Predicting Student Churn in Massive Open Online Courses

Zach Quinn

Bellevue University

Introduction

Massive Online Open Courses: A Primer

The increasing demand for proficiency across a spectrum of industries as well as a proliferation of online resources from academics at all levels has spawned an entirely new approach to asynchronous education: The Massive Open Online Course (MOOC). While a conventional collegiate-level course typically spans ten to twelve weeks on average, MOOCs appeal to working professionals because these individuals can dedicate two to five hours a week to learning compared to a national average of 17 hours a week for high school and university students (Pierre, 2014). This means that the average college student, according to this guidance, spends over 200 hours on homework, assignments, tests and participation exercises throughout a semester that might represent only a sliver of their overall degree progress. Conversely, those who participate in MOOCs can typically expect to spend 10 – 100 hours on an entire course that offers an authentic certification and, in many cases, real-world value (Chung, 2015). The asynchronous nature of these programs means students can complete these courses on their own schedules, with more than 45% out of a 4500 survey sample size reporting that they work on MOOCs in the evening after a full day of work (Chung, 2015).

It must be noted, however, that these programs only differ from degree programs in availability, not in content. The vehicles used to deliver content in MOOCs are the same kinds of educational instruments a student may find in a conventional online classroom: Discussion boards, short projects, quizzes, exams and, occasionally, a live interaction with the instructor. Unlike conventional courses, MOOCs tend to focus on a detailed examination of a particular subject. MOOCs also tend to assume a base level of knowledge. For example, instead of a course that teaches Python programming in a comprehensive manner, since it involves a shorter time frame, a MOOC might cover a specific use case for Python such as algorithmic trading or Python for business intelligence. MOOCs present themselves to students as all-inclusive approaches to learning. While a conventional online course might require a textbook or other materials that would represent out-of-pocket costs, an MOOC often contains all information, files and resources within the course shell. MOOCs also involve a bit more academic guidance, as the courses are organized sequentially with an instructor walking students through concepts, assignments, assessments and answers. However, the predominant issue with MOOCs is not a matter of convenience. Instead, like school itself, the larger business problem for MOOCs is a significant attrition rate.

The Problem of Attrition in Online Education

Even incredibly reputable and renowned institutions, like Harvard and MIT, are struggling with MOOC attrition rates upward of 96% over a five-year time span (Murray, 2019). With tens of thousands of students enrolling (estimates suggest 40 – 60,000 per course), name recognition only assists with recruitment and marketing. Across multiple platforms such as EDX (the partnership between Harvard and MIT), Coursera and Udemy, the proportion of students who actually complete these courses ranges from 1 – 5% (Koller, Ng & Chen, 2013). Analysts and academics have coined a term for the goal of successful MOOCs: The retention funnel, and note that if these proportions occurred and were observed in an in-person environment, the

school would be shut down. Even in smaller sample size studies the attrition rate was nearly 70%. According to the article 'Exploring the Factors Affecting MOOC retention: A Survey Study' published in the peer-reviewed journal *Computers and Education*, out of a sample size of 400 Egyptian students, only 32% completed an entire course (Hone & Said, 2016). Part of the reason students do not complete these courses is because while universities dedicate entire offices to advertising the tangible financial benefit of an undergraduate or graduate degree, MOOCs make a much more ambiguous promise to 'build job skills' without actually defining these skill sets (Lederman, 2019). According to the article Why MOOCs Didn't Work, in 3 Data Points published in the journal *Inside Higher Ed*, efforts to democratize an education normally reserved for certain socioeconomic groups has largely failed because it has difficulty articulating such a grandiose message in a manner that makes it obvious what the benefit of such a program is: "A strategy that depends upon bringing new learners into higher education cannot succeed if educational institutions cannot support learners in converting their time and financial investment into completing a course to earn a credential with labor market value" (Lederman, 2019, 1). The author contends that elite universities seeking to democratize education must pivot, instead, to offering affordable pathways for accessing and continuing conventional degree programs. Simply offering supplemental, boutique courses is not enough to intrigue consumers. While MOOC retention has fallen over the past decade, there are studies such as that conducted by Jordan (2015) that suggest that a more realistic estimate for completion rate would be around 12 percent (Jordan, 2015). However, Jordan also suggests that while MOOC enrollment has fallen overall, completion rates have increased proportionally. The author employed a linear regression methodology in an attempt to determine whether MOOC course completion rates were changing over time. The author employed multiple regression to determine how multiple variables impacted enrollment, engagement and completion.

Business Problem: Churn

Customer retention is both the goal and predominant challenge for a subscription-based business model (Altexsoft, 2019). Broadly defined, churn is a tendency of customers to abandon a product or discontinue being loyal to a brand due to boredom, incompatibility or dissatisfaction (Altexsoft, 2019). According to analyst estimates, customer churn costs U.S. businesses nearly 140 billion dollars per year and that nearly a third of consumers revealed that they would stop doing business with a company after only one bad customer experience (Horstmeyer, 2020). With the market for MOOCs growing exponentially, it is more crucial than ever for these content providers to explore strategies to preemptively target users who would otherwise be tempted to churn. Market analysts predict that the MOOC market value in the educational technology sector will grow by 40 percent within the next seven years according to market research firm Data Bridge (Data Bridge, 2020). The prevalence of distance learning due to the pandemic has caused many companies to enter the MOOC/e-learning industry. Increased competition among these companies and a consumer need to refine and develop job skills during a period of historic layoffs will likely result in rapid growth within the sector (Data Bridge, 2020). In the MOOC space students have a variety of options including Coursera, EDX and Udemy. Like brick and mortar schools themselves these institutions range from delivering an elite, privatized experience (EDX, Harvard, MIT) to a more specific, niche experience on skill building platforms like Udemy (Altexsoft, 2019). Unlike many other business problems, churn is a fairly intuitive issue: If customers become disengaged from a product, they will seek stimulation and, by extension,

