



CY TECH SCIENCES ET TECHNIQUES

Learning Analytics of MOOCs: A systematic review and meta-analysis

Author

Buu NGUYEN

December 2021

Contents

1	Introduction	2
2	Methods	2
3	Results	3
3.1	Observation of learners' engagement in MOOC Effectuation	3
3.2	Viewed videos between Genders	4
3.3	Viewed videos between Country's HDI groups	6
3.4	Genders and Country's HDI groups	8
3.5	Course completion and Viewed videos between Gender and Country's HDI group	10
4	Discussion	14
	References	15

List of Figures

1	Proportion of learners' engagement level	4
2	Differences in Viewed videos between Male and Female	6
3	Differences in Viewed videos across Country's HDI levels	8
4	Mosaic plot of Gender and Country's HDI	9
5	Completion odds ratios of Gender and Country's HDI	11
6	Survival plot between Country's HDI groups	12
7	Survival plot between Male and Female learners	13

List of Tables

1	Number of registrants by engagement level and iteration	3
2	Independent sample t-test	5
3	Welch ANOVA test	7
4	Levene's test	7
5	Games-Howell post-hoc comparisons	7
6	Chi-Square test	9
7	Course completion and Viewed videos to Quit	10

1 Introduction

It has been almost a decade since the evolution of learning analytics, a field of research and practice that collects data generated by learners during their learning interactions and analyzes them. Learning analytics serves to optimize learning and leverages decision-making related to learning, teaching, and educational management. As a result, governments, universities, and MOOCs provider organizations have collected millions of terabytes of data about learners and how they learn. In this study, we focus on analyzing data of online learners. Online learning has shown significant growth over the last decade, as the internet and education combine to provide people with the opportunity to gain new skills (Koksal, 2020). Numerous online learning platforms such as Coursera, MITx, edX, Udacity, Udemy, Lynda, and Skillshare serve millions of learners.

2 Methods

The data was collected from [MOOC Effectuation](#), a five weeks-long entrepreneurship course taught by Professor Philippe Silberzahn, EMLYON Business School. It was hosted by a MOOC agency that used the open-source LMS Canvas from Instructure. Currently, it is organized on Coursera.

All three iterations of usages data and surveys were merged using dplyr library in R. Participants were categorized based on their level of engagement based on the typology of learners (Kizilcec, Piech, & Schneider, 2013):

- **Completers** obtained a certificate by passing an exam and submitting an assignment.
- **Disengaging learners** submitted at least one quiz or assignment but did not complete the course.
- **Auditing learners** had viewed more than 10 percent of available videos but did not submit any quiz or assignment.
- **Bystanders** fell below the 10 percent threshold thus concluded as not participate in the course.

This study was wrapped up in R. Inferential statistical analysis was performed using hypothesis tests such as Student’s t-test, ANOVA, and Chi-Square. Logistic Regression model was used to explore the Odd Ratios of completing the course between categorical groups. In addition, survival analysis was implemented to investigate the viewed videos between categorical groups using Kaplan-Meier method.

3 Results

3.1 Observation of learners' engagement in MOOC Effectuation

	1 st iteration	2 nd iteration	3 rd iteration
Bystanders	3725	1901	2146
Auditing learners	587	302	311
Disengaging learners	2323	1191	1042
Completers	2351	780	753
Total	8986	4174	4252

Table 1: Number of registrants by engagement level and iteration

As seen in Table 1, we observe that number of registrants dropped significantly from the first iteration to the following iterations, from 8986 to 4174 and 4252.

