Overview:

The speed dating dataset was collected by Columbia Business School (<http://www.stat.columbia.edu/~gelman/arm/examples/speed.dating/>) from 2002 to 2004. The experiment covers a series of events (waves) where people meet people from the opposite sex in 4-minute dates, and, afterwards, rate the partner and decide whether there is any interest in meeting the partner again (in a real date). There is a well-established process in which participants are required to answer questions (repeatedly) in different phases of the experiment such as “What do you think MOST of your fellow men/women look for in the opposite sex?” or “What do you think the opposite sex looks for in a date?” or even questions about how the person thinks he stands: “How do you think you measure up?”

All questions asked the participants to rate six attributes (five if the participant is talking about himself): attractiveness, sincerity, intelligence, fun, ambition, and shared interests. In a few rounds of the experiment, the rating measurement was altered from distributing 100 points between the attributes to rating each attribute from 1-10.

Database mapping (from original CSV dataset)

To start with, each “contender” is asked to fill a background questionnaire regarding various points such as age, race, interests and how important he rates a few matters such as same religion or same race – the list long but simple. After each date, the participant is asked to rate the partner’s attributes culminating in one variable: whether the participant wants (or not) to have a date with the partner. This is mapped in the Database in entities: PERSON, SD\_WAVE, SD\_WAVE\_PERSON, SD\_WAVE\_PAIR.

Throughout the event and up to 3-4 weeks later, the participant is asked to continue answering generic questions (always rating the same attributes listed above). Many of the questions proposed repeat a few times and seem more related to behavior than necessarily to the choices and decisions during the event. That is very interesting data though – especially if the answer for the same question can be mapped over time showing the progress. Such questions and their answers are mapped in the following Database entities: PREFERENCE\_PERCEPTION and SURVEY\_QUESTION.

A initial ERD proposal is shown below – this might change slightly according to the data we find in the dataset.

[Skecht\_ERD]

Main tables’ samples roughly obtained from the original CSV are also displayed below. The data might change slightly (addition/removal of fields), but the whole idea of how the data is stored should be the same.

[Sample\_Person\_Data]

[Sample\_SD\_Wave\_Person]

[Sample\_SD\_Wave\_Pair]

[Sample\_Preference\_Perception]