Parking Violation Data Analysis for Policy Development

Introduction

This project focuses on analysing parking violation data from September 2024 to identify trends and key insights. The goal is to use these findings to shape regulations that help reduce daily violations and improve compliance. The results aim to support better parking management.

Source of the data set:

Data was sourced from Data.gov (https://catalog.data.gov/dataset/parking-violations-issued-in-september-2024)

The dataset comprises 31 fields, detailing ticket information (e.g., ticket number, violation type, issue time and date), geolocation (latitude, longitude), the issuing agency, and penalties. It includes both categorical data (e.g., violation type, issuing agency) and numerical data (e.g., penalties, coordinates). While the data is generally consistent with, some columns contain missing or incomplete values (e.g., latitude, longitude) and mixed data types that may require standardization (e.g., penalty columns).

To streamline the analysis and focus on the most impactful variables relevant to the project objective, I removed columns that did not directly support the core analysis.

Removed columns are as follows:

```
# indentified a list of columns that are not required or might not contribute for the analysis
Columns_to_be_removed = ['VEHICLE_TYPE','MULTI_OWNER_NUMBER','DISPOSITION_CODE','DISPOSITION_TYPE','DISPOSITION_DESC','DISPOSITION_DATE',
                         'TOTAL_PAID','PENALTY_1','PENALTY_2','PENALTY_3','PENALTY_4','PENALTY_5','XCOORD','YCOORD','LATITUDE','LONGITUDE','MAR_ID','GIS_LAST_MOD_DTTM']

# removed the columns that are not required or might not contribute for the analysis
Clean_Parking_Violation= Clean_Parking_Violation_Set.drop(columns=Columns_to_be_removed,errors='ignore')
```

Preparing Time-Based Data for Analysis

To effectively analyse and visualize parking violations by location, I will split the date field into separate columns for year, month, and day. This breakdown will help identify time-based patterns and improve the clarity of visualizations. This action resulted in the creation of three new columns, which were subsequently used in the analysis and visualizations.

Created new columns:

```
# Extract year, month, and day into 3 separate columns
Clean_Parking_Violation['ISSUE_YEAR'] = Clean_Parking_Violation['ISSUE_DATE'].dt.year
Clean_Parking_Violation['ISSUE_MONTH'] = Clean_Parking_Violation['ISSUE_DATE'].dt.month
Clean_Parking_Violation['ISSUE_DAY'] = Clean_Parking_Violation['ISSUE_DATE'].dt.day
```

Identifying Top Locations for Parking Violations

After preparing the data for analysis and visualization, i then proceed with the initial analysis, which is fundamental to understanding where parking violations occur most frequently, the locations with the highest frequency of parking violations. This will be achieved through the following steps:

1. **Counting Violations by Location:**
   Using the violation ***"location"*** column, I calculated the number of violations at each unique location. This metric will provide a detailed breakdown of violation frequencies, helping identify areas with the highest infraction rates.
2. **Filtering to Top N Locations:**
   To ensure clarity in visualizations, I will focus on the top N locations with the highest violation counts. This filtering will highlight key hotspots while avoiding clutter, making it easier to identify significant trends.

The process of performing the above mentioned.

To analyse parking violation data and identify and rank the most frequent violation locations. We start with

Calculate Location Frequencies

- Use value_counts() on the ***"location"*** column to count the occurrences of each unique location.
- Convert the result into a DataFrame using reset_index().

Rename Columns

- Rename the columns in the newly created DataFrame to ***"location"*** (for the location names) and **"violation_count"** (for the frequency counts).

Filter Top N Locations (Optional)

- Use nlargest() to select the top N locations with the highest violation counts (e.g., top 10 locations). This step is optional based on your analysis requirements.

