SAMSUNG

Together for Tomorrow!
Enabling People
Education for Future Generations

OULAD Learning Path Recommender System

Customizing User Experience

SIC 6 | Team 6

The Learning Process is no longer linear and " One-Size Fits All "

Meet The Team

Supervised By

Eng. Haneen

Yousef Albasel

Computer Science Student
Cairo University

Ali Amr

Data Science Student
Alexandria University

Mahmoud Salama

Data Science Student
Alexandria University

Table of contents

Problem Statement

Students enrolled in online distance education have diverse learning needs. Many online platforms doesn't address to these individual differences

Level Differences	Students are often left to navigate through content that is either too advanced or too basic
Learning Pace	Some students move quickly through course material, while others need more time
Retention Rates	Negative impact on retention rates and performance
Competition	The education technology space is becoming increasingly competitive, with many platforms offering similar courses.

Goal

Our goal is to understand each student's learning preferences and performance levels to tailor learning path based on these metrics

01

Improve learning outcomes

02

Increase retention rates

03

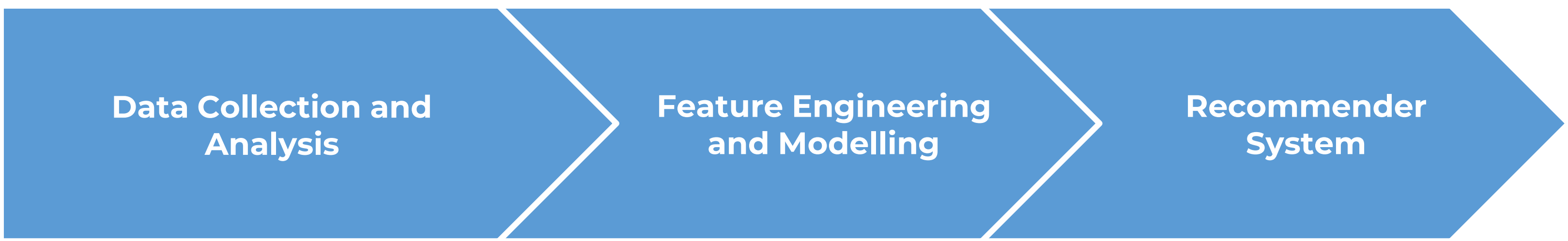
Enhancing user experience

04

Boosting business profitability

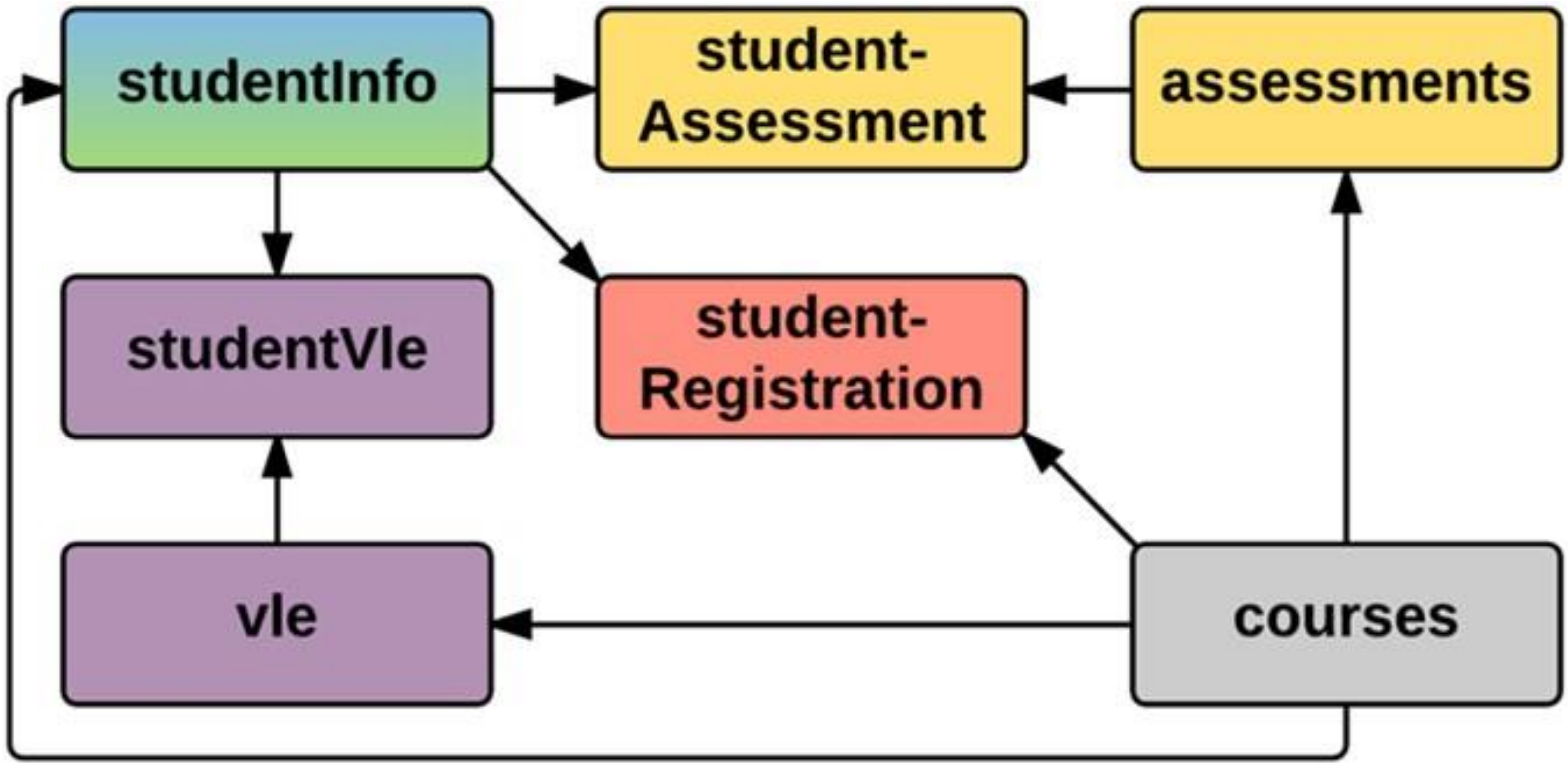
Solution

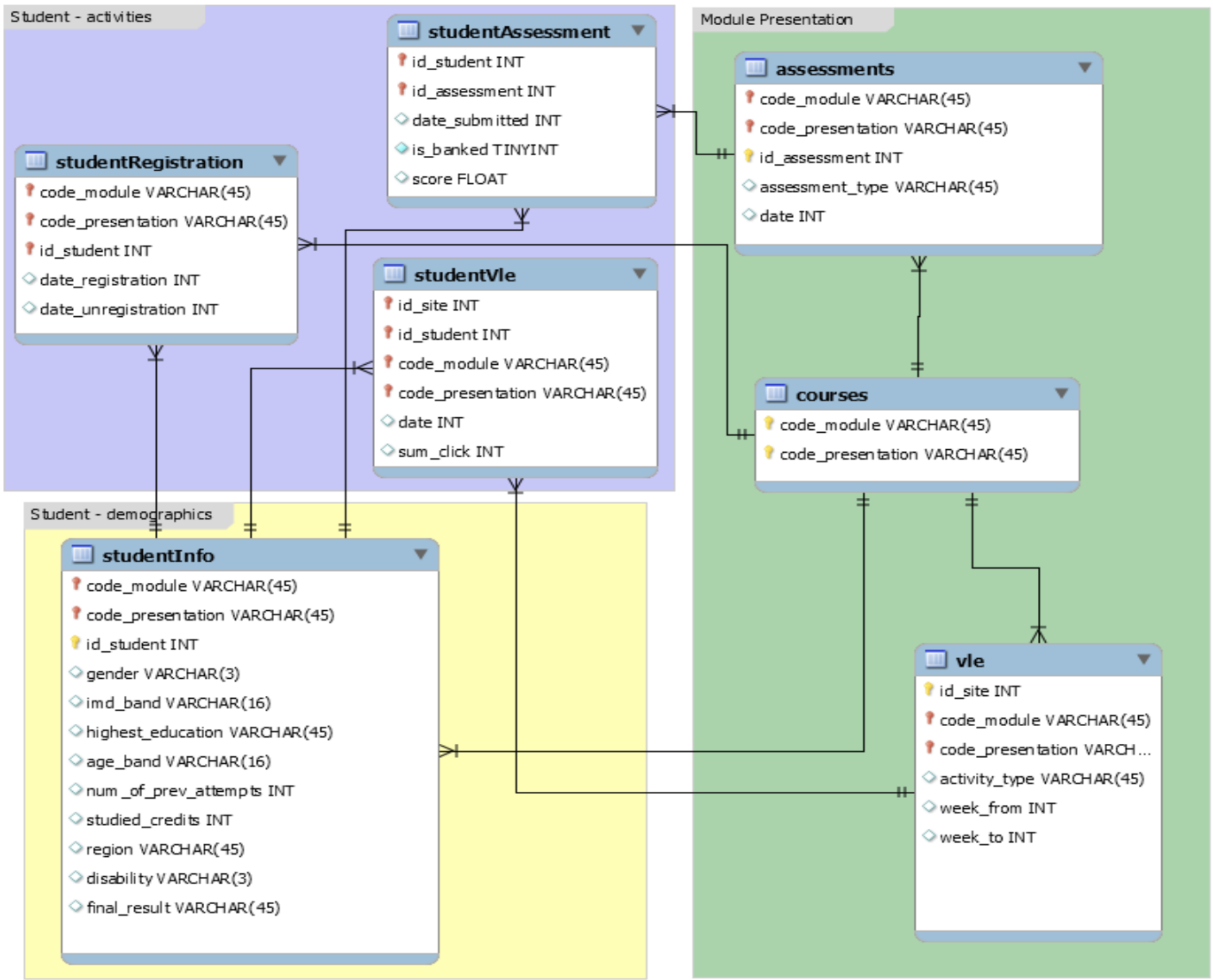
Develop a recommender system that evaluates student engagement and performance metrics to make recommendations on learning paths and material.



Dataset Description

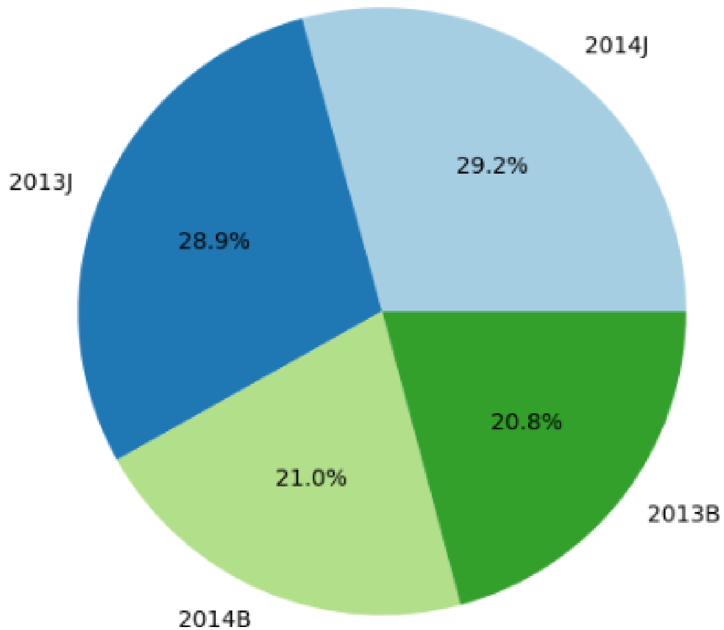
The Open University, a major British university with the largest undergraduate student population in the UK, provides the dataset through its Online Learning Platform (VLE). This platform supports off-campus students by offering access to course materials, forums, assignments, assessments, and grades.


studentInfo
student-
Assessment
assessments
studentVle
student-
Registration
vle
courses


