

IDENTIFYING THE DROP-OUT DECISION MOMENT IN PROGRAMMING MOOC STUDENT PROGRESS DATA

Jakub Swacha

Department of IT in Management, University of Szczecin (Poland)

Abstract

Although Massive Open Online Courses (MOOCs) on programming are considered a proven way of teaching this difficult subject to large audiences, they also suffer from high dropout rates, often exceeding 90%, which leads to ineffective allocation of instructors' effort, wasted on providing support to students who are not willing to finish the course anyway. This paper analyzes course progress data gathered from students attending a MOOC on introduction to programming in Python to pinpoint the course sections at which learners most often drop out as well as investigate the correlation between the students' performance at earlier course sections and their dropping out at the later sections. The key findings include: large differences in dropout rates among respective course sections; the lack of course sections that the students' low performance at which could be used as a good predictor of their later dropout; and the relatively high variety in the predictability of dropout at respective course sections depending on the students' performance at the earlier course sections. The findings bring practical implications for both MOOC developers and instructors.

Keywords: *Student dropout analysis, student dropout prediction, programming MOOC, MOOC management.*

1. Introduction

Massive Open Online Course (MOOC) is defined as “an online course open to anyone without restrictions, usually structured around a set of learning goals in an area of study, which often runs over a specific period of time on an online platform which allows interactive possibilities that facilitate the creation of a learning community” (EC, 2014, p. 2). The MOOC's advantages made it a highly popular form of distance learning, with over 220 million users worldwide having enrolled in at least one MOOC in 2024, according to a recent report (MGR, 2026). Although it has been confirmed that MOOC-based education is capable of significantly improving academic achievements and motivation (Wang and Li, 2022), it has a weak side, which consists in the high dropout rate (Huang, Jew, & Qi, 2023), which may be as high as 95% (Alghamdi, Soh, & Li, 2025, and works cited therein). Despite this phenomenon has been the subject of many versatile studies (Cişmaşu, Cibu, Cotfas, & Delcea, 2025), there are many aspects that require further investigation. This paper contributes to this vein of research, taking into focus the example of a computer programming MOOC, and investigating students' progress in the course to reveal: the topical sections at which learners most often drop out and the potential relationships between students' low performance at specific course sections and their later drop-out decision.

2. Data and methods

2.1. Data source

The study is based on course progress data obtained from a MOOC on introduction to programming in Python 3 whose design and evaluation has been described in our earlier work (Swacha, 2021). The MOOC runs on the Open edX platform (Open edX, 2026) which provides access to individual grades attained by every student at the end of each course section. In the course, there are 14 modules, titled respectively: First contact with the Python language, Character strings, Programs, Sequences, Loops, Sets and dictionaries, Functions, Object-oriented programming, Python standard modules, Data processing, Algorithms in Python, Storage of data, Use of PYPI modules, and Python in practical applications.

Although as of February 2026 there is an eighth edition of the course ongoing, due to updates of the platform, which is maintained by a third party, access to a part of the student progress data has been lost. Consequently, we were able to retrieve complete data of two course editions only and one edition's partial data, altogether comprising 5542 registered participants.

2.2. Data preprocessing

As is the case with many other MOOCs, most of the participants leave the course early, not passing even its first section. While such cases can be interesting for research, e.g., in the context of zero-time dropout predictors (Swacha, Muszyńska, 2023), they are of no use to this study, which seeks to identify course sections where the students' performance at can be a good indicator of their later dropping out. Consequently, all records of such early dropouts have been removed from the database, leaving 1326 student records for the analysis; these include 974 students who dropped out before the final test. The overall dropout rate in the course dataset analyzed (including the early dropouts) was 87.5%.

