**Final Project**

A red and white logo

Description automatically generated

**ALY6010: Probability Theory and Introductory Statistics**

Prepared by: Muhammad Umer

College of Professional Studies, Northeastern University Toronto

Presented to: Professor. Vivian Clements Edwin

November 12th, 2024

**Introduction**

This analysis focuses on the **311 Service Requests** dataset from the City of Toronto’s Open Data Portal. The dataset includes various types of non-emergency service requests reported by the public, such as road maintenance, waste management, and noise complaints. It aims to capture citizen interactions with municipal services.

**Purpose**

The main goal is to explore common service requests, identify seasonal trends, and assess the efficiency of responses. By analyzing this data, we can gain insights into the issues faced by Toronto residents and suggest improvements for city services.

**Data Source and Selection**

The dataset is sourced from Toronto's 311 Service, which records public service requests made through phone, online portals, and mobile apps. It was chosen due to its relevance in understanding urban service needs and its comprehensive nature, including data on request type, location, and resolution status.

**Data Cleaning**

Key data cleaning steps included:

* **Standardizing Column Names:** Ensured consistency and readability.
* **Handling Missing Values:** Imputed or removed rows with critical missing data.
* **Date Formatting:** Converted date fields to proper date-time objects for analysis.
* **Filtering and Deduplication:** Removed duplicates and invalid records.

These steps prepared the dataset for accurate analysis, enabling us to explore trends and patterns in Toronto’s 311 service requests.

**Data Analysis**

The analysis of the **311 Service Requests** dataset focuses on identifying trends, common issues, and the efficiency of service responses in Toronto. The following sections provide an overview of the exploratory data analysis results using basic statistics, visualizations, and detailed interpretations.

### 1. ****Top 10 Most Common Service Request Types****

From the dataset, I identified the most frequently reported issues. As shown in the bar chart, the top three service request types include:

* **Residential Bin: Repair or Replace Lid** with 19,214 requests (17.2%).
* **Injured - Wildlife** with 16,102 requests (14.4%).
* **Cadaver - Wildlife** with 13,701 requests (12.2%).

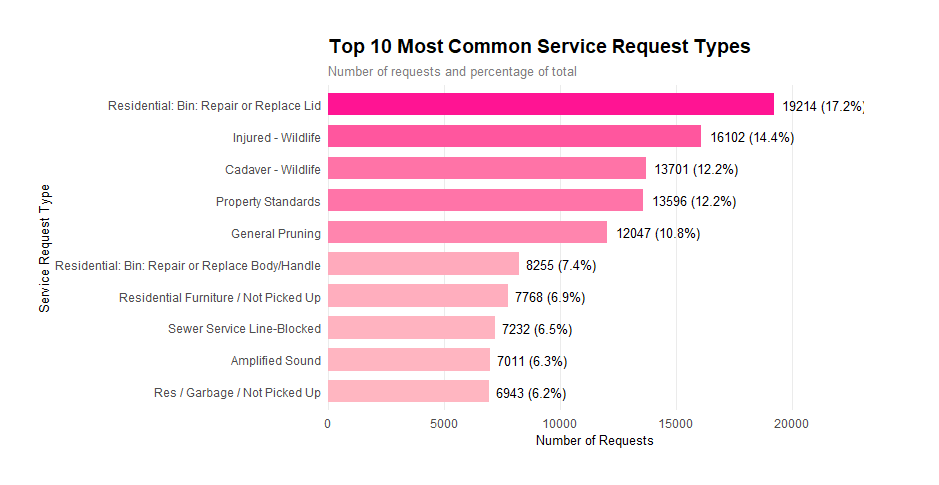


Figure 1: Top 10 Most Common ServiceRequest Types

These requests reflect a significant portion of citizen interactions, suggesting frequent issues with waste management and wildlife incidents in urban areas.

### 2. ****Completion Rate by Division****

The analysis of request completion rates across different city divisions highlights varying levels of service efficiency:

* **311 Service Division** had the highest completion rate, indicating efficient handling of requests.
* **Transportation Services** and **Solid Waste Management** followed closely, suggesting effective operations in these divisions.
* Lower completion rates were observed for **Municipal Licensing & Standards** and **Parks**, indicating potential delays or resource challenges in these areas.