Student - activities
studentAssessment
id_student INT
id_assessment INT
date_submitted INT
is_banked TINYINT
score FLOAT
studentRegistration
code_module VARCHAR(45)
code_presentation VARCHAR(45)
id_student INT
date_registration INT
date_unregistration INT
studentVle
id_site INT
id_student INT
code_module VARCHAR(45)
code_presentation VARCHAR(45)
date INT
sum_click INT
Module Presentation
assessments
code_module VARCHAR(45)
code_presentation VARCHAR(45)
id_assessment INT
assessment_type VARCHAR(45)
date INT
courses
code_module VARCHAR(45)
code_presentation VARCHAR(45)
Student - demographics
studentInfo
code_module VARCHAR(45)
code_presentation VARCHAR(45)
id_student INT
gender VARCHAR(3)
imd_band VARCHAR(16)
highest_education VARCHAR(45)
age_band VARCHAR(16)
num_of_prev_attempts INT
studied_credits INT
region VARCHAR(45)
disability VARCHAR(3)
final_result VARCHAR(45)
vle
id_site INT
code_module VARCHAR(45)
code_presentation VARCH...
activity_type VARCHAR(45)
week_from INT
week_to INT

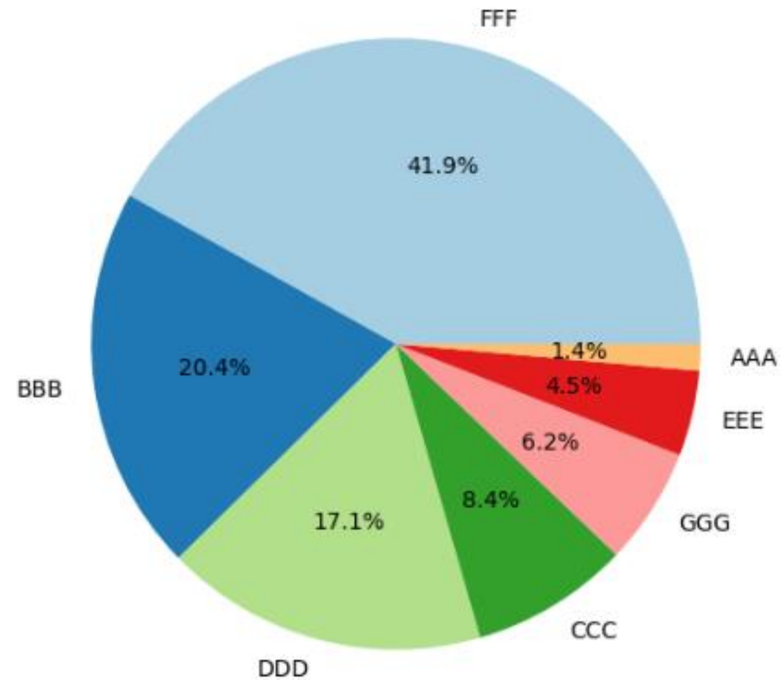
Code Presentation

Identification code of the presentation.
February(B) and October(J)



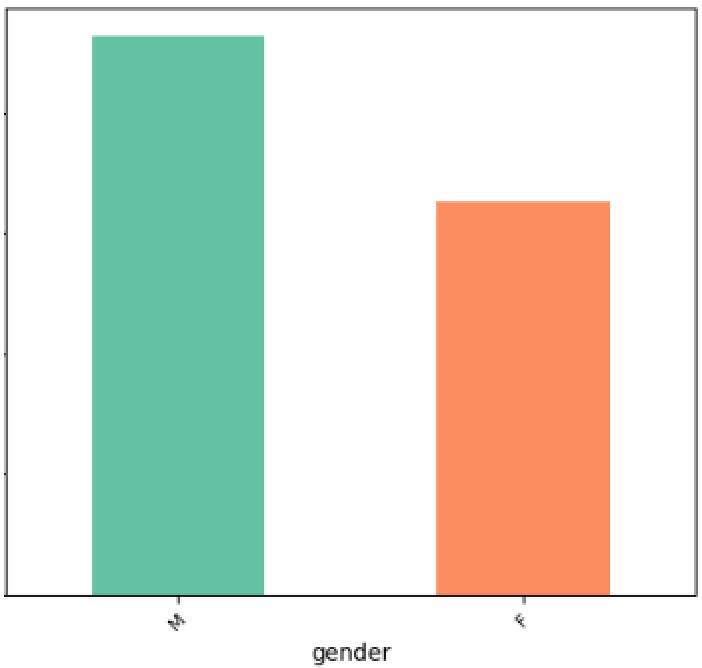
Code Module

identification code of the module



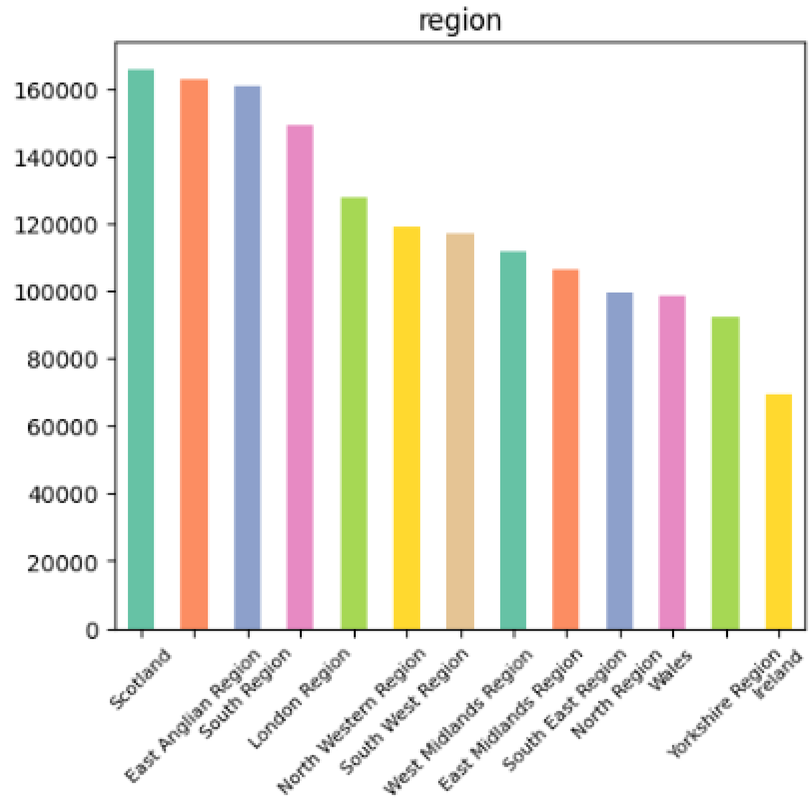
Gender

Student's Gender



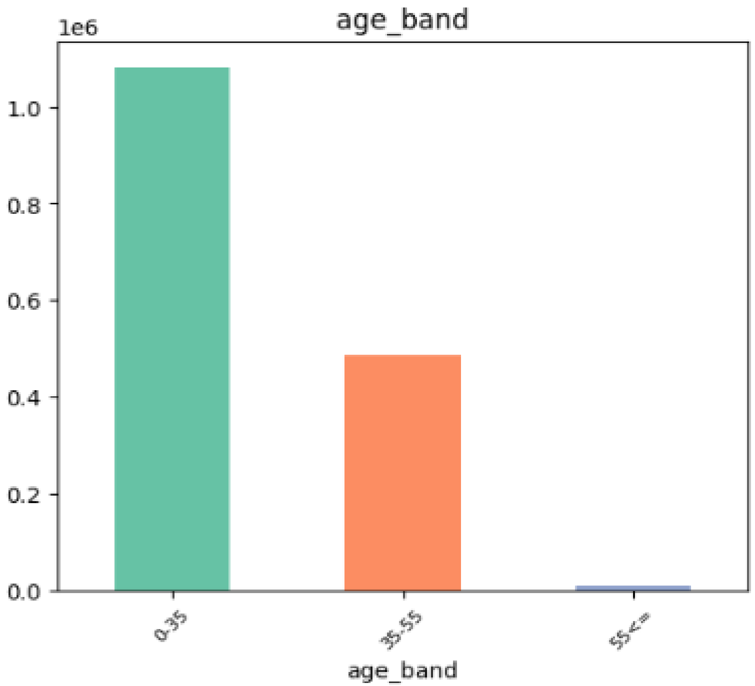
Region

Geographic region, where the student lived while taking the module-presentation.



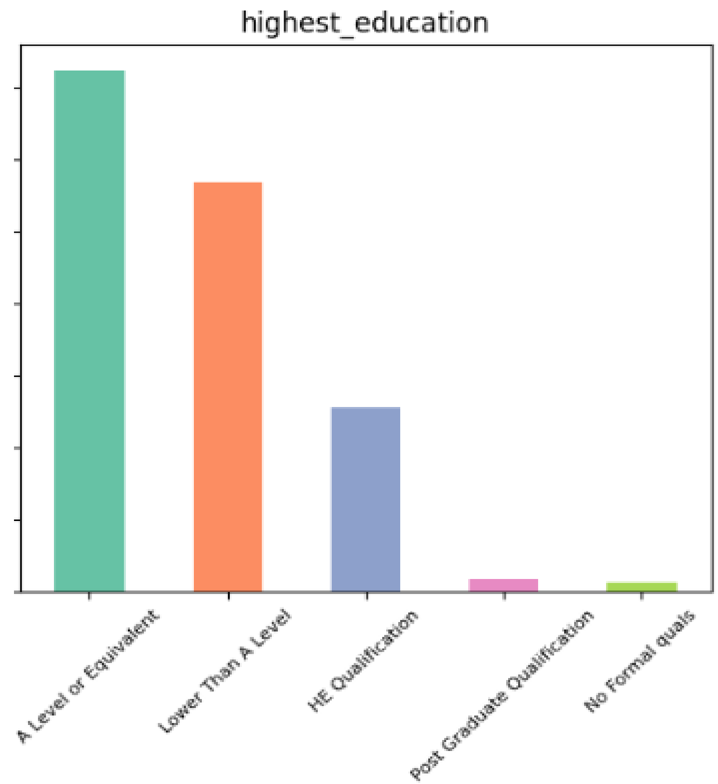
Age

band of the student's age.



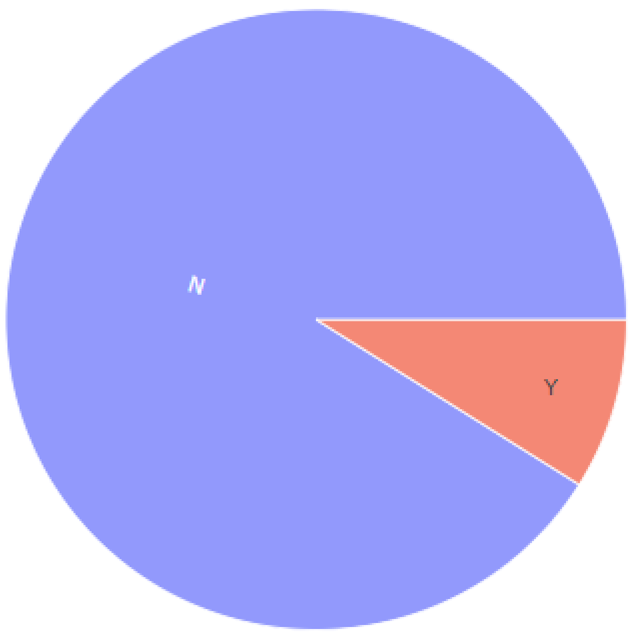
Highest Education

highest student education level on entry to the module presentation.



