**Project Proposal: Data Analysis Project**

**ITC686**

**1. Source of Data:**

The dataset was obtained from the Chicago Data Portal and provided by the Chicago Police Department. The primary data source for this project is accessible at

https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2/about\_data

It is readily available for download and analysis purposes.

**2. Dataset Size:**

The dataset comprises 1.76 GB of records, providing substantial data for meaningful analysis and insights.

**3. Description of Dataset:**

The dataset contains reported crime incidents in Chicago from 2001 to the present, excluding the most recent seven days. It is updated daily, and crime classifications may undergo changes pending further investigation. Structured to facilitate analysis, it comprises 8.03 million rows and 22 columns, containing relevant fields essential for extracting insights.

**4. List of Fields:**

The pertinent fields within the dataset include:

1. **ID:** Unique identifier for the record.
2. **Case Number:** The Chicago Police Department RD Number (Records Division Number) is unique to the incident.
3. **Date:** Date when the incident occurred. This is sometimes the best estimate.
4. **Block:** The partially redacted address where the incident occurred, placing it on the same block as the actual address.
5. **IUCR:** IUCR is the Illinois University Crime Reporting Code. It is directly linked to the Primary Type and Description.
6. **Primary Type:** The primary description of the IUCR code.
7. **Location Description:** Description of the location where the incident occurred.
8. **Arrest:** Indicates whether an arrest was made.
9. **Domestic:** Indicates whether the incident was domestic-related as defined by the Illinois Domestic Violence Act.
10. **District:** Indicates the police district where the incident occurred.
11. **Community Area:** Indicates the community area where the incident occurred.
12. **FBI Code:** This code indicates the crime classification outlined in the FBI's National Incident-Based Reporting System (NIBRS).
13. **Year:** The year the incident occurred.
14. **Updated On:** Date and time the record was last updated.
15. **Location:** The location where the incident occurred in a format that allows for creating maps and other geographic operations on this data portal. This location is shifted from the actual location for partial redaction but falls on the same block.

**5. Queries**

1. Trend Analysis: Determine if there's an increasing or decreasing trend in specific types of crimes over the years.

- Query: "Retrieve the total number of incidents for each year for the primary type 'Robbery'."

2. Spatial Analysis: Identify high-crime areas by analyzing incidents in different community areas or districts.

- Query: "Find the top 5 community areas with the most reported incidents."

3. Temporal Analysis: Analyze patterns in crime occurrence based on time, such as monthly, weekly, or hourly trends.

- Query: "Show the distribution of incidents by day of the week for the primary type 'Burglary' in 2023."

4. Arrest Rate: Investigate the effectiveness of law enforcement by examining the proportion of arrests made for different types of crimes.

- Query: "Calculate the arrest rate for the primary type 'Assault' for 2022."

5. Domestic Violence: Explore the prevalence of domestic-related incidents compared to overall crime rates.

-Query: "Determine the percentage of domestic-related incidents for the primary type 'Domestic Battery' in the past five years."

6. Hotspot Identification: Identify specific locations where certain crimes are most concentrated.

- Query: "Identify the top 10 blocks with the highest number of reported incidents for the primary type 'Narcotics' in the current year."

**Project Overview**

The project aims to analyze crime incidents reported in Chicago from 2001 to the present using a comprehensive dataset provided by the Chicago Police Department. This dataset, available on the Chicago Data Portal, contains 1.76 GB of data, comprising 8.03 million rows and 22 columns. It includes essential fields such as the type of crime, location, date, arrest status, and more, which will be analyzed to derive meaningful insights and patterns related to crime in Chicago over the years.

**Objectives of the Project**

The project's primary objectives are multifaceted, aiming to provide a comprehensive understanding of crime dynamics in Chicago. Firstly, the project seeks to identify trends in specific types of crimes over the years to determine whether there are increasing or decreasing patterns. Secondly, spatial analysis will pinpoint high-crime areas across community areas or police districts. Thirdly, temporal patterns will be explored to understand when crimes are most likely to occur based on time, such as monthly, weekly, or hourly trends. Fourthly, the project will assess the effectiveness of law enforcement by calculating the arrest rate for different types of crimes. Additionally, the prevalence of domestic-related incidents will be analyzed compared to overall crime rates. Lastly, hotspot identification will be performed to identify specific locations with high concentrations of certain crimes.

**Potential Analyses and Insights**

The project offers various avenues for insightful analyses based on the dataset. For instance, analyzing the number of crime incidents over the years can reveal which types of crimes have seen significant increases or decreases. Spatial analysis can help identify areas requiring more attention from law enforcement agencies, allowing for targeted resource allocation. Temporal analysis can provide insights into crime patterns based on time, potentially uncovering correlations between specific times or days and certain types of crimes. Calculating the arrest rate for different crime types can offer insights into the effectiveness of law enforcement efforts. Moreover, analyzing domestic-related incidents can shed light on the prevalence and nature of domestic violence in the city. Finally, hotspot identification can help deploy resources more efficiently to tackle crime hotspots, contributing to crime prevention and control efforts.

**What We Can Do**

Several steps need to be undertaken to accomplish the objectives and derive meaningful insights from the dataset. Initially, data cleaning will be crucial to handle missing values, outliers, and inconsistencies in the dataset. Exploratory data analysis (EDA) will then be performed to better understand the data distribution and relationships between variables. Data visualization techniques, such as charts, graphs, and maps, will be employed to represent the findings effectively. Statistical analysis will be conducted using appropriate methods to test hypotheses, identify correlations, and derive insights. Machine learning models may also be implemented for predictive analysis or clustering to identify patterns and trends in the data. Finally, the findings, insights, and visualizations will be compiled into a comprehensive report, and a presentation will be prepared to communicate the results effectively to stakeholders.

**Tools and Technologies**

Various tools and technologies will be utilized throughout the project to facilitate data analysis, visualization, and reporting. Python libraries such as Pandas and NumPy and R and SQL will be employed for data analysis and manipulation. For data visualization, tools like Matplotlib, Seaborn, Tableau, and GIS tools for spatial analysis will be used. Machine learning libraries such as Scikit-learn and TensorFlow may also be utilized for predictive analysis or clustering tasks. Lastly, Microsoft Word and PowerPoint, or similar tools, will be used to compile the report and create the presentation to effectively communicate the project's findings.

**6. Group members:**

|  |  |
| --- | --- |
| **Name** | **CMULoginID** |
| Riad Hossain | hossa1r |
| Mahmod Besher Alrez | alrez1m |