BI PROJECT REPORT

Drive Link:

https://drive.google.com/drive/folders/1k408tT9PDguTSRShcdaqvxS0BI05aqIC?usp=sharing

Group Members:

Zehra Ahmed 26965

Farah Inayat 26912

Zuha Aqib 26106

Kisa Fatima 27076

Description of Dataset: 2
- Key Features: 2
- Industry Background: 3

Summary of Data Wrangling/EDA steps and consolidated results 4
- Data Wrangling 4
- EDA & Statistical Testing 7

Summary of Design Thinking Framework 8
- 1. Empathize: 8
- 2. Define: 10
- 3. Ideate: 11
- 4. Prototypes: 16
- 5. Testing: 16

BI REPORT: 18
- Title Page: 18
- Problem Statement: 18
- Dashboard 1: Overall Flight Disruption Status: 19
- Dashboard 2: Root Cause Analysis of Cancellations: 20
- Dashboard 3: Root Cause Analysis of Delays : 21
- Dashboard 4: Delay Drill-Throughs 22
- Dashboard 5: Airport Analysis 23
- Dashboard 6: Airline $ Route Analysis 24
- BI Charts: 24

Data Story: 36

Conclusion: Targeted Action for Maximum Impact 42

Work Contribution: 42

Description of Dataset:

This dataset contains detailed records of airline flight operations within the United States during the early months of the COVID-19 pandemic (Jan to June 2020). The data captures a wide range of flight-level attributes, including scheduling, delays, cancellations, and causative factors. It is designed to help understand disruptions in air travel during a time of unprecedented global impact.

Key Features:

- Flight Details: Includes scheduled and actual departure/arrival times, airline codes, tail numbers, and route identifiers (origin/destination).
- Delay Metrics: Offers granular insights into delay durations and causes, such as:
    - Carrier delay
    - Weather delay
    - National Aviation System (NAS) delay
    - Security and late aircraft delays
- Cancellations: Binary indicators for cancellations, along with categorized reasons (e.g., Weather, Carrier).
- Temporal Data: Includes full date, day of the week, time blocks, and time-of-day groupings.
- Distance and Time: Covers scheduled and actual flight duration, airtime, taxi in/out times, and great-circle distances.
- The dataset uses local time for all time-based fields and supports block-level analysis with pre-binned time intervals and distance groups for easier segmentation.

Industry Background:

The global aviation industry was among the hardest-hit sectors during the COVID-19 pandemic. Between March and June 2020, airlines saw drastic reductions in demand, leading to:

- Widespread cancellations of domestic and international flights,
- Major schedule disruptions and delays due to reduced staffing, airport protocol changes, and aircraft unavailability,
- Operational bottlenecks caused by varying lockdown policies across states and regions.

Summary of Data Wrangling/EDA steps and consolidated results

Data Wrangling

The dataset initially contained 47 columns with raw, inconsistent, and missing data requiring standardization and transformation to enable meaningful business intelligence insights.

Key Observations:

- Fields like YEAR and QUARTER had constant or derived values
- Time columns such as DEP_TIME and ARR_TIME were in HHMM format and had missing values for cancelled flights
- Cancellation and delay reasons were poorly encoded or incomplete
- Many fields needed renaming for understanding meaning

Steps Followed:

- Columns with a single value (e.g., YEAR) were dropped
- Time fields were standardized (handling 2400 edge case and HHMM conversion)
- Categorical fields were re-encoded/mapped into readable labels
- Missing values were logically filled or replaced with placeholders
- Testing included value validation, consistency checks, and null rechecks

Results:

<u>Columns Dropped:</u>

- YEAR (It is 6 month data from Jan to June so single value in entire col)
- QUARTER (could be picked from FL_DATE)
- DEP_DELAY_NEW, ARR_DELAY_NEW: Duplicate of DEP_DELAY, ARR_DELAY logic
- DEP_TIME_BLK, ARR_TIME_BLK: Redundant once hour fields were derived

Time Fields Centralized:

- FL_DATE → converted to datetime and renamed FLIGHT_DATE
- CRS_DEP_TIME and CRS_ARR_TIME → created SCHEDULED_DEP_HOUR / SCHEDULED_ARR_HOUR
- DEP_TIME → int, filled -1 for cancelled flights, fixed 2400, renamed ACTUAL_DEP_TIME
- WHEELS_OFF / WHEELS_ON → renamed DEP_TAKEOFF_TIME, ARR_LANDING_TIME, converted to int
- ARR_TIME → fixed 2400, added ACTUAL_ARR_HOUR

Categorical & Identification Fields

- DAY_OF_WEEK → mapped to weekday names (Mon, Tue, etc.)
- MKT_UNIQUE_CARRIER → renamed to MARKETING_CARRIER_CODE
- TAIL_NUM → filled missing values as: CANCELLED == 1 → keep null, else 'UNKNOWN'
- Renamed airport/state codes (e.g., ORIGIN → ORIGIN_AIRPORT_CODE, ORIGIN_STATE_ABR → ORIGIN_STATE_CODE)

Delay-Related Fields

- DEP_DELAY, ARR_DELAY:
    1. Cancelled → 9999 (placeholder)
    2. Others: Calculated where possible from actual - scheduled
    3. Converted to int
- DEP_DEL15, ARR_DEL15:
    1. Cancelled → -1, mapped as 'Cancelled'
    2. Mapped buckets: 0 → 'Less than 15', 1 → 'Greater than 15'
    3. Renamed to DEP_DELAY_15_MIN, ARR_DELAY_15_MIN
- DEP_DELAY_GROUP, ARR_DELAY_GROUP:
    1. Cancelled → 9999, mapped to 'Cancelled'
    2. Binned by 15-minute increments, all mapped to strings

Taxi, Elapsed Time & In-Air Time

- TAXI_OUT, TAXI_IN:
    1. Cancelled → -1
    2. Converted to int, renamed to ORIGIN_TAXI_TIME, DEST_TAXI_TIME

- ACTUAL_ELAPSED_TIME:
    1. Cancelled → 0
    2. Else: Calculated from actual arrival - departure
- AIR_TIME:
    1. Cancelled → 0
    2. Else: WHEELS_ON - WHEELS_OFF
    3. Renamed to IN_AIR_DURATION

Cancellation Handling

- CANCELLED: Mapped to string 'Yes'/'No' → new column cancelled_c
- CANCELLATION_CODE:
    1. Mapped: A → Carrier, B → Weather, C → NAS, D → Security
    2. Missing filled with 'Not Cancelled'
    3. Renamed to CANCELLATION_REASON

Distance Fields

- DISTANCE: No change
- DISTANCE_GROUP: Mapped integers to ranges like '500–750 miles', converted to string

Delay Cause Columns

- CARRIER_DELAY, WEATHER_DELAY, NAS_DELAY, SECURITY_DELAY, LATE_AIRCRAFT_DELAY:
    1. Cancelled → -1
    2. Nulls in operational flights → 0
    3. Converted to int

Testing columns again:

Apart from the usual routine checks such as looking at data types and checking if missing values exists again, these are the more data specific validations that we did:

- Logic Tests:
    1. If CANCELLED == 1, then DEP_TIME == -1
    2. If WHEELS_OFF exists but WHEELS_ON == -1, → flagged as diverted
    3. Verified ACTUAL_ELAPSED_TIME equals ACTUAL_ARR_TIME - ACTUAL_DEP_TIME
    4. Verified AIR_TIME equals WHEELS_ON - WHEELS_OFF

EDA & Statistical Testing

1. Normality Testing Using Shapiro-Wilk Test: All tested columns, including time-based fields (e.g., scheduled and actual departure/arrival times), delay durations, and distance measures, showed significant deviation from a normal distribution. This lack of normality is expected in operational datasets where rare events like long delays or cancellations create skewed distributions.
2. Outlier Analysis Using Z-score method ±3: The outliers represent real operational anomalies, such as unusually long taxi times or severe weather-related delays. Rather than removing them, these cases require further investigation and so were left in the dataset, as they could point to systemic weaknesses, overburdened routes, or underperforming aircraft. Outliers in distance suggest a mix of short-haul and long-haul routes, which should be analyzed separately when studying delay behavior.
3. Correlation Analysis using Pearson Correlation Coefficient: The following insights were captured:
    a. Departure delays almost directly lead to arrival delays (DEP_DELAY & ARR_DELAY). This reinforces that early-stage issues ripple through the flight cycle.
    b. Delays highly predict cancellations (DEP_DELAY & CANCELLED).Flight cancellations are often preceded by extreme delays, indicating a threshold effect.
4. Chi Square Test: The following insights were captured:
    a. Given that carrier code and performance indicators like delay groups and cancellations had a high association, it can be seen that certain aircraft or airlines are more prone to delays or cancellations.
    b. Given that delay groups and categorical variables like ORIGIN had a high association, it can be said that origin and destination geography significantly impact delay likelihood, perhaps due to weather, airport congestion, or route complexity.

Summary of Design Thinking Framework

The idea was to create a human-centered, insight-driven solution tailored to the challenges of airline operations during the COVID-19 period.

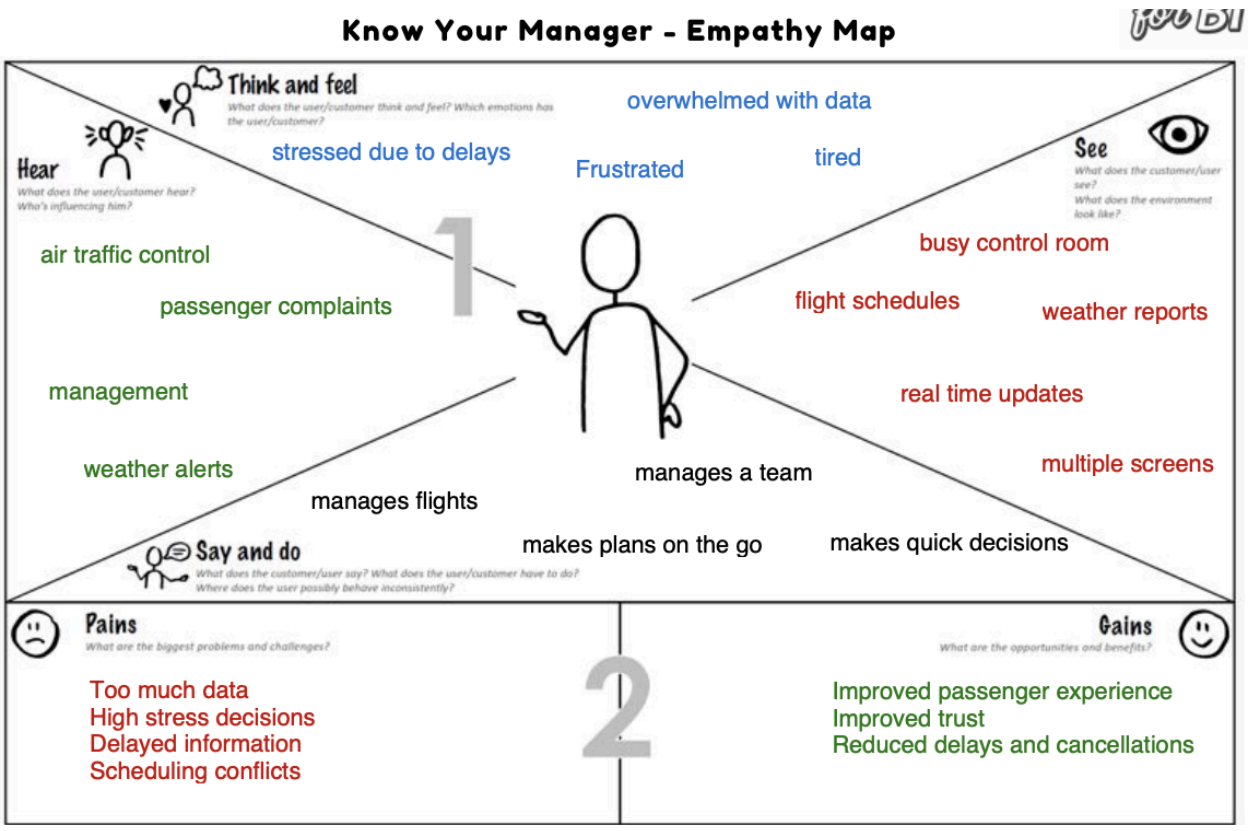
1. **Empathize:**

   We used AI to act as a Civil Aviation Officer (someone that dealt with high-pressure decisions, information overload, and real-time crisis management).

