

Review of MOOC Platforms

Jennifer Bochenek, Joan Kibaara, Toby Obunadike

Information Visualization

INFO 633

Spring 2020

Abstract- In this paper we are examining course enrollments, course types, and exactly what course topics are covered by three major MOOC platforms: Coursera, EdX, and Udemy. We examined student enrollment by course difficulty and over time. We found that beginner level courses or those that have no difficulty level have the highest enrollment and that course enrollment has been increasing over time. Then we looked at what course categories were popular with students and which course categories were more common on each of the platforms. Each platform had an issue with course offerings and student enrollment mismatch, where the proportion of students enrolling in a category of classes either far outnumber the classes offered or vice versa. Finally, we looked at course titles to determine which topics were more common on each platform and how that builds on the course catalog offerings. The results of that analysis backed up our previous results and added more detail. Our conclusion was that each platform had different offerings, but they do not seem to be providing additional content based on user needs and interests.

Index Terms- MOOC, Education, Udemy, Coursera, EdX

I. INTRODUCTION

MOOCs exist only because of advances in internet-based technology and a need for education and credentialing that was not being met by the higher education system. Students are increasingly turning to MOOC platforms for upskilling and credentialing. However, what types of courses and how many students can vary across the different MOOC types and platforms. In this paper, we will examine student enrollment and what types of courses are offered at three major MOOC platforms.

II. RELEVANT LITERATURE

The term MOOC was first used in 2008 by Dave Cormier to describe Stephen Downes' and George Siemens' 'connectivist' or distributed peer learning model (Baturay, 2015; Yuan & Powell, 2013). The first MOOC content library based on this idea was launched in 2011 as free access to the course 'Introduction to Artificial Intelligence' taught by Stanford University professors Sebastian Thrun and Peter Norvig (Baturay, 2015; Belleflamme & Jacqmin, 2014; Yuan & Powell, 2013). Two other Stanford University professors, Andrew Ng and Daphne Koller, founded the first dedicated MOOC platform in 2012 (Belleflamme & Jacqmin, 2014). They named this platform Coursera. Sebastian Thrun went on to found Udacity in late 2012 (Baturay, 2015). Udemy was founded in 2010, but based on review of the literature has not been considered part of the birth MOOC (Conache, Dima, & Mutu, 2016). This could be in part that it is described in the literature as a portal for MOOC (Spyropoulou, Pierrakeas, & Kameas, 2014) or a "consortium that allows anyone to teach and participate" (Conache, Dima, & Mutu, 2016).

There are a variety of MOOC providers, in addition to the Udacity, Coursera, and Udemy mentioned previously. The fourth major MOOC platform is considered EdX, a joint venture between Harvard and MIT that has since expanded to other University partners (Ibn El Ahrache, Badir, Tabaa, & Medouri, 2013; Bucovetchi & Stanciu, 2018; Nagasampige & Nagasampige, 2017). Other MOOC providers include FutureLearn, Khan Academy, LinkedIn Learning, iversity, Kadenze, Peer to Peer University, and more. MOOCs are either for-profit (such as Coursera, Udemy, Udacity, Kadenze, iversity, etc) or non-profit (for example EdX, Khan Academy, Peer to Peer University, etc) and they offer free or paid courses or a mixture of the two.

The MOOC learning model is based on either connectivist theory (called cMOOCs) or traditional learning methods (called xMOOCs) (Laxmi & Gure, 2015; Baturay, 2015; Belleflamme & Jacqmin, 2014). There are four main activities that distinguish connectivist MOOCs according to Stephen Downes (2011): aggregation of content that each student can select which content seems important based on the course material, associated remixing of material, repurposing concepts and ideas students learn, and feeding forward student materials, which means sharing student-generated materials. MOOCs are dependent on two features, open access - anyone can join - and scalability - courses are designed to support a large number of participants.

Multiple reports compare the different features (Ejreaw & Drus, 2017; Ibn El Ahrache, Badir, Tabaa, & Medouri, 2013; Ong & Grigoryan, 2015; Tsironis, Katsanos, & Xenos, 2016) or the number of users (Conache, Dima, & Mutu, 2016; Bucovetchi & Stanciu, 2018) across MOOC platforms, but none really cover what types of content MOOCs are offering and how that compares to the number of users that consume that content. The closest we could find compared math course offerings across four MOOCs (Intuit, Coursera, EdX, and Udemy) but only briefly touched on the users (Semenikhina, Drushlyak, Bondarenko, Kondratiuk, & Ionova, 2019). Instead, they focused on the quality of the math course material (Semenikhina, Drushlyak, Bondarenko, Kondratiuk, & Ionova, 2019). Their narrow focus and lack of description about the comparative enrollment in students in math courses compared to other course offerings is what spurred this research topic.

