Datasheet for Committee of Adjustment Applications Dataset*

Understanding more about the dataset

Angel Xu

December 3, 2024

Extract of the questions from Gebru et al. (2021).

The dataset discussed in this datasheet is 'About Committee of Adjustment Applications' (City of Toronto Open Data Portal 2024).

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 urban planning applications submitted to Toronto's Committee of Adjustment. It identifies patterns influencing approvals based on application type, submission year, and planning district, addressing the gap in urban planning research for transparent decision-making.
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 compiled and maintained by the City of Toronto's Committee of Adjustment, published through the City of Toronto Open Data Portal.
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.*
    - It maintained as part of municipal operations by the City of Toronto, likely funded by local government resources.
4. *Any other comments?*
    - None

*Code and data are available at: https://github.com/Anjojoo/Committee-of-Adjustment-Applications.

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 formal applications submitted to the City of Toronto's Committee of Adjustment. Each instance corresponds to a unique application, detailing the specific request for zoning variances or land use consents.
    - Multiple types of instances in the dataset primarily include:
        - Applications for Minor Variances (MV): Requests for small adjustments to zoning bylaws, such as building height limits or property setbacks.
        - Applications for Consents (CO): Requests related to land subdivision or adjustments required for property transactions or development.
    - In addition to these core application types, the dataset also includes related information about the:
        - Applicants: Individuals or entities submitting the applications, such as property owners, developers, or authorized agents.
        - Properties: Details about the location and zoning characteristics of the properties associated with the applications.
        - Decisions: The outcomes of the applications (e.g., approved, refused) once processed by the Committee of Adjustment.

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

    - There are 3559 instances in total and will continue increase as new data is updated.

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 contains all recorded applications submitted to the City of Toronto's Committee of Adjustment during the specified timeframe. It is a comprehensive record rather than a sample, systematically maintained by the city to include all relevant applications processed by the committee.
    - While the dataset encompasses all formal applications within its defined scope, it represents a subset of the broader context of land-use adjustments in Toronto. This broader context may include informal requests resolved without a formal application, adjustments handled by other municipal departments, and activities not requiring

Committee of Adjustment approval. Additionally, applications outside the specified timeframe are excluded.

- The dataset is representative of formal applications submitted to the Committee of Adjustment, reflecting geographic and temporal variations across all planning districts in Toronto. However, its focus on formal submissions means it does not account for informal or administrative adjustments. The representativeness of the dataset is validated through its complete coverage of applications within the committee's jurisdiction, ensuring systematic inclusion of all eligible instances during the specified 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 raw dataset represents a formal application submitted to the City of Toronto's Committee of Adjustment and consists of structured data in the form of features. These features capture key attributes of each application, including:
        - Application Details: Type of application (Minor Variance or Consent), submission year, and purpose.
        - Property Information: Geographic location (planning district), zoning designations, and property identifiers.
        - Administrative Data: Applicant details (e.g. owner, developer, or agent), decision status (e.g. approved or refused), and times for submission and resolution.
        - Descriptive Features: Project descriptions and specific requests related to zoning variances or consents.

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

    - Yes, each instance has a label or target associated with it, the decision outcome of the application.
    - The decision outcome is a variable indicating the final application decision by the Committee of Adjustment. This label reflects the final resolution of the application process, determined after administrative review, public hearings, and committee deliberation. It serves as the primary target for analysis in the cleaned dataset, where it is further processed into a binary format (1 for approved, 0 for refused) to enable predictive modeling and statistical analysis.

- The decision outcome encapsulates the core purpose of the dataset: to analyze patterns and factors influencing urban planning decisions in Toronto.

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.*

- Some instances in the raw dataset contain missing information due to incomplete submissions or administrative oversights. Missing data may include applicant details, project descriptions, property information, or time spots, often resulting from errors during data entry or non-mandatory fields in the submission system. External factors, such as unregistered properties or unclear zoning bylaws, also contribute to these gaps.
- These missing data points reflect the challenges of maintaining comprehensive administrative records. The cleaned dataset addressed these issues through preprocessing, such as filtering and reconstructing variables, ensuring the data was complete and reliable for analysis.

