Analysis on NYC Pedestrian Ramp Complaints: Final Presentation

April 22, 2025

Carlos Correa

CC95353N@PACE.EDU

PRACTICAL DATA SCIENCE

MS in Data Science

Seidenberg School of Computer Science and Information Systems

Pace University

Agenda

- Project Plan Recap
- Modeling Methods
- Findings
- Recommendations

Executive Summary

Context:

As part of a settlement between the City of New York and plaintiffs regarding ADA accessibility, the City agreed to invest $1.55 Billion in the next decade. To make New York City more accessible.

Problem:

Pedestrian ramps and corner complaints can pose safety and accessibility issues.

This project aims to answer:

How can the city of New York improve the efficiency and effectiveness of pedestrian ramp repairs to reduce complaint volumes and minimize repair times?

Project Plan Recap

Deliverable	Due Date	Status
Data & EDA	3/25/25	Complete
Methods, Findings and Recommendations	4/1/25	Complete
Final Presentation	4/22/25	In Progress

Data

Dataset overview

Pedestrian Ramp Complaints Data Set available on NYC Open Data:

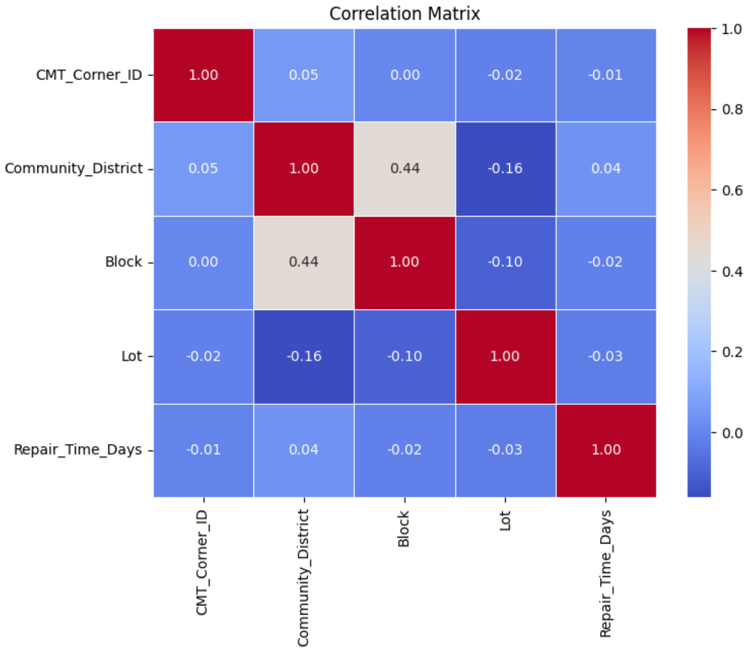
- 4,583 rows. Each row is a complaint or represent a bulk complaint if more than one person made a report about the same corner ramp on the same day.
- Earliest Data point is from September 11, 2018
- Available data since 2018 last updated March 5, 2025
- Dataset was made public on June 20, 2024

Exploratory Data Analysis

Correlation

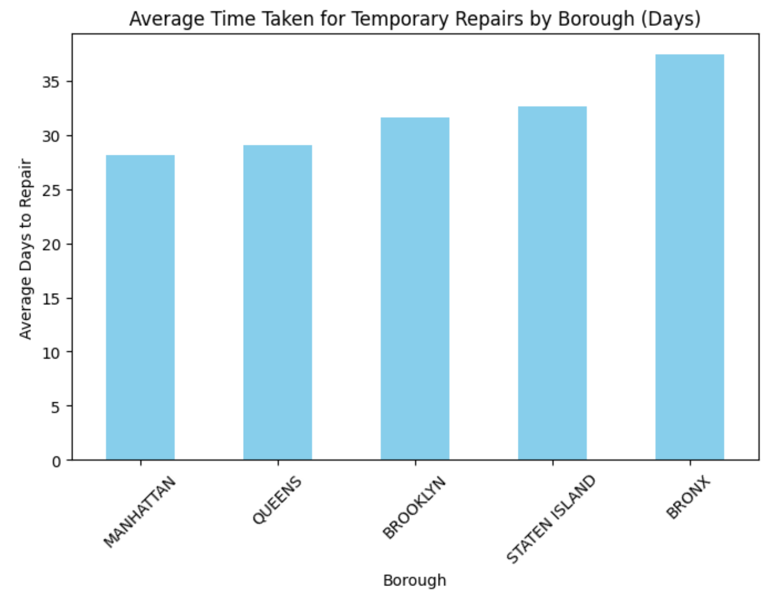
- What does this mean?
- The weak correlations indicate that relationships between variables are nonlinear or influenced by hidden factors.

- Key Takeaway:
- For the purpose of this project, maybe a decision tree model will help to uncover hidden patterns.



How Long Does It Take To Repair a Ramp in Each Borough?

