The goal of this project is to predict flight delays based on flight information delay information gathered from the Kaggle web site: <https://www.kaggle.com/usdot/flight-delays/version/1>

In addition to this data, airport operations (i.e., how busy an airport is) is to be incorporated into the model per departure and destination airport and region.

One significant ‘next step’ would be to add weather information to the model.

**Context**

The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics tracks the on-time performance of domestic flights operated by large air carriers. Summary information on the number of on-time, delayed, canceled, and diverted flights is published in DOT's monthly Air Travel Consumer Report and in this dataset of 2015 flight delays and cancellations.

**Acknowledgements**

The flight delay and cancellation data were collected and published by the DOT's Bureau of Transportation Statistics.

1. Data Files

airlines.csv

this is a small file that contains 14 rows of data.

The airline IATA code (global airport unique identifier) and the airline name are the only two columns.

An IATA airport code, also known as an IATA location identifier, IATA station code or simply a location identifier, is a three-letter code designating many airports around the world, defined by the International Air Transport Association (IATA).

airports.csv

This file contains 322 rows of airport information.

The airport IATA code, airport name, city, state, country, latitude and longitude are in each row.

flights.csv

this file is a dataset of 2015 flight delays and cancellations

5,819,081 rows of data (includes column row and last row is empty)

31 columns described below

5819079 rows × 31 columns (per pandas)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column name | Column description | Sample data | Example format or type or units | Questions |
|  |  |  |  |  |
| YEAR | Year of the flight trip | 2015 | YYYY |  |
| MONTH | Month of the flight trip | 1 | M {d, dd} |  |
| DAY | Day of the flight trip | 1 | D {d, dd} |  |
| DAY\_OF\_WEEK | Day of week of the flight trip | 4 | {1 – 7} |  |
| AIRLINE | Airline identifier | AS |  |  |
| FLIGHT\_NUMBER | Flight identifier | 98 |  |  |
| TAIL\_NUMBER | Aircraft identifier | N407AS |  |  |
| ORIGIN\_AIRPORT | Starting airport | ANC |  |  |
| DESTINATION\_AIRPORT | Destination airport | SEA |  |  |
| SCHEDULED\_DEPARTURE | Planned departure time | 0005 |  |  |
| DEPARTURE\_TIME | WHEEL\_OFF – TAXI\_OUT | 2354 |  |  |
| DEPARTURE\_DELAY | Total delay on departure | -11 |  |  |
| TAXI\_OUT | The time duration elapsed between departure from the origin airport gate and wheels off | 21 |  |  |
| WHEELS\_OFF | The time point that the aircraft wheels leave the ground | 0015 |  |  |
| SCHEDULED\_TIME | Planned time amount needed for the flight trip | 205 |  |  |
| ELAPSED\_TIME | AIR\_TIME + TAXI\_IN + TAXI\_OUT | 194 |  |  |
| AIR\_TIME | The time duration between wheels\_off and wheels\_on\_time | 169 |  |  |
| DISTANCE | Distance between the two airports | 1448 | miles |  |
| WHEELS\_ON | The time point that the aircraft’s wheels touch on the ground | 0404 |  |  |
| TAXI\_IN | The time duration elapsed between wheels-on and gate arrival at the destination airport | 4 |  |  |
| SCHEDULED\_ARRIVAL | Planned arrival time | 0430 | HHMM | UTC or Local? |
| ARRIVAL\_TIME | WHEELS\_ON + TAXI\_IN | 0408 | HHMM |  |
| ARRIVAL\_DELAY | ARRIVAL\_TIME – SCHEDULED\_ARRIVAL | -22 | Integer |  |
| DIVERTED | Aircraft landed on airport that out of schedule | 0 | Boolean {0, 1} |  |
| CANCELED | Flight canceled (1 = cancelled) | 0 | Boolean {0, 1} |  |
| CANCELLATION\_REASON | Reason for cancellation of flight: A – airline/carrier B – weather C – national air system D – security | A | Single character {A, B, C, D} |  |
| AIR\_SYSTEM\_DELAY | Delay caused by air system | 25 | Integer |  |
| SECURITY\_DELAY | Delay caused by security | 20 | Integer |  |
| AIRLINE\_DELAY | Delay caused by the airline | 0 | Integer |  |
| LATE\_AIRCRAFT\_DELAY | Delay caused by aircraft | 0 | Integer |  |
| WEATHER\_DELAY | Delay caused by weather | 0 | Integer |  |