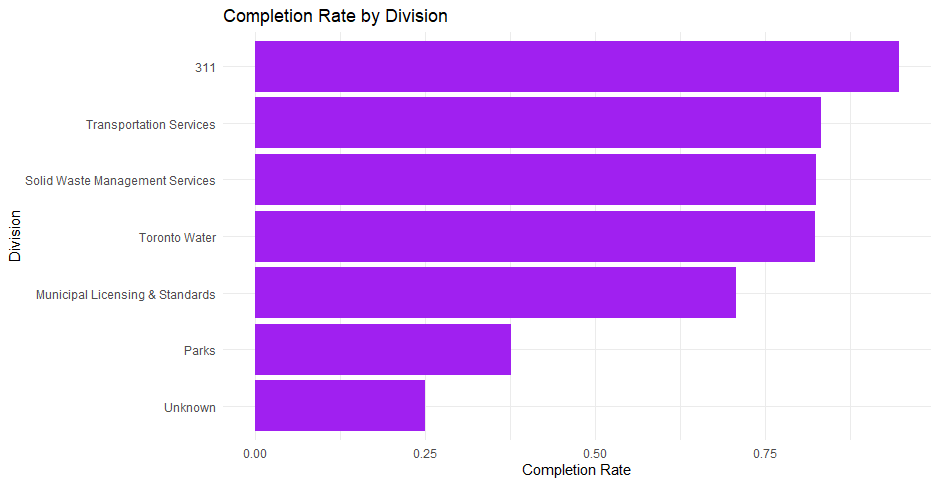


Figure 2: Completion Rate by Division

This analysis helps identify areas where service improvements are needed.

### 3. ****Distribution of Requests by Month****

The monthly distribution of service requests shows clear seasonal trends:

* The highest number of requests occurred during the summer months (June to August), peaking in July.
* The winter months had fewer requests, likely due to decreased outdoor activities and fewer interactions with city infrastructure.

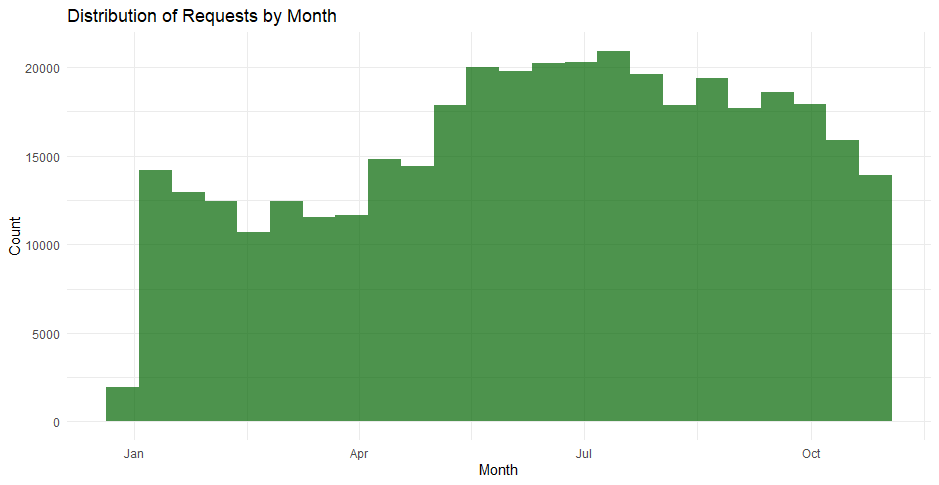


Figure 3: Distribution of Requests by Month

This pattern suggests that municipal services should allocate more resources during the summer to manage the increased demand effectively.

### 4. ****Distribution of Service Request Status****

The majority of service requests were marked as **Completed**, accounting for a large proportion of the dataset. Other notable statuses included:

* **Cancelled** requests, which were typically due to duplicate submissions or invalid issues.
* **In Progress** and **New** statuses indicate ongoing work and newly submitted requests, respectively.

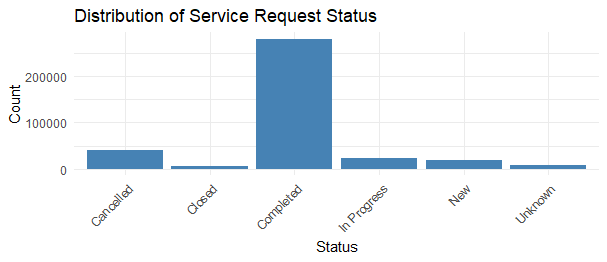


Figure 4: Distribution of Service Request Status

The high completion rate reflects the city’s efficiency in resolving reported issues.