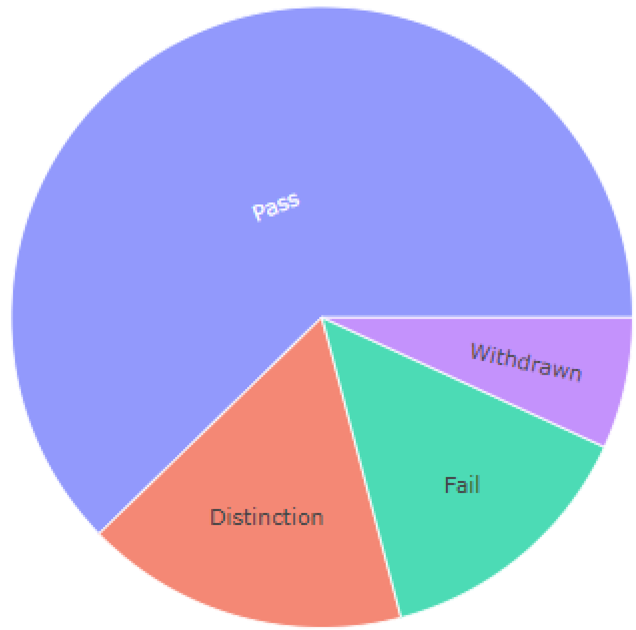
Disability

If a student is disabled



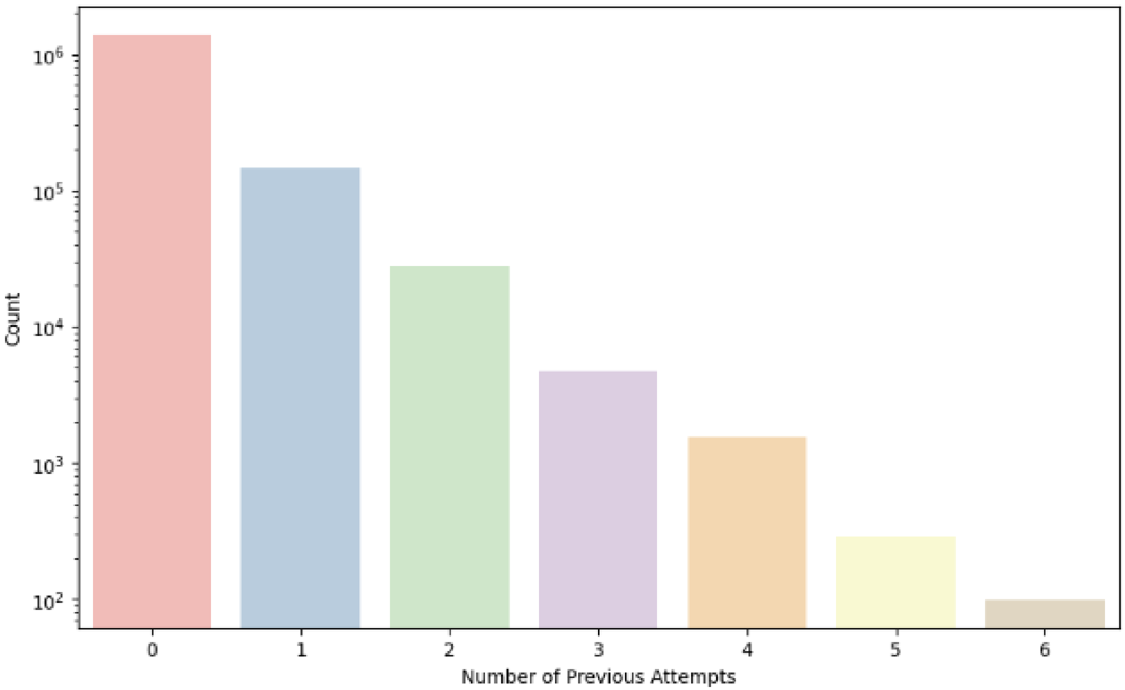
Final Result

student's final result in the module-presentation.



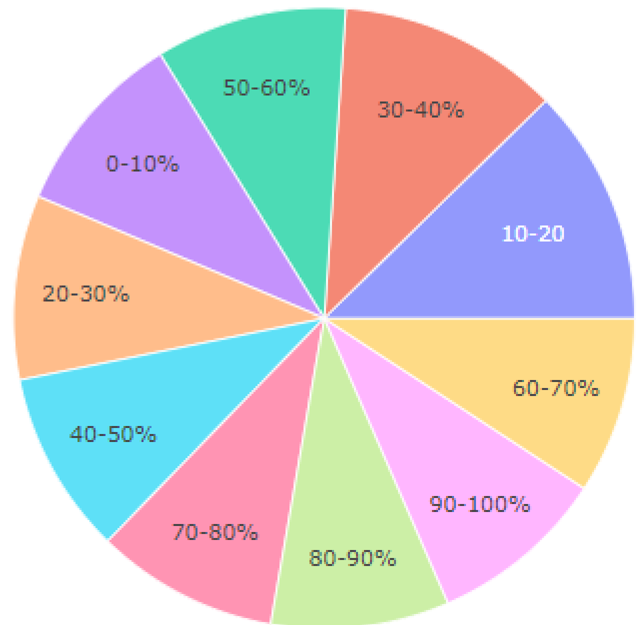
Number of Previous Attempts

the number times the student has attempted this module.



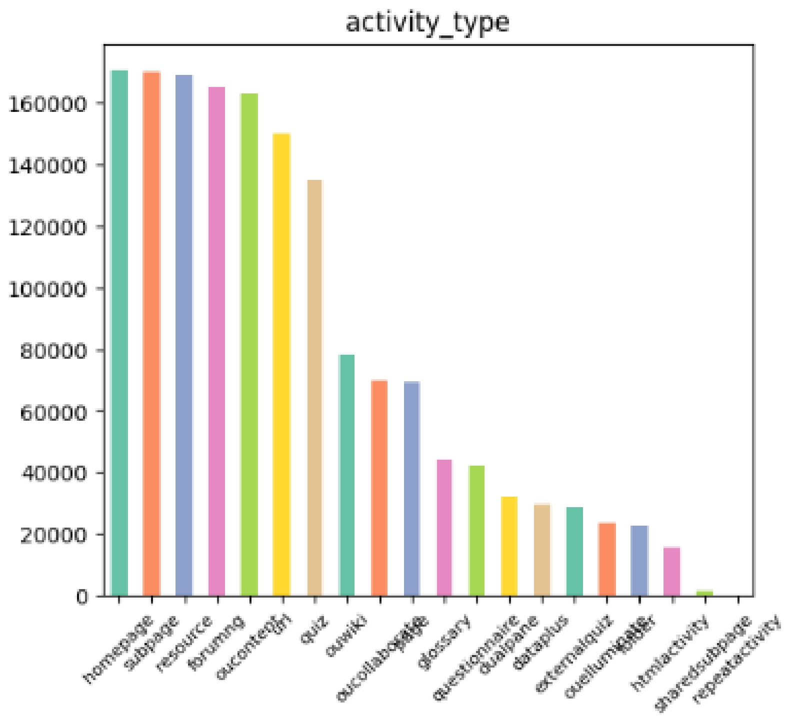
IMD Band

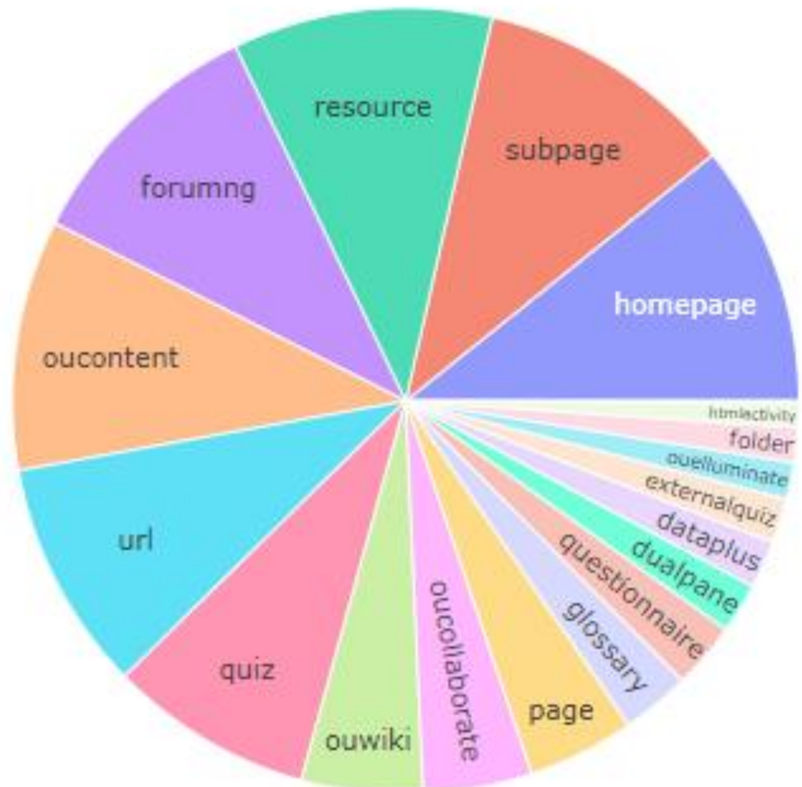
specifies the Index of Multiple Depravation band of the place where the student lived during the module-presentation.



Activity Type

the role associated with the module material.



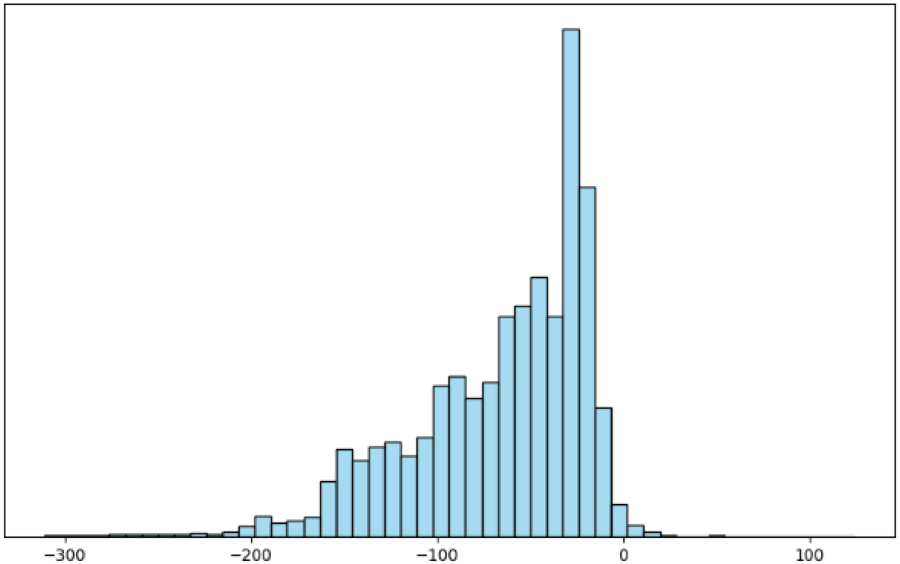


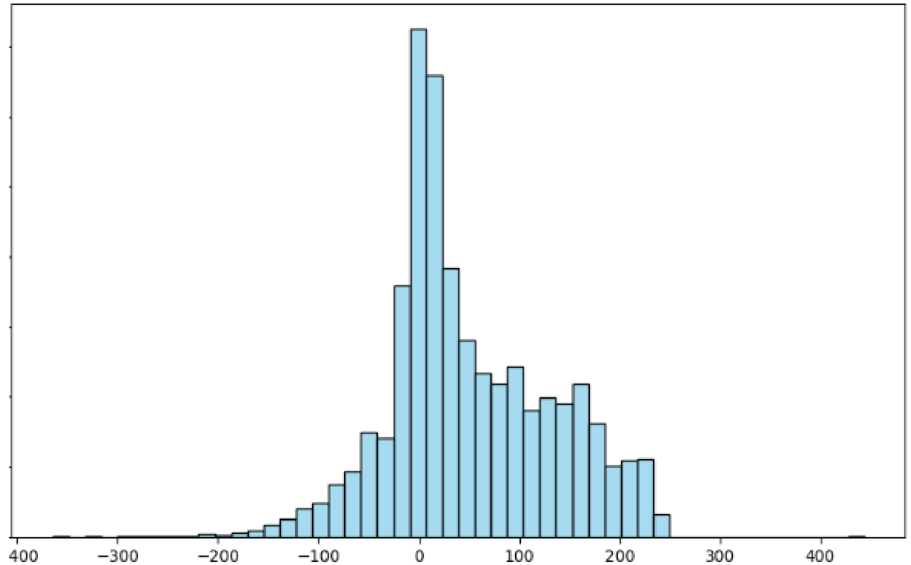
Date Registration

Date of student's registration on the module presentation, measured relative to the start of the module-presentation

Date Unregistration

Date of student unregistration from the module presentation, measured relative to the start of the module-presentation.