Group by Location and Aggregate Violations (Extension):

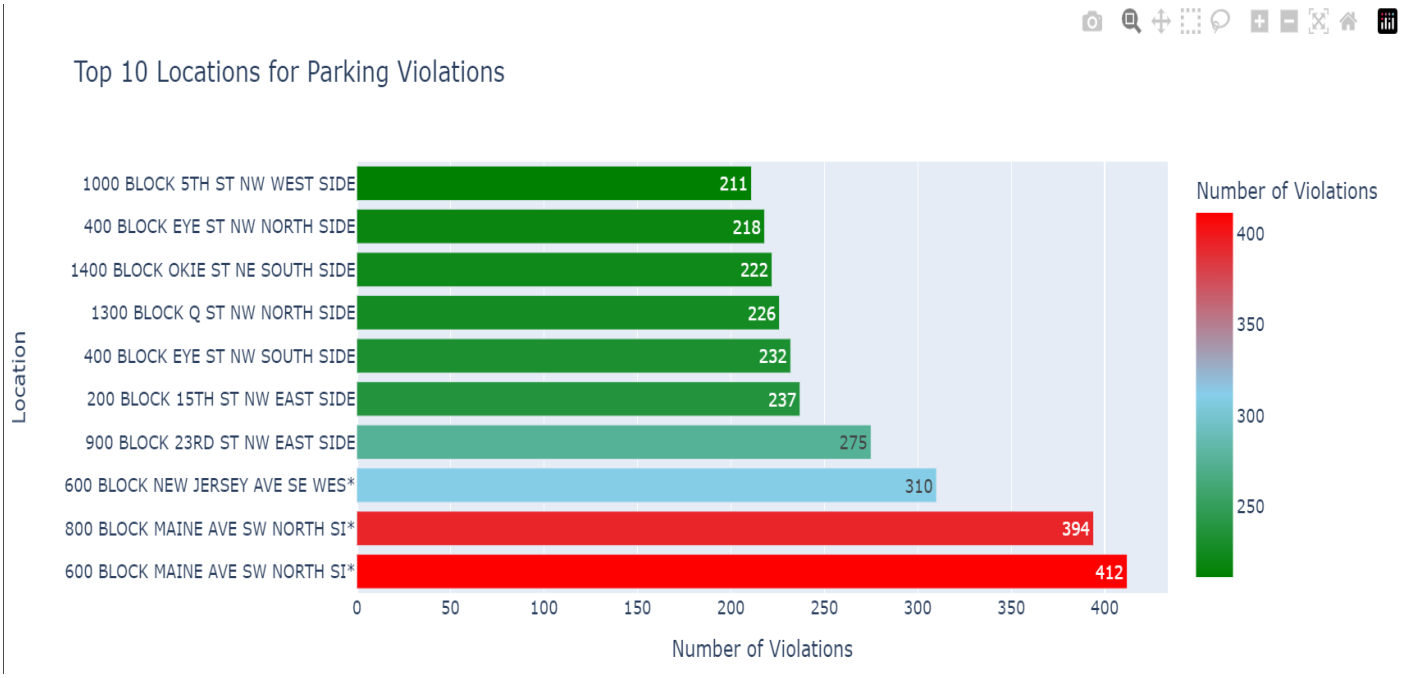
- Use groupby to group the data by ***"location"*** .
- Count the number of **"Ticket_Number"** entries for each location using .count() to explicitly calculate violations for grouped locations.
- Use reset_index() to convert the result back into a DataFrame.

Sort Locations by Violation Count:

- Use sort_values() to sort the location_counts DataFrame by the **"Ticket_Number"** column in descending order to rank locations by the number of violations.

Visual Created

The outlined steps have been visualized through an interactive horizontal bar chart showcasing the top 10 locations with the highest parking violations, based on the aggregated and sorted **location_counts** data. The chart highlights these locations on the y-axis, with the number of violations represented on the x-axis. Dynamic coloring, ranging from green for low violations to red for high violations. This visual effectively aids in identifying high-violation areas, providing a clear foundation for targeted analysis or intervention.



