

# Datasheet-Flight Delays

Zixi Song

2024-03-31

## Datasheet for the Dataset of 2015 Flight Delays and Cancellations

### Motivation

**Purpose of Dataset Creation:** The 2015 Flight Delays and Cancellations dataset was compiled with the aim of providing insight into the on-time performance of domestic flights operated by large air carriers in the United States. The dataset serves as a resource for analyzing trends in flight delays and cancellations, aiding travelers in making informed decisions and supporting the aviation industry in improving service reliability.

**Creators:** The dataset was collected and published by the U.S. Department of Transportation's Bureau of Transportation Statistics as part of their ongoing efforts to monitor and report on the quality of air travel.

### Composition

**Data Instances:** The dataset includes detailed records for domestic flights in 2015, featuring data on delays, cancellations, and operational aspects of the flights. Attributes include delay time, terminal, airline, capacity, temperature, rain presence, wind gust speed, and wind direction.

**Number of Instances:** The dataset encompasses a comprehensive collection of flight records throughout 2015, with each record corresponding to a unique flight instance.

**Sampling:** Data represents an exhaustive collection of reported flights, capturing a wide array of conditions and outcomes to accurately reflect the year's flight operations.

**Data Type:** Quantitative measurements (e.g., delay in minutes, temperature, wind gust speed) and categorical data (e.g., terminal, airline, rain presence).

**Labels:** Records are identifiable by attributes such as date, flight number, airline, and terminal, ensuring detailed traceability.

**Data Splits:** Not applicable, as the dataset provides a holistic view of the year's flight operations without division into training or testing sets.

### Collection Process

**Data Collection:** Information was aggregated from official flight operation reports, ensuring accuracy and reliability. The collection process adhered to the DOT's standards for data reporting and integrity.

### Preprocessing/Cleaning/Labeling

**Preprocessing Done:** Data underwent rigorous cleaning and structuring, aligning with the International Classification of Diseases for categorizing delays and ensuring uniformity across records.

**Raw Data Availability:** While the processed dataset is openly accessible, raw data may be obtained through official requests to the DOT.

### Uses

**Current Uses:** The dataset is utilized by travelers, researchers, and industry analysts for understanding and predicting flight delay patterns, enhancing travel planning, and facilitating improvements in airline operational efficiency.

**Potential Uses:** Future applications may extend to machine learning models for delay prediction, optimization studies for airport operations, and academic research in aviation management and policy development.

**Restrictions on Use:** Distributed under CC0: Public Domain License, allowing unrestricted use, distribution, and modification.

## **Distribution**

**Distribution Channels:** Available for download through the Bureau of Transportation Statistics website and related data platforms.

**Licence:** CC0: Public Domain.

## **Maintenance**

**Maintainers:** Oversight is provided by the Bureau of Transportation Statistics, ensuring the dataset's relevance and accuracy.

**Update Schedule:** The dataset specifically covers the year 2015 and does not specify an update frequency.

## **Ethical Considerations**

**Data Privacy:** Aggregated data ensures individual flight details are not personally identifiable, prioritizing privacy while offering valuable insights.

**Ethical Oversight:** Compiled with consideration for ethical standards in data reporting and public dissemination, focusing on the collective benefit of improved air travel understanding.

## **Disclaimer**

**Limitations:** The dataset's scope is confined to flights in 2015, and interpretations should consider temporal and geographical context.

**Errata and Updates:** Users are encouraged to consult the Bureau of Transportation Statistics for the latest data and corrections.

This datasheet provides a comprehensive overview of the 2015 Flight Delays and Cancellations dataset, underlining its significance, composition, and practical applications. It aims to foster informed utilization of the data, supporting endeavors to advance the understanding and management of flight operations.