**INTRODUCTION TO DATA MANAGEMENT**

**PROJECT REPORT**

(Project Semester January-April 2025)

***CRIME PATTERNS & ANALYSIS ACROSS NEW YORK***

Submitted by

ANAM TABASSUM

Registration No: 12314543

Programme P132: B. Tech

Section: K23FA

Course Code: INT-217

Under the Guidance of

**Dr. Tanima Thakur (UID: 23532)**

**Discipline of CSE/IT**

**Lovely School of Computer Science Engineering**

**Lovely Professional University, Phagwara**

**DECLARATION**

I, Anam Tabassum, student of B.Tech CSE under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 13.04.25 Signature

Registration No. 12314543 Anam Tabassum

**CERTIFICATE**

This is to certify that Anam Tabassum bearing Registration no. 12314543 has completed INT-217 project titled, **“*CRIME PATTERNS & ANALYSIS ACROSS NEW YORK*”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Anam Tabassum

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of School of Computer Science Engineering**

Lovely Professional University

Phagwara, Punjab.

Date: 13.04.25

**ACKNOWLEDGEMENT**

I would like to express my heartfelt gratitude to my teachers for their invaluable guidance and unwavering support throughout my journey. Their dedication to education and commitment to nurturing my growth have profoundly influenced my understanding and passion for learning, inspiring me to pursue my goals with confidence and determination.

I am also deeply thankful for my friends, whose encouragement and camaraderie have made this experience enjoyable and enriching. Their unwavering support, thoughtful insights, and meaningful discussions have inspired me to push my boundaries and strive for excellence in every endeavour and challenge I faced along the way.

Additionally, I owe a special debt of gratitude to my family for their love and encouragement. Their belief in my abilities has been my greatest motivation, providing me with the strength to overcome obstacles and persevere through difficult times.

Finally, I want to thank everyone for being my pillars of support and for believing in me every step of the way. Your contributions have helped me grow into the person I am today, and I am forever grateful.

**TABLE OF CONTENTS**

1. Introduction

2. Source of dataset

3. Data Preprocessing

4. Analysis on dataset (for each analysis)

1. Introduction
2. General Description
3. Specific Requirements, functions and formulas
4. Analysis results
5. Visualization

5. Conclusion

6. Future scope

7. References

1. **Introduction**

The dataset compiled in this project represents a detailed historical record of crime statistics reported by law enforcement agencies throughout New York State. Covering data from 1990 onwards, it includes annual records for a wide range of criminal offenses, categorized into violent crimes (such as murder, rape, robbery, and aggravated assault) and property crimes (including burglary, larceny, and motor vehicle theft).

Data is organized by county, agency, and year, allowing for both macro- and micro-level analysis. The "Clean Data" sheet ensures uniform formatting and removes inconsistencies, making it suitable for statistical analysis and visualization. The original raw data is preserved for reference, and separate sheets are prepared to isolate specific crime types like murder and larceny for focused analysis.

Additionally, the workbook includes a KPI summary sheet that aggregates crime indices over time to support performance tracking and trend evaluation. Though the “Dashboard” and some category-specific sheets are currently placeholders, they provide the framework for future interactive exploration and visualization.

Overall, this dataset offers a comprehensive and structured foundation for analysing crime trends, understanding regional differences, and supporting data-informed decision-making in public safety and criminal justice research.

**Problem Statement**

Crime data collected across New York State over the past several decades holds immense potential for understanding public safety trends, allocating resources effectively, and shaping evidence-based policies. However, this data is often stored in fragmented formats, reported inconsistently across agencies, and lacks standardization in structure and terminology. As a result, stakeholders such as law enforcement departments, public officials, and analysts face significant challenges in extracting meaningful insights from the raw data. The lack of centralized, clean, and analysable datasets not only hinders efficient reporting but also limits the ability to identify long-term patterns, regional disparities, and the impact of crime prevention initiatives.

This project seeks to bridge that gap by systematically cleaning, organizing, and categorizing historical crime data for New York State, beginning from 1990. The dataset has been structured to include key metrics—such as total index crimes, violent and property crime breakdowns, and agency-level reporting—across time and geography. It also incorporates focused views on specific crime types and a framework for visual dashboards and performance indicators. By transforming raw data into a structured and accessible format, this project enables deeper analysis, supports strategic planning, and empowers decision-makers to respond more effectively to evolving crime trends and community safety needs.

