Understanding Factors Contributing to Application Approval in Urban Planning Decisions (2016-2024)*

Analysis using data from Toronto's Committee of Adjustment applications using Bayesian Logistic Regression

Angel Xu

November 30, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

Table of contents

---

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

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2….

1.1 Estimand

2 Data

2.1 Overview

Our data (City of Toronto Open Data Portal 2024) is extracted from opendatatoronto (Open Data Toronto n.d.). This dataset focuses on applications submitted to the City of Toronto's Committee of Adjustment, a governing body responsible for reviewing minor variance and consent applications related to land use and development. These applications are submitted by property owners and developers seeking permission to deviate from existing zoning regulations or to subdivide land. The Committee's decisions—whether to approve or refuse an application—have significant implications for urban development, community planning, and city growth.

The Committee of Adjustment plays a key role in Toronto's urban planning ecosystem. It ensures that deviations from zoning bylaws align with the city's broader development objectives, maintaining a balance between flexibility for property owners and adherence to long-term planning goals. The decisions made by the committee are influenced by various factors, including the type of application, geographic considerations (e.g. planning districts), and the timing of submission. Understanding these decisions provides insight into urban planning dynamics and the regulatory challenges faced by a growing city like Toronto.

The analysis presented in this paper were conducted using R programming language (R Core Team 2023). The `tidyverse` packages (Wickham et al. 2019), `dplyr` package (Wickham et al. 2023), and `arrow` package (Richardson et al. 2024) were used in the process of data simulation, testing beforehand. After the original raw data was downloaded by using `tidyverse` package (Wickham et al. 2019), and `dplyr` package (Wickham et al. 2023), data cleaning process was done by using `tidyverse` package (Wickham et al. 2019), `dplyr` package (Wickham et al. 2023), and `arrow` package (Richardson et al. 2024). The code style is corrected by `here` package (Müller 2020), `lintr` package (Hester et al. 2024), and `styler` package (Müller and Walthert 2024). Then, models were constructed using `tidyverse` package (Wickham et al. 2019), and `rstanarm` (Goodrich et al. 2022) package. Graphs were made with `ggplot2` package (Wickham 2016) and `tidybayes` package (Kay 2024). Tables were constructed with `knitr` package (Xie 2023). The model results are then presented by `modelsummary` (Arel-Bundock 2022) package.

Following Alexander (2023), this paper uses Bayesian Logistic Model to explore factors that affect the application results.

Although several similar datasets with overlapping themes were considered but ultimately not used due to limitations in scope, structure, or relevance. For instance, the Toronto Application Information Centre Dataset provides information on broader planning applications like rezoning requests and site plan approvals, but its complexity and inclusion of unrelated application

types make it unsuitable for isolating patterns related to minor variances and consents. The Ontario Municipal Board (OMB) Decisions Dataset includes appeal outcomes, but its focus on contested cases introduces bias and omits the routine applications central to this study.

Consequently, focusing on the Committee of Adjustment Applications dataset (City of Toronto Open Data Portal 2024) stands out for its granularity and relevance, enabling a focused analysis of approval trends and their influencing factors, which directly aligns with the study's goals.

2.2 Measurement

The dataset captures information on applications submitted to the City of Toronto's Committee of Adjustment, providing a structured record of decisions made on zoning variances and consents. The dataset's structure results from Toronto's well-defined application process. When property owners or developers submit applications to the Committee of Adjustment, they provide detailed information about their project, including its location, purpose, and the specific variances or consents requested. These details are reviewed by committee staff and entered into a digital database, ensuring consistency and standardization.

Applications to the Committee of Adjustment are typically submitted by property owners, developers, or their authorized agents (e.g., architects or planners). Eligibility to submit an application requires ownership or legal rights to the property in question. For some application types, such as consents, additional prerequisites include demonstrating compliance with existing zoning bylaws or providing documentation showing that the proposed subdivision aligns with city planning objectives. Minor variance applications, on the other hand, require applicants to show that the requested adjustments are "minor" in nature, do not conflict with broader zoning intentions, and maintain compatibility with the surrounding area. As a result, individuals submitting these applications are often professionals with expertise in urban planning, real estate, or development, or property owners working in consultation with such professionals.

The data collection process begins with property owners or developers submitting applications through a centralized system. Each application includes essential details, which are systematically verified and recorded by administrative staff. Following public hearings, committee decisions (approval or refusal) are added to the database, completing the data entry process. The dataset is then maintained and periodically published by the City of Toronto, ensuring it remains accessible and accurate for public and analytical use.

The dataset thus provides a structured lens through which to examine urban planning decisions. By distilling real-world phenomena—such as zoning adjustments and land use approvals—into discrete variables, the dataset enables systematic analysis of trends and relationships within Toronto's regulatory framework. This structured approach ensures that the dataset is both practical for analysis and representative of the nuanced processes underlying urban development.

2.3 Data Cleaning

To ensure the dataset was suitable for analysis, several high-level cleaning steps were performed. First, unused decision outcomes were filtered out, retaining only applications that were explicitly approved or refused. The `decision` variable was then converted into a binary format, with `1` representing approvals and `0` representing refusals, simplifying its use in modeling.

A new variable, `year`, was constructed by extracting the year from the `in_date` field. This allowed for a clearer examination of temporal trends while maintaining the granularity of the original data. Applications from very early years with minimal counts were excluded to avoid bias and ensure a representative analysis of contemporary decision-making trends.

Additionally, key variables are selected out in cleaned dataset for convenience. More details will be explored in appendix.

2.4 Features

The dataset provides details of applications submitted to the Committee of Adjustment, focusing on key variables:

- Decision: Indicates whether an application was approved or refused.
- Application Type: Categorizes the application as either a minor variance or a consent.
- Year: The year when the application was submitted.
- Planning District: The geographic area within Toronto where the application originated.

Table 1: Preview of Key Variables

application_type	year	planning_district	decision
MV	2016	North York	0
MV	2016	North York	1
MV	2016	North York	0
MV	2016	North York	1
MV	2016	North York	1
MV	2016	Toronto East York	1

The first six rows of the cleaned dataset, displayed in Table 1, offer a clear preview of these key variables.

These variables allow for an in-depth analysis of the factors influencing application decision results and provide a quantitative basis for evaluating trends in urban planning decisions. Specific and further discussion about these variables are included in Section 2.5 and Section 2.6.

2.5 Outcome variables

The outcome variable, decision, captures whether an application submitted to the Committee of Adjustment was approved (represented as 1) or refused (represented as 0).
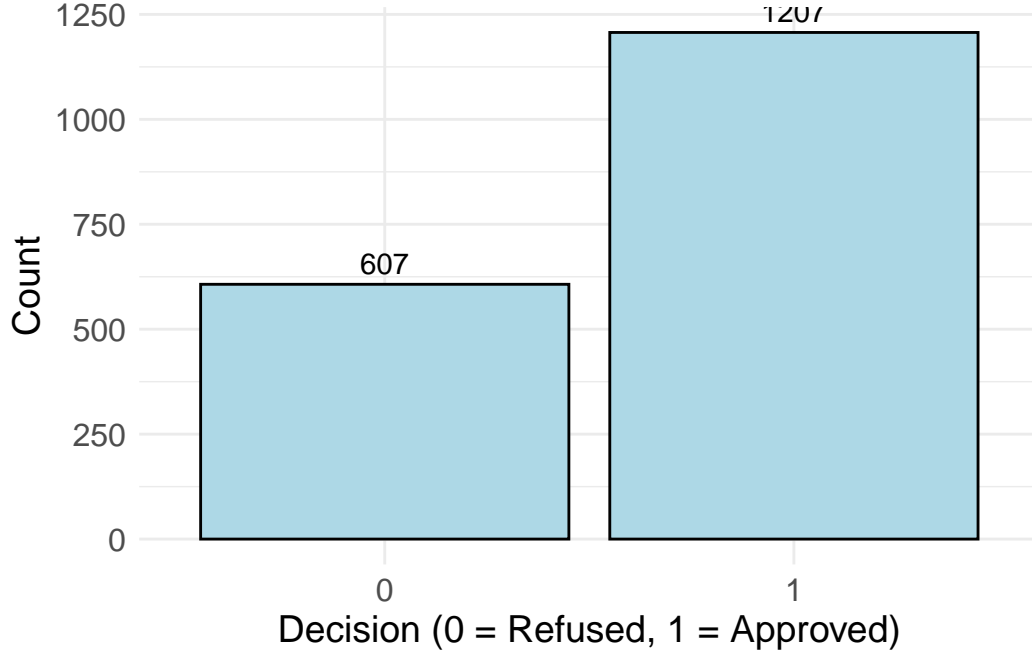

Figure 1: Distribution of Application Decisions

Figure 1 illustrates the distribution of decision outcomes. Out of the total applications, 1,207 (66.5%) were approved, while 607 (33.5%) were refused. This demonstrates that the majority of applications submitted to the Committee of Adjustment are approved, suggesting that most requests align with zoning regulations or are deemed acceptable following the committee's review process.

