

UNIVERSITY OF SOUTHERN DENMARK

TEKNISK FAKULTET

MÆRSK MC-KINNEY MØLLER INSTITUTTET

NYC Traffic Enforcement: A Data Visualization Representation

Authors:

Sigurd Vind	sivin20@student.sdu.dk
Andreas Gade	angad20@student.sdu.dk
Lauge Solvang	lasol21@student.sdu.dk
Asma Khatun	askha21@student.sdu.dk

Exam nr.

493836
494221
4110095
4104409

Advisor:

Parisa Niloofar	parni@mmmi.sdu.dk
-----------------	-------------------

Semester:	E23
Course code:	T520040102
Hand-in:	28-12-2023

1 Background & Motivation

1.1 Exploring Patterns and Trends

Our primary objective is to uncover significant patterns and trends within the extensive [New York City Parking Tickets dataset](#). Through meticulous data analysis, we aim to extract insights that extend beyond numerical values, offering a nuanced understanding of parking violations dynamics within the city. This exploration seeks to reveal not only statistical correlations but also the contextual relevance of the data.

1.2 Rationale for Dataset Selection

The selection of the [New York City Parking Tickets dataset](#) is deliberate and rooted in its real-world relevance, inherent complexity, and alignment with our academic course objectives. This strategic choice serves as a conduit for students to bridge theoretical knowledge with practical application, fostering a profound comprehension of data analysis methodologies and their tangible implications in real-world scenarios.

1.3 Project Inspiration: A Fresh Perspective

Our project, focused on visualizing parking violations data, stems from a blend of interests and the ambition to explore a dataset often overlooked in the data visualization realm. We aim to move away from commonly used datasets that often lead to repetitive themes and insights. By delving into parking violations, we venture into a less-trodden domain, aiming to uncover new perspectives and visualize fresh insights.

1.4 The Rich Dynamics of Parking Violations

Parking violations present a compelling mix of legal, societal, and urban dynamics, offering a rich area for exploration. The dataset's inclusion of temporal and geographical variables adds depth to our project. The temporal dimension allows us to create dynamic visualizations that reveal patterns and trends across different times of day, days of the week, and even seasonal variations. Similarly, the geographical aspect enables spatial analysis, aiding in identifying violation hotspots and understanding geographical distributions. The dataset's comprehensive categorical variables also offer a plethora of opportunities for diverse visualizations.

1.5 Technical Challenges and Learning Opportunities

Furthermore, the technical challenge of processing and visualizing a complex dataset with variables like time and location is an attractive aspect of this project. It offers us an opportunity to hone our data processing and visualization skills, contributing to our growth in this field.

1.6 Academic Significance

The academic significance of our analytical pursuits extends beyond routine data exploration. This endeavor seamlessly integrates theoretical concepts with practical application, providing a holistic view of the subject matter. The outcomes of our analysis not only enrich the collective body of academic knowledge but also provide cross-disciplinary insights into urban studies, transportation

management, and policy implications. Our exploration of the [New York City Parking Tickets dataset](#) contributes to the evolving discourse within our academic community.

1.7 Educational Value of Data Visualization

Moreover, the incorporation of data visualization into our analytical toolkit elevates the educational value of this report. It transcends traditional boundaries by providing a visually immersive experience. The utilization of various visualization tools not only enhances the comprehension of complex data structures but also cultivates essential skills in communication. As students navigate the intricate landscape of visual representation, they are not merely analyzing data but crafting narratives that bridge the gap between raw information and actionable insights.

1.8 Preparing for Data-Driven Environments

This approach, rooted in the fusion of theoretical knowledge with hands-on experience, underscores the paramount importance of data literacy in the context of future data-driven environments. By fostering an interactive and collaborative learning experience, we aim not only to support multi-perspective analysis but also to instill a passion for data-driven inquiry among students. In essence, this report signifies a journey that extends beyond the confines of academia—a holistic educational experience equipping students with the skills and mindset necessary for the complex landscapes of data analysis and visualization in the real world.

