Chicago Crime & Max Temp
Data Mining

CSPB 4502
Team 1 :
Patrick Ridley
Shrey Shah

Chicago Crime and Max Temp - Overview & Objectives

Project Overview:

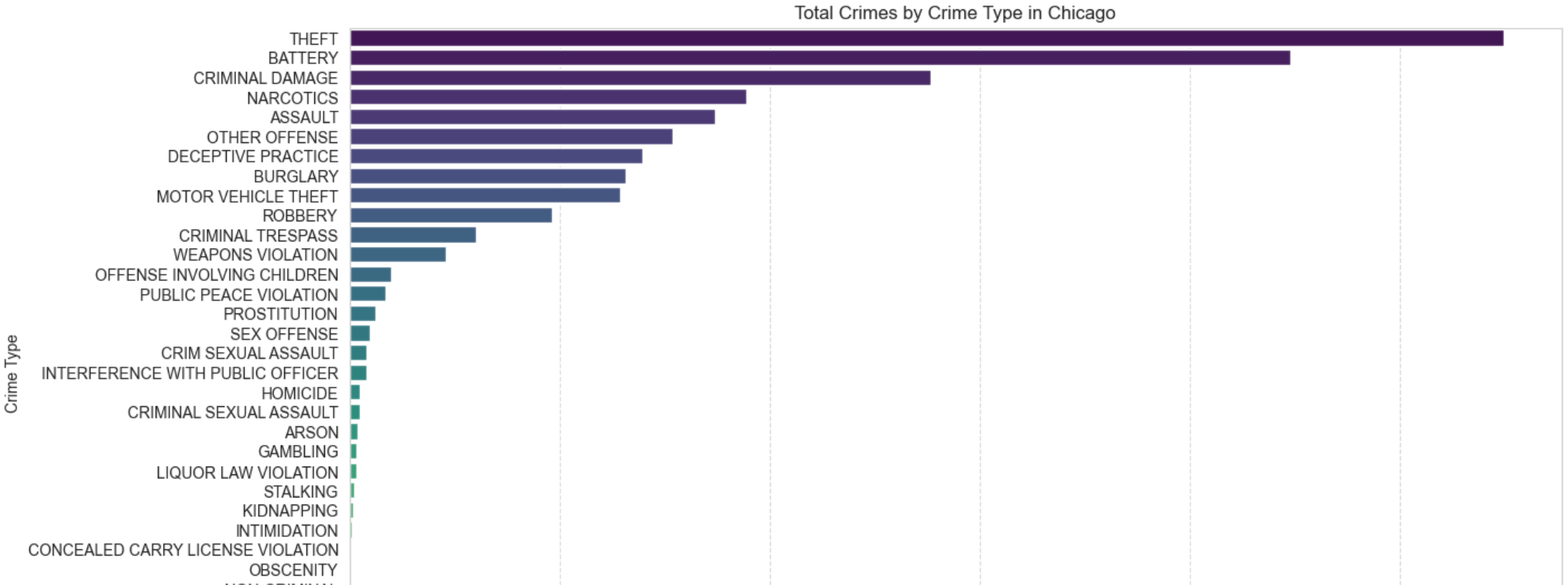
- Analyzing the relationship between temperature variations and violent crime occurrences in Chicago (2010–2025).
- Leveraging historical crime and weather datasets to identify patterns and trends.

Primary Objectives:

- Identify how temperature fluctuations influence violent crime rates, specifically assaults, batteries, homicides, and sex-related offenses.
- Explore seasonal crime patterns and determine how weather conditions correlate with crime surges or reductions.
- Provide actionable insights for crime prevention and public safety improvement.

Chicago Crime and Max Temp - Importance for Policy & Community

- **Why This Matters:**
  - Enables law enforcement to strategically allocate resources based on predictable temperature-related crime spikes.
  - Assists urban planners and policymakers in designing community safety initiatives and interventions tailored to seasonal crime patterns.
  - Improves emergency response preparedness during identified high-risk temperature periods (moderate warmth, first warm days following colder seasons).
  - Enhances public awareness regarding the environmental impact on community safety, promoting proactive community engagement and preventive measures.

