Visualization of Complex Data

Enhancing Park Safety Through Data-Driven Analysis in NYC Parks

Project Report

Introduction

We embarked on this project with the aim of analyzing the prevalence of crime in New York parks, spurred by various factors. Media reports frequently highlight incidents of crime in public spaces, raising concerns about the safety of these areas, particularly parks frequented by vulnerable populations such as children and the elderly. Our objective is to harness data analysis to identify the safest parks for public use, support governmental efforts in bolstering security measures in parks with higher crime rates, and enhance public awareness regarding park safety. Personally invested as individuals who often seek relaxation and recreation in parks, we are keen on comprehending and ameliorating the safety of these spaces to ensure their continued accessibility and security for all. Additionally, this analysis could aid law enforcement agencies in allocating resources more efficiently to safeguard these crucial community areas.

SMART Questions

1. **How has crime density varied across different boroughs from 2015 to 2023?**

Utilizing our "Cumulative Crime Data by Borough: 2015-2023" chart, we intend to explore trends and fluctuations in crime rates across boroughs over the years to discern any significant changes or persistent hotspots.

2. **Is there a correlation between the park size and the incidence of crimes?**

Through our "Crime Density Analysis: Comparing Park Size to Crime Incidents in 2023," we aim to ascertain whether larger parks tend to experience more crimes due to their size and complexity, or if smaller parks are disproportionately affected.

3. **What are the trends in felony assaults within NYC parks over the past years?**

With our analysis titled "Felony Assault Trends in NYC Parks (2015-2023)," we will examine the progression of serious violent crimes within park boundaries to evaluate the efficacy of current safety measures and the necessity for targeted interventions.

Team Members and Contributions

1. **Ashwin Muthuraman:**

Responsibilities: Ashwin led the initial stages of the project, focusing primarily on data preparation and cleaning. He meticulously processed the data using Python, ensuring it was ready for analysis. Additionally, Ashwin took on the responsibility of compiling the final project report, skillfully integrating all findings and analyses into a cohesive document.

2. **Tenneti Srinivas Saiteja:**

Responsibilities: Saiteja played a pivotal role in conducting the Exploratory Data Analysis (EDA). Leveraging various statistical tools and techniques, he discovered patterns and insights within the dataset. Following the analysis, Saiteja crafted the project presentation, adeptly translating complex data insights into clear, visual slides that facilitated easy comprehension and discussion.

3. **Bharath Genji Mohana Ranga:**

Responsibilities: Bharath spearheaded the visualization efforts by creating the Tableau dashboard. He conceptualized and implemented several charts, including bar charts, area charts, and scatter plots, to visually depict the crime trends and analysis. His contributions were instrumental in enabling both the team and viewers to interact with the data findings intuitively.

Description of the Data

For our project, we utilized the NYC Park Crime dataset, a comprehensive record of crime incidents reported within New York City parks. The dataset contains 41,545 rows and 14 columns. This dataset covers various types of crimes, including murder, rape, robbery, felony assault, burglary, grand larceny, and grand larceny of motor vehicles. Each entry in the dataset includes details such as park names, boroughs, park sizes, and total incident counts, offering a detailed overview of safety within these public spaces.

We accessed the data from the official website of the NYPD, where crime statistics are regularly updated. Specifically, we obtained the dataset from https://www.nyc.gov/site/nypd/stats/crime-statistics/park-crime-stats.page. The data was provided in quarterly Excel files, spanning from the first quarter of 2015 to the fourth quarter of 2023. We downloaded these files for each quarter to ensure a continuous dataset for our analysis.

Additional information about the dataset's structure and usage can be found on New York City's open data portal, https://data.cityofnewyork.us/Public-Safety/NYC-Park-Crime-Data/ezds-sqp6/about_data.

Upon downloading the Excel files, we began data preparation by consolidating all quarterly data into a single CSV file. This consolidated dataset underwent cleaning to remove inconsistencies and prepare it for analysis. Using Python's Pandas library, we performed tasks such as data cleansing, normalization, and transformation. The resulting cleaned dataset facilitated our exploratory data analysis (EDA) and served as the data source for our Tableau dashboard, enabling us to visually represent our findings and insights effectively.

Analysis of Data Quality

Sample data:

PARK	BOROUGH	SIZE (ACRES)	CATEGORY	MURDER	RAPE	ROBBERY	FELONY AS	BURGLARY	GRAND LA	GRAND LA	TOTAL	QUARTER	YEAR
Pelham Bay Park	BRONX	2771.747	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Van Cortlandt Park	BRONX	1146.43	ONE ACRE OR LARGER	0	0	0	0	1	0	0	1	Q1	2015
Rockaway Beach And Boardwalk	QUEENS	1072.564	ONE ACRE OR LARGER	0	0	1	0	0	0	0	1	Q1	2015
Freshkills Park	STATEN ISLAND	913.32	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Flushing Meadows Corona Park	QUEENS	897.69	ONE ACRE OR LARGER	0	0	0	1	1	1	1	4	Q1	2015
Latourette Park & Golf Course	STATEN ISLAND	843.97	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Marine Park	BROOKLYN	798	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Belt Parkway/Shore Parkway	BROOKLYN/QUEENS	760.43	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Bronx Park	BRONX	718.373	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Franklin D. Roosevelt Boardwalk And Beach	STATEN ISLAND	644.35	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Alley Pond Park	QUEENS	635.514	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Prospect Park	BROOKLYN	526.25	ONE ACRE OR LARGER	0	0	1	0	0	1	0	2	Q1	2015
Forest Park	QUEENS	506.86	ONE ACRE OR LARGER	0	0	1	0	0	1	0	2	Q1	2015
Grand Central Parkway	QUEENS	460.16	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Ferry Point Park	BRONX	413.8	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Coney Island Beach & Boardwalk	BROOKLYN	399.203	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Cunningham Park	QUEENS	358	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Richmond Parkway	STATEN ISLAND	350.983	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Cross Island Parkway	QUEENS	326.895	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Great Kills Park	STATEN ISLAND	315.094	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Wolfe'S Pond Park	STATEN ISLAND	302.693	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Conference House Park	STATEN ISLAND	286.382	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Randall'S Island Park	MANHATTAN	256.111	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Rockaway Community Park	QUEENS	255.4	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Kissena Park	QUEENS	237.147	ONE ACRE OR LARGER	0	0	0	0	0	0	0	0	Q1	2015
Riverside Park	MANHATTAN	235.765	ONE ACRE OR LARGER	0	0	0	0	0	2	0	2	Q1	2015

Variable Descriptions:

1. **PARK:**
   - Type: Categorical
   - Description: Name of the park.
   - Distinct Categories: Names of different parks.

