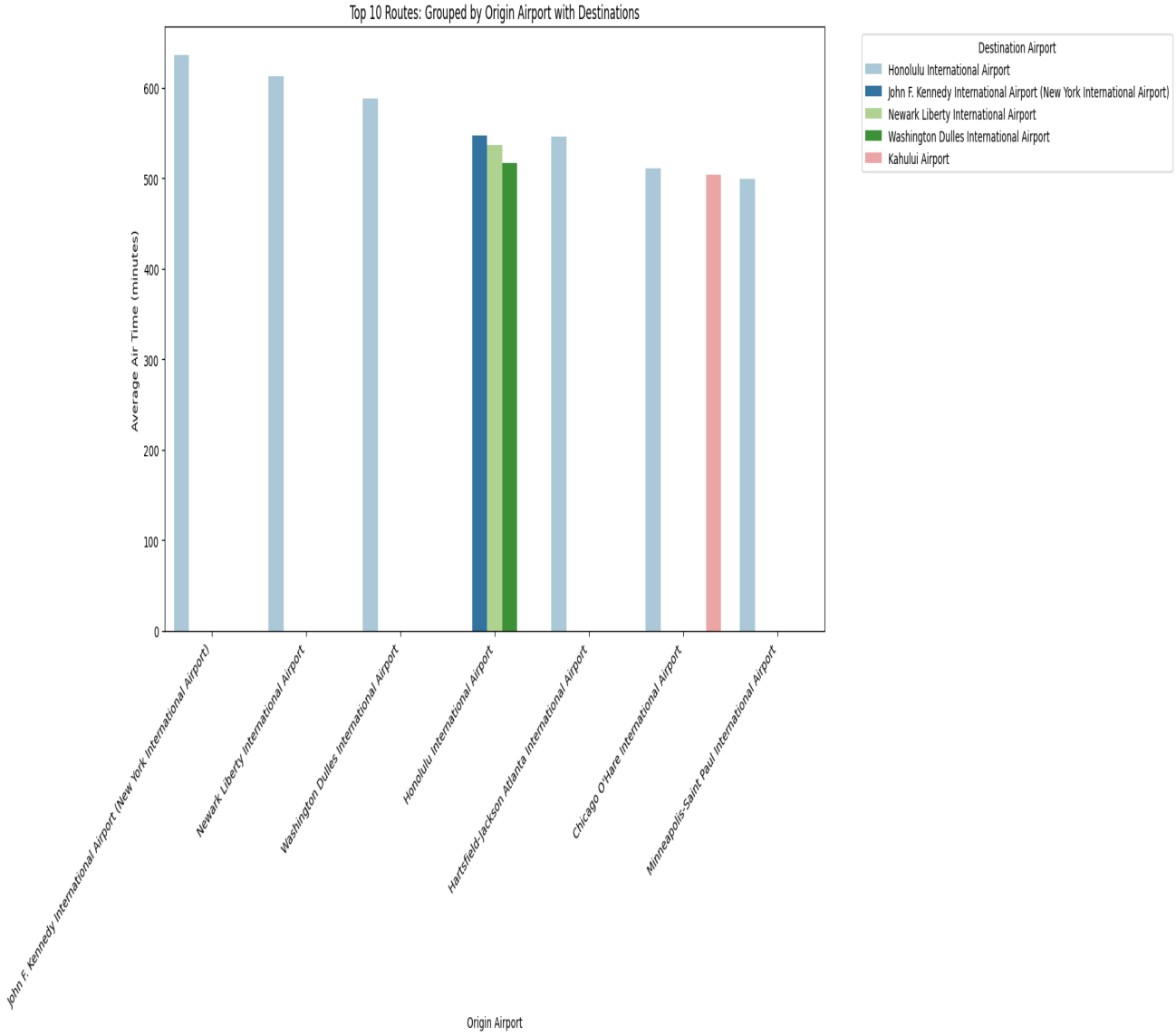


Hypothesis6:Flights scheduled between 00:08 and 01:52 (midnight to 2 AM) show higher average air times, often ranging between 4.9 to 5.3 hours. This trend is especially notable on Wednesdays (Day 3) and Fridays (Day 5).



