



MOOC VISUALIZATION AND ANALYTICS

Jason Baik



Huzefa Rangwala,
Jessica Lin,
Jill K. Nelson

Background

- MOOCs offer low-cost and convenient education to students
- Huge problem in MOOCs is the low-completion rates
- Widely cited dropout rate is 90%

Objectives

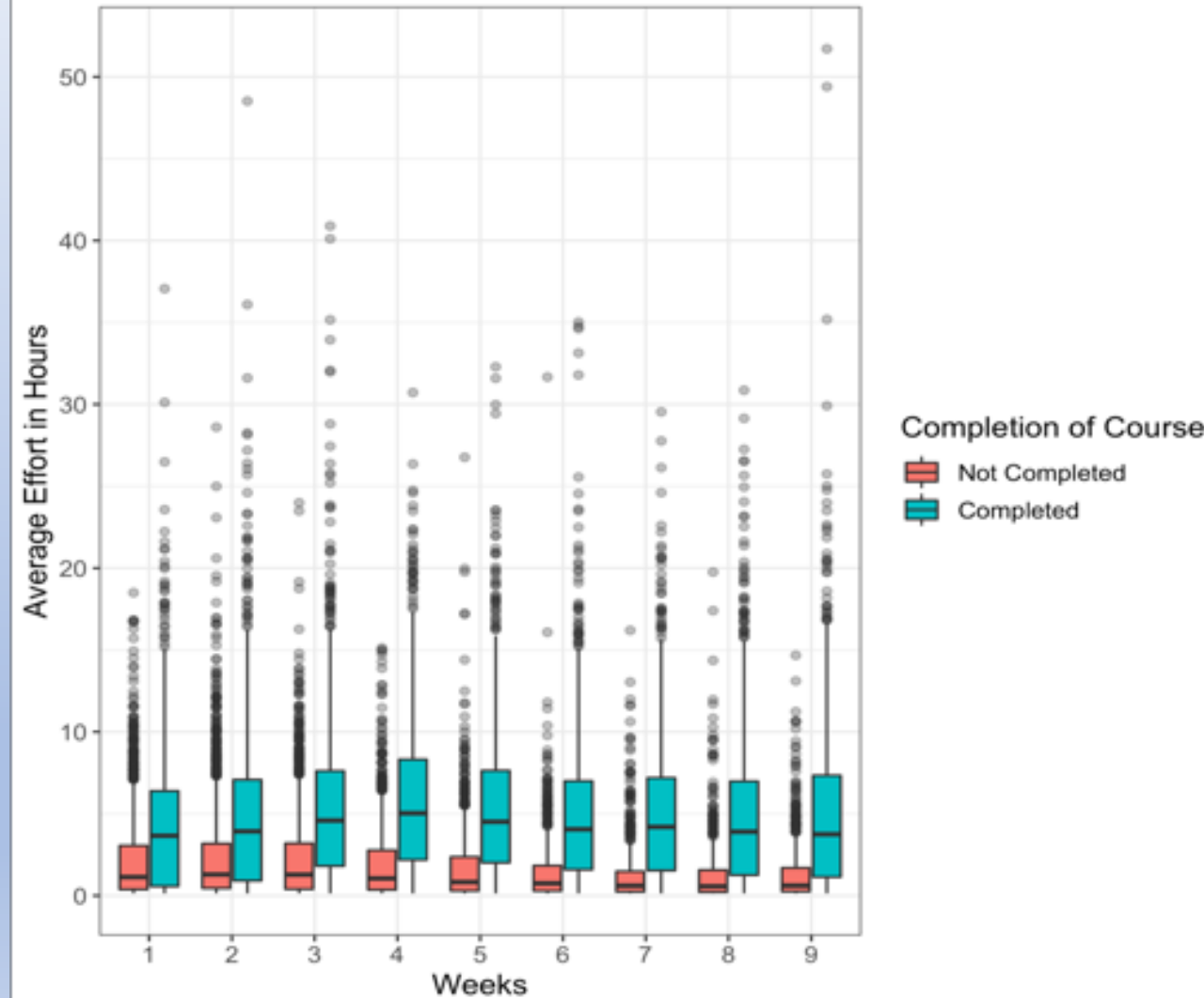
- Explore four datasets from Stanford edX's Statistics in Medicine
- Employ Different K-Means Clustering
- Provide an interactive web application using R for instructors and course administrators

