Analyzing Parking Violations in NYC: A Precinct-Wise Approach

Team 19

Spriha Bankata Mishra, McKenzie Campbell, Divij Mishra, Hilary Present, Saksham Sudershan, Spiros Valouxis

1 INTRODUCTION AND MOTIVATION

Urban areas face many pressing issues like traffic congestion and the efficient use of public space, making parking policies a pivotal factor in addressing these challenges. Inadequate policies and insufficient parking infrastructure exacerbate traffic congestion and lead to longer search times for available parking, profoundly affecting urban life.

Therefore, our research aims to address this underexplored area of study, analyzing the influence of factors such as precinct, time of day, time of year, and vehicle type on NYC parking behavior and subsequent parking violations. Our analysis aims to unravel the rationality behind these violations, shedding light on precinct-specific patterns and potential solutions.

The goal of our study is to better understand parking patterns, predict availability, and identify the key factors contributing to parking violations. By doing so, our research offers profound insights into the complexities of urban parking and opens doors for technology-driven improvements in the city. By merging diverse datasets, featuring precinct-focused analysis, time series forecasting, and interactive visualizations, we anticipate our findings will provide policymakers and researchers with the necessary foundation to comprehensively tackle urban parking challenges moving forward.

In summary, our research significantly contributes to urban planning and policymaking, offering a comprehensive analysis of parking violations in New York City. It provides valuable insights into the intricate realm of urban parking and lays the foundation for future research in this domain.

2 PROBLEM DEFINITION

Our problem statement comprises three parts. We aim to:

(1) perform a precinct-wise analysis of key determinants that lead to traffic violations and parking tickets in New York City,
(2) generate precinct-wise time-series forecasts for traffic violations, and
(3) create an informative and user-friendly interactive visualization to increase the accessibility of our findings.

3 LITERATURE REVIEW

Parking policies are essential for efficient urban planning and traffic management, influencing car usage [12, 14, 15], traffic congestion, safety [2, 3], and residential housing prices [14]. Confusing policies and inadequate parking contribute to escalated traffic congestion and increased time spent looking for parking [7, 13]. Surveys indicate that up to 45% of traffic flow in urban areas is attributed to drivers searching for available curb-side parking [7, 13].

Parking violation fines, or "tickets," serve as a significant revenue source for the city [12], despite often being only partially collected. In New York City, for instance, only slightly more than half of the fines are paid in full [11]. Additionally, a study in Freiburg, Germany [5] concluded that, from an economic perspective, it is rational for an individual to not pay parking fees approximately 86.1% of the time, meaning the likelihood of a parking violation getting detected is low. What can be done to remedy such rational rule-breaking?

To address this line of inquiry, researchers have investigated patterns in parking ticket data. Kim and Wang [9] explored specific attributes of individual vehicles that lead to double-parking citations and identified hotspots for double-parking incidents through geospatial analysis. Their study shows that commercial vehicles are more likely to be repeat offenders. Subsequent research has shown that illicit commercial parking, exacerbated by urban congestion, further contributes to the shortage of available parking spaces [18]. This study uses a weighted distance decay regression model to quantitatively assess the relationship between illegal commercial parking, parking supply, and demand, and their findings can be expanded to encompass personal vehicles as well.

Numerous studies also leverage machine learning in contexts relevant to our research. Jelen et al. [7] use methods such as random forest models to predict parking availability, concluding that the search for open spots can increase traffic congestion. They also suggest that in-ground parking sensors can help increase efficiency. Badii et al. [1] suggest that tracking sensors and parking apps improve efficiency further, and apply Bayesian regularized neural networks to identify available parking spots, notably in city garages in Florence and Tuscany. It is worth noting, however, that an increase in dedicated parking spots does not always prevent illegal parking [19]. Karantaglis et al. [8] demonstrated the prediction of parking violation rates through the use of deep residual neural networks. Gao and Ozbay [4] also use regression models to predict parking violations, but with a specific focus on double-parking frequencies. In another study, Gao et al. [16] use multiple regression models to predict parking violations based on NYC data from 2017 and 2018. Our study builds upon this by combining and utilizing larger and more recent data sets. Studies have also explored geospatial data to investigate correlations between certain neighborhoods, points of interest, and increased parking violations [10] [6]. Notably, points of interest – such as landmarks and lane width – are significant contributors to these patterns [20].

