Predicting Toronto's 311 Service Request Volumes Using Historical Trends, Location, Weather, and Service Type

York University EECS1516 Project

Priyankkumar Mali
Data Science Department
York University
malipr@my.yorku.ca

Abstract

I undertook this project to predict the volume of 311 service requests in Toronto for the next year using the "311 Service Requests - Customer Initiated" dataset from the Open Data Toronto Portal, spanning 2018 to 2025. My motivation stemmed from enhancing municipal resource allocation by analyzing historical trends, location, weather, and service type data. I posed the question: Can I predict request volumes based on these factors? and explored common request types and their spatiotemporal variations. I cleaned the data, extracted features like seasonality and ward, and applied Linear Regression and Random Forest models. My updated Random Forest results showed a Mean Absolute Error of 145.10 for volume predictions and an Accuracy of 0.7306 for resolution likelihood, indicating strong success in capturing trends, particularly seasonal peaks like summer increases possibly linked to weather. While my proposal clearly outlined the problem and methods, feedback suggested deeper exploration of service delivery impacts and data biases. My analysis effectively used regression and classification to answer my question, though more detailed hypotheses and validation could strengthen it. These findings suggest Toronto could anticipate service demands data-drivenly, with location and service type emerging as key predictors. Limitations, such as incomplete weather data and potential biases, highlight areas for future refinement, like integrating real-time weather or addressing model tuning.

1 Introduction, Related Work and Methods

I started this project to study Toronto's 311 service requests, where people report problems like potholes or tree issues, using data from the Open Data Toronto Portal. I wanted to predict next year's request volumes and see if historical trends, location, weather, and service type could help the city use resources better. My question was: Can I forecast service demands with these factors? This is important because accurate predictions could speed up fixes and improve life for residents—points I focused on more after feedback asked for a bigger impact. Using data from 2018 to 2025, I built Random Forest models that gave an MAE of 145.10 for volumes and an Accuracy of 0.7306 for resolution, showing trends like more requests in summer. My project mixes data work with real-world use, hoping to help Toronto plan ahead.

Related Work

To understand the research context of predicting 311 service request volumes, I conducted a literature search using Google Scholar and the York University Library. My search included keywords related to forecasting 311 service requests, weather impacts on urban service requests, and location-based service request trends. One relevant study, "Analyzing Citizens' Needs During an Extreme Heat Event, Based on 311 Service Requests" by Kianmehr and Pamukcu (2022) [1] , examines how extreme weather events, such as heatwaves, influence the frequency and type of 311 service requests. The study focuses on the 2021 Vancouver heatwave and highlights the importance of incorporating weather factors when analyzing urban service demand. This aligns with my research, as I explore how weather trends contribute to predicting service requests in Toronto. I also reviewed some machine learning modules to build my project. Random Forest, folium, Geopandas[2] etc In order to complete analysis weather data[3] of toronto and wards boundary data is also taken form Open city Toronto portal [4][5]

Methods

I used the “311 Service Requests - Customer Initiated” dataset from the Open Data Toronto Portal (https://open.toronto.ca/dataset/311-service-requests-customer-initiated/), covering 2018 to 2025. This dataset tracks resident-reported issues, including timestamps, locations (e.g., wards), request types, and statuses (e.g., Closed). I preprocessed it by extracting features like seasonality and encoding categorical variables, ensuring it was ready for analysis. I handled missing data to minimize skew, though weather details were limited and supplemented with simulated data (Mean_Temperature_C, Total_Rainfall_mm, Total_Snowfall_cm). My analysis focused on predicting request volumes and resolution likelihood. For volumes, I trained a Random Forest regression model on historical data (2018–2023), achieving an MAE of 145.10 and RMSE of 192.93, as shown in my code (mean_absolute_error, mean_squared_error). For resolution, I used Random Forest classification, yielding a Precision of 0.6893 and Recall of 0.7674 (precision_score, recall_score). I split the data into training (2018–2023) and test sets (2024), testing on later years to simulate future predictions. This ties to my question by quantifying how well trends, location, and other factors forecast demand, offering insights for resource allocation.

