Campaign Zero: 911 Call Data Patterns

In collaboration with BU Spark!

By:

Dang Dinh (MSDS '26)
dddinh@bu.edu

Damayanti Gupta (MSDS '26)
dami123@bu.edu

Jaiveer Raikhy (MSDS '26)
jaiveerr@bu.edu

Johanna Ludviksdottir (MSDS '26)
jolea@bu.edu

Peng Wang (MSDS '26)
pengwang@bu.edu

San Yan (MSDS '26)
sanyan@bu.edu

For partial fulfillment of course requirements for
DS 701 Tools for Data Science

Table of Contents

1. Introduction

Police-related 911 call data provides critical insights into public safety needs, community concerns, and law enforcement practices. However, this data is currently fragmented across individual city portals, each with its own formats, variables, and documentation standards. As a result, no comprehensive, multi-city dataset exists, which limits the ability to compare trends or assess broader public safety dynamics.

This project was conducted in collaboration with Campaign Zero, a public safety focused non-profit organization working to end police violence through data transparency and policy change. Campaign Zero's work includes aggregating policing data, advocating for alternatives to traditional law enforcement responses, and developing publicly accessible platforms like Mapping Police Violence (MPV). The project was guided by input from Andrerw Zaharia, the director of data science at Campaign Zero, to ensure alignment with the organization's broader mission.

Building on this objective, this report begins with exploratory data analyses of police-related calls from three initial cities to examine how call patterns vary across different urban contexts and social conditions. These cities, New York City, Detroit, and Seattle, represent the first phase of the project, with the ultimate goal of scaling the data pipeline to collect, standardize, and analyze 911 call data from the majority of major cities in the U.S.

To further support this broader objective, we developed a standardized data pipeline that collects, synchronizes, and analyzes 911 call data across the three cities, enabling meaningful cross-city comparisons and deeper insights into public safety needs and outcomes. The analysis focuses on identifying temporal, geographic, and categorical trends in calls, including where and when calls occur, and what types of incidents drive call volume.

2. Dataset Description

The datasets we used in this project were publicly available and provided detailed records of police-related 911 calls from three cities: New York City, Detroit, and Seattle.

New York City

For New York, we derived the dataset from the NYPD's Internal Computer Aided Dispatch (ICAD) system, which is used by call takers and dispatchers to manage communication between the public and NYPD personnel. Each record represents an entry into the ICAD system, including both public-initiated and officer-initiated incidents. The data provides a comprehensive view of the range of issues to which the NYPD responds, from citizen-reported emergencies to self-initiated police activity. Key variables include call type, call priority, dispatch times, arrival times, and location information. The dataset encompasses over **50 million** records spanning from **January 2018 through October 2025**, representing one of the most comprehensive municipal 911 call datasets publicly available.

Original Data Files:

- Historical data (2018-2024): ~47 million records / 15 GB size in CSV format
- Current data (2025): ~5.3 million records / 1.5 GB size in CSV format
- Source: NYC Open Data Portal
    - The dataset has 18 columns in total, and the data dictionary is also included in the portal.

Detroit

The Detroit dataset includes all 911 police emergency response calls for service since **September 2016, to the present**. These calls originate from public reports, the non-emergency Detroit Police Department's (DPD) Telephone Crime Reporting (TCR) line, and ShotSpotter gunshot detection alerts. The dataset contains approximately **2.6 million records**, where each record represents a single call for service. Each record includes key details such as the time received, nature of the call, assigned priority, response precinct, and response times. To protect individual privacy, incident locations are reported by their nearest intersection.

- Data (2016 - 2025) ~2.6 million records
- Source: City of Detroit Open Data Portal

Seattle

The Seattle dataset captures police response activity logged by the Seattle Police Department (SPD) Communications Center. Each record represents a single call for service, including calls from the public as well as calls logged by officers based on field observations. The dataset spans from **April 2018 to the present** and contains approximately **10.6 million records**. To protect privacy, call locations are reported at the "beat" level, which is the most granular unit used for patrol deployment. This dataset is updated on a daily basis, with full refreshes twice a year to reconcile changes.

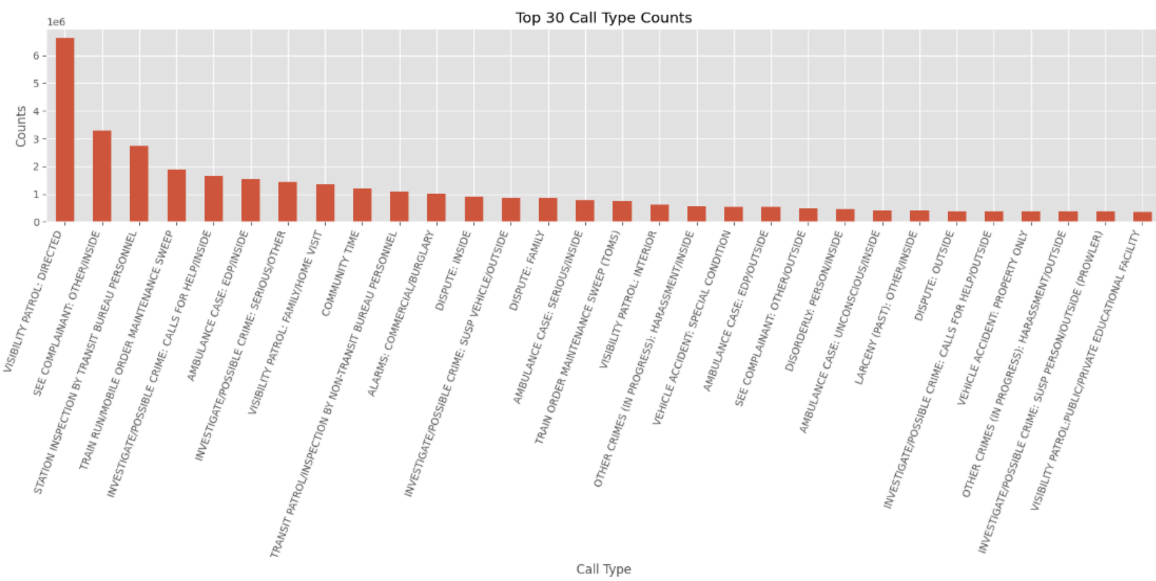
- Data (2018-2025) ~10.6 million records
- Source: Seattle Open Data Portal

3. Exploratory Data Analysis

3.1 Call Type Distribution

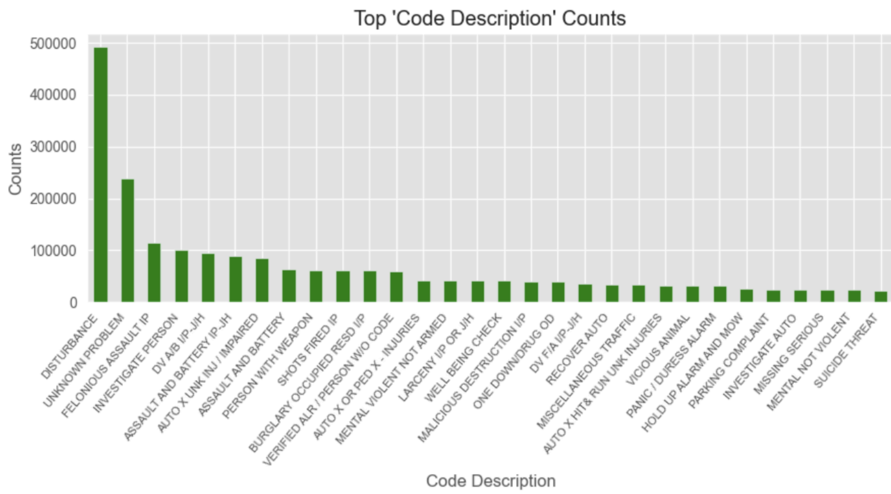
3.1.1 New York City

The top call types for New York mainly involve routine oversight and monitoring tasks. “Visibility Patrol: Directed” is by far the most frequent type of calls, followed by “See Complaint: Other/Inside” and other transit-related activities such as station inspections and train-run checks. These patterns reflect the city’s busy urban environment and large transit system, where officers spend much of their time patrolling public spaces and responding to non-emergency complaints. Overall, the call types show that everyday patrol and monitoring make up a major part of New York’s 911 activity.



