Analyzing Parking Violations in New York City

Miriam Palmetshofer*
University of Applied Sciences Upper Austria
Hagenberg, Austria
miriam.palmetshofer@students.fh-hagenberg.at

Lukas Zandomeneghi*
University of Applied Sciences Upper Austria
Hagenberg, Austria
lukas.zandomeneghi@students.fh-hagenberg.at

ABSTRACT

This project analyzes over 9 million NYC parking violations from fiscal year 2014 to uncover spatial, temporal, and agency-level patterns. Using Apache Spark, Elasticsearch, and Kibana, we find that repeat offenders—primarily commercial vehicles like UPS trucks—cluster in Manhattan's busiest streets. Violations decline on Sundays, holidays, and during snowstorms, and peak during weekday business hours. Manhattan leads in violation density, while the "TRAFFIC" agency dominates enforcement. Despite some data quality issues, the analysis offers valuable insights for improving parking policy and enforcement of parking violations.

KEYWORDS

Big Data, Apache Spark, Kibana, Parking Tickets, Urban Analytics, NYC OpenData

1 INTRODUCTION

In a densely populated and heavily trafficked city like New York, effective parking enforcement is critical to maintaining order and accessibility. Each year, the NYC Department of Finance issues over 10 million parking tickets, creating a rich dataset that reflects patterns of driver behavior, enforcement practices, and spatial-temporal dynamics in the city.

With the growing availability of public data and scalable processing tools, there is an opportunity to move toward data-driven insights. By analyzing large-scale parking violation records, our aim is to uncover trends in parking behavior and assess the consistency and coverage of enforcement agencies.

1.1 Research Questions

This study explores the following research questions.

- **Do repeat offenders cluster in certain areas or vehicle types?** Understanding repeat violations can inform targeted awareness campaigns or increased monitoring.
- **Are parking violations more common during specific times or days?** Identifying temporal patterns can help optimize enforcement schedules.
- **Can violation hotspots be identified?** Detecting spatial clusters of violations may support urban planning and infrastructure improvements.
- **How does enforcement strength vary between agencies?** Analyzing the frequency of the tickets by agency can reveal differences in the rigor of the enforcement across regions.

To answer these questions, we applied big data technologies, specifically Apache Spark for distributed data processing and Kibana for interactive visual analytics, to NYC parking violation data.

*Both authors contributed equally to this research.

2 DATASET

2.1 Source and Scope

The dataset utilized in this study is sourced from the NYC Open Data portal, specifically the "Parking Violations Issued - Fiscal Year 2014" dataset. This data set coversovers all parking violations issued in New York City during the fiscal year 2014, which spans from July 1, 2013, to June 30, 2014. The data is provided by the New York City Department of Finance and includes detailed information about each parking violation recorded during this period. The dataset is publicly available[1] and can be accessed through the NYC Open Data portal. It is provided in a CSV format, which facilitates ease of use with various data processing tools and platforms.

2.2 Structure and Features

The dataset comprises over 9 million records, each representing a single parking violation. Each record contains 43 fields that provide comprehensive details about the violation. Key fields include a unique identifier for each ticket, the license plate number of the vehicle, the state in which the vehicle is registered, Issue Date, Violation Code etc.

These fields enable a multifaceted analysis of parking violations, examining temporal patterns, spatial distribution, vehicle characteristics, and enforcement agency behaviors.

2.3 Data Quality and Limitations

Although the data set set is comprehensive, certain limitations should be acknowledged. Some records contain missing or incomplete information. This can impact the accuracy and completeness of certain analyses. Additionally, inconsistencies in data entry—such as variations in how vehicle types or plate types are recorded—can introduce noise and reduce the reliability. Another notable issue is data duplication. In some cases, records for the same location, time and plate appear multiple times. This was taken into account as far as possible when pre-processing the data.

Despite these limitations, the dataset remains a valuable resource for understanding parking violation patterns and enforcement practices.

3 METHODOLOGY AND TECHNOLOGY STACK

3.1 Apache Spark

Apache Spark was used as the primary data processing engine for handling the large-scale dataset. The Spark job was containerized using Docker and deployed in a standalone cluster mode, consisting of a master node and a dedicated worker. We implemented our data ingestion, transformation, and aggregation scripts in PySpark, and

[1]NYC Open Data – Parking Violations Issued – Fiscal Year 2014, available at: https://data.cityofnewyork.us/City-Government/jt7v-77mi

mounted them into the container via a shared volume. The worker was configured with 1 core and 1 GB of memory, sufficient for our exploratory analysis workload.

3.2 Elasticsearch

Elasticsearch served as the storage and indexing layer for the processed data. Spark job outputs were structured and exported in a format compatible with Elasticsearch. We used Elasticsearch in a single-node mode for simplicity, with security features disabled to ease local development. Data was persisted via a Docker-managed volume to retain indices between container restarts.

3.3 Kibana

Kibana was used as the visualization layers. It connected directly to the Elasticsearch instance and was accessible via a web interface. The visualizations were created using Lens to support answering the project's research questions.

4 EXPLORATORY DATA ANALYSIS

4.1 R1 – Temporal Trends

To analyze daily patterns in parking violations, we focused on the period from November 10, 2013, to March 3, 2014. This timeframe was selected due to its inclusion of Christmas and Thanksgiving.

A clear trend emerges: parking violations drop significantly on Sundays and public holidays. For instance, Thanksgiving (November 28, 2013) and Christmas Day (December 25, 2013) show sharp declines in daily ticket counts. This pattern probably reflects lower traffic volumes and fewer parking enforcement officers.

More interestingly, several dips in violation counts occurred on weekdays that were neither holidays nor Sundays. Upon investigation, these dates aligned with major snowstorms that impacted New York City. (see Appendix 1) These events included:

- **Winter Storm Hercules** on January 2–3, 2014 [2]
- **January Blizzard** on January 21–22, 2014 [4]
- **Heavy snowfall** during Winter Storm Pax on February 13, 2014 [1]
- **February Nor'easter** on February 13–14, 2014 [3]

During these events, parking violations dropped sharply—likely due to suspended enforcement and reduced traffic. The results show that Sundays, public holidays and environmental influences have a significant impact on the number of park tickets issued.

