

Understanding Factors Contributing to Application Approval in Urban Planning Decisions*

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

Angel Xu

November 27, 2024

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

Table of contents

1	Introduction	1
1.1	Estimand	2
2	Data	2
2.1	Overview	2
2.2	Measurement	3
2.3	Data Cleaning	4
2.4	Features	4
2.5	Outcome variables	5
2.6	Predictor variables	6
2.6.1	Application type	6
2.6.2	Year	7
2.6.3	Planning District	8
3	Model	10
3.1	Model set-up	10
3.1.1	Model justification	11
3.1.2	Model Assumption	12

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

3.1.3	Potential Limitation and Addressing Violation	13
4	Results	13
5	Discussion	15
5.1	First discussion point	15
5.2	Second discussion point	15
5.3	Third discussion point	15
5.4	Weaknesses and next steps	15
	Appendix	16
A	Additional data cleaning details	16
B	Model details	17
B.1	Posterior predictive check	17
B.2	Diagnostics	17
	References	18

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). 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 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

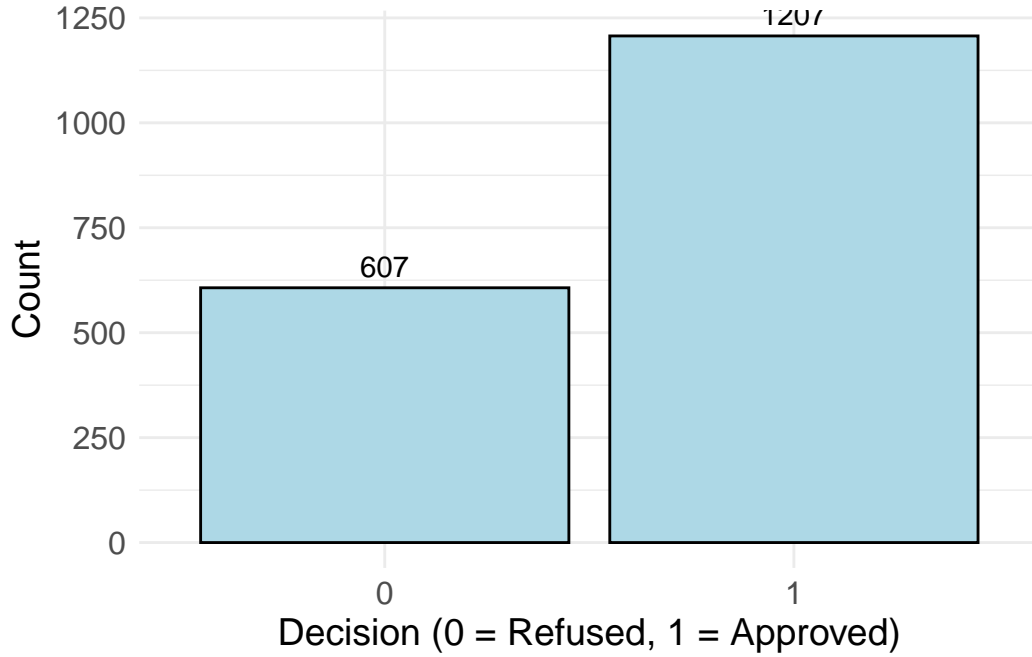


Figure 1: Distribution of Application Decisions

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.

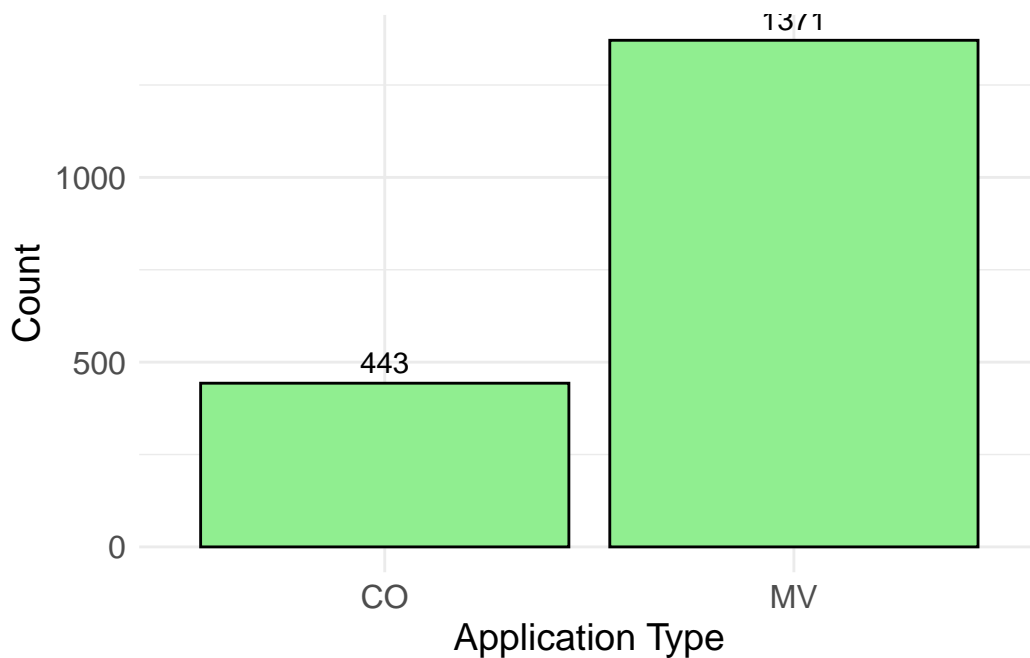


Figure 2: Distribution of Application Types

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.

