VANDERBILT | M.S. DATA SCIENCE

Capstone Development (DS-5999)

Project Proposal

Please answer the following questions to help guide you in scoping out your Capstone project.

1. **Who are you working with on this problem? This could be a company, a faculty member, a research group, etc. If it is just you, that's fine, just say "NA" here.**
   For my capstone project, I will collaborate with Peter Harry Haslag and Joshua Tyler White, who are professors at the Owen School and specialize in finance.

2. **Will you need to meet with me during the semester? If you said "NA" above, the answer to this question is "Yes." Not meeting with me requires that some other faculty member or practitioner volunteer to provide mentoring and/or oversight. If you say "No" here, please let me know who this person is who will perform this task. It is perfectly fine for you to meet with both me and someone else if you prefer that, and you can change your mind later.**
   I can provide the following:
   a. **Accountability on timelines, planning and meeting deadlines.**
   b. **Help in how to solve difficult logical coding/analysis challenges.**
   c. **Assist in formulating and calibrating metrics that effectively communicate results.**
   d. **Scoping out your problem and ensuring it is not too big or too small.**
   e. **Anything having to do with analytics, statistics, and classification/regression ML tasks.**
   f. **I will be of less help in the following areas: Image recognition, NLP as well as specifics in deep learning.**
   I will be conducting biweekly meetings with Peter every Tuesday throughout the spring semester to provide updates and discuss progress. In the event of new questions or challenges, I will seek assistance from Jesse Blocher as necessary.

3. **Describe the problem you are solving. Be sure to include why this problem is unique or novel. If you are working in a research group or team, be sure to detail your precise contribution in relation to what the whole group is doing.**
   The project question is "Does a CEO's prior work experience distort internal labor markets and create higher turnover costs in non-related segments?" The motivation behind this is that "many articles have pointed out that the accounting profession has had a difficult time hiring new auditors. This has been due to generational shifts, lack of excitement around the job tasks, and lower relative wages. One unexplored explanation is that the experience of the CEOs who head accounting firms has changed from auditing to consulting as the Big 4 accounting firms have grown their consulting arms. Given the CEO's prior work experience, does that dictate what area they emphasize and are there

signs of inefficiency? In other words, does having a CEO who previously worked in the consulting arm of the firm imply that they direct fewer resources toward the auditing arm and either

    a. prevent them from hiring in that segment OR

    b. provide fewer incentives for employees to stay, which will increase turnover"

Alongside Peter, I will be investigating this question but can also focus on one aspect that I find interesting if desired.

4. **Describe the data you need for your project. Be as detailed as you can be here – even including key column names you require is encouraged.**
NOTE: You should be sure that you have the necessary access to this data before the beginning of the semester. Access to data is a key point of delay in these projects!

    For this project, I will have access to multiple data sets. "Overall I will use employment history and job postings to understand the internal and external labor markets. For internal labor markets, these would include elements like the probability and duration of promotion, separation rates, and on-the-job training. Ultimately the goal is to document how the CEO's experience affects these markets."

    For more in-depth about the datasets I will have the following access:

a. Resume data (EMSI):
    i. Data is broken up into 3 sections/files: employment, education, and profile. The ID variable is what ties all these 3 files together.
        1. Education: Includes information on degree types, major, graduation year, school IPEDS, and location. Using IPEDS there is the following data: college characteristics, pricing, tuition, admissions, completion, and FASFA usage. The IPEDS data is a panel at the college-year level.
        2. Employment: Includes information on job title, job function, occupation, firm name, company identifiers, and start/end dates.
        3. Profile: Includes more recent job and current location (last known location)
        4. Addition data: Includes
            a. BLS: wage distribution and employment numbers of each occupation at the MSA-year level. These reports occur about twice a year.
            b. IPEDS: school tuition, applications/acceptances/enrollment, and FAFSA.
            c. ONet: for each occupation, quantifies the relevance across different tasks, knowledge, skills, abilities, work activities, work context, work values, and cross-references to other job title codes. For example, it quantifies the importance of Design knowledge, peripheral vision ability, and critical thinking skills for a Marine Engineer.

b. Glassdoor employee ratings

c. Job postings data (LinkUp)
    i. Posting data: Includes for each post we see the text, date opened, date removed, company identifier
    ii. Company information: Includes the website, industry, and name.

5. **Describe your approach or primary task. What are you going to do with the data above to answer the question above?**

My first task is to understand the data and how they relate to one another. I will then classify people at each of the Big4 into one of the following categories:

   a. audit-focused
   b. consulting-focused
   c. Other

This can be done using job titles, occupation codes from ONet, or other methods that are reasonable given the data we have, or take on a pure-play auditing or consulting firm and use those as a guide. Next, I will create various metrics around the internal labor market or organizational structure to see if their are shifts when a new CEO enters.

6. **Describe what you have done so far, along with an estimate of how much time you have invested in this project already.**

I haven't dedicated time to this project yet, but I intend to allocate the next few weeks before the Spring semester begins to immerse myself in the data. During this period, I plan to extensively review articles that explore similar ideas, aiming to enhance my understanding of the methodology and investigative approaches relevant to my question. This preparatory phase is crucial to ensure that I don't commence analysis blindly. By gaining insights beforehand, I aim to enter the spring semester with a clear idea of how to initiate analysis and effectively work with the data.

7. **Describe what will be the biggest challenges you will face in executing this project. Identify 1-3 challenges.**

My initial challenge lies in comprehending the intricacies of the data and business aspects, particularly understanding the dynamics of internal and external labor markets and how they interrelate. Given my technical background rather than a business-oriented one, I recognize the need for thorough research to bridge this gap. Another challenge I anticipate is the potential for certain methods or metrics, once created or tested, to prove ineffective, potentially leading to time loss as I explore and experiment with alternative methodologies.