4.2 R2 – Repeat Offenders

To identify patterns among drivers with frequent parking violations, we defined "repeat offenders" as vehicles with more than five recorded violations within the dataset. After filtering accordingly, we focused our analysis on the top seven license plate IDs with the highest violation counts. (see Appendix 2)

To our surprise, the most frequently fined vehicle had accumulated over 1,400 parking violations during the fiscal year. Upon closer inspection, we observed that in many cases the vehicle received multiple violation codes for the same time and location, which contributed to the high count.

Curious about the types of vehicles incurring such high numbers of tickets, we analyzed their registered plate types. (see Appendix 3) All top offenders were marked as `COM` — indicating commercial vehicles. To gain further insight, we used the online tool *lookupaplate.com*[2] to gather technical details like brand and specific model.

The results revealed that the top three offenders were UPS delivery trucks. For some of the remaining vehicles, no match could be found; however, at least two were identified as Peterbilt trucks — a brand typically associated with commercial vehicles. This strongly suggests that frequent violators tend to be part of delivery fleets.

To further support this assumption, we examined the timestamps of the violations. (see Appendix 4) Interestingly, none of the high-frequency offenders received tickets on weekends. All recorded violations occurred Monday through Friday, which aligns with typical business days.

To understand where these violations were most commonly issued, we examined the distribution of tickets across street names. (see Appendix 5) The most frequent locations were Broadway, 3rd Avenue, Madison Avenue, 5th Avenue, and Lexington Avenue — all of which are major Manhattan thoroughfares. These streets are characterized by high traffic volume, extensive commercial activity, designated delivery zones, and strict parking enforcement — conditions that make it particularly challenging for delivery drivers to comply with all regulations.

4.3 R3 – Violation Hotspots

To identify where and when parking violations are most concentrated, we performed an analysis of geographic and temporal distribution patterns.

Geographic Hotspots. Violation frequencies were aggregated by street name and county. The main streets—Broadway, 3rd Avenue, 5th Avenue, Madison Avenue, and Lexington Avenue—align with dense commercial zones in Manhattan and highly correlate to the violations per street of the repeated offenders from R2 (see Appendix 5 and Appendix 6). At the county level (Appendix 7), Manhattan (New York) had the highest ticket volume, followed by Brooklyn (Kings), Queens, and the Bronx. Staten Island (Richmond) had the fewest ticket volume. Violations that were not associated with any county are grouped together to "OTHER".

Temporal Hotspots. Hourly violation trends showed consistent peaks between 6 AM and 6 PM, corresponding to active enforcement periods and business hours. Violations were significantly lower during early mornings, evenings and nights (see Appendix 8. Monthly counts dipped in January and February 2014, likely due to severe winter weather (see R1, for detailed analysis).

4.4 R4 – Agency Strength

To evaluate the enforcement strength of different agencies, we computed the daily ticket issuance statistics for each issuing agency. These statistics included the average, maximum, minimum, and standard deviation of daily ticket counts. We enriched the data using a lookup table that mapped agency codes (e.g., "T", "P") to their full agency names (e.g., "TRAFFIC", "POLICE DEPARTMENT").

Appendix 9 visualizes the values of these four metrics across the top seven agencies by ticket volume. The "TRAFFIC" agency stands out overwhelmingly, with a daily average exceeding 15,000

[2] https://www.lookupaplate.com/new-york

tickets and a daily maximum above 40,000 tickets. It also exhibits a relatively high standard deviation, suggesting variability in enforcement activity.

Other agencies such as the "DEPARTMENT OF TRANSPORTATION" and "POLICE DEPARTMENT" issue significantly fewer tickets per day—both in terms of average and peak values. The "DEPARTMENT OF SANITATION" "OTHER/UNKNOWN AGENCIES" and "NYS COURT OFFICERS" all show low statistics, which could be due to relatively limited enforcement roles.

The chart suggests a heavily centralized enforcement landscape where a few agencies—especially "TRAFFIC"—dominate ticket issuance. This may hint towards a broader jurisdiction and operational scale within New York City.

An analysis of issued tickets grouped by county and agency shows that mainly one agency ("DEPARTMENT OF TRANSPORTATION") is responsible for the "OTHER" entries in the county (Appendix 10).

REFERENCES

[1] Time Magazine. 2014. *Winter Storm Pax Dumps Snow, Ice Across South, Northeast.* https://time.com/7222/winter-weather-storm-pax-school-closings/ Accessed: 2025-05-17.

[2] Office of the Mayor, NYC. 2014. *Mayor de Blasio Issues Winter Storm Update.* https://www.nyc.gov/office-of-the-mayor/news/009-14/mayor-de-blasio-issues-winter-storm-update Accessed: 2025-05-17.

[3] Wikipedia contributors. 2014. *February 2014 nor'easter.* https://en.wikipedia.org/wiki/February_2014_nor%27easter Accessed: 2025-05-17.

[4] Wikipedia contributors. 2014. *February 2014, North American Blizzard.* https://en.wikipedia.org/wiki/January_20%E2%80%9322,_2014_North_American_blizzard Accessed: 2025-05-17.

