Data-Driven Approaches to Parking Management in New York City

Introduction

In fiscal year 2023, the NYC Department of Finance (2023) reported 17,245,489 parking tickets and camera violations, reflecting an 11.36% increase from the previous year. With a population of approximately 8.26 million residents (NYC Department of City Planning, 2023), this statistic translates to nearly two tickets per person on average. Excluding individuals without a driver's license would likely result in an even higher average. Given the substantial impact of parking issues on daily life, it is imperative to investigate how big data analytics can enhance parking management and mitigate violations.

Parking management can be divided into two primary components: the cost-effective detection of parking violations and the strategic planning of parking infrastructure and land use to meet the needs of both residents and tourists. By analyzing data related to parking violations in conjunction with information on parking infrastructure and land use, it becomes possible to identify patterns of parking violations and discern the key factors influencing these infractions.

Related Research

A multitude of studies have focused on mitigating parking challenges through the optimization of parking systems, utilizing advanced guidance systems (Chai et al., 2019) and implementing Internet of Things (IoT) technologies (Osoba, 2012; Singh, 2024). However, illegal parking remains prevalent and presents significant challenges for resolution. In the context of traffic management, studies indicate that illegal parking is one of the leading causes of urban nonrecurring congestion, behind crashes and construction (Han, L. D. et al., 2005).

And illegal parking would even affect autonomous vehicles (Kuo and Lin, 2024). Consequently, there is a pressing need for data-driven methodologies to ascertain the fundamental causes of issues related to illegal parking.

Traditionally, methods for detecting illegal parking have relied on street occupation sensors to ascertain whether a vehicle is parked in a designated space. These systems typically utilize either underground sensors or advanced vision-based technologies (Yang et al., 2017), even combined with deep learning (Liu et al., 2024). However, these detection systems are often costly and more effective in constrained environments, making their installation in public spaces or entire cities particularly challenging. Using indirect measurements—such as location, time, and historical patterns—may provide valuable insights for addressing this issue.

Some studies have partially addressed these indirect factors. For instance, the research conducted by Kawamura et al. (2014) examined truck parking citations in Chicago, utilizing geographic information systems and regression analysis, and revealed the prevalence of illegal parking among delivery vehicles. But there are still some other patterns and relationships with parking violations that need to be figured out. Furthermore, these insights will provide practical solutions for city planners and policymakers, promoting a more efficient and equitable urban environment.

Data Processing

The data of parking violation issues from fiscal years 2021 to 2024 was collected from NYC Open Data. And for the matching with daily life propose, the data was re-organized into natural years from 2021 to 2023. And all the years mentioned below are natural years. The data of street distribution, Street Name Dictionary (SND), parking infrastructures, and land use were also all from NYC Open Data platform. All the row data has some dirty data which needs to be cleaned.

The data of parking violation where both street name and street code are empty will be regarded as invalid data and be deleted directly. Additionally, the street code and street name in the raw dataset has some missing, misspelled, and irregular abbreviations, which cannot match the shapefiles of street map well. So, both the raw dataset and the shapefile of road networks need further processing. At first, some irregular abbreviations in the street name of violation will be normalized, and the street cord will be completed with the street name of the same name. Second, the data in the shapefile of street networks will be compared with SND using street name to match the street code. However, the street names in SND have some different organization methods with the street name in shapefile. So, the Levenshtein Distance Algorithm will be applied to match the proper street name in SND with the street name in the shapefile of roads. This algorithm has a high accuracy in matching street names (Innerhofer-Oberperfler & Augsten, 2004). Finally, the raw dataset can be matched with the shapefile of streets using street code.

And the violation time in raw dataset of parking violation also has different data structure and some error value which exceed 24 hours, it will be re-organized into 24-hour clock system if it can be repaired, and the error data will be replaced to null value to avoid the misleading of analysis.

Violation Analysis

General Analysis

From 2021 to 2023, there are a total of 52,392,189 valid parking violation issues records in New York, of which 15,004,334 are from 2021, 16,523,915 are from 2022, and 20,863,940 are from 2023. And the increasing rates of 2022 and 2023 are 10.13% and 26.27%, respectively. It is obvious that the violation numbers are increasing year by year. It should be noted that the New York Cities declared state of emergency due to the COVID-19 and partially relaxed the lockdown order on November 10th, 2020. New York City was fully reopened on July 1st, 2021. The affection of pandemic in 2021 cannot be ignored. Some research indicates that the lockdown implemented during the pandemic is likely to result in a decrease in the demand for parking (Mesfin, B. G. et al., 2024).

Among all these violation issues, 38,966,908 issues were committed by local vehicles, which consists 74.38% of total issues; and 13,425,281 issues were committed by non-local vehicles, which consists 25.62% of total issues. Considering that the base number of non-local vehicles is minor compared to local vehicles, the non-local vehicles may be more likely to violate parking regulations.

About the violation type, the most common violation code is 36, which means illegal parking in or near a designated school zone. About the issuing agency, most of the issues were issued by Traffic Enforcement Agents (TEA).

Issues Number Every Month

In general, the number of parking violation issues in the third quarter is the highest from 2021 to 2023 and decreases towards the end of the year (Fig 1), especially in 2023. And generally, the violation numbers every month are increasing year by year. This kind of pattern may be because the third quarter often includes summer months when people are more active, leading to increased traffic and parking demand, especially in tourist areas or during events.

Furthermore, it can be observed that the impact of COVID-19 on the number of violations in 2021 appears to be minimal.

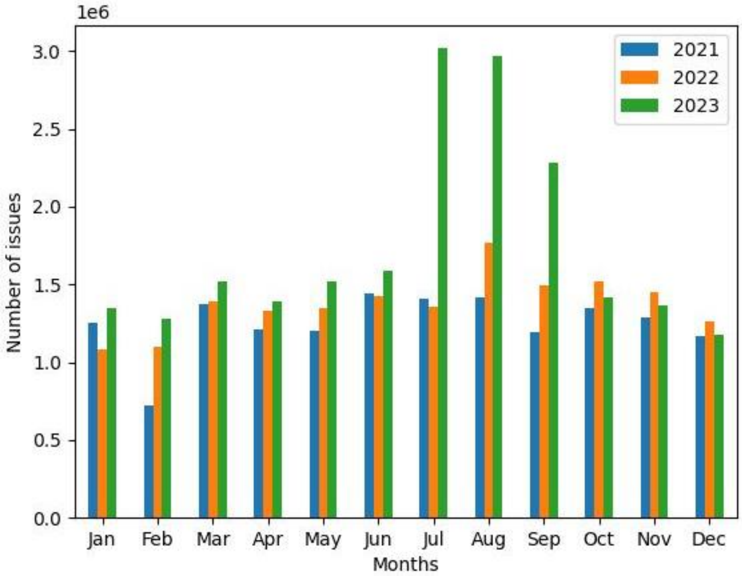


Fig 1. Number of Parking Violation Issues from 2021 to 2023

Issues Number Every Hour

From 2021 to 2023, the incidence of parking violations predominantly occurred between 7:00 AM and 1:00 PM, with a notable decline observed after midnight (Fig 2). This trend may be associated with the operational schedule of law enforcement agencies. Previous findings indicate that the majority of violations are issued by the TEA, whose personnel typically commence their duties at 7:00 AM. Given that field staff are contracted for 40 hours per week, this translates to approximately six hours of work per day, according to TEA's recruiting documents. This working arrangement coincidently aligns their active enforcement period with the aforementioned time frame of 7:00 AM to 1:00 PM. And during the midnight hours, most vehicles have typically returned to their residences, where most owners have access to garages. Additionally, the ending of the TEA's and other law enforcement agencies' field operation hours further contributes to the rapid decline in the number of parking violations during this time.

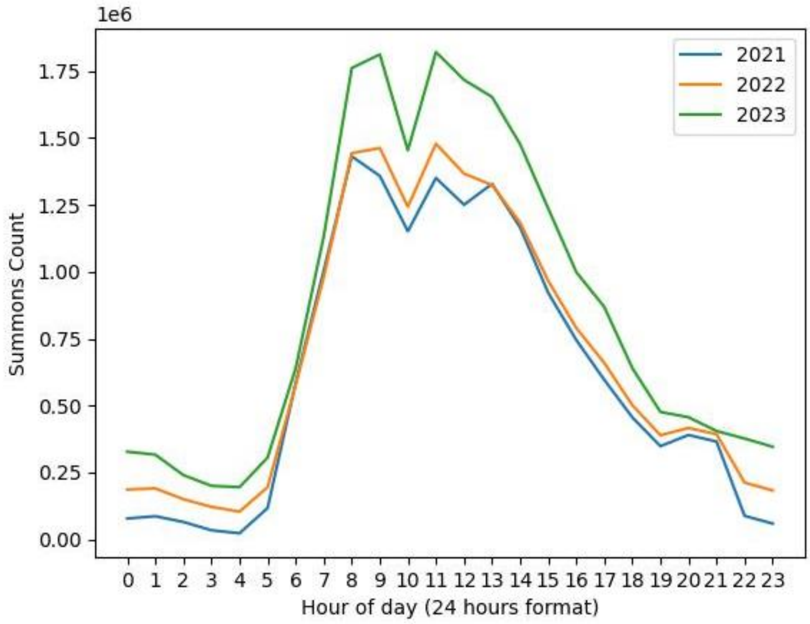


Fig 2. Number of Parking Violation Issues Every Hour from 2021 to 2023

Furthermore, the period around 9:00 AM coincides with the morning rush hour, during which a substantial volume of vehicles is present on the road. Consequently, the incidence of parking violations significantly decreases during this time. And the period of 8:00 PM after the end of evening peak, during which a significant volume of vehicles returned home, leads to a slight increase in violation issues number. Considering the effects of COVID-19 in 2021, it can be observed that the pandemic's impact on the number of violations during that year was minimal. The trends in violations over the three-year period exhibit a remarkable degree of similarity.

Violation Location

From 2021 to 2023, most of the parking violation issues are issued at south parts of Manhattan, Park Avenues of Manhattan, Midwood of Brooklyn, Midwood of Brooklyn, and some main road of New York City (Fig 3). It is easy to find that hot tourist areas, main road areas and high-density residential areas have more parking violations than other areas.