   Come up with more questions and jot them down. Make sure to keep the chat safe as you will be submitting it.

   Act as a civil aviation manager at an international airport, managing flight delays & cancellations.
   Some more Questions:
   How do you monitor and manage flight delays and cancellations?
   Do you use historical data to predict future trends in air traffic or delays?
   How important is real-time data for your daily operations?
   Are there any decisions you wish could be automated with the help of BI tools?
   https://chatgpt.com/share/683496cd-1b44-8002-9494-edf71a5cad25

   We identified their core challenges were “Managing unpredictable delays and cancellations”, “Handling stressed passengers and tight schedules”, and “Navigating data from various, often fragmented, sources (e.g., weather, security alerts)”

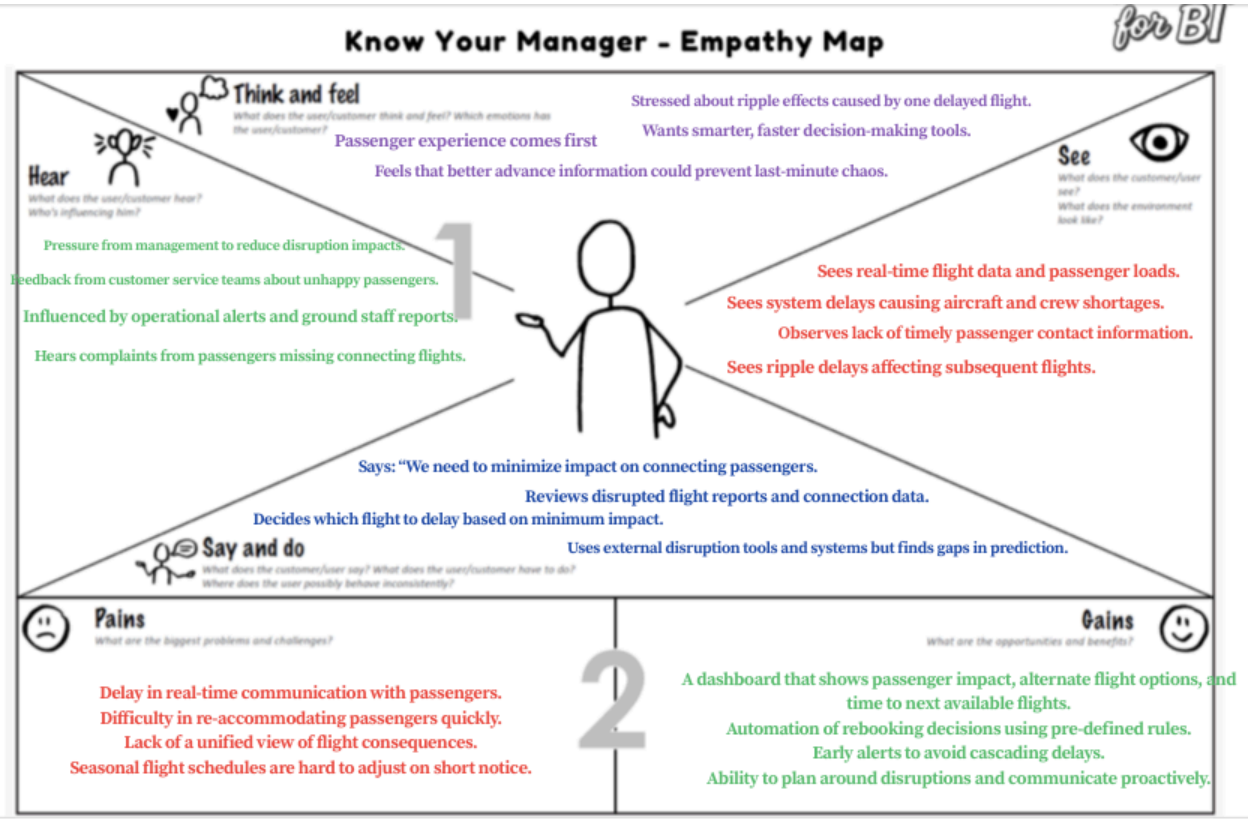


We also interviewed Agha Muhammad Majid Hussain, Head of Network Planning and Market Research at PIA to gather the following insights:

- When a delay occurs, the first data point reviewed is the number of passengers with onward connections. Flights with fewer connecting passengers are preferred for delays to minimize the cascading impact on travel plans. This shows that the priority are Connecting Passengers.
- There is an operational challenge caused by long-term disruption effects such as Covid or the current Pak-India escalations.Flight schedules are fixed seasonally. A single disruption can trigger a domino effect on subsequent flights using the same aircraft, making it hard to recover quickly. Long gaps between flights also lead passengers to cancel and switch airlines.
- Real-Time Data & Communication is Crucial. Airlines generally have access to the necessary data, but quick access and timely action are critical. Having complete passenger contact information at the time of ticket booking helps enable proactive communication to reduce inconvenience.
- What an Ideal Dashboard Should Show:

- Details of the disrupted flight
- Total passengers and number at risk of missing onward flights
- Best alternate flights, their available seats, and new routing options
- Time to next available flight
- Consequential disruptions – other flights likely to be affected due to the current delay

- Recommendation for Automation? Airlines can benefit from disruption management tools that auto-recommend rebooking options and highlight ripple effects across the network using pre-defined parameters.



2. Define:

We then defined How might we Questions to figure out Potential BI Queries and frame and reframe how we viewed the problem.

So what's the business problem?

"[Stakeholder], a [role] from [company], needs a way to [need] because [insight]. Fulfilling this need will help the organization [business impact]."

Fatima, a civil aviation officer from Jinnah International Airport, needs a way to manage excessive flight delays and cancellations due to recent political turmoil. Fulfilling this need will help the organization improve customer satisfaction and ensure smoother flight schedules.

How might we... (1 or 2 statements - Focus on a clear BI problem)

How might we provide Fatima with a real-time BI dashboard that integrates flight status, weather, and political event alerts to proactively manage delays and cancellations and use historical and live data to predict high-risk flights and automate decision-making for gate assignments and passenger communication, ensuring smoother operations.

3. **Ideate:**

We each took time to craft our own solutions to our own set of potential BI queries, and came back and discussed. We delay trend visualizations over time and by airline; categorical breakdowns of delay causes (Carrier, Weather, NAS, etc.; and highlighting high-risk routes.

BI DESIGN SPRINT - IDEATE

Design Thinking for BI

TEAM MEMBER 1 NAME: Kisa Fatima

Come up with as many ideas. One idea per sticky note.

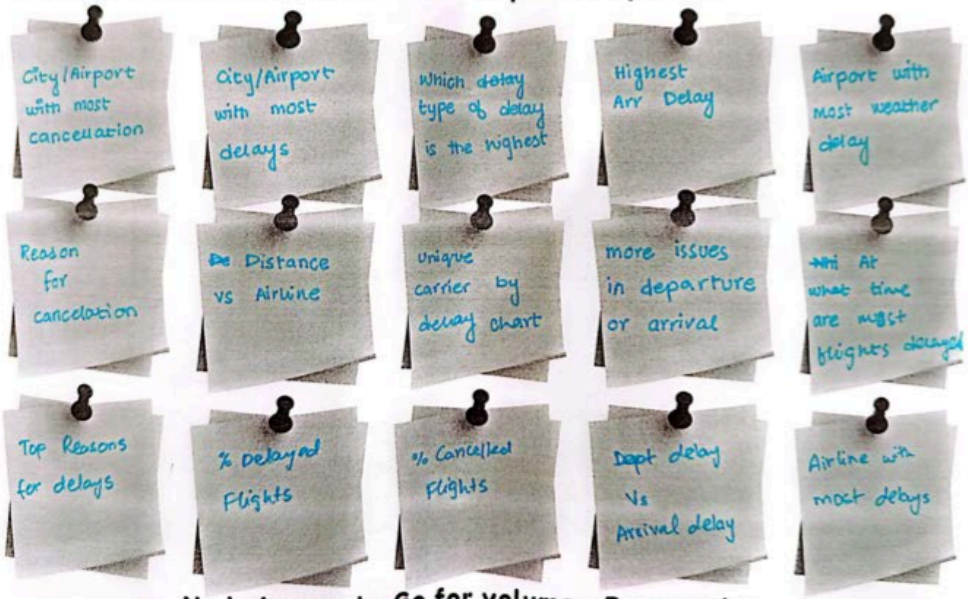
No judgement - Go for volume - Be creative

BI DESIGN SPRINT - IDEATE

Design Thinking for BI

TEAM MEMBER 2 NAME: Farah Inayat

Come up with as many ideas. One idea per sticky note.



No judgement - Go for volume - Be creative

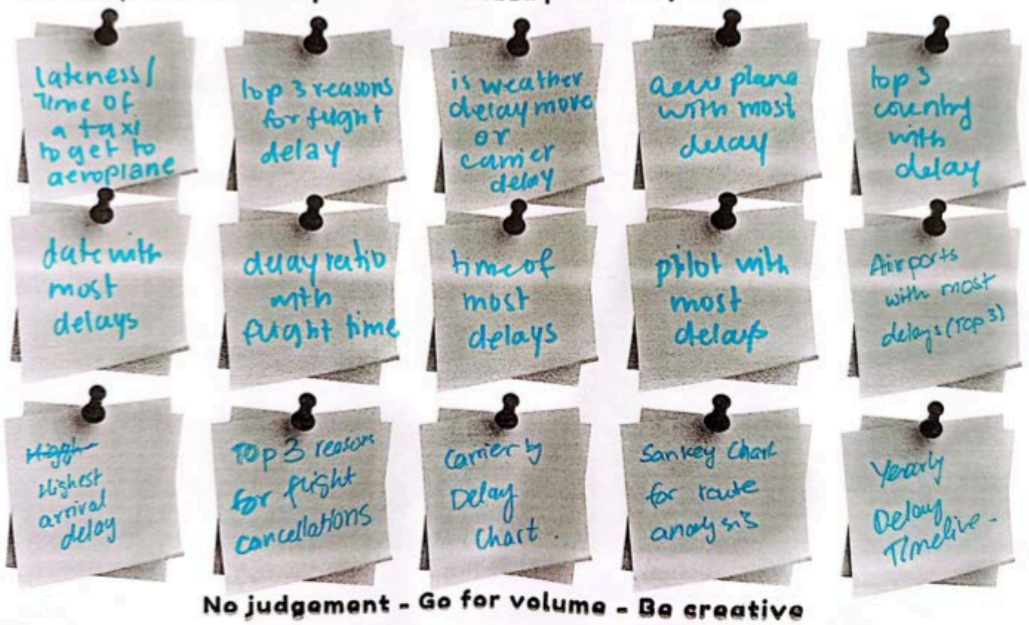
(WE [illegible] AND WE DON'T JUDGE)

BI DESIGN SPRINT - IDEATE

Design Thinking for BI

TEAM MEMBER 3 NAME: Zuha Aqib

Come up with as many ideas. One idea per sticky note.