content, from another provider (Altexsoft, 2019). Part of the trouble with MOOCs is that, with the diversity of content offered, it is difficult to determine what exactly constitutes course engagement, attrition and completion. Author Jonathan Haber contends that the attrition crisis for MOOCs is not an objective truth. In fact, due to the variety of methods for calculating enrollment and completion, rates can fluctuate as much as 20 – 30% from their initial values (Haber, 2013). For instance, Haber profiles a professor whose course engagement rate jumped from 15 to 48% simply when he lowered the criteria for engagement to 'watched one video' or 'attended one lecture' (Haber, 2013). Haber's argument that data related to MOOCs may be skewed due to the lack of a universal metric for engagement will be taken into account throughout the duration of this project. Therefore, in order to identify, intervene and preempt customer churn incidents this paper will clarify what metrics constitute engagement in the context of this project.

The Project

By crafting a coherent and comprehensive definition of churn as it relates to MOOC data and framing realistic outcomes that align with research presented in the first part of this essay, this project endeavors to identify the variables that indicate retention and leverage an understanding of these factors to conceive, design and deploy multivariable predictive analytic models in order to preempt short and long-term student attrition rates evident in MOOCs. While customer churn is a phenomenon that has been identified and studied for decades, applying predictive strategies to preempt customer churn in MOOCs represents an emerging, somewhat niche concentration of scholarship and business application. MOOCs present a particular challenge to a conventional churn analysis because, like so many other tech-based platforms, these courses operate on a so-called 'freemium' business model. While platforms offer paid content, including full-fledged 'micro degree programs', they also tout a student's ability to audit as many courses as he or she would like, free of charge. Although students who audit courses are not able to submit assignments for grades or participate in certain activities, they still have access to an entire course. Throughout the course a student is presented with several opportunities to pay for a certificate of completion or other credential but, as Lederman (2019) noted, many students do not understand the marketability or tangible financial value of such a certification. In other words, they find themselves in a difficult business predicament: They do not know what their money is paying for. The combination of a flawed business model and unclear message regarding the value of the product these platforms is selling results in a disproportionate number of students dropping out after only one week of engagement, according to MOOC data obtained for this project.

Although customer retention analysis is valued by large businesses, such a project would be invaluable to MOOC instructors. With the exception of school-sponsored course content hosted on platforms like EDX or Coursera, many of these instructors are business professionals who are seeking to pass on knowledge. Platforms like Udemy pride themselves on delivering instruction that is grounded in an instructor's real-world experience. However, unlike larger institutions like Harvard and MIT that can afford to structure programs to account for attrition, instructors of smaller, more niche classes suffer a more substantial financial loss when students opt out of paying for the full version of a course. A lack of engaged students also results in another problem that is unique to user-centric platforms: A lack of individuals to leave substantiated, objective reviews of the content. Since, on sites like Udemy, many of these

instructors are not affiliated with postsecondary institutions and do not benefit from name recognition, one of the best ways to attract new students is promotion through detailed, transparent reviews by pleased or displeased past students. The important component to note about these reviews is that, regardless of whether the student was satisfied or dissatisfied, they have at least completed the course and are, therefore, qualified to provide an objective evaluation of its quality and larger application to the professional world. User-generated reviews are a feature that will be considered and potentially integrated into the model associated with this project.

Hypothesis

This project aims to apply conventional multivariable churn analysis techniques to the field of online boutique instruction in order to gauge and inspire greater engagement in such courses. The hypothesis for this experiment is: By identifying and combining numeric demographic, engagement, and content-based features in a multivariable logistic regression model, it will be possible to make a binary prediction of which students earned or did not earn a certificate of completion in a course on EDX or Udemy. In order to present the findings to a non-technical audience, the resulting classification report will be displayed on a heat map and other universally recognized intuitive visualizations.

10 Supplemental Research Questions

With more and more schools converting to distance learning formats and unemployed adults refining core job skills, MOOCs fulfill a crucial role in the provision of supplemental education. Additionally, the global MOOC market size is projected to grow at a rate of over 40% from 3.9 billion in 2018 to 20.8 billion by 2023 (Sugla, 2020). Although this sector is profitable, MOOCs also have an attrition rate well above the worst performing educational institutions, at an astronomical 95% by some estimates. Consequently, I am proposing the following research questions:

- Does a student's completion of one MOOC increase the likelihood of completing another course offered by the same platform (EDX)?

Justification: As an occasional MOOC student, I noticed that the most thorough and glowing reviews for courses or instructors are written by students who have completed multiple courses. I am seeking to determine if there is a strong correlation between MOOC history and attrition.

- Does the number of weeks enrolled impact the probability that a student will complete an EDX course?

Justification: The research I included in my annotated bibliography suggests that student engagement drops off after a certain point. I am trying to determine where the threshold for engagement lies (or if a universal drop off point exists) and how to advise MOOC creators on intervention strategies.

- Does course duration determine whether or not a student will complete an MOOC through EDX? Does course size?

Justification: As a gut reaction, I would think that shorter courses would have more successful completion rates. This question attempts to explore that informal hypothesis.

- Do demographic features such as gender or age impact one's likelihood of completing a course?

Justification: In any kind of advertising or business problem demographic variables typically influence the outcome of a given intervention strategy.

- Does the percentage of completion for paid students make an accurate predictor for whether or not a student will finish a given course?

Justification: This is a somewhat intuitive question that seeks to analyze past user history in order to predict future engagement.

- Does the percentage audited make for a reliable and accurate predictor for whether or not a student will pay for receive a certification?

Justification: Given the low bar of entry for these courses (they are free and require no prior credentials), it would make sense that attrition occurs within the group that are auditing and not paying for a course.

- Does prior education level influence one's decision (or lack thereof) to finish a particular EDX or Udemy course?