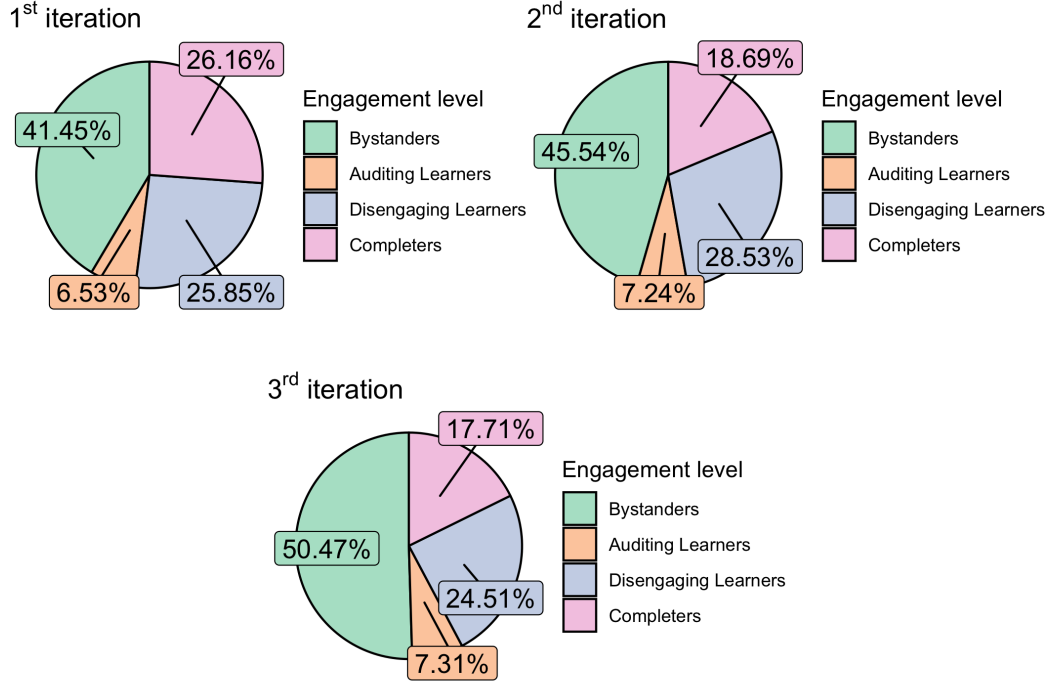


Figure 1: Proportion of learners' engagement level

From Figure 1, we can see that the majority of people who registered for the MOOC Effectuation were Bystanders. Their proportion was 41.45% in the first iteration and increased to 45.54% and 50.47% in the second and third iteration, respectively.

In all three iterations, the proportion of Auditing and Disengaging learners was stable at around 7% and 26%.

In the first iteration, 26.16% of learners finished the MOOC and obtained a certificate. However, Completers' proportion diminished to 18.69% in the second iteration and 17.71% in the third iteration.

3.2 Viewed videos between Genders

We want to investigate if the the number of viewed videos differs across Male and Female learners. We have the hypothesis:

H_1 : There is a significant difference in Viewed videos between Male and Female learners.

.y.	group1	group2	n1	n2	statistic	df	p
last.video	Male	Female	6103	2990	0.23	5764.39	0.82

Table 2: Independent sample t-test

An Independent-samples t-test was conducted to compare the Viewed videos for Male and Female (See Table 2). There were no significant differences ($t(5764.39) = 0.23$, $p = 0.82$) in scores for Male ($M = 20.84$, $SD = 13.46$) and Female ($M = 20.77$, $SD = 13.91$). The magnitude of the differences in the means (mean difference = 0.07, 95% CI: -0.53 to 0.67) was very small. Hence, $\mathbf{H_1}$ was not supported. Figure 2 summarizes the result.

