

# Analysis of domestic US flights

## Project – Description and Demo

### SOURCE OF THE DATA

The data used for this project corresponds to on-time performance of domestic US flights departing from San Francisco, CA from October 2014 to September 2015. This data was obtained from the Bureau of Transportation Statistics

[http://www.transtats.bts.gov/DL\\_SelectFields.asp?Table\\_ID=236](http://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236)

The screenshot shows the TranStats website interface. The header includes the United States Department of Transportation logo and navigation links. The main content area is titled "On-Time Performance" and features a search bar, download instructions, and a table of available fields for selection. The table lists fields such as Time Period, Airline, and Carrier, each with a description and a "Get Lookup Table" link. The interface also includes filters for Geography, Year, and Period, and a download button.

Field Name	Description	Support Table
<b>Time Period</b>		
<input type="checkbox"/> Year	Year	
<input type="checkbox"/> Quarter	Quarter (1-4)	<a href="#">Get Lookup Table</a>
<input type="checkbox"/> Month	Month	<a href="#">Get Lookup Table</a>
<input type="checkbox"/> DayofMonth	Day of Month	
<input type="checkbox"/> DayOfWeek	Day of Week	<a href="#">Get Lookup Table</a>
<input type="checkbox"/> FlightDate	Flight Date (yyyymmdd)	
<b>Airline</b>		
<input type="checkbox"/> UniqueCarrier	Unique Carrier Code. When the same code has been used by multiple carriers, a numeric suffix is used for earlier users, for example, PA, PA(1), PA(2). Use this field for analysis across a range of years.	<a href="#">Get Lookup Table</a>
<input type="checkbox"/> AirlineID	An identification number assigned by US DOT to identify a unique airline (carrier). A unique airline (carrier) is defined as one holding and reporting under the same DOT certificate regardless of its Code, Name, or holding company/corporation.	<a href="#">Get Lookup Table</a>
<input type="checkbox"/> Carrier	Code assigned by IATA and	<a href="#">Get Lookup Table</a>

One CSV file per month was downloaded with the fields:

- YEAR
- MONTH
- DAY\_OF\_WEEK
- FL\_DATE
- UNIQUE\_CARRIER
- FL\_NUM
- ORIGIN\_CITY\_NAME
- ORIGIN\_STATE\_ABR
- DEST\_CITY\_NAME
- DEST\_STATE\_ABR
- CRS\_DEP\_TIME
- DEP\_DELAY\_NEW

- |                     |                       |
|---------------------|-----------------------|
| - ARR_DELAY_NEW     | - CARRIER_DELAY       |
| - CANCELLED         | - WEATHER_DELAY       |
| - CANCELLATION_CODE | - NAS_DELAY           |
| - DIVERTED          | - SECURITY_DELAY      |
| - AIR_TIME          | - LATE_AIRCRAFT_DELAY |
| - DISTANCE          |                       |

These 12 CSVs were merged in the terminal into one single CSV called “merged.csv”

Additionally, another CSV called “L\_UNIQUE\_CARRIERS” was downloaded to convert the code of the airlines to their names.

### DATABASE TABLE STRUCTURE

A table called “flights” was created in the database with the following fields and formats:

- f\_number INT
- carrier TEXT
- dates JSON (including 'YEAR', 'MONTH', 'DAY\_OF\_WEEK', 'FL\_DATE', 'CRS\_DEP\_TIME')
- route JSON (including 'ORIGIN\_CITY\_NAME', 'DEST\_CITY\_NAME', 'AIR\_TIME', 'DISTANCE')
- cancel JSON (including 'CANCELLED', 'CANCELLATION\_CODE', 'DIVERTED')
- dep\_delay FLOAT
- arr\_delay FLOAT
- delay\_cause TEXT

The table contains information of the 160,000+ domestic flights that have departed from San Francisco in the last year.

### SERVER TEST / USEFUL INFORMATION RETRIEVED

The information obtained from the web server helps compare different airlines and provide us with useful information to choose the best airline to fly. Note that, for security reasons, the server does not remain active but you can reproduce it with the accompanying code (just substitute the IP in the below addresses).

#### 1. Average delay (in minutes) per carrier:

We can first compare airlines by average arrival delay

[http://54.67.16.179:5000/avg\\_delay/carrier](http://54.67.16.179:5000/avg_delay/carrier)

Looks like companies like Airtran, US Airways or Delta are the most punctual overall.

