The Impact of Gender on Canadian Unemployment*

2018 to 2023

Fatimah Yunusa

April 18, 2024

This paper examines the impact of gender on unemployment rates across various demographics from 2018 to 2023, employing a comprehensive dataset from the Labor Force Survey to unravel the interplay between gender and overall economic participation.We also use a linear regression model to help disect the influence of gender, among other predictors, on the unemployment rates of Generatinon Z. Our findings reveal significant disparities in unemployment trends, with distinct patterns emerging between different genders and age groups, hilighting the persistent gender gap in the labour market. This reaserch is important because it shows the underlying factors that contribute to unemployment disparities, offering insights into the effectiveness of current policies and the need for targeted interventions.This paper enhances our understanding of the labour market's complexities, advocating for inclusinve economic strategies to bridge the gender divide in employment opportunities.

Table of contents

1 Introduction **2**

2 Data **3**

2.1 Data Analysis 3
2.2 Data source & Features 4
2.3 Data Measurement & Limitations 5
2.4 Cleaning and Preparation 6

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

3 Model **7**
3.1 Model Set-up . . . 7
3.2 Model Justification . . . 8

4 Results **8**
4.1 Model Results . . . 8

5 Discussion **9**
5.1 Implications of these results . . . 9
5.2 Possible Causes . . . 9
5.2.1 Sector Vulnerability . . . 9
5.2.2 Digital & Remote Work Readiness . . . 10
5.2.3 Educational Disruptions . . . 10
5.2.4 Global Economic Slowdown . . . 10
5.3 Weaknesses and next steps . . . 10

6 Conclusion **11**

A Appendix **11**
A.1 Datasheet for Labour Force Survey . . . 11

References **20**

1 Introduction

In todays rapidly changing economy, understanding the dynamics of unemployment through the perspective of gender provides crucial insights into the structural inequalities that exist in the labour market. There has always been a persistent gender gap in unemployment rates as a result of normalised gender roles in society. However, in the recent decade, their is an increasing decrease in the size of the gender gap. This can largely be attributed to increased educational attainment among women, changing gender norms, increased childcare support and flexibility, economic shift from manufacturing to service sector, and corporate inclusion efforts. Globally, finding a job as a woman is much harder and when women finally get jobs, they statistically work in low quality jobs (Organization (2017)).

By examining the interplay between gender and unemployment rates, this paper aims to uncover the underlying patterns that continue to disadvantage one gender over another, in order to offer more efficient and equitable economic policies that decrease the unemployment gap. A Addressing this gap not only positively contributes to economic theory but also has practical implications for improving workforce inclusivity and economic resilience.

This paper leverages a very comprehensive dataset derived from the `Labour Force Survey` (Canada 2024b) spanning from 2018 to 2023, a period that is notable for its socio-economic turbulence and technological disruptions because of Covid-19 and the increase in artificial intelligence.

Current research in this area has an extensive gap that needs to be filled because despite extensive research in the area, the interplay between gender, age, and unemployment within the specific context of Generation Z remains inadequately explored.

To address this gap, we employ the use of a linear regression model to scrutinize the influence of gender alongside other predictors of unemployment rates. The estimand in our study is the differential impact of gender on the likelihood of unemployment for individuals within the Generation Z demographic, holding other variables constant.

Our statistical analysis shows that employment trends exhibit significant disparities. These findings affirm the existence of a gender gap but also highlight its persistence and complexity within the labour force.

There are many implications of this research.Understanding the underlying factors of the gender gap in unemployment allows us to evaluate the effectiveness of current Canadian labour force policies. Ultimately, this paper contributes to discussions about labour economics in Canada by advocating for more inclusive policies that aim at decreasing the unemployment gap.

The paper is structured as follows: Following this introduction, we describe the methodology and data sources used in the analysis in (Section 2). (Section 3)describes the model utilized in this paper, and results (Section 4) relays the results of our statistical analysis. The paper concludes with a discussion (Section 5) of the implications of the results, possible causes, weaknesses and next steps.