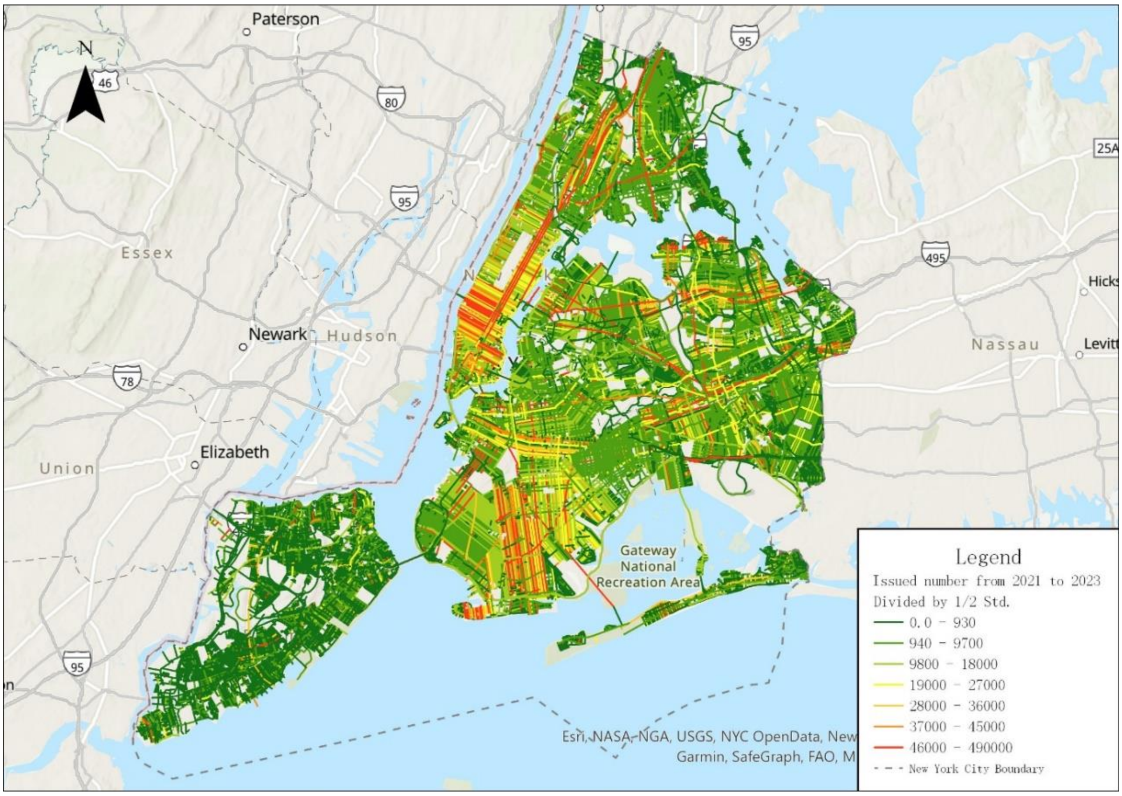


Fig 3. Parking Violation Issues from 2021 to 2023

An analysis of year-over-year differences reveals that most areas of New York City do not experience a significant increase in the number of parking violations. However, regions characterized by a higher overall volume of parking violations demonstrate a marked upward trend over the past three years (Fig 4, Fig 5). This trend may be attributed to the fact that areas with a greater total of parking violations often experience heightened economic activity and urban development, resulting in increased demand for parking and, consequently, a rise in violations. This trend seems to be that the impact of COVID-19 on the number of violations in 2021 appears to be minimal.

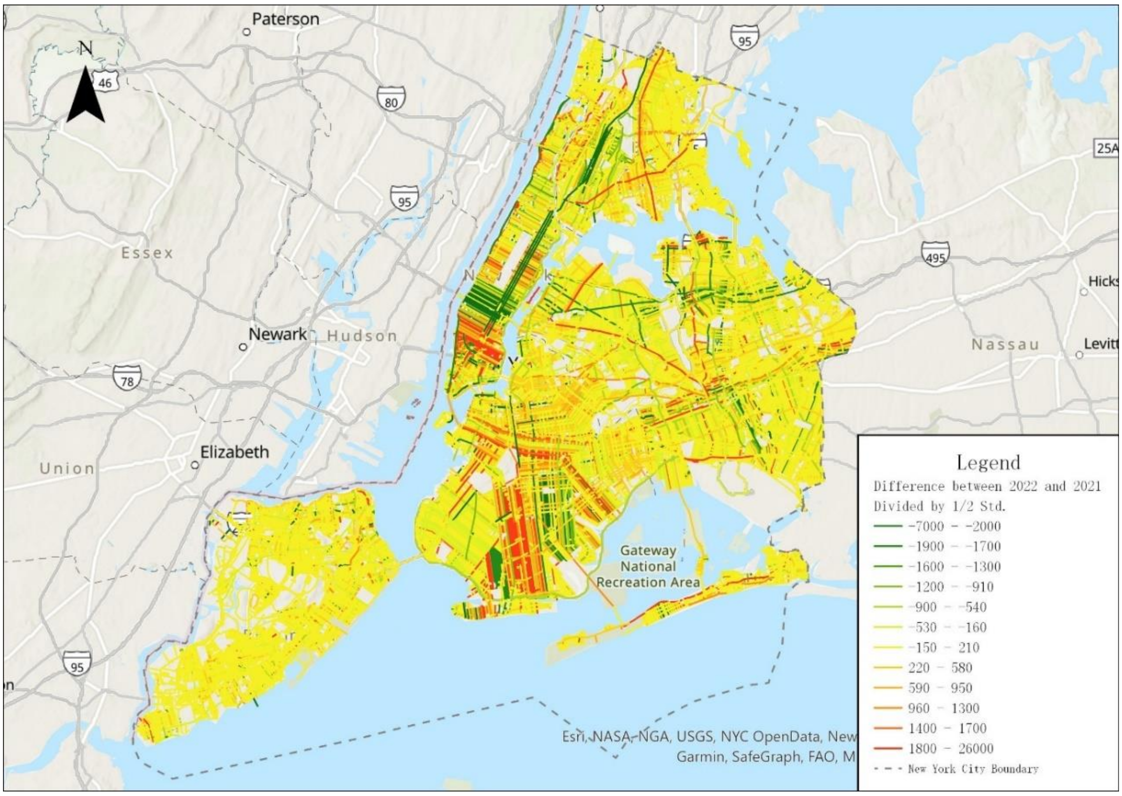


Fig 4. Difference of Parking Violation Issues between 2022 and 2021

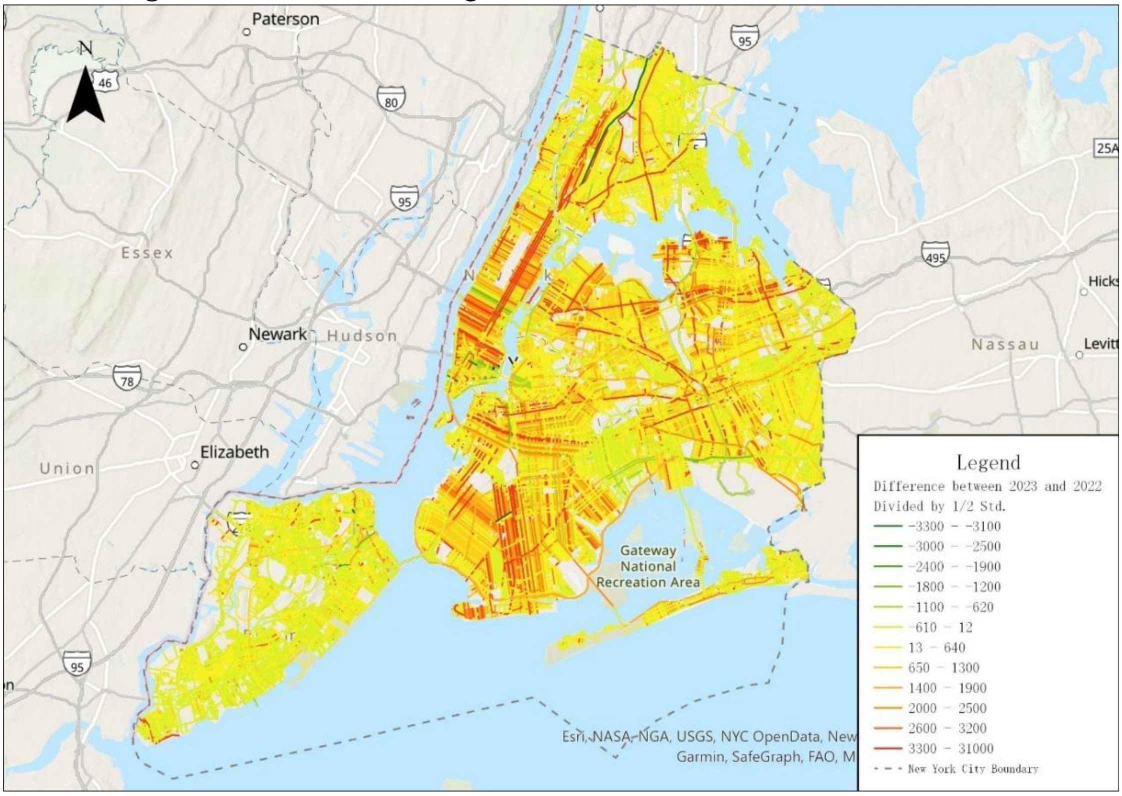


Fig 5. Difference of Parking Violation Issues between 2023 and 2022

Violation Relationship

Relationship with Driver

There are many methods to predict violation patterns. For example, machine learning (ML) models which range from random forests (RF) (Gao et al., 2019) to powerful residual neural networks (Karantaglis et al., 2022) have been used to estimate the number of illegally parked cars. However, applying graph neural networks to this problem is challenging due to the incompatibility of most datasets with the required format. Additionally, the sparsity of annotations and missing data in on-street parking datasets further complicates their effectiveness. Using semi-supervised graph convolutional networks may solve these problems (Karantaglis et al., 2024), but this method empathizes more on the violation location but ignores the role of personal characteristics.

In summary, RF is a good algorithm to analysis the inner relationship between behavior and characteristics. It is a supervised ML algorithm which is wildly used in classification tasks such as the impact of loan features on bank loan prediction (Madaan et al., 2021; Dansana et al., 2024), and it achieves 80% accuracy in classification (Madaan et al., 2021). Because of the similarities of the prediction function between parking violation and loan overdue repayment, both are using classification model to map samples in a dataset to given category, RF can be used to predict the parking violation patterns related to drivers.

RF algorithm uses ensemble learning to builds multiple decision trees and aggregates their results to make predictions more accurate and stable (Fig 6). RF predicts outcomes by combining the results of multiple decision trees, either by selecting the most common class for classification or calculating the mean for regression. The methodology has been introduced by Dansana et al. (2024).