We anticipate that merging the NYC parking violation data set with the points of interest data, as well as NYC parking meter and signage data, will allow us to comprehensively cover parking demand hotspots. This will enable us to investigate the influence of fee pricing and clear signage on parking violations. Furthermore, we noticed a research gap on the impact of car dimensions on parking violations. As such, we plan to explore potential correlations between larger car dimensions and increased occurrences of parking violations.

4 METHODS

Understanding traffic and parking violations is a crucial component of maintaining order and efficiency in urban areas. Therefore, through this project, we aim to present a comprehensive precinct-wise analysis of parking violation rates in New York City. Our study incorporates four major innovations that we believe both shed new light on this issue and serve as a valuable resource for policymakers and researchers:

1. Novel Data and Features. Our research is focused on a combination of datasets and features that have not been previously explored in the context of parking violations. The goal here is to unearth correlations that have not yet been discovered, as well as provide a platform for potential avenues of investigation for future research. Our primary data source is the Parking Violation dataset published by the NYC Department of Finance, which is a detailed account of parking and traffic violations in New York City from 2013 to 2023. This dataset includes a plethora of information including violation precinct, issue date, time of day, vehicle type, and kind of violation, among many more.

The second set of datasets we analyzed is from the extensive NYPD archives of Collisions and Summonses Traffic Data, which spans from 2011 to 2023. When investigated in conjunction with our central Parking Violation dataset, our primary aim is to identify any previously undiscovered correlations between variables that could contribute to parking violations in each precinct and those that might exacerbate collision rates. Discerning these correlations bears significant implications for future policymaking and urban planning.

Additionally, we incorporated data from the "Violation Codes, Fines, Rules & Regulations" dataset from the NYC Department of Finance to extract information about parking violation fines throughout NYC and match them with their respective codes from our main dataset. The merged dataset offers valuable insights into diverse parking regulations, associated fines, and factors shaping parking behavior in NYC. This integration allows us to comprehensively examine the interplay between parking violations, fines, and specific regulations across different city areas.

2. Precinct-Wise Analysis. Another distinctive aspect of our method is the precinct-wise analysis. While prior research has examined parking violations at different levels, ranging from city-wide to borough-level to individual street-level analyses, our study centers on precincts. This approach enables us to look at the city on a more granular scale, investigating whether this

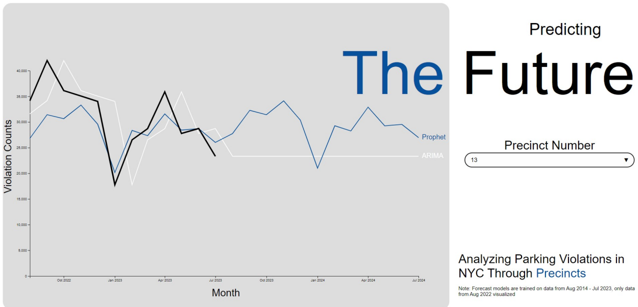

Figure 1: Example of Precinct Forecast Models (GitHub Pages Link)

depth of analysis reveals substantial variations in inter-precinct comparisons. In certain variables, such as most common violation type and average time of violation, neighboring precincts showed significant disparities in their outcomes. This suggests that policymakers should prioritize precinct-level initiatives to more effectively address parking challenges.

3. Time Series Forecasting. From an algorithmic perspective, our research employs both the ARIMA model and Meta's Prophet model [17] to provide time series forecasts for parking violation rates. We initially used the ARIMA model to predict monthly averages for the next recording year, however, due to dampening, ARIMA tends to predict very close to the overall average for all the months. Therefore, since this was a longer predicting interval, we decided to supplement our research with the Prophet model. This dual-model approach aims to provide users such as policymakers with a robust baseline for assessing the impact of their policies. Our research, emphasizing long-term trends, equips users with the tools to make informed decisions and adjust strategies as evolving patterns continue to unfold.

4. Interactive Visualization. To enhance the accessibility and interpretability of our findings, the primary component of our visualization is a choropleth map of the 77 precincts throughout New York City. Users can select the specific feature they wish to analyze through a drop-down menu and manipulate month and year parameters using sliders. This interactive interface enables users to comprehensively explore the data, revealing precinct-level variations and trends tailored to their distinct concerns. Our visual representation is not only informative but also user-friendly, making it an essential tool for both policymakers and future researchers.

