Statistical/Hypothetical Question

In this discussion, the primary question that was explored in this analysis was: What are the relationships between incident locations, times, and their impact on police district activity? Specifically, correlations between spatial (e.g., MapX, MapY) and temporal (TimeCreate, TimeClosed) variables were examined in an attempt to identify any significant patterns, and evaluate the role of SelfInitiated incidents in district-level distributions.

Outcome of the EDA

From the exploratory data analysis (EDA), several key insights emerged:

1. **Spatial and Temporal Relationships**:
    - Weak linear correlations were found between spatial variables (MapX and MapY), suggesting little relationship between the X-Y coordinates.
    - Strong positive correlations between TimeCreate and TimeClosed (Pearson correlation = 1.00) indicated a near-linear dependency due to their temporal progression.
2. **District-Level Distributions**:
    - The probability mass functions (PMFs) showed uneven distributions of incidents across police districts, with some districts experiencing higher proportions of incidents.
3. **Regression Analysis**:
    - A linear regression model on TimeClosed (dependent) against TimeCreate (independent) explained 100% of the variance ($R^2 = 1.00$). This reflects the direct relationship between incident creation and closure times.

Missed Opportunities

The analysis primarily focused on spatial and temporal variables. While the distributions of incidents across districts and the relationships between variables were explored, **interaction effects** and **non-linear relationships** may have been overlooked. For instance, time-of-day trends or geographical clustering could provide deeper insights.

Variables That Could Have Helped

The analysis could have benefited from:

- **Incident severity or priority levels**: Understanding the gravity of incidents would contextualize resource needs.

- **Weather or event data**: External factors could explain spikes in incidents in specific districts.

Questionable Assumptions

1. **Normality Assumption**:
    - Fitting normal and exponential distributions to spatial data (e.g., MapX) assumed theoretical alignment, which was not well-supported due to heavy skewness and outliers.
2. **Linearity Assumption**:
    - Correlation and regression analyses assumed linearity, potentially ignoring non-linear relationships or interactions.

Conclusion

While the EDA provided foundational insights into spatial and temporal incident data, further investigation with additional contextual variables and more flexible modeling approaches (e.g., clustering or non-linear regression) could yield deeper and more actionable insights.