



*Section Data & Humanités Digitales
Sciences-Po Saint-Germain-en-Laye
CY Tech*

Regard sur les plateformes de MOOCS : le poids de l'HDI et du Genre dans
l'Investissement des Apprenants.

Florian Grolleau 3A-DD

Sous la supervision de : Matthieu Cisel (P.h.D)

Janvier 2025

Résumé

Cet article traite des différences d'investissements des apprenants sur les plateformes d'apprentissage en ligne, qui dispensent de nombreuses formations : des « MOOCS ». Le MOOC de notre étude, nommé MOOC Effectuation¹ a été la formation en ligne sur laquelle nous nous sommes appuyés. Il est constitué de vidéo à regarder ainsi que de différents quizzes à compléter. Nous discuterons ici de la mise en évidence de profils d'apprenants, particulièrement affectés par certaines variables socio-démographiques. En effet nous démontrerons que les variables socio-démographiques qui s'appliquent aux apprenants ont des impacts disparates dans la réussite et la complétion d'un MOOC par un utilisateur, et que celle-ci explique en partie l'existence de profil (à titre d'exemple : « Bystander », « Completer »). Ainsi nous considérerons que le visionnage de vidéo et la réalisation de quiz constituent des témoins de complétion du MOOC en question. Afin de mettre ces résultats en évidence, nous nous sommes appuyés sur un panel de test statistiques varié : test de Student, Chi2, Pearson, Poisson non-inflated etc. Il apparaît alors que l'effet de l'HDI dans la complétion d'un cours est probante, et des résultats visibles d'inégalités de complétion ont pu être observées entre différents groupes d'HDI (bas-intermédiaire-haut). Cependant, il importe de mentionner que l'effet du genre sur l'investissement des apprenants est peu significatif, et n'explique que très peu les différences de visionnages de vidéos. On peut supposer qu'au regard du poids de l'HDI dans l'investissement des apprenants, l'importance du genre est effacé, quand bien même on retrouve statistiquement un impact dudit genre.

Très rapide remise en contexte

Unknown Author
13/01/2025 14:05

Quelques mots sur la méthode

Unknown Author
13/01/2025 14:09

Présentation succincte des résultats

Unknown Author
13/01/2025 14:10

Interprétation rapide des résultats

Unknown Author
13/01/2025 14:13

Table of Contents

Résumé	2
Introduction.....	5
1. Méthodes et données.....	7
1.1 Sources des données.....	7
1.2 Gestion des données aberrantes.....	8
1.3 Techniques employées.....	8
2. Résultats.....	10
1.1 Introduction.....	10
1.2 Des variables qui ne sont pas indépendantes.....	11
1.3 Analyse des différences entre les moyennes des groupes selon le genre.....	11
1.4 Des variables qui mesurent la complétion très corrélées entre elles.....	12
1.5 Une interaction entre le Genre et l'HDI modérée.....	12
3. Discussion.....	14
3.1 Plan de la discussion.....	14
3.2 Les limites.....	14
3.3 Interprétation des résultats.....	15
3.4 Perspectives.....	16
4. Conclusion.....	16
5. Références.....	17

Table of Figures

Figure 1 : Proportion de données manquantes par variable.....	7
Figure 2: Distribution du nombre de vidéos visionnées par apprenant.....	8
Figure 3: Le processus de merging des données.....	9
Figure 4: Proportion des types d'apprenants par catégorie et itération.....	10
Figure 5: Résultats du test du Chi2 entre le Genre et l'HDI.....	11
Figure 6: Test de Welch sur le Genre.....	12
Figure 7: Corrélation entre les vidéos vues et les quiz réalisés.....	12
Figure 8: ANOVA avec le Genre et l'HDI.....	13
Figure 9: Résultats du modèle de régression logistique sur l'effet du genre et de l'HDI sur la probabilité de réussite à l'examen.....	14

Introduction

Avec le lancement de la plateforme numérique publique « FUN MOOC » en 2013 en France, les initiatives publiques comme privées pour proposer des formations en ligne certifiantes et souvent reconnus par l'État ont explosé. Plus généralement inscrite dans un objectif mondial de s'affranchir des contraintes que pose les cours en présentiel dans les formations, l'univers du numérique a trouvé un moyen d'accomplir cet objectif : les MOOCs. Les MOOCs (de l'acronyme massive open online courses) sont des cours en ligne et des formations en ligne, le plus souvent dispensés par des université ou des entreprises. Ceux-ci permettent aux professeurs, acteurs privés de partager à un plus grand nombre d'élèves leurs analyses sur différents sujets. L'engouement est tel que certaines plateformes comme COURSERA, proposent des formations certifiantes voire diplômantes (pour beaucoup onéreuses) de prestigieuses universités (Stanford, le MIT, Harvard).

