

Datasheet for Toronto Neighbourhood Profiles Data*

A deeper look into the 2021 Canadian Census Data

Julia Lee

December 3, 2024

The following contains information about the Toronto Neighbourhoods Profile Data. Extract of the questions are provided by Gebru and colleagues (2021, <https://doi.org/10.1145/3458723>)

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 Neighbourhood Profiles dataset was created to gain a better understanding of the populations living in the city of Toronto. By highlighting various demographic and socio-economic characteristics such as education, income, and employment of Toronto residents, this data can be used to understand how Toronto is changing and to better serve all Toronto residents.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
 - This dataset is a collection of data from the 2021 Canadian Census conducted and provided by Statistics Canada. To organize population information for Toronto specifically, the Social Development, Finance & Administration of Open Data Toronto created this dataset.
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.*
 - The creation of the dataset is funded by the city of Toronto, and it is made accessible to everyone through the city's open data portal.

*Code and data are available at: https://github.com/jjlee-lee/Toronto_Neighbourhood_Income.git

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*
 - The instances that comprise the dataset represent Toronto residents and the various demographic and socio-economic characteristics they had at the time of the 2021 Census.
2. *How many instances are there in total (of each type, if appropriate)?*
 - There are over 2000 categories (e.g. education attainment, household size, etc.) of data provided for each of the 158 neighbourhoods in Toronto.
3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*
 - The short-form census is sent out to all households in a given region, and the long-form census is sent out to only 25% of households in a given region. With this, the 2021 Census data reflected in the Neighbourhood Profiles data can be a representative sample of all residents in a specific region.
4. *What data does each instance consist of? “Raw” data (for example, unprocessed text or images) or features? In either case, please provide a description.*
 - Each instance consists of unprocessed text in an excel file.
5. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*
 - Missing data values can be imputed and the aggregation of individual household characteristics or experiences can lead to incorrect assumptions about certain regions and their populations.