# 311 Service Request Analysis & Prediction

• • •

Springboard - Data Science Career Track

Tasmia Kayenat
July 10, 2025

# Content

- Problem Statement
- Project goals
- Dataset Overview
- Data Wrangling
- EDA - Exploratory Data Analysis
- Preprocessing and Modeling
- Recommendations

# Problem Statement:

- NYC's 311 data reflects how the city responds to resident concerns.

- Delays and inconsistencies in complaint resolution reduce public satisfaction.

- Analyze trends and build a model to predict delayed service requests.

# Project Goals

## Goal 1

Identify high-volume and slow-responding complaint types

## Goal 2

Highlight underperforming agencies, underserved boroughs and predict which complaints are likely to be delayed

## Goal 3

Deliver actionable recommendations for improving city responsiveness

# Understanding the Data

# Data Overview

- Source: NYC Open Data — 311 Service Requests

- Key fields: Created Date, Complaint Type, Agency, Borough, Response Time

- ~185358 rows after cleaning

# Data Wrangling

- Dropped irrelevant columns
- Converted time columns to datetime
- Feature Engineered "Response Time (Hours)" and "Is Closed"
- Standardized categorical variables
- Focused on top 3 boroughs
- Removed duplicates

# EDA - Exploratory Data Analysis

## Common complaints

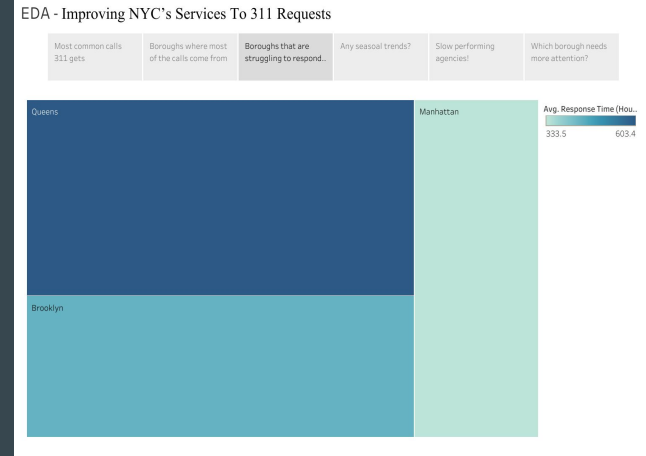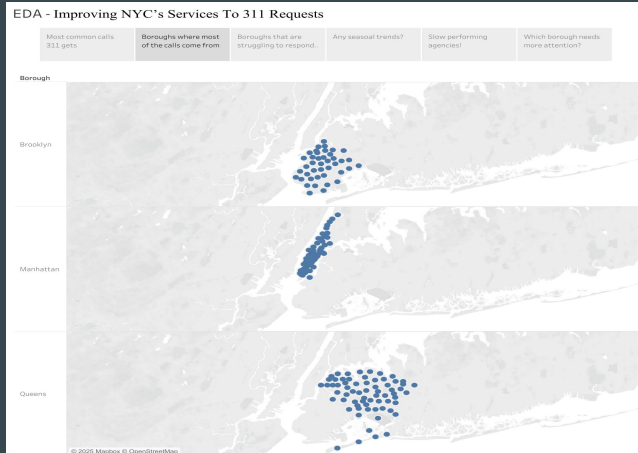Key insight: Noise complaint dominate, especially Noise - Residential.
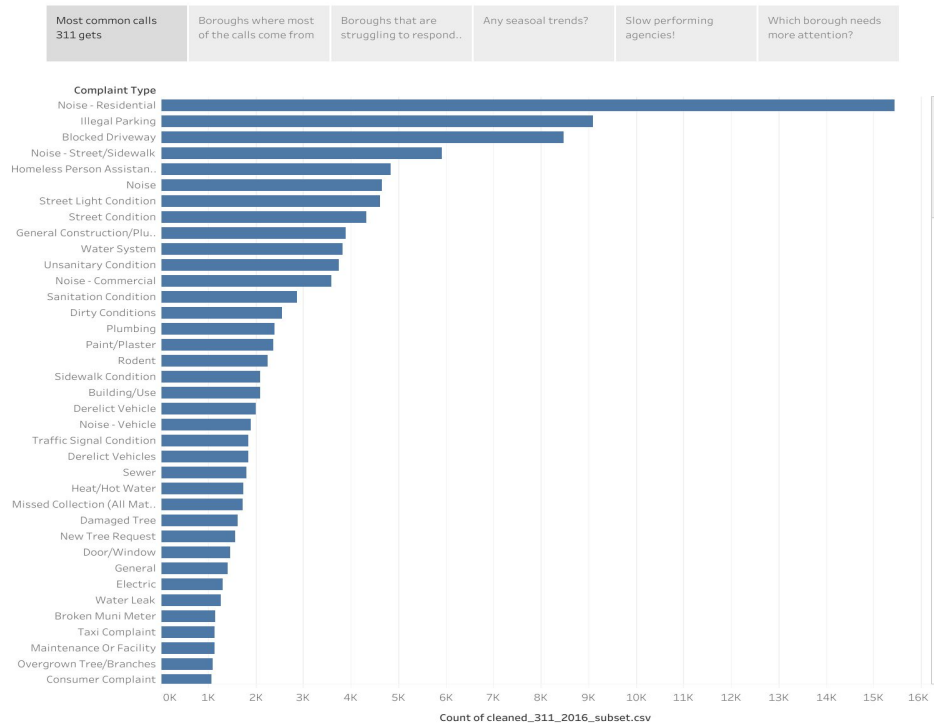
# EDA - Exploratory Data Analysis