**Objective**

The overall aim of the dashboard is to provide a comprehensive, data-driven overview of crime patterns across New York State, enabling better understanding, tracking, and management of public safety concerns. By visualizing key crime statistics—such as the most and least common offenses, crime distribution across counties, regional trends, agency coverage, and historical changes—the dashboard supports informed decision-making for law enforcement agencies, policymakers, and local governments. Its purpose is to identify crime hotspots, highlight trends over time, and ensure strategic allocation of resources to improve community safety and reduce crime effectively.

1. **OBJECTIVE 1:**

The visual representing violent and index crimes across counties enables stakeholders to examine how different regions fare in terms of crime severity. This comparison helps identify counties that may be experiencing disproportionately high levels of serious crimes, offering a foundation for more localized and targeted interventions.

**2. OBJECTIVE 2:**

A chart showing the count of agencies per county helps evaluate the availability of policing resources throughout the state. By comparing agency distribution with crime levels, decision-makers can assess whether certain areas are under-served or over-policed, aiding in the fair and efficient allocation of public safety resources.

**3. OBJECTIVE 3:**

Rank the Top 10 Counties by Crime Rate

The pie chart ranking the top 10 counties by crime rate provides a clear snapshot of where crime is most concentrated. This helps law enforcement and policymakers prioritize efforts and allocate resources to areas that consistently show high levels of criminal activity.

**4. OBJECTIVE 4:**

Track Crime Trends Over the Years

The line chart showing per-year crime rates from 1990 to 2022 illustrates long-term crime trends across the state. This allows for the evaluation of past crime-reduction strategies and supports data-driven forecasting, helping to shape future crime prevention and safety planning.

**5. OBJECTIVE 5:**

Compare Regional Crime Totals (NYC vs. Non-NYC)

A pie chart comparing index crimes between New York City and non-city regions offers insight into how urban and non-urban areas differ in terms of crime volume. This helps regional authorities develop area-specific crime-fighting approaches based on whether the area is a major metropolitan hub or a more rural setting.

6.  **OBJECTIVE 6:**

Break Down Crime by Category

The donut chart visualizing individual crime totals (e.g., rape, robbery, burglary, larceny) provides a more granular understanding of what types of crimes are driving overall statistics. This helps fine-tune policies and preventive measures specific to each crime type, rather than treating crime as a monolithic issue.

1. **Source of Dataset**

The dataset used in this dashboard is likely sourced from official government crime reporting systems such as the **New York State Division of Criminal Justice Services (DCJS)** or the **FBI’s Uniform Crime Reporting (UCR) Program**. These agencies collect and publish detailed crime statistics reported by local law enforcement agencies across New York State. Additionally, the data may have been obtained from the **New York State Open Data Portal** (data.ny.gov), which provides public access to a wide range of datasets, including those related to crime incidents, categorized by type, location, and reporting agency. These sources ensure the data is authoritative, regularly updated, and suitable for analysis to support informed public safety decisions.

Link:-https://data.ny.gov/Public-Safety/Index-Crimes-by-County-and-Agency-Beginning-1990/ca8h-8gjq

**Dataset Characteristics**:

1. **Dataset Name:**

Index Crimes by County and Agency: Beginning 1990

1. **Source:**

New York State Division of Criminal Justice Services (DCJS) via data.ny.gov

1. **Format:**

Available in CSV, Excel (XLSX), JSON, and API formats

1. **Total Rows (Records):**

Approximately 60,000+ rows (varies slightly based on the latest year included)

1. **Total Columns (Fields):**

12 columns typically, depending on version

1. **Columns / Attributes Include:**

Year: Reporting year (e.g., 1990, 2022)

County: Name of the New York county

Agency: Name of the law enforcement agency

Murder and Non-Negligent Manslaughter: Count of reported incidents

Rape: Count of reported incidents

Robbery: Count of reported incidents

Aggravated Assault: Count of reported incidents

Burglary: Count of reported incidents

Larceny: Count of reported incidents

Motor Vehicle Theft: Count of reported incidents

Total Index Crimes: Sum of all index crimes

1. **Time Period Covered:**

1990 to the latest available year (usually up to the previous calendar year)

