

# What makes students withdraw courses at Open University?

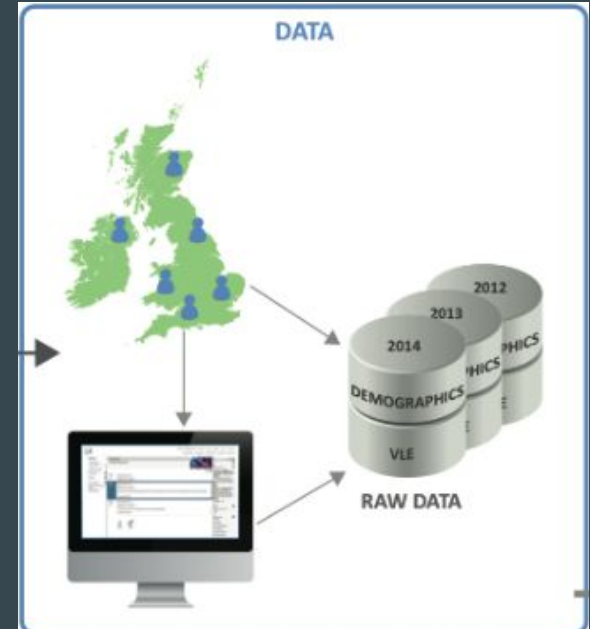
...

Liam Wang, April 8th, 2017

# Overview of dataset

Open University is an online education provider at England.

The dataset that we are using is the anonymised Open University Learning Analytics Dataset (OULAD). It contains data about courses, students and their interactions with Virtual Learning Environment (VLE) for seven selected courses.



# Current situation of Open University courses:

## High Withdrawal Rate

31.1%

of students

**WITHDREW**

from the course.

## High Failure Rate

21.6 %

of students

**FAILED**

the course.

## Low Passing Rate

47.3%

of students

**PASSED**

the course.

# Project objective:

To eventually improve the **retention** of students at Open University, by analyzing factors that are closely related to students performances (especially withdrawals).

---

# Some sample questions to be answered:

1. Do students' demographic information affect their performances?
2. Do passing students interact with material more than failing or withdrawal students?
3. Do courses with certain VLE (contents) attract students more, hence lower failure/withdrawal rate?

---

# Project Significance

By analyzing which factors have potential positive or negative relations with the students' performances, we are able to improve contents and structure of the courses, provide personalized recommendations for students, and offer extra study assistance if students are in need.



# Overview of Method

Exploratory Analysis with 1) statistical analysis, 2) classification model, and 3) visualization validation.

1. Feature Extraction with data preprocessing
2. Statistical Tests: correlation, ANOVA, etc.
3. Classification prediction models: Naive Bayes, Random Forest, etc.
4. Results summarized
5. Validation with visualizations

# Features extraction: Demographics, Assessments, Interaction and Contents

## Demographics:

- Age groups
- Deprivation Index Percentage
  - indication of the the deprivation level of student's location
- Education Level
- etc.

## Assessments:

- Mean Final Exam score
- Mean CMA score
  - CMA: Computer Marked Assessment
- Mean TMA score
  - TMA: Tutor Marked Assessment



# Features extraction: Demographics, Assessments, Interaction and Contents

## Interaction:

- Sum of clicks of student
- Average daily clicks of each student
- Number of days that student stays interacted with the course
- Days of early registration:
  - How early did the student register

## Course Contents:

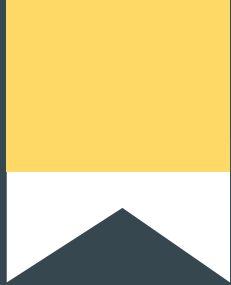
- Count of total web pages
- Count of total quizzes
- Count of forums
- Count of urls provided
- Count of resources provided
- etc.

# Results Summarized:



After series of tests and data analysis, most significant factors were summarized by counting the times they appear in top 10 features of each method's findings. Such voting mechanism finally leads to TEN most important features:

- Education level
- Average Daily Clicks
- Sum of Clicks
- Days Interacted
- Count of Subpages
- Count of Quizzes
- Count of URLs
- Count of forums



