Class CENTRAL: Making INFORMED ONLINE LEARNING DECISIONS

Compiled by Kailun Cheng

TOPICS ExPLORED:

Background:

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

Objectives:

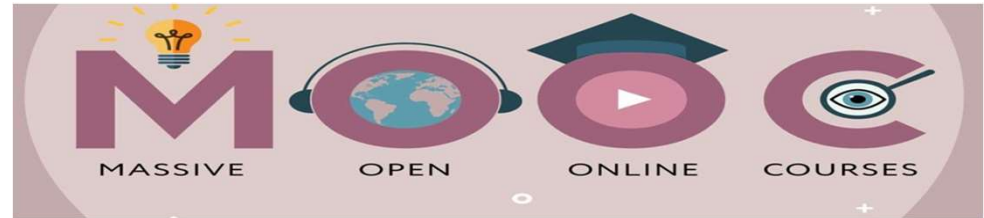- Student and Class Central Perspectives

Courses Categorical Variables EDA

Courses Numerical Variables Discussion:

- Duration Ambiguity: Feature Engineering
- Stars and Reviews: Feature Ranking
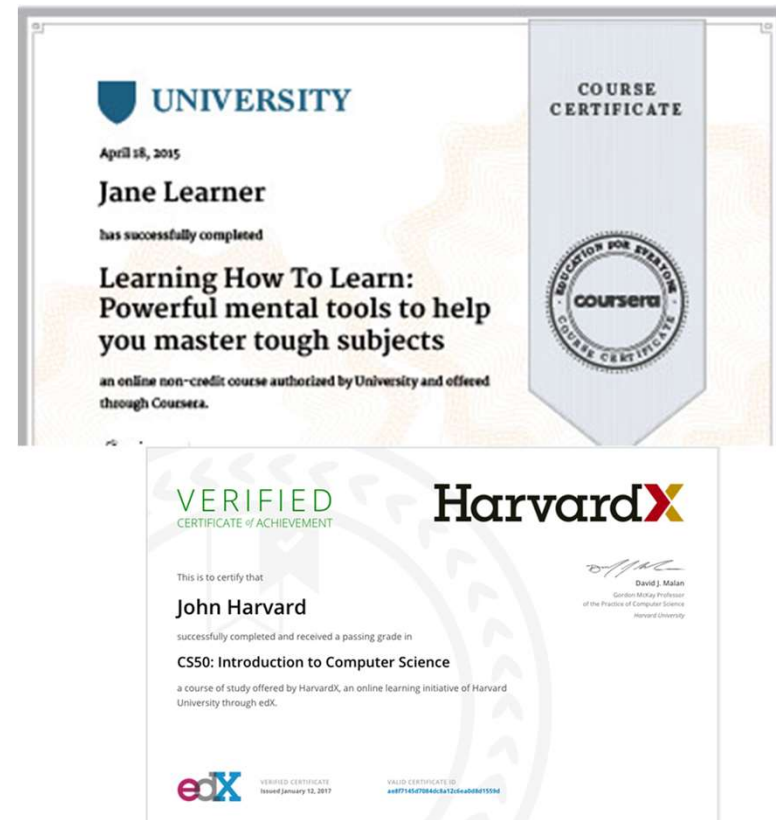
Subpopulation Case Study

- Course duration of top 50 English courses with certificates

Background:MooCs Growth and Current Statistics

MOOCs Advantages

- Feedback comes in the form of exams or projects that require the learner the complete to obtain a certification

