One-Pager: Analysis of NYPD Crime-Related Call Data

## Introduction

The dataset used in this analysis is derived from the New York Police Department (NYPD) and contains detailed records of various calls for service throughout New York City. Each call record provides multiple attributes, including the **incident date and time**, **precinct involved,** a brief **description of the issue** and **geographic coordinates** among others. The dataset consists of around **37.9 million records** and spans from the year 2019-till date, capturing both **crime** and **non-crime** incidents. This analysis focuses on isolating crime-related activities to explore the distribution and trends in the dataset.

We are focusing on data from 2019 to 2023 to analyze trends before and after significant events such as the COVID-19 pandemic. In 2019, the number of arrests was lower, whereas post-2020 the trends started to change. Hence, this timeframe will help us to understand the reason behind increasing arrests based on the 911 calls where NYPD were involved.

## Data Cleaning

During the data cleaning phase, several issues were identified that required attention. Notably:

* **Missing Values**: The OBJECTID column present in the historic data had 82% null values, hence it was removed. The **Borough** column had records with null values, which were subsequently removed for clarity in geographical analysis. Other columns such as **arrival times** and **closing times** had missing values, but they were not removed as they were not essential for the analysis at this stage.
* **Incorrect Time Data**: Some call records had incorrect timestamps, notably 1899-12-30T00:00:00.000 for incident times. This value was corrected by replacing it with 00:00:00 where necessary to maintain consistency when analyzing the time of incidents.
* **Outliers**: While the overall dataset is massive, outliers in the form of extreme or suspiciously low/high call volumes in some months were inspected. These anomalies were retained to highlight real trends such as seasonal spikes in crime rates.
* **Inconsistency in Data**: The Year-to-date (YTD) data for 2024 had 18 columns and the historic data from 2019 had 20 columns. The data was made consistent by removing OBJECTID column which had a lot of null values and Location column which has redundant information. Post this step the YTD and historic data were merged.

## Interesting Insights from Exploratory Data Analysis (EDA)

The EDA process aimed to uncover key insights regarding crime-related calls across time and geographical locations. Here are the major steps and findings:

1. **Crime vs. Non-Crime Calls**: A critical first step was determining the proportion of calls that were crime related. It was found that **39.26% of all calls** were related to crime, which amounted to nearly **14.9 million records**. The remaining 60.74% were related to non-crime activities such as civil matters, medical emergencies, or general assistance.
2. **Analysis by Borough**: Calls were distributed across New York’s five boroughs: **Brooklyn**, **Bronx**, **Manhattan**, **Queens**, and **Staten Island**. Brooklyn had the highest number of crime-related calls, followed closely by Bronx and Manhattan. This geographic distribution provides insight into where crime rates are highest across the city.
3. **Trend Analysis**: Time series analysis was performed to study the trends of crime-related calls on **monthly** and **yearly** levels:
   * **Monthly Trends**: Calls tended to peak in the summer months, particularly around **June and July**, a trend consistent with historical crime patterns. There was a notable dip in calls during colder months.
   * **Yearly Trends**: The data spanned multiple years, with fluctuations in the volume of calls year over year. An overall increase in crime-related calls was observed over the analyzed period, with some annual declines that could potentially align with law enforcement efforts or external factors like the COVID-19 pandemic.
4. **Incident Types**: Within crime-related calls, common types included **disputes**, **larceny**, and **investigations of possible crimes**. An in-depth look into the "Main Type" and "Sub Type" of calls provided a breakdown of which criminal activities were most frequently reported.

The data didn’t have any numerical columns. Hence the numerical analysis of the dataset was limited. Going further when the data is transformed to conduct complex analysis, numerical columns will be derived.

A graph and diagram of a graph

Description automatically generated with medium confidence

**Figure 1: Some Interesting insights from EDA**

## Conclusion

This analysis of NYPD call data, particularly crime-related incidents, provides valuable insights into the distribution and nature of criminal activity across New York City. By isolating crime-related calls and analyzing trends over time and space, we identified key patterns, such as seasonal and geographic crime spikes, which can be useful for future crime prevention efforts. Although some challenges with missing and incorrect data were encountered, these were addressed effectively, allowing for a clearer understanding of the dataset's structure and the underlying trends it revealed.