IE6600 Computation and Visualization

Spring Semester 2025

Project 3

Topic: Moving Violation – Feb 2025 (Washington, DC)

Group 3

Palak Tanwar

Ganapriya Hiriyur Shivakumar

Sushritha Bharadwaj Deshakulkarni Srikantha

Abstract

This project explores geospatial patterns and insights derived from the dataset “Moving Violations Issued in February 2025”, sourced from data.gov. The primary goal is to analyze the spatial distribution of traffic violations within Washington, D.C., using interactive visualizations created with the Plotly library. The dataset contains detailed records of violations including location data, violation types, enforcement methods, and issuing agencies, offering a rich foundation for geographical and behavioral analysis.

The project workflow includes meticulous data inspection, cleaning, and transformation to ensure accurate spatial representation. We resolved issues such as missing geolocation values and inconsistent formatting. Leveraging Plotly’s capabilities, we developed a series of interactive maps and visual dashboards that reveal hotspots of violations, time-based trends, and correlations with enforcement types and vehicle information.

Key findings highlight clusters of violations in densely populated or high-traffic zones, time-of-day patterns, and areas with frequent enforcement activity. These insights offer practical implications for urban planning, traffic enforcement strategies, and public safety awareness.

All interactive outputs are saved as standalone HTML files, facilitating accessible and immersive exploration of the geospatial patterns. The final deliverables include a comprehensive Jupyter Notebook documenting the entire analytical process, a report detailing the methodology and key insights, and an optional website for public dissemination of results.

Table of Contents

1. Introduction

This project focuses on the geospatial analysis and interactive visualization of the dataset titled "Moving Violations Issued in February 2025," retrieved from data.gov. The dataset captures detailed records of moving violations issued across Washington, D.C., during a single month, with attributes including violation types, vehicle details, issuing agencies, and most importantly, geographic coordinates (latitude and longitude).

The objective is to clean, process, and visualize this data to identify spatial clusters of traffic violations, highlight high-risk zones, and explore potential correlations between location and violation type. To achieve this, we utilize the Plotly library in Python to create interactive maps and dashboards, which enhance user engagement and allow dynamic exploration of the data.

2. Data Acquisition and Inspection

The dataset used in this project, titled “Moving Violations Issued in February 2025,” was sourced from data.gov, a public repository for government datasets across the United States. The dataset is provided in CSV format and contains detailed information about traffic violations issued within Washington, D.C., including spatial data fields such as latitude and longitude.

Upon downloading the dataset, it was imported into a Python-based analysis environment using libraries such as pandas, numpy, and plotly for manipulation and visualization. The initial inspection focused on understanding the structure, types of attributes, and overall data quality. The dataset comprises several key columns, including:

- ticket_number: A unique identifier for each violation.
- issue_date: The date the violation was issued.
- violation_code and violation_description: Identifiers and descriptions of the type of violation.
- fine_amount: The monetary penalty associated with the violation.

- location, latitude, longitude: Spatial data used for geolocation.
- issuing_agency: The law enforcement agency that issued the ticket.

2.1 Dataset Overview

Below is an overview of the dataset:

```
Dataset Shape: (121228, 29)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 121228 entries, 0 to 121227
Data columns (total 29 columns):
 #   Column                  Non-Null Count   Dtype  
---  ------                  --------------   -----  
 0   OBJECTID                121228 non-null  int64  
 1   LOCATION                121228 non-null  object 
 2   XCOORD                  119207 non-null  float64
 3   YCOORD                  119207 non-null  float64
 4   ISSUE_DATE              121228 non-null  object 
 5   ISSUE_TIME              121228 non-null  int64  
 6   ISSUING_AGENCY_CODE     121228 non-null  int64  
 7   ISSUING_AGENCY_NAME     121228 non-null  object 
 8   ISSUING_AGENCY_SHORT    121228 non-null  object 
 9   VIOLATION_CODE          121228 non-null  object 
 10  VIOLATION_PROCESS_DESC  121228 non-null  object 
 11  PLATE_STATE             121228 non-null  object 
 12  ACCIDENT_INDICATOR      2381 non-null    object 
 13  DISPOSITION_CODE        121228 non-null  int64  
 14  DISPOSITION_TYPE        121228 non-null  object 
 15  DISPOSITION_DATE        874 non-null     object 
 16  FINE_AMOUNT             121228 non-null  int64  
 17  TOTAL_PAID              121228 non-null  int64  
 18  PENALTY_1               121228 non-null  int64  
 19  PENALTY_2               121228 non-null  int64  
 20  PENALTY_3               121228 non-null  int64  
 21  PENALTY_4               121228 non-null  int64  
 22  PENALTY_5               121228 non-null  int64  
 23  RP_MULT_OWNER_NO        121228 non-null  int64  
 24  BODY_STYLE              0 non-null       float64
 25  LATITUDE                119207 non-null  float64
 26  LONGITUDE               119207 non-null  float64
 27  MAR_ID                  111662 non-null  float64
 28  GIS_LAST_MOD_DTTM       121228 non-null  object 
dtypes: float64(6), int64(12), object(11)
memory usage: 26.8+ MB
```

