

**TASK**

**Exploratory Data Analysis on the Crime Stats Dataset**

[](http://www.hyperiondev.com/portal/)

**Introduction**

This report presents an in-depth analysis of the Crime Stats Dataset, shedding light on the intricate dynamics of crime trends and distributions in South Africa. The dataset encompasses crime statistics from the financial year 2011/2012 to 2022/2023, as released by the South African Police Services. This comprehensive dataset comprises a variety of crime incidents, offering insights into the landscape of criminal activities across the nation.

**Summary of the Data Set**

The Crime Stats Dataset provides a comprehensive view of crime incidents in South Africa, spanning over a decade of records. The dataset encompasses the following categories:

Crime Incidents by Category: This section encapsulates a broad spectrum of crime incidents, providing a comprehensive overview of criminal activities in South Africa.

Contact Crime Incidents by Category: Focusing on crimes involving direct contact between perpetrators and victims, this category provides insights into physical confrontations.

Sexual Offence Incidents by Category: These incidents encompass various forms of sexual offenses, offering insights into this critical aspect of criminal behaviour.

Aggravated Robbery Incidents by Category: This category delves into cases of aggravated robberies, providing insights into incidents involving theft and violence.

Contact Related Crime Incidents by Category: Highlighting crimes associated with contact-based scenarios, this category offers insights into interpersonal confrontations.

Property Related Crime Incidents by Category: This section focuses on crimes revolving around property, providing insights into thefts, burglaries, and other property-related offenses.

Other Serious Crime Incidents by Category: Covering a range of serious criminal activities beyond the previously mentioned categories, this section offers insights into various complex crime scenarios.

Crimes Detected as a Result of Police Action by Category: Providing valuable insights into the efficacy of police actions, this category highlights crimes that have been detected and acted upon by law enforcement agencies.

**Summary from the Source**

The dataset's source, the South African Police Services, emphasizes its significance as a comprehensive compilation of crime statistics for the given period. These statistics enable researchers, policymakers, and stakeholders to gain a deep understanding of crime trends, facilitating informed decisions and strategies to address security challenges.

In the subsequent sections of this report, we delve into the exploratory data analysis performed on the dataset. This includes data cleaning procedures, data stories, and visualizations that offer valuable insights into crime trends, geographical distributions, category comparisons, and more.

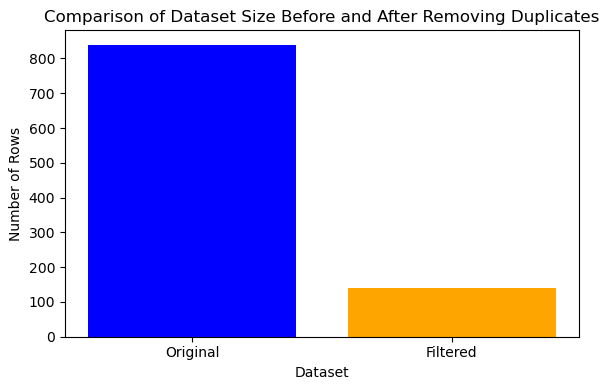
**DATA CLEANING**

**1. Cleaning Financial Year:**

You filtered the dataset to include only the financial years 2021/2022 and 2022/2023 by applying fuzzy matching to the "Financial Year" column. This ensured that only rows with these specific financial years were retained in the dataset.