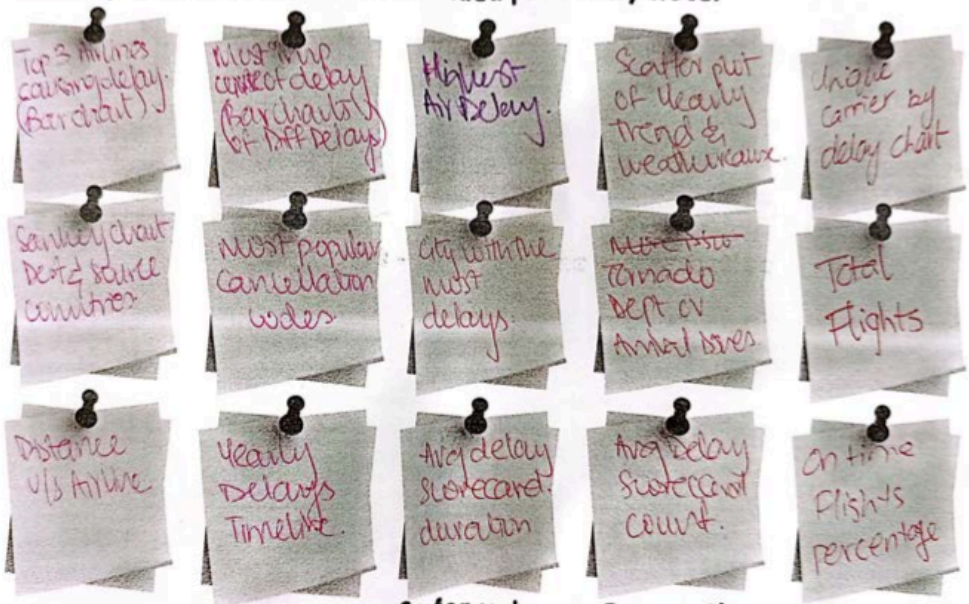
No judgement - Go for volume - Be creative

BI DESIGN SPRINT - IDEATE

Design Thinking for BI

TEAM MEMBER 4 NAME: Zehra Ahmed.

Come up with as many ideas. One idea per sticky note.



No judgement - Go for volume - Be creative

BI DESIGN SPRINT - IDEATE

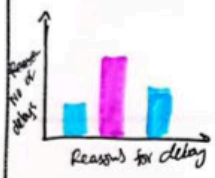
Design Thinking for BI

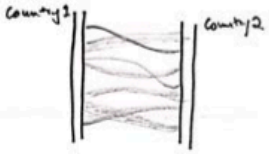
Crazy 4s - Sketch ideas that might be a BI solution
Each member will create this individually.

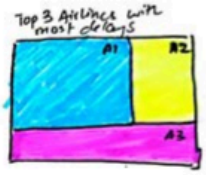
1

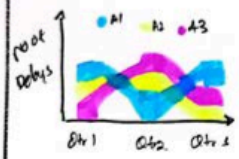
Avg Flight Delay

Flight cancellation %









Unique carriers & flights by

Team member Name

Kisa

Fatima

2

3

4

BI DESIGN SPRINT - IDEATE

Design Thinking for BI

Crazy 4s - Sketch ideas that might be a BI solution
Each member will create this individually.

2

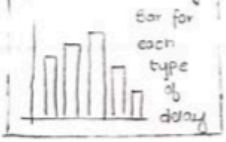
Airport with most delays

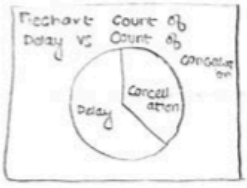
Airport with most cancellations

Highest Arr Delay

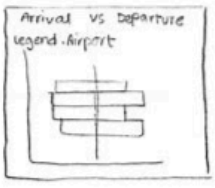
Percentage of flights with Arr. Delay

Bar Chart for Delays

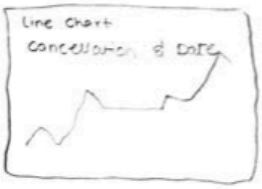




Arrival vs Departure legend. Airport



Line chart Cancellation of Date



Team member Name

Farah

Inayat

1

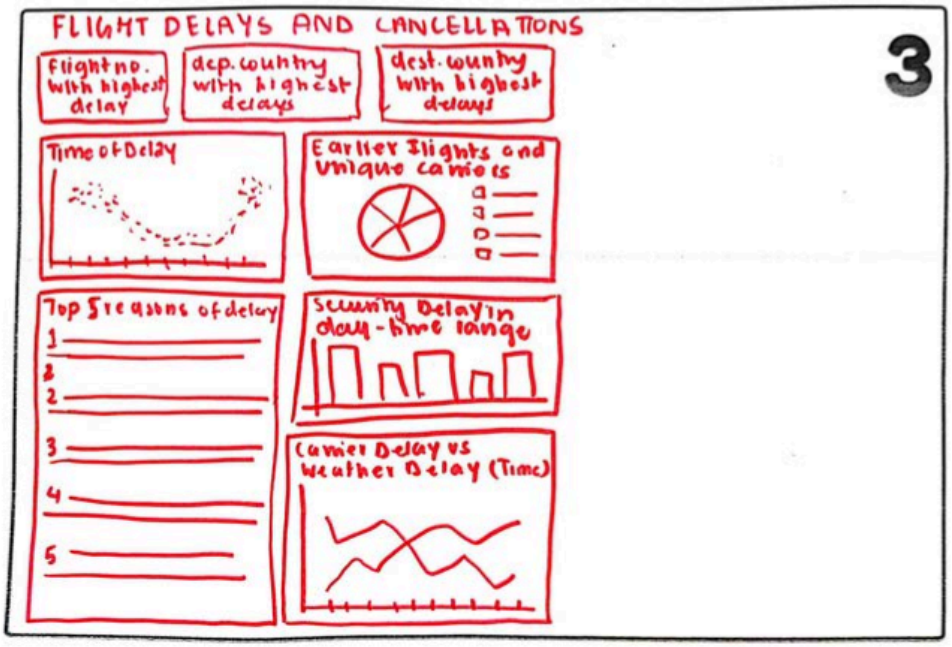
3

4

BI DESIGN SPRINT - IDEATE

Design Thinking for BI

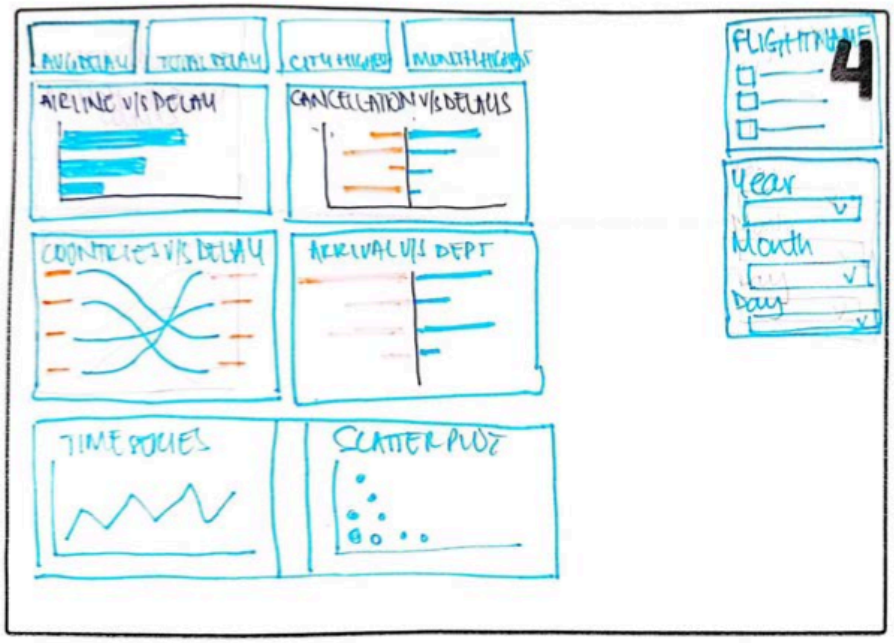
Crazy 4s - Sketch ideas that might be a BI solution
Each member will create this individually.



BI DESIGN SPRINT - IDEATE

Design Thinking for BI

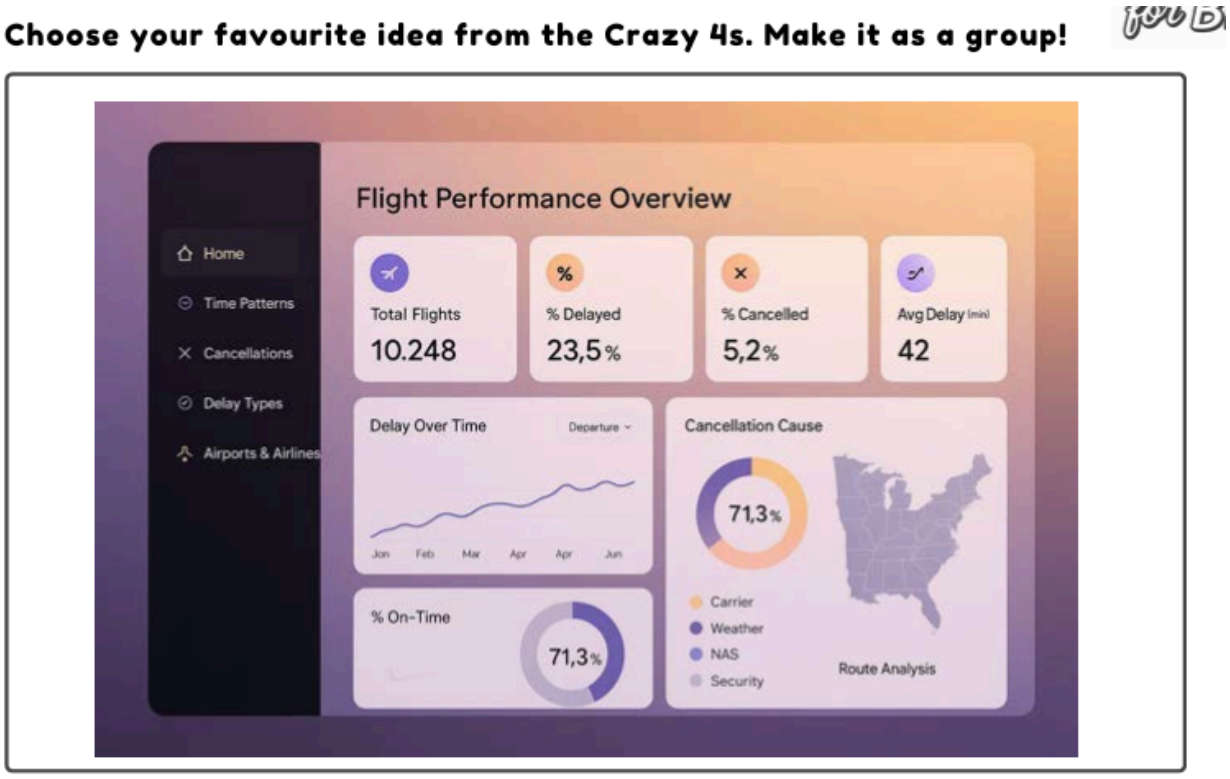
Crazy 4s - Sketch ideas that might be a BI solution
Each member will create this individually.