- Ranking: No ranking.

2. **BOROUGH:**
   - Type: Categorical
   - Description: Borough where the park is located.
   - Distinct Categories: Bronx, Queens, Staten Island, Brooklyn, Brooklyn/Queens, Manhattan.
   - Ranking: No ranking.
3. **SIZE (ACRES):**
   - Type: Quantitative
   - Description: Size of the park in acres.
   - Measurement: Continuous numerical measurement in acres.
4. **CATEGORY:**
   - Type: Categorical
   - Description: Category of the park based on size.
   - Distinct Categories: Categorized based on park size: ONE ACRE OR LARGER, PLAYGROUND LESS THAN ONE ACRE, BASKETBALL & PLAYGROUND LESS THAN ONE ACRE, BASKETBALL & POOL & PLAYGROUND LESS THAN ONE ACRE, POOL & RECREATION CENTER LESS THAN ONE ACRE, PLAYGROUND & POOL LESS THAN ONE ACRE, POOL LESS THAN ONE ACRE, BASKETBALL & POOL LESS THAN ONE ACRE, RECREATION CENTER LESS THAN ONE ACRE, BASKETBALL & REC CENTER LESS THAN ONE ACRE
   - Ranking: No ranking.
5. **MURDER:**
   - Type: Quantitative
   - Description: Number of murder incidents.
   - Measurement: Count of incidents.
6. **RAPE:**
   - Type: Quantitative

- Description: Number of rape incidents.
- Measurement: Count of incidents.

7. **ROBBERY:**
    - Type: Quantitative
    - Description: Number of robbery incidents.
    - Measurement: Count of incidents.
8. **FELONY ASSAULT:**
    - Type: Quantitative
    - Description: Number of felony assault incidents.
    - Measurement: Count of incidents.
9. **BURGLARY:**
    - Type: Quantitative
    - Description: Number of burglary incidents.
    - Measurement: Count of incidents.
10. **GRAND LARCENY:**
    - Type: Quantitative
    - Description: Number of grand larceny incidents.
    - Measurement: Count of incidents.
11. **GRAND LARCENY OF MOTOR VEHICLE:**
    - Type: Quantitative
    - Description: Number of grand larcenies of motor vehicle incidents.
    - Measurement: Count of incidents.
12. **TOTAL:**
    - Type: Quantitative
    - Description: Total number of incidents.
    - Measurement: Count of incidents.

13. QUARTER:

- Type: Categorical
- Description: Quarter of the year when the data was recorded.
- Distinct Categories: Q1, Q2, Q3, Q4.
- Ranking: No ranking.

14. YEAR:

- Type: Categorical
- Description: Year when the data was recorded.
- Distinct Categories: Years from 2015 to 2023.
- Ranking: No ranking.

Data Types:

- **Categorical Variables:** PARK, BOROUGH, CATEGORY, QUARTER, YEAR
- **Numerical Variables:** SIZE (ACRES), MURDER, RAPE, ROBBERY, FELONY ASSAULT, BURGLARY, GRAND LARCENY, GRAND LARCENY OF MOTOR VEHICLE, TOTAL

Descriptive Statistics:

	SIZE (ACRES)	RAPE	ROBBERY	FELONY ASSAULT \
count	41545.000000	41545.000000	41545.000000	41545.000000
mean	24.997487	0.004646	0.074064	0.051607
std	123.272937	0.073118	0.401264	0.367388
min	0.000000	0.000000	0.000000	0.000000
25%	0.915000	0.000000	0.000000	0.000000
50%	1.581000	0.000000	0.000000	0.000000
75%	4.945000	0.000000	0.000000	0.000000
max	2771.747000	3.000000	10.000000	21.000000

	GRAND LARCENY	GRAND LARCENY OF MOTOR VEHICLE	TOTAL
count	41545.000000	41545.000000	41545.000000
mean	0.098183	0.001829	0.238175
std	0.964672	0.057606	1.438132
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000
75%	0.000000	0.000000	0.000000
max	69.000000	4.000000	71.000000

Numeric Variables:

- **SIZE (ACRES):**
    - Mean: 24.997 acres
    - Standard Deviation: 123.273 acres
    - Minimum: 0 acres
    - Maximum: 2771.747 acres
- **Incidents:**
    - Mean incidents per category are very low, with most incidents having a mean close to 0.
    - Standard deviation varies significantly, indicating variance in incident counts.
    - The maximum count for each incident type is relatively low, but there are outliers.
- **TOTAL (Total Incidents):**
    - Mean: 0.238 incidents
    - Standard Deviation: 1.438 incidents
    - Minimum: 0 incidents
    - Maximum: 71 incidents

Categorical Variables:

- **BOROUGH:**
    - Contains six unique boroughs.
    - Bronx has the highest number of records.
    - There are some missing values denoted by 'nan'.
- **QUARTER:**
    - Contains four quarters (Q1, Q2, Q3, Q4).
    - Each quarter has data, indicating regular data collection.
- **YEAR:**
    - Data spans from 2015 to 2023, with no missing years.

Main analysis

Data Preparation

1. Logging Setup

- **Purpose:** Sets up a logging mechanism to track the execution and any issues of the script.
- **Details:** It logs the events into a file named data_processing.log at the INFO level, and the log entries include a timestamp, the log level, and a message.

