Data Analysis

Research Project

The analysis should be done based on the data collected by students or downloaded from open sources. The dataset should contain not less than 10 variables with different scales of measurement and not less than 100 cases. If needed, preliminary data transformation could be done using recoding, calculation of new variables based on existing variables, filters, aggregation, ranking, etc.

The data should be analyzed using descriptive statistics (frequency analysis, statistical characteristics, graphs), analysis of relationship between variables (crosstabs, Chi-squared test, correlations), regression analysis, PCA and cluster analysis. The report must contain at least one regression model. If certain data analysis methods are not applicable to analyze the data, it should be explained why they are not applicable.

The report should be completed in Jupyter Notebook. It should contain the research questions and hypotheses as well as the main results of the quantitative data analysis. The report could be prepared individually or in the group of 2 students. The main results of analysis should be presented on the slides (not less than 10 slides). In December each group will present the report.