Wed 09/17

- Intros
- Decide on scenario
    - Considerations: team's strengths, potential ease of finding a dataset
    - Vy submitted once decided: 2, 1, 3
- Decide on potential recurring meeting slot: Wednesdays 2-3pm
- Akhil not available over fall break, Paul, Shyon, and Vy can

Wed 09/24

- Milestone 1 planning

Sat 09/27

- Review drafts of exec summary sections
    - Q1-3: @Vy to draft up
    - Q4: @Paul @Shyon Security measures research, based off of real websites if possible
    - Q5: @all, each person comes up with list of things you wanna have (outcomes/deliverables) and associated methodologies. We can come up with methodologies after deciding on deliverables we want
    - @Akhil to look deeper into SEI insider threat datasets
- Decide on sections 4-5 especially

Wed 10/01

- Finalize and submit Milestone 1

Sun 10/12

- Game plan for final deliverables:
    - **Week 1 (fall break): Scoping & Data Foundation (i.e. get the heavy stats stuff out of the way)**
        - Define key insider threat scenarios (e.g., contractor exfiltrating M&A data, trader manipulating models).
        - Select SEI CERT features relevant to BillyBank's roles.
        - Preprocess data (filter, normalize behavioral indicators, …)
        - Make synthetic dataset if needed.
        - Apply linear regression to identify key predictors of insider behavior
        - Run Monte Carlo simulation to estimate Expected Annual Loss (EAL) by role, region, insider type, etc.
        - Generate static charts and/or heatmap as see fit
    - **Week 2:**
        - Turn the stats into an interactive visual tool

        - Try to get concrete math perhaps on how features affect metrics
        - Create dashboard prototype (Streamlit?)
        - Integrate EAL + regression outputs to display live metrics
        - Add sliders/toggles for mitigation actions (training coverage, PAM adoption, detection sensitivity, ..)
        - Visualize “before vs after” financial risk reduction

Wed 10/15

- Akhil’s findings on dataset and features of interest
    - Logon csv (privileged admin, odd login times, etc), device csv, http, psychometric (user’s behaviors) (are they open to sharing, agreeableness, prob from survey)
    - Build out synthetic dataset, or getting script from AI to clean existing, can get that done within week
    - Look into case study of similar financial institution, take something close to our org, say losses can be in certain range (min and max) then random in between. Can choose specific numbers for Monte Carlo later.
- Shyon’s Streamlit experiment
    - Whipped up something simple, should be simple
- Paul looked into remediations for later. Also Monte Carlo ideas. Get figures from real breaches, litigation fees, lawyer fees, etc. ex if insider leaked data, and if there’s no logging, how much would have to pay out. Also add legally mandated stuff for training portion.
- Maybe deliverable be survey that would give outcomes on whether someone is prone to insider threat, then can use that to tailor trainings
    - Akhil can try mapping emotional responses (5) to actual attributes → written by Friday

Sat 10/25

- Akhil’s dataset info + synthetic dataset
- Continue planning
    - If is_malicious flag on, map some financial loss associated, then run Monte Carlo on that.
    - Paul to look into case studies: financial losses associated with Chase.
        - potential impact: can start from mapping actors to which type of threats they may pose, map to which role as well, which can help with risk per role for ML
    - Akhil can look into ML stuff to predict how much of a risk is a person within a 30-day period
        - Likelihood

- Aggregate loss range for Monte Carlo
- Shyon + Vy: Figuring which factors to think about for questionnaire, make a subset of behavior factors

Sat 11/1:

- Finalize and Vy to submit Milestone 2
- Paul: Software that Chase-scale companies use for non-reputation costs 80-90k/year
- Opportunity score → how much deviates from baseline → deduce probability. Model for number of working days per year, using more manageable # of employees. For particular role, region, probability someone is malicious?

Sun 11/9:

- Debrief Paul's sheets on impact case studies and remediations
- Paul's goal for this week: compile each role and how much lost they'd incur, min (doing 1 least costly thing) and max (all malicious can do), average of how much, so Akhil can use those values to total annual loss per role for region, script will be separate so that when checkbox is clicked so can recalc Monte Carlo accordingly, default 1
- Interactive dashboard: Heatmap, Monte Carlo for every role in a given year, checkboxes of solutions you can apply, if click, rerun Monte Carlo, if nothing checked multiply by some scaler 1, otherwise sum of probabilities of what you're clicking (shouldn't exceed 0.4). Summary statistics and percentiles.
- Shyon to try front-end
    - Akhil to give script
    - Integrate front end and Python plots

Sun 11/16:

- *(Vy to let Paul join via Zoom)*
- Debrief Vy's proposed presentation flow: everyone needs to address (1) What I did (2) Why it matters (3) CRM connection
    - Vy (PM) recaps client problem/scenario, overview of what team did (name drop course concepts like FAIR and MC)
    - Akhil (Data lead): screenshot of synthetic dataset (billybank_activity_updated.xlsx), heatmap
    - Paul (Risk/Remediation lead): talk about mapping role-level probabilities into monetary impact by building loss bands for likely incident types using industry refs and JPMC-like case studies, then run Monte Carlos that gives us EAL, etc. percentiles?

    - Shyon (UI lead): Pull up interactive front-end that ties to model and Monte Carlo, explain features available
    - Vy: Audience input stuff here? Then go into final CoAs/recs, tie back to BillyBank's mission. Can be a bit salesy on a particular rec (contractor PAM, targeted UEBA). Maybe reference KPIs they can track
- Decide on final blueprint format: quick video to play at start of presentation, or short final report
- Submit for early feedback possible?
- Deliverables to finish:
    - GitHub repo w/ any relevant notebooks, code, etc.
    - Deployed front-end w/ link added to slide for ppl to follow along
    - CRM Sources Referenced
    - Team 3 – CRM Project Final Presentation
- Next meeting
    - Akhil to send scripts over to Shyon by Monday to flesh out front-end, printing Monte Carlo and JSONs, vizs in JPEG

Sat 11/22:

- *(Vy to open Zoom room)*
- Shyon's streamlit frond-end, see if can deploy, add link/QR to deck below
- Debrief Vy's drafted deck: Team 3 – CRM Project Final Presentation
    - Decide on final recs
- To do: Update README to submit as blueprint to tie everything together

Mon 11/24:

- *(Vy to open Zoom room for Paul)*
- Rehearse presentation, decide what to cut to stay under time limit
- Final feedback for Shyon's streamlit frond-end