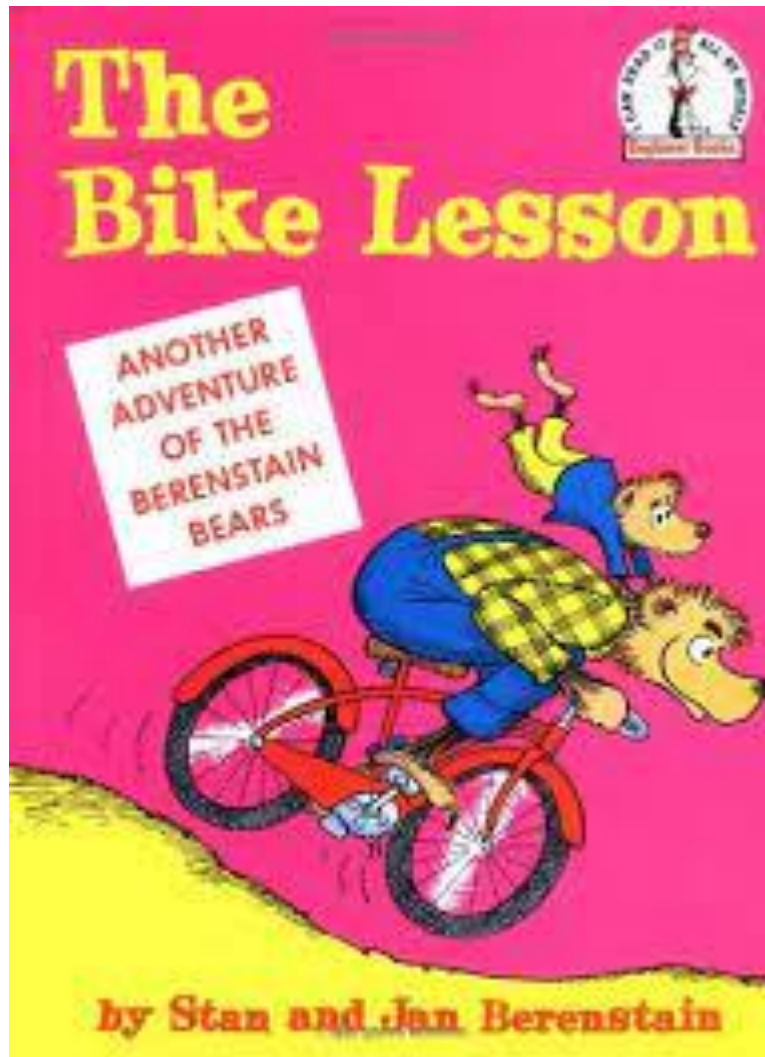# Is the stock market insane? I can't tell you!

Or, how I spent 4 months learning what not to do in data science.

Things I did:

- Import and collate two distinct datasets (SEC filing data and stock price data) ▲
- Include too much price information and mess up my predictions ▼
- Figure out that my data isn't normal at all ▲
- Make it a little better with a percent-change transform ▲
- Forget to use the percent-change transformed data when I tried the predictions again.
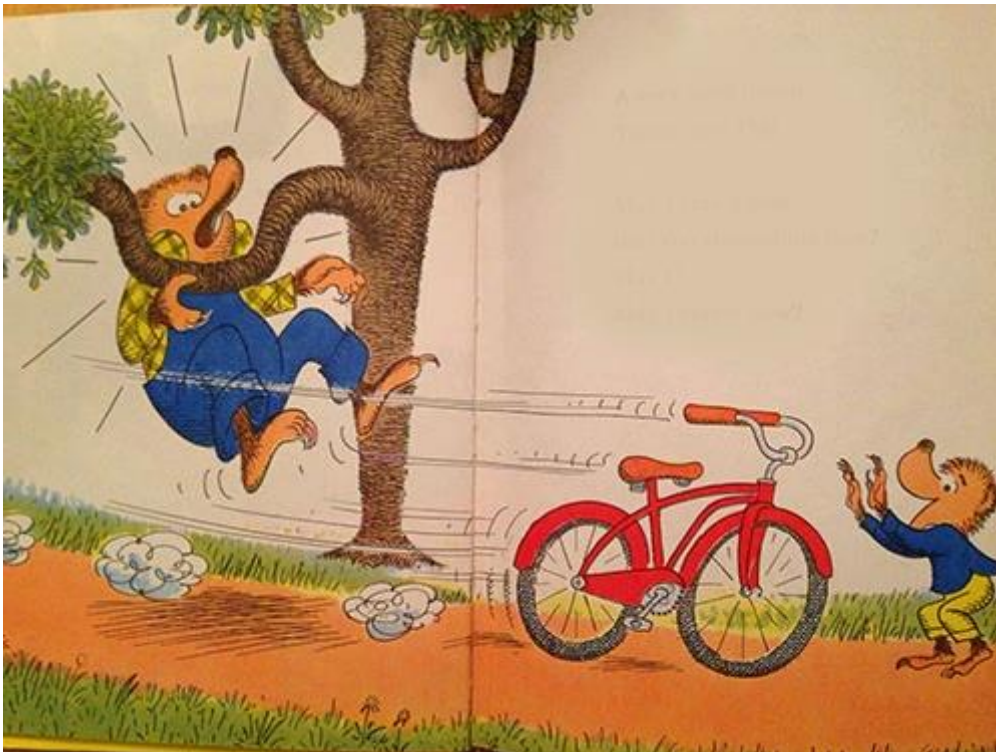
Things I did not do:

- Deeply clean my data and account for the presence of outliers.
- Apply transformations on a stock-by-stock basis (I did do this for percent-changes but not for other transforms like log and min/max)
- Treat my data as time-series information and correct accordingly.
- Remind myself regularly of my project proposal and the way I thought I was going to answer my questions.

Things I wish I hadn't done:

- Mix up data belonging to various stocks, even if they were in a similar industry.
- Use linear regression on known non-normal data.

Things I would do if I did this again:

- Check and correct for time-series concerns
  - Inflation! My data spanned 10 years or more
  - Seasonality, especially Q1-Q3 vs year-end results
  - Detrending, if inflation didn't correct it all
- Be more firm about choosing a prediction method
- Do more cleanup work
- Normalize data within each stock before any aggregating for prediction
- Keep my eyes on the project plan!