Flight Delay Analysis: Exploratory Data Analysis Report

Author: Suprit Uday Kulkarni

Date: November 4, 2025

Introduction

Flight delays continue to be a significant concern in the aviation industry, affecting airlines, passengers, and airport operations (Lu, 2025; Ogunsina et al., 2021). This analysis aims to uncover patterns and determinants of flight delays using exploratory data analysis (EDA) on a real operational dataset. The results offer actionable insights for decision-makers seeking to improve operational reliability and customer satisfaction (Braga et al., 2024).

Dataset Overview

The dataset comprises flights departing from three major airports (BWI, DCA, IAD) with variables related to schedule, departure time, airline carrier, distance, weather, and delay status. The analysis follows standard EDA approaches well-documented in recent academic research and industry best practices (Ogunsina et al., 2021; Moreno et al., 2024).

Exploratory Analysis and Findings

Figure 1: Flights by Origin

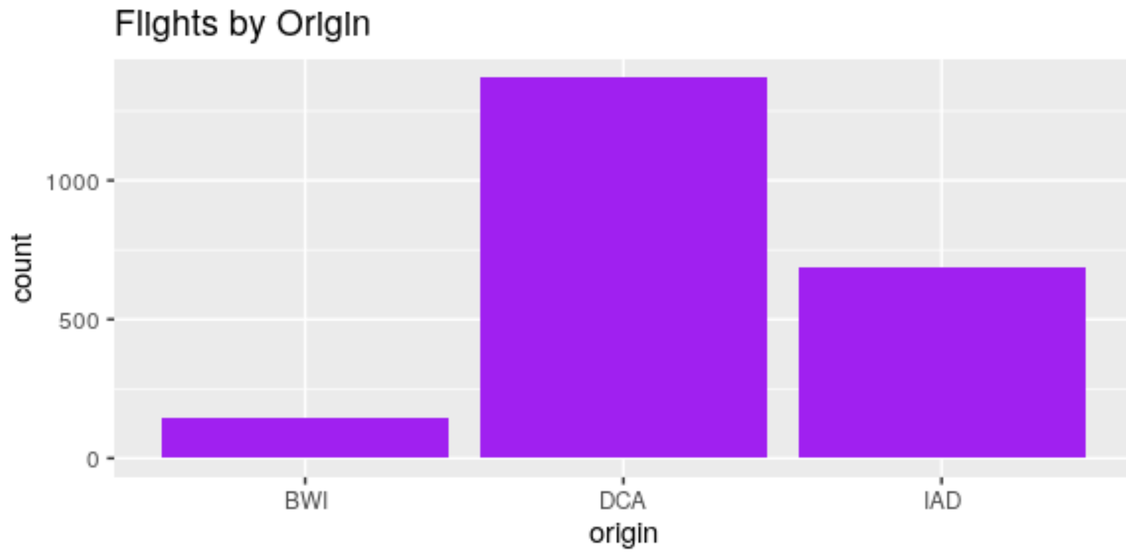