Table 2: Summary Statistics of Decisions

Decision	Percentage
Refused (0)	33.5%
Approved (1)	66.5%

Table 2 provides a clear picture of the approval and refusal rates within the dataset, highlighting a relatively high likelihood of approval for submitted applications.

2.6 Predictor variables

2.6.1 Application type

The application type variable categorizes each application submitted to the Committee of Adjustment into two types:

Consent (CO): Requests related to the subdivision of land into separate parcels, often for new developments or property sales.

Minor Variance (MV): Applications seeking small adjustments to zoning regulations, such as building setbacks, height limits, or lot coverage.

These categories represent distinct regulatory processes, with Consents generally requiring more extensive evaluation compared to Minor Variances due to their implications on land use.

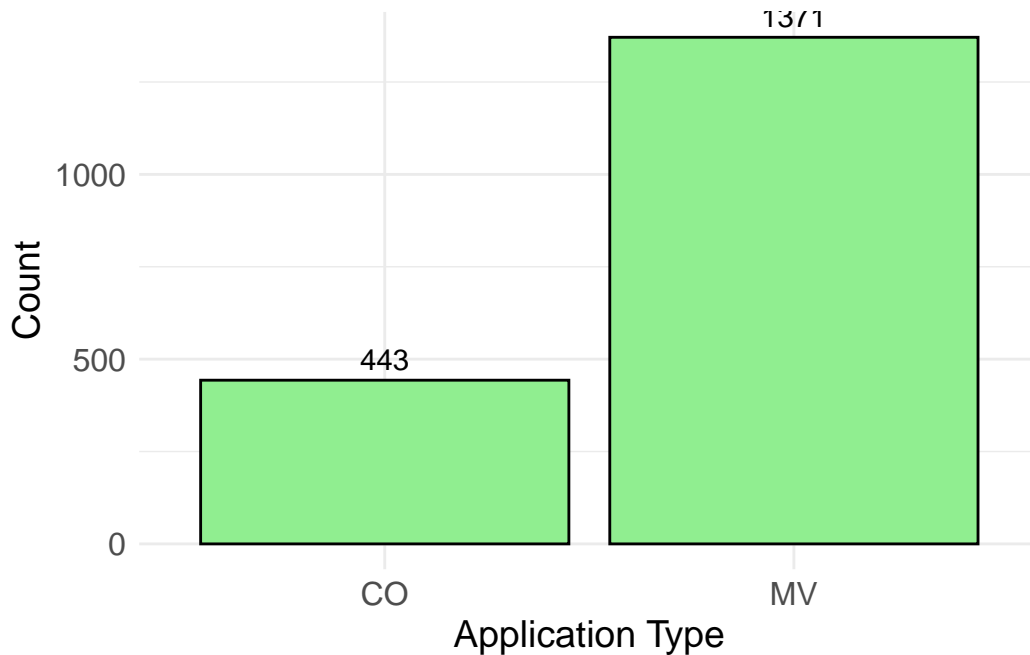Figure 2: Distribution of Application Types

Figure 2 presents the distribution of applications by type. Minor Variances account for the majority of applications which reach to 1371 (75.6%), reflecting the routine nature of these requests in urban development. In contrast, Consents are less frequent which only have 443 counts (24.4%), likely due to their more specific requirements and potential complexity.

Table 3: Summary Statistics of Application Types

Application Type	Percentage
Consent (CO)	24.4%
Minor Variance (MV)	75.6%

Table 3 emphasize the dominance of Minor Variances in the dataset, providing context for their role in urban development.

The relationship between application_type and the outcome variable decision will be explored further to determine if approval rates differ between Minor Variances and Consents. Preliminary trends suggest that Minor Variances may have higher approval rates due to their less complex nature. Additional analyses may examine whether these patterns vary by planning district or year.

2.6.2 Year

The year variable represents the year of submission for each application in the dataset. It provides insight into temporal trends in applications submitted to the Committee of Adjustment. This variable is instrumental in identifying changes in the volume of applications over time, which may reflect policy shifts, urban development trends, or external factors like economic conditions.

Figure 3 clarifies the distribution of applications from 2016 to 2024, with notable fluctuations in annual counts. Fewer applications were submitted during 2016-2019, with counts declining steadily from 198 applications in 2016 to a low of 93 in 2019. This period may reflect a stabilization phase or lower activity in urban development processes. Afterwards, application counts increased significantly, peaking in 2023 with 378 applications, followed by a slight decline to 249 in 2024. The rise in 2023 may be due to heightened development activity or easing of restrictions from prior external events, such as the COVID-19 pandemic.

Table 4: Summary Statistics of Years

Years	Percentage
2016	10.9%
2017	9.8%
2018	6.6%
2019	5.1%
2020	8.6%
2021	14.3%
2022	10.1%

Table 4: Summary Statistics of Years

Years	Percentage
2023	20.8%
2024	13.7%

Table 4 states how each year contributes to the total volume of applications.

The year variable is likely related to the decision outcome and application type. As approval rates may have varied over time due to changing city planning policies or committee priorities. Besides, temporal trends may reveal shifts in the proportion of Minor Variances (MV) versus Consents (CO), reflecting evolving development demands.

2.6.3 Planning District

The planning district variable identifies the geographic district within Toronto where an application was submitted. The city is divided into four planning districts, each with its own development characteristics and zoning regulations. These districts are crucial for understanding spatial trends in application submissions and outcomes, as local policies and priorities can influence decision-making processes.

Figure 4 displays the distribution of applications across four planning districts. We can observe that North York has the highest number of applications, accounting for a significant proportion of the dataset. This could reflect the area's high density of development projects or active property owners seeking zoning adjustments. Scarborough has the fewest applications, possibly due to fewer development activities or stricter zoning compliance. Etobicoke York and Toronto East York have moderate numbers of applications, reflecting their roles as active but less prominent districts compared to North York.

Table 5: Summary Statistics of Planning Districts

Planning Districts	Percentage
Etobicoke York	16.9%
North York	43.8%
Scarborough	12.7%
Toronto East York	26.6%

Table 5 provide a view of how applications are distributed spatially across the city.