2 Data

2.1 Data Analysis

Table 1: Unemployed Data from 2019 to 2023 by sex

Reference_Period	GenZ_Unemployment	Adult_Unemployment	Seniors_Unemployment
2018	21.3	10.2	9.8
2019	21.4	10.2	9.1
2020	40.1	16.9	14.9
2021	27.0	12.9	13.3
2022	20.1	9.1	8.9
2023	21.5	9.7	8.3

Table 1 shows a yearly comparison of unemployment rates segmented by age groups, showing the general state of the labour market in relation to Genz, adult and senior unemployment in Canada from 2018-2023. From this, we can see an improvement in the employment landscape.

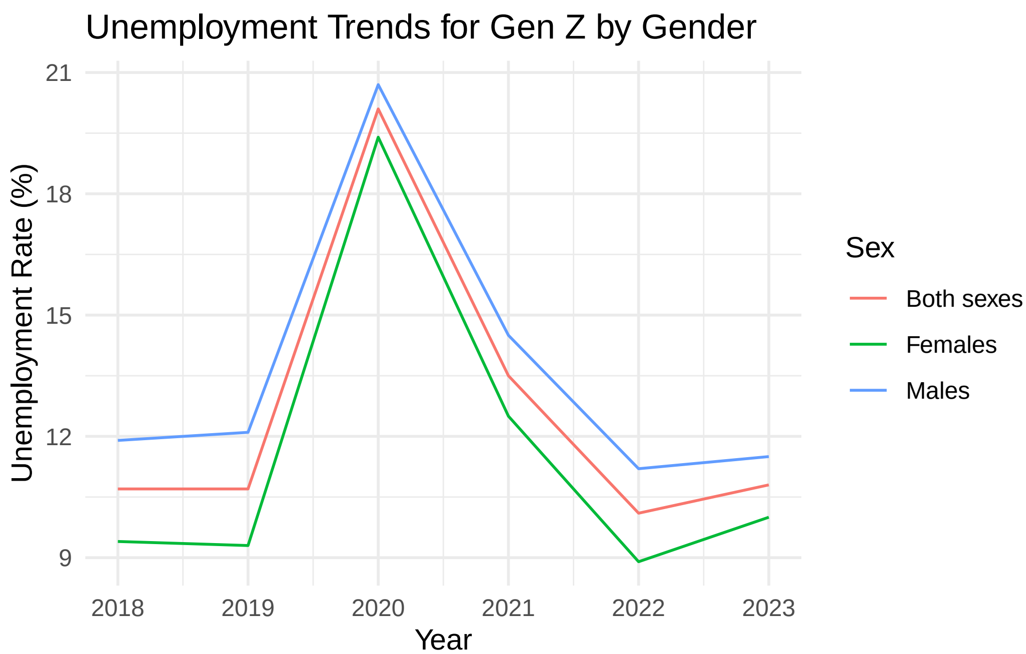

Figure 1: Total Unemployment Rates by Age Group for All Genders

Figure 1 presents a visual representation of unemployment rates from 2013 to 2023, broken down by gender.A peak is observed in 2020 for all categories which is attributed to the COVID-19 pandemic. Post 2020, the labour force unemployment rates seem to trend downwards, suggesting a recovery phase in the job market.

2.2 Data source & Features

The primary and only dataset used in this paper was sourced from the `Labour Force Survey` (Canada 2024b) that was conducted by the national statistics office of Canada (Statistics Canada). This annual survey captures captures a wide range of information relating to the employment status of Canadians across various demographics, including age, gender,education level, and employment status. The LFS is designed to give a comprehensive snapshot of the workforce and has been a critical source of the data used by policymakers to analyze market conditions.

To carry out our analysis, we have extracted data from the `Labour Force Survey` (Canada 2024b) between 2018 to 2023. The key variables are:

- Sex: Categorized as males, females, both sexes.
- Age group: Divided into three groups: 15-24, 25-44, and 45 and over.
- Unemployment Rate: Calculated as the percentage of the unemployed in the labour force.

These categories allow us to dissect the impact of age and gender on unemployment rates. The dataset includes a six-year series, capturing fluctuations that may correlate with economic cycles or policy changes.

2.3 Data Measurement & Limitations

The `Labour Force Survey` (Canada 2024b) provides estimates of employment and unemployment. Here, we give more context to the general survey methods used.