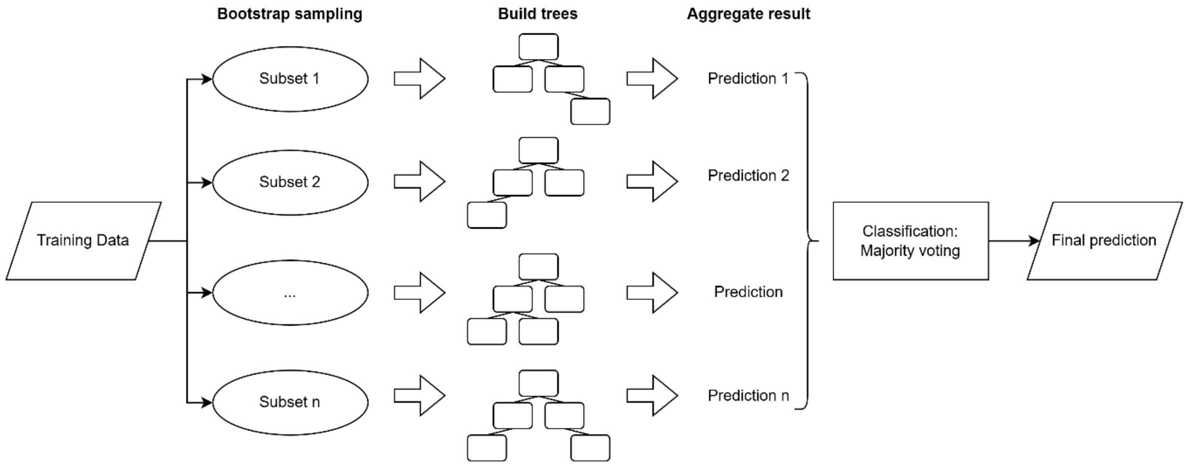


Fig 6. Random Forest Classifier

To analysis the which characters of the drivers will affect the violation type, the registration state, plate type, vehicle body type, vehicle make, and violation time of drivers are chosen as the variables, and violation code issued by agencies are selected as the prediction result. Due to the data are manual input by enforcement officers, there are some data have misspelling. To enhance the accuracy of the model, values in variables that occur with a frequency lower than one hundred will be classified as invalid data and subsequently excluded from the dataset used for modeling. And 20% of the data are selected randomly as test data. The number of trees is set to 128, because further increasing the number of trees will not result in any further improvement of the accuracy of the RF except for increasing the computational load (Oshiro et al., 2012).

The result of RF model shows that the violation time has the most significant impact on the parking violation type, followed by vehicle body type (Fig 7). And the accuracy of modeling is 42.11%.

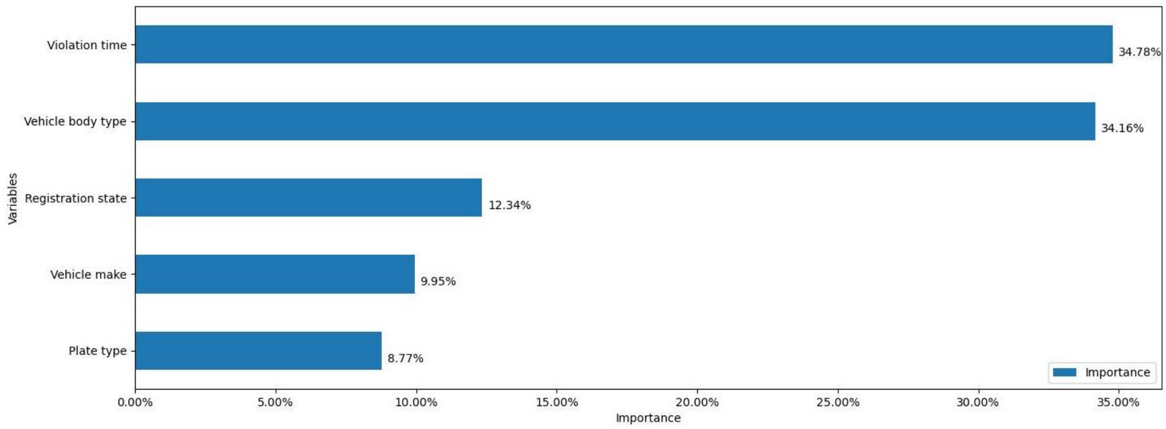


Fig 7. The Influence Degree of Different Characters

An analysis of the influence of land use over the years (Fig 8 to Fig 10) reveals that the patterns of parking violations associated with driver characters have not been significantly affected by the passage of time and the infection of COVID-19. The violation time and the vehicle body type of drivers are the two most significant factors influencing the occurrence of violations.

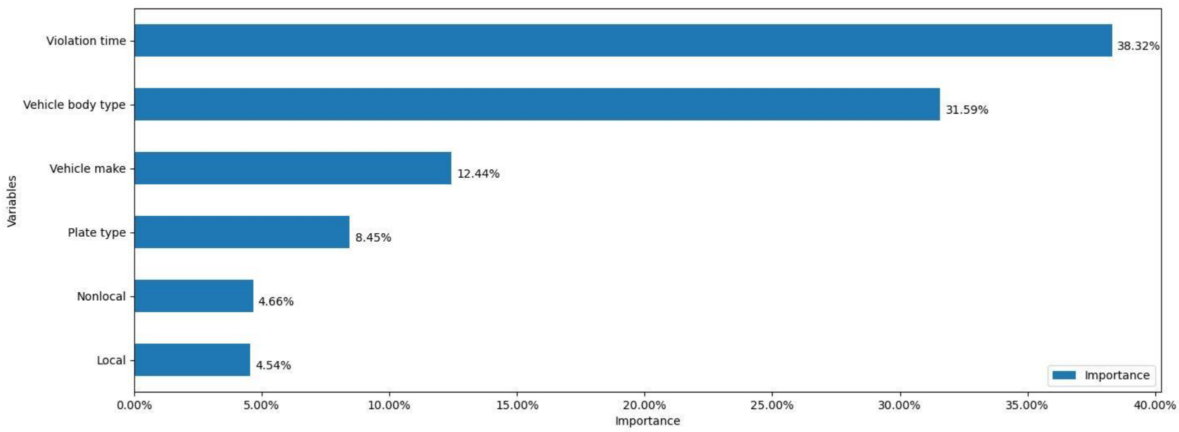


Fig 8. The Influence Degree of Different Characters in 2021

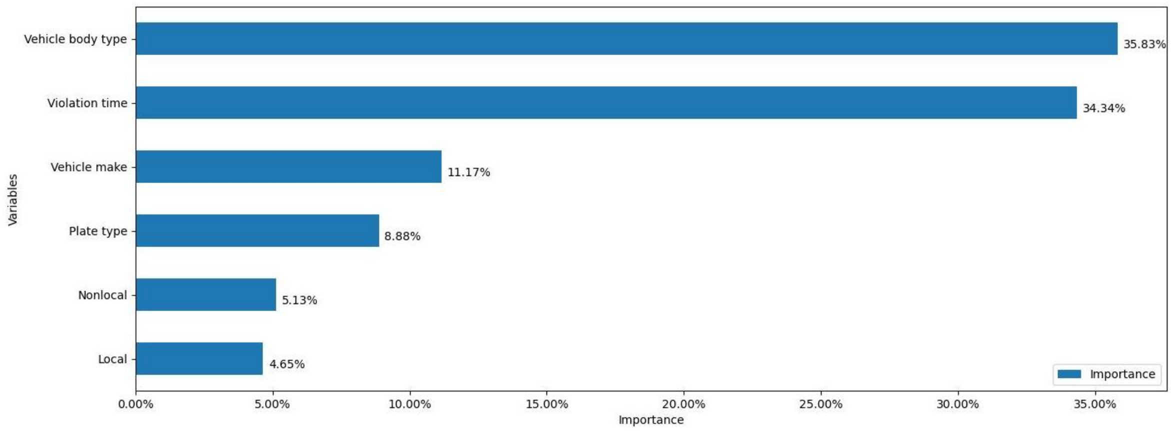


Fig 9. The Influence Degree of Different Characters in 2022

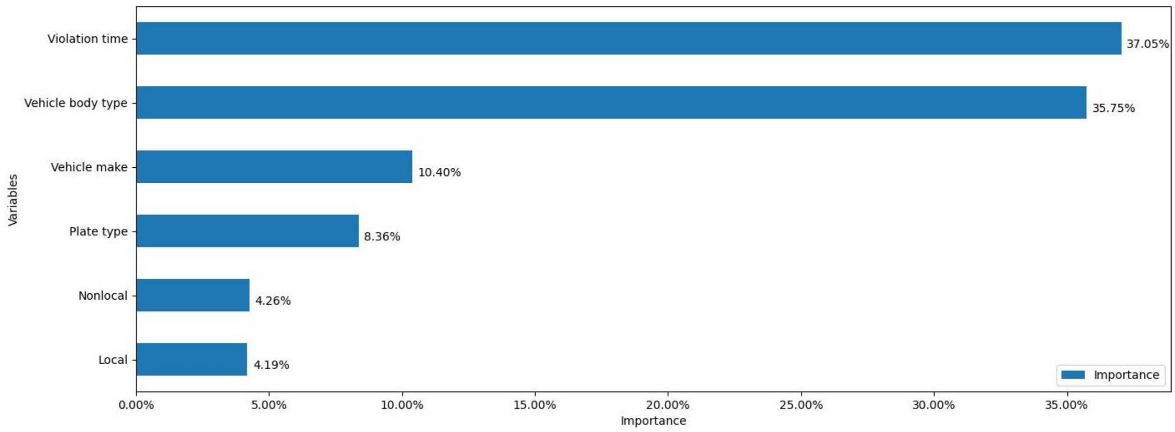


Fig 10. The Influence Degree of Different Characters in 2023

So, the violation time and vehicle body type from 2021 to 2023 can be converted into dummy variables through One-Hot encoding and subsequently reapplying RF model, the result shows that the four violation times exerting the most significant influence on parking violation types are 8:00 AM, 9:00 AM, 11:00 AM and 7:00 PM (Fig 11). The model achieved an accuracy of 36.08%. It is evident that both the morning and evening rush hour periods significantly affect the types of violations observed.

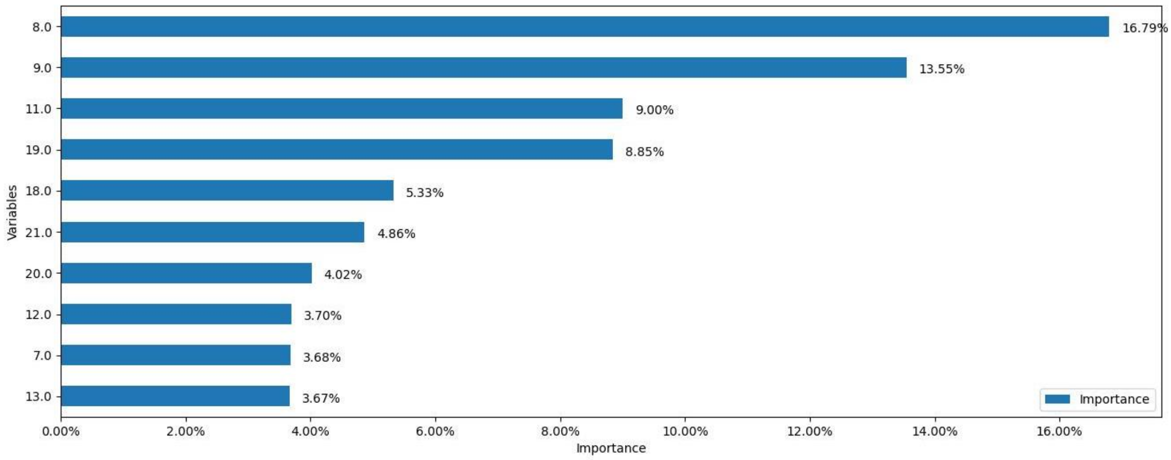


Fig 11. The Influence Degree of Top Ten Violation Time

Due to the inherent challenges associated with machine learning algorithms, such as RF, which are often characterized by the "black box" problem (Azodi et al., 2020), it remains difficult to interpret the individual subtrees generated during the RF process. Despite various efforts (Guloway & Woźniak, 2021, Bastani et al., 2017) to extract interpretable decision tree ensembles from RF models, achieving clarity and understanding in these interpretations continues to pose significant challenges. But the scatter plot of original data can reflect the relationship between variables and prediction result to some extent.

