**CS5803 Data Visualization**

**Coursework (Reassessment) for 2023/24**

**Student ID:** 2358963

**Name:** Aman Khokhar

**Project:** Analyzing and Visualizing Crime Data in London

**Introduction**

Crime analysis is essential for law enforcement agencies to understand patterns, allocate resources effectively, and implement strategies to reduce crime rates. This analysis focuses on crime data for Barking and Dagenham, a borough in London, covering the period from July 2022 to June 2024. The data is sourced from reliable government portals, providing detailed insights into various crime categories and their occurrences over time.

**Description of the Dataset**

**Dataset Source**

1. **Recorded Crime Summary**:
   * **Source URL**: https://data.london.gov.uk/dataset/recorded\_crime\_summary
   * **Description**: This dataset provides a summary of recorded crimes in London, categorized by crime type and borough. It includes monthly counts of different crimes, allowing for trend analysis and comparison across boroughs.
2. **Statistical GIS Boundary Data**:
   * **Source URL**: https://www.data.gov.uk/dataset/statistical-gis-boundary-files-for-london
   * **Description**: This dataset contains geographical boundaries, names, longitude and latitude coordinates, area in hectares, and geographical maps of London boroughs. It is useful for spatial analysis and mapping crime data to specific locations.

**Data Dictionary**

* **Crime**: The broad category of the crime (e.g., ARSON AND CRIMINAL DAMAGE, BURGLARY).
* **Sub-Crime**: The specific type of crime within the broader category (e.g., ARSON, CRIMINAL DAMAGE).
* **Borough Name**: The location where the crime was recorded (in this case, Barking and Dagenham).
* **Monthly Columns (22-Jul to 24-Jun)**: The number of occurrences of the specific crime in each month from July 2022 to June 2024.
* **GIS Boundary Data**:
  + **Borough Name**: The name of the borough.
  + **Longitude**: The longitudinal coordinate of the borough.
  + **Latitude**: The latitudinal coordinate of the borough.
  + **Area (hectares)**: The area of the borough in hectares.
  + **Geographical Map**: The map representing the geographical boundaries of the London borough.

**User Persona and Questions**

**Background**: Ben has a background in criminology and data science. He is responsible for analyzing crime data across London to support policy-making and resource allocation. John works closely with law enforcement agencies, local councils, and community organizations to develop strategies for crime reduction and public safety improvement.  
**Goals**:

* To identify and analyze crime trends across London over time.
* To understand the geographical distribution of crime within London.
* To determine which boroughs are most affected by different types of crime.

**Planned Questions**

1. **What are the crime rate trends in London over time?**
   * Analysis of overall crime trends in London from July 2022 to June 2024.
   * Identification of any increasing or decreasing patterns in specific crime categories.
2. **How is crime spread out through London?**
   * Examination of the geographical distribution of crime across different London boroughs.
   * Mapping of crime hotspots and areas with lower crime rates.
3. **Which boroughs are affected the most by crime?**
   * Identification of boroughs with the highest and lowest crime rates.
   * Comparison of crime severity and frequency across boroughs to determine the most affected areas.
   * Analysis of specific crime types that are particularly prevalent in certain boroughs.

**Requirements for Data Visualization**

**Functional Requirements**

The project requires a variety of visualizations to effectively communicate the trends and distribution of crime across London. **Time series plots** are essential for illustrating the trends in crime rates over time for each category. These line charts will help identify increasing or decreasing patterns, aiding in resource allocation and strategic planning. Additionally, **monthly comparison charts** such as bar charts or heatmaps will allow for an in-depth analysis of crime rates within each borough, highlighting patterns and anomalies that may require further investigation.

**Geographical distribution visualizations** are crucial for understanding how crime is spread across different boroughs. Choropleth maps will visually represent the intensity of crime in each area, utilizing GIS boundary data for accuracy. These maps will help identify crime hotspots, which can be further analyzed using density plots or cluster maps to pinpoint specific areas requiring attention.

For a more detailed analysis, borough-specific visualizations will be created. Ranking charts will show which boroughs have the highest and lowest crime rates, providing a clear comparison. Detailed profiles for each borough, displayed on interactive dashboards, will break down different crime types over time, offering a comprehensive view of crime patterns in each area.

Interactive dashboards will enhance the user experience by allowing users to filter data by crime type, borough, and time period. These dashboards will also feature drill-down options for more granular analysis. Hover and tooltip information will provide additional details and statistics, making the data more accessible and informative.