2 Project Objective

The Project visualizes graphical demonstration of important analysis and relationships among the variables of Parking violation dataset. The aim of the project is to find useful information from the dataset and apply some visualization tools to represent graphical views which can be used in future.

2.1 Questions

By analyzing the dataset, a set of questions has been found which convey almost all the information about New York parking violations.

- What is the distribution of violations in New York, and where are tickets most issued?
- When do parking violations mostly occur?
- Which types of parking violations are issued the most?
- Are there any seasonal or weekly patterns in the issuing of parking violations?
- What were the top five violation counties? What were the top five vehicle types with the highest percentages of violation for each county?
- Which brands of vehicles got the most parking tickets?
- What were the top ten parking violations in New York city in 2016?
- How did the distribution of parking violations change over the course of the year of 2016?

2.2 Objective

The objective of the project is learning and demonstrating the valuable insights of Parking violation dataset. The project provides a better understanding of geographical distributions of parking violations, pick hour of getting parking tickets, categorizing the violation types, vehicle types, vehicle body types, vehicle making brands and so on. Here, we come up with all the useful details of the dataset and represent as graphs.

3 Data

3.1 Data Source

Our data was sourced from a comprehensive dataset available on Kaggle, specifically the “NYC Parking Tickets” dataset for the fiscal year 2016. This dataset, titled *Parking_Violations_Issued_-_Fiscal_Year_2016.csv*, can be accessed at Kaggle: NYC Parking Tickets Dataset.

3.2 Data Description

We use 9 different variables that offer insight into parking violations across New York City for the specified period. These variables include:

- **Violation.Code:** A numeric identifier for each type of parking violation, such as parking in a no-parking zone or at an expired meter.
- **Violation.Time:** The specific time the parking violation was recorded, formatted in a 12-hour clock system.
- **Issue.Date:** The date of issuance for the parking violation ticket, following a month/day/year format.
- **Type:** Classification of the vehicle’s plate, indicating whether it is for passenger, commercial use, etc.
- **VehicleMake:** Brand or manufacturer of the vehicle that was issued the parking violation, such as “TOYOT” for Toyota or “HONDA” for Honda.
- **VehicleBodyType:** The type of vehicle body, categorized as “SUBN” for SUV/Suburban, “4DSD” for a four-door sedan, “VAN” for a van, etc.
- **County:** The abbreviated name of the county where the violation occurred, such as “Q” for Queens, “K” for Brooklyn, or “NY” for Manhattan.
- **Violation.Precinct:** The precinct number of the police department responsible for the area where the violation took place.

The dataset is extensive, encompassing 10 million records and occupying a size of 2.15 GB.

3.3 Data Processing

The initial dataset comprising 10 million records was sampled down to 200,000 entries using a custom Python script. This random sampling was necessary to manage the dataset’s size for more efficient analysis. The processing stage required minimal cleanup, predominantly involving the correction of a few records with incorrect date and time formats. The majority of the dataset was used as-is, without the need for extensive cleaning or reformatting.

4 Visualization

4.1 Heatmap

The heatmap in Figure 1 serves as an effective visualization for displaying parking violations data over time. It maps the frequency of violations across weeks on the x-axis and days of the week on the y-axis, using color intensity to indicate the number of violations. This design allows for easy identification of temporal trends, such as weekly cycles and seasonal changes. The use of the viridis color palette ensures the map is accessible and interpretable, with darker colors signaling more violations. Interactivity, enabled by plotly, allows users to hover over cells for detailed information, enhancing the utility of the heatmap. Special dates are also annotated, providing context for any deviations in the pattern due to holidays or weather events. Overall, the heatmap offers a clear and interactive way to analyze the distribution of parking violations over time.