By plotting the relationship between violation time and violation code, it can be found that nearly all violation codes are encompassed within these time periods (Fig 12). This indicates that these periods are associated with a heightened risk of parking violations, even though the absolute number of some periods – like 7:00 PM – is not high. But all these times are turning point in violation number (Fig 2).

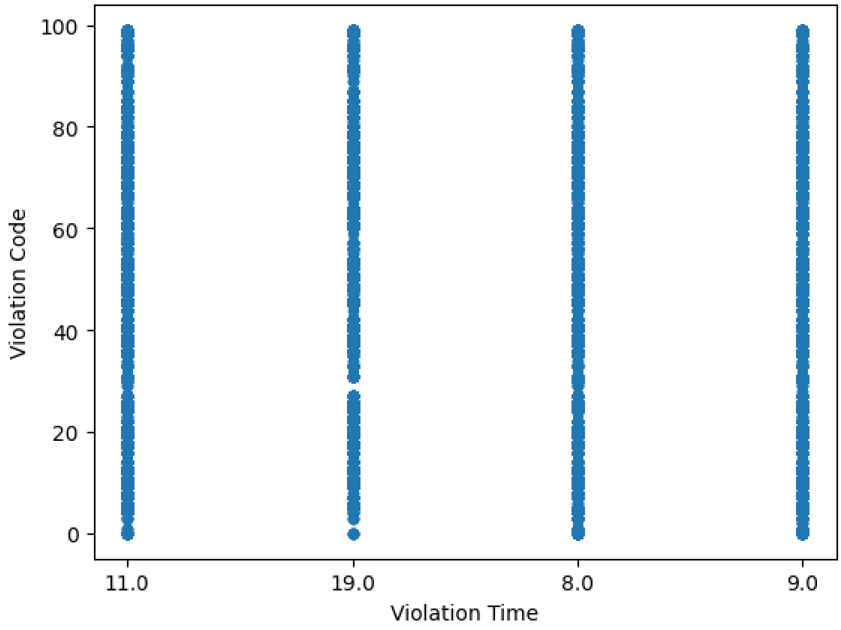


Fig 12. Scatter Graph of Violation Time and Violation Code

The vehicle body types that exert the most significant influence on the incidence of parking violations include vans, utility vehicles, delivery vehicles, and standard duty vehicles (Fig 13). A total of 185 distinct vehicle body types have been documented in the records of parking violation issues; consequently, only the top ten vehicle body types are presented herein. This indicates that these categories of vehicles are associated with a diverse range of parking violations. Furthermore, it is noteworthy that all these vehicles can be classified as commercial vehicles.

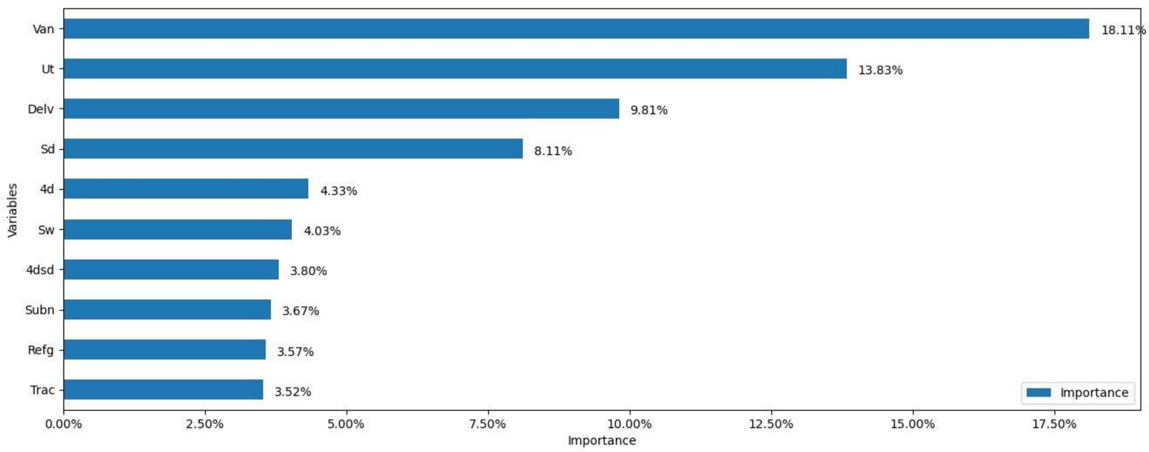


Fig 13. The Influence Degree of Top Ten Vehicle Body Type

By analyzing the correlation between vehicle body type and violation codes, it becomes evident that delivery vehicles and vans exhibit a propensity to violate nearly all codes. In contrast, standard duty vehicles and utility vehicles demonstrate distinct patterns of violations (Fig 14).

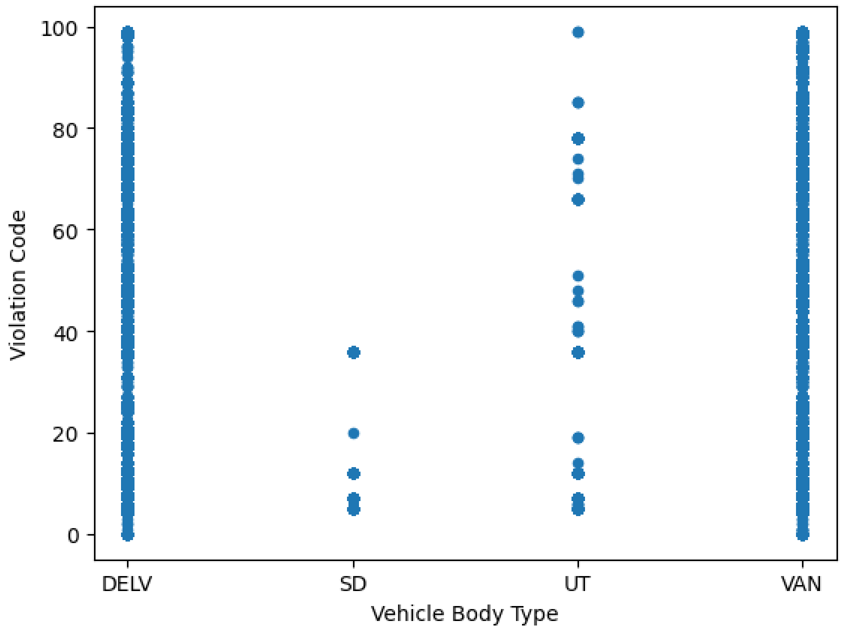


Fig 14. Scatter Graph of Violation Time and Vehicle Body Type

Standard duty vehicles tend to predominantly violate codes between 0 and 20, which are primarily related to violations of parking sign regulations, as well as codes ranging from 35 to 40, primarily associated with exceeding time limits .In contrast, utility vehicles exhibit a more complex pattern of violations compared to standard duty vehicles. Notably, utility vehicles rarely violate codes between 20 and 30, which pertain to illegal parking in designated spaces, as well as codes ranging from 55 to 65, typically related to obstructions and highway regulations, and codes from 85 to 95, which are generally associated with issues concerning vehicle registration documentation.

Relationship with Parking Infrastructure

To analysis the influence of parking infrastructure such as parking lots and parking meters in parking violation issues, the buffer was used to calculate how many parking infrastructures are near streets. The buffer was set as dishes shape and the distances are set in four categories, 50 meters, 100 meters, 200 meters, 500meters, 1000 meters, 2000 meters and 5000 meters. Due to simple regression models are not suitable for time-series data (Özen, 2024) relating to parking violation issues, the RF regression was used to find out which distance has the significant impact of parking violation issues. RF can be not only used in classification, but also in regression. The diversity in decision trees helps capture different relationships between inputs and the target variable, decreasing correlation among trees and enhancing the model's robustness to noise and outliers (Özen, 2024). Given that the most recent data on parking lots and parking meters is from 2022, and to mitigate the potential influence of COVID-19, the analysis is conducted using data from 2022.

First is the type of parking infrastructures and the distance between the streets and them. The result (Fig 15) shows that the parking meters within 500 meters and the parking lots within 5000 meters around the streets has a significant impact on the number of parking violation issues, while the impact of parking meters within 50 meters and the parking lots within 1000 meters is relatively weaker.

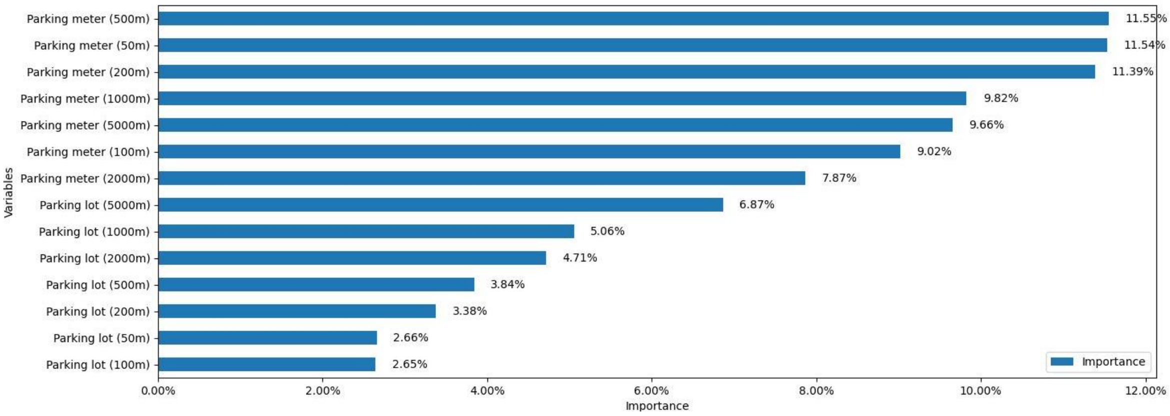


Fig 15. The Influence Degree of Parking Infrastructure

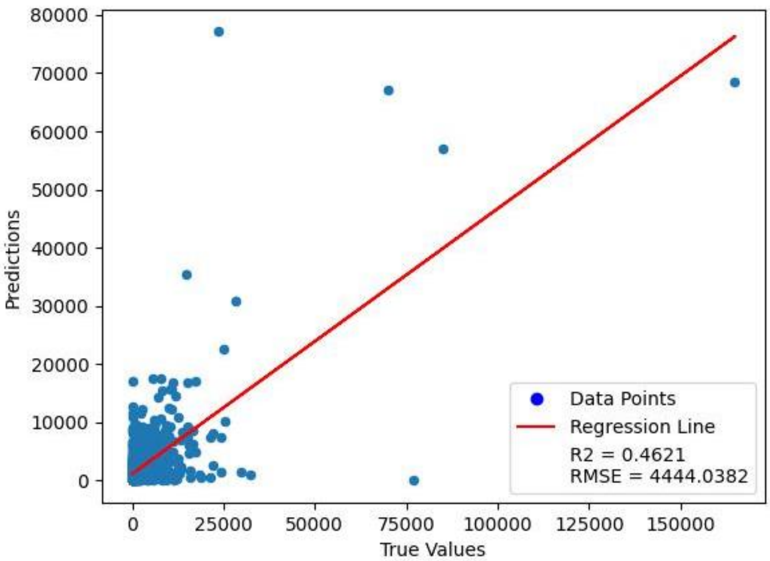


Fig 16. The Accuracy of RF model of Parking Infrastructure

And in general, parking meters have more significant impact on the number of parking violation issues than parking lots. Furthermore, an intriguing phenomenon has been observed: the quantity of parking meters located near streets exerts a more considerable influence on parking violation occurrences compared to the number of parking meters situated in larger, more expansive areas. Conversely, the number of parking lots located in larger areas appears to have a greater impact on parking violations than the number of parking lots situated in close proximity to streets.