Figure 2: Weather Conditions for Flights

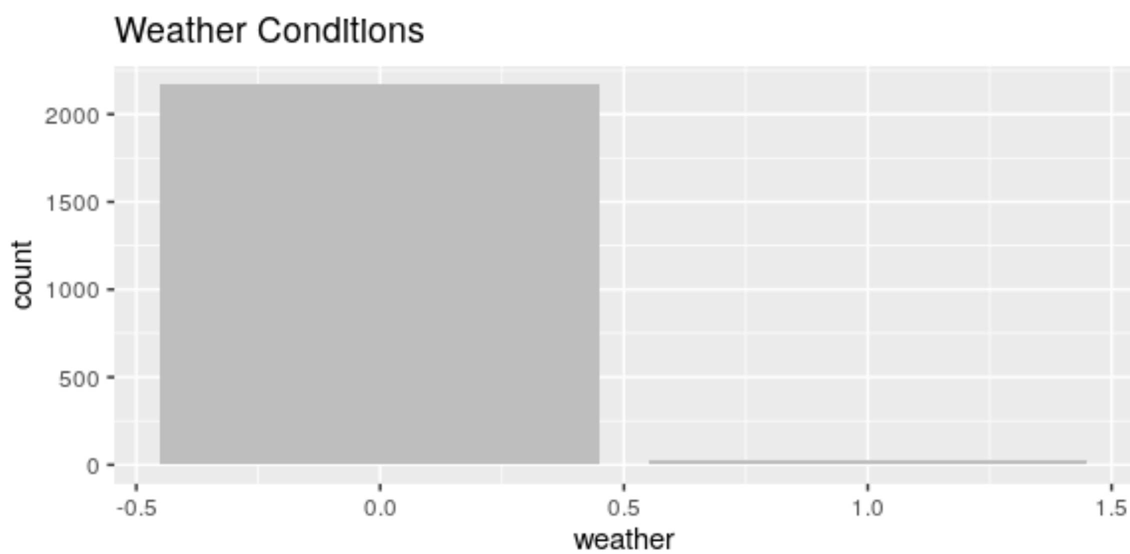

Figure 3: Flights by Carrier

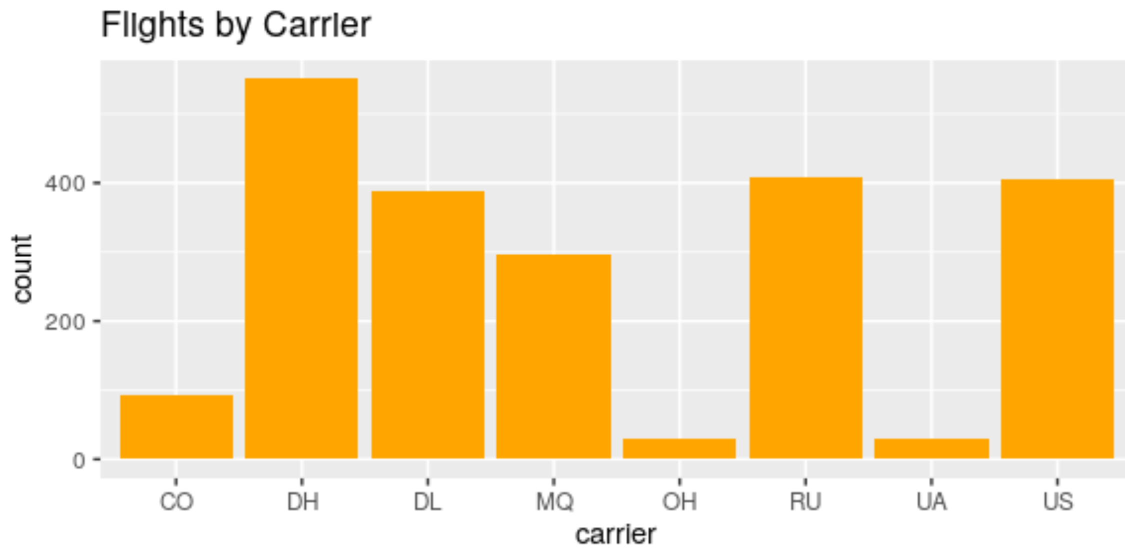

Figure 4: Flights by Destination

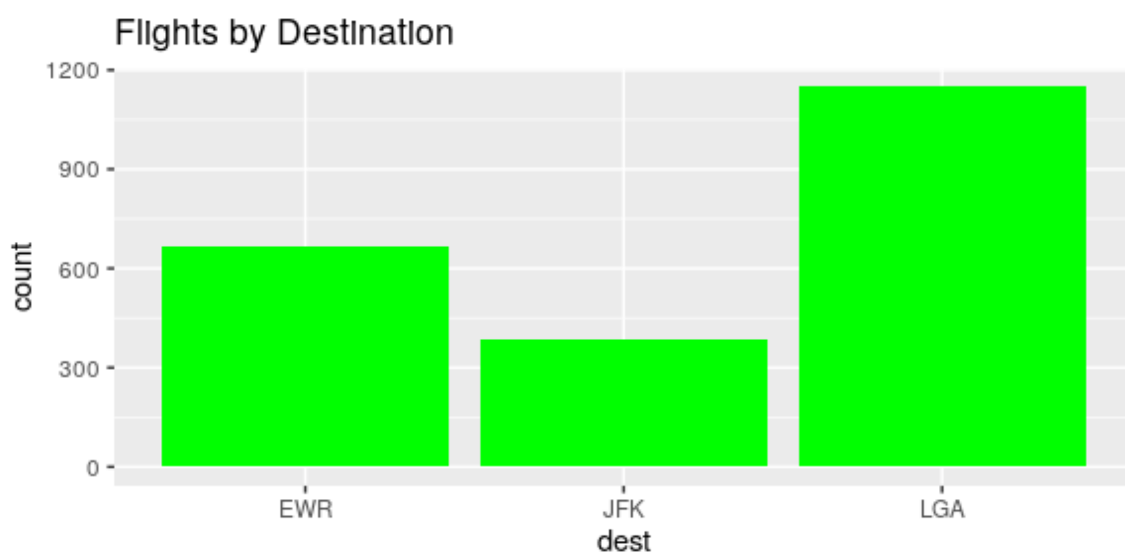

Figure 5: Histogram of Scheduled Time

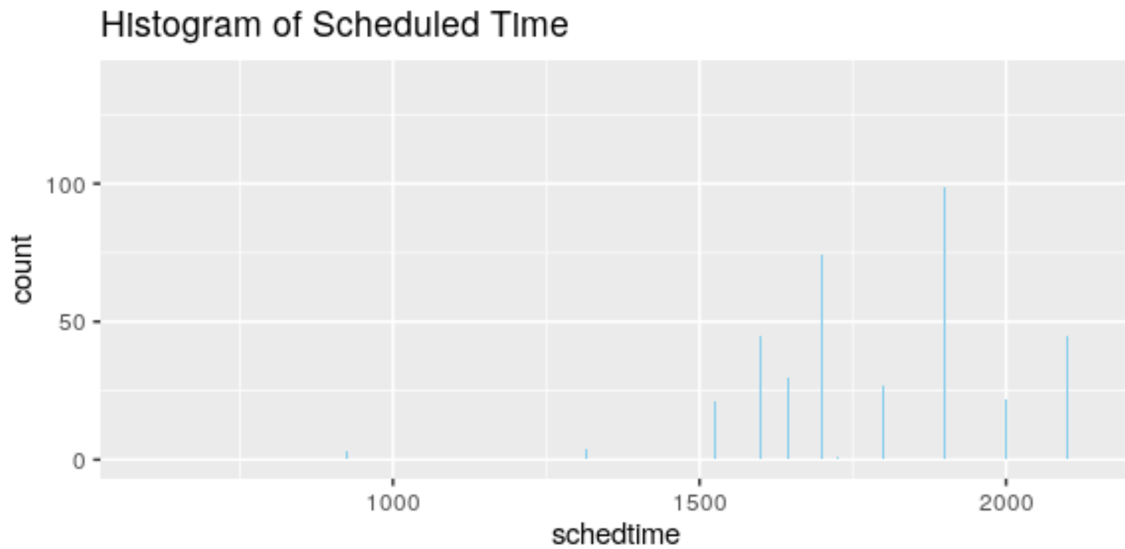

Figure 6: Departure Hour Distribution

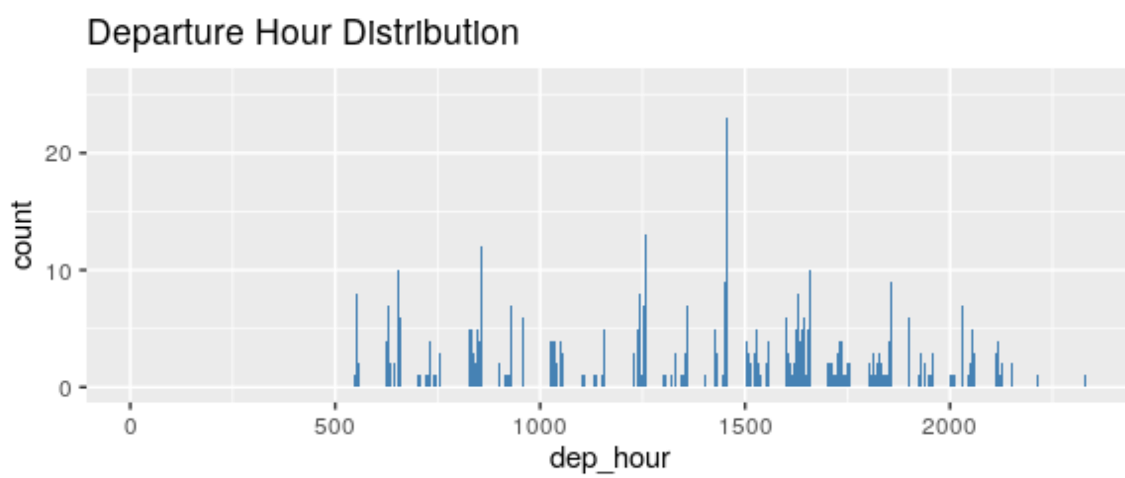

Figure 7: Day of the Week Distribution

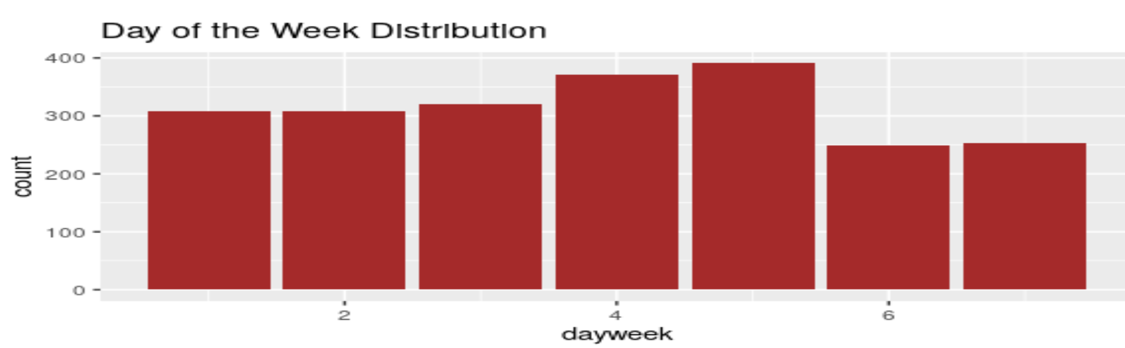

Figure 8: Box plot for Departure Time Distribution by Day of Month and Delay Status