Exploratory Data Analysis I

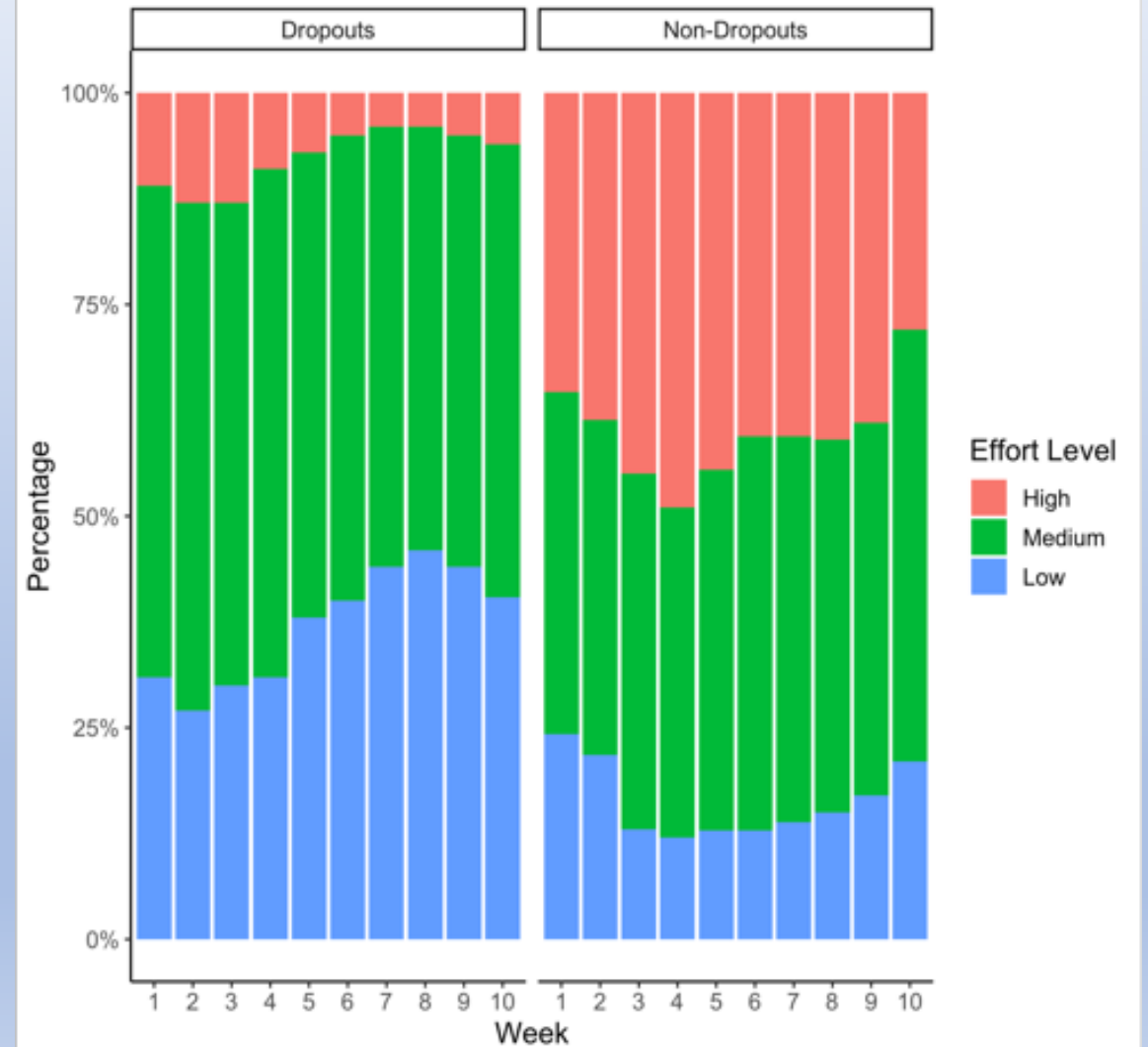
- Number of students who are in all four datasets: 7659
- Number of students who dropped out: 4996
- Dropout Rate: 65%