Summary of Evolving Project Innovations. As our project evolved, we successfully incorporated additional datasets, including car type information (e.g., sedan vs. SUV) and NYC traffic violation codes and fine amounts, enhancing the richness of our analysis. Our commitment to comprehensive data accessibility is reflected in our prioritization of explainability, such as including definitions for each of the most common types of violations.

In the later stages of our project, we conducted detailed analyses of each feature, providing further insights into their relationships with parking violations. The initial list of features we investigated, including issue date, time of day, type of violation, number of violations, issuing precinct, and precinct collision rate, among others, yielded valuable findings that shaped our understanding of parking behavior in New York City.

Furthermore, we successfully developed ARIMA and Prophet models to generate one-year forecasts for precinct-wise parking violation rates, which are described in more detail in our Model Evaluation and Selection section below. These forecasts are now accessible through our interactive visualization, empowering users to plan for the future based on predicted trends.

In summary, our precinct-wise analysis of parking violations in New York City contributes valuable insights to urban planning and policymaking. By leveraging an innovative combination of data sources, a precinct-centric approach, time series forecasting, and an interactive visualization, users such as policymakers are equipped with the tools they need to effectively address parking challenges. This innovative combination not only addresses current issues but also sparks curiosity and encourages additional exploration in this field

5 DESCRIPTION OF TESTBED AND EXPERIMENTS

Model Evaluation and Selection: Our experiments involve the use of real-world datasets related to parking violations in New York City. The testbed includes the Parking Violation Dataset, NYPD Archives, and the Violation Codes, Fines, Rules & Regulations dataset from the NYC Department of Finance. Our raw dataset spans ten years, with individual files from 2014 to 2023 totaling 24.7 GB. Each file contains 10-20 million rows, resulting in 119.4 million timestamped data points, each with 43 features. Following data cleaning, which involved removing unnecessary features and addressing missing or faulty data points, the dataset was refined to 4.3 GB. Despite this reduction in size, the number of data points remained largely unchanged as only a small fraction was deemed faulty.

For correlation analysis and time series forecasts, we grouped the data by month (120 months over ten years) and precinct (77 precincts). This aggregation yielded a processed dataset of 9240 rows, forming the basis for our subsequent visualizations.

In our time series forecasting, we employed two models: the AutoRegressive Integrated Moving Average (ARIMA) model and the Prophet model developed by Facebook's Core Data Science team. The ARIMA model, a standard time series forecasting approach, represents the forecasted value as a weighted average of previous values. The Prophet model, on the other hand, is an additive model that considers piecewise trend components, seasonal components, and holiday components.

To fine-tune and evaluate these models, we trained them on data from 2014-2022 and used 2023 data for Root Mean Squared Error (RMSE) evaluation. Both the ARIMA and Prophet models resulted in similar RMSEs, but the ARIMA models were found to be unsuitable for long-term forecasts. A grid search for ARIMA hyperparameters revealed that most resulting models were ARIMA(1/2/3, 0/1, 0) with no seasonality. Given that the next forecasted value for ARIMA(p, d, 0) relies on the preceding p+d values, our fit parameters rendered the ARIMA models unsuitable for long-term forecasts. This can be seen in our visualization, where the ARIMA forecasts tend to plateau soon after the recorded data ends.

In contrast, the Prophet model, with its out-of-the-box hyperparameters, proved effective for long-term forecasts and planning. Although it tended to overshoot and undershoot, the Prophet model demonstrated the ability to predict changes in trends, particularly in identifying the locations of peaks and valleys in the forecasts.

During the data analysis, two significant outliers were identified: a surge in parking violations in January 2015 – with an unclear cause – and a dip in violations from March 2020 to July 2020 – attributed to the impact of COVID-19. To enhance the precision of our models, we removed the COVID-19 data as outliers. Despite this, no notable difference in forecast quality was observed, leading us to use the models trained on the full dataset for our final visualization.