- Characterized by flexibility

110M Students

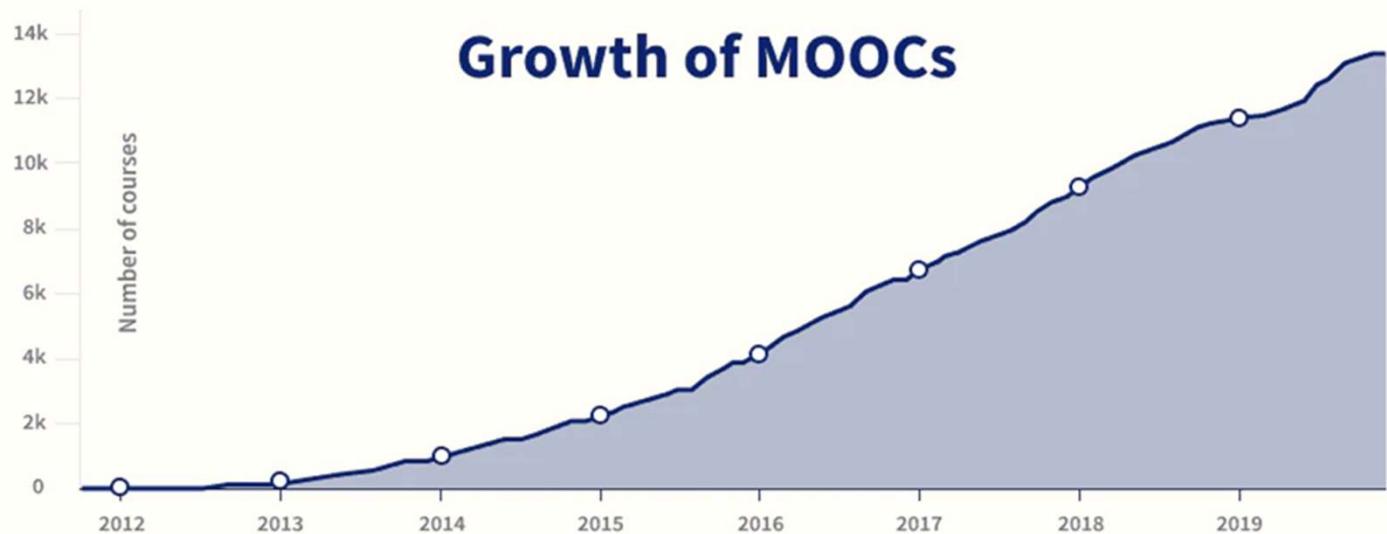**900+** Universities

13.5k Courses

820 Microcredentials

50 MOOC-based degrees

20 19 *The Year in Review*

Growth of MOOCs

(ClassCentral.com)

Number of courses

14k
12k
10k
8k
6k
4k
2k
0

2012 2013 2014 2015 2016 2017 2018 2019

	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

Objective:Course Planning

Course Planning:

Questions from the student's perspective:

How many courses to take at the same time?

- Do courses of various subjects/institutions/with, without certificates/languages differ in duration?

…Let's explore the variables
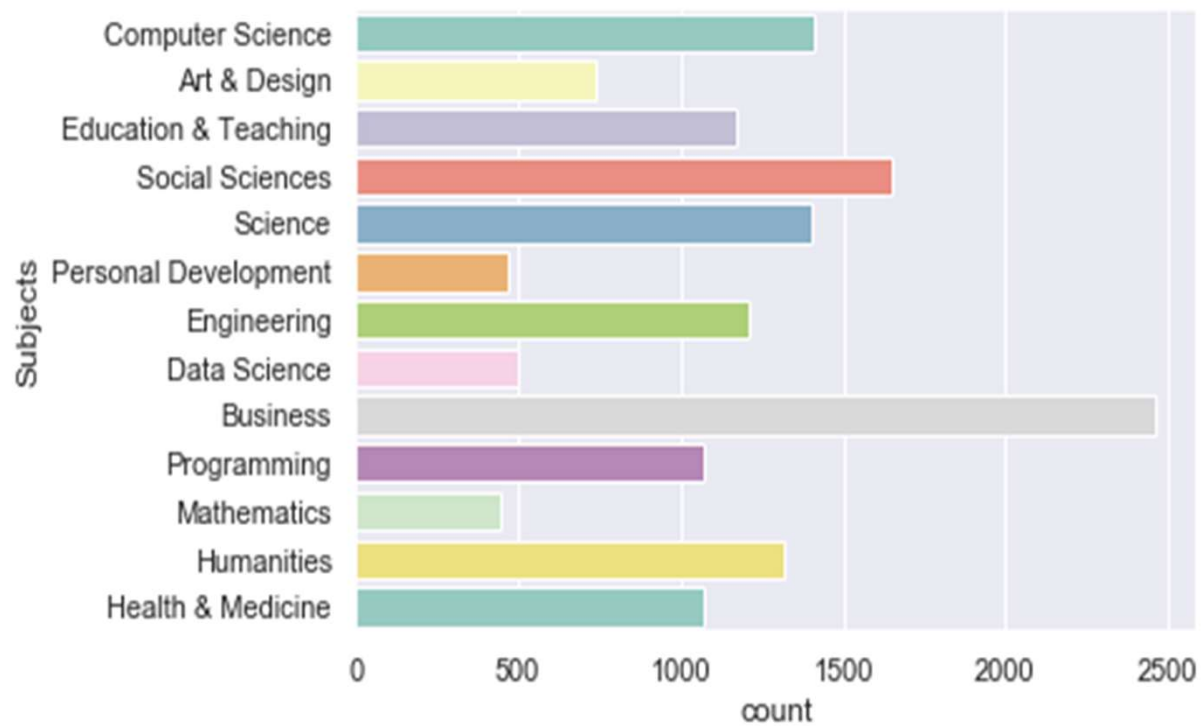
Questions from Class Central's purposes:

*Disclaimer: Class Central is learner-supported. When you buy through links on our site, we may earn an affiliate commission.