4. Prototypes:

The selected ideas were translated into a Power BI prototype. Core features included:

- Interactive filters by airline, delay type, and airport
- Clean layouts separating delays from cancellations
- Use of time blocks and grouped delay intervals for intuitive reading



We ensured the prototype mirrored real use-case scenarios and supported quick decision-making.

5. Testing:

We gave the prototype back to GPT and asked it to evaluate our progress

Upload the prototype picture & ask your AI expert for feedback!

Label axes and charts more clearly (e.g., Y-axis for bar/line graphs).
Replace “Qtr 1, Qtr 2…” with actual months or dates for clarity.
Add titles/captions to each graph for quick interpretation.
Consider tooltips or hover info in the final version for extra data detail.
Group related metrics together visually (e.g., all delay-related charts in one area).

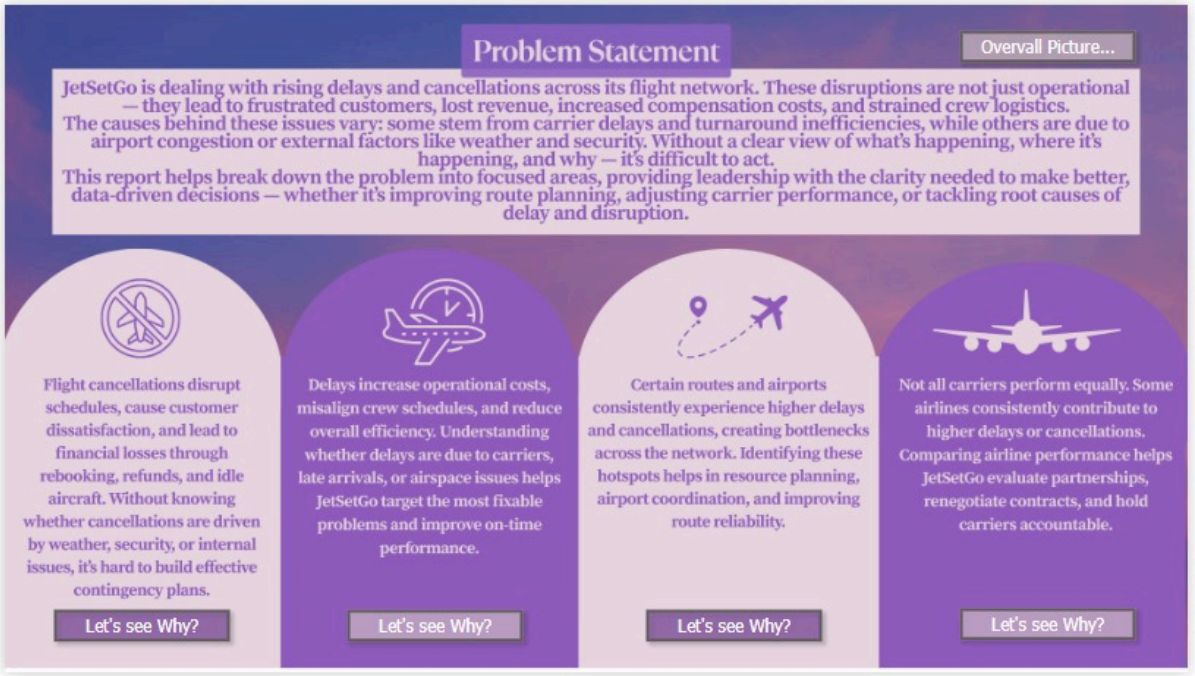
What can you improve? Reflect as a team!

Use the good practices we learned in class.
Make a readable and clean dashboard
There is a repetition between filters and charts that would act as a filter.
Positioning of the charts.
Add more charts.

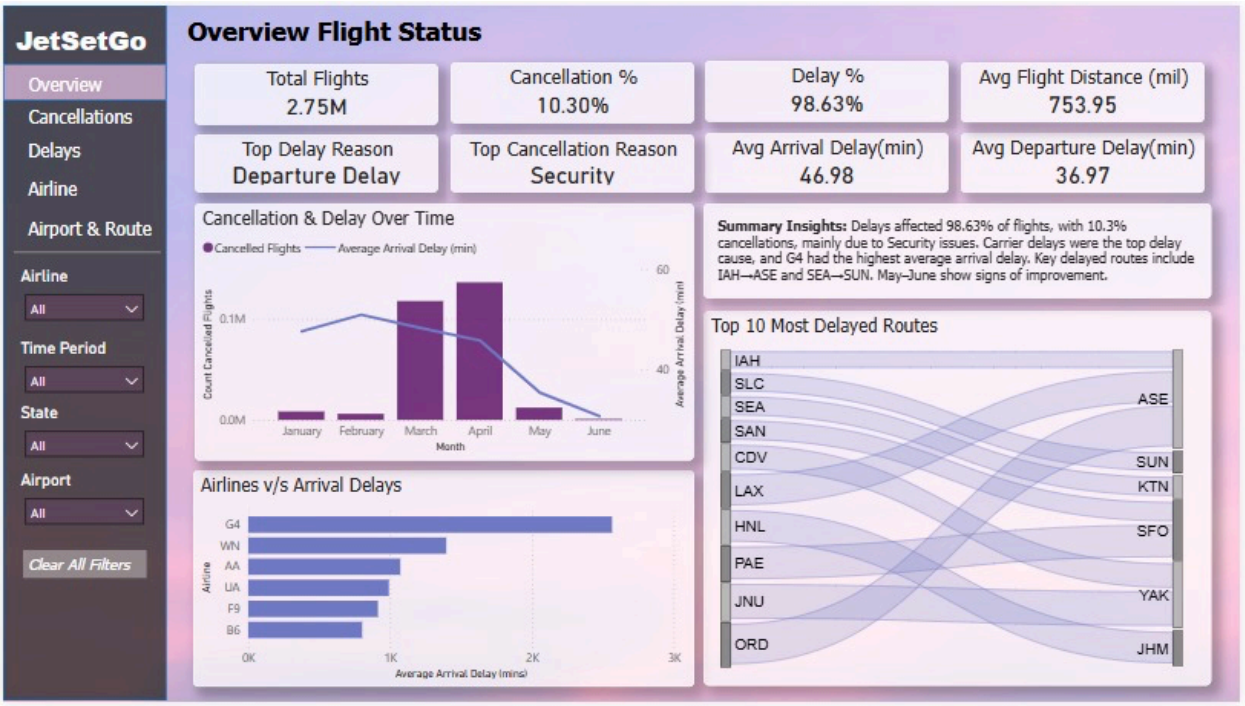
BI REPORT:

Title Page:

Problem Statement:

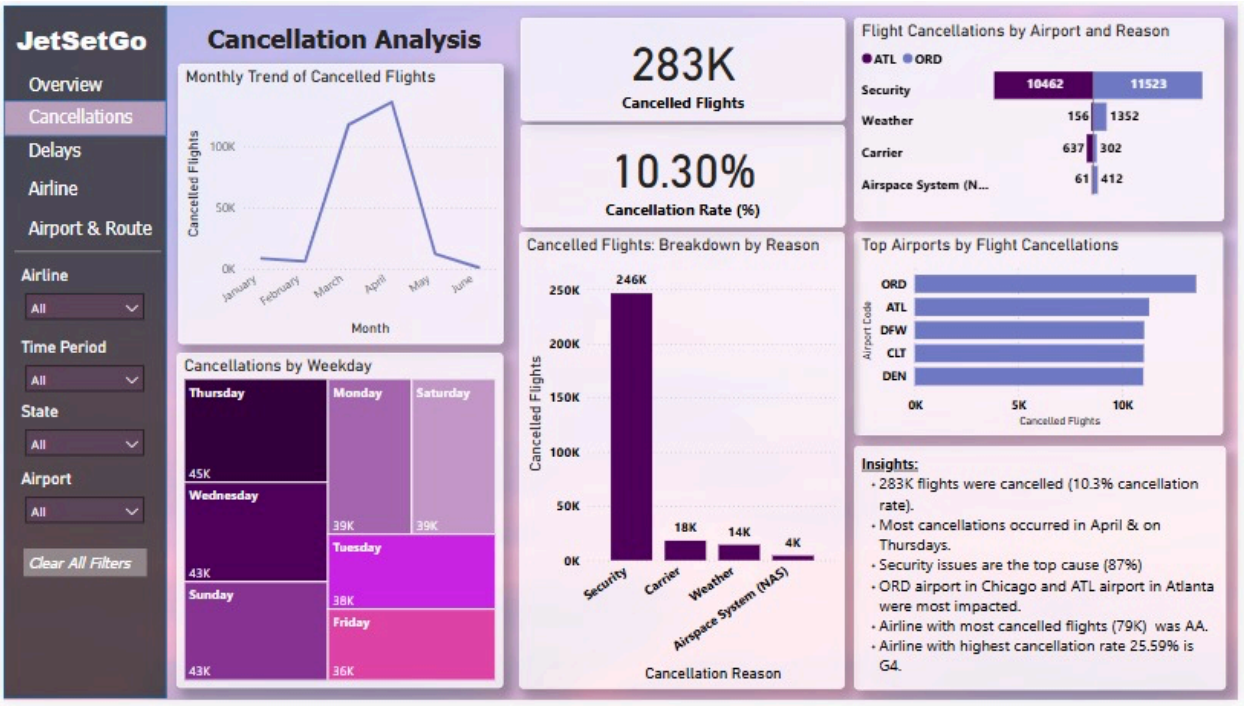


Dashboard 1: Overall Flight Disruption Status:



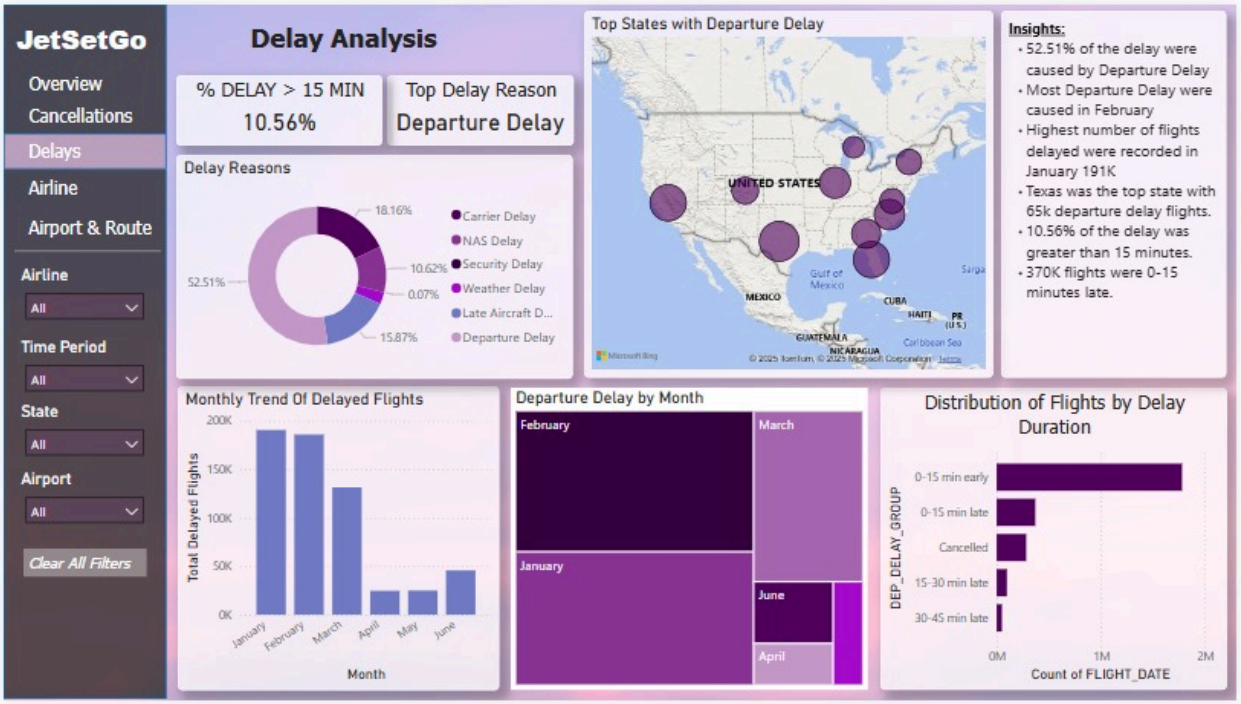
The purpose of this page is to give an executive level overview of the data at hand - that is the overall picture of flight cancellations, delays, and the top most reasons, problematic airlines and routes. This addresses two main questions, firstly “How are flight operations performing overall” followed by “Where should leadership focus attention nex?” Each Visual has a drillthrough attached that leads to the next page of insights and explorations.

Dashboard 2: Root Cause Analysis of Cancellations:



The purpose of this page is to identify what is causing Cancellations, how severe the issues is, and when Cancellations are most prone - in terms of time, state, airport, arrivals vs departures.

Dashboard 3: Root Cause Analysis of Delays :



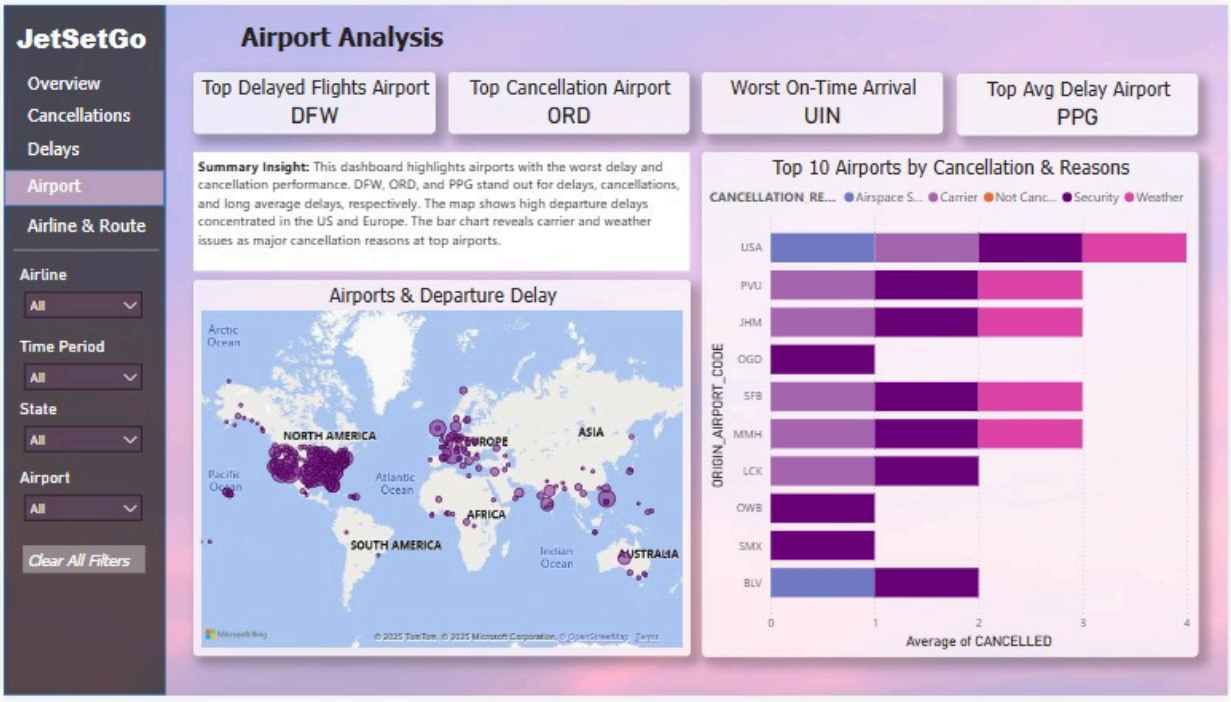
The purpose of this page is to identify what is causing Delays, how severe the issues are, and when delays are most prone - in terms of time, state, airport, arrivals vs departures.

In both Dashboard 2 and 3, knowing the root cause can help stakeholder in upcoming strategic planning to solve the problem from the base.

Dashboard 4: Delay Drill-Throughs

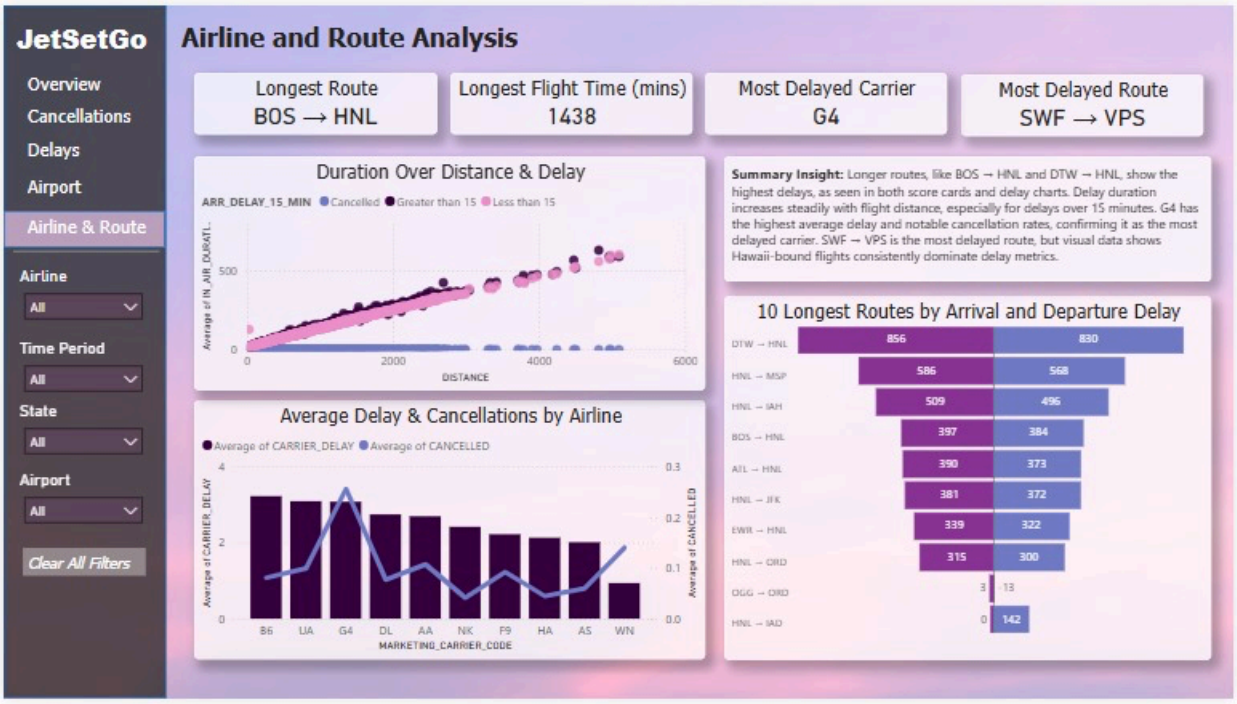
We drill through Month wise to get a holistic view of how delays are spread across geographical and weekly + hourly patterns.

Dashboard 5: Airport Analysis



This dashboard highlights airports with the worst delay and cancellation performance. DFW, USA, and PPG stand out for delays, cancellations, and long average delays, respectively. The map shows high departure delays concentrated in the US and Europe. The bar chart reveals carrier and weather issues as major cancellation reasons at the top airports with cancellations.

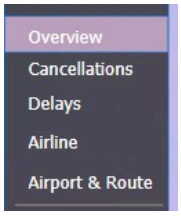
Dashboard 6: Airline $ Route Analysis



Longer routes, like BOS → HNL and DTW → HNL, show the highest delays, as seen in both score cards and delay charts. Delay duration increases steadily with flight distance, especially for delays over 15 minutes. G4 has the highest average delay and notable cancellation rates, confirming it as the most delayed carrier. SWF → VPS is the most delayed route, but visual data shows Hawaii-bound flights consistently dominate delay metrics.

BI Charts:

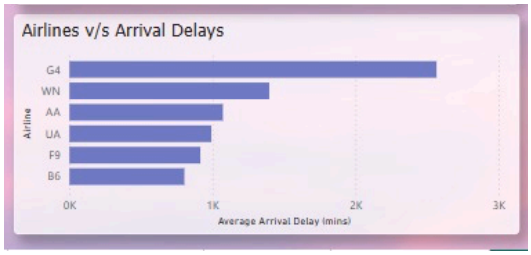
Dashboard 1:



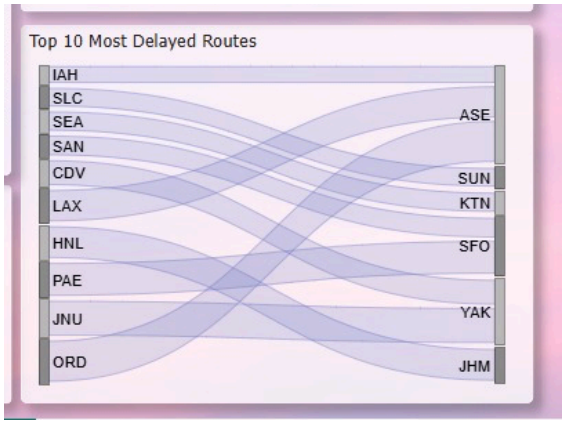
There is a side tab which allows for easy navigation through the multi-page report. This is in addition to the drill through links.

The KPI cards present the most urgent metrics, such as Total Flights (this provides the baseline context into the volume of flights being handled), % of flights cancelled, % of flights facing delay of at least 15 mins, the average distance of flights, top delay reasons, avg arrival and departure delays.

