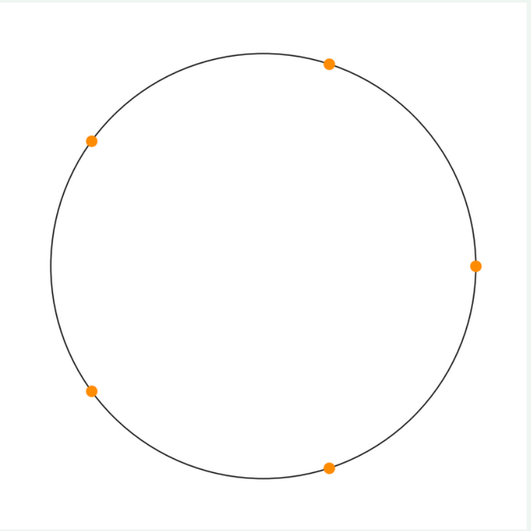
UNVEILING THE DRIVERS OF ACADEMIC SUCCESS

A Machine Learning Approach



24 March 2024

Team 4 - The 5 Yutes of ℝunity

INTRODUCTION

Aim: investigate factors that correlate with academic performance and to analyse recommendations from previous students on how to succeed in the course.

We understand that the data provided by the company contains sensitive information about students, and thus, we ensure utmost confidentiality and privacy protection throughout our analysis. We saw that there were some questions that could induce bias into the model, such as GPA score and mothers education level. Therefore, we prioritised fairness and equity in our models, actively mitigating biases and ensuring that our predictions did not perpetuate existing inequalities or disadvantage any particular group of students.

DATA PRE-PROCESSING

- Analysed individual random data sets to understand key columns that would further benefit our analysis.
- Collaborated with each other to identify relevant data.
- Identified NaN values and removed columns and rows we deemed unnecessary.
- Standardised data.
- Changed multi variable responses while manually scraping the questions.

KEY METRICS

- student identifier (student_id)
- chapter number (chapter_number)
- section number (section_number)
- number of correct answers in end of chapter test (n_correct)
- number of possible answers in end of chapter test (n_possible)
- number of total questions attempts in the end of chapter test (n_attempt)
- proportion of correct answers for end of chapter test (EOC)
- the amount of time user was considered to be engaged (engaged)
- the psychological construct measured by the pulse check item (Cost_Response, Expectancy_Response, Intrinsic_Value_Response, Utility_Value_Response).
- end of survey responses (response)
- end of survey questions (prompt)

HOW DO WE MEASURE ACADEMIC PERFORMANCE?