In addition to analyzing the significance of various parking infrastructures at different distances from the streets, as previously established, partial dependency plots (PDP) can be employed to elucidate the dependence between the target response and a set of input features of interest (Zhao et al., 2024) on parking violation occurrences. These plots can also identify the inflection points that characterize the impact of these predictors. The partial dependence function is defined as follows:

$$\hat{f}_{x_s}(x_s) = E_{x_c}\left[\hat{f}(x_s, x_c)\right] = \int \hat{f}(x_s, x_c)\, d\mathbb{P}(x_c) \quad (1)$$

Each variable S has its own partial dependence function $\hat{f}_{x_s}$, which can calculate the average value of $\hat{f}$. x_s is fixed and varies over its marginal distribution $d\mathbb{P}(x_c)$. Both $\hat{f}$ and $d\mathbb{P}(x_c)$ are uncertain (Zhao et al., 2024).

It is easy to interpret from the PDP result (Fig 17) that the number of parking meters within 200 meters, especially of those within 100 meters, around the streets have a positive correlation with the number of parking violation issues. This may because that more meters can indicate a higher demand for parking spaces. When spaces are limited, drivers may be more likely to park illegally (e.g., in no-parking zones or in front of fire hydrants) to secure a spot. Additionally, areas with more parking meters often have stricter enforcement policies. This means that parking attendants or enforcement officers are more likely to be present, leading to more tickets being issued. This also includes violations related to parking meters, such as failing to display the municipal meter receipt.

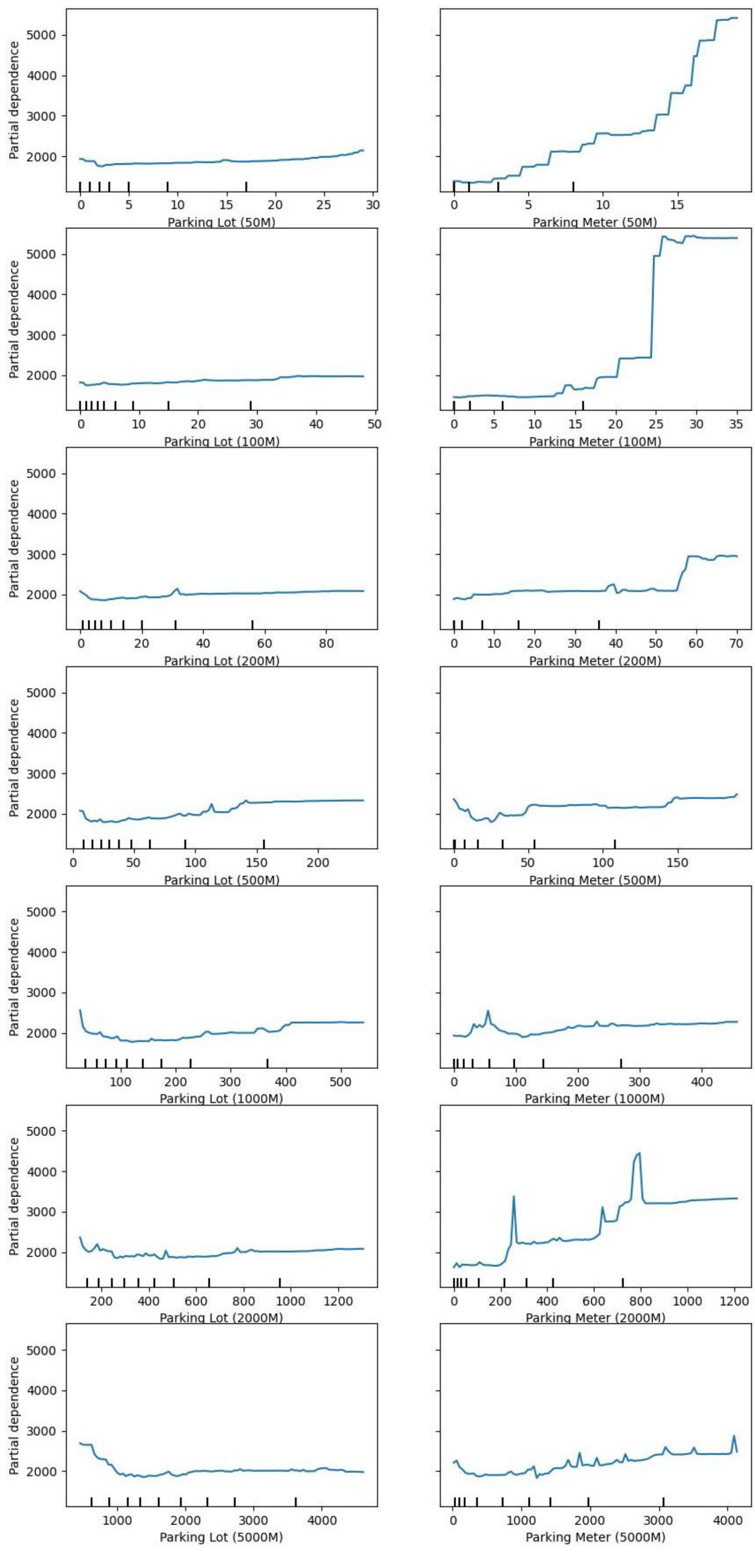


Fig 17. PDP Result of the Correlation between Parking Infrastructure and Violation Issues

But for the parking meters within 500 meters which has the most significant impact on the number of parking violation issues, the results show that the number of meters has a negative correlation with the number of parking violation issues when the number is less than 50. These streets may have appropriate roadside parking spaces with a low demand for parking spaces.

In general (Fig 17), the number of parking lots around streets has a negative correlation with parking violation issues, and it has a Marginal Diminishing Effect if the number exceed a certain value (e.g. 2 parking lots within 50 meters around the street). This phenomenon may be attributed to the notion that when there are sufficient parking lots to meet the appropriate demand, the incidence of parking violations tends to be low. However, once the number of parking lots surpasses a certain threshold—typically indicative of high parking demand in the adjacent streets—there may be a lag in the growth of parking lot availability relative to demand. For instance, if all nearby parking lots are consistently occupied, drivers may resort to illegal parking, regardless of the number of available parking spaces.

According to the result of RF Regression (Fig 15), the parking lots within 5000 meters around the streets has a significant impact on the number of parking violation issues. And the number of it has a negative correlation with the parking violation issues (Fig 17). But counterintuitively, the changes in the number of parking lots have almost no correlation with the changes in the number of parking violation issues (Fig 18). And if the violation issues number was separated into negative and positive parts and do simple regression again (Fig 19), the results show that the increasing number of parking lots can both increase and decrease the number of violation issues.

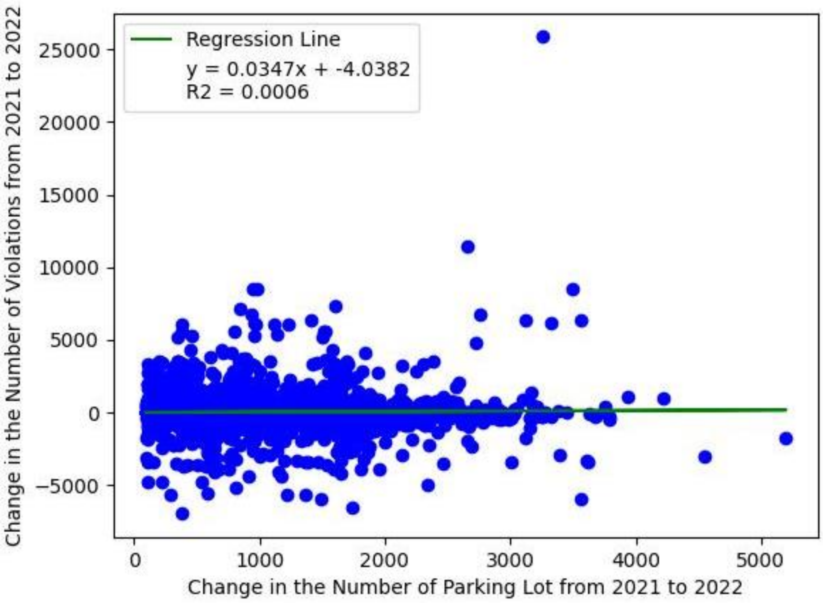
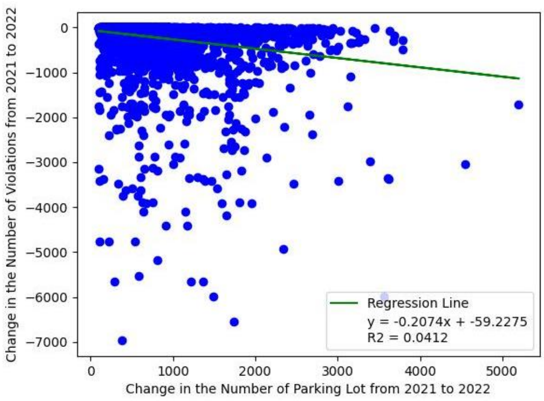


Fig 18. Correlation between Changes in the Number of Parking lots and Violation Issues from 2021 to 2022



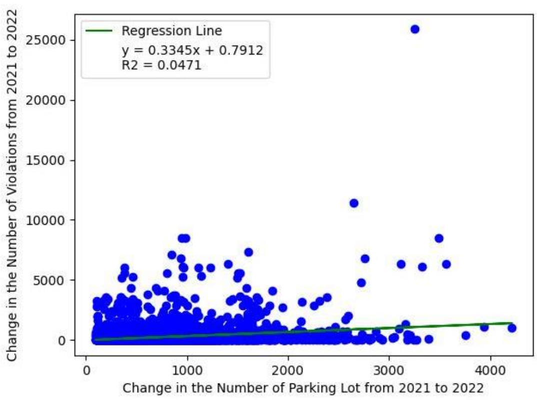


Fig 19. Correlation between Changes in the Number of Parking lots and Violation Issues (Separate by Negative and Positive) from 2021 to 2022

The combination of these phenomenon indicates that the number of parking lots and parking meters may have limited impact on the number of parking violation issues, especially when the number of parking infrastructure around the streets exceeds a certain value.