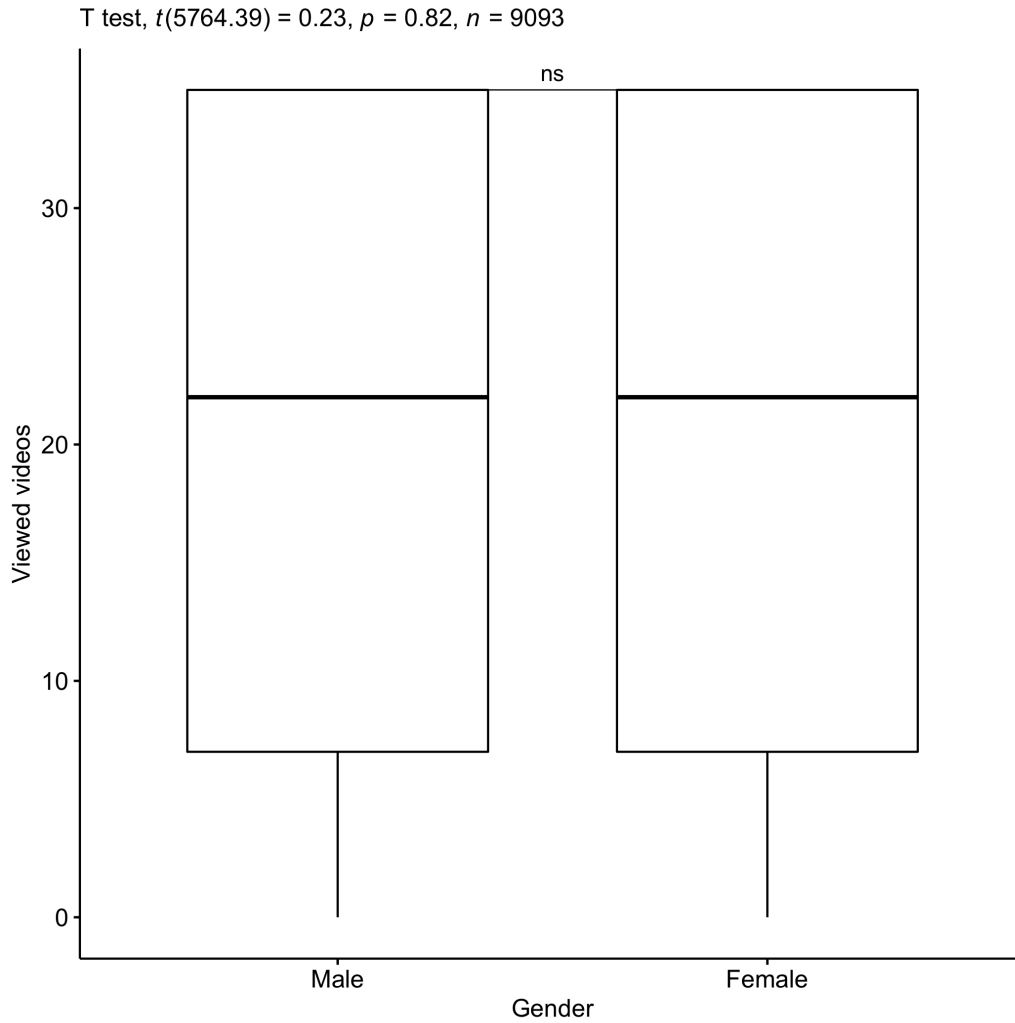


Figure 2: Differences in Viewed videos between Male and Female

3.3 Viewed videos between Country's HDI groups

We want to investigate if the the number of viewed videos differs across different levels of Country's HDI (Very High, Intermediate, and Low). We have the hypothesis:

H₁: There are significant differences in Viewed videos across different Country's HDI levels.

.y.	n	statistic	DFn	DFd	p	method
last.video	8963	83.45	2.00	1310.25	****	Welch ANOVA

Table 3: Welch ANOVA test

A Welch one-way ANOVA test was performed to evaluate if the the number of viewed videos differs across different Country's HDI levels. Learners were divided into three groups: Very High ($n = 7264$), Intermediate ($n = 667$), and Low ($n = 1032$). The results from Table 3 suggest that the the number of viewed videos of the groups differ significantly ($F(2, 1310.25) = 83.45$, $p < 0.0001$). Hence, H_1 was supported.

df1	df2	statistic	p
2	8960	11.85	***

Table 4: Levene's test

Since the Levene's Statistic from Table 4 is significant ($p < 0.001$), the equal variance was not assumed. Therefore, to check for individual differences between groups, post-hoc comparisons were assessed using Games-Howell.

group1	group2	estimate	conf.low	conf.high	p
Very High	Intermediate	-3.57	-4.88	-2.25	****
Very High	Low	-5.03	-6.03	-4.04	****
Intermediate	Low	-1.47	-3.03	0.09	0.071

Table 5: Games-Howell post-hoc comparisons

The test from Table 5 indicated that the mean score for Very High group ($M = 21.67$, $SD = 13.58$) was significantly different from Intermediate group ($M = 18.10$, $SD = 13.85$) and Low group ($M = 16.63$, $SD = 12.64$). The mean differences were significant at 0.05 level. However, no significant differences were detected between Intermediate group and Low group. Figure 3 summarizes the result.