Attirant des dizaines de milliers d'apprenants, voire des millions pour les formations les plus prisées, certaines de ces cours en ligne sont suivis par des citoyens de tous les pays du monde. Se pose alors un problème : l'accessibilité de ces formations. Si au premier abord l'aspect numérique devrait faciliter leur accès et même permettre une complétion plus aisée de ces formations, en réalité, le suivi d'une formation en ligne peut présenter différentes difficultés. Mais si la formation est dématérialisée, quels sont donc les facteurs qui peuvent affecter sa (non)-complétion ?

Remise en contexte sur le sujet, de quoi parle-t-on ?

Unknown Author
13/01/2025 14:47

Evocation du problème

Unknown Author
13/01/2025 14:54

Bien que l'état de l'art sur le sujet soit très peu fourni, certaines études (Njingang Mbadjoin, & Chaker, 2021) ont mis en lumière des inégalités de complétion liées à des facteurs socio-démographiques tels que l'âge, le niveau d'éducation de l'apprenant, mais également des variables contextuelles comme le soutien social influencent l'investissement des apprenants Bourdon et Bélisle, 2014; Leturcq, 2017; Vonthron et coll. 2007) ou encore, la vie familiale, les tâches domestiques et la disponibilité temporelle sont également citées comme facteurs pouvant « affecter (ou non) la réussite ou la persistance dans des dispositifs de formation en ligne » (Allan et Lewis, 2009).

Très rapide revue de l'état de l'art sur le sujet + éléments qui nous conduisent à notre question

Unknown Author
13/01/2025 14:52

Knowledge gap mis en évidence ici

Unknown Author
13/01/2025 14:53

Cependant l'impact de certaines variables socio-démographiques n'a été que très peu traité, c'est le cas du genre et de l'IDH (indice de développement humain) chez les apprenants. Or, comme nous l'avons évoqué plus haut, ces formations s'affranchissent des frontières, et par l'usage de l'anglais et des outils de traduction, d'une langue d'enseignement. C'est pourquoi il nous semble nécessaire de nous pencher sur ce sujet de l'inégalité de complétion en fonction du développement du pays d'origine (qui influe lui-même sur de nombreuses variables socio-économiques) étant les disparités géographiques des apprenants. Quant au genre des apprenants, peu d'études s'intéressent à l'impact de celui-ci sur leur investissement. Pourtant certains facteurs qui découlent de cette variable qu'est le genre reflète déjà des disparités sur l'engagement des apprenants (tel que la situation familiale, le temps passé aux tâches domestiques fortement inégales en fonction du genre).

Pourquoi l'étude de ces deux variables est pertinente

Unknown Author
13/01/2025 15:06

Dans quelle mesure le genre et l'IDH de l'apprenant influencent-ils son niveau d'engagement dans le suivi d'un MOOC ?

Problématique

Unknown Author
13/01/2025 15:12

Pour ce faire, nous avons utilisé l'IDE Rstudio pour réaliser nos études statistiques ainsi que nos visualisations. Par ailleurs il nous a fallu importer un jeu de données concernant les MOOCs, il s'agit en l'occurrence du jeu de données MOOC : MOOC Effectuation ou MOOC1. S'agissant d'un jeu de données il a fallu trier celles-ci ainsi que réaliser un merge des différents csv qui étaient inclus dedans, en conservant uniquement les données qui nous intéressaient. Par ailleurs, nous avons principalement éprouvé nos données en utilisant des tests statistiques.

Explication rapide de la méthode

Unknown Author
13/01/2025 16:04

Dans notre approche, nous avons formulés les hypothèses suivantes pour guider notre étude. Tout d'abord, nous avons supposé qu'il existait une influence significative de l'IDH du pays d'origine en effet celui-ci nous renseigne en partie sur les possibilités de l'apprenant à avoir un accès Internet stable, qui nous le supposons, affecte le taux de complétion. Par ailleurs nous avons aussi supposé que les apprenants ayant déjà une expérience avec les plateformes numériques montraient un engagement plus élevé, autrement dit, les pays à IDH élevés desquels proviennent principalement les MOOCs seraient plus sensibles à un investissement fort dans ces mêmes MOOCs.

Formulation de quelques hypothèses de départ, qui seront très liées à notre partie discussion.

Unknown Author
13/01/2025 16:13

Afin de réaliser cette étude, il a été essentiel d'établir une méthodologie efficace tant dans l'analyse que dans la constitution d'une base de données exploitable et l'élaboration de graphiques et tests statistiques.

1. Méthodes et données

1.1 Sources des données

Les données utilisées proviennent pour l'essentiel d'un jeu de données nommée « MOOC1 » ou « MOOC Effectuation ». De ce jeu de données, nous avons extrait les parties qui portaient sur des « learnings analytics » issus de différentes itérations de ce MOOC.

Dans nos méthodes, afin de mieux saisir l'importance de certains tests statistiques nous avons également utilisé la bibliothèque « IRIS » sur les pétales de fleurs, disponible sur Rstudio. Par ailleurs nous avons également employé le contenu (à des fins similaires que la bibliothèque IRIS) le jeu de données «maths_competition_awards_data.csv ».