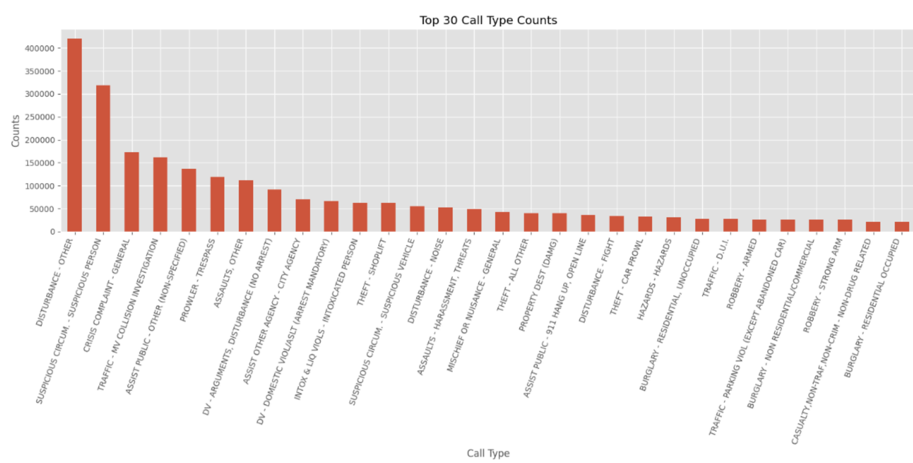
3.1.2 Detroit

Detroit’s leading call types prominently feature disturbances and some categories related to violent crime. While “Disturbance” is by far the most frequent call type and “Unknown Problem” is also prevalent, serious incidents such as “Felonious Assault IP” appear near the top of the distribution. This mix shows that although violent crime is a significant component of Detroit’s emergency calls, broad disturbance calls still make up the majority share overall.



3.1.3 Seattle

Seattle's top call types are dominated by nonviolent, community-level concerns, with categories like "Disturbance - Other" and "Suspicious Circumstance/Suspicious Person" appearing more frequently than serious crime reports. This suggests that much of Seattle's police activity centers on addressing public disorder, welfare checks, and general community complaints rather than violent incidents.

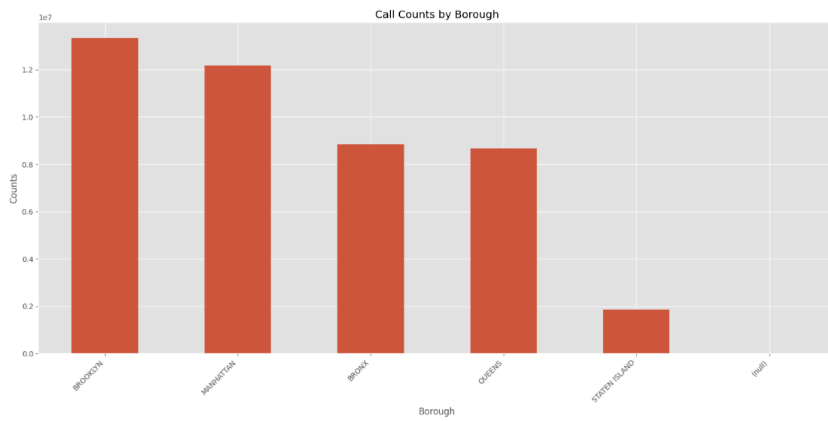


Across the three cities, call type distributions reveal how local conditions shape distinct policing demands. New York's calls are dominated by oversight and transportation system monitoring, reflecting the city's dense urban environment and extensive public infrastructure. Detroit shows a pattern that is led by disturbances and unknown problem calls, but with several violent crime categories appearing more prominently than in the other cities. Seattle shows yet another pattern, with calls mainly focused on nonviolent community concerns, disturbances, and welfare checks. These different patterns highlight how each city's unique social and spatial dynamics influence the nature of its emergency response workload.

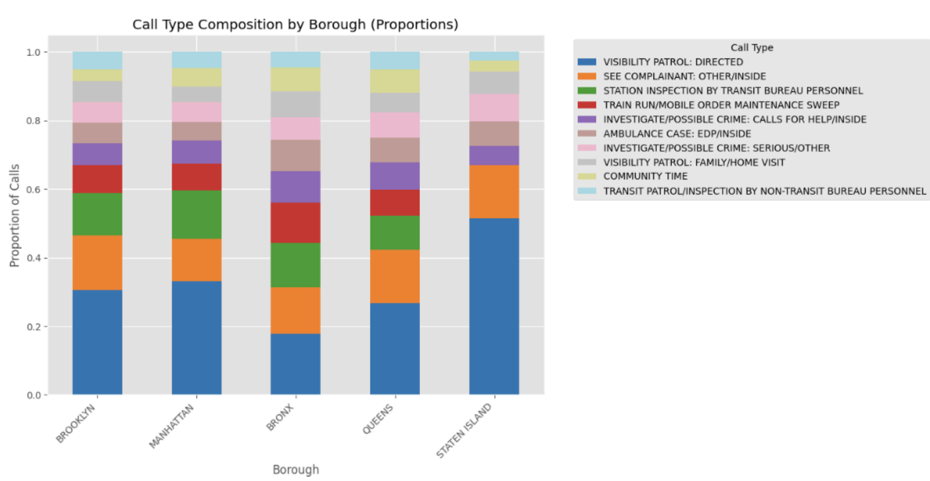
3.2 Distribution of Call Counts by Boroughs and Neighborhoods

3.2.1 New York City

In New York, the call counts, aggregated by borough, show a pronounced concentration in Brooklyn and Manhattan, with the Bronx and Queens following at lower but still substantial levels. Staten Island has significantly fewer calls than any other borough. This can be attributed to the demographic and population density differences across boroughs, as well as potential variations in service demand and urban activity.

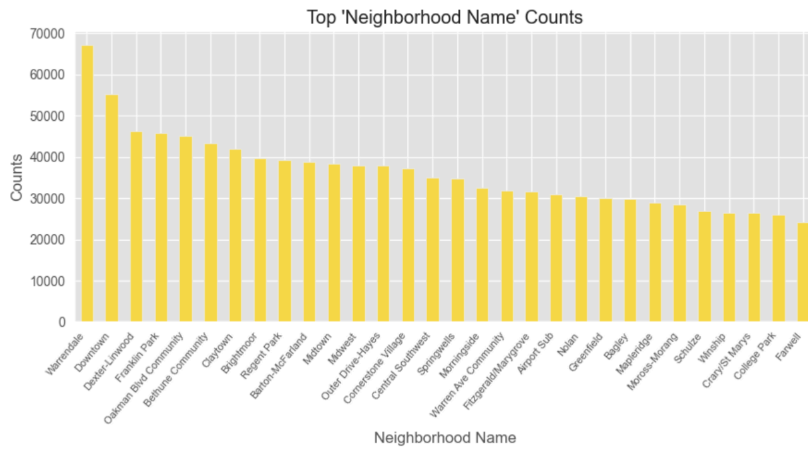


When we analyzed the call type composition by borough, Staten Island has a much larger share of directed patrol calls than any other borough. In contrast, Brooklyn, Manhattan, the Bronx, and Queens have a more balanced mix of call types, with many calls involving complaints from residents and transit-related inspections. More specialized or investigative calls make up smaller but fairly steady proportions across all boroughs. Overall, these patterns show that while each borough has different needs, the general structure of police work across the city is still quite similar.



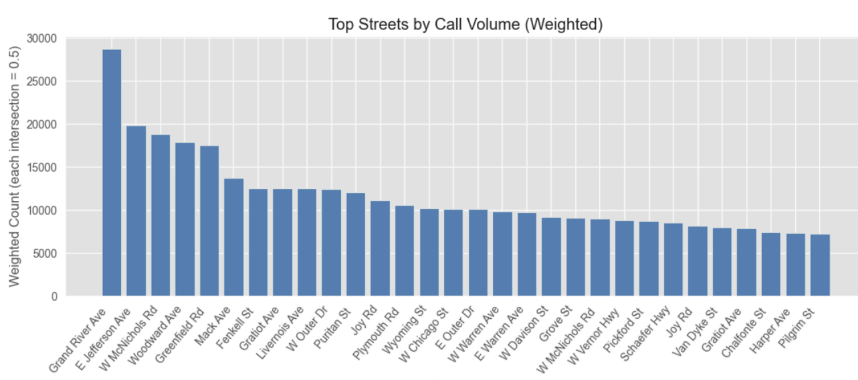
3.2.2 Detroit

In Detroit, call volumes are spread across multiple neighborhoods with the highest concentrations in Warrendale, Downtown, and Dexter-Linwood, indicating localized hotspots of service demand. There is a gradual decline in call counts per neighborhood, which suggests a relatively even spread of calls citywide rather than calls being heavily focused in only a few areas.