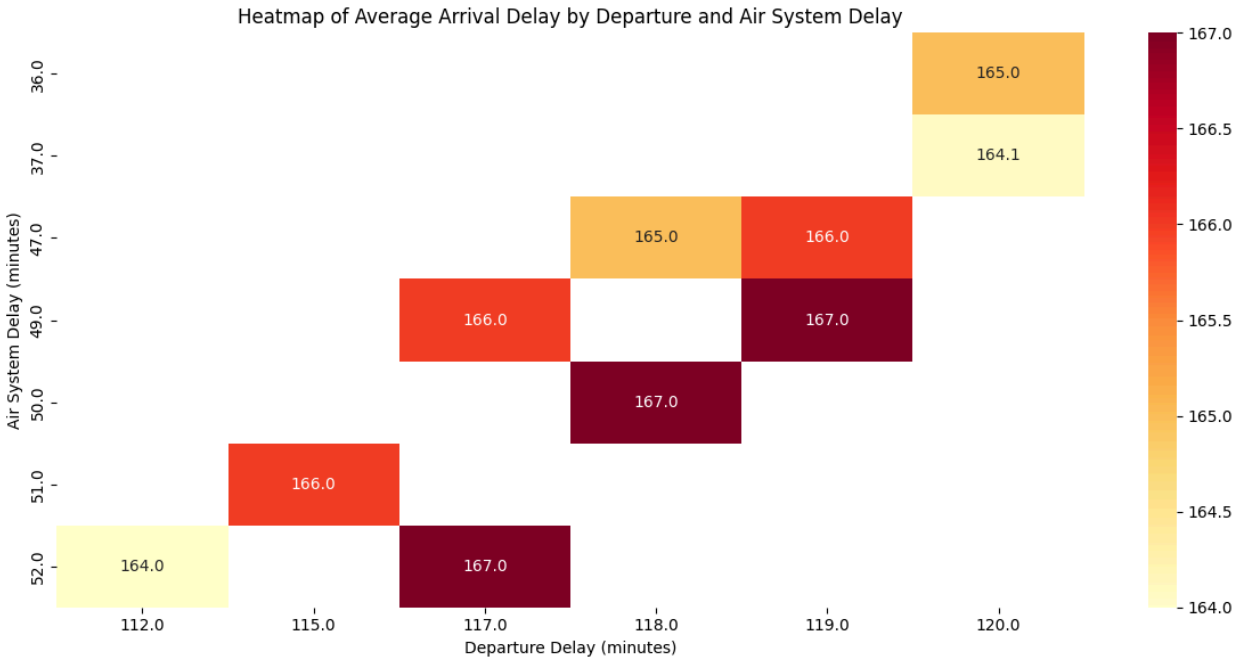
Hypothesis7 : Flights originating from John F. Kennedy International Airport (JFK) ,Newark Liberty International Airport to destination Honolulu International Airport (HNL)

