DSCI 5360 – Project Part II

Data Visualization for Analytics

Project Group: 8

Group Members:

- Pooja Kondagoli
- Jhansi Rani Minda
- Venkata Sai Jyothi Ravilla
- Chimata Nagabala Bhavani
- Sai Kiran Thirigabathi Ramesh
- Praneeth Vamshi Tarlapalli

1. Introduction and Background

Motivation:

There are many challenges within cities, such as congestion and mobility in urban areas with large populations, like New York City. Hence this topic is chosen making use of the NYC Parking Violation Data Set. Owing to it being available as a dataset and the issues arising from implementing such a policy, this study seems to be attractive research for analysis, showing how parking behaviors and enforcement arise. Moreover, openness to the public creates an environment of openness and invites people to develop the opinion about urban problems. Through the in-depth examination of this data source, the scientists can obtain multidisciplinary perspectives and be able to develop ameliorated parking management techniques.

Context and Relevance:

The project analyzing the NYC Parking Violation Dataset is vital to urban planning, transportation, and data science due to the presiding parking woes in highly dense cities such as New York. Parking congestion is a mobility, business activity and environmental sustainability problem. Researchers can obtain their insights from this dataset to shape evidence-based policies for city planners and policymakers. The aim of the project is to ensure transparency and community involvement, thereby improving urban efficiencies and sustainability.

Objectives and Queries:

Primary objectives:

The project aims to examine NYC parking violations, evaluate enforcement effectiveness, measure regulatory impact, investigate demographic discrepancies, and suggest effective management strategies. Below are a few key points.

- Data Exploration: Discover the general rival distribution and trends by analyzing the dataset of parking.
- Predictive Modeling: Create models for anticipating parking availability or probable finding of parking spot in facets at some times.
- Optimization: Looking for alternatives for the underuse and overuse management of parking facilities with ways that well including a correct pricing and placement
- Spatial Analysis: Draw up maps of the parking demand and supply patterns to detect the overloading and potential improvements.

- Customer Behavior Analysis: In our increasing knowledge of users' behavior that includes parking options, duration of stay and payment methods to achieve better user-experience.

Project questions:

The questions are directed to pinpointing violation trends, evaluating enforcement means, assessing regulatory impact, uncovering demographic disparities, and developing recommendations.

- What are the peak hours and days of parking demand in different areas of New York City?
- How does parking availability vary across different neighborhoods and boroughs?
- Are there any trends or patterns in parking violations or ticketing?
- What factors influence the likelihood of finding parking, such as time of day, day of the week, and weather conditions?
- What is the impact of parking regulations and pricing on parking behavior and congestion?
- How effective are current parking policies and regulations in managing parking demand and congestion?
- Are there any correlations between parking availability and socioeconomic factors like income or population density?
- What are the most common parking violations, and which areas have the highest rates of violations?
- How do different transportation modes (e.g., public transit, ridesharing) impact parking demand and availability?

Goal:

- Provide meaningful tips on parking behavior and enforcement dynamics.
- Contribute to the creation of congestion-free and more equitable parking management in New York City.
- Enable data-driven decision-making in urban planning and transportation policy.
- Understand parking practices to improve traffic flow and urban mobility.

Benefit:

Through the project, there is a huge scope of modification in urban planning and transportation system by providing parking violation trends and effectiveness of the

enforcement policies in NYC. This may result in a decreased level of congestion, better mobility becomes available, and increased quality of urban life. Also, by working on and regulations and the disparity in population bodies that can strengthen the fairness of the enforcement procedures. In general, the discovery of the project will be influential in designing improved and true parking management systems which will benefit us locally and worldwide.

Datasets

Origins and Support

The dataset of NYC Parking Violation Data Set is managed and runs under the budget of NYC Department of Finance; and this is an annual basis. The unit enforces parking regulations, issuing any violation tickets, and on a yearly basis reports the number of parking tickets issued in the city. A great deal of assistance also can be available through the government departments, academic institutions and advocacy organizations that are craving to leverage the dataset for the research and policy development in urban planning projects.

Dataset Owner: NYC OpenData

Data Collection: DOF Parking Violations Issued

Dataset Information (Agency): Department of Finance (DOF)

Date Created: July 22, 2016

Data Last Updated: March 16, 2024

(Parking Violations Issued - Fiscal Year 2024 | NYC Open Data (cityofnewyork.us)) ([ref:1,2])

Variable Insights:

The Raw Data set consists of 42 Columns capturing different features of New York Parking Violation as mentioned in the previous document (Project Part 1). Only the required columns for visualization have been filtered from the raw dataset and below are the columns used in our visualization.

1. Summons-Number: Unique identifiers of summons
2. Plate ID: Registered plate ID
3. Issue date: Issue date
4. Violation Code: Type of Violation

5. Vehicle Body Type: Body type of the vehicle
6. Issuing Agency: Issuing Agency code
7. Street-code1: Geographical Street code
8. Violation Time: Time violation occurred
9. Violation county: County of violation
10. House Number: Number of the house
11. Street name: Street name of summons issued
12. Intersecting street: Violation near the intersecting street
13. Violation legal code: Type of legal code
14. Violation Description: Description of violation
15. Year: Year of violation
16. Violation Code: Code for the violation
18. Fine Amount: Amount to be paid for breaking the law
19. All Other fines: Other fines
20. Violation Category: Category of violation.

These parameters are the keys to such visualizations as temporal trends in violation issuance, spatial distribution of violations across the city, demographic analysis of the vehicles involved, and comparison of enforcement effectiveness across different departments or periods. They also equip us to answer pivotal questions on parking behavior, enforcement dynamics, and policy implications that lie within the dataset.

Scope and Dimension:

The Dataset is downloaded from the City of New York open data website. The original yearly dataset consists of violation data for the fascial year followed by NYC Traffic department that is from previous year July-current year June of each year. So, we have downloaded dataset of 2023 and 2024 to consolidate to get complete data for year 2023.

