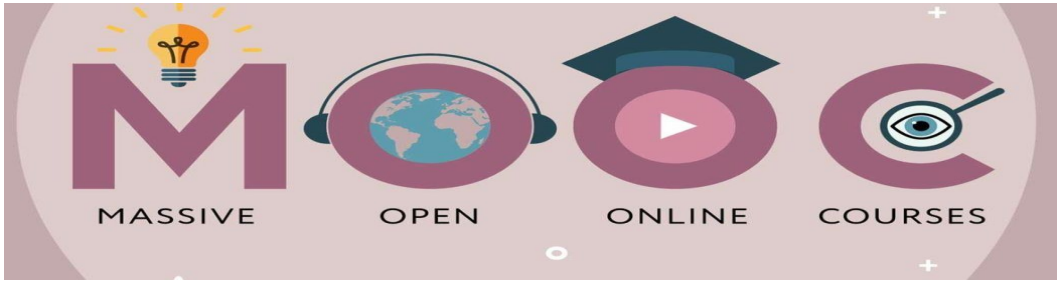
Class Central: Making informed online learning decisions

Compiled by Kailun Cheng

TOPICS EXPLORED:



Background:

- MOOCs Growth and Summary Statistics 2019
- Courses Platforms and Institutions

Courses Categorical Variables EDA:

- Subjects
- Institutions
- Certificate Offered and Language

Courses Numerical Variables Discussion:

- Duration Ambiguity: Feature Engineering
- Stars and Reviews: Feature Ranking

Subpopulation Case Study

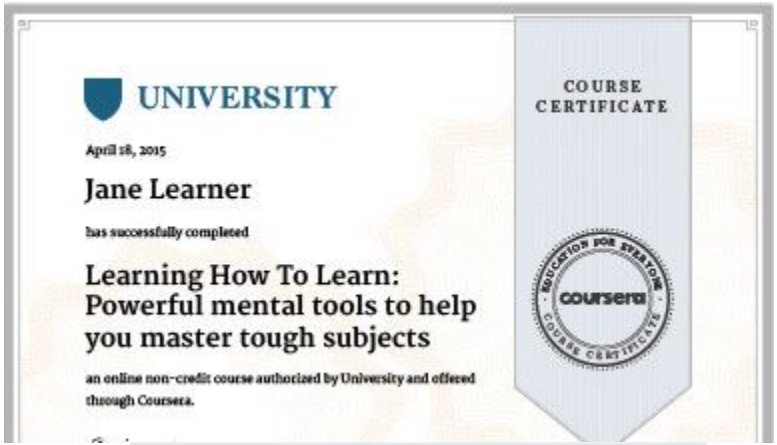
- Prediction of Course duration of top 50 English courses with certificates

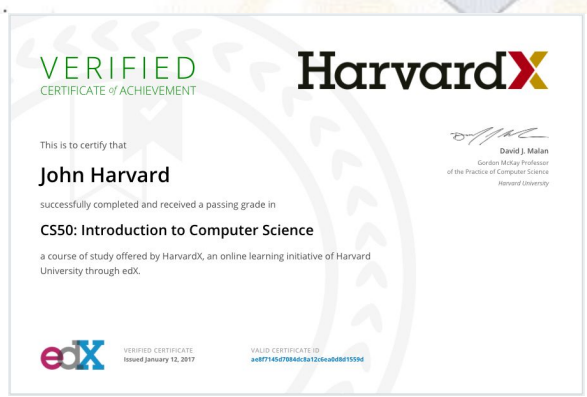
Background: MOOCs Growth and Current Statistics

MOOCs Advantages

- MOOCs mainly differ from online lecture notes or videos by learner feedback
- Feedback comes in the form of exams or projects that require the learner the complete to obtain a certification

