

Airline Statistics Data Description

Please refer to the referenced Bureau of Transportation Statistics for full details on underlying data. The csv file given is a summary report pulled from the referenced web site. It reports to you the causes of delays, the minutes of delay for each cause and the number of flights delayed (prorated when multiple causes are assigned) by cause.

Briefly, here are the data fields described:

arr_flights is total number of flights arriving

arr_del15 is the total number of flights delayed more than 15 minutes (official definition of a delay is at least 15 minutes)

carrier_ct, weather_ct, nas_ct, security_ct, late_aircraft_ct, are the number of flights delayed for each of these causes (see the website for descriptions of each cause). These can have decimal values because they prorate the cause when there is more than one. In other words, if a flight is delayed for carrier and NAS reasons, each is given a 0.5 value (See the note under the pie chart on the referenced web page from the assignment description). Notice that these values sum up to the value of arr_del15

arr_cancelled, arr_diverted, number of flights cancelled or diverted respectively (does not count as a delayed flight).

arr_delay, total minutes delayed for all delays

carrier_delay, weather_delay, nas_delay, security_delay, late_aircraft_delay, number of minutes of delay due to each of these causes.

Hope this helps. When in doubt, read the documentation.