exhibit consistently high average air time greater than 10 hours

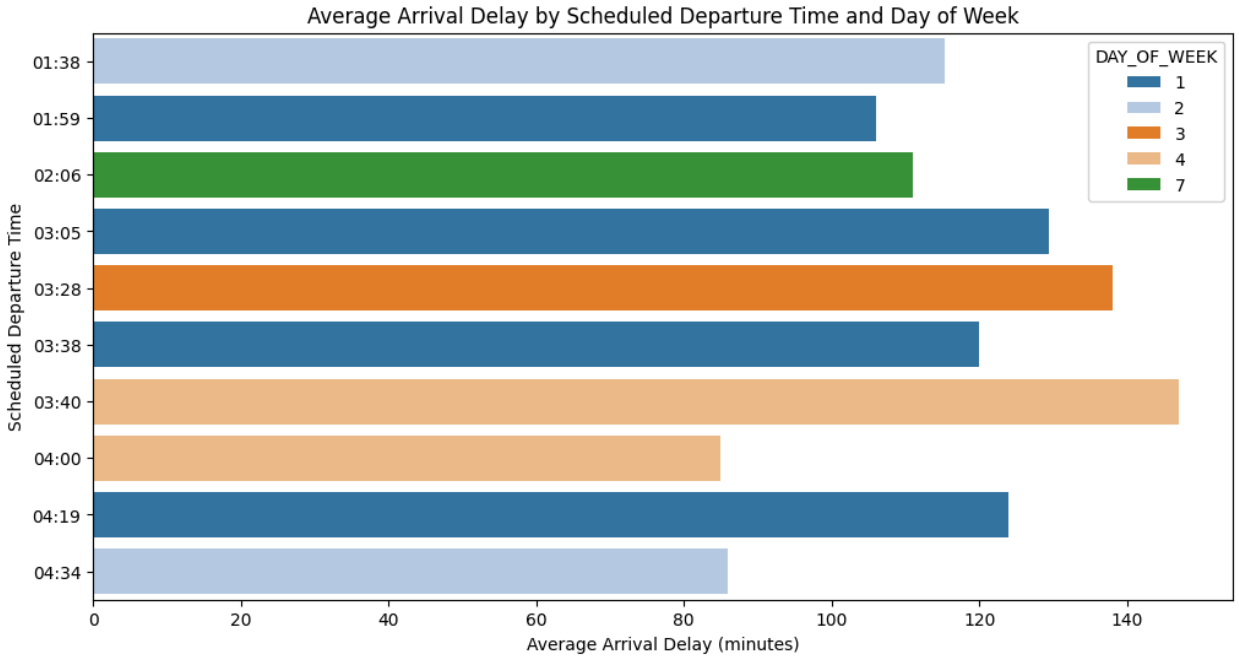


- **Arrival Segment**

Hypothesis8: Departure delay and air system delay significantly affect arrival delay, particularly in cases where both delays are high.

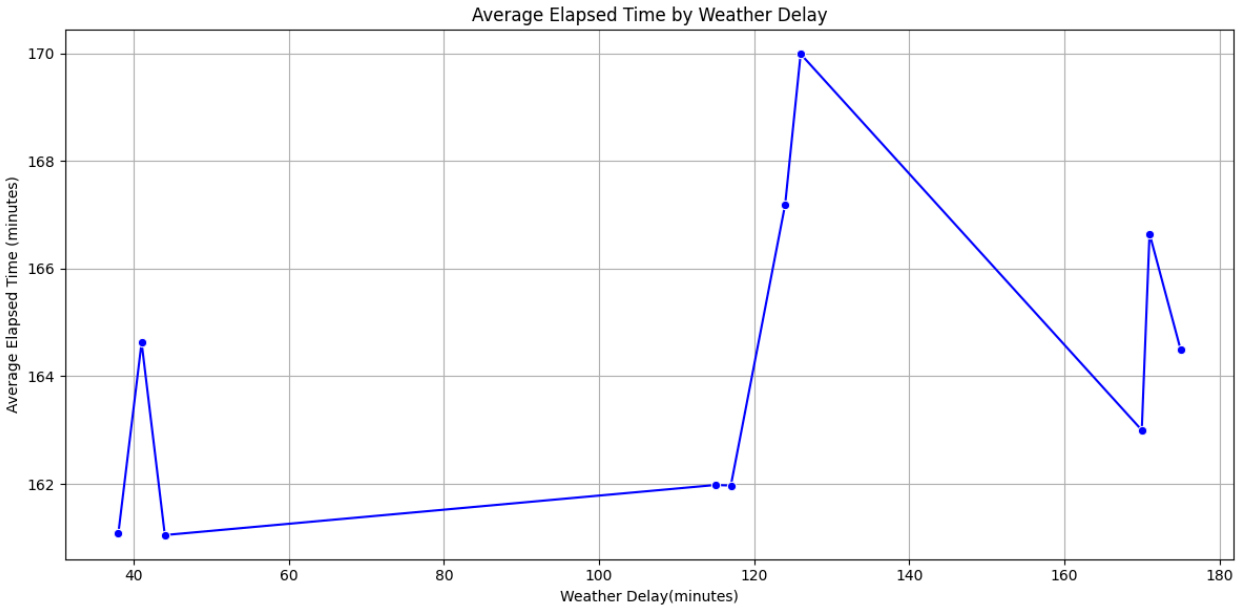


Hypothesis9: Flights departing between 03:28 and 03:40 tend to have the highest average arrival delays, especially on Day 4 (Thursday)