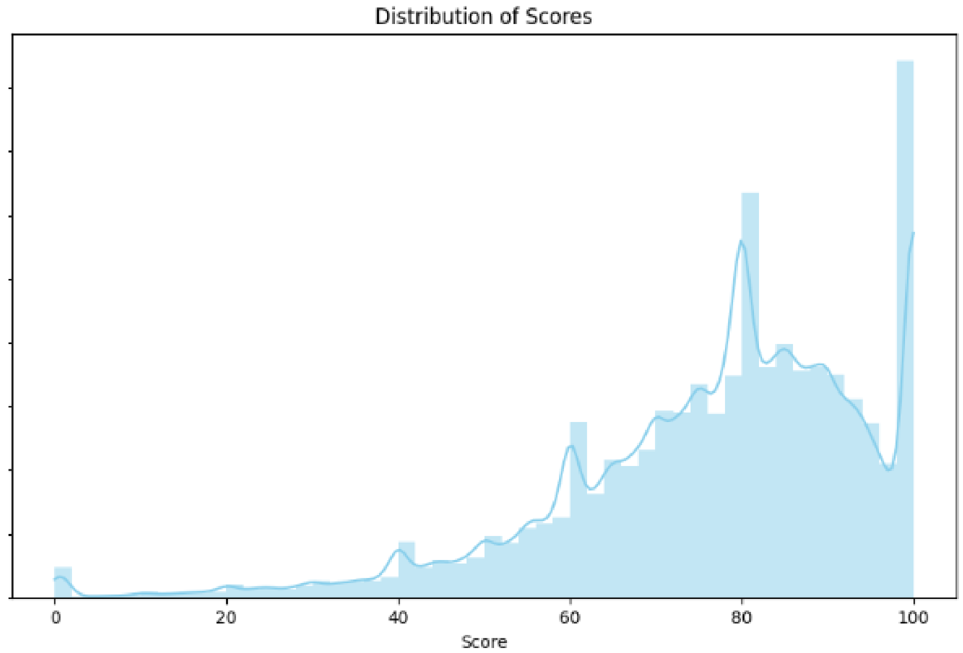


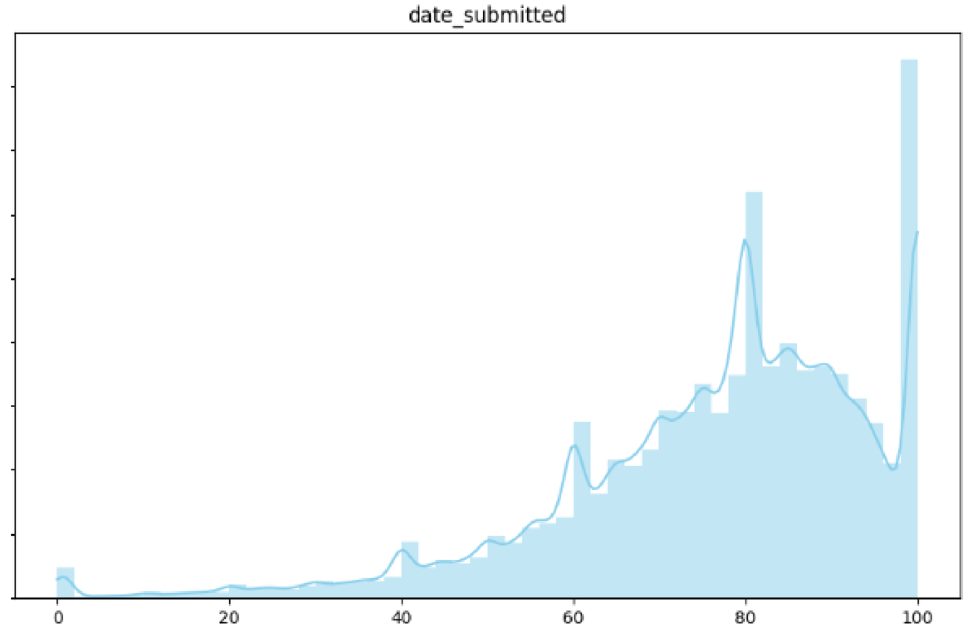
Score

The student's score in this assessment. The range is from 0 to 100. The score lower than 40 is interpreted as Fail.

Date Submitted

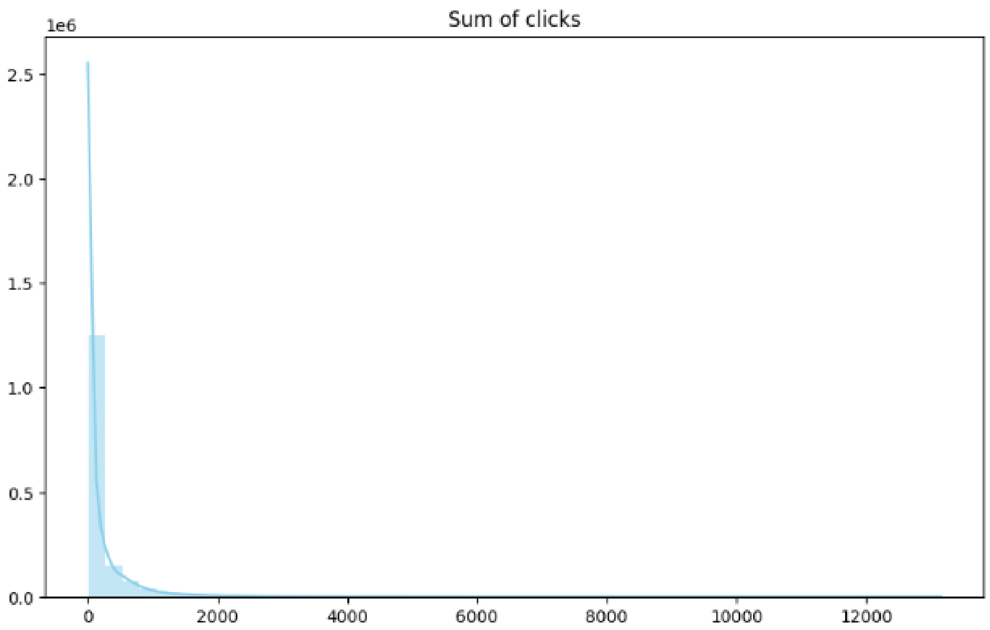
The date of student submission, measured as the number of days since the start of the module presentation.





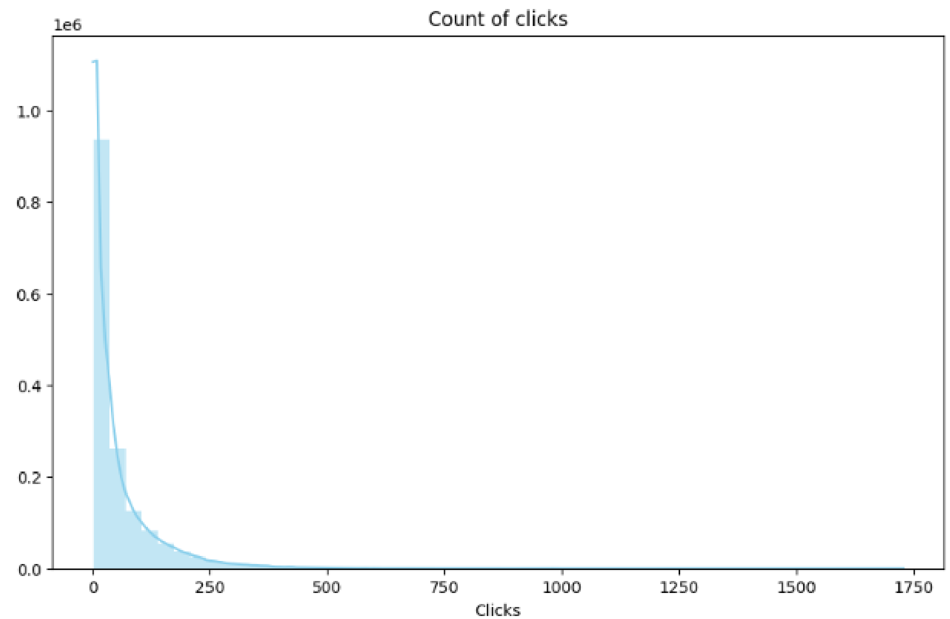
Sum of Clicks

Total number of clicks each student clicked



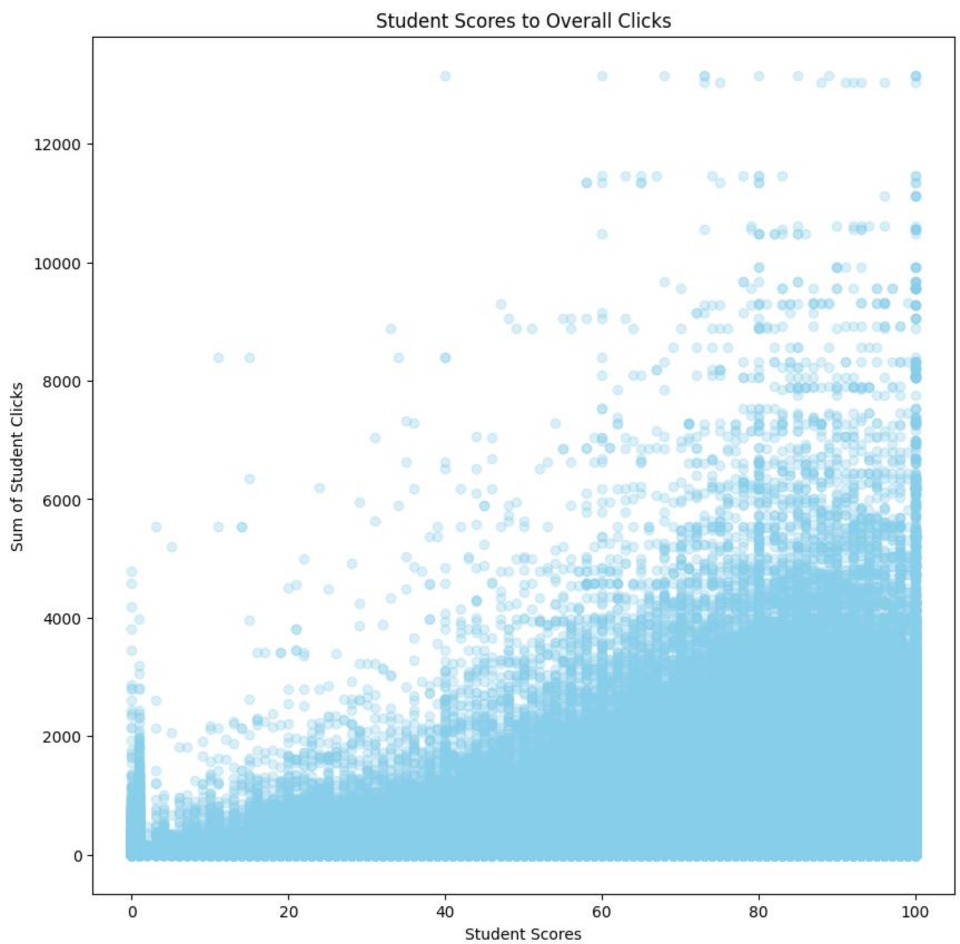
Count of Clicks

The number of times a student interacted with a material



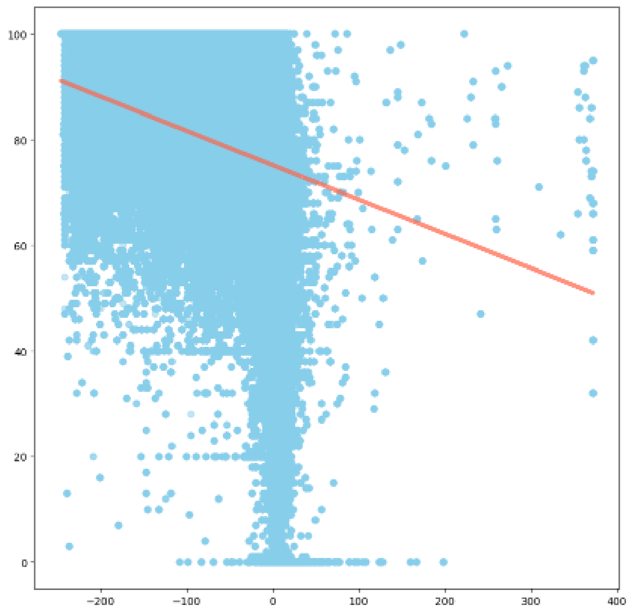
Student's Score increases as he interacts more with VLE

