DATA WRANGLING

Exploring New York City's 311 Service Requests

Author 1
Niklas Lystrup Poulsen
Vrije Universiteit Amsterdam

Author 2
Mafalda de Macedo Pinto Candal
Vrije Universiteit Amsterdam

I.Introduction

New York City never sleeps, and neither do its complaints. From noise complaints to potholes, illegal parking to power outages, New York's 311 system serves as the primary access-point to non-emergency city services and information about government programs (*NYC311 · NYC311*, n.d.). It has done so since 2003 and, in 2016, NYC311 received over 35 million requests for services and information through its phone number, mobile app, and social media (Kontokosta et al., n.d.). The system covers multiple agencies, each handling specific types of complaints - for example, the New York Police Department responds to illegal fireworks, while the Department of Health and Mental Hygiene addresses food poisoning incidents at restaurants. Technological progress since the service's inception has now allowed the city to harness the sheer volume of collective information running through these requests, prompting a data-driven approach that iteratively improves services and uncovers essential insights into public needs, resource allocation, and patterns of city life. The last complete year of data is 2024, and it constitutes the best starting point for a brief descriptive and exploratory analysis of New York City's 311 Service Requests.

MRQ What underlying trends can be identified in the NY 311 service requests for 2024?

RQ1 What were the most common types of service requests in 2024, and how did their frequency vary throughout the year?
RQ2 How were service requests distributed among the responding agencies?
RQ3 How did the placement of service requests differ across boroughs?
RQ4 Are there specific times of day or days of week when certain complaint types are more common?

II.Data Sources - Where did we get our data from?

Originally, we were interested in doing our project on the newly established Congestion Pricing Program for traffic in NYC. We wanted to uncover how this initiative had so far impacted the traffic and subsequently the city. We didn't end up doing this as it's a fairly new program and data is therefore sparse. With New York City in mind, we ended up looking elsewhere for data on the city and came across the city's very own data platform, NYC OpenData, where New York City's agencies and other partners publish free public data on just about anything in the city. We chose to delve into the NYC311 Service Request dataset as it was extensive enough to investigate a wide range of day-to-day issues faced by residents and to uncover deeper insights into how a high-energy metropolis combats these issues with a unique service like 311. To accompany and support one of the analysis steps in RQ3, we also incorporated population census data from 2022, available on the website population.de (population.de, n.d.).

The dataset contains data leveraged from 3,458,285 service requests over the course of 2024. Among the original 41 features, we selected only those that were of utmost importance to our main research question, enabling us to keep the project manageable in both scope and scale with respect to the trends under investigation. Specifically, our final set of features includes: Unique Key, Created Date, Closed Date, Status, Agency, Agency Name, Complaint Type, City, Borough and Location Type.

III.Pre-Processing stage

1. Data collection

The first stage of our analysis was uploading the data to our jupyter notebook. NYC OpenData offers the possibility to add filters (like choosing a timeframe or a specific borough of interest) and choose columns through the platform. Additionally, it is possible to download the data (either as a csv, xml,...) or to obtain an API endpoint. Initially, we attempted to use the web API, but this led to issues related to a limit of 1000 rows imposed and the overall size of the dataset. After restricting the amount of columns we wished to keep, we opted to download the data as a csv file and avoid crowding the jupyter notebook.

2. Data cleaning

Once the data was successfully uploaded, we faced the data preprocessing stage. We started by using commands like df.head() and df.info() to get an overview of the structure of the dataset. We then chose to move some columns for better organisation and moved on to formatting and handling missing data. Every service request is attributed a 'Unique Key', so we started by checking that all the values under that column were unique, meaning there were no undesirable duplicates. We then computed the count of null values per column. There were only null values for three columns (‘Closed Date’, ‘City’ and ‘Location Type’), all in the order of hundreds of thousands. This led to the decision to delete the columns ‘City’ and ‘Location Type’, whose incompleteness made them a worse option than analysing by ‘Borough’.