III. METHODS

A. MOOC platforms

Coursera was started in 2012 as a for-profit company. Coursera partners with many universities to offer its content. This platform offers courses that can be taken on their own and series of courses known as specializations. Both courses and specializations can result in certification after passing. Courses can be audited for free but must be paid in order to earn certificates or progress toward a specialization. Coursera recently adopted a pay by month feature for

specializations called Coursera Plus, which allows learners to pay a set rate per month for a specialization, rather than a flat fee. Courses typically include quizzes, videos, readings, and projects. The projects are peer-graded.

EdX was launched as a non-profit in May 2012 as a joint project between Harvard and MIT. It has since expanded to over 120 institutional partners. EdX offers solitary courses in addition to programs that cover multiple courses. Courses and programs are free, but users may pay for a verified certificate at the beginning of the course. Content is presented via lectures while grades are determined by quizzes and/or projects.

Udemy is a for-profit enterprise launched in 2010 and differs from the other platforms mentioned here in that the content is not created by university partners but by individual contributors. Instructors determine their course contents and price and upload the materials to the platform. Learners then can choose which courses they would like to take. Some courses might be linked by theme and instructor, but there are no official programs or specializations offered. Grades are determined by quizzes, while projects are ungraded. Udemy also has Udemy for Business, which is directed at employers to help upskill their employed force.

B. Datasets

1. Coursera data

The Coursera dataset is an excel file that contains information about the courses offered by Coursera. This dataset is a combination of two datasets and has 570 instances and 10 attributes which can also be referred to as the columns. As previously mentioned, the first data contained information about the courses like who organizes the course. However, it lacked some information we needed in order to compare the three datasets. Then we joined this dataset with a second Coursera dataset that contained more information like the course category.

2. Udemy data

The Udemy data is a CSV file that contains information about the courses offered. The dataset has 3679 instances and 12 columns. The Udemy dataset gives an

in-depth statistical view of the courses being offered. To elaborate, the dataset focuses more on the courses than the users of the course. Columns like the number of reviews, content duration, and the number of subscribers, that it gives us more information about a particular course.

3. EdX data

The EdX data is a CSV file which contains information about the EDX offered courses, and also information about the users of the platform. When compared to the other two datasets, this is the most detailed, in relation to the user information. This dataset contains 23 columns and 291 instances. It is more detailed because it tells us what percentage of the users watched a video or accessed 50% of the course content. Another important factor of this EDX data is that it also contains data about the last time a course was offered which is useful in determining what's on-demand and why the courses are offered. In relation to the users, this dataset gives us a percentage of male and female users/students.

4. Combined data

In order to facilitate data analysis, we combined similar variables across each dataset and merged the data end-to-end. Additionally, we further cleaned the datasets by reducing the number of variable levels that were introduced by this merger. For example, for the difficulty levels of courses, one had named it 'beginner' and another had called it 'beginner level'. We simplified that down to just beginner to ease comparisons across the platforms. One area that needed cleaning for our analyses was with the course category names. Coursera originally had 11-course categories, Udemy had five, and EdX had four. Using EdX's four as a base, we mapped the Coursera and Udemy categories onto EdX's. This mapping resulted in six total categories that will be discussed later in the paper.

popular for the given categories. Last but not least, we analyzed the courses offered in each platform, to determine which topics were offered.

