

COM-480: Milestone 1

Jérémy Baffou, Camille Pittet, Renuka Singh Virk

March 21, 2025

1 Problematic

"What am I doing here?" This question necessarily popped into your head you at least once in your life. May it be while doing the queue at Migros with crying children, or in the middle of an exam. Switzerland is a wonderful country, yet for some reasons, we cannot help but dream of distant lands that may revealed to be the perfect place. The sunny summers along the lake tend to push these thoughts away. But the harsh reality of exercises series in the Rolex during the long winter nights bring them rushing back. *"Why should I stay in the cold and the fog when I could be in Bangkok, New-York or Acapulco!"*. Once the excitement gone, reality usually sets back in: The world is vast. Leaving Switzerland is one thing, but the true challenge remains. How to find the perfect land for you? How can you be sure to be aligned with the culture, the people and the values in a given country? Is the grass genuinely greener on the other side? Well, we've heard your thoughts and they sounded just like a mission for the Data-Vizards! With lots of visualizations, and a bit of our magic, we will build the dream tool to help you discover your country-soulmate.

In more technical terms, the aim of this project will be to perform a world-wide analysis of the opinion of people on a large set of key topics, from ethical values to attitudes towards science and technology. To provide an engaging and personalized experience, users will answer a series of questions to generate their "avatar", a simple representation of their values. We will then identify the countries that are the most aligned with their profile. Furthermore, additional analysis will create a mapping of the different values through the countries. This interactive feature will let users explore and understand the global landscape of cultural perceptions, ultimately helping them uncover the country of their dream.

2 Dataset

2.1 General

Our project will leverage data from the World Values Survey database [1]. This database comprises multiple survey waves conducted in nearly 70 countries since 1980, covering a wide range of subjects, from social values to political interests and security. Our analysis will specifically focus on the latest available survey wave, covering the period from 2017 to 2022. If time permits, we may also investigate trends and changes in responses over time. Given the extensive set of questions available in the latest wave (294 questions), we have decided to narrow our scope and concentrate on a selected subset of topics:

- *Demographics.*
- *Ethical values and norms.*
- *Social capital, trust and organizational membership.*
- *Social values, attitudes and stereotypes.*

As we progress in the processing of the data, we may integrate other sections to have a more global picture of the panel and to support our idea presented in section 1.

The dataset is very qualitative and has been used in a variety of studies [2],[3],[4]. We performed a preliminary data analysis that is presented in 2.2.

2.2 Preprocessing

The data was collected in each country between 2017 and 2022. We will assume this time period to be approximately stable and that no drastic changes have occurred within this time. All countries will be kept to provide a worldwide analysis of the values across countries. One of the most challenging part will be selecting which questions we want to keep for our visualization, while avoiding inducing bias. We exclude questions that have too few respondents in at least one country. We will avoid value imputation as the data is already complicated and the risk of adding noise or artifacts to the data is high.

3 Exploratory Data Analysis

3.1 Missing Values



Figure 1: Missing Values

The first step of our preprocessing pipeline is assessing the distribution of missing values. Several features exhibit missing data (see Fig 1). The number of missing values is consistently either 60,000 or 90,000, a pattern that appears due to certain questions being specific to a single country, thus being answered only by a subset of participants. To proceed with the analysis, these questions are excluded, which drops 31% of the features.

3.2 Feature Type

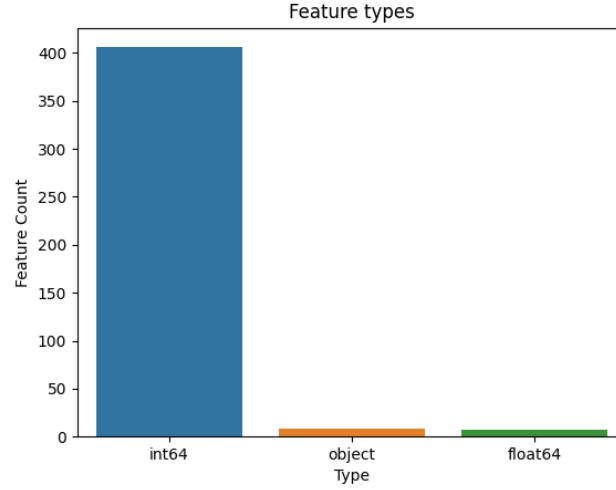


Figure 2: Feature Types

Even after filtering out features with missing values, the data still contains 421 features. To ensure they have the expected format, we look at the distribution of feature types (see Fig. 2). We observe that the major part of the data is of type int64. In fact, the WVS [1] (World Value Survey) collects data in the form of multiple choice questions, where each integer is mapped an answer. The answers are listed in a pdf file and can be accessed by parsing the document. The remaining data types are either information about the participant (e.g. age information as type float) or information about the survey wave (e.g. doi information as type object).

