Analyzing Boston 311 Service Requests: Efficiency, Equity, and Seasonality

By Joel Thomas Zachariah, Bikramjeet Singh, Khanh Huynh

Contents

Introduction:

The dataset used for this project is the "311 Service Requests" dataset, obtained from the Boston Data Portal (https://data.boston.gov/dataset/311-service-requests). The dataset contains detailed records of 311 service requests submitted by residents of Boston from 2018 to 2023. The merged dataset includes over one million records with variables capturing request details such as case title, neighborhood, and time-to-close metrics.

Research Questions:

The analysis aims to address the following questions:

1. What are the most common service request types, and how do they vary by season?

2. How do neighborhoods differ in their average time to close service requests?

3. What factors predict the time it takes to close a service request?

Outcome Variables:

1. **Time to Close Requests**: The time (in days) between the open date and closed date for each service request.

2. **Seasonal Distribution of Request Types**: Frequency of service requests by season and type.

Predictor Variables:

Using the dataset columns, the following predictors are most relevant:

1. **Case Title (case_title)**: Type of service requested.

2. **Closure Reason (closure_reason)**: Provides context for request resolution.

3. **Neighborhood (neighborhood_services_district)**: Geographic location of requests.

4. **Department (department)**: Responsible department.

5. **Source (source)**: Submission method.

6. **Season (open_dt)**: Derived variable (Winter, Spring, Summer, Fall).

7. **Day of the Week (open_dt)**: Weekday or weekend.

8. **Days to Acknowledge (target_dt - open_dt)**: Time before acknowledgment.

9. **Volume of Requests**: Aggregate daily/seasonal requests.

Feature Engineering:

1. **Season Variable**: Derived from the open_dt variable using the month to categorize requests into Winter, Spring, Summer, and Fall.
2. **Cleaned Case Titles**: Grouped infrequent case titles into "Other" and combined similar categories (e.g., "Street Light Outages" and "Street Light Knockdowns" were grouped into "Street Lights").
3. **Binary Variables**: Future analysis may include converting continuous variables such as time_to_close into binary categories (e.g., requests resolved quickly vs. not).

Sub-Groups for Analysis:

1. **By Season**: Analyze trends for request types and frequencies.
2. **By Neighborhood**: Compare average time to close and frequencies of request types.
3. **By Source**: Evaluate differences in time-to-close metrics based on submission channels.

Descriptive Statistics Tables:

Table 1: Summary Statistics by neighborhood

neighborhood	mean_time_to_close	median_time_to_close
South Boston	50.08068	0.3356308
Charlestown	38.21812	0.6473090
Boston	32.91186	0.4553009
Fenway / Kenmore / Audubon Circle / Longwood	32.75324	0.5854167
South Boston / South Boston Waterfront	32.36812	0.2490162
Downtown / Financial District	30.33602	0.3183449
Jamaica Plain	29.39188	0.7737384
West Roxbury	28.59652	1.1990104
Allston	28.17908	0.7383449
Back Bay	28.04876	0.4204167
Greater Mattapan	26.28508	0.8659086
Hyde Park	26.18937	0.9127546
Brighton	26.08055	0.6471644
Mattapan	24.53408	0.9169850

neighborhood	mean_time_to_close	median_time_to_close
Mission Hill	24.20384	0.5315394
Beacon Hill	24.12336	0.2719850
Roslindale	23.25524	0.8534259
Allston / Brighton	22.73853	0.5909086
South End	22.43589	0.2298495
East Boston	22.33916	0.5215336
Roxbury	20.26923	0.5614236
Dorchester	20.09554	0.6931134
Chestnut Hill	10.32684	0.6470312

Table 2: Grouped Descriptive Statistics by Season

season	mean_time_to_close	median_time_to_close	total_requests
Winter	24.72728	0.5614583	320,866
Spring	28.62248	0.5914815	353,614
Summer	26.45733	0.5834838	408,275
Fall	23.77183	0.5578588	369,738

Table 3: Grouped Descriptive Statistics by Year

year	mean_time_to_close	median_time_to_close	total_requests
2018	38.17743	0.6589236	239,680
2019	35.20734	0.6481134	232,615
2020	33.85252	0.5222222	228,343
2021	22.69980	0.4671759	251,578
2022	13.46532	0.5452778	253,456
2023	13.98960	0.5803819	246,821

Analytical Plans and Methods:

1. **Linear Regression**

- Model 1: Time to Close Requests (Outcome 1)

- Predictors: Case Title, Neighborhood, Source, Season
- Objective: Identify significant predictors of time-to-close metrics.