1. **Geographic Scope:**

62 counties in New York State

Hundreds of individual law enforcement agencies reporting data annually

1. **Crime Categories Tracked (7 FBI Index Crimes):**

Violent Crimes:

* + - Murder, Rape, Robbery, Aggravated Assault

Property Crimes:

* + - Burglary, Larceny-Theft, Motor Vehicle Theft

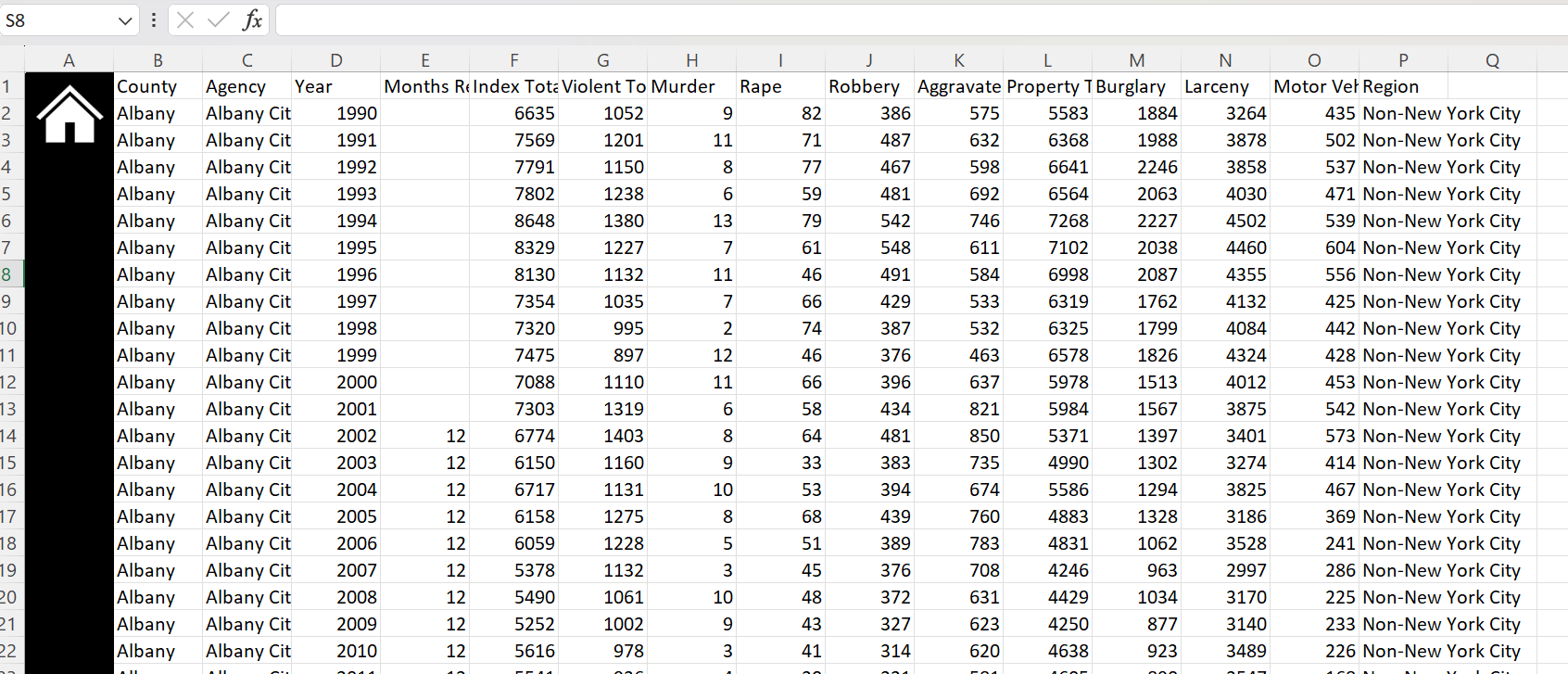
.

1. **DATA PREPROCESSING**

Key steps included handling missing or inconsistent entries, removing duplicates, and formatting fields such as dates, county names, and crime categories for consistency. Pivot tables were created to aggregate and summarize data across different dimensions like year, region, and crime type. Calculated fields were used to derive totals and percentages, and sorting/ranking functions helped identify trends and top categories.