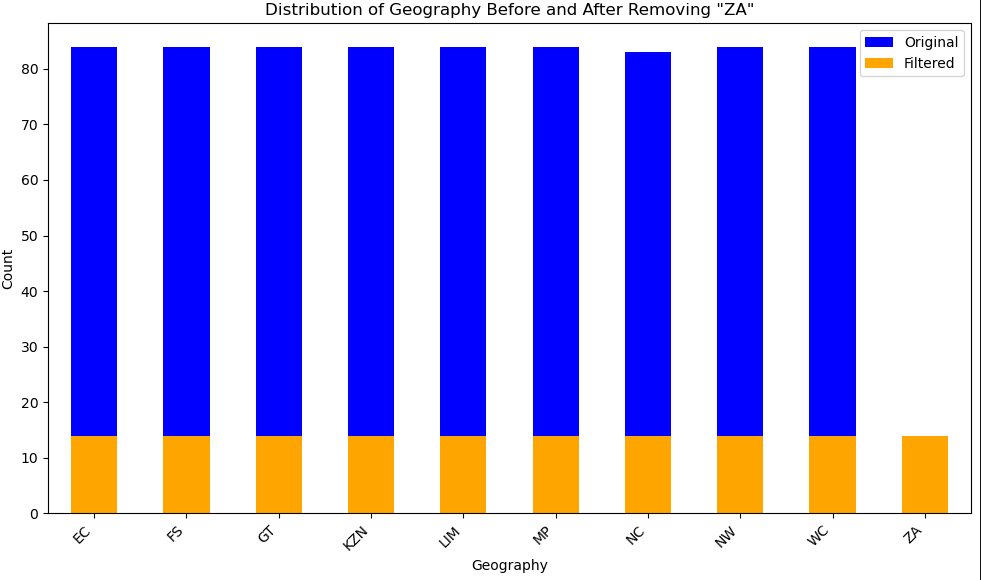
**2. Removing Duplicates:**

You identified and removed any duplicate rows from the dataset, ensuring that each unique combination of "Geography," "Crime Category," "Financial Year," and "Count" appeared only once.



**3. Removing 'ZA' from Geography:**

You removed rows with 'ZA' from the "Geography" column, effectively excluding the data associated with the entire country of South Africa.



**MISSING DATA**

Based on the dataset you provided, it appears that there are no explicit missing values (NaN or None) in the columns you mentioned: "Geography," "Crime Category," "Financial Year," and "Count." However, when we were cleaning the dataset earlier, there were some rows with 'None' in the "Financial Year" column. These rows were filtered out, so the current dataset seems to be cleaned and does not have any missing values.

