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: https://cran.r-project.org/web/packages/nycflights13/index.html; but can be enriched with newer data, e.g., from https://www.transtats.bts.gov/DataIndex.asp) To do this, you get database tables comprising flight information, airport information, plane information, carrier information and weather information. Your goal is to use Tableau for analyzing this dataset and creating visual insights.

You can think of:

- Evaluating the delays for the different airlines
- Evaluating the delays depending on the destination airports and distances
- Evaluating reasons for delays
- Changes in delays over time
- Plot the worst routes (routes with highest delays)
- …

Part of the evaluation is made based on your originality and effort put into analyzing the dataset (see grading rubric on last page). The options above are not the only ones you should have a look at! Wondering about the questions to get insight into data is also part of a data scientist job!

You can use SQLite studio as well to create some more detailed queries that are difficult to perform in Tableau, and subsequently export them as CSV and load them into Tableau.

<u>Required Output</u>

The output required is twofold

- A text file that acts as a technical report, in which some logic is explained. (queries/joins you performed for each graph, other preprocessing on the variables, overall logic of the story, SQL code that you used to perform extra queries …)
- 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 it ideally should be a packaged workbook (see class and https://help.tableau.com/current/pro/desktop/en-us/save_savework_packagedworkbooks.htm). If the files are too large to submit on IESEG online, you can share for instance a onedrive folder or something similar.

- A publicly available Tableau workbook on Tableau public.

The Tableau file should contain a mix of simple worksheets, dashboards and one story.

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, e.g. 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

Grading rubric

		1	2	3	4	5
Data modelling	10%	The data is not prepared correctly. The tables are not joined correctly. Data types and/or values are erroneous. No extra queries performed	There are still small errors on joining, data types or values. No extra queries performed	The data are all represented in a correct way, and joins are performed well. No extra queries performed	The data are all represented in a correct way, and joins are performed well. Simple extra queries performed.	The data are all represented in a correct way, and joins are performed well. Use of more advanced queries when needed.
Visualizations	45%	Few visualizations are used and they are very simple. The visualizations contain errors such as overink, distortions of the axes, …	There are more visualizations, but there are still errors and not best fit for the purpose	There are sufficient but simple data visualizations that have been executed correctly with simple options and good fit	There graphs are diverse and include options; most of the graphs fit their purpose and do not contain graphical errors	The visualization contains sophisticated and diverse graphs and options that work correctly. No errors
Business analysis	30%	The analysis is just a gathering of information. Absence of analysis-story and/or data modelling is not coherent with analysis-story.	Some smaller parts can be discerned in the presentation, although there is no overall story. Some separate analyses	There is a basic story in the data, although it is not very well developed. The analysis could go deeper	The story is clear, but not all parts are relevant or provide insight	The analyst has developed a perfect story with factual data that provides real insight. The analyses are relevant and coherent
Presentation	15%	Visualizations are not in function of the story to be told. The story jumps from a to b without clear links; no use of graphics tools/useful graphics to support the story	There is a story, but it is not well introduced nor finished. The link between graphs and story is not always clear and hampers the story.	The story is clear, but work is needed to help the graphs explain the story. There is a conclusion.	Most of the graphs and story fit. The story is presented in a coherent and logical way with logical conclusions	The graphs fully support the story, and the options help to focus on the most important parts. After this presentation, the audience gets actionable insights