A Comprehensive Analysis of Parking Violations in New York City: Unraveling Patterns and Implications

Shubham Laxmikant Deshmukh
Department of Computer Science
Virginia Tech
shubhamd23@vt.edu

Sai Manas Rao Pulakonti
Department of Computer Science
Virginia Tech
saimanasraop@vt.edu

Sriharsha Vemuri
Department of Computer Science
Virginia Tech
sriharshavemuri7@vt.edu

ABSTRACT

This research study provides a thorough examination of parking infractions in New York City spanning the years 2019 to 2023, with specific emphasis on the subset of data from 2023. The study starts with an initial examination of the NYC parking tickets dataset through exploratory data analysis (EDA). This analysis investigates the distribution of tickets across different boroughs, temporal trends, and the overall financial consequences of parking offenses on the city. The study also examines the impact of the Covid-19 period on violation codes, ticket counts, and longitudinal patterns across different boroughs.

The second phase of the study presents a machine learning algorithm that can accurately forecast violation codes. This algorithm provides government entities with a helpful tool to improve revenue management and allocate resources more effectively. The prediction algorithms utilize past trends and variables such as location and time to aid in the early detection of possible infractions, hence facilitating effective and fair law enforcement.

The research offers a sophisticated comprehension of the changing urban dynamics and the capacity for data-driven approaches in public policy and governance.

KEY WORDS:

Parking violations, New York City, Exploratory Data Analysis (EDA), Temporal patterns, Financial impact, Covid-19 era, Machine learning, Violation codes prediction, Revenue management, Resource allocation, Neural Networks, Random Forest, Decision Tree, KNN

1. INTRODUCTION:

Urban regions are greatly affected by parking infractions, which have substantial consequences for city planning, income generating, and resource allocation. This research study provides a thorough examination of parking infractions in New York City spanning from 2014 to 2023, with specific emphasis on the subset of data from 2023. The study starts with an initial examination of the NYC parking tickets dataset, focusing on exploring the distribution of citations among different boroughs, identifying temporal trends, and assessing the overall financial consequences of parking offenses on the city. The study also examines the impact of the Covid-19 period on violation codes, ticket counts, and longitudinal patterns in different boroughs.

The study's second phase presents a machine learning model that can accurately forecast violation codes. This model serves as a significant tool for government agencies to improve revenue management and allocate resources more effectively. The prediction algorithms utilize past trends and elements such as location and time to aid in the early detection of probable infractions, hence facilitating efficient and fair law enforcement.

2. RELATED RESEARCH:

The SpotAngels blog [5] is very similar to this project, it analyzed New York City parking ticket data and found that between October 2020 and September 2021, there were over 4 million tickets issued totaling nearly $260 million in revenue. The most common violation was for street cleaning. According to SpotAngels' analysis, the neighborhoods with the highest rate of parking tickets per restricted parking spots are Kips Bay, Battery Park City, and Downtown Brooklyn, while the Theater District, Chelsea, and Jamaica had some of the most ticketed individual blockfaces, with over 300 tickets per spot. SpotAngels noted that their parking map app, which shows real-time parking rules, availability, and ticket risk, can help drivers avoid expensive tickets. The post also previewed that SpotAngels offers alerts when drivers receive new parking tickets.

Another research [6] presents an analysis of taxi flow data in Manhattan, NYC using various data mining techniques. The goals are to characterize locations based on taxi pickup/dropoff demand profiles, build a probabilistic graph model of taxi flows to uncover patterns, identify interesting or unusual locations in the graph, and detect spatio-temporal anomalies or events. The methods used include clustering geographic areas by activity profiles, constructing an average probabilistic flow graph, extracting features to find "locations of interest", role extraction to find common structural behaviors, and change detection on roles to identify outlier days/locations. The discovered knowledge can help with traffic management, infrastructure planning, and identifying events that significantly alter normal taxi flow patterns.

Another interesting paper[7] analyzes the parking violations of United Nations diplomats in New York City from 1997-2005 to develop a revealed preference measure of corruption. Prior to 2002, diplomats had immunity from parking enforcement. The authors find that unpaid parking violations are strongly correlated with existing country corruption ratings, suggesting persistent cultural norms related to corruption. After 2002, enforcement increased and unpaid violations dropped 98%, indicating legal enforcement also plays a major role. The results suggest both cultural norms and legal enforcement are important determinants of corruption.

3. METHODOLOGY:

3.1 Exploratory Data Analysis (EDA):

The exploration of the NYC parking tickets dataset from 2014 to 2023 initiates with meticulous attention to detail. Data preprocessing is undertaken to ensure the dataset's readiness, with a specific focus on the 2023 subset. The analysis delves into multiple dimensions, commencing with a panoramic review of ticket distribution across the five boroughs. Monthly variations are scrutinized to identify temporal patterns, and further insights are derived by investigating weekday and hourly ticket distribution.