We can see that there is a positive correlation between score and sum of clicks for students, Interestingly there is a spike of scores at 0 where it looks like students with no clicks do better than some students with a number of clicks

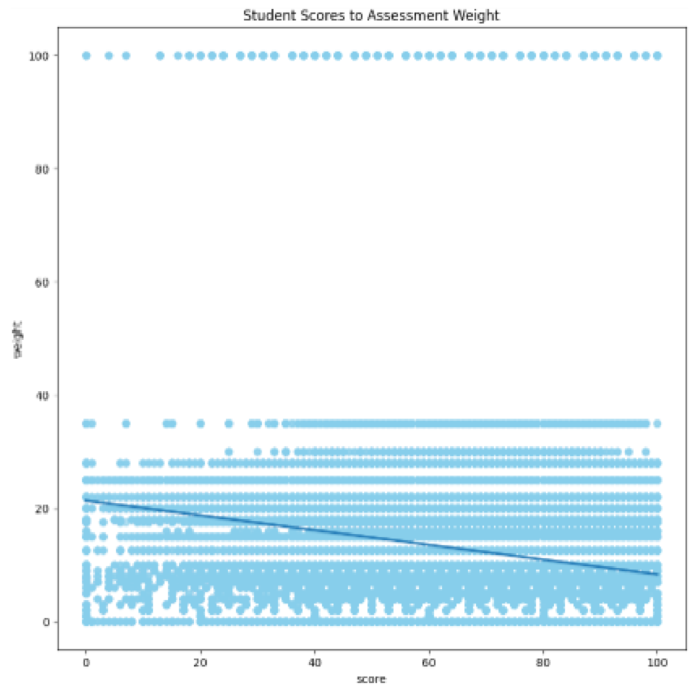


Late Submission Negatively impacts student's score

As a student submits the assessment sooner he get a better score

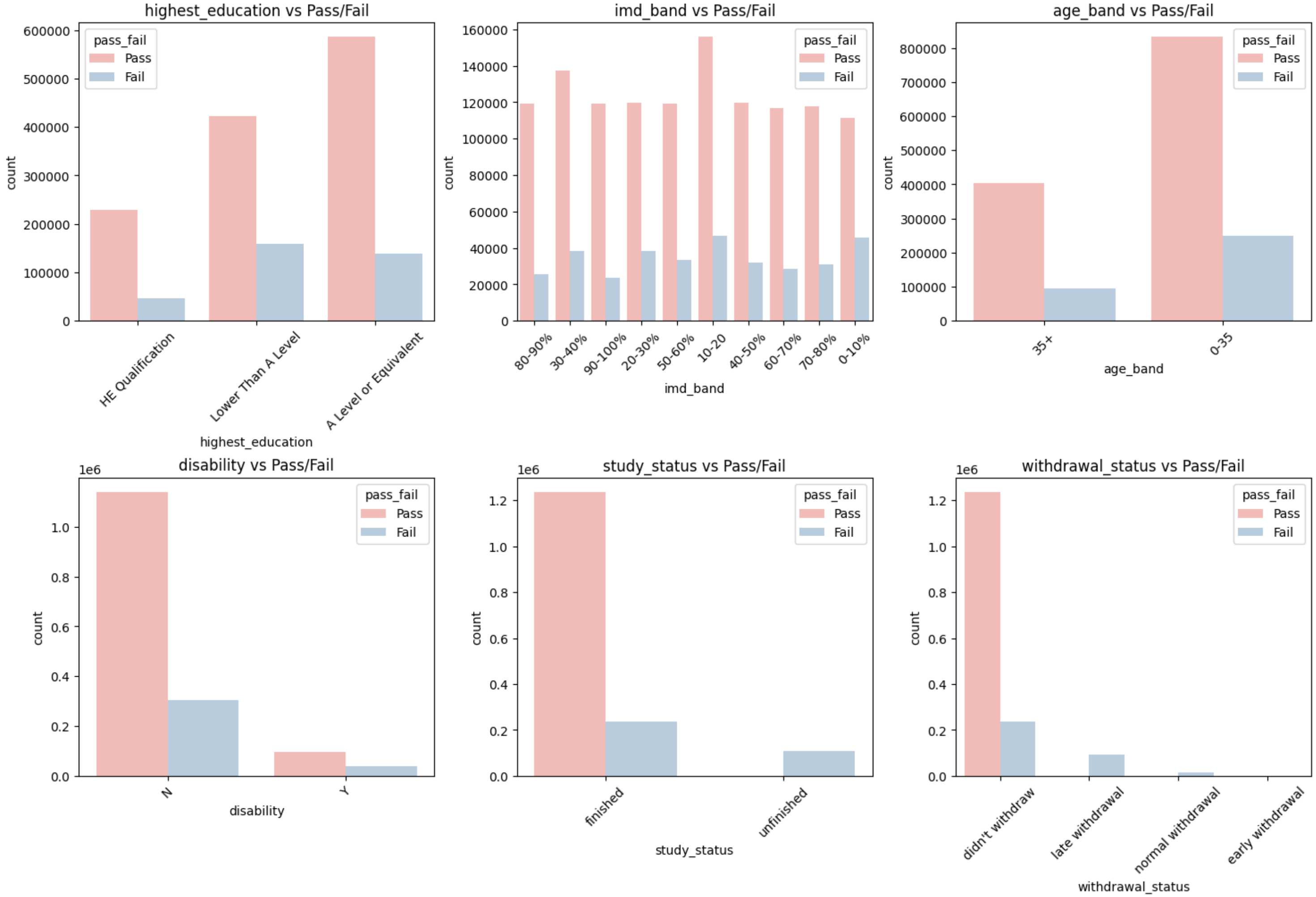


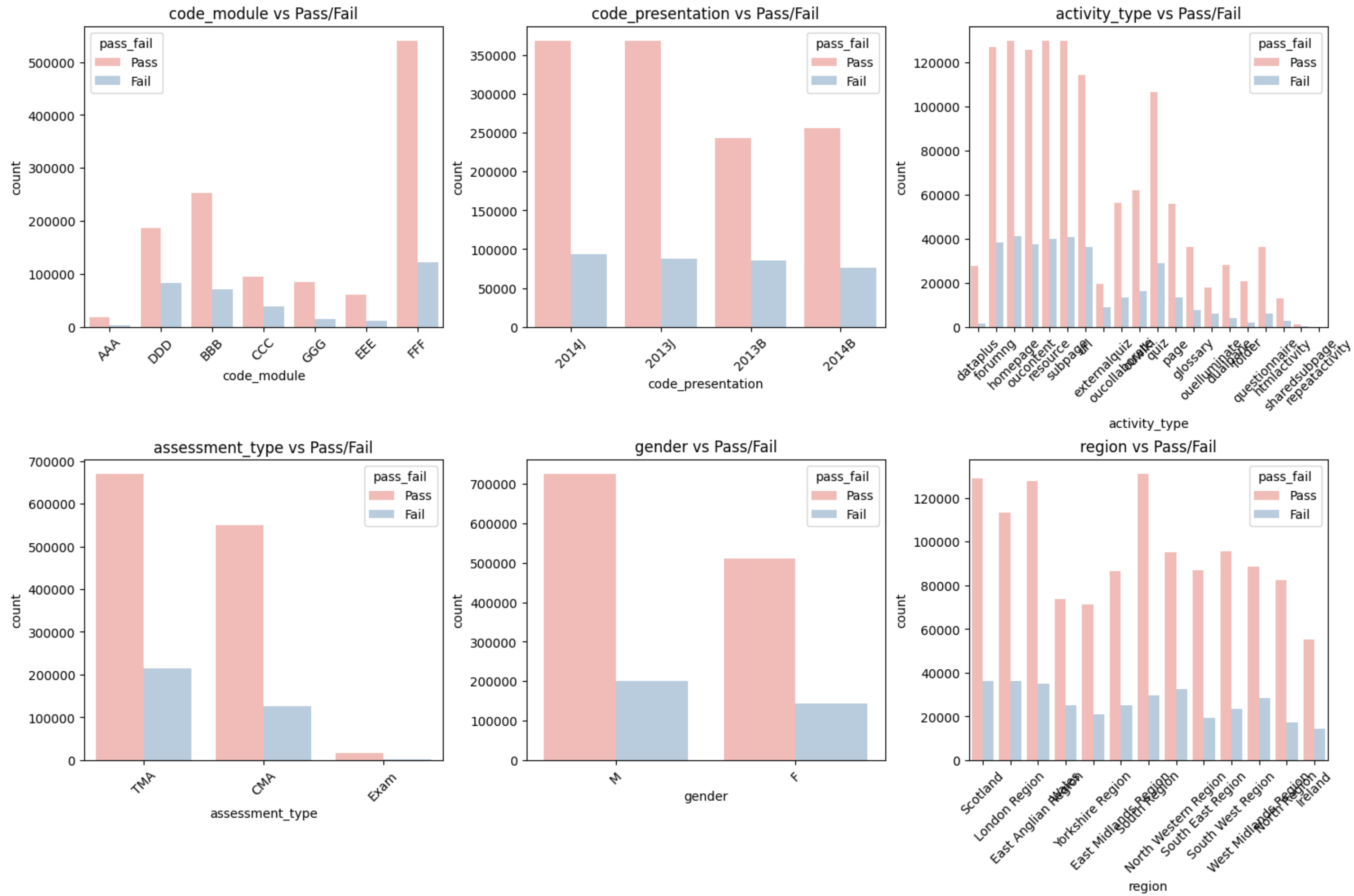
Assessment weight doesn't seem to have that much effect on student's score



