

Southampton Hackathon

Overview

The Hackathon challenges your ability to forecast a complex time series over a multi-period horizon. Specifically, you'll produce load-factor forecasts for 50 airlines, system-wide (how full on average will the airline's planes be). To produce your forecasts, we've provided several datasets including:

- Airline Type - The general business model followed by the airline.
- Seats - How many seats will be offered by the airline at the prediction point one year into the future aggregated by year, month, airline, departing and arriving country
- Covid - A variety of covid related datapoints offered at the daily resolution for countries
- GDP - Quarterly reported and forecasted Gross Domestic Product for selected countries (includes several aggregated indexes).
- Training & Validation Load Factors - For several airlines, you'll receive a mixture of reported load factor information at the yearly, quarterly and monthly level. The data does have gaps, and reporting lags.
- Lookup tables for airlines & countries to help join datasets.

On the submission, you'll have to forecast both airlines that have reported load factor data available for modeling and airlines that do not share load factor data. Further complicating efforts, there are lags and variability in reported load factors. To help modelers, the actual data for 10 airlines is provided as a validation dataset. Submissions will be evaluated on MAPE. The top 6 teams will be given an opportunity to present their approach and findings to a panel of judges who will determine winning teams.

Datasets

Airlines (airlines.csv)

Includes an airline type that can be used in modeling.

Field	Type	Description
airline_code_iata	varchar	IATA code associated with the airline. Note there can be duplicate codes issued by IATA.
airline_type	varchar	There are several options: <ul style="list-style-type: none">• Low-cost• Regional• Mainline• Leisure• Cargo• Airline ACMI (Leasing Arrangement to provide aircraft, crew, maintenance & insurance)

Country (country.csv)

Lookup table that includes iata and iso codes for joining data.

Field	Type	Description
country	varchar	Name of the country
iata	varchar	2 letter country code issued by IATA
iso	varchar	3 letter country code
numeric	integer	Numeric code associated with country

Covid (covid.csv)

Compiled covid data by date.

Field	Type	Description
iso_code	varchar	Iso code of the country

location	varchar	Country name
date	date	Date of reported information
total_cases	float	Total confirmed cases of Covid-19.
new_cases	float	New confirmed cases of Covid-19
total_deaths	float	Total attributed Covid-19 deaths
new_deaths	float	New attributed Covid-19 deaths
stringency_index	float	Composite measures index based school closures, workplace closures, and travel bans. 0-100, where 100 is most strict.
reproduction_rate	float	Estimate of the effective reproduction rate (R) of Covid-19
total_tests	float	Total Covid-19 tests performed
new_tests	float	New Covid-19 tests performed
positive_rate	float	Share of Covid-19 test that are positive in 7 day moving avg.
total_vaccinations	float	Total vaccinations administered
population	float	Population
population_density	float	Population divided by land area (sq km).

GDP (gdp.csv)

Field	Type	Description
location_iso	varchar	Country iso code or index: <ul style="list-style-type: none"> • OECD - Org for Economic Co-Operation and Development • OECDE - OECD Europe • G-7 • G-20 • EA19 - Euro Area (19 Countries) • EU27_2020 - EU 27 Countries in 2020
indicator	varchar	Indicator reported - only QGDP is made available
subject	varchar	Only TOT - Total is made available
measure	varchar	Two types: <ul style="list-style-type: none"> • PC_CHGPP = Percent Change over prior period • PC_CHGPY = Percent Change over prior year
frequency	varchar	<ul style="list-style-type: none"> • A = Annual • Q = Quarter
time	varchar	Depending on frequency: <ul style="list-style-type: none"> • Year yyyy (for A) • Year-Q(1-4) (for Q)
value	float	Floating point increase (negative decrease)

Load Factors (load_factor_data.csv)

A collection of reported load factors by airline and accounting period. It is expected that there will be gaps in reported load factors and overlaps (annual and monthly for example). There is a time lag in reporting load factors (meaning that it can take several months for a load factor to be reported).

Field	Type	Description
organization_code_iata	varchar	IATA code of the airline
organization_code_icao	varchar	ICAO code of the airline
period_start_date	date	Starting date of the load factor estimate
period_end_date	date	Ending date of the load factor estimate
accounting_period	varchar	Description of the type of load factor reported: <ul style="list-style-type: none"> • Annual • Quarter (1-4) • Month (1-12)
passenger_load_factor_pct	float	0-100, the load factor for the airline, system-wide over the time range.

Seats (seats.csv)

The seats table provides the number of scheduled seats moving between departure and arrival country. Domestic traffic is represented as departure_country_iata = arrival_country_iata.

Field	Type	Description
operating_airline	varchar	IATA code of the airline
departure_country_iata	varchar	IATA code of the departure airport's country
arrival_country_iata	varchar	IATA code of the arrival airport's country
seats_year	integer	Year (%Y, yyyy) for the seating record
seats_month	integer	Month (%m, 1-12) for the seating record
total_seats	integer	Total number of scheduled seats moving from departure country to arrival country

Validation (validation.csv)

Ground truth for 10 airlines to be used in training.

Field	Type	Description
load_year	integer	yyyy - Year of the forecast (2022)
load_month	integer	mm - Month of the forecast (1-12)
organization_code_iata	varchar	IATA code of the airline for which the load factor forecast is generated.
passenger_load_factor_pct	float	The team provided submission for the year, month and airline. Should be between 0 and 100.

Submission (submission.csv)

Your data submission involves filling in the passenger_load_factor_pct for 600 rows in the submission file. Any blank or unfilled submissions will be set to 0.0. The accuracy portion of the contest will be judged on MAPE (mean absolute percentage error).

Field	Type	Description
load_year	integer	yyyy - Year of the forecast (2022)

load_month	integer	mm - Month of the forecast (1-12)
organization_code_iata	varchar	IATA code of the airline for which the load factor forecast is generated.
passenger_load_factor_pct	float	The team provided submission for the year, month and airline. Should be between 0 and 100.