# Validation with visualizations

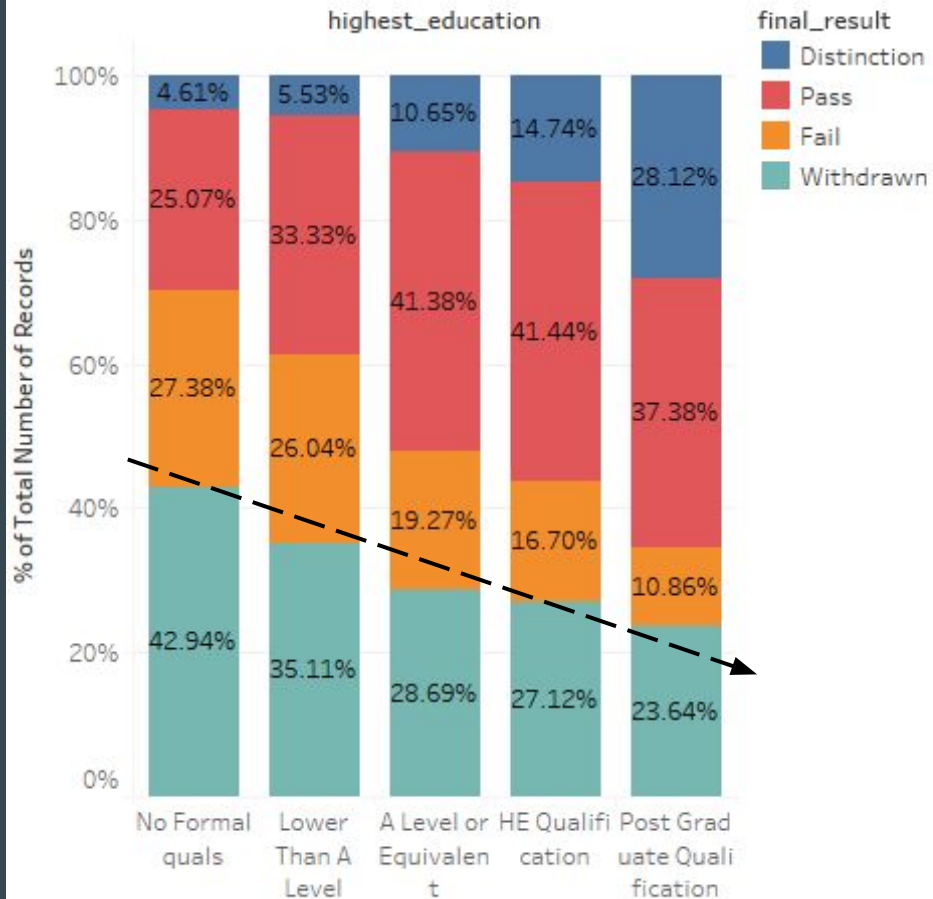
With visualizations generated by Tableau software, most findings were confirmed, with some additional findings updated.

# Demographic finding 1:

Students with higher education level perform better than less educated students.

The withdrawal rate is the shown in light blue color in the bottom block of each column bar. It decreases as education level goes up to the right.

Higher education -> lower withdraw rate

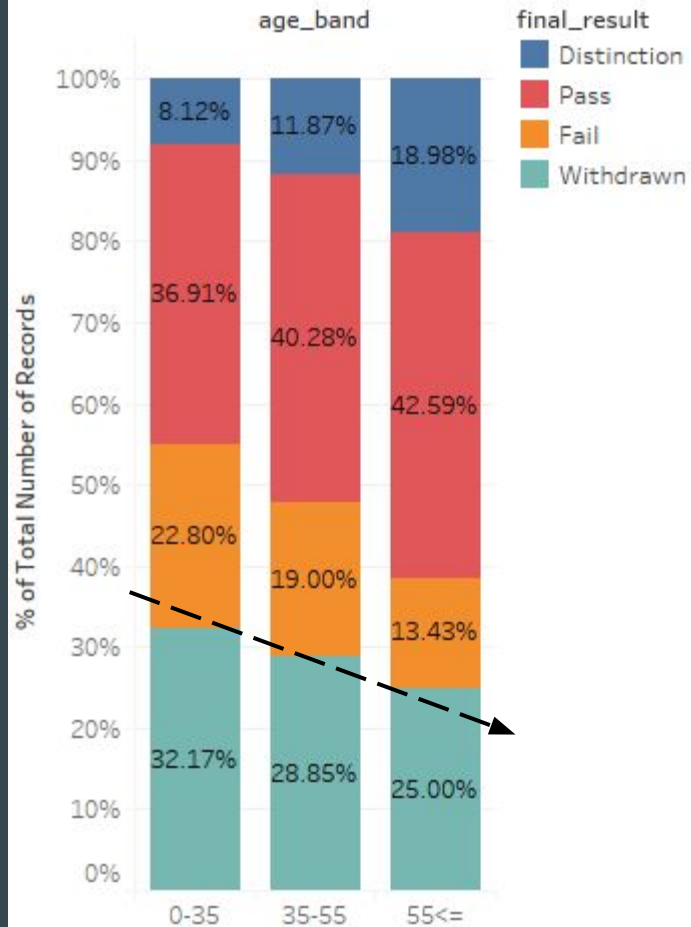


## Demographic finding 2:

Older students generally have less withdrawal rate than younger generation students.

