Amsterdam Airport Schiphol - Data Science Assignment

This is it, your moment to shine! We kindly ask you to work on the two parts of this assignment and to send us your code and a summary of your results in an easily readable format (e.g. a Jupyter notebook). We don't expect the results to be perfect, final or complete, but rather will use them to assess your technical skills as well as your understanding and flexibility when handling new datasets. Please show us your code and explain how you arrived at your answers. When judging your assessment we will look at the approach you took to solve the problem at hand, but also as at your code. Feel free to make assumptions about the problems, just clearly state and justify them. Good luck!

Part 1: Predictive Modeling

Predicting delays of departing aircrafts is one of the projects that we work on here at Schiphol - and this is your chance to give it a go as well! We provide you with more than six months' worth of publicly available departure data (`flights.csv`) and would like you to show off your technical and analytical skills by addressing the following points:

- What are typical delays? What factors could be good predictors? Any insights? Perform any cleaning, visualization, exploratory analysis, and/or machine learning that you see fit to answer these questions. Be creative! (Note: You can obtain the delays by combining scheduled date and scheduled time and subtracting this from the actual off block time, which is the time the plane actually left. Look here for further information: https://developer.schiphol.nl/apis/flight-api/overview, you will need to register an account)

- Build a predictive model for delay on the short term! This mean we want to be able to predict 2 hours ahead whether a flight will be delayed or not. What could serve as a good baseline and how does your model perform? Which metric did you choose and why? Are there any features with predictive power? How do you encode highly categorical features? Do you take into account any information from earlier on the day? What can we learn?

- Would you recommend your model to be used by the business? And could you think of any use-cases where being able to forecast delay could be a meaningful input?

The data was pulled from the Schiphol Flight API. Under the following link you can find a full description for all the fields and also request access to the API yourself: https://developer.schiphol.nl/apis/flight-api/overview Feel free to use any additional data source that is publicly available to improve your model, i.e. data from the Schiphol Flight API or any other website.

Part 2: Your own work

We love to see what interests you. Therefore we ask you to share a machine learning project that you are most proud of. Send us a github link, a zipped archive or whatever format you like. It would be great if you accompany your code with a small readme that describes what the project is. We don’t need to be able to run it, we just want to understand what makes you tick as a data scientist!