Justification: One of the features included in the data set is a column for 'Bachelor's degree or higher.' It is possible that education level could make a reliable predictor for one's ability to finish an MOOC offering.

- Does the scope of content influence a student's choice to earn a certificate?

Justification: There are variables in the dataset that include metrics for the minutes of video watched and number of forums participated in, which are fairly standard KPIs for a platform that hosts online digital content.

- Does the number of subscribers or published reviews influence a student's decision to earn or opt out of earning a certificate of completion?

Justification: Without name recognition, courses offered by individuals and not institutions disproportionately benefit from evidence that a course has been subscribed to and thoroughly reviewed.

- Do free course offerings (audits) have higher or lower attrition rates than paid courses, certifications or micro degree programs on the EDX platform?

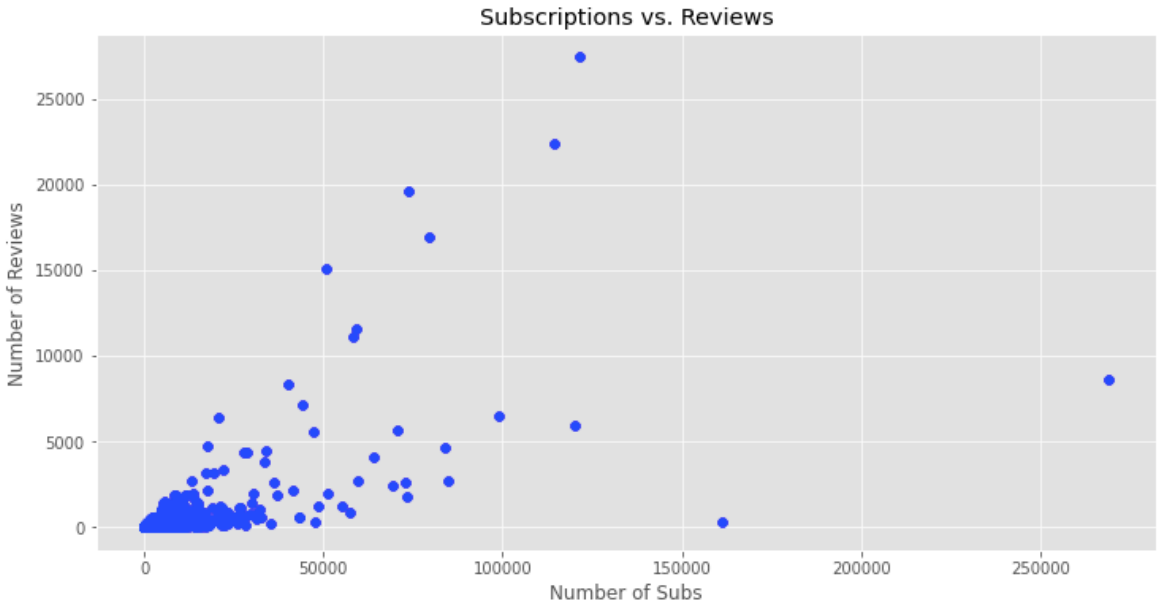
Justification: This is the research question I began my inquiry with and still consider it to be the core problem. Intuitively, it would seem like the price of a course would impact completion, i.e. free courses would have a higher enrollment rate but lower completion rate and paid courses would have a lower enrollment rate but higher completion rate because the student is paying for a product.

The Data

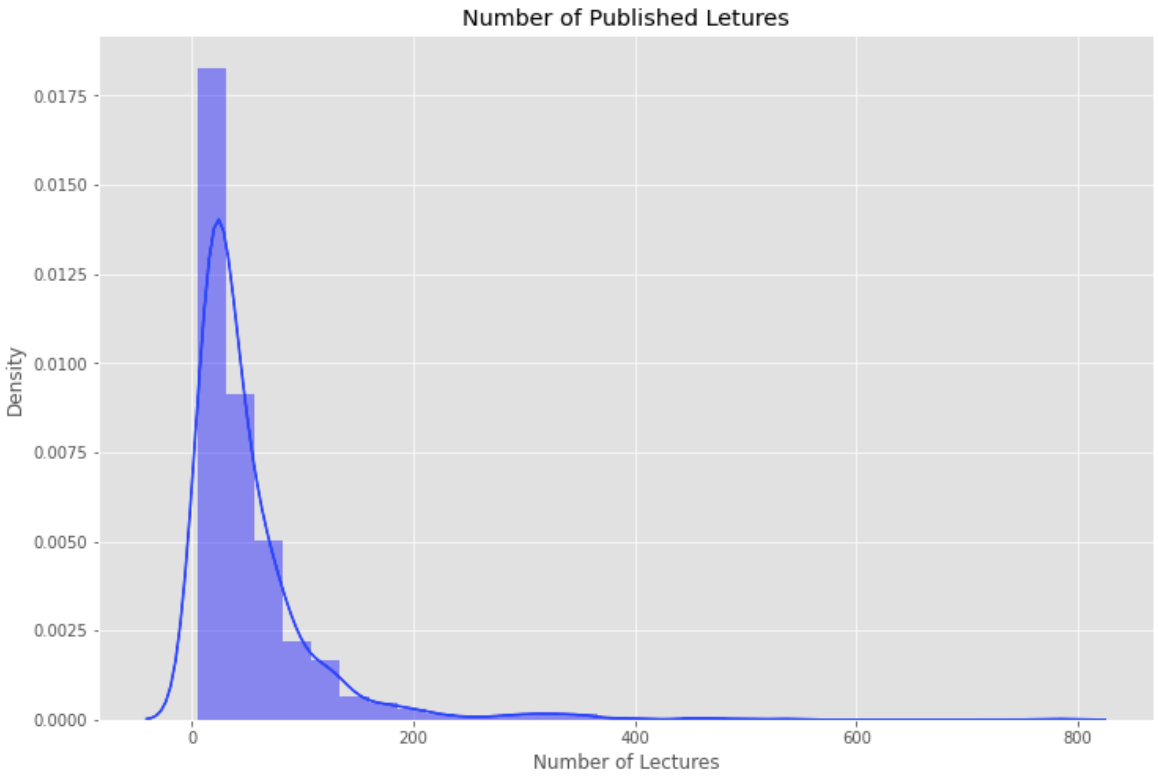
The dataset for this project is a combination of four data sets. While the aggregation of the data is a novel approach, the data itself was obtained from open repositories Kaggle and Data.World. Three datasets relate to the platform EDX. These are edx_courses, mooc_2016, and mooc_2017 respectively. The fourth dataset offers insight into user behavior within a specific subfield of course offering: Web development on Udemy. This dataset is known as, appropriately, 'web_dev.' All four datasets are flat format, CSV files that were imported into a Jupyter Notebook. When the final dataset was created it was converted to a CSV file for use in other Python-based environments, R Studio and Tableau for later visualization. The four data frames were concatenated using Pandas. The NA values were imputed as 0s in order to maintain the structure and integrity of the data. The final dataset contains over 400,000 rows and 67 columns. Since the data will be used for the purpose of creating models that will be assessed by classification algorithms, specifically logistic regression, the variables of concern were exclusively numeric (int and float, respectively).

