Homework 10: group G5 report

Introduction to Data Science (LTAT.02.002)

Sadu Maria Jaarman, Kristina Kivistik, Laura Lall

Task 2. Business understanding

1. Identifying your business goals

1.1 Background

Human values play a foundational role in shaping individual perspectives and collective societal trends, influencing everything from personal decision-making to broader societal trends. In Estonia, a country shaped by its history and evolving democratic system, understanding values is also key to understanding what drives political behavior. Therefore, this project seeks to explore the connection between human values and political self-placement on the left-right scale among Estonian participants using European Social Survery (ESS) round 10 data.

In 2019, the Development Monitoring Center (Arenguseire Keskus) conducted a study on the values of Estonians, exploring how these values influence societal development.[1] One finding of the study highlights that while sociodemographic characteristics are linked to people's values, it is these values that play a key role in shaping choices in political preferences (Strenze, 2019: 119).

While the Development Monitoring Center also examined values and political self-determination in its report, their analysis focused specifically on respondents' attitudes towards the socio-political issues of inequality, same-sex rights and immigration (Strenze, 2019: 117). Our project takes a broader approach to human values, examining individuals' self-reported subjective human values and their perceived placement on the left-right political spectrum.

[1] Ainsaar, M., & Strenze, T. (2019). Väärtused kui inimvara ja nende seos ühiskonna arenguga. *Tallinn, Tartu: Arenguseire Keskus, Tartu Ülikool*.

Exploring the link between personal values and political alignment offers valuable insights into the factors shaping political behavior, which can contribute to addressing challenges such as polarization.

1.2 Business goals

The primary objectives include identifying key human values that influence individuals' position on the political left-right spectrum, understanding the differences between far-left and far-right perspectives, and uncovering potential areas of shared values. By analyzing these correlations, the project seeks to uncover the societal factors underlying political orientation, contributing to deeper insights into Estonia's sociopolitical dynamics.

1.3 Business success criteria

Success will be determined by the ability to generate clear insights about the key human values shaping individuals' political orientations on the left-right spectrum. This includes mapping and examining the distinct patterns of values that set far-right and far-left participants apart, as well as highlighting areas where common or overlapping values emerge across political divides.

2. Assessing your situation

2.1 Inventory of resources

The project is conducted using Python as the primary tool, utilizing access to ESS 10 dataset alongside data visualization and statistical analysis libraries (such as pandas, numpy etc).

2.2 Requirements, assumptions, and constraints

Our project does not have any requirements in the area of legal / security obligations, as the data we will be using is from an open access source (ESS) and is anonymized. The main requirement for us is by the deadline of the project, on the basis of our research problem, to have analyzed the data and drawn accurate conclusions for it in a fair and appropriate way.

When working with our data, we will be operating under the assumption that at least the majority of the respondents are familiar with the concept of the left-right political scale and position themselves, to at least some extent, on the basis of their political views instead of giving random answers.

The main constraint of our project is the subjectivity of the main variable, self-perceived political placement – what respondents perceive their political placement as might not be in actual accordance with how all of their political views would really position on the ideological scale. However, this isn't a major concern, as through including the aspect of self-perception in the interpretation of our results, there should be no obstacle to our analysis of the data yielding actual relevant conclusions.

2.3 Risks and contingencies

Risks	Contingencies
a. Difficulty finding meaningful clusters due to overlapping or ambiguous value patterns among participants	a. Use various different clustering techniques and validate results with statistical methods (eg. t-test)
b. Limited experience with Python and data mining tools may slow down progress and/or introduce unwanted errors	b. Use the knowledge and know-how gained throughout the course and online resources to resolve technical difficulties
c. Tight timeline could affect the quality of the project	c. Break the project into smaller tasks with clear deadlines

2.4 Terminology

Here are the definitions of two central terms of our project.

- political alignment - people's subjective self-definition on the political left-right scale.
- human values - the deeply rooted, abstract orientations that help to guide, justify or explain people's opinions, attitudes and actions.

2.5 Costs and benefits

Costs	Benefits
a. There are no costs associated with this study, as the required data is freely available for non-commercial use through the ESS Data Portal, and the analytical knowledge is provided as part of this course	a. The project could inform policy makers about societal polarization and ways to address it. b. The analysis of human values and political alignment can contribute to ongoing academic discussions and research.

3. Defining your data-mining goals

3.1 Data-mining goals

The primary goal of data mining is to uncover patterns and correlations between human values and political self-placement among Estonian participants in the ESS 10 dataset. This involves employing clustering techniques to participants based on shared values and political orientations, identifying key distinguishing factors between far-left, centrist, and far-right groups, and exploring areas of overlap that transcend political divides.

3.2 Data-mining success criteria

Success will be measured by the ability to extract meaningful insights using clustering methods to group participants effectively and identify statistically significant relationships between human values and political orientation. Success lies in defining clear clusters and understanding their unique characteristics.

Task 3. Data understanding

1. Gathering data

1.1 Outline data requirements

Our project requires a dataset which has a large enough sample size to guarantee the statistical relevance of our results, secondly, we have a duty to ensure that the participants are not too similar in terms of demographic variables (e.g. avoiding a sample where 75% of participants are over the age of 65) to guarantee a balanced dataset.

