

NYC 311 Flood Complaint Analysis

May 7, 2025

Abstract

Abstract: This study investigates sewer-related complaints in New York City using 2024 data from the NYC 311 Service Requests system, focusing on street flooding and catch basin complaints. Combining exploratory analysis, hypothesis testing, logistic regression, and negative binomial models, the analysis identifies inequalities in complaint resolution times driven by complaint type, borough, socioeconomic conditions, and timing factors. Catch basin complaints are recognized as strong predictors of reports of flooding, and delays happen more often in poorer communities and on weekends. Findings favor targeted operation reform and better complaint data practices to advance urban flood response equity and efficiency.

Keywords: NYC 311, Street Flooding, Catch Basin, Response Time, Logistic Regression, Negative Binomial Regression

1 Introduction

New York City’s sewer system is a vital element of maintaining urban resilience, particularly in shielding the effects of flooding brought about by extreme weather patterns. With the effect of climate change progressively raising precipitation intensity and frequency, urban drainage systems face growing operating concerns. Although previous studies have documented trends in service complaints and capacity of infrastructure, such studies are often lack complaint type analysis regarding local conditions and response metrics.

This research fills these gaps by examining complaint response time, seasonal trends, and socio-demographic impacts using data from NYC’s 311 system for 2024. Focusing particularly on Street Flooding (SF) and Catch Basin Clogged/Flooding (CB) complaints, the research incorporates spatial, temporal, and socioeconomic variables in the analysis to ascertain not only where and when complaints are occurring, but how quickly they are being addressed.

The remainder of the paper is organized as follows. Section 2 describes the data sources and preprocessing steps. Section 3 outlines the statistical methods and modeling

techniques. Simulation results and application outcomes are presented in Section 4. Section 5 discusses implications and concludes the paper.

2 Data

2.1 Overview

This project uses the NYC 311 Service Requests dataset filtered to sewer-related complaints in 2024. The dataset includes 9,483 records, specifically focusing on *Street Flooding (SF)* and *Catch Basin Clogged/Flooding (CB)* issues. Each record contains information on time, location, type of complaint, and resolution status.

Such complaints are most commonly reported and are directly associated with rainfall events and infrastructure operation. SF and CB complaints are a good indicator of immediate flood-related hazards affecting residents' mobility, safety, and property. Their abundance, operational utility, and close connection to weather conditions make them susceptible to both descriptive and predictive modeling for urban flood risk.

2.2 Variables and Structure

The dataset contains 41 columns. Key variables used in this analysis include:

- `descriptor` — Identifies the type of complaint (SF or CB).
- `created_date`, `closed_date` — Used to compute resolution time.
- `borough`, `incident_zip` — Geographical identifiers.
- `latitude`, `longitude` — Used for spatial mapping.

2.3 Cleaning and Preprocessing

- Columns with 100% missing values (e.g., `vehicle_type`, `due_date`) were removed.
- Redundant fields such as `location`, `agency_name`, and state plane coordinates were dropped.
- ZIP codes and boroughs were cleaned using lookup dictionaries; unspecified or missing entries were filled where possible.
- Timestamps were standardized; anomalies (e.g., `closed_date` before `created_date`) were flagged.

The cleaned dataset was saved in Apache Feather format, reducing the size by 0.682 and improving loading speed.

2.4 Missing Data Summary

Variable	Missing (%)	Comment
closed_date	2.19%	Open complaints
incident_address	23.6%	Incomplete location data
latitude, longitude	1.0%	Excluded from spatial maps
intersection.street.1/2	76.4%	Mostly missing

Table 1: Summary of key missing data fields.

2.5 Complaint Distribution and Visualization

Figure 1 visualizes the geographic distribution of complaints across New York City, colored and shaped by complaint type.