In summary, while parking infrastructure plays a role in violation rates, its impact may be limited when the number of facilities exceeds specific levels, indicating a complex relationship between parking availability and compliance.

Relationship with Land Use

To analysis the influence of land use around the street in parking violation issues, the buffer was used to calculate how many land uses (resident, public, transportation, industry, and commerce) are near streets. The buffer was set as dishes shape and the distances are set in four categories, 50 meters, 100 meters, 200 meters, 500meters, 1000 meters, 2000 meters and 5000 meters. And RF regression is used to analysis the importance of the type of land uses and the distance between the streets and them to the parking violation issues.

The result of data from 2021 to 2023 (Fig 20) shows that commercial land within 1000 meters around the streets have the most significant infection on the parking violation issues, while the impact of other land use like industry land within 5000 meters and public land within 500 meters is relatively weaker. And resident land within 5000 meters also has some impact on the parking violation issues.

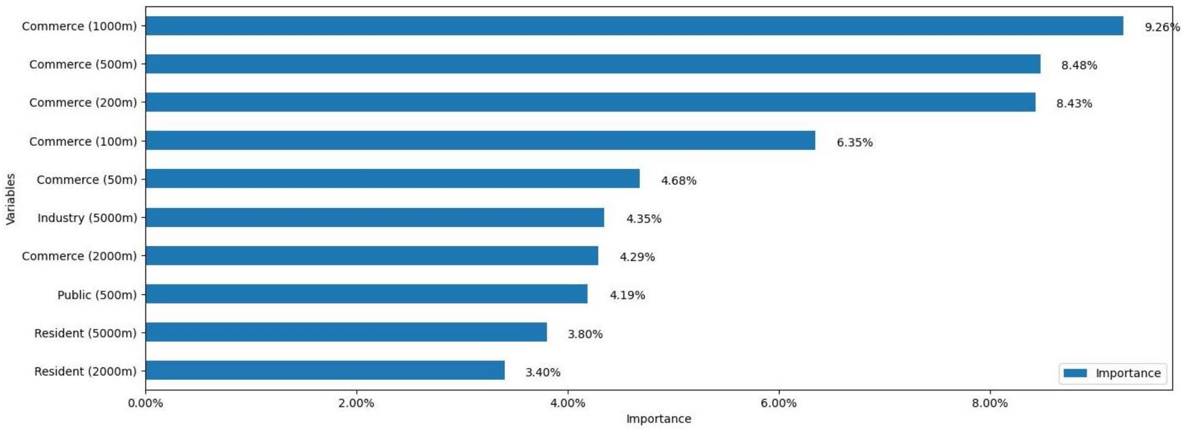


Fig 20. The Influence Degree of Top Ten Land Use from 2021 to 2023

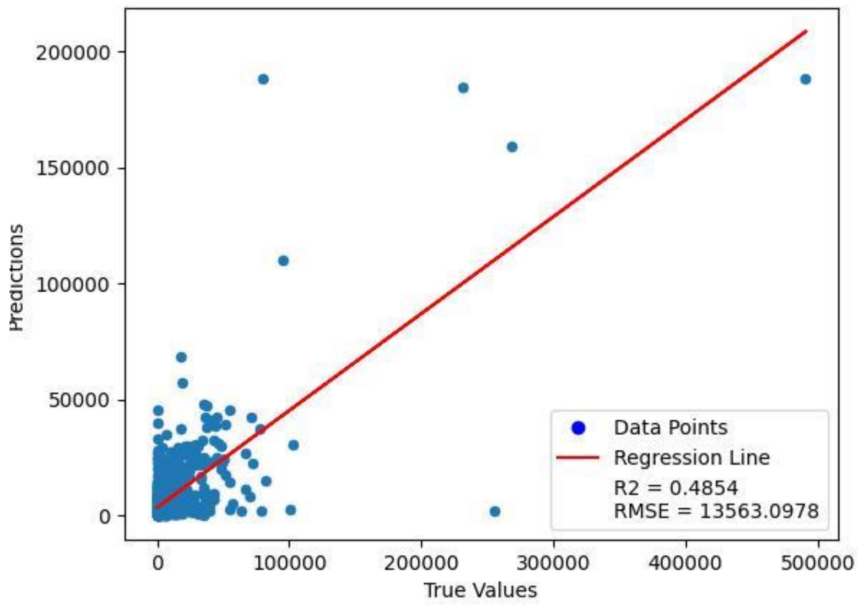


Fig 21. The Accuracy of RF model of Land Use from 2021 to 2023

An analysis of the influence of land use over the years (Fig 22 to Fig 24) reveals that the patterns of parking violations associated with different land uses have not been significantly affected by the passage of time and the infection of COVID-19. The number of commercial lands has the most significant impact on the number of violation issues.

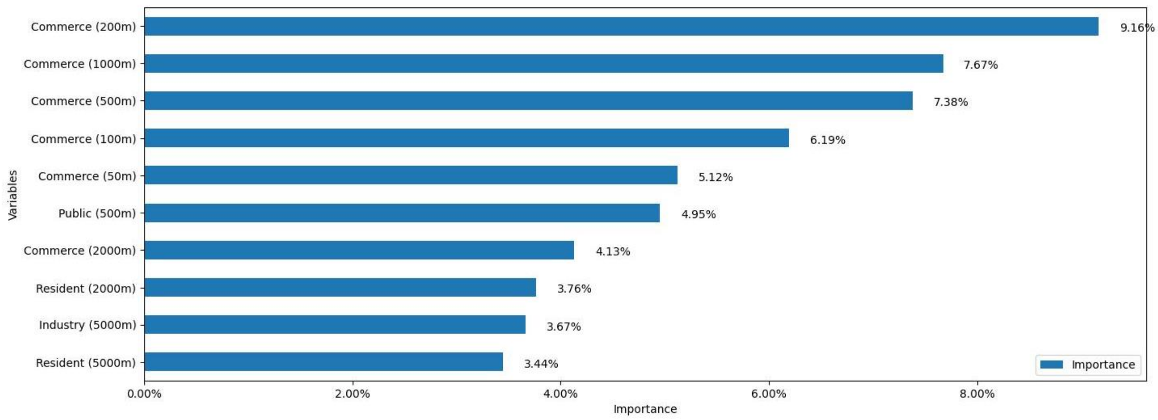


Fig 22. The Influence Degree of Top Ten Land Use in 2021

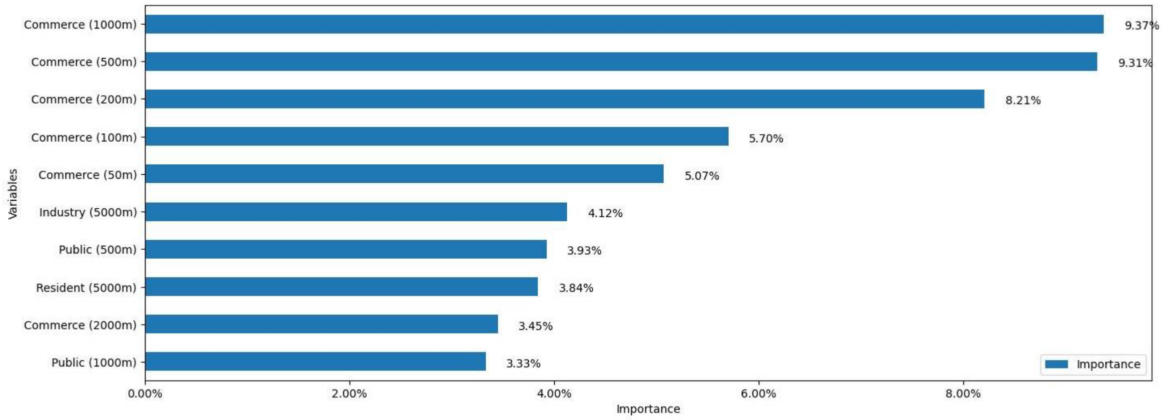


Fig 23. The Influence Degree of Top Ten Land Use in 2022

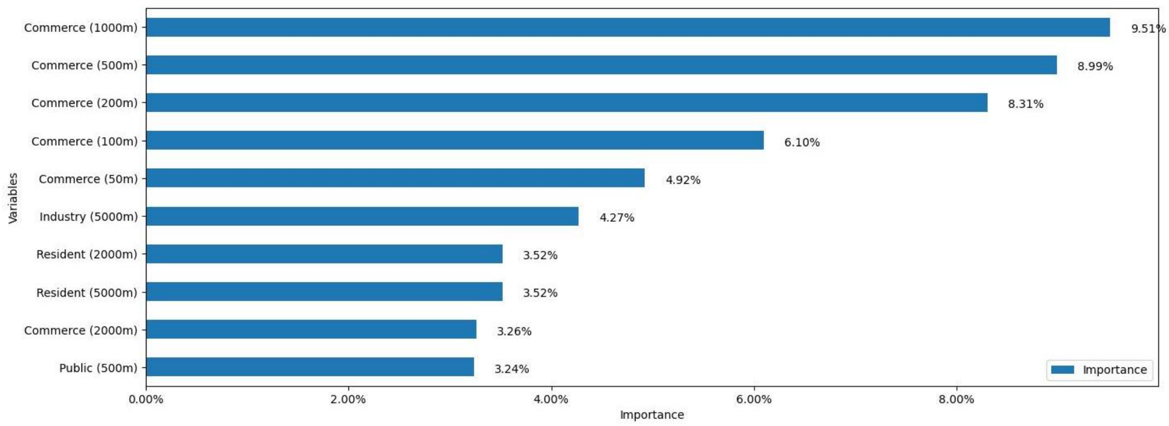


Fig 24. The Influence Degree of Top Ten Land Use in 2023

Utilizing PDP to understand the marginal effects of individual predictors of land use on the parking violation issues (Fig 25), it can be found that the number of commercial lands and transportation lands have a significant positive correlation with the number of parking violation issues no matter how far it is from the streets.