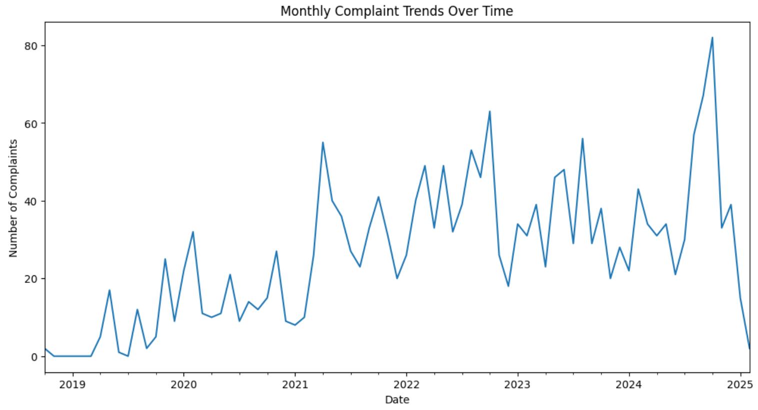
- Key Takeaway:
- Repairs in the Boroughs of Manhattan and Queens advance at a faster pace than the five-borough average of 31 days.
- Assumption:
- The boroughs of Brooklyn, the Bronx and Staten Island are caused by bureaucratic delays.
-



Are Complaints Seasonal?

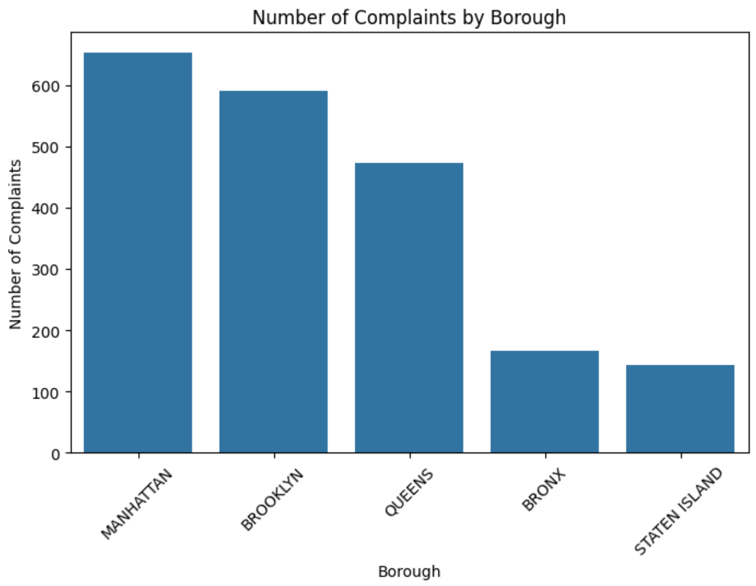
- Key Takeaway:
- This graph clearly shows there is a seasonality trend in the spikes and valleys of the complaints.
- Summer and Winters seasons are tied to spikes in complaints.

- Conclusion:
- The city should prepare in advance for the summer and winter to avoid complaints and repairs backlog.



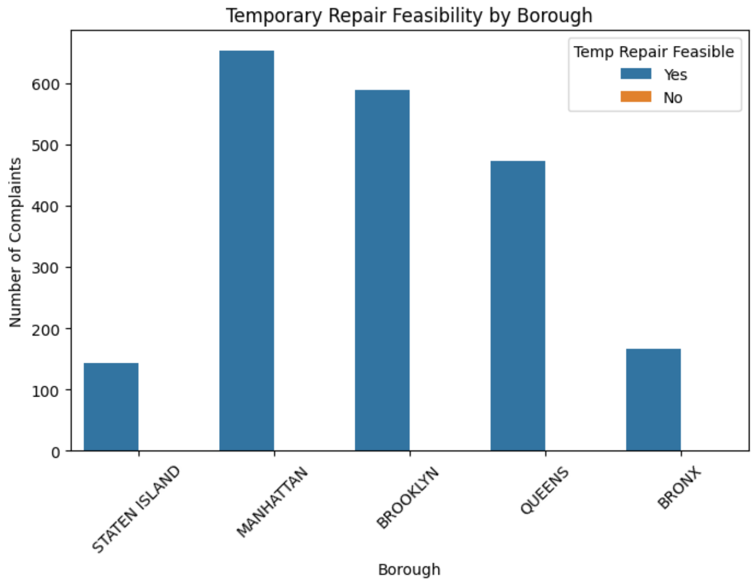
Number of Complaints by Borough

- Key Takeaway:
- Manhattan and Brooklyn lead by number of complaints. Which means that higher amount of works needs to be done in these boroughs.
- Assumption:
- The high numbers of complaints in Manhattan, Brooklyn and Queens are a result of higher number of ramps, higher population and higher foot traffic.



