Table of Contents

Executive Summary

Introduction

The comprehensive dashboards created from UK Northeast Cities' crime data provide a detailed overview of the regional crime trends and hotspots. They offer critical insights into the types and distribution of criminal activities from January to October 2023. These visual analytics serve as a vital tool for informing law enforcement strategies and enhancing public safety initiatives.

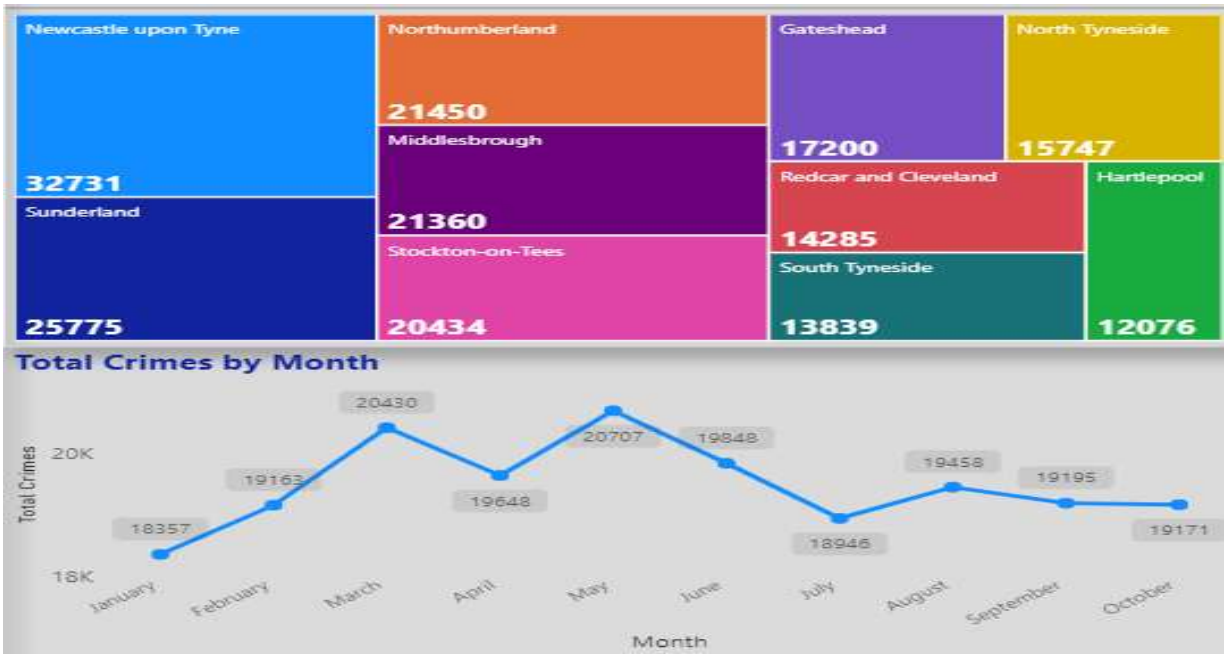

Figure 1 Monthly Total Crime for Cities

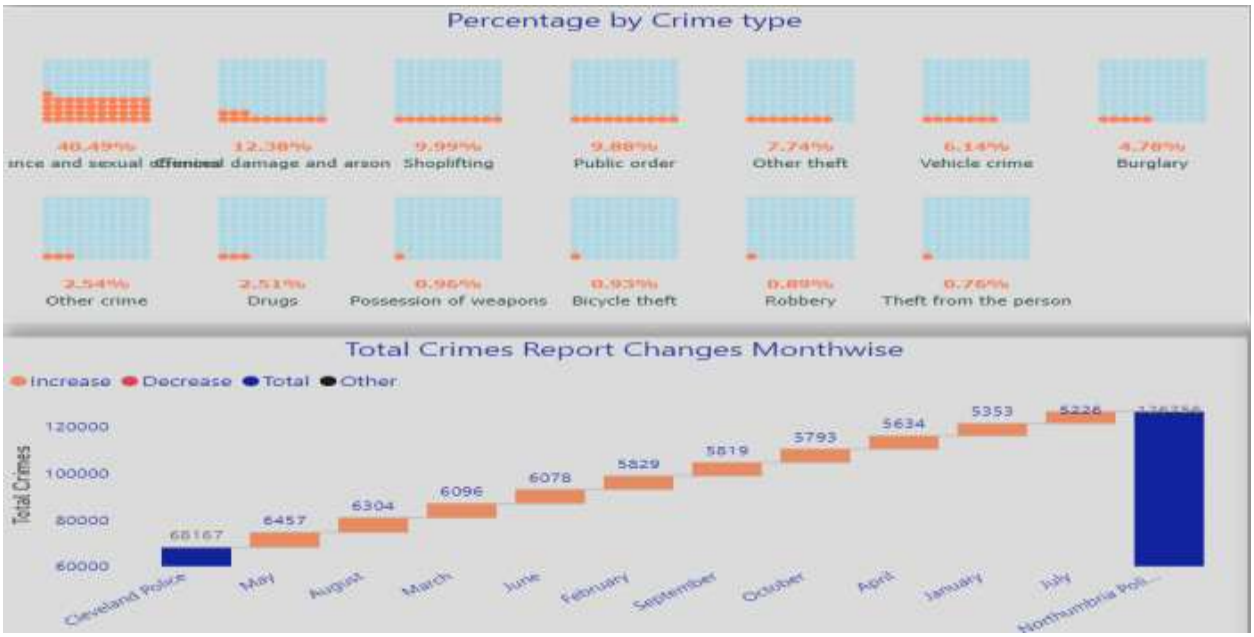

Figure 2 Crime Type Percentage

Area	Location	Sum of Crime Count	Last outcome category
Newcastle upon Tyne	On or near	1482	Investigation complete; no suspect identified
Newcastle upon Tyne	On or near Parking Area	805	Investigation complete; no suspect identified
Newcastle upon Tyne	On or near	774	Status update unavailable
Newcastle upon Tyne	On or near Shopping Area	711	Investigation complete; no suspect identified
Northumberland	On or near	628	Investigation complete; no suspect identified
Sunderland	On or near	624	Investigation complete; no suspect identified
Northumberland	On or near	611	Status update unavailable
Stockton-on-Tees	On or near	555	Investigation complete; no suspect identified
Northumberland	On or near Parking Area	503	Investigation complete; no suspect identified
Gateshead	On or near	472	Investigation complete; no suspect identified
Middlesbrough	On or near	447	Unable to prosecute suspect
Northumberland	On or near Parking Area	444	Under investigation
Newcastle upon Tyne	On or near Parking Area	416	Under investigation
Middlesbrough	On or near	382	Investigation complete; no suspect identified
Total		**194923**	