The planning_district variable may also interacts with application type and decision. Approval rates may vary by district, reflecting differences in local policies or development dynamics.
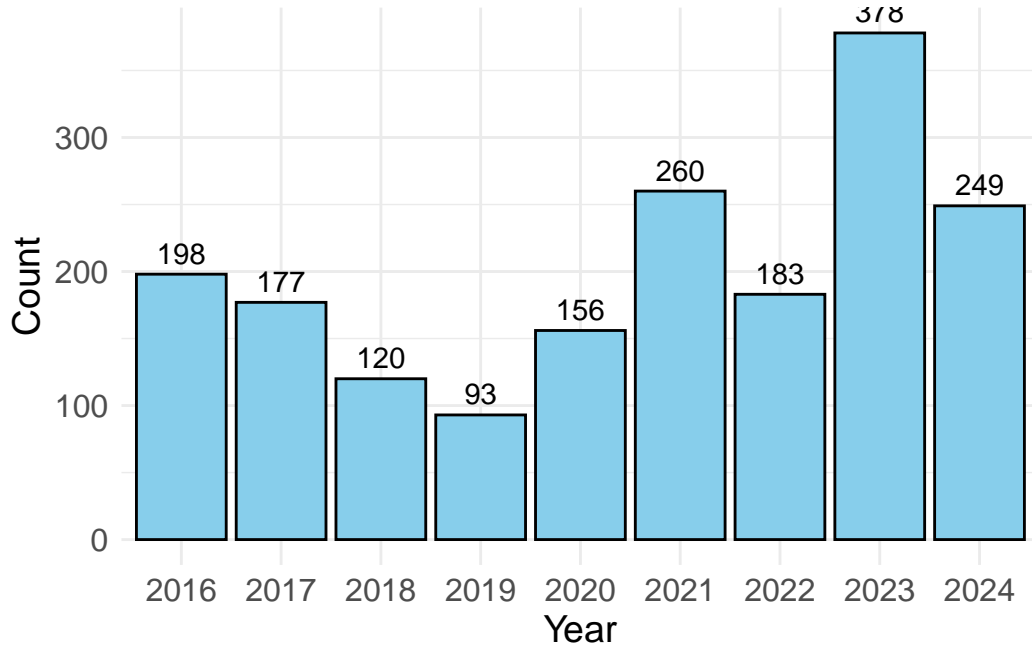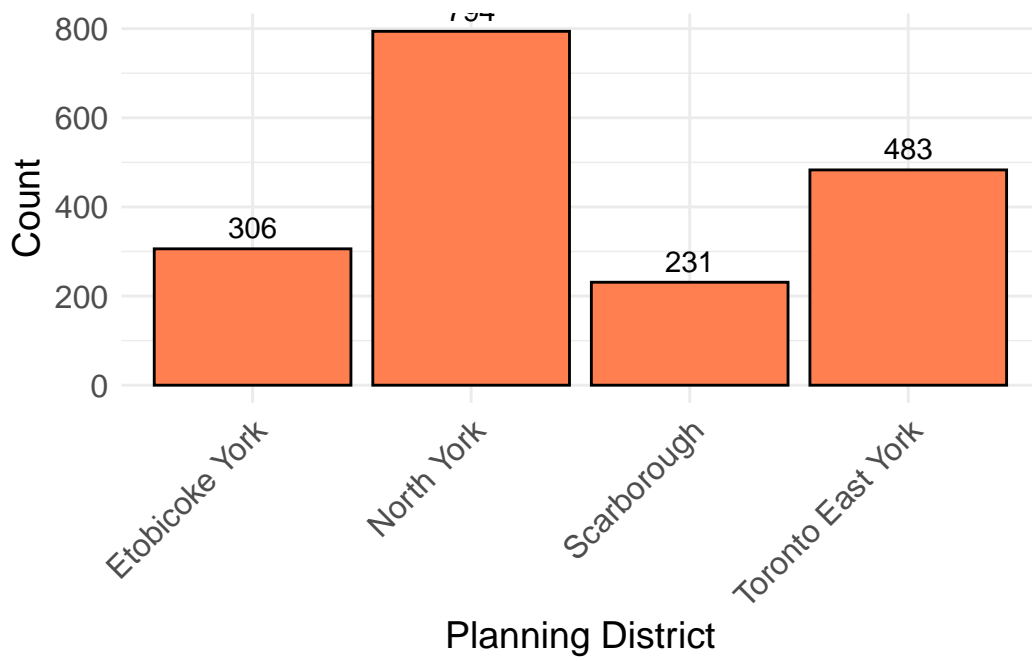
Figure 3: Distribution of Applications Years

Figure 4: Distribution of Planning District

The proportion of Minor Variances (MV) versus Consents (CO) may differ across districts, indicating distinct development priorities or zoning challenges.

3 Model

In this paper, a Bayesian logistic model will be utilized to understand how factors such as application type, submission year, and planning district influence the decisions of the application committee. This model estimates the probability of an application being approved (P(Approval)) based on the given predictors.

The goal of this modeling strategy is twofold: (1) to estimate how each predictor affects the likelihood of approval, and (2) to quantify the uncertainty associated with these estimates. This approach is well-suited for incorporating prior knowledge while addressing the hierarchical nature of the predictors, such as geographic regions.

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). `rstanarm` (Goodrich et al. 2022) is used to set up the model.`tidybayes` package (Kay 2024) is used to plot predicted probabilities in for result analysis.

Here we briefly describe the Bayesian analysis model used to investigate, and include justification for model and the variables, as well as discuss underlying assumptions, potential limitations, and evidence of model validation and checking.

Background details and diagnostics are included in Appendix B.

3.1 Model set-up

In the Bayesian logistic regression model, the response variable y follows a Bernoulli distribution, reflecting its binary nature. Coefficients β_k (where k=0,1,2,3) of every predictor variables follow a Normal distribution. Specifically:

$$y_i|p_i \sim \text{Bern}(p_i) \tag{1}$$
$$\beta_0 \sim \text{Normal}(0, 2.5) \tag{2}$$
$$\beta_1 \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta_2 \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\beta_3 \sim \text{Normal}(0, 2.5) \tag{5}$$

Combining all the components, the complete model can be expressed as:

$$\log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 \times \text{application type} + \beta_2 \times \text{year} + \beta_3 \times \text{planning district} \tag{6}$$

- p_i: the probability for the binary outcome variable to be 1 (1 = approval, 0 = refusal).
- β_0: Intercept, representing the baseline log-odds of approval for the reference levels of the predictors.
- β_1: Coefficients for application type, dummy-coded with CO as the reference category.
- β_2: Coefficient for year, measured as the year of application to capture temporal trends.
- β_3: Coefficients for planning district, dummy-coded with North York as the reference category.

In the above Bayesian logistic regression model, the coefficients $\beta = \beta_0, \beta_1, \beta_2, \beta_3$ are treated as random variables, each following a specified prior distribution. These priors represent our initial beliefs about the plausible values of the coefficients before observing the data. In this model, each β_k (where k=0,1,2,3) is assigned a normal distribution, with a zero mean and 2.5 standard deviation. The choice of these parameters reflects a weakly informative prior, centered around zero to express no strong prior belief about the direction or magnitude of the effects, while allowing for reasonable variability in the coefficient estimates.

The normal distribution is particularly suitable for these priors because it reflects a belief that most effects are likely to be small or moderate, centered around zero, while allowing for deviations in either direction. The standard deviation of 2.5 is a practical choice for weakly informative priors, offering a balance between constraining the parameters and allowing the data to influence the estimates. This ensures that the model avoids overfitting, particularly when data is sparse or multicollinearity exists among predictors.

After observing the data, the posterior distributions of β are obtained by combining these priors with the likelihood of the observed outcomes. The resulting posterior distributions provide updated beliefs about the coefficients, incorporating both prior knowledge and evidence from the data. This approach allows for robust inference, offering a clear quantification of parameter uncertainty and supporting more nuanced interpretations of the predictors' effects.

3.2 Model justification

The Bayesian logistic regression model is an appropriate choice for analyzing how application type, submission date, and planning district influence the likelihood of application approval. This choice is driven by the binary nature of the response variable, the model's ability to incorporate prior knowledge, and its probabilistic framework, which quantifies uncertainty more effectively than frequentist alternatives. Logistic regression naturally constrains predicted probabilities to the [0,1] interval, ensuring interpretability as likelihoods of approval, while the Bayesian approach enhances flexibility and provides richer insights through posterior distributions.

Compared to traditional generalized linear models (GLM), the Bayesian approach allows for the inclusion of prior beliefs through weakly informative priors ($N(0, 2.5)$) for all coefficients. These priors help stabilize the model, particularly in cases where data is sparse or multi-collinear, without overwhelming the contribution of the observed data. Additionally, Bayesian

methods yield full posterior distributions rather than single-point estimates, allowing for a clearer understanding of parameter uncertainty and a better capacity to incorporate uncertainty into decision-making processes. These advantages make the Bayesian logistic regression model preferable to a standard GLM, which provides point estimates and relies on asymptotic approximations for inference.

The model also stands out against simpler approaches, such as simple linear regression (SLR). While SLR might be used for binary outcomes in certain contexts, it is theoretically inappropriate because it does not constrain predictions to the [0,1] interval. This can result in nonsensical predicted probabilities outside this range. Moreover, SLR assumes a linear relationship between predictors and the response variable, which is unsuitable for binary outcomes. The logit transformation in logistic regression, by contrast, ensures a proper probabilistic framework while preserving interpretability in terms of log-odds.

The predictors included in the model further justify the choice of Bayesian logistic regression. Application type captures fundamental differences in how various categories of applications, such as minor variances and consents, might affect approval outcomes. Submission date is treated as a continuous variable, reflecting temporal trends without arbitrary grouping, while planning district accounts for geographic variability through categorical indicators. These predictors are naturally suited to a logistic framework, and the Bayesian approach accommodates any inherent variability in their effects.

In summary, Bayesian logistic regression provides a robust, clarified, and flexible framework for analyzing the factors influencing application decisions. Its advantages over GLM and SLR include the ability to incorporate prior information, quantify uncertainty, and handle the non-linear nature of the binary response variable. By balancing methodological rigor with practical transparency, this model ensures that the analysis remains both statistically sound and actionable for stakeholders.

3.3 Model Assumption

This Bayesian model relies on several assumptions that ensure the validity of its predictions and the transparency of its coefficients. While these assumptions are generally less restrictive than those for simpler models like linear regression, they remain critical to the robustness of the results. The key assumptions are listed below:

1. Binary Response Variable The model assumes that the response variable (y) is binary, taking values of either 1 (approval) or 0 (refused). This assumption aligns with the nature of the data, as application decisions are dichotomous outcomes. Any deviations from binary coding would invalidate the logistic framework and require alternative modeling approaches.