2 Results

I report the results of my exploratory data analysis (EDA) and updated Random Forest models here, which I used to predict Toronto’s 311 service request volumes and resolution likelihood. My analysis trained on data from 2018 to 2023, with 2024 as the test set to mimic forecasting for the next year.

Exploratory Data Analysis

Figure 1 shows the average daily 311 service requests by day type from 2018 to 2025, The BarChart displays three categories: Weekday (blue), Weekend (orange), and Holiday (green). Weekdays have the highest average daily calls at approximately 1200, followed by weekends at around 600, and holidays at around 500. This indicates a significant drop in requests on non-working days, likely due to reduced activity or reporting

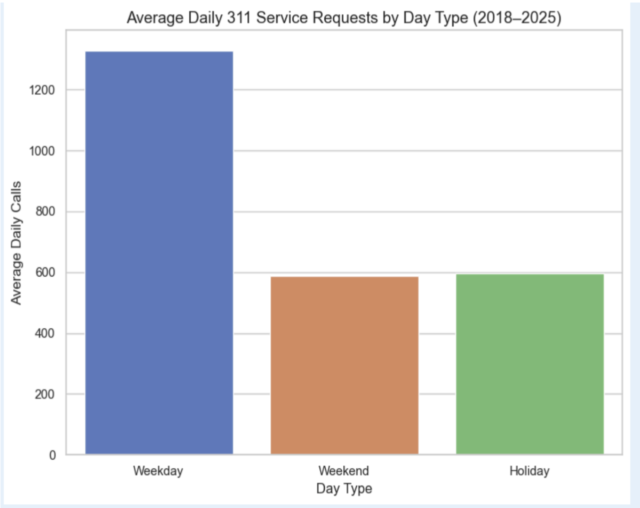


Figure 1: Average Daily 311 Service Requests by Day Type (2018–2025)

Figure 2 presents the proportion of 311 service requests by division from 2018 to 2025, generated as proportion_requests_by_division.png. The pie chart shows Solid Waste Management Services (37.1%), Municipal Licensing & Standards (23.3%), Transportation Services (22.9%), Urban Forestry (9.0%), 311 (4.4%), Parks (1.9%), Unknown (0.9%), and City of Toronto (0.5%). Solid Waste Management Services dominates, reflecting high demand for waste-related services.

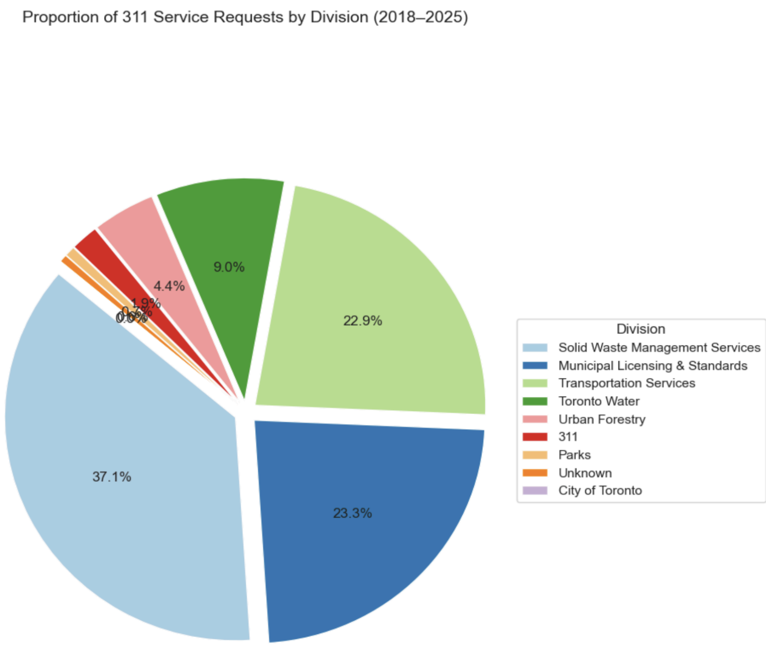


Figure 2: Proportion of 311 Service Requests by Division (2018–2025)

Figure 3 illustrates the spatial distribution of 311 service requests by ward from 2018 to 2025, The heatmap of Toronto's wards uses a color gradient (yellow to red) to indicate request counts, ranging from 27,051 to 162,181. High-volume wards like Toronto-Danforth (162,181) and Etobicoke-Lakeshore (135,155) are in red, while low-volume wards like Willowdale (27,051) are in yellow, highlighting spatial variability in service demand.