User Feedback for Interactive Choropleth Visualization. To gauge the effectiveness of our model and visualization, we presented our work to a sample of students at Georgia Tech and requested their feedback regarding our visualizations on a scale of 1-5. This evaluation helped us fine-tune our approach, allowing us to revise our visualization to optimize for attractiveness, understandability, and ease-of-use. In doing so, we ensured that our visualizations are straightforward and informative, allowing our research to serve its intended purpose. The feedback received from users on the scale of 1-5 is as follows:

- Visual Appeal (Average Rating): 3.94
- Interaction Ease (Average Rating): 4.24
- Understandability (Average Rating): 4.12

This feedback allowed us to make necessary revisions in subsequent iterations of our visualizations. For example, to enhance the visual appeal, we opted for consistent color schemes in our choropleths across all features in the dropdown menu, ensuring uniformity for a more polished appearance. Furthermore, we made our visualizations compatible with computer screens of varying sizes, promoting accessibility across different devices. Additionally, we heightened the contrast between the choropleth background and the accompanying descriptions, improving overall clarity. To enhance ease-of-use, we increased the size of the feature dropdown box, as well as the size of the month and year sliders. Lastly, to improve understandability, we added a tooltip feature that appears when users hover their mouse over a precinct to reinforce the key takeaways for the chosen feature and that specific precinct.

Overall, users found the interface intuitive and the information presented valuable for understanding parking violation trends in New York City. These insights were invaluable in refining the user interface to better align with user expectations and preferences.

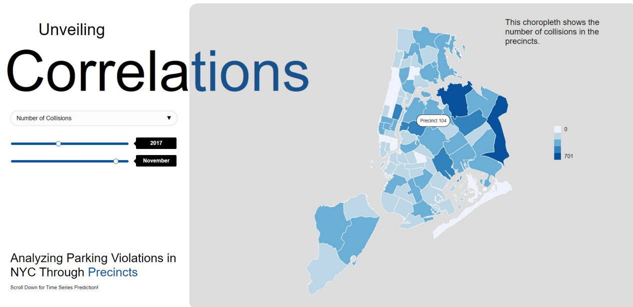

Figure 2: Precinct-Wise Choropleth Visualization (GitHub Pages Link)

6 OBSERVATIONS

In our analysis, it is essential to acknowledge a significant limitation stemming from the absence of normalizing data, such as the total number of vehicles of a specific type or the total number of vehicles parked in a specific precinct at a given time. This absence complicates our ability to draw definitive conclusions from our findings, as we lack a holistic contextual benchmark for the observed patterns. Therefore, our observations are considered in relation to each other rather than being contextualized within the broader context of urban traffic.

Despite this limitation, examining the patterns of parking violations over the past 11 years reveals noteworthy trends. For example, the most prevalent type of violation across all years, months, and precincts was Street Cleaning. This violation occurred when a vehicle was parked in an area or during a time of day where signage explicitly indicated that parking was not allowed. The fine for such an infraction is $65. Cumulatively, the average sum of fines for these violations amounted to $633,284,273.60, shedding light on the substantial financial consequences of these violations.

As part of our analytical approach, we incorporated the use of the estimated moving average (EMA) with a 3-day window and a threshold based on the average daily violation count of each individual precinct for each month. Specifically, any month with more than 3 out of the last 5 days of the month surpassing the daily average was considered to be indicative of a potential end-of-month increase in violations, possibly driven by an effort to meet a predetermined quota for filed violations. Notably, precincts 78 and 80 emerged as the precincts most frequently exhibiting this trend, followed by precincts 122, 120, 108, and 13. This suggests a consistent pattern of adjusting violation counts to align with specific quotas in these areas.

The analysis of collision data exposes insights into the relationship between violations and vehicular incidents. In our analysis, precincts 105, 109, and 75 emerge as the areas with the highest number of collisions, reporting 52,605, 51,268, and 47,519 collisions, respectively. Notably, the collision figures showcase a significant decline during the COVID-19 years. While the immediate drop during the pandemic is unsurprising due to lockdowns and restricted movement, the intriguing aspect is the indication that collisions remained comparatively lower in the post-COVID period compared to pre-pandemic. This raises questions about potential lasting effects of the pandemic on driving habits or broader shifts in transportation patterns. Analyzing such effects and trends could provide valuable insights for urban planning, transportation policies, and understanding the evolving dynamics of vehicular incidents in these precincts and how these relate to parking violations.

Furthermore, certain precincts exhibited patterns in ticketing frequencies that correspond to the type of vehicle involved, with passenger vehicles consistently receiving the highest number of tickets. However, certain months still saw heightened ticketing for commercial vehicles in particular precincts – specifically, precincts 1, 22, and 19.