	OBJECTID	LOCATION	XCOORD	YCOORD	ISSUE_DATE	ISSUE_TIME	ISSUING_AGENCY_CODE	ISSUING_AGENCY_NAME	ISSUING_AGENCY_SHORT	VIOLATIO
0	40968849	550 BLOCK GALLOWAY ST NW	NaN	NaN	2025/02/06 05:00:00+00	751	57	METRO POLICE	MTP	
1	40968850	550 BLOCK GALLOWAY ST NW	NaN	NaN	2025/02/06 05:00:00+00	808	57	METRO POLICE	MTP	
2	40968851	201 BLOCK GALLOWAY ST NW	NaN	NaN	2025/02/07 05:00:00+00	737	57	METRO POLICE	MTP	
3	41288420	201 BLK GALLOWAY ST NW	NaN	NaN	2025/02/07 05:00:00+00	748	57	METRO POLICE	MTP	
4	41288421	201 BLOCK GALLOWY ST NW	NaN	NaN	2025/02/07 05:00:00+00	304	57	METRO POLICE	MTP	

5 rows × 29 columns

2.2 Data Cleaning

As part of our initial data inspection, we examined the extent of missing values in the dataset to determine which columns might require cleaning or imputation.

```
naPercent = df.isna().sum()/ len(df)*100
naPercent
```

```
OBJECTID                   0.000000
LOCATION                   0.000000
XCOORD                     1.667107
YCOORD                     1.667107
ISSUE_DATE                 0.000000
ISSUE_TIME                 0.000000
ISSUING_AGENCY_CODE        0.000000
ISSUING_AGENCY_NAME        0.000000
ISSUING_AGENCY_SHORT       0.000000
VIOLATION_CODE             0.000000
VIOLATION_PROCESS_DESC     0.000000
PLATE_STATE                0.000000
ACCIDENT_INDICATOR        98.035932
DISPOSITION_CODE           0.000000
DISPOSITION_TYPE           0.000000
DISPOSITION_DATE          99.279044
FINE_AMOUNT                0.000000
TOTAL_PAID                 0.000000
PENALTY_1                  0.000000
PENALTY_2                  0.000000
PENALTY_3                  0.000000
PENALTY_4                  0.000000
PENALTY_5                  0.000000
RP_MULT_OWNER_NO           0.000000
BODY_STYLE               100.000000
LATITUDE                   1.667107
LONGITUDE                  1.667107
MAR_ID                     7.890916
GIS_LAST_MOD_DTTM          0.000000
dtype: float64
```

Cleaned data:

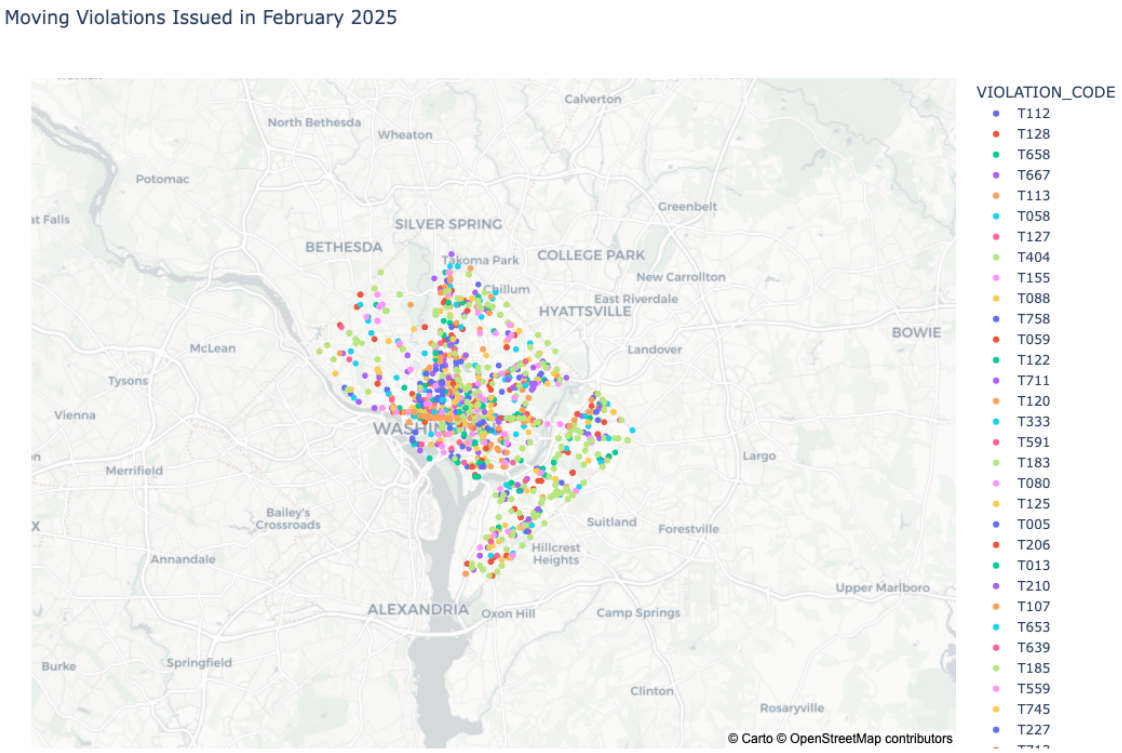
```
cleandf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 121228 entries, 0 to 121227
Data columns (total 25 columns):
 #   Column                  Non-Null Count   Dtype  
---  ------                  --------------   -----  
 0   OBJECTID                121228 non-null  int64  
 1   LOCATION                121228 non-null  object 
 2   XCOORD                  119207 non-null  float64
 3   YCOORD                  119207 non-null  float64
 4   ISSUE_DATE              121228 non-null  object 
 5   ISSUE_TIME              121228 non-null  int64  
 6   ISSUING_AGENCY_CODE     121228 non-null  int64  
 7   ISSUING_AGENCY_NAME     121228 non-null  object 
 8   ISSUING_AGENCY_SHORT    121228 non-null  object 
 9   VIOLATION_CODE          121228 non-null  object 
 10  VIOLATION_PROCESS_DESC  121228 non-null  object 
 11  PLATE_STATE             121228 non-null  object 
 12  DISPOSITION_CODE        121228 non-null  int64  
 13  DISPOSITION_TYPE        121228 non-null  object 
 14  FINE_AMOUNT             121228 non-null  int64  
 15  TOTAL_PAID              121228 non-null  int64  
 16  PENALTY_1               121228 non-null  int64  
 17  PENALTY_2               121228 non-null  int64  
 18  PENALTY_3               121228 non-null  int64  
 19  PENALTY_4               121228 non-null  int64  
 20  PENALTY_5               121228 non-null  int64  
 21  RP_MULT_OWNER_NO        121228 non-null  int64  
 22  LATITUDE                119207 non-null  float64
 23  LONGITUDE               119207 non-null  float64
 24  GIS_LAST_MOD_DTTM       121228 non-null  object 
dtypes: float64(4), int64(12), object(9)
memory usage: 23.1+ MB
```

3. Exploratory Data Analysis using Plotly

As part of our Exploratory Data Analysis (EDA), we used Plotly, an interactive graphing library, to create dynamic and insightful visualizations. Plotly allows for the creation of high-quality charts and maps that make it easier to explore patterns, trends, and outliers in the data. Its interactive features, such as zooming and hovering, provided deeper insights and enhanced our ability to interpret complex datasets effectively.

Mapbox

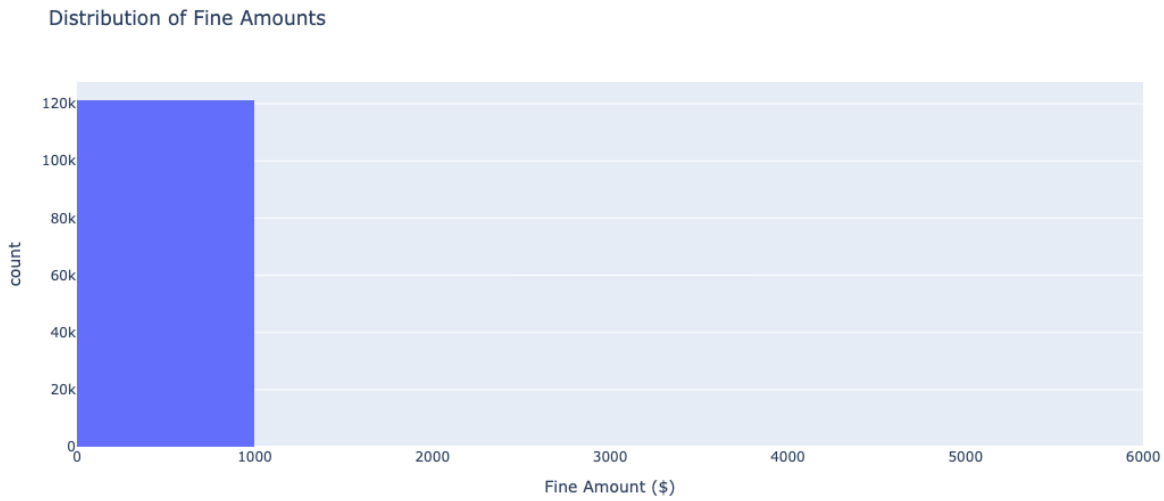


The map reveals clear clusters of moving violations, highlighting high-traffic zones or areas where traffic rules are frequently broken. Different colors represent various types of violations, helping identify which offenses are more common in specific regions. The concentration of certain violations also points to areas of active enforcement by specific agencies. Additionally, regions