In order to ensure smooth analysis, we converted the 'Created Date' and 'Closed Date' columns to datetime format and made sure the data covered the entirety of the year 2024 as we intended. We first extract the 'Created Date' of the first and last rows to make sure that these corresponded to the first and last seconds of the year, and we then counted the number of rows per month to confirm that the data was roughly well distributed throughout the year.

Lastly, we decided to analyse the unique values for the ‘Complaint Type’ column. Firstly, we noticed that there was little uniformity in case, so we converted all the entries to mixed case, and standardised disjunctive complaint types by replacing all ‘ or ’ by ‘/’. We also realised that at least a few types had subcategories signalled by a ‘ - ’, e.g. ‘Noise - Residential’, ‘Noise - Commercial’. For this reason, we printed all values with ‘ - ’ and concluded that this was the case for 4 main categories. We opted to generalise these 4 main categories by deleting the specifications and keeping only the prefixe. Once this step was complete, we were ready to tackle our research questions.

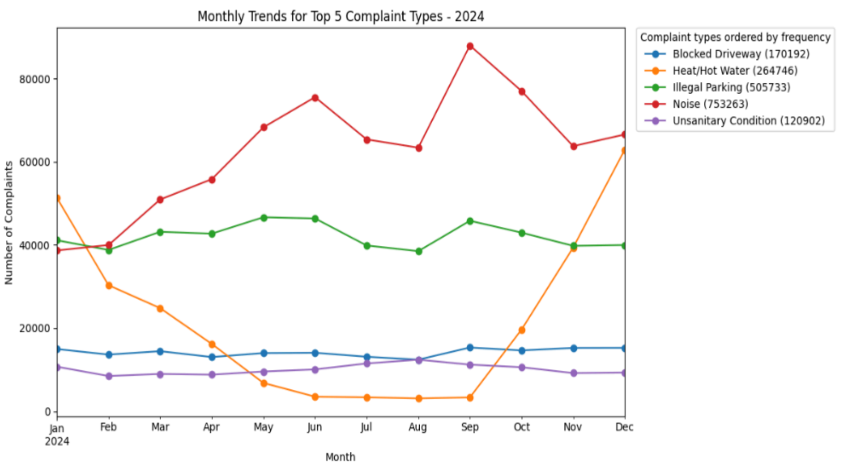
IV. Processing stage and findings

3. Data aggregation and visualisation

RQ1 What were the most common types of service requests in 2024, and how did their frequency vary throughout the year?

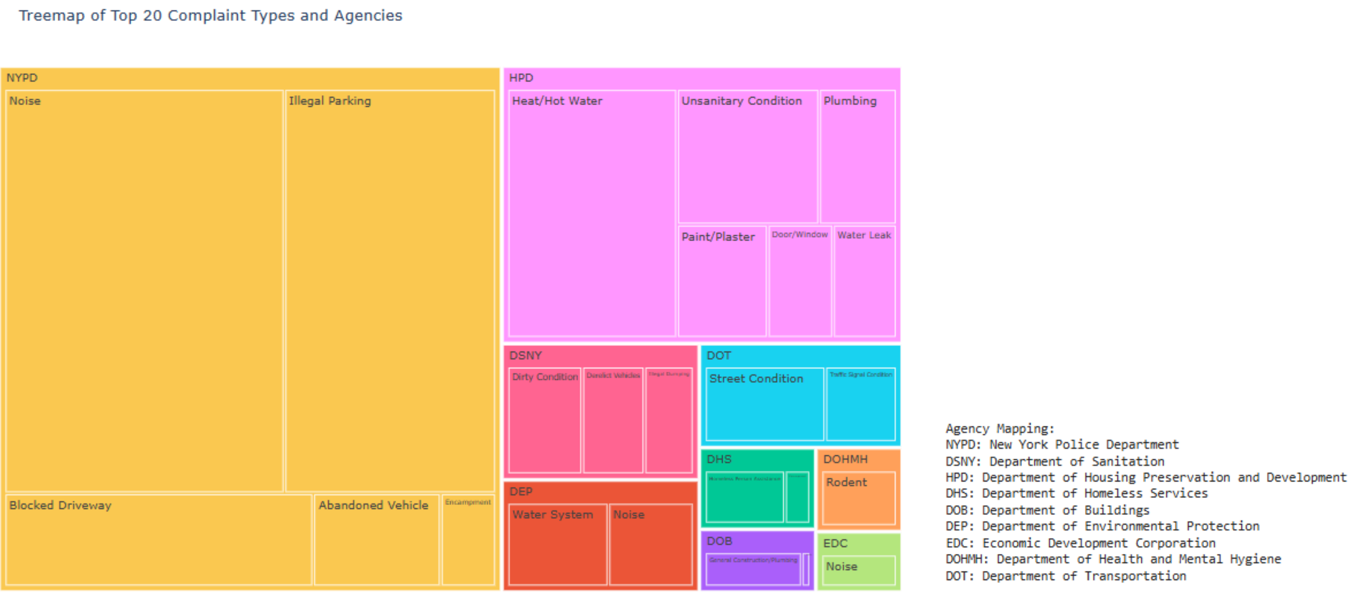
This question focuses on the most common types of service request. We started by exploring the frequency of each complaint type. But we further wanted to understand how these frequencies varied throughout the year, so we opted for a monthly analysis. We grouped and counted the number of entries by 'Month' and 'Complaint Type'. We agreed that this question would be well-answered with a **line chart** which helped understand the monthly evolution of the 5 most common complaint types (we attempted 10, but it already jeopardised readability).