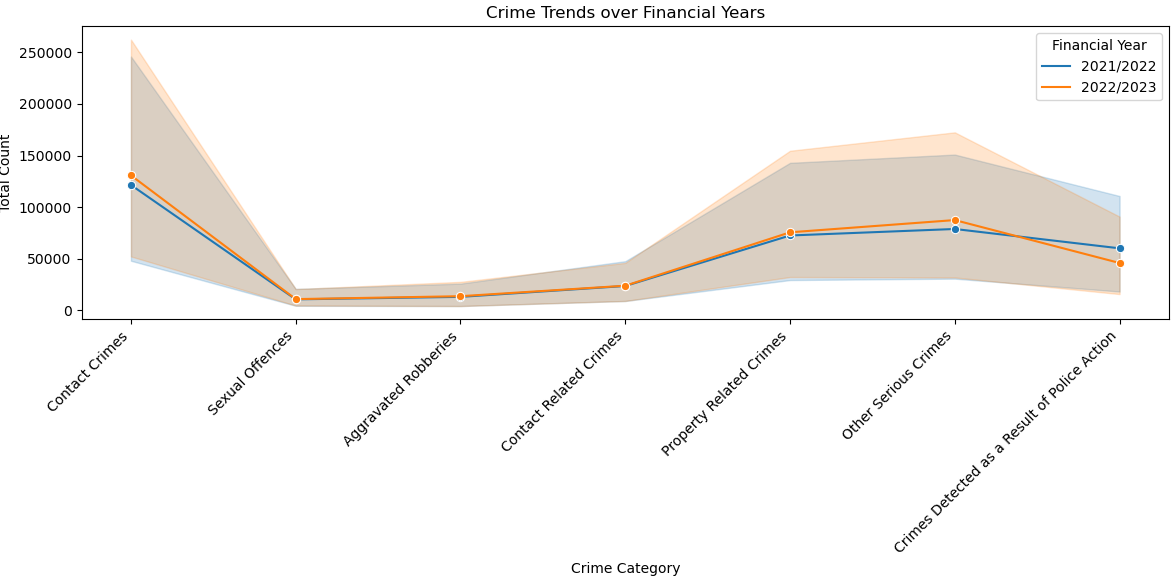
**DATA STORIES AND VISUALISATIONS**

**DATA STORIES AND VISUALISATIONS**

This section presents a comprehensive exploration of the Crime Stats Dataset through a series of data stories and visualizations. These data stories unveil meaningful insights and assumptions based on the visualizations of the dataset.

**Data Story 1: Crime Trends over Financial Years**

Understanding crime trends across financial years is crucial for identifying evolving criminal dynamics. We delve into the variations of different crime categories between the financial years 2021/2022 and 2022/2023.



The filtered dataset focuses exclusively on the target financial years, i.e., 2021/2022 and 2022/2023. A line plot, constructed using Seaborn, illustrates the trends in crime categories across these years. Each crime category is color-coded, aiding comparison.

Observations:

Contact Crimes: A slight increase is observed from 2021/2022 to 2022/2023.

Sexual Offences: Counts remain relatively consistent across both years.

Aggravated Robberies: This category mirrors the trend of sexual offences.

Contact Related Crimes: Similar to contact crimes, a marginal increase is seen.

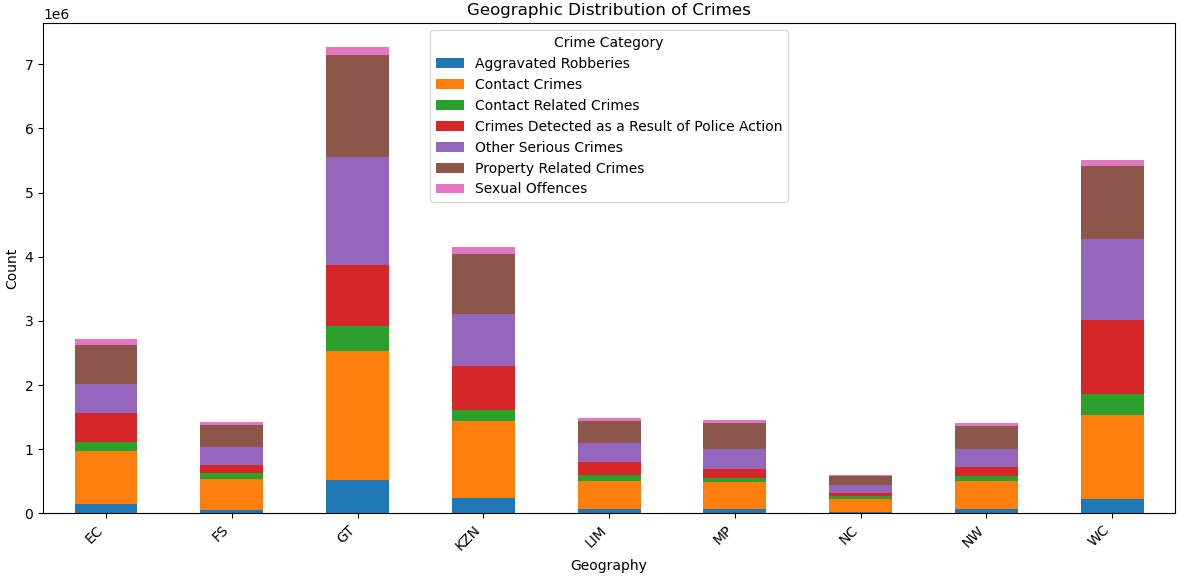
Property Related Crimes: A slight increase in 2022/2023 is noted.

Other Serious Crimes: There's a notable increase compared to previous years.

Crimes Detected as a Result of Police Action: A decline is evident in 2022/2023.