2. Independence of Observations: The model assumes that all observations are independent of one another. This is reasonable for application data, as each decision is typically

made independently by the committee. However, if clustering or dependence exists (e.g., decisions within the same planning district are correlated), the model may need to incorporate hierarchical or random effects to address these dependencies.

3. No Perfect Multicollinearity Logistic regression assumes that the predictors are not perfectly correlated, as this would prevent the model from estimating unique coefficients. For example, if a predictor is a linear combination of other predictors, the model would fail to converge. In this analysis, categorical variables of application type and planning district are appropriately encoded to avoid such issues.

4. Proper Model Specification The model assumes that all relevant predictors are included and correctly specified. Omitting important variables or including irrelevant ones could lead to biased estimates or reduced transparency. For example, excluding interaction terms when they are theoretically justified might result in incomplete understanding of the predictors' effects.

3.4 Potential Limitation and Addressing Violation

Violations of the above assumptions can impact this model's validity while severl solution are listed below:

1. Non-linearity: Adding polynomial terms or splines could address non-linear relationships if diagnostic checks suggest deviations.

2. Dependence: If observations are not independent (e.g. decisions clustered by region), hierarchical models or random effects may be introduced.

3. Multicollinearity: Variance inflation factors or correlation matrices can help identify multicollinearity issues, which can be resolved by removing or combining correlated predictors.

By adhering to these assumptions and addressing potential violations, the Bayesian logistic regression model provides a robust and effective framework for analyzing the factors influencing application outcomes.

3.5 Model Validation

To ensure the robustness and reliability of the Bayesian logistic regression model, a series of validation and diagnostic steps were conducted. Detailed plots and results, including posterior predictive checks (PPC), posterior vs. prior comparisons, and convergence diagnostics, are provided in the appendix for further examination. Below, the focus is on other key validation methods:

1. **Out-of-Sample Testing**

The dataset was split into a training set (80%) and a test set (20%) to evaluate the model's predictive performance on unseen data. The model was trained on the training set, and predictions were made on the test set. Classification accuracy was calculated as the proportion of correctly classified outcomes in the test set. This approach assesses the model's generalizability.

2. Confusion Metrix

3. Sensitivity Analysis The sensitivity of the model was assessed by varying the priors for the coefficients to examine whether the results were robust to different prior specifications. The predictions remained stable across different prior choices, indicating that the model is not overly influenced by prior assumptions.

4. Predictive Power (AUC-ROC) The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) was computed to evaluate the model's ability to discriminate between approved and refused applications. A high AUC-ROC value (>0.8) confirmed the model's strong discriminatory performance.

4 Results

Our results are summarized in Table 6.

The result highlights how the predictors influence the likelihood of application approval while controlling for other variables. The coefficients, presented as log-odds, provide insights into both the direction and magnitude of each predictor's impact on the approval decision.

The intercept estimate of -0.06 represents the baseline (an application of CO subimited in 2016 at Etobicoke York) log-odds of approval when all predictors are at their reference levels. This corresponds to a baseline probability close to 0.5, indicating nearly equal odds of approval or refusal in the absence of other influencing factors. This serves as a neutral starting point for understanding how the predictors shift the likelihood of approval.

The variable application_typeMV, with a coefficient of -0.20, suggests that applications for minor variances (MV) are less likely to be approved compared to consents (CO), which act as the reference category. In terms of odds, the transition from a consent to a minor variance reduces the odds of approval by approximately 18% ($e^{-0.20} \approx 0.82$). This finding aligns with expectations, as minor variances may face stricter scrutiny due to their potentially greater impact on local zoning adjustments.

The year variables reveal significant temporal variations in approval likelihood, using 2016 as the reference year. These coefficients highlight how approval dynamics have shifted over time. For instance, in 2017, the coefficient of 0.74 indicates a 109% increase in the odds of approval compared to 2016 ($e^{0.74} \approx 2.09$), which suggests a substantial policy or procedural shift. In 2018, the strongest positive effect is observed, with a coefficient of 0.96, reflecting a 161%

15

Table 6: Model results of Application Approval Probability

	Model
(Intercept)	−0.06
	(0.22)
application_typeMV	−0.20
	(0.13)
year2017	0.74
	(0.22)
year2018	0.96
	(0.25)
year2019	0.22
	(0.26)
year2020	0.56
	(0.23)
year2021	0.34
	(0.20)
year2022	0.28
	(0.21)
year2023	0.99
	(0.20)
year2024	1.24
	(0.22)
planning_districtNorth York	0.41
	(0.15)
planning_districtScarborough	0.16
	(0.18)
planning_districtToronto East York	0.33
	(0.16)
Num.Obs.	1814
R2	0.045
Log.Lik.	−1121.092
ELPD	−1134.6
ELPD s.e.	15.9
LOOIC	2269.3
LOOIC s.e.	31.8
WAIC	2269.2
RMSE	0.46

increase in odds relative to 2016 ($e^{0.96} \approx 2.61$). However, in 2019, the coefficient drops to 0.22, representing a smaller increase of 25% in odds ($e^{0.22} \approx 1.25$), suggesting that this period may have faced stricter reviews or procedural adjustments. By 2020, the coefficient rebounds to 0.56, translating to a 75% increase in odds ($e^{0.56} \approx 1.75$).

For more recent years, 2021 sees a coefficient of 0.34, showing a 40% increase in odds ($e^{0.34} \approx 1.40$). In 2022, the coefficient is 0.28, reflecting a 32% increase ($e^{0.28} \approx 1.32$). In 2023, approvals surged again, with a coefficient of 0.99, representing a 170% increase in odds ($e^{0.99} \approx 2.69$). Finally, 2024 showed the highest increase, with a coefficient of 1.24, translating to a 245% increase in approval odds compared to the baseline year of 2016 ($e^{1.24} \approx 3.46$). This steady upward trend reflects systemic improvements in the approval process, possibly due to increased efficiency, evolving zoning policies, or external influences favoring development.

The model captures geographic variability in approval rates across Toronto's planning districts. North York, with a coefficient of 0.41, shows a 50% increase in odds compared to the baseline district ($e^{0.41} \approx 1.50$), highlighting its role as a development hub and its alignment with urban growth priorities. Scarborough, on the other hand, has the lowest approval likelihood, with a coefficient of 0.16, indicating a modest 17% increase in odds ($e^{0.16} \approx 1.17$), suggesting challenges related to stricter zoning rules or community resistance to changes. Toronto East York has a coefficient of 0.33, showing a 39% increase in odds ($e^{0.33} \approx 1.39$), reflecting its mix of heritage preservation and urban development priorities. The results emphasize that approval rates vary significantly across districts due to localized zoning policies, development pressures, and socio-political factors influencing decision-making.

In summary, the model reveals key patterns in how application type and submission year influence the likelihood of approval. Minor variance applications face slightly lower odds of approval compared to consents, while temporal trends indicate an overall increase in approval rates from 2016 to 2020. These findings provide a foundation for understanding the dynamics of approval decisions and highlight areas for further exploration, such as geographic variability or interactions between predictors.

Figure 5 demonstrates variability in the predicted probability of approval between consent applications (CO) and minor variances (MV). Consent applications exhibit a slightly higher median predicted probability of approval compared to minor variances. The interquartile range (IQR) for both application types is comparable, indicating a similar spread of probabilities. However, minor variance applications show a broader lower whisker, suggesting that certain cases have substantially lower probabilities of approval. This supports the interpretation that minor variances may face stricter scrutiny due to their greater potential impact on zoning.