2. Function: process_yearly_data(year, base_path)

- **Purpose:** Processes data for a specific year by loading data from Excel files for each quarter, cleaning, and combining them.
- **Details:**
  - Loop through Quarters: Iterates through the four quarters of the year.
  - Error Handling: Uses a try-except block to handle file reading or data processing errors.
  - Data Loading: Reads Excel files with varying rows to skip based on specific years and quarters.
  - Data Cleaning:
    - Normalizes column names to uppercase.
    - Filters out summary rows where the 'PARK' column equals "Total".
    - Adds 'QUARTER' and 'YEAR' as additional columns.
  - Concatenation: Combines data from all quarters into a single DataFrame.
  - Further Cleaning: Fills missing values, standardizes text format, and strips whitespace.

```python
def process_yearly_data(year, base_path):
    quarterly_data = []

    for quarter in ['Q1', 'Q2', 'Q3', 'Q4']:
        try:
            file_path = os.path.join(base_path, f'nyc-park-crime-stats-{quarter}-{year}.xlsx')

            # Adjust skiprows based on specific years and quarters
            if (year == 2015 and quarter in ['Q1', 'Q2', 'Q3']) or \
               (year == 2018 and quarter == 'Q2') or \
               (year == 2021 and quarter == 'Q1'):
                skiprows = 4
            else:
                skiprows = 3

            # Load the data
            data = pd.read_excel(file_path, skiprows=skiprows)

            # Normalize column names to uppercase
            data.columns = data.columns.str.upper()

            # Filtering out rows where PARK equals "Total" (after normalizing to uppercase)
            data = data[data['PARK'] != 'Total']

            # Add quarter and year information
            data['QUARTER'] = quarter
            data['YEAR'] = year

            quarterly_data.append(data)
```

3. **Main Script Execution**

- **Purpose:** Executes data processing for each year from 2015 to 2023 and combines all the results.
- **Details:**
    - Data Processing: Calls process_yearly_data for each year and collects the yearly DataFrames.
    - Concatenation and Saving:
        - Combines all the yearly DataFrames into one final DataFrame.
        - Normalizes column names again.
        - Converts 'YEAR' and 'QUARTER' into categorical data types for better analysis performance.
        - Filters out any remaining rows where 'PARK' is "Total".
        - Saves the final consolidated data to a CSV file.
    - Logging: Logs the completion of data processing or errors if no data was processed.

```python
# Merge and save the processed data
if all_years_data:
    final_data = pd.concat(all_years_data, ignore_index=True)

    # Normalize column names for the final DataFrame if not already done
    final_data.columns = final_data.columns.str.upper()

    # Convert 'YEAR' and 'QUARTER' to categorical types for analysis
    final_data['YEAR'] = pd.Categorical(final_data['YEAR'])
    final_data['QUARTER'] = pd.Categorical(final_data['QUARTER'], categories=['Q1', 'Q2', 'Q3', 'Q4'], ordered=True)

    final_data = final_data[final_data['PARK'].str.upper() != 'TOTAL']

    # Save the consolidated data
    final_file_path = os.path.join(processed_path, 'nyc_park_crime_data_2015_to_2023.csv')
    final_data.to_csv(final_file_path, index=False)
    logging.info("Data processing completed successfully and saved.")
else:
    logging.error("No data processed.")
```

Data Cleaning

Check for Missing Values and Removal: The initial check using final_data.isnull().sum() identifies any columns with missing values. Removing missing data (final_data.dropna(subset=['PARK', 'BOROUGH'], inplace=True)) ensures that subsequent analyses aren't skewed by null entries, particularly in key categorical fields like 'PARK' and 'BOROUGH'.

Conversion of Data Types and Handling Errors: Conversion of 'MURDER' and 'BURGLARY' columns to numeric types using pd.to_numeric() with errors='coerce' ensures that all data in these columns are numerical and non-numeric entries are set to NaN. This step is crucial for accurate statistical calculations and avoids data type mismatch errors during analysis.

Data Pre-processing

Setting Categorical Data: Transforming the 'YEAR' and 'QUARTER' columns into categorical data types (pd.Categorical()) helps in treating these columns appropriately during visualization and model building. It maintains the logical ordering of data (e.g., quarters from Q1 to Q4), which is essential for time-series analysis or seasonal comparisons.

Handling Ordered Categories: By explicitly ordering the 'QUARTER' data (categories=['Q1', 'Q2', 'Q3', 'Q4']), you maintain a meaningful sequence which is crucial for any analysis that may need to recognize this inherent order, such as trend analysis across quarters.

Feature Engineering

Calculation of Crime Rate Per Acre: Adding a new column 'CRIME RATE PER ACRE' as final_data['TOTAL'] / final_data['SIZE (ACRES)']. This metric provides a standardized measure of crime density, which can be particularly useful for comparative analysis across parks of varying sizes and can influence decisions on law enforcement resource allocation.

Exploratory Data Analysis (EDA)

1. Total Crimes per Borough Over the Years

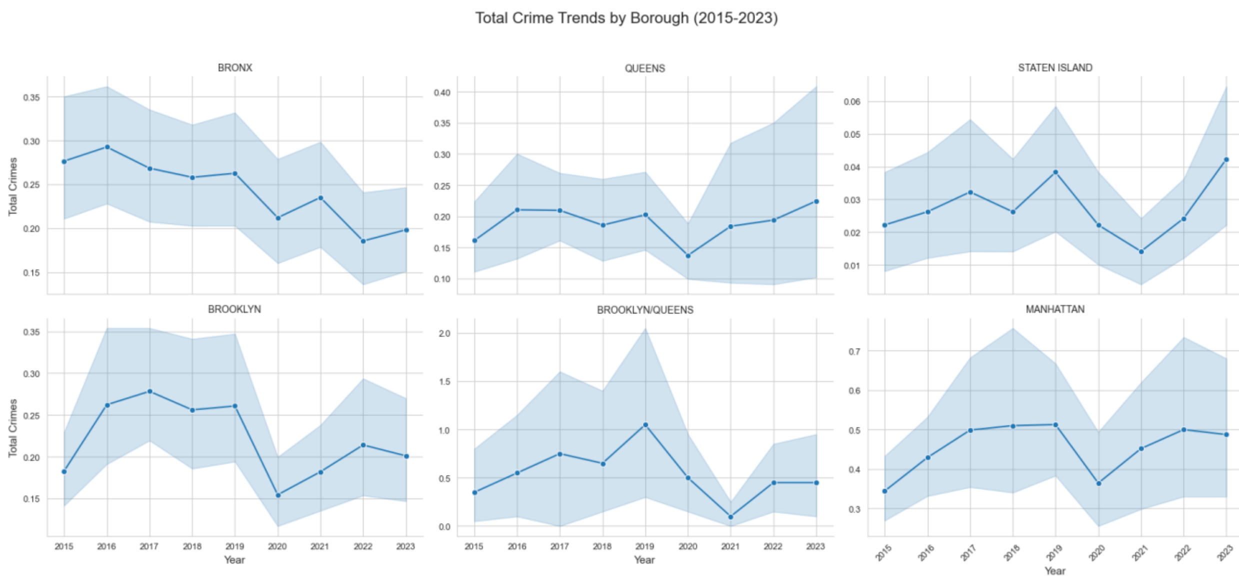
Findings:

- Varying Crime Rates: Each borough displays unique trends in crime rates over the years, indicating localized crime patterns influenced by borough-specific factors.
- Temporal Fluctuations: Notable fluctuations within each borough over time suggest changes in law enforcement policies, socio-economic factors, or other events affecting crime rates.

Process:

- Used Seaborn's FacetGrid to visualize total crime trends by borough from 2015 to 2023.
- Set custom formatting for better readability and aesthetics.

```python
# Set the aesthetic style of the plots
sns.set_style("whitegrid")

# Create the FacetGrid Line Plot
g = sns.FacetGrid(final_data, col='BOROUGH', col_wrap=3, height=4, sharey=False, aspect=1.5)
g.map(sns.lineplot, 'YEAR', 'TOTAL', marker='o')

# Enhance the plot
g.fig.suptitle('Total Crime Trends by Borough (2015-2023)', fontsize=16, y=1.03)
g.set_titles('{col_name}', fontsize=12)
g.set_axis_labels('Year', 'Total Crimes')

# Improve the aesthetics
g.set(xticks=final_data['YEAR'].unique())
plt.xticks(rotation=45)

# Iterate through axes to set custom formatting
for ax in g.axes.flatten():
    ax.tick_params(labelsize=9)
    ax.yaxis.get_label().set_fontsize(11)
    ax.xaxis.get_label().set_fontsize(11)

# Adjust the layout
g.tight_layout()

# Show the plot
plt.show()
```



2. Correlation Matrix of Crime Types

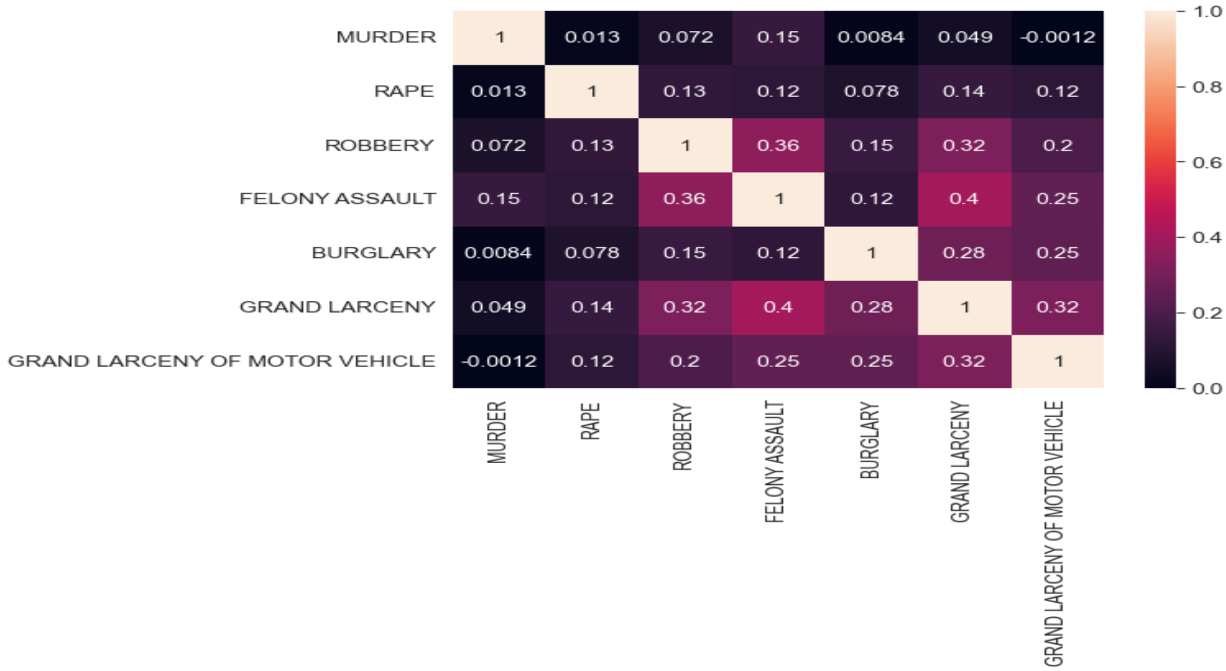
Findings:

- **Strength of Correlations:** Identified notable correlations between certain crime types, such as robbery and felony assault, and grand larceny and grand larceny of motor vehicle.
- **Color Coding:** Used a color gradient to enhance visual interpretation of correlations.

Process:

- Created a correlation matrix using Seaborn's heatmap function.
- Annotated the heatmap for clarity.

```
# 2. Correlation Matrix of Crime Types
crime_types = ['MURDER', 'RAPE', 'ROBBERY', 'FELONY ASSAULT', 'BURGLARY', 'GRAND LARCENY', 'GRAND LARCENY OF MOTOR VEHICLE']
corr_matrix = final_data[crime_types].corr()
sns.heatmap(corr_matrix, annot=True)
plt.show()
```



3. Crime Rate by Borough

Findings:

- **Differential Crime Rates:** Manhattan exhibits the highest crime rate per acre, while Staten Island has the lowest.
- **Comparison Among Boroughs:** Identified variations in crime rates among different boroughs.

