

FLYR Labs Data Science Home Assignment

Description

The goal of this task is to evaluate your data preparation, analysis and modeling skills. You receive on-time airline performance data and the question we would like you to answer is:

What departure delay is expected for any scheduled flight in the test set? What is an error of the model's prediction on the test set?

Things to note:

1. Data consists of 14 months, where 12 months (07.2018 - 06.2019) are for training/validation and 2 months (07.2019 - 08.2019) for testing purposes.
2. DepDelay (in minutes) is your target variable.
3. From an operational perspective, flight is classified as delayed, when it is late 15 minutes or more.
4. Scheduled departure time of the flight is the last available moment for doing prediction.
5. For modeling/analysis use a subset of 8 carriers, which are the most common in the dataset.
6. Feel free to use any publicly available information/dataset for feature engineering.
7. Don't forget about exploratory data analysis (at least 3 questions & answers).

There is no need for intense hyperparameter tuning (if training takes too long). We will evaluate your work as a whole, so a creative and thoughtful approach to the data and modeling is more important than getting a small improvement in your evaluation metric.

Our preferred format is Jupyter notebook (with printed results & your notes), but if you use other tools, then any form including source code + explanation would be accepted.

Dataset

COLUMN	DESCRIPTION
FL_DATE	Flight Date (yyyymmdd)
OP_UNIQUE_CARRIER	Unique Carrier Code
OP_CARRIER_FL_NUM	Flight Number
ORIGIN	Origin Airport Code
DEST	Destination Airport Code
CRS_DEP_TIME	CRS Departure Time (local time: hhmm)
DEP_TIME	Actual Departure Time (local time: hhmm)
DEP_DELAY	Difference in minutes between scheduled and actual departure time. Early departures show negative numbers.
CRS_ARR_TIME	CRS Arrival Time (local time: hhmm)
ARR_TIME	Actual Arrival Time (local time: hhmm)
ARR_DELAY	Difference in minutes between scheduled and actual arrival time. Early arrivals show negative numbers.
CANCELLED	Cancelled Flight Indicator (1=Yes)
CANCELLATION_CODE	Specifies The Reason For Cancellation
CRS_ELAPSED_TIME	CRS Elapsed Time of Flight, in Minutes
ACTUAL_ELAPSED_TIME	Elapsed Time of Flight, in Minutes
AIR_TIME	Flight Time, in Minutes
DISTANCE	Distance between airports (miles)
CARRIER_DELAY	Carrier Delay, in Minutes
WEATHER_DELAY	Weather Delay, in Minutes
NAS_DELAY	National Air System Delay, in Minutes
SECURITY_DELAY	Security Delay, in Minutes
LATE_AIRCRAFT_DELAY	Late Aircraft Delay, in Minutes