A FIGURES

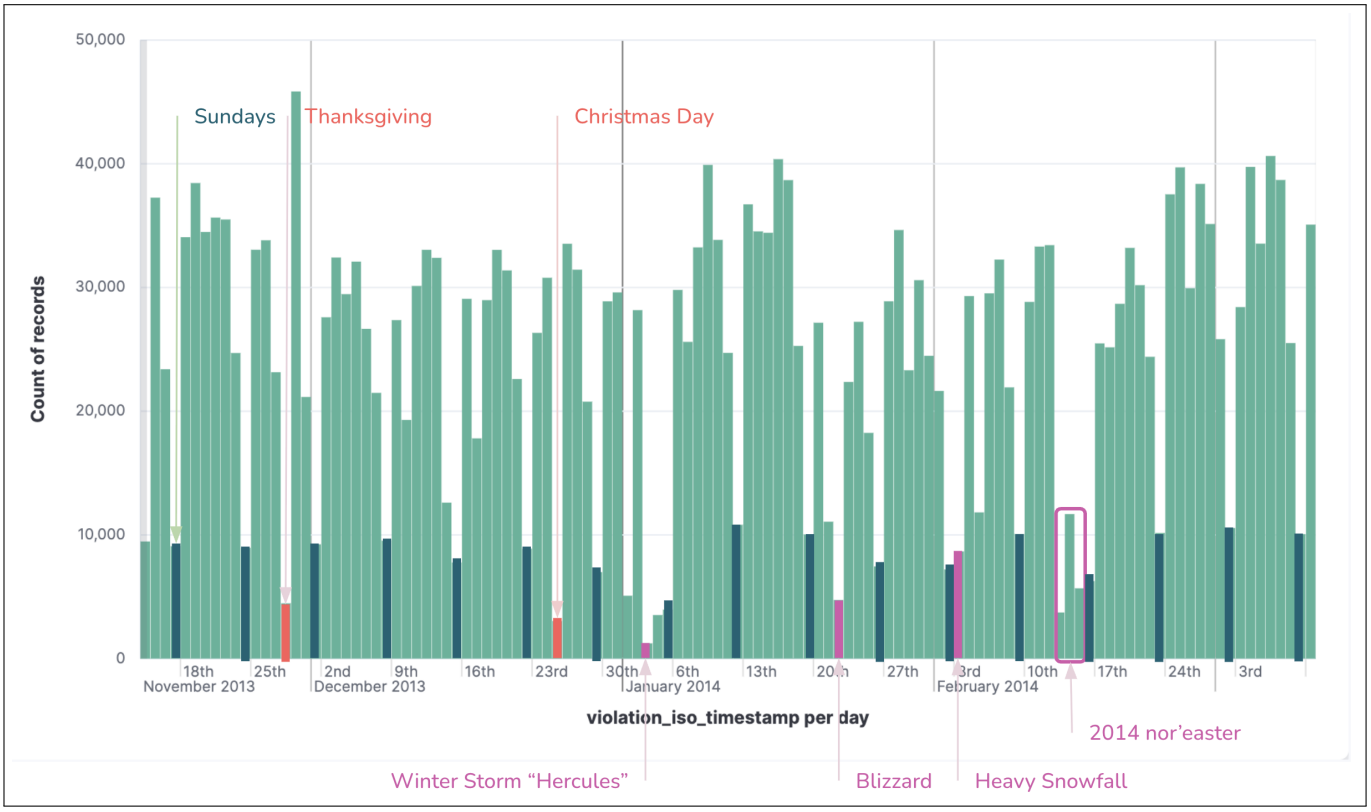

Figure 1: Daily parking violations between November 2013 and March 2014, annotated with Sundays, holidays, and major snowstorms.