Seasonal Rhythms and Weekly Patterns in Parking Violations

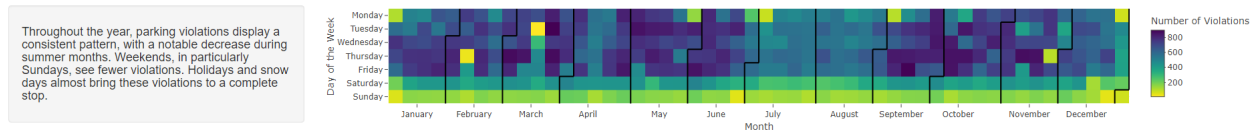


Figure 1: Heatmap Visualizing Parking Violations by Day and Week

4.2 Categories by Hour

The bar chart visualization in Figure 2 represents parking violations over the course of a day by category. The x-axis represents different times of the day, rounded to the nearest half-hour, and the y-axis shows either the count or the percentage of violations, based on user selection. Users can choose to view absolute numbers or proportional data, and they can filter by violation categories through a dropdown menu. Each bar is color-coded to represent different categories of violations, providing a clear visual distinction.

The Morning Rush

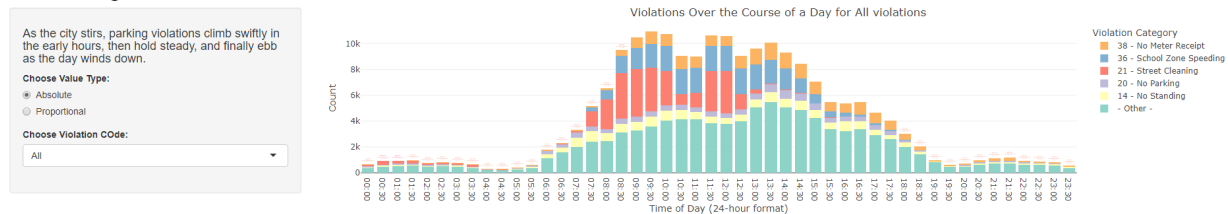


Figure 2: Hourly Distribution of Parking Violations by Category

The chart is interactive, with hover functionality that reveals more details about each time slot and violation category. This enables users to glean specific insights about the frequency and distribution of parking violations throughout the day. By stacking the bars, the chart also allows for a comparison of the relative proportions of each violation category at each time point.

