Flight delays

Background

You have been hired by Newark airport to investigate the effect of weather on aeroplane departure delays.

They believe poor weather conditions are causing too many delays and want to invest in improving facilities, so that aircraft can take off in more types of weather. However, they do not fully understand how serious weather related delays are, and are not sure what type of weather they should be most concerned about. As part of investigating the effect of weather you should investigate other factors to understand how important weather is in comparison to them

They also want to understand how they compare to other New York airports.

Data

The data is available across several CSV files. The most important is `flights.csv`. It has the following variables:

- year, month, day: Date of departure
- dep_time,arr_time: Actual departure and arrival times, local tz.
- sched_dep_time,sched_arr_time: Scheduled departure and arrival times, local tz.
- dep_delay,arr_delay: Departure and arrival delays, in minutes. Negative times represent early departures/arrivals.
- hour, minute: Time of scheduled departure broken into hour and minutes.
- carrier: Two letter carrier abbreviation. See airlines to get name
- tailnum: Plane tail number
- flight: Flight number
- origin, dest: Origin and destination. See airports for additional metadata.
- air_time: Amount of time spent in the air, in minutes
- distance: Distance between airports, in miles
- time_hour: Scheduled date and hour of the flight as a POSIXct date. Along with origin, can be used to join flights data to weather data.

We also have metadata about airports in `airports.csv`.

- faa: FAA airport code
- name: Usual name of the airport
- lat,lon: Location of airport
- alt: Altitude, in feet
- tz: Timezone offset from GMT
- dst: Daylight savings time zone. A = Standard US DST: starts on the second Sunday of March, ends on the first Sunday of November. U = unknown. N = no dst.
- tzone: IANA time zone, as determined by GeoNames webservice

Airline names are normally given as codes. The dataset `airlines.csv` translates the codes to names.

- carrier: Two letter abbreviation
- name: Full name

For many of the planes we have additional meta data in `planes.csv`.

- tailnum: Tail number
- year: Year manufactured
- type: Type of plane
- manufacturer, model: Manufacturer and model
- engines, seats: Number of engines and seats
- speed: Average cruising speed in mph
- engine: Type of engine

And finally we have weather data for all three NYC airports in `weather.csv`.

- origin: Weather station. Named origin to facilitate merging with flights data
- year, month, day, hour: Time of recording
- temp, dewp: Temperature and dewpoint in F
- humid: Relative humidity
- wind_dir,wind_speed,wind_gust: Wind direction (in degrees), speed and gust speed (in mph)
- precip: Precipitation, in inches
- pressure: Sea level pressure in millibars
- visib: Visibility in miles
- time_hour: Date and hour of the recording as a POSIXct date