- Model 2: Seasonal Distribution of Service Request Types (Outcome 2)
  - Predictors: Season, Case Title, Neighborhood, Source
  - Objective: Understand how request types vary seasonally.

2. **Chi-Square Test**

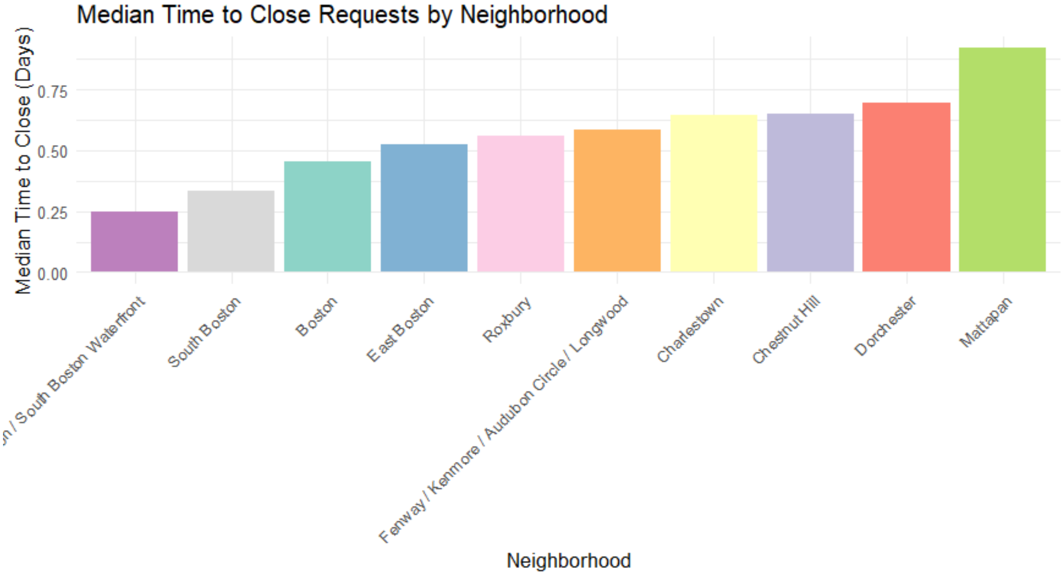
- Used to evaluate the relationship between case_title and season.

3. **Visualization Methods**

- Seasonal trends for top request types using bar charts.
- Neighborhood disparities in time-to-close metrics using grouped bar plots.
- Interactive facets to compare service types across seasons.

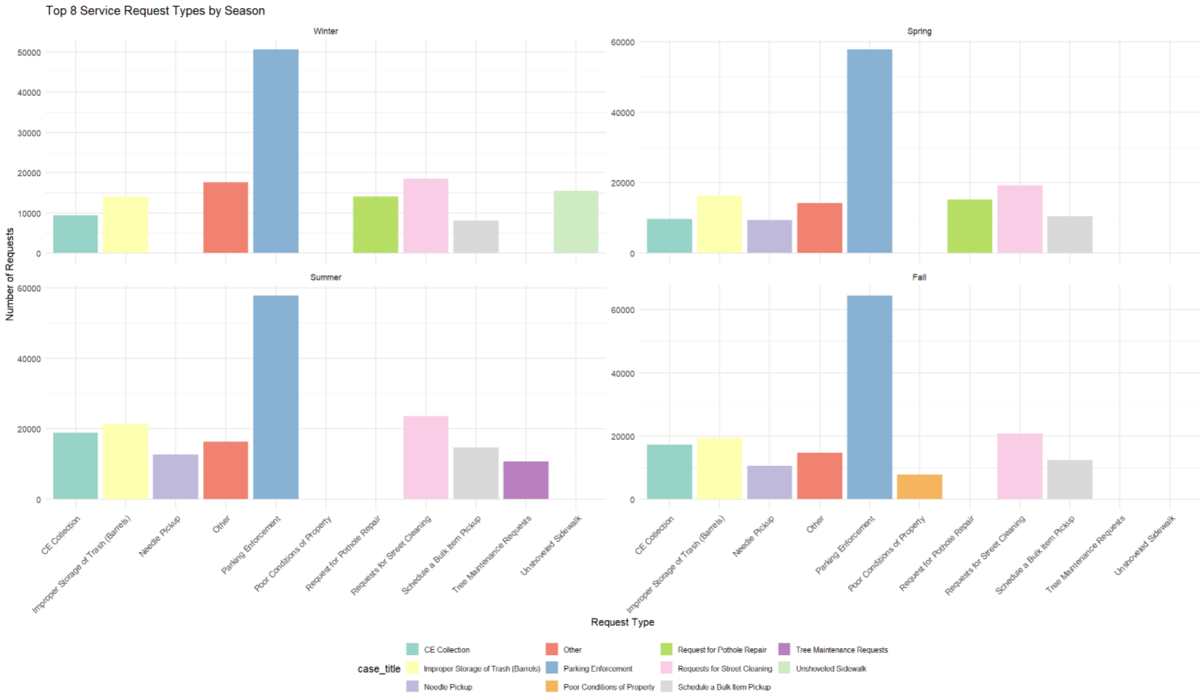
Visualization and Analysis:

Neighborhood Disparity in Time to Close



The bar chart titled **"Median Time to Close Requests by Neighborhood"** displays the median time taken to close service requests across various neighborhoods. The x-axis represents different neighborhoods, while the y-axis shows the median time to close in days. The chart reveals that the **Seaport/South Boston Waterfront** neighborhood has the shortest median closure time, while **Mattapan** has the longest. Other neighborhoods, such as South Boston and Boston, have relatively shorter closure times compared to areas like Dorchester and Chestnut Hill, which experience longer delays. This visualization provides insights into the efficiency of service request handling across different neighborhoods.

Top 8 Service Request Based Seasonally



The bar chart titled **"Top 8 Service Request Types by Season"** presents the distribution of service requests across Winter, Spring, Summer, and Fall. The x-axis represents different service request types, while the y-axis shows the number of requests. **Parking Enforcement** consistently has the highest number of requests across all seasons, significantly surpassing other categories. Other commonly reported issues include **Improper Storage of Trash (Barrels), Requests for Street Cleaning, and CE Collection**, which maintain relatively steady numbers across seasons. Seasonal variations are also observed, with **Unshoveled Sidewalk** requests peaking in Winter and maintenance-related requests showing consistency throughout the year. The chart provides insight into the demand trends for various municipal services across different times of the year.

Methods Used to Analyze Data

Predicting Time to Close and Seasonal Trends

1. Linear Regression for Time-to-Close Service Requests

To understand the factors influencing, multiple linear regression model was applied with **time_to_close** as the dependent variable. The predictor variables included **case title (type of request), neighborhood, source of submission, and season**. The regression identified which factors, particularly **neighborhood** and **case title**, had the strongest association with **time_to_close**.

The regression output provides statistical significance for each predictor. The findings indicate:

- **Neighborhood and case title** have the strongest impact on the time required to resolve a service request.
- **Submission source** (phone, mobile app, or website) also influences response time, with phone-based requests tending to close faster.
- **Seasonal effects** are evident, with some request types taking significantly longer in winter months compared to summer.

This analysis helps in identifying inefficiencies and areas where service response can be optimized.

2. Regression Analysis of Seasonal Trends in Request Volume

A second regression model was constructed to analyze how **seasonality affects the frequency of service requests**. The dataset was preprocessed to ensure categorical variables, such as season, case title, and neighborhood, were properly formatted. To investigate trends, **service request counts** were aggregated by **season, case title, neighborhood, and submission source**. The regression model was defined as:

The regression results confirm the following insights:

- **Certain request types** (e.g., parking enforcement, streetlight issues) exhibit strong seasonal patterns.
- **Seasonal fluctuations** were statistically significant, demonstrating peaks in winter months for snow-related services and in summer for street maintenance.
- **Neighborhood effects** remain consistent, with high-volume request areas maintaining seasonal variations.

The results underscore the need for **season-specific resource allocation**, ensuring efficient response times across different times of the year.

To ensure the robustness of our seasonal trend regression model, we conducted VIF analysis to assess potential multicollinearity among predictor variables.

```
> print(vif_values)
                 GVIF  Df GVIF^(1/(2*Df))
season       1.065167   3        1.010578
case_title   5.458198 236        1.003602
neighborhood 1.229598  22        1.004708
source_var   4.312820   6        1.129527
> 
```

The **VIF analysis** confirms that multicollinearity is not a concern in the regression model. All values are close to 1, indicating minimal correlation between predictors. This ensures the reliability of the regression results.