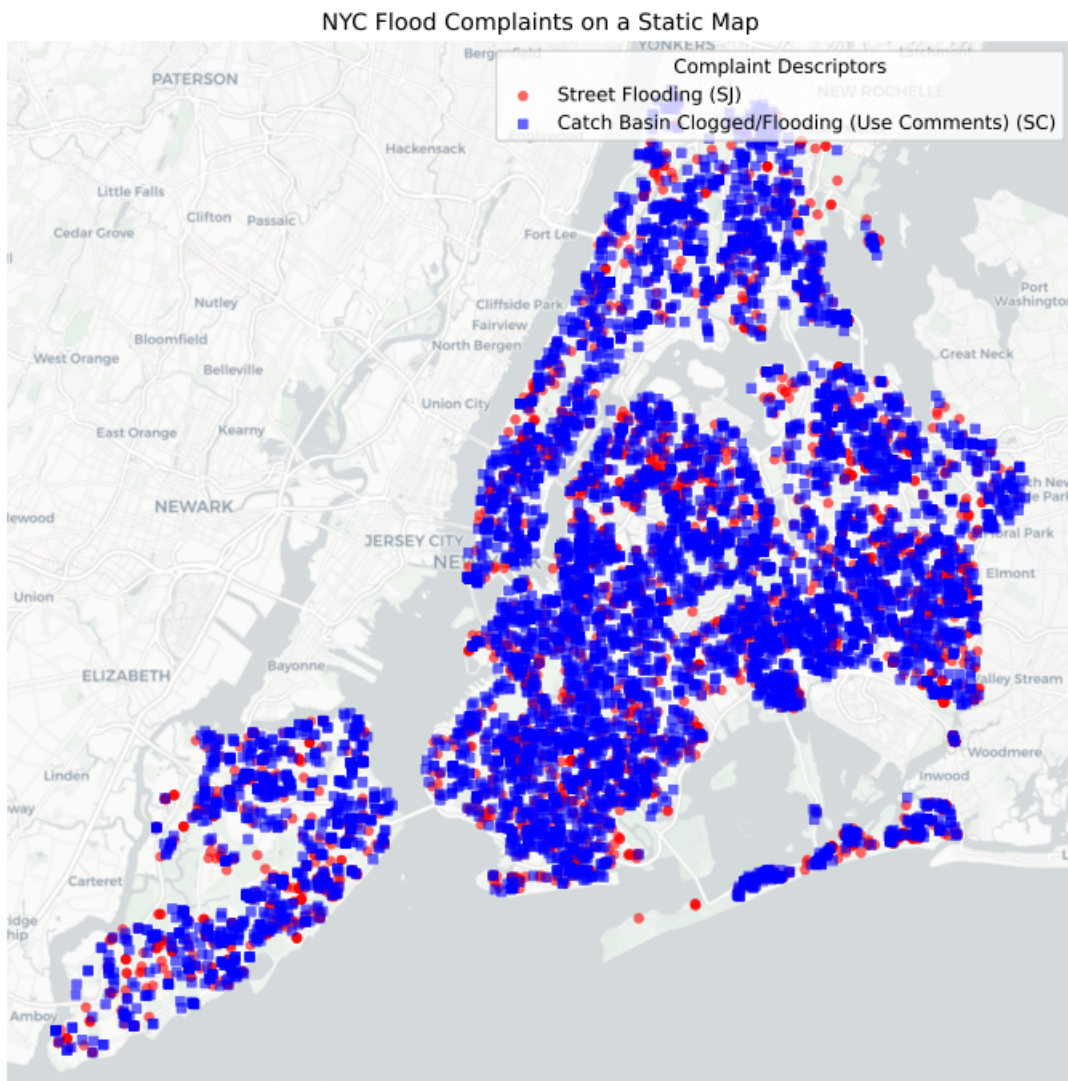


Figure 1: Map of Sewer Complaints by Type Across NYC (2024).

Figure 2 compares the total count of complaints for Catch Basin vs. Street Flooding.