Figure 6 explores temporal trends in the predicted probability of approval from 2016 to 2024. The median predicted probability of approval rises steadily from 2016 to 2024. While 2016 exhibits the lowest probabilities, years such as 2018 and 2024 display higher medians, indicating increased approval likelihood in recent years. Variability within years remains relatively stable, with slight fluctuations in the spread across the interquartile range. This pattern aligns with
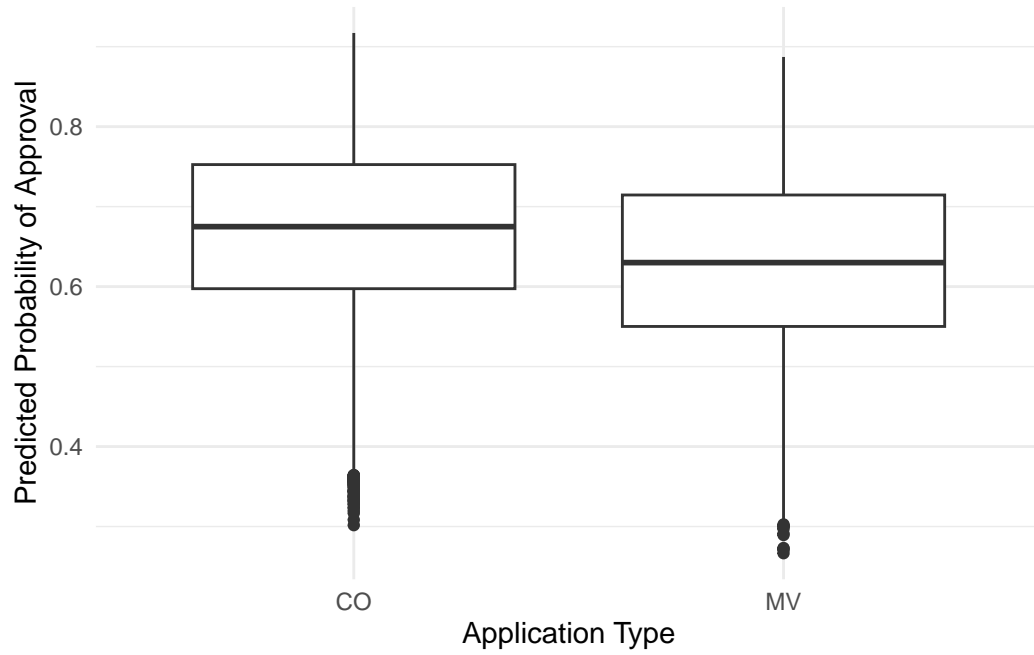
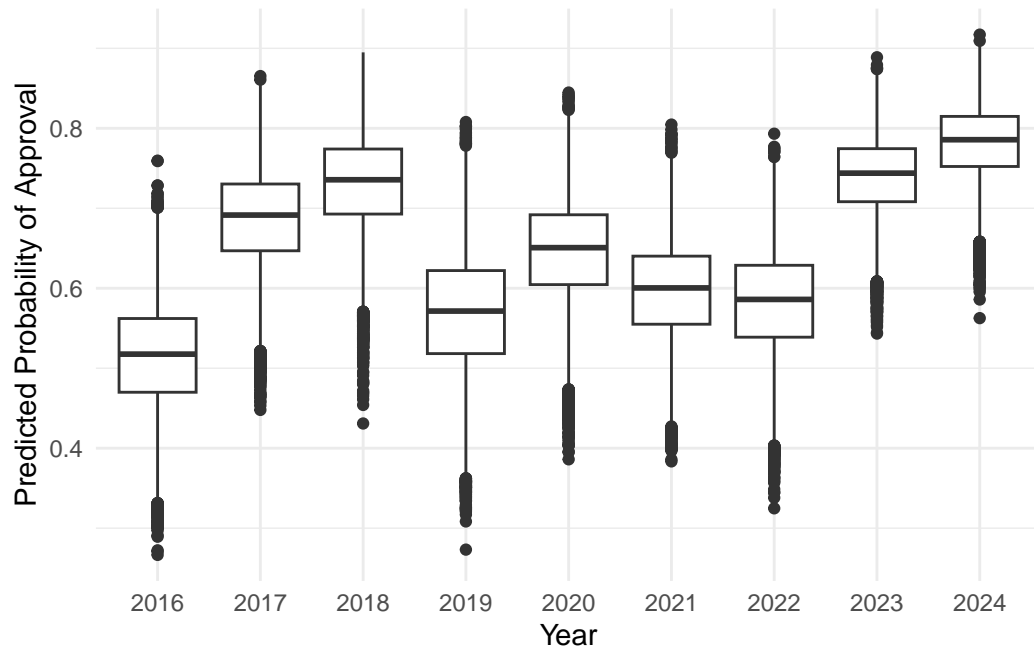Figure 5: Predicted Probability of Approval by Application Type

Figure 6: Predicted Probability of Approval by Year

findings from the regression analysis, where temporal predictors indicated a gradual improvement in approval rates over time. This likely reflects policy or procedural changes that have positively influenced the decision-making process.
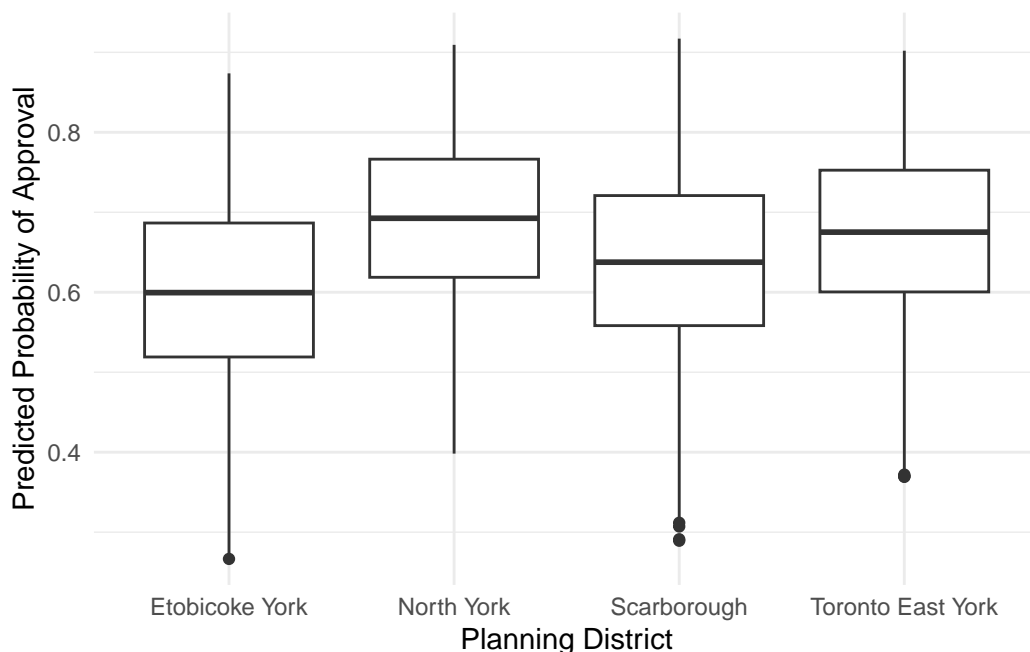

Figure 7: Predicted Probability of Approval by Planning District

Figure 7 evaluates geographic variability in approval probabilities across the four planning districts: Etobicoke York, North York, Scarborough, and Toronto East York. North York consistently shows the highest median predicted probability of approval, reaffirming its status as a development hub with favorable zoning outcomes. In contrast, Scarborough exhibits the lowest median and a wider spread of probabilities, with outliers indicating challenging cases. Toronto East York and Etobicoke York present intermediate medians, suggesting more balanced approval probabilities in these regions. These geographic differences highlight the influence of localized development dynamics and zoning considerations on approval outcomes.

5 Discussion

5.1 Overview

This paper investigates the factors influencing application approvals submitted to the City of Toronto's Committee of Adjustment, using a Bayesian logistic regression model. By analyzing predictors such as application type, submission year, and planning district, the study provides insights into the dynamics of urban planning decisions. The model captures temporal trends,

spatial variability, and procedural differences, offering a structured understanding of approval outcomes. Visualizations, including coefficient plots and effect plots, complement the analysis, helping to interpret the relationships between predictors and approval likelihood.

5.2 Temporal Trends in Approval Rates

One key finding is the temporal trend in approval rates. The analysis reveals that approval odds have increased steadily over time, peaking in 2018. This trend may reflect evolving planning policies, greater efficiency in committee processes, or a shift in zoning priorities. The data suggests that the Committee of Adjustment has become more permissive in granting approvals, which could indicate a growing alignment between applications and city planning objectives. This result highlights how administrative processes adapt over time to urban development needs.

5.3 Geographic Disparities in Approval Decisions

Spatial variability plays a significant role in application approvals. The findings show that North York consistently exhibits the highest approval rates, likely due to its prominence as a development hub with infrastructure supporting large-scale projects. Conversely, Scarborough has the lowest odds of approval, suggesting stricter zoning enforcement or less alignment with development objectives in this area. This geographic disparity underscores the influence of local contexts and priorities on urban planning decisions, revealing how different districts balance growth and zoning compliance.

5.4 Weaknesses

While this paper provides insights into the factors influencing application approvals, there exists several limitations. These weaknesses cover data-related issues, methodological constraints, and broader contextual considerations.

1. **Data Limitations** The analysis relies solely on the dataset provided by the City of Toronto's Committee of Adjustment, which, while complete in its coverage of applications, lacks certain critical variables. Key contextual factors, such as community objections, project complexity, or the presence of political or economic pressures, are not captured. For instance, the dataset does not provide information about why certain applications are refused or approved beyond the recorded decision, limiting the ability to analyze qualitative or situational influences.

