**Discipline: Big Data in Law Enforcement**

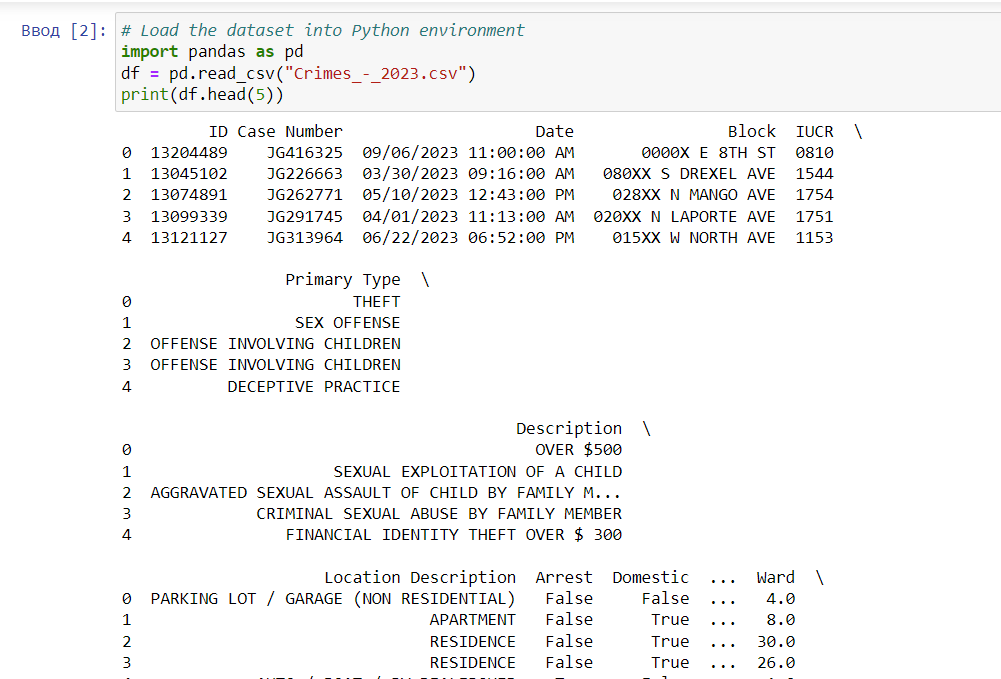
**Endterm. Report of a Project**

Berikova Malika, Kurmangazynova Asiya

The objective of this project is to analyze a real dataset related to law enforcement using Numpy, Pandas, SQL, and Apache PySpark. For this analysis, we have chosen the “Crimes – 2023” dataset of Chicago crimes (<https://data.cityofchicago.org/Public-Safety/Crimes-2023/xguy-4ndq/about_data>), which provides comprehensive information about various crime incidents reported in Chicago over the years. The goal of this analysis is to gain insights into crime trends, identify high-crime areas, understand demographic patterns, and provide valuable information for law enforcement agencies to enhance public safety measures.

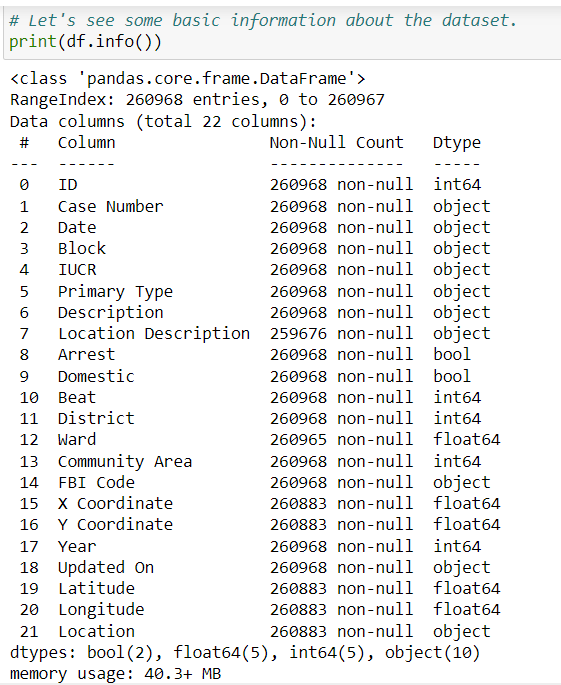
***Data Preparation:***

Load Dataset:

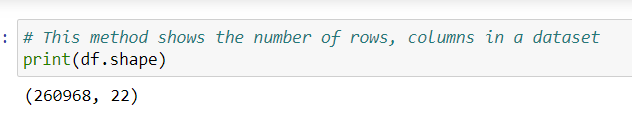


The dataset has been successfully loaded into the Python environment, and initial data cleaning and preprocessing have been performed to ensure data quality and consistency using Pandas and Numpy.

