DATA SCIENCE THREAT INVESTIGATION.

**Data Analysis Approach:**

**1.Data Collection and Integration:** Gather relevant data from network logs (firewall, intrusion detection systems), employee access records (login/logout times, resource access), email communications (internal/external emails), and potentially other sources like file server activity, printing logs, or physical access records. Crucially, ensure data is collected in a forensically sound manner, preserving its integrity for potential legal proceedings. Integrate this data into a centralized platform for analysis.

**2.Baseline Establishment:** Establish a baseline of "normal" behavior for each employee. This involves analyzing historical data to understand typical work patterns, resource usage, communication patterns, and access times. Statistical methods like calculating means, standard deviations, and identifying common access patterns are useful here.

**3.Anomaly Detection:** Employ anomaly detection techniques to identify deviations from the established baselines. This could include:

**Unusual Access Patterns:** Accessing sensitive data outside of normal working hours, accessing resources unrelated to their job function, or attempting to access restricted systems.

**Communication Anomalies:** Sudden increase in email volume, communication with external or suspicious email addresses, use of code words or unusual language.

**Network Activity Deviations:** Large data transfers, access to unusual ports or protocols, or connections to known malicious IP addresses.

**Behavioral Changes:** Changes in login frequency, file access patterns, or printing activity.

1. **Correlation and Contextualization:** Correlate anomalies across different data sources to build a comprehensive picture of potentially malicious activity. For example, an unusual network access pattern combined with an email to a competitor would raise a higher level of suspicion. Contextual information, such as employee performance reviews, project assignments, or known grievances, can be valuable in interpreting anomalies.

**Challenges.**

**Distinguishing Legitimate vs. Malicious:** Differentiating between genuine work-related activity and malicious actions is a significant challenge. False positives are common, and careful analysis is needed to avoid accusing innocent employees.

**Data Volume and Variety:** The sheer volume and variety of data sources can be overwhelming. Efficient data processing and analysis techniques are essential.

**Data Silos:** Data may be spread across different systems and departments, making integration and analysis difficult.

**Evolving Threats:** Insider threats can be sophisticated and adapt their tactics to avoid detection. Continuous monitoring and refinement of detection methods are necessary.

**Balancing Security And Privacy.**

**Data Minimization:** Collect only the data that is absolutely necessary for the investigation.

**Purpose Limitation:** Use the collected data only for the purpose of investigating the suspected insider threat.

**Access Control:** Restrict access to the collected data to a small team of authorized personnel.

**Transparency (where possible):** Inform employees about the data collection and analysis process, while respecting the confidentiality of the investigation.

**Data Retention:** Retain the collected data only for as long as necessary.

**Legal Compliance:** Adhere to all relevant data privacy regulations (e.g., GDPR, CCPA).

**Transparency And Ethical Standards.**

**Establish Clear Policies:** Develop clear policies regarding data collection, monitoring, and employee privacy.

**Document Everything:** Maintain detailed records of the investigation process, including data sources, analysis methods, and findings.

**Independent Review:** Involve an independent party to review the investigation findings and ensure that ethical standards are being followed.

**Employee Communication:** Communicate the findings of the investigation to relevant stakeholders in a clear and transparent manner.

**Communication With Stakeholders.**

**Tailor Communication:** Adapt the communication style and level of detail to the specific audience (technical vs. non-technical).

**Visualizations:** Use charts and graphs to effectively communicate complex data and findings.

**Focus on Key Findings:** Highlight the most important findings and their implications.

**Provide Context:** Explain the background of the investigation and the methodology used.

**Be Objective:** Present the findings in a clear and objective manner, avoiding speculation or bias.