2. **Methodological Constraints** The Bayesian logistic regression model assumes that the relationships between predictors and the outcome variable are additive and linear on the log-odds scale. This assumption may oversimplify real-world interactions between

variables, such as how specific planning districts may interact with application types or temporal trends. Furthermore, the model does not fully account for potential multicollinearity among predictors, which could distort the estimated coefficients. While priors were carefully chosen, their influence on the results might vary with a different prior specification, introducing some degree of subjectivity into the analysis.

3. **Simplified Representation of Approval Dynamics** The approval process involves a combination of bureaucratic procedures, stakeholder negotiations, and technical considerations that are not fully captured in this analysis. For example, public hearings, lobbying efforts, and input from local communities likely play a significant role in shaping decisions but are absent from the dataset. This simplification means that the model captures only part of the story, potentially overlooking nuanced dynamics in urban planning and governance.

4. **Geographic and Temporal Generalizability** The analysis focuses exclusively on the City of Toronto, meaning the findings are context-specific and may not generalize to other municipalities with different urban planning frameworks, policies, or demographic pressures. Temporal trends in approval rates, for example, could be influenced by unique local events or policy changes that are not representative of broader patterns. Similarly, the observed geographic variability in approval rates might reflect Toronto-specific zoning challenges rather than universal principles.

5.5 Nest Steps

Building on the findings and addressing the limitations of this paper, there are several avenues for future research and analysis that could provide a deeper understanding of the factors influencing application approvals in urban planning contexts.

One critical next step is to expand the dataset by incorporating additional variables that capture contextual and qualitative factors influencing decisions. Data on community feedback and objections during public hearings is a good choice since it could provide insights into how local resistance or support impacts approval likelihood. Information on project complexity, such as the scale, type, and potential impact of proposals, could reveal patterns in decision-making. Besides, access to committee meeting notes or further exploration in meeting minutes would help clarify the reasoning behind specific approvals or refusals. Furthermore, integrating economic and political context, such as local economic indicators or changes in political leadership, would offer a broader understanding of approval trends.

The current model assumes additive relationships between predictors and approval likelihood, which may oversimplify real-world dynamics. Future research could explore nonlinear relationships, such as whether certain predictors have diminishing or exponential effects on approval rates. Interaction effects between predictors, such as the interplay between planning districts and application types, could also provide more nuanced insights. For instance, minor variances may be treated differently in densely populated urban areas compared to suburban districts.

Extending this analysis to other municipalities or regions would help determine whether the observed trends in Toronto are unique or indicative of broader urban planning practices. Cross-regional comparisons could identify common principles across jurisdictions, highlight local deviations driven by distinct policies or demographics, and establish benchmarks for understanding how Toronto's process aligns with other cities. Such studies could enhance the generalizability of findings and provide practical lessons for urban planners in different contexts.

Despite to future improvements for addressing limitations, further exploration can be continued based on the current result. The temporal trends identified in this paper suggest shifts in approval rates over time, but the reasons behind these changes remain unexplored. Future work could examine specific policy changes or regulatory updates to understand their impact on approval trends. Additionally, external events such as economic downturns, housing crises, or infrastructure investments could be analyzed for their influence on application volumes and decisions. A difference-in-differences approach could also help isolate the effects of significant policy interventions, providing a clearer picture of causal relationships.

Given the importance of geographic variability in approval rates, future research could adopt spatial analysis techniques. Spatial regression models could account for spatial autocorrelation and clustering effects in planning districts, while network analysis could study relationships between districts, such as spillover effects of development activity. Additionally, analyzing proximity to key infrastructure, such as transit hubs or commercial centers, could provide insights into spatial patterns of decision-making.

Finally, the findings from this paper could inform the development of predictive tools to assist urban planners and developers. These tools could provide probabilistic forecasts of approval likelihood based on project attributes and contextual factors. Additionally, they could help committees identify systemic biases or inefficiencies in their decision-making processes and aid policymakers in designing more equitable and transparent approval frameworks.

In summary, future research should prioritize richer data integration, explore more sophisticated modeling techniques, and broaden the scope of analysis to other regions and temporal contexts. These steps would not only enhance academic understanding of urban planning decisions but also provide actionable insights for policymakers, developers, and urban planners working toward effective and equitable processes.

Appendix

A Additional data cleaning details

The dataset underwent a structured cleaning process to ensure its suitability for analysis, focusing on preparing the data while minimizing bias and enhancing interpretability. Below is a detailed description of the key cleaning steps:

1. **Filtering Unused Decision Outcomes**

The dataset originally included a range of decision outcomes beyond approvals and refusals, such as applications that were deferred or withdrawn. These were excluded to focus solely on cases where the Committee of Adjustment made a definitive decision. This exclusion is done by filtering decisions with "approve", "approval", and "refused" words in it while conditionally approved are all default as approval to enlarge the data size. This ensured a binary structure for the decision variable, aligning with the study's objectives.

2. **Conversion of decision to Binary Format**

The decision variable was converted into a binary format for consistency and ease of analysis:

1: Representing approved applications. 0: Representing refused applications. This transformation provided a clear and standardized representation of application outcomes.

3. **Creation of the year Variable**

A new variable, year, was created by extracting the year component from the date field. This allowed for an analysis of temporal trends without losing the temporal resolution of the data. The original date variable remained intact for potential use in supplementary analyses.

4. **Filtering Applications from Earlier Years** Applications before 2016 were removed from the dataset due to their low counts. Retaining these outliers could lead to biased results, particularly in trend analyses or modeling. By excluding these years, the dataset better represents the more recent and consistent patterns of decision-making.

5. **Removing Missing Values** Records with missing values in the chosen variables (decision, application_type, year, and planning_district) were removed.

B Model details

B.1 Posterior predictive check

In Figure 8a we implement a posterior predictive check. This shows the model's ability to generate data consistent with the observed data. By overlaying the observed data and simulated replicated datasets generated from the posterior predictive distribution, we assess the model's adequacy. The close alignment between the observed data and the majority of the simulated replicated datasets indicates that the model captures the underlying structure of the data well. Deviations would signal potential misfits or areas where the model assumptions may need to be revised.

In Figure 8b we compare the posterior with the prior. This comparison highlights how the data influenced parameter estimates. Large shifts from the prior to the posterior, such as seen for certain planning districts and years, suggest strong data-driven updates. Conversely, parameters where the posterior closely resembles the prior indicate minimal influence from the observed data. This evaluation helps assess the informativeness of the priors and the robustness of the parameter estimates, ensuring the model is both data- and prior-consistent.

B.2 Diagnostics

Figure 9a is a trace plot. It shows he sampled values for each chain of the Markov Chain Monte Carlo (MCMC) simulation over iterations. The chains appear to mix well and explore the parameter space without any evident trends or stickiness, suggesting convergence. The consistent overlap of chains indicates that the posterior distribution has been effectively sampled, reducing the risk of bias due to poor chain mixing or non-convergence.

Figure 9b is a Rhat plot. It evaluates the Gelman-Rubin convergence diagnostic. All parameters show $\hat{R}$ values close to 1, which indicates that the between-chain and within-chain variances are nearly identical. This confirms that the MCMC chains have converged and are sampling from the same target distribution, validating the reliability of the posterior estimates.

