City Crime Analysis Using Power BI

1. Project Overview and Objective

This project involves cleaning, transforming, and analysing raw data using **Excel** and creating an interactive **Power BI dashboard** to derive meaningful business insights.
The main objective is to demonstrate data pre-processing techniques using Excel and an interactive Power BI dashboard visualization to make informed decisions.

---

2. Data Sources

- **Source Description and Timeline:** Google dataset/2025.
- **Domain**: Environment

---

3. Problem Statement

- Crime data is large and complex, making it difficult for stakeholders to quickly understand overall crime trends and patterns across districts and time periods.
- There is a lack of a centralized and interactive system to identify high-crime districts, wards, and offense types for effective public safety planning.
- Decision-makers struggle to analyze how crime varies by time (year and shift), limiting their ability to deploy resources efficiently.
- Existing reports do not provide drill-down capabilities to explore detailed crime information from a high-level summary.
- Without clear visual insights and trend analysis, it is challenging to monitor changes in crime levels and identify emerging hotspots.

---

4. Attribute (Column /Features) Details:

Attribute Name	**Attribute Nam**	**Description**
CCN	Text / Integer	Central Complaint Number – unique identifier for each crime incident
REPORT_DAT	Date/Time	Date and time when the crime was reported
Report year	Integer	Year in which the crime was reported
SHIFT	Text	Time shift when the crime occurred

		(DAY, EVENING, MIDNIGHT)
OFFENSE	Text	Type of crime committed (e.g., Theft, Assault, Robbery)
METHOD	Text	Method used to commit the crime
DISTRICT	Text	Police district where the crime occurred
WARD	Integer / Text	Administrative ward of the crime location
PSA	Integer	Police Service Area responsible for the location
NEIGHBORHOOD_CLUSTER	Text	Grouped neighborhood classification
BLOCK	Text	Street block where the crime occurred
LATITUDE	Decimal Number	Latitude coordinate of the crime location
LONGITUDE	Decimal Number	Longitude coordinate of the crime location
XBLOCK	Decimal Number	X-coordinate for spatial mapping
YBLOCK	Decimal Number	Y-coordinate for spatial mapping

---

5. Tools & Technologies

- **Excel:** Data cleaning, transformation, and Pivot Tables.

- **Power BI:** Data modelling, DAX calculations, visualization, and interactive dashboard creation.

---

6. Data Pre-Processing (Excel / Power Query)

Tasks Performed:

- **Data Cleaning & Transformation:** Removed duplicates, handled missing values, standardized formats, and created calculated fields.
- **Filtering & Sorting:** Organized data to focus on relevant records.
- **Pivot Tables:** Generated Pivot Tables for data summarisation and initial insights.

---

7. Data Modelling and DAX (Power BI)

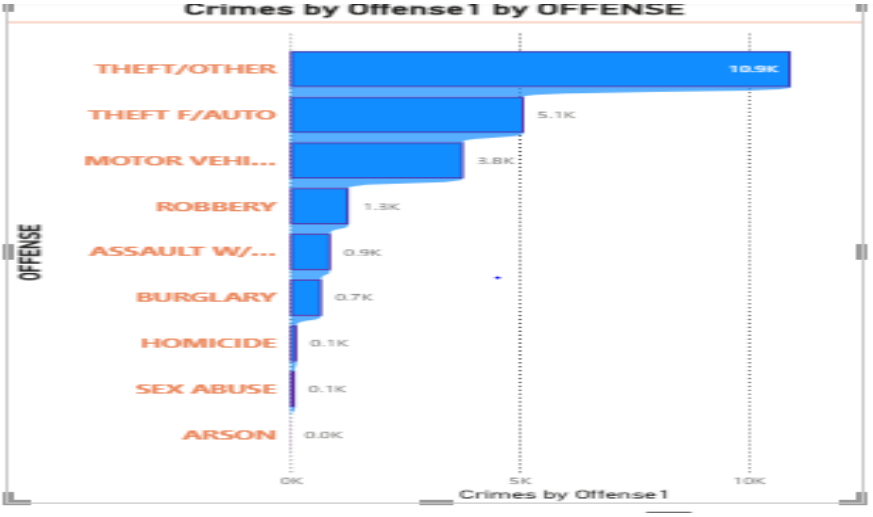
- **Calculated Columns & DAX Measures:** Implemented DAX formulas for key metrics, such as total Crimes, Crimes by District, and Crimes Trend by year.
  - Total Crimes
  - Crimes by Offense
  - Crimes by District
  - Crimes by shift
  - Crimes Trend by year
  - Crimes by Ward