Now let’s see some basic information about the dataset with these functions:



This method provides a concise summary of the DataFrame, including the column names, data types, non-null counts, and memory usage. It's useful for quickly understanding the structure of the dataset and identifying any missing values.



This attribute returns a tuple representing the dimensions of the DataFrame, i.e., the number of rows and columns.

Изображение выглядит как текст, снимок экрана, Шрифт, число

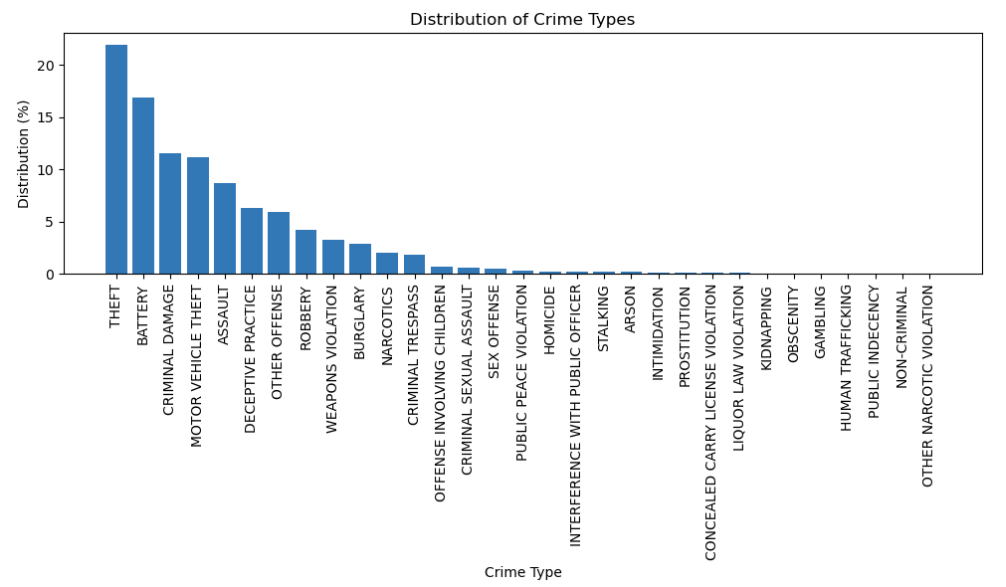
Автоматически созданное описание

This method generates descriptive statistics for numerical columns in the DataFrame, such as count, mean, standard deviation, minimum, maximum, and quartiles. It's helpful for understanding the distribution of numerical data.

***Data Analysis:***

1. *Crime Type Analysis:*





Here we used the **SQLite** to do the analysis, also the **pandas** and **matplotlib.**

The bar chart illustrates the distribution of different crime types, with "Theft" being the most common crime type, followed by "Battery" and "Criminal Damage." The Crime Type Analysis provides valuable insights into the distribution and frequency of different types of crimes occurring within a specific area or jurisdiction, it serves as a crucial tool for understanding the nature and extent of crime in a community, guiding efforts to prevent and address crime effectively, and ultimately contributing to improved public safety and well-being.

1. *Location-Based Analysis:*

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Implementation of this code gives us insights into the distribution of crimes across different location descriptions. It helps identify the most common locations where crimes occur, allowing law enforcement agencies to focus their resources and efforts on areas with higher crime rates.