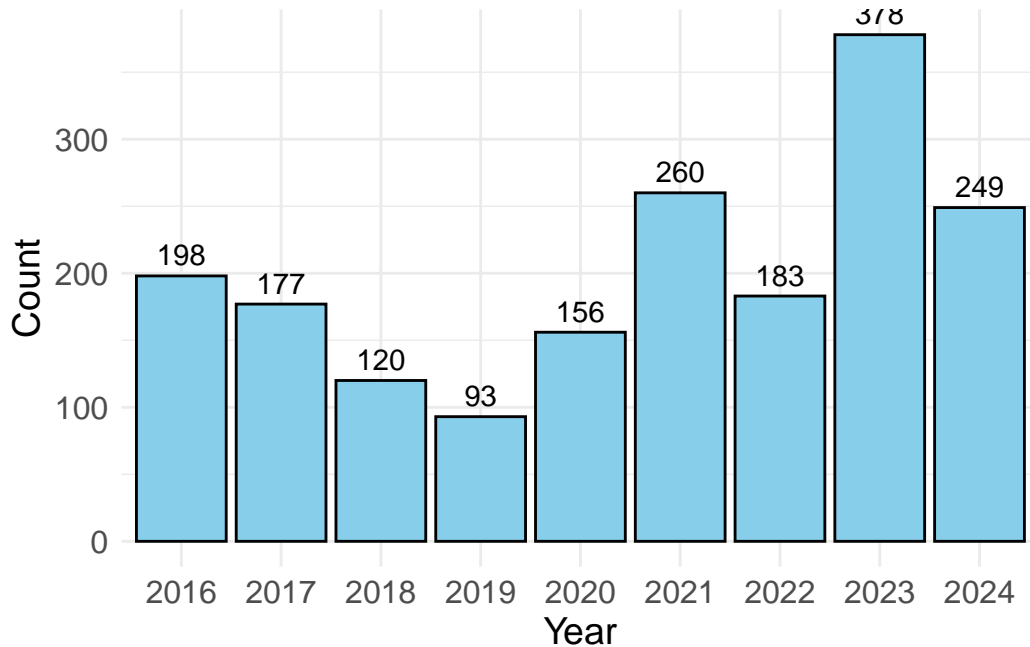


Figure 3: Distribution of Applications Years

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%
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.

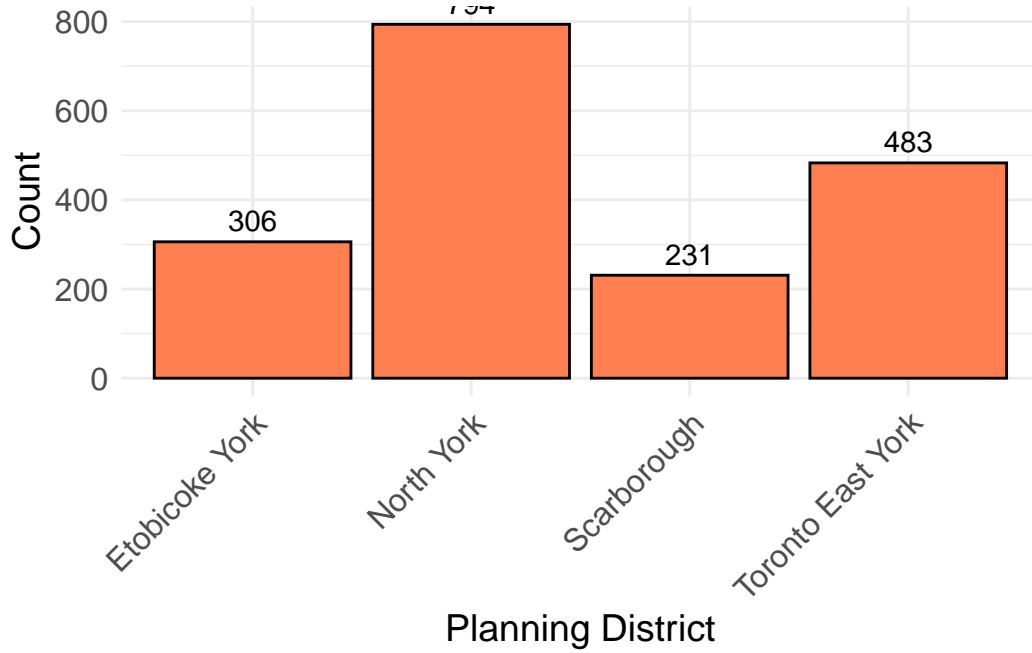


Figure 4: Distribution of Planning District

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. 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 appli-

cation committee. This model estimates the probability of an application being approved ($P(\text{Approval})$) 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 (`rstanarm?`). We use the default priors from `rstanarm`.