A. Student Enrollment

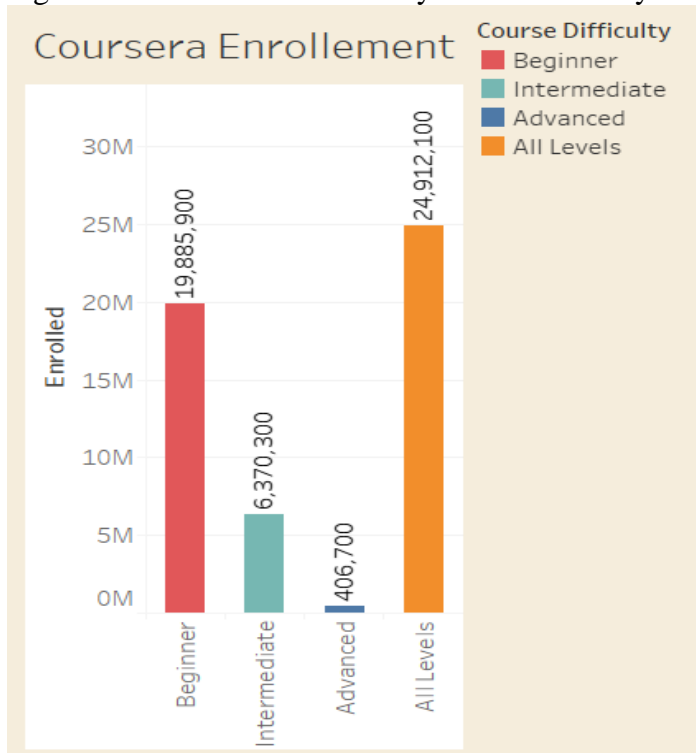
1. Coursera

Coursera platform was launched in 2012. In our analysis, the Coursera dataset has the most current course offerings since the data was obtained in May 2020. However, we do not have a detailed breakdown of each course's launch date. To date, the platform has over 50 million enrollment. The most popular courses are those that are offered to all level students with approximately 25 million enrollment. The beginner level courses are the second in popularity with about 20 million enrollment. Intermediate courses rank third in popularity with about 6 million subscriptions. The least popular courses are advanced-level courses, with only 400,000 enrollment to date.

IV. ANALYSES

We performed different analytics for each of the MOOC platforms. Our goal was to, first, understand the number of enrollment or subscriptions in each platform over a period of time. Next, we compared the course offerings in each of the platforms. We wanted to understand which platform was most popular and least

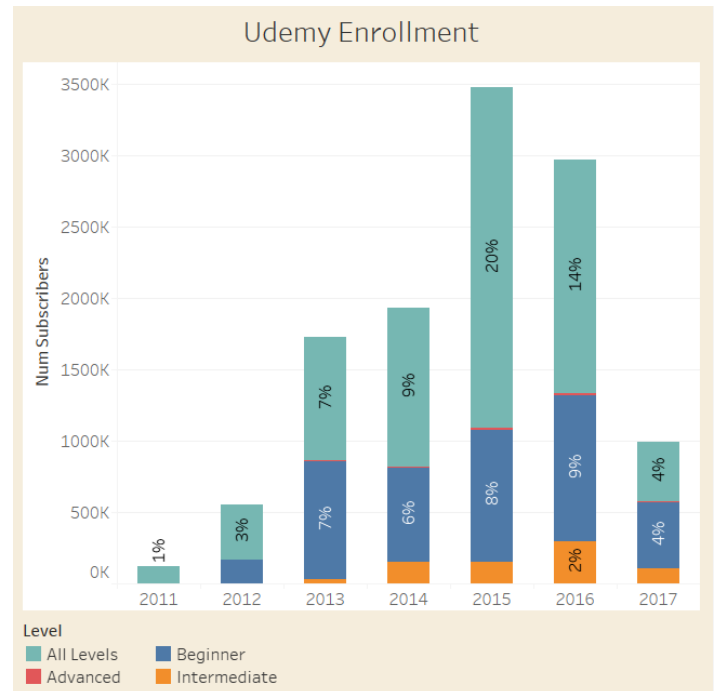
Figure 1. Coursera enrollment by course difficulty



2. Udemy

As mentioned, the first Udemy courses were launched in 2010. The available data is for each course's publication timestamp which ranges from 2011 to 2017. In the first five years of business, the total number of subscriptions increased each year. 2015 had the most number of subscribers. However, in 2016 and 2017, the total number of subscribers declined. Most subscriptions were on courses that can be taken by all levels. The first beginner-level course subscription was in 2012. These are the second most popular courses. Intermediate level courses are the third most popular courses, with the first course subscriptions in 2013. In 2016, the decrease in total subscriptions was due to a 7% decrease in the all level course subscriptions. However, there was about a 1% increase in both beginner and intermediate level courses subscriptions in the same year. The number of advanced-level courses subscription is less than 1% since their launch in 2013.