**Data Story 2: Geographic Distribution of Crimes**

Unearthing crime distribution across geographical regions aids in pinpointing areas with higher crime rates. The analysis and visualization below elucidate this aspect.



The dataset is grouped by 'Geography' and 'Crime Category,' and total counts are aggregated. 'ZA' is excluded from the 'Geography' column. The resulting data is pivot-tabled, producing a matrix suitable for a stacked bar plot. The plot showcases the distribution of crime categories across diverse geographical areas.

Notable Insights:

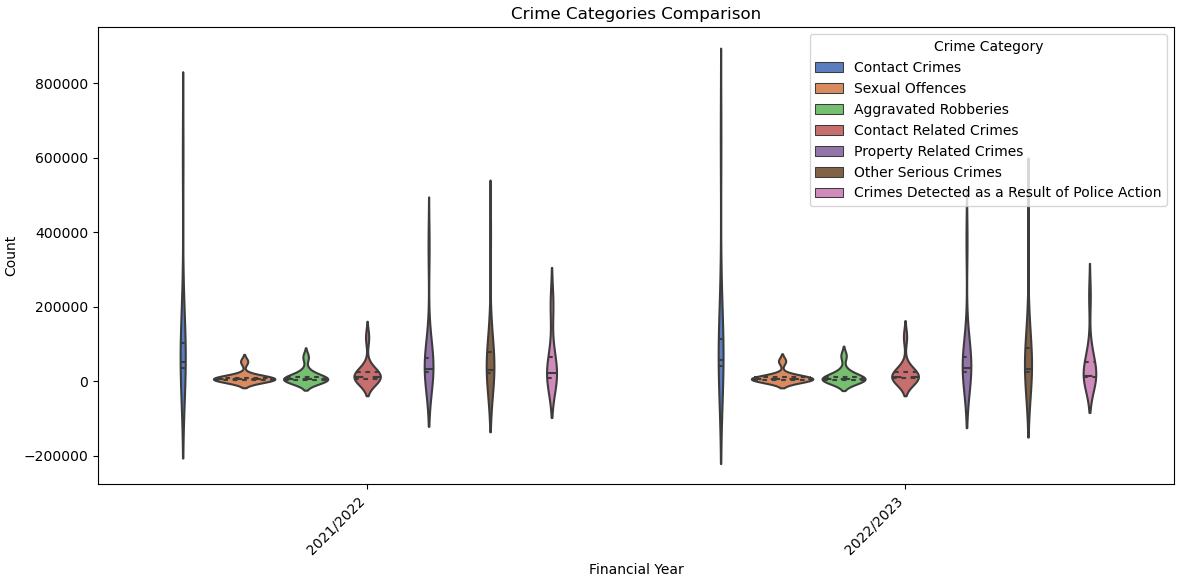
Gauteng (GT) and Western Cape (WC) consistently exhibit higher crime counts across multiple categories.

Western Cape (WC) stands out with significant counts in various categories, highlighting its crime distribution.

The visualization underscores the distribution pattern of crimes, enabling targeted interventions and policies.

**Data Story 3: Crime Categories Comparison**

Comparing crime category distributions within each financial year is pivotal in identifying any shifts or trends.



The dataset is filtered to encompass the target financial years. A violin plot, generated using Seaborn, effectively compares the distributions of different crime categories. The y-axis represents crime category counts for both 2021/2022 and 2022/2023, while each category is color-coded for easy differentiation.