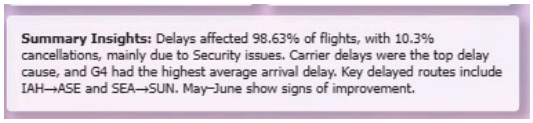
This chart shows both cancellation volume and average arrival delay over the 6 month period. Both are mapped on the same visual to show correlation between cancellations and delays such that is improving one impacting the other, or are the two linked together? The time trend presents an insight if our performance has been improved or not over time.



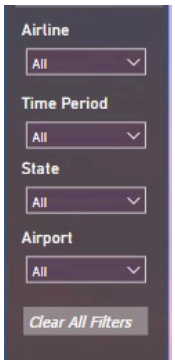
This gives the top 6 most problematic airlines in terms of arrival delays overall. It shows that G4 has significantly more delays then the rest of the flights on average.



Sankey Chart provides an insight into the 10 most problematic routes based on average delay time.



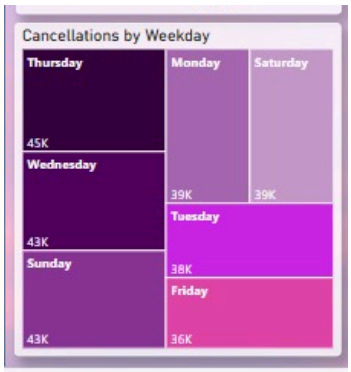
The stakeholder is provided with byte sized insight summary of the entire dashboard to make sense of the overall picture at a glance.



All visuals and KPIs are completely dynamic showing cancellations, delays and cause of concern based on Airline, State, Airport, Time.

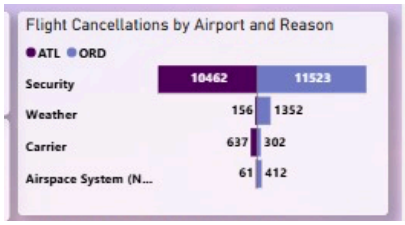
Dashboard 2:

This chart shows month wise trend of cancellations, now had it been for a general dataset we could have said it was to show month over month trend of cancellations, however we see that during Covid cancellations spiked in March and peaked in April before returning to normal in May.



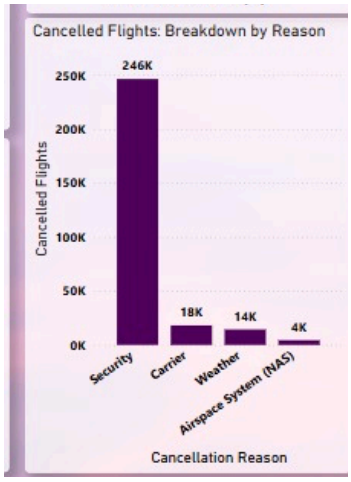
Weekday vs Weekend Trend Analysis of Cancellations, with most number of delays on Thursday and least on Friday. No such pattern seen between Weekend and Weekday.

These KPIs are to show severity of cancellations at a glance.

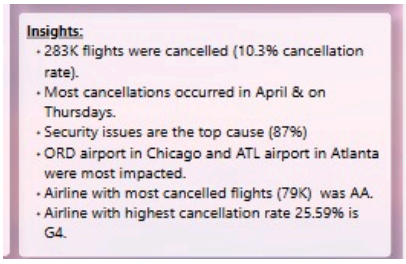


Cancellation Reason analysis of top two worsts airports shows that security reasons are the singlest most impactful factor for flight cancellations, significantly higher than the other.

Figuring out the top 5 worst airports with the most cancellations.

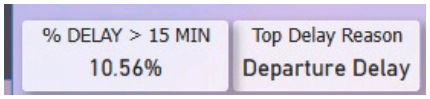


This shows that security is the single most important cause of cancellation overall and not just for the top 2 worst airports.

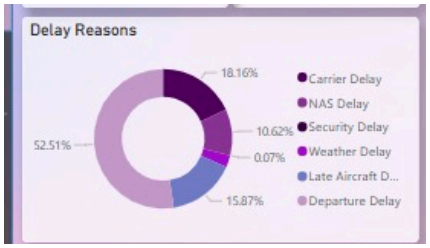


The stakeholder is provided with byte sized insight summary of the entire dashboard to make sense of the overall picture at a glance.

Dashboard 3:



These KPIs are to show severity of delay at a glance and also shows that departures cause more delays than arrivals.



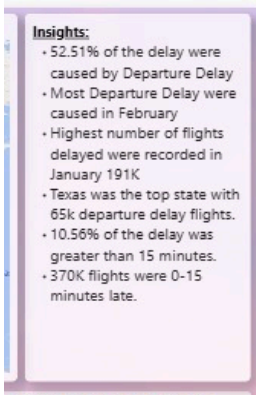
This shows severity of the various causes of delays.

State wise analysis of the most impacted geographical areas by delay.

This shows delays peaked at January and declined in April and May. This is in stark contrast to cancellations month-wise pattern. Very likely because due to cancellations, flights weren't there to get delayed.

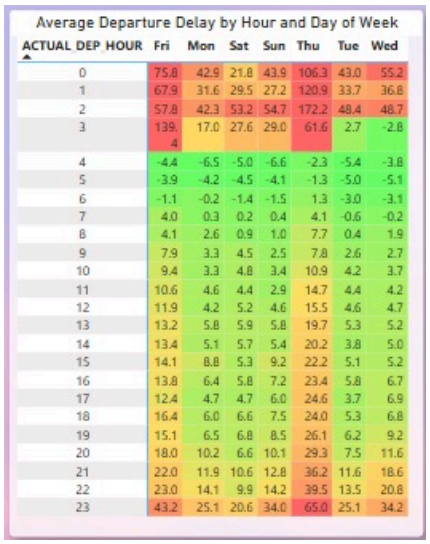
Here we are looking only specifically at departure delays by month. Previous chart was both arrival and departure delays combined.

The most common time delay is of at most 15 mins. This shows that even if delays take place they arent that severe, however they do get accumulated overtime creating a ripple effect.



The stakeholder is provided with byte sized insight summary of the entire dashboard to make sense of the overall picture at a glance.

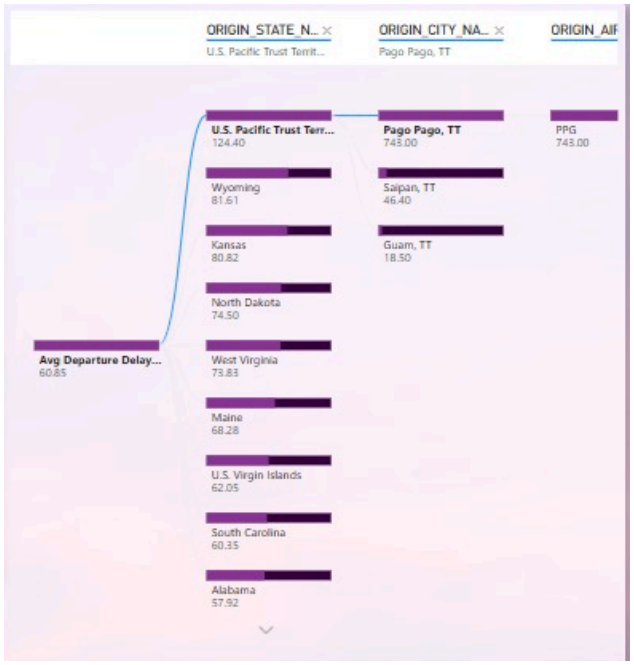
Dashboard 4:



Average Departure Delay by Hour and Day of Week

ACTUAL_DEP_HOUR	Fri	Mon	Sat	Sun	Thu	Tue	Wed
0	75.8	42.9	21.8	43.9	106.3	43.0	55.2
1	67.9	31.6	29.5	27.2	120.9	33.7	36.8
2	57.8	42.3	53.2	54.7	172.2	48.4	48.7
3	139.4	17.0	27.6	29.0	61.6	2.7	-2.8
4	-4.4	-6.5	-5.0	-6.6	-2.3	-5.4	-3.8
5	-3.9	-4.2	-4.5	-4.1	-1.3	-5.0	-5.1
6	-1.1	-0.2	-1.4	-1.5	1.3	-3.0	-3.1
7	4.0	0.3	0.2	0.4	4.1	-0.6	-0.2
8	4.1	2.6	0.9	1.0	7.7	0.4	1.9
9	7.9	3.3	4.5	2.5	7.8	2.6	2.7
10	9.4	3.3	4.8	3.4	10.9	4.2	3.7
11	10.6	4.6	4.4	2.9	14.7	4.4	4.2
12	11.9	4.2	5.2	4.6	15.5	4.6	4.7
13	13.2	5.8	5.9	5.8	19.7	5.3	5.2
14	13.4	5.1	5.7	5.4	20.2	3.8	5.0
15	14.1	8.8	5.3	9.2	22.2	5.1	5.2
16	13.8	6.4	5.8	7.2	23.4	5.8	6.7
17	12.4	4.7	4.7	6.0	24.6	3.7	6.9
18	16.4	6.0	6.6	7.5	24.0	5.3	6.8
19	15.1	6.5	6.8	8.5	26.1	6.2	9.2
20	18.0	10.2	6.6	10.1	29.3	7.5	11.6
21	22.0	11.9	10.6	12.8	36.2	11.6	18.6
22	23.0	14.1	9.9	14.2	39.5	13.5	20.8
23	43.2	25.1	20.6	34.0	65.0	25.1	34.2

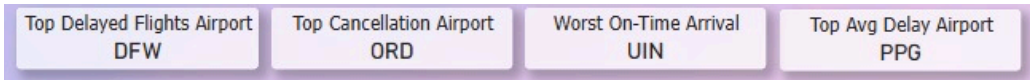
This heatmap shows severity by Average departure delay across hour wise distribution for each day of the week.



We move through hierarchy on Departure State > CIty > Airport to find precisely where the problem lies and how to address at the very root level moving up.

Dashboard 5:

Score Cards



Top Delayed Flights Airport: DFW

Dallas/Fort Worth (DFW) has the highest number of delayed flights.

Top Cancellation Airport: USA

The airport with the code "USA" experiences the most cancellations overall.

Worst On-Time Arrival: UIN

Quincy Regional (UIN) has the worst record for on-time flight arrivals.

Top Avg Delay Airport: PPG

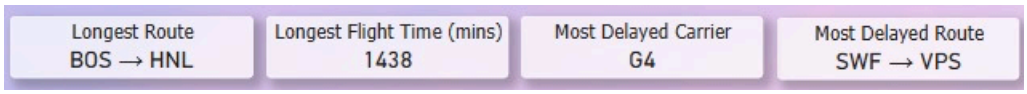
Pago Pago International (PPG) has the highest average delay time per flight.

A global map visualizing airports by departure delay. The size of the circle represents the magnitude of departure delays. There are denser clusters in the US and Europe.

A horizontal stacked bar chart showing the top 10 airports with the most cancellations, broken down by specific reasons like weather, carrier issues, or airspace constraints. Security is a reason for cancellation in all the top 10 airports.

