1. Import libraries
2. Sneak peek of the raw data
3. Load the entire csv file (~61 Mega Bytes)
   1. From a preview, decide what to drop and what to keep
      1. sp500 = pd.read\_csv("data/sp500.csv",index\_col='Symbol', usecols=[0, 2, 3, 7])
   2. And also this is time to think about the meaning of features (columns)
      1. Decided to merge ‘listing’ and ‘time\_on\_site’ -> ‘listing\_per\_day’
4. Seen from the header of the data, there are sellers with ZERO activity
   1. Easy to sort out
5. After this step, I tried to observe the data in various ways.
   1. Scatter plot, bar chart and etc.
6. **Principal Component Analysis**