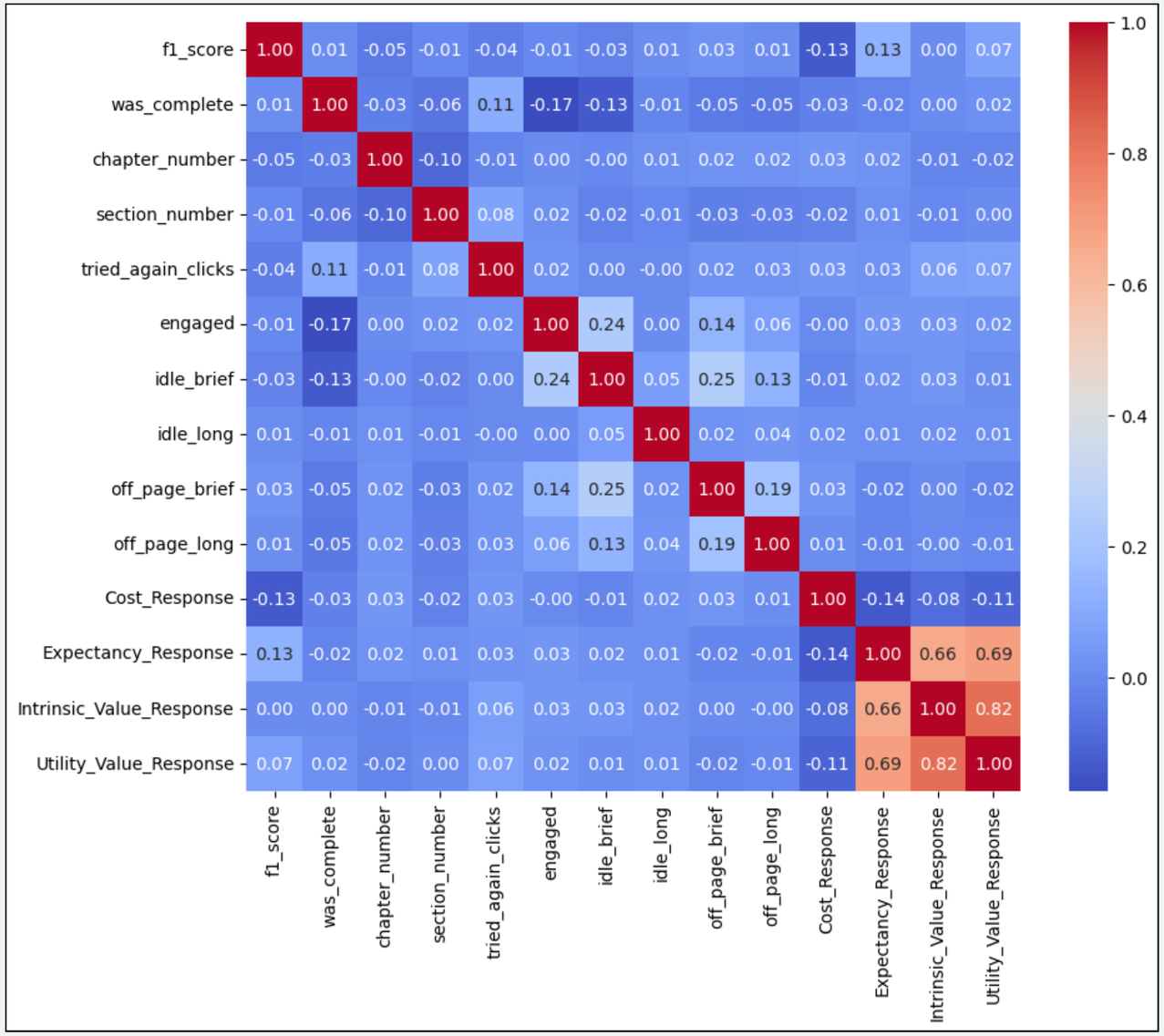
Problem: EOC metric was not informative for academic performance.

Solution: Adding penalties for getting questions wrong, taking inspiration from F1 score.

$$\text{accuracy} = \frac{\text{number correct}}{\text{number possible}} \qquad \text{efficiency} = \frac{\text{number correct}}{\text{number of attempts}}$$

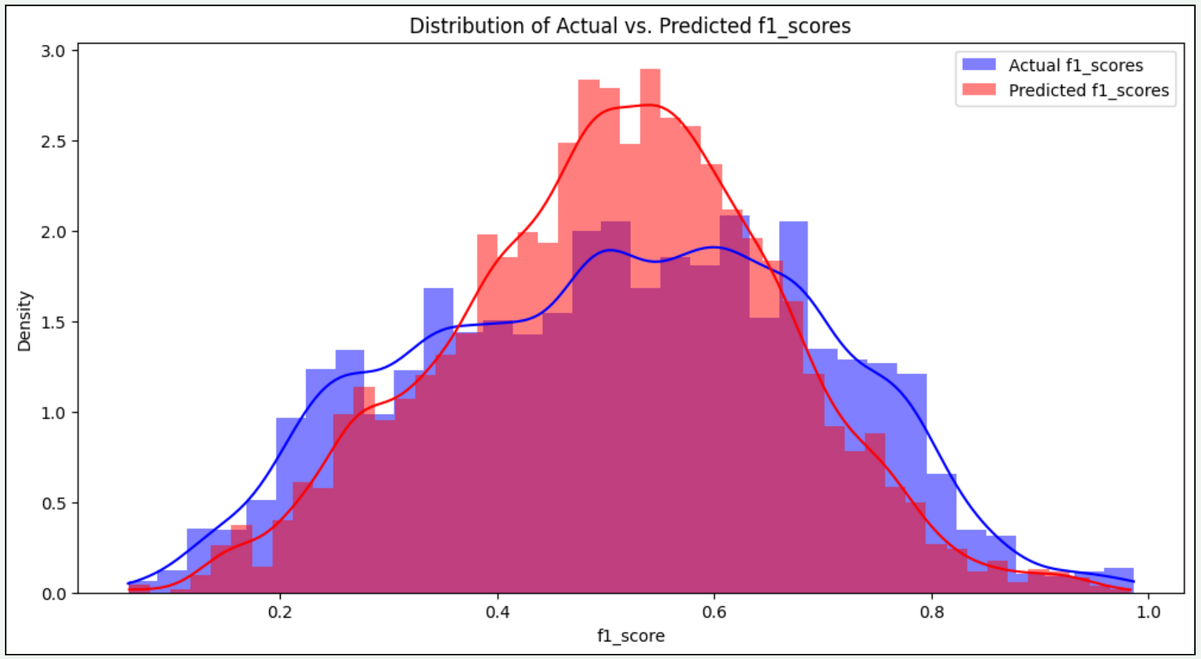
$$\text{F1} = \frac{2 \times \text{accuracy} \times \text{efficiency}}{\text{accuracy} + \text{efficiency}}$$