Figure 3 Total Crime Count by Area and Outcome

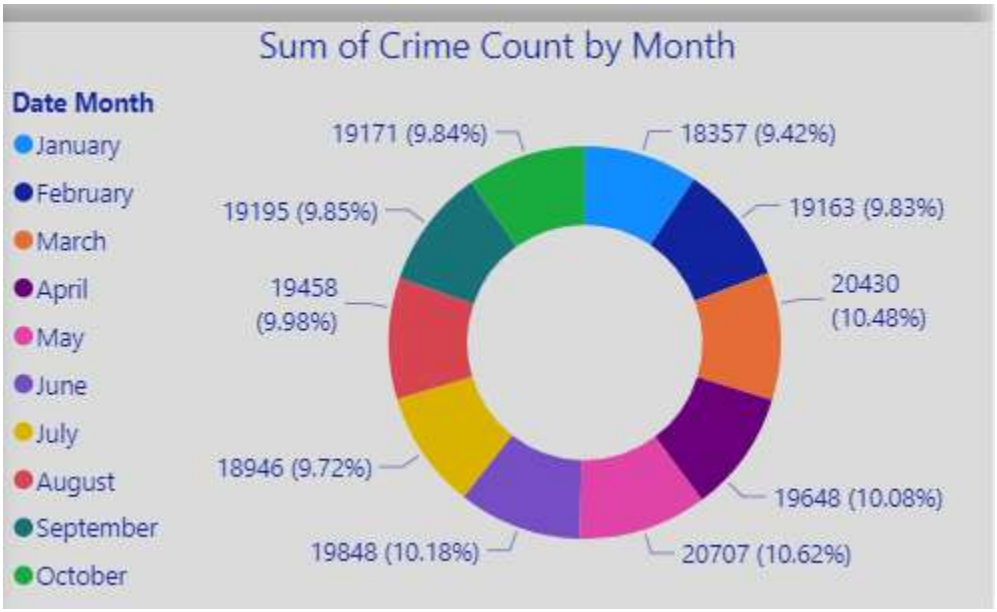

Figure 4 Monthly Total Crime

Key Findings

Here are the key findings from the analysis of the UK Northeast Cities crime data dashboards:

High Incidence of Violence and Sexual Offenses: The most common crimes reported were related to violence and sexual offenses.

Concentration of Crimes in Urban Areas: Crime incidents are densely clustered within urban centers, indicating specific hotspots that require focused attention.

Variation in Crime Over Time: There appears to be a temporal pattern or trend in crime rates, with certain months showing higher incidences.

Localization of Crime Types: Certain areas are more prone to specific types of crime, as evidenced by the detailed mapping on the crime spotting dashboard.

Volume of Crime Reports: Some areas have a high number of repeat incidents, which may indicate persistent problems or effective crime reporting mechanisms.

Distribution of Crime by Type: The dashboards provide a clear hierarchy of the prevalence of different crime types.

Resolution of Crimes: There is an extensive breakdown of crime outcomes, which helps to understand the effectiveness of the criminal justice response.

Predictive Indicators for Law Enforcement: The month-wise changes in crime reports could serve as predictive indicators for law enforcement to anticipate and prepare for future trends.

These findings can help to guide policy, improve public awareness, and enhance the strategic deployment of law enforcement resources.

Recommendations

Based on the key findings from the UK Northeast Cities crime data dashboards, here are some recommendations:

Targeted Intervention for Violence and Sexual Offenses: Implement specialized campaigns and increase patrols in areas with high incidences of violence and sexual offenses. Consider community programs that educate on prevention and support victims.

Focused Resources on Urban Crime Hotspots: Allocate more resources to urban centers with high crime rates. This could include increased police presence, community policing initiatives, and the installation of surveillance equipment.

Seasonal Crime Prevention Strategies: Analyze the variation in crime over time to develop and implement seasonal or monthly crime prevention strategies that anticipate and mitigate potential increases in crime rates.

Localized Crime Type Response: Tailor law enforcement training and response plans to the specific types of crimes that are prevalent in certain areas to improve the effectiveness of crime handling.

Enhanced Reporting and Analysis of Repeat Incidents: Investigate areas with high numbers of repeat incidents to understand and address the root causes. Improve crime reporting mechanisms to ensure accurate and timely data collection.

Crime Type Prioritization: Use the distribution of crime by type to prioritize which crimes to address first based on their prevalence and impact on the community.

Review and Improvement of Crime Resolution Processes: Assess the current crime resolution outcomes and processes to identify areas for improvement, ensuring that cases are resolved efficiently and justly.

Development of Predictive Policing Capabilities: Leverage the data on month-wise changes in crime reports to develop predictive policing models that can forecast and prevent future crimes, optimizing resource deployment.

By implementing these recommendations, law enforcement and community leaders can work towards reducing crime rates, improving public safety, and fostering a more secure environment for all residents in the UK Northeast Cities.

Introduction

To understand the crime dynamics within the northeast of England, particularly in areas like Newcastle, Middlesbrough, Redcar and others, I have embarked on a Business Intelligence (BI) project using data obtained from the data.Police.uk website. This project aims to delve into crime patterns in these regions, with a special focus on street crimes.

Dataset Description

The dataset spans from January to October 2023 and covers two police jurisdictions: **Cleveland and Northumbria Police**. It is comprehensive, including columns such as Crime ID, Month, Reporting Agency, Geographic Coordinates (Longitude, Latitude), Location, Lower Layer Super Output Area (LSOA) code and name, Crime Type, Last Outcome Category, and Context, and total number of **247,864** rows in data sets.

Column Name	Type	Description
Crime ID	String	Unique identifier for each crime incident
Month	Date	Month when the crime was reported
Reported by	String	Authority or department reporting the crime
Falls within	String	Jurisdiction or area of the crime
Longitude	float64	Longitude coordinates of the crime location
Latitude	float64	Latitude coordinates of the crime location
Location	String	Textual description or address of the crime location
LSOA code	String	Lower Layer Super Output Area code
LSOA name	String	Name of the Lower Layer Super Output Area
Crime type	String	Category of the crime
Last outcome category	String	Latest status or outcome of the crime
Context	float64	Numerical data linked to the crime's context

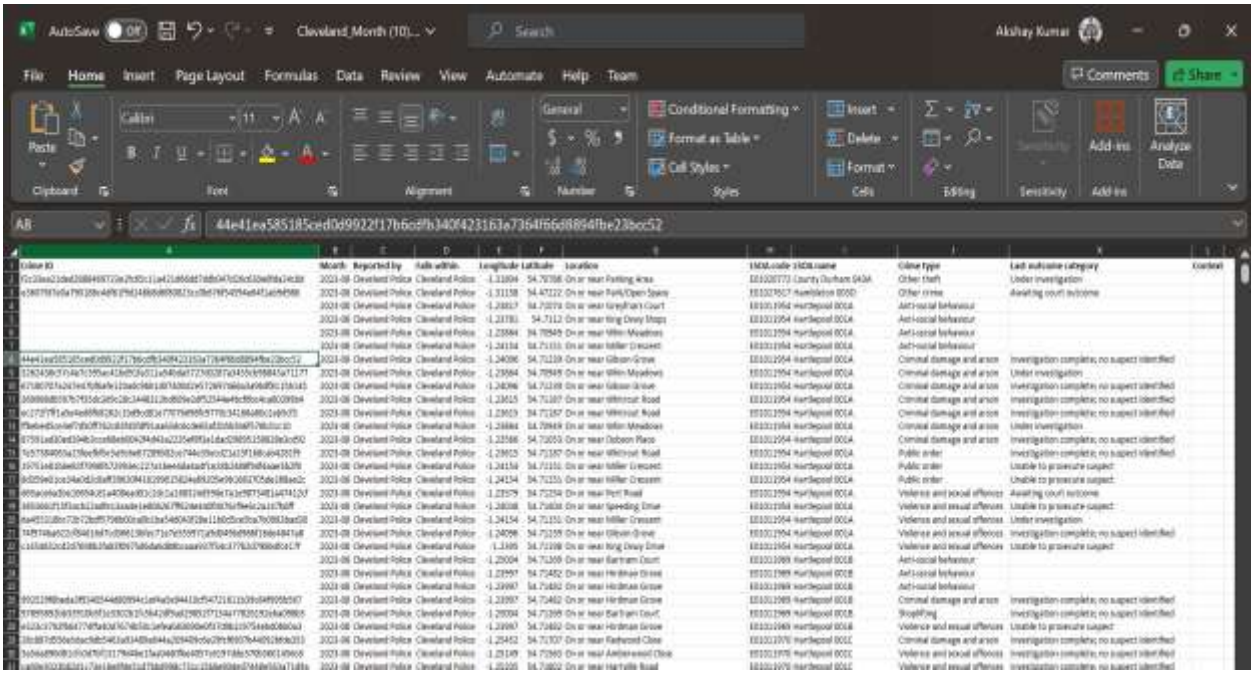

Figure 5 Raw Image to Dataset file

BI Requirements

Why Did I Choose This Dataset?

