

BFA Enrollment and Curriculum with Data Analysis

Afolabi Isiaka
School of Graduate Professional
Studies of Pennsylvania State
University
Great Valley, PA
avi5117@psu.edu

Abstract—This study investigates the enrollment trends and curriculum structure of the Bachelor of Fine Arts (BFA) program at the University of Diploma Printing in Romanigstan's Department of Arts. The department, known for its diverse and distinguished faculty, offers a wide range of specialized courses. However, recent years have seen a stagnation in student enrollment, prompting a comprehensive review of the curriculum. The analysis involves preprocessing and cleaning student enrollment records and course data from the past five years. The study identifies discrepancies between the course catalog and actual enrollments, uncovering unmatched courses that may no longer be relevant. Using cluster analysis, the research reveals significant enrollment patterns, guiding recommendations for curriculum redesign. The findings suggest reducing total credit hours, grouping electives into concentration areas, and streamlining the course offerings to better align with contemporary student interests and academic trends. These insights aim to revitalize the BFA program, ensuring its continued relevance and appeal to prospective students.

Keywords—*elective courses, data analysis, cluster analysis*

The University of Diploma Printing in Romanigstan's Department of Arts, a key part of the College of Arts and Sciences, features a distinguished faculty recognized for their contributions to the arts. The Department offers a Bachelor of Fine Arts (BFA) program with diverse areas such as puppetry, adventure recreation, and fishing sciences. The program includes seven core and six elective courses, totaling 39 credit hours, with 50 courses available to art and non-art majors. However, despite its strengths, the program has seen stagnant enrollment over the past five years, prompting a need for a comprehensive curriculum review and redesign.

I. OBJECTIVES

A. Reduce the number of credit hours

Decrease the total credit hours required for the BFA program from 39 to 33, streamlining the curriculum while maintaining its integrity and comprehensiveness.

B. Define concentration areas for elective courses

Group elective courses into coherent concentration areas that reflect student interests and contemporary disciplinary trends, providing clearer pathways for students to specialize in their areas of interest.

C. Reduce the number of offered electives

Identify and retain the most relevant and popular elective courses, reducing the overall number of electives to create a more focused and manageable curriculum.

II. DATASET DESCRIPTION, INITIAL PROCESSING

A. Data Mining and Analysis Framework

This investigation will be performed within the framework of the Cross-Industry Standard Process for Data Mining (CRISP-DM) [1].

B. Utilized Software and Systems

Data cleaning, processing, and analysis steps were performed with the use of python for data analytics.

C. Data Description

The course dataset contains information on courses offered in the Bachelor of Fine Arts (BFA) program at the University of Diploma Printing in Romanigstan. It includes foundational courses, a 21-credit core curriculum, and a range of elective courses. Core courses cover diverse subjects like Environmental Systems, Computer Linear Algebra, and Contemporary African Art. Elective options allow students to explore various topics, from American Social Policy to Elementary Arabic.

This registration dataset is organized into categories that reflect the structure of the curriculum, making it easy to analyze. The data is useful for understanding the course individual students are registering for as well as the semester.

D. Initial Pre-Processing

Initial Pre-processing required changing file type of course data to 'docx' for easier implementation of 'python-docx' library. Data from word files could be extracted and stored this way.

III. DATA PREPROCESSING

The analysis of the dataset provides crucial insights into the structure and content of the student enrollment records. Before gather insights we need to look for duplicates and Null values in our datasets.

We will handle duplicate and null values that could cause are data to be inaccurate by dropping the rows that contain these items. The heatmap below shows where are null values exist and how frequent in each dimension. We observe that course name contains null values in *figure 1*.