Feasibility of Temporary Repairs

- Key Takeaway:
- Plan for quick and low on resources temporary repairs based on feasibility.
- Deploy more repair crews where temporary repairs will help reduce the complaint backlog.
- Assuming non-temporary repairs require more work and resources, assign “heavy duty” crews where permanent needed are high and aim for “a one and done repair”.



Model Selection

Model Selection

Initially we found out that the features are not linearly correlated.

Key Takeaways:

What it does:
•Flags complaints as "High Priority" if they match urgency signals like:

- High community demand
- Infeasible repairs
- Bulk reporting

.

Model Type: Transparent Decision Tree

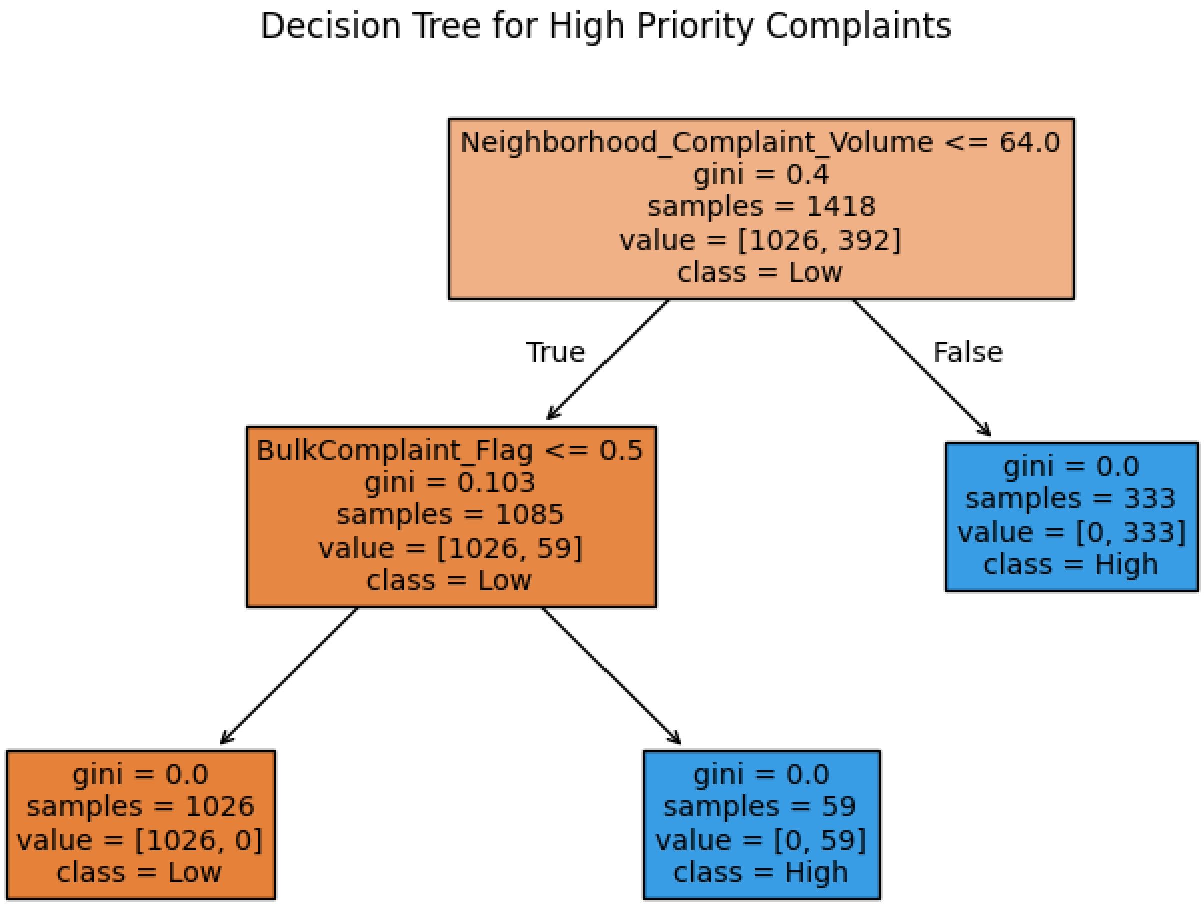
Why this approach?

- Clear logic: The model works like a flowchart — anyone can follow it
- Fast decisions: Makes instant classifications using a simple "yes/no" rule path
- Fair and consistent: Ensures all complaints are evaluated by the same objective criteria
- Explainable: Helps operations teams understand why a complaint is considered high priority

Model Selection

The model determined that a Neighborhood Complaint Volume of 64 is the cutoff between low and high priority

Bulk Complaint Flag can either have a value of 0 or 1 so the 0.5 is used to determine whether it is a bulk complaint or not.



[Technicalities]

Feature Engineering

Why? The dataset didn't contain optimal features, but combining some features allowed us to create new features to work on.

