

Datasheet for ‘Labour force characteristics by sex and detailed age group, annual’*

2019-2023

Fatimah Yunusa

April 16, 2024

This Labour Force Survey (LFS) datasheet provides information about Canada’s labor market, offering data on employment trends, demographics, and key indicators such as occupation and earnings. Conducted monthly by Statistics Canada, this dataset is a valuable resource for researchers, policymakers, and businesses seeking to understand and navigate the dynamics of the Canadian workforce.

Extract of the questions from @gebru2021datasheets.

The information provide below is sourced from: <https://www.statcan.gc.ca/en/reference/refcentre/index>

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 survey aimed to produce information about unemployment rates and other labour market indicators within the Canadian economy. This survey serves the purpose of providing data to several Canadian Government levels to help them evaluate and plan for different employment programmes in Canada.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
 - Statistics Canada
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.*

*Code and data are available at: <https://github.com/fatimahsy/Gender-Unemployment-.git>.

- This dataset was funded by the Canadian Government. Statistics Canada, as the national statistical agency, operates under the mandate of the federal government, receiving its funding primarily through the federal budget.
4. *Any other comments?*
- No

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 primarily represent individuals within households across the country. Each instance corresponds to a person who is part of the surveyed population, providing data on various aspects of their employment status, including whether they are employed, unemployed, and not in the labour force. The survey covers demographic details such as age, gender, and education, along with job-related information such as occupation, industry, and working hours.
2. *How many instances are there in total (of each type, if appropriate)?*
 - In total, approximately 100,000 individuals are surveyed.
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 LSF is a sample of the people in Canada that are non-institutionalized population aged 15 and over. This larger set includes all people who are either employed, unemployed, or not in the labor force, excluding those in the military or institutional residents. The LFS uses a stratified random sampling design. This method involves dividing the entire population into distinct subgroups (strata) based on characteristics like geography and then randomly selecting samples from each stratum.
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 in the Labour Force Survey (LFS) of Canada consists of structured and processed data, not raw data like unprocessed text or images. The data collected is in the form of features derived from responses to a structured questionnaire. These features include both quantitative and categorical data. It includes

demographic information, employment details, economic information, geographic information.

5. *Is there a label or target associated with each instance? If so, please provide a description.*
 - yes, Employed: Individuals who performed work for pay or profit during the reference week (the week preceding the survey) or who had a job but were temporarily absent from work. Unemployed: Individuals who were not employed during the reference week but were actively looking for work and were available for work. Not in the Labour Force: Individuals who were neither employed nor unemployed during the reference week. This category includes students, retirees, homemakers, and individuals who are not actively seeking employment.
6. *Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*
 - There is no missing information.
7. *Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.*
 - In the Labour Force Survey (LFS) of Canada, relationships between individual instances are not made explicit within the dataset.
8. *Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.*
 - The dataset is provided as a single continuous time series and can be split up if needed.
9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*
 - The dataset is clean and does not contain any errors, noise, or redundancy.
10. *Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.*

- The dataset is highly self contained and does not require any external resources.
11. *Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.*
 - No, all the participants have provided permission for their data to be used.
 12. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*
 - No
 13. *Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*
 - Yes, the Labour Force Survey (LFS) of Canada identifies and captures data on various sub-populations based on demographic characteristics such as age, gender, education level, and geographic location. These subpopulations are crucial for understanding labor market dynamics and disparities across different groups within the population.
 14. *Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.*
 - No it is not.
 15. *Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.*
 - The Labour Force Survey (LFS) of Canada does not contain data that might be considered highly sensitive in nature. However, it does collect demographic information and employment-related data that, while not inherently sensitive, could still be considered private or personal.
 16. *Any other comments?*
 - No

Collection process

1. *How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey*

responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.

- LFS interviews are conducted by telephone in English or French by interviewers working out of a regional office CATI (Computer Assisted Telephone Interview) site or by personal visit from a field interviewer.
2. *What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?*
- Normal administration software, and validation checks.
3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*
- The Labour Force Survey (LFS) of Canada employs a probabilistic sampling strategy to select a representative sample of households from the larger population. This sampling strategy involves the use of probability sampling methods to ensure that every household in the target population has a known and non-zero chance of being selected for inclusion in the survey.
4. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?*
- Not mentioned.
5. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.*
- Collected continuously through the year.
6. *Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.*
- No.
7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*
- The dataset was collected directly from Statistics Canada.

8. *Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.*

- Yes, the survey is optional so agreeing to do the survey is seen as consent.

9. *Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.*

- explicit consent from individuals is not typically obtained in the same manner as in academic research studies. Instead, respondents are informed about the purpose of the survey, the voluntary nature of participation, and how their information will be used and protected.

10. *If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).*

- Yes

11. *Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.*

- No it has not.

12. *Any other comments?*

- No

Preprocessing/cleaning/labeling

1. *Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.*

- NO

2. *Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the “raw” data.*

- No

3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*

- No

4. *Any other comments?*

- No

Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*

- Yes, the Labour Force Survey (LFS) of Canada dataset has been extensively used for various tasks and analyses by researchers, policymakers, economists, and other stakeholders.

2. *Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.*

- No

3. *What (other) tasks could the dataset be used for?*

- Analysis about the health of the economy or analysis about the competence of the current government.

4. *Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (for example, stereotyping, quality of service issues) or other risks or harms (for example, legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?*

- No

5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*

- No

6. *Any other comments?*

- No

Distribution

1. *Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.*

- It permits unrestricted use, provided that the original work is cited.
2. *How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?*
 - On the Statistics Canada Website.
 3. *When will the dataset be distributed?*
 - TBD
 4. *Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/ or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.*
 - Not mentioned.
 5. *Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.*
 - Not specified.
 6. *Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.*
 - No
 7. *Any other comments?*
 - No

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*
 - Statistics Canada
2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*
 - Email infostats@statcan.gc.ca Telephone (toll free) 1-800-263-1136 (international) 1-514-283-8300
3. *Is there an erratum? If so, please provide a link or other access point.*
 - No

4. *Will the dataset be updated (for example, to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (for example, mailing list, GitHub)?*
 - Yes, it is updated yearly.
5. *If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.*
 - All info gathered was anonymous.
6. *Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.*
 - Yes.
7. *If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.*
 - No
8. *Any other comments?*
 - No