Etant donnée qu'il s'agissait de larges volumes de données, les données manquantes étaient nombreuses et certaines itérations ne contenait que peu d'informations qui les rendaient parfois inutilisables. Voici la répartition des données manquantes dans le jeu de données :

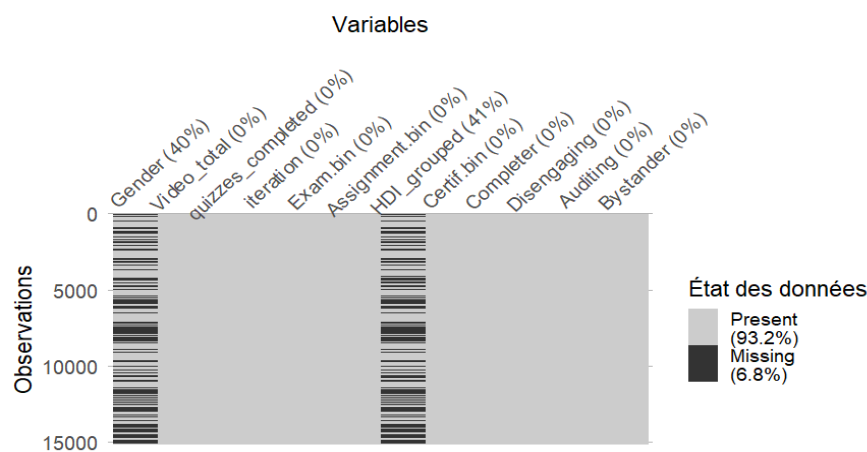
Données principales

Unknown Author
13/01/2025 16:31

Détails des données annexes uniquement utiles à la méthodologie

Unknown Author
13/01/2025 16:31

Figure 1 : Proportion de données manquantes par variable



On remarque dans la Figure 1. que les données manquantes, bien que très spécifiques à certaines variables sont néanmoins présentes à des taux très élevés pour les variables affectées. En l'occurrence on s'aperçoit que les deux variables sur lesquelles nous souhaitons orienter notre étude sont les plus affectées, il s'agit du genre (avec 40 % de données manquantes) et de l'HDI, avec 41 % de données manquantes. Ainsi la fiabilité des données peut être amplement remise en question bien que les données, étant ici issues de formulaire numérique, lors de l'inscription à un MOOC sont en elles-mêmes fiables. Pour autant le volume de données étant relativement élevé, le diviser par deux (comme c'est le cas pour certaines variables) n'entrave pas ou peu la qualité de l'étude.

Commentaires sur les données manquantes

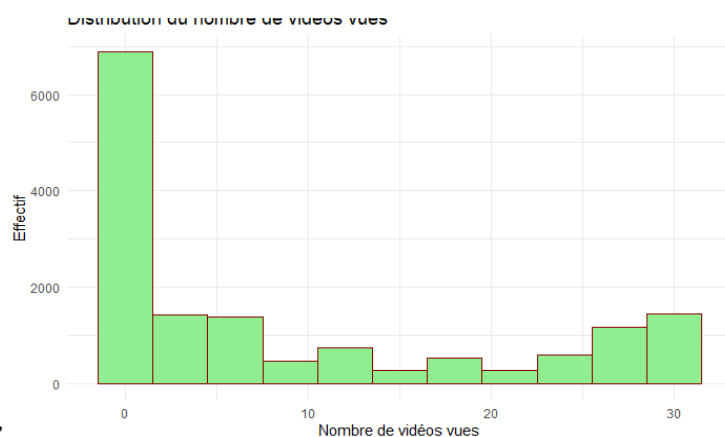
Unknown Author
13/01/2025 16:42

1.2 Gestion des données aberrantes

- On retrouve une quantité non négligeable de données manquantes dans notre jeu de données, mais également des valeurs et données aberrantes. Pour exemple la figure 2. ci-dessous :

Explication d'un outlier

Unknown Author
13/01/2025 17:02



En effet, remarque dans Figure 2: Distribution du nombre de vidéos visionnées par apprenant la Figure 2.

que le nombre de vidéos visionnées par apprenant lors de la réalisation du MOOC est très hétérogène et que de nombreuses valeurs se concentrent autour de 0, c'est-à-dire aucune vidéo visionnée. De fait la Figure 2. nous montre qu'on a près de 7000 apprenants qui n'ont jamais visionné de vidéo, contre moins de 2000 qui en ont regardé 30, soit la totalité des vidéos. Plusieurs explications peuvent être données à ce phénomène : la principale est qu'un très grand nombre d'utilisateurs et d'apprenants du MOOC n'ont procédé qu'à la création du compte, mais n'y ont jamais véritablement pris part et n'ont donc jamais regardé de vidéo concernant le MOOC.