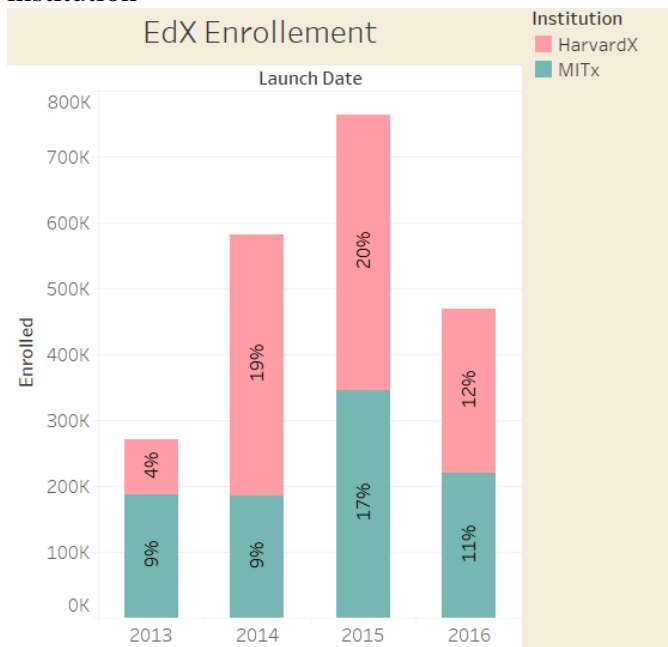
Figure 2. Udemy enrollment by course launch year and by difficulty



3. EdX

EdX was first launched in 2013. The data we have gives us information about each course's launch date. We included the institution that offers the course in our analysis of student enrollment. As we see, MIT and Harvard both offered courses in 2013, with MIT having 5% more enrollment than Harvard. In 2014, the enrollment to Harvard courses increased significantly by 15%, while MIT stayed constant at 9%. 2015 had the most enrollment with Harvard having 3% more enrollment than MIT. In 2016, there was a decline in total enrollment for both institutions. Harvard enrollment had the highest decrease of 8% compared to MIT which decreased by 6%.

Figure 3. EdX enrollment by course launch year and by institution



4. Conclusion on Enrollment

Coursera has over 51 million students, which is the largest across all platforms. Udemy ranks second with 11.8 million students, while EdX has the least with 4.4 million. Coursera data is more recent compared to EdX and Udemy which is based on 2017 and 2016 data respectively. 2015 had the most number of student enrollment in both Udemy and EdX. When comparing the popularity of courses by their levels, we see the same trend for both Coursera and Udemy. The all-level courses are the most popular courses, followed by beginner level, intermediate level and advanced level as the least popular course levels.

B. Course Offerings

We also looked at what types of courses were offered by each platform. In order to facilitate this, we used EdX's broad categories and collapsed the more fine-grain Coursera and Udemy categories onto the EdX categories. As a result, we found that there were six categories total, two of which (Business Finance and Musical Instruments) had no analog in the EdX categories.

We examined the course offerings in two different ways, the first by how many subscribers there were within each category by platform and the second by

how many courses were offered within each category by the platform. The comparison of these two tells us where the platforms are putting their efforts in terms of content creation compared to how their users are consuming the content.

Figure 4. Course Categories by Number of Subscribers and Platform

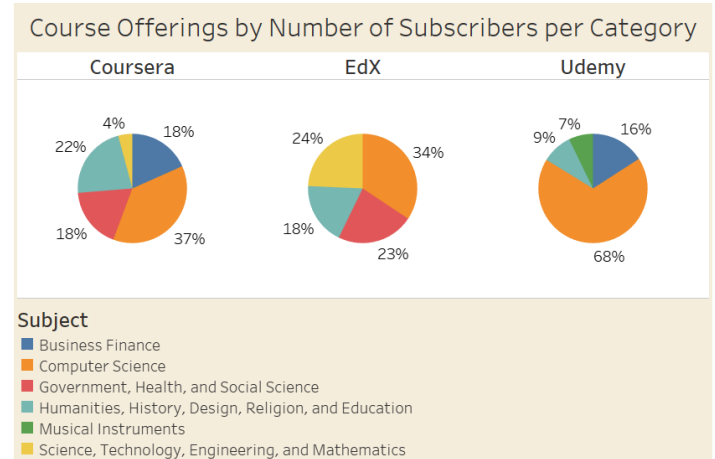
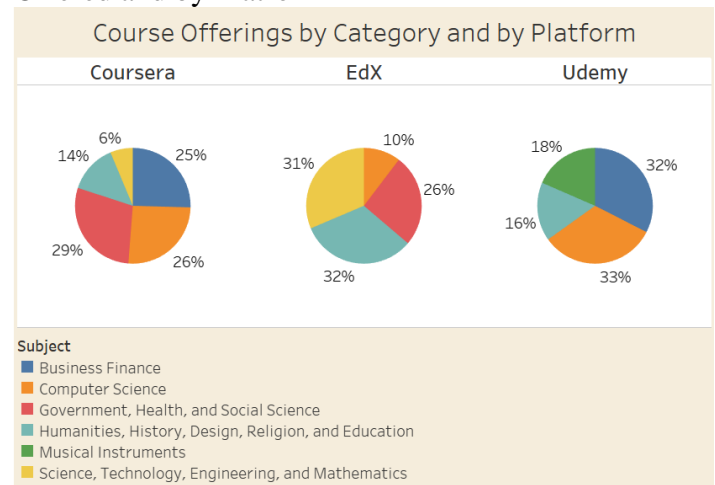