Key Takeaways:

More complaints = higher attention needed

Feature: Neighborhood_Complaint_Volume

• Counted how many complaints came from each Community_District.

•Why:

High complaint volume may signal:

- Systemic infrastructure issues
- Greater public demand
- Potential equity gaps in service delivery

Business Insight: More complaints = higher attention needed

Feature Engineering

Key Takeaways:

Business Insight: Bulk = potentially higher operational urgency.

Feature: BulkComplaint_Flag (1 = Yes, 0 = No)

•Converted the text field BulkComplaint ("Yes"/"No") into a binary numeric value.

•Why:

Assumed Bulk complaints typically come from citywide audits or coordinated community efforts.

- •Represent many problem sites
- •Often require faster and more strategic responses

Feature Engineering

Key Takeaways:

Business Insight: No temp fix = high impact = higher urgency.

Feature: Temp_Repair_Feasible_Flag (1 = Yes, 0 = No)

- Converted Temp_Repair_Feasible from text to binary numeric.

- Why:

If a temporary repair is not feasible, it usually means:

  - The ramp is inaccessible until full repair
  - The situation is more urgent and impactful for residents with disabilities

Feature Engineering

Key Takeaways:

Technical Insight: Prevents large values from dominating the model.

Feature: Normalized_Volume (scaled between 0 and 1)
• Applied min-max scaling to Neighborhood_Complaint_Volume to normalize it for modeling.

•Why:
Some models (especially scoring or regression-based) perform better when numeric values are scaled evenly.

Feature Engineering

Key Takeaways:

This simulates real-world prioritization using measurable features.

•It allows us to train the model to learn patterns from known high-priority complaints.

•Once trained, the model can be used to predict High_Priority = 1 for new complaints, helping prioritize them automatically

Feature: Priority_Score (custom formula)

•Combined all engineered features into a scoring formula. Ultimately this is the outcome variable of the model

•Why: This formula reflects business intuition about what makes a complaint more urgent.

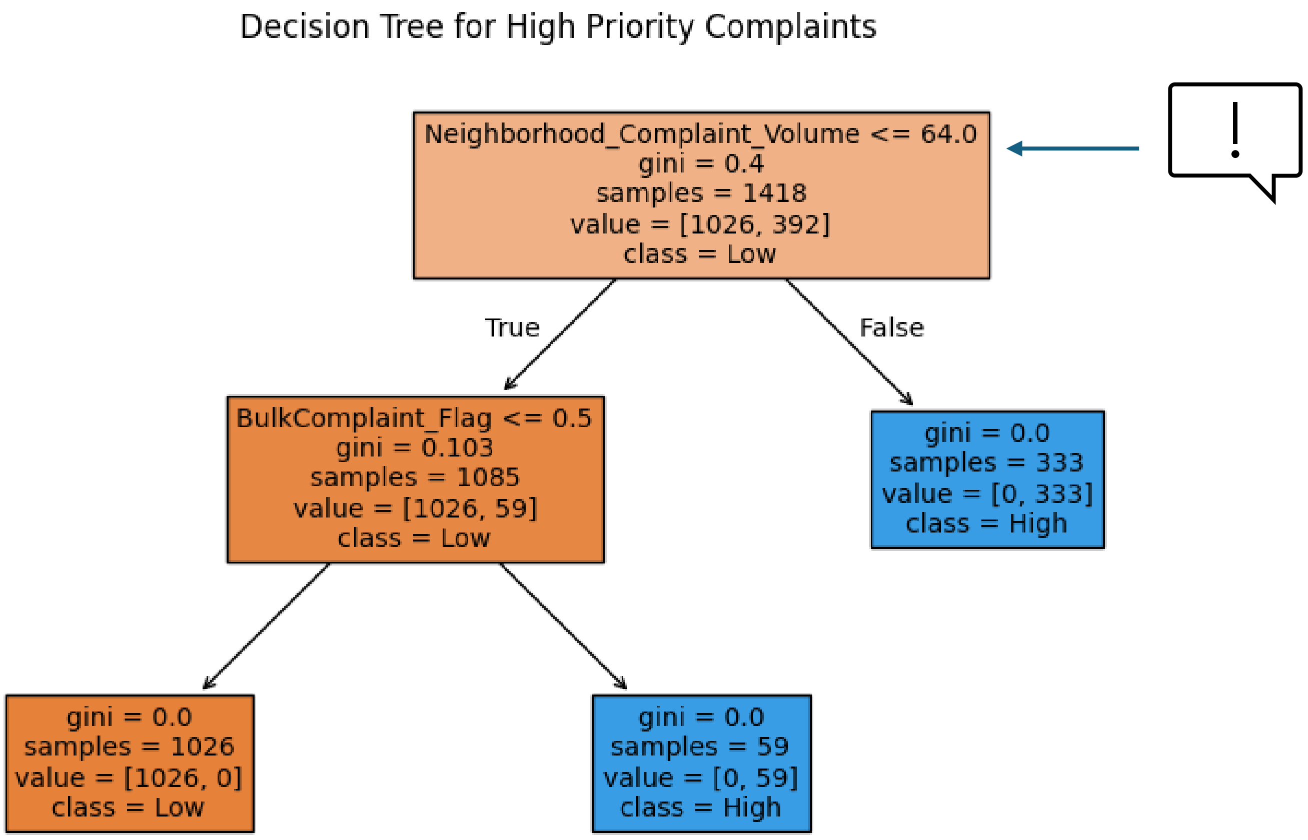
•Each factor is weighted based on perceived importance.