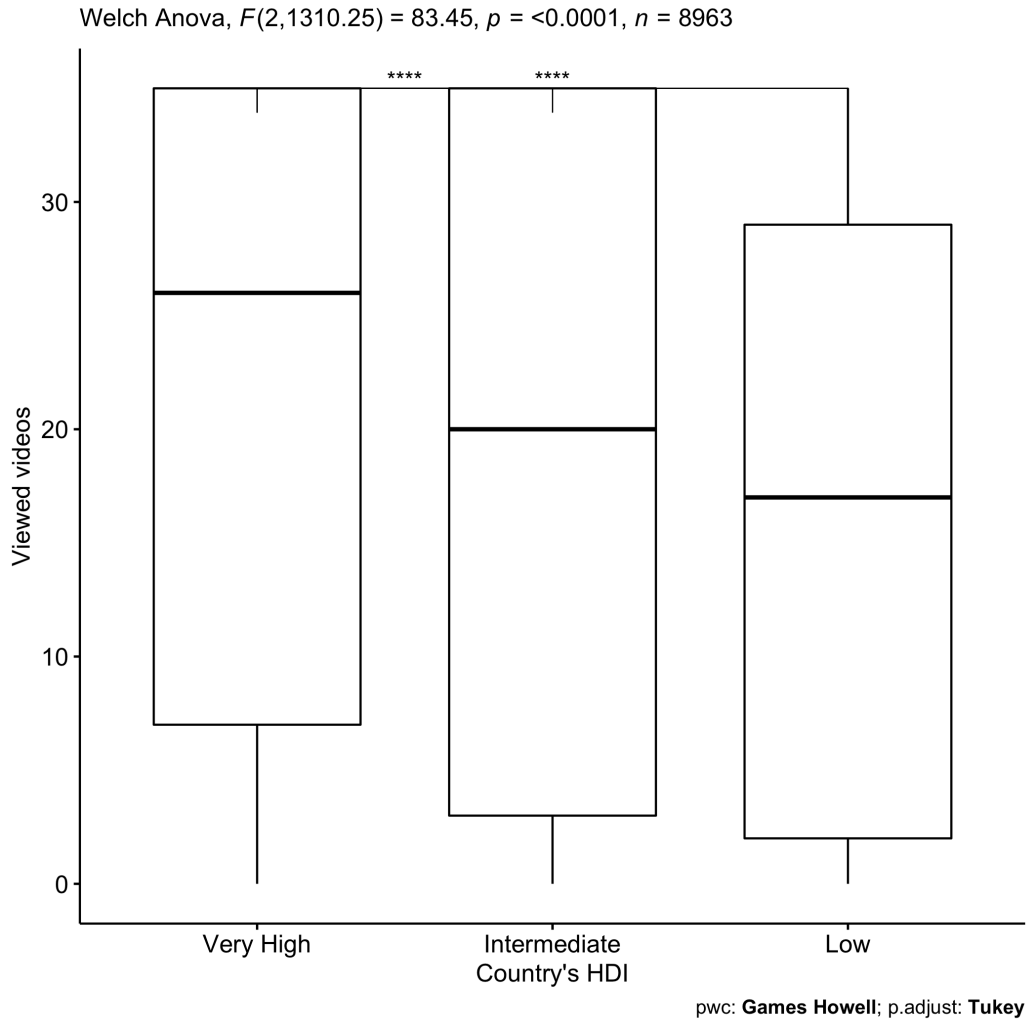


Figure 3: Differences in Viewed videos across Country's HDI levels

3.4 Genders and Country's HDI groups

From Figure 4, we detect 4 noticeable associations: Male-Very High, Male-Low, Female-Very High, and Female-Low. We use a Chi-square test of independence to evaluate if these associations are significant or not.

