



DataFest Student Infos and Guidelines

Welcome

- FOLLOW US ON TWITTER: @DataFestGermany. Any updates, announcements, or surprises will be posted on twitter. You may also post questions or comments to twitter. Note that we will NOT answer detailed questions twitter---to get help, you must be physically present.
- We have arranged for consultants to stay for the entire 44 hours.
- We have quite a few “VIP” consultants visiting. These are “data professionals” from around the country, and also academics and researchers. They are here to help, and so ask them if you get stuck on anything or are puzzled by anything. Their goal is to talk to you, to see what you are thinking about, and to see if they can help.
- This is a friendly competition. Feel free to ask for help, but don't be surprised if no one knows the answer! We'll do our best, and if at any time we learn the answer, we'll share it with you. In exchange, if you solve a problem that another team might be having, please share it with them.
- There are students from twelve universities here. Mingle! You may be co-workers someday or classmates at graduate school.
- Most of all: have fun. Meet some new people, allow yourself to get into some irrelevant conversations, don't stress out, learn something new, and have a good time.

Rules

- You can come and go as you please, but all work must be completed on premises.
- PLEASE help keep the building clean. We are visitors here, and so treat the building as you would a friend's home. A friend who will not clean up after you.
- The data will be provided by a private company. Without explicit instructions from the provider, it must be used onsite at DataFest only. Usage of the data apart from DataFest is not permitted. At the end of DataFest, delete all data from thumb drives, hard drives, etc. You MAY, however, share your results and presentations with others.

- We will make photos at the event and also the local television may come. Therefore we require you to give us full rights on all photographic and video material we make during the event.
- You must return the DataFest USB when handing in your final presentation.
- Please wear your name tag at all times so that we know who belongs and who doesn't.
- Should members of your team drop out at the last minute, you might be merged with another team who is also missing members.
- At all times between 9am-12midnight there will be a friendly Consultant present. These are faculty, PhD students, members of the p3 group, or other professionals with field specific knowledge on the dataset. They all have different areas of expertise, so if you get stuck on something, ask someone else later. Feel free to ask anything. This is not an exam, but a collaboratory competition. Do not expect the Consultants to write code for you, or do data management, etc. They are there to help point you in the right direction, but you're responsible for getting there on your own.

Judging

- Each team will have 4 minutes to present their findings to the judges.
- Each team will be allowed **at most three slides**. Three. So at some point on Sunday, you might want to set aside time to think about what you want the judges to know. The 4 minute time limit will be strictly enforced. All team members must be present for the presentation, but not all team members need to actually speak (given the time limitation).
- Your presentation must be emailed to **datafest@stat.uni-muenchen.de** by 11:30am on Sunday. Allowed formats: **PDF**. If using a web-based tool like GoogleDocs or Prezi, please export to PDF and send the PDF as your submission, you will not have time to log on/off to your account during the presentations.
- **Your file with your slides must have the same name as your team.**
- Awards will be given in three categories:
 - Best Insight: This is the main prize.
 - Best Visualization
 - Best Use of Outside Data

Supplies

We strongly recommend that every member of the team bring a laptop, if possible. We recommend that you make sure beforehand that the software you will be using throughout the weekend is properly installed and running on your computer. You will be working with a large dataset so make sure that you have the space for it on your drive.

We will have snacks and munchies. Feel free to bring anything additional you might want. You are of course free to come and go as you please, but particularly the first night (up until midnight) will be fairly structured.

You might want to bring some favorite statistical or computational reference books, if you have them, or bookmark some pages that you routinely refer to.

Large Data Advice

The dataset you will be working with is quite large. If you type a variable name to view it, it will take a while to display. Therefore, remember these R commands: `head()`, `tail()`, `str()`.

We strongly recommend you create a small data set that you can use to test things on. Then, if it works out, you can apply your procedure to the large dataset. Some procedures can take a frustratingly long time to run on large data sets, and so it will be comforting to know that your procedure works (because you tested it on a smaller data set) while you wait. We recommend taking a random sample of rows from the original data set, but there might be other approaches you find useful.

Working with large and complex data is hard. We have made sure that many competent consultants are on site, so please use these resources and ask any remaining questions. It is your task to come up with a great solution, but we are there to help you out with hints and advice.