1 Overview

The goal of this session is to get familiar with core data mining tasks and the data mining process *without* diving into the details of how algorithms work. We provide JUPYTER notebooks, but feel free to use any tool you know/like: e.g. WEKA, MATLAB, EXCEL, GNUPLOT, your own programs and scripts, as long as they help gain insights into the data.

2 Exercise 1: Hospital Data

Download and unpack `session1.zip` from Toledo. Run the `eda-hospitals.ipynb` notebook. Using the guidelines, get familiar with the data and identify its interesting properties. You will look at attribute distributions, relations between attributes, missing data, and so on.

3 Exercise 2: Bike Rental Usage Analysis

Run the `bikes.ipynb` notebook. The dataset you will be working with is related to the two-year historical log corresponding to years 2011 and 2012 from the Capital Bikeshare system, Washington D.C., USA. In addition, the corresponding weather and seasonal information have been added to the dataset.[1] Try to:

1. Discuss possible use cases of these data.[2]
2. Complete the notebook to answer the questions about the data.

[1] Raw data is is publicly available on `http://capitalbikeshare.com/system-data`
Weather information is extracted from `http://www.freemeteo.com`

[2] See, for example, the following articles making use of similar data: `http://toddwschneider.com/posts/a-tale-of-twenty-two-million-citi-bikes-analyzing-the-nyc-bike-share-system/`