

Airline Delay and Cancellation Analysis

Himanshu Gandhi (1770138) Venkata Shreya Kala (1764875) Shubham Sahoo (1824661)

In the fast-paced world of air travel, it's crucial for airlines to handle delays and cancellations. The project explores a complex dataset ([Source](#)) with 28 features and 7.21M instances describing diverse aspects of flight information crucial in determining the performance and reliability of air transportation.

The dataset under consideration comprises diverse features that capture crucial aspects of airline operations and performance. Time-related variables, such as scheduled departure and arrival times, actual departure and arrival times, and associated delays, offer an overall understanding of temporal dynamics in flight schedules. Specific information related to aircraft, such as carrier details, flight numbers, and diversion indicators, helps identify distinctive patterns for different carriers and instances of unexpected route changes. Metrics indicating operational efficiency, like taxi times and elapsed times, provide insights into the effectiveness of ground operations and the total duration of flights. Moreover, the dataset details geospatial information, including the origin and destination airports and other influential factors such as the regional weather conditions. The classifications for delays, for example, carrier delay, weather delay, and late aircraft delay offer a detailed breakdown of delay factors which enable a detailed analysis of the underlying causes behind disruptions in airline schedules.

Research Question and Objectives:

The primary aim of this research initiative is to delve into the intricate dynamics of historical airline data, unravel the underlying patterns, and identify the multifaceted factors that contribute to delays and cancellations in air travel.

By thoroughly studying the dataset, we aim to gain insights into the relationships among variables, covering both external factors like weather conditions and internal ones such as airline operations.

In this study, we will employ advanced statistical and machine learning techniques, with a primary focus on time series analysis and predictive modelling. Time series analysis will be a pivotal tool in uncovering temporal patterns and trends within our historical airline dataset, enabling us to discern recurring patterns in delays and cancellations. Additionally, predictive modelling will play a central role in our methodology. By establishing relationships between the diverse set of features and key outcome variables, such as delays and cancellations, our models aim to quantify the impact of various factors. Through this approach, we aspire to develop accurate predictive models that offer valuable insights for airlines and passengers alike.