2023-year data consists of over 10M records which is approximately equals to 3GB of data. It has 42 columns and approximately 10M rows.

Utility and Constraints:

The New York City Parking Violation Data Set assists projects to scale up by providing parking behaviors analysis and evaluation of enforcement. On the contrary, inaccurate data, the limited time scale, and privacy issues imply that the application is not immune from several difficulties. Despite these limitations, however, the significance of the dataset for the purpose of developing evidence-based urban mobility and parking management policies is still not constrained. Making sure that these limitations are delineated properly is of paramount significance to uphold the integrity of the findings.

Data Cleaning and Preprocessing:

- As raw data was for each fascial year, we had to extract the required month data from multiple fascial datasets.
- Final parking violation 2023 dataset consists of 15 columns and over 10M rows and Violation category dataset consists of 5 columns and 490 rows.
- Selected only required columns from the raw data set.
- Converted to required datatypes.
- We have converted the time column into readable format.
- As there were separate columns for data and time, we had to merge into one so that we could use this column for visualization.
- As there over 100 violation codes we have categorized into six violation groups:
    - Parking
    - Lane violation
    - Traffic signals
    - Vehicle meter expiration
    - Vehicle number plate expiration
    - Others
- We created aggregated datasets.
    - Violation counts per day.
    - Total fine amount per violation category
    - Issuing agency count
    - Top 2023 violation category
    - Violation time vs summons number (Ticket number)
    - County revenue
    - Issuing agency details
- All the above aggregated datasets are created using the python pandas. The complete aggregated datasets are in the below Jupiter notebook.
  [ref:13]

Design principles and improvement:

Below are the links which are used for design principles and key points are taken for these references and are used in our visualization.

[ref: 2,3,4,5,6]

Chart Design Principles | Hands-On Data Visualization (handsondataviz.org)

A Complete Guide to Line Charts | Atlassian

Basic_Charts.pptx: DSCI 5360 Section 007 - Data Visualization (Spring 2024 1) (instructure.com)

Basic Visualizations S.pptx: DSCI 5360 Section 007 - Data Visualization (Spring 2024 1) (instructure.com)

Gemini - chat to supercharge your ideas (google.com)

Line chart:

- Choosing an appropriate measurement interval
- Using a valid baseline
- Keeping an average line for better understanding
- Not plotting too many line
- Using a proper value for the color marks

Bar Graph.

- Include value annotations.
- Sorting the bars for better understanding
- Use color wisely.
- Use a common zero-valued baseline.
- Maintain rectangular forms for your bars.

Bubble chart.

- Scale bubble area by value.
- Limit number of points to plot
- Include a legend.
- Label the bubble and its value.
- Use color wisely.

Map Graph

- Making sure the coordinates are proper.
- Using the Text label from marks.
- If there are any unidentified areas, then filter or edit the data accordingly to match the same.
- Use color gradients to represent different data ranges (e.g., light to dark for low to high values).

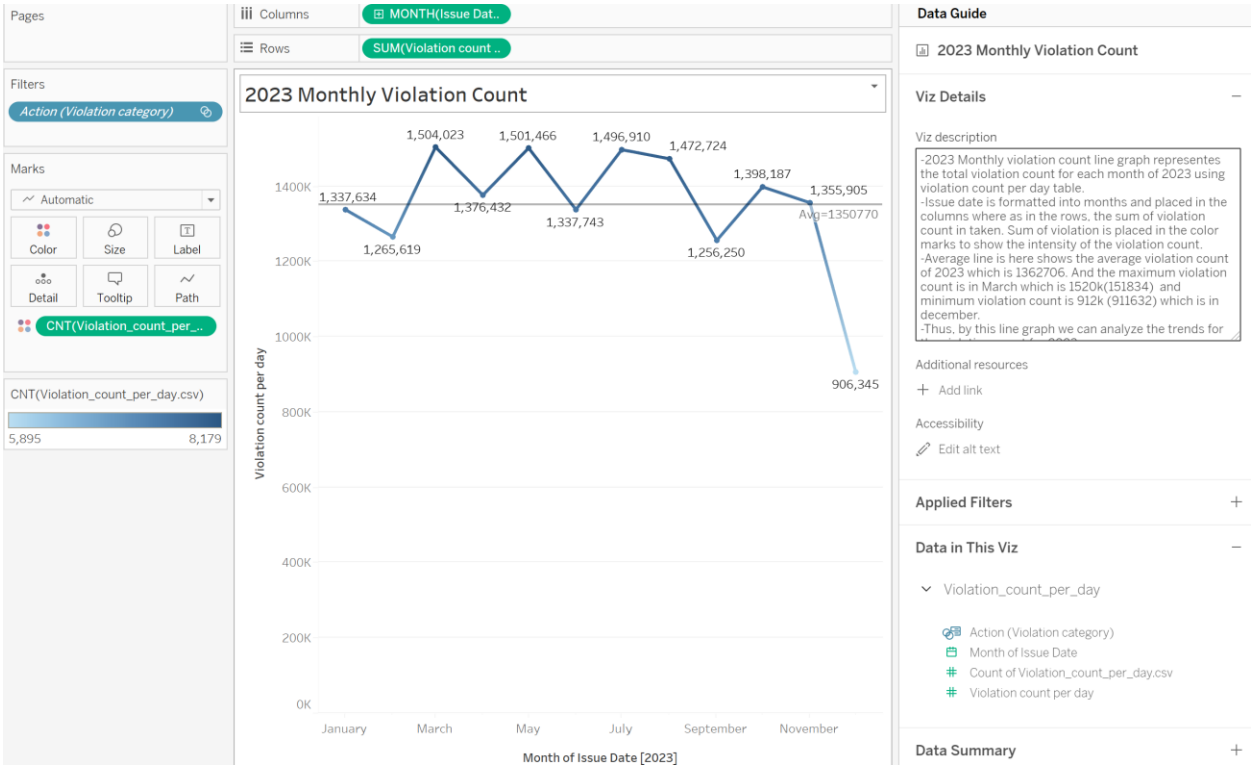
Highlight tables.