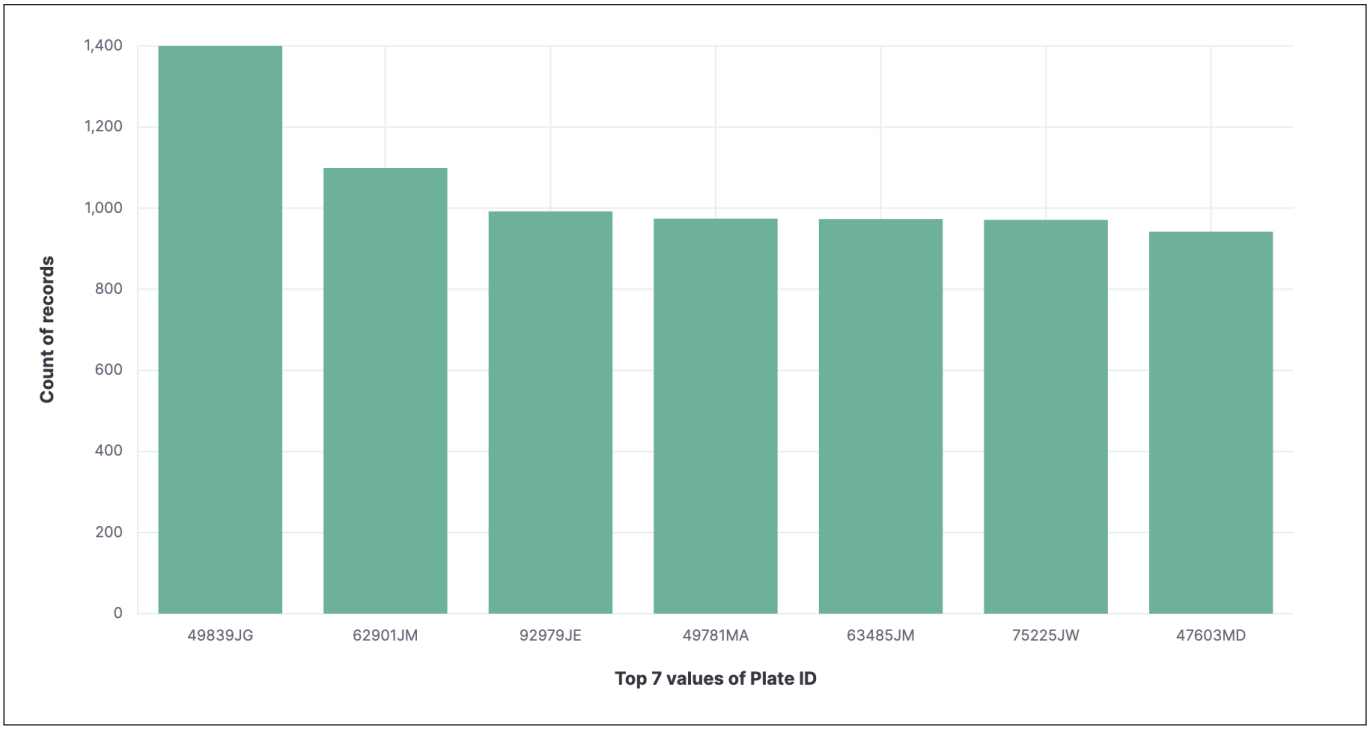

Figure 2: Analysis of top seven repeated offenders by plate ID

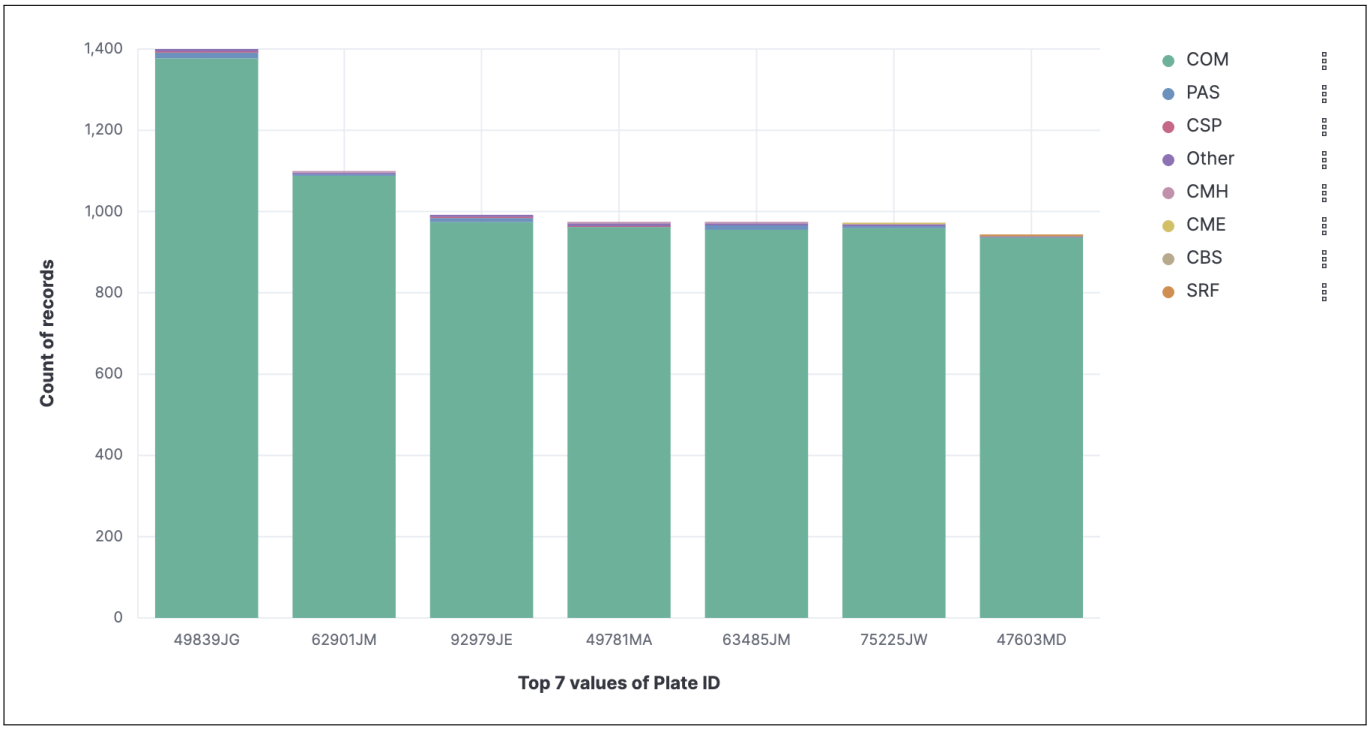

Figure 3: Analysis of top seven repeated