PMP Plan

Fritz, Mason, Sara, Susan

- Label subway alerts with delay or not delay (1/23) (Susan and Fritz)
    - Include train line and direction/borough
        - Get train stations affected by the delay
    - Try coming up with a regular expression versus manual labeling (Mason)
    - Possibly NLP??
- Clean complaint data (1/23)(Sara)
    - (keep after 2017)(remove irrelevant complaints)(include station)(datetime)
    - Assign each crime to the closest train station using geopy
- Add delayed column to complaint data (yes or no for train delayed nearby)(1/24)(Mason)
    - For two hours after each delay
- EDA (1/24)(Susan)
- Create machine learning model(1/26)(Fritz)
    - Select a model
    - Predict how much crime will happen based on a delay
    - Refine model and increase accuracy
- SQL database (1/26)(Sara)
    - databricks/azure data factory?
    - Create tables?
- Use kafka and adf(1/27)(Fritz)
    - Make producer using Susan's web scraper
    - Make consumer in a databrick to put data into storage
    - Put data through model, and output to dashboard
- ETL report(1/30)(Mason/Susan)
- Build a dashboard(Mason/Sara)
    - Napkin drawing for visualization (1/27)
    - Napkin drawing for dashboard (1/27)
    - Power bi or dash(1/31)
- Get feedback
- Write technical report(1/31)(All)
- Create presentation(2/1)(All)