- Clarity & Focus: Maintain the core principles of highlight tables: easy readability and emphasis on important values.
- Data-Ink Ratio: Maximize the space for data and minimize chart elements. Focus on color gradients or conditional formatting within cells.
- Number Formatting: Apply consistent number formatting (e.g., number of decimal places, currency symbols) for easy comparison.

- Text Alignment: Left-align text and right-align numbers for consistent reading.
- Sorting & Filtering: Enable sorting by columns to highlight trends or filtering data to focus on specific categories.

Narrative Integration

2023 Monthly Violation Count



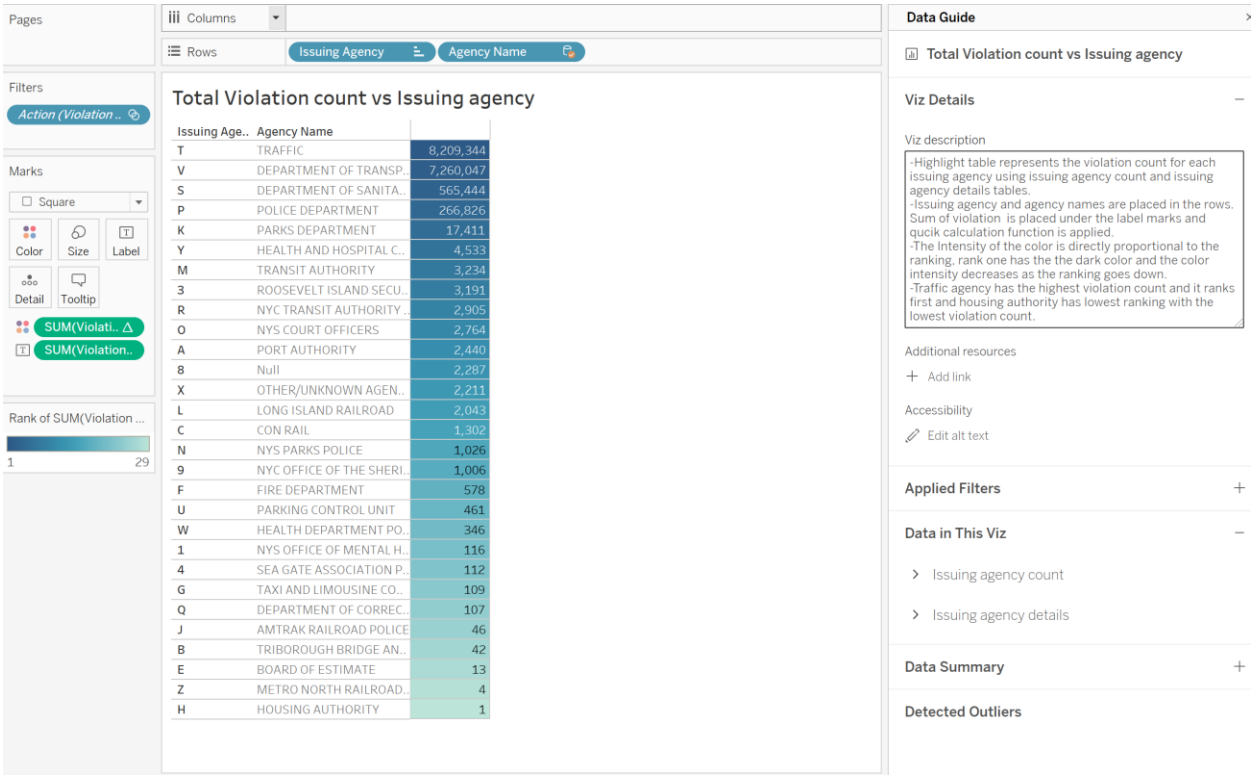
2023 Monthly violation count line graph represents the total violation count for each month of 2023 using violation count per day table. Issue date is formatted into months and placed in the columns whereas in the rows, the sum of violation counts in taken. Sum of violation is placed in the color marks to show the intensity of the violation count. The average line here shows the average violation count of 2023 which is 1362706. And the maximum violation count is in March which is 1520k (151834) and minimum violation count is 912k (911632) which is in December. Thus, by this line graph we can analyze the trends for the violation count for 2023.

Proposed Question

Describe the violation trends over the month? Which month has the highest and lowest violation count?

2023 Monthly violation count line graph represents the total violation count for each month of 2023 using violation count per day table. The average line here shows the average violation count of 2023 which is 1362706. And the maximum violation count is in March which is 1520k (151834) and minimum violation count is 912k (911632) which is in December. Thus, by this line graph we can analyze the trends for the violation count for 2023.

Total Violation count vs Issuing agency.



