Anomaly Detection Project

Presented by: Woody Hayes, Cristina Lucin & Chris Rosenberger

Anomalies in Curriculum Data: Executive Summary

1. **Which lesson appears to attract the most traffic consistently across cohorts (per program)?**
    a. PHP: HTML-css, Java I, Laravel intro, Javascript I, Javascript II
    b. Java: Javascript I, HTML-css, jquery, Java-I, Spring
    c. Data Science: Fundamentals, SQL/mySQL-overview, Classification (Overview, Scaling), Python
2. **Is there a cohort that referred to a lesson significantly more than other cohorts seemed to gloss over?**

- Data Science: Darden viewed 'Classification Overview' 1109 times, more than double all others cohorts
- Web Development (PHP/Java): Most lessons seemed balanced, search features used more by some

3. **Are there students who, when active, hardly access the curriculum? If so, what information do you have about these students?**

- Yes, though the majority of these students may have left the program before completion. Some students (user 832 accessed curriculum twice in the middle of program. Some users only accessed curriculum towards the end of the program (user 56).

5. At some point in 2019, the ability for students and alumni to access both curriculums (web dev to ds, ds to web dev) should have been shut off. Do you see any evidence of that happening? Did it happen before?

- With students being able to access other program curriculums after 2019, we have limited evidence to support that the cutoff implemented was not effective. Records located indicate that Data Science cohorts access Web Development course information as late as 2020 and beyond.

7. Which lessons are least accessed?

- PHP: Laravel, mySQL, PHP II, PHP III Data Science: Classification, Python, Pandas, Time Series Java: PHP I, PHP II, PHP III

Final Project Notebook (GitHub): https://github.com/codeupdonuts/anomaly-project/blob/main/final_project_notebook.ipynb