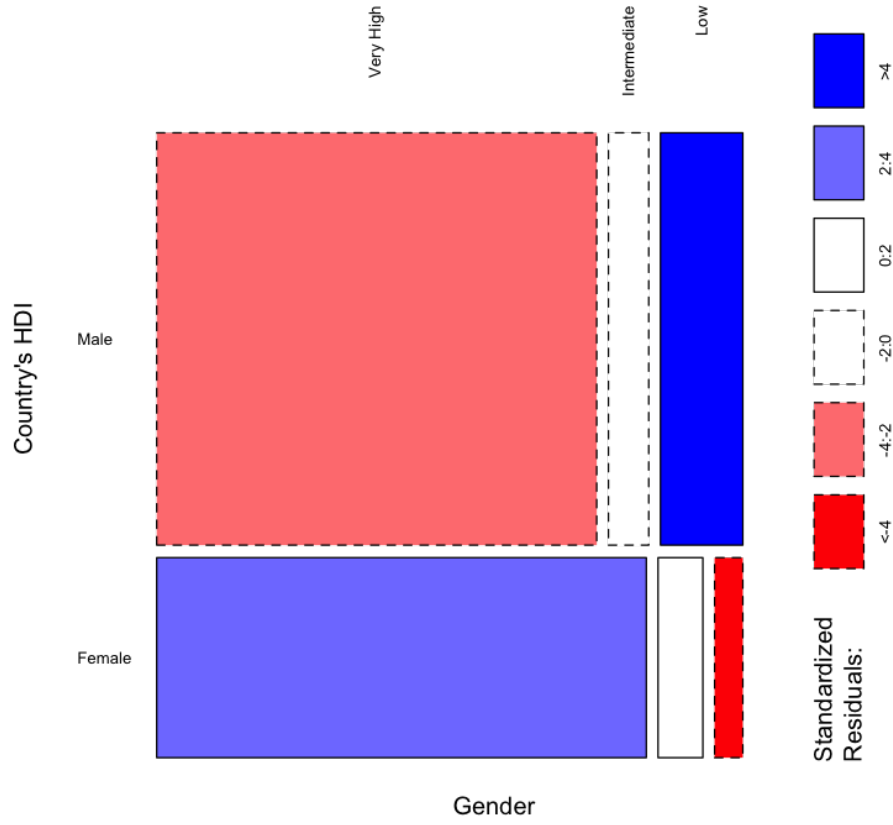


Figure 4: Mosaic plot of Gender and Country's HDI

We want to identify the association between gender of learners and their country's HDI. We have the hypothesis:

H_1 : There is a significant association between Gender and Country's HDI group.

n	statistic	p	df	method
8957	179.05	****	2	Chi-square test

Table 6: Chi-Square test

Chi-Square statistics were used to examine association between learners' gender and their country's HDI (See Table 6). There was significant relationship at 0.05 significance level between Gender and Country's HDI group ($\chi^2 = 179.05$, $df = 2$, $p < 0.0001$). Hence, H_1 was supported.

3.5 Course completion and Viewed videos between Gender and Country's HDI group

Characteristic	Course completion			View videos to Quit		
	OR ¹	95% CI ¹	p-value	HR ¹	95% CI ¹	p-value
Country_HDI						
Very High	—	—		—	—	
Intermediate	0.68	0.57, 0.80	<0.001	1.41	1.28, 1.55	<0.001
Low	0.42	0.36, 0.49	<0.001	1.76	1.63, 1.90	<0.001
Gender						
Male	—	—		—	—	
Female	0.99	0.90, 1.08	0.8	1.02	0.96, 1.08	0.6
¹ OR = Odds Ratio, CI = Confidence Interval, HR = Hazard Ratio						