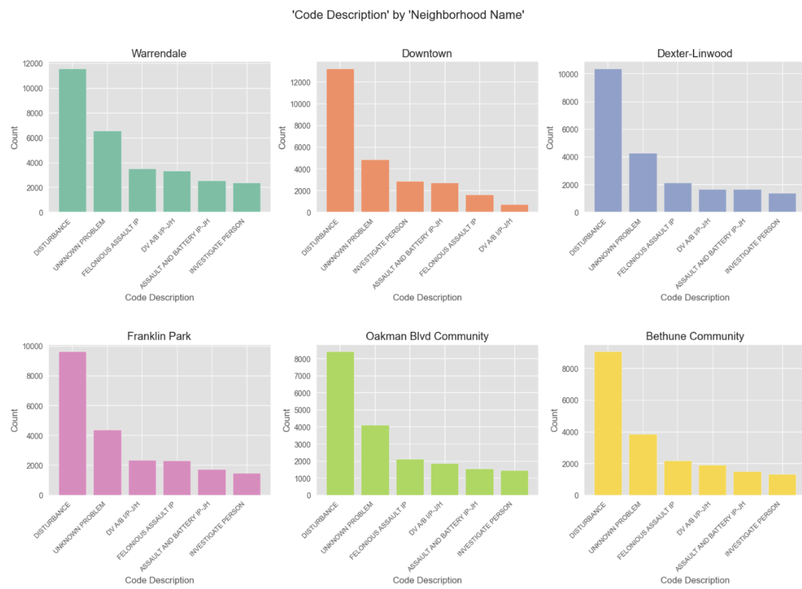


Call demand clusters more sharply in New York's boroughs than Detroit's neighborhoods because of the differences in city size, population distribution, and urban characteristics.

Looking closer at Detroit, the street-level data shows which specific roads have the most calls. Grand River Avenue has the highest call volume by far, with East Jefferson Avenue, Woodward Avenue, and Greenfield Road also having a lot of calls. We do not have the neighborhoods for each street here, but knowing these top streets helps add more detail to what we saw about calls in different parts of the city.

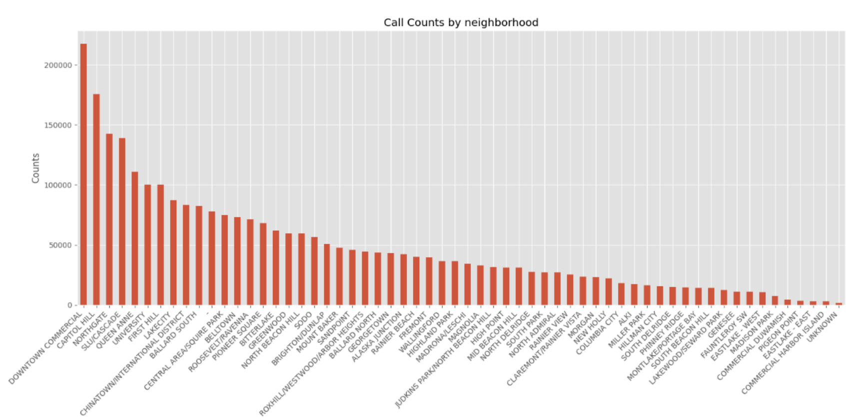


Across the following six neighborhoods in Detroit, "Disturbance" is the most common call type by far. "Unknown Problem" is usually the second most frequent, though the number of calls varies by area. Assault related calls, domestic violence incidents, and investigative responses also appear consistently across neighborhoods, indicating shared patterns of community safety concerns. While the most frequent categories remain similar, the differences in call volumes across neighborhoods highlight localized variations in service demand and public safety dynamics.

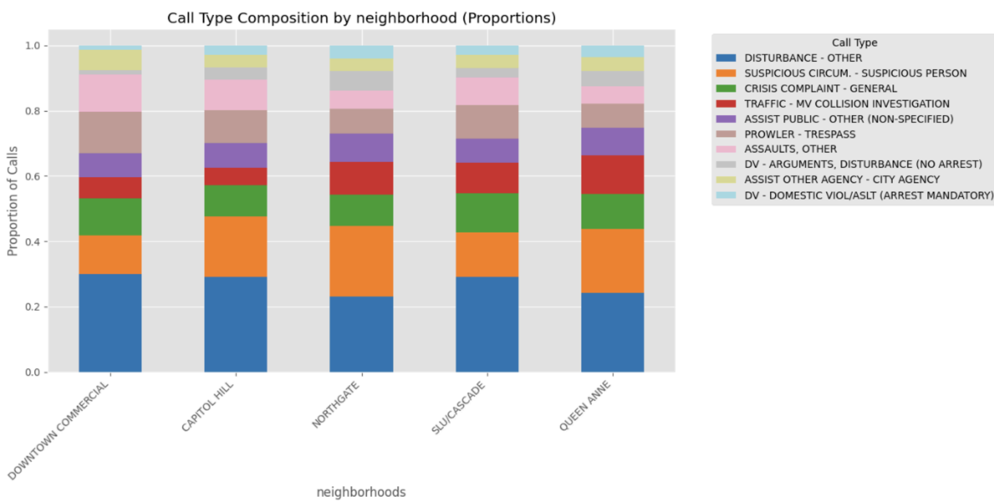


3.2.3 Seattle

In Seattle, call volumes vary significantly across neighborhoods, with Downtown Commercial, Capitol Hill, and Northgate receiving the highest number of calls. These areas represent major commercial, nightlife, and transit hubs, which help explain their elevated service demand. Beyond the top neighborhoods, call counts decline more gradually, indicating that activity is distributed across a wide range of residential and mixed-use areas rather than concentrated in only a few locations.



Across Seattle neighborhoods, call type composition is consistent overall, with “Disturbance - Other” and “Suspicious Circumstance/Suspicious Person” making up the majority of calls in every neighborhood. Downtown Commercial and Capitol Hill show slightly higher proportions of disturbance-related complaints, which reflects their dense activity and nightlife. Calls from Northgate and Queen Anne are more evenly distributed, including significant amounts of crisis complaints and traffic investigations. Even though the specific proportions vary slightly, the overall pattern shows that Seattle’s emergency service demand is shaped primarily by nonviolent community concerns and routine public assistance needs across neighborhoods.

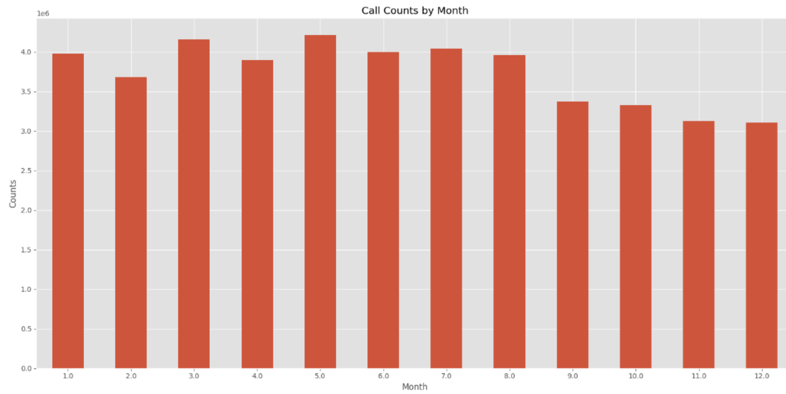


Across all three cities, call volumes accumulate in major commercial or high-activity areas, while more residential neighborhoods show lower counts. New York displays clear borough-level differences, Detroit shows a broadly distributed pattern with only modest peaks in a few neighborhoods, and Seattle’s highest volumes appear in central urban districts. Overall, the geographic distribution of calls reflects how population density and neighborhood function shape service demand.

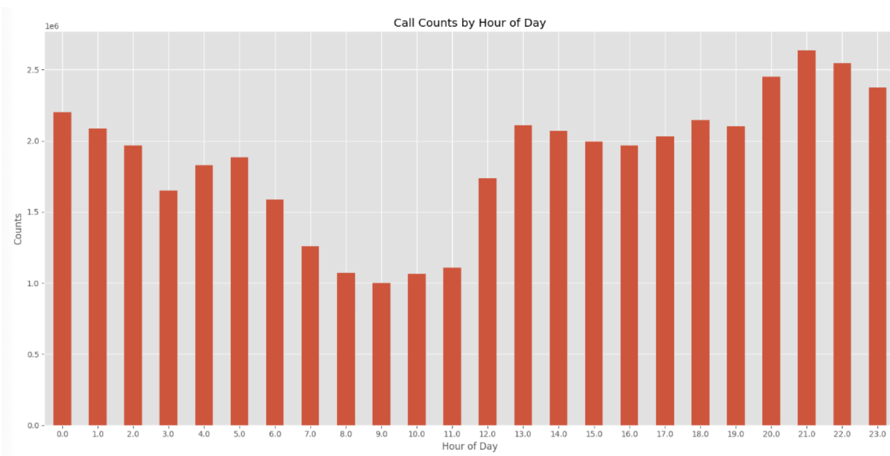
3.3 Temporal Distributions of Calls

3.3.1 New York City

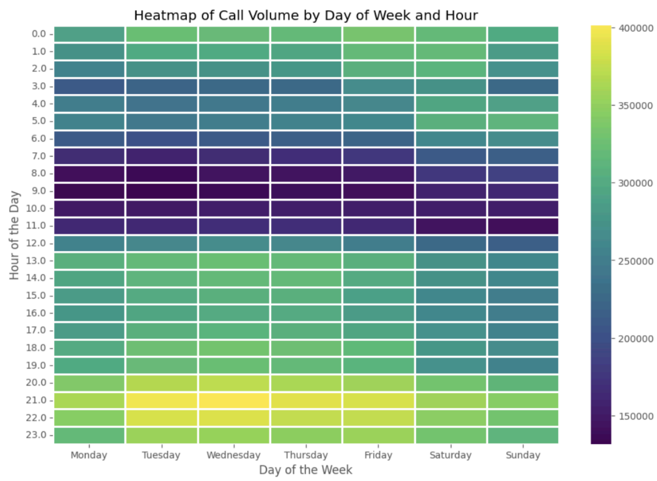
The monthly distribution of calls in New York shows moderate seasonal variation, with call volumes generally highest in the late spring and early summer. March, May, June, and July stand out with the greatest activity, while the fall and early winter months show a gradual decline. Although the overall pattern does not show dramatic seasonal swings, there is a consistent tendency for warmer months to generate more calls.