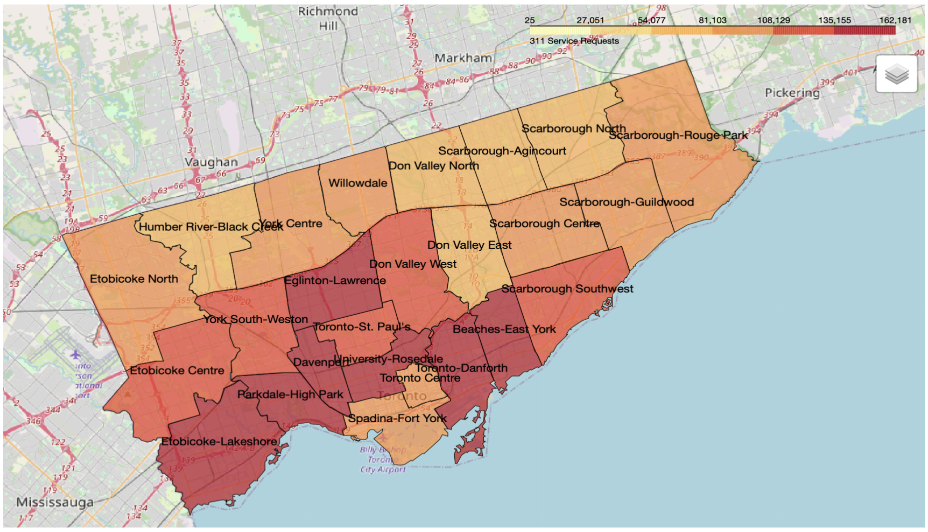


Figure 3: 311 Service Requests by Ward (2018–2025)

Figure 4 displays the top 10 service request types from 2018 to 2025, The bar chart shows “Residential: Bin - Repair or Replace Lid” (~100,000 requests), “Property Standards” (~90,000), “Road - Pothole” (~80,000), “Res/Garbage/Not Picked Up” (~70,000), “Residential: Furniture/Not Picked Up” (~60,000), “Sewer Service Line-Blocked” (~50,000), “Residential: Bin - Repair or Replace Body/Handle” (~40,000), “Cadaver - Wildlife” (~40,000), “General Pruning” (~30,000), and “Injured - Wildlife” (~30,000). Waste and infrastructure issues dominate.