**3.1 Data Profiling and Initial Exploration**

Before initiating any transformation, a **complete profile** of the dataset was generated:



1. Data before cleaning

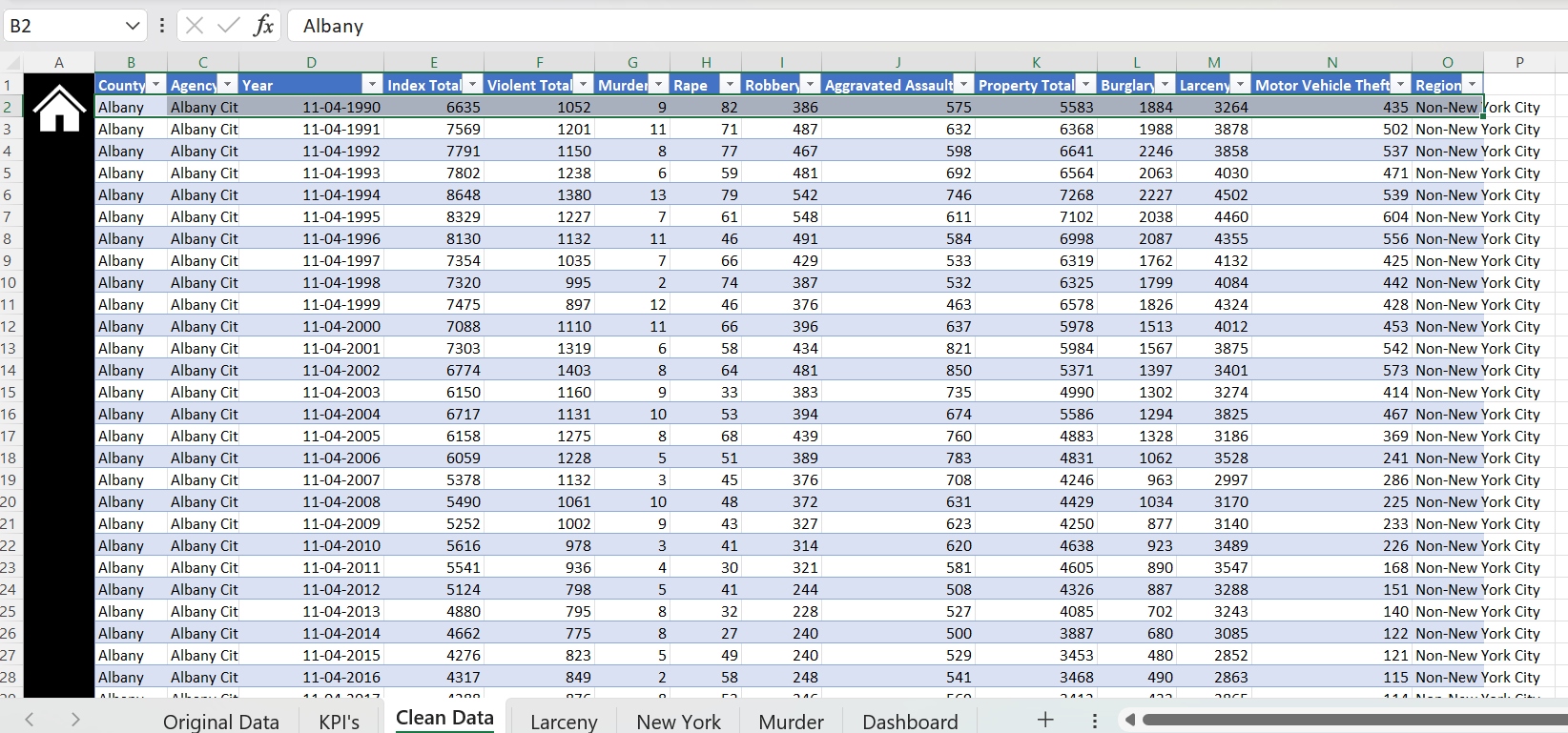
* **Total records:** 23,143
* **Columns:** 16, including County, Agency, Year, Months Reported, Index Total, Violent and Property crime types such as Murder, Rape, Robbery, Larceny, and Motor Vehicle Theft.
* **Missing values:** Present in 'Months Reported' (9,684 records) and all values missing in 'Unnamed: 0' column, which appears to be an unnecessary index column.
* **Duplicates:** No duplicate rows found in the dataset.