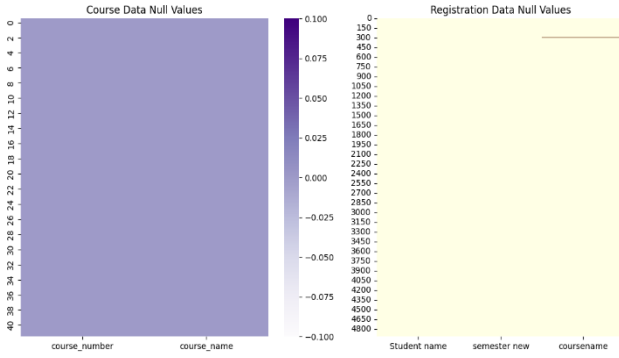


Figure 1.

In addition to observing Null values, investigation of Registration data revealed 1249 duplicates while Course data contains no duplicate rows.

After data cleaning we see the results of cleansing null values in *figure 2*. Course and registration data has also been checked for duplicate rows, and if existing, removed. Data is cleaned therefore we can perform exploratory analysis to gather accurate insights into our data.

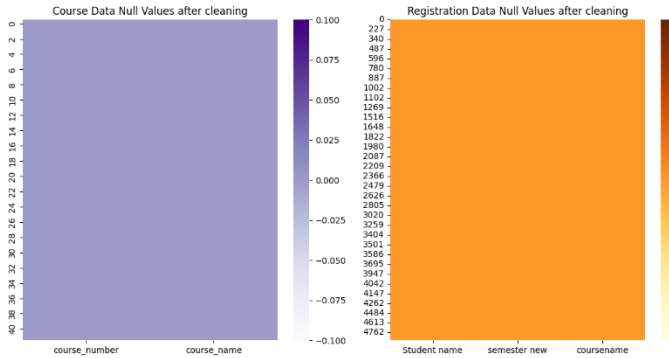


Figure 2.

IV. DATA UNDERSTANDING

By merging the datasets using unique course names, we identified a discrepancy between the course catalog and enrollment data, resulting in 139 unmatched courses. Notably, some unmatched courses, such as 'DRUGS, BRAIN AND MIND' and 'FORENSIC ANTHROPOLOGY', had student registrations. This suggests that these courses may not have been offered in the last five years or are planned/proposed courses without enrollment records.

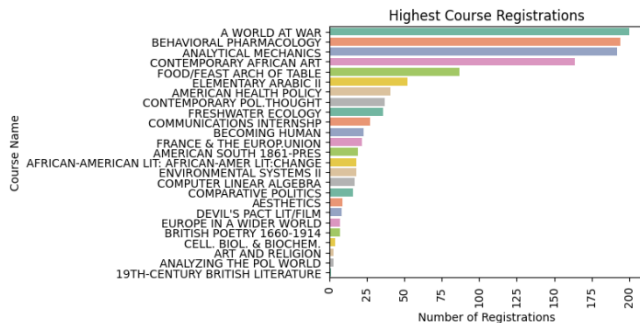


Figure 3.

Figure 3 illustrates the enrollment trends for each course over the past five years. The chart highlights the most popular courses, with 'A WORLD AT WAR' consistently showing high enrollment numbers. This indicates areas of strength in

the current curriculum and potential foundation courses for concentration areas.

The dataset combines two main components: a list of current courses offered by the Department of Arts and a record of student enrollments for the last five years. Results of merger are seen in Table I.

TABLE I. DATA DICTIONARY

Column Name	Data Type	Description
student_name	object	Name of the student
course_number	object	Unique identifier or number assigned to the course
course_name	object	Name of the course
semester	object	Semester in which the student enrolled (e.g., Fall, Spring)
year	object	Academic year in which the student enrolled (e.g., 2019, 2020)

V. MODEL SELECTION AND DATA PREPROCESSING

To unlock insights from the enrollment data, we embarked on a thorough data preparation journey. Our approach consisted of the following key steps:

A. Data Preprocessing

1) *One-Hot Encoding*: leveraged the OneHotEncoder from scikit-learn to transform categorical columns into numerical representations. This enabled us to tap into the hidden patterns and relationships within the data.

