

Project IGR204

Speed dating



Antoine Boulat | Simon Delarue | Mohammed El Yaagoubi | Mathias Nourry | Lingli Zhan

MS BDG / IA

I. Choice of dataset

We choose the dataset Speed Dating. This dataset is composed by columns and rows that are contained in Comma Separated Values format.

This dataset can help us to answer some of the following questions :

1. Evolution of the criteria of women/men according to their age
2. Analysis of proportions of matches between men/woman according to criteria
3. Reproduction analysis (are people looking for the same profiles as they are)
4. The bias of the financial situation (will rich people attach themselves to different profiles?)
5. Impact of money, is the financial situation crowding out the other criteria
6. Does the match rate increase with the profile completion rate?
7. ...and other questions might come after studying more deeply the dataset

Context:

A sociologist seeks to study the criteria of human interactions and more specifically how do couples come together and the alteration of these criteria by societal biases (money impact, social reproduction ...). This study must be very broad in scope because it will appear in a press article.

What are readers background ?

We assume readers with a wide variety of profiles.

So, the aim of this article is to provide a general information as well as comprehensive overview.

What is the objective of the sociologist ?

His goal is to analyze human behavior, confront the clichés and hence help readers to develop a more defined understanding of nowadays gender expectations.

And why not even made them think about the implications of all this for our society.

Are visualization tools used by the sociologist aimed primarily at *exploring* or *communicating* the data?

They will be used at exploring potential correlations in the data. The purpose of this approach is to analyse interactions between groups of people having in mind all the above questions.

Then, they will be used by the sociologist to communicate his findings to readers, especially make the understanding easier for non-specialists.

II. Description of dataset

The dataset contains 8379 entries for 195 variables. Yet, data is missing for almost 26% of the whole dataset (or is non relevant to be filled - for example if the candidates had to choose between specific items to fill).

In the data, we have got information about candidates from all around the world, like gender, age, background (studies) but also about what they expect from the speed-dating meeting, i.e their goal. Finally, for each candidate we have the answers about questionnaires that were given to them, regarding their feelings about themselves, the attributes they put in their

scorecards about the candidate they met and feeling about the event. Most of this information is already encoded as numerical values. A smaller part of the variables are still qualitative.

We have 551 unique candidates for the whole dataset, 49.94% female and 50.06% male. Data have been gathered on 21 waves of speed-dating.