

Topic: Study of Propagation of Contagious Diseases using Air Schedule data

There are likely many factors explaining the propagation and control of contagious diseases. In this topic, we suggest to look at how they can be related to air travel, through the proxy of air flight schedules. Hence, during this week, students could include flight schedules as an additional element in the analysis.

Some aspects that students could work on during the week include:

- Quick state of the art review: how does the transport industry influence contagious diseases propagation? What has been proposed so far?
- Modelling of diseases propagation: apply AI knowledge and models to study this, taking into account the provided data
- Simulation tools to forecast future diseases spreading. Eg., what-if there is a new virus first detected in Congo in 2025, with some parameters like $\rho = 1.2$, etc ? How this will spread? Which countries, and when?

It is important to note that these are suggestions, and we are of course open to new angles and conclusions proposed by the students.

Amadeus Schedule Data Summary

The dataset is a summary of the full 2019 worldwide flight schedule. More precisely, it is a detailed count of the number of aircraft seats scheduled, broken down by the following dimensions:

- **Departure city** and **arrival city**. Cities are represented by their 3-letter IATA code. If a city has several airports (such as Orly and Roissy CDG for Paris), those will be merged at the city level
- **Departure week**. This is the week of flight departures, in the YYYYWW format. For example, 201901 means the first week of 2019, 201902 the second week of 2019, and so on
- **Total seats**. This is the total number of seats flown on that city origin-destination (O&D) pair during that week. It's relevant to note this is not the number of passengers, but the physical capacity of aircraft – those two values are nevertheless correlated

The full schedule data set is about 71MB, and 2.6 million rows, in one csv file, for close to 4000 arrival and departure cities. The format follows the dimensions above. A sample is below:

Departure City	Arrival City	Departure Week	Total Seats
BOM	BHU	201949	490
JUH	XMN	201949	484
CWB	BHZ	201950	826

Supporting data

Additional information on the cities (country, coordinates, etc) can be found in the open travel data (optd) repository: <https://github.com/opentraveldata/opentraveldata/tree/master/opentraveldata>

Public data on epidemics such as covid-19 can be found on several repositories. One typical example is the “our world in data” repository around covid, covering aspects such as tests and case numbers: <https://github.com/owid/covid-19-data/tree/master/public/data>

Contact

This dataset and project is proposed by the Amadeus AI Research and Technology team (“ART”). You can contact us at this address: research@amadeus.com

The people in the team responsible for this project are Thierry Delahaye and Rodrigo Acuna-Agost.