- Survey Data Sources and Methodology
  - The target population for the Survey is the non-institutionalized population 15 years of age and above. They use a stratified multi-stage sampling approach in order to consolidate the collected data from various households.
- Error Detection
  - Some editing was carried out in the process of conducting these interviews. When inputting information into the computer that is out of range, the interviewe is prompted to modify the information.
  - They also identified lgically inconsistent or missing information items and modified them.
- Estimation
  - The sample data are weighted to make sure tabulations of estimates at the national, provincial, and sub-provincial levels are aggregated. The last adjustment to the weight is created to correct for coverage error in the survey.
- Revisions and Seasonal Adjustments
  - Seasonal adjustment is used to eliminate seasonal variations from almost 3,000 series in the survey. They do this using the X-12-ARIMA method.
  - At the beginning of each year, the survey revises its estimates.
- Data Accuracy

  - Since the Labour Force Survey is a sample survey, all estimates are subject to both sampling and non-sampling errors. Non-response to the survey tends to average about 10% of eligable households. A weight adjustment is usually applied to account for the households that did not respond.
  - Sampling errors are measured using coefficients of variation.

- Data Limitations

  - The Labour Force Survey uses the unemployment rate as a key measure. This rate is the international standard for labour force metrics. Nevertheless, the self-reported nature of the survey data may introduce some bias, a limitation that has to be acknowledged. Even with this limitation, the Labour Force Survey still remains the most useful and accurate dataset available.

A detailed datasheet is also provided in the Appendix (Section A)

2.4 Cleaning and Preparation

The dataset required minimal cleaning, primarily because of the accuracy of the dataset. There were no missing values, which indicates a comprehensive data collection process by Statistics Canada.

To carry out statistical analysis, this paper uses the Statistics of Canada `Labour force characteristics by sex and detailed age group, annual` (Canada 2024a) dataset. We are using the `R programming language` (R Core Team 2023) to conduct the analysis along with `readR` (Wickham, Hester, and Bryan 2024), `lubridate` (Grolemund and Wickham 2011), `tidyverse` (Wickham et al. 2019), `dplyr` (Wickham et al. 2023), `tidyr` (Wickham, Vaughan, and Girlich 2023), `knitr` (Xie 2014), `janitor` (Firke 2023), `scales` (Wickham, Pedersen, and Seidel 2023), `RColorBrewer` (Neuwirth 2022), `ggplot2` (Wickham 2016), `kableExtra` (Zhu 2024), `here` (Müller 2020), `arrow` (Richardson et al. 2024), `rstanarm` (Brilleman et al. 2018), `modelsummary` (Arel-Bundock 2022), and `lme4` (Bates et al. 2015).

3 Model

3.1 Model Set-up

The statistical model used to analyze Generation Z unemployment is represented by the following equation:

$$\begin{aligned}\text{Gen Z Unemployment} = \beta_0 &+ \beta_1 \cdot \text{Sex} \\ &+ \beta_2 \cdot \text{Reference Period} \\ &+ \beta_3 \cdot \text{Adult Unemployment} \\ &+ \beta_4 \cdot \text{Seniors Unemployment} + \epsilon\end{aligned}$$

Where:

- `Gen z Unemployment` is the dependent variable we want to predict. unemployment rate for Generation Z (people between the ages of 15 to 24)
- `Sex` is an independent dummy variable indicating gender.
- `Reference Period` is another independent variable that is a time variable, to account for economic trends and cycles.
- `Adult Unemployment`is an independent variable referring to the unemployment rate for adults aged 25 to 44.
- `Seniors Unemployment` is an independent variable referring to the unemployment rate for seniors aged 45 and above.
- β_0 is the y-intercept, representing the expected value of `Gen z Unemployment` when all the independent variables are zero.
- $\beta_1, \beta_2, \beta_3, \beta_4$ are coefficients for each independent variable, representing the change in `Gen z Unemployment` for a one-unit change in the respective independent variable, holding all other variables constant.
- ϵ represents the error term, accounting for the variability in `Gen z Unemployment` not explained by the model.

