Group Assignment Business Reporting Tools

NYC Flights data

Assignment Description

You are a group of general airline analysts, and you want to investigate different aspects of delays from the different New York City airports (the data stem from 2013, but can be enriched with newer data). To do this, you get database comprising flight information, airport information, plane information, carrier information and weather information. Your goal is to make use of SQL for any preprocessing of the dataset, and use Tableau for creating visual insights. (Thus, avoid for instance creating new columns in Tableau, merging, …).

You can think of:

* Evaluating the delays for the different airlines
* flights and carriers -fabian
* Evaluating the delays depending on the destination airports and distances
* airports and flights -fabian
* Evaluating reasons for delays
* planes and flights-usman
* Changes in delays over time
* plot avg delay of month over years-usman
* Plot the worst routes (routes with highest delays)
* Iva
* impact of weather on delays
* weather and flights-Iva
* Think of best merge for two basetables
* …

Required Output

The output required is threefold

* A text file that acts as a technical report, in which some logic is explained. Not all SQL queries should be mentioned in the document, but the code with SQL queries should also be attached
* A Tableau that works correctly on a computer that does not have the dataset available. So when I open it, it has to function correctly. This means if you create a database, it has to be accessible for the teacher as well.
* A powerpoint with a business presentation concerning your results

Compared to the individual assignment, more weight should be given to the SQL code and the SQL code will also be evaluated. Also, the tableau should be more advanced (e.g., includes a dashboard and compelling story compared to only a dashboard for the individual assignment).

Data Description

The data comprise several datasets (which can be seen as a database) related to air flights, all leaving from one of NYC airports in 2013.

(Note: the data are gathered from the R package nycflights13, with original data coming from freely available data sources; see package nycflights13 for more info.

This also offers the possibility to enrich these data, eg with potential delay reasons)

Airlines

|  |  |
| --- | --- |
| **Variable** | **Description** |
| carrier | Two letter abbreviation |
| name | Full name |

Airports

|  |  |
| --- | --- |
| **Variable** | **Description** |
| faa | FAA airport code |
| name | Usual name of the airport |
| lat | Latitude location of the airport |
| lon | Longitude location of the 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 |

Flights

|  |  |
| --- | --- |
| **Variable** | **Description** |
| year | Year of departure |
| month | Month of departure |
| day | Day of departure |
| dep\_time | Actual departure times (format HHMM or HMM), local tz. |
| arr\_time | Actual arrival times (format HHMM or HMM), local tz. |
| sched\_dep\_time | Scheduled departure times (format HHMM or HMM), local tz. |
| sched\_arr\_time | Scheduled arrival times (format HHMM or HMM), local tz. |
| dep\_delay | Departure delays, in minutes. Negative times represent early departures. |
| arr\_delay | Arrival delays, in minutes. Negative times represent early arrivals. |
| hour | Time of scheduled departure (hour) |
| minute | Time of scheduled departure (minutes) |
| carrier | Two letter carrier abbreviation. See airlines to get name |
| tailnum | Plain tail number |
| flight | Flight number |
| origin | Origin (see airports for additional metadata) |
| dest | Destination (see airports for additional metadata) |
| air\_time | Amount of time spent in the air, in minutes distance |
| 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. |

Planes

|  |  |
| --- | --- |
| **Variable** | **Description** |
| tailnum | Tail number |
| year | Year manufactured |
| type | Type of plane |
| manufacturer | Plane manufacturer |
| model | Plane model |
| engines | Number of engines |
| seats | Number of seats |
| speed | Average cruising speed in mph |
| engine | Type of engine |

Weather

|  |  |
| --- | --- |
| **Variable** | **Description** |
| origin | Weather station. Named origin to facilitate merging with flights data |
| year | Time of recording (year) |
| month | Time of recording (day) |
| day | Time of recording (month) |
| hour | Time of recording (hour) |
| temp | Temperature in F |
| dewp | Dewpoint in F |
| humid | Relative humidity |
| wind\_dir | Wind direction (in degrees) |
| wind\_speed | Wind speed (mph) |
| wind\_gust | Wind gust speed (mph) |
| precip | Precipitation, in inches pressure |
| pressure | Sea level pressure in millibars |
| visib | Visibility in miles |
| time\_hour | Date and hour of the recording as a POSIXct date |