Data science internship audition project :: Feb 2018

Task

Demonstrate your data science skills by exploring a dataset from a Pearson e-learning product. Prepare a report for the product team developing this product in which you'll present actionable insights that you think would help the team improve their product.

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

- your **report** (in any form you like: slides, markdown, etc);
- all **artifacts** that will show us your **thought process**, e.g. planning notes, code, batch scripts, tests, visualizations, documentation, etc. In other words, we'd like to see anything that can **show us how you approached** this task.

Note: You can use **any tool** you like to complete this task, e.g. R, Python, Excel, etc. Since we work mainly in R, you will score **bonus points** if you show us that you can complete this task using **using R** and such **tidyverse packages** as tidyr, dplyr, and ggplot2. So even if you don't know R yet, it might be worth spending a weekend to learn the basics using such free sources as this one.

Data dictionary

The dataset comes from a Pearson e-learning platform for English language learners. It's an online workbook with automatically-graded activities that they used alongside a paper textbook. The activities are either assigned by the teacher as homework or completed voluntarily by learners as extra practice.

- **learner_id**: anonymized learner identifier
- **country**: country code of the learner
- **in_course**: "t" if the learner belongs to course taught by a teacher (as opposed to studying alone)
- **chapter**: number or name of a chapter in the workbook
- **avg_score**: a learner's avg. 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 learner 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.