**KARACHI CRIME FILES PROJECT REPORT**

# Introduction

This project is an end-to-end analysis of crime trends in Karachi, based on a dataset obtained from Kaggle.

The primary objective was to extract meaningful insights from raw, unstructured data and present them visually using Excel dashboards.

The project demonstrates practical application of Power Query, Pivot Tables, and various Excel charting tools to clean, analyse, and visualize complex data for informed decision-making.

# 1. Data Collection

The original dataset was sourced from Kaggle and contained multiple years worth of reported crimes in Karachi.

It included crime categories, district-wise data, age group segmentation, and educational background of individuals involved.

However, the dataset was poorly structured with missing values, inconsistent headers, extra formatting artifacts, and misaligned data entries which made it unsuitable for direct analysis.

# 2. Data Cleaning and Preparation

All data cleaning was carried out using Power Query in Excel (Mac). The following major steps were performed:

* Column Cleanup: Removed irrelevant and empty columns, standardized column names for easy reference.
* Row Filtering: Removed blank and duplicate rows.
* Data Type Correction: Ensured all numeric columns (e.g., crime counts) were in the correct number format, and date/time fields (if any) were properly formatted.
* Handling Missing Values: Null or empty values were filled with defaults like 0 or Unknown, depending on the column context.
* Unpivoting & Structuring: Transformed cross-tabbed data into tabular form using "Unpivot Columns", enabling easier aggregation and charting.
* Merging & Appending: Combined multiple data tables where necessary (e.g., demographic + crime type) using Power Query Append and Merge functions.

Once cleaned, the data was loaded back into Excel for further analysis using Pivot Tables.

# 3. Analysis and Pivot Table Creation

Several Pivot Tables were created to answer specific analytical questions:

* Monthly Trends: To analyse how crime volume fluctuated month-over-month.
* Area Comparison: District-wise comparison to identify zones with the highest or lowest crime rates.
* Most Reported Crimes: Aggregated data to find the top crime types reported in each district.
* Age vs Crime: Identified which age groups were most affected by which types of crime.
* Education vs Crime: Analysed how education level (Matric vs Graduate) correlated with different crime categories.

These Pivot Tables served as the backend for dynamic chart visualizations.

# 4. Dashboard Development

The visual dashboard was built on top of the Pivot Tables using multiple types of Excel charts:

* Line Charts: For month-over-month crime trends.
* Clustered Column Charts: To compare crime by district and crime type.
* Pie Charts / Doughnut Charts: To show proportional distribution of education level vs. crime types.
* Heatmaps: To highlight concentration of crimes across areas and categories.
* Bar Charts: For age group vs. crime type analysis.

Field buttons were enabled for interactivity, and slicers were added (where applicable) for filtering by year, area, or crime type.

# 5. Key Insights

* Digital Fraud is one of the most reported crimes, especially among graduates.
* Nazimabad, Landhi, and Other districts reported higher crime volumes.
* Individuals aged 2540 years are most frequently involved in or affected by crimes.
* Graduates show higher involvement in cyber-related offenses, while Matric-educated individuals are more connected to traditional crimes like robbery or pickpocketing.
* Some regions (e.g., Clifton, Defence) remain stable, while others show volatile patterns.**6. Challenges Faced**
* Limited field button functionality on Excel for Mac made dynamic interactivity slightly restricted.- Cleaning such a messy dataset required iterative use of Power Query steps like Fill Down, Replace Errors, and Unpivot.
* Geographic categorization was inconsistent (e.g., many records Other

# 7. Recommendations

* Digital literacy programs to reduce cybercrime.
* Awareness campaigns targeting high-crime age groups.
* Improved data entry standards in crime databases to minimize Other classifications.
* More structured data release practices on public platforms like Kaggle for better usability.

# Conclusion

This project illustrates the full cycle of data-driven crime analysing from unstructured raw data,

through robust cleaning using Power Query, analysis using Pivot Tables, and ending in a visually engaging dashboard.

It demonstrates the power of Excel (even on Mac) for transforming messy datasets into actionable insights.