**Project Report: Crime in Chicago (2001-Present)**

**Group Members:**

Riad Hossain - hossa1r

Mahmod Besher Alrez - alrez1m

**Project Overview:**

The project holds immense significance as it embarks on analyzing crime incidents reported in Chicago from 2001 to the present. This analysis is based on a comprehensive dataset, a testament to the credibility of our findings provided by the Chicago Police Department. The dataset 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:**

1. Identify trends in specific types of crimes over the years.
2. Conduct spatial analysis to pinpoint high-crime areas across community areas or police districts.
3. Explore temporal patterns to understand when crimes are most likely to occur.
4. Assess the effectiveness of law enforcement by calculating arrest rates for different types of crimes.
5. Analyze the prevalence of domestic-related incidents compared to overall crime rates.
6. Perform hotspot identification to identify specific locations with high concentrations of certain crimes.

**Potential Analyses and Insights:**

* Trends in crime incidents over the years.
* Spatial analysis to identify high-crime areas.
* Temporal patterns to understand when crimes occur.
* Effectiveness of law enforcement through arrest rate analysis.
* Prevalence and nature of domestic-related incidents.
* Identification of crime hotspots for targeted interventions.

**What We Can Do:**

1. Data cleaning to handle missing values, outliers, and inconsistencies.
2. Exploratory data analysis (EDA) to understand data distribution and relationships.
3. Data visualization using charts, graphs, and maps.
4. Statistical analysis to test hypotheses and identify correlations.
5. Machine learning models for predictive analysis or clustering.
6. Compilation of findings into a comprehensive report and presentation.

**Source of Data:**

The dataset was obtained from the Chicago Data Portal provided by the Chicago Police Department. It is readily available for download and analysis purposes.

**Dataset Size:**

1.76 GB comprising 8.03 million rows and 22 columns, providing substantial data for analysis.

**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 contains relevant fields essential for extracting insights.

**List of Fields:**

* ID
* Case Number
* Date
* Block
* IUCR
* Primary Type
* Location Description
* Arrest
* Domestic
* District
* Community Area
* FBI Code
* Year
* Updated On
* Location

**Tools and Technologies:**

* Python libraries such as Pandas and NumPy for data analysis.
* Matplotlib and Seaborn for data visualization.
* Google Colab for project evaluation.
* Microsoft Word and PowerPoint for report compilation and presentation.

**Analysis of Results:**

The analysis provided insights into significant fluctuations in crime rates over the years, prevalent crime categories, arrest rates for different crime types, geographical distribution of crimes, and relationships between variables such as community area and arrest rates. These insights can guide law enforcement agencies in formulating data-driven policies and interventions to address crime challenges effectively.

**Queries and EDA:**

First, we imported the necessary libraries, mounted the drive and connected the drive with the ipynb file, by which we executed our dataset.

We used the .describe() and .head() functions to describe the dataset.

Query 1: We did the query to get the “Total Number of Incidents by Year,” we observed that the highest number of incidents occurred in 2002, which is 486807.

Query 2: We did the query to get the “Number of Arrests vs. Non-Arrests,”.

Query 3: This query was about “Distribution of Crime Types,” and we see the highest incident is THEFT, which is 1614940.

Query 4: “Arrest Rate by Crime Type”, where the arrest rate is highest for DOMESTIC VIOLENCE, which is 100%.

Query 5: “Most Common Crime Location by Year” where the most common crime is held in STREET in every year.

Query 6: “Crime Trend Analysis,” where we analyze the incidents that occur over the years and the percentage of the changes by years.

Query 7: “Weekly pattern for robbery incidents,” Where we first converted the “Date” column and then Filtered the data for the crime type "Robbery." Later, we will create a new column to extract the day of the week. Then, group the data by day of the week, count the number of incidents and sort the data by the day of the week. Finally, we visualize the weekly pattern for robbery incidents.

Query 8: In this query, we extract the Top 5 Community Areas with the Highest Domestic Violence Incidents.

Query 9: We finally extract the Hotspot Zones for Narcotics Crimes.

EDA: For EDA (Exploratory Data Analysis) representation, we first import some libraries such as Matplotlib and Seaborn to visualize the data analysis.

* 1. In this analysis, we overview the dataset and its schema.
  2. In this analysis, we visualize the Crime Trends Over the Years.
  3. In this analysis, we visualize Crime Distribution by Type and observe that theft is the most significant crime.
  4. In this analysis, we visualize the Arrest Rate Analysis. Here, the arrest rate is highest for domestic violence.
  5. In this analysis, we visualize Crime Distribution by Location. We observe that in the street, the crimes held most.
  6. In this analysis, we first Filter the dataset for narcotics-related crimes and Group by Year and calculate the mean arrest rate. Then, we Calculate the correlation between the Year and Arrest Rate for Narcotics Crimes
  7. Scatter plot of Arrest Rate for Narcotics crimes Over the Years
  8. In this analysis, we again Filter the dataset for assault-related crimes and Group by Community Area and calculate the mean arrest rate. Then, we Calculate the Correlation between Community Area and Arrest Rate for Assault Crimes.
  9. Scatter plor for Arrest Rate for Assault Crimes by Community Area.

**Conclusion:**

The comprehensive analysis of the Chicago crime dataset has furnished valuable insights and actionable recommendations for enhancing law enforcement strategies and fostering safer communities. The findings are a blueprint for crafting data-driven policies, interventions, and collaborative initiatives to fortify public safety and reduce crime rates across Chicago.

**SourseCode:**

The original file located:

https://colab.research.google.com/drive/1OQ5lB2hA5lbHJlbfNgYXPGCRcBPsk0Fx?authuser=2#scrollTo=lRzQap0F646g

**References:**

1. Crime in Chicago 2001-present

<https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2/about_data>

1. Quick EDA using Pandas Plots

<https://jyoti05iitd.medium.com/quick-eda-using-pandas-plots-407fa4d8ecf0>

1. Python Pandas data examples

<https://github.com/naomifridman/learn_python_pandas?tab=readme-ov-file>

1. Chicago Crime EDA: SQL and Pandas

<https://www.kaggle.com/code/boldy717/chicago-crime-eda-sql-and-pandas>

1. Chicago Crime Data Cleaning and EDA

<https://medium.com/@ahsanzafar222/chicago-crime-data-cleaning-and-eda-a744c687a291>