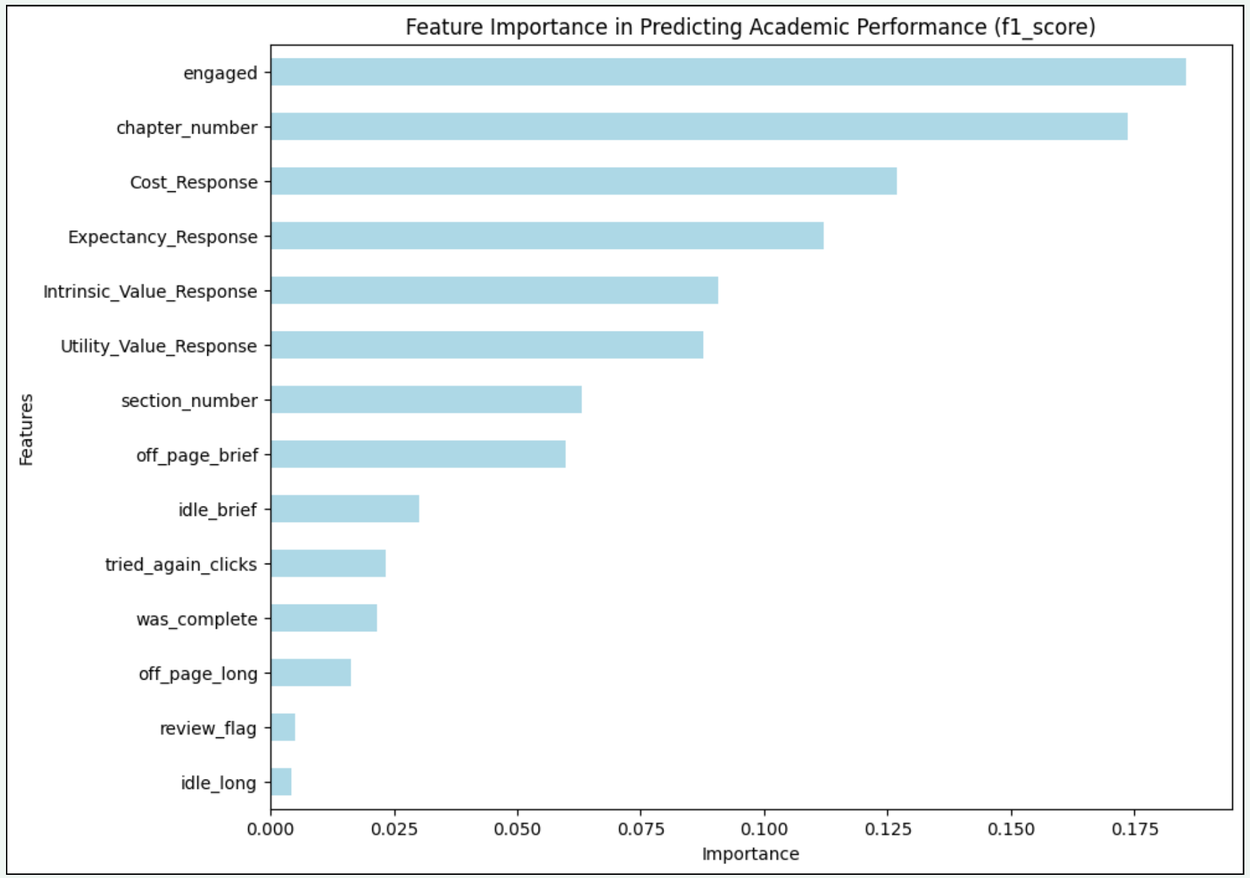
PRINCIPLE COMPONENT ANALYSIS



- PCA is used to find the most important features when making a prediction.
- The heat map shows small correlations as these types of variables are different
- We cannot use multiple linear regression and had to turn to random forest regression.
- Random forest regression is widely used for this type of problem as it uses decision trees to capture more complex relationships between different types of variables.