Finally, comparison visualizations will be used to identify differences and trends across various dimensions. Side-by-side comparisons will make it easy to compare different boroughs or crime types, while year-over-year comparisons will highlight long-term trends. These visualizations will support a thorough and nuanced understanding of crime dynamics in London.

**Non-Functional Requirements**

The visualizations must be highly usable, with a user-friendly interface that is easy to navigate and understand, even for users without a technical background. Performance is also critical; the visualizations should be responsive and capable of handling large datasets efficiently. Accuracy is paramount to ensure that the data is represented correctly, providing reliable insights for decision-making. Finally, the visualizations should be accessible, complying with relevant standards to ensure that they can be used by individuals with disabilities.

**Design Analysis**

To effectively communicate the crime data analysis, we will compare the original prototype design with the final implementation. This comparison will illustrate the evolution of the design and the rationale behind any changes.

**Original Prototype Design (Paper Landscape)**

The original prototype, presented as hand-drawn sketches or wireframes, provided a basic layout for visualizing crime data. The header was simply designed with a title and basic navigation elements. The crime trend charts, intended to show trends over time, were basic line charts with monthly data points for different crime categories. For geographical distribution, the prototype included a simple choropleth map of London, color-coded by crime intensity. Borough comparisons were depicted through bar charts or tables, showing crime rates by borough and ranking them based on total crime rates. Interactive dashboards were basic, with simple filters and dropdown menus to select crime types and time periods.

**Annotated Screenshot of the Final Implementation**

The final implementation features a significantly enhanced design compared to the prototype. The header has been upgraded to include a logo, a search functionality, and a multi-level navigation system for easier access to various sections. The crime trend charts have evolved into interactive line charts that include zoom and hover functionalities, trendlines, and customizable time ranges, providing a more detailed and user-friendly view of crime trends.

The geographical distribution visualization has been improved with a detailed choropleth map that offers interactive features such as zooming, filtering by crime type, and pop-up tooltips with crime statistics. This change allows for a more nuanced spatial analysis of crime data. Borough comparisons have been refined with dynamic bar charts that include sorting and comparison features, complemented by heatmaps and ranking charts. These enhancements make it easier to compare crime rates across different boroughs and identify trends more effectively.

The interactive dashboards have been transformed into comprehensive tools with advanced filtering options and drill-down capabilities. The dashboards now feature user-friendly sliders, multi-select options, and interactive views that allow for a deeper exploration of the data. These improvements enhance the user experience by making it easier to perform detailed analyses and understand complex crime patterns.

**Comparison and Rationale**

The transition from the original prototype to the final implementation involved substantial upgrades in functionality and usability. The header’s enhancement with a logo and search functionality improves navigation and accessibility. The crime trend charts were made interactive to provide users with a more engaging and detailed exploration of trends. The geographical distribution visualization was upgraded to include interactive elements, facilitating a more comprehensive spatial analysis of crime data. The borough comparison charts were refined to offer dynamic comparisons and additional visualization types, making it easier to analyze crime rates across different boroughs.

The interactive dashboards underwent significant improvements to offer advanced filtering and exploration capabilities, enhancing the overall user experience. These changes reflect the need for more sophisticated tools to support in-depth analysis and effective decision-making in crime prevention and resource allocation.

Overall, the final design provides a richer, more interactive experience, allowing users to gain deeper insights into crime patterns and trends, thereby supporting more informed decisions for crime reduction and public safety strategies.

**Implementation**

To implement the crime data visualization in Tableau, I followed these steps:

**1. Data Import and Preparation:**

I imported the crime statistics and borough boundary datasets from the provided sources into Tableau. Initially, I cleaned the data in Microsoft Excel to ensure accuracy. This involved removing null values and correcting typos in the columns. After cleaning, I merged the datasets using the borough name as the key identifier, creating a unified dataset that combined crime data with geographical information.

**2. Data Transformation:**

In Tableau, I pivoted the date columns from the crime dataset to convert the wide format (where each month was a separate column) into a long format. This transformation created a single "Date" column and a corresponding "Crime Count" column, which streamlined the analysis of crime trends over time.

**3. Parameters and Calculated Fields:**

I set up parameters for start and end dates, allowing users to select a time range for the analysis. Using these parameters, I created a calculated field to sum the crime occurrences within the selected date range. This dynamic calculation enabled users to query and view crime data for any specified period.