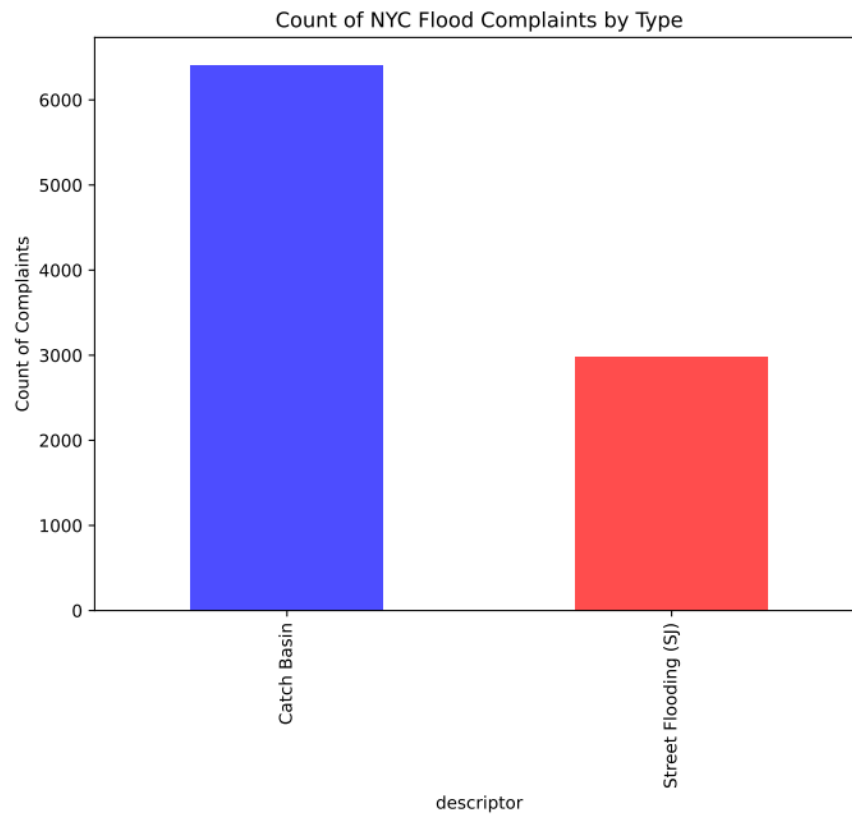


Figure 2: Count of Catch Basin vs. Street Flooding Complaints.

Figure 3 provides a breakdown of the percentage share of each complaint type.

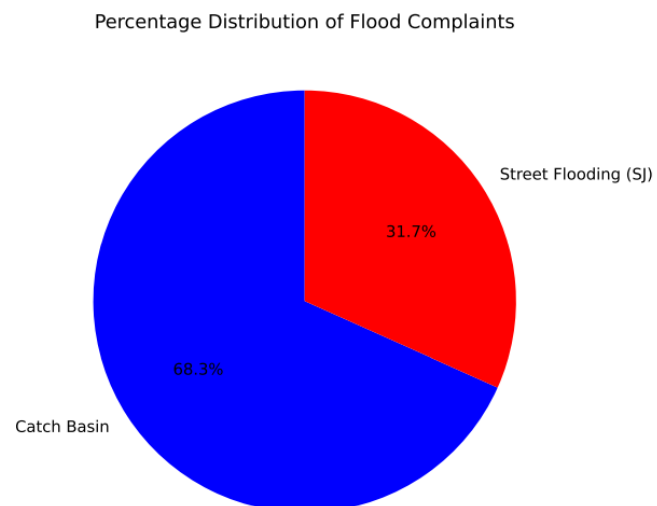


Figure 3: Percentage Distribution of Complaint Types.

2.6 Observations

Based on Figure 2 and Figure 3, Complaints of Catch Basins surpass Street Flooding complaints by more than two to one. This makes sense and is with the nature of these complaints: infrastructure congestion is more persistent, while street flooding is typically complained about during specific weather conditions.

The spatial map in Figure 1 highlights higher complaint densities in Queens and Brooklyn. Manhattan shows relatively fewer reports, likely due to better infrastructure and drainage systems.

The spatial map in Figure 1 shows higher complaint densities in Queens and Brooklyn. Manhattan, comparatively, shows relatively fewer reports. This is an observation from visual density of complaint markers by borough, where Manhattan obviously has fewer points per unit area than outer boroughs. Despite differences in area and population, context levels of infrastructure investment put it into perspective, Manhattan has an advantage due to generally having better developed and centrally served sewer and drainage systems since it is economically central. This aligns with earlier studies (e.g., Agonafr et al. 2021), where lower complaint frequencies were associated with areas with better infrastructure and maintenance coverage. Therefore, the somewhat lower complaint rate in Manhattan can reflect true conditions and system performance as well as not reflect underreporting.

2.7 Engineered Variables

To support further modeling, new features were created:

- `response_time` (hours between created and closed dates)
- `over3d` (binary flag for complaints taking 72+ hours)
- `weekend_holiday`, `rush_hour`, `season`
- Merged ACS 2023 data: median income, education, population density

These features were used in later stages to explain and predict resolution delays. Full modeling results are in Section 3 and 4.

3 Methods

3.1 Summary of Statistical Techniques

Table 2 summarizes the statistical methods applied across different tasks.

Table 2: Summary of Statistical Methods Used

Method	Purpose
Exploratory Data Analysis (EDA)	Visualized spatial and temporal trends in complaints using maps, bar charts, and boxplots.
Welch’s T-test	Compared mean response times between Street Flooding (SF) and Catch Basin (CB) complaints.
One-Way ANOVA	Tested whether average response times differed across boroughs.
Two-Way ANOVA	Tested whether the effect of complaint type on response time varies by borough.
Logistic Regression	Modeled the likelihood of complaints exceeding a 72-hour resolution time.
Negative Binomial Regression	Modeled ZIP-level daily CB complaint counts, accounting for overdispersion.

3.2 Response Time Modeling (Logistic Regression)

A logistic regression model was used to predict whether a complaint would remain unresolved for more than 72 hours. The dependent variable `over3d` was regressed on the following predictors:

- **Temporal:** season (dummy-coded), hour of day, rush hour, weekend/holiday.
- **Complaint type:** descriptor (SF or CB).
- **Location:** borough (dummy-coded).
- **Neighborhood characteristics:** population density, education index, unemployment rate, housing affordability index.