Figure 5. Course Catalog by Number of Courses Offered and by Platform



In figures 4 and 5 above, Coursera has the most categories with five total, with the most subscribers taking computer science classes (37%), while computer science was only the second-largest category in terms of numbers of courses offered (26%). The largest category of courses offered was government, health, and social science (29%), but that accounted for only 18% of subscribers. Only one other category besides computer science had a larger subscription base than the number of courses offered and that was in the humanities, history, design, and education portion

(accounts for 14% of courses offered, but 22% of subscriptions). This is interesting because Coursera could be doing a better job providing relevant content for their users by increasing the number of computer science and humanities classes.

EdX has a similar issue. For this platform, there were only four course categories, with no business classes offered, unlike the two other platforms. Although computer science accounts for only 10% of the courses offered, they also account for 34% of subscriptions. In comparison, the humanities, history, design, religion, and education category is 32% of the courses offered but only 10% of the subscriptions. The offerings and subscriptions for government, history, design, religion, and education courses as well as science, technology, engineering, and mathematics courses are more balanced.

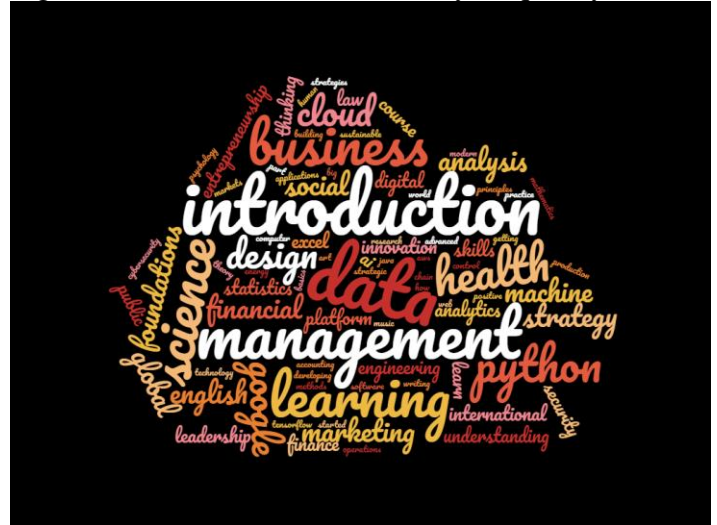
Udemy, as previously mentioned, does not create or curate their own content. Instead, the content is created and managed by individual instructors. Theoretically, in this more free-market style MOOC platform, instructors would be making content where the subscribers desire to see them. However, Udemy has the greatest imbalance of the three platforms. 68% of the subscriptions were in the computer science category, while that accounts for only 33% of courses. This is the biggest need gap seen across any of the platforms. Of note, Udemy is also the only platform of the three that includes musical instruments as a category, while this could have conceivably been collapsed onto the humanities category, we felt that would be losing a key category that is unique to Udemy. That difference could account for the fact that musical instruments make up 18%, but only 7% of subscriptions. Instead of Udemy being driven by a free-market type of economy, it seems to instead be backed by instructors creating content based on their passions and interests rather than to capitalize on where the most subscribers are.

C. Course topics

One method of examining the course offerings for each platform was to take the term frequencies of words in the course titles. Term frequencies are found by tabulating a count of each word (excluding common or

‘stop’ words such as ‘and’, ‘the’, ‘a’, ‘from’, etc) across all the course titles within each platform. The results are then plotted as word clouds where the size of the word indicates greater frequency.

