**QUESTION 1**

**Approach to Analyzing Diverse Data Sources**

***1.Data Collection & Integration***

- Collect network logs (IP connections, unusual traffic volumes).

- Collect employee access records (login times, locations, resource usage).

- Collect email communications metadata (frequency, recipients, anomalies in attachments or timing – but limit deep content      analysis unless warranted).

- Integrate these into a unified data model for correlation.

***2. Feature Engineering & Anomaly Detection***

- Network logs: detect unusual outbound connections, access to restricted servers, or data exfiltration attempts.

- Access records: flag abnormal login times (e.g., 3 AM logins), geographic mismatches, or excessive failed attempts.

- Email data: look for unusual spikes in external communications, sensitive keywords in subject lines, or large attachments     
  sent outside the organization.

- Use baselines of normal behavior per employee/team and apply machine learning anomaly detection or rule-based                     
   thresholds.

**Challenges  I Might  Face in Distinguishing Legitimate vs Malicious Actions**

1. Overlap between normal and malicious behavior:  
    Example: A developer accessing large amounts of source code could be doing legitimate work or exfiltrating data.

2. False positives: Anomalies (e.g., late-night work) may simply be due to flexible schedules.

3. Dynamic roles: Employees’ access needs change (e.g., cross-project work), making baselines tricky.

4. Data context: Logs show what happened, but not always why it happened.

**Balancing Cybersecurity with Privacy Rights Of Employees**

1. Principle of proportionality: Only collect and analyze the minimum data necessary for the investigation.

2. Metadata focus: Prioritize analyzing behavioral patterns and metadata over reading personal email content.

3. Anonymization/pseudonymization: Use anonymized identifiers during early-stage analysis; only de-anonymize if     
     suspicious activity is confirmed.

4. Access controls: Limit who can view sensitive data (e.g., emails) to authorized investigators.

**Maintaining Transparency and Ethical Standards**

1.Clear policies: Ensure employees are aware of monitoring practices via acceptable-use policies.

2.Audit trail: Maintain detailed records of all investigative steps for accountability.

3.Ethics board / compliance oversight: Have investigations reviewed by legal and compliance teams.

4.Fairness: Avoid targeting individuals without strong evidence; always consider alternative explanations.

**How To Communicate Findings**

***1.For technical stakeholders*** (IT Security, Data Teams):

- Provide detailed anomaly detection reports, raw logs, and correlation visualizations.

- Use dashboards/heatmaps to show risk scores by user or department.

***2. For non-technical stakeholders*** (Executives, HR, Legal):

- Summarize risks in plain language: e.g., “Unusual data transfer outside business hours from employee X.”

- Focus on business impact (data loss risk, compliance breach, reputational harm).

- Present only confirmed patterns, not raw personal data, to avoid privacy overreach.

**QUESTION 2**

import matplotlib.pyplot as plt  
import numpy as np

months = ['07/2019', '08/2019', '09/2019', '10/2019', '11/2019']  
searches = [50, 53, 59, 56, 62]  
direct = [39, 47, 42, 51, 51]  
social\_media = [70, 80, 90, 87, 92]

x = np.arange(len(months))  
width = 0.25

fig, ax = plt.subplots()  
ax.bar(x - width, searches, width, label='Searches', color='blue')  
ax.bar(x, direct, width, label='Direct', color='pink')  
ax.bar(x + width, social\_media, width, label='Social Media', color='orange')

ax.set\_xlabel('Months')  
ax.set\_ylabel('Visitors in thousands')  
ax.set\_title('Visitors by web traffic sources')  
ax.set\_xticks(x)  
ax.set\_xticklabels(months)  
ax.legend()

plt.show()