

Solve our challenge and showcase your passion for data science, your creativity and your innovative technical skills. Send your solution to challenge@datatonic.com and become part of our growing team!



Challenge : airline 'big data' analytics

Background

In the dataset you will find all domestic commercial passenger flights in the United States between 2010 and 2012. You will find departure, arrival, time and delays.

Business case

It's up to you. Try putting yourself in the shoes of someone interested in getting insights out of this dataset.

- + What is the trend over time in number of flights and in delay times? Total US and per region?
- + What are the top airports, which states have the most flights (internal/external), which routes are the most popular?
- + Which airports/states/routes have the most delays?
- + Is there any seasonality in the flights, and in the delays?

By all means this is just one angle. Your creative juices might find other ways to do something with this data.

Dataset

Source: <http://www.transtats.bts.gov/>

date	STRING	date of the flight
airline	STRING	airline name
airline_code	STRING	airline code
departure_airport	STRING	airport code
departure_state	STRING	US state
departure_lat	FLOAT	departure latitude (for maps, or distance calculations)
departure_lon	FLOAT	departure longitude (for display on maps, or distance calculations)
arrival_airport	STRING	airport code
arrival_state	STRING	US state
arrival_lat	FLOAT	arrival latitude (for maps, or distance calculations)
arrival_lon	FLOAT	arrival longitude (for display on maps, or distance calculations)
departure_schedule	INTEGER	scheduled departure time HHMM as integer
departure_actual	INTEGER	actual departure time HHMM as integer
departure_delay	FLOAT	departure delay in minutes as decimal
arrival_schedule	INTEGER	scheduled arrival time HHMM as integer
arrival_actual	INTEGER	actual arrival time HHMM as integer
arrival_delay	FLOAT	arrival delay in minutes as decimal. negative values means the flight is early

Success criteria

Some ideas..

- + Focus on efficient transformation and integration of the data,
- + Or focus on visualising the insights, either static or dynamic,
- + Or focus on an efficient statistical algorithm,
- + Or focus on an effective communication of your findings,
- + Or use a graph to represent the data,
- + Or calculate the shortest paths between any point A and point B,
- + Or a combination of all of the above

Output/Deliverable

No restrictions. Go wild.