Process:

- Calculated the mean crime rate per acre for each borough.
- Visualized the results using a heatmap for better comparison.

```
# 2. Correlation Matrix of Crime Types
crime_types = ['MURDER', 'RAPE', 'ROBBERY', 'FELONY ASSAULT', 'BURGLARY', 'GRAND LARCENY', 'GRAND LARCENY OF MOTOR VEHICLE']
corr_matrix = final_data[crime_types].corr()
sns.heatmap(corr_matrix, annot=True)
plt.show()
```

```
BOROUGH
BRONX                   0.058767
BROOKLYN                0.056148
BROOKLYN/QUEENS         0.003235
MANHATTAN               0.090631
QUEENS                  0.041874
STATEN ISLAND           0.008628
Name: Crime Rate per Acre, dtype: float64
```

4. Quarterly Crime Totals

Findings:

- **Seasonal Variability:** Q3 shows the highest crime total, suggesting a seasonal peak in crime rates during this period.
- **Comparison Across Quarters:** Notable fluctuations in crime rates across quarters, with Q3 consistently showing higher crime rates.

Process:

- Grouped the data by quarter and calculated the total number of crimes reported in each quarter.
- Visualized the results using a bar chart for easy comparison.

```
# 5. Total Crimes per Quarter
total_crimes_per_quarter = final_data.groupby('QUARTER')['TOTAL'].sum()

print(total_crimes_per_quarter)
```

```
QUARTER
Q1       984.0
Q2      2699.0
Q3      4098.0
Q4      2114.0
Name: TOTAL, dtype: float64
```

5. Total Crimes per Quarter

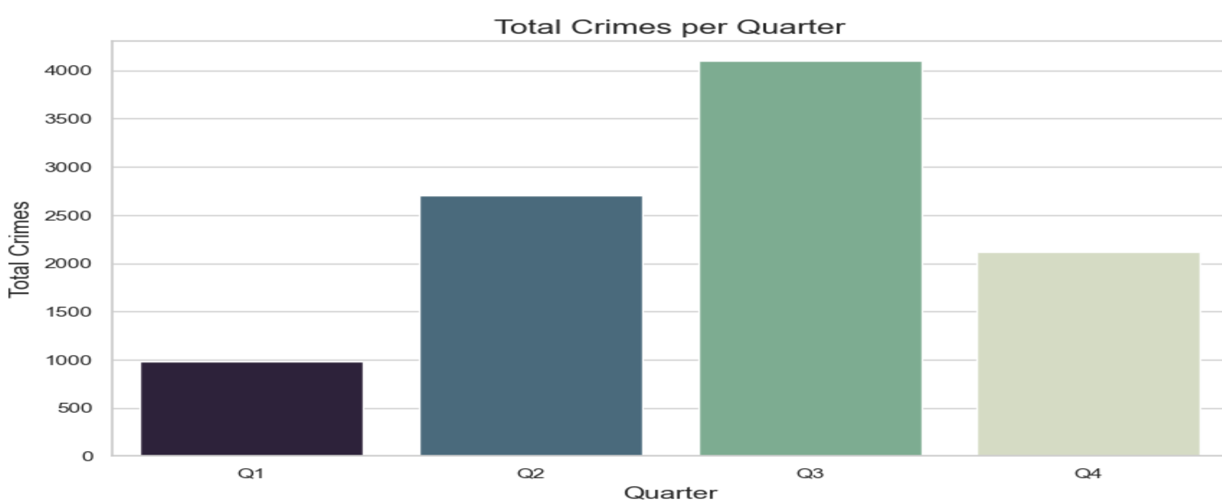
Findings:

- **Seasonal Variability:** Identified a significant increase in crime during Q3, indicating a seasonal peak in criminal activity.
- **Comparison Across Quarters:** Noted fluctuations in crime rates across quarters, with Q3 consistently showing the highest rates.

Process:

- Visualized total crimes per quarter using a bar chart to highlight seasonal patterns.

```python
sns.set(style="whitegrid")

# Creating a cubehelix palette that starts light and ends dark
palette = sns.cubehelix_palette(start=.5, rot=-.75, as_cmap=False, reverse=True, n_colors=len(total_crimes_per_quarter))

# Plotting total crimes per quarter with the custom palette
plt.figure(figsize=(10, 6))
sns.barplot(x=total_crimes_per_quarter.index, y=total_crimes_per_quarter.values, palette=palette)
plt.title('Total Crimes per Quarter', fontsize=16)
plt.xlabel('Quarter', fontsize=14)
plt.ylabel('Total Crimes', fontsize=14)
plt.show()
```



6. Mean Crime Rate per Acre for Each Borough

Findings:

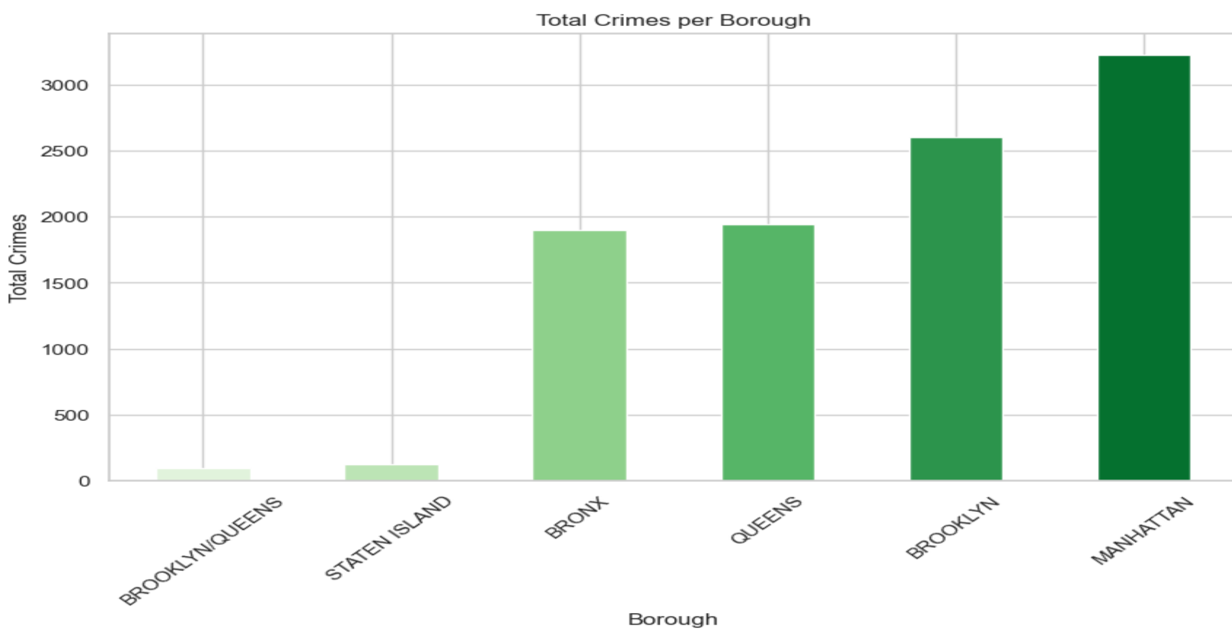
- **Differential Crime Rates:** Manhattan has the highest average crime rate per acre, significantly surpassing other boroughs.