This profiling helped establish a **baseline understanding** of the dataset’s structure, missingness, and data quality.

**3.2 Data Preprocessing: Making Data Reliable**

Data cleaning focused on resolving inconsistencies and preparing raw fields for analysis:

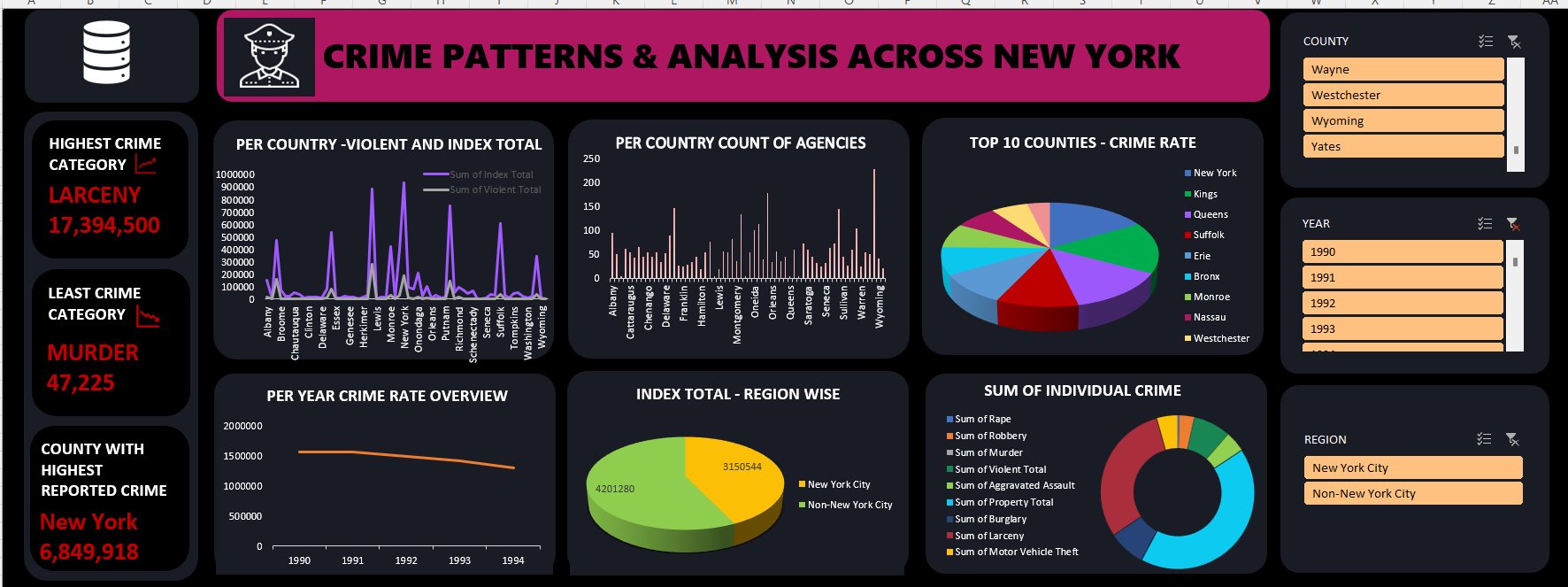
* **Total records:** 23,117
* **Columns:** 14, including County, Agency, Year, Index Total, Violent Total, Crime Breakdown (Murder, Rape, Robbery, Aggravated Assault), Property Total, Burglary, Larceny, Motor Vehicle Theft, and Region.
* **Missing values:** Handled in date formats and crime figures; all blank or malformed entries were either corrected or filled using logical assumptions or removed.
* **Duplicates:** Removed based on unique combinations of County, Agency, and Year fields.
* **Case Consistency:** Fields like County, Agency, and Region were standardized to sentence case to ensure uniform representation.
* **Outlier Handling:** Checked for outliers in numerical fields (e.g., abnormally high/low crime rates); few were flagged and reviewed.