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:

$$y_i | p_i \sim \text{Bern}(p_i) \quad (1)$$

$$\beta_0 \sim \text{Normal}(0, 2.5) \quad (2)$$

$$\beta_1 \sim \text{Normal}(0, 2.5) \quad (3)$$

$$\beta_2 \sim \text{Normal}(0, 2.5) \quad (4)$$

$$\beta_3 \sim \text{Normal}(0, 2.5) \quad (5)$$

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

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

- 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.

3.1.1 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 multicollinear, 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.

3.1.2 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.

3.1.3 Potential Limitation and Addressing Violation

Violations of the above assumptions can impact this model's validity while several solutions 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.

4 Results

Our results are summarized in Table 6.

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

Table 6: Explanatory models of flight time based on wing width and wing length

	First 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

5.2 Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

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 5a we implement a posterior predictive check. This shows...

In Figure 5b we compare the posterior with the prior. This shows...

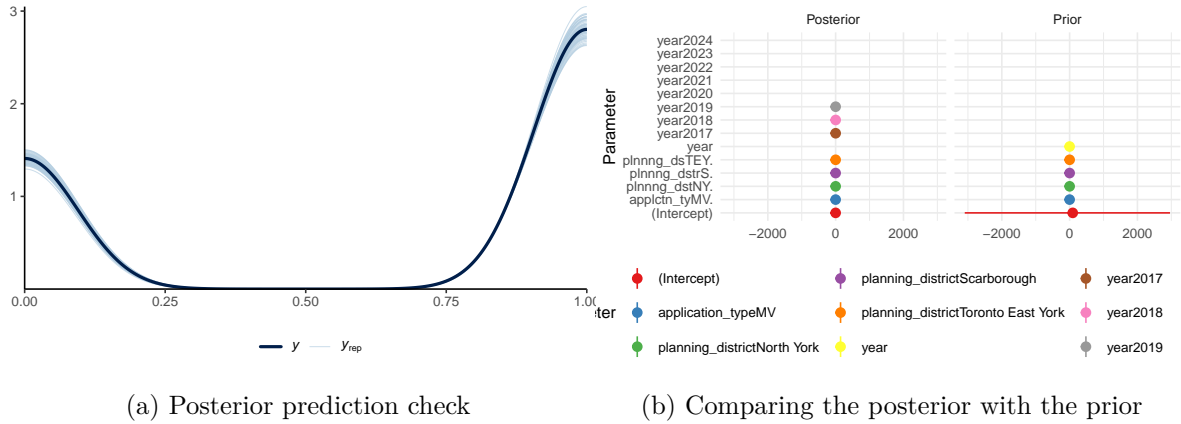


Figure 5: Examining how the model fits, and is affected by, the data

B.2 Diagnostics

Figure 6a is a trace plot. It shows... This suggests...

Figure 6b is a Rhat plot. It shows... This suggests...

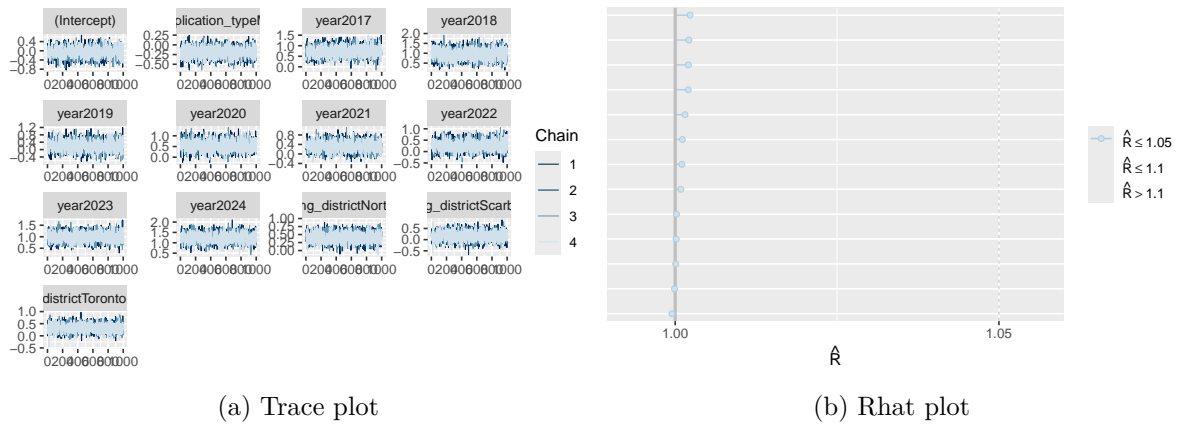


Figure 6: Checking the convergence of the MCMC algorithm

References

- Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. <https://tellingstorieswithdata.com/>.
- Arel-Bundock, Vincent. 2022. “modelssummary: 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>.
- 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, Dragoş 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 Golemund, 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>.