3.3 Country Diversity

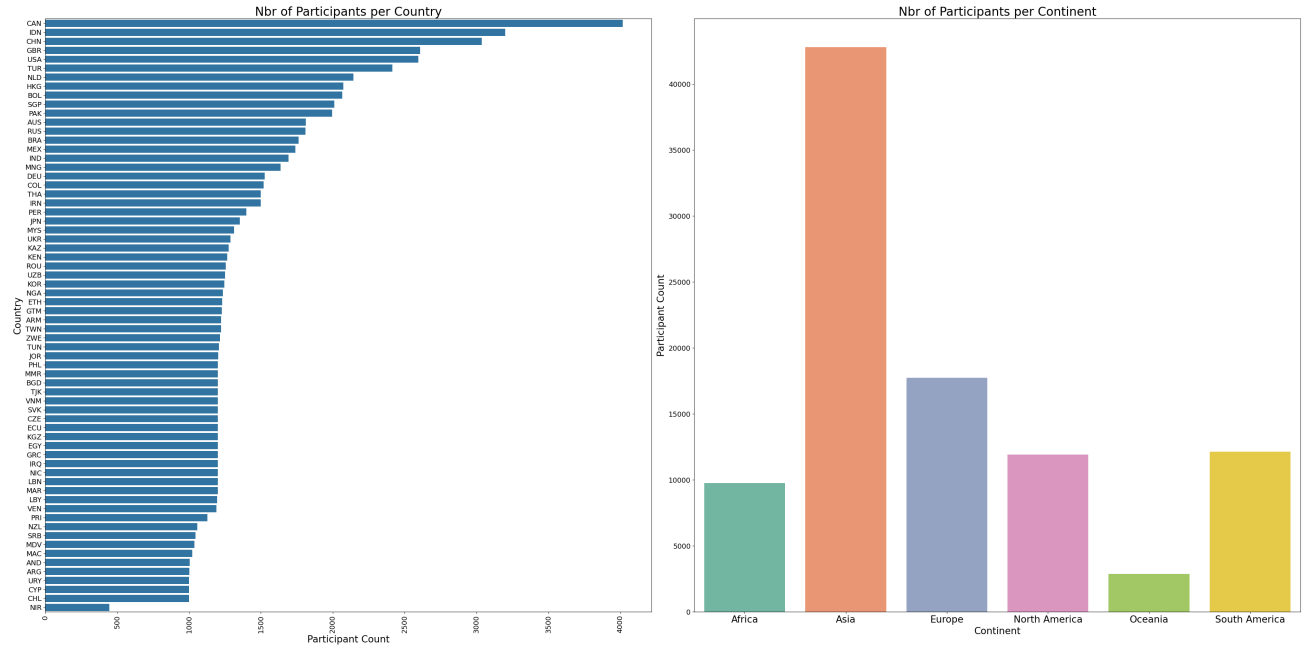


Figure 3: Participants per country and continent

This project aims to visualize various mindsets worldwide, thus ensuring country diversity is essential. We observe (see Fig. 3) that only 66 out of the 195 countries [5] are represented in our dataset, with participant numbers varying greatly across countries. Where more populated countries appear with more participants.

At a continental level, all five populated continents are included in the survey. The distribution of participants aligns with the relative population sizes of these continents, except for Africa which is notably underrepresented. The majority of respondents come from Asia, followed by Europe, South America, North America, Africa and finally Oceania.