B.3 Idealized Methodology and Survey

B.3.1 Overview

This section outlines an idealized methodology and survey design to supplement the observational dataset used in this study. The objective is to enhance the analysis of urban planning decisions by incorporating a carefully designed survey targeting key stakeholders in Toronto's zoning and land-use processes.
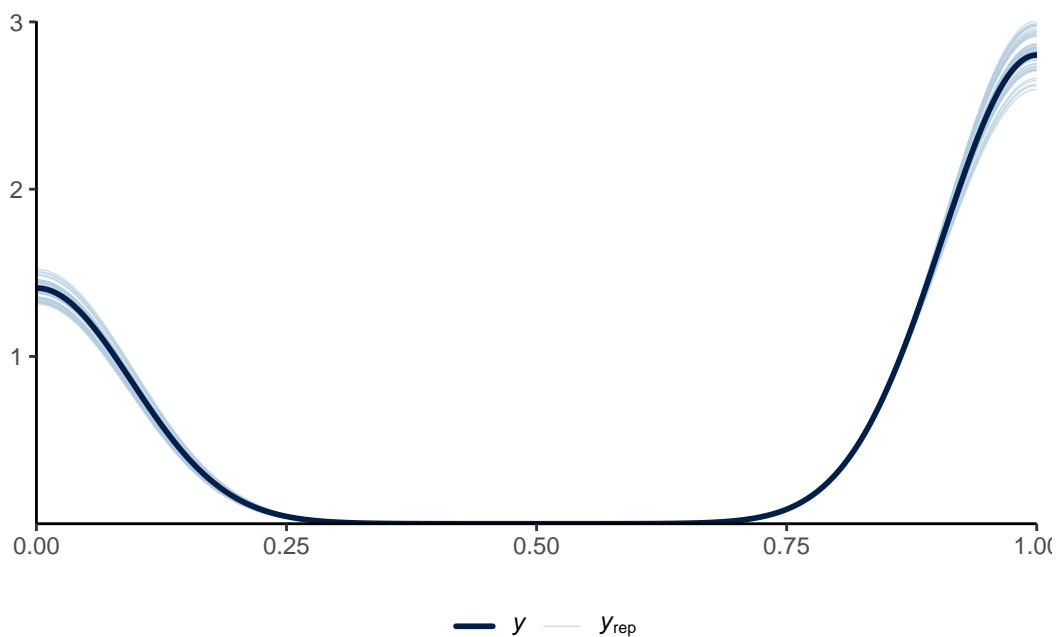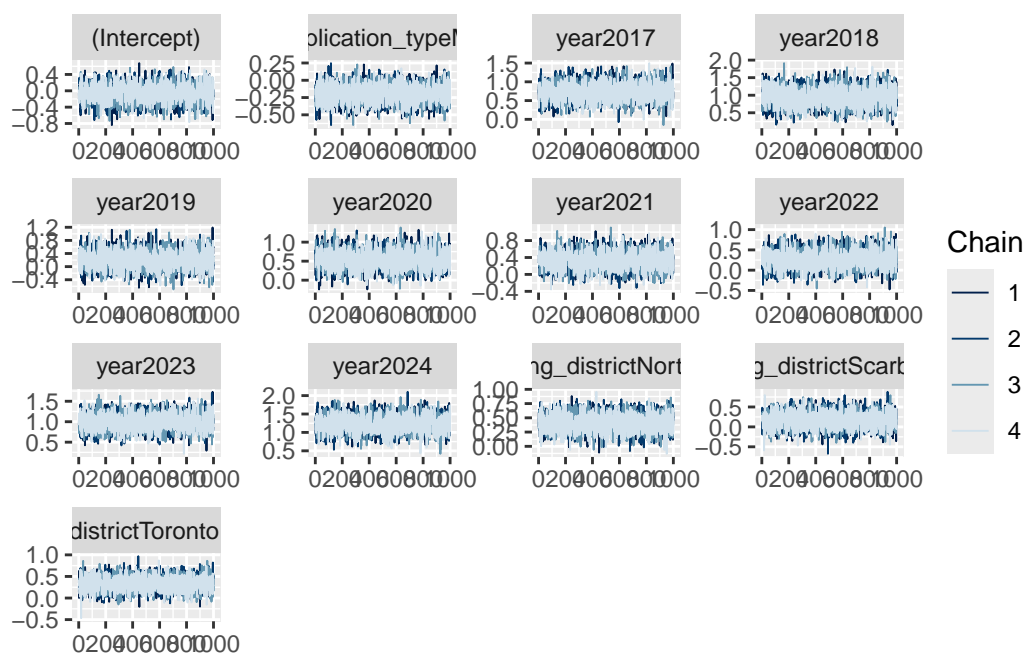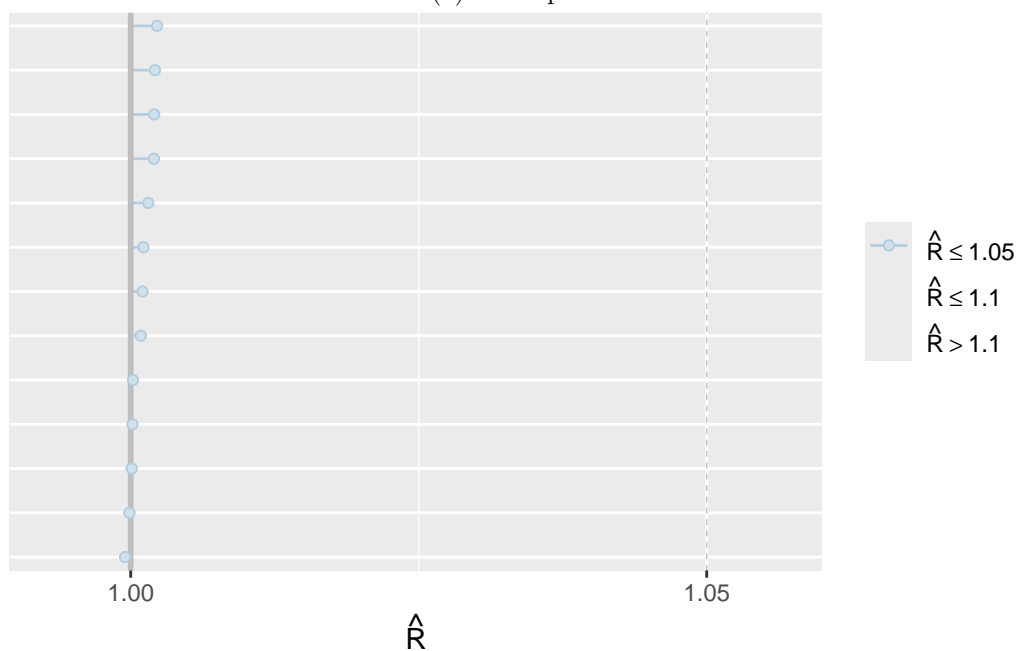
(a) Posterior prediction check
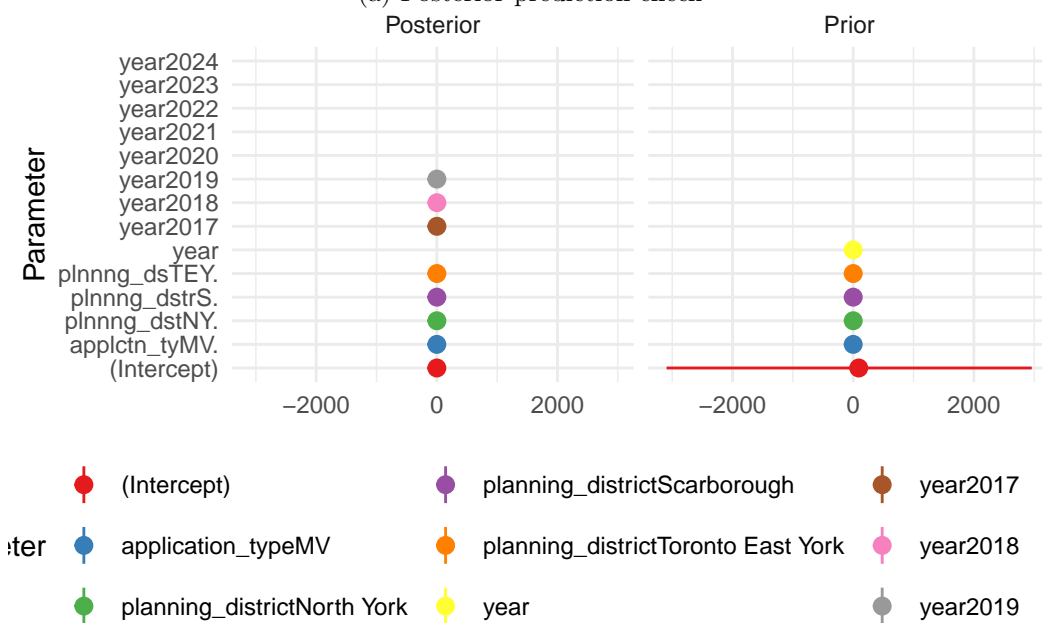
(b) Comparing the posterior with the prior

Figure 8: Examining how the model fits, and is affected by, the data

(a) Trace plot

(b) Rhat plot

Figure 9: Checking the convergence of the MCMC algorithm

B.3.2 Sampling Approach

B.3.2.1 Target Population

The target population comprises stakeholders involved in the Committee of Adjustment's decision-making process, including:

- **Property Owners and Developers**: Individuals or entities who submit applications for minor variances or consents.
- **Urban Planners and Architects**: Professionals who prepare applications or advise clients on zoning requirements.
- **Committee Members and Staff**: Officials responsible for evaluating applications and making decisions.
- **Community Members**: Residents who may provide feedback during public hearings.

B.3.2.2 Sampling Frame

The sampling frame would be constructed using publicly available application records and contact information from municipal directories. This ensures representation from all planning districts and captures diverse perspectives across geographic and professional domains.

B.3.2.3 Sample Size

Using stratified random sampling, the sample size would aim for approximately 3000 respondents, ensuring statistical validity while considering feasibility.

