



As the next step of the Blue Yonder hiring process, we kindly ask applicants to hand in a solution to the small programming task below to demonstrate their professional approach to developing software. For us, this includes appropriate unit tests and the application of clean code principles. We value both technical correctness and coding style.

Please be assured that there is no time limit, so you are free to tackle this task whenever it suits you best. Enjoy coding, we are looking forward to receiving your response. If we like your solution, we will invite you for interviews, where we will also talk with you about your code.

If you have any questions, please do not hesitate to contact us.

Part 1

Please make an explorative data analysis and build a prediction model for the hourly utilization “cnt” of this data set: <https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset>. Your solution will consist of a short analysis (text ~ 1 DIN A4 with additional pages for plots) and the relevant code your report is about (source files).

- In your report, present only one model, that you think is most suitable for a business-case. Summarize your reasons for choosing this model. Report the mean absolute deviations.
- Assume that the code you are writing is used in production in a daily prediction service and maintained by your colleagues (what could that mean?)

Please do not spend more than a few hours on the topic. Your code and predictive model are of equal importance to us.

Part 2

Above, you were asked to write a model for a small-to-medium data set. Can you outline how a solution would look like that is able to scale up? (half a page up to 1 page).

- What are the scaling properties of your model, if you assume that the amount of data you need to handle go up to several terabytes? Do you see any problems?

- How would you address these problems? Are there technologies for data storage/predictive modelling you can build upon?

Describe how the technologies you mention solve the scaling problems you see with model.

- What are the limits and drawbacks for your new approach?

- Do you have hands-on experience with such technologies? Which ones? For how long?

PLEASE COMPLETE THE TASK USING PYTHON