The data also reveals a noteworthy shift in the body types of vehicles with the highest violation percentages. Initially, sedans had the lead in 2014, but subsequently, suburban types (including SUVs) consistently held the top position. This suggests a potential change in the vehicle landscape receiving parking violations over the years. The sustained dominance of suburban types aligns with the broader trend of SUVs gaining popularity. Despite this overall trend, there are intriguing instances where delivery trucks or vans temporarily dominated specific precincts. These anomalies highlight the localized nature of parking challenges and regulations. The precinct-level analysis becomes crucial in deciphering such nuances.

Time-related patterns further underscore the nuances of parking violations. Our analysis showed that morning hours (8am – 12pm) tended to be the most common time for receiving tickets across precincts, followed by afternoons (12pm – 4pm). Noteworthy exceptions were observed, with a few precincts consistently reporting the nighttime (8pm – 8am) as the highest ticketed period. Consistently, evening (4 – 8pm) infractions were relatively rare. Interestingly, many precincts maintained consistent patterns regardless of the observed month and year, indicating a certain stability in ticketing behaviors and further underscoring the need for a precinct-wise analysis of traffic and parking violations in New York City.

The color of vehicles also appeared to play a role in ticketing patterns, with gray and black being the most frequently ticketed colors in the majority of precincts. However, anomalies were observed in precincts 1 – 20 and 109, where white vehicles were predominant in receiving tickets. In the same precincts, vans emerged as the most commonly ticketed vehicle type, suggesting a potential correlation between vehicle color and vehicle type in these specific areas.

In summary, our analysis uncovers numerous different multifaceted patterns in parking violations, illuminating the most common violations, financial implications, collision correlations, temporal trends, and even potential associations between vehicle color and type. However, the absence of normalizing data emphasizes the need for caution in drawing conclusive insights from our observations, and highlights how future studies could build upon our observations.

7 CONCLUSIONS AND DISCUSSION

In conclusion, our research – including a precinct-wise analysis, time series forecasting, and interactive visualizations – aims to provide a holistic understanding of parking violations in New York City. Through the utilization of diverse datasets, we were able reveal many of the intricate patterns and correlations that can significantly contribute to the formulation of effective urban policies.

Our approach transcends traditional analyses, offering precinct-focused insights, forecasting future trends, and presenting a user-friendly visualization tool. This culmination represents a valuable resource for policymakers and researchers concerned with the intricate challenges of urban parking. The insights garnered from our research hold the potential to guide targeted policies at the precinct level, addressing specific factors contributing to parking violations.

Moving forward, there is room for further research to build upon our work. Incorporating additional information, such as parking meter density throughout New York City or demographic details about each precinct, could deepen our understanding factors that lead to parking violations in urban areas and broaden the scope of policy recommendations. Our research sets the stage for ongoing exploration, providing a foundation for future investigations into the multifaceted dynamics of city parking.

All team members have contributed a similar amount of effort to ensure a balanced and collaborative approach to the project.

REFERENCES

[1] Claudio Badii, Paolo Nesi, and Irene Paoli. 2018. Predicting Available Parking Slots on Critical and Regular Services by Exploiting a Range of Open Data. *IEEE Access* 6 (2018), 44059–44071. https://doi.org/10.1109/ACCESS.2018.2864157

[2] Allison L. C. De Cerreño. 2004. Dynamics of On-Street Parking in Large Central Cities. *Transportation Research Record* 1898, 1 (2004), 130–137. https://doi.org/10.3141/1898-16 arXiv:https://doi.org/10.3141/1898-16

[3] Yu-Chiun Chiou and Chiang Fu. 2015. Modeling crash frequency and severity with spatiotemporal dependence. *Analytic Methods in Accident Research* 5-6 (2015), 43–58. https://doi.org/10.1016/j.amar.2015.03.002

[4] Jingqin Gao and Kaan Ozbay. 2017. A Data-driven Approach to Predict Double Parking Events Using Machine Learning Techniques.