**4. Visualizations:**

* **Geographical Map:** I created a geographical map of London to visualize crime distribution. The map used color coding to represent crime intensity across boroughs based on the calculated crime sum, helping users identify high-crime areas.
* **Line Chart:** I plotted a line chart to display crime trends over time. This chart used the calculated fields to show how crime rates changed month-by-month within the selected date range, making it easy to observe trends and fluctuations.
* **Pie Chart:** A pie chart was used to illustrate the proportion of different crime types. This visualization showed the relative share of various crime categories within the dataset, providing insight into the distribution of crime types.
* **Bar Graph:** I created a bar graph to compare crime rates by borough. This graph displayed the total crime rate for each borough, facilitating comparisons and highlighting which areas experienced higher or lower crime rates.

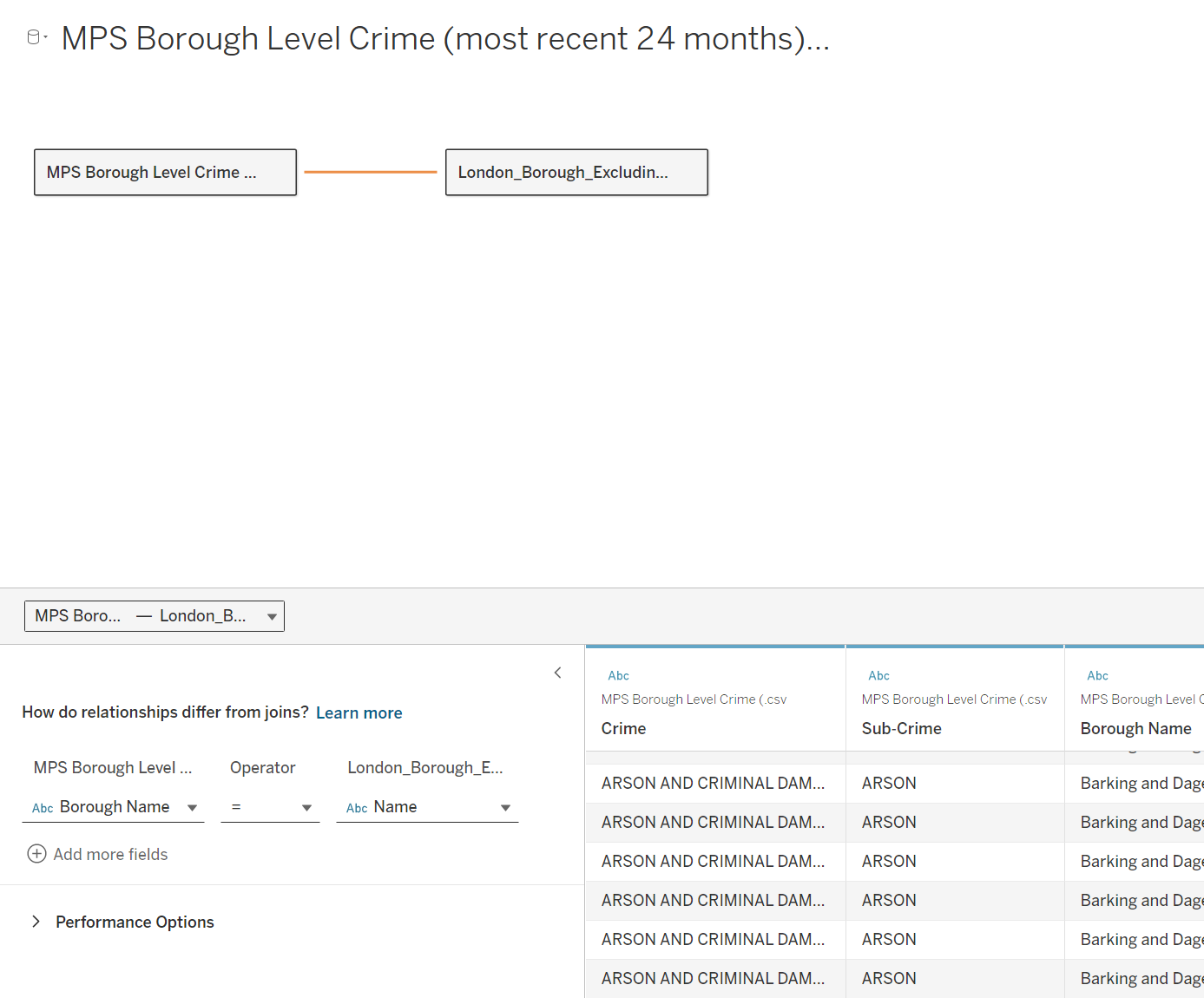
**5. Interactivity:**

To enhance user experience, I incorporated interactive elements such as filters, tooltips, and dropdown menus. These features allowed users to select different crime types, boroughs, and time periods, providing a more flexible and detailed exploration of the data.