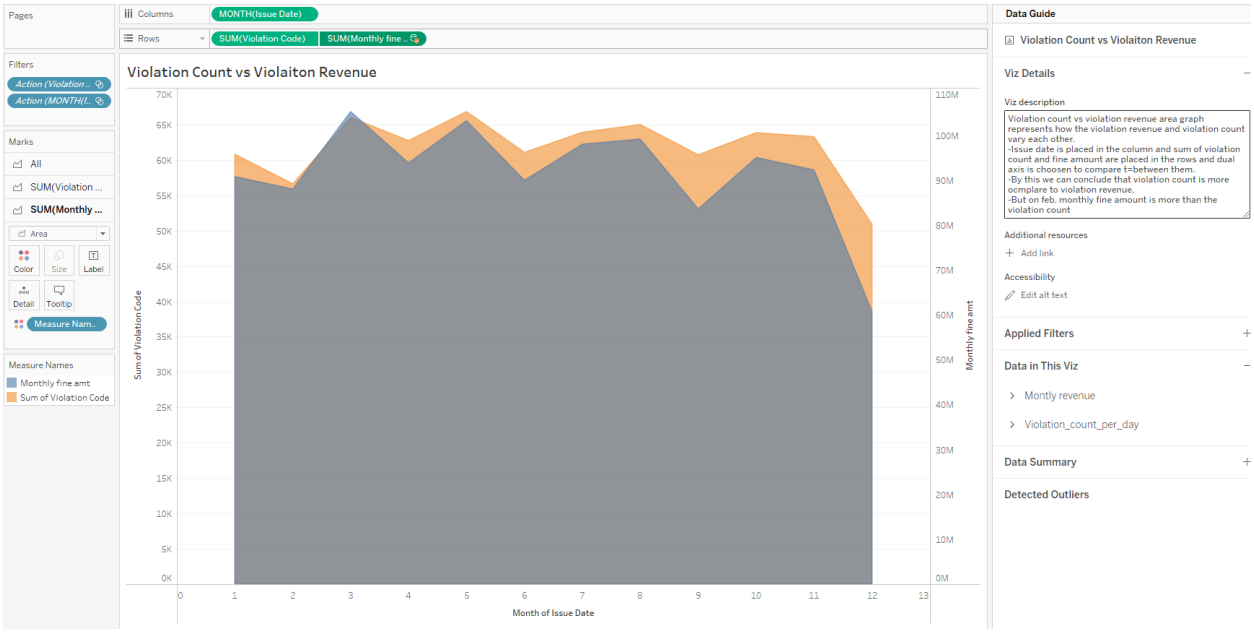
Highlight table represents the violation count for each issuing agency using issuing agency count and issuing agency details tables. Issuing agency and agency names are placed in the rows. Sum of violation is placed under the label marks and quick calculation function is applied. The Intensity of the color is directly proportional to the ranking, rank one has the the dark color and the color intensity decreases as the ranking goes down. Traffic agency has the highest violation count and it ranks first and housing authority has lowest ranking with the lowest violation count. Traffic agency has the highest violation count and it ranks first and housing authority has lowest ranking with the lowest violation count.

Proposed Question

Which Issuing agency has the highest violation count?

Traffic agency has the highest violation count and it ranks first and housing authority has lowest ranking with the lowest violation count.

Violation Count vs Violation Revenue



The Violation count vs violation revenue area graph represents how the violation revenue and violation count vary from each other. Issue date is placed in the column and sum of violation count and fine amount are placed in the rows and dual axis is chosen to compare between them. By this we can conclude that violation count is more compared to violation revenue, but in February, monthly fine amount is more than the violation count.

Proposed Question

Describe the relationship between monthly violation revenue and violation count?

Violation count for every month is more than the violation fine amount, only in February, monthly fine amount is more than the violation count

Total fine amount per violation category

The packed bubble graph represents the total fine amount for each violation category. The violation category is placed in the color marks, sum of total fine amount is placed in the size, violation category and total fine amount is placed in the text marks. The highest fine is for the parking category which is $414.66M and lowest is for lane violation which is $70.1M. Traffic violation category has the second highest violation which is $333.78M and all other category has around $100M total fine amount. Hence by this we can conclude Parking and traffic signals are the most violated category and they have maximum fine amount.

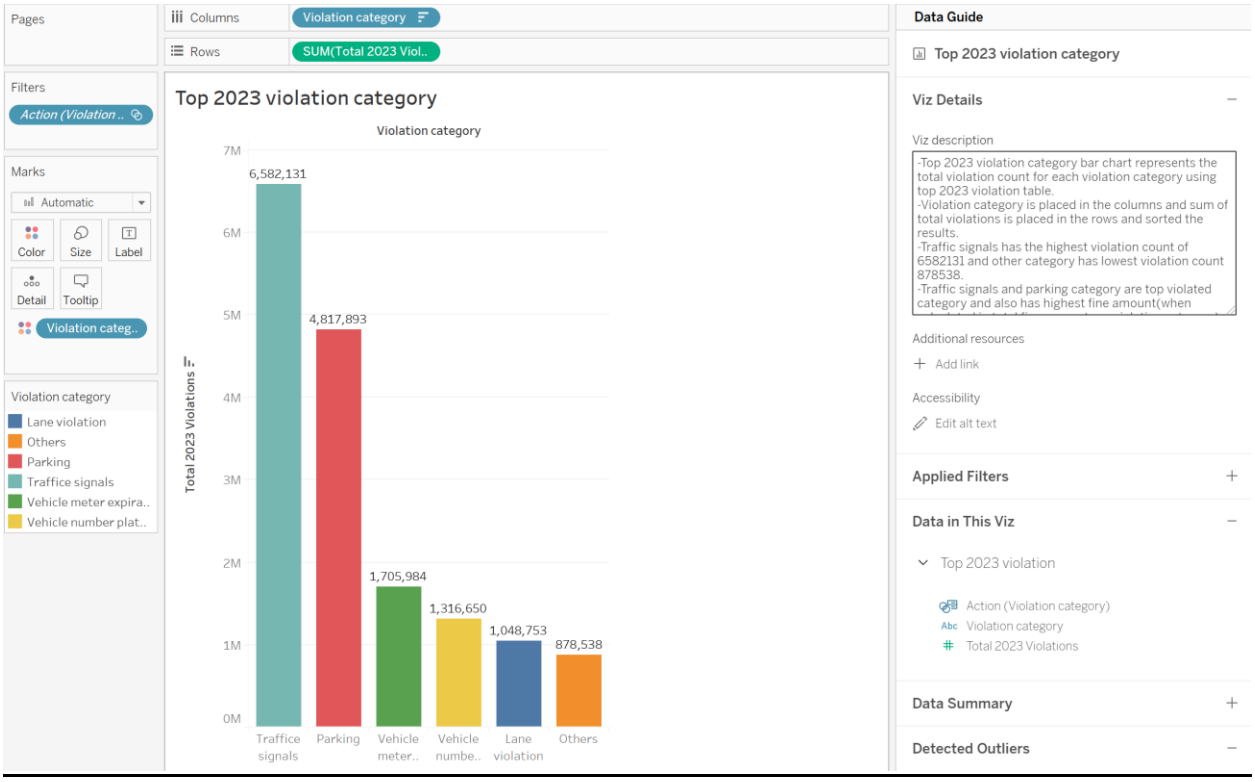
[ref:11]

Proposed Question

Which violation category has the highest violation revenue?