Analysis With Final Result

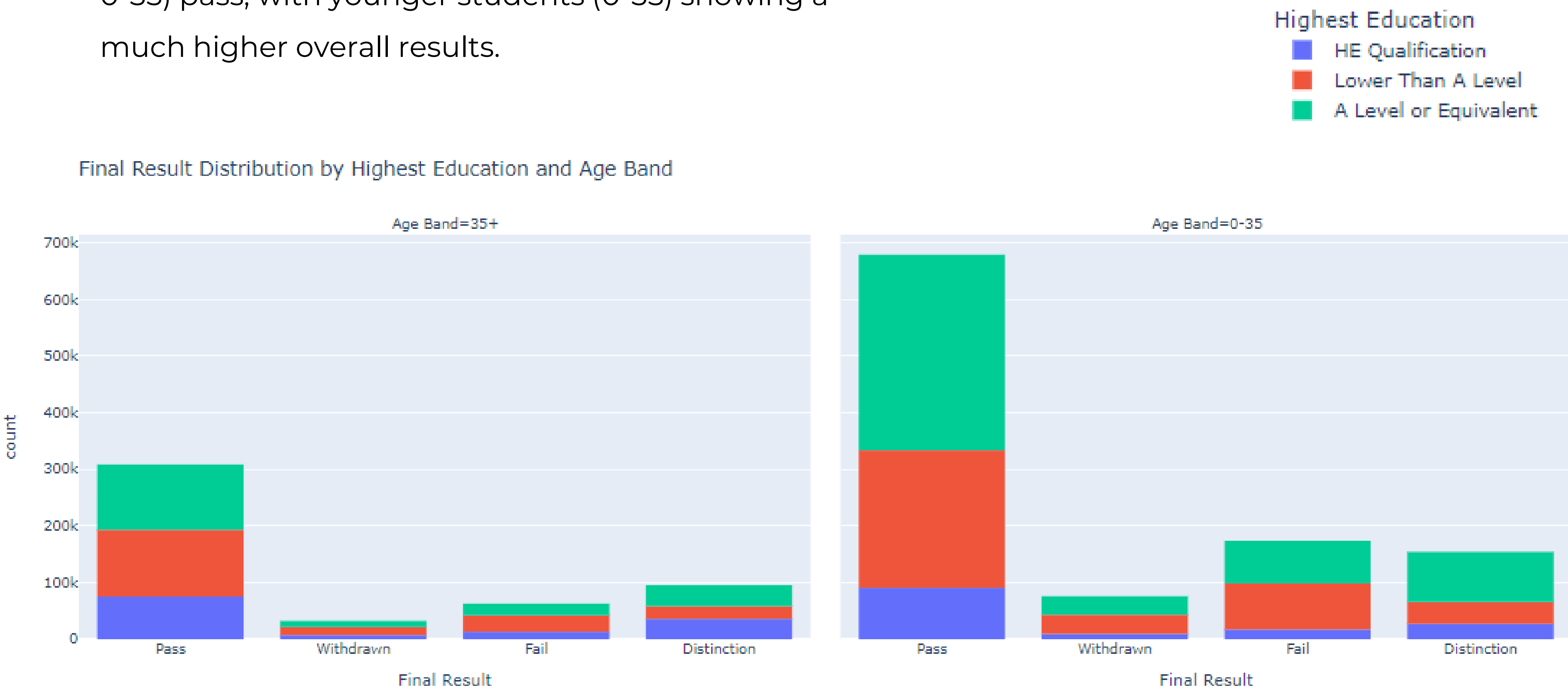
Analyzing the distribution of different categorical columns against the final result provides insights into the dominant categories, particularly those with scores above the 40% passing grade.




code_module vs Pass/Fail
pass_fail
Pass
Fail
count
code_module
AAA
DDD
BBB
CCC
GGG
EEE
FFF
code_presentation vs Pass/Fail
code_presentation
2014J
2013J
2013B
2014B
activity_type vs Pass/Fail
activity_type
dataplus
forumng
homepage
oucontent
resource
subpage
url
externalquiz
quiz
page
glossary
folder
questionnaire
htmlactivity
sharedsubpage
repeatactivity
assessment_type vs Pass/Fail
assessment_type
TMA
CMA
Exam
gender vs Pass/Fail
gender
M
F
region vs Pass/Fail
region
Scotland
London Region
Yorkshire Region
North Western Region
South East Region
South West Region
Ireland

Different Ages Highest Education vs Final Result

The majority of students in both age groups (35+ and 0-35) pass, with younger students (0-35) showing a much higher overall results.



Use our OULAD Dashboard for further detailed analysis

Feature Engineering

Feature Engineering is a key part in our project. This will enable us to derive new features that can create a truly tailored learning experience for each student.

01

Using Data Description To Extract Features

Using the dataset description and domain knowledge

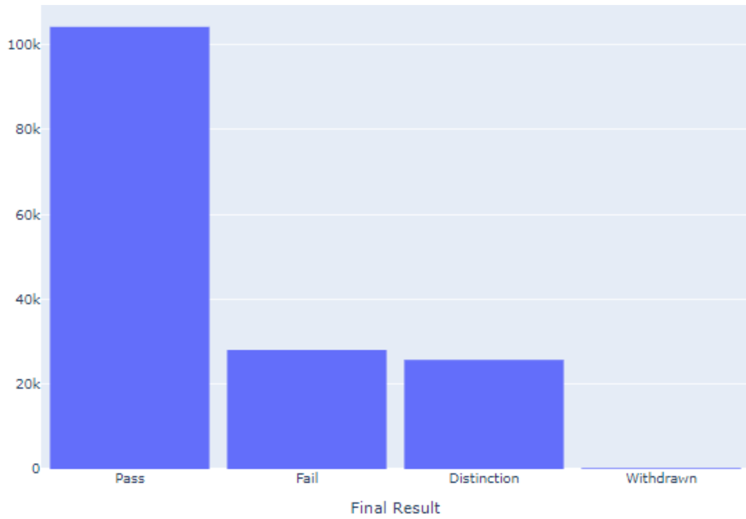
02

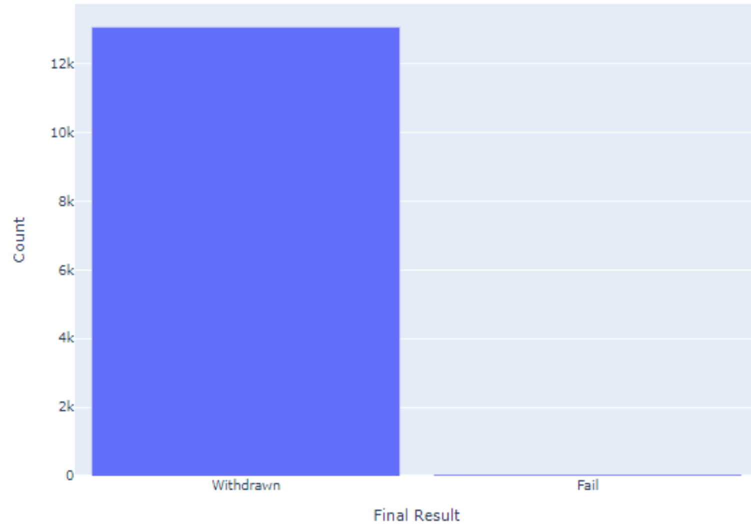
Deriving New Features

involves creating additional, more informative features from existing ones

1. Deriving Study_Status From Date_Unregsiteration

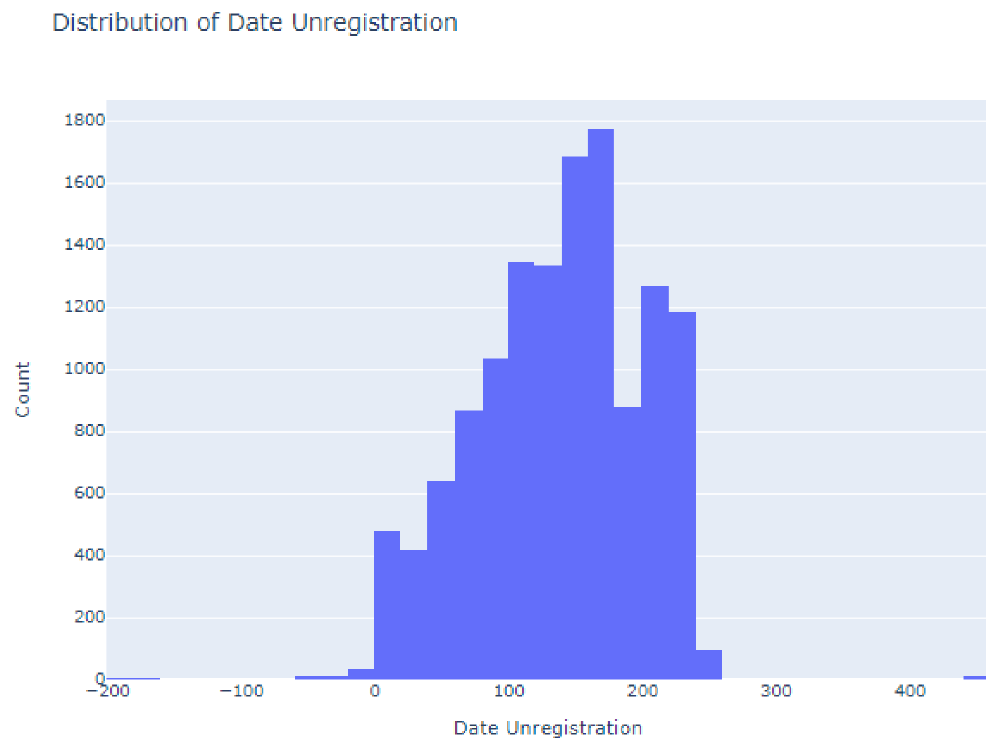
This column contained so many null values, we know that date_unregistered is null when a student doesn't complete his study





2. Binning withdrawal status based on previous analysis

We mentioned that only withdrawal status and a very few number of fails are present in date_unregisteration column, so we can create different bins to represent early-late withdrawals



3. Extracting Study Method Feature

student's preferred study method based on the types of activities they engage in

Poor Conditions:

Low activity across all types, indicating offline content consumption.

Interactive:

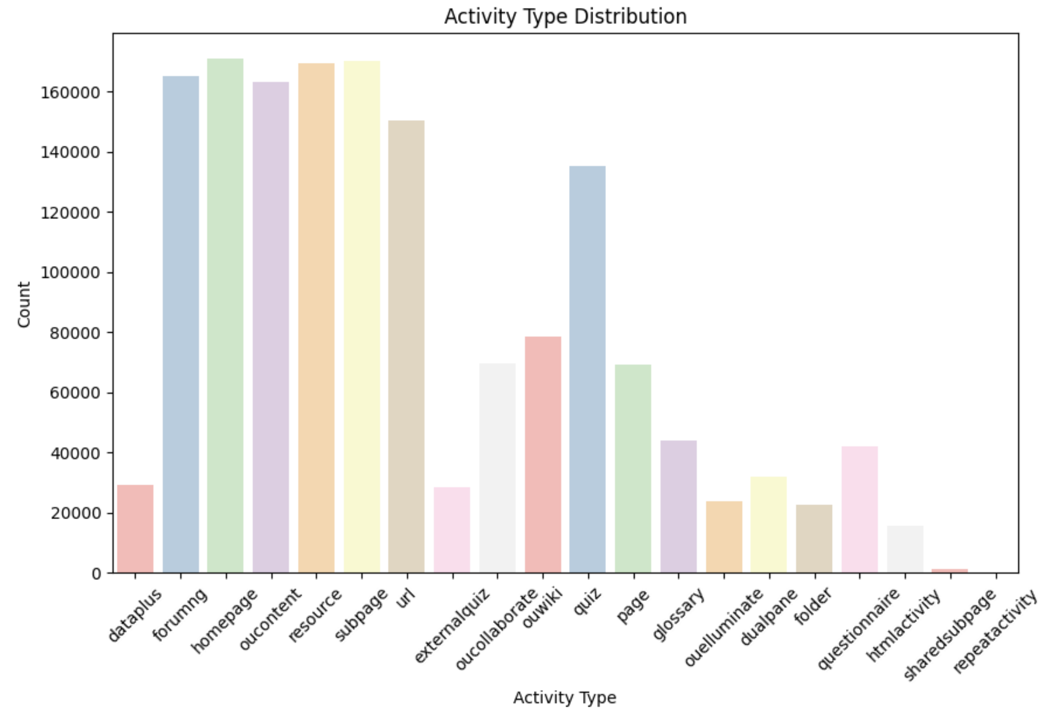
High engagement with quizzes, repeat activities, or questionnaires.