Findings

Neighborhood Volume Drives Urgency

Key Takeaways:

Complaint clusters signal widespread infrastructure problems:
The model prioritizes areas where multiple residents are raising concerns

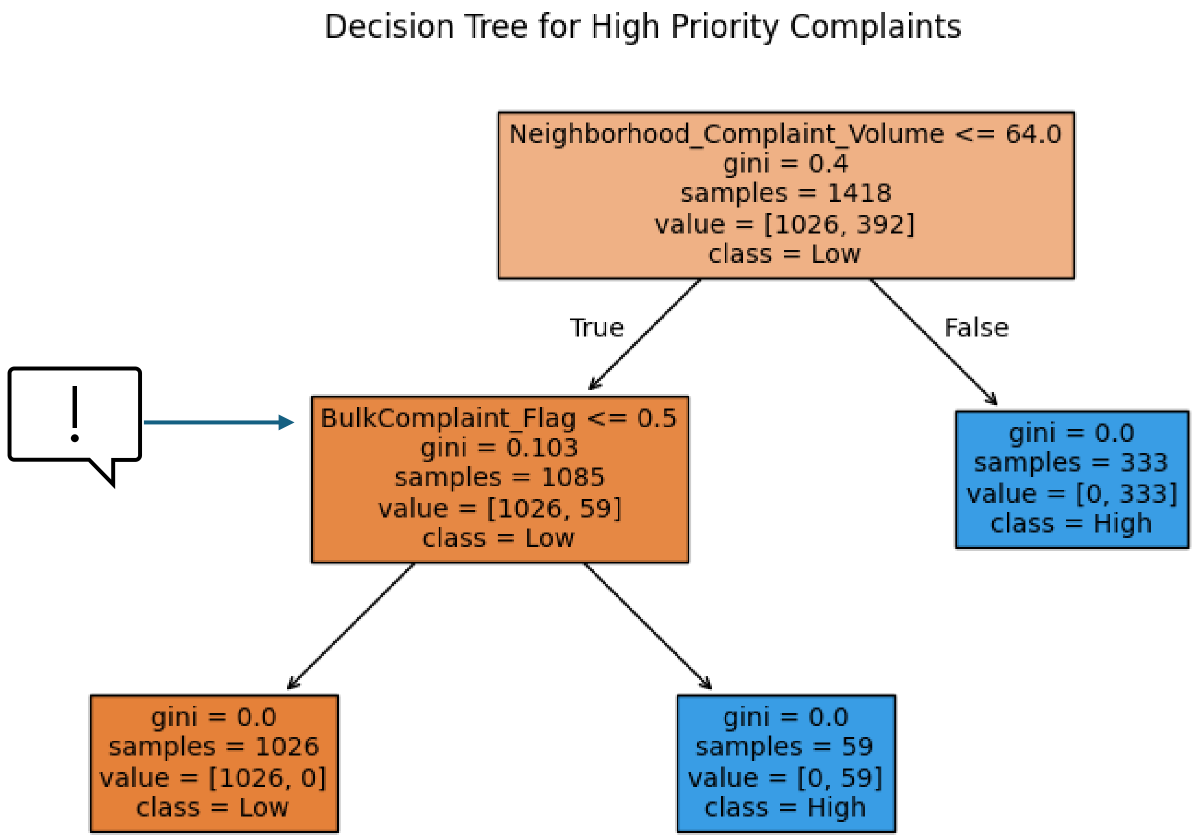


- The first and most important split in the tree was based on total complaints in a community district
- Neighborhoods with more than 64 complaints were automatically flagged high-priority

Bulk Complaints Drive Action

Key Takeaways:

The city can improve efficiency by addressing grouped complaints first: Maximize impact with less resources.



