**Property Crime Database**

# **Introduction**

In this data engineering project, the goal is to build a comprehensive data analytics table by merging crime data, housing dataset, and climate data of Chicago. The project focuses on the city of Chicago, utilizing various datasets to conduct a thorough analysis. The primary objectives include understanding the relationship between crime rates, housing prices, and climate conditions.

**Project Goal**

The central aim of this project is to create a robust data analytics table by combining crime data, housing dataset, and climate data by state. This integration enables a holistic analysis, shedding light on potential correlations and patterns among these diverse datasets.

**Data Sources**

To achieve the project goal, five distinct datasets were employed:

Zillow Dataset for Housing Prices:

The Zillow dataset provides housing prices for neighborhoods in Chicago. The Zillow Home Value Index (ZHVI) for All Homes is utilized, offering insights into the median house prices in specific areas.

Chicago Crimes Dataset:

Publicly available on the official website of the City of Chicago, this dataset contains information about various crimes. The analysis considers crime counts, allowing for an exploration of crime trends in different neighborhoods.

2010 Census Data:

To normalize crime counts based on neighborhood populations, the 2010 Census data at the neighborhood level is employed. This ensures a fair analysis, accounting for population differences across neighborhoods.

GlobalLandTemperaturesByMajorCity:

Downloaded this data for MajorCity of the world as globally and then extracted Chicago data from it.

Neighborhood Boundary Data:

Latitude-longitude coordinates in the crime dataset are mapped to Zillow-defined neighborhoods using neighborhood boundary data obtained from a GitHub repository. This mapping facilitates a more granular analysis, connecting crime incidents to specific neighborhoods.

**Ethical Considerations**

It's crucial to note ethical considerations in using the datasets. The Zillow dataset reports the Zillow Home Value Index (ZHVI) for All Homes, ensuring privacy by not disclosing individual property prices or details about homeowners or buyers. Similarly, the crime data protects privacy by excluding personal identifiers such as names, ages, genders, races, or religions of both criminals and victims.

**Analysis**

The analysis phase explores the relationship between crime rates, housing prices, and climate conditions.

Crime Rates and Housing Prices:

Investigate whether there is any correlation between crime rates and housing prices in different neighborhoods.

Explore patterns and outliers that may indicate areas with unique characteristics.

Climate Impact on Crime Rates:

Analyze the influence of climate data on crime rates.

Identify potential trends or seasonality in crime incidents based on climate conditions.

**Data Engineering Process**

The project involves creating a database to store CSV files, leveraging both SQL and Spark for data engineering tasks. The provided Python code illustrates the steps for creating an SQLite database and tables for the crime and climate datasets. The code uses the pandas library for data manipulation and SQLite for database management.

The Jupyter notebook containing the data preparation steps is linked for reference, providing transparency into the project's workflow.

**Conclusion**

This data engineering project serves as a foundation for further data analysis, offering insights into the intricate relationships between crime rates, housing prices, and climate conditions. By ethically handling and integrating diverse datasets, the project ensures privacy while enabling a nuanced exploration of patterns and correlations that contribute to a better understanding of the factors influencing urban dynamics in Chicago

**Data link:**  
<https://www.zillow.com/research/data/>

<https://data.cityofchicago.org/Public-Safety/Crimes-2019/w98m-zvie>

<https://datahub.cmap.illinois.gov/dataset/community-data-snapshots-raw-data/resource/8c4e096e-c90c-4bef-9cf1-9028d094296e?inner_span=True>

https://www.zillow.com/research/zhvi-methodology-2019-highlights-26221/