

**IEOR 142 Final Project Proposal**  
**Fanice Nyatigo, Haibin Lim, Dorothy Leung, Kana Mishra, Ashleigh Purvis**

1. Motivation: What problem are you solving? Why?
  - Analyzing diversity and pay among top companies
  - To determine whether or not diversity levels among top companies has an association to revenue earnings for the company as a whole
  - Look at trend between diversity and compensation
2. Data: What is the nature of the data used? How do we plan on collecting and processing the data?
  - Measure quantitative data collected through observation from different companies
  - Look for open source data sets
  - Scrape Glassdoor for average reported salaries (ones with high sample sizes)
  - Look up diversity disclosure reports for corresponding companies
  - Clean data to be on the same scale (for salary) and group by company, type of job, etc
3. Analytics models: What techniques do we plan to use?
  - Data cleaning (data wrangling, handling null values and categorical features)
  - Data visualization (observe trends and patterns from the data graphically)
  - CART or regression based on the data we collect (build a model that analyzes the diversity and pay)
  - Possibly look into time series for trends in pay and diversity at individual companies
4. Impact: What is the (potential) impact of your work with regard to the problem that you are trying to solve? What is the overall goal?(if you could build a perfect model, what would it be able to do?). How might you expand the scope of your analysis to improve its impact even more?
  - Our work can potentially impact revenue and wages at top companies
  - The overall goal is to delineate wage gaps not just among gender but also among other forms of diversity
  - A perfect model would ideally show the trend of the wage gap and how it's not really being fixed, or it would show how long it would take for the gap to be eliminated at the rate it's going at (which we would then hope to make the gap be eliminated faster by advertising our findings)
  - A perfect model would also be able to predict the wages of people at up and coming companies too, as well as the wage gaps of the diverse groups present in the company
  - To create a larger impact, we might expand this to look at companies internationally to get a sense of what we might be doing wrong in America