Exploratory Data Analysis II

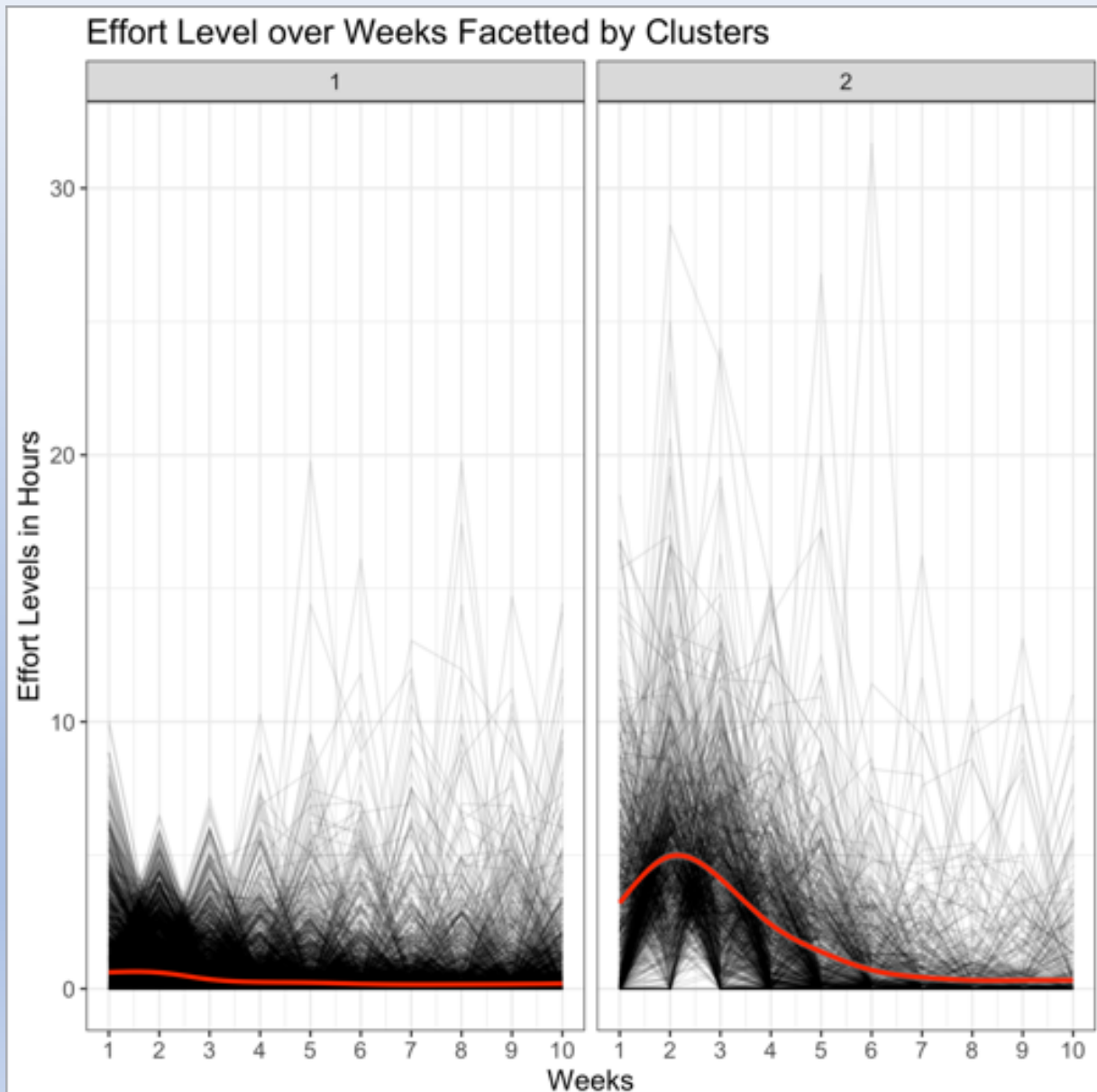
Distribution of Average Effort in Hours Per Week
By Completion of Course



