DataInSociety_FromSrinivasan

Maria Mangru

Reflection

During world war II, the nazis used census data to identify Jewish populations. Hollerith machines, developed by IBM, were used to process and sort the information gathered across various cities. Data collected on the 'race' of people were used as markers to locate large portions of the Jewish population. As a result of this data collection, the Nazi regime was able to commit systematic targeting of Jewish communities across Germany. Although historical instances are enlightening, they also pave the way for understanding the relevance of data misuse in the modern era. More recently, data misuse has been used through systems like credit scoring that reinforce socioeconomic disparities. Typically, algorithms are used to assess credit scores based on financial history; these include credit card usage, borrower's past loan repayment and other financial habits. People from lower-income families are more likely to have limited access to traditional banking services, which contributes to shorter credit history. Because of the widespread use of credit scores in institutions, those with lower credit scores will face higher interest rates on loans and difficulty renting or purchasing a home. Although credit scoring is not intended to be oppressive, the results often resemble systematic discrimination. Often, credit scoring systems reflect inherent biases in our economic system.

After observing historical and contemporary instances of data exploitation, it becomes evident that the quality and depth of data significantly influence society's viewpoints. Reflecting on Amia Srinivasan's quote (Cowen 2021), we can characterize 'thin' data as lacking depth and diversity and 'weak' data as poor quality, outdated, or inaccurate. ' thin' data will often lead to overgeneralizations or conclusions that do not accurately represent a population. For example, in healthcare, samples that do not reflect a diverse group of people may fall short of solutions that are only effective for some. Moreover, 'weak' data, outdated or inaccurate, can have widespread consequences on public policy and urban planning. Chapter 6 of D'Ignazio and Klein's work outlines important ideas relating to Amia Srinivasan's quote. In particular, the chapter highlights the necessity of context when analyzing data. The author emphasizes that data are not neutral but the result of uneven social relations. These ideas relate to Srinivasan's quote, which underscores that inadequate statistics should be avoided as they could reinforce social prejudices. For example, FiveThirtyEight's misreading of kidnapping statistics displays

the dangers of interpreting data without proper context. The chapter emphasizes the need for a more nuanced view of data, acknowledging its inherent biases and the necessity of context.

Furthermore, the convergence of Srinivasans' cautions against flimsy data that supports oppressive beliefs and Chapter 6 underscores the need for education and awareness of data literacy. By promoting data literacy, we may better traverse and confront the historical and context biases that influence how we interpret data. Doing this ensures better moral and knowledgeable use of data by addressing the issues and misuses that D'Ignazio, Klein, and Srinivasan have pointed out.

https://encyclopedia.ushmm.org/content/en/article/locating-the-victims
https://www.jstor.org/stable/2808153#:~:text=URL%3A%20https%3A%2F%2Fwww.jstor.org%2Fstable%2
https://www.marketplace.org/shows/marketplace-tech/credit-scores-and-the-bias-behind-them/

Relation To Dataset

Using these ideas, we will map them to a dataset of our choice: the Toronto Shelter System Flow dataset. This data set provides insights into the demographics and movement of homeless individuals utilizing Toronto's shelter system. The dataset records its data using a Shelter Management Information System (SMIS) through self-reporting. The categories of data collected are age groups, gender, and population groups (chronic homelessness, refugees, families, youth). The data also contains information on active homeless individuals, those who have used the shelter system in the last three months but have not moved to permanent housing. However, the data has limitations; for instance, it does not consider those who sleep outdoors or use non-city-funded homelessness services. Moreover, the dataset's limited scope could lead to policies that do not adequately address the abovementioned needs.

Fundamental ideas from Chapter 6 also apply to the Toronto Shelter System Flow dataset, emphasizing the value of contextual understanding in data analysis. Without context, the dataset's numbers, for example, might not fairly represent the complexities and subtleties of homelessness, such as why people choose not to use shelter services. According to the chapter, data should always be linked to the social environment in which they were created because of uneven social relations. This is especially crucial when examining the Toronto Shelter System Flow dataset since a comprehensive understanding of homelessness's economic, sociological, and political aspects must be obtained. In light of these factors, it is possible to read the dataset's depiction of homelessness as a mirror of the more significant structural and societal problems brought up by D'Ignazio and Klein as well as Srinivasan, requiring a more sophisticated approach to data interpretation and policy development.