This graph indicates that from these, only noise and heat/hot water complaints suffer major fluctuations throughout the year. Noise complaints seem to increase with the arrival of good weather, then have a decline consistent with the most popular vacation months and see a rapid increase with the beginning of the school year, lowering steadily until the winter holidays. The heat/hot water complaints seem to have a strong reverse correlation with warm weather, having a peak at the start of winter and then lowering until they stabilise during the summer months, having a rapid increase from September onwards. The remaining categories appear to have a roughly steady frequency throughout the year.



RQ2 How were service requests distributed among the responding agencies?

This question focuses on the role of the different agencies. There are two similar columns in this dataset: ‘Agency’ for the initials of the agency and ‘Agency Name’ for the full name of the entity. For simplicity and reliability, we opted to use the ‘Agency’ column. Somewhat similarly to the previous RQ, we started by grouping and counting the number of entries by 'Complaint Type' and 'Agency'. We then filtered the DataFrame to include only the top 20 most frequent complaint types and colour-coded treemap for visualization. The treemap is interactive to allow us to explore the various complaints and precise values. This treemap featured 9 agencies, and 24 rectangles, because while many complaint types are mostly specific to a singular agency, others like noise, encampments, and plumbing are addressed by multiple agencies. NYPD responded to over half of the top 20 most frequent complaints, followed by HPD with roughly a fourth, and DSNY with roughly a sixteenth. In general, the complaint types each agency answered to had a clear connection or similarity, possibly except for the DEP (Department of Environmental Protection) which answered to Water System complaints, but also a sizable portion of the noise complaints.

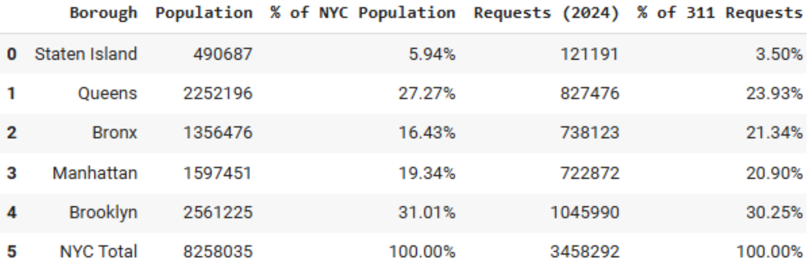


Treemap of Top 20 Complaint Types and Agencies

Agency Mapping:
NYPD: New York Police Department
DSNY: Department of Sanitation
HPD: Department of Housing Preservation and Development
DHS: Department of Homeless Services
DOB: Department of Buildings
DEP: Department of Environmental Protection
EDC: Economic Development Corporation
DOHMH: Department of Health and Mental Hygiene
DOT: Department of Transportation

