Final Group Project

Evidence-Based Strategy Design for Innovation and Operational Excellence

"Reducing Flight Delays Through Data-Driven Analysis: Improving U.S. Airline Operational Efficiency"

ISM 6155: Enterprise Information Systems Management

Dr. Nasir M Uddin

Group 1:

Aishwarya Gopal

Andrey Martynenko

Debjani Sarma

Dushyanth Asrani Gopal

Saketh Polavarapu

Executive Summary

Flight delays and cancellations have become a growing concern for the U.S. aviation industry, especially after the COVID-19 pandemic. These disruptions affect customer satisfaction and also create operational inefficiencies and financial burdens for airlines. This project analyzes airline performance data from 2019 to 2023 using a dataset from Kaggle, originally derived from the U.S. Department of Transportation's Bureau of Transportation Statistics. The analysis focuses on identifying delay trends across time, carriers, airports, and causes to pinpoint the most critical factors behind recurring disruptions. By combining data visualizations in Tableau with strategic business frameworks, the project outlines opportunities to improve turnaround operations, enhance schedule reliability, and offer more consistent service. Our main recommendation emphasizes data-driven adjustments to help airlines strengthen customer trust and optimize resources in a competitive post-pandemic travel environment.

Problem Definition

Since the pandemic, airlines have faced an ongoing challenge with delays and cancellations. As demand for air travel continues to rise, carriers are finding it difficult to maintain reliable schedules, largely due to staffing shortages, airspace congestion, and weather-related disruptions. For airline alliances trying to stay competitive, improving operational efficiency is a top priority. This report seeks to answer the question: how can airlines reduce delays and cancellations using data insights to guide smarter operational planning? By examining which variables, such as airline, airport, season, or time of day contribute most to disruptions, the goal is to provide specific and actionable recommendations for improving turnaround times and flight reliability while maintaining cost-efficiency and customer satisfaction.

