**Massive Online Open Course Analysis – Udemy**

# Introduction

Udemy is a massive online open course (MOOC) platform that offers both free and paid courses. The company has a variety of courses tailored to subscribers with different level of skills and interests.

Problem Statement

The dataset provided contains 3,682 records of courses from 4 subjects (Business Finance, Graphic Design, Musical Instruments and Web Design) spanning seven years. The task is to generate trends of the Udemy dataset to identify problems and areas of interest, and then make recommendations based on these analyses.

# Objectives

* Prepare the data for analysis (Data Cleaning)
* Analyze the data
* Interpret trends and identify key areas
* Make recommendations or propose solutions

# Data Cleaning

* Formatted columns to fit the type of the values they hosted
* Created new column to calculate revenue

# Analysis and Insights

## Descriptive Statistics

To get a general overview of the data being worked with, the following were identified over the period under consideration:

* Lectures
  + 229 distinct lectures
  + 148,000 total administered lectures
* Subscribers
  + Total of 12 million subscribers
  + 8 million on paid courses
  + 4 million on unpaid courses
* Courses
  + 2 course types-paid and unpaid
  + 3658 distinct courses
  + 3368 paid courses
  + 310 unpaid courses
  + Average course duration is 4.09 hours
* Price
  + 66.05 average price per course
  + 200 highest course price
  + 20 lowest course price
* Reviews
  + 578,000 total reviews received
* Revenue
  + $885,000,000 in total revenue

## Analysis By Subject

This focuses on generating charts that give life to the data from the perspective of each subject. This provides an opportunity to establish and compare relationships between variables. To do this, the following were determined:

* Total subscribers by subject
* Total lectures by subject
* Average price by subject
* Total reviews by subject
* Average duration by subject

Web Development was the most subscribed subject (8million subscribers) while Graphic design and Musical Instrument tied for the least subscribed (1 million subscribers each). Web development was the only subject on the platform in 2011, so this may have marginally contributed to the high total number of subscribers over the review period. Also, Web Development and Business Finance were the most expensive subjects on average, and this may have been because they had the greatest number of lectures. A Business Finance course for instance cost $68.55 on average and has 39,000 lectures in total, whereas a Musical Instrument course cost $49.56 and has a total of 26,000 lectures. The highest average price for all subjects was observed in 2017, on the high end a Web Development course cost $85.78 on average, while on the low end a Musical Instrument course cost $51.34.

Finally, Web Development was the most reviewed subject while Musical Instrument was the least reviewed. We posit that the high number of subscribers for web development may explain the high number of reviews and vice versa for musical instrument subject. The positive correlation of 0.6 suggests that when number of subscribers increase, the number of reviews increase. All subjects received their lowest number of reviews in 2017. It is worth noting that the proportion of reviews to subscribers is abysmal across all subjects. For the most subscribed subject only 2.3% of subscribers gave a review and for the least subscribed, 3.17% reviewed Musical Instrument while 4.50% reviewed Graphic design.

Analysis by subject indicates that 2 of the four subjects are very popular, only a small proportion of subscribers leave reviews. The most popular subjects Web Development and Business Finance are also the most expensive on average.

The slicers for these analyses were date (year).

## Analysis By Level

To further establish relationships within our dataset, a comparison based on the level of subscriber skill was analyzed. The following were the parameters for such comparisons:

* Average duration of course by level
* Average price by level
* Total subscribers by level
* Total lectures by level

The most subscribed to course level is all levels for the period under review. An observation of the average price across levels indicates that a subscriber gets more value for money by opting to go through all levels than picking a specific level. An all-levels course cost 73.02 on average over the period, while an expert level course for instance cost 91.12 on average. All level was consistently the most subscribed, except in 2017 where beginner level courses were the most subscribed (46.8%). Across subjects and the various years all levels have the highest total lectures while expert level has the least total lectures.

In 2011, all-levels was the only option available on the platform and in 2012 beginner and intermediate levels were introduced, then finally expert level was introduced in 2013. The beginner level was initially the most expensive on average until the expert level was introduced, then beginner level transitioned to being the least expensive level on average. Although all-levels courses have the highest average duration and expert level have the lowest average duration over the period. The average course duration across the years for the various levels has been fluctuating. With respect to expert level, the shortest average course duration (1.63) was observed in 2014, all-levels shortest course duration (4.11) occurred in 2017 while intermediate and beginner level shortest average course duration happed in 2015.

For expert level courses when the price dropped in 2014 the number of lectures available and average content duration dropped. Udemy operates a revenue sharing system with Instructors and the expert level is where instructors make the most money per course, as such dropping prices mean Instructors have less incentive to publish premium/detailed content at discounted prices. Instead, some opt out of publishing while others reduce the content duration. Subject wise this was observed for Business Finance and Musical Instruments at expert level.