In New York, call activity follows a rather clear daily pattern. Calls are highest late at night and in the early morning hours, especially between 8 PM and 2 AM. Then call volume gradually declines, dropping to its lowest levels in the late morning. After this mid-day low point, call activity rises again through the afternoon and evening. This pattern reflects increased police demand during nighttime hours when public activity and potential conflicts are more common.

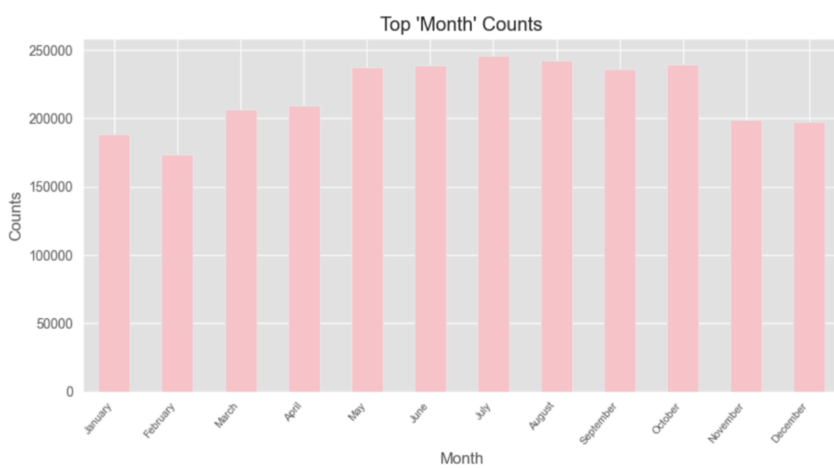


The following day-of-week and hourly heatmap for New York's police activity. Across all days of the week, calls are highest in the late evening and early morning hours, and lowest during mid-morning. Activity then gradually increases throughout the afternoon before rising again at night. Weekends follow the same pattern, with Friday and Saturday nights having slightly higher call volume than other days. Overall, the heatmap highlights the steady nighttime demand for police services in New York.

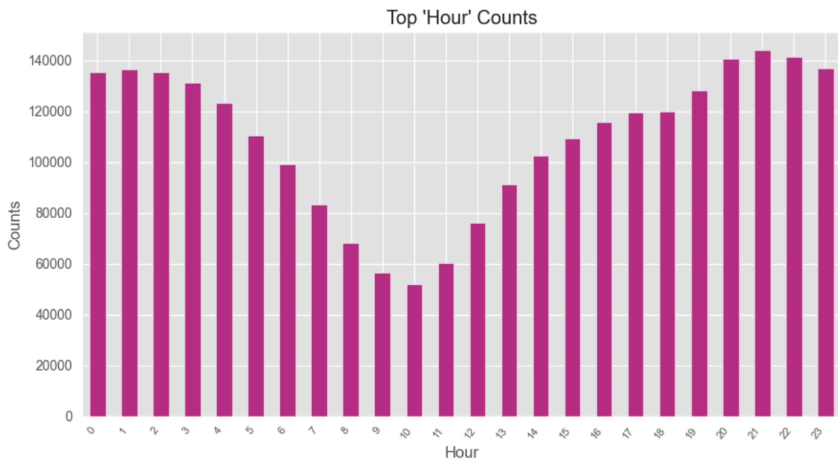


3.3.2 Detroit

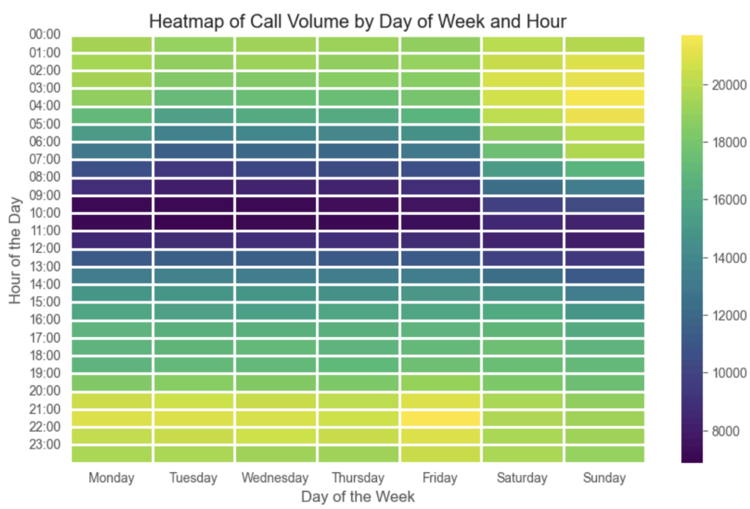
Detroit’s 911 call activity shows a modest seasonal variation, with higher call volumes during the warmer months and lower volumes in the colder months. July records the highest number of calls, while February has the lowest, consistent with a general increase in activity during summer when outdoor activity and public presence are at their peak.



Detroit’s hourly call distribution shows that police demand is highest during the late evening and early morning. Call volumes are noticeably higher after 6 PM, stay high during the night, and peak around 2 AM. The lowest call volume is between 5 AM and 9 AM. Overall, this pattern suggests that police activity in Detroit increases at night, when public activity and potential disturbances are more common.

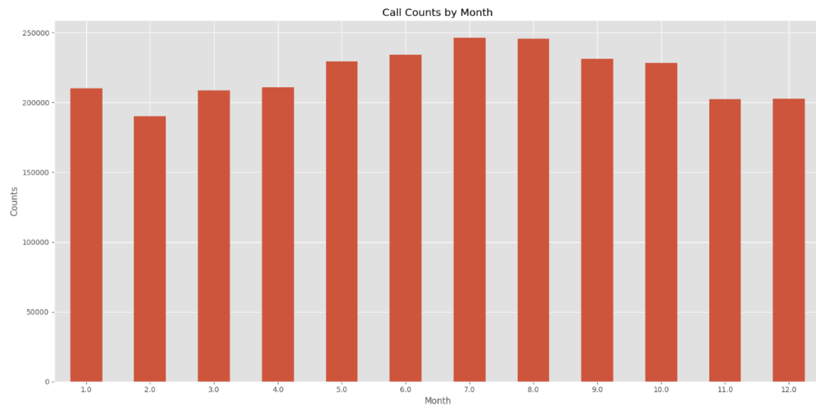


The following day-of-week and hourly heatmap for Detroit shows that call activity is consistently highest during the late evening and early morning hours across all days, especially between 8 PM and 2 AM. In contrast, call frequency decreases during the morning hours, reaching its lowest levels around 9 to 11 AM. Weekends, more precisely Friday and Saturday, show slightly higher late-night activity compared to weekdays. This suggests that increased social activity contributes to elevated demand for emergency services during these periods.