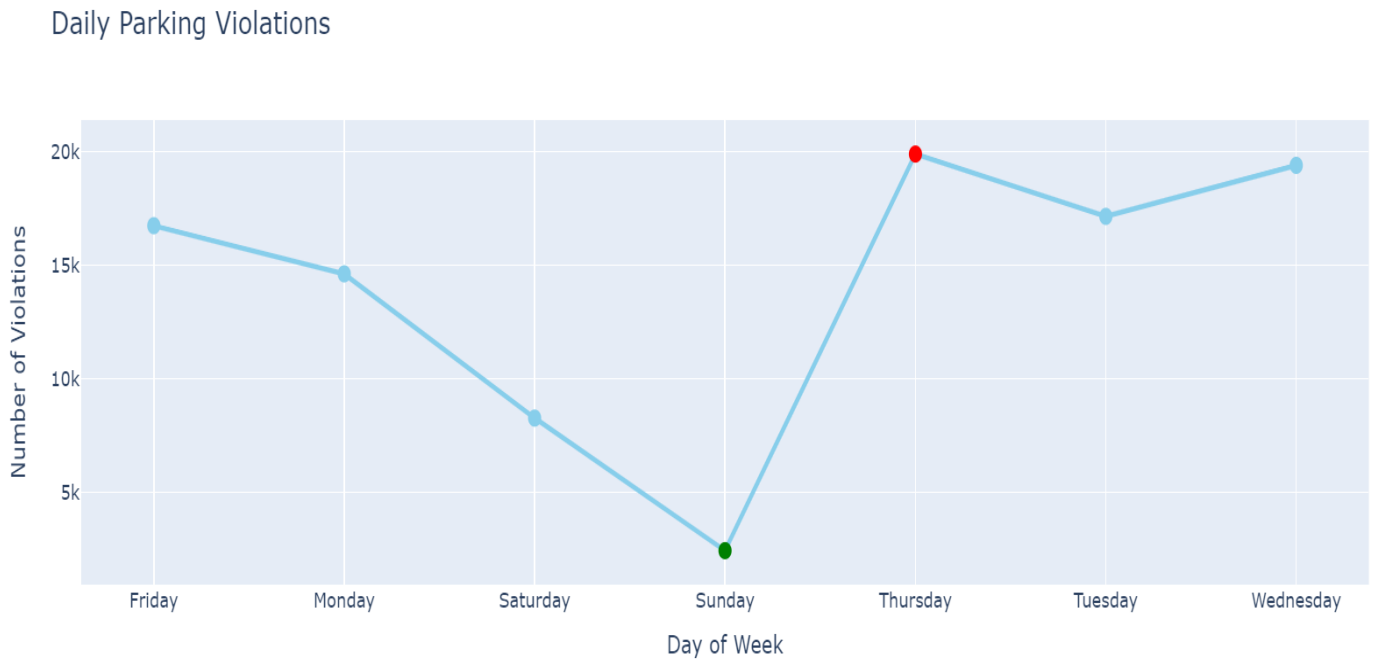
Daily Parking Violations: Interactive Line Chart

This line chart illustrates the daily parking violations reported throughout the week, highlighting key patterns and anomalies. It visualizes parking violations for each day, making it easy to observe trends such as sudden dips or spikes, which can provide insights into parking behaviour and enforcement activities. Key points are emphasized, with the day having the highest violations marked by a green dot to spotlight peak activity, and the day with the lowest violations marked by a red dot to emphasize the least active day. This approach ensures a clear and engaging representation of the data.

The visual is an **interactive line chart** that represents the number of parking violations **(TICKET_NUMBER)** across each day of the week **(ISSUE_DAY).** The design includes:

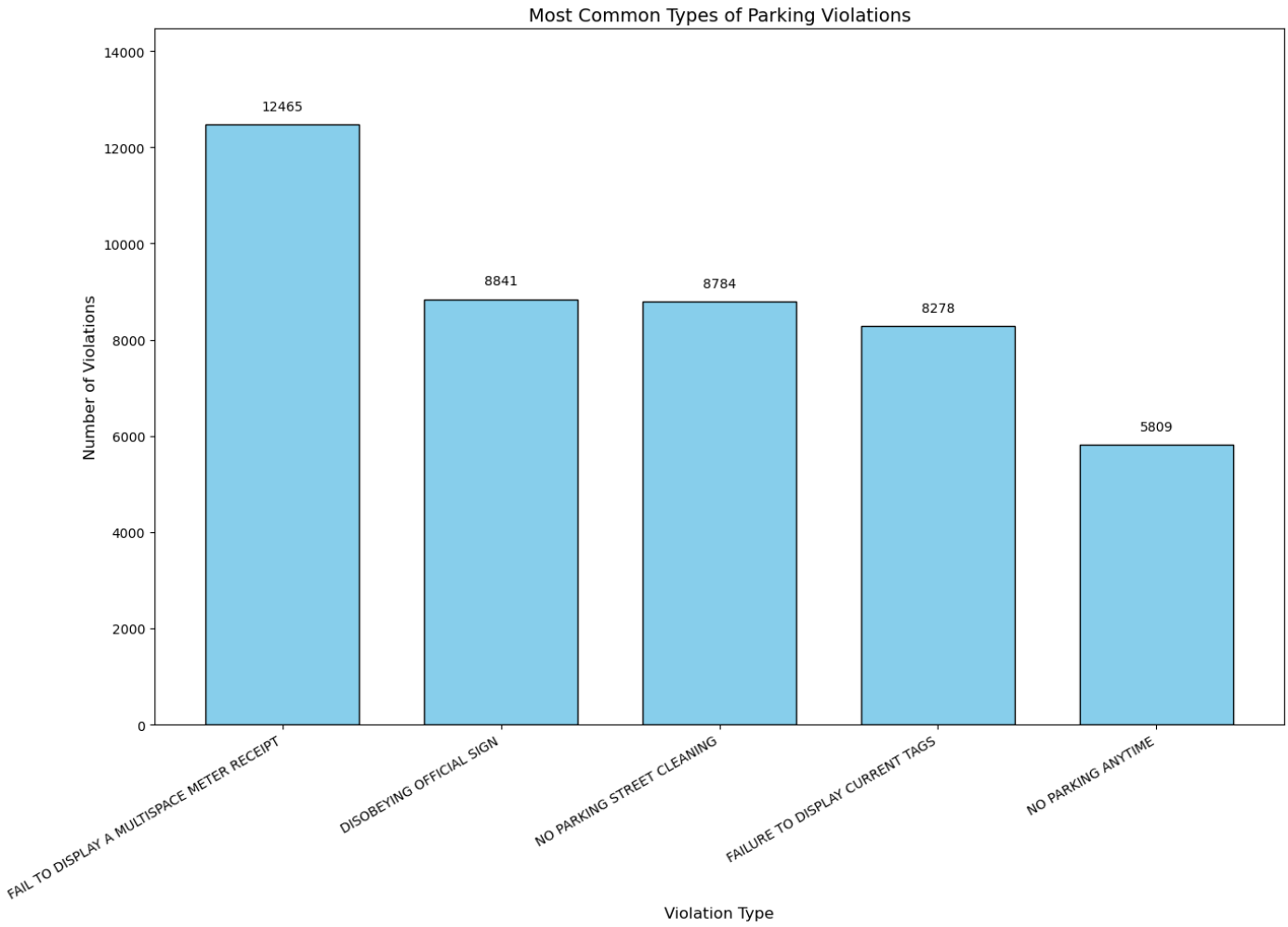
- **Line Trend**: A smooth line connects the points, showing the overall pattern of parking violations throughout the week.
- **Markers**: Each point on the line corresponds to a specific day's violation count.
    - **Green Marker**: Highlights the day with the lowest number of violations.
    - **Red Marker**: Highlights the day with the highest number of violations.
    - **Blue Markers**: Represent other days for comparison.
- **Custom Colours and Layout**: Enhances readability with larger markers, neutral line colour, and clear axis labels.
- **Interactive Features**: Allows users to hover over points for detailed data, making it intuitive for analysis.

This design effectively conveys both the trend and key highlights in parking violations over the week.