2) *Data Concatenation*: By concatenating the one-hot encoded features with the original DataFrame, we created a comprehensive dataset that paved the way for in-depth analysis.

3) *Streamlining the Data*: To ensure efficiency and prevent redundancy, we removed the original categorical columns from the dataset.

4) *Pivot Table Creation*: Through the groupby method, we aggregated the dummy variables and created a pivot table with student names as the index. This facilitated the analysis of course enrollment patterns and trends.

B. Model Selection

With our prepared data in hand, we selected the following models to uncover meaningful insights:

- Cluster Analysis: Uncovering Hidden Patterns

By clustering students with similar enrollment patterns, we verified the consistency of our results and gained confidence in our recommendations.

VI. RESULTS AND EVALUATION

To identify optimal groupings of elective courses, we employed a K-means cluster analysis. Parameters for are cluster is k (number of groupings).

Using the elbow method to determine the optimal number of clusters, k. The elbow plot (Figure 4) reveals an optimal cluster count of 3, indicating 3 distinct groups of elective courses.

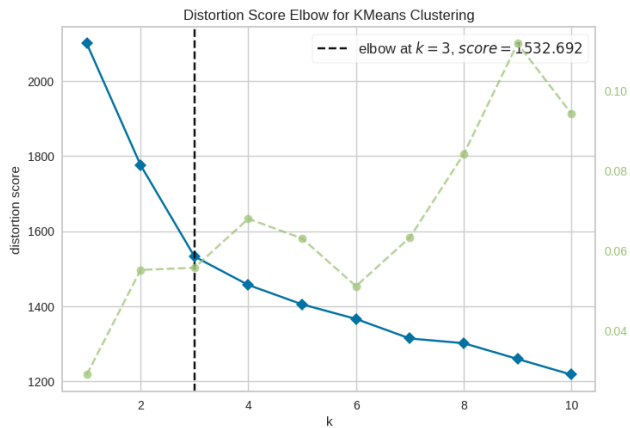


Figure 4.

A. Cluster Analysis

Our cluster analysis yielded three clusters, each characterized by a set of elective courses. For instance:

- Cluster 0 consists of courses like "BECOMING HUMAN", "ELEMENTARY ARABIC" suggesting a concentration area focused on Global and Historical Studies.
- Cluster 1 comprises courses like "COMPUTER SYSTEMS," "ALGORITHMS," and "DATA STRUCTURES," indicating a concentration area focused on Arts and Applied Sciences
- Cluster 2 includes courses like "HUMANITIES SEMINAR," "PHILOSOPHY," and "HISTORY," suggesting a concentration area focused on European and Political Studies.

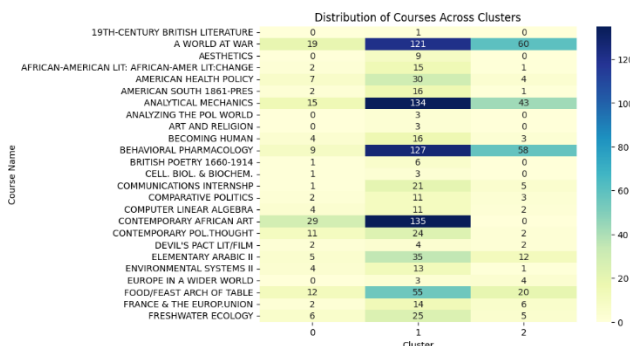


Figure 5.

VII. SUMMARY, DISCUSSIONS, AND LIMITATIONS

The overall performance of the data analysis was satisfactory, yielding actionable insights that can guide the curriculum redesign. The findings support the need to streamline elective courses and define concentration areas to make the program more focused and relevant to current student interests and industry trends. However, the analysis also highlighted the need for ongoing monitoring of course offerings to ensure they remain aligned with student needs and institutional goals.