The Color of Compliance: Painting New York's Parking Landscape

Dense Parking Violations in Lower and Midtown Manhattan Spare Central Park

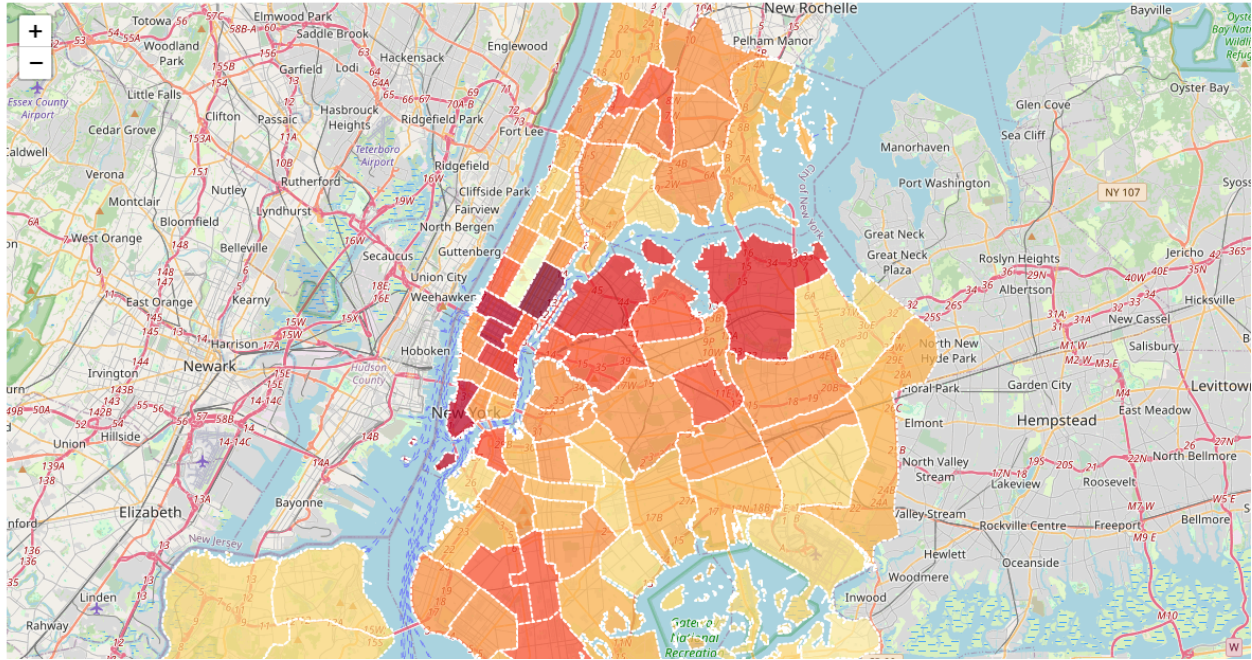


Figure 3: Map of Parking Violations by Precinct

4.3 Violations Map

The interactive map in Figure 3, created using the leaflet library, presents a geospatial visualization of parking violations across New York precincts. The map is a choropleth that assigns colors to precincts based on the number of parking violations, with the color intensity reflecting the magnitude of violations. This visual representation employs the viridis color palette.

A key feature of the map is its interactivity; users can click on individual precincts to receive information about the number of parking violations in that precinct. Highlighting functionality enhances this interaction by outlining the selected precinct, making the user experience more intuitive.

To facilitate clearer distinctions between precincts with varying violation counts, the application uses a root transformation for the color scale breaks. This approach linearizes the distribution of violation counts, allowing for more nuanced visual differentiation between areas with small differences in violations.

4.4 Violations by Vehicle Plate Type

Asma

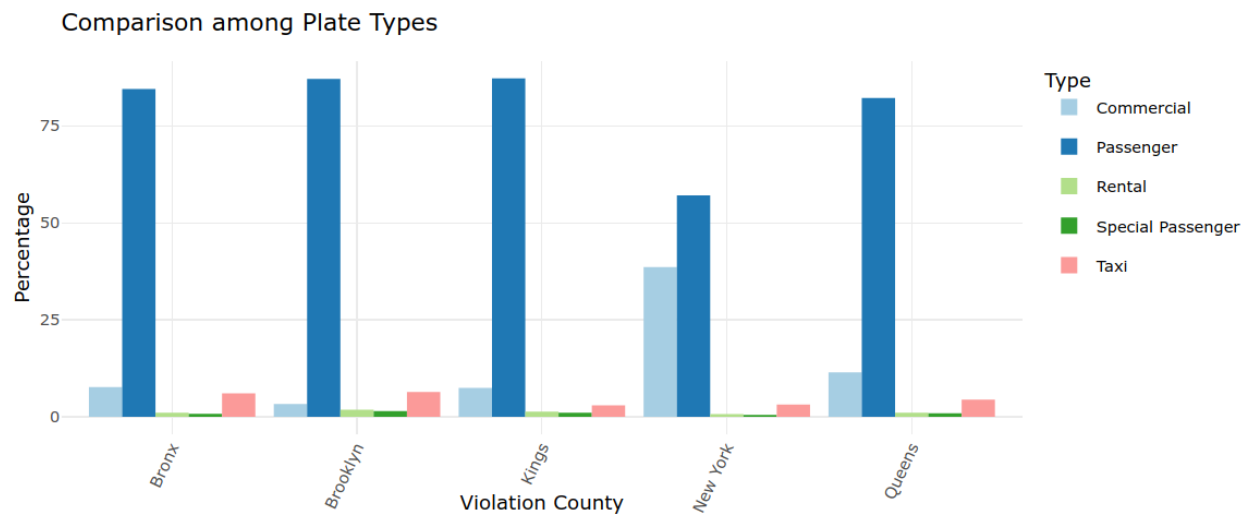


Figure 4: Caption

4.5 Violations by Vehicle Make

The bar chart visualized in Figure 5 shows the distribution of parking violations by vehicle make. The bar is interactive, allowing the user to choose whether to see it ascending or descending, which increases the readability of the graph. It was made with plotly, and also features a hover functionality which allows you to see the exact percentage that that brand of vehicles makes out of the total amount of vehicles that got parking tickets in 2016.