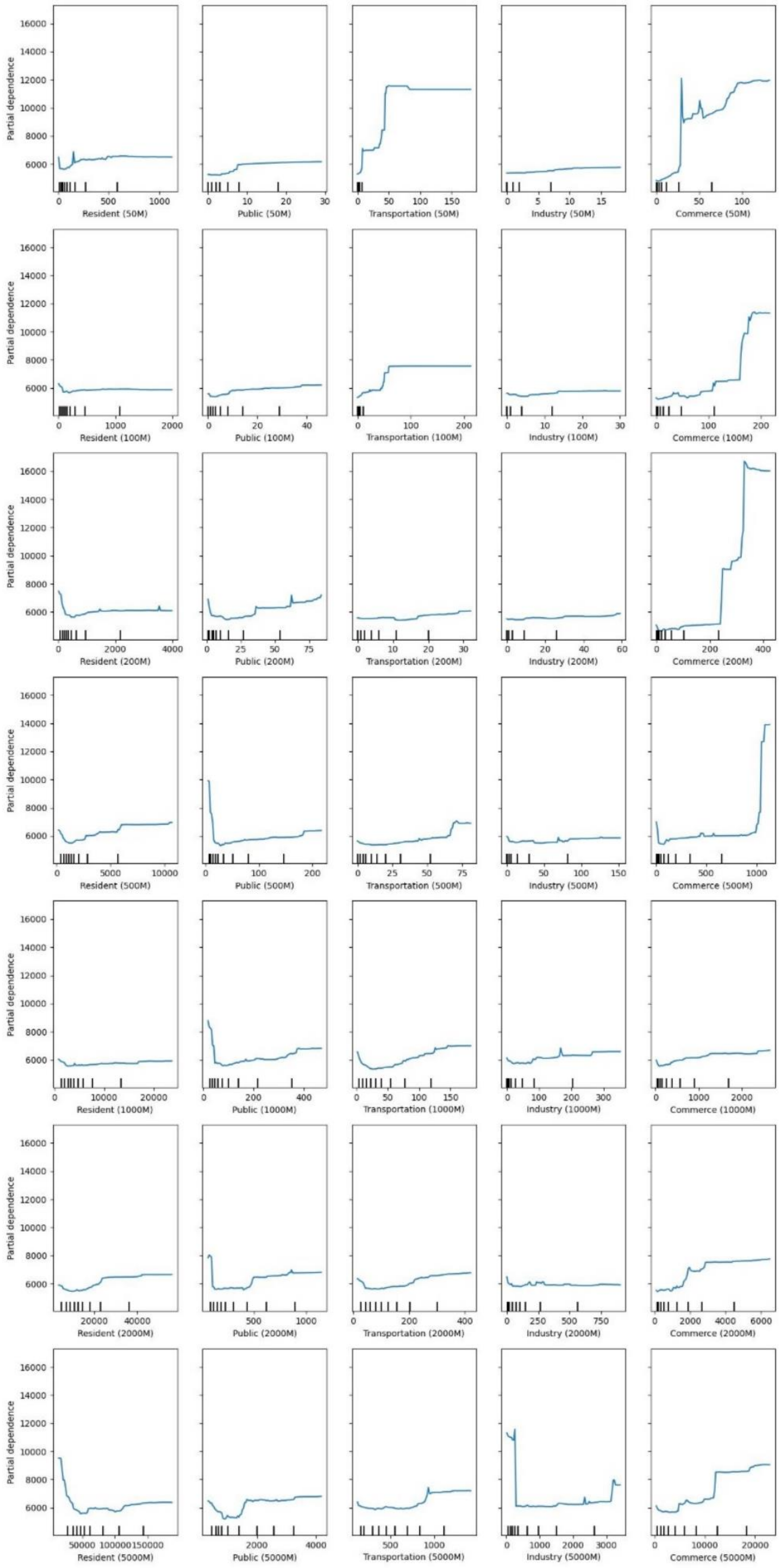


Fig 25. PDP Result of the Correlation between Land Use and Issues from 2021 to 2023

This relationship between the number of commercial lands and issues is intuitive; as the density of commercial establishments in proximity to the streets increases, the demand for parking also rises, consequently leading to higher rates of parking violations.

The findings regarding the relationship between parking infrastructure and violations, as previously discussed, indicate a positive correlation between the number of parking meters and the frequency of parking violations. This correlation may be attributed to the notion that an increased number of parking meters signifies a heightened demand for parking spaces. The strong linear correlation between the quantity of commercial lands and the number of parking meters, as evidenced by an R-squared value of 0.9406 (Fig 26), and the positive correlation between commercial lands and parking violation issues, further substantiates this hypothesis. In contrast, the relationship between the number of parking lots and the extent of commercial land shows a weaker linear correlation (Fig 26). It is important to note that parking lots can accommodate significantly more parking needs legally compared to parking meters. This capacity suggests that parking lots are more effective in alleviating issues related to insufficient legal parking availability than parking meters. Consequently, while the number of parking lots has slight liner correlation with the number of commercial lands, it is still associated with a negative correlation to parking violation occurrences. This conclusion aligns with the observation that the number of parking lots does not exhibit a purely positive correlation with the frequency of parking violations. This distinction highlights the complex dynamics of parking infrastructure and its implications for urban planning and policy development.

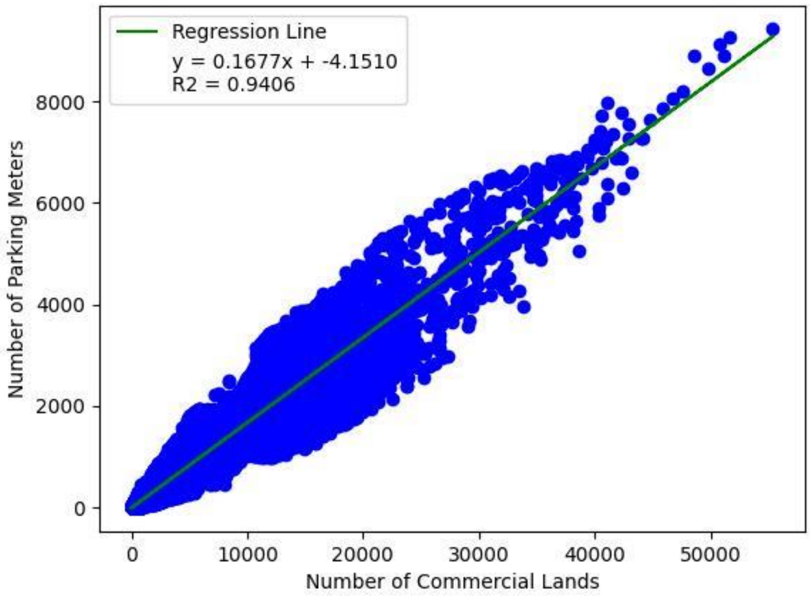


Fig 26. Relationship between the Number of Parking Meters and Commercial Lands

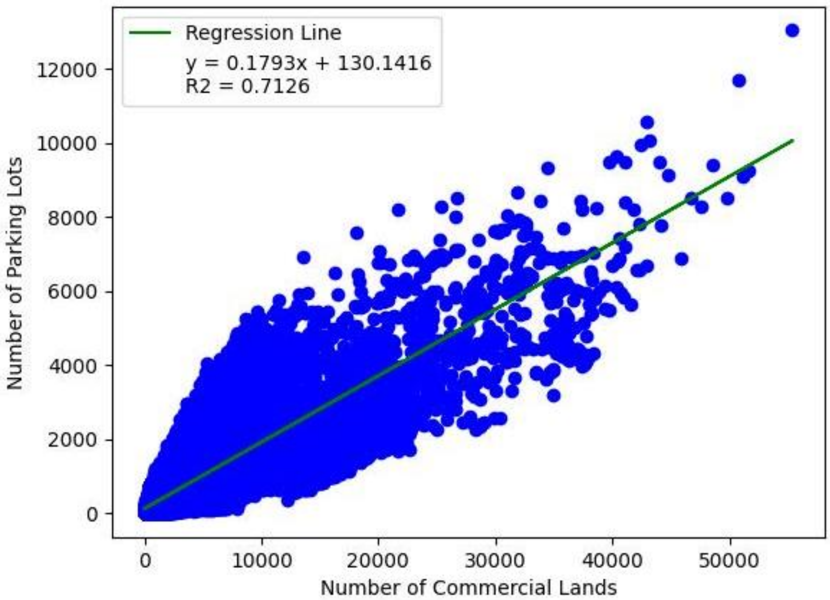


Fig 27. Relationship between the Number of Parking Lots and Commercial Lands

The observed positive correlation between the quantity of transportation infrastructure and the incidence of parking violations may initially seem counterintuitive. Typically, an expansion of transportation infrastructure is associated with improved public transportation options, which can lead to a decreased reliance on private vehicles and, consequently, a reduction in the frequency of illegal parking. However, the strong linear correlation between the quantity of commercial properties and the number of transportation facilities, as indicated by an R-squared value of 0.8335 (Fig 28), demonstrates that areas with improved transportation lands often undergo increased development and commercial activity. This influx can attract a greater number of visitors and residents, thereby elevating the demand for parking and potentially resulting in a higher incidence of violations as individuals seek convenient parking options. The reduction in the number of private vehicles facilitated by transportation infrastructure is insufficient to mitigate the influx of vehicles associated with commercial land use.

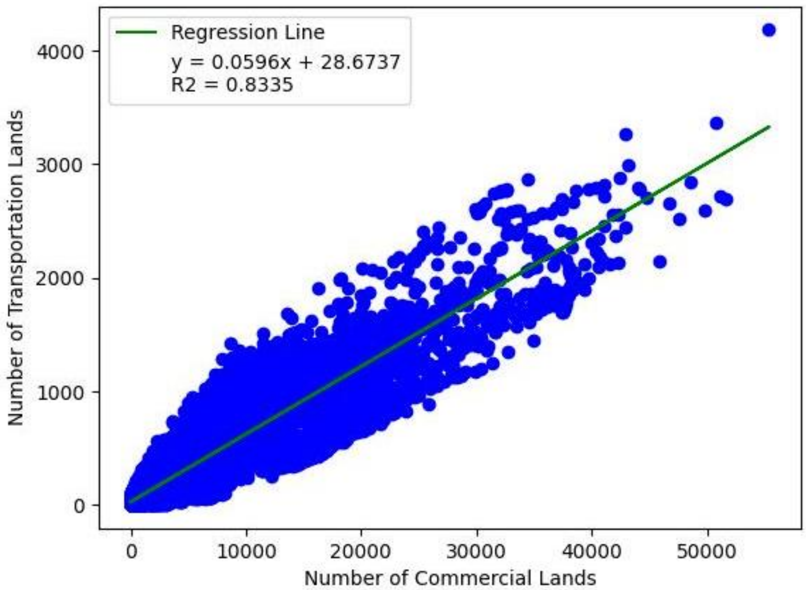


Fig 28. Relationship between the Number of Transportation Lands and Commercial Lands

For the industry land. The number of industry lands has no significant infection on the number of parking violation issues until the search range expands to 5000 meters. In this context, the quantity of industrial land adjacent to the streets exhibits a negative correlation with the number of issues up to a threshold of 1,000. This pattern resonates with Kawamura et al.'s (2014) observation on truck parking violations, which is that there is no significant association between industrial or transportation businesses and high truck parking violations, and fewer violations in long-established industrial zones due to street design and low density. Beyond the threshold, the correlation shifts to a positive relationship. This phenomenon may be attributed to the fact that, when the density of industrial buildings is optimal, an increase in industrial land use corresponds to a decrease in other types of land use. Furthermore, the nature of industrial properties necessitates a lower demand for public parking, resulting in a reduced need for parking spaces. However, once the density surpasses a certain threshold, the demand for public parking—driven by the needs of employees and transportation—exceeds the capacity of the available parking facilities in the surrounding area, ultimately leading to an elevated rate of parking violations.