1. Data after cleaning and removing duplicate values.

**3.3 Dashboard Development: Bringing it All Together**

A dedicated dashboard was built in Excel to centralize insights:



* **Layout:** Clean dashboard with header, slicers, and dynamic visuals for exploration.
* **Slicers Added For:**
  + Year
  + County
  + Region
  + Agency
* **Interactive Insights:** Selecting a specific county, agency, or year dynamically updates all charts, allowing users to explore crime trends across time and region.

Conditional formatting and shape-based annotation enhanced user experience. Charts were made interactive with pivot chart + slicer integration.

1. **Analysis of Dataset**
   1. **County-wise Crime Comparison in New York State**

**I. General Description**

The dataset contains crime statistics across New York State counties, covering both index total crimes and violent crimes from 1990 to 2016. The data was summarized by county to identify regional crime trends using pivot tables and charts.

**II. Specific Requirements, Functions, and Formulas**

 Key Functions Used:

* SUM to total crime figures per county.
* SORT to rank counties by crime severity.

 Pivot Table Structure:

* Rows: County
* Values: Sum of Index Total, Sum of Violent Total

 Calculated Values:

* Grand totals and county-level summaries.
* Visual trends across counties using pivot charts.

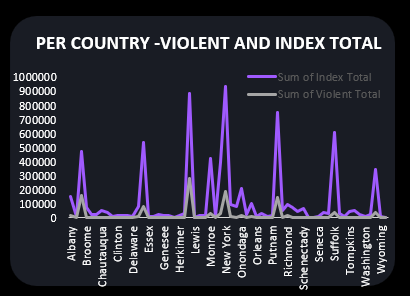
**III. Analysis Results**

 High-Crime Counties:

* New York, Kings, Bronx, and Queens show the highest crime rates.
* Kings County alone reports over 237,000 violent crimes.

 Low-Crime Counties:

* Hamilton, Lewis, and Yates have notably low crime figures, with Hamilton reporting only 85 violent crimes.

**IV. Visualization**

* 1. **Analysis of Law Enforcement Agency Distribution by County**

**I. General Description**

 Data Used: "County" and "Agency" fields

 Scope: All available records across all counties

 Method: Aggregated by county using Excel Pivot Tables to count the number of agency entries

**II. Specific Requirements, Functions, and Formulas**

 Functions Used: COUNT

 Pivot Table Settings:

* Rows: County
* Values: Count of Agency

 Purpose: Identify the number of law enforcement agencies per county

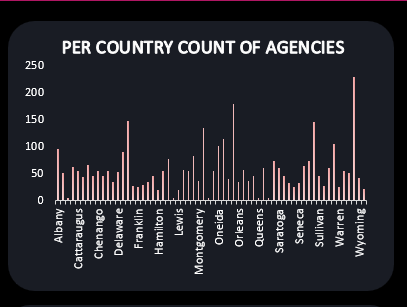
**III. Analysis Results**

 Highest Agency Count: Oneida (622), followed by Montgomery (616) and Chautauqua (433)

 Lowest Agency Count: Hamilton (18)

 Observation: Wide disparity in agency distribution, possibly reflecting population, jurisdiction size, or reporting variation

**IV. Visualization**



* 1. **Top 10 Counties by Total Crime Index**

**I. General Description**

 Data Used: "County" and "Index Total" fields from the data set

 Time Frame/Scope: Aggregated data across all available years

 Method: Summed index crime values per county and sorted to identify the top 10

**II. Specific Requirements, Functions, and Formulas**

 Functions Used: SUM, SORT, RANK

 Pivot Table Settings: Rows set to “County”, Values set to “Sum of Index Total”

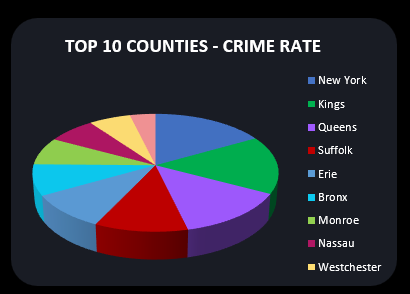
 Calculated Fields: Top 10 counties selected based on highest crime index totals