- **Comparison Among Boroughs:** Noted variations in crime rates per acre among different boroughs.

Process:

- Calculated the mean crime rate per acre for each borough.
- Visualized the results using a bar chart for comparison.

```
# Grouping data by 'BOROUGH' and summing the 'TOTAL' crimes
total_crimes_per_borough = final_data.groupby('BOROUGH')['TOTAL'].sum().sort_values()

# green color palette
palette = sns.color_palette("Greens", n_colors=len(total_crimes_per_borough))

# Plotting the total crimes per borough
plt.figure(figsize=(10, 6))
total_crimes_per_borough.plot(kind='bar', color=palette)
plt.title('Total Crimes per Borough')
plt.xlabel('Borough')
plt.ylabel('Total Crimes')
plt.xticks(rotation=45)
plt.show()
```



7. Crime Rate per Acre Distribution

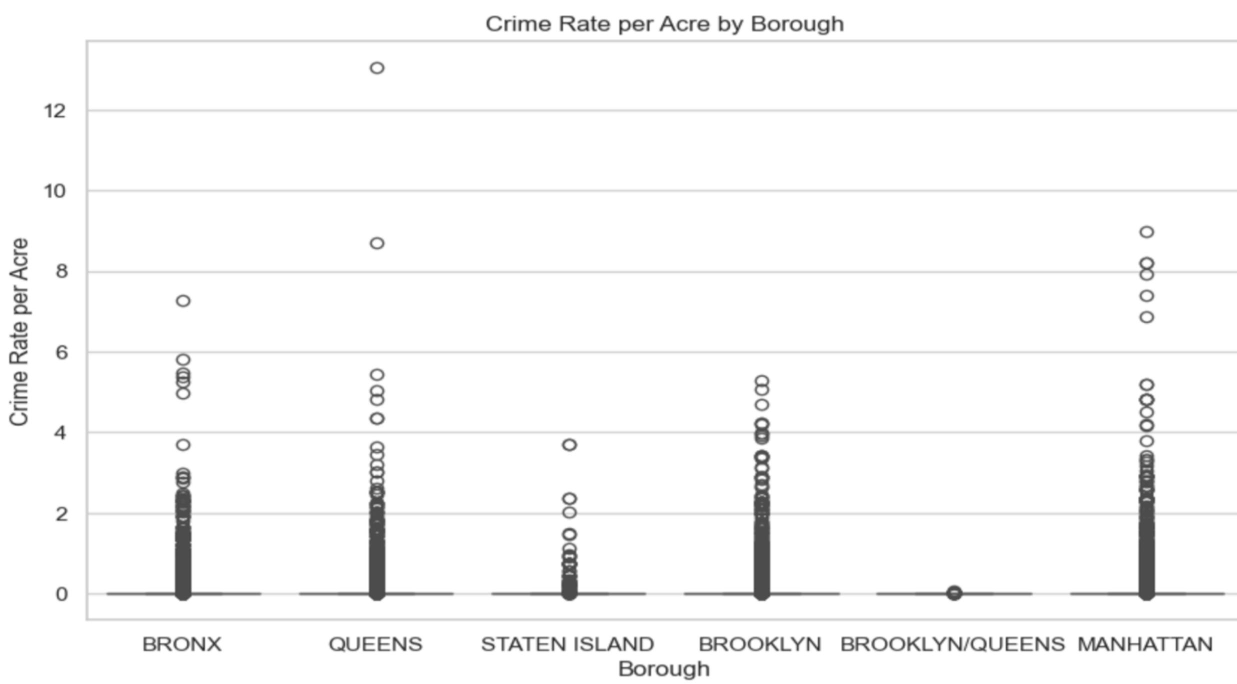
Findings:

- **Distribution and Variability:** Manhattan exhibits the widest range of crime rates, while Staten Island and the combined Brooklyn/Queens area display tighter clusters of data points.
- **Comparative Analysis:** Compared the distribution and variability of crime rates among different boroughs.

Process:

- Visualized crime rate per acre distribution using a boxplot to show variability and range of data points.

```
plt.figure(figsize=(10, 6))
sns.boxplot(data=final_data, x='BOROUGH', y='CRIME RATE PER ACRE')
plt.title('Crime Rate per Acre by Borough')
plt.xlabel('Borough')
plt.ylabel('Crime Rate per Acre')
plt.show()
```



8. Heatmap of Crimes by Year and Quarter

Findings:

- **Temporal Trends and Seasonal Patterns:** Identified consistent patterns of higher crime rates in Q3 across most years, suggesting a seasonal increase in crime during these months.
- **Year-to-Year Comparison:** Noted variations in crime rates across years, with notable peaks and drops in certain years.