. The highest fine is for the parking category which is $414.66M and lowest is for lane violation which is $70.1M. Traffic violation category has the second highest violation which is $333.78M and all other category has around $100M total fine amount.

Top 2023 violation category



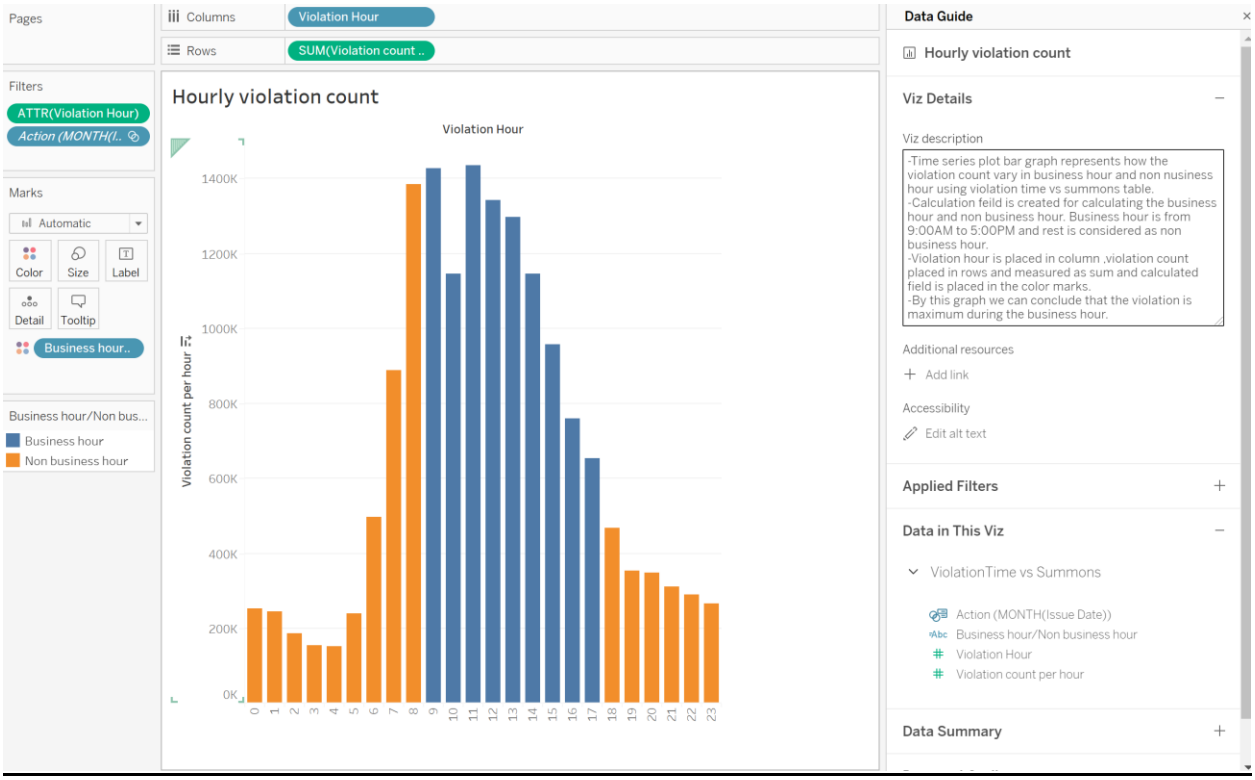
Top 2023 violation category bar chart represents the total violation count for each violation category using top 2023 violation table. The violation category is placed in the columns and sum of total violations is placed in the rows and sorted the results. Traffic signals have the highest violation count of 6582131 and the other category has the lowest violation count 878538.Traffic signals and parking category are top violated category and has highest fine amount (when calculated in total fine amount per violation category)

Proposed Question

Which violation category has the highest violation count?

Traffic signals have the highest violation count of 6582131 and the other category has the lowest violation count 878538.