PCA FINDINGS





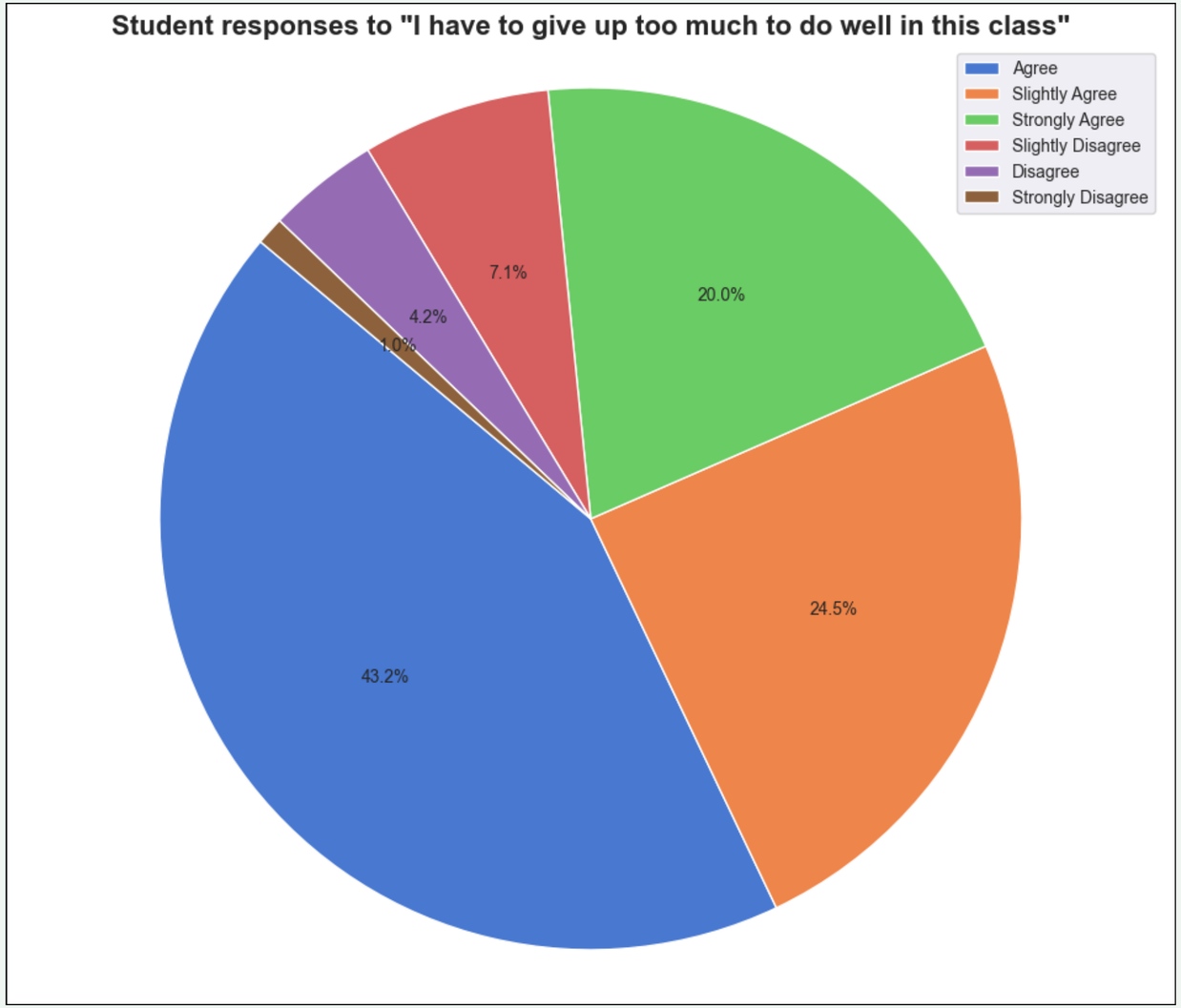
- **Left plot:** accuracy of our random forest regression model (r^2 ~ 0.6).
- **Right plot:** the most important features for predicting our derived F1 score.