Beyond the immediate spatial and temporal aspects, the study assesses the overall financial impact of parking violations on the city. This includes a nuanced examination of fine breakdowns across precincts, an exploration of the most prevalent reasons leading to ticket issuances, and a comprehensive analysis of issuing agencies. A notable emphasis is dedicated to unraveling the influence of the Covid-19 era on violation codes, ticket counts, and longitudinal trends across boroughs, providing a nuanced understanding of the evolving urban dynamics from 2019 to 2022.

3.2 Machine Learning Model Prediction:

Predicting violation codes through machine learning offers a valuable opportunity for government agencies to enhance revenue management and resource allocation via more proactive handling of traffic violations, such as parking tickets. This approach optimizes fine collection while enabling resource reallocation towards addressing more severe crimes. Prediction algorithms, using historical patterns and factors like location and time, assist in early identification of potential violations, enhancing enforcement and compliance. Overall, this data-driven method promotes efficient and equitable law enforcement, aligning with broader public policy and governance objectives.

3.2.1 Neural Networks

The model built here is the implementation of deep learning[4] concepts using the Keras library in Python. The model is an instance of an artificial neural network (ANN), which is a machine learning technique that employs linked nodes or neurons arranged in layers, drawing inspiration from the human brain. The model is trained using a dataset of traffic violations to forecast the violation code by considering multiple features, including registration state, plate type, issue date, vehicle body type, vehicle make, issuing agency, violation precinct, issuer precinct, violation hour, violation county, street name, and vehicle color.

Algorithm breakdown:

Data Preprocessing : CSV data loaded, columns selected, and split into 80% training and testing sets. Categorical variables are one-hot encoded for model compatibility.

Model Architecture[4] : Sequential model design with various layers:

- Input Layer : Receives input data with neurons matching input features.
- Dense Layers : Fully connected layers (256, 128 neurons) initialized using 'he_uniform' for weight initialization.

  Equation: y=activation (Wx+b)

  where `W` represents the weights, `x` is the input, `b` is the bias, and `activation` is the activation function applied to the layer.

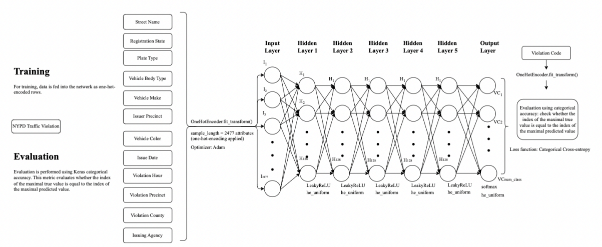
- LeakyReLU Layers : Activations introducing non-linearity and mitigating the dying ReLU problem.

  Equation: $f(x) = f(x) = \{x,\ if\ x > 0\ \alpha x,\ otherwise$

  where `α` is a small constant. In this case, `α` is set to the default value of 0.3.

- Batch Normalization Layers : Normalizes activations within each batch for improved neural network performance.
- Dropout Layers : Mitigates overfitting by randomly setting input units to 0 during training.
- Output Layer : Utilizes softmax activation for multi-class classification.

Softmax Function: $\sigma(z) = \frac{e^{zj}}{\sum_{k=1}^{K} e^{zk}}$



figure(i): Neural Network Architecture

Model Compilation : Adam optimizer for adaptive learning rate and categorical cross-entropy loss for multi-class classification.

Model Training : Fit method used with batch size 64 for 150 epochs. Early stopping and model checkpoint callbacks prevent overfitting and retain the best model.

Model Evaluation : Performance assessed on test data, printing the model's accuracy.

Model Prediction : Generates predictions on test data, converting predicted violation codes back to their original form post one-hot encoding.

This process enables the model to effectively predict violation codes based on diverse traffic violation data, aiding law enforcement in identifying and managing violations efficiently.

3.2.2 Random Forest

We built a comprehensive pipeline for a Random Forest Classifier, a robust ensemble method creating multiple decision trees and consolidating their predictions. This strategy elevates accuracy while curbing overfitting. Using the RandomForestClassifier from the sklearn.ensemble module, we instantiated a model with 100 decision trees (n_estimators=100) to ensure a diverse set. To maintain result reproducibility, we set a fixed random state (random_state=42). The model's performance was evaluated through 5-fold cross-validation, dividing the training set into 5 subsets. Four of these subsets were used for training, while the fifth served as validation. This cycle was repeated 5 times, varying the validation fold each time, offering a robust assessment of predictive performance. The formal representation of the random forest model involves the aggregation of numerous decision trees, consolidating their predictions for enhanced accuracy and reliability.

$$f(x) = \frac{1}{B}\sum_{b=1}^{B} f_b x$$

where (f_b(x)) is the (b)-th tree,and (B) is the total number of trees.

