1. Collect data (function so I can collect from multiple tickers)
   1. Ticker open, close, etc
   2. GT for ticker name search
   3. General market aggregate measure (S&P 500?)
2. EDA - .5 days
   1. Validate data against another data source
   2. Check volatility
   3. Check correlations, seasonaility, stationarity,
   4. Tbd other tasks included in stock eda practices
   5. Lags
      1. Autocorrelation plots
3. Preprocessing 1 days
   1. Combine GT, ticker, aggregate data
      1. align on date index and equalize frequencies
   2. generate any other covid features
      1. research how stock analysis has dealt with covid
      2. use cut off date to compare before/after
         1. look up ‘event study’
   3. generate percent change columns
4. Modeling 2 days
   1. Train test split
      1. Training on 2018-2020
      2. Testing on first half of 2021
   2. Build baseline time series model (not using GT)
      1. Decide what models to use
         1. Interpretable model
      2. Test training time for models selected
      3. Build and train models selected
      4. evaluate and interpret models
         1. calculate error metrics
         2. analyze residuals
         3. plot predictions vs actuals
         4. interpret predictions
   3. build time series on complete data (using GT)
      1. decide which models to use
      2. test training time
      3. build, train models
      4. interpret
5. conclusions 2 days
   1. For each ticker find the difference between the best performing GT and non GT model
   2. Hypothesis test to see if difference between GT/non GT is significant
      1. Analyze the residuals and extremes:
         1. Where were there high gains/low gains from using GT. Did certain tickers lose out or have big gains from using GT.
   3. answer whether GT improved modeling accuracy (answer problem statement)
      1. Focus on whether GT can improve the best model.