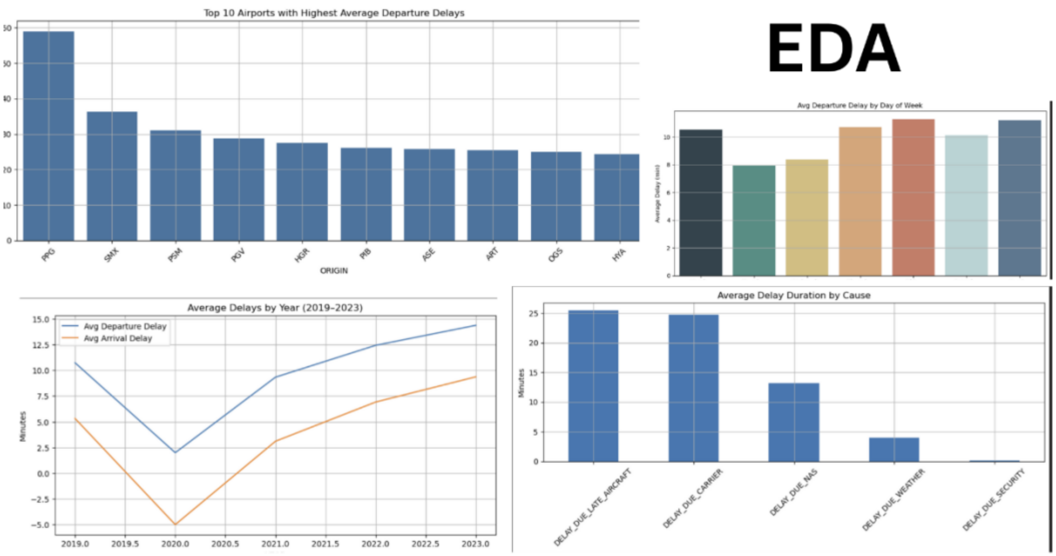
Data Overview

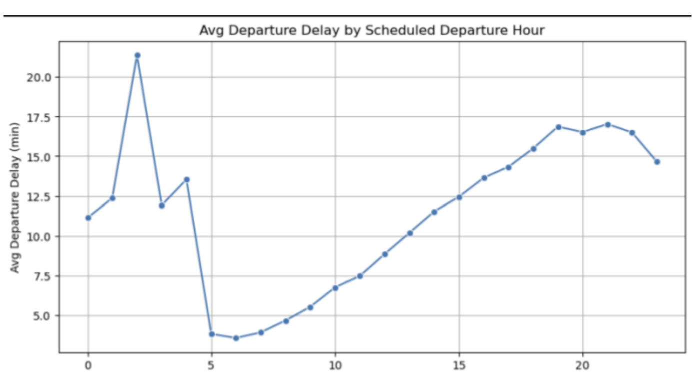
The dataset used for this project comes from Kaggle and includes over 7 million U.S. domestic flight records spanning from 2019 to 2023. It contains detailed information such as flight dates, airline names, airport identifiers, scheduled and actual departure/arrival times, delay durations, cancellation markers, causes of delay, distance, and taxi times. This depth of information makes the dataset suitable for analyzing performance across time, region, and airline, allowing for more accurate conclusions about the U.S. airline industry's reliability and efficiency.

Methodology

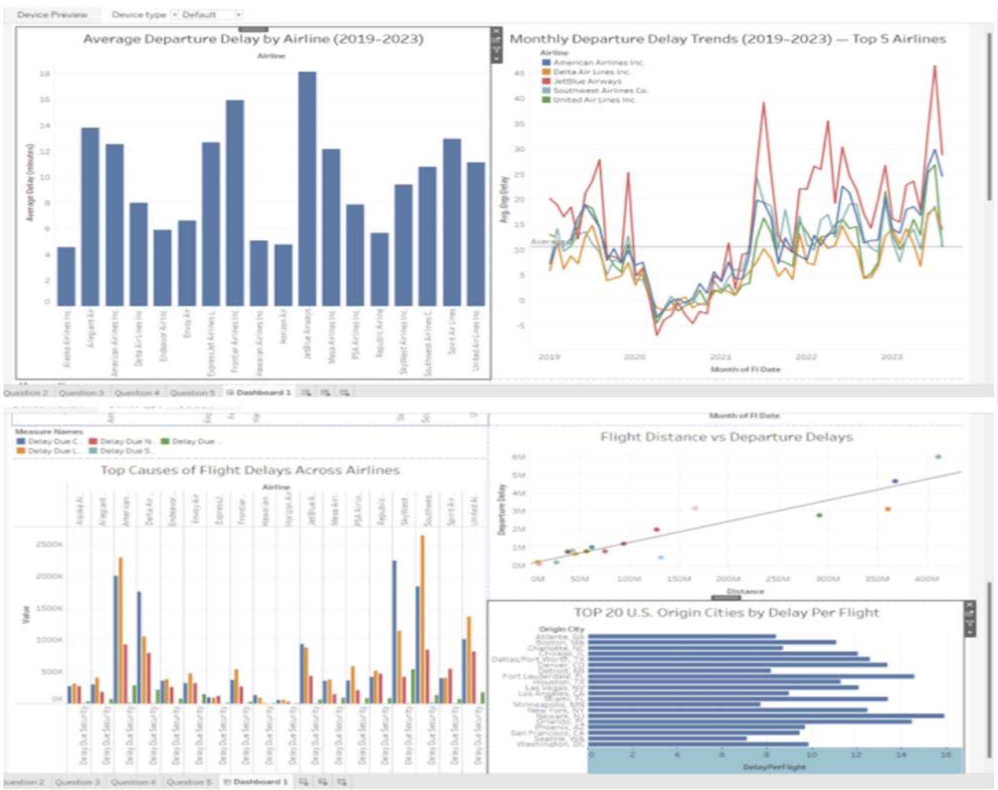
The analysis is based on five years of U.S. domestic flight data, made available through Kaggle and originally sourced from the Department of Transportation. Using Python and Tableau, the data was cleaned and processed to prepare for descriptive and visual analysis. Unused columns were removed, and rows with missing values in key fields were dropped. New fields were created, such as a binary cancellation flag and a Month-Year field, to support trend-based visualizations. The cleaned dataset was exported as a CSV file for use in Tableau. These steps helped isolate core delay metrics and compare performance across multiple dimensions, such as airline, airport, month, and time of day.