Process:

- Created a heatmap to visualize total crimes by year and quarter for temporal trend analysis.

```
# Preparing data for heatmap
crime_heatmap_data = final_data.pivot_table(values='TOTAL', index='YEAR', columns='QUARTER', aggfunc='sum')

# Plotting the heatmap
plt.figure(figsize=(12, 8))
sns.heatmap(crime_heatmap_data, annot=True, cmap='Blues', fmt='g')
plt.title('Heatmap of Total Crimes by Year and Quarter', fontsize=16)
plt.xlabel('Quarter', fontsize=14)
plt.ylabel('Year', fontsize=14)
plt.show()
```

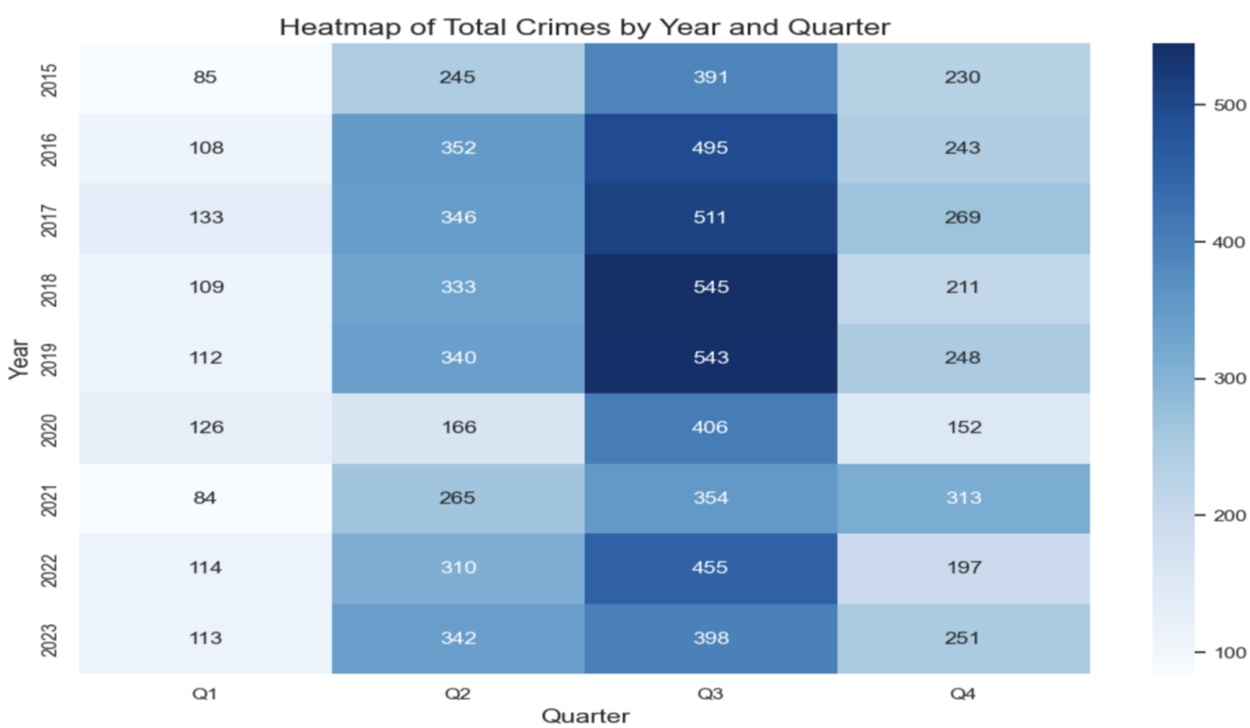
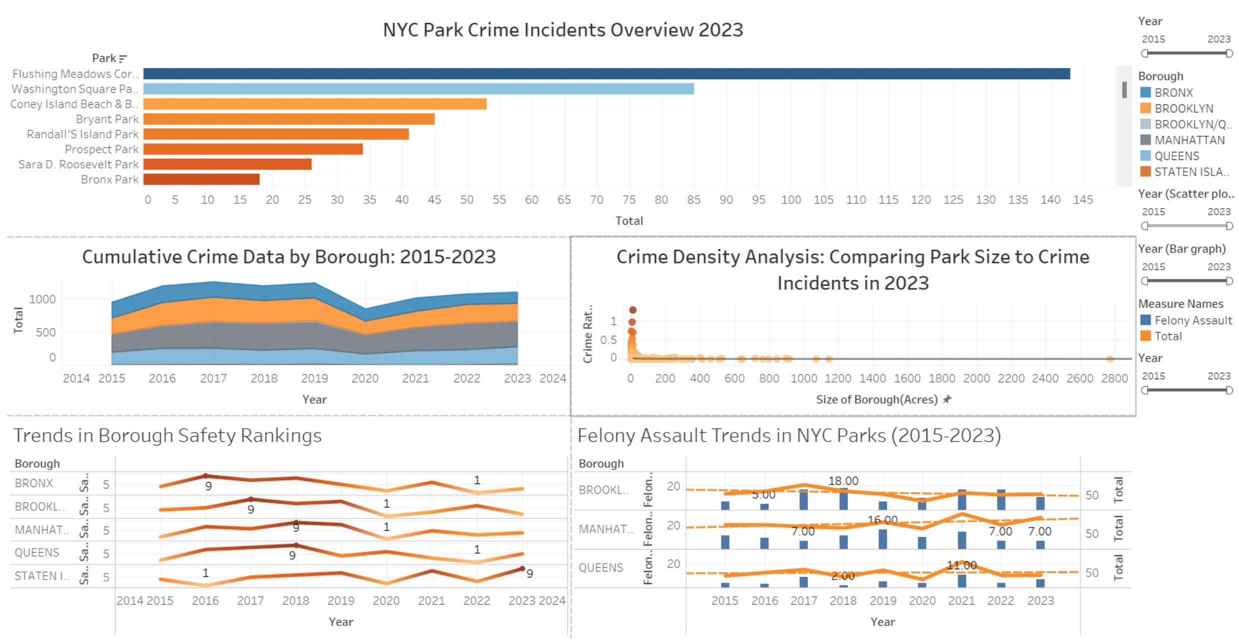


Tableau Analysis

Dashboard



1. NYC Park Crime Incidents Overview 2023

- **Findings**:
    - **Variation in Crime Rates Across Parks**: Notable differences in crime incidents were observed, with Flushing Meadows Corona Park, Washington Square Park, and Coney Island Beach & Boardwalk recording the highest numbers. This indicates these larger, more frequented parks are prone to higher crime rates.
    - **Low Crime Incidence Parks**: Parks like Fort Washington Park, Franz Sigel Park, and Battery Park showed significantly fewer incidents, suggesting that these less crowded or smaller parks might be safer.
- **Process**:
    - **Visualization Technique**: Used a bar chart to effectively display the number of incidents per park, allowing for easy comparison across various locations.
    - **Data Preparation**: Ensured accurate park names and incident counts to maintain data integrity for the visualization.