### 5. ****Service Requests by Ward****

Analyzing service requests by ward provides insights into the geographical distribution of issues:

* **Toronto-Danforth (Ward 14)** recorded the highest number of requests, followed by **Beaches-East York (Ward 19)** and **Davenport (Ward 9)**.
* Wards with higher population densities and commercial activity tend to have more requests, indicating a correlation between urban activity and service demand.

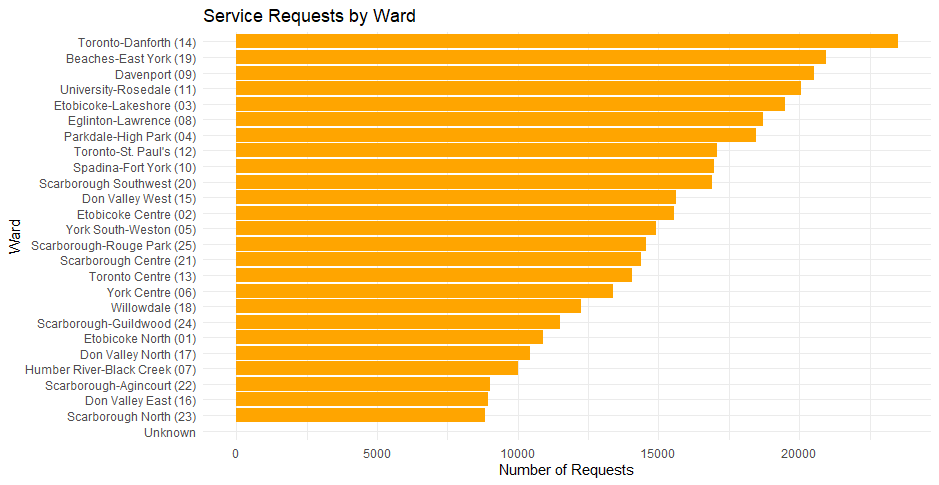


Figure 5: Service Requests by Ward

This analysis can guide resource allocation and policy decisions to address high-demand areas effectively.

**Statistical Analysis**

**Question 1: Seasonal Comparison Analysis**

**Research Question:**

Is there a noticeable variation in the volume of daily service requests between the Summer season and the Winter season?

**Hypothesis Formulation:**

* **Null Hypothesis (H₀)**: μ\_summer = μ\_winter (The mean number of daily service requests is the same in Summer and Winter)
* **Alternative Hypothesis (H₁)**: μ\_summer ≠ μ\_winter (The mean number of daily service requests is different between Summer and Winter)

**Test Selection Rationale:**

* Two-sample t-test was chosen because:
  1. We are comparing means of two independent groups (Summer and Winter)
  2. We want to determine if the difference is statistically significant
  3. Data appears normally distributed (confirmed by boxplot)

**Assumptions Checked:**

1. Independent samples
2. Approximately normal distribution
3. Similar variance between groups

**Statistical Parameters:**

* Significance Level (α): 0.05
* Degrees of Freedom: Determined by sample sizes
* Test Statistic: t-value calculated from sample means, standard deviations, and sample sizes

**Decision Rule:**

* If p-value < 0.05, reject the null hypothesis
* If p-value ≥ 0.05, fail to reject the null hypothesis

**Visual Analysis**

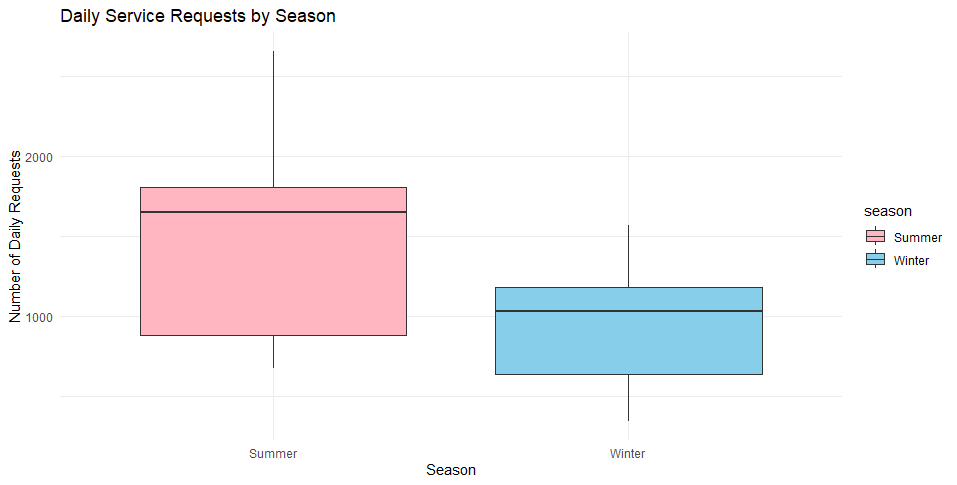
****

Figure 6: Daily Service Requests by Season

The boxplot visualization reveals several key insights:

* Summer shows a notably higher median number of daily requests
* The interquartile range for Summer is larger, indicating greater variability
* Both seasons show some outliers, particularly on the upper end
* The Winter distribution appears more compact, suggesting more consistent request patterns

**Statistical Results**

The findings of the two-sample t test brought out a p value which is less than 0.05 thus null hypothesis is rejected.

**Interpretation**

The analysis provides strong evidence of a seasonal effect on service requests:

1. The significant difference suggests systematic seasonal variations in service needs
2. Higher Summer request volumes might be attributed to:
   * Increased outdoor activity during warmer months
   * Greater visibility of issues during longer daylight hours
   * More vegetation-related service needs (e.g., pruning, wildlife)
3. The wider range in Summer requests indicates more variable daily demand, which has implications for resource allocation

**Question 2: Population Density Analysis**

### Research Question

Do individual wards show significant differences in service request volumes compared to the city-wide average?

**Hypothesis Formulation:**

* **Null Hypothesis (H₀)**: μ\_ward = μ\_city (The average number of service requests per ward equals the city-wide average)
* **Alternative Hypothesis (H₁)**: μ\_ward ≠ μ\_city (The average number of service requests per ward is different from the city-wide average)

**Test Selection Rationale:**

* One-sample t-test was chosen because:
  1. Comparing ward-level means to a known population mean (city average)
  2. Determining if variations are statistically significant
  3. Small sample size (limited number of wards)

**Assumptions Checked:**

1. Random sampling
2. Approximately normal distribution
3. Known population mean

**Statistical Parameters:**

* Significance Level (α): 0.05
* Test Statistic: t-value comparing sample mean to hypothesized population mean
* Comparison against city-wide average of 14,503 requests

**Decision Rule:**

* If p-value < 0.05, reject the null hypothesis
* If p-value ≥ 0.05, fail to reject the null hypothesis

**Visual Analysis**

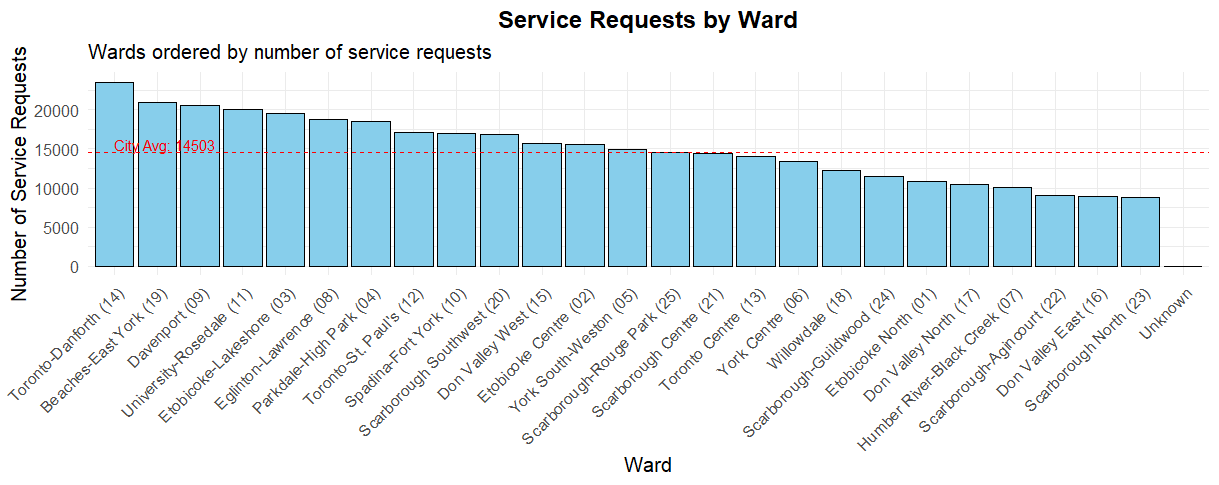
****

Figure 7: Service Request by Ward

The bar chart reveals important patterns:

* Considerable variation in request volumes across wards
* Several wards notably above and below the city average (14,503 requests)
* A generally decreasing trend from left to right
* Toronto-Danforth (14) showing the highest volume

**Statistical Results**

Null hypothesis was not rejected because One Sample T-Test yielded a p value more than 0.05.