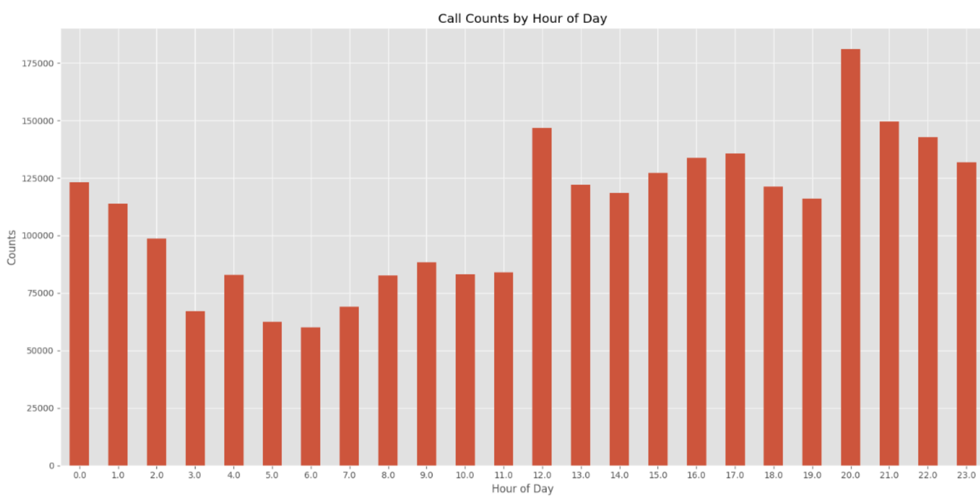


3.3.3 Seattle

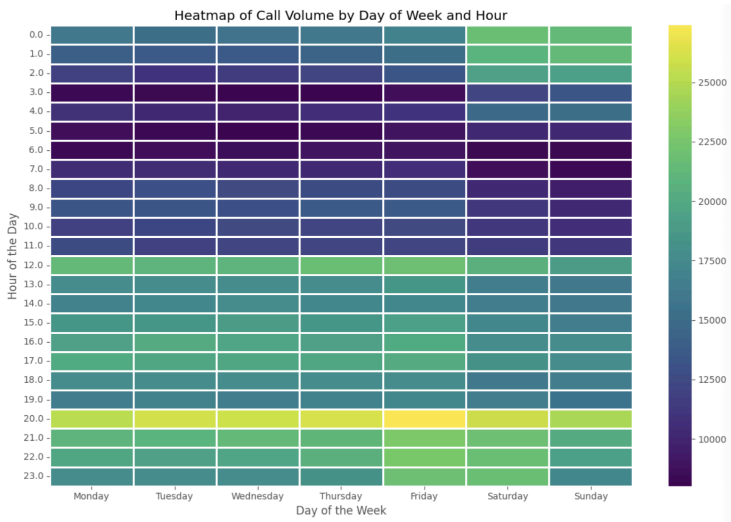
Seattle's monthly call volumes show a seasonal pattern, with the highest activity occurring during mid-summer, particularly in July and August. Call volume is noticeably lower in colder months, with February having noticeably fewer calls. This trend suggests that police service demand in Seattle increases during warmer months when outdoor activity and public presence are higher.



Seattle's hourly call patterns show a noticeable increase in activity during the late afternoon and evening hours. Call volumes begin to significantly rise after 11 AM, peaking at 8 PM, and then gradually decrease. In contrast, the lowest call activity occurs in the early morning hours, especially from 4 to 7 AM. This pattern suggests that emergency demand in Seattle is closely tied to daytime and evening activity levels, with fewer incidents occurring overnight.



Seattle's call volume shows clear differences throughout the day, with the lowest call volume occurring in the early morning hours before gradually increasing through midday and afternoon. The highest call activity appears in the evening, especially between 7 PM and 10 PM, and this pattern holds across the entire week. Friday and Saturday evenings show slightly higher peaks than other days, suggesting that increased social activity and nightlife contribute to the rise in calls during those times. Overall, the heatmap shows a consistent pattern where call demand builds throughout the day and reaches its peak in the evening.

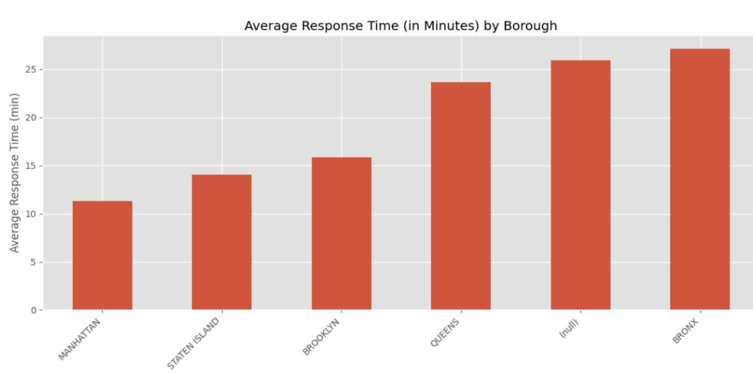


Across all three cities, the temporal patterns of 911 calls reveal consistent daily and seasonal fluctuations. Call volumes tend to be lower in the early morning hours, increase steadily through the day, and peak in the evening and late-night periods. Monthly patterns show modest seasonal variation, with warmer months generally associated with higher call volumes. Although each city displays its own specific patterns, they all demonstrate that the need for police services is closely tied to people's daily and weekly patterns of activity and seasonal shifts. These temporal trends offer important insight into when policing resources are likely to face the greatest demand.

3.4 Response Times

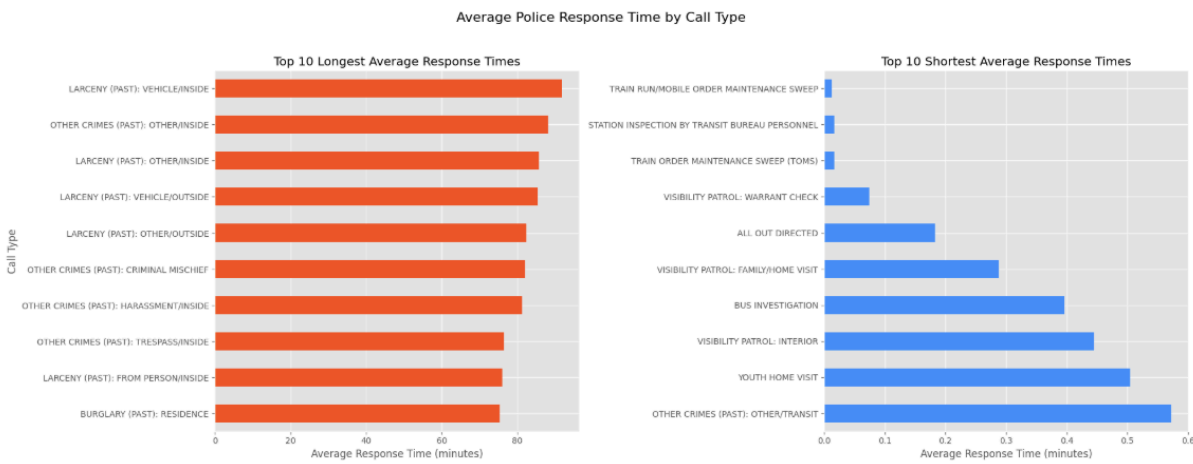
3.4.1 New York City

The average police response time in New York City differs noticeably across boroughs. Manhattan has the fastest response, at roughly 11-12 minutes, followed by Staten Island, about 14 minutes, and Brooklyn, about 16 minutes. Queens and the Bronx face much longer waits, with Queens averaging about 23-24 minutes and the Bronx showing the slowest response times at about 27 minutes. These differences likely reflect variations in population density, call volume, and travel constraints across boroughs.



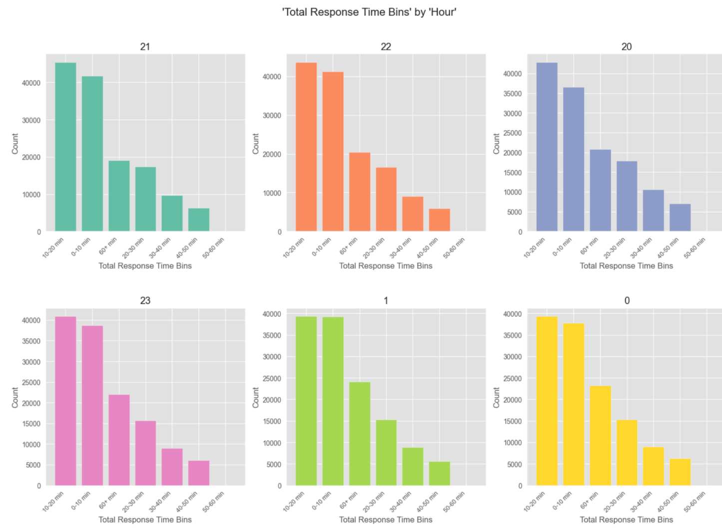
Response times in New York also vary widely depending on the type of call. The call types with the longest average response times, often exceeding 90 minutes, are mostly non-urgent, past-tense crimes such as larceny, trespass, harassment, and burglary. These incidents are of lower priority because they do not involve an active threat or immediate danger.

In contrast, the shortest response times, often well under 1 minute, occur for calls linked to transit maintenance, inspections, patrol activities, and certain administrative or proactive policing duties. These tasks are likely handled quickly because they are scheduled, routine, or already involve officers who are on-site or nearby.