- Characterized by flexibility





110M Students

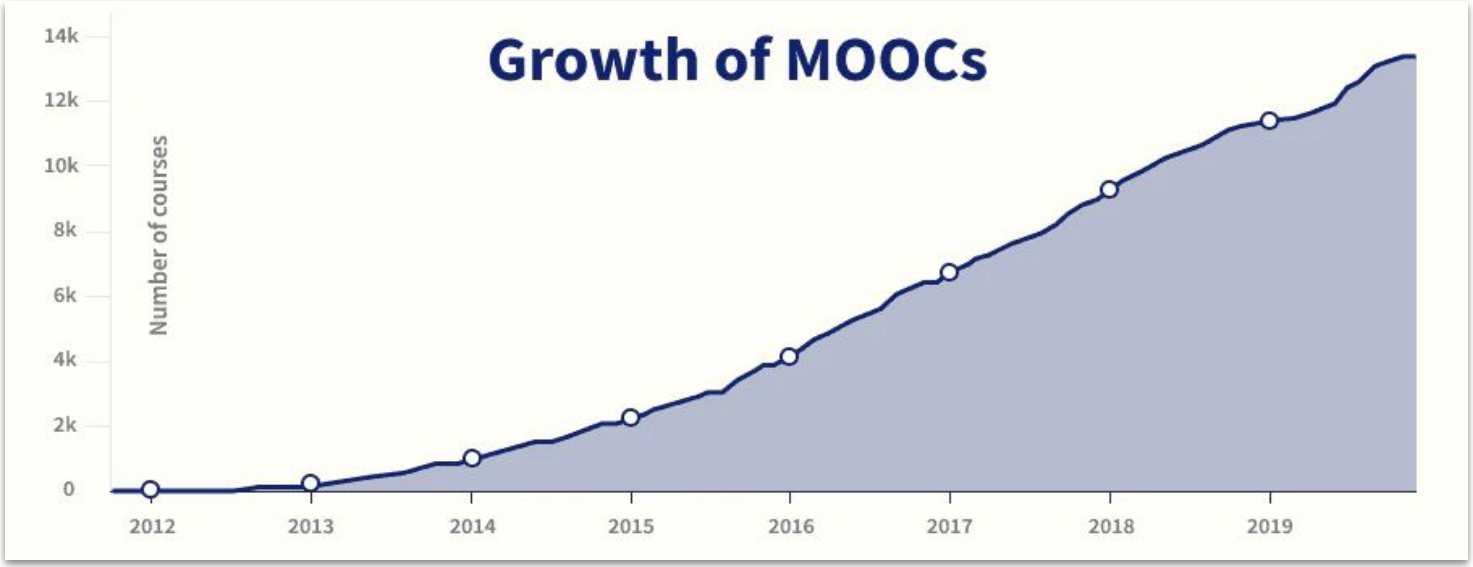
900+ Universities

13.5k Courses

820 Microcredentials

50 MOOC-based degrees

2019 The Year in Review

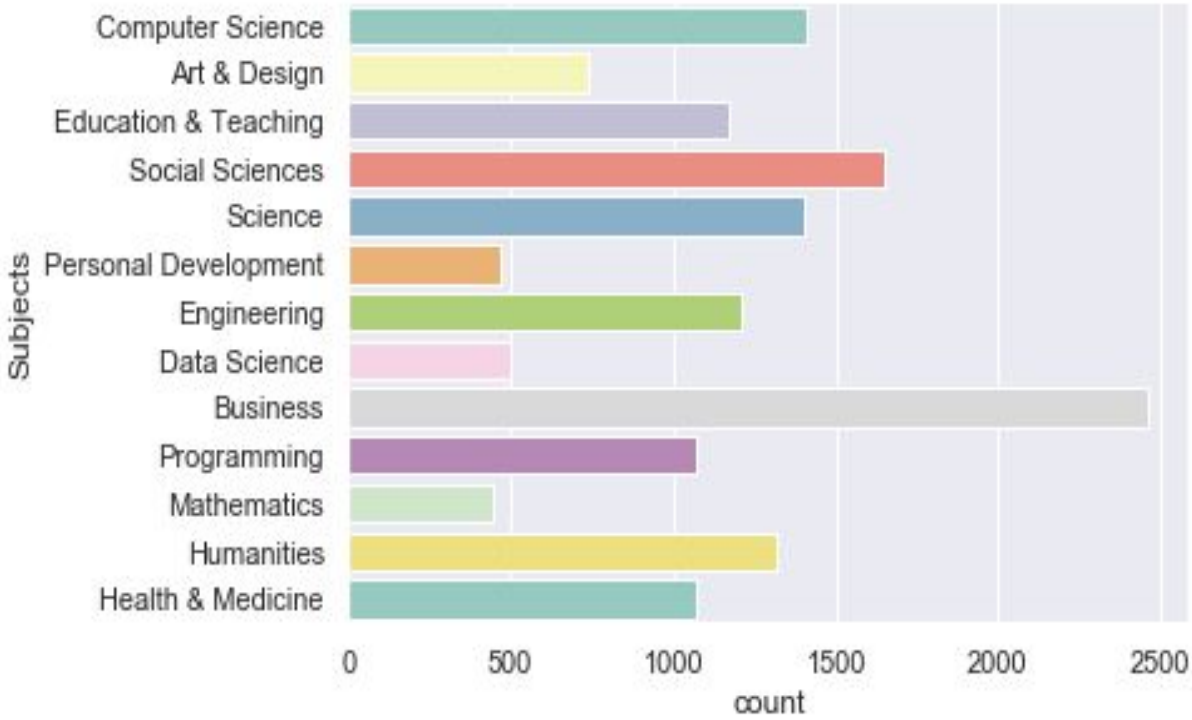


(ClassCentral.com)

	Learners	Courses Available	Microcredentials Offered	Master's Degrees Conferred
Coursera	45 million	3,800	420	16
edX	24 million	2,640	292	10
Udacity	11.5 million	200	40	1
FutureLearn	10 million	880	49	23
Swayam	10 million	1,000	0	0

