

Course Competition in Large-scale Learning Platforms

Chrysanthos Dellarocas, Bin Gu, Boston University

Yutong Han, Peking University

Abstract

Learning platforms are digital marketplaces that connect learners with courses and programs offered by multiple providers. Examples include Coursera, edX, Pluralsight, Skillsoft, Udacity, and Udemy. These platforms have the potential to disrupt the education industry in ways similar to how digital platforms have disrupted other industries such as hospitality, transportation, journalism, and retail. The purpose of this study is to explore the dimensions of competition in a large learning platform that currently hosts almost a quarter of a million courses in a wide variety of topics. In popular topic categories (e.g. Python, Java) several hundred, or even thousand, courses, essentially on the same subject, compete for learners' attention. Nevertheless, when one looks at how enrollment distributes among those courses, one sees a steep power law where enrollment is concentrated at a small number of popular “blockbuster” courses. Understanding how prospective learners select courses among so many choices, and how such “blockbuster” courses emerge, is a question that is interesting in itself, and may offer important insights that can inform how education institutions can best navigate the coming era of education platforms.

Background

Learning platforms are digital marketplaces that connect learners with courses and programs offered by multiple providers. Examples include Coursera, edX, Pluralsight, Skillsoft, Udacity, and Udemy. These platforms have the potential to disrupt the education industry in ways similar to how digital platforms have disrupted other industries such as hospitality, transportation, journalism, and retail. (Dellarocas and Hagiu 2022).

Learning platforms are already prevalent for non-degree courses (e.g., Udemy, Skillsoft, LinkedIn Learning) and professional education (e.g., Emeritus, Masterschool, Udacity), and some are already making inroads into the degree program space (e.g., Coursera, edX).

As other industries have already found out (Hagiu and Wright 2021), the environment on large platforms becomes increasingly competitive over time: This is the result of natural growth (platforms with more customers also attract more merchants) and can also stem from platform design changes (e.g., the recommendation algorithm places a higher weight on offerings with lower prices). Moreover, growing platforms do not just attract existing brands; they also make it possible for a long tail of new and highly

specialized offerings by formerly “amateur” providers to become viable. The entry of such new providers in formerly brand-dominated markets (e.g., Airbnb) further intensifies competition.

Research Questions

The purpose of this study is to explore the dimensions of competition in a large learning platform that currently hosts almost a quarter of a million courses in a wide variety of topics ranging from data science to languages, to wellness. Many of the most popular topics are related to Information Technology training. Table 1 lists the number of available courses on the top five Information Technology-related topics. The most popular topic is Python programming, with close to 5,000 available courses in February 2023. Nevertheless, when one looks at how enrollment distributes among those courses, one sees a steep power law where enrollment is concentrated at a small number of popular “blockbuster” courses (Figure 1). Understanding how prospective learners select courses among so many choices, and how such “blockbuster” courses emerge, is a question that is interesting in itself, and may offer important insights that can inform how institutions, such as ours, can best navigate the coming era of education platforms.

In this work, we aim to:

- uncover the factors that propel a course to the top of the learners’ preference in a crowded category
- understand the dimensions of course differentiation that matter to learners
- distill strategies that education providers can employ to successfully position their courses, especially in crowded categories with many other competing offerings.

The key methodological challenges we face are 1) to develop a machine learning approach to identify dimensions of course differentiation, 2) to develop a statistical model to capture user choice process in a large-scale platform and 3) to leverage new technology introductions as natural experiments to study the dynamics of course differentiation and entry strategies.

Data

We have access to a unique data set from a large learning platform. That particular platform sells individual courses directly to consumers, as well as bundles of courses to organizations. Individual courses are priced from \$9.99 to \$199.99, with an average course price of \$20. Our data set contains full historical information about all courses published on that platform since 2011. This information includes general course data (course title, course topics, course syllabus, etc.), learner data (how many enrolled and when), and course reviews data.

Preliminary results

We constructed a sub-dataset containing the 2,379 courses on the topic of Python that were published before 2021. We conducted a linear regression where the independent variable is a course's relative rank within the sub-dataset in terms of the total learners who enrolled in the course during the calendar year 2022 (i.e., at least one year after the publication of the most recent course in the sub-dataset). Larger rank numbers map to larger numbers of course enrollments.

Table 2 lists the results of the regression. We find that a higher course rank (i.e., higher enrollments during 2022) has a statistically significant positive relationship with:

- Free courses
- Courses with higher average learner ratings
- Courses with higher total enrollments before 2022
 - Ratings and past enrollments are visible to new prospective learners who appear to take both into consideration when selecting courses
- Courses with earlier publication dates
 - It seems that learners have a preference for “seasoned” courses that have been around longer; this effect is on top of the effect of reviews and past enrollments and will be explored in depth in our ongoing work
- Courses with more recent update dates
 - Since technologies change rapidly, learners appear to seek up-to-date courses
- Higher course instructor ratings at the time of the course's publication
- Higher total number of enrolled students in the instructor's other courses
 - In the platform we study, courses are not created by institutions with recognizable brands, but rather, by individuals or small organizations. An instructor's “brand” is her ratings and the number of enrolled students in previously launched courses.