## Borough Trends

Key insight: Queens had the highest volume
and the slowest response time.

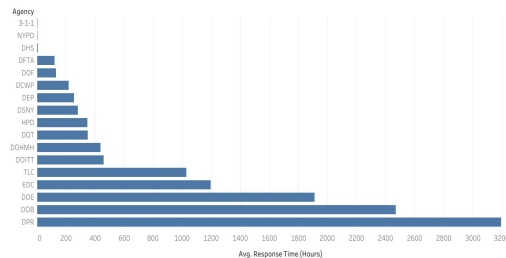# EDA - Exploratory Data Analysis

## Agency Performance & Time Trends

Key insight:

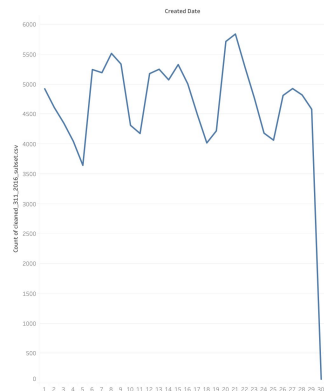NYPD = Slowest agency

Fewer complaints on Fridays and on weekends

# Preprocessing & Modeling
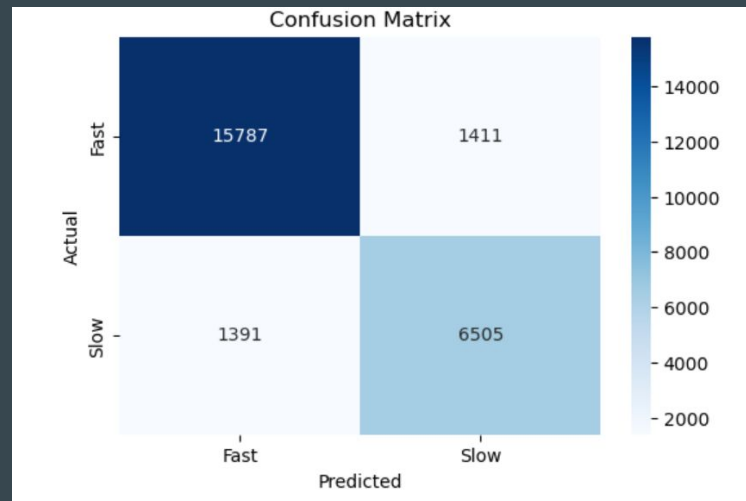
Model: Logistic Regression
Target: Slow response
Features: Complaint Type, Borough, Agency

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.92      | 0.92   | 0.92     | 17198   |
| 1            | 0.82      | 0.82   | 0.82     | 7896    |
|              |           |        |          |         |
| accuracy     |           |        | 0.89     | 25094   |
| macro avg    | 0.87      | 0.87   | 0.87     | 25094   |
| weighted avg | 0.89      | 0.89   | 0.89     | 25094   |

```
Confusion Matrix:
 [[15787  1411]
 [ 1391  6505]]
ROC AUC Score: 0.9539899022211413
```