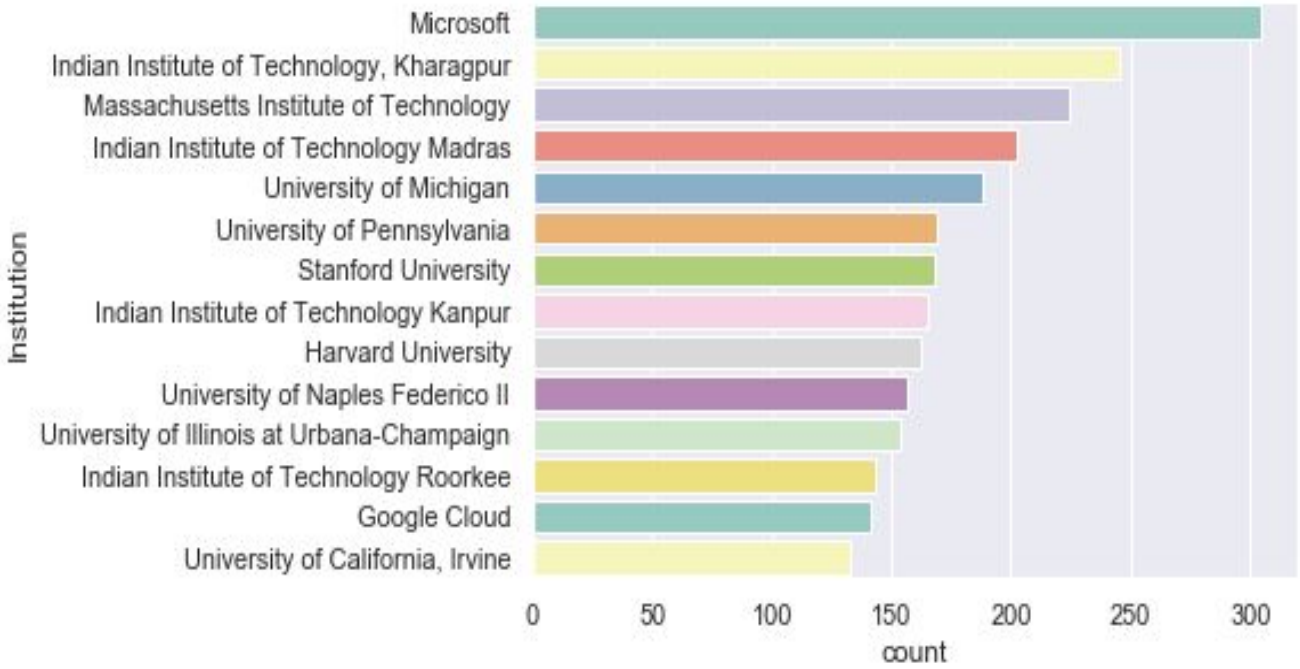
Classcentral.com/2019

Courses Categorical Variables EDA



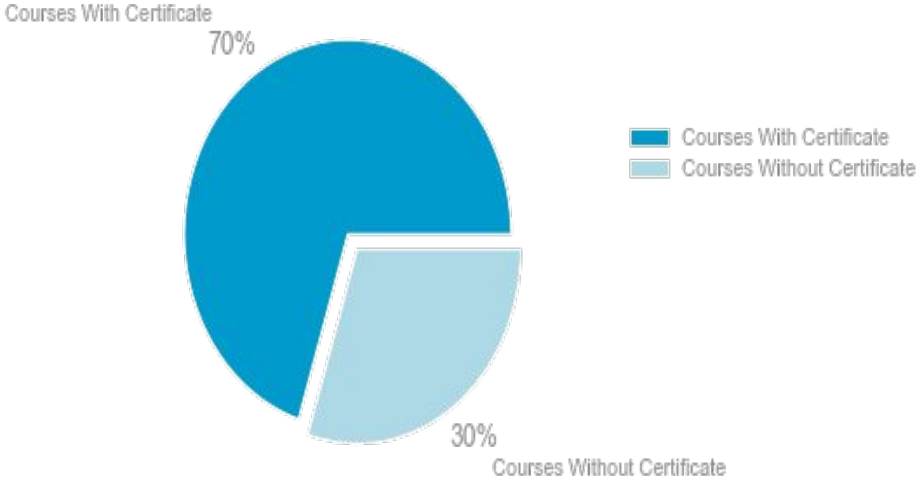
- **Count:** 14914
- **Unique:** 13
- **Top:** Business
- **Most Frequent Occurrence:** 2462

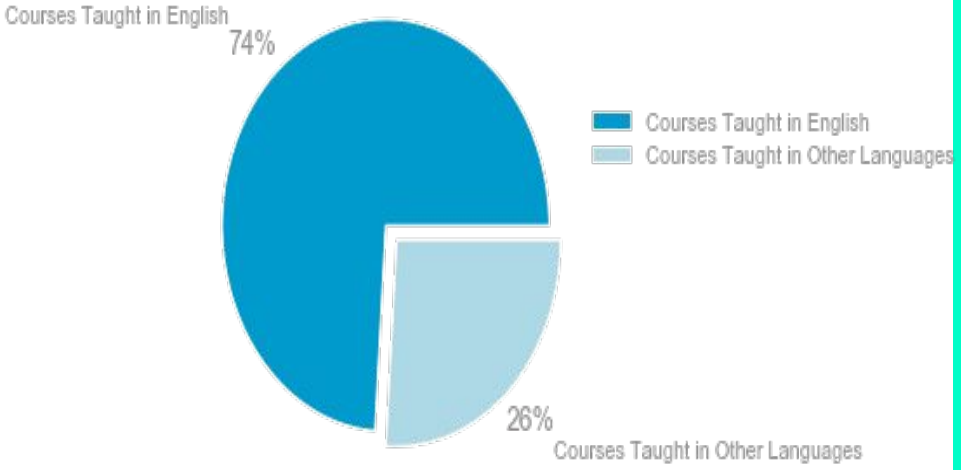
Bar graph with the different subjects



- **Count:** 14914
- **Unique:** 13
- **Top:** Microsoft
- **Most Frequent Occurrence:** 304

Bar graph with Different Institutions





With or Without Certificate: [10525:4485]

Languages: English or not English

Courses CONTINUOUS Variables Discussion

Duration: Online Flexibility

Consider physically as standalone entity:

- The playback time of the course
- Homework completion time
- Exam completion time

Distribution with bounds

Consider in reality as interactions with people:

- Interruptions of daily activities cause repeated or slower playback
- Assignments might take longer because of learning inconsistencies