2.3. Data analysis

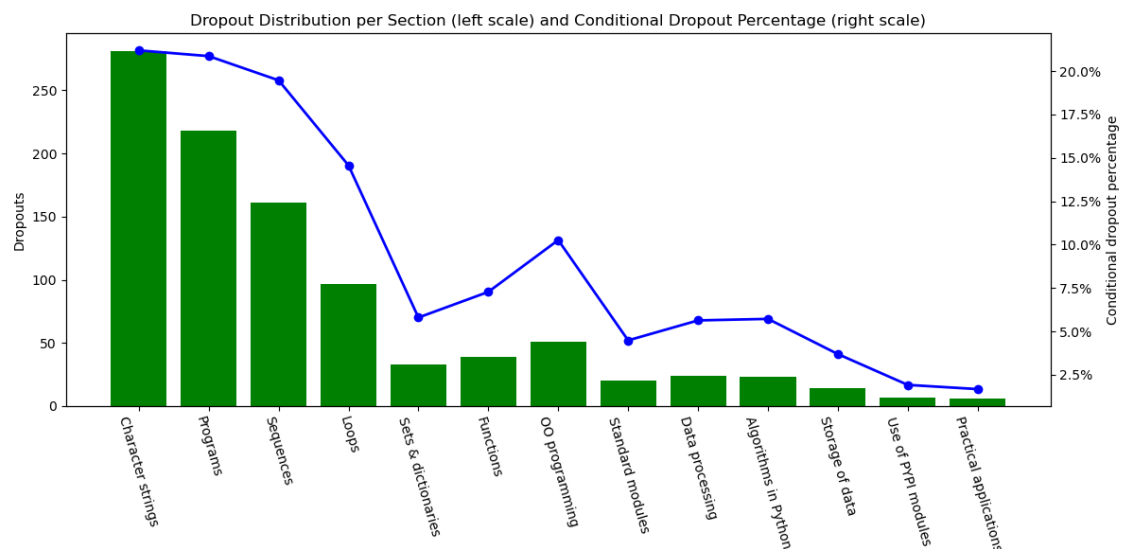
Data preprocessing and analysis have been performed using scripts written in the Python language. The main results were obtained using visual analysis of charts generated with the help of Python's Pyplot and Seaborn libraries. To visualize the heatmap, the Pearson correlation coefficient has been calculated using Python's Pandas library. The Python code for data preprocessing and visualization has been developed with the assistance of Microsoft Copilot.

3. Results

3.1. Dropouts per course section

Figure 1 presents bars indicating the number of students dropping out of the course at its specific sections and a line indicating the percentage of students dropping out at a specific section, considering only the students still remaining in the course at that point. Note that the chart begins with section 2 as the early dropouts (which happened at section 1) have been removed during data preprocessing (and their inclusion in the chart would make it unreadable due to the out-of-proportion number of early dropouts).

Figure 1. Number and percentage of dropouts per course section (covering sections 2-14).