Distribution of Parking Violations by Vehicle Make

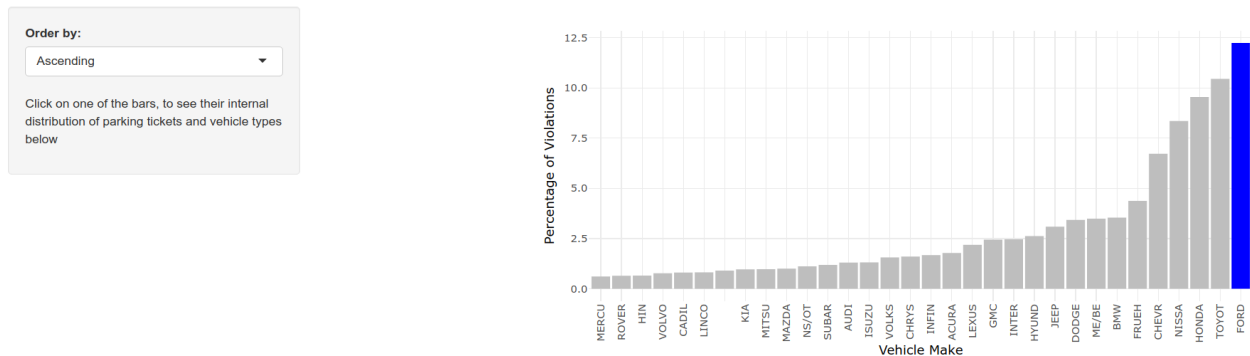


Figure 5: Distribution of parking violations by vehicle make

Additionally the bars on the graph are clickable which changes the data of the pie chart and the donut chart that is placed below, as seen in Figures 7 & 6. The bar graph features a popout, as clicking on e.g. "TOYOT" changes the color of the corresponding bar in the bar graph, and then lets you see the body type distribution and most likely parking tickets for that brand of vehicles.

The pie chart as previously mentioned and seen in Figure 6, shows the distribution of vehicle body types. This distribution has been translated from shorthand abbreviations to real text, to further increase the readability of the data. I.e. the abbreviation "4DSD" means "4 Door Sedan",

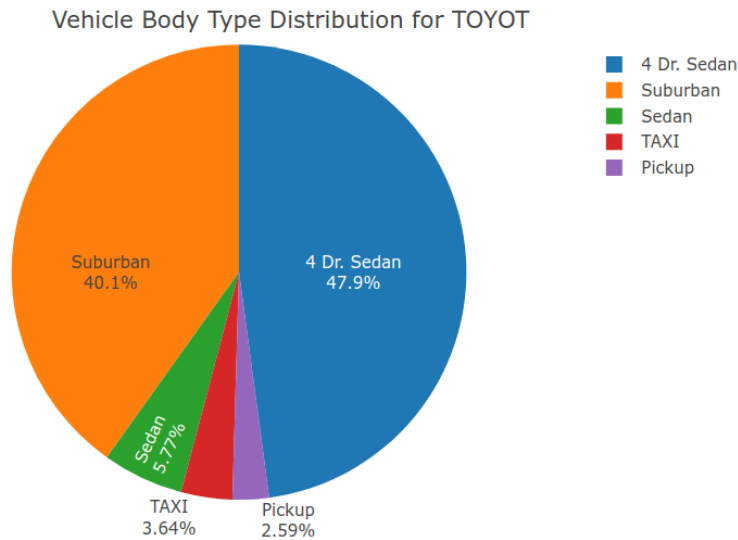


Figure 6: Vehicle body type distribution

and as an example is responsible for about 48% of parking violations for Toyotas in New York.

The donut chart, as previously mentioned and seen in Figure 7, shows what ticket a given vehicle make is most likely to get based on the dataset used. Here it uses the violation codes combined with the violation description to be more descriptive in which violation was given. As mentioned earlier in section

Andreas should write something about his donut graph?