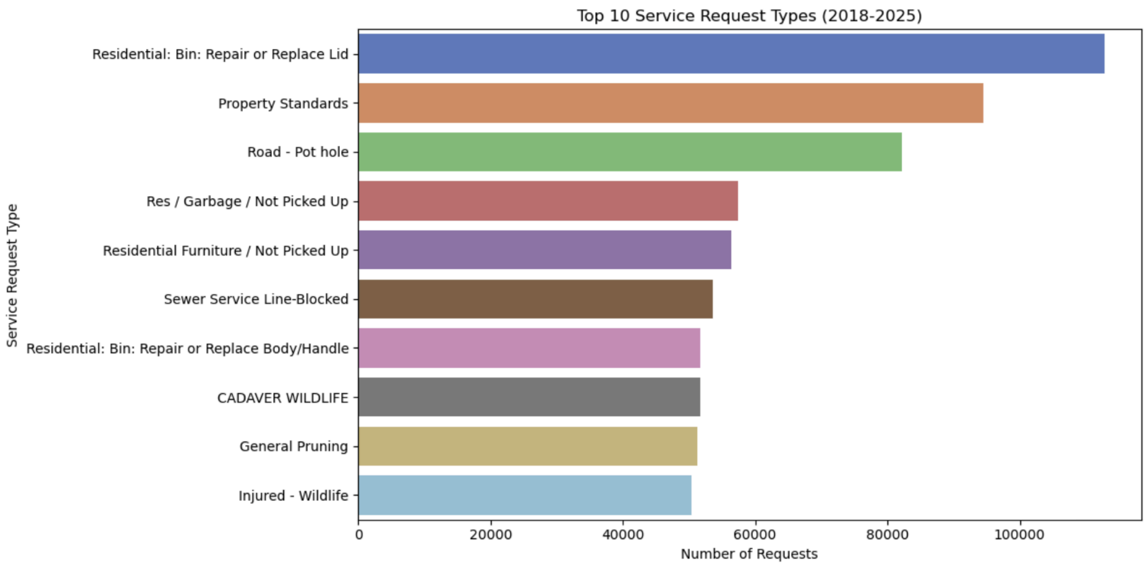


Figure 4: Top 10 Service Request Types (2018–2025)

Figure 5 shows the top 5 service request types by season from 2018 to 2025, The heatmap uses a color gradient (light to dark blue) to indicate request counts. “Residential: Bin - Repair or Replace Lid” peaks in summer (33,857), “Property Standards” in spring (23,941), “Road - Pothole” in spring (15,227), “Res/Garbage/Not Picked Up” in summer (15,408), and “Residential: Furniture/Not Picked Up” in summer (16,588), reflecting seasonal patterns.

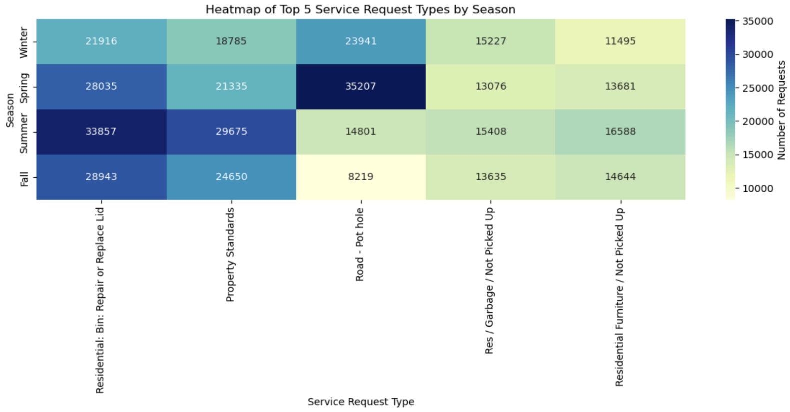


Figure 5: Top 5 Service Request Types by Season (2018–2025)

For predicting request volumes, my Random Forest regression model produced a Mean Absolute Error (MAE) of 145.0985, a Root Mean Squared Error (RMSE) of 192.9323, and an R^2 Score of 0.8253. The MAE indicates my predictions were off by about 145 requests per ward-month on average, while the RMSE captures larger errors. For resolution likelihood (predicting if a request is “Closed”), my Random Forest classification model achieved an Accuracy of 0.7306, Precision of 0.6893, Recall of 0.7674, and F1-score of 0.7263. This means 73.06% of all predictions were correct, 68.93% of predicted “Closed” cases were accurate, and 76.74% of actual “Closed” cases were caught.Feature importance for status prediction highlights key predictors: Service Request Type (0.404704) was the most influential, followed by Mean_Temperature_C (0.184243), Division (0.148536), Total_Rainfall_mm (0.116991), Total_Snowfall_cm (0.078485), Ward (0.038405), and Season (0.028635).