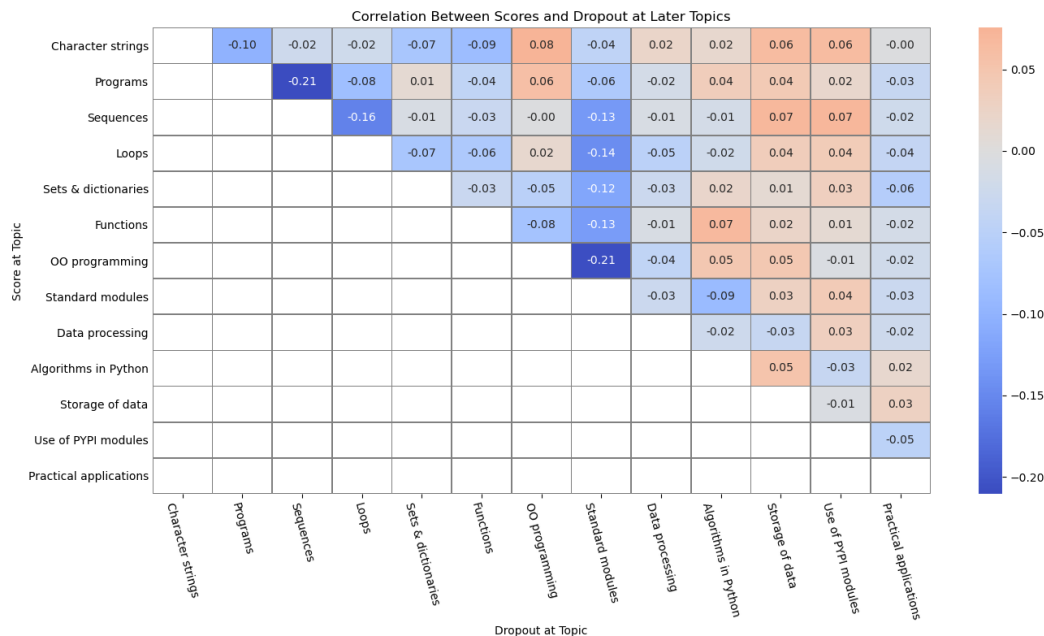


As can be seen in Figure 1, the first five bars (representing course sections 2-6) follow a clearly decreasing pattern which can be caused by both the increasing difficulty of subsequent programming topics and the sufficient time spent in the course to realize it is not the right course for those who might have imagined it differently. The height of the remaining bars varies, which can be, at least to some extent, explained by the varied difficulty of the specific course sections. Note, e.g., the relatively high dropout percentage for the OO programming section and low for the Standard modules section.

3.2. Course section performance vs. later dropouts

Figure 2 presents a heatmap showing the correlation between the score received by students in the test covering the respective course section's topical scope and the dropout happening at any of the later course sections. Note that negative correlation indicates that a higher score attained at a specific course section reduces dropout likelihood at the other section, whereas positive correlation indicates that a higher score attained at a specific course section increases dropout likelihood at the other section.

Figure 2. Correlation between students' learning effects at specific course section and dropout at a specific section.



While the level of correlation shown in Figure 2 is very weak in most cases, an unexpected observation can be made: that the patterns tend to form vertically, not horizontally. This means that despite the key idea motivating for the presented study was the expectation that there could be course topics the students' performance at which could indicate their higher probability of later dropout and no such cases were found in the analyzed data, instead, course topics have been identified the students' dropout at which could be better predicted by their performance at previous course sections (in the studied course: Loops, Functions, and Standard modules). Note that, considering also Figure 1, these are not course sections for which the dropout rate was particularly low or high (in which case, a simple explanation for this finding could be provided).

In contrast, negligible yet positive correlation is visible primarily in two columns corresponding to two late course sections: Storage of data and Use of PYPI modules. Such a counter-intuitive result suggests that dropout at these sections of the course is most probably not caused by their difficulty (as the students performing better in earlier course sections were slightly more prone to drop out at these sections).

4. Discussion

The study adds a new aspect to the prior research on MOOC dropout. While the earlier studies were focused on why the students drop out (Bezerra & Silva, 2017), this one strives to answer when it is most probable to happen. While the general trend of dropout percentage is decreasing with course progress, the results presented in section 3.1 clearly show its notable variation, most probably caused by the difference in difficulty of the respective course sections – itself confirmed by prior study on the number of students' solution submissions till one was accepted (Szydłowska, Miernik, Ignasiak, & Swacha, 2022). This provides a practical implication for MOOC course designers that they should avoid rapid changes in the difficulty of subsequent course sections, as it may lead to students dropping out.

More unexpected are the results presented in section 3.2. Regarding the very weak correlation, it must be noted that it is not contradictory to prior research (e.g., Panagiotakopoulos, Kotsiantis, Kostopoulos, Iatrellis, & Kameas, 2021) showing high correlation between assessment results and dropout, as those authors consider dropout anywhere in the course, not in a specific section later to the one at which the student's learning progress was assessed. While the results have not indicated course topics the students'

performance at which would be a particularly good predictor of their later dropout, they have indicated course topics the students' dropout at which could be predicted by students' performance at prior course sections easier than for other topics. This provides a clear practical implication for MOOC course instructors that they should identify such course sections and provide additional support for the low-performing students at those sections to reduce dropout risk.

5. Conclusions and future work

The study contributes both to the research on MOOC dropout, indicating that dropout predictability varies depending on course sections, and to the practice of managing MOOCs, for which it provides simple means for identifying course sections at which low-performing students are more prone to dropout and should then be given additional support.

We believe that this study will motivate further research to confirm the presented findings (using data from other courses) and put them in context (by interviewing students to identify their dropout motives). Another interesting direction of future work is investigating the factors that contributed to the dropout decision at specific course section, thus extending the vein of studies which investigated the factors that affect MOOC effectiveness (e.g., Yu, Xu, & Sukjairungwattana, 2024; Huang, & Qi, 2025).