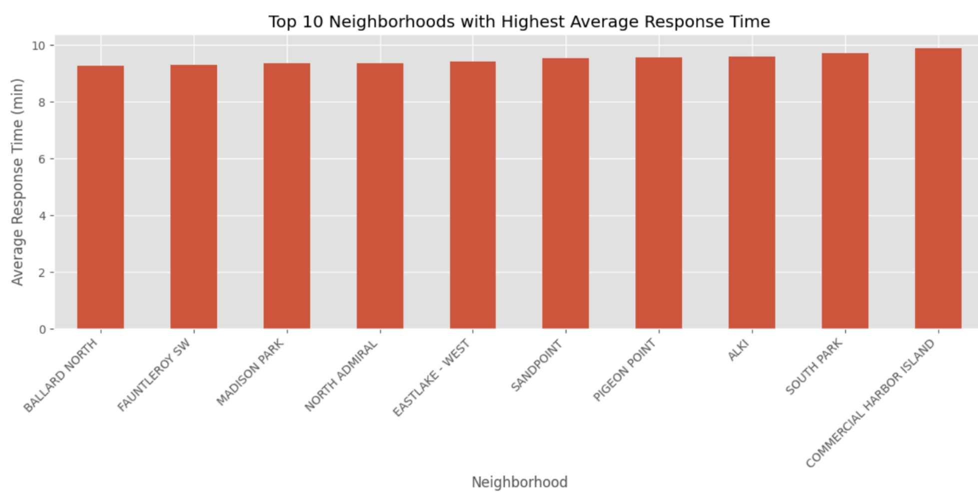
3.4.2 Detroit

The charts for Detroit analyze the distribution of 911 call response times during the peak late-night and early-morning hours (specifically 8 PM, 9 PM, 10 PM, 11 PM, 12 AM, and 1 AM). A strong and consistent pattern emerges across all of these high-volume hours: the most common response time is 0-10 minutes, which has the highest call count in every chart. The data also shows a clear trend where the number of calls steadily decreases as the response time gets longer, meaning response times of 20-30 minutes are less common than 10-20 minutes, and so on. This indicates that during these peak periods, the vast majority of 911 calls in Detroit are responded to in under 10 minutes, with very long response times being progressively rarer.



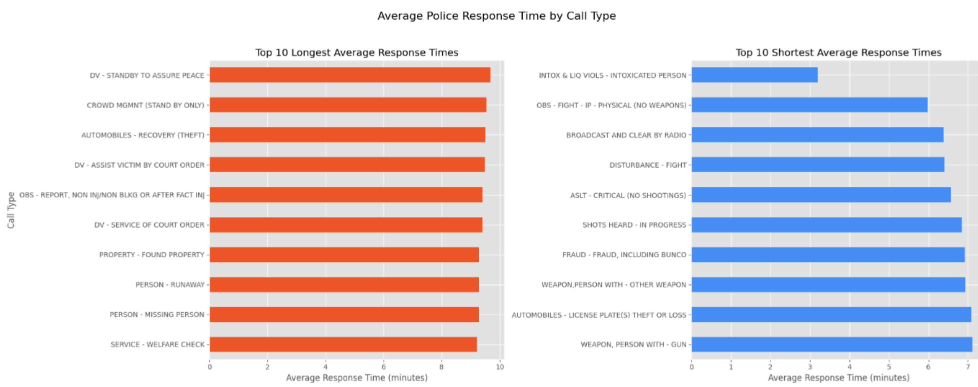
3.4.3 Seattle

Across Seattle's neighborhoods, average police response times are fairly consistent, generally falling between 9 and 10 minutes among the neighborhoods with the longest waits. The difference between neighborhoods is relatively small, but areas such as Commercial Harbor Island and South Park show slightly higher averages, nearing 10 minutes. This indicates that response time variation in Seattle is less dramatic than in larger cities like New York, with only modest differences across neighborhoods.



In Seattle, response times vary depending on the call, but the overall range is narrower than in some other cities. The longest average response time, around 9 to 10 minutes, is associated with lower urgency calls such as welfare checks, standby or court-order-related requests, missing or runaway persons, and reports taken after the fact. These call types usually involve no immediate danger, which explains why they receive slower responses.

In contrast, the shortest response times, typically between 3 and 7 minutes, are tied to high-priority, in-progress incidents such as disturbances, assaults, shots fired, weapons-related calls, and other situations that require immediate attention. This pattern highlights how Seattle prioritizes urgent, potentially dangerous events while allocating less time-sensitive cases to a slower response queue.



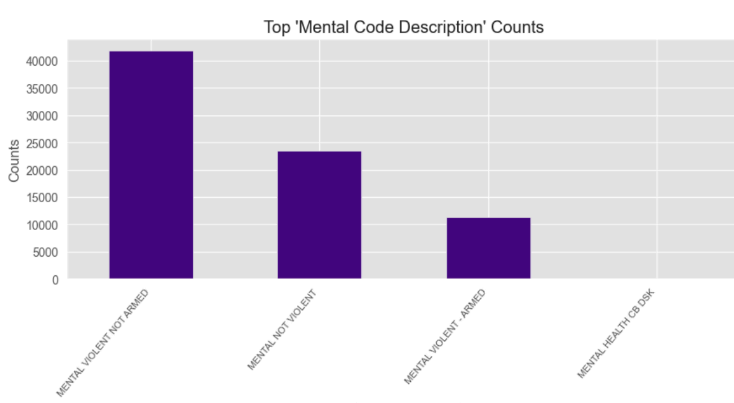
3.5 Mental Health Related Calls

3.5.1 New York

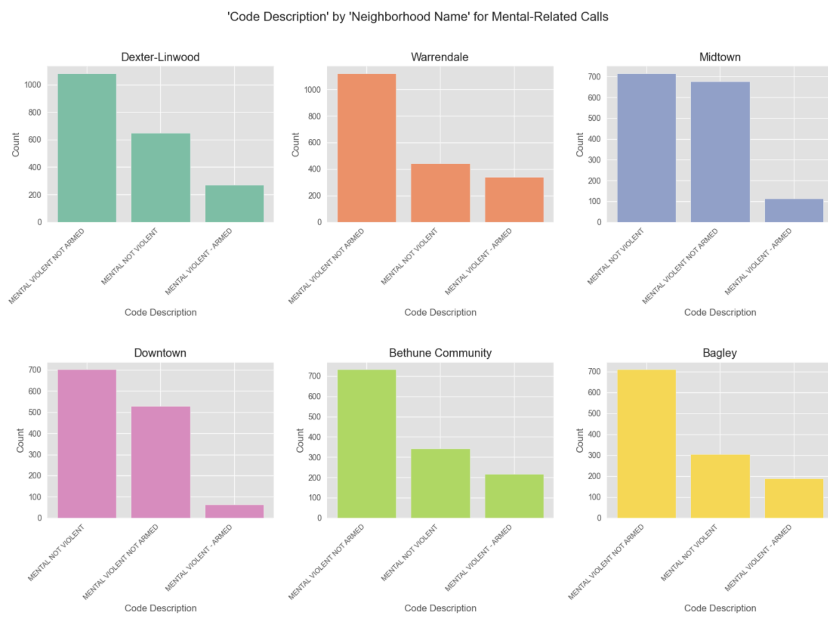
While NYC's raw data lacks explicit mental health indicators, we derived a mental health flag through keyword detection. But due to the data not existing, no meaningful analysis can be made

3.5.2 Detroit

The data shows that most mental health-related 911 calls in Detroit are classified as "Mental Violent Not Armed," followed by "Mental Not Violent" and "Mental Violent - Armed." A smaller number of incidents are categorized as "Mental Health C.B. Desk." This pattern indicates that the majority of mental health-related emergencies involve individuals in crisis without weapons involved.



When examining neighborhood-level patterns, Dexter-Linwood and Warrendale report the highest counts of mental health-related calls, while Midtown, Downtown, Bethune Community, and Bagley show similar but lower volumes. Across all neighborhoods, “Mental Violent Not Armed” consistently represents the most common classification, whereas “Mental Violent - Armed” remains the least frequent. These findings suggest localized concentrations of mental health-related incidents, with certain neighborhoods experiencing greater emergency activity.