Public lands exhibit similar trends; however, the threshold for the search radius decreases to 200 meters. This phenomenon may be attributed to the higher demand for public parking associated with public lands compared to industrial areas. Specifically, there tends to be a greater influx of individuals from outside the local area seeking to access public facilities located on public land, as opposed to those visiting industrial facilities situated in industrial zones. Additionally, a notable distinction arises in that the incidence of parking violations surrounding public lands decreases more significantly than that observed in industrial areas prior to the density of public lands reaching the threshold of density. This may be because that the public lands itself can provided some public parking lots to resolve some parking violations.

And the number of residential lands also shows a negative correlation with the number of parking violation issues no matter how far it is from streets before it meets the threshold of density. Typically, areas characterized by low residential density tend to have adequate parking facilities, such as private garages. Conversely, regions with high residential density are often comprised of apartment complexes, which offer limited parking availability. These characteristics align with the observed patterns of parking violations. Kawamura et al.'s (2014) observation on truck parking violations, which is that high residential turnover rates also increased parking violation numbers, can also prove this pattern.

Overall, the findings underscore the complex interplay between land use types and parking violation dynamics, emphasizing the need for strategic urban planning to mitigate violations.

Conclusion

This study has provided a comprehensive analysis of parking violation patterns in New York City from 2021 to 2023, revealing significant trends and insights that can inform future parking management strategies. The data indicates a notable increase in parking violations over the years, with non-local vehicles contributing high percents of these infractions. This trend highlights the need for targeted interventions aimed at non-local drivers to help them to avoid violating parking regulations.

Temporal analysis indicates that parking violations reach their peak during the summer months, suggesting that heightened activity during this period may result in increased demand for parking. Additionally, the number of parking violations also stays high during morning rush hour. The analysis of relationship between driver behavior and violation types also demonstrates that the time of violation significantly influences the types of parking violations, particularly during the morning and evening rush hours. However, the analysis of issues number every hour also indicates that the number of violations recorded during the evening rush hour is considerably lower than that observed during the morning rush hour. These differences in the concentration of parking violations number during morning hours and evening rush hours may be associated with the operational hours of law enforcement agencies. Nevertheless, the relationship between the working hours of law enforcement agencies and the number of tickets issued per hour requires further investigation. Overall, it would be better for law enforcement agencies to implement a more effective arrangement of road patrols after 1:00 PM. These insights can inform city planners and enforcement agencies in optimizing their resources and strategies during peak periods.

Geographically, the concentration of violations in high-density areas, particularly in popular tourist destinations such as southern Manhattan and high-density resident areas, underscores the challenges of parking management in urban environments. The law enforcement agencies should pay more attention to these areas.

The analysis of violation relationship also identified other key factors influencing parking violations number, including vehicle characteristics, the impact of parking infrastructure and land use. The findings indicate that specific vehicle types, especially commercial vehicles, are more prone to engage in illegal parking. This underscores the necessity for tailored enforcement strategies that consider the distinct behaviors associated with various categories of vehicles. However, the model's accuracy is not remarkably high, suggesting that the relationships between vehicle characteristics and violation patterns are complex and challenging to predict. And the results also suggest that purely increasing the number of parking lots and parking meters is useless and not necessarily better.

The correlation between parking violations and proximity to parking meters and parking lots indicates that the availability of legal parking options has little help in reducing the number of parking violations, especially when the number exceed the marginal. Furthermore, the study highlighted the significant impact of land use on parking violations, with commercial areas showing a strong positive correlation. This relationship emphasizes the importance of considering land use patterns in parking management strategies, as areas with high commercial activity are likely to experience greater parking demand and, consequently, more violations. Therefore, the best way to reduce parking violations is to start from early urban planning. The high density of land use, especially high-density commercial areas, has a strong influence on parking violations, which is hard to mitigate by building new parking facilities.

In conclusion, the findings of this study underscore the necessity for a multifaceted approach to parking management in New York City. By leveraging big data analytics and machine learning models, city planners and policymakers can develop more effective strategies that address the underlying causes of parking violations. Future research should continue to explore the dynamic interactions between vehicle characteristics, land use, and parking infrastructure, as well as the potential for innovative technologies to enhance parking enforcement and compliance. Although preliminary analyses indicate that COVID-19 had a minimal impact on the patterns of violations, it remains a subject that requires continuous monitoring and analysis through the establishment of long-term tracking of future data.

Ultimately, a more efficient and equitable parking management system can contribute to improved urban mobility, reduced congestion, and a better quality of life for residents and visitors alike.

Appendix

Analysis Source Code: https://github.com/LeiterStosszahn/Course-Project-in-PolyU/

Reference

Azodi, C. B., Tang, J., & Shiu, S. H. (2020). Opening the black box: interpretable machine learning for geneticists. *Trends in genetics*, *36*(6), 442-455. https://doi.org/10.1016/j.cjca.2021.09.004

Bastani, O., Kim, C., & Bastani, H. (2017). Interpreting blackbox models via model extraction. *arXiv preprint arXiv:1705.08504*. https://doi.org/10.48550/arXiv.1705.08504

Chai, H., Ma, R., & Zhang, H. M. (2019). Search for parking: A dynamic parking and route guidance system for efficient parking and traffic management. *Journal of Intelligent Transportation Systems*, *23*(6), 541-556. https://doi.org/10.1080/15472450.2018.1488218s

Dansana, D., Patro, S. G. K., Mishra, B. K., Prasad, V., Razak, A., & Wodajo, A. W. (2024). Analyzing the impact of loan features on bank loan prediction using Random Forest algorithm. *Engineering Reports*, *6*(2), e12707. https://doi.org/10.1002/eng2.12707

Gao, S., Li, M., Liang, Y., Marks, J., Kang, Y., & Li, M. (2019). Predicting the spatiotemporal legality of on-street parking using open data and machine learning. *Annals of GIS*, *25*(4), 299-312. https://doi.org/10.1080/19475683.2019.1679882

Gulowaty, B., & Woźniak, M. (2021, July). Extracting interpretable decision tree ensemble from random forest. In *2021 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE. https://doi.org/10.1109/IJCNN52387.2021.9533601

Han, L. D., Chin, S. M., Franzese, O., & Hwang, H. (2005). Estimating the impact of pickup- and delivery-related illegal parking activities on traffic. *Transportation Research*

Record, *1906*(1), 49-55. https://doi.org/10.1177/0361198105190600106

Innerhofer-Oberperfler, R., & Augsten, T. N. (2004). Using Approximate String Matching Techniques to Join Street Names of Residential Addresses. https://bit.ly/3AbwEtC

Isles, P. D. (2024). A random forest approach to improve estimates of tributary nutrient loading. *Water Research*, *248*, 120876. https://doi.org/10.1016/j.watres.2023.120876

Karantaglis, N., Passalis, N., & Tefas, A. (2022). Predicting on-street parking violation rate using deep residual neural networks. *Pattern Recognition Letters*, *163*, 82-91. https://doi.org/10.1016/j.patrec.2022.09.023

Karantaglis, N., Passalis, N., & Tefas, A. (2024). Semi-supervised learning for on-street parking violation prediction using graph convolutional networks. *Neural Computing and Applications*, 1-10. https://doi.org/10.1007/s00521-024-10248-5

Kawamura, K., Sriraj, P. S., Surat, H. R., & Menninger, M. (2014). Analysis of factors that affect the frequency of truck parking violations in urban areas. *Transportation Research Record*, *2411*(1), 20-26. https://doi.org/10.3141/2411-03

Kuo, L. C., & Lin, H. Y. (2024, June). Illegal Parking Detection Based on Multi-Task Driving Perception. In *2024 IEEE Intelligent Vehicles Symposium (IV)* (pp. 1865-1870). IEEE. https://doi.org/10.1109/IV55156.2024.10588798

Liu, D., Sun, Y., Li, Z., Wang, Z., & Zhang, Y. (2024, February). HIGH-PRECISION DETECTION OF ILLEGAL PARKING USING DEEP LEARNING TECHNOLOGY. In *The 7th International scientific and practical conference "Professional development: theoretical basis and innovative technologies"(February 20-23, 2024) Paris, France. International Science Group. 2024. 427 p.* (p. 326). https://doi.org/10.46299/ISG.2024.

1.7

Madaan, M., Kumar, A., Keshri, C., Jain, R., & Nagrath, P. (2021). Loan default prediction using decision trees and random forest: A comparative study. In *IOP conference series: materials science and engineering* (Vol. 1022, No. 1, p. 012042). IOP Publishing. https:// doi.org/10.1088/1757-899X/1022/1/012042

Mesfin, B. G., Sun, D., & Peng, B. (2022). Impact of COVID-19 on urban mobility and parking demand distribution: a global review with case study in Melbourne, Australia. *International Journal of Environmental Research and Public Health*, *19*(13), 7665. https://doi.org/10.3390/ijerph19137665

NYC Department of Finance, Division of Treasury & Payment Services (Ed.) (2023). Annual report of New York city parking tickets and camera violations fiscal year 2023. NYC Department of Finance, Division of Treasury & Payment Services. https://bit.ly/4daVSG2

NYC Department of City Planning, Population Division (Ed.) (2023). New York city's current population estimates and trends. NYC Department of City Planning, Population Division. https://bit.ly/4euVO4U

Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How many trees in a random forest?. In *Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, MLDM 2012, Berlin, Germany, July 13-20, 2012. Proceedings 8* (pp. 154-168). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31537-4_13

Osoba, S. B. (2012). Appraisal of parking problems and traffic management measures in central business district in Lagos, Nigeria. *Journal of sustainable development*, *5*(8), 105.

http://dx.doi.org/10.5539/jsd.v5n8p105

Özen, F. (2024). Random forest regression for prediction of Covid-19 daily cases and deaths in Turkey. *Heliyon*, *10*(4). https://doi.org/10.1016/j.heliyon.2024.e25746

Singh, J. (2024). Autonomous Vehicles and Smart Cities: Integrating AI to Improve Traffic Flow, Parking, and Environmental Impact. *Journal of AI-Assisted Scientific Discovery*, *4*(2), 65-105. https://scienceacadpress.com/index.php/jaasd/article/view/186

Yang, C. F., Ju, Y. H., Hsieh, C. Y., Lin, C. Y., Tsai, M. H., & Chang, H. L. (2017). iParking–a real-time parking space monitoring and guiding system. *Vehicular Communications*, *9*, 301-305. https://doi.org/10.1016/j.vehcom.2017.04.001

Zhao, X., Yang, H., Yao, Y., Qi, H., Guo, M., & Su, Y. (2022). Factors affecting traffic risks on bridge sections of freeways based on partial dependence plots. *Physica A: Statistical Mechanics and its Applications*, *598*, 127343. https://doi.org/10.1016/j.physa.2022.127343