Interestingly, we find a negative relationship between rank and the number of an instructor's other courses, which seems to suggest that learners avoid instructors who spread themselves too thin.

Finally, we observe a curvilinear relationship between rank and course length. Learners tend to avoid very short and very long courses, preferring courses that have the “right” length to offer them substance, while keeping them engaged. It appears that the sweet spot for Python courses is a course duration of about 1,000 to 3,000 minutes or 15 to 45 hours of video.

We conducted similar regressions with courses in other popular IT topics (SQL, Power BI) and found consistent results.

Ongoing work

We have barely scratched the surface of course competition in education platforms. In our ongoing work, we are looking at the effect of additional course features, such as keywords in a course’s title and description, the choice of topics covered in a course’s syllabus, and information contained in the trailer and sample course videos. We are also looking at how courses in crowded categories can differentiate themselves by bridging across topics. Finally, we examine how courses launched over time when new topics (e.g. new information technologies) emerge position themselves to stand out from the competition.

References

Chrysanthos Dellarocas and Andrei Hagiu. How Higher Ed Can Contend with Learning Platforms. Harvard Business Review, December 21, 2022.

Andrei Hagiu and Julian Wright. Don’t Let Platforms Commoditize your Business. Harvard Business Review, May-June 2021.

Tables and Figures

Topic	Number of courses
Python	4958
Excel	3235
Microsoft Certification	2512
Javascript	2360
Wordpress	2343

Table 1. Number of available courses in the top five information technology topics (as of February 2023).

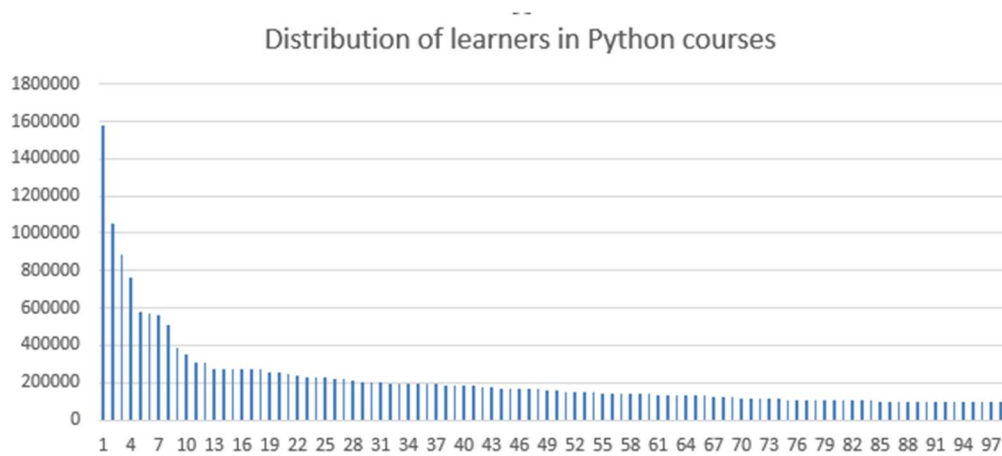


Figure 1. Total number of enrolled learners in the top 100 courses on Python

Variable	Estimate	Std_Error	t_value	Pr(> t)	
(Intercept)	-1.10E+04	1.09E+03	-10.116	<2e-16	***
Average course rating	1.96E+02	1.97E+01	9.949	<2e-16	***
Paid course	-8.91E+02	9.87E+01	-9.034	<2e-16	***
Past course enrollments	1.738E-03	1.242E-04	13.990	<2e-16	***
Published date	-1.65E-01	2.40E-02	-6.867	8.38E-12	***
Last updated date	4.23E-01	2.32E-02	18.226	<2e-16	***
All levels course	1.27E+02	5.14E+02	0.248	0.804345	
Beginners course	5.05E+01	5.14E+02	0.098	0.921726	
Intermediate course	6.82E+01	5.19E+02	0.131	0.895474	
Advanced course	9.61E+01	5.14E+02	0.187	0.851681	
Course length	2.71E-01	3.25E-02	8.332	<2e-16	***
Course length ^ 2	-3.39E-05	6.85E-06	-4.946	8.13E-07	***
Has certificate	-5.89E+00	9.26E+01	-0.064	0.949291	
Has close caption	8.50E+01	3.21E+01	2.646	0.008199	**
Is marketing boost agreed	1.44E+01	1.37E+02	0.105	0.916262	
Instructor's other course rating	1.00E+02	2.90E+01	3.454	0.000561	***
Instructor's other courses	-5.14E-01	7.42E-02	-6.927	5.56E-12	***
Instructor's other course enrollments	3.19E-04	5.81E-05	5.494	4.36E-08	***
Course price	-1.94E+00	8.23E-01	-2.353	0.018705	*

Residual standard error: 511.3 on 2327 degrees of freedom
 Multiple R-squared: 0.4561, Adjusted R-squared: 0.4456
 F-statistic: 43.37 on 45 and 2327 DF, p-value: < 2.2e-16

Table 2. Determinants of a Python course's rank in terms of total enrollments during 2022
 (higher rank implies more learners)