Confusion Matrix

|        | Fast  | Slow |
|--------|-------|------|
| Fast   | 15787 | 1411 |
| Slow   | 1391  | 6505 |

# Preprocessing and Modeling



Top 15 Most Important Features (Logistic Regression)
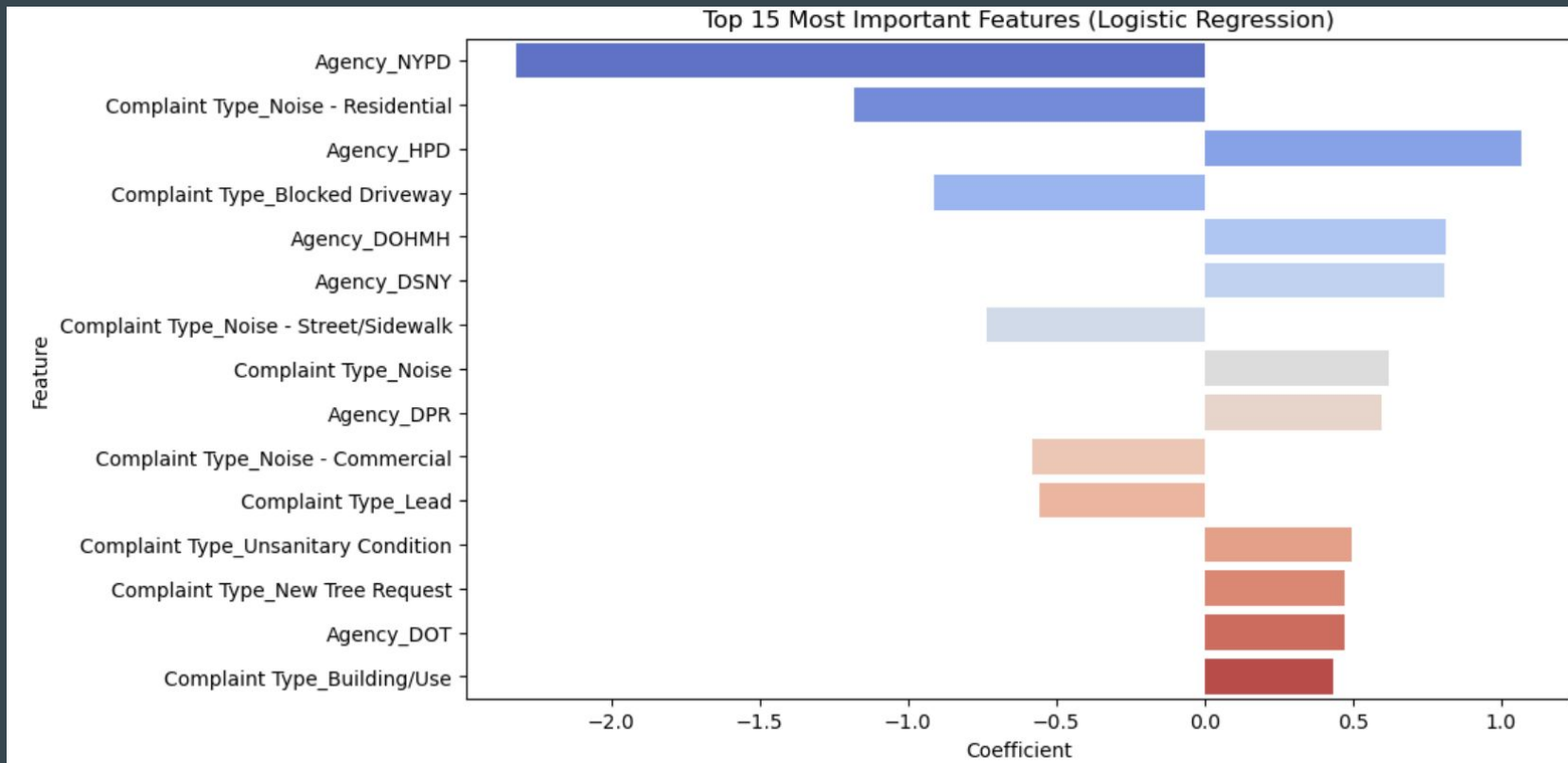
# Recommendations:

- Improve NYPD's responses for noise complaints

- Allocate more resources to Queens and Brooklyn

- Deploy a predictive dashboard to prioritize at risk complaints.

# Thank you

Tasmia Kayenat

———

aashatasmia1999@gmail.com