Interprétation de ce problème de valeurs aberrantes

Unknown Author
13/01/2025 17:05

Ceci pourrait fausser nos résultats statistiques et notamment nos tests de Poisson, nous avons donc dû nous adapter et nous orienter sur des tests dit de zero-inflated : Zero Inflated Negative Binomial (ZINB) ou encore Zero-inflated Poisson (ZIP).

Solutions trouvées

Unknown Author
13/01/2025 17:09

1.3 Techniques employées

Le jeu de données se présentant sous la forme de sub-jeu de données plus petit, correspondant à différentes itérations du MOOC Effectuation, nous avons dû, pour recréer un unique jeu de données contenant les différentes itérations, procéder à un « merging » de ces données.

Il y avait au total 6 csv : 3 de type usages.effec c'est à dire les logs d'une des 3 itération du MOOC, et 3 de type effectuation qui concerne les résultats aux questionnaires du MOOC. Il a donc été nécessaire de réaliser le processus suivant, exposé dans la figure 3.

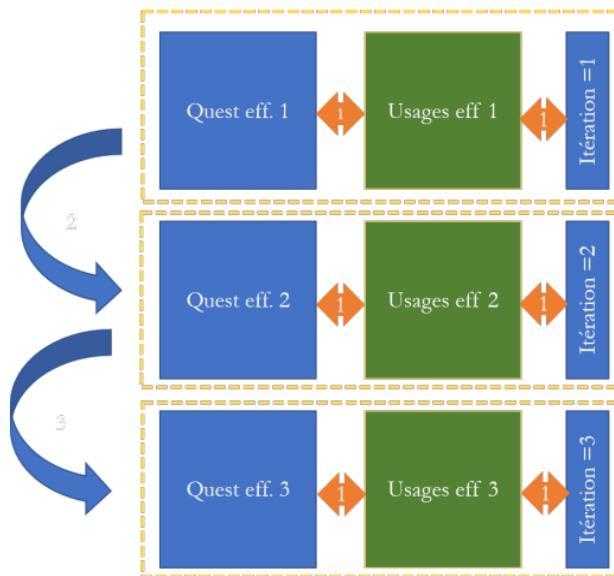


Figure 3: Le processus de merging des données

Comme on peut le remarquer dans la

Figure 3. des « rows binds » et des « columns binds » ont été réalisés afin de constituer une unique base de données (au format csv, puis stockée dans un data frame). L'objectif était d'abord de regrouper les données qui concernait une même itération, auxquelles on ajoutait également une colonne avec la valeur de leur itération (1,2 ou 3).

Par ailleurs, par soucis de clarté et d'efficacité, nous avons fait le choix de supprimer les catégories concernant l'HDI que sont Moyen (M) et Haut (H) pour les regrouper dans une catégorie Intermédiaire (I).

Enfin, afin de créer des profils d'apprenants, nous avons choisis de créer une typologie des apprenants de tel sorte que:

- Si l'examen a été passé ou le certificat obtenu l'apprenant devient un « Completer »
- Si un quiz a été fait ou un devoir soumis mais le certificat n'a pas été obtenu ou l'examen n'a pas été réalisé, alors il s'agit d'un « disengaging learner »

Explication de la structure des données et comment elles ont été traitées

Unknown Author
13/01/2025 17:23

Termes techniques liés à R

Unknown Author
13/01/2025 17:25

- Si aucun quiz n'a été réalisé et aucun devoir soumis mais qu'au moins 6 vidéos ont été visionnées alors on parle « d'auditing learner »
- si moins de 6 vidéos ont été visionnées et qu'aucun quiz n'a été réalisé alors il s'agit d'un « bystander »

Présentation de la typologie d'apprenant utilisée
Unknown Author
13/01/2025 20:23

Comme le présente la Figure.4

Itération	Completer (%)	Disengaging (%)	Auditing (%)	Bystander (%)
1	31.07	27.61	0.72	39.42
2	26.99	25.42	1.49	44.79
3	25.66	21.85	0.94	49.99

Table 2: *

Note. Les proportions sont exprimées en pourcentage pour chaque rôle (Completer, Disengaging, Auditing, Bystander) et sont calculées en moyenne par itération.

Figure 4: Proportion des types d'apprenants par catégorie et itération

Afin de réaliser nos modèles statistiques, et de maintenir une cohérence graphique de ceux-ci nous avons utilisé l'IDE Rstudio, le fichier généré étant un notebook.

Détails des bibliothèques et logiciels utilisés
Unknown Author
13/01/2025 18:41

Les librairies statistiques et de visualisations suivantes se sont avérées très utiles :

- **Dplyr** : Une bibliothèque pour la manipulation de données. Elle permet de filtrer, trier, regrouper, et transformer facilement des tableaux de données.
- **ggplot2** : Une bibliothèque pour créer des visualisations poussées en R.
- **Tidyr** : Utilisée pour nettoyer et transformer les données en organisant les tableaux.
- **Naniar** : Spécialisée dans la gestion et la visualisation des valeurs manquantes dans les jeux de données.
- **Vcd** : Permet de visualiser les données catégorielles, notamment via des graphiques comme les diagrammes de mosaïque et les graphiques d'associations.
- **MASS** : Contient des fonctions pour les analyses statistiques avancées, notamment les modèles linéaires généralisés (GLM) et les méthodes d'estimation robustes.
- **PscI** : Fournit des outils pour modéliser des données de comptage, y compris les modèles Zero-Inflated Poisson (ZIP) et Zero-Inflated Negative Binomial (ZINB).