<u>Chi-Square Test Between Season and Request Type</u>

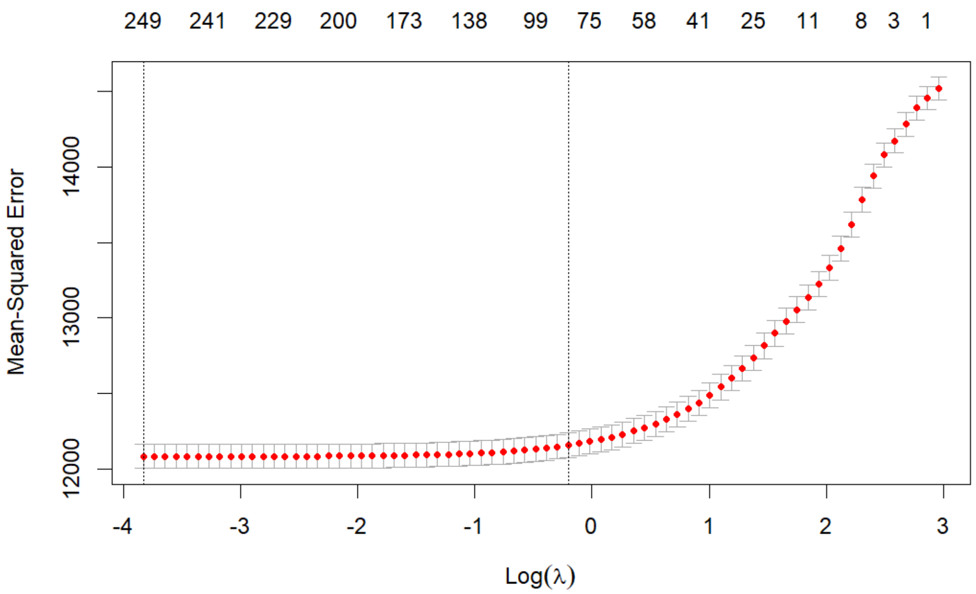
To evaluate whether there is a significant relationship between season and service request type, a Chi-square test for independence was conducted. The test assesses whether the frequency of different service request types varies significantly across seasons, , identifying potential seasonal trends in service demand.

```
> print(chi_sq_test_cleaned)

        Pearson's Chi-squared test

data:  season_case_table_cleaned
X-squared = 112867, df = 708, p-value < 0.00000000000000022
```

Since the p-value is extremely small (< 0.001), the results indicate a strong statistical association between season and request type. This means that service request distributions are not random but significantly influenced by seasonal trends.

LASSO Regression

To enhance the predictive performance of the regression model and reduce overfitting, LASSO regression was used:




```
> print(lasso_model)

Call:  cv.glmnet(x = x, y = y, alpha = 1)

Measure: Mean-Squared Error

    Lambda Index Measure    SE Nonzero
min 0.0217    74   12082 79.55     249
1se 0.8164    35   12157 81.20      80
> |
```

The LASSO model confirms that certain variables are not significant predictors of the response variable, allowing for feature reduction and improved model interpretability.

Analysis of Licensing Data and Parking Complaints Across Zip Codes

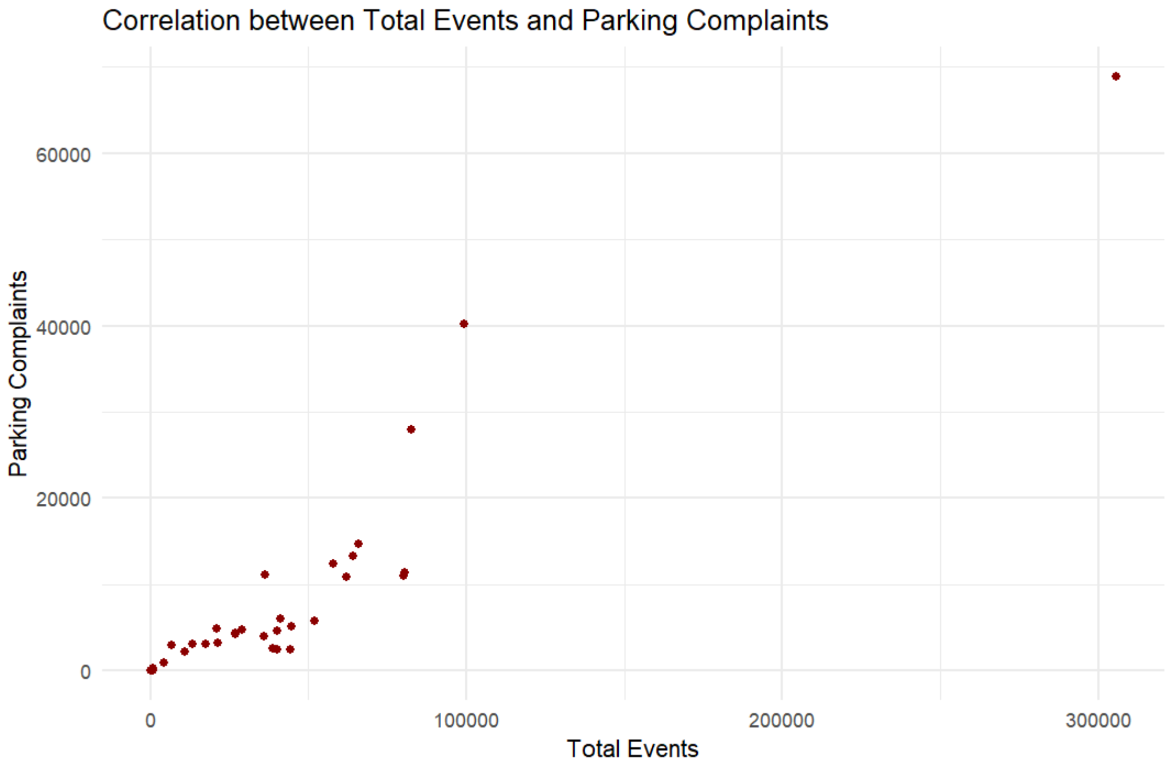
To expand the scope of analysis, an additional dataset containing licensing information was integrated with the 311 service requests dataset. This integration enables a comparative analysis of business licensing activity and service request patterns, particularly focusing on parking enforcement complaints.