offenders by plate ID and plate type

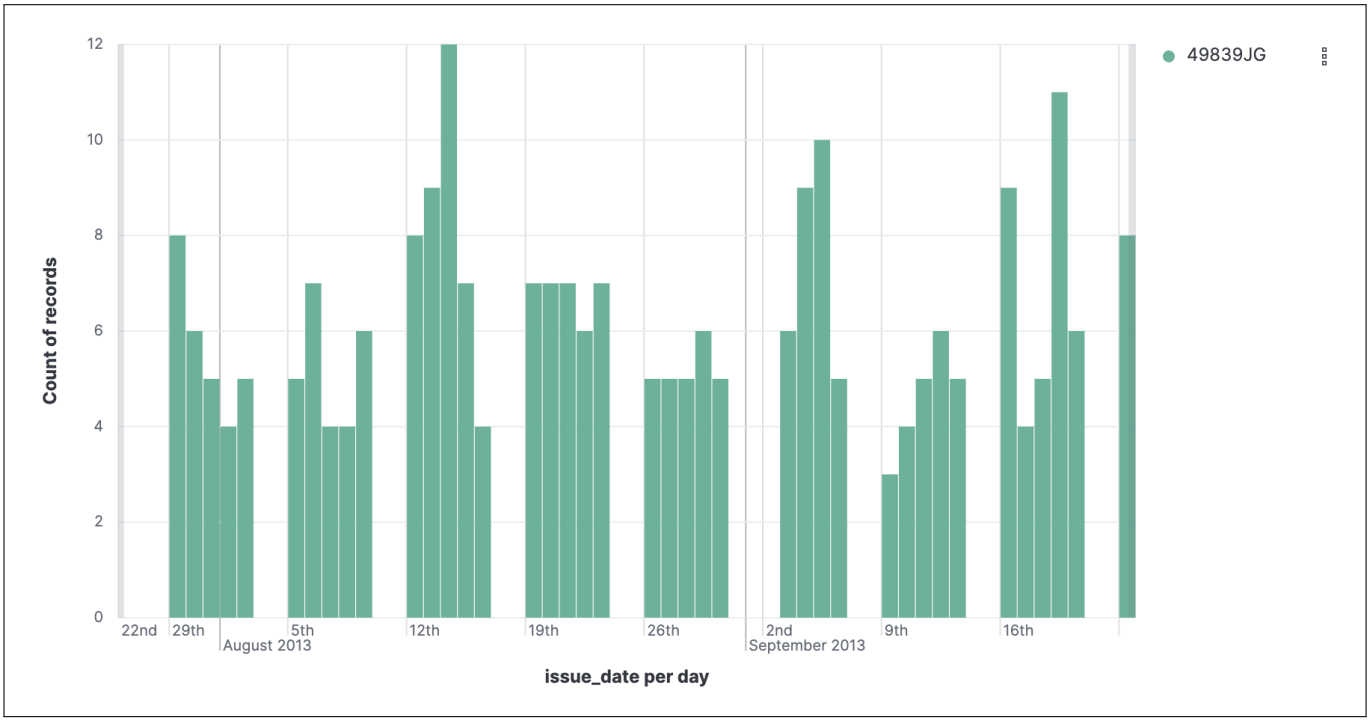

Figure 4: Analysis of violation time of top repeated offender

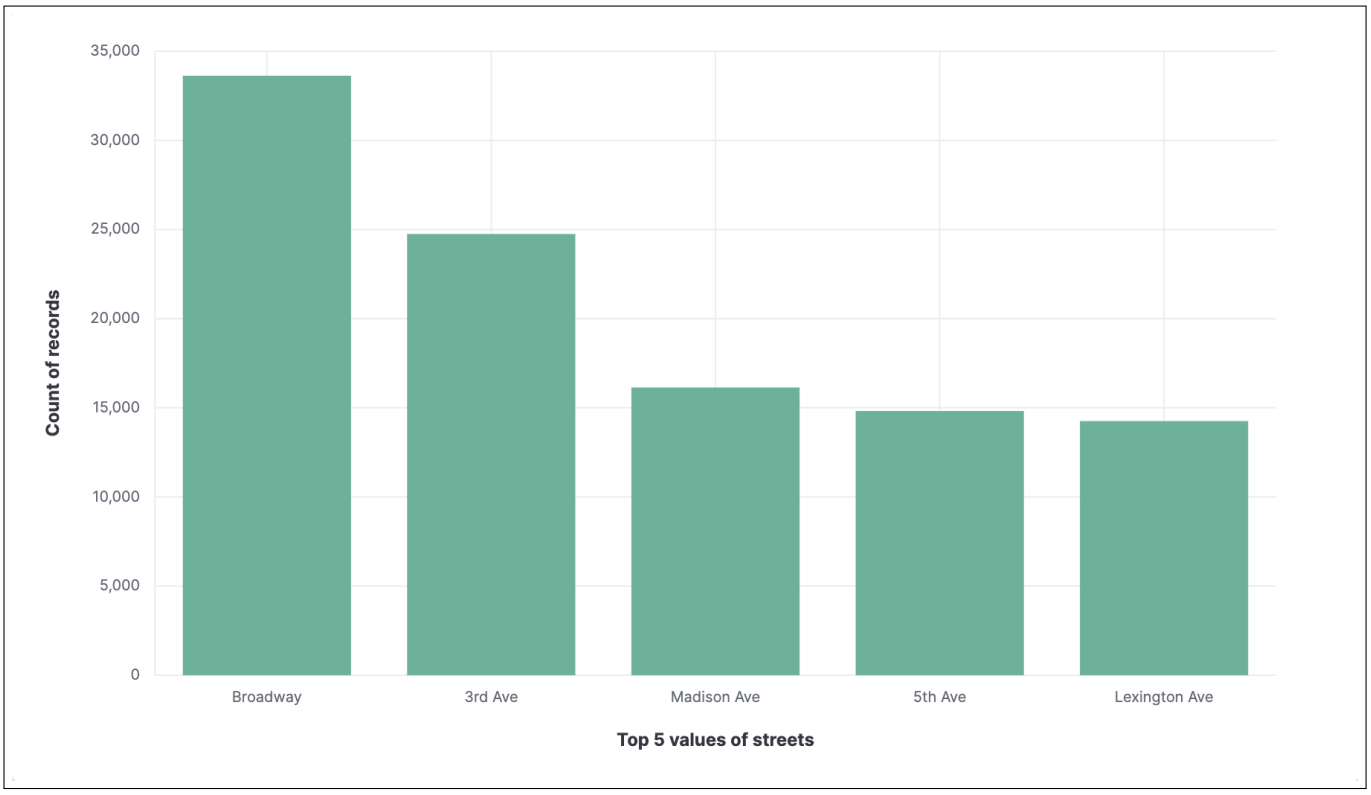