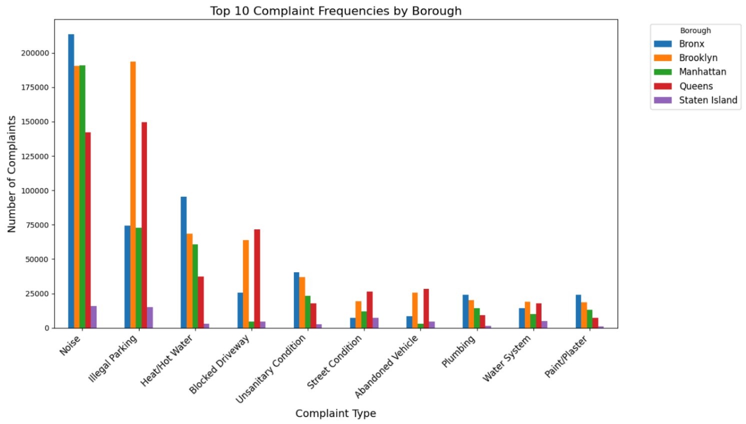
RQ3 How did the placement of service requests differ across boroughs?

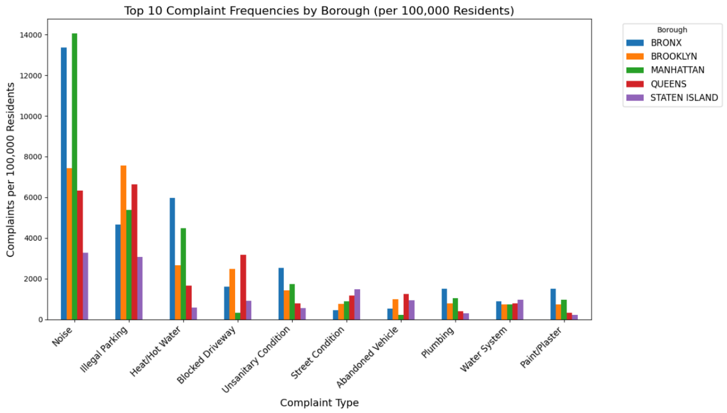
This question explores the spatial component of this dataset, by focusing on boroughs as units of analysis. We started by printing the unique values for the column ‘Boroughs’ to better understand the data saved in it. Following this, we counted the total number of service requests by borough. Although the differences were large, the number of requests by borough correlated heavily with the population of that said borough, as expected. A larger population within a borough resulted in a higher number of requests. The correlation is shown in our table below where we calculated each borough’s population as a percentage against NYC’s total population. We also calculated the percentage of 311 requests for each borough. Ultimately, we aimed to determine which boroughs deviated from the expected trend—identifying those that submit a disproportionately high or low number of requests relative to their share of New York City's population. The population census data is retrieved through the website city population.de and the population numbers are from 2022.



	Borough	Population	% of NYC Population	Requests (2024)	% of 311 Requests
0	Staten Island	490687	5.94%	121191	3.50%
1	Queens	2252196	27.27%	827476	23.93%
2	Bronx	1356476	16.43%	738123	21.34%
3	Manhattan	1597451	19.34%	722872	20.90%
4	Brooklyn	2561225	31.01%	1045990	30.25%
5	NYC Total	8258035	100.00%	3458292	100.00%