The licensing dataset was first cleaned and standardized to ensure consistency in variable names. To assess overall business activity, the total number of licenses issued across zip codes was analyzed, providing insights into city-wide business operations. A zip code-level breakdown was generated to identify areas with high business density. Additionally, parking enforcement complaints were examined by filtering and grouping reports based on zip codes, allowing the identification of locations with the highest concentration of parking violations. To further contextualize complaint patterns, event frequency per zip code was analyzed by computing the total number of service requests in each area. By merging this data with parking complaints, a direct comparison between event volume and parking-related concerns was conducted, highlighting correlations between high-activity areas and increased parking enforcement issues.

A comparative assessment was conducted to understand how business density, event volume, and parking violations relate.

- High-business-density zip codes report more parking enforcement complaints,
- Event-heavy areas with frequent service requests show increased parking-related reports, suggesting parking constraints in high-traffic zones.

A Correlation Between Events and Parking Complaints was found to be **0.937**, indicating a strong positive correlation. This suggests that areas with higher service request volumes tend to experience more parking-related complaints, likely due to increased traffic and parking constraints. These findings highlight the need for improved parking management strategies in high-activity zones. Plot was created for the correlation.



Conclusion and Recommendations:

Findings:

- *Seasonal Trends*: Service request patterns vary by season. Parking Enforcement dominates year-round, while unsholved sidewalks peak especially in the winter. Street cleaning requests rise in summer and fall, indicating necessary seasonal cleanups.
- Neighborhood Disparities: Median time to close service requests across neighborhoods. The South Boston Waterfront records the fastest response times, while Mattapan has the slowest with closure times more than triple those of the most efficient neighborhoods.
- Parking Complaints and Business Activity: A strong positive correlation (0.937) was found between total service requests and parking enforcement complaints, highlighting that high-business-density and event-heavy zip codes face greater parking constraints.

Recommendations:

- Increasing staff and resources during peak seasons for specific service types. For instance, deploy additional groups for unshoveled sidewalks in winter and street cleaning in summer and fall.
- Investigate the causes of delays in Mattapan and other slower neighborhoods. Implement strategies such as improved workflow systems, increased funding to reduce response times, or better workforce distribution.
- Set baseline performance standards for request solution times across neighborhoods as a metric to deliver more equitable services.
- Use real-time parking monitoring systems in busy commercial and event areas to manage congestion. Increase enforcement in high-violation zones, especially where parking issues are frequent. Implement zoning policies to ensure enough parking space as businesses grow.

Wrap-Up:

This study highlights the importance of seasonal and geographic variations in municipal service demands. By implementing data-driven policy changes, the City of Boston can enhance service efficiency, reduce disparities in response times, and improve overall resident satisfaction. Proactive resource allocation and targeted interventions will be crucial in ensuring a fair and effective 311 service request system.

References

City of Boston. (n.d.). 311 Service Requests. Retrieved from https://data.boston.gov/dataset/311-service-requests

City of Boston. (n.d.). Special Event License Applications. Retrieved from https://data.boston.gov/dataset/special-event-license-applications

Kaggle. (n.d.). Boston 311 Dataset. Retrieved from https://www.kaggle.com/datasets/sandy1112/boston-311

Boston.gov. (n.d.). Boston 311. Retrieved from https://www.boston.gov/departments/boston-311

City of Boston. (n.d.). Open Data Challenge. Retrieved from https://github.com/CityOfBoston/opendatachallenge