the coefficients $\beta_1, \beta_2, \beta_3, \beta_4$ identify the relationship between each independent variable and `Gen z Unemployment`:

- β_1 represents the differential impact of gender on Generation Z unemployment. If positive, suggests that being female is associated with higher unemployment rates compared to being male.
- β_2 shows the impact of time on unemployment rates.
- β_3 measures how unemployment rates in adults (25-44) are related to youth unemployment. A positive value would suggest that higher adult unemployment rates are associated with higher unemployment rates for Generation Z. A negative value would suggest that higher adult unemployment rates are associated with lower unemployment rates for

Generation Z, potentially showing a labour market where different age cohorts no not directly compete for the same jobs.

- β_4 reflects the relationship between seniors' unemployment rates (45 and above) and Generation z unemployment. If this coefficient is positive, it could imply that when older age groups face unemployment, Generation Z also tends to experience higher unemployment. It it is negative, it could imply a counter-cyclical effect where higher unemployment rates in older populations might correlate with lower unemployment rates among Generation Z, perhaps due to increased part-time or entry-level job openings.

3.2 Model Justification

The choice to use a multiple linear regression model to analyze the relationship between age groups, education level, and peoples unemployment status is underpinned by the following considerations.We have chosen to utilize Generation Z as the dependent variable because it has become a growing area of concern mainly as a result of Generation Z's openness and transparency regarding unemployment and its stark realities (Telford (2024)). Research consistently demonstrates gender disparities in labour market outcomes. Incorporating 'Sex' allows for explicit testing of these disparities within Generation Z.

The model assumes linearity, homoscedasticity, independence, and normality of the error term. It presumes that the relationships between the predictors and the overall outcome are linear and add up. Choosing to use a multiple linear regression allows us to evaluate each variable and its effect independently while controlling for potential impacts of other things. A non-linear model or the inclusion of interaction terms could have been used. The strengths of this model line in the fact that it is a very simple model to interpret, but this might be seen as a weakness as it poses a threat of oversimplifying the already complex labour market.

4 Results

4.1 Model Results

- **Intercept**: The model's intercept is estimated at approximately -235.13, suggesting that when all other variables are at zero, the predicted Gen Z unemployment rate would be negative, which is not meaningful in this context.
- **Sex**: The coefficients for SexFemales and SexMales suggest a comparison to generation z unemployment. SexFemales has a coefficient of approximately -0.60 , and SexMales has a coefficient of approximately 0.62 . Neither of these are statistically significant at conventional levels, indicating that there is no strong evidence of a difference in unemployment rates between females or males compared to the baseline when controlling for other factors.

- **Reference_Period**: The coefficient for Reference_Period is about 0.12. This suggests that there is no clear evidence of a trend over time in the Gen Z unemployment rate when other variables are held constant.
- **Adult_Unemployment**: The coefficient for adult_unemployment is the most statistically significant with an estimate of about 3.12. This suggests that a one-unit increase in the adult unemployment rate is associated with an increase of approximately 3.12 units in the Gen Z unemployment rate, holding other factors constant.
- **Seniors_Unemployment**: The coefficient for seniors_unemployment is approximately -0.61 , indicating that the association between seniors' unemployment rates and Gen Z's unemployment rates is not statistically significant at the conventional levels.

The model appears to strongly suggest that the unemployment rates among adults are a significant predictor of unemployment rates in Generation Z, while changes in seniors' unemployment rates and the effects of gender do not appear to have a significant association in this model. The lack of statistical significance in several variables could be due to a variety of factors, such as a small sample size, lack of variability in the predictors, or simply that there is no strong relationship between these predictors and the response variable in the population from which the sample was drawn.

5 Discussion

5.1 Implications of these results

Based on our analysis, we see that adult unemployment rates have really good predictive power over the unemployment rates among Generation Z. This relationship shows that younger generations are potentially vulnerable because of their lower experience level. When adults face unemployment, generation z individuals are likely to encounter employment challenges.This result is backed by historical trends that show generation z is often less employed than older cohorts (Koop (2021)). Due to longer periods of unemployment Gen Z will be missing out on allot of entry-level experience. This may affect them later in life, as they will often be playing catch up.