3.2.3 Decision Tree

We built a comprehensive pipeline for a Decision Tree Classifier model using scikit-learn. Decision Trees employ recursive feature partitioning, creating a hierarchical structure to sequentially determine class labels or outputs based on feature thresholds. This model predicts violation codes using a dataset and undergoes 5-fold cross-validation to assess its performance across diverse data subsets. The model's architecture involves recursively segmenting the feature space to make decisions at each node, eventually predicting the target variable. This process is iterated 5 times with different validation folds, ensuring a robust evaluation of predictive performance. The equation for Information Gain, fundamental in Decision Trees, facilitates the selection of the most informative features for optimal splitting.

$$Gain(T, X) = Entropy(T) - Entropy(T, X) \ ,$$

where (T) is the Target Variable, (X)=Feature to be split on, Entropy(T,X)= The entropy calculated after the data is split on feature (X).

3.2.4 K Nearest Neighbors

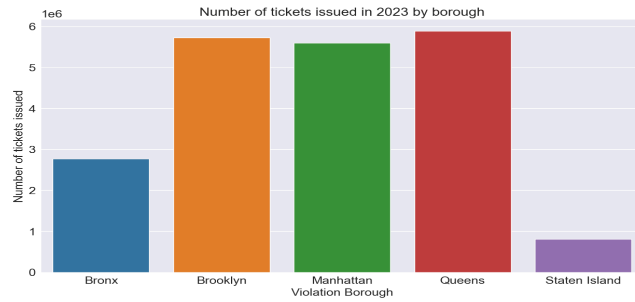
We've created a comprehensive pipeline for a K Nearest Neighbors (KNN) Classifier model using scikit-learn. KNN is a supervised learning method for classification and regression, predicting output based on the majority class or average value of its K nearest neighbors in the training dataset. This model predicts violation codes using a dataset and undergoes 5-fold cross-validation to evaluate performance across diverse data subsets. Repeating this process 5 times with varying validation folds ensures a reliable assessment of predictive accuracy.

4. EXPERIMENTS:

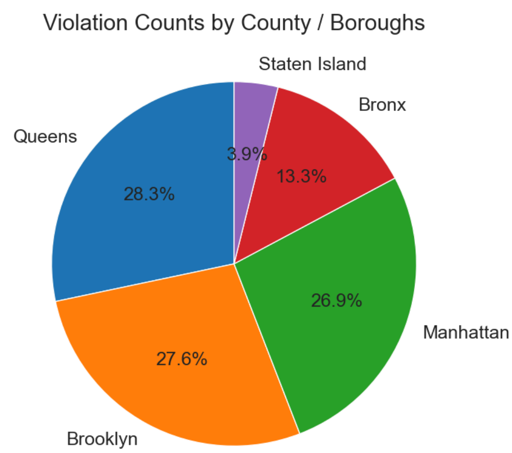
In this section, we discuss the various data exploratory techniques employed to analyze the dataset. Our investigation commences with an exploration of the distribution of tickets across boroughs in 2023, aiming to elucidate the number of tickets issued and identify potential spatial patterns.

4.1 Distribution of tickets borough wise:

To initiate our exploration, we examine the number of tickets issued in 2023 across different boroughs. This analysis serves as a foundational step in understanding the spatial distribution of parking violations and provides insights into potential hotspots or variations in enforcement practices.



figure(ii): Number of tickets issued in 2023 by borough



figure(iii): Violation Counts by County/Boroughs

In 2023, our analysis reveals that Queens emerged as the foremost issuer of parking tickets in New York City, while Staten Island recorded the lowest incidence. Queens, renowned for its status as one of the most densely populated areas in NYC with over 22,000 people per square mile, presents a unique urban landscape characterized by a scarcity of parking spaces and exorbitant costs associated with private parking garages, reaching several hundred and even thousands of dollars per month. This high population density, coupled with the inherent challenges of parking in Queens, likely contributes to the heightened issuance of parking tickets in this borough.

Conversely, Staten Island, characterized by a more rural ambiance with a population density just over 8,000 people per square mile, exhibits a fundamentally different urban infrastructure. Many homes in Staten Island feature driveways and garages, and shopping centers are often equipped with spacious parking lots —

features notably absent in the other boroughs of New York City. These distinguishing characteristics of Staten Island's urban layout may account for the comparatively lower incidence of parking violations, as residents and visitors benefit from more accommodating parking options, thus mitigating the need for extensive ticket enforcement measures.

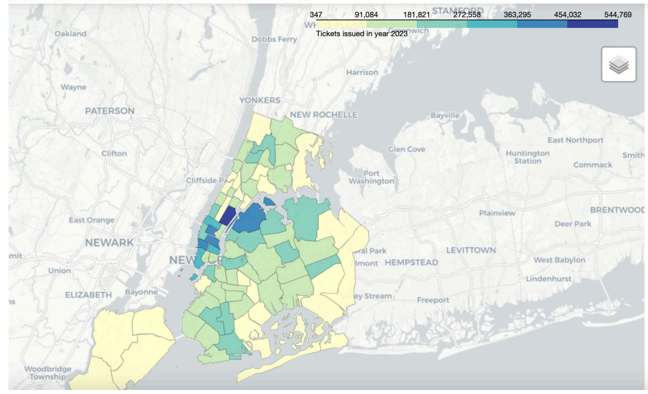
4.2 Distribution of tickets across the precincts