Table 7: Course completion and Viewed videos to Quit

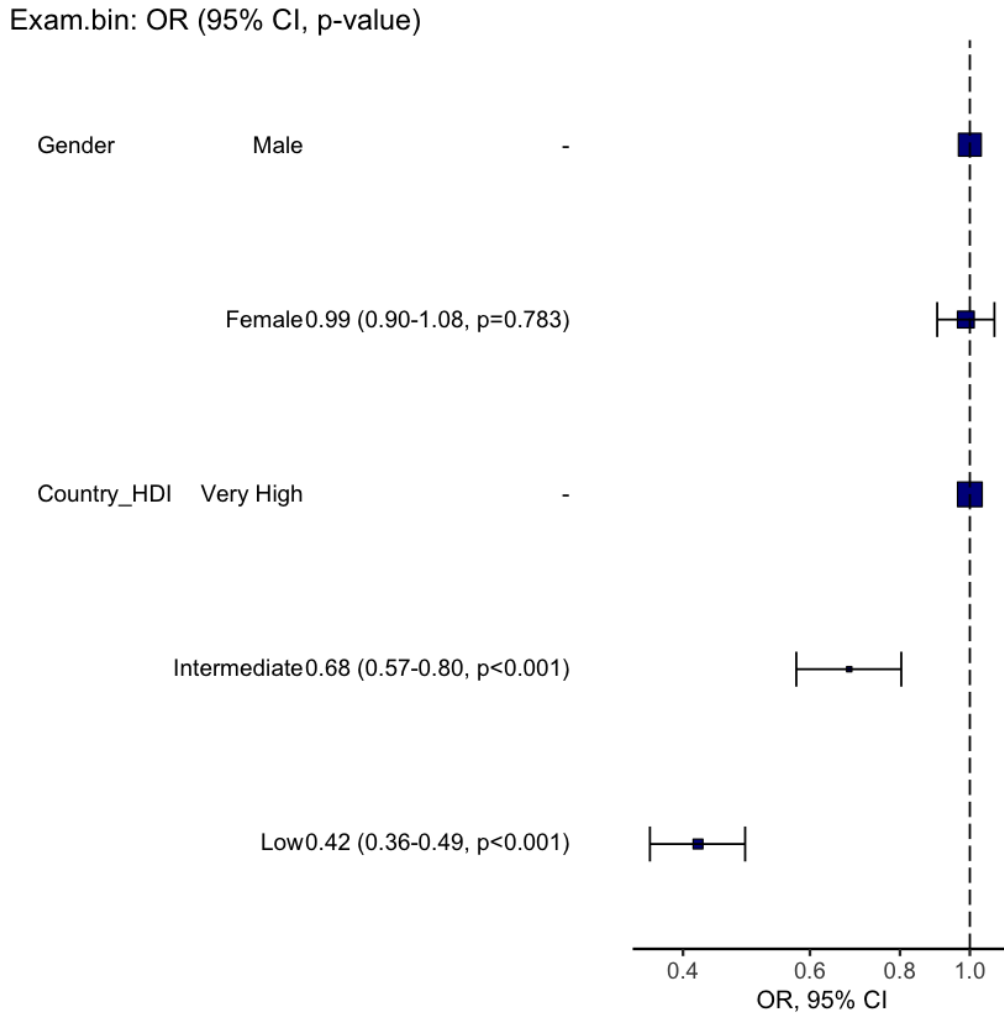


Figure 5: Completion odds ratios of Gender and Country's HDI

From Table 7 and Figure 5, we see that the difference in accomplishment's odds ratio between Male and Female learners was insignificant ($OR = 0.99$, $p = 0.783$). In addition, taking learners from Very High HDI countries as references, the odds of completing the course of learners from Intermediate HDI countries and Low HDI countries were smaller by a factor of 0.68 ($OR = 0.68$, $p < 0.001$) and by a factor of 0.42 ($OR = 0.42$, $p < 0.001$), respectively.