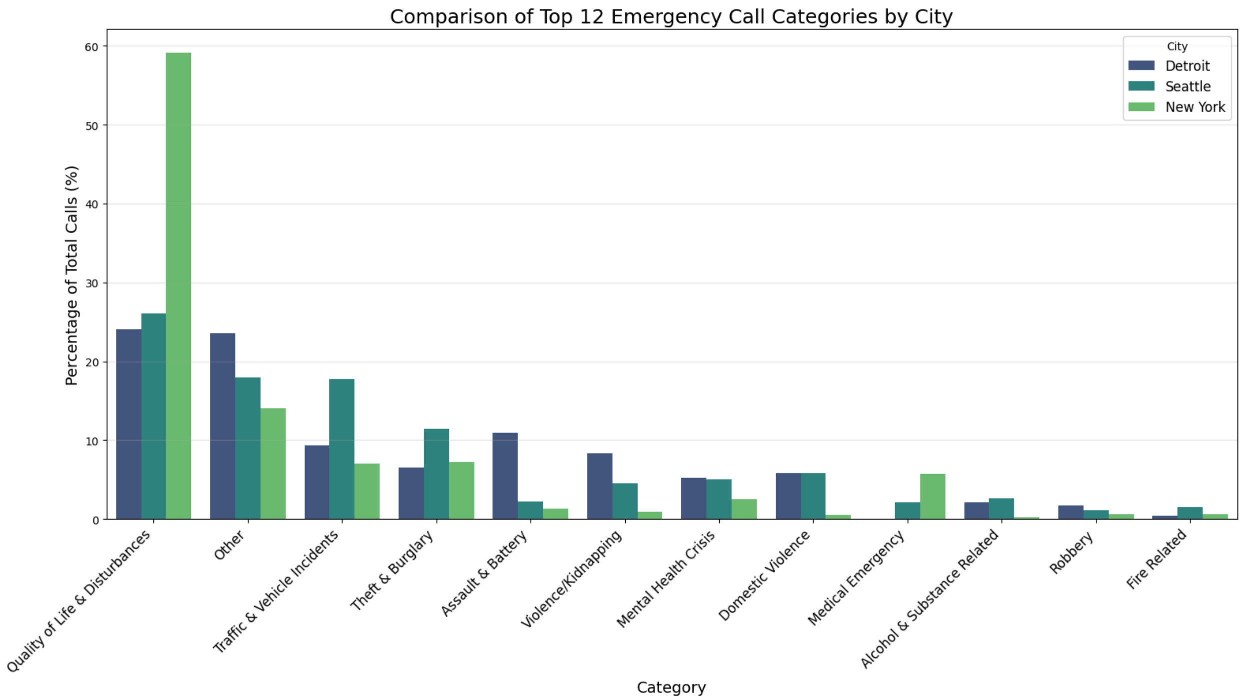
3.5.3 Seattle

Although Seattle has columns that correspond to mental health (CARE-related columns), there is not enough data present to make meaningful conclusions.

3.6 Cross City Analyses

3.6.1 Call Types

We compared the percentage distribution of the top 12 emergency call categories across Detroit, Seattle, and New York, showing clear differences in call patterns. New York is overwhelmingly dominated by “Quality of Life & Disturbances,” which makes up nearly 60% of calls, far higher than Detroit and Seattle, indicating a strong focus on non-violent, community-disruption issues. Detroit shows a more balanced spread, with relatively high proportions of “Other,” “Assault & Battery,” and “Violence/Kidnapping,” suggesting a heavier emphasis on serious crime-related calls compared to the other cities. Seattle stands out for higher shares in “Traffic & Vehicle Incidents” and “Theft & Burglary,” reflecting more activity in transportation and property-related issues. Across all three cities, categories such as “Fire Related,” “Robbery,” and “Medical Emergency” account for smaller portions of total calls, though Seattle and New York have slightly higher medical-related shares than Detroit. Overall, the graph highlights how each city's emergency call profile reflects different urban challenges, with New York focused more on quality-of-life concerns, Detroit showing higher violent crime-related calls, and Seattle leaning toward traffic and property crime issues.

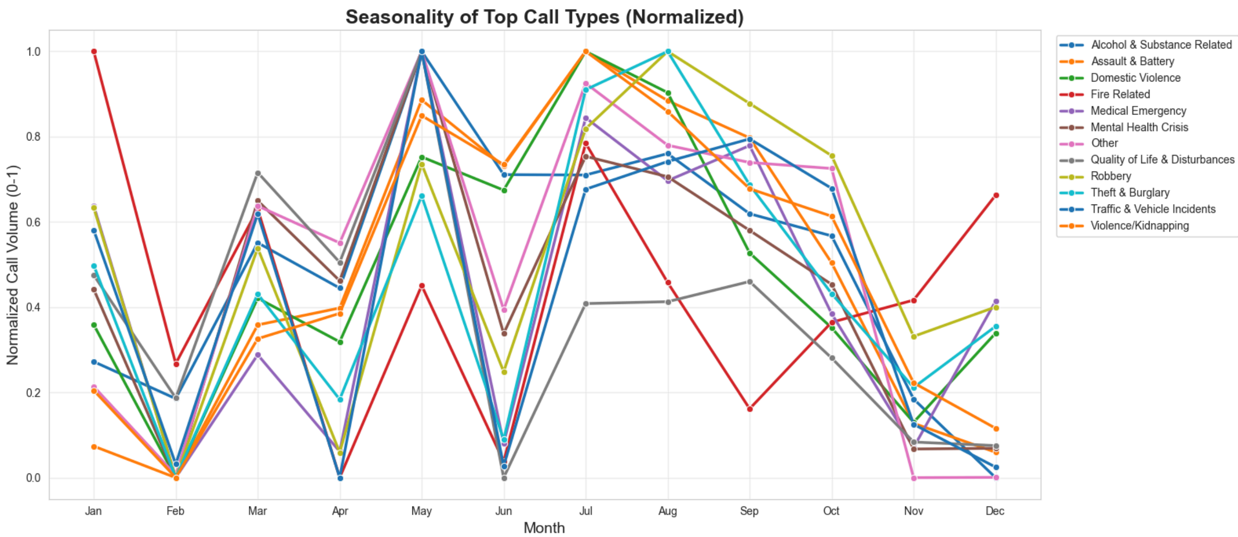


3.6.1 Seasonality of Call Types

Clear seasonal trends emerge across different emergency call types throughout the year. Most categories, including assault & battery, domestic violence, theft & burglary, robbery, traffic incidents, and quality-of-life disturbances, show a steady rise from late spring, reaching their highest normalized levels during summer (especially July and August). This suggests warmer months are associated with increased outdoor activity, travel, and social interaction, which correlates with higher incident and complaint volumes.

In contrast, many call types dip noticeably during late winter and early spring (February and April) and again during late fall and winter (November and December), when overall public activity tends to decrease. An important exception is fire-related calls, which show strong peaks in January and again toward the end of the year, likely tied to heating equipment use, holiday decorations, and seasonal fire hazards. Mental health and medical emergency calls appear more stable than other categories, but still follow a mild summer increase. Overall, the chart demonstrates that emergency call demand is highly cyclical, with most public-safety and quality-of-life issues intensifying in warmer months and easing in colder periods, while fire-related risks rise during winter.

The observed dip in call volume across all categories in June appears to be due to incomplete source data for that specific month, rather than a deviation from the overall seasonal trend of increased summer activity.



4. Data Analysis

4.1 Methodology

Data Preparation - Pipeline

To facilitate the standardization of the datasets, we developed an ETL pipeline that creates datasets that allow cross-city comparison. The pipeline normalizes heterogeneous datasets that vary in terms of the structure of the data (e.g., the structure of each column) and the classification scheme used to categorize a call into a common analysis-ready format. The pipeline is extensible such that new cities can be ingested with minimal code refactoring.

The core architecture is configuration-driven for greater scalability. The pipeline is controlled through a YAML format file, which maps the inputs and outputs and provides rules to map and filter columns, instead of hardcoding city-specific logic in the processing script itself. The abstraction layer lets the research team add new cities or new data fields through a configuration file update, greatly cutting the development time needed to expand the system.

The processing workflow is implemented as a four-stage execution:

1. **Ingestion (Read)**: The raw data from commissioned open datasets from target cities such as Austin, Chicago, Detroit, and New York is in various formats, such as CSV, Excel, etc. The ingestion pipeline processes larger datasets in chunks to avoid using too much memory when reading larger files, as is the case with New York.
2. **Preprocessing (Clean)**: Once ingested, the data goes through an initial cleaning process where the column headers are normalized, timestamp errors are corrected, and the data types are standardized.
3. **Standardization and Mapping**: This is the critical transformation phase where city-specific schemas are mapped to the project's unified data model.
    - Geospatial Alignment: Location data is processed and mapped to standardized identifiers, such as GEOIDs, to facilitate spatial analysis.
    - Metric Computation: Derived metrics necessary for downstream analysis, such as response times, are calculated dynamically during this stage.
4. The final processed data is serialized into compressed Parquet files. This format was chosen to optimize storage efficiency and query performance for subsequent analysis.

Unique Challenge for New York Data

The NYC dataset presented unique challenges due to its scale and structure. The data was provided in two separate files requiring consolidation before analysis could begin. However, both of these datasets are too big, and when combined, the total number of rows exceeded 53 million. Therefore, it exceeds the memory capacity of standard computing hardware. Loading 53 million records simultaneously requires approximately 20 gigabytes of RAM, making conventional processing approaches unfeasible.

We employed chunked processing, as mentioned, reading and processing data in batches of 500,000 records. This approach balanced memory efficiency (each chunk required only 2-3GB RAM) against processing speed (minimizing file I/O overhead). This chunk size was determined through testing to optimize performance on systems with 32GB of total memory. For over 50 million records, we processed 105 chunks in total, then combined all the chunks into one final single file for the ETL across cities.

