My Hackathons Recommended

SkyHack 3.0: United Airlines

Event Dates :

3 October 2025 - 5 October 2025

My Progress | Team Details | Event Details

Registration

Round Ended | Round 01

Approved

SkyHack Challenge

Ends in 1 Day | Round 02

Pending

Problem Statement

Frontline teams at United Airlines are responsible for ensuring every flight departs on time and is operationally ready. However, not all flights are equally easy to manage. Certain flights pose greater complexity due to factors such as limited ground time, higher volumes of checked or carry-on baggage, and specific customer service needs that often increase with passenger load.

Currently, identifying these high-difficulty flights relies heavily on personal experience and local team knowledge. This manual approach is inconsistent, non-scalable, and risks missing opportunities for proactive resource planning across the airport.

To address this, you are tasked with developing a Flight Difficulty Score that systematically quantifies the relative complexity of each flight using the datasets provided, which span two weeks of departures from Chicago O'Hare International Airport (ORD).

Objective

The goal is to design a data-driven framework that:

- Calculates a Flight Difficulty Score for each flight using flight-level, customer, and station-level data.
- Identifies the primary operational drivers contributing to flight difficulty to enable proactive planning and optimized resource allocation.

Deliverables

1. Exploratory Data Analysis (EDA)

- 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?

- Are high special service requests flights also high-delay after controlling for load?

2. Flight Difficulty Score Development

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.

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?

Feature Exploration Guidance

You are encouraged to explore and engineer features from across the datasets. Some examples of feature categories include:

- Ground time constraints: How scheduled ground time compares with expected turnaround standards
- Flight-specific characteristics: Number of checked bags, children on board, haul, aircraft etc.
- Passenger service needs: Requests such as wheelchair assistance or special seating

These are not exhaustive—you should explore the datasets further to identify additional operational features that could influence flight complexity.
Creative ideas are welcome! — feel free to explore additional approaches or external datasets
Things to Note:

- Demonstrate your ability to write SQL/Python/R for the above stated tasks. Submit your code files along with instructions to run them.
- Summarize your approach, calculations and findings in a concise, data-driven presentation (7-8 slides)
- As this case study is intended to assess your critical thinking, individual analytical, and coding capabilities, you must not use generative AI tools such as ChatGPT, Claude, or Grok for assistance

Submissions Guidelines:

Submission Requirements

To submit your project, please provide the following in a single ZIP file:

- Report: A PDF or PPT file that includes your findings, calculations, and insights. You can use visualizations such as charts and diagrams to support your report.
- Difficulty Score File: A CSV file named test_<your_name>.csv (e.g., test_johndoe.csv). This file must contain the following: flight details, features used for calculation, and final difficulty score (or class/rank order).
- Code: The Python, SQL or R scripts you used. Please include instructions on how to run them. If you used a code repository, you can share the link instead of the files.
- Additional Materials: Any other materials you used to create your report.

Upload Instructions

- Combine all your files (the report, CSV, code, and any other materials) into a single ZIP folder.
- The total size of the ZIP file must not exceed 50 MB.
- You can save a draft of your submission before the final deadline.
- Drafts can be updated by uploading a new ZIP file.
- Once you submit your project, you will not be able to make any changes.
- Any drafts you have saved will be automatically submitted after the deadline.

Important Notes - Please note that there is no limit on the number of slides/pages. The structure of your report should include -table of contents, executive summary, headings, references, footnotes, and appendices.

Data Dictionary:

Flight Level Information

Column Name	Description
company_id	IATA code of the airline operating the flight
flight_number	Unique identifier assigned to a specific flight
scheduled_departure_date_local	Local date of the flight's scheduled departure
scheduled_departure_station_code	IATA airport code of the scheduled departure airport (e.g., "JFK", "LAX")

	departure airport (e.g., "JFK", "LAX")
scheduled_arrival_station_code	IATA airport code of the scheduled arrival airport (e.g., "JFK", "LAX")
scheduled_departure_datetime_local	Scheduled local date and time of departure from the origin airport
scheduled_arrival_datetime_local	Scheduled local date and time of arrival at the destination airport
actual_departure_datetime_local	Actual local date and time when the flight departed from the origin airport
actual_arrival_datetime_local	Actual local date and time when the flight arrived at the destination airport
total_seats	Total number of passenger seats available on the aircraft for the flight
fleet_type	Type model of aircraft used for the flight
carrier	Distinction between Mainline and Express
scheduled_ground_time_minutes	Planned duration (in minutes) the aircraft is scheduled to spend on the ground between flights
actual_ground_time_minutes	Actual time available (in minutes) for ground operations between flights
minimum_turn_minutes	Minimum required turnaround time (in minutes) for the aircraft between flights

PNR Flight Level Information

Column Name	Description
company_id	IATA code of the airline operating the flight
flight_number	Unique identifier of the flight associated with the PNR
scheduled_departure_date_local	Local date of the flight’s scheduled departure
scheduled_departure_station_code	IATA airport code of the scheduled departure airport (e.g., "JFK", "LAX")
scheduled_arrival_station_code	IATA airport code of the scheduled arrival airport (e.g., "JFK", "LAX")
record_locator	Unique identifier for the PNR, used to reference a passenger booking
pnr_creation_date	Date when the PNR was created
total_pax	Total number of passengers associated with the PNR on this flight
lap_child_count	Number of lap children (infants not occupying a seat) included in the PNR
is_child	Indicates whether the passenger is classified as a child
basic_economy_pax	Number of passengers in the PNR booked in basic economy fare class
is_stroller_user	Indicates whether the passenger is a stroller user

PNR Remarks Information

Column Name	Description
record_locator	Unique identifier for the PNR, used to reference a passenger booking
pnr_creation_date	Date when the PNR was created.
flight_number	Unique identifier of the flight associated with the PNR
special_service_request	Description of the requested special service (e.g., wheelchair assistance)

Airports Information

Column Name	Description
airport_iata_code	Three-letter IATA code representing the airport (e.g., "LAX", "JFK").
iso_country_code	Two-letter country code where the airport

	is located (e.g., "US", "CA").

Bag Level Information

Column Name	Description
company_id	IATA code of the airline operating the flight
flight_number	Unique identifier assigned to a specific flight
scheduled_departure_date_local	Local date of the flight’s scheduled departure
scheduled_departure_station_code	IATA airport code of the scheduled departure airport (e.g., "JFK", "LAX")
scheduled_arrival_station_code	IATA airport code of the scheduled arrival airport (e.g., "JFK", "LAX")
bag_tag_unique_number	Unique identifier for the bag tag
bag_tag_issue_date	Date the bag tag was issued
bag_type	Type of bag (e.g., "Checked", “Transfer”) **Hot transfer bags are transfer bags with a connection time of less than 30 minutes*

Data Dictionary:

Airports Data.csv
Bag Level Data.csv
Flight Level Data.csv
PNR Flight Level Data.csv
PNR Remark Level Data.csv

Join Our Telegram Group in case of any queries.
Link to Telegram

Submit Your Zip File