Methodology

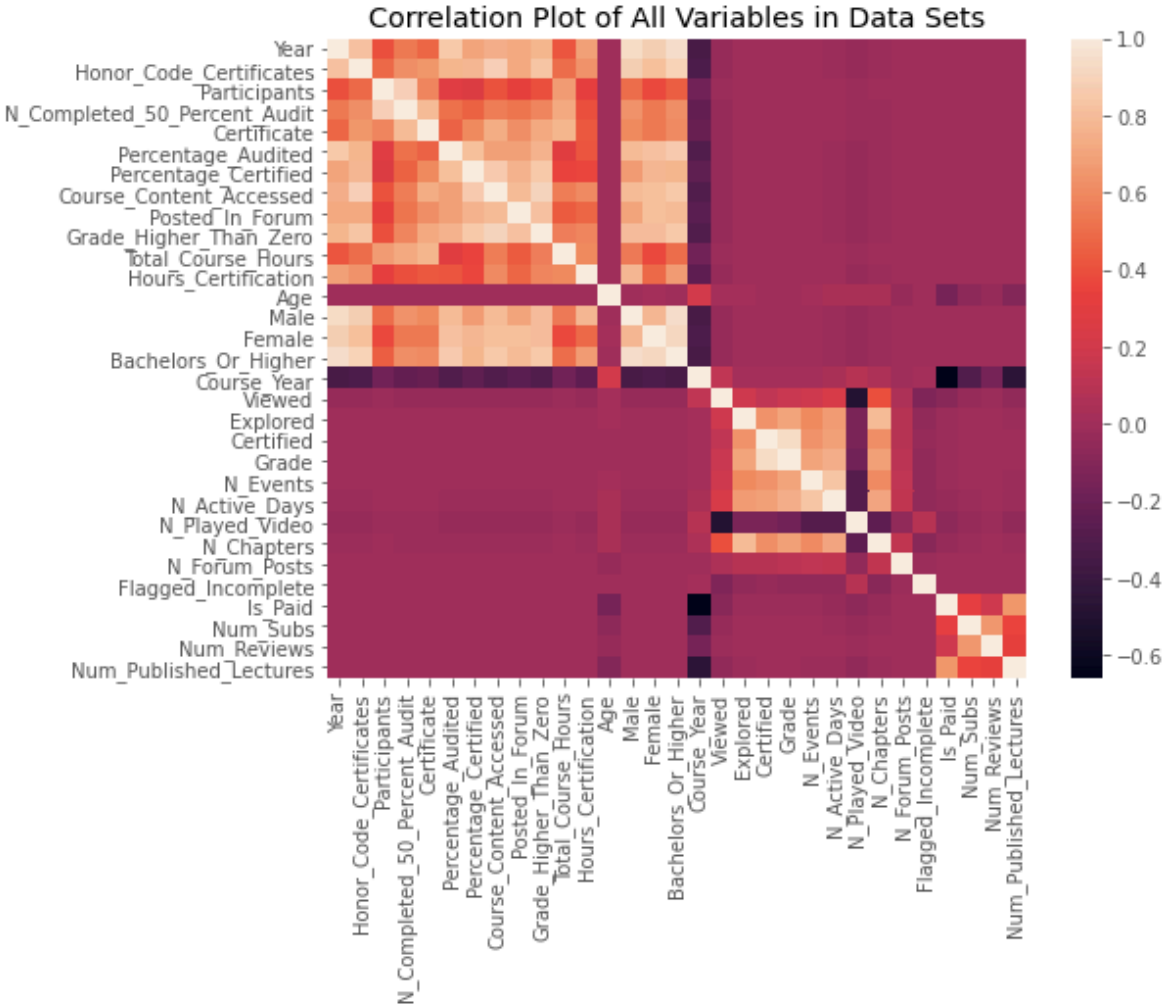
After loading the four files, an exploratory data analysis was conducted in order to identify and leverage univariate instances and multivariate correlations for later model creation.