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.*

    - Relationships between individual instances are not explicitly represented in the raw dataset. Each instance corresponds to an independent application submitted to the Committee of Adjustment, with no direct links or interactions recorded between applications.
    - However, implicit relationships can be inferred through shared attributes such as planning districts, submission years, or application types. For example, multiple applications from the same district may reflect localized development trends or zoning challenges. These relationships are not explicitly encoded but can be analyzed through groupings or aggregations of shared variables in 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.*

    - There are no predefined splits in the raw dataset while a data split is applied on the cleaned dataset for model validation. The dataset is divided into training (80%) and testing (20%) subsets.
    - The rationale for this split is to use the training set for model building and parameter estimation, while the testing set evaluates the model's performance on unseen data. This ensures the model generalizes well and avoids overfitting.

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

    - No.

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.

- Self-contained.

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, any personally identifiable information about applicants, such as names, contact details, or specific property addresses, is either excluded or anonymized in the dataset.

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. The dataset does not explicitly identify sub-populations based on demographic attributes such as age or gender. Instead, sub-populations are represented through administrative and geographic groupings relevant to urban planning decisions.

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.

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?*

    - None.

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 for each instance was primarily directly observable and recorded as part of the formal application process to the City of Toronto's Committee of Adjustment. Key data points, such as application type, geographic location, and submission dates, were submitted directly by applicants through a centralized online system or physical forms. Decision outcomes were added by committee staff following administrative review and public hearings.
- Some data, such as zoning compliance or project descriptions, can be considered reported by subjects, as applicants provide this information when submitting their applications. These submissions are validated by committee staff to ensure they align with zoning regulations and meet the submission requirements.

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 was collected through a centralized online application system maintained by the City of Toronto's Committee of Adjustment. Applicants submit their applications via this digital platform or through physical forms, which are then digitized by administrative staff.
   - Administrative staff manually input application details from physical submissions and verify the accuracy of online submissions. Besides, submitted data is stored in a centralized digital database that standardizes and organizes records, ensuring consistency across entries. Submission and resolution dates are automatically records in the system.
   - For validation, submitted applications are reviewed by committee staff to ensure compliance with zoning bylaws and completeness of information. Missing or unclear details are flagged for clarification with applicants. Additionally, the centralized system enforces mandatory fields during the submission process, reducing the likelihood of incomplete records. The database is also maintained and updated regularly to reflect accurate and current information, ensuring its reliability for public use and analysis.

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 is not a sample

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 data collection process was conducted by administrative staff and officials at the City of Toronto's Committee of Adjustment. These staff members are municipal employees who are compensated as part of their regular salaries funded by the city government. They were responsible for verifying, recording, and maintaining the data submitted by applicants.
- Applicants submitted their applications voluntarily as part of the formal urban planning process. They were not compensated for providing their data, as the submission is a required step for obtaining zoning variances or consents. Application fees paid by the applicants help fund the administrative operations of the Committee of Adjustment.

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 raw dataset includes applications submitted from 1999 to the present. This timeframe matches the creation timeframe of the data, as the dataset records information at the time applications are submitted, reviewed, and decided upon. The times for submissions, hearings, and decisions accurately reflect the actual events as they occurred, ensuring no significant temporal discrepancies between the creation and collection of data.
   - The cleaned dataset filtered out earlier years, which have sparse data, to improve analytical reliability, but the timeframe still aligns with the period during which the applications were originally created.

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. The data was collected and maintained by the City of Toronto's Committee of Adjustment as part of routine administrative operations. Since the dataset primarily contains anonymized and aggregated administrative data without sensitive personal identifiers, it does not fall under categories that typically require ethical review.

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 was collected directly from individuals and entities submitting applications to the City of Toronto's Committee of Adjustment.

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, individuals submitting applications to the City of Toronto's Committee of Adjustment were notified about the data collection. The submission process, whether online or in-person, includes a notification that the information provided will be used for processing the application and may be recorded in public records as part of municipal governance procedures.

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.*

    - Individuals submitting applications to the City of Toronto's Committee of Adjustment implicitly consented to the collection and use of their data as part of the formal application process. Consent is typically provided by agreeing to terms of service or privacy statements for online submissions or signing disclaimers included on physical forms.

