EDA Task

This task consisted of getting data for alerts and incidents from AWS Athena and visually analyzing the data in order to find patterns.

The approach:

* 1. Started working with 2022 December data.

i. For 01/12/2022, most incidents and alerts happened at ACCT0011200 site\_id.

ii. The highest difference between the number of alerts and incidents occurred at sub site ACCT0011884

iii. Pivoted to getting only sev 1 incidents. Found max at sub site ACCT0013161.

iv. Manual go-through of the data lead to following conclusions:

* + - 1. switch cpu over utilization, ap memory over utilization, clarity auth failure and aggregated device disconnections taking place together
      2. rogue ap and uxi taking place together
  1. Pivoted to working on the latest data (Feb – Apr)

i. For sub site ACCT0012484, ACCT0011385, ACCT0012924, ACCT0012557, ACCT0012287, ACCT0012893

* + - 1. plotted scatter plots of incident description and alert description with respect to epoch, along with their priorities
      2. plotted scatter plot for sev 1 and sev 2 incidents separately with the incident categories and alert categories w.r.t epoch
  1. Pivoted to working only on THD sev 1 incidents

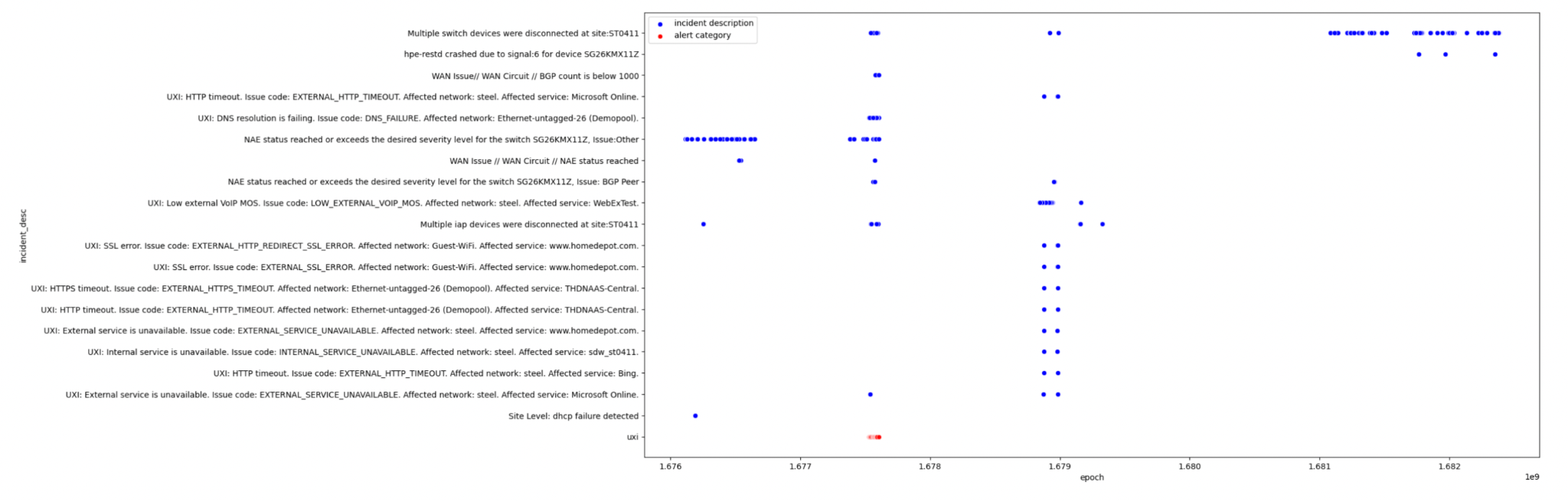
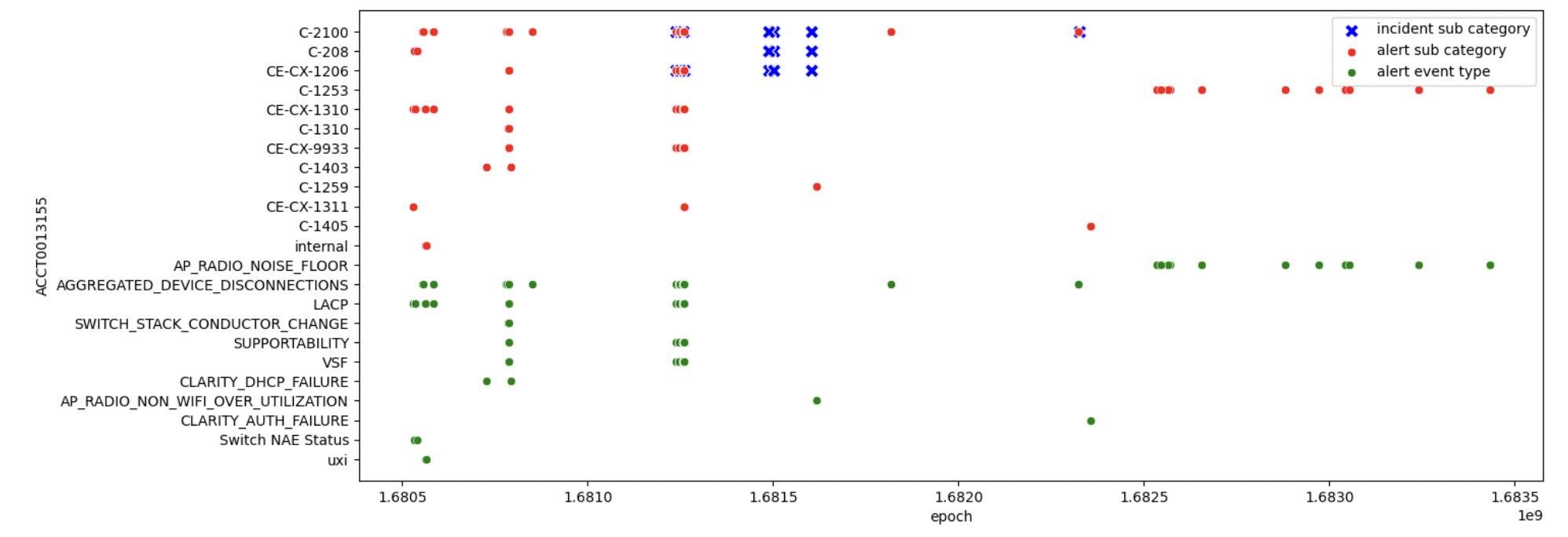
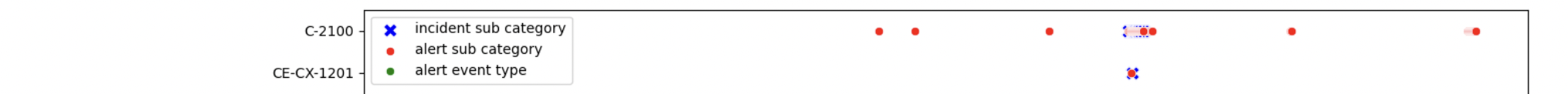
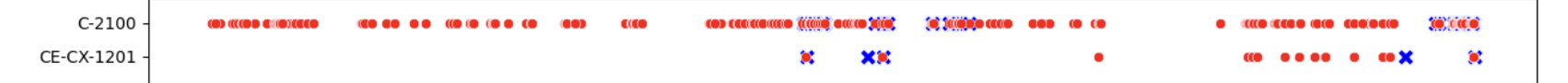
i. Additionally fetched the data for alert sub categories, incident sub categories and alert event type from Athena

ii. Plotted scatter plot for incident and alert sub categories and alert event types w.r.t epoch for the complete time duration

iii. Plotted scatter plot for the same at a gap of 5 days, thrice

4. Plotted scatter plots for the month of April to check if the alerts are being generated as they should.

Conclusions:

1. No particular patterns in the data for the months of February, March and April
2. Clear patterns for Module reboot, Partner out of sync and Peer timeout taking place together
3. UXI HTTP timeout, External and internal service unavailable, SSL error taking place at the same time
4. Incidents C-2100, C-208, CE-CX-1206 taking place in the same time frame at a few sub sites
5. C-2100 and CE-CX-1201 taking place in the same time frame at multiple sub sites

For more information and plots, please refer to the following github repo:

<https://github.com/kartikchandna2/data-analysis-task/tree/main>