Datasheet for Alberta Death Cause from 2001 to 2022 Dataset*

Xincheng Zhang   Zhijun Zhong

Invalid Date

Mortality rates and annual death numbers are critical indicators of population dynamics, reflecting both the growth and contraction of the population. In this report, we conducted a complete investigation into the various causes of death in the year 2020, developing multiple models to elucidate the six leading causes of mortality and deducing the formula to predict future situations. Our findings indicate that Governments need to concentrate on environmental issues and raise citizen awareness to minimize the number of neoplasms, organic dementia, and COVID-19 deaths. Our objective is to enhance overall well-being and elevate human happiness by Unveiling the escalation in mortality attributable to diverse factors.

Extract of the questions from (**gebru2021datasheets?**).

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 enable analysis of Australian politicians. We were unable to find a publicly available dataset in a structured format that had the biographical and political information on Australian politicians that was needed for modelling.

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

   - Dataset is created by service Alberta an authoritative government organization on behalf of the whole citizen in the province.

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.*

   - No grant. A government organization

---

*Code and data are available at: https://github.com/JerrZzzz/Demystify-Demise.

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 that comprised the dataset is the resident(people) in the Alberta, Canada. There is no rating on this dataset. But we can be sure that this dataset is authoritative.

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

   - The dataset is based on Alberta citizens from year 2001 to 2022 and icludes 663 rows.

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).*

   - It is possible that it is selected from a larger dataset. But death information are private information which is not the best idea to publish all dataset or observations. The dataset limit the geographic area to the Alberta, Canada, which means that it is not a set of several diverse range of instances.

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 of the instances consists the year, the cause of death, total death because of this cause and the ranking of this cause in this year.

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

   - each column are well labelled. We can get an idea of what each column means by a glimpse.

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.*

- No. But for intentionally removed information, we cannot be sure. We believe there is a removed information. But the removal is because the information is personalized.

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.*

   - No

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.*

   - There are no data splits.

9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*

   - No error are found 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.*

    - No link to external resourses. There is a license where tells us that we can download or republish the dataset. but we have to follow the rules here https://open.alberta.ca/licence.

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 data is republishable.

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.*

- No, Only population in Alberta.

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.*

    - Citizens in Alberta

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.*

    - No.

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.*

    - No. We cannot access the information about how the data is acquired.

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?*

    - We don't know.

3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*

    - No, we don't know

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)?*

    - Government workers are involved in collection process and we don't know the information of compensation.

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.*

   - From 2001 to 2022, which is match the creation timeframe of the data associated with the instances.

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.*

   - Collection process are not shown on the webpage.

7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*

   - No, we collect data from a government organization.

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.*

   - We do not have access to the above information.

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.*

   - We do not have access to the above information.

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).*

    - We do not have access to the above information.

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.*

    - We do not have access to the above information.

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.*

   - We do not have access to the above information.

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.*

   - We do not have access to the above information.

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

   - We do not have access to the above information.

4. *Any other comments?*

   - No.

Uses

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

   - No.

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?*

   - Dataset can be used for compare with the future and make suggestion for any changes for financial budget on different area decide by 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, the data does not shown any personal information.

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

   - The dataset should not be used for any of the 5 things below:

   a. Personal Information;
   b. Information or Records that are not accessible under applicable laws;
   c. third party rights the Information Provider is not authorized to license;
   d. the names, crests, logos, or other official symbols of the Information Provider; and
   e. Information subject to other intellectual property rights, including patents, trademarks and official marks.

   These information can be found on license.

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.*

   - No, all third parties can download data from the website.

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

   - It was published on government's website. No DOI can be found.

3. *When will the dataset be distributed?*

   - DATE CREATED is on 2015-05-13 and update on 2023-09-22

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.*

   - No license can be found regarding distribution. There is a license for using 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.*

   - Governing Law: This licence is governed by the laws of the province of Alberta and the applicable laws of Canada. Legal proceedings related to this licence may only be brought in the courts of Alberta.

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 restrictions. But we should not do:

   a. Personal Information;
   b. Information or Records that are not accessible under applicable laws;
   c. third party rights the Information Provider is not authorized to license;
   d. the names, crests, logos, or other official symbols of the Information Provider; and
   e. Information subject to other intellectual property rights, including patents, trademarks and official marks.

7. *Any other comments?*

   - No

Maintenance

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

   - Government of Alberta

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

   - CONTACT EMAIL: sa.opendata-rs@gov.ab.ca

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. The newest update is on 2023-09-22

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 limits on the retentaion of the data. We do not have access to the above information

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.*

- Government of Alberta will keep working on the data and update them. But we do not have information on whether or not they will continue doing it in the future.

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, it is a collection of death on a government base. We do not expect people to extend on the dataset. Government of Alberta will not accept extended dataset as their new updated dataset.

8. *Any other comments?*

   - No

1 References