3.4 Questions Diversity

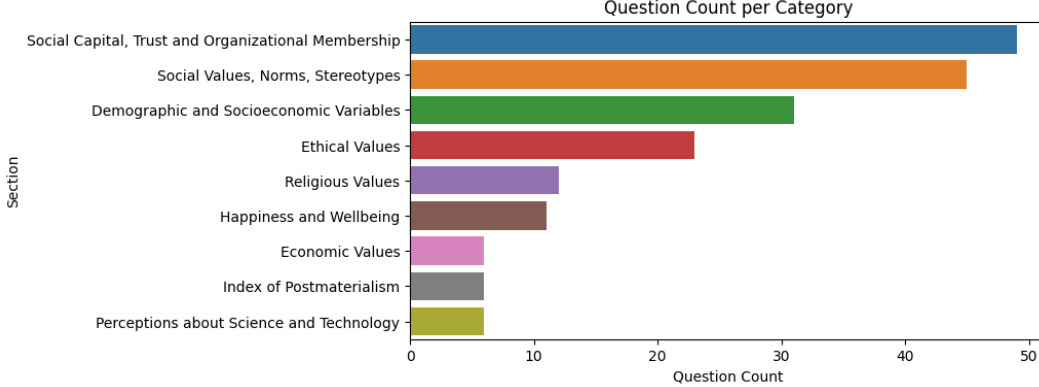


Figure 4: Question Count per Category

To determine the most suitable country for you, only the following survey sections are considered (see Fig. 3). While there are disparities in the number of questions across categories, this is not problematic since 7 questions per category are already sufficient to determine a potential match for you. To maximize the relevance of our recommendations, all questions will be carefully reviewed, retaining only the most meaningful ones. Additionally, subsections may be created to get more explicit and well-defined topics.

4 Related work

This dataset has been used extensively throughout the years due to its very high quality with around 1000 publications according to the [World Value Survey Network](#). On the academic side, most of the studies have found applications in human sciences, economics and psychology [2, 3, 6, 7]. Recent papers use the database to compare the behavior of different architectures of Large Language Models (LLMs) [4, 8]. Closer to our application a few blogs and articles provide insights on the database [9, 10, 11]. They often focus on the evolution of features over time or narrow their analysis to a very small set of values. The World Value Survey Network provides also a serie of [webinars](#) presenting interesting insights extracted from the database.

Our approach differs from existing one on several aspects. First, to the best our knowledge, the notion of an avatar as a projection of the user values has not been performed on this dataset. We thus provide a website with an engaging and personalized experience while still providing a comprehensive approach to the different features present in the dataset. Furthermore, we will implement an additional layer consisting of an interactive map to visualize the cultural differences over the world. This spatial visualization of several values across countries has also not yet been conducted in a readable way. We thus propose an interactive framework that first provide a personalized approach to the user, and then allow them to gain a deeper understanding of the datasets through the spatial map.

As for the sources of inspirations, we have listed a few useful ressources. Regarding the avatar section, we could implement something like [this](#). With a focus on the social and ethic values rather than on the physical aspect. For general purpose illustration, the visualizations in [11] are interesting with dynamic elements over time and representation of population size. Finally, the map section could

be introduced as a globe in 3d such as [there](#). Also flat maps can be interesting, [epsilon](#) make usually very nice visualizations.

References

- [1] C. Haerpfer, R. Inglehart, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin, and B. Puranen, “World values survey wave 7 (2017-2022) cross-national data-set,” 2024.
- [2] A. Tausch, “The civic culture of the arab world: A comparative analysis based on world values survey data,” *Middle East Review of International Affairs*, vol. 20, no. 1, pp. 35–59, 2016.
- [3] R. Inglehart, *Human beliefs and values: A cross-cultural sourcebook based on the 1999-2002 values surveys*. Siglo XXI, 2004.
- [4] C. Li, M. Chen, J. Wang, S. Sitaram, and X. Xie, “Culturellm: Incorporating cultural differences into large language models,” *Advances in Neural Information Processing Systems*, vol. 37, pp. 84799–84838, 2024.
- [5] T. Saunders, “How many countries are there in 2025?,” *BBC Science Focus*, 2025.
- [6] M. Minkov and G. Hofstede, “Hofstede’s fifth dimension: New evidence from the world values survey,” *Journal of cross-cultural psychology*, vol. 43, no. 1, pp. 3–14, 2012.
- [7] M. Minkov, *What makes us different and similar: A new interpretation of the World Values Survey and other cross-cultural data*. Klasika i Stil Publishing House Sofia, Bulgaria, 2007.
- [8] E. Durmus, K. Nguyen, T. I. Liao, N. Schiefer, A. Askell, A. Bakhtin, C. Chen, Z. Hatfield-Dodds, D. Hernandez, N. Joseph, *et al.*, “Towards measuring the representation of subjective global opinions in language models,” *arXiv preprint arXiv:2306.16388*, 2023.
- [9] “Western values are steadily diverging from the rest of the world’s,” *The Economist*, Aug 2023.
- [10] W. n. Institute for Futures Studies in Stockholm, “Wvs cultural map,” *Youtube*, Aug 2023.
- [11] P. Francesco, “Prejudice free,” -, Aug 2023.