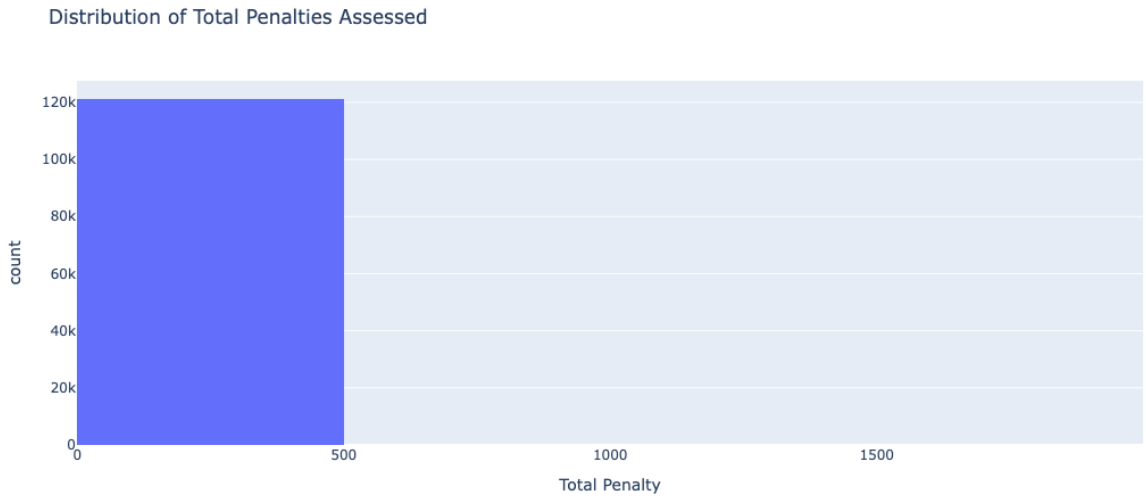
with repeated violations may indicate a need for infrastructure or traffic management improvements.

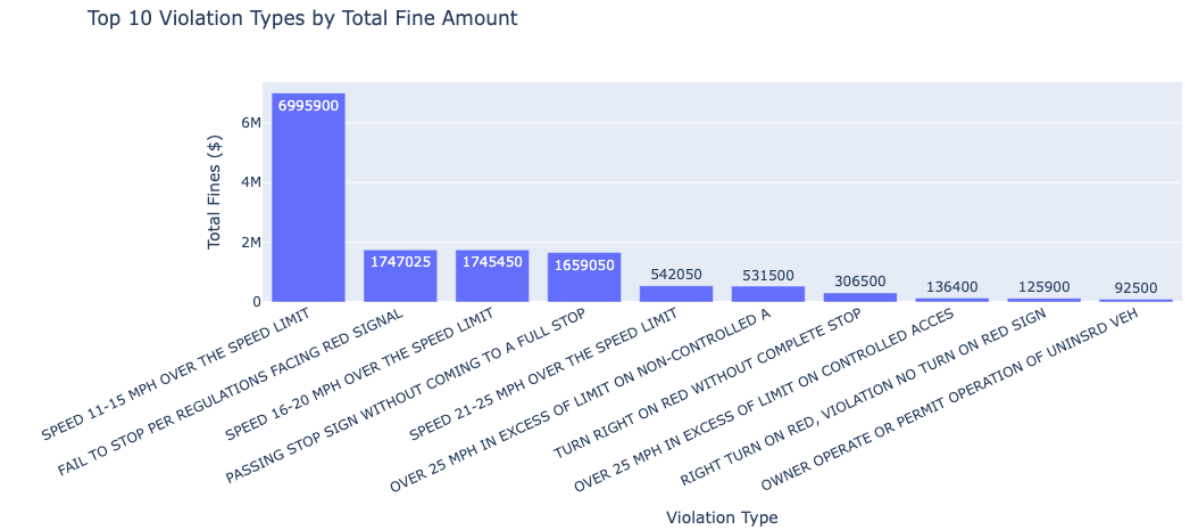
Bar Graph

This bar chart highlights the top 10 most common moving violations issued in February 2025. The most frequent offense by far was speeding 11–15 MPH over the limit, with nearly 70,000 tickets issued, significantly higher than any other violation. Other common violations included running stop signs, failing to stop at red signals, and turning violations. This data provides insight into the most frequent types of unsafe driving behavior and can help prioritize traffic enforcement and safety interventions.



