

GRAPHICAL MODELS TO REPRESENT CONSUMER BEHAVIOR



THESIS
PROJECT

ABOUT POINTLOGIC

Pointlogic, A Nielsen Company helps customers with decision making in the area of media and marketing. Our main assets are Nielsen data and advanced analytics capabilities.

Pointlogic's head office from Rotterdam has a data science team of about 20 people with backgrounds in econometrics, computer science and mathematics. We offer **data science students** the possibility to either to work solely on internal projects, or as a thesis project (as part of their MSc curriculum).

YOUR PROFILE

- Currently enrolled in a Master in Computer science, Econometrics, Mathematics;
- Good social and communication skills;
- Good command in spoken and written English;
- Available for at least 20 hours per week;
- Available for a period of at least 6 months.

CONTACT INFORMATION

If you are interested in this project or other projects, feel free to contact us at careers@pointlogic.com. We would like to find out if you are up for this challenge, so please include your CV and motivation letter, explaining why did you choose this assignment, what related courses you've been enrolled in and what similar work you've done around this specific topic during your education.

ABOUT THE PROJECT

The goal of this project is to represent a very high-dimensional data set with a graphical model (a Bayesian network or a Markov random field). The data is derived from a survey in which respondents voluntarily answer a large questionnaire. The questionnaire includes questions on leisure activities, purchase behaviors, demographics, social media and more (as illustrated below). The total number of variables obtained from the data is about 10,000. The yearly sample size for the questionnaire is about 50,000.

Working on this project, you will help us to reproduce not only the marginals of the multivariate distribution, but also the covariance structure. Examples of queries we would like the network to answer are:

- ▶ From those who are watching sports programs on TV, who is also visiting sport websites?
- ▶ What percentage of TV programs are people between 20 and 30 viewing online?
- ▶ What is the composition of people following a diet by age, gender and education? How does that vary across different type of diets?

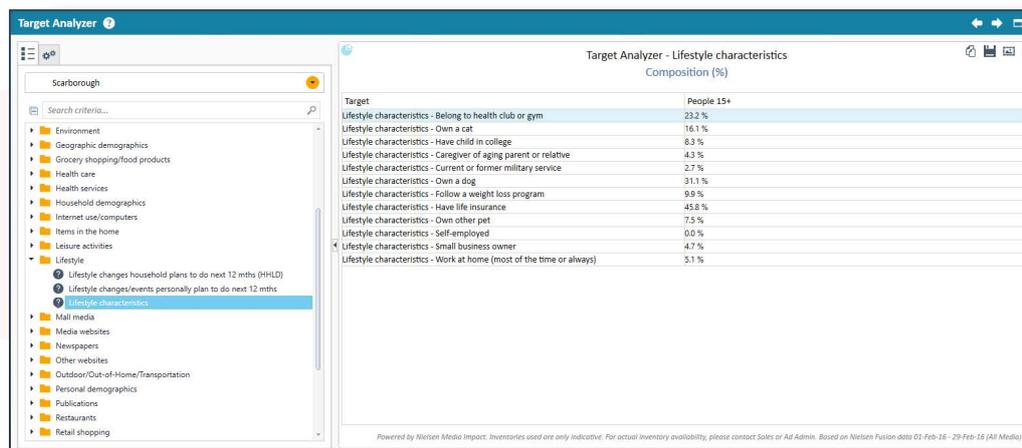


Figure 1. Lifestyle characteristics a measured in our survey. A graphical model can represent the relations between this variable and other variables measured in the survey.