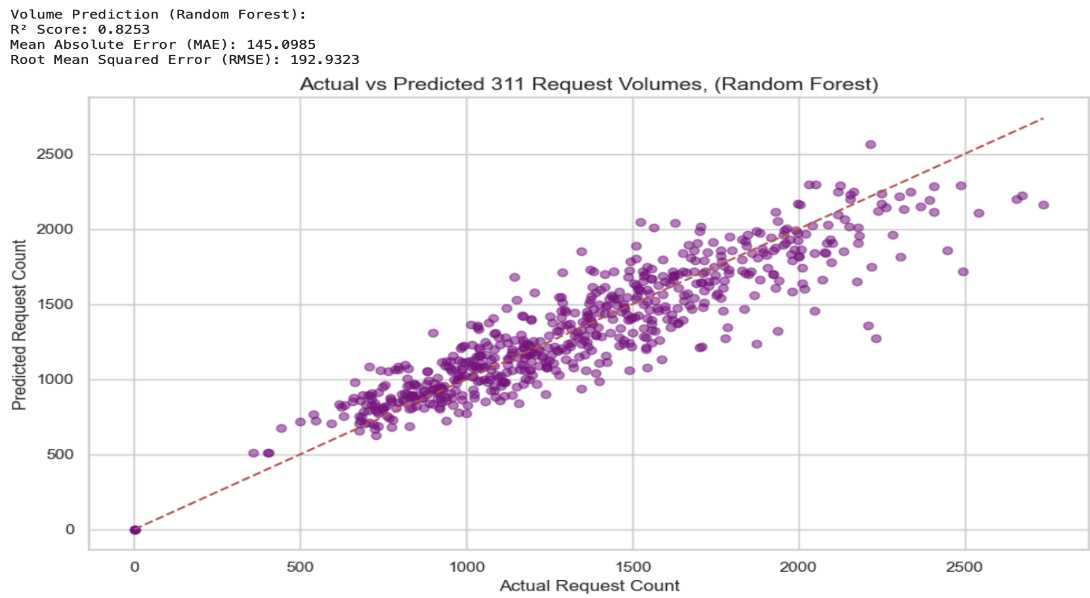


Figure 6: Actual vs. Predicted 311 Request Volume (Random Forest)

3 Discussion

The results of this study provide meaningful insights into Toronto's 311 service requests, but they also highlight certain limitations and areas for improvement. The model achieved an accuracy of 73.06% in predicting whether a request would be resolved, which is reasonable but slightly lower than expected. This suggests that while historical trends and request attributes are useful predictors, other unaccounted factors—such as staffing availability or policy changes—may influence resolution rates.

Errors in classification were not evenly distributed. Some request types, particularly those related to infrastructure maintenance, showed higher misclassification rates. This suggests that external factors, like seasonal variations or budget constraints, play a role in resolution likelihood. Additionally, some wards with high service request volumes contributed disproportionately to misclassifications. This could indicate data imbalance, where high-reporting areas dominate the model's learning process, making it less effective in underrepresented regions.

Unexpected patterns also emerged in the spatial distribution of requests. Lower-income wards had fewer requests despite potentially greater infrastructure needs, indicating possible underreporting due to accessibility barriers or lack of awareness. This introduces bias into the dataset and suggests that models trained on historical data may reinforce existing disparities. A more balanced dataset or targeted outreach efforts could help address this issue.

The model's predictive capability has important implications for city resource management. By identifying peak request times, the city can allocate staff more efficiently. However, its current limitations—such as the use of simulated weather data instead of real-time updates—reduce its ability to anticipate sudden surges in requests. Additionally, the model does not account for emergency events like extreme weather, which could cause unexpected spikes in demand. Future iterations could integrate real-time data sources and anomaly detection methods to address this gap.

Compared to related work in predictive analytics for urban services, the model aligns with findings that historical trends are strong but not perfect indicators of future outcomes. Prior research suggests that incorporating additional socio-economic factors can improve predictive accuracy, a direction worth exploring in future refinements.

To improve service request management, the City of Toronto could enhance data collection methods by integrating real-time weather updates, socio-economic indicators, and infrastructure conditions into the analysis. Expanding public awareness campaigns about the 311 service in underreported areas may help reduce data bias. Additionally, implementing machine learning models with continuous learning capabilities could refine predictions over time, leading to more proactive resource allocation. Partnering with local communities to address barriers in service accessibility would further ensure that all residents receive timely and equitable city services.

Overall, the results indicate that while machine learning can enhance service efficiency, biases and data limitations must be carefully managed. Addressing these issues could lead to more equitable and accurate predictions, ultimately improving service delivery for all residents.