The model was trained using an 80/20 train-test split. The 80/20 split is a common standard in predictive modeling, balancing sufficient training data for model learning and sufficient test data for generalization performance estimation. Accuracy, precision, recall, F1 score, and AUC were used as the evaluation metrics. Also, feature importance was explained using SHAP values and odds ratios.

3.3 Complaint Count Modeling (Negative Binomial Regression)

Catch Basin complaint counts were modeled using negative binomial regression due to overdispersion in the data. The dependent variable was the number of CB complaints per ZIP code per day. Predictors included:

- **Weather:** current and lagged 1-day precipitation.
- **Temporal dependence:** lagged CB complaint count.

- **Landscape:** slope, elevation, impervious surface percentage.
- **Demographics:** education index, unemployment rate.

Poisson regression was also fit for comparison. Deviance, AIC, and residual plots were used to assess overdispersion and model fit.

3.4 Hypothesis Testing

All of the testing has been conducted using a significance level of 0.05. This is standard statistical inference and has a 5% risk of Type I error, with a balance between sensitivity and specificity. It is widely used and accepted in all statistical analysis to test evidence against null hypotheses.

Three hypotheses were tested using the following methods:

1. H1: Mean response times differ significantly between CB and SF complaints — Welch’s T-test.
2. H2: Response times vary by borough — One-way ANOVA.
3. H3: The effect of complaint type on response time varies by borough — Two-way ANOVA.

All tests used a significance threshold of 0.05. Results led to rejection of the null in each case.

3.5 Model Evaluation

The logistic model achieved an AUC of 0.73 on both training and testing data, but the F1 score was low due to class imbalance. Odds ratios and SHAP values showed complaint type, borough, and education index were key predictors.

The negative binomial model outperformed Poisson regression in deviance and residual dispersion, confirming its suitability for modeling overdispersed count data.

4 Results

4.1 Differences in Response Times

Several statistical tests were conducted to examine response time differences. Welch’s T-test showed that catch basin (CB) complaints had significantly longer response times than street flooding (SF) complaints ($p < 0.001$). One-way ANOVA confirmed that average response times differed across boroughs ($p < 0.001$). A two-way ANOVA showed

a significant interaction effect between borough and complaint type, indicating that some boroughs prioritize complaint types differently ($p < 0.001$). These findings are visualized in the boxplot in Figure 4, which displays response times by borough and complaint type.

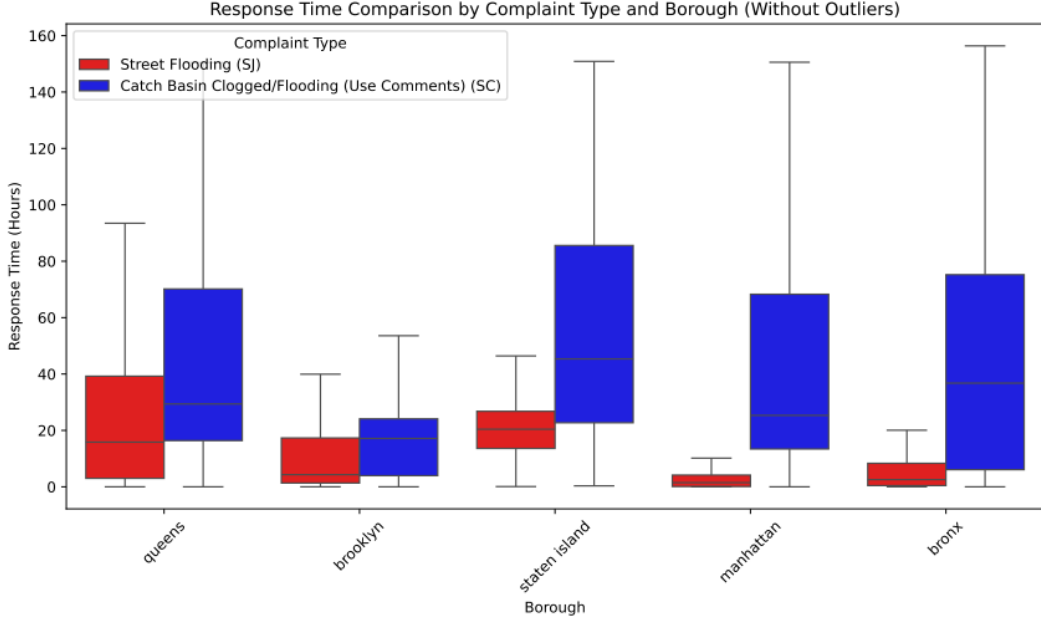


Figure 4: Boxplot of response times by borough and complaint type (outliers removed).