Data Analysis Summary





The exploratory analysis revealed several patterns from 2019 to 2023. Both average departure and arrival delays increased over time, peaking in 2023, which signals ongoing strain in post-pandemic recovery efforts. Regional airports like Pago Pago (PPG) and Santa Maria (SMX) showed particularly high average delays, likely due to limited resources or geographic challenges. Flights scheduled early in the morning had the fewest delays, while evening flights saw higher disruptions, especially after 6 PM. Mondays, Fridays, and Sundays were identified as the days with the most delays. The most frequent causes were related to late-arriving aircraft and airline operations, with each averaging over 25 minutes. Severe delays defined as over 60 minutes nearly tripled in frequency between 2020 and 2023. These findings directly influenced the strategic recommendations proposed later in the report.



Summary of Dashboard Insights (2019–2023)

- JetBlue and Frontier had the highest average delays; Alaska and Endeavor performed best.
- JetBlue's delays spiked post-2021, while all top airlines showed rising delay trends after COVID.
- Carrier-related issues were the top cause of delays across most airlines—more than weather or NAS.
- Longer flights showed a clear link to higher delays, suggesting greater operational complexity.
- Major hubs like Fort Lauderdale, Miami, and Atlanta faced the worst delays per flight, highlighting regional congestion hotspots

Market & Competitive Analysis

The U.S. airline industry remains highly competitive, with major players like American, Delta, and United operating alongside low-cost carriers such as Southwest, JetBlue, and Frontier. Post-pandemic travel recovery has made operational reliability a key factor in customer satisfaction and loyalty. Airline alliances are focusing more on shared delay reduction efforts. Additionally, the Department of Transportation is increasing pressure through stricter reporting rules and refund requirements for controllable delays. Some airlines are investing in digital tools and AI-based systems, while others still rely on outdated legacy infrastructure. This creates a competitive edge for carriers that can adopt innovation faster and more effectively.

Data Logic & Decision Frameworks (SWOT, TOWS, Value Chain)

Using the SWOT framework, strengths like data availability and post-COVID investments were balanced against weaknesses such as outdated infrastructure and limited predictive capabilities. Opportunities include real-time data feeds from FAA systems and the rise of AI-driven scheduling tools. However, threats from regulatory scrutiny and advanced competition from other airlines present real risks.

SWOT Analysis

Strengths:

- Large historic delay and weather datasets available
- Internal shift toward using analytics post-COVID 19
- Some investment already made in process improvement tools
- Senior leadership in some airlines supportive of digital transformation
- Data from FAA systems increasingly accessible

Weaknesses:

- Legacy dispatch, crew scheduling, and IT infrastructure
- Limited in-house AI and predictive modeling capabilities
- Ground teams understaffed at major hubs
- Resistance to operational change in unionized environments
- Data silos between departments reduce model effectiveness

Opportunities:

- FAA NextGen real-time weather and airspace data feeds

- Adoption of predictive delay tools across the industry
- Cost-effective SaaS platforms for real-time rerouting
- Chance to lead among mid-tier carriers lagging behind Delta/Alaska
- Potential partnerships with tech vendors for pilot programs

Threats:

- Regulatory pressure from DOT on controllable delays and refunds
- Legal risk if AI tools are not explainable or well-audited
- Faster digital adoption by competitors may widen performance gap
- Negative media coverage if delay problems persist without action
- Dependence on third-party tools could introduce vendor lock-in

TOWS Matrix

Opportunity + Strength Integrate FAA NextGen weather data to improve proactive scheduling	**Weakness + Opportunity** Partner with AI vendors while building internal capabilities
Threat + Strength Use existing data for explainable reporting to regulators	**Weakness + Threat** Shift to cloud-based systems to bypass legacy limitations

