

MANNHEIM MASTER OF APPLIED DATA SCIENCE & MEASUREMENT MASTER PROJECT PROPOSAL

Name of student(s) Deliverance Bougie Stefan Zimmermann
Name of supervisor and Email address. Please note: Your supervisor must be either a Professor or Honorary Professor, in accordance with §9 Paragraph 1 of the Exam Regulations. Prof Dr Florian Keusch, f.keusch@uni-mannheim.de Prof Dr Alexandru Cernat, alexandru.cernat@manchester.ac.uk (contacted and currently out of office)
Abstract <p>The Socio_Economic Panel (SOEP) is a longitudinal household survey that has been conducted throughout Germany since 1984. With its broad range of topics, such as employment, education, income, and attitudes, different questionnaires for each stage of life, household context, duration of the panel, and the many special sample groups (migrants, refugees, high income households, East Germans) comparisons over time can be presented with great accuracy using SOEP data.</p> <p>Our project aims to utilize this inventory of social science data by creating a platform that will put research results at the fingertips of non-scientists, such as journalists and modern historians. As this platform will be widely accessible, we hope to establish a new culture of using scientific results in non-scientific areas of society which do not have direct access to SOEP data. The proposed master project would establish a transfer interface utilizing R Shiny with aggregated data that non-scientists can use to answer current societal challenges outside the social science discipline.</p>
What is the (working) title of the Master project proposed? Development of a Data Visualization Platform for Non-scientists with SOEP Data.
What is the key question? How to create an interactive interface that provides information (basic statistical indicators and visualization) to the non-scientific community in a user-friendly way, making simple analyses possible for the public.

Which data source(s) do you use to answer this question?

The [Socio-Economic Panel \(SOEP\)](#) is one of the largest and longest-running multidisciplinary panel studies in the world and has been providing cross-sectional and longitudinal data [exclusively to researchers](#). In the age of „Fake News,“ there are data sources like SOEP that can answer many social, economic, and societal questions to educate society. However, access to the valuable information SOEP microdata can provide is only available to researchers and for teaching purposes.

Is the data you use already available? If so, where? If not, which data collection methods do you apply?

We would need to apply for data access with SOEP. Since Deliverance and I are part of a master's thesis/master project at an university, we need a project-based data sharing agreement between the university and SOEP:

https://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.88926.de/soep_application_contract.583953.pdf

This is how this process works:

https://www.diw.de/en/diw_01.c.601584.en/data_access.html#c_741347

The responsible project head will have to fill out the required document and request the data. We would like to have 100 percent of the SOEP EU v36 edition.

Which methods of data preparation, analysis, and evaluation do you plan to apply?

The data preparation will be conducted in Stata. This will include selecting the necessary variables, renaming variables, joining datasets, dealing with missing data, measuring data quality.

Aggregated tables will be generated after the data preparation. These tables will include simple statistics, such as counts, mean, median, confidence intervals, and/or proportions for each of the variables.

The aggregated tables will be used to create an interactive dashboard. This interface will allow the user to select variables for which they wish to see simple statistics, regressions, and visualizations.

Which skills do your team members need to work on this project?

- Knowledge of statistical localization and dispersion parameters
- Programming skills in data preparation in R and Stata
- Understanding about metadata and dataset documentation
- Programming skills in R Shiny to build an interactive data platform
- Knowledge and programming skills in the area of data visualization

What are the general tasks of each team mate?

Since the project is very ambitious and requires a lot of programming, all project members will have to work on these areas. However, splitting the work packages as soon as the idea is clearly defined is possible at any time.

- Data preparation in Stata
- Generation of aggregated tables in csv format
- Processing of aggregated tables in plotly graphs (R package)
- Creating the user interface in R Shiny