DataFest Visitor/Consultant & Judging Guidelines

Location

Ludwig-Maximilians-Universität München
Geschwister-Scholl-Platz 1, 80539 München
Rooms: Reception Desk: Senatssaal E106, Global events: Lecture hall E004

If you are judging, please be at the Senatssaal (room E106) no later than 11:30 on Sunday. We will walk together after some coffee and instructions to E004 for the final presentations.

Directions

By Car:

LMU Munich's historic Main Building is located in the heart of the city. Since parking facilities near the Downtown Campus are extremely limited, we recommend that you use Munich's excellent public transportation network.
Go to http://www.mvv-muenchen.de/

By Train:

From: München Hauptbahnhof
Go to http://www.mvv-muenchen.de/
From the Munich Central Train Station take the U-Bahn (subway) line U4 (direction Arabellapark) or U5 (direction Neuperlach Süd) to Odeonsplatz. Change train and take the U-Bahn line U3 (direction Olympia-Einkaufszentrum) or U6 (direction Garching-Forschungszentrum) get off at Universität. Walk along the platform in the train's direction of travel and take the stairway at this end of the station, which will bring you up directly in front of the LMU Main Building.

Your Task

The most important thing is to keep the mood light and encouraging! I suspect by sometime Saturday afternoon things might seem rather dire to some of the students.

On Friday night, students will be busy trying to make sense of the data. We expect that some of them might have technical problems with getting started (loading the data, viewing it, etc.). Throughout the weekend the teams will be on their own,

though we might have intermittent “check ins”. We imagine they will get stuck and need advice. Sometimes, the advice could be highly technical and, depending on your background, outside your expertise. Don't worry. They know that you are not there to solve their problems, but to offer advice. See if you can steer them towards standard problem-solving techniques: break the problem into smaller pieces, go online for advice, etc. Guide them to think about context. What sort of distribution do they expect? Why? What might cause that? How does that compare to what they saw?

This is a competition, but it is supposed to be friendly and collaborative, so don't worry about revealing any special knowledge. This is not an exam, and so if someone asks, please answer if you can, and don't worry about other teams over hearing. I'm hoping that, after the first evening, teams will be sharing basic technical advice on their own.

The Data

This year’s data is provided by Ticketmaster, a website that sells tickets for concerts, sports, arts, theater, and other events. Students will not know the source of the data until Friday evening. At that point, we will present the data and the outlines of the challenge.

Although the data provider points to a particular type of challenge, we will encourage the students to think about possible strategies to improve ticket sales, challenges in the analysis of web-based transaction data and questions of representativity in self-selected samples in as large a context as possible, and will certainly encourage them to explore any hypotheses they might generate.

Students will download the data on Friday, which contains information on:

- Ticket purchases over the last ten years with details about each event
- Advertisements via Google AdWords, including information about the bidding process for specific ads
- Customer online behavior when visiting the website

Students will also be provided a codebook along with some additional reference material. These data pose particular challenges:

- Real data from online transactions containing relevant and non-relevant variables
- Complex data structure with multiple files
- Large data (Gigabytes) prohibiting certain ways for analysis
- Handling of geographic and transactional data is not learned in class
- Concerns about representativity

A note about software

Most students will use R but we also expect many to use Stata, Matlab, Python, SAS, SPSS, Tableau, etc. as well. If asked for support on a platform you’re not familiar with, simply stating so is sufficient.

Judging

Judging will begin on Sunday at 12 noon .The presentation and judging schedule is as follows:

12:00 - 14:15	Presentations
14:15 - 14:30	Judges deliberations
14:30 - 15:00	Award ceremony

Teams will have 4 minutes and 3 slides to make their case. You will be provided note cards to help keep things straight. It's strongly recommended that you make note of teams that are especially strong as you're watching the presentations so that the deliberation conversation can be as efficient as possible.

There will be three prize categories:

- Best in Show
- Best Visualization
- Best Use of Outside Data

The latter category is meant to encourage students to find supporting information and/or data beyond what we give them.

Judges	
Jean-Paul Schmetz	Cliqz / Burda
Dr. Torsten Schön	Audi
Prof. Dr. Volker Tresp	Siemens
Dr. Jan Ittner	Boston Consulting Group
PD Dr. Matthias Schubert	LMU Munich - Computer Science
Prof. Dr. Bernd Bischl	LMU Munich - Informatics

Feedback

DataFest has become an annual event that is now being held at seven locations in the United States. With DataFest Germany we are the first to bring this format to Europe. We hope to grow the event further in the coming years. With that in mind, please send me any advice or constructive criticisms that will help us improve this event in the future.

Our primary goal is to provide a rewarding experience for the students, an experience that sharpens their analytical skills and gives them some confidence that they can take what they learn here out into the "real world".