I chose this dataset primarily because of my personal interest in the safety of the area where I live. The Northeast, particularly cities like Middlesbrough and Redcar, has a reputation for street crimes. This dataset provides an opportunity to analyze crime trends and patterns in these areas.

Does My Dataset Address the Big Data Problem?

The dataset addresses the big data problem by providing a large volume of varied data over a significant time span. It allows for a detailed analysis of crime types, locations, and their changes over time.

Which specific features are you going to focus on?

I will focus on crime types, specifically street crimes, and their geographical distribution within the cities. The dataset's longitudinal aspect also allows for the examination of temporal trends in crime rates.

Will this dataset help you in developing specific business skills?

Yes, analyzing this dataset will enhance my skills in data analytics, pattern recognition, and geographical information systems (GIS). It will also improve my ability to draw actionable insights from complex data sets.

What KPI questions do you seek to answer with your BI project?

The dashboards designed to analyze the UK Northeast Cities crime data aim to answer several Key Performance Indicator (KPI) questions that can help in evaluating the performance of

law enforcement and the effectiveness of crime prevention strategies. Here are some of the critical KPI questions these dashboards address:

1. What is the trend of overall crime rates over time in the region?

This KPI tracks the increase or decrease in crime over the months, providing insight into whether crime is becoming frequent.

2. Which types of crime are most prevalent, and how do they vary by area?

By breaking down crime by type and location, this KPI can help law enforcement prioritize resources and strategies for the most affected areas.

3. Are there specific hotspots where crime is significantly higher?

Identifying areas with high crime rates can help in deploying targeted interventions and preventative measures.

4. What are the patterns of violent and sexual offenses reported?

Given their severity, understanding the patterns of these offenses can aid in developing specialized responses and support services.

5. How effective are the crime resolution and outcome efforts?

This KPI assesses the success of the criminal justice system in resolving cases and can indicate areas for improvement.

6. What is the volume of repeat incidents in particular locations?

High volumes of repeat incidents can indicate areas where crime prevention measures may be failing, necessitating a review and adjustment of strategies.

7. How do crime types and rates correlate with socio-demographic and economic factors in various neighborhoods?

Analyzing crime data alongside socio-demographic information can reveal underlying factors that contribute to crime, guiding community development efforts.

8. What is the response time and effectiveness of law enforcement interventions?

This KPI measures how quickly and effectively law enforcement responds to and deals with reported crimes.

9. What seasonal or temporal factors affect crime rates, and how can this information improve policing schedules?

Understanding how crime rates change with seasons or times of the day can help in optimizing patrol schedules and resource allocation.

10. Can emerging trends in the data be used to predict and prevent future crimes?

This KPI explores the potential of using historical data to forecast crime trends and implement preventative measures proactively.

These KPI questions, when answered, can provide actionable insights for law enforcement agencies, policymakers, and community leaders to enhance public safety and crime prevention efforts.

Data Preprocessing

Download the Data Sets

The initial step involved downloading separate files for each month from the Police.co.uk website, for both Cleveland and Northumbria Police departments.

Figure 6 Dataset website

Combine Multiple Data Sets

To streamline the analysis, these monthly files were combined into a single dataset. This was accomplished using Power BI's Combine Load to select a folder of files and combine them upon loading, other option I have gone through merge and append functionalities, but I choose to select the combine and select folder option, one more thing we can achieve it by placing the upcoming data sets file that are not published yet, once it will published we can place the file in data sets folder and we can Scheduled Refresh In Power BI Service, we can set up a schedule to refresh the dataset at specific intervals. This is useful for keeping our data up to date without manual intervention. The frequency of refresh can vary depending on your Power BI service plan.

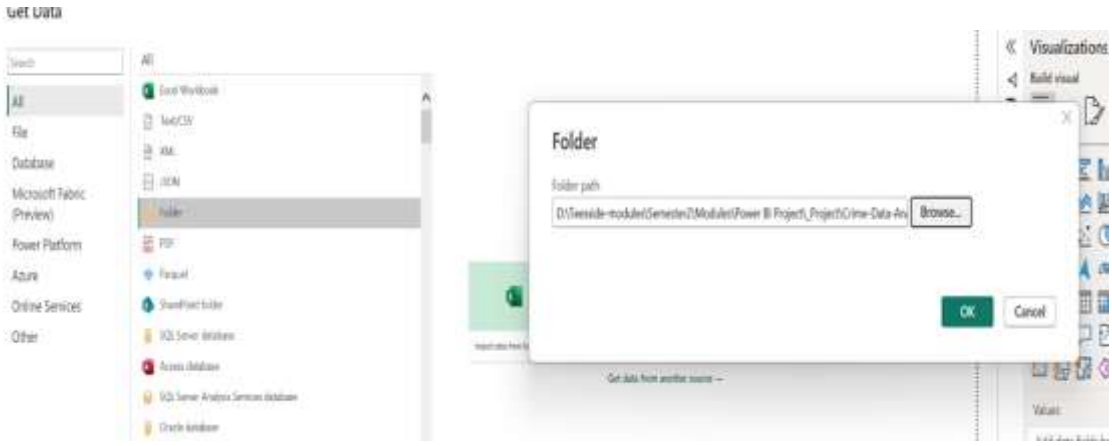

Figure 7 Load the excel data set folders.

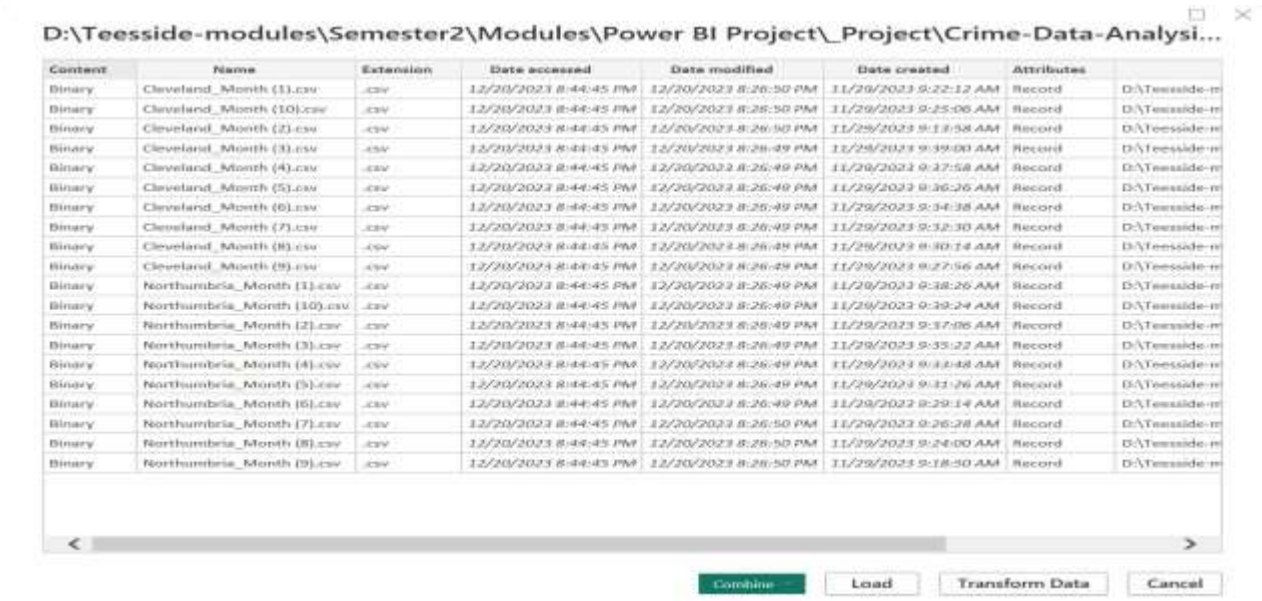

Figure 8 Load all excel file and combine it.