Scatter plots like the above example, detailing the linear relationship between subscriptions and reviews helped to visualize overarching trends within the data, which will be helpful with feature selection and engineering in the next phase of the project. The variables examined, while almost exclusively numeric, can be divided into three broader categories of demographics (age, gender, bachelors degree), content (number of lectures, number of reviews, number of subscriptions) and engagement (number of chapters consumed, number of video lectures watched, number of posts in discussion forums). With a large data set while the observation of bivariate relationships is important, it is also necessary to note the instances of individual variables like the following:



After exploring and noting the relationships inherent within the initial datasets, the data was concatenated. Since there were no similarly named columns to perform inner or outer joins on the data was stacked vertically. Zeroes were imputed for null or NA values and a Pearson's correlation test was performed. This test resulted in a correlation matrix, which yielded the heat map inserted below:



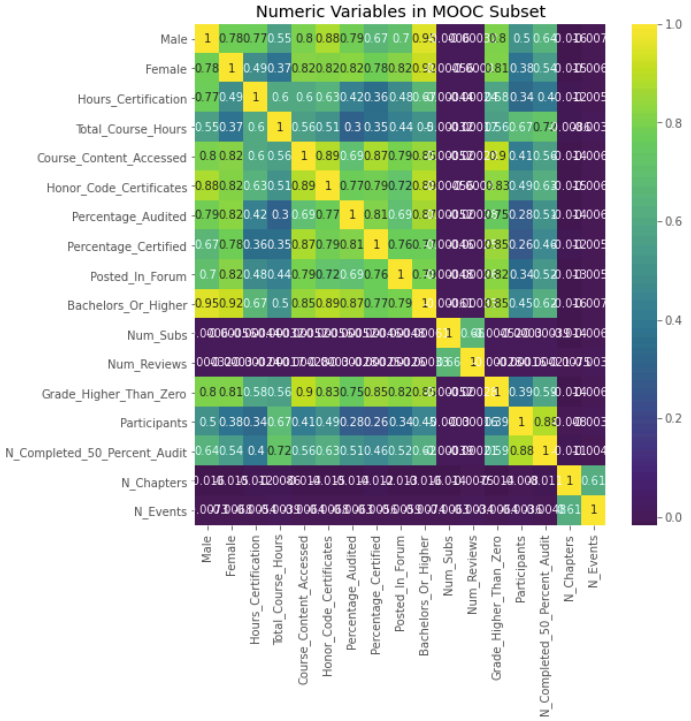
These 67 variables were then plotted on scatter plots, histograms and heat maps to determine further correlative trends. At this point in the exploratory data analysis several trends have emerged which are necessary to include below due to the overwhelming volume of information included in the above heat map.

- There is a positive correlation between those possessing a bachelor's degree or higher and those who have audited a portion or all of a given course (0.87).
- There is a significant correlation between the number of participants in a given course and the number of participants who have audited more than 50% of that course (0.88).
- There is a correlation between those who have posted in a discussion forum and have earned a grade above zero (0.82).
- There is a fairly strong correlation between number of chapters accessed and days invested in taking a course (0.71).
- There is a slight correlation with total course hours and number of certificates earned (0.6)

Following this step, the numeric variables identified as having positive correlations were combined as a subset into a separate correlation matrix and heat map. However, given the amount of variables in this rather large dataset, there were still columns that were being underutilized and were, frankly, irrelevant to the analysis. Therefore, another sub set was created containing numeric variables with high correlations. Variables integrated into this sub set include the following:

- Male
- Female
- Hours_Certification
- Total_Course_Hours
- Course_Content_Accessed
- Honor_Code_Certificates
- Percentage_Audited
- Percentage_Certified
- Posted_In_Forum
- Bachelors_Or_Higher
- Num_Subs
- Num_Reviews
- Grade_Higher_Than_Zero
- Participants
- N_Completed_50_Percent_Audit
- N_Chapters
- N_Events

The final correlation matrix, visualized on a heatmap, is inserted below:



The next step in this project will be to create a multivariable regression model to determine which variables can most reliably predict a binary outcome that will represent a student completing or not completing a course. Based upon the exploratory data analysis detailed in the above sections, the following predictor variables will be combined in several multivariable regression models:

- Hours_Certification
- Total_Course_Hours
- Course_Content_Accessed
- Percentage_Audited
- Percentage_Certified
- Posted_In_Forum
- Bachelors_Or_Higher
- Num_Subs
- Num_Reviews
- Grade_Higher_Than_Zero
- Participants
- N_Completed_50_Percent_Audit
- N_Chapters
- N_Events

Since between all four datasets there was no binary variable that specifically says 'Completed or not completed', the following variables will be tested as adequate targets:

- Certified
- Viewed
- Explored
- Honor_Code_Certificates

A multivariable regression model will be created using a 75-25 partition of the dataset. Given the large amount of data it makes more sense to create a larger test set to ensure precision on a section of data the model has not encountered prior to the test. The target variables will remain binary so that it is absolutely clear how exactly the predictor variables influence the targeted outcome. At least three multivariable logistic regression models will be constructed. The first will represent demographic characteristics and include a combination of 'bachelors or higher', 'participants' and 'N_Completed_50_Percent_Audit.' The second model will represent content such as 'Num_Subs', 'Num_Reviews', and 'Total_Course_Hours.' The final models will represent engagement and will likely include a combination of 'Course_Content_Accessed', 'Posted_In_Forum' and 'Grade_Higher_Than_Zero among other variables that have demonstrated high correlations in these areas. The hope is that by combining variables in smaller combinations the results will yield more insight into not only user behavior, but also content performance and user demographics. By obtaining analyses related to these areas it will be possible to present a more specific and comprehensive report of the model's performance in order to advise stakeholders on further retention measures and intervention strategies.

The model will be assessed by notable classification algorithms such as a Support Vector Machine methodology and Naïve Bayes. Since determining the influence of variables on an outcome is a classification problem, it makes sense to leverage such algorithms for the purpose of classification.