4.2 Logistic Model for Delays Over Three Days

The logistic regression model estimated the likelihood that a complaint remained unresolved after 72 hours ($\text{over3d} = 1$). The model achieved an AUC of 0.73 on both training and test sets. Although accuracy was high (79 percent), the F1 score for detecting delayed complaints was low (0.005), due to severe class imbalance. Precision was 33 percent for predicting delays, and recall was nearly 0.

Table 3 summarizes the confusion matrix and performance metrics.

Table 3: Logistic Model Evaluation (Test Set)

	Precision	Recall	F1 Score
Non-delayed (0)	0.79	1.00	0.88
Delayed (1)	0.33	0.00	0.01
Overall Accuracy		0.79	
AUC		0.73	

Table 4 shows the logistic regression coefficients and their corresponding odds ratios for the likelihood of response times exceeding 72 hours.

Table 4: Logistic Regression Results: Predictors of Response Time Over 72 Hours

Feature	Coefficient	Odds Ratio
season_summer	0.127473	1.135954
unemployment_rate	0.112623	1.119210
season_spring	0.069896	1.072397
borough_manhattan	0.056113	1.057717
rush_hour	0.016413	1.016549
borough_queens	0.001696	1.001698
borough_staten island	0.000243	1.000243
season_winter	-0.012604	0.987475
housing_affordability_index	-0.022640	0.977615
weekend_holiday	-0.075759	0.927039
education_index	-0.109260	0.896497
population_density	-0.109658	0.896140
descriptor_Street Flooding (SJ)	-0.596772	0.550586
borough_brooklyn	-0.717803	0.487823

Key Predictors. Odds ratios showed that:

- Complaints from Brooklyn were 51% less likely to be delayed.
- Street flooding complaints were 45% less likely to be delayed than catch basin complaints.
- Higher education levels and denser populations were associated with faster resolutions.
- Complaints during summer were 13.6% more likely to be delayed.

SHAP analysis confirmed these features as the most impactful. Figure 5 shows the SHAP summary plot.

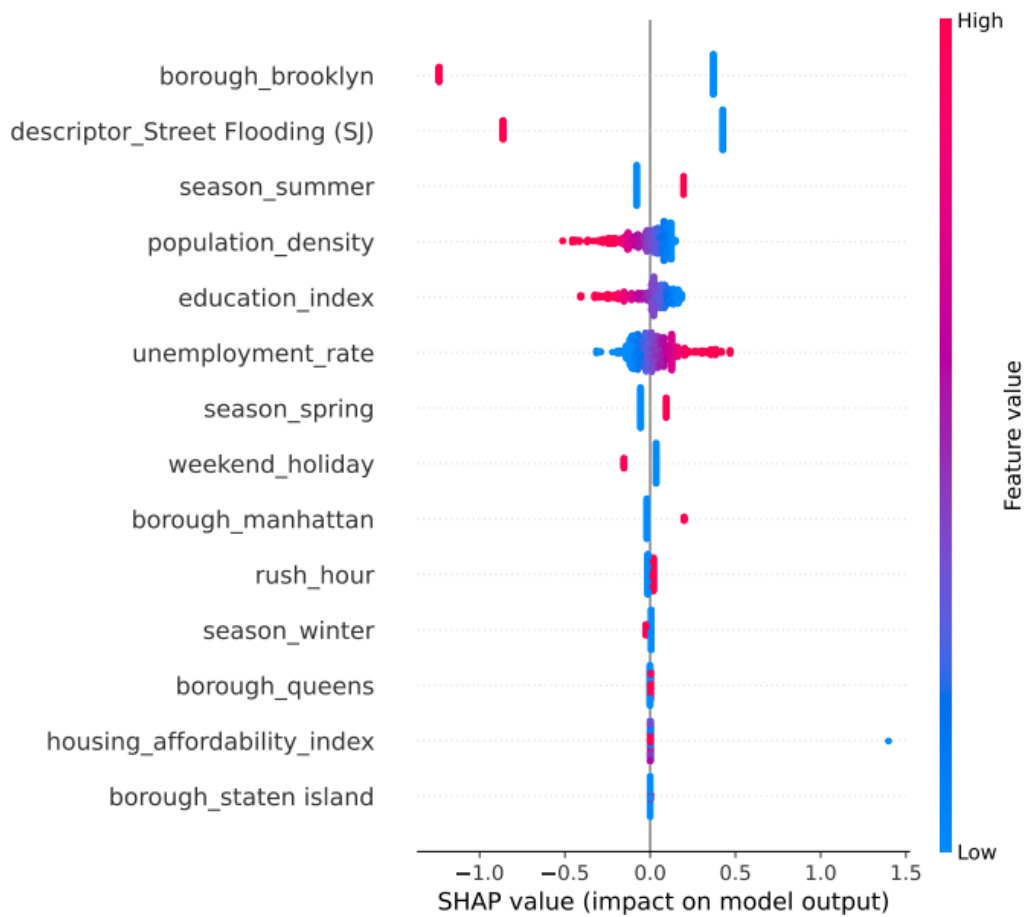


Figure 5: SHAP summary plot showing feature importance for predicting delays (> 72 hours).

