PROBLEM 2: DETECTING AND ANALYZING SECURITY THREATS

Dataset Description

File: `dataset2_threat_detection.csv`

The dataset contains 5,000 security threat records over 30 days with the following fields:

- `timestamp`: When the threat was detected
- `threat_type`: Type of security threat
- `severity`: Threat severity level (Low, Medium, High, Critical)
- `source_ip`: Source IP address
- `affected_system`: System affected by the threat
- `hostname`: Hostname of affected system
- `user_account`: User account involved
- `detection_method`: How the threat was detected
- `status`: Current status of the threat
- `response_time_minutes`: Time taken to respond
- `confidence_score`: Detection confidence (0-100)
- `event_count`: Number of related events
- Additional derived fields for analysis

Learning Objectives

- Understand threat landscape and severity distribution
- Analyze detection effectiveness and response times
- Identify vulnerable systems and users
- Evaluate security controls performance

Questions to Answer

Q1: Threat Landscape Overview

a) What are the most common types of threats detected?

b) What is the distribution of threat severity levels?

c) Create a visualization showing threat trends over the 30-day period

Q2: Severity Analysis

a) Which threat types are most likely to be Critical severity?

b) Compare the average response time across different severity levels

c) Create a treemap showing threat distribution by severity and type

Q3: Detection Performance

a) Which detection methods are most effective (highest confidence scores)?

b) What is the average confidence score by threat type?

c) Identify threats with low confidence scores (<70) that might be false positives

Q4: Response Time Analysis

a) What is the average response time for each severity level?

b) Are response times improving over the monitoring period?

c) Create a box plot showing response time distribution by severity

Q5: System Vulnerability Assessment

a) Which systems are most frequently targeted?

b) Which systems experience the highest severity threats?

c) Create a bubble chart showing systems by threat count and average severity

Q6: User Risk Analysis

a) Identify users associated with the most security threats

b) Which users have admin/privileged accounts involved in incidents?

c) Visualize user risk scores based on threat involvement

Q7: Status and Resolution Tracking

a) What percentage of threats are blocked vs. investigating vs. false positives?

b) How does resolution status vary by threat severity?

c) Create a Sankey diagram showing threat flow from detection to resolution

Q8: Internal vs. External Threats

a) Compare threats originating from internal vs. external IP addresses

b) Which threat types are more common from external sources?

c) Visualize the geographic or network distribution of threat sources