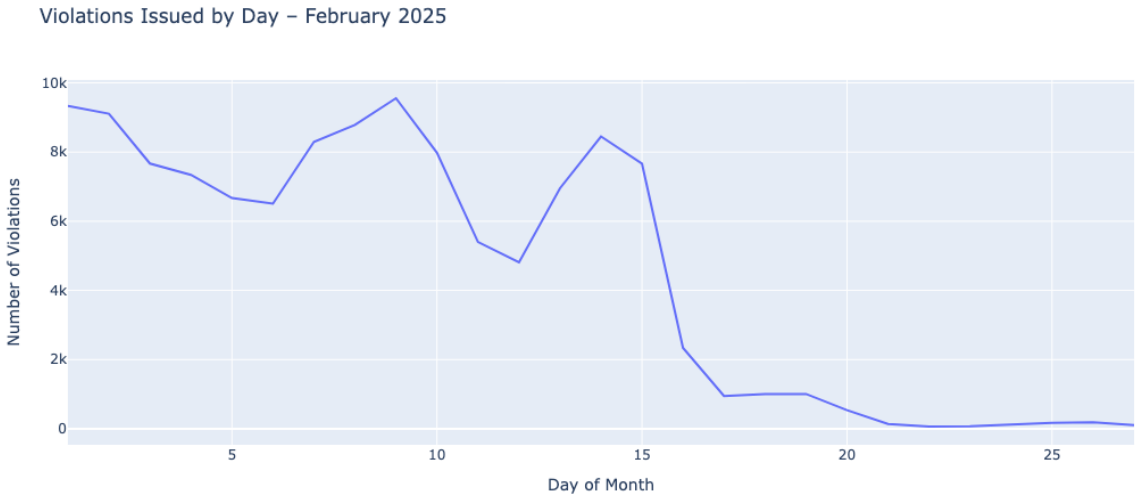
This histogram shows the distribution of fine amounts for moving violations. Most fines are under $1,000, indicating that most violations incur relatively low penalties. Higher fine amounts are extremely rare, suggesting that severe infractions are uncommon or less frequently enforced. This trend reflects a system focused on issuing fines for more common, moderate offenses.





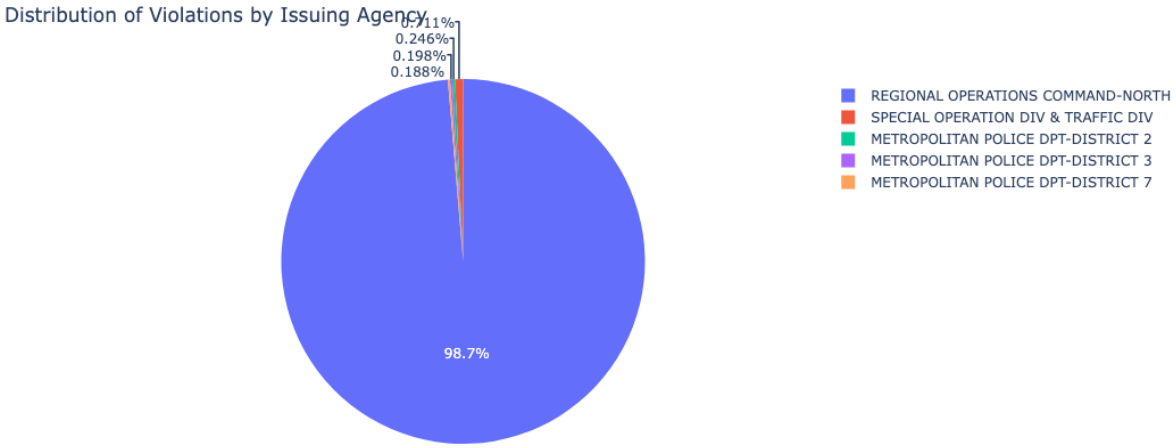
Speeding violations, particularly in the 11–15 mph and 16–20 mph ranges, account for most fines, with one category alone generating nearly $7 million. These fines serve as a major revenue source for traffic authorities, highlighting the need for stricter enforcement and public awareness. Targeted measures like speed cameras and education campaigns could help reduce violations and improve road safety over time.