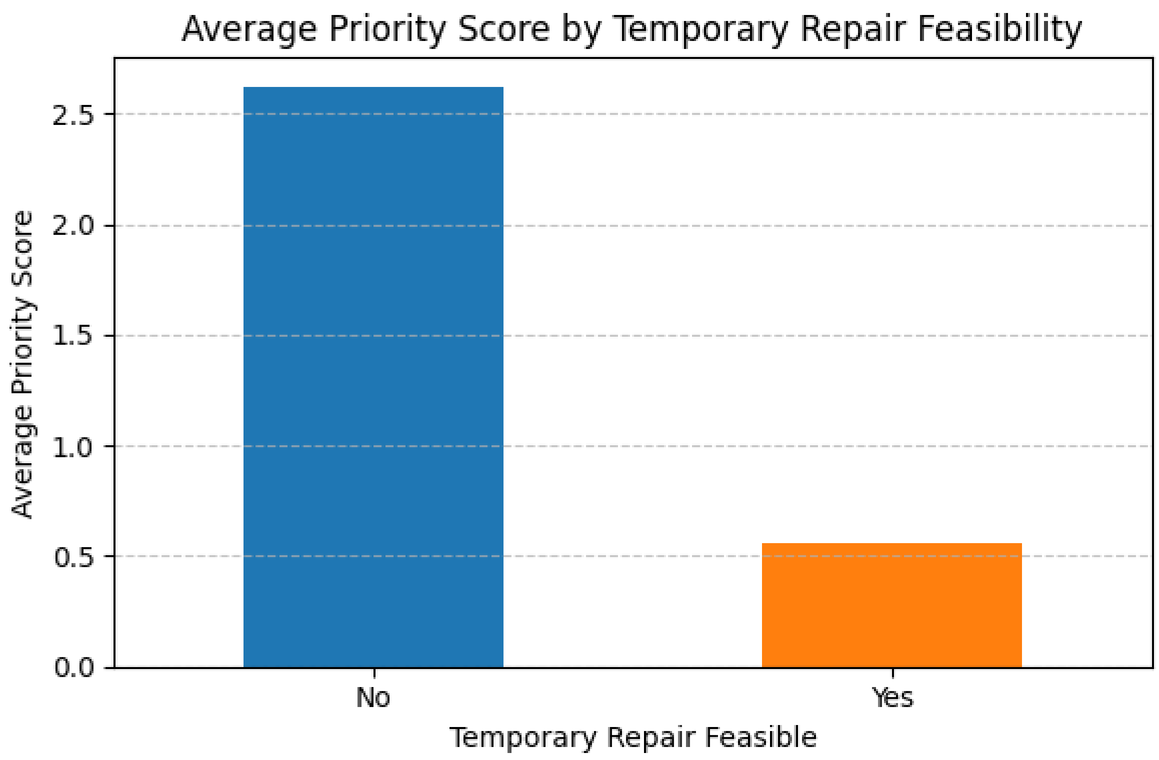
- In quieter districts (≤ 64 complaints), only complaints flagged as bulk were labeled high-priority
- Complaints not part of bulk reports were deprioritized

Temporary Repairs Not Feasible = Greater Urgency

Key Takeaways:

If a sidewalk ramp can't be fixed temporarily, this means these ramp locations will remain inaccessible and present compliance risks.

Important to fix it with enough time to avoid compliance risk during the yearly Summer and Winter spike in complaints.



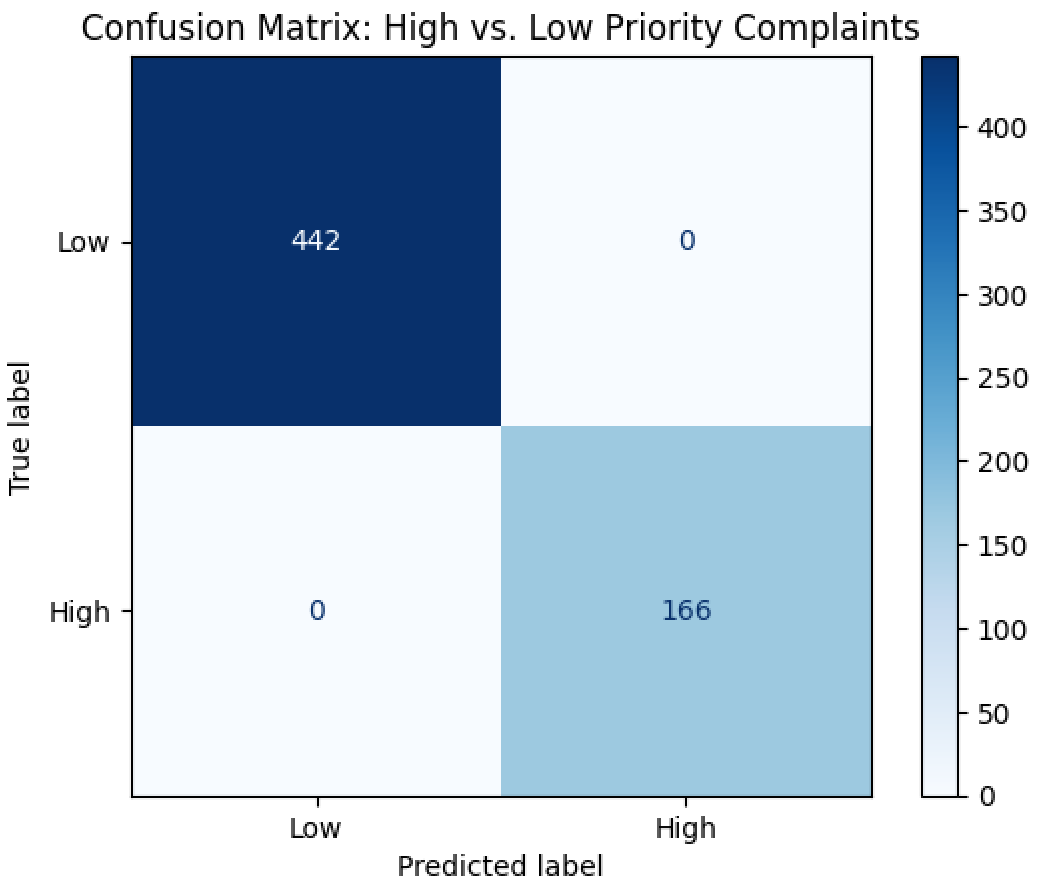
- Complaints where temporary repairs were not feasible had significantly higher priority scores
- These complaints were more likely to result in long-term inaccessibility