The withdrawal rate is shown in light blue color in the bottom block of each column bar. It decreases as age in groups goes up to the right.

Young students more likely to withdraw

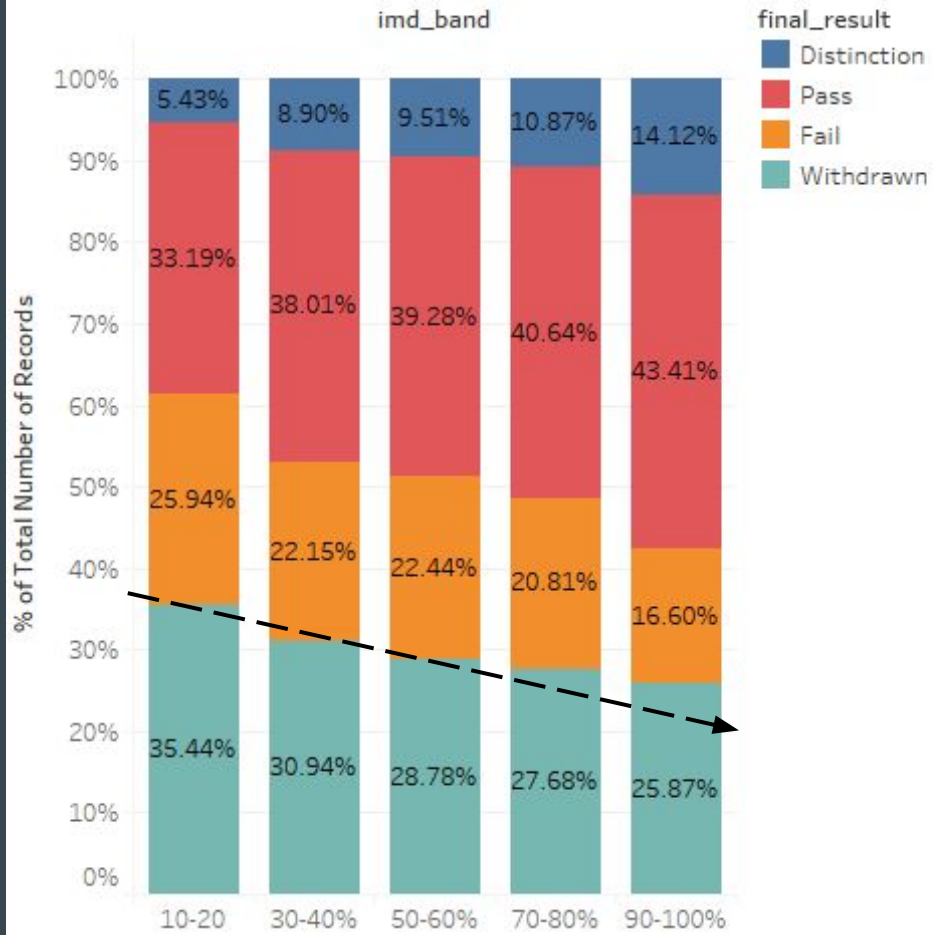


## Demographic finding 3:

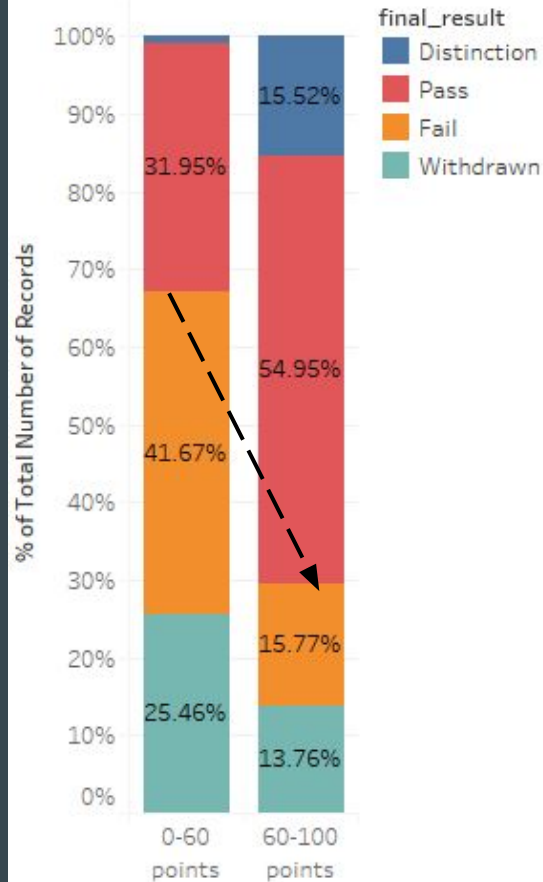
Students who come from less deprived areas have less withdrawal rate than more deprived areas.