Dashboard6:

Score Cards

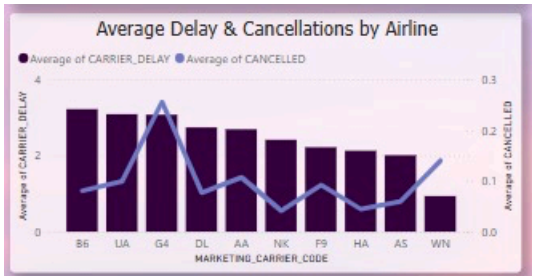


Longest Route (BOS → HNL): Indicates the farthest route in terms of distance.

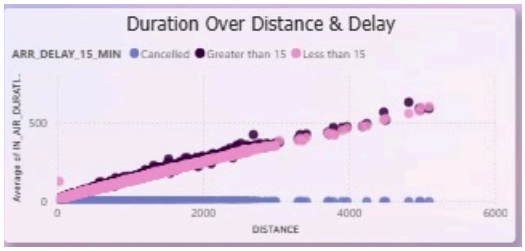
Longest Flight Time (1438 mins): Highlights the maximum airborne duration observed.

Most Delayed Carrier (G4): Shows the airline with the highest average delays.

Most Delayed Route (SWF → VPS): Identifies the route suffering the most from delays.



The bars represent the average carrier delay, and the line shows the average cancellation rate. G4 has the highest cancellation rate and a very high average carrier delay, making it the least punctual airline. WN has the lowest delays and moderate cancellations, showing better overall performance.

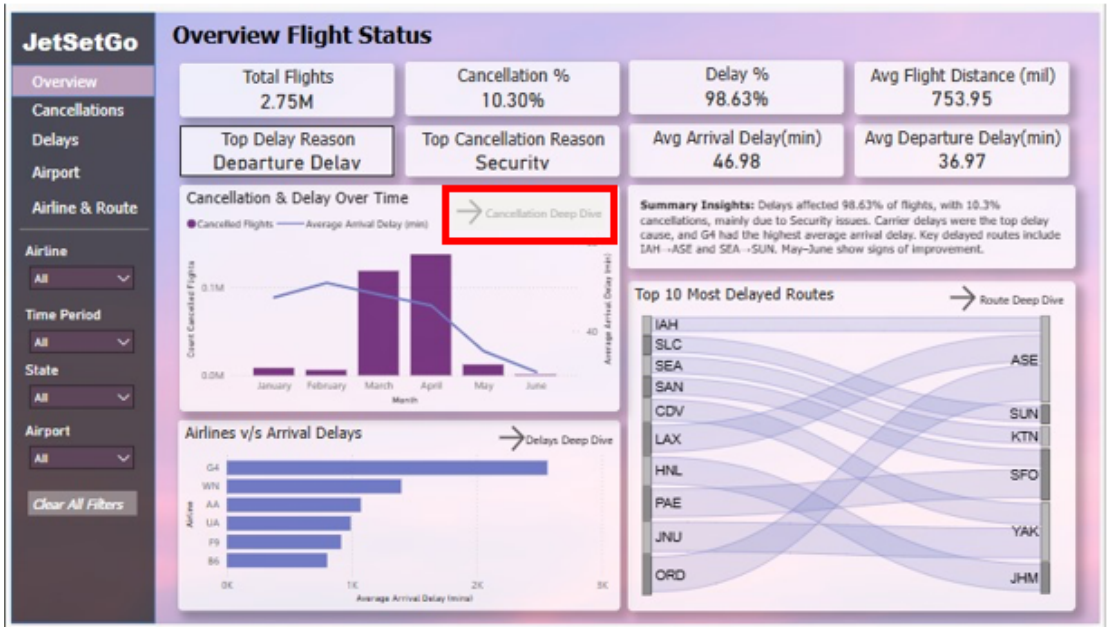


As flight distance increases, in-air duration and delays rise steadily. Flights delayed more than 15 minutes and less than 15 minutes have a similar trend. More than 15 minutes delayed flights always have a slightly higher duration over distance. Cancelled flights can be seen at the bottom as they have no in-air duration due to cancellation.

DTW → HNL has the highest cumulative delay time, followed closely by HNL → MSP and HNL→ IAH, all showing over 1000 minutes of total delays. Routes involving HNL (Honolulu) dominate the list, indicating that Hawaii-bound travel is consistently delayed. The chart reflects that longer flights are heavily impacted by both departure and arrival delays.

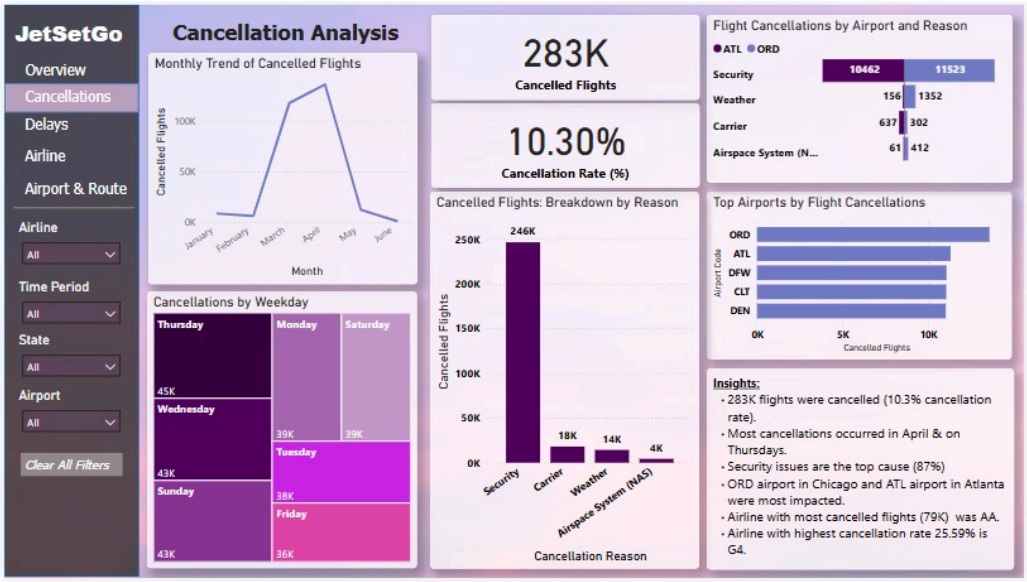
Data Story:

1. **Root Causes of Disruption: What's going wrong and why?**



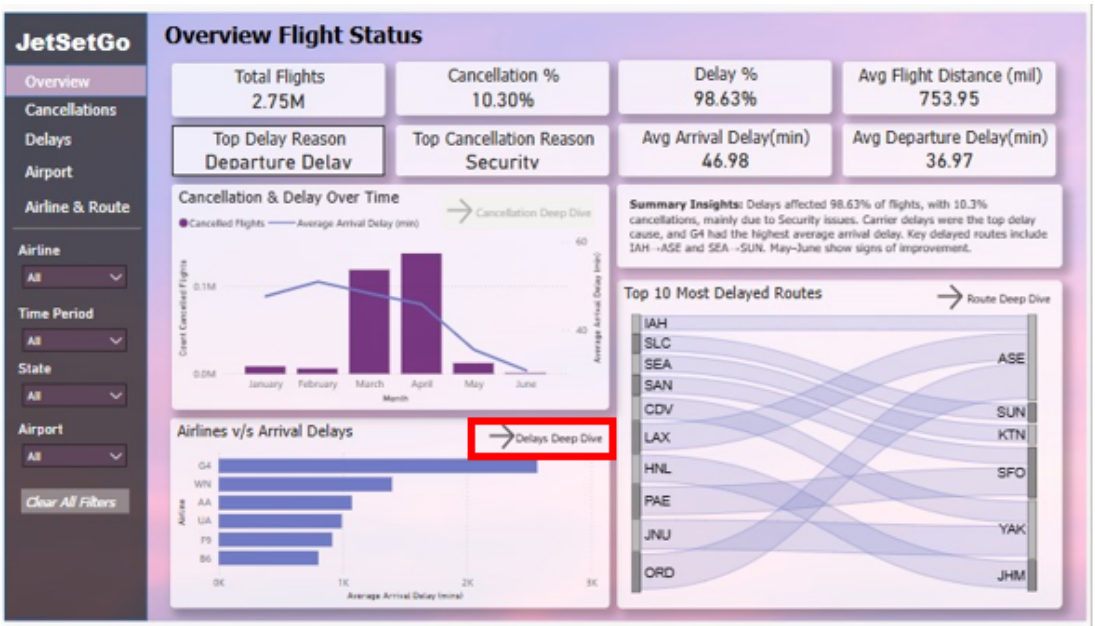
We start off from the overall dashboard above, We see that security and Departure Delay are the leading culprits behind travel disruptions, accounting for the majority of cancellations and delays respectively.

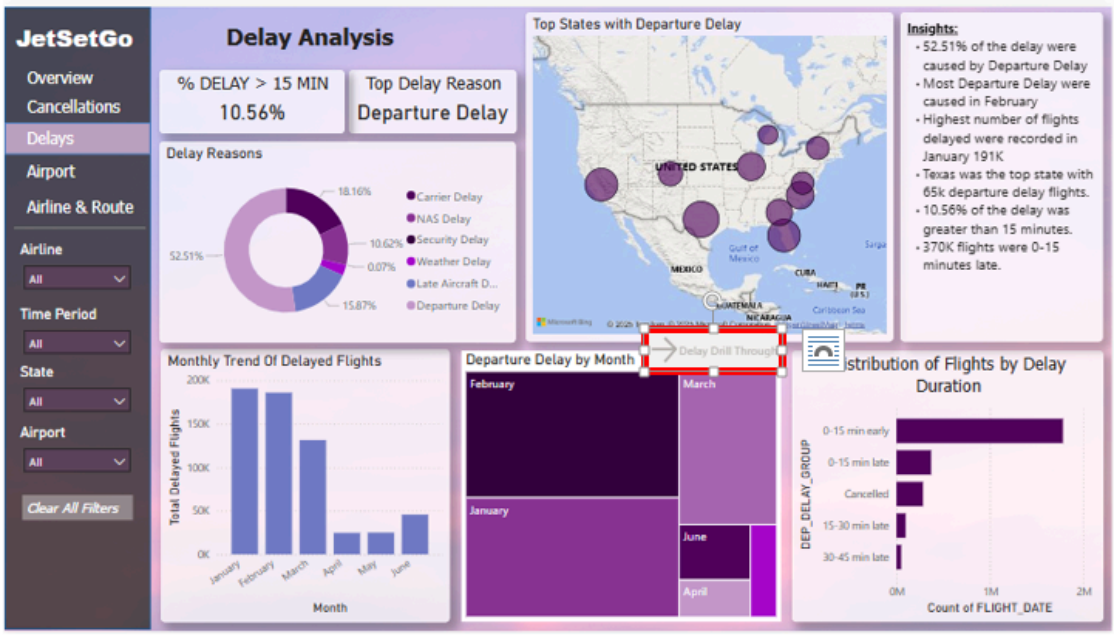
Therefore as a next step, we click on drill throughs for cancellations and delays and see exactly what the problem is.