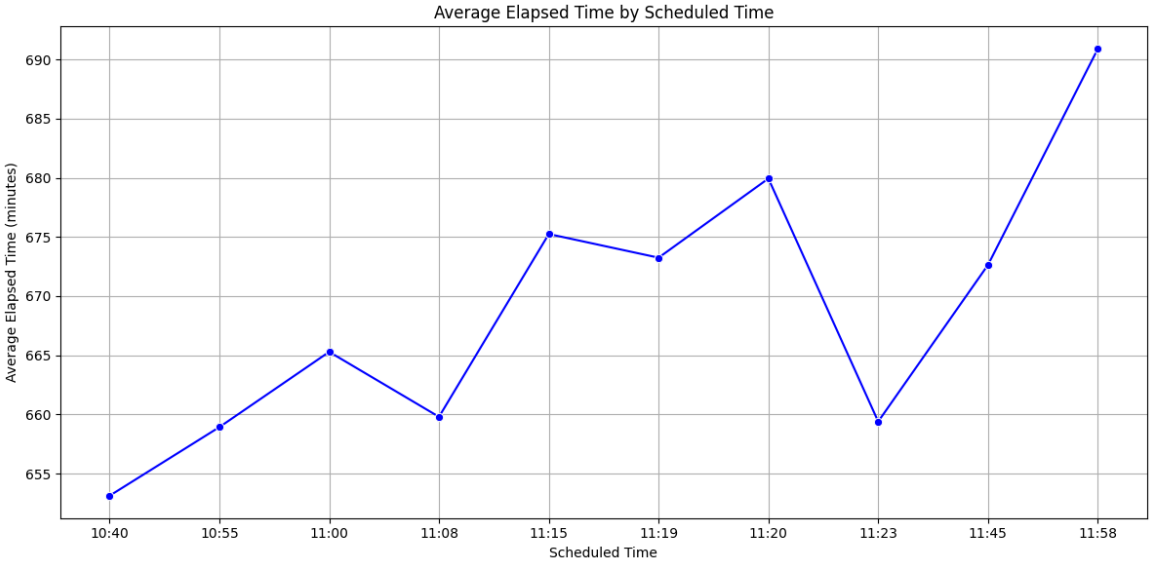


- **Total Flight Duration**

Hypothesis10: Flights with higher weather delays tend to have increased average elapsed times, especially when the delay exceeds a certain threshold (e.g., 120 minutes).



Hypothesis11:Flights scheduled around late morning (10:40 AM – 11:58 AM) tend to have slightly longer average elapsed times compared to flights scheduled earlier in the day.



Key Findings from Exploratory Data Analysis

- **Depature Segement**

  Airport-Specific Delays

  - Wilmington, Martha's Vineyard, and Barnstable show midweek delay peaks.
  - Newark to Minneapolis and Pago Pago on Wednesdays face extreme airline delays.
  - JFK, LGA, and Bangor have high taxi-out times, especially in summer and February.
  - St. Cloud Regional and Wilmington top total delays; Gustavus and Southwest Oregon see high arrival delays.

  Time and Day Patterns

  - Late-night (00:00–05:30 AM) and afternoon (2 PM) flights show highest delays.
  - Mondays, Saturdays, and Sundays worsen early-morning taxi-out times.
  - Thursdays and Fridays peak taxi-out delays at JFK/LGA.

  Aircraft and Airline Insights

  - Tail numbers N121UA, N843MH, and N79011 consistently report high delays.
  - Spirit Airlines leads in departure delays; Hawaiian Airlines the lowest.
  - Flight 6839 and Atlantic Southeast/Frontier show high airline delays.

