



Data Science Exercise

This exercise is designed to test your machine learning and coding skills. In addition, we want to evaluate your ability to formulate interesting questions and present your results.

Download the Yelp data set (https://www.yelp.com/dataset_challenge/dataset) and formulate one or a few interesting questions you want to answer using the data. Use any number of machine learning algorithms of your choice to answer your question(s). You should use Python, R or Scala to solve this exercise. Please make your script(s) executable as a Docker container, such that we can reproduce your results.

Once you are done, create a short presentation containing your question(s) and explaining your findings. Publish your code on GitHub and include a README, explaining how to run your code using Docker. Send the link to your repository and the presentation to jobs@newyorker.de once you have completed the exercise.

Good luck!