In our investigation of the distribution of parking tickets across precincts in 2023, we referenced the New York Police Department's public information office to establish valid precinct ranges for each borough. https://www.nytimes.com/2017/03/17/nyregion/nypd-precincts.html.The precinct ranges were identified as follows: 1-34 for Manhattan, 40-52 for the Bronx, 60-94 for Brooklyn, 100-115 for Queens, and 120-123 for Staten Island. Notably, precinct reassignments resulted in non-numbered precincts, such as the 22nd becoming the Central Park Precinct, the 14th transforming into Midtown South, and the 18th becoming Midtown North.

However, during our analysis, we encountered challenges, including over 9 million NaN entries for precinct information and a substantial number of precincts in each county that deviated from the standard precinct numbers. Data cleansing revealed that these irregular values were often associated with non-police entities, such as MTA, Parking, or court departments. After cleaning the data, the majority of tickets aligned with valid precincts for each borough.

To visualize this distribution, we obtained precinct GeoJson data from the NYC Open Data website and filtered out valid precinct numbers. The precinct boundaries are accessible at https://data.cityofnewyork.us/Public-Safety/Police-Precincts/78dh-3ptz. About 11 million tickets were found in valid precinct numbers, and we created a dictionary specifying the valid precinct range for each borough. This structured approach sets the stage for subsequent spatial analysis and visualization of parking ticket distribution across precincts in New York City.

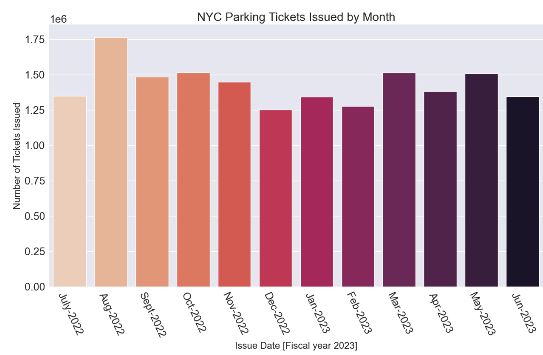
Following the above data analysis, we utilized the Folium library in Python to create a map illustrating the number of tickets issued in New York City for valid precinct numbers within the dataset. Folium, a powerful tool for data manipulation and visualization, enabled us to seamlessly integrate and visualize the information on a Leaflet map. This map provides a clear representation of the spatial distribution of parking tickets across valid precincts in each borough, offering valuable insights into the enforcement patterns and hotspots within New York City.



figure(iv): Tickets issued in New York City, in 2023.

4.3 NYC Parking Tickets Issued by Month

In our exploration of NYC parking tickets issued by month, we initially addressed challenges related to diverse date formats within the dataset. Through careful cleaning and preprocessing, we organized the temporal data using the "time_df" dataframe. Subsequently, our analysis aimed to predict the monthly distribution of parking tickets. This investigation into the temporal patterns of ticket issuance provides valuable insights into the variations and trends in parking violations over the course of the year.

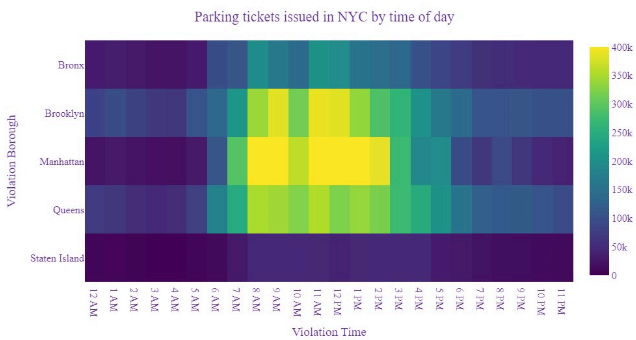


figure(v): NYC Parking Tickets issued by month.

Our analysis of the NYC parking ticket data reveals that no singularly dominant month stands out prominently. However, upon examination, we observe that August emerges with the highest ticket issuance, surpassing 1.75 million tickets for the month. In contrast, December records the lowest number of tickets issued, with nearly 1.24 million. This observed dip in December may be attributed to a decrease in enforcement activities during the holiday season, possibly influenced by reduced vehicular traffic or people vacationing. These nuanced variations in monthly ticket issuance contribute to a comprehensive understanding of the temporal dynamics surrounding parking violations in New York City.

4.4 What hours of the day are most tickets issued:

Continuing our temporal analysis using the "time_df" dataframe, we delved into understanding the hours of the day when parking tickets are most frequently issued. To achieve this, we segmented our results by violation county, providing a granular view of ticketing patterns. The presentation of our findings was enhanced through the use of a heatmap, a visual representation that aids in identifying peak ticketing hours across different counties. Employing the "unstack" function after grouping by operation facilitated the creation of a structured presentation akin to a pivot table, enabling a comprehensive examination of the busiest hours for ticket issuance in New York City.



figure(iv): Parking Tickets in NYC by time of day.