In this walkthrough, I will detail the process followed to analyze and interpret the crime data for London, emphasizing key findings through various Tableau visualizations, including a geographical plot that highlights crime distribution across boroughs.

**1. Data Preparation and Integration:**

The first step involved importing crime statistics and borough boundary datasets into Tableau. After cleaning the data in Microsoft Excel—removing null values and correcting errors—I merged the datasets using the "Borough Name" field. This created a unified dataset that included both crime data and geographical boundaries. I then pivoted the date columns into a long format with "Date" and "Crime Count" columns to facilitate time-series analysis.

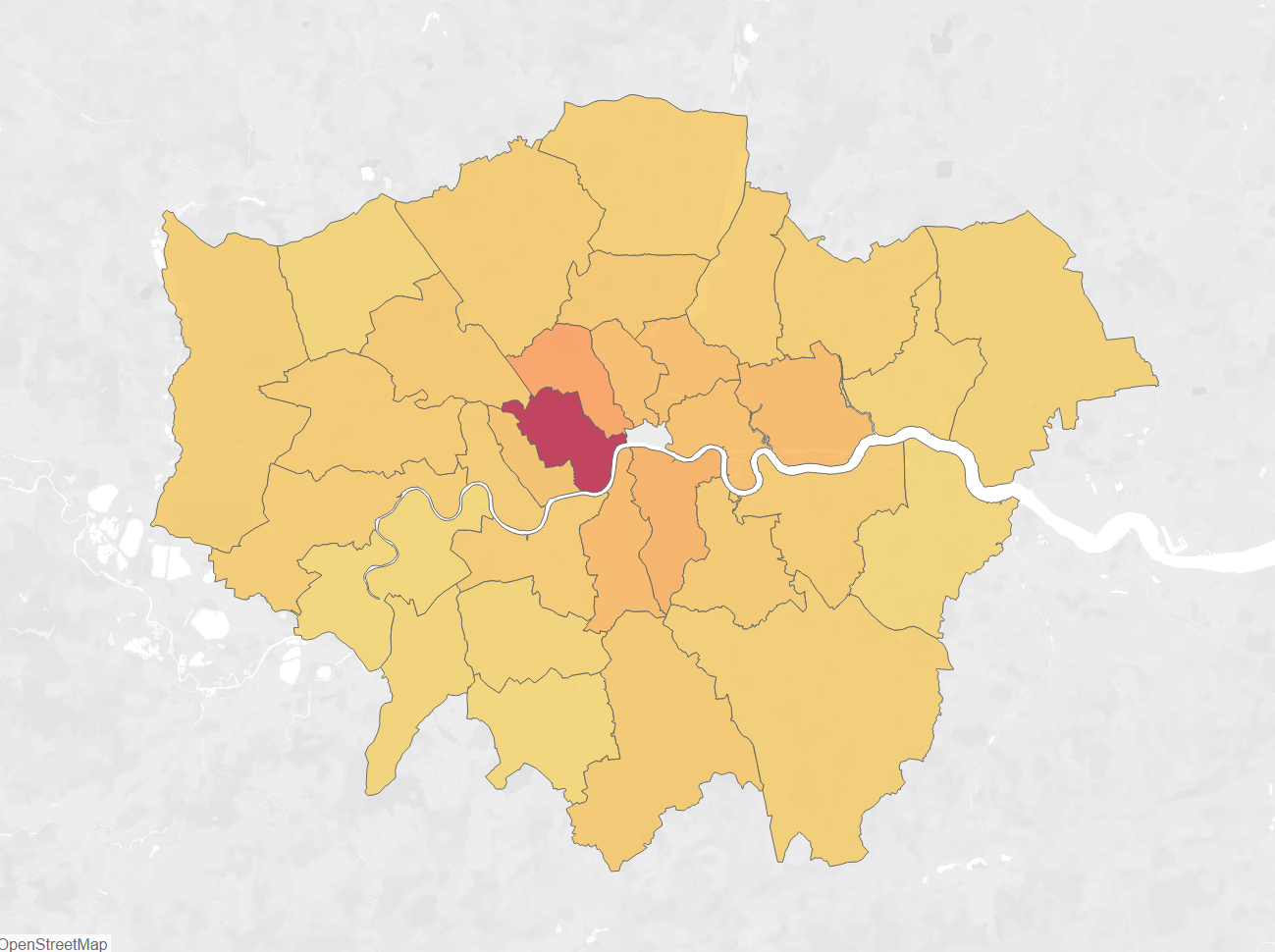


**2. Visualizations and Analysis:**

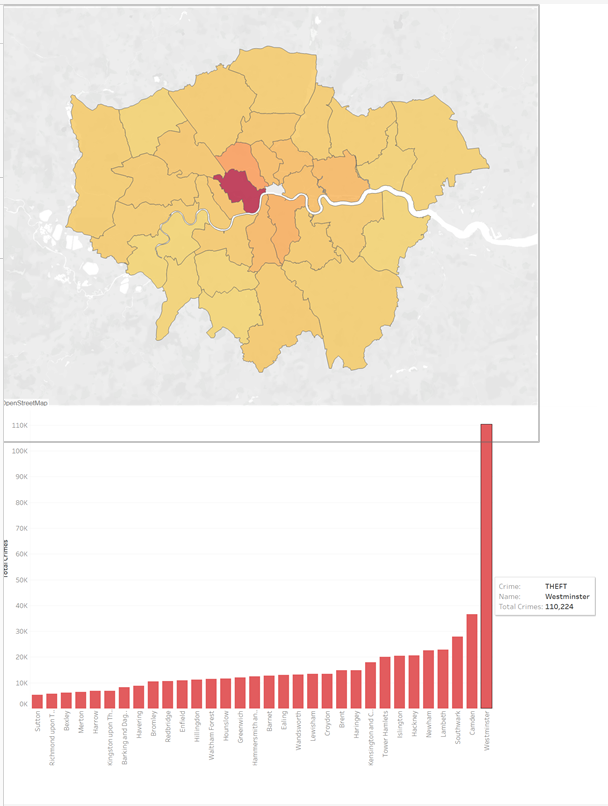
