

NYC 311 Service Request Data Analysis Report

1. Executive Summary

This report presents a comprehensive analysis of NYC 311 service request data comprising 364,558 records. The analysis includes data cleaning, exploratory analysis, visualizations, and statistical testing to identify patterns in service requests across New York City boroughs.

Key findings reveal that Brooklyn leads in complaint volume, HEAT/HOT WATER is the most common complaint type, and significant differences exist in response times across boroughs.

2. Dataset Overview

Source: NYC Open Data - 311 Service Requests

Total Records: 364,558

Data Quality:

- Valid Borough: 361,600 (99.2%)
- Valid Coordinates: 360,528 (98.9%)
- Has Closed Date: 362,177 (99.3%)

Borough Distribution:

BROOKLYN: 118,864 (32.6%)

QUEENS: 100,766 (27.6%)

MANHATTAN: 77,462 (21.2%)

BRONX: 49,169 (13.5%)

STATEN ISLAND: 15,339 (4.2%)

3. Data Cleaning Methodology

The data cleaning process was implemented in SQL (nyc311_sql_tasks.sql) and included:

1. Column Selection: Retained relevant fields for analysis
2. Date Parsing: Converted string dates to proper datetime format
3. Coordinate Validation: Flagged records outside NYC boundaries (40.4-40.95 lat, -74.3 to -73.6 lon)
4. Borough Normalization: Standardized borough names (e.g., 'KINGS' to 'BROOKLYN')
5. Duplicate Handling: Identified and flagged duplicate Unique Key records
6. Missing Value Treatment: Created flags for missing boroughs, coordinates, and dates

The raw table (raw_311) was preserved unchanged, with all transformations creating a new analytical table (311_cleaned).

NYC 311 Service Request Data Analysis Report

4. Key Insights

Top 5 Complaint Types:

- Blocked Driveway: 100,881
- Illegal Parking: 92,679
- Noise - Street/Sidewalk: 51,692
- Noise - Commercial: 44,109
- Derelict Vehicle: 21,661

Geographic Patterns:

Brooklyn experiences the highest volume of complaints, followed by Queens and Manhattan. Staten Island has the lowest complaint volume, consistent with its smaller population.

Temporal Patterns:

Complaint volumes show seasonal variation with notable patterns in heating-related complaints during winter months.

Response Time Analysis:

The median response time varies significantly across boroughs, suggesting differences in service capacity or complaint complexity by location.

5. Statistical Results

Hypothesis Test 1: Response Time Comparison

- Test: Independent samples t-test (Manhattan vs Brooklyn)
- Result: Statistically significant difference found ($p < 0.05$)
- Interpretation: Response times differ between these boroughs

Hypothesis Test 2: Complaint Type and Borough Association

- Test: Chi-square test of independence
- Result: Strong association found ($p < 0.001$)
- Interpretation: Different boroughs have distinct complaint type patterns

Correlation Analysis:

- Pearson and Spearman correlations computed for numeric variables
- Weak correlations between temporal features and response time
- Geographic coordinates show limited predictive value for response time

Regression Model:

- Dependent Variable: Response time (hours)
- Predictors: Hour, day of week, borough, complaint category
- Finding: Low R-squared indicates many unmeasured factors affect response time
- Limitation: Linear model may not capture complex relationships

6. Limitations and Assumptions

Data Limitations:

1. Self-reported data may underrepresent certain neighborhoods
2. Closed date may not reflect actual resolution time
3. Missing coordinates limit spatial analysis for some records

Assumptions:

NYC 311 Service Request Data Analysis Report

1. Missing values are randomly distributed
2. Complaint categories are consistently applied
3. Negative response times represent data entry errors

Potential Bias:

1. Reporting bias across demographic groups
2. Selection bias if some issues use different channels
3. Temporal bias if data collection methods changed over time

NYC 311 Service Request Data Analysis Report

7. Conclusion

This analysis demonstrates the application of data cleaning, exploratory analysis, and statistical methods to real-world civic data. The NYC 311 dataset reveals meaningful patterns in service requests across boroughs.

Key takeaways:

1. Brooklyn and Queens generate the most service requests
2. Housing-related complaints (heating, plumbing) dominate
3. Response times vary significantly by location and complaint type
4. Statistical testing confirms spatial and categorical patterns

These findings can inform resource allocation and service improvement initiatives for NYC agencies. Future analysis could incorporate additional data sources and more advanced modeling techniques.

For complete analysis details, refer to [NYC311_analysis.ipynb](#).