Our analysis reveals a consistent trend in parking ticket issuance across all five boroughs, with the most tickets being issued during the hours between 7 AM and 3 PM. This pattern holds true uniformly, indicating a shared temporal dynamic in ticketing practices throughout New York City. Notably, Staten Island stands out in the heatmap, exhibiting comparatively fewer tickets issued regardless of the time of day. This distinction in ticketing frequency suggests a unique enforcement landscape in Staten Island, potentially influenced by distinct urban characteristics or community behaviors that contribute to lower ticket issuance rates across various hours.

4.5 What is the overall income that NYC makes on tickets:

To determine the overall income generated by parking tickets in New York City, we initiated a multi-step process. First, we retrieved a CSV file from NYC Open Data containing violation descriptions and associated fines for each violation code. The file's URL is 'https://data.cityofnewyork.us/api/views/ncbg-6agr/rows.csv?accessType=DOWNLOAD'. After acquiring the data, we renamed the relevant columns for consistency and merged the two dataframes. This comprehensive approach allows us to calculate the cumulative income generated by parking tickets in New York City, providing insights into the financial impact of parking violations on the municipal level.

```
merged_df.Fine.sum()
```

```
1565764240.0
```

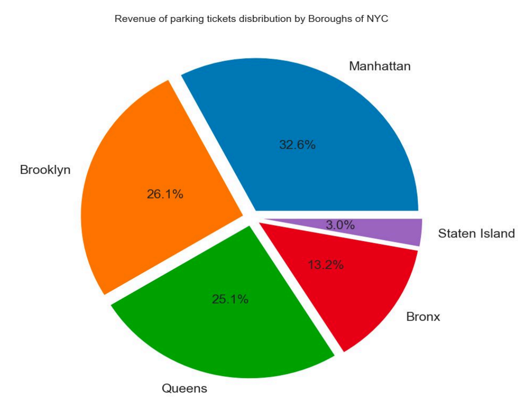
figure(vi): Cumulative fines from parking tickets in all five boroughs of New York City.

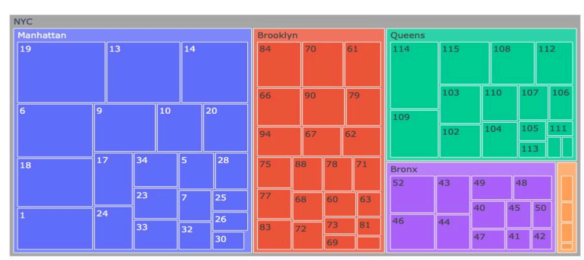
Our analysis reveals that the cumulative fines from parking tickets in all five boroughs of New York City amount to a substantial $1.5 billion. This substantial financial impact underscores the significance of parking violations as a revenue source for the city and emphasizes the economic implications of parking enforcement practices on a municipal scale.

4.6 What is the breakdown of fines across boroughs and precincts:

To obtain a detailed breakdown of fines across boroughs and precincts, our approach involved several key steps. First, we constructed a dataframe containing information on the borough, precinct, violation description, and the associated fine. Subsequently, we utilized this data to create a pie chart illustrating the distribution of fines in dollars and percentages across the different boroughs. This visual representation provides a clear overview of the relative contributions of each borough to the total fine amount.

Additionally, to enhance our exploration, we employed a treemap visualization that captures the hierarchical structure of fines, depicting the relationships between the city, boroughs, and precincts. This treemap serves as a comprehensive and visually informative tool for understanding the finer details of fine distribution across New York City's diverse boroughs and precincts.





figure(vii & viii): Revenue of Parking Tickets Distribution in NYC by boroughs.

The pie chart visually represents the percentage breakdown of fines across each borough, offering a clear and concise overview of the relative contributions to the total fine amount. This graphical representation facilitates a quick understanding of the proportional distribution of fines among the five boroughs.

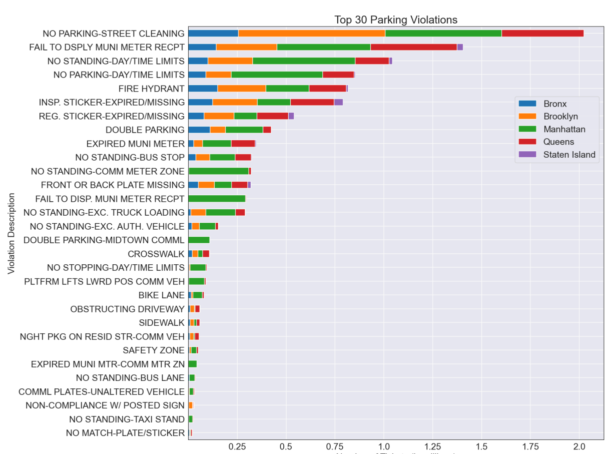
Furthermore, the treemap provides an interactive and hierarchical breakdown of fines for the entire city, individual boroughs, and specific precincts. Hovering over different sections of the treemap reveals detailed information, offering insights into the fine distribution at various levels. This interactive visualization enhances the user experience by providing nuanced information about the contribution of each borough and precinct to the overall fines imposed in New York City.

