

Datasheet for 'Income Inequality and Happiness'

Carl Fernandes

Invalid Date

This study examines the relationship between income inequality, as measured by the Gini Index, and national happiness scores. Incorporating other socio-economic factors such as the Democracy Index and Inequality-Adjusted Human Development Index, we explore the depth of this relationship. The study aims to inform policy by identifying the broader determinants of happiness beyond economic measures.

Extract of the questions from @gebru2021datasheets.

Motivation

1. *For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*
 - The dataset was curated to explore the hypothesis that income inequality, as quantified by the Gini Index, has a discernible impact on the perceived happiness within nations, taking into account factors like democratic governance and human development.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
 - Dataset was created by Carl Fernandes for University of Toronto
3. *Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*
 - N/A
4. *Any other comments?*
 - N/A

*Code and data are available at: <https://github.com/Carl-1506/Happiness-and-inequality.git>

Composition

1. What do the instances that comprise the dataset represent?

The dataset instances represent individual countries, with each instance encapsulating various socio-economic indicators for the year 2022.

2. How many instances are there in total?

There are 100 instances, each corresponding to a unique country.

3. Does the dataset contain all possible instances or is it a sample of instances from a larger set?

The dataset is a sample of countries from the year 2022, selected based on the availability of complete data for the variables of interest.

4. What data does each instance consist of?

Each instance consists of the country's name, the Gini Index, Democracy Index, Inequality-Adjusted Human Development Index (IHDI), and the corresponding Happiness Score for 2022.

5. Collection process

6. How was the data associated with each instance acquired? The data was acquired from publicly available databases, including the World Happiness Report, the World Bank, Our World in Data, and the UNDP.

7. What mechanisms or procedures were used to collect the data?

Data were collected through automated scripts pulling from the aforementioned databases' APIs or data download sections.

Preprocessing/cleaning/labeling

1. Was any preprocessing/cleaning/labeling of the data done?

Preprocessing involved cleaning for missing values, ensuring consistency across indicators, and standardizing country names.

Uses

1. Has the dataset been used for any tasks already?

The dataset has been used for an initial exploratory data analysis and regression modeling to ascertain the relationships between happiness and the selected socio-economic indicators.

Distribution

1. **How will the dataset be distributed?** The dataset, along with the code, will be available in a GitHub repository, with a README explaining the structure and use-cases.

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*
 - Github