Distribution without bounds

Duration: Feature Engineering

Physical: equations with slight randomness

- Hours per week based on Normal distribution bounded by scraped variable
- Selected Normal distribution because of the Central Limit Theorem
- Mean of the sample modelled by Normal distribution is the maximum likelihood estimate of the sample

Interaction with people:

- Use physical variable as an estimate for planning purposes
- Need to create a categorical variable called **Flexible_Deadlines**
- **Flexible_Deadlines** can be an additional categorical variable for courses and indicates which courses need additional data from source

Stars And Number of REviews: Ranking

Stars Ratings:

- Calculated by weighted average of the ratings for each star level

Possible Inconsistencies:

- Low number of reviews for unpopular courses
- New courses have yet to be reviewed
- Biased ratings

Possible Solutions:

- Rank the courses based on the “sort” option of available star ratings
- Gather more data from source websites

Subpopulation Case Study

Subpopulation Characteristics:

Useful characteristics:

- With Certificate
- Taught in English
- Recently started or starting soon with two weeks
- Top 50 ranked for each subject
- Without flexible deadline

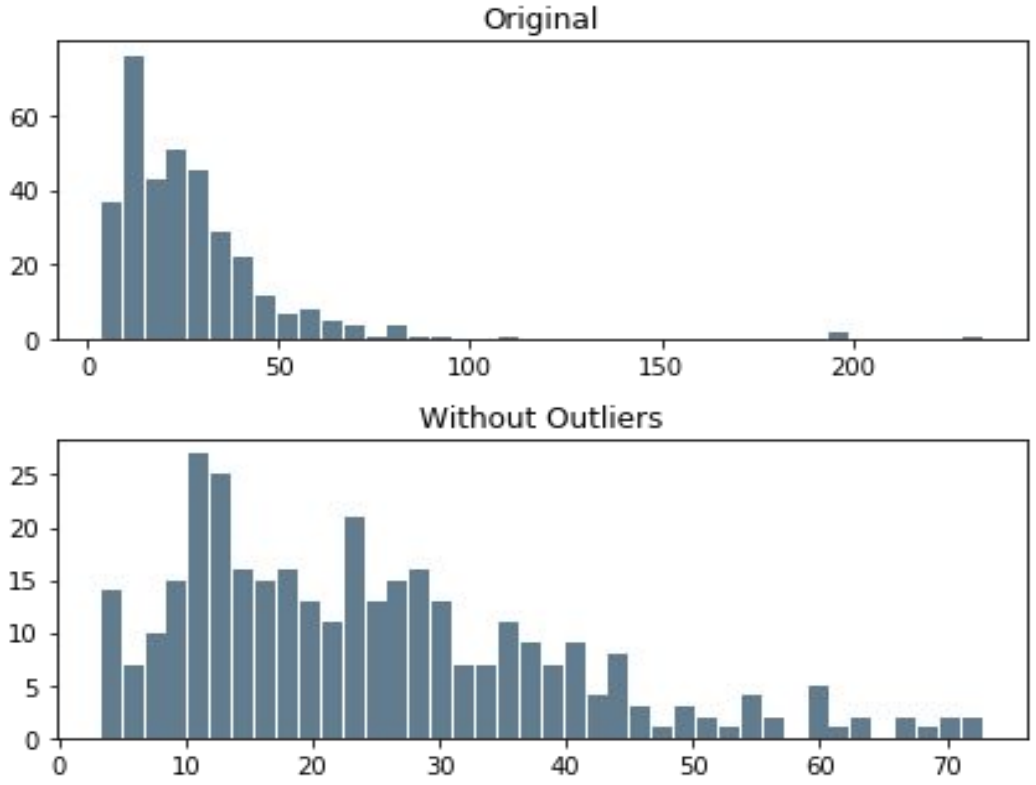
Questions to explore:

Based on these characteristics, how does course duration vary?

Duration: 'Course _Duration _Physical (hours)'

Summary Stats and Remove Outliers

mean	27.510394
std	24.129930
min	3.247697
25%	12.936404
50%	22.979072
75%	34.685296
max	234.184144



Future work:

- Identify more variables into features
- Scrape more data from MOOC source website
- Reviews