Create Single Table

After combining the files, a single table was created, named "Street Crime," to facilitate a unified analysis.

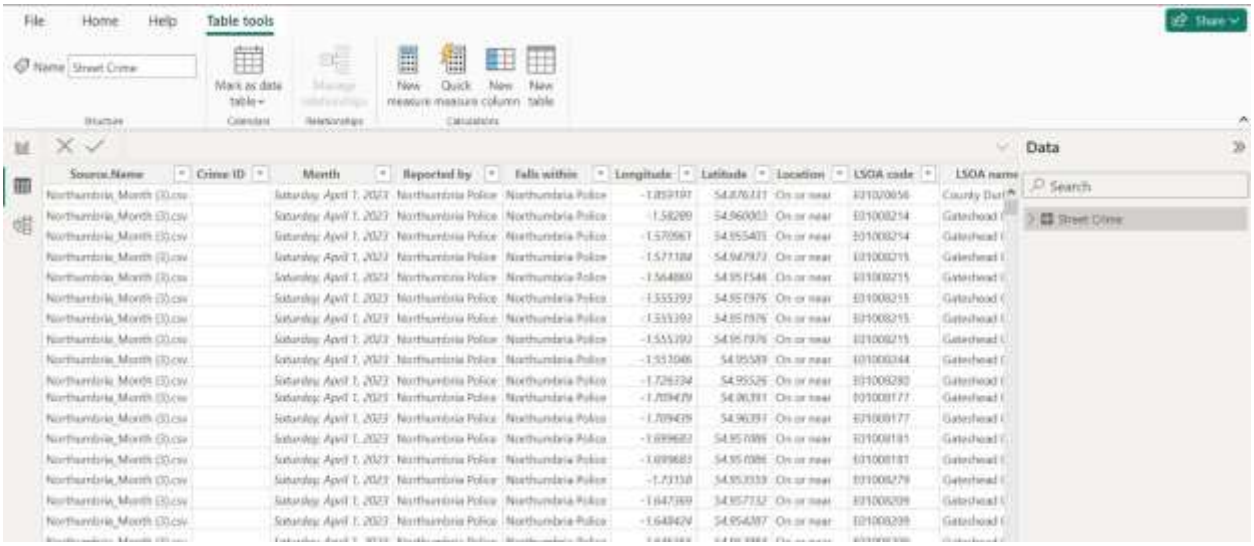

Figure 9 Create the single data set from multiple file

Removing Duplicates

Duplicates in the dataset were identified and removed. This step is crucial to ensure the accuracy and reliability of the analysis.

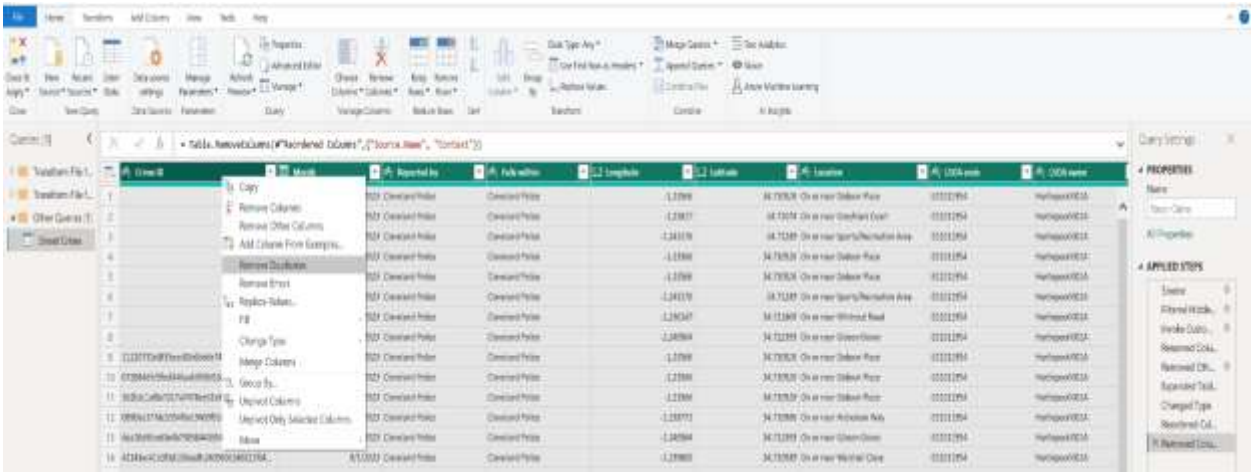

Figure 10 Remove duplicates in all columns

Removing Empty Rows

Empty rows, which can skew results and analyses, were identified and eliminated from the dataset.

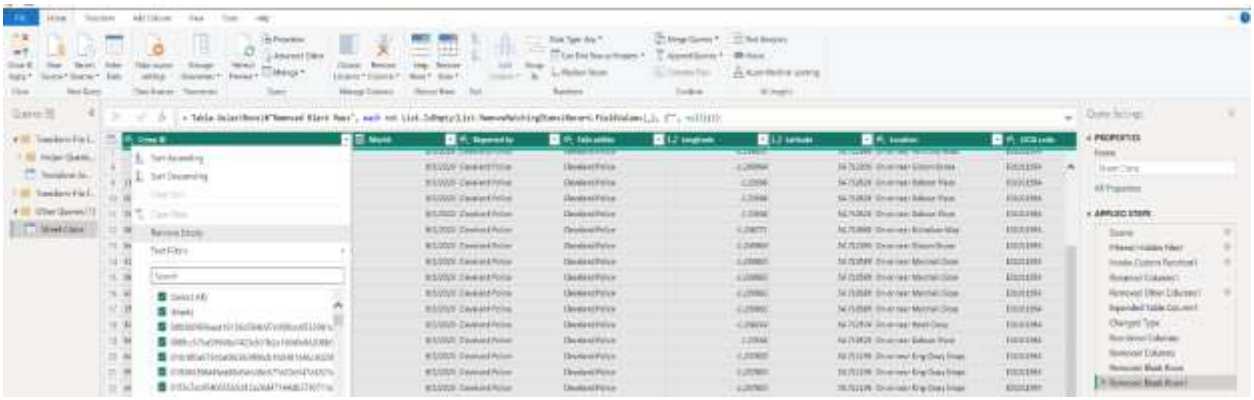

Figure 11 Remove empty rows in all columns.

Removing Blank Rows and Null Data

Further data cleaning involved the removal of blank rows and null data across all columns, enhancing the dataset's integrity.

