

## Instructions:

- Using the data at [https://reveliolabs.s3.us-east-2.amazonaws.com/Sample\\_Workforce\\_Data\\_0720.csv](https://reveliolabs.s3.us-east-2.amazonaws.com/Sample_Workforce_Data_0720.csv) , create a Jupyter Notebook that asks and answers a question about the workforce of any, or all, of the companies listed. You may address just one company, an industry, the 10 largest or all of them... there are no requirements on what question you may ask other than that it is interesting (Does Facebook have more software engineers than Walmart is not interesting).
- This project is intended to be open ended. You will all start with the same dataset on employment data for xxx companies. The question you choose to answer with that data, however, is up to you, as are the methods you use.
- Unlike the problem sets, **there is no specific "checklist" you have to follow in order to get a good grade on this**, as long as you demonstrate the following:
- Comprehensive use of python code & concepts learned over the semester thus far (30%)
- Depth of project analysis, i.e. how many scenarios/"experiments" did you test (30%)
- Interpretation of data results (30%)
- Readability/reproducibility of code/visualizations. How easy is it for someone else to understand what you are doing with your code? Would someone be able to replicate it for another project? Are your graphs legible? Do they have titles, appropriate axis, easily understood legends (10%)?
- The code behind the graphs and statistics included in articles you read online is often made available to the public so it can assist others in their work and to help with reproducibility of results. Your Notebook should be created with this in mind, though you do not need to write paragraphs explaining every for-loop; simple notes on what a certain group of lines do is satisfactory.