



# Coding exercise

This coding exercise is meant to assess your development standards and best practices when developing an ETL for a new dataset. Please write a Python script that performs every step of a data processing pipeline, from raw data to clean data ready to be used for statistical modeling purposes.

Your code should process two example datasets that will be provided to you.

1. **airline\_flights.csv**
2. **big\_tech\_stock\_prices.txt**

What the steps of the data processing pipeline are is on you. It should contain at least some reading, validating, and saving of data steps. The code you write and hand-in will be discussed during the coding interview.

**Important:** do not spend more than three hours on this. This coding assignment is only a means of getting an idea of your coding and data science skills. You should not spend too much time on studying what the best setup for a data processing pipeline should be.

**Hint:** If you do not have enough time to finish everything you would like to code, only write the outline of your program. This way you show your thought process during development.