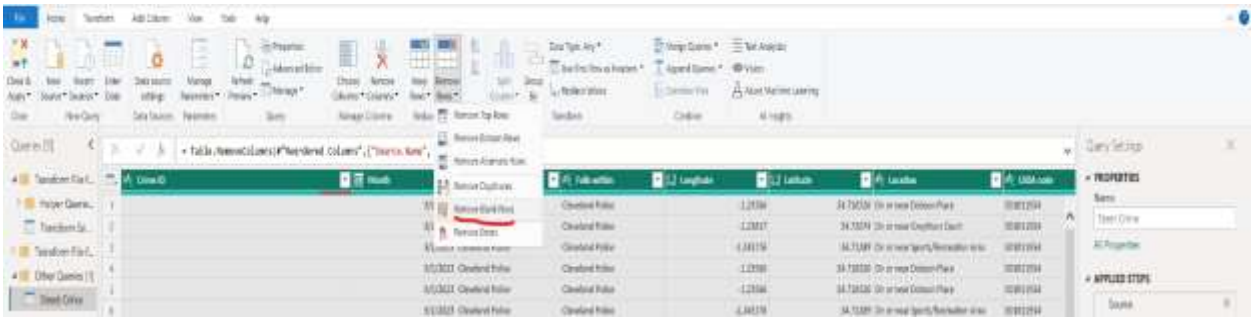

Figure 12 Remove null and blank rows.

Changing Data Types

Each column's data type was examined and adjusted as necessary to align with the nature of the data it contained.

Removing Columns

Columns that were deemed irrelevant, such as 'Source.Name' and 'Context', were removed. This step streamlined the dataset for more focused analysis.

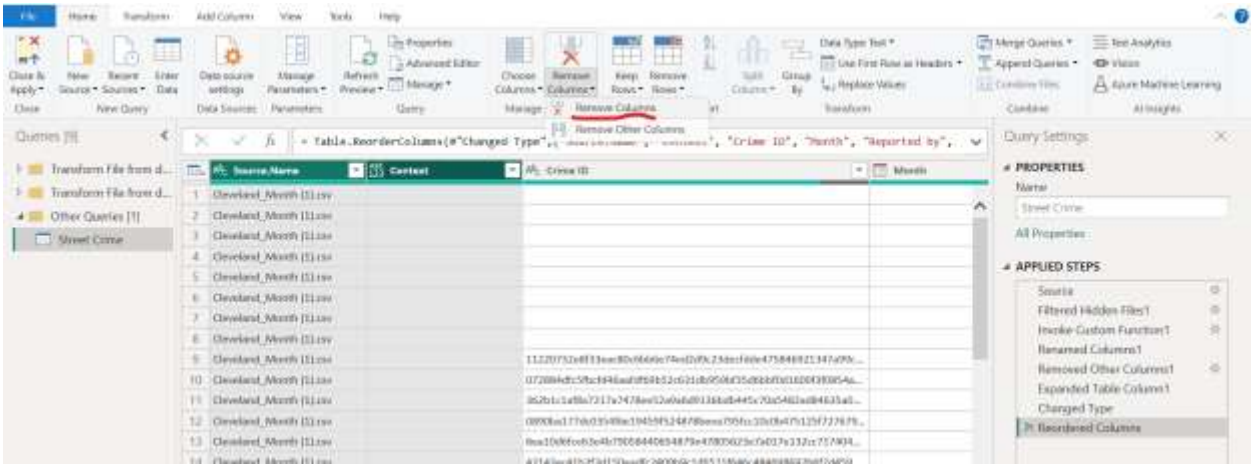

Figure 13 Remove Columns

Renaming Columns

The 'Month' column was renamed to 'Date' for better clarity and to reflect the nature of the data more accurately.

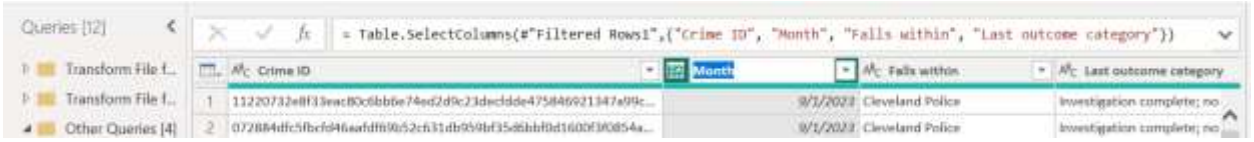

Figure 14 Rename column Month to Date

Having meticulously completed the extensive process of data processing and cleaning, we have successfully refined our dataset to a total of **194,923** rows. This comprehensive effort ensures the accuracy and reliability of the data, paving the way for more effective and insightful analysis.

Column Manipulation

Calculated Columns

Area: dataset includes a column labeled 'LSOA Name', which stands for Lower Layer Super Output Area, a geographical classification used extensively in the United Kingdom. This column contains names that are rich in geographical context, typically encompassing specific area or neighborhood names. The aim here is to distill this detailed information into a more focused 'Area' column, which will specifically represent the area names extracted from the 'LSOA Name'.

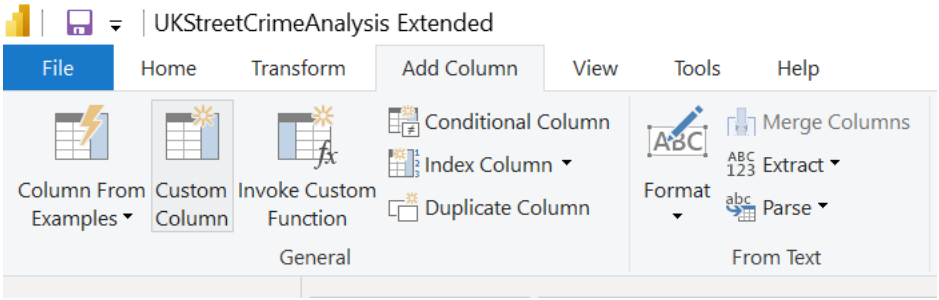

Figure 15 Create Custom column.

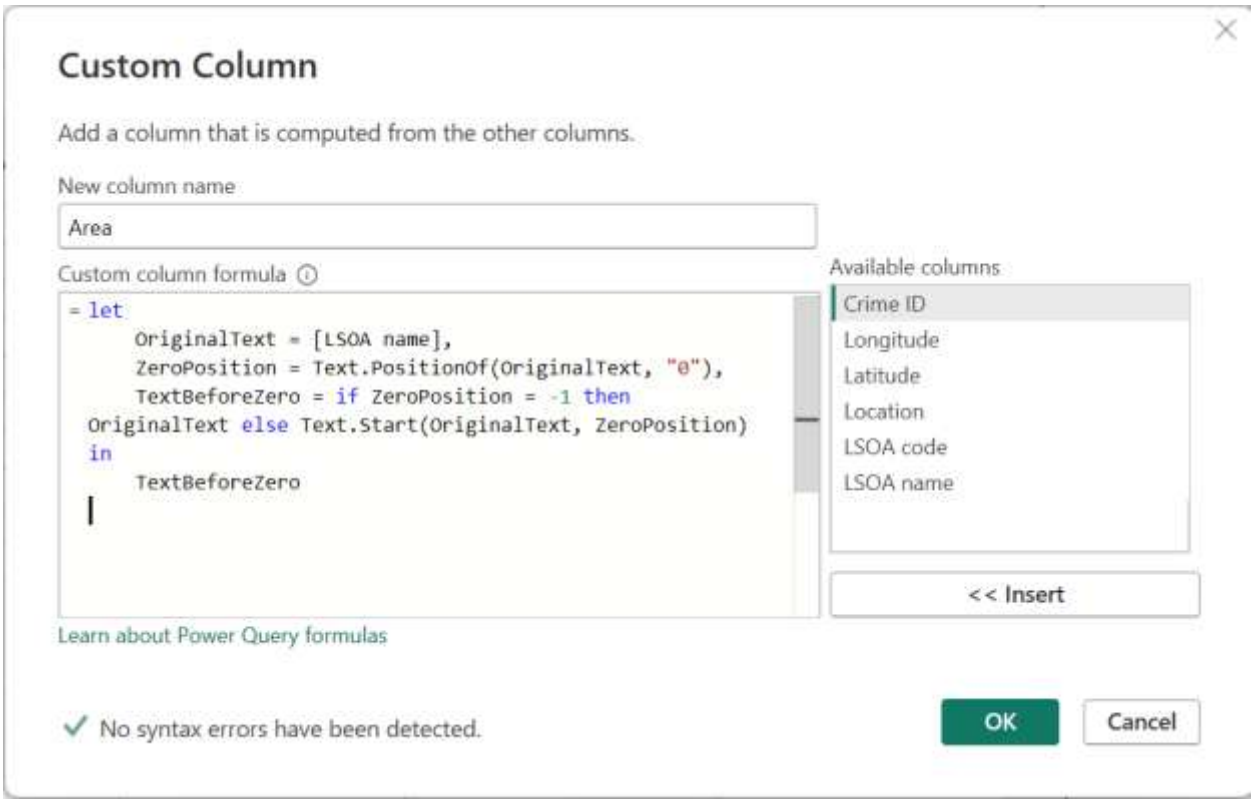

Figure 16 Get area name in LOSA Name Column using M Language Formula.