Figure 5: Analysis of most common streets for repeated offenders

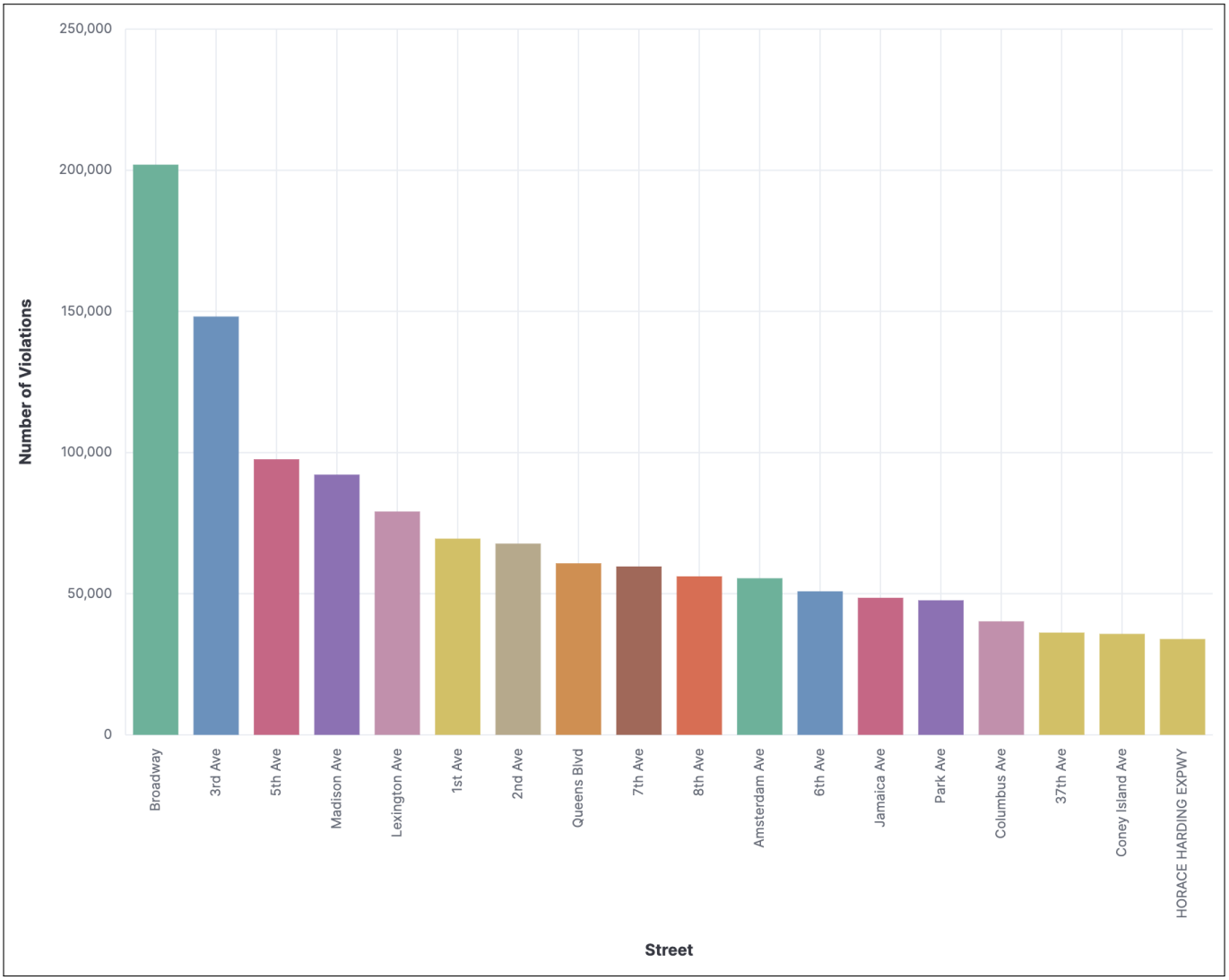

Figure 6: Analysis of the streets with the most parking violations