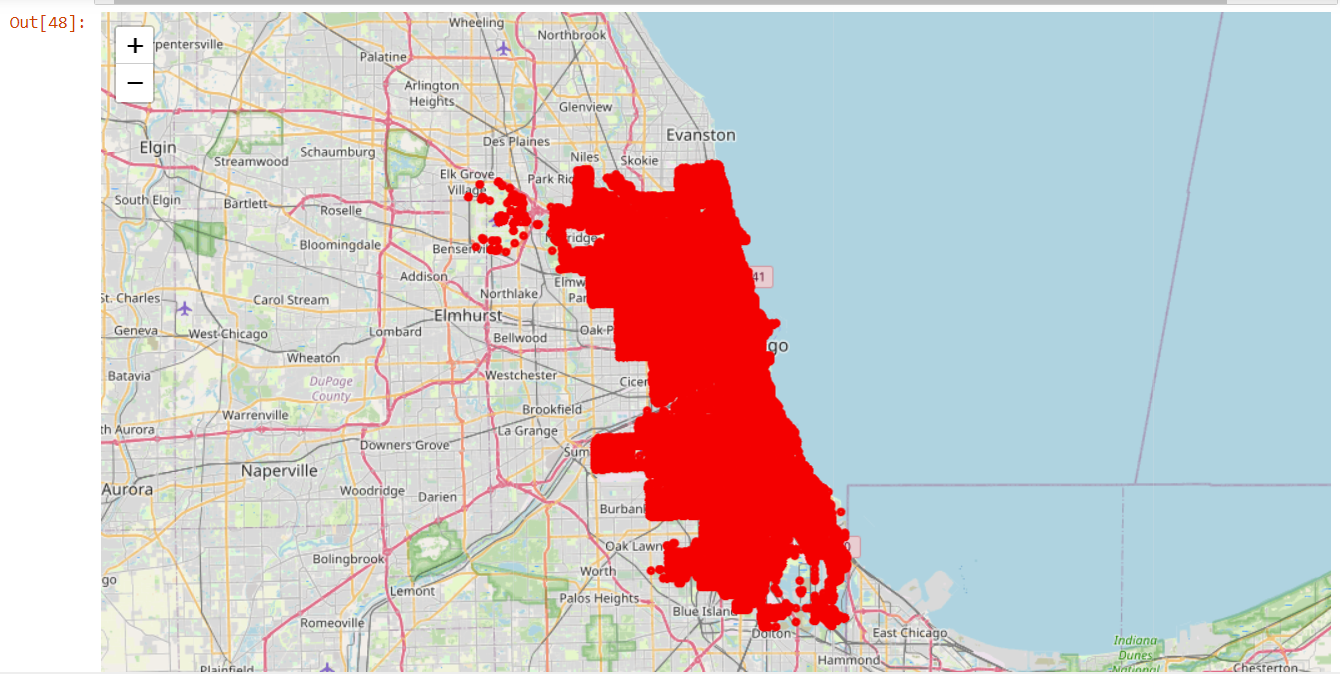
This analysis can also inform city planning and community safety initiatives by highlighting areas that may require additional attention or intervention.

1. *Spatial Analysis of Crime Hotspots:*

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

This code leverages Apache Spark for data loading and preprocessing and utilizes Folium for spatial visualization, allowing for efficient analysis and visualization of crime hotspots in the dataset.



The Spatial Analysis of Crime Hotspots provides valuable insights into the geographical distribution and concentration of crime within a specific area or jurisdiction. It is critical for understanding the spatial distribution of crime, guiding targeted interventions and resource allocation, and fostering collaboration between law enforcement, policymakers, urban planners, and community stakeholders to create safer and more resilient communities.

***Conclusion:***

In conclusion, the analyses conducted on the crime dataset “Crimes\_–\_2023.csv” for Chicago in 2023 have yielded valuable insights into various aspects of crime within the city. The Crime Type Analysis revealed the distribution and frequency of different types of crimes, aiding in prioritizing resources for crime prevention. The Location-Based Analysis provided insights into the spatial patterns of criminal activity, guiding targeted interventions in high-crime areas. The Spatial Analysis of Crime Hotspots identified specific geographic areas with high concentrations of criminal activity, informing strategies to improve public safety. Overall, these analyses contribute to a better understanding of crime trends and support evidence-based decision-making for crime prevention and intervention efforts in Chicago.

***The GitHub link:***

<https://github.com/molik-molik/Endterm_Big_Data>