

Efficacy analytics audition project :: Oct 2016

Task

Show us your data science skills by exploring the attached dataset, and preparing a presentation of your findings and/or actionable insights you think are relevant.

We're not only interested in the **solution** but also in the **process** you went through to arrive at it. Send us a public Git repository that will include both:

- the presentation of your findings (in any form you like);
- all artifacts that will show us your thought process, for example: planning notes, code, batch scripts, tests, visualizations, documentation, etc. Anything that can show us how you approached the given task and the presentation of your findings would be useful.

Please submit a link to a public Git repository to your HR contact within 48 hours of receiving these instructions and data. Thank you!

Data dictionary

The dataset comes from a Pearson e-learning platform, an online workbook that is used alongside a paper textbook. Students complete activities, either assigned by the teacher or chosen by themselves.

- **learner_id**: anonymized student identifier
- **country**: country code of the student
- **in_course**: "t" if the student belongs to course taught by a teacher (as opposed to studying alone)
- **unit**: number or name of a unit (chapter) in the workbook
- **avg_score**: average percentage score on all activities within a given unit
- **completion**: the percentage of activities completed in a given unit, out of all activities available in that unit
- **inv_rate**: This is the extent to which a student deviates from the suggested order of activities by the pedagogy experts within a given unit. A value of zero indicates no departure from the suggested order, a value of one indicates a complete reversal of the order.