B.3.2.4 Stratified Random Sampling

Stratified random sampling enhances the accuracy of survey estimates because it ensures proportional representation of different demographic groups. This minimizes the risk of sampling bias, where certain groups may be overrepresented or underrepresented, skewing the results.

Hence, to achieve a representative sample, stratified random sampling will be employed. This method involves dividing the population into homogeneous subgroups based on key demographic characteristics, then randomly selecting samples from each stratum proportionate to its size in the population. The strata will include:

- Application types: Categorize applicants based on the purpose of their submission. (Minor Variance and Consent)
- Planning districts: Stratify by Toronto's four major planning districts to capture geographic diversity. (North York, Scarborough, Toronto East York, and Etobicoke York)
- Applicant roles: Group participants by their involvement in the application process. (Individual Property Owners, Developers, and Consultants)

Within these strata, respondents will be randomly selected to align with the distribution of these categories in the broader population. The process begins by grouping the population into strata defined by characteristics such as application type, planning district, and applicant role. Afterward, random samples are drawn independently from each subgroup, ensuring proportional representation. In cases where smaller groups are critical to the study, adjustments can be made to ensure adequate representation. Finally, the selected samples from each subgroup are combined to create the overall sample, ensuring that all significant categories are included.

This method reduces sampling variability, enhances representation, and provides more accurate estimates compared to simple random sampling, as the unique characteristics of each subgroup are reflected in the final sample.

B.3.2.5 Trade-offs

While stratified random sampling ensures proportional representation and enhances the precision of survey estimates, it is not without challenges. One notable trade-off is the complexity involved in defining and managing multiple strata, particularly when the strata are based on multiple dimensions such as application type, planning district, and applicant role. This complexity increases the time and resources required for sample design and execution. For instance, ensuring proportional representation across strata like **Scarborough developers** or **North York consultants** may necessitate oversampling in smaller groups, leading to additional effort in recruitment and logistical coordination.

Another trade-off lies in the potential for overstratification, which can result in small sample sizes within certain strata. If some strata, such as **Scarborough Consent applications**, have very few members, it may be difficult to draw statistically meaningful conclusions. This could necessitate adjustments, such as combining smaller strata or applying post-stratification weighting, which could introduce bias or reduce interpretability. Additionally, the method assumes that the characteristics used to define strata are adequately known and relevant, which may not account for unobserved variables that influence the outcomes, such as the socio-political environment or project complexity. Despite these challenges, the benefits of stratified random sampling in terms of improved accuracy and subgroup analysis make it a robust choice for this study.

B.3.3 Data Validation

Survey responses would be validated through cross-referencing with the cleaned dataset. For instance:

Matching survey-reported outcomes with recorded decisions. Verifying consistency between applicant-reported application details and dataset attributes (e.g., application type, planning

district). To ensure representativeness, response rates would be monitored, and additional recruitment efforts would target underrepresented groups or districts.

B.3.4 Survey

This survey will be conducted using Google Forms, which is an effective platform for data collection. The survey can be accessed by the link, Google Form Survey.

B.3.4.1 Survey Structure

Title:

Understanding Urban Planning Decisions

Introduction:

I appreciate your participation in this survey, which seeks to better understand the factors influencing urban planning decisions and application outcomes in Toronto. Your responses are crucial for enhancing the research and contributing to insights that could inform future urban planning processes.

Please note:

- Your answers will be treated with complete confidentiality.
- Participation in this survey is voluntary.
- You are encouraged to provide honest and thoughtful responses.
- The survey is estimated to take about 5 minutes to complete.
- If you have any questions or concerns, feel free to reach out to me at anjojoo.xu@mail.utoronto.ca (Angel Xu).

Thank you for your valuable contribution! As a token of our appreciation, each participant will receive $5 upon completion of the survey.

Section 1: Application Details

What type of application did you submit?

- MV (Minor Variance)

- CO (Consent)

- Prefer not to say

Which planning district was your application associated with?

- North York

- Scarborough

- Etobicoke York
- Toronto East York
- Prefer not to say

Which year did you submit your application?

- 2016
- 2017
- 2018
- 2019
- 2020
- 2021
- 2022
- 2023
- 2024
- Other (please specify)

What was the outcome of your application?

- Approved
- Refused
- Prefer not to say
- Others (please specify)

Section 2: Applicant Background

What role best describes your involvement in the application process?

- Property Owner
- Developer
- Consultant
- Prefer not to say
- Others (please specify)

How many years of experience do you have in urban planning or related fields?

- <1 year

- 1–5 years

- 5–10 years

- 10 years

- Prefer not to say

Section 3: Process and Challenges

What was the primary purpose of your application?

- Residential Adjustment

- Commercial Development

- Mixed-Use Development

- Prefer not to say

- Others (please specify)

What challenges did you encounter during the application process?

- Zoning Compliance

- Documentation Requirements

- Public Objections

- Prefer not to say

- Others (please specify)

Section 4: Influences on Decisions

Do you believe community feedback impacted the outcome of your application?

- Yes

- No

Were there any external factors (e.g., economic conditions, policy changes) that influenced the decision?

- Yes

- No

Do you have any further comments or insights about the factors that might affect the application decision?

[answer box]

Section 5: Feedback and Improvements

What improvements would you recommend for the application process?

- Streamlining Documentation
- Clearer Guidelines
- Enhanced Community Engagement
- Others (please specify)

Do you have any further comments or insights about the application process?

d[answer box]

End Message: Thank You for Completing the Survey!

We sincerely appreciate your time and thoughtful responses. Your participation is essential in helping us gain a deeper understanding of the factors influencing urban planning decisions in Toronto. Your insights will contribute to research aimed at improving zoning application processes and understanding decision dynamics.

If you have any questions or would like to learn more about this study, please don't hesitate to contact our research team at anjojoo.xu@mail.utoronto.ca (Angel Xu).

As a token of our gratitude, each participant will receive a $5 reward shortly after completing the survey. Thank you again for your valuable contribution!

B.3.5 Surevey Design Consideration

When designing the survey, several key factors were considered to ensure representativeness, clarity, and relevance to the study's objectives. The survey targets individuals involved in Toronto's urban planning processes, including property owners, developers, and consultants.

Question phrasing was tailored to align with terminology familiar to the target respondents, minimizing ambiguity. Efforts were also made to limit the survey's length to encourage participation while ensuring sufficient data for meaningful analysis. A balance was struck between closed-ended questions, which facilitate quantitative analysis, and open-ended questions, which capture nuanced qualitative insights. Additionally, the survey design accounted for ethical considerations, such as maintaining confidentiality and emphasizing voluntary participation.

B.3.6 Strength

The survey design has several strengths. Stratified random sampling ensures that the results are representative of the population, reducing sampling bias. By aligning strata with key demographic and process variables, the survey facilitates detailed subgroup analysis, enhancing the depth of insights. The inclusion of both quantitative and qualitative questions allows for an overall understanding of the factors influencing urban planning decisions. Closed-ended questions enable statistical analysis, while open-ended responses provide richer context and identify factors not captured in the observational dataset.

The survey also addresses gaps in the cleaned dataset, such as perceptions of community feedback, challenges during the application process, and external influences on decisions. These insights complement the quantitative data, offering a more holistic view of the decision-making process.

B.3.7 Weakness

Despite its strengths, the survey has some limitations. Stratified random sampling requires precise knowledge of the population's composition, which may not always be accurate, particularly for less-documented strata. Smaller strata, such as applications from certain districts or less common applicant roles, may result in small sample sizes, limiting statistical power. Additionally, reliance on self-reported data introduces the potential for response bias, as participants may overstate or understate certain factors, particularly regarding the impact of community feedback or external influences.

The survey's limited length, while improving completion rates, may constrain the depth of responses, especially for complex topics. Furthermore, the survey results are context-specific, reflecting Toronto's unique regulatory framework and urban planning dynamics, which may limit the generalizability of findings to other cities or contexts. Despite these challenges, the survey is a valuable tool for supplementing the dataset and addressing research objectives.

References

Alexander, Rohan. 2023. *Telling Stories with Data.* Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

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.

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

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

Hester, Jim, Florent Angly, Russ Hyde, Michael Chirico, Kun Ren, Alexander Rosenstock, and Indrajeet Patil. 2024. *Lintr: A 'Linter' for r Code.* https://CRAN.R-project.org/package=lintr.

Kay, Matthew. 2024. *tidybayes: Tidy Data and Geoms for Bayesian Models.* https://doi.org/10.5281/zenodo.1308151.

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

Müller, Kirill, and Lorenz Walthert. 2024. *Styler: Non-Invasive Pretty Printing of r Code.* https://CRAN.R-project.org/package=styler.

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/.

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.

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.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.* https://CRAN.R-project.org/package=knitr.