- How to get users to plan for as many courses as possible?
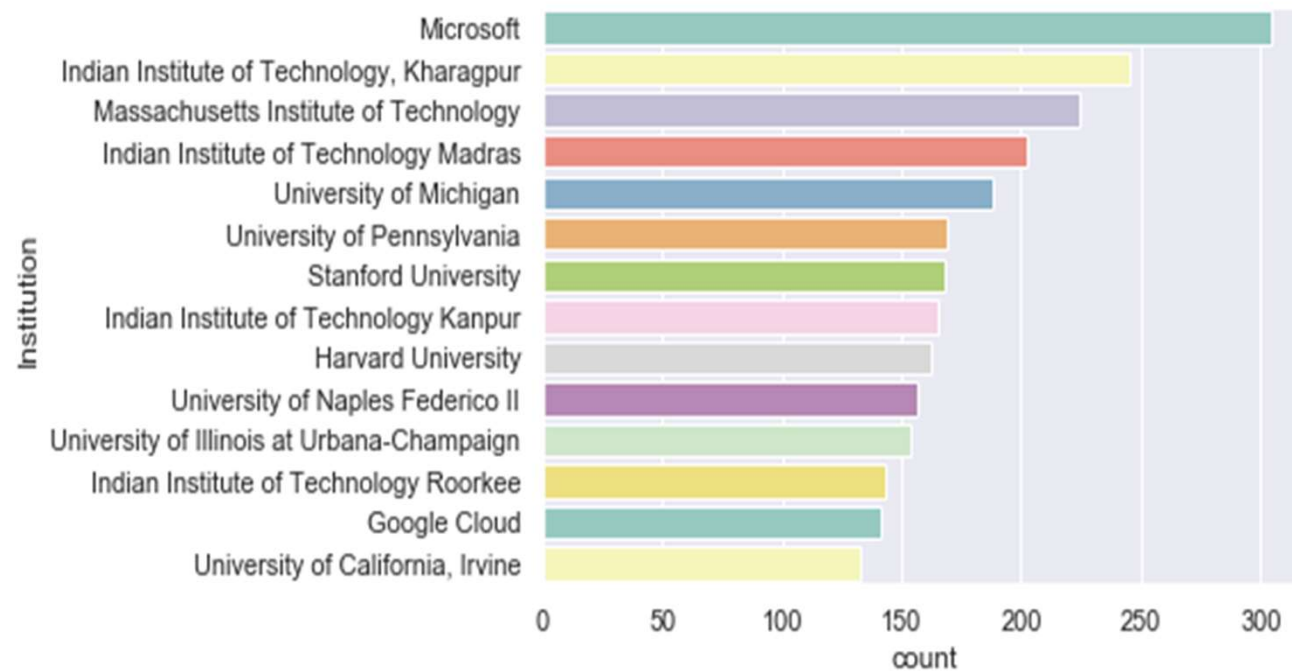
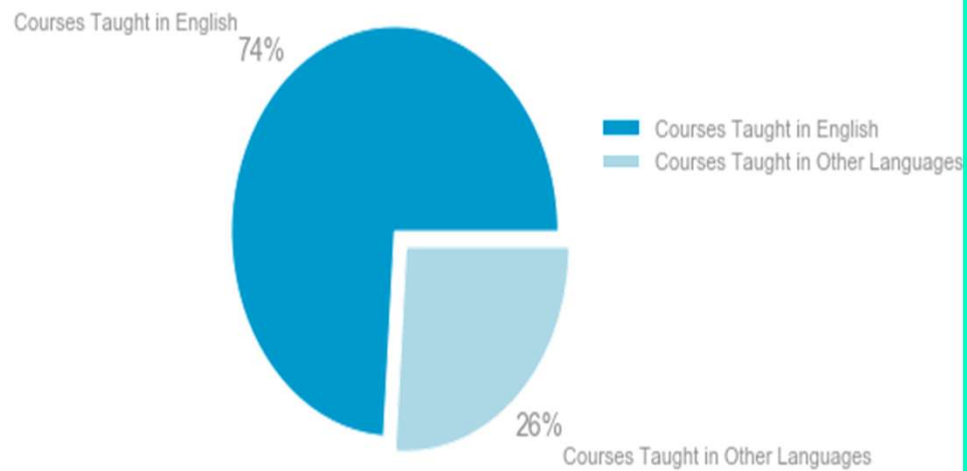# Courses Categorical Variables EDA:Subjects, Insitution, Certificate, Language
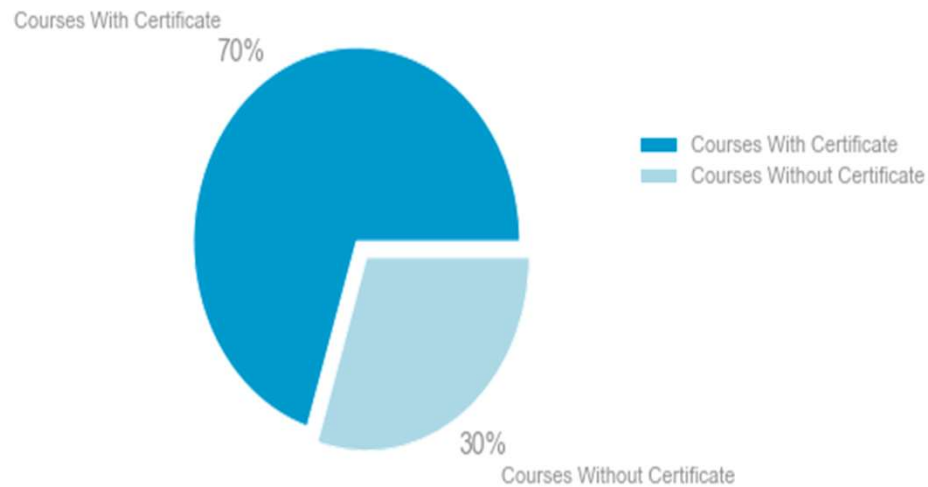
- **Count:** 14914
- **Unique:** 13
- **Top:** Business
- **Most Frequent Occurrence:** 2462