Hourly violation count



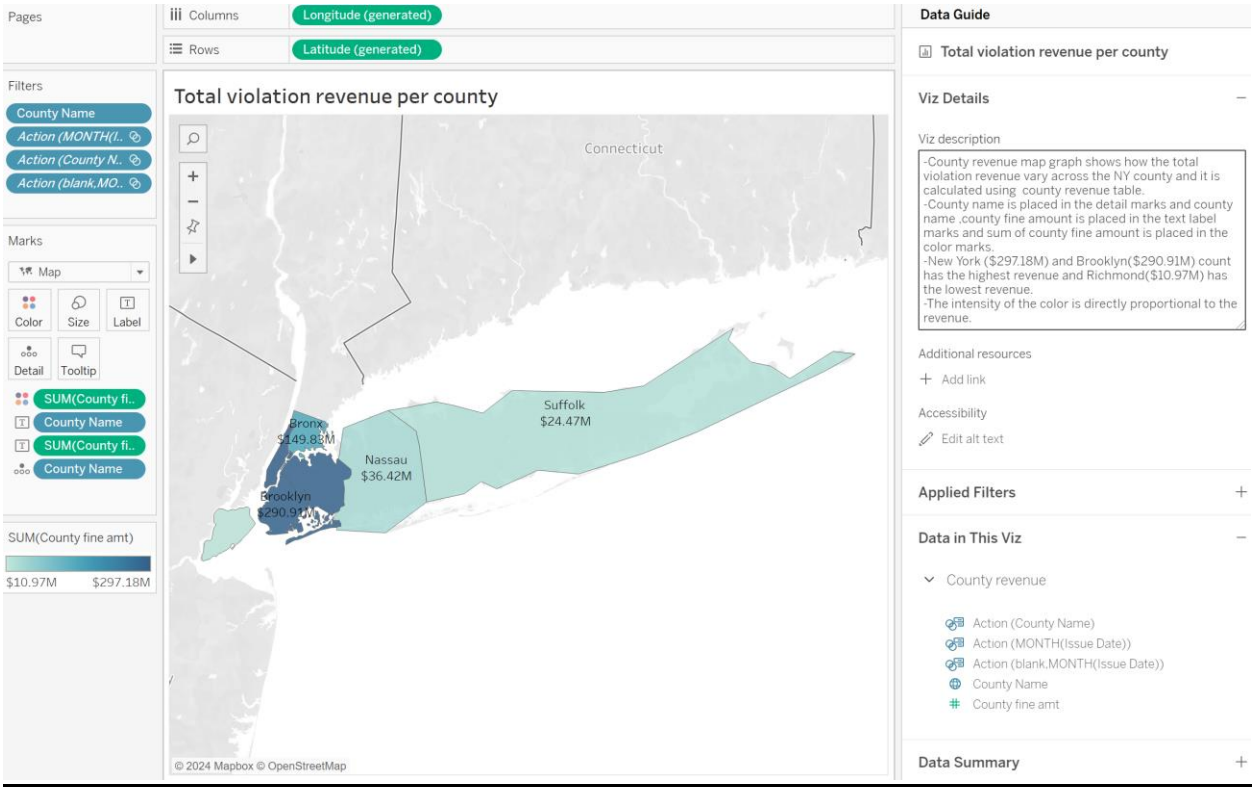
Time series plot bar graph represents how the violation count varies in business hour and non-business hour using violation time vs summons table. The calculation field is created for calculating the business hours and non-business hour. Business hours are from 9:00AM to 5:00PM and the rest is considered as non-business hours. Violation hour is placed in column, violation count placed in rows and measured as sum and calculated field is placed in the color marks. By this graph we can conclude that the violation is maximum during the business hour.

Proposed Question

During which hour the violation count is higher?

The violation is maximum during the business hour i.e. around 8-9:00AM and 11:00AM

Total violation revenue per county



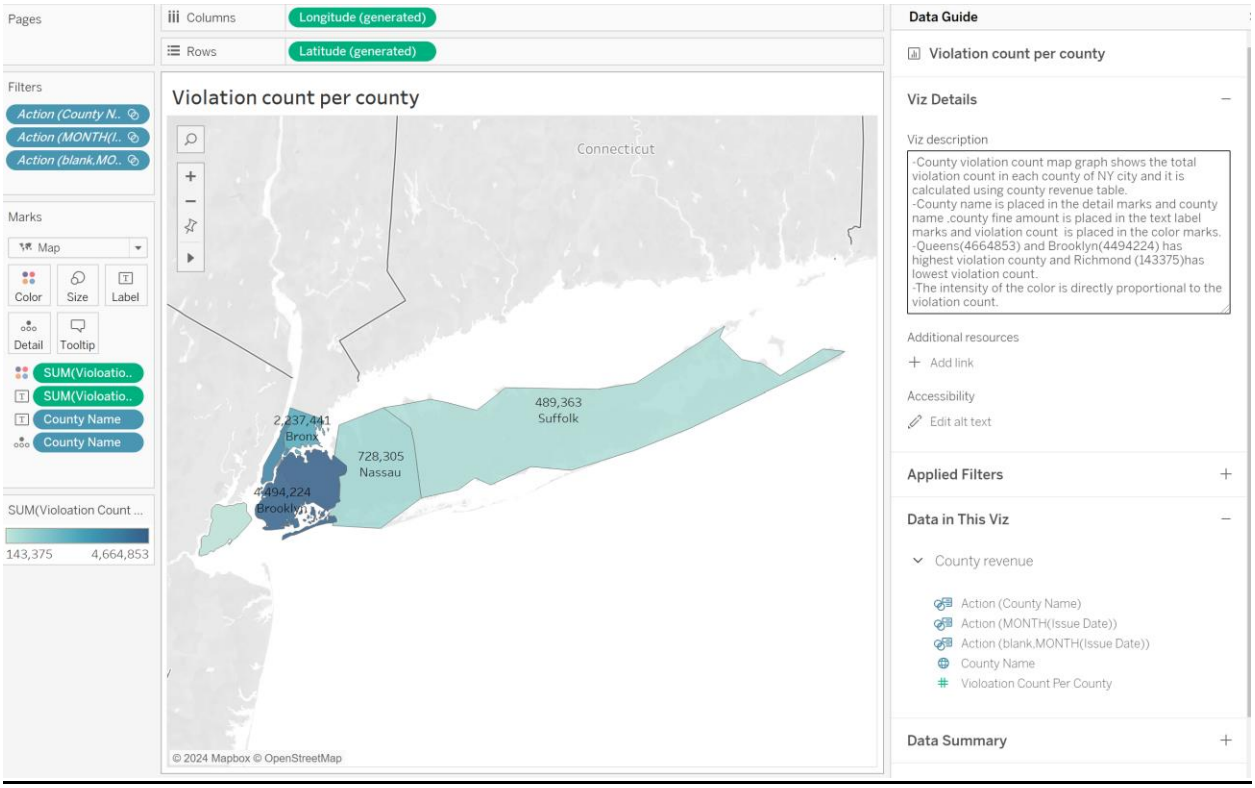
County revenue map graph shows how the total violation revenue varies across the NY county and it is calculated using county revenue table. County name is placed in the detail marks and county name, county fine amount is placed in the text label marks and sum of county fine amount is placed in the color marks. New York ($297.18M) and Brooklyn($290.91M) count has the highest revenue and Richmond($10.97M) has the lowest revenue. The intensity of the color is directly proportional to the revenue.

Proposed Question

Which county has the highest revenue?

New York ($297.18M) and Brooklyn($290.91M) count has the highest revenue and Richmond($10.97M) has the lowest revenue.

