PEOPLE ANALYTICS: HR MANAGEMENT AND DATA

Please do not circulate without Prof. De Stefano's permission

Prof. Federica De Stefano[1]

Team Assignment Instructions

Logistics

During the "directed teamwork" session in class, teams will analyse the case data and draw initial insights from their analyses. The professor will guide and support teams during the session. At the end of the day, all teams should have a draft of the power point presentation of their final insights.

Teams will then have a few days, to finalize and formalize their recommendations in an 8-slide Power Point presentation (+appendix if necessary).

The Assignment

You have been asked to help PalliserCorp to identify the correct applicants to hire for their analyst job.

To allow you to make rigorous recommendations, they have provided you with two datasets. Your assignment is to analyze the data and answer the following questions (please summarize your answers in a power point presentation):

1. Which three applicants would you recommend that the company hires?
2. What formula, algorithm or heuristic did you use to pick these three applicants, and how did you arrive at this approach?
3. What concerns or caveats do you have about your recommendations?

The Datasets

Candidates012020.csv provides data on 21 different applicants. Your goal is to pick the three applicants that you expect to have the best performance at PalliserCorp

---

[1] Developed in collaboration with Matthew Bidwell (Wharton, University of Pennsylvania).

Hiringperformance012020.csv provides data on the attributes of current employees, as well as a performance rating. This data can be used to understand which characteristics best predict performance in the analyst job.

Both datasets are in "comma separated values" format. They can be opened using excel or R (or pretty much any other data analysis package that you choose.

Variable Definitions

PalliserCorp has identified the following attributes of employees and applicants that they believe may be useful in making hiring decisions

ID	This is a unique numeric identifier attached to each candidate
Age	This is candidate age in years
Gender	This is candidate gender (1= female, 0 = male)
Selective College	This variable takes a value of 1 if the candidate went to a selective college and 0 otherwise
GPA	This is the GPA that the candidate achieved during their undergraduate degree
Extra-curriculars	This is an evaluation of the value of extra-curricular activities that might indicate potential (e.g. membership of sports teams; leadership roles in voluntary organizations) It is on a 1-5 scale, where 1 is the poorest score and 5 is the best.
Modelling	These are the scores that the applicants received on a test of modelling skills that PalliserCorp implemented for all applicants and current employees
Grit, Maturity, Collaborativeness, Rigor, Ambition	These are scores that are taken from structured interviews. Each score is agreed by the interviewers, and represents their assessment of how the candidate scored on this dimension. The possible scores run from 1-5 with 1 being the worst and 5 being the best score.
Outside experience	This is the number of years that the candidate has worked in other companies. For existing employees, it is the number of years that they worked in other companies before joining PalliserCorp.
Tenure	This is the number of years that individuals have worked at PalliserCorp. It is zero for all candidates.
Performance	This is the score that current employees received on their most recent evaluation. It is on a 1-7 scale with 7 being the best possible score. (Hint – you might consider exploring what characteristics of employees predict performance).

Programs

We have also uploaded a file "R_commands_basics_NEW.R" that you can use to upload the data into R and conduct basic analyses. To get the most use out of it:

- Make sure that you put the program file in the same directory as the data. Otherwise you will need to edit line 10 in order to describe the path to the datafile.
- Experiment with different variables. We have provided example code for calculating correlations and running regressions with just two variables. You should experiment with including additional variables to gain a clearer picture of what drives performance in this context.

Please note that YOU DO NOT NEED TO USE R FOR THIS ASSIGNMENT. You can use Excel (or indeed any other package) if you are more comfortable with it. We provided this data purely as additional help for those that want to experiment with R.