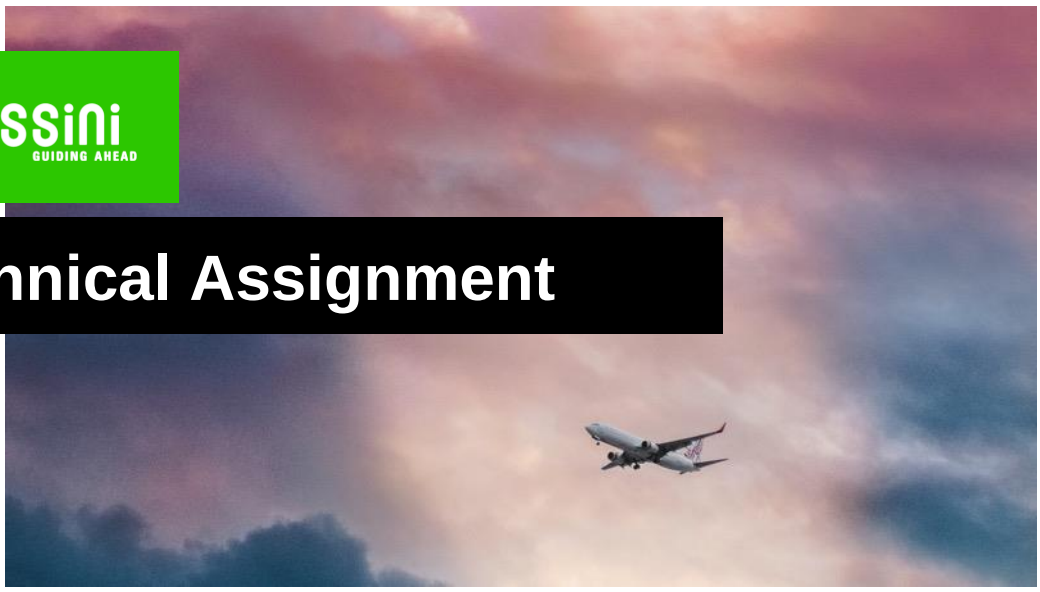

Technical Assignment

For the first part of the exercise, we ask you to download (with a script if you can) the complete data from www.transtats.bts.gov for the entire year 2019 and write a script that imports the data into a database with adjustable specifications.

For the second part, your objectives are to create a script that performs a full descriptive statistics for the flight delay in San Francisco. Try to offload as much of the calculations as possible to the database level and investigate what the most common factors (*Cause of Delay*) for a departure flight delay at the San Francisco airport (SFO) are. Analyse if there is a significant difference of delay between the carriers at the San Francisco airport in 2019. If so, which carriers are better and which worse? Conduct a complete analysis for this part (including centrality, variability and shape). Please write a reproducible report (a report that generates at compilation a document with output based on provided data). You can choose for this task one or more scripting languages, such as R, Python, or Julia (including all libraries that are available for that language e.g., C libraries in R) and send all script files (without the data files) to Matthias Mörch at least a day before the interview. The raw files need to be able to compile on any given computer that has the languages and dependencies installed.

Finally, create a presentation (in German) to explain your findings (not the mathematical foundations) to a board of directors of American Airlines (AA), that do not have any knowledge of statistics, but need to decide on how to tackle the delay problem of their airline.
You will present your findings in the interview for about 20 min.

You will have until your interview to complete the exercise, but you may return the exercise at any point during that period. This exercise should not take too much time – we give you ample time to help you fit this exercise into your schedule. Please keep in mind that this exercise is not designed to be overly burdensome in terms of the time commitment.

If you have questions during the exercise, please email Matthias Mörch.