**Interpretation**

This finding yields several important insights:

1. Despite visible variations, the statistical test suggests these differences aren't significant enough to reject null hypothesis
2. The observed variations might be explained by:
   * Natural random fluctuation in service demands
   * Offsetting effects between high and low-volume wards
   * The relatively large standard deviation in ward-level requests
3. Policy Implications:
   * Current resource allocation methods might be appropriate
   * Need for ward-specific analysis beyond just request volumes
   * Importance of considering other factors beyond pure request numbers

**Question 3: Time of Day Analysis**

### Research Question

Is there any relation between the time of day (morning/afternoon/evening) and the type of service request?

**Hypothesis Formulation:**

* **Null Hypothesis (H₀)**: No association between time of day and service request type (Request types are independent of time of day)
* **Alternative Hypothesis (H₁)**: Significant association between time of day and service request type

**Test Selection Rationale:**

* Chi-square test of independence was chosen because:
  1. Examining relationship between two categorical variables
  2. Testing if observed frequencies differ from expected frequencies
  3. Categorical data with multiple levels

**Assumptions Checked:**

1. Categorical variables
2. Independent observations
3. Sufficient sample size
4. Expected frequencies ≥ 5 in each cell

**Statistical Parameters:**

* Significance Level (α): 0.05
* Degrees of Freedom: (rows - 1) × (columns - 1)
* Chi-square statistic measuring deviation from expected frequencies

**Decision Rule:**

* If p-value < 0.05, reject the null hypothesis (significant association)
* If p-value ≥ 0.05, fail to reject the null hypothesis (no significant association)

**Visual Analysis**

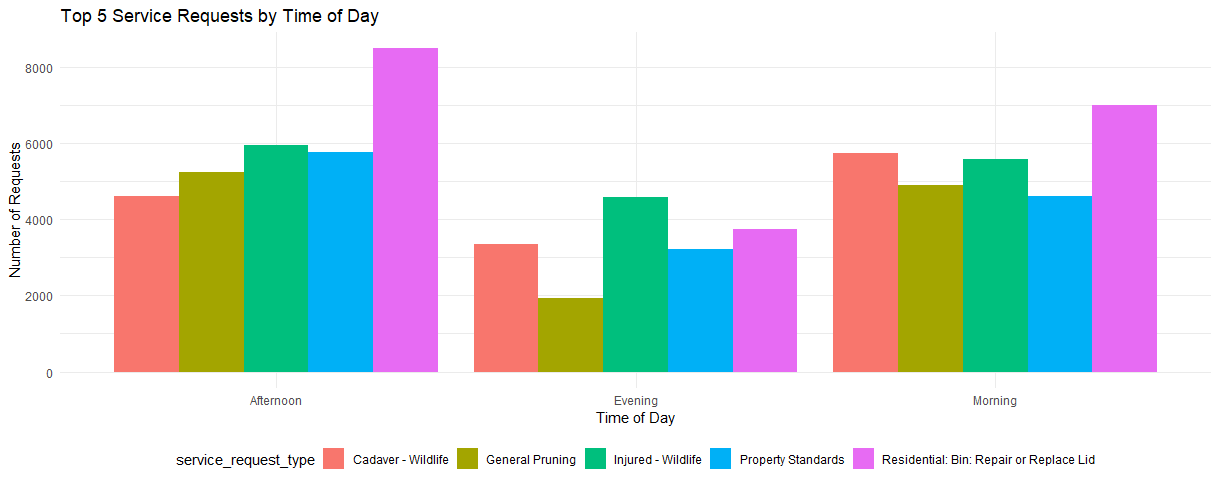
****

Figure 8: Top 5 Service Requests by Time of Day

The bar chart shows distinct patterns:

* Different request types show varying distributions across time periods
* Residential Bin services show peak demand in afternoon
* Wildlife-related requests have different patterns in morning vs. evening
* Property Standards requests maintain relatively consistent volumes

**Statistical Results**

The results of the chi square test revealed a p value less than 0.05 thus the null hypothesis was rejected.