Crime Count: I have introduced an additional column named 'Crime Count', wherein I have uniformly assigned the value of 1 to every row. This deliberate enhancement is instrumental in developing a new measure named 'Count By Crime'. This measure is pivotal as it simplifies the process of quantifying crime incidents, thereby laying a solid foundation for an in-depth and methodical analysis of crime occurrences. By transforming each incident into a countable entity, it becomes more straightforward to analyze and understand crime trends and distributions."

Calculated Measure

Several calculated measures were introduced for enhanced analysis:

Crime Type Table:

Total Crime by Type

COUNTROWS(RELATEDTABLE('Street Crime')). This expression counts the number of crime incidents from the 'Street Crime' table that correspond to each crime type, effectively quantifying crimes by their types. It's a succinct way to analyze the distribution of various crime categories in the dataset.

Unique Crime Types

DISTINCTCOUNT (CrimeType[Crime type]). This measure calculates the total number of distinct crime types in the dataset by counting the unique entries in the 'Crime type' column of the 'CrimeType' table. It's useful for identifying the variety of crime categories present in the data.

Street Crime Table:

1. Count by Crime

CALCULATE(SUM('Street Crime'[Crime Count]), ALLEXCEPT('Street Crime', 'CrimeType'[Crime type], 'Street Crime'[Falls within])) calculates the total number of each crime type within specific areas or jurisdictions. It sums up crime counts from the 'Street Crime' table while keeping the filters for crime type and area intact. This measure is crucial for analyzing crime frequency by type and location.

2. Crime by Type

COUNTAX(RELATEDTABLE(CrimeType), CrimeType[Crime type]) counts the unique crime types from the 'CrimeType' table, providing insights into the variety of crimes in the dataset.

3. Percentage

DIVIDE([Count By Crime], CALCULATE(SUM('Street Crime'[Crime count]), ALLEXCEPT('Street Crime', 'Street Crime'[Falls within])), 0) * 100. calculates the proportion of each crime type as a percentage of the total crimes in a specific area, ensuring safe division and providing a clear understanding of crime distribution by type within each area.

4. Total Crime

COUNTROWS('Street Crime'). calculates the total number of crime records in the 'Street Crime' table. This straightforward measure provides an overall count of all crime incidents recorded in the dataset.

5. Total Outcomes

COUNTA('Street Crime'[Last outcome category]). This measure calculates the total number of entries in the 'Last outcome category' column of the 'Street Crime' table, providing a count of all recorded outcomes for the crime incidents.

Location Table:

1. Highest Crime Type and Count

```
Highest Crime Type and Count by Area =
VAR CrimeCountPerType =
    SUMMARIZE(
        'Street Crime',
        'Location'[Area],
        'CrimeType'[Crime type],
        "CrimeCount", COUNT('Street Crime'[Crime ID])
    )
VAR HighestCrimePerArea =
    TOPN(
        1,
        CrimeCountPerType,
        [CrimeCount],
        DESC
    )
VAR Result =
    SELECTCOLUMNS(
        HighestCrimePerArea,
        "CrimeType", 'CrimeType'[Crime type],
        "Count", [CrimeCount]
    )
RETURN
    IF(
        ISBLANK(SUMX(HighestCrimePerArea, [CrimeCount])),
        BLANK(),
        CONCATENATEX(Result, [CrimeType] & " (" & [Count] & ")", ", ")
    )
```

Figure 17 Highest Crime Type and Count using DAX Expression

the most frequent crime type in each area. It calculates crime counts per type per area, then uses TOPN to find the crime type with the highest count in each area, and presents this information, showing the crime type and its count for each specific area.

2. Least Crime Type and Count

```
Least Crime Type and Count by Area =
VAR CrimeCountPerType =
    SUMMARIZE(
        'Street Crime',
        'Location'[Area],
        'CrimeType'[Crime type],
        "CrimeCount", COUNT('Street Crime'[Crime ID])
    )
VAR LowestCrimePerArea =
    TOPN(
        1,
        CrimeCountPerType,
        [CrimeCount],
        ASC,
        [CrimeCount],
        0
    )
VAR Result =
    SELECTCOLUMNS(
        LowestCrimePerArea,
        "CrimeType", 'CrimeType'[Crime type],
        "Count", [CrimeCount]
    )
RETURN
    IF(
        ISBLANK(SUMX(LowestCrimePerArea, [CrimeCount])),
        BLANK(),
        CONCATENATEX(Result, [CrimeType] & " (" & [Count] & ")", ", ")
    )
```

Figure 18 Lowest Crime Type and Count using Dax Expression

Crime type with the lowest count in each area, summarizing crime data and using TOPN to identify the least frequent crime type, presenting it along with its count for each area.

3. Total Crime by Location

COUNTROWS(RELATEDTABLE('Street Crime')). calculates the number of crime incidents associated with each specific location. This measure counts the rows in the 'Street Crime' table that are related to each entry in the 'Location' table, giving a total count of crimes for every distinct location in the data.

4. Unique Locations

DISTINCTCOUNT(Location[Location]). calculates the total number of distinct locations in the 'Location' table. This measure provides a count of unique geographical locations where crimes have been reported in the dataset

Report Table:

1. Total Report

COUNTROWS(Report). calculates the total number of entries in the 'Report' table. This measure gives the overall count of crime reports recorded in the dataset

These calculated columns and measures are instrumental in providing a nuanced understanding of the data, allowing for a detailed analysis of street crime trends and patterns in the covered areas.

Data Model

The process of normalizing the data involved breaking down the single 'Street Crime' table into three separate tables: 'Location', 'Crime Type', and 'Report'.
This approach helps in organizing the data into a more manageable and efficient structure.

Normalization

Normalization was achieved by following these steps:

Street Crime: Retained essential columns like 'Crime ID', 'Date', 'Falls within', and 'Last outcome category', making it the central fact table that holds the key data points for each crime incident.
Location: This dimension table includes 'Crime ID' (as a foreign key), 'Latitude', 'Longitude', 'Location', 'LSOA code', and 'LSOA name'. These columns contain geographical data that describe where each crime occurred.

Crime Type: Contains 'Crime ID' (as a foreign key) and 'Crime type'. This table categorizes each crime incident, allowing for analysis based on the type of crime.

Report: Includes 'Crime ID' and 'Reported by', linking each crime to the entity that reported it.

By duplicating the original 'Street Crime' table and retaining only the necessary columns for each of the new tables, a clean and normalized set of tables was created.

Create Relationship

Power BI's auto-detection feature recognized the relationships between these tables based on the 'Crime ID' column, which acts as a primary key in the 'Street Crime' table and a foreign key in the other tables. The relationships are likely one-to-many, with 'Street Crime' being the one side, indicating that each crime ID is unique to the 'Street Crime' table and can be related to multiple entries in the other tables if there are multiple reports, locations, or types associated with a single crime.