2. Résultats

1.1 Introduction

Dans cette partie, il s'agira d'exposer les résultats statistiques de nos tests, via de multiples tables de reportings. Les tests statistiques en question s'attachent à la fois aux associations entre les variables, mais également à leur importance propre dans l'engagement des apprenants dans le MOOC.

Phrase introductive des résultats

Unknown Author
13/01/2025 20:23

1.2 Des variables qui ne sont pas indépendantes

Tout d'abord les tests statistiques réalisés ont mis en évidence que les variables du genre et de l'HDI n'étaient pas totalement indépendantes l'une de l'autre en outre d'après la figures 5 ci dessous :

Test	Statistique	df	p-value
Test du Chi2	179.24	2	0.001
V de Cramer	0.1415		

Figure 5: Résultats du test du Chi2 entre le Genre et l'HDI

Ainsi d'après la Figure 5. le test du Chi2 révèle que le X-squared = 179,24 et que la p-value = $2,2 \times 10^{-16} < 0,001$. Ainsi, l'on peut en conclure que les deux variables ne sont pas indépendantes.

Dans la Figure.5 on peut remarquer que le V de Cramer (qui mesure la force d'association) n'est pas égal à 0 et que donc il existe bel est bien une association entre le Genre et l'HDI, cependant celle ci reste malgré tout peu significative : V de Cramer = 0,1415.

Commentaires sur les résultats sans interprétation

toujours une citation de la table « Figure n° » dans les commentaires

Unknown Author
13/01/2025 20:42

1.3 Analyse des différences entre les moyennes des groupes selon le genre

Les analyses statistiques effectuées permettent de comparer les moyennes de la variable Video_total entre les deux groupes définis par le genre : "un homme" et "une femme". Le Welch Two Sample t-test a été réalisé pour évaluer si la différence observée entre ces moyennes est statistiquement significative.

Phrase introductive pour les résultats

Unknown Author
13/01/2025 20:46

Statistique	Valeur	Intervalle de confiance (95%)
<i>t</i>	-3.7603	2*[-1.458, -0.459]
<i>df</i>	5879.2	
<i>p</i>	0.0001713	
Moyenne (un homme)		13.18
Moyenne (une femme)		14.14

Note. Les résultats montrent une différence significative entre les moyennes des deux groupes ($p < 0.001$). Les valeurs moyennes sont exprimées sur l'échelle de la variable Video.total.

Figure 6: Test de Welch sur le Genre

D'après la Figure 6, et le **test t de Welch** on a révélé une différence significative entre les moyennes des deux groupes ($t=-3.76$, $df=5879.2$, $p=0.0001713$). La ppp-valeur obtenue est inférieure au seuil alpha standard ($\alpha=0.05$), ainsi l'hypothèse nulle selon laquelle les moyennes des deux groupes sont identiques est rejetée.

Les moyennes observées sont les suivantes :

- Moyenne pour le groupe "un homme" : 13.18
- Moyenne pour le groupe "une femme" : 14.14

Ainsi, les femmes obtiennent en moyenne une valeur de totale de vidéos regardées plus élevée que les hommes. La différence moyenne estimée entre les deux groupes est comprise dans l'**intervalle de confiance à 95%** : [-1.46, -0.46]. Cela signifie qu'il est peu probable que cette différence moyenne soit nulle dans la population. Cela implique que le genre a une influence probante sur le nombre total de vidéos visionnées mais cet écart entre genre est petit, elle est en effet d'un peu plus de 1 vidéo de plus regardée pour une femme par rapport à un homme.

Lecture des résultats + commentaires

Unknown Author
13/01/2025 20:50

1.4 Des variables qui mesurent la complétion très corrélées entre elles

Type de corrélation	Coefficient	Interprétation
Corrélation de Pearson	0.8056	Forte corrélation positive
Corrélation de Spearman	0.7999	Forte corrélation positive

Figure 7: Corrélation entre les vidéos vues et les quiz réalisés

D'après la figure 7, les analyses de corrélation effectuées entre les variables « Videos_Vues » et « Quiz_Realisés » révèlent une association positive et forte. La corrélation de Pearson ($r=0.8056$) indique une relation linéaire significative entre ces deux variables : à mesure que le nombre de vidéos visionnées augmente, le nombre de quizz réalisés tend également à augmenter de manière proportionnelle.