With this overview in mind, we can now dive a bit deeper and look at how the top 10 complaint types are divided across the boroughs throughout 2024, instead of simply looking at ALL service requests like we did in the previous section. Below, our visualizations communicate this in two different ways: absolute complaint counts (left) versus per-capita normalization (right). On the left-hand chart raw complaint totals are displayed, revealing Bronx and Manhattan leading in almost every category. For example, both boroughs are extremely noisy, likely due to their high population. However, there is a counterpoint to this: Queens is much less noisy but has about the same population count as Manhattan and Brooklyn. In this visualization, the smaller boroughs are represented more clearly. Even though this visualization shows the per-capita differences, some of the same patterns still emerge, at least for large categories such as “illegal parking” and “noise”. Staten Island climbs this visualization slightly more in the noisy category but is still quite low compared to the other boroughs. The reason for this might be because the population density in the borough is much lower compared to the other boroughs. One would assume that pop. density is a big factor in the reporting of noisiness and this is simply a variable that is not taken into account in these visualizations, therefore leading to these discrepancies even in the per-capita visualization. For some of the smaller complaint categories such as “street condition” and “water system”, Staten Island is on par with the other boroughs and even leads in these two specific complaint categories. Another interesting take away from the per-capita visualization is that the Bronx, being the 2nd smallest borough and appearing as less dominant in raw numbers, now show some of the highest - if not the highest - per-capita complaint rates, indicating that residents in these boroughs submit 311 requests at a disproportionately high rate. The “small” underlying factors, such as population density, remain hidden within the data but become more apparent through a per-capita lens. Given the limitations of this paper, an exploration of more assumptions and contributing factors is beyond reach.



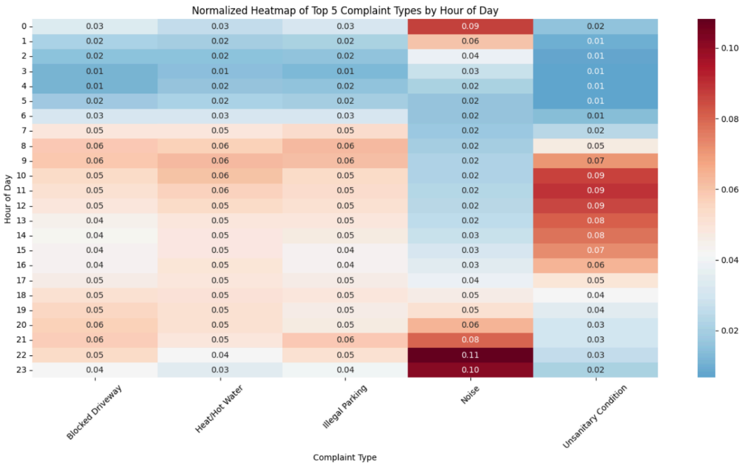


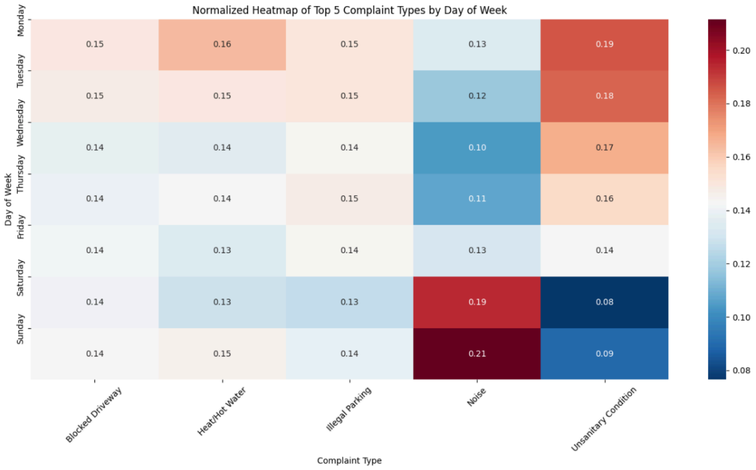
RQ4 Are there specific times of day or days of week when certain complaint types are more common?

Having concluded our section on spatial analysis, we move on to temporal analysis, exploring trends at specific times of day and days of week. We started by creating a pivot table with counts for each complaint type by hour. But finding trends in such big values is challenging, so we decided to normalise the counts, by dividing by the total count for that complaint type. This way, we had a unit that focused more on temporal trends per complaint type than absolute values. We opted to create two heatmaps to visualise this data and used a colour ramp with white (neutral) for the average value (1/24 ~ 0.042 for time of day and 1/7 ~ 0.143 for days of week), red for higher values, and blue for lower.