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).*

    - There is no indication that individuals submitting applications to the City of Toronto's Committee of Adjustment are provided with a formal mechanism to revoke their consent after submission.

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.

12. *Any other comments?*

    - None.

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, not mentioned.

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, can be accessed in `data/01-raw_data`.

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 (R Core Team 2023) was used. For cleaning process of used analysis dataset in this paper, it can be accessed in `scripts/03-clean_data.R`.

4. *Any other comments?*

    - None.

Uses

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

    - Yes, the dataset has been used to analyze factors influencing urban planning decisions in Toronto by Angel Xu. No other use of the dataset is found.

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?*

    - The dataset has the potential to support various tasks related to urban planning, governance, and policy analysis. It could be used for predictive modeling to estimate application outcomes based on geographic, temporal, and contextual factors, providing insights into decision-making trends. Additionally, the dataset could be applied to process optimization by identifying inefficiencies in the application review and approval process and suggesting administrative improvements. It could also be used to explore spatial analysis to uncover geographic disparities in approval rates, highlighting districts with higher or lower likelihoods of approval and identifying potential inequities or regional planning challenges.

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?*

    - Yes, the dataset reflects historical decision-making processes, which may include systemic biases. Using this dataset without accounting for such factors could perpetuate unfair treatment or stereotyping of specific regions or application types.

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

    - Yes, the dataset should not be used to identify or profile individual applicants since it does not contain personal identifiers and was not designed for such purposes. The dataset should not be used to draw conclusions about zoning or planning decisions in regions outside Toronto as well.

6. *Any other comments?*

    - None.

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.*

    - Yes, the dataset is distributed publicly through the City of Open Data Toronto Portal (Open Data Toronto n.d.).

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 is distributed through the City of Open Data Toronto Portal (Open Data Toronto n.d.) as downloadable files (can be download as CSV, JSON or XML format). Users can access the data directly from the portal's website, where it is periodically updated and maintained by the city. However, the dataset does not have a DOI.

3. *When will the dataset be distributed?*

    - It is updated daily to reflect new applications and decisions, ensuring that the data remains current and relevant. Users can access the dataset at any time.

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.*

    - Yes, the dataset is distributed under the City of Toronto's Open Data License, which allows public use with certain restrictions. The license permits users to copy, modify, publish, and share the data, provided that proper attribution is given to the City of Toronto. It prohibits using the data in ways that misrepresent its source or for unlawful purposes.
    - No fees is associated with accessing or using the dataset under this license.

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.*
    - No.
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?*
    - None.

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*
    - The dataset is supported, hosted, and maintained by the City of Toronto, specifically through its Open Data Portal (Open Data Toronto n.d.). The City's municipal departments are responsible for collecting, verifying, updating, and publishing the data.
    - Regular updates are performed to ensure accuracy and reflect the latest submissions and decisions by the Committee of Adjustment.
2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*
    - The owner, curator, or manager of the dataset can be contacted through the City of Toronto's Open Data team or the Committee of Adjustment.
    - For general inquiries about the dataset, users can email the City of Toronto's Open Data team at opendata@toronto.ca. For specific questions about the Committee of Adjustment applications, additional contact details are available on the Committee of Adjustment website.
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 dataset is refreshed dailty while updated when new applications occur. Updates are managed by the City of Toronto's Open Data team in collaboration with the Committee of Adjustment. Updates are communicated to dataset consumers through the Open Data Toronto Portal (Open Data Toronto n.d.).

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.*

    - The dataset relates to formal applications rather than personal data. Thus, no explicit limits on the retention of data are specified.

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.*

    - No.

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.*

    - There is no formal mechanism for external parties to directly extend, augment, or contribute to the dataset. However, others can contact the Open Data team at opendata@toronto.ca if they wish to provide feedback or suggest improvements.

8. *Any other comments?*

    - None.

References

City of Toronto Open Data Portal. 2024. "Committee of Adjustment Applications." https://open.toronto.ca/dataset/committee-of-adjustment-applications/.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. "Datasheets for Datasets." *Communications of the ACM* 64 (12): 86–92.

Open Data Toronto. n.d. "Open Data Toronto." https://open.toronto.ca/.

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