De plus, la corrélation de Spearman ($p=0.7999$) confirme cette association en montrant qu'elle est également monotone, ce qui signifie que, globalement, lorsque l'une des variables augmente, l'autre

Présentation des résultats de la table et commentaires de ceux-ci, toujours une citation de la table « Figure n° » dans les commentaires

Unknown Author
13/01/2025 21:03

suit le même comportement, même en cas de relation légèrement non linéaire ou en présence de valeurs extrêmes.

1.5 Une interaction entre le Genre et l’HDI modérée.

Une analyse de variance (ANOVA) prenant en compte les interactions entre le genre et l’HDI a montré que ces interactions ne sont pas significatives. Cela suggère que l’effet de l’HDI sur le nombre de vidéos visionnées est indépendant du genre (mais pas nécessairement l’inverse comme nous l’avons mis en évidence plus haut). Ainsi, le niveau de développement humain du pays d’origine des participants constitue un déterminant majeur du nombre de vidéos visionnées, sans que cette relation soit influencée par le genre.

Annonce succincte des résultats

Unknown Author
13/01/2025 21:31

Prédicteur	Estimation	Erreur standard	t	p
(Intercept)	6.2333	0.3712	16.794	0.001 ***
Genre (femme)	0.5354	0.9825	0.545	0.586
HDI (groupe I)	4.7111	0.6476	7.275	0.001 ***
HDI (groupe TH)	8.4854	0.4045	20.980	0.001 ***
Interaction Genre (femme) : HDI (groupe I)	-1.8618	1.3300	-1.400	0.162
Interaction Genre (femme) : HDI (groupe TH)	-0.2671	1.0193	-0.262	0.793

Note. Les résultats montrent des effets significatifs pour les prédicteurs *HDI (groupe I)* et *HDI (groupe TH)* sur la variable dépendante *Video_total* ($p < 0.001$). Les interactions entre le genre et les groupes HDI ne sont pas significatives ($p > 0.05$).

Figure 8: ANOVA avec le Genre et l’HDI

Table de reporting

Unknown Author
13/01/2025 21:32

D’après la Figure 8 : Le genre, pris isolément, n’a pas d’effet significatif sur le nombre de vidéos visionnées ($b=0.535, p=0.586$). Les hommes et les femmes visionnent un nombre de vidéos similaire en moyenne. En revanche, l’HDI a un effet principal significatif. Les participants des pays à HDI intermédiaire visionnent en moyenne 4.711 vidéos de plus que ceux des pays à HDI faible ($p<0.001$). Cet effet est encore plus prononcé pour les pays à HDI très élevé, où les participants consomment 8.485 vidéos de plus en moyenne par rapport aux pays à HDI faible ($p<0.001$).

Les résultats en questions commentés

Unknown Author
13/01/2025 21:32

Les interactions entre le genre et l’HDI ne sont pas significatives, que ce soit pour les pays à HDI intermédiaire ($b=-1.862, p=0.162$) ou pour les pays à HDI très élevé ($b=-0.267, p=0.793$). Cela indique que l’effet de l’HDI sur le nombre de vidéos visionnées est le même pour les hommes et les femmes. Par exemple, dans les pays à HDI intermédiaire, les femmes et les hommes consomment environ 4.7 vidéos de plus que ceux des pays à HDI faible, sans différence notable entre les genres. De manière similaire, dans les pays à HDI très élevé, hommes et femmes consomment environ 8.5 vidéos de plus que leurs homologues des pays à HDI faible, là encore sans interaction significative entre le genre et l’HDI.

Ces résultats mettent en évidence que l’HDI est un facteur clé de la consommation de vidéos, avec une augmentation marquée pour les pays à HDI intermédiaire et très élevé.

Par ailleurs l’impact du HDI a aussi été mis en évidence dans la réussite globale (et donc la complétion) du MOOC, En effet comme le présente la Figure 9 :

Transition vers la Figure 9

Unknown Author
13/01/2025 21:32

Variable	Référence	OR	IC (lower)	IC (upper)	p-value	Signif.
Intercept	Réf.	-	0.139	0.197	2.62×10^{-90}	***
Genre (Femme)	-	1.122	1.001	1.256	4.72×10^{-2}	*
HDI (I)	-	1.119	0.852	1.465	4.16×10^{-1}	ns
HDI (TH)	-	1.372	1.144	1.656	7.86×10^{-4}	***

Table 2: *

Note. OR = Odds Ratio ; IC = Intervalle de Confiance à 95%. Les résultats montrent des effets significatifs pour le Genre (Femme) et l’HDI (Très Haut).

Figure 9: Résultats du modèle de régression logistique sur l’effet du genre et de l’HDI sur la probabilité de réussite à l’examen

Figure 9

Unknown Author
13/01/2025 21:32

Le modèle de régression logistique exposé par la Figure 9 met en évidence que le genre et l’HDI sont des facteurs significatifs pour prédire la probabilité de réussite à l’examen. Les femmes ont une probabilité légèrement supérieure de réussir par rapport aux hommes (OR=1.122, p=0.047). Concernant l’HDI, les participants provenant de pays à HDI très élevé ont une probabilité significativement plus grande de réussir à l’examen (OR=1.372, p<0.001). En revanche, l’effet pour les pays à HDI intermédiaire n’est pas significatif (OR=1.119, p=0.416). Ces résultats indiquent une influence positive de l’HDI élevé sur les performances académiques, indépendante du genre.