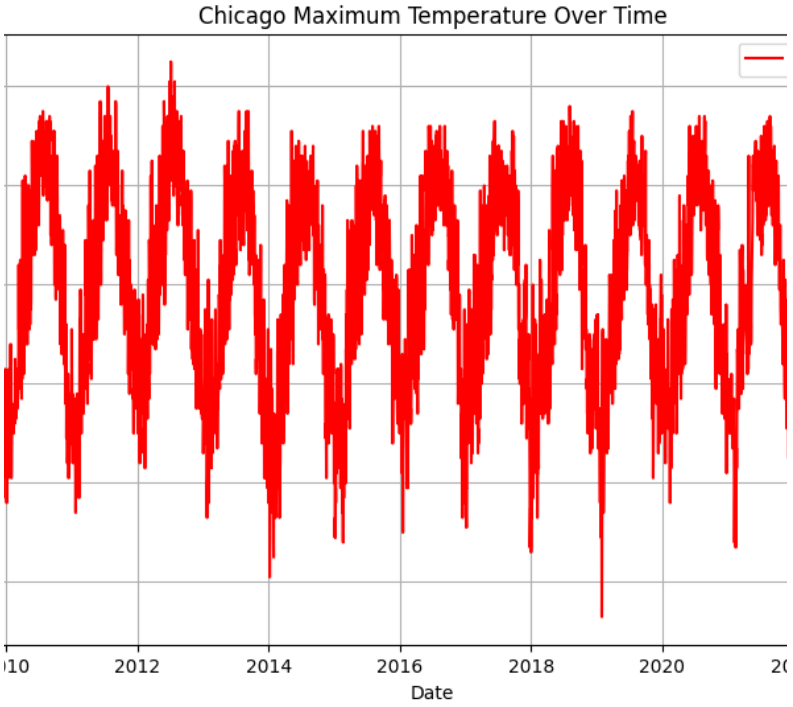


Crime/Max Temp Data Sources:

- **Data Sources:**
  - Chicago Crime (2010–2025, 1 M+ incident records)
  - NOAA Weather (daily TMAX, TMIN)

Crime/Max Temp - Data

- **Preprocessing:**
  - **Crime:**
    - Imported into MySQL; filtered to violent offenses only (Battery, Assault, Homicide, Sex-related)
    - Converted Arrest from text→Boolean; standardized date to adjDate (YYYY-MM-DD)
  - **Weather:**
    - Cleaned in pandas: dropped rows missing TMAX/TMIN; exported cleaned CSV for import



Crime/Max Temp - Data

- **Integration & Warehousing:**
- Aggregated daily counts by type in SQL
- **Joined** on weather_data.Date = chi_crime.adjDate → new table crime_weather_data
- Primary index on Date; secondary indexes on Primary Type & severity for OLAP

```
Date  TMAX  TMIN  Crime_Count  Assault_Count
Battery_Count  \
0  2008-01-01   31    10        1832           63          303
1  2008-01-02   26     6         919           44          112
2  2008-01-03   25     8         916           33          126
3  2008-01-04   37    25        1099           57          154
4  2008-01-05   45    37        1183           61          191

   Homicide_Count  Sex_Crime_Count
0               2              199
1               1               13
2               1               10
3               1               16
4               2               15
/var/folders/gl/c9dr3r7j047d6wnxp52m68wr0000gn/T/ipyke
rnel_24897/2305868975.py:27: UserWarning: pandas only
supports SQLAlchemy connectable (engine/connection) or
database string URI or sqlite3 DBAPI2 connection. Other
DBAPI2 objects are not tested. Please consider using
SQLAlchemy.
  df = pd.read_sql(query, conn)
```

Crime/Max Temp - Data

Data Mining Techniques:

Exploratory Data Analysis (EDA): crime-by-type bar charts; TMAX vs. total-crime time series
Clustering (K-Means): group days into low/moderate/high-temperature crime patterns
Association Rule Mining (Apriori, FP-Growth): find frequent "weather→crime" rules
Time-Series Forecasting (ARIMA, seasonal decomposition): model crime rate trends

Tools & Environment:

MySQL (Workbench), Python (pandas, NumPy, Matplotlib) in Jupyter Notebook

Visuals to include:

Schema & Query: outline of crime_weather_data table creation and LEFT JOIN SQL
Bar Chart: crime frequency by "Primary Type"
Time Series: daily TMAX vs. total violent-crime counts
Smoothed Trend: inverted-U relationship of TMAX bins vs. assaults
Heatmap: Pearson correlation between TMAX/TMIN and each crime count

Chicago Crime / TMAX Key Finding

Crime Trends with Weather:

Violent crime rates notably increase during periods of moderate warmth (approx. 68°F to 83°F).

A noticeable decline in crime rates is observed during extreme heat conditions (above 85°F).

Crime occurrences peak significantly on comfortable, warm days rather than on extremely hot or cold days.

Methodology:

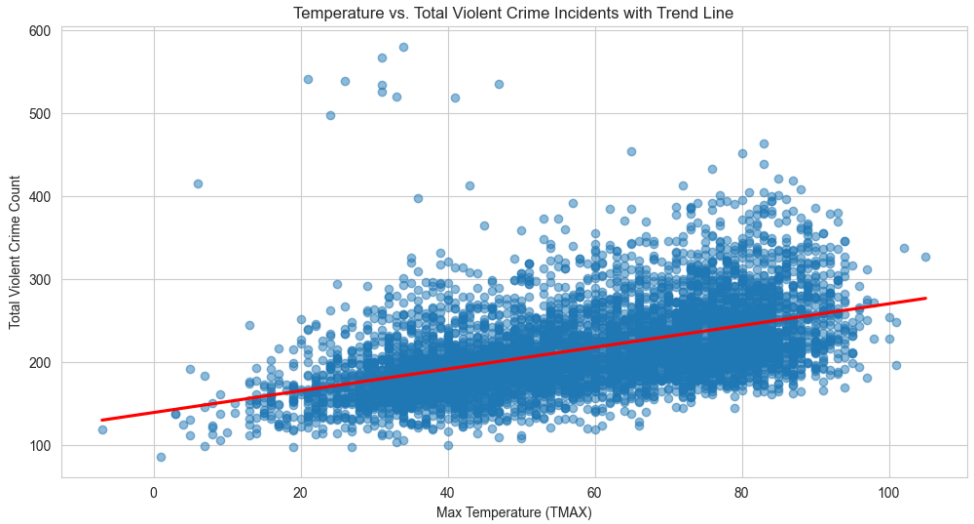
Data integrated crime counts and daily temperature TMAX using SQL joins.