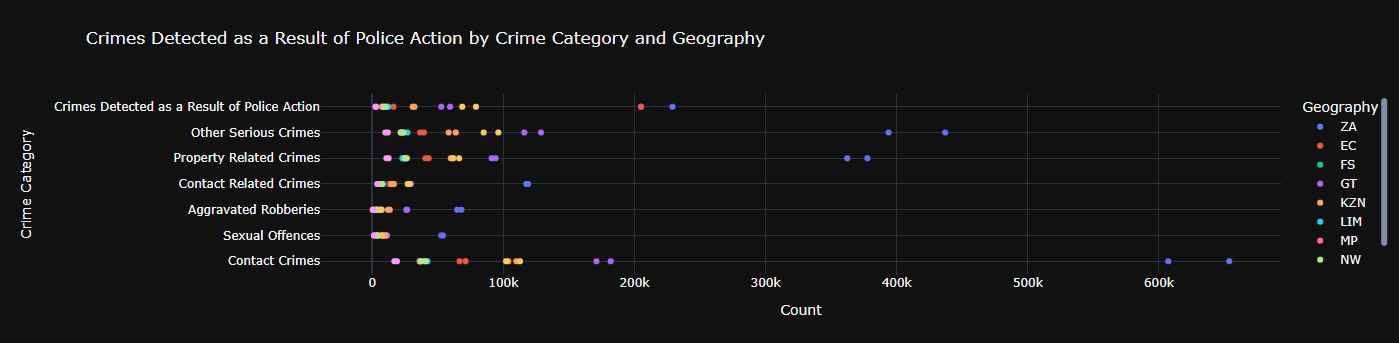
Key Observations:

The violin plot provides insights into how crime categories evolved over the financial years.

Visual comparisons of distribution reveal trends or variations in criminal activities.

**Data Story 4: Crimes Detected as a Result of Police Action**

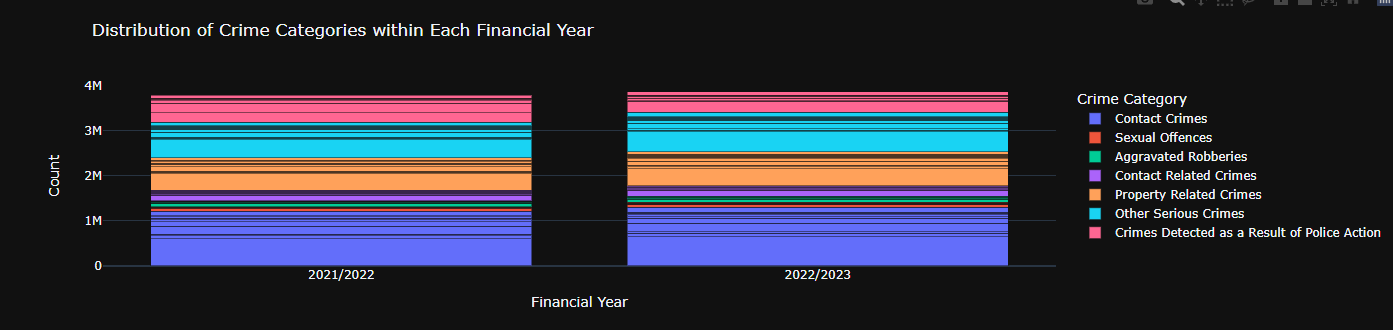
Visualizing the connection between crimes detected due to police action and distinct crime categories across regions yields informative insights.



Plotly Express is utilized to craft an interactive scatter plot that delineates the relationship between police-detected crimes, crime categories, and geographic regions. Each point denotes a specific crime category's count, with colours signifying different regions. The interactive nature allows detailed information to be obtained through point hovering.

**Data Story 5: Crime Categories Stacked Bar Plot**

The distribution of crime categories within each financial year is concisely depicted through a stacked bar plot.



Plotly Express constructs the stacked bar plot, showcasing each financial year's distribution. Colours represent different crime categories, and the bars' heights correspond to total crime counts. The upper y-axis range is adjusted to enhance visualization.

These data stories, backed by visualizations, provide a holistic understanding of crime trends, geographic distributions, category comparisons, and more. This comprehensive exploration empowers data-driven decision-making and effective strategies to address crime-related challenges.

**THIS REPORT WAS WRITTEN BY : Irshaad Alladin**