---

8. Analysis and Visualizations (Power BI)

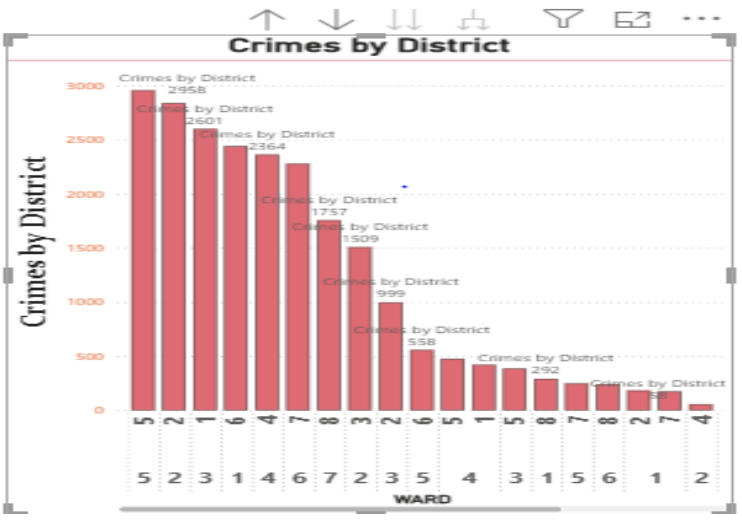
Dashboard Features:

- **Multiple Visualizations based on problem statement:** Bar charts, line charts, pie charts, cards, and tables to communicate insights.

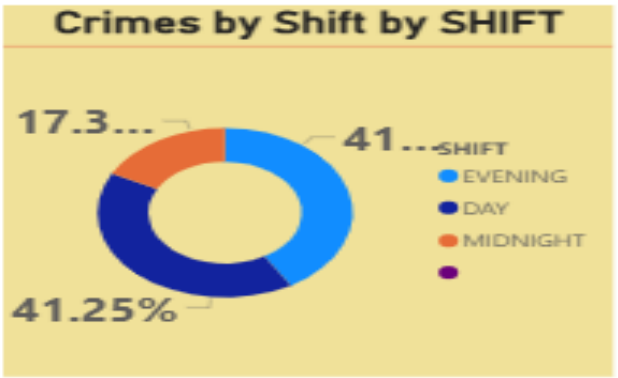


- **THEFT/OTHER** is the highest offense category with **~10.5K cases**, followed by **THEFT FROM AUTO (~5.1K)**.
- **THEFT FROM AUTO (5.1K)** and **MOTOR VEHICLE THEFT (3.8K)** together form a significant portion of offenses.

- Serious violent offenses like **ROBBERY (1.3K)**, **ASSAULT WITH WEAPON (0.9K)**, and **HOMICIDE (0.1K)** are comparatively low.
- **Homicide, Sex Abuse, and Arson** show very low counts, but their **social impact is extremely high.**



- The top **3 districts** record **~2,600–2,950 crimes each**, significantly higher than the rest.
- After the top **5 districts**, crime counts fall rapidly from **~1,500 to below 1,000**.
- The lowest districts report **fewer than 200 crimes**.
- Since crime volume varies greatly by district, **resource allocation should be district-specific** rather than uniform.
- Focusing enforcement and prevention strategies on **high-crime districts** could significantly reduce overall crime.