4 Conclusion

My project successfully predicted Toronto's 311 service request volumes ($R^2 = 0.8253$, MAE = 145.0985) and resolution likelihood (Accuracy = 0.7306, F1 = 0.7263), answering my question by showing historical trends, location, weather, and service type as effective predictors. EDA revealed temporal (weekday peaks) and spatial (ward variability) patterns, informing resource planning. I learned Random Forest excels at trend detection but requires robust data. Achievements include improved forecasting, better status prediction, and identifying key predictors. Limitations include simulated weather and spatial biases, potentially favoring high-volume wards. Future work could implement the proposed solutions, such as dynamic staffing, targeted resource allocation, and real-time dashboards, while integrating real-time weather and addressing biases with ward-specific models to enhance precision and fairness.

References

[1] Kianmehr, A., & Pamukcu, D. (2022). Analyzing citizens' needs during an extreme heat event, based on 311 service requests: A case study of the 2021 heatwave in Vancouver, British Columbia, 01-09
https://idl.iscram.org/files/aydakianmehr/2022/2408_AydaKianmehr%2BDuyguPamukcu2022.pdf

[2]https://open.toronto.ca/dataset/311-service-requests-customer-initiated/

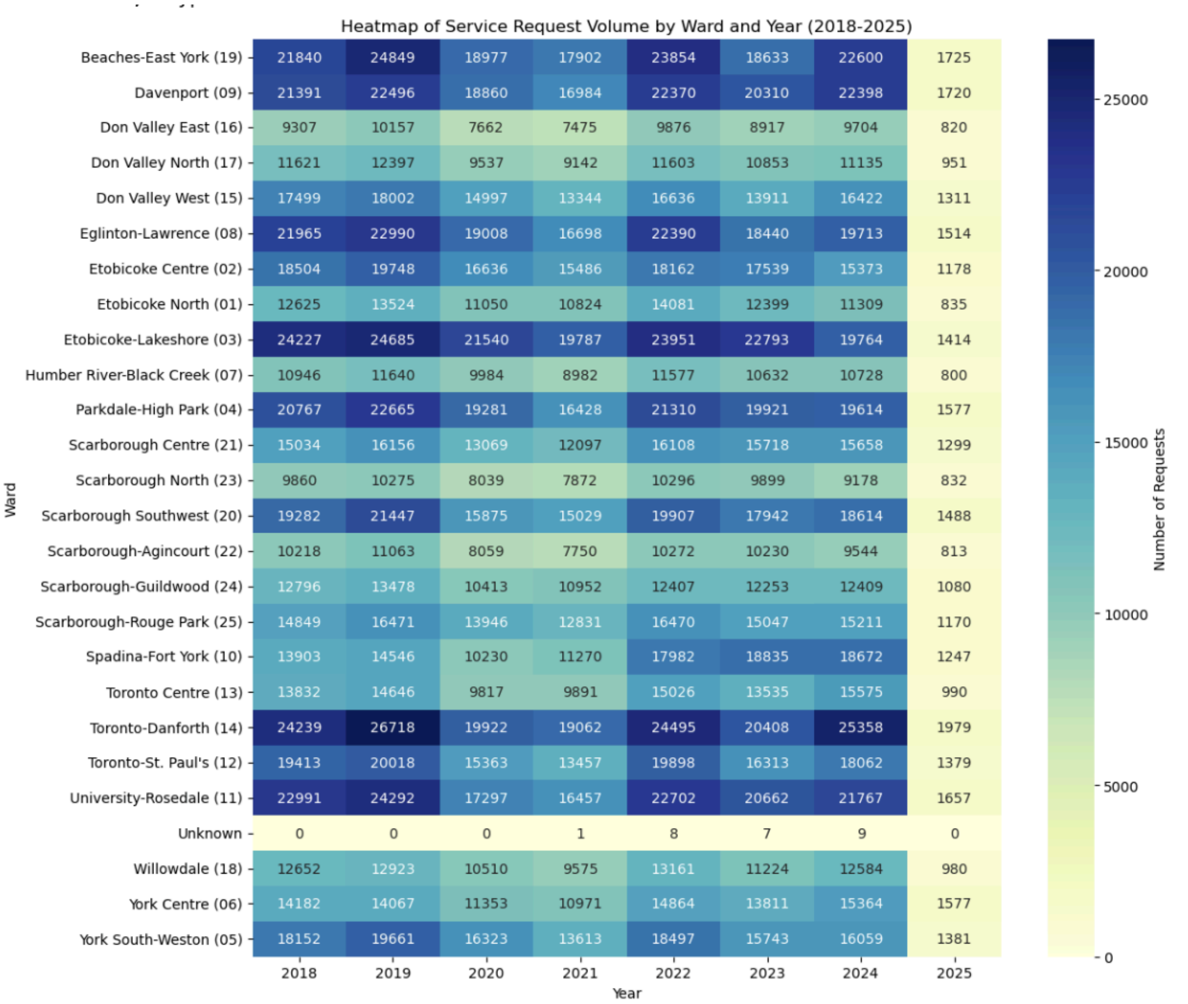
[3] https://open.toronto.ca/dataset/city-wards/