Utilized correlation analysis and clustering techniques (K-Means) to identify distinct crime patterns linked to specific temperature ranges.

Validated findings through visualizations like bar graphs and heatmaps, confirming higher crime rates during comfortable weather conditions.

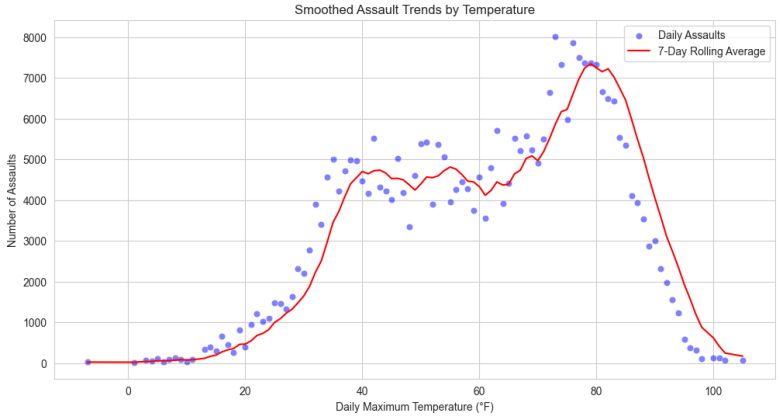
Trend Line Analysis

- **Crime Rate vs Temperature:**
  - The trend line analysis reveals a clear seasonal pattern.
  - Crime rates rise notably during warmer periods, with distinct peaks in the spring and early summer months.
  - Visual analysis supports the hypothesis that moderate weather conditions are conducive to higher crime rates.



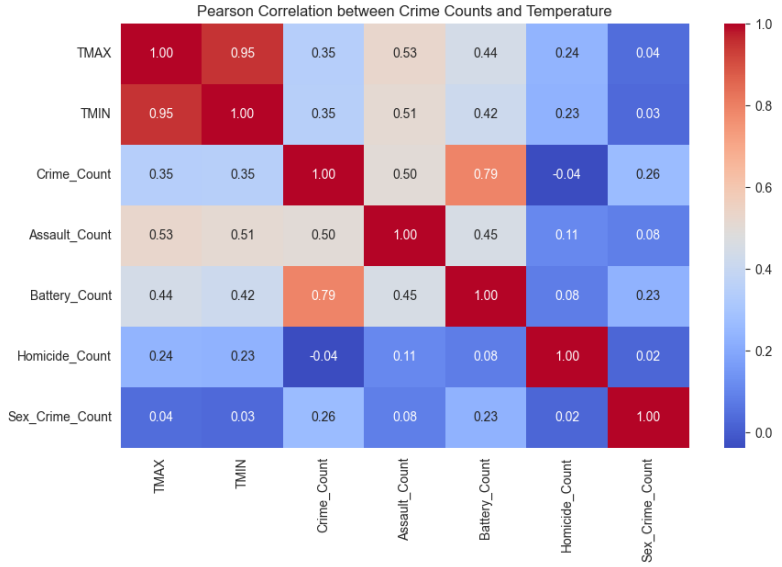
Smoothed Assault Trends by Temperature

- **Assault Trends:**
  - Assault incidents display a distinct inverted-U shape relationship with temperature.
  - Assaults peak within moderate temperature ranges (approximately 70°F to 80°F).
  - Incidents decrease noticeably during extreme heat or cold temperatures, likely due to reduced outdoor activity.



Pearson Correlation Analysis

- **Correlation Between Crime Counts and Temperature:**
- Assault and battery incidents show strong positive correlations with daily maximum temperatures (TMAX).
- Moderate correlations exist between total crime counts and temperature.
- Minimal or weak correlations found for homicides and sex-related crimes, indicating these crime types are less sensitive to temperature variations.
- Pearson heatmap visually confirms the strength and direction of these relationships, highlighting areas for targeted preventive measures.



Conclusion & Interpretation

Insights on Temperature and Crime:

While TMAX provides a useful measure, it does not directly drive crime. Rather, nicer, more comfortable weather conditions are associated with increased crime activity.

The correlation likely stems from increased social interactions and outdoor activity during pleasant weather conditions rather than the temperature itself.

Data Mining Interpretation:

Clustering analysis indicated higher crime groupings during moderate temperatures, likely reflecting periods of increased social interaction rather than direct temperature effects.

Association rule mining highlighted frequent co-occurrences between moderate temperatures and violent crimes, reinforcing the presence-of-people hypothesis.

Recommendations & Future Directions