Analysis by level suggest that as level progresses courses become more focused hence, they have a shorter duration. They however become more expensive, indicating the premium placed on knowing more.

The date (year) and subject were used as filters when a more focused analysis was required.

Analysis By Year

To establish performance and observe trends across the years, the following parameters were used:

* Number of subscribers and lectures year
* Number of reviews by year
* Average course price by year
* Average content duration

The number of subscribers and lectures consistently increased until 2015 for subscribers and 2016 for number of lectures. Meanwhile in relation to subjects, Graphic design and Web development were the subjects that didn’t experience consistent increase in number of subscribers until 2015. For Graphic design, number of subscribers slumped in 2013 by 50.2%, rose in subsequent years then dipped again in 2016 and 2017. However, for Web development the slumps occurred in 2014, 2016 and 2017.

Average price over the period under review has a zig zag pattern, the lowest price observation ($40.78) was in 2012 and the highest ($75.64) in 2017. For intermediate courses average price trended upward over the period. The biggest price dip occurred in 2014 for expert and beginner level but in 2012 for all-level. Highest average price across the four various levels occurred in 2017, except the beginner level whose highest average price was observed in 2012. Average course duration trended downward until 2016 when it increased by 9.90% then dipped again the subsequent year by 15.40%.

Although price rose and fell over the period, number of lectures continually increasing until 2016 suggests that the consistent growth in number of subscribers compensated enough for Instructors to keep making money. However, following the drop in number of subscribers in 2016 to 14.63% then to 66.67% in 2017, the number of lectures available in 2017 dropped by 50.36%. This suggests that the increase in price, did not compensate for lost revenue due to subscriber decline hence some instructors left the platform, thus explaining the drop in available lectures at the end of 2017.

Number of reviews increased year on year until 2015, with the highest increase (127.08%) happening in 2015. For subject specific observations, Web development was the only subject that followed the pattern of consistent increase in the number of reviews year on year until 2015. Across levels, pattern of number of reviews per year have been mixed. Highest number of reviews for beginner and intermediate level occurred in 2016 while the highest number for all-levels and expert level occurred in 2015

Analysis by year suggest that content duration, number of reviews, subscribers and lectures are on the decline while prices are inclining. 2017 was the worst year in terms of subscriber numbers across levels and subject and in effect revenue while 2015 was the best year for revenue and subscriber count.

The subject and course levels were used as filters when a more focused analysis was required.

Analysis By Revenue

To establish monetary performance over the period across levels and subjects, the following parameters were used:

* Revenue by subject
* Revenue by level
* Revenue by year

Revenue was on an upward trend until 2015 and this is consistent with the pattern observed in subscriber growth. The best performing year by revenue was 2015 ($315 million) and the sharpest drop 67.86% occurred in 2017. In terms of subject Web Development and Business Finance generated the most revenue and in terms of level, all-level and beginner level generated the most profit. Again, this is consistent with these subjects and levels having the highest number of subscribers. Overall revenue generation in each specific year by subject and level was consistent with their number of subscribers. Thus, the most subscribed subject/level per year generated the most revenue while the least subscribed subject/level per year generated the smallest revenue.

# Recommendations

Review content duration, this is an indication of content quality. On average across levels and subjects content duration has been on the decline. Producing quality content may help grow subscriber base again.

Implement mid-course reviews, this will help push course review numbers up and possibly attract new students in situations where the reviews are positive.

Review pricing and content duration for Graphic design and Musical Instrument. Prices trended upward on average while course duration trended downward. Improving subject content/duration to merit upward trending price may help make these subjects more popular.

Review pricing strategy at expert level, as it’s the least subscribed level. Overly discounting courses at this level adversely impacts content duration and number of available lectures, meaning subscribers do not get a lot of value out of this level.

# Conclusions

Analysis by year suggest that content duration, number of reviews, subscribers and lectures are on the decline while prices are inclining.

Analysis by level suggest that, as level progresses courses become more focused hence, they have a shorter duration. They however become more expensive, indicating the premium placed on knowing more. Discounting expert level courses also contributes to an even shorter course duration.

Analysis by subject indicates that 2 of the four subjects are very popular and only a small proportion of subscribers leave reviews for all subjects. The most popular subjects Web Development and Business Finance are also the most expensive on average.

The EdTech market has become highly competitive, as such, it has become increasingly important to deliver high quality content as well as value for money. Udemy needs to pay attention to course quality as the data suggests this may be one of the reasons for a decline in subscriber base and as a result revenue.