Violation counts per county



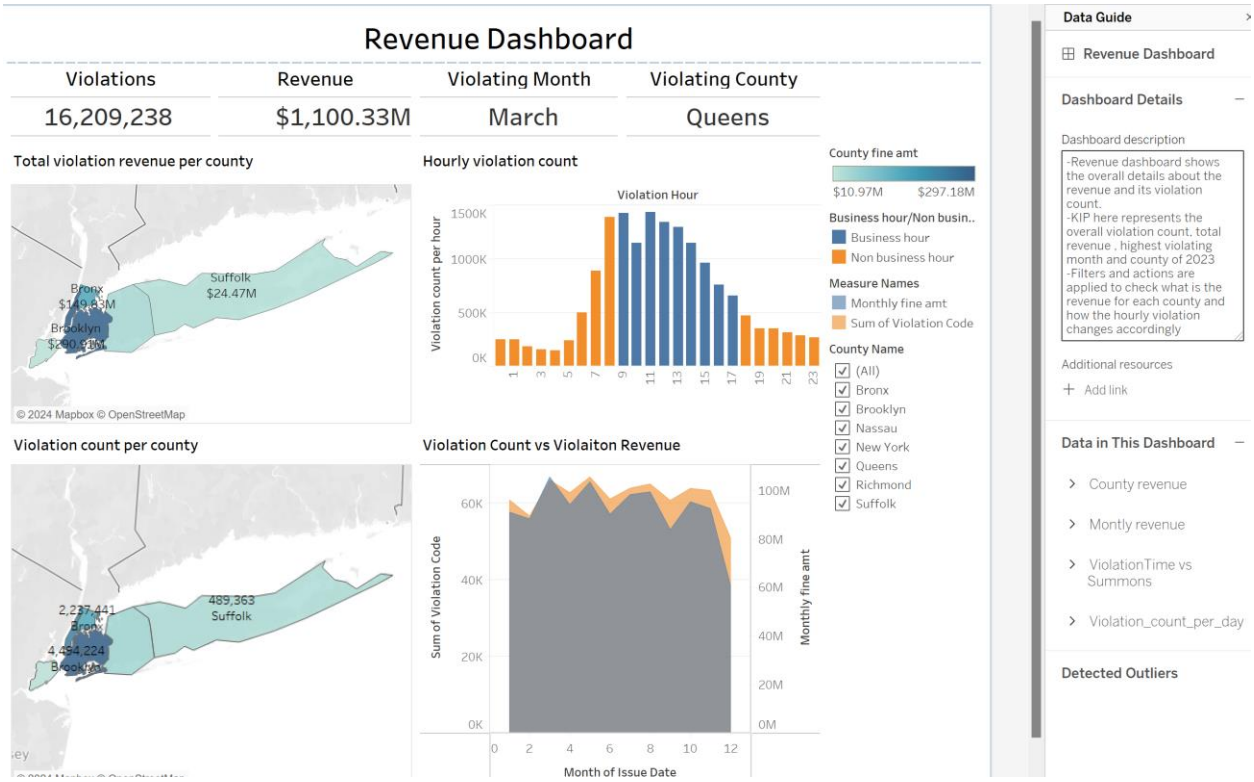
The county violation count map graph shows the total violation count in each county of NY city and it is calculated using county revenue table. County name is placed in the detail marks and county name, county fine amount is placed in the text label marks and violation count is placed in the color marks. Queens (4664853) and Brooklyn (4494224) have the highest violation county and Richmond (143375) has lowest violation count.The intensity of the color is directly proportional to the violation count.

Proposed Question

Which county has the highest violation count?

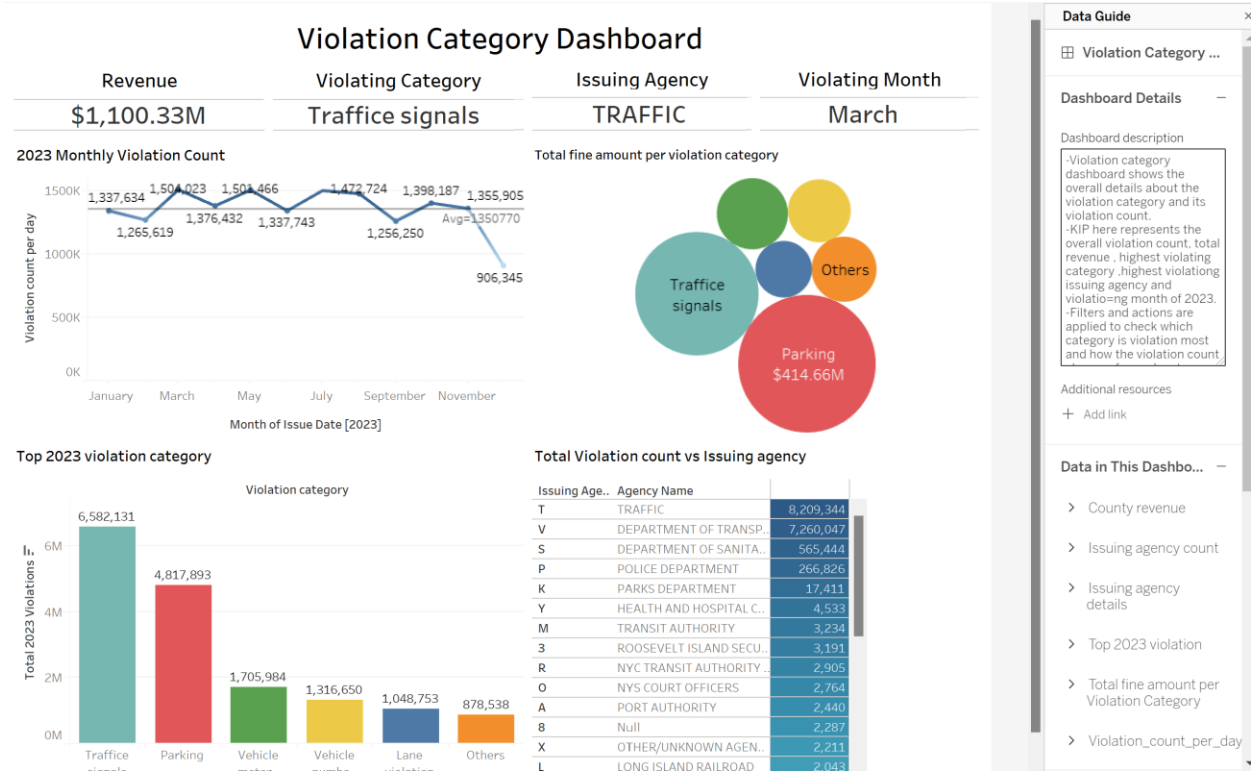
Queens (4664853) and Brooklyn (4494224) have the highest violation county and Richmond (143375) has lowest violation count.