The withdrawal rate is the shown in light blue color in the bottom block of each column bar. It decreases with the percentile of deprivation index increases. The higher the percentile, the wealthier the area is.

Students from less deprived areas have lower withdraw rate



TMA average  
score: 0-60 vs.  
60-100

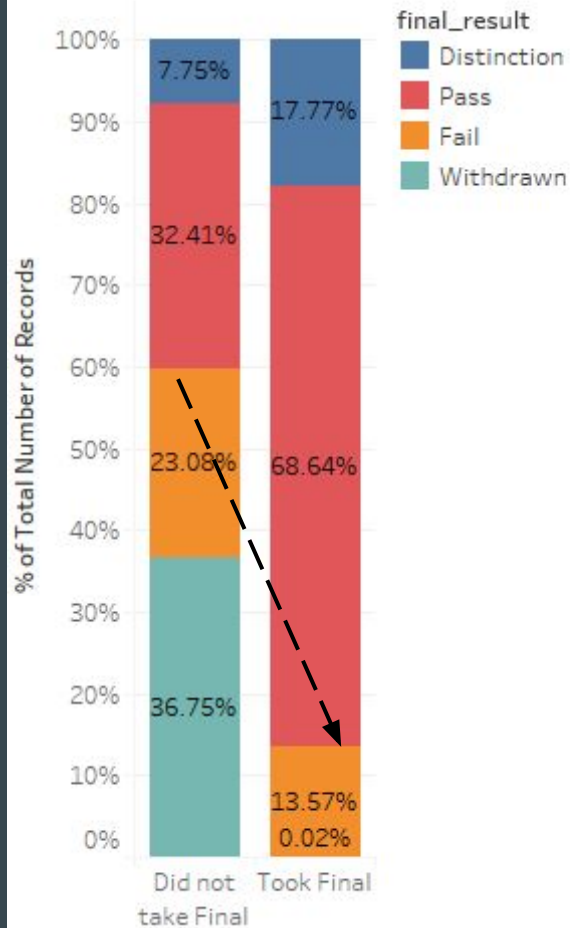


## Assessment finding 1:

**70% of the students who averages at least 60 points in Tutor Marked Assessments(TMA) have passed the course, comparing to 33% in those averages less than 60.**

The passing rate is shown as the sum of top two blocks of each column bar, in blue and red. The blue means distinction. The red means pass. Significant differences are shown here.

## Final Exam: Not Take vs Take



## Assessment finding 2:

**87% of those who took the final exam passed, while only 40% of the students passed without taking final exam.**

The passing rate is shown as the sum of top two blocks of each column bar, in blue and red. The blue means distinction. The red means pass. Significant differences are shown here.

