



Data Science Coding Challenge

Thank you for your interest to join our data science team! Before you join us we like to give you a chance to experience the kind of data and problems we work on. Your submission gives us an opportunity to get an impression of your thinking, communication, and coding style.

Please note that we are interested in getting to know how you approach the problem more than in the actual results. We respect that you have a busy life, so please focus on quality over quantity. We expect that this challenge will not take more than 3-6 hours of your time.

Challenges

We prepared two challenges - please select one challenge for your submission.

Critically examine the dataset of your choice while paying attention to data problems such as missing data or small sample size.

Time-series Anomaly Detection

The dataset contains readings from a temperature sensor. The goal is to detect outliers and anomalies in the dataset. Please take care when exploring the dataset to note what you observe and provide an outlook what next steps could be that follow after your solution.

Customer Lead Generator

The columns 'b_in_kontakt_gewesen', 'b_gekauft_gesamt' will help you to identify the training data and the labels. The goal is to train a model that can be used to support customer lead generation. Please take care when exploring the dataset to note what you observe and provide an outlook what next steps could be that follow after your solution.

Remarks

- Explain your code and results, using inline code documentation and/or notebook text cells
- Please state and explain simplifying assumptions you may make
- Please remember testing - both the data and your code

Deliverables

Provide your results and the necessary code to create it as a single .ZIP file via Email. If you can not attach it please provide a download link to a hosting service such as OneDrive, GDrive, Dropbox, etc.

We are a team of data scientist that work with Python and Jupyter on a daily basis. A self-contained submission that we can run inside a container or virtual environment is our preferred working mode. We like to track our code with git and write tests for (helper) functions and packages we use to analyze data and train models.