4.7 What are the most common reasons tickets are issued:

In our exploration of the most common reasons tickets are issued, we followed a systematic process. First, we merged the dataframe with violation descriptions to associate each ticket with its corresponding violation type. Next, we strategically grouped the data for effective plotting, ensuring clarity and relevance in the presentation.

To provide context and regional insights, we correlated these ticket categories with the respective boroughs/counties. The analysis focused on identifying and highlighting the top 30 ticket types out of a total of 97.

This approach allows for a streamlined and informative representation of the primary reasons for ticket issuances in New York City, shedding light on the most prevalent violations and their distribution across different boroughs.



figure(ix): Top 30 Parking Violations.

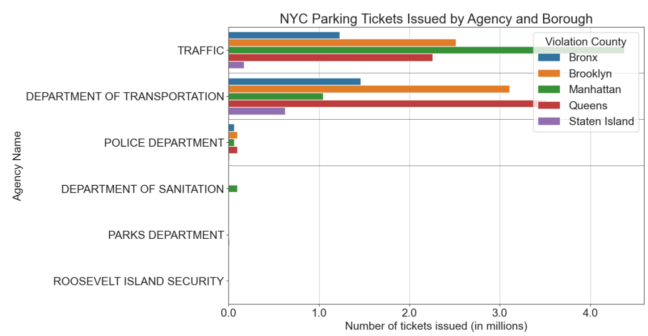
Our analysis reveals that the most common reasons for ticket issuances in New York City are related to street cleaning violations and expired meters. These two categories stand out prominently as the primary drivers of parking tickets across the city.

A notable observation during our analysis is the presence of a significant number of non-parking-related tickets in the dataset before eliminating NaN entries for precinct and county. These non-parking-related tickets are issued by cameras set up at intersections and school zones to monitor speed, representing a distinct category of violations within the dataset. This distinction underscores the importance of careful consideration and filtering in data analysis, especially when dealing with diverse sources of ticket information that may encompass various types of infractions beyond traditional parking violations.

4.8 What is the breakdown of 'Issuing Agencies' in the data:

Our exploration of the breakdown of 'Issuing Agencies' involved scraping the [NYC Open Data Website](https://data.cityofnewyork.us/City-Government/Parking-Violations-Issued-Fiscal-Year-2021/pvqr-7yc4) using the BeautifulSoup library to parse the HTML. The extracted information was then converted into a pandas dataframe. Subsequently, we merged this dataset with our 'Tickets Issued' and 'Violation Fines' dataframe for comprehensive analysis.

The resulting breakdown provides a detailed overview of the issuing agencies associated with parking violations in New York City. This analysis contributes valuable insights into the distribution of ticket issuances among different agencies, facilitating a nuanced understanding of the varied contributors to the overall parking enforcement landscape.



figure(x): NYC Parking Tickets issued by Agency and Borough.

Our analysis of the breakdown of 'Issuing Agencies' unveils a concentrated distribution, with four NYC agencies responsible for issuing a substantial 96.22% of all tickets, despite the presence of 34 agencies in total. Notably, Manhattan's tickets are predominantly issued by the traffic agency and police departments, underscoring their significant role in enforcement within the borough. Similarly, Brooklyn and Queens experience substantial ticketing from the Department of Transportation, emphasizing the varied roles played by different agencies in the parking enforcement landscape across the city. This concentrated distribution sheds light on the key contributors to the issuance of parking tickets in New York City and highlights the notable influence of specific agencies in different boroughs.

4.9 Analyzing a pattern in the Top 5 Violation Codes during Covid - 19 Era (2019- 2022)

During the Covid-19 pandemic, we focused on the top 5 parking violation codes by count, specifically related to parking tickets in

NYC. Our analysis aimed to uncover patterns in whether these top 5 violation codes changed before and after the pandemic. To assess this, we gathered datasets from fiscal years 2019, 2020, 2021, and 2022. Surprisingly, after filtering criteria for the analysis, we discovered that the top 5 violation code categories remained consistent throughout the pandemic, despite significant disruptions and changes in human behavior.