#### 2. Average delay (in minutes) per destination:

Now by destination:

[http://54.67.16.179:5000/avg\\_delay/destination](http://54.67.16.179:5000/avg_delay/destination)

Almost 35 minutes average delay when flying to Madison, WI!

#### 3. Average delay (in minutes) per day of week

Which day of the week is better to fly?

[http://54.67.16.179:5000/avg\\_delay/dayofweek](http://54.67.16.179:5000/avg_delay/dayofweek)

Saturday is the best day to fly (to arrive on time).

**4. Average delay (in minutes) per departure time**

What time of the day is best to catch a flight?

[http://54.67.16.179:5000/avg\\_delay/departure-hour](http://54.67.16.179:5000/avg_delay/departure-hour)

Better to fly early in the morning

**5. Rank airlines per percentage of flights cancelled**

Which airline is more likely to cancel a flight?

[http://54.67.16.179:5000/perc\\_cancelled](http://54.67.16.179:5000/perc_cancelled)

Seems that Skywest cancels more than 3% of their flights.

**6. Rank airlines per percentage of flights delayed**

What's the percentage of flights with a delay of 10 minutes or more and the delay cause is carrier-related (either carrier delay or late-aircraft delay)?

[http://54.67.16.179:5000/perc\\_delayed\\_carrier](http://54.67.16.179:5000/perc_delayed_carrier)

Interesting, Hawaiian Airlines ends up having delays one third of the times.

**7. Comparing airlines by destination (destination provided by user)**

Now comes the interesting part, let's say we pick a destination and we want to obtain detailed information about flights to this particular destination and compare all the airlines flying there. This may help us decide which company to fly with.

For instance, let's say we decide to fly to Honolulu, HI, we type

<http://54.67.16.179:5000/destination/Honolulu,%20HI>

and we obtain one dict per airline, containing:

1. Total\_flights
2. Average\_delay\_minutes
3. Maximum\_delay\_hours
4. Percentage\_delayed\_by\_carrier
5. Percentage\_cancelled
6. Percentage\_diverted

In the particular case of Honolulu, we obtain:

```
{
  "Delta Air Lines Inc.": {
    "1.Total_flights": 171,
    "2.Average_delay_minutes": 8.4,
    "3.Maximum_delay_hours": 1.9,
    "4.Percentage_delayed_by_carrier": "11.11%",
    "5.Percentage_cancelled": "0.0%",
    "6.Percentage_diverted": "0.0%"
  },
  "Hawaiian Airlines Inc.": {
    "1.Total_flights": 365,
    "2.Average_delay_minutes": 22.99,
    "3.Maximum_delay_hours": 15.9,
    "4.Percentage_delayed_by_carrier": "36.99%",
    "5.Percentage_cancelled": "0.0%",
    "6.Percentage_diverted": "0.27%"
  },
  "United Air Lines Inc.": {
    "1.Total_flights": 1579,
    "2.Average_delay_minutes": 20.69,
    "3.Maximum_delay_hours": 8.6,
    "4.Percentage_delayed_by_carrier": "24.13%",
    "5.Percentage_cancelled": "1.52%",
    "6.Percentage_diverted": "0.13%"
  }
}
```

}  
}

And we can see, for instance, that Delta Airlines has much better numbers than the other two carriers flying to Honolulu. The maximum delay was 15 hours in Hawaiian Airlines, 8 hours in United, and only 1.9 hours in Delta. We might prefer to fly with Delta and have less probability that the flight spoils our vacation.

#### **8. Get percentage of flights “problematic” by cause (destination provided by user)**

To finish up, we may want to identify what percentage of the flights to the destination we choose are problematic, which is defined as either more than 10 minutes delay, or cancelled or diverted.

You can type any destination. For example, let’s see what happens in Honolulu and Madison:

[http://54.67.16.179:5000/problem\\_by\\_destination/Honolulu,%20HI](http://54.67.16.179:5000/problem_by_destination/Honolulu,%20HI)

[http://54.67.16.179:5000/problem\\_by\\_destination/Madison,%20WI](http://54.67.16.179:5000/problem_by_destination/Madison,%20WI)

Delays in Honolulu are mostly due to the carrier. However, Madison gets delays due to the National Aviation System. So the conclusion is choose a better airline in case of Honolulu, while try to avoid Madison as a destination if you can.