Codebook for Flight Dataset

# Data Overview

## Credentials

This data set can be downloaded from Kaggle data sets.

The direct link to data is: <https://www.kaggle.com/giovamata/airlinedelaycauses/data>

## Business goal

This data was collected to answer the question:

What are the main reasons a flight arrives prior or in delay to its scheduled arrival time.

## Data description

The data set is a data frame of 30 variables over 1,936,758 rows. Each row tracks the on-time performance of domestic flights operated by large air carriers in the USA.

# Variables description

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Name** | **Description** | **Type** | **Possible values** |
| **Year** | The year of the flight | Integer | 2008 |
| **Month** | The month of the flight | Integer | 1-12, (1 represents January) |
| **DayofMonth** | The day of month of the flight | Integer | 1-31, max is 31 |
| **DayOfWeek** | The weekday of the flight | Integer | 1-7, (1 represents Sunday) |
| **DepTime** | Actual departure time | Double | 1-2400, (1 represents 00:01 in local time) |
| **CRSDepTime** | Scheduled departure time | Integer | 0-2359, (0 represents 00:00 in local time) |
| **ArrTime** | Actual arrival time | Double | 1-2400, (1 represents 00:01 in local time) There are 7110 records missing data |
| |  | | --- | | **CRSArrTime** | | Scheduled arrival time | Integer | 0-2400, (0 represents 00:00 in local time) |
| **UniqueCarrier** | Air Carrier codes | Character | "WN","XE","YV" ect |
| **FlightNum** | flight number | Integer | 1-9742 |
| **TailNum** | plane tail number | Character | "N798SW","N211WN"  ect |
| **ActualElapsedTime** | Actual elapsed time in minutes | Double | 14-1114, There are 8387 records missing data |
| **CRSElapsedTime** | Scheduled elapsed time in minutes | Double | -25-660, There are 198 records missing data |
| **AirTime** | Air time in minutes | Double | 0-1091, There are 8387 records missing data |
| **ArrDelay** | Arrival delay in minutes | Double | -109-2461, There are 8387 records missing data |
| **DepDelay** | Departure delay in minutes | Double | 6-2467 |
| **Origin** | Origin airport code (IATA) | Character | "IDA","TPA" ect |
| **Dest** | Destination airport code (IATA) | Character | "IDA","TPA" ect |
| **Distance** | The distance of flight in miles | Integer | 11-4962 |
| **TaxiIn** | taxi in time, in minutes | Double | 0-240 There are 7110 records missing data |
| **TaxiOut** | taxi out time, in minutes | Double | 0-422 There are 455 records missing data |
| **Cancelled** | was the flight cancelled? | Integer | 0(No)-1(Yes) |
| **CancellationCode** | reason for cancellation (A = carrier, B = weather, C = NAS, D = security) | Character | N(No) ,  A (yes-carrier),  B (yes-weather),  C (yes-NAS)  D = security |
| **Diverted** | Diverted flight | Integer | 0(No)-1(Yes) |
| **CarrierDelay** | The time delay in minutes because Carrier | Double | 0-2436 There are 689270records missing data |
| **WeatherDelay** | The time delay in minutes because weather | Double | 0- 1352 There are 689270  records missing data |
| **NASDelay** | The time delay in minutes because NAS | Double | 0- 1357  There are 689270  records missing data |
| **SecurityDelay** | The time delay in minutes because security | Double | 0-392 There are 689270  records missing data |
| **LateAircraftDelay** | The time delay in minutes because aircraft | Double | 0-1316 There are 689270  records missing data |