2. Cumulative Crime Data by Borough: 2015-2023

- **Findings**:
  - **Borough-specific Trends**: Brooklyn reported the highest crime rates consistently over the years, while Staten Island had the fewest incidents. The data also revealed a significant drop in crime rates across all boroughs in 2020, likely influenced by the COVID-19 pandemic.
- **Process**:
  - **Visualization Technique**: Employed a stacked area chart to depict changes over time within each borough, highlighting trends and anomalies.
  - **Analytical Focus**: Paid special attention to external factors such as policy changes and public health crises that might influence crime trends.

3. Crime Density Analysis: Comparing Park Size to Crime Incidents in 2023

- **Findings**:
  - **Higher Crime Rates in Smaller Parks**: Smaller parks tend to have higher crime rates per acre, suggesting that the density of visitors and smaller areas concentrate crime incidents more than in larger parks.
  - **Decrease in Crime Rate with Increasing Park Size**: Larger parks generally show lower crime rates per acre, supporting the notion that open spaces with fewer crowd interactions per area unit reduce crime opportunities.
- **Process**:
  - **Visualization Technique**: Utilized a scatter plot to correlate park size with crime rates, providing a clear visual of how crime density varies with park area.
  - **Data Handling**: Calculated crime rates per acre for each park and then plotted these against park sizes to ensure precise and relevant insights.

4. Felony Assault Trends in NYC Parks (2015-2023)

- **Findings**:
  - **Variable Trends Across Boroughs**: Brooklyn saw fluctuations in felony assault cases with peaks around 2019, while Manhattan displayed a steadier trend. Queens showed fewer incidents overall with a peak in 2021.

    - **General Decline in Recent Years**: Recent data suggest a downward trend in felony assaults, indicating potentially improved park safety or effective policing strategies.
- **Process**:
    - **Visualization Technique**: Line charts with dual axes were used to display both the raw counts of felony assaults and the total crimes for context, enhancing understanding of trends against the broader crime backdrop.
    - **Analytical Approach**: Analyzed year-over-year changes and examined potential causes for fluctuations, such as law enforcement efforts or community initiatives.

5. Trends in Borough Safety Rankings

- **Findings**:
    - **Fluctuating Rankings Over Time**: Safety rankings for boroughs fluctuated annually, with Brooklyn and Manhattan alternating between safest and least safe over the period reviewed.
    - **Current Trends in 2023**: Recent data show Brooklyn and Staten Island as less safe, with the Bronx and Queens perceived as safer, highlighting shifting safety dynamics within NYC.
- **Process**:
    - **Visualization Technique**: Used a line chart to track safety rankings over time, providing a clear view of the upward and downward trends for each borough.
    - **Contextual Analysis**: Considered historical data and current events that might influence safety perceptions and actual crime rates across boroughs.

Key Findings

1. **Variation in Crime Rates Across Parks:**

Not all parks in NYC are equally safe. Parks like Flushing Meadows Corona Park and Washington Square Park exhibit higher crime rates, likely due to their size and popularity. Conversely, parks such as Fort Washington Park and Battery Park experience fewer incidents, possibly owing to their smaller size and lower foot traffic.

2. **Borough Crime Trends:**

Brooklyn consistently tops the list with the highest number of reported crimes, while Staten Island consistently reports the fewest. A notable decrease in crime rates across all boroughs around 2020 suggests the influence of external factors such as policy changes and the COVID-19 pandemic on crime.

3. **Crime Density in Parks:**

Analysis indicates that smaller parks tend to have higher crime rates per acre compared to larger parks. This implies that park size may play a role in crime prevention, with larger parks offering more space for people to disperse, potentially reducing opportunities for criminal activity.

4. **Felony Assault Trends:**

While fluctuations in felony assault cases have been observed across Brooklyn, Manhattan, and Queens, all boroughs show a general decline in recent years. This trend suggests potential improvements in park safety measures or law enforcement strategies.

5. **Borough Safety Rankings:**

Safety rankings for boroughs have varied over time. In 2023, Brooklyn and Staten Island are perceived as less safe, while the Bronx and Queens are considered safer. These rankings underscore the diverse safety conditions across different areas of NYC.

Overall, the analysis unveils the intricate dynamics of crime in NYC parks, influenced by factors such as park size, borough demographics, and external events like the COVID-19 pandemic.

Conclusion

Limitations: While our analysis has provided valuable insights, it is essential to acknowledge several limitations. Firstly, aggregating data from multiple boroughs may obscure specific trends and factors affecting crime rates in each distinct area. Disaggregating this data could offer more precise insights into localized crime dynamics. Additionally, limited discussion on seasonal adjustments raises questions about the true underlying trends, highlighting the need for statistical adjustments to account for seasonal effects. Moreover, the lack of integration of contextual factors such as socio-economic, demographic, or policy changes leaves gaps in understanding the comprehensive drivers of crime trends.

Future Directions: To address these limitations, future analyses could incorporate additional contextual data such as economic indicators, population density changes, or law enforcement strategies to gain a more comprehensive understanding of crime dynamics. Longitudinal studies would also be beneficial in assessing the long-term effects of policies and global events on crime trends. Furthermore, applying predictive analytics to crime data could facilitate the forecasting of future trends, aiding in proactive crime prevention and resource allocation.

Lessons Learned: Our analysis has highlighted the profound impact of external factors such as the COVID-19 pandemic and associated policy changes on crime dynamics, emphasizing the need for adaptive law enforcement strategies. Additionally, the utility of visual data representation in conveying complex information and trends quickly is underscored, emphasizing its importance for policymakers, law enforcement agencies, and the public. Moreover, insights into seasonal crime fluctuations emphasize the importance of planning law enforcement strategies and public safety measures, accordingly, ensuring effective crime prevention efforts throughout the year.