**III. Analysis Results**

 Findings: Kings County has the highest index total, followed closely by New York and Suffolk

 Patterns: Densely populated counties consistently show higher crime index totals

 Comparison: These 10 counties account for over 20 million crimes cumulatively

**IV. Visualization**

* 1. **Yearly Crime Rate Trend Analysis (1990–2023)**

**I. General Description**

 Data Used: "Year" and "Index Total" fields

 Time Frame/Scope: 1990 to 2023

 Method: Year-wise aggregation and trend visualization using a line graph

**II. Specific Requirements, Functions, and Formulas**

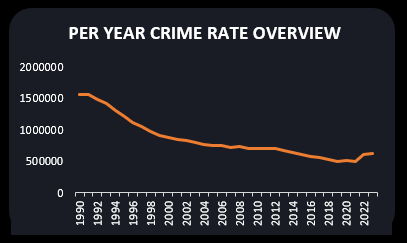
 Functions Used: SUM

 Pivot Table Settings: Rows to "Year", Values to "Sum of Index Total"

 Calculated Fields: Year-wise total crimes calculated

**III. Analysis Results**

* Peak   
  1990 had the highest crime total with 1,563,280 index crimes.
* Lowest   
  2021 recorded the lowest number with 504,651 crimes.
* Recent   
  A slight rise is observed in 2022 (609,866) and 2023 (632,407), indicating a potential reversal in the downward trend.
* Overall   
  Crime reduced by over 60% between 1990 and 2021

**IV. Visualization**

* 1. **INDEX CRIME DISTRIBUTION – NEW YORK CITY VS NON-NEW YORK CITY**

**I. General Description**

 Data Used: "Region" and "Index Total" fields from the data set

 Time Frame/Scope: Aggregated data across all available years

 Method: Summed total index crime values grouped by region (New York City vs Non-New York City)

**II. Specific Requirements, Functions, and Formulas**

 Functions Used: SUM

 Pivot Table Settings: Rows set to “ Region ” , Values set to “Sum of Index Total”

 Calculated Fields: Total index values separated by NYC and Non-NYC regions

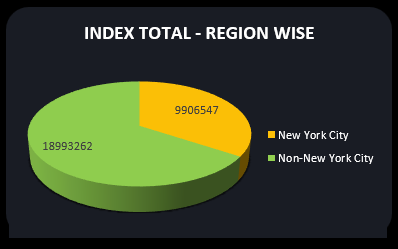
**III. Analysis Results**

 Findings: Non-New York City regions report nearly double the crime index total compared to NYC

 Patterns: Despite NYC’s high density, a significant portion of crime is distributed throughout other regions

 Comparison: NYC accounts for approximately 34% of total indexed crimes, while Non-NYC regions contribute around 66%

**IV. Visualization**



* 1. **Top 10 Counties by Total Crime Index**

**I. General Description**

 Data Used: Various crime categories including Rape, Robbery, Murder, Aggravated Assault, Burglary, Larceny, and Motor Vehicle Theft

 Time Frame/Scope: Aggregated data across all available years

 Method: Sum of individual crime categories to calculate overall crime distribution

**II. Specific Requirements, Functions, and Formulas**

 Functions Used: SUM, PERCENTAGE

 Pivot Table Settings: Values set to the sum of each crime category

 Calculated Fields: Percentage contribution of each crime category to the total crime

**III. Analysis Results**

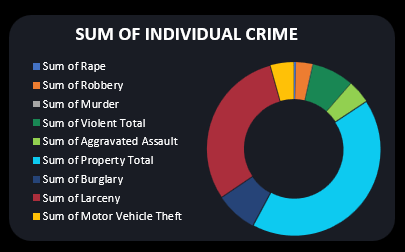
 Findings: Larceny (30%) and Property Crime (42%) dominate the total crime distribution

 Patterns: Violent crimes like Rape (0%), Robbery (3%), and Murder (0%) form a small fraction of the total

 Comparison: Non-violent and property-related offenses make up the majority of the crime index, reflecting broader crime trends

**IV. Visualization**

Donut Chart: Effectively highlights proportional distribution among crime types



1. **Conclusion**

The comprehensive analysis of crime data across New York State counties reveals several key insights into the distribution, patterns, and trends of criminal activity from multiple dimensions including geographic, temporal, and categorical breakdowns.