A. Recommendations

1) *Reduction of Total Credit Hours*: It is recommended by research to decrease the total credit hours required for the BFA program from 39 to 33.[3] This reduction will streamline the curriculum while preserving its essential components. Focus should be on maintaining courses that align with student interests and current industry trends.

2) *Course Catalog Review*: A comprehensive review of the current course catalog should be undertaken to identify and remove outdated or irrelevant courses. This will ensure that the catalog reflects actual enrollment patterns and that all active courses are accurately listed. Courses with consistently high enrollment, such as 'A WORLD AT WAR,' should be considered as core courses in the program (See IV).

3) *Removal of Unmatched Courses*: Courses that were identified as unmatched between the catalog and enrollment data should be carefully evaluated. If they are no longer relevant or have not been offered recently, they should be removed from the program to streamline the curriculum (See IV).

4) *Elective Grouping into Concentrations*: Elective courses should be grouped into coherent concentration areas that reflect current disciplinary trends and student interests. This will provide students with clearer pathways for specialization, enhancing the program's appeal and relevance Global and Historical Studies, European and Political Studies, Arts and Applied Sciences (See VI).

B. Discussions

1) *Algorithm Evaluation*: One of the key metrics used to evaluate the performance of the clustering algorithm was the Silhouette Score, which was calculated to be 0.1036 on average.

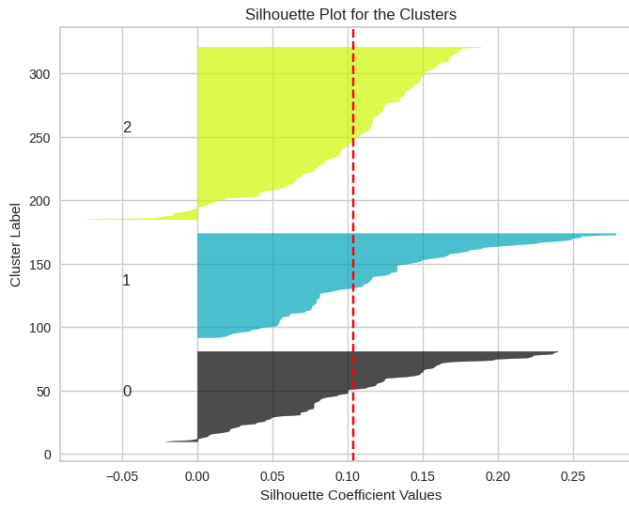


Figure 6.

As shown in Figure 6. Cluster 2 shows relatively better separation, while Clusters 0 and 1 exhibit significant overlap, indicating possible misclassification [2]. Further exploration into different algorithms may be needed.

2) *Anomalies*: A notable finding during the analysis was the occurrence of students registering for courses without a corresponding course number or name (See IV). This anomaly raises concerns about the accuracy and completeness of the enrollment records. Further research should be conducted to understand why students are registering for classes with missing course information

C. Limitations

One of the primary limitations was the quality and completeness of the data. The presence of missing course numbers and names in the registration data indicates potential issues with data entry or system errors. These anomalies limit the ability to conduct a fully comprehensive analysis, as they may obscure true patterns in student enrollment and course popularity. Additionally, the occurrence of unmatched courses between the catalog and enrollment records further complicates the analysis, as it raises questions about the accuracy and alignment of the data sources.

REFERENCES

- [1] S. Tripathi et al., "Ensuring the robustness and reliability of data-driven knowledge discovery models in production and manufacturing," *Frontiers in artificial intelligence*, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8236533/>
- [2] Selecting the number of clusters with silhouette analysis on KMeans clustering — scikit-learn 0.21.2 documentation," *Scikit-learn.org*, 2019
- [3] J. Bound, M. F. Lovenheim, and S. Turner, "Increasing Time to Baccalaureate Degree in the United States," *Educ Finance Policy*, Vol. 7, no. 4, 375-424, Oct. 2012.