Focusing first on the time of day trends, we conclude that the 'Noise' column stands out with a peak from 8pm to 2am (consistent with nightlife events) and residual complaints for the remaining time, while the remaining columns show residual complaints during the night. 'Blocked Driveway' and 'Illegal Parking' have soft peaks morning and afternoon, when many people are attempting to leave or enter their residences. For 'Heat/Hot Water', there seems to exist a steady inflow of complaints throughout the day, and for 'Unsanitary Condition' there are more complaints during the usual 9-5 schedule.

Regarding the day of week trends, the first three columns show little variation, achieving a max of 0.03 variation, with a slight skew towards the beginning of the week. The last two columns have much clearer and, interestingly, opposing trends. 'Noise' complaints are much more frequent during the weekend, and have its lowest on Wednesday, having slightly higher values towards the beginning and end of the working week. Conversely, 'Unsanitary Condition' complaints are not frequent during the weekend, and start the weekend at its highest, with a soft decline throughout the working week.





V.Limitations

The analysis focuses on the top 5, 10, or 20 complaint types, which avoids insights from less frequent data categories. Additionally, using only one year of data restricts the ability to generalize trends over time, and therefore limits the usefulness of these trends. Analyzing borough-level spatial data lacks the granularity needed for more detailed insights into localized patterns. Furthermore, the repetition of similar code throughout the analysis could have been streamlined using reusable snippets, such as defining top complaints and filtering data.

VI.Conclusion

New York’s 311 system serves as the primary access-point to non-emergency city services, covering 147 unique complaint types and multiple responding agencies. In 2024 alone, New York’s 311 recorded 3,458,285 service requests. Given the primary role of this platform, the NYC311 Service Request dataset poses as a remarkably complete and reliable source to analyse essential trends in public needs, resource allocation for agencies, and patterns of city life. The main research question that this report hoped to answer was: *What underlying trends can be identified in the NY 311 service requests for 2024?* With limited resources and skills, we opted to tackle 4 sub-questions which focused on 1) the types and frequency of service requests, 2) the responding agencies, 3) brief spatial analysis by boroughs, 4) brief temporal analysis by times of day and days of week.

It appears fundamental to summarise some of the main findings. The 5 most frequent complaint types are ‘Blocked Driveway’, ‘Heat/Hot Water’, ‘Illegal Parking’, ‘Noise’ and ‘Unsanitary Condition’. Noise complaints rise with warm weather, dip during peak vacation months, surge at the start of the school year, and decline towards winter holidays. Heat/hot water complaints peak in early winter, decrease through summer, and rise again from September onward. These findings also highlighted the role of different agencies in addressing complaints, with the NYPD handling a significant share of noise-related and other public disturbance reports, while the HPD managed most heat/hot water concerns. Spatially, complaint distribution was closely aligned with borough population sizes, but per-capita analysis revealed that certain boroughs, like the Bronx, reported disproportionately high complaint rates. When it comes to temporal trends, complaints such as noise peaked during late-night hours and weekends, while others, like illegal parking, followed commuting patterns.

Ultimately, this analysis underscores the value of the NYC311 system not only as a civic engagement tool but also as a rich dataset for identifying urban patterns and guiding resource allocation. Future studies could deepen this exploration by incorporating additional variables such as weather conditions, socioeconomic data, or response times to further refine insights into public service needs and city management efficiency.

References

Kontokosta, C. E., Weiss, M., Snively, C., Gulick, S., & Weiss, M. B. (n.d.). *NYC311 - Case - Faculty & Research*. Harvard Business School. Retrieved February 2, 2025, from https://www.hbs.edu/faculty/Pages/item.aspx?num=53497

NYC311 · NYC311. (n.d.). NYC.gov. Retrieved February 2, 2025, from https://portal.311.nyc.gov/article/?kanumber=KA-02498

population.de. (n.d.). *New York City Boroughs (USA): Boroughs - Population Statistics, Charts and Map*. City Population. Retrieved February 2, 2025, from https://www.citypopulation.de/en/usa/newyorkcity/