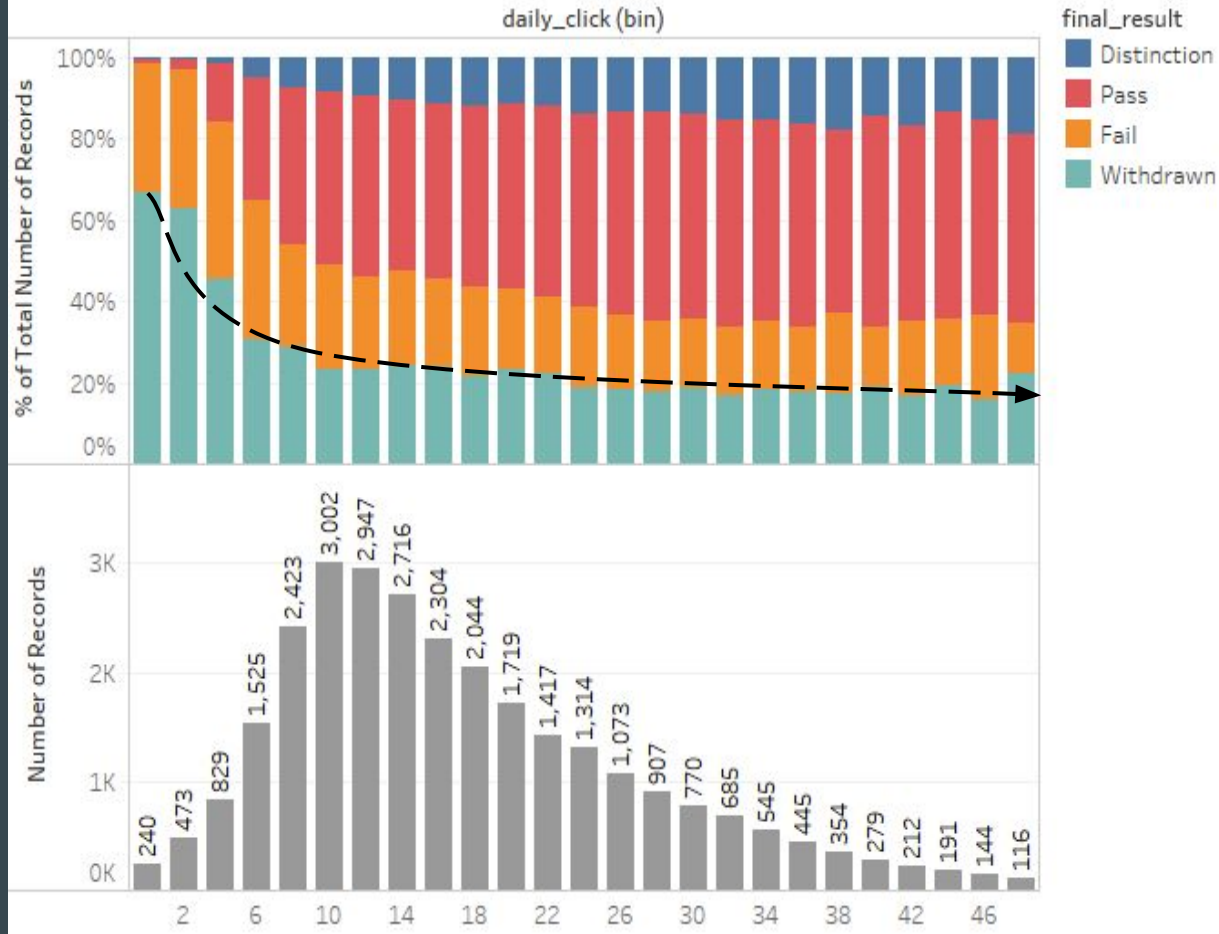


# Interaction finding 1:

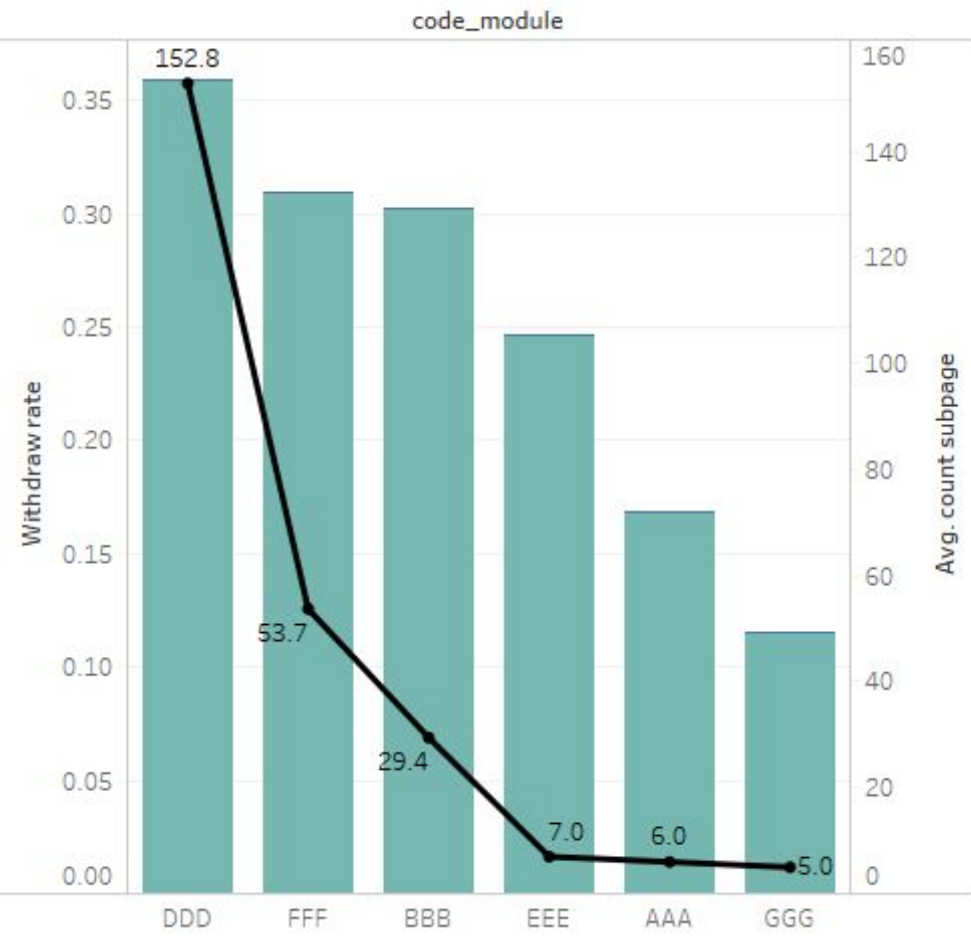
Students with more daily clicks have less withdrawal rates than less interactive students.