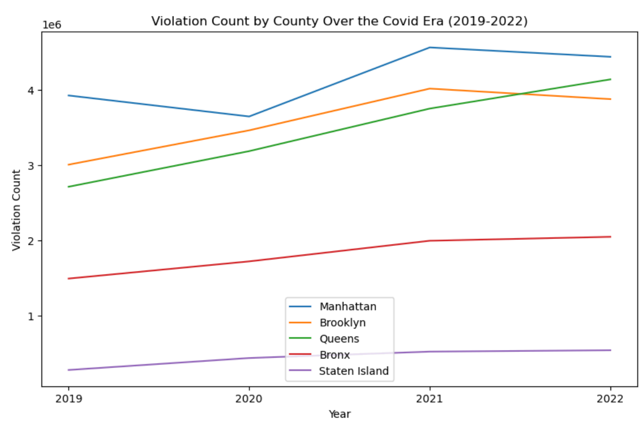
	Violation County	Bronx	Brooklyn	Manhattan	Queens	Staten Island	sum
Violation Code	Violation Description						
21	NO PARKING-STREET CLEANING	262906	682769	304362	436236	329	1686602
38	FAIL TO DSPLY MUNI METER RECPT	134163	254058	327383	398829	19447	1133880
14	NO STANDING-DAY/TIME LIMITS	98370	218853	503067	163091	12835	996216
20	NO PARKING-DAY/TIME LIMITS	104597	146596	439990	163444	6726	861353
40	FIRE HYDRANT	176206	259979	214398	182103	11050	843736

figure(xi). These are the Top 5 Violated Codes and their description of the ticket that remained constant from year 2019 to 2023 with the same order.

This observation challenges expectations, as one would anticipate a shift in violation patterns during a time when a large portion of the population was either isolated or hospitalized due to Covid-19. The persistent order and codes from 2019 to 2023 suggest an anomalous behavior that contradicts conventional expectations during a pandemic.

4.10 Distribution of tickets by borough over the years in the Covid Era (2019-23)

Again for finding out a pattern during the era of before, during and after the Covid -19 pandemic we are plotting a line plot for the number of tickets issued during the years of 2019 to 2022.



figure(xii): This is the line plot for the number of parking tickets issued by the Violation County / Boroughs during the Covid -19 era.

From the Fig you can observe that only in the case of Manhattan the number of tickets dropped during the peak of Covid -19 July 2020 era dropped as Manhattan had most of the Covid -19 cases in NYC [4]. But all other boroughs or Violation County followed the same trend as they were following from the time before Covid-19 started (2018-2019). This signifies an anomalous behavior as most of the people should have been quarantined for everyone's safety so most of them did not use any vehicle that frequently as in the year 2019 so the number should increase rather than increasing linearly with time.

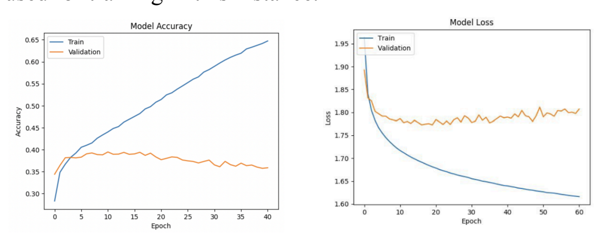
4.11 Comparison of Predictive Models

Model	Neural Network	Random Forest	KNN	Decision Tree
Accuracy	40.19 %	33.75 %	28.4 %	26.17%

figure(xiii): Comparison table of Predictive Models

In the comparison of the four models - Neural Network, Random Forest, Decision Tree, and K-Nearest Neighbors (KNN) - the Neural Network model has achieved the highest accuracy at 40.19%. This is followed by the Random Forest model at 33.75%, the KNN model at 28.4%, and finally, the Decision Tree model at 26.17%.

It's important to note that this comparison is based on a multi-label classification task involving 63 violation codes. In such a scenario, the probability of correctly guessing a violation code by chance is 1/63. Therefore, the accuracy levels achieved by the models, although seemingly low, are significantly higher than what would be expected by random guessing. The models were trained on a dataset of 100,000 rows. Training the models on a larger dataset could potentially improve their accuracy. However, due to system limitations, the full dataset could not be used for training in this instance.



figure(ix & x): Accuracy vs Epochs, Model Loss vs Epochs

DISCUSSIONS:

An examination of the NYC parking tickets dataset spanning from 2019 to 2023 provides valuable information about the geographical and chronological patterns of parking infractions and their economic consequences for the city. Queens has the highest frequency of parking fines being issued, most likely because to its elevated population density and limited availability of parking places, whilst Staten Island demonstrates the lowest occurrence of parking offenses. Tickets are predominantly issued between 7 AM to 3 PM in all boroughs, with August seeing the largest ticket issuance and December the lowest.

The utilization of a machine learning model enables the early detection of prospective infractions by analyzing past trends and considering aspects such as location and time. This process enhances the effectiveness of enforcement and compliance efforts. This data-centric approach facilitates effective and fair law enforcement, in line with wider public policy and governance goals. The Neural Network model has attained the utmost accuracy, reaching 40.19%.

Government agencies may utilize these findings to improve revenue management and resource allocation by proactive management of traffic offenses, such as parking citations, hence enhancing decision-making and planning. This strategy enhances the efficiency of gathering minor offenses while allowing for the redistribution of resources to focus on more serious crimes. Nevertheless, it is important to examine ethical factors to guarantee the responsible utilization of both the data and the model, while avoiding any discriminatory practices or violations of individuals' rights. The proceeds derived from parking citations ought to be used towards community welfare, such as enhancing public transit or infrastructure.