The data model diagram shows how the data is connected. Having such a normalized data model is beneficial for several reasons:

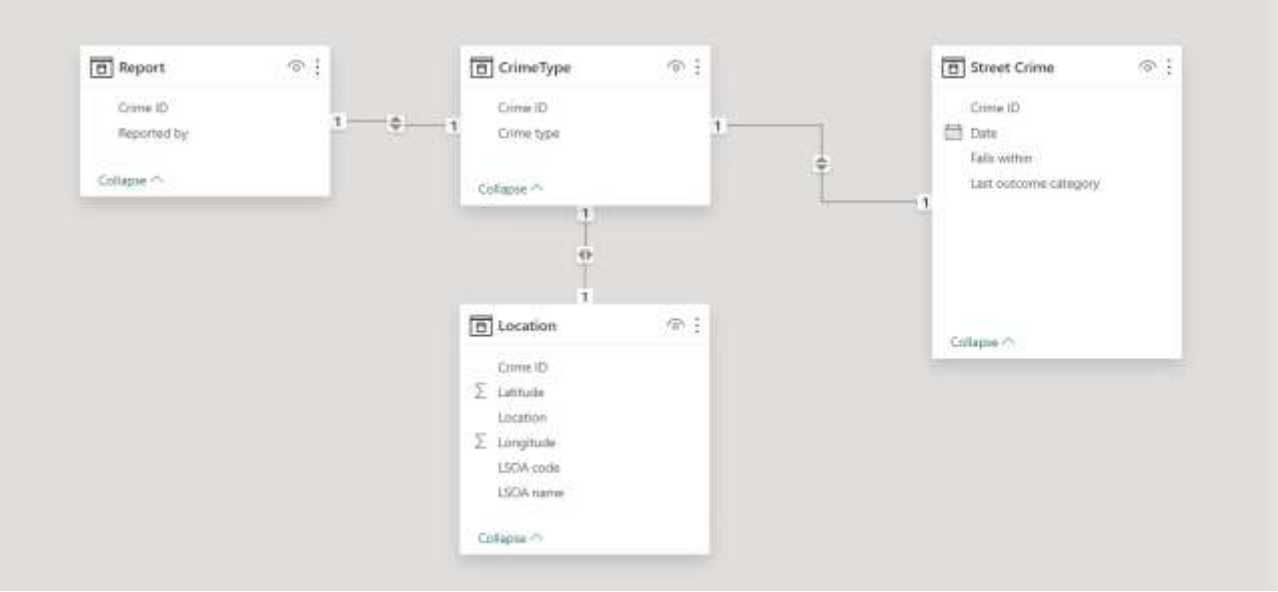

Figure 19 Entity Relationship Diagram

It reduces redundancy and inconsistency, ensuring that each data element is stored only once. It simplifies the maintenance of the data because updates, deletions, and insertions are made in just one place. It enhances data integrity and accuracy, which are crucial for reliable analysis and reporting. With the relationships established, we can now perform a wide range of analyses using Power BI's powerful data visualization tools. we can explore crime trends over time, analyze crime by location, and compare the frequency of different crime types, among other insights. This normalized data model is a strong foundation for a comprehensive BI solution.