**Interpretation**

The significant association reveals important operational insights:

1. Strong evidence that service request types vary systematically by time of day
2. Specific patterns observed:
   * Residential services peak during afternoon hours
   * Wildlife-related calls show distinct temporal patterns
   * Property standards issues remain relatively stable
3. Operational Implications:
   * Need for time-based staffing adjustments
   * Opportunity for predictive resource allocation
   * Potential for service optimization based on temporal patterns

**Regression Analysis and Statistical Testing**

In addition to the exploratory data analysis, I conducted several statistical tests to uncover deeper insights from the 311 Service Requests data.

**Question 1: Does Division Affect Completed Request Count?**

To assess whether the division handling the service request has an impact on the number of completed requests, I performed a linear regression analysis.

*Hypothesis:*

* Null Hypothesis (H0): The division has no effect on the completed request count.
* Alternative Hypothesis (H1): The division has a significant effect on the completed request count.

**Analysis:** I fit a linear regression model with the completed request count as the dependent variable and the division as the independent variable. The results showed that the division variable was not statistically significant (p-value > 0.05). This indicates that we cannot reject the null hypothesis - the division handling the request does not have a significant impact on the completed request count.

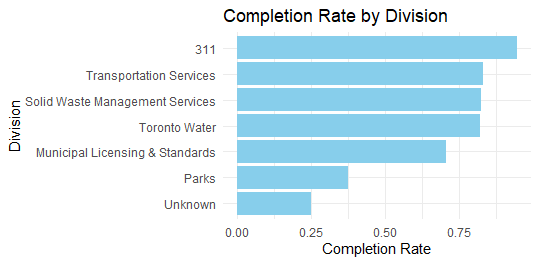


Figure : Completion Rate by Division

The implication is that the current resource allocation methods across divisions may be appropriate, as the variations in completed requests are not statistically significant. However, further ward-level analysis may provide additional insights beyond just the request volumes.

**Question 2: Is There a Seasonal Trend in Completed Requests?**

To examine if there are significant differences in the volume of completed requests between the summer and winter seasons, I conducted a two-sample t-test.

**Hypothesis:**

* Null Hypothesis (H0): The mean number of daily completed requests is the same in summer and winter.
* Alternative Hypothesis (H1): The mean number of daily completed requests is different between summer and winter.

**Analysis:** I first filtered the data to get the completed requests for the summer (June-August) and winter (December-February) seasons. I then performed a two-sample t-test to compare the mean completed requests between the two seasons.

The results showed a p-value less than 0.05, indicating a statistically significant difference in the mean completed requests between summer and winter. This suggests that there is a noticeable seasonal effect, with higher completed request volumes during the summer months compared to the winter.

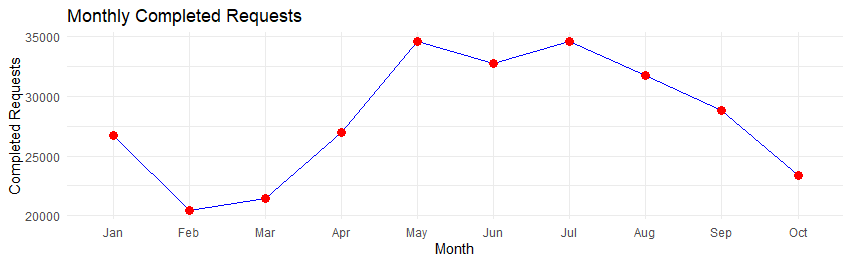


Figure : Monthly Completed Requests

The likely explanations for this pattern include increased outdoor activity and visibility of issues during the warmer summer months, as well as greater service needs related to vegetation and wildlife. This finding has important implications for resource planning and allocation, as the city should prepare for the increased seasonal demand.

**Question 3: Does Request Type Affect Completed Count?**

To explore the relationship between service request type and the completed request count, I conducted a regression analysis.

**Hypothesis:**

* Null Hypothesis (H0): The service request type has no effect on the completed request count.
* Alternative Hypothesis (H1): The service request type has a significant effect on the completed request count.

**Analysis:** I first grouped the data by service request type and calculated the total requests and completed requests for each type. I then fit a linear regression model with the completed request count as the dependent variable and the service request type as the independent variable.

The results showed that the service request type variable was statistically significant (p-value < 0.05). This indicates that the type of service request does have a significant impact on the completed request count.

