# Short description of how my sampling works

1. Takes top 25 most active users
2. Takes 3 logs from each email.csv, device.csv, file.csv, http.csv, logon.csv (normal\_sample.csv)
3. Takes 10 malicious users
4. Takes malicious logs from the r4.2 answers file (malicious\_sample.csv)
5. Combine all logs and filter by user (combined\_dataset.csv)
6. Convert all user logs into prompts (user\_prompts.json)

{

        "role": "user",

        "content": "User: AJR0319\nActivities:\nAccessed file a file at 14:06\nFile content: 33 vegas campaign patrick undefeated hit lack or 26 influenced hometown recalled including assigned allowing met punch 8 from network increased together germany silver mat 1973 chicago hit straight effort erik czech boards which 2 left 33 spent old at became near easily asked segment strength going ... during business hours\nAccessed file a file at 08:38\nFile content: course already now surveillance 1974 roosevelt notorious joyce signed presentation 7 identification wanted cell department hill 1997 holding las patrols guard helicopters face rainbow newkill thick speed either bail chancellor 100 aunt assatas related listened short asylum faculty washington permitt... during business hours\nAccessed file a file at 08:33\nFile content: 8 havana robbed revolutionaries accounts these thick relocated entrance controversy harlem out short truth 26 graduation traffic free foster struggle march arraigned front through paid was pope her confrontation man hotel wounded other neither i renamed change insufficient stakeouts here duty still ... during business hours\nVisited website http://list-manage.com/Society\_of\_the\_Song\_Dynasty/kaifeng/sverqrffregcrgtebbzvatjnyxvatfgvpx841773898.php at 08:54 during business hours\nVisited website http://ebay.com/Hamilton\_Ontario/stelcos/nhgbercnveivqrbfobngenpvatfrncbeg1029042913.php at 15:23 during business hours\nVisited website http://sidereel.com/Miniopterus\_griveaudi/griveaudi/ivfnovplpyrpebffpbhageljbeyqfrevrf1447879033.htm at 08:58 during business hours\nLogged off at 15:40 during business hours\nLogged off at 15:53 during business hours\nLogged off at 15:37 during business hours\nLogged in at 13:24 during business hours\nLogged in at 07:42 during business hours\nLogged in at 07:30 during business hours\nWhat is the risk level?"

    },

1. Each user has one prompt which looks something like this
2. The malicious users should have suspicious activity that includes http link access at (wikileaks.org) or access pc outside business hours multiple times