Data Visualization

Dashboards

Crime Overview Dashboard

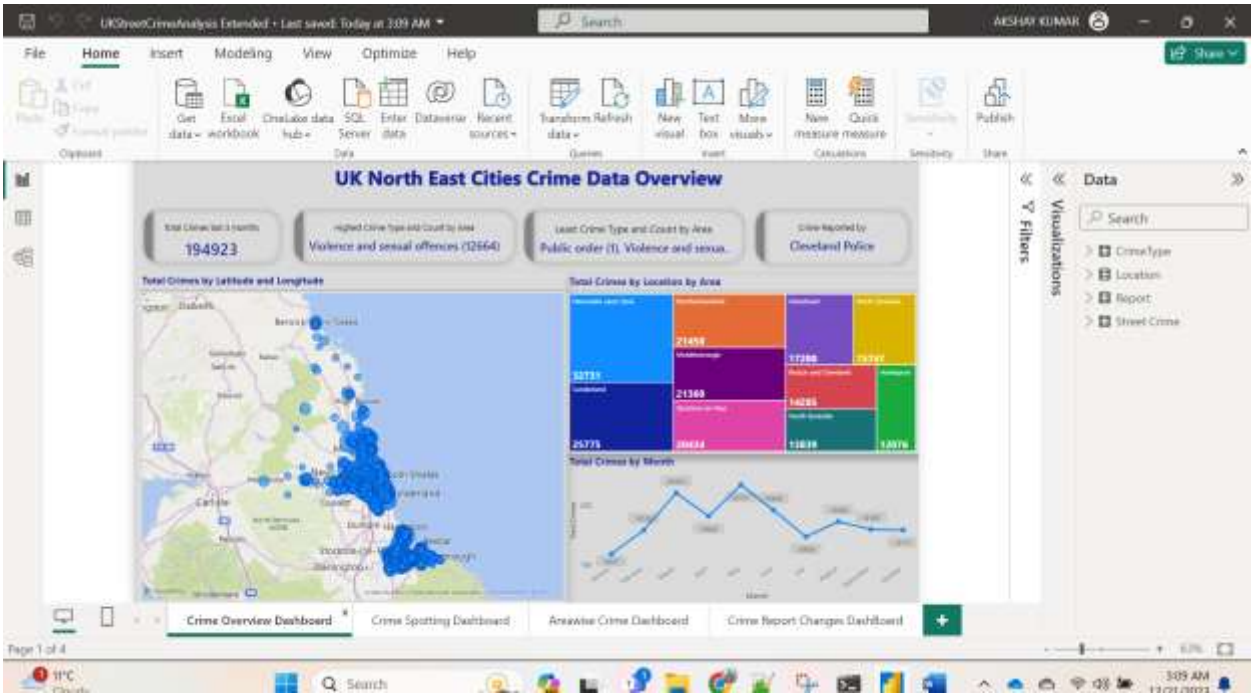

Figure 20 Crime Overview Dashboard

Crime Sporting Dashboard

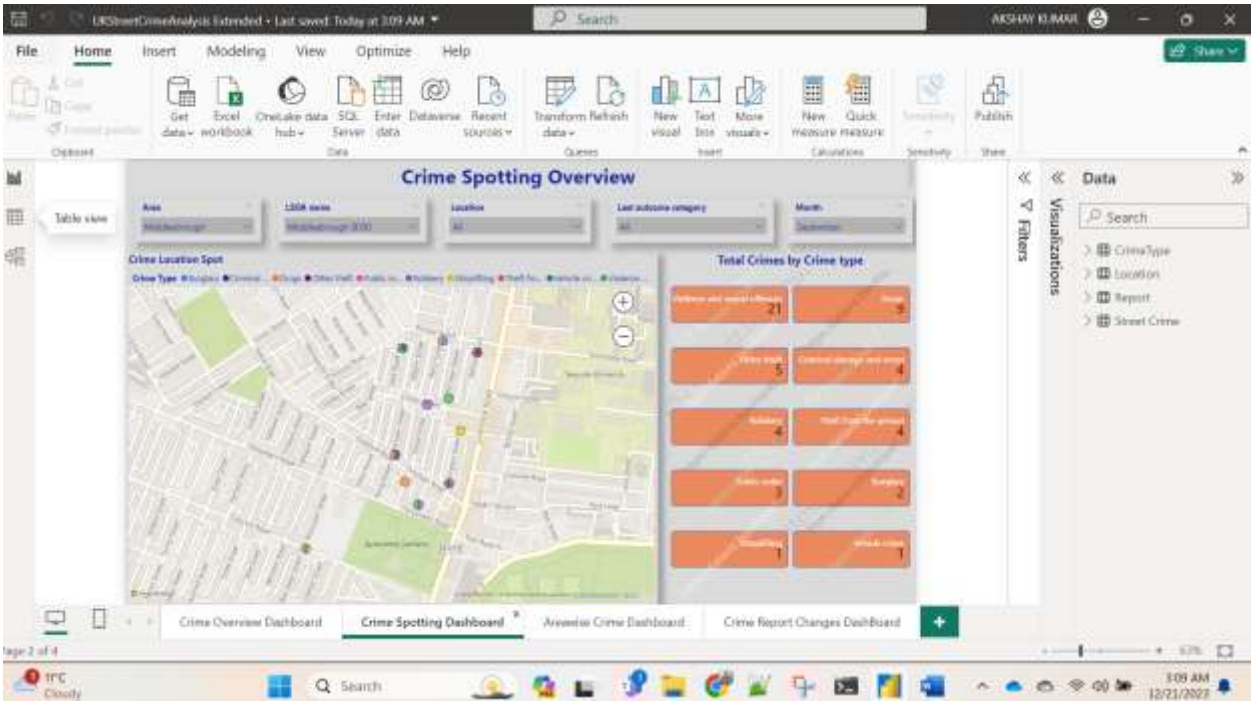

Figure 21 Crime Spotting Dashboard

Area wise Crime Dashboard

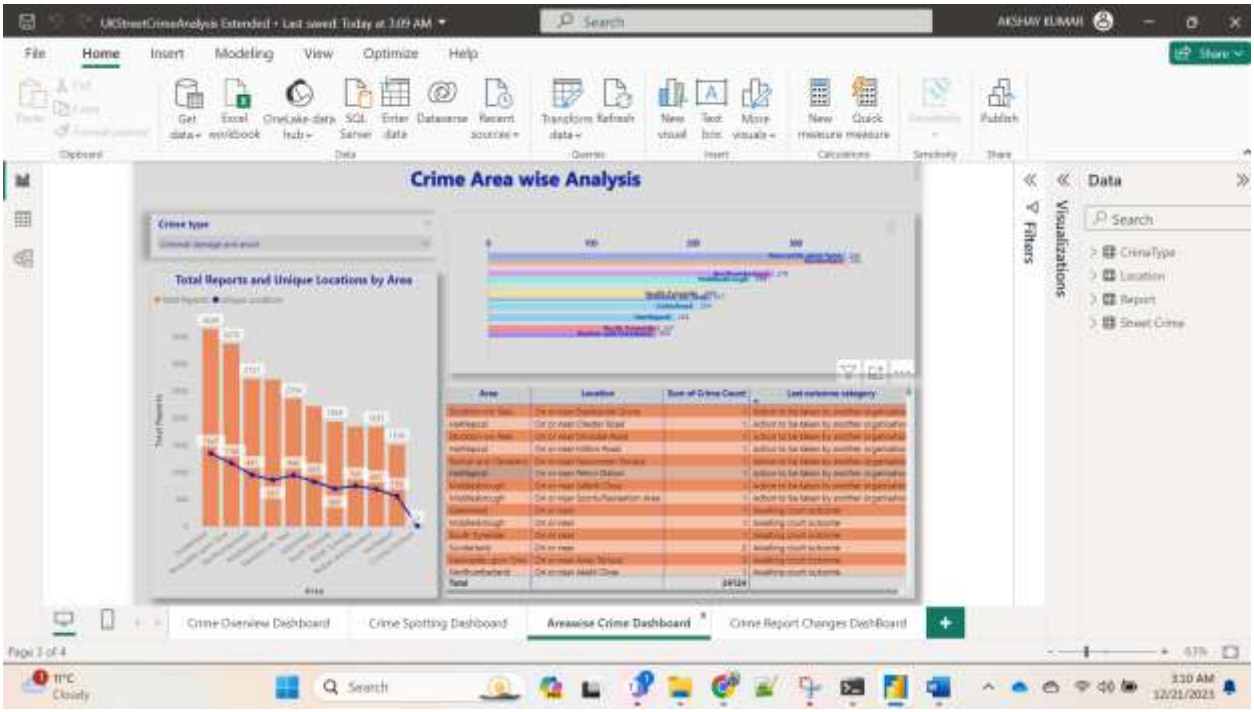

Figure 22 Area wise Crime Dashboard

Crime Report Changes Dashboard

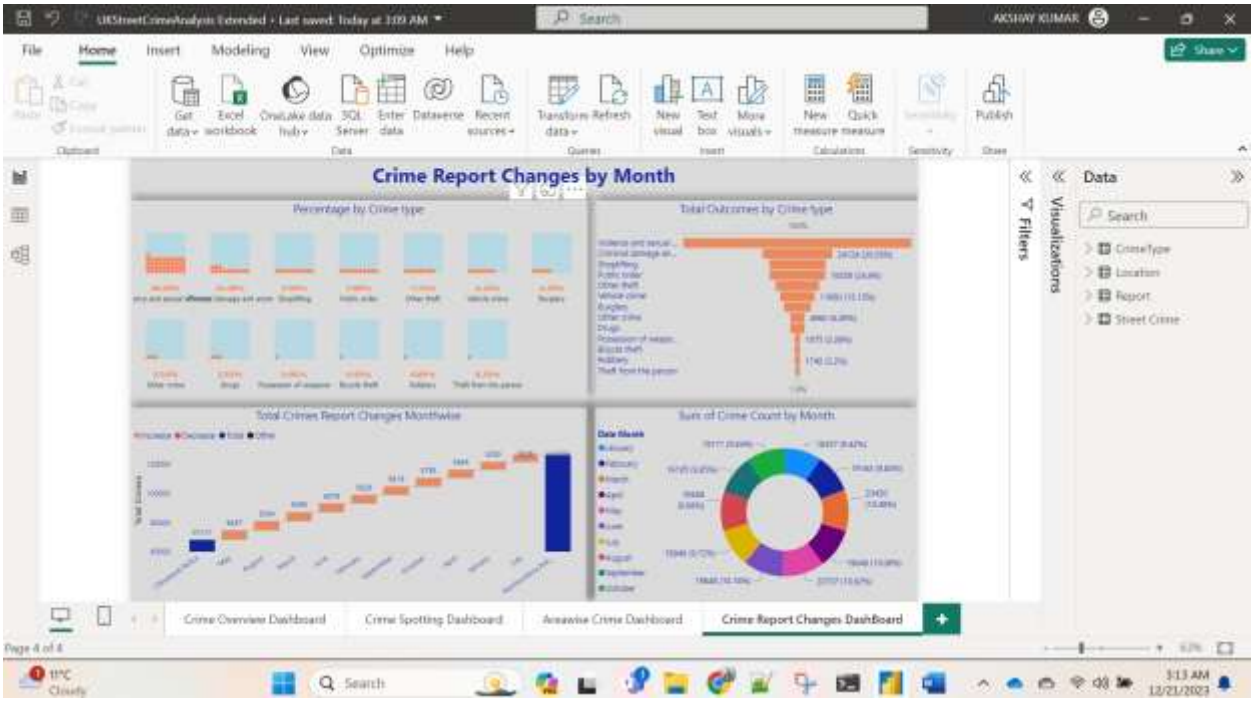

Figure 23 Crime Report changes Dashboard

References

1. https://www.youtube.com/watch?v=-4Hn8th7a4E&t=162s
2. https://github.com/jigneshk5/Crime-Spotting-Dashboard