**Geographical Plot of London Boroughs:**

The geographical plot of London boroughs provides a visual representation of crime distribution across the city. This map uses color coding to depict crime intensity, with varying shades representing different levels of crime occurrences.

* **Westminster:** The map clearly shows that Westminster has the highest concentration of crime, particularly theft. This high crime rate in Westminster can be attributed to its bustling commercial and tourist areas, which present more opportunities for theft and other crimes. The dark shading on the map highlights this borough as a significant hotspot for criminal activity.



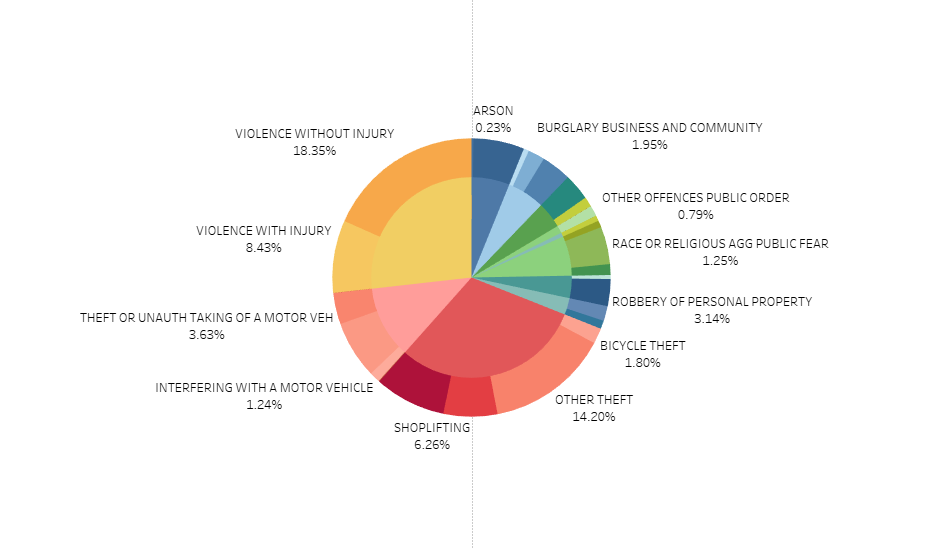
* **Geographical Disparities:** Other boroughs display lower crime rates, with varying shades indicating less intense crime levels. This geographic disparity in crime distribution is crucial for understanding where to focus law enforcement efforts and resources. The map allows for quick identification of high-crime areas and supports strategic planning for crime prevention.