Call Categorization

The largest obstacle that was encountered in this project was that different cities had vastly different terminology. This was especially true in the case of the 911 call types/categories. For instance, when it comes to armed robbery calls, the three cities use notably different terminology. Detroit uses timing codes like *'Robbery armed ip-any' (in progress)*, *'Robbery armed jh-any weapon' (just happened)*, and *'Robbery armed report' (after the fact)*. New York specifies both timing and location with descriptions like *'Robbery (in progress): commercial'* or *'Robbery (past): other/outside'*. Seattle uses simpler categories like *'Robbery - armed'*, *'Robbery - commercial'*, and *'Robbery - street'* that focus on type and location rather than timing. Another example would be looking at vehicular accidents. New York data refers to these incidents as '*Vehicle Accident*', whereas Seattle calls these '*Traffic-Accident*'.

One of the key methodological challenges to engaging in meaningful cross-jurisdictional analysis came with the disparate naming conventions and terminologies used across the three datasets. In fact, this results in a lack of standardization in call type classification systems, which fundamentally obstructs our ability to perform comparative studies that might otherwise generate new insight into patterns of urban public safety and resource allocation strategies across different metropolitan contexts.

In order to overcome this analytical obstacle, we designed and applied a structured methodology for standardizing the heterogeneous categorical data into one standard taxonomic structure. More precisely, we created a Python mapping algorithm that converts the city-specific call type designations to a uniform classification schema with 15 higher-order categorical groups. The script performs lexical analysis of the original call descriptions by using keyword detection and pattern matching to find semantic similarities between the different terminologies. It uses these identified linguistic markers within the call type descriptions to systematically assign each discrete city-specific category to its appropriate standardized classification.

This methodological intervention proved essential in providing the shared analytical ground necessary for robust cross-city comparisons of the same measures and operational patterns. In this, we turned what was essentially three incommensurable datasets into one coherent analytical frame that allows for the comparison not only of differences and similarities, but also broader trends in public safety call patterns across the three different urban jurisdictions investigated.

Geographic Enrichment

The client had emphasized several times that they require geographical data in the prepared dataset to assess whether there is a relation between call types and geographical locations. After discussing with the client, we decided to use Census block group identifiers (GEOIDs) as the geographical data format. This is because they already use this in their current system. The pipeline enriches call data with GEOIDs through a spatial join process. First, Census TIGER/Line shapefiles containing geographic boundaries for block groups are downloaded from the U.S. Census Bureau's website using the state FIPS code (e.g., "53" for Washington). The shapefile is then loaded using geopandas and converted to the WGS84 coordinate reference system (EPSG:4326) to match standard latitude/longitude coordinates.

For records with valid latitude and longitude coordinates, the pipeline creates a GeoDataFrame by converting coordinate pairs into point geometries. A spatial join operation is then performed using the "within" predicate, which matches each call location point to the Census block group polygon that contains it. This spatial join assigns the corresponding GEOID to each call record based on its geographic location. Records with missing coordinates are left without GEOIDs. The process enables neighborhood-level analysis by linking each emergency call to standardized Census geographic units, facilitating cross-city comparisons and demographic correlations.

Visualization techniques

In the Exploratory Data Analysis, a comprehensive collection of visualization techniques was used. This was facilitated by using the MatPlotLib, SeaBorn, and Folium Python libraries. Bar Charts were chosen as the primary method to compare most metrics, such as call category counts and response times, as they allow for quick and easy interpretation. Color gradient heatmaps were utilized to display temporal patterns, such as the relationship between day of week and call volumes, or hour of day and response times. Stacked bar charts show the mix of call types across different neighborhoods in a given city. Geographic visualizations using Folium with HeatMap features provided spatial analysis of call concentrations across urban areas.

5. Answers to Key Project Questions

- **Who is calling 911, and under what circumstances?**

The datasets we used do not include individual-level demographic information about callers; we can still draw meaningful insights by examining temporal, geographic, and incident-level characteristics present in the data. In this context, we interpret "who is calling 911" as the circumstances and recurring behavioral patterns that result in 911 calls instead of the personal identity of callers.

Across all three cities, most calls involve non-violent incidents such as disturbances, quality-of-life concerns, and welfare checks. Call activity follows consistent temporal patterns, with increasing volumes throughout the afternoon and peaking during evening and late night hours, particularly in warmer months. Geographic data shows that calls are most concentrated in high-density, high-activity areas such as commercial districts and downtown neighborhoods. Residential areas show lower but steady levels of demand.

- **Are there noticeable patterns in call types by city, season, or demographic group or geographic area?**

After comparing the analyses of the three cities, certain conclusions can be made. First of all, it is clear from the results that the summer months, from June to August, show the highest number of calls. This is true for all three cities examined. This can be attributed to the fact that the agreeable weather during these months prompts more people to spend time outside, increasing the chances of incidents happening.

Another consistent pattern is that the volume of calls peaks in the late evening, roughly from 8 pm to 10 pm. This is likely due to the fact that this time represents a combination of the tail-end of rush hour—where there is still a lot of movement and activity—and the period when people start going out for the evening. The influence of rush hour also explains why weekdays have a higher number of calls compared to weekends. While these patterns were common across all the cities, there are variances due to the different natures of each city, which have been discussed above.

Furthermore, New York's calls are heavily dominated by quality-of-life and disturbance issues, Detroit shows relatively higher shares of violent crime–related calls like assault and violence, and Seattle has a larger proportion of traffic incidents and theft/burglary, indicating different urban priorities across the cities.

- **How does 911 call volume correlate with fatal police violence or neighborhood-level indicators?**

Since we didn’t have any data about fatal police violence or any neighborhood-level indicators, we were not able to answer this question.

- **What does standardized 911 data reveal that siloed data does not?**

Before this standardization effort, 911 call data existed as "incommensurable datasets" trapped within isolated municipal repositories, where unique formats and terminologies obscured broader public safety insights. By harmonizing these disparate sources into a single coherent analytical frame, standardized data reveals shared operational patterns that remain invisible in siloed contexts. For example, while individual city portals prevent direct comparison due to incompatible categorical systems, such as Detroit’s reliance on timing codes versus Seattle’s type-based classifications, the standardized dataset exposes universal human behaviors that transcend geography, such as the consistent summer peak in call volumes and the surge in activity during the late evening transition period across all three cities. Furthermore, this unified lens clarifies distinct "urban safety profiles," rigorously demonstrating how New York is structurally dominated by quality-of-life and transit concerns, whereas Detroit and Seattle show significantly higher relative demands for violent crime and property issues, respectively. Ultimately, standardization transforms fragmented local records into a powerful tool for comparative analysis, allowing researchers to distinguish between localized service demands and broader, systemic trends in policing.

6. Conclusions and Recommendations

This project successfully demonstrated the technical and analytical feasibility of harmonizing massive, heterogeneous 911 datasets comprising over 50 million records into a unified, analysis-ready format. The cross-city analysis leads to the primary conclusion that while the nature of emergency demand is deeply local—shaped by specific urban characteristics such as New York's transit density or Detroit's specific safety challenges—the temporal dynamics of policing are remarkably consistent. By normalizing timestamps across time zones, the standardized data reveals that all cities exhibit identical summer peaks and late-night activity surges, suggesting that police demand is driven by universal human activity cycles rather than local policy.

Consequently, the study confirms that "one-size-fits-all" assumptions about public safety are insufficient. Instead, the unified taxonomy clarifies the unique "urban fingerprints" of each municipality: New York is structurally dominated by quality-of-life and transit concerns (nearly 60% of calls), whereas Detroit and Seattle show significantly higher relative demands for violent crime and property issues, respectively. Furthermore, the standardization process exposed critical variations in data collection methodologies. We found that Detroit's explicit multi-category system provides richer data than NYC's keyword-based approach, and that the absence of standardized mental health crisis tracking across cities prevents systematic analysis of a critical public safety issue.

Based on these findings, we recommend scaling the data pipeline to ingest 911 records from the majority of major U.S. cities, leveraging the extensible, configuration-driven architecture established in this pilot phase. Scaling this research would allow agencies to move from isolated decision-making to evidence-based policy transfer, identifying effective practices in peer cities with similar urban profiles. To further enhance the analytical value of this repository, future iterations should prioritize the integration of demographic and outcome-level data—enabled by our pipeline's standardization of incidents to Census Block Groups (GEOIDs)—to allow for critical assessments of equitable policing. Finally, we recommend the broader adoption of the standardized 15-category classification schema developed in this study, which proved essential in providing the "shared analytical ground" necessary to benchmark public safety operations across diverse metropolitan contexts.

Team Member contributions in cleaning, preprocessing, and performing analysis on data

1) Jaiveer - Cleaning and analysis for New York City and Seattle; Seattle Pipeline
2) Johanna - Detroit and Seattle final column mapping; data interpretation for Seattle, Detroit, and New York
3) San (Will) - Analyzed Boston EDA before dropping, New York cleaning pipeline, tested city pipelines, and New York final column mapping.
4) Dang - Merging strategy and New York Pipeline, develop geoid converters that can work across cities.
5) Damayanti - Detroit EDA, Detroit Pipeline, Categorization code, Pipeline Configuration by YAML
6) Peng - Boston EDA, insights before dropping; Cleaning and analysis for Detroit