Commentaires sur les résultats de la Figure 9

Unknown Author
13/01/2025 21:34

3. Discussion

3.1 Plan de la discussion

Afin d’apporter un regard critique et un approfondissement de nos résultats, nous observerons tout d’abord une analyse critique de nos résultats. Par la suite nous tenterons de présenter des modèles explicatifs de tels résultats d’une part pour l’HDI, d’autre part pour les langues d’écriture. Enfin nous essayerons d’établir des perspectives futures de réalisation d’études similaires portant sur l’engagement dans les MOOCs.

Annonce du plan de la discussion

Unknown Author
13/01/2025 22:10

3.2 Les limites

Nous avons tout d'abord pu exposer un problème de fiabilité des données dans la partie méthodologie. Ceci étant, les données manquantes et les problèmes d'outliers (ou valeurs aberrantes) ont pu être traités comme mentionné plus haut.

En revanche, certaines limites subsistent, c'est le cas notamment du MOOC en lui-même. En effet ce MOOC ne représente que ses propres statistiques et il n'est en aucun possible de généraliser son cas, même si la mise en lumière de certains patterns peut nous permettre de déceler des tendances que nous évoquerons dans l'explication et l'interprétation des résultats. De fait suivant la qualité et l'objet des MOOCs, les apprenants n'observent pas forcément la même assiduité et ceci ne peut être retranscrit dans l'étude d'un seul MOOC. Pour exemple, un cours de management avec aucun prérequis pourra trouver un public probablement beaucoup plus large (et donc beaucoup plus enclin à contenir des « disengaging learners » qu'un cours d'entrepreneuriat poussé (Cisel, 2015) ou un cours de développement web. En ce sens, il nous est donc difficile de généraliser le cas de notre MOOC à tous les MOOCs des plateformes numériques.

Limite 1 (très inspiré de « Two Tale of MOOCS »)

Unknown Author
13/01/2025 22:32

Cependant on pourrait proposer des résultats similaire à des MOOCS analogues, et réaliser une étude sur la place de la communication, de la stratégie marketing vis-à-vis du cours, au fur et à mesure des itérations (ce qui change nécessairement son attractivité et ses taux de complétion), (Cisel, 2015).

Limite 2

Unknown Author
13/01/2025 22:33

Enfin, on pourrait également ajouter qu'avec une absence de comparaison avec d'autres formats d'apprentissage, il nous est difficile de véritablement retranscrire l'importance des désengagements, des disparités entre les groupes, comparé à d'autres types de formation.

Limite 3

Unknown Author
13/01/2025 22:33

De plus on peut également supposer qu'il existe un véritable manque de données qualitative. En effet, nous nous sommes axés sur un très petit nombre de paramètres pour mesurer une complétion ou un abandon, à savoir les vidéos visionnées, les quizzes complétés, le genre et l'HDI alors même que de très nombreux facteurs externes pourrait revoir à la baisse (ou à la hausse) l'impact de ces deux variables. Tandis que les vidéos visionnées et les quizzes complétés ne rendent pas nécessairement compte de tout l'engagement possible d'un apprenant, certains préférant lire, échanger sur les forums mis à disposition (les chats des apprenants).

Limite 4

Unknown Author
13/01/2025 22:33

3.3 Interprétation des résultats

Nous avons pu tout d'abord constater que l'importance de l'HDI était majeur, et ce quelque soit le niveau de développement humain du pays d'origine. Pour rappel, les apprenants issus de pays développés ou très développés regardaient 66 % de fois plus de vidéos que ceux issu des pays pauvres (HDI bas). De même les personnes issues de pays d'HDI Intermédiaire (I) possédaient des taux de complétion bien supérieur à ceux des pays en voie de développement. Pour autant il existe une différence également très importante entre les HDI élevés et intermédiaires. On pourrait expliquer ceci par le fait que les taux de complétion en fonction de l'HDI est une vitrine des disponibilités technologiques des pays, et que certains pays, plus assujettis à des coupures de courant, à des aléas dans le bon fonctionnement des télécommunications, ne permettent pas de

suivre avec aisance une formation en ligne (parfois sur plusieurs semaines). A l'inverse la culture du numérique étant de plus en plus importante dans les pays à HDI élevé, l'écart d'engagement peut aussi en ce sens se creuser et se ressentir, avec des apprenants issus de pays développés qui cherchent à s'affranchir de formations en présentiel pour se tourner vers des formations en lignes faciles d'accès.

