San Francisco Crime Data Analysis with SQL & BigQuery

Project Overview

This project focused on analyzing crime incidents in San Francisco using SQL queries on a public dataset stored in Google BigQuery. The analysis was conducted in a Databricks notebook environment, combining SQL queries with markdown explanations.

Tools and Technologies Used

- Databricks (SQL notebook)
- Google BigQuery (Public Dataset)
- SQL (for querying and data exploration)
- Markdown (for documenting and interpreting results)

Dataset Description

The dataset used includes detailed records of crime incidents reported in San Francisco from 1936 onward. Key fields include the date and time of the incident, police district, category and type of crime, and resolution status.

Key Analyses Performed

- Explored crime distribution by year to understand long-term trends
- Analyzed most common crime types across different neighborhoods
- Identified the districts with the highest crime rates (e.g., Tenderloin, Southern)
- Evaluated crime resolution status to understand which types are more frequently solved

Main Insights

- The Tenderloin and Southern districts consistently had the highest crime reports
- Common offenses included Larceny/Theft, Assault, and Drug/Narcotic violations
- A large portion of crimes were unresolved, with many marked as 'NONE' under resolution
- Crime rates showed fluctuations across decades, revealing distinct patterns over time

Conclusion

The project demonstrated the use of SQL for large-scale data exploration and analysis. It provided meaningful insights into the spatial and temporal patterns of crime in San Francisco. Further work could include predictive modeling or integrating additional datasets for deeper analysis.