**Pie Chart Analysis:**

The pie chart breaks down the proportion of different types of crimes across London:

* **Theft and Violent Crimes:** These are the most prevalent crime types, dominating the chart. This indicates that theft and violent crimes are significant concerns citywide.
* **Theft Concentration:** Westminster’s high rate of theft is evident, reinforcing the need for targeted strategies to address theft in high-risk areas.

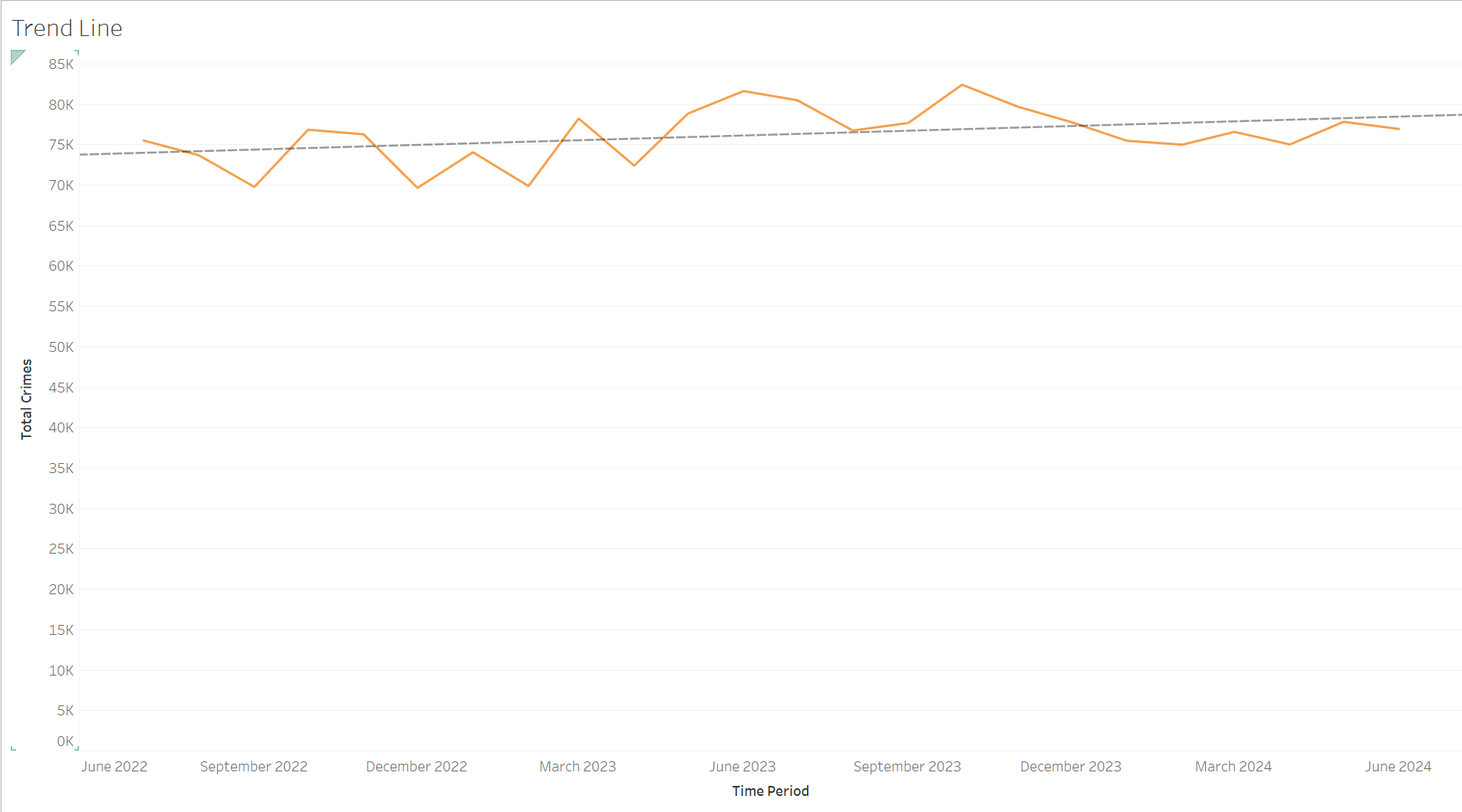


**Trend Analysis:**

**Line Chart Trends:**

The line chart visualizes crime trends over time:

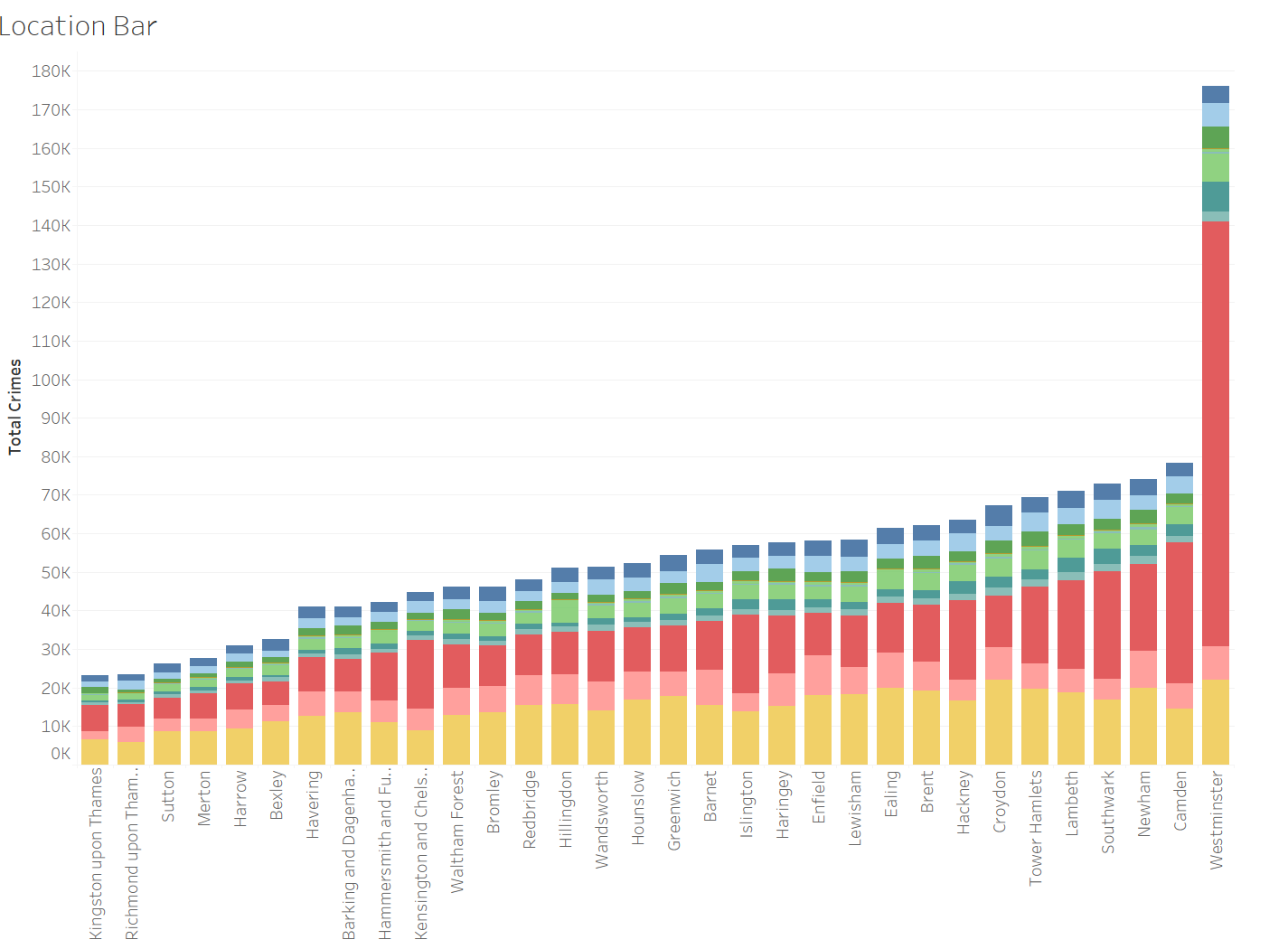
* **Upward Trend:** The upward-sloping trend line suggests a persistent or increasing crime rate overall, highlighting the need for more effective crime reduction strategies.
* **Declining Crime Types:** Specific crimes such as burglary, possession of weapons, and sexual assaults show a downward trend, indicating successful interventions in these areas.