ANALYSING STUDENT RESPONSES

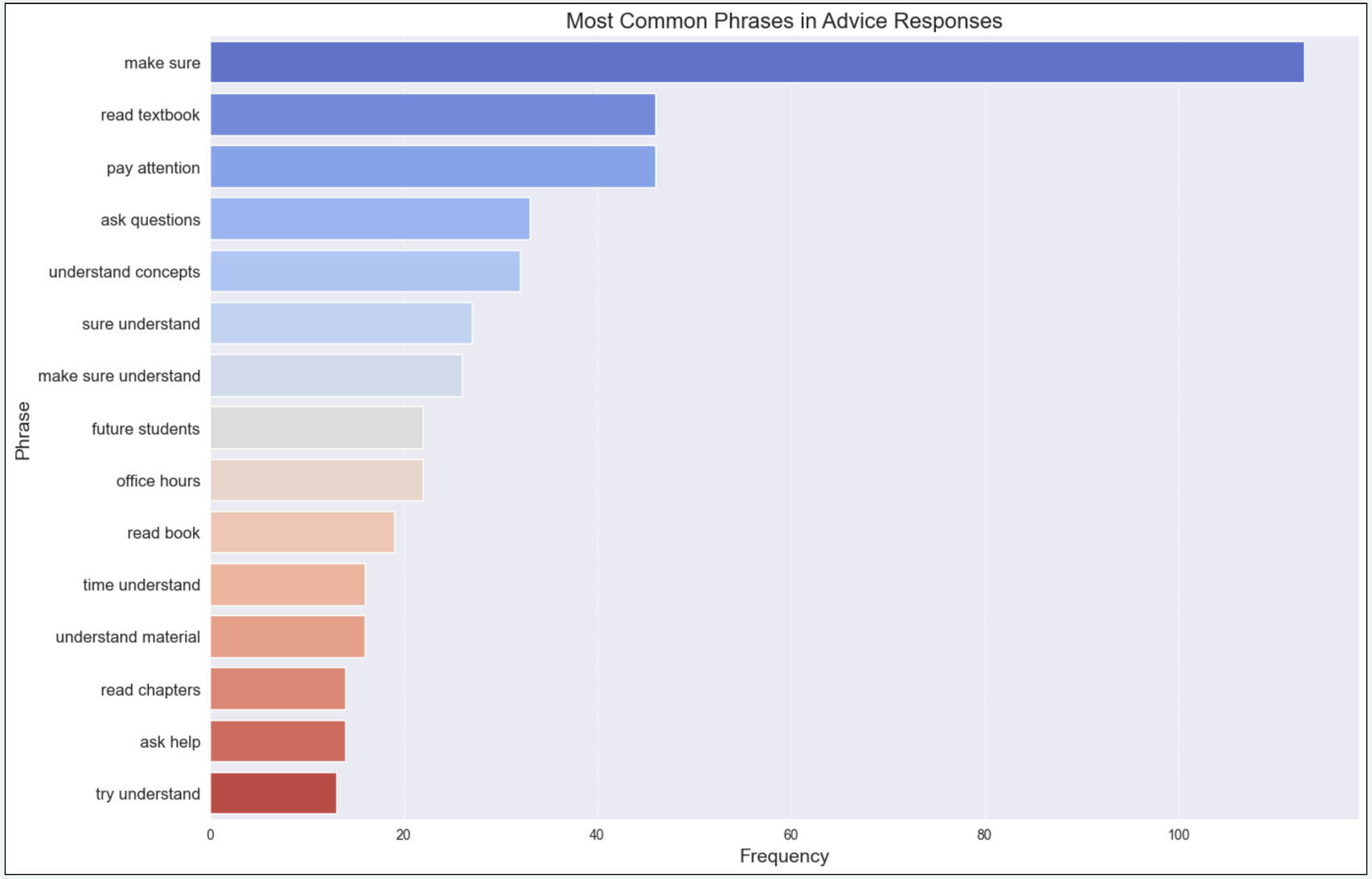
- We investigated responses students gave to the level of agreement statements in the end of course survey.

Our findings:

- Most students found the class interesting and relevant.
- Most students agreed that to do well in the class they had to give up too much.
- This agrees with our previous findings.



SENTIMENT ANALYSIS



DRAWING CONCLUSIONS

- Positive relations between direct student engagement and academic performance.
- Claim further reinforced by sentiment analysis.
- Current leading academic studies agree with our conclusion [1].

Advice to future students:

- Ask questions.
- Read textbooks.
- Attend office hours.
- But most importantly... stay engaged!

[1] https://www.researchgate.net/publication/324183400_Relationships_between_student_engagement_and_academic_achievement_A_meta-analysis

CONTINUATION OF ANALYSIS

Continuation of current analysis:

- Perform analysis on full data set.
- When performing PCA, we should incorporate media views data set.

Alternative routes of analysis:

- Give institutions tailored feedback.
- Work on a dashboard to visualise student performance based on key features identified.