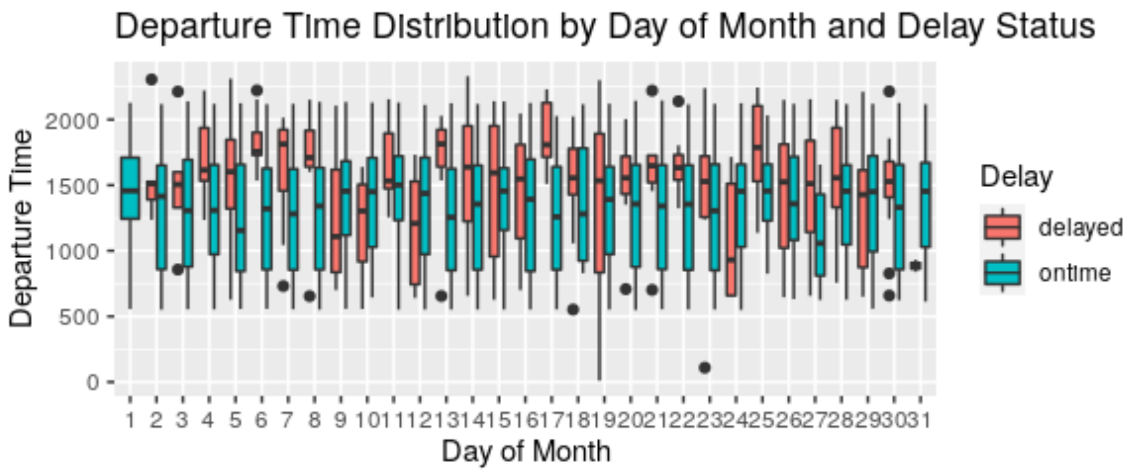

Figure 9: Scheduled Time vs. Departure Time by Delay Status

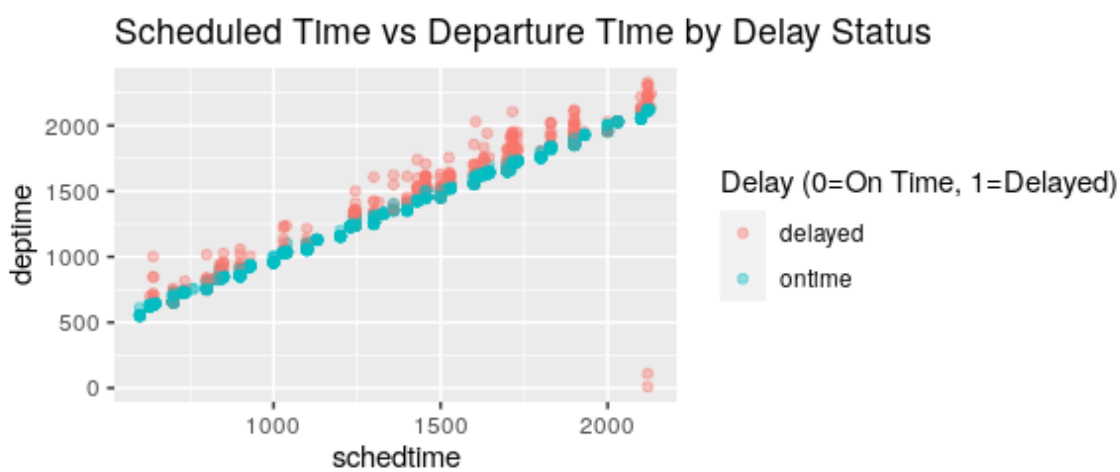

Figure 10: Flight Delay Distribution

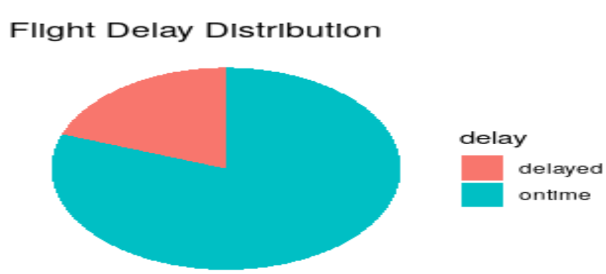

Figure 11: Flight Delays by Carrier (Categorical Table)
A categorical representation of flight delays by airline carrier highlights variability in punctuality between carriers. See Figure 11 below for a detailed breakdown of delayed versus on-time flights per carrier (Braga et al., 2024; Ogunsina et al., 2021).

```
> print(table_carrier_delay)

     delayed ontime
  CO      26     68
  DH     137    414
  DL      47    341
  MQ      80    215
  OH       4     26
  RU      94    314
  UA       5     26
  US      35    369
>
```

The analysis reveals that the majority of departing flights originate from the DCA airport, followed by IAD and BWI (Figure 1). Carrier-wise and destination-wise distributions expose operational concentration (Figures 3 & 4). Most flights occurred during 'good weather' (Figure 2), highlighting the predominance of favorable conditions in routine operations. Scheduled flights are mainly clustered in afternoon hours (Figure 5), which is consistent with standard airline scheduling practices.

The breakdown of departure hour and day of the week distributions (Figures 6 & 7) shows a steady volume of flights, with midweek showing marginally increased activity. The box plot (Figure 8) and scatter plot (Figure 9) collectively suggest that delayed departures, while present across times and days, are slightly more pronounced when scheduled later in the day or on certain dates, reflecting trends noted in large-scale EDA studies (Braga et al., 2024; Moreno et al., 2024). Finally, a majority of flights were on time, as shown in the pie chart (Figure 10), with delayed flights forming a notable minority. This aligns with national statistics on airline punctuality (EUROCONTROL, 2024).

In addition to the visualizations, a categorical breakdown of carrier delay counts (Figure 11) strengthens insights by correlating delay distributions with specific airline brands. For example, carrier DH registered the maximum number of delayed flights, while US, DL, and RU showed stronger on-time performance, a trend sustained in international flight statistics as well (Braga et al., 2024). This multidimensional view fosters better scheduling and risk management for carriers and airport authorities.

