

DATA SCIENCE COMPETITION 2020

Sponsor

Institute of Data Science – Texas A&M University

Challenge

Build models for predicting expected flight delays by airline and flight for the third quarter of 2019.

Data

Historical data from the Bureau of Transportation Statistics for 2018 and the first half of 2019. The data files, are CSV files containing records for almost 11 million flights in 2018 and 2019. The data dictionary is in an excel file FlightDataDescription.xlsx.

Data Files:

FlightDelays.csv: 10,915,495 flight logs for 2018 and the first half of 2019

Airports.csv: Airport locations and codes for 362 passenger airports in the United States

Routes.csv: Distance for 6,684 airport routes with origins and destinations in the U.S.

Airfares.csv: The average airfares between cities for 2018 and the first half of 2019.

Competition Details

Students will work in teams of up to 5 analyzing data involving passenger flights in 2018 and 2019. Teams have the option of using a faculty mentor to facilitate problem solving.

Contestants will have one month to complete their analysis and submit their data file and findings to be reviewed by the judging panel. From these entries, judges will work to narrow down and select the top ones to move forward to the final round of the competition. The chosen finalists will present their analysis to the judging panel as part of a one-day event where the ultimate winner be determined.

Written Solution

Each team is asked to submit a report describing their solution.