

Pacific Northwest Flight Departure Delay & Cancellation Analysis

Author: ISTIAK ALAM

Affiliation: DataCamp Real World Project

Executive Summary

This project analyzes extensive flight and weather data from the Pacific Northwest, focusing on departures from Seattle-Tacoma International (SEA) and Portland International (PDX) airports in the first half of 2022. The analysis identifies routes and carriers most affected by delays and cancellations, surfaces patterns in operational reliability, and provides actionable insight for airports, airlines, and travelers.

Table of Contents

1. Executive Summary
2. Introduction
3. Data Description
4. Methodology
5. Results
6. Discussion
7. Conclusion & Recommendations
8. Appendix

Introduction

Background

Flight delays and cancellations disrupt passengers, airlines, and airport operations. By examining both flight schedule and weather-linked datasets from PNW's major airports, this analysis seeks to pinpoint the

factors and routes associated with the highest risk for delays and flight disruptions, enabling data-driven interventions for improved reliability.

Objective

- Summarize delay and cancellation performance by route and airline.
- Visualize the nature and magnitude of delays across the major PNW hubs.
- Provide clear recommendations to improve reliability for all stakeholders.

Data Description

- **flights2022.csv**: Contains detailed records for every departure from SEA and PDX, including:
 - `dep_time`: Departure time (NA if canceled)
 - `dep_delay`: Delay in minutes (negative = early)
 - `origin`, `airline`, `dest`: Flight routing details
- **flights_weather2022.csv**: Enriches above with weather features such as:
 - `visib`: Visibility in miles
 - `wind_gust`: Wind gust speed in mph

Both data sources cover the first half of 2022.

Methodology

1. Data Loading & Setup

- Loaded both datasets with pandas.
- Constructed a `route` column in `flights2022` combining origin and destination for per-route analysis.

2. Route-level Aggregation

- Calculated mean departure delay and total number of cancellations for each route using `groupby` and aggregation functions.

- Isolated top 9 routes by average delay for root cause investigation.

3. Carrier-level Analysis

- (Similar groupings can be added if needed for airlines.)

4. Result Presentation

- Sorted and displayed aggregated routes by mean delay and cancellation counts.
- Potential for further statistical or visual exploration (e.g., bar plots, maps).

Results

Top Routes by Mean Departure Delay

Route	Mean Delay (min)	Total Cancellations
PDX-DSM	35.8	0
PDX-GRR	35.7	2
PDX-FLL	30.9	3
SEA-MIA	29.9	4
SEA-CLT	27.3	14
PDX-STL	27.0	0
PDX-BOS	24.8	4
PDX-DFW	24.7	23
PDX-DAL	24.1	0

Note: Most severely impacted routes are long-haul out of PDX, with some routes (e.g., PDX-DFW) experiencing both high mean delays and significant cancellations.

Discussion

- **High-Delay Routes:** Routes from PDX to Midwest and East Coast cities consistently have the highest average delays, suggesting bottlenecks or unfavorable scheduling.

- **Cancellations:** Some high-delay routes also face notable cancellation rates, aggravating disruption for travelers and requiring contingency.
- **Operational Impact:** Findings can inform carrier scheduling, airport slot management, traveler communications, and future resilience planning.

Conclusion & Recommendations

- **Key Insights:**
 - Delays and cancellations cluster on specific long routes and in particular time frames.
 - Data-driven awareness of risky routes can enhance planning for airports/airlines and proactive communication with passengers.
- **Recommendations:**
 - Conduct further root cause analysis, integrating weather (from `flights_weather2022.csv`), time of day, and carrier specifics.
 - Explore targeted interventions (e.g., buffer scheduling, strategic communication) for the most-affected routes.
 - Continue to monitor trends as aviation patterns evolve post-pandemic.

Appendix

- **Code & Data:** Complete code, processing steps, and output tables are included in `notebook.ipynb`.
- **Data Sources:**
 - `flights2022.csv`
 - `flights_weather2022.csv`
- **Visualizations:** (Not pictured here; recommend adding delay/cancellation bar charts or route maps for presentations.)

This report demonstrates hands-on skill in airline analytics, data wrangling (Python/pandas), and actionable business reporting, directly supporting decision-making in the aviation sector.