Line Plot



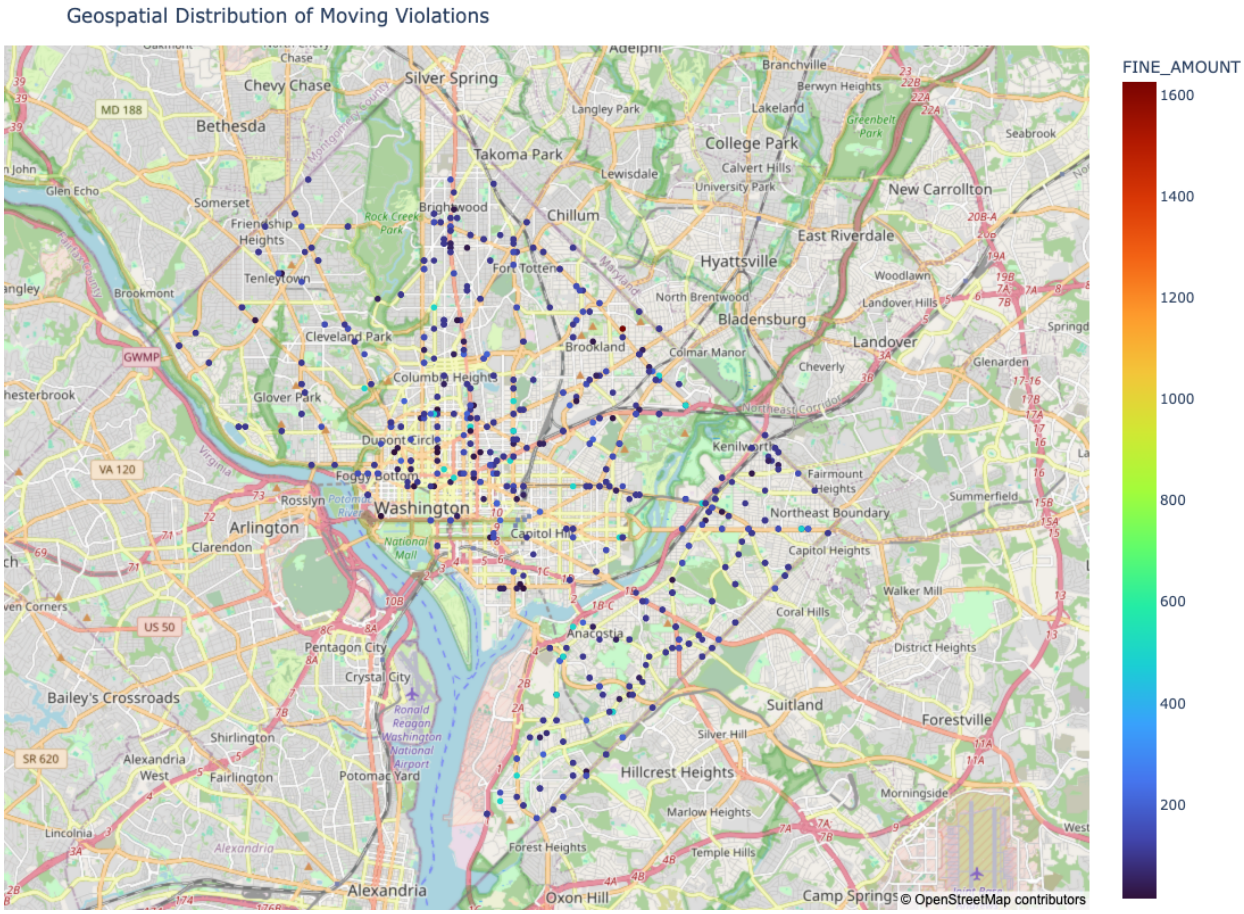
The number of violations peaked on February 1st with around 10,000 cases, followed by a steady decline until mid-month. From the 15th to the 25th, the trend stabilized with slight fluctuations. These patterns may be influenced by weekends, holidays, or shifts in enforcement efforts.

Pie Chart



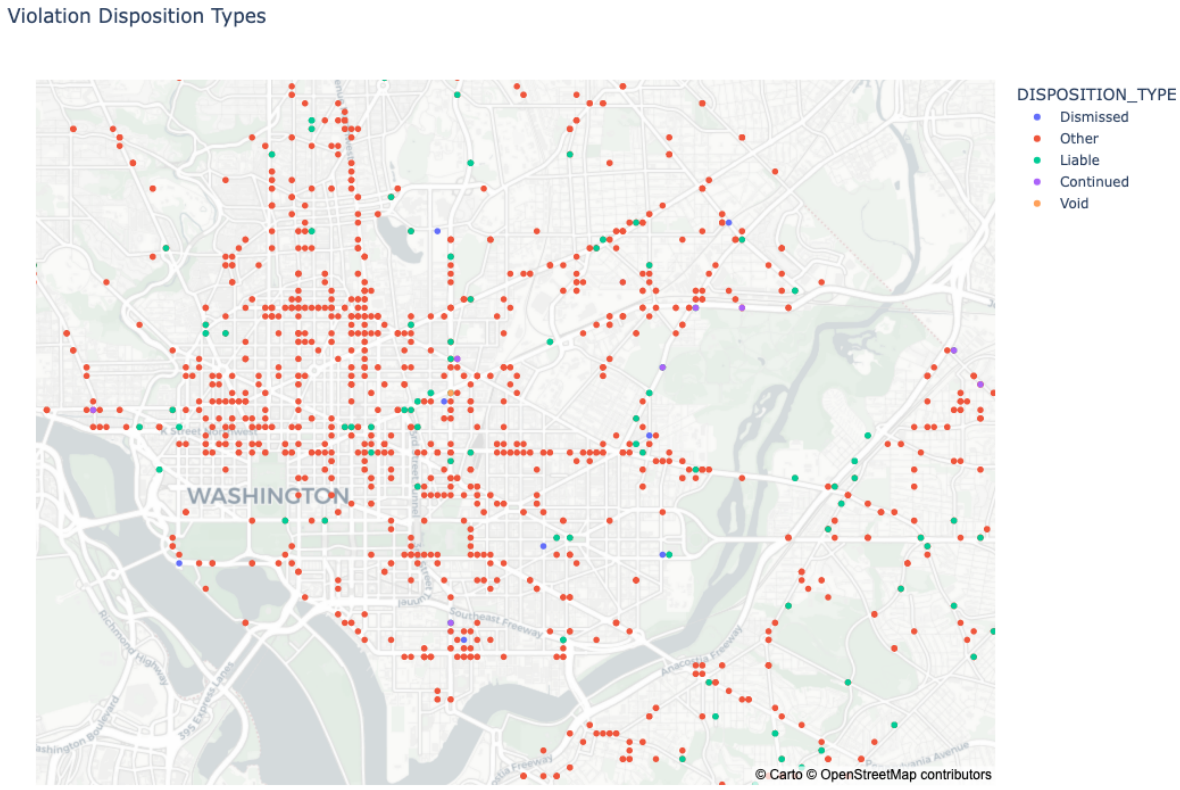
This pie chart shows the distribution of moving violations by issuing agency. The Regional Operations Command–North is responsible for most violations, issuing 98.7% of all tickets. Other agencies, such as the Special Operation Division & Traffic Division and various Metropolitan Police Districts, account for less than 1% each. This indicates that enforcement activity is highly centralized under a single agency.