Resource-Based:

Heavy use of resources

Collaborative:

Active participation in group tools



Deriving more Interaction Features

Feature	Explaination
submission_timeliness	The time difference between when the assessment was and the scheduled date of the assessment
score_per_weight	The score normalized by the weight of the assessment.
module_engagement_rate	The rate of engagement with the module, calculated as the ratio of sum of clicks
repeat_student	A binary indicator (1 or 0) indicating if the student is repeating the module (1 for repeat, 0 for first attempt).
weighted_engagement	The engagement score weighted by the assessment's weight
improvement_rate	The consistency (standard deviation) of engagement scores (sum) for each student across all their assessments.
learning_pace	The rate of score improvement for a student, calculated as the change in score between their first and last assessments divided by the number of assessments

Modelling

01 Clustering

Based on engagement metrics, we cluster students into groups

02 Classification

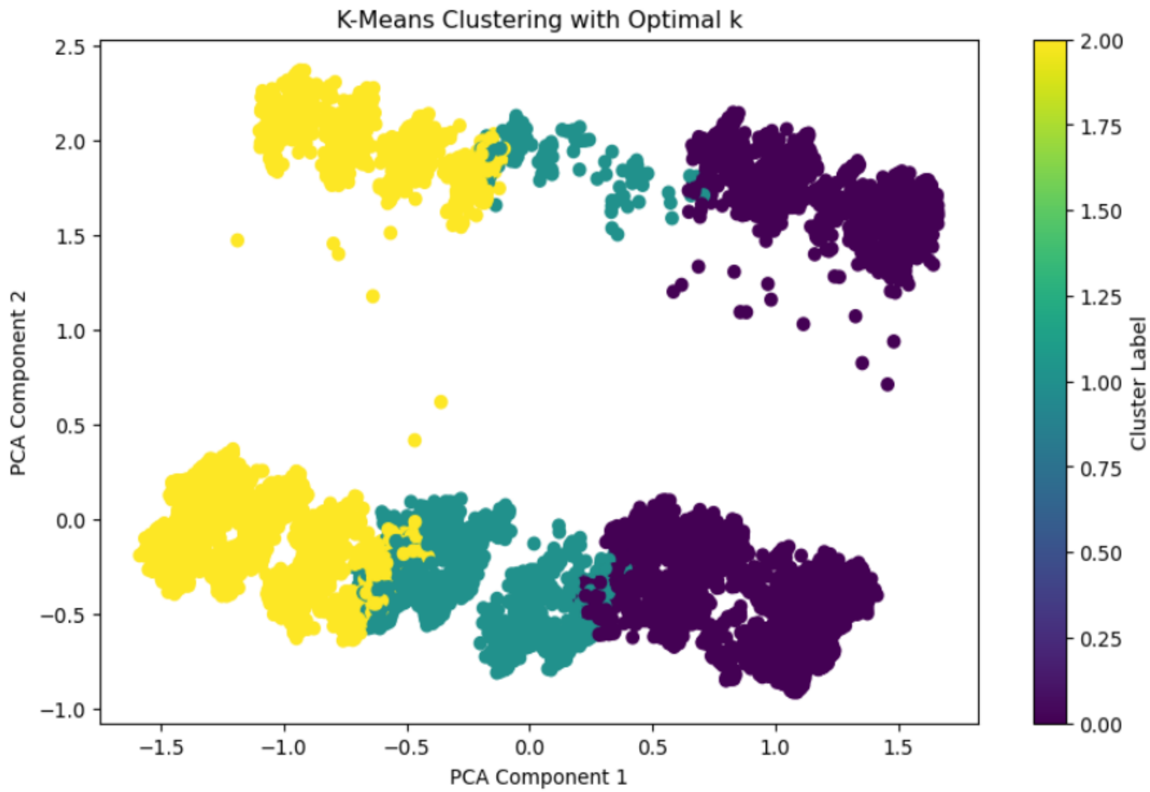
Predicting learning preference for a student

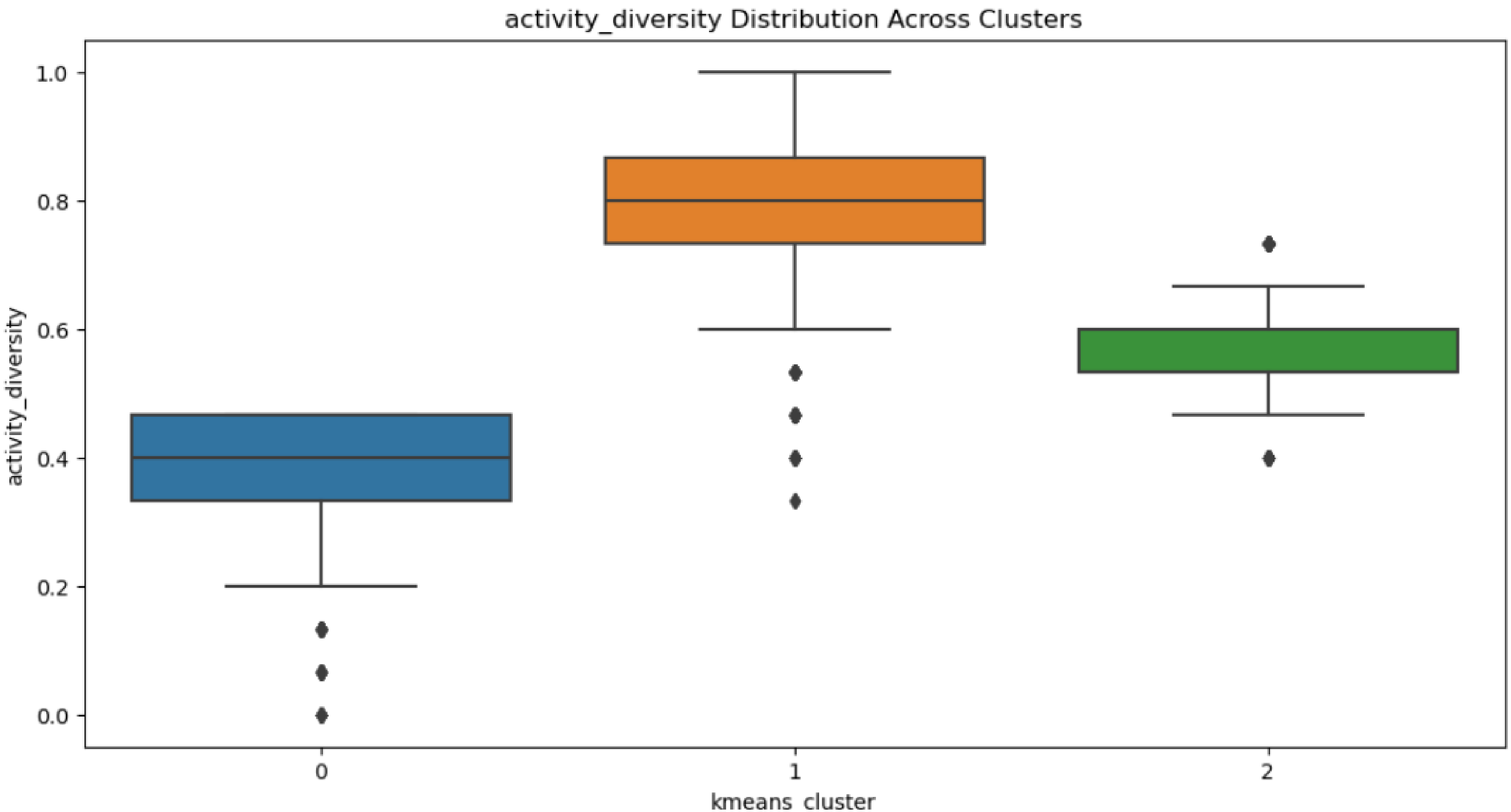
03 Recommender System

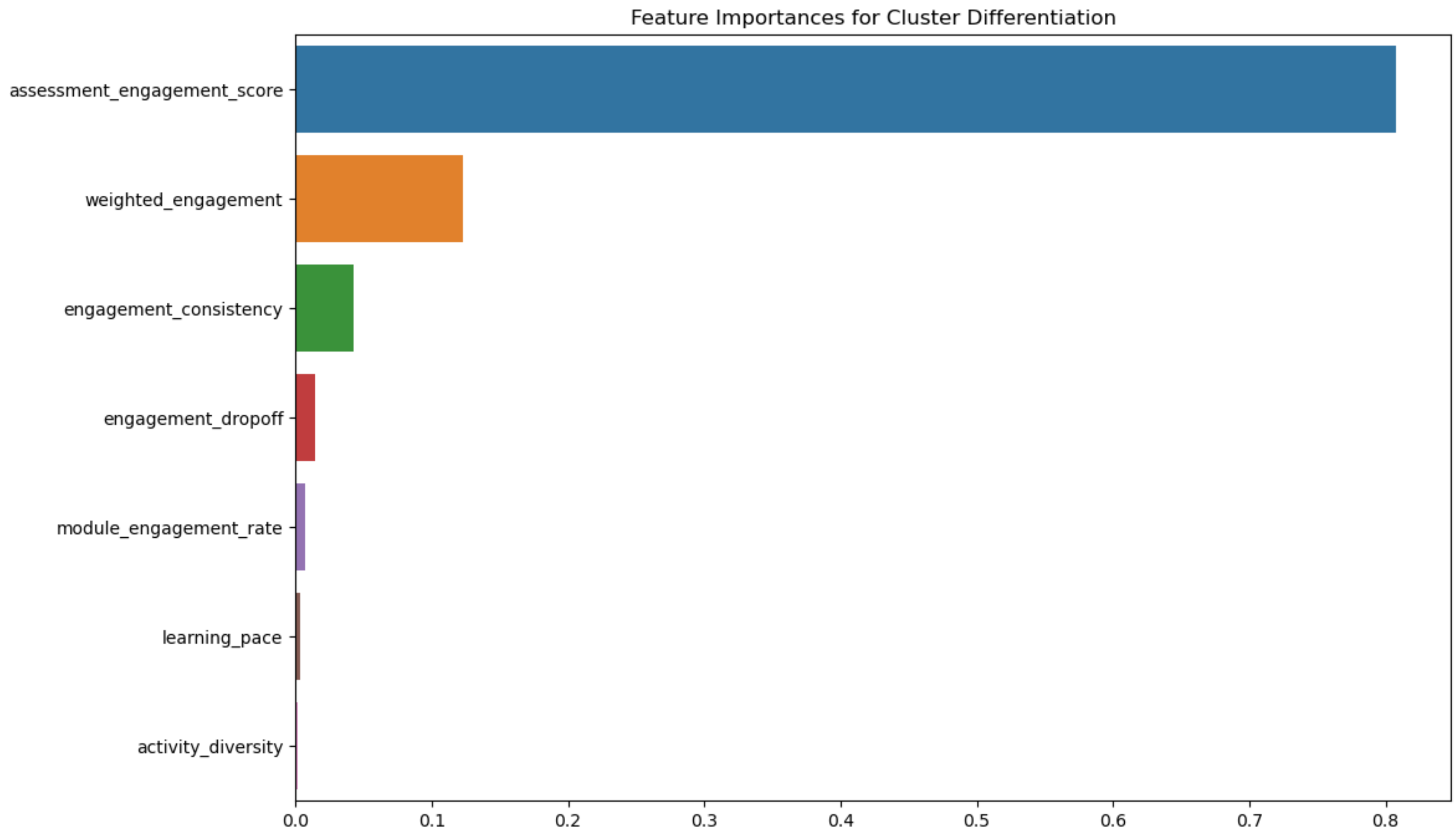
Personalized Learning path recommender

1. Clustering

After scaling and encoding, we perform clustering on engagement-based features to group students according to their interactivity, engagement, and learning levels.




activity_diversity Distribution Across Clusters
1.0
0.8
0.6
0.4
0.2
0.0
activity_diversity
0
1
2
kmeans_cluster


Feature Importances for Cluster Differentiation
assessment_engagement_score
weighted_engagement
engagement_consistency
engagement_dropoff
module_engagement_rate
learning_pace
activity_diversity
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