Model Summary: Transparent, Accurate, Actionable

Key Takeaways:

This model offers a fast, fair, and transparent way to label new complaints based on historical urgency.
With an F1 Score of 1.0, which means it had perfect recall.

It allows for an easy to explain and transparent process.



- The model correctly classified all high-priority and standard-priority complaints in the test set

- Built with just three engineered features which are fully explainable and reproducible (BulkComplaint_Flag, Temp_Repair_Feasible_Flag, Neighborhood_Complaint_Volume)

Recommendations

Integrate Spatial and Demographic Data

Key Takeaways:

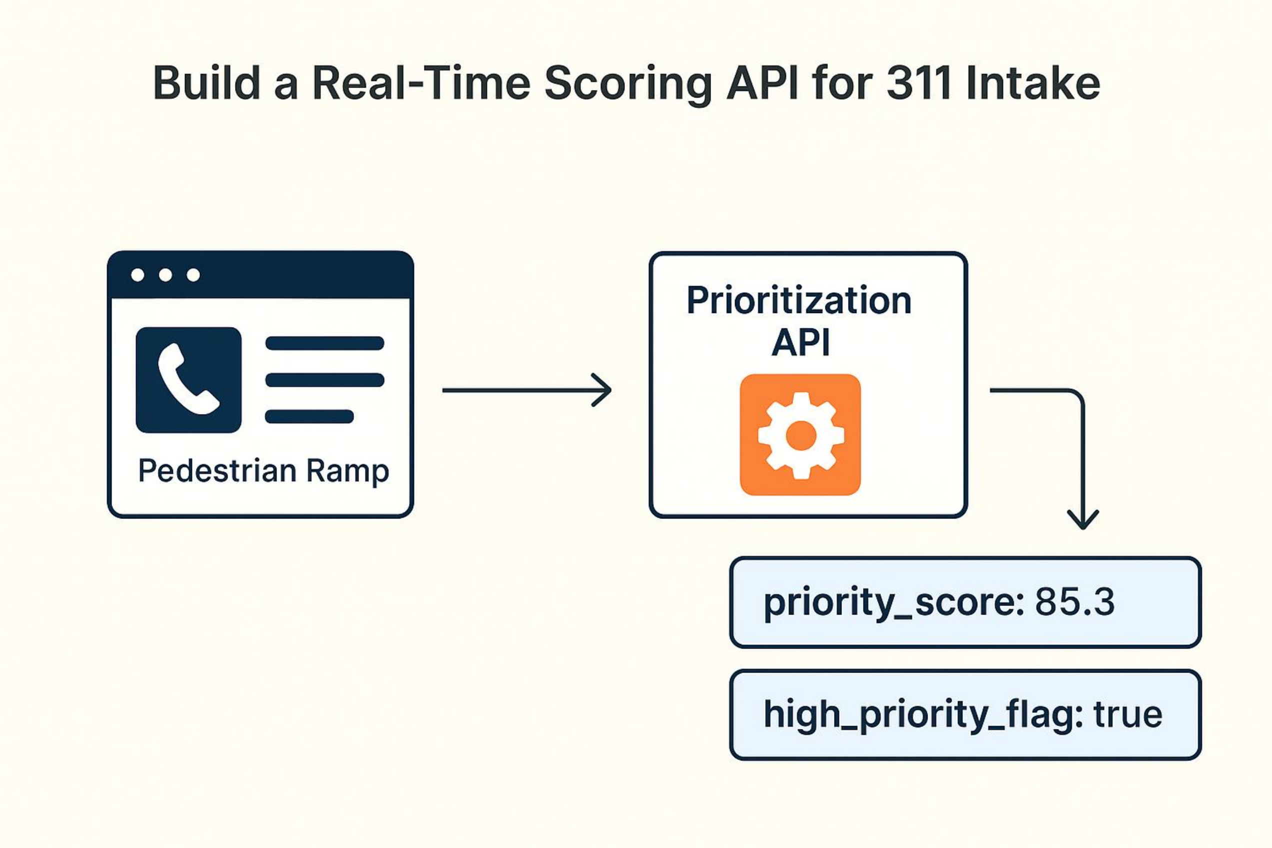
This would help the city target underserved, vulnerable areas, not just areas with high complaint volume.

