Problem Statement: Predicting Daily Ambulance Call Volume in NYC

Objective:

The objective is to predict daily ambulance call volumes for each NYC borough, leveraging historical emergency call data along with external features such as weather and holidays. Accurate forecasts can help optimize ambulance allocation, improve response times, and support resource planning.

Context:

NYC Emergency Medical Services (EMS) experiences variable daily demand across boroughs. Calls are influenced by temporal patterns (day of week, holidays), spatial factors (borough-level demand), and potentially weather. Historical data is available hourly, but volume and variability make direct analysis challenging.

Criteria for Success:

- Predict daily ambulance call volumes with reasonable accuracy across all boroughs.
- Capture key temporal patterns such as weekday/weekend effects and holiday anomalies.
- Identify trends that can inform EMS staffing and resource allocation.

Scope of Solution Space:

- Focus on daily aggregation of ambulance calls by borough.
- Include external predictors: weather features (temperature, precipitation, snowfall, snow depth) and holiday/weekend indicators are used.
- Weather and holiday data may have gaps or inconsistencies across boroughs.
- Focus is on predictive modeling, not causal inference.

Constraints:

- The original hourly dataset contains ~28 million rows, too large for direct processing. A subset of 500,000 rows (Jan 1, 2025 – Apr 15, 2025) is used.
- Weather and holiday data may have gaps or inconsistencies across boroughs.
- While it is technically possible to incorporate special event indicators (e.g., parades, sports events) by borough or ZIP code, doing so at a borough level introduces a bias - Localized events (e.g., in one ZIP code) could affect only a small area, but aggregating this signal to the entire borough may distort predictions by inflating borough-wide call volume estimates.
- This project focuses on predictive modeling, not causal inference, so model interpretability is secondary to forecasting accuracy.

Stakeholders:

- NYC EMS management and dispatch teams (resource allocation, staffing).
- Public health planners (identifying periods of high demand).
- Data analysts and data scientists building predictive models.

Data Sources:

- NYC Emergency Call Data via API (NYC Open Data API)
- External datasets:
    - Weather (temperature, precipitation, snowfall, snow depth)
    - Holiday calendars by borough.