The withdrawal rate is the shown in light blue color in the bottom block of each bar. It decreases as daily average number of clicks goes up to the right. The bars on the bottom is the whole distribution of daily clicks.

The more clicks student have each day, the less likely they withdraw



Courses with more subpages has higher withdraw rates



## Course Contents finding 1:

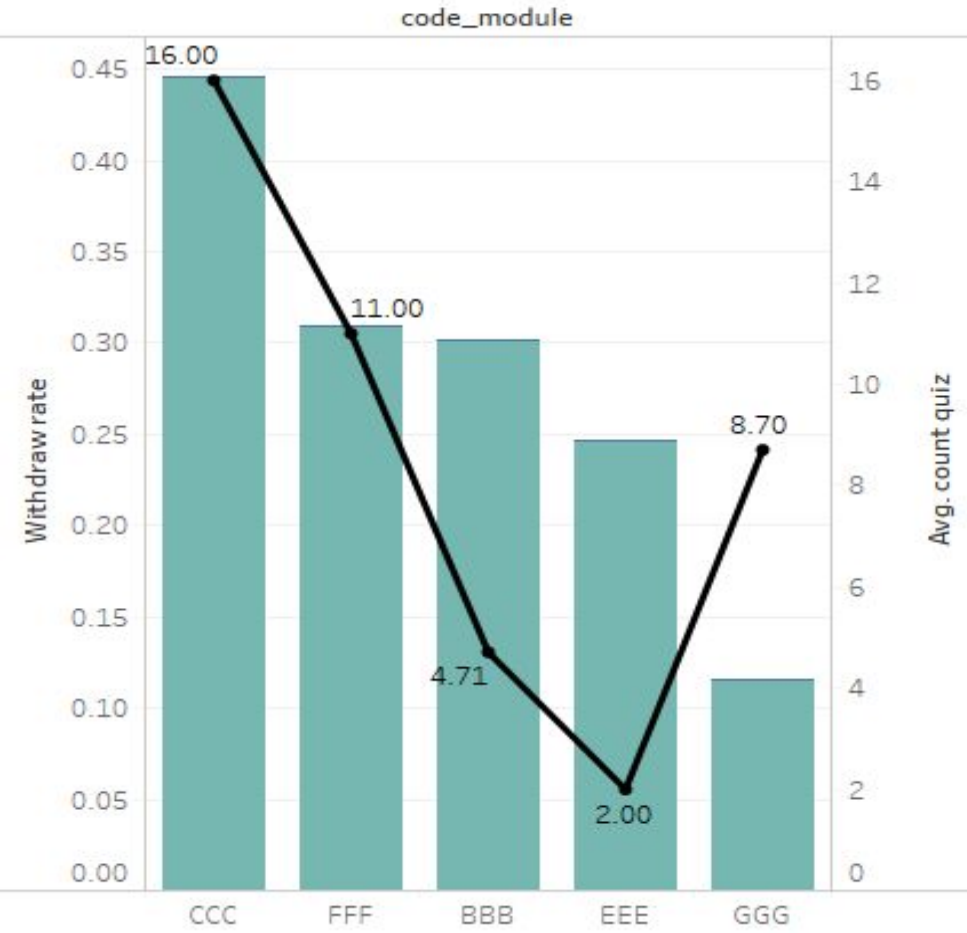
Courses with lower withdrawal rates have less subpages.

The bars in light blue are average withdrawal rates of each course.

The black line connects average numbers of subpages each course contains.

