

Data Intake Report

Name: Airline Delay and Cancellation Data

Report Date: 12.05.2023

Internship Batch: LISUM20

Version: 1.0

Data intake by: Guru Guha Narayanan Muthunarayanan

Data intake reviewer:

Data storage location:

Data Details:

1. Airline delay and cancellation data 2015-2018

Total Number of Observations	18707146
Total Number of Features	27
Base Format of File	csv
Size of Data	2.13 GB

2. Airline delay and cancellation data 2014

Total Number of Observations	5819811
Total Number of Features	27
Base Format of File	csv
Size of Data	680 MB

The datasets include a collection of flight data spanning multiple years from 2014 to 2018. It provides valuable insights into flight delays and cancellations over an extended period. The dataset contains the following columns.

1. FL_DATE: The date of the flight.
2. OP_CARRIER: The code for the operating carrier or airline.
3. OP_CARRIER_FL_NUM: The flight number assigned by the operating carrier.
4. ORIGIN: The code for the origin airport.
5. DEST: The code for the destination airport.
6. CRS_DEP_TIME: The scheduled departure time of the flight.
7. DEP_TIME: The actual departure time of the flight.
8. DEP_DELAY: The departure delay in minutes.
9. TAXI_OUT: The time duration from departure from the origin gate to wheels off.

10. WHEELS_OFF: The time when the aircraft's wheels left the ground.
11. WHEELS_ON: The time when the aircraft's wheels touched the ground at the destination.
12. TAXI_IN: The time duration from wheels-on to arrival at the destination gate.
13. CRS_ARR_TIME: The scheduled arrival time at the destination.
14. ARR_TIME: The actual arrival time at the destination.
15. ARR_DELAY: The arrival delay in minutes.
16. CANCELLED: Indicates whether the flight was cancelled (1) or not (0).
17. CANCELLATION_CODE: The code indicating the reason for flight cancellation.
18. DIVERTED: Indicates whether the flight was diverted (1) or not (0).
19. CRS_ELAPSED_TIME: The scheduled total flight duration in minutes.
20. ACTUAL_ELAPSED_TIME: The actual total flight duration in minutes.
21. AIR_TIME: The total time spent in the air during the flight.
22. DISTANCE: The distance travelled by the flight in miles.
23. CARRIER_DELAY: The delay caused by the carrier in minutes.
24. WEATHER_DELAY: The delay caused by weather conditions in minutes.
25. NAS_DELAY: The delay caused by the National Airspace System (NAS) in minutes.
26. SECURITY_DELAY: The delay caused by security issues in minutes.
27. LATE_AIRCRAFT_DELAY: The delay caused by late aircraft arrival in minutes.

These columns provide detailed information about flight dates, carriers, origin and destination airports, scheduled and actual departure/arrival times, delays, cancellations, diversions, flight durations, and reasons for delays or cancellations. Analysing this dataset can provide valuable insights into flight operations and performance over the years.