Revenue Dashboard



The revenue dashboard shows the overall details about the revenue and its violation count. KIP here represents the overall violation count, total revenue, highest violating month and county of 2023.Filters and actions are applied to check what is the revenue for each county and how the hourly violation changes accordingly.

Violation Category Dashboard



The violation category dashboard shows the overall details about the violation category and its violation count. KIP here represents the overall violation count, total revenue, highest violating category, highest violating issuing agency and violation month of 2023. Filters and actions are applied to check which category is violation most and how the violation count changes for each category.

Summary and Conclusions

The main goal was to analyze the New York parking violations, evaluate efficiency of the enforcement, consider the regulatory effect, contemplate the demographic gaps and suggest the management tactics. The findings have confirmed the presence of a seasonal effect, hotspot areas, violations happening more often, and peak accident times. The idea of this project for illustrating such parking behavior and law enforcement objectives is effectively in line with the thought of congestion- fewer roads and fair parking management. Through evidence-based decision making the planning and policy making of urban transport can be improved. Benefits such as reduced traffic,

greater efficiency and dramatic improvements in urban life are all possible. Enforcement equity is impacted by regulations and demographic unevenness, as well as other factors. Therefore, in summary, the project contributes to the development of sustainable parking management protocols that would be replicated both locally and on the globe. [ref:9]

References.

1) Parking Violations Issued - Fiscal Year 2024 | NYC Open Data (cityofnewyork.us)
2) https://www.kaggle.com/datasets/new-york-city/nyc-parking-tickets
3) Chart Design Principles | Hands-On Data Visualization (handsondataviz.org)
4) A Complete Guide to Line Charts | Atlassian
5) Basic_Charts.pptx: DSCI 5360 Section 007 - Data Visualization (Spring 2024 1) (instructure.com)
6) Basic Visualizations_S.pptx: DSCI 5360 Section 007 - Data Visualization (Spring 2024 1) (instructure.com)
7) Gemini - chat to supercharge your ideas (google.com)
8) Basic Calculations_S.pptx: DSCI 5360 Section 007 - Data Visualization (Spring 2024 1) (instructure.com)
9) ChatGPT (openai.com)
10) "Basic Visualizations_S.pptx: DSCI 5360 Section 007 - Data Visualization (Spring 2024 1)." Accessed: Feb. 24, 2024. [Online]. Available: https://unt.instructure.com/courses/104031/files/25950574?module_item_id=6421667
11) "Build a Packed Bubble Chart." Accessed: Feb. 24, 2024. [Online]. Available: https://help.tableau.com/current/pro/desktop/en-us/buildexamples_bubbles.htm
12) "Data & Analytics," Bespin Global. Accessed: Feb. 24, 2024. [Online]. Available: https://bespinglobal.us/services/data-analytics
13) "Data Preprocessing, Analysis, and Visualization for building a Machine learning model," InsideAIML. Accessed: Feb. 24, 2024. [Online]. Available: https://insideaiml.com/blog/Data-Preprocessing,-Analysis,-and-Visualization-for-building-a-Machine-learning-model-1166
14) "What Is Data Analysis: A Comprehensive Guide," Simplilearn.com. Accessed: Feb. 24, 2024. [Online]. Available: https://www.simplilearn.com/data-analysis-methods-process-types-article

Contributions:

Member Name	Contributions
Pooja Kondagoli	Conducting research to find a suitable dataset for analysis. Leading data preprocessing efforts to clean and transform the dataset into usable measures and dimensions using Python. Participating in brainstorming sessions with team members to develop effective visualization strategies. Designing Tableau visualizations and dashboards based on the agreed-upon strategies. Drafting the project plan and implementation proposal to outline objectives and tasks. Organizing team meetings and facilitating coordination among team members to ensure smooth progress. Reviewing and contributing to documentation to maintain clarity and consistency in project deliverables. Overall, played a crucial role in various stages of the project, from data acquisition and preprocessing to visualization design and project management.
Jhansi Rani Minda	Responsible for creating and maintaining comprehensive documentation for the project, ensuring that all aspects of the project are well-documented and easily accessible. Also committed to continuously updating and improving the documentation to ensure its accuracy and relevance.
Venkata Sai Jyothi Ravilla	Contributed to developing the tableau visualization and narration of viz insights and drafting references. Actively involved in brainstorming session.
Chimata Nagabala Bhavani	Responsible for creating and maintaining comprehensive documentation for the project, ensuring that all aspects of the project are well-documented and easily accessible. Also committed to continuously updating and

	improving the documentation to ensure its accuracy and relevance.
Sai Kiran Thirigabathi Ramesh	Participated in the creation of thorough documentation and had smooth communication and workflow collaboration. Involved in the selection of relevant data for project and offered helpful input during the brainstorming stage. Furthermore, actively participated in the creation of thorough documentation and made sure that my teammates and I had smooth communication and workflow collaboration.
Praneeth Vamshi Tarlapalli	Categorized the violation codes and groups. Actively participated in team discussions.