Figure 6. Coursera course names by frequency



This first word cloud shows the high-frequency words for course titles of classes on Coursera. From this visualization, we can see that the words ‘data’, ‘introduction’, and ‘management’ show up often in course offerings. Other high-frequency words include ‘science’, ‘health’, ‘business’, and ‘python’. This is a reflection of what we saw earlier in the course catalogs, where the largest group of classes were computer science classes (26%). The word introduction is so prevalent because it is a word that is used across the different categories, same as ‘learning’ and ‘science’ (e.g., health science, data science, computer science, etc.) while the word ‘business’ is exclusively associated with the business finance category (25% of course offerings). ‘Management’ could refer to either computer science or business finance (e.g., data management, construction management, etc.).

This word cloud supports our previous findings, particularly with regard to the levels of courses offered and what categories of courses are offered. The levels can be seen from the inclusion of the word ‘introduction’ which supports the previously stated finding that beginner level courses had the highest level of enrollment. In the image we can see ‘foundations’ which is a synonym for introduction and further supports that beginner level classes are prevalent. The high numbers for enrollment are being boosted by the

large number of classes with beginner level content as seen by courses with introduction in the title. Of the top three categories of courses offered (computer science, humanities, and business finance), two are clearly reflected in the course topics, while the last might be too broad of a category to be seen in course titles. Instead, the fourth largest category - government, health and social science - can be observed instead. This is being driven by how concrete and similar course titles are for the health domain (e.g., public health and health care are two common phrases in class names).

Figure 7. EdX course names by frequency

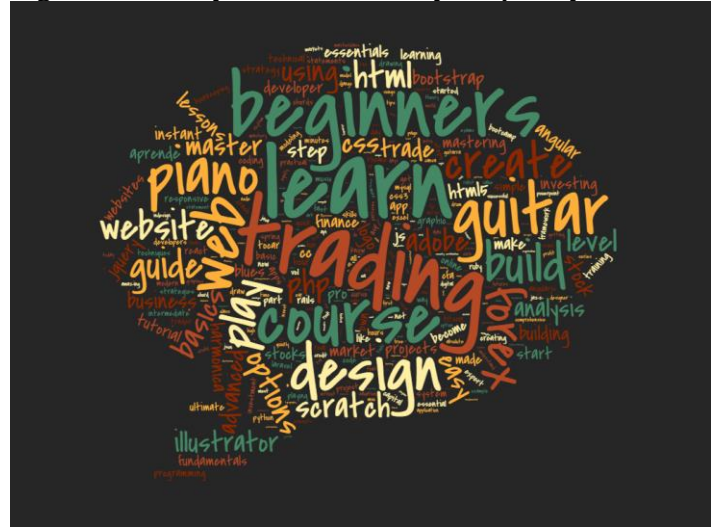


This word cloud is a visualization of the course titles available on EdX. Similar to Coursera, ‘introduction’ is the most popular word followed by ‘science’, ‘data’, ‘history’, and ‘analysis’. This is in line with course categories, as ‘science’, ‘data’ and ‘analysis’ all fall under the science, technology, engineering, and mathematics (31%) and all could also fall under computer science (10%). Together they account for 41% of the courses offered. This also highlights one of the limitations of the categories compared to course topics, in that course topics can fall under multiple categories and it is not a clear or clean distinction at times between what categories a word could belong to.

One interesting addition that this word cloud does add to our previous findings has to do with the word ‘introduction’. Previously we were not able to discuss findings for course difficulties for the EdX dataset as it is not an included variable; however, the word ‘introduction’ was the most common word. Thus we can infer that introductory, or beginners, classes are the

most common classes on EdX as they are on the two other MOOC platforms.

Figure 8. Udemy course names by frequency



The final word cloud covers Udemy's course offerings. This word cloud is very different in terms of word frequencies compared to the other two platforms. This may be attributed to the fact that course titles are at the discretion of the course creator/instructor as compared to being moderated by the platform as is with Coursera and EdX. The most frequent word was 'learn' as compared to the word 'introduction' that was most common on the other platforms. The third most frequent word is 'beginners' which can be likened to the word 'introduction' used by the two other MOOC platforms. The second most common word was 'trading' which was always associated with the business finance category. Udemy also had a strong showing of web development courses, as seen in the words 'web', 'php', 'website', 'css', and 'html', these fell under the computer science category which was 33% of the content but 68% of the subscriptions. Udemy's unique course category, musical instruments, is also displayed in this visualization with words like 'piano' and 'guitar' among the most frequent words.

