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 created to analyze mortality trends in Canada from 2000 to 2022, with a focus on age, sex, and cause of death. The purpose is to understand public health trends and guide policymakers to made decision

2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*

The dataset was created by Statistics Canada, a national statistical agency responsible for producing statistics to help better understand Canada, its population, resources, economy, society, and culture.

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 was funded by the Government of Canada, which provides funding to Statistics Canada to carry out its statistical activities

4. *Any other comments?*

N/A

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 in the dataset represent mortality cases in Canada from 2000 to 2022, categorized by age, sex, and cause of death. The dataset includes information on the total number of deaths, leading causes of death, age at the time of death, gender distribution, and partial places of residence.

2. *How many instances are there in total (of each type, if appropriate)?*

The exact number of instances is not provided.

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 dataset contains a sample of mortality instances from the larger population of deaths in Canada during the specified timeframe. It is representative of mortality trends in Canada during this period.

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 dataset included structured data representing various attributes related to mortality cases, such as age, sex, cause of death, and geographical information.

5. *Is there a label or target associated with each instance? If so, please provide a description.*

There is no explicit mention of labels or targets.

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.*

The excerpt does not mention any missing information in individual instances.

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.*

The excerpt does not mention any explicit relationships between individual instances,

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.*

No recommended data splits.

9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*
There is no explicit mention of errors, sources of noise, or redundancies in the dataset.

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 seems to be self-contained and sourced directly from Statistics Canada

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.*

The dataset contains information include personal details such as age, gender, and cause of death. While this information is not confidential.

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.*

The dataset itself does not contain content that is inherently offensive, insulting, threatening, or anxiety-inducing

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.*

e dataset identifies sub-populations based on age and gender. Age groups and gender distributions are used to analyse mortality trends within specific demographic part.

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.*

Since the dataset represents mortality cases at an aggregate level, it is not directly identify individuals.

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 dataset contains information related to health data, including causes of death, which may be considered sensitive

16. *Any other comments?*
     N/A

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.*

     The data used in the study was acquired from Statistics Canada.

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?*

The data is sourced from Statistics Canada, it is that the data collection involved governmental procedures and databases.It include official death records, medical certificates, and population censuses.

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 dataset covers a significant portion of mortality data in Canada during the specified timeframe.

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)?*
The individuals involved in the data collection process are 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.*

The data spans from 2000 to 2022, which aligns with the specified timeframe of the study.

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.*

There is no specific mention.

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 data used in the study was acquired 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.*

Not mentioned.

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.*

Given that the data is sourced from governmental records, individual consent may not have been obtained for its collection

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).*

individual consent is not allowed to obtain the data collection, there is no specific mechanism for individuals.

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.*

There is not mentioned

12. *Any other comments?*

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.*

Extract top 10 causes leading to mortality, perform ranking,the numbers of death,years

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.*

Yes.raw data saved is the purpose of supporting unanticipated future uses

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

R studio to clean data

4. *Any other comments?*

N/A

Uses

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

The examples of previous uses of the dataset are not provided. Policymakers etc. have utilised the dataset to analyse mortality trends in Canada and adjust public health strategies

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.*

There is no specific repository mentioned that links to papers or systems utilizing this dataset.

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

The dataset can be used for comparing mortality trends in Canada with other countries,Predicting future mortality trends based on historical data etc.

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?*

Users need to be cautious about extending conclusions beyond what the dataset can support or assuming cause-and-effect relationships without evidence. It's important to think about the ethical impact and make sure that analyzing the data doesn't promote stereotypes or unfairly target certain groups.

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

It is inappropriate to use the data to discriminate against specific populations based on health outcomes

6. *Any other comments?*

N/A

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.*

The dataset is publicly available on the Statistics Canada website and can be accessed by anyone without restrictions. It can be distributed to third parties outside of the entity that created it.

2. *How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?*

The dataset can be downloaded by everyone in statistics Canada.There is no DOI assigned to it.

3. *When will the dataset be distributed?*

It is already for public access.

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.*

The dataset is available under the Open Government License - Canada. This licence allows users to freely use, modify, and distribute the dataset.

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.*

There is no mention of third-party restrictions imposed on the dataset.Users are free to use the dataset without any restrictions.

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.*

There are no specific export controls or regulatory restrictions associated with the dataset.
7. *Any other comments?*
N/A

Maintenance

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

Statistics Canada will be responsible for supporting, hosting, and maintaining the dataset.

2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*

Contact information for Statistics Canada could be provided for inquiries about the dataset.

3. *Is there an erratum? If so, please provide a link or other access point.*

There is no mention of erratum

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)?*

The dataset will be update annually

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.*

There is no mention

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.*

There is no specific information about it.

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.*

The data sourced from government website, there is no mentioned about external contribution

8. *Any other comments?*
N/A