1.2 Verify data availability

The data required for our project is freely available on the open access ESS Data Portal, with the participant responses of ESS round 10 covering all necessary variables for the execution of our project – the possibility of only accessing the data collected from Estonian participants; participants' answers to their placement on the left-right political scale, and their opinions on different statements regarding human values.

1.3 Define selection criteria

For our project, we will use the ESS open access database. From there, using the Data Wizard tool, we will select a subset of the most recent round of survey that covers all aspects needed to fulfil the aims of our project. In the context of our project, that means the subset needs to contain answers of Estonian participants and their answers to questions on their political placement and how their human values manifest.

2. Describing data

The dataset is from The European Social Survey in the form of a CSV file. The original round 10 of ESS was 55.6 MB, which contained data from 31 countries and 14 different variable categories, which each contained tens of different variables. However, since our focus is on Estonia and the human values of its people, we have created a sub-dataset that exclusively includes relevant data about human values, political alignment specific to Estonia. Therefore, the dataset we are working with contains 1542 rows and 37 variables (listed below). The variable descriptions are sourced directly from the ESS website[2].

Identifier and weight variables (statistical weights for analysis):

- name - identifies the title of the used dataset
- essround - ESS round (survey round)
- edition - edition number
- proddate - production date of the dataset
- idno - respondent's identification number
- cntry - country
- dweight - design weight
- pweight - population size weight
- anweight - analysis weight

Political alignment:

- lrscale - placement on the left-right scale

Gender and age:

- gndr - gender
- agea - age of the respondent

[2] European Social Survey European Research Infrastructure (ESS ERIC) (2023) ESS10 - integrated file, edition 3.2 [Data set]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/ess10e03_2.

Human values:

- imprich - important to be rich, have money and expensive things
- ipeqopt - important that people are treated equally and have equal opportunities
- ipshabt - important to show abilities and be admired
- impsafe - important to live in secure and safe surroundings
- impdiff - important to try new and different things in life
- ipfrule - important to do what is told and follow rules
- ipudrst - important to understand different people
- ipmodst - important to be humble and modest, not draw attention
- ipgdtim - important to have a good time
- impfree - important to make own decisions and be free
- iphlppl - important to help people and care for others well-being
- ipsuces - important to be successful and that people recognise achievements
- ipstrgv - important that government is strong and ensures safety
- ipadvnt - important to seek adventures and have an exciting life
- ipbhprp - important to behave properly
- iprspot - important to get respect from others
- iplylfr - important to be loyal friends and devote to people close
- impenv - important to care for nature and environment
- imptrad - important to follow traditions and customs
- impfun - important to seek fun and things that give pleasure

Sample Design variables:

- prob - sampling probability
- stratum - sampling stratum
- psu - primary sampling unit

The data includes all the fields that we expect and that need to be there. The dataset includes the positioning of the individuals on the left-right scale, their age, gender and their opinions on the human values they possess.

3. Exploring data

Below we provide an overview of the values and the value ranges.

- Political alignment (lrscale) - This indicates respondents' self-placement on the political spectrum. Ranges from 0 to 10, 0 being far-left and 10 being far-right, with additional values for 77 (Refusal), 88 (Don't Know), and 99 (No Answer).
- Gender (gndr) - Gender of the respondent. 1- male, 2 - female, 9- no answer.
- Age (agea) - Age of the respondent. 999 - unavailable.
- Human values (variables from imprich to impfun) - Measures respondents' alignment with various human values. Ranges from 1 to 6, 1 - very much like me, 2 - like me, 3- somewhat like me, 4 - a little like me, 5 - not like me, 6 - not like me at all, with additional values for 7 -refusal, 8- don't know, 9 - no answer.
- Sample Design variables (prob, stratum, psu) - used for sampling validation.

In order to accurately analyze the data, lrscale values 77, 88 and 99 should be treated as missing values and not taken into account when analyzing the correspondence of political alignment with human values. Furthermore, values 7, 8, 9 in the human values variables and age of 999 should be treated the same.

4. Verifying data quality

The data is from ESS round 10, which was conducted in 2020, so the data is fairly recent. ESS is also one of the most popular and reliable cross-national data sources for the mapping of the attitudes and demographic data of the inhabitants of the European Union.

Task 4. Planning your project

1. Data preparation and project planning

The data preparation task consists of removing missing values, unnecessary variables etc from our data in order to transform it into a suitable dataset to be used for analysis. Project planning consists of all the tasks necessary for this report.

- Division of time: 6 hours per team member.

2. Exploring data

To explore the data, we will be using data visualization tools to uncover patterns, anomalies, and potential relationships within the dataset.

- Division of time: 6 hours per team member.

3. Clustering and statistical analysis

To perform the task of clustering and statistical analysis, we will be using clustering techniques to group participants based on shared values and political orientations and performing statistical tests to identify significant factors influencing political placement.

- Division of time: 8 hours per team member.

4. Interpretation of results

To interpret our results, we will analyze clustering results to identify distinct groups (far-left, centrist, far-right) and their defining characteristics in the area of human values.

- Division of time: 5 hours per team member.

5. Reporting

As the final part of our project, we will prepare to present our findings. For that, we will compile our key findings into a poster.

- Division of time: 5 hours per team member.