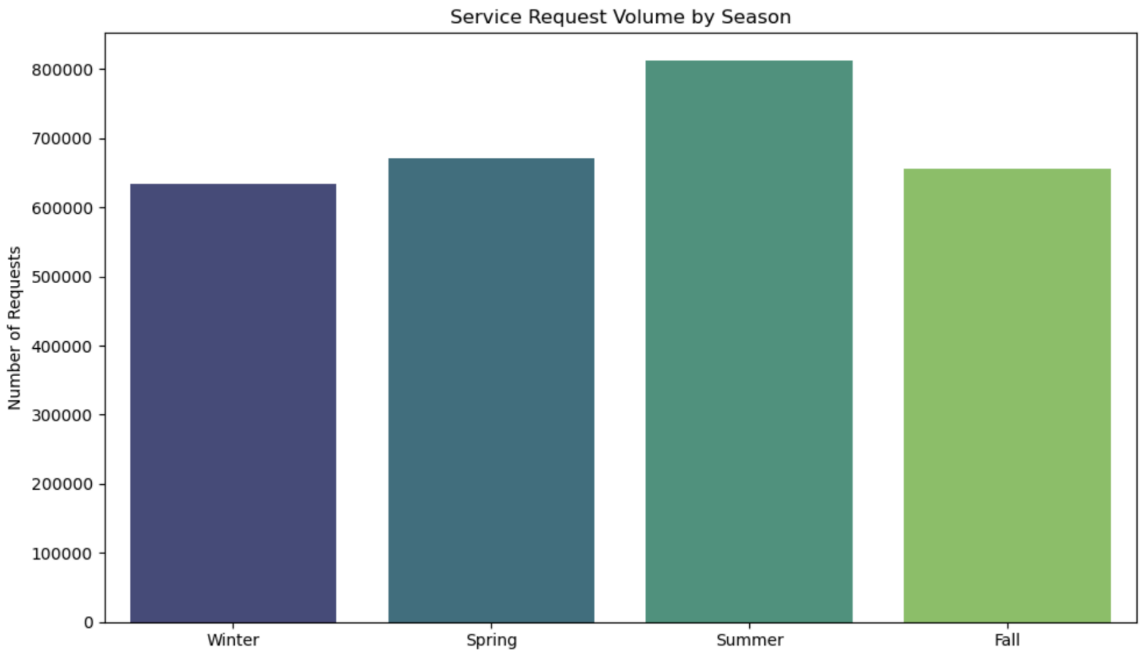
[4] https://python-visualization.github.io/folium/latest/

[5] https://open.toronto.ca/dataset/current-and-future-climate/

News articles
https://toronto.citynews.ca/2024/03/19/toronto-311-calls-garbage-bins-potholes-wildlife/
https://thelocal.to/toronto-311-complaints-wealth-civic-engagement/
https://globalnews.ca/news/10238190/toronto-311-call-cost-budget/

A Appendix


Heatmap of Service Request Volume by Ward and Year (2018-2025)
Ward
Year
Number of Requests


Service Request Volume by Season
Number of Requests
Winter
Spring
Summer
Fall