[5] Stefan Gössling, Andreas Humpe, Rafael Hologa, Nils Riach, and Tim Freytag. 2022. Parking violations as an economic gamble for public space. *Transport Policy* 116 (2022), 248–257. https://doi.org/10.1016/j.tranpol.2021.12.010

[6] Young-An Kim James C. Wo and Sarah E. Malone. 2023. Examining the spatial distribution of parking tickets in San Francisco neighborhoods: An overlooked form of urban inequality? *Journal of Urban Affairs* 0, 0 (2023), 1–32. https://doi.org/10.1080/07352166.2023.2239956 arXiv:https://doi.org/10.1080/07352166.2023.2239956

[7] Goran Jelen, Vedran Podobnik, and Jurica Babic. 2021. Contextual prediction of parking spot availability: A step towards sustainable parking. *Journal of Cleaner Production* 312 (2021), 127684. https://doi.org/10.1016/j.jclepro.2021.127684

[8] Nikolaos Karantaglis, Nikolaos Passalis, and Anastasios Tefas. 2022. Predicting on-street parking violation rate using deep residual neural networks. *Pattern Recognition Letters* 163 (2022), 82–91. https://doi.org/10.1016/j.patrec.2022.09.023

[9] Woojung Kim and Xiaokun (Cara) Wang. 2022. Double parking in New York city: a comparison between commercial vehicles and passenger vehicles. *Transportation* 49, 5 (01 Oct 2022), 1315–1337. https://doi.org/10.1007/s11116-021-10212-5

[10] Javad Koohpayma, Amir Tahooni, Mohammadreza Jelokhani-Niaraki, and Jamal Jokar Arsanjani. 2019. Spatial Analysis of Curb-Park Violations and Their Relationship with Points of Interest: A Case Study of Tehran, Iran. *Sustainability* 11, 22 (2019). https://doi.org/10.3390/su11226336

[11] Ruth Vassar Lazenby. 2020. *Who Pays? An Analysis of Fine Collection in New York City*. Technical Report. https://heinonline.org/HOL/Page?handle=hein.journals/yljfor130&div=11&g_sent=1&casa_token=HeVpK5jkObkAAAAA:QK9Xqz6wp2zlYGwcpFtTkAaZhs6d4yqpYoPtxeh_P935-lTAEnvwX8PfAPXJcDfTjiZUJ9w_5g&collection=journals

[12] Livia Mucciolo, Fay Walker, and Aravind Boddupalli. 2023. *The Cost of Parking*. Technical Report. https://policycommons.net/artifacts/3412918/the-cost-of-parking/

[13] D. Shoup. 2007. *Cruising for Parking*. Technical Report. https://escholarship.org/uc/item/6sn7s1x2

[14] Donald C. Shoup. 1997. The High Cost of Free Parking. *Journal of Planning Education and Research* 17, 1 (01 Sep 1997), 3–20. https://doi.org/10.1177/0739456X9701700102

[15] Jelena Simićević, Smiljan Vukanović, and Nada Milosavljević. 2013. The effect of parking charges and time limit to car usage and parking behaviour. *Transport Policy* 30 (2013), 125–131. https://doi.org/10.1016/j.tranpol.2013.09.007

[16] Yunlei Liang Joseph Marks Yuhao Kang Song Gao, Mingxiao Li and Moying Li. 2019. Predicting the spatiotemporal legality of on-street parking using open data and machine learning. *Annals of GIS* 25, 4 (2019), 299–312. https://doi.org/10.1080/19475683.2019.1679882 arXiv:https://doi.org/10.1080/19475683.2019.1679882

[17] Letham B. Taylor SJ. 2017. Forecasting at scale. *PeerJ Preprints* (2017). https://doi.org/10.7287/peerj.preprints.3190v2

[18] Adam Wenneman, Khandker M. Nurul Habib, and Matthew J. Roorda. 2015. Disaggregate Analysis of Relationships between Commercial Vehicle Parking Citations, Parking Supply, and Parking Demand. *Transportation Research Record* 2478, 1 (2015), 28–34. https://doi.org/10.3141/2478-04 arXiv:https://doi.org/10.3141/2478-04

[19] Xizhen Zhou, Xueqi Ding, Jie Yan, and Yanjie Ji. 2023. Spatial heterogeneity of urban illegal parking behavior: A geographically weighted Poisson regression approach. *Journal of Transport Geography* 110 (2023), 103636. https://doi.org/10.1016/j.jtrangeo.2023.103636

[20] Stefania Zoika, Panagiotis G. Tzouras, Stefanos Tsigdinos, and Konstantinos Kepaptsoglou. 2021. Causal analysis of illegal parking in urban roads: The case of Greece. *Case Studies on Transport Policy* 9, 3 (2021), 1084–1096. https://doi.org/10.1016/j.cstp.2021.05.009