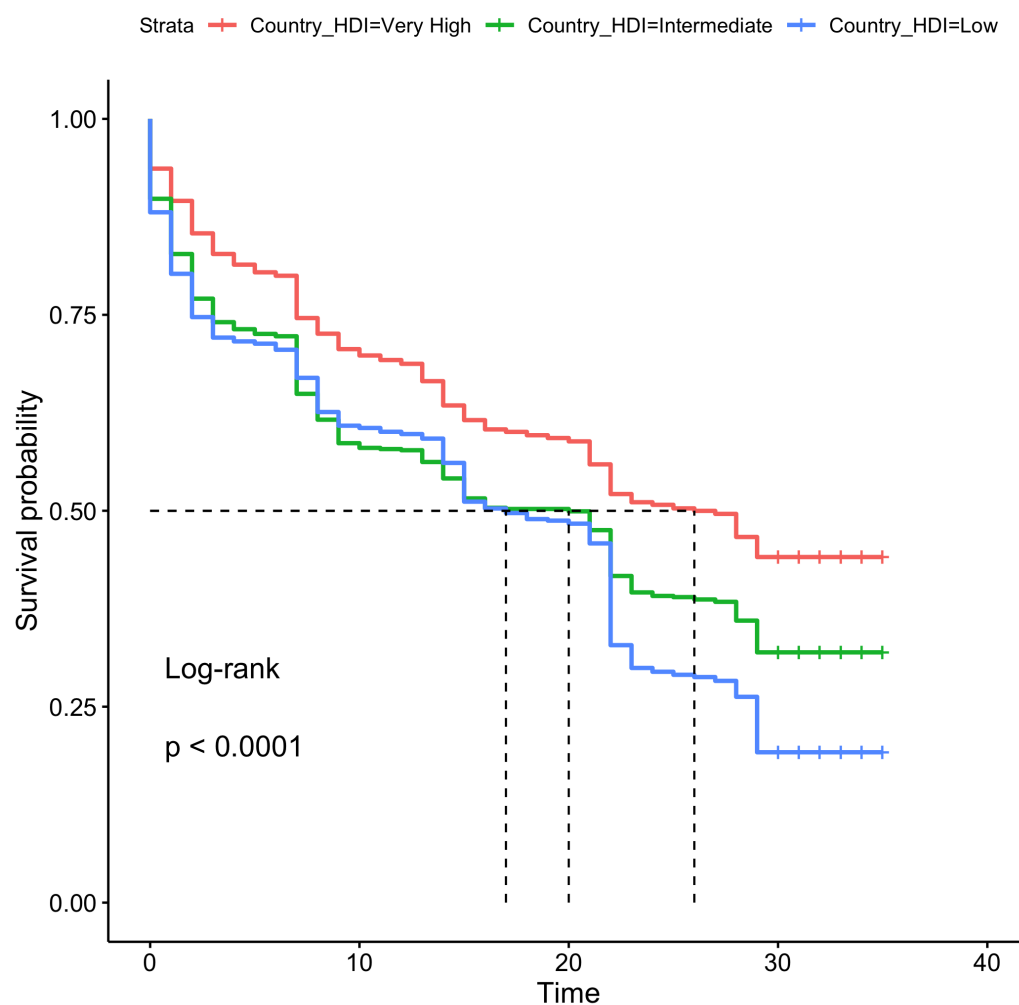


Figure 6: Survival plot between Country's HDI groups

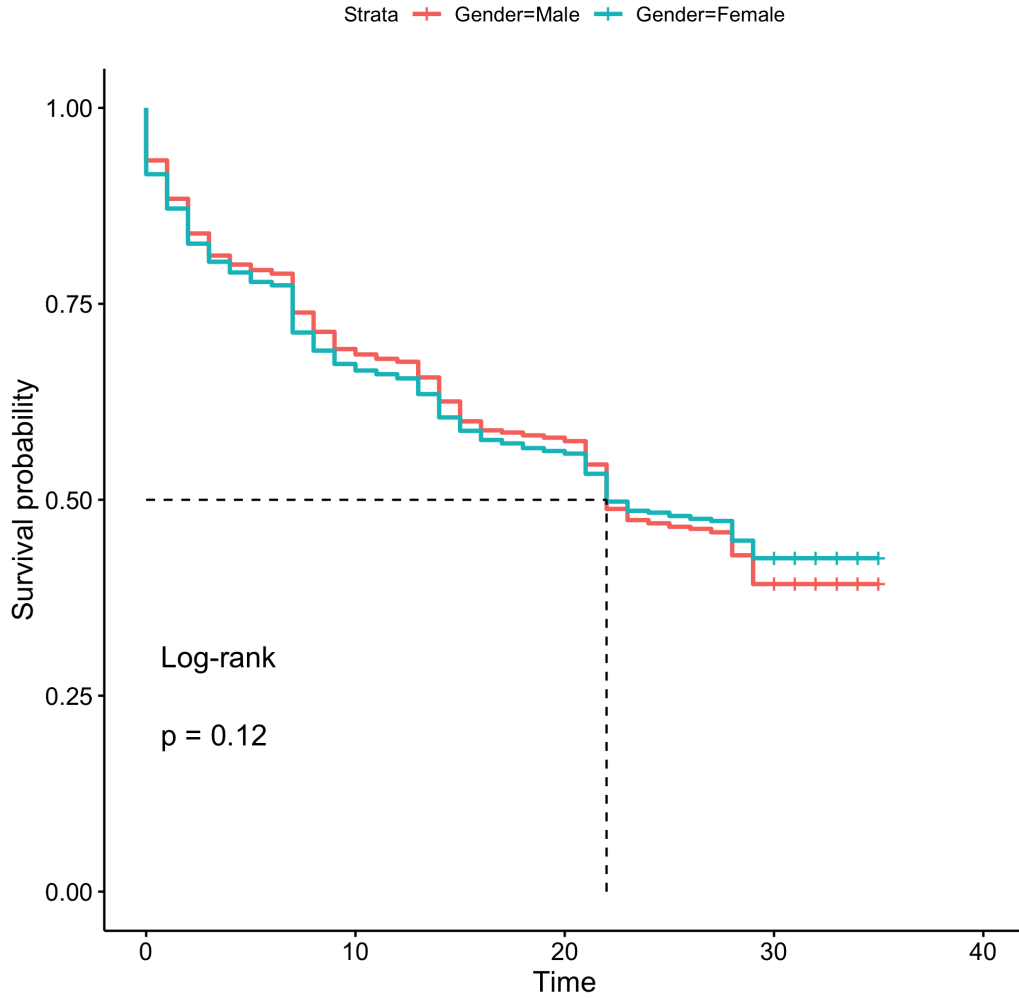


Figure 7: Survival plot between Male and Female learners

From Figure 6, we see that the median survival time for Very High group was around 26 videos, whereas the median for Intermediate group and Low group were around 20 videos and 17 videos, respectively. The median survival time for both Male and Female learners were around 22 videos, as seen in Figure 7.

From Table 7, taking learners from Very High HDI countries as references, learners from Intermediate HDI countries and Low HDI countries were 1.41 times ($HR = 1.41$, $p < 0.001$) and 1.76 time ($HR = 1.76$, $p < 0.001$) as likely to drop the course, respectively. In addition, there was no significant

difference in the likelihood of quitting the course between Male and Female learners (HR = 1.02, p = 0.6).

4 Discussion

The course's first iteration had around two times more registrants than the second and the third iteration, and the completion percentage was also the highest. Understandably, the first iteration's advertising campaign was invested more than the following iterations. Therefore, the considerable difference in registrants is explainable. In addition, the first iteration of a course may attract proportionally more engaged participants than following iterations because enthusiastic registrants are reasonably proactive than their counterparts. These honorable learners were likely aware of the launch of the course from the first iteration. So they registered sooner than less motivated learners, and consequently, their share decreased across iterations.

The completion rates of all three iterations were low. In fact, this phenomenon happens in almost all MOOCs in general. Of course, there are criticisms against MOOCs' low completion rates, but we need to understand that these courses are different from traditional on-campus counterparts. Even though their registrants do not obtain certificates, they still benefit from these courses. As a result, the demand for getting a certificate of MOOCs is obviously lower than a course from a university.

There was no gender inequality between participants in this MOOC regarding course completion and the ratio of drop out. However, we detect the distinction in completion rates between residence groups of learners, mainly in their country's socioeconomic status. Learners from developing countries were likely to drop the courses sooner than learners from developed countries. In addition, their completion rate was also lower. The living environments of learners may affect how they take the MOOC. Living in low condition life may decrease the motivation of continuous learning. The stable access to the MOOCs is also questionable for these learners because there are some clear disadvantages to access quality for lower-income families ([Herold, 2016](#)).

References

- Herold, B. (2016). Mobile-only internet access presents hurdles for families, survey finds. *Education Week*.
- Kizilcec, R., Piech, C., & Schneider, E. (2013, 04). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In (p. 170-179). doi: 10.1145/2460296.2460330
- Koksal, I. (2020, 05). The rise of online learning. *Forbes*.