Geospatial Plot



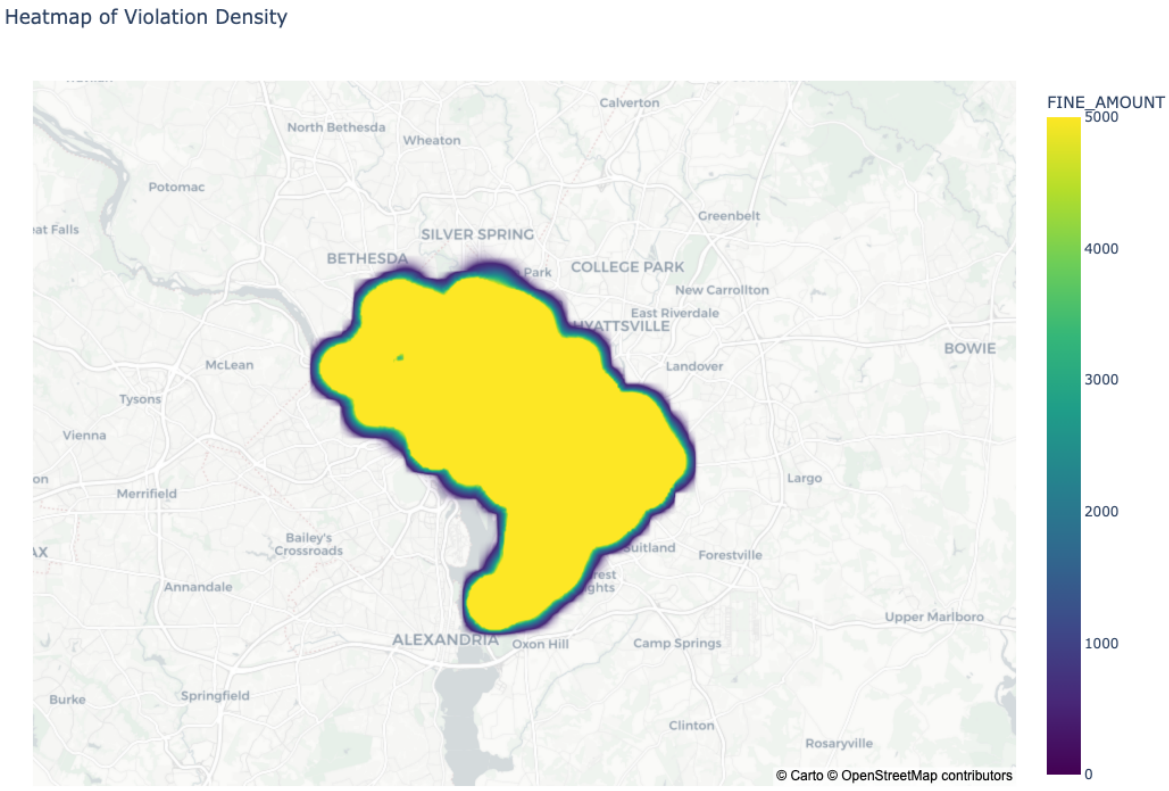
This geospatial map illustrates where moving violations occurred across Washington, D.C., with colors representing the corresponding fine amounts. Areas with higher fines, shown in warmer colors like orange and red, are concentrated in specific corridors, suggesting zones of severe or frequent violations. This pattern highlights priority regions for targeted enforcement and potential infrastructure improvements.