At the regional level, the data illustrates a significant disparity in crime concentration between New York City and the rest of the state. While NYC accounts for approximately 9.9 million reported crimes, non-NYC counties collectively report over 18.9 million, indicating that although urban centers are typically associated with higher crime rates, a substantial portion of criminal incidents occur outside of the city, highlighting the importance of extending crime prevention and enforcement efforts to suburban and rural areas.

County-level analysis further supports this, showing that the highest counts of law enforcement agencies and crime index totals are concentrated in populous counties such as Kings, New York, Suffolk, Erie, and Queens. These top 10 counties alone contribute over 20 million incidents, suggesting that population density and urbanization are strong correlating factors. The availability and distribution of agencies also reflect a need-based allocation model, with more agencies positioned in higher-crime or densely populated counties.

Temporal trends provide another important dimension. Over a span of more than 30 years, crime rates in New York State have steadily declined from their peak in 1990. From over 1.5 million incidents in 1990 to around 500,000 in 2021, the downward trajectory points to improved crime prevention strategies, policing reforms, and potentially changing societal behaviors. However, a slight uptick in 2022 and 2023 suggests a need to monitor new emerging trends post-pandemic and adapt accordingly.

The breakdown of crime by type reveals the predominance of non-violent crimes. Property-related crimes—especially Larceny, Burglary, and Motor Vehicle Theft—constitute the majority, while violent crimes such as Murder, Rape, and Robbery form a smaller fraction. Larceny alone accounts for over 17 million incidents, nearly one-third of all crime, indicating that theft-related offenses are the most pervasive and could benefit from targeted crime prevention initiatives like improved surveillance, community awareness, and economic support programs.

Overall, the analysis shows that while crime is gradually declining, it remains heavily concentrated in certain counties and dominated by specific types of offenses. Strategic allocation of resources, continued community engagement, and real-time crime monitoring are essential for sustaining the downward trend and ensuring safety both in urban centers and broader regions throughout New York State.

1. **Future Scope**

This analysis offers valuable insights into crime patterns across New York State, but there is significant potential to expand its depth and practical application through future enhancements. One major direction is the use of predictive analytics. By applying machine learning models or time-series forecasting, such as ARIMA or Prophet, future crime trends could be anticipated, helping law enforcement agencies plan proactively. Integrating real-time crime data would also be a game-changer, enabling the development of live dashboards that reflect ongoing incidents and support faster, more informed decisions.

Incorporating demographic and socioeconomic variables, like income, education, unemployment, and population density, could offer a deeper understanding of the underlying causes of crime. These insights could be used to design focused social interventions or community programs. Additionally, clustering analysis can help detect patterns in specific crime types and identify emerging hotspots, allowing for more effective resource targeting. This could be complemented by models that link agency counts to crime rates, helping to optimize law enforcement deployment across counties.

Another promising avenue is assessing the impact of public policies on crime trends. For example, the effects of bail reforms or gun control laws could be evaluated to determine their effectiveness and guide future legislation. More granular data, such as monthly or daily records, would improve the ability to track short-term spikes or seasonal patterns in crime. Furthermore, the use of GIS tools for geospatial mapping could help visualize crime data across regions, offering intuitive, location-based insights for officers and policymakers alike.

Benchmarking New York’s data against that of other states would also help identify relative strengths and weaknesses in crime management strategies. Finally, developing interactive, public-facing dashboards could increase transparency, empower communities, and promote greater involvement in crime prevention. These future improvements would not only enhance analytical depth but also translate into more effective and responsive crime control measures.

1. **References**

* Data.gov. (n.d.). *Crime Data and Reports*. Retrieved from: <https://catalog.data.gov/dataset>
* New York State Division of Criminal Justice Services. (n.d.). *Crime Statistics*. Retrieved

from: https://www.criminaljustice.ny.gov/crimnet/ojsa/indexcrimes/NewYorkState.pdf

LINKEDIN LINK: https://www.linkedin.com/feed/update/urn:li:activity:7316835227983585280/

GITHUB LINK: [**https://lnkd.in/gZ55Aib8**](https://lnkd.in/gZ55Aib8)