Plus encore l'existence d'une barrière de la langue ou encore de prérequis peut s'avérer être un frein non négligeable chez les apprenants de pays à HDI faible. Ceux-ci peuvent être naturellement découragés par le fait de devoir apprendre indépendamment du cours, des langues d'enseignement (français, anglais, espagnols) qui sont généralement les plus importantes dans les formations de type MOOCs.

La disponibilité professionnelle peut également être un facteur d'engagement ou désengagement des apprenants dans les MOOCs. Or celle-ci est directement corrélée à l'HDI du pays de résidence de l'apprenant. De fait des disponibilités professionnelles plus faibles entraînent naturellement des sensibilités à se désengager d'un cours beaucoup plus évidentes (Njingang Mbadjoin, & Chaker, 2021).

Pour ce qui est de la partie du Genre, les résultats conduisent à remarquer que son influence ne serait que minime dans l'engagement des apprenants. Cependant, on peut supposer que même si l'impact de l'HDI a tendance à éclipser celui du Genre, une différence subsiste, notamment au niveau du nombre de vidéos visionnées, légèrement plus élevé pour les femmes ainsi qu'un potentiel de réussite 12,2 % plus élevé aux examens.

Cependant il est possible qu'à une échelle plus globale le genre ai un impact plus conséquent dans l'assiduité lors du suivi d'un MOOC. De fait, on peut remarquer que Les femmes montrent une réussite supérieure à celle des hommes (respectivement $M = 69,01$; $ET = 27,91$ contre $M = 63,54$; $ET = 28,91$) de manière significative : $U = 22668$; $p = 0,022$ d'après une étude similaire, cette fois ci sur plusieurs MOOCs (Njingang Mbadjoin, & Chaker, 2021). Par ailleurs, l'étude en question explique néanmoins que certaines femmes, bien qu'une minorité, s'auto-censurent dans les premiers chapitres des MOOCs, souvent inconsciemment et exerce un retrait (non justifié) du cours.

Pas d'initiales lors de la citation

Unknown Author
13/01/2025 22:54

Pas d'initiales lors de la citation

Unknown Author
13/01/2025 22:54

3.4 Perspectives

Pour enrichir nos résultats et conduire une étude qui fait preuve de plus de représentativité de l'ensemble des MOOCs, il serait pertinent d'élargir l'analyse à plusieurs MOOCs variés afin d'identifier des tendances globales et non spécifiques à un cours unique. L'intégration de données qualitatives, comme les motivations et obstacles perçus par les apprenants (implémentés sous la forme d'un formulaire), pourrait également enrichir les interprétations. Une analyse de l'impact des langues d'enseignement et des contraintes contextuelles (accès technologique, obligations professionnelles) permettrait de mieux comprendre les disparités observées. Enfin, évaluer l'effet de stratégies pédagogiques comme la gamification (le développement de la pédagogie par le loisir) ou les forums interactifs pourrait aider à maximiser l'engagement et la complétion dans les MOOCs.

Partie perspectives qui contrebalancera avec la conclusion

Unknown Author
13/01/2025 23:39

4. Conclusion

Ainsi, malgré certaines limites qui surlignent la fragilité d'une telle étude quant à sa généralisation à tous les MOOCs, nous avons pu dans l'ensemble mettre en évidence des corrélations entre le pays d'origine des apprenants et leurs taux d'engagement dans les MOOCs. Pareillement, nous avons pu exposer la complexité de la variable du genre, qui bien que plus atténué dans son impact révèle cependant que des disparités existent et qu'elles sont explicables. Les disparités induites par le genre sont fortement corrélées à des réalités domestiques, en particulier à des mœurs inégalitaires. Tandis que les disparités plus fortes liées à l'HDI sont plutôt expliquées par des inégalités de facteurs socio-démographiques de plus grandes ampleurs.

Conclusion sans trop s'envoyer « de fleurs », reste sur un bilan mitigé

Unknown Author
13/01/2025 23:44

5. Références

- Matthieu Cisel, Mattias Mano, Rémi Bachelet, Philippe Silberzahn. A Tale of Two MOOCs: Analyzing Long-Term Course Dynamics. European Moocs Stakeholders Summit (eMOOCs), May 2015, Mons, Belgium. Ffhal-01635080
- Njingang Mbadjoin, T. & Chaker, R. (2021). Les liens entre les objectifs de formation, les facteurs sociodémographiques et la réussite chez des participants à un MOOC professionnalisant. McGill Journal of Education / Revue des sciences de l'éducation de McGill, 56(1), 149–170. <https://doi.org/10.7202/1087053ar>
- Allan, B. et Lewis, D. (2009). Time, individual careers and lifelong learning. Dans J. Field, J. Gallaher et R. Ingram (dir.), Researching transitions in lifelong learning (p. 40–49). Routledge
- Alstete, J. W. et Beutell, N. J. (2004). Performance indicators in online distance learning courses: A study of management education. Quality Assurance in Education, 12(1), 6–14. <https://doi.org/10.1108/09684880410517397>