DATA VISUALIZATION AND STORY TELLING

TEAM - 15

FINAL PROJECT PHASE-1

PRANAY DATTA KAVUKUNTLA, LOKESH VANGULA

INTRODUCTION:

Dataset: **Crime Incident Reports in Worcester**

- There are a total of 398,373 records in this collection, and they provide a comprehensive description of criminal activity in Worcester. There is a lot of thought put into the structure of the database, with fields like "INCIDENT_NUMBER," "DATE_LOGGED," "YEAR," "TIME_LOGGED," "DEPARTMENT," "INCIDENT_TYPE," "LOCATION," and "ZIPCODE" for each record. With this much information, we can examine crime trends from several angles, including when they occur (in terms of time and date), where they occur (in terms of location and zip code), what kinds of crimes are committed, and which police agencies are involved.

Key Characteristics:

- **Time Series:** YEAR and TIME_LOGGED offer insights on when crimes occur, potentially indicating patterns like seasonal changes or time-of-day tendencies.
- **Space-Based Analyses:** Crime can be mapped geographically using LOCATION and ZIPCODE, which is essential for locating problem regions and prioritizing police resources.
- **Typology of Crime:** INCIDENT_TYPE classifies each occurrence according to its nature, which is crucial for identifying patterns of crime in the city.
- **Police Response:** By using the DEPARTMENT attribute, reaction times and police manpower can be analyzed.

ATTRIBUTES:

- Each criminal incidence is given its own INCIDENT_NUMBER. It guarantees that each record is separate and may be referred to by its own unique identifier. This is necessary for tracing individual cases or checking for data replication.
- The DATE_LOGGED field stores the time and date when the issue was initially reported. It's crucial for things like analyzing patterns over time, seeing fluctuations, and pinpointing times of high and low crime rates.
- The YEAR feature, which reflects the year of the incident, is useful for looking at trends over time. It's crucial for tracking crime rates over time and analyzing trends over time.
- When an incident is reported, the time it was reported is recorded in the TIME_LOGGED field. Time data is crucial for identifying patterns such as the time of day when particular sorts of crimes are more likely to occur, which can affect police patrolling plans and community safety measures.
- DEPARTMENT is a field that designates the police division or department that received the report and dispatched officers to the scene. It can be used to see how cases are assigned, how resources are divided, and how efficient individual departments are.

- This attribute, named INCIDENT_TYPE, classifies the type of criminal occurrence that occurred. It's vital for identifying the types of crimes prevalent in different places, helping in adjusting crime prevention plans and public safety measures to specific types of criminal activity.
- Important for spatial analysis, "LOCATION" details the incident's general location or address. It allows crime hotspots to be mapped, geographic trends to be analyzed, and potential trouble locations to be located.
- ZIP Codes provide a more specific geographical identity than just a city or town. It is essential for analyzing certain regions and figuring out how crimes are distributed in various areas. Data collected via ZIPCODE can be utilized to better target public safety and community policing efforts.

NOTE: TO ACCESS MY DATASET WHICH WAS TAKEN FROM GOOGLE DIRECTORY PLEASE GO THROUGHT THIS LINK:
https://drive.google.com/drive/folders/1d7tAwkmVTlZthy7uEmdftfwnZhjW2AuX?usp=sharing

LIMITATIONS:

The dataset, despite its extensiveness, has some restrictions due to the very nature of the data. To begin, the accuracy of the reporting and recording of criminal offenses is a crucial factor in determining the quality and breadth of the data. Misclassifications or the failure to record events could introduce bias into the analysis. Second, there is a possibility that the numbers underestimate the actual crime rate because it only takes into account incidents that were reported to the authorities and not those that were not brought to their attention.

REASON FOR SELECTING THE DATASET:

I decided to use this dataset in order to gain a deeper understanding of the dynamics behind the criminal activity in Worcester. An investigation of this nature can be of use in developing effective strategies for law enforcement and the prevention of crime. This information enables the investigation of patterns over time, the identification of high-risk zones, and the knowledge of the most common sorts of events, all of which can be utilized to drive preventative policing approaches and community safety programs. In addition to this, it provides students with a real-world application of data visualization and narrative methodologies, making it an excellent tool for use in the classroom.

AMBIGUITY IN THE DATA:

The dataset has a number of different degrees of uncertainty. Even if the INCIDENT_TYPE is useful, it might not be specific enough or it might not be consistent enough between entries to be utilized for accurately classifying crimes. There is a possibility that temporal analysis will not

be accurate if the TIME_LOGGED attribute relates to the time that the occurrence was recorded rather than the time that it actually took place. It's possible that insufficient or imprecise LOCATION and ZIPCODE data will cause precision issues in spatial analysis. The analysis of the dataset is extremely limited because there is no incident-specific context, such as severity or outcomes. This further restricts the analysis.

INTENDED AUDIENCE:

The analysis of this dataset is tailored for a diverse audience, encompassing several key groups:

- **Law Enforcement and City Officials:** These parties can apply the data to make informed choices about crime prevention, resource allocation, and city-specific policy initiatives that will better benefit the people of Worcester. Law enforcement agencies can improve their patrol strategies and community outreach by reviewing crime statistics.
- **Academic and Research Community:** Researchers and scholars in fields including criminology, urban studies, public administration, and data science will find this information invaluable. It's a treasure trove of information for those studying urban crime, law enforcement, and public security.
- **Government Agencies:** Law enforcement agencies at all levels, from city to state to federal, can benefit from this data. It can be used as a basis for national policy changes, funding priorities, and safety actions. The data set provides a benchmark against which the prevalence of crime in different parts of the country may be assessed.
- **General Public:** A greater public interest in and involvement with public safety measures can result from making this data more widely available. Residents of a city can better protect themselves and actively involved in their neighborhood if they have a clear grip on the dynamics of local crime. Local law enforcement is also encouraged to be more transparent and accountable.
- **University Setting:** This dataset serves as a great resource for academic communities and may be utilized as a teaching tool to improve data analysis and visualization approaches. Colleges can use this data to better protect their students and reduce crime by gaining insight into local trends of criminal activity. This program helps with both academic requirements and campus security.

NOTE: we chose this dataset because it will allow us to work with reliable data and tackle important problems in the area around our university and all the visualizations would be based on police_incident_report_combined as that csv file is more refined as we have worked on it by including couple of columns and also refined the data more.