CONCLUSION:

In conclusion, the thorough examination of parking violations in New York City spanning from 2019 to 2023 has yielded useful observations on the allocation of citations across boroughs, temporal trends, and the total economic consequences of parking offenses on the city. The research also emphasized the impact of the Covid-19 period on infraction codes, ticket counts, and longitudinal patterns across different boroughs. The implementation of a machine learning model for forecasting violation codes has demonstrated its worth as a beneficial instrument for government agencies to improve revenue management and resource allocation. The prediction algorithms utilize past trends and elements such as location and time to aid in the early detection of probable infractions, hence facilitating efficient and fair law enforcement.

There is potential to enhance and broaden the machine learning models in future work, with the aim of improving forecast accuracy and integrating other elements that might impact parking infractions. Moreover, the study has the potential to be expanded to additional cities in order to analyze and juxtapose patterns and trends. It is also possible to examine the effects of policy changes on parking infractions. Finally, the research might investigate the possibilities of these data-driven techniques in other domains of public policy and governance, such as traffic control, urban design, and infrastructure advancement.

ACKNOWLEDGMENTS:

We would like to express our deep appreciation to Dr. Naren Ramakrishnan, our instructor for the course at Virginia Tech's NVC campus, for his important contributions and direction in shaping the design of the project's framework. His commitment to developing an encouraging learning atmosphere at the NVC campus has played a crucial role in the advancement of our research. We value the supportive academic environment at Virginia Tech, which has significantly influenced and enhanced the standard of our academic pursuits.

AUTHOR CONTRIBUTIONS:

We, **Sriharsha**, **Sai Manas Rao**, and **Shubham**, played equal and integral roles in our research paper. While everyone contributed equally on EDA, **Sriharsha** carefully crafted the Introduction and Conclusion sections, defining the research problem and summarizing key findings for Experiments. **Shubham, Sai Manas Rao** contributed substantially to the Related Research section, offering context through comprehensive citation and analysis. **Sai Manas Rao and Shubam**, led the development of the ML model and 'Violation Code' prediction model, shaping the Algorithms/Methodology and Experiments sections.

REFERENCES:

[1] Characterizing taxi flows in New York City - University of Illinois Chicago.(n.d.-a).https://www.cs.uic.edu/~urbcomp2013/urbcomp2015/papers/Characterizing-Taxi-Flows_Momtazpour.pdf

[2]Author links open overlay panelNikolaos Karantaglis, AbstractThe lack of available parking spaces can be among the most significant issues that can affect the quality of life of citizens in large cities. This has led to the development of on-street parking systems that typically ensure that parking spaces w, Djenouri, Y., Manessi, F., Selmi, Z., Tumrani, S., Wang, D., Yang, C.-F., Gao, J., Gao, S., & Glorot, X. (2022, October 4). *Predicting on-street parking violation rate using deep residual neural networks*. Pattern Recognition Letters. https://www.sciencedirect.com/science/article/abs/pii/S0167865522002938

[3]Wikimedia Foundation. (2023, November 22). *Timeline of the covid-19 pandemic in New York City*. Wikipedia. https://en.wikipedia.org/wiki/Timeline_of_the_COVID-19_pandemic_in_New_York_City

[4] *Tensorflow 2 quickstart for beginners : Tensorflow Core*. TensorFlow. (n.d.). https://www.tensorflow.org/tutorials/quickstart/beginner

[5]SpotAngels. (2022, April 6). *NYC parking tickets study, Tips and tricks*. SpotAngels Blog. https://www.spotangels.com/blog/nyc-parking-tickets-where-youre-likely-to-get-one-and-how-spotangels-can-pay-for-it/

[6] Characterizing taxi flows in New York City - virginia tech. (n.d.-b). https://people.cs.vt.edu/naren/papers/taxicabs-urbcomp15.pdf

[7] Corruption, norms, and legal enforcement: Evidence from ... - JSTOR. (n.d.-c). https://www.jstor.org/stable/10.1086/527495

[8] Johnson, P. (2014, June 4). *Open data saves New York City drivers from parking tickets*. Computerworld. https://www.computerworld.com/article/2695542/open-data-saves-new-york-city-drivers-from-parking-tickets.html

[9] Full article: Predicting the spatiotemporal legality of on-street ... (n.d.-d). https://www.tandfonline.com/doi/full/10.1080/19475683.2019.1679882

IMPORTANT LINKS

Github repo link:
https://github.com/Shubhs0411/A-Comprehensive-Analysis-of-Parking-Violations-in-New-York-City-Unraveling-Patterns-and-Implications
Kaggle dataset link used for EDA:
https://www.kaggle.com/datasets/shubhamdeshmukh17/nyc-parking-violations-updated-dataset-2020-2024
Dataset link for ML model training:
https://drive.google.com/file/d/1_wb7y59pv_sLiiwfmggd1CaPUdO4pH2z/view?usp=sharing