**Bar Graph of Borough Crime Rates:**

The bar graph compares crime rates across different boroughs:

* **High Crime Rates in Westminster:** The bar graph confirms that Westminster has higher crime rates compared to other boroughs, particularly for theft. This visualization supports the findings from the geographical plot, providing a clear comparison of crime rates by borough.
* **Regional Variations:** The graph highlights differences in crime rates, with some boroughs showing reduced crime rates, suggesting effective local crime prevention measures.



**Summary of Findings:**

1. **Prevalence of Theft and Violent Crimes:** The pie chart and geographical plot confirm that theft and violent crimes are prevalent across London, with theft being particularly high in Westminster.
2. **Regional Differences:** The geographical plot shows that Westminster is a major crime hotspot, while other boroughs experience lower crime rates. The bar graph reinforces this by displaying Westminster's high crime rates compared to other areas.
3. **Trends in Crime Rates:** The line chart reveals an upward trend in overall crime rates, despite successful reductions in specific crime categories like burglary and weapon possession.

These insights allow law enforcement and policymakers to focus resources and strategies on high-crime areas, such as Westminster, and apply successful interventions from areas with reduced crime rates to other regions.

**Discussion and Critical Evaluation**

The project showcased Tableau’s strengths in analyzing and visualizing crime data. The interactive dashboards effectively engaged users, allowing dynamic exploration through features like filters, tooltips, and drill-downs. This interactivity was crucial for examining crime trends and geographical distributions. The geographical plot highlighted Westminster as a major crime hotspot, providing clear visual evidence that directed attention to areas with the highest crime rates. Additionally, the line charts effectively demonstrated trends over time, showing a concerning upward trajectory in overall crime rates despite declines in specific categories such as burglary and weapon possession.

However, the project faced some limitations. The data was aggregated monthly, which could obscure shorter-term trends and fluctuations. More granular data could offer deeper insights into specific crime patterns. Additionally, the geographical analysis was limited to borough-level data, which may not capture variations within boroughs. Including more detailed spatial data could improve the understanding of crime distribution. The visualizations also lacked additional contextual information, such as socio-economic factors, which could enhance the interpretation of crime patterns.

On a personal level, this project provided valuable experience in integrating and transforming datasets, ensuring data accuracy, and creating effective visualizations. It enhanced my ability to interpret complex data and develop actionable insights. Looking forward, I plan to explore advanced analytical techniques such as predictive modeling and machine learning to further deepen data analysis. Additionally, incorporating more detailed and diverse data sources will be important for future projects.

Tableau’s user-friendly interface and powerful interactive capabilities were significant assets, though the high cost of licensing and the learning curve for advanced features pose challenges. Overall, the project demonstrated how Tableau can be used effectively for crime data analysis, providing meaningful insights while also highlighting areas for improvement and further exploration.

**Conclusion**

The objective of this project was to analyze and visualize crime data for London from July 2022 to June 2024, focusing on identifying trends, understanding crime distribution, and pinpointing high-crime areas. The approach involved integrating and cleaning datasets from government sources, transforming the data for analysis, and creating interactive visualizations in Tableau. The solution included geographical maps, line charts, pie charts, and bar graphs to effectively convey crime trends and patterns.

Key outcomes revealed that Westminster had the highest concentration of theft, highlighting it as a major crime hotspot. The overall crime rate showed an upward trend, signaling the need for more effective crime reduction strategies. Despite reductions in specific crime categories like burglary and weapon possession, the persistent increase in overall crime emphasized the importance of continued and targeted interventions. The project successfully utilized Tableau’s capabilities to provide clear, actionable insights for crime prevention and resource allocation.

**References**

1. **Recorded Crime Summary Dataset:** Greater London Authority. (n.d.). *Recorded Crime Summary*. Retrieved from https://data.london.gov.uk/dataset/recorded\_crime\_summary
2. **Statistical GIS Boundary Data:** Ordnance Survey. (n.d.). *Statistical GIS Boundary Files for London*. Retrieved from https://www.data.gov.uk/dataset/statistical-gis-boundary-files-for-london
3. **Tableau Software:** Tableau Software. (n.d.). *Tableau - Business Intelligence and Data Visualization Software*. Retrieved from https://www.tableau.com
4. **Data Cleaning and Preparation Techniques:** Microsoft. (n.d.). *Excel - Data Cleaning and Preparation*. Retrieved from https://support.microsoft.com/en-us/excel
5. **Visualization and Analysis Best Practices:** Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press.
6. **Understanding Crime Trends and Patterns:** Crime and Justice Research Institute. (2022). *Analyzing Crime Trends and Patterns: A Guide*. Retrieved from https://www.crimejustice.org/resources