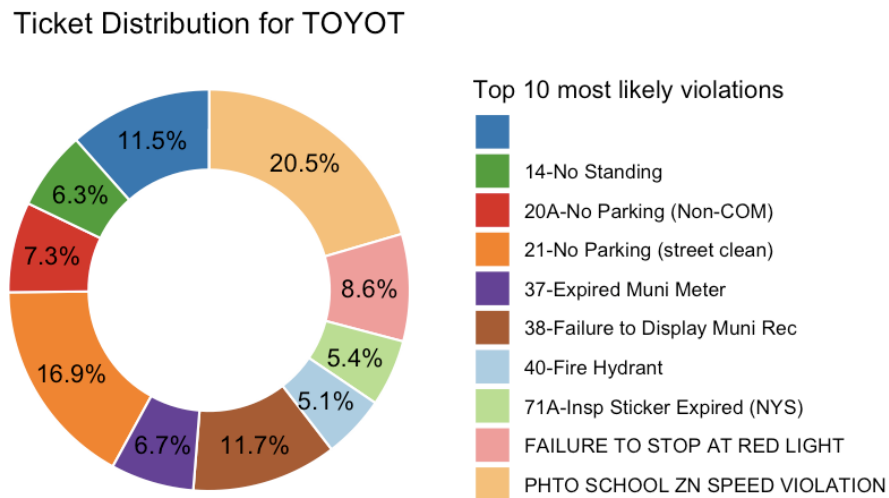


Figure 7: Donut chart for violations by vehicle make

4.6 Violation Trends over time

The use of an animated graph, as seen in Figure 8 adds an interactive and dynamic element to the visualization, making it more engaging and visually appealing for the users. It allows them to see the trend in violations over time in a more dynamic and intuitive way, enabling them to easily identify patterns, fluctuations, and any significant changes in the data.

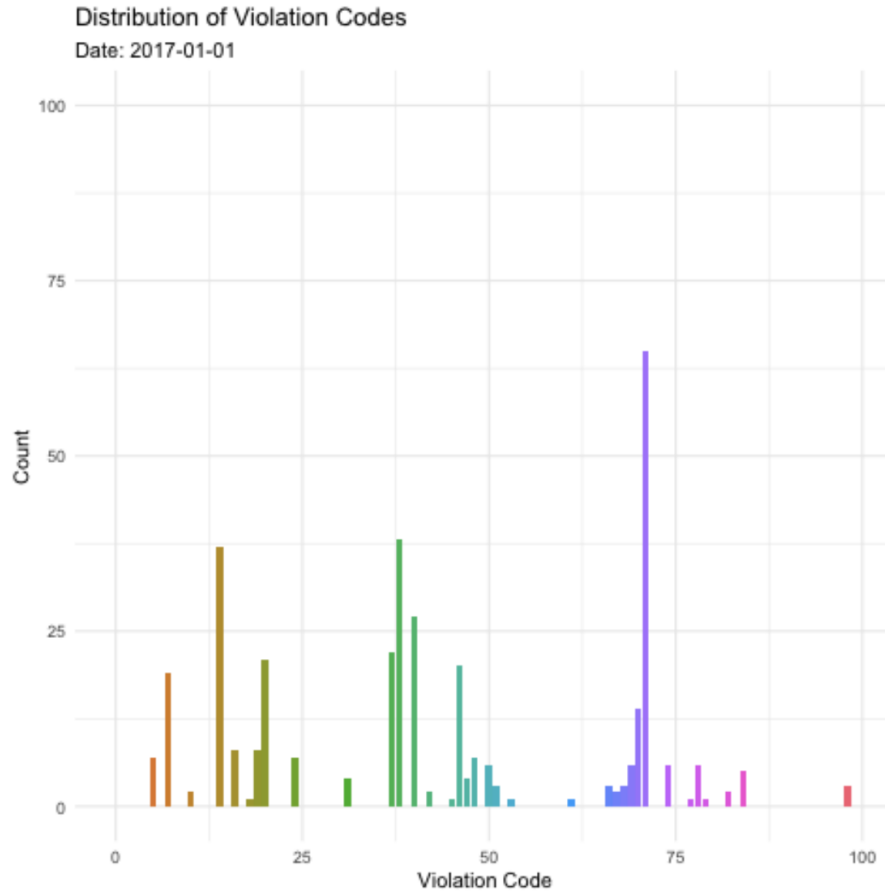


Figure 8: Animated Bar chart for Violation Trends

It was created by mapping violations to a bar chart that reflects the amount of tickets issued for a certain violation code per day. A temporal factor was then appended onto the bar chart to allow for a visual representation of the magnitude of violations over the time span that the dataset spans on.