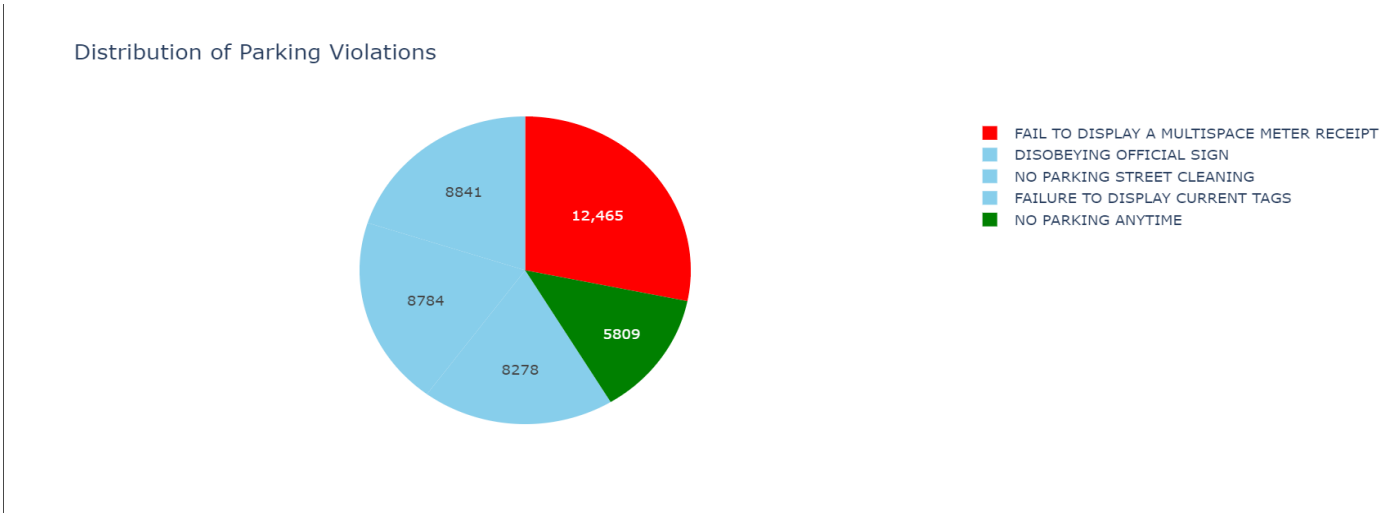
Most Common Types of Parking Violations: Insights from the Bar Chart

This bar chart highlights the top five most common parking violation types **(VIOLATION_PROC_DESC)** based on their frequency of occurrence. Each bar represents a specific violation type, with its height indicating the number of recorded violations, providing a clear visual comparison of the most prevalent infractions.



Distribution of Parking Violations: Insights from the Pie Chart

The pie chart visualizes the distribution of the top 5 parking violation types, with custom colors highlighting the highest (red) and lowest (green) counts, while the remaining types are styled in sky blue. Each slice represents a violation type proportional to its occurrence, with labels displaying exact counts and percentages for easy interpretation. The interactive design allows users to explore details by hovering over slices, while the title "Distribution of Parking Violations" adds clarity. This visualization effectively highlights the relative contribution of each violation type, emphasizing key trends for better analysis.



Stakeholder-Centric Insights

The insights presented in this report are tailored to prioritize the needs and decision-making requirements of primary stakeholders: city planners and law enforcement. Each section of the report is designed to address specific priorities for these audiences and can be used as follows:

City planners can leverage this analysis to enhance urban traffic and parking management. By identifying high-violation areas and peak violation times, they can revise parking regulations, such as implementing time restrictions or adjusting pricing to better align with demand. Geographic violation data can guide resource allocation, ensuring parking spaces, signage, and enforcement efforts are concentrated in areas with frequent violations. Additionally, observed trends in violation types and daily patterns provide a data-driven foundation for policy development, enabling the creation of more effective and relevant regulations that address specific urban parking challenges.

Law enforcement can use this analysis to optimize their enforcement strategies and enhance compliance with parking regulations. By identifying violation hotspots, they can allocate patrols or deploy monitoring systems, such as cameras, more effectively to areas with frequent infractions. Additionally, insights into the most common types of violations can guide focused training programs for officers and targeted public awareness campaigns, aiming to educate drivers and reduce these specific infractions. This data-driven approach enables more efficient enforcement and promotes adherence to parking rules.

Future Steps

To build upon this project, future work will focus on collecting and integrating more detailed and complete datasets, including demographic and economic factors, to enhance the depth and precision of the analysis. Implementing real-time monitoring systems will enable dynamic insights, allowing city planners and law enforcement to respond promptly to emerging trends. Additionally, policy simulation models can be employed to test and predict the impact of various parking management strategies, offering data-driven solutions for reducing violations. Engaging with the community through workshops or surveys will help incorporate public perspectives, ensuring that proposed recommendations are inclusive and effective. Finally, expanding the use of advanced analytics and visualization tools will cater to diverse stakeholder needs, improving the overall decision-making process.