Scatter Mapbox



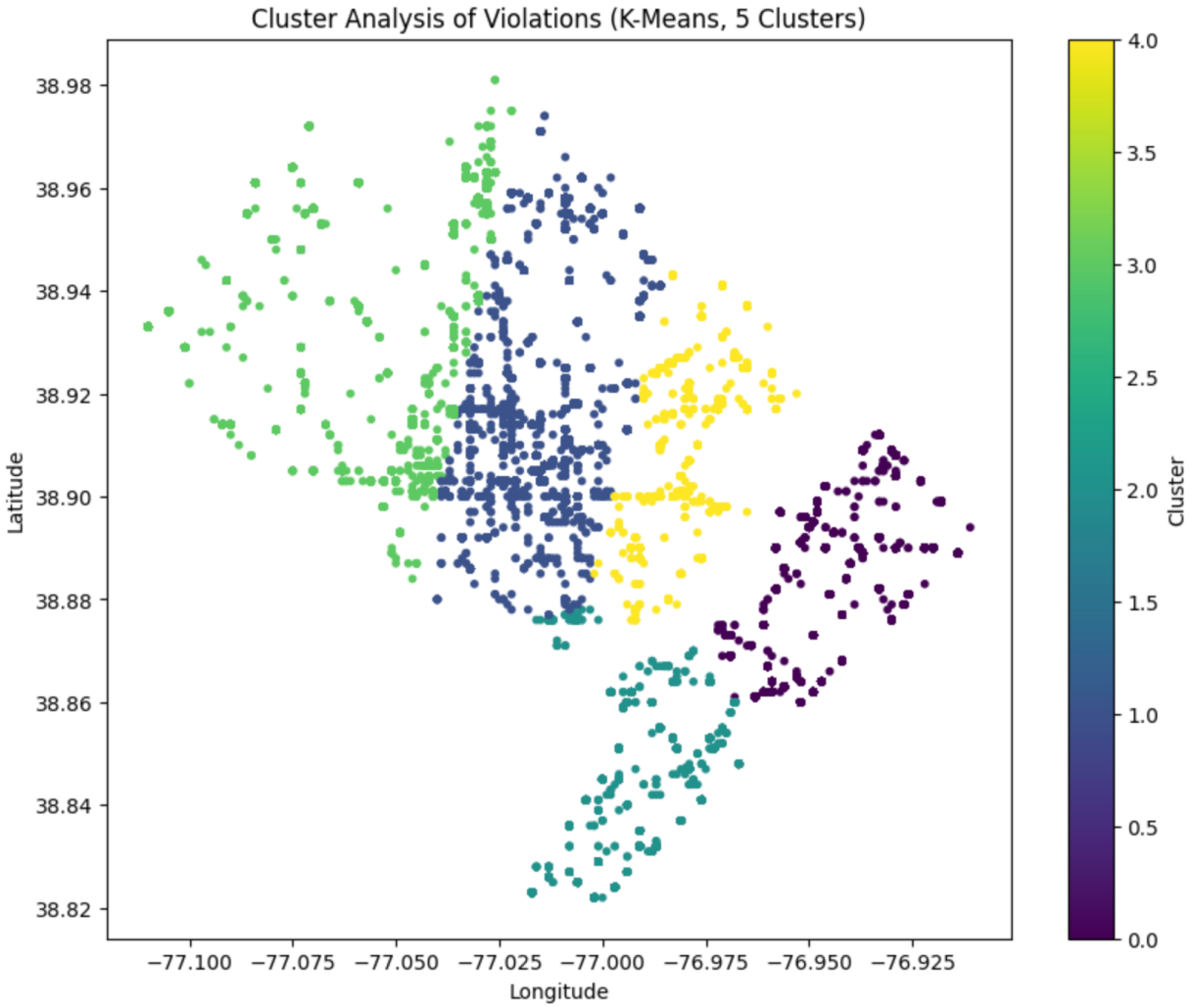
This map displays the spatial distribution of violation disposition types across Washington, D.C. Most violations are marked as "Liable" (in red), indicating they resulted in a fine or penalty. A smaller number are "Dismissed," "Continued," or "Void," shown in varying colors, highlighting areas where citations were contested or not enforced. This helps identify both enforcement hotspots and potential areas of legal leniency.

Density Mapbox



The heatmap highlights high violation hotspots in bright yellow, showing areas with dense traffic violations that may need increased enforcement. In contrast, darker purple zones indicate fewer violations, possibly due to lower traffic or underreporting. These patterns help prioritize targeted interventions like patrols or infrastructure improvements and assess the effectiveness of current enforcement strategies.

Scatter Plot



The cluster analysis shows that traffic violations are concentrated in specific areas rather than evenly spread across the region. These hotspots, like the green, blue, and purple zones, suggest consistent patterns of violations that may point to local traffic issues or enforcement gaps. Identifying these clusters helps guide resource allocation, such as patrols or cameras, and supports targeted policy or infrastructure changes to enhance safety and reduce violations.

4. Conclusion

This project provided valuable insights into moving violations issued across Washington, D.C., in February 2025. Through exploratory data analysis using Plotly, we identified key trends such as the high prevalence of speeding violations, the concentration of fines in specific regions, and the dominant role of certain enforcement agencies. Geospatial and temporal visualizations revealed clear hotspots and patterns, supporting the need for targeted interventions, policy adjustments, and increased public awareness. Overall, the findings can help inform more efficient traffic enforcement strategies and contribute to improving road safety across the city.