Results

The exploratory data analysis and analysis of correlation coefficients were integral in generating a high accuracy, precision and recall rate for the model. For a 75-25 split the multivariable models were very precise, with an accuracy that ranged between 95 and 99 percent. Three separate multivariable models were deployed to determine the binary outcomes of whether or not a student would complete a course ('certified') whether or not a student would earn an honor code certificate ('honor_code_certificate') and whether or not a student would view at least one module in a course ('viewed'). These designations were label encoded and represented as binary variables (0 or 1) for analysis. In pursuing the answer to this project's core research question, whether or not a student would finish a particular course, the model combined fourteen variables (listed in full on page twelve of this report) and fit a multivariable regression classifier to make the binary prediction. The first model was the weakest out of the three. While its accuracy rate was 97 percent, the model's precision and recall statistics were considerably lower, at 75 and 60 percent respectively. While a 97 percent accuracy sounds ideal, it is difficult to say, with absolute certainty, that this model is that accurate because the precision and recall rates mean that it did not reliably discern between true positives, true negatives, false positives and false negatives over 75 percent of the time.

The next model reduced the number of independent variables by more than half to six. This time, the model combined hours of certification, course content accessed, percentage audited, the history of a student posting in a discussion forum, the number of students with a bachelor's degree or higher and a grade higher than zero to predict whether they would earn an honor code certificate and, ultimately, pay for such a distinction (as opposed to conducting a free audit). The reduction of variables resulted in a far more reliable model with an accuracy rate of 96 percent, a precision of 99 percent, and a recall of 95 percent. The final model combined the least amount of variables, ingesting only four columns: number of chapters, number of events, number of reviews and number of subscriptions to predict how many students would view at least one module. Like the previous model, this combination of a few highly positively correlated variables yielded a high accuracy rate of 95 percent, but more importantly, a precision rate of 99 percent, and a recall rate of 95 percent. However, this model could better discern between actual and predicted labels than the previous two iterations.

Conclusion

With the massive online open course (MOOC) sector projected to grow by over 15 billion dollars in the next three years, there are both vested business and academic interests in studying the factors that contribute to the successful retention of students throughout the duration of a particular program, certification or micro degree program (Sugla, 2020). Following the COVID-19 pandemic, more and more institutions embraced and legitimized the idea of asynchronous and synchronous distance learning strategies. Since the U.S. hit a decades-high 13% unemployment

rate in March and April of 2020, adults at all stages of their careers have turned to MOOCs for supplemental learning, professional certification and exploration (Kochhar, 2020). However, in the nearly eight years of MOOC prominence a vexing trend has emerged: Students will begin a course and, shortly after enrollment, neglect its offerings or drop out entirely, at an attrition rate of nearly 96% over a period of the last five years according to estimates (Murray, 2019). Therefore, this project applied a multiple logistic regression methodology to a comprehensive data set of data for those enrolled in over 1,000 courses on the EDX platform to select features to create a model that will predict the likelihood of an individual completing a course on one of the world's largest and most reputable MOOC platforms. Discoveries made throughout the course of this project could result in recommendations that enable platform administrators, content designers and educators to proactively engage students who are more likely to continue for the duration of the course, rather than trying to simultaneously appeal to tens of thousands of online learners with disparate interest levels. This project has demonstrated that by emphasizing select demographic traits (such as the possession of a bachelor's degree), historic behavior (like how many modules one has consumed or forum participation) and content attributes (such as number of subscribers) it is entirely plausible to craft interventions that can preempt student churn in massive open online courses.

References

Altexsoft. (2019). Customer Churn Prediction Using Machine Learning: Main Approaches and Models. *KD Nuggets*. https://www.kdnuggets.com/2019/05/churn-prediction-machine-learning.html

Chung, C. (2015). MOOC-Takers Practice Flexibility: Class Central Survey Results. *The Report by Class Central*. https://www.classcentral.com/report/mooc-takers-practice-flexibility/

Data Bridge. (2020). Massive Open Online Course (MOOC) Market will grow at a rate of 40.50% by 2027: High Demand Due to COVID-19 Companies Profiled LinkedIn, FutureLearn, Udemy and More. *Data Bridge Market Research*. https://www.globenewswire.com/news-release/2020/08/24/2082887/0/en/Massive-Open-Online-Course-MOOC-Market-will-grow-at-a-rate-of-40-50-by-2027-High-Demand-Due-to-COVID-19-Companies-Profiled-LinkedIn-FutureLearn-NovoEd-Udemy-and-more.html

Haber, J. (2013). MOOC Attrition Rates – Running the Numbers. *The Huffington Post*. https://www.huffpost.com/entry/mooc-attrition-rates-runn_b_4325299

Hone, K. & Said, G. (2016). Exploring the Factors Affecting MOOC retention: A Survey Study. *Computers and Education* (98), pp. 157 – 168.

Horstmeyer, T. (2020). Customer Churn and Retention: Top 25 Stats You Need To Know. *Gravy Solutions*. https://www.gravysolutions.io/post/customer-churn-and-retention-top-25-stats-you-need-to-know#:~:text=Avoidable%20customer%20churn%20is%20costing%20U.S.%20businesses%20%24136%20billion%20a%20year.&text=9.,due%20to%20poor%20customer%20service.&text=10.

Jordan, K. (2015). Massive Open Online Course Completion Rates Revisited: Assessment, Length and Attrition. International Review of Research in Open and Distributed Learning, 16(3). https://files.eric.ed.gov/fulltext/EJ1067937.pdf

Koller, D., Ng, A. & Chen, Z. (2013). Retention and Intention in Massive Open Online Courses: In Depth. *Educause*. https://er.educause.edu/articles/2013/6/retention-and-intention-in-massive-open-online-courses-in-depth

Lederman, D. (2019). Why MOOCs Didn't Work, in 3 Data Points. *Inside Higher Ed*. https://www.insidehighered.com/digital-learning/article/2019/01/16/study-offers-data-show-moocs-didnt-achieve-their-goals

Murray, S. (2019). MOOCs Struggle to Lift Rock-Bottom Completion Rates. Financial Times. https://www.ft.com/content/60e90be2-1a77-11e9-b191-175523b59d1d