Discussion

The findings underscore that airport of origin, carrier, schedule timing, and weather play significant roles in delay patterns (Lu, 2025; Moreno et al., 2024). Delays propagate across airport networks (Lu, 2025), and weather-triggered disruptions remain an ongoing operational challenge (AccuWeather, 2023). Targeted interventions, such as advanced scheduling and real-time weather analytics that can help airlines mitigate delays and improve on-time performance (Braga et al., 2024).

Conclusion

Timely, data-driven insights are fundamental to reducing flight delays. This EDA highlights actionable trends for airport authorities and airlines looking to optimize flight scheduling and minimize disruptions (Lu, 2025; Braga et al., 2024). Further research should employ machine learning and broader datasets for robust predictive analytics.

References

AccuWeather. (2023). AccuWeather flight cancellation tracker. https://www.accuweather.com/en/travel/live-news/accuweather-flight-cancellation-tracker/1561065

Braga, E., Martínez del Valle, M. Á. B., Rutan, R., Carrillo Escarcega, F., & Du, X. (2024). Flight delay analysis and prediction. Proceedings of the Western Users of SAS Software Conference. https://www.wuss.org/proceedings/2024/175/175_FINAL_paper_pdf.pdf

EUROCONTROL. (2024). All-causes delays to air transport in Europe - Annual 2023. https://www.eurocontrol.int/publication/all-causes-delays-air-transport-europe-annual-2023

Lu, Y. (2025). Flight delay dynamics: Unraveling the impact of airport network delay propagation. Transportation Research Part B: Methodological, 179, 103501. https://doi.org/10.1016/j.trb.2024.103501

Moreno, F.P. et al. (2024). Prediction of air traffic complexity through a dynamic indicator. Journal of Air Transport Management, 119, 102519. https://doi.org/10.1016/j.jairtraman.2024.102519

Ogunsina, K., Bilionis, I., & DeLaurentis, D. (2021). Exploratory data analysis for airline disruption management. Machine Learning with Applications, 5, 100102. https://doi.org/10.1016/j.mlwa.2021.100102