IFDA midsemester feedback

The team is getting its arms around working with five client companies (members of IFDA) and an overall client in IFDA itself. Three companies have provided data of varying quality and two companies have not provided any data.

This is a big complicated project with no single point of contact for the client (instead there are six points of contact). I think the team now understands that an overall needs assessment (what data exists, what data would be needed, to answer basic questions about recruitment and retention, etc.) will be one of two important deliverables for the project. The other important deliverable will be some illustrative analyses using what data is available, showing how the data can help answer recruitment (and retention?) questions.

One potential impediment to progress is that there doesn't seem to be one person on the IFDA side who can (a) articulate sufficiently concrete research questions and (b) compel companies to share data with the team. The team is meeting more with member company points of contact, than the overall IFDA point of contact, and the companies clearly have different interests, motivations and willingness to participate. The lack of clear concrete research questions is somewhat inherent to the project, because no one (yet) knows what data is available to help make questions feasible to answer.

Another potential impediment to progress is that it is difficult for the team to go back to each one of the client companies and ask for more data, differently coded data, etc. For example, companies typically keep track of “termination” data but these are not terminations in the usual sense (firing for cause). Instead they are all instances when an employee separates from a company: firing, quitting for another job, retiring, etc. These reasons for termination are not in the data that has been shared by some of the companies with the team. Another example is that the data shared with the team has date of application for job applicants, but not what date the request for applications was published (so the team cannot distinguish early applicants from late ones), nor any links between applicants and subsequent employees (so the team cannot identify applicants who were subsequently employees for only a short time, vs applicants who were subsequently employees for a long time). This may in part be a power issue, and one of the teams advisors (me, Jamie or Anneka) may have to get involved to get the team the right data (if it exists). Alternatively the team could do the needs assessment just based on the data that it does or doesn’t have right now.

Despite all this I expect the team will make excellent progress on needs assessment and will also have some analyses to illustrate what might be possible with more complete data that is regularized across companies. I recommend that

1. The team articulates some useful questions about recruitment strategies. These questions should focus on relationships, such as
   1. identifying predictor variables that are already in the data, or that could reasonably be collected in the future, that predict important outcomes such as successful hiring, length of employment if hired, satisfactory performance on the job, separation other than firing, etc.
   2. Relationships between wage distributions for job categories within the food distribution industry (using data from the companies) and wage distributions for the same job categories in other industries competing for the same workers (e.g. possibly using data from BLS data, etc.)
   3. Etc.
2. The team builds the needs assessment to show what data exists, or is needed, to answer those questions. In describing data that is needed, the team should concentrate on
   1. Data that reasonably could be collected by the companies
   2. Variables that are useful for questions articulated in (a)
   3. Not just identifying variables that are needed (e.g. termination date) but also values of variables (e.g. reason for termination: firing, other job, retirement, etc.) that are needed to answer questions articulated in (a)
3. For some subset of the questions in (a) the team should either be able to either produce some illustrative analyses, and/or explain what data is missing that prevents them from answering the question(s).

The team is working very hard and putting many hours into data organization, cleaning and linking, as well as EDA. You are on track to produce a report or set of reports that will be useful for IFDA planning and policymaking in the future.

Brian Junker  
March 26, 2021