It can be seen that they have similar trends.

Courses with more quizzes have higher withdraw rates



## Course Contents finding 2:

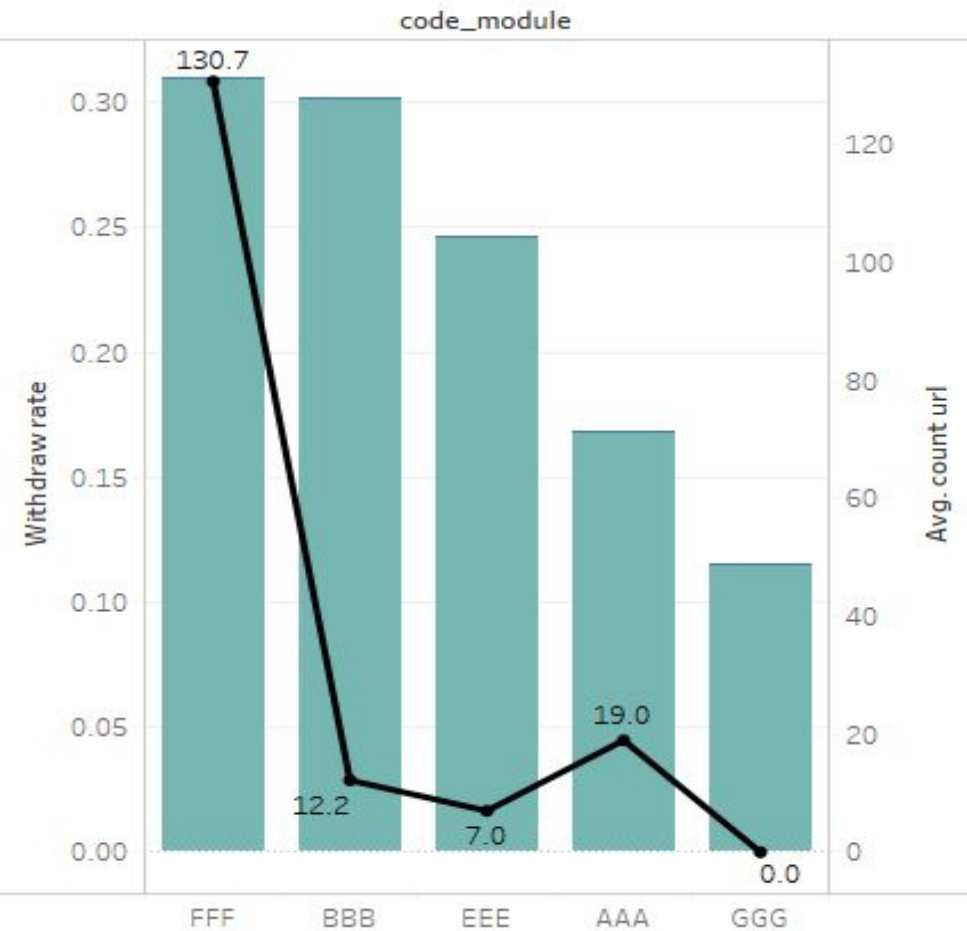
Courses with lower withdrawal rates have less quizzes.

The bars in light blue are average withdrawal rates of each course.

The black line connects average numbers of quizzes each course contains.

It can be seen that they have similar trends.

