PiC Challenges

Scott Shields to me 1 Jul

Hi Jamie,

Great to see you today.

Here are the challenges below - do send your work back by 08:00 on Friday 8th so Harry and I can review prior to our meeting.

FULL-STACK CHALLENGE

Given a series of datasets (in .csv format), create a simple web application (in a framework / language of your choice - use what you're comfortable with!) that a hypothetical student interested in going to university might use to assess their rough chances of getting in.

We leave the specifics to you - we really want to see the way you think given loose parameters - but we'd like the student / user to only have to put in two pieces of information about themselves. These inputs will obviously depend on the datasets you choose to use (we've provided a couple of links to the official UCAS website. Please only use these datasets in your work). The expected output is something along the lines of a percentage chance, but there's lots of room for interesting ways to dress this up! Again, we leave the specifics to you.

UCAS datasets:

Application / Acceptance (HE Provider level) - https://www.ucas.com/corporate/data-and-analysis/ucas-undergraduate-releases/ucas-undergraduate-end-cycle-data-resource-1

Application / Acceptance (applicant level) - https://www.ucas.com/corporate/data-and-analysis/ucas-undergraduate-releases/ucas-undergraduate-end-cycle-data-resource-4

Please include a link to the app you produce (hosted somewhere accessible - Heroku for example. No username and passwords please) and a link to the source code (ideally on Github).

Any questions let me know. We look forward to seeing your app!

DATA SCIENCE CHALLENGE

Your challenge is to get to the bottom of what bias exists in a recruitment process, i.e. what affects someone's chances of getting an offer - and any other insight you uncover in your analysis.

Your challenge is two-fold, one specific question and one more free-form / open question:

1. *What bias exists in the process, i.e. what will affect a candidate's chances of getting an offer (or dropping out at an intermediate stage)?*
2. *What other insight you can glean from the data?*

https://docs.google.com/spreadsheets/d/1t9VvFXdliDqPPMzItWpTCe5Pn_swaiVRKbECx8TY_x0/edit?usp=sharing

The link above is a representative set of PiC (client) data containing a candidate ID, PiC generated scores, indicators, and measures (detailed below), candidate diversity data, and the stage in the recruitment process the candidate got to.

Briefly, the scores (real numbers between 0 and 100 inclusive) generated by us:

- *"PiC Score" - this is the overall score given to a candidate, it is a linear combination of the two indicators (described below)*
- *"Achievement Indicator" - a score calculated from the relative academic performance of the candidate, the higher the score, the better the candidate did relative to their peers*
- *"Disadvantage Indicator" - a score based on the social background of the candidate, the higher the score, the more disadvantaged the candidate's social background*

In addition, we have included two lower-level "measures", both of which directly influence the Disadvantage Indicator:

- "Home Area Deprivation" - a score based on government data about how deprived / disadvantaged children growing up in various areas are on average. This is based on the candidate's postcode at age 16 (not provided). The higher the score, the more deprived / disadvantaged the candidate's background.
- "Home Area University Access" - a score based on government data around "Young Participation Rates" (essentially a number representing the proportion of young people in a given area that went to university). This is also based on the candidate's postcode at age 16. The higher the score, the worse the participation rate in their local area.

The output we would like from you is a client ready "mini-report" with your analysis and findings. As an optional extra, feel free to include any additional explanation the doesn't fit the client-side report, such as reasoning behind your methodologies, paths of analysis you initially explored but dismissed for whatever reason, etc.

We're primarily going to be impressed by two things:

1. Presentation - a clear and easy to understand (by a non-technical and technical audience) explanation of analyses
2. Approach - a logical and reasonable approach to the analysis itself