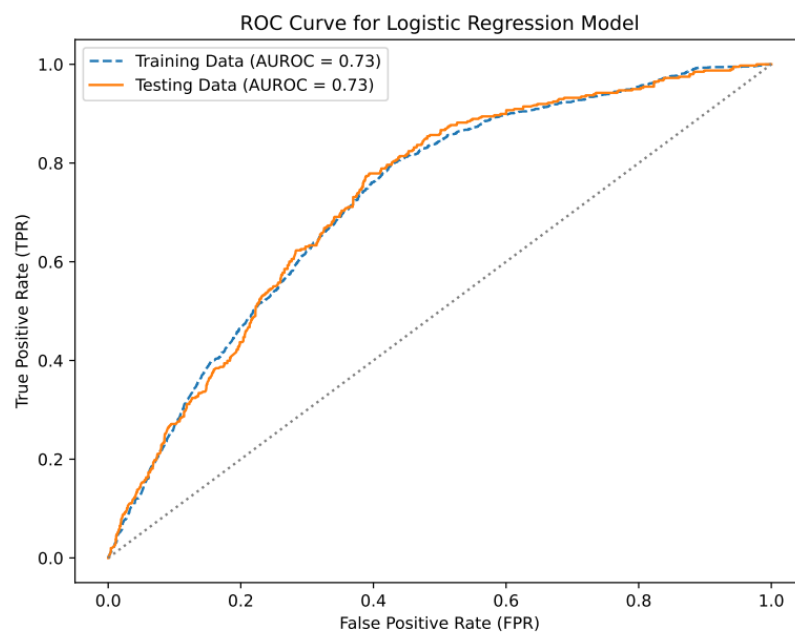


Figure 6: ROC Curve for Logistic Model (Test Set)

AUC (Area Under the Curve) is a metric for how well the model is able to distinguish between complaints that took more than three days to resolve and those that did not. Perfect predictions are assigned an AUC of 1.0, 0.5 is the same as a coin toss, and an AUC below 0.5 indicates the model performs worse than random chance. A high AUC indicates that the model is able to effectively identify delays, and an AUC near 0.5 indicates that the model is having a hard time distinguishing between slow and fast responses. Very low AUC indicates that the model is not working at all and even performs inverse predictions.

At 0.73 AUC for both training and test data, the model is preferable to random guessing but worse than perfect. The ROC curve in the figure above suggests the model does provide some predictive value, though, even if it still makes a lot of mistakes. As a predictive model for NYC commuters, this model has limited practical application, it does sometimes successfully predict delays, but isn't perfect enough to be of use for making regular decisions or public alert systems.

4.3 Complaint Count Modeling

A negative binomial regression model was used to estimate the number of daily CB complaints per ZIP code. The model significantly outperformed the Poisson model in terms of deviance (Table 5), indicating it better accounted for overdispersion in the data.

Table 5: Comparison of Regression Models

Model	AIC	Deviance
Poisson	270,894	0.3746
Negative Binomial	255,125	0.1167

Important Features. The most significant predictors of complaint volume included:

- **Lagged CB complaints:** Strong positive predictor; prior issues predict future complaints.
- **Precipitation (lagged and current):** Rainfall was associated with higher complaint volume.
- **Elevation and impervious surface:** Higher elevation and less permeable areas showed higher complaint counts.
- **Education index:** Better-educated areas had slightly fewer complaints, possibly reflecting reporting differences.

4.4 Spatial and Temporal Trends

Complaints varied seasonally and geographically. CB complaints were more common citywide and especially frequent in Queens and the Bronx. Manhattan had the shortest response times across all complaint types. Temporal trends showed spikes following rainfall events, consistent with infrastructure stress. Figure 7 show spatial clustering of complaints across NYC. Catch Basin complaints were highly concentrated in the Bronx and Queens, while Street Flooding complaints were more frequent in Brooklyn and south-eastern Queens.