Value Chain Impact

Stage	Option A (AI Platform)	Option B (Hub Diversification)
Network Planning	Minimal impact	Major changes to flight network
Scheduling	Medium flexibility required	Minor change
Dispatch Ops	High – real-time adjustments	Low impact
Ground Operations	Moderate	High – new staffing needed
Customer Communication	High – auto-updates	Moderate

Strategic Alternatives and Evaluation

Option A: AI Weather Prediction & Rerouting Platform	**Option B: Hub Diversification**
Real-time crew, gate, and aircraft adjustments	Faster deployment (6–9 months)
Moderate cost, SaaS-based	Forecasts delays using weather and scheduling data
Moves flights from congested hubs to secondary airports	Requires network redesign and workforce relocation
Longer timeline (12–24 months)	High cost and disruption risk

Option A is more flexible and quicker to implement. It helps directly with weather-related and carrier-driven delays the two leading causes we identified. It does not require major changes to infrastructure and avoids the cost of workforce relocation. It also meets growing regulatory expectations to improve delay transparency and accountability. Option B has long-term potential but comes with more operational disruption and a slower timeline. Based on these differences, Option A is the more practical and cost-effective path for reducing delays in the short term while still improving service quality.

Final Recommendation with Rationale

We recommend Option A. It is more affordable than other options and can be rolled out faster, which matters when delays are rising quickly. It also matches current FAA and DOT expectations for explainable systems, making it a safer choice from a compliance standpoint. Option A gives airlines more control over real-time operations without needing to make large changes to their network or workforce. It avoids the cost and disruption that come with moving flights or shifting staff. Based on the trends we analyzed, this solution brings the best balance between cost, time, and results. It offers a practical path to reduce delays using data airlines already have, without adding unnecessary complexity.

Ethical and Implementation Considerations

Option A should be implemented with attention to several key factors. First, the data used must be accurate, timely, and complete. Without good data, the system will not be reliable. The model itself should be explainable so that internal teams and regulators understand how decisions are made. This is especially important in the airline industry, where public trust and legal oversight matter. Any changes to operations should consider how they affect workers, especially dispatch teams and gate staff. If tasks become automated, support and training should be given to staff whose roles might change. Cybersecurity is also a concern, especially when outside vendors are involved. Contracts should include clear terms about data safety and support. Before full rollout, the system should be tested on a small scale to find problems early. During this phase, airlines should communicate clearly with staff, customers, and partners to avoid confusion. With the right planning, Option A can be deployed in a way that is smooth, fair, and effective.

Limitations

The dataset did not include valid cancellation data, which meant we could not directly analyze how often or why flights were canceled. Instead, we used severe delays defined as those over 60 minutes, as a substitute to measure major disruptions. This approach gave us a general sense of where service reliability was weakest. We also could not see detailed airport-level issues like ground crew shortages or temporary runway closures, which may have influenced specific delays. While these gaps limit how deep we could go in some parts of the analysis, the overall trends and findings still hold. They offer a useful view of how delays vary by time, location, and airline.

Reflection on Course Learning

Our recommendation aligns with core lessons from the Tesla and Jabil case studies, as well as our own simulation experience. From Tesla, we learned that innovation without reliable execution can lead to customer dissatisfaction, production issues, and compliance problems. A system that works in theory but fails in practice does more harm than good. Jabil's case showed how important it is to make sure predictive models are not only accurate but also understandable by both internal teams and outside regulators. Their leadership made it clear that decision-makers need to trust and explain what the model is doing. Our simulation added another layer by showing how delays are rarely caused by a single factor. Fixing one thing in isolation doesn't help if the rest of the system can't adjust. We found that delay reduction only works when all parts of the process: crew, gates, weather, and routing are managed together. Based on these lessons, we chose Option A because it supports real-time operations, allows explainable decisions, and works across the entire system without needing big structural changes.