  Seasonal and Security Factors

  - Summer and February spike taxi-out times; March, July, December raise airline delays.
  - Security delays are highest at Adak Airport, moderate at Gustavus.

- **In flight Segement**

  Weather Delays Drive OTAR Drops

  - Specific routes (e.g., Washington Dulles–Trenton Mercer) and times (late-night, 04:50 AM) face heavy weather delays, impacting departure and en-route times.

  Timing Matters

  - Late-night, early-morning, and February flights show higher delay risks.
  - Air times peak around midnight–2 AM on Wednesdays/Fridays.

  Route Challenges

  - Long air times from airports like Guam, Pago Pago, and JFK–Honolulu highlight network stress points.
- **Arrival Segment**

  Route-Specific Delays Hurt OTAR

  - Routes like Raleigh-Durham–Indianapolis face extreme arrival delays (>160 min), needing targeted fixes.

  Timing Inefficiencies

  - Departures between 03:28–03:40 AM (especially Thursdays) and early 4:22 AM arrivals show high delays, pointing to scheduling and staffing gaps.

  Systemic Delays Cascade

  - Early issues (departure, air traffic delays) trigger network-wide arrival delays.

  Hub Bottlenecks

  - High taxi-in times at hubs like Chicago O'Hare disrupt turnaround and OTAR.

- **Total Flight Duration**

  Weather Delays Disrupt OTAR

  - Flights with >120 min weather delays (e.g., Philadelphia avg 180 min) drive late arrivals (H41, H44).

Route-Specific Patterns

- Long flights (e.g., Honolulu 640 min) and short high-delay routes (e.g., Philadelphia) need tailored strategies (H43, H44).

Timing Impacts Efficiency

- Late morning flights (10:40–11:58 AM) show longer elapsed times, increasing delay risk (H40).

Impacts on On-Time Arrival Rate (OTAR)

- **Departure Segment**
  - Departure Delays: Major cause for arrival disruptions; midweek and overnight schedules need adjustment.
  - Taxi-Out Congestion: High taxi-out times at JFK, LGA, Bangor erode on-time performance; improving ground ops is key.
  - Aircraft Issues: Unreliable aircraft (e.g., N843MH) drag down OTAR predictability.
  - Vulnerable Routes/Times: Newark–Minneapolis route and early-morning flights severely impact OTAR.
  - Customer Trust: High total delays at St. Cloud, Wilmington, and seasonal peaks hurt passenger satisfaction and OTAR scores.

- **In flight segment**
  - Reduce Weather Delays: Focus on high-delay routes/times to cut >175 min delays (e.g., Dulles–Trenton Mercer in February).
  - Schedule Smarter: Avoid peak delay windows (late-night winter, early AM) and long air-time slots.
  - Gain Edge: Proactive delay management can boost IndiGo's reliability in the U.S. market.
- **Arrival Segment**
  - Fix Extreme Routes : Target Raleigh-Durham–Indianapolis to cut >160 min delays.
  - Adjust Schedules : Avoid peak delay windows (03:28–03:40 AM Thursdays).

    - Tackle Early Delays : Improve departure and air system coordination.

    - Boost Hub Efficiency : Streamline taxi-in at O'Hare for better network flow.
- **Total Flight Duartion**
    - Weather Delay Mitigation: Improve forecasting, contingency planning, and rerouting to minimize major weather-related delays.

    - Route Optimization: Assign faster aircraft and add schedule buffers for long or delay-prone routes to enhance punctuality.

    - Schedule Adjustment: Shift departures away from late morning peaks to reduce congestion and improve on-time arrivals.

    - Resource Planning: Increase staffing and ground operations support during high-delay periods to speed up turnaround times.

    - Passenger Communication: Proactively inform customers about potential delays to manage expectations and maintain satisfaction.

    - Ground Efficiency at Hubs: Streamline taxi-in and turnaround processes at major hubs to avoid cascading delays across the network.