Bar graph with the different subjects

- **Count:** 14914
- **Unique:** Top 15
- **Top:** Microsoft
- **Most Frequent Occurrence:** 304

Bar graph with Different Institutions

Courses With Certificate
70%

Courses With Certificate
Courses Without Certificate

30%
Courses Without Certificate

Courses Taught in English
74%

Courses Taught in English
Courses Taught in Other Languages

26%
Courses Taught in Other Languages

With or Without
Certificate:

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
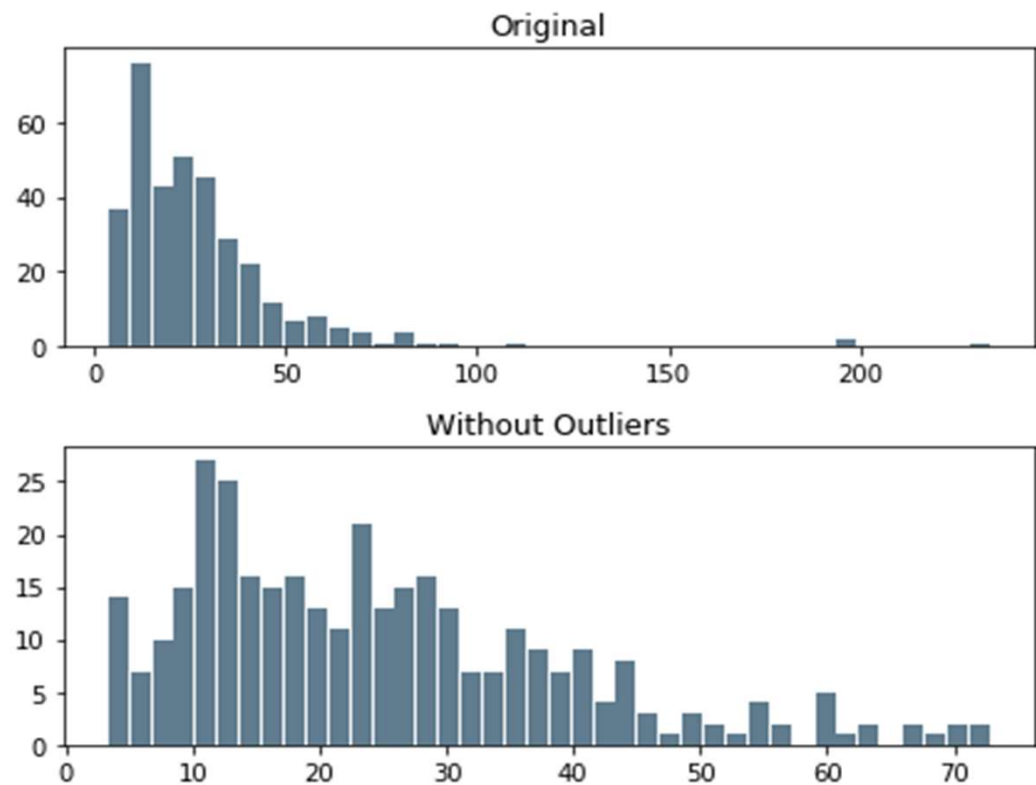- Without flexible deadline

Questions to explore:

Based on these characteristics, how do courses duration vary?

Duration: 'Course Duration Physical (hours)'

Summary Stats and Histogram with outliers removed

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