Possible sources for use cases:

1. Netflow
   1. Take a 24 hour period (or hourly sampling), compare current snapshot to previous/similar views, look for clustering
   2. System classification (unsupervised)
      1. MDS (distance plotting)
      2. k-nearest neighbors
         1. note, these two were taken directly from ML for hackers
   3. System classification (supervised)
      1. random forest
         1. I can collect a handful of “known bad” IP addresses and their function and we could pull/create a handful of profiles to train on.
2. Mod\_security logs
3. port scans / honeypot data
4. list of bad IPs (or IP addresses)
   1. ASN
   2. Geolocation
   3. DNS / Passive DNS
5. Nessus (or other) Vuln scan
6. Spam (general)
7. Bots
   1. Malware C&C
   2. netflow analysis of
8. Open DNS
9. Password lists
   1. Character analysis, etc
10. Meta-analysis
    1. internal vs external actor
11. Firewall logs
12. Patch coverage
13. Logins / LDAP queries
14. SNORT (or other) IDS
15. Phishing
16. VERIS data