5 Story & Results

This section aims to answer the questions presented in the in Section 2: Project Objective.

5.1 Addressing the project objectives

5.1.1 What is the distribution of violations in New York, and where are tickets most issued?

The interactive map clearly shows the distribution of parking violations around New York. Manhattan is shown as the place with the most parking violations, which we expected as Manhattan is also the most densely populated area of New York. Interestingly, specifically the area east of Central Park called Park Avenue has significantly more parking violations than the rest of the precincts in New York. We expected that the distribution was more even throughout Manhattan, and that the further from Manhattan you got, the less parking violations were present. This somewhat shows in the map, however Staten Island has less than the average amount of parking violations, which follows the trend of the population, as it has a much smaller amount of residents compared to Manhattan, Brooklyn, Queens, etc.

5.1.2 When do parking violations mostly occur, and which types of parking violations are issued the most?

Generally parking violations mostly happens between 0600 and 1900. The trend also shows that parking violations climb quickly during the morning, which might be an indication of the New York morning rush, where parents have to drop off their children at institutions and people have to hurry to work. It then falls a bit again from 10-11:30, which seems to be because Street Cleaning tickets are less issued. This might be because fewer street cleanings are scheduled in this timeslot.

Another big contributor to why there are more parking tickets between 08 and 15 is School Zone Speeding, which is something we expected as the schools are open in this timespan. The distribution of "No Parking", "No Standing" and "No Meter Receipt" tickets are somewhat evenly distributed in the hours that people are awake, which is also something we expected.

Another interesting fact is that you can easily connect outliers with either weather events like snowstorms or holidays like Memorial Day, Independence Day, etc. This is most likely because people are off work these days, meaning they do not drive to work or school.

5.1.3 Are there any seasonal or weekly patterns in the issuing of parking violations?

The heatmap of parking violations clearly shows seasonal and weekly rhythms in the issuance of parking violations. Saturdays and Sundays have a significantly lower amount of parking violations than the rest of the weekdays, which is expected as people are not working or at school during the weekends, but are much more likely to stay at home.

Additionally the data also shows that parking violations were much less present during the summer months of July and August, as well as the last days of December. This might be an indication of people going on holidays, which would mean that a lot more people travel out of New York for summer vacations or go home to their parents during Christmas. The summer months are always the months where most tourists travel to New York, but as the majority of tourists most likely travel to New York for leisure, many of them are more likely to take a taxi than to have brought their own cars or rented cars.

5.1.4 What were the top five violation counties? What were the top five vehicle types with the highest percentages of violation for each county?

Asma

5.1.5 Which brands of vehicles got the most parking tickets?

The distribution of which vehicle makes got the most parking violations in 2016 somewhat follows the expectations of the group to the dataset. Brands like Ford, Toyota, Honda, Nissan, Chevrolet were among the brands which were responsible for most parking violations, which follows the trend of most popular car brands in America. Sadly no data from 2016 was able to be procured, but as shown in the article by Carlogos [1], the top 5 vehicle makes in 2021 included exactly the 5 aforementioned brands. Interestingly Fruehauf is the 6th most responsible for parking violations, which is a company that sells both semi-trailers and trailers for trucks. By clicking on it, and viewing the distribution of vehicle body types, you can see that these are mostly Delivery trucks and Vans, which makes sense as they are also susceptible to break the parking laws.

5.1.6 How did the distribution of parking violations change over the course of the year of 2016?

Andreas

5.2 Further improvements

6 Discussion & Conclusion

References

- [1] “Popular car brands.” [Online]. Available: <https://www.carlogos.org/popular-car-brands/>

A Group contributions

The group contributed evenly to the project and report, as shown in Table 1 below.

Student	Percent of Contribution
Sigurd Vind	100%
Andreas Gade	100%
Lauge Solvang	100%
Asma Khatun	100%

Table 1: Table of contribution