Courses which provides more urls have higher withdraw rates



## Course Contents finding 3:

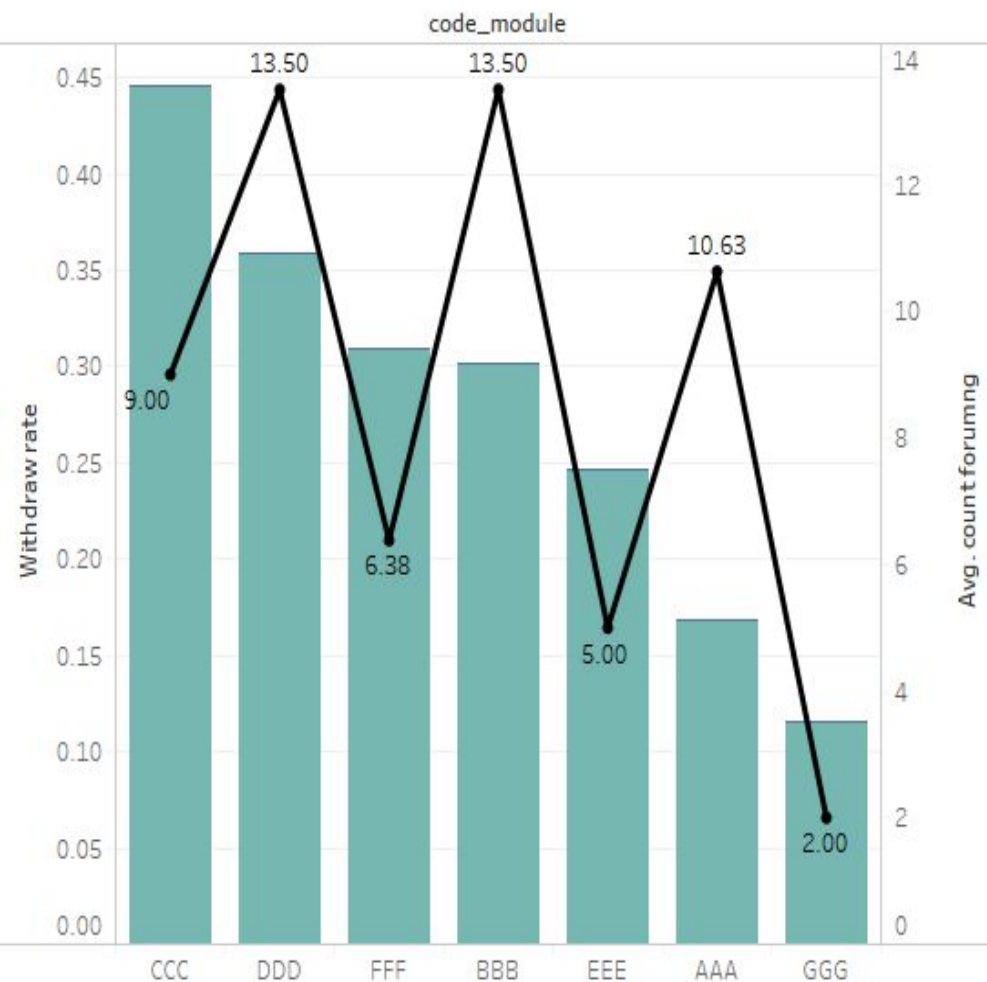
Courses with lower withdrawal rates have less urls.

The bars in light blue are average withdrawal rates of each course.

The black line connects average numbers of urls each course contains.

It can be seen that they have similar trends.

Courses with more forumng has higher withdraw rates



## Course Contents finding 4:

**Courses with lower withdrawal rates have less forums.**

The bars in light blue are average withdrawal rates of each course.

The black line connects average numbers of forums each course contains.

It can be seen that they have similar trends.

# Comparison of Good Courses and Less Good Courses

Firstly, 7 modules are separated into 2 groups according to their average withdrawal rates. As we can see from the table on the right, left is good courses and right is bad courses.

Clearly bad courses have higher number in counts of subpages, quizzes, urls, and forums.

Good courses vs. Less Good courses

|                    | Courses with lower<br>withdraw rates | Courses with higher<br>withdraw rates |
|--------------------|--------------------------------------|---------------------------------------|
| Avg. count subpage | 6.0                                  | 102.4                                 |
| Avg. count quiz    | 0.0                                  | 6.6                                   |
| Avg. count url     | 19.0                                 | 38.7                                  |
| Avg. count forumng | 10.6                                 | 11.7                                  |

| final_result | Courses with lower<br>withdraw rates | Courses with higher<br>withdraw rates |
|--------------|--------------------------------------|---------------------------------------|
| Distinction  | 5.88%                                | 8.23%                                 |
| Pass         | 65.11%                               | 31.82%                                |
| Fail         | 12.17%                               | 20.48%                                |
| Withdrawn    | 16.84%                               | 39.46%                                |

# Suggested Improvement plans based on findings:

- **Customer Strategies:**
  - Customized recommendations for different groups of students according levels of their education, age, and locations
  - Email notifications of study assistance when people struggle with class
  - Encourage students to take the final exam to increase the chance of passing
- **Course contents design:**
  - Make the course structure to be simpler and easier
  - Course pages to contain less subpages, quizzes, urls, forums

# Issues and suggestions for future research

1. Assessment results lack of data.

Something worth noticing is that over 50% people withdraw before they even start to take tests. Therefore, in reality, it is hard to predict behavior based on test results. Half of students have taken CMA, and most have taken TMA at least once, but only small percent took the final.

2. It would be interesting to investigate on repetitive registration and how students behave when retaking courses.
3. When segmenting students, different groups of students might have different preferences over different kinds of course structures - in terms of contents and materials. It is worth some deeper research.





**Thank you!**