Pierre, K. (2014). How much do you study? Apparently 17 hours a week is the norm. *USA Today*. https://www.usatoday.com/story/college/2014/08/18/how-much-do-you-study-apparently-17-hours-a-week-is-the-norm/37395213/

Bibliography

Kochhar, R. (2020). Unemployment rose higher in three months of COVID-19 than it did in two years of the Great Recession. *Pew Research Center*. https://www.pewresearch.org/fact-tank/2020/06/11/unemployment-rose-higher-in-three-months-of-covid-19-than-it-did-in-two-years-of-the-great-recession/

Sugla, S. (2020). MOOC Market by Component (Platforms (XMOOC and CMOOC), Services), Course (Humanities, Computer Science and Programming, and Business Management), User Type (High School, Undergraduate, Postgraduate, and Corporate) and Region - Global Forecast to 2023. *Markets and Markets*. https://www.marketsandmarkets.com/Market-Reports/massive-open-online-course-market-237288995.html

Appendix A: Datasets

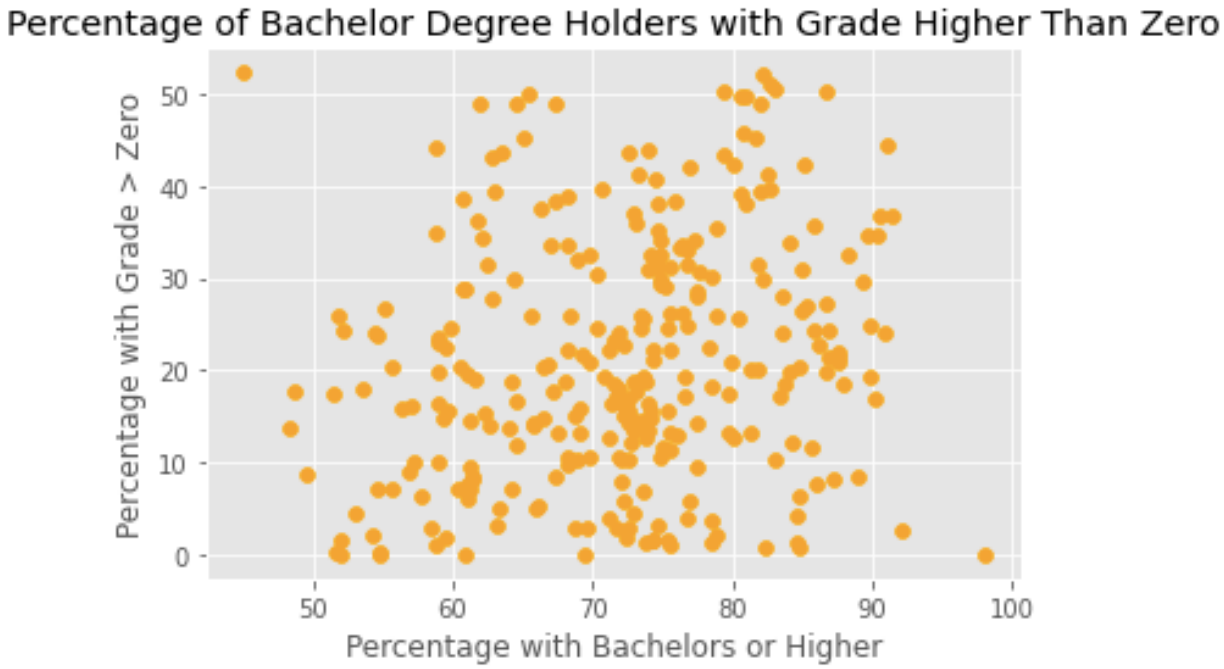
MOOC data set (Kaggle):
https://www.kaggle.com/kanikanarang94/mooc-dataset

Online Courses from MIT and Harvard (Kaggle):
https://www.kaggle.com/edx/course-study

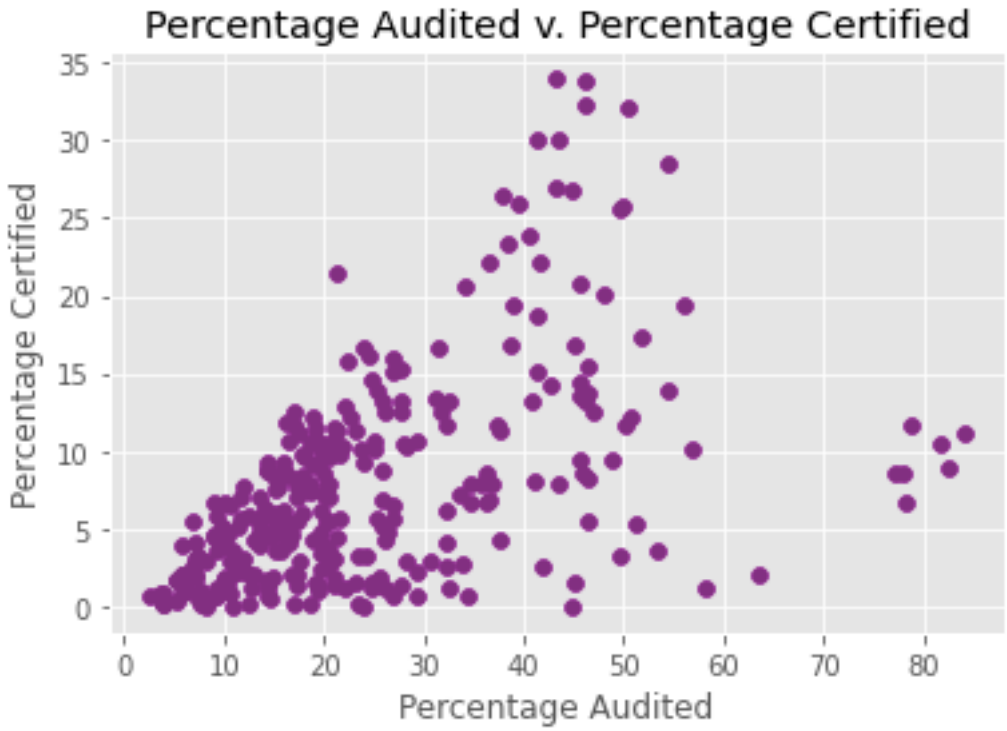
EDX Courses (Kaggle):
https://www.kaggle.com/imuhammad/edx-courses