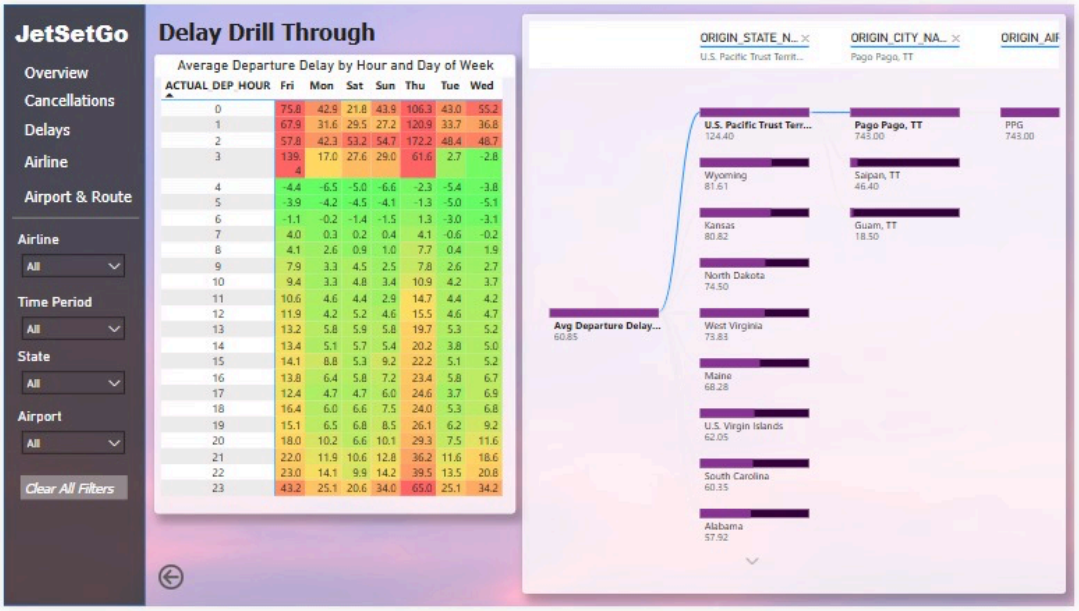
Now on this dashboard we are able to see that most cancellations happen on Thursday, and impact 283k flights. Top cancellation reason being security (246K out of 283K cancellations = ~87%).

Likewise, return back to the overall dashboard again, this time drilling through on delays...



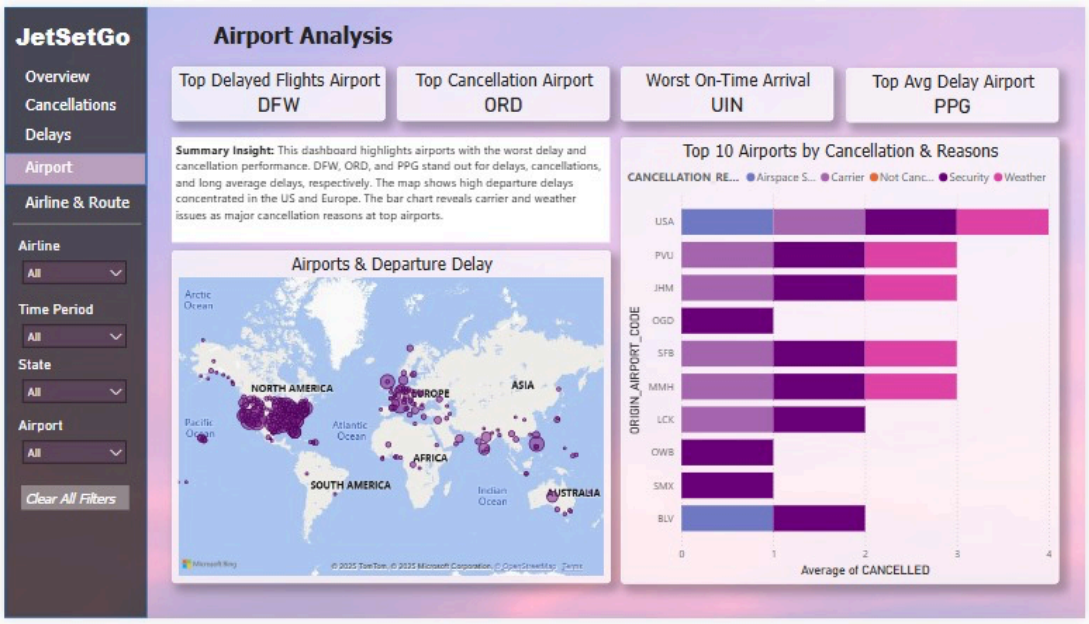


Now on this dashboard we see that severe Delays, which are greater than 15 min, impact around 10.56% of the flights and the top delay reason is delayed departures, which impacts an overall of 52.51% of delayed flights. We see that these are the most abundant in Feburary so let's drill down month wise and see the problematic geographical areas that is the most problematic State then city then airport. And we can also pin point exactly what hour and day of the week is contributing these the most.



Solving these two key aspects can help solve 10.3% cancellation rate, 98.63% flights delayed, avg. arrival delay: 46.98 mins thereby improving efficiency greatly.

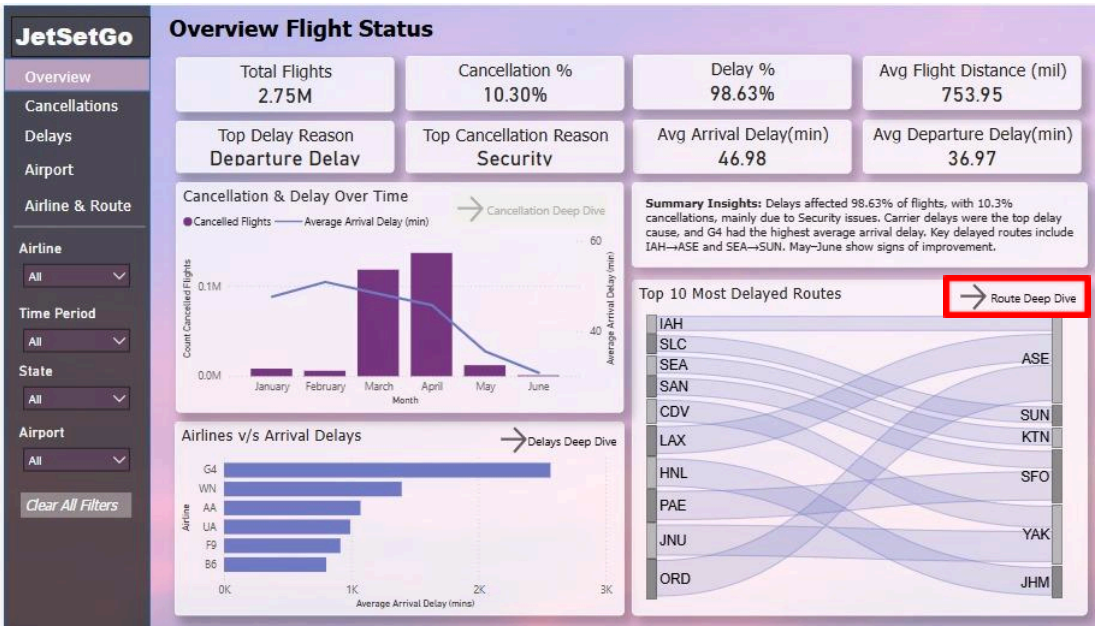
2. Airport Deep Dive



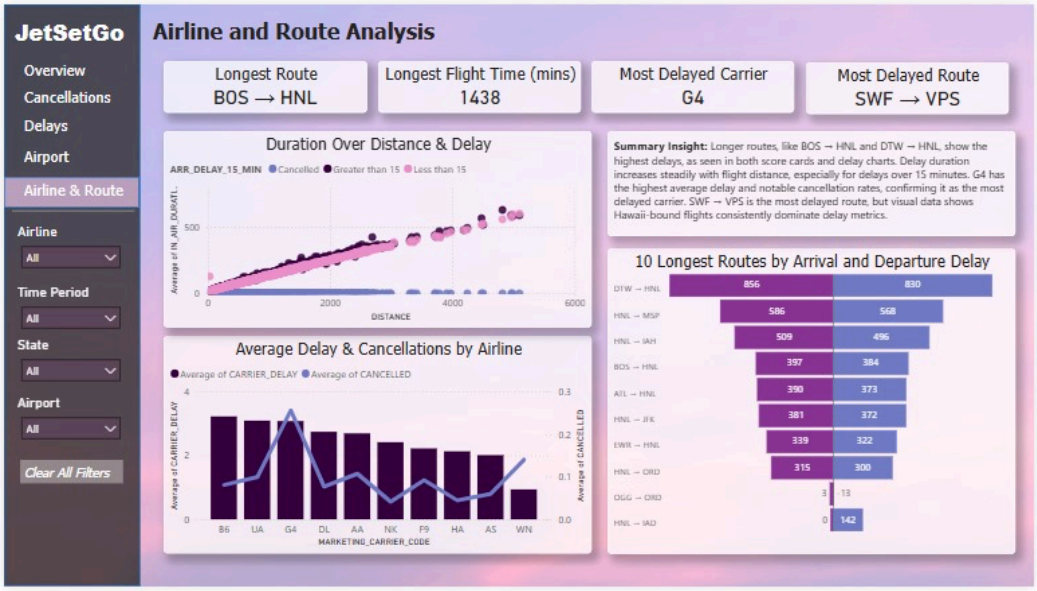
After understanding that delays are a major disruption in the overall view, we drill into the Airport Comparison Dashboard to see where these delays are happening most frequently.

We observe that DFW, ORD, UIN, and PPG stand out with the highest number of delays and cancellations, making them key contributors to overall disruptions. Delays are clustered around the USA and Europe. The top cancellation reason is Security, which is prevalent in all the top cancellation rate airports.

We return to the overall dashboard and now deep dive into route analysis.



3. Airline & Route Deep Dive



Shifting focus to the Airline & Route Analysis Dashboard, we dive deeper into which routes and airlines are driving the delays.

From the visuals, it's clear that G4 (Allegiant Air) has the highest average carrier delay, signaling a need to investigate its scheduling or operational issues. The most delayed route SWF → VPS. Additionally, routes like DTW → HNL and HNL → MSP consistently face the highest combined departure and arrival delays.

This dashboard helps us pinpoint which carriers and routes need urgent intervention to improve timeliness and passenger experience.

Conclusion: Targeted Action for Maximum Impact

By following a structured deep dive from identifying root causes to exploring delays, cancellations, airports, airlines, and routes, we now have a **clear map of disruption hotspots** in the flight ecosystem. Security issues and delayed departures stand out as the **primary drivers** of cancellations and delays, with specific airports (like ORD and DFW), routes (such as DTW → HNL), and carriers (notably G4) being consistently high-risk. Addressing just **a handful of critical factors** security protocols, departure efficiency, and underperforming routes/airlines could lead to significant improvements in travel reliability, reducing delays and cancellations, and ultimately enhancing the passenger experience across the board.

Work Contribution:

1. Kisa Fatima: Design Sprint + Airline Analysis Dashboard + Airport/Route Analysis Dashboard.
2. Zuha Aqib: EDA/Data Cleaning + Root Cause Delays Dashboard
3. Zehra Ahmed: Background Knowledge Template Document + Overview of Flight DIsruption Status Dashboard + Interview
4. Farah Inayat: Video + Report + Root Cause Cancellation Dashboard + Interview Empathy Map