From the datasets

Log-on info	id, user, pc, date, activity (log-on, log-off)	Can be used to spot abnormal working hours, logon frequency, or access from unusual machines
Email info	id, user, pc, to, from, date, subject, attachments, activity	To track internal and external communication, weird and high volume attachments
File info	id, user, pc, file, activity (Read / Write / Delete / Copy), date	What files are being accessed by whom, reading or modifying sensitive files
Web browsing info	id, user, pc, url, activity (uploading, downloading etc), date	Are they hitting suspicious links, maybe hitting phishing links, hitting competitor's links, uploading info at any site other than the org's domain.
External device info	id, user, pc, device, activity (Connect / Disconnect), date	Are they connecting any removable media for possible data theft
Psychometric	user, openness, conscientiousness, extraversion, agreeableness, neuroticism	Personality traits mapped to how close they are to insider risk.
HR events	user, event_type (Promotion / Warning / Termination), date	Maybe the event_type can prompt insider risk

Factors we can use for BillyBank

For simulation, we bring down the workforce into 5 behavioural access profiles:

Trader	Employees -> front office	Handles market and trading data - high exposure to financial records and client portfolios.
Analyst	Employees -> research / reporting / financial modeling	Access to confidential reports, pricing models, and client analytics.
IT_Admin	Employees -> privileged infrastructure and data custodians	Manages servers, logs, databases, and privileged accounts - potential for sabotage or privilege misuse.
Contractor	Contractors / temporary staff	Limited tenure but often huge data access - higher likelihood of using removable media.
Exec_Support	Executive Assistant to C-level / Regional Directors	Proximity to board materials, Merger Acquisition documents, and regulatory filings - sensitive communication exposure.

Psychometric (Predisposition)

Conscientiousness

- How organized, disciplined, and rule-abiding an employee is
- Low values -> higher chance of policy violations or careless behavior

Neuroticism

- How emotionally unstable or stress-reactive the person is
- High values -> higher stress response; when combined with HR flag, can trigger malicious behavior

Low conscientiousness -> negligence, shortcuts, and non-compliance
High neuroticism -> stress-driven or retaliatory acts under perceived pressure

HR factor (Stressor)

- Is_hr_flagged - a flag to determine the chance of an HR event
- Probabilities
  - Trader 2%
  - IT Admin 3%
  - Analyst 2%
  - Contractor 5%
  - Exec Assistant 1%
- This can push a high risk (psychometric) employee to perform insider threats.

Behavioral / Technical Activity Factors (Opportunity)

This is based on the factors from the 1st page (the ones in the dataset)

after_hours_logons	Number of logons outside normal business hours.	Possible unauthorized access or data gathering after supervision hours.
sensitive_file_reads	Count of accesses to protected or high-value directories (e.g., trading models, client PII).	Data exfiltration or reconnaissance.
usb_device_mounts	usb_device_mounts	Physical data theft or shadow transfer.
external_emails_sent	Outbound emails to non-corporate domains.	Potential data leak or client collusion.
emails_with_attachments	Internal or external emails containing file attachments.	means for data exfiltration.
cloud_upload_events	Web uploads to personal or unapproved cloud services.	Shadow IT usage / cloud exfiltration.
failed_logins	Count of failed authentication attempts.	Account compromise attempts or brute-force misuse.
files_deleted	Number of deletions of local or shared files.	Cover-up of malicious actions or sabotage.
http_competitor_visits	Visits to competitor or high-risk external sites.	Early indicators of job hunting or data sharing intent.

Probability of being malicious

- P(Malicious)=f(Predisposition,Stressor,Opportunity)