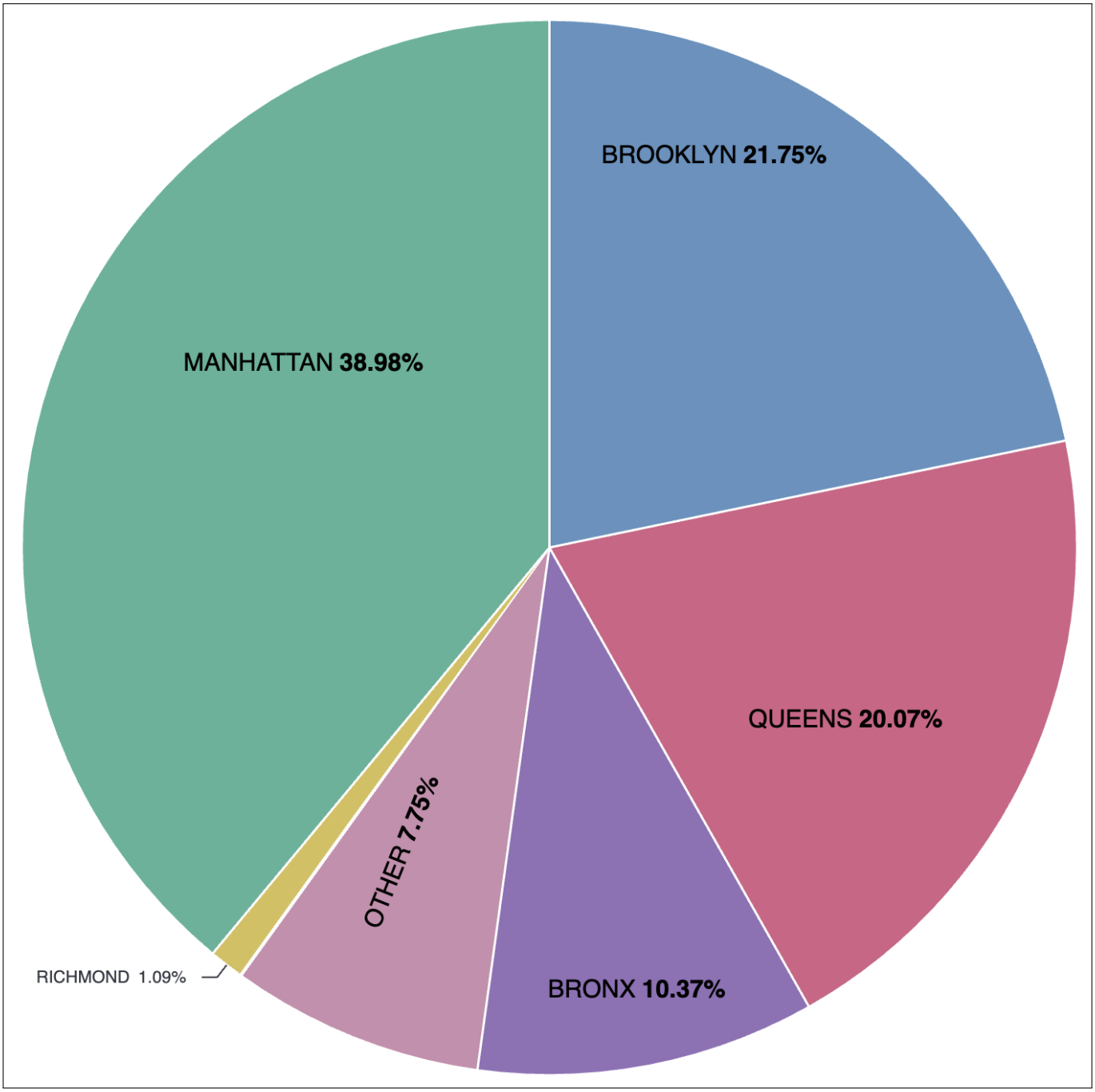

Figure 7: Analysis of the violation counts grouped by county

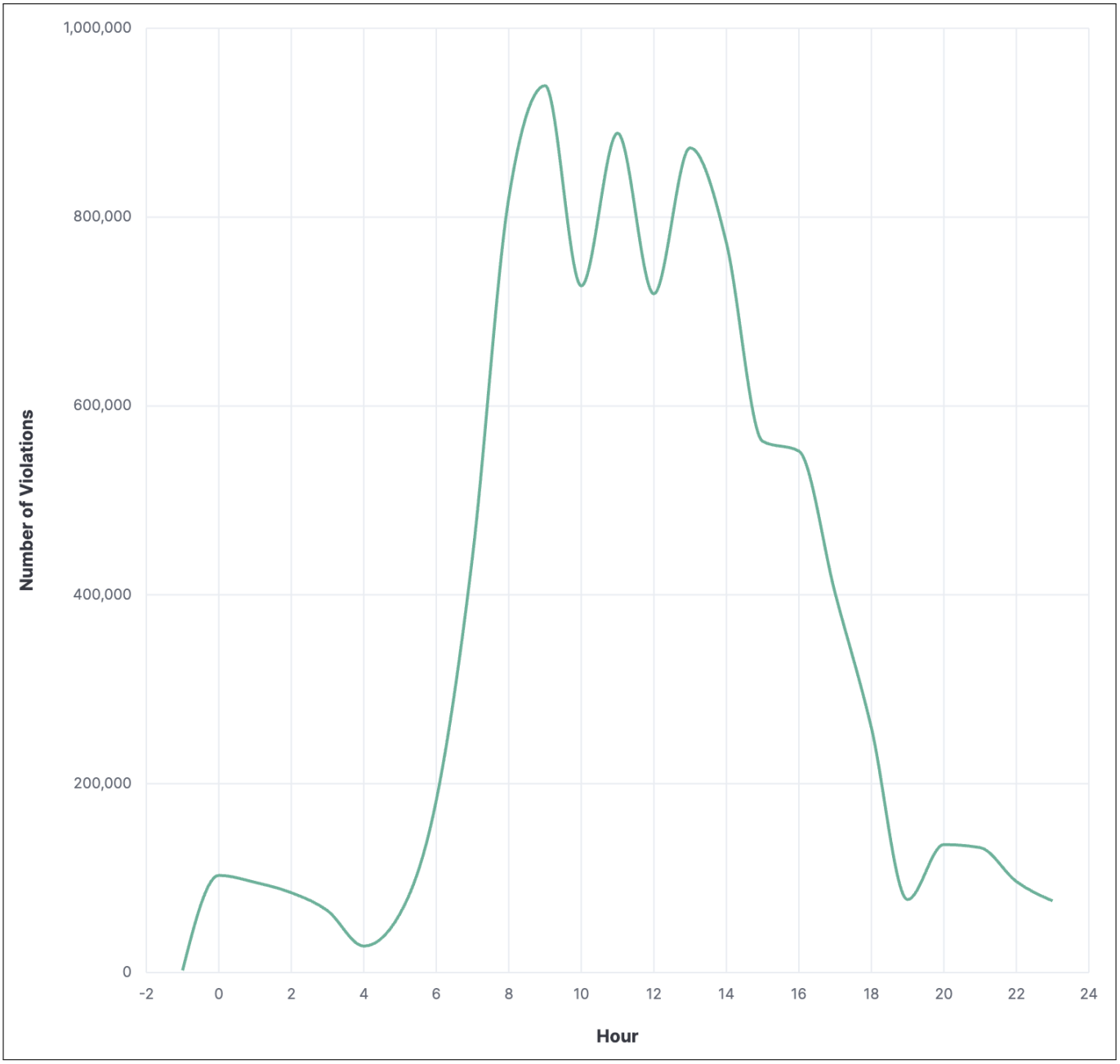

Figure 8: Analysis of the hourly violations (total counts)