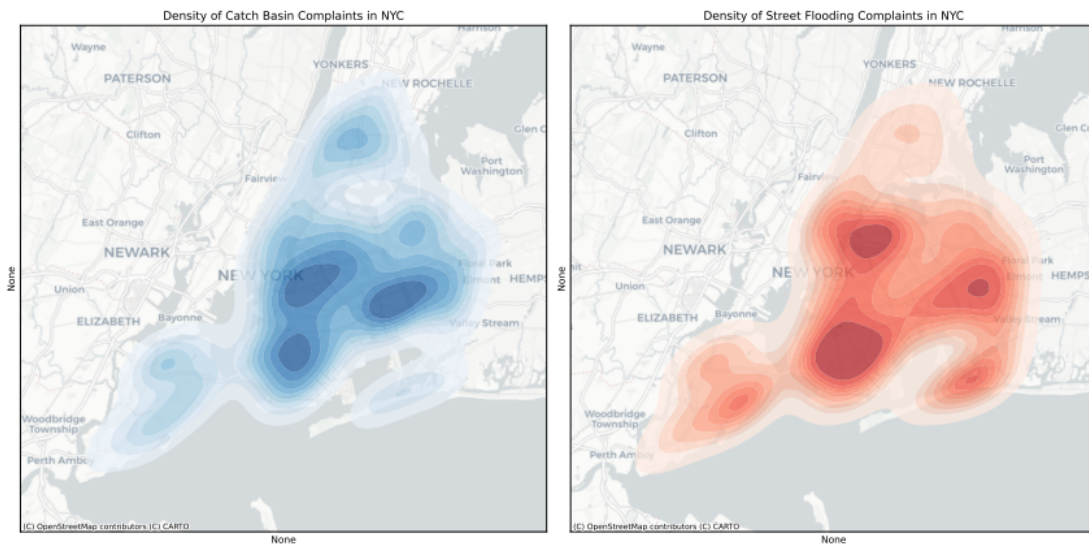


Figure 7: Left: Density map of Catch Basin complaints. Right: Density map of Street Flooding complaints.

4.5 Model Limitations

While the logistic model identified key drivers of delay, it struggled to predict actual delays due to class imbalance. The negative binomial model effectively captured count variability, but relied on ZIP-level aggregation, potentially masking street-level variation.

Overall Findings. Street flooding complaints receive quicker resolutions, particularly in Manhattan and Brooklyn. Catch basin complaints are more frequent and take longer to resolve, especially in lower-income and higher-elevation areas. Complaint volumes are sensitive to prior complaints and rainfall, supporting the need for proactive maintenance in susceptible neighborhoods.

5 Discussion

This study looked at NYC street flooding complaint patterns with a combination of hypothesis testing, regression modeling, and spatial analysis. Our findings align with and extend recent research that uses 311 data to evaluate urban flood risk, particularly the work of Agonafr et al. (2022)(2), which utilized LASSO and zero-inflated models to predict complaint patterns across ZIP codes in NYC.

5.1 Catch Basin Complaints as a Dominant Predictor

One of the vital findings of the 2022 study is the dominance of catch basin (CB) complaints as a predictor of street flooding complaints (SF). CB complaints accounted for over 41% of relative importance in the RF model, significantly larger than other factors. This is concurrent with the 2021 study(1), which found infrastructural conditions as the main drivers of SF variability. The prevalence of CB implies that blocked basins remain a leading maintenance issue and potential intervention point for city governments. Our own analysis, using an alternative modeling pipeline, produced very similar results: CB complaints were strong predictors of complaint volumes by day across ZIP codes, even when controlling for precipitation, elevation, and socioeconomic variables.

5.2 Topography and Socio-Demographic Factors

Topographic conditions of slope and elevation were also among the strongest predictors after catch basin (CB) complaints. These findings are consistent with the conclusion of Agonafr et al. (2021) that physical geography is central to flood risk distribution in New York City. Locations that are at lower elevations or slope bottoms tend to accumulate more water during a storm, increasing the chances of local flooding. Conversely, locations with higher elevations may have quicker flow but prove difficult in drainage when slope overruns system design. These findings confirm the role of physical geography in contributing to flood exposure and offer some insight into varying complaint patterns around the city.

One interesting insight provided by the 2022 paper is identifying socio-demographic variables as important predictors. The variable "commuters who drive alone" (COM02) was the second highest importance after CB in Model 1, which reflects SF complaint impact by commuting patterns and potentially 311 reporting participation bias. This is consistent with earlier concerns by Basiri et al (3). (regarding demographic bias in crowdsourced data. The logistic regression model also suggested borough-level variations, time-of-week trends (weekends and holidays), and socioeconomic conditions (education index, housing affordability) as predictors of longer delays. This extends the focus beyond frequency of complaint to response equity and system performance, factors less

emphasized in Agonafr et al.’s flood model.

5.3 Precipitation, Spatial Coordinates, and Complaint Timing

Spatial coordinates (latitude and longitude) and ZIP code area were significant predictors. Contrary to expectations, precipitation indicators did not appear in the leading predictors in the 2022 RF model, suggesting that city rainfall variability is less influential than land use and infrastructural characteristics. However, in our models, precipitation was a significant force behind complaint numbers in both logistic and negative binomial regression contexts.