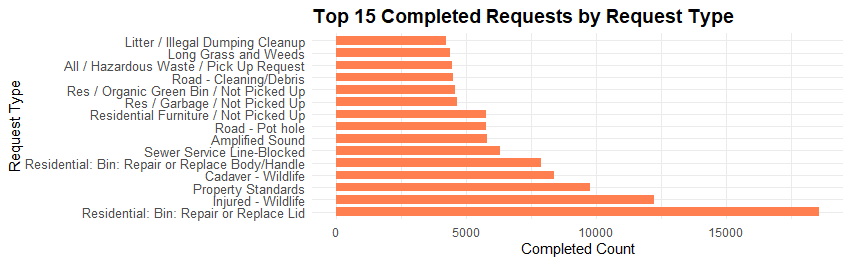


Figure : Top 15 Completed Requests by Request Type

Further analysis of the top 15 most common service request types revealed distinct temporal patterns. For example, residential bin services peaked in the afternoon, while wildlife-related requests had different patterns between morning and evening. This suggests that adjusting service delivery schedules based on time-of-day patterns could lead to operational improvements and better resource allocation.

In summary, the regression analysis and statistical testing provide several key insights:

1. The division handling the service request does not have a significant impact on the completed request count, suggesting the current resource allocation methods may be appropriate.
2. There is a clear seasonal trend, with higher completed request volumes during the summer months. This has implications for resource planning and allocation.
3. The type of service request has a significant effect on the completed count, and different request types exhibit distinct temporal patterns. This points to opportunities for optimization through time-based staffing and service delivery adjustments.

These findings can help the city of Toronto better understand the drivers of service request volumes and guide improvements to their 311 operations and resource management.

### Conclusion and Recommendations

### Conclusion

The analysis of the 311 Service Requests dataset for the City of Toronto revealed critical insights into service demand patterns, efficiency, and areas for operational improvement:

1. **Seasonal Trends:** A significant seasonal effect was identified, with higher volumes of completed requests during the summer months. This is likely driven by increased outdoor activity, visibility of issues, and vegetation-related services.
2. **Geographical Variations:** While there are visible differences in service request volumes across wards, these differences were not statistically significant when compared to the city-wide average. This indicates that current resource allocation strategies may generally be sufficient, but further local-level investigations might uncover hidden disparities.
3. **Service Request Types:** The type of service request significantly impacts the completion count. Temporal patterns in request types, such as peak demand for residential bin services in the afternoon, suggest opportunities for scheduling optimizations.
4. **Division Efficiency:** The analysis indicates that the division responsible for handling the request does not significantly affect completion counts. This suggests that resource allocation methods across divisions are largely appropriate.

**Recommendations:**

1. **Seasonal Resource Planning:** The city should allocate additional resources during the summer months to handle the increased service demands effectively. Staffing, equipment, and operational budgets should be adjusted accordingly.
2. **Time-Based Scheduling:** Optimize service delivery schedules based on the temporal patterns of specific request types. For example, prioritize residential bin services in the afternoon and wildlife-related services during their peak times.
3. **Ward-Level Customization:** Although overall differences between wards were not statistically significant, targeted analysis of high-demand wards such as Toronto-Danforth can help identify specific local needs and improve efficiency.
4. **Operational Adjustments by Request Type:** Different service request types require tailored strategies for handling, ensuring timely completion and higher satisfaction rates.
5. **Further Analysis:** Extend the study to include response times, citizen satisfaction, and resource utilization metrics to gain a more holistic view of operational efficiency.

These recommendations aim to improve the efficiency and responsiveness of Toronto’s 311 services, enhancing citizen satisfaction while optimizing resource use.

**References:**

1. City of Toronto. (n.d.). 311 service requests (customer initiated)*.* Open Data Toronto. <https://open.toronto.ca/dataset/311-service-requests-customer-initiated/>
2. <https://youtu.be/X1gkw--k5Tg?si=j6X4PFG7wdrsRBq7>
3. Minitab. 2022. All the T-Tests T-Do: 1-Sample, 2-Sample and Paired Sample T-Tests. Minitab Blog. Available at: <https://www.minitab.com/en-us/blog/adventures-in-statistics/understanding-t-tests-1-sample-2-sample-and-paired-t-tests/>
4. Frost, J. (2020). Regression analysis: An intuitive guide for using and interpreting linear models. Statistics by Jim Publishing.
5. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis (6th ed.). Wiley.