By using the word clouds to augment our previous analyses, we were able to find additional support for some of our initial findings. This is particularly true of the analyses concerning the level of difficulty of the course, and the course categories. It was also helpful to fill in some of the missing analyses concerning level of difficulty for the EdX courses that we could not have found any other way.

V. CONCLUSION

MOOC classes are a relatively new way to earn new skills and credentials. The classes are offered on a variety of platforms and for this paper, we looked at three: Coursera, EdX, and Udemy. By examining student engagement with the platforms through enrollment or subscription to course offerings we were able to learn that students are more likely to enroll beginning level courses. Additionally, in general, the number of students is increasing over time.

By looking at what types, or categories, of courses are offered we determined that there are six main categories, although one of the six (musical instruments) was only on the Udemy platform. The number of courses in each category is often at odds with the level of student interest in each category. This is most clearly explained by Udemy's computer science courses. These courses account for 68% of student subscriptions or enrollment but comprises 33% of all courses offered on Udemy.

When looking at course topics through word clouds that show common words in course titles, we see that across all three platforms, courses with beginners (Udemy) or introduction (Coursera or EdX) are among the most common. We can also see the slight differences across the three platforms reflected in the course titles, with Udemy having musical instruments and both Udemy and Coursera having business finance courses while EdX does not. Thus, each platform offers nuanced differences in what they are providing for students, but given the differences between the number of courses offered in each category and the number of enrollments in each category, there may be a student need that is not being met.

APPENDIX

Full-sized versions of the word clouds for ease of viewing are also included.

REFERENCES

- Edlearn14 Proceedings, 6981-6990. overview of the world of MOOCs. *Procedia-Social and Behavioral Sciences*, 174(12), 427-433.
- Belleflamme, P., & Jacqmin, J. (2016). An economic appraisal of MOOC platforms: business models and impacts on higher education. *CESifo Economic Studies*, 62(1), 148-169.
- Bucoveţchi, Ö., & Stanciu, R. D. (2018). Multi-layer Analysis Regarding Massive Online Open Courses Attractiveness Towards Academic Environment. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, 9(1), 44-49.
- Conache, M., Dima, R., & Mutu, A. (2016). A Comparative Analysis of MOOC (Massive Open Online Course) Platforms. *Informatica Economica*, 20(2).
- Downes, S. (2011). "Connectivism" and connective knowledge. *The Huffington Post*. Retrieved from https://www.huffpost.com/entry/connectivism-and-connecti_b_804653
- Ejreaw, A., & Drus, M. (2017). The challenges of massive open online courses (mooc)—a preliminary review. *ICOCI Kuala Lumpur. Univ. Utara Malaysia [Internet]*, 25-27.
- El Ahrache, S. I., Badir, H., Tabaa, Y., & Medouri, A. (2013). Massive Open Online Courses: a new dawn for higher education?. *International Journal on Computer Science and Engineering*, 5(5), 323.
- Laxmi, V., & Gure, G. S. (2015). Massive open online courses: new imperative in education. *International Journal of Research in Engineering, IT & Social Sciences*, 5(3).
- Nagasampige, M., & Nagasampige, K. (2017). A qualitative study on usage and effectiveness of Massive Open Online Courses (MOOCs) in Indian university education system. *European Journal of Open Education and E-learning Studies*.
- Ong, B. S., & Grigoryan, A. (2015). Technological Innovations in Learning: A Focus on MOOC. *Sains Humanika*, 5(2).
- Semenikhina, O. V., Drushlyak, M. G., Bondarenko, Y. A., Kondratiuk, S. M., & Ionova, I. M. (2019, May). Open Educational Resources as a Trend of Modern Education. In *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 779-782). IEEE.
- Tsironis A., Katsanos C., Xenos M. (2016). Comparative usability evaluation of three popular MOOC platforms. In *IEEE Engineering Education Conference, IEEE EDUCON 2016* (pp. 608-612), Abu Dhabi, UAE: IEEE, 10-13 Apr. 2016.
- Yuan, L., & Powell, S. J. (2013). MOOCs and open education: Implications for higher education.