- Add spatial joins to link each complaint to:
- **Census data** (e.g. % of population with disabilities, income levels)
- **Sidewalk accessibility scores** or known infrastructure gaps

Integrate Spatial and Demographic Data

Key Takeaways:

Enables instant, data-driven labeling: Allowing field crews, repair schedulers, and accessibility teams can act faster and smarter.



Right now, the model is static and used retrospectively. Turning it into a live tool would prioritize complaints as they come in.
How to do it:

- Wrap the model into an API that takes in 311 complaint data
- Have it return a priority_score + high_priority_flag
- Integrate this into the city's 311 or DOT complaint triage system

Appendix

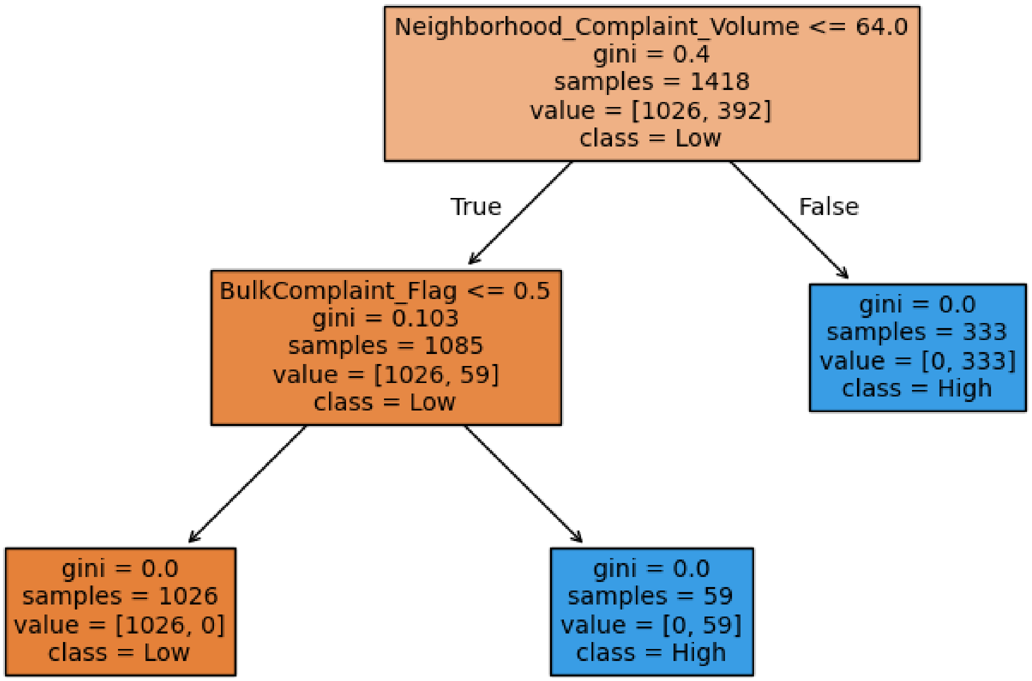
Model Selection

Key Takeaways:

Used For:
•Binary classification: High_Priority = 1 (Top 25% of complaints by urgency score)

Why not more complex models?
•Transparency is essential for public trust and decision-making
•Avoids “black box” concerns from city stakeholders. The “if-then” logic is transparent, and the model is defensible and equitable.

Decision Tree for High Priority Complaints



Why this model?
• It is easy to understand and visualize, ideal for public-sector applications
•Non-linear: Captures complex interactions between features like complaint volume and repair feasibility
•Fast & Lightweight: Works well with small datasets and doesn’t require heavy computational power
•Deterministic Rules: Produces repeatable logic that aligns with operational policies

GitRepo Link

https://github.com/cjcorrea12/NYC-Ramp

Dataset Source

https://data.cityofnewyork.us/Transportation/Pedestrian-Ramp-Complaints/jagj-gttd/about_data

PDF of Settlement for the Class Lawsuit

The settlement between the City of New York and the CENTER FOR INDEPENDENCE OF THE DISABLED, NEW YORK can be found at:

- https://www.nycpedramps.info/sites/default/files/2019-07/Pedestrian%20Ramp%20Settlement%20Agreement--Final%20Approved%207-23-2019.pdf