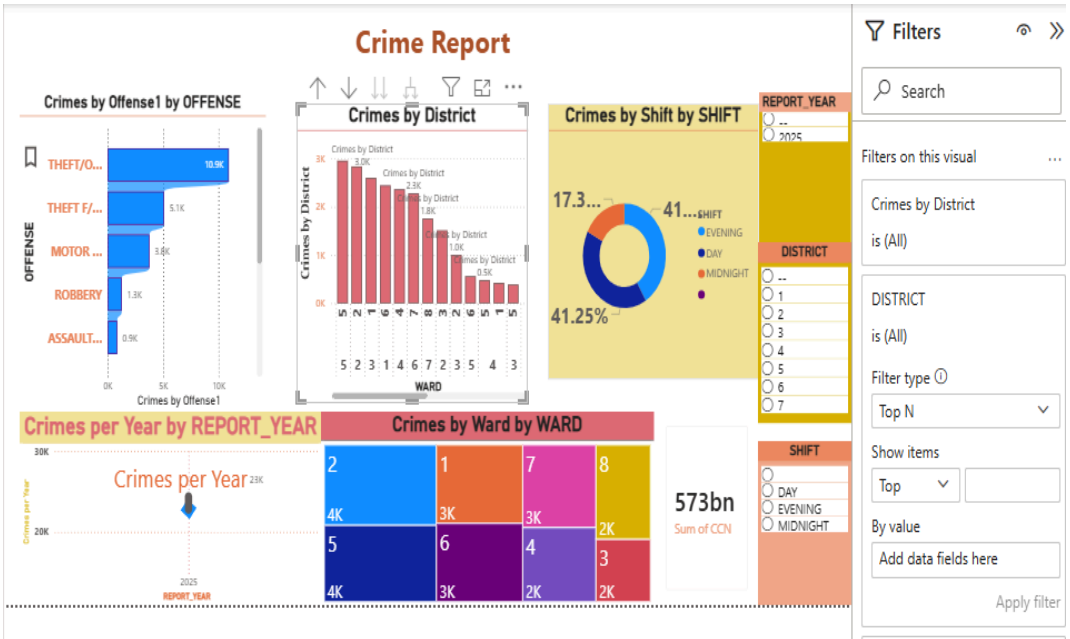


- The **Evening shift accounts for ~41% of total crimes**, making it the most crime-prone time period.
- The **Evening shift accounts for ~41% of total crimes**, making it the most crime-prone time period.
- The **Day shift also contributes ~41%**, nearly equal to the evening shift.
- Since **over 80% of crimes occur during Day and Evening shifts**, law enforcement and preventive measures should prioritize these periods.
- Strategic staffing during these shifts could yield the **maximum impact on crime reduction**.

- The dataset shows **approximately 23K crimes in the reported year (2025)**.
- Viewing crime data by year helps identify **macro-level patterns** rather than short-term fluctuations.
- The visualization indicates crime records are **heavily concentrated in one reporting year**, suggesting the dataset may represent **recent or current crime conditions**.
- Using **REPORT_YEAR as a slicer or drill-down level** enables users to analyze **districts, offenses, and shifts for that specific year**.
- This allows stakeholders to pinpoint **where and when crimes are increasing** within the year.

- **Wards 2 and 5** show the **highest crime counts (~4K each)**.
- **Wards 1, 6, and 7** each report **~3K crimes**.
- **Wards 3, 4, and 8** have **~2K crimes each**, which is lower comparatively but still significant.
- This suggests **crime is present across all wards**, not limited to just high-crime areas.
- The clear variation in crime across wards supports **ward-specific policing strategies**.
- Focusing interventions on **top wards (2 and 5)** can deliver the **highest impact on overall crime reduction**.

Consolidated Report /Dashboard.



9. Insights & Conclusions

Descriptive Analytics:

- **Purpose:** Summarizes historical crime data.
- The dataset is dominated by theft-related crimes, while violent crimes occur very rarely.
- Crime is unevenly distributed across districts, with a few districts showing much higher crime levels.
- Most crimes occur during the day and evening shifts, with the lowest crime during midnight hours.
- Overall crime volume is high, and certain wards experience significantly more crimes than others.

Diagnostic Analytics:

Purpose: Explains reasons behind crime patterns.

- High crime in certain districts may be due to **population density, commercial areas, or vehicle traffic**.
- Higher crime during **Day and Evening shifts** reflects **increased public activity**.
- Vehicle theft dominance indicates **parking and transit areas as vulnerable zones**.

Predictive Analytics:

Purpose: Forecasts future crime trends based on patterns.

- Districts and wards with historically high crime are **likely to remain hotspots**.
- Theft and vehicle-related crimes are expected to **continue dominating future incidents**.
- Day and Evening shifts may continue to record **higher crime volumes**.

Prescriptive Analytics:

Purpose: Recommends actions to reduce crime.

- Increase **patrol and surveillance** in top crime districts and wards.
- Strengthen **vehicle security measures** in high-risk areas.
- Allocate more staff during **Day and Evening shifts**.
- Implement **ward-specific and offense-specific prevention programs**.

---

10. Conclusions

- The analysis reveals that crime is predominantly driven by theft and vehicle-related offenses, with incidents concentrated in specific districts and wards. Most crimes occur during Day and Evening shifts, indicating higher risk during active public hours. These insights enable targeted, data-driven strategies for effective crime prevention and resource allocation.