Acknowledgements

The presented work has been co-financed by the Minister of Science of Poland under the Regional Excellence Initiative program.

References

- Alghamdi, S., Soh, B., & Li, A. (2025). A Comprehensive Review of Dropout Prediction Methods Based on Multivariate Analysed Features of MOOC Platforms. *Multimodal Technologies and Interaction*, 9(1), 3. <https://doi.org/10.3390/mti9010003>
- Bezerra, L. N. M., & Silva, M. T. da. (2017). A review of literature on the reasons that cause the high dropout rates in the MOOCs. *Revista Espacios*, 38(5), 11.
- Cişmaşu, I.-D., Cibu, B. R., Cotfas, L.-A., & Delcea, C. (2025). The Persistence Puzzle: Bibliometric Insights into Dropout in MOOCs. *Sustainability*, 17(7). <https://doi.org/10.3390/su17072952>
- EC. (2014). *Report on Web Skills Survey: Support Services to Foster Web Talent in Europe by Encouraging the use of MOOCs Focused on web Talent—First Interim Report*. European Commission. Retrieved February 9, 2026, from: <https://silo.tips/download/report-on-web-skills-survey>
- Huang, H., Jew, L., & Qi, D. (2023). Take a MOOC and then drop: A systematic review of MOOC engagement pattern and dropout factor. *Heliyon*, 9(4), e15220. <https://doi.org/10.1016/j.heliyon.2023.e15220>
- Huang, H., & Qi, D. (2025). Is MOOC really effective? Exploring the outcomes of MOOC adoption and its influencing factors in a higher educational institution in China. *PLOS ONE*, 20(2), e0317701. <https://doi.org/10.1371/journal.pone.0317701>
- MGR. (2026). *Massive Open Online Course (MOOC) Market Size, Share, Growth, and Industry Analysis, By Type (xMOOC Platforms, cMOOC Platforms), By Application (K-12 Education, University Education, Adult and Elderly Education, Corporate Training), Regional Insights and Forecast to 2035*. Market Growth Reports. Retrieved February 9, 2026, from: <https://www.marketgrowthreports.com/market-reports/massive-open-online-course-mooc-market-113506>
- Open edX (2026). *The Platform – Open edX*. Retrieved February 9, 2026, from: <https://openedx.org/the-platform/>
- Panagiotakopoulos, T., Kotsiantis, S., Kostopoulos, G., Iatrellis, O., & Kameas, A. (2021). Early Dropout Prediction in MOOCs through Supervised Learning and Hyperparameter Optimization. *Electronics*, 10(14), 1701. <https://doi.org/10.3390/electronics10141701>
- Swacha, J. (2021). Teaching Python programming with a MOOC: Course design and evaluation. *Proceedings of the Thirty-Seventh Information Systems Education Conference*, 131–137. Retrieved February 9, 2026, from: <http://proceedings.isecon.org/download/3xgemqw9tdmdi9u7rftn>
- Swacha, J., & Muszyńska, K. (2023). Predicting Dropout in Programming MOOCs through Demographic Insights. *Electronics*, 12(22), 4674. <https://doi.org/10.3390/electronics12224674>

- Szydłowska, J., Miernik, F., Ignasiak, M. S., & Swacha, J. (2022). Python Programming Topics That Pose a Challenge for Students. *Third International Computer Programming Education Conference (ICPEC 2022)*, *OASICS*, 102, 9. <https://doi.org/10.4230/OASICS.ICPEC.2022.7>
- Wang, M., & Li, G. (2022). A Meta-Analysis of MOOC-Based Academic Achievement, Engagement, Motivation, and Self-Regulation During the COVID-19 Pandemic: *International Journal of e-Collaboration*, 18(1), 1–17. <https://doi.org/10.4018/IJeC.301260>
- Yu, Z., Xu, W., & Sukjairungwattana, P. (2024). A meta-analysis of eight factors influencing MOOC-based learning outcomes across the world. *Interactive Learning Environments*, 32(2), 707–726. <https://doi.org/10.1080/10494820.2022.2096641>