5.2 Possible Causes

5.2.1 Sector Vulnerability

Younger generations have workforces that are more primarily mde up of service-oriented sectors like hospitality and leisure. These are the same industries that were widely affected by the pandemic. Their jobs are also often not secure because of their nature of being part-time.

5.2.2 Digital & Remote Work Readiness

There has also been a very abrupt shift to remote work. Although generation Z is tech-sravy, the sudden and abrupt shift may have also sidelined people whose jobs cant be performed remotely.

5.2.3 Educational Disruptions

A possible cause of this might be the unstudied effect of educational levels. Education levels play a very significant part in unemployment rates. There might be an unstudied shift in levels of education between these two age groups.

5.2.4 Global Economic Slowdown

Beyond the very immediate effects of the pandemic, the global economic slowdown affected international trade, overall investments and supply chains in general. This evidently caused a knock-on effect on job markets around the world.

5.3 Weaknesses and next steps

While our analysis has been effective in showing various trends, it has shown some limitations. The data chosen heavily relies on aggregate unemployment rates which does not show the underlying trends in the labour market. For example, within Gen Z, as mentioned in the prior section, there may be significant differences based on peoples education level, race, ethnicity,and location that the current analysis does not account for. Another limitation is the static nature of the Labour Force Survey. This data does not track individuals over time, rather it just takes a snapshot of individuals. This does not allow us to truly understand the actual trends.

In terms of next steps this analysis can take, future research would benefit from a dataset that is more robust in nature. Looking into datasets that have more variables like economic health metrics, industries, and educational attainment would provide more insight into the labour market in Canada. Incorporating macroeconomic factors like inflation, GDP, and changes in labour market policies could further contextualize the results within broader economic realities.

By addressing these weaknesses and looking at next steps forward, future research has the ability to be very informative and shape further labour policies.

6 Conclusion

Analyzing the relationship between gender and unemployment rates from 2018 to 2023 in Canada has revealed several trends within the labour market by using the Labour Force Survey and a linear regression model. The adult unemployment rate emerged as the most significant predictor of Generation Z's empowerment status, a reflection of the broader economic conditions that disproportionately affect the youngest entrants into the labour market.

Moving forward, it is important for policy makers to take into account the role that gener and age plays in the labour market. This paper shows that inclusive economic strategies that take into account various groups and disparities.To decrease the unemployment rates, it is important to take a multidisciplinary approach where all sectors are considered. Investments into education and equity will be crucial in forging a path forward.

A Appendix

A.1 Datasheet for Labour Force Survey

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.*
    - 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, aproximately 100,000 idividuals are suveyed.

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 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 anny errors, noise, or redundance.

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 compitence of the current governemnt.

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

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Brilleman, SL, MJ Crowther, M Moreno-Betancur, J Buros Novik, and R Wolfe. 2018. "Joint Longitudinal and Time-to-Event Models via Stan." https://github.com/stan-dev/stancon_talks/.

Canada, Statistics. 2024a. *Labour Force Characteristics by Sex and Detailed Age Group, Annual*. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410032701.

———. 2024b. *Labour Force Survey (LFS)*. https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3701.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

Koop, Avery. 2021. *Charted: The Gen z Unemployment Rate, Compared to Older Generations*. Visual Capitalist. https://www.visualcapitalist.com/gen-z-unemployment-rate-chart/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Neuwirth, Erich. 2022. *RColorBrewer: ColorBrewer Palettes*. https://CRAN.R-project.org/package=RColorBrewer.

Organization, International Labour. 2017. *The Gender Gap in Employment: What's Holding Women Back?* International Labour Organization. https://www.ilo.org/infostories/en-GB/Stories/Employment/barriers-women#intro.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Telford, Taylor. 2024. *No Job? No Shame. Younger Workers Are Opening up about Unemployment*. The Washington Post. https://www.washingtonpost.com/business/2024/02/10/gen-z-layoffs-tiktok-younger-workers/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023.

Dplyr: A Grammar of Data Manipulation. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Wickham, Hadley, Thomas Lin Pedersen, and Dana Seidel. 2023. *Scales: Scale Functions for Visualization*. https://CRAN.R-project.org/package=scales.

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2023. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.