The Data Science Job Search

Rachel Berryman

Agenda

- About me
- State of the market in Berlin
- How to get interviews
- Tailoring your application
- Nailing the coding challenge
- Your questions

Questions

Slido.com

#35884

BA Economics
MSc Sustainable Development

DATA SCIENCE RETREAT®
SINCE 2014

Data Science Bootcamp

pmOne

Data Science
Consultant

Energy Analyst

SQL

Data Scientist

TEMPUS

Applying in Berlin

- 2.5 years DS/Analytics Manager experience
- “Wohnsitz” in Berlin
- Experience hiring & leading DS team members

My Applications by the Numbers

State of the Market

- Research-oriented vs. pragmatic roles
- Junior vs. mid-level vs. Senior roles
- English vs. German speaking jobs
- Movement towards ML-Ops

Recruiter Tips

- More competition during Corona Crisis
- Coding challenge will be even more important
- Surge in companies building chatbots (NLP)
- DS for automation

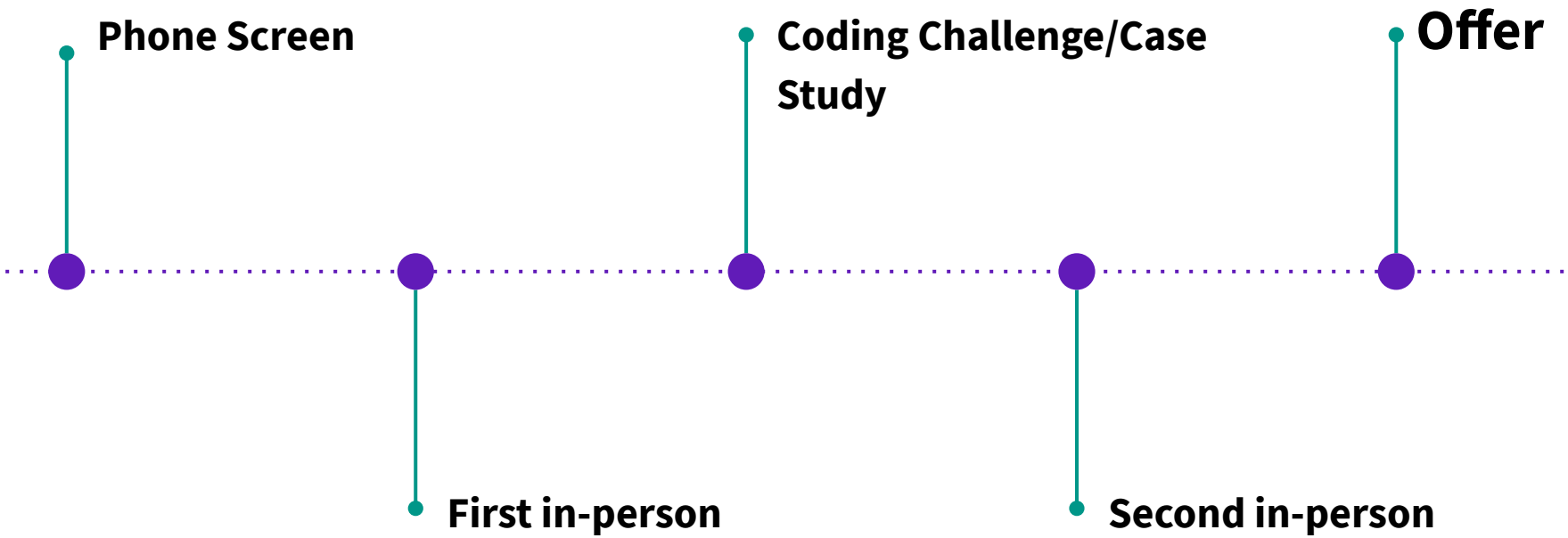
Making your application stand out

- Send a link to a project or paper
- Follow up with current employees
- Include a list of what you’re currently learning/doing
- Link to blog

Getting Interviews

- Meetups
- LinkedIn
- Alumni Organizations

Typical Application Process



Application Tips

- Tailor to the country and industry to which you're applying
- Numbers are worth 1000 words
- Apply through an employee wherever possible

Hello,

Hope this email finds you well. I am a recent computer science graduate from the American University of Be
Data Scientist on LinkedIn, and I found that it is in line with my keen interests in insert tech fields related. I a
in a prominent American startup, and I am looking to grow from this role into more challenging positions. Yo
work experience and background in the attached CV. I am looking forward to hearing from you in case an ap

Interview Tips

- Have your salary expectation and start date for that role ready
- Have questions for them
- Strong GitHub presence
- Experience speaking at Meetups
- Have a clear reason: why ***this*** role?
- Have a clear reason: why ***you***?

The Coding Challenge

- Get crystal clear on the asks
- Start as early as possible
- Submit early
- Refactor, refactor, refactor.
  - Functions
  - Basic tests
  - Clear visualizations
- Submit EXACTLY as instructed

The dataset contains **daily visitor data** from a shopping mall.
Dataset includes numeric and binary data columns. The column name is only available for 2 of the 15 features. The name of the rest of the features is unknown.
Weather data for the area where the shopping mall is located is also provided.
The task is to predict the column called 'label' for the test set.
Prediction error will be measured b=using MAPE.
You may complete the challenge using Python or R.
Please provide your (commented) source code, a report of your approach, where you describe in your own words:
* The libraries used and instructions of how to run your code
* An export (e.g. HTML, PDF) with the output of your code or notebook/report (if you used e.g. jupyter)
* How and why you transformed the data
* The approaches used to analyze and predict the daily visitors
* The evaluation as well as a comparison with a baseline approach
* The time you spend for solving the task (Do not spend more than 10 hours!)
Please send your final predictions of the test set (as csv file containing only the date and label column). Make sure that your final prediction contains an entry for every day (resulting in 363 predictions in total) and that the date column is formatted as Year-Month-Day (e.g. 2015-09-25; in python: %Y-%m-%d).

Hint: An appropriate baseline will give an MAPE of approx. 30.

Contact

rachelkberryman@gmail.com

https://www.linkedin.com/in/rachelberryman/