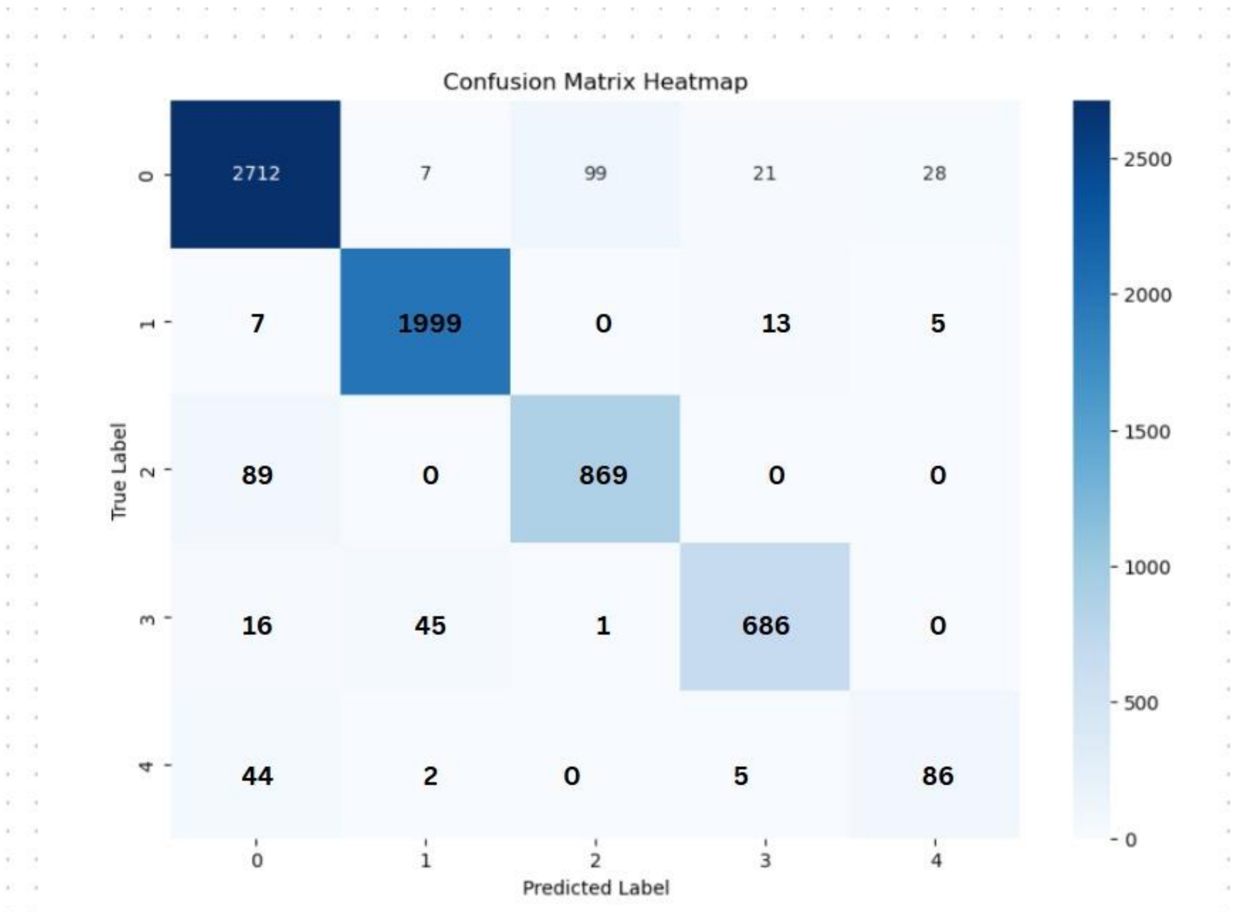
2. Classification

predict the most suitable learning style or content type for each student based on their cluster membership and other features.

```
array([[2712,    7,   99,   21,   28],
       [   7, 1999,    0,   13,    5],
       [  89,    0,  869,    0,    0],
       [  16,   45,    1,  686,    0],
       [  44,    2,    0,    5,   86]], dtype=int64)
```

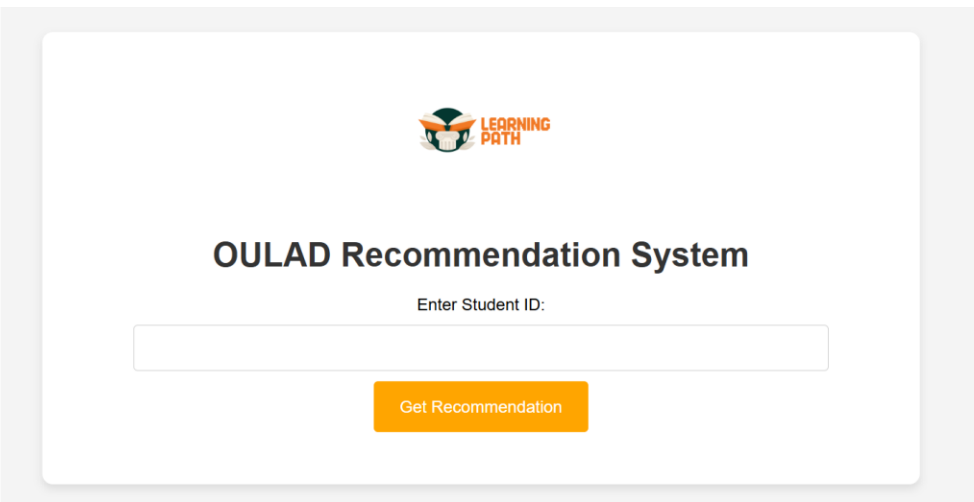
	Model	Train Accuracy	Test Accuracy	Train F1score	Test F1score
0	SVM	0.945643	0.940451	0.944881	0.939814
7	GradientBoost	0.958437	0.939709	0.957764	0.938616
5	Random Forest	0.999936	0.922186	0.999936	0.917442
1	Logistic Regression	0.921966	0.914464	0.917899	0.910224
4	Decision Tree	1.000000	0.905108	1.000000	0.905161
2	KNN	0.872828	0.844075	0.864576	0.834189
6	AdaBoost	0.701483	0.698990	0.731436	0.727228
3	GaussianNB	0.414741	0.413276	0.481092	0.479899

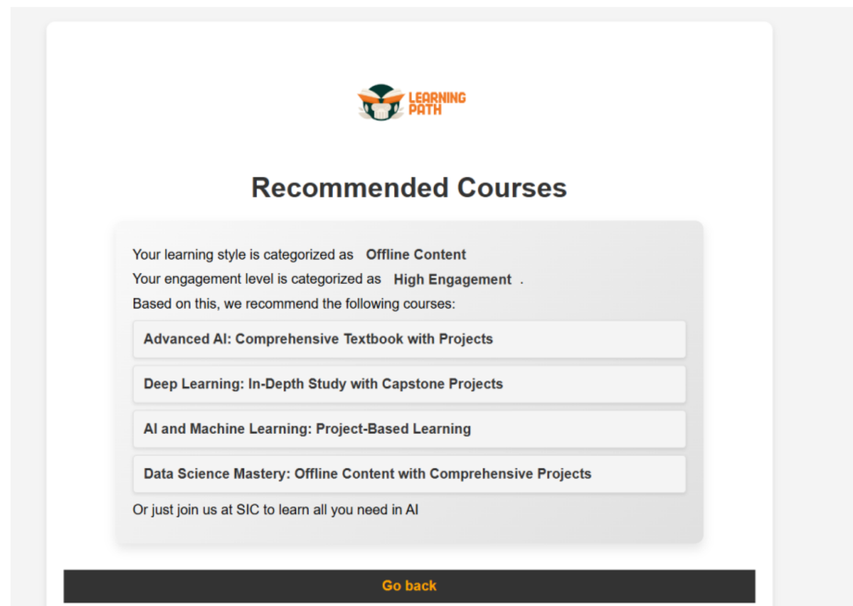
OULAD Learning Path Recommender System



3. Recommender System Deployed

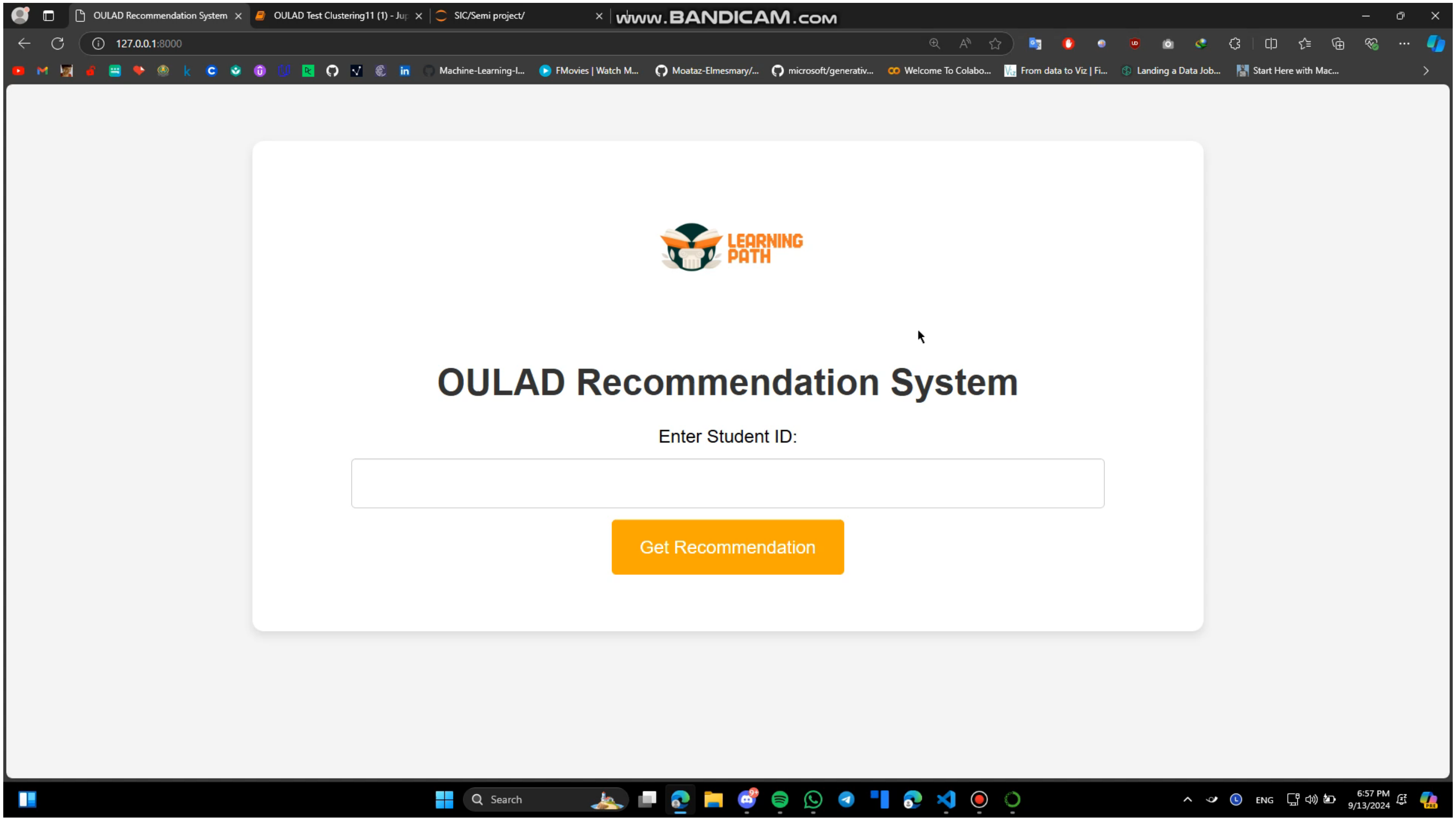
The recommender system uses clustering and classification results to suggest personalized learning materials and pathways.





Team 6

OULAD Learning Path Recommender System



Conclusion

By integrating clustering, classification, and recommender systems, the project can effectively segment students into meaningful groups, predict their learning preferences, and provide personalized recommendations.

- customized plans per individual are not cost-effective and require more effort
- reducing the costs associated with high dropout rates and low completion rates.

Future Plan:

By integrating text classification into our model by analyzing students’ feedbacks and reviews. Will help us to recognize students’ preferences and suitable learning path for each. Gave us more intuitive idea about what will be useful for them. Participating them in the process of model learning. Will give more solid answers.

SAMSUNG

Together for Tomorrow!
Enabling People

Education for Future Generations