Distribution of Effort Level per Week
Proportion of High Effort Level increases along with Week

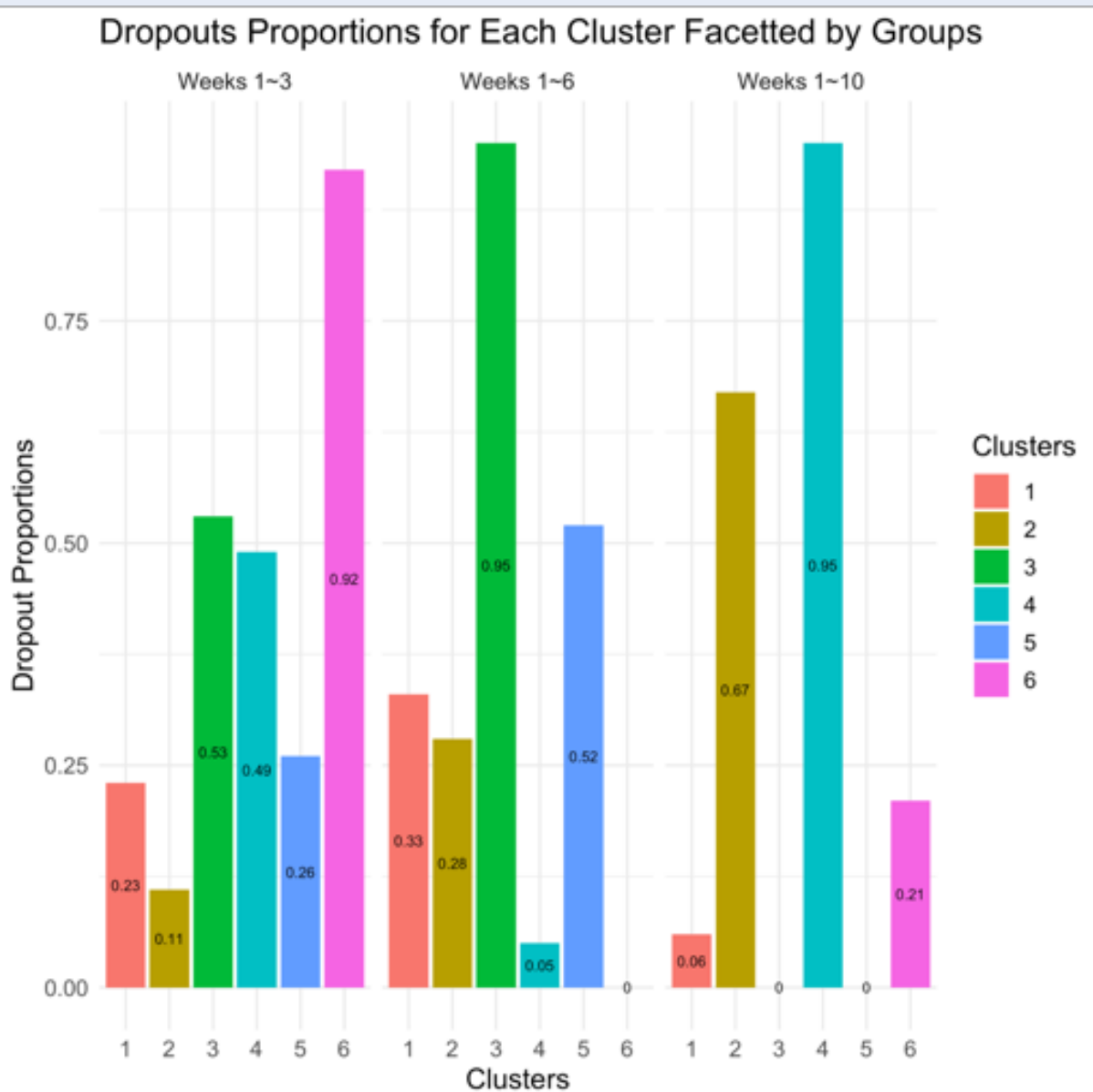


Exploratory Data Analysis III



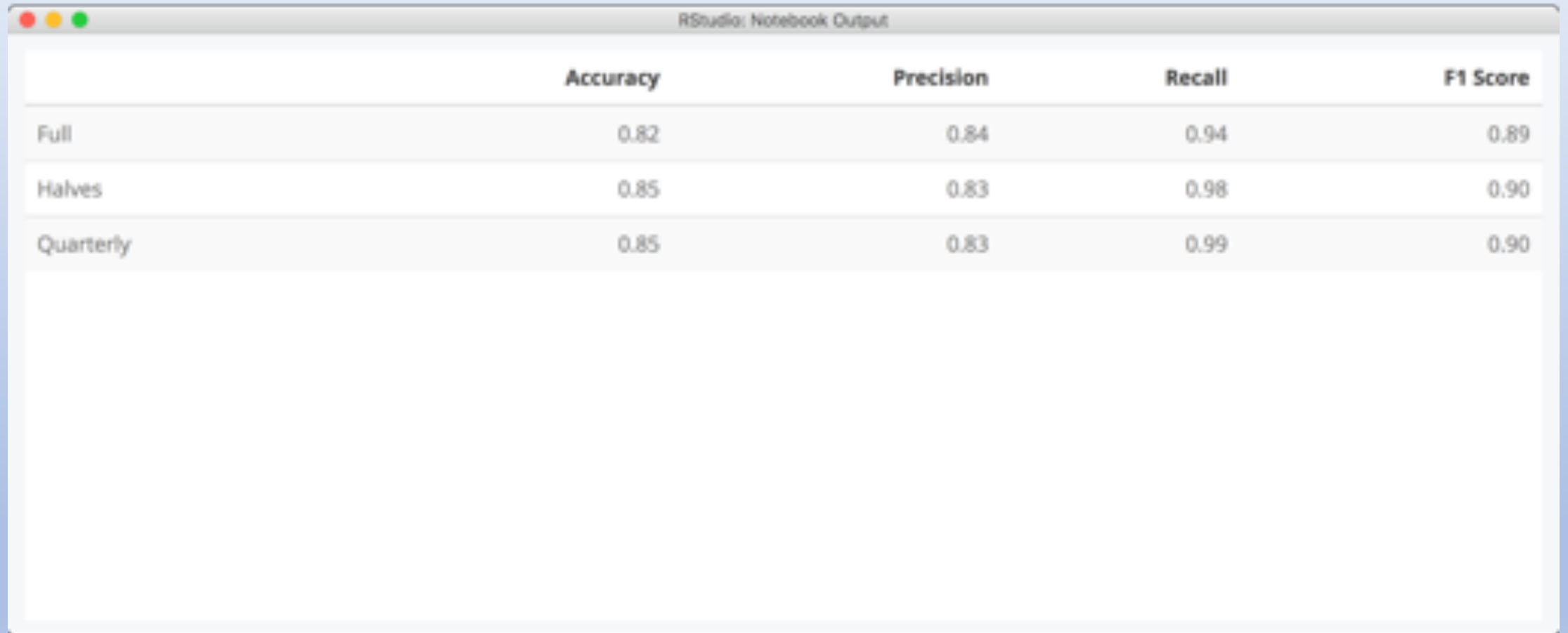
- (K=2) K-Means Clustering on Dropout Students
- Used Hours of Effort as Only Feature
- Among dropouts, observed two clusters of students with distinct trend lines of effort levels

(K=6) K-Means Clustering



- Used two features: Hours of Effort + Number of Times Student Played Video
- Divided into Three Time Periods
- Evolve across the different time periods
- Compared Students in Six Clusters to Ground Truth of Dropout Students.
- Three Different Clusters contained an extremely high proportion of Ground Truth Dropout Students

Clustering in Stages



The image shows a screenshot of the RStudio Notebook Output window. The window title is "RStudio: Notebook Output". It displays a table with five columns: Accuracy, Precision, Recall, and F1 Score. The rows represent different clustering stages: Full, Halves, and Quarterly. The data is as follows:

	Accuracy	Precision	Recall	F1 Score
Full	0.82	0.84	0.94	0.89
Halves	0.85	0.83	0.98	0.90
Quarterly	0.85	0.83	0.99	0.90

Shiny Dashboard Demo

https://jsonbaik.shinyapps.io/shiny_google_form/