Web Development (Data.World):
https://data.world/chasewillden/web-development-courses-from-udemy

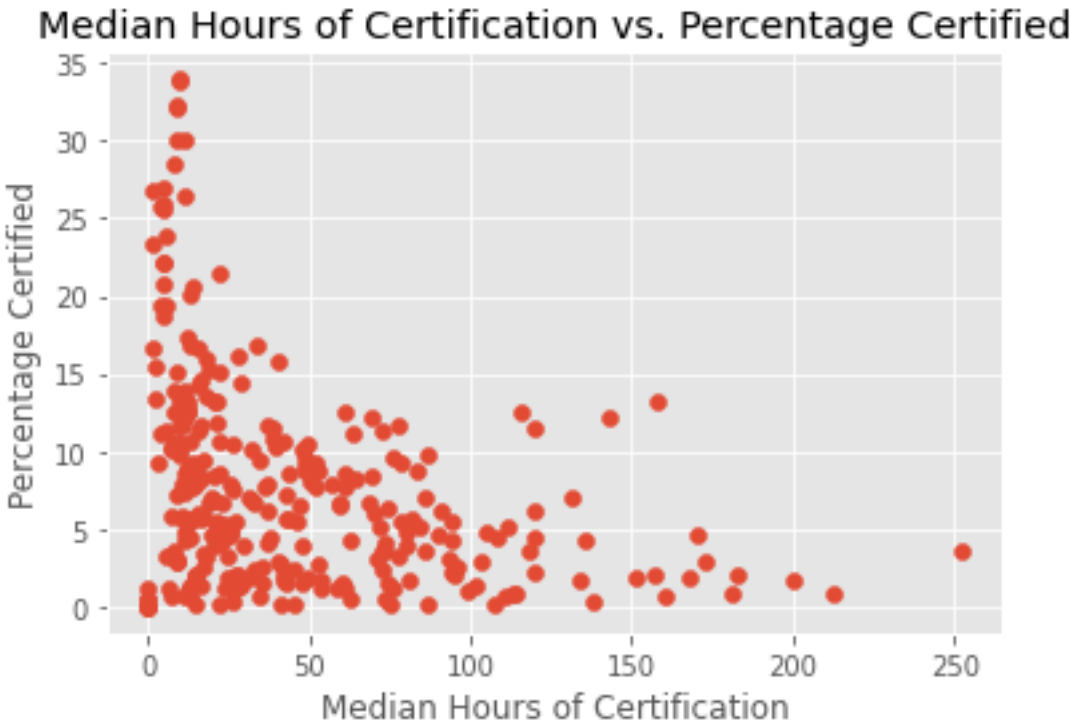
Appendix B: Influence of Bachelor's Degree on Grade Performance



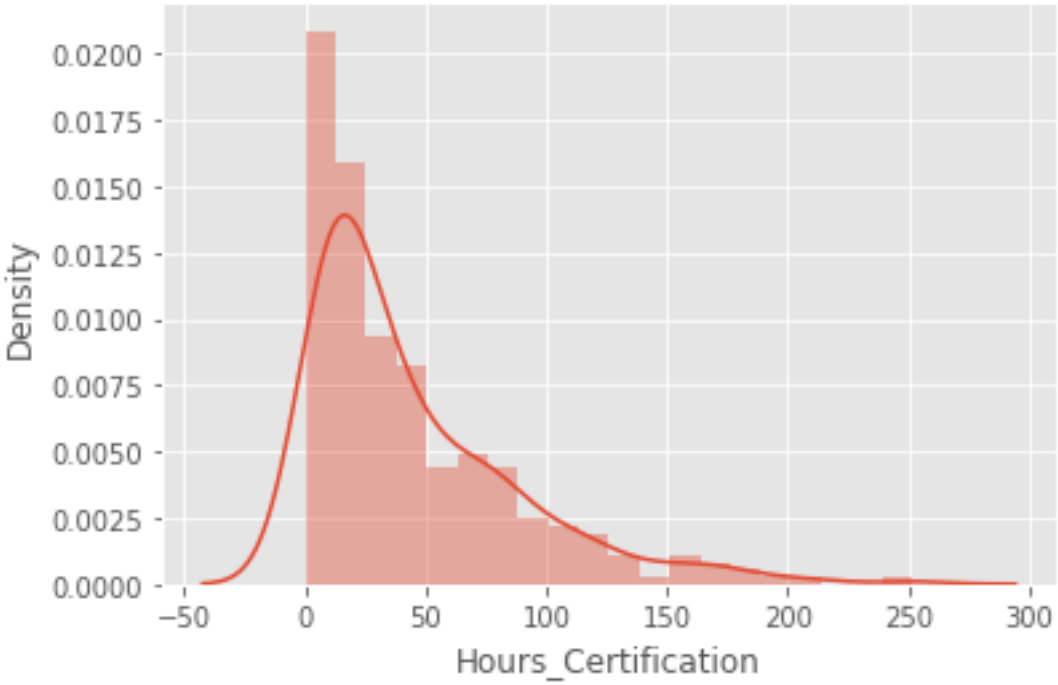
Appendix C: Percentage Audited vs. Percentage Certified



Appendix D: Median Hours Required for Certification vs. Population Certified



Appendix E: Distribution of Hours Required for Certification



Appendix F: Price of Course vs. Number of Published Lectures Available