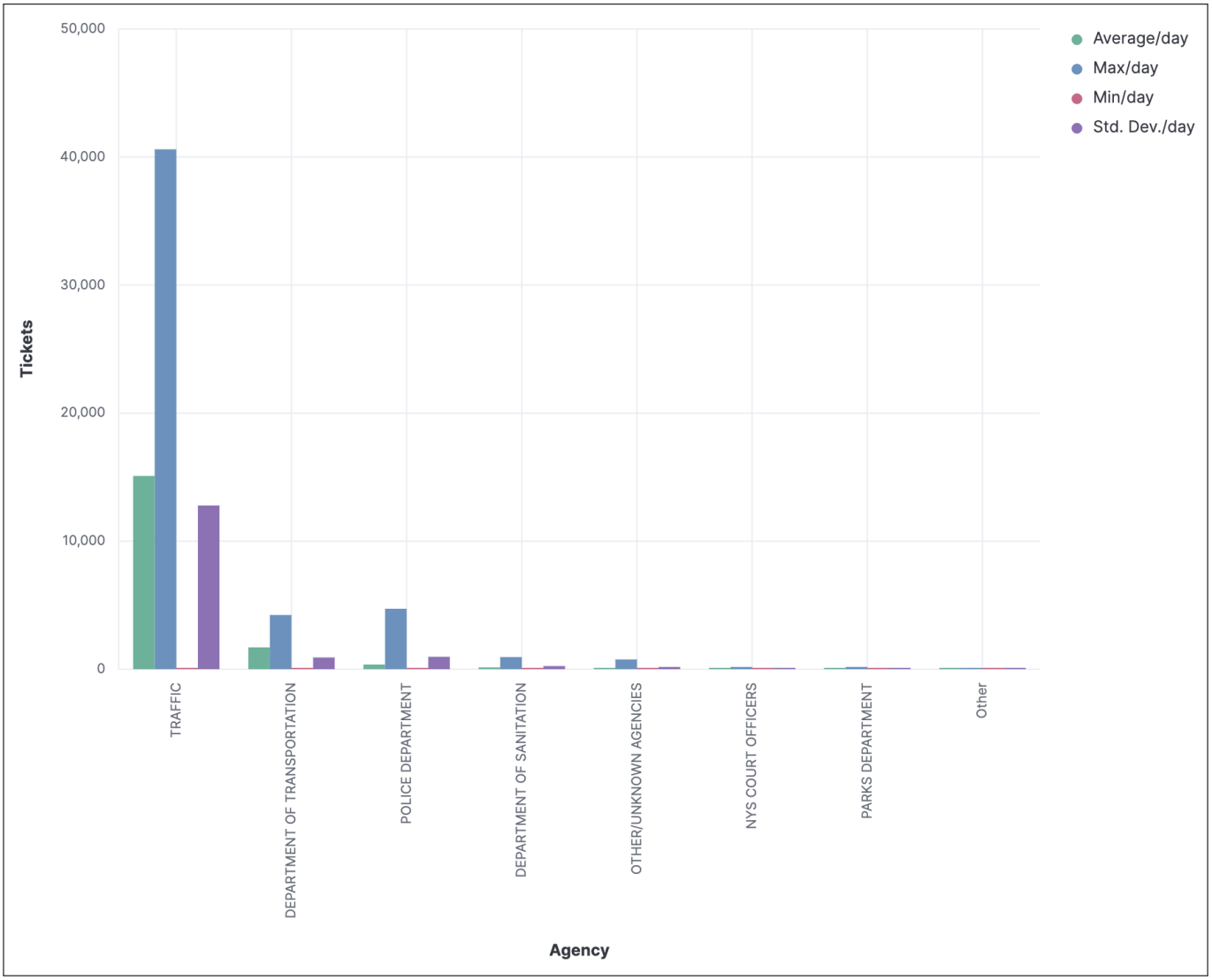

Figure 9: Tickets per day metrics grouped by the enforcing agencies

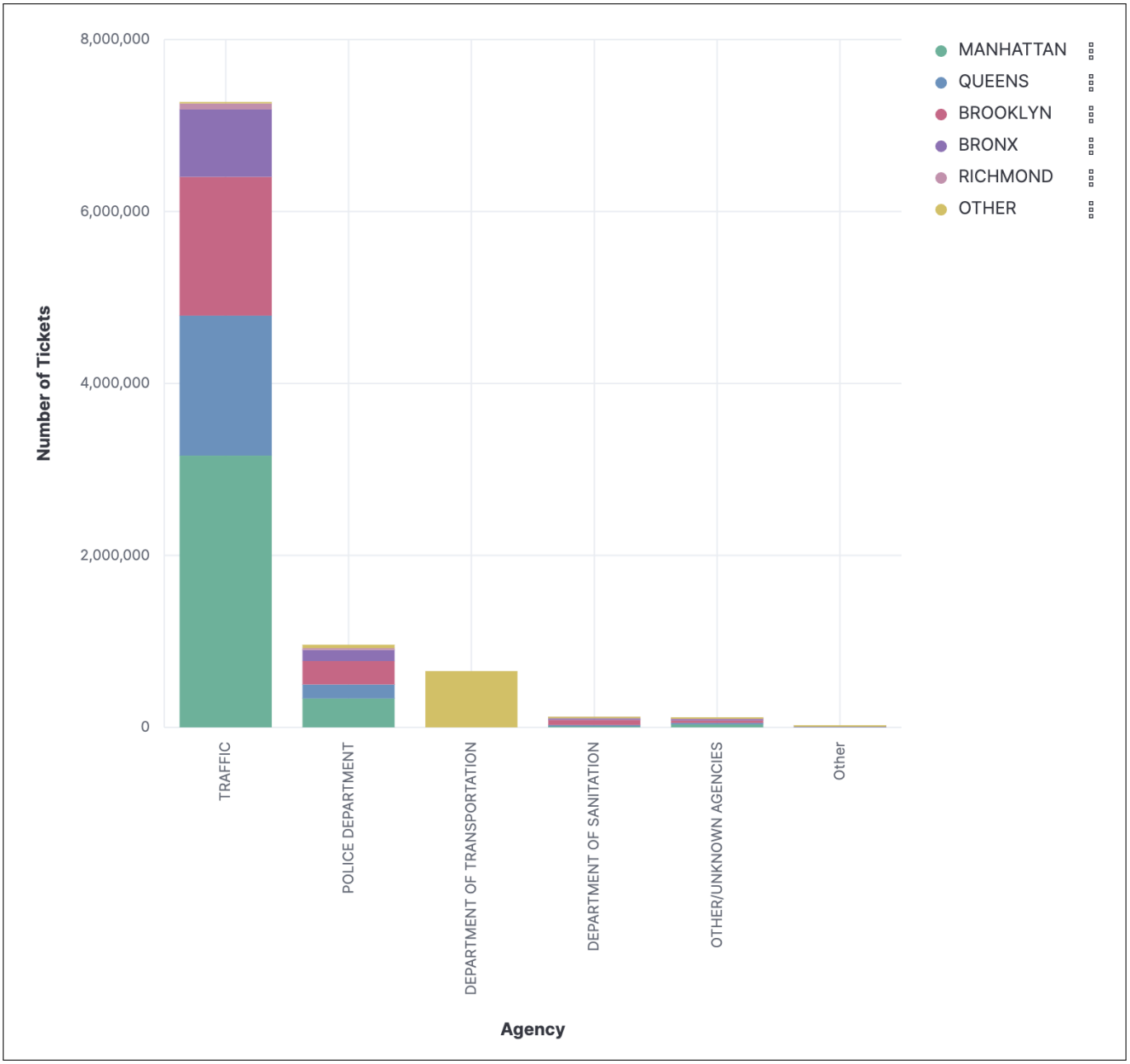

Figure 10: Issued Tickets per agency and county