5.4 Response Time Disparities and Delay Predictors

Three sets of results emerged from our extended analysis:

1. **Response time disparities:** SF complaints were addressed earlier than CB complaints on average. The difference was statistically significant by borough and interacted with geography, Staten Island and the Bronx, in particular, had longer delays for CB issues.
2. **Predictors of prolonged delays:** Logistic regression identified responses that took longer than 72-hour response times as more likely to occur when complaints had been reported over weekends, are in less affordable areas of housing, are better educated and when the area is highly dense. Borough was also a statistically significant predictor after other variables had been taken into account.
3. **Complaint frequency patterns:** The negative binomial model confirmed the overdispersion of complaint data. Precipitation and lagged complaint counts were highly predictive of daily volumes of CB complaints, as in previous results in Agonafr et al.(2). The model also outperformed a Poisson model by deviance and AIC.

5.5 Implications for Flood Risk Modeling and Policy

The findings imply that predictive models should incorporate not only hydrological and infrastructural parameters but also demographic and behavioral dimensions. This has implications for how cities deploy resources and issue warnings. Enhanced inspection frequencies for catch basins, targeted community engagement, and incorporating real-time crowdsourced data into forecasting models could improve resilience.

Our results also suggest that response procedures should be reconsidered, particularly for weekends and high density, lower-income neighborhoods. More specifically, procedures

might involve adaptive staffing or computerized systems where complaint are sorted by urgency and responded to accordingly to address temporal and spatial variation in demand. For example, increased field units on weekends, or selective prioritization of reports from regions that historically have had slower resolution, would increase the equity of services. Current parameters such as complaint intake and routing procedures may remain unchanged, yet scheduling, prioritization weights, and resource allocation logic may be adjusted based on complaint patterns, past delays, and neighborhood vulnerability profiles.

Furthermore, a potential analog to Newark’s “Adopt a Catch Basin” program (as referenced in Agonafr et al., 2022) could be piloted in NYC as a low-cost way to supplement municipal maintenance with public participation, particularly in boroughs like Staten Island and Queens that report high complaint rates.

5.6 Future Research

This research is good, and there are several waysthrough which additional research may extend it. One would be to incorporate more detailed weather information, including measurements of hourly precipitation totals, so as to observe in greater precision the impact storms have on the complaint pattern. This would then demonstrate whether complaint about flooding is predominantly occurring throughout days of great rainfall or if complaints occur with even lighter showers, perhaps by reason of underdeveloped infrastructure, in specific neighborhoods.

The next potential improvement would be to consider analyzing the data at a geographically narrower space. Instead of analyzing ZIP codes, the following analyses can be done looking at blocks or intersections. It will give better insight into just where things are breaking down. Adding more granularity to the analysis concerning the sewer system itself, i.e., pipe diameter, maintenance history, drainage capacity, might better quantify where places are more prone to flooding or taking too long a response time.

5.7 Conclusion

In conclusion the analysis shows that 311 complaint data is a good source of data to learn about trends in flood related services and delayed response areas. Catch basin complaints were more frequent and had longer delays compared to street flooding complaints especially in Staten Island and the Bronx. Logistic regression showed socioeconomic indicators like affordability of housing and education were contributing predictors of longer delays. The negative binomial analysis provided evidence that day-to-day volumes of complaints relied significantly on day-to-day complaint volumes and weather.

Beyond trends, the analysis identifies areas wherein service protocols as well as the design of infrastructures can improve. Outcomes show that delays is experienced more

often among low-income as well as density areas especially over weekends. They give awareness to fair and responsive concern in current complaint handling procedures. There may also be further policy in the future directed at dynamic staffing, forward looking maintenance within high risk zip codes and more integrated data usage to better focus flood response action. All of this underlies data driven practice utilization to leverage equity and efficiency in NYC urban flood management.

References

- [1] C. Agonafr, A. Ramirez Pabon, T. Lakhankar, R. Khanbilvardi, and N. Devineni, "Understanding New York City street flooding through 311 complaints," *Journal of Hydrology*, vol. 605, 2021, Art. no. 127300.
- [2] C. Agonafr et al., "A machine learning approach to evaluate the spatial variability of New York City's 311 street flooding complaints," *Computers, Environment and Urban Systems*, vol. 97, 2022, Art. no. 101854.
- [3] A. Basiri, M. Haklay, G. Foody, and P. Mooney, "Crowdsourced geospatial data quality: Challenges and future directions," *International Journal of Geographical Information Science*, vol. 33, no. 8, pp. 1588–1593, 2019.